

UNIVERSITY OF HELSINKI

MASTER'S THESIS

On Comparative Prime Number Theory

Author:
Joni Teräväinen

Supervisor:
Dr. Anne-Maria
Ernvall-Hytönen

July 30, 2014

Contents

1	Introduction	2
1.1	History and Motivation	2
1.2	Rubinstein and Sarnak's Paper and Recent Work	6
1.3	Aknowledgements	8
2	Explicit Formula and Its Corollaries	8
2.1	Functional Equation of the L -functions	8
2.2	Hadamard Products for Entire Functions	13
2.3	Number of Zeros of L -functions	15
2.4	Explicit Formula	16
2.5	Generalized Riemann Hypothesis	22
2.6	Grand Simplicity Hypothesis	24
3	Lemmas from Measure Theory and Probability Theory	25
3.1	Measure Theory	25
3.2	Probability Theory	29
3.3	Fourier transform	30
3.4	Bessel function	31
4	The Prime Races	32
4.1	Definitions and the Statement of the Main Theorem	32
4.2	Outline of the Proof	36
4.3	Existence of the Limiting Distribution	38
4.4	A formula for the Fourier transform of the Measure of the Prime Race	50
4.5	Comparison of Densities	55
4.6	Vanishing of the Bias as $q \rightarrow \infty$	62
4.7	Symmetries of the Prime Races	64
4.8	Asymptotics of the Measures of the Prime Races	65

1 Introduction

One of the most celebrated results in analytic number theory is the prime number theorem $\pi(x) \sim \text{Li}(x) := \int_2^x \frac{dt}{\log t}$ for the number of primes up to x . As Gauss already noticed, this should be a very good approximation, and it says that n should be prime "with probability $\frac{1}{\log n}$ ". The prime number theorem for arithmetic progressions is the generalization $\pi(x; q, a) \sim \frac{1}{\varphi(q)} \text{Li}(x)$, where $\pi(x; q, a)$ counts the primes up to x in an arithmetic progression $qn + a$ with q and a coprime. Bounding the difference between $\pi(x; q, a)$ and its asymptotic approximation is a central question in the theory of primes, and the famous generalized Riemann hypothesis predicts that $\pi(x; q, a)$ deviates from $\frac{1}{\varphi(q)} \text{Li}(x)$ by at most a constant times $\sqrt{x} \log x$, which is essentially the best that can be hoped for.

It is also important to understand the biases in the distribution of primes into residue classes, and this is what comparative prime number theory studies. In this thesis, we give a detailed exposition of Rubinstein and Sarnak's breakthrough paper from 1994 on comparative number theory. Among other things, Rubinstein and Sarnak assigned a (logarithmic) density to the positive integers m for which $\pi(m; q, a) > \pi(m; q, b)$, showed that it is always positive, and gave a simple criterion for determining which of the progressions $qn + a$ and $qn + b$ leads this "prime race" more often than the other. The work assumes the generalized Riemann hypothesis and the linear independence of the imaginary parts of the nontrivial zeros of L -functions over rational numbers.

1.1 History and Motivation

One of the simplest objects of study in comparative number theory is the function $\pi(x) - \text{Li}(x)$. It turns out to be negative for all x below a huge bound, which is possibly around 10^{300} , but Littlewood showed in 1914 that $\pi(x) - \text{Li}(x)$ is greater than $\frac{\sqrt{x}}{\log x}$ (and similarly smaller) for infinitely many integers x [2],[17].

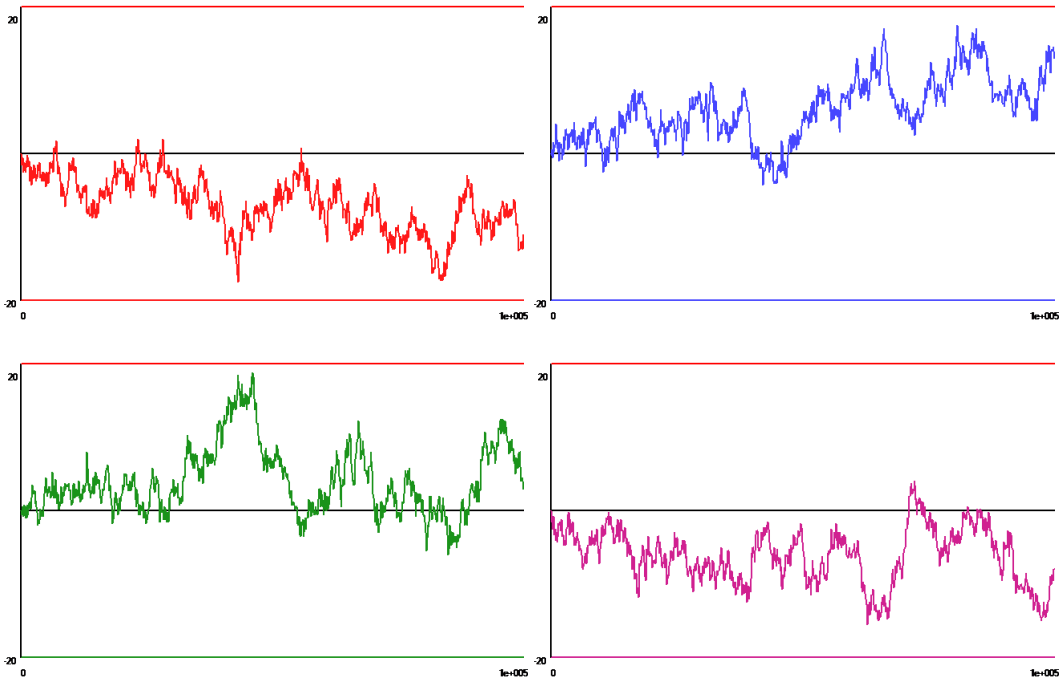
Nevertheless, one could suspect that the distribution of primes into two different arithmetic progressions with the same difference is very even, but surprisingly one sees a similar lack of uniformness of distribution. For the primes of the form $4n \pm 1$, this phenomenon is Chebyshev's bias, noticed by Chebyshev in 1853, and it is a special case of the "prime races" studied in this thesis. Chebyshev noticed that the progression $4n - 1$ tends to have more

primes than $4n + 1$, and similarly the primes of the form $3n - 1$ are usually in the lead over the primes $3n + 1$,¹, when counted up to a limit. Some data is presented below to make this surprising difference more concrete (computed with pari/gp).

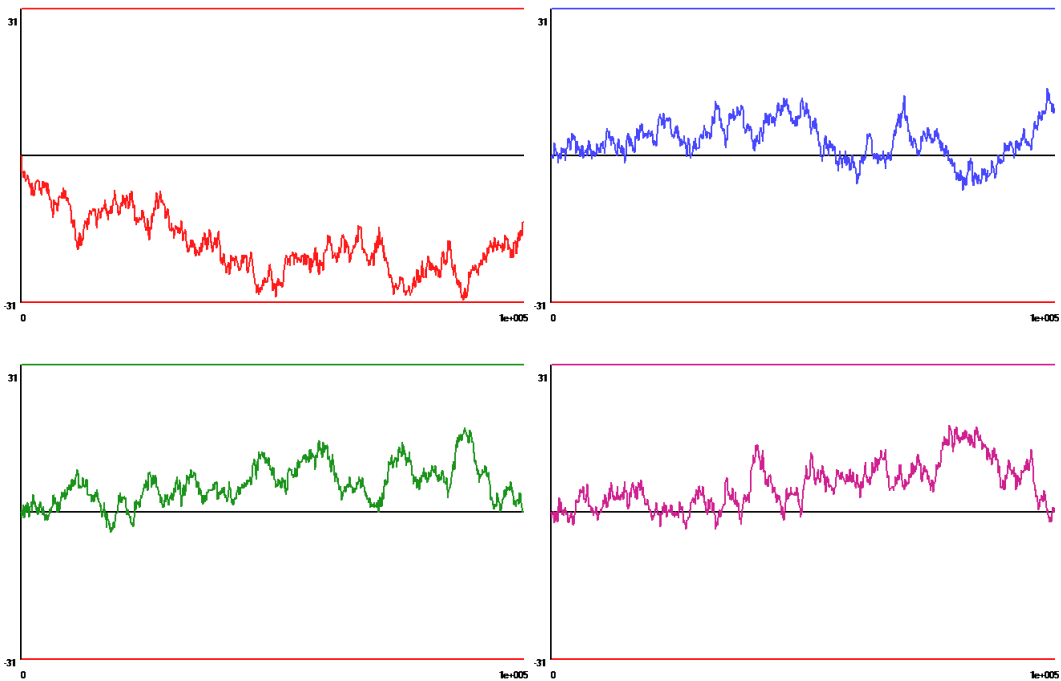
x	$\pi(x; 3, -1)$	$\pi(x; 3, 1)$	$\pi(x; 4, -1)$	$\pi(x; 4, 1)$
100	13	11	13	11
1000	87	80	87	80
10000	617	611	619	609
20000	1137	1124	1136	1125
30000	1634	1610	1633	1611
100000	4807	4784	4808	4783
200000	8995	8988	9006	8977
300000	13026	12970	13016	12960
1000000	39266	39231	39322	39175
10000000	332384	332194	332398	332180
50000000	1500653	1500480	1500681	1500452

According to the table, the primes $p \equiv -1 \pmod{3}$ and $p \equiv -1 \pmod{4}$ seem to always hold the lead. Actually, there are some points at which the primes $p \equiv 1 \pmod{3}$ are more numerous (the smallest one being 608981813209), but they are very rare [9]. For the primes $p \equiv 1 \pmod{4}$, there are also some points (such as 26861) at which they are ahead, but again the intervals are short and sparse. This is an interesting phenomenon, and one does not directly see any reason for the symmetry to break down. However, if one also studies other prime races, for instance races modulo 5 and 8, a pattern starts to form. By looking at the following plots, it seems that the progressions $5n \pm 1$ have less primes than $5n \pm 2$, and $8n + 1$ has less primes than the others $\pmod{8}$, which have roughly an equal number of them.

¹ Chebyshev did not mention the primes $3n \pm 1$ but he conjectured the unevenness in the distribution of the $4n - 1$ and $4n + 1$ primes in an interesting way: $\sum_p (-1)^{\frac{p-1}{2}} x^p$ should tend to $-\infty$ as $x \rightarrow 1$ [20].



Plots of $\pi(x; 5, a) - \frac{1}{4}\pi(x)$ for $a = 1, 2, 3, 4$ (in that order), in the range $x \in [0, 10^5], y \in [-20, 20]$.



Plots of $\pi(x; 8, a) - \frac{1}{4}\pi(x)$ for $a = 1, 3, 5, 7$ (in that order), in the range $x \in [0, 10^5], y \in [-31, 31]$.

$x \in [0, 10^5], y \in [-31, 31]$.

Remarkably, 1 and -1 are precisely the quadratic residues (mod 5), and similarly 1 is the only odd square (mod 8). This is not a coincidence, but explaining the bias towards the nonresidues rigorously is a difficult task. It was already shown by Littlewood [17] that both the primes $3n+1$ and $3n-1$ (and similarly modulo 4) hold the lead infinitely many times, but this says nothing about the relative frequencies of these events. Knapowski and Turán [15] formulated in their series of eight papers on comparative number theory many questions about the discrepancy of the number of primes in different residue classes, and obtained partial results, but many questions were left unanswered. In order to address these prime races more carefully, we must first quantify what it means for one sequence of primes to lead over the other "more of the time". This actually leads to measure theory, thus giving an interesting and a bit unexpected connection.

One way to measure the proportion of integers in a set A (which is to us the set of integers $x > 0$ for which $\pi(x; q, a) > \pi(x; q, b)$, denoted by $P_{q;a,b}$)² is the asymptotic density

$$d(A) := \lim_{x \rightarrow \infty} \frac{\#[[1, x] \cap A]}{x},$$

if the density exists ($\#B$ is the number of elements in B). This is of course the simplest interesting density for a subset of the positive integers. However, it turns out not to be the right density for our purposes; in fact, Kaczorowski [13] showed that it does not (always) exist for the sets $P_{q;a,b}$ we are interested in. This means in particular that the lower asymptotic density of these sets must be less than 1, giving rise to the question whether the densities of these sets are strictly between zero and one even if the appropriate density is used.

The numerics in the tables and graphs above give us some hint about the correct density. It seems that the prime races do not reveal themselves significantly on the linear scale, but rather on a logarithmic one. For example, the integers up to which the primes $3n+1$ lead over the primes $3n-1$ seem to be logarithmically sparse. This gives reason to suspect that the right way to measure the discrepancy is to use the logarithmic density, written in its

²Also the more general prime races between primes congruent to $a_1, \dots, a_r \pmod{q}$ are considered in this thesis, but for simplicity we concentrate on the case $r = 2$ now.

discrete form as

$$\delta(A) = \lim_{x \rightarrow \infty} \frac{1}{\log x} \sum_{\substack{a \in A \\ a \leq x}} \frac{1}{a},$$

where the "weight" $\frac{1}{a}$ has been assigned to an integer a . One of the reasons for using this density is that the excess of the $3n - 1$ primes in the logarithmic scale (meaning that the spacing on the x -axis grows exponentially) is $\frac{x}{e^{\frac{x}{2}}}(\pi(e^x; 3, -1) - \pi(e^x; 3, 1))$, which has an explicit formula involving trigonometric sums over the zeros of Dirichlet L -functions. Remarkably, we can control this quantity knowing only an estimate for the number of zeros of the L -functions in a given region and assuming that they have real part $\frac{1}{2}$ (unless they are trivial) and are not related to each other "in a too regular manner".

1.2 Rubinstein and Sarnak's Paper and Recent Work

We gave some heuristics above, but Rubinstein and Sarnak [28] were able to prove in 1994 the following beautiful theorem (and its generalization), assuming two standard hypotheses:

Theorem. *Assume the generalized Riemann hypothesis and the grand simplicity hypothesis. If a and b are positive integers, coprime to an integer $q \geq 1$, the set $P_{q;a,b} := \{x > 1 : \pi(x; q, a) > \pi(x; q, b)\}$ has a positive logarithmic density, and it satisfies $\delta(P_{q;a,b}) > \delta(P_{q;b,a})$ if and only if a is a quadratic nonresidue (mod q) and b is not. Moreover, if $\eta(x)$ is any function such that $\eta(x)$ tends to infinity as $x \rightarrow \infty$, the set $\{\pi(x; q, a) - \pi(x; q, b) > \frac{\sqrt{x}}{\eta(x) \log x}\}$ has the same logarithmic density as $P_{q;a,b}$ has.*

We will prove this and some other theorems from the paper [28] in this thesis, completing practically all the details. The Generalized Riemann hypothesis (GRH) asserts that all the nontrivial zeros of the Dirichlet L -functions have real part $\frac{1}{2}$. The other assumption, the grand simplicity hypothesis (GSH), says that the imaginary parts of all the nontrivial zeros of the L -functions (in the upper half-plane) are linearly independent over the rationals. Very roughly, the GRH allows us to use the explicit formula, established in Chapter 2, for the number of primes in a convenient way, while the GSH assures that we have some control over this explicit sum over the zeros of the L -functions.

In Chapter 4, we give an outline of the proof of the Main Theorem (which is a generalization of the theorem above to prime races between several residue classes and is also based on the paper [28]) of this thesis before proving it, so we say here only a couple of words about the proof. The GRH is employed to show that the normalized excess in the number of primes $a \pmod{q}$ compared to the primes $b \pmod{q}$ has a limiting distribution, or measure, $\mu_{a,b}$. The GSH tells that the Fourier transform $\hat{\mu}_{a,b}$ has an explicit formula, from which we can analyze its properties, and hence properties of $\mu_{a,b}$ itself. Questions about the logarithmic densities are reduced to questions about the symmetry of $\mu_{a,b}$. The formula also shows that the biases in the prime races diminish as $q \rightarrow \infty$.

Rubinstein and Sarnak [28] also used the formula for $\hat{\mu}_{a,b}$ to compute (assuming the GRH and the GSH) logarithmic densities of several sets $P_{q;a,b}$ and other prime races; in particular $\delta(P_{3,-1}) = 0.9990\dots$; $\delta(P_{4,-1}) = 0.9959\dots$. These very large densities arise from the facts that the smallest zeros of the corresponding L -functions are rather large in terms of imaginary part, but we will not deal with computational aspects here.

This is as far as we will go in this thesis, but many more questions arise, some of which have been answered, while others remain unsettled. A good survey of conjectures and theorems is [20]. We mention just one open problem: *When exactly is the prime race between the progressions $qn + a_1, \dots, qn + a_r$ even in the sense that each of the $r!$ possible orders has $\frac{1}{r!}$ as its logarithmic density?* These races for $r \geq 3$ are known as Renyi-Shanks races. Rubinstein and Sarnak determined the cases when this race is unbiased so that its distribution function is symmetric with respect to permutations, but this is a stronger condition than evenness of the race.

It is also worth mentioning that the prime races have been generalized in many ways; for example, Nathan Ng [26] considered in 2000 the prime races between the prime divisors of polynomials. It is known for example that the set of primes p that divide $x^3 + 2$ for some x has density $\frac{2}{3}$, and one may then consider the difference between the number of these primes and twice the number of other primes.

Very little is known about the prime races without one or both of the deep, but widely believed, conjectures GRH and GSH. There are actually reasons to believe that most results about the prime races could not be achieved unconditionally; Ford and Konyvagin [7] showed that if the GRH is false in a suitable way, then in any prime race between three progressions at least one

order occurs just finitely many times.

1.3 Acknowledgements

I am grateful to my supervisor Anne-Maria Ernvall-Hytönen for guidance and careful reading of numerous versions of the text and for helpful comments. I am also grateful to Esa Vesalainen, who is the second reader, for giving valuable feedback. I also want to mention that they were willing to check my thesis during their summer holidays. I also thank Jesse Jääsaari for voluntary reading of the entire manuscript and for giving useful comments. I was granted a scholarship by the Mathematics and Sciences Fund of the University of Helsinki while I was working on this thesis.

2 Explicit Formula and Its Corollaries

2.1 Functional Equation of the L -functions

Let χ be a Dirichlet character (mod q) for a positive integer $q \geq 1$; that is, $\chi : \mathbb{Z} \rightarrow \mathbb{C}$ satisfies $\chi(ab) = \chi(a)\chi(b)$ for all $a, b \in \mathbb{Z}$, $\chi(n) = 0$ if and only if n and q are not coprime, and χ has period q . We associate with χ the Dirichlet L -function defined by

$$L(s, \chi) := \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s} \quad (1)$$

for $\Re(s) > 1$. This is a Dirichlet series whose coefficients are $\chi(n)$, and it turns out to play a fundamental role in the study of the distribution of prime numbers (mod q), similar to the role of the Riemann zeta function $\zeta(s)$ in estimating the number of primes up to x , denoted by $\pi(x)$.³ We recall a few basic facts about the functions $L(s, \chi)$ before discussing the functional equation; the proofs of these properties can be found for example in Karatsuba's book [14] (pages 10, 64-69). More generally, everything in Chapter 2 can be found in one or more of the books by Davenport (pp. 59-102, 115-120) [5], Ingham (pp. 68-85) [11], Karatsuba (pp. 10, 102-124) [14], Murty (pp. 309-310, 331-370) [24].

The series defining a Dirichlet L -function converges absolutely and uniformly on compact subsets of $\{s : \Re(s) > 1\}$ since the integral $\int_1^\infty x^{-\sigma} dx$ converges

³In fact, $\zeta(s)$ is the L -function corresponding to the principal character (mod q), up to the factor $(1 - \frac{1}{p_1^s})^{-1} \dots (1 - \frac{1}{p_k^s})^{-1}$, where p_i are the primes not coprime to q .

for $\sigma > 1$ (we write $s = \sigma + it$ according to Riemann's classical notation). The L -function is therefore analytic in this region. When $\chi \neq \chi_0$ – the principal character satisfying $\chi(a) = 1$ for a and q coprime – we can actually say much more. Then it is easily seen that $\sum_{n=1}^q \chi(n) = 0$, so the sum $\sum_{n \leq x} \chi(n)$ is bounded in modulus by a constant, and thus the series defining $L(s, \chi)$ converges uniformly on compact subsets of $\{s : \Re(s) > 0\}$ by the standard partial summation formula

$$\sum_{n \leq x} a_n f(n) = A(x) f(x) - \int_1^x A(t) f'(t) dt, \quad (2)$$

where $A(t) = \sum_{n \leq t} a_n$, the summation is over the positive integers on $[1, x]$, and f is continuously differentiable. To see this, take $a_n = \chi(n)$ and $f(t) = t^{-s}$; this gives as $x \rightarrow \infty$

$$L(s, \chi) = \int_1^\infty \frac{\sum_{n \leq x} \chi(n)}{x^{s+1}} dx \ll 1, \quad \sigma \geq \sigma_0 > 1.$$

The partial summation formula will be applied frequently later on in this thesis. Due to locally uniform convergence, the series of an L -function defines an analytic function in the half-plane of positive real part.

In order to study the distribution of primes into residue classes with the help of the L -functions, we define the Chebyshev function corresponding to χ ;

$$\psi(x, \chi) := \sum_{n \leq x} \chi(n) \Lambda(n), \quad (3)$$

with $\Lambda(n) = \log p$ for $n = p^\alpha$ a prime power and 0 otherwise. The size of $\psi(x, \chi)$ for all $\chi \pmod{q}$ is very intimately related to the size of $\pi(x; q, a)$, the number of primes up to x that are congruent to $a \pmod{q}$ with a and q coprime, but the functions $\psi(x, \chi)$ are easier to deal with.

Let us also define the functions

$$\psi(x; q, a) := \sum_{\substack{n \leq x \\ n \equiv a \pmod{q}}} \Lambda(n) = \frac{1}{\varphi(q)} \sum_{\chi \pmod{q}} \bar{\chi}(a) \psi(x, \chi),$$

which are related to $\pi(x; q, a)$ by

$$\pi(x; q, a) = \sum_{n \leq x} \frac{1_{\mathbb{P}(q; a)}(n) \log n}{\log n} = \frac{\psi(x; q, a)}{\log x} + \int_2^x \frac{\psi(t; q, a)}{t \log^2 t} dt + O(\sqrt{x} \log x),$$

where $\mathbb{P}(q; a)$ is the set of primes congruent to $a \pmod{q}$, and we used partial summation with $a_n = 1_{\mathbb{P}(q; a)}(n)$, $f(t) = \frac{1}{\log t}$ and the fact that $\sum_{\substack{p \leq x \\ p \equiv a \pmod{q}}} \log p$

differs from $\psi(x; q, a)$ by at most $O(\sqrt{x} \log x)$.⁴ The study of $\pi(x; q, a)$ is therefore reduced to that of $\psi(x; q, a)$, and in particular the assumption $\psi(x; q, a) \sim \frac{x}{\varphi(q)}$ implies, by using the previous formula and applying partial integration to 1 and $\frac{1}{\log^2 t}$, that

$$\pi(x; q, a) \sim \frac{x}{\varphi(q) \log x} + \frac{1}{\varphi(q)} \int_2^x \frac{1}{\log^2 t} dt \sim \frac{1}{\varphi(q)} \int_2^x \frac{dt}{\log t} := \frac{1}{\varphi(q)} \text{Li}(x),$$

which actually proves to be a much better estimate than the approximation $\frac{x}{\varphi(q) \log x}$, although they are asymptotically equal.

As mentioned earlier, Gauss claimed that $\int_2^x \frac{dt}{\log t}$ should give a good estimate for the number of primes up to x on probabilistic grounds. Riemann confirmed Gauss' prediction assuming some properties of his zeta function ζ , but proving those properties was left for Hadamard, de la Vallée Poussin, and others. For more history, see [25].

The objective of Chapter 2 is to derive the following explicit formula, which will be applied to prove the Main Theorem of Chapter 4.

Theorem 2.1. (Explicit formula) *Let q be fixed, χ a primitive character $(\text{mod } q)$,⁵ and $T \geq 1$. If χ is an odd character (i.e. $\chi(-1) = -1$) and $x - \frac{1}{2}$ is an integer greater than 1, we have*

$$\psi(x, \chi) = - \sum_{|\rho| < T} \frac{x^\rho}{\rho} - \frac{L'(0, \chi)}{L(0, \chi)} + \sum_{m=1}^{\infty} \frac{x^{1-2m}}{2m-1} + O\left(\frac{x \log^2(xT)}{T}\right), \quad (4)$$

where the sum is taken over all the zeros of $L(s, \chi)$ in the region $\{s : 0 < \Re(s) \leq 1\}$ with absolute value less than T . In the case where χ is even (i.e. $\chi(-1) = 1$) and non-principal, and $x - \frac{1}{2}$ is an integer greater than 1, we have

$$\psi(x, \chi) = - \sum_{|\rho| < T} \frac{x^\rho}{\rho} - \log x - b(\chi) + \sum_{m=1}^{\infty} \frac{x^{-2m}}{2m} + O\left(\frac{x \log^2(xT)}{T}\right), \quad (5)$$

⁴The last fact follows just from the upper bound $\sqrt{x} + \sqrt[3]{x} + \dots + \sqrt[k]{x} \ll \sqrt{x} + (\log x) \sqrt[3]{x}$ for the number of perfect powers up to x , where $k = \lfloor \log_2 x \rfloor$.

⁵That is, it is not induced in the natural way by any character of smaller modulus, meaning that $\chi(n) = \chi'(n)$ whenever $(n, q) = 1$ for some character χ' with modulus $q' < q$.

where $b(\chi)$ is a constant [5].

