The Distribution of Prime Numbers

Dimitris Koukoulopoulos

To Jennifer

Contents

Preface

The main goal of this book is to introduce beginning graduate students to analytic number theory. In addition, large parts of it are suitable for advanced undergraduate students with a good grasp of analytic techniques.

Throughout, the emphasis has been put on exposing the main ideas rather than providing the most general results known. Any student wishing to do serious research in analytic number theory should broaden and deepen their knowledge by consulting some of the several excellent research-level books on the subject. Examples include: the books of Davenport [31] and of Montgomery-Vaughan [146] for classical multiplicative number theory; Tenenbaum's book [172] for probabilistic number theory and the saddle-point method; the book by Iwaniec-Kowalski [114] for the general theory of L-functions, of modular forms and of exponential sums; Montgomery's book [144] for the harmonic analytic aspects of analytic number theory; and the book of Friedlander-Iwaniec [59] for sieve methods.

Using the book

The book borrows the structure of Davenport's masterpiece *Multiplicative Number Theory* with several short- to medium-length chapters. Each chapter is accompanied by various exercises. Some of them aim to exemplify the concepts discussed, while others are used to guide the students to self-discover certain more advanced topics. A star next to an exercise indicates that its solution requires total mastery of the material.

The contents of the book are naturally divided into six parts as indicated in the table of contents. The first two parts study elementary and classical complex-analytic methods. They could thus serve as the manual for an

introductory graduate course to analytic number theory. The last three parts of the book are devoted to the theory of sieves: Part 4 presents the basic elements of the theory of the small sieve, whereas Part 5 explores the method of bilinear sums and develops the large sieve. These techniques are then combined in Part 6 to study the spacing distribution of prime numbers and prove some of the recent spectacular results about small and large gaps between primes. Finally, Part 3 studies multiplicative functions and the anatomy of integers, and serves as a bridge between the complex-analytic techniques and the more elementary theory of sieves. Topics from it could be presented either in the end of an introductory course to analytic number theory (Chapter 13 most appropriately), or in the beginning of a more advanced course on sieves (the most relevant material is contained in Chapters 14 and 15, as well as in Theorem 16.1).

Certain portions of the book can be used as a reference for an undergraduate course. More precisely, Chapters 1–8 can serve as the core of such a course, followed by a selection of topics from Chapters 14, 15, 17 and 21.

A short guide to the main theorems of the book. Below is a list of the main results proven and of their prerequisites.

Chebyshev's and Mertens' estimates are presented in Chapters 2 and 3, respectively. Their proofs rest on the material contained in Part 1.

The landmark *Prime Number Theorem* is proven in Chapter 8. Understanding it requires a good grasp of all preceding chapters.

The *Siegel-Walfisz theorem*, which is a uniform version of the Prime Number Theorem for arithmetic progressions, is presented in Chapter 12. Its proof builds on all of the material preceding it.

The *Landau-Selberg-Delange method* is a key tool in the study of multiplicative functions. It is presented in Chapter 13. Appreciating its proof requires a firm understanding of Chapters 1–8 for the main analytic tools, as well as of Chapter 12 for dealing with uniformity issues.

The foundations of *probabilistic number theory* are explained in Chapters 15 and 16, where the *Erdős-Kac theorem* and the *Sathe-Selberg theorem* are proven. The main prerequisites can be found in Part 1 and in Chapter 14. In addition, Chapter 13 is needed for the Sathe-Selberg theorem.

The *Fundamental Lemma of Sieve Theory* is proven in Chapter 19. Its proof uses ideas and techniques from Part 1 and Chapters 14–17.

Vinogradov's method, one of the foundations of modern analytic number theory, is presented in Chapter 23. It builds on the material of Chapters 1–12 and 19.

The *Hardy-Littlewood circle method* is presented in Chapter 24. It is used to detect additive patterns among the primes and, more specifically, to count ternary arithmetic progressions all of whose members are primes.

The *Bombieri-Vinogradov theorem*, often called the "Generalized Riemann Hypothesis on average", is established in Chapter 26. Understanding its proof requires mastery of Vinogradov's method (Chapter 23) and of the *large sieve* (Chapter 25).

Linnik's theorem provides a very strong bound on the least prime in an arithmetic progression. It is proven in Chapter 27 and its prerequisites are Chapters 1–12, 17–20, 22–23 and 25.

The breakthrough of Zhang-Maynard-Tao about the existence of infinitely many *bounded gaps between primes* is presented in Chapter 28. Its proof requires a firm understanding of the Fundamental Lemma of Sieve Theory (Chapter 19), of Selberg's sieve (Chapter 21) and of the Bombieri-Vinogradov theorem (Chapter 26).

The recent developments about *large gaps between primes* of Ford-Green-Konyagin-Tao and Maynard are presented in Chapter 29. Understanding them necessitates knowledge of the same concepts as the proof of the existence of bounded gaps between primes, with the addition of the results on smooth numbers presented in Chapters 14 and 16.

Maier discovered in 1985 that the distribution of prime numbers has certain unexpected irregularities. His results are presented in Chapter 30 and they assume knowledge of Linnik's theorem (and of its prerequisites), as well as of Buchstab's function (see Chapter 14 and, more precisely, Theorem 14.4).

Acknowledgments

Many people have helped me greatly in many different ways in writing this book.

I am indebted to Leo Goldmakher and James Maynard, with whom I discussed the contents of the book extensively at various stages of the writing process. In addition, an early version of the manuscript was used as a teaching reference by Wei Ho at the University of Michigan, and by Leo Goldmakher at Williams College. I am grateful to them and their students for the valuable feedback they provided.

I am obliged to Martin Čech, Tony Haddad, Youcef Mokrani, Alexis Leroux-Lapierre, Joëlle Matte, Kunjakanan Nath, Stelios Sachpazis, Simon St-Amant, Jeremie Turcotte and Peter Zenz, who patiently studied earlier versions of the book, catching various errors and providing many excellent comments.

I have had very useful mathematical conversations with Sandro Bettin, Brian Conrey, Chantal David, Ben Green, Adam Harper, Jean Lagacé and K. Soundararajan on certain topics of the book; I am grateful to them for their astute remarks. Furthermore, I would like to thank the anonymous reviewers for their suggestions that helped me improve the exposition of the ideas in the manuscript, especially those related to the bilinear methods presented in Part 5.

I am indebted to Kevin Ford and Andrew Granville, who taught me analytic number theory. Their influence is evident throughout the book.

A special thanks goes to Ina Mette, Marcia Almeida and Becky Rivard for guiding me through the publishing process. I would also like to thank Brian Bartling and Barbara Beeton for their assistance with several typesetting questions, as well as Alexis Leroux-Lapierre for his help with designing the figures that appear in the book.

Last but not least, I would like to thank my wife Jennifer Crisafulli for her love, support and companionship. This book could not have been written without her and I wholeheartedly dedicate it to her.

Funding. During the writing process, I was supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant 2018-05699) and by the Fonds de recherche du Québec—Nature et technologies (projet de recherche en équipe—256442). Part of the book writing took place during my visit at the Mathematical Sciences Research Institute of Berkeley in the Spring of 2017 (funded by the National Science Foundation under Grant No. DMS-1440140), at the University of Oxford in the Spring of 2019 (funded by Ben Green's Simons Investigator Grant 376201) and at the University of Genova in June 2019 (funded by the Istituto Nazionale di Alta Matematica "Francesco Severi"). I would like to thank my hosts for their support and hospitality.

Notation

Throughout the book, we make use of some standard and some less standard notation. We list here the most important conventions.

The symbols $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$ denote the sets of natural numbers (we do not include zero in $\mathbb{N}$), integers, rational numbers, real numbers and complex numbers, respectively. Furthermore, given an integer $n \geqslant 1$, we write $\mathbb{Z}/n\mathbb{Z}$ for the set of residues mod n, as well as $(\mathbb{Z}/n\mathbb{Z})^*$ for the set of reduced residues mod n.

We write $\mathbb{P}$ to indicate a probability measure, and $\mathbb{E}[X]$ and $\mathbb{V}[X]$ for the expectation and the variance, respectively, of a random variable X.

Given a set of real numbers A and a parameter y, we write $A_{\leqslant y}$ for the set of numbers $a \in A$ that are $\leqslant y$; similarly for $A_{>y}$, $A_{\geqslant y}$, $A_{<y}$. We also write $|A|$ or $\#A$ for the cardinality of A, whichever is more convenient.

The letter p always denotes a prime, and the letter n always denotes an integer (usually, a natural number). We write $d|n$ to mean that d divides n, and that $p^k\|n$ to mean that p^k is the exact power of p dividing n. Lastly, $d|n^\infty$ means that all prime factors of d appear in the factorization of n too.

When we write (a, b), we might mean the open interval with endpoints a and b, the pair of a and b, or the greatest common divisor of the integers a and b. The meaning will always be clear from the context. Similarly, the symbol $[a, b]$ will sometimes denote the closed interval with endpoints a and b, and some other times the least common multiple of the integers a and b.

We write $P^+(n)$ and $P^-(n)$ to denote the largest and smallest prime factors of n, respectively, with the convention that $P^+(1) = 1$ and $P^-(1) = \infty$. Given a parameter y and an integer $n \geqslant 1$, we say that n is *y-smooth* if all its prime factors are $\leqslant y$ (i.e., if $P^+(n) \leqslant y$). The set of y-smooth

numbers is denoted by $\mathcal{S}(y)$. Lastly, we say that n is *y-rough* if all its prime factors are $> y$ (i.e., if $P^-(n) > y$). Equivalently, $(n, P(y)) = 1$, where $P(y) := \prod_{p \leqslant y} p$.

The symbol log denotes the natural logarithm (base e). We also let $\mathrm{li}(x) = \int_2^x \mathrm{d}t/\log t$ denote the *logarithmic integral*.

Given $x \in \mathbb{R}$, we write $\lfloor x \rfloor$ for its integer part (defined to equal $\max \mathbb{Z}_{\leqslant x}$, and also called the "floor" of x), $\lceil x \rceil$ for the "ceiling" of x (defined to equal $\min \mathbb{Z}_{\geqslant x}$) and $\{x\}$ for the fractional part of x (defined to equal $x - \lfloor x \rfloor$).

Given $\alpha \in \mathbb{R}$, we write $\|\alpha\|$ to denote its distance from the nearest integer. On the other hand, if ψ is a bilinear form, then $\|\psi\|$ denotes its norm (see Chapter 25). Finally, if $\vec{v} \in \mathbb{C}^n$ or $f : \mathbb{N} \to \mathbb{C}$ is an arithmetic function, we write $\|\vec{v}\|_2$ and $\|f\|_2$ for their ℓ^2-norm.

The symbol $C^k(X)$, where $X \subseteq \mathbb{R}$ and $k \in \mathbb{Z}_{\geqslant 0} \cup \{\infty\}$, denotes the set of functions $f : X \to \mathbb{C}$ whose first k derivatives exist and are continuous.

We write 1_E to denote the indicator function of a set or of an event E. For example, $1_{[0,1]}$ denotes the indicator function of the interval $[0, 1]$ and $1_{(n,10)=1}$ denotes the indicator function of the event that n is coprime to 10. In particular, 1_P will denote the indicator function of the set of primes.

The letter s will usually denote a complex number, in which case we denote its real part by σ and its imaginary part by t following Riemann's original notation that has now become standard. In addition, non-trivial zeroes of the Riemann zeta function and of Dirichlet L-functions will be denoted by $\rho = \beta + i\gamma$. Notice that we also use the letter γ for the Euler-Mascheroni constant, whereas $\rho(u)$ will also refer to the Dickman-de Bruijn function. The precise meaning of each letter will be clear from the context.

We employ frequently the usual asymptotic notation $f = O(g)$, $f \ll g$, $f \asymp g$, $f \sim g$ and $f = o(g)$, whose precise definition is given in Chapter 1.

Finally, we list below some other symbols and the page of their definition:

And then there were infinitely many

Ever since Euclid's proof of the infinitude of prime numbers, the distribution of these fundamental objects has fascinated mathematicians. Unlike other special sets of integers that have a very regular structure, such as the set of perfect squares, the primes do not follow any apparent pattern. Consequently, guessing the exact location of the nth smallest prime number seems to be an impossible challenge as n grows to be larger and larger.[1]

Since the sequence of primes appears to be so chaotic, we can set the more modest goal of understanding what is the *approximate* location of the nth smallest prime, which we denote by p_n. Equivalently, we seek a good approximation for the counting function of prime numbers

$$\pi(x) := \#\{p \leqslant x\}.$$

Indeed, we have that $\pi(p_n) = n$, so that any approximation of $\pi(x)$ can be immediately translated to an approximation of p_n, and vice versa.

The study of the distribution of primes preoccupied the young Gauss. After examining tables of large primes, he observed that their density around x is about $1/\log x$. Translated into the language of Calculus, this means that a good approximation for $\pi(x)$ is given by the *logarithmic integral*

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

[1] Even the simpler question of deciding whether a given large integer is prime was proven to be a very hard challenge. It was only in 2004 that Agrawal, Kayal and Saxena [1] constructed a deterministic algorithm that solves this problem in polynomial time without relying on any unproven conjectures.

Using L'Hôpital's rule, we find that

$$\lim_{x\to\infty} \frac{\mathrm{li}(x)}{x/\log x} = 1,$$

so that Gauss's guess implies that $\pi(x)$ is approximately equal to $x/\log x$ for large x. Symbolically, we write

$$\pi(x) \sim \frac{x}{\log x} \qquad (x \to \infty) \tag{0.1}$$

meaning that the ratio of these two functions tends to 1 as $x \to \infty$. This notation will be discussed in greater length in Chapter 1. Equivalently, Gauss's guess (0.1) says that $p_n \sim n \log n$ as $n \to \infty$.

It took more than a century to prove Gauss's conjecture for $\pi(x)$. The path to the proof was outlined by his student Riemann in his epoque making mémoire *Über die Anzahl der Primzahlen unter einer gegebenen Grösse* published in 1859. In this work, Riemann explained how $\pi(x)$ is intimately connected to analytic properties of the function

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

now called the *Riemann zeta function*. He then proposed a program whose completion would lead to a profound understanding of the distribution of prime numbers. In particular, it would establish the existence of a constant $c > 0$ such that

$$|\pi(x) - \mathrm{li}(x)| \leqslant c\sqrt{x}\log x \qquad \text{for all} \quad x \geqslant 2, \tag{0.2}$$

a very strong form of confirmation of Gauss's guess. By 1895, Hadamard [81] and von Mangoldt [136] had proved rigorously all but one steps in Riemann's master plan. The last step however remains elusive to this date. It is the famous *Riemann Hypothesis* that we will discuss in Chapter 8. Nevertheless, in 1896, Hadamard [83] and de la Vallée Poussin [175] proved a weak form of the Riemann Hypothesis that was strong enough to lead to a proof of Gauss's conjecture, now called the Prime Number Theorem.

Prime Number Theorem. *As $x \to \infty$, we have that $\pi(x) \sim x/\log x$.*

We will give the proof of this fundamental result in Chapter 8.

Except for the size of $\pi(x)$, there are many other interesting questions about prime numbers that concern the existence of various patterns among them. To understand such patterns, we assume a probabilistic point of view.

Indeed, the absence of structure in the sequence of primes might lead one to expect that they behave as if they were random objects. Specifically, in 1936 Cramér proposed to model the statistical properties of prime numbers as follows: we consider a sequence of random variables $(X_1, X_2, X_3, \dots)$ that

we think of as a model of the indicator function of the primes. That is to say, X_n models the event that a "randomly chosen" integer n is prime. Hence X_n must be a Bernoulli random variable (i.e., it only takes the values 0 and 1). In addition, Gauss's guess that the density of primes around x is $1/\log x$ can be interpreted to mean that the chances of a random integer n being prime are about $1/\log n$. We thus take

$$\mathbb{P}(X_n = 1) = 1/\log n \qquad (n \geqslant 3), \tag{0.3}$$

so that $\mathbb{P}(X_n = 0) = 1 - 1/\log n$, and we set for completion $X_1 = 0$ and $X_2 = 1$. Finally, since knowledge of the primality of some integer n does not seem to offer much information about the primality of another integer n', we assume that the random variables X_n are independent of each other.

The sequence $(X_n)_{n=1}^{\infty}$ is called *Cramér's model*. It naturally gives rise to the set of "random primes" $\{\, n \in \mathbb{N} : X_n = 1 \,\}$. We denote its elements by $P_1 < P_2 < \cdots$. By construction, it easily follows that $P_n \sim n \log n$ with probability 1 as $n \to \infty$, that is to say, if we fix $\varepsilon > 0$ and take n large in terms of ε, then $|P_n - n \log n| \leqslant \varepsilon n \log n$ with probability 1. Actually, more is true: the analogue of $\pi(x)$ is the random variable

$$\Pi(x) = \#\{P_n \leqslant x\} = \sum_{n \leqslant x} X_n.$$

We have that

$$\mathbb{E}[\Pi(x)] = 1 + \sum_{3 \leqslant n \leqslant x} \frac{1}{\log n},$$

which is essentially a Riemann sum of the logarithmic integral $\mathrm{li}(x)$. In fact, a consequence of Theorem 1.10 below is that $|\mathbb{E}[\Pi(x)] - \mathrm{li}(x)| \leqslant 10$. Similarly, the independence of the random variables X_n implies that

$$\mathbb{V}[\Pi(x)] = \sum_{n \leqslant x} \mathbb{V}[X_n] \sim \frac{x}{\log x}$$

as $x \to \infty$. Applying the law of the iterated logarithm [117], we find that

$$\begin{aligned} |\Pi(x) - \mathrm{li}(x)| &\leqslant \sqrt{(2+\varepsilon)\mathbb{V}[\Pi(x)] \log\log(\mathbb{V}[\Pi(x)])} \\ &\sim \sqrt{(2+\varepsilon)x(\log\log x)/\log x} \end{aligned} \tag{0.4}$$

almost surely as $x \to \infty$, where ε is any fixed positive real number. Comparing this inequality with (0.2), we see that $\Pi(x)$, which is the random model of $\pi(x)$, satisfies the Riemann Hypothesis with probability 1.

If primes really do behave like a sequence of random variables such as $(X_n)_{n=1}^{\infty}$, then we should be able to find all sorts of patterns among them. For example, there should be many primes of the form $4n+1$, or of the form n^2+1. Moreover, the mutual independence of the variables X_n suggests that we should be able to make several integers prime simultaneously. For

example, there should be infinitely many n such that the integers n and $n+2$ are both primes, in which case they are called a pair of *twin primes*. Similarly, the triplet $(n, 2n+1, n^2+6)$ should have prime coordinates infinitely often. One should be careful not to take such arguments too far: the integers n and $n+1$ can be simultaneously prime only when $n = 2$, because at least one of them is even. More subtly, if $n > 3$, then we cannot make n and n^2+2 simultaneously prime, because $n \equiv \pm 1 \pmod 3$ and thus $n^2+2 \equiv 0 \pmod 3$. In Chapter 17, we will see a way to modify Cramér's model so that it takes into account such "local" (i.e., involving congruences) obstructions to primality.

Despite the limitations of Cramér's model, all indications we have so far support the hypothesis that primes behave as if they were random. Throughout this book, we will present various results that are in accordance with this hypothesis. Specifically, in Chapter 12, we will prove that there are infinitely many primes of the form $4n+1$. More generally, we will prove that every arithmetic progression $qn+a$ contains infinitely many primes, as long as the obvious necessary condition that a and q are coprime holds. As a matter of fact, we will show that primes are equidistributed among these reduced arithmetic progressions.

Prime Number Theorem for arithmetic progressions. *Let $q \geqslant 3$, and let $\varphi(q) = \#(\mathbb{Z}/q\mathbb{Z})^*$ be Euler's totient function. If $(a, q) = 1$, then*

$$\pi(x; q, a) := \#\{\, p \leqslant x : p \equiv a \pmod q \,\} \sim \frac{x}{\varphi(q) \log x} \quad (x \to \infty).$$

On the other hand, it is not known to this day whether there are infinitely many pairs of twin primes. We do have two partial substitutes of this conjecture: Chen [24, 25] proved that there are infinitely many primes p such that $p+2$ is the product of at most two primes. We will prove a weaker version of Chen's theorem in Chapter 18. In addition, Zhang [188] proved that there is some $h \in \mathbb{N}$ such that the tuple $(n, n+1, \dots, n+h)$ contains at least two primes for infinitely many integers n. Maynard [138] and Tao [171] improved this result by showing that for each m there is some $h = h(m)$ such that the tuple $(n, n+1, \dots, n+h)$ contains at least m primes for infinitely many n. We will present the results of Zhang-Maynard-Tao in Chapter 28.

Substantial progress has also been made on the existence of arbitrarily long arithmetic progressions among primes: in 2008, Green and Tao [77] proved that for each $k \geqslant 2$, there are infinitely many integers n and d such that the numbers n, $n+d$, $\dots$, $n+kd$ are all primes. We will prove the case $k = 2$ of this result in Chapter 24 (that essentially goes back to work of I. M. Vinogradov).

On the contrary, prime values of non-linear polynomials remain a mystery: there is not a single example of a univariate polynomial of degree

at least 2 that provably takes prime values infinitely often. However, we have robust methods of bounding from above the frequency with which a given polynomial takes prime values, as we will see in Chapter 19. In addition, there has been significant progress in multivariate polynomials in recent years, starting with the work of Friedlander and Iwaniec [58] and of Heath-Brown [98], and continuing with its extensions due to Heath-Brown, Li, Maynard and Moroz [99–101, 141].

As the above discussion shows, our knowledge about primes is rather sporadic, and the deeper and more complex properties of these fundamental objects seem to escape us despite the collective efforts of mathematicians since the time of Euclid. Proving that prime numbers behave pseudorandomly and can be located inside interesting arithmetic sequences is one of the holy grails of analytic number theory. The purpose of this book is to present some of our best tools towards this grand goal.

Part 1

First principles

Chapter 1

Asymptotic estimates

The functions we encounter in number theory are often irregular. It is then desirable to approximate them by simpler functions that are easier to analyze. As an example, consider the function $f(x)$ that counts the number of integers in the interval $[1, x]$ with $x \geqslant 1$. We can easily see that f is a step function with jumps of length 1 at all integers. This function may be written in terms of a more familiar function: the integer part of x, denoted by $\lfloor x \rfloor$. This is the unique integer satisfying the inequalities $\lfloor x \rfloor \leqslant x < \lfloor x \rfloor + 1$. It is then clear that $f(x) = \lfloor x \rfloor$, whence $f(x) = x + E(x)$ for some function $E(x)$ that is bounded by 1 in absolute value. We have thus approximated the step function $f(x) = \lfloor x \rfloor$ by the smooth function x, and the remainder term in this approximation is a bounded function. We express this via the *asymptotic formula*

$$\lfloor x \rfloor = x + O(1). \tag{1.1}$$

In general, given complex-valued functions f, g and h, and a subset I of their domain of definition, we write

$$f(x) = g(x) + O(h(x)) \qquad (x \in I) \tag{1.2}$$

and read "$f(x)$ equals $g(x)$ plus big-Oh of $h(x)$" if there is a constant $c = c(f, g, I)$ such that

$$|f(x) - g(x)| \leqslant c \cdot h(x) \quad \text{for each } x \in I.$$

We will often refer to $g(x)$ as the *main term* of (1.2), and to $O(h(x))$ as the *remainder term* or *error term*. In addition, we will often call the constant c *absolute* to mean that it does not depend on the argument of the functions f, g and h, nor on various other parameters that might be present.

Notice that in (1.1) the difference $x - \lfloor x \rfloor$ is the fractional part of x, denoted by $\{x\}$. However, it turns out that it is often simpler to ignore the exact value of the remainder term and to only keep track of the fact that it is a bounded function. Suppose, for example, that we want to approximate the expression $\sum_{n \leqslant x} \lfloor n\sqrt{2} \rfloor$. Applying (1.1) to each of the $\lfloor x \rfloor$ summands, we find that

$$\sum_{n \leqslant x} \lfloor n\sqrt{2} \rfloor = \sum_{n \leqslant x} \big(n\sqrt{2} + O(1)\big) = \sqrt{2} \sum_{n \leqslant x} n + O(x), \tag{1.3}$$

since the total error is the sum of $\lfloor x \rfloor$ functions of size $O(1)$. On the other hand, we know that $1 + 2 + \cdots + N = N(N+1)/2$. Applying this with $N = \lfloor x \rfloor = x + O(1)$, we conclude that

$$\sum_{n \leqslant x} \lfloor n\sqrt{2} \rfloor = \sqrt{2} \cdot \frac{(x + O(1)) \cdot (x + O(1))}{2} + O(x) = \frac{\sqrt{2}}{2} x^2 + O(x)$$

for $x \geqslant 1$, since $O(x) + O(x) + O(1) = O(x)$ when $x \geqslant 1$. Indeed, the notation $O(x) + O(x) + O(1)$ denotes a sum $f_1(x) + f_2(x) + f_3(x)$ for which there are absolute constants $c_1, c_2, c_3 \geqslant 0$ such that $|f_j(x)| \leqslant c_j x$ for $j = 1, 2$ and $|f_3(x)| \leqslant c_3$. Hence, $|f_1(x) + f_2(x) + f_3(x)| \leqslant (c_1 + c_2 + c_3)x$ for $x \geqslant 1$.

Remark 1.1. As we see in the above example, the power of the asymptotic notation is that it allows us to turn inequalities into equalities and it is thus amenable to algebraic manipulations. Beware though that the rules of addition and multiplication change when we use asymptotic notation. For example, $O(1)+O(1) = O(1)$, since the sum of two bounded functions is also bounded. Similarly, we have $O(1) \cdot O(1) = O(1)$ and $O(1) - O(1) = O(1)$. On the other hand, if we sum an unbounded number of bounded functions (as in (1.3)), the error term must reflect this by growing linearly in the number of summands. □

The asymptotic notation also allows us to compare the *order of magnitude* of different functions: if

$$f(x) = O(g(x)) \qquad (x \in I), \tag{1.4}$$

we say that "f has smaller or equal order of magnitude than g in I". Often, we express this relation using *Vinogradov's notation*

$$f(x) \ll g(x) \qquad (x \in I),$$

which has the exact same meaning as (1.4).

If $f(x) \ll g(x)$ and $g(x) \ll f(x)$ for $x \in I$, we write

$$f(x) \asymp g(x) \qquad (x \in I)$$

and we say that "f and g have the same order of magnitude in I".

Remark 1.2. The range I in which we compare the functions is important. For instance, $\sqrt{x} \ll x$ when $x \geqslant 1$, but $x \ll \sqrt{x}$ when $x \in [0,1]$. □

Remark 1.3. Sometimes, the functions f and g we are comparing depend on various parameters. It is then possible that the implied constant in the estimate $f(x) \ll g(x)$ depends on these parameters. If so, we will indicate this dependence by a subscript. For instance, for each fixed $\varepsilon > 0$, we have

$$\log x \ll_\varepsilon x^\varepsilon \qquad (x \geqslant 1).$$ □

There are two more related definitions of asymptotic notation that will be important throughout this book and they concern the limiting behavior of functions. We write

$$f(x) \sim g(x) \quad (x \to x_0) \qquad \Longleftrightarrow \qquad \lim_{x \to x_0} \frac{f(x)}{g(x)} = 1,$$

where $x_0 \in \widehat{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$ and g is non-zero in an open neighborhood of x_0. Under the same assumptions, we also introduce the notation

$$f(x) = o(g(x)) \quad (x \to x_0) \qquad \Longleftrightarrow \qquad \lim_{x \to x_0} \frac{f(x)}{g(x)} = 0$$

or, for brevity,

$$f(x) = o_{x \to x_0}(g(x)) \qquad \Longleftrightarrow \qquad \lim_{x \to x_0} \frac{f(x)}{g(x)} = 0.$$

Notice that if $f(x) = o(g(x))$ as $x \to x_0$, then $f(x)$ has genuinely smaller order of magnitude than $g(x)$ in the vicinity of x_0.

We give below some examples to illustrate the use of the above asymptotic notation.

Example 1.4. Often we have a composite expression that we want to evaluate asymptotically, such as $\log\lfloor x \rfloor$. Since $\lfloor x \rfloor = x + O(1)$, the Mean Value Theorem implies that

$$\log \lfloor x \rfloor = \log x + O(1) \cdot \frac{1}{c}$$

for some c between $\lfloor x \rfloor$ and x. Thus

$$\log \lfloor x \rfloor = \log x + O(1/x) \quad (x \geqslant 1).$$ □

Example 1.5. A simple application of the Mean Value Theorem is sometimes not sufficient because we need more precision in our approximation. We may then employ Taylor's theorem. For example, we have

$$\sqrt{x + \log x} = \sqrt{x} + \frac{\log x}{2\sqrt{x}} - \frac{\log^2 x}{8x^{3/2}} + O\Big(\frac{\log^3 x}{x^{5/2}}\Big) \quad (x \geqslant 1).$$ □

Example 1.6. The asymptotic notation can also be used to obtain asymptotic expansions of integrals that cannot be computed in terms of elementary functions. As an example, we analyze the logarithmic integral $\mathrm{li}(x) = \int_2^x \mathrm{d}y/\log y$. We integrate by parts repeatedly to find that

$$\begin{aligned}
\mathrm{li}(x) &= \frac{y}{\log y}\Big|_{y=2}^{x} - \int_2^x y\,\mathrm{d}\Big(\frac{1}{\log y}\Big) \\
&= \frac{x}{\log x} + O(1) + \int_2^x \frac{1}{\log^2 y}\,\mathrm{d}y \\
&= \frac{x}{\log x} + \frac{x}{\log^2 x} + O(1) + 2\int_2^x \frac{\mathrm{d}y}{\log^3 y} \\
&\;\;\vdots \\
&= \frac{x}{\log x} + \frac{x}{\log^2 x} + \cdots + \frac{(N-1)!x}{\log^N x} + O_N(1) + N!\int_2^x \frac{\mathrm{d}y}{\log^{N+1} y}.
\end{aligned}$$

The last integral is $\sim x/\log^{N+1} x$ as $x \to \infty$ by L'Hôpital's rule, so we arrive at the asymptotic formula

$$\mathrm{li}(x) = \frac{x}{\log x} + \frac{x}{\log^2 x} + \frac{2!x}{\log^3 x} + \cdots + \frac{(N-1)!x}{\log^N x} + O_N\Big(\frac{x}{(\log x)^{N+1}}\Big)$$

for $x \geqslant 2$. □

Summation by parts

Many theorems of analytic number theory can be phrased as asymptotic estimates for the *summatory function* of a sequence $(a_n)_{n=1}^{\infty}$ of complex numbers. This is the function

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

where $x \in \mathbb{R}_{\geqslant 1}$. For instance, if $a_n = 1_P(n)$ (the indicator function of prime numbers), then its summatory function $A(x)$ is the counting function of prime numbers $\pi(x)$.

The simplest case is when $a_n = f(n)$, with $f \in C^1([0, +\infty))$. If f does not vary too rapidly, it is reasonable to expect that

$$f(n) \approx \int_{n-1}^{n} f(t)\mathrm{d}t, \quad \text{whence} \quad \sum_{n \leqslant x} f(n) \approx \int_0^x f(t)\mathrm{d}t.$$

To make the above heuristic rigorous, we examine how close $f(n)$ is to $\int_{n-1}^{n} f(t)\mathrm{d}t$. We begin by writing

$$f(n) - \int_{n-1}^{n} f(t)\mathrm{d}t = \int_{n-1}^{n} (f(n) - f(t))\mathrm{d}t.$$

For any constant c, we have $\mathrm{d}t = \mathrm{d}(t-c)$. Integrating by parts yields that

$$f(n) - \int_{n-1}^{n} f(t)\mathrm{d}t = -(n-1-c)(f(n)-f(n-1)) + \int_{n-1}^{n} (t-c)f'(t)\mathrm{d}t.$$

If we take $c = n-1$, the "side term" $(n-1-c)(f(n-1)-f(n))$ vanishes. Since we also have that $t-(n-1) = \{t\}$ for $t \in [n-1, n)$, we conclude that

$$f(n) = \int_{n-1}^{n} f(t)\mathrm{d}t + \int_{n-1}^{n} f'(t)\{t\}\mathrm{d}t.$$

Summing the above identity formula over $n = M+1, M+2, \ldots, N$, where $M < N$ are two integers, we arrive at the *Euler-Maclaurin summation formula*:

$$\sum_{M<n\leqslant N} f(n) = \int_{M}^{N} f(t)\mathrm{d}t + \int_{M}^{N} f'(t)\{t\}\mathrm{d}t. \tag{1.5}$$

The first integral is the expected main term and the second term will be smaller if f' is of smaller order of magnitude than f, which is a quantitative way of saying that f does not vary too rapidly.

We demonstrate the versatility of the Euler-Mclaurin summation formula with a few examples. When $f(t) = t^2$, we have

$$\left| \int_{0}^{N} f'(t)\{t\}\mathrm{d}t \right| = \left| \int_{0}^{N} 2t\{t\}\mathrm{d}t \right| \leqslant \int_{0}^{N} 2t\,\mathrm{d}t = N^2.$$

Consequently,

$$\sum_{n=1}^{N} n = \int_{0}^{N} t^2\mathrm{d}t + \int_{0}^{N} 2t\{t\}\mathrm{d}t = \frac{N^3}{3} + O(N^2).$$

This should be compared with the well-known exact formula $\sum_{n=1}^{N} n^2 = N(N+1)(2N+1)/6$. However, it would be rather hard to guess such an exact formula for the sum $\sum_{n=1}^{N} n^{100}$ (though, see Exercise 1.3). Nevertheless, adapting the above argument implies readily that

$$\sum_{n=1}^{N} n^{100} = \frac{N^{101}}{101} + O(N^{100}).$$

A good exercise on the Euler-Maclaurin summation formula is to check that

$$\sum_{n=1}^{N} \frac{1}{n} = \log N + O(1) \sim \log N \quad (N \to \infty), \tag{1.6}$$

which is an estimate on the rate of divergence of the harmonic series. A more precise formula will be proven in Theorem 1.11 below.

The Euler-Maclaurin summation formula is a special case of a general identity. To prove this generalization, we introduce a tool called *summation*

by parts or *partial summation*, which is a discrete analogue of integration by parts. As it will become clear throughout this book, partial summation is one of the main workhorses of analytic number theory. It allows us to pass from estimates on $A(x) = \sum_{n \leqslant x} a_n$ to estimates for general sums of the form $\sum_{y<n\leqslant z} a_n f(n)$ with f a continuously differentiable function. The case when $a_n = 1$ for all n, for which $A(x) = \lfloor x \rfloor = x + O(1)$, corresponds to the Euler-Maclaurin formula.

To make the passage from $A(x)$ to $\sum_{y<n\leqslant z} a_n f(n)$, we use the theory of Riemann-Stieltjes integration (see Appendix A): note that $A(x)$ is a step function that is continuous from the right and that has jumps of length a_n at each integer n. For any continuous f, Theorem A.1(f) implies that

$$\sum_{y<n\leqslant z} a_n f(n) = \int_y^z f(t)\mathrm{d}A(t), \tag{1.7}$$

where the right-hand side is a Riemann-Stieltjes integral. If we further assume that f is continuously differentiable and integrate by parts (see Theorem A.1(d)), we arrive at the formula

$$\sum_{y<n\leqslant z} a_n f(n) = A(t)f(t)\Big|_{t=y}^{z} - \int_y^z A(t)f'(t)\mathrm{d}t. \tag{1.8}$$

Remark 1.7. A more elementary way to prove (1.8) that avoids the use of Riemann-Stieltjes integrals is to use *Abel's summation formula*: if $(a_n)_{n=1}^{\infty}$ and $(b_n)_{n=1}^{\infty}$ are two sequences of complex numbers, then

$$\sum_{n=M+1}^{N} a_n b_n = A_n b_{n+1}\Big|_{n=M}^{N} - \sum_{n=M+1}^{N} A_n(b_{n+1} - b_n) \tag{1.9}$$

for all integers $M \geqslant N \geqslant 1$, where $A_n = A(n) = \sum_{j=1}^{n} a_j$. Indeed, this can be proven by noticing that

$$\sum_{n=M+1}^{N} a_n b_n = \sum_{n=M+1}^{N} (A_n - A_{n-1})b_n = \sum_{n=M+1}^{N} A_n b_n - \sum_{n=M}^{N-1} A_n b_{n+1}.$$

In the special case when $b_n = f(n)$, we have that

$$A_n(b_{n+1} - b_n) = \int_n^{n+1} A(x)f'(x)\mathrm{d}x$$

because $A(x) = A_n$ when $x \in [n, n+1)$. Together with (1.9), this leads us to (1.8). We leave the details as an exercise. □

Example 1.8. For reasons we will explain later, we often count primes with a logarithmic weight. To this end, we define *Chebyshev's theta function*

$$\theta(x) := \sum_{p\leqslant x} \log p.$$

We may use relation (1.8) to go back and forth between $\pi(x)$ and $\theta(x)$: we have

$$\theta(x) = \int_1^x (\log y)\mathrm{d}\pi(y) = \pi(x)\log x - \int_1^x \frac{\pi(y)}{y}\mathrm{d}y.$$

Similarly, using that $\pi(x) = \sum_{2-\varepsilon<p\leqslant x} 1$ for each $\varepsilon > 0$, we have

$$\pi(x) = \int_{2^-}^x \frac{1}{\log y}\mathrm{d}\theta(y) = \frac{\theta(x)}{\log x} + \int_2^x \frac{\theta(y)}{y\log^2 y}\mathrm{d}y.$$

Note that we replaced 2^- by 2 in the rightmost integral, which is justified by the fact $\int_{a^-}^b f = \int_a^b f$ for any function f that is Riemann-integrable on the interval $[a, b]$. □

Often, we have at our disposal an asymptotic formula of the form

$$A(x) = M(x) + R(x) \qquad \text{for all } x \geqslant x_0, \tag{1.10}$$

where $M(x)$ is a continuously differentiable main term that approximates $A(x)$, and $R(x)$ is the remainder term to this approximation. For instance, when $a_n = 1$ for all n, we have $A(x) = \lfloor x\rfloor = x - \{x\}$. Another important example is when a_n is the indicator function of the primes for which $A(x) = \pi(x)$. We then write

$$\pi(x) = \mathrm{li}(x) + R(x), \tag{1.11}$$

with the Prime Number Theorem being equivalent to the estimate $R(x) = o(x/\log x)$ as $x \to \infty$, and with the Riemann Hypothesis yielding the much stronger estimate $R(x) = O(\sqrt{x}\log x)$ (see (0.2)).

For $z \geqslant y \geqslant x_0$, relations (1.7) and (1.10) imply that

$$\begin{aligned}\sum_{y<n\leqslant z} a_n f(n) &= \int_y^z f(t)\mathrm{d}(M(t) + R(t)) \\ &= \int_y^z f(t)M'(t)\mathrm{d}t + \int_y^z f(t)\mathrm{d}R(t) \\ &= \int_y^z f(t)M'(t)\mathrm{d}t + R(t)f(t)\Big|_{t=y}^z - \int_y^z R(t)f'(t)\mathrm{d}t, \end{aligned} \tag{1.12}$$

where we successively applied parts (b), (e) and (d) of Theorem A.1.

Example 1.9. Write $\pi(x) = \mathrm{li}(x) + R(x)$ as in (1.11). Applying (1.12) with $a_n = 1_P(n)$, $f(n) = \log n$, $z = x$ and $y = 1$, we find that

$$\theta(x) = x - 1 + R(x)\log x - \int_1^x \frac{R(t)}{t}\mathrm{d}t.$$

If for large t the remainder $R(t)$ is much smaller than $\mathrm{li}(t) \sim t/\log t$, we see that a good approximation to $\theta(x)$ is given by x (see Exercise 1.7). □

Recall that $A(x) = \lfloor x \rfloor$ when $a_n = 1$ for all n. Writing $A(x) = x - \{x\}$, we find that (1.12) implies the following generalization of (1.5).

Theorem 1.10 (Euler-Maclaurin summation formula)**.** *If $f \in C^1([y,z])$, then*

$$\sum_{y<n\leqslant z} f(n) = \int_y^z f(t)\mathrm{d}t - \{t\}f(t)\Big|_{t=y}^{z} + \int_y^z \{t\}f'(t)\mathrm{d}t.$$

In particular, if $f \in C^1([1,+\infty))$, then for every $x \geqslant 1$

$$\sum_{n\leqslant x} f(n) = f(1) + \int_1^x f(t)\mathrm{d}t - \{x\}f(x) + \int_1^x \{t\}f'(t)\mathrm{d}t.$$

We give two important applications of the Euler-Maclaurin formula. We start with an estimate of the growth of the harmonic series that sharpens the estimate in (1.6). The symbol γ in its statement denotes the *Euler-Mascheroni constant* that is defined by

$$\gamma = 1 - \int_1^\infty \frac{\{t\}}{t^2}\mathrm{d}t = 0.57721\ldots. \tag{1.13}$$

Theorem 1.11. *For $x \geqslant 1$, we have*

$$\sum_{n\leqslant x} \frac{1}{n} = \log x + \gamma + O\Big(\frac{1}{x}\Big).$$

Proof. By Theorem 1.10, we have that

$$\sum_{n\leqslant x} \frac{1}{n} = \int_1^x \frac{\mathrm{d}t}{t} + 1 - \frac{\{x\}}{x} - \int_1^x \frac{\{t\}}{t^2}\mathrm{d}t.$$

Since $0 \leqslant \{t\}/t^2 \leqslant 1/t^2$, the integral $\int_1^\infty \{t\}t^{-2}\mathrm{d}t$ converges absolutely. In addition, $\int_1^x \mathrm{d}t/t = \log x$. We may thus write

$$\sum_{n\leqslant x} \frac{1}{n} = \log x + \left(1 - \int_1^\infty \frac{\{t\}}{t^2}\mathrm{d}t\right) - \frac{\{x\}}{x} + \int_x^\infty \frac{\{t\}}{t^2}\mathrm{d}t.$$

Moreover, we have the inequalities

$$0 \leqslant \frac{\{x\}}{x} \leqslant \frac{1}{x} \quad \text{and} \quad 0 \leqslant \int_x^\infty \frac{\{t\}}{t^2}\mathrm{d}t \leqslant \int_x^\infty \frac{\mathrm{d}t}{t^2} = \frac{1}{x},$$

which complete the proof of the theorem. □

A more involved application of Theorem 1.10 is given in Stirling's approximation for the factorial function.

Theorem 1.12 (Stirling's formula)**.** *For $n \in \mathbb{N}$, we have*

$$n! = \Big(\frac{n}{e}\Big)^n \sqrt{2\pi n}\,(1 + O(1/n)).$$

Proof. Taking logarithms and applying the Euler-Maclaurin formula, we find that

$$\log(n!) = \sum_{j=1}^{n} \log j = \int_1^n \log x \, \mathrm{d}x + \log 1 - \{n\} \log n + \int_1^n \frac{\{t\}}{t} \mathrm{d}t$$
$$= n \log n - n + 1 + \int_1^n \frac{\{t\}}{t} \mathrm{d}t,$$

since $n \in \mathbb{N}$ here and thus $\{n\} = 0$. Next, set

$$F(x) = \int_0^x (\{t\} - 1/2) \mathrm{d}t.$$

Since $\{t\} - 1/2$ is a 1-periodic function of mean 0 over a complete period, we find that F is also 1-periodic. In particular, $F(n) = 0$ for all $n \in \mathbb{N}$, and $F(x) = O(1)$ for all $x \geqslant 1$. Integration by parts implies that

$$\int_1^n \frac{\{t\}}{t} \mathrm{d}t = \frac{\log n}{2} + \int_1^n \frac{\{t\} - 1/2}{t} \mathrm{d}t = \frac{\log n}{2} + \frac{F(t)}{t}\Big|_{t=1}^{n} + \int_1^n \frac{F(t)}{t^2} \mathrm{d}t$$
$$= \frac{\log n}{2} + \int_1^n \frac{F(t)}{t^2} \mathrm{d}t.$$

(Justify why we can integrate by parts even though F is not differentiable everywhere.) The integral $\int_1^\infty F(t) t^{-2} \mathrm{d}t$ converges absolutely by the estimate $F(t) = O(1)$ and its tails satisfy the bound

$$\left| \int_n^\infty \frac{F(t)}{t^2} \mathrm{d}t \right| \leqslant \int_n^\infty \frac{|F(t)|}{t^2} \mathrm{d}t \ll \int_n^\infty \frac{\mathrm{d}t}{t^2} = \frac{1}{n}.$$

This proves that

$$\log(n!) = (n + 1/2) \log n - n + c + O(1/n), \quad \text{where} \quad c = 1 + \int_1^\infty \frac{F(t)}{t^2} \mathrm{d}t.$$

Since $e^{O(1/n)} = 1 + O(1/n)$ by Taylor's theorem, the proof will be complete provided that we can show that $e^c = \sqrt{2\pi}$. Establishing this identity requires different means outlined in Exercise 1.11 below. Alternatively, see Theorem 1.13. □

The saddle-point method

One of the most useful methods for obtaining asymptotic evaluations of integrals is the *saddle-point method*.[1] In its simplest form that we present here it is also called *Laplace's method* and it is used to evaluate asymptotically integrals of the form $\int_a^b e^f$, where $f : [a, b] \to \mathbb{R}$ is a function. In practice, f depends on some parameters and our goal is to estimate $\int_a^b e^f$ in terms of these parameters.

[1] Other names for it are "method of the steepest descent" and "stationary-phase method".

If f has a unique maximum in $[a,b]$, say at c, then we may expect that most of the mass of the integral $\int_a^b e^{f(x)}\mathrm{d}x$ comes from values of x around c. If $c \in (a,b)$ and f is a smooth function, then c is a *stationary point* of f, that is to say, $f'(c) = 0$. Moreover, if $f''(c)$ does not vanish, then it must be negative by the maximality of $f(c)$. Using quadratic approximation, we find that

$$f(x) \approx f(c) - \frac{|f''(c)|}{2}(x-c)^2,$$

so that we might expect that

$$\int_a^b e^{f(x)}\mathrm{d}x \approx \int_{x\approx c} e^{f(c)-|f''(c)|(x-c)^2/2}\mathrm{d}x. \tag{1.14}$$

The integrand on the right-hand side of the above formula decays fast when x moves away from c. Hence, it seems reasonable to expect that

$$\int_a^b e^{f(x)}\mathrm{d}x \approx \int_{-\infty}^{\infty} e^{f(c)-|f''(c)|(x-c)^2/2}\mathrm{d}x = e^{f(c)}\sqrt{2\pi/|f''(c)|},$$

where we used the identity $\int_{\mathbb{R}} e^{-u^2/2}\mathrm{d}u = \sqrt{2\pi}$.

There are various subtleties and technicalities that we left out of the above discussion. Rather than trying to prove an abstract and general theorem that establishes rigorously the above formula, we demonstrate how to handle all the necessary details in a concrete example.

Our goal is to study the asymptotic behavior of Euler's *Gamma function* that is defined for $\mathrm{Re}(s) > 0$ by the formula

$$\Gamma(s) = \int_0^{\infty} e^{-x}x^{s-1}\mathrm{d}x.$$

This function extends the usual factorial function. Indeed, noticing that $e^{-x} = (-e^{-x})'$ and integrating by parts, we deduce the *functional equation of the Gamma function*

$$\Gamma(s+1) = s\Gamma(s), \tag{1.15}$$

valid whenever $\mathrm{Re}(s) > 0$. Iterating this formula implies that $\Gamma(n+1) = n!$ for all $n \in \mathbb{Z}_{\geqslant 0}$. Moreover, (1.15) can be used to meromorphically continue Γ to the entire complex plane: applying it $n+1$ times, we deduce that

$$\Gamma(s) = \frac{\Gamma(s+n+1)}{s(s+1)\cdots(s+n)}, \tag{1.16}$$

which can be taken as the definition of Γ for $\mathrm{Re}(s) > -n-1$. It is clear

from this formula that the only singularities of Γ are simple poles at 0, -1, -2, ... with $\operatorname{res}_{s=-n}\Gamma(s) = (-1)^n/n!$.[2]

Let us now use the saddle-point method to estimate $\Gamma(s)$ when $s \geqslant 1$. We note that

$$\Gamma(s) = \frac{\Gamma(s+1)}{s} = \frac{1}{s}\int_0^\infty e^{f(x)}\mathrm{d}x, \quad \text{where} \quad f(x) = -x + s\log x. \tag{1.17}$$

We have $f'(x) = -1 + s/x$ and $f''(x) = -s/x^2$. In particular, the function f has a unique maximum in the positive reals at $x = s$. If we can show that the integral in (1.17) is dominated by values of x very close to s and carry out the argument leading to (1.14), we will deduce that

$$\Gamma(s) \sim (s/e)^s\sqrt{2\pi/s},$$

which is a generalization of Stirling's formula for the factorial function.

In order to establish the above formula rigorously, we begin by showing that we may truncate the range of integration in (1.17) to values of x that are very close to s. This is done in two stages.

First, we show we can discard the portion of the integral over $[2s, +\infty)$ at the cost of a small error term. Indeed, for each $x \geqslant 2s$, we have $f'(x) \leqslant -1/2$. Hence, the Mean Value Theorem implies that $f(x) \leqslant f(2s) - (x - 2s)/2$. Consequently,

$$\int_{2s}^\infty e^{f(x)}\mathrm{d}x \leqslant 2e^{f(2s)} = 2e^{f(s)+(\log 2-1)s},$$

and thus

$$\Gamma(s) = \frac{1}{s}\int_0^{2s} e^{f(x)}\mathrm{d}x + O(e^{f(s)-s/4}).$$

Next, we show we can discard the portion of the integral over $E := \{x \in (0, 2s] : |x - s| \geqslant s^{2/3}\}$. Indeed, for all $x \in E$, Taylor's theorem implies there is some $c \in (0, 2s]$ such that $f(x) = f(s) + (x - s)f'(s) + (x - s)^2 f''(c)/2$. We have $f'(c) = -s/c^2 \leqslant -1/(4s)$ and $(x - s)^2 \geqslant s^{4/3}$. Therefore,

$$\int_E e^{f(x)} \leqslant \int_0^{2s} e^{f(s)-s^{1/3}/8}\mathrm{d}x = 2se^{f(s)-s^{1/3}/8}.$$

In conclusion, we have the asymptotic formula

$$\Gamma(s) = \frac{1}{s}\int_{|x-s|\leqslant s^{2/3}} e^{f(x)}\mathrm{d}x + O(e^{f(s)-s^{1/3}/10}). \tag{1.18}$$

[2]Recall that a function f that is analytic in the punctured disk $\{\, z \in \mathbb{C} : 0 < |z - a| < r \,\}$ has a Laurent series expansion $f(z) = \sum_{n\in\mathbb{Z}} c_n(z - a)^n$ about a. Its residue at a is defined to be c_{-1} and is denoted by $\operatorname{res}_{z=a} f(z)$. If there is an integer $m \geqslant 1$ such that $c_{-m} \neq 0$ and $c_n = 0$ for $n < -m$, then we say that f has a pole of order m at a. The pole is called simple when $m = 1$. A simple way to check whether f has a simple pole at a is to compute $\lim_{z\to a}(z - a)f(z)$. If this limit exists and is non-zero, then f has a simple pole at $z = a$ of residue $c_{-1} = \lim_{z\to a}(z - a)f(z)$.

Consider now the portion of the integral with $|x-s| \leqslant s^{2/3}$. For such x, we have $f'''(x) = 2s/x^3 = O(1/s^2)$ and thus

$$f(x) = s\log(s/e) - \frac{(x-s)^2}{2s} + O\Big(\frac{|x-s|^3}{s^2}\Big)$$

by Taylor's theorem. Consequently,

$$\begin{aligned} e^{f(x)} &= (s/e)^s e^{-(x-s)^2/2s + O(|x-s|^3/s^2)} \\ &= (s/e)^s e^{-(x-s)^2/2s}(1 + O(|x-s|^3/s^2)), \end{aligned}$$

where we used the formula $e^\delta = 1 + O(\delta)$ for bounded values of δ, a consequence of the Mean Value Theorem. We make the change of variables $x = s + y\sqrt{s}$ to find that

$$\int_{|x-s|\leqslant s^{2/3}} e^{f(x)}\mathrm{d}x = (s/e)^s\sqrt{s}\int_{|y|\leqslant s^{1/6}} e^{-y^2/2}(1 + O(|y|^3/\sqrt{s}))\mathrm{d}y.$$

Since $\int_{\mathbb{R}} e^{-y^2/2}\mathrm{d}y = \sqrt{2\pi}$ and $\int_{\mathbb{R}} |y|^3 e^{-y^3/2}\mathrm{d}y < \infty$, we conclude that

$$\int_{|x-s|\leqslant s^{2/3}} e^{f(x)}\mathrm{d}x = (s/e)^s\sqrt{s}\Big(\sqrt{2\pi} - R + O(1/\sqrt{s})\Big), \tag{1.19}$$

where

$$R := \int_{|y|>s^{1/6}} e^{-y^2/2}\mathrm{d}y \leqslant \int_{|y|>s^{1/6}} e^{-|y|/2}\mathrm{d}y = 4e^{-s^{1/6}/2}.$$

Together with (1.18), the above estimates imply that

$$\Gamma(s) = (s/e)^s\sqrt{2\pi/s}\,\big(1 + O(1/\sqrt{s})\big) \qquad (s \geqslant 2), \tag{1.20}$$

thus generalizing Theorem 1.12, only with a weaker error term.

The above formula can be further extended to complex values of s and the error term can be improved.

Theorem 1.13 (Stirling's formula II)**.** *Fix $\delta > 0$. Uniformly for $s \in \mathbb{C}$ with $|s| \geqslant 1$ and $|\arg(s)| \leqslant \pi - \delta$, we have that*

$$\Gamma(s) = (s/e)^s\sqrt{2\pi/s}\,(1 + O(1/|s|)).$$

Proof. For general complex values of s, the function $f(x) = -x + s\log x$ does not have a stationary point on the semiline $\mathbb{R}_{\geqslant 0}$. One approach to proving the theorem is to employ Cauchy's residue theorem to write

$$\Gamma(s) = \frac{1}{s}\int_L e^{-z}z^s\mathrm{d}z,$$

where L is the semiline $\{\, z \in \mathbb{C} : z = \lambda s,\ \lambda \geqslant 0\,\}$ traversed from 0 to ∞. The new contour contains the stationary point $z = s$ and we could use the ideas leading to (1.20) to estimate $\Gamma(s)$. This is rather complicated in practice (and an excellent exercise). Instead, we use a trick.

We begin by noticing that

$$\Gamma(s) = \lim_{n\to\infty} \frac{\Gamma(s+n+1)}{s(s+1)\cdots(s+n)}. \tag{1.21}$$

We will use the method of proof of (1.20) to show that

$$\Gamma(s+n+1) \sim e^{-n} n^{s+n} \sqrt{2\pi n} \quad (n\to\infty). \tag{1.22}$$

For the purposes of proving (1.22), s is considered fixed. We then have

$$\begin{aligned}\Gamma(n+s+1) &= \int_{|x-n|\leqslant n^{2/3}} e^{-x} x^{s+n} \mathrm{d}x + O\Big(\int_{|x-n|\geqslant n^{2/3}} e^{-x} x^{\sigma+n} \mathrm{d}x\Big) \\ &= \int_{|x-n|\leqslant n^{2/3}} e^{-x} x^{s+n} \mathrm{d}x + o_{n\to\infty}(n^{\sigma}(n/e)^n \sqrt{n}),\end{aligned}$$

where we bounded the error term using a variant of (1.18) with $n+\sigma$ in place of s. For the main term, we note that $x^s \sim n^s$ when $|x-n| \leqslant n^{2/3}$. On the other hand, we may estimate the integral of $e^{-x}x^n$ over $x \in [n-n^{2/3}, n+n^{2/3}]$ using (1.19) (with n in place of s). This proves (1.22).

Now, combining (1.21) and (1.22), we arrive at the formula

$$\log\Gamma(s) = \lim_{n\to\infty} \bigg((s+n)\log n - n + \log\frac{\sqrt{2\pi n}}{s} - \sum_{j=1}^{n} \log(s+j) \bigg). \tag{1.23}$$

We employ the Euler-Maclaurin formula to estimate the sum over j:

$$\begin{aligned}\sum_{j=1}^{n} \log(s+j) &= \int_0^n \log(s+x)\mathrm{d}x + \int_0^n \frac{\{x\}}{s+x}\mathrm{d}x \\ &= (s+n)\log(s+n) - s\log s - n + \frac{\log(s+n)-\log s}{2} \\ &\quad + \int_0^n \frac{\{x\}-1/2}{s+x}\mathrm{d}x.\end{aligned}$$

If we set $F(x) = \int_0^x (\{x\}-1/2)\mathrm{d}x \ll 1$ and integrate by parts as in the proof of Theorem 1.12, we find that

$$\begin{aligned}\sum_{j=1}^{n} \log(s+j) &= (s+n)\log(s+n) - s\log s - n + \frac{\log(s+n)-\log s}{2} \\ &\quad + \int_0^n \frac{F(x)}{(s+x)^2}\mathrm{d}x.\end{aligned}$$

Inserting the above formula into (1.23) yields that

$$\log\Gamma(s) = s\log s - s + \frac{\log(2\pi/s)}{2} + \int_0^{\infty} \frac{F(x)}{(s+x)^2}\mathrm{d}x.$$

It remains to show that the integral on the right-hand side of the above equality is $\ll 1/|s|$. Indeed, if $s = \sigma + it$ with $\sigma \geqslant 0$, then $|x+s| \asymp |x| + |s|$, so that

$$\int_0^\infty \frac{F(x)}{(s+x)^2}\mathrm{d}x \asymp \int_0^\infty \frac{1}{(x+|s|)^2}\mathrm{d}x \ll \frac{1}{|s|}.$$

Finally, if $s = \sigma + it$ with $\sigma \leqslant 0$, then our assumption that $|\arg(s)| \leqslant \pi - \delta$ implies that $|\sigma| \ll_\delta |t|$ and thus $|s| \asymp |t| + |\sigma| \asymp_\delta |t|$. Hence

$$\int_0^\infty \frac{F(x)}{(s+x)^2}\mathrm{d}x \ll \int_0^\infty \frac{\mathrm{d}x}{(x-|\sigma|)^2+t^2} \leqslant \int_0^{2|s|} \frac{\mathrm{d}x}{t^2} + \int_{2|s|}^\infty \frac{\mathrm{d}x}{(x/2)^2} \ll_\delta \frac{1}{|s|},$$

which completes the proof. □

We conclude our discussion on the Gamma function by proving that it can be represented by an infinite product. (See Exercise 1.14 for the rigorous definition of convergence of infinite products.)

Theorem 1.14. *For all $s \in \mathbb{C}$, we have that*

$$\Gamma(s) = \frac{1}{s}\prod_{n=1}^\infty \frac{(1+1/n)^s}{1+s/n} = \frac{e^{-\gamma s}}{s}\prod_{n=1}^\infty \frac{e^{s/n}}{1+s/n}.$$

In particular, Γ does not have any zeroes.

Proof. By (1.21) and (1.22) with N in place of n, we have that

$$\Gamma(s) = \lim_{N\to\infty} \frac{N!N^s}{s(s+1)\cdots(s+N)} = \frac{1}{s}\lim_{N\to\infty}\prod_{n=1}^N \frac{n}{s+n}\left(\frac{n+1}{n}\right)^s,$$

and the first equality follows. Finally, the second equality follows by noticing that $\prod_{n=1}^N (1+1/n) = (N+1) \sim e^{-\gamma}\prod_{n=1}^N e^{1/n}$ by Theorem 1.11. □

Exercises

Exercise 1.1. Consider the following functions:

$$f_1(x) = x^{1/\log\log x}, \quad f_2(x) = e^{\sqrt{\log x}}, \quad f_3(x) = (\log x)^A, \quad f_4(x) = \sqrt{x},$$
$$f_5(x) = e^x, \quad f_6(x) = \frac{x}{(\log x)^A}, \quad f_7(x) = \frac{x}{e^{\sqrt{\log x}}}, \quad f_8(x) = \log\log x,$$

where A is a fixed positive real number. Order the functions in terms of their order of magnitude as $x \to \infty$, namely find a permutation $\sigma \in S_8$ such that $f_{\sigma(1)}(x) \ll f_{\sigma(2)}(x) \ll \cdots \ll f_{\sigma(8)}(x)$ when $x \to \infty$.

Exercise 1.2. Show the following asymptotic estimates:

(a) $\log(1+\delta) = \delta + O(\delta^2)$ for $\delta \in [-1/2, 1/2]$.

(b) $\sqrt{x+1} = \sqrt{x} + O(1/\sqrt{x})$ for $x \geqslant 1$.

(c) $e^\delta = 1 + O(\delta)$ for $|\delta| \leqslant 1$.

(d) If $p > 1$, then $\sum_{n>x} 1/n^p \ll_p x^{1-p}$ for all $x \geqslant 1$.

(e) Let $\rho \in (0,1)$ and consider a sequence $(a_n)_{n=1}^{\infty}$ such that $0 \leqslant a_{n+1} \leqslant \rho a_n$ for all $n \geqslant 1$. Then

$$\sum_{n \geqslant N} a_n \asymp_\rho a_N.$$

Exercise 1.3. Show that there is a polynomial P_k of degree $k+1$ and of leading coefficient $1/(k+1)$ such that

$$\sum_{n=1}^{N} n^k = P_k(N) \qquad \text{for all } N \in \mathbb{N}.$$

[*Hint:* Use Abel's summation formula (1.9) and induction on k.]

Exercise 1.4.

(a) Prove that

$$\sum_{n \leqslant x} \sqrt{n} = \frac{2}{3}x^{3/2} + O(\sqrt{x}) \qquad (x \geqslant 1).$$

(b) Prove that there is some constant c such that

$$\sum_{n \leqslant x} \frac{1}{\sqrt{n}} = 2\sqrt{x} + c + O\Big(\frac{1}{\sqrt{x}}\Big) \qquad (x \geqslant 1).$$

Exercise 1.5.

(a) If $f : \mathbb{R}_{\geqslant 1} \to \mathbb{R}_{\geqslant 0}$ is decreasing, then show that

$$\int_{\lfloor x \rfloor + 1}^{\infty} f(t)\mathrm{d}t \leqslant \sum_{n>x} f(n) \leqslant \int_{\lfloor x \rfloor}^{\infty} f(t)\mathrm{d}t \qquad (x \geqslant 1).$$

(b) Prove that

$$\frac{1}{\delta x^\delta} \leqslant \sum_{n>x} \frac{1}{n^{1+\delta}} \leqslant \frac{1}{\delta(x-1)^\delta} \qquad (\delta > 0,\ x \geqslant 2)$$

and

$$\log x \leqslant \sum_{n \leqslant x} \frac{1}{n} \leqslant 1 + \log x \qquad (x \geqslant 1).$$

Exercise 1.6. A number n is called square-full if $p^2 | n$ for all primes $p | n$.

(a) Show that n is square-full if and only if it can be written as $n = a^2 b^3$ for some integers a, b.

(b) Prove that $\#\{\, n \leqslant x : n \text{ is square-full} \,\} \asymp \sqrt{x}$ for $x \geqslant 1$.

Exercise 1.7. Define Chebyshev's psi function

$$\psi(x) := \sum_{p^k \leqslant x} \log p.$$

(a) Prove that $|\psi(x) - \theta(x)| \leqslant \sqrt{x}\log x$ for all $x \geqslant 1$.

(b) Prove that the asymptotic relations $\pi(x) \sim x/\log x$, $\theta(x) \sim x$ and $\psi(x) \sim x$ are equivalent as $x \to \infty$.

(c) Prove that the asymptotic estimates $\pi(x) = \mathrm{li}(x) + O(\sqrt{x}\log x)$, $\theta(x) = x + O(\sqrt{x}\log^2 x)$ and $\psi(x) = x + O(\sqrt{x}\log^2 x)$ are equivalent in the range $x \in \mathbb{R}_{\geqslant 2}$.

(d) Fix $c > 0$. Prove that the estimates $\pi(x) = \mathrm{li}(x) + O(xe^{-c\sqrt{\log x}})$, $\theta(x) = x + O(xe^{-c\sqrt{\log x}}\log x)$ and $\psi(x) = x + O(xe^{-c\sqrt{\log x}}\log x)$ are equivalent in the range $x \in \mathbb{R}_{\geqslant 2}$.

Exercise 1.8. Let $(a_n)_{n=1}^{\infty}$ be a sequence of complex numbers, and define

$$M(x) = \frac{1}{x}\sum_{n\leqslant x} a_n \quad \text{and} \quad L(x) = \frac{1}{\log x}\sum_{n\leqslant x}\frac{a_n}{n}$$

to be its *mean value* and its *logarithmic mean value*, respectively.

(a) If $\lim_{x\to\infty} M(x) = \ell$, then show that $\lim_{x\to\infty} L(x) = \ell$ as well.

(b*) Construct a sequence of $a_n \in [0,1]$ for which $L(x)$ tends to a limit as $x \to \infty$, whereas $M(x)$ does not.

Exercise 1.9. Here we study the asymptotic behavior of a Poisson distribution of parameter $\lambda \to \infty$. Throughout we fix $\varepsilon > 0$ and $c \geqslant 2$.

(a) Recall Exercise 1.2(e), and let $Q(u) = u\log u - u + 1$. Show that:

$$\text{(i)} \qquad \sum_{n\geqslant u\lambda} \frac{e^{-\lambda}\lambda^n}{n!} \asymp_{\varepsilon,c} \frac{e^{-Q(u)\lambda}}{\sqrt{\lambda}} \qquad \text{if } 1+\varepsilon \leqslant u \leqslant c\,;$$

$$\text{(ii)} \qquad \sum_{0\leqslant n\leqslant u\lambda} \frac{e^{-\lambda}\lambda^n}{n!} \asymp_{\varepsilon} \frac{e^{-Q(u)\lambda}}{\sqrt{\lambda}} \qquad \text{if } \varepsilon \leqslant u \leqslant 1-\varepsilon.$$

(b) For fixed $\alpha < \beta$ and $\lambda \to \infty$, show that

$$\sum_{\lambda+\alpha\sqrt{\lambda}<n\leqslant\lambda+\beta\sqrt{\lambda}} \frac{e^{-\lambda}\lambda^n}{n!} \sim \frac{1}{\sqrt{2\pi}}\int_{\alpha}^{\beta} e^{-t^2/2}\mathrm{d}t\,.$$

[*Hint:* First, prove the estimate $e^{-\lambda}\lambda^n/n! \sim (2\pi\lambda)^{-1/2}e^{-(n-\lambda)^2/2\lambda}$ when $n = \lambda + O(\sqrt{\lambda})$ and $\lambda \to \infty$.]

Exercise 1.10. This exercise generalizes the proof of Theorem 1.12.

We define the sequence of the *Bernoulli polynomials* $B_n(x)$ and of the *Bernoulli numbers* B_n as follows: we let $B_0(x) = B_0 = 1$, $B_1 = -1/2$ and $B_1(x) = x + B_1$. We then let $B_2(x) = B_2 + 2\int_0^x B_1(x)\mathrm{d}x$, where B_2 is such that $\int_0^1 B_2(x)\mathrm{d}x = 0$, that is to say, $B_2 = 1/6$ and $B_2(x) = x^2 - x + 1/6$. In general, assuming we have defined $B_n(x)$, we let $B_{n+1}(x) = B_{n+1} + (n+1)\int_0^x B_n(t)\mathrm{d}t$, where B_{n+1} is such that $\int_0^1 B_{n+1}(x)\mathrm{d}x = 0$.

(a) For $n \neq 1$, show that $B_n(1) = B_n(0) = B_n$. Conclude that the function $x \to B_n(\{x\})$ is 1-periodic and continuous. In addition, show that $\int_0^x B_n(\{t\})\mathrm{d}t = (B_{n+1}(\{x\}) - B_{n+1})/(n+1)$ for all $n \geqslant 1$ and $x \in \mathbb{R}$.

(b) Given integers $a < b$ and $k \geqslant 1$, and a smooth function f, prove that

$$\sum_{a<n\leqslant b} f(n) = \int_a^b f(x)\mathrm{d}x + \sum_{\ell=1}^{k} \frac{(-1)^\ell B_\ell}{\ell!}(f^{(\ell-1)}(b) - f^{(\ell-1)}(a)) + (-1)^{k+1}\int_a^b \frac{B_k(\{x\})f^{(k)}(x)}{k!}\mathrm{d}x.$$

(c) Let $m \in \mathbb{Z}$ and $k \in \mathbb{N}$. Show that

$$\int_0^1 B_k(x)e^{-2\pi imx}\mathrm{d}x = -1_{m\neq 0}\frac{k!}{(2\pi im)^k}.$$

[*Hint:* Use part (b) when $m \neq 0$.] Conclude that

$$B_k(\{x\}) = -\frac{k!}{(2\pi i)^k}\sum_{m\neq 0}\frac{e^{2\pi imx}}{m^k} \quad \text{for } k \geqslant 2.$$

(d) For $k \geqslant 1$, show that $B_{2k+1} = 0$ and

$$B_{2k} = \frac{(-1)^{k-1}(2k)!}{2^{2k-1}\pi^{2k}}\sum_{m\geqslant 1}\frac{1}{m^{2k}} = \frac{(-1)^{k-1}(2k)!\zeta(2k)}{2^{2k-1}\pi^{2k}}.$$

(e) Show that $B_n(x) = \sum_{k=0}^n \binom{n}{k}B_{n-k}x^k$ for $n \geqslant 0$, and deduce that $B_n(x+1) = B_n(x) + nx^{n-1}$ for $n \geqslant 1$.

(f) Prove the recursion formula $B_n = -(n+1)^{-1}\sum_{k=2}^{n+1}\binom{n+1}{k}B_{n+1-k}$ for $n \geqslant 1$ and deduce from it that $|B_n| \leqslant (4/5)^n n!$.

(g) Consider the generating series $F(z,x) = \sum_{n=0}^\infty B_n(x)z^n/n!$. Prove that $\partial F/\partial x = zF$, as well as that $F(z,1) - F(z,0) = z$. Deduce that $F(z,x) = e^{zx}z/(e^z-1)$.

(h) Show that $F(z,0) + z/2 = z/(e^z-1) + z/2$ is an even function and give a new proof that $B_{2n+1} = 0$ for $n \geqslant 1$.

(i) Noticing that $z/(e^z-1) = 1/(1+\sum_{n=1}^\infty z^n/(n+1)!)$, give an explicit formula for B_n.

Exercise 1.11.* (a) For each $n \in \mathbb{Z}_{\geqslant 0}$, let $I_n := \int_0^{\pi/2}(\cos x)^n\mathrm{d}x$. Show that

$$I_{2k} = \frac{\pi}{2}\cdot\frac{1\cdot 3\cdots(2k-1)}{2\cdot 4\cdots(2k)} \qquad \text{and} \qquad I_{2k+1} = \frac{2\cdot 4\cdots(2k)}{1\cdot 3\cdots(2k+1)}.$$

(b) Show that $I_{n+1} \sim I_n$ as $n \to \infty$. [*Hint*: Show that most of the mass of the integral defining I_n is concentrated around $x = 0$.]

(c) If c is the constant from the proof of Theorem 1.12, show that $e^c = \sqrt{2\pi}$.

(d) Use the saddle-point method to prove that $I_n \sim \sqrt{\pi/(2n)}$.

Exercise 1.12. Fix $\delta > 0$, and let $s \in \mathbb{C}$ with $|s| \geqslant 1$ and $|\arg(s)| \leqslant \pi - \delta$.

(a) If $s = \sigma + it$, prove that

$$|\Gamma(s)| \asymp_\delta |s|^{\sigma-1/2}e^{-\sigma-|t\arg(s)|}.$$

In particular, if $|\sigma| \leqslant C$ and $|t| \geqslant 1$, then $|\Gamma(s)| \asymp_C |t|^{\sigma-1/2}e^{-\pi|t|/2}$.

(b) Show that

$$(\Gamma'/\Gamma)(s) = \log s - 1/(2s) + O(1/|s|^2).$$

[*Hint:* If $f(s) = \log\big(\Gamma(s)(e/s)^s(s/2\pi)^{1/2}\big)$ and ε is small enough in terms of δ, then $f'(s) = (2\pi i)^{-1}\oint_{|z|=\varepsilon|s|} z^{-2}f(s+z)\mathrm{d}z \ll_{\varepsilon,\delta} 1/|s|^2$.]

Exercise 1.13. For all $s \in \mathbb{C}$, show the *reflection* and the *duplication formula* of the Gamma function:

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)} \quad \text{and} \quad \Gamma(2s) = \pi^{-1/2}2^{2s-1}\Gamma(s)\Gamma(s+1/2).$$

[*Hint*: Show that $\Gamma(s)\Gamma(1-s)\sin(\pi s)$ and $4^s\Gamma(s)\Gamma(s+1/2)/\Gamma(2s)$ are entire, 1-periodic and bounded functions.]

Exercise 1.14. Given a sequence $(a_n)_{n=1}^\infty \subseteq \mathbb{C}$, consider the infinite product $\prod_{n=1}^\infty(1+a_n)$ and the partial products $P_{M,N} = \prod_{n=M}^N(1+a_n)$. We then define the following notions:

- $\prod_{n=1}^\infty(1+a_n)$ *diverges to zero* if $\lim_{N\to\infty} P_{M,N} = 0$ for all $M \in \mathbb{Z}_{\geqslant 1}$.
- $\prod_{n=1}^\infty(1+a_n)$ *converges* (conditionally) if there is some $p \neq 0$ and some $M \in \mathbb{Z}_{\geqslant 1}$ such that $\lim_{N\to\infty} P_{M,N} = p$. We then define $\prod_{n=1}^\infty(1+a_n) = p\prod_{n=1}^{M-1}(1+a_n)$ and say that $\prod_{n=1}^\infty(1+a_n)$ converges to $p\prod_{n=1}^{M-1}(1+a_n)$.
- $\prod_{n=1}^\infty(1+a_n)$ *converges absolutely* if $\prod_{n=1}^\infty(1+|a_n|)$ converges.

(a) Check that the definition of conditional convergence of $\prod_{n=1}^\infty(1+a_n)$ does not depend on the choice of M. [*Hint:* Verify first that if $\prod_{n\geqslant 1}(1+a_n)$ converges, then $1+a_n \neq 0$ for all sufficiently large n.]

(b) Assume that $\prod_{n=1}^\infty(1+a_n)$ converges. Show that $\prod_{n=1}^\infty(1+a_n) = 0$ if and only if there is some n such that $1+a_n = 0$.

(c) Show that if $\prod_{n=1}^\infty(1+a_n)$ converges, then $\lim_{n\to\infty} a_n = 0$.

(d) Show that $\prod_{n=1}^\infty(1+a_n)$ converges if and only if for each $\varepsilon > 0$ there is an integer $N = N(\varepsilon)$ such that $|P_{N_1,N_2} - 1| < \varepsilon$ for $N_2 \geqslant N_1 \geqslant N$.
[*Hint:* Use Cauchy's criterion for the convergence of sequences.]

(e) Show that if $\prod_{n=1}^\infty(1+a_n)$ converges absolutely, then it also converges conditionally. [*Hint:* Expand $\prod_{n=N_1}^{N_2}(1+a_n)-1$ and compare it with $\prod_{n=N_1}^{N_2}(1+|a_n|)-1$.]

(f) Show that $\prod_{n=1}^\infty(1+a_n)$ converges absolutely if and only if so does the series $\sum_{n=1}^\infty a_n$. [*Hint:* In both cases $\lim_{n\to\infty} a_n = 0$, whence $|a_n|/2 \leqslant \log(1+|a_n|) \leqslant |a_n|$ for large n.]

(g) Show that $\prod_{n=1}^\infty(1+a_n)$ converges if and only if there is some $N \in \mathbb{Z}_{\geqslant 1}$ such that $|a_n| < 1$ for $n \geqslant N$ and the series $\sum_{n=N}^\infty \log(1+a_n)$ converges, where log denotes the principal branch of the logarithm (i.e., $\log x \in \mathbb{R}$ for $x > 0$), as follows:

1) Note that if either $\prod_{n=1}^\infty(1+a_n)$ or $\sum_{n=1}^\infty a_n$ converges, the sequence $(a_n)_{n=1}^\infty$ converges to 0. Conclude that in either case there is $N_0 \in \mathbb{Z}_{\geqslant 1}$ such that $\mathrm{Re}(1+a_n) > 0$ and $|a_n| < 1$ for $n \geqslant N_0$.
2) Set $S_{N_1,N_2} = \sum_{n=N_1}^{N_2}\log(1+a_n)$ for $N_2 \geqslant N_1 \geqslant N_0$, where N_0 is as above. Show that there are integers k_{N_1,N_2} such that $\log P_{N_1,N_2} = S_{N_1,N_2} + 2\pi i k_{N_1,N_2}$ for $N_2 \geqslant N_1 \geqslant N_0$.

3) Assume that $\lim_{n\to\infty} a_n = 0$. Show that there is some $N_0' \in \mathbb{Z}_{\geqslant N_0}$ such that $|S_{N_1,N_2} - S_{N_1,N_2+1}| < \pi$ and $|\log P_{N_1,N_2} - \log P_{N_1,N_2+1}| < \pi$ for $N_2 \geqslant N_1 \geqslant N_0'$. Use induction on N_2 to conclude that $k_{N_1,N_2} = 0$ for $N_2 \geqslant N_1 \geqslant N_0'$.

4) Prove the equivalence stated in part (g).

(h) Show that if $\sum_{n=1}^{\infty} |a_n|^2 < \infty$, then the product $\prod_{n=1}^{\infty}(1+a_n)$ converges if and only if the series $\sum_{n=1}^{\infty} a_n$ converges.

(i) Assume that $a_n \neq -1$ for all n. Is it possible that $\prod_{n=1}^{\infty}(1+a_n)$ diverges and $\sum_{n=1}^{\infty} a_n$ converges? Is it possible that $\prod_{n=1}^{\infty}(1+a_n)$ converges and $\sum_{n=1}^{\infty} a_n$ diverges?

Chapter 2

Combinatorial ways to count primes

Perhaps the oldest way of counting primes is the *sieve of Eratosthenes*. Named after the ancient Greek mathematician Eratosthenes of Cyrene, it is an algorithm that determines all primes up to a given threshold x. In its core lies the fact that any composite integer $n > 1$ has a prime divisor $p \leqslant \sqrt{n}$. The steps of the algorithm are:

(1) List all integers in $[2, x]$.

(2) Circle the number 2 and delete all proper multiples of it.

(3) Find the smallest $n \in [3, \sqrt{x}\,]$ that has not been deleted nor circled yet and circle it. If such an n does not exist, circle all integers that have not been deleted yet and terminate the algorithm.

(4) Delete all proper multiples of n and return to step (3).

It is clear that after the termination of the algorithm the circled integers will be exactly the primes $\leqslant x$. The algorithm of Eratosthenes is called a "sieve" because the only integers that "do not pass through it", that is to say, are not deleted at any stage of the algorithm, are the primes $\leqslant x$.

The idea of Eratosthenes was further developed by Legendre, who used it to write down a formula for $\pi(x)$. Indeed, an integer $n \in (\sqrt{x}, x]$ is prime if and only if it has no prime factors $\leqslant \sqrt{x}$. We thus arrive at the formula

$$\pi(x) = \#\{\, n \leqslant x : (n, P(\sqrt{x}\,)) = 1 \,\} + O(\sqrt{x}\,), \tag{2.1}$$

where we recall that

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

Consider the more general problem of estimating the number of integers $\leqslant x$ that are coprime to some integer m. Since $(n+m,m)=(n,m)$, the condition that n and m are coprime is m-periodic. In particular, every interval of length m contains the same number of integers n coprime to m. This number is given by Euler's *totient function*

$$\varphi(m) := \#\{\, 1 \leqslant n \leqslant m : (n,m)=1 \,\} = \#(\mathbb{Z}/m\mathbb{Z})^*,$$

where $(\mathbb{Z}/m\mathbb{Z})^*$ denotes the group of reduced residues mod m. We will give a formula for the number of integers $\leqslant x$ that are coprime to m in terms of $\varphi(m)$, but first we establish some fundamental properties of the totient function.

The Chinese Remainder Theorem implies the group isomorphism

$$(\mathbb{Z}/ab\mathbb{Z})^* \cong (\mathbb{Z}/a\mathbb{Z})^* \times (\mathbb{Z}/b\mathbb{Z})^* \qquad \text{whenever} \quad (a,b)=1.$$

We infer from this relation that $\varphi(ab)=\varphi(a)\varphi(b)$ whenever $(a,b)=1$. Any function $f:\mathbb{N}\to\mathbb{C}$ satisfying the functional equation

$$f(ab)=f(a)f(b) \qquad \text{whenever} \quad (a,b)=1 \tag{2.2}$$

and the condition $f(1)=1$ is called *multiplicative*. If (2.2) holds for all $a,b\in\mathbb{N}$ without any restrictions on their greatest common divisor, then f is called *completely multiplicative*. Thus we see that φ is multiplicative but not completely multiplicative (for example, $\varphi(4)=2$ but $\varphi(2)=1$). Iterating (2.2) with $f=\varphi$ implies that

$$\varphi(m)=\prod_{p^k\|m}\varphi(p^k).$$

Hence calculating $\varphi(m)$ is reduced to finding its value on prime powers. The latter is easier, since the condition $(n,p^k)=1$ simplifies to the condition $p\nmid n$. Consequently,

$$\varphi(p^k)=p^k-\#\{\, 1\leqslant n\leqslant p^k : p|n \,\}=p^k-p^{k-1}.$$

We thus deduce the formula

$$\frac{\varphi(m)}{m}=\prod_{p|m}\Bigl(1-\frac{1}{p}\Bigr).$$

Now, we go back to the problem of estimating the counting function of integers coprime to m. As we already discussed, periodicity implies that each interval of length m contains exactly $\varphi(m)$ of such integers. If N is the unique integer satisfying the inequalities $Nm\leqslant x<(N+1)m$, then

$$N\cdot\varphi(m)\leqslant\#\{\, n\leqslant x : (n,m)=1 \,\}\leqslant(N+1)\cdot\varphi(m).$$

Noticing that $N = \lfloor x/m \rfloor = x/m + O(1)$, we find that

$$\begin{aligned} \#\{\, n \leqslant x : (n,m) = 1 \,\} &= (x/m + O(1)) \cdot \varphi(m) \\ &= x \prod_{p \mid m} \Big(1 - \frac{1}{p}\Big) + O(\varphi(m)). \end{aligned} \tag{2.3}$$

The remainder term in the above estimate can be improved significantly. To do so, we reappraise the sieve of Eratosthenes-Legendre from a purely combinatorial point of view: we have

$$\#\{\, n \leqslant x : (n,m) = 1 \,\} = \# \bigcap_{p \mid m} \{\, n \leqslant x : p \nmid n \,\}. \tag{2.4}$$

We apply the inclusion-exclusion principle to rewrite the right-hand side as

$$\begin{aligned} \# \bigcap_{p \mid m} \{\, n \leqslant x : p \nmid n \,\} &= \#\{n \leqslant x\} - \sum_{p \mid m} \#\{\, n \leqslant x : p \mid n \,\} \\ &\qquad + \sum_{\substack{p < p' \\ p, p' \mid m}} \#\{\, n \leqslant x : pp' \mid n \,\} \mp \cdots \end{aligned} \tag{2.5}$$

$$= \lfloor x \rfloor - \sum_{p \mid m} \lfloor x/p \rfloor + \sum_{\substack{p < p' \\ p, p' \mid m}} \lfloor x/(pp') \rfloor \mp \cdots . \tag{2.6}$$

The above formula has $2^{\#\{p \mid m\}}$ summands—one for each choice of a subset of the distinct prime factors of m. The quantity $\#\{p \mid m\}$ will reoccur several times throughout the book, so we give it a name:

$$\omega(m) := \#\{p \mid m\}, \quad \text{as well as} \quad \Omega(m) := \sum_{p^k \parallel m} k. \tag{2.7}$$

Inserting the approximation $\lfloor y \rfloor = y + O(1)$ into (2.6) and noticing that

$$1 - \sum_{p \mid m} \frac{1}{p} + \sum_{\substack{p < p' \\ p, p' \mid m}} \frac{1}{pp'} \mp \cdots = \prod_{p \mid m} \Big(1 - \frac{1}{p}\Big)$$

yields:

Theorem 2.1. *For $x \geqslant 1$ and $m \in \mathbb{N}$, we have*

$$\#\{\, n \leqslant x : (n,m) = 1 \,\} = x \prod_{p \mid m} \Big(1 - \frac{1}{p}\Big) + O\big(2^{\omega(m)}\big).$$

Remark 2.2. The above theorem has a natural probabilistic interpretation: for n to be coprime to m, we must have that $p \nmid n$ for each $p \mid m$. The chances that a randomly chosen integer n is a multiple of p are about $1/p$: indeed, we have $\#\{\, n \leqslant x : p \mid n \,\} = \lfloor x/p \rfloor \sim x/p$ as $x \to \infty$, so we see that a $1/p$ proportion of integers are divisible by p. But then, the chances that

an integer is not divisible by p are $1 - 1/p$. Assuming that divisibility by different primes are independent events, we are led to expect that the chances that an integer is coprime to x are about $\prod_{p|m}(1 - 1/p)$, as proven in Theorem 2.1 when x and m are in appropriate ranges. We will return to this probabilistic heuristic in Chapter 15. □

In order to appreciate the strength of Theorem 2.1 in the context of the sieve of Eratosthenes, we need to understand the product $\prod_{p|m}(1 - 1/p)$ when $m = P(\sqrt{x})$. The following lemma establishes an upper bound that is sharp up to a multiplicative constant (cf. Theorem 3.4). The idea of its proof goes back to Euler and will play a fundamental role in counting primes throughout the book.

Lemma 2.3. *For each $x \geqslant 2$,*

$$\prod_{p \leqslant x} \Big(1 - \frac{1}{p}\Big) \leqslant \frac{1}{\log x}.$$

Proof. Instead of bounding the product from above, we consider its reciprocal

$$\prod_{p \leqslant x} \Big(1 - \frac{1}{p}\Big)^{-1} = \prod_{p \leqslant x} \Big(1 + \frac{1}{p} + \frac{1}{p^2} + \cdots\Big).$$

Expanding the rightmost product, we see that the summands are in one-to-one correspondence with products of the form $1/(p_1^{a_1} \cdots p_r^{a_r})$, where $p_1 < \cdots < p_r \leqslant x$ and $a_i \geqslant 1$ (the empty product with $r = 0$ is also permitted). By the Fundamental Theorem of Arithmetic, this means that the summands can be reindexed as $1/n$, where the variable n runs over all integers that only have prime factors $\leqslant x$, that is to say, the set of x-smooth integers. In particular, this includes all integers $n \leqslant x$, so that

$$\prod_{p \leqslant x} \Big(1 - \frac{1}{p}\Big)^{-1} \geqslant \sum_{n \leqslant x} \frac{1}{n} \geqslant \sum_{n \leqslant x} \int_n^{n+1} \frac{\mathrm{d}t}{t} \geqslant \log x,$$

as claimed. □

Combining relation (2.1), Theorem 2.1 and Lemma 2.3, it seems like we can prove that $\pi(x) \lesssim 2x/\log x$. However, this is wishful thinking because the error term in Theorem 2.1 becomes way too big when $m = P(\sqrt{x})$: we have $\omega(m) = \pi(\sqrt{x})$, which we expect to be of size $\asymp \sqrt{x}/\log x$. The underlying reason for the failure of Theorem 2.1 in estimating $\pi(x)$ is that relation (2.6) has an enormous number of terms. As we will see in Theorem 3.4(c), this is not a mere technicality: the function $x \prod_{p \leqslant \sqrt{x}}(1 - 1/p)$, which is the alleged main term in Theorem 2.1 when $m = P(\sqrt{x})$, is not asymptotic to $\pi(x)$ as $x \to \infty$.

