Université de Montréal

# Prime Number Races

par

**Tony Haddad**

Département de mathématiques et de statistique
Faculté des arts et des sciences

Mémoire présenté en vue de l'obtention du grade de
Maître ès sciences (M.Sc.)
en mathématiques

Orientation mathématiques pures

Août 2020

# Université de Montréal

Faculté des arts et des sciences

Ce mémoire intitulé

## Prime Number Races

présenté par

## Tony Haddad

a été évalué par un jury composé des personnes suivantes :

*Andrew Granville*

(président-rapporteur)

*Dimitris Koukoulopoulos*

(directeur de recherche)

*Matilde Lalín*

(membre du jury)

# Résumé

Sous l'hypothèse de Riemann généralisée et l'hypothèse d'indépendance linéaire, Rubinstein et Sarnak ont prouvé que les valeurs de $x \geqslant 1$ pour lesquelles nous avons plus de nombres premiers de la forme $4n + 3$ que de nombres premiers de la forme $4n + 1$ en dessous de $x$ ont une densité logarithmique d'environ 99,59 %. En général, l'étude de la différence $\#\{p \leqslant x : p \in A\} - \#\{p \leqslant x : p \in B\}$ pour deux sous-ensembles de nombres premiers $A$ et $B$ s'appelle la course entre les nombres premiers de $A$ et de $B$. Dans ce mémoire, nous cherchons ultimement à analyser d'un point de vue numérique et statistique la course entre les nombres premiers $p$ tels que $2p + 1$ est aussi premier (aussi appelés nombres premiers de Sophie Germain) et les nombres premiers $p$ tels que $2p - 1$ est aussi premier. Pour ce faire, nous présentons au préalable l'analyse de Rubinstein et Sarnak pour pouvoir repérer d'où vient le biais dans la course entre les nombres premiers 1 (mod 4) et les nombres premiers 3 (mod 4) et émettons une conjecture sur la distribution des nombres premiers de Sophie Germain.

**Mots-clés: Courses de nombres premiers, Biais de Chebyshev, Formule explicite, Fonctions $L$, Nombres premiers de Sophie Germain, Crible de Selberg, Modèle de Cramér, Méthode du cercle, Théorie analytique des nombres**

# Abstract

Under the Generalized Riemann Hypothesis and the Linear Independence Hypothesis, Rubinstein and Sarnak proved that the values of $x$ which have more prime numbers less than or equal to $x$ of the form $4n+3$ than primes of the form $4n+1$ have a logarithmic density of $\approx 99.59\,\%$. In general, the study of the difference $\#\{p \leqslant x : p \in A\} - \#\{p \leqslant x : p \in B\}$ for two subsets of the primes $A$ and $B$ is called the prime number race between $A$ and $B$. In this thesis, we will analyze the prime number race between the primes $p$ such that $2p+1$ is also prime (these primes are called the Sophie Germain primes) and the primes $p$ such that $2p-1$ is also prime. To understand this, we first present Rubinstein and Sarnak's analysis to understand where the bias between primes that are 1 (mod 4) and the ones that are 3 (mod 4) comes from and give a conjecture on the distribution of Sophie Germain primes.

**Keywords: Prime number races, Chebyshev's bias, Explicit formula, $L$-Functions, Sophie Germain primes, Selberg sieve, Cramér's model, Circle method, Analytic number theory**

# Contents

# List of tables

# List of figures

# List of abbreviations and acronyms

| | |
|---|---|
| CLT | Central Limit Theorem |
| gcd | Greatest common divisor |
| GRH | Generalized Riemann Hypothesis |
| lcm | Least common multiple |
| LI | Linear Independence Hypothesis |
| LIL | Law of the Iterated Logarithm |
| RH | Riemann Hypothesis |
| PNT | Prime Number Theorem |

# Notation and conventions

In this thesis, the variable $p$ always represents a prime number unless explicitly stated otherwise. The sets of numbers $\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$ and $\mathbb{C}$, respectively, designate the natural numbers, the integers, the rational numbers, the real numbers and the complex numbers. The natural numbers are the positive integers, namely the numbers $1, 2, 3$, etc. We will have two different ways to write the indicator function. If $P(n)$ is a logical proposition, then $\mathbf{1}_{P(n)}$ returns 1 if $P(n)$ is true, and 0 if $P(n)$ is false. If $A \subset \Omega$, then $\mathbf{1}_A \colon \Omega \to \mathbb{R}$ is defined by $\mathbf{1}_A(x) = 1$ whenever $x \in A$ and $\mathbf{1}_A(x) = 0$ otherwise.

From elementary number theory, when $d$ and $n$ are integers, we write $d \,|\, n$ to mean that $n$ is divisible by $d$. If $p^\nu$ is a prime power, then $p^\nu \,\|\, n$ means that $p^\nu$ is the highest power of $p$ dividing $n$ and we say that $p^\nu$ divides exactly $q$. The gcd of two integers $a$ and $b$ is denoted by the symbol $(a, b)$ and their lcm is denoted by $[a, b]$. We have to note that $(a, b)$ and $[a, b]$ could also represent real intervals but there will not be any ambiguity from the context. For a finite set $A$, we use $|A|$ or $\#A$ to represent the number of elements in $A$. For $x \in \mathbb{R}$, the greatest integer smaller than $x$, also called the integer part, is noted $[x]$. The fractional part is defined by $\{x\} = x - [x]$ and the distance between $x$ and its closest integer is denoted by $\|x\| = \min_{n\in\mathbb{Z}} |x - n|$.

If $z$ is a complex number, then $\mathrm{Re}(z)$ denotes its real part and $\mathrm{Im}(z)$ denotes its imaginary part. As Riemann originally did, we will often use the variable $s$ to represent a complex number, and we will denote its real and imaginary parts by $\sigma$ and $t$, respectively, so that $s = \sigma + it$. In addition, the nontrivial zeros of any $L$-function or of the Riemann zeta function will have the special notation $\rho = \beta + i\gamma$. If we give a subscript to $s$ or $\rho$, the subscript is automatically given to the variables representing their real and imaginary parts. For example, $\mathrm{Re}(s_0) = \sigma_0$ and $\mathrm{Im}(\rho_\chi) = \gamma_\chi$.

The symbol $e(t)$ is another way to write $e^{2\pi it}$. Moreover, we write $\log x$ for the natural logarithm (with base $e = 2.71828\ldots$).

For asymptotic estimates, if $f, g$ are two functions with real or complex values and $a \in \widehat{\mathbb{R}} = \mathbb{R} \cup \{\pm\infty\}$, then

$$f(x) \sim g(x) \quad (x \to a) \qquad \Longleftrightarrow \qquad \lim_{x\to a} \frac{f(x)}{g(x)} = 1.$$

Also, if $g$ takes positive real values in a neighbourhood of $a$, then

$$f(x) = o(g(x)) \quad (x \to a) \qquad \Longleftrightarrow \qquad \lim_{x \to a} \frac{f(x)}{g(x)} = 0.$$

To describe bounds, we will use Vinogradov's notation $f(x) \ll g(x)$ or Landau's notation $f(x) = O(g(x))$ to mean that $|f(x)| \leqslant c \cdot g(x)$ for a positive constant $c$. The range of $x$ where the bound is valid will be specified. If the constant $c$ depends on a parameter $\alpha$, then we will write $f(x) \ll_\alpha g(x)$ or $f(x) = O_\alpha(g(x))$. If two positive functions $f, g$ have the same order of magnitude in the sense that $f(x) \ll g(x) \ll f(x)$, then we write $f(x) \asymp g(x)$.

Finally, if $f$ is a complex-valued function, then $f^*$ is defined by

$$f^*(x) = \lim_{\varepsilon \to 0} \frac{f(x - \varepsilon) + f(x + \varepsilon)}{2},$$

provided the limit exists.

# Remerciements

J’aimerais remercier mon directeur de recherche, Dimitris Koukoulopoulos, pour sa disponibilité, son soutien, sa patience et toutes les discussions à son bureau ou à travers Zoom. Je suis très reconnaissant pour le projet qu’il m’a donné qui n’a pas cessé de me fasciner et pour avoir été disponible pour répondre à mes questions. Ce mémoire aurait été très différent sans son aide. J’aimerais aussi le remercier pour son soutien financier, ainsi que le Département de Mathématiques et Statistique, le centre de recherche CICMA et l’ISM pour leurs bourses. De plus, tout ce que je raconte dans mon mémoire est basé sur des techniques que j’ai apprises dans mes cours de théorie des nombres analytique avec Dimitris et avec Andrew Granville. Je remercie aussi les membres du jury pour leur lecture attentive et leurs commentaires instructifs.

J’aimerais remercier mes amis à l’université, en particulier ma cohorte au baccalauréat en mathématiques pures. Je me suis toujours senti entouré de gens aussi curieux et intéressés par les mathématiques que moi. Je tiens aussi à remercier Stelios et Kunjakanan, que j’ai rencontrés pendant ma maîtrise et à leurs visites hebdomadaires à mon bureau. Merci à Mélanie, Marie, Andy, Susan, Katia et Kevin pour leur soutien constant et leurs amitiés. Merci à Daniel avec qui j’ai vécu à peu près tout et ses conseils aux moments où j’en avais le plus besoin. Merci à Hansheng, à ses longues discussions téléphoniques à propos de tout et de rien, on s’est développé une routine ensemble pour atteindre nos propres objectifs. Et un merci particulier à Mathieu pour toutes les conversations que j’ai eues avec lui, il s’intéressait toujours à ma maîtrise et était toujours là à discuter.

Je tiens à remercier mes cousins Michel, Mariane et Élina, ma grand-mère Rosette, ma grande-tante Marie-Rose et tous les autres membres de ma familles pour leur soutien. En particulier, merci à mon oncle Rafic et ma tante Salwa qui ont toujours essayer de satisfaire ma curiosité depuis que je suis tout petit, et de me laisser emprunter leurs livres pour une durée indéterminée. Un gros merci à Céline, Antoine, Yara, Pierre et Nada pour m’acceuillir chez eux et pour leur soupers, je ne sortais jamais de leur maison sans un sourire. En particulier à Céline pour son amitié, ses conseils et pour m’avoir aidé à passer à travers mes moments les plus difficiles.

Je remercie du fond de mon cœur Anik pour son amour, ses encouragements et sa joie de vivre. L'année 2020 n'est pas une année facile ni pour ce monde, ni pour moi, mais c'est beaucoup plus facile à gérer avec le sourire d'Anik à ses côtés.

Finalement, j'aimerais remercier mon frère Martin et mes parents Diana et Elie. Ils ont toujours cru en moi et ils m'ont ouvert toutes les portes pour que je me rende là où je suis. Je leur suis infiniment reconnaissant. Sans eux, il n'y aurait pas ce mémoire.

# Chapter 1

# As if they were random

## 1.1. Prime numbers

When we first start to master multiplication at the beginning of our elementary education, we notice that some positive integers can be taken apart into a product of two smaller positive integers. For example: $12 = 3 \times 4$ or $35 = 5 \times 7$. If the decomposition of $n \in \mathbb{N}$ was made using addition, it would have been very easy to find every single way to break it down: $12 = 1 + 11 = 2 + 10 = \ldots = 6 + 6$. This is because subtracting from $n$ any smaller number $m$ would result in another positive integer. With multiplication, though, factoring is not as effortless. When $1 < m < n$, the number $\frac{n}{m}$ is not always an integer, and this is where we usually first see a clear distinction in the way we treat addition and multiplication. The numbers in $\mathbb{N}$, which cannot be written as a product of two smaller natural numbers, are called *prime numbers*.

After breaking down a number (for example, $60 = 12 \times 5$), the factors themselves are not necessarily prime, and they might have a nontrivial decomposition of their own ($60 = 12 \times 5 = (4 \times 3) \times 5$). By inductively applying this reasoning at every step of the factorization, every nonprime integer greater than 2 (which is also called a *composite number*) can be broken down into a product of prime numbers:

$$60 = 12 \times 5 = 4 \times 3 \times 5 = 2 \times 2 \times 3 \times 5.$$

One of the most important results in elementary number theory, which is why prime numbers are so important when studying multiplication in integers, is that this decomposition into prime numbers is unique (up to the order of the factors). This prime number decomposition can characterize every integer, and it dictates the multiplicative relations between them. Prime numbers take this way the roles of the atoms in the integer universe.

## 1.2. Chaos in primes

One might expect that primes, which have such a simple definition, are well-behaved objects and that their distribution can be intuitive and predictable. However, we quickly see by enumerating them that there is no clear pattern emerging. Taking a large odd prime, we cannot predict easily where the next prime lies. The next odd number could have a lot of prime factors. The next odd number could also be a prime number (in which case we say they are *twin primes*). It is a famous open problem to prove that there are infinitely many twin primes.

This chaos and lack of understanding of how the primes behave has some surprising applications. For example, RSA cryptography is an algorithm widely used to secure data transmission. This algorithm uses a public key, meaning that the tools used to encrypt a message (which are huge integers in RSA) are not kept secret. The private key used to decrypt the message should be kept hidden; otherwise, there would be no point in having this type of cryptosystem since anyone could decipher an encrypted message.

In RSA cryptography, the public key is an integer $N$, which is the product of two primes, and the private key is $\phi(N) := \{1 \leqslant n \leqslant N : (n, N) = 1\}$, where $\phi$ is called the *Euler totient function*. If $p_1, p_2$ are distinct primes, then $\phi(p_1p_2) = (p_1 - 1)(p_2 - 1)$. Thus, we need to know how to factor $N$ into primes to compute $\phi(N)$, and there is no known polynomial-time algorithm for integer factorization. This is what makes RSA cryptography so strong. There is a correlation between our understanding of integer factorization into prime numbers and the strength of the encryption scheme.

Let now $p$ and $q$ be two primes satisfying the linear equation $q = 2p + 1$. In this case, we say that $p$ is a *Sophie Germain prime* and that $q$ is a *safe prime*. Safe primes are strong candidates for RSA cryptography. The reason is that John M. Pollard created in 1974 an algorithm (see [25]) to easily factor integers of the form $N = p_1p_2$ if either $p_1 - 1$ or $p_2 - 1$ are *smooth numbers*, that is, numbers with prime divisors lower than some fixed bound. By taking safe primes, we ensure that one cannot use Pollard's algorithm to find the private key and decipher our messages.

## 1.3. Probability and number theory

There exist some deterministic tests that detect in polynomial time if a number is prime or not[1]. However, to find prime numbers, we might have to take a more global and probabilistic approach. Perhaps a regularity starts forming if we look at the prime numbers as an ensemble instead of just looking at a prime and its neighbours. Some questions about the frequency and the behaviour of prime numbers can be answered this way.

[1] The first one found that did not rely on any unproven conjecture was the AKS primality test, which was created in 2002. See [1].

We have been in a never-ending quest to find these mystical objects' properties since Euclid's time. A fundamental question was answered by Euclid himself, who proved that there are infinitely many prime numbers. Indeed, if there were only a finite amount, we can get a new number by multiplying them all together and adding one:

$$N = \left( \prod_{p} p \right) + 1.$$

This number $N$ cannot be divisible by any prime, so it should be a prime itself. The contradiction is that we would have found a new prime bigger than any prime.

In this global view of the primes, the next question could naturally be: Do we know how many primes are there around some large real number $x$? The chaotic nature of the primes does not help us in having a straight answer. But looking at it through a probabilistic lens and based on a multitude of computations, Carl Friedrich Gauss famously conjectured at 15 or 16 years of age that the probability that a large number $x$ is a prime number is about $\frac{1}{\log x}$.

When looking at the prime numbers from a measure-theoretic standpoint, one has to construct a prime-counting measure $\mu$ described by

$$\mu(A) := \sum_{p \in A} 1 = \# \left\{ p \in A \right\}$$

and this measure can be characterized by the following cumulative distribution function:

$$\pi(x) := \mu((-\infty,x]) = \sum_{p \leqslant x} 1 = \# \left\{ p \leqslant x \right\}.$$

This function $\pi$ is a step-function that has jumps at every prime number. At first glance, we have only changed the tools we are using to describe prime numbers. The advantage of this point of view becomes clear when looking at the function $\pi$ in a graph on a large scale. The data indicate that we can approximate $\pi$ by a smooth function (see Figure 1.1).

In his only paper in number theory, Bernhard Riemann used the function $\zeta(s) := \sum_{n=1}^{\infty} n^{-s}$, first introduced by Leonhard Euler, to study the distribution of prime numbers. He proved that $(s-1)\zeta(s)$ admits an analytic continuation on the entire complex plane and provided a functional equation showing there is a symmetry between $\zeta(s)$ and $\zeta(1-s)$. The function $\zeta$ is now referred to as the *Riemann zeta function*. Riemann also made multiple conjectures about it in his paper. Here are three of them:

(1) If $N(T)$ is the number of zeros $\rho = \beta + i\gamma$ such that $|\gamma| \leqslant T$ and $0 < \beta < 1$, then

$$N(T) = \frac{T}{\pi} \log\!\left(\frac{T}{2\pi}\right) - \frac{T}{\pi} + O(\log T). \tag{1.1}$$

(2) The function $\pi(x) - \mathrm{Li}(x)$ has an explicit formula given in terms of the zeros of $\zeta$.

(3) If $\rho = \beta + i\gamma$ is a complex number with $0 \leqslant \beta \leqslant 1$ and $\zeta(\rho) = 0$, then $\beta = \frac{1}{2}$.

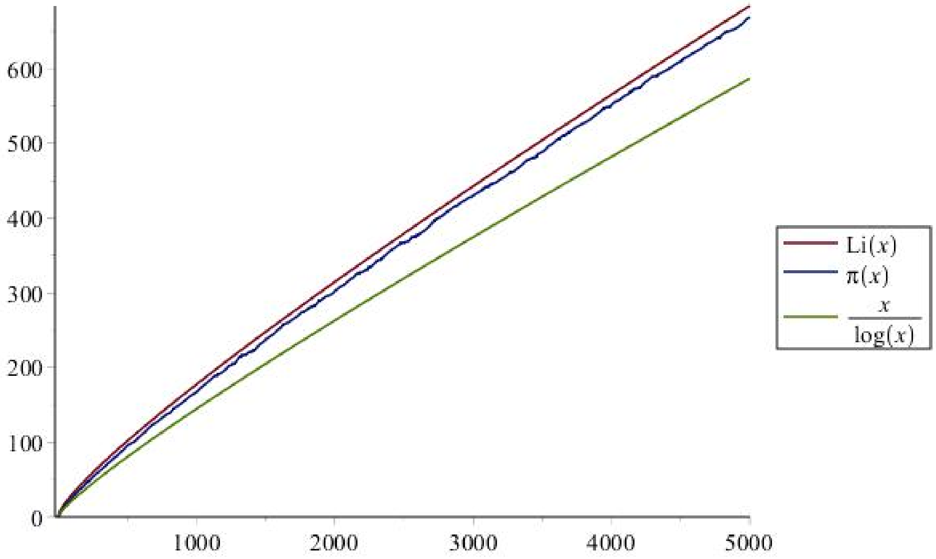

**Fig. 1.1.** Approximating $\pi(x)$ with a smooth function.

The $\zeta$ function has zeros at every nonpositive even integers $-2, -4, \ldots$ Those are called the *trivial zeros*. Every other zero of $\zeta$, often referred as a *nontrivial zero*, is in the set $\{s \in \mathbb{C} : 0 < \sigma < 1\}$ called the *critical strip*. We also call the set $\left\{s \in \mathbb{C} : \sigma = \frac{1}{2}\right\}$, the *critical line*. As per conjecture (2) above, knowing where the nontrivial zeros are located inside the critical strip helps us understand the error term $\pi(x) - \mathrm{Li}(x)$. Riemann's conjecture (3), known as the *Riemann Hypothesis* (RH), states that every nontrivial zero lies on the critical line. This problem remains wide open to this day.

An asymptotic formula for the function $\pi(x)$ was ultimately proven with the *Prime Number Theorem* (PNT). It gives us a way to estimate $\pi(x)$ with a smooth function by saying that

$$\pi(x) \sim \mathrm{Li}(x) \quad \text{as } x \to \infty,$$

where

$$\mathrm{Li}(x) := \int_2^x \frac{\mathrm{d}t}{\log t}.$$

In 1895, Hans von Mangoldt proved Riemann's conjecture (2) by giving the following explicit formula: For any $x > 1$ which is not a prime power, we have

$$\sum_{p^k \leqslant x} \log p = x - \lim_{T \to \infty} \sum_{|\gamma| \leqslant T} \frac{x^\rho}{\rho} - \log(2\pi) - \frac{1}{2}\log(1 - x^{-2}), \tag{1.2}$$

where $\rho = \beta + i\gamma$ runs over the nontrivial zeros of $\zeta$ with multiplicity. This formula shows how the location of the zeros dictates the distribution of the primes. Von Mangoldt also proved Riemann's conjecture (1) about the density of zeros inside the critical strip.

In 1896, Jacques Hadamard and Charles-Jean de la Vallée Poussin independently proved the Prime Number Theorem by finding a region in the critical strip without any zeros of $\zeta$. This allowed them to say that the right-hand side of the explicit formula (1.2) is $\sim x$. De la Vallée Poussin ultimately proved that

$$\pi(x) = \mathrm{Li}(x) + O\Big(xe^{-c\sqrt{\log x}}\Big) \quad \text{for } x \geqslant 2, \tag{1.3}$$

where $c$ is an absolute positive constant.

The above bound on the error term has been improved to $O\Big(xe^{-c'(\log x)^{3/5}(\log\log x)^{-1/5}}\Big)$ for another positive constant $c'$ by I. M. Vinogradov and N. M. Korobov in 1958, who enlarged the known zero-free region of the critical strip. But the data reveals that we could still be far away from the true order of the error term. Let $\alpha$ be a function on the interval $[10^4, 10^8]$ defined by the equation $\pi(x) = \mathrm{Li}(x) - x^{\alpha(x)}$. If we could not do better than Vinogradov and Korobov's estimate, than for any $\varepsilon > 0$, there would be infinitely many values of $x$ such that $|\pi(x) - \mathrm{Li}(x)| > x^{1-\varepsilon}$ because $x^{1-\varepsilon} = o(xe^{-c'(\log x)^{3/5}(\log\log x)^{-1/5}})$ as $x \to \infty$. So we would expect $\alpha(x)$ to sometimes wander just below 1 for $x$ large enough. However, this is not quite what happens as we see in Figure 1.2.

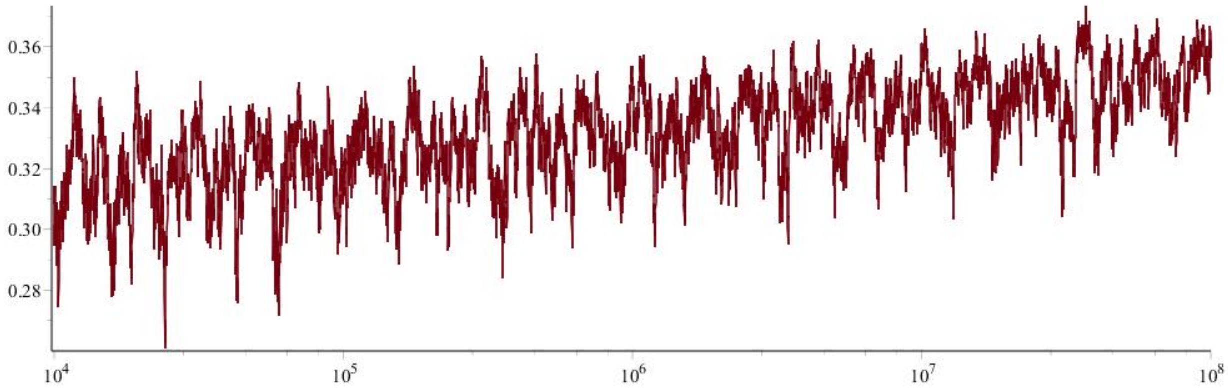

**Fig. 1.2.** Graph of the function $\alpha$ defined by $\pi(x) = \mathrm{Li}(x) - x^{\alpha(x)}$.

Indeed, we can understand $\alpha(x)$ in terms of the zeros of $\zeta$. The explicit formula implies that

$$\limsup_{x\to\infty} \frac{\log|\pi(x) - \mathrm{Li}(x)|}{\log x} \leqslant \sup\Big\{\beta \in [\tfrac{1}{2}, 1) : \zeta(\beta + i\gamma) = 0\Big\}. \tag{1.4}$$

The reversed inequality is also true, but uses different tools. Under the Riemann Hypothesis, the right-hand side in (1.4) becomes $\frac{1}{2}$. In fact, we get a more precise version of the Prime Number Theorem with $\pi(x) = \mathrm{Li}(x) + O(\sqrt{x}\log x)$.

David Platt and Timothy Trudgian have numerically verified this year in [24] that every $\rho = \beta + i\gamma$ with $\zeta(\rho) = 0$, $0 < \beta < 1$ and $|\gamma| \leqslant 3 \cdot 10^{12}$ are on the critical line and they counted more than 12 trillion of them. This means that even if $\alpha$ seems to stay concentrated between 0.3 and 0.36 in Figure 1.2[2], we should expect that $\alpha$ will take values which are close to $\frac{1}{2}$ for larger values of $x$.

We may also derive the weaker but simpler-to-compute estimate $\pi(x) \sim x/\log x$ from (1.3). Indeed, integration by parts gives

$$\mathrm{Li}(x) = \int_2^x \frac{\mathrm{d}t}{\log t} = \frac{x}{\log x} + \frac{2}{\log 2} + \int_2^x \frac{\mathrm{d}t}{\log^2 t} = \frac{x}{\log x} + O\left(\frac{x}{\log^2 x}\right), \tag{1.5}$$

where the bound $\int_2^x \frac{\mathrm{d}t}{\log^2 t} \ll \frac{x}{\log^2 x}$ can be found using l'Hôpital's rule on the ratio of the two sides and noting that it tends to 1. We can see in Figure 1.1 that we seem to have the inequality $\pi(x) > \frac{x}{\log x}$ for large enough $x$. This is shown by using exactly the same techniques as in (1.5) to find that for $x \geqslant 2$, we have

$$\pi(x) = \sum_{k \leqslant n} \frac{(k-1)!x}{\log^k x} + O_n\left(\frac{x}{\log^{n+1} x}\right).$$

By taking $n = 2$, we see that $\left(\pi(x) - \frac{x}{\log x}\right) \sim \frac{x}{\log^2 x}$ as $x \to \infty$. We also seem to have the inequality $\pi(x) < \mathrm{Li}(x)$ based on the Figure 1.1. However, John E. Littlewood showed in 1914 that this inequality is reversed infinitely many times as $x$ gets larger. We have yet to find the smallest $x_0$ such that $\pi(x_0) > \mathrm{Li}(x_0)$. In 1999, Carter Bays and Richard H. Hudson proved in [4] that there exists values of $x$ such that $\pi(x) > \mathrm{Li}(x)$ inside the interval

$$[1.398201 \times 10^{316}, 1.398244 \times 10^{316}].$$

Using the estimate $\pi(x) \approx \mathrm{Li}(x)$ to understand the rate of growth of the function $\pi$, we have that $(\mathrm{Li}(x))' = 1/\log x$. Since the primes have a way of behaving randomly, this could be an argument towards confirming Gauss' intuition since the proportion of primes around $x$ can be represented by

$$\frac{\pi(x+h) - \pi(x-h)}{2h}.$$

If $\pi$ would have been a differentiable function, then this value could be approximated by the derivative of $\pi$ for a small enough $h > 0$.

## 1.4. Cramér's model for primes

Let $p_1 < p_2 < \ldots$ be the list of prime numbers in ascending order. Using the prime number theorem and the fact that $\pi(p_n) = n$, we can find a formula approximating the $n^{\text{th}}$ prime. First, we need $x$ to be written as a function of $\pi(x)$

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

[2] The graph seems to be oscillating around the smooth curve $\frac{\log(\frac{1}{2}\mathrm{Li}(x^{1/2}) + \frac{1}{3}\mathrm{Li}(x^{1/3}))}{\log x}$ which is asymptotically equal to $\frac{1}{2} - \frac{\log\log x}{\log x} + O\left(\frac{1}{\log^2 x}\right)$. We will give more details about this in Remark 3.3.

where the error term has been shifted to the other side by using the fact that for small enough $y$, we have $\frac{1}{1+y} = 1 + O(y)$. We wish we had $\log \pi(x)$ instead of $\log x$, but we can correct this by using the prime number theorem again and get

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

and replace $\log x$ by $\log \pi(x)$ and find

$$x = \pi(x) \log \pi(x) \left(1 + O\left(\frac{\log\log x}{\log x}\right)\right) \implies p_n = n \log n \left(1 + O\left(\frac{\log\log p_n}{\log p_n}\right)\right).$$

Finally, since we always have that $n \leqslant p_n$ and that $\frac{\log\log x}{\log x}$ is decreasing, we get the formula $p_n = n(\log n + O(\log\log n))$. This gives us an approximate formula for the $n^{\text{th}}$ prime.

By an easy telescopic argument, we know that the gaps between the first $N$ primes are on average

$$\frac{1}{N-1} \sum_{n=1}^{N-1} (p_{n+1} - p_n) = \frac{p_N - 2}{N-1} = \log N + O(\log\log N). \tag{1.6}$$

Harald Cramér was investigating what the true order of $\max_{n \leqslant x}(p_{n+1} - p_n)$ is. He conjectured that

$$\max_{n \leqslant x}(p_{n+1} - p_n) \sim \log^2 x$$

based on a probability model of the primes numbers (see Proposition 29.1 of [16]).

This model is defined by considering a sequence $X_n$ of independent Bernoulli variables such that $\mathbb{P}(X_n = 1) = \frac{1}{\log n}$ when $n \geqslant 3$ (we can also set $X_1 = 0$ and $X_2 = 1$ almost surely). This sequence models the indicator function of the prime numbers. We can think of the sequence $a_n$ such that $a_n = 1$ if $n$ is prime and 0 otherwise as a "typical" outcome of this probability space. Notice that this model is in tune with what Gauss conjectured in his teenage years.

The random function $\Pi(x) := \sum_{n \leqslant x} X_n$ simulates the prime-counting function $\pi(x)$. By summation by parts, we have

$$\mathbb{E}[\Pi(x)] = \sum_{n \leqslant x} \mathbb{E}[X_n] = 1 + \sum_{2 < n \leqslant x} \frac{1}{\log n} = \mathrm{Li}(x) + O(1),$$

where the main term is exactly the smooth function we have estimated $\pi(x)$ with previously. Using the independence of the random variables, we find the variance of $\Pi(x)$ to be

$$\mathrm{Var}(\Pi(x)) = \sum_{n \leqslant x} \mathrm{Var}(X_n) = \sum_{2 < n \leqslant x} \left(\frac{1}{\log n} - \frac{1}{(\log n)^2}\right) = \frac{x}{\log x} + O\left(\frac{x}{\log^2 x}\right),$$

where the last equality follows again by partial summation. Using the Law of the Iterated Logarithm (LIL), we get the bound on the error term: $|\Pi(x) - \mathrm{Li}(x)| \ll (x \log\log x / \log x)^{1/2}$ almost surely as $x \to \infty$. So the Riemann Hypothesis holds almost surely for Cramér's model.

However, this model has its limitations. For any fixed $k \geqslant 1$, we have $\mathbb{E}[\sum_{n\leqslant x} X_n X_{n+k}] \sim \frac{x}{\log^2 x}$. When $k = 1$, this absurdly suggests that we should expect infinitely many pairs of prime numbers differing by one.

If we use Cramér's model to study the distribution of primes in the congruence class $a \pmod q$, we find that the primes are distributed evenly between every congruence class mod $q$. However, this is obviously false, and we do not have to look further than $q = 2$: Every prime is odd except the number 2, so the class $1 \pmod 2$ contains almost all primes. Andrew Granville suggested a refinement of the model that considers divisibility by small primes (in this case, prime factors of $q$). In Chapter 5, we will see an application of this new model to conjecture an asymptotic for the Sophie Germain primes.

## 1.5. Primes in arithmetic progressions

If we define

$$\pi(x; q, a) := \#\left\{p \leqslant x : p \equiv a \pmod q\right\},$$

then the *Prime Number Theorem for arithmetic progressions* describes the equidistribution of primes amongst the reduced residue classes mod $q$. The theorem can be stated as follows: If $x, q \geqslant 2$ and $(a, q) = 1$, then there exists an absolute positive constant $c$ such that

$$\frac{\pi(x; q, a)}{\pi(x)} = \frac{1}{\phi(q)} + O_q\Big(e^{-c\sqrt{\log x}}\Big). \tag{1.7}$$

Under a certain generalization of the Riemann Hypothesis called the *Generalized Riemann Hypothesis*, which we will explain in the next chapter, we can have the better bound $\ll \frac{\log^2 x}{\sqrt{x}}$ uniformly for $q \leqslant x$ in the error term in (1.7). It shows that, for a fixed $q$, the larger $x$ gets, the closer the primes less than $x$ get to be evenly distributed between the reduced residue classes.

By studying the function $E(x; q, a, b) := \frac{\log x}{\sqrt{x}}(\pi(x; q, a) - \pi(x; q, b))$, the Prime Number Theorem for arithmetic progressions under the Generalized Riemann Hypothesis only tells us that

$$E(x; q, a, b) \ll \log^2 x$$

for $2 \leqslant q \leqslant x$ and $a, b$ relatively prime to $q$. However, in a letter to M. Fuss in 1853, P. L. Chebyshev noticed that the quantity $E(x; 4, 3, 1)$ tends to oscillate around 1. This is unexpected if we thought that primality did not follow any arithmetic structure besides the fact that all the large primes resides in reduced residue classes, and are otherwise equidistributed. Chebyshev's remark leads us to conjecture that $\pi(x; 4, 3) > \pi(x; 4, 1)$ for more than half of the $x$ and this puts an asterisk on the uniform distribution that the Prime Number Theorem for arithmetic progressions gives. We can note that the normalisation $\log x/\sqrt{x}$ in the definition of $E(x; q, a, b)$ represents the order of magnitude of the standard deviation of $\pi(x;4,3)-\pi(x;4,1)$, and its expected value for that matter, in a particular probability space. The tendency of primes to be more frequent in one residue class over another is often called in the literature: *Chebyshev's bias.* Michael Rubinstein and Peter Sarnak proved in 1994, under the Generalized Riemann Hypothesis and the Linear Independence Hypothesis (which we will explain in Chapter 3), that we can measure the set of $x$ such that $\pi(x;4,3)>\pi(x;4,1)$ using the logarithmic density:

$$\lim_{X\to\infty}\frac{1}{\log X}\int_{[1,X]\cap S_0}\frac{\mathrm{d}t}{t}=0.9959\ldots\quad \text{where } S_0=\{x\geqslant 1:\pi(x;4,3)>\pi(x;4,1)\}.$$

This shows that $\pi(x;4,3)-\pi(x;4,1)$ changes signs infinitely many times since $\sup S_0=\sup S_0^c=+\infty$. As we will see in Chapter 3, this problem is intimately linked to Littlewood's problem about the sign of $\pi(x)-\mathrm{Li}(x)$.

More generally, let $\pi_A(x)=\#\{p\leqslant x:p\in A\}$ be the cumulative distribution function of primes in $A$, a subset of $\mathbb{N}$. The study of the sign of the quantity $\pi_A(x)-\pi_B(x)$ for two different natural number subsets $A,B$ is called a *prime number race.* It is as if, at every milestone $x$, the two sets are racing by trying to have the most primes below $x$. In Chebyshev's observation, the primes of the form $4n+3$ are more often in the lead against the primes $4n+1$.

This thesis will take a deep dive in understanding where Chebyshev's bias comes from and numerically comparing the $4n+1$ and $4n+3$ race to another race involving the Sophie Germain primes, which are the primes $p$ with $2p+1$ also prime, against its sibling sequence, the primes $p$ such that $2p-1$ is also prime. However, since so little is known about these last two sequences (we don't even know if there are infinitely many of them), we will also have to form conjectures to understand what their asymptotics might be.

From now on in this thesis, the primes $p$ such that $2p+1$ is also prime will be called the *Sophie Germain primes of the first kind*[3] and their counting function will be denoted by

$$\pi_+(x):=\{p\leqslant x:2p+1\text{ is also prime}\}.$$

Similarly, the primes $p$ such that $2p-1$ is also prime will be referred to as the *Sophie Germain primes of the second kind*, and their counting function will be denoted by

$$\pi_-(x):=\{p\leqslant x:2p-1\text{ is also prime}\}.$$

Studying the race between these two types of primes means studying the size and sign of $\pi_+(x)-\pi_-(x)$.

[3] A finite sequence of primes $(p_n)$ such that $p_{n+1}=2p_n+1$ is called a *Cunningham chain of the first kind.* If the condition were $p_{n+1}=2p_n-1$ instead, then it would be called a *Cunningham chain of the second kind.* This is why we use this terminology for Sophie Germain primes. It is believed that there are infinitely many Cunningham chains of both kind having length $k$ for any $k\geqslant 1$.

## 1.6. Structure of the thesis

Chapter 2 intends to serve as an introduction to many definitions and tools of classical analytic number theory. In particular, we dive into the study of Dirichlet series, and we present another version of the explicit formula (1.2) which is appropriate to prove the Prime Number Theorem for arithmetic progressions under a generalization of the Riemann hypothesis.

Chapter 3 explores the reason for Chebyshev's bias by using the explicit formula to rigorously explain why quadratic residues (mod $q$) have less "chance" of being primes than quadratic nonresidues (mod $q$) for $q = 4$ or any odd prime. This chapter will also be where we will discuss Littlewood's problem about the sign of $\pi(x) - \mathrm{Li}(x)$.

The next three chapters (Chapters 4, 5 and 6) will be devoted to convince the reader of the following conjecture:

**Conjecture 1.1.** *As $x \to \infty$, we have*

$$\pi_+(x) \sim \pi_-(x) \sim \frac{c_2 x}{\log^2 x}$$

*where*

$$c_2 = 2\prod_{p \geqslant 3}(1 - \frac{1}{(p-1)^2}) \approx 1.32032\ldots$$

The constant $c_2$ is often referred as the *twin prime constant* since the function $\pi_2(x) := \#\{p \leqslant x : p + 2 \text{ is prime}\}$ shares exactly the same conjectured asymptotic as $\pi_+(x)$ and $\pi_-(x)$. We will focus on the study of $\pi_+(x)$; all the results obtained can be transferred easily to $\pi_-(x)$.

In Chapter 4, we will prove a nontrivial upper bound on $\pi_+(x)$ using Selberg's sieve method, which is four times higher than the conjectured size.

In Chapter 5, we explain Conjecture 1.1 using Granville's refinement of Cramér's model with another sieving technique, which considers the divisibility by small primes.

In Chapter 6, we will explain the circle method developed by G. H. Hardy and J. E. Littlewood by changing the problem into the evaluation of an integral over the unit circle $\{z \in \mathbb{C} : |z| = 1\}$. Hardy and Littlewood conjectured that the main contribution of this integral comes from the *major arcs*, which consist of the $z$ in the unit circle such that $\arg(z)/\pi$ is close to a rational number with low denominator. We will prove that the major arcs' contribution is of the same size as the asymptotic in Conjecture 1.1. The rest of the arcs on the unit circle are called *minor arcs*. In certain different settings, one could prove that the minor arcs' contribution is negligible compared to the contribution from the major arcs. However, in our case, we will explain why minor arcs cannot be dealt with in the same manner.