In both cases, letting $T \rightarrow \infty$ we immediately obtain

Corollary 2.2. *With the notations of the previous theorem (in particular, x is half more than a positive integer),*

$$\psi(x, \chi) = - \sum_{\rho} \frac{x^{\rho}}{\rho} - (1-a) \log x - b(\chi) + \sum_{m=1}^{\infty} \frac{x^{a-2m}}{2m-a}, \quad (6)$$

where a equals 1 for an odd character χ and 0 for an even one, $b(\chi)$ is a constant, and the sum runs over the zeros of $L(s, \chi)$ in a symmetric way (that is, we are summing over the zeros ρ with $|\rho| < T$ and taking the limit).

Notice that all the terms except the sum over the zeros are rather irrelevant; the last sum is bounded by a constant by comparison with a geometric series, so everything except the main term is $O(\log x)$.

Although the formulas (4) and (5) are much more practical than the corollary, since they lack infinite sums over zeros, (6) is strictly speaking the explicit formula as it has no error term.

A few remarks about the explicit formula are in order. If x was allowed to be an arbitrary positive real number in (4) and (5), the formulas would not be valid, since the error term in these formulas actually depends on the inverse of the fractional part of x . In addition, the order of summation is relevant in (6) to assure the convergence of the series. In our applications a suitable choice of T in the explicit formula is of the order \sqrt{x} since then we get

$$\psi(x, \chi) = - \sum_{|\rho| < \sqrt{x}} \frac{x^{\rho}}{\rho} + O(\sqrt{x} \log^2 x),$$

where the error term is essentially as good as we want ⁶ but the sum over the zeros is still rather short. This is the setting which we apply at the end of the chapter to deduce $|\psi(x, \chi)| \ll \sqrt{x} \log^2 x$ from the generalized Riemann hypothesis.

To derive the explicit formula, we will establish a product formula for $L(s, \chi)$, or actually for a close relative of it. For this, we first need to extend $L(s, \chi)$ to an analytic function defined in the whole complex plane and to derive a functional equation for it.

⁶Even the assumption of the generalized Riemann hypothesis leaves the possibility that intervals of the form $[x, x + \sqrt{x}]$ might not contain prime powers (although they should contain); see [35].

Theorem 2.3. *Let χ be a primitive character (mod q) for $q > 1$, and let $a = 0$ if χ is even and $a = 1$ otherwise. Let*

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

Then ξ extends to an entire function satisfying the functional equation

$$\frac{i^a \sqrt{q}}{\tau(\chi)} \xi(s, \chi) = \xi(1-s, \bar{\chi}) \quad (7)$$

where for a character λ (mod q)

$$\tau(\lambda) := \sum_{m=1}^q \lambda(m) e^{\frac{2\pi im}{q}}$$

is a Gauss sum.

A proof of the functional equation that yields the analytic continuation follows the ideas already presented in Riemann's famous memoir of 1860 for deriving the functional equation of the ζ function [27], which was extended to the case of L -functions by de la Vallée Poussin.

The proof exploits the properties of the Jacobi theta function, but for the sake of brevity, we omit it.

From the functional equation it is evident in which region the zeros of $L(s, \chi)$ must lie. By the Euler product

$$L(s, \chi) = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}, \quad (8)$$

whose proof for $\sigma > 1$ follows easily from the fundamental theorem of arithmetic, one has $L(s, \chi) \neq 0$ for $\Re(s) > 1$ ([14], page 51). Hence also $L(s, \chi) \neq 0$ for $\Re(s) < 0$ unless χ is even and $s = 0, -2, -4, \dots$ or χ is odd and $s = -1, -3, -5, \dots$ since, as already mentioned, the Gamma function is zero-free and has poles precisely at the negative integers and at 0. The search for the zeros of the L -functions can therefore be restricted to the region $\{s : 0 \leq \Re(s) \leq 1\}$. Already the slight improvement that the boundary of this *critical strip* does not contain any zeros is essentially equivalent to the prime number theorem for arithmetic progressions.

2.2 Hadamard Products for Entire Functions

The objective of this subsection is to employ the Hadamard product formula to derive the product expansion

$$\xi(s, \chi) = e^{A(\chi) + B(\chi)s} \prod_{\rho} \left(1 - \frac{s}{\rho}\right) e^{\frac{s}{\rho}}, \quad (9)$$

where ξ is as in the previous subsection, $A(\chi)$ and $B(\chi)$ are constants depending only on χ , and the product runs through all the zeros ρ of $L(s, \chi)$, ordered according to increasing modulus.

After we have established this, by taking the logarithmic derivative, we get a sum identity connecting $\frac{L'(s, \chi)}{L(s, \chi)}$ to the zeros ρ of $L(s, \chi)$. The logarithmic derivative of $L(s, \chi)$ will, on the other hand, be seen to have a direct connection to $\psi(x, \chi)$ via an integral representation, so the results claimed above will lead to the explicit formula.

As the subsection title indicates, the product formula for $\xi(s, \chi)$ is a special case of a more general theorem, due to Hadamard, concerning representations of entire functions as products involving their zeros.

In order to have a suitable product representation for f , we must assume that f does not grow true rapidly, or more precisely that f is an entire function of finite order. This means that there exists a finite number α such that

$$\max_{|s|=R} |f(s)| \leq c_{\alpha} e^{R^{\alpha}}$$

for all R . If a is the infimum of such values of α , we say that a is the *order of f* . We will restrict ourselves to entire functions of order 1, as these are the only ones we need, but the theory presented here is quite similar for functions of any finite order.⁷

We are now ready to formulate the following theorem (see [33], pages 147-153 for a proof).

Theorem 2.4. (Hadamard) *Let f be an entire function of order 1 that has a zero of order $\ell \geq 0$ at 0, and the other zeros of f are a_1, a_2, \dots in increasing*

⁷In the following claim, the factors $\exp(\frac{s}{a_n})$ should be replaced with $\exp(\frac{s}{a_n} + \frac{1}{2} \left(\frac{s}{a_n}\right)^2 + \dots + \frac{1}{m} \left(\frac{s}{a_n}\right)^m)$ where m is the smallest integer not smaller than the order of f .

order by modulus, with multiplicities. Then there exist constants A and B such that

$$f(s) = s^\ell e^{A+Bs} \prod_{n=1}^{\infty} \left(1 - \frac{s}{a_n}\right) e^{\frac{s}{a_n}} \quad (10)$$

for all $s \in \mathbb{C}$.

The proof is omitted as it is quite long, entirely complex analytic and the result is rather well-known. Besides Hadamard's product formula, we will need Jensen's formula in the next subsection. It connects the number of zeros of f inside a disk to the average of $\log f(s)$ on that disk. Jensen's formula is also a useful ingredient in proving Hadamard's formula.

Lemma 2.5. (Jensen's formula). *Let f be an entire function that has no zeros on the circle $|s - z_0| = R$, and suppose $f(z_0) \neq 0$. Let $n(r)$ be the number of zeros of f within $B(z_0, r)$; then*

$$\frac{1}{2\pi} \int_0^{2\pi} \log |f(z_0 + Re^{i\varphi})| d\varphi - \log |f(z_0)| = \int_0^R \frac{n(r)}{r} dr.$$

For a proof, see the book [33] of Stein and Shakarchi, pages 154-156.

We now show that Hadamard's formula (10) leads to the important product representation (9) of $\xi(s, \chi)$ by straightforward estimation. Indeed, $\xi(s, \chi)$ was shown to be entire in the previous section, so we just need to show that it is of order 1. The functional equation $\xi(s, \chi) = \frac{\tau(\chi)}{i^a \sqrt{q}} \xi(1-s, \bar{\chi})$ with $a = 0$ or $a = 1$ allows us to restrict our considerations to the half-plane $\sigma \geq \frac{1}{2}$. From

$$\xi(s, \chi) = \pi^{-\frac{s+a}{2}} q^{\frac{s+a}{2}} \Gamma\left(\frac{s+a}{2}\right) L(s, \chi), \quad a = \begin{cases} 0 & \text{if } \chi \text{ is even} \\ 1 & \text{if } \chi \text{ is odd} \end{cases}$$

we conclude that $\xi(s, \chi)$ is the product of two exponential functions, the term

$$\Gamma\left(\frac{s+a}{2}\right) \leq e^{c_1 |s| \log |s|},$$

(using the Stirling's approximation $\log \Gamma(s) = s \log s - s + \frac{1}{2} \log s + O(1)$ for $|s| \rightarrow \infty$ and $\arg(s)$ bounded away from π ; for a proof see [14], pages 44-45), and the term

$$L(s, \chi) = s \int_1^{\infty} A(x) x^{-s-1} dx,$$

where $A(x) = \sum_{n \leq x} \chi(n)$ by partial summation. This gives $|L(s, \chi)| \leq 2\varphi(q)|s|$, so $|\xi(s, \chi)| \leq e^{C|s| \log s}$, implying that $\xi(s, \chi)$ is of order 1 (since the definition shows that $\xi(s, \chi)$ grows exponentially when s is real and positive). We thus obtained the product formula (9), which turns out to be a key element in proving the explicit formula later on.

2.3 Number of Zeros of L -functions

We formulate a fundamental formula for the zeros of an L -function. This will be used frequently in Chapter 4 while proving the Main Theorem.

Theorem 2.6. *For the number of nontrivial zeros of the function $L(s, \chi)$, with χ a primitive character and the modulus of the zeros not greater than T , we have*

$$N(T) = \frac{T}{\pi} \log \frac{qT}{2\pi} - \frac{T}{\pi} + O(\log T) \quad (11)$$

(The modulus q is assumed to be fixed, so there is no need to indicate dependence on it.)

We first record a few immediate but useful consequences.

Corollary 2.7. *We have*

$$\sum_{0 < |\gamma| < T} \frac{1}{\gamma} = O(\log^2 T) \quad (12)$$

$$\sum_{|\gamma| > T} \frac{1}{\gamma^2} = O\left(\frac{\log T}{T}\right) \quad (13)$$

$$N(T+c) - N(T) = O(\log T) \quad (14)$$

$$|\rho_n| \sim \frac{\pi n}{\log n}, \quad (15)$$

where γ denotes the imaginary part of a zero of $L(s, \chi)$, c is a constant and ρ_n is the n th nontrivial zero in the order of increasing modulus (it does not matter if some zeros have equal modulus).

Proof. The proof is just partial summation. See [11], pages 70-71 (there $N(T)$ counts the number of zeros of the ζ -function, but the proof is the same). \square

A proof of the formula for $N(T)$ can be found for instance in Davenport's book [5]. The proof is based on applying the argument principle to the ξ -function and estimating the contributions of different terms in the definition of $\xi(s, \chi)$.

2.4 Explicit Formula

The Hadamard product (9) for $\xi(s, \chi)$ will be the key to our proof of the explicit formula. By taking the logarithmic derivative of (9), we get

$$\frac{\xi'(s, \chi)}{\xi(s, \chi)} = B(\chi) + \sum_{\rho} \left(\frac{1}{s - \rho} + \frac{1}{\rho} \right),$$

and plugging this in the definition of $\xi(s, \chi)$, this leads to

$$\frac{L'(s, \chi)}{L(s, \chi)} = B(\chi) - \frac{1}{2} \log \frac{q}{\pi} - \frac{\Gamma'(\frac{s+a}{2})}{\Gamma(\frac{s+a}{2})} + \sum_{\rho} \left(\frac{1}{s - \rho} - \frac{1}{\rho} \right), \quad (16)$$

where $a = 0$ if χ is even and $a = 1$ if χ is odd. This formula connects the logarithmic derivative of the L -function to a sum over its zeros. Before we can utilize it, though, we need to derive an integral representation for $\psi(x, \chi)$ that connects it to the logarithmic derivative of $L(s, \chi)$.

This is a special case of the useful and well-known Perron's formula, which we now formulate.

Theorem 2.8. (Perron's formula). *Let $f(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$ be a Dirichlet series converging absolutely in the half-plane $\Re(s) = \sigma > 1$ such that*

- (i) $a_n \ll \Phi(n)$, where Φ is increasing, and
- (ii) $f(s) = O(\frac{1}{(\sigma-1)^\alpha})$ for some $\alpha > 0$ as $\sigma \rightarrow 1+$. Then for any $T \geq 1, 1 < c < c_0$ and x half more than a positive integer one has

$$A(x) := \sum_{n \leq x} a_n = \frac{1}{2\pi i} \int_{c-iT}^{c+iT} f(s) \frac{x^s}{s} ds + O\left(\frac{x^c}{T(c-1)^\alpha}\right) + O\left(\frac{x^c \Phi(2x) \log x}{T}\right),$$

and the constants inside the big O depends only on the upper bound c_0 for c .

A proof can be found in Karatsuba's book [14] on pages 64-66.

Perron's formula is often very useful as it tells that if one controls a Dirichlet series $f(s)$ well enough, that is, one is able to give upper bounds for it or to

apply the residue theorem to it, then one knows how the partial sums of the coefficients behave. We are going to apply Perron's formula to the Dirichlet series

$$\frac{L'(s, \chi)}{L(s, \chi)} = - \sum_{n=1}^{\infty} \frac{\chi(n)\Lambda(n)}{n^s}, \quad \sigma > 1,$$

which is obtained just by differentiating the Euler product $\log L(s, \chi) = \log \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}$. Its coefficients are bounded by $\log n$, and $\frac{L'(s, \chi)}{L(s, \chi)}$ is at most

$$\sum_{n=1}^{\infty} \frac{\log n}{n^\sigma} \ll \int_1^{\infty} \frac{\log t}{t^\sigma} dt \ll (\sigma - 1)^{-1}$$

as $\sigma \rightarrow 1+$ (actually, $L(1, \chi) \neq 0$, so $\frac{L'(s, \chi)}{L(s, \chi)}$ is bounded in a neighborhood of 1). Hence setting $c = 1 + \frac{1}{\log x}$, $c_0 = 3$, we arrive at

$$\psi(x, \chi) = -\frac{1}{2\pi i} \int_{c-iT}^{c+iT} \frac{L'(s, \chi)}{L(s, \chi)} \frac{x^s}{s} ds + O\left(\frac{x \log^2 x}{T}\right), \quad (17)$$

since then $x^c \ll x$. With the representation (17) of $\psi(x, \chi)$ in terms of the logarithmic derivative of the L -function now derived, we can exploit the sum formula for the logarithmic derivative of $L(s, \chi)$. In what follows, we follow the ideas presented in the books of Murty [24] and Davenport [5].

Assume that χ is odd. Let R be the rectangular curve whose vertices are $c \pm iT$ and $-K \pm iT$ (c is still $1 + \frac{1}{\log x}$), where K is an even positive integer that will go to infinity subsequently. Exploiting the representation of $\frac{L'(s, \chi)}{L(s, \chi)}$ as the sum (16), we get by the residue theorem

$$\begin{aligned} -\frac{1}{2\pi i} \int_R \frac{L'(s, \chi)}{L(s, \chi)} \frac{x^s}{s} ds &= - \sum_{\rho \in \bar{R}} \operatorname{Res}_{s=\rho} \left(\frac{L'(s, \chi)}{L(s, \chi)} \frac{x^s}{s} \right) - \sum_{1-2m \in \bar{R}} \operatorname{Res}_{s=1-2m} \left(\frac{L'(s, \chi)}{L(s, \chi)} \frac{x^s}{s} \right) \\ &= - \sum_{\rho \in \bar{R}} \lim_{s \rightarrow \rho} \left((s - \rho) \frac{L'(s, \chi)}{L(s, \chi)} \frac{x^s}{s} \right) \\ &\quad - \sum_{1-2m \in \bar{R}} \lim_{s \rightarrow 1-2m} \left((s - (1 - 2m)) \frac{L'(s, \chi)}{L(s, \chi)} \frac{x^s}{s} \right) \\ &= -\frac{L'(0, \chi)}{L(0, \chi)} - \sum_{\rho \in \bar{R}} \frac{x^\rho}{\rho} - \sum_{m=1}^{\frac{K}{2}} \frac{x^{1-2m}}{2m-1}, \end{aligned} \quad (18)$$

where the sum is over the nontrivial zeros ρ inside \bar{R} (where \bar{R} is the rectangle enclosed by R), and the latter sum arises from the trivial zeros. We just used the fact that $\frac{(x-x_0)f'(x)}{f(x)} \rightarrow 1$ as $x \rightarrow x_0$ where x_0 is a zero of f , which follows by writing $f(x) = (x-x_0)^k g(x)$ with $g(x_0) \neq 0$.

Let $R_1 = [-K + iT, -K - iT]$, $R_2 = [c + iT, -K + iT]$ and $R_3 = [-K - iT, c - iT]$. If we show that the integrals of $\frac{L'(s, \chi)}{L(s, \chi)} \frac{x^s}{x}$ over these three sides of R are of size $\ll \frac{x \log^2 x}{T}$, the explicit formula follows as the fourth side of the rectangle contributed $\psi(x, \chi) + O\left(\frac{x \log^2 x}{T}\right)$ by (17). By symmetry, it suffices to consider the integrals over R_1 and R_2 .

We first consider the integral over R_1 . We will use the following estimate.

Lemma 2.9. *For $\sigma \leq -1$, we have*

$$\left| \frac{L'(s, \chi)}{L(s, \chi)} \right| \ll \log |s| \quad (19)$$

when the distance of s to the trivial zeros of $L(s, \chi)$ is at least $\frac{1}{2}$.

Proof. We have $\frac{\xi'(s)}{\xi(s)} = \frac{\xi'(1-s)}{\xi(1-s)}$ by the functional equation, and since the logarithmic derivative of a product is the sum of logarithmic derivatives, $\frac{\xi'(s)}{\xi(s)} = \frac{L'(s, \chi)}{L(s, \chi)} + \frac{1}{2} \log \frac{q}{\pi} + \frac{\Gamma'(\frac{s+1}{2})}{\Gamma(\frac{s+1}{2})}$ (the middle term is the logarithmic derivative of $(\frac{q}{\pi})^{\frac{s+1}{2}}$). Stirling's formula (or one version of it; see [14], page 45) says $\frac{\Gamma'(z)}{\Gamma(z)} = \log z + O\left(\frac{1}{z}\right)$ for $|\arg z| < \pi$, so

$$\frac{L'(s, \chi)}{L(s, \chi)} = \frac{L'(1-s, \bar{\chi})}{L(1-s, \bar{\chi})} + O(\log |s|) \quad (20)$$

when s is uniformly bounded away from the poles of $\frac{\Gamma'(\frac{s+1}{2})}{\Gamma(\frac{s+1}{2})}$. For $\sigma \geq 2$ we may use the identity

$$\frac{L'(s, \chi)}{L(s, \chi)} = \sum_{n=1}^{\infty} \frac{\chi(n)\Lambda(n)}{n^s} \ll \sum_{n=1}^{\infty} \frac{\log n}{n^\sigma} \leq \sum_{n=1}^{\infty} \frac{\log n}{n^2} \ll 1$$

to deduce (19), and then (20) yields the lemma for $\sigma \leq -1$. \square

The integral over R_1 is now at most

$$\begin{aligned} \int_{-T}^T \left| \frac{L'(-K + it)}{L(-K + it)} \right| \frac{x^{-K}}{|K + iT|} &\ll \frac{1}{Kx^K} \int_{-T}^T \log(K^2 + t^2) dt \\ &\ll \frac{\log K}{Kx^K} \cdot T \log T. \end{aligned} \quad (21)$$

Similarly, the integral over $R_2 \cap \{\sigma \leq -1\}$ is not greater than

$$\begin{aligned} \int_{-K}^{-1} \left| \frac{L'(t+iT)}{L(t+iT)} \right| \frac{x^t}{|t+iT|} dt &\ll \frac{1}{T} \int_1^\infty \log(t^2 + T^2) x^{-t} dt \\ &\ll \frac{\log T}{T} \int_1^\infty (\log t) x^{-t} dt \\ &\ll \frac{\log T}{T}. \end{aligned} \tag{22}$$

To estimate the integral over the segment $[-1+iT, c+iT]$, we need a lemma.

Lemma 2.10. *There exists a sequence T_n tending to infinity, such that $n < T_n < n+1$ for all $n \in \mathbb{Z}_+$, and satisfying*

$$\left| \frac{L'(s, \chi)}{L(s, \chi)} \right| \ll \log^2 T_n$$

when $\Im(s) = T_n$ and $-1 \leq \sigma \leq 2$.

Proof. We employ formula (16) and Stirling's formula $\frac{\Gamma'(z)}{\Gamma(z)} = \log z + O\left(\frac{1}{z}\right)$ to see that

$$\frac{L'(s, \chi)}{L(s, \chi)} = \sum_{\rho} \left(\frac{1}{s-\rho} + \frac{1}{\rho} \right) + O(\log T)$$

when $-1 \leq \sigma \leq 2$ and $\Im(s) = T$ (notice that s is not close to the poles of Γ). We estimate the sum over the zeros in several parts:

$$\begin{aligned} \sum_{|\rho| \geq 2|s|} \left(\frac{1}{s-\rho} + \frac{1}{\rho} \right) &\ll |s| \sum_{|\rho| \geq 2|s|} \frac{1}{|\rho|^2 - |s||\rho|} \\ &\ll |s| \sum_{|\rho| > 2T} \frac{2}{|\rho|^2} \ll \log T, \end{aligned}$$

because of the relations $|s| \leq \frac{|\rho|}{2}$ and $T \leq |s| < T+2$, and Corollary 2.7.

Furthermore, again by exploiting the formula for $N(T)$ and the fact that the zeros ρ with $|s-\rho| = k$ have imaginary part between $k-1$ and $k+1$ or

$-k - 1$ and $-k + 1$ for $k \gg 1$, we compute

$$\begin{aligned}
\sum_{\substack{|s-\rho|>1 \\ \frac{1}{2}|s|\leq|\rho|\leq 2|s|}} \left(\frac{1}{s-\rho} + \frac{1}{\rho} \right) &\ll |s| \sum_{k\leq 3|s|} \sum_{\substack{|s-\rho|\in(k,k+1] \\ \frac{1}{2}|s|\leq|\rho|}} \frac{1}{k|\rho|} \\
&\ll \sum_{k\leq 3|s|} \frac{N(|s|+k+1) - N(|s|+k-1)}{k} \\
&\ll \log T \sum_{k=1}^{3|s|} \frac{1}{k} \ll \log^2 T,
\end{aligned}$$

We also bound

$$\begin{aligned}
\sum_{|\rho|\leq \frac{1}{2}|s|} \left(\frac{1}{s-\rho} + \frac{1}{\rho} \right) &\ll |s| \sum_{|\rho|\leq \frac{1}{2}|s|} \frac{1}{|\rho||s-\rho|} \\
&\ll \sum_{|\rho|\leq \frac{1}{2}|s|} \frac{1}{|\rho|} \ll \log T.
\end{aligned}$$

Lastly, in order to bound the sum where $|s - \rho| < 1$, we use the fact that the interval $[n, n + 1]$ can be partitioned into intervals so that $L(s, \chi)$ vanishes only when $\Im(s)$ is at an endpoint of an interval and so that their number is at most $c_q \log T$ for some $c_q > 0$ (this is immediate from Corollary 2.7, which said $N(T + 1) - N(T) \leq c_q \log T$). Choose T_n to be a midpoint of the longest of these subintervals of $[n, n + 1]$. Then $|\rho - T_n| \gg \frac{1}{\log n}$ for all zeros ρ , and therefore

$$\begin{aligned}
\sum_{\rho:|s-\rho|\leq 1} \left(\frac{1}{s-\rho} + \frac{1}{\rho} \right) &\ll |s| \sum_{|s-\rho|\leq 1} \frac{1}{|s||\rho-s|} \\
&\ll \log T \cdot \max_{\rho} \frac{1}{|s-\rho|} \ll \log^2 T
\end{aligned}$$

when $T = T_n = |s|$.

Putting it all together, we see that for the T_n chosen above, the lemma holds. \square

Now, returning to the proof of the explicit formula, we see that the integral

of $-\frac{L'(s, \chi)}{L(s, \chi)} \frac{x^s}{s}$ on the segment $[-1 + iT_n, 1 + \frac{1}{\log x} + iT_n]$ is bounded by

$$\begin{aligned} \int_{-1}^{1 + \frac{1}{\log x}} \log^2 T_n \frac{x^t}{|t + iT_n|} dt &\ll \frac{\log^2 T_n}{T_n} \int_{-1}^{1 + \frac{1}{\log x}} x^t dt \\ &\ll \frac{x \log^2 T_n}{T_n}. \end{aligned} \quad (23)$$

As the integral over R_1 approaches zero as $K \rightarrow \infty$, taking the limit in (18), we obtain

$$-\frac{1}{2\pi i} \int_C \frac{L'(s, \chi)}{L(s, \chi)} \frac{x^s}{s} ds = -\sum_{\rho \in \bar{C}} \frac{x^\rho}{\rho} - \frac{L'(0, \chi)}{L(0, \chi)} + \sum_{m=1}^{\infty} \frac{x^{1-2m}}{2m-1},$$

where C is otherwise as R but extends to infinity on the left, \bar{C} is again the region inside C , and the last term grows logarithmically in x . Notice that in the sum, \bar{C} can be replaced with $\bar{C} \cap \{s : 0 \leq \sigma \leq 1\}$ because the nontrivial zeros lie in this domain. Now if T_n is given by Lemma 2.10 and chosen so that $T_n \leq T \leq T_n + 1$, combining the previous equation with (17) and the estimates (21), (23) for the integral over R_2 , we conclude

$$\begin{aligned} \psi(x, \chi) &= -\int_{c-iT_n}^{c+iT_n} \frac{L'(s, \chi)}{L(s, \chi)} \frac{x^s}{s} ds + O\left(\frac{x \log^2 x}{T}\right) + O\left(\frac{x \log^2 T_n}{T_n}\right) \\ &= -\sum_{\rho \in \bar{C}} \frac{x^\rho}{\rho} - \frac{L'(0, \chi)}{L(0, \chi)} + \sum_{m=1}^{\infty} \frac{x^{1-2m}}{2m-1} + O\left(\frac{x \log^2(xT)}{T}\right) \\ &= -\sum_{|\rho| < T} \frac{x^\rho}{\rho} - \frac{L'(0, \chi)}{L(0, \chi)} + \sum_{m=1}^{\infty} \frac{x^{1-2m}}{2m-1} + O\left(\frac{x \log^2(xT)}{T}\right). \end{aligned} \quad (24)$$

We used the fact that the zeros ρ with $T_n \leq \Im(\rho) \leq T_n + 1$ satisfy $\left|\frac{x^\rho}{\rho}\right| \ll \frac{x}{T}$ and their number is $O(\log T)$. This finishes the proof of Theorem 2.1 for odd characters.