Even though the above discussion puts a cap on our expectations, the sieve of Eratosthenes-Legendre can still be used to prove that primes are sparse. The fundamental observation, made by Legendre, is that since we are only after an upper bound for $\pi(x)$, we may use the simple inequality

$$\pi(x) \leqslant \#\{\, n \leqslant x : (n, P(y)) = 1 \,\} + y,$$

where y is a parameter at our disposal. This inequality follows by noticing that every prime $p > y$ is coprime to $P(y)$. We then use Theorem 2.1 to find that

$$\pi(x) \leqslant x \prod_{p \leqslant y} \Big(1 - \frac{1}{p}\Big) + O(y + 2^{\pi(y)}). \tag{2.8}$$

To bound the right-hand side, we apply Lemma 2.3. Taking $y = \log x$ yields

$$\pi(x) \ll \frac{x}{\log\log x}. \tag{2.9}$$

Despite the fact that the above estimate is rather weak compared to what we expect to be the truth, at least it demonstrates that approximately 100% of the integers are composite (see also Exercise 2.3). On the other hand, taking logarithms in Lemma 2.3 and using Taylor's expansion for the function $\log(1-\delta)$ when $|\delta| \leqslant 1/2$, we find that

$$\sum_{p \leqslant x} \frac{1}{p} \geqslant \log\log x - O(1)$$

for all $x \geqslant 3$, which shows that primes are not too sparse.

Chebyshev's estimate

In 1852, Chebyshev discovered a completely different way to count primes and vastly improve (2.9). His argument was simplified significantly by Erdős. The key observation is that the central binomial coefficient $\binom{2n}{n}$ is an integer that is divisible by all primes $p \in (n, 2n]$. Indeed,

$$\binom{2n}{n} = \frac{2n(2n-1)\cdots(n+1)}{n!},$$

so if $p \in (n, 2n]$, then p divides the numerator and is coprime to the denominator. Thus $p \mid \binom{2n}{n}$ for all $p \in (n, 2n]$, as claimed. But then the product of all such primes divides $\binom{2n}{n}$, and we deduce that

$$\prod_{n < p \leqslant 2n} p \leqslant \binom{2n}{n} \leqslant \sum_{j=0}^{2n} \binom{2n}{j} = 4^n.$$

The rightmost inequality is almost sharp, since Stirling's formula implies that $\binom{2n}{n} \asymp 4^n/\sqrt{n}$. Taking logarithms and recalling the definition of $\theta(x)$ in Example 1.8, we find that

$$\theta(2n) - \theta(n) = \sum_{n<p\leqslant 2n} \log p \leqslant n \log 4$$

for all $n \in \mathbb{N}$. Applying the above inequality with $n = 2^j$, $0 \leqslant j \leqslant k$, and summing telescopically implies that

$$\theta(2^k) \leqslant 2^{k+1} \log 4.$$

For each $x \geqslant 1$, there is $k \in \mathbb{N}$ such that $2^{k-1} \leqslant x < 2^k$, whence

$$\theta(x) \leqslant \theta(2^k) \leqslant 2^{k+1} \log 4 \leqslant 4x \log 4 \leqslant 6x.$$

We may pass from the above inequality to an upper bound for $\pi(x)$ using partial summation, as in Example 1.8: for all $x \geqslant 2$, we have

$$\pi(x) = \frac{\theta(x)}{\log x} + \int_2^x \frac{\theta(y)}{y \log^2 y} \mathrm{d}y \leqslant \frac{6x}{\log x} + \int_2^x \frac{6y}{\log^2 y} \mathrm{d}y \ll \frac{x}{\log x}.$$

An analogous lower bound can also be established by studying the prime factorization of $\binom{2n}{n}$. The details of the proof are outlined in Exercise 2.10. This leads us to:

Theorem 2.4 (Chebyshev's estimate)**.** *For $x \geqslant 2$, we have*

$$\pi(x) \asymp \frac{x}{\log x}.$$

Exercises

Exercise 2.1. Let f be an arithmetic function. Show the following:

(a) f is multiplicative if and only if $f(n) = \prod_{p^k \| n} f(p^k)$ for all $n \in \mathbb{N}$.

(b) f is completely multiplicative if and only if $f(n) = \prod_{p^k \| n} f(p)^k$ for all $n \in \mathbb{N}$.

Exercise 2.2. A function $f : \mathbb{N} \to \mathbb{C}$ is called *additive* if

$$f(mn) = f(m) + f(n) \qquad \text{whenever} \qquad (m, n) = 1.$$

Show that the functions ω and Ω, defined in (2.7), are additive.

Exercise 2.3. For $x \geqslant y \geqslant 3$, prove that

$$\#\{x < p \leqslant x + y\} \ll \frac{y}{\log \log y}.$$

Exercise 2.4 (The square-free sieve)**.**

(a) Modify the sieve of Eratosthenes-Legendre to prove that

$$\#\{\, n \leqslant x : n \text{ is square-free} \,\} = x \cdot \prod_p \Big(1 - \frac{1}{p^2}\Big) + O(\sqrt{x}) \qquad (x \geqslant 1).$$

(b) Prove that $\prod_p(1-1/p^2) = 6/\pi^2$. [*Hint:* Show $\prod_{p\leqslant y}(1-1/p^2)^{-1} = \sum_{n\in\mathcal{S}(y)} 1/n^2$, where $\mathcal{S}(y) = \{\, n \in \mathbb{N} : p|n \Rightarrow p \leqslant y \,\}$ is the set of y-smooth numbers, and use Exercise 1.10(d) with $k = 1$.]

Exercise 2.5. Let $f(n) = \#\{(n_1, n_2) \in \mathbb{N}^2 : [n_1, n_2] = n\}$, where $[n_1, n_2]$ is the least common multiple of n_1 and n_2. Show that f is multiplicative and evaluate it at prime powers.

Exercise 2.6. Set $f(n) = \varphi(n)/n$, and let $\{n_k\}_{k=1}^{\infty}$ be the sequence of values n at which f attains a "record low", that is to say, $n_1 = 1$ and, for $k \geqslant 2$, n_k is defined as the smallest integer $> n_{k-1}$ with $f(n_k) < f(n)$ for all $n < n_k$. (For example, $n_2 = 2$ and $n_3 = 6$.) Find a general formula for n_k and $f(n_k)$.

Exercise 2.7. Recall the definition of Chebyshev's psi function from Exercise 1.7. Show that $|\psi(x) - \theta(x)| \ll \sqrt{x}$ for $x \geqslant 1$.

Exercise 2.8. Let $p_1 < p_2 < \cdots$ denote the sequence of primes, and let $P_k = p_1 p_2 \cdots p_k$ denote the kth *primorial*. The validity of the Prime Number Theorem can be assumed in solving this exercise.

(a) Show that $p_k \sim k \log k$ and $\log P_k \sim k \log k$ as $k \to \infty$.

(b) Show that $\omega(n) \lesssim \log n / \log\log n$ as $n \to \infty$. [*Hint*: What can you say about $\omega(n)$ if $n \leqslant P_k$?]

(c) Show that

$$\frac{\varphi(n)}{n} \sim \prod_{\substack{p|n \\ p\leqslant \log n}} \Big(1 - \frac{1}{p}\Big) \geqslant \prod_{p\leqslant \log n} \Big(1 - \frac{1}{p}\Big) \qquad (n \to \infty).$$

Exercise 2.9. Let $\tau(n) = \#\{d|n\}$ be the divisor function and, more generally, $\tau_k(n) = \#\{(d_1, \dots, d_k) \in \mathbb{N}^k : d_1 \cdots d_k = n\}$.

(a) Show that τ_k is multiplicative.

(b) For each prime power p^a, show that

$$k \leqslant \tau_k(p^a) \leqslant \min\{k^a, (a+1)^{k-1}\}.$$

Conclude that $k^{\omega(n)} \leqslant \tau_k(n) \leqslant \min\{k^{\Omega(n)}, \tau(n)^{k-1}\}$.

(c) For each prime power p^a, show the exact formula

$$\tau_k(p^a) = \binom{a+k-1}{k-1}.$$

(d) Assuming the Prime Number Theorem, find a sequence of integers n such that $\tau_k(n) = k^{(1+o(1))\log n/\log\log n}$. [*Hint*: How can you create an integer with lots of divisors?]

(e) For $y \geqslant 1$, let $\Omega(n; y) = \sum_{p^a \| n,\, p>y} a$. Show that

$$\Omega(n; y) \leqslant \log n / \log y.$$

(f) Show that

$$\tau_k(n) \leqslant \prod_{\substack{p^a \| n \\ p\leqslant y}} (a+1)^{k-1} \prod_{\substack{p^a \| n \\ p>y}} k^a \leqslant (2\log n + 1)^{(k-1)y} \cdot k^{\log n/\log y}.$$

Choose y appropriately to conclude that

$$\tau_k(n) \leqslant n^{(\log k + o(1))/\log\log n} \qquad (n \to \infty).$$

Exercise 2.10. Prove the lower bound in Theorem 2.4 as follows:

(a) If $v_p(m)$ denotes the p-adic valuation of m, that is to say, the highest power of p that divides m, show that

$$v_p(n!) = \sum_{k \geqslant 1} \lfloor n/p^k \rfloor.$$

(b) Show that $\lfloor 2x \rfloor - 2\lfloor x \rfloor$ is a 1-periodic function taking only the values 0 and 1. Conclude that

$$\binom{2n}{n} \leqslant (2n)^{\pi(2n)}.$$

(c) Prove that $\pi(x) \gg x/\log x$ for $x \geqslant 2$.

Exercise 2.11 (Nair [147])**.** Let

$$I_n = \int_0^1 x^n(1-x)^n \mathrm{d}x \qquad \text{and} \qquad M_n = \operatorname{lcm}[n+1, n+2, \ldots, 2n+1].$$

(a) Prove that $I_n \cdot M_n$ is a non-negative integer.

(b) Prove that $I_n \leqslant 4^{-n}$.

(c) Prove that $M_n \leqslant (2n+1)^{\pi(2n+1)}$.

(d) Deduce a new proof of the lower bound $\pi(x) \gg x/\log x$ for $x \geqslant 2$.

Exercise 2.12* Find the average value of the greatest common divisor of a and b asymptotically, as a and b range over all integers up to x.

Chapter 3

The Dirichlet convolution

The combinatorics of the sieve of Eratosthenes are naturally encoded in the *Möbius function* that is denoted by μ and defined by

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n \text{ is square-free and has } k \text{ distinct prime factors,} \\ 0 & \text{otherwise.} \end{cases}$$

The Möbius function can be easily seen to be multiplicative. Its connection to the sieve of Eratosthenes is revealed by observing that, since a natural number n equals 1 if and only if it has no prime factors, the inclusion-exclusion principle implies that

$$1_{n=1} = 1 - \sum_{p} 1_{p|n} + \sum_{p<p'} 1_{pp'|n} \mp \cdots = \sum_{d|n} \mu(d). \tag{3.1}$$

This formula is known as the *Möbius inversion formula*. Applying it with (n, m) in place of n, and noticing that $d|(n, m)$ if and only if $d|n$ and $d|m$, leads us to (2.6). The Möbius inversion formula sits naturally inside a general framework that we develop in this chapter.

The ring of arithmetic functions

We say that f is an *arithmetic function* if it is of the form $f : \mathbb{N} \to \mathbb{C}$. We write $\mathcal{A}$ for the set of all arithmetic functions. Given $f, g \in \mathcal{A}$, we define their *Dirichlet convolution* $f * g$ to be the arithmetic function defined by

$$(f * g)(n) = \sum_{ab=n} f(a)g(b) = \sum_{d|n} f(d)g(n/d) = \sum_{d|n} f(n/d)g(d).$$

The triplet $(\mathcal{A}, +, *)$ is a commutative unitary ring whose unit is the function $\delta(n) := 1_{n=1}$. In this set-up, the Möbius inversion formula states that μ is the *Dirichlet inverse* of the constant function 1, that is to say, its inverse with respect to the operation of the Dirichlet convolution. In general, a function f possesses a Dirichlet inverse if and only if $f(1) \neq 0$. In particular, all multiplicative functions are invertible in this ring.

Note that the Dirichlet convolution preserves multiplicativity: if f and g are multiplicative, then so is $f * g$. It can also be shown that if f is multiplicative, then so is its Dirichlet inverse. In particular, the operation $*$ renders the set of multiplicative functions an abelian group.

Proving the above affirmations about the Dirichlet convolution is a good exercise.

Convolution identities

As we will see shortly, an important technique for estimating averages of various arithmetic functions f has as its starting point a decomposition of f as the Dirichlet convolution of two simpler arithmetic functions. With this in mind, we study here some important examples of such decompositions.

One of the most classical convolution identities concerns the *divisor function*, for which we have

$$\tau(n) = \#\{d|n\} = (1 * 1)(n).$$

This formula can be generalized to all higher-order divisor functions, which we defined and studied in Exercise 2.9, by noticing that

$$\tau_k(n) = \#\{(d_1, \dots, d_k) \in \mathbb{N}^k : d_1 \cdots d_k = n\} = (\underbrace{1 * \cdots * 1}_{k \text{ times}})(n).$$

A related identity allows us to rewrite the "sum-of-divisors function"

$$\sigma(n) := \sum_{d|n} d = (\mathrm{id} * 1)(n),$$

where "id" denotes here the identity function on $\mathbb{N}$, that is to say, $\mathrm{id}(n) = n$.

A less obvious example of a convolution identity is

$$\varphi = \mu * \mathrm{id}. \tag{3.2}$$

There are two ways to prove (3.2): either we observe that both sides are multiplicative and compare them at prime powers, or we use that

$$\varphi(n) = \sum_{1 \leqslant n \leqslant m} \sum_{d|n,\, d|m} \mu(d).$$

Interchanging the order of summation yields (3.2).

Finally, it is possible to write down a convolution identity for (a close relative of) the indicator function of primes that encapsulates the Fundamental Theorem of Arithmetic. We start by expressing n in its prime factors, say $n = \prod_{p^a \| n} p^a$, and then take logarithms. This yields the formula

$$\log n = \sum_{p^a \| n} a \log p = \sum_{p^k | n} \log p$$

because if $p^a \| n$, then $p^k | n$ for $k \in \{1, \dots, a\}$. We have thus proven that

$$\log = 1 * \Lambda, \tag{3.3}$$

where

$$\Lambda(n) := \begin{cases} \log p & \text{if } n = p^k \text{ for some prime } p \text{ and some integer } k \geqslant 1, \\ 0 & \text{otherwise} \end{cases}$$

is *von Mangoldt's function*, which is a very convenient weighted variant of the indicator function of the sequence of primes. As a matter of fact, due to the identity (3.3), it is often easier to obtain results about primes by working with the summatory function of Λ, i.e., Chebyshev's psi function (see Exercise 1.7), instead of $\pi(x)$. We may then pass to $\pi(x)$ using Exercise 2.7 and the discussion in Examples 1.8 and 1.9.

Remark 3.1. Guessing relations (3.2) and (3.3) is far from trivial. In the next chapter, we will see a more systematic method of obtaining convolution identities. Using it will explain (3.2) and (3.3) in a more intuitive way. □

Dirichlet's hyperbola method

When an arithmetic function f is the Dirichlet convolution of two simpler functions g and h, we can estimate its partial sums using what we already know about the partial sums of g and h. The starting point is the identity

$$\sum_{n \leqslant x} f(n) = \sum_{n \leqslant x} \sum_{ab = n} g(a)h(b) = \sum_{ab \leqslant x} g(a)h(b). \tag{3.4}$$

There are several ways to rearrange the right-hand side of (3.4). An obvious one is to fix a and sum over b. This leads us to the formula

$$\sum_{n \leqslant x} f(n) = \sum_{a \leqslant x} g(a) \sum_{b \leqslant x/a} h(b).$$

The above arrangement of the summation is particularly effective when g is either supported on small integers a, or when g has small partial sums. We illustrate the details by estimating the partial sums of the totient function.

Theorem 3.2. *For $x \geqslant 2$, we have*

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

Proof. In the identity $\varphi = \mu * \mathrm{id}$, we note that the functions φ and id are much bigger in modulus than μ. We thus rearrange the sum as

$$\sum_{n\leqslant x} \varphi(n) = \sum_{a\leqslant x} \mu(a) \sum_{b\leqslant x/a} b.$$

We have $\sum_{b\leqslant y} b = y^2/2 + O(y)$ by the Euler-Maclaurin summation formula (Theorem 1.10), whence

$$\sum_{n\leqslant x} \varphi(n) = \sum_{a\leqslant x} \mu(a)\Big(\frac{x^2}{2a^2} + O(x/a)\Big) = \frac{x^2}{2} \sum_{a\leqslant x} \frac{\mu(a)}{a^2} + O\Big(x \sum_{a\leqslant x} \frac{|\mu(a)|}{a}\Big),$$

where we used the triangle inequality to bound the error term. The sum over a in the main term equals $c + O(\sum_{a>x} 1/a^2) = c + O(1/x)$ with $c = \sum_{a\geqslant 1} \mu(a)/a^2$, whereas the sum over a in the error term is $\leqslant \sum_{a\leqslant x} 1/a \ll \log x$. To complete the proof, it remains to prove that $c = 6/\pi^2$. This identity is a special case of relation (4.12) that we will prove in the next chapter. See also Exercise 3.8. □

Let us now use the above ideas to estimate $\sum_{n\leqslant x} \tau(n)$: we have

$$\sum_{n\leqslant x} \tau(n) = \sum_{n\leqslant x} (1*1)(n) = \sum_{a\leqslant x} \sum_{b\leqslant x/a} 1.$$

The innermost sum equals $x/a + O(1)$, whence

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

by Theorem 1.11. This is a genuine asymptotic formula, but the error term is only slightly smaller than the main term and we would like to do better. Reexamining our argument, we see that the approximation $\sum_{b\leqslant x/a} 1 = x/a + O(1)$ is not very good for large values of a. Instead, for large a, it would have been much better to switch the roles of a and b, by fixing b and summing first over a instead. More formally, given parameters $A, B \geqslant 1$ with $AB = x$, we can rearrange the sum as follows:

$$\sum_{n\leqslant x} \tau(n) = \sum_{ab\leqslant x} 1 = \sum_{\substack{ab\leqslant x\\ a\leqslant A}} 1 + \sum_{\substack{ab\leqslant x\\ a>A}} 1 = \sum_{a\leqslant A} \sum_{b\leqslant x/a} 1 + \sum_{b\leqslant B} \sum_{A<a\leqslant x/b} 1.$$

We write the rightmost sum over a as $\sum_{a\leqslant x/b} 1 - \sum_{a\leqslant A} 1$ to find that

$$\sum_{n\leqslant x} \tau(n) = \sum_{a\leqslant A} \sum_{b\leqslant x/a} 1 + \sum_{b\leqslant B} \sum_{a\leqslant x/b} 1 - \Big(\sum_{a\leqslant A} 1\Big)\Big(\sum_{b\leqslant B} 1\Big). \tag{3.5}$$

The estimate $\sum_{n\leqslant y} 1 = y + O(1)$ implies that

$$\sum_{n\leqslant x} \tau(n) = \sum_{a\leqslant A}(x/a + O(1)) + \sum_{b\leqslant B}(x/b + O(1)) - (A + O(1))(B + O(1)).$$

Gathering all remainder terms, we rewrite the above formula as

$$\sum_{n\leqslant x} \tau(n) = \sum_{a\leqslant A} \frac{x}{a} + \sum_{b\leqslant x} \frac{x}{b} - AB + O(A + B).$$

Applying Theorem 1.11 twice and recalling that $AB = x$, we deduce that

$$\begin{aligned}\sum_{n\leqslant x} \tau(n) &= x\big(\log(AB) + 2\gamma + O(1/A + 1/B)\big) - AB + O(A + B)\\ &= x\log x + (2\gamma - 1)x + O(A + B).\end{aligned}$$

The optimal choice is $A = B = \sqrt{x}$, which yields Dirichlet's famous estimate:

Theorem 3.3. *For $x \geqslant 1$, we have*

$$\sum_{n\leqslant x} \tau(n) = x\log x + (2\gamma - 1)x + O(\sqrt{x}).$$

The method of proof of Theorem 3.3 is called *Dirichlet's hyperbola method*. Its name is justified by a geometric reappraisal of it. The sum $\sum_{ab\leqslant x} 1$ counts the number of lattice points $(a, b) \in \mathbb{N} \times \mathbb{N}$ below the hyperbola $ab = x$. The way we rearranged this sum corresponds to writing the range of (a, b) as $X \cup Y$, where $X = \{(a, b) \in \mathbb{N}^2 : a \leqslant A\}$ and $Y = \{(a, b) \in \mathbb{N}^2 : b \leqslant B\}$. We then use inclusion-exclusion to infer that

$$\sum_{n\leqslant x} \tau(n) = |X \cup Y| = |X| + |Y| - |X \cap Y|,$$

which is relation (3.5). Dirichlet's hyperbola method is a key tool in analytic number theory and we will encounter it several times in this book.

Mertens' three estimates

We conclude this chapter with an application of the above circle of ideas to the theory of primes due to Mertens. Using the convolution identity (3.3), he discovered in 1872, several years before the Prime Number Theorem was established, a way to estimate various sums over primes.

Theorem 3.4 (Mertens' three estimates)**.** *For $x \geqslant 2$ we have:*

(a) $$\sum_{p \leqslant x} \frac{\log p}{p} = \log x + O(1);$$

(b) $$\sum_{p \leqslant x} \frac{1}{p} = \log\log x + c + O(1/\log x), \quad \textit{where } c \textit{ is a constant};$$

(c) $$\prod_{p \leqslant x} \Big(1 - \frac{1}{p}\Big) = \frac{e^{-\gamma}}{\log x}\big(1 + O(1/\log x)\big).$$

Proof. First, we prove (a). On the one hand, the identity $\log = \Lambda * 1$ yields

$$\sum_{n \leqslant x} \log n = \sum_{a \leqslant x} \Lambda(a) \sum_{b \leqslant x/a} 1 = \sum_{a \leqslant x} \Lambda(a) \cdot (x/a + O(1)) = x \sum_{a \leqslant x} \frac{\Lambda(a)}{a} + O(x),$$

where the error term was bounded using Chebyshev's estimate (Theorem 2.4). Since $\sum_{a=p^k,\, k \geqslant 2} \Lambda(a)/a = O(1)$, we conclude that

$$\sum_{n \leqslant x} \log n = x \sum_{p \leqslant x} \frac{\log p}{p} + O(1).$$

On the other hand, we know that

$$\sum_{n \leqslant x} \log n = x \log x - x + O(\log x)$$

by partial summation. This completes the proof of Mertens' first estimate.

To prove (b), we use (a) and partial summation. More precisely, let

$$\sum_{p \leqslant x} \frac{\log p}{p} = \log x + R(x),$$

so that $R(x) = O(1)$. We then have

$$\sum_{p \leqslant x} \frac{1}{p} = \int_{2^-}^{x} \frac{1}{\log t} \mathrm{d} \sum_{p \leqslant t} \frac{\log p}{p} = \int_{2}^{x} \frac{1}{t \log t} \mathrm{d}t + \int_{2^-}^{x} \frac{1}{\log t} \mathrm{d}R(t).$$

The first integral on the right-hand side equals $\log\log x - \log\log 2$. In the second integral, we integrate by parts to find that

$$\sum_{p \leqslant x} \frac{1}{p} = \log\log x - \log\log 2 + \frac{R(x)}{\log x} - \frac{R(2^-)}{\log 2} + \int_{2}^{x} \frac{R(t)}{t \log^2 t} \mathrm{d}t.$$

Since $R(2^-) = -\log 2$ and the integral $\int_2^\infty R(t)/(t\log^2 t)\mathrm{d}t$ converges absolutely by the estimate $R(t) \ll 1$, we conclude that

$$\begin{aligned}\sum_{p\leqslant x}\frac{1}{p} &= \log\log x + c + \frac{R(x)}{\log x} - \int_x^\infty \frac{R(t)}{t\log^2 t}\mathrm{d}t \\ &= \log\log x + c + O(1/\log x)\end{aligned}$$

with $c = -\log\log 2 + 1 + \int_2^\infty R(t)/(t\log^2 t)\mathrm{d}t$.

Finally, we prove (c). Using Taylor's theorem, we can write $\log(1-x) = -x - f(x)$, where $f(x) = O(x^2)$ when $|x| \leqslant 1/2$. In particular,

$$\log\prod_{p\leqslant x}\Big(1-\frac{1}{p}\Big) = -\sum_{p\leqslant x} 1/p - \sum_{p\leqslant x} f(1/p).$$

The series $\sum_p f(1/p)$ converges and its tail satisfies the estimate

$$\sum_{p>x} f(1/p) = \sum_{p>x} O(1/p^2) = O(1/x).$$

Together with part (b), this yields the estimate

$$\log\prod_{p\leqslant x}\Big(1-\frac{1}{p}\Big) = -\log\log x - \kappa + O(1/\log x), \tag{3.6}$$

where $\kappa := c + \sum_p f(1/p)$. It remains to show that $\kappa = \gamma$. This is proven using information about the analytic behavior of the Riemann zeta function around the point 1 (see Exercise 5.4). □

Remark 3.5. Theorem 3.4(c) implies that the alleged main term in Theorem 2.1 when $m = P(\sqrt{x})$ is $\sim e^{-\gamma}x/\log\sqrt{x} = 2e^{-\gamma}x/\log x$ as $x\to\infty$. But $2e^{-\gamma} = 1.12291\ldots > 1$, so we cannot have that $\pi(x) \sim 2e^{-\gamma}x/\log x$, for this would contradict Theorem 3.4(a) by partial summation. □

Corollary 3.6. *As $n\to\infty$, we have*

$$\varphi(n) \gtrsim e^{-\gamma} n/\log\log n.$$

Proof. This follows by Exercise 2.8(c) and Theorem 3.4(c). □

Exercises

Exercise 3.1. Let f be an arithmetic function. Prove that it has a Dirichlet inverse g if and only if $f(1)\neq 0$, in which case g can be calculated recursively by the formula $g(n) = -f(1)^{-1}\sum_{d|n,\, d>1} f(d)g(n/d)$.

Exercise 3.2. Let $\mathcal{A}$ and $\mathcal{M}$ denote the set of arithmetic and multiplicative functions, respectively. Prove that $(\mathcal{A},+,*)$ is a unitary commutative ring and that $(\mathcal{M},*)$ is an abelian group.

Exercise 3.3. Determine which of the following arithmetic functions are multiplicative:

$$f_1(n) = n\,; \quad f_2(n) = \log n\,; \quad f_3(n) = \mu^2(n)\,; \quad f_4(n) = \sum_{d|n} d\,;$$

$$f_5(n) = \tau_3(n)\,; \quad f_6(n) = (-1)^{n-1}\,; \quad f_7(n) = 1_{(n,30)=1}\,; \quad f_8(n) = \varphi(n)/n\,.$$

Exercise 3.4. Let f be a multiplicative function and g its Dirichlet inverse.

(a) For a prime p, calculate $g(p)$ and $g(p^2)$ in terms of the values of f.

(b) If f is completely multiplicative, show that $g = \mu f$.

Exercise 3.5. Prove the following variants of the Möbius inversion formula:

(a) Show that $f = 1 * g$ if and only if $g = \mu * f$.

(b) If $f = 1 * g$, then show that $g(p^k) = f(p^k) - f(p^{k-1})$ for all primes p and all integers $k \geqslant 1$.

(c) Let $F, G : \mathbb{R}_{\geqslant 1} \to \mathbb{C}$. Prove that $F(x) = \sum_{n \leqslant x} G(x/n)$ for all $x \geqslant 1$ if and only if $G(x) = \sum_{n \leqslant x} \mu(n) F(x/n)$ for all $x \geqslant 1$.

Exercise 3.6. For $x \geqslant 1$, show that $\sum_{n \leqslant x} \mu(n) \lfloor x/n \rfloor = 1$, and deduce that

$$\Big| \sum_{n \leqslant x} \frac{\mu(n)}{n} \Big| \leqslant 1.$$

Exercise 3.7. For each $k \in \mathbb{N}$, we define the kth generalized von Mangoldt function to be $\Lambda^{(k)} = \mu * \log^k$. Prove the following statements:

(a) $\Lambda^{(k+1)} = \Lambda^{(k)} \log + \Lambda^{(k)} * \Lambda$.

(b) Λ_k is supported on integers n with $\leqslant k$ distinct prime factors.

(c) If $n = p_1 \cdots p_k$ for some distinct primes $p_1, \dots, p_k$, then

$$\Lambda^{(k)}(n) = k!(\log p_1) \cdots (\log p_k).$$

(d) $0 \leqslant \Lambda^{(k)}(n) \leqslant 2^{k-1} (\log n)^k$ for each $n \in \mathbb{N}$.

Exercise 3.8. Find f such that $\mu^2 = 1 * f$, and deduce that

$$\#\{\, n \leqslant x : n \text{ is square-free} \,\} = cx + O(\sqrt{x}) \qquad (x \geqslant 1),$$

where $c = \sum_{n=1}^{\infty} \mu(n)/n^2$. Then, use Exercise 2.4 to prove that $c = 6/\pi^2$.

Exercise 3.9. Show that there are constants $c_1, c_2 \in \mathbb{R}$ such that

$$\sum_{n \leqslant x} \omega(n) = x \log\log x + c_1 x + O(x/\log x) \qquad (x \geqslant 3)\,;$$

$$\sum_{n \leqslant x} \Omega(n) = x \log\log x + c_2 x + O(x/\log x) \qquad (x \geqslant 3)\,.$$

Conclude that, if $\xi : \mathbb{N} \to \mathbb{R}_{>0}$ is such that $\lim_{n\to\infty} \xi(n) = \infty$, then $\#\{\, n \leqslant x : \Omega(n) \geqslant \omega(n) + \xi(n) \,\} = o_{x\to\infty}(x)$.

Exercise 3.10. Prove that, for every fixed integer $k \geqslant 3$, there is a polynomial P_k of degree $k-1$ and of leading coefficient $1/(k-1)!$ such that

$$\sum_{n \leqslant x} \tau_k(n) = x \cdot P_k(\log x) + O_k(x^{1-1/k}) \qquad (x \geqslant 1).$$

Exercise 3.11. Let $\mathcal{S}$ denote the set of square-full integers (see Exercise 1.6 for their definition).

(a) Show that $1_{\mathcal{S}}(n) = \sum_{a^2b^3=n} \mu^2(b)$.

(b) Show that there are constants $c_1, c_2 \in \mathbb{R}$ such that

$$\#\mathcal{S} \cap [1, x] = c_1 x^{1/2} + c_2 x^{1/3} + O(x^{1/5}).$$

Exercise 3.12.

(a) Show that $2^\omega = 1 * \mu^2$ and deduce that there is a constant c such that

$$\sum_{n \leqslant x} 2^{\omega(n)} = 6\pi^{-2} x \log x + cx + O(x^{2/3}) \qquad (x \geqslant 2).$$

(b*) Prove that the error term in (a) can be improved to $O(\sqrt{x} \log x)$.

Exercise 3.13. Estimate the sums $\sum_{p \leqslant x} (\log p)^k/p$ and $\sum_{p > x} 1/p^2$.

Exercise 3.14.* Prove that the Prime Number Theorem is equivalent to the existence of a constant c such that

$$\sum_{p \leqslant x} \frac{\log p}{p} = \log x + c + o(1) \qquad (x \to \infty). \tag{3.7}$$

Exercise 3.15* (Landau [124])**.** Recall the notation $\delta(n) = 1_{n=1}$. This exercise proves that the Prime Number Theorem is equivalent to the estimate

$$\sum_{n \leqslant x} \mu(n) = o(x) \quad (x \to \infty). \tag{3.8}$$

(a) (i) Show that $-\mu \log = \mu * (\Lambda - 1) + \delta$.

(ii) Assuming the Prime Number Theorem, prove (3.8). [*Hint:* Prove first that $\sum_{n \leqslant x} \mu(n) \log n = o(x \log x)$ as $x \to \infty$.]

(b) Let $f(n) = \log n - \tau(n) + 2\gamma$.

(i) Show that $\Lambda - 1 = \mu * f - 2\gamma\delta$.

(ii) Show that $\sum_{n \leqslant x} f(n) \ll \sqrt{x}$ for $x \geqslant 1$.

(iii) Assuming (3.8), prove the Prime Number Theorem.

Exercise 3.16.*

(a) Prove there is a choice of constants c_1, c_2 for which $\sum_{n \leqslant x} (\log^2 n - 2\tau_3(n) - c_1\tau(n) - c_2) \ll x^{2/3}$ uniformly for $x \geqslant 1$.

(b) Recall the function $\Lambda_2 = \mu * \log^2$ from Exercise 3.7. Prove that it satisfies the estimate $\sum_{n \leqslant x} \Lambda_2(x) = 2x \log x + O(x)$ for $x \geqslant 1$, and conclude that

$$\psi(x) \log x + \sum_{p \leqslant x} \psi(x/p) \log p = 2x \log x + O(x) \quad (x \geqslant 1).$$

(c) Show that

$$\limsup_{x \to \infty} (\psi(x)/x) + \liminf_{x \to \infty} (\psi(x)/x) = 2.$$

In particular, if $\lim_{x \to \infty} \psi(x)/x$ exists, then it must equal 1.

Exercise 3.17.* Show that the Prime Number Theorem is equivalent to the relation $\sum_{n=1}^{\infty} \mu(n)/n = 0$. [*Hint:* Exercises 3.6 and 3.15.]

Chapter 4

Dirichlet series

The ubiquity and importance of convolution identities in analytic number theory calls for a systematic way of discovering them. We can obtain a very satisfactory answer to this problem by developing the theory of Dirichlet series: to each arithmetic function f, we associate the generating function

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

called the *Dirichlet series* of f. We do not concern ourselves with the convergence of this series for now, an issue that we will address in the end of the chapter. Rather, we treat $F(s)$ as a formal infinite series.

We write $\mathcal{D}$ for the set of formal Dirichlet series. If $G(s) = \sum_{n=1}^{\infty} g(n)/n^s$ is another element of $\mathcal{D}$, then we define

$$F(s) + G(s) := \sum_{n=1}^{\infty} \frac{f(n) + g(n)}{n^s} \quad \text{and} \quad F(s)G(s) := \sum_{n=1}^{\infty} \frac{(f * g)(n)}{n^s},$$

with the latter definition motivated by the formal calculation

$$\sum_{a=1}^{\infty} \frac{f(a)}{a^s} \sum_{b=1}^{\infty} \frac{g(b)}{b^s} = \sum_{a,b \geqslant 1} \sum \frac{f(a)g(b)}{(ab)^s} = \sum_{n=1}^{\infty} \frac{1}{n^s} \sum_{ab=n} f(a)g(b). \tag{4.1}$$

Evidently, the triplet $(\mathcal{D}, +, \cdot)$ forms a ring that is isomorphic to the ring of arithmetic functions $(\mathcal{A}, +, *)$. Hence, the study of the ring of arithmetic functions is equivalent to that of formal Dirichlet series.

In view of the above discussion, if we are given the functions f and g with Dirichlet series F and G, respectively, then the function h solving the identity $f = g * h$ is the unique arithmetic function whose Dirichlet series is the quotient F/G. Hence, we are faced with the problem of inverting G.

When g is multiplicative, this problem has a particularly elegant solution. The reason is that in this case G satisfies the formal identity

$$\sum_{n=1}^{\infty} \frac{g(n)}{n^s} = \prod_p \Big(1 + \frac{g(p)}{p^s} + \frac{g(p^2)}{p^{2s}} + \cdots\Big), \tag{4.2}$$

called the *Euler product* of G. Before we discuss the formal proof of (4.2), note that it allows us to invert G rather easily, since the factors of its Euler product are Taylor series in $z = p^{-s}$ (see Example 4.3 below). Moreover, we can estimate the coefficients of $1/G$ using Cauchy's residue theorem (see Exercise 4.10).

To see (4.2), we expand its right-hand side. We then obtain a formal sum of all products of the form

$$\frac{g(p_1^{a_1}) \cdots g(p_r^{a_r})}{(p_1^{a_1} \cdots p_r^{a_r})^s},$$

where $p_1, \ldots, p_r$ are distinct prime numbers, $a_1, \ldots, a_r \in \mathbb{Z}_{\geqslant 1}$ and $r \in \mathbb{Z}_{\geqslant 0}$. By multiplicativity, the numerator can be written as $g(p_1^{a_1} \cdots p_r^{a_r})$. The Fundamental Theorem of Arithmetic implies that the products $p_1^{a_1} \cdots p_r^{a_r}$ are in one-to-one correspondence with all natural numbers. This gives a formal proof of (4.2). A rigorous version will be given in the next section.

Example 4.1. The most important Dirichlet series is arguably the Riemann zeta function

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

We will study it in great detail in Part 2. For now, note that

$$\zeta(s) = \prod_p \Big(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\Big) = \prod_p \Big(1 - \frac{1}{p^s}\Big)^{-1}. \tag{4.3}$$

To compute the inverse of $\zeta(s)$, we use the sequence of formal identities

$$\frac{1}{\zeta(s)} = \prod_p \Big(1 - \frac{1}{p^s}\Big) = \sum_{n \geqslant 1} \frac{\mu(n)}{n^s}. \tag{4.4}$$

This formula can be considered as an analytic version of the Möbius inversion formula (3.1). □

Example 4.2. An alternative way of proving (3.2) is by noticing that

$$\begin{aligned} F(s) = \sum_{n \geqslant 1} \frac{\varphi(n)}{n^s} &= \prod_p \Big(1 + \frac{p-1}{p^s} + \frac{p(p-1)}{p^{2s}} + \frac{p^2(p-1)}{p^{3s}} + \cdots\Big) \\ &= \prod_p \frac{p^s - 1}{p^s - p} = \prod_p \frac{1 - 1/p^s}{1 - 1/p^{s-1}} = \frac{\zeta(s-1)}{\zeta(s)}. \end{aligned}$$

□

Example 4.3. Let f be a multiplicative function. We will calculate its Dirichlet inverse g using (4.2). We write $F(s)$ and $G(s)$ for the formal Dirichlet series of f and g, respectively. Since $(f*g)(n) = 1_{n=1}$, the factors in the Euler product of $F(s)G(s)$ must all be equal to 1. Namely, we have

$$\sum_{k\geqslant 0} \frac{f(p^k)}{p^{ks}} \sum_{\ell\geqslant 0} \frac{g(p^\ell)}{p^{\ell s}} = \sum_{m\geqslant 0} \frac{1}{p^{ms}} \sum_{k+\ell=m} f(p^k)g(p^\ell) = 1.$$

Thus

$$\sum_{\ell\geqslant 0} \frac{g(p^\ell)}{p^{\ell s}} = \frac{1}{1+\sum_{k=1}^\infty f(p^k)/p^{ks}} = 1 + \sum_{j\geqslant 1} (-1)^j \bigg(\sum_{k\geqslant 1} \frac{f(p^k)}{p^{ks}} \bigg)^j. \tag{4.5}$$

Expanding the jth power and regrouping the summands according to the power of p^s in the denominator, we find that

$$g(p^\ell) = \sum_{j=1}^{\ell} (-1)^j \sum_{\substack{k_1+\cdots+k_j=\ell \\ k_1,\dots,k_j\geqslant 1}} \cdots \sum f(p^{k_1}) \cdots f(p^{k_j}). \tag{4.6}$$

Since the above calculations are purely formal, it might be reassuring to verify them in a more direct way. Indeed, using induction on ℓ and the fact that $\sum_{k+\ell=m} f(p^k)g(p^\ell) = 0$ for $m \geqslant 1$ yields a proof of (4.6), even in the case when $F(s)$ and $G(s)$ converge nowhere. □

Example 4.4. Taking logarithms formally in (4.3), we find that

$$\log \zeta(s) = \sum_p \log \Big(1 - \frac{1}{p^s}\Big)^{-1} = \sum_p \sum_{k\geqslant 1} \frac{1}{kp^{ks}}.$$

By formal differentiation, we are led to the formal identity

$$-\frac{\zeta'}{\zeta}(s) = \sum_p \sum_{k\geqslant 1} \frac{\log p}{p^{ks}}. \tag{4.7}$$

The right-hand side is the Dirichlet series of von Mangoldt's function Λ we saw in the previous chapter. On the other hand, we have the formal identity

$$-\zeta'(s) = \sum_{n=1}^\infty \frac{\log n}{n^s}.$$

We thus guess that the left-hand side of (4.7) is the Dirichlet series of $\mu * \log$. This leads us to the convolution identity

$$\Lambda = \mu * \log. \tag{4.8}$$

Möbius inversion thus yields

$$\log = 1 * \Lambda. \tag{4.9}$$

This is relation (3.3), which we proved in a more combinatorial way before.

Notice that we also have the variant of (4.8)

$$\Lambda = -1 * \mu \log . \tag{4.10}$$

This formula can be proven using Möbius inversion:

$$\Lambda(n) = \sum_{d|n} \mu(d) \log(n/d) = -\sum_{d|n} \mu(d) \log d.$$

Alternatively, we can also see (4.10) by formally differentiating $1/\zeta$. □

In conclusion, we can use formal manipulations of Dirichlet series to guess various convolution identities, which we can then also verify in a more direct way.

Analytic properties of Dirichlet series

We conclude this chapter with a study of the convergence of general Dirichlet series $\sum_{n=1}^{\infty} f(n)/n^s$. Following Riemann's notation, we always write

$$s = \sigma + it.$$

Note that $|n^s| = n^\sigma$. Thus, if $f(n) = O(n^\theta)$, then $\sum_{n=1}^{\infty} f(n)/n^s$ converges absolutely for $\sigma > \theta + 1$. Moreover, for each fixed $\varepsilon > 0$, it converges uniformly for $\sigma \geqslant \theta + 1 + \varepsilon$. Hence, it defines a holomorphic function in the half-plane $\sigma > \theta + 1$.

This simple argument can be vastly generalized: Dirichlet series converge in half-planes of the form $\sigma > \alpha$ and they define holomorphic functions in their domain of convergence.

Theorem 4.5. *Let $F(s) = \sum_{n=1}^{\infty} f(n)/n^s$ be a Dirichlet series. If $F(s_0)$ converges for some complex number $s_0 = \sigma_0 + it_0$, then $F(s)$ converges uniformly in compact subsets of the half-plane $\sigma > \sigma_0$. In particular, it defines a holomorphic function there.*

Proof. The proof is easier when the convergence at s_0 is absolute, so we first give it in this case. Note that $|f(n)/n^s| = |f(n)|/n^\sigma \leqslant |f(n)|/n^{\sigma_0}$ if $\sigma \geqslant \sigma_0$. Weierstrass's criterion then implies that the series $\sum_{n=1}^{\infty} f(n)/n^s$ converges absolutely and uniformly for $\sigma \geqslant \sigma_0$.

We now give the proof of the general case that is more delicate. We set $g(n) = f(n)/n^{s_0}$ and note that it suffices to show that the Dirichlet series $G(s) = \sum_{n \geqslant 1} g(n) n^{-s} = F(s + s_0)$ defines an analytic function in the half-plane $\sigma > 0$. For all $M \geqslant N \geqslant 1$, partial summation implies that

$$\sum_{N < n \leqslant M} \frac{g(n)}{n^s} = \frac{1}{M^s} \sum_{N < n \leqslant M} g(n) + s \int_N^M \sum_{N < n \leqslant x} g(n) \frac{\mathrm{d}x}{x^{s+1}}. \tag{4.11}$$

Since $\sum_{n=1}^{\infty} g(n)$ converges, for each ε we can find some N_0 such that

$$\Big| \sum_{N<n\leqslant M} g(n) \Big| < \varepsilon \qquad (M \geqslant N \geqslant N_0).$$

As a consequence, we have

$$\Big| \sum_{N<n\leqslant M} \frac{g(n)}{n^s} \Big| \leqslant \varepsilon + |s| \int_N^M \frac{\varepsilon}{x^{\sigma+1}} \mathrm{d}x \leqslant \varepsilon + \frac{\varepsilon |s|}{\sigma} \qquad (M \geqslant N \geqslant N_0,\ \sigma > 0).$$

This clearly proves that, viewed as a series of functions, $\sum_{n=1}^{\infty} g(n)/n^s$ converges uniformly in compact subsets of the half-plane $\sigma > 0$. Indeed, if K is such a compact set, then there are numbers $\delta > 0$ and $B \geqslant 1$ such that $\sigma \geqslant \delta$ and $|s| \leqslant B$ for all $s \in K$. The analyticity of G follows readily. □

The above theorem naturally leads us to attach to a Dirichlet series $F(s) = \sum_{n=1}^{\infty} f(n)/n^s$ the quantity

$$\sigma_c = \sigma_c(F) := \inf\{\, \sigma \in \mathbb{R} : \exists t \in \mathbb{R} \text{ such that } F(\sigma + it) \text{ converges} \,\},$$

called the *abscissa of convergence* of F. Theorem 4.5 implies that F defines a holomorphic function in the half-plane $\sigma > \sigma_c$. We further define the *abscissa of absolute convergence* of F by

$$\sigma_a = \sigma_a(F) := \inf\{\, \sigma \in \mathbb{R} : F(\sigma) \text{ converges absolutely} \,\}.$$

For example, if $F = \zeta$, then $\sigma_c = \sigma_a = 1$. The properties of σ_c and σ_a are studied in the exercises.

A lot of the formal calculations we saw earlier can be rigorously justified when the involved Dirichlet series converge absolutely. For example, this is true for relation (4.1). In particular, we may rigorously prove that

$$\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} = \frac{1}{\zeta(s)}, \tag{4.12}$$

where $\mathrm{Re}(s) > 1$. Taking $s = 2$ yields a more direct proof of the identity $\sum_{n=1}^{\infty} \mu(n)/n^2 = 6/\pi^2$ that we saw in Exercise 3.8.

Similarly, the Euler product representation of Dirichlet series of multiplicative functions can be rigorously proven in their domain of absolute convergence. Firstly, let us consider the case of the Riemann zeta function. If $\mathrm{Re}(s) > 1$, then the absolute convergence of the series $\sum_{n\geqslant 1} n^{-s}$ allows us to sum its terms in any order. In particular, if we let $\mathcal{N}(y,k) = \{\, n \in \mathbb{N} : p^\nu \| n \ \Rightarrow\ p \leqslant y \text{ and } \nu \leqslant k \,\}$, then

$$\zeta(s) = \lim_{y\to\infty} \lim_{k\to\infty} \sum_{n\in\mathcal{N}(y,k)} n^{-s} = \lim_{y\to\infty} \lim_{k\to\infty} \prod_{p\leqslant y} (1 + p^{-s} + p^{-2s} + \cdots + p^{-ks}).$$

This establishes relation (4.3) when $\mathrm{Re}(s) > 1$.

The above argument can be easily generalized. We leave the details as an exercise. (It is highly recommended to first solve Exercise 1.14.)

Theorem 4.6. *Let f be a multiplicative function and $s \in \mathbb{C}$. The series $\sum_{n=1}^{\infty} f(n)/n^s$ converges absolutely if and only if so does the double series $\sum_p \sum_{k=1}^{\infty} f(p^k)/p^{ks}$. When they both converge absolutely, we have*

$$\sum_{n=1}^{\infty} \frac{f(n)}{n^s} = \prod_p \Big(1 + \frac{f(p)}{p^s} + \frac{f(p^2)}{p^{2s}} + \cdots\Big).$$

Remark 4.7. (a) It is important to emphasize here that the assumption of absolute convergence is crucial to represent a Dirichlet series of a multiplicative function as an Euler product. For example, the function $f(n) = (-1)^{n-1}$ is multiplicative. Its Dirichlet series $F(s)$ converges absolutely for $\sigma > 1$ and conditionally for $\sigma > 0$ by (4.11) with $g = f$, since $\sum_{n \leqslant x} f(n) = O(1)$ for all $x \geqslant 1$. However, its Euler product $(1 - 1/2^s - 1/2^{2s} - \cdots) \prod_{p>2}(1 + 1/p^s + 1/p^{2s} + \cdots)$ diverges to ∞ for $s \in (0, 1]$, because $\sum_{p>2} 1/p = \infty$ by Theorem 3.4(b).

(b) Knowing that $F(s)$ can be written as an absolutely convergent Euler product at some point s makes it very easy to check whether $F(s)$ vanishes: we simply need to check whether one of the factors vanishes (see Exercise 1.14(b)). For example, since $\zeta(s) = \prod_p (1 - 1/p^s)^{-1}$ for $\sigma > 1$, we have that $\zeta(s) \neq 0$ for $\sigma > 1$. As we will see in the next chapter, the location of the zeroes of ζ is intimately related to the distribution of prime numbers. □

Exercises

Exercise 4.1. (a) Find f and g such that $\sigma = \varphi * f$ and $\varphi/\mathrm{id} = 1 * g$.

(b) Use (4.5) and (4.6) to calculate the Dirichlet inverses of μ^2, 2^Ω and φ.

Exercise 4.2. If f and g are Dirichlet inverses of each other, then find a non-recursive formula for the values of g in terms of the values of f.

Exercise 4.3. Let $F(s) = \sum_{n \geqslant 1} f(n)/n^s$ be a Dirichlet series, and let σ_c and σ_a be its abscissas of convergence and of absolute convergence, respectively.

(a) Prove that $\sigma_c \leqslant \sigma_a \leqslant \sigma_c + 1$.

(b) Prove that $\sigma_c < +\infty$ if and only if there is $\theta \in \mathbb{R}$ such that $f(n) = O(n^\theta)$ for all $n \in \mathbb{N}$.

Exercise 4.4. Compute the Dirichlet series associated to the functions $f_1, \ldots, f_8$ from Exercise 3.3; your answer could be given in terms of ζ. Then, determine their abscissas of convergence and of absolute convergence.

Exercise 4.5. Show that there is a constant $\theta \in (0,1)$ and a polynomial P of degree 3 and of leading coefficient $1/\pi^2$ such that

$$\sum_{n\leq x} \tau^2(n) = xP(\log x) + O(x^\theta).$$

[*Hint:* Write $\tau^2 = \tau_4 * f$ and use Exercise 3.10.]

Exercise 4.6. Let f be an arithmetic function, and let $F(s)$ be its formal Dirichlet series. Define $f'(n) := -f(n)\log n$, and let $F'(s)$ be its Dirichlet series. Prove that $(f*g)' = f'*g + f*g'$ and $(FG)' = F'G + FG'$.

Exercise 4.7. Let f be a multiplicative function with formal Dirichlet series F, and define Λ_f via the convolution identity

$$f\log = \Lambda_f * f.$$

(a) Prove that the formal Dirichlet series of Λ_f is $-F'/F$.

(b) Prove that Λ_f is supported on prime powers, and that $\Lambda_f(p) = f(p)\log p$ for all primes p. [*Hint:* If $F = \prod_p E_p$ is the Euler product of F, we have the formal identity $F'/F = \sum_p E_p'/E_p$.]

(c) Calculate Λ_f when f is completely multiplicative.

(d) Calculate Λ_f when $F(s) = \prod_p (1-1/p^s)^{-f(p)}$.

Exercise 4.8. Let $F(s) = \sum_{n=1}^\infty f(n)/n^s$ be a Dirichlet series with abscissa of convergence $\sigma_c < +\infty$. Prove that the abscissa of convergence for the series of derivatives $-\sum_{n=1}^\infty f(n)(\log n)/n^s$ is also σ_c. Deduce that $F'(s) = -\sum_{n=1}^\infty f(n)(\log n)/n^s$ when $\mathrm{Re}(s) > \sigma_c$.

Exercise 4.9. If $F(s) = \sum_{n\geqslant 1} f(n)/n^s$ and $G(s) = \sum_{n\geqslant 1} g(n)/n^s$ converge and are equal in the half-plane $\mathrm{Re}(s) > \alpha$, then prove that $f = g$.

Exercise 4.10. Let f be a multiplicative function, and let g be its Dirichlet inverse. Fix a prime p and assume that there is some $M > 0$ such that $|f(p^k)| \leqslant M^k$ for all $k \in \mathbb{Z}_{\geqslant 1}$.

(a) Show that the power series $\sum_{k\geqslant 0} f(p^k)z^k$ converges absolutely for $|z| < 1/M$ and does not vanish for $|z| < 1/(2M)$.

(b) If $0 < r < 1/(2M)$, then show that

$$g(p^k) = \frac{1}{2\pi i}\oint_{|z|=r} \frac{1}{1+f(p)z+f(p^2)z^2+\cdots}\cdot\frac{\mathrm{d}z}{z^{k+1}}.$$

(c) Let $\varepsilon > 0$. Prove that $g(p^k) \ll_{\varepsilon,M} (2M+\varepsilon)^k$ for all $k \in \mathbb{Z}_{\geqslant 1}$.

(d) When $f(n) = (-1)^{n-1}$, compute $g(p^k)$ for all primes p. What do you observe when you compare $g(p^k)$ with the estimate of part (c)?

Exercise 4.11*. Let $F(s) = \sum_{n=1}^\infty f(n)/n^s$ and $G(s) = \sum_{n=1}^\infty g(n)/n^s$ be two Dirichlet series with $F(s)G(s) = 1$. If F has abscissa of convergence $< +\infty$, is it true that G also has abscissa of convergence $< +\infty$?

Part 2

Methods of complex and harmonic analysis

Chapter 5

An explicit formula for counting primes

So far, we have seen various ways of counting primes using combinatorial devices. We now introduce a different approach that transforms the problem of estimating $\pi(x)$ into a problem in complex analysis. The key idea is to package the primes all together and form an appropriate generating function.

Given an arithmetic function f, the most common generating function attached to f is arguably its power series

$$A(z) = \sum_{n \geqslant 1} f(n) z^n.$$

This series converges to a holomorphic function in a disk $|z| < R$. Moreover, $f(n)$ can be recovered from $A(z)$ via Cauchy's residue formula that implies

$$f(n) = \frac{1}{2\pi i} \oint_{|z|=r} \frac{A(z)}{z^{n+1}} \mathrm{d}z \qquad (n \in \mathbb{N},\ 0 < r < R). \tag{5.1}$$

We apply (5.1) when $f = 1_P$, the indicator function of the sequence of primes. The associated power series is

$$Q(z) := \sum_{p} z^p.$$

Summing (5.1) for $n = 0, 1, \ldots, N$ when $f = 1_P$ yields the inversion formula

$$\sum_{p \leqslant N} 1 = \sum_{0 \leqslant n \leqslant N} \frac{1}{2\pi i} \oint_{|z|=r} \frac{Q(z)}{z^{n+1}} \mathrm{d}z = \frac{1}{2\pi i} \oint_{|z|=r} \frac{Q(z)(1 - z^{N+1})}{z^{N+1}(1 - z)} \mathrm{d}z \tag{5.2}$$

for any $r \in (0, 1)$. Hence, a good understanding of the analytic behavior of $Q(z)$ can lead us to precise estimates for the counting function of the primes.

The above strategy arrives quickly at a dead end because it is not clear how to control the function $Q(z)$ without already knowing a lot about primes. As a matter of fact, the same objection can be raised for any generating function associated to the sequence of primes: how is it possible to determine its asymptotic behavior without already having a good grasp of the distribution of primes?

To break the vicious cycle, we analyze $Q(z)$ more closely. This function is naturally tied to the additive structure of the sequence of prime numbers. For example, note that

$$Q(z)^k = \sum_{p_1,\dots,p_k} z^{p_1+\cdots+p_k} = \sum_{n\geqslant 0} g_k(n) z^n,$$

where $g_k(n)$ is the number of ways to write n as the sum of k primes. However, primes are multiplicative objects, so it is more natural to study them from a multiplicative point of view. To this end, we observe that the logarithmic function is a group isomorphism from $(\mathbb{R}_{>0}, \times)$ to $(\mathbb{R}, +)$. We are thus naturally led to consider the generating function

$$\sum_p z^{\log p}.$$

This is no longer a power series because the exponents are not integers.

Note that $z^{\log p} = p^{\log z}$. Working with the complex logarithm causes technical difficulties. For this reason, we make the change of variables $s = -\log z$, so that our generating function becomes the Dirichlet series

$$\mathscr{P}(s) := \sum_p \frac{1}{p^s}.$$

In view of Mertens' second estimate (Theorem 3.4), this Dirichlet series has abscissa of convergence 1. In particular, Theorem 4.5 tells us that it defines a holomorphic function in the half-plane $\mathrm{Re}(s) > 1$.

Let us now consider the kth power of $\mathscr{P}$: we have

$$\mathscr{P}(s)^k = \sum_{p_1,\dots,p_k} \frac{1}{(p_1\cdots p_k)^s} = \sum_{n\geqslant 1} \frac{r_k(n)}{n^s},$$

where $r_k(n)$ is the number of ways to write n as the product of k primes. In particular, r_k is supported on integers with $\leqslant k$ prime factors. In comparison, before we had no control over the support of g_k. We thus see right away that $\mathscr{P}(s)$ has better properties than $Q(z)$.

Taking the above argument one step further, Euler proved that $\mathscr{P}(s)$ can be written in terms of the Riemann zeta function $\zeta(s) = \sum_{n=1}^{\infty} 1/n^s$, which is for $\mathbb{N}$ what $\mathscr{P}(s)$ is for the sequence of primes. The key is Euler's

product formula

$$\zeta(s) = \prod_p \Big(1 - \frac{1}{p^s}\Big)^{-1} \qquad (\mathrm{Re}(s) > 1)$$

that we proved in the previous chapter. Taking logarithms, we infer that

$$\log \zeta(s) = \sum_p \sum_{m \geqslant 1} \frac{1}{m p^{ms}} = \sum_{m \geqslant 1} \frac{\mathscr{P}(ms)}{m} \tag{5.3}$$

which provides the link between $\mathscr{P}$ and ζ. The above formula is the starting point of analytic number theory, as it relates the function $\mathscr{P}$, for which we knew nothing about, to the function ζ. The latter is significantly simpler because it is defined as a summation over all integers, a very regular set. It thus seems plausible that we can obtain good estimates for $\mathscr{P}$ via this link.

As in the case of the function $Q(z)$ and the inversion formula (5.2), we want to find a passage from $\mathscr{P}(s)$ to $\pi(x) = \sum_{p \leqslant x} 1$. We start by writing

$$\mathscr{P}(s) = \int_1^\infty x^{-s} \mathrm{d}\pi(x) = s \int_0^\infty \pi(x) x^{-s-1} \mathrm{d}x. \tag{5.4}$$

Hence, we see that the function $\mathscr{P}(-s)/(-s)$ is the *Mellin transform* of the function $\pi(x)$. (A brief introduction to the necessary theory of the Mellin transform is given in the last section of Appendix B.) Mellin inversion allows us to go from (5.4) to the formula

$$\sum_{p<x} 1 + \frac{1_{x \text{ is prime}}}{2} = \frac{1}{2\pi i} \int_{(\alpha)} \mathscr{P}(s) \frac{x^s}{s} \mathrm{d}s, \tag{5.5}$$

where $\int_{(\alpha)} f(s)\mathrm{d}s$ denotes the *principal value* of $\int_{\mathrm{Re}(s)=\alpha} f(s)$, namely

$$\int_{(\alpha)} f(s)\mathrm{d}s = \lim_{T \to \infty} \int_{\substack{\mathrm{Re}(s)=\alpha \\ |\,\mathrm{Im}(s)| \leqslant T}} f(s)\mathrm{d}s. \tag{5.6}$$

Indeed, to see (5.5), we apply Theorem B.4 (whose hypotheses are met here with $\alpha_1 = -\infty$ and $\alpha_2 = -1$) and then make the change of variables $s \to -s$.

Jumping into the void

The inversion formula (5.5) expresses $\pi(x)$ in terms of the Riemann zeta function. However, it is not that useful as it stands for the estimation of $\pi(x)$. Indeed, we expect that there are about $x/\log x$ primes $\leqslant x$. On the other hand, we have $|x^s| = x^\alpha$ on the right side of (5.5). Since $\alpha > 1$, the size of x^s is bigger than the expected main term. This means that if we are to extract an asymptotic estimate for $\pi(x)$ from (5.5), we must understand the integrand in a way that is precise enough to establish significant cancellation among the different parts of the range of integration. Obtaining such sharp estimates on $\mathscr{P}$ without already controlling $\pi(x)$ seems impossible.

It thus seems that we have again reached an impasse. Riemann though had a brilliant idea to circumvent it. He realized that $\zeta(s)$ can be extended in a canonical way to values of s outside its domain of convergence using the theory of analytic and meromorphic continuation of complex analysis.[1] We do not need to delve too deeply into this theory; as we will see shortly, the special structure of ζ allows us to meromorphically continue it[2] to $\mathbb{C}$ relatively easily. The extension we obtain has only one singularity: a simple pole of residue 1 at $s = 1$. Such an extension must be unique by the identity principle. Thus ζ really is well-defined over $\mathbb{C}$. Using this fact and Cauchy's residue theorem, we can then replace the line of integration in (5.5) by a new contour that reaches to the left of the vertical line $\mathrm{Re}(s) = 1$, where x^s becomes of smaller magnitude than x. Hence, we can hope to obtain bounds for this new integral that are of genuinely smaller order than $x/\log x$. The main term to the approximation of $\pi(x)$ will arise from the singularities in the region encircled by the old and the new contour of integration. The end result of this calculation will be a formula for $\pi(x)$ in terms of the singularities of $\mathscr{P}$.

We devote the rest of this chapter to making the above discussion more precise and to laying Riemann's idea on rigorous mathematical grounds.

The meromorphic continuation of ζ

Perhaps the simplest way of meromorphically continuing ζ is to use the Euler-Maclaurin formula. Indeed, when $\mathrm{Re}(s) > 1$, $\zeta(s)$ is defined as the sum of the smooth function $1/n^s$ over $n \geqslant 1$, so Theorem 1.10 implies that

$$\zeta(s) = \frac{s}{s-1} - s\int_1^\infty \frac{\{y\}}{y^{s+1}}\mathrm{d}y. \tag{5.7}$$

The integral on the right side converges absolutely for $\mathrm{Re}(s) > 0$ because $\{y\}$ is bounded. Thus, the right side of (5.7) supplies a meromorphic continuation of ζ to the half-plane $\mathrm{Re}(s) > 0$. The only singularity of ζ in this half-plane is a simple pole at $s = 1$ of residue 1 (a reflection of the divergence of the harmonic series $\sum_{n=1}^\infty 1/n$).

More generally, Exercise 1.10(b) implies that

$$\zeta(s) = \frac{s}{s-1} + \sum_{\ell=1}^{k} \frac{B_\ell}{\ell!}\prod_{j=0}^{\ell-2}(s+j) - \frac{\prod_{j=0}^{k-1}(s+j)}{k!}\int_1^\infty \frac{B_k(\{x\})}{x^{s+k}}\mathrm{d}x \tag{5.8}$$

[1] In fact, this theory was partly pioneered by Riemann himself.

[2] The YouTube channel 3Blue1Brown has an excellent video about the meromorphic continuation of ζ that is called "Visualizing the Riemann hypothesis and analytic continuation". The video is located at the web address `https://www.youtube.com/watch?v=sD0NjbwqlYw`.

for $\mathrm{Re}(s) > 1$. Since the right side is meromorphic for $\mathrm{Re}(s) > -k+1$ with only a simple pole at $s = 1$ of residue 1, so is ζ. Letting $k \to \infty$ establishes the alleged meromorphic continuation of ζ to the entire complex plane.

Let us now examine what the above discussion tells us about the analytic character of $\mathscr{P}$. We start from relation (5.3). Since $\sum_{m\geqslant 2} \mathscr{P}(ms)/m = \sum_{m\geqslant 2,\, p} 1/(mp^{ms}) = O(1)$ for $\mathrm{Re}(s) \geqslant 1$, we find that $\mathscr{P}(s) = \log\zeta(s) + O(1)$ for $\mathrm{Re}(s) > 1$. In particular, $\mathscr{P}(s) \sim -\log(s-1)$ as $s \to 1$, that is to say, $\mathscr{P}(s)$ has a logarithmic singularity at $s = 1$. This type of singularity prohibits us from extending $\mathscr{P}$ to an analytic function around $s = 1$. In particular, we cannot apply Cauchy's residue theorem to an integral of the form $\int_C (\mathscr{P}(s)x^s/s)\mathrm{d}s$, where C is a closed contour going around 1. For this reason, extracting the main term for $\pi(x)$ from (5.5) is a bit hard (though certainly possible as Riemann himself explained in his 1859 manuscript).

The above obstacle is merely of a technical nature. To overcome it, recall that the asymptotic behavior of $\pi(x)$ can be extracted from that of Chebyshev's theta and psi functions

$$\theta(x) = \sum_{p\leqslant x} \log p \qquad \text{and} \qquad \psi(x) = \sum_{n\leqslant x} \Lambda(n).$$

Indeed, we saw in Examples 1.8 and 1.9 how to go back and forth between $\pi(x)$ and $\theta(x)$. In addition, Chebyshev's functions are very close to each other in virtue of Exercise 2.7 which implies that

$$|\theta(x) - \psi(x)| \ll \sqrt{x} \qquad (x \geqslant 2).$$

Therefore, instead of estimating $\pi(x)$, we may work with $\psi(x)$. We need an analogue of formula (5.5) for this function.

In general, a straightforward adaptation of the proof of (5.5) implies the following generalization: if f is an arithmetic function whose Dirichlet series F converges absolutely in the half-plane $\mathrm{Re}(s) > 1$, then

$$\sum_{n<x} f(n) + \frac{1_{x\in\mathbb{N}} f(x)}{2} = \frac{1}{2\pi i}\int_{(\alpha)} F(s)\frac{x^s}{s}\mathrm{d}s \qquad (x > 1,\ \alpha > 1). \tag{5.9}$$

This general identity is called the *Perron inversion formula*.

We apply (5.9) with $f = \Lambda$ whose summatory function is Chebyshev's psi function. The associated Dirichlet series is $-\zeta'/\zeta$. Since ζ is meromorphic over $\mathbb{C}$, so is $-\zeta'/\zeta$. They both have a simple pole of residue 1 at $s = 1$. Moreover, if z is a zero of ζ multiplicity m, then ζ'/ζ has a simple pole of residue m at $s = z$. Indeed, we may write $\zeta(s) = (s-z)^m g(s)$ with g analytic and non-zero in a neighborhood of z. Hence, $(\zeta'/\zeta)(s) = m/(s-z) + (g'/g)(s)$ and g'/g is analytic around z. This implies that

$$\mathrm{res}_{s=z}(\zeta'/\zeta)(s) = m. \tag{5.10}$$

As we will see in the next chapter, the zeroes of ζ fall under two categories: the *trivial zeroes*, which are located at $-2, -4, -6, \ldots$, and the *non-trivial zeroes*, which are located in the strip $0 \leqslant \mathrm{Re}(s) \leqslant 1$. We denote a generic non-trivial zero by[3] $\rho = \beta + i\gamma$.

Remarkably, there is an *explicit formula* for $\psi(x)$ in terms of the non-trivial zeroes of the Riemann zeta function.[4]

Theorem 5.1. *For all $x, T \geqslant 2$, we have*

$$\psi(x) = x - \sum_{|\gamma| \leqslant T} \frac{x^\rho}{\rho} + O\bigg(\frac{x \log^2(xT)}{T} + \log x\bigg), \tag{5.11}$$

where the sum runs over the non-trivial zeroes of ζ with each zero repeated as many times as its multiplicity.

Before we explain why Theorem 5.1 is true, let us momentarily pause and make a few comments about it. This astonishing result reveals that primes, an elementary arithmetic object, have a "dual" complex-analytic object associated to them: the zeroes of ζ. These two objects of seemingly unrelated nature are interconnected in a fundamental way: the main term on the right-hand side of (5.11) approximates $\psi(x)$ better and better as $T \to \infty$, similarly to the Fourier expansion of a periodic function. Hence, the zeroes of ζ encode in principle everything we need to know about the distribution of primes (and vice versa). We may think of the zeroes as "frequencies" with which the counting function of prime numbers resonates. For this reason, they are of fundamental importance in mathematics.

Theorem 5.1 will play a key role in the proof of the Prime Number Theorem. Indeed, to establish the asymptotic formula $\psi(x) \sim x$, it suffices to bound $\sum_{|\gamma| \leqslant T} x^\rho/\rho$ and prove that it is of negligible size compared to x. Since $|x^\rho| = x^\beta$, this essentially reduces the Prime Number Theorem to showing that β is a bit less than 1 for all zeroes of ζ.

Cauchy's residue theorem and the explicit formula

Let us now give a rough sketch of the proof of Theorem 5.1. The complete details will be given in Chapter 8, after having developed the necessary tools.

We present the argument in a more general context. Recall the Perron inversion formula (5.9), valid for any arithmetic function f whose Dirichlet series F converges absolutely to the right of the line $\mathrm{Re}(s) = 1$. Similarly to

[3] The letter γ here is not to be confused with Euler-Mascheroni's constant defined by (1.13). This ambiguous notation is customary in the literature.

[4] The contribution of the trivial zeroes has been absorbed into the error term. There is an even more precise version of the explicit formula that takes into account trivial zeroes (see Exercise 8.2(a) and [31, Chapter 17]). The version stated in Theorem 5.1 is sufficient for most applications.

ζ and ζ'/ζ, the Dirichlet series F of many interesting arithmetic functions can be meromorphically continued to a half-plane $\mathrm{Re}(s) > \alpha_0$ with $\alpha_0 < 1$. In this case, the integral on the right-hand side of (5.9) can be studied using complex analysis as we explain below.

Fix $\alpha' \in (\alpha_0, 1) \setminus \{0\}$ such that $F(s)$ has no poles when $\mathrm{Re}(s) = \alpha'$. Such an α' always exists because F has at most countably many singularities in any given open region. Moreover, let $T = T(x)$ be large enough so that

$$\sum_{n \leqslant x} f(n) = \frac{1}{2\pi i} \int_{\substack{\mathrm{Re}(s)=\alpha \\ |\mathrm{Im}(s)| \leqslant T}} F(s) \frac{x^s}{s} \mathrm{d}s + E, \tag{5.12}$$

with $E = o(\sum_{n \leqslant x} |f(n)|)$. The existence of such a T is guaranteed by (5.6). Furthermore, similarly to α', the parameter T can be chosen in such a way that F has no singularities on the lines $\mathrm{Im}(s) = \pm T$.

Let C_1 denote the contour of integration in (5.12), that is to say, the line segment from $\alpha - iT$ to $\alpha + iT$. We write symbolically $C_1 = [\alpha - iT, \alpha + iT]$. We deform C_1 to a new contour of integration consisting of the line segments $C_2 = [\alpha - iT, \alpha' - iT]$, $C_3 = [\alpha' - iT, \alpha' + iT]$ and $C_4 = [\alpha' + iT, \alpha + iT]$ (see Figure 5.1). We denote this new contour by $C_2 + C_3 + C_4$.[5] We claim that

$$\frac{1}{2\pi i} \int_{C_1} F(s) \frac{x^s}{s} \mathrm{d}s = \sum_{2 \leqslant j \leqslant 4} \frac{1}{2\pi i} \int_{C_j} F(s) \frac{x^s}{s} \mathrm{d}s + \sum_w \mathrm{res}_{s=w} \frac{F(s)x^s}{s}, \tag{5.13}$$

where the rightmost sum runs over all singularities of $F(s)/s$ in $\Omega := \{ s \in \mathbb{C} : \alpha' < \sigma < \alpha, |t| < T \}$. Indeed, the integrand $F(s)x^s/s$ is meromorphic in Ω and analytic in an open neighborhood of the boundary $\partial\Omega$. Since $\partial\Omega = C_1 - C_2 - C_3 - C_4$ when traversed counterclockwise, Cauchy's residue theorem implies that

$$\frac{1}{2\pi i} \oint_{\partial\Omega} F(s) \frac{x^s}{s} \mathrm{d}s = \sum_w \mathrm{res}_{s=w} \frac{F(s)x^s}{s}.$$

This proves our claim that (5.13) holds.

Combining (5.12) and (5.13), we infer that

$$\sum_{n \leqslant x} f(n) = \sum_w \mathrm{res}_{s=w} \frac{F(s)x^s}{s} + E + R,$$

where

$$R = \sum_{2 \leqslant j \leqslant 4} \frac{1}{2\pi i} \int_{C_j} F(s) \frac{x^s}{s} \mathrm{d}s.$$

We think of R as an error term because $|x^s/s| \leqslant x^\sigma/|t|$, so that the integrand $F(s)x^s/s$ is small on $C_2 \cup C_4$ because $|t| = T$ is large, and it is also small on

[5] In general, if C, C' are two contours with a given orientation, then $C + C'$ denotes the contour that first traces C and then C' in their respective orientation. Furthermore, $-C$ is the contour C traced in the opposite orientation.

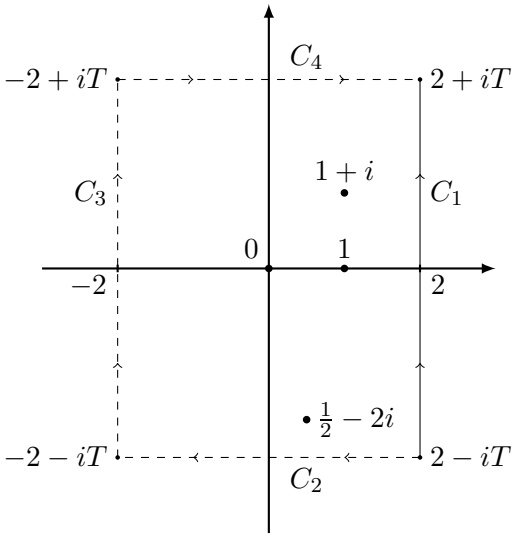

Figure 5.1. The poles of $\zeta(s)\zeta(s-i)\zeta(s+1/2+2i)/s$ inside the rectangle defined by the points $\pm 2 \pm iT$.

C_3 because $\sigma = \alpha' < 1$. In reality, we also need bounds on $F(s)$ to estimate R. Such estimates can be a bit tricky to obtain outside the region of absolute convergence. We will see methods of establishing them in Chapters 6, 8 and 11.

Assuming that R is indeed negligible, we are led to the guesstimate

$$\sum_{n \leqslant x} f(n) \approx \sum_{\substack{w \text{ is a pole of } F(s)/s \\ \alpha' < \mathrm{Re}(w) < \alpha, |\mathrm{Im}(w)| < T}} \mathrm{res}_{s=w} \frac{F(s)x^s}{s}. \tag{5.14}$$

Combining this heuristic with (5.10) explains why $\psi(x)$ should be closely approximated by the sum $x - \sum_{|\gamma| \leqslant T} x^\rho/\rho$ from Theorem 5.1. The rigorous proof of Theorem 5.1 will be given in Chapter 8, after having developed further the theory of the Riemann zeta function (in Chapter 6) and of the Perron inversion formula (in Chapter 7). We will then use the explicit formula for $\psi(x)$ together with a bound on the zeroes of ζ to establish the Prime Number Theorem in Chapter 8.

We conclude this chapter with some examples that showcase the utility and versatility of the ideas presented above.

Example 5.2. As a toy example, consider the function $f = 1$. We then have that $F = \zeta$, whose only singularity is a simple pole of residue 1 at $s = 1$. Thus, the only singularity of $\zeta(s)x^s/s$ in the half-plane $\mathrm{Re}(s) > 0$ is a simple pole of residue x at $s = 1$. This leads us to the prediction that

$\sum_{n\leqslant x} 1 \approx x$. This is of course true, since we know by elementary methods that $\sum_{n\leqslant x} 1 = x + O(1)$. □

Remark 5.3. In general, if F has a simple pole of residue r_w at a point w that is different than the origin, then

$$\operatorname{res}_{s=w} \frac{F(s)x^s}{s} = \frac{r_w x^w}{w}.$$

We can generalize this calculation further: if F has a pole of order m at $w \neq 0$, then there are coefficients $c_{w,0}, c_{w,1}, \dots, c_{w,m-1} \in \mathbb{C}$ such that

$$\operatorname{res}_{s=w} \frac{F(s)x^s}{s} = x^w\big(c_{w,m-1}(\log x)^{m-1} + c_{w,m-2}(\log x)^{m-2} + \cdots + c_{w,0}\big).$$

Indeed, let $F(s)/s = a_{w,m}/(s-w)^m + \cdots + a_{w,1}/(s-w) + \sum_{j\geqslant 0} b_{w,j}(s-w)^j$ be the Laurent expansion of $F(s)/s$ about $s = w$. In addition, we have the Taylor series expansion $x^s = x^w \sum_{j\geqslant 0}(s-w)^j(\log x)^j/j!$. Hence, the claimed formula for $\operatorname{res}_{s=w}(F(s)x^s/s)$ holds with $c_{w,j} = a_{w,j+1}/j!$. □

Example 5.4. Consider the divisor function τ, for which we have the convolution identity $\tau = 1 * 1$. Thus, its Dirichlet series is $\zeta(s)^2$, which has a meromorphic continuation to $\mathbb{C}$ with its only pole being a double pole of order 2 at $s = 1$. In view of relation (5.14) and Remark 5.3, we are led to predict that there are coefficients $c_0, c_1 \in \mathbb{C}$ such that

$$\sum_{n\leqslant x} \tau(n) \approx c_1 x \log x + c_0 x.$$

To calculate c_0 and c_1, note that $\zeta(s) = 1/(s-1) + \gamma + O(|s-1|)$ for $|s-1| \leqslant 1/2$ by Exercise 5.2, whereas $1/s = 1 - (s-1) + O(|s-1|^2)$. Hence

$$\frac{\zeta(s)^2}{s} = \frac{1}{(s-1)^2} + \frac{2\gamma - 1}{s-1} + O(1),$$

which implies that $c_1 = 1$ and $c_0 = 2\gamma - 1$. This agrees with Theorem 3.3. □

Example 5.5. Let f be the indicator function of square-full integers (see Exercise 1.6). In Exercise 3.11, we saw that the partial sums of f up to x have an asymptotic expansion with two main terms, of size $x^{1/2}$ and $x^{1/3}$, respectively. These terms can be guessed using (5.14): the multiplicativity of f implies that its Dirichlet series equals

$$F(s) = \prod_p \Big(1 + \frac{1}{p^{2s}} + \frac{1}{p^{3s}} + \cdots\Big) = \frac{\zeta(2s)\zeta(3s)}{\zeta(6s)} \tag{5.15}$$

for $\operatorname{Re}(s) > 1$. Since ζ has a meromorphic continuation to $\mathbb{C}$, so does F. In addition, the only singularities of F in the half-plane $\operatorname{Re}(s) > 1/6$ are simple

poles at the points $s = 1/2$ and $s = 1/3$. They both arise from the simple pole of ζ at $s = 1$. Relation (5.14) then leads us to the prediction that

$$\#\{\, n \leqslant x : n \text{ square-full} \,\} \approx \frac{\zeta(3/2)}{\zeta(3)} x^{1/2} + \frac{\zeta(2/3)}{\zeta(2)} x^{1/3}. \qquad \square$$

Exercises

Exercise 5.1. Prove that

$$\zeta(s) = \frac{1}{1 - 2^{-s+1}} \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{n^s} \qquad (\mathrm{Re}(s) > 0).$$

Exercise 5.2. When $0 < |s-1| \leqslant 1$, show the following estimates:

$$\begin{aligned} \zeta(s) &= \frac{1}{s-1} + \gamma + O(|s-1|)\,; \\ \log \zeta(s) &= -\log(s-1) + \gamma \cdot (s-1) + O(|s-1|^2) \qquad (s \notin [-1, 0])\,; \\ \frac{\zeta'}{\zeta}(s) &= -\frac{1}{s-1} + \gamma + O(|s-1|)\,. \end{aligned}$$

Exercise 5.3. Use (5.14) to predict the main term in the asymptotic formulas for $\sum_{n \leqslant x} \log n$, $\sum_{n \leqslant x} \varphi(n)$, $\sum_{n \leqslant x} \mu^2(n)$, $\sum_{n \leqslant x} \tau_3(n)$ and $\sum_{n \leqslant x} \tau(n)^2$. Compare your prediction with Theorems 1.12 and 3.2, and Exercises 3.8, 3.10 and 4.5, respectively.

Exercise 5.4*. Complete the proof of Theorem 3.4(c) as follows:

(a) Uniformly for $x \geqslant 2$ and $\varepsilon \in (0, 1]$, prove that

$$\sum_{p \leqslant x} \log\Big(1 - \frac{1}{p}\Big) = \sum_{p \leqslant x} \log\Big(1 - \frac{1}{p^{1+\varepsilon/\log x}}\Big) + O(\varepsilon).$$

(b) Uniformly for $x \geqslant 2$ and $\varepsilon \in (0, 1]$, prove that

$$\sum_{p > x} \log\Big(1 - \frac{1}{p^{1+\varepsilon/\log x}}\Big) = -\int_{\varepsilon}^{\infty} \frac{e^{-u}}{u} \mathrm{d}u + O\Big(\frac{1}{\log x}\Big).$$

[*Hint:* Taylor's theorem.]

(c) Deduce that the constant in (3.6) is $\kappa = \int_0^{\infty} u^{-1}(e^{-u} - 1_{[0,1]}(u))\mathrm{d}u$. [*Hint:* Use Exercise 5.2 to rewrite $\log \zeta(1 + \varepsilon/\log x)$.]

(d) Prove that $\gamma = \int_0^{\infty} u^{-1}(1_{[0,1]}(u) - e^{-u})\mathrm{d}u$.
[*Hint:* Note that $\gamma = \lim_{N\to\infty}(-\log N + \int_0^1 (1 + x + \cdots + x^{N-1})\mathrm{d}x)$ and let $x = 1 - u/N$.]

Chapter 6

The Riemann zeta function

The explicit formula (5.11) underlines the central role of the Riemann zeta function and of its zeroes in the study of prime numbers. We thus undertake a careful study of ζ in this chapter.

The functional equation

One of the most fundamental properties of ζ discovered by Riemann himself is that it possesses a symmetry with respect to the vertical line $\mathrm{Re}(s) = 1/2$. This symmetry is depicted in the *functional equation* of ζ: for all $s \in \mathbb{C}$, we have

$$\pi^{-s/2}\Gamma\Big(\frac{s}{2}\Big)\zeta(s) = \pi^{-(1-s)/2}\Gamma\Big(\frac{1-s}{2}\Big)\zeta(1-s), \tag{6.1}$$

where Γ is Euler's Gamma function. We can also rewrite (6.1) as[1]

$$\zeta(s) = \lambda(s)\zeta(1-s), \qquad \text{where} \qquad \lambda(s) = \pi^{s-1/2}\frac{\Gamma((1-s)/2)}{\Gamma(s/2)}. \tag{6.2}$$

Using Exercise 1.13, we find two alternative expressions for λ:

$$\lambda(s) = 2^{s-1}\pi^s \big/ \big(\Gamma(s)\cos(\pi s/2)\big) = 2^s\pi^{s-1}\Gamma(1-s)\sin(\pi s/2). \tag{6.3}$$

As we saw in Chapter 1, the Gamma function is very well understood. Hence it suffices to study ζ in the half-plane $\mathrm{Re}(s) \geqslant 1/2$ and then use the functional equation to pass to the entire complex plane.

[1] In the literature, the function λ is usually denoted by χ. Since we have reserved the letter χ for Dirichlet characters, we use the letter λ, which is the first letter of the Greek word λόγος that means ratio.

Let us now show (6.1). In the process of doing so, we will see another proof of the meromorphic continuation of ζ to $\mathbb{C}$.

At the heart of the proof of (6.1) lies the *Poisson summation formula* (Theorem B.3). Let $f : \mathbb{R} \to \mathbb{C}$ be as in Theorem B.3, that is to say, $f \in C^2(\mathbb{R})$ and $f^{(j)}(x) \ll 1/x^2$ for $|x| \geqslant 1$ and $j \in \{0, 1, 2\}$, so that its Fourier transform

$$\widehat{f}(\xi) = \int_{\mathbb{R}} f(x) e^{-2\pi i \xi x} \mathrm{d}x$$

satisfies the bound $\widehat{f}(\xi) \ll 1/\xi^2$ for $|\xi| \geqslant 1$. Assume further that f is even and consider its Mellin transform

$$F(s) = \int_0^\infty f(x) x^{s-1} \mathrm{d}x,$$

which is well defined for $0 < \mathrm{Re}(s) < 2$. The change of variables $x \to nx$ implies that

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

Summing this formula over all $n \geqslant 1$ when $1 < \mathrm{Re}(s) < 2$, we find that

$$\zeta(s) F(s) = \int_0^\infty S_f(x) x^{s-1} \mathrm{d}x \qquad \text{with} \qquad S_f(x) = \sum_{n \geqslant 1} f(nx). \tag{6.4}$$

Next, we use Poisson's summation formula (B.3) to deduce that

$$f(0) + 2S_f(x) = \sum_{n \in \mathbb{Z}} f(nx) = \frac{1}{x} \sum_{n \in \mathbb{Z}} \widehat{f}(n/x) = \frac{\widehat{f}(0) + 2S_{\widehat{f}}(1/x)}{x},$$

since f and $\widehat{f}$ are even. We then split the range of integration on the right side of (6.4) as $(1, +\infty) \cup [0, 1]$. We also make the change of variables $x \to 1/x$ to the portion over $[0, 1]$. We thus find that

$$\begin{aligned} \zeta(s) F(s) &= \int_1^\infty S_f(x) x^{s-1} \mathrm{d}x + \int_1^\infty S_f(1/x) x^{-s-1} \mathrm{d}x \\ &= \int_1^\infty S_f(x) x^{s-1} \mathrm{d}x + \int_1^\infty S_{\widehat{f}}(x) x^{(1-s)-1} \mathrm{d}x + \frac{\widehat{f}(0)}{2(s-1)} - \frac{f(0)}{2s}. \end{aligned}$$

In order to symmetrize the above formula, we choose $f(x) = 2e^{-\pi x^2}$ that is self-dual, that is to say, $\widehat{f} = f$. For this choice of f, we have that

$$F(s) = 2 \int_0^\infty e^{-\pi x^2} x^{s-1} \mathrm{d}x = \pi^{-s/2} \int_0^\infty e^{-y} y^{s/2-1} \mathrm{d}y = \pi^{-s/2} \Gamma(s/2),$$

so that

$$\pi^{-s/2} \Gamma\Big(\frac{s}{2}\Big) \zeta(s) = \int_1^\infty S(x) x^{s-1} \mathrm{d}x + \int_1^\infty S(x) x^{(1-s)-1} \mathrm{d}x + \frac{1}{s(s-1)}$$

with $S(x) = 2\sum_{n=1}^{\infty} e^{-\pi n x^2}$. The right-hand side of the above formula is clearly an analytic function for $s \in \mathbb{C}$ except for simple poles at $s = 1$ and at $s = 0$. It is also invariant with respect to the change of variables $s \to 1 - s$, thus proving the functional equation (6.1). Since Γ does not vanish in $\mathbb{C}$ (see Corollary 1.14) and has a simple pole at $s = 0$, we also deduce the meromorphic continuation of ζ to $\mathbb{C}$ with its sole singularity being a simple pole at $s = 1$.

