Elementary Analytic Number Theory

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 [3] but they usually depend on abstract algebra and complex analysis heavily. In contrast, it can be argued that both topics are entirely excluded from this book except the very basics. Probably the heaviest result from complex numbers used here is that the product of absolute values of two complex numbers equals to the absolute value of the product of those complex numbers. 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 [47, Part Two, Chapter III] 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 [47, Part two, Chapter III, §3] and Apostol [3, Chapter VI]. Thus, we avoid the unnecessarily confusing and counter-intuitive calculation presented by Landau [47, Theorem 152] while still keeping the proof of Dirichlet's theorem completely elementary. The other proof is due to Selberg [57]. Both will be included in Chapter 3. We keep the calculation by Landau for complex characters though since it is quite intuitive and historically, this case has always been the easier one.
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 [12] and Friedlander and Iwaniec [29] 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 [29] 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.
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 [29, Chapter VI] 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 [46] 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 Turán's proof of a weaker version of a theorem on normal order by Ramanujan-Hardy which essentially contains *Turán'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 4.
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 not go with the spirit of this book. Similarly, I did not follow APOSTOL [3] and discuss Dirichlet characters from a more general point of view using group theory.
8) The reader may omit Chapter 6 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 [47], INGHAM [35] 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 6, A Mathematical Dispute of Twentieth Century which will shed light on why this was such a difficult task.

Masum Billal
4:03 AM, 13 May, 2022

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

Notations

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

$\operatorname{lcm}(a, b)$ 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 $(a_1, \ldots, a_k)$ such that $\gcd(a_1, \ldots, a_k, n) = 1$ and $1 \leq a_1, \ldots, a_k \leq n$.

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

$\lfloor x \rfloor$ Floor of x, greatest integer not exceeding x.

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

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

$\lambda(n)$ Liouville function of n, $\lambda(n) = (-1)^{\Omega(n)}$ or Carmichael's universal exponent function.

$\Lambda(n)$ 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.

$\Re(s)$ Real part of the complex number s.

$\Im(s)$ Imaginary part 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.

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

γ Euler-Mascheroni constant.

1 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.

Definition 1.1 (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.

Definition 1.2 (Summatory Function) For an arithmetic function f, the *summatory function* of f is defined as

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

where $\mathbb{S}$ is some set possibly dependent on n.

When $\mathbb{S}$ 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

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

Associated with this is the average order of f.

Definition 1.3 (Average Order) For an arithmetic function f,

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

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 [32].

Definition 1.4 (Normal Order) Let f and F be arithmetic functions such that

$$(1 - \epsilon)F(n) < f(n) < (1 + \epsilon)F(n) \tag{1.1}$$

holds for *almost* all $n \le x$ 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 (1.1) is very small compared to x.

An interesting property in summatory functions is that

$$\begin{aligned} \sum_{i=1}^{n} F(i) &= \sum_{i=1}^{n} \sum_{d \mid i} f(d) \\ &= \sum_{i=1}^{n} \left\lfloor \frac{n}{i} \right\rfloor f(i) \end{aligned}$$

Here, the last equation is true because there are $\lfloor n/i \rfloor$ multiples of i not exceeding n.

1.1 Order of Some Arithmetic Functions

Recall that the number of divisor function

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

We can generalize this as follows.

Definition 1.5 (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.

Definition 1.6 (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.

Definition 1.7 (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 $0 \leq \{x\} < 1$.

Definition 1.8 (Small O) Let f and g be two real or complex valued functions. Then the following two statements are equivalent

$$f(x) = o(g(x)) \tag{1.2}$$

$$\lim_{x \to \infty} \frac{f(x)}{g(x)} = 0 \tag{1.3}$$

Some trivial examples are $1/x = o(1)$, $x = o(x^2)$ and $2x^2 \neq o(x^2)$. LANDAU [42, Page 883 (second volume is paged consecutively after first volume)] states that the symbol O had been first used by BACHMANN [5, Page 401]. HARDY [31] uses the notations $\prec$ and $\succ$ respectively but they are no longer in practice. HARDY and RIESZ [33] 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,

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

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.

Definition 1.9 (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 (1.1) is $o(x)$. We can also say, the number of n satisfying (1.1) 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 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 1.10 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.

$$\begin{aligned} S_k(x) &= \sum_{n\leq x}\sum_{d|n} d^k \\ &= \sum_{n\leq x} \left\lfloor \frac{x}{n} \right\rfloor n^k \\ &= \sum_{n\leq x} \left(\frac{x}{n} + O(1)\right) n^k \\ &= x\sum_{n\leq x} n^{k-1} + O\left(\sum_{n\leq x} n^k\right) \end{aligned}$$

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

$$\begin{aligned} \sum_{n\leq x} n^k &\leq \sum_{n\leq x} x^k \\ &= x^k \sum_{n\leq x} 1 \\ &= \lfloor x \rfloor x^k \\ &= (x + O(1))x^k \\ &= x^{k+1} + O(x^k) \end{aligned}$$

We have that $S_k(x) = x\left(x^k + O(x^{k-1})\right) + O(x^{k+1}) = O(x^{k+1})$. 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 [54], for an English translation, see Knoebel et al. [40]). Lehmer [49, Chapter II, Theorem 1] proves that

$$\mathfrak{S}(x,k) = \frac{x^{k+1}}{k+1} + \Delta \tag{1.4}$$

where $|\Delta| \leq x^k$. The reader may also be interested in MacMillan and Sondow [50].

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

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

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

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

Beumer [7, (§8)] also considers the generalization $\tau_k(n)$ in this exact form. Using this for T,

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

Thus, we have the recursive result

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

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

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

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 [1] states this formula which today is a cornerstone of analytic number theory.

Theorem 1.11 (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

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

are known, then

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

In particular, if f is an arithmetic function,

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

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 [55, Page 83, §4] 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

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

with respect to $\pi(x), \phi(x)$ and an integral where $\pi(x)$ is the number of primes not exceeding x. A consequence of Theorem 1.11 is the celebrated *Euler's summation formula.*

Theorem 1.12 (Euler's summation formula) Let f be a continuous differentiable function in $[y, x]$. Then

$$\sum_{y<n\le x} f(n) = \int_y^x f(t)dt + \int_y^x \{t\}f(t)dt + \{y\}f(y) - \{x\}f(x)$$

where $\{t\} = t - \lfloor t \rfloor$ is the fractional part of t.

Proof. ■

As an application of Euler's summation formula, we can derive a result similar to (1.4) taking $f(n) = n^k$

for $k \geq 0$.

$$\begin{aligned}\mathfrak{S}_k(x) &= \sum_{n\leq x} n^k \\ &= \int\limits_1^x t^k dt + k\int\limits_1^x t^{k-1}(t-\lfloor t\rfloor)dt + 1 - (x-\lfloor x\rfloor)x^k \\ &= \frac{x^{k+1}}{k+1} - \frac{1}{k+1} + O\left(k\int\limits_1^x t^{k-1}dt\right) + O(x^k) \\ &= \frac{x^{k+1}}{k+1} + O(x^k)\end{aligned}$$

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

$$\sum_{n\leq x}\frac{\tau_{k-1}(n)}{n} = \frac{T_{k-1}(x)}{x} - \int\limits_1^x -\frac{T_{k-1}(t)}{t^2}dt$$

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

$$\sum_{n\leq x}\tau(n) = \sum_{n\leq x}\left\lfloor\frac{x}{n}\right\rfloor$$

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 $\lfloor x/a\rfloor$ possible choices for a valid value of b. So, the number of pairs in the first case is

$$\sum_{a\leq\sqrt{x}}\left\lfloor\frac{x}{a}\right\rfloor$$

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 $\lfloor x/b\rfloor - \sqrt{x}$ choices for a valid value of a, the choices namely are

$$\lfloor x\rfloor + 1, \ldots, \left\lfloor\frac{x}{b}\right\rfloor$$

Then the number of pairs in this case is

$$\sum_{b\leq\sqrt{x}}\left\lfloor\frac{x}{b}\right\rfloor - \lfloor\sqrt{x}\rfloor$$

Thus, the total number of such pairs is

$$\sum_{a\leq\sqrt{x}}\left\lfloor\frac{x}{a}\right\rfloor + \sum_{b\leq\sqrt{x}}\left(\left\lfloor\frac{x}{b}\right\rfloor - \lfloor\sqrt{x}\rfloor\right) = 2\sum_{n\leq\sqrt{x}}\left\lfloor\frac{x}{n}\right\rfloor - \lfloor x\rfloor^2 \tag{1.5}$$

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

$$\begin{aligned}\sum_{n\le\sqrt{x}} \lfloor x/n \rfloor &= \sum_{n\le\sqrt{x}} \left(\frac{x}{n} + O(1)\right)\\ &= x\sum_{n\le\sqrt{x}} \frac{1}{n} + O(\sqrt{x})\\ &= xH(\sqrt{x}) + O(\sqrt{x})\end{aligned}$$

Setting $a_n = 1$ and $f(n) = 1/n$ in Theorem 1.11, 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,

$$\begin{aligned}H(x) &= 1 + O\left(\frac{1}{x}\right) + \int_1^x \left(\frac{1}{t} + \frac{O(1)}{t^2}\right) dt\\ &= 1 + O\left(\frac{1}{x}\right) + \int_1^x \frac{1}{t}dt + O\left(\int_1^x \frac{1}{t^2}dt\right)\\ &= 1 + O\left(\frac{1}{x}\right) + \log x + O\left(1 - \frac{1}{x}\right)\end{aligned}$$

Thus, we have the following result.

Theorem 1.13 (Divergence of Harmonic Sum) For $x \ge 1$,

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

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 1.14 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 [22] (republished in EULER [27]) used C and O in his original paper. MASCHERONI [51] used A and a. Today it is not known whether γ is even irrational. For now, we will not require the use of γ, so we will use Theorem 1.13. Applying this, we have

$$\begin{aligned}
\sum_{n\le\sqrt{x}} \left\lfloor \frac{x}{n} \right\rfloor &= xH(\sqrt{x}) + O(\sqrt{x}) \\
&= x\left(C + \log\sqrt{x} + O\left(\frac{1}{\sqrt{x}}\right)\right) + O(\sqrt{x}) \\
&= \frac{1}{2}x\log x + Cx + O\left(\frac{x}{\sqrt{x}}\right) + O(\sqrt{x}) \\
&= \frac{1}{2}x\log x + O(x)
\end{aligned}$$

We can now use this to get

$$\begin{aligned}
\sum_{n\le x}\tau(n) &= 2\sum_{n\le\sqrt{x}} \lfloor x/n\rfloor - \lfloor\sqrt{x}\rfloor^2 \\
&= x\log x + O(x)
\end{aligned}$$

Thus, we get the following result.

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

DIRICHLET [18] actually proves the more precise result given below.

Theorem 1.15 (Dirichlet's average order of τ theorem)

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

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 [55] points out in his paper that the error term $O(1/\sqrt{x})$ in Dirichlet's theorem can be improved to $O\left(x^{-2/3+\epsilon}\right)$ or $O\left(x^{-2/3}\log x\right)$ as LANDAU [44, Page 689] shows.

We can now get back to estimating T. Using Theorem 1.11, we were able to deduce

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

Using Theorem 1.15, $T(x) = x \log x + O(x)$, so

$$\begin{aligned}
T_3(x) &= O(T(x)) + x \int_1^x \frac{T(t)}{t^2} dt \\
&= O(x \log x) + x \int_1^x \frac{\log t + O(1)}{t} dt \\
&= O(x \log x) + x \int_1^x \frac{\log t}{t} dt + xO\left(\int_1^x \frac{1}{t} dt\right) \\
&= O(x \log x) + x \int_1^x \frac{\log t}{t} dt
\end{aligned}$$

Using integration by parts,

$$\begin{aligned}
\int \frac{\log t}{t} dt &= \log t \int \frac{1}{t} - \int \left(\frac{1}{t} \int \frac{1}{t} dt\right) dt \\
&= \log^2 t - \int \frac{\log t}{t} dt
\end{aligned}$$

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 1.16 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 [45, Page 2] 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 $S_k(x)$.

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

Here, we can see that the function

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

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

Definition 1.17 (Zeta Function) For a complex number s, the zeta function $\zeta(s)$ is defined as

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

We will discuss zeta function in details in Section 1.2. For now, let us establish a result similar to Theorem 1.15 for partial sums of ζ. Setting $f(n) := n^{-s}$ and $a_n = 1$ in Theorem 1.11, $A(x) = \lfloor x \rfloor = x + O(1)$

and

$$\begin{aligned}
\sum_{n \le x} \frac{1}{n^s} &= \lfloor x \rfloor x^{-s} - \int_1^x (t + O(1)) f'(t) dt \\
&= x^{1-s} + O\left(x^{-s}\right) + s \int_1^x t^{-s} dt + O\left(s \int_1^x t^{-s-1} dt\right) \\
&= x^{1-s} + \frac{s}{1-s}\left(x^{1-s} - 1\right) + O\left(\int_1^x t^{-s-1} dt\right) \\
&= \frac{x^{1-s}}{1-s} + C + O(x^{-s})
\end{aligned}$$

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

Theorem 1.18 Let s be a positive real number other than 1. Then

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

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

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

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

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

$$\begin{aligned}
S_k(x) &= \frac{x^{k+1}}{k+1} \sum_{n \le x} \frac{1}{n^{k+1}} + O\left(x^k \sum_{n \le x} \frac{1}{n^k}\right) \\
&= \frac{x^{k+1}}{k+1}\left(\frac{x^{-k}}{-k} + \zeta(k+1) + O(x^{-k-1})\right) + O\left(x^k\left(\frac{x^{1-k}}{1-k} + \zeta(k) + O(x^{-k})\right)\right) \\
&= \frac{x}{-k(k+1)} + \frac{x^{k+1}}{k+1}\zeta(k+1) + O(x^{k+1-k-1}) + \left(\frac{x}{1-k} + x^k\zeta(k) + O(1)\right) \\
&= \frac{x^{k+1}}{k+1}\zeta(k+1) + O(x) + O(1) + O(x + x^k)
\end{aligned}$$

From this, we finally get the following.

Theorem 1.19 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 $\varphi(n)$.

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

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

Definition 1.20 (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.

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

Lehmer [49] used the notation $\varphi_k(n)$ but today $J_k(n)$ is used more often.

Jordan [39, Page 95 – 97] first discussed this function and Lehmer [49] 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 [49, Theorem VI] proves the following which he calls the *fundamental theorem.*

$$J_k(mn) = J_k(n)\prod_{p^e\|m}\left(p^{ke} - p^{k(e-1)}\lambda(n,p)\right) \tag{1.6}$$

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.

$$\sum_{d|n} J_k(d) = n^k \tag{1.7}$$

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.

1.2 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 [4] is a very good read on this subject. Euler's first contribution in this matter is EULER [21] 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 [23] (republished in EULER [28]) proves the following fundamental result which essentially gives a new proof of infinitude of primes.

Theorem 1.21 (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 [23] is the following which we shall prove later.

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

Here, $\sim$ is the asymptotic equivalence we have already defined. Even though Euler is the main architect behind the development of ζ, RIEMANN [56] 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.

Definition 1.22 (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 [33, §1, Page 1] considers the following as *general Dirichlet series*

$$\sum_{n \geq 1} a_n e^{-\lambda_n s} \tag{1.8}$$

where (λ_n) is a strictly increasing sequence of real numbers that tend to infinity. Following this, HARDY and RIESZ [33] calls $\mathfrak{D}$ the *ordinary Dirichlet series* when $\lambda_n = \log n$. DIRICHLET [15] considers real values of s and proves a number of important theorems. As Hardy states, JENSEN [37, 38] discusses the first theorems where s is complex involving the nature of convergence of 1.8. CAHEN [11] 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.*

Definition 1.23 (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

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

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

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

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.

Definition 1.24 (Dirichlet Convolution) For two arithmetic functions f and g, the *Dirichlet product* or *Dirichlet convolution* of f and g is defined as

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

Theorem 1.25 Let f and g be multiplicative arithmetic functions. Then $f * g$ is also multiplicative.

Proof. ■

Theorem 1.26 (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 [53].

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

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 1.27 (Möbius Inversion) Let f be an arithmetic function and F be the summatory function

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

Then

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

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,

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

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 [12, Page 4, Theorem 1.2.3], let us define *dual convolution*.

Definition 1.28 (Divisor Closed Set) A set of positive integers $\mathbb{S}$ is a *divisor closed set* if $d \mid n$, then $d \in \mathbb{S}$ holds for all $n \in \mathbb{S}$.