Finally, in Chapter 7, we will present some figures graphing the prime number race for Sophie Germain primes and studying its periodicity. The graph for the Sophie Germain prime race will be closer to a random walk than to Chebyshev's prime number race.

The numerical constants in this paper have been calculated using Wolfram Mathematica 12.1 or MATLAB R2020a, and the graphs have been generated with MATLAB R2020a or Maple 2018.

Overall, this thesis will explore techniques to find or hypothesize densities of some subsets of natural numbers, especially types of prime numbers, to ultimately study the distribution of these seemingly random objects.

# Chapter 2

# The explicit formula

The absence of a simple formula dictating what would be the $n$th prime could blur our intuition about prime numbers and their structure. This chapter intends to introduce classical methods and tools used frequently in analytic number theory to get a clearer sight of the distribution of prime numbers.

## 2.1. Primes in arithmetic progressions

The positive integers are partitioned into the $q$ distinct congruence classes mod $q$ in a way such that the first $N$ numbers are close to being evenly distributed among the congruence classes: if we fix $a$ and $q$ such that $0 \leqslant a < q$, then we can find with simple combinatorial arguments that

$$\frac{\#\left\{n \leqslant N : n \equiv a \pmod{q}\right\}}{\#\{n \leqslant N\}} = \frac{1}{q} + O\left(\frac{1}{N}\right).$$

The way prime numbers split among the different congruence classes is not as obvious. First, any $n \in \mathbb{N}$ with $n \equiv a \pmod{q}$ is divisible by common factors of both $a$ and $q$. Therefore, a prime $p \equiv a \pmod{q}$ such that $(a, q) > 1$ would have to be divisible by $(a, q)$ and this means that there can only be at most one prime number in these particular congruence classes.

For the classes $a \pmod{q}$ with $a$ and $q$ relatively prime, Dirichlet proved in his 1837 memoir that they all contain infinitely many primes. His work is considered by many to mark the birth of analytic number theory.

As a matter of fact, the primes are equidistributed among the different reduced residue classes mod $q$. By this, we mean that if we fix $a$ and $q$ such that $(a, q) = 1$ and we let $\pi(x; q, a)$ be the number of prime numbers less than $x$ in the congruence class $a \pmod{q}$, then it is known that

$$\frac{\pi(x; q, a)}{\pi(x)} \to \frac{1}{\phi(q)} \quad \text{as } x \to \infty$$

where $\phi(q)$ is *Euler's totient function* returning the number of positive integers up to $q$ which are relatively prime to $q$. This result is called the *Prime Number Theorem for arithmetic progressions.*

A quantitative form of the prime number theorem in arithmetic progressions is the *Siegel-Walfisz theorem* which states that if we let $A > 0$, then there exists an absolute constant $c > 0$ such that for $1 \leqslant q \leqslant (\log x)^A$ and $(a, q) = 1$, we have

$$\pi(x; q, a) = \frac{\mathrm{Li}(x)}{\phi(q)} + O_A\Big(xe^{-c\sqrt{\log x}}\Big). \tag{2.1}$$

This theorem lets $q$ vary, and the bound on the error term is uniform with respect to $q$ inside the set of integers in the interval $[1, (\log x)^A]$.

In the following sections, we will prove a stronger result, but it will be entirely depending on a deep conjecture.

**Theorem 2.1.** *Let $A > 0$. Under the Generalized Riemann Hypothesis, we have*

$$\pi(x; q, a) = \frac{\mathrm{Li}(x)}{\phi(q)} + O_A\Big(\sqrt{x}\log x\Big)$$

*uniformly for $q \ll_A x^A$ and $(a, q) = 1$.*

The Riemann Hypothesis is arguably one of the most famous unsolved problems in all of modern pure mathematics. We will explain what this conjecture and its generalization mean at the end of Section 2.5, and we will show how they are intimately linked to the distribution of prime numbers.

This chapter will present the outline of the classical method to prove Theorem 2.1. It has multiple steps, and it will introduce a new set of tools from complex analysis, which will be useful in the rest of this memoir. We will define the Dirichlet $L$-functions and explain how a linear combination of them encodes the distribution of primes in the reduced arithmetic progressions of a modulus $q$. We will give an explicit formula to a function related to $\pi(x; q, a)$ in terms of the zeros of the $L$-functions. This will be the key to understand Chebyshev's bias in the next chapter.

## 2.2. Dirichlet series

Our goal will be to give an estimate for

$$\pi(x; q, a) = \sum_{\substack{p \leqslant x \\ p \equiv a \ (\mathrm{mod}\ q)}} 1.$$

This represents the sum up to $x$ of the sequence $(b_n)_{n\in\mathbb{N}}$ defined by

$$b_n = \begin{cases} 1 & \text{if } n \text{ is a prime congruent to } a \pmod{q}, \\ 0 & \text{otherwise.} \end{cases}$$

We cannot estimate the sum $\sum_{n\leqslant x} b_n$ by approximating it with an integral using partial summation since the terms $b_n$ are far from varying smoothly. However, another way to study sequences and their partial sums in number theory and combinatorics is to find and study a generating function of the sequence $b_n$ instead.

Let $f$ be an *arithmetic function* with *polynomial growth*, meaning that $f\colon \mathbb{N} \to \mathbb{C}$ and that $f(n) \ll n^k$ for some $k \in \mathbb{N}$. We can define the function

$$F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}$$

in the region of convergence of the series. This function is called the *Dirichlet series* of $f$.

If $f(n) \ll n^k$, then for $s$ such that $\sigma > k+1$ we have $f(n)n^{-s} \ll n^{k-\sigma}$, which means that $F(s)$ is absolutely convergent. Conversely, if there exists $s$ such that $F(s)$ converges absolutely, then the main term $f(n)n^{-s} \to 0$, which implies that $f(n) \ll n^{\sigma}$. Thus we can conclude that having a function $f(n)$ with polynomial growth is equivalent to saying that its Dirichlet series converges absolutely for some $s \in \mathbb{C}$.

Moreover, a special class of arithmetic functions are called *multiplicative function*. They are the functions $f$ such that $f(1) = 1$ and $f(n{\cdot}m) = f(n)f(m)$ whenever we have $(n,m) = 1$. It is known that we can rewrite $F(s)$ as a product over primes:

$$F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s} = \prod_p \left(1 + \frac{f(p)}{p^s} + \frac{f(p^2)}{p^{2s}} + \dots\right), \tag{2.2}$$

whenever the series is absolutely convergent. The right-hand side of (2.2) is called the *Euler product* of $F(s)$. This equation can be proved using the fundamental theorem of arithmetic.

If $f$ is *completely multiplicative*, that is to say, a function respecting $f(1) = 1$ and the property $f(n \cdot m) = f(n)f(m)$ without having any restriction on $n$ and $m$, then

$$F(s) = \prod_p \left(1 + \frac{f(p)}{p^s} + \left(\frac{f(p)}{p^s}\right)^2 + \dots\right) = \prod_p \left(1 - f(p)p^{-s}\right)^{-1} \tag{2.3}$$

in the region where the series converges absolutely since each term in the product becomes a geometric series.

***Example* 2.2.** One crucial example of a Dirichlet series is the *Riemann zeta function*

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_p \left(1 - \frac{1}{p^s}\right)^{-1}, \tag{2.4}$$

defined for $\sigma > 1$. It is the Dirichlet series of the constant function 1 and is the key to understand the distribution of prime numbers.

By applying the logarithmic function on (2.3), we can transform the Euler product into a sum over prime powers:

$$G(s) = \log F(s) = \sum_p \log(1 - f(p)p^{-s})^{-1} = -\sum_p \sum_{k \geqslant 1} \frac{f(p)^k}{kp^{ks}},$$

where the last equality is the expansion of the logarithm into its Taylor series (this can only be done when $\sigma$ is big enough so that $|f(p)p^{-s}| < 1$ for all primes $p$). Then $G(s)$ is the Dirichlet series of

$$g(n) = \begin{cases} -\frac{f(p)^k}{k} & \text{if } n = p^k \text{ is a prime power} \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

Like Cauchy's integral formula for the power series, we will eventually need an inversion formula involving the Dirichlet series to extract useful information on the sequence it represents. We will see in Section 2.4 that this inversion formula will be an integral over a vertical line of the complex plane. However, in practice, this integral is not easy to compute directly, and we will need the integrand to be meromorphic to use the residue theorem from complex analysis and give a value to the integral.

Each term of the series $F(s)$ is an analytic function of the form $s \mapsto f(n)n^{-s} = f(n)e^{-s\log n}$ since the exponential function have a power series representation which is convergent on the whole complex plane. To determine a region where $F(s)$ is analytic, it suffices to show that the series converges uniformly on every compact subsets of the region[1].

**Theorem 2.3.** *Let $\sum_{n=1}^{\infty} f(n)n^{-s_0}$ be a convergent series, $E = \{s \in \mathbb{C} : \sigma > \sigma_0\}$ and $E_1 = \{s \in \mathbb{C} : \sigma > \sigma_0 + 1\}$. Then the Dirichlet series $F(s) = \sum_{n=1}^{\infty} f(n)n^{-s}$ converges uniformly on every compact subsets of $E$ implying that $F(s)$ is analytic on $E$. In particular, the series $F(s)$ converges absolutely for $s \in E_1$.*

***Proof.*** Let $K$ be a compact subset of $E$, and let $\varepsilon > 0$. Since $F(s_0)$ is convergent, there exists $N_0$ such that for every $M > N \geqslant N_0$, we have

$$\left| \sum_{N < n \leqslant M} \frac{f(n)}{n^{s_0}} \right| < \varepsilon.$$

[1] See Theorem 1 of [2] p. 176 for a proof of this.

If we let $S(t) = \sum_{N_0 < n \leqslant t} f(n)n^{-s_0}$, then $|S(t)| < \varepsilon$ for every $t \geqslant 0$. By using the Riemann-Stieltjes integral and integrating by parts, we find that

$$\begin{aligned}\left|\sum_{N<n\leqslant M} f(n)n^{-s}\right| &= \left|\int_N^M \frac{\mathrm{d}S(t)}{t^{s-s_0}}\right| \\ &= \left|\frac{S(M)}{M^{s-s_0}} - \frac{S(N)}{N^{s-s_0}} + (s-s_0)\int_N^M S(t)t^{s_0-s+1}\,\mathrm{d}t\right| \leqslant \varepsilon\left(2 + \frac{|s-s_0|}{\sigma-\sigma_0}\right).\end{aligned}$$

The quantity $\left(2 + \frac{|s-s_0|}{\sigma-\sigma_0}\right)$ is uniformly bounded for $s \in K$ because of the compactness of $K$. Therefore, we can conclude that we have uniform convergence of the series on $K$.

Finally, if $F(s_0)$ converges, then $|f(n)n^{-s_0}| \to 0$, which implies that $f(n) \ll n^{\sigma_0}$. For $s \in E_1$, we have

$$\left|\frac{f(n)}{n^s}\right| \ll \frac{1}{n^{\sigma-\sigma_0}}$$

implying that $F(s)$ absolutely converges. □

Consequently, we define the *abscissa of convergence*

$$\sigma_c = \inf\left\{\sigma \in \mathbb{R} : \text{the series } F(\sigma) \text{ converges}\right\}$$

and the *abscissa of absolute convergence*

$$\sigma_a = \inf\left\{\sigma \in \mathbb{R} : \text{the series } F(\sigma) \text{ converges absolutely}\right\}.$$

We can be sure by Theorem 2.3 that $F(s)$ diverges for any $\sigma < \sigma_c$, converges conditionally[2] for $\sigma \in (\sigma_c, \sigma_a)$, converges absolutely for $\sigma > \sigma_a$, that $\sigma_a - \sigma_c \in [0,1]$ and that $f$ with polynomial growth is now equivalent to having $\sigma_c < +\infty$.

We know that for a complex number $a$, if two Taylor series at $a$ are equal on a set that has a limit point, then the Taylor series's coefficients are equal. An analogous uniqueness theorem exists for Dirichlet series:

**Theorem 2.4** (Uniqueness Theorem)**.** *If $f_1$ and $f_2$ are two arithmetic functions with polynomial growth and*

$$\sum_{n=1}^{\infty} \frac{f_1(n)}{n^{s_k}} = \sum_{n=1}^{\infty} \frac{f_2(n)}{n^{s_k}}$$

*for a sequence $s_k$ such that $\sigma_k \to +\infty$, then $f_1 = f_2$.*

***Proof.*** Aiming for a contradiction, let's suppose that $f_1 \neq f_2$. Let's define $g\colon \mathbb{N} \to \mathbb{C}$, $\alpha, N > 0$ such that

$$g(n) = f_1(n) - f_2(n), \quad g(n) \ll n^{\alpha} \quad \text{and} \quad N = \min\left\{n : g(n) \neq 0\right\}.$$

[2] The interval may be empty if $\sigma_c = \sigma_a$.

Let $G(s)$ be the Dirichlet series of $g$. For every $k \in \mathbb{N}$, we are supposed to have $G(s_k) = 0$ by hypothesis. On the other hand, however, we have

$$\sum_{n>N} \frac{g(n)}{n^{s_k}} \ll \sum_{n>N} \frac{1}{n^{\sigma_k - \alpha}} \ll \int_N^\infty \frac{\mathrm{d}t}{t^{\sigma_k-\alpha}} \ll_{\alpha,N} \frac{1}{\sigma_k N^{\sigma_k}},$$

and this implies that

$$N^{s_k} G(s_k) = N^{s_k} \sum_{n \geqslant N} \frac{g(n)}{n^{s_k}} = g(N) + O_{k,N}\left(\frac{1}{\sigma_k}\right) \implies \lim_{k\to\infty} N^{s_k} G(s_k) = g(N).$$

This contradicts $g(N) \neq 0$. □

We have proved that $F(s)$ is analytic on an open set as long as the Dirichlet series converges at some point of the complex plane. Using analytic continuation, we can generally extend the definition of $F(s)$ to points where the series does not converge. However, when it comes to $\log F(s)$, the complex logarithm is not a meromorphic function, and we have to take into account which branch of the logarithm we are working with. An alternative would be to work with the logarithmic derivative $(\log F(s))' = F'(s)/F(s)$, which is much easier to use from an analytic point of view.

If we have again the series $F(s) = \sum_{n=1}^\infty f(n) n^{-s}$, then by differentiating term by term we get the series $-\sum_{n=1}^\infty (f(n)\log n) n^{-s}$. Since $-f(n)\log n$ is of polynomial growth if $f$ is, then we can conclude that the series $-\sum_{n=1}^\infty (f(n)\log n) n^{-s}$ is a well-defined function for at least one point of the complex plane. Using Theorem 2.3 again, we can find an open region of the complex plane where this series converges uniformly. Therefore, we are allowed to say that when differentiating a Dirichlet series which converges at some point, it suffices to multiply every term by $-\log n$.

Coming back to the case where $f$ is completely multiplicative, we can find the coefficients of $F'(s)/F(s)$ by differentiating $\log F(s)$ in a region where $\sigma > \frac{\log|f(p)|}{\log p}$ for all $p$. By multiplying the coefficients (2.5) by $-\log n$, we get the coefficients of $-F'(s)/F(s) = \sum_{n=1}^\infty f(n)\Lambda(n) n^{-s}$ where

$$\Lambda(n) = \begin{cases} \log p & \text{if } n = p^k, \\ 0 & \text{otherwise.} \end{cases}$$

The function $\Lambda$ is commonly called in number theory the *von Mangoldt function*. What we have done here is defining another function $\Lambda$, which is only supported on prime powers and which can be used to filter the values of $f$ at prime powers. The arithmetic function $f \cdot \Lambda$ has an easy-to-use Dirichlet series given by $-F'(s)/F(s)$.

Instead of studying the sum

$$\pi(x; q, a) = \sum_{\substack{p \leqslant x \\ p \equiv a \ (\mathrm{mod}\ q)}} 1,$$

it might be easier to study the function

$$\psi(x;q,a) := \sum_{\substack{n\leqslant x \\ n\equiv a \ (\mathrm{mod}\ q)}} \Lambda(n).$$

This is *Chebyshev's psi function* for the arithmetic progression $a$ (mod $q$). We will see that since the powers greater or equal to 2 of the primes are sparser than the primes themselves, their contribution to the sum will be negligible. The logarithmic weight varies slowly, so we can remove it using partial summation and thus go from estimates on $\psi(x;q,a)$ to estimates on $\pi(x;q,a)$ (and vice versa).

## 2.3. Dirichlet convolution

We take a detour from understanding primes in arithmetic progressions to explain how multiplication of Dirichlet series works and develop an essential tool that will be used several times in the next chapters.

The set of power series is closed under multiplication, but multiplying together two power series is not as simple as multiplying the coefficients. For arithmetic functions $f$ and $g$ such that their respective power series both converge in a specific open set $U \subset \mathbb{C}$, then for $x \in U$, we have

$$\left(\sum_{n=0}^{\infty} f(n)x^n\right) \cdot \left(\sum_{n=0}^{\infty} g(n)x^n\right) = \sum_{n=0}^{\infty} h(n)x^n \quad \text{where } h(n) := \sum_{\substack{0\leqslant a,b\leqslant n \\ a+b=n}} f(a)g(b). \tag{2.6}$$

An analogous version can be given for functions on the real line. Let $\mathcal{F}(f)$ denote the Fourier transform of a function $f$, then $\mathcal{F}(f)\mathcal{F}(g) = \mathcal{F}(h)$ where

$$h(x) = \int_{-\infty}^{\infty} f(t)g(x-t)\,\mathrm{d}t.$$

Those two operations are often referred to as the *convolution* for power series[3] and the Fourier transform, respectively.

If $F(s)$ and $G(s)$ are the Dirichlet series of $f(n)$ and $g(n)$ respectively, then $F(s)+G(s)$ is the Dirichlet series of $f(n)+g(n)$ as expected. However, for the region where both series $F(s)$ and $G(s)$ are absolutely convergent, we have

$$F(s)\cdot G(s) = \left(\sum_{a=1}^{\infty} \frac{f(a)}{a^s}\right) \cdot \left(\sum_{b=1}^{\infty} \frac{g(b)}{b^s}\right) = \sum_{a=1}^{\infty}\sum_{b=1}^{\infty} \frac{f(a)g(b)}{(ab)^s} = \sum_{n=1}^{\infty} \frac{(f*g)(n)}{n^s}$$

$$\text{where } (f*g)(n) := \sum_{\substack{1\leqslant a,b\leqslant n \\ ab=n}} f(a)g(b) = \sum_{d\mid n} f(d)g(n/d). \tag{2.7}$$

[3] Sometimes, the convolution for the power series is sometimes called *Cauchy product*.

The function $f * g$ is called the *Dirichlet convolution.* It is a multiplicative analogue to the function $h$ defined in (2.6). With the Dirichlet convolution, the arithmetic functions with polynomial growth have a ring structure, making the Dirichlet convolution easy to manipulate.

**Theorem 2.5.** *Let $\mathcal{A} = \{f : \mathbb{N} \to \mathbb{C}$ with polynomial growth$\}$. Then the triplet $(\mathcal{A}, +, *)$ forms a commutative ring with unity $\varepsilon(n) = \mathbf{1}_{n=1}$. In addition, the set of multiplicative functions in $\mathcal{A}$ form a subgroup of the unit group $\mathcal{A}^*$.*

***Proof.*** Let $\mathcal{D}$ be the set of Dirichlet series converging somewhere on the complex plane. The fact that $(\mathcal{D}, +, \cdot)$ is a commutative ring is obvious. This ring's unity is the constant function 1, which could be represented as the Dirichlet series of $\varepsilon$. Let $\varphi : \mathcal{A} \to \mathcal{D}$ be the map defined by

$$\varphi(f) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}.$$

Then $\varphi(f+g) = \varphi(f) + \varphi(g)$, $\varphi(f * g) = \varphi(f) \cdot \varphi(g)$ because of (2.7) and $\varphi(\varepsilon) = 1$. The map $\varphi$ is bijective and its inverse $\varphi^{-1}$ is well-defined by the Uniqueness Theorem (Theorem 2.4). All the commutative ring axioms are respected with the triplet $(\mathcal{A}, +, *)$ since they follow the commutative ring axioms of $(\mathcal{D}, +, \cdot)$ via the bijective map $\varphi$. Therefore, the map $\varphi$ is a ring isomorphism between $\mathcal{A}$ and $\mathcal{D}$.

Finally, $f : \mathbb{N} \to \mathbb{C}$ being multiplicative and having polynomial growth is equivalent to having an Euler product representation for its Dirichlet series

$$F(s) = \prod_p \sum_{k=0}^{\infty} f(p^k)(p^{-s})^k \tag{2.8}$$

which implies that

$$\frac{1}{F(s)} = \prod_p \left( \sum_{k=0}^{\infty} f(p^k)(p^{-s})^k \right)^{-1} = \prod_p \sum_{k=0}^{\infty} g(p^k)(p^{-s})^k$$

where $g(1) = 1$ and we can define recursively

$$g(p^k) = -\sum_{j=0}^{k-1} f(p^{k-j}) g(p^j)$$

for $k \geqslant 1$. Since the product of two Euler products of the form (2.8) can also be represented as an Euler product by using (2.6) on the inner sum, the multiplicative functions with polynomial growth form a subgroup of $\mathcal{A}^*$. □

We give below some important examples of Dirichlet convolution used in the rest of the thesis.

***Example* 2.6.** Let $\tau_1$ be the constant function 1 and $\tau_n = \tau_{n-1} * 1$ for $n \geqslant 2$. The function $\tau_2$ is usually noted $\tau$ and is called the *divisor function*. This is because $\tau(n)$ counts the number of divisors of $n$. Generally, $\tau_k(n)$ counts the number of ways to write $n$ as a product of $k$ positive integers since

$$\tau_k(n) = \sum_{a_1 \cdots a_k = n} 1.$$

Its Dirichlet series is $\zeta(s)^k$.

***Example* 2.7** (Möbius inversion formula)**.** To find the Dirichlet inverse of the constant function 1, we can use the Euler product of the Riemann zeta function. For $\sigma > 1$,

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

which is the Euler product of the multiplicative function $\mu$ supported on squarefree integers defined by $\mu(p) = -1$ for every prime $p$. This function is called the *Möbius function*.

Since $\mu$ is the inverse of 1, we have the identity

$$\mathbf{1}_{n=1} = \varepsilon(n) = (1 * \mu)(n) = \sum_{d|n} \mu(d). \tag{2.9}$$

This is the *Möbius inversion formula*. One main application of this formula is to extract from a sum $\sum a_n$ the terms such that $n$ is coprime to some positive integer $m$: We have

$$\sum_{n\colon (n,m)=1} a_n = \sum_n a_n \sum_{d|(n,m)} \mu(d) = \sum_n a_n \sum_{\substack{d|n \\ d|m}} \mu(d) = \sum_{d|m} \mu(d) \sum_{n \equiv 0 \pmod d} a_n = \sum_{d|m} \mu(d) \sum_r a_{dr},$$

provided the series $\sum a_n$ absolutely converge. The last expression describes secretly the inclusion-exclusion principle from combinatorics. The above formula transforms a problem about a sum of terms with a coprimality condition into a problem about sums over arithmetic progressions. This is the base idea of sieve theory (see Chapter 4 and Section 5.2).

***Example* 2.8.** One idea where we can see the Möbius inversion formula in action is when studying Euler's totient function $\phi$. Let us recall that $\phi(m)$ counts the positive integers less than or equal to $m$ which are coprime to $m$. Then

$$\phi(m) = \sum_{n \leqslant m} \mathbf{1}_{(n,m)=1} = \sum_{n \leqslant m} \sum_{\substack{d|n \\ d|m}} \mu(d) = \sum_{d|m} \mu(d) \sum_{\substack{n \leqslant m \\ d|n}} 1 = \sum_{d|m} \mu(d) \frac{m}{d} = (\mu * \iota)(m)$$

where $\iota(n) = n$ is the inclusion map from $\mathbb{N}$ to $\mathbb{C}$. The Dirichlet series of $\iota$ is obviously $\zeta(s-1)$ which means that the Dirichlet series of $\phi$ is $\zeta(s-1)/\zeta(s)$. We can also find

$$\phi = \mu * \iota \implies \iota = 1 * \phi \implies n = \sum_{d|n} \phi(d),$$

which is useful is we use it to write the gcd of $n$ and $m$ as a sum over their common divisors:

$$(n,m)=\sum_{\substack{d\mid n\\ d\mid m}}\phi(d).$$

***Example* 2.9.** The last example involves the von Mangoldt function $\Lambda$ which we defined in the previous section. From what we have seen so far, we can say that the Dirichlet series of $\Lambda$ is $-\zeta'(s)/\zeta(s)$. Furthermore, differentiating a Dirichlet series means multiplying every coefficient by $-\log n$. For example, the Dirichlet series of log is $-\zeta'(s)$, whence

$$\zeta(s)\cdot\left(-\frac{\zeta'(s)}{\zeta(s)}\right)=-\zeta'(s)\implies 1*\Lambda=\log.$$

We could also prove this directly. If $\nu_p$ is a function defined as $\nu_p(n)=k$ where $p^k\parallel n$, then

$$\begin{aligned}(1*\Lambda)(n)&=\sum_{d\mid n}\Lambda(d)=\sum_{p^j\mid n}\log p=\sum_{p\mid n}\log p\sum_{j\leqslant\nu_p(n)}1\\&=\sum_{p\mid n}\nu_p(n)\log p=\log\left(\prod_{p\mid n}p^{\nu_p(n)}\right)=\log n.\end{aligned}$$

## 2.4. Perron's formula

For Dirichlet series to be useful in analytic number theory, we must be able to extract from them information about the sequence they represent, as it is for any type of generating functions. Let $F(s)=\sum_{n=1}^{\infty}f(n)n^{-s}$ be a Dirichlet series which is convergent at the point $s=c-2\pi i\xi$ where $c,\xi\in\mathbb{R}$ (we temporarily use this notation instead of the usual $s=\sigma+it$ for convenience). Our goal is to have a formula for $S(x)=\sum_{n\leqslant x}f(n)$. A clear link between $F$ and $S$ comes from partial summation that gives us the formula

$$F(s)=\sum_{n=1}^{\infty}\frac{f(n)}{n^s}=\int_0^{\infty}x^{-s}\,\mathrm{d}S(x)=s\int_0^{\infty}S(x)x^{-s-1}\,\mathrm{d}x.$$

This formula makes $F(s)/s$ the *Mellin transform*[4] of the function $S(1/x)$. The Mellin transform is the Fourier transform, written in a different form. To directly derive an inversion formula, we will make the change of variables $u=-\log x$:

$$\frac{F(c-2\pi i\xi)}{c-2\pi i\xi}=\int_{-\infty}^{\infty}S(e^{-u})e^{uc}e^{-2\pi iu\xi}\,\mathrm{d}u$$

We have that $\xi\mapsto\frac{F(c-2\pi i\xi)}{c-2\pi i\xi}$ is the Fourier transform of the function $u\mapsto S(e^{-u})e^{uc}$. Since $S(e^{-u})$ is a compactly supported step function, it is piecewise continuously differentiable function in $L^1(\mathbb{R})$, thus we can use Fourier's inversion formula:

$$S^*(e^{-u})e^{uc}=\lim_{T\to\infty}\int_{-T}^{T}\frac{F(c-2\pi i\xi)}{c-2\pi i\xi}e^{2\pi iu\xi}\,\mathrm{d}\xi.$$

[4] The Mellin transform of a function $f$ is defined to be $M_f(s)=\int_0^{\infty}f(x)x^s\,\mathrm{d}(\log x)=\int_0^{\infty}f(x)x^{s-1}\,\mathrm{d}x$.

The new function $S^*$ is simply redefined to satisfy the Fourier inversion formula by taking the average of the left and right limit at every discontinuity:

$$S^*(x) = \lim_{\varepsilon \to 0^+} \frac{S(x+\varepsilon) + S(x-\varepsilon)}{2}.$$

This means that $S^*(x) = S(x) - \frac{f(n)}{2}$ if $x \in \mathbb{N}$ and that $S^*(x) = S(x)$ otherwise. By changing back the variables in the integral from $\xi$ to $s$, then we get

$$S^*(e^{-u}) = \frac{1}{2\pi i} \lim_{T\to\infty} \int_{c-2\pi i T}^{c+2\pi i T} \frac{F(s)}{s} e^{-us} \,\mathrm{d}s.$$

Thus we have proved[5] that for every $x > 0$, we have

$$S^*(x) = \frac{1}{2\pi i} \int_{(c)} \frac{F(s)}{s} x^s \,\mathrm{d}s, \tag{2.10}$$

where $\int_{(c)}$ means that we are taking the limit as $T \to \infty$ of the integral along the path along the vertical segment $\gamma_{c,T}\colon [-T, T] \to \mathbb{C}$ defined by $\gamma_{c,T}(t) = c + it$. The equation (2.10) is known as *Perron's formula.*

To use this idea in practice, we need to evaluate the complex integral on the right-hand side of (2.10), and the main tool for this is Cauchy's residue theorem from complex analysis. We will approximate the integral over the vertical line by a contour integral over a closed loop.

As an example of a computation of Perron's formula, we follow a proof from §17 of [5] to give an integral which is a smooth approximation to the "almost" indicator function $\mathbf{1}^*_{(1,\infty)}(y)$ ("almost" in the sense that the function has the value $\frac{1}{2}$ at the point $y = 1$). This proposition will later be used in the proof of a truncated version of the integral in (2.10) (Lemma 2.14).

**Proposition 2.10.** *For $y, c, T > 0$, we have*

$$\left| \frac{1}{2\pi i} \int_{c-iT}^{c+iT} \frac{y^s}{s} \,\mathrm{d}s - \mathbf{1}^*_{(1,\infty)}(y) \right| \ll \begin{cases} \frac{y^c}{1+T|\log y|} & \text{if } y \neq 1, \\ \frac{c}{T} & \text{if } y = 1. \end{cases}$$

***Proof.*** If $y = 1$, then

$$\frac{1}{2\pi i} \int_{c-iT}^{c+iT} \frac{1}{s} \,\mathrm{d}s = \frac{1}{2\pi i} \int_{-T}^{T} \frac{i}{c+it} \,\mathrm{d}t = \frac{1}{2\pi i} \int_{-T}^{T} \frac{t+ic}{c^2+t^2} \,\mathrm{d}t = \frac{c}{\pi} \int_0^T \frac{\mathrm{d}t}{c^2+t^2}.$$

With the change of variables $u = t/c$, we get

$$= \frac{1}{\pi} \int_0^{T/c} \frac{\mathrm{d}u}{1+u^2} = \frac{1}{\pi} \int_0^{\infty} \frac{\mathrm{d}u}{1+u^2} + O\left( \int_{T/c}^{\infty} \frac{\mathrm{d}u}{u^2} \right) = \frac{1}{2} + O\left( \frac{c}{T} \right).$$

[5] These back and forth between variables $s$ and $\xi$ could be avoided by having in our arsenal the *Mellin inversion formula* $f^*(x) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} M_f(s) x^{-s} \,\mathrm{d}s$, which, as we just saw, is just a few changes of variables away from Fourier's inversion formula.

The main term in the last part of the equation is obtained by noticing that $\frac{1}{1+u^2}$ is the derivative of $\arctan(u)$.

If $y \neq 1$, the strategy is to shift the path in a part of the complex plane where the integral will be small; this way, the value of the integral is the sum of the residues of the poles that we encountered in the homotopy. Note that in our case, the meromorphic function $\frac{y^s}{s}$ has only one pole of residue 1 at $s = 0$.

In the case $0 < y < 1$, the integrand's size will be diminished if we push the path on the far right of the complex plane. However, we still have to take into account the fact that the endpoints of our path have to stay fixed in our homotopy, meaning that for all $M > c$, we have

$$\int_{c-iT}^{c+iT} \frac{y^s}{s}\,\mathrm{d}s = \int_{c-iT}^{M-iT} \frac{y^s}{s}\,\mathrm{d}s + \int_{M-iT}^{M+iT} \frac{y^s}{s}\,\mathrm{d}s - \int_{c+iT}^{M+iT} \frac{y^s}{s}\,\mathrm{d}s = \int_{c-iT}^{+\infty-iT} \frac{y^s}{s}\,\mathrm{d}s - \int_{c+iT}^{+\infty+iT} \frac{y^s}{s}\,\mathrm{d}s,$$

where the last part of the equality is due to the fact that $\frac{y^s}{s}$ tends to 0 as $\sigma \to +\infty$ uniformly with respect to the imaginary part of $s$ because

$$\left|\frac{y^s}{s}\right| \leqslant \frac{1}{\sigma}.$$

We can bound the last two integrals trivially

$$\left|\int_{c-iT}^{c+iT} \frac{y^s}{s}\,\mathrm{d}s\right| = \left|\int_{c-iT}^{+\infty-iT} \frac{y^s}{s}\,\mathrm{d}s - \int_{c+iT}^{+\infty+iT} \frac{y^s}{s}\,\mathrm{d}s\right| \leqslant \frac{2}{T}\int_c^{\infty} y^{\sigma}\,\mathrm{d}\sigma \leqslant \frac{2y^c}{|\log y|\,T}. \tag{2.11}$$

This bound is good, but we will eventually need a better one for values of $y$ close to 1. To remedy the situation, we can modify the path of the integral in another way. Let $C$ be the circle in the complex plane of radius $R = \sqrt{c^2 + T^2}$ centered at 0 and let $\gamma'$ be the circular arc of $C$ parametrized by arc length starting at $c - iT$, travelling counterclockwise and finishing at $c + iT$:

$$\left|\int_{c-iT}^{c+iT} \frac{y^s}{s}\,\mathrm{d}s\right| = \left|\int_{\gamma'} \frac{y^s}{s}\,\mathrm{d}s\right| \leqslant \frac{y^c}{R}\int_{\gamma'} \mathrm{d}s \leqslant \frac{y^c}{R}2\pi R \leqslant 2\pi y^c. \tag{2.12}$$

If $y > 1$, then we will use a similar strategy to bound the integral by modifying the contour in two ways by first pushing the integral to the left, and second, by travelling clockwise around the circle $C$ instead of counterclockwise. The only difference here is that we will consider the pole at $s = 0$ by adding its residue. □

The reason to use this approximation of an indicator function as an integral to prove Perron's formula becomes clear when looking at $S^*(x) = \sum_{n=1}^{\infty} a_n \mathbf{1}^*_{(1,\infty)}\left(\frac{x}{n}\right)$. In section 2.6, we are going to apply Proposition 2.10 in order to have an explicit formula for $\psi(x; q, a)$.

## 2.5. Dirichlet characters and $L$-functions

In general, the function $n \mapsto \mathbf{1}_{n\equiv a \pmod q}$ is not completely multiplicative, which means that we cannot do as we discussed at the end of Section 2.2 to find the Dirichlet series of

$$\Lambda(n)\mathbf{1}_{n\equiv a \pmod q}$$

since the indicator function is not completely multiplicative. The strategy will be to write down $\mathbf{1}_{n\equiv a \pmod q}$ as a linear combination of completely multiplicative functions.

Having a function being $q$-periodic and supported on integers relatively prime to $q$ is an advantage for us because it can be redefined on the unit group of the ring of congruence classes (mod $q$), which is usually denoted $(\mathbb{Z}/q\mathbb{Z})^*$. Looking at the group as if it was a probability space by assigning the same probability to each element of the group, then the complex function space $L^2((\mathbb{Z}/q\mathbb{Z})^*)$ containing every function $f\colon (\mathbb{Z}/q\mathbb{Z})^* \to \mathbb{C}$ is a Hilbert space with the inner product

$$\langle f, g\rangle = \frac{1}{\phi(q)} \sum_{a\in(\mathbb{Z}/q\mathbb{Z})^*} f(a)\overline{g(a)}.$$

This function space is of course of dimension $\phi(q)$ since the indicator functions $n \mapsto \mathbf{1}_{n\equiv a \pmod q}$ for $(a, q) = 1$ form a basis of $L^2((\mathbb{Z}/q\mathbb{Z})^*)$. Dirichlet introduced another orthonormal basis containing only completely multiplicative functions.

The *Dirichlet characters mod $q$* are arithmetic functions $\chi\colon \mathbb{Z} \to \mathbb{C}$ defined by being $q$-periodic completely multiplicative functions with $\chi(n) \neq 0$ if, and only if, $(n, q) = 1$. They can be seen as being the lifts to $\mathbb{Z}$ of the group homomorphisms $(\mathbb{Z}/q\mathbb{Z})^* \to \mathbb{C}^*$, where $\mathbb{C}^*$ is the multiplicative group of nonzero complex numbers. For example, the indicator function of the integers coprime to $q$ is a Dirichlet character. This particular character is called the *principal character mod $q$* and is denoted by $\chi_0$.

Let $\chi$ be a character mod $q$. Since $\chi$ is completely multiplicative, we have $\chi(1) = 1$. With Euler's theorem, Dirichlet character can only take roots of unity or zero as their values since for $(a, q) = 1$, we have $\chi(a)^{\phi(q)} = \chi(a^{\phi(q)}) = 1$.

Note that for $q = 1$, the only character is the principal character, which always returns 1. The principal character $\chi_0$ of a modulus $q \geqslant 2$ is the constant function 1 everywhere on the principal character's support. In general, when a character $\xi$ of modulus $m$ can be decomposed as

$$\xi(n) = \mathbf{1}_{(n,m)=1} \cdot \chi(n)$$

where $\chi$ is of modulus $q$ which is a factor of $m$, then we say that $\chi$ *induces* $\xi$. The smallest modulus of the character inducing $\xi$ is called the *conductor*. In the case where a character of smaller modulus cannot induce $\xi$, then we say that $\xi$ is a *primitive character*. This means that no principal characters of any modulus $q \geqslant 2$ can be primitive since it is induced by the constant function returning 1.

In general, we can construct the Dirichlet characters mod $q$ by decomposing $(\mathbb{Z}/q\mathbb{Z})^*$ into a direct product of cyclic groups. This is true, of course, because it is a finite abelian group. However, we can find an explicit decomposition using the Chinese Remainder Theorem:

$$(\mathbb{Z}/q\mathbb{Z})^* \cong (\mathbb{Z}/p_1^{\nu_1}\mathbb{Z})^* \times \ldots \times (\mathbb{Z}/p_k^{\nu_k}\mathbb{Z})^*$$

where $q = p_1^{\nu_1} \ldots p_k^{\nu_k}$ is the prime factorization of $q$. From elementary number theory, every multiplicative group modulo a prime power is cyclic except for powers of 2 greater than 8 where

$$(\mathbb{Z}/2^\nu\mathbb{Z})^* \cong (\mathbb{Z}/2\mathbb{Z}) \times (\mathbb{Z}/2^{\nu-2}\mathbb{Z})$$

if $\nu \geqslant 3$.