The case of an even character is almost identical, with only very minor changes: K is then odd, x^{1-2m} should be replaced with x^{-2m} in (18), and $-\frac{L'(0, \chi)}{L(0, \chi)}$ should be replaced with $-b(\chi) - \log x$, since it is the residue of $-\frac{L'(s, \chi)}{L(s, \chi)} \frac{x^s}{s}$ at $s = 0$ due to the expansions

$$\frac{L'(s, \chi)}{L(s, \chi)} = \frac{1}{s} + b(\chi) + \dots \quad \text{and} \quad \frac{x^s}{s} = \frac{1}{s} + \log x + \dots$$

In addition, as noticed at the beginning of the chapter, letting $T \rightarrow \infty$ proves the explicit formula (6) with an infinite sum over the zeros. \square

In the case of $\chi(n) = 1$ for all n (which is the previously excluded trivial character inducing the Riemann ζ function), very similar arguments turn out to work, and an analogous formula holds:

$$\psi(x) := \sum_{n \leq x} \Lambda(n) = x - \sum_{|\rho| < T} \frac{x^\rho}{\rho} + O\left(\frac{x \log^2(xT)}{T} + \log x\right); \quad (25)$$

this leads to the prime number theorem with error term, among other things. For a proof, see the book [14] of Karatsuba, for instance.

2.5 Generalized Riemann Hypothesis

With the explicit formula proved, we turn to some rather immediate consequences of it. The *generalized Riemann hypothesis*, or the GRH, states that

$$L(s, \chi) \neq 0 \quad \text{for } s \in \left\{s : 0 \leq \sigma < \frac{1}{2}\right\} \cup \left\{s : \frac{1}{2} < \sigma \leq 1\right\}$$

for every Dirichlet character χ (it would of course suffice to state that there are no zeros with $\frac{1}{2} < \sigma \leq 1$). This extremely important conjecture is directly connected to the size of the function $\psi(x; q, a)$ via the following result.

Theorem 2.11. *The GRH is equivalent to*⁸

$$\psi(x; q, a) = \frac{x}{\varphi(q)} + O(\sqrt{x} \log^2 x)$$

for any fixed coprime a and q .

Proof. First assume the GRH. We know that

$$\psi(x; q, a) = \frac{1}{\varphi(q)} \sum_{\chi \pmod{q}} \bar{\chi}(a) \psi(x, \chi), \quad (26)$$

⁸This result was proved in a similar way for $\psi(x)$ assuming the (ordinary) Riemann hypothesis by Helge von Koch in 1901, but it remains essentially the best result we know. One can thus say that the (generalized) Riemann hypothesis is true if and only if the primes (in arithmetic progressions) are distributed as uniformly as possible. [24]

and by choosing $T = \sqrt{x}$ in (25), we have for the term $\psi(x, \chi_0)$ that⁹

$$|\psi(x, \chi_0) - x| \leq \left| \sum_{|\rho| < \sqrt{x}} \frac{x^{\frac{1}{2}}}{\rho} \right| + O(\sqrt{x} \log^2 x),$$

since $\Re(\rho) = \frac{1}{2}$. Moreover,

$$\sum_{|\rho| < \sqrt{x}} \frac{1}{\rho} \ll \log^2 x.$$

by (12). Similarly, exploiting the truncated versions of the explicit formula, that is (4) and (5), with $T = \sqrt{x}$, we get

$$\psi(x, \chi) = O(\sqrt{x} \log^2 x)$$

for non-principal characters χ . Now one of the terms in (26) is $x + O(\sqrt{x} \log^2 x)$ and the others are $O(\sqrt{x} \log^2 x)$, which proves the claim.

Conversely, assume that $|\psi(x; q, a) - \frac{x}{\varphi(q)}| \ll \sqrt{x} \log^2 x$ for all coprime a and q . Then choosing $q = 1$, we get $|\psi(x, \chi_0) - x| \ll \sqrt{x} \log^2 x$ for the principal character $\chi_0 \pmod{q}$, so by (26) the functions

$$f_x(a) := \frac{1}{\varphi(q)} \sum_{\substack{x \pmod{q} \\ \chi \neq \chi_0}} \bar{\chi}(a) \psi(x, \chi)$$

are $O(\sqrt{x} \log^2 x)$. By reversing the order of summation, we see that

$$\sum_{a \in \mathbb{Z}_q^\times} \chi^*(a) f_x(a) = \psi(x, \chi^*)$$

for any character $\chi^* \neq \chi_0$ so $\psi(x, \chi) = O(\sqrt{x} \log^2 x)$ for non-principal χ .

Then we may use partial summation in the following way:

$$\begin{aligned} \frac{L'(s, \chi)}{L(s, \chi)} &= - \sum_{n=1}^{\infty} \frac{\chi(n) \Lambda(n)}{n^s} \\ &= \varphi(q) \int_1^{\infty} \frac{\psi(x, \chi)}{x^{s+1}} dx \\ &\ll \varphi(q) \int_1^{\infty} \frac{\log^2 x}{x^{\frac{1}{2} + \sigma}} dx < \infty \end{aligned}$$

⁹Notice that $\psi(x, \chi_0)$ and $\psi(x)$ differ by a bounded amount. Also notice that although the GRH does not directly speak about ζ , the ordinary Riemann hypothesis is just a special case, since $L(s, \chi_0)$ and $\zeta(s)$ are equal up to a finite number of factors $1 - \frac{1}{p_i^s}$.

for $\sigma > \frac{1}{2}$ and $\chi \neq \chi_0$. This means that $\frac{L'(s, \chi)}{L(s, \chi)}$ has no poles in that half-plane, so the L -functions corresponding to non-principal characters are zero-free there. The function $L(s, \chi_0)$ is just a finite product $\prod_{p|q} (1 - \frac{1}{p^s})$ times $\zeta(s)$, and we have

$$\begin{aligned} \frac{\zeta'(s)}{\zeta(s)} + \zeta(s) &= - \sum_{n=1}^{\infty} \frac{\Lambda(n) - 1}{n^s} \\ &= \int_1^{\infty} \frac{\psi(x) - x}{x^{s+1}} dx \\ &\ll \int_1^{\infty} \frac{\log^2 x}{x^{\frac{1}{2} + \sigma}} dx < \infty, \end{aligned}$$

so $\frac{\zeta'}{\zeta}(s) + \zeta(s)$ is also analytic in the half-plane $\sigma > \frac{1}{2}$, completing the proof. \square

In the main chapter of the thesis, the GRH is a fundamental assumption, as is mentioned in the Introduction.

2.6 Grand Simplicity Hypothesis

Another assumption that has to be made in Chapter 4 is the *grand simplicity hypothesis* (GSH):

$$\left\{ \gamma \geq 0 : L\left(\frac{1}{2} + i\gamma, \chi\right) = 0, \chi \text{ a character (mod } q) \right\}$$

is a set that is linearly independent over the rationals for any q (we also interpret that these zeros may not occur twice and must be simple).

Two very particular corollaries are that the zeros $\sigma = \frac{1}{2}$ are simple, and that $L(\frac{1}{2}, \chi) \neq 0$ but even both of these are still open problems, although they are widely believed to be true [19]. It has to be remarked that a rational dependence between some zeros of L -functions would be astonishing, since even no algebraic dependencies between the nontrivial zeros have been found or believed to exist, so the GSH is "likely" to be true. Furthermore, it is easily seen that the n -dimensional measure of the sequences $(x_1, \dots, x_n) \in \mathbb{R}^n$ having a rational linear relation among them is zero for any n , so there ought to be no rational dependence between the nontrivial zeros if they are "random enough". In the Prime Races chapter, the GSH is required to assure that trigonometric sums of the form

$$\sum_{\gamma > 0} c_{\chi} \frac{e^{i\gamma \log x}}{\frac{1}{2} + i\gamma},$$

where the sum is over the positive imaginary parts of the zeros of $L(s, \chi)$ on $\sigma = \frac{1}{2}$, behave like sums of random variables, and leads to a formula for the value distribution of the sum.

3 Lemmas from Measure Theory and Probability Theory

In this chapter, we recall some concepts from measure theory, probability theory and real analysis required in the next chapter on the prime races.

3.1 Measure Theory

All the facts presented in this subsection can be found in [30] or [34]. The reader is assumed to be familiar with basic measure theory, that is sigma algebras, measures (which to us are always nonnegative), measurability, integration with respect to a measure, and monotone and dominated convergence. We state the measure theoretic lemmas that will be applied in the next chapter to measures related to the prime races.

The following corollary of Fubini's theorem (or the Fubini–Tonelli theorem to be more accurate) is a useful relation between probability density functions and expectation.

Lemma 3.1. (A corollary of Fubini's theorem) *Let $f : X \rightarrow \mathbb{R}$ be a nonnegative μ -measurable function, where X is σ -finite. Then*

$$\int_X f d\mu = \int_0^\infty \mu(\{x \in X : f(x) > t\}) dt. \quad (27)$$

Borel measures and *distributions*, that is, continuous linear functionals on $C_0(\mathbb{R}^d)$ (the space of compactly supported continuous functions on \mathbb{R}^d)¹⁰, are closely related by the Riesz representation theorem, which we are going to use several times.

Lemma 3.2. (Riesz representation theorem) [30] *Let T be a positive linear functional on $C_0(\mathbb{R}^d)$; that is, T maps continuous compactly supported functions to real numbers and $T(f) \geq 0$ for $f \geq 0$. Then there exists a unique Borel measure μ for which*

$$T(f) = \int_{\mathbb{R}^d} f(x) d\mu(x) \quad (28)$$

¹⁰Distributions can be defined in other spaces as well, but this is the case that is relevant for our purposes.

for all continuous compactly supported functions f .

Two measures μ and ν on the same σ -algebra and measure space are in a sense comparable if one of them (say μ) is *absolutely continuous* with respect to the other: $\mu(A) = 0$ whenever $\nu(A) = 0$. In this case, if μ and ν are also σ -finite, we have the following theorem.

Lemma 3.3. (Radon-Nikodym theorem) [30] *If μ is absolutely continuous with respect to ν , one has*

$$\mu(A) = \int_A f d\nu \tag{29}$$

for all μ -measurable A and an (almost everywhere) unique nonnegative ν -measurable f .

This $f = \frac{d\mu}{d\nu}$ is called the *Radon-Nikodym derivative* of μ with respect to ν . Sometimes we denote $\mu = f(x)d\nu(x)$ to indicate that the measures are related by (29) (the Lebesgue measure can be denoted just by dx). If μ is a Borel measure on \mathbb{R}^d which is absolutely continuous with respect to the Lebesgue measure m_d , we denote the Radon-Nikodym derivative by $\mu(x)$. In this way, we have naturally assigned a function to the measure μ . Since

$$\mu(A) = \int_A \mu(x) dx$$

by definition, $\mu(x)$ can be called the *probability density function of the measure μ* if μ is a probability measure. Conversely, given a nonnegative integrable function $f : X \rightarrow \mathbb{R}$, formula (29) does indeed define a measure, so functions can be interpreted as measures. Absolute continuity of measures will be important when dealing with measures related to the prime races.

We are also going to need the notion of the *weak convergence* of measures. We say that the measures μ_n , $n = 1, 2, \dots$ on $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$ converge to a measure μ defined on the same measure space weakly if

$$\int_{\mathbb{R}^d} f d\mu_n \rightarrow \int_{\mathbb{R}^d} f d\mu \tag{30}$$

for all bounded continuous functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$.

A theorem of fundamental importance in the next chapter is the portmanteau theorem, which gives equivalent characterizations of the convergence of probability measures. For a proof, see [29].

Lemma 3.4. (Portmanteau theorem) *Let $\mu_n, n = 1, 2, \dots$ and μ be Borel probability measures on \mathbb{R}^d . Then the following are equivalent.¹¹*

(i) $\mu_n \rightarrow \mu$ weakly, that is, $\int_{\mathbb{R}^d} f d\mu_n \rightarrow \int_{\mathbb{R}^d} f d\mu$ for all bounded continuous functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$.

(ii) $\int_{\mathbb{R}^d} f d\mu_n \rightarrow \int_{\mathbb{R}^d} f d\mu$ for all bounded Lipschitz functions $f : \mathbb{R}^d \rightarrow \mathbb{R}$.

(iii) $\lim_{n \rightarrow \infty} \mu_n(A) = \mu(A)$ for all sets A with $\mu(\partial A) = 0$.

One more measure theoretic lemma is needed to prove Theorem 4.1 in the following chapter. If ν is a Borel probability measure with compact support, then for integrable functions f we can define the functional $\nu(f) = \int_{\mathbb{R}^d} f d\nu$ associated to ν .

Lemma 3.5. (Functionals arising from measures) *Let $\nu_n, n = 1, 2, \dots$ be Borel probability measures on \mathbb{R}^d with compact supports. Assume that the limit $\Lambda(f) := \lim_{n \rightarrow \infty} \nu_n(f)$ exists and is finite for all bounded Lipschitz functions $f \in Lip_b(\mathbb{R}^d)$. Then there exists a Borel probability measure μ such that $\Lambda(f) = \int_{\mathbb{R}^d} f d\mu$ for all $f \in Lip_b(\mathbb{R}^d)$*

This lemma will allow us to construct the weak limit of a sequence of measures related to the prime race. In the space $C_0(\mathbb{R}^d)$ the claim would just be a trivial consequence of the Riesz representation theorem, but we need the lemma for a bigger space. In $Lip_b(\mathbb{R}^d)$ functionals do not generally arise from measures, so we must exploit the condition. Notice that if we could choose f to be a characteristic function above, we would obtain $\Lambda(1_A) = \lim_{n \rightarrow \infty} \nu_n(A)$ for all Borel sets A , so $\Lambda(1_A)$ would be a measure by the Vitali-Hahn-Saks theorem (see [3]). Therefore we modify the proof of the Vitali-Hahn-Saks theorem in [3] to prove this lemma. We start with the following

Lemma 3.6. (Schur) *Let $a_{n,k}$ be nonnegative real numbers such that $a_k := \lim_{n \rightarrow \infty} a_{n,k}$ exist and*

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} a_{n,k} < \infty.$$

Then

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} |a_{n,k} - a_k| = 0.$$

¹¹The equivalence of (i) and (ii) is rather surprising as many functions such as $x \mapsto \sin x^2$ are bounded and continuous but not even uniformly approximable with Lipschitz functions.

For a proof, see [31]. Next we prove the original lemma about functionals.

Proof of Lemma 3.5. Evidently Λ is a positive linear functional. We may write $f = f^+ + f^-$ where $f^+ \geq 0$, $f^- \leq 0$ and the functions are bounded and Lipschitz, so, without loss of generality, $f \geq 0$. Let us prove that Λ satisfies the monotone convergence theorem in the following sense: Whenever $f_n \geq 0$ are bounded and Lipschitz, $f_{n+1} \geq f_n$ and $f_n \rightarrow f \in Lip_b(\mathbb{R}^d)$ pointwise, we have $\Lambda(f_n) \rightarrow \Lambda(f)$. For this, write $g_n = f_{n+1} - f_n \geq 0$, where $f_0 := 0$. Then $f_n = \sum_{k=0}^{n-1} g_k$, and by monotone convergence for the measures ν_n , we have

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} \nu_n(g_k) &= \lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} \int_{\mathbb{R}^d} g_k d\nu_n \\ &= \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f d\nu_n = \Lambda(f) < \infty. \end{aligned}$$

Hence Schur's lemma gives

$$\lim_{n \rightarrow \infty} \sum_{k=0}^{\infty} |\nu_n(g_k) - \Lambda(g_k)| = 0.$$

In particular, by monotone convergence for ν_n ,

$$\begin{aligned} 0 &= \lim_{n \rightarrow \infty} \left| \sum_{k=0}^{\infty} \nu_n(g_k) - \sum_{k=0}^{\infty} \Lambda(g_k) \right| \\ &= \lim_{n \rightarrow \infty} |\nu_n(f) - \sum_{k=0}^{\infty} \Lambda(g_k)| = |\Lambda(f) - \sum_{k=0}^{\infty} \Lambda(g_k)| \end{aligned}$$

This means that $\Lambda(f_N) = \sum_{k=0}^{N-1} \Lambda(g_k) \rightarrow \Lambda(f)$ as $N \rightarrow \infty$, proving the claim about monotone convergence.

Now let μ be a Borel measure such that for compactly supported continuous functions f we have $\Lambda(f) = \int_{\mathbb{R}^d} f d\mu$; such a measure exists by the Riesz representation theorem. Let $f \in Lip_b(\mathbb{R}^d)$ be nonnegative, and let f_n be compactly supported continuous functions such that $f_n \leq f_{n+1}$ and $f_n \rightarrow f$ pointwise (let f_n be equal to f in the ball $B(0, n)$ and supported on $B(0, n+1)$ in such a way that f_n is Lipschitz and satisfies $0 \leq f_n \leq f$). Then f is the monotone limit of the functions f_n , so $\Lambda(f_n) \rightarrow \Lambda(f)$. On the other hand, by monotone convergence for μ and the fact that f_n vanishes outside a compact set,

$$\lim_{n \rightarrow \infty} \Lambda(f_n) = \lim_{n \rightarrow \infty} \int_{\mathbb{R}^d} f_n d\mu = \int_{\mathbb{R}^d} f d\mu.$$

Thus Λ arises from a Borel measure, as wanted. By setting $f(x) = 1$ for all x , we see that μ is a probability measure. \square

3.2 Probability Theory

We need some results and concepts from probability theory subsequently. A good general reference concerning probability theory is [12]. Let $(\Omega, \mathcal{F}, \mu)$ be a measure space, where μ is a probability measure. A function $f : \Omega \rightarrow \mathbb{R}$ is called a *random variable* if it is measurable in this space.¹²

In \mathbb{R}^d we may define a natural density (when it exists) by

$$P(A) := \lim_{x \rightarrow \infty} \frac{1}{(2x)^d} m(\{A \cap [-x, x]^d\}), \quad (31)$$

where m is the Lebesgue measure on \mathbb{R}^d . Intuitively, $P(A)$ tells the "asymptotic proportion" of \mathbb{R}^d covered by A . Although P is finitely additive, it is unfortunately not a measure.

A sequence f_1, f_2, \dots of random variables in $(\mathbb{R}^d, \text{Bor}(\mathbb{R}^d))$ is said to be *asymptotically independent* if

$$P(\{f_{i_1} \in B_{i_1}, \dots, f_{i_n} \in B_{i_n}\}) = \prod_{k=1}^n P(\{f_{i_k} \in B_{i_k}\}) \quad (32)$$

for any indices $i_1 < \dots < i_n$ and Borel sets B_{i_1}, \dots, B_{i_n} . Intuitively, this means that the f_i 's do not correlate: knowledge of the value of f_i does not affect the value distribution of f_j for $i \neq j$ (and similarly for larger collections of f_i 's).

In connection with the grand simplicity hypothesis, which tells that the imaginary parts of the zeros of the L -functions are in a way random, the Kronecker-Weyl theorem turns out to be of crucial importance. We remark that in many sources it is only formulated in a special case.

Lemma 3.7. (Kronecker-Weyl theorem) [22] *Let $\gamma_1, \dots, \gamma_n \in \mathbb{R}$ and $G = \{(\gamma_1 y \pmod{1}, \dots, \gamma_n y \pmod{1}) : y \in \mathbb{R}\}$. Then the closure \bar{G} is a subtorus of $\mathbb{R}^n/\mathbb{Z}^n$ of dimension r (that is, isomorphic to $\mathbb{R}^r/\mathbb{Z}^r$), where r is the dimension of the \mathbb{Q} -vector space spanned by $\gamma_1, \dots, \gamma_n$, and moreover, G is uniformly distributed on this subtorus, meaning that*

$$\lim_{Y \rightarrow \infty} \int_0^Y f(\gamma_1 y \pmod{1}, \dots, \gamma_n y \pmod{1}) dy \rightarrow \int_{\bar{G}} f(y) d\mu_{\bar{G}}(y) \quad (33)$$

¹²A more standard notation in this setting would be P for the probability measure and X for a random variable. However, these notations are reserved in the following chapter.

for all bounded continuous functions f on \mathbb{R}^r . Here $\mu_{\bar{G}}$ is the normalized Haar measure on \bar{G} .¹³

In particular, the Kronecker-Weyl theorem implies the following lemma, which will be applied in the next chapter where γ_i will be the imaginary parts zeros of L -functions.

Lemma 3.8. (Independence of cosines) *The functions $1, \cos(\gamma_1 x), \dots, \cos(\gamma_n x)$ are asymptotically independent if $\gamma_1, \dots, \gamma_n$ are linearly independent over \mathbb{Q} .*

Proof. We must verify the equation (32), where $f_k(x) = \cos(\gamma_k x)$. It is well-known that it suffices to consider the cases $B_i = (-\infty, t_i)$ as these sets generate the Borel sigma algebra. We may assume $t_i < 1$ as otherwise $f_i(x) \in B_i$ for all x . Now $\cos(\gamma x) > t_i$ if and only if

$$\frac{\gamma x}{2\pi} \pmod{1} \in (-\arccos t_i, \arccos t_i).$$

If we denote $I_i = [-1, 1] \setminus (-\arccos t_i, \arccos t_i)$, our task is to prove

$$P\left(\left\{\frac{\gamma_1 x}{2\pi} \pmod{1} \in I_1, \dots, \frac{\gamma_n x}{2\pi} \pmod{1} \in I_n\right\}\right) = \prod_{k=1}^n P\left(\left\{\frac{\gamma_k x}{2\pi} \pmod{1} \in I_k\right\}\right),$$

and this is by Weyl's criterion on equidistribution (See [32], pages 108-112. There $n = 1$ but the general case is similar.) the same as claiming that the set

$$G := \left\{ \left(\frac{\gamma_1 x}{2\pi} \pmod{1}, \dots, \frac{\gamma_n x}{2\pi} \pmod{1} \right) : x \in \mathbb{R} \right\}$$

is equidistributed in $\mathbb{R}^n/\mathbb{Z}^n$ in the sense used in the Kronecker-Weyl theorem. The Kronecker-Weyl theorem now gives the claim, as $r = n$ in that theorem by our assumption, so the statement is proved. \square

3.3 Fourier transform

The Fourier transform of an integrable function $f : \mathbb{R}^d \rightarrow \mathbb{C}$ is

$$\hat{f}(\xi) := \int_{\mathbb{R}^d} e^{-ix \cdot \xi} f(x) dx. \tag{34}$$

¹³It can be shown that on any locally compact abelian group there is up to a constant unique measure that is translation invariant and for which measure of sets can be approximated using open or compact sets, and this is called the Haar measure. For a proof of the existence and uniqueness of the Haar measure, see [18].

We mention this because various normalizations are used depending on the author. The reader is assumed to know the basic formulas for the Fourier transform in \mathbb{R}^d , including the inversion formula and the Fourier transform of the Gaussian function (which is $\widehat{e^{-ax^2}} = \sqrt{\frac{\pi}{a}}e^{-\frac{\xi^2}{4a}}$ for $a > 0$). Furthermore, the reader should know that the Fourier transform of a convolution is a product of Fourier transforms, and that in $L^2(\mathbb{R}^d)$ the Fourier transform is a bijection.

We shall also need the *Fourier transform of a Borel measure* μ , defined by

$$\hat{\mu}(\xi) := \int_{\mathbb{R}^d} e^{-ix \cdot \xi} d\mu(\xi).$$

If μ is absolutely continuous with respect to the Lebesgue measure, and $\hat{\mu}$ is integrable, the function $\mu(x)$ is defined and can be recovered:

$$\mu(x) = \frac{1}{2\pi} \int_{\mathbb{R}^d} e^{ix \cdot \xi} \hat{\mu}(\xi) d\xi.$$

Finally, the Fourier transform is even defined for tempered distributions, that is, continuous linear functionals from $\mathcal{S}(\mathbb{R}^n)$ to \mathbb{C} where $\mathcal{S}(\mathbb{R}^n)$ is the Schwartz space consisting of functions whose all derivatives decay faster than polynomially. The definition is

$$\hat{T}(\varphi) = T(\hat{\varphi})$$

for tempered distributions T and Schwartz functions φ .¹⁴

We also need a theorem connecting the convergence of Fourier transforms to the convergence of the original functions.

Lemma 3.9. Levy's theorem [12]: *For random variables $f_n, f : \mathbb{R}^d \rightarrow \mathbb{R}$ the convergence $\hat{f}_n \rightarrow \hat{X}$ pointwise as $n \rightarrow \infty$ implies $f_n \rightarrow f$ in distribution¹⁵ if \hat{f} is continuous.*

3.4 Bessel function

In the next chapter we are going to need the basic Bessel function to compute Fourier transforms of certain measures. A good source on Bessel functions is

¹⁴This is seen to be a generalization of the earlier definitions using Parseval's formula.

¹⁵That is, the distribution functions of f_n converge to the distribution of X at the points where the distribution function of f is continuous

[16]. The *Bessel function* of order zero (henceforth referred to as the Bessel function) is by definition

$$J_0(z) := \sum_{m=0}^{\infty} \frac{(-1)^m}{4^m (m!)^2} z^{2m}, \quad (35)$$

a function that is analytic in the complex plane due to locally uniform convergence of the series. The Bessel function arises in computing several integrals; for instance

$$J_0(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ix \sin t} dt, \quad (36)$$

which will be utilized later on.