The zeroes of ζ and the Riemann Hypothesis

Let us now discuss the zeroes of the Riemann zeta function that are intimately related to the distribution of primes.

When $\mathrm{Re}(s) > 1$, we know that $\zeta(s)$ is given by an absolutely convergent Euler product whose factors do not vanish. In particular, $\zeta(s) \neq 0$ (see Remark 4.7(b)). In addition, Theorem 1.14 implies that Γ does not vanish at all in $\mathbb{C}$. As a consequence, the left-hand side of the functional equation (6.1) is non-zero for $\mathrm{Re}(s) > 1$. Hence, the right-hand side of (6.1) must also not vanish in the same region. Equivalently, $\zeta(s)\Gamma(s/2) \neq 0$ for $\mathrm{Re}(s) < 0$. However, note that $\Gamma(s/2)$ has simple poles at the points $-2, -4, \ldots$. Hence, ζ must have simple zeroes at $-2, -4, \ldots$, and no other zeroes when $\mathrm{Re}(s) < 0$. The pole of Γ at 0 does not induce a zero of ζ, because it is counterbalanced by the pole of ζ at 1. In fact, $\zeta(0) = -1/2$ by Exercise 6.3.

As we mentioned in Chapter 5, the zeroes of ζ at the negative even integers are called *trivial*. All other zeroes lie in the strip $0 \leqslant \mathrm{Re}(s) \leqslant 1$ and are called *non-trivial*. We denote them by $\rho = \beta + i\gamma$. These are the zeroes appearing on the right-hand side of the functional equation (5.11). For this reason, the strip $0 \leqslant \mathrm{Re}(s) \leqslant 1$ is called *the critical strip*.

The functional equation and the obvious symmetry $\zeta(\bar{s}) = \bar{\zeta}(s)$ imply that if ρ is a non-trivial zero of ζ, then so are the numbers $\bar{\rho}$, $1-\rho$ and $1-\bar{\rho}$. In his 1859 mémoire, Riemann postulated that all non-trivial zeroes of ζ lie on the line $\mathrm{Re}(s) = 1/2$, which is the line of symmetry of ζ. We thus refer to this line as *the critical line*. Riemann's conjecture is known today as the *Riemann Hypothesis*.

The Riemann Hypothesis is a very important conjecture because it offers us unparalleled control on the distribution of primes. To explain this claim, we go back to the explicit formula (5.11). If $\rho = 1/2 + i\gamma$ for all non-trivial zeroes of ζ, then

$$\Big| \sum_{|\gamma| \leqslant T} \frac{x^{1/2+i\gamma}}{1/2 + i\gamma} \Big| \leqslant x^{1/2} \sum_{|\gamma| \leqslant T} \frac{1}{\sqrt{1/4 + \gamma^2}}.$$

As we will prove later (see Lemma 8.2(a)), the number of zeroes with $|\gamma| \in [n, n+1]$ is $\ll \log(n+2)$, so that

$$\Big| \sum_{|\gamma| \leqslant T} \frac{x^{1/2+i\gamma}}{1/2+i\gamma} \Big| \ll x^{1/2} \sum_{0 \leqslant n \leqslant T} \frac{\log(n+2)}{n} \ll x^{1/2} \log^2 T. \tag{6.5}$$

Taking $T = x$ establishes the remarkably accurate estimate

$$\psi(x) = x + O(\sqrt{x} \log^2 x) \tag{6.6}$$

uniformly for all $x \geqslant 2$. By partial summation (see Exercise 1.7), this is equivalent to having that

$$\pi(x) = \mathrm{li}(x) + O(\sqrt{x} \log x) \tag{6.7}$$

uniformly for all $x \geqslant 2$, where we recall that $\mathrm{li}(x) = \int_2^x \mathrm{d}t/\log t$.

Juxtaposing (6.7) and (0.4), we see that under the Riemann Hypothesis the distribution of primes is as close to being "random" as we could hope for (up to factors of $\log x$ and $\log\log x$). In fact, Exercise 8.2(c) shows that we cannot replace $\sqrt{x}$ by a smaller power of x in (6.7), thus making (6.7) "the best of all possible worlds".

In contrast, the best known version of the Prime Number Theorem establishes (6.7) with a much weaker error of size $x \exp(-c(\log x)^{3/5}/(\log\log x)^{1/5})$ [114, Corollary 8.30]. In Chapter 8, we will show (6.7) with a remainder term of size $x \exp(-c\sqrt{\log x})$ by proving that ζ does not vanish too close to the line $\mathrm{Re}(s) = 1$.

Remarkably, (6.6) (and hence (6.7)) is equivalent to the Riemann Hypothesis: indeed, let $\psi(x) = x + E(x)$ and apply (1.12) with $a_n = \Lambda(n)$, $f(n) = 1/n^s$, $y = 1$ and $z \to +\infty$ to find that

$$-\frac{\zeta'}{\zeta}(s) = \frac{s}{s-1} + s \int_1^\infty \frac{E(u)}{u^{s+1}} \mathrm{d}u \qquad (\mathrm{Re}(s) > 1).$$

But if $E(u) = O(\sqrt{u} \log^2 u)$ for $u \geqslant 2$, the right-hand side of the above formula is meromorphic for $\mathrm{Re}(s) > 1/2$, thus providing a meromorphic continuation of $-\zeta'/\zeta$ to the half-plane $\mathrm{Re}(s) > 1/2$, with the only pole located at $s = 1$. In particular, ζ does not vanish in this half-plane. By the functional equation (6.1), it cannot vanish in the half-plane $\mathrm{Re}(s) < 1/2$ either, and the Riemann Hypothesis follows. We have thus established:

Theorem 6.1. *The Riemann Hypothesis is true if and only if* (6.6) *holds.*

The order of magnitude of ζ

In the previous chapter, we gave a rough outline of the proof of the explicit formula for $\psi(x)$. More generally, we saw how to estimate the partial sums of an arithmetic function f in terms of the singularities of its Dirichlet series

F. A crucial technical step that we set aside in this discussion is the need to bound F past its region of absolute convergence. We explain here how to do this when $F = \zeta$. In Chapter 8, we will develop additional tools that will also allow us to handle the quotient ζ'/ζ and establish Theorem 5.1.

So, let us suppose that we are given some $s \in \mathbb{C}$. We then want to understand the size of $\zeta(s)$. The functional equation (6.2) and Exercise 1.12 imply that

$$|\zeta(s)| \asymp_C |\zeta(1-s)| \cdot |t|^{1/2-\sigma} \quad (-C \leqslant \sigma \leqslant 1/2,\ |t| \geqslant 1). \tag{6.8}$$

Hence, it suffices to bound $|\zeta(s)|$ when $\mathrm{Re}(s) \geqslant 1/2$.

When $\mathrm{Re}(s) > 1$, this is relatively easy: since ζ is given by an absolutely convergent Euler product on this half-plane, we have

$$1/\zeta(\sigma) \leqslant |\zeta(\sigma + it)| \leqslant \zeta(\sigma). \tag{6.9}$$

Indeed, this follows by noticing that $1 - |z| \leqslant |1/(1-z)| \leqslant 1/(1-|z|)$ when $|z| < 1$. In particular, we conclude that $|\zeta(s)| \asymp_\varepsilon 1$ for $\sigma \geqslant 1 + \varepsilon$. We then also find by (6.8) that $|\zeta(s)| \asymp_{\varepsilon,C} |t|^{1/2-\sigma}$ for $\sigma \in [-C, -\varepsilon]$.

On the other hand, bounding ζ inside the critical strip $0 \leqslant \mathrm{Re}(s) \leqslant 1$ is much harder. It turns out we can use the information we have outside the critical strip to extrapolate a bound for ζ inside it. This uses the Phragmén-Lindelöf principle [159, Chapter 12], which is a generalization of the maximum modulus principle.

Theorem 6.2. *Let f be a function that is analytic in an open neighborhood of the vertical strip $\alpha_1 \leqslant \mathrm{Re}(s) \leqslant \alpha_2$, and for which there is an absolute constant C such that $f(s) \ll \exp\{|t|^C\}$ when $\alpha_1 \leqslant \mathrm{Re}(s) \leqslant \alpha_2$. Assume further that $f(\sigma_j + it) \ll (1+|t|)^{\theta_j}$ for $j = 1, 2$ and all $t \in \mathbb{R}$.*

Given $\sigma \in [\alpha_1, \alpha_2]$, there is a unique $u \in [0,1]$ such that $\sigma = u\alpha_1 + (1-u)\alpha_2$. We then have

$$f\big(\sigma + it\big) \ll (1+|t|)^{u\theta_1 + (1-u)\theta_2} \qquad (t \in \mathbb{R}).$$

We postpone the proof of this theorem momentarily because it is a bit technical and use it to study ζ. In fact, because of the pole of ζ at $s = 1$, we work instead with the function $f(s) = (s-1)\zeta(s)$ that is entire. Note that f grows at most polynomially in $|t|$, that is to say, $f(s) \ll_\sigma (1+|t|)^{O(1)}$, as it can be readily seen by relation (5.8). Since $\zeta(1+\varepsilon+it) \ll_\varepsilon 1$ and $\zeta(-\varepsilon + it) \ll_\varepsilon |t|^{1/2+\varepsilon}$ for $|t| \geqslant 1$, Theorem 6.2 implies that $\zeta(s) \ll |t|^{(1-\sigma+\varepsilon)/2}$ for $-\varepsilon \leqslant \sigma \leqslant 1+\varepsilon$ and $|t| \geqslant 1$.

To summarize the above discussion, we have shown the following result.

Theorem 6.3. *Fix $\varepsilon > 0$ and $C \geqslant 1$. For $s = \sigma + it$ with $\sigma \geqslant -C$ and $|t| \geqslant 1$, we have*

$$\zeta(s) \ll_{\varepsilon,C} \begin{cases} 1 & \text{if } \sigma \geqslant 1+\varepsilon, \\ |t|^{(1-\sigma+\varepsilon)/2} & \text{if } -\varepsilon \leqslant \sigma \leqslant 1+\varepsilon, \\ |t|^{1/2-\sigma} & \text{if } -C \leqslant \sigma \leqslant -\varepsilon. \end{cases}$$

Motivated by the above theorem, we define

$$\ell(\sigma) = \limsup_{|t|\to\infty} \frac{\log|\zeta(\sigma+it)|}{\log|t|} \tag{6.10}$$

for each $\sigma \in \mathbb{R}$, that is to say, $\ell(\sigma)$ is the smallest number such that

$$|\zeta(\sigma+it)| \ll_{\varepsilon,\sigma} |t|^{\ell(\sigma)+\varepsilon} \qquad (|t| \geqslant 1)$$

for each fixed $\varepsilon > 0$. The discussion in the beginning of the section implies that $\ell(\sigma) = 0$ for $\sigma > 1$, that $\ell(\sigma) = 1/2 - \sigma$ for $\sigma < 0$ and that

$$\ell(\sigma) = 1/2 - \sigma + \ell(1-\sigma).$$

Furthermore, Theorem 6.2 implies that $\ell(\sigma)$ is a convex function. In particular, it is continuous (see Exercise 6.7), so that $\ell(1) = 0$ and $\ell(0) = 1/2$. It is believed that

$$\ell(\sigma) = \begin{cases} 0 & \text{if } \sigma \geqslant 1/2, \\ 1/2 - \sigma & \text{if } 0 \leqslant \sigma \leqslant 1/2. \end{cases}$$

This is known as the *Lindelöf hypothesis*.

The convexity of $\ell(\sigma)$ reduces the Lindelöf hypothesis to the case when $\sigma = 1/2$. Any improvement of the exponent $1/4$ in the estimate $\zeta(1/2+it) \ll_\varepsilon |t|^{1/4+\varepsilon}$ of Theorem 6.3 is called a *subconvexity estimate*. In turn, this is essentially equivalent to proving that the sum $\sum_{n\leqslant x} n^{it}$ is small compared to x (i.e., it "exhibits cancellation") when x is in the vicinity of $|t|^{1/2}$ (see formula (7.18)). The current record is $\ell(1/2) \leqslant 13/84 \approx 0.154$ due to Bourgain [16].

Proof of Theorem 6.2. By a linear change of variables, we may assume that $\alpha_1 = 0$ and $\alpha_2 = 1$, so that our goal is to show that $f(\sigma+it) \ll (1+|t|)^{(1-\sigma)\theta_1+\sigma\theta_2}$. In addition, we may assume that $\theta_1, \theta_2 \geqslant 0$; otherwise, we replace $f(s)$ by $f(s)(s+1)^k$ for a large integer k.

To study f at height t, we consider the function $f(z+it)$ with $0 \leqslant \mathrm{Re}(z) \leqslant 1$. We further normalize $f(z+it)$ to be bounded on the boundary of the strip $0 \leqslant \mathrm{Re}(z) \leqslant 1$ by letting

$$g_t(z) := f(z+it) \,/\, \big[(1+|t|)^{(1-z)\theta_1+z\theta_2} \cdot (z+1)^N\big],$$

where $N = \max\{\lceil\theta_1\rceil, \lceil\theta_2\rceil\}$. Indeed, note that

$$|g_t(iy)| \ll (1+|t+y|)^{\theta_1} \big/ \big[(1+|t|)^{\theta_1} \cdot (1+|y|)^N\big] \leqslant 1$$

for $y \in \mathbb{R}$, by our assumption that $N \geqslant \theta_1 \geqslant 0$. Similarly, $|g_t(1+iy)| \ll 1$.

We have shown that g_t is uniformly bounded on the boundary of the strip $0 \leqslant \mathrm{Re}(z) \leqslant 1$. If we knew its maximum occurred on this boundary, the theorem would readily follow, since

$$|f(\sigma+it)| \leqslant 2^N(1+|t|)^{(1-\sigma)\theta_1+\sigma\theta_2} \cdot |g_t(\sigma)| \quad (0 \leqslant \sigma \leqslant 1). \tag{6.11}$$

The main idea of the Phragmén-Lindelöf principle is to construct an auxiliary function which is bounded and whose maximum does occur on the boundary of the strip $0 \leqslant \mathrm{Re}(z) \leqslant 1$. To this end, we let

$$h_\varepsilon(z) = \exp\{\varepsilon(e^{i\pi z/4} + e^{-i\pi z/4})\},$$

where $\varepsilon > 0$ is fixed for the moment. If $z = x+iy$, then note that

$$\mathrm{Re}(e^{i\pi z/4} + e^{-i\pi z/4}) = \cos(\pi x/4)(e^{-\pi y/4} + e^{\pi y/4}) \geqslant e^{\pi|y|/4}/\sqrt{2}$$

for $x \in [0,1]$ and $y \in \mathbb{R}$. Our assumption that $f(z+it) \ll \exp\{|t+y|^C\}$ implies that the function g_t/h_ε is bounded. In fact, we have $|(g_t/h_\varepsilon)(x+iy)| \leqslant 1$ for all $x \in [0,1]$, as long as y is large enough in terms of ε and t (we suppress the dependence on C, since we consider it fixed).

Let $Y = Y(\varepsilon, t)$ be such that $|(g_t/h_\varepsilon)(x+iy)| \leqslant 1$ for $y \geqslant Y$ and $x \in [0,1]$. For each $T \geqslant Y$, we consider the rectangle R_T with vertices $\pm iT$ and $1 \pm iT$. Note that g_t/h_ε is uniformly bounded on the boundary of R_T (independently of ε, t and T). The maximum modulus principle implies that $(g_t/h_\varepsilon)(z) \ll 1$ for all $z \in R_T$. Letting $T \to \infty$, we find that $g_t(z) \ll h_\varepsilon(z)$ for all z in the strip $0 \leqslant \mathrm{Re}(z) \leqslant 1$, uniformly in $t \in \mathbb{R}$ and $\varepsilon > 0$. We then let $\varepsilon \to 0^+$ to deduce that $g_t(z) \ll 1$, uniformly in t. Hence, the theorem readily follows by (6.11). □

Exercises

Exercise 6.1. Show that the Riemann Hypothesis is equivalent to knowing that $\mathrm{Re}(\rho) \leqslant 1/2$ for all non-trivial zeroes ρ.

Exercise 6.2. Show that $\zeta(\sigma) < 0$ when $-2 < \sigma < 1$.

Exercise 6.3. Prove that $\zeta(-n) = (-1)^n B_{n+1}/(n+1)$ for $n \geqslant 0$. [*Hint:* Exercise 1.10(d).]

Exercise 6.4. Use (6.1) and Theorem 1.14 to show that

$$\frac{\zeta'}{\zeta}(s) + \frac{\zeta'}{\zeta}(1-s) = \frac{1}{s} + \frac{1}{1-s} + \gamma + \log\pi + \sum_{n\geqslant 1}\Big(\frac{1}{2n+s} + \frac{1}{2n+1-s} - \frac{1}{n}\Big).$$

Conclude that $(\zeta'/\zeta)(0) = \log(2\pi)$ and $\zeta'(0) = -\log(2\pi)/2$.

Exercise 6.5. For $t \in \mathbb{R}$, let[2]

$$\vartheta(t) = \arg\Gamma(1/4 + it/2) - t(\log \pi)/2.$$

Show that *Hardy's function* $Z(t) := e^{i\vartheta(t)}\zeta(1/2 + it)$ is real valued.

Exercise 6.6. Let $\theta \in (0, 1)$ be such that $\sum_{n \leqslant x} \mu(n) = O(x^\theta)$ for all $x \geqslant 1$. Prove that $\zeta(s) \neq 0$ for $\mathrm{Re}(s) > \theta$.

Exercise 6.7. Prove that a convex function $f : [a, b] \to \mathbb{R}$ is continuous. [*Hint:* For each $x \in [a, b]$, show that the ratio $(f(y) - f(x))/(y - x)$ is increasing as a function of $y \in [a, b] \setminus \{x\}$.]

Exercise 6.8. Prove that the function $\ell(\sigma)$ is non-negative and decreasing.

Exercise 6.9. Let $f(s)$ be a bounded analytic function in an open neighborhood of the strip $0 \leqslant \mathrm{Re}(s) \leqslant 1$. If $M_\sigma = \sup_t |f(\sigma + it)|$, then show that $M_\sigma \leqslant M_0^{1-\sigma} M_1^\sigma$. [*Hint:* Consider $f_\varepsilon(s) = f(s)M_0^{s-1}M_1^{-s}/(1 + \varepsilon s)$.]

Exercise 6.10. When $\sigma_1 = -1$ and $\sigma_1 = 1$, show that we may relax the condition $f(s) \ll \exp\{|t|^C\}$ in Theorem 6.2 to $|f(s)| \leqslant \exp\{Ae^{B|t|}\}$, where $A \geqslant 0$ and $0 \leqslant B < \pi/2$. Furthermore, show that we cannot take $B \geqslant \pi/2$. [*Hint:* Consider the function $f(s) = \exp\{\cos(\pi s/2)\}$.]

[2]Since $\Gamma(s)$ does not vanish and is analytic for $\mathrm{Re}(s) > 0$, we may define $\log\Gamma(s)$ for $\mathrm{Re}(s) > 0$. We take the branch that is real valued for $s > 0$. Then $\arg\Gamma(s) = \mathrm{Im}\log\Gamma(s)$, as usual.

Chapter 7

The Perron inversion formula

If f is an arithmetic function whose Dirichlet series F can be meromorphically continued past its domain of absolute convergence, then we expect that the asymptotic behavior of the partial sums of f is determined by the singularities of F, as described in the guesstimate (5.14). The main goal of this chapter is to justify this heuristic.

In our discussion of Perron's inversion formula in Chapter 5, we ignored a subtle technical issue: the choice of the parameter T used to truncate the integral $\int_{(\alpha)}(F(s)x^s/s)\mathrm{d}s$ and approximate it by $\int_{\sigma=\alpha,\,|t|\leqslant T}(F(s)x^s/s)\mathrm{d}s$. In practice, it is important to have a quantitative form of Perron's inversion formula (5.9) that allows us to choose T as an appropriate function of x. To do so, we approach (5.9) from a slightly different point of view.

The key observation is that

$$\sum_{n<x} f(n) + \frac{1_{x\in\mathbb{N}} f(x)}{2} = \sum_{n\geqslant 1} f(n)\delta(n/x) \qquad \text{with} \qquad \delta(y) = 1_{0<y<1} + \frac{1_{y=1}}{2}.$$

The Mellin transform of δ is $\int_0^\infty \delta(y)y^{s-1}\mathrm{d}y = 1/s$, so that

$$\sum_{n<x} f(n) + \frac{1_{x\in\mathbb{N}} f(x)}{2} = \sum_{n\geqslant 1} f(n)\cdot\frac{1}{2\pi i}\int_{(\alpha)}\frac{(x/n)^s}{s}\mathrm{d}s \tag{7.1}$$

for each $\alpha > 1$ (or, even, for $\alpha > 0$). The next natural step is to interchange the order of summation and integration, which would yield (5.9). It is hard to justify this step directly because the integrals on the right side of (7.1) do not converge absolutely. We discuss two ways to circumvent this problem.

The first method: Truncating Perron's integral. Instead of using the exact formula $\delta(y) = (1/2\pi i)\int_{(\alpha)}(y^{-s}/s)\mathrm{d}s$ with $\alpha > 0$, we can use the following truncated form of it that offers substantial technical flexibility.

Lemma 7.1. *Uniformly for $y > 0$, $\alpha > 0$ and $T \geqslant 1$, we have*

$$\frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(s)=\alpha\\ |\mathrm{Im}(s)|\leqslant T}} \frac{y^{-s}}{s}\mathrm{d}s = \begin{cases} 1_{0<y<1} + O\big(y^{-\alpha}/\max\{1, T|\log y|\}\big) & \text{if } y \neq 1,\\ 1/2 + O(\alpha/T) & \text{if } y = 1. \end{cases}$$

We postpone the proof of the above result till the end of the chapter. Note that if $\sum_{n\geqslant 1}|f(n)|/n^\alpha$ converges for some $\alpha > 0$, then Lemma 7.1 implies

$$\sum_{n<x} f(n) + \frac{1_{x\in\mathbb{N}}f(x)}{2} = \frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(s)=\alpha\\ |\mathrm{Im}(s)|\leqslant T}} F(s)\frac{x^s}{s}\mathrm{d}s + O(x^\alpha \cdot R), \tag{7.2}$$

where

$$R = \sum_{n\geqslant 1}\frac{|f(n)|}{n^\alpha \max\{1, T|\log\frac{x}{n}|\}}.$$

The error term can be bounded if f does not grow too fast, thus yielding a quantitative version of (5.9) as follows.

Theorem 7.2. *Let f be an arithmetic function with Dirichlet series $F(s) = \sum_{n=1}^\infty f(n)/n^s$. Assume there are constants $A, C \geqslant 0$ and $\theta \geqslant -1$ such that*

$$|f(n)| \leqslant Cn^\theta(1+\log n)^A \qquad (n \in \mathbb{N}).$$

For $x, T \geqslant 2$ and $\alpha \geqslant \theta + 1 + 1/\log x$, we have that

$$\sum_{n\leqslant x} f(n) = \frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(s)=\alpha\\ |\mathrm{Im}(s)|\leqslant T}} F(s)\frac{x^s}{s}\mathrm{d}s + O\Big(\frac{x^\alpha(\log x)^{A+1}}{T} + x^{\alpha-1}(\log x)^A\Big);$$

the implied constant depends at most on A, C and θ.

We prove Theorem 7.2 at the end of the chapter, along with Lemma 7.1.

The second method: Using smooth cut-offs. We now discuss an alternative way to obtain a quantitative form of Perron's inversion formula. The underlying cause for the slow decay of the integrand in Perron's inversion formula (5.9) is that the function δ is discontinuous. This is a reflection of the uncertainty principle in harmonic analysis: the discontinuous function δ is *too* localized on the interval $[0, 1]$, so its Mellin transform must have relatively *heavy tails* (that is to say, it cannot decay too fast at infinity). In order to get around this issue, we approximate δ by a more delocalized function whose Mellin transform decays faster at infinity.

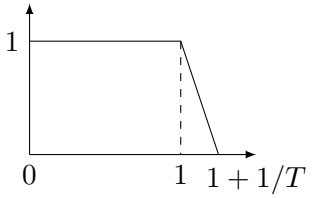

Figure 7.1. The graph of the function $\delta_T(y)$.

A concrete example is provided by the function

$$\delta_T(y) = \begin{cases} 1 & \text{if } 0 \leqslant y \leqslant 1, \\ T + 1 - Ty & \text{if } 1 < y < 1 + 1/T, \\ 0 & \text{if } y \geqslant 1 + 1/T, \end{cases}$$

where $T \geqslant 1$ is some large parameter that plays an analogous role to that of the truncation point in Lemma 7.1 (see Figure 7.1). By construction, we have

$$\sum_{n \leqslant x} f(n) = \sum_{n \geqslant 1} f(n)\delta_T(n/x) + O\Big(\sum_{x < n \leqslant x + x/T} |f(n)| \Big).$$

We then rewrite δ_T in terms of its Mellin transform, which is equal to

$$\int_0^\infty \delta_T(y) y^{s-1} \mathrm{d}y = \frac{(1 + 1/T)^{s+1} - 1}{s(s+1)/T}.$$

Notice that this is an absolutely integrable function on each line $\mathrm{Re}(s) = \alpha \neq 0$, as opposed to the Mellin transform of δ. In addition, Theorem B.4 (applied here with $\alpha_1 = 1$ and $\alpha_2 = \infty$) implies that

$$\delta_T(n/x) = \frac{1}{2\pi i} \int_{(\alpha)} \frac{(1 + 1/T)^{s+1} - 1}{s(s+1)/T} (x/n)^s \mathrm{d}s$$

for any $\alpha > 1$. As a consequence,

$$\sum_{n \leqslant x} f(n) = \frac{1}{2\pi i} \int_{(\alpha)} F(s) \frac{x^s}{s} \cdot \frac{(1 + 1/T)^{s+1} - 1}{(s+1)/T} \mathrm{d}s + O\Big(\sum_{x < n \leqslant x + x/T} |f(n)| \Big). \tag{7.3}$$

Notice that $(1 + 1/T)^{s+1} - 1 \sim (s+1)/T$ when $s = o(T)$, so that the integrands in (7.3) and in Theorem 7.2 are asymptotically the same for small s. In addition, the absolute convergence of the integral in (7.3) allows us to truncate it in a very straightforward way. The larger T is, the better we can control the error term $\sum_{x < n \leqslant x + x/T} |f(n)|$, but the worse bounds we have on the Mellin transform of δ_T due to the presence of T^{-1} in the denominator. Consequently, we have to choose the parameter T in an optimal way that

balances the gains and the losses. A similar situation arises when using Lemma 7.1 to truncate Perron's formula. We will see a concrete application of this version of Perron's formula later, in the proof of Theorem 13.2.

Taking the above idea one step further, we approximate the sharp cut-off function δ by a smooth function $\phi \in C^\infty(\mathbb{R}_{\geqslant 0})$ such that

$$\begin{cases} \phi(y) = 1 & \text{if } 0 \leqslant y \leqslant 1, \\ 0 \leqslant \phi(y) \leqslant 1 & \text{if } 1 < y < 1 + 1/T, \\ \phi(y) = 0 & \text{if } y \geqslant 1 + 1/T. \end{cases} \tag{7.4}$$

Example 7.3. A simple way to construct such a ϕ is to begin with a smooth function $g \geqslant 0$ that is supported on $[0,1]$ and for which $\int_0^1 g(x)\mathrm{d}x = 1$. We then set $g_T(x) = T \cdot g(Tx)$, which is supported on $[0, 1/T]$ and whose integral over $\mathbb{R}$ also equals 1, and take $\phi(y) = \int_{y-1}^{y+1/T} g_T(w)\mathrm{d}w$ for $y > 0$. Clearly $0 \leqslant \phi(y) \leqslant \int_{-\infty}^{\infty} g_T(w)\mathrm{d}w = 1$. Moreover, if $y \in [0,1]$, then $y - 1 \leqslant 0$ and $y + 1/T \geqslant 1/T$, so that $\phi(y) = \int_0^{1/T} g_T(w)\mathrm{d}w = 1$. Finally, if $y \geqslant 1 + 1/T$, then $y - 1 \geqslant 1/T$, so that $\phi(y) = 0$. This proves that ϕ satisfies (7.4).

Notice that for the constructed function ϕ we have

$$\|\phi^{(k)}\|_\infty \ll_k T^k. \tag{7.5}$$

This is the typical behavior for the kth derivative of functions ϕ satisfying (7.4), since they vary by 1 in an interval of length $1/T$. □

Given ϕ satisfying (7.4), we consider its Mellin transform

$$\Phi(s) = \int_0^\infty \phi(y) y^{s-1}\mathrm{d}y$$

that converges absolutely for $\sigma > 0$. Integrating by parts $k+1$ times, and noticing that $\phi^{(k+1)}$ is supported on $[1, 1 + 1/T]$, we find

$$\Phi(s) = \frac{(-1)^{k+1}}{s(s+1)\cdots(s+k)} \int_1^{1+1/T} \phi^{(k+1)}(y) y^{s+k}\mathrm{d}y. \tag{7.6}$$

This provides a meromorphic continuation of Φ to the entire complex plane. The only potential poles are at the points $s = -k$ for $k \in \mathbb{Z}_{\geqslant 0}$ of residue

$$\operatorname{res}_{s=-k} \Phi(s) = -\frac{1}{k!} \int_0^\infty \phi^{(k+1)}(y)\mathrm{d}y = \frac{\phi^{(k)}(0)}{k!}.$$

By (7.4), we have that $\phi(0) = 1$ and $\phi^{(k)}(0) = 0$ for $k \geqslant 1$, so that the only pole is at $s = 0$ and its residue equals 1.

Finally, using (7.6), we find that

$$\Phi(s) \ll_k T^{-1} \cdot |s|^{-k-1} \cdot \|\phi^{(k+1)}\|_\infty \cdot (1 + 1/T)^{\max\{\sigma, 0\}}$$

for $|\operatorname{Im}(s)| \geqslant 1$ and $k \in \mathbb{Z}_{\geqslant 0}$. If ϕ satisfies (7.5), then

$$\Phi(s) \ll_k (1+1/T)^{\max\{\sigma,0\}} \min_{0\leqslant j\leqslant k} \frac{T^j}{|s|^{j+1}} = \frac{(1+1/T)^{\max\{\sigma,0\}}}{|s|\max\{1,|s|/T\}^k} \tag{7.7}$$

for $|t| \geqslant 1$ and $k \in \mathbb{Z}_{\geqslant 0}$. In particular, $\Phi(s)$ starts decaying extremely fast as soon as $|s| > T$, which is in accordance with the uncertainty principle.

Let us know see how we can use the above discussion to estimate the partial sums of f. We start by observing that

$$\sum_{n\leqslant x} f(n) = \sum_{n\geqslant 1} f(n)\phi(n/x) + O\Big(\sum_{x<n\leqslant x+x/T} |f(n)|\Big).$$

For any $\alpha > 0$, Mellin's inversion formula (Theorem B.4) implies that

$$\phi(n/x) = \frac{1}{2\pi i}\int_{(\alpha)} \Phi(s)(x/n)^s \mathrm{d}s.$$

If $F(s)$ converges absolutely when $\operatorname{Re}(s) = \alpha$, then

$$\begin{aligned}\sum_{n\geqslant 1} f(n)\phi(n/x) &= \sum_{n\geqslant 1} \frac{f(n)}{2\pi i}\int_{(\alpha)} \frac{\Phi(s)x^s}{n^s}\mathrm{d}s \\ &= \frac{1}{2\pi i}\int_{(\alpha)} F(s)\Phi(s)x^s \mathrm{d}s,\end{aligned} \tag{7.8}$$

where the change of order of summation and integration is justified by Lebesgue's Dominated Convergence Theorem.

Similarly to ζ, it is often the case that F has a meromorphic continuation to a half-plane $\sigma > \alpha_0$ for some $\alpha_0 < \alpha$. In this case, F usually satisfies its own version of Theorem 6.3: for any fixed $\sigma > \alpha_0$, the function $|F(\sigma+it)|$ is bounded by a suitable power of $|t|$ when $|t| \to \infty$. On the other hand, (7.7) implies that $|\Phi(\sigma+it)|$ grows faster than *any* fixed power of $|t|$, so that $F(s)\Phi(s)$ is absolutely integrable on any vertical line $\operatorname{Re}(s) = \alpha'$ with $\alpha' > \alpha_0$. Using Cauchy's residue theorem, we arrive at the formula

$$\begin{aligned}\frac{1}{2\pi i}\int_{(\alpha)} F(s)\Phi(s)x^s\mathrm{d}s = \frac{1}{2\pi i}\int_{(\alpha')} F(s)\Phi(s)x^s\mathrm{d}s \\ + \sum_{\alpha'<\operatorname{Re}(w)<\alpha} \operatorname{res}_{s=w}(F(s)\Phi(s)x^s),\end{aligned} \tag{7.9}$$

where the sum on the last line runs over the singularities of $F(s)\Phi(s)$. Indeed, (7.9) follows by letting $T \to \infty$ in (5.13) with $\Phi(s)$ in place of $1/s$.

Finally, the integral on the right-hand side of (7.9) is estimated using (7.7) and analogues of Theorem 6.3 for $F(s)$. We give the necessary details in the second example below.

Two examples

We demonstrate how to use the Perron inversion formula in practice by employing it to study the partial sums of τ_k and to count square-full integers.

Averaging the divisor functions τ_k. We wish to estimate the summatory function of the kth divisor function τ_k, where $k \in \mathbb{Z}_{\geqslant 2}$. We have that

$$\sum_{n=1}^{\infty} \frac{\tau_k(n)}{n^s} = \zeta(s)^k,$$

which has a meromorphic continuation to $\mathbb{C}$ with its only singularity being a pole of order k at $s = 1$. Since $\tau_k(n) \leqslant n^{O_k(1/\log\log n)}$ for $n \geqslant 3$ by Exercise 2.9(f), we fix $\varepsilon > 0$ and apply Theorem 7.2 with $\theta = \varepsilon/2$, $A = 0$ and $\alpha = 1 + \varepsilon$. We have $1 + \varepsilon \geqslant \theta + 1 + 1/\log x$ for $x \geqslant e^{2/\varepsilon}$, whence

$$\sum_{n \leqslant x} \tau_k(n) = \frac{1}{2\pi i} \int_{\substack{\mathrm{Re}(s)=1+\varepsilon \\ |\mathrm{Im}(s)| \leqslant T}} \zeta(s)^k \, \frac{x^s}{s} \mathrm{d}s + O_{k,\varepsilon}\Big(\frac{x^{1+\varepsilon}\log x}{T}\Big) \tag{7.10}$$

uniformly for $x \geqslant T \geqslant 2$ and $x \geqslant e^{2/\varepsilon}$. We will apply (5.13) to replace the contour $[1 + \varepsilon - iT, 1 + \varepsilon + iT]$ by $C_1 + C_2 + C_3$, where

$$C_1 = [1+\varepsilon-iT, \alpha'-iT], \quad C_2 = [\alpha'-iT, \alpha'+iT], \quad C_3 = [\alpha'+iT, 1+\varepsilon+iT]$$

for some $\alpha' \in (0, 1)$ to be chosen later. All implied constants in what follows might depend on α', ε and k.

Consider the rectangle whose vertices are the points $1 + \varepsilon \pm iT$ and $\alpha' \pm iT$. The only pole of the integrand in (7.10) inside this rectangle is at $s = 1$ and has order k. Consequently,

$$\sum_{n \leqslant x} f(n) = \mathrm{res}_{s=1} \frac{x^s \zeta(s)^k}{s} + \frac{1}{2\pi i} \int_{C_1+C_2+C_3} \zeta(s)^k \, \frac{x^s}{s} \mathrm{d}s + O\Big(\frac{x^{1+\varepsilon}\log x}{T}\Big).$$

Remark 5.3 implies that there is a polynomial P_k of degree $k - 1$ and with leading coefficient $1/(k-1)!$ such that

$$\mathrm{res}_{s=1}(x^s\zeta(s)^k/x) = x \cdot P_k(\log x).$$

We treat the integral over $C_1 + C_2 + C_3$ as an error term and bound it crudely. Using Theorem 6.3, we have $\zeta(s) \ll_\varepsilon (1+|t|)^{(1-\sigma+\varepsilon)/2}$ for $|s-1| \gg 1$ and $0 \leqslant \sigma \leqslant 1 + \varepsilon$. Since we also have $|\alpha' + it| \asymp 1 + |t|$ for $\alpha' > 0$, we find

$$\begin{aligned} \int_{C_2} \zeta(s)^k \, \frac{x^s}{s} \mathrm{d}s &= i \int_{-T}^{T} \zeta(\alpha' + it)^k \, \frac{x^{\alpha'+it}}{\alpha' + it} \mathrm{d}t \\ &\ll \int_{-T}^{T} (1 + |t|)^{(1-\alpha'+\varepsilon)k/2-1} x^{\alpha'} \mathrm{d}t \\ &\ll x^{\alpha'} T^{(1-\alpha'+\varepsilon)k/2}. \end{aligned}$$

Similarly, we have that

$$\int_{C_1} \zeta(s)^k \frac{x^s}{s} \mathrm{d}s = \int_{\alpha'}^{1+\varepsilon} \zeta(\sigma - iT)^k \frac{x^{\sigma - iT}}{\sigma - iT} \mathrm{d}\sigma \ll \int_{\alpha'}^{1+\varepsilon} T^{(1-\sigma+\varepsilon)k/2-1} x^\sigma \mathrm{d}\sigma$$

$$\ll \max_{\alpha' \leqslant \sigma \leqslant 1+\varepsilon} T^{(1-\sigma+\varepsilon)k/2-1} x^\sigma, \tag{7.11}$$

and the same estimate holds for the integral over C_3. Assuming that $T \leqslant x^{2/k}$, the maximum in (7.11) occurs when $\sigma = 1 + \varepsilon$.

To conclude, we have proved that

$$\sum_{n \leqslant x} \tau_k(n) = x P_k(\log x) + O_{k,\varepsilon,\alpha'}\Big(x^{\alpha'} T^{(1-\alpha'+\varepsilon)k/2} + \frac{x^{1+\varepsilon} \log x}{T}\Big).$$

The error term increases when α' increases. Taking $\alpha' = \varepsilon$ yields

$$\sum_{n \leqslant x} \tau_k(n) = x P_k(\log x) + O_{k,\varepsilon}\big(x^\varepsilon (T^{k/2} + x/T) \log x\big).$$

We optimize this estimate by taking $T = x^{2/(k+2)}$. Replacing ε by $\varepsilon/2$, we arrive at the following result.

Theorem 7.4. *Fix $k \in \mathbb{Z}_{\geqslant 2}$ and $\varepsilon > 0$. There is a polynomial P_k of degree $k-1$ and of leading coefficient $1/(k-1)!$ such that*

$$\sum_{n \leqslant x} \tau_k(n) = x P_k(\log x) + O_{k,\varepsilon}(x^{k/(k+2)+\varepsilon}) \quad (x \geqslant 1).$$

Remark 7.5. Theorem 7.4 improves upon Exercise 3.10 when $k \geqslant 3$, while yielding a slightly weaker version of Theorem 3.3 when $k = 2$. Since in its proof we took α' very close to 0, it is tempting to examine what would happen had we chosen $\alpha' < 0$. The calculation is a bit different now, since $|\zeta(\alpha' + it)| \asymp_{\alpha'} |t|^{1/2-\alpha'}$ for $\alpha' < 0$ and $|t| \geqslant 1$. It turns out that this idea does not lead to an improvement of Theorem 7.4. We leave the verification of this claim as an exercise. □

Square-full integers. As in Example 5.5, let f be the characteristic function of square-full integers and

$$F(s) = \sum_{n \geqslant 1} \frac{f(n)}{n^s} = \frac{\zeta(2s)\zeta(3s)}{\zeta(6s)}.$$

Theorem 7.2 with $\theta = A = 0$ and $\alpha = 1 + 1/\log x$ implies that

$$\sum_{n \leqslant x} f(x) = \frac{1}{2\pi i} \int_{\substack{\mathrm{Re}(s) = 1 + 1/\log x \\ |\mathrm{Im}(s)| \leqslant T}} \frac{\zeta(2s)\zeta(3s)}{\zeta(6s)} \cdot \frac{x^s}{s} \mathrm{d}s + O\Big(\frac{x \log x}{T}\Big)$$

uniformly for $x \geqslant T \geqslant 2$. This formula puts us right away at a disadvantage: the error term should really be of size $O(x^{1/2+\varepsilon}/T)$, because $\sum_{n \leqslant x} f(n) \asymp \sqrt{x}$, that is to say, the parameter θ is $-1/2$ *on average*. We could prove a

more general version of Theorem 7.2 that would allow such an improvement, but it is significantly simpler to work instead with a smooth cut-off ϕ.

Let $T \geqslant 2$ be a parameter that we will choose later. By Example 7.3, there are functions $\phi^{\pm} \in C^{\infty}(\mathbb{R}_{\geqslant 0})$ such that

$$1_{[0,1-1/T]} \leqslant \phi^{-} \leqslant 1_{[0,1]} \leqslant \phi^{+} \leqslant 1_{[0,1+1/T]},$$

and $\|(\phi^{\pm})^{(k)}\|_{\infty} \ll_k T^k$ for each fixed k. Then

$$\sum_{n\geqslant 1} f(n)\phi^{-}(n/x) \leqslant \sum_{n\leqslant x} f(n) \leqslant \sum_{n\geqslant 1} f(n)\phi^{+}(n/x). \tag{7.12}$$

Let $\phi \in \{\phi^{-}, \phi^{+}\}$, and let Φ denote its Mellin transform, which satisfies (7.7). We use relations (7.8) and (7.9) to find that

$$\begin{aligned}\sum_{n\geqslant 1} f(n)\phi(n/x) = \frac{\zeta(3/2)}{\zeta(3)}\cdot\frac{\Phi(1/2)}{2}x^{1/2} + \frac{\zeta(2/3)}{\zeta(2)}\cdot\frac{\Phi(1/3)}{3}x^{1/3} \\ + \frac{1}{2\pi i}\int_{(\alpha')}\frac{\zeta(2s)\zeta(3s)}{\zeta(6s)}\Phi(s)x^s\mathrm{d}s\end{aligned} \tag{7.13}$$

for any $\alpha' \in (1/6, 1/3)$. To ease notation, let α' be α from now on.

To estimate the integral over the line $\mathrm{Re}(s) = \alpha$, we use Theorem 6.3 to find that

$$\zeta(2s)\zeta(3s) \ll_{\varepsilon,\alpha} (1+|t|)^{1/2-\alpha+\varepsilon/2}(1+|t|)^{1/2-3\alpha/2+\varepsilon/2} = (1+|t|)^{1-5\alpha/2+\varepsilon}$$

for any fixed $\varepsilon \in (0, 1/2]$. Finally, since $\mathrm{Re}(6s) = 6\alpha > 1$, we have $|\zeta(6s)| \geqslant 1/\zeta(6\alpha) \geqslant_{\alpha} 1$ from relation (6.9). Together with (7.7), this implies that

$$\frac{\zeta(2s)\zeta(3s)}{\zeta(6s)}\Phi(s)x^s \ll_{\varepsilon,\alpha,k} \frac{(1+|t|)^{-5\alpha/2+\varepsilon}x^{\alpha}}{\max\{1, |t|/T\}^k}$$

for any fixed $k \geqslant 0$. We use the above inequality to bound the integral in (7.13): we have

$$\int_{\substack{\mathrm{Re}(s)=\alpha\\ |\mathrm{Im}(s)|\leqslant T}} \frac{\zeta(2s)\zeta(3s)}{\zeta(6s)}\Phi(s)x^s\mathrm{d}s \ll \int_{-T}^{T}(1+|t|)^{-5\alpha/2+\varepsilon}x^{\alpha}\mathrm{d}t \ll T^{1-5\alpha/2+\varepsilon}x^{\alpha},$$

since we have assumed that $\alpha < 1/3$. On the other hand,

$$\int_{\substack{\mathrm{Re}(s)=\alpha\\ |\mathrm{Im}(s)|\geqslant T}} \frac{\zeta(2s)\zeta(3s)}{\zeta(6s)}\Phi(s)x^s\mathrm{d}s \ll \int_{|t|\geqslant T}\frac{|t|^{-5\alpha/2+\varepsilon}x^{\alpha}}{(|t|/T)^2}\mathrm{d}t \ll T^{1-5\alpha/2+\varepsilon}x^{\alpha},$$

assuming that $\varepsilon \leqslant 1/2$. Inserting these bounds into (7.13), we conclude that

$$\begin{aligned}\sum_{n\geqslant 1} f(n)\phi(n/x) = \frac{\zeta(3/2)}{\zeta(3)}\cdot\frac{\Phi(1/2)}{2}x^{1/2} + \frac{\zeta(2/3)}{\zeta(2)}\cdot\frac{\Phi(1/3)}{3}x^{1/3} \\ + O_{\alpha,\varepsilon}(x^{\alpha}T^{1-5\alpha/2+\varepsilon}).\end{aligned}$$

Finally, since $1_{[0,1-1/T]} \leqslant \phi \leqslant 1_{[0,1+1/T]}$, we have that

$$\Phi(s) = \int_0^1 y^{s-1}\mathrm{d}s + O(1/T) = 1/s + O(1/T)$$

for $s \in \{1/2, 1/3\}$, so that

$$\sum_{n\geqslant 1} f(n)\phi(n/x) = \frac{\zeta(3/2)}{\zeta(3)}x^{1/2} + \frac{\zeta(2/3)}{\zeta(2)}x^{1/3} + O\bigg(\frac{x^{1/2}}{T} + x^{\alpha}T^{1-5\alpha/2+\varepsilon}\bigg)$$

for $\phi = \phi^{\pm}$. We optimize the error term by taking $\alpha = 1/6 + \varepsilon/2$ and $T = x^{4/19}$. Together with (7.12) this implies the estimate

$$\sum_{n\leqslant x} f(n) = \frac{\zeta(3/2)}{\zeta(3)}x^{1/2} + \frac{\zeta(2/3)}{\zeta(2)}x^{1/3} + O(x^{11/38+\varepsilon}).$$

This recovers the main terms of Exercise 3.11, but has a worse error term.

We thus see that even though using Perron's inversion formula offers an intuitive way of establishing asymptotic formulas, it is sometimes possible to prove superior results using more elementary methods. Hence, it is important to be fluent in both ways of approaching a problem.

Truncating Perron's integral

We conclude the chapter by proving Lemma 7.1 and Theorem 7.2.

Proof of Lemma 7.1. First, we consider the case $0 < y < 1$. We fix a large $A > 0$ and apply (5.13) with $F(s) = 1$ and $\alpha' = -A$ to find that

$$\frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(s)=\alpha\\ |\mathrm{Im}(s)|\leqslant T}} \frac{(1/y)^s}{s}\mathrm{d}s = 1 + \frac{1}{2\pi i}(I_{-1} + I_0 + I_1), \tag{7.14}$$

where I_{-1} is the integral of y^{-s}/s over the line segment $[\alpha - iT, -A - iT]$, I_0 is over $[-A-iT, -A+iT]$ and I_1 is over $[-A+iT, \alpha+iT]$. What is important in the above formula is that the integrand is very small on the new contour of integration: either the denominator is large (in $I_{\pm 1}$, that are supported on the horizontal line segments $[-A \pm iT, \alpha \pm iT]$), or the numerator is small (in I_0, that is supported on the vertical line segment $[-A - iT, -A + iT]$) because $y > 1$. More concretely,

$$I_{\pm 1} = \int_{\alpha\pm iT}^{-A\pm iT} \frac{(1/y)^s}{s}\mathrm{d}s \ll \int_{-A}^{\alpha} \frac{(1/y)^{\sigma}}{|\sigma| + T}\mathrm{d}\sigma \leqslant \frac{1}{T}\int_{-A}^{\alpha} y^{-\sigma}\mathrm{d}\sigma \leqslant \frac{y^{-\alpha}}{T|\log y|},$$

whereas

$$I_0 = \int_{-A-iT}^{-A+iT} \frac{(1/y)^s}{s}\mathrm{d}s \ll \int_{-T}^{T} \frac{y^A}{A + |t|}\mathrm{d}t.$$

Letting $A \to \infty$, we have $I_0 \to 0$. This proves the lemma when $y \leqslant e^{-1/T}$.

When $e^{-1/T} \leqslant y < 1$, we use a variation of (5.13): we replace the line of integration $L = \{ s \in \mathbb{C} : \sigma = \alpha, |t| \leqslant T \}$ by the circular arc $C = \{ s \in \mathbb{C} : |s| = \sqrt{\alpha^2 + T^2}, \sigma \leqslant \alpha \}$ traversed clockwise. As in (7.14), Cauchy's theorem implies that $\int_L (y^{-s}/s)\mathrm{d}s = \int_C (y^{-s}/s)\mathrm{d}s + 2\pi i$. To bound the integral over C, we note that $|y^{-s}/s| \ll y^{-\alpha}/T$ for $s \in C$. Since the length of C is $\asymp T$, the lemma follows in this case too.

The case $y > 1$ is similar. However, instead of shifting the contour to the left, we shift it to the right, so that y^{-s}/s becomes smaller in magnitude. No pole is encountered this time. We leave the details as an exercise.

It remains to handle the case $y = 1$. We argue by direct computation:

$$\frac{1}{2\pi i} \int_{\substack{\mathrm{Re}(s)=\alpha \\ |\mathrm{Im}(s)|\leqslant T}} \frac{\mathrm{d}s}{s} = \frac{1}{2\pi} \int_0^T \Big(\frac{1}{\alpha + it} + \frac{1}{\alpha - it} \Big) \mathrm{d}t = \frac{\alpha}{\pi} \int_0^T \frac{\mathrm{d}t}{\alpha^2 + t^2}.$$

The rightmost integral equals $\arctan(T/\alpha)/\pi = 1/2 + O(\alpha/T)$, which completes the proof of the lemma. □

Proof of Theorem 7.2. Using the fact that $f(n) \leqslant Cn^\theta (1 + \log n)^A$ and (7.2), we have

$$\sum_{n \leqslant x} f(n) = \frac{1}{2\pi i} \int_{\substack{\mathrm{Re}(s)=\alpha \\ |\mathrm{Im}(s)|\leqslant T}} F(s) \frac{x^s}{s} \mathrm{d}s + O(x^\alpha \cdot E + x^\theta (\log x)^A),$$

where

$$E = \sum_{n \geqslant 1} \frac{n^\theta (1 + \log n)^A}{n^\alpha \max\{1, T|\log \frac{x}{n}|\}} \leqslant \sum_{n \geqslant 1} \frac{(1 + \log n)^A}{n^{1 + 1/\log x} \max\{1, T|\log \frac{x}{n}|\}}.$$

We write $E = E_1 + E_2 + E_3 + E_4$, where E_1 is the part of the sum with $|x - n| \leqslant 1$, E_2 is with $1 < |x - n| \leqslant x/T$, E_3 is with $\max\{1, x/T\} < |x - n| \leqslant x/2$ and E_4 is with $|x - n| > x/2$.

We clearly have that $E_1 \ll (\log x)^A/x$. The sum E_2 has non-empty range only when $x \geqslant T$, in which case we have

$$E_2 \leqslant \sum_{x - x/T \leqslant n \leqslant x + x/T} \frac{(1 + \log n)^A}{n^{1 + 1/\log x}} \asymp \frac{x}{T} \cdot \frac{(\log x)^A}{x}.$$

For the terms in the range of E_3, we note that $|\log(x/n)| \asymp |n - x|/x$, by Taylor's expansion of the logarithm about 1, so that

$$E_3 \ll \sum_{\max\{1, x/T\} \leqslant |n - x| \leqslant x/2} \frac{(\log x)^A}{T \cdot |n - x|} \leqslant \sum_{1 \leqslant 2^j \leqslant x/2} \sum_{2^j \leqslant |x - n| < 2^{j+1}} \frac{(\log x)^A}{T \cdot 2^j}.$$

Since there are $\ll 2^j$ integers n with $2^j \leqslant |x - n| < 2^{j+1}$, we deduce that $E_3 \ll (\log x)^{A+1}/T$. Finally, for the terms in the range of E_4 we note that

$|\log(x/n)| \gg 1$, whence

$$E_4 \ll \frac{1}{T}\sum_{n\geqslant 1}\frac{(1+\log n)^A}{n^{1+1/\log x}} \ll_A \frac{(\log x)^{A+1}}{T}$$

by an application of the Euler-Maclaurin summation formula (Theorem 1.10). (Alternatively, note that the contribution of $n \in [x^j, x^{j+1})$ to E_4 is $\ll T^{-1}(\log x)^{A+1}(j+1)^A e^{-j}$. Summing over all $j \geqslant 0$ proves the claimed bound.) Putting together the above estimates implies that $E \ll (\log x)^{A+1}/T + (\log x)^A/x$, thus completing the proof of the theorem. □

Exercises

Exercise 7.1. Consider an arithmetic function f with Dirichlet series $F(s)$ converging absolutely for $\mathrm{Re}(s) = \alpha > 0$. Prove that

$$\frac{1}{x}\int_0^x \sum_{n\leqslant y} f(n)\mathrm{d}y = \sum_{n\leqslant x} f(n)(1-n/x) = \frac{1}{2\pi i}\int_{(\alpha)} F(s)\frac{x^s}{s(s+1)}\mathrm{d}s$$

and

$$\int_1^x \sum_{n\leqslant y} f(n)\frac{\mathrm{d}y}{y} = \sum_{n\leqslant x} f(n)\log(x/n) = \frac{1}{2\pi i}\int_{(\alpha)} F(s)\frac{x^s}{s^2}\mathrm{d}s.$$

[*Hint:* Mellin inversion for $\phi(y) = 1_{y\leqslant 1}\cdot(1-y)$ and $\psi(y) = 1_{y\leqslant 1}\cdot\log(1/y)$.]

Exercise 7.2. Let $\phi \in C^\infty(\mathbb{R}_{\geqslant 0})$ be compactly supported, and let Φ denote its Mellin transform.

(a) Show that $\Phi(s)$ is analytic for $\mathrm{Re}(s) > 0$.

(b) If $\phi(y) = \phi_0$ when $y \in [0,1]$, show that Φ has a meromorphic continuation to $\mathbb{C}$ whose only singularity is a simple pole of residue ϕ_0 at $s = 0$.

(c) If $\mathrm{supp}(\phi) \subseteq [0,m]$, then prove that $\Phi(s) \ll_{\phi,A} m^{\max\{\sigma,0\}}/(1+|s|)^A$ for all $s \in \mathbb{C}$ and any fixed $A \geqslant 1$.

(d) Let f be an arithmetic function with Dirichlet series $F(s)$ converging absolutely for $\mathrm{Re}(s) = \alpha > 0$. For $x \geqslant 1$, prove that

$$\sum_{n\geqslant 1} f(n)\phi(n/x) = \frac{1}{2\pi i}\int_{(\alpha)} F(s)\Phi(s)x^s\mathrm{d}s.$$

Exercise 7.3. Let $f(n) = \mu^2(n)/\varphi(n)$ and let $F(s)$ be its Dirichlet series.

(a) Prove that $F(s) = \zeta(s+1)G(s)$, where $G(s)$ is a Dirichlet series that converges absolutely for $\mathrm{Re}(s) > -1/2$.

(b) Write G as an Euler product and calculate its logarithmic derivative G'/G. Deduce that $G(0) = 1$ and $G'(0) = \sum_p (\log p)/(p^2-p)$.

(c) For $1 \leqslant T \leqslant x$ and $\alpha \in (-1/2, 0)$, prove that

$$\sum_{n\leqslant x} f(n) = \frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(s)=1/\log x\\ |\,\mathrm{Im}(s)|\leqslant T}} F(s)\frac{x^s}{s}\mathrm{d}s + O((\log x)^2/T)$$
$$= \log x + \gamma + G'(0) + \frac{1}{2\pi i}\int_C F(s)\frac{x^s}{s}\mathrm{d}s + O((\log x)^2/T),$$

where C is the sum of the contours $[1/\log x - iT, \alpha - iT]$, $[\alpha - iT, \alpha + iT]$ and $[\alpha + iT, 1/\log x + iT]$.

(d) Use Theorem 6.3 to estimate the integral over C and deduce that

$$\sum_{n\leqslant x} f(n) = \log x + \gamma + G'(0) + O_\varepsilon(x^{-2/5+\varepsilon}) \qquad (x \geqslant 1)$$

for any fixed $\varepsilon > 0$.

Exercise 7.4. Use Theorem 7.2 and its variants to estimate $\sum_{n\leqslant x}\log n$, $\sum_{n\leqslant x}\mu^2(n)$ and $\sum_{n\leqslant x}\varphi(n)$. Compare your results with those obtained by Theorem 1.12, Exercise 3.8 and Theorem 3.2, respectively.

Exercise 7.5. An integer n is called cube-free if there is no prime p such that $p^3|n$. On the other hand, an integer is called cube-full if $p^3|n$ whenever $p|n$. Estimate the number of cube-free and cube-full integers in $[1, x]$.

Exercise 7.6. Show that there is a linear polynomial L, a quadratic polynomial Q and some $\delta > 0$ such that

$$\sum_{n\leqslant x} 2^{\omega(n)} = x\cdot L(\log x) + O(x^{1-\delta}) \qquad (x \geqslant 1)$$

and

$$\sum_{n\leqslant x} 2^{\Omega(n)} = x\cdot Q(\log x) + O(x^{1-\delta}) \qquad (x \geqslant 1).$$

Exercise 7.7.* Let $s = \sigma + it$ with $0 < \sigma \leqslant 1$ and $|t| \geqslant 2$.

(a) For $x \geqslant T \geqslant 2$ and $\alpha = 1 - \sigma + 1/\log x$, show that

$$\sum_{n\leqslant x}\frac{1}{n^s} = \frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(z)=\alpha\\ |\,\mathrm{Im}(z)|\leqslant T}} \zeta(s+z)\frac{x^z}{z}\mathrm{d}z + O\Big(\frac{x^{1-\sigma}\log x}{T}\Big).$$

(b) If $|\,\mathrm{Im}(z)| \leqslant |t|/2$ and $\varepsilon > 0$, then show that

$$\zeta(s+z) \ll_\varepsilon \begin{cases} |t|^{(1-\sigma-\mathrm{Re}(z))/2+\varepsilon} & \text{if } -\sigma \leqslant \mathrm{Re}(z) \leqslant \alpha,\\ |t|^{1/2-\sigma-\mathrm{Re}(z)+\varepsilon} & \text{if } \mathrm{Re}(z) \leqslant -\sigma. \end{cases}$$

(c) Let $\varepsilon > 0$ and $A \geqslant 1$ be fixed. Uniformly for $x \geqslant |t|$, prove that

$$\sum_{n\leqslant x}\frac{1}{n^s} = \zeta(s) + O_{A,\varepsilon}\big(|t|^{1/2-\sigma+\varepsilon}(|t|/x|)^A + x^{1-\sigma}|t|^{\varepsilon-1}\big). \tag{7.15}$$

Exercise 7.8.* Let ϕ be as in Exercise 7.2(b) with $\phi_0 = 1$.

(a) Consider $s \in \mathbb{C}$ with $0 \leqslant \sigma < 1$ and $|t| \geqslant 100$, as well as $x, y \geqslant 1$ with $xy = |t|/2\pi$. Show that

$$\sum_{n\geqslant 1} \frac{\phi(n/x)}{n^s} = x^{1-s}\Phi(1-s) + \zeta(s) + \frac{1}{2\pi i}\int_{(-3)} \zeta(s+z)\Phi(z)x^z \mathrm{d}z.$$

(b) Use (6.2) to write $\zeta(s+z) = \lambda(s+z)\zeta(1-s-z)$ and deduce the *approximate functional equation*

$$\zeta(s) = \sum_{n\geqslant 1} \frac{\phi(n/x)}{n^s} + \sum_{n\geqslant 1} \frac{\phi_s^*(n/y)}{n^{1-s}} + O_{A,\phi}(|t|^{-A}), \tag{7.16}$$

where

$$\phi_s^*(u) := -\frac{1}{2\pi i}\int_{(-3)} \Phi(z)\lambda(s+z)(u|t|/2\pi)^z \mathrm{d}z.$$

(c) Show that λ is meromorphic with its poles located at the odd positive integers with $\operatorname{res}_{z=2n+1} \lambda(z) = (-1)^{n-1}2^{2n+1}\pi^{2n}/(2n)!$. Finally, use Exercise 1.12 to show that $\lambda(a+ib) \ll_a \max\{|b|, 1\}^{1/2-a}$ when $a, b \in \mathbb{R}$ are such that $|a+ib-k| \geqslant 1/2$ for $k = 1, 3, 5, \ldots$.

(d) Let $\alpha \in (-\infty, 3/2] \setminus \{0, 1-\sigma\}$. Prove that

$$\phi_s^*(u) = -\frac{1}{2\pi i}\int_{(\alpha)} \Phi(z)\lambda(s+z)(u|t|/2\pi)^z \mathrm{d}z + 1_{\alpha>0}\cdot\lambda(s) - 1_{\alpha>1-\sigma}\cdot 2\Phi(1-s)(u|t|/2\pi)^{1-s}.$$

(e) Fix $\varepsilon > 0$ and $B > 0$. Show that

$$\phi_s^*(u) \ll_{\varepsilon,B} \begin{cases} u^{-B} & \text{if } u \geqslant |t|^\varepsilon, \\ |t|^{1/2-\sigma+3\varepsilon/2} & \text{if } 0 \leqslant u \leqslant |t|^\varepsilon. \end{cases}$$

[*Hint:* When $u \geqslant |t|^\varepsilon$, take $\alpha = -C$ in part (d) with C big enough in terms of ε and B. Otherwise, take $\alpha = 3/2$.]

(f) Combine parts (b) and (e) to prove that $\zeta(s) \ll_\varepsilon |t|^{(1-\sigma)/2+\varepsilon}$ for $|t| \geqslant 1$ and $0 \leqslant \sigma \leqslant 1$, thus recovering Theorem 6.3 inside the critical strip.

Exercise 7.9.*

(a) Let $z = a + ib$ with $a, b \in \mathbb{R}$ such that $|a| \ll |b|^{2/3}$ and $|z-k| \geqslant 1/2$ for $k = 1, 3, 5, \ldots$. Show that

$$|\lambda(z)| \asymp (2\pi)^a \max\{|b|, 1\}^{1/2-a}.$$

[*Hint:* Use the relations $\overline{\lambda}(z) = \lambda(\overline{z})$ and $\lambda(z) = 1/\lambda(1-z)$ to first reduce to the case when $b \geqslant 0$ and $a \geqslant 1/2$. When $a \geqslant 1/2$ and $b \geqslant 1$, use Exercise 1.12 noticing that $|z/2|^a = (b/2)^a(1 + O((a/b)^2))^a$ and $\arg(z/2) = \pi/2 - a/b + O((a/b)^3)$ for $a \ll b$. Similar estimates also hold for $|(1-z)/2|^a$ and $\arg((1-z)/2)$.]

(b) For $s \in \mathbb{C}$ with $0 \leqslant \sigma \leqslant 1$ and $|t| \geqslant 100$, and for $x, y \geqslant 1$ with $xy = |t|/2\pi$, show that

$$\zeta(s) = \sum_{n\leqslant x} \frac{1}{n^s} + \lambda(s)\sum_{n\leqslant y} \frac{1}{n^{1-s}} + O_\varepsilon((x^{-\sigma} + |t|^{1/2-\sigma}y^{\sigma-1})|t|^\varepsilon). \tag{7.17}$$

[*Hint:* Consider $\phi \in C^\infty(\mathbb{R}_{\geqslant 0})$ with Mellin transform Φ. Assume (7.4) and (7.7) with $T = |t|^{1/2-\varepsilon}$. For $z = a + ib$ with $|z| \geqslant |t|^{1/2} \geqslant |a|$ and $|b| \leqslant |t|/2$, show that

$$\Phi(z)\lambda(s+z)(|t|/2\pi)^z \ll_{\varepsilon,A} |t|^{-A}|1 + b/t|^{-a}(1+1/T)^{\max\{0,a\}}.$$

When $u \leqslant 1 - 2/T$, use this estimate to show that

$$\begin{aligned}\phi_s^*(u) &= \lambda(s) - \frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(z)=3/2\\ |\mathrm{Im}(s)|\leqslant|t|^{1/2}}} \Phi(z)\lambda(s+z)(u|t|/2\pi)^z \mathrm{d}z + O_A(|t|^{-A}) \\ &= \lambda(s) - \frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(z)=|t|^{1/2}\\ |\mathrm{Im}(s)|\leqslant|t|^{1/2}}} \Phi(z)\lambda(s+z)(u|t|/2\pi)^z \mathrm{d}z + O_A(|t|^{-A}),\end{aligned}$$

since $\Phi(z)\lambda(s+z)$ has no poles when $\mathrm{Re}(z) \geqslant 3/2$ and $|\mathrm{Im}(z)| \leqslant |t|^{1/2}$. Conclude that $\phi_s^*(u) = \lambda(s) + O_A(|t|^{-A})$. On the other hand, when $u \geqslant 1 + 2/T$, show that $\phi_s^*(u) \ll_A u^{-10}|t|^{-A}$ by moving the contour to the line $\mathrm{Re}(z) = -|t|^{1/2}$.]

Remark 7.6. When $0 < \mathrm{Re}(s) < 1$, formula (7.15) allows us to approximate $\zeta(s)$ accurately by a sum of $|t|^{1+\varepsilon}$ terms. On the other hand, taking $x = y = \sqrt{|t|/2\pi}$ in (7.17), we can write $\zeta(s)$ as a linear combination of two much shorter sums, each of length $\sqrt{|t|/2\pi}$. In particular,

$$\begin{aligned}\zeta(1/2+it) &= \sum_{n\leqslant\sqrt{|t|/2\pi}} \frac{1}{n^{1/2+it}} + e^{-2i\vartheta(t)} \sum_{n\leqslant\sqrt{|t|/2\pi}} \frac{1}{n^{1/2-it}} + O_\varepsilon(|t|^{-1/4+\varepsilon}) \\ &= e^{-i\vartheta(t)} \sum_{n\leqslant\sqrt{|t|/2\pi}} \frac{2\cos(\vartheta(t) - t\log n)}{\sqrt{n}} + O_\varepsilon(|t|^{-1/4+\varepsilon}), \qquad (7.18)\end{aligned}$$

where $\vartheta(t)$ is defined in Exercise 6.5. For this reason, formula (7.17) has significant applications. On a theoretical level, it is very useful when studying the value distribution and the moments of ζ. On a practical level, it allows us to calculate ζ fast inside the critical strip. Indeed, a variation of (7.18) was used by Riemann himself to calculate the first few non-trivial zeroes of ζ and verify they are on the line $\mathrm{Re}(s) = 1/2$. Riemann's exact variation of (7.18) was rediscovered by Siegel [164] when he was studying Riemann's handwritten notes at the University of Göttingen [155] and it is known today as the *Riemann-Siegel formula*. For a detailed discussion of this subject, see Chapter 7 of Edward's book on ζ [37]. □

Chapter 8

The Prime Number Theorem

Having developed the theory of the Riemann zeta function and of the Perron inversion, we use them to establish a quantitative version of the celebrated Prime Number Theorem.

Theorem 8.1. *There is a constant $c > 0$ such that*

$$\pi(x) = \mathrm{li}(x) + O\big(xe^{-c\sqrt{\log x}}\big) \qquad (x \geqslant 2).$$

Instead of working with $\pi(x)$, we work with Chebyshev's function $\psi(x) = \sum_{n \leqslant x} \Lambda(n)$. Our first goal is to establish the explicit formula (5.11). Subsequently, we will show that $\zeta(s) \neq 0$ when $\mathrm{Re}(s) \approx 1$. This will allow us to bound the sum over zeroes in (5.11) and obtain Theorem 8.1.

Proving the explicit formula

In order to use the techniques of the previous chapter, we need to control ζ'/ζ past its domain of absolute convergence. The key technical estimate is the following lemma, whose proof we postpone till the end of the chapter.

Lemma 8.2. *Let $s = \sigma + it \in \mathbb{C}$.*

(a) *There are $\ll \log(|t|+2)$ non-trivial zeroes $\rho = \beta + i\gamma$ of ζ with $|\gamma - t| \leqslant 1$, even when counted with multiplicity.*

(b) *If $|s + 2n| \geqslant 1/2$ for all $n \in \mathbb{N}$, then*

$$\frac{\zeta'}{\zeta}(s) = -\frac{1}{s-1} + \sum_{|\gamma - t| \leqslant 1} \frac{1}{s - \rho} + O(\log(|s| + 2))\,;$$

the sum runs over non-trivial zeroes of ζ listed with their multiplicity.

Proof of Theorem 5.1. Let $x, T \geqslant 2$. For technical reasons, we first prove the theorem when $|T-\gamma| \gg 1/\log T$ for all ρ. Our starting point is Theorem 7.2, which yields the formula

$$\psi(x)=\frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(s)=1+1/\log x\\ |\mathrm{Im}(s)|\leqslant T}}\left(-\frac{\zeta'}{\zeta}(s)\right)\frac{x^s}{s}\mathrm{d}s+O\left(\frac{x\log^2 x}{T}+\log x\right) \tag{8.1}$$

Next, we replace the contour of integration $[1+1/\log x-iT, 1+1/\log x-iT]$ by the contour $L_{-1}+L_0+L_1$, where we have set $L_{-1}=[1+1/\log x-iT, -2N-1-iT]$, $L_0=[-2N-1-iT, -2N-1+iT]$ and $L_1=[-2N-1+iT, 1+1/\log x+iT]$ for a fixed large integer $N \geqslant 1$. Indeed, relation (5.13) with $\alpha = 1+1/\log x$ and $\alpha' = -2N-1$ implies that

$$\begin{aligned}\psi(x) = &\sum_{-2N-1<\mathrm{Re}(w)<1+1/\log x}\mathrm{res}_{s=w}\left(\frac{(-\zeta'/\zeta)(s)x^s}{s}\right)\\ &+\frac{1}{2\pi i}\left(\int_{L_{-1}}+\int_{L_0}+\int_{L_1}\right)\frac{(-\zeta'/\zeta)(s)x^s}{s}\mathrm{d}s\\ &+O\left(\frac{x\log^2(xT)}{T}+\log x\right),\end{aligned}$$

where the sum over w runs over all poles of $f(s) := (-\zeta'/\zeta)(s)x^s/s$ in the rectangle formed by the points $1+1/\log x \pm iT$ and $-2N-1\pm iT$. Our next task is to locate all such poles.

The pole of ζ at $s=1$ induces a pole of f of residue x. Moreover, for each zero ρ of ζ of multiplicity m_ρ, we obtain a pole of residue $-m_\rho x^\rho/\rho$ (see relation (5.10) and the discussion preceding it). Finally, there is a pole at $s=0$ of residue $-\log(2\pi)$, and poles at $s=-2n \geqslant -2N$ of residue $x^{-2n}/(2n) \ll 4^{-n}$. Therefore

$$\begin{aligned}\psi(x) = x-\sum_{|\gamma|\leqslant T}\frac{x^\rho}{\rho}&+\frac{1}{2\pi i}\left(\int_{L_{-1}}+\int_{L_0}+\int_{L_1}\right)\frac{(-\zeta'/\zeta)(s)x^s}{s}\mathrm{d}s\\ &+O\left(\frac{x\log^2(xT)}{T}+\log x\right),\end{aligned} \tag{8.2}$$

where each zero ρ is repeated several times according to its multiplicity.

Next, we bound the contribution of the integrals over L_{-1} and L_1. On these integrals we have $|\mathrm{Im}(s)| = T$. Recall that we have assumed that $|T-\gamma| \gg 1/\log T$ for all γ. If $\beta+i\gamma$ is a zero of ζ, so is $\beta-i\gamma$, and we deduce that $|T+\gamma| \gg 1/\log T$. We thus find that $|s-\rho| \gg 1/\log T$ for $s \in L_{\pm 1}$ with $\sigma \geqslant -1$. Together with Lemma 8.2, this implies that

$$\frac{\zeta'}{\zeta}(s) \ll \log T + \sum_{|\gamma-t|\leqslant 1}\log T \ll \log^2 T$$

for all $s \in L_{\pm 1}$ with $\sigma \geqslant -1$. When $\sigma \leqslant -1$, we have the stronger bound $(\zeta'/\zeta)(s) \ll \log T$ since the distance of s to the zeroes of ζ is $\geqslant 1$. Consequently,

$$\int_{L_{\pm 1}} \frac{\zeta'}{\zeta}(s)\frac{x^s}{s}\mathrm{d}s \ll \int_{-1}^{1+1/\log x} \frac{(\log T)^2 x^\sigma}{T}\mathrm{d}\sigma + \int_{-2N-1}^{-1} \frac{(\log T)x^\sigma}{T} \ll \frac{x\log^2(xT)}{T}.$$

When $s \in L_0$, we have $|s+2n| \geqslant 1$ for all $n \in \mathbb{N}$ and $|s+\rho| \geqslant 2N+1$ for all ρ. Hence, $(\zeta'/\zeta)(s) \ll \log|s|$, which implies that

$$\int_{L_0} \frac{\zeta'}{\zeta}(s)\frac{x^s}{s}\mathrm{d}s \ll \int_{-T}^{T} \frac{(\log(N+|t|))x^{-2N-1}}{N+|t|}\mathrm{d}t = o_{N\to\infty}(1).$$

Inserting these estimates into (8.2) and letting $N \to \infty$ completes the proof of Theorem 5.1 when $|T-\gamma| \gg 1/\log T$ for all zeroes $\rho = \beta + i\gamma$ of ζ.

Finally, consider the general case. There are $\ll \log T$ zeroes of ζ in the horizontal strip $\{T \leqslant \mathrm{Im}(s) \leqslant T+1\}$. By the pigeonhole principle, there is $T' \in [T, T+1]$ such that $|T'-\gamma| \gg 1/\log T'$ for all zeroes. We then have

$$\psi(x) = x - \sum_{|\gamma|\leqslant T'} \frac{x^\rho}{\rho} + O\Big(\frac{x\log^2(xT')}{T'} + \log x\Big).$$

In addition, Lemma 8.2(b) implies that

$$\Big|\sum_{T\leqslant|\gamma|\leqslant T'} \frac{x^\rho}{\rho}\Big| \leqslant \sum_{T\leqslant|\gamma|\leqslant T+1} \Big|\frac{x^\rho}{\rho}\Big| \ll \frac{x\log T}{T}.$$

This proves Theorem 5.1 for all $T \geqslant 2$. □

A zero-free region and the Prime Number Theorem

In view of Theorem 5.1, the only ingredient missing from proving the Prime Number Theorem is showing that the terms x^ρ/ρ are small compared to the expected main term x. Since $|x^\rho| = x^\beta$, we need to prove that β is not too close to 1, namely that a certain region is *free of zeroes* of ζ. This is precisely the context of the next theorem.

Theorem 8.3. *There is a constant $c > 0$ such that $\zeta(s) \neq 0$ in the region*

$$\sigma \geqslant 1 - \frac{c}{\log(|t|+2)}.$$

Proof. Let $\rho_0 = \beta_0 + i\gamma_0$ be a non-trivial zero of ζ. We need to prove that

$$1 - \beta_0 > \frac{c}{\log(|\gamma_0|+2)}. \tag{8.3}$$

First of all, since ζ has a pole at 1, there is an absolute constant $\delta \in (0,1]$ such that $|\rho_0 - 1| \geqslant \delta$. In particular, if $|\gamma_0| < \delta/2$, then $1 - \beta_0 > \delta/2 \geqslant \delta/(4\log 2)$, so that (8.3) follows in this case provided that $c \leqslant \delta/4$, as we may certainly assume. For the rest of the proof, we assume that $|\gamma_0| \geqslant \delta/2$.

In order to explain the idea of the proof, we consider first the extreme case when $\beta_0 = 1$, i.e., $\zeta(1+i\gamma_0) = 0$. By the analyticity of ζ, we must have that $\zeta(\sigma + i\gamma_0) \sim a \cdot (\sigma - 1)$ for some $a \in \mathbb{C}$ as $\sigma \to 1$. On the other hand, we have $1/\zeta(\sigma + i\gamma_0) = \prod_p (1 - 1/p^{\sigma + i\gamma_0})$, and the only way this product can tend to infinity as $\sigma \to 1^+$ is if $p^{i\gamma_0}$ points towards -1 a lot of the time. But then $p^{2i\gamma_0} = (p^{i\gamma_0})^2$ should point often towards 1, thus implying that $\zeta(\sigma + 2i\gamma_0) = \prod_p (1 - 1/p^{\sigma + 2i\gamma_0})^{-1} \to \infty$ as $\sigma \to 1^+$, that is to say, ζ should have a pole at $1 + 2i\gamma_0$. This is impossible, since $\gamma_0 \neq 0$ here, and the only pole of ζ is at 1.

We formalize the above idea by introducing a family of metrics $\mathbb{D}_\sigma(\cdot,\cdot)$, $\sigma > 1$, on the set of multiplicative functions taking values in the unit circle that we define by

$$\mathbb{D}_\sigma(f,g)^2 = \frac{1}{2} \sum_p \sum_{m=1}^{\infty} \frac{|f(p^m) - g(p^m)|^2 \log p}{p^{m\sigma}}. \tag{8.4}$$

We think of σ as a parameter that will be optimized later in terms of γ_0.

By the triangle inequality,

$$\begin{aligned} \mathbb{D}_\sigma(1, n^{2i\gamma_0}) &= \mathbb{D}_\sigma(n^{-i\gamma_0}, n^{i\gamma_0}) \\ &\leqslant \mathbb{D}_\sigma(n^{-i\gamma_0}, \mu(n)) + \mathbb{D}_\sigma(\mu(n), n^{i\gamma_0}) \\ &= 2\mathbb{D}_\sigma(\mu(n), n^{i\gamma_0}). \end{aligned} \tag{8.5}$$

The above inequality is a rigorous way to see that if $p^{i\gamma_0} \sim -1$ on average, then $p^{2i\gamma_0} \sim 1$. We will prove though that $\mathbb{D}_\sigma(1, n^{2i\gamma_0})$ cannot be too small because ζ is analytic around $1 + 2i\gamma_0$, whereas a zero $\rho_0 = \beta_0 + i\gamma_0$ too close to $1 + i\gamma_0$ would make $\mathbb{D}_\sigma(\mu(n), n^{i\gamma_0})$ rather small.

We start by proving a lower bound on $\mathbb{D}_\sigma(1, n^{2i\gamma_0})$. Since $|1 - p^{it}|^2 = 2(1 - \mathrm{Re}(p^{-it}))$ for $t \in \mathbb{R}$, we have that

$$\mathbb{D}_\sigma(1, n^{2i\gamma_0})^2 = \sum_{p,m} \frac{(1 - \mathrm{Re}(p^{-2i\gamma_0 m}))\log p}{p^{m\sigma}} = -\frac{\zeta'}{\zeta}(\sigma) + \mathrm{Re}\Big(\frac{\zeta'}{\zeta}(\sigma + 2i\gamma_0)\Big).$$

We evaluate the last term on the right side using Lemma 8.2(b). Since $|\gamma_0| \geqslant \delta/2$, we have $|s + 2i\gamma_0 - 1| \gg 1$. In addition, we have

$$\mathrm{Re}\Big(\frac{1}{\sigma + 2i\gamma_0 - \rho}\Big) = \frac{\sigma - \beta}{(\sigma - \beta)^2 + (2\gamma_0 - \gamma)^2} \geqslant 0$$

for each $\rho = \beta + i\gamma$. As a consequence,

$$\operatorname{Re}\Big(\frac{\zeta'}{\zeta}(\sigma + 2i\gamma_0)\Big) \geqslant -O(\log(|\gamma_0| + 2)).$$

Since we also have that $(-\zeta'/\zeta)(\sigma) = 1/(\sigma - 1) + O(1)$, we conclude that

$$\mathbb{D}_\sigma(1, n^{2i\gamma_0})^2 \geqslant \frac{1}{\sigma - 1} - O(\log(|\gamma_0| + 2)).$$

Next, we deal with $\mathbb{D}_\sigma(\mu(n), n^{i\gamma_0})$. Similarly to before, we have that

$$\begin{aligned}\mathbb{D}_\sigma(\mu(n), n^{i\gamma_0})^2 &= -\frac{\zeta'}{\zeta}(\sigma) - \operatorname{Re}\Big(\frac{\zeta'}{\zeta}(\sigma + i\gamma_0)\Big) + O(1)\\ &= \frac{1}{\sigma - 1} - \sum_{|\gamma - \gamma_0| \leqslant 1} \frac{\sigma - \beta}{(\sigma - \beta)^2 + (\gamma_0 - \gamma)^2} + O(\log(|\gamma_0| + 2))\\ &\leqslant \frac{1}{\sigma - 1} - \frac{1}{\sigma - \beta_0} + O(\log(|\gamma_0| + 2))\end{aligned}$$

by dropping all the summands except for the one with $\rho = \rho_0$.

Combining the above estimates, we find that

$$\frac{1}{\sigma - 1} \leqslant \frac{4}{\sigma - 1} - \frac{4}{\sigma - \beta_0} + O(\log(|\gamma_0| + 2)).$$

If $\sigma = 1 + 1/(C\log(|\gamma_0| + 2))$ for some large enough C, we have

$$\frac{3.5}{\sigma - 1} > \frac{4}{\sigma - \beta_0}, \qquad \text{whence} \qquad 1 - \beta_0 > \frac{\sigma - 1}{7} = \frac{1}{7C\log(|\gamma_0| + 2)}.$$

Taking $c = \min\{\delta/4, 1/(7C)\}$ completes the proof of the theorem. □

Remark 8.4. The above proof recasts a classical argument due to Mertens. The original proof has as its starting point the relation

$$3 + 4\cos\theta + \cos(2\theta) = 2(1 + \cos\theta)^2 \geqslant 0, \tag{8.6}$$

often called the *3-4-1 inequality*. Setting $\theta = tm\log p$ and multiplying the above inequality by $p^{-m\sigma}\log p$ and summing it over all p and m yields (8.5). The proof of (8.5) we presented is due to Granville and Soundararajan [74] and fits within the framework of the theory of *pretentious multiplicative functions*. A full account of this theory is given in [75]. In addition, elements of it can be found in Chapters 13, 14, 22 and 27 of this book. □

We are finally ready to prove the Prime Number Theorem.

Proof of Theorem 8.1. We will estimate $\psi(x)$ instead; passing to $\pi(x)$ can be easily accomplished by partial summation (see Exercise 1.7(d)).

Combining the explicit formula (5.11) with Theorem 8.3, we find that there is an absolute constant $c_1 > 0$ such that

$$\psi(x) = x + O\Big(\sum_{|\gamma| \leqslant T} \frac{x^{1-c_1/\log T}}{|\rho|} + \frac{x \log^2 x}{T} \Big)$$

for any $T \in [2, x]$. Moreover, since $\zeta(0) \neq 0$, we have that $|\rho| \gg 1$ for all non-trivial zeroes of ζ. In particular, we have the estimate $|\rho| \asymp 1 + |\gamma|$. Finally, arguing as in (6.5), we find that $\sum_{|\gamma| \leqslant T} 1/(1+|\gamma|) \ll \log^2 T$. Putting everything together, we conclude that

$$\psi(x) = x + O\big(x^{1-c_1/\log T} (\log T)^2 + x(\log x)^2/T\big).$$

Taking $T = e^{\sqrt{\log x}}$ completes the proof. □

A bit of complex analysis

We conclude the chapter with the promised proof of Lemma 8.2. The starting point is a variation of the classical Borel-Carathéodory theorem.

Lemma 8.5. *Consider a function $f(z) = \sum_{n=0}^{\infty} c_n z^n$ that is analytic in the disk* [1] *$|z| \leqslant R$ with $f(0) = 0$. If $\operatorname{Re}(f(z)) \leqslant M$ whenever $|z| = R$, then*

$$|c_n| \leqslant \frac{8M}{R^n} \quad \text{for} \quad n \in \mathbb{Z}_{\geqslant 1}.$$

Furthermore, for $k \in \mathbb{Z}_{\geqslant 0}$ and $|z| \leqslant (1-\varepsilon)R$ with $0 < \varepsilon < 1$, we have

$$|f^{(k)}(z)| \leqslant \frac{8k!M}{\varepsilon^{k+1} R^k}.$$

Proof. Since f is analytic in the closed disk $|z| \leqslant R$, a compactness argument implies that it is also analytic in an open disk $|z| < R'$ with $R' > R$. In particular, its Taylor series $\sum_{n=0}^{\infty} c_n z^n$ converges absolutely when $|z| = R$.

Note that $c_0 = f(0) = 0$. Write $c_n = a_n + ib_n$ so that

$$\operatorname{Re}(f(Re^{i\theta})) = \sum_{n=0}^{\infty} R^n a_n \cos(n\theta) - \sum_{n=1}^{\infty} R^n b_n \sin(n\theta).$$

Fourier inversion (or Cauchy's residue theorem) then implies that

$$R^n a_n = \frac{1}{\pi} \int_0^{2\pi} \operatorname{Re}(f(Re^{i\theta})) \cos(n\theta) \mathrm{d}\theta \tag{8.7}$$

for $n \in \mathbb{Z}_{\geqslant 0}$. In particular, $\int_0^{2\pi} \operatorname{Re}(f(Re^{i\theta}))\mathrm{d}\theta = a_0 = 0$ and

$$|a_n| \leqslant \frac{1}{R^n \pi} \int_0^{2\pi} |\operatorname{Re}(f(Re^{i\theta}))| \mathrm{d}\theta$$

[1] This means that f is analytic in an open neighborhood of the disk $\{z \in \mathbb{C} : |z| \leqslant R\}$.

for $n \in \mathbb{Z}_{\geqslant 1}$. Hence

$$|a_n| \leqslant \frac{1}{R^n \pi} \int_0^{2\pi} (|\operatorname{Re}(f(Re^{i\theta}))| + \operatorname{Re}(f(Re^{i\theta})))\mathrm{d}\theta$$
$$= \frac{2}{R^n \pi} \int_0^{2\pi} \max\{\operatorname{Re}(f(Re^{i\theta})), 0\}\mathrm{d}\theta \leqslant \frac{4M}{R^n}.$$

A similar argument implies the same bound for $|b_n|$ and the claimed bound on $|c_n|$ follows. For the bound on $f^{(k)}(z)$, we simply note that $|f^{(k)}(z)| \leqslant \sum_{n \geqslant k} n(n-1)\cdots(n-k+1)|c_n|((1-\varepsilon)R)^{n-k}$ for $|z| \leqslant (1-\varepsilon)R$, and then insert our bound for c_n. □

If $f(z) = \log g(z)$ with $g(z) \neq 0$ on some disk $|z| \leqslant R$, then $\operatorname{Re}(f(z)) = \log|g(z)|$. Lemma 8.5 then allows us to translate an upper bound on $|g(z)|$ to bounds on f and its derivatives. This leads us to the following lemma, which is a generalization of the fact that a polynomial of degree d and of bounded coefficients grows roughly like $e^{O(d)}$ in the unit disk $|z| \leqslant 1$, while also having at most d roots there. Part (a) is due to Landau, and part (b) is a weak quantitative form of Jensen's formula from complex analysis.