The generators of $(\mathbb{Z}/2^\nu\mathbb{Z})^*$ are $-1$ and 5 if $\nu \geqslant 3$ and is only $-1$ when $\nu = 2$ (for the case $\nu = 1$, the group is trivial). The generators of a multiplicative group of integers modulo $n$ are called *primitive roots modulo n*. If we know that $g$ is a primitive root modulo $p$, then either $g$ or $g + p$ is a primitive root modulo $p^\nu$ for $\nu \geqslant 2$. A proof for all these statements can be found in most elementary number theory books (see Theorems 2.39, 2.40 and 2.43 in [22] pp. 102-105). The difficulty when constructing Dirichlet characters lies in finding a primitive root for the odd prime moduli since there is no known fast algorithm to find them.

The proof of the following proposition will show important properties of Dirichlet characters while giving a way to construct them if we know the primitive roots of every odd prime power dividing exactly $q$.

**Proposition 2.11.** *The Dirichlet characters mod q form an orthonormal basis of the space $L^2((\mathbb{Z}/q\mathbb{Z})^*)$.*

***Proof.*** Since $\dim(L^2((\mathbb{Z}/q\mathbb{Z})^*)) = \phi(q)$, then a set of $\phi(q)$ orthonormal functions is an orthonormal basis. Let $\chi$ be a Dirichlet character mod $q$. Since every group $(\mathbb{Z}/q\mathbb{Z})^*$ can be written as a finite direct product of cyclic group, there exists a finite number of generators $g_1, \ldots, g_k$ with respective orders $m_1, \ldots, m_k$ such that every element has a unique representation as $g_1^{a_1} \ldots g_k^{a_k}$ with $0 \leqslant a_j < m_j$. Thus a character is entirely defined by setting the values of $\chi(g_j)$ for every $j$. Furthermore, the product of the orders of the generators needs to match the size of $(\mathbb{Z}/q\mathbb{Z})^*$ so

$$m_1 \ldots m_k = \phi(q).$$

Since $\chi(g_j)^{m_j} = 1$, $\chi(g_j)$ needs to be a ${m_j}^{\text{th}}$ root of unity. Conversely, choosing any of the ${m_j}^{\text{th}}$ root of unity for our value of $\chi(g_j)$ will satisfy every properties that a Dirichlet character needs to have. Therefore, each $\chi(g_j)$ can take $m_j$ different values possible so there are $m_1 \dots m_k = \phi(q)$ different ways to define a Dirichlet character mod $q$.

Every character has norm 1 because a root of unity always has 1 as its absolute value:

$$\|\chi\|^2 = \frac{1}{\phi(q)} \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} |\chi(a)|^2 = 1.$$

Finally, to show the orthogonality between characters, we first notice that taking the complex conjugate of one character or taking the product of two characters mod $q$ also gives a character as a result since it satisfies all the properties defining Dirichlet characters. We then observe that for any group $G$ and a fixed element $b \in G$, the function $f \colon G \to G$ defined by $f(x) = bx$ is a permutation of the elements of $G$. This is true in particular for the group $(\mathbb{Z}/q\mathbb{Z})^*$, so if $\chi$ is a nonprincipal character mod $q$, then there exists an integer $b$ coprime to $q$ such that $\chi(b) \neq 1$ and

$$\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \chi(a) = \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \chi(ba) = \chi(b) \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \chi(a) \implies \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \chi(a) = 0. \tag{2.13}$$

We therefore conclude that if $\chi_1, \chi_2$ are two distinct Dirichlet characters mod $q$, then the product $\chi_1 \cdot \overline{\chi_2}$ is a nonprincipal character and

$$\langle \chi_1, \chi_2 \rangle = \frac{1}{\phi(q)} \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \chi_1(a)\overline{\chi_2(a)} = 0.$$

This means that the Dirichlet characters mod $q$ are an orthonormal set of function of $L^2((\mathbb{Z}/q\mathbb{Z})^*)$, so they are linearly independent. Since we have $\phi(q)$ of them (which matches the function space's dimension), we get an orthonormal basis. □

Any function in $L^2((\mathbb{Z}/q\mathbb{Z})^*)$ can be written as a linear combination of Dirichlet characters using the *multiplicative Fourier transform*.

***Example* 2.12.** The indicator function can be decomposed as

$$\mathbf{1}_{n \equiv a \pmod q} = \sum_{\chi \pmod q} c_\chi \chi(n)$$

where the sum is over all the Dirichlet characters mod $q$. The $c_\chi$ are the Fourier coefficients, and they can be deduced by noticing that for a particular character $\chi_1$, we have

$$c_{\chi_1} = \sum_{\chi \pmod q} c_\chi \langle \chi, \chi_1 \rangle = \langle \mathbf{1}_{n \equiv a \pmod q}, \chi_1 \rangle = \frac{\overline{\chi_1(a)}}{\phi(q)},$$

giving us the identity

$$\mathbf{1}_{n \equiv a \pmod q} = \frac{1}{\phi(q)} \sum_{\chi \pmod q} \chi(n)\overline{\chi(a)}. \tag{2.14}$$

***Example* 2.13.** Another example, which we will use in Section 6.1, is the function $n \mapsto e(n/q)\mathbf{1}_{(n,q)=1}$. This function is in $L^2((\mathbb{Z}/q\mathbb{Z})^*)$ which means that we can find its multiplicative Fourier transform. Using the same method as above, let $c_\chi$ be the Fourier coefficient defined as

$$e(n/q)\mathbf{1}_{(n,q)=1} = \sum_{\chi \pmod q} c_\chi \chi(n).$$

We find that for a particular Dirichlet character $\chi_1$, we have

$$c_{\chi_1} = \langle e(\cdot/q)\mathbf{1}_{(\cdot,q)=1}, \chi_1\rangle = \frac{1}{\phi(q)} \sum_{a\in(\mathbb{Z}/q\mathbb{Z})^*} \overline{\chi_1(a)}e(a/q) = \frac{\mathcal{G}(\overline{\chi}_1)}{\phi(q)},$$

where we define

$$\mathcal{G}(\chi) := \sum_{1\leqslant a\leqslant q} \chi(a)e(a/q).$$

This object is called the *Gauss sum of* $\chi$. Thus we say that

$$e(n/q)\mathbf{1}_{(n,q)=1} = \frac{1}{\phi(q)} \sum_{\chi \pmod q} \mathcal{G}(\overline{\chi})\chi(n).$$

To find the Dirichlet series $\Lambda(n)\mathbf{1}_{n\equiv a \pmod q}$, we use (2.14) to get the equation

$$\Lambda(n)\mathbf{1}_{n\equiv a \pmod q} = \frac{1}{\phi(q)} \sum_{\chi \pmod q} \overline{\chi(a)}(\Lambda(n)\chi(n)). \tag{2.15}$$

We define the *Dirichlet L-functions* as $L(s,\chi) := \sum_{n=1}^{\infty} \chi(n)n^{-s}$. Hence the Dirichlet series of $\Lambda(n)\mathbf{1}_{n\equiv a \pmod q}$ can be written as

$$\sum_{n=1}^{\infty} \frac{\Lambda(n)\mathbf{1}_{n\equiv a \pmod q}}{n^s} = \frac{1}{\phi(q)} \sum_{\chi \pmod q} \overline{\chi(a)} \sum_{n=1}^{\infty} \frac{\Lambda(n)\chi(n)}{n^s} = -\frac{1}{\phi(q)} \sum_{\chi \pmod q} \overline{\chi(a)} \frac{L'(s,\chi)}{L(s,\chi)}. \tag{2.16}$$

In the case when $q = 1$, the only character is the constant function 1, meaning that the associated Dirichlet $L$-function is $\zeta(s)$.

The series defining the $L$-functions converges absolutely for $\sigma > 1$ since $|\chi(n)| \leqslant 1$ for all $n$. In addition, if $\chi$ is nonprincipal, then we have convergence for $\sigma > 0$ because (2.13) implies that $S_N = \sum_{n\leqslant N} \chi(n)$ is a bounded sequence and $n^{-\varepsilon}$ is a decreasing sequence tending to 0 for every $\varepsilon > 0$ so by Theorem 2.3, we have convergence of the series if $\sigma > 0$.

The important Dirichlet $L$-functions to study are the $L(s,\chi)$ where $\chi$ is primitive. This is because if $\xi \pmod m$ is induced by $\chi \pmod q$ where $q$ is the conductor and $\chi$ is primitive, we know that the Euler product of $L(s,\xi)$ can be written as

$$L(s,\xi) = \prod_p (1-\xi(p)p^{-s})^{-1} = \prod_p (1-\mathbf{1}_{(p,m)=1}\chi(p)p^{-s})^{-1} = L(s,\chi)\prod_{p|m}(1-\chi(p)p^{-s}).$$

In particular, if $\chi_0$ is the principal character mod $q$, then

$$L(s,\chi_0) = \prod_{p|q}(1-p^{-s})\zeta(s).$$

We are eventually going to need an analytic continuation of Dirichlet $L$-functions over the complex plane.

To first find an analytic continuation of $\zeta(s)$, we can use partial summation to deduce that if $\sigma > 1$, then

$$\zeta(s) = \int_1^\infty \frac{\mathrm{d}t}{t^s} - \int_{1^-}^\infty \frac{\mathrm{d}\{t\}}{t^s} = \frac{s}{s-1} - s\int_1^\infty \frac{\{t\}}{t^{s+1}}\,\mathrm{d}t. \tag{2.17}$$

Since the right-hand side is meromorphic and the integral is convergent when $\sigma > 0$, then we obtain an analytic continuation of $\zeta(s)$ when $\sigma > 0$. Furthermore, Riemann proved in his only paper about number theory (see [26] pp. 135-144) that the Riemann zeta function has an analytic continuation on the whole complex plane with only one simple pole at 1. He gave a functional equation which shows the symmetry of $\zeta(s)$ with respect to the critical line. A convenient way to right this functional equation is

$$\zeta(s) = 2^s\pi^{s-1}\sin(\pi s/2)\Gamma(1-s)\zeta(1-s). \tag{2.18}$$

where $\Gamma$ is the *gamma function*[6]. With (2.17), we were able to extend the definition of $\zeta(s)$ with $\sigma > 0$, and the functional equation (2.18) gives values for $\zeta(s)$ on the whole complex plane (except for the simple pole of residue 1 at $s = 1$).

The series defining $L(s,\chi)$ is convergent for $\sigma > 0$. In 1882, Hurwitz gave a generalization of the functional equation for $L$-functions of primitive characters $\chi$ with conductor $q \geqslant 2$. He showed that $L(s,\chi)$ also has symmetry with respect to the critical line, thus giving values for $L(s,\chi)$ on the whole complex plane.

The only values that $\chi(-1)$ can take are $\pm 1$ since $\chi(-1)^2 = 1$. Let $\mathfrak{a} = 0$ if $\chi(-1) = 1$ (these characters are called *even characters*) and, in the other case, let $\mathfrak{a} = 1$ if $\chi(-1) = -1$ (which are called *odd characters*). Let $\mathcal{G}(\chi)$ be the Gauss sum, which was defined in Example 2.13. Then we can define the function

$$\xi(s,\chi) := (\pi/q)^{-\frac{s+\mathfrak{a}}{2}}\Gamma\left(\frac{s+\mathfrak{a}}{2}\right)L(s,\chi). \tag{2.19}$$

The function $\xi$ can be proven to be entire. We have excluded the case where $q = 1$ because we usually multiply by $\frac{s(s-1)}{2}$ at the end of the definition to cancel the poles of $\Gamma$ and $\zeta$ at

[6] The gamma function is a nice way to analytically generalize the factorial. It is defined by $\Gamma(s) = \int_0^\infty x^{s-1}e^{-x}\,\mathrm{d}x$ for $\sigma > 0$. By integrating by parts, we can see that it respects the functional equation $\Gamma(s+1) = s\Gamma(s)$ and with this we can extend the $\Gamma$ to be a meromorphic function with simple poles at every nonpositive integers.

0 and 1 respectively[7]. The *functional equation* for the $L$-function of a primitive character $\chi$ can be written as

$$\xi(1-s,\overline{\chi}) = \frac{i^{\mathfrak{a}}\sqrt{q}}{\mathcal{G}(\chi)}\xi(s,\chi). \tag{2.20}$$

The proof of (2.20) is given in §12 of [5]. Now that we have defined Dirichlet $L$-functions and that we know that they are all meromorphic over the complex plane, we can state the Riemann Hypothesis and its generalization, which both remain unproven:

- The *Riemann Hypothesis* (RH) asserts that the real part of every nontrivial zero of $\zeta(s)$ is $\frac{1}{2}$.
- The *Generalized Riemann Hypothesis* (GRH) asserts that, for every primitive character $\chi$, the real part of every nontrivial zero of $L(s,\chi)$ is $\frac{1}{2}$.

All that remains is taking the Dirichlet series (2.16) and using Perron's formula to have an explicit formula for $\psi(x;q,a)$.

## 2.6. The explicit formula

First, using (2.15), we get

$$\psi(x;q,a) = \sum_{n\leqslant x}\Lambda(n)\mathbf{1}_{n\equiv a \ (\mathrm{mod}\ q)} = \frac{1}{\phi(q)}\sum_{\chi \ (\mathrm{mod}\ q)}\overline{\chi(a)}\psi(x,\chi) \tag{2.21}$$

where

$$\psi(x,\chi) = \sum_{n\leqslant x}\Lambda(n)\chi(n).$$

By using Perron's formula, we have

$$\psi^*(x,\chi) = \frac{1}{2\pi i}\int_{(c)} -\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}\,\mathrm{d}s$$

for $c>1$. The explicit formula comes from evaluating this integral. We will use the same idea as in the proof of Proposition 2.10 by truncating the tails of the integral and pushing the integral to the far left of the complex plane while picking up residues along the way. However, to use the residue theorem, one detail we have omitted in this process is that one needs to have an integrand that is holomorphic inside the region enclosed by the closed curve we are integrating along, except maybe a finite number of isolated singularities.

To get the explicit formula, we want to get an estimate of the form

$$\psi(x,\chi) \approx \frac{1}{2\pi i}\int_{c-iT}^{c+iT} -\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}\,\mathrm{d}s \approx \sum \mathrm{res}, \tag{2.22}$$

[7] Note that $\Gamma$ also has simple poles at every negative integers but they are canceled out since $\zeta(-2n)=0$ for $n\geqslant 1$. To prove this, we can simply take the functional equation (2.20) since it does not matter if $\xi$ is entire or not to use the equation.

where the sum over the residues of the singularities $z$ with $\mathrm{Re}(z) < c$ and $|\mathrm{Im}(z)| < T$. The following lemma gives us an error term of the truncated version of Perron's formula to apply it in practice:

**Lemma 2.14.** *If $A(s) = \sum_{n=1}^{\infty} a_n n^{-s}$ is a Dirichlet series with $|a_n| \ll \log n$ for $n \geqslant 2$, then for $x, T \geqslant 2$ and $c = 1 + \frac{1}{\log x}$ we have*

$$\sum_{n\leqslant x} a_n = \int_{c-iT}^{c+iT} A(s)\frac{x^s}{s}\,\mathrm{d}s + O\!\left(\frac{x\log^2 x}{T} + \log x\right).$$

***Proof.*** First of all, let $S(x) = \sum_{n\leqslant x} a_n$, then

$$\begin{aligned} S(x) &= S^*(x) + O(\log x) \\ &= \sum_{n=1}^{\infty} a_n \mathbf{1}^*_{(1,\infty)}\left(\tfrac{x}{n}\right) + O(\log x) \\ &= \int_{c-iT}^{c+iT} A(s)\frac{x^s}{s}\,\mathrm{d}s + \sum_{n=1}^{\infty} a_n E(\tfrac{x}{n}, T) + O(\log x)\,, \end{aligned}$$

where $E(y,T)$ is the error term in the Proposition 2.10. The Dirichlet series in the last equality is obtained by interchanging an infinite series and an integral. This can be justified because for a fixed $x \geqslant 2$, the series $\sum_{n=1}^{\infty} a_n(\frac{x}{n})^s s^{-1}$ converges uniformly for $s$ on the vertical line $\sigma = c$ by the Weierstrass M-test. All that is left to show is that $\sum_{n=1}^{\infty} a_n E(\frac{x}{n}, T) \ll \frac{x\log^2 x}{T} + \log x$.

We are going to decompose the sum $\sum_{n=1}^{\infty} a_n E(\frac{x}{n}, T)$ with

$$\sum_{n\geqslant 1} = \sum_{1<|n-x|\leqslant \frac{x}{T}} + \sum_{\max\{1,\frac{x}{T}\}<|n-x|\leqslant \frac{x}{2}} + \sum_{|n-x|\geqslant \frac{x}{2}} + O(\log x)\,,$$

where the three sums on the right side are denoted $E_1, E_2$ and $E_3$ respectively. Note that the first sum may be empty and that the error term is composed of the sum over the values of $n$ respecting $|n-x| \leqslant 1$, which contains at most 3 terms. Let us also note that $x^c = ex \asymp x$ since we are going to use it repetitively to bound every sum $E_j$.

For the interval of the first sum where $1 < |n-x| \leqslant \frac{x}{T}$, we use the bound $E(\frac{x}{n}, T) \ll \left(\frac{x}{n}\right)^c$ since $\log(x/n)$ is getting too close to 0 to have a good bound with it in the denominator. This means that by applying this bound and $a_n \ll \log n$, we get

$$E_1 \ll \sum_{|n-x|\leqslant \frac{x}{T}} \log n \left(\frac{x}{n}\right)^c \ll \left(\frac{x}{x(1-\frac{1}{T})}\right)^c \log(x(1+\tfrac{1}{T})) \sum_{|n-x|\leqslant \frac{x}{T}} 1 \ll \frac{x}{T}\log x.$$

In the second sum, using Taylor's theorem, we know that $|\log(1+y)| \asymp |y|$ in any compact subset of $(-1,1)$. Thus we can use the estimate

$$|\log(x/n)| = |\log(n/x)| = \left|\log\left(1 + \tfrac{n-x}{x}\right)\right| \asymp \frac{|n-x|}{x}$$

because $\frac{|n-x|}{x} \leqslant \frac{1}{2}$. This also implies that $n \asymp x$ in the range of $n$. We can then bound $E_2$ with

$$E_2 \ll \sum_{\max\{1,\frac{x}{T}\}<|n-x|\leqslant\frac{x}{2}} \log n \frac{x(x/n)^c}{T\,|n-x|} \asymp \frac{x\log x}{T} \sum_{\max\{1,\frac{x}{T}\}<|n-x|\leqslant\frac{x}{2}} \frac{1}{|n-x|}.$$

We can see that the last sum is close to the summation defining a harmonic number, so we should expect the sum to contribute for about $\log x$ to the bound. To prove this, we can use dyadic decomposition and show that

$$\ll \frac{x\log x}{T}\sum_{1\leqslant 2^j\leqslant\frac{x}{2}}\ \sum_{2^j<|n-x|\leqslant 2^{j+1}}\frac{1}{|n-x|} \ll \frac{x\log x}{T}\sum_{1\leqslant 2^j\leqslant \frac{x}{2}}\ \sum_{2^j<|n-x|\leqslant 2^{j+1}}\frac{1}{2^j}\ll\frac{x\log^2 x}{T}.$$

In the intervals of the third sum defined by $|n-x| \geqslant \frac{x}{2}$, we have

$$E(\tfrac{x}{n},T) \ll \frac{(x/n)^c}{1+T\,|\log(x/n)|} \ll \frac{x}{T}n^{-c}$$

because $|\log(x/n)| \gg 1$. By applying this bound and $a_n \ll \log n \ll n^{\varepsilon}/\varepsilon$ to $E_3$ for $\varepsilon = \frac{1}{2\log x}$, we get

$$E_3 \ll \frac{x}{\varepsilon T}\sum_{n=1}^{\infty}\frac{1}{n^{c-\varepsilon}}.$$

We can bound the last sum by an integral since the terms are decreasing, which would give us

$$\ll \frac{x}{\varepsilon T}\int_1^{\infty}\frac{\mathrm{d}t}{t^{c-\varepsilon}} = \frac{x}{\varepsilon T}\cdot\frac{1}{c-\varepsilon-1} \ll \frac{x\log^2 x}{T},$$

and this finally proves that $\sum_{n=1}^{\infty} a_n E(\frac{x}{n},T) \ll \frac{x\log^2 x}{T}$. This completes the proof of the lemma. □

If we apply Lemma 2.14 to $\psi(x,\chi)$, we find

$$\psi(x,\chi) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT} -\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}\,\mathrm{d}s + O\!\left(\frac{x\log^2 x}{T} + \log x\right). \tag{2.23}$$

We apply the residue theorem by shifting the contour to the far left:

$$\begin{aligned}\psi(x,\chi) = \sum \mathrm{res} + \frac{1}{2\pi i}\int_{-\infty+iT}^{c+iT} -\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}\,\mathrm{d}s \\ - \frac{1}{2\pi i}\int_{-\infty-iT}^{c-iT} -\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}\,\mathrm{d}s + O\!\left(\frac{x\log^2 x}{T} + \log x\right),\end{aligned} \tag{2.24}$$

where $\sum\mathrm{res}$ is defined as in equation (2.22). Note that $\frac{-L'(s,\chi)}{L(s,\chi)}\cdot\frac{x^s}{s} \ll \frac{T\log(qk)}{kx^{2k}}$ uniformly on the vertical line $\mathrm{Re}(s) = 2k+\mathfrak{a}+1$ for any positive integer $k$, which is why we could push the contour to the far left without any problem. A proof of this and of the other bounds on $\frac{L'(s,\chi)}{L(s,\chi)}$ are in §19 of [5].

Furthermore, since we do not want to encounter a nontrivial zero of $L(s,\chi)$ when forming our contour, we need to choose carefully $T$ such that the distance between the contour and the closest zero is controlled. By using the argument principle, we have a formula for the density of the nontrivial zeros of $L(s,\chi)$:

**Theorem 2.15.** *Let $N(T,\chi)$ be the number of zeros $\rho_\chi = \beta_\chi + i\gamma_\chi$ in the critical strip of $L(s,\chi)$ with $|\gamma_\chi| \leqslant T$. Then*

$$N(T,\chi) = \frac{T}{\pi}\log\left(\frac{qT}{2\pi e}\right) + O(\log(qT))\,.$$

The proof is in §16 of [5]. This ensures that we can choose $T$ arbitrarily large such that $|\gamma_\chi - T| \gg \frac{1}{\log(qT)}$. Finally, the contribution of the two integrals in (2.24) is $O\Big(\frac{x\log^2(qxT)}{T}\Big)$. The proof is also in §19 of [5].

The residues of $\frac{-L'(s,\chi)}{L(s,\chi)} \cdot \frac{x^s}{s}$ come from the zeros of $L(s,\chi)$ and the origin (there is an extra simple pole at $s=1$ in the case $q=1$). The pole at the origin is simple since we have assumed GRH, and its residue is $-L'(0,\chi)/L(0,\chi)$ if $\chi$ is an odd character or if $q=1$, and is of the form $\log x + b(\chi)$ if $\chi$ is even with conductor $q \geqslant 2$. Both $L'(0,\chi)/L(0,\chi)$ and $b(\chi)$ are $O(\log q)$ under GRH. The residue at the zero $\rho$ of $L(s,\chi)$ is $-mx^\rho/\rho$, where $m$ is the multiplicity of the zero. Since sum of the residues of the trivial zeros is $\ll \log x$, this leads us to the following explicit formula:

**Theorem 2.16.** *Let $q, T \geqslant 1$, $x \geqslant 2$ and $\chi$ be a character mod $q$. Then*

$$\psi(x,\chi) = x\mathbf{1}_{\chi=\chi_0} - \sum_{|\gamma_\chi|\leqslant T}\frac{x^{\rho_\chi}}{\rho_\chi} + O\bigg(\frac{x\log^2(qxT)}{T} + \log(qx)\bigg)\,,$$

*where the $\rho_\chi = \beta_\chi + i\gamma_\chi$ represents the zeros of $L(s,\chi)$ with real part between $0$ and $1$, and each zero in the sum is counted with its multiplicity.*

# 2.7. The Prime Number Theorem for arithmetic progressions

Taking together the density of zeros with the explicit formula, we can have a sense of the size of the error term in Theorem 2.1. Let

$$\beta_0 = \max_{\rho=\beta+i\gamma}\beta,$$

where the maximum is taken over every nontrivial zero of $L$-functions of Dirichlet characters mod $q$, then if $q \ll_A x^A$ and $T = x$, we have

$$\sum_{|\gamma_\chi|\leqslant T}\frac{x^\rho}{\rho} \ll x^{\beta_0}\sum_{|\gamma_\chi|\leqslant T}\frac{1}{|\rho|} \ll x^{\beta_0}\int_{1/q}^{T}\frac{\log qt}{t}\,\mathrm{d}t = x^{\beta_0}(\log qT)^2$$

$$\implies \psi(x,\chi) = x\mathbf{1}_{\chi=\chi_0} + O_A\Big(x^{\beta_0}\log^2 x\Big)$$

for a primitive character $\chi$. In the case where $\xi \pmod{m}$ is induced by $\chi \pmod{q}$ where $q$ is the conductor and $\chi$ is primitive, we obtain the same estimate as above since

$$\psi(x,\xi) = \sum_{n\leqslant x} \Lambda(n)\xi(n) = \sum_{\substack{n\leqslant x\\(n,m)=1}} \Lambda(n)\chi(n) = \psi(x,\chi) + O\left(\sum_{\substack{p^k\leqslant x\\p\mid m}} \log p\right)$$
$$= \psi(x,\chi) + O(\log m \log x) = \psi(x,\chi) + O_A\Big(\log^2 x\Big)\,.$$

Thus, with the Riemann Hypothesis for $L$-functions of Dirichlet characters mod $q$ (meaning that $\beta_0 = 1/2$) and using (2.21), we have the estimate

$$\psi(x;q,a) = \frac{1}{\phi(q)} \sum_{\chi \pmod{q}} \overline{\chi(a)}\psi(x,\chi) = \frac{x}{\phi(q)} + O_A\Big(x^{1/2}\log^2 x\Big)\,,$$

where the main term come from the estimate on the principal character. If we let $\beta_0$ be free, we see why we have (1.4).

To transition to the function $\pi(x)$, we decompose the sum forming $\psi$ into the distinct powers of primes

$$\psi(x;q,a) = \sum_{\substack{p^k\leqslant x\\p^k\equiv a \pmod{q}}} \log p = \theta(x;q,a) + \sum_{\substack{p^k\leqslant x\\p^k\equiv a \pmod{q}\\k\geqslant 2}} \log p,$$

where

$$\theta(x;q,a) = \sum_{\substack{p\leqslant x\\p\equiv a \pmod{q}}} \log p$$

is *Chebyshev's theta function.* Since

$$\sum_{\substack{p^k\leqslant x\\p^2\equiv a \pmod{q}\\k\geqslant 2}} \log p \ll \sum_{p\leqslant x^{1/2}} \log p \sum_{2\leqslant k\leqslant \log x/\log p} 1 \ll x^{1/2}\log x, \tag{2.25}$$

we get $|\psi(x;q,a) - \theta(x;q,a)| \ll \sqrt{x}\log x$. Finally, we can get

$$\theta(x;q,a) = \frac{x}{\phi(q)} + O_A\Big(x^{1/2}\log^2 x\Big) \implies \pi(x;q,a) = \frac{\mathrm{Li}(x)}{\phi(q)} + O_A\Big(x^{1/2}\log x\Big)$$

with partial summation which proves Theorem 2.1. Let's note that if we take $q = 1$, this also proves the Prime Number Theorem under RH:

$$\pi(x) = \mathrm{Li}(x) + O\Big(x^{1/2}\log x\Big)\,.$$

# Chapter 3

# Chebyshev's bias

In this chapter, we will assume that GRH is true. The previous chapter established that the prime numbers are approximately uniformly distributed in the reduced arithmetic progressions mod $q$. The words "approximately uniformly distributed" mean in this context that $\pi(x;q,a)$ can be approximated by a function which is independent of the value $a$, as long as we choose $a$ to be relatively prime to $q$. However, Chebyshev noticed that there was a bias in the race between the prime numbers of the form $4n+1$ and the primes of the form $4n+3$. The difference $\pi(x;4,3)-\pi(x;4,1)$ seemed to take positive values for a large proportion of values of $x$. Let's recall the function $E(x;q,a,b)=\frac{\log x}{\sqrt{x}}(\pi(x;q,a)-\pi(x;q,b))$. Figure 3.1 illustrates the behavior of $E(x;4,3,1)$ in the interval $[10^4,10^8]$.

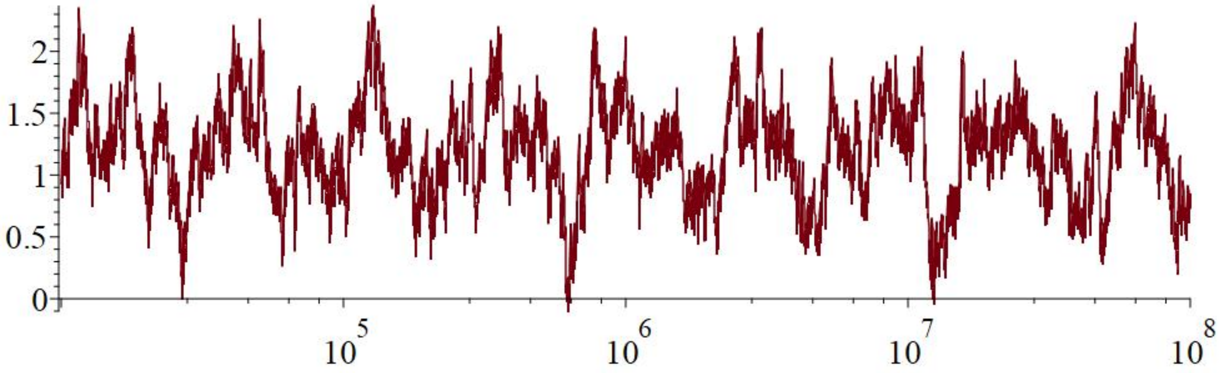

**Fig. 3.1.** Graph of the function $E(x;4,3,1)=\frac{\log x}{\sqrt{x}}(\pi(x;4,3)-\pi(x;4,1))$.

This graph shows three spikes falling below the $x$-axis. They represent the times where the primes 1 (mod 4) take the lead in the race against the primes 3 (mod 4). The function $E(x;4,3,1)$ is always positive for every $x\leqslant 10^8$ except on a subset of the union of intervals $[26861,26863)\cup[616841,633798)\cup[12306137,12382326)$.

At first glance, the graph seems erratic. However, the map $u\mapsto E(e^u;4,3,1)$ follows a quasi-periodic structure since, as we will show, this function can be written as a superposition of waves[1]. The frequencies and amplitudes of these waves can be determined by the zeros of $L(s,\chi_1)$ where $\chi_1$ is the only nonprincipal character mod 4. But with Theorem 2.1 we do not have enough information to study the bias and all we can say is that $E(x;4,3,1) \ll (\log x)^2$. Revisiting the steps of the proof where we are bounding functions which contribute to the error term of the theorem, we will give more details to (2.25) for the two reduced congruence classes of the modulus $q=4$. This is the starting point of the study of Chebyshev's bias.

## 3.1. Being a quadratic residue makes a difference

The sum defining Chebyshev's $\psi$ function can be decomposed into the different powers of prime numbers:

$$\psi(x;4,a) = \sum_{k\leqslant \frac{\log x}{\log 2}} \sum_{\substack{p\leqslant x^{1/k} \\ p^k\equiv a \pmod 4}} \log p. \tag{3.1}$$

Chebyshev wrote in his 1853 letter to Fuss (in French):

> *"En cherchant l'expression limitative des fonctions qui déterminent la totalité des nombres premiers de la forme* $4n+1$ *et ceux de la forme* $4n+3$*, pris au dessous d'une limite très grande, je suis parvenu à reconnaître que ces deux fonctions diffèrent notablement entre elles par leurs seconds termes, dont la valeur, pour les nombres* $4n+1$*, est plus grande que celle pour les nombres* $4n+1$*; ainsi, si de la totalité des nombres premiers de la forme* $4n+3$*, inférieurs à une limite quelconque* $x$*, on retranche celle des nombres premiers de la forme* $4n+1$*, et que l'on divise ensuite cette différence par la quantité* $\frac{\sqrt{x}}{\log x}$*, on trouvera plusieurs valeurs de* $x$ *telles, que ce quotient s'approchera de l'unité aussi près qu'on le voudra."*

In the second part of this quote, Chebyshev essentially writes that $E(x;4,3,1)$ has multiple values around 1, which falls in line with what we can see from Figure 3.1. In the first part of the quote however, he says that he recognized that the bias comes from the fact that the term $k=2$ in the equation (3.1) is completely different if we compare the two cases with $a$ being 1 or 3. This comes from the fact that every square of an odd prime is 1 (mod 4).

With the explicit formula for $\psi$, we will present how Rubinstein and Sarnak proved in [27] that $E(e^u;4,3,1)$ is close to a sum of trigonometric functions oscillating around 1. For our further calculations in this chapter concerning the limiting distribution and in Chapter 7 when studying the periodicity of the function, we only need a bound on the norm of the error term in $L^2$ which is easily computable since $L^2$ is a Hilbert space:

[1] This is why the graph of $E$ in Figure 3.1 is presented with the $x$-axis on a logarithmic scale.

**Theorem 3.1.** *For $u, T$ and $U$ respecting $u, \log T \in [0, U]$, then assuming GRH we can write*

$$E(e^u; 4, 3, 1) = 1 + \sum_{0<\gamma\leqslant T} \frac{\cos(\gamma u) + 2\gamma \sin(\gamma u)}{\frac{1}{4} + \gamma^2} + \varepsilon(u; T)$$

*where $\gamma$ are the imaginary parts of the zeros of $L(s, \chi_1)$ with $\chi_1$ being the nonprincipal character mod 4 and*

$$\int_0^U \varepsilon(u; T)^2 \,\mathrm{d}u \ll \frac{U \log^2 T}{T} + 1.$$

***Proof.*** Let $\chi_1$ be the nonprincipal Dirichlet character mod 4. The difference $\pi(x; 4, 3) - \pi(x; 4, 1)$ is essentially a sum of values of $\chi_1$ over primes since we can use Proposition 2.11 to find that

$$\mathbf{1}_{p\equiv 3 \pmod 4} - \mathbf{1}_{p\equiv 1 \pmod 4} = -\chi_1(p) \tag{3.2}$$

and ultimately rewrite $E(e^u; 4, 3, 1)$ as

$$E(e^u; 4, 3, 1) = -ue^{-u/2} \sum_{p\leqslant e^u} \chi_1(p).$$

From the discussion at the end of Section 2.2, studying a sum of values of a completely multiplicative function over primes is easier if we transition to the von Mangoldt function $\Lambda$ instead and use the theory of $L$-functions from the previous chapter. To do this, we can decompose the sum into three parts and notice that $\chi_1(p)^2 = 1$ for every odd prime to find the equation

$$\sum_{p\leqslant x} \chi_1(p) = \sum_{n\leqslant x} \frac{\Lambda(n)\chi_1(n)}{\log n} - \frac{1}{2} \sum_{3\leqslant p\leqslant \sqrt{x}} 1 - \sum_{3\leqslant k\leqslant \frac{\log x}{\log 2}} \frac{1}{k} \sum_{p\leqslant x^{1/k}} \chi_1(p)^k. \tag{3.3}$$

The oscillations in the theorem come from the first sum on the right-hand side of (3.3), the second sum will make the bias and the third double sum will contribute to the error term.

For the first sum on the right-hand side of (3.3), we can use partial summation twice. For any Dirichlet character $\chi$, let $G(x, \chi) = \int_2^x \psi(t, \chi) \,\mathrm{d}t$. Then we have

$$\begin{aligned}\sum_{n\leqslant x} \frac{\Lambda(n)\chi_1(n)}{\log n} &= \int_{2^-}^x \frac{\mathrm{d}\psi(t, \chi_1)}{\log t} = \frac{\psi(x, \chi_1)}{\log x} + \int_2^x \frac{\psi(t, \chi_1)}{t \log^2 t} \,\mathrm{d}t \\ &= \frac{\psi(x, \chi_1)}{\log x} + \frac{G(x, \chi_1)}{x \log^2 x} + \int_2^x \frac{G(t, \chi_1)(\log t + 2)}{t^2 \log^3 t} \,\mathrm{d}t. \end{aligned} \tag{3.4}$$

We can give a bound on $G(x, \chi_1)$ by integrating the associated explicit formula for $\psi(x, \chi_1)$ with $T = x$ and noting that the density of the zeros of $L(s, \chi_1)$ from Theorem 2.15 implies that $\sum \frac{1}{\gamma^2}$ is a convergent series:

$$G(x, \chi_1) = \int_2^x \sum_{|\gamma|\leqslant x} \frac{t^{1/2+i\gamma}}{\frac{1}{2} + i\gamma} + O\left(\frac{t \log^2(tx)}{x} + \log t\right) \mathrm{d}t \ll x^{3/2} \sum_{|\gamma|\leqslant x} \frac{1}{\gamma^2} \ll x^{3/2}.$$

Applying the bound to (3.4), we obtain the estimate

$$\sum_{n\leqslant x}\frac{\Lambda(n)\chi_1(n)}{\log n}=\frac{\psi(x,\chi_1)}{\log x}+O\bigg(\frac{\sqrt{x}}{\log^2 x}\bigg).$$

The second sum in (3.3) is simply evaluated with the Prime Number Theorem since $\frac{1}{2}(\pi(\sqrt{x})-1)=\frac{\sqrt{x}}{\log x}\left(1+O\left(\frac{1}{\log x}\right)\right)$. Finally, the final sum in (3.3) can be bounded by

$$\sum_{3\leqslant k\leqslant\frac{\log x}{\log 2}}\frac{1}{k}\sum_{p\leqslant x^{1/k}}\chi_1(p)^k\ll\sum_{3\leqslant k\leqslant\frac{\log x}{\log 2}}\frac{\pi(x^{1/k})}{k}\ll\sum_{3\leqslant k\leqslant\frac{\log x}{\log 2}}\frac{x^{1/k}}{\log x}\ll x^{1/3}.$$

We can combine these estimates in (3.3) to get

$$E(x;4,3,1)=-\frac{\log x}{\sqrt{x}}\sum_{p\leqslant x}\chi_1(p)=1-\frac{\psi(x,\chi_1)}{\sqrt{x}}+O\bigg(\frac{1}{\log x}\bigg).$$

With the explicit formula from Theorem 2.16, we can define the function

$$E(e^u;4,3,1)=1+\sum_{|\gamma|\leqslant T}\frac{e^{i\gamma u}}{\frac{1}{2}+i\gamma}+\sum_{T<|\gamma|\leqslant e^U}\frac{e^{i\gamma u}}{\frac{1}{2}+i\gamma}+O\Big(e^{u/2-U}(U^2+u^2)+\frac{1}{u}\Big)\tag{3.5}$$

We have separated the sum over the zeros of $L(s,\chi_1)$ into two parts since we only need a bound on the standard deviation if we replaced $u$ by a uniform random variable on the interval $[\log 2,U]$. Since $\chi_1$ is a real character, then the position of the zeros of $L(s,\chi_1)$ are symmetric with respect to reflection along the real axis. Thus, we can group the zeros in pairs in the following way:

$$\frac{e^{i\gamma u}}{\frac{1}{2}+i\gamma}+\frac{e^{-i\gamma u}}{\frac{1}{2}-i\gamma}=\frac{\cos(\gamma u)+2\gamma\sin(\gamma u)}{\frac{1}{4}+\gamma^2}.$$

All that is left to prove is to bound the mean square of the error term

$$\varepsilon(u;T)=\sum_{T<|\gamma|\leqslant e^U}\frac{e^{i\gamma u}}{\frac{1}{2}+i\gamma}+O\Big(e^{u/2-U}(U^2+u^2)+\frac{1}{u}\Big).$$

By taking the square and using the Cauchy-Schwarz inequality, we get

$$\varepsilon(u;T)^2\leqslant\sum_{T<|\gamma_1|,|\gamma_2|\leqslant e^U}\frac{2e^{i(\gamma_1-\gamma_2)u}}{(\frac{1}{2}+i\gamma_1)(\frac{1}{2}-i\gamma_2)}+O\Big(e^{u-2U}(U^4+u^4)+\frac{1}{u^2}\Big).\tag{3.6}$$

By integrating over the interval $[0,U]$ and by using the fact that $\int_0^U e^{i(\gamma_1-\gamma_2)u}\,\mathrm{d}u\ll\min\{U,1/|\gamma_1-\gamma_2|\}$ for arbitrary nontrivial zeros of $L(s,\chi_1)$, we can get the bound

$$\int_0^U\varepsilon(u;T)^2\,\mathrm{d}u\ll\left(\sum_{T<|\gamma_1|,|\gamma_2|\leqslant e^U}\frac{\min\{U,1/|\gamma_1-\gamma_2|\}}{|\gamma_1\gamma_2|}\right)+1.$$

Using the symmetry in the terms, it suffices to get an upper bound for the sum over the pairs of zeros such that $T<\gamma_1\leqslant\gamma_2\leqslant e^U$ and for the sum over the ones such that

$\gamma_1, (-\gamma_2) \in (T, e^U]$, and since we have

$$\sum_{\substack{T<\gamma_1\leqslant e^U\\ -e^U\leqslant\gamma_2<-T}} \frac{1}{|\gamma_1\gamma_2(\gamma_1-\gamma_2)|} = \sum_{T<\gamma_1,\gamma_2\leqslant e^U} \frac{1}{\gamma_1\gamma_2(\gamma_1+\gamma_2)} \ll \left(\sum_{T<\gamma_1\leqslant e^U} \frac{1}{\gamma_1^{3/2}}\right)^2 \ll 1,$$

then we only need to bound

$$\begin{aligned}\int_0^U \varepsilon(u;T)^2\,\mathrm{d}u &\ll \left(\sum_{T<\gamma_1\leqslant\gamma_2\leqslant e^U} \frac{\min\{U, 1/(\gamma_2-\gamma_1)\}}{\gamma_1\gamma_2}\right) + 1\\ &= \left[\sum_{\gamma_1>T}\frac{1}{\gamma_1}\left(\sum_{\gamma_2\in[\gamma_1,\gamma_1+\frac{1}{U}]}\frac{U}{\gamma_2} + \sum_{\gamma_2\in[\gamma_1+1/U, e^U]}\frac{1}{\gamma_2(\gamma_2-\gamma_1)}\right)\right] + 1.\end{aligned}$$

With partial summation and Theorem 2.15, we arrive to the desired upper bound on $\int_0^U \varepsilon(u;T)^2\,\mathrm{d}u$.