4 The Prime Races

4.1 Definitions and the Statement of the Main Theorem

Let $a_1, \dots, a_r \in \mathbb{Z}_q^\times$ be distinct and define

$$P_{q;a_1, \dots, a_r} = \{x \geq 2 : \pi(x; q, a_1) > \pi(x; q, a_2) > \dots > \pi(x; q, a_r)\}$$

as the set of real numbers $x \geq 2$ for which the number of primes up to x is distributed into residue classes (mod q) in such a way that the number of primes a_i (mod q) is greater than that of the primes a_{i+1} (mod q) for $i = 1, \dots, r-1$. The set $P_{q;a_1, \dots, a_r}$ is called the prime race between the residue classes a_1, \dots, a_r , in this order. One of the main questions in comparative prime number theory, and the main topic of this thesis, is the study of the sizes of these sets. It turns out that the right measure for the sets is the logarithmic density $\delta(P)$ which is defined for Lebesgue measurable sets $P \subset [1, \infty)$ by first setting

$$\delta_*(P) = \liminf_{X \rightarrow \infty} \frac{1}{\log X} \int_{[1, X] \cap P} \frac{dt}{t} \quad (37)$$

$$\delta^*(P) = \limsup_{X \rightarrow \infty} \frac{1}{\log X} \int_{[1, X] \cap P} \frac{dt}{t} \quad (38)$$

(these are the lower and upper logarithmic densities, respectively) and then $\delta(P) = \delta_*(P) = \delta^*(P)$ whenever the limits coincide. The measure $\nu = \frac{dt}{t}$, with respect to which we integrate in the above formulas, is called logarithmic since $\nu([a, b]) = \int_a^b \frac{dt}{t} = \log b - \log a$ for $b \geq a \geq 1$. Note that if P consists

of intervals $[P_i, P_i + 1)$, where P_i 's are integers (this happens for example if $P = P_{q;a_1, \dots, a_r}$), then

$$\delta(P) = \lim_{x \rightarrow \infty} \frac{1}{\log x} \sum_{P_i \leq x} \log \left(1 + \frac{1}{P_i} \right) = \lim_{x \rightarrow \infty} \frac{1}{\log x} \sum_{P_i \leq x} \frac{1}{P_i} + O \left(\lim_{x \rightarrow \infty} \frac{1}{\log x} \sum_{P_i \leq x} \frac{1}{P_i^2} \right)$$

by the Taylor series of $\log(1 + y)$, and the last term is just zero. Also notice that $\delta([1, \infty)) = 1$, which explains the word density.

The size of the density $\delta(P_{q;a_1, \dots, a_r})$ is closely related to whether a_i are squares or nonsquares $(\text{mod } q)$, but to be more precise, we will need the function

$$c(q, a) := -1 + \sum_{\substack{b \in [0, q-1] \\ b^2 \equiv a \pmod{q}}} 1.$$

This quantity is equal to -1 if and only if a is a quadratic nonresidue $(\text{mod } q)$. In addition, if $q = 4$ or q is a prime power or twice a prime power, there exists a primitive root $(\text{mod } q)$, and using this we see that $c(q, a)$ is just the Legendre symbol. It is also important to note that if a is a quadratic residue $(\text{mod } q)$, then $c(q, a) = c(q, 1)$, so the value does not depend on a . This is elementary number theory; see Gauss' book [8], pages 67-71. We occasionally use the notations R and N for the quadratic residues and nonresidues $(\text{mod } q)$, respectively (so, for instance, $P_{q;R,N}$ describes the race between the primes in R and the primes in N).

With these definitions, the Main Theorem of this thesis says

Main Theorem. *Assume the generalized Riemann hypothesis (GRH) and the grand simplicity hypothesis (GSH). Then $\delta(P_{q;a_1, \dots, a_r})$ exists and is strictly positive. In fact, for any function $\eta(x)$ tending to infinity, the set*

$$\left\{ x \geq 2 : \pi(x; q, a_{i+1}) > \pi(x; q, a_i) + \frac{\sqrt{x}}{\eta(x) \log x}, i = 1, \dots, r - 1 \right\}$$

also has $\delta(P_{q;a_1, \dots, a_r}) > 0$ as its logarithmic density.

Moreover, we have $\delta(P_{q;a,b}) > \frac{1}{2}$ if a is a quadratic nonresidue $(\text{mod } q)$ and b is a quadratic residue $(\text{mod } q)$, and $\delta(P_{a,b}) = \frac{1}{2}$ if a and b are both quadratic residues or nonresidues.

If $r = 3$ and there exists $\omega \in \mathbb{Z}_q^\times$ such that $\omega^3 \equiv 1 \pmod{q}$, $a_2 \equiv a_1 \omega \pmod{q}$ and $a_3 \equiv a_1 \omega^2 \pmod{q}$, the prime race is even in the sense that $\delta(P_{q;a_{\sigma(1)}, a_{\sigma(2)}, a_{\sigma(3)}}) =$

$\frac{1}{6}$ for any permutation σ of $\{1, 2, 3\}$.¹⁶

For $q = 4, p^\alpha, 2p^\alpha$, where p is a prime, the race between the residues and the nonresidues evens out as the modulus increases, in the sense that $\delta(P_{q;N,R}) \rightarrow \frac{1}{2}$ as $q \rightarrow \infty$.

Rubinstein and Sarnak proved the claims above in their seminal paper and with the same assumptions calculated that, for example, $\delta(P_{4;3,1}) = 0.9959\dots$ and $\delta(P_{3;2,1}) = 0.9990\dots$, so the prime races (mod 4) and (mod 3) are led by the nonresidues extremely often, but still the quadratic residues are way ahead infinitely many times. Although the Main Theorem cannot be formulated in terms of the asymptotic density, we note that the lower asymptotic densities of $P_{q;a_1,\dots,a_r}$ must be positive, for otherwise the logarithmic densities would also be zero. The difference $\frac{\sqrt{x}}{\eta(x)\log x}$ between the prime counting functions is almost the best one can hope for, since the GRH implies that the differences are $\ll \sqrt{x}\log x$. Finally, we remark that Rubinstein and Sarnak proved that any prime race evens out so that $\delta(P_{q;a_1,\dots,a_r}) \rightarrow \frac{1}{r!}$ as $q \rightarrow \infty$, but the proof for $\delta(P_{q;N,R})$ is similar but less technical.

Throughout this chapter, we follow Rubinstein and Sarnak's paper [28], but our exposition is much longer due to the amount of detail. We do not give numerics for the densities, but parts (i) and (ii) already give some exact densities, such as $\delta(P_{5;1,-1}) = \frac{1}{2}$ and $\delta(P_{7;1,2,4}) = \frac{1}{6}$, again under GRH and GSH. We also do not address generalizations to prime ideals or prime geodesics unlike in Rubinstein and Sarnak's paper.

In order to compare the numbers of primes congruent to $a_i \pmod{q}$, define the auxiliary function

$$E(x; q, a) = \frac{\varphi(q)\pi(x; q, a) - \pi(x)}{\sqrt{x}} \log x, \quad (39)$$

which is the scaled excess or deficit of primes congruent to $a \pmod{q}$, and the vector-valued function

$$\mathcal{E}_{q;a_1,\dots,a_r}(x) = (E(x; q, a_1), \dots, E(x; q, a_r)),$$

which is scaled so that by the GRH its norm is bounded by $O(\log^2 x)$ (q -dependence is not indicated). We are studying the prime race in the logarithmic scale, so we also set

$$E_{q;a_1,\dots,a_r}(y) := \mathcal{E}_{q;a_1,\dots,a_r}(e^y).$$

¹⁶It is a conjecture that the the two cases mentioned are the only ones in which the prime race is even, that is $\delta(P_{q;a_{\sigma(1)},\dots,a_{\sigma(r)}}) = \frac{1}{r!}$ for any permutation σ ; see [20].

We use the shorthand notation $E(y)$ whenever there is no danger of confusion.

The existence of the limiting distribution of $E_{q;a_1,\dots,a_r}$ is a crucial step in proving the existence of the logarithmic densities. This limiting distribution is the measure, or distribution, defined as

$$\mu(A) := \lim_{x \rightarrow \infty} \frac{1}{x} m(\{y \in (0, x) : E_{q;a_1,\dots,a_r}(y) \in A\})$$

whenever the limit exists. This is the asymptotic probability that $E_{q;a_1,\dots,a_r}(y) \in A$. However, we do not know a priori at all when this limit exists (except in the trivial case if A or its complement has measure zero). We will prove that the limiting distribution exists for all regular enough sets by first constructing a measure $\mu_{q;a_1,\dots,a_r}$ related to the distribution of $E_{q;a_1,\dots,a_r}$ and then showing that this measure is actually the limiting distribution. Therefore we need the following theorem.

Theorem 4.1. *With the assumption of the GRH, there exists a probability measure $\mu_{q;a_1,\dots,a_r}$ defined on the Borel sets of \mathbb{R}^r such that*

$$\lim_{Y \rightarrow \infty} \frac{1}{Y} \int_2^Y f(E_{q;a_1,\dots,a_r}(y)) dy = \int_{\mathbb{R}^r} f(x) d\mu_{q;a_1,\dots,a_r}(x) \quad (40)$$

for all bounded continuous functions $f : \mathbb{R}^r \rightarrow \mathbb{R}$.¹⁷

Notice that if we could choose f to be the characteristic function of the set $\{x \in \mathbb{R}^r : x_1 > x_2 > \dots > x_r\}$, we would obtain

$$\delta(P_{q;a_1,\dots,a_r}) = \mu_{q;a_1,\dots,a_r}(\{x \in \mathbb{R}^r : x_1 > x_2 > \dots > x_r\}),$$

which would not only prove that $P_{q;a_1,\dots,a_r}$ has a logarithmic density but also give information on its density in terms of $\mu_{q;a_1,\dots,a_r}$. However, a major obstacle for using this is that f was assumed to be bounded and continuous. Hence, we will derive an explicit formula for the Fourier transform $\hat{\mu}_{q;a_1,\dots,a_r}(\xi)$ of the measure and use it to study the properties of $\mu_{q;a_1,\dots,a_r}(x)$.¹⁸

¹⁷We will see that this convergence of integrals is actually equivalent to $\mu_{q;a_1,\dots,a_r}$ being the distribution function of $E_{q;a_1,\dots,a_r}$ if $\mu_{q;a_1,\dots,a_r}$ is shown to be absolutely continuous with respect to the Lebesgue measure.

¹⁸A measure μ naturally gives rise to a function $\mu(x)$, as remarked in Chapter 3, if it is absolutely continuous with respect to the Lebesgue measure. We are able to show this for $\mu_{q;a_1,\dots,a_r}$ assuming the GSH.

4.2 Outline of the Proof

With enough definitions available, we give an outline of the proof of the Main Theorem and the estimate for $\mu_{q;a_1,\dots,a_r}$, which is formulated in subsection 4.8, as their proofs involve numerous steps. The formulations of all the theorems and lemmas involved are found from the places where they are needed.

In order to prove the Main Theorem,

1. We will first prove Theorem 4.1; for this

1.1 We will derive an explicit formula for $E(x; q; a)$ with a small error term, owing to our assumption of the GRH (Lemma 4.2).

1.2 We define $E^T(y)$ for $T \geq 1$ by truncating the explicit formula for the components of $E(y)$. We shall show that the error ε in the approximation $E(y) = E^T(y) + \varepsilon(y)$ has a small mean square (Lemma 4.3).

1.3 We will show (Lemma 4.4) that for the the functions $E^T(y)$ we can find corresponding Borel probability measures ν^T such that

$$\lim_{Y \rightarrow \infty} \frac{1}{Y} \int_2^Y f(E^T(y)) dy = \int_{\mathbb{R}^r} f(x) d\nu^T(x) \quad (41)$$

for all bounded continuous functions $f : \mathbb{R}^r \rightarrow \mathbb{R}$. **1.4** We show (Lemma 4.5) that if f is Lipschitz, then the left-hand side of (40) is equal to $\lim_{T \rightarrow \infty} \nu^T(f)$, where $\nu^T(f)$ is the distribution corresponding to ν^T , and the latter limit is shown to exist and to be finite.

Finally, we use Lemma 3.5 to obtain a measure $\mu_{q;a_1,\dots,a_r}$ so that (40) holds for Lipschitz functions. We get the general case by the portmanteau theorem (Lemma 3.4). After that, the proof of Theorem 4.1 is finished.

2. Next, we prove Theorem 4.7, which is a formula for $\hat{\mu}_{q;a_1,\dots,a_r}$ in terms of Bessel function products, using GSH. In order to do this,

2.1 We start with Lemma 4.6 that will allow us to deal with the Fourier transform of the value distributions of trigonometric polynomials whose frequencies are rationally independent.

2.2 We notice that the components of $E(y)$ are limits of trigonometric polynomials with rationally independent frequencies (by the GSH), and we take Fourier transforms of them and apply the uncorrelatedness result. By simplifying the formula, we arrive at the Bessel product.

3. We prove Theorem 4.8, which extracts some important properties of $\mu_{q;a_1,\dots,a_r}$ from those of its Fourier transform. In particular, $\mu_{q;a_1,\dots,a_r}$ is shown to be absolutely continuous with respect to the r -dimensional Lebesgue measure (or the Lebesgue measure on the hyperplane $x_1 + \dots + x_{\varphi(q)} = 0$ in the

case $r = \varphi(q)$), and the Radon-Nikodym derivative $\mu_{q;a_1,\dots,a_r}(x)$ is shown to extend to an entire function on \mathbb{C}^r (or to an analytic function on a hyperplane if $r = \varphi(q)$). Furthermore, $\mu_{q;a_1,\dots,a_r}$ is seen to be symmetric with respect to the point $(-c(q; a_1), \dots, -c(q; a_r))$, and this is crucial in proving several parts of the Main Theorem.¹⁹

After that, we are in a position to prove all the parts of the Main Theorem except the vanishing of the bias as $q \rightarrow \infty$. We already concluded that $\mu_{q;a_1,\dots,a_r}$ is absolutely continuous, and this together with the portmanteau theorem tells that $\mu_{q;a_1,\dots,a_r}$ is actually the limiting distribution of $E_{q;a_1,\dots,a_r}(y)$ for sets whose boundary has zero Lebesgue measure (in particular, the limiting distribution exists in this case). Moreover, the Radon-Nikodym derivative does not vanish in a "large" set (as it is seen to be entire), so we may use its symmetry to prove the desired inequalities for the densities. We get the stronger statement about the set where $\pi(x; q, a_{i+1}) - \pi(x; q, a_i) > \frac{\sqrt{x}}{\eta(x) \log x}$ almost for free, owing to our scaling of $E_{q;a_1,\dots,a_r}(x)$.

The reason for the part of the main theorem concerning $r = 3$ to be true is that then $\mu_{q;a_1,\dots,a_r}(x)$ is invariant under permutations of the a_i 's.²⁰

4. Lastly, we show that the bias in $P_{q;N;R}$ (which should be the most biased prime race (mod q)) vanishes as $q \rightarrow \infty$. To show this, we estimate the logarithm of $\hat{\mu}_{q;N,R}(\xi)$ (which also has a Bessel product formula), and show that it approaches $-\frac{\xi^2}{2}$ as $q \rightarrow \infty$ uniformly on compact sets. We conclude that $\mu_{q;N,R}(\xi)$ approaches the Gaussian $\frac{1}{\sqrt{2\pi}}e^{-\frac{\xi^2}{2}}$ in the L^1 -norm. As a corollary, we see that $\delta(P_{q;N,R}) \rightarrow \frac{1}{2}$ as $q \rightarrow \infty$.

In addition to the Main Theorem, we shall prove

5. The tail of the measure $\mu_{q;a_1,\dots,a_r}$, that is $\mu_{q;a_1,\dots,a_r}(|x| \geq R)$ where R is large, is bounded from above by $\exp(-C_q\sqrt{R})$ and from below by $\exp(-C_q \exp(R))$. This is Theorem 4.10.

5.1 The upper bound follows just by considering the measures ν^T constructed in Lemma 4.4.

5.2 For the lower bound, which is only proved for $\mu_{q;N,R}$ as the proof is technical, we study the function $F_\varepsilon(\xi)$, which is the average of $R(e^x)$ over the range $[\xi - \frac{\varepsilon}{2}, \xi + \frac{\varepsilon}{2}]$ and $R(y)$ is the scaled excess of quadratic nonresidue primes up to y . The parameter ε is very small. We show (Lemma 4.11) that

¹⁹ f is symmetric with respect to a if $f(a-x) = f(a+x)$ for all $x \in \mathbb{R}^r$.

²⁰ We could also show that these are the only cases when $\mu_{q;a_1,\dots,a_r}$ is symmetric in the a_i 's, but the lack of symmetry would tell nothing about the densities.

if

$$\max_{0 \leq \gamma \leq \varepsilon^{-1}} \left\| \gamma m \frac{\varepsilon}{2} \right\| \leq \frac{d}{\log \varepsilon^{-1}}, \quad (42)$$

then $F_\varepsilon(\frac{m+1}{2}\varepsilon) > c\varepsilon^{-1} \log \varepsilon^{-1}$, where $c, d > 0$ are some constants and $\|\cdot\|$ denotes the distance to the nearest multiple of 2π and γ denotes a zero of the function $L(s, \chi_1)$.

5.3 We denote the set of integers $m \in [\frac{2}{\varepsilon}, \frac{M}{\varepsilon}]$ satisfying (42) by G_M and apply the pigeonhole principle to show that the size of G_M grows linearly in M .

5.4 We use Lemma 4.12 to prove that for $m \in G_M$ the interval $[(m+\frac{1}{2})\frac{\varepsilon}{2}, (m+\frac{3}{2})\frac{\varepsilon}{2}]$ has a subset of not too small measure in which $R(e^x)$ is large.

5.5 The previous lemma together with the bound on $|G_M|$ proves that $R(e^x)$ is large in a subset of $[0, M]$, whose measure we can bound from below.

5.6 Letting $M \rightarrow \infty$, the measure $\mu_{q;N,R}$ arises as the limiting distribution of $R(e^x)$, and we get the desired estimate for it.

4.3 Existence of the Limiting Distribution

Recall from Chapter 2 the definition

$$\psi(x; \chi) = \sum_{n \leq x} \chi(n) \Lambda(n)$$

where Λ is von Mangoldt's function. From Chapter 2 we have the explicit formula

$$\psi(x, \chi) = - \sum_{|\rho| \leq X} \frac{x^\rho}{\rho} + O\left(\frac{x \log^2(xX)}{X} + \log x\right)$$

for $x \geq 2$ and $X \geq 1$, where $\rho = \beta_\chi + i\gamma_\chi$ runs through all the zeros of $L(s, \chi)$ on $0 < \Re(s) < 1$. Assuming the GRH, we may suppose $\beta_\chi = \frac{1}{2}$ to get

$$\psi(x, \chi) = -\sqrt{x} \sum_{|\gamma_\chi| \leq X} \frac{x^{i\gamma_\chi}}{\frac{1}{2} + i\gamma_\chi} + O\left(\frac{x \log^2(xX)}{X} + \log x\right). \quad (43)$$

Let $E(x; q, a) = \frac{\log x}{\sqrt{x}}(\varphi(q)\pi(x; q, a) - \pi(x))$ and $c(q, a)$ as before. We need the following lemma

Lemma 4.2. *We have* ²¹

$$E(x; q, a) = -c(q, a) + \sum_{x \neq x_0} \bar{\chi}(a) \frac{\psi(x, \chi)}{\sqrt{x}} + O\left(\frac{1}{\log x}\right). \quad (44)$$

²¹In what follows, q is a constant so we are not concerned with the q -dependence of error terms.

Proof. First observe that

$$\begin{aligned}\psi(x; q, a) &= \sum_{\substack{n \leq x \\ n \equiv a \pmod{q}}} \Lambda(n) = \frac{1}{\varphi(q)} \sum_{n \leq x} \Lambda(n) \sum_{\chi \pmod{q}} \chi(n) \bar{\chi}(a) \\ &= \frac{1}{\varphi(q)} \sum_{\chi \pmod{q}} \bar{\chi}(a) \psi(x, \chi)\end{aligned}$$

since $\frac{1}{\varphi(q)} \sum_{\chi \pmod{q}} \bar{\chi}(a) \chi(n) = 1$ if $n \equiv a \pmod{q}$ and zero otherwise.

Denoting

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

and $\theta(x) := \theta(x; 1, 0)$, we have

$$\begin{aligned}\psi(x; q, a) - \theta(x; q, a) &= \sum_{\substack{p^2 \equiv a \pmod{q} \\ p \leq \sqrt{x}}} \log p + O(\sqrt[3]{x} \log^2 x) \\ &= (c(q, a) + 1) \frac{\sqrt{x}}{\phi(q)} + O\left(\frac{\sqrt{x}}{\log x}\right)\end{aligned}\quad (45)$$

by the prime number theorem in arithmetic progressions, footnote 4 and the definition of $c(q, a)$. Also, applying partial summation to the characteristic function $1_{\mathbb{P}}$ of primes and to the characteristic function $1_{\mathbb{P}_{q,a}}$ of the primes in an arithmetic progression, we get

$$\begin{aligned}E(x; q, a) &= \frac{\log x}{\sqrt{x}} (\varphi(q) \pi(x; q, a) - \pi(x)) \\ &= \frac{\log x}{\sqrt{x}} \sum_{n \leq x} (\varphi(q) 1_{\mathbb{P}_{q,a}}(n) \log n - 1_{\mathbb{P}}(n) \log n) (\log n)^{-1} \\ &= \frac{\varphi(q) \theta(x; q, a) - \theta(x)}{\sqrt{x}} + \frac{\log x}{\sqrt{x}} \int_2^x \frac{\varphi(q) \theta(t; q, a) - \theta(t)}{t \log^2 t} dt.\end{aligned}\quad (46)$$

In the above formula, using (45) and the orthogonality of characters, the first term is

$$\begin{aligned}&\frac{\varphi(q) \psi(x; q, a) - \sqrt{x} (c(q, a) + 1) - \psi(x) + \sqrt{x} + O\left(\frac{\sqrt{x}}{\log x}\right)}{\sqrt{x}} \\ &= -c(q, a) + \frac{\sum_{\chi \pmod{q}} \bar{\chi}(a) \psi(x, \chi) - \psi(x)}{\sqrt{x}} + O\left(\frac{1}{\log x}\right) \\ &= -c(q, a) + \frac{\sum_{\chi \neq \chi_0 \pmod{q}} \bar{\chi}(a) \psi(x, \chi)}{\sqrt{x}} + O\left(\frac{1}{\log x}\right),\end{aligned}\quad (47)$$

which is what Lemma 1 requires. We now show that the integral term of (46) can be included into the error term $O\left(\frac{1}{\log x}\right)$. We have

$$\begin{aligned} & \frac{\log x}{\sqrt{x}} \int_2^x \frac{\varphi(q)\theta(t; q, a) - \theta(t)}{t \log^2 t} dt \\ &= \frac{\log x}{\sqrt{x}} \int_2^x \frac{-c(q, a) + t^{-\frac{1}{2}} \sum_{\chi \neq \chi_0 \pmod{q}} \bar{\chi}(a) \psi(t, \chi) + O\left(\frac{1}{\log t}\right)}{\sqrt{t} \log^2 t} dt. \end{aligned} \quad (48)$$

using (47). Since

$$\begin{aligned} \int_2^x \frac{1}{\sqrt{t} \log^2 t} dt &= \left[\frac{2\sqrt{t}}{\log^2 t} \right]_2^x + \int_2^x \frac{4}{\sqrt{t} \log^3 t} \\ &\ll \frac{\sqrt{x}}{\log^2 x} + \sum_{k \leq \log_2 x} \int_{2^{k-1}}^{2^k} \frac{1}{\sqrt{t} \log^3 t} \\ &\ll \frac{\sqrt{x}}{\log^2 x} + \sum_{k \leq \log_2 x} \frac{2^{\frac{k}{2}}}{k^3} \ll \frac{\sqrt{x}}{\log^2 x}, \end{aligned} \quad (49)$$

we can neglect the terms $-c(q, a)$ and $O\left(\frac{1}{\log t}\right)$ in the numerator of the integrand in (48). Dealing with the rest of the terms inside the integral must be done a bit more carefully, since the inequality $\psi(x, \chi) \ll \sqrt{x} \log^2 x$ is not sufficiently strong now. However, we show that $\psi(x, \chi)$ is on average $\ll \sqrt{x}$, that is,

$$\psi_1(x, \chi) := \int_2^x \psi(t, \chi) dt \ll x^{\frac{3}{2}}. \quad (50)$$

Indeed, choosing $X = x$ in (43) and integrating over $[2, x]$ yields

$$\begin{aligned} \psi_1(x, \chi) &= \sum_{|\gamma_\chi| \leq x} \int_2^x \frac{t^{\frac{1}{2} + i\gamma_\chi}}{\frac{1}{2} + i\gamma_\chi} dt + O(x \log^2 x) \\ &\ll x^{\frac{3}{2}} \sum_{|\gamma_\chi| \leq x} \frac{1}{\left| \frac{1}{2} + i\gamma_\chi \right|^{\frac{3}{2} + i\gamma_\chi}} + O(x \log^2 x) \ll x^{\frac{3}{2}} \end{aligned}$$

since $\sum_{\gamma_\chi} \frac{1}{\frac{1}{4} + \gamma_\chi^2}$ converges.