Lemma 8.6. *Assume that $g(z)$ is analytic in the disk $|z| \leqslant 4R$ with $g(0) \neq 0$, and let $z_1, \dots, z_k$ be its zeroes in the disk $|z| \leqslant 2R$ listed with multiplicity.*
(a) *If $M \geqslant 0$ is such that $|g(z)| \leqslant e^M \cdot |g(0)|$ when $|z| = 4R$, then*

$$\left|\frac{g'(z)}{g(z)} - \sum_{\ell=1}^{k} \frac{1}{z - z_\ell}\right| \leqslant \frac{16M}{R} \qquad \text{for} \quad |z| \leqslant R.$$

(b) *If $M' \geqslant 0$ is such that $|g(z)| \leqslant e^{M'} \cdot |g(0)|$ when $|z| = 2R$, then*

$$\#\{\, 1 \leqslant \ell \leqslant k : |z_\ell| \leqslant R \,\} \leqslant 2M'. \tag{8.8}$$

Proof. (a) The function $G(z) = g(z)/\prod_{\ell=1}^{k}(z - z_\ell)$ is analytic for $|z| \leqslant 4R$ and non-zero for $|z| \leqslant 2R$. Thus $f(z) := \log(G(z)/G(0))$ is analytic in the disk $|z| \leqslant 2R$. It also vanishes at the origin. The maximum modulus principle implies that

$$\max_{|z|=2R} |G(z)/G(0)| \leqslant \max_{|z|=4R} |G(z)/G(0)| = \max_{|z|=4R} \left|\frac{g(z)}{g(0)} \prod_{\ell=1}^{k} \frac{z_\ell}{z - z_\ell}\right| \leqslant e^M,$$

since $|z_\ell| \leqslant 2R \leqslant |z - z_\ell|$ on the circle $|z| = 4R$. We then bound $f'(z)$ using Lemma 8.5 to complete the proof of part (a).

(b) We could use Jensen's formula to prove the second part. Instead, we give a direct proof following [146, Lemma 6.1].

Consider the auxiliary function

$$h(z) = g(z) \prod_{\ell=1}^{k} \frac{2R - z\bar{z}_\ell/2R}{z - z_\ell}$$

that is analytic for $|z| \leqslant 2R$. In addition, $|h(z)| = |g(z)|$ for $|z| = 2R$, since $z/2R = 2R/\bar{z}$ for such z and thus $|2R - z\bar{z}_\ell/2R| = |z - z_\ell|$. The maximum modulus principle then implies that

$$\max_{|z|=2R} |g(z)| \geqslant |h(0)| = |g(0)| \cdot \prod_{\ell=1}^{k} \frac{2R}{|z_\ell|} \geqslant |g(0)| \cdot 2^{\#\{\,1 \leqslant \ell \leqslant k : |z_\ell| \leqslant R\,\}},$$

and the proof is complete. □

Proof of Lemma 8.2. Any set of zeroes of ζ we consider in this proof is implicitly a multiset with each zero listed as many times as its multiplicity.

(a) Notice that $\zeta(s) \ll 1 + |t|$ for $|s - 1| \geqslant 1$ and $\sigma \in [1/2, 5]$ by (5.7). Moreover, relation (5.8) with $k = 3$ implies that $\zeta(s) \ll 1 + |t|^3$ for $|s - 1| \geqslant 1$ and $\sigma \in [-3/2, 1/2]$. We then apply Lemma 8.6(b) with $g(z) = \zeta(2 + it + z)(1 + it + z)$ and $R = 3$. We note that $|g(z)/g(0)| = O(1 + |t|^4)$ for $|z| \leqslant 4R$, since $1/\zeta(2 + it) = O(1)$ by (6.9). Therefore, if $A = \{\,\rho : |\rho - 2 - it| \leqslant 3\,\}$, then Lemma 8.6(b) implies that $|A| \ll \log(2+|t|)$. Since all zeroes with $|\gamma - t| \leqslant 1$ are in A, part (a) of the lemma follows.

(b) When $\sigma \geqslant 2$, the result is trivially true, since $(\zeta'/\zeta)(s) = O(1)$ for such s, as well as $|s - \rho| \geqslant 1$ for all zeroes with $|t - \gamma| \leqslant 1$ (and there are $O(\log(|t| + 2))$ such zeroes).

Next, assume that $-1 \leqslant \sigma \leqslant 2$, so that $|s - 2 - it| \leqslant 3$. Let $A' = \{\,\rho : |\rho - 2 - it| \leqslant 6\,\}$ and $A'' = \{\,\rho : |t - \gamma| \leqslant 1\,\}$. Lemma 8.6(a) (applied again to $g(z) = \zeta(2 + it + z)(1 + it + z)$ with $R = 3$) implies that

$$\left| \frac{\zeta'}{\zeta}(s) + \frac{1}{s-1} - \sum_{\rho \in A'} \frac{1}{s - \rho} \right| \ll \log(2 + |t|).$$

Note that $|s - \rho| \geqslant 1$ when $\rho \in A' \setminus A''$, and there are $\leqslant |A'| \ll \log(2 + |t|)$ such zeroes. This completes the proof of part (b) when $\sigma \in [-1, 2]$.

Finally, assume that $\sigma \leqslant -1$ and $|s + 2n| \geqslant 1/2$ for all $n \in \mathbb{N}$. Then we have $|\cot(\pi s/2)| \ll 1$. Thus, the functional equation (6.2) and Exercise 1.12(b) imply that

$$(\zeta'/\zeta)(s) = -(\zeta'/\zeta)(1 - s) - \log|1 - s| + O(1) = O(\log|s|). \tag{8.9}$$

This completes the proof of Lemma 8.2 in all cases. □

Exercises

Exercise 8.1. Let $x, T \geqslant 2$ and $\tau \in \mathbb{R}$.

(a) Adapt the proof of Theorem 5.1 to prove that

$$\sum_{n \leqslant x} \Lambda(n) n^{i\tau} = \frac{x^{1+i\tau}}{1+i\tau} - \sum_{|\gamma+\tau| \leqslant T} \frac{x^{\rho+i\tau}}{\rho+i\tau} + O\bigg(\frac{x \log^2(x(|\tau|+T))}{T} + \log x\bigg).$$

(b) Assuming the Riemann Hypothesis, prove that

$$\sum_{n \leqslant x} \Lambda(n) n^{i\tau} = \frac{x^{1+i\tau}}{1+i\tau} + O(\sqrt{x} \log^2(x+|\tau|)).$$

Exercise 8.2.

(a) Show that if $N \in \mathbb{Z}_{\geqslant 1}$ is not a prime power, then $\psi(N) = N - \sum_{|\gamma| \leqslant N^2} N^\rho/\rho - \log(2\pi) + o_{N\to\infty}(1)$. [*Hint:* First, prove a version of Theorem 7.2 with a better error term. (Solution in [31, Chapter 17].)]

(b) Show that ζ must have infinitely many non-trivial zeroes.

(c) For each $\varepsilon > 0$, show that there is at least one non-trivial zero ρ with $\mathrm{Re}(\rho) \geqslant 1/2 - \varepsilon$. Conclude that we *cannot* have $\psi(x) = x + O(x^{1/2-\varepsilon})$ for all $x \geqslant 1$.

Exercise 8.3.

(a) Consider ϕ and Φ as in Exercise 7.2(b). For $x \geqslant 1$ and $s \in \mathbb{C}\backslash\{1\}$ not coinciding with any zero of ζ show that

$$\sum_{n \geqslant 1} \frac{\Lambda(n)\phi(n/x)}{n^s} = x^{1-s}\Phi(1-s) - \sum_\rho x^{\rho-s}\Phi(\rho-s) - \phi_0 \frac{\zeta'}{\zeta}(s) - \sum_{n=1}^\infty x^{-2n-s}\Phi(-2n-s).$$

(b) When $s = 0$ and $x = 1$, simplify the above formula to

$$\sum_{n \geqslant 1} \Lambda(n)\phi(n) = \int_0^\infty \phi(y)\mathrm{d}y - \sum_\rho \Phi(\rho) - \phi_0 \log(2\pi) + \frac{1}{2}\int_1^\infty \phi'(y)\log(1-1/y^2)\mathrm{d}y.$$

Deduce that ζ has infinitely many non-trivial zeroes.

Exercise 8.4. Write c_1 for the constant in Theorem 8.3 and let $\delta_t = c_1/\log(|t|+2)$.

(a) For $\sigma \geqslant 1 - 0.5\delta_t$ and $|t| \geqslant 3$, show that $(\zeta'/\zeta)(s) \ll \log^2|t|$.

(b) In the same range of s, improve the above bound to

$$(\zeta'/\zeta)(s) \ll \log|t|.$$

[*Hint:* Prove that $\mathrm{Re}(-(\zeta/\zeta')(s)) \leqslant O(\log|t|)$ for $\sigma > 1-\delta_t$ and that $(\zeta'/\zeta)(1+\delta_t+it) \ll \log|t|$. Then, use Lemma 8.5.]

(c) For $\sigma \geqslant 1 - 0.5\delta_t$ and $|t| \geqslant 3$, prove that

$$|\log \zeta(s)| \leqslant \log\log|t| + O(1).$$

[*Hint:* When $\sigma \leqslant 1 + \delta_t$, show that $\log\zeta(s) = \log\zeta(1 + \delta_t + it) + O(1)$.]

(d) Show that a constant $c_2 > 0$ such that

$$\sum_{n \leqslant x} \mu(n) \ll x e^{-c_2\sqrt{\log x}} \quad (x \geqslant 2).$$

[*Hint:* Start with Theorem 7.2 with $\alpha = 1 + 1/\log x$, and then replace the contour $[\alpha - iT, \alpha + iT]$ by the contour

$$L = [\alpha - iT,\ \alpha' - iT] \ + \ [\alpha' - iT, \alpha' + iT] \ + \ [\alpha' + iT, \alpha + iT],$$

where $\alpha' = 1 - 0.5\delta_T$. To bound $1/\zeta$ on L, note that $|1/\zeta| \leqslant \exp\{|\log\zeta|\}$.]

Exercise 8.5. Assume the Riemann Hypothesis and fix $\varepsilon \in (0, 1/2)$.

(a) Adapt the argument of Exercise 8.4 to prove that

$$(\zeta'/\zeta)(s),\ \log\zeta(s) \ll_\varepsilon \log|t| \quad \text{for } |t| \geqslant 2,\ \sigma \geqslant 1/2 + \varepsilon.$$

(b) For $|t| \geqslant 2$ and $\sigma \geqslant 1/2 + \varepsilon$ prove that

$$\log\zeta(s),\ (\zeta'/\zeta)(s) \ll_\varepsilon (\log|t|)^{2\max\{1-\sigma,0\}+\varepsilon}.$$

Infer the Lindelöf Hypothesis. [*Hint:* Adapt the proof of Theorem 6.2.]

(c) Use Exercise 8.3(a) to give an alternative solution to part (b). [*Hint:* Rearranging the terms in Exercise 8.3(a) gives an expression for $(\zeta'/\zeta)(s)$. If $\mathrm{supp}(\phi) \subseteq [0, 2]$, then $\sum_n \Lambda(n)\phi(n/x)/n^s \ll x^{1-\sigma}\log x$. On the other hand, using an estimate of the form (7.7), the sum over zeroes ρ is $O_{\varepsilon,\phi}(x^{1/2-\sigma}\log|t|)$. Optimize x.]

Exercise 8.6. Prove that the Riemann Hypothesis is equivalent to:

(a) For each $\varepsilon > 0$, we have $\psi(x) = x + O_\varepsilon(x^{1/2+\varepsilon})$ uniformly in $x \geqslant 1$.

(b) For each $\varepsilon > 0$, we have $\sum_{n \leqslant x} \mu(n) \ll_\varepsilon x^{1/2+\varepsilon}$ uniformly in $x \geqslant 1$.

Exercise 8.7* ([31, Chapters 11–12])**.** Let $\xi(s) = \pi^{-s/2}\Gamma(s/2)\zeta(s)s(s-1)/2$.

(a) Prove that ξ is entire, satisfies the functional equation $\xi(s) = \xi(1 - s)$ and its zeroes are precisely the non-trivial zeroes of ζ, occurring with the same multiplicity.

(b) Prove that $|\xi(s)| \leqslant \exp\{0.5|s|\log|s| + O(|s|)\}$ for $|s| \geqslant 1$. [*Hint:* It suffices to consider the case when $\mathrm{Re}(s) \geqslant 1/2$.]

(c) Prove that the *Hadamard product* $h(s) = \prod_\rho (1 - s/\rho)e^{s/\rho}$ converges absolutely and uniformly on compact subsets of $\mathbb{C}$ (see Exercise 1.14). Deduce that there is an entire function Q such that $\xi = he^Q$.

(d) Prove that $\xi(s)/h(s) \ll \exp\{O(|s|\log^2|s|)\}$ for $|s| \geqslant 3$, as follows:
   (i) Let n_s denote the number of zeroes of ξ in the disk $\{\, z : |z - s| \leqslant 1 \,\}$ counted with multiplicity. Show that $n_s \ll \log|s|$.
   (ii) Show that there is $r \in [0, 1]$ such that all zeroes of ξ are at distance $\geqslant 1/(2n_s + 2) \gg 1/\log|s|$ from the circle $\{\, z : |z - s| = r \,\}$.

(iii) Fix z on the circle $|z-s|=r$. If $|\rho| \geqslant 2|z|$, prove that $|(1-z/\rho)e^{z/\rho}| = e^{O(|z|^2/|\rho|^2)}$; if $|\rho| \leqslant 2|z|$, prove that $|(1-z/\rho)e^{z/\rho}| \gg 1/(|s|\log|s|)$.

(iv) If $|z-s|=r$, prove that $\xi(z)/h(z) \ll \exp\{O(|s|\log^2|s|)\}$. Then use the maximum modulus principle to bound $\xi(s)/h(s)$.

(e) Use Lemma 8.5 to prove that $Q(s) \ll |s|\log^2|s|$ for $|s| \geqslant 3$. Conclude that $Q(s) = A + Bs$ for some $A, B \in \mathbb{C}$. [*Hint:* If $Q(s) = \sum_{n=0}^{\infty} c_n s^n$, then $c_n = (1/2\pi i)\oint_{|z|=R} Q(s)s^{-n-1}\mathrm{d}s$ for *any* R.]

(f) Show that $e^A = \xi(0) = 1/2$ and $B = (\xi'/\xi)(0) = (-\gamma + \log(4\pi))/2 - 1$.

(g) Prove that $-(\xi'/\xi)(0) = (\xi'/\xi)(1) = B + \sum_\rho (1/\rho + 1/(1-\rho))$. Conclude that $\lim_{T\to\infty}\sum_{|\gamma|\leqslant T} 1/\rho = (\gamma - \log(4\pi))/2 + 1$.

Exercise 8.8* ([31, Chapter 15])**.** Let $N(T)$ be the number of zeroes of $\zeta(s)$ in the rectangle $0 < \sigma < 1$, $0 < t < T$, and assume that T does not coincide with the ordinate of a zero.

(a) Let C be a contour that does not self-intersect, parametrized by the map $\phi : [0,1] \to C$ (i.e., ϕ is surjective and $\phi|_{(0,1)}$ injective). Moreover, let f be a holomorphic function that is defined in an open neighborhood of C and does not vanish on C.

(i) Show that there is an open, simply connected domain Ω such that $\Omega \cap C = C \setminus \{\phi(0), \phi(1)\}$ and $f(s) \neq 0$ for all $s \in \Omega$. In particular, we may define $\log f(s)$ on Ω.

(ii) Define the variation of the argument of f along C by

$$\Delta_C \arg f(s) := \mathrm{Im}\left(\log f(\phi(1^-)) - \log f(\phi(0^+))\right) = \mathrm{Im}\int_C \frac{f'}{f}(s)\mathrm{d}s. \tag{8.10}$$

Show that $\Delta_C \arg f(s) = \int_C (f'/f)(s)\mathrm{d}s$. In particular, $\Delta_C \arg f(s)$ is independent of the choice of ϕ and of the branch of $\log f$.

(b) Let $\xi(s)$ be as in Exercise 8.7, and let R be the rectangle with vertices 2, $2+iT$, $-1+iT$ and -1, traversed counterclockwise. Prove that $\xi(s) > 0$ for $s \in \mathbb{R}$, as well as that

$$2\pi N(T) = \Delta_R \arg \xi(s).$$

(c) If $L = [2, 2+iT] + [2+iT, 1/2+iT]$, then prove that

$$\Delta_R \arg \xi(s) = 2\Delta_L \arg \xi(s).$$

[*Hint:* Show that $\xi(\sigma+it) = \overline{\xi}(1-\sigma+it)$.]

(d) Use Stirling's formula to prove that

$$\Delta_L \arg \Gamma(s/2+1) = \frac{T}{2}\log\frac{T}{2e} + \frac{3\pi}{8} + O(1/T).$$

(e) Prove that $\Delta_L \arg \zeta(s) = O(\log T)$, and conclude that

$$N(T) = \frac{T}{2\pi}\log\frac{T}{2\pi e} + O(\log T).$$

[*Hint:* Note that $\log\zeta(2+iT) = O(1)$. Then use Lemma 8.2 to control $\Delta_{[2+iT,1/2+iT]} \arg \zeta(s)$.]

Chapter 9

Dirichlet characters

Having obtained a firm understanding of the frequency of occurrence of primes, we turn to other aspects of their distribution. Specifically, we would like to know what kind of patterns occur among them. Perhaps the simplest such question one can ask is whether there are primes in a given arithmetic progression $a \pmod q$, that is to say, primes of the form $qn + a$, with q and a being fixed and n varying. A natural restriction is that a and q must be coprime but, other than that, there is no reason *a priori* why the primes should have any preference for any particular reduced residue class mod q. Thus, Occam's razor leads us to the prediction that

$$\pi(x; q, a) = \#\{\, p \leqslant x : p \equiv a \pmod q \,\} \sim \frac{x}{\varphi(q) \log x} \qquad (x \to \infty) \tag{9.1}$$

for each pair of fixed and coprime natural numbers a and q.

Given the success of the Dirichlet series approach to the study of $\pi(x)$, it is tempting to introduce the series $\sum_{p \equiv a \pmod q} 1/p^s$. However, it is not obvious how to analyze this function because the condition $p \equiv a \pmod q$ does not behave well under multiplication and thus there is no obvious analogue of ζ. We will circumvent this problem using the ring structure of $\mathbb{Z}/q\mathbb{Z}$.

To explain the idea, suppose we want to count primes $p \equiv 1 \pmod 4$. Instead of counting them on their own, we note that they have a natural counterpart, the primes $p \equiv 3 \pmod 4$. Since every prime $p > 2$ is either $1 \pmod 4$ or $3 \pmod 4$, we have the linear relation

$$\pi(x; 4, 1) + \pi(x; 4, 3) = \pi(x) - 1 \sim x/\log x.$$

Thus, instead of showing that $\pi(x; 4, 1) \sim x/(2 \log x)$, it suffices to prove that $\pi(x; 4, 1) \sim \pi(x; 4, 3)$ or, equivalently, that $\pi(x; 4, 1) - \pi(x; 4, 3) =$

$o(x/\log x)$. We write

$$\pi(x;4,1) - \pi(x;4,3) = \sum_{p \leqslant x} \varepsilon(p),$$

where $\varepsilon(2) = 0$, $\varepsilon(p) = 1$ if $p \equiv 1 \pmod 4$ and $\varepsilon(p) = -1$ if $p \equiv 3 \pmod 4$. The function ε extends naturally to a 4-periodic function: we let $\varepsilon(0) = \varepsilon(2) = 0$, $\varepsilon(1) = 1$ and $\varepsilon(3) = -1$, and then define $\varepsilon(n)$ according to the remainder of $n \bmod 4$. The key observation is that ε is completely multiplicative over $\mathbb{Z}$, i.e., $\varepsilon(mn) = \varepsilon(m)\varepsilon(n)$ for all $m, n \in \mathbb{Z}$. Hence the Dirichlet series $\sum_p \varepsilon(p)/p^s$ is closely related to the logarithm of $E(s) = \sum_{n=1}^{\infty} \varepsilon(n)/n^s$. Furthermore, the periodicity and multiplicativity of ε allows us to get our hands on $E(s)$ much like we did with $\zeta(s)$.

Suppose now more generally that we want to study the primes in some reduced arithmetic progression mod q. Instead of considering one residue class on its own, we consider simultaneously all of them. Given complex numbers c_a indexed by $a \in (\mathbb{Z}/q\mathbb{Z})^*$, we form the linear combination

$$\sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} c_a \pi(x;q,a) = \sum_{p \leqslant x} f(p),$$

where $f(p) = 0$ if $p|q$ and $f(p) = c_a$ if $p \equiv a \pmod q$ with $(a,q) = 1$. We extend f to a q-periodic function over $\mathbb{Z}$ letting $f(n) = 0$ if $(n,q) > 1$ and $f(n) = c_a$ if $n \equiv a \pmod q$ with $(a,q) = 1$. We wish to find choices of coefficients c_a for which f is a completely multiplicative function over $\mathbb{Z}$, similarly to the function ε above.

We are thus naturally led to the concept of *Dirichlet characters*: given $q \in \mathbb{N}$, we say that the function $\chi : \mathbb{Z} \to \mathbb{C}$ is a Dirichlet character mod q if:

- χ is q-periodic;
- $\chi(n) \neq 0$ if and only if $(n,q) = 1$;
- χ is completely multiplicative over $\mathbb{Z}$.

As their name and the preceding discussion indicate, these objects were introduced by Dirichlet in his pioneering work on primes in arithmetic progressions. In the language of group theory, Dirichlet characters are in one-to-one correspondence with group homomorphisms from $(\mathbb{Z}/q\mathbb{Z})^*$ to $\mathbb{C}^*$, namely 1-dimensional representations of the group $(\mathbb{Z}/q\mathbb{Z})^*$. This correspondence is given by associating to each χ the group homomorphism $\widetilde{\chi} : (\mathbb{Z}/q\mathbb{Z})^* \to \mathbb{C}^*$ defined by $\widetilde{\chi}(n \pmod q) = \chi(n)$. Basic group theory implies that $\widetilde{\chi}$ takes values on the unit circle. In particular, $|\chi| \leqslant 1$.

As we will see in the next chapter, there are exactly $\varphi(q)$ Dirichlet characters χ mod q and they provide an orthonormal basis for the Hilbert space

of functions $f : (\mathbb{Z}/q\mathbb{Z})^* \to \mathbb{C}$ equipped with the inner product

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

This means that each character χ introduces an independent linear relation

$$\sum_{a\in(\mathbb{Z}/q\mathbb{Z})^*} \chi(a)\pi(x;q,a) = \sum_{p\leqslant x} \chi(p). \tag{9.2}$$

Moreover, the aforementioned orthonormality of the Dirichlet characters makes it easy to invert these linear relations: we have

$$\pi(x;q,a) = \frac{1}{\varphi(q)} \sum_{\chi \,(\mathrm{mod}\, q)} \overline{\chi}(a) \sum_{p\leqslant x} \chi(p). \tag{9.3}$$

Examining the above formula, we discover a Dirichlet character mod q that stands out: the function $n \to 1_{(n,q)=1}$. We call it the *principal character* mod q and denote it by χ_0. Its contribution to $\pi(x;q,a)$ equals

$$\frac{1}{\varphi(q)} \sum_{p\leqslant x,\ p\nmid q} 1 = \frac{\pi(x) + O(\log q)}{\varphi(q)}, \tag{9.4}$$

since there are $\leqslant \log q/\log 2$ prime divisors of q. We thus see that χ_0 naturally provides us with the conjectured main term in (9.1). Consequently, proving (9.1) amounts to showing that

$$\sum_{p\leqslant x} \chi(p) = o_{x\to\infty}(\pi(x)) \qquad \text{for} \quad \chi \neq \chi_0. \tag{9.5}$$

To estimate the left-hand side of (9.5), we introduce the Dirichlet series

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

This series is called the *Dirichlet L-function* associated to χ. Since χ is periodic and multiplicative, the behavior of $L(s,\chi)$ can be analyzed using analogous tools to the ones used to study ζ. In particular, we will prove that $L(s,\chi)$ can be analytically continued to the entire complex plane when $\chi \neq \chi_0$. For now, we note that $L(s,\chi)$ converges absolutely for $\mathrm{Re}(s) > 1$ because $|\chi| \leqslant 1$. In particular, the complete multiplicativity of χ and Theorem 4.6 imply that

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

Taking logarithms, we find that $\sum_p \chi(p)/p^s \approx \log L(s,\chi)$ for $\mathrm{Re}(s) > 1$, which provides the link between the sum in (9.5) and $L(s,\chi)$.

As in the proof of the Prime Number Theorem, it is more convenient to work with the logarithmic derivative of $L(s,\chi)$, for which we have

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

Instead of $\sum_{p\leqslant x}\chi(p)$ and $\pi(x;q,a)$, we then estimate

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

To proceed with this task, we use Perron's formula (Theorem 7.2) to write $\psi(x,\chi)$ in terms of the Dirichlet series $(-L'/L)(s,\chi)$. The analogy with the theory of $\psi(x)$ is now clear: the zeroes of $L(s,\chi)$ determine the poles of $(-L'/L)(s,\chi)$ and hence the asymptotic behavior of $\psi(x,\chi)$. In fact, in Chapter 11 we will show a generalization of the explicit formula (5.11):

$$\psi(x,\chi) = -\sum_{|\gamma|\leqslant T} \frac{x^\rho - 1}{\rho} + O\Big(\frac{x\log^2(qx)}{T}\Big) \tag{9.6}$$

uniformly for $x \geqslant T \geqslant 2$ and non-principal characters $\chi \,(\mathrm{mod}\, q)$. As in (5.11), we write $\rho = \beta + i\gamma$ for a *non-trivial zero* of $L(s,\chi)$, which necessarily lies inside the critical strip $0 \leqslant \mathrm{Re}(s) \leqslant 1$. In addition, zeroes are summed according to their multiplicities. There is a small difference though: the function $L(s,\chi)$ could have a zero at $s = 0$, which is the reason why the summands in (9.6) are slightly modified compared to those of (5.11).

Given (9.6), proving (9.1) is reduced to showing a zero-free region for $L(s,\chi)$.

The strategy we outlined above is carried out in the subsequent three chapters: Chapter 10 is dedicated to the study of character theory of finite abelian groups and its applications to Dirichlet characters, and Chapter 11 to the study of Dirichlet L-functions and to the proof of (9.6). Finally, in Chapter 12, we prove the necessary zero-free regions for Dirichlet L-functions to establish a uniform version of (9.1).

Exercises

Exercise 9.1. Find all Dirichlet characters mod 2, 3, 4, 5, 8 and 15. In each case, calculate $\sum_{\chi \,(\mathrm{mod}\, q)} \chi(a)$ for all $a = 0, 1, \ldots, q-1$, as well as $\sum_{a=0}^{q-1}\chi(a)$ for all $\chi \,(\mathrm{mod}\, q)$.

Exercise 9.2. Prove that if χ is a Dirichlet character and $(n,q) = 1$, then $|\chi(n)| = 1$. [*Hint:* Show that there is some integer k such that $\chi(n)^k = 1$.]

Exercise 9.3. Let χ be a Dirichlet character mod q.

(a) Prove that $\chi(-1) \in \{-1, 1\}$.

(b) Prove that χ is a real valued function if and only if $\chi^2 = \chi_0$.

(c) If χ is real valued and q is prime, then prove that either $\chi = \chi_0$ or χ is the Legendre symbol $(\cdot|q)$. [*Hint:* Evaluate $\chi(n^2)$.]

Exercise 9.4.

(a) If χ_j is a Dirichlet character mod q_j for $j = 1, 2$, then show that $\chi = \chi_1\chi_2$ is a character mod $[q_1, q_2]$.

(b) Conversely, if χ is a Dirichlet character mod q, and $q = q_1q_2$ with $(q_1, q_2) = 1$, construct characters $\chi_j \,(\mathrm{mod}\, q_j)$ for $j = 1, 2$ such that $\chi = \chi_1\chi_2$. [*Hint:* For each $n \in \mathbb{Z}$, there is a unique class $a \,(\mathrm{mod}\, q)$ such that $a \equiv n \,(\mathrm{mod}\, q_1)$ and $a \equiv 1 \,(\mathrm{mod}\, q_2)$. Define $\chi_1(n) := \chi(a)$.]

(c) If p is an odd prime, we know that the group $(\mathbb{Z}/p^k\mathbb{Z})^*$ is cyclic for each $k \geqslant 1$. Construct all Dirichlet characters mod p^k.

(d) Fix $k \geqslant 3$. We know that for each odd n, there are unique integers $a \in \{0, 1\}$ and $b \in \{0, 1, \dots, 2^{k-2}\}$ such that $n \equiv (-1)^a 5^b \,(\mathrm{mod}\, 2^k)$. Use this fact to construct all Dirichlet characters mod 2^k.

(e) Construct all Dirichlet characters mod q and deduce that there are exactly $\varphi(q)$ of them.

Exercise 9.5. A character $\chi \,(\mathrm{mod}\, q)$ is called *faithful* if $\chi(m) = \chi(n) \neq 0$ implies that $m \equiv n \,(\mathrm{mod}\, q)$. Otherwise, χ is called *unfaithful*. If q is prime, then show that there are $\varphi(q-1)$ faithful Dirichlet characters mod q.

Exercise 9.6. Let χ be a Dirichlet character mod q. An integer $d \geqslant 1$ is called a *period* of χ if $\chi(m) = \chi(n)$ when $m \equiv n \,(\mathrm{mod}\, d)$ and $(mn, q) = 1$.

(a) Show that d is a period of χ if and only if $\chi(n) = 1$ whenever $n \equiv 1 \,(\mathrm{mod}\, d)$ and $(n, q) = 1$.

(b) Show that if d_1, d_2 are periods of χ, then so is (d_1, d_2). [*Hint:* If m, n, k, $\ell \in \mathbb{Z}$ are such that $(mn, q) = 1$ and $m = n + kd_1 + \ell d_2$, then show that there is an integer a such that $(n + (k - ad_2)d_1, q) = 1$.]

(c) Show that there is a divisor $d^* \geqslant 1$ of q such that the set of periods of χ is the set of multiples of d^*. We call d^* the *conductor* of χ.

(d) If χ is faithful, show that $d^* = q$, but the converse is not always true.

(e) Let $\chi = \chi_1\chi_2$ be as in Exercise 9.4(a) with $(q_1, q_2) = 1$, and let d^*, d_1^* and d_2^* be their conductors, respectively. Prove that $d^* = d_1^*d_2^*$.

Exercise 9.7. Let χ be a real, non-principal character mod p^k with p prime.

(a) If $p > 2$, prove that $\chi(n) = (n|p)$ and that its conductor is p. [*Hint:* Use Hensel's lemma to study the congruence $x^2 \equiv n \,(\mathrm{mod}\, p^k)$.]

(b) If $q = 8$, prove there are three possibilities for χ: two of conductor 8 and one of conductor 4.

(c) If $p = 2$ and $k > 3$, then prove that $\chi(n) = \psi(n \,(\mathrm{mod}\, 8))$, where ψ is one of the three characters in part (b).

Chapter 10

Fourier analysis on finite abelian groups

As we saw in the previous chapter, Dirichlet characters are in correspondence with group homomorphisms from $(\mathbb{Z}/q\mathbb{Z})^*$ to $\mathbb{C}^*$. More generally, a *character of an abelian group* $(G, \cdot)$ is a group homomorphism $\chi : G \to \mathbb{C}^*$. We write $\widehat{G}$ for the set of all characters of G. The constant function 1 is obviously a character, called the *principal character* of G and often denoted by χ_0. All other characters are called *non-principal*.

The set $\widehat{G}$ admits a natural group structure with the operation being the multiplication of complex-valued functions. The group $(\widehat{G}, \cdot)$ is called the *dual group* of G or the *group of characters* of G.

From now on, we assume that G is finite. In this case, $\widehat{G}$ is also finite. Indeed, if $n = |G|$, then $g^n = 1$ from Lagrange's theorem. In particular, $\chi(g)^n = \chi(g^n) = \chi(1) = 1$. We infer that $\widehat{G}$ is finite and that $1/\chi = \overline{\chi}$.

The set $\widehat{G}$ is a subset of the set of functions from G to $\mathbb{C}$. We denote the latter set by $L^2(G)$ because it naturally forms a Hilbert space over $\mathbb{C}$ with respect to the inner product

$$\langle \alpha, \beta \rangle_G := \frac{1}{|G|} \sum_{g \in G} \alpha(g)\bar{\beta}(g).$$

Clearly, $\dim_{\mathbb{C}}(L^2(G)) = |G|$. A fundamental property of $\widehat{G}$ is that it forms an orthonormal basis of $L^2(G)$.

We begin by showing that $\widehat{G}$ is an orthonormal set, that is to say, $\langle \chi, \psi \rangle_G = 1_{\chi=\psi}$ for all characters $\chi, \psi \in G$. The case $\chi = \psi$ follows immediately by the fact that χ takes values on the unit circle. On the other

hand, if $\chi \neq \psi$ and we set $\xi = \chi\bar{\psi}$, then we must prove that $\sum_{g\in G} \xi(g) = 0$. Indeed, our assumption that $\chi \neq \psi$ implies the existence of an $h \in G$ such that $\xi(h) \neq 1$. Since $hG = G$, we have

$$\xi(h) \sum_{g\in G} \xi(g) = \sum_{g\in G} \xi(hg) = \sum_{g\in G} \xi(g),$$

from which we infer that $\langle \chi, \psi\rangle_G = |G|^{-1} \sum_{g\in G} \xi(g) = 0$.

Since $\widehat{G}$ is an orthonormal set of $L^2(G)$, it is an independent set. In particular, we have $|\widehat{G}| \leqslant \dim_{\mathbb{C}}(L^2(G)) = |G|$. To show that $\widehat{G}$ is a basis of $L^2(G)$, it suffices to establish the relation

$$|\widehat{G}| = |G|. \tag{10.1}$$

This relation is obvious if G is cyclic, say $G = \langle g\rangle$ of order n: in this case, every character is uniquely determined by its value at g which, as we saw above, must be an nth root of unity. Conversely, every nth root of unity $e^{2\pi ia/n}$ gives rise to a character χ via the relation $\chi(g^j) := e^{2\pi iaj/n}$, so (10.1) follows in this case. In the general case of a finite abelian group, relation (10.1) follows by writing G as the direct product of cyclic groups, say

$$G \cong \mathbb{Z}/d_1\mathbb{Z} \times \cdots \times \mathbb{Z}/d_k\mathbb{Z}, \tag{10.2}$$

and applying the following lemma whose proof is left as an exercise.

Lemma 10.1. *Let $(G_1, \cdot)$ and $(G_2, \cdot)$ be abelian groups with direct product G. The function $\phi : \widehat{G}_1 \times \widehat{G}_2 \to \widehat{G}$ associating the pair (χ_1, χ_2) to the character $G \ni (g_1, g_2) \to \chi_1(g_1)\chi_2(g_2) \in \mathbb{C}^*$ is a group isomorphism.*

The fact that $\widehat{G}$ is an orthonormal basis of $L^2(G)$ allows us to do Fourier analysis on G: for each $f : G \to \mathbb{C}$, we define the function $\widehat{f} : \widehat{G} \to \mathbb{C}$ by

$$\widehat{f}(\chi) = \langle f, \chi\rangle_G. \tag{10.3}$$

This is the *Fourier transform* of f and it satisfies the *inversion formula*

$$f = \sum_{\chi\in\widehat{G}} \widehat{f}(\chi) \cdot \chi. \tag{10.4}$$

Specializing to the function $f(g) = 1_{g=h}$, where h is a given element of G, we find that

$$1_{g=h} = \frac{1}{|G|} \sum_{\chi\in\widehat{G}} \chi(g)\overline{\chi}(h). \tag{10.5}$$

Finally, we have *Parseval's identity*

$$\sum_{\chi\in\widehat{G}} |\widehat{f}(\chi)|^2 = \frac{1}{|G|} \sum_{g\in G} |f(g)|^2. \tag{10.6}$$

Indeed, since $|z|^2 = z \cdot \bar{z}$, we have

$$\sum_{\chi \in \widehat{G}} |\widehat{f}(\chi)|^2 = \frac{1}{|G|^2} \sum_{\chi \in \widehat{G}} \sum_{g \in G} f(g)\overline{\chi}(g) \sum_{h \in G} \bar{f}(h)\chi(h)$$
$$= \frac{1}{|G|^2} \sum_{g,h \in G} f(g)\bar{f}(h) \sum_{\chi \in \widehat{G}} \overline{\chi}(g)\chi(h),$$

and (10.6) follows from (10.5).

Additive and multiplicative characters mod q

We study now in more detail the cases when $G = \mathbb{Z}/q\mathbb{Z}$ and $G = (\mathbb{Z}/q\mathbb{Z})^*$, the second one corresponding to Dirichlet characters. Since the operation in $\mathbb{Z}/q\mathbb{Z}$ is addition, we call the characters of this group the *additive characters* mod q. Similarly, we also refer to Dirichlet characters, that is to say, the characters of $(\mathbb{Z}/q\mathbb{Z})^*$, as the *multiplicative characters* mod q.

Since $\mathbb{Z}/q\mathbb{Z}$ is a cyclic group, the discussion in the proof of (10.1) implies that the additive characters mod q are the functions $n \to e(an/q)$ indexed by $a \in \{0, 1, \dots, q-1\}$, where we have introduced the symbol

$$e(x) := e^{2\pi i x}. \tag{10.7}$$

In particular, the character group of $\mathbb{Z}/q\mathbb{Z}$ is canonically isomorphic to $\mathbb{Z}/q\mathbb{Z}$.

On the other hand, the construction of the multiplicative characters mod q is explained in Exercise 9.4. In addition, a more detailed discussion is presented in [31, Chapter 4] and in [146, Section 4.2].

The Fourier transform on $\mathbb{Z}/q\mathbb{Z}$ is called the *additive Fourier transform* mod q. Using the explicit description of the character group of $\mathbb{Z}/q\mathbb{Z}$, we may identify the additive Fourier transform of $f : \mathbb{Z}/q\mathbb{Z} \to \mathbb{C}$ with the function $\widehat{f} : \mathbb{Z}/q\mathbb{Z} \to \mathbb{C}$ given by the formula

$$\widehat{f}(a) = \frac{1}{q} \sum_{n \in \mathbb{Z}/q\mathbb{Z}} f(n)e(-an/q).$$

In fact, since there is a natural correspondence between functions on $\mathbb{Z}/q\mathbb{Z}$ and q-periodic functions, we can think of $\widehat{f}$ as a q-periodic function from $\mathbb{Z}$ to $\mathbb{C}$, defined for any q-periodic function $f : \mathbb{Z} \to \mathbb{C}$.

Of particular importance is the interaction between additive and multiplicative characters. Recall that a Dirichlet character mod q is a q-periodic function. Hence, it has an additive Fourier transform mod q given by

$$\widehat{\chi}(a) = \frac{1}{q} \sum_{1 \leqslant n \leqslant q} \chi(n)e(-an/q).$$

A quantity that plays a key role in the study of $\widehat{\chi}$ is the *Gauss sum*

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

The multiplicativity of χ implies the relation

$$\sum_{1 \leqslant n \leqslant q} \chi(n) e(an/q) = \mathcal{G}(\chi)\overline{\chi}(a) \quad \text{whenever } (a,q) = 1. \tag{10.8}$$

This follows simply by making the change of variables $m = an$, which is invertible mod q when $(a,q) = 1$.

Setting $\lambda_\chi = \overline{\chi}(-1)\mathcal{G}(\chi)/q = \chi(-1)\mathcal{G}(\chi)/q$, relation (10.8) can also be written as $\widehat{\chi}(n) = \lambda_\chi \cdot \overline{\chi}(n)$ whenever $(n,q) = 1$, which evaluates the additive Fourier transform of χ at frequencies n that are coprime to q. In the next section we shall see that this formula can be expanded to all frequencies n for an important class of Dirichlet characters called primitive characters. That is to say, when χ is a primitive Dirichlet character mod q, we will show that $\widehat{\chi} = \lambda_\chi \cdot \overline{\chi}$. This demonstrates that primitive characters are conjugate eigenvectors of the additive Fourier transform mod q.

Primitive characters

Each Dirichlet character χ mod q naturally generates a new Dirichlet character ξ at every modulus m that is a multiple of q via the relation

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

We then say that χ *induces* ξ. Inverting our point of view, we also say that ξ is a *lift* of χ.

Given a character χ mod q, a natural question is whether it is the lift of some character ψ mod d with d a proper divisor of q. If this is the case, we say that χ is *imprimitive* and call the smallest such d the *conductor* of χ. On the other hand, if such a character ψ does not exist, then we say that χ is *primitive*. In the latter case, we define the conductor of χ to simply be its modulus q.

For example, the principal character χ_0 mod q is induced by the principal character mod 1, that is to say, the constant function 1 on $\mathbb{Z}$. Therefore, if $q > 1$, then χ_0 is imprimitive and has conductor 1.

Being an imprimitive character χ mod q means that there is a proper divisor of q (i.e., $d < q$) that is a period of χ in the sense of Exercise 9.6. In particular, the above definition of the conductor agrees with the one given in Exercise 9.6(c), as the following lemma shows.

Lemma 10.2. *Let χ be a Dirichlet character* mod q*. Then, χ is imprimitive if and only if there is a proper divisor d of q such that $\chi(m) = \chi(n)$ whenever $m \equiv n \pmod{d}$ and $(mn, q) = 1$.*

Proof. It is clear that if χ is induced by a character mod d, then $\chi(m) = \chi(n)$ whenever $(mn, q) = 1$ and $m \equiv n \pmod{d}$. Let us now prove that the converse statement is also true.

We define a function $\psi : \mathbb{Z} \to \mathbb{C}$ as follows: if $(a, d) > 1$, we set $\psi(a) = 0$. On the other hand, if $(a, d) = 1$, we note that there is some $k \in \mathbb{Z}$ such that $(a + kd, q) = 1$. We then define $\psi(a) = \chi(a + kd)$, which is independent of the choice of k in virtue of our assumption on χ. The function ψ is clearly a character mod d inducing χ, thus proving that χ is imprimitive. □

Using the above lemma, we prove the following fundamental property of primitive characters.

Theorem 10.3. *Let χ be a primitive Dirichlet character* mod q*. For all $n \in \mathbb{Z}$, we have*

$$\chi(n) = \frac{1}{\mathcal{G}(\overline{\chi})} \sum_{1 \leqslant a \leqslant q} \overline{\chi}(a) e(an/q).$$

Proof. When $(n, q) = 1$, this follows by (10.8). Assume now that $(n, q) = m > 1$, in which case we need to show that

$$\sum_{1 \leqslant a \leqslant q} \overline{\chi}(a) e(an/q) = 0.$$

Write $n = \ell m$ and $q = dm$, so that $(\ell, d) = 1$, and note that

$$\begin{aligned} \sum_{1 \leqslant a \leqslant q} \overline{\chi}(a) e(an/q) &= \sum_{j=1}^{m} \sum_{b=1}^{d} \overline{\chi}(b + dj) e((b + dj)\ell/d) \\ &= \sum_{b=1}^{d} e(b\ell/d) \sum_{j=0}^{m-1} \overline{\chi}(b + dj). \end{aligned}$$

So it suffices to show that

$$\sum_{j=0}^{m-1} \overline{\chi}(b + dj) = 0 \tag{10.9}$$

for all $b \in \{1, 2, \dots, d\}$. Since χ is primitive, there is some $j_0 \in \mathbb{Z}$ for which the number $r = 1 + j_0 d$ satisfies the relations $(r, q) = 1$ and $\chi(r) \neq 1$. (To see this, combine Lemma 10.2 and Exercise 9.6(a).) In particular, when

reduced mod q, the numbers $r \cdot (b + dj)$ with $0 \leqslant j <$ are a permutation of the numbers $b + dj$ with $0 \leqslant j < m$. Consequently,

$$\overline{\chi}(r) \sum_{j=0}^{m-1} \overline{\chi}(b + dj) = \sum_{j=0}^{m-1} \overline{\chi}(r(b + dj)) = \sum_{j=0}^{m-1} \overline{\chi}(b + dj).$$

Since $\chi(r) \neq 1$, relation (10.9) follows. This completes the proof of the theorem. □

The above theorem and Parseval's formula (10.6) allow us to determine the size of the Gauss sum for primitive Dirichlet characters.

Theorem 10.4. *If χ is a primitive Dirichlet character* mod q*, then*

$$|\mathcal{G}(\chi)| = \sqrt{q}.$$

Character sums

Notice in Theorem 10.4 that, even though $\mathcal{G}(\chi)$ is defined as a sum of $\varphi(q)$ complex numbers on the unit circle, its modulus equals $\sqrt{q}$, which is approximately the square-root of $\varphi(q)$ (see Corollary 3.6). This means that the numbers $\chi(n)e(n/q)$ are sufficiently randomly placed around the unit circle so that they annihilate each other when added all together. This kind of "square-root cancellation" is typical for averages involving Dirichlet characters. We demonstrate it in two other settings.

We start by showing that Theorem 10.3 can be used to show a generalization of the Poisson summation formula (Theorem B.3).

Theorem 10.5. *Let $f \in C^2(\mathbb{R})$ such that $f^{(j)}(x) \ll 1/x^2$ for $j \in \{0, 1, 2\}$ and $|x| \geqslant 1$, so that $\widehat{f}(\xi) \ll 1/\xi^2$ for $|\xi| \geqslant 1$.*

If χ is a primitive character mod q *and $N \in \mathbb{R}_{>0}$, then*

$$\sum_{n \in \mathbb{Z}} \chi(n) f(n/N) = \frac{\chi(-1)N}{\mathcal{G}(\overline{\chi})} \sum_{n \in \mathbb{Z}} \overline{\chi}(n) \widehat{f}(nN/q).$$

Proof. It suffices to prove the theorem when $N = 1$. The general case follows by noticing that the Fourier transform of $x \to f(x/N)$ is the function $\xi \to N\widehat{f}(N\xi)$.

We use Theorem 10.3 to write χ in terms of its additive Fourier expansion and find that

$$\begin{aligned} \sum_{n \in \mathbb{Z}} \chi(n) f(n) &= \frac{1}{\mathcal{G}(\overline{\chi})} \sum_{n \in \mathbb{Z}} f(n) \sum_{1 \leqslant a \leqslant q} \overline{\chi}(a) e(an/q) \\ &= \frac{1}{\mathcal{G}(\overline{\chi})} \sum_{1 \leqslant a \leqslant q} \overline{\chi}(a) \sum_{n \in \mathbb{Z}} f(n) e(an/q). \end{aligned}$$

We then apply the Poisson summation formula (Theorem B.3) to the function $g(x) = f(x)e(ax/q)$, whose Fourier transform is $\widehat{g}(\xi) = \widehat{f}(\xi - a/q)$. Thus

$$\sum_{n\in\mathbb{Z}} \chi(n)f(n) = \frac{1}{\mathcal{G}(\overline{\chi})} \sum_{1\leqslant a\leqslant q} \overline{\chi}(a) \sum_{n\in\mathbb{Z}} \widehat{f}(n - a/q)$$
$$= \frac{\chi(-1)}{\mathcal{G}(\overline{\chi})} \sum_{1\leqslant a\leqslant q} \overline{\chi}(-a) \sum_{n\in\mathbb{Z}} \widehat{f}((qn - a)/q).$$

Since $\chi(-a) = \chi(qn - a)$ for all $n \in \mathbb{Z}$, the theorem follows. □

If we let $N \in [1, q]$ in Theorem 10.5 and we assume that $\mathrm{supp}(f) \subseteq [0, 1]$, then the sum $\sum_{n\in\mathbb{Z}} \chi(n)f(n/N)$ is supported on integers $n \in [0, N]$. Since $\widehat{f}(\xi) \ll 1/|\xi|^2$ for $|\xi| \geqslant 1$, the dominant contribution to the *dual sum* $\sum_{n\in\mathbb{Z}} \widehat{f}(Nn/q)$ comes from integers $n = O(q/N)$. Hence, roughly speaking, Theorem 10.5 transforms a sum of length $\asymp N$ to a sum of length $\asymp q/N$. In particular, if $N > \sqrt{q}$, then the new sum is shorter, so that bounding it trivially yields a non-trivial bound on the sum we began with.

More precisely, since $\widehat{f}(\xi) \ll \min\{1, 1/|\xi|^2\}$ for all ξ, the sum on the right-hand side of Theorem 10.5 is

$$\sum_{n\in\mathbb{Z}} \widehat{f}(Nn/q)\chi(n) \ll \sum_{|n|\leqslant q/N} 1 + \sum_{|n|>q/N} \frac{1}{(Nn/q)^2} \ll q/N.$$

Together with Theorems 10.4 and 10.5, this implies that

$$\sum_{n\in\mathbb{Z}} f(n/N)\chi(n) \ll \frac{N}{\sqrt{q}} \cdot \frac{q}{N} = \sqrt{q}, \tag{10.10}$$

another occurrence of square-root cancellation, at least when $N \asymp q$.

In practice, we often need an estimate like (10.10) but with the integers $n \in [0, N]$ weighted by 1 and not by the smooth weight $f(n/N)$. Such an estimate is provided by the *Pólya-Vinogradov inequality*, which we prove using a variation of the argument leading to Theorem 10.5.

Theorem 10.6 (Pólya-Vinogradov inequality)**.** *Let χ be a non-principal character* mod q*. For $M \in \mathbb{R}$ and $N \in \mathbb{R}_{\geqslant 0}$, we have*

$$\sum_{M<n\leqslant M+N} \chi(n) \ll \sqrt{q}\log q.$$

Proof. Since χ is non-principal, we must have that $q > 1$. We may also assume that $M, N \in \mathbb{Z}$. First, we prove the theorem when χ is primitive. By Theorem 10.3 and the periodicity of χ, we have

$$\chi(n) = \frac{1}{\mathcal{G}(\overline{\chi})} \sum_{-q/2<a\leqslant q/2} \overline{\chi}(a)e(an/q).$$

Summing the above formula over $n \in (M, M+N]$, we find

$$\sum_{M<n\leqslant M+N} \chi(n) = \frac{1}{\mathcal{G}(\overline{\chi})} \sum_{-q/2<a\leqslant q/2} \overline{\chi}(a) \sum_{M<n\leqslant M+N} e(an/q). \tag{10.11}$$

Notice that we may assume that $a \neq 0$, since $\chi(0) = 0$. The sum of $e(an/q)$ over $n \in (M, M+N]$ is a geometric series with ratio of consecutive terms $e(a/q)$. We thus have

$$\Big| \sum_{M<n\leqslant M+N} e(an/q) \Big| = \left| \frac{1-e(Na/q)}{1-e(a/q)} \right| \leqslant \frac{1}{2|a/q|} \tag{10.12}$$

for $1 \leqslant |a| \leqslant q/2$, since $|1-e(x)| = |e(-x/2)-e(x/2)| = 2|\sin(\pi x)| \geqslant 4|x|$ for $x \in [-1/2, 1/2]$. Combining (10.11) and (10.12) with the fact that $|\mathcal{G}(\overline{\chi})| = \sqrt{q}$, we conclude that

$$\Big| \sum_{M<n\leqslant M+N} \chi(n) \Big| \leqslant \frac{2\sqrt{q}}{\pi} \sum_{1\leqslant |a|\leqslant q/2} \frac{1}{|a|} \leqslant \frac{4\sqrt{q}}{\pi}(1+\log(q/2)).$$

This completes the proof in the case that χ is primitive.

Finally, if χ is induced by the primitive character $\psi \pmod d$ with $d|q$, then

$$\chi(n) = \psi(n) 1_{(n,q)=1} = \psi(n) 1_{(n,q/d)=1}$$

because ψ is supported on integers coprime to d. Möbius inversion then implies that

$$\begin{aligned} \sum_{M<n\leqslant M+N} \chi(n) = \sum_{\substack{M<n\leqslant M+N \\ (n,q/d)=1}} \psi(n) &= \sum_{M<n\leqslant M+N} \psi(n) \sum_{a|(n,q/d)} \mu(a) \\ &= \sum_{a|q/d} \mu(a) \sum_{\substack{M<n\leqslant M+N \\ a|n}} \psi(n). \end{aligned}$$

In the inner sum, we write $n = ma$ and note that $\psi(n) = \psi(a)\psi(m)$, so that $\psi(a)$ can be factored outside the summation. Applying the Pólya-Vinogradov inequality to the primitive character ψ yields the estimate

$$\sum_{M<n\leqslant M+N} \chi(n) \ll \sum_{a|q/d} \sqrt{d} \log d.$$

Every divisor a of q/d comes with a complementary divisor $(q/d)/a$. At least one of these divisors is $\leqslant \sqrt{q/d}$, so that the total number of permissible values of a is $\leqslant 2\sqrt{q/d}$. (Or simply use the bound in Exercise 2.9(f).) This completes the proof in this case as well. □

Remark 10.7. Comparing the right-hand side in Theorem 10.6 with that in (10.10), we see we have an extra logarithm. This is caused by the fact that we have replaced the smooth cut-off $f(n/N)$ by the sharp cut-off $1_{(M,M+N]}(n)$.

Indeed, in the proof of Theorem 10.6, the sum $\sum_{M<n\leqslant M+N} e(\alpha n)$ with $\alpha = a/q \in [-1/2, 1/2]$ decays like $1/|\alpha|$. Had we weighted the integers $n \in (M, M+N]$ smoothly, we would have had faster decay, say $\ll 1/|\alpha|^2$.

Superficially, this extra logarithm seems like a technical and insignificant matter. After all, it is of negligible size compared to $\sqrt{q}$. However, improving upon the Pólya-Vinogradov inequality is very hard and is related to some very deep conjectures about Dirichlet characters. Paley [149] showed the existence of an infinite set of primitive quadratic characters χ for which

$$M(\chi) := \sup_{x\in[1,q]} \Big|\sum_{n\leqslant x} \chi(n)\Big| \gg \sqrt{q}\log\log q,$$

with q denoting the conductor of χ. On the other hand, Montgomery and Vaughan proved that $M(\chi) \ll \sqrt{q}\log\log q$ assuming a suitable generalization of the Riemann Hypothesis called the *Generalized Riemann Hypothesis* that we will discuss in the next chapter.

Remarkably, when χ has odd order g as an element of the group of Dirichlet characters mod q, Granville and Soundararajan [72] showed that the Pólya-Vinogradov inequality can be improved. Their results were sharpened by Goldmakher [61], who established the estimate $M(\chi) \ll_{g,\theta} \sqrt{q}(\log q)^\theta$ for each fixed $\theta > (g/\pi)\sin(\pi/g)$. A further improvement of this result was announced more recently by Lamzouri and Mangerel [123]. □

Exercises

Exercise 10.1. Show that the function

$$f(q) = \sum_{n\in(\mathbb{Z}/q\mathbb{Z})^*} e(n/q)$$

is multiplicative and calculate it.

Exercise 10.2.

(a) Calculate all primitive characters mod 3, 4, 5, 8 and 15.

(b) If q is a prime, then show that there are $q-2$ primitive characters.

(c) Show that a faithful character, defined in Exercise 9.5, is also primitive.

(d) Calculate all primitive and all faithful characters when q is the product of two distinct primes, and when $q = p^2$ with p prime.

(e) If $\mathcal{C}_q$ denotes the set of Dirichlet characters mod q, and $\mathcal{C}_q^*$ denotes the set of primitive elements of $\mathcal{C}_q$, then calculate $|\mathcal{C}_q^*|$. [*Hint:* Prove that $|\mathcal{C}_q| = \sum_{d|q} |\mathcal{C}_d^*|$.]

(f) Let $\chi = \chi_1\chi_2$ be as in Exercise 9.4(a) with $(q_1, q_2) = 1$. Show that χ is primitive if and only if χ_1 and χ_2 are primitive.

(g) If χ is a real primitive character mod q, then show that $q = 2^k q'$, where $k \in \{0, 2, 3\}$ and q' is odd and square-free.[1] Moreover, if $k \in \{0, 2\}$, then there is exactly one real primitive character mod q, whereas if $k = 3$, then there are exactly two real primitive characters mod q.

Exercise 10.3. Given two primitive characters $\chi, \psi \,(\mathrm{mod}\, q)$ and an integer $a \in \mathbb{Z}$ coprime to q, show that

$$\sum_{n \in \mathbb{Z}/q\mathbb{Z}} \chi(n+a)\overline{\psi}(n) = \frac{\chi(-a)\overline{\psi}(-a)\mathcal{G}(\chi)\mathcal{G}(\overline{\chi}\psi)}{\mathcal{G}(\psi)}.$$

Simplify the above expression when $\psi = \chi$. [*Hint:* Use Exercise 10.1.]

Exercise 10.4.* Given a non-principal character χ mod q and $N \in [1, q/2] \cap \mathbb{Z}$, show that

$$\sum_{|n| \leqslant N} \chi(n)(1 - |n|/N) \ll \sqrt{q}.$$

[*Hint:* $\sum_{|n| \leqslant N}(1 - |n|/N)e(\alpha n) = N^{-1}(\sin(N\pi\alpha)/\sin(\pi\alpha))^2$.]

Exercise 10.5.* If $\chi \,(\mathrm{mod}\, q)$ is induced by $\psi \,(\mathrm{mod}\, d)$, then prove that[2]

$$\mathcal{G}(\chi) = \mu(m)\psi(m)\mathcal{G}(\psi) \tag{10.13}$$

with $m = q/d$, as follows:

(a) For any $k \in \mathbb{Z}$, show that

$$\mathcal{G}(\chi) = \sum_{\substack{1 \leqslant a \leqslant q \\ (a+kd,q)=1}} \psi(a)e((a+kd)/q),$$

and conclude that

$$\mathcal{G}(\chi) = \frac{1}{m} \sum_{1 \leqslant a \leqslant q} \psi(a)e(a/q) \sum_{\substack{1 \leqslant k \leqslant m \\ (a+kd,q)=1}} e(k/m).$$

(b) When $(a, d) = 1$, show that

$$\sum_{\substack{1 \leqslant k \leqslant m \\ (a+kd,q)=1}} e(k/m) = 1_{(d,m)=1}\mu(m)e(-a\bar{d}/m),$$

where $\bar{d}$ is the multiplicative inverse of $d \,(\mathrm{mod}\, m)$.

(c) When $(d, m) > 1$, show that both sides of (10.13) are zero.

(d) Assume that $(d, m) = 1$. If $\overline{m}$ denotes the multiplicative inverse of $m \,(\mathrm{mod}\, d)$, show that $\overline{d}/m + \overline{m}/d \equiv 1/q \,(\mathrm{mod}\, 1)$, and complete the proof of (10.13).

[1] There is an important connection between real Dirichlet characters and the theory of binary quadratic forms, presented in Chapters 5 and 6 of Davenport's book [31].

[2] See [31, p. 67] for an alternative proof.

Chapter 11

Dirichlet L-functions

We now turn to the study of the infinite series $L(s,\chi) = \sum_{n=1}^{\infty} \chi(n)/n^s$, namely the L-function corresponding to the Dirichlet character χ. Since $|\chi| \leqslant 1$, this series converges absolutely when $\mathrm{Re}(s) > 1$, and for such s we have the Euler product representation

$$L(s,\chi) = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}. \tag{11.1}$$

When $\chi \neq \chi_0$, the summatory function $\sum_{n\leqslant x} \chi(n)$ is uniformly bounded by the Pólya-Vinogradov inequality. Partial summation then implies that $L(s,\chi)$ converges conditionally in the half-plane $\mathrm{Re}(s) > 0$. It is easily seen that $L(s,\chi)$ diverges at $s = 0$, so that $L(s,\chi)$ has abscissa of convergence 0 (but abscissa of absolute convergence 1).

We will show that $L(s,\chi)$ can be extended to an entire function. Moreover, we will prove that it enjoys various special properties similar to the ones possessed by ζ.

The analytic continuation and the functional equation

Notice that if $\chi \pmod q$ is induced by the character $\psi \pmod d$, then

$$L(s,\chi) = \prod_{p\nmid q} \left(1 - \frac{\psi(p)}{p^s}\right)^{-1} = L(s,\psi) \prod_{p\mid q} \left(1 - \frac{\psi(p)}{p^s}\right). \tag{11.2}$$

Because of this relation, the properties of $L(s,\psi)$ transfer (with appropriate modifications) to $L(s,\chi)$, so we often restrict our attention to the study of Dirichlet L-functions attached to primitive characters whose theory is

simpler. In particular, Dirichlet L-functions attached to primitive characters satisfy a functional equation analogous to (6.1) for the Riemann zeta function.

The functional equation of $L(s,\chi)$ changes slightly according to the value of $\chi(-1)$. Characters with $\chi(-1)=1$ are called *even*, whereas characters with $\chi(-1)=-1$ are called *odd*. We then take

$$a = \begin{cases} 0 & \text{when } \chi \text{ is even,} \\ 1 & \text{when } \chi \text{ is odd,} \end{cases} \tag{11.3}$$

and we introduce the so-called *completed L-function*

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

which is analogous to the function $\xi(s)$ that we defined in Exercise 8.7. The functional equation for $L(s,\chi)$ also involves the quantity

$$\varepsilon(\chi) := \frac{\mathcal{G}(\chi)}{i^a\sqrt{q}},$$

called its *root number*. Note that $|\varepsilon(\chi)|=1$ in virtue of Theorem 10.4. Furthermore, it is easy to check that $\varepsilon(\overline{\chi})=1/\varepsilon(\chi)$.

Theorem 11.1. *Let χ be a primitive, non-principal character. The functions $L(s,\chi)$ and $\xi(s,\chi)$ can be continued analytically to the entire complex plane. Moreover, for all $s\in\mathbb{C}$, their extensions satisfy the functional equation*

$$\xi(s,\chi) = \varepsilon(\chi)\cdot\xi(1-s,\overline{\chi}).$$

Proof. The key to the proof of this theorem is the variation of the Poisson summation formula given in Theorem 10.5. The argument is very similar to the one leading to (6.1), so we only sketch it.

We take $f(x)=2x^a e^{-\pi x^2}$, which has the same parity as χ, i.e., $f(-x)=\chi(-1)f(x)$, and note that its Mellin transform is

$$F(s)=\int_0^\infty f(y)y^{s-1}\mathrm{d}y = \pi^{-(s+a)/2}\Gamma\Big(\frac{s+a}{2}\Big).$$

Arguing as in (6.4), we find

$$q^{-a/2}\xi(s,\chi) = q^{s/2}L(s,\chi)F(s) = \int_0^\infty \sum_{n=1}^\infty \chi(n)f(ny/\sqrt{q})y^{s-1}\mathrm{d}y \tag{11.4}$$

for $\mathrm{Re}(s)>1$. Since $\chi(0)=0$, and f and χ have the same parity, we have

$$S_\chi(y) := \sum_{n=1}^\infty \chi(n)f(ny/\sqrt{q}) = \frac{1}{2}\sum_{n\in\mathbb{Z}}\chi(n)f(ny/\sqrt{q}).$$

We apply Theorem 10.5 to the function f with $N = \sqrt{q}/y$ to find that

$$S_\chi(y) = \frac{\chi(-1)\sqrt{q}}{2y\mathcal{G}(\overline{\chi})} \sum_{n\in\mathbb{Z}} \overline{\chi}(n)\widehat{f}(ny^{-1}/\sqrt{q}).$$

For our choice of f, we note that $\widehat{f} = \chi(-1)i^a f$ so that

$$S_\chi(y) = \frac{S_{\overline{\chi}}(1/y)}{\varepsilon(\overline{\chi})y} = \frac{\varepsilon(\chi)S_{\overline{\chi}}(1/y)}{y}.$$

We insert the above transformation formula into the part of the integral over $y \in (0,1)$ in (11.4). As for the Riemann zeta function, this proves at the same time that $\xi(s,\chi)$ can be extended to an entire function, as well as that it satisfies the claimed functional equation. We leave the verification of the details of this claim as an exercise. Finally, we note that, since Γ does not vanish anywhere, $L(s,\chi)$ itself extends to an entire function. □

When χ is a primitive, non-principal character, Theorem 11.1 allows us to obtain some information regarding the location of the zeroes of $L(s,\chi)$ which, in view of the explicit formula (9.6), rule the distribution of primes in arithmetic progressions. When $\mathrm{Re}(s) > 1$, the Euler product representation (11.1) implies that $L(s,\chi)$ does not vanish. Thus neither does $\xi(s,\chi)$ and, by the functional equation, we also have $\xi(s,\chi) \neq 0$ for $\mathrm{Re}(s) < 0$. Since $\Gamma((s+a)/2)$ has simple poles at the points $-2n-a$, $n \in \mathbb{Z}_{\geqslant 0}$, we deduce that $L(s,\chi)$ must have simple zeroes at -1, -3, -5, ... when χ is odd, and at -2, -4, -6, ... when χ is even, but no other zeroes when $\mathrm{Re}(s) < 0$. Moreover, $L(0,\chi) = 0$ when $\chi(-1) = 1$. In Theorem 12.8, we will show that $L(1,\chi) \neq 0$, which implies that 0 must be a simple zero of $L(s,\chi)$ when χ is even, and that $L(0,\chi) \neq 0$ when χ is odd.

The zeroes of $L(s,\chi)$ at the points $-2n-a$ with $n \in \mathbb{Z}_{\geqslant 0}$ are called the *trivial zeroes* of $L(s,\chi)$. All other zeroes of $L(s,\chi)$, which are in correspondence with the zeroes of $\xi(s,\chi)$ and necessarily lie in the critical strip $0 \leqslant \mathrm{Re}(s) \leqslant 1$, are called *non-trivial*. They are usually denoted by $\rho = \beta + i\gamma$, where $0 \leqslant \beta \leqslant 1$, or by $\rho_\chi = \beta_\chi + i\gamma_\chi$ when we want to underline their dependence on χ. We note that the functional equation and the obvious symmetry $\overline{L(s,\chi)} = L(\overline{s},\overline{\chi})$ imply that if $\rho = \beta + i\gamma$ is a non-trivial zero of $L(s,\chi)$, then so is $1 - \overline{\rho} = 1 - \beta + i\gamma$. It is widely believed that an extension of the Riemann Hypothesis holds, often called the *Generalized Riemann Hypothesis*. This conjecture postulates that all non-trivial zeroes of $L(s,\chi)$ lie on the critical line $\mathrm{Re}(s) = 1/2$.

Finally, we consider the case when χ is an imprimitive character induced by $\psi \pmod d$. Recall the factorization (11.2). In particular, all zeroes of $L(s,\psi)$ are also zeroes of $L(s,\chi)$. Notice that $L(s,\chi)$ might have some additional zeroes, at points s with $p^s = \psi(p)$ for some $p|q$. All such zeroes

are on the line $\mathrm{Re}(s) = 0$ and we consider them to be trivial zeroes of $L(s,\chi)$, together with the trivial zeroes of $L(s,\psi)$ at $s = -2n - a$ (with the caveat that $s = 0$ is excluded if $\psi = 1$, that is to say, if χ is principal). All other zeroes of $L(s,\chi)$ are considered non-trivial; the summation in (9.6) runs over them.

Bounds for $L(s,\chi)$

As in the case of the Riemann zeta function, it is very useful to have bounds on Dirichlet L-functions. By the functional equation in Theorem 11.1, we may restrict our attention to the half-plane $\mathrm{Re}(s) \geqslant 1/2$. We could use Theorem 6.2, but we present instead a different method that is simpler and thus more flexible, even though it yields weaker results. For the application of Theorem 6.2 to $L(s,\chi)$, see Exercise 11.1.

Lemma 11.2. *Let χ be a non-principal Dirichlet character* $\bmod q$. *For $j \in \mathbb{Z}_{\geqslant 0}$, $s = \sigma + it$ with $1/2 \leqslant \sigma \leqslant 2$, we have*

$$|L^{(j)}(s,\chi)| \ll_j Q^{\max\{0,1-\sigma\}}(\log Q)^{j+1} \qquad \text{with} \quad Q = \sqrt{q}(|t|+2).$$

Proof. We estimate $L^{(j)}(s,\chi) = \sum_{n=1}^{\infty} \chi(n)(-\log n)^j/n^s$ by inserting the Pólya-Vinogradov bound on $\sum_{n\leqslant x}\chi(n)$ via partial summation. However, we have to be careful because partial summation yields poor bounds for small n. There are two reasons for this: firstly, the function $x \to x^s$ oscillates a lot for small x, with its derivative sx^{s-1} getting under control only for $x > |s|$. Secondly, the Pólya-Vinogradov bound is non-trivial only for character sums of length $> \sqrt{q}\log q$. For these reasons, we bound the summands with small n trivially, noticing that

$$\left|\sum_{n=1}^{N} \frac{\chi(n)(\log n)^j}{n^s}\right| \leqslant N^{\max\{0,1-\sigma\}}(\log N)^j \sum_{n=1}^{N}\frac{1}{n} \ll N^{\max\{0,1-\sigma\}}(\log N)^{j+1}.$$

To the terms with $n > N$, we apply partial summation:

$$\begin{aligned}\sum_{n>N}\frac{\chi(n)(\log n)^j}{n^s} &= \int_N^{\infty}\frac{(\log y)^j}{y^s}\mathrm{d}\sum_{N<n\leqslant y}\chi(n)\\ &= \int_N^{\infty}\sum_{N<n\leqslant y}\chi(n)\frac{s(\log y)^j - j(\log y)^{j-1}}{y^{s+1}}\mathrm{d}y.\end{aligned}$$

The Pólya-Vinogradov inequality then yields the estimate

$$\begin{aligned}\sum_{n>N}\frac{\chi(n)(\log n)^j}{n^s} &\ll_j |s|\sqrt{q}(\log q)\int_N^{\infty}\frac{(\log y)^j}{y^{\sigma+1}}\mathrm{d}y\\ &\ll |s|\sqrt{q}(\log q)(\log N)^j N^{-\sigma}.\end{aligned}$$

Taking $N = 1 + \lfloor |s|\sqrt{q}\rfloor$ completes the proof. □

Proving the explicit formula for $L(s,\chi)$

We conclude this chapter with a proof of the explicit formula (9.6) for $L(s,\chi)$, which we restate in a slightly more general form.

Theorem 11.3. *For $x \geqslant T \geqslant 2$ and $\chi \pmod q$, we have*

$$\psi(x,\chi) = 1_{\chi=\chi_0}x - \sum_{|\gamma|\leqslant T} \frac{x^\rho - 1}{\rho} + O\Big(\frac{x\log^2(xq)}{T}\Big). \tag{11.5}$$

We start with a technical lemma, analogous to Lemma 8.2.

Lemma 11.4. *Consider $s = \sigma + it$ and a primitive, non-principal Dirichlet character χ mod $q \geqslant 3$. Furthermore, let $a \in \{0,1\}$ be as in (11.3). All implied constants below are absolute. Moreover, the zeroes of $L(s,\chi)$ are listed and counted with their multiplicity.*

(a) *There are $\ll \log\big[q(|t|+1)\big]$ non-trivial zeroes of $L(s,\chi)$ with $|\gamma - t| \leqslant 1$.*

(b) *Assume that $|s-z| \geqslant 1/2$ for all trivial zeroes z of $L(s,\chi)$. Then*

$$\frac{L'}{L}(s,\chi) = \sum_{|\gamma-t|\leqslant 1} \frac{1}{s-\rho} + O\big(\log\big[q(|s|+1)\big]\big).$$

Proof. Lemma 11.2 implies the bound $L(s,\chi) \ll (q(|s|+1))^2$ when $1/2 \leqslant \mathrm{Re}(s) \leqslant 3/2$ (in fact, it implies a better bound, but this will be sufficient). In addition, we have $|L(s,\chi)| \leqslant \zeta(3/2) = O(1)$ when $\mathrm{Re}(s) \geqslant 3/2$. By the functional equation of $L(s,\chi)$ and Exercise 1.12(a), this "polynomial growth" of $L(s,\chi)$ in $q(|s|+1)$ can be extended to the half-plane $\mathrm{Re}(s) \geqslant -10$. Employing Lemma 8.6(b) as in the proof of Lemma 8.2 yields part (a). Similarly, we also obtain part (b) when $\sigma \geqslant -10$. Finally, when $\sigma \leqslant -10$, we note that, analogously to (8.9), we have $(L'/L)(s,\chi) = -(L'/L)(1-s,\overline{\chi}) + O\big(\log(q(|s|+1))\big)$. Part (b) follows in this case as well by the trivial bound $(L'/L)(1-s,\overline{\chi}) = O(1)$. □

Proof of Theorem 11.3. First of all, note that

$$\sum_{n\leqslant x,\,(n,q)>1} \Lambda(n) \leqslant \sum_{p|q,\,k\geqslant 1\,:\,p^k\leqslant x} \log p \leqslant \sum_{p|q} \log x \ll (\log q)(\log x). \tag{11.6}$$

Hence, if $\xi \pmod m$ is the primitive character inducing $\chi \pmod q$, then

$$\psi(x,\chi) = \psi(x,\xi) + O((\log q)(\log x)). \tag{11.7}$$

Since $L(s,\chi)$ and $L(s,\xi)$ share the same non-trivial zeroes, (11.5) follows for χ if we can prove it for ξ. This means that, without loss of generality, we may assume that χ is primitive.

In addition, as in the proof of (5.11), we may reduce the proof to the case when T is such that $|T-\gamma| \gg 1/\log x$ for all zeroes of $L(s,\chi)$.

We let $\alpha = 1 + 1/\log x$ and use Theorem 7.2 to write

$$\psi(x,\chi) = \frac{-1}{2\pi i}\int_{\substack{\mathrm{Re}(s)=\alpha\\ |\mathrm{Im}(s)|\leqslant T}} \frac{L'}{L}(s,\chi)\frac{x^s}{s}\mathrm{d}s + O\bigg(\frac{x(\log x)^2}{T} + \log x\bigg).$$

Due to the potential presence of zeroes of $L(s,\chi)$ at or close to $s = 0$, it is convenient to modify the integrand in the above formula and remove the pole at the origin from the factor x^s/s: using Lemma 7.1, we see that $\int_{\mathrm{Re}(s)=\alpha,\,|\mathrm{Im}(s)|\leqslant T} n^{-s}s^{-1}\mathrm{d}s = O(1/(Tn^\alpha\log n))$ for $n \geqslant 2$. Consequently,

$$\psi(x,\chi) = \frac{-1}{2\pi i}\int_{\substack{\mathrm{Re}(s)=\alpha\\ |\mathrm{Im}(s)|\leqslant T}} \frac{L'}{L}(s,\chi)\frac{x^s-1}{s}\mathrm{d}s + O\bigg(\frac{x(\log x)^2}{T} + \log x\bigg).$$

Similarly to the proof of Theorem 5.1 in Chapter 8, we move the contour to the line $\mathrm{Re}(s) = -N - 1/2$ with N a large integer, picking up contributions from the zeroes of $L(s,\chi)$ (and the pole at $s = 1$ if $\chi = 1$). In the case when $\chi = 1$, we use Lemma 11.4 to control the logarithmic derivative of $L(s,\chi) = \zeta(s)$ on the new contour; otherwise, we use Lemma 8.2. We leave the details as an exercise. □

Exercises

Exercise 11.1. Fix $\varepsilon > 0$ and $C \geqslant 1$. For all primitive, non-principal Dirichlet characters $\chi \,(\mathrm{mod}\, q)$ and all $s = \sigma + it$ with $\sigma \geqslant -C$, show that

$$L(s,\chi) \ll_{\varepsilon,C} \begin{cases} 1 & \text{if } \sigma \geqslant 1+\varepsilon,\\ [q(|t|+2)]^{(1-\sigma+\varepsilon)/2} & \text{if } -\varepsilon \leqslant \sigma \leqslant 1+\varepsilon,\\ [q(|t|+2)]^{1/2-\sigma} & \text{if } -C \leqslant \sigma \leqslant -\varepsilon. \end{cases}$$

Exercise 11.2. Assuming the Generalized Riemann Hypothesis, prove that:

(a) $\psi(x,\chi) = 1_{\chi=\chi_0}x + O(x^{1/2}\log^2(qx)) \qquad (x \geqslant 1,\ \chi \,(\mathrm{mod}\, q))$.

(b) $\psi(x;q,a) = x/\varphi(q) + O(x^{1/2}\log^2(qx)) \qquad (x \geqslant 1,\ a \in (\mathbb{Z}/q\mathbb{Z})^*)$.

Exercise 11.3. Let χ be a primitive, non-principal Dirichlet character.

(a) Show that the Riemann Hypothesis for $L(s,\chi)$ is equivalent to knowing that $\mathrm{Re}(\rho) \leqslant 1/2$ for all non-trivial zeroes ρ of $L(s,\chi)$.

(b) Show that the Riemann Hypothesis for $L(s,\chi)$ is equivalent to knowing that for each fixed $\varepsilon > 0$ we have $\psi(x,\chi) \ll_{\varepsilon,\chi} x^{1/2+\varepsilon}$ uniformly for $x \geqslant 1$.

(c) Show that $L(s,\chi)$ must have infinitely many non-trivial zeroes.

(d) Fix $\theta < 1/2$. Show that we *cannot* have $\psi(x,\chi) \ll_\chi x^\theta$ for all $x \geqslant 1$.

Exercise 11.4. Let χ be a primitive, non-principal character mod q.

(a) Let ϕ and Φ be as in Exercise 7.2(b), $s \in \mathbb{C}$ and $x \geqslant 1$. If $\phi_0 \neq 0$, we also assume that $s \notin \{ z : L(z,\chi) = 0 \}$. Then

$$\sum_{n \geqslant 1} \frac{\Lambda(n)\chi(n)\phi(n/x)}{n^s} = -\phi_0 \frac{L'}{L}(s,\chi) - \sum_z x^{z-s}\Phi(z-s),$$

where z runs over all zeroes of $L(s,\chi)$ (trivial and non-trivial).

(b) Assume the Riemann Hypothesis for $L(s,\chi)$. Show that

$$\frac{L'}{L}(s,\chi),\ \log L(s,\chi) \ll_\varepsilon (\log(q|t|+q))^{2\max\{1-\sigma,0\}+\varepsilon}$$

for $\mathrm{Re}(s) \geqslant 1/2+\varepsilon$ and $\varepsilon > 0$, in two ways: firstly, use part (a) as per Exercise 8.5(c); secondly, use a suitable adaption of Theorem 6.2.

(c) Assume the Riemann Hypothesis for $L(s,\chi)$ and fix $\varepsilon > 0$. Show that

$$\sum_{n \leqslant x} \mu(n)\chi(n) \ll q^\varepsilon x^{1/2+\varepsilon} \qquad \text{for all } x \geqslant 1$$

with the implied constant depending at most on ε.

Exercise 11.5. Assume the Generalized Riemann Hypothesis. Given real numbers $x, q, T \geqslant 3$ and a residue class $a \in (\mathbb{Z}/q\mathbb{Z})^*$, we define

$$B(q,a) = \#\{ n \,(\mathrm{mod}\, q) : n^2 \equiv a \,(\mathrm{mod}\, q) \},$$

$$\theta(x;q,a) = \sum_{\substack{p \leqslant x \\ p \equiv a \,(\mathrm{mod}\, q)}} \log p, \qquad Z_T(x;q,a) = \sum_{\chi \,(\mathrm{mod}\, q)} \overline{\chi}(a) \sum_{|\gamma_\chi| \leqslant T} \frac{x^{i\gamma_\chi}}{1/2+i\gamma_\chi}.$$

(a) For $x \geqslant T^{3/2} \geqslant 2$, prove that

$$\theta(x;q,a) = \frac{x}{\varphi(q)} - \frac{\sqrt{x}}{\varphi(q)}(B(q,a) + Z_T(x;q,a)) + O_q\Big(\frac{x\log^2 x}{T}\Big).$$

[*Hint:* $\theta(x;q,a) = \psi(x;q,a) - \sum_{p \leqslant \sqrt{x},\, p^2 \equiv a \,(\mathrm{mod}\, q)} \log p + O(x^{1/3})$.]

(b) For $x \to \infty$, prove that[1]

$$\theta(x;4,3) - \theta(x;4,1) = \sqrt{x}\Bigg(1 + \sum_{|\gamma_\chi| \leqslant x^{2/3}} \frac{x^{i\gamma_\chi}}{1/2+i\gamma_\chi}\Bigg) + o(\sqrt{x}), \tag{11.8}$$

where χ is the unique non-principal character mod 4.

Exercise 11.6* ([31, Chapter 16])**.** Let $\chi \,(\mathrm{mod}\, q)$ be a primitive, non-principal character and write $N(T,\chi)$ for the number of non-trivial zeroes ρ of $L(s,\chi)$ with $|\gamma| \leqslant T$, counted with multiplicity. Throughout, T is chosen so that $L(s,\chi) \neq 0$ when $\mathrm{Im}(s) = \pm T$.

[1] If we let $x = e^u$, then $\int_0^{\log x} e^{i\gamma u}\mathrm{d}u = O(1/|\gamma|) = o(\log x)$ for $\gamma \neq 0$. Hence, we expect that the sum over zeroes in (11.8) is $o(1)$ on average over x. The presence of the term 1 on the right-hand side of (11.8) then means that most of the time we have $\theta(x;4,3) \geqslant \theta(x;4,1) + \delta\sqrt{x}$ with $\delta > 0$, meaning there are slightly more primes in the residue class 3 (mod 4) than in 1 (mod 4). This discrepancy is called *Chebyshev's bias* and it is explained in detail in [157] and [70].

(a) Let R be the rectangle with vertices $5/2 \pm iT$ and $-3/2 \pm iT$ traversed counterclockwise, and let L be its right half. Prove that

$$\pi \cdot (N(T,\chi)+1) = \Delta_L \arg \xi(s,\chi).$$

(b) Adapting the argument of Exercise 8.8, prove that

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

Exercise 11.7* ([31, Chapters 11–12]). Let χ be a primitive, non-principal Dirichlet character such that[2] $\rho \neq 0$ for all non-trivial zeroes of $L(s,\chi)$.

(a) Prove that the Hadamard product $h(s,\chi) = \prod_\rho (1-s/\rho)e^{s/\rho}$, defined over all non-trivial zeroes ρ of $L(s,\chi)$, converges absolutely and uniformly on compact subsets of $\mathbb{C}$.

(b) Prove that $\xi(s,\chi) = e^{A_\chi + B_\chi s} h(s,\chi)$ for some $A_\chi, B_\chi \in \mathbb{C}$ by adapting the argument of Exercise 8.7.

(c) Show that $e^{A_\chi} = \xi(0,\chi)$, that $B_\chi = (\xi'/\xi)(0,\chi)$ and that $\mathrm{Re}(B_\chi) = -\sum_\rho \mathrm{Re}(1/\rho)$.

Exercise 11.8*. Consider a primitive, non-principal character χ mod q, and ϕ, Φ as in Exercise 7.2(b) with $\phi_0 = 1$. Let $a \in \{0,1\}$ be as in (11.3) and

$$\lambda_a(s) = \begin{cases} 2^s\pi^{s-1}\Gamma(1-s)\sin(\pi s/2) = 2^{s-1}\pi^s/(\Gamma(s)\cos(\pi s/2)) & \text{if } a=0,\\ 2^s\pi^{s-1}\Gamma(1-s)\cos(\pi s/2) = 2^{s-1}\pi^s/(\Gamma(s)\sin(\pi s/2)) & \text{if } a=1. \end{cases}$$

(a) For each $s \in \mathbb{C}$, show that $L(s,\chi) = \varepsilon(\chi) q^{1/2-s}\lambda_a(s) L(1-s,\overline{\chi})$.

(b) From now on, consider $s \in \mathbb{C}$ with $0 \leqslant \sigma \leqslant 1$, as well as $x, y \geqslant 1$ with $xy = q\tau/2\pi$, where $\tau = \max\{|t|,1\}$. Adapt the argument of Exercise 7.8(a,b) to show the *approximate functional equation*

$$L(s,\chi) = \sum_{n\geqslant 1}\frac{\chi(n)\phi(n/x)}{n^s} + \varepsilon(\chi)q^{1/2-s}\sum_{n\geqslant 1}\frac{\overline{\chi}(n)\phi^*_{a,s}(n/y)}{n^{1-s}},$$

where

$$\phi^*_{a,s}(u) := -\frac{1}{2\pi i}\int_{(-3)} \Phi(z)\lambda_a(s+z)(u\tau/2\pi)^z \mathrm{d}z.$$

(c) Show that $L(s,\chi) \ll_\varepsilon (q\tau)^{(1-\sigma)/2+\varepsilon}$ by adapting the argument of Exercise 7.8(c–f). (This reproves Exercise 11.1 inside the critical strip.)

[2] We will show later, in Theorems 12.3 and 12.8, that $L(1,\chi) \neq 0$. Hence, the hypothesis that $\rho \neq 0$ for the non-trivial zeroes of $L(s,\chi)$ follows by the functional equation (Theorem 11.1).

Chapter 12

The Prime Number Theorem for arithmetic progressions

The pinnacle of the theory of Dirichlet L-functions is a quantitative form of the Prime Number Theorem for arithmetic progressions that is known in the literature as the *Siegel-Walfisz theorem*.

Theorem 12.1 (Siegel-Walfisz)**.** *Let $A > 0$. There exists an absolute constant $c > 0$ such that if $1 \leqslant q \leqslant (\log x)^A$ and $a \in (\mathbb{Z}/q\mathbb{Z})^*$, then*

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

An important feature of this result is that it proves that primes are equidistributed in $(\mathbb{Z}/q\mathbb{Z})^*$ when the modulus q tends to infinity with x at a rate that is polynomial in $\log x$, something that is very important in applications. The range of q and x can be significantly enlarged under the assumption of the Generalized Riemann Hypothesis (see Exercise 11.2).

To achieve the required uniformity in q and x, we must keep track of the dependence on q in the various estimates we prove. However, it might be easier to think of q as fixed, say $q = 3$, at the first passage of this chapter.

A zero-free region for Dirichlet L-functions

In view of the explicit formula (11.5), the bulk of the proof of Theorem 12.1 is establishing a zero-free region for Dirichlet L-functions. We start with a simple corollary of Lemma 11.4.

Lemma 12.2. *Let χ be a Dirichlet character* $\bmod q$*, $s = \sigma + it$ with $\sigma \in [1, 2]$, and $\mathcal{Z}$ a sublist of the non-trivial zeroes of $L(s, \chi)$ with $|\gamma - t| \leqslant 1$, possibly containing some zeroes multiple times. Then*

$$\operatorname{Re}\left(\frac{L'}{L}(s,\chi)\right) \geqslant -\operatorname{Re}\left(\frac{1_{\chi=\chi_0}}{s-1}\right) + \sum_{\rho\in\mathcal{Z}} \operatorname{Re}\left(\frac{1}{s-\rho}\right) - O(\log(q|t|+2q)).$$

Proof. Let $\psi \,(\mathrm{mod}\, d)$ be the primitive character inducing χ. Then

$$\frac{L'}{L}(s,\chi) = \frac{L'}{L}(s,\psi) + \sum_{p|q}\sum_{k=1}^{\infty} \frac{\psi(p)^k \log p}{p^{ks}} = \frac{L'}{L}(s,\psi) + O(\log q) \tag{12.1}$$

by (11.2). We then apply Lemma 8.2(b) or 11.4(b) to $L(s, \psi)$, according to whether $\psi = 1$ or $\psi \neq 1$, respectively. The lemma then follows by noticing that $\operatorname{Re}(1/(s-\rho)) \geqslant 0$ when $\sigma \geqslant 1$. □

Next, we prove a generalization of the zero-free region for ζ. Note that our result leaves the possibility for the existence of certain *exceptional zeroes* close to 1. These potential violations to the Generalized Riemann Hypothesis require different arguments that we present in the subsequent section.