Definition 1.29 (Dual Convolution) Let f and g be arithmetic functions. Then the dual convolution of f and g is the arithmetic function h defined as

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

where $\mathbb{D}$ is a divisor closed set.

Theorem 1.30 (Dual Möbius Inversion) Let f be an arithmetic function and F be the summatory function

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

where $\mathbb{D}$ is a divisor closed set. Then

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

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

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

For a prime p, if $\nu_p(k) > 1$, then $\mathbb{D}_{nk} \in \mathbb{D}_{np}$, so we don't need to consider any of $F(nk)$ separately for

$nk \in \mathbb{D}$. We only need to consider the set of sets

$$\{\mathbb{D}_{nq} : q \mid P(n)\}$$

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

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

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 $\sum_{q|P(n)} f(nq)\mu(q)$ and

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

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

While discussing inversion, we should also mention Dirichlet inverse.

Definition 1.31 (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.

$$\begin{aligned} g(1) &= \frac{1}{f(1)} \\ g(n) &= -\frac{1}{f(1)} \sum_{\substack{d|n \\ d<n}} f\left(\frac{n}{d}\right) g(d) \end{aligned}$$

Haukkanen [34, Theorem 2.2] proves the following closed formula to find the Dirichlet inverse of an arithmetic function f which we do not prove here.

Theorem 1.32 Let f be an arithmetic function such that $f(1) = 1$. Then the Dirichlet inverse of f is

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

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

1.3 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

$$\sum_{n\le x}\tau(n)=2\sum_{n\le\sqrt{x}}\left\lfloor\frac{x}{n}\right\rfloor-\left\lfloor\sqrt{x}\right\rfloor^2$$

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

$$\sum_{n\le x}\sigma(n)=\frac{1}{2}\left(\sum_{n\le\sqrt{x}}\left\lfloor\frac{x}{n}\right\rfloor+\sum_{n\le\sqrt{x}}(2n+1)\left\lfloor\frac{x}{n}\right\rfloor-\left\lfloor\sqrt{x}\right\rfloor^2-\left\lfloor\sqrt{x}\right\rfloor^3\right)$$

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

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

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

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

Specially when $a=b$,

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

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

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

We can easily prove the following.

Theorem 1.34 (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 1.35 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

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

1.4 A Variation of Generalized Convolution

We will now consider a slight variation of generalized convolution.

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

Theorem 1.36 Let f and g be arithmetic functions. Then

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

We can also write it as

$$(f \diamond g)(x) = \sum_{n \leq \sqrt{x}} g(n) \left(F\left(\left\lfloor \frac{x}{n} \right\rfloor\right) - F\left(\left\lfloor \frac{x}{n+1} \right\rfloor\right)\right) + f(n) g\left(\left\lfloor \frac{x}{n} \right\rfloor\right) - \mathfrak{B}(x)$$

where

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

Proof. Consider the integers

$$\left\lfloor \frac{x}{1} \right\rfloor, \ldots, \left\lfloor \frac{x}{\lfloor x \rfloor} \right\rfloor$$

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

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

$$\begin{aligned} F(x) &= \sum_{n \le x} f(n) \\ &= O(x^{\xi}) \\ g(x) &= \lfloor x \rfloor^k \end{aligned}$$

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

$$\begin{aligned} (f \diamond g)(x) &= \sum_{n \le x} f(n) \left\lfloor \frac{x}{n} \right\rfloor^k \\ &= \sum_{n \le \sqrt{x}} n^k \left(F\left(\left\lfloor \frac{x}{n} \right\rfloor \right) - F\left(\left\lfloor \frac{x}{n+1} \right\rfloor \right) \right) + f(n) \left\lfloor \frac{x}{n} \right\rfloor^k - \mathfrak{B}(x) \\ &= \sum_{n \le \sqrt{x}} n^k O\left(\left\lfloor \frac{x}{n} \right\rfloor^{\xi} \right) + f(n) \left\lfloor \frac{x}{n} \right\rfloor^k - \mathfrak{B}(x) \\ &= x^{\xi} O\left(\sum_{n \le \sqrt{x}} \frac{n^k}{n^{\xi}} \right) + \sum_{n \le \sqrt{x}} f(n) \left(\left(\frac{x}{n} \right) + O(1) \right)^k - \mathfrak{B}(x) \\ &= x^{\xi} O\left(\sum_{n \le \sqrt{x}} \frac{n^k}{n^{\xi}} \right) + \sum_{n \le \sqrt{x}} f(n) \left(\left(\frac{x}{n} \right)^k + O\left(\left(\frac{x}{n} \right)^{k-1} \right) \right) - \mathfrak{B}(x) \\ &= x^{\xi} O\left(\sum_{n \le \sqrt{x}} \frac{n^k}{n^{\xi}} \right) + x^k \sum_{n \le \sqrt{x}} \frac{f(n)}{n^k} + O\left(x^{k-1} \sum_{n \le \sqrt{x}} \frac{f(n)}{n^{k-1}} \right) - \mathfrak{B}(x) \end{aligned}$$

Now we need to focus on the following two sums.

$$\begin{aligned} \mathfrak{M}_s(x) &= \sum_{n \le x} \frac{n^s}{n^{\xi}} \\ \mathfrak{G}_s(x) &= \sum_{n \le x} \frac{f(n)}{n^s} \end{aligned}$$

where $s \ge 1$. Then

$$(f \diamond g)(x) = x^{\xi} O\left(\mathfrak{M}_k(\sqrt{x}) \right) + x^k \mathfrak{G}_k(\sqrt{x}) + O\left(x^{k-1} \mathfrak{G}_{k-1}(\sqrt{x}) \right) - \mathfrak{B}(x) \tag{1.9}$$

We can use Theorem 1.18 for computing $\mathfrak{M}$. If $s+1 < \xi$,

$$\begin{aligned} \mathfrak{M}_s(x) &= \sum_{n \le x} \frac{1}{n^{\xi - s}} \\ &= \frac{x^{1+s-\xi}}{1+s-\xi} + \zeta(\xi - s) + O\left(x^{s-\xi} \right) \end{aligned} \tag{1.10}$$

If $s+1=\xi$, from Theorem 1.13,

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

The case $\xi - 1 < s < \xi$ is possible if and only if ξ is not an integer and $s = \lfloor \xi \rfloor$ which can be taken care of in the same manner as (1.10). We can now assume $s \geq \xi$. In this case, $s - \xi \geq 0$ and from Theorem 1.19

$$\begin{aligned}\mathfrak{M}_s(x) &= \sum_{n\leq x} n^{s-\xi} \\ &= \frac{x^{s-\xi+1}}{s-\xi+1} + O\left(x^{s-\xi}\right)\end{aligned}$$

For handling $\mathfrak{G}$, we will use Theorem 1.11.

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

Thus, we have

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

Plugging these back in (1.9), we get the following result.

Theorem 1.37 Let f and g be arithmetic functions such that $F(x) = \sum_{n\leq x} f(n) = O\left(x^{\xi}\right), g(x) = \lfloor x \rfloor^k$. Then

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

1.5 Generalization of General Convolution

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

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

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

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 $(x_1/a_1, \dots, x_k/a_1)$, $\lfloor \mathbf{x}/\mathbf{a} \rfloor$ denote the vector $(\lfloor x_1/a_1 \rfloor, \dots, \lfloor x_k/a_k \rfloor)$ and

$$\begin{aligned} f(\mathbf{x}) &= f(x_1, \dots, x_k) \\ f\left(\frac{\mathbf{x}}{n^{\mathbf{a}}}\right) &= f\left(\frac{x_1}{n^{a_1}}, \dots, \frac{x_k}{n^{a_k}}\right) \\ f\left(\left\lfloor \frac{\mathbf{x}}{n^{\mathbf{a}}} \right\rfloor\right) &= f\left(\left\lfloor \frac{x_1}{n^{a_1}} \right\rfloor, \dots, \left\lfloor \frac{x_k}{n^{a_k}} \right\rfloor\right) \end{aligned}$$

Definition 1.38 (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

$$(\alpha \bullet f)(\mathbf{x}, \mathbf{a}) = \sum_{n^{\mathbf{a}} \le \mathbf{x}} \alpha(n)\, f\left(\frac{\mathbf{x}}{n^{\mathbf{a}}}\right) \tag{1.11}$$

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

Proposition 1.39 (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,

$$\begin{aligned} (\beta \bullet f)(\mathbf{x}, \mathbf{a}) &= \sum_{m^{\mathbf{a}} \le \mathbf{x}} \beta(m) \left(\frac{x_1}{m^{a_1}}, \dots, \frac{x_k}{m^{a_k}}\right) \\ (\alpha \bullet (\beta \bullet f))(\mathbf{x}, \mathbf{a}) &= \sum_{n^{\mathbf{a}} \le \mathbf{x}} \alpha(n) \left((\beta \bullet f)\left(\frac{\mathbf{x}}{n^{\mathbf{a}}}, \mathbf{a}\right)\right) \\ &= \sum_{n^{\mathbf{a}} \le \mathbf{x}} \alpha(n) \left((\beta \bullet f)\left(\frac{x_1}{n^{a_1}}, \dots, \frac{x_k}{n^{a_k}}\right)\right) \\ &= \sum_{n^{\mathbf{a}} \le \mathbf{x}} \alpha(n) \sum_{m^a \le \mathbf{x}/n^{\mathbf{a}}} \beta(m) f\left(\frac{x_1}{m^a n^a}, \dots, \frac{x_k}{m^a n^a}\right) \end{aligned}$$

We can collect the m and n together and write

$$\begin{aligned}
(\alpha \bullet (\beta \bullet f))(\mathbf{x}, \mathbf{a}) &= \sum_{(mn)^{\mathbf{a}} \leq \mathbf{x}} \alpha(n)\beta(m) f\left(\frac{x_1}{m^a n^a}, \ldots, \frac{x_k}{m^{a_k} n^{a_k}}\right) \\
&= \sum_{n^{\mathbf{a}} \leq \mathbf{x}} \left(\sum_{d|n} \alpha(d)\beta\left(\frac{n}{d}\right)\right) f\left(\frac{x_1}{n^{a_1}}, \ldots, \frac{x_k}{n^{a_k}}\right) \\
&= \sum_{n^{\mathbf{a}} \leq \mathbf{x}} (\alpha * \beta) f\left(\frac{\mathbf{x}}{n^{\mathbf{a}}}\right) \\
&= (\alpha * \beta) \bullet f(\mathbf{x}, \mathbf{a})
\end{aligned}$$

■

Proposition 1.40 (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. ■

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.

Proposition 1.41 Let f and g be arithmetic functions and $h = f * g$. If

$$F(\mathbf{x}, \mathbf{a}) = \sum_{n^{\mathbf{a}} \le \mathbf{x}} f(n)$$
$$G(\mathbf{x}, \mathbf{a}) = \sum_{n^{\mathbf{a}} \le \mathbf{x}} g(n)$$
$$H(\mathbf{x}, \mathbf{a}) = \sum_{n^{\mathbf{a}} \le \mathbf{x}} h(n)$$

then we have

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

Proof. We can write H as follows.

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

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

As a corollary, we have the following theorem.

Proposition 1.42 Let f be an arithmetic function. If

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

then we have

$$\sum_{n\le \sqrt[\mathbf{a}]{\mathbf{x}}} \sum_{d|n} f(d) = \sum_{n\le \sqrt[\mathbf{a}]{\mathbf{x}}} \left\lfloor \frac{\sqrt[\mathbf{a}]{\mathbf{x}}}{n} \right\rfloor f(n)$$
$$= \sum_{n^{\mathbf{a}}\le \mathbf{x}} F\left(\frac{\sqrt[\mathbf{a}]{\mathbf{x}}}{n}\right)$$

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

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

Define the function u as

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

From (1.7),

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

Using Möbius Inversion, we also get that

$$\mu * u = J \tag{1.12}$$

Let $F(\mathbf{x})$ be the number of vectors of positive integers $(a_1, \ldots, a_k)$ such that $1 \le a_i \le x_i$ and $\gcd(a_1, \ldots, a_k) = 1$. 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 $(a_1, \ldots, a_k)$. If $g = \gcd(a_1, \ldots, a_k) > 1$, then every a_i has to be divisible by g. Then the number of such vectors is

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

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

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

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

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

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 $(a_1,\ldots,a_k)$ 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 $(a_1,\ldots,a_k)$ such that $1\le a_1\le\ldots\le a_k\le n$ and $\gcd(a_1,\ldots,a_k)=1$. If for a positive integer m, $\mathbf{m}=(m,\ldots,m)$ 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

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

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

$$S(\mathbf{x}) = \sum_{n\le\mathbf{x}} J_s(n)\prod_{i=1}^{k}\left\lfloor\frac{x_i}{n}\right\rfloor \tag{1.13}$$

$$= \sum_{n\le\mathbf{x}} \mu(n)\left(\sum_{i\le \mathbf{x}/n} i^s\prod_{j=1}^{k}\left\lfloor\frac{x_j}{ni}\right\rfloor\right) \tag{1.14}$$

(1.13) follows from Associativity of Generalized Convolution and (1.12).

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

So, we will only prove (1.14). 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 $(\mu\bullet f)(\mathbf{x},\mathbf{1})$. Each of these vectors contribute g^s to the sum, so for a particular g, the contribution of g in the sum is

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

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

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

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

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

Tchebyscheff's Theorems | 2

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 [43, Page 37]) around 1792 or 1793 that

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

TCHEBYSCHEFF [60] is the first one to make any substantial progress on the matter. In fact, Tchebyscheff was close to proving the prime number theorem himself. TCHEBYSCHEFF [59] 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

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

exists, then it is 1. The only problem was to actually prove that this limit indeed exists. One of our objectives in this book is prove the prime number theorem without any serious analysis from the scratch. We will discuss some relevant results first that give us better insight into the structure of primes before doing that. EULER [20] proved this first although his method was a bit questionable.

Theorem 2.1 (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 inspired by LANDAU [47, Theorem 114].

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

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

and that this sum diverges. Now,

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

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

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

we have

$$\begin{aligned}
\log\left(\sum_{n\geq 1} \frac{1}{n}\right) &= \sum_p \left(\eta + \frac{\eta^2}{2} + \frac{\eta^3}{3} + \dots\right) \\
&< \sum_p \left(\eta + \eta^2 + \dots\right) \\
&= \sum_p \frac{\eta}{1-\eta} \\
&< 2\sum_p \eta \\
&= 2\sum_p \frac{1}{p}
\end{aligned}$$

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

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

for a constant C. Thus,

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

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

MERTENS [52] actually proved that

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

This is the first formal proof since Euler's method wasn't exactly clean. EULER [24, Page 228] uses

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

and sets $x := 1$ to conclude

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

This is indeed true, however, Euler's statement is vague. So this is not usually considered a rigorous proof of this result.

Theorem 2.2 If $x \to \infty$, then $\pi(x) = O\left(\frac{x}{\log\log x}\right)$. A weaker statement is $\pi(x) = o(x)$.

Proof. Let ξ be a real number such that $2 < \xi < x$ and the primes not exceeding ξ are $p_1, \ldots, p_k$. The number of positive integers not divisible by any of $p_1, \ldots, p_k$ in the interval $[\xi + 1, \ldots, x]$ is

$$\varphi(x,\xi) = x - \sum\left\lfloor\frac{x}{p_i}\right\rfloor + \sum_{i<j}\left\lfloor\frac{x}{p_ip_j}\right\rfloor - \ldots$$

If $\xi = \sqrt{x}$, then we have $\varphi(x,\xi) = \pi(x) - k + 1$. In general, $\pi(x) - r + 1 \leq \varphi(x,\xi)$ holds. Now,

$$\begin{aligned}\varphi(x,\xi) &\leq x - \sum\frac{x}{p_i} + 1 + \sum\frac{x}{p_ip_j} + 1 - \ldots \\ &= x\prod_{i=1}^{k}\left(1 - \frac{1}{p_i}\right) + 2^k \\ &< x\prod_{p\leq\xi}\left(1 - \frac{1}{p}\right) + 2^\xi\end{aligned}$$

Note that the choice of ξ is arbitrary and we can easily choose $2^\xi + r - 1 = o(x)$ or $\xi = c\log x$ for some constant c. Also,

$$\begin{aligned}\prod_{p\le x}\left(1-\frac{1}{p}\right)^{-1} &= \prod_{p\le x}\left(1+\frac{1}{p}+\frac{1}{p^2}+\dots\right)\\ &> \sum_{n\le x}\frac{1}{n}\\ &> \log x\end{aligned}$$

so

$$\begin{aligned}\prod_{p\le\xi}\left(1-\frac{1}{p}\right) &< \frac{1}{\log c\log x}\\ &= \frac{1}{\log c+\log\log x}\\ &< \frac{1}{\log\log x}\end{aligned}$$

Therefore,

$$\begin{aligned}\pi(x) &\le \varphi(x,\xi)+r-1\\ &< \frac{x}{\log\log x}+o(x)\\ \pi(x) &= O\left(\frac{x}{\log\log x}\right)\end{aligned}$$

and evidently, $\pi(x)=o(x)$ since

$$\lim_{x\to\infty}\frac{\frac{x}{\log\log x}}{x}=0$$

■

We can now start proving some results by Tchebyscheff [60] and Tchebyscheff [59]. Before we discuss Tchebyscheff's functions, we should discuss the following first which gives us some insight into why Tchebyscheff's functions are important.