□

## 3.2. The race $\pi(x)$ vs. $\mathrm{Li}(x)$

For any $q$ which is an odd prime, the Legendre symbol $n \mapsto \left(\frac{n}{q}\right)$ is a primitive character mod $q$. Hence, we can use the same techniques as the proof of Theorem 3.1 to pit in a prime number race the primes that are quadratic residues mod $q$ against the ones that are not and obtain a very similar theorem. One other prime number race, which does not really respect the definition of the prime number race that we gave in Section 1.5, is letting the primes collectively race against the smooth function $\mathrm{Li}(x)$. The Figure 3.2 illustrates how $\mathrm{Li}(x)$ seem to be always bigger than $\pi(x)$. To this day, there is no known number $x$ such that $\mathrm{Li}(x) - \pi(x)$ is negative. But Littlewood showed that if we let $x \to \infty$, we will find an infinite number of sign changes, we just haven't determined any yet. Just as Chebyshev's

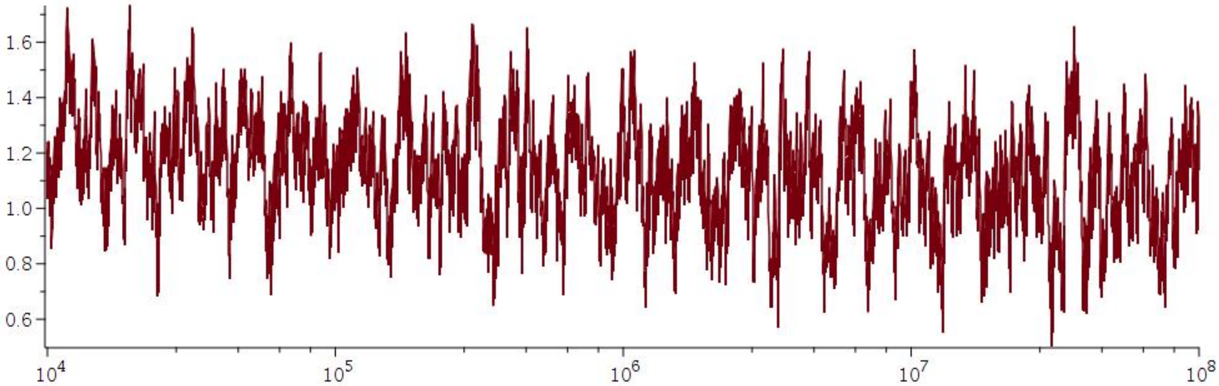

**Fig. 3.2.** Graph of $\frac{2}{\mathrm{Li}(\sqrt{x})}(\mathrm{Li}(x) - \pi(x))$.

race, $|\psi(x)-\theta(x)| \ll \sqrt{x}\log x$ is the crude bound blinding us from the reason of the bias in the race between $\pi$ and Li.

Following the proof of Theorem 3.1, we will show this using the explicit formula for $\psi(x)$. We will study the function

$$\Pi(x) = \sum_{n\leqslant x}\frac{\Lambda(n)}{\log n} = \sum_{k\leqslant \frac{\log x}{\log 2}}\frac{1}{k}\pi(x^{1/k}). \tag{3.7}$$

Using Möbius inversion, we can get

$$\pi(x) = \sum_{k\leqslant \frac{\log x}{\log 2}}\frac{1}{k}\pi(x^{1/k})\sum_{dr=k}\mu(d) = \sum_{d\leqslant \frac{\log x}{\log 2}}\frac{\mu(d)}{d}\sum_{r\leqslant \frac{\log x}{d\log 2}}\frac{1}{r}\pi((x^{1/d})^{1/r}) = \sum_{d\leqslant \frac{\log x}{\log 2}}\frac{\mu(d)}{d}\Pi(x^{1/d}).$$

If there is no bias in the race $\psi(x)$ vs. $x$, then we can assume that $\Pi(x)$ vs. $\mathrm{Li}(x)$ is also unbiased by partial summation, which means that a better way to approximate $\pi(x)$ would be

$$\pi(x) \approx \sum_{d\leqslant \frac{\log x}{\log 2}}\frac{\mu(d)}{d}\mathrm{Li}(x^{1/d}). \tag{3.8}$$

In fact, the function $\mathrm{Li}(x)-\frac{1}{2}\mathrm{Li}(\sqrt{x})-\pi(x)$ seems to already have much more sign changes than $\mathrm{Li}(x)-\pi(x)$ as Figure 3.2 illustrates with the multiple times the graph crosses the line $y=1$.

We can write $\Pi(x)$ in different ways. First, from (3.7) and the PNT with the Riemann Hypothesis, we can write

$$\Pi(x) = \pi(x) + \frac{1}{2}\pi(\sqrt{x}) + \sum_{3\leqslant k\leqslant \frac{\log x}{\log 2}}\frac{1}{k}\pi(x^{1/k}) = \pi(x) + \frac{1}{2}\mathrm{Li}(\sqrt{x}) + O\Big(x^{1/3}\Big).$$

In a different manner, we can write $\Pi(x)-\mathrm{Li}(x)$ as a Riemann-Stieltjes integral in the following way:

$$\Pi(x)-\mathrm{Li}(x) = \int_{2^-}^{x}\frac{\mathrm{d}(\psi(x)-x)}{\log x} \implies \mathrm{Li}(x)-\pi(x) = \frac{1}{2}\mathrm{Li}(\sqrt{x}) - \int_{2^-}^{x}\frac{\mathrm{d}(\psi(x)-x)}{\log x} + O\Big(x^{1/3}\Big).$$

Since the only difference between the explicit formula of $\psi(x)-x$ and the one of $\psi(x,\chi_1)$ is the location of the zeros of their associated $L$-functions, we can also arrive to a theorem of the same form and with an identical proof as that of Theorem 3.1:

**Theorem 3.2.** *For $u, T$ and $U$ respecting $u, \log T \in [0, U]$, then assuming the Riemann Hypothesis we can write*

$$\frac{2}{\mathrm{Li}(e^{u/2})}(\mathrm{Li}(e^u)-\pi(e^u)) = 1 + \sum_{0<\gamma\leqslant T}\frac{\cos(\gamma u)+2\gamma\sin(\gamma u)}{\frac{1}{4}+\gamma^2} + \varepsilon(u;T)$$

*where $\gamma$ are the imaginary parts of the zeros of $\zeta$ and*

$$\int_0^U \varepsilon(u;T)^2\,\mathrm{d}u \ll \frac{U\log^2 T}{T}+1.$$

**Remark 3.3.** The approximation (3.8) is the reason why the values of the function $\alpha(x)$ defined in Figure 1.2 seem to oscillate around the smooth function

$$\frac{\log(\frac{1}{2}\,\mathrm{Li}(x^{1/2})+\frac{1}{3}\,\mathrm{Li}(x^{1/3}))}{\log x}. \tag{3.9}$$

This function is asymptotic to $\frac{1}{2}-\frac{\log\log x}{\log x}$ which does not contradict the Riemann Hypothesis. We only considered three terms in the sum in (3.8) since it seemed sufficient to model the curve around which we have our oscillations.

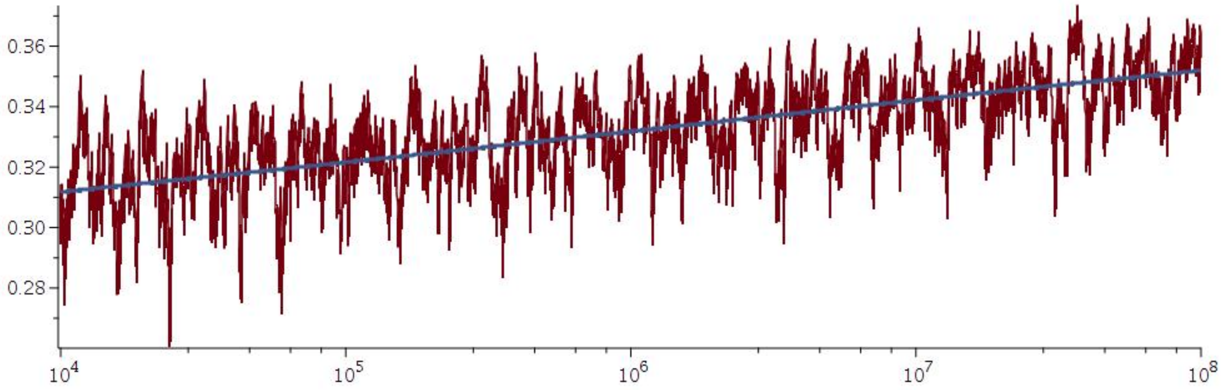

**Fig. 3.3.** Graph of the function $\alpha$ defined by $\pi(x)=\mathrm{Li}(x)-x^{\alpha(x)}$ (in red) and the smooth approximation (3.9) around which $\alpha(x)$ oscillates (in blue).

## 3.3. Why use the logarithmic density?

For a random positive real number $x$, could we predict the probability of having the primes 1 (mod 4) in the lead over the primes 3 (mod 4) for this particular value of $x$? In other words, what is the probability that $E(x;4,3,1)<0$? The words "a random positive real number $x$" are ambiguous since we did not specify the probability measure. There is a very natural way to understand which probability measure is appropriate for Chebyshev's race, the $\pi(x)$ vs. $\mathrm{Li}(x)$ race and any race where we end up with similar theorems as Theorem 3.1 and Theorem 3.2.

If we were to ask instead for a random number in the interval $[0,1)$, then most people would naturally assume that we are talking about the uniform distribution defined by the Lebesgue measure. The reason for this is that the uniform distribution treats every points in the interval with perfect equality, with no room for any bias. Every point in the interval is a mode of the uniform distribution. However, there is obviously no probability distribution where the support is $[0,\infty)$ and where every point is a mode.

Instead of trying to find a perfect distribution supported on $[0,\infty)$, we can interpret a random positive number as a uniformly distributed random variable in the interval $[0,X]$ for a very large $X$. If the limit

$$d(A) := \frac{1}{X}\int_0^X \mathbf{1}_A(y)\,\mathrm{d}y$$

exists, then we say that the *natural density* of $A$ is $d(A)$, and it acts as a probability measure. In 1962, Stanisław Knapowski and Pál Turán conjectured in [15] that $d(\{x>0 : E(x;4,3{,}1)<0\}) = 0$. This turned out to be false, and Jerzy Kaczorowski [13] showed that the natural density does not exist. Perhaps the natural density is not the right way to measure $\{x>0 : E(x;4,3{,}1)<0\}$.

The Lebesgue measure $\lambda$ is what we call an *invariant measure under translations* in the group $\mathbb{R}/\mathbb{Z}$ because for any $E\subset\mathbb{R}/\mathbb{Z}$, we have $\lambda(E)=\lambda(E+a)$ for all $a\in\mathbb{R}/\mathbb{Z}$ where the set $E+a=\{x+a : x\in E\}$. Any periodic function $f\colon\mathbb{R}\to\mathbb{R}$ with period $T$ can be entirely defined by its values over $[0,T)$. Thus if we would like to know the distribution of the values of $f(x)$ for a real number $x$ chosen uniformly at random, we can restrict our function on the period and use the uniform distribution on the interval $[0,T)$.

Let $g\colon\mathbb{R}\to\mathbb{R}$, if $g\circ f$ is absolutely integrable over the period, then

$$\frac{1}{T}\int_0^T (g\circ f)(x)\mathrm{d}x = \frac{1}{nT}\int_0^{nT}(g\circ f)(x)\mathrm{d}x$$

for every positive integer $n$. Thus for $X\geqslant T$, we have

$$\frac{1}{X}\int_0^X (g\circ f)(x)\,\mathrm{d}x = \frac{1+O(T/X)}{T}\int_0^T (g\circ f)(x)\,\mathrm{d}x.$$

So for a $T$-periodic function $f$, we can study its values with the probability distribution $\mu$ such that

$$\lim_{X\to\infty}\frac{1}{X}\int_0^X (g\circ f)(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} g(y)\,\mathrm{d}\mu(y). \tag{3.10}$$

For $g(y)=\mathbf{1}_{y>0}$, the integral above would be the natural density of the set $\{x>0 : f(x)>0\}$. The *logarithmic density* of $A$ is defined by

$$\delta(A) := \lim_{X\to\infty}\frac{1}{\log X}\int_{[1,X]\cap A}\frac{\mathrm{d}t}{t}.$$

With the change of variables $t=e^x$ in the integral of the left-hand side of (3.10), we get that the logarithmic density of the set $\{t>1 : f(\log t)>0\}$ exists if $f$ is a periodic function. The idea is to think as $E(e^u;4,3,1)$ as something close to a periodic function, which would imply that the logarithmic density of $\{t>1 : E(t;4,3,1)>0\}$ exists.

With Theorem 3.1, we transformed $E(e^u;4,3,1)$ into a sum of periodic functions. Any sum of $k$ periodic functions with respective period $T_1,\ldots,T_k>0$ whose span over the rationals is of dimension 1 is also a periodic function since there are integers $m_2,\ldots,m_k$

and $n_2, \ldots, n_k$ such that $T_j = \frac{m_j}{n_j} T_1$ and it would imply that the sum of periodic functions would have a period of $[m_2, \ldots, m_k] \cdot T_1$. If however the dimension of the span over the rationals is greater than 1, then we cannot have a periodic function and the numbers $T_1/T_j$ are not all rationals. However, with the rationals being dense in the real numbers, we could approximate any irrational number by a close enough rational number.

**Lemma 3.4** (Dirichlet's approximation theorem for simultaneous approximation)**.** *For every $(\alpha_1, \ldots, \alpha_d) \in \mathbb{R}^d$ and $N \in \mathbb{N}$, there exists fractions $\frac{a_1}{q}, \ldots, \frac{a_d}{q}$ with $1 \leqslant q \leqslant N^d$ such that*

$$\left|\alpha_i - \frac{a_i}{q}\right| < \frac{1}{qN}$$

*for all $i \leqslant d$.*

***Proof.*** Let's consider the set

$$\left\{([qN\alpha_1] \mod N, \ldots, [qN\alpha_d] \mod N) : 1 \leqslant q \leqslant N^d\right\},$$

which could be interpreted as a subset of the direct sum of $d$ copies of the group $\mathbb{Z}/N\mathbb{Z}$. By the pigeonhole principle, there exists $0 \leqslant q' < q'' \leqslant N^d$ such that

$$[q'N\alpha_i] \equiv [q''N\alpha_i] \pmod{N}$$

for all $i \leqslant d$. Let $q = q'' - q'$. This means that for each $i \leqslant d$, there exists a number $a_i$ such that

$$Na_i = [q''N\alpha_i] - [q'N\alpha_i] = qN\alpha_i + (\{q_1N\alpha_i\} - \{q_2N\alpha_i\}) \implies |qN\alpha_i - Na_i| < 1,$$

which implies the theorem by dividing both sides by $qN$. □

Fix differentiable periodic functions $f_1, \ldots, f_k$ with respective period $T_1, \ldots, T_k$. Let $\delta > 0$ and choose $N \in \mathbb{N}$ big enough such that $N \geqslant T_j/\delta$ for all $2 \leqslant j \leqslant k$. Then by the previous lemma, we can say that there exist fractions $\frac{a_2}{q}, \ldots, \frac{a_k}{q}$ with $1 \leqslant q \leqslant N^{k-1}$ such that for every $2 \leqslant j \leqslant k$, we have

$$\left|\frac{T_1}{T_j} - \frac{a_j}{q}\right| < \frac{1}{qN} \implies |qT_1 - a_jT_j| < \delta.$$

Hence, the sum of the periodic functions $S(x) = \sum f_j(x)$ are *quasiperiodic* in the sense that

$$S(x + qT_1) = \sum_{j=1}^{k} f_j(x + qT_1) = \sum_{j=1}^{k} f_j(x + a_jT_j + O(\delta)) = \sum_{j=1}^{k} f_j(x + O(\delta)) = S(x) + O_k(\delta)\,.$$

We can conclude that $E(e^u; 4, 3, 1)$ is indeed very close to a periodic function.

In fact, putting the $x$-axis on a logarithmic scale in Figure 3.1 makes the periodic structure apparent in the graph. It suggests that maybe there is a probability measure $\mu$ such that

$$\lim_{X\to\infty}\frac{1}{\log X}\int_1^X g(E(x;4,3,1))\frac{\mathrm{d}x}{x}=\lim_{X\to\infty}\frac{1}{X}\int_0^X g(E(e^u;4,3,1))\,\mathrm{d}u=\int_{-\infty}^{\infty}g(y)\,\mathrm{d}\mu(y).$$

## 3.4. Limiting distribution

In this section, we are going to prove the following theorem to ultimately show that the logarithmic density of the set $\{x>1 : E(x;4,3,1)>0\}$ exists:

**Theorem 3.5.** *Assuming GRH, there exists a probability measure $\mu$ such that*

$$\lim_{X\to\infty}\frac{1}{\log X}\int_1^X g(E(x;4,3,1))\frac{\mathrm{d}x}{x}=\int_{-\infty}^{\infty}g(y)\,\mathrm{d}\mu(y)$$

*for every Lipschitz continuous functions $g\colon\mathbb{R}\to\mathbb{R}$.*

To key to prove this theorem is the following classical lemma for which its proof is given in [12]:

**Lemma 3.6** (Kronecker-Weyl Theorem)**.** *Let $(\alpha_1,\dots,\alpha_d)\in\mathbb{R}^d$ and let's consider the set $A$ which is the closure of*

$$B=\{(e(t\alpha_1),\dots,e(t\alpha_d)) : t\in\mathbb{R}\}$$

*in $\mathbb{T}^d=\{(z_1,\dots,z_d) : |z_i|=1 \text{ for all } i\leqslant d\}$ is the $d$-torus. Then $A$ is an $r$-torus where $r$ is the dimension of the span of $\alpha_1,\dots,\alpha_d$ over $\mathbb{Q}$ and $B$ is equidistributed in $A$, which means that for any continuous function $h\colon\mathbb{R}^d\to\mathbb{R}$ we have*

$$\lim_{X\to\infty}\frac{1}{X}\int_0^X h(e(t\alpha_1),\dots,e(t\alpha_d))\,\mathrm{d}t=\int_A h(y)\,\mathrm{d}\nu(y),$$

*where $\nu$ is the normalized Haar measure on $A$.*

**Remark 3.7.** Note that the *Haar measure* of a locally compact abelian topological group is the unique measure, up to multiplication by a scalar, that is nonnegative, regular and invariant under translations. In the case where the measure is finite, a *normalized Haar measure* is to choose the Haar measure so that we get a probability measure (the whole space is of measure 1). In particular for the $d$-torus, let $\varphi\colon\mathbb{T}^d\to(-\frac{1}{2},\frac{1}{2}]^d$ be defined by

$$\varphi(z_1,\dots,z_d)=\frac{1}{2\pi}(\arg(z_1),\dots,\arg(z_d)),$$

then the normalized Haar measure of a set $E\subset\mathbb{T}^d$ is equal to the Lebesgue measure of the set $\varphi(E)$. In other words, the natural density of the values of $t>0$ such that $(e(t\alpha_1),\dots,e(t\alpha_d))$ is in some subset $S\subset\mathbb{T}^d$ is equal to the proportion of the $d$-torus this subset $S$ occupies.

***Proof of Theorem 3.5.*** For $T > 0$, let $k = \frac{1}{2}N(T,\chi_1)$, the imaginary parts of the first $k$ zeros of $L(s,\chi_1)$ above the $x$-axis in ascending order $0 \leqslant \gamma_1 \leqslant \ldots \leqslant \gamma_k$ and

$$G_T(y_1,\ldots,y_k) = g\left(1 + \sum_{j=1}^{k} \frac{\operatorname{Re}(y_j) + 2\gamma_j \operatorname{Im}(y_j)}{\frac{1}{4}+\gamma_j^2}\right). \tag{3.11}$$

Let's consider the set $A$, which is the closure of the set

$$B = \left\{(e^{iu\gamma_1},\ldots,e^{iu\gamma_k}) : u \in \mathbb{R}\right\}.$$

Since the function $G_T$ is continuous on $A$, the Kronecker-Weyl theorem implies that $A$ is an $r$-torus where $r$ is the dimension of the span of the first $k$ zeros of $L(s,\chi_1)$ and we can define the normalized Haar measure on $A$ as $\mathrm{d}a$ to get

$$\int_A G_T(a)\,\mathrm{d}a = \lim_{U\to\infty}\frac{1}{U}\int_0^U G_T(e^{iu\gamma_1},\ldots,e^{iu\gamma_k})\,\mathrm{d}u = \lim_{U\to\infty}\frac{1}{U}\int_0^U g(E(e^u;4,3,1) - \varepsilon(u;T))\,\mathrm{d}u.$$

If we let $c_g$ be a Lipschitz constant for $g$, then

$$\int_A G_T(a)\,\mathrm{d}a = \lim_{U\to\infty}\frac{1}{U}\int_0^U g(E(e^u;4,3,1)) + O(c_g\,|\varepsilon(u;T)|)\,\mathrm{d}u.$$

By the Cauchy-Schwarz inequality, the error term inside the limit is

$$\ll \frac{c_g}{\sqrt{U}}\left(\int_0^U |\varepsilon(u;T)|^2\,\mathrm{d}u\right)^{1/2} \ll c_g\left(\frac{\log T}{\sqrt{T}} + \frac{1}{\sqrt{U}}\right)$$

which means that by letting $U \to \infty$, we get

$$\int_A G_T(a)\,\mathrm{d}a + O\!\left(\frac{c_g \log T}{\sqrt{T}}\right) \leqslant \liminf_{U\to\infty}\frac{1}{U}\int_0^U g(E(e^u;4,3,1))\,\mathrm{d}u$$
$$\leqslant \limsup_{U\to\infty}\frac{1}{U}\int_0^U g(E(e^u;4,3,1))\,\mathrm{d}u \leqslant \int_A G_T(a)\,\mathrm{d}a + O\!\left(\frac{c_g\log T}{\sqrt{T}}\right).$$

Hence

$$\limsup_{T\to\infty}\int_A G_T(a)\,\mathrm{d}a \leqslant \liminf_{U\to\infty}\frac{1}{U}\int_0^U g(E(e^u;4,3,1))\,\mathrm{d}u$$
$$\leqslant \limsup_{U\to\infty}\frac{1}{U}\int_0^U g(E(e^u;4,3,1))\,\mathrm{d}u \leqslant \liminf_{T\to\infty}\int_A G_T(a)\,\mathrm{d}a,$$

which means that the following limits exist and are equal:

$$\lim_{T\to\infty}\int_A G_T(a)\,\mathrm{d}a = \lim_{U\to\infty}\frac{1}{U}\int_0^U g(E(e^u;4,3,1))\,\mathrm{d}u. \tag{3.12}$$

By the change of variables $u = \log x$, we can finish the proof of the theorem. □

**Corollary 3.8.** *If we assume GRH and that $\mu(\{0\}) = 0$ where $\mu$ is the probability measure defined in Theorem 3.5, then the logarithmic density of $S_0 := \{x > 1 : \pi(x;4,3) > \pi(x;4,1)\}$*

*defined by*

$$\delta(S_0) = \lim_{X\to\infty} \frac{1}{\log X} \int_{[1,X]\cap S_0} \frac{\mathrm{d}t}{t}$$

*exists and is equal to* $\mu((0,\infty))$.

***Proof.*** This is a consequence of the Portmanteau Theorem from probability theory (see Theorem 13.16 of [14] for the general version of the theorem), which says that if we let a bounded sequence of positive probability measures $\mathbb{P}_n$ converge weakly to a measure $\mathbb{P}$ on a metric space, in the sense that for any Lipschitz continuous random variables $X$ we have $\mathbb{E}_n[X] \to \mathbb{E}[X]$, then we have $\mathbb{P}_n[A] \to \mathbb{P}[A]$ for any continuity sets $A$, meaning that those sets have the property $\mu(\partial A) = 0$. Since we assumed that $\mu(\{0\}) = 0$, then $(0,\infty)$ is a continuity set with respect to the measure $\mu$. □

## 3.5. Primes 1 (mod 4) take the lead

To show that the logarithmic density $\delta(S_0) > 1/2$, we can first prove that $\delta(S_1) = 1/2$, where $S_1 := \{x > 1 : E(x; 4, 3, 1) > 1\}$, hence quantifying the fact that the function $E(e^u; 4, 3, 1)$ is oscillating around 1. To understand the symmetry, we need the Fourier transform of the measure $\mu$ defined in Theorem 3.5. We are going to need the imaginary parts of the first $k$ zeros of $L(s,\chi_1)$ to be linearly independent over $\mathbb{Q}$ in order to fix the set $A = \mathbb{T}^k$ in the Kronecker-Weyl Theorem (Lemma 3.6). Unfortunately, this has not been proved, but it is conjectured that for any fixed $q \geqslant 1$, the set

$$\bigcup_{\substack{\chi \ (\mathrm{mod}\ q) \\ \chi \text{ primitive}}} \{\gamma : L(\beta + i\gamma, \chi) = 0, 0 < \beta < 1, \gamma \geqslant 0\}$$

is linearly independent over $\mathbb{Q}$. This conjecture is called the *Linear Independence Hypothesis* (LI).

**Theorem 3.9.** *Let $\mu$ be defined as the measure in Theorem 3.5. Assuming GRH and LI, the Fourier transform of $\mu$ is*

$$\widehat{\mu}(\xi) = e(-\xi) \prod_{\gamma>0} J_0\left(\frac{4\pi\xi}{\sqrt{\frac{1}{4}+\gamma^2}}\right),$$

*where the product is over the imaginary parts of the zeros of $L(s,\chi_1)$ with $\chi_1$ being the nonprincipal character mod* 4*. The function $J_0$ is the Bessel function of the first kind defined by*

$$J_0(z) = \int_0^1 e^{-iz\cos(2\pi t)}\, \mathrm{d}t.$$

***Proof.*** Let $g(y) = e(-\xi y)$ and let $T, k, \gamma_1, \ldots, \gamma_k$ and $G_T$ as in the beginning of the proof of Theorem 3.5. We can define $G_T$ as in (3.11). Since $g$ is a Lipschitz function, the Fourier transform of $\mu$ is

$$\widehat{\mu}(\xi) = \int_{-\infty}^{\infty} g(y)\,\mathrm{d}\mu(y) = \lim_{T\to\infty}\int_{\mathbb{T}^k} G_T(a)\,\mathrm{d}a$$

by Theorem 3.5 and (3.12). The measure d$a$ is the normalized Haar measure of $\mathbb{T}^k$, we have

$$\begin{aligned}\widehat{\mu}(\xi) &= \int_0^1 \dots \int_0^1 e\left(-\xi\left(1+\sum_{j=1}^{k}\frac{\cos(2\pi x_j)+2\gamma_j\sin(2\pi x_j)}{\frac{1}{4}+\gamma_j^2}\right)\right)\mathrm{d}x_1\,\dots\,\mathrm{d}x_j\\ &= e(-\xi)\prod_{j=1}^{k}\int_0^1 e\left(-\frac{\xi}{\frac{1}{4}+\gamma_j^2}(\cos(2\pi x)+2\gamma_j\sin(2\pi x))\right)\mathrm{d}x. \quad (3.13)\end{aligned}$$

Let $T_j$ be a right triangle with its two legs being of lengths 1 and $2\gamma_j$. We can define $\theta_j$ as the acute angle adjacent to the leg of length 1. By the Pythagorean theorem, the hypotenuse of $T_j$ is of length $2\sqrt{\frac{1}{4}+\gamma_j^2}$. Thus, we have

$$\cos(2\pi x-\theta_j) = \cos(\theta_j)\cos(2\pi x)+\sin(\theta_j)\sin(2\pi x) = \frac{\cos(2\pi x)+2\gamma_j\sin(2\pi x)}{2\sqrt{\frac{1}{4}+\gamma_j^2}}.$$

We can apply this in (3.13) and use the fact that we are integrating over the period of the function $x\mapsto\cos(2\pi x)$ to obtain

$$\widehat{\mu}(\xi) = e(-\xi)\prod_{j=1}^{k}\int_0^1 e\left(-\frac{2\xi\cos(2\pi(x-\theta/2\pi))}{\sqrt{\frac{1}{4}+\gamma_j^2}}\right)\mathrm{d}x = e(-\xi)\prod_{j=1}^{k}J_0\left(\frac{4\pi\xi}{\sqrt{\frac{1}{4}+\gamma_j^2}}\right).$$

□

We can understand any probability measure if we understand its Fourier transform[2]. Since $e(\xi)\widehat{\mu}(\xi)$ is an even function and is real-valued for real values of $\xi$, we know that the distribution of $\mu$ is symmetric around 1 by Fourier inversion, which means that $\delta(S_1)=\frac{1}{2}$. More specifically, we can use Fourier inversion to find a formula for $\mu([1,1+x))$ for every $x>0$. Indeed, since the Fourier transform of the indicator function $\mathbf{1}_{(-x,x)}$ is $\frac{\sin(2\pi x\xi)}{\pi\xi}$, then for every $x>0$, we have

$$\mu((1-x,1+x))+\frac{1}{2}(\mu(\{1-x\})+\mu(\{1+x\})) = \int_{-\infty}^{\infty}\frac{\sin(2\pi x\xi)}{\pi\xi}\prod_{\gamma>0}J_0\left(\frac{4\pi\xi}{\sqrt{\frac{1}{4}+\gamma^2}}\right)\mathrm{d}\xi.$$

If we assume $\mu\ll\lambda$, then by symmetry, we obtain

$$\mu([1,1+x)) = \int_0^{\infty}\frac{\sin(2\pi x\xi)}{\pi\xi}\prod_{\gamma>0}J_0\left(\frac{4\pi\xi}{\sqrt{\frac{1}{4}+\gamma^2}}\right)\mathrm{d}\xi. \quad (3.14)$$

**Lemma 3.10.** *The function $x\mapsto\mu([1,1+x))$ is real analytic for every $x>0$.*

[2] In probability theory, we would look at the mapping $x\mapsto E(x;4,3,1)$ as a random variable in the sample space $(1,\infty)$ with the logarithmic density $\delta$ as its probability measure. The event space would be the $\sigma$-algebra generated by $E(x;4,3,1)$. In this context, $\mu$ is the probability distribution and $\widehat{\mu}(\xi)$ is the characteristic function of the random variable $E(x;4,3,1)$.

***Proof.*** A characterization of real analytic functions is given in Lemma 1.2.9 of [17]: A function $f$ is real analytic on an open set $U$ if, and only if, it is real smooth and for every compact set $K \subset U$, there exists $C > 0$ such that

$$f^{(k)}(x) \ll C^k k!.$$

for every $x \in K$ and every $k \geqslant 0$. To apply this characterization, we need to prove that the function $f\colon (0,\infty) \to [0,\infty)$ defined by $f(x) := x \mapsto \mu([1, 1+x))$ is a real smooth function and find all of its derivatives.

First, we will need a bound on the infinite product. The Bessel function has the following asymptotic: as $x > 0$, we have

$$|J_0(x)| \leqslant \min\left\{1, \sqrt{\tfrac{2}{\pi x}}\right\}. \tag{3.15}$$

This is a consequence of Theorem 5.1 in [8]. Since every imaginary parts $\gamma$ of the nontrivial zeros of $L(s, \chi_1)$ are $|\gamma| > 6$ (see [29]), there exists a positive constant $c$ such that

$$\log \prod_{\gamma>0} \left| J_0\left(\frac{4\pi\xi}{\sqrt{\frac{1}{4}+\gamma^2}}\right)\right| \leqslant c \sum_{6<\gamma\leqslant\xi} \log(\gamma/\xi) = \frac{c}{2}\int_6^\xi \log(t/\xi)\,\mathrm{d}N(t,\chi_1) = -\frac{c}{2}\int_6^\xi \frac{N(t,\chi_1)}{t}\,\mathrm{d}t.$$

Thus we can conclude that $\prod_{\gamma>0} J_0(4\pi\xi/\sqrt{\frac{1}{4}+\gamma^2}) \ll e^{-c'\xi}$ for some other absolute positive constant $c'$ by using Theorem 2.15. Hence, the integrand of (3.14) is Lebesgue-integrable for every fixed $x > 0$.

We should first note that if $g(x,\xi) = \sin(2\pi x\xi)$, then

$$\left|\frac{\partial^k g}{\partial x^k}(x,\xi)\right| \leqslant (2\pi\xi)^k.$$

Let $M > 0$. To prove that all the derivatives of $f$ exist for every $0 < x < M$ and every $\xi > 0$, we can find the upper bound

$$\frac{\sin(2\pi x\xi)}{\pi\xi} \prod_{\gamma>0} J_0\left(\frac{4\pi\xi}{\sqrt{\frac{1}{4}+\gamma^2}}\right) \ll_{\chi_1} Me^{-c'\xi}.$$

Since this upper bound is integrable over $(0,\infty)$, then by the dominated convergence theorem, we can conclude that $f$ is differentiable for every $x \in (0, M)$. Furthermore, for every $k \geqslant 1$, we have

$$\frac{1}{\pi\xi} \cdot \frac{\partial^k g}{\partial x^k}(x,\xi) \prod_{\gamma>0} J_0\left(\frac{4\pi\xi}{\sqrt{\frac{1}{4}+\gamma^2}}\right) \ll_{\chi_1} (2\pi\xi)^{k-1}e^{-c'\xi}.$$

which is also integrable on $(0,\infty)$. Thus we can say that $f$ is real smooth over $(0, M)$ and its derivative is

$$f^{(k)}(x) = \int_0^\infty \frac{1}{\pi\xi} \cdot \frac{\partial^k g}{\partial x^k}(x,\xi) \prod_{\gamma>0} J_0\left(\frac{4\pi\xi}{\sqrt{\frac{1}{4}+\gamma^2}}\right) \mathrm{d}\xi \ll \int_0^\infty (2\pi\xi)^{k-1}e^{-c'\xi}\,\mathrm{d}\xi \ll (2\pi/c')^k k!.$$

Hence we can conclude that $f(x)$ is real analytic over $(0, M)$ and since $M$ can be arbitrarily large, this proves the theorem. □

Having a real analytic distribution means that $\mu([1, 1+x))$ cannot be identically zero on an open interval $I \subset (0, \infty)$. Thus $\mu([1, 1+x))$ is strictly decreasing for $x > 0$, and we arrive to the following theorem:

**Theorem 3.11.** *Assuming GRH, LI and* $\mu \ll \lambda$*, the logarithmic density of* $S_0 = \{x > 1 : \pi(x; 4, 3) > \pi(x; 4, 1)\}$ *is strictly between* $1/2$ *and* $1$*.*

From this theorem, we deduce that $\sup S_0^c = \infty$, which means that in the prime number race, the primes 1 mod 4 take the lead infinitely many times. Since we can generalize the theorem to the $\pi(x)$ vs. $\mathrm{Li}(x)$ race, we can also conclude that under the assumptions in 3.11, there must necessarily exists an unbounded sequence of real numbers $x_n$ such that $\pi(x_n) > \mathrm{Li}(x_n)$, even with the fact that we still don't know where this sequence could start. Finally, using the Fourier transform from Theorem 3.9, Rubinstein and Sarnak computed in Section 4 of [27] that the logarithmic density of the set $A_1 := \{x > 1 : \mathrm{Li}(x) > \pi(x)\}$ is $0.99999973\ldots$

In general for $q \geqslant 2$, let $A_q$ be the set of $x > 1$ such that there are more primes less than $x$ which are quadratic nonresidues mod $q$ than quadratic residues mod $q$. In particular, $A_4 = S_0$ from Corollary 3.8. Rubinstein and Sarnak also computed that

$$\delta(A_3) = 0.9990\ldots$$
$$\delta(A_4) = 0.9959\ldots$$
$$\delta(A_5) = 0.9954\ldots$$
$$\delta(A_7) = 0.9782\ldots$$
$$\delta(A_{11}) = 0.9167\ldots$$
$$\delta(A_{13}) = 0.9443\ldots$$

It seems like $\delta(A_q)$ seems to gets closer to $1/2$ as $q$ gets bigger. We will see that this is the case if $q$ is restricted to odd primes[3].