Theorem 12.3. *Let $q \geqslant 3$ and $Z_q(s) = \prod_{\chi \,(\mathrm{mod}\, q)} L(s, \chi)$. There is an absolute constant $c_1 > 0$ (i.e., c_1 is independent of q) such that the region of $s = \sigma + it$ with*

$$\sigma \geqslant 1 - \frac{c_1}{\log(q\tau)}, \quad \text{where} \quad \tau = \max\{1, |t|\}, \tag{12.2}$$

contains at most one zero of Z_q. Furthermore, if this exceptional zero exists, then it is necessarily a real simple zero of Z_q, say $\beta_1 \in [1 - c_1/\log q, 1]$, and there is a real, non-principal character $\chi_1 \,(\mathrm{mod}\, q)$ such that $L(\beta_1, \chi_1) = 0$.

Proof. By Theorem 8.3 and relation (11.2), we may restrict our attention to zeroes of Z_q corresponding to non-principal characters χ. As in the proof of Theorem 8.3, the idea is that if $L(s, \chi)$ has a zero close to $1 + it$, then $\chi(p) \sim -p^{it}$ for most primes p. Therefore, $\chi^2(p) \sim p^{2it}$, which yields a pole of $L(s, \chi^2)$ at $s = 1+2it$. This can only happen if $\chi^2 = \chi_0$ and $t = 0$, so that this pole corresponds to that of ζ at $s = 1$. Real zeroes of real characters are then handled using a modification of this argument.

We now give the details of the above sketch. For convenience, we let

$$\mathcal{L}_t = \log\big(q \max\{|t|, 1\}\big).$$

Let $\rho = \beta + i\gamma$ be a zero of $L(s, \chi)$. If χ is real, we further assume that either $\gamma \neq 0$, or that ρ has multiplicity > 1 in $L(s, \chi)$. We want to show that ρ lies outside the region (12.2). Assume for the sake of contradiction that $\beta \geqslant 1 - c_1/\mathcal{L}_\gamma$. We will show this is impossible if c_1 is small enough.

Set $c_1 = 1/M^2$ and $\sigma = 1 + 1/(M\mathcal{L}_\gamma)$, and note that

$$\sigma - \beta \leqslant \sigma - 1 + \frac{1}{M^2\mathcal{L}_\gamma} = \frac{M+1}{M^2\mathcal{L}_\gamma} \leqslant \frac{1}{(M-1)\mathcal{L}_\gamma}. \tag{12.3}$$

Recall the distance function $\mathbb{D}_\sigma(f,g)$ defined in (8.4). The triangle inequality implies that

$$\begin{aligned}\mathbb{D}_\sigma(\chi(n)n^{-i\gamma}, \overline{\chi}(n)n^{i\gamma}) &\leqslant \mathbb{D}_\sigma(\chi(n)n^{-i\gamma}, \mu(n)) + \mathbb{D}_\sigma(\mu(n), \overline{\chi}(n)n^{i\gamma}) \\ &= 2\mathbb{D}_\sigma(\chi(n), \mu(n)n^{i\gamma}).\end{aligned}$$

By a straightforward computation (consult the proof of Theorem 8.3), and using that $\sum_{p|q}(\log p)/p = O(\log q)$, we have

$$\mathbb{D}_\sigma(\chi(n)n^{-i\gamma}, \overline{\chi}(n)n^{i\gamma})^2 = -\frac{\zeta'}{\zeta}(\sigma) + \operatorname{Re}\Big(\frac{L'}{L}(\sigma + 2i\gamma, \chi^2)\Big) + O(\log q)$$

and

$$\mathbb{D}_\sigma(\mu(n), \chi(n)n^{i\gamma})^2 = -\frac{\zeta'}{\zeta}(\sigma) - \operatorname{Re}\Big(\frac{L'}{L}(\sigma + i\gamma, \chi)\Big) + O(\log q).$$

Since $(-\zeta'/\zeta)(\sigma) = 1/(\sigma-1) + O(1) = M\mathcal{L}_\gamma + O(1)$, we infer that

$$4\operatorname{Re}\Big(\frac{L'}{L}(\sigma + i\gamma, \chi)\Big) + \operatorname{Re}\Big(\frac{L'}{L}(\sigma + 2i\gamma, \chi^2)\Big) \leqslant (3M + O(1))\mathcal{L}_\gamma. \tag{12.4}$$

We now bound from below the two summands on the left-hand side of (12.4).

Lemma 12.2 with $\mathcal{Z} = \emptyset$ implies that

$$\begin{aligned}\operatorname{Re}\Big(\frac{L'}{L}(\sigma + 2i\gamma, \chi^2)\Big) &\geqslant -\operatorname{Re}\Big(\frac{1_{\chi^2=\chi_0}}{\sigma - 1 + 2i\gamma}\Big) - O(\mathcal{L}_\gamma) \\ &\geqslant -(\delta(\chi,\rho)M + O(1))\mathcal{L}_\gamma,\end{aligned} \tag{12.5}$$

where $\delta(\chi,\rho) = 1$ if $|\gamma| \leqslant 1/(2M\log q)$ and χ is real, and $\delta(\chi,\rho) = 1/2$ otherwise. Similarly,

$$\operatorname{Re}\Big(\frac{L'}{L}(\sigma + i\gamma, \chi)\Big) \geqslant \frac{1}{\sigma - \beta} - O(\mathcal{L}_\gamma) = (M - O(1))\mathcal{L}_\gamma \tag{12.6}$$

by Lemma 12.2 with $\mathcal{Z} = \{\rho\}$, and by (12.3). Inserting (12.5) and (12.6) into (12.4), and choosing a large M leads to a contradiction when $\delta(\chi,\rho) = 1/2$.

It remains to treat the case when $\delta(\chi,\rho) = 1$, that is to say, when χ is real and $|\gamma| \leqslant 1/(2M\log q)$. There are two sub-cases. First, if ρ is a multiple real zero, then Lemma 12.2 with $\mathcal{Z} = \{\rho,\rho\}$, and relation (12.3) imply that

$$\operatorname{Re}\Big(\frac{L'}{L}(\sigma + i\gamma, \chi)\Big) \geqslant \frac{2}{\sigma - \beta} - O(\mathcal{L}_\gamma) \geqslant (2M - O(1))\mathcal{L}_\gamma$$

Together with (12.4) and (12.5), this leads us to a contradiction when M is large.

Finally, assume that χ is real and that $0 < |\gamma| \leqslant 1/(2M \log q)$. In this case, the obvious symmetry $\overline{L(s,\chi)} = L(\overline{s},\chi)$ implies that $\overline{\rho}$ is also a zero that is different from ρ. Applying Lemma 12.2 with $\mathcal{Z} = \{\rho, \overline{\rho}\}$ then yields

$$\begin{aligned}\operatorname{Re}\Big(\frac{L'}{L}(\sigma+i\gamma,\chi)\Big) &\geqslant \frac{1}{\sigma-\beta} + \frac{\sigma-\beta}{(\sigma-\beta)^2+\gamma^2} - O(\mathcal{L}_\gamma)\\ &\geqslant (1.8M - O(1))\mathcal{L}_\gamma\end{aligned}$$

by (12.3). As before, this leads to a contradiction when M is large enough.

We have thus proven that the only possible zero in the region (12.2) is a real, simple zero, and it can only arise as a zero of some Dirichlet L-function $L(s,\chi)$ of a real, non-principal Dirichlet character $\chi \pmod q$. It remains to prove that at most one such χ exists. Assume for contradiction that there are two different such characters, say χ_1 and χ_2, and let β_1 and β_2, respectively, be their zeroes in the region (12.2). By the triangle inequality,

$$\mathbb{D}_\sigma(\chi_1,\chi_2) \leqslant \mathbb{D}_\sigma(\chi_2,\mu) + \mathbb{D}_\sigma(\mu,\chi_2). \tag{12.7}$$

Since χ_1, χ_2 and $\chi_1\chi_2$ are all real, non-principal characters mod q, Lemma 12.2 with $\mathcal{Z} = \emptyset$ yields

$$\begin{aligned}\mathbb{D}_\sigma(\chi_1,\chi_2)^2 &= -\frac{\zeta'}{\zeta}(\sigma) + \frac{L'}{L}(\sigma,\chi_1\chi_2) + O(\log q)\\ &\geqslant \frac{1}{\sigma-1} - O(\log q) = (M - O(1))\log q,\end{aligned}$$

where $\sigma = 1 + 1/(M\log q)$ and $c_1 = 1/M^2$ as before. On the other hand, arguing as in (12.6), we have

$$\begin{aligned}\mathbb{D}_\sigma(\chi_1,\mu)^2 &= -\frac{\zeta'}{\zeta}(\sigma) - \frac{L'}{L}(\sigma,\chi_1) + O(\log q)\\ &\leqslant \frac{1}{\sigma-1} - \frac{1}{\sigma-\beta_1} + O(\log q) \ll \log q.\end{aligned}$$

The analogous upper bound holds for $\mathbb{D}_\sigma(\chi_1,\mu)^2$ too. Inserting the above estimates into (12.7) and taking M to be large enough leads to a contradiction. This completes the proof of the theorem. □

Theorem 12.3 allows the potential existence of extreme violations to the Generalized Riemann Hypothesis. Before discussing this issue, let us see what we can infer from Theorem 12.3 about prime numbers.

Theorem 12.4. *There is an absolute constant $c_2 > 0$ such that*

$$\psi(x;q,a) = \frac{x - \chi_1(a)x^{\beta_1}}{\varphi(q)} + O\Big(xe^{-c_2\sqrt{\log x}}\Big)$$

uniformly for $x \geqslant q \geqslant 3$ and $(a,q) = 1$, where the term $\chi_1(a)x^{\beta_1}$ is present only if there is an exceptional zero in Theorem 12.3.

Proof. We establish the theorem with $c_2 < 1$. As a consequence, we may assume that $q \leqslant e^{\sqrt{\log x}}$; otherwise, we may simply use the trivial bound $\psi(x;q,a) \ll x(\log x)/q$.

We use the orthogonality of Dirichlet characters (see (10.5)) and the explicit formula (11.5) to write

$$\begin{aligned}\psi(x;q,a) &= \frac{1}{\varphi(q)} \sum_{\chi \,(\mathrm{mod}\, q)} \overline{\chi}(a)\psi(x,\chi) \\ &= \frac{x}{\varphi(q)} - \sum_{\chi \,(\mathrm{mod}\, q)} \overline{\chi}(a) \sum_{|\gamma_\chi| \leqslant T} \frac{x^{\rho_\chi}-1}{\rho_\chi} + O\Big(\frac{x\log^2 x}{T}\Big), \end{aligned} \tag{12.8}$$

where $T \in [3, \sqrt{x}]$ will be chosen later. Let $\rho_\chi = \beta_\chi + i\gamma_\chi$ be a zero of $L(s,\chi)$ different from the exceptional zero β_1, and with imaginary part γ_χ in $[-T,T]$. We claim that there is an absolute constant $c > 0$ such that

$$\frac{x^{\rho_\chi}-1}{\rho_\chi} \ll \frac{x^{1-c/\log(qT)}}{1+|\gamma_\chi|} \qquad \text{when} \quad \rho_\chi \neq \beta_1,\ |\gamma_\chi| \leqslant T. \tag{12.9}$$

Indeed, if $\beta_\chi \leqslant 1/3$, we use the trivial bound $(x^{\rho_\chi}-1)/\rho_\chi \ll x^{1/3}\log x \ll x^{5/6}(\log x)/T$; otherwise, we use Theorem 12.3 to find that $\beta_\chi \leqslant 1 - c_1/\log(qT)$ for some absolute constant $c_1 > 0$. Since we also have that $|\rho_\chi| \asymp 1 + |\gamma_\chi|$ in this case, relation (12.9) readily follows with $c = \min\{c_1, 1/6\}$.

Inserting the bound (12.9) into (12.8), we conclude that

$$\psi(x;q,a) = \frac{x}{\varphi(q)} - \frac{\overline{\chi}_1(a)x^{\beta_1}}{\beta_1} + R$$

with a remainder term R of size

$$\begin{aligned} R &\ll \frac{1}{\varphi(q)} \sum_{\chi \,(\mathrm{mod}\, q)} \sum_{|\gamma_\chi| \leqslant T,\, \rho_\chi \neq \beta_1} \frac{x^{1-c_1'/\log(qT)}}{1+|\gamma_\chi|} + \frac{x\log^2 x}{T} \\ &\ll x^{1-c/\log(qT)}\log^2(qT) + \frac{x\log^2 x}{T}, \end{aligned}$$

where we used Lemma 11.4(a) to bound the sum over the zeroes of $L(s,\chi)$. We choose $T = e^{\sqrt{\log x}}$ and recall that $q \leqslant e^{\sqrt{\log x}}$. Consequently, $R \ll xe^{-c_2\sqrt{\log x}}$ with $c_2 = c/3$. Since $\overline{\chi}_1(a) = \chi_1(a)$ by the fact that χ_1 is a real character (if it exists at all), the proof is complete. □

Since for the moment we know nothing about β_1, it could be the case that $\beta_1 = 1$. In this extreme case, and assuming that q is fixed and $x \to \infty$, Theorem 12.4 implies that

$$\psi(x;a,q) = \begin{cases} (2+o(1))x/\varphi(q) & \text{if } \chi_1(a) = -1, \\ o(x/\varphi(q)) & \text{if } \chi_1(a) = 1. \end{cases} \tag{12.10}$$

We thus see that an exceptional zero of $\prod_{\chi \,(\mathrm{mod}\, q)} L(s,\chi)$ at $s=1$ forces a very uneven distribution of the primes among arithmetic progressions mod q, with half of the reduced classes containing twice as many primes as they should, and the rest containing very few primes.

More generally, it can be proven that if there is an exceptional zero at $\beta_1 = 1 - 1/(M \log q_1)$, then the residue classes $a \,(\mathrm{mod}\, q)$ with $\chi_1(a) = -1$ contain $(2 + o_{x,u_1,u_2\to\infty}(1))x/(\varphi(q)\log x)$ primes of size $\leqslant x$, provided that x is in the range $[q^{u_1}, q^{M/u_2}]$. This result is due to Linnik. We will prove a weak form of it in Chapter 27.

Exceptional characters

The characters χ whose L-function has a zero β in the region (12.2) are called *exceptional*, and the zero β is called an *exceptional zero* or a *Landau-Siegel zero*. This definition should not be taken too literally because it depends on the choice of the unspecified constant c_1 in (12.2). Strictly speaking, the rigorous definition of Landau-Siegel zeroes concerns a sequence of characters $\chi_j \,(\mathrm{mod}\, q_j)$ such that no product $\chi_j\chi_k$ with $j \neq k$ is principal, and for which there are real numbers

$$\beta_j = 1 - o_{j\to\infty}(1/\log q_j) \quad \text{such that} \quad L(\beta_j, \chi_j) = 0.$$

Disproving the existence of Landau-Siegel zeroes is a major open problem in analytic number theory. We establish some partial results about them.

We begin by showing the following result due to Landau, which proves that exceptional zeroes "repel" each other.

Theorem 12.5 (Landau)**.** *Let $\chi_1 \,(\mathrm{mod}\, q_1)$ and $\chi_2 \,(\mathrm{mod}\, q_2)$ be two real, non-principal characters that are not induced by the same primitive character, and both of whose L-functions have real zeroes β_1 and β_2, respectively. There is an absolute constant $c > 0$ such that*

$$\min\{\beta_1, \beta_2\} \leqslant 1 - \frac{c}{\log(q_1 q_2)}.$$

Proof. The theorem follows by a simple modification of the last part of the proof of Theorem 12.3, starting with (12.7). We leave the details as an exercise. □

We record two important corollaries of Theorem 12.5, both of which demonstrate the scarcity of Landau-Siegel zeroes.

Corollary 12.6. *Let $\chi_j \,(\mathrm{mod}\, q_j)$ be a sequence of primitive characters of strictly increasing moduli such that $L(\beta_j, \chi_j) = 0$ for some $\beta_j > 1 - c/\log q_j$. If c is small enough, then $q_{j+1} > q_j^{100}$ for all j.*

Corollary 12.7 (Page)**.** *There is an absolute constant $c > 0$ such that among all real, primitive characters of conductor $\leqslant Q$, there is at most one whose Dirichlet L-function has a real zero $> 1 - c/\log Q$.*

Corollary 12.7, known in the literature as *Page's theorem*, does not exclude the possibility that there is one character $\chi \pmod q$ with a zero at 1 that would yield the very uneven distribution of primes described in (12.10). To show there are no such zeroes, we use a different argument.

Let $\chi \pmod q$ be a real, non-principal character such that $L(\beta, \chi) = 0$ for some $\beta \leqslant 1$. We have

$$L(1,\chi) = \int_\beta^1 L'(\sigma,\chi)\mathrm{d}\sigma \ll (1-\beta)q^{(1-\beta)/2}\log^2 q \tag{12.11}$$

by Lemma 11.2. When $\beta \geqslant 1 - 1/\log q$, this implies that

$$1 - \beta \gg L(1,\chi)/\log^2 q. \tag{12.12}$$

Hence we could show that β is not too close to 1 by proving a lower bound for $L(1,\chi)$. A weak but uniform such bound follows.

Theorem 12.8. *If χ is a non-principal real Dirichlet character $\bmod q$, then*

$$L(1,\chi) \gg \frac{1}{\sqrt{q}\log^2 q}.$$

In particular, there is an absolute constant $c > 0$ such that

$$L(\sigma,\chi) \neq 0 \quad \textit{when} \quad \sigma > 1 - \frac{c}{q^{1/2}\log^4 q}.$$

Proof. The second part of the theorem follows readily from the first part and (12.12). Now, to bound $L(1,\chi)$ from below we consider the function $1 * \chi$. We note that

$$(1*\chi)(p^m) = \begin{cases} m+1 & \text{if } \chi(p) = 1, \\ 1_{2|m} & \text{if } \chi(p) = -1, \\ 1 & \text{if } \chi(p) = 0. \end{cases}$$

Using multiplicativity, we infer that[1] $(1*\chi)(n) \geqslant 0$ and $(1*\chi)(n^2) \geqslant 1$ for all n, whence $\sum_{n\leqslant x}(1*\chi)(n) \gg \sqrt{x}$. On the other hand, we have

$$S := \sum_{n\leqslant x}(1*\chi)(n) = \sum_{a\leqslant x}\chi(a)\left\lfloor x/a\right\rfloor.$$

We thus see that the expected main term is $x\sum_{a\leqslant x}\chi(a)/a \sim xL(1,\chi)$ for large enough x, which should allow us to get a lower bound for $L(1,\chi)$.

[1] When χ is a primitive character, $(1*\chi)(n)$ counts ideals of norm n in the ring of integers of the quadratic field $\mathbb{Q}(\sqrt{\theta q})$, where $\theta \in \{-1,+1\}$ is an appropriate sign (see [137, Chapter 7]). The inequality $(1*\chi)(n^2) \geqslant 1$ has a more conceptual proof in this context, since it is a consequence of the fact that there is always at least one ideal of norm n^2 (the principal ideal (n)).

However, it is hard to estimate S with a remainder term smaller than $\sqrt{x}$, which is the size of our lower bound for S. To bypass this technical issue, we work with the *smoothened* sum

$$T := \sum_{n\leqslant x}(1*\chi)(n)(1-n/x) = \sum_{a\leqslant x}\chi(a)\sum_{b\leqslant x/a}\Big(1-\frac{b}{x/a}\Big).$$

The Euler-Maclaurin summation formula (Theorem 1.10) implies that

$$\sum_{n\leqslant x}(1-n/x) = \frac{x}{2} - \frac{1}{x}\int_0^x \{t\}\mathrm{d}t.$$

Consequently,

$$\begin{aligned} T &= \sum_{a\leqslant x}\chi(a)\bigg(\frac{x}{2a} - \frac{1}{x/a}\int_0^{x/a}\{t\}\mathrm{d}t\bigg) \\ &= \frac{x}{2}\sum_{a\leqslant x}\frac{\chi(a)}{a} - \frac{1}{x}\int_0^x \{t\}\sum_{a\leqslant \min\{x,x/t\}} a\chi(a)\mathrm{d}t. \end{aligned}$$

Using the Pólya-Vinogradov inequality and partial summation, we find that

$$\sum_{a>x}\frac{\chi(a)}{a} \ll \frac{\sqrt{q}\log q}{x} \qquad \text{and} \qquad \sum_{a\leqslant y} a\chi(a) \ll y\sqrt{q}\log q$$

uniformly for $x, y \geqslant 1$. Consequently,

$$\sum_{n\leqslant x}(1*\chi)(n)(1-n/x) = \frac{xL(1,\chi)}{2} + O\big(\sqrt{q}(\log q)(\log x)\big).$$

On the other hand,

$$\sum_{n\leqslant x}(1*\chi)(n)(1-n/x) \geqslant \sum_{m^2\leqslant x}(1-m^2/x) \gg \sqrt{x}.$$

Comparing the above estimates when $x = cq(\log q)^4$ for a large enough constant c completes the proof of the theorem. □

Theorems 12.4 and 12.8 establish Theorem 12.1 uniformly for all moduli $q \leqslant (\log x)(\log\log x)^{-8}$. In order to handle larger q, we need a strengthening of Theorem 12.8 due to Siegel.

Theorem 12.9 (Siegel)**.** *Let $\varepsilon > 0$. There is a constant $c(\varepsilon) > 0$ such that for all real, primitive, non-principal Dirichlet characters $\chi \pmod q$, we have*

$$L(\sigma,\chi) \neq 0 \quad \textit{when} \quad \sigma > 1 - c(\varepsilon)q^{-\varepsilon},$$

with the possible exception of one character $\chi_1 \pmod{q_1}$.

Proof. We follow an argument due to Goldfeld [60]. Clearly, by taking $c(\varepsilon) \leqslant \varepsilon/2$, we may assume that there is at least one primitive, non-principal character whose L-function has a zero in $[1-\varepsilon/2, 1]$. Let χ_1 be such a character of minimal conductor $q_1 > 1$.

Now let $\chi \pmod q$ be a different real, primitive, non-principal Dirichlet character. If $q < q_1$, then the claimed zero-free region follows by the minimality of q, so assume that $q \geqslant q_1$.

We argue similarly to Theorem 12.8, only this time we replace $1 * \chi$ by $f = 1 * \chi * \chi_1 * \chi\chi_1$. This function is also non-negative; the easiest way to see this is by examining the logarithm of its Dirichlet series[2]

$$F(s) = \zeta(s)L(s,\chi_1)L(s,\chi)L(s,\chi\chi_1).$$

Let β_1 be the rightmost zero of $L(s,\chi_1)$ in the interval $[1-\varepsilon/2, 1]$, let $\phi \in C^\infty(\mathbb{R}_{\geqslant 0})$ such that $1_{[0,1]} \leqslant \phi \leqslant 1_{[0,2]}$, and consider the auxiliary sum

$$S = \sum_{n=1}^{\infty} \frac{f(n)\phi(n/x)}{n^{\beta_1}}.$$

On the one hand, we have the trivial lower bound $S \geqslant 1$ by dropping all summands except for the one with $n = 1$. On the other hand, we can evaluate S using Mellin inversion: Exercise 7.2(d) implies that

$$S = I_2, \quad \text{where} \quad I_\alpha := \frac{1}{2\pi i}\int_{(\alpha)} F(s+\beta_1)x^s\Phi(s)\mathrm{d}s$$

with Φ denoting the Mellin transform of ϕ. Since $F(\beta_1) = 0$, the only pole of the integrand is at $s = 1-\beta_1$, which is a positive number by Theorem 12.8. Shifting the contour to the line $\mathrm{Re}(s) = -1$, we find that

$$S = x^{1-\beta_1}L(1,\chi_1)L(1,\chi)L(1,\chi\chi_1)\Phi(1-\beta_1) + I_{-1}.$$

When $s = -1+it$, we have $|F(s+\beta_1)| \ll \max\{q,|t|\}^{c_1}$ for some absolute constant $c_1 > 0$ by Exercise 11.1 (the characters χ, χ_1 and $\chi\chi_1$ all have conductor $\leqslant q_1 q \leqslant q^2$). Since $|\Phi(-1+it)| \ll 1/(1+|t|)^{c_1+2}$ by Exercise 7.2(c), we find that $I_{-1} = O(q^{c_1}/x)$. Taking $x = c_2 q^{c_1}$ for a large enough constant c_2 makes $|I_{-1}| \leqslant 1/2$. Recalling that $S \geqslant 1$, we infer that

$$c_2 q^{c_1(1-\beta_1)}L(1,\chi_1)L(1,\chi)L(1,\chi\chi_1)\Phi(1-\beta_1) \geqslant 1/2.$$

Next, we note that $L(1,\chi\chi_1) \ll \log q$ by Lemma 11.2, $\Phi(1-\beta_1) \leqslant \int_0^2 y^{-\beta_1}\mathrm{d}y \leqslant 2/(1-\beta_1)$, and $L(1,\chi_1) \ll (1-\beta_1)q^{(1-\beta_1)/2}\log^2 q$ by (12.11). Since we also have $1-\beta_1 \leqslant \varepsilon/2$, we conclude that

$$L(1,\chi) \gg q^{-(c_1+1/2)(1-\beta_1)}/\log^3 q \gg_\varepsilon q^{-(c_1+1/2)\varepsilon}.$$

[2]This is the Dedekind function of the biquadratic field $\mathbb{Q}(\sqrt{\theta q}, \sqrt{\theta_1 q_1})$, where θ, θ_1 are appropriate signs (see footnote 1 on page 124).

Together with (12.11), this proves that $L(\sigma,\chi)\neq 0$ for $\sigma\geqslant 1-O_\varepsilon(q^{-(c_1+1)\varepsilon})$. Replacing ε by $\varepsilon/(c_1+1)$ completes the proof. □

The possible exceptional character χ_1 in Theorem 12.9 causes a subtle but significant problem: since we know nothing about it, we can at best use Theorem 12.8 to say that $L(\sigma,\chi_1)$ has no zeroes when $\sigma > 1 - O(1/(\sqrt{q_1}\log^4 q_1))$ for some $c_1 > 0$. Of course, q_1 here is a constant (the conductor of the hypothetical unique exceptional character χ_1), so there is some constant $\tilde{c}=\tilde{c}(\varepsilon,q_1)$ such that $L(\sigma,\chi_1)$ has no zeroes for $\sigma > 1-\tilde{c}q_1^{-\varepsilon}$. However, since we have no control over q_1, it is impossible to compute a specific value of $\tilde{c}$. Notice that this is not due to a lack of computing power, but because of the argument producing $\tilde{c}$. We then say that "$\tilde{c}$ cannot be computed effectively". We have thus arrived at the ineffective form of Theorem 12.9, known as *Siegel's theorem*.

Theorem 12.10 (Siegel)**.** *Let $\varepsilon > 0$. There is a constant $c(\varepsilon) > 0$ (that cannot be computed effectively) such that*

$$L(\sigma,\chi)\neq 0 \quad \textit{when} \quad \sigma > 1 - c(\varepsilon)q^{-\varepsilon}$$

for all real, non-principal Dirichlet characters $\chi \pmod q$.

Proof. We have already treated the primitive characters. The non-primitive characters are dealt with via formula (11.2). □

Combining Siegel's theorem with Theorem 12.4 completes the proof of Theorem 12.1. Note, however, that the ineffectivity of Theorem 12.10 transfers to the implicit constant in Theorem 12.1. As a consequence, results proven using Theorem 12.1 are generally not amenable to numerical analysis. There are some exceptions to this rule, as it is sometimes possible to isolate the influence of the exceptional character χ_1 in Theorem 12.9.

We will revisit exceptional characters in Chapters 22 and 27.

Exercises

Exercise 12.1. Adapt the argument of Exercise 8.4 to prove that there is a constant $c > 0$ such that for each fixed $A > 0$ we have

$$\sum_{n\leqslant x}\mu(n)\chi(n) \ll_A xe^{-c\sqrt{\log x}} \quad (1\leqslant q\leqslant (\log x)^A,\ \chi \pmod q).$$

Exercise 12.2. Let χ be a Dirichlet character mod q and $x\geqslant q\geqslant 3$.

(a) Prove that there is an absolute constant $c > 0$ such that

$$\psi(x,\chi) = 1_{\chi=\chi_0}x - \frac{x^{\beta_1}}{\beta_1} + O\Big(xe^{-c\sqrt{\log x}} + (\log q)^2 x^{1-c/\log q}\Big),$$

where the term x^{β_1}/β_1 is present only when χ is the exceptional character from Theorem 12.3.

(b) Let ϕ and Φ be as in Exercise 7.2. Prove that

$$\psi(x,\chi) = 1_{\chi=\chi_0}\Phi(1)x - x^{\beta_1}\Phi(\beta_1) + O_\phi\Big(xe^{-c\sqrt{\log x}} + (\log q)x^{1-c/\log q}\Big),$$

where the term $x^{\beta_1}\Phi(\beta_1)$ is present only when χ is the exceptional character from Theorem 12.3.

Exercise 12.3* ([114, Lemma 18.4]). Let χ be a real, non-principal character mod q, and let $\beta \in [1/2, 1]$ be a zero of $L(s,\chi)$.

(a) For $x \geqslant y \geqslant 1$, prove that

$$\Big(\sum_{y<p\leqslant x}\frac{1+\chi(p)}{p}\Big)\Big(\sum_{n\leqslant y}\frac{(1*\chi)(n)}{n}\Big) \leqslant \sum_{y<m\leqslant xy}\frac{(1*\chi)(m)}{m}.$$

(b) For $N \geqslant q$, prove that

$$\sum_{n\leqslant N}\frac{(1*\chi)(n)}{n} = L(1,\chi)(\log N + \gamma) + L'(1,\chi) + O(q^{1/4}N^{-1/2}\log N)$$

using the hyperbola method.

(c) If $\phi \in C^\infty(\mathbb{R})$ is such that $1_{[0,1/2]} \leqslant \phi \leqslant 1_{[0,1]}$, then show that

$$\sum_{n\leqslant y}\frac{(1*\chi)(n)}{n} \geqslant y^{\beta-1}\sum_{n\geqslant 1}\frac{(1*\chi)(n)\phi(n/y)}{n^\beta} \asymp \frac{L(1,\chi)}{1-\beta},$$

as long as $y \geqslant q^2$ and q is large enough depending only on ϕ.

(d) Deduce that

$$\sum_{y<p\leqslant x}\frac{1+\chi(p)}{p} \ll (1-\beta)\log x \qquad (x \geqslant y \geqslant q^3).$$

Exercise 12.4 (Alternative proof of a weak version of Theorem 12.3). Let $\chi \pmod q$ be a Dirichlet character, $t \in \mathbb{R}$ and $\tau = \max\{|t|, 2\}$. Assume that either χ is complex or $|t| \geqslant 1$.

(a) Using the 3-4-1 inequality (8.6), prove that

$$L(\sigma,\chi_0)^3|L(\sigma+it,\chi)|^4|L(\sigma+2it,\chi^2)| \geqslant 1 \quad \text{for } \sigma > 1.$$

[*Hint:* Recall that $\log|z| = \mathrm{Re}(\log z)$.]

(b) For $\sigma \in (1,2]$, show that $|L(\sigma+it,\chi)| \gg (\sigma-1)^{3/4}/\log^{1/2}(q\tau)$.

(c) For $\sigma, \sigma' \geqslant 1 - 1/\log(q\tau)$, show that

$$L(\sigma'+it,\chi) = L(\sigma+it,\chi) + O(|\sigma'-\sigma|\log^2(q\tau)).$$

Conclude that there is an constant $c > 0$ such

$$|L(\sigma+it,\chi)| \gg 1/\log^8(q\tau) \quad \text{for} \quad \sigma > 1 - c/\log^{10}(q\tau).$$

Part 3

Multiplicative functions and the anatomy of integers

Chapter 13

Primes and multiplicative functions

There is a strong connection between the distribution of primes and the average behavior of multiplicative functions. Landau proved that the Prime Number Theorem is *elementarily equivalent* to the relation $\sum_{n\leqslant x}\mu(n) = o_{x\to\infty}(x)$ (see Exercise 3.15). In a similar vein, the Riemann Hypothesis is equivalent to the bound $|\sum_{n\leqslant x}\mu(n)| \leqslant x^{1/2+o(1)}$ as $x\to\infty$ (see Exercise 8.6), whereas the Generalized Riemann Hypothesis amounts to showing the same estimate for the partial sums of $\mu\chi$ for each character $\chi \pmod q$.

In order to understand better the interplay between primes and multiplicative functions, we assume a more general point of view. Much of what we have done so far can be roughly described as follows: we are given an interesting arithmetic sequence indexed by primes, say $f(2)$, $f(3)$, $f(5)$, ..., and we want to understand its partial sums. To accomplish this, we consider a special generating function: the Dirichlet series $\sum_p f(p)/p^s$. In certain fortuitous situations, this series is related to the logarithm of the Dirichlet series $F(s) = \sum_{n=1}^\infty f(n)/n^s$ of a "nice" multiplicative function f. Analyzing averages of $f(n)$ thus gives us information on averages of $f(p)$.

We now explore the converse direction: assuming we have good bounds on $\sum_{p\leqslant x} f(p)$, we seek estimates for $\sum_{n\leqslant x} f(n)$. We will accomplish this goal when the sequence $f(p)$ is roughly constant on average. More precisely, we assume that there is some $\kappa\in\mathbb{C}$ and some parameter $Q\geqslant 2$ such that

$$\sum_{p\leqslant x} f(p)\log p = \kappa x + O_A\big(x/(\log x)^A\big) \qquad (x\geqslant Q) \tag{13.1}$$

for each fixed $A>0$. We think of κ as being fixed and Q as varying.

We must further impose a growth condition on f that prevents abnormally large values from ruling its partial sums. A simple way of achieving this is to assume that there is a fixed $k \in \mathbb{N}$ such that

$$|f| \leqslant \tau_k. \tag{13.2}$$

We call such a function *divisor-bounded*.

For instance, if $f = \mu$, then (13.1) and (13.2) hold with $\kappa = -1$, $k = 1$ and $Q = 2$, whereas when $f = \mu\chi$ for a non-principal character $\chi \,(\mathrm{mod}\, q)$, then $\kappa = 0$, $k = 1$ and $Q = \exp\{q^\varepsilon\}$ by the Siegel-Walfisz theorem (Theorem 12.1). A more unusual but still natural example is given by the function $\tilde{\mu}(n) = \mu^2(n)(-1)^{\omega(n;5,1)}$, where $\omega(n;5,1) = \#\{\, p|n : p \equiv 1 \,(\mathrm{mod}\, 5)\,\}$. Indeed, $\tilde{\mu}$ satisfies (13.1) with $\kappa = 1/2$ and $Q = 2$, whereas (13.2) holds with $k = 1$. We already have good predictions for the partial sums of μ and $\mu\chi$, but what about the average behavior of $\tilde{\mu}$?

Generalized divisor functions

The study of the above general class of functions f can be reduced to certain canonical representatives. These are generalizations of the combinatorially defined divisor functions τ_m, $m \in \mathbb{N}$. Recall that the Dirichlet series of τ_m is $\zeta(s)^m$. We then define τ_κ for $\kappa \in \mathbb{C}$ to be the arithmetic function whose Dirichlet series is $\zeta(s)^\kappa$. Using the Euler product representation of ζ and the Taylor series expansion of $(1-x)^{-\kappa}$ about $x = 0$, we find that

$$\tau_\kappa(p^a) = \binom{\kappa + a - 1}{a}. \tag{13.3}$$

In particular, $\tau_\kappa(p) = \kappa$, so that (13.1) holds by the Prime Number Theorem.

To relate a general function f satisfying (13.1) to the function τ_κ, we go back to the basics. Note that the average value of $f(p) - \tau_\kappa(p)$ is zero. Hence, if we write $f = \tau_\kappa * g$, then $g(p)$ is zero on average. We might thus guess that g has small partial sums. Dirichlet's hyperbola method would then suggest that f and τ_κ behave very similarly on average.

A more analytic way to view the above argument is to consider $F(s)$, the Dirichlet series of f. We then roughly have $F(s)\zeta(s)^{-\kappa} \approx \exp\{\sum_p (f(p) - \kappa)/p^s\}$. For this reason, (13.1) is essentially equivalent to the function $F\zeta^{-\kappa}$ having a C^∞-extension to the half-plane $\mathrm{Re}(s) \geqslant 1$ (see Lemma 13.5(a) below). For simplicity, let us assume momentarily a stronger version of (13.3), with an error term of size $O(x^{1-\varepsilon})$. Then, we can analytically continue $F\zeta^{-\kappa}$ to the half-plane $\mathrm{Re}(s) > 1 - \varepsilon$. Hence, we see from (5.14) that the partial sums of $g = f * \tau_{-\kappa}$ should indeed be rather small (there are no residue contributions to the right side of (5.14)). This allows us to relate averages of f and τ_κ via the hyperbola method.

Let us now study the partial sums of the prototypical function τ_κ. We begin, as usual, by invoking Perron's inversion formula: for any $x \notin \mathbb{Z}$ and any $\alpha \in (1, 1 + 1/\log x]$, we have

$$\sum_{n \leqslant x} \tau_\kappa(n) = \frac{1}{2\pi i} \int_{(\alpha)} \zeta(s)^\kappa \frac{x^s}{s} \mathrm{d}s. \tag{13.4}$$

When κ is an integer, ζ^κ has a meromorphic continuation to $\mathbb{C}$, so the usual contour shifting argument can be used to estimate $\sum_{n \leqslant x} \tau_\kappa(n)$. However, when $\kappa \notin \mathbb{Z}$, the function ζ^κ is only defined where $\log \zeta(s)$ is. In particular, since ζ has a pole at $s = 1$, we can only define $\zeta(s)^\kappa$ in a *simply connected* set of the complex plane that does not contain 1, nor any of the zeroes of ζ. There is no such domain containing a punctured disk centered at 1. Hence, it is not possible to employ Cauchy's residue theorem to study the contribution of the singularity at $s = 1$ to the partial sums of τ_κ, which means that we must develop a new method to deal with the integral in (13.4).

The LSD method

The main idea for estimating the integral in (13.4) goes back to work of Landau and was further developed by Selberg and Delange. Here, we present an adaptation of their technique that appeared in [68], which builds upon ideas in [114, Section 2.4]. The original method of Landau-Selberg-Delange (called the *LSD method* for brevity) is presented in great detail in Chapter II.5 of Tenenbaum's book [172]. We also outline it in Exercise 13.6.

For simplicity, assume that $\kappa > 1$. Note that the integrand $\zeta(s)^\kappa x^s/s$ blows up to ∞ when $s \to 1$. On the other hand, Exercise 8.4 shows that $\zeta(s)^\kappa x^s/s \ll x|t|^{-1/2} = o_{|t| \to \infty}(x)$ when $s = \sigma + it$ with $1 < \sigma \leqslant 1 + 1/\log x$. Thus, if we take α sufficiently close to 1, it seems reasonable to expect that most of the contribution to the integral in (13.4) comes from s close to 1. For such s, we have that $\zeta(s)^\kappa/s \sim 1/(s-1)^\kappa$. This leads us to guess that

$$\sum_{n \leqslant x} \tau_\kappa(n) \approx \frac{1}{2\pi i} \int_{(\alpha)} \frac{x^s}{(s-1)^\kappa} \mathrm{d}s. \tag{13.5}$$

The right-hand side of the above formula can be computed using Lemma 13.1 below, which is called *Hankel's formula*.

Lemma 13.1 (Hankel's formula)**.** *Let $x > 1$, $\alpha > 0$ and $\mathrm{Re}(\kappa) > 1$. Then*

$$\frac{1}{2\pi i} \int_{(\alpha)} \frac{x^s}{s^\kappa} \mathrm{d}s = \frac{(\log x)^{\kappa-1}}{\Gamma(\kappa)}.$$

If, in addition, $\alpha > 1$, then we have

$$\frac{1}{2\pi i} \int_{(\alpha)} \frac{x^s}{s(s-1)^\kappa} \mathrm{d}s = \frac{1}{\Gamma(\kappa)} \int_1^x (\log y)^{\kappa-1} \mathrm{d}y.$$

Proof. We have $\int_1^\infty (\log x)^{\kappa-1} x^{-s-1} \mathrm{d}x = s^{-\kappa}\Gamma(\kappa)$ for $\mathrm{Re}(s) > 0$. (Justify why this is true *for all* s with $\mathrm{Re}(s) > 0$.) Since $\mathrm{Re}(\kappa) > 1$, the function $1/s^\kappa$ is absolutely integrable on every vertical line $\mathrm{Re}(s) = \alpha$ with $\alpha \neq 0$. Thus, the Mellin inversion formula (Theorem B.4) implies that $(1/2\pi i)\int_{(\alpha)} y^s s^{-\kappa} \mathrm{d}s = 1_{y>1}(\log y)^{\kappa-1}/\Gamma(\kappa)$ for any $\alpha > 0$, which proves the first part of the lemma. Integrating over $y \in [0, x]$ proves the second part too with $\alpha + 1$ in place of α. □

The above discussion leads us to conjecture that

$$\sum_{n \leqslant x} \tau_\kappa(n) \approx \frac{x(\log x)^{\kappa-1}}{\Gamma(\kappa)} \quad (x \to \infty). \tag{13.6}$$

Note that this agrees with Theorem 7.4 when $\kappa \in \mathbb{N}$. Remarkably, it also agrees with what we know for the Möbius function. Indeed, when $\kappa = -1$, we have $\tau_{-1} = \mu$, for which we know that $\sum_{n \leqslant x} \mu(n) = o_{x\to\infty}(x)$. On the other hand, the right-hand side of (13.6) vanishes because of the pole of the Gamma function at $s = -1$.

The main goal of this chapter is to establish an appropriate version of (13.6) for *all* multiplicative functions f satisfying (13.1) and (13.2). Under the same assumptions, we will show that the asymptotic behavior of $\sum_{n \leqslant x} f(n)$ is determined by the analytic behavior of the Dirichlet series $F(s)$ when $s \approx 1$.

As we mentioned earlier, $F(s)\zeta(s)^\kappa$ admits a C^∞-extension to the half-plane $\mathrm{Re}(s) \geqslant 1$ under the assumptions of (13.1) and (13.2). Thus, the same must be true for the function $F(s)(s-1)^\kappa$ because $\zeta(s)(s-1)$ is analytic and non-zero in an open neighborhood of the plane $\mathrm{Re}(s) \geqslant 1$. We then let

$$c_j = \frac{\mathrm{d}^j}{j!\mathrm{d}s^j}\Big|_{s=1} (s-1)^\kappa F(s) \quad \text{and} \quad \widetilde{c}_j = \frac{\mathrm{d}^j}{j!\mathrm{d}s^j}\Big|_{s=1} \frac{(s-1)^\kappa F(s)}{s} \tag{13.7}$$

be the Taylor coefficients about 1 of the functions $(s-1)^\kappa F(s)$ and $(s-1)^\kappa F(s)/s$, respectively. Since $s = 1 + (s-1)$ and $1/s = 1 - (s-1) + (s-1)^2 \mp \cdots$ for $|s-1| < 1$, these coefficients are linked by the relations

$$\widetilde{c}_j = \sum_{a=0}^{j} (-1)^a c_{j-a} \quad \text{and} \quad c_j = \widetilde{c}_j + \widetilde{c}_{j-1} \qquad \text{for} \quad j = 0, 1, \dots \tag{13.8}$$

with the convention that $\tilde{c}_{-1} = 0$. Moreover, since $\zeta(s) \sim 1/(s-1)$ as $s \to 1^+$ and f is multiplicative, we have that

$$c_0 = \widetilde{c}_0 = \prod_p \Big(1 + \frac{f(p)}{p} + \frac{f(p^2)}{p^2} + \cdots\Big)\Big(1 - \frac{1}{p}\Big)^\kappa. \tag{13.9}$$

We will prove in Lemma 13.5(a) that $c_j, \tilde{c}_j \ll_{j,k} (\log Q)^{j+2k}$.

With the above notation, our main theorem is the following.

Theorem 13.2. *Fix $\varepsilon > 0$ and $J \in \mathbb{N}$. If f, c_j and $\tilde{c}_j$ are as above, then*

$$\sum_{n\leqslant x} f(n) = \int_2^x \sum_{j=0}^{J-1} c_j \frac{(\log y)^{\kappa-j-1}}{\Gamma(\kappa-j)} \mathrm{d}y + O\Big(\frac{x(\log Q)^{2k+J-1}}{(\log x)^{J+1-\mathrm{Re}(\kappa)}}\Big) \tag{13.10}$$

$$= x \sum_{j=0}^{J-1} \widetilde{c}_j \frac{(\log x)^{\kappa-j-1}}{\Gamma(\kappa-j)} + O\Big(\frac{x(\log Q)^{2k+J-1}}{(\log x)^{J+1-\mathrm{Re}(\kappa)}}\Big) \tag{13.11}$$

for $x \geqslant e^{(\log Q)^{1+\varepsilon}}$. The implied constants depend at most on k, J, ε and the implied constant in (13.1) *for A large enough in terms of k, J and ε.*

Remark 13.3. Note that when $\kappa \in \mathbb{Z}_{\leqslant 0}$, then *all* the main terms in (13.10) vanish because of the poles of the Gamma function at $0, -1, -2, \dots$. Hence, for each fixed $\varepsilon > 0$ and $A > 0$, we infer that

$$\sum_{n\leqslant x} f(n) \ll_{A,\varepsilon} x/(\log x)^A \qquad \big(x \geqslant e^{(\log Q)^{1+\varepsilon}}\big). \tag{13.12}$$

On the other hand, when $c_0 \neq 0$ and $\kappa \notin \mathbb{Z}_{\leqslant 0}$, we see that $\sum_{n\leqslant x} f(n)$ is much larger, of size $x(\log x)^{\mathrm{Re}(\kappa)-1}$.

For example, for the function $\tilde{\mu}(n) = \mu^2(n)(-1)^{\omega(n;5,1)}$ that we saw above, we have $\kappa = 1/2$ and $c_0 > 0$. Hence, $\sum_{n\leqslant x} \tilde{\mu}(n)$ is of size $x/(\log x)^{1/2}$. This might be a bit surprising because it is in stark contradiction with a common heuristic argument for the Riemann Hypothesis.

Indeed, given a square-free integer n, note that $\tilde{\mu}(n) = 1$ when $\omega(n;5,1)$ is even, while $\tilde{\mu}(n) = -1$ when $\omega(n;5,1)$ is odd. A similar situation is true for the Möbius function, with $\omega(n;5,1)$ replaced by $\omega(n)$. Since there is no reason to suspect any bias for the parity of the functions $\omega(n;5,1)$ and $\omega(n)$, we may be tempted to model μ and $\tilde{\mu}$ by a sequence of random, independent and equiprobable assignments of $+1$ or -1 to each square-free integer. The Central Limit Theorem would then predict that $|\sum_{n\leqslant x} \mu(n)| \leqslant x^{1/2+o(1)}$ and $|\sum_{n\leqslant x} \tilde{\mu}(n)| \leqslant x^{1/2+o(1)}$. While the former estimate is believed to be true in virtue of the Riemann Hypothesis, the second one is very far from the truth.

In conclusion, we should be very careful when using probabilistic arguments of the above sort to analyze partial sums of multiplicative functions, because their values are interdependent in a fundamental way. For instance, if n is odd, then we *always* have that $f(2n) = f(2)f(n)$, which means that the values $f(2n)$ and $f(n)$ are highly correlated. □

Before going on to prove Theorem 13.2, we record an important consequence of it.

Corollary 13.4. *Fix $A, C \geqslant 1$ and $\varepsilon \in (0, 1/2]$. Let $x \geqslant 2$ and $q, m \in \mathbb{N}$ with $q \leqslant (\log x)^C$ and $\omega(m) \leqslant \exp\{(\log x)^{1-\varepsilon}\}$. Then*

$$\sum_{\substack{n \leqslant x,\ (n,m)=1 \\ n \equiv a \,(\mathrm{mod}\, q)}} \mu(n) \ll_{\varepsilon, A, C} \frac{x}{(\log x)^A}.$$

Proof. Let $d = (a, q)$ and write $a = da_1$ and $q = dq_1$. If $n \equiv a \,(\mathrm{mod}\, q)$, then $d|n$. Hence, for the sum in the statement of the corollary to have any terms, we must have $(d, m) = 1$. In this case, we write $n = dr$, so that

$$\sum_{\substack{n \leqslant x,\ (n,m)=1 \\ n \equiv a \,(\mathrm{mod}\, q)}} \mu(n) = \sum_{\substack{r \leqslant x/d,\ (r,m)=1 \\ r \equiv a_1 \,(\mathrm{mod}\, q_1)}} \mu(dr) = \mu(d) \sum_{\substack{r \leqslant x/d,\ (r,dm)=1 \\ r \equiv a_1 \,(\mathrm{mod}\, q_1)}} \mu(r).$$

Since $(a_1, q_1) = 1$, we may expand the condition $r \equiv a_1 \,(\mathrm{mod}\, q_1)$ using Dirichlet characters. We thus find that

$$\sum_{\substack{n \leqslant x,\ (n,m)=1 \\ n \equiv a \,(\mathrm{mod}\, q)}} \mu(n) = \frac{\mu(d)}{\varphi(q_1)} \sum_{\chi \,(\mathrm{mod}\, q_1)} \overline{\chi}(a_1) \sum_{\substack{r \leqslant x/d \\ (r,dm)=1}} \mu(r)\chi(r). \tag{13.13}$$

Fix $\chi \,(\mathrm{mod}\, q_1)$ and note that $x/d \geqslant x/q \geqslant x/(\log x)^C$. We shall apply Theorem 13.2 with $f(n) = 1_{(n,dm)=1}\mu(n)\chi(n)$. For this function, Theorem 12.1 implies that

$$\begin{aligned} \sum_{p \leqslant w} f(p) \log p &= -\sum_{p \leqslant w} \chi(p) \log p + O(\omega(dm) \log w) \\ &= -1_{\chi = \chi_0} w + O_M\big(we^{-c\sqrt{\log w}} + \omega(dm) \log w\big) \end{aligned}$$

for all $w \geqslant \exp(q_1^{1/M})$, where $M > 0$ is arbitrarily large but fixed, and c is an absolute positive constant. Note that $\omega(d) \ll \log q$ for $d|q$. Hence, taking $M = (1 + \varepsilon)C$ yields (13.1) with parameters $\kappa = -1_{\chi=\chi_0}$ and $\log Q = \max\big\{q^{1/((1+\varepsilon)C)}, (\log \omega(m))^{1/(1-\varepsilon^2)}\big\}$. Moreover, (13.2) clearly holds with $k = 1$. Notice that our assumptions on x, m and q imply that

$$\log x \geqslant \max\{q^{1/C}, (\log \omega(m))^{1/(1-\varepsilon)}\} = (\log Q)^{1+\varepsilon}.$$

Consequently, Theorem 13.2 implies that

$$\sum_{\substack{r \leqslant x/d,\ (r,dm)=1}} \mu(r)\chi(r) \ll \frac{x/d}{(\log(x/d))^A} \ll \frac{x}{(\log x)^A}$$

(all the main terms vanish because either $\kappa = 0$ or $\kappa = -1$ here, whence $\Gamma(\kappa - j) = \infty$ for all $j \in \mathbb{Z}_{\geqslant 0}$). Inserting the above estimate into (13.13) completes the proof. □

Estimating Perron integrals without shifting contours

We now turn to the proof of Theorem 13.2. The argument leading to (13.6) can be made rigorous using the methods of Chapter 7, at least when $\mathrm{Re}(\kappa) > 1$. However, it cannot produce an approximation for $\sum_{n\leqslant x}\tau_\kappa(n)$ that is strong enough to detect all the lower order terms in the asymptotic estimation of $\sum_{n\leqslant x}\tau_\kappa(n)$. To prove Theorem 13.2 we need one additional idea.

Instead of estimating $\sum_{n\leqslant x}\tau_\kappa(n)$, we work with the weighted average $\sum_{n\leqslant x}\tau_\kappa(n)(\log n)^m$, where m is a fixed integer at our disposal. It is easy to go back and forth between these two sums using partial summation. Moreover, the Dirichlet series of $\tau_\kappa(n)(\log n)^m$ is $(-1)^m(\zeta^\kappa)^{(m)}(s)$, where $(\zeta^\kappa)^{(m)}$ denotes the mth derivative of ζ^κ. Hence

$$\sum_{n\leqslant x}\tau_\kappa(n)(\log n)^m = \frac{(-1)^m}{2\pi i}\int_{(\alpha)}(\zeta^\kappa)^{(m)}(s)\frac{x^s}{s}\mathrm{d}s \tag{13.14}$$

for $x\notin\mathbb{Z}$ and any $\alpha>1$. Using Exercise 8.4(c), it is possible to show that $(\zeta^\kappa)^{(m)}(s)/s \ll_m |t|^{-1/2}$ for $\sigma\geqslant 1$ and $|t|\geqslant 1$, which tends to 0 when $|t|\to\infty$, no matter how large m is. On the other hand, for s close to 1, we have $(\zeta^\kappa)^{(m)}(s)/s \sim (-1)^m\kappa(\kappa+1)\cdots(\kappa+m-1)/(s-1)^{\kappa+m}$. Choosing m large enough ensures that our integrand is much bigger for small $|t|$ than for large $|t|$. This allows for a much better estimation of the integral on the right-hand side of (13.14). We provide the necessary details below.

We begin with an auxiliary result. We postpone its proof till the end of the chapter because it is rather technical in the general case, while being easy in the prototypical and important case when $f=\tau_\kappa$ for which (13.1) holds with $Q=e$: the analyticity and the non-vanishing of $\zeta(s)(s-1)$ when $\mathrm{Re}(s)=1$ yields parts (a), (b) and (d), whereas Exercise 8.4 yields part (c).

Lemma 13.5. *Let f and c_j be as in the statement of Theorem* 13.2*, and let F be the Dirichlet series of f. All implied constants might depend on k and the implicit constants in* (13.1)*.*

(a) *$F(s)(s-1)^\kappa$ has a C^∞-extension to the half-plane $\mathrm{Re}(s)\geqslant 1$.*

(b) *For $j=0,1,2,\ldots$, we have $c_j \ll_j (\log Q)^{j+2k}$.*

(c) *For $m\in\mathbb{Z}_{\geqslant 0}$ and $\varepsilon>0$, we have*

$$F^{(m)}(s) \ll_{m,\varepsilon} |t|^\varepsilon + (\log Q)^{m+3k} \quad \text{for } \sigma\geqslant 1,\ |t|\geqslant 1/\log Q.$$

(d) *For $m,J\in\mathbb{Z}_{\geqslant 0}$ and $|s-1|\leqslant 2/\log Q$ with $\sigma\geqslant 1$, we have*

$$(-1)^m F^{(m)}(s) = \sum_{0\leqslant j<J} c_j\frac{\Gamma(\kappa-j+m)}{\Gamma(\kappa-j)}(s-1)^{j-m-\kappa} + O_{m,J}((\log Q)^{J+2k}|s-1|^{J-m-\mathrm{Re}(\kappa)}).$$

Proof of Theorem 13.2. All implied constants might depend on k, J and ε. In addition, they might depend on the size of κ. However, note that for (13.1) and (13.2) both to hold, we must have $|\kappa| \leqslant k$. So the dependence on κ can be absorbed into the dependence on k. The implied constants will also depend on an integer m we will choose later in terms of k, J and ε.

Instead of estimating the partial sums of f, we work with the function $f(n)(\log n)^m$. We will use a smooth variant of Perron inversion to rewrite the partial sums of this function. Let

$$T = (\log x)^{2k+J+1} \quad \text{and} \quad w(s) = T \cdot \big[(1+1/T)^{s+1} - 1\big]/(s+1). \tag{13.15}$$

Then (7.3) implies that

$$\sum_{n\leqslant x} f(n)(\log n)^m = \frac{(-1)^m}{2\pi i} \int_{\mathrm{Re}(s)=1+1/\log x} F^{(m)}(s) w(s) \frac{x^s}{s} \mathrm{d}s + R, \tag{13.16}$$

where $|R| \leqslant \sum_{x<n\leqslant x+x/T} |f(n)|(\log n)^m$. Since $|f| \leqslant \tau_k$ and T is a power of $\log x$, Theorem 7.4 implies that

$$|R| \ll x(\log x)^{m+k-1}/T \leqslant x(\log x)^{m-k-1-J}. \tag{13.17}$$

Next, we turn to the main term in (13.16), which we write as $I_1 + I_2 + I_3$ with I_1 denoting the portion of the integral with $|\operatorname{Im}(s)| \leqslant 1/\log Q$, I_2 being the portion with $1/\log Q < |\operatorname{Im}(s)| \leqslant T^2$, and I_3 being the remaining part.

First, we bound I_3. For $s = 1 + 1/\log x + it$, we note that $|F^{(m)}(s)| \leqslant \sum_{n\geqslant 1} |f(n)|(\log n)^m / n^{1+1/\log x}$. Using our assumption that $|f| \leqslant \tau_k$ and Theorem 7.4 (that we insert via partial summation), we infer that $F^{(m)}(s) \ll (\log x)^{m+k}$. In addition, we have $|w(s)| \ll T/|t|$. As a consequence,

$$I_3 \ll x(\log x)^{m+k}/T = x(\log x)^{m-J-k-1} \tag{13.18}$$

by the choice of T.

To bound I_2, we note that if $1/\log Q \leqslant |t| \leqslant T^2$ and m is large enough, then $w(s) \ll T$ and $F^{(m)}(s) \ll |t|^{1/2} + (\log Q)^{m+3k} \ll (\log x)^{m-J-k-1}/T^3$ by Lemma 13.5(b), since $\log x \geqslant (\log Q)^{1+\varepsilon}$. Consequently,

$$I_2 \ll x(\log x)^{m-J-k}. \tag{13.19}$$

It remains to estimate I_1. We use Lemma 13.5 to find that

$$I_1 = \sum_{0\leqslant j<J} \frac{\Gamma(\kappa-j+m)}{\Gamma(\kappa-j)} \cdot \frac{c_j}{2\pi i} \int_{\mathcal{L}} w(s)(s-1)^{j-m-\kappa} \frac{x^s}{s} \mathrm{d}s$$
$$+ O\Big(x(\log Q)^{J+2k} \int_{|t|\leqslant 1/\log Q} |s-1|^{J-m-\mathrm{Re}(\kappa)} \mathrm{d}t\Big),$$

where $\mathcal{L}$ denotes the vertical line segment $[1+1/\log x-i/\log Q, 1+1/\log x+i/\log Q]$. For $s\in\mathcal{L}$ and $j\in\mathbb{Z}\cap[0,J]$, we have $w(s)=1+O(1/T)$ and

$$(s-1)^{j-m-\kappa}\ll|s-1|^{j-m-\mathrm{Re}(\kappa)}\leqslant(\log x)^{m+\mathrm{Re}(\kappa)-j}.$$

Consequently,

$$I_1=\sum_{j=0}^{J-1}\frac{\Gamma(\kappa-j+m)}{\Gamma(\kappa-j)}\cdot\frac{c_j}{2\pi i}\int_{\mathcal{L}}\frac{x^s}{s(s-1)^{m+\kappa-j}}\mathrm{d}s+O(E),$$

where

$$E=x(\log x)^{m+\mathrm{Re}(\kappa)-J}(\log Q)^{J+2k-1}.$$

Assuming that $m\geqslant J+2+k\geqslant J+2+|\kappa|$, we have

$$\int_{\substack{\mathrm{Re}(s)=1+1/\log x\\ |t|\geqslant 1/\log Q}}\frac{x^s}{s(s-1)^{m+\kappa-j}}\mathrm{d}s\ll x(\log Q)^{m+\mathrm{Re}(\kappa)-j-1}\qquad(0\leqslant j<J).$$

Thus, Lemma 13.1 implies that

$$\frac{1}{2\pi i}\int_{\mathcal{L}}\frac{x^s}{s(s-1)^{m+\kappa-j}}\mathrm{d}s=\int_2^x\frac{(\log y)^{m+\kappa-j-1}}{\Gamma(m+\kappa-j)}\mathrm{d}y+O(x(\log Q)^{m+\mathrm{Re}(\kappa)-j-1}),$$

so that

$$I_1=\int_2^x P(\log y)(\log y)^m\mathrm{d}y+O(E)\qquad\text{with}\qquad P(w)=\sum_{j=0}^{J-1}\frac{c_jw^{\kappa-j-1}}{\Gamma(\kappa-j)}.$$

Combining this formula with (13.16)–(13.19), we deduce that

$$\sum_{n\leqslant x}f(n)(\log n)^m=\int_2^x P(\log y)(\log y)^m\mathrm{d}y+O(E).$$

Finally, we remove the weight $(\log n)^m$ with a simple partial summation argument to establish (13.10). Relation (13.11) then follows by expanding $\int_2^x((\log y)^{\beta-1}/\Gamma(\beta))\mathrm{d}y$ into an asymptotic series using integration by parts several times, much like we did in Example 1.6. □

Proof of Lemma 13.5. (a) The function $\zeta(s)(s-1)$ is analytic and nonzero in an open neighborhood of the half-plane $\mathrm{Re}(s)\geqslant 1$. Hence, it suffices to show that $F\zeta^{-\kappa}$ has a C^∞-extension to the half-plane $\mathrm{Re}(s)\geqslant 1$. We write $F\zeta^{-\kappa}=GH$, where

$$G(s)=\prod_p(1-1/p^s)^{\kappa-f(p)}\quad\text{and}\quad H(s)=\prod_p\frac{1+f(p)/p^s+f(p^2)/p^{2s}+\cdots}{(1-1/p^s)^{-f(p)}}.$$

The factors of $H(s)$ are $1+O(1/p^{2\sigma})$ by Taylor's theorem and (13.2). In particular, $H(s)$ is analytic for $\sigma>1/2$ and each derivative $H^{(m)}(s)$ is

uniformly bounded in the half-plane $\sigma \geqslant 1$. To establish the C^∞-extension of G, it is more convenient to work with its logarithm, for which we have

$$(\log G)^{(m)}(s) = \sum_{p,\ a\geqslant 1} \sum \frac{(-a\log p)^m (f(p)-\kappa)}{ap^{as}}. \tag{13.20}$$

The series on the right-hand side converges uniformly in compact subsets of the half-plane $\mathrm{Re}(s) \geqslant 1$ by (13.1) and partial summation (see the proof of Theorem 4.5). This completes the proof of part (a).

(b,c) By the above discussion, both parts will follow if we show that

$$G^{(m)}(s) \ll_m \max\{|t|^\varepsilon, \log Q\}^{m+2k} \quad \text{for } \sigma \geqslant 1,\ t \in \mathbb{R}. \tag{13.21}$$

Indeed, all derivatives of $\zeta^\kappa(s)(s-1)^\kappa$ are bounded in the vicinity of $s=1$. Together with (13.21), this yields part (b). To prove (c), we separate two cases. When $1/\log Q \leqslant |t| \leqslant 1$, we note that $(\zeta^\kappa)^{(m)}(s) \ll_m (\log Q)^{m+k}$, whereas when $|t| \geqslant 1$ we use the bound $(\zeta^\kappa)^{(m)}(s) \ll_m |t|^\varepsilon$, which is a consequence of Exercise 8.4(c). Together with (13.21), these estimates establish part (c) in all cases.

Let us now prove (13.21). We write $G = e^L$ and note that $G' = L'e^L$ and $G'' = L''e^L + (L')^2 e^L$. In general, $G^{(m)}$ is a finite linear combination of terms of the form $L^{(m_1)} \cdots L^{(m_r)} e^L$ with $m_1 + \cdots + m_r = m$ and $m_1, \dots, m_r \in \mathbb{Z}_{\geqslant 1}$. This reduces (13.21) to proving that

$$|L(s)| \leqslant 2k \log\log N + O_\varepsilon(1), \quad L^{(m)}(s) \ll_{m,\varepsilon} (\log N)^m \quad (m \geqslant 1) \tag{13.22}$$

for $\sigma \geqslant 1$ and $|t| \geqslant 1$, where $N = \exp(\max\{|t|^\varepsilon, \log Q\})$.

To prove (13.22), we adapt the proof of Theorem 11.2: we fix $A > \max\{m, \varepsilon^{-1}\}$, and use partial summation and (13.1) to find that

$$\sum_{p>N} \frac{(f(p)-\kappa)(-\log p)^m}{p^s} \ll_{m,A} (1+|t|/(\log N)^A)(\log N)^m \leqslant 2(\log N)^m.$$

Since $|f(p)-\kappa| \leqslant k + |\kappa| \leqslant 2k$ by (13.2), we also trivially have that

$$\Big|\sum_{p\leqslant N} \frac{(f(p)-\kappa)(-\log p)^m}{p^s}\Big| \leqslant \begin{cases} 2k\log\log N + O(1) & \text{if } m=0, \\ O((\log N)^m) & \text{if } m \geqslant 1, \end{cases}$$

as well as

$$\sum_{p,\ a\geqslant 2} \sum \frac{(-a\log p)^m (f(p)-\kappa)}{ap^{as}} = O_m(1).$$

This completes the proof (13.22), and hence of (13.21).

(d) Taylor's theorem implies that

$$\widetilde{F}(s) := F(s)(s-1)^\kappa = \sum_{0\leqslant j<J} c_j (s-1)^j + E(s), \tag{13.23}$$

where the remainder can be written as

$$E(s) = \int_1^s \widetilde{F}^{(J)}(z)\frac{(s-z)^{J-1}}{(J-1)!}\mathrm{d}z.$$

Dividing both sides of (13.23) by $(s-1)^\kappa$ and differentiating m times, we see that part (d) will follow if we can show that

$$E^{(\ell)}(s)(s-1)^{-\kappa-m+\ell} \ll_m |s-1|^{-\operatorname{Re}(\kappa)-m+J}(\log Q)^{J+2k} \tag{13.24}$$

when $|s-1| \leqslant 2/\log Q$ and $\sigma \geqslant 1$. Indeed, by induction on ℓ, we have

$$E^{(\ell)}(s) = \begin{cases} \int_1^s \widetilde{F}^{(J)}(z)(s-z)^{J-1-\ell}\mathrm{d}z/(J-1-\ell)! & \text{if } \ell \leqslant J-1, \\ \widetilde{F}^{(\ell)}(s) & \text{if } \ell \geqslant J. \end{cases}$$

Since $\widetilde{F}^{(n)}(s) \ll_n (\log Q)^{n+2k}$ for $z \in [1,s]$ by (13.21), we find that

$$E^{(\ell)}(s)(s-1)^\ell \ll (|s-1|\log Q)^{\max\{\ell,J\}}(\log Q)^{2k} \ll |s-1|^J(\log Q)^{J+2k}$$

for $|s-1| \leqslant 2/\log Q$. This shows (13.24), and thus part (d) of the lemma. □

Exercises

Exercise 13.1. Let $\kappa \in \mathbb{C}$. Estimate $\sum_{n\leqslant x}\kappa^{\omega(n)}$ and $\sum_{n\leqslant x}\kappa^{\omega(n)}\varphi(n)$.

Exercise 13.2. Fix $\varepsilon > 0$ and $A \geqslant 1$. Prove that, uniformly for $m \in \mathbb{N}$ and $x \geqslant 2 + \exp\{(\log\omega(m))^{1+\varepsilon}\}$, we have

$$\#\{\, n \leqslant x : (n,m) = 1 \,\} = x\prod_{p\mid m}\Big(1-\frac{1}{p}\Big) + O_A\Big(\frac{x}{(\log x)^A}\Big).$$

Exercise 13.3. Fix $\kappa \in \mathbb{C}$ and $\varepsilon > 0$. Given $m \in \mathbb{N}$, let $L(m) = \log(2+\omega(m))$. Uniformly for $m \in \mathbb{N}$ and $x \geqslant \exp\{L(m)^{1+\varepsilon}\}$, prove that

$$\sum_{n\leqslant x,\,(n,m)=1}\kappa^{\omega(n)} = \frac{c_\kappa f_\kappa(m)}{\Gamma(\kappa)}x(\log x)^{\kappa-1} + O_{\kappa,\varepsilon}(xL(m)^{2|\kappa|}(\log x)^{\operatorname{Re}(\kappa)-2}),$$

where $c_\kappa = \prod_p(1+\frac{\kappa-1}{p})(1-1/p)^{\kappa-1}$ and $f_\kappa(m) = \prod_{p\mid m}(1+\kappa/(p-1))^{-1}$.

Exercise 13.4 (Landau)**.** Let $b(n)$ be the indicator function of those $n \in \mathbb{N}$ that can be written as the sum of two squares. Prove that there is a constant $c > 0$ such that $\sum_{n\leqslant x} b(n) \sim cx/\sqrt{\log x}$ as $x \to \infty$. [*Hint:* $b(n) = 1$ if and only if ν is even whenever $p^\nu \| n$ with $p \equiv 3 \pmod 4$.]

Exercise 13.5. For $r > 0$ and $\varepsilon > 0$, we let $C_r(\varepsilon)$ be the contour $\{\, |s| = r : |\arg(s)| \leqslant \pi - \varepsilon \,\}$ traced counterclockwise. We then define the contour

$$\mathcal{H}_r(\varepsilon) = (-\infty - ir\sin\varepsilon, re^{-i(\pi-\varepsilon)}] \;+\; C_r(\varepsilon) \;+\; [re^{i(\pi-\varepsilon)}, -\infty + ir\sin(\varepsilon)).$$

The limit of $\mathcal{H}_r(\varepsilon)$ when $\varepsilon \to 0^+$ is denoted by $\mathcal{H}_r$ and is called a *Hankel contour* (see Figure 13.1). By convention, $\int_{\mathcal{H}_r} = \lim_{\varepsilon\to 0^+}\int_{\mathcal{H}_r(\varepsilon)}$. Prove that

$$\frac{1}{2\pi i}\int_{\mathcal{H}_r}\frac{x^s}{s^\kappa}\mathrm{d}s = \frac{1}{\Gamma(\kappa)}.$$

[*Hint:* Show that both sides are entire functions of κ.]

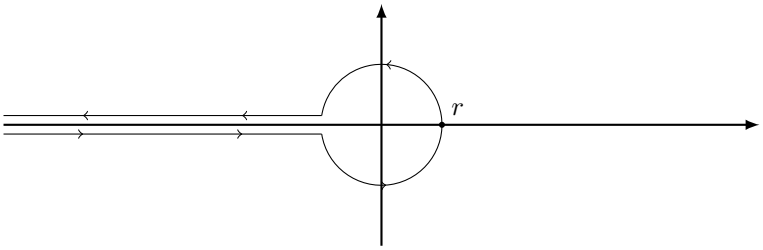

Figure 13.1. A Hankel contour

Exercise 13.6* ([162], [172, Chapter II.5]). Fix $\kappa \in \mathbb{C}$. Let $x \geqslant 3$, $\alpha = 1 + 1/\log x$ and $T \in [100, e^{\sqrt{\log x}}]$, and define $w(s)$ by (13.15).

(a) Prove that

$$\sum_{n \leqslant x} \tau_\kappa(n) = \frac{1}{2\pi i} \int_{\substack{\mathrm{Re}(s)=\alpha \\ |\mathrm{Im}(s)| \leqslant T^2}} \frac{\zeta^\kappa(s) w(s)}{s} x^s \mathrm{d}s + O_\kappa\Big(\frac{x(\log x)^{|\kappa|}}{T}\Big).$$

(b) Let $\delta = 0.5c_1/\log(2+T)$ with c_1 as in Exercise 8.4. In addition, consider $r \in (0, \delta)$ and let $\mathcal{H}'_r$ denote the truncated Hankel contour that goes from $-\delta - i0^+$ to $-r - i0^+$, then traces a circle of radius r to $-r + i0^+$, and finally goes to $-\delta + i0^+$. Prove that

$$\sum_{n \leqslant x} \tau_\kappa(n) = \frac{1}{2\pi i} \int_{\mathcal{H}'_r} \frac{\zeta^\kappa(s+1)}{s+1} x^{s+1} \mathrm{d}s + R,$$

where $|R| \leqslant x(\log x)^{O_\kappa(1)}(1/T + x^{-\delta})$. [*Hint:* After making the change of variables $s \to s+1$ in part (a), replace the contour $[1/\log x - iT^2, 1/\log x + iT^2]$ by the contour of Figure 13.2.]

(c) Develop $\zeta(s+1)^\kappa s^\kappa/(s+1)$ into Taylor series about $s = 0$ to give a new proof of Theorem 13.2 when $f = \tau_\kappa$.

Exercise 13.7.*

(a) When $\sigma \in (0,1)$, show that $\log \zeta(\sigma \pm i\varepsilon) \sim \log|\zeta(\sigma)| \mp i\pi$ as $\varepsilon \to 0^+$. [*Hint:* First, analyze $\log \zeta(s)$ when $s \sim 1$.]

(b) Use Exercise 13.6(b) to show there is a constant $c > 0$ such that

$$\sum_{n \leqslant x} \tau_{1/2}(n) = \frac{1}{\pi} \int_{1/2}^{1} \frac{|\zeta(\sigma)|^{1/2}}{\sigma} x^\sigma \mathrm{d}\sigma + O\big(xe^{-c\sqrt{\log x}}\big) \qquad (x \geqslant 2).$$

(c) Assume the Riemann Hypothesis and fix $\varepsilon > 0$. Show that

$$\sum_{n \leqslant x} \tau_{1/2}(n) = \frac{1}{\pi} \int_{1/2}^{1} \frac{|\zeta(\sigma)|^{1/2}}{\sigma} x^\sigma \mathrm{d}\sigma + O_\varepsilon(x^{1/2+\varepsilon}) \qquad (x \geqslant 2).$$

Exercise 13.8* (A partial converse to Theorem 13.2 [120]). Let f be a multiplicative function such that $|f| \leqslant 1$ and

$$\sum_{n \leqslant x} f(n) \ll_A x/(\log x)^A \qquad (x \geqslant 2) \tag{13.25}$$

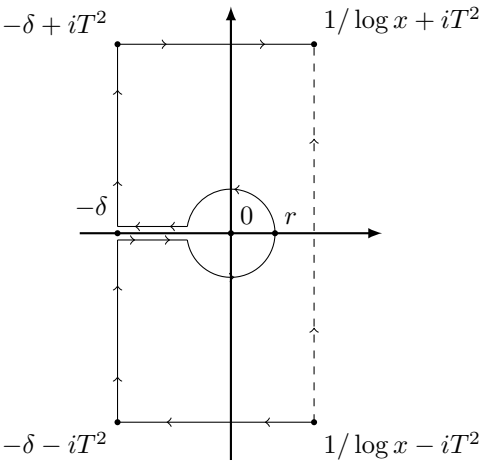

Figure 13.2. Deforming the contour $[1/\log x - iT^2, 1/\log x + iT^2]$

for all $A > 0$. Assume further that there is some $\delta > 0$ such that

$$\sum_{p\leqslant x} \frac{\mathrm{Re}(f(p)p^{-it})}{p} \geqslant (-1+\delta)\log\log x + O_B(1) \qquad (|t| \leqslant (\log x)^B,\ x \geqslant 2)$$

for each fixed $B > 0$. Then prove that

$$\sum_{p\leqslant x} f(p)\log p \ll_C x/(\log x)^C \qquad (x \geqslant 2)$$

for each fixed $C > 0$. [*Hint:* Let $F(s) = \sum_{n=1}^{\infty} f(n)/n^s$. Show that $\sum_{p>x} 1/p^{1+1/\log x} = O(1)$ and $\sum_{p\leqslant x}(p^{1/\log x} - 1)/p = O(1)$, and thus $|F(1 + 1/\log x + it)| = \exp\{\sum_{p\leqslant x} \mathrm{Re}(f(p)p^{-it})/p\} \gg_B (\log x)^{\delta-1}$ for $|t| \leqslant (\log x)^B$. Conclude that $|F(1 + 1/\log x + it)| \gg_B (\log x)^{\delta-1}$ for $|t| \leqslant (\log x)^B$ and hence $(F'/F)^{(m)}(1 + 1/\log x + it) \ll (\log x)^{m/2}$ for large $m \in N$.]

Chapter 14

Evolution of sums of multiplicative functions

The LSD method allows us to handle partial sums of multiplicative functions whose prime values are very regular. However, the information we have at our disposal is often more limited. In this chapter, we develop a technique that allows us to get a hold on partial sums of multiplicative functions under much weaker conditions. We mainly focus on non-negative functions, as they are easier to handle while still being a large enough class. For more advanced topics, see [38, Chapters 6 and 9], [75] and [172, Chapter III.4].

The underlying principle of the method we will use is very simple: if we know the average behavior of f over integers $n \leqslant x/2$, and over prime powers $p^k \leqslant x$, then we also know the average behavior of f over integers $m \leqslant x$. Indeed, any integer $m \leqslant x$ can be written as $m = p^k n$, with $p^k \leqslant x$ and n an integer $\leqslant x/p^k \leqslant x/2$ that is coprime to p, in which case we also have $f(m) = f(n)f(p^k)$. This simple fact should in principle imply that $S(x) = \sum_{n \leqslant x} f(n)$ obeys a recurrence relation involving the quantities $S(y)$ with $y \leqslant x/2$ and the numbers $f(p^k)$ with $p^k \leqslant 2x$.

An elegant way to derive the claimed recurrence begins with the obvious identity $F' = (F'/F) \cdot F$, where F is the Dirichlet series of f. Hence

$$f \log = \Lambda_f * f, \tag{14.1}$$

where Λ_f is the arithmetic function associated to the Dirichlet series $-F'/F$. This function generalizes von Mangoldt's function that satisfies (14.1) with $f = 1$. For instance, the definition of Λ_f readily implies that

$$\Lambda_f(p) = f(p) \log p.$$

In addition, similarly to Λ, the function Λ_f is supported on prime powers (see Exercise 4.7).

Example 14.1. (a) Let $\kappa \in \mathbb{C}$. If τ_κ is the function defined in (13.3), then $F(s) = \zeta(s)^\kappa$, whence $F'/F = \kappa\zeta'/\zeta$. Consequently, $\Lambda_f = \kappa\Lambda$.

(b) If $f = \mu^2\kappa^\omega$, then we have $F(s) = \prod_p(1 + \kappa/p^s)$, whence $\log F(s) = \sum_p \sum_{m=1}^\infty (-1)^{m-1}\kappa^m/(mp^{ms})$. We infer that $\Lambda_f(p^m) = (-1)^{m-1}\kappa^m \log p$, which grows exponentially fast in m. □

Now, using (14.1), we arrive at the formula

$$\sum_{n\leqslant x} f(n)\log n = \sum_{a\leqslant x} \Lambda_f(a) \sum_{b\leqslant x/a} f(b). \tag{14.2}$$

The slow growth of the logarithm implies that the left-hand side is $\approx (\log x)\sum_{n\leqslant x} f(n)$. Hence, (14.2) allows us to write $S(x) = \sum_{n\leqslant x} f(n)$ as a weighted average of $S(y)$ with $y \leqslant x/2$, where the weight is controlled by the values $f(p^k)$ with $p^k \leqslant x$. This establishes the claimed recurrence for the partial sums of any multiplicative function. We give two applications of this fundamental principle of multiplicative functions in Theorems 14.2 and 14.3.

Before we continue, we need to make some technical preparation. We face the problem that Λ_f can grow very rapidly even if f is divisor-bounded (i.e., $|f| \leqslant \tau_k$ for some $k \geqslant 0$). For instance, when $f = \mu^2\kappa^\omega$ with $|\kappa| > 1$, Example 14.1(b) shows that $\Lambda_f(p^m)$ grows exponentially fast in m. On the other hand, we saw in Example 14.1(a) that Λ_f is very tame when $f = \tau_\kappa$. Motivated by these observations, we introduce the function τ_f, defined as the arithmetic function whose formal Dirichlet series is given by $\prod_p(1 - 1/p^s)^{-f(p)}$. In other words, τ_f is the multiplicative function with

$$\tau_f(p^m) = \binom{f(p) + m - 1}{m} \tag{14.3}$$

for all prime powers p^m.

Working with τ_f alleviates various technicalities. For instance, we immediately see that its Dirichlet inverse equals τ_{-f}, and that Λ_{τ_f} is given by the simple formula $\Lambda_{\tau_f}(p^m) = f(p)\log p$ for all prime powers p^m. Moreover, we can easily relate f and τ_f with a simple convolution trick: if we write

$$f = \tau_f * r_f,$$

then the function r_f is supported on square-full integers. Since these numbers are very sparse (see Exercise 1.6), r_f often satisfies the inequality

$$\sum_{n\leqslant x} |r_f(n)| \ll x^{1-\delta} \quad (x \geqslant 1) \tag{14.4}$$

for some fixed $\delta > 0$. If, for instance, $|f| \leqslant \tau_k$ as in Chapter 13, then $|r_f| = |f * \tau_{-f}| \leqslant \tau_{2k}$, whence (14.4) holds for any $\delta < 1/2$ (see Exercises 2.9(f) and 1.6(b)). Exercise 14.6 establishes (14.4) in many more cases.

Assuming (14.4), we use Dirichlet's hyperbola method to find that

$$\sum_{n\leqslant x} f(n) = \sum_{a\leqslant y} r_f(a) \sum_{b\leqslant x/a} \tau_f(b) + R \tag{14.5}$$

for all $x \geqslant y \geqslant 1$, where

$$R = \sum_{b\leqslant x/y} \tau_f(b) \sum_{y<a\leqslant x/b} r_f(a) \ll x^{1-\delta} \sum_{b\leqslant x/y} \frac{|\tau_f(b)|}{b^{1-\delta}} \leqslant \frac{x}{y^\delta} \sum_{b\leqslant x/y} \frac{\tau_{|f|}(b)}{b}.$$

If we extend the summation to all integers b all of whose prime factors are $\leqslant x$, we arrive at the bound

$$\begin{aligned} R &\ll \frac{x}{y^\delta} \prod_{p\leqslant x} \left(1 + \frac{\tau_{|f|}(p)}{p} + \frac{\tau_{|f|}(p^2)}{p^2} + \cdots\right) \\ &= \frac{x}{y^\delta} \prod_{p\leqslant x} \left(1 - \frac{1}{p}\right)^{-|f(p)|}, \end{aligned} \tag{14.6}$$

where we used Taylor's theorem to obtain the last equality. If we also know that $|f| \leqslant \tau_k$, then $R \ll_k xy^{-\delta}(\log x)^k$ (and we can take any $\delta < 1/2$, as we discussed above). Exercise 14.4 establishes similar results when f satisfies weaker versions of the growth inequality $|f| \leqslant \tau_k$.

Our first application of (14.2) is a general purpose upper bound for the partial sums of non-negative multiplicative functions. What we will demonstrate is that, under some mild conditions, the mean value $x^{-1}\sum_{n\leqslant x} f(n)$ is controlled by the logarithmic mean value $(\log x)^{-1}\sum_{n\leqslant x} f(n)/n$. This is a rather special property of multiplicative functions (cf. Exercise 1.8(b)).

Notice that Theorem 14.2 below is sharp in the generality in which it is stated, since taking $f = \tau_k$ makes both sides of size $\asymp_k x(\log x)^{k-1}$. Notice also that Theorem 14.2 can be used to prove upper bounds when sieving the integers $n \leqslant x$ with a set of primes $\mathcal{P}$ (see Exercise 14.4).

Theorem 14.2. *If f is a multiplicative function such that $0 \leqslant f \leqslant \tau_k$, then*

$$\sum_{n\leqslant x} f(n) \ll_k x \cdot \exp\left\{\sum_{p\leqslant x} \frac{f(p)-1}{p}\right\}.$$

Proof. We first prove the theorem in the special case when $\tau_f = f$. In particular, we have $\Lambda_f(p^m) = f(p)\log p$, so that $\Lambda_f \leqslant k\Lambda$. Together with

(14.2) and Chebyshev's estimate, this implies that

$$\sum_{n\leqslant x} f(n)\log n \leqslant k\sum_{a\leqslant x} f(a)\sum_{b\leqslant x/a}\Lambda(b) \ll kx\sum_{a\leqslant x}\frac{f(a)}{a}.$$

We could now use partial summation to remove $\log n$ from the leftmost sum. More simply, note that

$$\begin{aligned}(\log x)\sum_{n\leqslant x} f(n) &= \sum_{n\leqslant x} f(n)\log n + \sum_{n\leqslant x} f(n)\log(x/n)\\ &\leqslant \sum_{n\leqslant x} f(n)\log n + \sum_{n\leqslant x} f(n)\cdot\frac{x}{n},\end{aligned}$$

whence

$$\sum_{n\leqslant x} f(n) \ll_k \frac{x}{\log x}\sum_{n\leqslant x}\frac{f(n)}{n}.$$

To complete the proof, we use the idea leading to (14.6): we have that

$$\sum_{n\leqslant x}\frac{f(n)}{n} \leqslant \sum_{p|n\ \Rightarrow\ p\leqslant x}\frac{f(n)}{n} = \prod_{p\leqslant x}\Big(\sum_{m=0}^{\infty}\frac{f(p^m)}{p^m}\Big) = \prod_{p\leqslant x}\Big(1-\frac{1}{p}\Big)^{-f(p)}.$$

Since $(1-t)^{-1}\leqslant e^{t+2t^2}$ for $t\in[0,1/2]$ by Taylor's theorem applied to the function $\log(1-t)$, and $f(p)\leqslant k$ for all p by assumption, we conclude that

$$\sum_{n\leqslant x}\frac{f(n)}{n} \leqslant e^{2ck}\exp\Big\{\sum_{p\leqslant x}\frac{f(p)}{p}\Big\}, \tag{14.7}$$

where $c=\sum_p 1/p^2$. Mertens' second estimate (Theorem 3.4(b)) then completes the proof when $f=\tau_f$.

Finally, we consider the general case. As we discussed above, (14.4) holds for any $\delta<1/2$ when $|f|\leqslant\tau_k$. Moreover, the remainder term R in (14.5) satisfies the bound $R\ll_k xy^{-\delta}(\log x)^k$. Taking $y=\sqrt{x}$ implies that

$$\begin{aligned}\sum_{n\leqslant x} f(n) &= \sum_{a\leqslant\sqrt{x}} r_f(a)\sum_{b\leqslant x/a}\tau_f(b) + O_k(x/\log x)\\ &\ll_k x\sum_{a\leqslant\sqrt{x}}\frac{|r_f(a)|}{a}\exp\Big\{\sum_{p\leqslant x/a}\frac{f(p)-1}{p}\Big\} + \frac{x}{\log x}.\end{aligned}$$

The term $x/\log x$ is $\ll x\exp\{\sum_{p\leqslant x}(f(p)-1)/p\}$ because $f\geqslant 0$ here. For the sum over a, note that Merten's second estimate implies that $\sum_{x/a<p\leqslant x}1/p = O(1)$ when $a\leqslant\sqrt{x}$. Moreover, $\sum_{a=1}^{\infty}|r_f(a)|/a$ converges by (14.4) and partial summation. The claimed estimate on $\sum_{n\leqslant x} f(n)$ thus follows. □

Our second application of (14.2) is a result due to Wirsing [186,187] that should be compared with Theorem 13.2. The idea underlying its proof is that if $f(p)\approx\kappa$ on average, then (14.2) implies that $\sum_{n\leqslant x} f(n)$ satisfies an

approximate differential equation. For technical reasons, it is much easier to work with logarithmic averages.