Due to this, partial integration and (49) give

$$\begin{aligned}
\int_2^x \frac{\psi(t, \chi)}{t \log^2 t} dt &= \left[\frac{\psi_1(t, \chi)}{t \log^2 t} \right]_2^x + \int_2^x \frac{\psi_1(t, \chi)(\log t + 2)}{t^2 \log^3 t} dt \\
&\ll \frac{\sqrt{x}}{\log^2 x} + \int_2^x \frac{1}{\sqrt{t} \log^2 t} dt \\
&\ll \frac{\sqrt{x}}{\log^2 x}
\end{aligned}$$

using (49). Hence all the terms in the integral in (48) are $\ll \frac{\sqrt{x}}{\log^2 x}$. This completes the proof of the lemma. \square

By combining the explicit formula for $\psi(x; \chi)$ with the previous lemma, we obtain

$$E(x; q, a) = -c(q, a) - \sum_{x \neq x_0} \bar{\chi}(a) \sum_{|\gamma_\chi| \leq T} \frac{x^{i\gamma_\chi}}{\frac{1}{2} + i\gamma_\chi} + \varepsilon_a(x, T, X)$$

for $2 \leq x \leq T \leq X$, where

$$\varepsilon_a(x, T, X) := - \sum_{x \neq x_0} \bar{\chi}(a) \sum_{T \leq |\gamma_\chi| \leq X} \frac{x^{i\gamma_\chi}}{\frac{1}{2} + i\gamma_\chi} + O\left(\frac{\sqrt{x} \log^2 X}{X} + \frac{1}{\log x}\right). \quad (51)$$

We shall prove that the remainder term $\varepsilon_a(x, T, X)$ has a small mean square.

Lemma 4.3. *For $Y \geq 2$, we have the inequality*

$$\int_2^Y |\varepsilon_a(e^y, T, e^Y)|^2 dy \ll Y \frac{\log^4 T}{T} + \frac{\log^4 T}{T}.$$

Proof. When we substitute $x = e^y$, $X = e^Y$ and expand the square of (51), we get the square of the sum term, the square of the error term, and twice their product. The integral of the square of the error term is

$$\ll \int_2^Y \left(\frac{Y^2 e^{\frac{y}{2}}}{e^Y} + \frac{1}{y} \right)^2 dy \ll \int_2^Y \left(\frac{1}{e^{\frac{y}{2}}} + \frac{1}{y} \right)^2 dy \ll 1.$$

The integral of the product of the sum term and the error term multiplied by two is by the Cauchy-Schwarz inequality

$$\begin{aligned} &\ll 2 \int_2^Y \left(\frac{1}{e^{\frac{y}{2}}} + \frac{1}{y} \right) \sum_{\chi \neq \chi_0} \bar{\chi}(a) \sum_{T \leq |\gamma_\chi| \leq e^Y} \frac{e^{i\gamma_\chi y}}{\frac{1}{2} + i\gamma_\chi} dy \\ &\ll \left(\int_2^Y \left(\frac{1}{e^{\frac{y}{2}}} + \frac{1}{y} \right)^2 dy \right)^{\frac{1}{2}} \left(\int_2^Y \left| \sum_{\chi \neq \chi_0} \bar{\chi}(a) \sum_{T \leq |\gamma_\chi| \leq e^Y} \frac{e^{i\gamma_\chi y}}{\frac{1}{2} + i\gamma_\chi} \right|^2 dy \right)^{\frac{1}{2}}. \end{aligned}$$

The first factor above is bounded by a constant while the second one is the square root of the main term.

Hence it is enough to estimate the main term — the integrated square of the sum in (51):

$$\begin{aligned} &\int_2^Y \left| \sum_{\chi \neq \chi_0} \bar{\chi}(a) \sum_{T \leq |\gamma_\chi| \leq e^Y} \frac{e^{iy\gamma_\chi}}{\frac{1}{2} + i\gamma_\chi} \right|^2 dy \\ &= \int_2^Y \sum_{\substack{\chi \neq \chi_0 \\ \lambda \neq \chi_0}} \bar{\chi}(a) \lambda(a) \sum_{\substack{T \leq |\gamma_\chi| \leq e^Y \\ T \leq |\gamma_\lambda| \leq e^Y}} \frac{e^{iy(\gamma_\chi - \gamma_\lambda)}}{(\frac{1}{2} + i\gamma_\chi)(\frac{1}{2} - i\gamma_\lambda)} \\ &\ll \sum_{\substack{\chi \neq \chi_0 \\ \lambda \neq \chi_0}} \sum_{\substack{T \leq |\gamma_\chi| \leq e^Y \\ T \leq |\gamma_\lambda| \leq e^Y}} \frac{1}{|\gamma_\chi| |\gamma_\lambda|} \min \left(Y, \frac{1}{|\gamma_\chi - \gamma_\lambda|} \right) \end{aligned} \quad (52)$$

by interchanging the sum and the integral and using the standard estimate $|\int_a^b e^{it} dt| \leq \min(|b - a|, \frac{1}{|a|} + \frac{1}{|b|})$.

We use the asymptotic formula from Chapter 2 for the n th zero in the order of magnitude: $|\gamma_\chi(n)| \sim \frac{n\pi}{\log n}$ for the imaginary part of the n th zero in the order of increasing magnitude to find that the sum (52) is less than

$$\begin{aligned} &\ll \sum_{\substack{\chi \neq \chi_0 \\ \lambda \neq \chi_0}} \sum_{\substack{T \leq |\gamma_\chi(m)| \\ T \leq |\gamma_\lambda(n)|}} \frac{1}{|\gamma_\chi(m)| |\gamma_\lambda(n)|} \min \left(Y, \frac{1}{|\gamma_\chi(m) - \gamma_\lambda(n)|} \right) \\ &\ll \sum_{\substack{\chi \neq \chi_0 \\ \lambda \neq \chi_0}} \sum_{\substack{m/\log m \geq T \\ n/\log n \geq T}} \frac{\log m \log n}{mn} \min \left(Y, \frac{1}{|\gamma_\chi(m) - \gamma_\lambda(n)|} \right) \\ &\ll \sum_{\substack{\chi \neq \chi_0 \\ \lambda \neq \chi_0}} \sum_{\substack{m \geq T \log T \\ n \geq T \log T}} \frac{\log m \log n}{mn} \min \left(Y, \frac{1}{|\gamma_\chi(m) - \gamma_\lambda(n)|} \right). \end{aligned} \quad (53)$$

As the asymptotic $\gamma_\chi(n) \sim \frac{\pi n}{\log n}$ for the n th zero is not strong enough to give any bounds on $|\gamma_\chi(n) - \gamma_\lambda(m)|^{-1}$ (and ignoring the minimum in (52) produces a divergent series), one needs to be a bit careful in estimating the sum.

The asymptotic formula for $N(T)$ together with the observation $n = N(|\gamma_\chi(n)|) + O(1)$ immediately gives

$$\pi n + O(\log n) = |\gamma_\chi(n)|(\log |\gamma_\chi(n)| + c(q))$$

for some constant $c(q)$, and we solve

$$|\gamma_\chi(n)| = \frac{\pi n + O(\log n)}{W(e^{c(q)}(\pi n + O(\log n)))},$$

where W is the Lambert function defined as the real solution to $W(x)e^{W(x)} = x$ for $x \geq -e^{-1}$.²²

Implicit differentiation gives $W'(x) = \frac{1}{e^{W(x)} + x} = \frac{1}{x + \frac{x}{W(x)}}$. In particular,

$$W(x+h) - W(x) = \frac{h}{x + \xi + \frac{x+\xi}{W(x+\xi)}}$$

for some $\xi \in [0, h]$ by the mean value theorem. If $h \ll \log x$, we get $|W(x+h) - W(x)| \ll \frac{\log x}{x}$. Therefore,

$$|\gamma_\chi(n)| = \frac{\pi n}{W(e^{c(q)}\pi n) + O(\frac{\log n}{n})} + O(\log n) = \frac{\pi n}{W(e^{c(q)}\pi n)} + O(\log n).$$

As the derivative of $h(x) := \frac{x}{W(ax)}$ is $\frac{1}{W(ax)+1}$, we get for $a > 0$, x, y large enough, and $x - y \geq \log^3 x$ that

$$h(x) - h(y) \geq \frac{x-y}{1+W(ax)} \gg \frac{x-y}{\log x}$$

using the elementary estimate $W(x) \ll \log x$. We thus arrived at the following useful bound:

$$\begin{aligned} \frac{1}{|\gamma_\chi(m) - \gamma_\lambda(n)|} &\ll \frac{1}{h(\pi m) - h(\pi n) + O(\log m)} \\ &\ll \frac{1}{\frac{m-n}{\log m} + \log m} \\ &\ll \frac{\log m \log n}{|m-n|} \end{aligned} \tag{54}$$

²²Indeed, $A = x(\log x + c)$ is equivalent to $Ae^c = e^c x \log(e^c x)$, which by substituting $e^y = e^c x$ becomes $ye^y = Ae^c$, so $y = W(Ae^c)$ and $x = e^{-c} e^{W(Ae^c)} = \frac{A}{W(Ae^c)}$. Clearly $W(x) \asymp \log x$.

for $|m - n| \geq \log^3(\max\{m, n\})$ and m, n greater than a constant. By symmetry, it suffices to consider the terms $n \leq m$, and the contribution of the terms with $n \in [m - \log^3 m, m]$ in (53) is at most

$$\begin{aligned}
&\ll Y \sum_{m \geq T \log T} \sum_{n \in [m - \log^3 m, m]} \frac{\log m \log n}{mn} \\
&\ll Y \int_{T \log T}^{\infty} \int_{x - \log^3 x}^x \frac{\log x \log y}{xy} dx dy \\
&\ll Y \int_{T \log T}^{\infty} \frac{\log x}{x} (\log^3 x) \frac{\log x}{x} dx \\
&\ll Y \int_{T \log T}^{\infty} \frac{\log^5 x}{x^2} dx \\
&\ll Y \frac{\log^4 T}{T}.
\end{aligned}$$

To see the validity of the last estimate, we consider more generally $I_n(X) = \int_X^{\infty} \frac{\log^n x}{x^2} dx$. Applying the substitution $x = e^y$ and integrating by parts we have

$$\begin{aligned}
I_n(X) &= \int_{\log X}^{\infty} y^n e^{-y} dy = O\left(\frac{\log^n X}{X}\right) + \int_{\log X}^{\infty} n y^{n-1} e^{-y} dy \\
&= n I_{n-1}(X) + O\left(\frac{\log^n X}{X}\right),
\end{aligned}$$

so $I_0(X) \ll \frac{1}{X}$ implies $I_n(X) \ll \frac{\log^n X}{X}$, and in particular $I_5(T \log T) \ll \frac{\log^4 T}{T}$.

Thus it suffices to consider the terms satisfying (54).

Using (54) for $|m - n| \geq 1$ and denoting $T' = T \log T$, (53) is bounded by

$$\begin{aligned}
&\ll Y \frac{\log^4 T}{T} + \sum_{\substack{m, n \geq T' \\ m \neq n}} \frac{\log m \log n}{mn} \cdot \frac{\log m \log n}{|m - n|} \\
&\ll Y \frac{\log^4 T}{T} + \iint_{\substack{x, y \geq T' \\ |y - x| > 1}} \frac{\log^2 x \log^2 y}{xy|y - x|} dx dy. \tag{55}
\end{aligned}$$

We may assume that T' is large enough and consider the integral only for

$y > x$ by symmetry. We have

$$\begin{aligned} \int_{\substack{x, y \geq T' \\ y \geq x+1}} \frac{\log^2 x \log^2 y}{xy|y-x|} dx dy &= \int_{T'}^{\infty} \frac{\log^2 x}{x} \int_{x+1}^{\infty} \frac{\log^2 y}{y|y-x|} dy dx \\ &= \int_{T'}^{\infty} \frac{\log^2 x}{x} \int_1^{\infty} \frac{\log^2(t+x)}{(t+x)t} dt dx. \end{aligned} \quad (56)$$

Here the inner integral is bounded as follows:

$$\begin{aligned} \int_1^{\infty} \frac{\log^2(t+x)}{(t+x)t} dt &= \int_1^x \frac{\log^2(t+x)}{(t+x)t} dt + \int_x^{\infty} \frac{\log^2(t+x)}{(t+x)t} dt \\ &\leq \log^2(2x) \int_1^x \frac{1}{(t+x)t} dt + \int_x^{\infty} \frac{2\log^2(t+x)}{(t+x)^2} dt \\ &\ll \frac{\log^2 x}{x} \int_1^x \frac{1}{t} dt + \frac{\log^2 x}{x} \\ &\ll \frac{\log^3 x}{x}, \end{aligned}$$

where we used $\frac{1}{t(t+x)} \leq \frac{2}{(t+x)^2}$ for $t \geq x$ and $I_2(X) \ll \frac{\log^2 X}{X}$. Thus the integral (56) is dominated by

$$\int_{T'}^{\infty} \frac{\log^5 x}{x^2} dx = I_5(T') \ll \frac{\log^5 T'}{T'} \ll \frac{\log^4 T}{T}$$

since $T' = T \log T$. This completes the proof. \square

Based on the previous lemmas, we approximate $E(y)$ with

$$E^T(y) := (E^T(e^y; q, a_1), \dots, E^T(e^y; q, a_r)), \quad (57)$$

where $E^T(e^y; q, a_i)$ is the truncated sum

$$E^T(e^y; q, a_i) = -c(q, a_i) - \sum_{x \neq x_0} \bar{\chi}(a_i) \sum_{|\gamma_x| \leq T} \frac{e^{iy\gamma_x}}{\frac{1}{2} + \gamma_x}.$$

In order to prove Theorem 4.1, we will show that

$$\lim_{Y \rightarrow \infty} \frac{1}{Y} \int_2^Y f(E^T(y)) dy \quad (58)$$

is given by integration against a Borel measure, and we use this later to deduce the same formula for $E(y)$.

Lemma 4.4. *For every $T \geq 2$, there is a Borel probability measure ν^T such that*

$$\lim_{Y \rightarrow \infty} \frac{1}{Y} \int_2^Y f(E^T(y)) dy = \int_{\mathbb{R}^r} f(x) d\nu^T(x). \quad (59)$$

for every bounded continuous function $f : \mathbb{R}^r \rightarrow \mathbb{R}$. Moreover, the support of ν^T is contained in $B(0, c_q \log^2 T)$ for some constant $c_q > 0$.

Proof. Let us first derive a representation of $E^T(y)$ as a finite sum of complex exponentials, which depend on the zeros of the L -functions. Let $\gamma_1, \dots, \gamma_N$ be the imaginary parts of all the zeros of the functions $L(s, \chi)$ on $\Re(s) = \frac{1}{2}$ where $\chi \neq \chi_0$ and $0 \leq \gamma_k \leq T$. Since $L(\frac{1}{2} + i\gamma_k, \chi_k) = 0$ (we denote by χ_k the character corresponding to the zero $\frac{1}{2} + i\gamma_k$) if and only if $L(\frac{1}{2} - i\gamma_k, \bar{\chi}_k) = 0$, we may pair up the zeros $\frac{1}{2} + i\gamma_k$ with their conjugates so that $E^T(e^y; q, a_i)$ becomes

$$\begin{aligned} & -c(q, a_i) - \sum_{k=1}^N \left(\bar{\chi}_k(a_i) \frac{e^{i\gamma_k y}}{\frac{1}{2} + i\gamma_k} - \chi_k(a_i) \frac{e^{-i\gamma_k y}}{\frac{1}{2} - i\gamma_k} \right) \\ & = -c(q, a_i) + \sum_{k=1}^N \left(2\Re \left(\bar{\chi}_k(a_i) \frac{e^{i\gamma_k y}}{\frac{1}{2} + i\gamma_k} \right) \right). \end{aligned}$$

Hence

$$E^T(y) = 2\Re \left(\sum_{k=1}^N b_k e^{i\gamma_k y} \right) + b_0, \quad (60)$$

where

$$b_0 = (-c(q, a_1), \dots, -c(q, a_r))$$

and

$$b_k = - \left(\bar{\chi}_k(a_1) \frac{1}{\frac{1}{2} + i\gamma_k}, \dots, \bar{\chi}_k(a_r) \frac{1}{\frac{1}{2} + i\gamma_k} \right),$$

and the real part of a vector naturally means the vector formed by the real parts of its components. Let $g : \mathbb{R}^N / \mathbb{Z}^N \rightarrow \mathbb{R}$ be the continuous 1-periodic function defined by

$$g(y_1, \dots, y_N) = f \left(2\Re \left(\sum_{k=1}^N b_k e^{2\pi i y_k} \right) + b_0 \right). \quad (61)$$

Then (58) is the limit

$$\lim_{Y \rightarrow \infty} \frac{1}{Y} \int_2^Y g\left(\frac{\gamma_1 y}{2\pi}, \dots, \frac{\gamma_N y}{2\pi}\right) dy. \quad (62)$$

The Kronecker-Weyl theorem (see the previous chapter) tells us that if we denote by G the subgroup

$$G = \left\{ \left(\frac{\gamma_1 y}{2\pi}, \dots, \frac{\gamma_N y}{2\pi} \right) \pmod{1} : y \in \mathbb{R} \right\}$$

of the torus $\mathbb{R}^N/\mathbb{Z}^N$ and by \bar{G} is its topological closure, then (62) converges to

$$\int_{\bar{G}} g(x) dm_{\bar{G}}(x)$$

where $m_{\bar{G}}$ is the normalized Haar measure on the set \bar{G} (which is itself isomorphic to some torus $\mathbb{R}^k/\mathbb{Z}^k$, where k is the dimension of the space spanned by $\gamma_1, \dots, \gamma_N$, again by the Kronecker-Weyl theorem).

Now the map

$$f \mapsto \int_{\bar{G}} g(x) dm_{\bar{G}}(x)$$

is a positive linear functional defined on bounded continuous functions (f and g are related by (61)), so by the Riesz representation theorem (see Chapter 3), the functional may be written as

$$\int_{\bar{G}} g(x) dm_{\bar{G}}(x) = \int_{\mathbb{R}^r} f(x) d\nu^T(x), \quad (63)$$

where ν^T is a Borel measure (the definition of g depends on T) and f is compactly supported and continuous. Since

$$|E^T(y)| \ll \sum_{|\gamma_x| \leq T} \frac{1}{|\frac{1}{2} + i\gamma_x|} \ll \log^2 T,$$

the definition of ν^T is irrelevant outside some ball $B(0, c_q \log^2 T)$, so we may assume that the support of the measure is contained in such a ball. For the same reason, the values of f outside some compact set are irrelevant in (58), so we get the case of bounded continuous functions f . By choosing $f \equiv 1$ (that is, $g \equiv 1$) in (63), we see that ν^T is a probability measure. This finishes the proof. \square

We will first prove Theorem 4.1 in the case where f is a Lipschitz function, and subsequently use it to prove the general case.

Lemma 4.5. For Lipschitz functions $f : \mathbb{R}^r \rightarrow \mathbb{R}$ we have

$$\lim_{Y \rightarrow \infty} \frac{1}{Y} \int_2^Y f(E(y)) dy = \lim_{T \rightarrow \infty} \nu^T(f),$$

where

$$\nu^T(f) := \int_{\mathbb{R}^r} f(x) d\nu^T(x)$$

is the distribution associated with ν^T and the limits are finite.

Proof. Now that we have Lemmas 4.2 and 4.3, we may calculate, denoting $\varepsilon^T(y) := E(y) - E^T(y)$ and c_f for a Lipschitz constant of f , that

$$\begin{aligned} & \frac{1}{Y} \int_2^Y f(E(y)) dy \\ &= \frac{1}{Y} \int_2^Y f(E^T(y) + \varepsilon^T(y)) dy \\ &= \frac{1}{Y} \int_2^Y f(E^T(y)) dy + O\left(\frac{c_f}{Y} \int_2^Y |\varepsilon^T(y)| dy\right) \\ &= \frac{1}{Y} \int_2^Y f(E^T(y)) dy + O\left(\frac{c_f}{\sqrt{Y}} \left(\int_2^Y |\varepsilon^T(y)|^2 dy\right)^{\frac{1}{2}}\right) \end{aligned}$$

using Jensen's inequality for the function $x \mapsto x^2$. By Lemma 4.2, we can continue to estimate

$$\begin{aligned} \frac{1}{Y} \int_2^Y f(E(y)) dy &= \frac{1}{Y} \int_2^Y f(E^T(y)) dy + O\left(c_f \left(\frac{\log^2 T}{\sqrt{T}} + \frac{\log^2 T}{\sqrt{YT}}\right)\right) \\ &\rightarrow \nu^T(f) + O\left(c_f \left(\frac{\log^2 T}{\sqrt{T}}\right)\right) \end{aligned} \tag{64}$$

as $Y \rightarrow \infty$, because of Lemma 4.3. Therefore,

$$\begin{aligned} & \nu^T(f) + O\left(c_f \left(\frac{\log^2 T}{\sqrt{T}}\right)\right) \\ &\leq \liminf_{Y \rightarrow \infty} \frac{1}{Y} \int_2^Y f(E(y)) dy \\ &\leq \limsup_{Y \rightarrow \infty} \frac{1}{Y} \int_2^Y f(E(y)) dy \\ &\leq \nu^T(f) + O\left(c_f \left(\frac{\log^2 T}{\sqrt{T}}\right)\right), \end{aligned}$$

which implies that the lim inf and lim sup are finite, and as $T \rightarrow \infty$ we find that

$$\lim_{Y \rightarrow \infty} \frac{1}{Y} \int_2^Y f(E(y)) dy$$

exists and equals

$$\lim_{T \rightarrow \infty} \nu^T(f),$$

and both limits are finite since the difference of (58) from $\nu^T(f)$ is at most a constant for a fixed T . \square

Now we derive Theorem 4.1 for Lipschitz functions with the help of the lemma on functionals from Chapter 3.²³ We may choose $\nu_n := \nu^n$ (ν^T only depends on $\lfloor T \rfloor$), and then Lemma 4.5 together with Lemma 3.5 tells that $\lim_{T \rightarrow \infty} \nu^T(f)$ arises from a Borel probability measure μ , so Theorem 4.1 follows in the Lipschitz case.

Now we prove the general case. We know that

$$\lim_{Y \rightarrow \infty} \frac{1}{Y} \int_2^Y f(E(y)) dy = \lim_{T \rightarrow \infty} \int_{\mathbb{R}^r} f(x) \nu^T(x) dx = \int_{\mathbb{R}^r} f(x) d\mu(x) \quad (65)$$

for Lipschitz functions $f : \mathbb{R}^r \rightarrow \mathbb{R}$. Let μ_Y be the Borel probability measure

$$\mu_Y(A) := \frac{1}{Y} m(\{y \in [0, Y] : E(y) \in A\}).$$

It is a Borel measure because the Lebesgue measure m is, and it is clearly a probability measure as well. Using the definition of the integral, we see that

$$\int_{\mathbb{R}^r} 1_A(x) d\mu_Y(x) = \frac{1}{Y} \int_0^Y 1_A(E(y)) dy,$$

where 1_A is the characteristic function of A . By linearity, the same holds for simple functions, and applying monotone convergence yields

$$\int_{\mathbb{R}^r} f(x) d\mu_Y(x) = \frac{1}{Y} \int_0^Y f(E(y)) dy,$$

and when this is substituted to (65), the portmanteau theorem tells that (65) holds also for all bounded continuous functions on \mathbb{R}^r , completing the proof of Theorem 4.1, with $\mu_{q; a_1, \dots, a_r} := \mu$. \square

²³The following argument and the lemma on functionals are not written explicitly in Rubinstein and Sarnak's paper.

4.4 A formula for the Fourier transform of the Measure of the Prime Race

As mentioned in the outline of the proof of the Main Theorem, we derive a representation for the Fourier transform $\hat{\mu}_{q;a_1,\dots,a_r}$ as a product involving the Bessel function and the zeros of the L -functions. This formula allows us to prove several properties of $\mu_{q;a_1,\dots,a_r}$, which will be used to prove inequalities about the logarithmic densities. It is not a coincidence that a product formula exists for the Fourier transform of the density function of a sum, in this case $E(y)$, whose terms are expected to behave randomly. We formulate a lemma on the Fourier transforms of asymptotically independent random variables, which we will soon apply to the explicit formula.

Lemma 4.6. *Let $X_1, \dots, X_n : \mathbb{R} \rightarrow \mathbb{R}^r$ be random variables. Then*

$$E_P(e^{i(X_1+\dots+X_n)\cdot\xi}) = \prod_{k=1}^n E_P(e^{iX_k\cdot\xi}), \quad (66)$$

where E_P is the expectation with respect to the asymptotic density on \mathbb{R} , $\xi \in \mathbb{R}^r$, and \cdot is the inner product.

Proof. By definition,

$$E_P(e^{i(X_1+\dots+X_n)\cdot\xi}) = \int_{\mathbb{R}^r} e^{ix\cdot\xi} p_{X_1+\dots+X_n}(x) dx,$$

where p_Z is the density function of Z with respect to P . Since X_1, \dots, X_n are independent with respect to P , it is well-known that $p_{X_1+\dots+X_n} = p_{X_1} * \dots * p_{X_n}$ (see [1], pages 7-8; the claim is formulated for measures there, but the proof does not use countable additivity). Since the Fourier transform of a convolution is the product of the Fourier transforms, we obtain

$$E_P(e^{i(X_1+\dots+X_n)\cdot\xi}) = \widehat{p_{X_1}} \cdots \widehat{p_{X_n}}.$$

Again by definition $\widehat{p_{X_k}} = E_P(e^{iX_k\cdot\xi})$, so the claim follows. \square

Now we formulate the product formula for the Fourier transform of the measure $\mu_{q;a_1,\dots,a_r}$.

Theorem 4.7. *Under the GRH and the GSH, for $\xi \in \mathbb{R}^r$ we have*

$$\hat{\mu}_{q;a_1,\dots,a_r}(\xi) = \exp\left(i \sum_{j=1}^r c(q, a_j) \xi_j\right) \prod_{\substack{x \pmod{q} \\ x \neq x_0}} \prod_{\gamma_x > 0} J_0\left(\frac{2|\sum_{j=1}^r \chi(a_j) \xi_j|}{\sqrt{\frac{1}{4} + \gamma_x^2}}\right), \quad (67)$$

where $J_0(z)$ is the Bessel function (see Chapter 3) and $c(q, a_j)$ is as before. Here and on later occurrences, γ_χ denotes the imaginary part of a nontrivial zero of $L(s, \chi)$ on $\Re(s) = \frac{1}{2}$.