[3] This statement is not true if we let $q$ be any natural number. Daniel Fiorilli proved in [6] that there exists a sequence of natural numbers $q_n$ such that $\delta(A_{q_n}) \to 1$. For example, $\delta(A_{4849845}) = 0.999999928\ldots > \delta(A_1)$. This heavier bias comes from the fact that a number with a lot of distinct prime factors has much more quadratic nonresidues classes than quadratic residues classes (in the example, $4849845 = 3\cdot5\cdot7\cdot11\cdot13\cdot17\cdot19$). The only moduli for which we have the same number of quadratic residue and nonresidue classes are 4, odd prime powers, or numbers which are two times an odd prime power.

## 3.6. CLT-like theorem

In most introductory statistics class, one usually learns that if $n$ is large (in applications, this $n$ has to be generally larger than 30, see [19] p. 280) and $x_1, \ldots, x_n$ are $n$ random observations independently drawn from the same distribution with finite mean $\mu$ and finite variance $\sigma^2$, whatever this distribution is, then the sample mean $\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j$ is approximately normally distributed with mean $\mu$ and variance $\sigma^2/n$. This comes from the *Central Limit Theorem* (CLT) from probability. Formally, the classical version of the CLT can be stated as follows: Let $X_1, X_2, \ldots$ be a sequence of independent and identically distributed such that $\mathbb{E}[X_1] = 0$ and $\mathrm{Var}(X_1) = 1$ and let $S_n = \sum_{j=1}^{n} X_j$. Then $S_n/\sqrt{n}$ converges in distribution to a standard normal distribution, which have the density function $(2\pi)^{-1/2}e^{-x^2/2}$.

Similarly as what we discussed at the beginning of Section 3.2, we can replace 4 by any odd prime $q$ in every statement of this chapter and the theorems would still hold, except that we would be working with the nontrivial zeros of the $L$-function $L(s, \chi_q)$ where $\chi_q$ is the Legendre symbol mod $q$. In particular, if we let $E_{N,R}(x;q) := \frac{\log x}{\sqrt{x}} \sum_{p \leqslant x} \left(\frac{p}{q}\right)$, then for every fixed $q$, $E_{N,R}(x;q)$ have a limiting distribution $\mu_q$, similar to the one for $E(x;4,3,1)$ in Theorem 3.5, and the Fourier transform of $\mu_q$ is

$$\widehat{\mu}_q(\xi) = e(-\xi) \prod_{\gamma_{\chi_q} > 0} J_0\left(\frac{4\pi\xi}{\sqrt{\frac{1}{4} + \gamma_{\chi_q}^2}}\right)$$

where the product is now over the imaginary parts $\gamma_{\chi_q}$ of the nontrivial zeros of $L(s, \chi_q)$. Like the CLT, the limiting distribution of $E_{N,R}(x;q)$ is approximately a normal distribution of mean 1 and variance $\log q$ as $q \to \infty$ over primes. This implies that the bias dissipates in the prime number race between residues mod $q$ and nonresidues mod $q$ as $q \to \infty$. Here is the formal statement of the theorem:

**Theorem 3.12.** *Let $\mu_q^*$ be the limiting distribution of $(E_{N,R}(x;q) - 1)/\sqrt{\log q}$, in the sense that*

$$\int_{-\infty}^{\infty} g(y)\, \mathrm{d}\mu_q^*(y) = \lim_{X \to \infty} \frac{1}{\log X} \int_1^X g\left(\frac{E_{N,R}(x;q) - 1}{\sqrt{\log q}}\right) \frac{\mathrm{d}x}{x}$$

*for every Lipschitz continuous functions $g\colon \mathbb{R} \to \mathbb{R}$. If we let $q_k$ be the $k^{th}$ odd prime, then $\mu_{q_k}^*$ converges in measure to a standard normal distribution as $k \to \infty$.*

***Proof.*** As in the proof of the Central Limit Theorem, the key is Lévy's continuity theorem from probability which states that a sequence of random variables $(X_n)$ converges in distribution to $X$ if, and only if, the characteristic functions of the $X_n$ converge pointwise to the characteristic function of $X$.

For an arbitrary odd prime $q$, we have that

$$\widehat{\mu}_q^*(\xi) = \prod_{\gamma_{\chi_q}>0} J_0\left(\frac{4\pi\xi}{\sqrt{\log q(\frac{1}{4}+\gamma_{\chi_q}^2)}}\right)$$

by the scaling and translation properties of the Fourier transform. Thus, for every $\xi > 0$, we can use $\log J_0(z) = -z^2/4 + O(z^4)$ for $z \to 0$ from the Taylor series of $J_0(z)$ and $\log(1-z)$ to obtain

$$\log \widehat{\mu}_q^*(\xi) = -\frac{4\pi^2\xi^2}{\log q}\sum_{\gamma_{\chi_q}>0}\frac{1}{\frac{1}{4}+\gamma_{\chi_q}^2} + O_\xi\left(\frac{1}{\log^2 q}\sum_{\gamma_{\chi_q}>0}\frac{1}{(\frac{1}{4}+\gamma_{\chi_q}^2)^2}\right). \tag{3.16}$$

The sum $\sum(\frac{1}{4}+\gamma_{\chi_q}^2)^{-1}$ can be deduced from p. 83 of [5]:

$$\sum_{\gamma_{\chi_q}>0}\frac{1}{\frac{1}{4}+\gamma_{\chi_q}^2} = \frac{\log q}{2} + \frac{L'(1,\chi_q)}{L(1,\chi_q)} + O(1) = \frac{\log q}{2} + O(\log\log q) \tag{3.17}$$

where the last bound ($L'(1,\chi_q)/L(1,\chi_q) \ll \log\log q$) is given by Littlewood in [18] under GRH. Furthermore, since $(\frac{1}{4}+\gamma_{\chi_q}^2)^{-1} \leqslant 4$, then

$$\sum_{\gamma_{\chi_q}>0}\frac{1}{(\frac{1}{4}+\gamma_{\chi_q}^2)^2} \leqslant 4\sum_{\gamma_{\chi_q}>0}\frac{1}{\frac{1}{4}+\gamma_{\chi_q}^2} \ll \log q. \tag{3.18}$$

By applying the estimate (3.17) and the bound (3.18) in the equation (3.16), we finally obtain

$$\log \widehat{\mu}_q^*(\xi) = -2\pi^2\xi^2 + O_\xi\left(\frac{\log\log q}{\log q}\right) \implies \widehat{\mu}_q^*(\xi) = e^{-2\pi^2\xi^2} + O_\xi\left(\frac{\log\log q}{\log q}\right).$$

Thus since $q_k$ represents the $k^{\text{th}}$ odd prime, then for any fixed $\xi$, we have

$$\lim_{k\to\infty}\widehat{\mu}_{q_k}^*(\xi) = e^{-2\pi^2\xi^2}$$

which is the Fourier transform of the density function of the standard normal distribution. We can finally prove the theorem by applying Lévy's continuity theorem. □

We can observe in Figure 3.4 how the standard deviation gets larger as $q$ grows, and how the bias is really apparent for $q = 11$, but not for $q = 100003$.

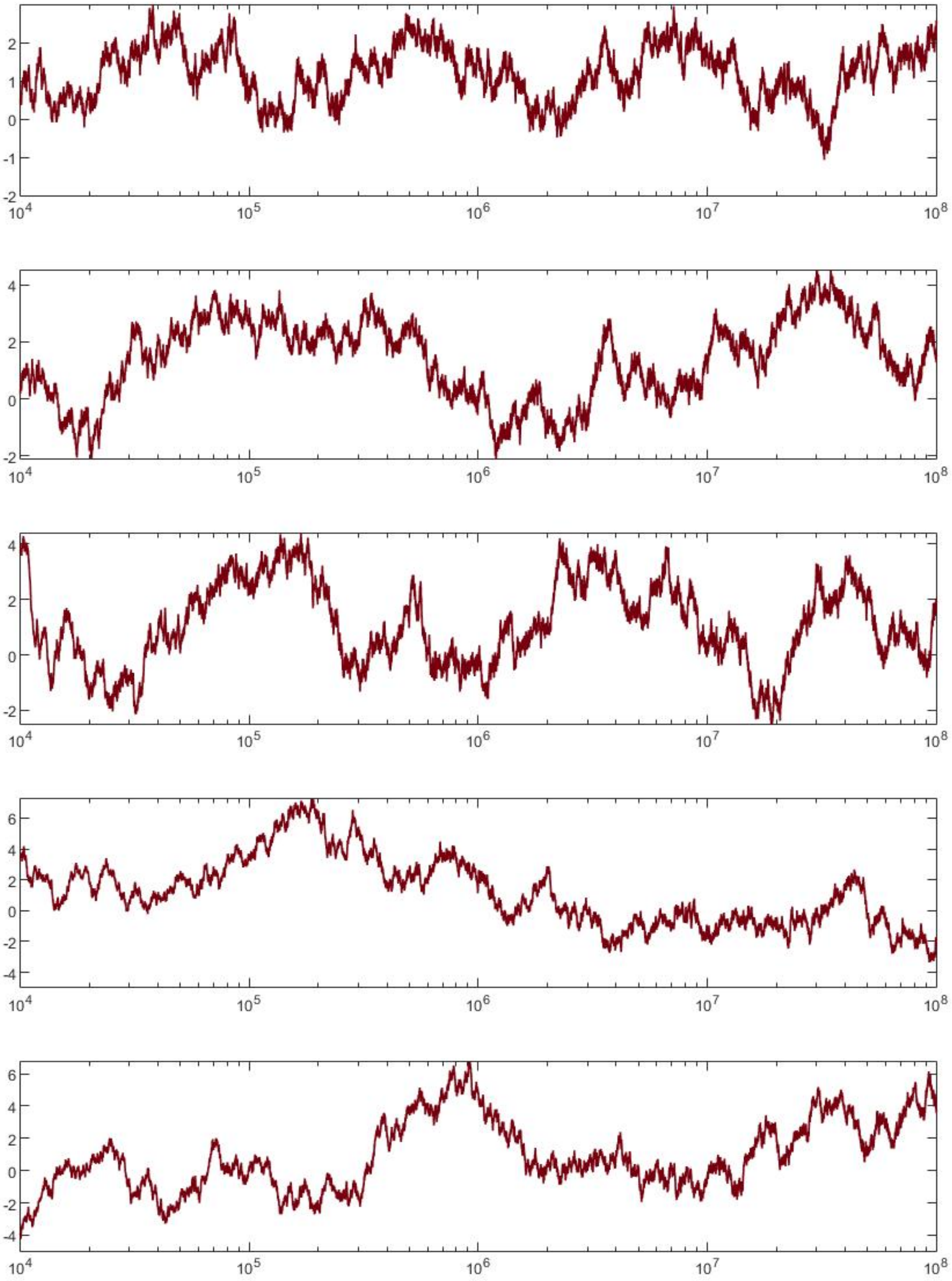

**Fig. 3.4.** Graphs of the function $E_{N,R}(x;q) = \frac{\log x}{\sqrt{x}} \sum_{p\leqslant x} \left(\frac{p}{q}\right)$ for $q = 11, 101, 1009, 10007$ and 100003 (from top to bottom).

# Chapter 4

# Selberg's sieve method

To theoretically study the race of $\pi_+(x)$ vs. $\pi_-(x)$, we need to know the asymptotic size of $\pi_+(x)$ and $\pi_-(x)$. However, it still has not been proven if both these functions tend to infinity. In the next three chapters, we will try to make sense of Conjecture 1.1.

The Dirichlet series ideas of Chapter 2 cannot be directly used to estimate the size of the set of Sophie Germain primes up to $x$ since the indicator function of $\{n \in \mathbb{N} : 2n+1 \text{ is prime}\}$ is neither multiplicative nor periodic. In this chapter, we will prove an upper bound on $\pi_+(x)$ and $\pi_-(x)$, which is of the same order of magnitude as their conjectured asymptotic size.

**Theorem 4.1.** *For $x \geqslant 3$, we have*

$$\pi_+(x), \pi_-(x) \leqslant \left(4c_2 + O\left(\frac{\log\log x}{\log x}\right)\right)\frac{x}{\log^2 x}.$$

The proof method is due to Selberg, who showed his technique to obtain an upper bound in [28]. We will directly apply his method to our situation. To have a more general perspective of sieve methods, we refer the reader to Part 4 of [16].

## 4.1. Sieve methods

One of the earliest known algorithms generating a list containing every prime numbers is called the *sieve of Eratosthenes*. It comes from the simple idea that every composite number $n \geqslant 4$ has a prime factor less than or equal to $\sqrt{n}$.

Let $I_n := \left(2^{2^{n-1}}, 2^{2^n}\right]$ for $n \geqslant 1$. We start with a list $\mathcal{P}_0$ containing only the number 2 (we immediately deduce its primality by noticing that there is no room for another positive factor between 1 and 2). Assuming we have created the finite list $\mathcal{P}_{n-1}$, we remove every multiple of every prime in $\mathcal{P}_{n-1}$ from the interval $I_n$. We then create a new list $\mathcal{P}_n$ containing $\mathcal{P}_{n-1}$ and the remaining integers in $I_n$. This way, by induction, $\mathcal{P}_n$ is a list of every prime $p \leqslant 2^{2^n}$.

The word "sieve" comes from the idea of sifting the elements of a set. The integers with small prime factors will pass through the sieve, leaving behind only integers with large prime factors, which we are interested in. As in the sieve of Eratosthenes, if we know where the primes less than $\sqrt{x}$ are, then we can have a formula for $\pi(x)$ of the form

$$\pi(x) - \pi(\sqrt{x}) = \#\{p \in (\sqrt{x}, x]\} = \#\left\{n \leqslant x : n \text{ is coprime to } \prod_{p \leqslant \sqrt{x}} p\right\} - 1. \tag{4.1}$$

The $-1$ in the equation comes from the fact that 1 is coprime to every positive integer.

For notational simplicity, we will note from now on

$$P(y) := \prod_{p \leqslant y} p.$$

Saying that a number $n$ is coprime to $P(y)$ is equivalent to saying that $n$ is a *$y$-rough number*, which means that $n$ has no prime factors less than or equal to $y$. Also, the relation $d \mid P(y)$ is equivalent to $d$ being squarefree and *$y$-smooth*, that is to say, every prime factor of $d$ is less than or equal to $y$.

The formula (4.1) is useful because of Möbius inversion, since it gives us have a formula for the number of primes without the use of the Prime Number Theorem (by using the trivial estimate $\pi(\sqrt{x}) \leqslant \sqrt{x}$):

$$\pi(x) = \sum_{n \leqslant x} \mathbf{1}_{(n, P(\sqrt{x}))=1} + O\Big(\sqrt{x}\Big) = \sum_{n \leqslant x} \sum_{\substack{d \mid P(\sqrt{x}) \\ dr = n}} \mu(d) + O\Big(\sqrt{x}\Big) \tag{4.2}$$

and by changing the order of summation, we get

$$\pi(x) = \sum_{d \mid P(\sqrt{x})} \mu(d) \left[\frac{x}{d}\right] + O\Big(\sqrt{x}\Big).$$

Note that this equation is essentially the inclusion-exclusion principle where we are subtracting 1 from the set $\{n \leqslant x\}$ for every multiple of every prime $p \leqslant \sqrt{x}$, and adding 1 for every multiple of every product of two distinct primes $p_1 < p_2 \leqslant \sqrt{x}$ since we've removed too much at the first step, and subtracting 1 for product of three distinct primes, and so on and so forth.

We can approach Sophie Germain primes the same way. Let's define

$$\pi_+(x; y) := \#\left\{p \leqslant x : (2p+1, P_3(y)) = 1\right\}$$

where $y \leqslant \sqrt{2x+1}$ and $P_3(y)$ is defined[1] as

$$P_3(y) := \prod_{3 \leqslant p \leqslant y} p.$$

[1] We could use $P(y)$ instead of $P_3(y)$ but it is not necessary since $2p+1$ can never be divisible by 2.

If $y = \sqrt{2x+1}$, then $\pi_+(x; \sqrt{2x+1})$ counts Sophie Germain primes between $\sqrt{2x+1}$ and $x$. We introduced the parameter $y$ because we want control over how many integers our sieve is retaining.

If we find a way to upper bound $\pi_+(x; y)$, then we could try to get an upper bound on $\pi_+$ by using the simple observation that

$$\pi_+(x) \leqslant \pi_+(x; y) + O\left(\frac{y}{\log y}\right) \quad \text{for } y \leqslant \sqrt{2x+1}, \tag{4.3}$$

since $p$ and $2p+1$ being both primes implies either that $2p+1$ does not have a prime factor less than $y$ or that $p \leqslant \frac{y-1}{2}$.

We can have an exact formula for $\pi_+(x; y)$ by using the same idea as in (4.2):

$$\pi_+(x; y) = \sum_{p \leqslant x} \mathbf{1}_{(2p+1, P_3(y))=1} = \sum_{p \leqslant x} \sum_{\substack{d \mid 2p+1 \\ d \mid P_3(y)}} \mu(d) = \sum_{d \mid P_3(y)} \mu(d) N_d(x) \tag{4.4}$$

where $N_d(x) = \#\{p \leqslant x : 2p \equiv -1 \pmod{d}\}$. Since we are summing over odd values of $d$, we know that 2 has a multiplicative inverse mod $d$ which we call $2^{-1}$ and $N_d(x) = \pi(x; d, -2^{-1})$. Hence, by assuming GRH and asking for $y \ll \log(x)$, we obtain $P_3(y) \ll_A x^A$ for some positive $A$, and Theorem 2.1 leads us to

$$\pi_+(x; y) = \sum_{d \mid P_3(y)} \mu(d) \pi(x; d, -2^{-1}) = \mathrm{Li}(x) \sum_{d \mid P_3(y)} \frac{\mu(d)}{\phi(d)} + O\Big(2^{\pi(y)} \sqrt{x} \log x\Big). \tag{4.5}$$

Since $\mu(d)/\phi(d)$ is a multiplicative function, we have

$$\pi_+(x; y) = \mathrm{Li}(x) \prod_{3 \leqslant p \leqslant y} \left(1 - \frac{1}{p-1}\right) + O\Big(2^{\pi(y)} \sqrt{x} \log x\Big). \tag{4.6}$$

To estimate the product, we can use the logarithm function

$$\begin{aligned} \log \prod_{3 \leqslant p \leqslant y} \left(1 - \frac{1}{p-1}\right) = \sum_{3 \leqslant p \leqslant y} \log \left(1 - \frac{1}{p-1}\right) &= - \sum_{3 \leqslant p \leqslant y} \left(\frac{1}{p} + O\left(\frac{1}{p^2}\right)\right) \\ &= -\log\log y + c + O\Big(e^{-c'\sqrt{\log y}}\Big) \end{aligned}$$

where $c, c'$ are constants and the last equality comes from partial summation and the Prime Number Theorem[2]. By exponentiating, we retrieve the product and prove that

$$\pi_+(x; y) = \frac{e^c \,\mathrm{Li}(x)}{\log y} + O\left(2^{\pi(y)} \sqrt{x} \log x + \frac{x e^{-c''\sqrt{\log y}}}{\log x}\right)$$

[2] Mertens proved, without the Prime Number Theorem, that there exists a constant $c'$ such that $\sum_{p \leqslant x} 1/p = \log\log x + c' + O(1/\log x)$ using the convolution $1 * \Lambda = \log$ proved in Example 2.9.

for some positive constant $c''$. But asking for $y \ll \log x$ cannot lead us to Theorem 4.1, and the best upper bound we can get with this strategy is

$$\pi_+(x) \ll \frac{x}{\log x \log\log x}.$$

The problem was that we had to sum too many terms together in (4.5).

## 4.2. Sieve weights

The key idea of sieve methods is to replace the Möbius function with another function that is easier to work with. Going back to (4.4), if we find a sequence $\rho_d$ supported on the divisors of $P_3(y)$ such that

$$\mathbf{1}_{(n,P_3(y))=1} \leqslant \sum_{d|n} \rho_d \tag{4.7}$$

with the extra condition that $\rho_d = 0$ if $d > D$ for another parameter $D$, we would be able to control how many times we are sieving our set while simulating the behaviour of $\mu(d)$ and get an upper bound on $\pi_+(x; y)$.

Selberg's sieve constructs the weights $\rho_d$ by optimizing the inequality

$$\mathbf{1}_{(n,P_3(y))=1} \leqslant \left( \sum_{d|n} \lambda_d \right)^2. \tag{4.8}$$

Here, $\lambda_d$ is a sequence of real numbers satisfying $\lambda_1 = 1$ and supported on

$$\mathcal{D} := \left\{ d : d \mid P_3(y) \text{ and } d \leqslant \sqrt{D} \right\}.$$

Expanding the right-hand side in (4.8), we get an upper bound sieve by having

$$\rho_d = \sum_{[d_1,d_2]=d} \lambda_{d_1}\lambda_{d_2} \tag{4.9}$$

in Equation (4.7).

We can now apply our sieve on $\pi_+(x; y)$:

$$\pi_+(x; y) = \sum_{p \leqslant x} \mathbf{1}_{(2p+1,P_3(y))=1} \leqslant \sum_{p \leqslant x} \sum_{d|2p+1} \rho_d = \sum_{d_1,d_2 \in \mathcal{D}} \lambda_{d_1}\lambda_{d_2}\pi(x; [d_1, d_2], -2^{-1}). \tag{4.10}$$

Note that if we were to study the function $\pi_2(x) = \{p \leqslant x : p + 2 \text{ is prime}\}$ instead of $\pi_+(x)$, then we would arrive at the same bound by this method since we would only have to replace $\pi(x; [d_1, d_2], -2^{-1})$ by $\pi(x; [d_1, d_2], -2)$ and these two quantities are of the same size as seen in Chapter 3. If we define $r_{d_1,d_2}$ via the equation

$$\pi(x; [d_1, d_2], -2^{-1}) = \frac{\mathrm{Li}(x)}{\phi([d_1, d_2])} + r_{d_1,d_2}$$

and we use this relation in (4.10), we get

$$\pi_+(x;y) \leqslant \mathrm{Li}(x) \sum_{d_1,d_2\in\mathcal{D}} \frac{\lambda_{d_1}\lambda_{d_2}}{\phi([d_1,d_2])} + \sum_{d_1,d_2\in\mathcal{D}} \lambda_{d_1}\lambda_{d_2} r_{d_1,d_2}. \tag{4.11}$$

Notice that equation (4.11) is true, no matter how we choose our sequence $\lambda_d$. If we look at every $\lambda_d$ in the support as free variables, we get a quadratic form, and Selberg's strategy was to choose the ideal sequence that minimizes the main term.

For a multiplicative function $f$ and any two squarefree numbers $a$ and $b$, we prove the propriety

$$f(a)f(b) = f(a)f\left(\tfrac{b}{(a,b)}\right)f((a,b)) = f([a,b])f((a,b)).$$

Thus, by using this property on $\phi$ in (4.11), we change the lcm into a gcd:

$$\frac{1}{\phi([d_1,d_2])} = \frac{\phi((d_1,d_2))}{\phi(d_1)\phi(d_2)} = \frac{(1*\mu*\phi)((d_1,d_2))}{\phi(d_1)\phi(d_2)} = \frac{1}{\phi(d_1)\phi(d_2)} \sum_{m|d_1,d_2} (\mu*\phi)(m).$$

This means that we can rewrite the sum in the main term of (4.11) as

$$\begin{aligned} Q := \sum_{d_1,d_2\in\mathcal{D}} \frac{\lambda_{d_1}\lambda_{d_2}}{\phi([d_1,d_2])} &= \sum_{d_1,d_2\in\mathcal{D}} \frac{\lambda_{d_1}\lambda_{d_2}}{\phi(d_1)\phi(d_2)} \sum_{m|d_1,d_2} (\mu*\phi)(m) \\ &= \sum_{m\in\mathcal{D}} (\mu*\phi)(m) \sum_{a_1,a_2\in\mathcal{D}} \frac{\lambda_{ma_1}\lambda_{ma_2}}{\phi(ma_1)\phi(ma_2)} = \sum_{m\in\mathcal{D}} \frac{(\mu*\phi)(m)}{\phi(m)^2} \left( \sum_{a\in\mathcal{D}} \frac{\lambda_{ma}}{\phi(a)} \right)^2. \end{aligned} \tag{4.12}$$

By making the change of variables

$$\xi_m = \sum_{a\in\mathcal{D}} \frac{\lambda_{ma}}{\phi(a)},$$

we diagonalize $Q$:

$$Q = \sum_{m\in\mathcal{D}} \frac{(\mu*\phi)(m)\xi_m^2}{\phi(m)^2}.$$

Note that, as it was for $\lambda_d$, the sequence $\xi_m$ is also supported on $\mathcal{D}$. The change of variables is invertible since, by Möbius inversion, we have

$$\lambda_n = \sum_{m\in\mathcal{D}} \frac{\lambda_{mn}}{\phi(m)} \sum_{d|m} \mu(d) = \sum_{d\in\mathcal{D}} \frac{\mu(d)}{\phi(d)} \sum_{a\in\mathcal{D}} \frac{\lambda_{dna}}{\phi(a)} = \sum_{d\in\mathcal{D}} \frac{\mu(d)\xi_{dn}}{\phi(d)},$$

where we used the fact that if $m$ is squarefree and $m = da$, then $(a,d) = 1$, and implies that $\phi(m) = \phi(a)\phi(d)$. We can use the Cauchy-Schwarz inequality to minimize $Q$ with the condition $\lambda_1 = 1$:

$$1 = \lambda_1 = \sum_{m\in\mathcal{D}} \mu(m)\sqrt{g(m)} \cdot \frac{\xi_m}{\sqrt{g(m)}\phi(m)} \leqslant \sqrt{L\cdot Q} \implies Q \geqslant \frac{1}{L},$$

where $g$ is the multiplicative function supported at odd squarefree integers such that $g(p) := \frac{1}{p-2}$ at every odd prime and

$$L := \sum_{m\in\mathcal{D}} g(m)$$

In addition, we obtain the equality $Q = \frac{1}{L}$ if we chose our $\lambda_d$ such that the ratio

$$\frac{\mu(m)\xi_m}{g(m)\phi(m)} = k \tag{4.13}$$

over every $m \in \mathcal{D}$ where $k$ is a constant. To satisfy the condition $\lambda_1 = 1$, we take $k = \frac{1}{L}$, since we will then have

$$\lambda_1 = \sum_{d\in\mathcal{D}} \frac{\mu(d)\xi_d}{\phi(d)} = k\sum_{d\in\mathcal{D}} g(d) = kL = 1.$$

To obtain a formula for $\lambda_n$ with $n \in \mathcal{D}$, we have to take into account that $\xi_{dn} \neq 0$ implies that $dn \in \mathcal{D}$:

$$\lambda_n = \sum_{d\in\mathcal{D}} \frac{\mu(d)\xi_{dn}}{\phi(d)} = \frac{\mu(n)g(n)\phi(n)}{L}\sum_{d\colon dn\in\mathcal{D}} g(d),$$

and since $g(p)\phi(p) = 1 + \frac{1}{p-2} = g(1) + g(p) = (1 * g)(p)$ for odd primes, then we can write

$$\lambda_n = \frac{\mu(n)}{L}\sum_{\delta|n} g(\delta)\sum_{d\colon dn\in\mathcal{D}} g(d) = \frac{\mu(n)}{L}\sum_{\delta|n}\sum_{d\colon dn\in\mathcal{D}} g(\delta d). \tag{4.14}$$

Since $n \in \mathcal{D}$, the product $(\delta, d) \mapsto \delta d$ is a one-to-one correspondence from $\{\delta : \delta \mid n\} \times \{d : dn \in \mathcal{D}\}$ to $\{a \in \mathcal{D} : [a, n] \leqslant y\}$. Thus we can write

$$\lambda_n = \mu(n)\left(1 - \frac{1}{L}\sum_{\substack{a\in\mathcal{D}\\ [a,n]>y}} g(a)\right) \quad \text{for } n \in \mathcal{D}.$$

Coming back to (4.3) and (4.11), we get the upper bound

$$\pi_+(x) \leqslant \frac{\mathrm{Li}(x)}{L} + \sum_{d_1,d_2\in\mathcal{D}} \lambda_{d_1}\lambda_{d_2} r_{d_1,d_2} + O\left(\frac{y}{\log y}\right) \tag{4.15}$$

for $y \leqslant \sqrt{2x+1}$.

## 4.3. Estimating the sum $L$

We need to get a lower bound on $L$ in (4.15). By taking $D = y^2$, we can rewrite $L$ as

$$L = \sum_{m\leqslant y} g(m)$$

Since $g(p) \approx \frac{1}{p}$ at every odd prime, we can try to write every term as a Dirichlet convolution $\frac{1}{g(m)} = \sum_{ab=m} \frac{h(a)}{b}$ where $h$ is relatively small if $m$ is odd. To explicitly find $h$, we can use a Dirichlet series: For $\sigma > 0$, where every series and products below absolutely converge, we have

$$\begin{aligned}\sum_{m=1}^{\infty}\frac{g(m)}{m^s} &= \prod_{p\geqslant 3}\left(1+\frac{1}{(p-2)p^s}\right)\\ &= \prod_{p\geqslant 3}\left(1+\frac{2}{(p-2)p^{s+1}}-\frac{1}{(p-2)p^{2s+1}}\right)\left(1-\frac{1}{p^{s+1}}\right)^{-1}\\ &= \left(1-\frac{1}{2^{s+1}}\right)\prod_{p\geqslant 3}\left(1+\frac{2}{(p-2)p^{s+1}}-\frac{1}{(p-2)p^{2s+1}}\right)\prod_{p}\left(1-\frac{1}{p^{s+1}}\right)^{-1}\\ &= H(s)\zeta(s+1)\end{aligned}$$

where $H$ is the Dirichlet series of $h$, which is a multiplicative function supported on cubefree integers not divisible by 4 defined by $h(2)=-\frac{1}{2}$, $h(p)=\frac{2}{p(p-2)}$ and $h(p^2)=-\frac{1}{p(p-2)}$ whenever $p$ is an odd prime.

Looking at the Euler product, we see that $H(s)$ converges absolutely for $\sigma>-\frac{1}{2}$. Hence for such a $\sigma$, we have

$$\sum_{a>z}|h(a)| \ll z^{\sigma}\sum_{a>z}|h(a)|\,a^{-\sigma} \ll_{\sigma} z^{\sigma} \tag{4.16}$$

Thus for $\varepsilon>0$ we have

$$\sum_{a\leqslant z}h(a)=H(0)+O_{\varepsilon}\Big(z^{-\frac{1}{2}+\varepsilon}\Big)=\frac{1}{c_2}+O_{\varepsilon}\Big(z^{-\frac{1}{2}+\varepsilon}\Big). \tag{4.17}$$

Using Dirichlet's convolution, we can get the following estimate for $L$:

**Proposition 4.2.** *If $D=y^2$ in the definition of $\mathcal{D}$, then for $y\geqslant 1$ and for every $\varepsilon>0$, we have*

$$L=\frac{\log y+\gamma}{c_2}+H'(0)+O_{\varepsilon}\Big(y^{-\frac{1}{3}+\varepsilon}\Big)$$

*where $\gamma$ is the Euler-Mascheroni constant, and $H$ is the Dirichlet series defined above.*

***Proof.*** With the convolution that we found above,

$$L=\sum_{ab\leqslant y}\frac{h(a)}{b}=\sum_{a\leqslant y^{2/3}}h(a)\sum_{b\leqslant \frac{y}{a}}\frac{1}{b}+\sum_{b\leqslant y^{1/3}}\frac{1}{b}\sum_{y^{2/3}<a\leqslant\frac{y}{b}}h(a). \tag{4.18}$$

This decomposition was found using *Dirichlet's hyperbola method*. The functions $y^{1/3}$ and $y^{2/3}$ in (4.18) were chosen such that their product equals $y$ and would ultimately minimize the error term.

Using partial summation, we get

$$\sum_{a\leqslant y^{2/3}}h(a)\sum_{b\leqslant\frac{y}{a}}\frac{1}{b}=\sum_{a\leqslant y^{2/3}}h(a)\left(\log y-\log a+\gamma+O\Big(\frac{a}{y}\Big)\right).$$

Using the same type of bounds as in (4.16), we get the bounds

$$\sum_{a>z} |h(a)| \log a \ll_\varepsilon z^{-\frac{1}{2}+\varepsilon} \quad \text{and} \quad \sum_{a\leqslant z} a\,|h(a)| \ll_\varepsilon z^{\frac{1}{2}+\varepsilon}$$

for $\varepsilon > 0$. Thus we get

$$\sum_{a\leqslant y^{2/3}} h(a) \sum_{b\leqslant \frac{y}{a}} \frac{1}{b} = \frac{\log y + \gamma}{c_2} + H'(0) + O_\varepsilon\Big(y^{-\frac{1}{3}+\varepsilon}\Big)\,.$$

For the second double sum, we can use (4.16) again and get

$$\sum_{b\leqslant y^{1/3}} \frac{1}{b} \sum_{y^{2/3}<a\leqslant \frac{y}{b}} h(a) \ll \sum_{b\leqslant y^{1/3}} \frac{1}{b} \sum_{a>y^{2/3}} |h(a)| \ll_\varepsilon y^{-\frac{1}{3}+\varepsilon}.$$

This proves the proposition. □

## 4.4. The Bombieri-Vinogradov theorem

All that is left is to understand (4.15) is to choose $y$ properly to get the quantity

$$\mathcal{R} = \sum_{d_1,d_2\leqslant y} \lambda_{d_1}\lambda_{d_2} r_{d_1,d_2}$$

as small as needed. It is easier to sum over the possible values of $[d_1, d_2]$:

$$\mathcal{R} = \sum_{\substack{d\leqslant y^2 \\ d|P_3(y)}} R_d \sum_{\substack{d_1,d_2\leqslant y \\ [d_1,d_2]=d}} \lambda_{d_1}\lambda_{d_2} \tag{4.19}$$

where $R_d$ is the error term defined by the equation

$$\pi(x; d, -2^{-1}) = \frac{\mathrm{Li}(x)}{\phi(d)} + R_d.$$

The inner sum in the last equality of (4.19) has at most $3^{\omega(d)}$ terms where the *prime omega function* $\omega(d)$ counts the number of distinct prime factors of $d$. This is because for every prime factor $p$ of a squarefree number $d$, we can construct every pair of $d_1$ and $d_2$ such that $[d_1, d_2] = d$ by choosing either that $p \mid d_1$ but $p \nmid d_2$, that $p \nmid d_1$ but $p \mid d_2$ or that $p \mid d_1$ and $p \mid d_2$. Since $|\lambda_{d_1}\lambda_{d_2}| \leqslant 1$, we can conclude that

$$\mathcal{R} \ll \sum_{\substack{d\leqslant y^2 \\ d|P_3(y)}} |R_d|\, 3^{\omega(d)}. \tag{4.20}$$

This is why one could ask for a bound on the error term for $\pi(x; q, a)$ that is uniform for $q$ in an interval that depends on $x$.

If we assume GRH, then $R_d \ll \sqrt{x}\log x$ uniformly for every odd $d \leqslant x$ by 2.1. Consequently, by taking $y \leqslant \sqrt{x}$ we get

$$\mathcal{R} \ll \sqrt{x}\log x \sum_{\substack{d\leqslant y^2 \\ d|P_3(y)}} 3^{\omega(d)} \tag{4.21}$$

We know that $3^{\omega(d)} \leqslant \tau_3(d)$ where $\tau_3 = 1 * 1 * 1$ since both functions are multiplicative and the inequality obviously holds for prime powers. Then

$$\sum_{d\leqslant y^2} 3^{\omega(d)} \leqslant \sum_{d\leqslant y^2} \tau_3(d) = \sum_{abc\leqslant y^2} 1 \leqslant \sum_{a\leqslant y^2} \sum_{b\leqslant y^2} \sum_{c\leqslant \frac{y^2}{ab}} 1 \ll y^2 \log^2 y. \tag{4.22}$$

Hence, by taking $y = \frac{x^{1/4}}{\log^3 x}$ in (4.15) and using the estimate $L = \frac{\log y}{c_2} + O(1)$ that we found in 4.2, we obtain

$$\pi_+(x) \leqslant \frac{4c_2 \operatorname{Li}(x)}{\log x + O(\log\log x)} + O\left(\frac{x}{\log^3 x}\right) = \frac{4c_2 x}{\log^2 x}\left(1 + O\left(\frac{\log\log x}{\log x}\right)\right) \tag{4.23}$$

which exactly leads to the result for $\pi_+(x)$ in Theorem 4.1.

But to prove the theorem, we do not need to bring up an unsolved conjecture such as GRH. A naive approach would be to bound $\mathcal{R}$ by taking $y = \log^A x$ for some positive number $A$ and using the Siegel-Walfisz theorem, but it can only lead us to

$$\pi_+(x) \ll_A \frac{x}{\log\log x}.$$

We do not necessarily need a pointwise bound on $R_d$, simply a bound on its average. Enrico Bombieri and A. I. Vinogradov proved the following theorem, and it is sometimes referred to as "The Riemann Hypothesis on average". It gives an upper bound for $\sum_{d\leqslant Q} |R_d|$ for $Q$ depending on $x$.

**Theorem 4.3** (Bombieri-Vinogradov Theorem). *For $A \geqslant 1$ and $x \geqslant 2$, we have*

$$\sum_{q\leqslant Q} \max_{y\leqslant x} \max_{\substack{a\leqslant q \\ (a,q)=1}} \left|\pi(x;q,a) - \frac{\operatorname{Li}(x)}{\phi(q)}\right| \ll_A \frac{x}{\log^A x}$$

*where $Q = \frac{\sqrt{x}}{\log^{A+2} x}$.*

A proof of this theorem can be found in Chapter 26 of [16]. Using the Cauchy-Schwarz inequality on (4.20), we get

$$\mathcal{R}^2 \ll \sum_{d\leqslant y^2} |R_d| \sum_{\substack{d\leqslant y^2 \\ d|P_3(y)}} 9^{\omega(d)} |R_d| \tag{4.24}$$

For the second sum, we can use the trivial bound $R_d \ll \frac{x}{d}$ uniformly for $d \ll x$. The weight $9^{\omega(d)}$ can be bounded above by $\tau_9(d)$ and by using exactly the same method as in (4.22) to find that if $y \ll \sqrt{x}$, then

$$\sum_{\substack{d \leqslant y^2 \\ d \mid P_3(y)}} 9^{\omega(d)} |R_d| \ll x \sum_{d \leqslant y^2} \frac{9^{\omega(d)}}{d} \ll x \sum_{a_1 \dots a_9 \leqslant y^2} \frac{1}{a_1 \dots a_9} \ll x \left( \sum_{a \leqslant y^2} \frac{1}{a} \right)^9 \ll x \log^9 y.$$

Combining this result and the Bombieri-Vinogradov theorem with $A = 15$ on the first sum in (4.24), we get

$$\mathcal{R} \ll \frac{x \log^{9/2} y}{\log^{A/2} x} \ll \frac{x}{\log^3 x}$$

by choosing $y = x^{1/4}/(\log^{17/2} x)$. Thus by choosing this value for $y$ and inserting it in (4.15), we get exactly the same result as in (4.23), which proves Theorem 4.1 unconditionally for $\pi_+(x)$. We can apply the same procedure to prove the theorem for $\pi_-(x)$.