Theorem 2.3 (Tchebyscheff) Let x be a positive real number. Then there are constants a and A such that

$$a\frac{x}{\log x}\le\pi(x)\le A\frac{x}{\log x}$$

Actually, Tchebyscheff gave a more precise statement that

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

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 [47, Theorem 112]; which is a translation of the first section of the first volume of Landau [46].

Proof. For any $x \geq 0$,

$$\lfloor x \rfloor - 2\left\lfloor \frac{x}{2} \right\rfloor \leq 1$$

since

$$\begin{aligned} \lfloor x \rfloor - 2\left\lfloor \frac{x}{2} \right\rfloor &< x - 2\left(\frac{x}{2} - 1\right) \\ &= 2 \end{aligned}$$

Let $n \geq 2$. For every $p \leq 2n$, let r denote the largest positive integer such that $p^r \leq 2n$ (which is $\lfloor \log 2n / \log p \rfloor$). We will first show that

$$\prod_{n<p\leq 2n} p \mid \mathfrak{N} \tag{2.1}$$

$$\mathfrak{N} \mid \prod_{p\leq 2n} p^r \tag{2.2}$$

where $\mathfrak{N} = \binom{2n}{n}$ and the product runs through the primes only. For a prime p,

$$\nu_p(\mathfrak{N}) = \nu_p((2n)!) - 2\nu_p(n!)$$

Since $p > n$, $\nu_p(n!) = 0$ and $\nu_p(\mathfrak{N}) = \nu_p((2n)!) \geq \nu_p\left(\prod_{n<p\leq 2n} p\right)$. On the other hand, for $p \leq 2n$, using Legendre's formula,

$$\begin{aligned} \nu_p(\mathfrak{N}) &= \nu_p((2n)!) - 2\nu_p(n!) \\ &= \sum_{i\geq 1} \left\lfloor \frac{2n}{p^i} \right\rfloor - 2\left\lfloor \frac{n}{p^i} \right\rfloor \end{aligned}$$

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

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

Thus, for every prime p,

$$p^{\nu_p(\mathfrak{N})} \mid p^{\lfloor \log_p 2n \rfloor}$$

and evidently

$$\mathfrak{N} \mid \prod_{p \le 2n} p^{\lfloor \log_p 2n \rfloor}$$

So $\mathfrak{N} \le \prod_{p \le 2n} 2n$. Now, there are $\pi(2n) - \pi(n)$ primes in the interval $(n, 2n)$ each of which are greater than n. So,

$$\begin{aligned} n^{\pi(2n)-\pi(n)} &\le \prod_{n<p\le 2n} p \\ &\le \mathfrak{N} \\ &\le \prod_{p \le 2n} 2n \\ &\le (2n)^{\pi(2n)} \end{aligned}$$

Taking logarithm,

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

Next, we have that

$$\begin{aligned} \mathfrak{N} &\le \sum_{i=0}^{2n} \binom{2n}{i} \\ &= 2^{2n} \\ \binom{2n}{n} &= \frac{(n+1)\cdots(2n)}{1\cdots n} \\ &\ge \prod_{i=1}^{n} \frac{n+i}{i} \\ &\ge 2^n \end{aligned}$$

giving

$$(\pi(2n) - \pi(n)) \log n \le \log \mathfrak{N} \le \pi(2n) \log 2n$$

Thus, we have

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

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 $x \geq 2$, setting $\eta = \lfloor x/2 \rfloor$ and $x = 2\eta + r$ with $0 \leq r < 2$, we have $x \leq 3\eta$ and $\log 2\eta \leq \log x$.

$$\begin{aligned}\pi(x) &\geq \ \pi\left(2\eta\right)\\ &\geq c\frac{2\eta}{\log 2\eta}\\ &= \frac{2c}{3}\frac{3\eta}{\log 2\eta}\\ &\geq a\frac{x}{\log x}\end{aligned}$$

for a positive constant a. On the other hand, since $x < 2 + 2\left\lfloor \frac{x}{2} \right\rfloor$,

$$\begin{aligned}\pi(x) - \ \pi\left(\frac{x}{2}\right) &= \pi(x) - \ \pi\left(\left\lfloor \frac{x}{2} \right\rfloor\right)\\ &\leq 2 + \ \pi\left(2\left\lfloor \frac{x}{2} \right\rfloor\right) - \ \pi\left(\left\lfloor \frac{x}{2} \right\rfloor\right)\\ &\leq 2 + 2\log 2\frac{\left\lfloor \frac{x}{2} \right\rfloor}{\log\left\lfloor \frac{x}{2} \right\rfloor}\\ &< d\frac{x}{logx}\end{aligned}$$

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

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

If $2^{m+1} \leq 2^{v+1}x < 2^{v+2}$ (that is, $v + 1 = \lfloor \log_2 x \rfloor$),

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

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

$$\begin{aligned}
\log x\pi(x) &= \sum_{m=0}^{v} \left(\log \frac{x}{2^m}\, \pi\left(\frac{x}{2^m}\right) - \log \frac{x}{2^{m+1}}\, \pi\left(\frac{x}{2^{m+1}}\right)\right) \\
&< d_2 x \sum_{m=0}^{v} \frac{1}{2^m} \\
&= Ax
\end{aligned}$$

■

A corollary is the following.

Theorem 2.4 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 **??**. Since

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

we only need to show that

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

diverges. I leave this as an exercise as well.

2.1 Tchebyscheff Functions

Tchebyscheff [59] defines two functions that are today known as Tchebisceff's ϑ and ψ functions.

Definition 2.5 (Tchebyscheff's ϑ function) For a real number x, Tchebyscheff's ϑ function is defined as

$$\vartheta(x) = \sum_{p \leq x} \log p$$

Definition 2.6 (Tchebyscheff's ψ function) For a real number x, Tchebyscheff's ψ function is defined as

$$\psi(x) = \sum_{p^i \leq n} \log p$$

Note that we can write ψ in different ways.

$$\begin{aligned}\psi(x) &= \sum_{p \le x} \log p \sum_{p^i \le x} 1 \\ &= \sum_{p \le x} \lfloor \log_p x \rfloor \log p\end{aligned}$$

Also,

$$\begin{aligned}\psi(x) &= \sum_{i \ge 1} \sum_{p^i \le x} \log p \\ &= \sum_{i \ge 1} \sum_{p \le \sqrt[i]{x}} \log p \\ &= \sum_{i \ge 1} \vartheta(\sqrt[i]{x})\end{aligned}$$

At this point, we should introduce the Von Mangoldt function.

Definition 2.7 (Von Mangoldt function) For a positive integer n,

$$\begin{aligned}\Lambda(n) &= \log p \text{ if } n = p^k \text{ for a prime } p \\ &= 0 \text{ otherwise}\end{aligned}$$

Then we can write ψ as

$$\psi(x) = \sum_{n \le x} \Lambda(n)$$

The following theorem immediately shows us the importance of these functions.

Theorem 2.8 If any of the three limits

$$\lim_{x\to\infty} \frac{\pi(x)}{\frac{x}{\log x}}, \lim_{x\to\infty} \frac{\vartheta(x)}{x}, \lim_{x\to\infty} \frac{\psi(x)}{x}$$

exists, then all three limits are equal.

Proof. Let the upper limits of the three be l_1, l_2, l_3 respectively. Obviously,

$$\begin{aligned}\vartheta(x) &\leq \psi(x) \\ &= \sum_{p\leq x} \left\lfloor \frac{\log x}{\log p} \right\rfloor \log p \\ &\leq \sum_{p\leq x} \frac{\log x}{\log p} \log p \\ &= \log x \sum_{p\leq x} 1 \\ &= \pi(x) \log x\end{aligned}$$

Then, for all x,

$$\frac{\vartheta(x)}{x} \leq \frac{\psi(x)}{x} \leq \frac{\pi(x)}{\frac{x}{\log x}}$$

So, $l_2 \leq l_3 \leq l_1$. Now, for any $0 < \epsilon < 1$, $\vartheta(x^\epsilon) \geq 0$ so $\vartheta(x) - \vartheta(x^\epsilon) \leq \vartheta(x)$.

$$\begin{aligned}\vartheta(x) - \vartheta(x^\epsilon) &= \sum_{x^\epsilon < p \leq x} \log p \\ &> \sum_{x^\epsilon < p \leq x} \log x^\epsilon \\ &= \log x^\epsilon \sum_{x^\epsilon < p \leq x} 1 \\ &= \epsilon(\pi(x) - \pi(x^\epsilon)) \log x\end{aligned}$$

Here, $\pi(x^\epsilon) < x^\epsilon$ so $\pi(x) - \pi(x^\epsilon) \geq \pi(x) - x^\epsilon$ and

$$\frac{\vartheta(x)}{x} > \epsilon \left(\frac{\pi(x)}{\frac{x}{\log x}} - \frac{\log x}{x^{1-\epsilon}} \right)$$

Taking $x \to \infty$, $\lim_{x\to\infty} \frac{\log x}{x^{1-\epsilon}} = 0$ so

$$\lim_{x\to\infty} \frac{\vartheta(x)}{x} \geq \epsilon \frac{\pi(x)}{\frac{x}{\log x}}$$

Therefore, $l_2 \geq l_1$ so $l_1 = l_2$ and so $l_3 = l_2 = l_1$. ∎

ϑ and ψ are discussed in every book on analytic number theory, however, it is never discussed why Tchebyscheff would consider these functions to begin with. Like most notable mathematical discovery, this was not a blind attempt by Tchebyscheff and he did not suddenly receive divine knowledge one night either. So there must be some explanation of how he thought of these functions and why they are so crucial in the study of prime numbers. Initially, I wanted to discuss a rationalization how Tchebyscheff might have thought of them. But I think it is more appropriate if we leave it as an open question to the reader to come up with such a rationalization how we might come up with such functions if we were to prove the prime number theorem. INGHAM [35, pp. 13] says the following about this matter:

> It happens (as will appear more clearly in §7) that, of the three functions π, ϑ, ψ, the one which arises most naturally from the analytical point of view is the most remote from the original problem, namely ψ. For this reason, it is usually most convenient to work in the first instance with ψ and to use Theorem 3 (or more precise relations corresponding to the degree of approximation contemplated) to deduce the results about π. This is a complication which seems inherent in the subject, and the reader should familiarize himself at the outset with the function ψ, which is to be regarded as the fundamental one.

One of the goals of Tchebyscheff's work was to prove a postulate by BERTRAND [6].

Conjecture 2.9 (Bertrand's postulate) For any real number $x > 1$, there is a prime p such that $n < p < 2n$.

Note that Theorem 2.3 already implies Bertrand's postulate if we can show that $A \leq 2a$. Because then $\vartheta\left(\frac{Ax}{a}\right) > \vartheta(x)$ so there must be a prime between x and cx for some constant $c \leq 2$. However, we will show that the direct approach in the proof of Theorem 2.3 does not produce a proof for this postulate. With the help of Theorem 2.8, it is enough to show that there exist constants a and A such that $A \leq 2a$ and

$$ax < \vartheta(x) < Ax$$

We again consider the binomial coefficient $\mathfrak{N} = \binom{2n}{n} < 2^{2n}$ and use the trivial fact $\mathfrak{N} < 2^{2n}$. If p is a prime such that $n < p < 2n$, then

$$\begin{aligned} \nu_p(\mathfrak{N}) &= \nu_p((2n)!) - 2\nu_p(n!) \\ &= \nu_p((2n)!) \end{aligned}$$

So, $\mathfrak{N}$ is divisible by all such primes p. Thus,

$$\mathfrak{N} \geq \prod_{n<p<2n} p$$

Since $\log \mathfrak{N} < 2n \log 2$,

$$\begin{aligned} 2n \log 2 &> \sum_{n<p<2n} \log p \\ &= \vartheta(2n) - \vartheta(n) \end{aligned}$$

The right side can be telescoped by setting $n := 2^i$ for $0 \leq i \leq k-1$.

$$\sum_{i=0}^{k-1} 2^{i+1} \log 2 > \vartheta(2^k)$$

Since $2^k > 1 + \ldots + 2^{k-1}$,

$$2^{k+1} \log 2 > \vartheta(2^k)$$

For any $x > 1$, taking $2^{k-1} \le x < 2^k$, $2^{k+1} \le 4x$ so $\vartheta(x) \le \vartheta(2^k) < 4x \log 2$ and $\vartheta(x) < Ax$ for some $A \le 4 \log 2$.

Again, for a prime p, letting r be the largest positive integer such that $p^r \le 2n$, similar to the proof of Theorem 2.3, $\nu_p(\mathfrak{N}) \le r$. From the definition,

$$e^{\psi(2n)} = \prod_{p \le 2n} p^r$$

so $\mathfrak{N} \mid e^{\psi(2n)}$. Also, $(2n+1)\mathfrak{N} \ge 2^{2n}$ so

$$(2n+1)e^{\psi(2n)} \ge 2^{2n}$$

Taking logarithm,

$$\log(2n+1) + \psi(2n) \ge 2n \log 2$$

Letting $\left\lfloor \frac{x}{2} \right\rfloor = n$, we have $\psi(x) \ge \psi(2n)$, $2n > x - 2$ and

$$\psi(x) \ge (x-1)\log 2 - \log(x+1)$$

We can now show that $\psi(x) \ge ax$ for some $a \ge \log 2$. However, this only gives us the bound $\frac{A}{a} \le 4$ which does not prove Bertrand's postulate. Tchebyscheff [59, §4, eqn. (5) pp. 376] showed that $A \le \frac{6}{5}a$ where

$$a \ge \log \frac{2^{\frac{1}{2}} 3^{\frac{1}{3}} 5^{\frac{1}{5}}}{30^{\frac{1}{30}}}$$

It was Erdős [19] who introduced himself to the mathematical world by proving Bertrand's postulate in a completely elementary manner using only properties of binomial coefficients. The reader can consult Aigner and Ziegler [2, §2] for an English translation. We will use some ideas we have already discussed on the matter.

Proof. The crucial idea behind Erdős's proof was to show that if there was no prime between n and $2n$, then $\mathfrak{N} = \binom{2n}{n}$ would not be as large as needs be.