Proof. Let us write the formula (57) for $E^T(e^y; q, a_j)$ in a form where the terms are real, using the fact that $L(\frac{1}{2} + i\gamma, \chi) = 0$ if and only if $L(\frac{1}{2} - i\gamma, \bar{\chi}) = 0$:

$$E^T(e^y; q, a_j) = -c(q, a_j) - \sum_{\substack{0 < \gamma_\chi \leq T \\ \chi \neq \chi_0}} \left(\frac{\bar{\chi}(a_j)}{\frac{1}{2} + i\gamma_\chi} e^{i\gamma_\chi y} + \frac{\chi(a_j)}{\frac{1}{2} - i\gamma_\chi} e^{-i\gamma_\chi y} \right)$$

We will use the formula

$$\begin{aligned} & (\alpha + i\beta)(a + bi)e^{ix} + (\alpha - i\beta)(a - bi)e^{-ix} \\ &= \alpha((a + bi)e^{ix} + (a - bi)e^{-ix}) + i\beta((a + bi)e^{ix} - (a - bi)e^{-ix}) \\ &= 2\alpha\sqrt{a^2 + b^2} \sin(x + \varphi) + 2\beta\sqrt{a^2 + b^2} \cos(x + \varphi), \end{aligned}$$

where we applied the identity $A \sin x + B \cos x = \sqrt{A^2 + B^2} \sin(x + \varphi)$, where $\tan \varphi = \frac{b}{a}$, and $\arctan x + \arctan \frac{1}{x} = \frac{\pi}{2}$ and $\cos x = \sin(\frac{\pi}{2} - x)$.

Denoting $\bar{\chi}(a_j) = \alpha_j + i\beta_j$ and employing the previous formula with $\alpha + i\beta = \bar{\chi}(a_j)$ and $a + bi = (\frac{1}{2} + i\gamma_\chi)^{-1}$, we arrive at

$$E^T(e^y; q, a_j) = -c(q, a_j) - \sum_{\substack{0 < \gamma_\chi \leq T \\ \chi \neq \chi_0}} \frac{2}{\sqrt{\frac{1}{4} + \gamma_\chi^2}} (\alpha_j \cos(\gamma_\chi y + \varphi) + \beta_j \sin(\gamma_\chi y + \varphi))$$

for some φ depending on γ_χ . Let $\gamma_1, \gamma_2, \dots$ be the imaginary parts of the nontrivial zeros of the L -functions (mod q), where $\chi \neq \chi_0$, in increasing order and with $\gamma_k > 0$. Also let $X_0 = (-c(q; a_1), \dots, -c(q; a_r))$ and let X_j for $j = 1, 2, \dots$ be the vector

$$\frac{-2}{\sqrt{\frac{1}{4} + \gamma_j^2}} (\alpha_1 \cos(\gamma_j y + \varphi) + \beta_1 \sin(\gamma_j y + \varphi), \dots, \alpha_r \cos(\gamma_j y + \varphi) + \beta_r \sin(\gamma_j y + \varphi)).$$

Then the random variables X_0, X_1, X_2, \dots are independent with respect to P . Indeed, Lemma 3.8 tells that under the GSH, the functions $\sin(\gamma_j y)$ are asymptotically independent, so the functions $\sqrt{a_k^2 + b_k^2} \sin(\gamma_j y + \psi_j)$ are asymptotically independent as well for any a_k, b_k, ψ_j , since $\sin(\gamma_j y + \psi_j)$ is just a shifted version of $\sin(\gamma_j y)$. By the formula $A \cos x + B \sin x = \sqrt{A^2 + B^2} \sin(x + \psi)$, where ψ is chosen suitably, we see that $a_k \cos(\gamma_j y +$

$\varphi_j) + b_k \sin(\gamma_j y + \varphi_j)$ are asymptotically independent. These functions are the components of X_j , so the asymptotic independence is verified (X_0 is constant and hence independent of everything). We may thus exploit Lemma 4.6.

Let us compute the left-hand side of (66), where X_j are as above and n is chosen so that $\gamma_j \leq T$ if and only if $j \leq n$. If P_Y is the truncated version of P , that is

$$P_Y(A) := \frac{1}{Y} m(\{A \cap [0, Y]\}),$$

one has

$$\begin{aligned} \int_{\mathbb{R}} e^{-iE^T(y) \cdot \xi} dP(y) &= \lim_{Y \rightarrow \infty} \int_{\mathbb{R}} e^{-iE^T(y) \cdot \xi} dP_Y(y) \\ &= \lim_{Y \rightarrow \infty} \frac{1}{Y} \int_0^Y e^{-iE^T(y) \cdot \xi} dy \\ &= \int_{\mathbb{R}^r} e^{-ix \cdot \xi} d\nu^T(x) = \widehat{\nu^T}(\xi) \end{aligned}$$

by the definition (31) of the asymptotic density and the formula (59) applied to $x \mapsto e^{ix \cdot \xi}$ ((59) holds for complex functions as well since we may apply it to the real and imaginary parts).

Next, we compute the right-hand side of (66). The density function of X_0 with respect to P is the Dirac measure δ_v , $v = -(c(q, a_1), \dots, c(q, a_r))$, whose Fourier transform is $e^{-i\xi \cdot v}$. We also need the asymptotic density of $\cos \gamma y$ (for $\gamma > 0$). We have $P(\cos \gamma y > t) = (\frac{2\pi}{\gamma})^{-1} (\frac{2}{\gamma} \arccos t) = \frac{1}{\pi} \arccos t$ for $t \in [-1, 1]$, since the function $y \mapsto \cos \gamma y$ has period $\frac{2\pi}{\gamma}$ and for $y \in (-\frac{\pi}{\gamma}, \frac{\pi}{\gamma})$ the inequality $\cos \gamma y > t$ is equivalent to $y \in (-\frac{1}{\gamma} \arccos x, \frac{1}{\gamma} \arccos x)$. The density function is given by the derivative

$$\frac{d}{dt} P(\{y \in \mathbb{R} : \cos(\gamma y) < t\}) = \frac{d}{dt} \left(1 - \frac{1}{\pi} \arccos x \right) = \frac{1}{\pi \sqrt{1-t^2}}$$

when $t \in [-1, 1]$ and 0 otherwise. The density function of $\cos(\gamma y + \varphi)$ is identical to this, as the function is the same up to a shift. When $\cos(\gamma y + \varphi) = x$, we have $\sin(\gamma y + \varphi) = s \sqrt{1-x^2}$, where $s = 1$ with asymptotic probability

$\frac{1}{2}$ and $s = -1$ also with probability $\frac{1}{2}$, so by the definition of X_j

$$\begin{aligned} E_P(e^{iX_j \cdot \xi}) &= \int_{\mathbb{R}^r} e^{-iX_j(x) \cdot \xi} dP(x) \\ &= \frac{1}{2} \int_{-1}^1 \exp\left(\frac{2i}{\sqrt{\frac{1}{4} + \gamma_j^2}} \sum_{k=1}^r \xi_k (\alpha_k t + \beta_k \sqrt{1-t^2})\right) \frac{dt}{\pi \sqrt{1-t^2}} \\ &\quad + \frac{1}{2} \int_{-1}^1 \exp\left(\frac{2i}{\sqrt{\frac{1}{4} + \gamma_j^2}} \sum_{k=1}^r \xi_k (\alpha_k t - \beta_k \sqrt{1-t^2})\right) \frac{dt}{\pi \sqrt{1-t^2}}, \end{aligned}$$

for $j \geq 1$, because making the substitution $x = \cos(\gamma_k t + \varphi)$ means that $dP(x)$ must be replaced with $dP(\cos \gamma_k t < t)$, which was computed above. With the formula $e^{i(a+b)} + e^{i(a-b)} = 2e^{ia} \cos b$ and the substitution $t = \sin \theta$, that is $\frac{dt}{\sqrt{1-t^2}} = d\theta$, this becomes

$$\frac{1}{\pi} \int_{-\pi}^{\pi} \exp\left(\frac{2i}{\sqrt{\frac{1}{4} + \gamma_j^2}} \sum_{k=1}^r \xi_k \alpha_k \sin \theta\right) \cos\left(\frac{2}{\sqrt{\frac{1}{4} + \gamma_j^2}} \sum_{k=1}^r \xi_k \beta_k \cos \theta\right) d\theta.$$

In order to simplify this integral, we show that the formula

$$\frac{1}{\pi} \int_{-\pi}^{\pi} e^{iA \sin \theta} \cos(B \cos \theta) d\theta = J_0(\sqrt{A^2 + B^2})$$

holds. To see this, observe that

$$\begin{aligned} &\int_{-\pi}^{\pi} e^{iA \sin \theta + iB \cos \theta} d\theta + \int_{-\pi}^{\pi} e^{iA \sin \theta - iB \cos \theta} d\theta \\ &= \int_{-\pi}^{\pi} e^{i\sqrt{A^2 + B^2} \sin(\theta + \varphi)} d\theta + \int_{-\pi}^{\pi} e^{i\sqrt{A^2 + B^2} \sin(\theta + \varphi')} d\theta \\ &= 2\pi J_0(\sqrt{A^2 + B^2}) \end{aligned}$$

by the formula for combining sines and cosines and by the integral representation (36) of the Bessel function. We obtained

$$\int_{\mathbb{R}^r} e^{-iX_j(x) \cdot \xi} dP(x) = J_0\left(\frac{2|\sum_{j=1}^r \xi_j \bar{\chi}(a_j)|}{\sqrt{\frac{1}{4} + \gamma_j^2}}\right),$$

as $\sqrt{A^2 + B^2} = |A + Bi| = |2\sum_{k=1}^r \xi_k (\alpha_k + i\beta_k)|$ and $\alpha_k + i\beta_k = \bar{\chi}(a_k)$ in the case under consideration. Furthermore, we evaluated the Fourier transform

of X_0 , so

$$\int_{\mathbb{R}^r} e^{-iX_0(x)\cdot\xi} dP(x) = \exp\left(i\sum_{j=1}^r \xi_j c(q, a_j)\right),$$

and hence the right-hand side of (66) becomes

$$\exp\left(i\sum_{j=1}^r \xi_j c(q, a_j)\right) \prod_{\substack{0 < \gamma_\chi \leq T \\ \chi \neq \chi_0}} J_0\left(\frac{2|\sum_{j=1}^r \xi_j \bar{\chi}(a_j)|}{\sqrt{\frac{1}{4} + \gamma_\chi^2}}\right).$$

This converges to the desired Bessel product as $T \rightarrow \infty$, since the product is seen to be convergent from $J_0(z) = 1 - \frac{z^2}{2} + O(z^4)$ and the fact that $\prod_k (1 - a_k)$ converges (to a nonzero value) if $\sum_k a_k$ converges (this follows from $1 - x \leq e^{-x}$). Therefore it remains to show that $\widehat{\nu}^T(\xi)$, the right-hand side of (66), converges to $\hat{\mu}_{q;a_1,\dots,a_r}(\xi)$. We know that $\nu^T \rightarrow \mu_{q;a_1,\dots,a_r}$ weakly since $\mu(f) = \lim_{T \rightarrow \infty} \nu^T(f)$ Lemma 4.5 and Theorem 4.1 for bounded Lipschitz functions, and the portmanteau theorem gives the same convergence for bounded continuous functions. This implies $\widehat{\nu}^T \rightarrow \hat{\mu}_{q;a_1,\dots,a_r}$ weakly. However, now $\widehat{\nu}^T$ converges pointwise to the Bessel product by the above argument and weakly to $\hat{\mu}_{q;a_1,\dots,a_r}$, so the latter convergence is also pointwise and the limits coincide. This proves Theorem 4.7. \square

By repeating the proofs presented so far, mostly just changing the notations, we get for $q = 4, p^\alpha$ or $2p^\alpha$ that there is a measure $\mu_{q;R,N}$ for the race between the quadratic residues and nonresidues satisfying Theorem 4.1 and Theorem 4.7 in the form

$$\hat{\mu}_{q;R,N}(\xi) = e^{i\xi} \prod_{\gamma > 0} J_0\left(\frac{2\xi}{\sqrt{\frac{1}{4} + \gamma_\chi}}\right), \quad (68)$$

where γ runs through the nontrivial zeros of $L(s, \chi_1)$ and χ_1 is the real, non-principal character (mod q). The analogue of Theorem 4.1 is (80), and the argument of Theorem 4.7 works up to notation for the distribution of any trigonometric series with linearly independent frequencies γ_k . The constant coefficient of (80) is -1 , and this accounts for the factor $e^{i\xi}$, similarly as the constant $-c(q, a)$ in (44) accounts for the exponential in (67). In fact, the proof of Theorem 4.7 for $\hat{\mu}_{q;R,N}$ is even a bit easier than for $\hat{\mu}_{q;a_1,\dots,a_r}$ in the sense that we do not need probability density functions of vector valued random variables.

4.5 Comparison of Densities

We will soon see that the function $\hat{\mu}_{q;a_1,\dots,a_r}$ is symmetric with respect to the point $(-c(q, a_1), \dots, -c(q, a_r))$, and consequently the measure of the prime races is symmetric with respect to the same point, which will lead to bias towards nonsquares (mod q). Also some other properties of the measure are needed, and they are stated in the following theorem (the claims of the theorem are mentioned in [28] but not explicitly proved there).

Theorem 4.8. *Assume first $r < \varphi(q)$. We have*

- (i) $|\hat{\mu}_{q;a_1,\dots,a_r}(\xi)| \ll e^{-c_q|\xi|^2}$ for some constant $c_q > 0$ when $\xi \in \mathbb{R}^r$.
- (ii) $\mu_{q;a_1,\dots,a_r}$ is absolutely continuous with respect to the r -dimensional Lebesgue measure, and its Radon-Nikodym derivative $\mu_{q;a_1,\dots,a_r}(x)$ is an entire function in \mathbb{C}^r (that is, each component function is entire in its complex variable).
- (iii) $\mu_{q;a_1,\dots,a_r}$ does not vanish on any subset of \mathbb{R}^r of positive r -dimensional Lebesgue measure (here \mathbb{R}^r is interpreted as a subset of \mathbb{C}^r consisting of those vectors that have real components).

Assume now $r = \varphi(q)$. Then

- (i) $|\hat{\mu}_{q;a_1,\dots,a_r}(\xi)| \ll e^{-c_q|\xi|^2}$ for some constant $c_q > 0$ when $\xi_1 + \dots + \xi_r = 0$.
- (ii) $\mu_{q;a_1,\dots,a_r}$ is absolutely continuous with respect to the Lebesgue measure on $T = \{\xi \in \mathbb{R}^r : \xi_1 + \dots + \xi_r = 0\}$, and its Radon-Nikodym derivative is analytic in $T' = \{\xi \in \mathbb{C}^r : \xi_1 + \dots + \xi_r = 0\}$.
- (iii) $\mu_{q;a_1,\dots,a_r}$ does not vanish on any subset of T of positive $r-1$ -dimensional Lebesgue measure.

In addition, in both cases we have

- (iv) $\mu_{q;a_1,\dots,a_r}(x)$ is symmetric with respect to the point $(-c(q, a_1), \dots, -c(q, a_r))$.

24

Proof. (i) Notice that in the formula (67) for $\hat{\mu}_{q;a_1,\dots,a_r}(\xi)$, the first factor has absolute value 1, and by an elementary computation, the Bessel function satisfies $J_0(x) \leq 1 - \frac{x^2}{10}$ for $x \in [-2, 2]$ and $|J_0(x)| \leq 1$ for all x . Therefore,

$$\begin{aligned} |\hat{\mu}_{q;a_1,\dots,a_r}(\xi)| &\ll \prod_{\substack{x \pmod{q} \\ x \neq x_0}} \prod_{\gamma_\chi > 0} \left(1 - \frac{|\sum_{j=1}^r \chi(a_j) \xi_j|^2}{5(\frac{1}{4} + \gamma_\chi^2)} \right) \\ &\leq \exp \left(- \sum_{\substack{x \pmod{q} \\ x \neq x_0}} \sum_{\gamma_\chi > 0} \frac{|\sum_{j=1}^r \chi(a_j) \xi_j|^2}{5(\frac{1}{4} + \gamma_\chi^2)} \right), \end{aligned}$$

²⁴Note that symmetry with respect to a point is different from symmetry with respect to permutations of variables, which is what one part of the Main Theorem is about.

because of the inequality $1 - x \leq e^{-x}$. We choose a constant $M_q > 0$ such that $\sum_{\gamma_\chi > 0} \frac{1}{5(\frac{1}{4} + \gamma_\chi^2)} \geq M_q$ for all characters $\chi \pmod{q}$. Then our expression is at most

$$\exp \left(- \sum_{\substack{\chi \pmod{q} \\ \chi \neq \chi_0}} M_q \left| \sum_{j=1}^r \chi(a_j) \xi_j \right|^2 \right). \quad (69)$$

This seems to decay like a Gaussian function, but actually it is not even trivial that it is integrable. Indeed, functions such as $\exp(-(\xi_1 + \dots + \xi_r)^2)$ are not integrable since this function is greater than e^{-1} in the set $\{(\xi_1, \dots, \xi_r) : |\xi_1 + \dots + \xi_r| < 1\}$ which has infinite r -dimensional Lebesgue measure.

We will first show that the system of linear equations

$$\sum_{j=1}^r \chi(a_j) \xi_j = 0, \quad \chi \in \{\chi_1, \dots, \chi_{\varphi(q)-1}\}, \quad (70)$$

where χ_i are the non-principal characters \pmod{q} , does not have any nonzero solution, and then we employ this fact to show that for some constant $c_q > 0$ there always exists some $\chi \neq \chi_0$ such that

$$\left| \sum_{j=1}^r \chi(a_j) \xi_j \right| \geq c_q |\xi|. \quad (71)$$

To prove the first claim, let $m_1, \dots, m_{\varphi(q)-1}$ be a set of non-congruent integers \pmod{q} that are coprime to q and contain a_1, \dots, a_r . Then a nontrivial solution to (70) implies a nontrivial solution to the more general system

$$\begin{aligned} \chi_1(m_1)x_1 + \dots + \chi_1(m_{\varphi(q)-1})x_{\varphi(q)-1} &= 0 \\ \chi_2(m_1)x_1 + \dots + \chi_2(m_{\varphi(q)-1})x_{\varphi(q)-1} &= 0 \\ &\dots \\ \chi_{\varphi(q)-1}(m_1)x_1 + \dots + \chi_{\varphi(q)-1}(m_{\varphi(q)-1})x_{\varphi(q)-1} &= 0, \end{aligned} \quad (72)$$

where $x_1, \dots, x_{\varphi(q)-1} \in \mathbb{C}$ (we just added some columns to the system). In other words, the existence of a nonzero solution tells that the matrix $A = (\chi_i(m_j))_{i,j=1}^{\varphi(q)-1}$ is not invertible. Let m_0 be a number coprime to q that is not congruent to $m_1, \dots, m_{\varphi(q)-1}$, and denote by χ_0 the principal character. Let B be the matrix $B = (\chi_i(m_j))_{i,j=0}^{\varphi(q)-1}$, which is obtained from A by adding one row and one column. We first show that B is invertible.

Observe that the vectors $(\chi_0(m_i), \chi_1(m_i), \dots, \chi_{\varphi(q)-1}(m_i))$ are pairwise orthogonal since their inner products are of the form

$$\sum_{\chi \pmod{q}} \chi(m_i) \bar{\chi}(m_j) = 0$$

by the orthogonality of characters, so the columns of B are linearly independent (because pairwise orthogonal vectors are linearly independent). Hence the rank of B is $\varphi(q)$, so B is invertible. Moreover, the first column of B consists of ones only, so by Cramer's rule $\det B = \det A$, meaning that A is invertible since B is. However, then the system (70) has no nontrivial solution.

Our next task is to show that not all the sums in (70) are small. Suppose that all the sums under consideration are smaller than $\frac{|\xi|}{C_q \sqrt{\varphi(q)-1}}$. Then the linear system

$$\sum_{j=1}^{\varphi(q)-1} \chi_i(a_j) \xi_j = b_i \quad i = 1, \dots, r$$

has a solution for some b_i such that $|b_i| < \frac{|\xi|}{C_q \sqrt{\varphi(q)-1}}$, and at least one b_i is non-zero. Choose $C_q = \|A_1^{-1}\|$ (the operator norm of A_1^{-1}), where $A_1 = (\chi_i(a_j))_{i,j=1}^r$ is now known to be invertible. the other hand, if $b = (b_1, \dots, b_r)$ and $\xi = (\xi_1, \dots, \xi_r)$, the solution (to the system (72)) is $\xi = A_1^{-1}b$, but then $|\xi| = |A_1^{-1}b| \leq \|A_1^{-1}\| |b| < C_q \sqrt{b_1^2 + \dots + b_r^2} \leq |\xi|$, a contradiction. Therefore (71) follows, and thus we have proved (i).

(ii) We first show the existence of the Radon-Nikodym derivative. By (i), we know that $\hat{\mu}_{q;a_1, \dots, a_r} \in L^2$. The Fourier transform is a bijection in L^2 , so $\widehat{\hat{\mu}}_{q;a_1, \dots, a_r}(-x) \in L^2$. However, as a distribution, this is the measure $\mu_{q;a_1, \dots, a_r}$ (in the previous chapter we have given measures an interpretation as distributions), so this is the Radon-Nikodym derivative (see previous chapter), also denoted by $\mu_{q;a_1, \dots, a_r}$. It remains to show that it is entire, or equivalently that the Fourier transform of $\hat{\mu}_{q;a_1, \dots, a_r}(\xi)$ is entire in \mathbb{C}^r . By (i), $\hat{\mu}_{q;a_1, \dots, a_r}(\xi)$ decays like $e^{-c_q |\xi|^2}$ for $\xi \in \mathbb{R}^r$, so the truncated integral

$$\int_{B(0,r)} \hat{\mu}_{q;a_1, \dots, a_r}(\xi) e^{ix \cdot \xi} d\xi$$

can be differentiated repeatedly under the integral sign with respect to any ξ_i (these are the components of ξ) and it converges uniformly to the Fourier

transform of $\hat{\mu}_{q;a_1,\dots,a_r}$ as $R \rightarrow \infty$. It is well-known that a uniform limit of analytic functions is analytic (see [33], page 53-54), so $\hat{\mu}_{q;a_1,\dots,a_r}(\xi)$ is entire in each of its variables, completing the proof.

(iii) This feature is common to all entire functions in \mathbb{C}^r . Assume that $\mu_{q;a_1,\dots,a_r}$ vanishes in $A \subset \mathbb{R}^r$ where $m_r(A) > 0$. Then, by Fubini's theorem, for each coordinate direction e_i in \mathbb{R}^r there exists a line ℓ_i parallel to e_i and intersecting A in a set of positive one dimensional measure (for otherwise integrating over all the segments in one direction would produce zero, while the integral of $\mu_{q;a_1,\dots,a_r}$ is 1 as it is a probability measure). However, the restriction of $\mu_{q;a_1,\dots,a_r}$ onto ℓ_i is entire as a function of one complex variable, so it cannot vanish on an interval unless it is constant. We conclude that $\mu_{q;a_1,\dots,a_r}$ is constant in each of its variables. This is a contradiction since $\mu_{q;a_1,\dots,a_r}$ is a probability density function, so its integral is 1. Hence the counter assumption was false.

(iv) To see that $\mu_{q;a_1,\dots,a_r}$ is symmetric with respect to the point $A := (-c(q, a_1), \dots, -c(q, a_r))$ (regardless whether $r < \varphi(q)$), notice that

$$\hat{\mu}_{q;a_1,\dots,a_r}(-\xi) = \exp\left(-2i \sum_{j=1}^r c(q, a_j)\xi_j\right) \hat{\mu}_{q;a_1,\dots,a_r}(\xi).$$

By taking the inverse Fourier transform on both sides, we obtain

$$\mu_{q;a_1,\dots,a_r}(\xi) = \mu_{q;a_1,\dots,a_r}(2A - \xi),$$

so $\mu_{q;a_1,\dots,a_r}(\xi)$ is symmetric with respect to A .

We are left with the case $r = \varphi(q)$.

(i) Again (69) is an upper bound for the Fourier transform $\hat{\mu}_{q;a_1,\dots,a_r}(\xi)$, but it no longer decays in every direction as it gets the same value at ξ and $\xi + \lambda(e_1 + \dots + e_r)$ for any λ , where (e_i) is the standard basis of \mathbb{R}^r . This is due to the fact that $\sum_{i=1}^{\varphi(q)} \chi(a_i) = 0$ for $\chi \neq \chi_0$. Nevertheless, we know that the system

$$\sum_{j=1}^{\varphi(q)} \chi(a_j)\xi_j, \quad \chi \in \{\chi_0, \dots, \chi_{\varphi(q)-1}\} \quad (73)$$

has no nontrivial solution since the corresponding matrix is the invertible

matrix B mentioned above. This leads to

$$\left| \sum_{j=1}^{\varphi(q)} \chi(a_j) \xi_j \right| \geq c_q |\xi|$$

for some $c_q > 0$ and for $\chi \neq \chi_0$ whenever $\xi_1 + \dots + \xi_{\varphi(q)} = 0$. Indeed, one has

$$\begin{aligned} |\xi| &= |B^{-1}B\xi| \leq \|B^{-1}\| |B\xi| \\ &= \|B^{-1}\| \sqrt{\sum_{i=1}^{\varphi(q)} \left| \sum_{j=1}^{\varphi(q)} \chi_i(a_j) \xi_j \right|^2} \\ &\leq C_q \max_{i \leq \varphi(q)} \left| \sum_{j=1}^{\varphi(q)} \chi_i(a_j) \xi_j \right|, \end{aligned}$$

and since $\xi_1 + \dots + \xi_{\varphi(q)} = 0$, we have $i \neq \varphi(q)$ in the maximum, so the condition (71) is fulfilled for at least one $\chi \neq \chi_0$.