## 4.5. A different perspective

Instead of studying $\pi_+(x; y)$, we could have studied the function

$$\pi'_+(x; y) := \# \{ n \leqslant x : (n(2n+1), P(y)) = 1 \}.$$

As in (4.3), we can get the following upper bound

$$\pi_+(x) \leqslant \pi'_+(x; y) + O(y) \quad \text{where } y \leqslant \sqrt{2x+1}, \tag{4.25}$$

since for every Sophie Germain prime of the first kind $p$, we have either $p \leqslant y$ or $p(2p+1)$ without any prime factor less than $y$. We can then use Selberg's sieve method as in the previous sections. The best upper bound on $\pi'_+(x; y)$ we can do using the same framework as (4.10) is to replace the constant 4 by 8 in Theorem 4.1, which is not as good but still gives us a bound of the right order of magnitude. Sieve methods can be applied in different settings. Generally, a problem has to respect a set of axioms to consider using sieve methods for processing an estimate or a bound.

Let $\mathcal{A}$ be a finite set of integers and $\mathcal{P}$ be a finite set of primes. We define

$$S(\mathcal{A}, \mathcal{P}) := \# \{ n \in \mathcal{A} : n \text{ has no prime factors in } \mathcal{P} \}.$$

Let's also define $A_d := \# \{ n \in \mathcal{A} : d \mid n \}$ and $P := \prod_{p \in \mathcal{P}} p$. These are the axioms of sieve theory, which are the starting point for estimating $S(\mathcal{A}, \mathcal{P})$:

**Axiom 1.** *We can define a parameter $X$, a multiplicative function $f$ with $0 \leqslant f(p) < 1$ for every prime $p$ and a sequence $r_d$ such that*

$$A_d = f(d)X + r_d \quad \textit{for } d \mid P.$$

**Axiom 2.** *There exist constants $\kappa, k \geqslant 0$ and $0 < \varepsilon \leqslant 1$ such that $f(p) \leqslant 1 - \varepsilon$ and $f(p) \leqslant k/p$ for primes $p \in \mathcal{P}$, and we have*

$$\sum_{\substack{p \in \mathcal{P} \\ p \leqslant z}} f(p) \log p = \kappa \log z + O(1) \quad \textit{for } z \leqslant \max \mathcal{P}.$$

**Axiom 3.** *There exist constants $A > 0$, $m \in \mathbb{N}$ and $D \geqslant 1$ such that*

$$\sum_{\substack{d \leqslant D \\ d \mid P}} \tau_m(d) \, |r_d| \leqslant \frac{X}{\log^A X}.$$

Let us explain how these axioms represent sieving. We think of $\mathcal{A}$ as a set that grows whenever a variable $x$ grows, e.g., $\{n \leqslant x\}$ and $\{p + 2 : p \leqslant x\}$. The goal is to sieve out all the multiples of $\mathcal{P}$ from $\mathcal{A}$. Let's suppose that $\#\mathcal{A} \approx X = X(x)$ as $x \to \infty$, and that we can find a multiplicative function $f$ such that $A_d \approx f(d)X$ as $x \to \infty$. We then define the sequence of remainders $r_d$ by setting $r_d := A_d - f(d)X$. They satisfy Axiom 1 by definition. Let's also suppose that the $r_d$ satisfy Axiom 3 with fixed $A, m$ and $D$.

Instead of using Möbius inversion to find a formula for $\mathbf{1}_{(n,P)=1}$, we seek two sequences $\rho_d^-$ and $\rho_d^+$ such that $\left|\rho_d^-\right|, \left|\rho_d^+\right| \leqslant \tau_m(d)$ (we say that these sequences are *divisor bounded*), which are supported on $\{d \mid P : d \leqslant D\}$, and such that

$$\sum_{d \mid n} \rho_d^- \leqslant \mathbf{1}_{(n,P)=1} \leqslant \sum_{d \mid n} \rho_d^+. \tag{4.26}$$

Then we say that the sequence $\rho_d^+$ is an *upper bound sieve* of level $D$ for the set of primes $\mathcal{P}$, and we write $(\rho_d^+)_d \in \Lambda^+(D, \mathcal{P})$. We can similarly say that $\rho_d^-$ is a *lower bound sieve* of level $D$ for the set of primes $\mathcal{P}$, and we can write that $(\rho_d^-)_d \in \Lambda^-(D, \mathcal{P})$.

If we want to get an upper bound on $S(\mathcal{A}, \mathcal{P})$, then

$$S(\mathcal{A}, \mathcal{P}) = \sum_{n \in \mathcal{A}} \mathbf{1}_{(n,P)=1} \leqslant \sum_{d \mid P} \rho_d^+ A_d = X \sum_{d \mid P} \rho_d^+ f(d) + O\!\left(\frac{X}{\log^A X}\right),$$

and similarly if we want a lower bound

$$S(\mathcal{A}, \mathcal{P}) \geqslant X \sum_{d \mid P} \rho_d^- f(d) + O\!\left(\frac{X}{\log^A X}\right).$$

In practice, the sieves $\rho_d^-$ and $\rho_d^+$ should behave like $\mu(d)\mathbf{1}_{d \mid P}$ since from (4.26), we have

$$\sum_{d \mid n} \rho_d^- \leqslant \sum_{d \mid n} \mu(d) \mathbf{1}_{d \mid P} \leqslant \sum_{d \mid n} \rho_d^+.$$

The only difference is that we control the size of the support with the parameter $D$. Without this control, we may be handling too many terms, which would lead us to a poor bound on the error term, such as in (4.5). In this chapter, Selberg's sieve method exploited squares' nonnegativity to construct a family of upper bound sieves $\rho_d^+$. This family is noted by $\Lambda^2(D, \mathcal{P})$.

No matter how we choose $\mathcal{A}$ and $\mathcal{P}$, Axiom 1 is useless on its own since by simply fixing $X$ and $f$, there always exists a sequence of remainders $r_d$. Axiom 1 is not really an axiom per se, but it defines $X$, $f$ and $r_d$ for the other axioms to make sense.

Axiom 2 gives an average value to our function $f$ on primes. The function $f$ could be understood as the proportion of congruence classes mod $d$ which are sieved out. The parameter $\kappa$ is called the *sifting dimension.* It represents the average number of congruence classes that we are sieving for every prime modulus $q$.

Finally, Axiom 3 takes care of the remainders $r_d$. With sieve methods, we only need to bound remainders on average. The parameter $D$ of this axiom is called the *level of distribution.*

In our two problems about $\pi_+(x; y)$ and $\pi'_+(x; y)$, we can understand it as a sieve problem in Table 4.1, and we can use Selberg's sieve method with these parameters.

| **Problems** | | $\pi_+(x; y)$ | $\pi'_+(x; y)$ |
|---|---|---|---|
| **Defining** $S(\mathcal{A}, \mathcal{P})$ | $\mathcal{A}$ | $\{2p + 1 : p \leqslant x\}$ | $\{n(2n + 1) : n \leqslant x\}$ |
| | $\mathcal{P}$ | $\{3 \leqslant p \leqslant y\}$ | $\{p \leqslant y\}$ |
| **Axiom 1** | $X$ | $\mathrm{Li}(x)$ | $x$ |
| | $f(p)$ | $1/\phi(p) = 1/(p - 1)$ | $(2 - \mathbf{1}_{p=2})/p$ |
| **Axiom 2** | $\kappa$ | 1 | 2 |
| | $k$ | 3/2 | 2 |
| | $\varepsilon$ | 1/2 | 1/3 |
| **Axiom 3** | $A$ | 2 | 3 |
| | $m$ | 3 | 3 |
| | $D$ | $\sqrt{x}/\log^{17} x$ | $x/\log^8 x$ |

**Table 4.1.** Parameters of the sieve problems $\pi_+(x; y)$ and $\pi'_+(x; y)$.

**Remark 4.4.** The fact that every value in Table 4.1 makes the three axioms valid is trivial, except for how both values of $D$ satisfy Axiom 3. We chose $D = \sqrt{x}/\log^{17} x$ for the problem $\pi(x; y)$ for the same reason that we chose $y = x^{1/4}/\log^{17/2} x$ in the previous section. For the problem $\pi'(x; y)$, the $r_d$ are bounded above by the number of congruence classes satisfying $n(2n + 1) \equiv 0 \pmod{d}$, which is itself bounded above by $2^{\omega(d)}$. Hence, our choice $D =$

$x/\log^8 x$ was to satisfy

$$\sum_{\substack{d\leqslant D\\ d|P(y)}} \tau_3(d) r_d \ll \sum_{\substack{d\leqslant D\\ d|P(y)}} \tau_3(d) 2^{\omega(d)} = \sum_{\substack{d\leqslant D\\ d|P(y)}} 6^{\omega(d)} \ll \sum_{d\leqslant D} \tau_6(d) \ll D\log^5 D \ll \frac{x}{\log^3 x},$$

where we found the upper bound on the sum of $\tau_6(d)$ in the same fashion as in (4.22).

We can generalize the Selberg sieve from the previous sections with the following theorem.

**Theorem 4.5.** *For $\mathcal{A}$ and $\mathcal{P} \subset \{p \leqslant y\}$, if the Axioms 1, 2 and 3 hold with $A = \kappa + 1$, $m = 3$, $D \geqslant y^2$ and $\log X \geqslant \log y$, then*

$$S(\mathcal{A},\mathcal{P}) \leqslant \frac{X}{\log^\kappa y}\left(1 + O_{\kappa,k,\varepsilon}\left(\frac{1}{\log y}\right)\right)\Gamma(\kappa+1)\prod_{p\in\mathcal{P}}(1-f(p))\prod_{p\leqslant y}\left(1-\frac{1}{p}\right)^{-\kappa}.$$

A proof of this theorem is given in its full generality in [16] Theorem 21.2. Taking $y = \sqrt{D}$ in Theorem 4.5 and the parameters defined in Table 4.1, we can easily obtain

$$\pi_+(x;y) \leqslant \left(4c_2 + O\left(\frac{\log\log x}{\log x}\right)\right)\frac{x}{\log^2 x}$$

and

$$\pi'_+(x;y) \leqslant \left(8c_2 + O\left(\frac{\log\log x}{\log x}\right)\right)\frac{x}{\log^2 x}.$$

We can notice that the products, when put together, absolutely converge to $c_2$ with an error term $O_\kappa(1/\log y)$. We have $X/\log^\kappa y \sim 4x/\log^2 x$ as $x \to \infty$ for both sieve problems. The only difference comes from the $\Gamma(\kappa+1)$, which means that the sifting dimension is why we have a worse bound with $\pi'_+(x;y)$. We start with a bigger set and have to sieve more congruence classes making our process less precise.

## 4.6. Sums of reciprocals

A consequence of Theorem 4.1, which indicates the sparsity of Sophie Germain primes, comes around when looking at the sum over the reciprocals.

In calculus, the first test to determine whether a series converges or not is by looking at the limit of the main term. One learns that if $\sum a_n$ converges, then $a_n \to 0$ as $n \to \infty$. To avoid any student incorrectly using the converse statement as a test for convergence in their final exam, the teacher sometimes follows with the statement “This does not mean that $a_n \to 0$ implies that the series $\sum a_n$ is convergent.” The intent is to introduce the student to the rigorous syntax of mathematical statements and provoke a thought about the true meaning of logical implications.

The first counterexample to the converse is usually the harmonic series. One way to prove the divergence is to say that for $N \geqslant 1$

$$\sum_{n\leqslant N}\frac{1}{n}=\int_1^N\frac{\mathrm{d}t}{t}-\int_{1^-}^N\frac{\mathrm{d}\{t\}}{t}=\log N+1-\int_1^N\frac{\{t\}}{t^2}\,\mathrm{d}t=\log N+\gamma+O\Big(\frac{1}{N}\Big).$$

where $\gamma = 1-\int_1^\infty\frac{\{t\}}{t^2}\,\mathrm{d}t$ is the *Euler-Mascheroni constant*.

It is not necessary to take all the terms of the harmonic series to have divergence. Euler showed that even though the primes are rare in the integers[3], the series $\sum\frac{1}{p}$ also diverges. Since we have the Prime Number Theorem at our disposal, we can use partial summation to show this. If $R(x) := \pi(x)-\mathrm{Li}(x)$, which is $O\Big(xe^{-c\sqrt{\log x}}\Big)$ for an absolute positive constant $c$, then

$$\sum_{p\leqslant N}\frac{1}{p}=\int_{2^-}^N\frac{\mathrm{d}\pi(t)}{t}=\int_2^N\frac{\mathrm{d}t}{t\log t}+\int_{2^-}^N\frac{\mathrm{d}R(t)}{t}=\log\log N+A+O\Big(e^{-c'\sqrt{\log N}}\Big)$$

where $A$ is a constant and $c'$ is another absolute positive constant.

When Brun was studying twin primes, he arrived at the conclusion that the sum of reciprocals of twin primes converges by giving a slightly weaker upper bound than ours on $\pi_2(x)$. For any subset $A\subset\mathbb{N}$, if $\{n\leqslant x : n\in A\}\asymp\frac{x}{(\log x)^\alpha}$, then $\sum_{n\in A}\frac{1}{n}$ converges if, and only if, $\alpha>1$. This is shown in the proof of the following corollary:

**Corollary 4.6.** *The series $\sum\frac{1}{p}$ over the Sophie Germain primes of the first kind is convergent.*

***Proof.*** Using Theorem 4.1, we need to evaluate using partial summation

$$\sum_{\substack{p>N\\ 2p+1\text{ is prime}}}\frac{1}{p}=\int_N^\infty\frac{\mathrm{d}\pi_+(t)}{t}=-\frac{\pi_+(N)}{N}+\int_N^\infty\frac{\pi_+(t)}{t^2}\,\mathrm{d}t\ll\frac{1}{\log^2 N}+\int_N^\infty\frac{1}{t\log^2 t}\,\mathrm{d}t\ll\frac{1}{\log N}$$

and the corollary follows[4]. □

**Remark 4.7.** We can note that Conjecture 1.1 would imply that we cannot have a faster rate of convergence than

$$\sum_{\substack{p>N\\ 2p+1\text{ is prime}}}\frac{1}{p}\sim\frac{c_2}{\log N}.$$

[3] In the sense that $\frac{\pi(n)}{n}\to 0$ as $n\to\infty$.

[4] Another proof of Corollary 4.6 is by dyadic decomposition: Let $S_j := \sum\frac{1}{p}$, where the sum is over the Sophie Germain primes of the first kind in the interval $(2^{j-1},2^j]$. Since $S_j\ll\pi_+(2^j)/2^{j-1}\ll 1/j^2$ and $\sum_{j=1}^\infty 1/j^2$ converges, then $\sum_{j=1}^\infty S_j$ also converges and the corollary follows.

# Chapter 5

# Cramér's model refined

To make a conjecture about the size of $\pi_+$ and $\pi_-$, we go back to the Bernoulli variables $X_n$ of Cramér's model defined in Section 1.4. We mainly use Cramér's construction when we try to interpret the primes' indicator as just a generic outcome of the sequence $X_n$. In general, the law of large numbers would imply that we should expect a sum of random variables to be quite close to its expected value. To give a bound on the error term for any sum of uniformly essentially bounded random variables, we can use the following lemma:

**Lemma 5.1** (Hausdorff's estimate)**.** *Let $Y_n$ be a sequence of independent real random variables such that $\mathbb{E}[Y_n] = 0$ and $\|Y_n\|_\infty \leqslant 1$ for all $n \in \mathbb{N}$. Then, for $\varepsilon > 0$, we have*

$$\sum_{j \leqslant n} Y_j \ll_\varepsilon n^{\frac{1}{2}+\varepsilon}$$

*almost surely.*

***Proof.*** Let $S_n = \sum_{j \leqslant n} Y_j$. By taking the $2k^{\text{th}}$ moment of $S_n$ for some $k > \frac{1}{2\varepsilon}$, expanding the sum and taking into account that every $Y_i$ have mean zero, we can use the multinomial theorem to get that

$$\mathbb{E}[S_n^{2k}] = \sum_{\substack{j_1+\dots+j_n=2k \\ j_\ell \neq 1 \ \forall \ell}} \binom{2k}{j_1, \dots, j_n} \mathbb{E}[Y_1^{j_1}] \dots \mathbb{E}[Y_n^{j_n}] \ll_k \sum_{\substack{j_1+\dots+j_n=2k \\ j_\ell \neq 1 \ \forall \ell}} 1 \tag{5.1}$$

since the multinomial coefficients have $(2k)!$ as a uniform upper bound. For every $n \geqslant k$, if we want to write $2k$ as a sum of $n$ non-negative integers without any ones, we know that at least $n-k$ of these integers are zero. Thus, we can give a bound by first counting the number of ways to place $n-k$ zeros amongst the $n$ possible positions and then counting the number of ways to write $2k$ as a sum of the $k$ nonnegative integers for the remaining positions. Thus, we get the bound

$$\mathbb{E}[S_n^{2k}] \ll_k \binom{n}{k} \binom{3k-1}{k-1} \ll_k n^k.$$

Using Markov's inequality, we get that for $\alpha = \frac{1}{2} + \varepsilon$, we have

$$\mathbb{P}\left(|S_n/n^\alpha| \geqslant 1\right) = \mathbb{P}(S_n^{2k} \geqslant n^{2k\alpha}) \leqslant \frac{\mathbb{E}[S_n^{2k}]}{n^{2k\alpha}} \ll_k \frac{1}{n^{k(2\alpha-1)}}.$$

Thus with the first Borel-Cantelli lemma, we get that $|S_n/n^\alpha| < 1$ for $n$ big enough almost surely because $\sum_{n=1}^{\infty} \frac{1}{n^{k(2\alpha-1)}}$ is convergent and this completes the proof. □

To model the behaviour of $\pi_+$, we need to think about the indicator of the primes as a specific outcome $\omega$ of the probability space generated by Cramér's model to say that $X_n(\omega) = \mathbf{1}_{n \text{ is prime}}$ and that

$$\pi_+(x) = \sum_{n \leqslant x} \mathbf{1}_{n \text{ is prime}} \mathbf{1}_{2n+1 \text{ is prime}} = \sum_{n \leqslant x} X_n(\omega) X_{2n+1}(\omega).$$

Let $Y_n = X_n X_{2n+1} - \frac{1}{\log n \log(2n+1)}$ for $n \geqslant 3$ and $\alpha = 1/2 + \varepsilon$ for any $\varepsilon > 0$ in Lemma 5.1. We then obtain almost surely that

$$\begin{aligned} \sum_{n \leqslant x} X_n X_{2n+1} &= \sum_{3 \leqslant n \leqslant x} \frac{1}{\log n \log(2n+1)} + O\Big(x^{\frac{1}{2}+\varepsilon}\Big) \\ &= \sum_{3 \leqslant n \leqslant x} \left( \frac{1}{\log n \log 2n} + O\bigg(\frac{1}{n \log^3 n}\bigg) \right) + O\Big(x^{\frac{1}{2}+\varepsilon}\Big) \\ &= \mathcal{L}(x) + O\Big(x^{\frac{1}{2}+\varepsilon}\Big) \end{aligned} \tag{5.2}$$

where

$$\mathcal{L}(x) = \int_2^x \frac{\mathrm{d}t}{\log t \log 2t}$$

and the last equality is obtained by partial summation. With l'Hôpital's rule used similarly as in (1.5), we arrive to the estimate $\mathcal{L}(x) \sim \frac{x}{\log^2 x}$.

If we assume that $\omega$ is in the event $\{\omega : \text{Equation (5.2) is true}\}$ (which is of measure 1), then it would lead to the conjecture that $\pi_+(x) \sim \mathcal{L}(x) \sim \frac{x}{\log^2 x}$.

This is great news since $\frac{x}{\log^2 x}$ is, up to a constant, the upper bound obtained on $\pi_+(x)$ in the previous chapter using Selberg's sieve method (see Theorem 4.1). However, looking at Figure 5.1, the ratio $\frac{\pi_+(x)}{\mathcal{L}(x)}$ seems to converge to $c_2$ instead of 1 as $x \to \infty$.

## 5.1. The Cramér-Granville model

As we discussed in the last paragraph of Section 1.4, the Cramér model can lead us to absurd contradictions. In 1985, Maier found that if $y = \log^A x$ for $A > 1$, then

$$\limsup_{x \to \infty} \frac{\pi(x+y) - \pi(x)}{y / \log x} > 1 \quad \text{and} \quad \liminf_{x \to \infty} \frac{\pi(x+y) - \pi(x)}{y / \log x} < 1. \tag{5.3}$$

This is surprising because one would not reach the same conclusion by making a prediction using Cramér's model. Maier found (5.3) by first sieving out the integers with a small prime factor and then using density arguments. Granville used this idea to have a more accurate

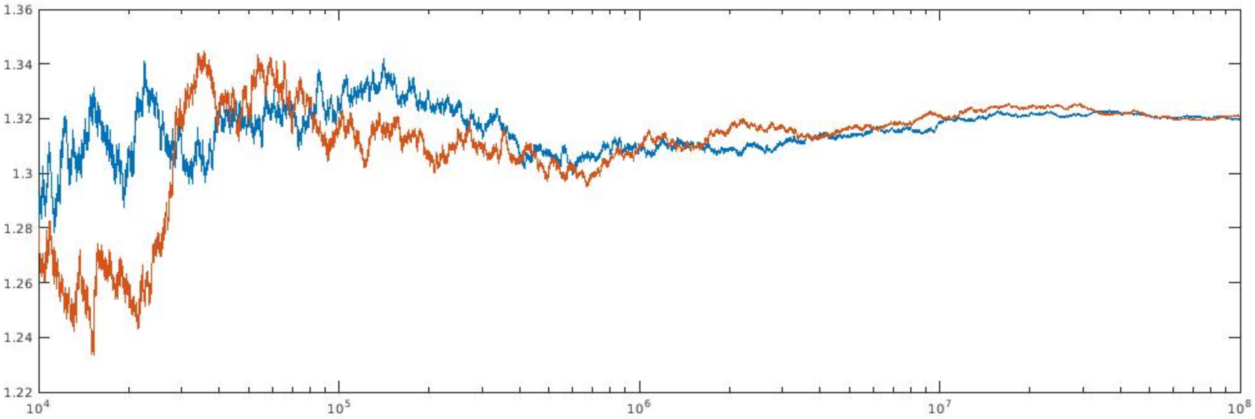

**Fig. 5.1.** Graph of $\pi_+(x)/\mathcal{L}(x)$ (in blue) and of $\pi_-(x)/\mathcal{L}(x)$ (in orange).

way to predict the behaviour of primes. He started by first sieving the primes less than a parameter $y$, before applying Cramér's ideas and put a probability measure proportional to $1/\log n$ on the $y$-rough integers.

When removing multiples of small primes to adjust Cramér's model, we must have a more accurate description of the primes. Let $\mathcal{P}$ be a set of primes, $P$ the products of all primes in $\mathcal{P}$ and $\alpha$ which only depends on $\mathcal{P}$. We define a sequence of independent Bernoulli random variables $Z_n$ with

$$\mathbb{P}(Z_n = 1) = \frac{\alpha}{\log n}\mathbf{1}_{(n,P)=1}$$

for $n$ large enough such that this probability is $\leqslant 1$. Then to model the primes, we would need to respect the prime number theorem, which means that

$$\mathbb{E}\left[\sum_{n\leqslant x} Z_n\right] = \alpha \sum_{\substack{n\leqslant x\\(n,P)=1}} \frac{1}{\log n} \sim \alpha \prod_{p\in\mathcal{P}}\left(1-\frac{1}{p}\right)\mathrm{Li}(x)$$

as long as $\max\mathcal{P} = \frac{\log x}{\log 2} - \xi(x)$ where $\xi(x) \to +\infty$ as $x \to \infty$. We prove this by using partial summation, coupled with the fact that

$$\sum_{\substack{n\leqslant x\\(n,P)=1}} 1 = \sum_{n\leqslant x}\sum_{\substack{d|n\\d|P}}\mu(d) = \sum_{d|P}\mu(d)\sum_{\substack{n\leqslant x\\d|n}} 1 = \sum_{d|P}\mu(d)\left[\frac{x}{d}\right]$$
$$= x\sum_{d|P}\frac{\mu(d)}{d} + O\Big(2^{\max\mathcal{P}}\Big) \sim x\prod_{p\in\mathcal{P}}\left(1-\frac{1}{p}\right)$$

as $x \to \infty$. This means it is convenient to choose

$$\alpha = \prod_{p\in\mathcal{P}}\left(1-\frac{1}{p}\right)^{-1}.$$

This is the idea of *presieving* a set of primes. Franz Mertens proved that

$$\prod_{p\leqslant y}\left(1-\frac{1}{p}\right)=\frac{e^{-\gamma}}{\log y}\left(1+O\left(\frac{1}{\log y}\right)\right) \tag{5.4}$$

where $\gamma$ is the Euler-Mascheroni constant. This means that we can define a probability measure for $n>(\max\mathcal{P})^2$. Every $Z_n$ with $n\leqslant(\max\mathcal{P})^2$ can be almost surely the primes' indicator function.

***Example* 5.2.** If we want to model the primes $a$ mod $q$ for $(a,q)=1$ and give a conjecture for the Prime Number Theorem for arithmetic progressions mod $q$ for a fixed $q$, then we have to take into account that those primes have to be coprime to $q$; thus it would be useful to presieve prime factors of $q$. Then we would expect $\pi(x;q,a)$ to be

$$\sum_{\substack{n\leqslant x\\ n\equiv a \pmod q}}\mathbb{E}[Z_n]=\prod_{p\mid q}\left(1-\frac{1}{p}\right)^{-1}\sum_{\substack{q^2<n\leqslant x\\ n\equiv a \pmod q}}\frac{1}{\log n}+O_q(1)\sim\frac{\mathrm{Li}(x)}{\phi(q)}$$

for fixed $q$ as $x\to\infty$.

Often, presieving is done with a set of primes of the form $\{p\leqslant y\}$ for a parameter $y$. For example, if we study the set of Sophie Germain primes of the first kind, then we expect $\pi_+(x)$ to be

$$\sum_{n\leqslant x}\mathbb{E}[Z_nZ_{2n+1}]=\sum_{\substack{2\leqslant n\leqslant x\\ (n(2n+1),P(y))=1}}\frac{\alpha^2}{\log n\log(2n+1)}+O\left(\frac{y^2}{\log^2 y}\right). \tag{5.5}$$

In order to estimate the sum in (5.5), we need to understand the function $\pi'(x;y)$ from Section 4.5. Using partial summation and Proposition 5.3, which we will prove in the next section, we can get

$$\begin{aligned}&=2\prod_{3\leqslant p\leqslant y}\left(1-\frac{2}{p}\right)\left(1-\frac{1}{p}\right)^{-2}\mathcal{L}(x)+O\left(\frac{x}{\log^2 x\log y}+\frac{y^2}{\log^2 y}\right)\\ &=\left(c_2+O\left(\frac{\log\log x}{\log x}\right)\right)\frac{x}{\log^2 x}.\end{aligned} \tag{5.6}$$

if we take $y=x^{1/8\log\log x}$. We would get the same result if we tried to model $\pi_-(x)$. This agrees with Conjecture 1.1.

## 5.2. Brun's pure sieve

For a finite set of integers $\mathcal{A}$ and $\mathcal{P}=\{p\leqslant y\}$ for some parameter $y$, we wish to prove an estimate on the value $S(\mathcal{A},\mathcal{P})$ defined in Section 4.5. We will present Brun's ideas, who proved that the series of reciprocals of twin primes converges. His method differs from the Selberg sieve by providing an upper and lower bound; however, the upper bound will not be as good as the one given in the previous chapter. This will give us an estimate on $\pi'_+(x;y) = \#\{n \leqslant x : (n(2n+1), P(y)) = 1\}$.

Brun's idea to get an upper and lower bound sieve is to exploit the fact that the Möbius inversion is an inclusion-exclusion process. Let $\mathcal{A}_d = \{n \in \mathcal{A} : d \mid n\}$ and $A_d$ be its size. Since

$$S(\mathcal{A},\mathcal{P}) = \#\left(\mathcal{A} \setminus \bigcup_{p\leqslant y} \mathcal{A}_p\right) = A_1 - \#\bigcup_{p\leqslant y} \mathcal{A}_p$$

and $\mathcal{A}_d$ is multiplicative in $d$ in the sense that $\mathcal{A}_{d_1} \cap \mathcal{A}_{d_2} = \mathcal{A}_{d_1 d_2}$ for coprime $d_1$ and $d_2$, then we can use Bonferroni's inequalities to stop the inclusion-exlusion process after $k$ steps and get

$$\begin{aligned} S(\mathcal{A},\mathcal{P}) \leqslant A_1 + \sum_{j=1}^{k} (-1)^j \sum_{p_1<\ldots<p_j\leqslant y} A_{p_1\ldots p_j} &= \sum_{n\in\mathcal{A}} \left(1 + \sum_{j=1}^{k}(-1)^j \sum_{p_1<\ldots<p_j\leqslant y} \mathbf{1}_{p_1\ldots p_j \mid n}\right) \\ &= \sum_{n\in\mathcal{A}} \sum_{\substack{d\mid P(y)\\ d\mid n \\ \omega(d)\leqslant k}} \mu(d) = \sum_{\substack{d\mid P(y)\\ \omega(d)\leqslant k}} \mu(d) A_d \end{aligned}$$

if $k$ is even, and we get the reversed inequality if $k$ is odd. We have an upper and lower bound sieve with

$$\rho_d^+ = \mu(d)\mathbf{1}_{\substack{d\mid P(y)\\ \omega(d)\leqslant k}} \quad \text{if } k \text{ is even and} \quad \rho_d^- = \mu(d)\mathbf{1}_{\substack{d\mid P(y)\\ \omega(d)\leqslant k}} \quad \text{if } k \text{ is odd.}$$

Thus, if we define $X, f$ and $r_d$ such that our problem satisfies Axiom 1, we have the estimate

$$\begin{aligned} S(\mathcal{A},\mathcal{P}) &= \sum_{\substack{d\mid P(y)\\ \omega(d)<k}} \mu(d) A_d + O\left(\sum_{\substack{d\mid P(y)\\ \omega(d)=k}} A_d\right) \\ &= X \prod_{p\leqslant y} (1 - f(p)) + O\left(\sum_{\substack{d\mid P(y)\\ \omega(d)\leqslant k}} |r_d| + X \sum_{\substack{d\mid P(y)\\ \omega(d)=k}} f(d)\right) \qquad (5.7) \end{aligned}$$

as long as we keep $k \leqslant \pi(y)$. When trying to study $\pi'_+(x;y)$ with the parameters from Table 4.1, we get Axiom 1 with

$$A_d = \#\{n \leqslant x : n(2n+1) \equiv 0 \pmod{d}\} = x\frac{\nu(d)}{d} + O(\nu(d)) \quad \text{where } \nu(p) = \begin{cases} 1 & \text{if } p = 2 \\ 2 & \text{if } p \geqslant 3 \end{cases}$$

and $\nu$ is multiplicative. With the multiplicity, we can get the bound $\nu(d) \ll 2^{\omega(d)}$. This function $\nu(d)$ represents the number of roots of the polynomial $n(2n+1)$ in $\mathbb{Z}/d\mathbb{Z}$. Putting this in (5.7), we get

$$\pi'_+(x;y) = \frac{x}{2}\prod_{3\leqslant p\leqslant y}\left(1-\frac{2}{p}\right) + O\left(2^k\left(\sum_{\substack{d|P(y)\\ \omega(d)\leqslant k}} 1 + x\sum_{\substack{d|P(y)\\ \omega(d)=k}}\frac{1}{d}\right)\right).$$

The main term is $\frac{x}{\log^2 y}$. For the sums in the remainder, we have

$$2^k\sum_{\substack{d|P(y)\\ \omega(d)=k}}\frac{1}{d} = 2^k\sum_{p_1<\ldots<p_k\leqslant y}\frac{1}{p_1\ldots p_k} \leqslant \frac{2^k}{k!}\left(\sum_{p\leqslant y}\frac{1}{p}\right)^k.$$

This last expression looks like $e^{\lambda}\mathbb{P}(Z=k)$ where $Z$ is a Poisson random variable of mean $\lambda = \sum_{p\leqslant y} 2/p \approx 2\log\log y$. Poisson random variables have a standard deviation of $\sqrt{\lambda}$ and their distribution resembles the normal distribution as $\lambda \to \infty$. This means that taking $k$ as a large multiple of $\log\log y$ would give us an error term really small. If $k = [7.94\log\log y]$, we would have

$$\ll \left(\frac{2e}{k}\sum_{p\leqslant y}\frac{1}{p}\right)^k \ll \left(\frac{2e\sum_{p\leqslant y}\frac{1}{p}}{7.94\log\log y}\right)^k \ll (0.685)^{7.94\log\log y} = (\log y)^{7.94\log 0.685} \ll \frac{1}{\log^3 y}$$

where we used $n! \gg (n/e)^n$ from Stirling's formula.

The other sum in the remainder can be seen as the number of products of at most $k$ distinct primes less than or equal to $y$; this has to be less than or equal to $y^k$. If $3 \leqslant y \leqslant x^{\frac{1}{8\log\log x}}$, then

$$(2y)^k\log^3 y \ll x^{\frac{7.94}{8}}\log^3 x \ll x \implies (2y)^k \ll \frac{x}{\log^3 y}.$$

We get the following estimate for $\pi'(x;y)$.

**Proposition 5.3.** *For $3 \leqslant y \leqslant x^{1/8\log\log x}$, we have*

$$\pi'_+(x;y) = \frac{x}{2}\prod_{3\leqslant p\leqslant y}\left(1-\frac{2}{p}\right)\left(1+O\left(\frac{1}{\log y}\right)\right).$$

We get an estimate on the distribution of objects which are the basis of the Cramér-Granville model.

## 5.3. Variance of the Sophie Germain prime race

The indicator of the Sophie Germain primes of the first kind is $\mathbf{1}_{n\text{ is prime}}\mathbf{1}_{2n+1\text{ is prime}}$ and the indicator of the Sophie Germain primes of the second kind is $\mathbf{1}_{n\text{ is prime}}\mathbf{1}_{2n-1\text{ is prime}}$. We can write the difference $\pi_+(x) - \pi_-(x)$ as sum of indicator functions:

$$\begin{aligned}\pi_+(x) - \pi_-(x) &= \left(\sum_{n\leqslant x} \mathbf{1}_{n \text{ is prime}} \mathbf{1}_{2n+1 \text{ is prime}}\right) - \left(\sum_{n\leqslant x} \mathbf{1}_{n \text{ is prime}} \mathbf{1}_{2n-1 \text{ is prime}}\right) \\ &= \sum_{n\leqslant x} \mathbf{1}_{n \text{ is prime}} (\mathbf{1}_{2n+1 \text{ is prime}} - \mathbf{1}_{2n-1 \text{ is prime}}).\end{aligned}$$

Thus we can try modelling the race with

$$M_x = \sum_{n\leqslant x} Z_n (Z_{2n+1} - Z_{2n-1}),$$

where the $Z_n$ are the Bernoulli random variables of the Cramér-Granville model at the end of Section 5.1 with the same parameter $y = x^{1/8 \log\log x}$.

To understand how far from the mean we should expect the prime race to be, we can calculate the variance. In the sum, the only of terms after $n$ which are dependant with the $n^{\text{th}}$ term are the terms $n+1$, $2n-1$ and $2n+1$. By using the formula for the variance of a sum, we get

$$\begin{aligned}\operatorname{Var}(M_x) = \sum_{n\leqslant x} [&\operatorname{Var}(Z_n Z_{2n+1}) + \operatorname{Var}(Z_n Z_{2n-1}) - 2\operatorname{Cov}(Z_n Z_{2n+1}, Z_n Z_{2n-1}) \\ &- 2\operatorname{Cov}(Z_n Z_{2n+1}, Z_{n+1} Z_{2n+1}) - 2\operatorname{Cov}(Z_n Z_{2n-1}, Z_{2n-1} Z_{4n-1}) \\ &+ 2\operatorname{Cov}(Z_n Z_{2n-1}, Z_{2n-1} Z_{4n-3}) + 2\operatorname{Cov}(Z_n Z_{2n+1}, Z_{2n+1} Z_{4n+3}) \\ &- 2\operatorname{Cov}(Z_n Z_{2n+1}, Z_{2n+1} Z_{4n+1})].\end{aligned} \tag{5.8}$$

To evaluate the covariances, we can note that if $W_1, W_2, W_3$ are three independent Bernoulli variables of parameters $p_1, p_2, p_3$ repectively, then

$$\operatorname{Cov}(W_1 W_2, W_1 W_3) = \mathbb{E}[W_1 W_2 W_3] - \mathbb{E}[W_1 W_2]\mathbb{E}[W_1 W_3] = p_1 p_2 p_3 (1 - p_1). \tag{5.9}$$

Since $n(n+1)(2n+1)$ is always even and $n(2n+1)(2n-1), n(2n-1)(4n-1)$ and $n(2n+1)(4n+1)$ are always multiples of 3, then from the six covariances above, we only need to keep the fourth and the fifth because the variables in the other covariances will be uncorrelated whenever we sieve out the multiples of 2 and 3.

The polynomials $n(2n-1)(4n-3)$ and $n(2n+1)(4n+3)$ have one solution in $\mathbb{Z}/2\mathbb{Z}$, two solutions in $\mathbb{Z}/3\mathbb{Z}$ and three solutions in $\mathbb{Z}/p\mathbb{Z}$ for $p \geqslant 5$. When applying (5.9) on the fourth and fifth covariances in (5.8), then we get that the contributions of the covariances in the sum is of the size $18\prod_{p\geqslant 5}(1 - \frac{3p-1}{(p-1)^3})\frac{x}{\log^3 x}$. Since the product of two independent Bernoulli variables is a Bernoulli variable itself, then we obtain

$$\begin{aligned}\operatorname{Var}(M_x) &= \sum_{n\leqslant x} \operatorname{Var}(Z_n Z_{2n+1}) + \operatorname{Var}(Z_n Z_{2n-1}) + O\biggl(\frac{x}{\log^3 x}\biggr)\\ &= \sum_{n\leqslant x} \mathbb{E}[Z_n Z_{2n+1}] + \mathbb{E}[Z_n Z_{2n-1}] + O\Biggl(\frac{x}{\log^3 x} + \sum_{n\leqslant x}\Bigl(\mathbb{E}[Z_n Z_{2n+1}]^2 + \mathbb{E}[Z_n Z_{2n-1}]^2\Bigr)\Biggr)\\ &= \frac{2c_2 x}{\log^2 x} + O\Biggl(\frac{x}{\log^3 x} + \sum_{n\leqslant x}\Bigl(\mathbb{E}[Z_n Z_{2n+1}]^2 + \mathbb{E}[Z_n Z_{2n-1}]^2\Bigr)\Biggr).\end{aligned}$$

To give an upper bound on the sum in the error term, we have

$$\sum_{n\leqslant x} \mathbb{E}[Z_n Z_{2n+1}]^2 \asymp \sum_{\substack{n\leqslant x\\ (n(2n+1),P(y))=1}} \frac{\alpha^4}{\log^4 n} \ll \frac{x}{\log^4 x} \prod_{3\leqslant p\leqslant y}\left(1-\frac{2}{p}\right)\left(1-\frac{1}{p}\right)^{-4} \ll \frac{x\log^2 y}{\log^4 x}$$

with Mertens' estimate (5.4). Since we previously set the parameter $y = x^{1/8\log\log x}$ to our Cramér-Granville model, then

$$\operatorname{Var}(M_x) = \left(2c_2 + O\biggl(\frac{1}{(\log\log x)^2}\biggr)\right)\frac{x}{\log^2 x}. \tag{5.10}$$

We can standardize $M_x$ by dividing it by an approximation of the standard deviation. Thus we can have a clearer image of the race by studying instead the quantity $\frac{\log x}{\sqrt{x}}(\pi_+(x) - \pi_-(x))$ as we will see ahead in Figure 7.1.