In order to show that, one of the first results Erdős uses is for any odd prime p such that $\frac{2}{3}n < p < n$, $2n < 3p$ so, $\nu_p((2n)!) = 2$ and $\nu_p(n!) = 1$. Therefore, $\nu_p(\mathfrak{N}) = \nu_p((2n)!) - 2\nu_p(n!) = 0$. In a similar manner, if $n+1 < p \le 2n+1$, then $\nu_p(\mathfrak{N}) = 1$. So,

$$\prod_{n+1 < p \le 2n+1} p \mid \binom{2n+1}{n}$$

In a similar fashion as above,

$$\begin{aligned}
(1+1)^{2n+1} &= \sum_{i=0}^{2n+1} \binom{2n+1}{i} \\
&\geq \binom{2n+1}{n} + \binom{2n+1}{n+1} \\
&\geq 2\binom{2n+1}{n}
\end{aligned}$$

Then $\binom{2n+1}{n} \leq 2^{2n}$. Erdős uses the last two facts to establish another elementary result: the product of primes not exceeding n does not exceed 4^n. Induction is the easiest way to prove this. For very small n, say $n \leq 10$, the result is obvious. When $n > 10$, if n is even, then it cannot be prime so

$$\begin{aligned}
\prod_{p \leq n} p &= \prod_{p \leq n-1} p \\
&\leq 4^{n-1} \\
&< 4^n
\end{aligned}$$

When n is odd,

$$\begin{aligned}
\prod_{p \leq 2m+1} p &= \left(\prod_{p \leq m+1} p\right) \cdot \left(\prod_{m+2 \leq p \leq 2m+1} p\right) \\
&\leq 4^{m+1} \cdot \binom{2m+1}{m} \\
&\leq 4^{m+1} \cdot 2^{2m} \\
&= 4^{2m+1}
\end{aligned}$$

We have already shown that

$$p^{\nu_p(\mathfrak{N})} \leq 2n$$

Finally, we can divide the factorization of $\mathfrak{N}$ using primes not exceeding $2n$ the following way

$$\begin{aligned}
\mathfrak{N} &= \left(\prod_{p \leq \sqrt{2n}} p^{\nu_p(\mathfrak{N})}\right) \cdot \left(\prod_{\sqrt{2n} < n \leq \frac{2n}{3}} p\right) \cdot \left(\prod_{\frac{2n}{3} < p < n} p\right) \cdot \left(\prod_{n \leq p \leq 2n} p\right) \\
&\leq \left(\prod_{p \leq \sqrt{2n}} 2n\right) \cdot \left(\prod_{\sqrt{2n} < p \leq \frac{2n}{3}} p\right) \cdot \left(\prod_{n \leq p \leq 2n} p\right)
\end{aligned}$$

since no prime in the region $(\frac{2n}{3}, n)$ divide $\mathfrak{N}$. If there is no prime between n and $2n$ either, then we have

$$\mathfrak{N} \leq (2n)^{\sqrt{2n}} \cdot 4^{\frac{2n}{3}}$$

because $\pi(\sqrt{2n}) \leq \sqrt{2n}$. However, $\mathfrak{N} \geq \frac{4^n}{2n}$ so

$$4^n \leq (2n)^{1+\sqrt{2n}} 4^{\frac{2n}{3}}$$
$$4^{\frac{n}{3}} \leq (2n)^{1+\sqrt{2n}}$$

We can easily see $4^{\frac{n}{3}}$ grows much faster than $(2n)^{1+\sqrt{2n}}$. We leave this to the reader to show that for large enough n, this does not hold true. ∎

In fact, you can show easily that $n \leq 4000$. Also, notice that we can easily show that primes exist in the region $n < p < 2n$ for $n < 4000$ in at most 12 steps which is also easy.

MERTENS [52] gave more precise results than **??**.

Theorem 2.11 (Mertens' theorems) Let x be a positive real number. As $x \to \infty$,

$$\sum_{p \leq x} \frac{1}{p} = \log\log x + B + O\left(\frac{1}{x}\right) \tag{2.3}$$

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + O(1) \tag{2.4}$$

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log x} \tag{2.5}$$

where B is a constant and γ is Euler-Mascheroni constant.

Proof. From Stirling's approximation formula,

$$\log n! = n \log n - n + O(n)$$

Also, from Legendre's formula,

$$\begin{aligned}
\log n! &= \sum_{p^i \leq n} \left\lfloor \frac{n}{p^i} \right\rfloor \log p \\
&= \sum_{i \leq n} \Lambda(i) \left\lfloor \frac{n}{i} \right\rfloor \\
&= \sum_{i \leq n} \Lambda(i) \frac{n}{i} + O(\Lambda(i)) \\
&= n \sum_{i \leq n} \frac{\Lambda(i)}{i} + O\left(\sum_{i \leq n} \Lambda(i)\right) \\
&= n \sum_{i \leq n} \frac{\Lambda(i)}{i} + O(\psi(n))
\end{aligned}$$

Since $\psi(n) = O(n)$,

$$n \sum_{i \leq n} \frac{\Lambda(i)}{i} + O(n) = n \log n - n + O(n)$$

Thus, we have

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

Notice that this is awfully similar to the original sum in question. Indeed, the sum is very closely related to this.

$$\begin{aligned}
\left| \sum_{n \leq x} \frac{\Lambda(n)}{n} - \sum_{p \leq x} \frac{\log p}{p} \right| &\leq \sum_{p \leq x} \left(\frac{1}{p^2} + \frac{1}{p^3} + \dots \right) \log p \\
&= \sum_{p \leq x} \frac{1}{p^2} \left(\frac{p}{p-1} \right) \log p \\
&= \sum_{p \leq x} \frac{\log p}{p(p-1)}
\end{aligned}$$

Since for all large enough $p > N$, $\log p \leq \frac{p-1}{\sqrt{p}}$,

$$\begin{aligned}
\left| \sum_{n \leq x} \frac{\Lambda(n)}{n} - \sum_{p \leq x} \frac{\log p}{p} \right| &\leq \sum_{N < p \leq x} \frac{1}{p\sqrt{p}} + O(1) \\
&< \sum_{i \geq 1} \frac{1}{i^{\frac{3}{2}}} + O(1) \\
&= \zeta \left(\frac{3}{2} \right) + O(1)
\end{aligned}$$

Since $\zeta(s)$ converges for $s > 1$, we have

$$\begin{aligned}
\sum_{p \leq x} \frac{\log p}{p} &= \sum_{n \leq x} \frac{\Lambda(n)}{n} + O(1) \\
&= \log x + O(1)
\end{aligned}$$

We can now apply Abel partial summation formula with $a_n := \frac{\log p_n}{p_n}$.

$$\begin{aligned}
\sum_{p_n \leq x} \frac{1}{p_n} &= \sum_{p_n \leq x} \frac{a_n}{\log p_n} \\
&= \frac{A(x)}{\log x} + \int_2^x \frac{A(t)}{t \log^2 t} dt
\end{aligned}$$

where $A(x) = \sum_{p_n \le x} \frac{\log p_n}{p_n} = \log x + O(1)$.

$$\sum_{p_n \le x} \frac{1}{p_n} = 1 + O\left(\frac{1}{\log x}\right) + \int_2^x \frac{1}{t \log t} dt + O\left(\int_2^x \frac{1}{t \log^2 t}\right) dt$$

We leave the calculation of the integrals to the reader to show that

$$\sum_{p_n \le x} \frac{1}{p_n} = \log \log x + B + O\left(\frac{1}{\log x}\right)$$

■

Since $\psi(x) = \sum_{n \le x} \Lambda(n)$, we can write

$$\sum_{n \le x} \left\lfloor \frac{x}{n} \right\rfloor \Lambda(n) = \sum_{n \le x} \psi\left(\frac{x}{n}\right)$$

As a corollary, we get the following.

Corollary 2.12 For all $x \ge 1$,

$$\sum_{n \le x} \psi\left(\frac{x}{n}\right) = x \log x - x + O(\log x)$$

Two Elementary Proofs of Legendre-Dirichlet Prime Number Theorem | 3

EULER [25, pp. 241] proved that the sequence $a + nd$ contains infinitely many primes for $a = 1$. This can be proven using cyclotomic polynomials (see BILLAL and RIASAT [8, §1.4, Theorem 1.47]). LEGENDRE [48, pp. 404] conjectured that $a + nd$ contains infinitely many primes when $\gcd(a, d) = 1$ although failed to prove it as GAUSS [30, §29, pp. 505 − 508] noted. DIRICHLET [13] proved in his famous paper that $a + nd$ contains infinitely many primes though the proof is complete only when d is a prime. DIRICHLET [14] (also see DIRICHLET [17]) completes this proof with Dirichlet's class number formula. This is why today it is known as *Dirichlet's theorem on arithmetic progression* or *Dirichlet's prime number theorem.*

3.1 Dirichlet Characters

Dirichlet's idea of proving Legendre's conjecture essentially comes from Euler's proof of the divergence of the sum of reciprocal of primes, except Dirichlet wanted to prove the same when restricted by the constraint that $p \equiv a \pmod{k}$. Euler's proof gives us another way to prove there are infinitely many primes because the sum of reciprocals of primes diverge. If there were finitely many primes, that would not be the case. So, if we could prove

$$\sum_{p \equiv a \pmod{k}} \frac{1}{p}$$

diverges as well, the theorem would be complete. This is where Dirichlet introduced the crucial idea of *characters*. The motivation Dirichlet had was to define a *periodic* function in a certain way that would allow one to

1) differentiate when a is relatively prime to k.
2) behaves the same as the reduced set of residues $\pmod{k}$.
3) behaves the same as the complex roots of unity (this is actually very important, for example, the sum of all roots of unity is 0.)
4) can differentiate when $b \equiv a \pmod{k}$ for $\gcd(a, k) = 1$.

The intuition behind finding a periodic function is to note that $\gcd(a, k) = 1$ implies $\gcd(a + k, k) = 1$. As we will see later, Dirichlet characters satisfy all of these properties. It will become apparent later how and why they play such a crucial role. LEGENDRE [48, pp. 186] defined the *Legendre symbol* which plays a pivotal role in elementary number theory. JACOBI [36] generalized Legendre symbol which is now known as *Jacobi symbol*. KRONECKER [41, pp. 770] generalized this to *Kronecker symbol* which is unfortunately less known today. This is because DIRICHLET [13] had already introduced (also see DIRICHLET [16]) a generalization of Kronecker symbol which is far more insightful. Let k be a fixed positive integer throughout this chapter.

Definition 3.1 (Dirichlet Character) An arithmetic function χ is called a *character* $\pmod{k}$ if

1) $\chi(a) = 0$ if $\gcd(a, k) > 1$.
2) $\chi(1) \neq 0$.
3) $\chi(ab) = \chi(a)\chi(b)$ for all positive integers a, b.
4) $\chi(a) = \chi(b)$ if $a \equiv b \pmod{k}$.

Note that Legendre symbol $\left(\frac{a}{p}\right)$ for prime p, Jacobi symbol $\left(\frac{a}{n}\right)$ and Kronecker symbol $\left(\frac{a}{l}\right)$ are all characters $\pmod{p}$, $\pmod{n}$ and $\pmod{l}$ respectively where p is prime, n is a positive integer and $l \equiv 0, 1 \pmod{4}$ for square-free l. 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 $\chi(1) = 1$.

It is still not entirely clear how χ satisfies the properties we mentioned above. The next three theorems show why $\chi(n)$ behaves like complex roots of unity.

Proposition 3.2 If $\gcd(n, k) = 1$, then $\chi(n)^{\varphi(k)} = 1$. So $\chi(n)$ is actually a complex number which is an $\varphi(k)$-th root of unity and $|\chi(n)| = 1$.

Proof. From Euler's theorem,

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

So, $\chi(n)^{\varphi(k)} = \chi(n^{\varphi(k)}) = \chi(1) = 1$. ■

Definition 3.3 (Principal character) $\chi_0(n) = 1$ when $\gcd(n, k) = 1$ is the *principal character* $\pmod{k}$. If $\gcd(n, k) > 1$, $\chi_0(n) = 0$.

Since the characters $\pmod{k}$ are complex roots of unity, the complex conjugate of $\chi(n)$, $\bar{\chi(n)}$ is also a root of unity, hence; the following.

Proposition 3.4 If $\chi(n)$ is a character $\pmod{k}$, $\bar{\chi(n)}$ is also a character $\pmod{k}$.

Proposition 3.5 For any Dirichlet character χ,

$$\sum_{0 \leq a \leq k-1} \chi(a) = \begin{cases} \varphi(k) & \text{if } \chi = \chi_0 \\ 0 & \text{otherwise} \end{cases}$$

Proof. If $\chi(n) = \chi_0(n)$ is the principal character, then

$$\sum_{i=0}^{k-1} \chi_0(i) = \sum_{\substack{i=0 \\ \gcd(i,k)=1}}^{k-1} \chi_0(i) + \sum_{\substack{i=0 \\ \gcd(i,k)>1}}^{k-1} \chi_0(i)$$
$$= \sum_{\substack{i=0 \\ \gcd(i,k)=1}}^{k-1} 1$$
$$= \varphi(k)$$

If $\chi(n)$ is not the principal character, for any l prime to k such that $\chi(l) \neq 1$, the set of residues $\{i : 1 \leq i \leq k-1\}$ and $\{il : 1 \leq i \leq k-1\}$ coincide. Then

$$\sum_{i=1}^{k-1} \chi(i) = \sum_{i=1}^{k-1} \chi(il)$$
$$= \sum_{i=1}^{k-1} \chi(i)\chi(l)$$
$$= \chi(l) \sum_{i=1}^{k-1} \chi(i)$$

Since $\chi(l) \neq 1$,

$$\sum_{i=1}^{k-1} \chi(i) = 0$$

■

The next two theorems show why χ behaves like a complete/reduced set of residues.

Proposition 3.6 If $\chi_1(n)$ and $\chi_2(n)$ are both characters, then so is $\chi_1(n)\chi_2(n)$.

Proof. Left as an exercise. ■

Proposition 3.7 If $\chi'(n)$ is a character and $\chi(n)$ runs over all characters of k, $\chi'(n)\chi(n)$ runs over all characters of k as well.

Proof. If $\gcd(n,k) \neq 1$, $\chi'(n)\chi(n)$ and $\chi(n)$ both are 0. For $\gcd(n,k) = 1$, if $\chi'(n)\chi(n) = \chi''(n)\chi'(n)$ then $\chi(n) = \chi''(n)$ since neither is 0. By Proposition 3.6, $\chi(n)\chi'(n)$ produces $\varphi(k)$ different characters so are a permutation of the original characters $\chi'(n)$. ■

Proposition 3.8 If a is a positive integer prime to k such that $a \not\equiv 1 \pmod{k}$, then there exists a character $\chi(a) \neq 1$.

Proof. If p is a prime divisor of k and $k = p^e l$ such that $p \nmid l$ then $p \nmid a$. For any e, there exists a primitive $\lambda(p^e)$-th root $g \pmod{p^e}$ where $\lambda(n)$ is Carmichael's universal exponent function (see Billal and Riasat [8, pp. 90]). For any a not divisible by p, there is a unique non-negative integer u not exceeding $\lambda(p^e)$ such that

$$a \equiv g^u \pmod{p^e}$$

Setting $\zeta := \exp\left(\frac{2i\pi}{\lambda(p^e)}\right)$ and $\chi(a) = \zeta^u$, we can easily see for $a \equiv g^u \pmod{p^e}$ and $b \equiv g^v \pmod{p^e}$,

$$\begin{aligned} ab &\equiv g^{u+v} \pmod{p^e} \\ \chi(ab) &= \zeta^{u+v} \\ &= \zeta^u \zeta^v \\ &= \chi(a)\chi(b) \end{aligned}$$

The other properties are trivially satisfied. Now it remains to see that for $a \not\equiv 1 \pmod{k}$, if $a \equiv g^u \pmod{p^e}$ then $\lambda(p^e) \nmid u$ so $\zeta^u \neq 1$. ∎

Proposition 3.9 If $C(k)$ is the number of characters $\pmod{k}$, then

$$\sum_{\chi} \chi(a) = \begin{cases} C(k) \text{ if } a \equiv 1 \pmod{k} \\ 0 \text{ otherwise} \end{cases}$$

where the sum ranges through all the characters.

Proof. If $a \equiv 1 \pmod{k}$,

$$\begin{aligned} \sum_{\chi} \chi(a) &= \sum_{\chi} 1 \\ &= C(k) \end{aligned}$$

If $\gcd(a, k) > 1$, then $\chi(a) = 0$ for all χ so $\sum_{\chi} \chi(a) = 0$. Now, $\gcd(a, k) = 1$ and $a \not\equiv 1 \pmod{k}$. By Proposition 3.8, there is a character χ' such that $\chi'(a) \neq 1$. By Proposition 3.7,

$$\sum_{\chi} \chi(a) = \sum_{\chi} \chi(a)\chi'(a)$$

Since $\chi'(a) \neq 1$, we have $\sum_{\chi} \chi(a) = 0$. ∎

Proposition 3.10 There are $\varphi(k)$ characters $\pmod{k}$.