(ii) For brevity, let us write $\mu = \mu_{q;a_1, \dots, a_r}$. We concluded that $|\hat{\mu}(\xi)| \ll e^{-c_q|\xi|^2}$ when $r = \varphi(q)$ and $\xi \in T := \{x_1 + \dots + x_r = 0\}$. Let $\Lambda : \mathbb{R}^{r-1} \rightarrow T$ be the linear transformation $\Lambda(x_1, \dots, x_{r-1}) = (x_1, \dots, x_{r-1}, -x_1 - \dots - x_{r-1})$ whose inverse projects the hyperplane T onto \mathbb{R}^{r-1} . Then $\hat{\mu}(\Lambda(x)) \in L^2(\mathbb{R}^{r-1})$. On the other hand, if $\mu_1(A) := \mu(\Lambda(A))$ for all measurable A , then $c\hat{\mu} \circ \Lambda = \hat{\mu}_1 \in L^2(\mathbb{R}^{r-1})$ for some c (that arises from the change of variables $y = \Lambda(x)$). By the same argument as in the case $r < \varphi(q)$, we get $\hat{\mu}_1(-x) \in L^2(\mathbb{R}^{r-1})$ and that as a distribution this equals μ_1 . Therefore μ_1 is absolutely continuous with respect to m_{r-1} with Radon-Nikodym derivative belonging to $L^2(\mathbb{R}^{r-1})$, so its variant μ is absolutely continuous with respect to the Lebesgue measure m_T on the plane, and $\mu(x) \in L^2(T)$. Since $\hat{\mu}(x)$ decays like a Gaussian function in T , the same holds for $\hat{\mu}_1$ in \mathbb{R}^{r-1} , so for the same reason as before, $\mu_1(x)$ is analytic in \mathbb{C}^{r-1} , and hence $\mu(x)$ is analytic in T' .

(iii) As $\mu_{q;a_1, \dots, a_r}(x)$ is analytic in T' , a very similar proof as for $r < \varphi(q)$ shows that it cannot vanish on any set $E \subset T$ with $m_T(E) > 0$, or equivalently $m_{r-1}(\Lambda^{-1}(E)) > 0$. Thus the theorem is proved. \square

We are now in a position to prove the claims of the Main Theorem about the densities. Defining

$$\mu_Y(A) := \frac{1}{Y} m(\{y \in [0, Y] : E(y) \in A\}),$$

we can finally say that $\mu_{q;a_1,\dots,a_r}$ gives the value distribution of $E(y)$ in the sense that

$$\mu_{q;a_1,\dots,a_r}(A) = \lim_{Y \rightarrow \infty} \mu_Y(A) \quad (74)$$

for all Borel sets A with $m_r(\partial A) = 0$ if $r < \varphi(q)$ and for all Borel sets A with $m_T(\partial A) = 0$ if $r = \varphi(q)$. Indeed, Theorem 4.1 can be written in the form

$$\int_{\mathbb{R}^r} f(x) d\mu_Y(x) \xrightarrow{Y \rightarrow \infty} \int_{\mathbb{R}^r} f(x) d\mu_{q;a_1,\dots,a_r}(x) \quad (75)$$

for all bounded continuous functions $f : \mathbb{R}^r \rightarrow \mathbb{R}$, which means $\mu_Y \rightarrow \mu_{q;a_1,\dots,a_r}$ weakly. If $r < \varphi(q)$, the portmanteau theorem then gives (75) whenever $\mu_{q;a_1,\dots,a_r}(\partial A) = 0$, and the absolute continuity proved above gives (75) when $m_r(\partial A) = 0$. If $r = \varphi(q)$, the formula (74) holds for all Borel sets $A \subset T$ with $\mu_{q;a_1,\dots,a_r}(\partial A) = 0$ and by the previous theorem this is satisfied when $m_T(\partial A) = 0$.

In particular, if $r < \varphi(q)$, we may choose $A = \{x_1 > \dots > x_r\}$. We get

$$\begin{aligned} \delta(P_{q;a_1,\dots,a_r}) &= \lim_{X \rightarrow \infty} \frac{1}{\log X} \int_{P_{q;a_1,\dots,a_r} \cap [\log 2, X]} \frac{dx}{x} \\ &= \lim_{Y \rightarrow \infty} \frac{1}{Y} \int_2^Y 1_{\{x_1 > \dots > x_r\}}(E(y)) dy \\ &= \lim_{Y \rightarrow \infty} \frac{1}{Y} m(\{y \in [2, Y] : E(y) \in \{x_1 > \dots > x_r\}\}) \\ &= \mu_{q;a_1,\dots,a_r}(\{x_1 > \dots > x_r\}) \\ &= \int_{\{x_1 > \dots > x_r\}} \mu_{q;a_1,\dots,a_r}(x) dx, \end{aligned} \quad (76)$$

and if this was equal to zero, $\mu_{q;a_1,\dots,a_r}$ would vanish on a subset of \mathbb{R}^r of infinite measure, which is impossible by part (iii) of Theorem 4.8. The same computation up to notation together with Theorem 4.8 works if $r = \varphi(q)$ (we choose $A = T \cap \{x_1 > \dots > x_r\}$). Therefore $\delta(P_{q;a_1,\dots,a_r}) > 0$.

To show the claim that the set

$$Q_{a_1,\dots,a_r} := \left\{ x \geq 2 : \pi(x; q, a_{i+1}) - \pi(x; q, a_i) > \frac{\sqrt{x}}{\eta(x) \log x}, i = 1, \dots, r-1 \right\}$$

has logarithmic density equal to $\delta(P_{q;a_1,\dots,a_r})$, one considers the set

$$A_\varepsilon := \{(x_1, \dots, x_r) \in \mathbb{R}^r : x_{i+1} - x_i > \varepsilon, i = 1, \dots, r-1\},$$

whose boundary has r -dimensional measure zero, so $\mu_{q;a_1,\dots,a_r}(A_\varepsilon) = \lim_{Y \rightarrow \infty} \mu_Y(A)$ holds. Choose $\varepsilon > 0$, and let M be such that $\frac{1}{\eta(x)} < \varepsilon$ for $x \geq M$. The logarithmic density assigns value zero to any bounded set, so the set $Q_{q;a_1,\dots,a_r} \cap [M, \infty)$ has the same logarithmic density as the original set. Mimicking the computation (76), we have

$$\begin{aligned} \delta(Q_{q;a_1,\dots,a_r}) &= \delta(Q_{q;a_1,\dots,a_r} \cap [M, \infty)) \\ &\geq \lim_{Y \rightarrow \infty} \frac{1}{Y} \int_M^Y 1_{A_\varepsilon}(E(y)) dy \\ &= \mu_{q;a_1,\dots,a_r}(A_\varepsilon). \end{aligned}$$

since $1_{A_\varepsilon}(E(y)) = 1$ implies $y \in Q_{q;a_1,\dots,a_r}$ for $Y \geq M$. However, the last expression converges to $\mu_{q;a_1,\dots,a_r}(A_0)$ as $\varepsilon \rightarrow 0$ by the continuity of measures. We already know that $\mu_{q;a_1,\dots,a_r}(A_0) = \delta(P_{q;a_1,\dots,a_r})$, so $\delta(Q_{q;a_1,\dots,a_r}) \geq \delta(P_{q;a_1,\dots,a_r})$. Since the converse inequality is trivial, the claim follows.

Next, let a be a quadratic nonresidue and b a quadratic residue (mod q). From part (iv) of Theorem 4.8, we know that $\mu_{a,b}$ is symmetric with respect to the point $(-c(q, a), -c(q, b))$. Now if we draw a line through $(-c(q, a), -c(q, b))$ that is parallel to the line $x_2 = x_1$, the plane \mathbb{R}^2 is divided into two half-planes H_1 and H_2 (H_1 is the upper one of these) such that

$$\int_{H_1} \mu_{a,b}(x) dx = \int_{H_2} \mu_{a,b}(x) dx = \frac{1}{2},$$

by symmetry. Since $-c(q, a) = 1$ and $-c(q, b) < 0$, the center of symmetry $(-c(q, a), -c(q, b))$ lies in the set $\{(x_1, x_2) : x_1 > x_2\}$. Therefore

$$\int_{\{x_1 > x_2\}} \mu_{a,b}(x) dx = \int_{H_1} \mu_{a,b}(x) dx + \int_{\{x_1 > x_2\} \setminus H_1} \mu_{a,b}(x) dx \geq \frac{1}{2}.$$

If equality held, the integral of $\mu_{a,b}$ would be zero on a strip of infinite two-dimensional Lebesgue measure, which contradicts the part (iii) of Theorem 4.8 and proves $\delta(P_{q;a,b}) > \frac{1}{2}$. One also sees that $\delta(P_{q;N,R}) > \frac{1}{2}$ since the Bessel product formula for $\hat{\mu}_{q;R,N}$ is symmetric with respect to $\xi = -1$ and $\mu_{q;R,N}$ is an entire function by a similar but easier argument than above (the claims of Theorem 4.8 can be verified explicitly in this case).

Assume then that a and b are both quadratic residues or nonresidues (mod q). As we remarked on page 33, $c(q, a) = c(q, b)$ in this case. Therefore, by part (iv) of Theorem 4.8, $\hat{\mu}_{a,b}$ is symmetric with respect to the point (r, r) , where

$r = -c(q, a)$. Therefore we can compute

$$\begin{aligned}
1 &= \int_{\{x_1 > x_2\}} \mu_{q;a,b}(x_1, x_2) dx_1 dx_2 + \int_{\{x_1 < x_2\}} \mu_{q;a,b}(x_1, x_2) dx_1 dx_2 \\
&= \int_{\{x_1 > x_2\}} \mu_{q;a,b}(x_1, x_2) dx_1 dx_2 + \int_{\{r-x_1 < r-x_2\}} \mu_{q;a,b}(r-x_1, r-x_2) dx_1 dx_2 \\
&= 2 \int_{\{x_1 > x_2\}} \mu_{q;a,b}(x_1, x_2) dx_1 dx_2 = 2\delta(P_{q;a,b}),
\end{aligned}$$

where we used the symmetry with respect to (r, r) in the penultimate step. Thus $\delta(P_{q;a,b}) = \frac{1}{2}$. \square

4.6 Vanishing of the Bias as $q \rightarrow \infty$

We shall now prove the part of the Main Theorem, which says that the logarithmic density of the set where the quadratic nonresidue primes $(\bmod q)$ have the lead approaches $\frac{1}{2}$ as $q \rightarrow \infty$ (q is still of the form $4, p^\alpha$ or $2p^\alpha$).

Fix some constant M , and restrict ξ to be at most M in modulus. Taking the logarithm of the formula (68) for the Fourier transform of $\mu_{q;R,N}$, we get for large values of q

$$\begin{aligned}
\log \hat{\mu}_{q;R,N}\left(\frac{\xi}{\sqrt{\log q}}\right) &= \frac{i\xi}{\sqrt{\log q}} + \sum_{\gamma > 0} \log J_0\left(\frac{2\xi}{\sqrt{\frac{1}{4} + \gamma^2} \sqrt{\log q}}\right) \\
&= \frac{i\xi}{\sqrt{\log q}} + \sum_{\gamma > 0} \log\left(1 - \frac{\xi^2}{(\log q)\left(\frac{1}{4} + \gamma^2\right)} + D(\xi)\right)
\end{aligned}$$

where $|D(\xi)| \ll \frac{\xi^4}{\log^2 q} \ll \frac{M^4}{\log^2 q}$ by the power series of J_0 . Then by the Taylor series of the logarithm,

$$\log \hat{\mu}_{q;R,N}\left(\frac{\xi}{\sqrt{\log q}}\right) = \frac{i\xi}{\sqrt{\log q}} - \frac{\xi^2}{\log q} \sum_{\gamma > 0} \frac{1}{\frac{1}{4} + \gamma^2} + O\left(\frac{M^4}{\log^2 q} \sum_{\gamma > 0} \frac{1}{\left(\frac{1}{4} + \gamma^2\right)^2}\right).$$

The sum $\sum_{\gamma > 0} \frac{1}{\frac{1}{4} + \gamma^2}$ is related to the logarithmic derivative of $L(s, \chi_1)$ by

$$\sum_{\gamma > 0} \frac{1}{\frac{1}{4} + \gamma^2} = \frac{1}{2} \log q + \frac{L'}{L}(1, \chi_1) + O(1)$$

where C stands for Euler's constant. Indeed, this is formula (16) from Chapter 2 for the real character χ_1 when we notice that

$$\sum_{\rho} \left(\frac{1}{1-\rho} + \frac{1}{\rho} \right) = \sum_{\gamma>0} \frac{1}{\frac{1}{4} + \gamma^2}$$

and $B(\chi) = -\sum_{\gamma>0} \frac{1}{\frac{1}{4} + \gamma^2}$ (for this, see Davenport's book [5], pages 82-83).

In addition, one has $\frac{L'}{L}(1, \chi_1) \ll \log \log q$ (the following proof is from [10]). To see this, we take the logarithmic derivative of the Euler product of $L(s, \chi_1)$, and then set $s = 1$, to get

$$\frac{L'}{L}(1, \chi_1) = -\sum_p \frac{\chi_1(p) \log p}{p - \chi_1(p)} = -\sum_p \frac{\chi_1(p) \log p}{p} + O(1),$$

where we just applied $\frac{1}{x} - \frac{1}{x+h} = \frac{h}{x(x+h)}$. By partial summation, this gives

$$\frac{L'}{L}(1, \chi_1) = \int_2^{\infty} \frac{\theta(x, \chi_1)}{x^2} dx + O(1) = \int_y^{\infty} \frac{\theta(x, \chi_1)}{x^2} dx + O(\log y) + O(1), \quad (77)$$

where $\theta(x, \chi_1) = \sum_{p \leq x} \chi_1(p) \log p$. We choose $y = \log^4 q$ and use the estimate $\theta(x, \chi) \ll \sqrt{x} \log^2(qx)$, which is implied by the GRH (in Chapter 3 this was proved without indicating q -dependence for $\psi(x, \chi)$, which differs from $\theta(x, \chi)$ by at most $2\sqrt{x} \log x$. A proof of the result with q -dependence can be found in [24] on page 370). We obtain the result $\frac{L'}{L}(1, \chi_1) \ll \log \log q$ from (77) since

$$\int_y^{\infty} \frac{(\log q + \log x)^2}{x^{\frac{3}{2}}} dx = O\left(\frac{\log^2 q}{\sqrt{y}} + \frac{(\log q) \log \log q}{\sqrt{y}} + 1\right).$$

Because of

$$\sum_{\gamma>0} \frac{1}{(\frac{1}{4} + \gamma^2)^2} \leq 4 \sum_{\gamma>0} \frac{1}{(\frac{1}{4} + \gamma^2)},$$

we now get, for $|\xi| \leq M$,

$$\begin{aligned} \log \hat{\mu}_{q,R,N} \left(\frac{\xi}{\sqrt{\log q}} \right) &= \frac{i\xi}{\sqrt{\log q}} - \frac{\xi^2}{2} \left(1 + O\left(\frac{\log \log q}{\log q}\right) + O\left(\frac{M^4}{\log q}\right) \right) \\ &= -\frac{\xi^2}{2} + O\left(\frac{M}{\sqrt{\log q}} + \frac{M^2 \log \log q}{\log q} + \frac{M^6}{\log q}\right), \end{aligned}$$

which approaches $-\frac{\xi^2}{2}$ as $q \rightarrow \infty$. Therefore $\hat{\mu}_{q;R,N}(\frac{\xi}{\sqrt{\log q}}) \rightarrow e^{-\frac{\xi^2}{2}}$ as $q \rightarrow \infty$ pointwise, so by Levy's theorem from Chapter 3, $\sqrt{\log q} \mu_{q;R,N}(\frac{x}{\sqrt{\log q}})$ converges in distribution to the probability density function of the normal distribution, which is $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$. In particular,

$$\int_0^\infty \mu_{q;R,N}(x) dx = \int_0^\infty \sqrt{\log q} \mu_{q;R,N}(\frac{x}{\sqrt{\log q}}) dx \rightarrow \frac{1}{\sqrt{2\pi}} \int_0^\infty e^{-\frac{x^2}{2}} dx = \frac{1}{2}$$

as $q \rightarrow \infty$. The integral above is $\delta(P_{q;R;N})$, so we have proved that it approaches $\frac{1}{2}$. Obviously $\delta(P_{q;N;R}) = \delta([1, \infty)) - \delta(P_{q;R;N})$, so this also approaches $\frac{1}{2}$, proving the statement. \square

4.7 Symmetries of the Prime Races

We will demonstrate the remaining part of the Main Theorem, namely that the condition given for $r = 3$ in the Main Theorem implies that the prime race is even; more precisely,

Theorem 4.9. *The function $\mu_{q;a_1,\dots,a_r}(x_1, \dots, x_r)$ is invariant under permutations of (x_1, \dots, x_r) if one of the following holds:*

- (i) $r = 2$ and $a_1 a_2$ is a quadratic residue
- (ii) $r = 3$ and there is an element $\omega \in \mathbb{Z}_q^\times$ such that

$$\omega^3 \equiv 1 \pmod{q}, \quad a_2 \equiv a_1 \omega \pmod{q}, \quad a_3 \equiv a_2 \omega \pmod{q}.$$

Notice that $\delta(P_{q;a_1,\dots,a_r})$ is symmetric in the a_i 's if $\mu_{q;a_1,\dots,a_r}$ is. Indeed,

$$\delta(P_{q;a_1,\dots,a_r}) = \int_A \mu_{q;a_1,\dots,a_r}(x) dx,$$

where $A = \{(x_1, \dots, x_r) \in \mathbb{R}^r : x_1 > x_2 > \dots > x_r\}$, so the symmetry of $\mu_{q;a_1,\dots,a_r}$ implies that the order of a_i 's does not matter for the density. It is by no means clear that if $\mu_{q;a_1,\dots,a_r}$ is not symmetric in the a_i 's, then all of the densities $\delta(P_{q;a_1,\dots,a_r})$ cannot be equal. This is believed to be true, though, but it is an open problem; see [20].

Proof. We want to show that the Bessel product formula for $\mu_{q;a_1,\dots,a_r}$ is symmetric in its variables in the cases mentioned above. One sees by multiplying out that $|\chi(a_1)\xi_1 + \chi(a_2)\xi_2|^2 = |\chi(a_2)\xi_1 + \chi(a_1)\xi_2|^2$, so the Bessel product factor is symmetric in the case (i). In the case (ii), if σ is a permutation of $\{1, 2, 3\}$, then by comparing the coefficients, one sees that

$|\chi(a_1)\xi_1 + \chi(a_2)\xi_2 + \chi(a_3)\xi_3|^2 = |\chi(a_{\sigma(1)})\xi_1 + \chi(a_{\sigma(2)})\xi_2 + \chi(a_{\sigma(3)})\xi_3|^2$ for all $\xi_1, \xi_2, \xi_3 \in \mathbb{R}$ if and only if

$$\Re(\chi(a_i)\bar{\chi}(a_j)) = \Re(\chi(a_{\sigma(i)})\bar{\chi}(a_{\sigma(j)}))$$

for all $1 \leq i < j \leq 3$. If condition (ii) is satisfied, we need to prove

$$\Re(w^{i-j}) = \Re(w^{\sigma(i)-\sigma(j)})$$

for all $i < j$, where $w = \chi(\omega)$ is a cubic root of unity. If σ is an identical permutation or (123) or (231), this certainly holds. If σ is a transposition, say (12), then the claim is true because of $\Re(w) = \Re(w^{-1})$. Hence the claim holds for all permutations.

If condition (i) of the theorem holds, then also the exponential factor in the Bessel product of (46) is symmetric in ξ_1, ξ_2 , so $\hat{\mu}_{q;a_1,a_2}(\xi_1, \xi_2)$ is also symmetric, and then $\mu_{q;a_1,a_2}$ is symmetric as well. If condition (ii) holds, then $c(q, a_3) = c(q, a_2) = c(q, a_1)$ since $c(q, a)$ depends only on whether a is a quadratic residue (mod q) and $a_3 \equiv \omega^2 a_1, a_2 \equiv \omega^{-2} a_3 \pmod{q}$. This means that the exponential factor of the formula for $\hat{\mu}_{q;a_1,a_2,a_3}$ is symmetric, and we already observed that the Bessel product is symmetric. Since $\hat{\mu}_{q;a_1,a_2,a_3}(\xi)$ is now symmetric in its variables, so is $\mu_{q;a_1,a_2,a_3}(\xi)$. This finishes the proof of the last assertion of the Main Theorem. \square

4.8 Asymptotics of the Measures of the Prime Races

We consider the rate of decay of $\mu_{q;a_1,\dots,a_r}$ at infinity, or more precisely, the decay of $\mu_{q;a_1,\dots,a_r}(\{|x| \geq R\})$ as a function of R . It turns out that this function goes to zero at least subexponentially, but at most double-exponentially. Based on what is known for special prime races (see [23] for the race between $\pi(x)$ and $\text{Li}(x)$, which is not discussed in this thesis but behaves similarly), it is suspected that an estimate of the form $\exp(-\exp(-c_q\sqrt{R}))$ is close to reality [28]. This explains why the behavior of the prime races (for example the race between the primes $3n-1$ and $3n+1$ mentioned in the introduction) is seen only in extremely large scales. For simplicity, the lower bound is proved only for the race between the quadratic nonresidues and residues.

Theorem 4.10. *Assuming the GRH and the GSH, one has*

$$\mu_{q;a_1,\dots,a_r}(\{|x| \geq R\}) \ll \exp(-c_q\sqrt{R}) \quad (78)$$

for some $c_q > 0$. For q equal to 4, an odd prime power or twice an odd prime power,

$$\mu_{q;R,N}(\{x \geq R\}) \gg \exp(-\exp(c_q R)). \quad (79)$$

The same holds for $\mu_{q;N,R}$.

The proof of the upper bound follows by combining some earlier lemmas and theorems. Indeed, we have by (64) the inequality

$$\mu(f) = \nu^T(f) + O\left(c_f \frac{\log^2 T}{T}\right),$$

where c_f is a Lipschitz constant for f , ν^T is as in Lemma 4.4, and

$$\mu(f) := \lim_{Y \rightarrow \infty} \int_2^Y \frac{1}{Y} f(E(y)) dy.$$

We choose $f(x) = 1 - \min\{1, d(x, A)\}$, where $A = B(0, R)^c$ is the complement of the ball $B(0, R)$ and $d(x, A)$ is the distance of x from A . Then f is Lipschitz with constant 1, so

$$|\mu(f) - \nu^T(f)| \leq c \frac{\log^2 T}{T}$$

for some absolute constant c . By Lemma 4.4, ν^T is supported in the ball $B(0, c_q \log^2 T)$ for some c_q , so if $R = 1 + c_q \log^2 T$, the supports of f and ν^T are disjoint. Then

$$\mu(f) \leq c \frac{\log^2 T}{T}.$$

Since $f(x) \geq 1_{B(0,R)^c}(x)$, the definition of $\mu(1_{B(0,R)^c})$ gives

$$\lim_{Y \rightarrow \infty} \frac{1}{Y} m(\{y \in (0, Y] : E(y) \in B(0, R)^c\}) \leq c \frac{\log^2 T}{T}.$$

The limit on the left is equal to $\mu_{q;a_1,\dots,a_r}(B(0, R)^c)$ by (74), so our choice of R gives

$$\mu_{q;a_1,\dots,a_r}(B(0, R)^c) \ll e^{-c'_q \sqrt{R}},$$

which was to be shown. □

The lower bound is significantly more tedious to prove. Let χ_1 be the non-trivial real character (mod q), and let

$$R(x) := \frac{\log x}{\sqrt{x}} \sum_{p \leq x} \chi_1(p) = \frac{\log x}{\sqrt{x}} (\pi_{q;R}(x) - \pi_{q;N}(x))$$

be the comparison function of this prime race. With this notation, along with the notations used before, Lemma 4.1 results in

$$\begin{aligned}
R(x) &= \frac{1}{\varphi(q)} \left(\sum_{a \in R} E(x; q, a) - \sum_{a \in N} E(x; q, a) \right) \\
&= \frac{1}{\varphi(q)} \sum_{a \in R} \left(-c(q, a) + \sum_{\chi \neq \chi_0} \bar{\chi}(a) \frac{\psi(x, \chi)}{\sqrt{x}} \right) \\
&\quad - \frac{1}{\varphi(q)} \sum_{a \in N} \left(-c(q, a) + \sum_{\chi \neq \chi_0} \bar{\chi}(a) \frac{\psi(x, \chi)}{\sqrt{x}} \right) + O\left(\frac{1}{\log x}\right) \\
&= -1 + \frac{1}{\varphi(q)} \sum_{\chi \neq \chi_0} \left(\sum_{a \in R} \bar{\chi}(a) - \sum_{a \in N} \bar{\chi}(a) \right) \frac{\psi(x, \chi)}{\sqrt{x}} + O\left(\frac{1}{\log x}\right) \\
&= -1 + \frac{\psi(x, \chi_1)}{\sqrt{x}} + O\left(\frac{1}{\log x}\right),
\end{aligned}$$

where we used the fact that $|R| = |N|$, $c(q, a) = 1$ for $a \in R$ and $c(q, a) = -1$ for $a \in N$,²⁵ interchanged summations, and used the following:

$$\sum_{a \in R} \bar{\chi}(a) - \sum_{a \in N} \bar{\chi}(a) = - \sum_{a \in \mathbb{Z}_q^\times} \bar{\chi}(a) + 2 \sum_{a \in R} \bar{\chi}(a) = \sum_{x \in \mathbb{Z}_q^\times} \bar{\chi}^2(x) = \begin{cases} 0 & \text{if } \chi^2 \neq \chi_0 \\ \varphi(q) & \text{if } \chi^2 = \chi_0. \end{cases}$$

Using the explicit formula (5) from Chapter 2 for $\psi(x, \chi_1)$, we can further write

$$R(e^y) = -1 - \sum_{|\gamma| \leq X} \frac{e^{i\gamma y}}{\frac{1}{2} + i\gamma} + O\left(\frac{e^{\frac{y}{2}} \log^2(e^y X)}{X}\right), \quad (80)$$

where $\frac{1}{2} + i\gamma$ runs through the zeros of $L(s, \chi_1)$ on $0 < \Re(s) < 1$.²⁶ Since

²⁵This holds generally only for $q = 2, 4$, a prime power or twice a prime power

²⁶The occurrence of -1 in this formula produces the factor $e^{i\xi}$ in the Bessel product (68).