## 5.4. Twin primes vs. Sophie Germain primes

To pass from (5.5) to (5.6), we had to use partial summation and the estimate for $t \geqslant 2$

$$\frac{1}{\log t\log(2t+1)} = \frac{1}{\log t\log(2t)} + O\biggl(\frac{1}{t\log^3 t}\biggr) \implies \int_2^x \frac{1}{\log t\log(2t+1)} = \mathcal{L}(x) + O(1)$$

to ultimately get the conjecture $\pi_+(x) = c_2\mathcal{L}(x)$. Since the logarithmic function is *slowly oscillating*, meaning that for every positive real number $k$ we have $\log(kt) \sim \log t$ as $t \to \infty$, why not replace the $\log(2t)$ by $\log t$ in the denominator of the integrand. The reason is that this will get the error term too large. Indeed, if $\mathrm{Li}_2(x) = \int_2^x \frac{\mathrm{d}t}{\log^2 t}$, then $(\mathrm{Li}_2(x) - \mathcal{L}(x)) = \int_2^x \frac{\log 2\ \mathrm{d}t}{\log^2 t\log(2t)} \sim \frac{x\log 2}{\log^3 x}$, which is very big compared to the standard deviation of the Sophie Germain race.

As a matter of fact, the twin prime constant $c_2$ does not have this appellation for no reason. Using the Cramér-Granville model, we get that $\sum_{n\leqslant x} \mathbb{E}[Z_n Z_{n+2}] = c_2\,\mathrm{Li}_2(x) + O\Bigl(x^{1/4\log\log x}\Bigr)$ which means that the twin prime counting function $\pi_2(x)$ is asymptotic to $\frac{c_2 x}{\log^2 x}$, as is $\pi_+(x)$. However, if we would race the two functions against each other, then we could conjecture that

$$\frac{\log^3 x}{x}(\pi_2(x)-\pi_+(x)) \approx \frac{\log^3 x}{x}\left(\sum_{n\leqslant x}\mathbb{E}[Z_n(Z_{n+2}-Z_{2n+1})]\right) \sim \frac{c_2\log^3 x}{x}(\mathrm{Li}_2(x)-\mathcal{L}(x))$$

which would mean that for $x$ sufficiently large, there is always more twin primes than Sophie Germain primes less than $x$ since

$$\lim_{x\to\infty}\frac{\log^3 x}{x}(\pi_2(x)-\pi_+(x)) = c_2\log 2 \approx 0.9151786\ldots$$

as seen in Figure 5.2.

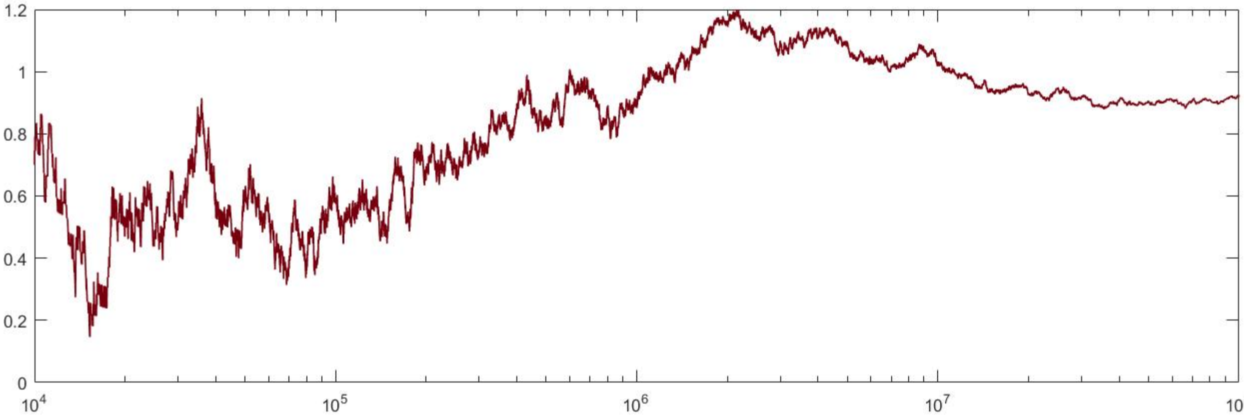

**Fig. 5.2.** Graph of $(\pi_2(x)-\pi_+(x))/\int_2^x \frac{\mathrm{d}t}{\log^2 t\log(2t)}$.

# Chapter 6

# The circle method

To study the number of Sophie Germain primes of the first kind, it suffices to estimate

$$\psi_+(x) := \sum_{n \leqslant x} \Lambda(n)\Lambda(2n+1).$$

The proper prime powers contribute to a negligible amount to this sum, and the logarithmic weight can be easily removed to have an estimate for $\pi_+(x)$. This can be compared to the techniques we used in Section 2.7 to transform our estimate for $\psi(x; q, a)$ into an estimate for $\pi(x; q, a)$. Similarly, for Sophie Germain primes of the second kind, we need to estimate the quantity

$$\psi_-(x) := \sum_{n \leqslant x} \Lambda(n)\Lambda(2n-1).$$

Since $\int_0^1 e(k\alpha)\,\mathrm{d}\alpha = \mathbf{1}_{k=0}$, for $x \geqslant 2$ we can rewrite $\psi_+(x)$ as

$$\begin{aligned}\psi_+(x) = \sum_{\substack{n \leqslant x \\ m \leqslant 2x+1 \\ m=2n+1}} \Lambda(n)\Lambda(m) = \sum_{\substack{n \leqslant x \\ m \leqslant 2x+1}} \Lambda(n)\Lambda(m) \int_0^1 e((2n-m+1)\alpha)\,\mathrm{d}\alpha \\ = \int_0^1 S_x(2\alpha) S_{2x+1}(-\alpha) e(\alpha)\,\mathrm{d}\alpha, \quad (6.1)\end{aligned}$$

where

$$S_x(\alpha) := \sum_{n \leqslant x} \Lambda(n) e(n\alpha).$$

Similarly,

$$\psi_-(x) = \int_0^1 S_x(2\alpha) S_{2x-1}(-\alpha) e(-\alpha)\,\mathrm{d}\alpha.$$

In 1918, G. H. Hardy and Srinivasa Ramanujan developed a method to study the partition function that counts the number of ways to write a number as a sum of positive integers. In the 1920s, Hardy and Littlewood proved multiple additive number theory problems by transforming the problem at hand into evaluating an integral of Fourier series over the unit circle[1], similar to the one given in (6.1). They noticed that the Fourier series $S_x(\alpha)$ is larger for $\alpha$ near rational numbers with low denominators. Such values ought to give the dominant contribution of the integral in (6.1).

Let us explain their idea in detail. Let $P$ and $Q$ be positive real parameters such that[2] $Q \geqslant 2P^2$. For $q \leqslant P$, $1 \leqslant a \leqslant q$ and $(a,q)=1$, we define $\mathfrak{M}(q,a)$ as the set of $\alpha$ such that $\|\alpha - a/q\| \leqslant 1/Q$, where $\|\alpha\|$ is the distance between $\alpha$ and the closest integer[3]. We can define the *major arcs* as

$$\mathfrak{M} = \bigcup_{q \leqslant P} \bigcup_{\substack{1 \leqslant a \leqslant q \\ (a,q)=1}} \mathfrak{M}(q,a)$$

and the *minor arcs* as

$$\mathfrak{m} = [0,1] \setminus \mathfrak{M}.$$

In the next section, we will prove that

**Theorem 6.1.** *For $x \to \infty$ and $A > 1$, we have*

$$\int_{\mathfrak{M}} S_x(2\alpha) S_{2x+1}(-\alpha) e(\alpha) \mathrm{d}\alpha = c_2 x + O_A\left(\frac{x}{\log^{A-1} x}\right)$$

*where $c_2$ is the twin prime constant. The same estimate holds if we replace $e(\alpha)$ by $e(-\alpha)$ in the integrand.*

Let $\mathcal{W} = \left\{p^k : k \geqslant 2\right\}$ be the set of proper prime powers. To show how Theorem 6.1 connects with Conjecture 1.1, we have to remove the terms such that either $n$ or $2n+1$ is in $\mathcal{W}$:

$$\sum_{\substack{n \leqslant x \\ n \in \mathcal{W} \text{ or } 2n+1 \in \mathcal{W}}} \Lambda(n)\Lambda(2n+1) \ll \sum_{\substack{n \leqslant x \\ n \in \mathcal{W}}} \Lambda(n)\Lambda(2n+1) + \sum_{\substack{n \leqslant x \\ 2n+1 \in \mathcal{W}}} \Lambda(n)\Lambda(2n+1)$$

$$\ll \log x \left( \sum_{\substack{n \leqslant x \\ n \in \mathcal{W}}} \Lambda(n) + \sum_{\substack{n \leqslant x \\ 2n+1 \in \mathcal{W}}} \Lambda(2n+1) \right) \ll \log x \left( \sum_{\substack{n \leqslant x \\ n \in \mathcal{W}}} \Lambda(n) + \sum_{\substack{n \leqslant 2x+1 \\ n \in \mathcal{W}}} \Lambda(n) \right) \ll \sqrt{x} \log^2 x$$

where the last bound comes from (2.25). Thus we can say

$$\psi_+(x) = \theta_+(x) + O\Big(\sqrt{x} \log x\Big) \quad \text{where } \theta_+(x) = \sum_{\substack{p \leqslant x \\ 2p+1 \text{ is prime}}} \log p \log(2p+1).$$

[1] Let $S^1 = \{z \in \mathbb{C} : |z| = 1\} = \{e(\alpha) : \alpha \in \mathbb{R}\}$, then we can write $S_x(\alpha)$ as a function over $S^1$ which means that we can look at the integral in (6.1) as a complex integral over $S^1$ with a change of variable.

[2] This condition is necessary to keep the major arcs from overlapping, as we will see with (6.2).

[3] This function can be seen as a metric in $S^1$. For any $\alpha, \beta \in \mathbb{R}$, $\|\alpha - \beta\|$ represent the length of the smallest arc on the unit circle between $e(\alpha)$ and $e(\beta)$ divided by $2\pi$. We can note that $\mathfrak{M}(q,a)$ can also be defined by the $\alpha$ such that $|\alpha - a/q| \leqslant 1/Q$, except for $\mathfrak{M}(1,1)$.

Then if we assumed that the contribution of (6.1) did come from the major arcs, then we could get

$$\theta_+(x) = \psi_+(x) + O\Big(\sqrt{x}\log x\Big) = c_2 x + O_A\bigg(\frac{x}{\log^{A-1} x}\bigg).$$

Using partial summation, we can get

$$\pi_+(x) = \int_{2^-}^x \frac{\mathrm{d}\theta_+(t)}{\log t \log(2t+1)} = c_2 \int_2^x \frac{\mathrm{d}t}{\log t \log(2t+1)} + O_A\bigg(\frac{x}{\log^{A+1} x}\bigg),$$

which would lead us to the conjecture

$$\pi_+(x) = \frac{c_2 x}{\log^2 x} + O\bigg(\frac{x}{\log^3 x}\bigg).$$

However, we still did not manage to prove that the minor arcs offer a negligible contribution.

## 6.1. Major arcs

We are going to prove Theorem 6.1 following the ideas of Chapter 26 in [5]. Let $A > 1$. We are setting the parameters defining the major arcs to $P = \log^A x$ and $Q = x/\log^A x$. First, we show that all the $\mathfrak{M}(q, a)$ with $q \leqslant P$ are pairwise disjoint subsets of $[0, 1]$. Let $a/q \neq a'/q'$ be two rational numbers with $q, q' \leqslant P$ in the additive group $\mathbb{R}/\mathbb{Z}$. Then

$$\left|\frac{a}{q} - \frac{a'}{q'}\right| = \frac{|aq' - a'q|}{qq'} \geqslant \frac{1}{qq'} \geqslant \frac{1}{P^2} \geqslant \frac{2}{Q} \tag{6.2}$$

as long as $x$ is big enough to have $x/\log^{3A} x \geqslant 2$. Thus we cannot have two distinct major arcs overlapping.

The reason why it is easy to estimate $S_x(\alpha)$ on major arcs is because, for $\alpha \in \mathfrak{M}(a, q)$, we can approximate $S_x(\alpha)$ by $S_x(a/q)$, which is easily manageable: Using Example 2.13 and the Siegel-Walfisz theorem, we have

$$\begin{aligned}\sum_{\substack{n\leqslant x\\(n,q)=1}} \Lambda(n) e(na/q) &= \frac{1}{\phi(q)} \sum_{n\leqslant x} \Lambda(n) \sum_{\chi \pmod q} \mathcal{G}(\overline{\chi})\chi(na) = \frac{1}{\phi(q)} \sum_{\chi \pmod q} \mathcal{G}(\overline{\chi})\chi(a)\psi(x,\chi)\\ &= \frac{\mathcal{G}(\chi_0)x}{\phi(q)} + O_A\Big(qxe^{-c_1\sqrt{\log x}}\Big) = \frac{\mathcal{G}(\chi_0)x}{\phi(q)} + O_A\Big(xe^{-c_2\sqrt{\log x}}\Big),\end{aligned} \tag{6.3}$$

where $c_1$ and $c_2$ are two positive constants and we used the trivial bound $\mathcal{G}(\overline{\chi}) \ll q$. By using the trivial bound $\omega(q) \leqslant q \leqslant P$, we also have:

$$\sum_{\substack{n\leqslant x\\(n,q)>1}} \Lambda(n) e(na/q) = \sum_{p|q} \log p \sum_{k \leqslant \frac{\log x}{\log p}} e(p^k a/q) \ll \omega(q) \log x \ll \log^{A+1} x.$$

Thus we obtain the estimate

$$S_x(a/q) = \sum_{\substack{n\leqslant x\\(n,q)=1}} \Lambda(n)e(na/q) + \sum_{\substack{n\leqslant x\\(n,q)>1}} \Lambda(n)e(na/q) = \frac{\mu(q)x}{\phi(q)} + O_A\Big(xe^{-c_2\sqrt{\log x}}\Big)\,, \tag{6.4}$$

where we used the fact that $\mathcal{G}(\chi_0) = \mu(q)$. This can be proved using Möbius inversion and the fact that $\sum_{r=1}^{R} e(ar/R) = 0$ if $R \nmid a$:

$$\sum_{\substack{b\leqslant q\\(b,q)=1}} e(b/q) = \sum_{b\leqslant q} e(b/q) \sum_{\substack{d|b\\d|q}} \mu(d) = \sum_{d|q} \mu(d) \sum_{r\leqslant q/d} e(dr/q) = \mu(q). \tag{6.5}$$

For $a$ and $q$ fixed and $\alpha \in \mathfrak{M}(q,a)$, we can now study $S_{2x}(\alpha)$ by defining $\beta := \alpha - a/q \pmod 1$ and writing $e(n\alpha)$ as $e(na/q)e(n\beta)$. Inside the major arc $\mathfrak{M}(a,q)$, we are now working with a variable $\beta$ such that $|\beta| < 1/Q \pmod 1$. Let $B_{a,q}(t) := S_t(a/q)$, $R_{a,q}(t) := B_{a,q}(t) - \mu(q)[t]/\phi(q)$ and $T_x(\beta) := \sum_{n\leqslant x} e(n\beta)$. By using (6.4), we obtain the upper bound

$$R_{a,q}(t) \ll_A te^{-c_2\sqrt{\log t}} \ll_A xe^{-c_2\sqrt{\log x}}$$

uniformly for $1 \leqslant t \leqslant x$. Thus we can use partial summation to change our problem about primes into a problem about natural numbers:

$$\begin{aligned} S_x(\alpha) &= \sum_{n\leqslant x} \Lambda(n)e(na/q)e(n\beta) \\ &= \int_{1^-}^{x} e(t\beta)\,\mathrm{d}B_{a,q}(t) \\ &= \frac{\mu(q)}{\phi(q)} \int_{1^-}^{x} e(t\beta)\,\mathrm{d}[t] + \int_{1^-}^{x} e(t\beta)\,\mathrm{d}R_{a,q}(t) \\ &= \frac{\mu(q)}{\phi(q)} T_x(\beta) + R_{a,q}(x)e(x\beta) - 2\pi i\beta \int_{1}^{x} R_{a,q}(t)e(t\beta)\,\mathrm{d}t \\ &= \frac{\mu(q)}{\phi(q)} T_x(\beta) + O_A\Big((1+x/Q)xe^{-c_2\sqrt{\log x}}\Big) \\ &= \frac{\mu(q)}{\phi(q)} T_x(\beta) + O_A\Big(xe^{-c_3\sqrt{\log x}}\Big)\,, \end{aligned} \tag{6.6}$$

for an absolute positive constant $c_3$. By the same techniques, we can find

$$S_x(-\alpha) = \frac{\mu(q)}{\phi(q)} T_x(-\beta) + O_A\Big(xe^{-c_3\sqrt{\log x}}\Big)\,. \tag{6.7}$$

However, we have to be more vigilant with $S_x(2\alpha)$. The Fourier transform of $n \mapsto e(2na/q)\mathbf{1}_{(n,q)=1}$ in $L^2((\mathbb{Z}/q\mathbb{Z})^*)$ is

$$\frac{1}{\phi(q)} \sum_{\chi \pmod q} \mathcal{G}_2(\overline{\chi})\chi(n) \quad \text{where } \mathcal{G}_2(\chi) := \sum_{a\leqslant q} \chi(a)e(2a/q).$$

Similarly as in (6.4), we have

$$S_x(2a/q) = \frac{\mathcal{G}_2(\chi_0)x}{\phi(q)} + O_A\Big(xe^{-c_2\sqrt{\log x}}\Big) = \frac{\mu_2(q)x}{\phi(q)} + O_A\Big(xe^{-c_2\sqrt{\log x}}\Big)\,,$$

where $\mu_2$ is a multiplicative function[4] with Dirichlet series $(1+2^{1-s})\zeta(s)^{-1}$. We can see that $\mathcal{G}_2(\chi_0) = \mu_2(q)$ by using the same techniques as in (6.5). Hence we have

$$S_x(2\alpha) = \frac{\mu_2(q)}{\phi(q)}T_x(2\beta) + O_A\Big(xe^{-c_3\sqrt{\log x}}\Big)\,. \tag{6.8}$$

By putting (6.6), (6.7) and (6.8) in the integral over a major arc, we get

$$\begin{aligned}\int_{\mathfrak{M}(q,a)} & S_x(2\alpha)S_{2x+1}(-\alpha)e(\alpha)\,\mathrm{d}\alpha \\ &= \frac{\mu_2(q)\mu(q)}{\phi(q)^2}e(a/q)\int_{-1/Q}^{1/Q} T_x(2\beta)T_{2x+1}(-\beta)e(\beta)\,\mathrm{d}\beta + O\bigg(\frac{x^2}{Q}e^{-c_3\sqrt{\log x}}\bigg)\end{aligned}$$

which implies that when we bring the major arcs together, there exists a positive constant $c'$ such that

$$\begin{aligned}\int_{\mathfrak{M}} & S_x(2\alpha)S_{2x+1}(-\alpha)e(\alpha)\,\mathrm{d}\alpha \\ &= \Bigg(\sum_{q\leqslant P}\frac{\mu_2(q)\mu^2(q)}{\phi(q)^2}\Bigg)\int_{-1/Q}^{1/Q} T_x(2\beta)T_{2x+1}(-\beta)e(\beta)\,\mathrm{d}\beta + O\Big(xe^{-c'\sqrt{\log x}}\Big)\,. \quad (6.9)\end{aligned}$$

**Lemma 6.2.** *For $x \geqslant 2$ and $Q = \frac{x}{\log^A x}$, we have*

$$\int_{-\frac{1}{Q}}^{\frac{1}{Q}} T_x(2\beta)T_{2x+1}(-\beta)e(\beta)\,\mathrm{d}\beta = x + O\bigg(\frac{x}{\log^A x}\bigg)\,.$$

***Proof.*** To understand the integral above, we want to estimate it with the same integral on the interval $[0,1]$. This allows us to transform the calculation of this integral into a combinatorial problem like in (6.1):

$$\int_0^1 T_x(2\beta)T_{2x+1}(-\beta)e(\beta)\,\mathrm{d}\beta = \#\{(n,m) : n \leqslant x \text{ and } m = 2n+1\} = x + O(1)\,. \tag{6.10}$$

The only major arc for the integers (where the Fourier series $T_x(\beta)$ is at its largest) is around 1 on the unit circle. To see this, $T_x(\beta)$ can be understood as a geometric sum: For

[4] We prove the multiplicativity of $\mu_2$ since it is the Dirichlet convolution of two multiplicative functions. The values of $\mu_2$ coincides with the ones for the Möbius function $\mu$ at odd integers and have the values $\mu_2(2) = 1$, $\mu_2(4) = -2$ and $\mu_2(2^k) = 0$ for $k \geqslant 3$. In general, one might have to work with the sum $c_q(n) = \sum_{a\leqslant q}\chi_0(a)e(na/q)$ which is called the *Ramanujan's sum*. We can notice that $c_q(1) = \mu(q)$ and $c_q(2) = \mu_2(q)$. By applying the same methods as we did to understand the multiplicativity and the Dirichlet series of $\mu_2$, we can say that $q \mapsto c_q(n)$ is multiplicative and find the formula $c_q(n) = \frac{\mu(q/(n,q))\phi(q)}{\phi(q/(n,q))}$ by seeing that it coincides at prime powers.

$\beta \notin \mathbb{Z}$, we have

$$T_x(\beta) = \sum_{n \leqslant x} e(n\beta) = \frac{e([x+1]\beta) - e(\beta)}{e(\beta) - 1} \ll \frac{1}{|e(\beta) - 1|} \ll \frac{1}{\|\beta\|} \tag{6.11}$$

uniformly for $x > 0$. Thus, when $2\beta \notin \mathbb{Z}$, then $T_x(2\beta)T_{2x+1}(-\beta)e(\beta) \ll \|2\beta\|^{-1}\|\beta\|^{-1}$ and

$$\int_{\frac{1}{Q}}^{\frac{1}{2}-\frac{1}{Q}} T_x(2\beta)T_{2x+1}(-\beta)e(\beta)\,\mathrm{d}\beta \ll \int_{\frac{1}{Q}}^{\frac{1}{2}-\frac{1}{Q}} \frac{\mathrm{d}\beta}{\|2\beta\|\,\|\beta\|} = \frac{1}{2}\int_{\frac{2}{Q}}^{1-\frac{2}{Q}} \frac{\mathrm{d}\beta}{\|\beta\|\,\|\beta/2\|} = 2\int_{\frac{2}{Q}}^{\frac{1}{2}} \frac{\mathrm{d}\beta}{\beta^2} \ll Q. \tag{6.12}$$

We can get the same upper bound using the same method for the same integral in the interval $[1/2 + 1/Q, 1 - 1/Q]$. For the integral on an interval very close to $1/2$, we have the trivial bound $T_x(2\beta) \ll x$ and $T_{2x+1}(\beta) \ll 1$ by (6.11), which leads us to

$$\int_{\frac{1}{2}-\frac{1}{Q}}^{\frac{1}{2}+\frac{1}{Q}} T_x(2\beta)T_{2x+1}(-\beta)e(\beta)\,\mathrm{d}\beta \ll \frac{x}{Q}. \tag{6.13}$$

By putting together (6.10), (6.12) and (6.13), we can prove the lemma. □

The summation in (6.9) converges if $P \to \infty$ which means that we can estimate it using the infinite series. The Euler product representation of the Dirichlet series of $f(q) = \frac{\mu_2(q)\mu^2(q)}{\phi^2(q)}$ is

$$F(s) = \left(1 + \frac{1}{2^s}\right)\prod_{p \geqslant 3}\left(1 - \frac{1}{(p-1)^2 p^s}\right),$$

which converges absolutely for $\sigma > -1$. Thus if we let $\varepsilon = \frac{1}{A}$, then

$$\sum_{q \leqslant P} \frac{\mu_2(q)\mu^2(q)}{\phi^2(q)} = F(0) + O\left(\sum_{q > P} |f(q)|\right) = c_2 + O\left(\frac{1}{P^{1-\varepsilon}} \sum_{q > P} \frac{|f(q)|}{q^{-1+\varepsilon}}\right) = c_2 + O\left(\frac{1}{\log^{A-1} x}\right).$$

Placing this estimate and Lemma 6.2 into (6.9), we get Theorem 6.1. The proof is very similar for the Sophie Germain primes of the second kind.

## 6.2. Minor arcs in ternary problems

The study of $\psi_+(x)$ is a *binary problem*, in the sense that we are studying the summatory function of $\Lambda(n)\Lambda(m)$ over pairs of integers $(n, m)$ satisfying one linear condition. In contrast, a *ternary problem* would be the study of a summatory function $\Lambda(n_1)\Lambda(n_2)\Lambda(n_3)$ over triplets of integers $(n_1, n_2, n_3)$ satisfying one linear condition. We have a hard time understanding the contribution of minor arcs integral in (6.1), and this characteristic is shared with most binary problems. In this section, we will show a way to deal with the minor arcs in a ternary problem, and we will explain in the next section why we cannot use the same techniques to understand binary problems and show that the minor arcs have a negligible contribution compared to the major arcs. For example, the *Goldbach's conjecture*, stating that every even integer greater or equal to 4 can be written as a sum of two primes, is a binary problem and remains unsolved since Christian Goldbach stated the problem in a letter to Euler in 1742. However, *Goldbach's weak conjecture*, which says that every odd integer greater or equal to 7 can be written as a sum of three primes, is a ternary problem. It has been solved, and the main argument of the proof uses the circle method[5].

An example of a ternary problem is the study of prime pairs $p_1, p_2$ such that $p_1 + p_2 + 1$ is also prime. We can notice that if we have a pair $(p_1, p_2)$ with $p_1 = p_2$ and with $p_1 + p_2 + 1$ prime, then $p_1$ is a Sophie Germain prime of the first kind. If a proof of Goldbach's conjecture existed, we would immediately deduce the infinitude of pairs $(p_1, p_2)$ with $p_1 + p_2 + 1$ prime. We will prove this unconditionally using the circle method:

**Theorem 6.3.** *For $x \geqslant 2$, the number of unordered pairs of primes $p_1, p_2 \leqslant x$ such that $p_1 + p_2 + 1$ is also prime is*

$$\left(\frac{K}{2} + O\left(\frac{\log\log x}{\log x}\right)\right)\frac{x^2}{\log^3 x}$$

*where*

$$K = \prod_p \left(1 + \frac{1}{(p-1)^3}\right) \approx 2.30096\ldots$$

We will eventually have to evaluate an integral around the unit circle like before. Still, in order to get an upper bound on the integral over the minor arcs, we first need Vinogradov's nontrivial upper bound on the Fourier series $S_x(\alpha) = \sum_{n \leqslant x} \Lambda(n) e(n\alpha)$:

**Lemma 6.4.** *For $x \geqslant 2$, $\alpha \in \mathbb{R}$, $q > 0$ and an irreducible fraction $\frac{a}{q}$ such that $\left|\alpha - \frac{a}{q}\right| < \frac{1}{q^2}$, we have*

$$S_x(\alpha) \ll (xq^{-1/2} + x^{4/5} + (xq)^{1/2}) \log^{5/2} x.$$

A proof of this is given in Chapter 23 of [16]. To use this lemma, we will need to know that for every real number $\alpha$, there is a rational number that is close enough relative to the magnitude of the denominator.

In particular, using Dirichlet's approximation theorem (Lemma 3.4 with $d = 1$), we know that for every $\alpha \in \mathbb{R}$ there always exists an irreducible fraction $a/q$ such that $|\alpha - a/q| < q^{-2}$. We will now use the same setup as in Section 6.1 for the proof of Theorem 6.3.

***Proof of Theorem 6.3.*** As in the case of Sophie Germain primes, let us define

$$R(x) := \sum_{n,m \leqslant x} \Lambda(n)\Lambda(m)\Lambda(n+m+1).$$

[5] In 1922, Hardy and Littlewood proved that, assuming the Generalized Riemann Hypothesis, Goldbach's weak conjecture holds for every sufficiently large integer. In 1937, I. M. Vinogradov gave a proof of this unconditionally. Finally, in 2013, Harald A. Helfgott proved Goldbach's weak conjecture for every integer $n \geqslant 7$.

It suffices to prove that $R(x) = Kx^2(1 + O(1/\log x))$. Later in the proof, we will show how we can deduce the theorem from the result on $R(x)$. Using the same method as in (6.1), we have

$$\begin{aligned} R(x) &= \int_0^1 S_x(\alpha)^2 S_{2x+1}(-\alpha)e(\alpha)\,\mathrm{d}\alpha \\ &= \int_{\mathfrak{M}} S_x(\alpha)^2 S_{2x+1}(-\alpha)e(\alpha)\,\mathrm{d}\alpha + \int_{\mathfrak{m}} S_x(\alpha)^2 S_{2x+1}(-\alpha)e(\alpha)\,\mathrm{d}\alpha \end{aligned}$$

with the major and minor arcs defined in exactly the same way as in the previous sections with $P = \log^9 x$ and $Q = [x/\log^9 x]$. We follow the same procedure as Section 6.1 to obtain

$$\int_{\mathfrak{M}} S_x(\alpha)^2 S_{2x+1}(-\alpha)e(\alpha)\,\mathrm{d}\alpha = Kx^2\left(1 + O\left(\frac{1}{\log^8 x}\right)\right).$$

By Lemma 3.4 there exists an irreducible fraction $a/q$ such that $q \leqslant Q$ and $|\alpha - a/q| < 1/(qQ) \leqslant q^2$ and if $\alpha \in \mathfrak{m}$, then we know that $q > P$. Using Lemma 6.4, we know that for $\alpha \in \mathfrak{m}$

$$S_{2x+1}(-\alpha) \ll (xP^{-1/2} + x^{4/5} + (xQ)^{1/2})\log^{5/2} x = \frac{x}{\log^2 x}. \tag{6.14}$$

Furthermore, if we look at $S_x(\alpha)$ as a Fourier series, then Parseval's identity gives us the average value of $|S_x(\alpha)|^2$:

$$\int_0^1 |S_x(\alpha)|^2\ \mathrm{d}\alpha = \sum_{n\leqslant x} \Lambda(n)^2 = x\log x + O(x)\,. \tag{6.15}$$

since proper prime powers have a contribution $\ll \sqrt{x}\log x$ and partial summation leads us to

$$\sum_{p\leqslant x} \log^2 p = \int_1^x \log t\,\mathrm{d}\theta(t) = \theta(x)\log x - \int_1^x \frac{\theta(t)}{t}\,\mathrm{d}t = x\log x + O(x)\,.$$

Combining (6.14) and (6.15), we can get an upper bound on the contribution of the minor arcs:

$$\int_{\mathfrak{m}} S_x(\alpha)^2 S_{2x+1}(-\alpha)e(\alpha)\,\mathrm{d}\alpha \ll \int_{\mathfrak{m}} |S_x(\alpha)|^2\,|S_{2x+1}(-\alpha)|\ \mathrm{d}\alpha \ll \frac{x^2}{\log x}. \tag{6.16}$$

If we now combine the results of the major and minor arcs, we obtain

$$R(x) = \sum_{n,m\leqslant x} \Lambda(n)\Lambda(m)\Lambda(n+m+1) = Kx^2 + O\left(\frac{x^2}{\log x}\right).$$

To remove the contribution of proper prime powers, we use the same technique as before and say that if we let $\mathcal{W} = \left\{p^k : k \geqslant 2\right\}$, then

$$\sum_{\substack{n,m\leqslant x\\ n\in\mathcal{W}\text{ or}\\ m\in\mathcal{W}\text{ or}\\ n+m+1\in\mathcal{W}}} \Lambda(n)\Lambda(m)\Lambda(n+m+1) \leqslant \sum_{\substack{n,m\leqslant x\\ n\in\mathcal{W}}} \Lambda(n)\Lambda(m)\Lambda(n+m+1)$$

$$+ \sum_{\substack{n,m\leqslant x\\ m\in\mathcal{W}}} \Lambda(n)\Lambda(m)\Lambda(n+m+1) + \sum_{\substack{n,m\leqslant x\\ n+m+1\in\mathcal{W}}} \Lambda(n)\Lambda(m)\Lambda(n+m+1)$$

$$\ll \log^2 x \left( x \sum_{\substack{n\leqslant x\\ n\in\mathcal{W}}} \Lambda(n) + x \sum_{\substack{m\leqslant x\\ m\in\mathcal{W}}} \Lambda(m) + \sum_{\substack{n,m\leqslant x\\ n+m+1\in\mathcal{W}}} \Lambda(n+m+1) \right).$$

The number of pairs $n, m \leqslant x$ such that $n + m + 1 = k$ is less than $x$, which implies that we have

$$\ll x\log^2 x \left( \sum_{\substack{n\leqslant x\\ n\in\mathcal{W}}} \Lambda(n) + \sum_{\substack{m\leqslant x\\ m\in\mathcal{W}}} \Lambda(m) + \sum_{\substack{3\leqslant k\leqslant 2x+1\\ k\in\mathcal{W}}} \Lambda(k) \right) \ll x^{3/2}\log^3 x.$$

The last bound comes from (2.25), and it leaves us with

$$B(x) := \sum_{\substack{p_1,p_2\leqslant x\\ p_1+p_2+1 \text{ is prime}}} \log p_1 \log p_2 \log(p_1+p_2+1) = Kx^2 + O\left(\frac{x^2}{\log x}\right). \tag{6.17}$$

Finally, we cannot use partial summation as usual to remove the logarithmic weights since the sum is over two variables[6]. First,

$$B(x) = \sum_{\substack{p_1,p_2\leqslant x\\ p_1+p_2+1 \text{ is prime}}} \log p_1 \log p_2 \log(p_1+p_2+1) \leqslant r(x)\log^3 x \left(1 + O\left(\frac{1}{\log x}\right)\right) \tag{6.18}$$

where

$$r(x) := \sum_{\substack{p_1,p_2\leqslant x\\ p_1+p_2+1 \text{ is prime}}} 1.$$

Let

$$r_\delta(x) := \sum_{\substack{x^{1-\delta}<p_1,p_2\leqslant x\\ p_1+p_2+1 \text{ is prime}}} 1.$$

By the PNT, we have for $x$ sufficiently large and $\delta < 1/4$:

$$r(x) - r_\delta(x) \leqslant 2 \sum_{\substack{p_1,p_2\leqslant x\\ p_1\leqslant x^{1-\delta}\\ p_1+p_2+1 \text{ is prime}}} 1 \leqslant 2\pi(x^{1-\delta})\pi(x) \leqslant \frac{3x^{2-\delta}}{\log^2 x},$$

[6] The rest of the proof, about deducing the theorem from (6.17), is inspired from the proof of Theorem 8.1 in [21] (pp. 228-230).

which implies

$$B(x) \geqslant \sum_{\substack{x^{1-\delta}<p_1,p_2\leqslant x \\ p_1+p_2+1 \text{ is prime}}} \log p_1 \log p_2 \log(p_1+p_2+1) \geqslant (1-\delta)^3 \log^3(x) r_\delta(x)$$

$$\geqslant (1-O(\delta))\log^3 x \left(r(x) - \frac{3x^{2-\delta}}{\log^2 x}\right) \implies \frac{B(x)}{r(x)\log^3 x} \geqslant (1-O(\delta))\left(1 - \frac{3x^{2-\delta}}{r(x)\log^2 x}\right).$$

We have the lower bound $r(x)\log^2 x \gg x^2/\log x$ by combining (6.17) and (6.18) which ultimately gives us

$$\frac{B(x)}{r(x)\log^3 x} \geqslant 1 + O\left(\delta + \frac{\log x}{x^\delta}\right).$$

By choosing $\delta = 2\log\log x/\log x$ and taking the upper bound (6.18) into account[7], we obtain

$$B(x) = r(x)\log^3 x\left(1 + O\left(\frac{\log\log x}{\log x}\right)\right) \implies r(x) = \frac{B(x)}{\log^3 x}\left(1 + O\left(\frac{\log\log x}{\log x}\right)\right)$$
$$= \frac{Kx^2}{\log^3 x}\left(1 + O\left(\frac{\log\log x}{\log x}\right)\right),$$

which proves of the theorem when applying the symmetry arguments since almost[8] every unordered pair $p_1, p_2 \leqslant x$ in the sum $r(x)$ has been counted twice. □

## 6.3. Minor arcs in binary problems

Let's come back to the distribution of Sophie Germain primes of the first kind. To ultimately prove Conjecture 1.1 and have an asymptotic for $\pi_+(x)$, we need to estimate

$$\int_{\mathfrak{m}} S_x(2\alpha) S_{2x+1}(-\alpha) e(\alpha)\,\mathrm{d}\alpha. \qquad (6.19)$$

From (6.15), we know that the expected value of $|S_x(\alpha)|^2$ if $\alpha$ is uniformly distributed on the interval $[0,1]$ is asymptotic to $x\log x$ as $x \to \infty$, thus we should expect that

$$|S_x(\alpha)| = x^{1/2+o(1)}$$

for a generic $\alpha \in [0,1]$. Let $\lambda$ be the Lebesgue measure. We can note that, in practice, we choose the major arcs such that $\lambda(\mathfrak{M}) = o(1)$ and $\lambda(\mathfrak{m}) = 1 - o(1)$, so we expect $|S_x(\alpha)| = x^{1/2+o(1)}$ for most $\alpha \in \mathfrak{m}$. Then no matter how we choose our major and minor arcs, we cannot expect to bound the contribution over the minor arcs trivially, otherwise we get a bound as big as the main term found in Theorem 6.1. With the Cauchy-Schwarz inequality, we can prove an upper bound on the expected value of $|S(\alpha)|$ rigorously, if $\alpha$ was

[7] Note that choosing $\delta$ this way means that $x^{1-\delta} = x/\log^2 x$.

[8] The only exceptions are when $p_1 = p_2$ are both Sophie Germain primes of the first kind. They have a negligible contribution to the sum since there is so few of them.

uniformly distributed in $\mathfrak{m}$:

$$\left(\int_{\mathfrak{m}} |S_x(\alpha)| \, \mathrm{d}\alpha\right)^2 \leqslant \left(\int_0^1 |S_x(\alpha)| \, \mathrm{d}\alpha\right)^2 \leqslant \int_0^1 |S_x(\alpha)|^2 \, \mathrm{d}\alpha \sim x \log x$$
$$\implies \frac{1}{\lambda(\mathfrak{m})} \int_{\mathfrak{m}} |S_x(\alpha)| \, \mathrm{d}\alpha \leqslant \sqrt{x \log x}(1 + o(1)).$$

If we were to use the circle method to prove Conjecture 1.1, we would need to find parts of the integral in (6.19) which cancel each other.