Proof. If χ runs through all the characters, by Proposition 3.9

$$\begin{aligned}\sum_{i=0}^{k-1}\sum_{\chi}\chi(i) &= 0+\sum_{\chi}\chi(1)+\sum_{i=2}^{k-1}\sum_{\chi}\chi(i)\\ &= C(k)\end{aligned}$$

On the other hand, by Proposition 3.5

$$\begin{aligned}\sum_{i=0}^{k-1}\sum_{\chi}\chi(i) &= \sum_{i=0}^{k-1}\chi_0(i)+\sum_{i=0}^{k-1}\sum_{\substack{\chi\\ \chi\neq\chi_0}}\chi(i)\\ &= \varphi(k)\end{aligned}$$

Thus, $C(k)=\varphi(k)$. ■

Proposition 3.11 Let l be a positive integer prime to k. Then for a positive integer a,

$$\sum_{\chi}\frac{\chi(a)}{\chi(l)} = \begin{cases}\varphi(k) \text{ if } a\equiv l \pmod{k}\\ 0 \text{ otherwise}\end{cases}$$

Proof. Let t be a positive integer such that $lt \equiv 1 \pmod{k}$. Such t exists since l is prime to k. Then $\chi(l)\chi(t)=\chi(lt)=\chi(1)=1$.

$$\begin{aligned}\sum_{\chi}\frac{\chi(a)}{\chi(l)} &= \sum_{\chi}\chi(a)\chi(t)\\ &= \sum_{\chi}\chi(at)\end{aligned}$$

According to Proposition 3.5, if $at\equiv 1 \pmod{k}$, then $\sum_{\chi}\chi(at)=\varphi(k)$ otherwise $\sum_{\chi}\chi(at)=0$. Since $at\equiv 1\equiv lt \pmod{k}$ and $\gcd(t,k)=1$, we have $a\equiv l \pmod{k}$. ■

Definition 3.12 χ is a *character of the first kind* if $\chi=\chi_0$ is the principal character. If χ is real but not principal, then χ is a *character of the second kind.* Otherwise, χ can sometimes be complex and called *character of the third kind.*

The next result is a very crucial one and was the intuition behind defining a function like χ with the properties we discussed.

Proposition 3.13 If χ is not a character of the first kind, then $\sum_{i=a}^{b}\chi(i)\le\frac{\varphi(k)}{2}$ for $1\le a\le b$.

Proof. Notice that, $\sum_{i=0}^{k-1}\chi(i)=0$ if χ is not of the first kind. So we can only consider $1\le a\le b\le a+k-1$. Also, $|\chi(a)|=1$ for $\varphi(k)$ residues that prime to k. If the number of residues i for which $|\chi(i)|=1$ and $a\le i\le b$ does not exceed $\frac{\varphi(k)}{2}$, then we are done. Otherwise, there are more than $\frac{\varphi(k)}{2}$ residues i for which $|\chi(i)|=1$ and $a\le i\le b$. In that case, the number of residues i for which $|\chi(i)|=1$ and $b+1\le i\le a+k-1$

does not exceed $\frac{\varphi(k)}{2}$.

$$\sum_{i=a}^{b} \chi(a) = \sum_{i=a}^{a+k-1} \chi(i) - \sum_{i=b+1}^{a+k-1} \chi(i)$$

Since $\sum_{i=a}^{a+k-1} \chi(i) = 0$,

$$\begin{aligned} \left| \sum_{i=a}^{b} \chi(a) \right| &= \left| \sum_{i=b+1}^{a+k-1} \chi(i) \right| \\ &< \frac{\varphi(k)}{2} \end{aligned}$$

■

We now prove a very insightful result.

Proposition 3.14 Let χ be a character of the second or third kind $\pmod k$ and f be a function such that $f(x) \geq 0$ and $f'(x)$ is continuous for all $x \geq x_0$ for some positive x_0. Then

$$\sum_{a<n\leq b} \chi(n) f(n) = O\left(f(a)\right)$$

When I said earlier that it would become apparent how someone would think of functions likes characters, this lemma was one of the results I had in my mind. Let $P(x)$ denote the set of primes not exceeding x. If for a positive integer $n > 1$, $P_a(x,n)$ is the set of primes not exceeding that are $a \pmod n$ and $r_1, \ldots, r_{\varphi(n)}$ are the numbers prime to n not exceeding n, then obviously $P(x) = P_{r_1}(x,n) \cup \ldots P_{r_{\varphi(n)}}(x,n)$ where $P_i(x,n)$ and $P_j(x,n) = \{\}$ are disjoint if $i \neq j$. Now, we know that there are infinitely many primes to $P(x) \to \infty$ as $x \to \infty$. So, it is highly likely we also have $P_a(x,n) \to \infty$ for fixed a and n as well. However, we need to somehow sift them out in a way that allows us to retain only primes of the form $a \pmod n$. Imagine something like this: we already know $\sum_{p\leq x} 1 \to \infty$ and we want to show $\sum_{\substack{p\leq x \\ p\equiv a \pmod n}} 1 \to \infty$ as $x \to \infty$. What if we could somehow relate these two? In order to do that, we should definitely look at something like $\sum_{p\leq x} f(a,p)$ where $f(a,p)$ would be 0 unless $p \equiv a \pmod n$. Again, it is very difficult to get an idea how to exactly handle this directly. One thing we could do is to consider something like $\sum_{p\leq x} f(a,p)g(p)$ where $g(p)$ would be known to us and we would have a good idea about both $\sum_{p\leq x} g(p)$. Then if we somehow had an estimation for $\sum_{p\leq x} f(a,p)g(p)$ itself, we could single out $\sum_{p\leq x} f(a,p)$ itself by using something like Abel partial summation formula. Now, clearly, we want $f(a,p) = 1$ when $p \equiv a \pmod n$ otherwise $f(a,p) = 0$. In order to be able to use Abel partial summation formula, we should consider something like $\sum_{n\leq x} f(a,n)$. Now, since we only want primes we must also have $f(a,n) = 0$ whenever $\gcd(a,n) > 1$. Another crucial intuition is that we can be certain $\sum_{\substack{p\leq x \\ p\equiv a \pmod n}}$ would be surely related to $\sum_{p\leq x} 1$, most likely bounded by a factor. In other words, we are conjecturing $\sum_{\substack{p\leq x \\ p\equiv a \pmod n}} = O\left(\sum_{p\leq x} 1\right)$. Obviously, this will turn out to be true, but we are pretending we still do not know if it is true or false. Then we should look for a function that can make $\sum_{p\leq x} f(a,p)g(p)$ disappear except for a few cases. This suggests us that the partial sums $\sum_{n\leq x} f(a,n)$ should be bounded, in other words we should look for functions that would give us $\sum_{y\le n\le x} f(a,n) = O(1)$ ideally. This might be a little too much too ask for, so we could even look for something like $\sum_{y\le n\le x} f(a,n)g(n)$ to be bounded by $g(x)$ so $\sum_{y\le n\le x} f(a,n)g(n) = O(g(x))$ would also be desirable. This is where the idea of complex roots come. If you recall, the sum of non-unit complex roots vanishes. Therefore, we should definitely look into sort of one to one correspondence of complex roots. This actually makes more sense if you recall that the only root that does not contribute to the vanishing sum is 1. This all sounds a bit handwavy so we will now jump into the formal proof again.

Proof. Since χ is non-principal,

$$\sum_{i=1}^{k-1} \chi(i) = 0$$

In Abel partial summation formula, setting $a_n := \chi(n)$, we have $A(k) = 0$. By periodicity, $A(nk) = 0$ for all positive integer n. Thus, if $\lfloor x \rfloor = kq + r$ with $0 \le r < k$, $A(x) = A(r) < \varphi(k)$ so $A(x) = O(1)$.

$$\begin{aligned}
\sum_{a<n\le b} \chi(n)f(n) &= A(b)f(b) - A(a)f(a) - \int_a^b A(t)f'(t)dt \\
&= O(f(b)) - O(f(a)) - O\left(\int_a^b f'(t)dt\right) \\
&= O(f(a))
\end{aligned}$$

∎

Since $\lim_{b\to\infty} \sum_{a<n\le b} \chi(n)f(n) = \sum_{n\ge 1} \chi(n)f(n) - \sum_{n\le a} \chi(n)f(n)$, we have the following.

Proposition 3.15 If $\lim_{x\to\infty} f(x) = 0$, then

$$\sum_{n\le a} \chi(n)f(n) = \sum_{n\ge 1} \chi(n)f(n) + O(f(a))$$

3.2 Dirichlet's *L*-Series

We mentioned earlier that Riemann considered $\zeta(s)$ for complex s. Here, we do something similar except we do not require any complex analysis except for the most basic facts.

Definition 3.16 (L-Series) Let s be a complex number. If k is a positive integer and χ is a character (mod k), Dirichlet's L-series $L(s,\chi)$ associated with χ and s is defined as

$$L(s,\chi) = \sum_{a\ge 1} \frac{\chi(a)}{a^s}$$

So, L-series is the Euler Product when $f(n) := \chi(n)$. Since χ is completely multiplicative, using properties of Euler product, we immediately have the following.

Proposition 3.17 For a complex s and a character χ,

$$L(s,\chi) = \prod_p \frac{1}{1 - \dfrac{\chi(p)}{p^s}}$$

Proposition 3.18 For any of $\varphi(k)$ characters χ and complex number s with $\Re(s) > 1$, $L(s,\chi)$ converges.

Proof. For a complex number $s = \sigma + it$ and $a = e^b$, $|a^s| = a^\sigma |a^{it}| = a^\sigma |e^{bit}|$. Since $e^{ix} = \cos x + i \sin x$, $|e^{ix}| = \cos^2 x + \sin^2 x = 1$. Thus, $|a^s| = a^\sigma$ and by triangle inequality,

$$\begin{aligned}
|L(s,\chi)| &\le \left| \sum_{a \ge 1} \frac{\chi(a)}{a^s} \right| \\
&\le \sum_{a \ge 1} \left| \frac{\chi(a)}{a^s} \right| \\
&= \sum_{a \ge 1} \frac{|\chi(a)|}{a^\sigma} \\
&\le \sum_{a \ge 1} \frac{1}{a^\sigma} \\
&= \zeta(\sigma)
\end{aligned}$$

If $\sigma > 1$, then $\zeta(\sigma)$ converges, hence, so does $L(s,\chi)$. ■

Proposition 3.19 For a non-principal χ, $L'(s,\chi)$ converges for $s > 1$.

Proof. This is comparatively easy to prove now using Proposition 3.13. ■

Proposition 3.20 If χ is a character of the second kind and

$$\varsigma(n) = \sum_{d|n} \chi(d)$$

Then $\varsigma(n) \ge 0$ for all n and $\varsigma(n) \ge 1$ for square n.

Proof. Since χ is completely multiplicative and ς is the summatory function on divisors, ς is multiplicative also. So we mainly need to look at

$$\begin{aligned}
\varsigma(p^e) &= \sum_{d|p^e} \chi(d) \\
&= 1 + \sum_{i=1}^{e} \chi(p^i) \\
&= 1 + \sum_{i=1}^{e} (\chi(p))^i
\end{aligned}$$

If $\chi(p) = 0$, then $\varsigma(p^e) = 1$. If $\chi(p) = 1$, then $\varsigma(p^e) = e + 1$. Otherwise $\chi(p) = -1$ so $\varsigma(p^e) = 1$ if e is even or $\varsigma(p^e) = 0$ if e is odd. If $n = p_1^{e_1} \cdots p_r^{e_r}$, then $\varsigma(n) = \varsigma(p_1^{e_1}) \cdots \varsigma(p_r^{e_r})$. Since each $\varsigma(p_i^{e_i}) \geq 0$, we have $\varsigma(n) \geq 0$ and if n is square, then e_i is even for $1 \leq i \leq r$ so $\varsigma(p_i^{e_i}) \geq 1$ and $\varsigma(n) \geq 1$. ■

Proposition 3.21 If χ is of the second kind, then $L(1, \chi) \neq 0$.

The proofs of $L(1, \chi) \neq 0$ for different kinds of χ presented here are primarily due to Apostol [3, Chapters 6] and Landau [47, Theorems 150-151]. Landau [47, pp. 121, Theorem 152] says the following about this theorem:

> This is the deepest of all of the lemmas that are necessary for Dirichlet's proof. Dirichlet proved it only by the considerably roundabout method of using the so-called theory of the class number of quadratic forms.

The reader is encouraged to try and prove Proposition 3.21. For example, a question is whether Abel partial summation formula works here with the following decomposition or not:

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

Then taking $x \to \infty$ seems very promising but there is a catch. Figuring out this catch is left to the reader as an exercise.

Proof. Let

$$\Upsilon(x) = \sum_{n \leq x} \frac{\varsigma(n)}{\sqrt{n}}$$

By Proposition 3.20, $\varsigma(n) \geq 1$ for square n, so

$$\Upsilon(x) \geq \sum_{n \leq \sqrt{x}} \frac{1}{n}$$

Clearly $\Upsilon(x) \to \infty$ as $x \to \infty$. Now,

$$\begin{aligned}\Upsilon(x) &= \sum_{n \le x} \frac{\sum_{d|n} \chi(d)}{\sqrt{n}} \\ &= \sum_{de \le x} \frac{\chi(d)}{\sqrt{de}}\end{aligned}$$

We can use Dirichlet Hyperbola Method with $a = b = \sqrt{x}$, $f(n) = \frac{\chi(n)}{\sqrt{n}}$ and $g(n) = \frac{1}{\sqrt{n}}$.

$$\Upsilon(x) = \sum_{n \le \sqrt{x}} \frac{\chi(n)}{\sqrt{n}} G\left(\frac{x}{n}\right) + \sum_{n \le \sqrt{x}} \frac{1}{\sqrt{n}} F\left(\frac{x}{n}\right) - F(\sqrt{x})G(\sqrt{x})$$

where $F(x) = \sum_{n \le x} f(n)$ and $G(x) = \sum_{n \le x} g(n)$. Setting $f(n) := \frac{1}{\sqrt{n}}$ in Proposition 3.15,

$$\sum_{n \le x} \frac{\chi(n)}{\sqrt{n}} = \sum_{n \ge 1} \frac{\chi(n)}{\sqrt{n}} + O\left(\frac{1}{\sqrt{x}}\right)$$

From Cauchy's convergence criteria, we can see that

$$\sum_{n \ge 1} \frac{\chi(n)}{\sqrt{n}}$$

converges. Since $\lim_{x \to \infty} \frac{1}{\sqrt{x}} = 0$, $F(x) = C + O\left(\frac{1}{\sqrt{x}}\right)$. Setting $s := \frac{1}{2}$ in Theorem 1.18,

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

Using these, we get

$$\begin{aligned}\Upsilon(x) &= \sum_{n \le \sqrt{x}} \frac{1}{\sqrt{n}} \left(\chi(n) \left(2\sqrt{\frac{x}{n}} + D + O\left(\sqrt{\frac{n}{x}}\right) + C + O\left(\sqrt{\frac{n}{x}}\right)\right)\right) - F(\sqrt{x})G(\sqrt{x}) \\ &= 2\sqrt{x} \sum_{n \le \sqrt{x}} \frac{\chi(n)}{n} + D \sum_{n \le \sqrt{x}} \frac{\chi(n)}{\sqrt{n}} + \frac{1}{\sqrt{x}} O\left(\sum_{n \le \sqrt{x}} \chi(n)\right) \\ &+ C \sum_{n \le \sqrt{x}} \frac{1}{\sqrt{n}} + \frac{1}{\sqrt{x}} O\left(\sum_{n \le \sqrt{x}} 1\right) - F(\sqrt{x})G(\sqrt{x}) \\ &= 2\sqrt{x} \sum_{n \le \sqrt{x}} \frac{\chi(n)}{n} + D \cdot F(\sqrt{x}) + C \cdot G(\sqrt{x}) + O(1) - F(\sqrt{x})G(\sqrt{x})\end{aligned}$$

Note that $ab - ac - bd = (b-c)(a-d) - cd$. Setting $a := F(\sqrt{x}), b := G(\sqrt{x})$, we see that

$$F(\sqrt{x})G(\sqrt{x}) - D \cdot F(\sqrt{x}) - C \cdot G(\sqrt{x}) = O(1) + O\left(\frac{1}{\sqrt{x}}\right)$$

Taking $x \to \infty$, we have

$$\begin{aligned}\lim_{x\to\infty} \Upsilon(x) &= 2\sqrt{x} \sum_{n \leq \sqrt{x}} \frac{\chi(n)}{n} + O(1) \\ &= 2\sqrt{x} L(1, \chi) + O(1)\end{aligned}$$

If $L(1, \chi) = 0$, then $\lim_{x\to\infty} \Upsilon(x^2) = O(1)$ which is impossible since $\Upsilon(x) \to \infty$ as $x \to \infty$. Therefore, we must have $L(1, \chi) \neq 0$. ■

Lemma 3.22 For real ϵ, t such that $0 < \epsilon < 1$,

$$(1-\epsilon)^3 |1 - \epsilon e^{it}|^4 |1 - \epsilon e^{2it}|^2 < 1$$

Proof. Since $|z|^2 = z\bar{z}$ and $e^{ix} = \cos x + i \sin x$, $1 - \epsilon e^{it} = 1 - \epsilon \cos t - i\epsilon \sin t$.

$$\begin{aligned}|1 - \epsilon e^{it}|^2 &= (1 - \epsilon \cos t)^2 + \epsilon^2 \sin^2 t \\ &= 1 - 2\epsilon \cos t + \epsilon^2\end{aligned}$$

Similarly, $1 - \epsilon e^{2it} = 1 - \epsilon \cos 2t - i\epsilon \sin 2t$, and

$$|1 - \epsilon e^{2it}|^2 = 1 - 2\epsilon \cos 2t + \epsilon^2$$

From the arithmetic-geometric mean inequality,

$$|1 - \epsilon e^{it}|^2 |1 - \epsilon e^{it}|^2 |1 - \epsilon e^{2it}|^2 \leq \left(1 - \frac{2}{3}\epsilon(2\cos t + \cos 2t) + \epsilon^2\right)^3$$

Now, $2\cos x + \cos 2x \geq -3/2$ due to

$$2\cos x + \cos 2x = -\frac{3}{2} + 2\left(\cos t + \frac{1}{2}\right)^2$$

Thus,

$$\begin{aligned}|1 - \epsilon e^{it}|^4 |1 - \epsilon e^{2it}|^2 &\leq (1 + \epsilon + \epsilon^2)^3 \\ &= \frac{1}{(1-\epsilon)}^3\end{aligned}$$

This proves the claim. ■

Lemma 3.23 For $s > 1$,

$$L(s, \chi_0)^3 |L(s, \chi)|^4 |L(s, \chi^2)|^2 \geq 1$$

Proof. We can write characters as $\chi = \zeta^i$ for some $0 \leq i < k$ where ζ is a complex root of unity. Then $\chi(p) = e^{it}$ for some real t. Take prime p such that $p \nmid k$. Setting $\epsilon = 1/p^s$,

$$\left(1 - \frac{\chi_0(p)}{p^s}\right)^3 \left|1 - \frac{\chi(p)}{p^s}\right|^4 \left|1 - \frac{\chi^2(p)}{p^s}\right|^2 \leq 1$$

If $p \mid k$, then we have equality. Then we see that the claim is true since

$$L(s, \chi) = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}$$

■

Proposition 3.24 $L(1, \chi) \neq 0$ for a non-principal character χ.