Theorem 14.3. *Fix $k \geqslant 0$, $c \in [0,1)$ and $\kappa \in \mathbb{C}$, and consider a multiplicative function f such that $|f| \leqslant \tau_k$,*

$$\sum_{p \leqslant x} \frac{f(p)\log p}{p} = \kappa \log x + O(1) \qquad (x \geqslant 2) \tag{14.8}$$

and

$$\sum_{p \leqslant x} \frac{|f(p)| - \operatorname{Re}(f(p))}{p} \leqslant c \log\log x + O(1) \qquad (x \geqslant 2). \tag{14.9}$$

We then have

$$\sum_{n \leqslant x} \frac{f(n)}{n} = \frac{\mathfrak{S}(f)}{\Gamma(\kappa+1)} \cdot (\log x)^{\kappa} + O((\log x)^{\operatorname{Re}(\kappa)+c-1}) \qquad (x \geqslant 2),$$

where $\mathfrak{S}(f) = \prod_p (1-1/p)^{\kappa}(1 + f(p)/p + f(p^2)/p^2 + \cdots)$. The implied constant depends at most on k, the distance of c from 1, and the implied constants in (14.8) *and* (14.9).

Proof. As in Theorem 14.2, it suffices to consider the case when $f = \tau_f$.

Let $S(x) := \sum_{n \leqslant x} f(n)/n$ and note that

$$\begin{aligned} \sum_{n \leqslant e^w} \frac{f(n)\log n}{n} &= \sum_{m \leqslant e^w} \frac{f(m)}{m} \sum_{a \leqslant e^w/m} \frac{\Lambda_f(a)}{a} \\ &= \sum_{m \leqslant e^w} \frac{f(m)}{m} \big(\kappa \log(e^w/m) + O(1)\big) \end{aligned} \tag{14.10}$$

for all $w \geqslant 1$. In addition,

$$\sum_{m \leqslant e^w} \frac{f(m)}{m} \log(e^w/m) = \sum_{m \leqslant e^w} \frac{f(m)}{m} \int_m^{e^w} \frac{\mathrm{d}y}{y} = \int_1^{e^w} \frac{S(y)}{y} \mathrm{d}y$$

by interchanging the order of summation and integration. Lastly, note that

$$\Big| \sum_{m \leqslant e^w} \frac{f(m)}{m} \Big| \leqslant \sum_{m \leqslant e^w} \frac{|f(m)|}{m} \ll w^{\operatorname{Re}(\kappa)+c} \qquad (w \geqslant 1) \tag{14.11}$$

by (14.7), since $\sum_{p \leqslant x} |f(p)|/p \leqslant (\operatorname{Re}(\kappa)+c)\log\log x + O(1)$ by (14.8) and (14.9). Consequently,

$$S(e^w) = \kappa \int_1^{e^w} \frac{S(y)}{y} \mathrm{d}y + O(w^{\operatorname{Re}(\kappa)+c}) \qquad (w \geqslant 1). \tag{14.12}$$

On the other hand, partial summation implies that $\sum_{n\leqslant e^w} f(n)(\log n)/n = wS(e^w) - \int_1^{e^w} (S(y)/y)\mathrm{d}y$. Thus

$$wS(e^w) = (\kappa+1)\int_1^{e^w} \frac{S(y)}{y}\mathrm{d}y + O(w^{\mathrm{Re}(\kappa)+c}) \qquad (w \geqslant 1).$$

We bound the part of the integral over $y \in [1,e]$ trivially by $\ll 1$. In addition, we let $y = e^u$ and

$$S(e^u) = u^\kappa g(u).$$

Hence, we arrive at the formula

$$w^{\kappa+1}g(w) = (\kappa+1)\int_1^w u^\kappa g(u)\mathrm{d}u + O(w^{\mathrm{Re}(\kappa)+c}) \qquad (w \geqslant 1).$$

Notice that if we had an exact equality $w^{\kappa+1}g(w) = (\kappa+1)\int_1^w u^\kappa g(u)\mathrm{d}u$ and g were a differentiable function, then we would immediately infer that $g'(u) = 0$, that is to say, g is a constant function. We will give an asymptotic version of this argument.

Let

$$E(w) = g(w) - \frac{\kappa+1}{w^{\kappa+1}}\int_1^w u^\kappa g(u)\mathrm{d}u, \tag{14.13}$$

so that

$$E(w) = O(w^{c-1}) \qquad (w \geqslant 1). \tag{14.14}$$

We multiply $E(w)$ by $1/w$ and integrate over $w \in [1,z]$ to find that

$$\begin{aligned}\int_1^z \frac{E(w)}{w}\mathrm{d}w &= \int_1^z \frac{g(w)}{w}\mathrm{d}w - \int_1^z u^\kappa g(u)\int_u^z \frac{\kappa+1}{w^{\kappa+2}}\mathrm{d}w\mathrm{d}u\\ &= \frac{1}{z^{\kappa+1}}\int_1^z u^\kappa g(u)\mathrm{d}u.\end{aligned}$$

Together with (14.13), this implies that

$$g(z) = E(z) + (\kappa+1)\int_1^z E(w)\frac{\mathrm{d}w}{w}. \tag{14.15}$$

By (14.14) and our assumption that $c < 1$, the integral on the right-hand side of (14.15) converges and its tails are $\ll z^{c-1}$. Hence

$$g(z) = \lambda + O(z^{c-1}) \qquad \text{with} \qquad \lambda := (\kappa+1)\int_1^\infty E(w)\frac{dw}{w}. \tag{14.16}$$

Taking $z = \log x$ completes the proof of the theorem, as long as we can show that $\lambda = \mathfrak{S}(f)/\Gamma(\kappa+1)$. To do this, we compute $\mathfrak{S}(f)$ in two ways.

Let F be the Dirichlet series attached to f and note that

$$\mathfrak{S}(f) = \lim_{\sigma\to 1^+} F(\sigma)\zeta(\sigma)^{-\kappa}.$$

Since $\zeta(\sigma) \sim 1/(\sigma-1)$ when $\sigma \to 1^+$, we can rewrite the above relation as

$$\mathfrak{S}(f) = \lim_{\sigma \to 1^+} F(\sigma)(\sigma-1)^\kappa. \tag{14.17}$$

In addition, partial summation and (14.16) imply that

$$\begin{aligned} F(\sigma) &= (\sigma-1)\int_0^\infty S(e^u)e^{-(\sigma-1)u}\mathrm{d}u \\ &= (\sigma-1)\int_0^\infty (\lambda u^\kappa + O(u^{\mathrm{Re}(\kappa)+c-1}+1))e^{-(\sigma-1)u}\mathrm{d}u \\ &= \big(\lambda\Gamma(\kappa+1) + O((\sigma-1)^{1-c})\big)(\sigma-1)^{-\kappa}. \end{aligned}$$

Since $c<1$, we conclude that $\lim_{\sigma\to 1^+} F(\sigma)(\sigma-1)^\kappa = \lambda\Gamma(\kappa+1)$. Comparing this relation to (14.17) yields our claim that $\lambda = \mathfrak{S}(f)/\Gamma(\kappa+1)$. □

Delay differential equations

Wirsing's theorem capitalizes on the idea that the evolution of the function $S(x) = \sum_{n\leqslant x} f(n)$ is controlled by a differential equation, a consequence of (14.2). In fact, an important aspect of (14.2) is that, since $\Lambda_f(1) = 0$, the right-hand side only involves values of $S(t)$ with $t \leqslant x/2$. There is thus a certain *delay* on the right-hand side of (14.2). This feature is amplified if we consider functions f that are supported on integers free of prime factors $\leqslant y$, since then the right-hand side of (14.2) only involves values of $S(t)$ with $t \leqslant x/y$. The simplest such example is the indicator function of y-*rough integers*, namely integers all of whose prime factors are $> y$. We denote their summatory function by

$$\Phi(x,y) := \#\{\, n \leqslant x : P^-(n) > y \,\},$$

where we recall that $P^-(n)$ is the smallest prime factor of n with the convention that $P^-(1) = \infty$.

The function $\Phi(x,y)$ is closely related to the sieve of Eratosthenes-Legendre. In particular, when $y = \sqrt{x}$, we see that $\Phi(x,\sqrt{x}) \sim x/\log x$ by the Prime Number Theorem. On the other extreme, Theorem 2.1 with $m = \prod_{p\leqslant y} p$ implies that $\Phi(x,y) \sim e^{-\gamma}x/\log y$ when y tends to infinity at a rate such that $y \leqslant \log x$, since then $\omega(m) \leqslant \pi(y) \ll \log x/\log\log x$. We want to fill in the gap and understand how $\Phi(x,y)$ evolves when y goes from $\sqrt{x}$ to $\log x$.

Using (14.2) with $f(n) = 1_{P^-(n)>y}$, for which $\Lambda_f(p^k) = 1_{p>y}\log p$, we find that

$$\sum_{n\leqslant x,\, P^-(n)>y} \log n = \sum_{p^k\leqslant x,\, p>y}\sum \Phi(x/p^k, y)\log p. \tag{14.18}$$

The left-hand side should be roughly $\Phi(x,y)\log x$, while prime powers p^k with $k \geqslant 2$ should not contribute significantly to the right-hand side. We should thus have

$$(\log x)\Phi(x,y) \approx \sum_{y<p\leqslant x} \Phi(x/p,y)\log p.$$

Note that $\Phi(x/p,y) = 1$ when $p > x/y$, since the only integer $\leqslant x/p < y$ free of prime factors $\leqslant y$ is the number 1. Consequently,

$$(\log x)\Phi(x,y) \approx x + \sum_{y<p\leqslant x/y} \Phi(x/p,y)\log p. \tag{14.19}$$

If we pretend for a moment that $x \to \Phi(x,y)$ is a continuously differentiable function, the Prime Number Theorem suggests that the sum over p is $\approx \int_y^{x/y} \Phi(x/w,y)\mathrm{d}w$. Letting $x/w = y^t$ and $u = \log x/\log y$, we find that

$$(\log x)\Phi(x,y) \approx x + x\int_1^{u-1} \frac{\Phi(y^t,y)}{y^t/\log y}\mathrm{d}t. \tag{14.20}$$

This relation suggests that there is a function B such that

$$\Phi(x,y) \sim \frac{xB(u)}{\log y} \quad (x = y^u,\ u \geqslant 1,\ y \to \infty). \tag{14.21}$$

For consistency with the estimate $\Phi(x,y) \sim x/\log x$ when $\sqrt{x} \leqslant y \leqslant x/\log x$, and with the recursive relation (14.20), we must have that

$$B(u) = \frac{1}{u} \quad (1 \leqslant u \leqslant 2), \quad uB(u) = 1 + \int_1^{u-1} B(v)\mathrm{d}v \quad (u \geqslant 2). \tag{14.22}$$

The above relations together define a unique continuous function $B : \mathbb{R}_{\geqslant 1} \to \mathbb{R}$ called *Buchstab's function*. It is usually denoted by ω, but we use the letter B here to avoid confusion with the arithmetic function $\omega(n)$.

For consistency with Theorem 2.1 and Mertens' third estimate, we should have that $\lim_{u\to\infty} B(u) = e^{-\gamma}$. Exercises 14.7 and 14.11 give two ways of proving this guess rigorously. For now, we note that $1/u \leqslant B(u) \leqslant 1$ for $u \geqslant 1$, as can be seen by (14.22) and induction on $\lfloor u \rfloor$. Moreover, B is differentiable in $(1,2) \cup (2,+\infty)$ and its derivative satisfies the *delay differential equation*

$$uB'(u) = B(u-1) - B(u) \quad \text{for } u > 2.$$

As we will see again later on, the solutions to delay differential equations rule the asymptotic behavior of various sieve-theoretic functions.

We now prove that (14.21) is indeed true.

Theorem 14.4. *Fix $u > 1$. If $x = y^u$, then*

$$\Phi(x,y) = \frac{xB(u)}{\log y} + O_u\Big(\frac{x}{(\log x)^2}\Big).$$

Proof. We argue by induction on $\lfloor u \rfloor$. When $1 < u \leqslant 2$, the theorem follows by the Prime Number Theorem. Assume now its validity when $u \leqslant N$ for some $N \in \mathbb{Z}_{\geqslant 2}$, and consider $u \in (N, N+1]$.

We begin by simplifying (14.18). Note that

$$\sum_{k\geqslant 2} \sum_{p^k \leqslant x/y,\, p>y} \Phi(x/p^k, y) \log p \leqslant \sum_{p\geqslant y} \sum_{k\geqslant 2} \frac{x \log p}{p^k} = \sum_{p\geqslant y} \frac{x \log p}{p(p-1)} \ll \frac{x}{y},$$

where the last inequality follows by Chebyshev's estimate (Theorem 2.4) and partial summation. In addition, Theorem 14.2 with $f(n) = 1_{P^-(n)>y}$ implies that $\Phi(x,y) \ll x/\log y$ for $x \geqslant y \geqslant 2$. Together with partial summation, this yields the estimate $\sum_{n\leqslant x,\, P^-(n)>y} \log(x/n) \ll x/\log y$. We combine this inequality with relation (14.18) to deduce that

$$\begin{aligned}(\log x) &= \sum_{n\leqslant x,\, P^-(n)>y} \log n + O(x/\log y)\\ &= \sum_{y<p\leqslant x} \Phi(x/p, y) \log p + O(x/\log y).\end{aligned}$$

Since $\Phi(x/p, y) = 1$ for $p > x/y$, we conclude that

$$(\log x)\Phi(x,y) = \sum_{y<p\leqslant x/y} \Phi(x/p, y) \log p + x + O(x/\log y).$$

Finally, note that $x/p \leqslant y^{u-1}$ when $p > y$, so that the induction hypothesis can be applied to estimate $\Phi(x/p, y)$. Consequently,

$$(\log x)\Phi(x,y) = \frac{x}{\log y} \sum_{y<p\leqslant x/y} B\Big(\frac{\log(x/p)}{\log y}\Big)\frac{\log p}{p} + x + O_u\Big(\frac{x}{\log y}\Big).$$

Since B is continuous and $t \to \theta(y^t) = \sum_{p\leqslant y^t} \log p$ is a step function with jumps of length $\log p$ whenever $t = \log p/\log y$, we have

$$\sum_{y<p\leqslant x/y} B\Big(\frac{\log(x/p)}{\log y}\Big)\frac{\log p}{p} = \int_1^{u-1} B(u-t)\frac{\mathrm{d}\theta(y^t)}{y^t}. \tag{14.23}$$

Next, we write $\theta(y^t) = y^t(1 + \delta(y^t))$, and integrate by parts to find that[1]

$$\begin{aligned}\int_1^{u-1} B(u-t)\frac{\mathrm{d}\theta(y^t)}{y^t} &= (\log y)\int_1^{u-1} B(u-t)\mathrm{d}t + B(u-t)\delta(y^t)\Big|_{t=1}^{u-1}\\ &\quad + \int_1^{u-1} (B'(u-t) + B(u-t)\log y)\delta(y^t)\mathrm{d}t.\end{aligned} \tag{14.24}$$

[1] Strictly speaking, we have to treat separately the integrals over $[1, u']$ and $[u', u-1]$, where $u' = \min\{2, u-1\}$, because of the discontinuity of B' at 2. Formula (14.24) remains valid though.

Since $\delta(y^t) \ll 1/\log(y^t)$ by the Prime Number Theorem, we conclude that

$$\sum_{y<p\leqslant x/y} B\Big(\frac{\log(x/p)}{\log y}\Big)\frac{\log p}{p} = (\log y)\int_1^{u-1} B(u-t)\mathrm{d}t + O_u(1).$$

Plugging this formula into (14.23) and using (14.22) proves that $\Phi(x,y) = xB(u)/\log y + O_u(x/\log^2 x)$ for $u \in (N, N+1]$. This completes the inductive step, and hence the proof of the theorem. □

The "dual" to y-rough numbers are *y-smooth numbers*,[2] which are integers all of whose prime factors are $\leqslant y$. These numbers also play a central role in number theory, and we will enounter them again when we develop sieve methods in Part 4. Here, we use ideas from the theory of multiplicative functions to study their counting function

$$\Psi(x,y) := \#\{\, n \leqslant x : P^+(n) \leqslant y \,\},$$

where we recall that $P^+(n)$ denotes the largest prime factor of n with the convention that $P^+(1) = 1$.

Arguing heuristically and writing $x = y^u$ as before, we find that

$$(\log x)\Psi(x,y) \approx \sum_{n\leqslant x,\, P^+(n)\leqslant y} \log n \approx \sum_{p\leqslant y} \Psi(x/p, y)\log p \tag{14.25}$$

$$\approx \int_1^y \Psi(x/t, y)\mathrm{d}t$$

$$= (\log y)\int_{u-1}^u \frac{\Psi(y^v, y)}{y^v}\mathrm{d}v. \tag{14.26}$$

This relation leads us to conjecture that there is a function ρ such that

$$\Psi(x,y) \sim x\rho(u) \quad (x = y^u,\ u \geqslant 0,\ y \to \infty). \tag{14.27}$$

For consistency with the estimate $\Psi(x,y) = \lfloor x \rfloor = x + O(1)$ when $y > x$, and with (14.26) when $u \geqslant 1$, we must have that

$$\rho(u) = 1 \quad (0 \leqslant u \leqslant 1), \qquad u\rho(u) = \int_{u-1}^u \rho(v)\mathrm{d}v \quad (u \geqslant 1). \tag{14.28}$$

Together, these relations define a unique differentiable function $\rho : \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ called the *Dickman-de Bruijn function*. Note that differentiating the second formula in (14.28) yields the delay differential equation

$$u\rho'(u) = -\rho(u-1). \tag{14.29}$$

[2]The reason for this terminology comes from looking at the sequence of divisors of an integer. If n is y-smooth, then every interval $[z, zy]$, $z < x/z$, contains a divisor of n. On the other hand, if n is y-rough, then its divisors can have a much more singular distribution.

Unlike Buchstab's function that is of order of magnitude 1, the Dickman-de Bruijn function decays extremely rapidly. For instance, Exercise 14.10(c) states that $\rho(u) = e^{O(u)}(u \log u)^{-u}$.

We conclude this chapter with a proof of our guess (14.27).

Theorem 14.5. *Fix $u > 0$. For $x = y^u$, we have*

$$\Psi(x, y) = x\rho(u) + O_u(x/\log x).$$

Proof. Unlike (14.18), where $x/p^k \leqslant y^{u-1}$ in its range of summation, in (14.25) we only have the bound $x/p^k \leqslant x/2$. Hence, we cannot use this relation to induct on $\lfloor u \rfloor$. We could induct on $\lfloor \log x/\log 2 \rfloor$ (see Exercise 14.8, where such an induction is performed in another problem), but we present instead a proof that uses a different recursive formula due to Buchstab.

Note that if $z > y$ and n is z-smooth but not y-smooth, then $P^+(n) \in (y, z]$. Hence, we may uniquely write $n = pm$, where $p \in (y, z]$ and $P^+(m) \leqslant p$ (we simply take $p = P^+(n)$). This leads us to *Buchstab's identity*

$$\Psi(x, z) - \Psi(x, y) = \sum_{y < p \leqslant z} \Psi(x/p, p). \tag{14.30}$$

When $y \geqslant \sqrt{x}$ and $z = x$, we have $x/p \leqslant p$ for all $p > y$, so that (14.30) becomes

$$\begin{aligned}\Psi(x, y) &= \lfloor x \rfloor - \sum_{y < p \leqslant x} \lfloor x/p \rfloor \\ &= x - x \sum_{y < p \leqslant x} \frac{1}{p} + O(\pi(x)) \\ &= x(1 - \log u) + O(x/\log x)\end{aligned}$$

by Mertens' and Chebyshev's estimates (see Theorems 3.4(b) and 2.4, respectively). This establishes the theorem for $u \in [1, 2]$.

For the general case, we apply (14.30) with $z = \sqrt{x}$ to find that

$$\Psi(x, y) = x\rho(2) - \sum_{y < p \leqslant \sqrt{x}} \Psi(x/p, p) + O(x/\log x).$$

Since $\log(x/p)/\log y < u - 1$ for $p > y$, we may now induct on $\lfloor u \rfloor$ to establish the theorem, arguing similarly to the proof of Theorem 14.4. We leave the details as an exercise. □

Exercises

Exercise 14.1. Let f be as in Theorem 14.2. Show that

$$\sum_{n > x} \frac{f(n)}{n^2} \ll_{k,\lambda} \frac{1}{x} \cdot \exp\Big\{ \sum_{p \leqslant x} \frac{f(p) - 1}{p} \Big\} \qquad (x \geqslant 1).$$

[*Hint:* If M denotes the right-hand side of the above inequality, show that $\sum_{xe^j<n\leqslant xe^{j+1}} f(n)/n^2 \ll M\cdot e^{-j}\cdot(1+j/\log x)^{\max\{k-1,0\}}$ uniformly for $j\geqslant 0$.]

Exercise 14.2. For fixed $r\in\mathbb{R}$ and $k\in\mathbb{N}$, show that

$$\sum_{n\leqslant x}\tau_k(n)(\varphi(n)/n)^r \asymp_{r,k} x(\log x)^{k-1} \qquad (x\geqslant 2).$$

[*Hint*: To get the lower bound, use Theorem 7.4 and Hölder's inequality.]

Exercise 14.3. Let g be such that $\mu^2/\varphi = g*(1/\operatorname{id})$, where $\operatorname{id}(n)=n$.

(a) Calculate g on all prime powers. Then use Exercise 14.1 to show that

$$\sum_{n>x}|g(n)| \leqslant \sum_{ab^2>x}\frac{\mu^2(a)\mu^2(b)}{a\varphi(a)b\varphi(b)} \ll \frac{1}{\sqrt{x}} \qquad (x\geqslant 1).$$

(b) Prove that there is some constant c such that

$$\sum_{n\leqslant x}\frac{\mu^2(n)}{\varphi(n)} = \log x + c + O\big((\log x)/\sqrt{x}\big) \qquad (x\geqslant 2).$$

Exercise 14.4. Uniformly for $m\in\mathbb{N}$ and $x\geqslant 1$, prove that

$$\#\{\,n\leqslant x : (n,m)=1\,\} \ll x\cdot\prod_{p\mid m,\, p\leqslant x}(1-1/p).$$

Exercise 14.5. Let f be a multiplicative function with $0\leqslant f\leqslant \tau_k$.

(a) If g is such that $f*g=\tau_k$, then prove that $\mu^2(n)g(n)\geqslant 0$ and

$$\sum_{n\leqslant x}\frac{\mu^2(n)f(n)}{n}\sum_{n\leqslant x}\frac{\mu^2(n)g(n)}{n} \geqslant \sum_{n\leqslant x}\frac{\mu^2(n)\tau_k(n)}{n} \qquad (x\geqslant 1).$$

(b) Prove that

$$\sum_{n\leqslant x}\frac{f(n)}{n} \asymp_k \exp\Big\{\sum_{p\leqslant x}\frac{f(p)}{p}\Big\} \qquad (x\geqslant 1).$$

(c) If $\sum_{p\leqslant y} f(p)\geqslant cy/\log y$ for $y\in[\sqrt{x},x]$, then prove that

$$\sum_{n\leqslant x} f(n) \asymp_{k,c} x\cdot\exp\Big\{\sum_{p\leqslant x}\frac{f(p)-1}{p}\Big\}.$$

[*Hint:* Note that $\sum_{n\leqslant x} f(n)\geqslant \sum_{m\leqslant x^{1/3}} f(m)\sum_{x^{1/3}<p\leqslant x/m} f(p)$.]

(d) Let $b(n)$ be as in Exercise 13.4. Give a new proof of the estimate

$$\sum_{n\leqslant x} b(n) \asymp x/\sqrt{\log x} \qquad (x\geqslant 2).$$

Exercise 14.6.* Let f be a multiplicative function, and write $f=\tau_f * r_f$.

(a) For each $\varepsilon>0$, use Hölder's inequality to prove that

$$\sum_{n\leqslant x}|r_f(n)| \ll_\varepsilon x^{1-\varepsilon/(2+2\varepsilon)}\Big(\sum_{n\leqslant x}|r_f(n)|^{1+\varepsilon}/n\Big)^{1/(1+\varepsilon)}.$$

[*Hint:* Recall that r_f is supported on square-full integers.]

(b) Assume there are constants $k \geqslant 0$ and $\lambda \in [1,2)$ such that $|f(p^\nu)| \leqslant k\lambda^{\nu-1}$ for all prime powers p^ν. Prove that $|r_f(n)| \leqslant n^{o(1)}\lambda^{\Omega(n)}$. Deduce that (14.4) holds for some $\delta > 0$.

(c) Assume there is $\theta > 1$ such that $\sum_{p\leqslant x}\sum_{\nu\geqslant 1}|f(p^\nu)|^\theta/p^\nu \ll \log\log x$ for all $x \geqslant 3$.
  (i) If $\varepsilon < \theta - 1$, prove that $\sum_{\nu\geqslant 1}\tau_{|f|}(p^\nu)^{1+\varepsilon}/p^\nu \ll_{\varepsilon,\theta} |f(p)|^{1+\varepsilon}/p$. [*Hint:* Show that $f(p) \ll (p\log\log(p+1))^{1/\theta}$ and use Exercise 1.2(e).]
  (ii) Show that $|r_f(n)|^{1+\varepsilon} \leqslant \tau(n)^\varepsilon \sum_{ab=n}|f(a)|^{1+\varepsilon}\tau_{|f|}(b)^{1+\varepsilon}$, and conclude that (14.4) holds for any $\delta < (1-\theta)/(4-2\theta)$.

(d) Let f be as in part (c). Assume further that (14.8) and (14.9) hold. Prove that f satisfies the conclusions of Theorems 14.2 and 14.3.

(e) Let f be as in part (c). Assume further that (13.1) holds with $Q = 2$. Prove that f satisfies the conclusions of Theorem 13.2. [*Hint:* Use Hölder's inequality to bound $\sum_{x<n\leqslant x+x/T}|f(n)|$.]

Exercise 14.7.* Let $\kappa \in \mathbb{C}$, $c \in [0,1)$, $k, C \geqslant 0$ and $Q \geqslant 3$. Consider a multiplicative function f such that $\tau_f = f$ and $|f| \leqslant \tau_k$. Assume further that

$$\Big|\sum_{p\leqslant x}\frac{f(p)\log p}{p} - \kappa\log x\Big| \leqslant \log Q \quad\text{and}\quad \sum_{p\leqslant x}\frac{|f(p)| - \mathrm{Re}(f(p))}{p} \leqslant c\log\log x + C.$$

for all $x \geqslant Q$, and let $\mathfrak{S}(f)$ be as in Theorem 14.3. All implied constants below may depend on κ, k, c and C, but they must be uniform in x and Q.

(a) For $x \geqslant Q$, prove the following estimates:
  (i) $\sum_{p>x}(f(p)-\kappa)/p = O(\log Q/\log x)$.
  (ii) $\sum_{m\leqslant x}|f(m)|/m \ll |\mathfrak{S}(f)|(\log x)^{\mathrm{Re}(\kappa)+c}$ for $x \geqslant Q$.
  (iii) $\sum_{n\leqslant x}f(n)/n = (\mathfrak{S}(f)/\Gamma(\kappa+1))\cdot(\log x)^\kappa\cdot(1+O((\log Q)(\log x)^{c-1}))$. [*Hint:* Improve (14.14) to $E(w) \ll |\mathfrak{S}(f)|(\log Q)w^{c-1}$ for $w \geqslant \log Q$.]

(b) Prove that $\lim_{u\to\infty}B(u) = e^{-\gamma}$.

Exercise 14.8.* Let $f : \mathbb{N} \to \mathbb{C}$ be a multiplicative function such that $|f| \leqslant \tau_k$ for some $k \in \mathbb{N}$. Assume further that

$$\Big|\sum_{p\leqslant x}f(p)\log p\Big| \leqslant Mx/(\log x)^A \qquad (x \geqslant 2)$$

for some $M, A > 0$. Show that there is $M' = M'(k, A, M)$ such that

$$\Big|\sum_{n\leqslant x}f(n)\Big| \leqslant M'x/(\log x)^{A-k+1} \qquad (x \geqslant 2) \tag{14.31}$$

as follows. Firstly, reduce to the case when $f = \tau_f$. Then, prove that

$$\sum_{n\leqslant x}f(n)\log n = \sum_{ab\leqslant x,\, a\leqslant x^\varepsilon}\Lambda_f(a)f(b) + O_{\varepsilon,k,A,M}(x/(\log x)^{A-k})$$

for all $x \geqslant 2$ and each fixed $\varepsilon \in (0,1)$. Finally, induct on the dyadic interval $(2^{j-1}, 2^j]$ containing x to prove (14.31).

Exercise 14.9.* Let $f : \mathbb{N} \to \mathbb{C}$ be a multiplicative function satisfying (13.1) for some *fixed* $A > 1$ and $\kappa \in \mathbb{C}$. Assume further that (13.2) holds. Estimate the partial sums of f. [*Hint:* Write $f = \tau_\kappa * g$ and use Exercise 14.8 on g.]

Exercise 14.10. We extend the Dickman-de Bruijn function to all u by letting $\rho(u) = 0$ for $u < 0$. Note then that $u\rho(u) = \int_{u-1}^{u} \rho(v)\mathrm{d}v$ for all u.

(a) Show that $0 < \Gamma(\sigma) < 1$ when $1 < \sigma < 2$, as well as that $\Gamma(\sigma)$ is increasing for $\sigma \geqslant 2$. [*Hint:* Use Theorem 1.14 to show that $\log\Gamma$ is convex on $\mathbb{R}_{>0}$.]

(b) For $u > 0$, show that $0 < \rho(u) < 1/\Gamma(u+1)$. [*Hint:* Argue by contradiction and consider the smallest u that does not satisfy the inequalities $0 < \rho(u) < 1/\Gamma(u+1)$.]

(c) Prove that $\rho(u) = e^{-u\log(u\log u)+O(u)}$ for $u \geqslant 1$. [*Hint:* Note that if $ce^{-ta(t)} \leqslant \rho(t) \leqslant Ce^{-tb(t)}$ for all $t < u$, where $a, b\colon \mathbb{R}_{\geqslant 1} \to \mathbb{R}_{\geqslant 1}$ are increasing, then $c\int_{u-1}^{u} e^{-ta(u)}\mathrm{d}t \leqslant u\rho(u) \leqslant C\int_{u-1}^{\infty} e^{-tb(u-1)}\mathrm{d}t$.]

(d) Consider the Laplace transform of the Dickman-de Bruijn function

$$\widehat{\rho}(s) = \int_0^\infty \rho(t)e^{-st}dt \quad \text{for } \mathrm{Re}(s) > 0.$$

Show that $\widehat{\rho}\,'(s) = \widehat{\rho}(s)(e^{-s}-1)/s$, and conclude that there is a constant $c \in \mathbb{C}$ such that $\widehat{\rho}(s) = e^{c-f(s)}$, where $f(s) = \int_0^s z^{-1}(1-e^{-z})\mathrm{d}z$.

(e) Show that $s\widehat{\rho}(s) \to 1$ when $s \to \infty$ over positive real numbers. Use Exercise 5.4(c) to conclude that $c = \gamma$.

Exercise 14.11.

(a) For $u > 2$, show that $uB'(u) = -\int_{u-1}^{u} B'(v)\mathrm{d}v$ and $|B'(u)| \leqslant \rho(u)$.

(b) Show that $\lim_{u\to\infty} B(u)$ exists and equals $1 + \int_1^\infty B'(u)\mathrm{d}u$.

(c) If $\widehat{B}(s) = \int_1^\infty B(v)e^{-sv}\mathrm{d}v$, then show that $\widehat{B}\,'(s) = -e^{-s}s^{-1}(\widehat{B}(s)+1)$. Conclude that there is a constant $d \in \mathbb{C}$ such that $\widehat{B}(s) + 1 = s^{-1}e^{d+f(s)}$ with $f(s) = \int_0^s z^{-1}(1-e^{-z})\mathrm{d}z$. In particular, $\lim_{s\to 0^+} s\widehat{B}(s) = e^d$.

(d) Let $s \to \infty$ in the formula $\widehat{B}(s) + 1 = s^{-1}e^{d+f(s)}$ and use Exercise 5.4(c) to conclude that $d = -\gamma$. In particular, $\widehat{\rho}(s)(\widehat{B}(s)+1) = 1/s$.

(e) Let $L(s) = \int_1^\infty B'(v)e^{-sv}\mathrm{d}v$. Show that $\widehat{B}(s) = (e^{-s} + L(s))/s$. Deduce that $\lim_{u\to\infty} B(u) = e^{-\gamma}$.

(f) Show that the difference $E(u) := B(u) - e^\gamma$ changes signs infinitely often as $u \to \infty$. [*Hint:* Show that $E(u) = -\int_u^\infty B'(v)\mathrm{d}v = O(e^{-u})$. On the other hand, substituting $B = E + e^{-\gamma}$ in the formula $uB'(u) = -\int_{u-1}^{u} B'(v)\mathrm{d}v$, show that $uE(u) = -\int_{u-1}^{\infty} E(v)\mathrm{d}v$ for $u > 2$.]

Chapter 15

The distribution of multiplicative functions

To obtain a better understanding of multiplicative functions it is not enough to know their asymptotic behavior. We also must study the distribution of their values. A probabilistic framework thus arises.

Given a set $\mathcal{A} \subseteq \{n \leqslant x\}$, we write

$$\mathbb{P}_{n\leqslant x}(\mathcal{A}) = |\mathcal{A}| \,/\, \lfloor x \rfloor$$

for the probability that a *randomly chosen* integer $n \leqslant x$ lies in $\mathcal{A}$. The underlying σ-algebra is naturally the power set of $\mathbb{N}_{\leqslant x}$. Given a random variable $Z : \mathbb{N}_{\leqslant x} \to \mathbb{C}$, we write

$$\mathbb{E}_{n\leqslant x}[Z] = \frac{1}{\lfloor x \rfloor} \sum_{n\leqslant x} Z(n).$$

Within this framework, the value distribution of a real-valued function f is determined by the distribution function $\mathbb{R} \ni u \to \mathbb{P}_{n\leqslant x}(f(n) \leqslant u)$.

As in the previous chapters, we shall focus on multiplicative functions f with $f(p) \sim \kappa$ on average. We then expect that $f(n)$ is roughly $\kappa^{\omega(n)}$ on average, so that the value distribution of f is reduced to that of ω.

The Kubilius model

To study ω, we note that $\omega(n) = \sum_{p|n} 1 = \sum_p 1_{p|n}$. We then define the key random variables $B_d(n) := 1_{d|n}$ for $d \in \mathbb{N}$, so that

$$\omega = \sum_{p\leqslant x} B_p \tag{15.1}$$

on the probability space $\mathbb{N}_{\leqslant x}$. The functions B_d are Bernoulli random variables. Since there are exactly $\lfloor x/d \rfloor$ multiples of d up to x, we have

$$\mathbb{P}_{n\leqslant x}(B_d(n)=1)=\frac{\lfloor x/d\rfloor}{\lfloor x\rfloor}=\frac{x/d+O(1)}{x+O(1)}=1/d+O(1/x).$$

When d is fixed and $x\to\infty$, the above expression tends to $1/d$. In addition, if $p_1,\dots,p_m$ are distinct primes, then

$$\begin{aligned}\mathbb{P}_{n\leqslant x}(B_{p_1}(n)=\cdots=B_{p_m}(n)=1)&=\mathbb{P}_{n\leqslant x}(B_{p_1\cdots p_m}(n)=1)\\&=\frac{1}{p_1\cdots p_m}+O(1/x).\end{aligned}\tag{15.2}$$

Therefore, if we let $x\to\infty$, we find that

$$\mathbb{P}_{n\leqslant x}(B_{p_1}(n)=\cdots=B_{p_m}(n)=1)\sim\prod_{j=1}^{m}\mathbb{P}_{n\leqslant x}(B_{p_j}(n)=1).$$

We are thus led to the conclusion that the random variables B_p for p prime are *approximately independent* from each other.

The above analysis and relation (15.1) imply that ω is the sum of quasi-independent random variables. It is thus tempting to use tools from probability theory to study its value distribution. There is an obvious problem with this approach: most of the standard probabilistic tools apply to truly independent random variables.

To circumvent this problem, we introduce new Bernoulli random variables K_p (living in some ambient probability space) that are completely independent from each other, and for which we have the exact equality $\mathbb{P}(K_p=1)=1/p$. The random variables K_p are idealized models of B_p. Collectively, they form the *Kubilius model of the integers*.

Let us consider now the sum $S=\sum_{p\leqslant x}K_p$, whose distribution models the function ω on the space $\mathbb{N}_{\leqslant x}$. Since S is the sum of independent random variables, we may apply to it well-established probabilistic tools. We can then hope to transfer the results on S to the deterministic setting of ω. However, it should be noted that the Kubilius model has its limits, as Remark 2.2 reveals. We will return to this important point in Chapter 18 and discuss it in more detail.

We conclude our introductory discussion of the Kubilius model showing that it is possible to construct the random variables K_p in a very natural and concrete way. We take as our probability space the set of y-smooth integers $\mathcal{S}(y)$, which we employ with the probability measure

$$\mathbb{P}_{\mathcal{S}(y)}(\mathcal{A}):=\prod_{p\leqslant y}\Big(1-\frac{1}{p}\Big)\sum_{n\in\mathcal{A}}\frac{1}{n}.$$

Notice that the random variables $(B_p)_{p\leqslant y}$ we saw before become completely independent of one another in this new probability space: if $p_1 < \cdots < p_k \leqslant y$ and $a_1, \dots, a_k \in \mathbb{Z}_{\geqslant 1}$, then

$$\mathbb{E}_{\mathcal{S}(y)}[B_{p_1}^{a_1} \cdots B_{p_k}^{a_k}] = \prod_{p\leqslant y} \Big(1 - \frac{1}{p}\Big) \sum_{\substack{n\in\mathcal{S}(y)\\ p_1,\dots,p_k \mid n}} \frac{1}{n} = \frac{1}{p_1 \cdots p_k}.$$

A Central Limit Theorem for ω

We now use the Kubilius model to study the value distribution of ω. Similarly to (15.1), we have that $\omega = \sum_{p\leqslant y} B_p$ on the space $\mathcal{S}(y)$, that is to say, ω is a sum of independent random variables. Its mean value is

$$\mathbb{E}_{\mathcal{S}(y)}[\omega] = \sum_{p\leqslant y} \frac{1}{p} = \log\log x + O(1)$$

by Mertens's second estimate (Theorem 3.4(b)). Similarly, its variance equals

$$\mathbb{V}_{\mathcal{S}(y)}[\omega] = \sum_{p\leqslant y} \mathbb{V}_{\mathcal{S}(y)}[B_p] = \sum_{p\leqslant y} \left(\frac{1}{p} - \frac{1}{p^2}\right) = \log\log x + O(1),$$

where we used the independence of the variables B_p. Since the B_p's are uniformly bounded and the variance of ω tends to infinity, Lindeberg's Central Limit Theorem (see Theorem 27.2 in [7] or Theorem 2.1.5 in [168]) implies that, for any fixed $\alpha \leqslant \beta$, we have

$$\lim_{y\to\infty} \mathbb{P}_{n\in\mathcal{S}(y)}\Big(\alpha \leqslant \frac{\omega(n) - \log\log y}{\sqrt{\log\log y}} \leqslant \beta\Big) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-t^2/2}\mathrm{d}t. \tag{15.3}$$

The random variables B_p are approximately independent with respect to the probability measure $\mathbb{P}_{n\leqslant x}$ too. Thus, we might expect a similar result to hold for this measure. This was indeed proved by Erdős and Kac in 1940.

Theorem 15.1 (Erdős-Kac)**.** *For each fixed $\alpha \leqslant \beta$, we have that*

$$\lim_{x\to\infty} \mathbb{P}_{n\leqslant x}\Big(\alpha \leqslant \frac{\omega(n) - \log\log x}{\sqrt{\log\log x}} \leqslant \beta\Big) = \frac{1}{\sqrt{2\pi}} \int_\alpha^\beta e^{-t^2/2}\mathrm{d}t.$$

We will use the so-called *method of moments* to prove Theorem 15.1. The key result is Theorem 15.2, whose proof is given in Appendix C. We write $\mathcal{N}(0,1)$ for the standard normal distribution.

Theorem 15.2. *Let $(X_j)_{j=1}^\infty$ be a sequence of real-valued random variables.*
(a) *Assume that*

$$\lim_{j\to\infty} \mathbb{E}[X_j^k] = \mathbb{E}[\mathcal{N}(0,1)^k] \quad \textit{for all } k \in \mathbb{Z}_{\geqslant 1}. \tag{15.4}$$

Then $(X_j)_{j=1}^\infty$ converges in distribution to $\mathcal{N}(0,1)$.

(b) *Conversely, assume that* $(X_j)_{j=1}^\infty$ *converges in distribution to* $\mathcal{N}(0,1)$. *If, in addition,* $\sup_{j\geqslant 1}\mathbb{E}[X_j^{2k}]<\infty$ *for all* $k\in\mathbb{N}$, *then* (15.4) *holds.*

We already know that $(\omega-\log\log y)/\sqrt{\log\log y}$ converges in distribution to $\mathcal{N}(0,1)$ with respect to the measure $\mathbb{P}_{\mathcal{S}(y)}$ when $y\to\infty$. We now check that the second hypothesis of Theorem 15.2(b) holds for it.

Lemma 15.3. *Uniformly for* $y\geqslant 2$ *and* $k\in\mathbb{Z}_{\geqslant 1}$, *we have*

$$\mathbb{E}_{n\in\mathcal{S}(y)}\Big[\Big|\frac{\omega(n)-\log\log y}{\sqrt{\log\log y}}\Big|^k\Big]\ll k!.$$

Proof. Note that $|t|^k\leqslant k!e^{|t|}\leqslant k!(e^t+e^{-t})$ for any $t\in\mathbb{R}$. It thus suffices to prove that

$$\mathbb{E}_{n\in\mathcal{S}(y)}[e^{\alpha\omega(n)/\sqrt{\log\log y}}]\ll e^{\alpha\sqrt{\log\log y}}$$

for $\alpha=\pm 1$. The independence of the Bernoulli random variables B_p implies that

$$\mathbb{E}_{\mathcal{S}(y)}[z^\omega]=\prod_{p\leqslant y}\mathbb{E}_{\mathcal{S}(y)}[z^{B_p}]=\prod_{p\leqslant y}\Big(1+\frac{z-1}{p}\Big)\leqslant\exp\Big\{\sum_{p\leqslant y}\frac{z-1}{p}\Big\}$$

for all $z>0$. We take $z=e^{\alpha/\sqrt{\log\log y}}$, for which we have

$$z-1=e^{\alpha/\sqrt{\log\log y}}-1=\alpha/\sqrt{\log\log y}+O(1/\log\log y).$$

Hence, Mertens' second estimate (Theorem 3.4(b)) completes the proof. □

We are now ready to prove the Erdős-Kac theorem.

Proof of Theorem 15.1. We follow an idea due to Billingsley [7, Section 30]. Throughout, we let $\lambda_x=\log\log x$.

By Theorem 15.2(a), it suffices to prove that the moments $\mathbb{E}_{n\leqslant x}[(\omega(n)-\lambda_x)^k/\lambda_x^{k/2}]$ converge to $\mathbb{E}[\mathcal{N}(0,1)^k]$ when $x\to\infty$. However, we already know that $\mathbb{E}_{n\in\mathcal{S}(y)}[(\omega(n)-\lambda_y)^k\lambda_y^{-k/2}]$ tends to $\mathbb{E}[N(0,1)^k]$ when $y\to\infty$ by Theorem 15.2(b), which is applicable in view of relation (15.3) and Lemma 15.3. Consequently, it suffices to show that

$$\mathbb{E}_{n\leqslant x}\Big[\Big(\frac{\omega(n)-\lambda_x}{\sqrt{\lambda_x}}\Big)^k\Big]=\mathbb{E}_{n\in\mathcal{S}(y)}\Big[\Big(\frac{\omega(n)-\lambda_y}{\sqrt{\lambda_y}}\Big)^k\Big]+o_{x\to\infty}(1)\tag{15.5}$$

for each fixed $k\in\mathbb{Z}_{\geqslant 1}$, where $y=y(x)$ is an appropriate function of x going to infinity.

An integer $n\leqslant x$ has $\leqslant\log x/\log y$ prime factors $>y$ (see Exercise 2.9(e)). So, if we let $y=x^{1/\log\log\log x}$ and $\omega(n;y):=\#\{p|n:p\leqslant y\}$, then

$$\omega(n)=\omega(n;y)+O(\log\log\log x)\qquad(n\leqslant x).$$

Since we also have that $|\lambda_y - \lambda_x| = o_{x\to\infty}(\sqrt{\lambda_x})$, relation (15.5) is reduced to showing that

$$\mathbb{E}_{n\leqslant x}\big[(\omega(n;y)-\lambda_x)^k\big] = \mathbb{E}_{n\in\mathcal{S}(y)}\big[(\omega(n)-\lambda_x)^k\big] + o_{x\to\infty}(\lambda_x^{k/2}). \tag{15.6}$$

Note that

$$\begin{aligned}\mathbb{E}_{n\leqslant x}\big[(\omega(n;y)-\lambda_x)^k\big] &- \mathbb{E}_{n\in\mathcal{S}(y)}\big[(\omega(n;y)-\lambda_x)^k\big]\\ &= \sum_{j=0}^{k}\binom{k}{j}(-\lambda_x)^{k-j}\big(\mathbb{E}_{n\leqslant x}[\omega(n;y)^j] - \mathbb{E}_{n\in\mathcal{S}(y)}[\omega(n)^j]\big).\end{aligned}$$

Hence, it suffices to show that, for each fixed $j\in\mathbb{Z}\cap[0,k]$, we have

$$\mathbb{E}_{n\leqslant x}[\omega(n;y)^j] - \mathbb{E}_{n\in\mathcal{S}(y)}[\omega(n)^j] = o_{x\to\infty}(\lambda_x^{j-k/2}).$$

We begin by noticing that $\omega(\cdot;y)^j = (\sum_{p\leqslant y} B_p)^j = \sum_{p_1,\dots,p_j\leqslant y} B_{p_1}\cdots B_{p_j}$. Taking expectations implies that

$$\mathbb{E}_{n\leqslant x}[\omega(n;y)^j] = \sum_{p_1,\dots,p_j\leqslant y}\mathbb{P}_{n\leqslant x}(\cap_{i=1}^{j}\{B_{p_i}(n)=1\}).$$

We then apply a variant of (15.2) to conclude that

$$\begin{aligned}\mathbb{E}_{n\leqslant x}[\omega(n;y)^j] &= \sum_{p_1,\dots,p_j\leqslant y}\frac{\lfloor x/[p_1,\dots,p_j]\rfloor}{\lfloor x\rfloor}\\ &= \sum_{p_1,\dots,p_j\leqslant y}\frac{1}{[p_1,\dots,p_j]} + O(\pi(y)^j/x),\end{aligned}$$

where $[p_1,\dots,p_j]$ denotes the least common multiple of the primes p_1, …, p_j. A similar calculation implies that

$$\mathbb{E}_{n\in\mathcal{S}(y)}[\omega(n)^j] = \sum_{p_1,\dots,p_j\leqslant y}\frac{1}{[p_1,\dots,p_j]},$$

whence $\mathbb{E}_{n\leqslant x}[\omega(n;y)^j] - \mathbb{E}_{n\in\mathcal{S}(y)}[\omega(n)^j] \ll \pi(y)^j/x = o_{x\to\infty}(\lambda_x^{j-k/2})$ by the choice of y. This completes the proof of Theorem 15.1. □

Dissecting sums of multiplicative functions

Now that we have a good understanding of the distribution of ω, we use it to analyze the finer structure of $\sum_{n\leqslant x}\kappa^{\omega(n)}$. Our goal is to determine which values of ω give the dominant contribution to this sum. More concretely, we want to identify those sets $\mathcal{I}(x)\subset\mathbb{R}_{\geqslant 0}$ with the property

$$\sum_{n\leqslant x,\,\omega(n)\in\mathcal{I}(x)}\kappa^{\omega(n)} \sim \sum_{n\leqslant x}\kappa^{\omega(n)} \quad (x\to\infty). \tag{15.7}$$

Of course, we could take $\mathcal{I}(x) = \mathbb{R}_{\geqslant 0}$, but this is not so insightful. We want to find $\mathcal{I}(x)$ that is as small as possible, while still satisfying (15.7).

Since ω has mean value $\log\log x$ and standard deviation $\sqrt{\log\log x}$ over the probability space $\mathbb{N}_{\leqslant x}$ a natural guess is the set

$$\mathcal{I}_1(x) = [\log\log x - \xi(x)\sqrt{\log\log x}, \log\log x + \xi(x)\sqrt{\log\log x}],$$

where $\xi(x)$ is a function tending to infinity slowly. However, note that $\kappa^{\omega(n)} = (\log x)^{\log\kappa + o(1)}$ on the set $\mathcal{A}_1(x) := \{\, n \leqslant x : \omega(n) \in \mathcal{I}_1(x) \,\}$, whence

$$\sum_{n\in\mathcal{A}_1(x)} \kappa^{\omega(n)} = (\log x)^{\log\kappa+o(1)} \cdot |\mathcal{A}_1(x)| = x(\log x)^{\log\kappa+o(1)} \tag{15.8}$$

as $x \to \infty$. If $\kappa \neq 1$, we find that $\log\kappa < \kappa - 1$. Hence, Theorem 13.2 implies that the contribution of integers $n \in \mathcal{A}_1(x)$ to $\sum_{n\leqslant x} \kappa^{\omega(n)}$ is negligible.

From the above discussion, we conclude that the sum $\sum_{n\leqslant x} \kappa^{\omega(n)}$ with $\kappa \neq 1$ is dominated by integers $n \leqslant x$ with "atypical" values of $\omega(n)$ with respect to the measure $\mathbb{P}_{n\leqslant x}$. As a matter of fact, for the purpose of identifying $\mathcal{I}(x)$, it is more natural to switch to the weighted probability measure

$$\mathbb{P}^{\kappa}_{n\leqslant x}(\mathcal{A}) := \frac{\sum_{n\in\mathcal{A}} \kappa^{\omega(n)}}{\sum_{n\leqslant x} \kappa^{\omega(n)}}. \tag{15.9}$$

We write $\mathbb{E}^{\kappa}_{n\leqslant x}[Z]$ for the expectation of the random variable $Z : \mathbb{N}_{n\leqslant x} \to \mathbb{R}$ with respect to this measure. Finding a set satisfying (15.7) then amounts to understanding the distribution of ω with respect to the measure $\mathbb{P}^{\kappa}_{n\leqslant x}$.

It turns out that ω is approximately Gaussian with respect to $\mathbb{P}^{\kappa}_{n\leqslant x}$ as well, but with expectation and variance $\sim \kappa\log\log x$. We may then take

$$\mathcal{I}_\kappa(x) = [\kappa\log\log x - \xi(x)\sqrt{\log\log x}, \kappa\log\log x + \xi(x)\sqrt{\log\log x}], \tag{15.10}$$

where $\xi(x) \to \infty$. The details of this argument are outlined in Exercise 15.2.

Exercises

Exercise 15.1. Deduce that Theorem 15.1 holds with Ω in place of ω too. [*Hint:* Use Exercise 3.9.]

Exercise 15.2. Fix $\kappa > 0$, and let $\mathbb{P}^{\kappa}_{n\leqslant x}$ be defined by (15.9). Define also $\mathbb{P}^{\kappa}_{\mathcal{S}(y)}(\mathcal{A}) = \prod_{p\leqslant y}(1+\kappa/(p-1))^{-1}\sum_{n\in\mathcal{A}}\kappa^{\omega(n)}/n$.

(a) For $d \in \mathcal{S}(y)$, prove that $\mathbb{P}^{\kappa}_{n\in\mathcal{S}(y)}(B_d(n) = 1) = \kappa^{\omega(d)}d^{-1}/\prod_{p|d}(1+\frac{\kappa-1}{p})$.

(b) If $\omega(d) \leqslant k$ for some fixed $k \in \mathbb{N}$ and $x \geqslant d^2$, then prove that

$$\mathbb{P}^{\kappa}_{n\leqslant x}(B_d(n) = 1) = \mathbb{P}^{\kappa}_{n\in\mathcal{S}(y)}(B_d(n) = 1) + O_{k,\kappa}(1/(d\log x)).$$

[*Hint:* Use Theorem 13.2 on the function $f_d(m) := \kappa^{\omega(dm)-\omega(d)}$.]

(c) For each fixed $\alpha \leqslant \beta$, prove that

$$\lim_{x\to\infty} \mathbb{P}^{\kappa}_{n\leqslant x}\Big(\alpha \leqslant \frac{\omega(n)-\kappa\log\log x}{\sqrt{\kappa\log\log x}} \leqslant \beta\Big) = \frac{1}{\sqrt{2\pi}}\int_{\alpha}^{\beta} e^{-t^2/2}\mathrm{d}t.$$

In particular, (15.7) holds for the set $\mathcal{I}_\kappa(x)$ defined by (15.10).

Exercise 15.3. Let $f : \mathbb{N} \to \mathbb{R}_{\geqslant 0}$ be multiplicative. Determine necessary conditions under which ω satisfies an analogue of the Erdős-Kac theorem with respect to the measure $\mathbb{P}^f_{n\leqslant x}(\mathcal{A}) = \sum_{n\in\mathcal{A}} f(n)/\sum_{n\leqslant x} f(n)$.

Exercise 15.4. (a) (Landau [126]) For each fixed $k \in \mathbb{N}$, prove that

$$\mathbb{P}_{n\leqslant x}(\omega(n)=k) \sim \frac{1}{\log x}\cdot\frac{(\log\log x)^{k-1}}{(k-1)!} \qquad (x\to\infty).$$

(b) (Hardy-Ramanujan [95]) Show there are constants A and B such that

$$\mathbb{P}_{n\leqslant x}(\omega(n)=k) \leqslant \frac{A}{\log x}\cdot\frac{(\log\log x+B)^{k-1}}{(k-1)!},$$

uniformly for $x \geqslant 2$ and $k \in \mathbb{N}$.

[*Hint:* When $\omega(n) = k+1$, prove that there are at least k ways to write $n = p^a m$ with $p \leqslant n^{1/(a+1)}$ and $\omega(m) = k$. Then, induct on k.]

Exercise 15.5. Let $q \in \mathbb{N}$ and $\mathbb{P}_q$ denote the uniform counting measure on the set of Dirichlet characters mod q (i.e., $\mathbb{P}_q(\mathcal{X}) := |\mathcal{X}|/\varphi(q)$ for any set $\mathcal{X}$ of Dirichlet characters mod q).

Let $Z(x;\chi) = \pi(x)^{-1/2}\sum_{p\leqslant x}\chi(p)$. If $x, q \to \infty$ at a rate such that $x = q^{o(1)}$, and χ is sampled with respect to the measure $\mathbb{P}_q$, then prove that the random variables $\chi \to \sqrt{2}\,\mathrm{Re}(Z(x;\chi))$ and $\chi \to \sqrt{2}\,\mathrm{Im}(Z(x;\chi))$ both converge in distribution to the standard normal distribution.[1] [*Hint:* Model $\chi(p)$ by $e^{i\theta_p}$, where θ_2, θ_3, θ_5, ... is a sequence of independent random variables that are uniformly distributed on $[0, 2\pi]$.]

[1] In fact, $\sqrt{2}\cdot(\mathrm{Re}(Z(x;\chi)), \mathrm{Im}(Z(x;\chi)))$ converges to a 2-dimensional Gaussian: when $q \to \infty$ and $x = q^{o(1)}$, the quantity $\mathbb{P}_q(\mathrm{Re}(Z(x;\chi)) \leqslant u/\sqrt{2},\ \mathrm{Im}(Z(x;\chi)) \leqslant v/\sqrt{2})$ tends to $\mathbb{P}(N(0,1)\leqslant u)\cdot\mathbb{P}(N(0,1)\leqslant v)$. Proving this result requires a 2-dimensional analogue of Theorem 15.2 and is an excellent exercise on the method of moments.

Chapter 16

Large deviations

Let X be a real-valued random variable with expectation $\mu = \mathbb{E}[X]$. In many cases, most of the mass of X is concentrated around μ. To measure this concentration, we seek to estimate the rate of decay to 0 of the probabilities $\mathbb{P}(X > \mu + u)$ and $\mathbb{P}(X < \mu - u)$, that is to say, how "heavy" are the right and left tails of the distribution of X.

If X is exponential, i.e. it has density $1_{t \geqslant 0} \cdot e^{-t}$, then $\mu = 1$, $\mathbb{P}(X > 1 + u) = e^{-u-1}$ and $\mathbb{P}(X < 1 - u) = 0$ for $u \geqslant 1$. On the other hand, the tails of $\mathcal{N}(0,1)$ are of size $\approx \exp(-u^2/2)$, which is much smaller than e^{-u-1}. Hence, a Gaussian is more concentrated than an exponential distribution.

As another example, consider the distribution of ω with respect to the measure $\mathbb{P}^{\kappa}_{n \leqslant x}$ defined in (15.9). Understanding wherein lies the mass of this distribution is essentially equivalent to finding a set $\mathcal{I}(x)$ satisfying (15.7).

The study of tails of distributions is the subject matter of *Cramér's theory of large deviations* ([168, Section 1.3], [7, Chapter 9]). For simplicity, let us assume that X is normalized so that $\mu = 0$. We focus on understanding the frequency of occurrence of the event $\{X > u\}$. Since $\{X < -u\} = \{-X > u\}$, this treats left tails as well upon replacing X by $-X$.

The main tool in the study of $\{X > u\}$ is the Laplace transform of X,

$$\mathcal{L}_X(s) := \mathbb{E}[e^{sX}],$$

which is typically defined in some vertical strip $c_1 < \mathrm{Re}(s) < c_2$. The simplest way of using $\mathcal{L}_X(s)$ to estimate $\mathbb{P}(X > u)$ is via Chernoff's inequality (which is a simple consequence of Markov's inequality): for any $\sigma \in I := (c_1, c_2) \cap (0, +\infty)$, we have

$$\mathbb{P}(X > u) = \mathbb{P}(e^{\sigma(X-u)} > 1) \leqslant \mathbb{E}[e^{\sigma(X-u)}] = e^{-\sigma u} \cdot \mathcal{L}_X(\sigma). \tag{16.1}$$

If the function $e^{-\sigma u}\cdot\mathcal{L}_X(\sigma)$ is minimized at $\alpha=\alpha(u)\in I$, then its derivative must vanish at $\sigma=\alpha$, whence

$$\mathcal{L}'_X(\alpha)=u\mathcal{L}_X(\alpha). \tag{16.2}$$

We assume for simplicity that the equation $\mathcal{L}'_X(\sigma)=u\mathcal{L}_X(\sigma)$ has a unique solution in I, so that α is determined by (16.2).

The above method can often yield lower bounds on $\mathbb{P}(X>u)$ as well: in many cases it turns out that for the optimal choice of $\sigma=\alpha$ the integral $\mathcal{L}_X(\sigma)=\int e^{\sigma X}\mathrm{d}\mathbb{P}$ is dominated by values of $X\approx u$, that is to say,

$$\mathcal{L}_X(\alpha)\approx\int_{X\approx u}e^{\alpha X}\mathrm{d}\mathbb{P}\approx e^{\alpha u}\mathbb{P}(X\approx u).$$

Hence, we also have the rough lower bound $\mathbb{P}(X>u)\gtrapprox e^{-\alpha u}\mathcal{L}_X(\alpha)$.

A more sophisticated approach is to use the inverse Laplace transform: for $c\in I$, Perron's inversion formula implies that

$$\mathbb{P}(X>u)=\mathbb{E}\bigg[\frac{1}{2\pi i}\int_{(c)}\frac{e^{s(X-u)}}{s}\mathrm{d}s\bigg]=\frac{1}{2\pi i}\int_{(c)}\mathcal{L}_X(s)e^{-su}\frac{\mathrm{d}s}{s}, \tag{16.3}$$

provided, of course, that we can justify an application of Fubini's theorem. The choice of c is crucial here, and we take $c=\alpha$. This is because if we write $\mathcal{L}_X(s)e^{-su}=e^{f(s)}$, then $f'(\alpha)=0$, so that the integrand has a *stationary point* at $s=\alpha$. Under some mild conditions, the integral in (16.3) is then dominated by values of $s\approx\alpha$. We may thus obtain an asymptotic evaluation for $\mathbb{P}(X>u)$ using Taylor's theorem for the integrand, much like we did in the proof of formula (1.20).

We present below three applications of this circle of ideas.

Dissecting sums of multiplicative functions: Encore

First, we use the method of large deviations to obtain information on the multiplicative structure of a "typical" integer with respect to the probability measure $\mathbb{P}^{\kappa}_{n\leqslant x}$. Results of this kind fall under the subfield of number theory called *Anatomy of Integers*. To state our theorem, we let $p_1(n)<p_2(n)<\cdots<p_{\omega(n)}(n)$ denote the sequence of the distinct prime factors of n in increasing order.

Theorem 16.1. *Fix $\kappa,\varepsilon>0$ and a function $\xi:\mathbb{R}_{>0}\to\mathbb{R}_{>0}$ tending to infinity. Let $\mathcal{A}(x)$ be the set of integers $n\leqslant x$ with*

$$\frac{j}{\kappa+\varepsilon}\leqslant\log\log p_j(n)\leqslant\frac{j}{\kappa-\varepsilon}\quad(\xi(x)\leqslant j\leqslant\omega(n))$$

and $\omega(n)/\log\log x\in[\kappa-\varepsilon,\kappa+\varepsilon]$. Then $\mathbb{P}^{\kappa}_{n\leqslant x}(\mathcal{A}(x))=1-o_{x\to\infty}(1)$.

Proof. All implied constants might depend on κ. Recall the notation

$$\omega(n;y) = \#\{\, p|n : p \leqslant y \,\} = \sum_{p\leqslant y} B_p(n).$$

Since $\omega(n; p_j(n)) = j$, the theorem will follow if we can show that, for each fixed $\varepsilon \in (0,\kappa)$ and each fixed function $\psi : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ with $\lim_{x\to\infty} \psi(x) = \infty$, the event

$$(\kappa-\varepsilon)\log\log y \leqslant \omega(n;y) \leqslant (\kappa+\varepsilon)\log\log y \qquad (\psi(x)\leqslant y\leqslant x) \tag{16.4}$$

occurs with probability $1 - o_{x\to\infty}(1)$ with respect to the measure $\mathbb{P}^\kappa_{n\leqslant x}$. Since $\mathbb{E}^\kappa_{n\leqslant x}[\omega(n;y)] \sim \kappa\log\log y$, this amounts to showing a simultaneous concentration-of-measure inequality for all of the random variables $\omega(\cdot;y)$.

To accomplish the above task, we use Chernoff's inequality (16.1) (also called *Rankin's trick* in this context): for all $u \geqslant \kappa$ and all $\sigma > 0$, we have

$$\mathbb{P}^\kappa_{n\leqslant x}\big(\omega(n;y) > u\log\log y\big) \leqslant (\log y)^{-\sigma u}\cdot \mathbb{E}^\kappa_{n\leqslant x}[e^{\sigma\omega(n;y)}]. \tag{16.5}$$

In addition, we have

$$\mathbb{E}^\kappa_{n\leqslant x}[e^{\sigma\omega(n;y)}] \asymp \frac{1}{x(\log x)^{\kappa-1}}\sum_{n\leqslant x}\kappa^{\omega(n)}e^{\sigma\omega(n;y)} \ll_\sigma (\log y)^{\kappa(e^\sigma-1)}, \tag{16.6}$$

where the first estimate follows from Theorem 13.2, and the second one from Theorem 14.2 (applied to the function $f(n) = \kappa^{\omega(n)}e^{\sigma\omega(n;y)}$, for which $f(p) = \kappa + \kappa(e^\sigma - 1)1_{p\leqslant y}$) and from Mertens' second estimate (Theorem 3.4(b)). We insert (16.6) into (16.5) and optimize the resulting upper bound by taking $\sigma = \log(u/\kappa)$. This yields the inequality

$$\mathbb{P}^\kappa_{n\leqslant x}\big(\omega(n;y) > u\log\log y\big) \ll (\log y)^{-\kappa Q(u/\kappa)}$$

uniformly for $y \in [2,x]$ and $\kappa \leqslant u \leqslant 100\kappa$, where $Q(t) := t\log t - t + 1$.

A very similar argument also proves that

$$\mathbb{P}^\kappa_{n\leqslant x}\big(\omega(n;y) < u\log\log y\big) \ll (\log y)^{-\kappa Q(u/\kappa)} \quad (2\leqslant y\leqslant x,\ 0<u\leqslant\kappa).$$

We thus arrive at the concentration-of-measure inequality

$$\mathbb{P}^\kappa_{n\leqslant x}\big(\big|\omega(n;y) - \kappa\log\log y\big| > \varepsilon\log\log y\big) \ll_\kappa (\log y)^{-\delta(\kappa,\varepsilon)}, \tag{16.7}$$

where $\delta(\kappa,\varepsilon) = \kappa\min\{Q(1+\varepsilon/\kappa), Q(1-\varepsilon/\kappa)\} > 0$. This establishes the theorem for a fixed value of y.

We pass to a result for all values of y using a simple trick: we fix the check points $y_j = \min\{\psi(x)^{e^j}, x\}$, $j \geqslant 0$, and let J be such that $y_{J-1} < x = y_J$. Then, the union bound and relation (16.7) with $\varepsilon/2$ in place of ε imply that

$$\begin{aligned}\mathbb{P}^\kappa_{n\leqslant x}\Big(\bigcup_{j=0}^{J}\big\{\big|\omega(n;y_j) - \kappa\log\log y_j\big| > \frac{\varepsilon}{2}\log\log y_j\big\}\Big) &\ll_\kappa \sum_{j=0}^{J}(\log y_j)^{-\delta(\kappa,\varepsilon/2)}\\ &\ll_{\kappa,\varepsilon} (\log\psi(x))^{-\delta(\kappa,\varepsilon/2)}.\end{aligned}$$

Now, let $n \leqslant x$ be such that $|\omega(n; y_j) - \kappa \log\log y_j| \leqslant 0.5\varepsilon \log\log y_j$ whenever $0 \leqslant j \leqslant J$. We know that this holds with probability $1 - o_{x\to\infty}(1)$. In addition, if $y \in [\psi(x), x]$, then there is $j \in \{1, \dots, J\}$ such that $y_{j-1} \leqslant y \leqslant y_j$. Hence

$$\omega(n; y) \geqslant \omega(n; y_j) \geqslant (\kappa - \varepsilon/2) \log\log y_{j-1} \geqslant (\kappa - \varepsilon) \log\log y,$$

provided that x is large enough (so that $y \geqslant \psi(x)$ is also large enough). Starting with the inequality $\omega(n; y) \leqslant \omega(n; y_j)$ we may also prove that $\omega(n; y) \leqslant (\kappa + \varepsilon) \log\log y$. This completes the proof of the theorem. □

The saddle-point method: Encore

In Theorem 15.1, we let the value of ω vary in certain wide intervals. Now, we study the proportion of integers $n \leqslant x$ for which $\omega(n)$ takes a given value k. This is a rare event, so we will study it using the method of large deviations and the theory of the inverse Laplace transform.

Theorem 16.2. *Fix $C > 0$. For $x \geqslant 1$ and $k \in \mathbb{Z} \cap [1, C \log\log x]$, we have*

$$\mathbb{P}_{n\leqslant x}(\omega(n) = k) = \frac{G(\alpha)}{\Gamma(\alpha+1)} \cdot \frac{(\log\log x)^{k-1}}{(k-1)! \log x} \big(1 + O_C\big(k/(\log\log x)^2\big)\big),$$

where $\alpha = (k-1)/\log\log x$ and

$$G(z) := \prod_p \left(1 + \frac{z}{p-1}\right)\left(1 - \frac{1}{p}\right)^z = \prod_p \left(1 + \frac{z-1}{p}\right)\left(1 - \frac{1}{p}\right)^{z-1}.$$

Proof. We may assume that $k \geqslant 2$, with the case $k = 1$ following from the Prime Number Theorem.

Since ω takes values in $\mathbb{Z}_{\geqslant 0}$, it is easy to invert the Laplace transform here: using Cauchy's residue theorem, we readily find that

$$\mathbb{P}_{n\leqslant x}(\omega(n) = k) = \mathbb{E}_{n\leqslant x}\Big[\frac{1}{2\pi i}\oint_{|z|=r} \frac{z^{\omega(n)}\mathrm{d}z}{z^{k+1}}\Big] = \frac{1}{2\pi i}\oint_{|z|=r} \mathbb{E}_{n\leqslant x}[z^{\omega(n)}]\frac{\mathrm{d}z}{z^{k+1}}$$

for any $r > 0$. We estimate the integrand using Theorem 13.2: we have

$$\mathbb{E}_{n\leqslant x}[z^{\omega(n)}] = H(z) z (\log x)^{z-1} + O_r((\log x)^{\mathrm{Re}(z)-2}) \tag{16.8}$$

uniformly for $|z| = r$, where $H(z) := G(z)/\Gamma(z+1)$. As a consequence,

$$\mathbb{P}_{n\leqslant x}(\omega(n) = k) = \frac{1}{2\pi i}\oint_{|z|=r} \frac{H(z)(\log x)^{z-1} + O_r((\log x)^{z-2})}{z^k}\mathrm{d}z. \tag{16.9}$$

The function $H(z)$ is entire and has bounded derivatives. Hence, it will not affect the order of magnitude of the integral. If it were not present, nor did we have an error term, we could use Cauchy's theorem to find that

$$\frac{1}{2\pi i}\oint_{|z|=r} \frac{(\log x)^{z-1}}{z^k}\mathrm{d}z = \frac{1}{\log x} \cdot \frac{(\log\log x)^{k-1}}{(k-1)!}. \tag{16.10}$$

The main idea is to choose r in a way that the mass of the integrals in (16.9) and in (16.10) is concentrated around the point $z = r$, because then we can replace $H(z)$ by $H(r)$ in (16.9) at the cost of a small error, and then apply (16.10).

To carry out the above strategy, we use the saddle-point method: we pick $r = \alpha = (k-1)/\log\log x$, so that, if we write $(\log x)^{z-1}/z^{k-1} = e^{\ell(z)}/z$, then $\ell'(\alpha) = 0$. (We have left a z in the denominator because $\mathrm{d}z/z$ is the natural invariant measure on the circle $|z| = r$.) For this choice of r, it can be seen using quadratic approximation that most of the mass of the integral in (16.10) is on the arc $|z - \alpha| \ll \alpha/\sqrt{k}$.

Now, we write $\mathbb{P}_{n\leqslant x}(\omega(n) = k) = I_1 + I_2$, where

$$I_1 := \frac{1}{2\pi i}\oint_{|z|=\alpha} \frac{H(\alpha)(\log x)^{z-1}}{z^k}\mathrm{d}z = \frac{H(\alpha)}{\log x}\cdot\frac{(\log\log x)^{k-1}}{(k-1)!}$$

by Cauchy's theorem, and

$$\begin{aligned} I_2 &:= \frac{1}{2\pi i}\oint_{|z|=r} \frac{(H(z) - H(\alpha) + O_C(1/\log x))(\log x)^{z-1}}{z^k}\mathrm{d}z \\ &= \frac{1}{2\pi i}\oint_{|z|=\alpha} \frac{[(z-\alpha)H'(\alpha) + O_C(|z-\alpha|^2 + 1/\log x)](\log x)^{z-1}}{z^k}\mathrm{d}z \end{aligned}$$

by Taylor's theorem. The integral of $(z-\alpha)(\log x)^{z-1}/z^k$ vanishes by Cauchy's theorem and the choice of α. We bound the remaining part of I_2 by making the change of variables $z = \alpha e^{i\theta}$ with $\theta \in [-\pi, \pi]$. In conclusion, we have arrived at the estimate

$$I_2 \ll_C \int_{-\pi}^{\pi} \frac{(\log x)^{\alpha\cos\theta - 1}(\alpha^2|e^{i\theta} - 1|^2 + 1/\log x)}{\alpha^{k-1}}\mathrm{d}\theta.$$

For all $\theta \in [-\pi, \pi]$, we have $|e^{i\theta} - 1| \asymp |\theta|$, by Taylor's theorem. In addition,

$$(\log x)^{\alpha\cos\theta} \leqslant e^{k\cos\theta} \leqslant e^{k(1-\theta^2/2+\theta^4/24)} \leqslant e^{k(1-c\theta^2)},$$

where $c = (1 - \pi^2/12)/2 \approx 0.0887$. Therefore,

$$\begin{aligned} I_2 &\ll_C \frac{e^k}{\alpha^{k-1}\log x}\int_{-\pi}^{\pi}(\alpha^2\theta^2 + 1/\log x)e^{-ck\theta^2}\mathrm{d}\theta \\ &\ll_C \frac{e^k}{\alpha^{k-1}\log x}\cdot\left(\alpha^2 k^{-3/2} + k^{-1/2}/\log x\right) \\ &\ll_C \frac{\alpha^2}{k}\cdot\frac{1}{\log x}\cdot\frac{(\log\log x)^{k-1}}{(k-1)!} \end{aligned}$$

by Stirling's formula and the choice of α. Finally, we note that $H(\alpha) \asymp_C 1$ for $\alpha \in [0, C]$ to complete the proof. □

Smooth numbers

The theory of large deviations can be used to obtain strong quantitative bounds for the number of y-smooth numbers $\leqslant x$. Notice that

$$\mathbb{E}_{n\in\mathcal{S}(y)}[\log n] = \mathbb{E}_{n\in\mathcal{S}(y)}\Big[\sum_{d\mid n}\Lambda(d)\Big] = \sum_{d\in\mathcal{S}(y)}\frac{\Lambda(d)}{d} = \sum_{p\leqslant y}\frac{\log p}{p-1} \sim \log y$$

as $y\to\infty$. Thus, if n is y-smooth and $> x = y^u$, then $\log n$ is approximately u times larger than its expected size. This offers a heuristic explanation why the Dickman-de Bruijn function decays so fast.

Optimizing Chernoff's inequality (16.1) (often referred to as Rankin's trick in this context) leads us to the following general theorem.

Theorem 16.3. *Let f be a multiplicative function such that $0\leqslant f\leqslant\tau_k$ for some $k\in\mathbb{Z}_{\geqslant 1}$. Let $x\geqslant y\geqslant 3$ and $u=\log x/\log y$. If $y\geqslant(\log x)^{2+\delta}$ for some $\delta>0$, then*

$$\sum_{n\in\mathcal{S}(y),\, n>x}\frac{f(n)}{n} \leqslant \frac{e^{O_{k,\delta}(u)}}{(u\log(2u))^u}\cdot\exp\Big\{\sum_{p\leqslant y}\frac{f(p)}{p}\Big\}.$$

Proof. All implied constants might depend on δ and k. We may assume that u is large enough and that $\delta\leqslant 1/2$. Let $\varepsilon\in[1/\log y, 1/2-\delta/5]$ to be chosen later. We have

$$\sum_{n\in\mathcal{S}(y),\, n>x}\frac{f(n)}{n} \leqslant \sum_{n\in\mathcal{S}(y)}\frac{(n/x)^\varepsilon f(n)}{n} = x^{-\varepsilon}\prod_{p\leqslant y}\Big(1+\sum_{m=1}^{\infty}\frac{f(p^m)}{p^{m(1-\varepsilon)}}\Big).$$

To each factor, we apply the inequality $1+t\leqslant e^t$. Our assumptions that $0\leqslant f\leqslant\tau_\kappa$ and that $\varepsilon\leqslant 1/2-\delta/5$ imply that $\sum_{p,\,m\geqslant 2} f(p^m)/p^{m(1-\varepsilon)}\ll 1$. Consequently,

$$\sum_{n\in\mathcal{S}(y),\, n>x}\frac{f(n)}{n} \ll \exp\Big\{-\varepsilon\log x+\sum_{p\leqslant y}\frac{f(p)}{p^{1-\varepsilon}}\Big\}. \tag{16.11}$$

Next, we write $\varepsilon = w/\log y$, so that $1\leqslant w\leqslant(1/2-\delta/5)\log y$. For the primes $p\leqslant y^{1/w}$, we note that $p^{w/\log y}=1+O(w\log p/\log y)$. Thus

$$\sum_{p\leqslant y^{1/w}}\frac{f(p)}{p^{1-w/\log y}} = \sum_{p\leqslant y^{1/w}}\frac{f(p)}{p}+O(1) \leqslant \sum_{p\leqslant y}\frac{f(p)}{p}+O(1),$$

by Mertens's estimates (Theorem 3.4). For the bigger primes, we use Chebyshev's estimate (Theorem 2.4) and partial summation to find that

$$\sum_{y^{1/w}<p\leqslant y}\frac{f(p)}{p^{1-w/\log y}} \leqslant \sum_{y^{1/w}<p\leqslant y}\frac{k}{p^{1-w/\log y}} \ll 1+\frac{e^w}{\log y}+\int_{y^{1/w}}^{y}\frac{t^{w/\log y}}{t\log t}\mathrm{d}t.$$

Making the change of variables $t^{w/\log y} = e^z$, we deduce that the right-hand side is $\ll e^w/w$.

Putting the above estimates together, we conclude that

$$\sum_{n\in\mathcal{S}(y),\, n>x} \frac{f(n)}{n} \ll \exp\Big\{ -uw + O(e^w/w) + \sum_{p\leqslant y} \frac{f(p)}{p} \Big\}.$$

We choose $w \geqslant 1$ implicitly via the formula $e^{w-1}/w = u$. Taking logarithms, we find that $w - 1 - \log w = \log u$. In particular, $w \asymp \log u$, whence $\log w = \log\log u + O(1)$. We appeal again to the identity $w - 1 - \log w = \log u$ to find that $w = \log(u\log u) + O(1)$. If we can show that $w \leqslant (1/2 - \delta/5)\log y$ for this choice of w, the theorem will follow. This inequality is true if

$$\begin{aligned} \frac{e^{(1/2-\delta/5)\log y - 1}}{(1/2-\delta/5)\log y} \geqslant u \quad &\iff \quad y^{1/2-\delta/5} \geqslant e(1/2-\delta/5)\log x \\ &\iff \quad y \geqslant c\cdot(\log x)^{2/(1-2\delta/5)}, \end{aligned}$$

where $c = (e(1/2-\delta/5))^{2/(1-2\delta/5)}$. The last inequality is indeed satisfied by our assumption that $y \geqslant (\log x)^{2+\delta}$, thus concluding the proof. □

Using the method of proof of Theorem 14.2, we can deduce from Theorem 16.3 an analogous result for the arithmetic mean of a multiplicative function over y-smooth integers. Taking $f = 1$ we find a uniform bound for the function $\Psi(x,y)$ that we encountered in Chapter 14.

Theorem 16.4. *Assume the set-up of Theorem* 16.3. *Then*

$$\sum_{n\in\mathcal{S}(y)\cap[1,x]} f(n) \leqslant \frac{e^{O_{k,\delta}(u)}}{(u\log(2u))^u}\cdot x\cdot \exp\Big\{\sum_{p\leqslant y}\frac{f(p)-1}{p}\Big\}.$$

Proof. Without loss of generality, $\delta \leqslant 1/2$. In addition, in virtue of Theorem 14.2, we may assume that u is large. Lastly, the condition $y \geqslant (\log x)^{2+\delta}$ allows us to also assume that x and y are large. As in the proof of Theorem 16.3, we consider a parameter $\varepsilon \in [1/\log y, 1/2 - \delta/5]$.

We use a variation of the proof of Theorem 14.2 that links mean values of multiplicative functions to logarithmic mean values. Mimicking the argument there, we start by writing

$$(\log x) \sum_{n\in\mathcal{S}(y)\cap[1,x]} f(n) = S_1 + S_2, \tag{16.12}$$

where

$$S_1 = \sum_{n\in\mathcal{S}(y)\cap[1,x]} f(n)\log(x/n) \qquad \text{and} \qquad S_2 = \sum_{n\in\mathcal{S}(y)\cap[1,x]} f(n)\log n.$$

For S_1, we note that $\log(x/n) \leqslant (x/n)^{1-\varepsilon}/(1-\varepsilon) \leqslant 2(x/n)^{1-\varepsilon}$ and thus

$$S_1 \leqslant 2x^{1-\varepsilon} \sum_{n\in\mathcal{S}(y)} \frac{f(n)}{n^{1-\varepsilon}}. \tag{16.13}$$

Next, we bound S_2. We do not have control on Λ_f because we have not assumed that $\tau_f = f$. Instead, we note that

$$f(n)\log n = f(n)\sum_{p^a\|n} \log(p^a) = \sum_{\substack{p^a m=n\\ p\nmid m}} f(m)f(p^a)\log(p^a),$$

whence

$$S_2 \leqslant \sum_{\substack{m\in\mathcal{S}(y),\, p\leqslant y,\, a\geqslant 1\\ p^a m\leqslant x}}\sum\sum af(p^a)f(m)\log p = \sum_{\substack{m\in\mathcal{S}(y),\, p\leqslant y\\ pm\leqslant x}}\sum f(m)\log p \sum_{\substack{a\geqslant 1\\ p^a m\leqslant x}} af(p^a).$$

For each fixed m and p, we have $1 \leqslant a \leqslant \log(x/m)/\log p$ and $f(p^a) \leqslant \tau_k(p^a) \ll_k a^{k-1} \leqslant (\log(x/m)/\log p)^{k-1}$. Hence

$$S_2 \ll_k \sum_{m\in\mathcal{S}(y)\cap[1,x/2]} f(m) \sum_{2\leqslant p\leqslant \min\{x/m,y\}} \frac{(\log(x/m))^{k+1}}{(\log p)^k}.$$

For any $z \geqslant 2$, Chebyshev's estimate and partial summation imply that

$$\sum_{2\leqslant p\leqslant z} \frac{(\log(x/m))^{k+1}}{(\log p)^k} \ll_k z\cdot\left(\frac{\log(x/m)}{\log z}\right)^{k+1}.$$

If $z = \min\{x/m, y\}$, then $\log(x/m)/\log z \leqslant u$ and $z \leqslant (x/m)^{1-\varepsilon}y^{\varepsilon}$. We thus conclude that

$$S_2 \ll_k u^{k+1}y^{\varepsilon}x^{1-\varepsilon} \sum_{m\in\mathcal{S}(y)} \frac{f(m)}{m^{1-\varepsilon}}.$$

Together with (16.12) and (16.13), this implies that

$$\sum_{n\in\mathcal{S}(y)\cap[1,x]} f(n) \ll_k \frac{u^k y^{\varepsilon}x^{1-\varepsilon}}{\log y} \sum_{n\in\mathcal{S}(y)} \frac{f(n)}{n^{1-\varepsilon}}. \tag{16.14}$$

We then follow the proof of Theorem 16.3 to bound the right side of (16.14) (taking $\varepsilon = w/\log y$ with $e^{w-1}/w = u$). This completes the proof. □

It is also possible to use the theory of large deviations to obtain an asymptotic estimate for $\Psi(x,y)$. This is done using an analogue of (16.3) obtained by Theorem 7.2, which implies that

$$\Psi(x,y) = \frac{1}{2\pi i}\int_{(\alpha)} \prod_{p\leqslant y}\Big(1-\frac{1}{p^s}\Big)^{-1}\frac{x^s}{s}\mathrm{d}s.$$

We choose α satisfying the analogue of (16.3), which is the equation

$$\sum_{p\leqslant y}\frac{\log p}{p^\alpha-1}=\log x.$$

This argument was carried out by Hildebrand and Tenenbaum [103].

Theorem 16.5. *For $x\geqslant y\geqslant 2$ with $u=\log x/\log y$ and α as above, we have*

$$\Psi(x,y)=\frac{x^\alpha\prod_{p\leqslant y}(1-1/p^\alpha)^{-1}}{\alpha\sqrt{2\pi L}}\Big(1+O\Big(\frac{1}{u}+\frac{\log y}{y}\Big)\Big),$$

where $L=-\frac{\mathrm{d}}{\mathrm{d}\sigma}\big|_{\sigma=\alpha}\sum_{p\leqslant y}(\log p)/(p^\sigma-1)=\sum_{p\leqslant y}p^\alpha(\log p)^2/(p^\alpha-1)^2$.

More details on the subject of smooth numbers can be found in Chapter III.5 of Tenenbaum's book [172], as well as in the survey article [104] of Hildebrand and Tenenbaum.

Exercises

Exercise 16.1. Estimate the sum $\sum_{1<n\leqslant x}1/\omega(n)$ in two ways: (a) use Theorem 16.2; (b) use concentration-of-measure inequalities obtained by the method of large deviations.

Exercise 16.2. Fix $C>0$ and let $\lambda_y=\prod_{p\leqslant y}(1-1/p)^{-1}$. For $y\geqslant 1$ and $k\in\mathbb{Z}$ with $0\leqslant k\leqslant C\lambda_y$, prove that

$$\mathbb{P}_{n\in\mathcal{S}(y)}(\omega(n)=k)=G(k/\lambda_y)e^{-\lambda_y}\lambda_y^k k!^{-1}\big(1+O_C(k/\lambda_y^2)\big),$$

where $G(z)$ is defined in Theorem 16.2.

Exercise 16.3. Let $k\geqslant 0$, and let f be a multiplicative function such that $0\leqslant f\leqslant\tau_k$.

(a) Use Theorem 16.3 to show that there is some constant $C=C(k)$ such that

$$\sum_{n\in\mathcal{S}(y)\cap[1,x]}\frac{f(n)}{n}\asymp_k\exp\Big\{\sum_{p\leqslant y}\frac{f(p)}{p}\Big\}\qquad\text{for } x\geqslant y^C\geqslant 1.$$

(b) Give a new proof of the lower bound from Exercise 14.5(b) that states that $\sum_{n\leqslant x}f(n)/n\gg_k\exp\{\sum_{p\leqslant y}f(p)/p\}$ for all $x\geqslant 1$.

Part 4

Sieve methods

Chapter 17

Twin primes

The theory of Dirichlet L-functions allows us to make significant progress on our understanding of prime numbers. However, there are numerous important questions about primes that seem to be intractable using Dirichlet series because they are fundamentally of non-multiplicative character. To study them, we go back to the basics and employ the most fundamental way of detecting primes: the sieve of Eratosthenes-Legendre. We illustrate some of the main ideas by discussing the famous twin prime conjecture.

Twin primes arise naturally when studying the spacing distribution of primes. The first ten primes are 2, 3, 5, 7, 11, 13, 17, 19, 23 and 29, and the spacings between them are 1, 2, 2, 4, 2, 4, 2, 4, 6. The number 1 will never appear again as a spacing because all primes $p \geqslant 3$ are odd, and thus $p+1$ cannot be prime because it is even and > 2. By the same argument, no odd number > 1 will ever appear as a spacing between two consecutive primes. On the other hand, there is no obvious reason why the even numbers should not keep reoccurring. Already the number 2 appears four times in the above list, and the number 4 appears three times. The number 6 appears once, but this is only because we have not looked far enough yet for a second appearance. Indeed, 31 and 37 are both primes and they differ by 6.