# Chapter 7

# Numerical investigations

In this chapter, we want to understand if the Sophie Germain prime number race shares similarities with Chebyshev's race or if it behaves more like a random walk. Under GRH, the proof of the existence of a limiting distribution for the function $E(x; 4, 3, 1)$ comes from the fact that we were able to prove that $E(e^u; 4, 3, 1)$ is close to a sum of periodic functions in the function space $L^2$ (see Theorem 3.1). This sum can be decomposed into a constant function, which led to the bias and a superposition of trigonometric functions. We obtained the sum of trigonometric functions by using the explicit formula for $\psi(x, \chi_1)$ where $\chi_1$ was the nonprincipal character mod 4. We know that $\psi(x, \chi_1)$ is a step-function that jumps at every prime power, and squares' contribution is not negligible in the study of this prime number race. However, the fact that every square of primes is 1 mod 4 creates this imbalance. In other words, numbers that are 3 mod 4 have more chance to be prime since they cannot be square numbers. Maybe the squares also lead to a bias in the Sophie Germain prime number race.

The graph of $\frac{\log x}{\sqrt{x}}(\pi_+(x) - \pi_-(x))$ in Figure 7.1 looks quite different from the races in Figure 3.1 or in Figure 3.2. We are going to study this prime number race with a logarithmic

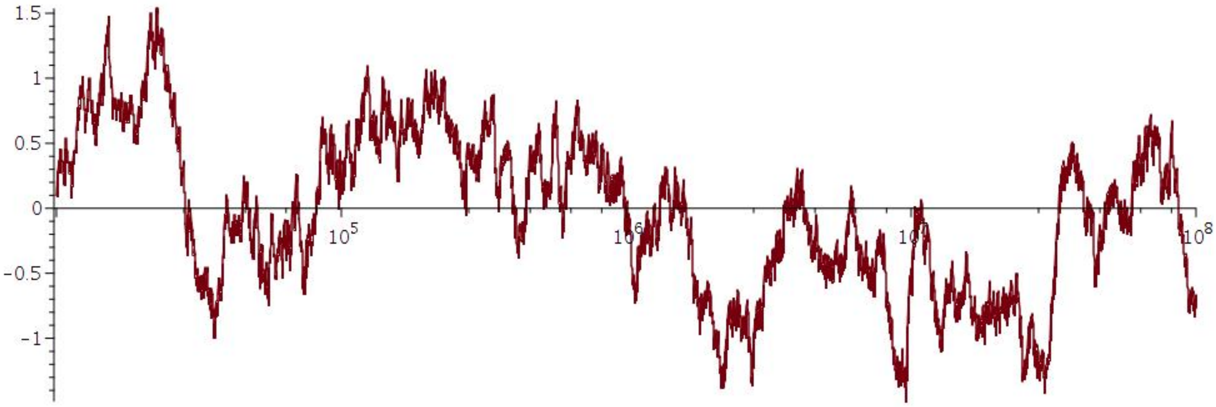

**Fig. 7.1.** Graph of $\frac{\log x}{\sqrt{x}}(\pi_+(x) - \pi_-(x))$.

scale. This is because primes are on average logarithmically spaced, which is why it is not really surprising to see in Chapter 3 that the right way to measure the distribution of the values of $E(x;4,3,1)$ is through the logarithmic density. Maybe the Sophie Germain prime number race is periodic on the logarithmic scale.

## 7.1. Do squares make a difference?

In Chebyshev's race, the bias essentially came from the fact that every square of primes is of the form 1 (mod 4). When we look at our race $\pi_+$ vs. $\pi_-$, the number $2p+1$ can never be a square because, for any prime $p \geqslant 3$, we have

$$2p+1 \text{ is a square } \implies 2p+1 \equiv 1 \pmod 4 \implies 4 \mid 2p$$

which is impossible. However, for any prime $p$,

$$2p-1 \text{ is a square } \iff \exists k \text{ such that } 2p-1=(2k+1)^2 \iff p = 2k^2+2k+1.$$

By using the Cramér-Granville model from Chapter 5, we can then conjecture that the function

$$\xi(x) = \#\left\{p \leqslant x : \exists k \text{ such that } p = 2k^2+2k+1\right\} \sim \sum_{k \leqslant (x/2)^{1/2}+O(1)} \mathbb{E}[X_{2k^2+2k+1}] \sim \frac{C\sqrt{x}}{\log x}, \tag{7.1}$$

where

$$C = \sqrt{2}\prod_p \left(1-\frac{\chi_1(p)}{p-1}\right) \approx 1.9414\dots$$

with $\chi_1$ being the only nonprincipal character mod 4. The second product is convergent because if we let $B(x)=\sum_{p\leqslant x}\chi_1(p)$ which is $\ll xe^{-c\sqrt{\log x}}$ for a positive constant $c$ by the PNT for arithmetic progressions, then with partial summation we get

$$\begin{aligned}\sum_{p>y}\chi_1(p)\log\left(1-\frac{1}{p}\right) &= -B(y)\log\left(1-\frac{1}{y}\right) - \int_y^\infty \frac{B(t)}{t(t-1)}\,\mathrm{d}t \\ &\ll e^{-c\sqrt{\log y}} + \int_y^\infty \frac{e^{-c\sqrt{\log t}}}{t}\,\mathrm{d}t \ll e^{-c'\sqrt{\log y}}\end{aligned}$$

so that

$$\begin{aligned}\prod_{p>y}\left(1-\frac{\chi_1(p)}{p-1}\right) &= \prod_{\substack{p>y \\ p\equiv 1 \ (\mathrm{mod}\ 4)}}\left(1-\frac{1}{(p-1)^2}\right)\prod_{p>y}\left(1-\frac{1}{p}\right)^{\chi_1(p)} \\ &= (1+O(1/y))\left(1+O\left(e^{-c'\sqrt{\log y}}\right)\right) = 1+O\left(e^{-c'\sqrt{\log y}}\right)\end{aligned}$$

for a constant $c' < c$.

If the Sophie Germain race was anything like the Chebyshev race, the fact that $2p+1$ can never be a square should make it more probable to be a prime than $2p-1$. However, Figure 7.1 shows that the Sophie Germain prime race has a very different structure than Chebyshev's race.

If we suppose that, like in the Chebyshev race, the function $c_2x$ is a good approximation for $\sum_{n\leqslant x}\Lambda(n)\Lambda(2n+1)$ in the sense that the race between these two functions do not have any bias, then so will $c_2\mathcal{L}(x)$ against $\Pi_+(x) := \sum_{n\leqslant x}\frac{\Lambda(n)\Lambda(2n+1)}{\log n\log(2n+1)}$. We can decompose $\Pi_+(x)$ into three different sums. The first is over the $n$ which are Sophie Germain primes of the first kind, the second is over the $n$ which are squares of primes and $2n+1$ is prime, and the third is over the $n$ such that either $n$ or $2n+1$ is in $\mathcal{W}_3 := \left\{p^k : k\geqslant 3\right\}$. Note that the second sum equals to $1/2$ since for every $p\neq 3$, we have $3\mid 2p^2+1$.

$$\Pi_+(x) = \pi_+(x) + \frac{1}{2} + \sum_{\substack{n\leqslant x\\ n\in\mathcal{W}_3 \text{ or}\\ 2n+1\in\mathcal{W}_3}}\frac{\Lambda(n)\Lambda(2n+1)}{\log n\log(2n+1)}.$$

We can bound the sum above by

$$\sum_{\substack{n\leqslant x\\ n\in\mathcal{W}_3 \text{ or}\\ 2n+1\in\mathcal{W}_3}}\frac{\Lambda(n)\Lambda(2n+1)}{\log n\log(2n+1)} \ll \sum_{\substack{n\leqslant x\\ n\in\mathcal{W}_3 \text{ or}\\ 2n+1\in\mathcal{W}_3}} 1 \ll \sum_{\substack{n\leqslant x\\ n\in\mathcal{W}_3}} 1 + \sum_{\substack{n\leqslant x\\ 2n+1\in\mathcal{W}_3}} 1 \ll \sum_{p\leqslant x^{1/3}}\sum_{k\leqslant\frac{\log x}{\log p}} 1 \ll x^{1/3}.$$

Thus $\Pi_+(x) = \pi_+(x) + O\left(x^{1/3}\right)$ which means that $c_2\mathcal{L}(x)$ would also be a good approximation for $\pi_+(x)$ since the error term $O\left(x^{1/3}\right)$ is less than the standard deviation of the race.

However, if we do the same process for $\Pi_-(x) := \sum_{n\leqslant x}\frac{\Lambda(n)\Lambda(2n-1)}{\log n\log(2n-1)}$, then we have to take into account whenever $n$ or $2n-1$ is the squares of a prime which would give us

$$\Pi_-(x) = \pi_-(x) + \frac{1}{2}\xi_1(x) + \frac{1}{2}\xi_2(x) + \frac{1}{4}\xi_3(x) + O\left(x^{1/3}\right)$$

where

$$\begin{aligned}\xi_1(x) &:= \#\left\{p\leqslant x : 2p-1 \text{ is the square of a prime}\right\}\\ &= \#\left\{p\leqslant\sqrt{2x-1} : \frac{p^2+1}{2} \text{ is prime}\right\},\\ \xi_2(x) &:= \#\left\{p\leqslant\sqrt{x} : 2p^2-1 \text{ is prime}\right\}\\ \text{and}\quad \xi_2(x) &:= \#\left\{p\leqslant\sqrt{x} : 2p^2-1 \text{ is the square of a prime}\right\}.\end{aligned}$$

Let $\mathcal{A} = \left\{n(n^2+1)/2 : n\leqslant\sqrt{2x-1}\right\}$, $\mathcal{B} = \{n(2n^2-1) : n\leqslant\sqrt{x}\}$, $\mathcal{P} = \{p\leqslant x^{1/5}\}$, and let's use the same definitions as in Section 4.5. By applying Selberg's sieve method from Chapter 4 or by directly using Theorem 4.5, we have the upper bounds

$$\xi_1(x) \leqslant S(\mathcal{A},\mathcal{P}) + O\left(x^{1/5}\right) \ll \frac{\sqrt{x}}{\log^2 x}$$

and

$$\xi_2(x), \xi_3(x) \leqslant S(\mathcal{B},\mathcal{P}) + O\Big(x^{1/5}\Big) \ll \frac{\sqrt{x}}{\log^2 x}.$$

The main reason why we have a better bound than the trivial bound is because of the sifting dimension. We know that $n(2n^2-1) \equiv 0 \pmod{p}$ has three solutions if $p \equiv 1, 3 \pmod{8}$, and it has only one solution if $p \equiv 5, 7 \pmod{8}$. We also know that $n(n^2+1)/2 \equiv 0 \pmod{p}$ has three solutions for $p \equiv 1 \pmod{4}$ and only one for $p \equiv 3 \pmod{4}$. Thus both problems $S(\mathcal{A},\mathcal{P})$ and $S(\mathcal{B},\mathcal{P})$ respect the Axiom 2 with $\kappa = 2$.

We can conclude that

$$\Pi_+(x) - \Pi_-(x) = \pi_+(x) - \pi_-(x) + O\left(\frac{\sqrt{x}}{\log^2 x}\right).$$

Since the error term is negligible compared to the conjectured standard deviation $\sqrt{2c_2x}/\log x$ from (5.10), we conclude that squares do not influence the Sophie Germain race like they do in the Chebyshev race.

## 7.2. Periodicity

Following the discussion in Section 3.3, a finite superposition of trigonometric function is almost a periodic function. The next theorem quantifies this "almost periodicity" by giving a bound on the second moment of $E(e^{U+h};4,3,1) - E(e^U;4,3,1)$ where $U$ is a uniform random variable in the interval $[0, X]$ for an $X$ very large.

**Theorem 7.1.** *Let $\chi_1$ be the nonprincipal character mod 4, $T \geqslant 1$, $\delta > 0$ and let $h > 0$ such that $\|\gamma h/2\pi\| < \delta$ for every nontrivial zero of $L(s,\chi_1)$ with imaginary part $0 < \gamma \leqslant T$. Then, assuming GRH, we have*

$$\frac{1}{X}\int_0^X (E(e^{u+h};4,3,1) - E(e^u;4,3,1))^2 \,\mathrm{d}u \ll \delta^2 \log^4 T + \frac{\log^2 T}{T} + \frac{1}{X}.$$

***Proof.*** To prove this theorem, we can use the trigonometric approximation of $E(e^u;4,3,1)$ that we found in Theorem 3.1. We have the following identities:

$$\begin{aligned}\cos(\gamma(u+h)) - \cos(\gamma u) &= -2\sin(\gamma h/2)\sin(\tfrac{\gamma}{2}(2u+h)),\\ \sin(\gamma(u+h)) - \sin(\gamma u) &= 2\sin(\gamma h/2)\cos(\tfrac{\gamma}{2}(2u+h)).\end{aligned}$$

Thus we can use these equations to find that

$$\sum_{0<\gamma\leqslant T} \frac{\cos(\gamma(u+h)) - \cos(\gamma u) + 2\gamma(\sin(\gamma(u+h)) - \sin(\gamma u))}{\frac14 + \gamma^2}$$
$$= \sum_{0<\gamma\leqslant T} \frac{4\sin(\gamma h/2)\sin(\frac{\gamma}{2}(2u+h) - \theta_\gamma)}{\sqrt{\frac14+\gamma^2}},$$

where $\theta_\gamma$ is defined as the angle in any right triangle such that the adjacent leg to $\theta_\gamma$ is of length 1 and its opposite leg is of length $2\gamma$ (similarly as in the proof of Theorem 3.9).

For $X \geqslant h$, let's denote the norm in $L^2$ by $\|f(u)\|_2 = \left(\int_0^X f(u)^2 \,\mathrm{d}u\right)^{1/2}$ for any function $f \in L^2([0,X])$ (not to be confused with $\|\cdot\|$ as the distance to the closest integer in the statement of the theorem). We can define the function $\varepsilon(u;T)$ as in Theorem 3.1, and use Minkowski's inequality to find that

$$\begin{aligned}
&\left\| E(e^{u+h};4,3,1) - E(e^u;4,3,1) \right\|_2 \\
&\qquad \leqslant \sum_{0<\gamma\leqslant T} \frac{4\left|\sin(\gamma h/2)\right| \left\|\sin(\frac{\gamma}{2}(2u+h)-\theta_\gamma)\right\|_2}{\sqrt{\frac{1}{4}+\gamma^2}} + \|\varepsilon(u+h;T)\|_2 + \|\varepsilon(u;T)\|_2 \\
&\qquad\qquad \ll \delta\sqrt{X}\sum_{0<\gamma\leqslant T}\frac{1}{\gamma} + \frac{\log T\sqrt{X}}{\sqrt{T}} + 1
\end{aligned}$$

where we used the bound $|\sin(\pi x)| \leqslant \pi\|x\|$ for every real number $x$. We can finally use the density of the zeros of $L$-functions (Theorem 2.15) to have the bound $\sum_{0<\gamma\leqslant T} 1/\gamma \ll \log^2 T$ and prove the theorem. □

This theorem shows that the "almost periods" of $E(e^u;4,3,1)$ are characterized by the numbers $h$ such that $\gamma h/2\pi$ is close to an integer for every first few zeros of $L(s,\chi_1)$. Figure 7.2 gives out the second moment of the difference between the graph and different translations of the graph 3.1 on the logarithmic scale.

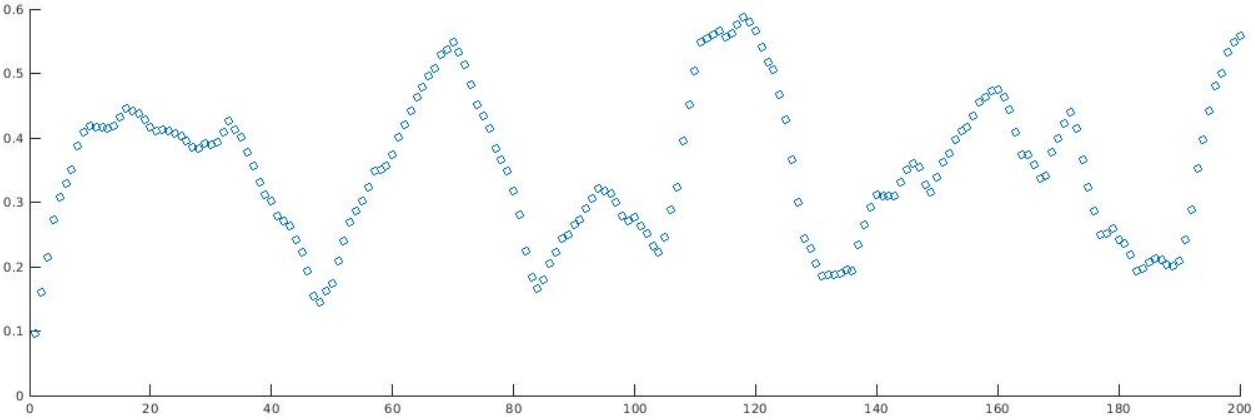

**Fig. 7.2.** Second moment of $E(10^{U+x/100};4,3,1) - E(10^U;4,3,1)$ for a random variable $U$ which is uniform in $[0,10^6]$.

To show that the depressions in the graph corresponds to the "almost periods" of $E(e^u;4,3,1)$ discussed above, Table 7.1 gives the values of $x$ where the second moment in Figure 7.2 is low and their corresponding values for $\frac{x\gamma\log 10}{200\pi}$ for the first five nontrivial zeros of $L(s,\chi_1)$. The five zeros of $L(s,\chi_1)$ in the table are taken out of [29].

| Values of $x$ | Second moment in Figure 7.2 | Zeros of $L(s,\chi_1)$ | | | | |
|---|---|---|---|---|---|---|
| | | **6.02095** | **10.24377** | **12.98810** | **16.34261** | **18.29199** |
| 48 | 0.1455 | 1.05911 | 1.80193 | 2.28467 | 2.87474 | 3.21764 |
| 84 | 0.1658 | 1.85345 | 3.15337 | 3.99816 | 5.03079 | 5.63088 |
| 104 | 0.2220 | 2.29474 | 3.90417 | 4.95011 | 6.22860 | 6.97156 |
| 131 | 0.1854 | 2.89049 | 4.91776 | 6.23523 | 7.84564 | 8.78149 |
| 167 | 0.3360 | 3.68483 | 6.26920 | 7.94873 | 10.0017 | 11.1947 |
| 183 | 0.1926 | 4.03787 | 6.86984 | 8.71029 | 10.9599 | 12.2673 |

**Table 7.1.** Values of $\frac{x\gamma \log 10}{200\pi}$ for the corresponding nontrivial zero $\gamma$ of $L(s,\chi_1)$.

The colours in the tables are there to highlight the values close to an integer. If the value $\left\|\frac{x\gamma \log 10}{200\pi}\right\| \leqslant 0.05$, then it is in red. If it is between 0.05 and 0.10, then it is in magenta. If it is between 0.10 and 0.15, then it is in blue.

We can notice that the low spikes in Figure 3.1, where the graph falls below the $x$-axis and have been characterized at the beginning of Chapter 3, seem to be evenly spaced on the logarithmic scale. In fact

$$\frac{100(\log 616841 - \log 26861)}{\log 10} \approx 136.105 \quad \text{and} \quad \frac{100(\log 12306137 - \log 616841)}{\log 10} \approx 129.995.$$

Those values are close to $x = 131$.

We can also see that the second moment for the value $x = 167$ is high, but it is around a local minimum. This shows that the first zeros' contribution for a low second moment is more important than the next ones. We can understand this directly from Theorem 3.1 since the smaller zeros represent waves with higher amplitudes.

In the Sophie Germain race, we can create a graph similar to Figure 7.2. We can compute the second moment in Figure 7.3 of the difference between translations of the graph over the logarithmic scale in Figure 7.1. We see that the Figure 7.3 points do not oscillate as much and do not go as low relatively to their mean value as the ones in Figure 7.2. It means that we should not expect to have a trigonometric representation that approximates our race. We can't use the same techniques as in Section 3.4 to prove that the logarithmic density exists since the race is not "almost periodic".

Chebyshev's race $\pi(x;4,3)-\pi(x;4,1)$ can be written as a sum of a Dirichlet character over primes as we saw in (3.2). As we discussed in Section 2.2, the Dirichlet series $\sum_p \chi_1(p)p^{-s}$ is not easy to use. However, we used instead $L(s,\chi_1) = \sum_{n\geqslant 1} \Lambda(n)\chi_1(n)n^{-s}$ which was useful to obtain the trigonometric representation of $E(e^u;4,3,1)$ in Theorem 3.1. With Perron's

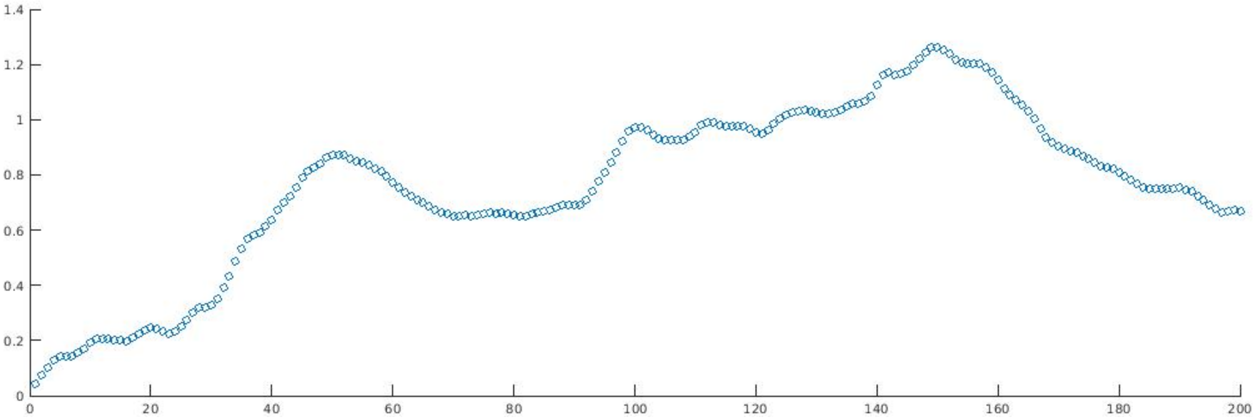

**Fig. 7.3.** Second moment of $f(10^{U+x/100})-f(10^U)$ for a random variable $U$ which is uniform in $[0,10^6]$ and $f(x)=\frac{\log x}{\sqrt{x}}(\pi_+(x)-\pi_-(x))$.

formula, we arrived at an explicit formula for $\psi(x,\chi_1)$ in Theorem 2.16 in terms of the zeros of $L(s,\chi_1)$. To have this explicit formula, we needed an analytic continuation for $L(s,\chi_1)$ over the whole complex plane and the functional equation 2.20 provided that. However, what made finding the distribution of $\pi_+(x)$ hard is that there is no Dirichlet series representing the Sophie Germain primes at our disposal that is meromorphic on the whole complex plane like $L$-functions are.

As we saw in 3.6, the race between the quadratic residues and the quadratic nonresidues mod $q$ for a large prime $q$ is approximately normally distributed with a larger variance than Chebyshev's race. We also saw that the bias dissipates as $q\to\infty$ over primes. These are consequences of the fact that the nontrivial zeros of $L(s,\chi_q)$ where $\chi_q$ is the Legendre symbol mod $q$ become denser as $q$ gets larger. Since denser zeros means that the amplitudes and frequencies are closer together, then we obtain a race which behaves randomly. In fact, it is likely that the expression $\frac{\log x}{\sqrt{x}}(\pi_+(x)-\pi_-(x))$ can be written as a superposition of many waves of similar amplitudes. János Pintz has a method of finding an explicit formula that points to this direction in [23]. This is why we can suspect that $\frac{\log x}{\sqrt{x}}(\pi_+(x)-\pi_-(x))$ behaves like a random walk.

To model the Sophie Germain prime number race, we take inspiration from Cramér's model. For every prime, let $X_p$ be a sequence of independent discrete random variable supported in $\{-1,0,1\}$ such that $\mathbb{P}(X_p=-1)=\mathbb{P}(X_p=1)=1/\log(2p)$. We can also define

$$S(x):=\frac{\log x}{\sqrt{x}}\sum_{p\leqslant x}X_p.$$

Figure 7.4 presents eight different outcomes of $S(x)$. As we can see, Figure 7.1 could fit in with the graphs of Figure 7.4 more than the structured "almost periodic" graph in Figure 3.1, and the Sophie Germain race seems more like a "typical outcome" of the probability space. In Figure 7.5, we also have the graphs of the second moments of the different translations of

$S(x)$ over the logarithmic scale. These graphs look more are not as oscillatory as the graph for Chebyshev's race in Figure 7.2. The Sophie Germain race looks more like a random walk than the Chebyshev prime number race.

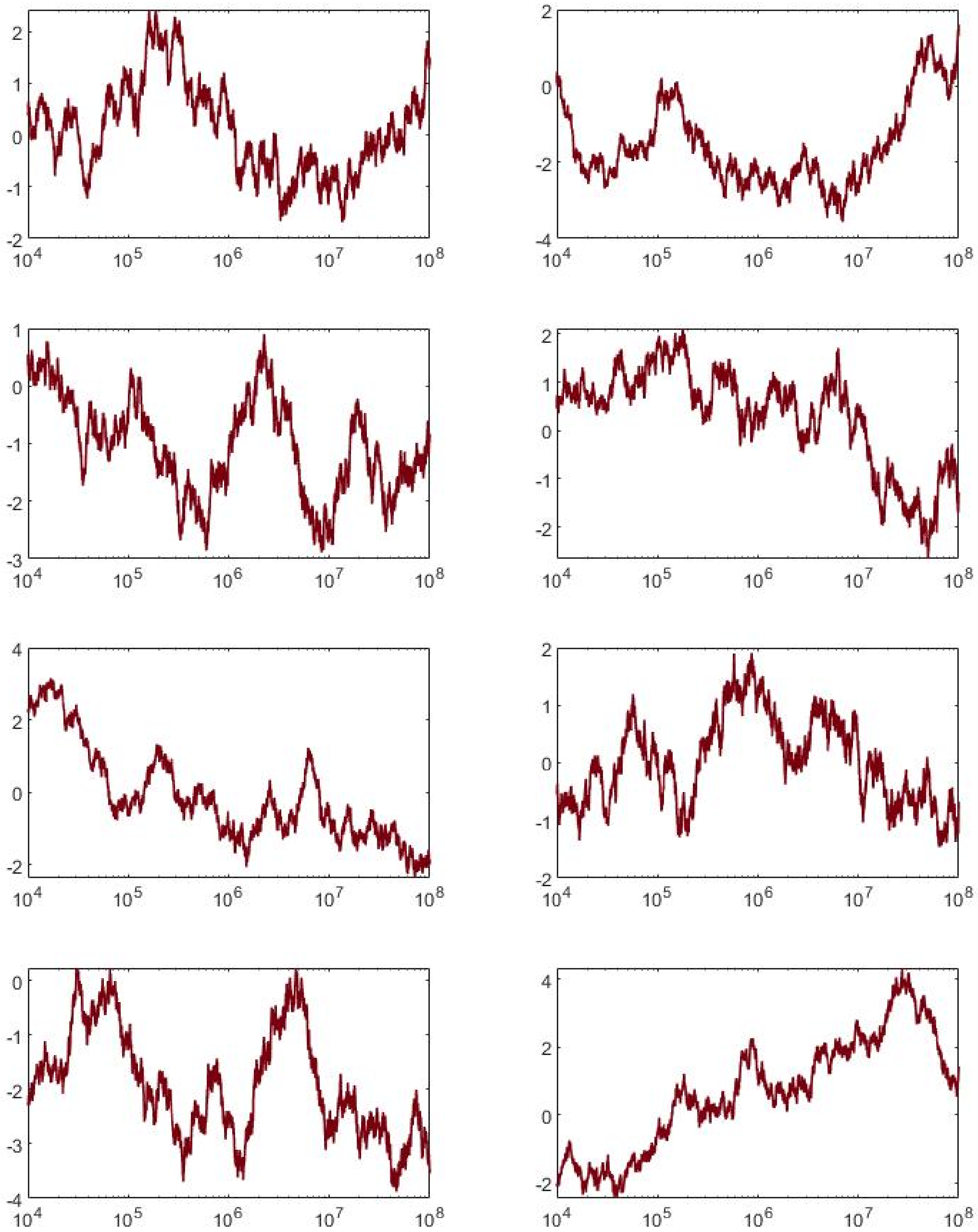

**Fig. 7.4.** Different outcomes of the random walk $S(x)$.

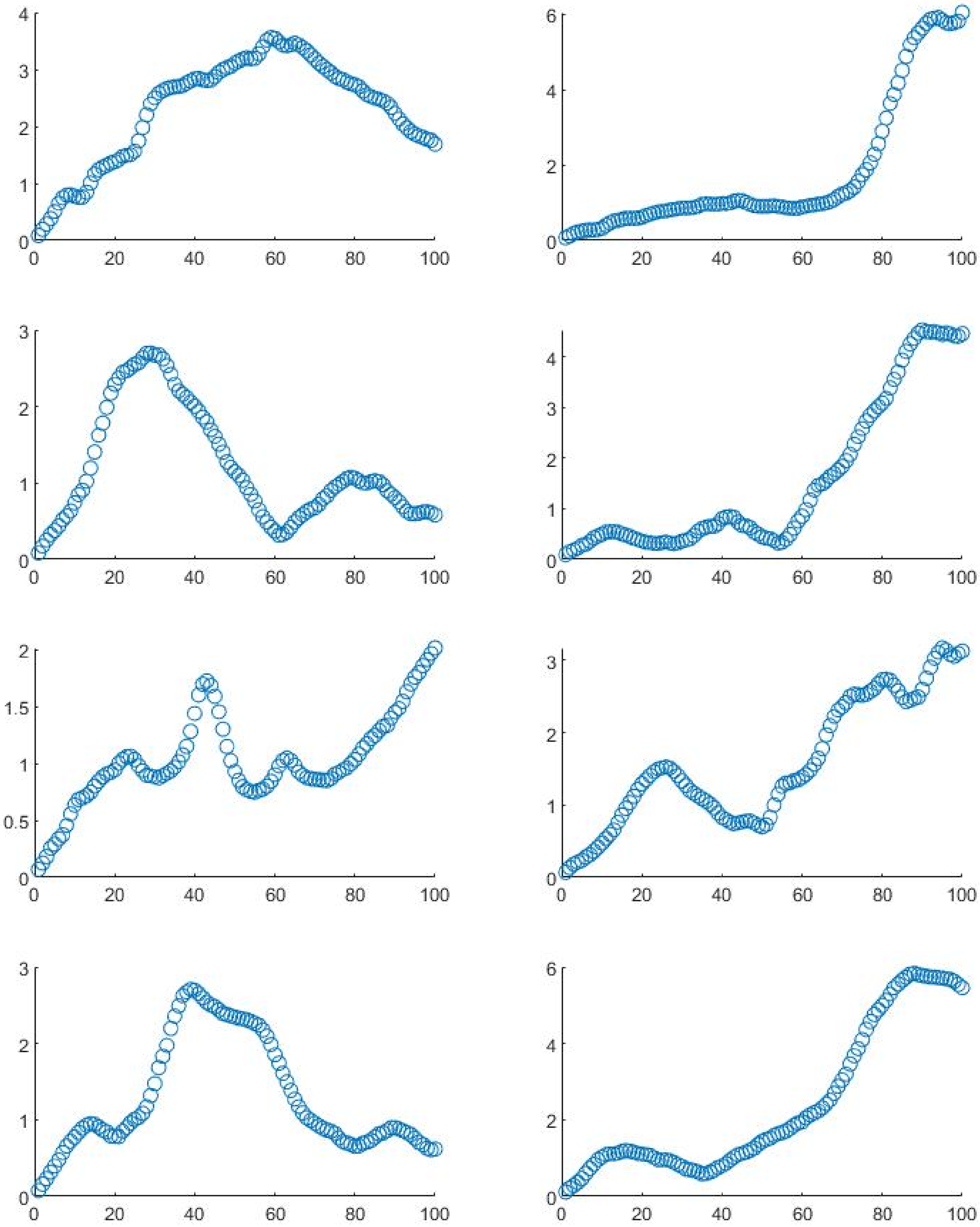

**Fig. 7.5.** Second moment of $S(10^{U+x/100}) - S(10^U)$ for a random variable $U$ which is uniform in $[0, 10^6]$ where $S(x)$ is the outcome in the same position in Figure 7.4.

# Chapter 8

# Conclusion

This thesis investigated the differences between the structured mod 4 prime number race and the more random race between two kinds of Sophie Germain primes.

We established an explicit formula for different prime counting functions using the theory of $L$-functions, and we proved the Prime Number Theorem for arithmetic progressions. Using this explicit formula, we approximated the prime number race with a sum of trigonometric functions, and we established that the race is biased to the primes 3 mod 4 because they are quadratic nonresidues. We discussed that we could measure this bias using the logarithmic density.

We also wanted to establish a conjecture (Conjecture 1.1) on the distribution of Sophie Germain prime counting function:

$$\pi_+(x) \sim \frac{c_2 x}{\log^2 x}$$

as $x \to \infty$. We gave an upper bound on $\pi_+(x)$ with the same order of magnitude as the conjecture using Selberg's sieve method. We presented Granville's refinement of Cramér's model for primes and used it to find that we should expect the conjecture to be true. We also used the Hardy-Littlewood circle method to find an asymptotic on the major arcs of the integral representing the function $\psi_+(x)$ which matches the conjecture, and we also proved unconditionally that there exists infinitely many primes of the form $p+q+1$ where $p, q$ are primes. In Figure 5.1, we show that the data supports Conjecture 1.1. These techniques are efficient to establish conjectures on the distribution of subsets of primes defined by a linear equation (like the Sophie Germain primes or the twin primes), even if we have never proved if these subsets are infinite.

We finally studied the Sophie Germain prime number race in the same way we studied the mod 4 race. We established that, even though $2p+1$ can never be a square, but we can use the Cramér-Granville model to predict that there are infinitely many squares of the form $2p-1$, it does not seem like this fact have any contribution to generate a bias, otherwise we would have seen that the graph in Figure 7.1 would have a mean value which is over $0.9707\ldots$, which is not the case. We also studied the periodicity of this graph and compared it to the periodicity of random walks and the mod 4 race to conclude that the Sophie Germain prime race looks more like a random walk.

# Index

# References

[1] M. Agrawal, N. Kayal, and N. Saxena. PRIMES is in P. *Ann. of Math. (2)*, 160(2):781–793, 2004.

[2] L. V. Ahlfors. *Complex analysis.* McGraw-Hill Book Co., New York, third edition, 1978. An introduction to the theory of analytic functions of one complex variable, International Series in Pure and Applied Mathematics.

[3] A. Akbary, N. Ng, and M. Shahabi. Limiting distributions of the classical error terms of prime number theory. *Q. J. Math.*, 65(3):743–780, 2014.

[4] C. Bays and R. H. Hudson. A new bound for the smallest $x$ with $\pi(x) > \mathrm{li}(x)$. *Math. Comp.*, 69(231):1285–1296, 2000.

[5] H. Davenport. *Multiplicative number theory*, volume 74 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, third edition, 2000. Revised and with a preface by Hugh L. Montgomery.

[6] D. Fiorilli. Highly biased prime number races. *Algebra Number Theory*, 8(7):1733–1767, 2014.

[7] D. Fiorilli and G. Martin. Inequities in the Shanks-Rényi prime number race: an asymptotic formula for the densities. *J. Reine Angew. Math.*, 676:121–212, 2013.

[8] G. B. Folland. *Fourier analysis and its applications.* The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1992.

[9] J. Friedlander and H. Iwaniec. *Opera de cribro*, volume 57 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2010.

[10] A. Granville. Harald Cramér and the distribution of prime numbers. *Scand. Actuar. J.*, (1):12–28, 1995. Harald Cramér Symposium (Stockholm, 1993).

[11] A. Granville and G. Martin. Prime number races. *Amer. Math. Monthly*, 113(1):1–33, 2006.

[12] E. Hlawka. *The theory of uniform distribution.* A B Academic Publishers, Berkhamsted, 1984. With a foreword by S. K. Zaremba, Translated from the German by Henry Orde.

[13] J. Kaczorowski. On the Shanks-Rényi race problem. *Acta Arith.*, 74(1):31–46, 1996.

[14] A. Klenke. *Probability theory*. Universitext. Springer, London, second edition, 2014. A comprehensive course.

[15] S. Knapowski and P. Turán. Comparative prime-number theory. I. Introduction. *Acta Math. Acad. Sci. Hungar.*, 13:299–314, 1962.

[16] D. Koukoulopoulos. *The distribution of prime numbers*, volume 203 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, [2019] ©2019.

[17] S. G. Krantz and H. R. Parks. *A primer of real analytic functions*, volume 4 of *Basler Lehrbücher [Basel Textbooks]*. Birkhäuser Verlag, Basel, 1992.

[18] J. E. Littlewood. On the Class-Number of the Corpus $P(\surd - k)$. *Proc. London Math. Soc. (2)*, 27(5):358–372, 1928.

[19] W. Mendenhall, R. J. Beaver, B. M. Beaver, and S. E. Ahmed. *Introduction to Probability and Statistics*. Nelson Education Ltd., third canadian edition, 2014.

[20] H. L. Montgomery and R. C. Vaughan. *Multiplicative number theory. I. Classical theory*, volume 97 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2007.

[21] M. B. Nathanson. *Additive number theory*, volume 164 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1996. The classical bases.

[22] I. Niven, H. S. Zuckerman, and H. L. Montgomery. *An introduction to the theory of numbers*. John Wiley & Sons, Inc., New York, fifth edition, 1991.

[23] J. Pintz. A new explicit formula in the additive theory of primes with applications i. the explicit formula for the goldbach and generalized twin prime problems, 2018.

[24] D. Platt and T. Trudgian. The riemann hypothesis is true up to $3 \cdot 10^{12}$, 2020.

[25] J. M. Pollard. Theorems on factorization and primality testing. *Proc. Cambridge Philos. Soc.*, 76:521–528, 1974.

[26] B. Riemann. *Collected papers*. Kendrick Press, Heber City, UT, 2004. Translated from the 1892 German edition by Roger Baker, Charles Christenson and Henry Orde.

[27] M. Rubinstein and P. Sarnak. Chebyshev's bias. *Experiment. Math.*, 3(3):173–197, 1994.

[28] A. Selberg. On an elementary method in the theory of primes. *Norske Vid. Selsk. Forh., Trondhjem*, 19(18):64–67, 1947.

[29] The LMFDB Collaboration. The L-functions and Modular Forms Database, $L(\chi,s)$, where $\chi$ is the Dirichlet character with label 4.3. `https://beta.lmfdb.org/L/Character/Dirichlet/4/3/`, 2020. [Online; accessed 21 August 2020].