Proof. We have already showed this for χ of the second kind. If χ is of the third kind, notice that χ^2 is also a non-principal character. Because if $\chi^2 = 1$, then $\chi \in \{1, -1\}$ which would make χ real. Using Proposition 3.14,

$$\begin{aligned} |L(s, \chi)| &= \sum_{n \geq 1} \frac{\chi(n)}{n^s} \\ &\leq \sum_{n \geq 1} \chi(n) \\ &\leq \varphi(k) \end{aligned}$$

For $1 < s < 2$,

$$\begin{aligned} L(s, \chi_0) &\sum_{n \geq 1} \frac{\chi_0(n)}{n^s} \\ &= \sum_{\substack{n \geq 1 \\ \gcd(n,k)=1}} \frac{1}{n^s} \\ &< 1 + \sum_{n \geq 1} \frac{1}{n^s} \\ &< \frac{s}{s-1} \\ &< \frac{2}{s-1} \end{aligned}$$

Then $|L(s,\chi_0)|^3 < 8/(s-1)^3$. If $L(1,\chi)=0$ indeed holds,

$$\begin{aligned}
|L(s,\chi)| &= |L(s,\chi) - L(1,\chi)| \\
&= \left|\int_1^s L'(t,\chi)dt\right| \\
&< (s-1)\varphi(k)
\end{aligned}$$

Multiplying them, we have for $1 < s < 2$ and using $|L(s,\chi^2)| < \varphi(k)$,

$$|L(s,\chi)|^4 L(s,\chi_0)^3 |L(s,\chi^2)|^2 < (s-1)^4\varphi(k)^4 \frac{8}{(s-1)^3}\varphi(k)^2$$

By the lemma, $|L(s,\chi)|^4 L(s,\chi_0)^3|L(s,\chi^2)|^2 \geq 1$, so for all $1 < s < 2$,

$$8(s-1)\varphi(k)^6 > 1$$

For $k > 2$, $\varphi(k) > 1$, so setting

$$s := 1 + \frac{1}{8\varphi(k)^6}$$

which still satisfies $1 < s < 2$, we get $1 > 1$ which is impossible. Thus, $L(1,\chi) \neq 0$ cannot hold. ■

Proposition 3.25 If $s > 1$,

$$L'(s,\chi) = -\sum_{i\geq 1} \frac{\chi(i)\log i}{i^s}$$

and $L'(s,\chi)$ converges.

Proof. It is a straightforward differentiation $da^{-s} = (-a^{-s}\log a)da$. Next, we show that the absolute value of $L'(s,\chi)$ is bounded above for $s = 1+\epsilon$ where $\epsilon > 0$.

$$\begin{aligned}
\sum_{i\geq 1}\left|\frac{\chi(i)\log i}{i^s}\right| &= \sum_{i\geq 1}|\chi(i)|\left|\frac{\log i}{i^s}\right| \\
&= \sum_{i\geq 1}\left|\frac{\log i}{i^s}\right| \\
&\leq \sum_{i\geq 1}\frac{\log i}{i^{1+\epsilon}}
\end{aligned}$$

We leave it as an exercise to show that

$$\sum_{i\geq 1}\frac{\log i}{i^{1+\epsilon}}$$

diverges if $\epsilon = 0$ and converges if $\epsilon > 0$. ■

3.3 First Proof of Legendre-Dirichlet Theorem

We already discussed that we want to use the fact that the sum of reciprocals of primes diverges. If there are indeed infinitely many primes of the form $kn+l$ for $\gcd(k,l)=1$, we expect that

$$\lim_{x\to\infty} \sum_{\substack{p\le x \\ p\equiv l \pmod k}} \frac{1}{p} \to \infty$$

We already know that

$$\lim_{x\to\infty} \sum_{p\le x} \frac{1}{p} = \log\log x$$

Another similar result we established is

$$\lim_{x\to\infty} \sum_{p\le x} \frac{\log p}{p} = \log x \tag{3.1}$$

This actually gives us more room to work with because we can see that this sum is very closely related to

$$\lim_{x\to\infty} \sum_{n\le x} \frac{\Lambda(n)}{n}$$

This expression is not quite the same as the left side of (3.1). In fact,

$$\begin{aligned}
\sum_{n\le x} \frac{\Lambda(n)}{n} &= \sum_{p\le x} \frac{\log p}{p} + \sum_{p\le x} \log p \sum_{i\ge 2} \frac{1}{p^i} \\
&= \sum_{p\le x} \frac{\log p}{p} + \sum_{p\le x} \frac{\log p}{p(p-1)}
\end{aligned}$$

Now since

$$\sum_{p\le x} \frac{\log p}{p(p-1)} < \sum_{n\le x} \frac{\log n}{n(n-1)}$$

which obviously converges, we can see $\sum_{n\le x} \frac{\Lambda(n)}{n} \to \infty$ and $\sum_{p\le x} \frac{\log p}{p} \to \infty$ are equivalent and one would imply the other. It will be enough to prove the same with the additional constraint that p must be of the form $kn+l$. We can actually consider the more general sum

$$\sum_{\substack{n\le x \\ n\equiv l \pmod k}} \frac{\Lambda(n)}{n^s}$$

and then take $s \to 1$. Furthermore,

$$\sum_{p \le x} \frac{\log p}{p} = \sum_{\substack{1 \le l \le k \\ \gcd(l,k)=1}} \sum_{\substack{p \le x \\ p \equiv l \pmod{k}}} \frac{\log p}{p}$$

So, if we assume for the time now that the distribution of primes $\pmod{k}$ is uniform, meaning any choice l prime to k yields the same number of primes when $x \to \infty$, we expect that

$$\lim_{x \to \infty} \sum_{\substack{p \le x \\ p \equiv l \pmod{k}}} \frac{\log p}{p} = \frac{1}{\varphi(k)} \lim_{x \to \infty} \sum_{p \le x} \frac{\log p}{p}$$

since there are $\varphi(k)$ such l for which $\gcd(k,l) = 1$.

Next, we should contemplate how we can restrict the primes in the sum $\sum_{p \le x} \frac{\log p}{p}$ to only primes of the form $kn+l$. Here we can see how the characters come into the whole equation using Proposition 3.11. But before doing that if we recall $\log n = \sum_{d|n} \Lambda(n)$, there should be some connection between sums $\sum_{n \ge 1} \frac{\log n}{n^s}$, $\sum_{p \le x} \frac{\log p}{p}$ and $\sum_{n \ge 1} \frac{\Lambda(n)}{n^s}$. In fact, we have a better connection due to $L'(s,\chi)$ which even involves χ.

Proposition 3.26 For a complex s and a character χ,

$$\sum_{n \ge 1} \frac{\chi(n)\Lambda(n)}{n^s} = -\frac{L'(s,\chi)}{L(s,\chi)}$$

Proof. This is actually quite straightforward using the definitions and Dirichlet product.

$$\begin{aligned} L(s,\chi) \sum_{n \ge 1} \frac{\chi(n)\Lambda(n)}{n^s} &= \sum_{n \ge 1} \frac{\chi(n)}{n^s} \sum_{n \ge 1} \frac{\chi(n)\Lambda(n)}{n^s} \\ &= \sum_{n \ge 1} \frac{\chi(n)}{n^s} \sum_{d|n} \Lambda(n) \\ &= \sum_{n \ge 1} \frac{\chi(n) \log n}{n^s} \\ &= -L'(s,\chi) \end{aligned}$$

∎

Proposition 3.27 If χ is the principal character, then $\frac{L'(s,\chi)}{L(s,\chi)}$ converges if $s > 1$ and it diverges if $s \to 1$.

Proof. For $s > 1$, the proof is straightforward once we notice that $|\chi(n)\Lambda(n)| \le \log n$. For $s \to 1$,

$$-\frac{L'(s,\chi)}{L(s,\chi)} = \sum_{n \ge 1} \frac{\Lambda(n)}{n}$$

We have already shown that the divergence of $\sum_{n\le x}\frac{\Lambda(n)}{n}$ and $\sum_{p\le x}\frac{\log p}{p}$ are equivalent. ■

Proposition 3.28 If χ is not the principal character, $\frac{L'(s,\chi)}{L(s,\chi)}$ converges for $s>1$.

We omit the proof since it is straightforward at this stage. The stage is now set to finally prove Legendre's conjecture. Now we simply filter the sums related to $L(s,\chi)$ and $L'(s,\chi)$ to sieve out the primes of the form $kn+l$.

Proposition 3.29 Let l be relatively prime to k. Then

$$-\sum_{\chi}\frac{1}{\chi(l)}\frac{L'(s,\chi)}{L(s,\chi)} = \varphi(k)\sum_{\substack{n\ge 1\\ n\equiv l \pmod k}}\frac{\Lambda(n)}{n^s}$$

where χ runs over all the characters of k.

Proof. We will use the expansion of $\frac{L'(s,\chi)}{L(s,\chi)}$ along with Proposition 3.11.

$$\begin{aligned}
-\sum_{\chi}\frac{1}{\chi(l)}\frac{L'(s,\chi)}{L(s,\chi)} &= -\sum_{\chi}\frac{1}{\chi(l)}\sum_{n\ge 1}\frac{\chi(n)\Lambda(n)}{n^s}\\
&= \sum_{n\ge 1}\frac{\Lambda(n)}{n^s}\sum_{\chi}\frac{\chi(n)}{\chi(l)}\\
&= \sum_{\substack{n\ge 1\\ n\equiv l \pmod k}}\varphi(k)\frac{\Lambda(n)}{n^s}\\
&= \varphi(k)\sum_{\substack{n\ge 1\\ n\equiv l \pmod k}}\frac{\Lambda(n)}{n^s}
\end{aligned}$$

■

Theorem 3.30 (Dirichlet's Theorem on Arithmetic Progressions) For any l prime to k, there are infinitely many primes of the form $kn+l$.

Proof. Take $s\to 1$ such that $s=1+\epsilon$ for $\epsilon>0$.

$$-\sum_{\chi}\frac{1}{\chi(l)}\frac{L'(s,\chi)}{L(s,\chi)} = \varphi(k)\sum_{\substack{n\ge 1\\ n\equiv l \pmod k}}\frac{\Lambda(n)}{n^s}$$

On the left side, if χ is of the first kind, then $\frac{L'(s,\chi)}{L'(s,\chi)}\to\infty$ for $s\to 1$. Otherwise, $\frac{L'(s,\chi)}{L(s,\chi)}$ is finite. Thus, the left side is not bounded, hence, neither is the right side.

$$\sum_{\substack{n\geq 1 \\ n\equiv l \pmod{k}}} \frac{\Lambda(n)}{n^s} = \lim_{x\to\infty} \sum_{\substack{p^i\leq x \\ p\equiv l \pmod{k}}} \frac{\log p}{p^{is}}$$

We divide the primes that appear in the sum in two parts based on whether $i = 1$ or $i > 1$. For $i > 1, n \geq 1$ and $s \to 1$ from the right side, $n^{is} > n^i$ and $\frac{n}{2} < n$.

$$\begin{aligned}
\sum_{p^i\equiv l \pmod{k}} \frac{\log p}{p^{is}} &< \sum_{n\geq 1} \frac{\log n}{n^{is}} \\
&< \sum_{n\geq 1} \frac{\log n}{n^i} \\
&= \sum_{n\geq 2} \log n \sum_{i\geq 2} \frac{1}{n^i} \\
&= \sum_{n\geq 2} \frac{\log n}{n(n-1)} \\
&< 2\sum_{n\geq 2} \frac{\log n}{n^2}
\end{aligned}$$

Since the last sum converges,

$$\lim_{x\to\infty} \sum_{\substack{i\geq 2 \\ p^i\leq x \\ p\equiv l \pmod{k}}} \frac{\log p}{p^{is}} = O(1)$$

Therefore,

$$\lim_{x\to\infty} \sum_{\substack{p^i\leq x \\ p\equiv l \pmod{k}}} \frac{\log p}{p^{is}} \to \infty$$

$$\lim_{x\to\infty} \sum_{\substack{p\leq x \\ p\equiv l \pmod{k}}} \frac{\log p}{p^{s}} + \lim_{x\to\infty} \sum_{\substack{i\geq 2 \\ p^i\leq x \\ p\equiv l \pmod{k}}} \frac{\log p}{p^{is}} \to \infty$$

$$\lim_{x\to\infty} \sum_{\substack{p\leq x \\ p\equiv l \pmod{k}}} \frac{\log p}{p^{s}} + O(1) \to \infty$$

$$\lim_{x\to\infty} \sum_{\substack{p\leq x \\ p\equiv l \pmod{k}}} \frac{\log p}{p^{s}} \to \infty$$

Thus, there must be an infinite number of such $p \equiv l \pmod{k}$. Otherwise the sum would also be finite. ■

Note. It may not be apparent at first where we used the fact $L(1, \chi) \neq 0$. Do ponder on it in case you missed that we did not state it explicitly. But it is still a necessary prerequisite for this proof to work.

3.4 Second Proof by Selberg

Today there are many variations of the proof of Dirichlet's theorem on arithmetic progressions. But most proofs essentially have the same core idea. A much lesser known and studied proof due to SELBERG [57] has a little different approach to it. The idea itself is primarily a modification of *Selberg's asymptotic formula* which will be discussed heavily in Section 4.1. In the last proof, we tried to connect sums of prime powers of the form $kn + l$ with another sum that would not converge. Then we separated the primes only so that the prime powers that are at least squares contribute only to a convergent sum. Selberg's proof does something similar except instead of using the divergence of L series, he used a characteristic function for product of at most two prime powers. Here, we will lay the foundation of Selberg's asymptotic formula and we will prove the actual formula that will be used to prove the prime number theorem later. The calculations in this proof are pretty heavy but we will try to show as much of the calculation as possible here.

For a real x, let

$$\lambda_d(x) = \mu(d) \log^2 \frac{x}{d}$$

and ϕ be the summatory function of λ,

$$\phi_n(x) = \sum_{d|n} \lambda_d(x)$$

Lemma 3.31 We have

$$\phi_n(x) = \begin{cases} \log^2 x & \text{if } n = 1 \\ \log p \log^2 \frac{x}{p} & \text{if } n = p^u \\ 2 \log p \log q & \text{if } n = p^u q^v \\ 0 & \text{otherwise} \end{cases}$$

where p, q are distinct primes and u, v are positive integers.