In 1849, de Polignac conjectured that any even number should appear infinitely many times as the gap between two consecutive primes. The pairs of primes that differ by 2 (and which are necessarily consecutive) are called *twin primes*. To study them, we define their counting function

$$\pi_2(x) := \#\{\, n \leqslant x : n, n+2 \text{ are both primes} \,\}.$$

The *twin prime conjecture*, which is a special case of Polignac's conjecture, states that $\pi_2(x) \to \infty$ as $x \to \infty$.

Counting twin primes is a so-called *additive problem*: we are asking for solutions of the equation $q - p = 2$, where both p and q are prime numbers. Hence, the Dirichlet series approach, which was crucially based on the Euler product representation of the Dirichlet L-functions, is of limited use here. To make progress towards the twin prime conjecture, we revisit the combinatorial ideas of Chapter 2.

Sieving for twin primes

We begin by rewriting $\pi_2(x)$ using the sieve of Eratosthenes-Legendre: if for some $n \in (\sqrt{x+2}, x]$ the product $n(n+2)$ has no prime factors $\leqslant \sqrt{x+2}$, none of n and $n+2$ have prime divisors smaller than their square-root, and so they must both be prime. The converse is also true. Hence,

$$\pi_2(x) = \#\{\, n \leqslant x : (n(n+2), P(\sqrt{x+2}\,)) = 1 \,\} + O(\sqrt{x})$$

with $P(y) = \prod_{p \leqslant y} p$ as usual. We would like to use the inclusion-exclusion principle to estimate $\pi_2(x)$ but the most direct application of this argument produces trivial bounds (see the discussion following Theorem 2.1). Limiting our goal to an upper bound for $\pi_2(x)$, we use Legendre's idea to find that

$$\pi_2(x) \leqslant \pi_2(x, y) + O(y), \tag{17.1}$$

where

$$\pi_2(x, y) = \#\{\, n \leqslant x : (n(n+2), P(y)) = 1 \,\}$$

and y is a parameter $\leqslant \sqrt{x+2}$ that we are free to choose. We then apply inclusion-exclusion as in (2.5) to find that

$$\pi_2(x, y) = \sum_{d | P(y)} \mu(d) \cdot N_2(x; d), \tag{17.2}$$

where

$$N_2(x; d) = \#\{\, n \leqslant x : d | n(n+2) \,\}.$$

To estimate the right-hand side of (17.2), we note that each interval of length d contains exactly $\nu_2(d)$ numbers n such that $d | n(n+2)$, where

$$\nu_2(d) := \#\{\, n \in \mathbb{Z}/d\mathbb{Z} : n(n+2) \equiv 0 \,(\mathrm{mod}\, d) \,\}.$$

Adapting the argument leading to (2.3), we deduce the formula

$$N_2(x; d) = x \cdot \frac{\nu_2(d)}{d} + O(\nu_2(d)). \tag{17.3}$$

The function ν_2 is multiplicative by the Chinese Remainder Theorem and on primes it equals

$$\nu_2(p) = \begin{cases} 1 & \text{if } p = 2, \\ 2 & \text{if } p > 2. \end{cases}$$

In particular, $\nu_2(d) \leqslant 2^{\omega(d)} = \tau(d)$ for square-free integers d.

The above discussion leads us to the asymptotic formula

$$\begin{aligned}\pi_2(x,y) &= x\sum_{d|P(y)}\frac{\mu(d)\nu_2(d)}{d}+O\Big(\sum_{d|P(y)}2^{\omega(d)}\Big)\\ &= \frac{x}{2}\prod_{3\leqslant p\leqslant y}\Big(1-\frac{2}{p}\Big)+O(3^{\pi(y)}).\end{aligned}\tag{17.4}$$

As in the proof of (2.9), we are forced to choose y to be a multiple of $\log x$. We thus arrive at the estimate

$$\pi_2(x)\ll\frac{x}{(\log\log x)^2}.$$

Note, however, that this bound is worse than the trivial inequality $\pi_2(x)\leqslant\pi(x)\ll x/\log x$.

It seems that we have quickly reached an impasse. This remained the state of affairs for more than a hundred years following Legendre's work on the sieve of Eratosthenes. The great breakthrough in sieve theory that turned it from an interesting observation to an indispensable part of modern number theory was undertaken by Viggo Brun in 1915. His starting point was the realization that it is possible to replace the exact formula (17.4) by upper and lower bounds that involve a lot fewer summands, thus making the remainder terms much more manageable.

Brun's first improvement of the sieve of Eratosthenes-Legendre arises from a better understanding of the mechanics of the inclusion-exclusion principle. Recall that $\pi_2(x,y)$ counts the number of $n\leqslant x$ that are in the complement of the union of the sets $\mathcal{N}_2(x;p)=\{\,n\leqslant x:p|n(n+2)\,\}$ with $p\leqslant y$. By the union bound, we have $x-\pi_2(x,y)\leqslant T_1(x,y)$, where $T_1(x,y)=\sum_{p\leqslant y}N_2(x;p)$. The expression $x-T_1(x,y)$ then serves as a first approximation to $\pi_2(x,y)$ that always underestimates its size, because there are numbers lying in the intersection of two of the sets $\mathcal{N}_2(x;p)$. We then add to $x-T_1(x,y)$ the quantity $T_2(x,y)=\sum_{p_1<p_2\leqslant y}N(x;p_1p_2)$. This leads to an overestimation of $\pi_2(x,y)$, the reason being that there are numbers lying in the intersection of three of the sets $\mathcal{N}_2(x;p)$. At the next step we thus subtract from the expression $x-T_1(x,y)+T_2(x,y)$ the quantity $T_3(x,y)=\sum_{p_1<p_2<p_3\leqslant y}N_2(x;p_1p_2p_3)$. The resulting expression $x-T_1(x,y)+T_2(x,y)-T_3(x,y)$ underestimates $\pi_2(x,y)$.

Continuing in the above fashion, we arrive at the Bonferonni inequalities

$$\sum_{j=0}^{2\ell-1}(-1)^jT_j(x,y)\leqslant\pi_2(x,y)\leqslant\sum_{j=0}^{2\ell}(-1)^jT_j(x,y)\tag{17.5}$$

for any $\ell \in \mathbb{Z}_{\geqslant 1}$, where $T_j(x,y) = \sum_{p_1<\dots<p_j\leqslant y} N_2(x; p_1\cdots p_j)$. We rewrite these inequalities in terms of the Möbius function as

$$\sum_{\substack{d\mid P(y)\\ \omega(d)\leqslant 2\ell-1}} \mu(d)N_2(x;d) \leqslant \pi_2(x,y) \leqslant \sum_{\substack{d\mid P(y)\\ \omega(d)\leqslant 2\ell}} \mu(d)N_2(x;d). \tag{17.6}$$

We must choose ℓ so that the following two requirements are met:

- ℓ must be small enough, so that the upper and lower bounds in (17.6) have a lot fewer terms than the right-hand side of (17.2). This will allow us to estimate $\pi_2(x,y)$ for y much larger than $\log x$.
- ℓ must be large enough, so that the lower and upper bounds in (17.6) are close to the real size of $\pi_2(x,y)$.

With the above requirements in mind, we note that (17.6) implies that

$$\pi_2(x,y) = \sum_{\substack{d\mid P(y)\\ \omega(d)\leqslant 2\ell-1}} \mu(d)N_2(x;d) + O\Big(\sum_{\substack{d\mid P(y)\\ \omega(d)=2\ell}} N_2(x;d)\Big).$$

Since $N_2(x;d) = x\cdot\nu_2(d)/d + O(\nu_2(d))$ by (17.3), as well as $\nu_2(d)\leqslant 4^\ell$ when d is square-free with $\omega(d)\leqslant 2\ell$, we infer that

$$\pi_2(x,y) = x \sum_{\substack{d\mid P(y)\\ \omega(d)\leqslant 2\ell-1}} \frac{\mu(d)\nu_2(d)}{d} + O\Big(\sum_{\substack{d\mid P(y)\\ \omega(d)\leqslant 2\ell}} 4^\ell + \sum_{\substack{d\mid P(y)\\ \omega(d)=2\ell}} \frac{4^\ell x}{d}\Big). \tag{17.7}$$

We must choose ℓ large enough so that

$$\sum_{\substack{d\mid P(y)\\ \omega(d)\leqslant 2\ell-1}} \frac{\mu(d)\nu_2(d)}{d} \sim \prod_{p\leqslant y}\Big(1-\frac{\nu_2(p)}{p}\Big) = \frac{1}{2}\prod_{3\leqslant p\leqslant y}\Big(1-\frac{2}{p}\Big). \tag{17.8}$$

Theorem 16.1 implies that, when weighted with $\kappa^{\omega(n)}$, a "random" integer n tends to have $\sim \kappa\log\log n$ prime factors. Motivated by this fact, we will eventually choose $\ell = c\log\log y$ for a large enough constant c. To prove that (17.8) holds for such a choice, we start by observing the inequalities

$$\sum_{\substack{d\mid P(y)\\ \omega(d)\leqslant 2\ell-1}} \frac{\mu(d)\nu_2(d)}{d} \leqslant \prod_{p\leqslant y}\Big(1-\frac{\nu_2(p)}{p}\Big) \leqslant \sum_{\substack{d\mid P(y)\\ \omega(d)\leqslant 2\ell}} \frac{\mu(d)\nu_2(d)}{d}, \tag{17.9}$$

which are analogous to (17.6). (In fact, they follow from (17.6) with $x = P(y)$, because we then have $N_2(x;d)/x = \nu_2(d)/d$ and $\pi_2(x,y)/x = \prod_{p\leqslant y}(1-\nu_2(p)/p)$. See also Exercise 17.2.) Using (17.7) and (17.9), we find that

$$\pi_2(x,y) = x\prod_{p\leqslant y}\Big(1-\frac{\nu_2(p)}{p}\Big) + O\Big(\sum_{\substack{d\mid P(y)\\ \omega(d)\leqslant 2\ell}} 4^\ell + \sum_{\substack{d\mid P(y)\\ \omega(d)=2\ell}} \frac{4^\ell x}{d}\Big). \tag{17.10}$$

Since $\nu_2(2) = 1$ and $\nu_2(p) = 2$ for $p \geqslant 3$, Mertens' third estimate (Theorem 3.4(c)) implies that the main term of (17.10) has size $\asymp x/(\log y)^2$.

The first remainder term in (17.10) controls how many summands there are in the truncated version of the inclusion-exclusion principle (17.6). To bound it, we simply note that if $d|P(y)$ and $\omega(d) \leqslant 2\ell$, then $d \leqslant y^{2\ell}$. Hence

$$\sum_{\substack{d|P(y)\\ \omega(d)\leqslant 2\ell}} 4^\ell \leqslant 4^\ell y^{2\ell}.$$

This is small compared to the main term if $(2y)^{2\ell} \leqslant x/(\log x)^3$, say.

Finally, the second remainder term in (17.10) measures how close the upper and lower bounds in (17.6) are. It should thus become small when ℓ becomes large enough. Indeed, we have

$$\sum_{\substack{d|P(y)\\ \omega(d)=2\ell}} \frac{4^\ell x}{d} \leqslant \sum_{p_1<\dots<p_{2\ell}\leqslant y} \frac{4^\ell x}{p_1\cdots p_{2\ell}} \leqslant \frac{4^\ell x}{(2\ell)!}\Big(\sum_{p\leqslant y}\frac{1}{p}\Big)^{2\ell} \tag{17.11}$$

by rearranging the 2ℓ primes in all $(2\ell)!$ possible ways. The right side of (17.11) is $\asymp x\cdot(\log y)^2\cdot\mathbb{P}(Z=2\ell)$, where Z is a Poisson random variable of parameter $\lambda = \sum_{p\leqslant y} 2/p \sim 2\log\log y$. Since Z has mean value λ and is concentrated around its mean value with high probability (see Exercise 1.9), we can make $\mathbb{P}(Z=2\ell)$ as small as we want by letting the ratio $\ell/\sum_{p\leqslant y}1/p$ be large enough. More concretely, using the inequality $n! \geqslant n^n/e^n$, we have

$$\sum_{\substack{d|P(y)\\ \omega(d)=2\ell}} \frac{4^\ell x}{d} \leqslant x\cdot\Big(\frac{e\sum_{p\leqslant y}1/p}{\ell}\Big)^{2\ell} \ll \frac{x}{(\log y)^3}$$

for $\ell \geqslant 3.99\sum_{p\leqslant y}1/p \sim 3.99\log\log y$. In addition, we must have $(2y)^{2\ell} \leqslant x/(\log x)^3$. Such an ℓ exists as long as $y \leqslant x^{1/(8\log\log x)}$ and x is large enough.

To summarize, we have proved that

$$\pi_2(x,y) = \big\{1+O(1/\log y)\big\}\frac{x}{2}\prod_{3\leqslant p\leqslant y}\Big(1-\frac{2}{p}\Big) \tag{17.12}$$

when $2 \leqslant y \leqslant x^{1/(8\log\log x)}$ and x is large. As an immediate corollary, we have the following remarkable result due to Brun.

Theorem 17.1. *For $x \geqslant 2$, we have*

$$\pi_2(x) \ll \frac{x(\log\log x)^2}{(\log x)^2}.$$

In particular, the series $\sum_{p,\,p+2 \text{ twin primes}} 1/p$ converges.

Proof. The first part follows from (17.1) and (17.12) with $y = x^{1/(8\log\log x)}$. For the second part, we use partial summation. Alternatively, note that if $\mathcal{P}_2$ denotes the set of primes p such that $p+2$ is also prime, then we have $\sum_{p\in\mathcal{P}_2\cap(2^j,2^{j+1}]} 1/p \leqslant 2^{-j}\pi_2(2^{j+1}) \ll (\log j)^2/j^2$ by the first part. Summing this inequality over all j proves the convergence of $\sum_{p\in\mathcal{P}_2} 1/p$. □

Remark 17.2. The value of the series in the statement of Theorem 17.1 is called *Brun's constant* and its numerical calculation has an interesting history, as it led to the discovery of a bug in Intel's® PentiumTM microprocessor by Nicely (see [184]). □

The Cramér-Granville model

It is important to take a moment and understand the quality of our bound on $\pi_2(x)$. Namely, we want to understand what the expected size of $\pi_2(x)$ is and how this compares to the estimate $\pi_2(x) \ll x(\log\log x)^2/(\log x)^2$.

To answer these questions, we go back to Cramér's model. Recall the basic set-up: $(X_n)_{n\geqslant 1}$ is a sequence of independent Bernoulli random variables such that $\mathbb{P}(X_n = 1) = 1/\log n$ for $n \geqslant 3$. This sequence is presumed to model the indicator function of the primes.

Consider now the random variable $\Pi_2(x) = \sum_{n\leqslant x} X_n X_{n+2}$ as a random model of $\pi_2(x)$. A straightforward calculation reveals that $\mathbb{E}[\Pi_2(x)] \sim x/\log^2 x$ as $x \to \infty$, thus suggesting that $\pi_2(x) \sim x/\log^2 x$. However, as we mentioned when we originally introduced Cramér's model (see page 4), the random variables X_n are insensitive to arithmetic information. We should thus be careful when using them because they may lead us to false conclusions. For example, the same argument as above also suggests that the number of $n \leqslant x$ such that both n and $n+1$ are primes is $\sim x/\log^2 x$, a conclusion that is blatantly false.

To get around this issue, we modify Cramér's model following an idea due to Granville. To capture the arithmetic structure of primes modulo small integers, our new model will consist of random variables $(Y_n)_{n=1}^{\infty}$ supported on $\mathcal{N} = \{\, n > y^2 : (n, P(y)) = 1 \,\}$, where y is a large parameter to be chosen later. Theorem 2.1 implies that $\mathcal{N}$ contains approximately $\alpha := \prod_{p\leqslant y}(1-1/p)$ proportion of $\mathbb{N}$. Since an integer n in $\mathcal{N}$ is *presieved* with all primes $\leqslant y$, its chances of being prime are $\sim \alpha^{-1}/\log n > 1/\log n$.

In view of the above discussion, we define the Cramér-Granville model to be a sequence of independent Bernoulli random variables $(Y_n)_{n=1}^{\infty}$ with

$$\mathbb{P}(Y_n = 1) = \alpha^{-1} 1_{n\in\mathcal{N}}/\log n. \tag{17.13}$$

The corresponding model for the number of twin primes up to x is $\widetilde{\Pi}_2(x) = \sum_{n\leqslant x} Y_n Y_{n+2}$, for which we have

$$\mathbb{E}[\widetilde{\Pi}_2(x)] = \alpha^{-2} \sum_{y^2<n\leqslant x} \frac{1_{(n(n+2),P(y))=1}}{\log(n)\log(n+2)}.$$

If $y \leqslant x^{1/(9\log\log x)}$, then (17.12) and partial summation imply that

$$\mathbb{E}[\widetilde{\Pi}_2(x)] \sim \frac{x}{(\log x)^2} \prod_{p\leqslant y} \Big(1-\frac{\nu_2(p)}{p}\Big)\Big(1-\frac{1}{p}\Big)^{-2}.$$

The product over primes converges absolutely as $y \to \infty$, since its factors are $1 + O(1/p^2)$. Letting $y \to \infty$ leads us to conjecture that

$$\pi_2(x) \sim c_2 \cdot \frac{x}{(\log x)^2}, \qquad \text{where} \qquad c_2 = 2\prod_{p\geqslant 3} \Big(1-\frac{2}{p}\Big)\Big(1-\frac{1}{p}\Big)^{-2}.$$

The constant c_2 is called the *twin prime constant*.

In view of the above discussion, our bound on $\pi_2(x)$ is off by a factor of $(\log\log x)^2$. To remove this extra factor, we must find more efficient versions of (17.6), where the parameter y is allowed to be even larger than $x^{1/(8\log\log x)}$, while still being able to control the total error after inserting (17.3). Doing so is a delicate task that requires a good understanding of which integers d we can discard from the formula $\pi_2(x,y) = \sum_{d|P(y)} \mu(d) N_2(x;d)$ without losing too much information. In turn, this relies on a good grasp of the distribution of multiplicative functions that we studied in Chapters 15 and 16. We note, however, that we will not be able to obtain a non-trivial lower bound on $\pi_2(x,y)$ when $y = \sqrt{x+2}$, which is what would be required to settle the twin prime conjecture.

Exercises

Exercise 17.1 (The Bonferonni inequalities)**.** Let $A_1,\dots,A_k$ be subsets of a finite set X. If $A = A_1^c \cap \cdots \cap A_k^c$, then show that

$$|A| = |X| - \sum_{1\leqslant k_1\leqslant k} |A_{k_1}| \pm \cdots + (-1)^r \sum_{1\leqslant k_1<\cdots<k_r\leqslant k} |A_{k_1}\cap\cdots\cap A_{k_r}| + \Delta_r$$

for all $r \in \mathbb{Z}_{\geqslant 0}$, where $(-1)^{r+1}\Delta_r \geqslant 0$. [*Hint:* Show the identity $|A| = |X| - \sum_{k_1=1}^k |A_{k_1} \setminus \bigcup_{k_1<\ell\leqslant k} A_\ell|$ by dividing the elements of $X\setminus A$ according to the largest index k_1 such that $a \in A_{k_1}$. Then iterate this identity.]

Exercise 17.2. Let $\tilde{\nu}_2$ denote the completely multiplicative function with $\tilde{\nu}_2(p) = \nu_2(p)$, and define a probability measure on the set $\mathcal{S}(y)$ of y-smooth integers by

$$\mathbb{P}(\mathcal{A}) = \prod_{p\leqslant y} \Big(1-\frac{\nu_2(p)}{p}\Big) \sum_{n\in\mathcal{A}} \frac{\tilde{\nu}_2(n)}{n}.$$

If $\mathcal{A}_d = \{ n \in \mathcal{S}(y) : d|n \}$ with $d|P(y)$, prove $\mathbb{P}(\mathcal{A}_d) = \nu_2(d)/d$ and deduce (17.9). [*Hint:* The Bonferonni inequalities have a measure-theoretic version.]

Exercise 17.3. Adapt Brun's method to prove the following estimates:

(a) If m is $x^{1/(4\log\log x)}$-smooth, then

$$\#\{ n \leqslant x : (n, m) = 1 \} \sim x \cdot \frac{\varphi(m)}{m} \qquad (x \to \infty).$$

(b) $\#\{x - y < p \leqslant x\} \ll y \log\log y / \log y$ for $x \geqslant y \geqslant 3$.

(c) $\#\{ n \leqslant x : n^2 + 1 \text{ is prime} \} \ll x \log\log x / \log x$.

Exercise 17.4. Let $\mathbf{h} = (h_1, \dots, h_k)$ be a k-tuple of distinct integers. For each prime p, define $\nu_{\mathbf{h}}(p)$ to be the number of congruence classes mod p occupied by the numbers $h_1, \dots, h_k$, and set

$$\mathfrak{S}(\mathbf{h}) := \prod_p \Big(1 - \frac{\nu_{\mathbf{h}}(p)}{p}\Big)\Big(1 - \frac{1}{p}\Big)^{-k}.$$

(a) Show that $\mathfrak{S}(\mathbf{h})$ is an absolutely convergent Euler product.

(b) The k-tuple $\mathbf{h}$ is called *admissible* if $\nu_{\mathbf{h}}(p) < p$ for all primes p. Show that this is equivalent to having $\mathfrak{S}(\mathbf{h}) > 0$.

(c) If $\mathbf{h}$ is an admissible k-tuple, the *Hardy-Littlewood conjecture* states that

$$\#\{ n \leqslant x : n + h_1, \dots, n + h_k \text{ are all primes} \} \sim \mathfrak{S}(\mathbf{h}) \cdot \frac{x}{(\log x)^k} \tag{17.14}$$

when $x \to \infty$. Use the Cramér-Granville model to justify this conjecture.

Exercise 17.5. Let $N \in \mathbb{Z}_{\geqslant 1}$. Use the Cramér-Granville model to predict an asymptotic formula for the number of pairs of primes (p, q) such that $p + q = 2N$.

Exercise 17.6 (Montgomery's conjecture)**.** Use a suitable version of the Cramér-Granville model to argue that, for each fixed $\varepsilon > 0$, we have

$$\pi(x; q, a) = \frac{\mathrm{li}(x)}{\varphi(q)} + O_\varepsilon\big(x^\varepsilon (x/q)^{1/2}\big)$$

uniformly for $x \geqslant q \geqslant 1$ and $(a, q) = 1$. In particular,

$$\pi(x; q, a) \sim \frac{x}{\varphi(q) \log x}$$

if x and q tend to infinity at a rate such that $q \leqslant x^{1-\varepsilon}$.

Chapter 18

The axioms of sieve theory

The general problem in sieve theory asks for bounds on the quantity

$$S(\mathcal{A}, \mathcal{P}) := \#\{\, a \in \mathcal{A} : (a, \mathcal{P}) = 1 \,\},$$

where $\mathcal{A}$ is a finite set of integers, $\mathcal{P}$ is a finite set of primes, and the notation $(a, \mathcal{P}) = 1$ means that a has no prime factors from $\mathcal{P}$.

It is convenient to generalize further this set-up. Given a sequence of weights $\mathcal{A} = (a_n)_{n=1}^{\infty} \subset \mathbb{R}_{\geqslant 0}$ with $\sum_{n=1}^{\infty} a_n < \infty$, we define

$$S(\mathcal{A}, \mathcal{P}) = \sum_{(n,\mathcal{P})=1} a_n.$$

This incorporates the quantity $\#\{\, a \in \mathcal{A} : (a, \mathcal{P}) = 1 \,\}$ by taking $a_n = 1_{\mathcal{A}}(n)$. We will switch back and forth between the two definitions, using the more general one when discussing theoretical aspects of the sieve, and the more specialized one when discussing concrete applications. This ambiguity in the notation helps us avoid the introduction of unnecessary new symbols.

Various important questions can be written in the above language.

Example 18.1. If $\mathcal{A} = \{x - y < n \leqslant x\}$ for some $x \geqslant y + 1 \geqslant 2$ and $\mathcal{P} = \{p \leqslant \sqrt{x}\}$, then $S(\mathcal{A}, \mathcal{P})$ counts primes in the interval $(x - y, x]$. If, for instance, $x = N^2$ and $y = 2N$ for some $N \in \mathbb{Z}_{\geqslant 3}$, then proving that $S(\mathcal{A}, \mathcal{P}) > 0$ is equivalent to Landau's conjecture that there is always a prime number between $(N-1)^2$ and N^2. □

Example 18.2. If $\mathcal{A} = \{ n(n+2) : 1 < n \leqslant x \}$ and $\mathcal{P} = \{p \leqslant \sqrt{x+2}\}$, then $S(\mathcal{A}, \mathcal{P})$ counts integers $n \in (\sqrt{x+2}, x]$ such that both n and $n+2$ are prime numbers, that is to say, $(n, n+2)$ is a pair of twin primes. □

Example 18.3. Generalizing the above examples, let $\mathcal{A} = \{ f(n) : x - y < n \leqslant x \}$, where $x \geqslant y \geqslant 1$ and f is a polynomial over $\mathbb{Z}$. Assume further that $f = f_1 \cdots f_r$, where $f_1, \ldots, f_r$ are irreducible polynomials over $\mathbb{Z}$ (i.e., they are primitive and irreducible over $\mathbb{Q}$), and let $\mathcal{P} = \{p \leqslant z\}$, where $z = \max\{ |f_j(n)|^{1/2} : x - y < n \leqslant x, 1 \leqslant j \leqslant r \}$. (Note that $z \sim cx^{d/2}$ when $x \to \infty$, where $d = \max_{1 \leqslant j \leqslant r} \deg(f_j)$ and c is some appropriate positive constant.) Then $S(\mathcal{A}, \mathcal{P})$ counts integers $n \in (x - y, x]$ such that $f_1(n), \ldots, f_r(n)$ are all primes $> z$. Since $f_j(n) \ll n^{\deg(f_j)}$, any such n must satisfy the inequalities $x \geqslant n \gg x^{d/(2d')}$, where $d' = \min_{1 \leqslant j \leqslant r} \deg(f_j)$. Note that for this range to be non-empty we must have that $d' \geqslant d/2$.

For instance, if $f(x) = x^2 + 1$, $y = x$ and $\mathcal{P} = \{p \leqslant \sqrt{x^2+1}\}$, then proving that $S(\mathcal{A}, \mathcal{P}) > 0$ for infinitely many values of x would imply Landau's conjecture that there are infinitely many primes of the form $n^2 + 1$. □

Example 18.4. We can also count twin primes using an alternative set-up: we take $\mathcal{A} = \{ p+2 : p \leqslant x \}$ and $\mathcal{P} = \{p \leqslant \sqrt{x+2}\}$, so that $S(\mathcal{A}, \mathcal{P})$ counts primes $p \leqslant x$ such that $p + 2$ is a prime $> \sqrt{x+2}$. As we will see later on, this alternative formulation yields better results on twin primes. □

Example 18.5. If $\mathcal{A} = \{ 2N - p : p \leqslant N \}$ and $\mathcal{P} = \{p \leqslant \sqrt{2N}\}$ for some integer $N \geqslant 2$, then $S(\mathcal{A}, \mathcal{P})$ counts primes $p \leqslant N$ such that $2N - p$ is also prime. In particular, $S(\mathcal{A}, \mathcal{P}) > 0$ if and only if we can write $2N$ as the sum of two primes (the smallest one of which we take to be p). Proving this statement for all $N \geqslant 2$ is *Goldbach's conjecture*. □

Example 18.6. If $\mathcal{A} = \{ p - 1 : p \leqslant x \}$ and $\mathcal{P} = \{ p' \leqslant x : p' \equiv 3 \pmod 4 \}$, then $S(\mathcal{A}, \mathcal{P})$ counts primes $p \leqslant x$ such that $p - 1$ has no prime factors $\equiv 3 \pmod 4$. In particular, $p - 1$ can be written as the sums of two squares.

Using a trick due to Iwaniec, we can reduce the size of primes in $\mathcal{P}$: we take $\mathcal{A} = \{ p - 1 : p \leqslant x, p \equiv 3 \pmod 8 \}$ and $\mathcal{P} = \{ p' \leqslant \sqrt{x} : p' \equiv 3 \pmod 4 \}$. We claim that all primes p counted by $S(\mathcal{A}, \mathcal{P})$ are such that $p - 1$ is the sum of two squares. It suffices to prove that $p - 1$ has no prime factors that are $3 \pmod 4$. Note that $(p-1)/2 \equiv 1 \pmod 4$. Hence, the number $(p-1)/2$ is divisible by an even number of primes $\equiv 3 \pmod 4$ (counted with multiplicity). But $p - 1 \leqslant x - 1$ can have at most one prime factor $> \sqrt{x}$. We thus conclude that if $p \equiv 3 \pmod 8$ and $(p - 1, \mathcal{P}) = 1$, then $p - 1$ has no prime factors $\equiv 3 \pmod 4$. In particular, $p - 1$ can be written as the sum of two squares. □

Typically, we study $S(\mathcal{A}, \mathcal{P})$ in a general axiomatic framework. We introduce and discuss each of the three sieve axioms in the following sections.

Axiom 1: Generalizing the Kubilius model

By Möbius inversion, we have

$$S(\mathcal{A},\mathcal{P}) = \sum_n a_n \sum_{d|(n,\mathcal{P})} \mu(d) = \sum_{d|\mathcal{P}} \mu(d) A_d, \tag{18.1}$$

where the notation $d|\mathcal{P}$ means that $d|\prod_{p\in\mathcal{P}} p$ (i.e., d is square-free and all of its prime factors lie in $\mathcal{P}$), and

$$A_d := \sum_{n\equiv 0 \,(\mathrm{mod}\, d)} a_n = \sum_m a_{dm}. \tag{18.2}$$

In the important case when we are sieving a set $\mathcal{A}$ instead of a sequence, we have

$$A_d = \#\{\, a \in \mathcal{A} : a \equiv 0 \,(\mathrm{mod}\, d)\,\}.$$

In order to proceed further, we must estimate A_d asymptotically. We work out such an estimate in each of the examples discussed above:

Example 18.1: Here $\mathcal{A} = \{x - y < n \leqslant x\}$, so $A_d = y/d + O(1)$.

Example 18.2: We have $\mathcal{A} = \{\, n(n+2) : n \leqslant x\,\}$. Thus, relation (17.3) implies that $A_d = x\cdot\nu_2(d)/d + O(\nu_2(d))$, where $\nu_2(d)$ counts the roots of the polynomial $x(x+2)$ mod d.

Example 18.3: Since $\mathcal{A} = \{f(n) : x - y < n \leqslant x\}$ with $x \geqslant y \geqslant 1$ and $f(x) \in \mathbb{Z}[x]$, a straightforward generalization of (17.3) implies that $A_d = y\cdot\nu_f(d)/d + O(\nu_f(d))$ with $\nu_f(d) = \#\{n \in \mathbb{Z}/d\mathbb{Z} : f(n) \equiv 0 \,(\mathrm{mod}\, d)\}$.

Example 18.4: Here we have $\mathcal{A} = \{\, p + 2 : p \leqslant x\,\}$, so that $A_d = \pi(x; d, -2)$. If d is even, then $A_d = 1$. On the other hand, if d is odd and $\leqslant (\log x)^C$ for some fixed $C > 0$, the Siegel-Walfisz theorem (Theorem 12.1) implies that $A_d = \mathrm{li}(x)/\varphi(d) + O_C(xe^{-c\sqrt{\log x}})$, where c is an absolute positive constant. Note that if we assume the Generalized Riemann Hypothesis, Exercise 11.2 implies the improved estimate $A_d = \mathrm{li}(x)/\varphi(d) + O(\sqrt{x}\log x)$ for odd d.

Example 18.5: Here we have $\mathcal{A} = \{\, 2N - p : p \leqslant N\,\}$. Consequently, $A_d = \pi(N; d, 2N)$. When $(d, 2N) > 1$, any prime $p \equiv 2N \,(\mathrm{mod}\, d)$ must divide $2N$, whence $A_d \leqslant \omega(2N) \ll \log(2N)$. On the other hand, if $(d, 2N) = 1$ with $d \leqslant (\log N)^C$, we have the estimate $A_d = \mathrm{li}(N)/\varphi(d) + O_C(Ne^{-c\sqrt{\log N}})$.

Example 18.6: In the second part, we took $\mathcal{A} = \{\, p - 1 : p \leqslant x,\, p \equiv 3 \,(\mathrm{mod}\, 8)\,\}$ and $\mathcal{P} = \{\, p' \leqslant \sqrt{x} : p' \equiv 3 \,(\mathrm{mod}\, 4)\,\}$. Therefore, $A_d = \pi(x; 8d, a_d)$ whenever $d|\mathcal{P}$, where a_d is determined by the Chinese Remainder Theorem and the congruences $a_d \equiv 1 \,(\mathrm{mod}\, d)$ and $a_d \equiv 3 \,(\mathrm{mod}\, 8)$. We thus conclude that $A_d = \mathrm{li}(x)/\varphi(8d) + O_C(xe^{-c\sqrt{\log x}})$ for $d \leqslant (\log x)^C$ with $d|\mathcal{P}$.

Observe that in all of the above examples there is a quantity X and a multiplicative function ν such that

$$A_d \sim \frac{\nu(d)}{d} \cdot X \quad \text{for all small enough } d|\mathcal{P}. \tag{18.3}$$

We summarize in the following table the values of X and ν for each of our six examples:

Example	X	$\nu(d)$
18.1	y	1
18.2	x	$\nu_2(d)$
18.3	y	$\nu_f(d)$

Example	X	$\nu(d)$
18.4	$\mathrm{li}(x)$	$1_{2\nmid d} \cdot d/\varphi(d)$
18.5	$\mathrm{li}(N)$	$1_{(d,2N)=1} \cdot d/\varphi(d)$
18.6	$\mathrm{li}(x)/4$	$d/\varphi(d)$

Note that

$$\nu(p) < p \quad \text{for all} \quad p \in \mathcal{P} \tag{18.4}$$

in each of Examples 18.1–18.6 (with the possible exception of Example 18.3 if there is a prime p such that $x^p - x | f(x)$ over the finite field $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$). Relation (18.4) means that A_p is asymptotically smaller than $X \sim A_1$, which we certainly need if we are to extract elements a_n of the sequence $\mathcal{A}$ with n having no prime factors in $\mathcal{P}$.

We denote the remainder term in the approximation (18.3) by

$$r_d := A_d - X \cdot \nu(d)/d.$$

Since $d|\mathcal{P}$ in all summands of (18.1), we only need to control r_d when $d|\mathcal{P}$. We thus arrive at the first axiom of sieve theory.

Axiom 1. There is a multiplicative function ν, a parameter X and a sequence of remainders $(r_d)_{d|\mathcal{P}}$ such that

$$A_d = \frac{\nu(d)}{d} \cdot X + r_d \qquad \text{for all} \quad d|\mathcal{P}$$

and

$$\nu(p) < p \quad \text{for all} \quad p \in \mathcal{P}.$$

In the spirit of the Kubilius model of the integers, the function $\nu(d)/d$ can be interpreted as a multiplicative density function that we denote by

$$\delta(d) := \frac{\nu(d)}{d} \in [0,1] \quad \text{for } d|\mathcal{P}. \tag{18.5}$$

Indeed, if we employ $\mathbb{N}$ with the probability measure

$$\mathbb{P}(E) := \frac{\sum_{n\in E} a_n}{\sum_{n\geqslant 1} a_n},$$

then the event $E_d = \{\, n \in \mathbb{N} : d|n \,\}$ occurs with probability

$$\mathbb{P}(E_d) = A_d/A_1 \sim \delta(d)$$

when (18.3) holds with sufficiently small remainder r_d. Hence, our assumption that ν is multiplicative means that the events $(E_p)_{p\in\mathcal{P}}$ are quasi-independent. This leads us to guess that

$$S(\mathcal{A},\mathcal{P}) \sim X\prod_{p\in\mathcal{P}}(1-\delta(p)) = X\prod_{p\in\mathcal{P}}\Big(1-\frac{\nu(p)}{p}\Big). \tag{18.6}$$

The same relation can also be seen by replacing A_d by $\delta(d)X + r_d$ in (18.1) and ignoring all the remainder terms.

In Theorem 2.1, we saw that the above guess is true when $\mathcal{A} = \{n \leqslant x\}$ and $\mathcal{P} \subset [1, \log x]$. This theorem will be improved and generalized in the next chapter thus establishing (18.6) when $\mathcal{P} \subset [1, X^{o(1)}]$ for various sequences $\mathcal{A}$. On the other hand, relation (18.6) fails when $\mathcal{A} = \{n \leqslant x\}$ and $\mathcal{P} = \{p \leqslant \sqrt{x}\}$, as we discussed in Remark 3.5. This reflects the failure of the independence hypothesis for the divisibility by large primes: indeed, if $p_1 > p_2 > p_3 > x^{1/3}$, then there is no integer $n \leqslant x$ that is simultaneously divisible by p_1, p_2 and p_3. In particular, $E_{p_1} \cap E_{p_2} \cap E_{p_3} = \emptyset$, which means that the events E_{p_1}, E_{p_2}, E_{p_3} are interrelated.

Axiom 2: The sifting dimension

A very useful and intuitive way to think of the quantity $\nu(p)$ is as the number of residue classes we must "remove/sieve out" modulo p in order to capture elements of our sequence $\mathcal{A}$ that are primes (or products of a few primes). For instance, in Example 18.2, we want to make both n and $n+2$ to be primes. Hence, we must sieve out all integers lying in the congruence classes $0 \,(\mathrm{mod}\, p)$ and $-2 \,(\mathrm{mod}\, p)$. Correspondingly, we have

$$\nu(p) = \#\Big(\{0 \,(\mathrm{mod}\, p)\} \cup \{-2 \,(\mathrm{mod}\, p)\}\Big) = \begin{cases} 1 & \text{if } p = 2, \\ 2 & \text{if } p \geqslant 3. \end{cases} \tag{18.7}$$

Similarly, consider the set-up of Example 18.3 with $f(x) = x^2 + 1$. In order to capture prime values of this polynomial, we must remove from the set $\{n \leqslant x\}$ all integers that lie in a congruence class $a \,(\mathrm{mod}\, p)$ such that $a^2 + 1 \equiv 0 \,(\mathrm{mod}\, p)$ for some prime $p \leqslant \sqrt{x^2+1}$. We then find that

$$\nu(p) = \begin{cases} 1 & \text{if } p = 2, \\ 2 & \text{if } p \equiv 1 \,(\mathrm{mod}\, 4), \\ 0 & \text{if } p \equiv 3 \,(\mathrm{mod}\, 4). \end{cases} \tag{18.8}$$

Naturally, the larger $\nu(p)$ gets the harder it is to estimate $S(\mathcal{A},\mathcal{P})$. A simple way of ensuring that $\nu(p)$ does not get too large is to assume the existence of a parameter $k > 0$ such that

$$\nu(p) \leqslant k \quad \text{for all} \quad p \in \mathcal{P}. \tag{18.9}$$

For each of our six examples, we have:

Example	$\nu(p)$	k
18.1	1	1
18.2	$2 - 1_{p=2}$	2
18.3	$\nu_p(f)$	$\deg(f)$

Example	$\nu(p)$	k
18.4	$1_{p>2} \cdot p/(p-1)$	3/2
18.5	$1_{p\nmid 2N} \cdot p/(p-1)$	3/2
18.6	$p/(p-1)$	3/2

where in Example 18.3 we used the fact that a polynomial of degree d has $\leqslant d$ roots over the finite field $\mathbb{F}_p$, and in Example 18.6 that $\min \mathcal{P} = 3$.

As we remarked above, the smaller k is, the easier it is to estimate $S(\mathcal{A}, \mathcal{P})$. With this in mind, note that in Examples 18.4–18.6 we may take $k \sim 1$ when p is large enough. However, in Example 18.3, the inequality $\nu_p(f) \leqslant \deg(f)$ is sharp in full generality, because a polynomial $f(x)$ factors completely mod p for a positive proportion of the primes by the Chebotarev Density Theorem (see [148, Theorems 8.3 (p. 47) and 13.4 (p. 545)], as well as relation (18.8) above for a concrete example).

It turns out that for many applications we only need an averaged form of (18.9) that allows us to reduce the value of k. There are various ways of averaging (18.9). A very useful one is the following.

Axiom 2. There are constants $\kappa \geqslant 0$ and $C > 0$ such that

$$\prod_{p \in \mathcal{P} \cap (y_1, y_2]} \left(1 - \frac{\nu(p)}{p}\right)^{-1} \leqslant \left(1 + \frac{C}{\log y_1}\right) \left(\frac{\log y_2}{\log y_1}\right)^{\kappa}$$

uniformly for $3/2 \leqslant y_1 \leqslant y_2 \leqslant \max \mathcal{P}$.

Axiom 2 is often called the *Iwaniec condition*;[1] the infimum of the values of κ satisfying it is called the *sifting dimension* of $\mathcal{A}$ with respect to the set of primes $\mathcal{P}$. Note that if (18.9) holds and there is some $\varepsilon \in (0, 1]$ such that $\nu(p)/p \leqslant 1 - \varepsilon$ for all $p \in \mathcal{P}$, then Axiom 2 holds with $\kappa = k$ and $C = C(k, \varepsilon)$ by Mertens' third estimate (Theorem 3.4(c)).

For each of our six examples, we have:

Example	$\nu(p)$	κ
18.1	1	1
18.2	$2 - 1_{p=2}$	2
18.3	$\nu_p(f)$	r

Example	$\nu(p)$	κ
18.4	$1_{p>2} \cdot p/(p-1)$	1
18.5	$1_{p\nmid 2N} \cdot p/(p-1)$	1
18.6	$p/(p-1)$	1/2

where r is the number of irreducible factors of f over $\mathbb{Q}$, and in Example 18.6 we have $\kappa = 1/2$ because $\mathcal{P} \subset \{p \equiv 3 \pmod 4\}$.

Often, we must assume a more precise version of Axiom 2.

[1] Often, the Iwaniec condition refers to a slightly weaker version of Axiom 2, where the factor $1 + C/\log y_1$ is replaced by some absolute constant C'.

Axiom 2′. There are constants κ, $k \geqslant 0$ and $\varepsilon \in (0,1]$ such that

$$\sum_{p\in\mathcal{P}\cap[1,w]} \frac{\nu(p)\log p}{p} = \kappa \log w + O(1) \quad \text{for all } w \leqslant \max \mathcal{P}$$

and

$$\nu(p) \leqslant \min\{(1-\varepsilon)p, k\} \quad \text{for all } p \in \mathcal{P}.$$

It is easy to show that Axiom 2′ implies Axiom 2 for some $C = C(\varepsilon, k, \kappa)$.

Axiom 3: The level of distribution of $\mathcal{A}$

In order for Axiom 1 to be meaningful, we must be able to show that the quantities r_d are small compared to the alleged main term $X \cdot \nu(d)/d$. In practice, we only need to show such an estimate *on average*. The precise condition that we will need is the following.

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

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

In Chapter 19 we will appeal to this axiom with $m = 1$, whereas in Chapter 21, we will use it with $m = 3$. The quantity D is called the *level of distribution* of the sequence $\mathcal{A}$. It is a measure of how well we can control the distribution of $\mathcal{A}$ among the progressions $0 \pmod d$.

Example 18.7. When $\mathcal{A} = \{ f(n) : x - y < n \leqslant x \}$ for a polynomial $f(x) \in \mathbb{Z}[x]$, as in Example 18.3 above (which incorporates Examples 18.1 and 18.2 as well), then $r_d = O(\nu_f(d))$. Since $\nu_f(p) \leqslant \deg(f) =: k$, we find that $r_d = O(\tau_k(d))$ for $d|\mathcal{P}$, so that

$$\sum_{d\leqslant D,\, d|\mathcal{P}} \tau_m(d)|r_d| \ll \sum_{d\leqslant D} \tau_m(d)\tau_k(d) \ll_{k,m} D\cdot(\log D)^{km-1}$$

by Theorem 14.2. Recalling that $X = y$ here, Axiom 3 holds with

$$D \asymp_{k,m} y/(\log y)^{A+km-1}. \tag{18.10}$$

□

Example 18.8. In Example 18.4, we noted that $r_d = O_C(xe^{-c\sqrt{\log x}})$ for $d \leqslant (\log x)^C$. This allows us to verify Axiom 3 with $D = (\log x)^C$ for any fixed $C > 0$. However, this is a much smaller level of distribution than the one we obtained in relation (18.10). This poses a serious hurdle if we want to study twin primes using the set-up of Example 18.4.

On the other hand, if we assume the Generalized Riemann Hypothesis, we have $r_d = O(\sqrt{x}\log x)$. We may thus verify Axiom 3 with $D \asymp \sqrt{x}/(\log x)^{A+m-1}$. Remarkably, Bombieri [8, 10] and A. I. Vinogradov [176]

proved unconditionally (i.e., without the assumption of any unproven hypotheses) that we have a level of distribution that is almost as strong.

Theorem 18.9 (The Bombieri-Vinogradov theorem)**.** *Fix $A \geqslant 0$. For $x \geqslant 2$ and $1 \leqslant Q \leqslant x^{1/2}/(\log x)^{A+3}$, we have*

$$\sum_{q \leqslant Q} \max_{y \leqslant x} \max_{(a,q)=1} \left| \pi(y; q, a) - \frac{\mathrm{li}(y)}{\varphi(q)} \right| \ll_A \frac{x}{(\log x)^{A+1}}. \tag{18.11}$$

This landmark result yields Axiom 3 with $m = 1$ and $D \asymp \sqrt{x}/(\log x)^{A+3}$ in Examples 18.4 and 18.6, and with $m = 1$ and $D \asymp \sqrt{N}/(\log N)^{A+3}$ in Example 18.5. We will prove it in Chapter 26. □

Remark 18.10. It is believed that the Bombieri-Vinogradov theorem can be extended significantly. More precisely, Elliott and Halberstam [41] conjectured the following improvement.

The Elliott-Halberstam conjecture. *Fix $A, \varepsilon > 0$. Relation* (18.11) *holds uniformly for $x \geqslant 2$ and $1 \leqslant Q \leqslant x^{1-\varepsilon}$.*

The Elliott-Halberstam conjecture is very deep, going well beyond the reach of the Generalized Riemann Hypothesis. Among other things, it implies that the level of distribution is $D = x^{1-\varepsilon}$ in Examples 18.4 and 18.6, and $D = N^{1-\varepsilon}$ in Example 18.5. Partial results towards it have been proven by Bombieri, Iwaniec, Fouvry, Friedlander and Zhang [11–13, 49–52, 188]. On the other hand, Friedlander, Granville, Hildebrand and Maier [53–55] disproved (18.11) when $Q = x/\exp(A(1-\varepsilon)(\log\log x)^2/\log\log\log x)$ building on the earlier work of Maier [134] that we will discuss in Chapter 30. □

The fundamental lemma of sieve theory

Assuming Axioms 1–3, our goal is to substitute the exact identity

$$S(\mathcal{A}, \mathcal{P}) = \sum_{d | \mathcal{P}} \mu(d) A_d$$

by upper and lower bounds

$$\sum_{d \in \mathscr{D}^-} \mu(d) A_d \leqslant S(\mathcal{A}, \mathcal{P}) \leqslant \sum_{d \in \mathscr{D}^+} \mu(d) A_d, \tag{18.12}$$

where $\mathscr{D}^{\pm}$ are certain subsets of $\{d | \mathcal{P}\}$ for which both sides of (18.12) can be bounded asymptotically. We will accomplish this goal in Chapter 19 by extending and improving the ideas of Brun presented in Chapter 17.

Replacing A_d by Axiom 1 in (18.12), we find that

$$X \sum_{d \in \mathscr{D}^-} \frac{\mu(d)\nu(d)}{d} + \sum_{d \in \mathscr{D}^-} \mu(d) r_d \leqslant S(\mathcal{A}, \mathcal{P}) \leqslant X \sum_{d \in \mathscr{D}^+} \frac{\mu(d)\nu(d)}{d} + \sum_{d \in \mathscr{D}^+} \mu(d) r_d.$$

In order to be able to apply Axiom 3 and estimate the sum of the remainder terms, we must assume that $\mathscr{D}^{\pm} \subseteq \{d|\mathcal{P},\, d \leqslant D\}$. On the other hand, the sets $\mathscr{D}^{\pm}$ must be chosen in a way that

$$\sum_{d\in\mathscr{D}^{\pm}} \frac{\mu(d)\nu(d)}{d} \sim \sum_{d|\mathcal{P}} \frac{\mu(d)\nu(d)}{d} = \prod_{p\in\mathcal{P}} \Big(1-\frac{\nu(p)}{p}\Big). \tag{18.13}$$

Let $y = \max \mathcal{P}$, so that $\mathcal{P} \subseteq \{p \leqslant y\}$. Since the sets $\mathscr{D}^{\pm}$ can only contain integers $\leqslant D$, considerations based on Theorems 16.3 and 16.4 suggest that (18.13) can be accomplished as long as the ratio $\log D/\log y$ is large enough. The following theorem confirms this heuristic.

Theorem 18.11 (The Fundamental Lemma of Sieve Theory)**.** *Consider $\mathcal{A}$ and $\mathcal{P}$ satisfying Axioms 1 and 2 for some $\kappa, C > 0$. Set $y = \max \mathcal{P}$ and $u_\kappa = 1 + 2/(e^{0.53/\kappa} - 1)$, and note that $1 < u_\kappa < 1 + 3.8\kappa$.*

(a) *Uniformly for $u \geqslant 1$, we have*

$$S(\mathcal{A},\mathcal{P}) = (1+O_{\kappa,C}(u^{-u/2}))X\prod_{p\in\mathcal{P}}\Big(1-\frac{\nu(p)}{p}\Big) + O\Big(\sum_{d\leqslant y^u,\, d|\mathcal{P}} |r_d|\Big).$$

(b) *Assume Axiom 3 with $m = 1$, $A = \kappa + 1$ and $D \geqslant y^{u_\kappa}$. If $\log X \gg \log y$ and D, X are large enough in terms of κ and C, then*

$$\frac{X}{100}\prod_{p\in\mathcal{P}}\Big(1-\frac{\nu(p)}{p}\Big) \leqslant S(\mathcal{A},\mathcal{P}) \leqslant 5X\prod_{p\in\mathcal{P}}\Big(1-\frac{\nu(p)}{p}\Big).$$

We will prove Theorem 18.11 in Chapter 19. In part (b) of its statement, the crucial quantity is D^{1/u_κ} because it determines the maximum size of primes we can sieve with. To get a sense of the quality of our result when κ and D vary, we discuss the case of twin primes.

Example 18.12. Recall that we have two set-ups for getting our hands on twin primes. In the first set-up, given in Example 18.2, we have $\kappa = 2$ and $D = x^{1-o(1)}$, so that $D^{1/u_\kappa} = x^{1/u_2-o(1)} \approx x^{1/7.59}$. On the other hand, in Example 18.4 we have $\kappa = 1$ and $D = x^{1/2-o(1)}$, so that $D^{1/u_2} = x^{0.5/u_1-o(1)} \approx x^{1/7.72}$. Hence, the first set-up allows us to sieve with larger primes. However, when $\kappa = 1$, it is possible to establish a version of Theorem 18.11(b) valid for $y \leqslant D^{1/2-\varepsilon}$, which becomes $y \leqslant x^{1/4-\varepsilon}$ in the set-up of Example 18.4 (see [59, Chapter 12]). On the contrary, the best known version of Theorem 18.11 when $\kappa = 2$ is valid for $y \lessapprox D^{1/4.2664}$, which becomes $y \lessapprox x^{1/4.2664}$ in the set-up of Example 18.2 (see [33, Theorem 6.1] or [34]). □

Corollary 18.13. *For $x \geqslant 2$, we have*

$$\pi_2(x) = \#\{\, p \leqslant x : p+2 \text{ is prime}\,\} \ll x/(\log x)^2$$

and

$$\#\{\, p \leqslant x : \Omega(p+2) \leqslant 7 \,\} \gg x/(\log x)^2.$$

Proof. For the first part, note that

$$\pi_2(x) \leqslant S(\mathcal{A}, \mathcal{P}) + O(x^{1/7.8})$$

with $\mathcal{A} = \{\, p+2 : p \leqslant x \,\}$ and $\mathcal{P} = \{p \leqslant x^{1/7.8}\}$. By our analysis of Example 18.4, we may apply Theorem 18.11(b) with $X = \mathrm{li}(x)$, $\nu(d) = 1_{(d,2)=1} d/\varphi(d)$ and level of distribution $D \asymp \sqrt{x}/(\log x)^{100}$ (assuming the Bombieri-Vinogradov theorem). Consequently,

$$S(\mathcal{A}, \mathcal{P}) \asymp \mathrm{li}(x) \prod_{p \in \mathcal{P}} \left(1 - \frac{\nu(p)}{p}\right) \sim \frac{x}{\log x} \prod_{3 \leqslant p \leqslant x^{1/7.8}} \left(1 - \frac{1}{p}\right) \asymp \frac{x}{(\log x)^2}$$

by Mertens' third estimate. The claimed upper bound on $\pi_2(x)$ then follows.

For the second part, we may assume that x is large enough. We have

$$\#\{\, p \leqslant x : \Omega(p+2) \leqslant 7 \,\} \geqslant S(\mathcal{A}, \mathcal{P}) \gg \frac{x}{(\log x)^2}.$$

Indeed, if $p \leqslant x$ is counted by $S(\mathcal{A}, \mathcal{P})$, then all prime factors of $p+2$ are $> x^{1/7.8}$. But an integer $\leqslant x+2$ can have at most seven prime factors $> x^{1/7.8}$. This completes the proof. □

Remark 18.14. Chen [24,25] proved that the second part of Corollary 18.13 is true with the number 7 replaced by 2. A proof of this result that comes remarkably close to the twin prime conjecture is presented in [59, Section 25.6] and in [86, Chapter 11]. □

Exercises

Exercise 18.1. For $x \geqslant y \geqslant 3$, use Theorem 18.11 to prove:

(a) If m is $x^{1/u}$-smooth, then

$$\#\{\, n \leqslant x : (n, m) = 1 \,\} = (1 + O(e^{-100u})) \cdot x\varphi(m)/m\,.$$

(b) $\#\{x - y < p \leqslant x\} \ll y/\log y\,.$

(c) $\#\{\, x - y < p \leqslant x : (p^2+1)/2 \text{ is prime} \,\} \ll y/(\log y)^2\,.$

(d) $\#\{\, n \leqslant x : \Omega(n^2+1) \leqslant 7 \,\} \gg x/\log x\,.$

(e) If $\mathbf{h} = (h_1, \dots, h_k)$ is a fixed admissible k-tuple, then

$$\#\{\, n \leqslant x : n + h_1, \dots, n + h_k \text{ are all primes} \,\} \ll_{\mathbf{h}} x/(\log x)^k\,.$$

(f) $\#\{\, p \leqslant x : p - 1 \text{ is the sum of two squares} \,\} \ll x/(\log x)^{3/2}.$

Chapter 19

The Fundamental Lemma of Sieve Theory

In Chapter 17, we saw how Brun used some simple facts about the inclusion-exclusion principle to obtain upper and lower bounds for $\pi_2(x, z)$. In the present chapter, we will generalize and improve these bounds with our end goal being to establish the Fundamental Lemma of Sieve Theory.

Given an integer n and a set of primes $\mathcal{P}$, let us write $\mathcal{P}^-(n)$ to denote the smallest prime factor of n from the set $\mathcal{P}$ with the convention that $\mathcal{P}^-(n) = 1$ if $(n, \mathcal{P}) = 1$. Given any sequence $\mathcal{A} = (a_n)_{n=1}^{\infty} \subset \mathbb{R}_{\geqslant 0}$ with $\sum_{n\geqslant 1} a_n < \infty$, we have

$$\begin{aligned} S(\mathcal{A}, \mathcal{P}) = \sum_{(n,\mathcal{P})=1} a_n &= \sum_{n\geqslant 1} a_n - \sum_{p_1\in\mathcal{P}} \sum_{\mathcal{P}^-(n)=p_1} a_n \\ &= \sum_{n\geqslant 1} a_n - \sum_{p_1\in\mathcal{P}} \sum_{\mathcal{P}^-(m)\geqslant p_1} a_{p_1 m}. \end{aligned} \tag{19.1}$$

This formula is called *Buchstab's identity* and its importance is that it allows us to perform inclusion-exclusion one step at a time. To see why it is true, note that if $n \geqslant 1$ is such that $(n, \mathcal{P}) > 1$, then there is a unique $p_1 \in \mathcal{P}$ with $\mathcal{P}^-(n) = p_1$. Equivalently, $p_1 | n$ and $\mathcal{P}^-(n/p_1) \geqslant p_1$. Setting $n = mp_1$ completes the proof of (19.1).

Recall the notation A_d defined in (18.2). The first term on the right side of (19.1) equals A_1, so it can be estimated using Axiom 3. Next, we want to estimate the double sum over p_1 and m in (19.1). For each fixed $p_1 \in \mathcal{P}$, we are asking for a bound on $S(\mathcal{A}_{p_1}, \mathcal{P} \cap [2, p_1))$, where $\mathcal{A}_{p_1} = (a_{p_1 m})_{m=1}^{\infty}$. Getting such a bound might be impossible for certain p_1. For instance, if

p_1 is bigger than the level of distribution D, then we cannot say anything meaningful about $\sum_{m\geqslant 1} a_{p_1 m} = A_{p_1}$. For this reason, we will discard certain "inconvenient" primes p_1. In general, given any set $\Pi_1 \subseteq \mathcal{P}$, we have the upper bound

$$S(\mathcal{A},\mathcal{P}) \leqslant A_1 - \sum_{p_1\in\Pi_1} \sum_{\mathcal{P}^-(m)\geqslant p_1} a_{p_1 m}.$$

We now iterate the above argument: applying Buchstab's identity (19.1) with $\mathcal{P}\cap[2,p_1)$ in place of $\mathcal{P}$, and with $(a_{p_1 m})_{m=1}^{\infty}$ in place of $(a_n)_{n=1}^{\infty}$, yields

$$\sum_{\mathcal{P}^-(m)\geqslant p_1} a_{p_1 m} = A_{p_1} - \sum_{p_2\in\mathcal{P}\cap[2,p_1)} \sum_{\mathcal{P}^-(m)\geqslant p_2} a_{p_1 p_2 m}.$$

Hence, our upper bound for $S(\mathcal{A},\mathcal{P})$ can be rewritten as

$$S(\mathcal{A},\mathcal{P}) \leqslant A_1 - \sum_{p_1\in\Pi_1} A_{p_1} + \sum\sum_{\substack{p_2<p_1\\ p_1\in\Pi_1,\, p_2\in\mathcal{P}}} \sum_{\mathcal{P}^-(m)\geqslant p_2} a_{p_1 p_2 m}.$$

The "unknown" rightmost sum has non-negative weight now, so we cannot drop any potentially inconvenient terms from it. We rewrite it using a new application of Buchstab's identity (19.1), this time with $\mathcal{P}\cap[2,p_2)$ as our set of primes, and with $(a_{p_1 p_2 m})_{m=1}^{\infty}$ in place of $(a_n)_{n=1}^{\infty}$. Thus

$$\begin{aligned} S(\mathcal{A},\mathcal{P}) \leqslant A_1 - \sum_{p_1\in\Pi_1} A_{p_1} + \sum\sum_{\substack{p_2<p_1\\ p_1\in\Pi_1,\, p_2\in\mathcal{P}}} A_{p_1 p_2} \\ - \sum\sum\sum_{\substack{p_3<p_2<p_1\\ p_1\in\Pi_1,\, p_2,p_3\in\mathcal{P}}} \sum_{\mathcal{P}^-(m)\geqslant p_3} a_{p_1 p_2 p_3 m}. \end{aligned}$$

We may now choose any set $\Pi_3 \subseteq \mathcal{P}\times\mathcal{P}\times\mathcal{P}$ and obtain an upper bound:

$$\begin{aligned} S(\mathcal{A},\mathcal{P}) \leqslant A_1 - \sum_{p_1\in\Pi_1} A_{p_1} + \sum\sum_{\substack{p_2<p_1\\ p_1\in\Pi_1,\, p_2\in\mathcal{P}}} A_{p_1 p_2} \\ - \sum\sum\sum_{\substack{p_3<p_2<p_1\\ p_1\in\Pi_1,\, (p_1,p_2,p_3)\in\Pi_3}} \sum_{\mathcal{P}^-(m)\geqslant p_3} a_{p_1 p_2 p_3 m}. \end{aligned}$$

Continuing this way, we find that, given any choice of sets $\Pi_{2j-1} \subseteq \mathcal{P}^{2j-1}$, $j\in\mathbb{Z}_{\geqslant 1}$, we have the general upper bound

$$S(\mathcal{A},\mathcal{P}) \leqslant \sum_{d\in\mathscr{D}^+} \mu(d) A_d, \tag{19.2}$$

where

$$\mathscr{D}^+ = \left\{ d = p_1\cdots p_r : \begin{array}{l} p_1 > \cdots > p_r,\ p_j\in\mathcal{P} \text{ for all } j, \\ (p_1,\dots,p_j)\in\Pi_j \text{ for all odd } j\leqslant r \end{array} \right\}. \tag{19.3}$$

Similarly, given any choice of sets $\Pi_{2j} \subseteq \mathcal{P}^{2j}$, $j \in \mathbb{Z}_{\geqslant 1}$, iterating Buchstab's identity and dropping certain terms at every even step, leads us to the lower bound

$$S(\mathcal{A}, \mathcal{P}) \geqslant \sum_{d \in \mathscr{D}^-} \mu(d) A_d, \tag{19.4}$$

where

$$\mathscr{D}^- = \left\{ d = p_1 \cdots p_r : \begin{array}{l} p_1 > \cdots > p_r,\ p_j \in \mathcal{P} \text{ for all } j, \\ (p_1, \dots, p_j) \in \Pi_j \text{ for all even } j \leqslant r \end{array} \right\}. \tag{19.5}$$

Evidently, this construction offers a great deal of flexibility. For example, the choice $\Pi_j = \mathcal{P}^j$ for $j \leqslant 2\ell$ and $\Pi_j = \emptyset$ for $j > 2\ell$ corresponds to the Bonferonni inequalities that led to (17.5) (see also Exercise 17.2). This choice is often called *Brun's pure sieve*.

Generally speaking, the upper and lower bounds for $S(\mathcal{A}, \mathcal{P})$ we obtained in (19.2) and (19.4) constitute part of the theory of the so-called *combinatorial sieve*. This is not the only way of producing bounds for $S(\mathcal{A}, \mathcal{P})$, as we will see in Chapter 21.

Brun's sieve

Brun introduced more sophisticated choices of sets Π_j that can be motivated by considering what the prime factors of a typical integer look like. We present a variation of his argument below.

Throughout, we let

$$y = \max \mathcal{P} \quad \text{and} \quad D = y^u.$$

Recall that $\mathcal{S}(y)$ denotes the set of y-smooth numbers. In view of relation (18.13) and the discussion surrounding it, our goal is to choose the sets Π_j in such a way that

$$\sum_{d \in \mathscr{D}^{\pm}} \frac{\mu(d)\nu(d)}{d} \sim \sum_{d \mid \mathcal{P}} \frac{\mu(d)\nu(d)}{d} = \sum_{d \in \mathcal{S}(y)} \frac{\mu(d)\nu(d) 1_{d \mid \mathcal{P}}}{d}, \tag{19.6}$$

while ensuring that $\mathscr{D}^{\pm} \subseteq [1, D]$.

If we assume that $\nu(p) 1_{p \in \mathcal{P}} \sim \kappa$ on average (e.g. we assume Axiom 2′), then $\nu(d) 1_{d \mid \mathcal{P}}$ behaves similarly to $\kappa^{\omega(d)}$ on average. Now, let $p_1 > p_2 > \cdots > p_r$ be the prime factors of d in decreasing order. A variation of Theorem 16.1 implies that, when we weigh $d \in \mathcal{S}(y)$ with $\kappa^{\omega(d)}/d$, the sequence $\{\log p_1, \dots, \log p_r\}$ typically decays exponentially with ratio of consecutive terms $\approx \exp(-1/\kappa)$. In addition, for the largest prime factor, we typically have $\log p_1 \asymp \log y$. Hence, the typical asymptotic behavior of the prime factors of d is

$$\log p_j \approx (\log y) \cdot e^{-j/\kappa}.$$

Here we are weighing d with $\mu(d)\nu(d)1_{d|\mathcal{P}}/d$ that has alternating signs, but we still expect a similar behavior for the prime factors of d: almost all the weight of $\sum_{d\in\mathcal{S}(y)}\mu(d)\nu(d)1_{d|\mathcal{P}}/d$ should be supported on integers for which $\log p_j/\log y \approx \exp(-cj/\kappa)$ for some appropriate c.

Motivated by the above discussion, we set

$$\Pi_j = \{\,(p_1,\dots,p_j)\in\mathcal{P}^j : p_1>\cdots>p_j,\ p_j\leqslant y_j\,\}, \tag{19.7}$$

where y_j are certain cut-off parameters that decay doubly exponentially. Their precise definition is a bit technical: given an integer $J\geqslant 0$ and real numbers $\alpha\in(0,1)$ and $w\in[2,y]$, we let

$$y_{2j-1}=y_{2j}=\begin{cases} y & \text{if } j\leqslant J,\\ y^{\alpha^{j-J}} & \text{if } J<j\leqslant K,\\ w & \text{otherwise,}\end{cases} \tag{19.8}$$

where K is the largest integer such that $y^{\alpha^{K-J}}\geqslant w$.

Note that $y_j=y$ for $j\leqslant 2J$, so that the sets Π_j do not restrict the first $2J$ prime factors of integers $d\in\mathscr{D}^{\pm}$. This will ensure (19.6) when $J\to\infty$, provided that α is close enough to 1. In addition, note that the sets Π_j do not restrict the prime factors $\leqslant w$ of integers $d\in\mathscr{D}^{\pm}$. This last condition is of a more technical nature and the reason why we insert it will become clearer later on (see relation (19.13) below).

Choosing J, α and w appropriately, we prove:

Theorem 19.1 (The Fundamental Lemma of Sieve Theory, II)**.** *Let $\kappa>0$, $C\geqslant 1$, $y\geqslant 1$, $\mathcal{P}\subseteq\{p\leqslant y\}$ and $D=y^u$ with $u\geqslant u_\kappa=1+2/(e^{0.53/\kappa}-1)$. If D is large enough in terms of κ and C, then there are two arithmetic functions $\lambda^{\pm}$ such that:*

(a) $\lambda^{\pm}(1)=1$, $|\lambda^{\pm}|\leqslant 1$, $\mathrm{supp}(\lambda^{\pm})\subset\{\,d|\mathcal{P} : d\leqslant D\,\}$;

(b) $(1*\lambda^-)(n)\leqslant 1_{(n,\mathcal{P})=1}\leqslant(1*\lambda^+)(n)$ *for all* $n\in\mathbb{N}$;

(c) *if ν is any multiplicative function such that $0\leqslant\nu(p)<p$ for all $p\in\mathcal{P}$, and which satisfies Axiom 2 with parameters κ and C, then*

$$\frac{11}{10^3}\prod_{p\in\mathcal{P}}\Big(1-\frac{\nu(p)}{p}\Big)\leqslant\sum_{d|\mathcal{P}}\frac{\lambda^-(d)\nu(d)}{d}\leqslant\sum_{d|\mathcal{P}}\frac{\lambda^+(d)\nu(d)}{d}\leqslant 4.9\prod_{p\in\mathcal{P}}\Big(1-\frac{\nu(p)}{p}\Big)$$

and

$$\sum_{d|\mathcal{P}}\frac{\lambda(d)\nu(d)}{d}=\{1+O_{\kappa,C}(u^{-u/2})\}\prod_{p\in\mathcal{P}}\Big(1-\frac{\nu(p)}{p}\Big)\quad\text{for }\lambda\in\{\lambda^+,\lambda^-\}.$$

Firstly, let us prove how this more technical form of the fundamental lemma allows us to deduce Theorem 18.11.

Proof of Theorem 18.11 assuming Theorem 19.1. Let us consider $\lambda^{\pm}$ as in the statement of Theorem 19.1, with $\mathcal{P}, y, \kappa, C$ as in Theorem 18.11 and $D = y^u$. Since $1_{(n,\mathcal{P})=1} \leqslant (1 * \lambda^+)(n)$, we have $S(\mathcal{A}, \mathcal{P}) \leqslant \sum_n a_n \sum_{d|n} \lambda^+(d)$. Interchanging the order of summation and applying Axiom 1 yields the upper bound

$$S(\mathcal{A}, \mathcal{P}) \leqslant X \sum_d \frac{\lambda^+(d)\nu(d)}{d} + R^+, \quad \text{where} \quad R^+ := \sum_d \lambda^+(d) r_d. \tag{19.9}$$

Similarly, we have the lower bound

$$S(\mathcal{A}, \mathcal{P}) \geqslant X \sum_d \frac{\lambda^-(d)\nu(d)}{d} + R^-, \quad \text{where} \quad R^- := \sum_d \lambda^-(d) r_d. \tag{19.10}$$

Since $|\lambda^{\pm}| \leqslant 1$ and $\operatorname{supp}(\lambda^{\pm}) \subset \{d | \mathcal{P}, d \leqslant D\}$, we have $|R^{\pm}| \leqslant \sum_{d|\mathcal{P},\, d \leqslant D} |r_d|$. If we assume Axiom 3 with $A = \kappa + 1$ and $m = 1$, we thus have $|R^{\pm}| \leqslant X/(\log X)^{\kappa+1}$. If we further suppose that $\log X \gg \log y$ and that X is large enough, then Axiom 2 implies that $|R^{\pm}| \leqslant 10^{-100} X \prod_{p \in \mathcal{P}} (1 - \nu(p)/p)$.

By the above discussion and Theorem 19.1, both parts of Theorem 18.11 follow immediately when $u \geqslant u_\kappa$. It remains to prove part (a) when $u \leqslant u_\kappa$.

Let $z = D^{1/u_\kappa} \leqslant y$ and $\mathcal{P}_{\leqslant z} = \mathcal{P} \cap [2, z]$. We then have

$$0 \leqslant S(\mathcal{A}, \mathcal{P}) \leqslant S(\mathcal{A}, \mathcal{P}_{\leqslant z}) \ll_{\kappa, C} X \prod_{p \in \mathcal{P}_{\leqslant z}} (1 - \nu(p)/p)$$

by the portion of part (a) already proven. In view of Axioms 1 and 2, we have

$$1 \leqslant \frac{\prod_{p \in \mathcal{P}_{\leqslant z}} (1 - \nu(p)/p)}{\prod_{p \in \mathcal{P}} (1 - \nu(p)/p)} \leqslant (1 + C/\log z) \cdot (u_\kappa/u)^\kappa \ll_{\kappa, C} 1.$$

Hence, $0 \leqslant S(\mathcal{A}, \mathcal{P}) \ll X \prod_{p \in \mathcal{P}} (1 - \nu(p)/p)$, and Theorem 18.11(a) follows in this case too by assuming the implicit constant in its statement is large enough. □

Proof of Theorem 19.1. Let $y^* = y^*(\kappa, C)$ be a large enough constant to be chosen later. If $y \leqslant y^*$, we simply take $\lambda^{\pm}(d) = \mu(d) \cdot 1_{d|\mathcal{P}}$. We then trivially have

$$\sum_{d|\mathcal{P}} \frac{\lambda^{\pm}(d)\nu(d)}{d} = \prod_{p \in \mathcal{P}} \Big(1 - \frac{\nu(p)}{p}\Big).$$

In addition, any d in the support of $\lambda^{\pm}$ satisfies $d \leqslant \prod_{p \leqslant y^*} p \leqslant D$ provided that D is large enough. This proves the theorem in this case.

Assume now that $y \geqslant y^*$. Let $\varepsilon = \varepsilon(\kappa, C)$ be a small enough constant, and let $u^* = u^*(\kappa, C, \varepsilon)$ be a large enough constant, both to be chosen later.

We then define $\lambda^\pm(d) := 1_{d\in\mathscr{D}^\pm}\mu(d)$ with the sets Π_j given by (19.7) and the parameters y_j satisfying (19.8) with

$$w = \frac{\varepsilon \log y}{2}, \qquad \begin{cases} \alpha = e^{-0.53001/\kappa}, \quad J = 0 & \text{if } u_\kappa \leqslant u \leqslant u^*, \\ \alpha = e^{-5/u}, \quad J = \lceil 3u/10 \rceil & \text{if } u > u^*. \end{cases}$$

To check that $\lambda^\pm$ satisfy condition (a), we must verify that $\mathscr{D}^\pm \subset [1, D]$. If $d = p_1 \cdots p_s d_0 \in \mathscr{D}^-$ with $p_1 > p_2 > \cdots > p_s > w$ and $d_0 | \prod_{p\in\mathcal{P}\cap[2,w]} p$, then $p_1, \dots, p_{2J+1} \leqslant y$ and $p_{2j+1} \leqslant p_{2j} \leqslant y_{2j}$ for all $j \geqslant J+1$. In addition, if y^* is large enough, then our choice of w and the Prime Number Theorem imply that $d_0 \leqslant \prod_{p\leqslant w} p \leqslant y^\varepsilon$ for all $y \geqslant y^*$. Hence,

$$\frac{\log d}{\log y} \leqslant \frac{\log d_0}{\log y} + 2J + 1 + 2\sum_{j\geqslant J+1} \alpha^{j-J} \leqslant \varepsilon + 2J + 1 + \frac{2}{\alpha^{-1} - 1}. \tag{19.11}$$

The right side of (19.11) is $\leqslant u$ if $u \geqslant u^*$ and u^* is large enough, because $1/(\alpha^{-1} - 1) \sim u/5$ in this case. In addition, the right side of (19.11) is $\leqslant u$ if $u \in [u_\kappa, u^*]$ and ε is small enough, by the definition of u_κ. Hence, there are choices of u^* and ε such that $\mathscr{D}^- \subset [1, D]$.

Similarly, if $d = p_1 \cdots p_s d_0 \in \mathscr{D}^+$ with $p_1 > p_2 > \cdots > p_s > w$ and $d_0 | \prod_{p\in\mathcal{P}\cap[2,w]} p$, then $p_1, \dots, p_{2J} \leqslant y$, $p_{2j} \leqslant p_{2j-1} \leqslant y_{2j-1}$ for all $j \geqslant J+1$, and $d_0 \leqslant y^\varepsilon$. Arguing as above, we infer that

$$\frac{\log d}{\log y} \leqslant \varepsilon + 2J + 2\sum_{j\geqslant J+1} \alpha^{j-J} \leqslant u,$$

provided that ε and u^* are chosen appropriately. As a consequence, we also have $\mathscr{D}^+ \subset [1, D]$. This establishes condition (a).

To check that (b) is satisfied, we follow the argument leading to (19.2) and (19.4), but this time starting from the indicator version of Buchstab's identity that reads

$$1_{(n,\mathcal{P})=1} = 1 - \sum_{p\in\mathcal{P}} 1_{p_1|n} \cdot 1_{\mathcal{P}^-(n/p_1)\geqslant p_1}. \tag{19.12}$$

Alternatively, we may simply note that (b) follows by applying (19.2) and (19.4) to the sequence $\mathcal{A}$ defined by $a_k = 1_{k=n}$.

It remains to prove that the functions $\lambda^\pm$ satisfy condition (c). To this end, set

$$V(z) = \sum_{d|\mathcal{P}\cap[2,z)} \frac{\mu(d)\nu(d)}{d} = \prod_{p\in\mathcal{P},\, p<z} \Big(1 - \frac{\nu(p)}{p}\Big),$$

as well as

$$\beta = -\kappa \log \alpha = \begin{cases} 0.53001 & \text{if } u_\kappa \leqslant u \leqslant u^*, \\ 5\kappa/u & \text{if } u > u^*. \end{cases}$$

Since $\nu(p) < p$ for all $p \in \mathcal{P}$, we have $V(z) > 0$ for all z. In addition, since $y_{2j-h} = \max\{y^{\alpha^{j-J}}, w\}$ for $j > J$ and $h \in \{0,1\}$, Axiom 2 implies that

$$\frac{V(y_{2j-h})}{V(y)} = \prod_{p \in \mathcal{P} \cap (y_{2j-h}, y]} \left(1 - \frac{\nu(p)}{p}\right)^{-1} \leqslant \left(1 + \frac{C}{\log w}\right) \alpha^{-\kappa(j-J)}$$

$$\leqslant \exp((j-J)\beta + \varepsilon), \tag{19.13}$$

since we may assume that y^* is large enough so that $w = 0.5\varepsilon \log y \geqslant e^{C/\varepsilon}$ for all $y \geqslant y^*$. We remark that (19.13) also implies that

$$\sum_{p \in \mathcal{P} \cap (y_{2j-h}, y]} \frac{\nu(p)}{p} \leqslant (j-J)\beta + \varepsilon \qquad (j > J,\ h \in \{0,1\}), \tag{19.14}$$

as it can be seen using the inequality $\nu(p)/p \leqslant -\log(1 - \nu(p)/p)$.

Now, let us consider $d | \mathcal{P}$ such that $d \notin \mathscr{D}^+$. If we write $d = p_1 \cdots p_r$, then there is a unique integer $j > J$ such that

$$p_{2j-1} > y_{2j-1} \quad \text{and} \quad p_{2k-1} \leqslant y_{2k-1} \quad (1 \leqslant k < j). \tag{19.15}$$

Hence, there is a unique way to write

$$d = p_1 \cdots p_{2j-1} d',$$

where $d' | \prod_{p \in \mathcal{P} \cap [2, p_{2j-1})} p$ and the primes $p_1, \ldots, p_{2j-1}$ are a strictly decreasing sequence of elements of $\mathcal{P}$ satisfying (19.15). Since $\mu(d) = -\mu(d')$ and ν is multiplicative, we conclude that

$$V(y) - \sum_{d | \mathcal{P}} \frac{\lambda^+(d)\nu(d)}{d} = -\sum_{j > J} V_{2j-1}, \tag{19.16}$$

where we have set

$$V_m = \sum_{\substack{y_m < p_m < \cdots < p_1 \leqslant y \\ p_1, \ldots, p_m \in \mathcal{P} \\ p_i \leqslant y_i \ (i < m,\ i \equiv m \,(\mathrm{mod}\, 2))}} \frac{\nu(p_1) \cdots \nu(p_m)}{p_1 \cdots p_m} \cdot V(p_m).$$

Similarly, we have

$$V(y) - \sum_{d | \mathcal{P}} \frac{\lambda^-(d) g(d)}{d} = \sum_{j > J} V_{2j}. \tag{19.17}$$

Next, we fix an integer $m > 2J$ and proceed to the estimation of V_m. Note that $0 \leqslant V(p_m) \leqslant V(y_m)$ for $p_m > y_m$. Since the function ν is non-negative, we infer that

$$0 \leqslant V_m \leqslant V(y_m) \sum_{\substack{y_m < p_m < \cdots < p_1 \leqslant y \\ p_1, \ldots, p_m \in \mathcal{P}}} \frac{\nu(p_1) \cdots \nu(p_m)}{p_1 \cdots p_m}.$$

By rearranging the primes $p_1, \dots, p_m$ in all possible $m!$ ways, we find that

$$V_m \leqslant \frac{V(y_m)}{m!} \sum_{\substack{p_1,\dots,p_m \in \mathcal{P}\cap(y_m,y] \\ \text{distinct}}} \cdots \sum \frac{\nu(p_1)\cdots\nu(p_m)}{p_1\cdots p_m} \leqslant \frac{V(y_m)}{m!}\Bigg(\sum_{p\in\mathcal{P}\cap(y_m,y]} \frac{\nu(p)}{p}\Bigg)^m.$$

Writing $m = 2j-h$ with $h \in \{0,1\}$ and $j > J$, and applying (19.13) and (19.14), we arrive at the inequality

$$V_{2j-h} \leqslant V(y)\cdot \frac{e^{(j-J)\beta+\varepsilon}((j-J)\beta+\varepsilon)^{2j-h}}{(2j-h)!}.$$

If we let $j = J+\ell$ and sum the above inequality over all $\ell \geqslant 1$, we find that

$$\sum_{j>J} V_{2j-h} \leqslant V(y) \sum_{\ell=1}^{\infty} \frac{e^{\beta\ell+\varepsilon}(\beta\ell+\varepsilon)^{2J+2\ell-h}}{(2J+2\ell-h)!}. \tag{19.18}$$

For the first part of the theorem, note that $\beta\ell+\varepsilon \leqslant 0.53002\ell$ as long as u^* is large enough and ε is small enough. In particular, the summands on the right-hand side of (19.18) are decreasing as functions of J, and we deduce that

$$\frac{\sum_{j>J} V_{2j-h}}{V(y)} \leqslant \sum_{\ell=1}^{\infty} \frac{e^{0.53002\ell}(0.53002\ell)^{2\ell-h}}{(2\ell-h)!} \leqslant \begin{cases} 1-11/10^3 & \text{if } h=0,\\ 3.9 & \text{if } h=1,\end{cases}$$

where the last inequality is verified numerically. Together with (19.16) and (19.17), this completes the proof of the first part of condition (c).

For the second part of (c), we may assume that $\varepsilon \leqslant \kappa$ and $u \geqslant u^*$, with u^* large enough so that $\beta = 5\kappa/u \leqslant \min\{1/6, \varepsilon/2\}$ and $J = \lceil 3u/10\rceil \geqslant 2\varepsilon/\beta$. Since $n! \geqslant (n/e)^n$ for all $n \in \mathbb{Z}_{\geqslant 0}$, we have

$$\frac{\sum_{j>J} V_{2j-h}}{V(y)} \leqslant \sum_{\ell=1}^{\infty} \frac{e^{\beta\ell+\varepsilon}(\beta\ell+\varepsilon)^{2J+2\ell-h}}{(2J+2\ell-h)!} \leqslant \sum_{\ell=1}^{\infty}\left(\frac{e^{\beta/2+1}(\beta\ell+\varepsilon)}{2J+2\ell-h}\right)^{2J+2\ell-h}.$$

In addition, noticing that $e^{1+\beta/2} \leqslant e^{13/12} \leqslant 3$ and that $\beta\ell+\varepsilon \leqslant 2\max\{\beta\ell,\varepsilon\}$, we find that

$$\frac{e^{\beta/2+1}(\beta\ell+\varepsilon)}{2J+2\ell-h} \leqslant \begin{cases} 3\varepsilon/J \leqslant 10\kappa/u & \text{if } \ell \leqslant \varepsilon/\beta,\\ 3\beta = 15\kappa/u & \text{if } \ell > \varepsilon/\beta.\end{cases}$$

In any case, the right-hand side is $\leqslant 15\kappa/u$. Assuming that $u^* \geqslant 30\kappa$ as we may, we conclude that

$$\frac{\sum_{j>J} V_{2j-h}}{V(y)} \leqslant \sum_{\ell=1}^{\infty} (15\kappa/u)^{2J+2\ell-h} \ll (15\kappa/u)^{2J} \ll u^{-u/2}.$$

This completes the proof of the theorem. □

Sieve weights

A careful reexamination of the proof of the Fundamental Lemma of Sieve Theory reveals that its most crucial component is the construction of the arithmetic functions $\lambda^{\pm}$ from Theorem 19.1. These functions replace the exact Möbius inversion formula

$$1_{(n,\mathcal{P})=1} = \sum_{d|(n,\mathcal{P})} \mu(d) \tag{19.19}$$

with an upper and a lower bound of the form

$$\sum_{d|n} \lambda^{-}(d) \leqslant 1_{(n,\mathcal{P})} \leqslant \sum_{d|n} \lambda^{+}(d). \tag{19.20}$$

In general, a function λ^{+} that is supported on $\{\, d|\mathcal{P} : d \leqslant D \,\}$ and that satisfies the right inequality of (19.20) for all n is called an *upper bound sieve* of level D for the set of primes $\mathcal{P}$. We then write $\lambda^{+} \in \Lambda^{+}(D, \mathcal{P})$. Similarly, an arithmetic function $\lambda^{-} : \mathbb{N} \to \mathbb{R}$ that is supported on $\{\, d|\mathcal{P} : d \leqslant D \,\}$ and that satisfies the left inequality of (19.20) for all n is called a *lower bound sieve* of level D for the set of primes $\mathcal{P}$, and we write $\lambda^{-} \in \Lambda^{-}(D, \mathcal{P})$.

Given any choice of sets $\Pi_j \subseteq \mathcal{P}^j$, the functions $\lambda^{\pm}(d) = 1_{d \in \mathscr{D}^{\pm}} \mu(d)$ with $\mathscr{D}^{\pm}$ defined by (19.3) and (19.5) are in the classes $\Lambda^{\pm}(D, \mathcal{P})$. Indeed, this assertion follows from relation (19.12) and the discussion surrounding it.

All sieves $\lambda^{\pm} \in \Lambda^{\pm}(D, \mathcal{P})$ yield bounds for $S(\mathcal{A}, \mathcal{P})$ as per (19.9) and (19.10). A good choice of $\lambda^{\pm}$ should have the additional property that the upper bound in (19.9) and the lower bound in (19.10) are as close to each other as possible. This roughly means that the convolutions $w_n^{\pm} = (1 * \lambda^{\pm})(n)$ behave on average similarly to $1_{(n,\mathcal{P})=1}$.

The above point of view of sieve methods will be very useful when studying gaps between primes in Chapters 28 and 29, where our goal will be to construct a sieve weight w_n that correlates strongly with many of the integers $n, n+1, \dots, n+H$ being prime. In particular, the non-negativity of the sieve weights $1 * \lambda^{+}$ will be crucial.

Sifting limits and the beta sieve

There is a construction of sieve weights that yields a version of Theorem 18.11 with a smaller constant in place of u_κ: given a parameter β, we let

$$\Pi_j = \{\, (p_1, p_2, \dots, p_j) \in \mathcal{P}^j : p_1 > \cdots > p_j,\ p_1 \cdots p_j < D/p_j^{\beta} \,\}. \tag{19.21}$$

The upper and lower bound sieves produced from these sets Π_j are together called the *beta sieve*. It was introduced by Rosser and was fully developed by Iwaniec. To explain why we choose the sets Π_j in this specific way, we must introduce the concept of the *sifting limit*.

Given a dimension κ, let β_κ be the infimum of all numbers β such that

$$S(\mathcal{A},\mathcal{P}) \gg X\prod_{p\in\mathcal{P}}\left(1-\frac{\nu(p)}{p}\right) \tag{19.22}$$

whenever the pair $(\mathcal{A},\mathcal{P})$ satisfies Axioms 1–3, the second one with dimension κ and the third one with level of distribution $D \geqslant (\max\mathcal{P})^\beta$. That is to say, β_κ is the infimum of all numbers with which we can replace u_κ in Theorem 18.11 and still have a version of the lower bound there.