$\frac{1}{2} + i\gamma$ is a zero of $L(s, \chi_1)$ if and only if $\frac{1}{2} - i\gamma$ is, we can rewrite the sum as

$$\begin{aligned}
R(e^y) - O\left(1 + \frac{e^{\frac{y}{2}}(y + \log X)^2}{X}\right) &= - \sum_{0 \leq \gamma \leq X} \left(\frac{e^{i\gamma y}}{\frac{1}{2} + i\gamma} + \frac{e^{-i\gamma y}}{\frac{1}{2} - i\gamma} \right) \\
&= -2 \sum_{0 \leq \gamma \leq X} \Re \left(\frac{e^{i\gamma y}}{\frac{1}{2} + i\gamma} \right) \\
&= -2 \sum_{0 \leq \gamma \leq X} \left(\frac{\gamma \sin \gamma y}{\frac{1}{4} + \gamma^2} + \frac{\frac{1}{2} \cos \gamma y}{\frac{1}{4} + \gamma^2} \right) \\
&= -2 \sum_{0 \leq \gamma \leq X} \frac{\gamma \sin \gamma y}{\frac{1}{4} + \gamma^2} + O(1) \\
&= -2 \sum_{0 \leq \gamma \leq X} \frac{\sin \gamma y}{\gamma} + O(1)
\end{aligned}$$

because $\sum_{\gamma} \frac{1}{\gamma^2} = O(1)$. This representation of $R(e^y)$ in terms of a weighted sum of the sines of the zeros γ is central to the proof of the lower bound for the tail of the measure, since

$$\mu_{q;R,N}(\{x \geq \lambda\}) = \lim_{x \rightarrow \infty} \frac{1}{x} m(\{y \in [0, x] : R(e^y) \geq \lambda\}) \quad (81)$$

by the same argument that we used to prove this formula for $\mu_{q;a_1, \dots, a_r}$. From the formula for $R(e^y)$ one can guess, although it is not trivial to prove, that $R(e^y)$ oscillates on both sides of 0 as y varies, meaning that both the quadratic nonresidue primes and the residue primes hold the lead for some time. This was already shown in the proof of the Main Theorem.

To consider $R(e^y)$ more carefully, let ε be a small positive number, m a positive integer, and $\xi \geq 2 + \frac{\varepsilon}{2}$. Define $F_\varepsilon(\xi)$ as the average of $R(e^y)$ over $y \in [\xi - \frac{\varepsilon}{2}, \xi + \frac{\varepsilon}{2}]$. We show that these averages are large assuming a certain condition, so the function $R(e^y)$ is also large on some intervals. We formulate this as a lemma.

Lemma 4.11. *We have $F_\varepsilon(\frac{m+1}{2}\varepsilon) > c \log \varepsilon^{-1}$ when*

$$\max_{0 \leq \gamma \leq \varepsilon^{-1}} \left\| \gamma m \frac{\varepsilon}{2} \right\| \leq \frac{d}{\log \varepsilon^{-1}},$$

where $c, d > 0$ are certain constants and $\| \cdot \|$ denotes the distance to the nearest multiple of 2π and γ denotes a zero of the function $L(s, \chi_1)$.

Proof. We begin by simplifying the definition of $F_\varepsilon(\xi)$.

$$\begin{aligned}
F_\varepsilon(\xi) &:= \frac{1}{\varepsilon} \int_{\xi - \frac{\varepsilon}{2}}^{\xi + \frac{\varepsilon}{2}} R(e^y) dy \\
&= -\frac{2}{\varepsilon} \int_{\xi - \frac{\varepsilon}{2}}^{\xi + \frac{\varepsilon}{2}} \sum_{0 \leq \gamma \leq X} \frac{\sin \gamma y}{\gamma} dy + \frac{1}{\varepsilon} O \left(\int_{\xi - \frac{\varepsilon}{2}}^{\xi + \frac{\varepsilon}{2}} \left(1 + \frac{e^{\frac{y}{2}} (y + \log X)^2}{X} \right) dy \right) \\
&= \frac{2}{\varepsilon} \sum_{0 \leq \gamma \leq X} \frac{\cos((\xi - \frac{\varepsilon}{2})\gamma) - \cos((\xi + \frac{\varepsilon}{2})\gamma)}{\gamma^2} \\
&\quad + \frac{1}{\varepsilon} O \left(\varepsilon + \frac{\left(e^{\frac{1}{2}(\xi + \frac{\varepsilon}{2})} - e^{\frac{1}{2}(\xi - \frac{\varepsilon}{2})} \right) (\xi + 1 + \log X)^2}{X} \right) \\
&= \frac{4}{\varepsilon} \sum_{0 \leq \gamma \leq X} \frac{\sin \gamma \xi \sin \frac{\gamma \varepsilon}{2}}{\gamma^2} + \frac{1}{\varepsilon} O \left(\varepsilon + \frac{\left(e^{\frac{1}{2}(\xi + \frac{\varepsilon}{2})} - e^{\frac{1}{2}(\xi - \frac{\varepsilon}{2})} \right) (\xi + 1 + \log X)^2}{X} \right) \\
&\rightarrow \frac{4}{\varepsilon} \sum_{0 \leq \gamma} \frac{\sin \gamma \xi \sin \frac{\gamma \varepsilon}{2}}{\gamma^2} + O(1)
\end{aligned}$$

as $X \rightarrow \infty$. We define

$$F_\varepsilon^*(\xi) := \frac{4}{\varepsilon} \sum_{0 \leq \gamma \leq \varepsilon^{-2}} \frac{\sin \gamma \xi \sin \frac{\gamma \varepsilon}{2}}{\gamma^2}; \quad (82)$$

then $F_\varepsilon^*(\xi) = F_\varepsilon(\xi) + O(1)$. The function $F_\varepsilon^*(\xi)$ is an "almost periodic" version of the average function (as a trigonometric polynomial, F_ε^* has a certain set of "almost periods" that are repeated with arbitrary precision, though not exactly).

Notice that for $0 < \varepsilon < 1$,

$$\begin{aligned}
F_\varepsilon^*\left(\frac{\varepsilon}{2}\right) &= \frac{4}{\varepsilon} \sum_{0 \leq \gamma \leq \varepsilon^{-2}} \frac{\sin^2 \frac{\gamma \varepsilon}{2}}{\gamma^2} \\
&\geq \frac{4}{\varepsilon} \sum_{0 \leq \gamma \leq \varepsilon^{-1}} \frac{\sin^2 \frac{\gamma \varepsilon}{2}}{\gamma^2} \\
&\geq 4\varepsilon \cos \frac{1}{2} \sum_{0 \leq \gamma \leq \varepsilon^{-1}} \left(\frac{\sin \frac{\gamma \varepsilon}{2}}{\gamma \varepsilon} \right)^2 \\
&\geq 4\varepsilon \cos \frac{1}{2} \sum_{0 \leq \gamma \leq \varepsilon^{-1}} \frac{1}{3^2} \geq c_0 \log \varepsilon^{-1},
\end{aligned}$$

since $|\frac{\gamma\varepsilon}{2}| \leq \frac{1}{2}$ implies $|\frac{\sin\frac{\gamma\varepsilon}{2}}{\gamma\varepsilon}| > \frac{1}{3}$, and the number of γ 's on $[0, \varepsilon^{-1}]$ is $\gg \varepsilon^{-1} \log \varepsilon^{-1}$ by the asymptotic formula for the number of zeros. We remark that here Rubinstein and Sarnak's paper makes a very minor mistake in considering $F_\varepsilon^*(\varepsilon)$, which does not behave as nicely as $F_\varepsilon^*(\frac{\varepsilon}{2})$ because there is much oscillation. In my MathOverflow question [21], I was informed that the matter is fixed just by considering $F_\varepsilon^*(\frac{\varepsilon}{2})$.

Using $|\sin x| \leq x$, we estimate

$$\begin{aligned} & |F_\varepsilon^*((m+1)\frac{\varepsilon}{2}) - F_\varepsilon^*(\frac{\varepsilon}{2})| \\ &= \frac{4}{\varepsilon} \left| \sum_{0 \leq \gamma \leq \varepsilon^{-2}} \frac{(\sin((m+1)\frac{\gamma\varepsilon}{2}) - \sin\frac{\gamma\varepsilon}{2}) \sin\frac{\gamma\varepsilon}{2}}{\gamma^2} \right| \\ &\leq 2 \sum_{0 \leq \gamma \leq \varepsilon^{-2}} \left| \frac{(\sin((m+1)\frac{\gamma\varepsilon}{2}) - \sin\frac{\gamma\varepsilon}{2})}{\gamma} \right| \\ &\leq 2 \max_{0 \leq \gamma \leq \varepsilon^{-2}} \|\gamma m \frac{\varepsilon}{2}\| \sum_{0 \leq \gamma \leq \varepsilon^{-2}} \frac{1}{\gamma} \\ &\leq \max_{0 \leq \gamma \leq \varepsilon^{-2}} \|\gamma m \frac{\varepsilon}{2}\| c_1 \log^2 \varepsilon^{-1}, \end{aligned}$$

where $\|\cdot\|$ is the distance to the nearest multiple of 2π , and we used $|\sin a - \sin b| \leq \|a - b\|$, which holds for $a, b \in [0, \pi)$ by the mean value theorem and can be extended to other values by the periodicity of the sine.

Now we choose $d = \frac{c_0}{2c_1}$ in the lemma we are proving. Then $F_\varepsilon^*((m+1)\frac{\varepsilon}{2}) \geq \frac{c_0}{2} \log \varepsilon^{-1}$. This implies, for some constant $c_2 > 0$ and small values of ε , the bound

$$F_\varepsilon\left(\frac{m+1}{2}\varepsilon\right) > c_2 \log \varepsilon^{-1},$$

so the lemma is proved. \square

Let m be an integer on $[\frac{2}{\varepsilon}, \frac{M}{\varepsilon}]$, where M is an integer variable tending to infinity later on. We would like to find many values of m such that the condition of Lemma 4.11 is satisfied, since this would give us intervals on which $R(e^y)$ is large (we formulate this soon as Lemma 4.12). Therefore, define

$$G_M := \left\{ m \in \left[\frac{2}{\varepsilon}, \frac{M}{\varepsilon} \right] : \max_{0 \leq \gamma \leq \varepsilon^{-2}} \|\gamma m \frac{\varepsilon}{2}\| \leq D \right\},$$

where $D = \frac{d}{\log \varepsilon^{-1}}$ and d is as before. We use pigeonhole principle to derive a lower bound on $|G_M|$. Consider the vectors $x_m = (\frac{\gamma_1 m}{4\pi}, \dots, \frac{\gamma_N m}{4\pi}) \bmod 1$, which lie in $[0, 1]^N$. Divide this cube into boxes of side length $\frac{1}{k}$, where k is the smallest integer such that $\frac{1}{k} \leq \frac{D}{2}$. The number of these boxes is $k^N \leq (\frac{3}{D})^N$ for small values of ε (recall that D depends on ε). Thus, there will be at least

$$\left\lceil \frac{M - 2}{\varepsilon(6c_1 c_0^{-1} \log \varepsilon^{-1})^N} \right\rceil$$

vectors x_m in some box. If m_1, \dots, m_ℓ are the corresponding indices, then the components of the vectors $x_{m_i} - x_{m_1}, i = 1, \dots, \ell$ satisfy the required bound for their distance from a multiple of 2π , so

$$|G_M| \geq \left\lceil \frac{M - \log 2}{\varepsilon(6c_1 c_0^{-1} \log \varepsilon^{-1})^N} \right\rceil. \quad (83)$$

As a function of M , which is a free (large) parameter, this bound is linear.

To transform bounds on $|G_M|$ to information about the positivity of $R(e^y)$, we also consider the squares of the quadratic averages of $R(e^y)$ defined by

$$q_m := \frac{1}{\varepsilon} \int_{\frac{m}{2}\varepsilon}^{\frac{m+2}{2}\varepsilon} R(e^y)^2 dy.$$

We need the following lemma.

Lemma 4.12. *If $m \in G_M$, the set*

$$\left\{ y \in \left[\frac{m}{2}\varepsilon, \frac{m+2}{2}\varepsilon \right] : R(e^y) \geq \frac{1}{2}c_2 \log \varepsilon^{-1} \right\}$$

has Lebesgue measure not less than

$$\frac{\varepsilon c_2^2 \log^2 \varepsilon^{-1}}{4q_m}.$$

Proof. Denote

$$\nu(\lambda) = \varepsilon^{-1} m \left(\left\{ y \in \left[\frac{m}{2}\varepsilon, \frac{m+2}{2}\varepsilon \right] : R(e^y) \leq \lambda \right\} \right),$$

which is the probability distribution function for $R(e^y)$ on the interval under consideration. By definition, $\int_{-\infty}^{\infty} \nu'(\lambda) d\lambda = 1$, and the expectation of ν' is

$$\int_{-\infty}^{\infty} \lambda \nu'(\lambda) d\lambda = \int_{-\infty}^{\infty} \nu(\lambda) d\lambda = \frac{1}{\varepsilon} \int_{\frac{m}{2}\varepsilon}^{\frac{m+2}{2}\varepsilon} R(e^y) dy = F_\varepsilon((m+1)\frac{\varepsilon}{2}) \geq c_2 \log \varepsilon^{-1}$$

by partial integration (ν is compactly supported) and a corollary of Fubini's theorem (Lemma 3.1) if $m \in G_M$. The variance is given by

$$\int_{-\infty}^{\infty} \lambda^2 \nu'(\lambda) = q_m,$$

again by Fubini's theorem. Because of

$$\int_{-\infty}^{\frac{c_2}{2} \log \varepsilon^{-1}} \lambda \nu'(\lambda) \leq \frac{c_2}{2} \log \varepsilon^{-1} \int_{-\infty}^{\frac{c_2}{2} \log \varepsilon^{-1}} \nu'(\lambda) = \frac{c_2}{2} \log \varepsilon^{-1} \nu \left(\frac{1}{2} c_2 \log \varepsilon^{-1} \right)$$

and the fact $\nu(\lambda) \leq 1$, we must have

$$\int_{\frac{1}{2} c_2 \log \varepsilon^{-1}}^{\infty} \lambda \nu'(\lambda) \geq \frac{1}{2} c_2 \log \varepsilon^{-1}.$$

Applying the Cauchy-Schwartz inequality, we deduce

$$\begin{aligned} \frac{1}{2} c_2 \log \varepsilon^{-1} &\leq \int_{\frac{1}{2} c_2 \log \varepsilon^{-1}}^{\infty} \lambda \sqrt{\nu'(\lambda)} \cdot \sqrt{\nu'(\lambda)} d\lambda \\ &\leq \left(\int_{\frac{1}{2} c_2 \log \varepsilon^{-1}}^{\infty} \lambda^2 \nu'(\lambda) \right)^{\frac{1}{2}} \left(\int_{\frac{1}{2} c_2 \log \varepsilon^{-1}}^{\infty} \nu'(\lambda) \right)^{\frac{1}{2}} \\ &\leq \sqrt{q_m} \nu \left(\frac{1}{2} c_2 \log \varepsilon^{-1} \right)^{\frac{1}{2}}, \end{aligned}$$

or

$$\nu \left(\frac{1}{2} c_2 \log \varepsilon^{-1} \right) \geq \frac{c_2^2 \log^2 \varepsilon^{-1}}{4q_m},$$

which proves the claim. \square

This lemma gives us a lower bound on the measure of the set in which the the function $R(e^y)$ is "large";

$$m(\{y \in [2, \frac{M+2}{2}\varepsilon] : R(e^y) \geq \frac{1}{2} c_2 \log \varepsilon^{-1}\}) \geq \sum_{m \in G_M} \frac{\varepsilon c_2^2 \log^2 \varepsilon^{-1}}{4q_m}. \quad (84)$$

Applying the inequality

$$\sum_{m \in G_M} \frac{1}{q_m} \sum_{m \in G_M} q_m \geq |G_M|^2,$$

which is a special case of the Cauchy-Schwartz inequality, we get for (84) a lower bound of

$$\geq |G_M|^2 \frac{1}{\sum_{m \in G_M} q_m} \frac{\varepsilon c_2^2 \log^2 \varepsilon^{-1}}{4}. \quad (85)$$

One has

$$\sum_{2 \leq m \varepsilon \leq M} q_m \leq \frac{1}{\varepsilon} \int_2^{\frac{M+2}{2}\varepsilon} R(e^y)^2 dy \leq \frac{c_3 M}{\varepsilon},$$

since applying Cauchy-Schwarz in the form $(\sum_{i=1}^n a_i)^2 \leq n \sum_{i=1}^n a_i^2$ one gets

$$\int_2^M R(e^y)^2 dy \ll_q \sum_{q \in \mathbb{Z}_q^\times} \int_2^M E(e^y; q, a)^2 dy,$$

and by choosing T to be a constant in Lemma 4.3, the upper bound of order M is obtained.

Consequently $\sum_{m \in G_M} q_m \leq c_3 M \varepsilon^{-1}$, so the previous estimate (85) is at least

$$\begin{aligned} &\geq \frac{|G_M|^2 \varepsilon^2 c_2^2 \log^2 \varepsilon^{-1}}{c_3 M \cdot 4} \\ &\geq \left[\frac{M-2}{\varepsilon (6c_1 c_0^{-1} \log \varepsilon^{-1})^N} \right]^2 \frac{\varepsilon^2 c_2^2 \log^2 \varepsilon^{-1}}{4c_3 M} \end{aligned}$$

by (83). We conclude, combining (84) with the previous inequality and with $\lfloor x \rfloor \geq \frac{x}{2}$, that

$$\frac{2M}{\varepsilon (M-2)^2} m(\{y \in [2, \frac{M+2}{2}\varepsilon] : R(e^y) \geq \frac{1}{2} c_2 \log \varepsilon^{-1}\}) \geq \frac{c_2^2 \log^2 \varepsilon^{-1}}{4c_3 (6c_1 c_0^{-1} \log \varepsilon^{-1})^{2N}}.$$

As $M \rightarrow \infty$, we know by (81) that the left-hand side approaches

$$\mu_{q;R,N}([\frac{1}{2} c_2 \log \varepsilon^{-1}, \infty)).$$

Choosing $\lambda = \frac{1}{2} c_2 \log \varepsilon^{-1}$, we have, as before, $N = N(\varepsilon^{-2}) \ll \varepsilon^{-2} \log \varepsilon^{-2} \ll \exp(A\lambda)$ for some constant $A > 0$. Therefore,

$$\mu_{q;R,N}([\lambda, \infty)) = \mu_{q;N,R}([\frac{1}{2} c_2 \log \varepsilon^{-1}, \infty)) \geq \frac{A_1 \lambda^2}{(A_2 \lambda)^{2 \exp(A_3 \lambda)}} \geq \exp(\exp(-a\lambda))$$

for a suitable constant $a > 0$ depending on q and any large λ . This completes the proof of the lower bound for the logarithmic density of the set on which the nonresidue primes lead over the residue primes. The proof of the lower bound of $\mu_{q,N;R}((-\infty, \lambda))$ is the same, except we have $F_\varepsilon^*(-\frac{\varepsilon}{2}) \ll -\varepsilon^{-1} \log \varepsilon^{-1}$ by the same argument, and then we consider $F_\varepsilon^*(-(m+1)\frac{\varepsilon}{2})$ for the same values of m . \square

References

- [1] Michael N. Barber, B. W. Niham: *Random and Restricted Walks*, Gordon and Breach (1970)
- [2] Carter Bays, Richard H. Hudson, *A new bound for the smallest x with $\pi(x) > li(x)$* , Mathematics of Computation, Vol. 69, No. 231, pp. 1285-1296
- [3] James K. Brooks, *On the Vitali-Hahn-Saks and Nikodým theorems*, Proc. Natl. Acad. Sci. USA. 64(2), pp. 468-471 (1969)
- [4] P. L. Chebyshev, *Lettre de M. le professeur Tchébichev à M. Fuss, sur un nouveau théoreme relatif aux nombres premiers contenus dans la formes $4n+1$ et $4n+3$* , Bull. de la Classe phys.-math. de l'Acad. Imp. des Sciences St. Petersburg 11, p. 208 (1853)
- [5] Harold Davenport: *Multiplicative Number Theory*, Springer(1980)
- [6] Andrey Feuerverger and Greg Martin: *Biases in the Shanks-Rényi prime number race*, Experimental Mathematics, Vol. 9, No. 4 (2000)
- [7] Kevin Ford, Sergei Konyvagin, *Chebyshev's conjecture and the prime number race*, <http://www.math.uiuc.edu/~ford/wwpapers/racefk.pdf> (2002)
- [8] C. F. Gauss, *Disquisitiones Arithmeticae*, translated by Arthur C. Clarke, Yale university press (1966)
- [9] Andrew Granville and Greg Martin: *Prime Number Races*, American Mathematical Monthly, Monthly, Vol. 113, No. 1, pp. 1-33 (2006).
- [10] Andrew Granville and H. M. Stark: *ABC implies no "Siegel Zeros" for L-functions of characters with negative discriminant*, Inventiones Mathematicae Vol. 139, No. 3, pp. 509-523 (2000)

- [11] A. E. Ingham: *The Distribution of Prime Numbers*, Cambridge University Press (1932)
- [12] Jean Jacod, Philip Protter, *Probability Essentials*, Springer (2004)
- [13] J. Kaczorowski, *A contribution to the Shanks-Rényi race problem*, Quart. J Math., Vol. 44, pp. 451-458 (1993)
- [14] Anatolij A. Karatsuba: *Basic Analytic Number Theory*, Springer (1993)
- [15] S. Knapowski, P. Turán: *Comparative Prime Number Theory I-VIII*, Acta Mathematica Hungarica (1962-1963)
- [16] N. N. Lebedev, *Special Functions and Their Applications*, Dover, Ch. 5 (1972)
- [17] J. E. Littlewood, *Distribution Des Nombres Premiers*, C. R. Acad. Sci. Paris, Vol. 175 (1914), pp. 1896-1872
- [18] Lynn H. Loomis, *An Introduction to Abstract Harmonic Analysis*, D. Van Nostrand Company, Ch. VI (1953)
- [19] Greg Martin, *Linear Independence of Zeros of L-functions*, slides at <http://www.math.ubc.ca/~gerg/slides/Toronto-7Oct11.pdf> (2011)
- [20] Greg Martin, Justin Sacrify, *Comparative prime number theory: a survey*, <http://arxiv.org/pdf/1202.3408v1.pdf> (2012)
- [21] MathOverflow question 161284: *A sum over zeros of L-functions in the paper "Chebychev's Bias"*, asked by Joni Teräväinen on March 24, 2014; answered by "GH from MO" and Igor Rivin on March 24-25, 2014. <http://mathoverflow.net/questions/161284/a-sum-over-zeros-of-l-functions-in-the-paper-chebychevs-bias>
- [22] MathOverflow question 162875: *Reference for Kronecker-Weyl theorem in full generality*, asked by Greg Martin on April 9, 2014; answered by Peter Humphries on April 9, 2014. <http://mathoverflow.net/questions/162875/reference-for-kronecker-weyl-theorem-in-full-generality>
- [23] Hugh L. Montgomery, *The Zeta Function and Prime Numbers*, Proceedings of the Queen's Number Theory Conference, pp. 12-24 (1980)
- [24] M. Ram Murty: *Problems in Analytic Number Theory*, Springer (2001)

- [25] W. Narkiewicz, *Development of Prime Number Theory: From Euclid to Hardy and Littlewood*, Springer (2000)
- [26] Nathan Ng, *Limiting distributions and zeros of Artin L-functions*, Ph.D. Thesis, University of British Columbia, www.cs.uleth.ca/~nathanng/RESEARCH/phd.thesis.pdf (2000)
- [27] Bernhard Riemann, *Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse*, <http://www.claymath.org/sites/default/files/zeta.pdf> (1860)
- [28] Michael Rubinstein and Peter Sarnak: *On Chebyshev's Bias*, *Experimental Mathematics*, Vol. 3, No. 3 (1994)
- [29] A. W. van der Vaart: *Asymptotic statistics*, Cambridge University Press, Ch. 2 (1998)
- [30] Walter Rudin: *Real and Complex Analysis*, McGraw-Hill (1987)
- [31] J. Schur, *Über lineare Transformationen in der Theorie der unendlichen Reihen*, *Journal für die reine und angewandte Mathematik*, 151, pp. 79-111 (1921).
- [32] Elias M. Stein, Rami Shakarchi, *Fourier Analysis: An Introduction*, Princeton University Press(2003)
- [33] Elias M. Stein, Rami Shakarchi, *Complex Analysis*, Princeton University Press (2003)
- [34] Elias M. Stein, Rami Shakarchi, *Measure Theory, Integration and Hilbert Spaces*, Princeton University Press (2005)
- [35] Cem Yalçm Yıldırım, *A Survey of Results on Primes in Short Intervals*, <http://www.math.boun.edu.tr/instructors/yildirim/paper/paper9.pdf>