Proof. Assume that n is square-free. If $n = p_1 \cdots p_r$ for distinct primes $p_1, \ldots, p_r$,

$$\phi_n(x) = \phi_{n/p_i}(x) - \phi_{n/p_i}(x/p_i) \tag{3.2}$$

Then by induction, the claim follows. If $n = p_1^{u_1} \cdots p_r^{u_r}$, then we can either resort to induction or simply notice that (3.2) is independent of u_i. ■

This ϕ can be considered a characteristic function for numbers with at most two distinct prime factors much like Mangoldt's Λ. However, ϕ is also dependent on x. Although, we can consider a generalization of von Mangoldt function. We know that $\log n = \sum_{d|n} \Lambda(d)$. From Möbius inversion,

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

This representation enables us to define the generalization.

Definition 3.32 (Generalized von Mangoldt Function) The r-th von Mangoldt function $\Lambda_r(n)$ for a non-negative integer r and positive integer n is

$$\Lambda_r(n) = \sum_{d|n} \mu(d) \log^r \frac{n}{d}$$

We can replace log with an arbitrary arithmetic function f that has the same property as logarithm $f(mn) = f(m) + f(n), f(m/n) = f(m) - f(n)$ and consider the summatory function

$$F_r(n) = \sum_{d|n} \mu(d)\, f^m\left(\frac{n}{d}\right) \tag{3.3}$$

We leave the following generalization as an exercise.

Proposition 3.33 If n has exactly r prime divisors, then with the notations in (3.3),

$$F_r(n) = r! \prod_{p^e \| n} \left(f(p^e) - f(p^{e-1})\right)$$

where $p^e \| n$ denotes $p^e \mid n$ but $p^{e+1} \nmid n$.

For the rest of the section, χ is a character of the second kind. When $\gcd(k, l) = 1$,

$$\begin{aligned}\sum_{\substack{n \le x \\ n \equiv l \pmod k}} \phi_n(x) &= +O\left(\log^2 x\right) + \sum_{\substack{p^u \le x \\ p^u \equiv l \pmod k}} \phi_{p^u}(x) + \sum_{\substack{p^u q^v \le x \\ p^u q^v \equiv l \pmod k}} \phi_{p^u q^v}(x) \\ &= O\left(\log^2 x\right) + \sum_{\substack{p^u \le x \\ p^u \equiv l \pmod k}} \log p \log \frac{x^2}{p} + \sum_{\substack{p^u q^v \le x \\ p^u q^v \equiv l \pmod k}} \log p \log q\end{aligned}$$

Here, we are not actually omitting the factor 2. Rather we are simply considering that the contribution from $p^u q^v$ is $2 \log p \log q$ whereas $q^v p^u$ is 0 since both are actually same even though by order only which does not matter. So instead we can simply consider it as if both $p^u q^v$ and $q^v p^u$ contribute $\log p \log q$ which actually makes our calculation easier while keeping the total invariant. Again, we will treat $u = 1$ and $u > 1$ separately.

$$\sum_{\substack{p^u \le x \\ p^u \equiv l \pmod k}} \log p \log \frac{x^2}{p} = \sum_{\substack{p \le x \\ p \equiv l \pmod k}} \log p \log \frac{x^2}{p} + \sum_{\substack{p^u \le x, u \ge 2 \\ p^u \equiv l \pmod k}} \log p \log \frac{x^2}{p}$$

You can probably guess by now that the reason we treat them separately is because largest contribution towards the sum is made by the $p \le x$ rather than $p^2 \le x, p^3 \le x, \ldots$ since they are much rarer.

$$\begin{aligned}
\sum_{\substack{p \le x \\ p \equiv l \pmod k}} \log p \log \frac{x^2}{p} &= \sum_{\substack{p \le x \\ p \equiv l \pmod k}} \log p (2 \log x - \log p) \\
&= 2 \sum_{\substack{p \le x \\ p \equiv l \pmod k}} \log p \log \frac{x}{p} + \sum_{\substack{p \le x \\ p \equiv l \pmod k}} \log^2 p
\end{aligned}$$

Using partial summation, $\sum_{p \le x} \log^2 p = \left(\sum_{p \le x} \log p\right) \log x - \int_2^x \frac{\vartheta(t)}{t} dt$.

$$\begin{aligned}
\sum_{p \le x} \log p \log \frac{x}{p} &= \left(\sum_{p \le x} \log p\right) \log x - \sum_{p \le x} \log^2 p \\
&= \vartheta(x) \log x - \sum_{p \le x} \log^2 p \\
&= \int_2^x \frac{\vartheta(t)}{t} dt \\
&= O(x)
\end{aligned}$$

since we already know from Tchebyscheff's theorems that $\vartheta(x) = O(x)$ and $\psi(x) = O(x)$. Thus,

$$\sum_{\substack{p \le x \\ p \equiv l \pmod k}} \log p \log \frac{x^2}{p} = \sum_{\substack{p \le x \\ p \equiv l \pmod k}} \log^2 p + O(x)$$

For $u \ge 2$,

$$\begin{aligned}
\sum_{\substack{p^u \le x, u \ge 2 \\ p^u \equiv l \pmod k}} \log p \log \frac{x^2}{p} &\le 2 \big(\sum_{p \le \sqrt{x}} \log p \big) \log x \\
&\le 2\pi(\sqrt{x}) \log^2 x \\
&= O\left(\frac{\sqrt{x}}{\log x}\right) \log^2 x \\
&= O\left(\sqrt{x} \log x\right)
\end{aligned}$$

Next, for when $u \geq 2$ and $n = p^u q^v$,

$$\sum_{\substack{p^u q^v \leq x \\ u \geq 2 \\ p^u q^v \equiv l \pmod k}} \log p \log q \leq \sum_{\substack{p^u q^v \leq x \\ u \geq 2}} \log p \log q$$

$$= \sum_{\substack{p^u \leq x \\ u \geq 2}} (\log p) \sum_{q^v \leq x/p^u} \log q$$

$$= \sum_{\substack{p^u \leq x \\ u \geq 2}} (\log p) \left(\log_q \left\lfloor \frac{x}{p^u} \right\rfloor \log q \right)$$

Using $\log_a b \log bc = \log_a c$,

$$\sum_{\substack{p^u q^v \leq x \\ u \geq 2 \\ p^u q^v \equiv l \pmod k}} \log p \log q = \sum_{\substack{p^u \leq x \\ u \geq 2}} \log p \log \left\lfloor \frac{x}{p^u} \right\rfloor$$

$$= O\left(x \sum_{\substack{p^u \leq x \\ u \geq 2}} \frac{\log p}{p^u} \right)$$

$$= O\left(x \sum_{p \leq x} \frac{\log p}{p(p-1)} \right)$$

Again, $\sum \frac{\log p}{p(p-1)}$ converges, so the sum is simply $O(x)$. Combining all of these,

$$\sum_{\substack{n \leq x \\ n \equiv l \pmod k}} \phi_n(x) = \sum_{\substack{p \leq x \\ p \equiv l \pmod k}} \log^2 p + \sum_{\substack{pq \leq x \\ pq \equiv l \pmod k}} \log p \log q + O(x)$$

Since ϕ is the summatory function of λ,

$$\begin{aligned}
\sum_{\substack{n\le x\\ n\equiv l \pmod k}} \phi_n(x) &= \sum_{\substack{n\le x\\ n\equiv l \pmod k}} \sum_{d|n} \lambda_d(x) \\
&= \sum_{\substack{n\le x\\ \gcd(n,k)=1}} \lambda_n(x) \sum_{\substack{m\le x/n\\ mn\equiv l \pmod k}} 1 \\
&= \sum_{\substack{n\le x\\ \gcd(n,k)=1}} \lambda_n(x) \left\lfloor \frac{x}{nk} \right\rfloor \\
&= \frac{x}{k} \sum_{\substack{n\le x\\ \gcd(n,k)=1}} \frac{\lambda_n(x)}{n} + O\left(\sum_{n\le x} |\lambda_n(x)|\right) \\
&= \frac{x}{k} \sum_{\substack{n\le x\\ \gcd(n,k)=1}} \frac{\lambda_n(x)}{n} + O\left(\sum_{n\le x} \log^2 \frac{x}{n}\right)
\end{aligned}$$

Now, $\log^2 \frac{x}{n} = (\log x - \log n)^2 = \log^2 x - 2\log x \log n + \log^2 n$. Using Abel's partial summation formula, $\sum_{n\le x} \log n = x\log x - x + O(\log x)$. Combining this and Abel's partial summation again on $\sum_{n\le x} \log^2 n$, we can verify that

$$\sum_{n\le x} \log^2 \frac{x}{n} = O(x)$$

Thus,

$$\sum_{\substack{n\le x\\ n\equiv l \pmod k}} \phi_n(x) = \frac{x}{k} \sum_{\substack{n\le x\\ \gcd(n,k)=1}} \frac{\lambda_n(x)}{n} + O(x)$$

$$\sum_{\substack{p\le x\\ p\equiv l \pmod k}} \log^2 p + \sum_{\substack{pq\le x\\ pq\equiv l \pmod k}} \log p \log q = \frac{x}{k} \sum_{\substack{n\le x\\ \gcd(n,k)=1}} \frac{\lambda_n(x)}{n} + O(x)$$

This is now much more friendly to work with because we have reduced the restriction to $\gcd(n,k) = 1$

only which is free of l.

$$\begin{aligned}
\sum_{\substack{n=1\\ \gcd(n,k)>1}}^{x} \sum_{\substack{p\le x\\ p\equiv n \pmod k}} \log^2 p &\le \sum_{p\le x}\sum_{\substack{n=1\\ p|n}}^{k} \log^2 p \\
&= \sum_{p\le x} \log^2 p \left\lfloor \frac{k}{p} \right\rfloor \\
&\le k \sum_{p\le x} \frac{\log^2 p}{p} \\
&= O\left(\log^2 x\right)
\end{aligned}$$

We also need to calculate the same for $\log p \log q$.

$$\begin{aligned}
\sum_{\substack{n=1\\ \gcd(n,k)>1}}^{x} \sum_{\substack{pq\le x\\ pq\equiv n \pmod k}} \log p \log q &\le 2\sum_{pq\le x}\sum_{\substack{n=1\\ p|n}}^{k} \log p \log q \\
&= 2\sum_{pq\le x} \log p \log q \left\lfloor \frac{k}{p} \right\rfloor \\
&\le 2k \sum_{pq\le x} \frac{\log p}{p} \log q \\
&= 2k \sum_{p\le x} \frac{\log p}{p} \sum_{q\le x/p} \log q \\
&= O\left(\sum_{p\le x} \frac{\log p}{p} \sum_{q\le x/p} \log(x/p) \right) \\
&= O\left(\sum_{p\le x} \frac{\log p}{p} \log(x/p) O\left(\frac{x/p}{\log(x/p)} \right) \right) \\
&= O\left(x \sum_{p\le x} \frac{\log p}{p^2} \right) \\
&= O\left(x\right)
\end{aligned}$$

We leave it as an exercise for the reader to show that

$$\sum_{p\le x} \frac{\log p}{p^2} = O(1)$$

This is once again a desirable result for us. The right side is independent of the choice of l. Therefore,

$$\sum_{\substack{p\le x\\ p\equiv l \pmod k}} \log^2 p + \sum_{\substack{pq\le x\\ pq\equiv l \pmod k}} \log p \log q = \frac{1}{\varphi(k)}\left(\sum_{p\le x}\log^2 p + \sum_{pq\le x}\log p\log q\right) + O(x)$$

So far, it hasn't been very clear where exactly this line of thinking was going. Now we see that we finally have a result relating primes not exceeding x which are of the form $kn+l$ to primes not exceeding x in general. This last result we have actually gives us more insight. One way is to write $\log^2 p = p\left(\frac{\log^2 p}{p}\right)$ and use partial summation formula.

$$\sum_{\substack{p\le x\\ p\equiv l \pmod k}} \frac{\log^2 p}{p} = \frac{1}{x}\sum_{\substack{p\le x\\ p\equiv l \pmod k}} \log^2 p + \int_1^x \left(\sum_{\substack{p\le t\\ p\equiv l \pmod k}} \log^2 p\right)\frac{1}{t^2}dt$$

In a similar manner,

$$\sum_{\substack{pq\le x\\ pq\equiv l \pmod k}} \frac{\log p\log q}{pq} = \frac{1}{x}\sum_{\substack{pq\le x\\ pq\equiv l \pmod k}} \log p\log q + \int_1^x \left(\sum_{\substack{pq\le t\\ pq\equiv l \pmod k}} \log p\log q\right)\frac{1}{t^2}dt$$

Summing them,

$$\begin{aligned}
\sum_{\substack{p\le x\\ p\equiv l \pmod k}} \frac{\log^2 p}{p} + \sum_{\substack{pq\le x\\ pq\equiv l \pmod k}} \frac{\log p\log q}{pq} &= \frac{1}{x}\left(\sum_{\substack{p\le x\\ p\equiv l \pmod k}} \log^2 p + \sum_{\substack{pq\le x\\ pq\equiv l \pmod k}} \log p\log q\right)\\
&+ \int_1^x \frac{1}{t^2}\left(\sum_{\substack{p\le t\\ p\equiv l \pmod k}} \log^2 p + \sum_{\substack{pq\le t\\ pq\equiv l \pmod k}} \log p\log q\right)dt\\
&= \frac{1}{x\varphi(k)}\left(\sum_{p\le x}\log^2 p + \sum_{pq\le x}\log p\log q\right) + O(1)\\
&+ \int_1^x \left(\frac{1}{t^2\varphi(k)}\left(\sum_{p\le x}\log^2 p + \sum_{pq\le x}\log p\log q\right) + O\left(\frac{1}{t}\right)\right)dt\\
&= \frac{1}{x\varphi(k)}\left(\sum_{p\le x}\log^2 p + \sum_{pq\le x}\log p\log q\right)\\
&+ \frac{1}{\varphi(k)}\int_1^x \frac{\sum_{p\le t}\log^2 p + \sum_{pq\le t}\log p\log q}{t^2}dt + O(\log x)
\end{aligned}$$

This is where we use the substitution we mentioned earlier.

$$\sum_{p\le x}\log^2 p=\sum_{p\le x}p\left(\frac{\log^2 p}{p}\right)$$

Using partial summation formula here,

$$\sum_{p\le x}\log^2 p=x\sum_{p\le x}\frac{\log^2 p}{p}-\int_1^x\sum_{p\le t}\frac{\log^2 p}{p}dt$$

Similarly,

$$\begin{aligned}\sum_{pq\le x}\log p\log q&=\sum_{pq\le x}pq\left(\frac{\log p\log q}{pq}\right)\\&=x\sum_{pq\le x}\frac{\log p\log q}{pq}-\int_1^x\sum_{pq\le t}\frac{\log p\log q}{pq}dt\end{aligned}$$

Summing them together,

$$\begin{aligned}\sum_{p\le x}\log^2 p+\sum_{pq\le x}\log p\log q=x&\left(\sum_{p\le x}\frac{\log^2 p}{p}+\sum_{pq\le x}\frac{\log p\log q}{pq}\right)\\&-\int_1^x\left(\sum_{p\le t}\frac{\log^2 p}{p}+\sum_{pq\le t}\frac{\log p\log q}{pq}\right)dt\end{aligned}$$

We again leave this calculation as an exercise to finally reach that

$$\sum_{\substack{p\le x\\ p\equiv l \pmod k}}\frac{\log^2 p}{p}+\sum_{\substack{pq\le x\\ pq\equiv l \pmod k}}\frac{\log p\log q}{pq}=\frac{1}{\varphi(k)}\left(\sum_{p\le x}\frac{\log^2 p}{p}+\sum_{pq\le x}\frac{\log p\log q}{pq}\right)+O(\log x)$$

Here we can again get a better estimation separating $\frac{\log p\log q}{pq}$.

$$\begin{aligned}\sum_{pq\le x}\frac{\log p\log q}{pq}&=\sum_{p\le x}\frac{\log p}{p}\sum_{q\le x/p}\frac{\log q}{q}\\&=\sum_{p\le x}\frac{\log p}{p}\log\frac{x}{p}+O\left(\log x\right)\\&=\left(\sum_{p\le x}\frac{\log p}{p}\right)\log x-\sum_{p\le x}\frac{\log^2 p}{p}+O(\log x)\\&=\log^2 x-\sum_{p\le x}\frac{\log^2 p}{p}+O(\log x)\end{aligned}$$

Therefore,

$$\sum_{p\le x}\frac{\log^2 p}{p}+\sum_{pq\le x}\frac{\log p\log q}{pq}=\log^2 x+O(\log x)$$
$$\sum_{\substack{p\le x\\ p\equiv l \pmod k}}\frac{\log^2 p}{p}+\sum_{\substack{pq\le x\\ pq\equiv l \pmod k}}\frac{\log p\log q}{pq}=\frac{\log^2 x}{\varphi(k)}+O(\log x)$$

Notice that we already know partial sum of $\frac{\log p}{p}$ so it makes sense to use it again, however, $\frac{\log p\log q}{pq}$ makes it difficult. We can simply ignore this term and write

$$\sum_{\substack{p\le x\\ p\equiv l \pmod k}}\frac{\log^2 p}{p}\le\frac{\log^2 x}{\varphi(k)}+O(\log x)$$

Using partial summation formula,

$$\begin{aligned}
\sum_{\substack{p\le x\\ p\equiv l \pmod k}}\frac{\log p}{p} &= \sum_{\substack{p\le x\\ p\equiv l \pmod k}}\frac{\log^2 p}{p\log p}\\
&=\frac{\sum_{p\equiv l \pmod k}^{p\le x}\frac{\log^2 p}{p}}{\log x}+\int_1^x\frac{\sum_{p\equiv l \pmod k}^{p\le t}\frac{\log^2 p}{p}}{t\log^2 t}dt\\
&\le\frac{\log x}{\varphi(k)}+\int_1^x\frac{\frac{\log^2 t}{\varphi(k)}+O(\log t)}{t\log^2 t}dt\\
&=\frac{\log x}{\varphi(k)}+\frac{1}{\varphi(k)}\int_1^x\frac{1}{t}dt+O\left(\int_1^x\frac{1}{t\log t}dt\right)\\
&=\frac{2\log x}{\varphi(k)}+O(1)+I
\end{aligned}$$

Letting $\log t=z$, $dt=e^z dz$,

$$\begin{aligned}
I&=\int_1^{\log x}\frac{e^z dz}{ze^z}\\
&=\log\log x+O(1)
\end{aligned}$$

Thus, we have

$$\sum_{\substack{p\le x\\ p\equiv l \pmod k}}\frac{\log p}{p}\le\frac{2\log x}{\varphi(k)}+O\left(\log\log x\right)$$

Same as before, we also calculate

$$\sum_{\substack{pq\le x\\ pq\equiv l \pmod k}} \frac{\log p\log q}{pq} \le \sum_{\substack{p\le \sqrt[3]{x}\\ q\le \sqrt[3]{x}\\ pq\equiv l \pmod k}} \frac{\log p\log q}{pq} + 2\sum_{\substack{\sqrt[3]{x}<p\le x\\ pq\equiv l \pmod k}} \frac{\log p\log q}{pq}$$

$$= \sum_{p\le \sqrt[3]{x}} \sum_{\substack{q\le \sqrt[3]{x}\\ pq\equiv l \pmod k}} \frac{\log p\log q}{pq} + 2\sum_{\sqrt[3]{x}<p\le x} \frac{\log p}{p} \sum_{\substack{q\le x/p\\ pq\equiv l \pmod k}} \frac{\log q}{q}$$

For the second part of the right sum, note that

$$\sum_{p\le z} \frac{\log p}{p}\log(x/p) = \log x\log z - \frac{\log^2 z}{2} + O(\log x)$$

Then

$$\sum_{\sqrt[3]{x}<p\le x} \frac{\log p}{p} \sum_{\substack{q\le x/p\\ pq\equiv l \pmod k}} \frac{\log q}{q} \le \sum_{\sqrt[3]{x}<p\le x} \frac{2\log p\log(x/p)}{p\varphi(k)} + O\left(\sum_{\sqrt[3]{x}<p\le x} \frac{\log p}{p}\log\log x\right)$$

$$= \frac{2}{\varphi(k)} \sum_{\sqrt[3]{x}<p\le x} \frac{\log p}{p}\log(x/p) + O\left(\log x\log\log x\right)$$

$$= \frac{2}{\varphi(k)} \left(\sum_{p\le x} \frac{\log p}{p}\log(x/p) - \sum_{p\le \sqrt[3]{x}} \frac{\log p}{p}\log(x/p)\right) + O\left(\log x\log\log x\right)$$

$$= \frac{2}{\varphi(k)} \left(\frac{1}{2}\log^2 x - \frac{1}{3}\log^2 x + \frac{1}{18}\log^2 x + O(\log x)\right) + O(\log x\log\log x)$$

$$= \frac{4\log^2 x}{9\varphi(k)} + O(\log x\log\log x)$$

Using this

$$\sum_{\substack{pq\le x\\ pq\equiv l \pmod k}} \frac{\log p\log q}{pq} \le \sum_{p\le \sqrt[3]{x}} \sum_{\substack{q\le \sqrt[3]{x}\\ pq\equiv l \pmod k}} \frac{\log p\log q}{pq} + \frac{8\log^2 x}{9\varphi(k)} + O(\log\log x)$$

Recall that we have already established

$$\sum_{\substack{p\le x\\ p\equiv l \pmod k}} \frac{\log^2 p}{p} + \sum_{\substack{pq\le x\\ pq\equiv l \pmod k}} \frac{\log p\log q}{pq} = \frac{\log^2 x}{\varphi(k)} + O(\log x)$$

This gives us

$$\sum_{\substack{p\le x\\ p\equiv l \pmod k}} \frac{\log^2 p}{p} \ge \frac{\log^2 x}{9\varphi(k)} - \sum_{\substack{p,q\le \sqrt[3]{x}\\ pq\equiv l \pmod k}} \frac{\log p \log q}{pq} + O(\log\log x)$$

Obviously,

$$\sum_{\substack{p\le x\\ p\equiv l \pmod k}} \frac{\log^2 p}{p} \le \log x \sum_{\substack{p\le x\\ p\equiv l \pmod k}} \frac{\log p}{p}$$

Therefore,

$$\begin{aligned}
\sum_{\substack{p\le x\\ p\equiv l \pmod k}} \frac{\log p}{p} &\ge \frac{\log x}{9\varphi(k)} - \frac{1}{\log x}\sum_{\substack{p,q\le \sqrt[3]{x}\\ pq\equiv l \pmod k}} \frac{\log p \log q}{pq} + O(\log\log x)\\
\frac{1}{\log x}\sum_{\substack{p\le x\\ p\equiv l \pmod k}} \frac{\log p}{p} &\ge \frac{1}{9\varphi(k)} - \frac{1}{\log^2 x}\sum_{\substack{p,q\le \sqrt[3]{x}\\ pq\equiv l \pmod k}} \frac{\log p \log q}{pq} + O\left(\frac{\log\log x}{\log x}\right)\\
&= \frac{1}{9\varphi(k)} - \frac{1}{\log^2 x}\sum_{\substack{p\le \sqrt[3]{x}\\ p\equiv u \pmod k}} \frac{\log p}{p} \sum_{\substack{q\le \sqrt[3]{x}\\ q\equiv v \pmod k\\ uv\equiv l \pmod k}} \frac{\log q}{q} + O(1)
\end{aligned}$$

Letting

$$Q_l(x) := \frac{1}{\log x}\sum_{\substack{p\le x\\ p\equiv l \pmod k}} \frac{\log p}{p}$$

we get

$$\begin{aligned}
Q_l(x) &\ge \frac{1}{9\varphi(k)} - \frac{1}{9}\sum_{uv\equiv l \pmod k}\left(\frac{1}{\log \sqrt[3]{x}}\sum_{\substack{p\le \sqrt[3]{x}\\ p\equiv u \pmod k}} \frac{\log p}{p}\right)\left(\frac{1}{\log \sqrt[3]{x}}\sum_{\substack{q\le \sqrt[3]{x}\\ q\equiv v \pmod k}} \frac{\log q}{q}\right) + O(1)\\
&\ge \frac{1}{9\varphi(k)} - \sum_{uv\equiv l \pmod k} Q_u(\sqrt[3]{x})Q_v(\sqrt[3]{x}) + O(1)
\end{aligned}$$

Two Elementary Proofs of the Prime Number Theorem 4

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 6. For now, we will ignore the history and focus only on the proof.

4.1 Selberg's Asymptotic Formula

4.2 First Proof by Erdős and Selberg

4.3 Second Proof by Selberg

5 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

$$\begin{aligned}\pi(x)-\pi(\sqrt{x})+1 &= \lfloor x\rfloor - \sum_{p\le x}\left\lfloor\frac{x}{p}\right\rfloor + \sum_{p_1<p_2\le x}\left\lfloor\frac{x}{p_1p_2}\right\rfloor - \sum_{p_1<p_2<p_3\le x}\left\lfloor\frac{x}{p_1p_2p_3}\right\rfloor + \dots\\ &= \sum_{n\le x}\mu(n)\left\lfloor\frac{x}{n}\right\rfloor\end{aligned}$$

Now, $\lfloor x\rfloor = x + O(1)$, so

$$\begin{aligned}\pi(x)-\pi(\sqrt{x})+1 &= x\sum_{\substack{n\le x\\ \rho(n)\le\sqrt{x}}}\frac{\mu(n)}{n} + O\left(\sum_{\substack{n\le x\\ \rho(n)\le\sqrt{x}}}\mu(n)\right)\\ &= x\prod_{p\le\sqrt{x}}\left(1-\frac{1}{p}\right) + O(2^{\pi(\sqrt{x})})\end{aligned}$$

The last line is true since there are $\pi(\sqrt{x})$ primes not exceeding $\sqrt{x}$ and $|\mu(n)| = 1$ for all square-free $n \le x$ such that $\rho(n) \le \sqrt{x}$. 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. $\mathfrak{M}, \mathfrak{G}$

5.1 Brun's Sieve

Brun [9] also see Brun [10]

5.2 Selberg's Sieve

5.3 Turán's Method

6 A Mathematical Dispute of Twentieth Century

The prime number theorem was first conjectured by Gauss in his letter to the astronomer Encke as pointed out by LANDAU [43, pp. 37].

This page is intentionally left blank

Articles

Niels Henrik Abel

1826 "Untersuchungen über die Reihe: $1+(m/1)x+m\cdot(m-1)/(1\cdot2)\cdots x^2+m\cdot(m-1)\cdot(m-2)/(1\cdot 2\cdot 3)\cdots x^3+\ldots$", *Journal für Math.*, vol. I, pp. 311-339, DOI: 10.1515/9783112347386-030.

Raymond Ayoub

1974 "Euler and the zeta function", *The American Mathematical Monthly*, vol. LXXXI, no. 10, pp. 1067-1086, DOI: 10.2307/2319041.

Joseph Bertrand

1845 "Mémoire sur le nombre de valeurs que peut prendre une fonction quand on y permute les lettres qu'elle renferme", *Journal de l'École Royale Polytechnique*, vol. XVIII, pp. 123-140.

Martin Beumer

1962 "The arithmetical function $\tau_k(n)$", *The American Mathematical Monthly*, vol. LXIX, no. 8, pp. 777-781, DOI: 10.2307/2310778.

Viggo Brun

1915 "Über das Goldbachsche Gesetz und die Anzahl der Primzahlpaare", *Archiv for Mathematik og Naturvidenskab*, vol. XXXIV, no. 8, pp. 1-19.

Eugène Cahen

1894 "Sur la fonction $\zeta(s)$ de Riemann et sur des fonctions analogues", fr, *Annales scientifiques de l'École Normale Supérieure*, vol. XI, 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.

1839 "Recherches sur diverses applications de l'analyse infinitesimale à la théorie des nombres", *Journal für die reine und angewandte Mathematik (Crelles Journal)*, vol. MDCCCXXXIX, no. 19, pp. 324-369, DOI: 10.1515/crll.1839.19.324.

Paul Erdős

1932 "Beweis eines Satzes von Tschebyschef", *Acta Litt. Univ. Sci., Szeged, Sect. Math.*, vol. V, pp. 194-198.

Leonhard Euler

1737 "Variae Observationes circa series Infinitas", *Commentarii Academiae Scientiarum Petropolitanae*, vol. IX, pp. 160-188, DOI: 10.1090/spec/098/41.

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

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

1744 "Variae Observationes circa series infinitas", *Commentarii academiae scientiarum Petropolitanae*, vol. IX, 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*, vol. XLVIII, 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*, vol. VI, no. 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*, vol. XXX, pp. 127-136, DOI: 10.1515/crll.1846.30.166.

Johan Ludwig William Valdemar Jensen

1884 "OM RÆKKERS KONVERGENS", *Tidsskrift for mathematik*, 5th ser., vol. II, pp. 63-72, ISSN: 09092528, 24460737, http://www.jstor.org/stable/24540057.

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*, vol. XXII, no. 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*, vol. XIX, pp. 687-772.

1912b “Über eine idealtheoretische funktion”, *Transactions of the American Mathematical Society*, vol. XIII, no. 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*, vol. XXII, no. 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*, vol. CXVIII, no. 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 (Crelles Journal)*, vol. LXXVIII, pp. 46-62, DOI: 10.1515/crll.1874.78.46.

August Ferdinand Möbius

1832 “Über eine besondere art von Umkehrung der Reihen.” *Journal für die reine und angewandte Mathematik (Crelles Journal)*, vol. IX, 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*, vol. III, 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*, vol. L, no. 2, pp. 297-304, DOI: 10.2307/1969454.

Pafnutï Lvovitch Tchebyscheff

1852a “Mémoire sur les nombres premiers”, *Journal de Mathématiques Pures et Appliquées*, vol. XVII, no. 1, pp. 366-390.

1852b “Sur la totalité des nombres premiers inférieurs à une limite donnée”, *Journal de Mathématiques Pures et Appliquées*, vol. XVII, no. 1, pp. 341-365.

Books

Martin Aigner and Günter M. Ziegler

1999 *Proofs from the book*, 3rd ed., Springer.

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 *Die analytische zahlentheorie*, Teubner, vol. II.

Masum Billal and Samin Riasat

2021 *Integer sequences: Divisibility, Lucas and Lehmer sequences*, Springer Nature, DOI: 10.1007/978-981-16-0570-3.

Viggo Brun

1920 *Le crible d'eratosthène et le théorème de goldbach*, no. 3, Skrifter utgit av videnskapsselskapet i Kristiania, mat.-naturv.

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. I, pp. 313-342, DOI: 10.1017/cbo9781139237321.012.

1897b "Recherches sur diverses applications de l'analyse infinitésimale à la théorie des nombres", in, *G. Lejeune Dirichlet's Werke*, ed. by Leopold Kronecker and László Fuchs, Druck Und Verlag Von Georg Reimer., vol. I, pp. 411-496, DOI: 10.1017/cbo9781139237338.030.

1897c "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. II, pp. 49-66, DOI: 10.1017/cbo9781139237345.007.

Leonhard Euler

1748 *Introductio in Analysin Infinitorum*, vol. I.

1783 "De summa seriei ex numeris primis formatae $1/3 - 1/5 + 1/7 + 1/11 - 1/13 - 1/17 + 1/19 + 1/23 - 1/29 + 1/31$ etc. ubi numeri primi formae $4n - 1$ habent signum positivum, formae autem $4n + 1$ signum negativum", in, *Opuscula analytica*, St. Petersburg: Imperial Academy of Sciences, vol. II.

Joshn Friedlander and Henryk Iwaniec

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

Carl Friedrich Gauss

1801 *Disquisitiones Arithmeticae. Auctore D. Carolo Friderico Gauss*, In commissis apud Gerh. Fleischer, DOI: 10.1007/978-1-4939-7560-0.

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. II.

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

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

Adrien Marie Legendre

1798 *Essai sur la theéorie 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.

James Stirling

1730 *Methodus differentialis: Sive, tractatus de summatione et interpolatione serierum infinitarum*, 1st ed., Typis G. Bowyer.

Appendix

Theorem .1 (Legendre's formula) The exact power of prime p in $n!$ is

$$\nu_p(n!) = \sum_{n \geq 1} \left\lfloor \frac{n}{p^i} \right\rfloor$$

Theorem .2 (Cauchy's convergence criteria) Let $(a_{n\geq})$ be a sequence of complex numbers. Then $(a_{n\geq 1})$ converges if and only if for any positive real number ϵ, there exists a positive integer N such that $s_n - s_m < \epsilon$ for all $n \geq m \geq N$ where

$$s_k = \sum_{i=1}^{k} a_i$$

Theorem .3 (Stirling's approximation formula) For positive integer n,

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n+1}} < n! < \sqrt{2\pi n}\left(\frac{n}{e}\right)^n e^{\frac{1}{12n}}$$

Stirling [58] actually showed an exact series for $\log n!$ but usually a weaker statement like this suffices. We can also write it as

$$\log n! = n \log n - n + O\left(\log n\right)$$

Index