Now, given a pair $(\mathcal{A},\mathcal{P})$ as above and some primes $p_1 > \cdots > p_j$ from the set $\mathcal{P}$, we have

$$A_{p_1\cdots p_j d} = \frac{\nu(d)}{d}\cdot\frac{\nu(p_1\cdots p_j)}{p_1\cdots p_j}X + r_{p_1\cdots p_j d}$$

whenever $d|\prod_{p\in\mathcal{P}\cap[2,p_j)} p$. We thus see that Axiom 1 holds for the sequence $\mathcal{A}_{p_1\cdots p_j} := (a_{p_1\cdots p_j m})_{m=1}^\infty$ and the set of primes $\mathcal{P}\cap[2,p_j)$ with $X\prod_{i=1}^j \nu(p_i)/p_i$ in place of X, $r_{p_1\cdots p_j d}$ in place of r_d and with the same multiplicative density $\nu(d)/d$. It is reasonable to expect that $\mathcal{A}_{p_1\cdots p_j}$ has level of distribution $D/(p_1\cdots p_j)$ (for instance, consider the case when $\mathcal{A} = \{x-y<n\leqslant x\}$). Hence, if we assume that $\beta > \beta_\kappa$, then

$$S(\mathcal{A}_{p_1\cdots p_j},\mathcal{P}\cap[2,p_j)) \gg X\prod_{\substack{p\in\mathcal{P}\\ p<p_j}}\left(1-\frac{\nu(p)}{p}\right) \qquad \text{when} \quad D/(p_1\cdots p_j)\geqslant p_j^\beta$$

by our hypothesis (19.22). Assume, now, we want to construct a lower bound sieve for $S(\mathcal{A},\mathcal{P})$ using iterations of Buchstab's identity (19.1) as explained earlier in this chapter: we have

$$\begin{aligned} S(\mathcal{A},\mathcal{P}) &= A_1 - \sum_{p_1\in\mathcal{P}} S(\mathcal{A}_{p_1},\mathcal{P}\cap[2,p_1)) \\ &= A_1 - \sum_{p_1\in\mathcal{P}} A_{p_1} + \sum_{p_1,p_2\in\mathcal{P},\, p_2<p_1}\sum S(\mathcal{A}_{p_1p_2},\mathcal{P}\cap[2,p_2)). \end{aligned}$$

Since we cannot control the terms with $D/(p_1p_2) < p_2^\beta$, we drop them and set

$$\Pi_2 = \{\,(p_1,p_2)\in\mathcal{P}^2 : p_1>p_2,\ p_1p_2 < D/p_2^\beta\,\}.$$

Continuing as above, and dropping each time the terms with $D/(p_1\cdots p_j) < p_j^\beta$, we arrive at the choice (19.21) for the sets Π_j.

Notice that our hypothesis that relation (19.22) holds when $\max\mathcal{P}\leqslant D^{1/\beta}$ is fed into itself, thus becoming a "self-fulfilling prophecy". This is a typical feature of sieve-theoretic functions, whose asymptotic behavior is often ruled by delay differential equations. We already saw this phenomenon in the study of smooth and rough numbers. In the case of the beta sieve,

Iwaniec proved that choosing the sets Π_j by (19.21) for an appropriate value of β leads to the inequalities

$$f(u)+o(1) \leqslant \frac{S(\mathcal{A},\mathcal{P})}{\prod_{p\in\mathcal{P}}(1-\nu(p)/p)} \leqslant F(u)+o(1) \quad (D,X\to\infty)$$

under Axioms 1–3, where $\mathcal{P}\subseteq[1,y]$, $u=\log D/\log y$ and the functions f and F are the solutions to the following system of delay differential equations:

$$\begin{cases} u^\kappa F(u)=A & \text{if } u\leqslant\beta+1,\\ u^\kappa f(u)=B & \text{if } u\leqslant\beta,\end{cases} \quad \begin{cases} (u^\kappa F(u))'=\kappa u^{\kappa-1}f(u-1) & \text{if } u>\beta+1,\\ (u^\kappa f(u))'=\kappa u^{\kappa-1}F(u-1) & \text{if } u>\beta\end{cases}$$

for certain parameters A and B. In particular, we have

$A>1$,	$B>0$,	$\beta=1$	when $\kappa<1/2$,
$A=2(e^\gamma/\pi)^{1/2}$,	$B=0$,	$\beta=1$	when $\kappa=1/2$,
$A>1$,	$B=0$,	$1<\beta<2$	when $\kappa>1/2$,
$A=2e^\gamma$,	$B=0$,	$\beta=2$	when $\kappa=1$,
$A\approx 21.7484437308$,	$B=0$,	$\beta\approx 4.8339865967$	when $\kappa=2$,

where the calculation of A and β in the last two lines is due to S. Blight.

A comprehensive discussion of the beta sieve can be found in Chapter 11 of the book by Friedlander and Iwaniec [59]. In particular, Section 11.19 there gives numerical approximations for A and β for more values of κ.

Exercises

Exercise 19.1. Given a finite set of primes and multiplicative density function $\delta:\mathbb{N}\to[0,1]$, we set

$$V_\delta(\mathcal{P})=\prod_{p\in\mathcal{P}}(1-\delta(p)).$$

If, in addition, $\delta(p)<1$ for all $p\in\mathcal{P}$, we define the *relative* density function

$$\delta^*(q)=\prod_{p|q}\frac{\delta(p)}{1-\delta(p)} \qquad \text{for all} \quad q|\mathcal{P}. \tag{19.23}$$

(a) If λ is an arithmetic function supported on $\{q|\mathcal{P}\}$, prove that

$$\sum_{q|\mathcal{P}}\lambda(q)\delta(q)=V_\delta(\mathcal{P})\sum_{q|\mathcal{P}}\delta^*(q)(1*\lambda)(q).$$

(b) If $\lambda^\pm\in\Lambda^\pm(D,\mathcal{P})$ and δ_1,δ_2 are two multiplicative functions with $0\leqslant\delta_1(p)\leqslant\delta_2(p)<1$ for all $p\in\mathcal{P}$, then prove the *monotonicity principles*

$$\frac{\sum_{q|\mathcal{P}}\lambda^-(q)\delta_2(q)}{V_{\delta_2}(\mathcal{P})}\leqslant\frac{\sum_{q|\mathcal{P}}\lambda^-(d)\delta_1(q)}{V_{\delta_1}(\mathcal{P})}\leqslant 1$$

and

$$\frac{\sum_{q|\mathcal{P}}\lambda^+(q)\delta_2(q)}{V_{\delta_2}(\mathcal{P})}\geqslant\frac{\sum_{q|\mathcal{P}}\lambda^+(q)\delta_1(q)}{V_{\delta_1}(\mathcal{P})}\geqslant 1.$$

Exercise 19.2. It is often easier to construct upper bound sieves rather than lower bound ones. This exercise shows how to pass from a collection of upper bound sieves to a lower bound sieve.

Consider a number $D \geqslant 1$ and a set of primes $\mathcal{P}$. Suppose that for each prime $p \in \mathcal{P}$ we are given a sieve $\lambda_p^+ \in \Lambda^+(D/p, \mathcal{P} \cap [2,p))$. Show that the function

$$\lambda^-(d) = \begin{cases} 1 & \text{if } d = 1, \\ -\lambda_p^+(d/p) & \text{if } d|\mathcal{P},\ d > 1,\ p = P^-(d), \\ 0 & \text{otherwise.} \end{cases}$$

is a lower bound sieve of level D for the set of primes $\mathcal{P}$.

Exercise 19.3* (The Brun-Hooley sieve)**.** This exercise develops a variation of Brun's pure sieve that leads to results of the same strength as Theorem 18.11. Throughout, $\mathcal{A}$ and $\mathcal{P}$ satisfy Axioms 1–3 with sifting dimension κ, level of distribution D and $A = \kappa + 1$. In addition, $y = \max \mathcal{P}$ and $u = \log D/\log y$, with u assumed to be large enough in terms of κ and C.

(a) If $\mathcal{P} = \bigcup_{r=1}^R \mathcal{P}_r$ is a partition of the set of primes $\mathcal{P}$, then prove that

$$1_{(n,\mathcal{P})=1} \leqslant \prod_{r=1}^R \sum_{\substack{d_r|\mathcal{P}_r \\ \omega(d_r)\leqslant 2\ell_r}} \mu(d_r)$$

for any choice of integers ℓ_r.

(b) Fix $\varepsilon > 0$ and $\lambda > 1$. Set $y_r = y^{\lambda^{1-r}}$ and let R be the biggest integer such that $y_R \geqslant e^{C/\varepsilon}$. Then define $\mathcal{P}_R = \mathcal{P} \cap [2, y_R]$ as well as $\mathcal{P}_r = \mathcal{P} \cap (y_{r+1}, y_r]$ when $1 \leqslant r \leqslant R-1$. Finally, set

$$\mathscr{D}^+ = \{\, d = d_1 \cdots d_R : d_r|\mathcal{P}_r,\ \omega(d_r) \leqslant 2\ell_r\ (1 \leqslant r \leqslant R-1) \,\}$$

with $\ell_r = \lfloor r(u-u_0)(1-1/\lambda)^2/2 \rfloor$, where $u_0 = \log(\prod_{p\in\mathcal{P}_R} p)/\log y$. Prove that the function $\lambda^+(d) = 1_{d\in\mathscr{D}^+}\mu(d)$ is in the class $\Lambda^+(D,\mathcal{P})$.

(c) Let ν be the function from Axiom 1 and set

$$V_r^+ = \sum_{d|\mathcal{P}_r,\, \omega(d)\leqslant 2\ell_r} \frac{\mu(d)\nu(d)}{d}, \quad V_r = \sum_{d|\mathcal{P}_r} \frac{\mu(d)\nu(d)}{d}$$

and $E_r = V_r^+ - V_r$, with the convention that $\ell_R = \infty$. Prove that

$$0 \leqslant E_r \leqslant \sum_{d|\mathcal{P}_r,\, \omega(d)=2\ell_r+1} \frac{\nu(d)}{d} \leqslant \frac{(\kappa\log\lambda+\varepsilon)^{2\ell_r+1}}{(2\ell_r+1)!} \quad \text{for } r = 1,\dots,R-1.$$

(d) Prove that

$$\prod_{r=1}^R V_r^+ - \prod_{r=1}^R V_r = \sum_{r=1}^{R-1} V_1\cdots V_{r-1}E_rV_{r+1}^+\cdots V_{R-1}^+V_R \leqslant Se^S V_1\cdots V_R$$

with $S = \sum_{r=1}^{R-1} E_r/V_r$. Conclude that

$$S(\mathcal{A},\mathcal{P}) \leqslant (1 + O_{\kappa,C}((u+1)^{-u/2} + 1/\log X))X \prod_{p\in\mathcal{P}} \Big(1 - \frac{\nu(p)}{p}\Big).$$

(e) Use Buchstab's identity to show that

$$S(\mathcal{A},\mathcal{P}) \geqslant (1 - O_{\kappa,C}(u^{-u/2} + 1/(\log X)^C))X \prod_{p\in\mathcal{P}} \Big(1 - \frac{\nu(p)}{p}\Big).$$

Exercise 19.4*. Assume the notation and assumptions of Theorem 19.1. In particular, $\lambda^\pm$ are the sieve weights constructed in its proof, $y = \max\mathcal{P}$ and $D = y^u$. All implied constants below may depend on the parameters κ and C.

(a) Let $f : \mathbb{N} \to \mathbb{C}$ be an arithmetic function for which there is a multiplicative function ν as in Theorem 19.1(c) and some $S \geqslant 0$ such that

$$\Big|\sum_{d|\mathcal{P}\cap[2,z)} \frac{\mu(d)f(dm)}{d}\Big| \leqslant S\nu(m) \prod_{p\in\mathcal{P},\, p<z} \Big(1 - \frac{\nu(p)}{p}\Big)$$

for all $m|\prod_{p\in\mathcal{P},\, p\geqslant z} p$ and all $z \in [1, y]$. For $\lambda \in \{\lambda^+, \lambda^-\}$, prove that

$$\sum_{d|\mathcal{P}} \frac{\lambda(d)f(d)}{d} = \sum_{d|\mathcal{P}} \frac{\mu(d)f(d)}{d} + O\Big(\frac{S}{u^{u/2}} \prod_{p\in\mathcal{P}} \Big(1 - \frac{\nu(p)}{p}\Big)\Big).$$

(b) Let ν be as in Theorem 19.1. Assume further there is some $k \geqslant 0$ such that $0 \leqslant \nu(p) \leqslant k$ for all $p \in \mathcal{P}$. For $\lambda \in \{\lambda^+, \lambda^-\}$ and $r \in \mathbb{N}$, prove that

$$\sum_{d|\mathcal{P}} \frac{\lambda(d)\nu(d)(\log d)^r}{d} = \sum_{d|\mathcal{P}} \frac{\mu(d)\nu(d)(\log d)^r}{d} + O\Big(\frac{(\log y)^r}{u^{u/2}} \prod_{p\in\mathcal{P}} \Big(1 - \frac{\nu(p)}{p}\Big)\Big)$$

with the implied constant depending also on r and k.

(c) Let ν be as in Theorem 19.1. For $\lambda \in \{\lambda^+, \lambda^-\}$ and $x \geqslant y$, prove that

$$\sum_{\substack{n\leqslant x\\(n,\mathcal{P})=1}} \frac{\nu(n)\mu^2(n)}{n} = \sum_{n\leqslant x} \frac{(\lambda * 1)(n)\nu(n)\mu^2(n)}{n} + O\Big(\frac{1}{u^{u/2}} \prod_{\substack{p\leqslant x\\p\notin\mathcal{P}}} \Big(1 + \frac{\nu(p)}{p}\Big)\Big).$$

[*Hint:* Use part (a) with $f(d) = \sum_{a\leqslant x/d} \nu(da)\mu^2(da)/a$, $\nu^*(d) = \prod_{p|d} \nu(p)/(1+\nu(p)/p)$ in place of ν, and $S = \prod_{p\leqslant x}(1 + \nu(p)/p) = \prod_{p\leqslant x}(1 - \nu^*(p)/p)^{-1}$.]

Exercise 19.5* (A study of the beta sieve)**.** Let $\mathcal{P}$ be a set of primes and $y = 1 + \max\mathcal{P}$. Given $D = y^u$ with $u \geqslant 2$, let $\lambda^\pm(d) = 1_{d\in\mathscr{D}^\pm}\mu(d)$ be the beta sieve weights (i.e., $\mathscr{D}^\pm$ are given by (19.3) and (19.5) with the sets Π_j given by (19.21)). In addition, let ν be a multiplicative function such that $0 \leqslant \nu(p) \leqslant \min\{p - 1, k\}$ for all $p \in \mathcal{P}$.

(a) Let $m = 2j - h$ with $h \in \{0, 1\}$. Assume that $p_1 > p_2 > \cdots$ are some primes in $\mathcal{P}$ such that $p_1 \cdots p_{m-1}p_m^{\beta+1} > D$ and $p_1 \cdots p_{n-1}p_n^{\beta+1} \leqslant D$ for all $n < m$ with $n \equiv h \pmod 2$. Prove that

$$p_1p_2 \cdots p_{2i-h-1} \leqslant Dy^{-(u-1)\left(\frac{\beta-1}{\beta+1}\right)^{i-1}} \quad (1 \leqslant i \leqslant j),$$

and deduce that $p_m > y^{\delta_m}$ with $\delta_m = \frac{u-1}{\beta+1}\big(\frac{\beta-1}{\beta+1}\big)^{j-1} \geqslant \frac{1}{\beta+1}\big(\frac{\beta-1}{\beta+1}\big)^{m/2}$.

(b) If $V(z) = \prod_{p\in\mathcal{P},\, p<z}(1-\nu(p)/p)$ and

$$V_m = \sum_{\substack{y^{\delta_m}<p_m<\cdots<p_1\leqslant y\\ p_1,\dots,p_m\in\mathcal{P}\\ p_1\cdots p_{m-1}p_m^{\beta+1}>D}}\cdots\sum \frac{\nu(p_1\cdots p_m)}{p_1\cdots p_m}\cdot V(p_m),$$

prove that $V(y) = \sum_{d|\mathcal{P}}\mu(d)\nu(d)/d$ and

$$V(y) - \sum_{j\geqslant 1} V_{2j} \leqslant \sum_{d|\mathcal{P}} \frac{\lambda^-(d)\nu(d)}{d} \leqslant \sum_{d|\mathcal{P}} \frac{\lambda^+(d)\nu(d)}{d} \leqslant V(y) + \sum_{j\geqslant 1} V_{2j-1}.$$

(c) For any $\varepsilon\in[0,1)$, use Rankin's trick to prove that

$$V_m \leqslant \frac{k^m}{m!y^{\varepsilon(u-\beta)}}\Big(\sum_{y^{\delta_m}<p\leqslant y}\frac{1}{p^{1-\varepsilon}}\Big)^m V(y^{\delta_m}).$$

(d) Show that there is a choice of $\beta\leqslant 1+4k$ such that

$$\sum_{d|\mathcal{P}}\frac{\lambda^\pm(d)\nu(d)}{d} = (1+O_k(u^{-u}))V(y).$$

[*Hint:* Choose $\varepsilon = w/\log y$ as in the proof of Theorem 16.3.]

(e) If $u\geqslant 1+2/(e^{0.5295/k}-1)$ and all primes of $\mathcal{P}$ are large enough in terms of k, prove that

$$\sum_{d|\mathcal{P}}\frac{\lambda^-(d)\nu(d)}{d} \geqslant \frac{V(y)}{40}.$$

[*Hint:* Take $\varepsilon = 0$ in part (c).]

(f) Assume that $\log D \geqslant c + (1+2/(e^{0.5295/k}-1))\log y$ with c large enough in terms of k. Construct a sieve $\lambda^*\in\Lambda^-(D,\mathcal{P})$ that satisfies the conclusion of part (e) even when $\mathcal{P}$ contains small primes. [*Hint:* Partition $\mathcal{P} = \mathcal{P}_1\cup\mathcal{P}_2$, where $\mathcal{P}_1 = \mathcal{P}\cap[1,y_0]$ and $\mathcal{P}_2 = \mathcal{P}\cap(y_0,y]$. Any $d|\mathcal{P}$ can be uniquely written as $d = d_1d_2$ with $d_j|\mathcal{P}_j$. Take $\lambda^*(d) = \mu(d_1)\lambda^-(d_2)$ with λ^- a lower bound beta sieve of level D/e^c for $\mathcal{P}_2$.]

Chapter 20

Applications of sieve methods

Sieve methods are a versatile tool that can be employed in a great variety of ways. We demonstrate their utility by presenting several results where they play a key role.

Primes in short arithmetic progressions

When $y/q \geqslant x^{\varepsilon}$, a strengthening of Montgomery's conjecture (see Exercise 17.6) states that

$$\#\{\, x-y < p \leqslant x : p \equiv a \,(\mathrm{mod}\, q)\,\} \sim \frac{y}{\varphi(q)\log x} \qquad (x \to \infty). \tag{20.1}$$

This statement is well beyond the reach of the Generalized Riemann Hypothesis, which is not sufficient to detect primes in $(x-y, x]$ when $y \leqslant \sqrt{x}$, nor primes $p \equiv a \,(\mathrm{mod}\, q)$ that are $\leqslant q^2$. Nevertheless, we can use a sieve to prove an upper bound of the expected order of magnitude.

Theorem 20.1 (The Brun-Titchmarsch inequality)**.** *Uniformly for $q \in \mathbb{N}$, $a \in (\mathbb{Z}/q\mathbb{Z})^*$ and $x \geqslant y \geqslant q$, we have*

$$\#\{\, x-y < p \leqslant x : p \equiv a \,(\mathrm{mod}\, q)\,\} \ll \frac{y}{\varphi(q)\log(2y/q)}.$$

Proof. When $y \leqslant 10q$, the result follows trivially by the fact that there are $\leqslant y/q + 1$ integers in the arithmetic progression $a \,(\mathrm{mod}\, q)$ that also lie in the interval $(x-y, x]$. Let us now assume that $y > 10q$ and set $\mathcal{A} = \{\, x-y < n \leqslant x : n \equiv a \,(\mathrm{mod}\, q)\,\}$ and $\mathcal{P} = \{p \leqslant z\}$ with $z = (y/q)^{1/4}$.

Any prime $p \in (x-y, x]$ that is in the congruence class $a \pmod q$ is either $\leqslant z$, or is counted by $S(\mathcal{A}, \mathcal{P})$. Hence,

$$\#\{\, x-y < p \leqslant x : p \equiv a \pmod q \,\} \leqslant S(\mathcal{A}, \mathcal{P}) + z. \tag{20.2}$$

We estimate $S(\mathcal{A}, \mathcal{P})$ using Theorem 18.11(b). We must first verify Axioms 1–3. We have

$$A_d = \#\{\, x-y < n \leqslant x : n \equiv 0 \pmod d,\ n \equiv a \pmod q \,\}.$$

If $(d, q) > 1$, there are no integers n in the intersection of the congruence classes $0 \pmod d$ and $a \pmod q$, because $(a, q) = 1$. We thus have $A_d = 0$ when $(d, q) > 1$. Assume now that $(d, q) = 1$. For such integers d, the Chinese Remainder Theorem implies that there is a unique congruence class $a_d \pmod{qd}$ such that $A_d = \#\{\, x-y < n \leqslant x : n \equiv a_d \pmod{qd} \,\}$. Therefore, $A_d = y/(qd) + r_d$ with $|r_d| \leqslant 2$. In conclusion, Axiom 1 holds with $X = y/q$, $\nu(d) = 1_{(d,q)=1}$ and $|r_d| \leqslant 2$. Axiom 2 then obviously holds with $\kappa = 1$, and Axiom 3 with $m = 1$, $A = 2$ and $D = (y/q)/(\log(y/q))^3$. We may thus apply Theorem 18.11(b) and Mertens' third estimate (Theorem 3.4(c)) to deduce that

$$S(\mathcal{A}, \mathcal{P}) \ll \frac{y}{q} \prod_{p \leqslant z,\, p \nmid q} \Big(1 - \frac{1}{p}\Big) \asymp \frac{y}{q \log z} \prod_{p \leqslant z,\, p \mid q} \Big(1 - \frac{1}{p}\Big)^{-1}. \tag{20.3}$$

The last product is $\leqslant \prod_{p \mid q} (1 - 1/p)^{-1} = q/\varphi(q)$. This completes the proof. □

The Titchmarsch-Linnik divisor problem

Two consecutive integers are always coprime. More generally, we expect their multiplicative structure to be more-or-less uncorrelated. Thus, even if p is a prime number, then the integer $p - 1$ should still have the anatomy of a "typical" integer as described in Theorem 16.1, except for obvious restrictions such as the fact that $p - 1$ is even if $p > 2$, or that $p - 1$ cannot be congruent to $2 \pmod 3$ if $p > 3$. It is then reasonable to guess that

$$\sum_{p \leqslant x} \tau_k(p-1) \approx \frac{1}{\log x} \sum_{n \leqslant x} \tau_k(n) \asymp_k x (\log x)^{k-2}.$$

Titchmarsch studied the sum on the left-hand side when $k = 2$ and evaluated it asymptotically under the assumption of the Generalized Riemann Hypothesis. Subsequently, Linnik removed this assumption, so that the following result now holds unconditionally.

Theorem 20.2. *For $x \geqslant 3$, we have that*

$$\sum_{p \leqslant x} \tau(p-1) = \frac{\zeta(2)\zeta(3)}{\zeta(6)} x + O\Big(\frac{x \log\log x}{\log x}\Big).$$

Proof. We follow an argument due to Rodriguez [156]. Note that

$$\tau(n) = \sum_{ab=n} 1 = 1_{n=\square} + 2 \sum_{a|n,\, a<\sqrt{n}} 1.$$

Therefore,

$$\begin{aligned}\sum_{p\leqslant x} \tau(p-1) &= 2 \sum_{p\leqslant x} \sum_{\substack{a<\sqrt{p-1}\\ a|p-1}} 1 + O(\sqrt{x}) \\ &= 2 \sum_{a<\sqrt{x-1}} \big(\pi(x;a,1) - \pi(a^2+1;a,1)\big) + O(\sqrt{x})\end{aligned}$$

by interchanging the order of summation of a and p. We have the crude bound $\pi(a^2+1;a,1) = O(a^2/[\varphi(a)\log(2a)])$ from the Brun-Titchmarsch inequality. This bound is $\ll x^{1/4}$ when $a \leqslant x^{1/4}$, whereas it is $\ll (x^{1/2}/\log x)\cdot a/\varphi(a)$ when $a \in (x^{1/4}, x^{1/2}]$. As a consequence,

$$\sum_{a<\sqrt{x-1}} \pi(a^2+1,a,1) \ll x^{1/4}\cdot x^{1/4} + \frac{x^{1/2}}{\log x}\cdot x^{1/2} \ll \frac{x}{\log x}.$$

with the last estimate following by Theorem 14.2 applied with $f(a) = a/\varphi(a)$. In addition, the Bombieri-Vinogradov theorem (Theorem 18.9) implies that

$$\sum_{p\leqslant Q} \left|\pi(x,a,1) - \frac{\mathrm{li}(x)}{\varphi(a)}\right| \ll \frac{x}{\log x}$$

with $Q = \sqrt{x}/(\log x)^3$. Consequently,

$$\sum_{p\leqslant x} \tau(p-1) = 2\sum_{a\leqslant Q} \frac{\mathrm{li}(x)}{\varphi(a)} + 2 \sum_{Q<a\leqslant\sqrt{x-1}} \pi(x,a,1) + O\bigg(\frac{x}{\log x}\bigg).$$

Now, applying Wirsing's theorem (Theorem 14.3) with $f(a) = a/\varphi(a)$, we find that

$$\sum_{a\leqslant Q} \frac{1}{\varphi(a)} = \frac{\zeta(2)\zeta(3)}{\zeta(6)}\log Q + O(1) = \frac{\zeta(2)\zeta(3)}{2\zeta(6)}\log x + O(\log\log x).$$

Since $\mathrm{li}(x) = x/\log x + O(x/\log^2 x)$, we conclude that

$$\sum_{p\leqslant x} \tau(p-1) = \frac{\zeta(2)\zeta(3)}{\zeta(6)}x + 2 \sum_{Q<a\leqslant\sqrt{x-1}} \pi(x,a,1) + O\bigg(\frac{x\log\log x}{\log x}\bigg).$$

Finally, we bound crudely the sum over $a \in [Q, \sqrt{x-1}]$, taking advantage that this is a short interval if we rescale it logarithmically (we have $\log a = \log x + O(\log\log x)$ when $a \in [Q, \sqrt{x-1}]$). Indeed, using the

Brun-Titchmarsch inequality and Wirsing's theorem once again with $f(a) = a/\varphi(a)$, we have that

$$\sum_{Q<a\leqslant\sqrt{x-1}} \pi(x,a,1) \ll \sum_{Q<a\leqslant\sqrt{x-1}} \frac{x}{\varphi(a)\log x} \ll \frac{x\log\log x}{\log x}.$$

This completes the proof of the theorem. □

Multiplicative functions over short arithmetic progressions

Our last application of sieve methods is an analogue of the Brun-Titchmarsch inequality for multiplicative functions that greatly generalizes Theorem 14.2. It was proved by P. Shiu [165].

Theorem 20.3. *Fix $k \in \mathbb{N}$ and $\varepsilon > 0$. Given any choice of $q \in \mathbb{N}$, $a \in (\mathbb{Z}/q\mathbb{Z})^*$, real numbers $x \geqslant y \geqslant 1$ with $y/q \geqslant x^\varepsilon$, and a multiplicative function f such that $0 \leqslant f \leqslant \tau_k$, we have*

$$\sum_{\substack{x-y<n\leqslant x\\ n\equiv a\,(\mathrm{mod}\,q)}} f(n) \ll_{k,\varepsilon} \frac{y}{q}\exp\Big\{\sum_{\substack{p\leqslant x\\ p\nmid q}}\frac{f(p)-1}{p}\Big\}.$$

Proof. All implied constants might depend on k and ε without further notice. We may assume that x is large enough in terms of them. We begin by showing a preliminary estimate.

Set $z = y/q \in [x^\varepsilon, x]$ and note that $\sum_{z^{1/u}<p\leqslant x} 1/p \leqslant \log u + O(1)$ for $u \geqslant 1$. We thus infer the bound

$$\sum_{\substack{X-Y<n\leqslant X\\ n\equiv b\,(\mathrm{mod}\,q)\\ P^-(n)>z^{1/u}}} 1 \ll \frac{Y}{q}\prod_{\substack{p\leqslant z^{1/u}\\ p\nmid q}}\Big(1-\frac{1}{p}\Big) \ll \frac{uY}{q}\exp\Big\{-\sum_{\substack{p\leqslant x\\ p\nmid q}}\frac{1}{p}\Big\} \tag{20.4}$$

uniformly for $X \geqslant Y \geqslant y/z^{1/2}$, $u \geqslant 4$ and $(b,q) = 1$, as it can be seen by (20.3) applied with $z^{1/u}$ and b in place of z and a, respectively. Let us now show how to deduce the general case of the theorem from (20.4).

Call S the sum in the statement of the theorem. The rough idea is to fix some small parameter $\delta > 0$ and to decompose each integer n in the range of S as $n = mm'$ with $P^+(m) \leqslant z^\delta < P^-(m')$. (Such a decomposition always exists and is unique.) Anatomical considerations based on Theorems 16.1 and 16.4 suggest that $m \leqslant z^{1/2}$ for a "typical" n, as long as δ is small enough. Fixing such an m (which must also be coprime to q), we see that $m' \in (x/m - y/m, x/m]$ and $m' \equiv a\overline{m}\,(\mathrm{mod}\,q)$, where $\overline{m}$ denotes the inverse of $m\,(\mathrm{mod}\,q)$. In addition, since $0 \leqslant f \leqslant \tau_k$ and $\Omega(m') \leqslant \log x/\log(z^\delta) = O(1/\delta)$ (see Exercise 2.9(e) for the last inequality), we have $f(n) \ll f(m)$. Thus, for each fixed $m \leqslant z^{1/2}$, we should be able to estimate the sum over

m' using (20.4). The problem is that there are various "atypical" integers n for which $m' > z^{1/2}$. Dealing with them creates various technicalities. We give the details below.

First of all, since $f(n) \leqslant \tau_k(n) \ll n^{\varepsilon/4}$, the contribution of integers $\leqslant z^{1/2}$ to S is $\ll z^{1/2}x^{\varepsilon/4} \leqslant z^{3/4}$. We consider now an integer $n > z^{1/2}$ in the range of S. We decompose it in its prime factors, say $n = \prod_{i=1}^{R} p_i^{\nu_i}$ with $p_1 < p_2 < \cdots < p_R$. If we let $n_r = \prod_{i=1}^{r} p_i^{\nu_i}$, then there is a unique integer $r \in [1, R]$ such that $n_{r-1} \leqslant z^{1/2} < n_r$. We then write $m = n_{r-1}$, $p' = p_r$, $\nu = \nu_r$ and $m' = n/m$, so that $P^-(m') = p'$, $P^+(m) < p'$ and $(p')^\nu m > \sqrt{z} \geqslant m$. Since $0 \leqslant f \leqslant \tau_k$ and $\Omega(m') \leqslant \log x/\log p' \leqslant \log(z^{1/\varepsilon})/\log p'$, we have

$$f(n) = f(m)f(m') \leqslant f(m)k^{\varepsilon^{-1}\log z/\log p'}. \tag{20.5}$$

Moreover, the relation $mm' \equiv a \pmod q$ and our assumption that $(a, q) = 1$ imply that $(m, q) = 1$ and $m' \equiv a\overline{m} \pmod q$.

From the above discussion, we infer that

$$S \leqslant S_1 + S_2 + S_3 + O(z^{3/4}),$$

where S_1 is the part of S with $p' > z^{1/4}$, S_2 is the part with $p' \leqslant z^{1/4}$ and $m > z^{1/4}$ and S_3 is the part with $m, p' \leqslant z^{1/4}$. Note that $\nu \geqslant 2$ in S_3, since $z^{(\nu+1)/4} \geqslant (p')^\nu m > z^{1/2}$ for its summands. For this reason, the main contribution to S comes from S_1 and S_2. We estimate each sum individually below.

To bound S_1, we apply (20.5) and then (20.4) to find that

$$S_1 \leqslant k^{4/\varepsilon} \sum_{\substack{m\leqslant\sqrt{z}\\(m,q)=1}} f(m) \sum_{\substack{(x-y)/m<m'\leqslant x/m\\ m'\equiv a\overline{m}\,(\mathrm{mod}\,q)\\ P^-(m')>z^{1/4}}} 1 \ll \sum_{\substack{m\leqslant\sqrt{z}\\(m,q)=1}} f(m)\,\frac{z/m}{\exp\{\sum_{p\leqslant x,\,p\nmid q} 1/p\}},$$

since $z = y/q$. We then use (14.7) to arrive at the estimate

$$S_1 \ll \lambda z, \qquad \text{where} \qquad \lambda := \exp\Bigg\{\sum_{\substack{p\leqslant x\\ p\nmid q}} \frac{f(p)-1}{p}\Bigg\}.$$

Next, we estimate S_2. We introduce the checkpoints $z_j = z^{2^{-j}}$. There is a unique $J \in \mathbb{N}$ such that $z_{J+1} < (\log z)^3 \leqslant z_J$. For $j < J$, we let $\mathcal{N}_j$ be those integers n in the range of S_2 that also satisfy the inequality $z_{j+1} < P^+(m) \leqslant z_j$. Finally, we let $\mathcal{N}_J$ be the set of z_J-smooth integers n in the range of S_2. We also write $S_{2,j}$ for the contribution of $n \in \mathcal{N}_j$ to S_2.

First, we estimate $S_{2,j}$ for $j < J$. Since $p' > P^+(m) > z_{j+1} = z^{2^{-j-1}}$ in its range, an adaptation of the argument we used to bound S_1 implies that

$$\begin{aligned} S_{2,j} &\leqslant k^{2^{j+1}/\varepsilon} \sum_{\substack{z^{1/4}<m\leqslant\sqrt{z}\\(m,q)=1,\, P^+(m)\leqslant z_j}} f(m) \sum_{\substack{(x-y)/m<m'\leqslant x/m\\ m'\equiv a\overline{m}\,(\mathrm{mod}\, q)\\ P^-(m)>z_{j+1}}} 1 \\ &\ll k^{2^{j+1}/\varepsilon} \sum_{\substack{z^{1/4}<m\leqslant\sqrt{z}\\(m,q)=1,\, P^+(m)\leqslant z_j}} f(m)\frac{2^j z/m}{\exp\{\sum_{p\leqslant x,\, p\nmid q} 1/p\}}. \end{aligned}$$

We then apply Theorem 16.3 with $f(m)1_{(m,q)=1}$ in place of f to deduce that $S_{2,j} \ll \lambda z/e^j$.

In the sum $S_{2,J}$, we do not have any precise information about the position of p'. We will thus estimate the sum over m' trivially. The gains will come from the fact that m is z_J-smooth, and here $z_J \leqslant (\log z)^6$. More precisely, using the fact that $f(n) \leqslant \tau_k(n) \ll z^{0.01}$ for $n \leqslant x \leqslant z^{1/\varepsilon}$, we have

$$\begin{aligned} S_{2,J} &\ll z^{0.01} \sum_{\substack{z^{1/4}<m\leqslant\sqrt{z}\\ P^+(m)\leqslant(\log z)^6}} \sum_{\substack{x/m-y/m<m'\leqslant x/m\\ m'\equiv a\overline{m}\,(\mathrm{mod}\, q)}} 1 \\ &\ll z^{1.01} \sum_{\substack{m>z^{1/4}\\ P^+(m)\leqslant(\log z)^6}} \frac{1}{m} \ll z^{1.01-1/24+o(1)} = o_{z\to\infty}(\lambda z) \end{aligned}$$

by Theorem 16.3. We conclude that $S_2 \leqslant \sum_{j=1}^{J} S_{2,j} \ll \lambda z$.

Finally, it remains to bound S_3. Note that $(p')^\nu z^{1/4} \geqslant (p')^\nu m > z^{1/2}$, so that $(p')^\nu > z^{1/4}$ in its range. Since we also have that $p' \leqslant z^{1/4}$, there must exist some integer $\mu \geqslant 2$ such that $z^{1/4} < (p')^\mu \leqslant z^{1/2}$. Writing $n = (p')^\mu n'$, and observing that $p' \nmid q$ and that $f(n) \leqslant \tau_k(n) \ll z^{0.01}$ for $n \leqslant x$, we arrive at the estimate

$$\begin{aligned} S_3 &\ll z^{0.01} \sum_{\mu\geqslant 2} \sum_{\substack{z^{1/4}<(p')^\mu\leqslant z^{1/2}\\ p'\nmid q}} \sum_{\substack{(x-y)/(p')^\mu<n'\leqslant x/(p')^\mu\\ n'\equiv\overline{(p')^\mu}a\,(\mathrm{mod}\, q)}} 1 \\ &\ll z^{0.01} \sum_{\mu\geqslant 2} \sum_{(p')^\mu>z^{1/4}} \frac{z}{(p')^\mu} \leqslant \sum_{\mu\geqslant 2}\sum_{p'} \frac{z^{0.01+11/12}}{(p')^{2\mu/3}} \ll z^{1.01-1/12} \end{aligned}$$

by Rankin's trick, since $(p')^{\mu/3} \geqslant z^{1/12}$ when $(p')^\mu > z^{1/4}$. We thus see that the contribution of S_3 to S is negligible.

Putting together the above estimates proves that $S \ll \lambda z$, thus completing the proof of the theorem. □

Exercises

Exercise 20.1. Let $x = y^u$. Given $n \in \mathbb{N}$, we write n_y for its y-smooth part, that is to say, $n_y := \prod_{p^k \| n,\, p \leqslant y} p^k$. Uniformly for all $\mathcal{A} \subseteq \mathcal{S}(y)$, prove that

$$\mathbb{P}_{n \leqslant x}(n_y \in \mathcal{A}) = \mathbb{P}_{n \in \mathcal{S}(y)}(n \in \mathcal{A}) + O(e^{-u}).$$

Use the above relation to give an alternative proof of the Erdős-Kac theorem.

Exercise 20.2.*

(a) For $x \geqslant 2$, prove that

$$\sum_{p \leqslant x} \tau_3(p-1) \asymp x \log x.$$

[*Hint:* Show that $\sum_{d|n, d \leqslant x^{1/3}} \tau(d) \leqslant \tau_3(n) \leqslant 3 \sum_{d|n, d \leqslant x^{2/3}} \tau(d)$ when $n \leqslant x$.]

(b) Assume the Elliott-Halberstam conjecture. Prove that there is a constant $c > 0$ such that

$$\sum_{p \leqslant x} \tau_3(p-1) = cx \log x + O(x) \quad (x \geqslant 1).$$

Exercise 20.3.* Let f be a multiplicative function with $0 \leqslant f \leqslant \tau_k$.

(a) Adapt the proof of Shiu's theorem to show that

$$\sum_{p \leqslant x} f(p-1) \ll_k x \exp\Big\{ \sum_{p \leqslant x} \frac{f(p)-2}{p} \Big\}.$$

[*Hint:* You will need an estimate for $\#\{ n \leqslant x : P^-(n(2an+1)) > y \}$ uniformly in $a \in \mathbb{N}$ and $x \geqslant y \geqslant 1$.]

(b) Fix $C \geqslant 10$. Uniformly for $2 \leqslant k \leqslant C \log\log x$, show that

$$\#\{ p \leqslant x : \omega(p-1) = k \} \ll_C \sqrt{k} \cdot \frac{x(\log\log x)^k}{(\log x)^2 k!}.$$

[*Hint:* Use Chernoff's inequality (a.k.a. Rankin's trick).]

(c) Show an estimate analogous to the one in part (a) for the sum

$$\sum_{\substack{x-y < p \leqslant x \\ p \equiv 1+2a \,(\mathrm{mod}\, 2q)}} f(p-1)$$

when $(a(1+2a), q) = 1$ and $y/q \geqslant x^{\varepsilon}$ for some fixed $\varepsilon > 0$.

Chapter 21

Selberg's sieve

In 1947, Selberg introduced a different approach to sieving based on the simple fact that squares are non-negative. This allowed him to construct in one stroke a very general class of upper bound sieves often called Λ^2-sieves.

We start with a set of primes $\mathcal{P}$ and a function $\lambda : \mathbb{N} \to \mathbb{R}$ that is supported on integers $d|\mathcal{P}$ and satisfies the condition $\lambda(1) = 1$. Then

$$1_{(n,\mathcal{P})=1} \leqslant \Big(\sum_{d|n} \lambda(d)\Big)^2. \tag{21.1}$$

Indeed, if $(n, \mathcal{P}) = 1$, then both sides of (21.1) equal 1; otherwise, the left side is 0 whereas the right one is non-negative.

Opening the square in (21.1), we find that the right side equals $(1 * \lambda^+)(n)$, where

$$\lambda^+(d) = \sum_{[d_1,d_2]=d} \lambda(d_1)\lambda(d_2).$$

Hence, λ^+ is an upper bound sieve for the set of primes $\mathcal{P}$. If, in addition, we assume that $\mathrm{supp}(\lambda) \subset [1, \sqrt{D}]$, then $\mathrm{supp}(\lambda^+) \subset [1, D]$. We denote this special class of upper bound sieves λ^+ by $\Lambda^2(D, \mathcal{P})$. Lower bound sieves can also be obtained using Exercise 19.2.

We have thus produced a general class of sieve weights λ^+ for which the inequality $(1 * \lambda^+)(n) \geqslant 1_{(n,\mathcal{P})=1}$ is automatically satisfied. Optimizing the choice of λ^+ then becomes a calculus problem. Indeed, using Axiom 1 and the argument leading to (19.9), we find that

$$S(\mathcal{A}, \mathcal{P}) \leqslant X \sum_{d_1,d_2} \frac{\lambda(d_1)\lambda(d_2)\nu([d_1, d_2])}{[d_1, d_2]} + \sum_{d_1,d_2} \lambda(d_1)\lambda(d_2) r_{[d_1,d_2]}. \tag{21.2}$$

Ignoring the error term for now, we focus on minimizing the main term

$$Q := \sum_{d_1,d_2} \frac{\lambda(d_1)\lambda(d_2)\nu([d_1,d_2])}{[d_1,d_2]}.$$

This is a quadratic form in the variables $\lambda(d)$ with $d \leqslant \sqrt{D}$ and $d|\mathcal{P}$, under the restriction that $\lambda(1) = 1$.

Selberg's solution to this minimization problem was to diagonalize Q, since then finding the optimal choice of λ becomes trivial. To motivate his argument, let us consider first the special case when $\nu = 1$. We then have

$$Q = \sum_{d_1,d_2} \frac{\lambda(d_1)\lambda(d_2)}{[d_1,d_2]} = \sum_{d_1,d_2} \frac{\lambda(d_1)\lambda(d_2)}{d_1 d_2}(d_1,d_2). \tag{21.3}$$

A natural thing to do next is to set $m = (d_1, d_2)$, so that $d_j = ma_j$ with $(a_1, a_2) = 1$. We then find that

$$Q = \sum_m \frac{1}{m} \sum_{(a_1,a_2)=1} \frac{\lambda(ma_1)\lambda(ma_2)}{a_1 a_2}.$$

The problem is that the variables a_1 and a_2 on the right-hand side are tangled via the condition $(a_1, a_2) = 1$. If we did not have this condition, the double sum would factor as a perfect square thus diagonalizing Q.

We could replace the condition $(a_1, a_2) = 1$ using the Möbius inversion formula $1_{(a_1,a_2)=1} = \sum_{d|a_1,a_2} \mu(d)$. Instead, we use a trick that untangles a_1 and a_2 in a simpler way: we go back to (21.3) and rewrite its right-hand side using the convolution identity

$$(d_1,d_2) = \sum_{m|(d_1,d_2)} \varphi(m) = \sum_{m|d_1,d_2} \varphi(m).$$

Together with the change of variables $d_j = ma_j$, this implies that

$$Q = \sum_m \frac{\varphi(m)}{m^2} \sum_{a_1,a_2} \frac{\lambda(ma_1)\lambda(ma_2)}{a_1 a_2} = \sum_m \frac{\varphi(m)}{m^2} \Big(\sum_a \frac{\lambda(ma)}{a} \Big)^2$$

Setting $\xi(m) = \sum_a \lambda(ma)/a$ diagonalizes Q, which allows us to minimize it easily in terms of the new variables ξ.

We now generalize the above idea to arbitrary functions ν. First of all, note that we may assume that λ is supported on integers $d|\mathcal{P}'$ with

$$\mathcal{P}' := \{\, p \in \mathcal{P} : \nu(p) > 0 \,\}.$$

This restriction is justified by simply observing that $\nu([d_1, d_2]) = 0$ whenever either d_1 or d_2 is a square-free integer with at least one prime factor from $\mathcal{P} \setminus \mathcal{P}'$.

Now, for any choice of $d_1, d_2 | \mathcal{P}'$, the number $[d_1, d_2]$ is square-free. Hence,

$$\frac{\nu([d_1,d_2])}{[d_1,d_2]} = \frac{\nu(d_1)\nu(d_2)}{d_1 d_2} \cdot \frac{(d_1,d_2)}{\nu((d_1,d_2))} = \frac{\nu(d_1)\nu(d_2)}{d_1 d_2} \sum_{m | d_1, d_2} \frac{\varphi^*(m)}{\nu(m)},$$

where[1]

$$\varphi^*(m) := m \prod_{p|m} \Big(1 - \frac{\nu(p)}{p}\Big). \tag{21.4}$$

We then find that

$$\begin{aligned} Q &= \sum_{d_1,d_2} \lambda(d_1)\lambda(d_2) \cdot \frac{\nu(d_1)\nu(d_2)}{d_1 d_2} \sum_{m|d_1,d_2} \frac{\varphi^*(m)}{\nu(m)} \\ &= \sum_m \frac{\nu(m)\varphi^*(m)}{m^2} \Big(\sum_a \frac{\lambda(ma)\nu(a)}{a} \Big)^2, \end{aligned}$$

where we let $d_j = ma_j$ for each j. It is now clear that making the change of variables

$$\xi(m) = 1_{m \in \mathscr{D}} \sum_a \frac{\lambda(ma)\nu(a)}{a} \qquad \text{with} \qquad \mathscr{D} := \{\, d \leqslant \sqrt{D} : d | \mathcal{P}' \,\}.$$

diagonalizes Q. We need to show that this is an invertible change of variables, which we accomplish by an application of Möbius inversion. Since λ is also supported on $\mathscr{D}$, for each d we have

$$\begin{aligned} \sum_{m \in \mathscr{D},\, d|m} \frac{\mu(m/d)\nu(m)\xi(m)}{m} &= \sum_{m \in \mathscr{D},\, d|m} \mu(m/d) \cdot \frac{\nu(m)}{m} \sum_a \frac{\lambda(ma)\nu(a)}{a} \\ &\overset{n=ma}{=} \sum_{n \in \mathscr{D},\, d|n} \frac{\lambda(n)\nu(n)}{n} \sum_{m:\ d|m|n} \mu(m/d). \end{aligned}$$

Making the change of variables $m = de$ in the innermost sum, we find that it equals $\sum_{e|n/d} \mu(e) = 1_{n=d}$. We thus arrive at the inversion formula

$$\frac{\lambda(d)\nu(d)}{d} = \sum_{m \in \mathscr{D},\, d|m} \frac{\mu(m/d)\nu(m)\xi(m)}{m}. \tag{21.5}$$

This proves our claim about the invertibility of our change of variables.

Recall our constraint $\lambda(1) = 1$ which, in view of (21.5), becomes

$$\sum_{m \in \mathscr{D}} \frac{\mu(m)\nu(m)\xi(m)}{m} = 1. \tag{21.6}$$

We have thus transformed our task to minimizing the quadratic form $Q = \sum_m \nu(m)\varphi^*(m)\xi(m)^2/m^2$ under condition (21.6), with ξ supported on $\mathscr{D}$.

[1]Note that $\nu/\varphi^* = \delta^*$, where δ^* is the multiplicative function we saw in Exercise 19.1.

We can solve the above minimization problem using Lagrange multipliers. Alternatively, we can use the Cauchy-Schwarz inequality: applying it with coefficients $(\nu(m)\varphi^*(m))^{1/2}\xi(m)/m$ and $1_{m\in\mathscr{D}}\mu(m)(\nu(m)/\varphi^*(m))^{1/2}$, we find that

$$1 = \Big(\sum_{m\in\mathscr{D}} \frac{\mu(m)\nu(m)\xi(m)}{m}\Big)^2 \leqslant Q\sum_{m\in\mathscr{D}} \frac{\mu^2(m)\nu(m)}{\varphi^*(m)}.$$

We know that the above inequality is an equality exactly when there is a constant $L\neq 0$ such that

$$\xi(m) = \frac{1}{L}\cdot 1_{m\in\mathscr{D}}\cdot\frac{\mu(m)m}{\varphi^*(m)} \tag{21.7}$$

for all m. To calculate the value of L, we use (21.6). This yields $L = \sum_{m\in\mathscr{D}}\nu(m)/\varphi^*(m)$, where we used that $\mathscr{D}$ contains only square-free integers (so that $\mu^2(m)=1$ for each $m\in\mathscr{D}$). The minimal value of Q is thus

$$Q = \frac{1}{L^2}\sum_{m\in\mathscr{D}}\frac{\nu(m)}{\varphi^*(m)} = \frac{1}{L}.$$

The above calculations lead us to the following fundamental result.

Theorem 21.1. *Let $\mathcal{A}$ and $\mathcal{P}$ satisfy Axiom 1. If $D\geqslant 1$ and φ^* is defined by* (21.4)*, then*

$$S(\mathcal{A},\mathcal{P}) \leqslant \frac{X}{L} + \sum_{\substack{d\leqslant D\\ d|\mathcal{P}}} 3^{\omega(d)}|r_d| \qquad \textit{with} \qquad L = \sum_{\substack{m\leqslant\sqrt{D}\\ m|\mathcal{P}}}\frac{\nu(m)}{\varphi^*(m)}.$$

Proof. Let ξ be defined by (21.7), where we recall that $\mathscr{D} = \{\, d\leqslant\sqrt{D} : d|\mathcal{P}'\,\}$ with $\mathcal{P}' = \{\, p\in\mathcal{P} : \nu(p)>0\,\}$. Then, we define $\lambda(d)$ via relation (21.5) when $d\in\mathscr{D}$, whereas we set $\lambda(d)=0$ otherwise. We claim that

$$|\lambda(d)|\leqslant 1 \quad \text{whenever} \quad d\in\mathscr{D}. \tag{21.8}$$

Before proving this inequality, let us see how it establishes the theorem.

Indeed, the first term on the right-hand side of (21.2) equals X/L by the discussion preceding Theorem 21.1, whereas the second term is $\leqslant \sum_{d_1,d_2\in\mathscr{D}}|r_{[d_1,d_2]}|$ by virtue of (21.8). Given a square-free integer d, there are $3^{\omega(d)}$ ways to write it as $d=[d_1,d_2]$. This completes the proof, assuming the validity of (21.8).

To prove (21.8), we first calculate $\lambda(d)$. For any $d \in \mathscr{D}$, relation (21.5) and our choice of ξ imply that

$$\begin{aligned}\lambda(d) &= \frac{d}{\nu(d)} \sum_{m\in\mathscr{D},\, d|m} \frac{\mu(m/d)\nu(m)\xi(m)}{m} \\ &= \frac{d}{L\nu(d)} \sum_{m\in\mathscr{D},\, d|m} \frac{\mu(m/d)\mu(m)\nu(m)}{\varphi^*(m)} \\ &\overset{m=da}{=} \frac{\mu(d)}{L} \cdot \frac{d}{\varphi^*(d)} \sum_{a:\, da\in\mathscr{D}} \frac{\nu(a)}{\varphi^*(a)}.\end{aligned}$$

For each $d \in \mathscr{D}$ (that is necessarily square-free), we have the formula $d/\varphi^*(d) = \sum_{b|d} \nu(b)/\varphi^*(b)$. We thus find that

$$\lambda(d) = \frac{\mu(d)}{L} \sum_{a:\, da\in\mathscr{D}} \sum_{b|d} \frac{\nu(a)}{\varphi^*(a)} \cdot \frac{\nu(b)}{\varphi^*(b)}.$$

The products ab with a and b as above are all distinct from each other, with each one of them determining a unique integer $n \in \mathscr{D}$. Since $\nu/\varphi^* \geqslant 0$, we conclude that

$$\sum_{a:\, da\in\mathscr{D}} \sum_{b|d} \frac{\nu(a)}{\varphi^*(a)} \cdot \frac{\nu(b)}{\varphi^*(b)} \leqslant \sum_{n\in\mathscr{D}} \frac{\nu(n)}{\varphi^*(n)} = L.$$

This proves our claim (21.8), thus completing the proof of the theorem. □

The sum L from the statement of Theorem 21.1 can be estimated asymptotically using Theorem 14.3 and Axiom 2′. The resulting bound for $S(\mathcal{A}, \mathcal{P})$ is given below.

Theorem 21.2. *Let $\mathcal{A}$ and $\mathcal{P} \subseteq \{p \leqslant y\}$. Assume that Axioms 1, 2′ and 3 hold, the second one with parameters κ, k and ε, and the third one with $m = 3$, $A = \kappa + 1$ and level of distribution $D \geqslant y^2$. If, in addition, $\log X \gg \log y$, then*

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

Proof. All implied constants might depend on κ, k and ε. By Theorem 21.1 with $D = y^2$ and our assumptions on $(\mathcal{A}, \mathcal{P})$, it suffices to show that

$$L := \sum_{\substack{m\leqslant y \\ m|\mathcal{P}}} \frac{\nu(m)}{\varphi^*(m)} = \frac{e^{-\kappa\gamma}}{\Gamma(\kappa+1)} \prod_{p\in\mathcal{P}} \left(1 - \frac{\nu(p)}{p}\right)^{-1} + O((\log y)^{\kappa-1}), \tag{21.9}$$

Indeed, since $\prod_{p\in\mathcal{P}}(1-\nu(p)/p)^{-1} \gg (\log y)^\kappa$ from Axiom 2′, inverting (21.9) yields $L^{-1} = (e^{\kappa\gamma}/\Gamma(\kappa+1) + O(1/\log y)) \prod_{p\in\mathcal{P}}(1-\nu(p)/p)$ as needed.

To prove (21.9), we define the multiplicative function f by the relation

$$f(p^m) = \begin{cases} 1_{p\in\mathcal{P}} \cdot 1_{m=1} \cdot \nu(p)/(1-\nu(p)/p) & \text{if } p \leqslant y, \\ \tau_\kappa(p^m) & \text{if } p > y, \end{cases}$$

where τ_κ is defined by (13.3). In particular, $L = \sum_{m\leqslant y} f(m)/m$. We evaluate this sum using Theorem 14.3, whose conditions hold for f by Axiom 2′ and Mertens' estimates (Theorem 3.4(c)). Hence,

$$S = \frac{(\log y)^\kappa}{\Gamma(\kappa+1)} \prod_p \left(1 + \frac{f(p)}{p} + \frac{f(p^2)}{p^2} + \cdots\right)\left(1-\frac{1}{p}\right)^\kappa + O((\log y)^{\kappa-1}).$$

The factors with $p > y$ are all equal to 1. On the other hand, we have

$$\prod_{p\leqslant y}\left(1 + \frac{f(p)}{p} + \frac{f(p^2)}{p^2} + \cdots\right) = \prod_{p\in\mathcal{P}}\left(1-\frac{\nu(p)}{p}\right)^{-1}.$$

Using Mertens' third estimate to evaluate $\prod_{p\leqslant y}(1-1/p)^\kappa$ completes the proof of (21.9) and hence of the theorem. □

As a direct corollary of Theorem 21.2 with $\mathcal{P} = \{p \leqslant \sqrt{x}/(\log x)^{2k}\}$, we have an estimate for the number of prime values of an admissible k-tuple. (See Exercise 17.4 for the definition of an admissible k-tuple.)

Corollary 21.3. *Let $\mathbf{h} = (h_1, \dots, h_k)$ be an admissible k-tuple of distinct integers, and define $\nu_{\mathbf{h}}(p) = \#\{\, h_j \pmod p : 1 \leqslant j \leqslant k \,\}$. Then*

$$\#\{\, n \leqslant x : n+h_1, \dots, n+h_k \text{ are all primes}\,\} \leqslant (2^k k! + \varepsilon)\frac{\mathfrak{S}(\mathbf{h})x}{(\log x)^k}$$

with $\varepsilon = O_{k,\mathbf{h}}(\log\log x/\log x)$ and

$$\mathfrak{S}(\mathbf{h}) = \prod_p \left(1-\frac{\nu_{\mathbf{h}}(p)}{p}\right)\left(1-\frac{1}{p}\right)^{-k}.$$

This result should be compared with the Hardy-Littlewood conjecture (17.14). In particular, taking $h_1 = 0$ and $h_2 = 2$, Corollary 21.3 implies that the number of twin primes is at most 8 times the expected amount. In Exercise 21.2 we will see an improvement of this result when $k = 2$.

Our final application of Selberg's sieve is an explicit version of the Brun-Titchmarsch inequality which shows that the number of primes $\leqslant x$ in the progression $a \pmod q$ is at most $2+\varepsilon$ times the expected number when $\log x/\log q \to \infty$. In Exercise 22.2, we will see that improving this factor to $2-\varepsilon$ would imply that there are no exceptional zeroes in Theorem 12.3. However, this cannot be done using only Axioms 1–3 and their variations, as we discuss after the proof of Theorem 21.4.

Theorem 21.4 (The Brun-Titchmarsch inequality, II)**.** *Uniformly for $q \in \mathbb{N}$, $a \in (\mathbb{Z}/q\mathbb{Z})^*$ and $x \geqslant y \geqslant q$, we have*

$$\#\{\, x-y < p \leqslant x : p \equiv a \,(\mathrm{mod}\, q)\,\} \leqslant \frac{(2+\varepsilon)y}{\varphi(q)\log(2y/q)}$$

with $\varepsilon = O(\log\log(3y/q)/\log(y/q))$.

Proof. There are two ways to obtain the claimed upper bound, both involving a trick to get the required uniformity in q. The more standard proof is quite similar to the proof of Theorem 20.1, and it is outlined in Exercise 21.1. We present here an alternative proof that uses the monotonicity principle (see Exercise 19.1).

Let $\mathcal{A} = \{\, x-y < n \leqslant x : n \equiv a \,(\mathrm{mod}\, q)\,\}$ and $\mathcal{P} = \{p \leqslant z\}$ with z to be determined. As in the proof of Theorem 20.1, the pair $(\mathcal{A}, \mathcal{P})$ satisfies Axiom 1 with $X = y/q$, $\nu(d) = 1_{(d,q)=1}$ and $|r_d| \leqslant 2$. Hence, for any function λ supported on square-free integers $\leqslant z$, (21.2) implies that

$$\begin{aligned}\#\{\, x-y < p \leqslant x : p \equiv a \,(\mathrm{mod}\, q)\,\} &\leqslant S(\mathcal{A},\mathcal{P}) + z \\ &\leqslant \frac{y}{q} \sum_{(d_1d_2,q)=1} \frac{\lambda(d_1)\lambda(d_2)}{[d_1,d_2]} + 2z^2\|\lambda\|_\infty^2,\end{aligned}$$

where $\|\lambda\|_\infty$ is the supremum norm of λ. The sum over d_1, d_2 in the main term can be written as

$$\sum_m \frac{\lambda^+(m) 1_{(m,q)=1}}{m},$$

where $\lambda^+(m) = \sum_{[d_1,d_2]=m} \lambda(d_1)\lambda(d_2)$ is an upper bound sieve. Hence, Exercise 19.1(b) with $\delta_1(m) = 1_{(m,q)=1}/m$ and $\delta_2(m) = 1/m$ implies that

$$\begin{aligned}0 \leqslant \sum_{(d_1d_2,q)=1} \frac{\lambda(d_1)\lambda(d_2)}{[d_1,d_2]} &\leqslant \prod_{p\leqslant z,\, p\mid q} \Big(1-\frac{1}{p}\Big)^{-1} \sum_{d_1,d_2} \frac{\lambda(d_1)\lambda(d_2)}{[d_1,d_2]} \\ &\leqslant \frac{q}{\varphi(q)} \sum_{d_1,d_2} \frac{\lambda(d_1)\lambda(d_2)}{[d_1,d_2]}.\end{aligned}$$

We now choose the weights λ to be the optimal weights with respect to the function $\nu(d) = 1$ and the set of primes $\mathcal{P} = \{p \leqslant z\}$, in which case $\|\lambda\|_\infty \leqslant 1$ by (21.8) and

$$\sum_{d_1,d_2} \frac{\lambda(d_1)\lambda(d_2)}{[d_1,d_2]} = 1 \Big/ \sum_{m\leqslant z} \frac{\mu^2(m)}{\varphi(m)},$$

since $\varphi^* = \varphi$ here. By Theorem 14.3, we have $\sum_{m\leqslant z} \mu^2(m)/\varphi(m) = \log z + O(1)$. We thus conclude that

$$\#\{\, x-y < p \leqslant x : p \equiv a \,(\mathrm{mod}\, q)\,\} \leqslant \frac{y}{\varphi(q)(\log z + O(1))} + O(z^2).$$

Taking $z = (y/q)^{1/2}(\log(2y/q))^{-100}$ completes the proof. □

The parity problem of sieve methods

As we discussed above, improving the constant 2 in Theorem 21.4 would have the spectacular consequence of eliminating Landau-Siegel zeroes. However, Selberg proved that this is not possible using sieve methods and the mere assumption of Axioms 1–3 and their variations. To do so, he constructed sets $\mathcal{A}$ that satisfy Axioms 1–3 and for which the true size of $S(\mathcal{A}, \mathcal{P})$ matches the upper bound provided by Theorem 21.2.

Indeed, let $\mathcal{P} = \{p \leqslant \sqrt{x}\}$,

$$\mathcal{A}^{(1)} = \{\, n \leqslant x : \Omega(n) \text{ is odd} \,\} \quad \text{and} \quad \mathcal{A}^{(0)} = \{\, n \leqslant x : \Omega(n) \text{ is even} \,\}.$$

Note that

$$\begin{aligned} A_d^{(j)} &= \sum_{\substack{n \leqslant x \\ d \mid n}} \frac{1 + (-1)^{j+\Omega(n)}}{2} \\ &= \frac{\lfloor x/d \rfloor}{2} + \frac{(-1)^{j+\Omega(d)}}{2} \sum_{m \leqslant x/d} (-1)^{\Omega(m)} \\ &= \frac{x}{2d} + O\Big(\frac{x}{d} e^{-c\sqrt{\log(x/d)}}\Big) \end{aligned}$$

for some absolute constant $c > 0$, where the error is bounded using Exercise 8.4(d) and a convolution trick (i.e., we write $(-1)^{\Omega} = \mu * f$). Theorem 13.2 can also be used if we settle for a weaker error term.

The above estimate implies that Axiom 1 is satisfied with $X = x/2$ and $\nu(d) = 1$. In addition, Axiom 2′ holds with $\kappa = k = 1$, whereas Axiom 3 holds with $m = 3$, $A = 2$ and $D = x/e^{(\log\log x)^3}$. We may thus apply Theorem 21.2 to conclude that

$$0 \leqslant S(\mathcal{A}^{(j)}, \sqrt{x}) \leqslant (1 + o(1)) \frac{x/2}{\log \sqrt{x}} = (1 + o(1)) \frac{x}{\log x} \tag{21.10}$$

as $x \to \infty$.

On the other hand, we can calculate $S(\mathcal{A}^{(j)}, \sqrt{x})$ directly. We know that any integer $n > \sqrt{x}$ counted by it must be prime. Therefore

$$S(\mathcal{A}^{(1)}, \sqrt{x}) = \pi(x) + O(\sqrt{x}) \sim \frac{x}{\log x}$$

as $x \to \infty$ from the Prime Number Theorem, whereas

$$S(\mathcal{A}^{(0)}, \sqrt{x}) = O(\sqrt{x}) = o\Big(\frac{x}{\log x}\Big).$$

So we see that, up to an error of size $o(x/\log x)$, the upper bound in (21.10) is sharp when $j = 1$, and the lower bound is sharp when $j = 0$. In particular, we cannot hope to improve upon (21.10) unless we impose an extra axiom that eliminates the above examples.

Because of the shape of Selberg's extremal examples, the inability to improve upon (21.10) under Axioms 1–3 is called the *parity problem of sieve methods*. The underlying reason that causes this obstruction is that the sieve weights we have constructed under Axioms 1–3 do not correlate with the Möbius function, so they cannot differentiate between integers with an even and an odd number of prime factors.

In Chapter 23, we will present a method going back to Vinogradov that allows us to "break the parity barrier" for certain sequences $(a_n)_{n=1}^{\infty}$ using bilinear form methods. The book of Harman [96] and the paper of Friedlander and Iwaniec [56] study such "parity-breaking sieves" in a much more systematic way. Let us briefly mention that the axiom imposed on $\mathcal{A}$ in [56] concerns (roughly) bilinear sums of the form $\sum_{k\leqslant K,\,\ell\leqslant L} a_{k\ell}\mu(k\ell)$. Showing that there is cancellation in such sums means that the sequence $\mathcal{A}$ does not correlate with the Möbius function.

Remark 21.5. In light of the above discussion, Chen's theorem [24,25] that there are infinitely many primes p such that $p+2$ is the product of at most two primes is the best possible result we can hope for using sieve methods and the mere assumption of Axioms 1–3. □

Exercises

Exercise 21.1. Let $q\in\mathbb{N}$ and $z\geqslant 1$.

(a) Note that $q/\varphi(q)=\sum_{d|q}\mu^2(d)/\varphi(d)$ and $\mu^2(m)/\varphi(m)=\prod_{p|m}(1+1/p+1/p^2+\cdots)$. Conclude that

$$\frac{q}{\varphi(q)}\sum_{m\leqslant z,\,(m,q)=1}\frac{\mu^2(m)}{\varphi(m)}\geqslant\sum_{m\leqslant z}\frac{\mu^2(m)}{\varphi(m)}\geqslant\sum_{m\leqslant z}\frac{1}{m}\geqslant\log(\lfloor z\rfloor+1).$$

(b) Use the above inequalities to prove that

$$\#\{\,x-y<p\leqslant x : p\equiv a\,(\mathrm{mod}\,q)\,\}\leqslant\frac{y}{\varphi(q)\log(\lfloor z\rfloor+1)}+2z^2+z$$

for all $x\geqslant y\geqslant 1$, and deduce Theorem 21.4.

Exercise 21.2. Let h be an even integer, and let $L_h=\sum_{p|h}(\log p)/p$.

(a) Show that $L_h\leqslant\log(\omega(h))+O(1)$.

(b) Show that there are two absolute constants M_1 and M_2 such that

$$\#\{\,p\leqslant x : p+h\text{ prime}\,\}\leqslant\left(1+\frac{M_1(L_h+\log\log x)}{\log x}\right)\frac{4c_2x}{(\log x)^2}\prod_{p|h,\,p>2}\frac{p-1}{p-2}$$

when $x\geqslant e^{M_2L_h}$, where c_2 denotes the twin prime constant as usual. [*Hint:* Use Exercise 14.7.]

Chapter 22

Sieving for zero-free regions

Sieve methods provide an alternative way of establishing zero-free regions for Dirichlet series that are of the same strength as Theorems 8.3 and 12.3. The idea is as follows: assume that we want to obtain a zero-free region for ζ close to $1+it$, where $|t| \geqslant 2$. For any $\sigma \in (0,1)$, we have

$$\zeta(\sigma+it) = \zeta(1+it) - \int_{\sigma}^{1} \zeta'(\alpha+it)\mathrm{d}\alpha.$$

Now, let $\delta \in (0,1)$ be such that the quantity $M := \sup_{1-\delta\leqslant\sigma\leqslant 1} |\zeta'(\sigma+it)|$ satisfies the inequality $\delta M \leqslant |\zeta(1+it)|/2$. We then infer that $1/2 \leqslant |\zeta(\sigma+it)|/|\zeta(1+it)| \leqslant 3/2$ for $\sigma \in [1-\delta, 1]$. Hence, we have reduced proving a zero-free region to an upper bound for $\zeta'(\sigma+it)$ and a lower bound for $\zeta(1+it)$, so that we can determine the largest δ for which $\delta M \leqslant |\zeta(1+it)|/2$.

To bound ζ', we argue as in Theorem 11.2, but we need to be a bit more careful because the Dirichlet series representation of ζ is not valid inside the critical strip. We apply the Euler-Maclaurin summation formula to obtain a generalization of (5.7): for each $N \in \mathbb{Z}_{\geqslant 1}$ and for $\mathrm{Re}(s) > 1$, we have

$$\zeta(s) = \sum_{n=1}^{N} \frac{1}{n^s} + \sum_{n>N} \frac{1}{n^s} = \sum_{n=1}^{N} \frac{1}{n^s} + \frac{N^{1-s}-1}{1-s} - s\int_{N}^{\infty} \frac{\{y\}}{y^{s+1}}\mathrm{d}y. \tag{22.1}$$

Now, both $\zeta(s)$ and the rightmost expression in (22.1) are well-defined for $\mathrm{Re}(s) > 0$. In addition, they share the same singularities in this region (a simple pole of residue 1 at $s=1$). Since they are equal for $\mathrm{Re}(s) > 1$, they must also be equal for $\mathrm{Re}(s) > 0$ by the identity principle. Differentiating

yields the formula

$$\zeta'(s) = -\sum_{n=1}^{N} \frac{\log n}{n^s} + \frac{N^{1-s}-1}{(1-s)^2} - \frac{N^{1-s}\log N}{1-s} - \int_N^\infty \frac{\{y\}(1-s\log y)}{y^{s+1}}\mathrm{d}y,$$

valid for all $\mathrm{Re}(s) > 0$ and all $N \in \mathbb{Z}_{\geqslant 1}$. We take $N = \lfloor |t| \rfloor$, put absolute values everywhere and argue as in the proof of Lemma 11.2 to find that $\zeta'(s) = O(\log^2 |t|)$ for $\sigma > 1 - 1/\log|t|$ and $|t| \geqslant 2$.

On the other hand, we have $|\log \zeta(1+it)| \leqslant \log\log|t| + O(1)$ by Exercise 8.4(c), whence $|\zeta(1+it)| \gg 1/\log|t|$. This leads to a zero-free region of the form $\sigma \geqslant 1 - O(1/\log^3|t|)$ which is weaker than the one in Theorem 8.3.

To understand why we arrived at a weaker zero-free region, we must reexamine the above argument. Notice that to bound $\zeta'(s)$ we estimated trivially all the summands with $n \leqslant |t|$. In fact, for the upper bound $\zeta'(s) = O(\log^2|t|)$ to be achieved, we must have that $n^{it} \approx 1$ for all $n \leqslant |t|$. But then $p^{it} \approx 1$ for all primes $p \leqslant |t|$, so that $\prod_{p\leqslant|t|} |1 - 1/p^{1+it}|^{-1} \approx \log|t|$. This suggests we should be able to replace the lower bound $|\zeta(1+it)| \gg 1/\log|t|$ by the stronger estimate $|\zeta(1+it)| \gg \log|t|$, which would then lead us to a zero-free region of the same strength as the one in Theorem 8.3.

Sifted Dirichlet series

To get around the above issue, we introduce a truncated version of ζ. Instead of truncating in an archimedean way and considering the sum $\sum_{n>N} 1/n^s$, we truncate it multiplicatively and work with

$$\zeta_y(s) := \sum_{P^-(n)>y} \frac{1}{n^s} = \prod_{p>y} \left(1 - \frac{1}{p^s}\right)^{-1}.$$

This "multiplicatively truncated" version of ζ has the advantage of possessing an Euler product representation. More generally, given a Dirichlet series $F(s) = \sum_{n=1}^\infty f(n)/n^s$, we set

$$F_y(s) = \sum_{P^-(n)>y} \frac{f(n)}{n^s}. \tag{22.2}$$

When $F(s)$ is a Dirichlet L-function, we have the following crucial estimate.

Theorem 22.1. *Let χ be a Dirichlet character* $\bmod q$*, $s = \sigma + it$ and $y \in \mathbb{R}$ such that $y \geqslant \max\{10, q(|t|+1)\}$ and $\sigma \geqslant 1 - 1/\log y$. If χ is principal, we further assume that $|t| \geqslant 1/\log y$. For $j \in \{0,1\}$, we have*

$$L_y^{(j)}(s,\chi) \ll (\log y)^j. \tag{22.3}$$

Before embarking on the proof of Theorem 22.1, let us see how we can use it to extract a zero-free region for $L(s,\chi)$. Set

$$q_{\chi,t} := \max\{q(|t|+1), 10, \exp(1_{\chi=\chi_0}/|t|)\}. \tag{22.4}$$

When $y \geqslant q_{\chi,t}$, we also have $|t| \geqslant 1_{\chi=\chi_0}/\log y$. Thus, Theorem 22.1 implies that $L_y'(\sigma+it,\chi) \ll \log y$ for $\sigma \geqslant 1 - 1/\log y$. As a consequence,

$$L_y(\sigma+it,\chi) - L_y(1+it,\chi) = \int_1^\sigma L_y'(\alpha+it,\chi)\mathrm{d}\alpha \ll |1-\sigma|\log y.$$

We thus infer that there is an absolute constant $c>0$ such that

$$|L_y(\sigma+it,\chi)| \asymp |L_y(1+it,\chi)| \quad \text{for} \quad |\sigma-1| \leqslant c\cdot\frac{|L_y(1+it,\chi)|}{\log y}. \tag{22.5}$$

In particular, we see that the size of $|L_y(1+it,\chi)|$ controls the quality of the zero-free region we can obtain. Moreover, if the upper bound $|L_y(1+it,\chi)| \leqslant O(1)$ from Theorem 22.1 is the true order of magnitude of $|L_y(1+it,\chi)|$, then we recover the zero-free region for $L(s,\chi)$ given in Theorem 12.3.

We now prove Theorem 22.1. After this task has been completed, we will see how we can control the size of $L_y(s,\chi)$.

Sifted character sums

The key to proving Theorem 22.1 is the Fundamental Lemma of Sieve Theory, which allows us to estimate character sums running over y-rough numbers. Going from such an estimate to a bound for $L_y(s,\chi)$ is then accomplished by a routine partial summation argument.

Lemma 22.2. *For $t\in\mathbb{R}$ and $x\geqslant y\geqslant \max\{q(|t|+1),10\}^{100}$, we have*

$$\sum_{\substack{n\leqslant x\\ P^-(n)>y}} \chi(n)n^{it} = 1_{\chi=\chi_0}\cdot\frac{x^{1+it}}{1+it}\prod_{p\leqslant y}\Big(1-\frac{1}{p}\Big) + O\Big(\frac{x^{1-110/\log y}}{\log y}\Big).$$

Proof. Let $\lambda^\pm$ be the sieve weights from Theorem 19.1 with $D=\sqrt{x}$, $\mathcal{P}=\{p\leqslant y\}$ and $\kappa=1$. Set $\delta=\lambda^+-\lambda^-$ and note that $\delta*1\geqslant 0$, as well as that $(\lambda^+*1)(n)-(\delta*1)(n)\leqslant 1_{P^-(n)>y}\leqslant(\lambda^+*1)(n)$. Consequently,

$$\sum_{\substack{n\leqslant x\\ P^-(n)>y}}\chi(n)n^{-it} = \sum_{n\leqslant x}\chi(n)n^{-it}(\lambda^+*1)(n) + O\Big(\sum_{n\leqslant x}(\delta*1)(n)\Big). \tag{22.6}$$

For the error term, we have

$$\sum_{n\leqslant x}(\delta*1)(n) = \sum_{m\leqslant\sqrt{x}}\delta(m)\Big\lfloor\frac{x}{m}\Big\rfloor = x\sum_{m\leqslant\sqrt{x}}\frac{\lambda^+(m)-\lambda^-(m)}{m} + O(\sqrt{x}).$$

Since $\log D/\log y = 0.5\log x/\log y$ here, Theorem 19.1 yields the estimate

$$\sum_{n\leqslant x}(\delta * 1)(n) \ll x^{1-110/\log y}/\log y. \tag{22.7}$$

For the main term, we have

$$\sum_{n\leqslant x}\chi(n)n^{-it}(\lambda^+ * 1)(n) = \sum_{d\leqslant\sqrt{x}}\lambda^+(d)\chi(d)d^{-it}\sum_{m\leqslant x/d}\chi(m)m^{-it}.$$

We apply partial summation to remove the factor m^{-it} from the inner sum. Note that $\sum_{n\leqslant w}\chi(n) = 1_{\chi=\chi_0}w\varphi(q)/q + O(q)$, by periodicity. Hence,

$$\sum_{m\leqslant w}\chi(m)m^{-it} = 1_{\chi=\chi_0}\frac{\varphi(q)}{q}\cdot\frac{w^{1-it}}{1-it} + O((|t|+1)q\log(2w))$$

uniformly for $w \geqslant 1$. In turn, this implies that

$$\begin{aligned}\sum_{n\leqslant x}\chi(n)n^{-it}(\lambda^+ * 1)(n) = 1_{\chi=\chi_0}\frac{\varphi(q)}{q}\cdot\frac{x^{1-it}}{1-it}\sum_{d\leqslant\sqrt{x}}\frac{\lambda^+(d)\chi_0(d)}{d}\\ + O(\sqrt{x}q(|t|+1)\log x).\end{aligned} \tag{22.8}$$

Since we have assumed that $x \geqslant y \geqslant \max\{q(|t|+1), 10\}^{100}$, we have

$$\sqrt{x}q(|t|+1)\log x \leqslant x^{0.51} \leqslant x^{1-1.1/\log(10)} \leqslant x^{1-110/\log y}.$$

In addition, note that

$$\frac{\varphi(q)}{q}\sum_{d\leqslant\sqrt{x}}\frac{\lambda^+(d)\chi_0(d)}{d} = (1+O(x^{-110/\log y}))\prod_{p\leqslant y}\Big(1-\frac{1}{p}\Big)$$

by Theorem 19.1, since $y \geqslant q$ here. Combining the above estimates with (22.6) and (22.7) completes the proof of the lemma. □

Proof of Theorem 22.1. Let χ and t satisfy the hypotheses of the theorem, and recall the definition of $q_{\chi,t}$ from (22.4). First, we prove (22.3) when $y \geqslant q_{\chi,t}^{100}$ and $\sigma \in [1-100/\log y, 2]$. From Lemma 22.2 we know that

$$\sum_{n\leqslant x}1_{P^-(n)>y}\chi(n)n^{-it} = \alpha\cdot\frac{x^{1-it}}{1-it} + R(x) \qquad \text{for all } x \geqslant y, \tag{22.9}$$

where $\alpha = 1_{\chi=\chi_0}\prod_{p\leqslant y}(1-1/p)$ and $R(x) = R_{\chi,t,y}(x) \ll x^{1-110/\log y}/\log y$. Hence, for all $w \in \mathbb{C}$ with $\mathrm{Re}(w) > 1$, partial summation implies that

$$\begin{aligned}L_y(w+it,\chi) &= 1 + \alpha\int_y^\infty\frac{\mathrm{d}x}{x^{w+it}} - \frac{R(y)}{y^w} + w\int_y^\infty\frac{R(x)}{x^{w+1}}\mathrm{d}x\\ &= 1 + \frac{\alpha}{w+it-1} - \alpha\int_1^y\frac{\mathrm{d}x}{x^{w+it}} - \frac{R(y)}{y^w} + w\int_y^\infty\frac{R(x)}{x^{w+1}}\mathrm{d}x\end{aligned} \tag{22.10}$$

In view of (22.9), the right-hand side of (22.10) is meromorphic for $\mathrm{Re}(w) > 1 - 110/\log y$. In addition, it has the same singularities as $L_y(w+it,\chi)$, i.e., a simple pole at $w = 1-it$ of residue α. Hence, (22.10) must hold for $\mathrm{Re}(w) > 1 - 110/\log y$. Differentiating it j times and setting $w = \sigma$, we infer that

$$(-1)^j L_y^{(j)}(s,\chi) = 1_{j=0} + \frac{\alpha}{(s-1)^2} - \alpha\int_1^y \frac{(\log x)^j \mathrm{d}x}{x^s} - \frac{R(y)(\log y)^j}{y^\sigma} + \int_y^\infty \frac{R(x)[\sigma(\log x)^j - j(\log x)^{j-1}]}{x^{\sigma+1}}\mathrm{d}x.$$

Since $R(x) \ll x^{1-110/\log y}/\log y$ and $\sigma \geqslant 1 - 100/\log y$, the two terms involving R on the right-hand side of the above identity are $\ll (\log y)^j$. If $\chi \neq \chi_0$, this proves (22.3) because $\alpha = 0$. Finally, if $|t| \geqslant c_1/\log y$, we note that $\alpha \ll 1/\log y$ and $|s-1| \geqslant |t| \geqslant c_1/\log y$, as well as $\int_1^y (\log x)^j x^{-s}\mathrm{d}x \ll (\log y)^j$. Putting together these estimates completes the proof of (22.3) in this case as well.

Finally, we prove (22.3) when $q_{\chi,t} \leqslant y \leqslant q_{\chi,t}^{100}$ and $\sigma \geqslant 1 - 1/\log y$. The case $\sigma \geqslant 2$ is trivial by the absolute convergence of $L^{(j)}(s,\chi)$. Assume now that $1 - 1/\log y \leqslant \sigma \leqslant 2$ and let $z = y^{100}$, so that $z \geqslant q_{\chi,t}^{100}$ and $\sigma \in [1 - 100/\log z, 2]$ Thus, the results we proved above apply with z in place of y, that is to say, $L_z(s,\chi) \ll 1$ and $L_z'(s,\chi) \ll \log z$. In addition, note that

$$|L_y(s,\chi)| = |L_z(s,\chi)| \prod_{y<p\leqslant z} |1-\chi(p)/p^s|^{-1} \asymp |L_z(s,\chi)| \ll 1, \tag{22.11}$$

where the product over p was bounded by observing that $1 - 1/p \leqslant |1 - \chi(p)/p^s| \leqslant 1 + 1/p$ and then applying Mertens' third estimate (Theorem 3.4(c)). Similarly, we have

$$\begin{aligned} L_y'(s,\chi) &= L_z'(s,\chi) \prod_{y<p\leqslant z} \Big(1 - \frac{\chi(p)}{p^s}\Big)^{-1} - L_y(s,\chi) \sum_{y<p\leqslant z}\sum_{m\geqslant 1} \frac{\chi(p^m)\log p}{p^{ms}} \\ &\ll |L_z'(s,\chi)| + |L_y(s,\chi)|\log z \ll \log y, \end{aligned}$$

where we used Mertens' second and third estimates to obtain the first inequality. This completes the proof of Theorem 22.1. □

Pretentious multiplicative functions

Relation (22.5) reduces the proof of a zero-free region for $L(s,\chi)$ to understanding the size of $L_y(1+it,\chi)$. We shall accomplish the latter task using the theory of *pretentious multiplicative functions*. Our starting point is the following lemma.

Lemma 22.3. *Let f be a completely multiplicative function with $|f| \leqslant 1$, and let F denote its Dirichlet series. For $y \geqslant 2$ and $\sigma > 1$, we have*

$$\log F_y(s) = \sum_{y<p\leqslant x} \frac{f(p)}{p^{1+it}} + O(1) \quad \textit{with} \quad x = \max\{y, e^{1/(\sigma-1)}\}.$$

Proof. Note that $F_y(s) = \prod_{p>y}(1-f(p)/p^s)^{-1}$ by the complete multiplicativity of f. Since $|f| \leqslant 1$, we infer that

$$\log F_y(s) = \sum_{p>y}\sum_{m\geqslant 1} \frac{f(p)^m}{mp^{ms}} = \sum_{p>y} \frac{f(p)}{p^s} + O(1). \tag{22.12}$$

Now, Chebyshev's estimate and partial summation imply that

$$\sum_{p>e^{1/(1-\sigma)}} \frac{1}{p^\sigma} \ll 1. \tag{22.13}$$

This proves the lemma when $\sigma \geqslant 1 + 1/\log y$ (since $x = y$ then). On the other hand, if $\sigma < 1 + 1/\log y$, so that $\sigma = 1 + 1/\log x$, then (22.12) and (22.13) imply that

$$\log F_y(s) = \sum_{y<p\leqslant x} \frac{f(p)}{p^{1+1/\log x+it}} + O(1).$$

Finally, observe that $p^{1/\log x} = 1 + O(\log p/\log x)$ for $p \leqslant x$, and recall that $\sum_{p\leqslant x}(\log p)/p \ll \log x$. This completes the proof of the lemma. □

For future reference, we record the following one-sided bound for the averages of $\chi(p)/p^{1+it}$, which is obtained as a direct corollary of Theorem 22.1 and Lemma 22.3.

Corollary 22.4. *Let χ be a Dirichlet character* mod q *and $t \in \mathbb{R}$. Then*

$$\sum_{u<p\leqslant v} \frac{\mathrm{Re}(\chi(p)p^{-it})}{p} \leqslant O(1) \quad \textit{for all } u \geqslant v \geqslant q_{\chi,t}.$$

Lemma 22.3 relates $L_y(1+it,\chi)$ to logarithmic averages of $\chi(p)p^{-it}$ and demonstrates that for $L_y(1+it,\chi)$ to be small, the quantities $\chi(p)p^{-it}$ must predominantly have negative real part. The most extreme case would be when $\chi(p)p^{-it} \approx -1$ for most p, in which case we can think of $\chi(n)n^{-it}$ as "pretending to be" the Möbius function. To study more rigorously this type of arguments, we introduce the distance function

$$\mathbb{D}(f,g;u,v)^2 = \frac{1}{2}\sum_{u<p\leqslant v} \frac{|f(p)-g(p)|^2}{p},$$

which is a variant of the distance function used in the proofs of Theorems 8.3 and 12.3. Note that if $|f(p)| = |g(p)| = 1$ for all $p \in (u, v]$, then

$$\mathbb{D}(f, g; u, v)^2 = \sum_{u<p\leqslant v} \frac{1 - \mathrm{Re}(f(p)\bar{g}(p))}{p}. \tag{22.14}$$

Together with Lemma 22.3, this establishes the connection of $L_y(s, \chi)$ to the distance function.

With the above notation, we have the following significant strengthening of Corollary 22.4 that shows that there exists a parameter Y controlled by the size of $L_y(1+it, \chi)$ such that the average behavior of $\chi(p)p^{-it}$ undergoes a phase transition when $p \approx Y$: it is 0 on average when $p \geqslant Y$, whereas it is -1 on average when $p < Y$.

Theorem 22.5. *Let χ be a Dirichlet character* mod q*, $t \in \mathbb{R}$ and $y \geqslant q_{\chi,t}$. There exists some $Y = Y(\chi, t) \in [y, +\infty]$ such that*

$$\mathbb{D}(\chi(n), \mu(n)n^{it}; u, v) = O(1) \quad \textit{when} \quad [u, v] \subseteq [y, Y) \tag{22.15}$$

and

$$\sum_{u<p\leqslant v} \frac{\mathrm{Re}(\chi(p)p^{-it})}{p} = O(1) \quad \textit{when} \quad [u, v] \subseteq [Y, +\infty). \tag{22.16}$$

In fact, we have $\log Y/\log y \asymp 1/|L_y(1 + it, \chi)|$.

Proof. We take $Y = \max\{y, y^{1/|L_y(1+it,\chi)|}\}$. Since $|L_y(1 + it, \chi)| \ll 1$ from Theorem 22.1, we have that $\log Y/\log y \asymp 1/|L_y(1 + it, \chi)|$.

To deal with the potential issue of having $L(1 + it, \chi) = 0$,[1] we let $Y_\varepsilon = y^{1/|L_y(1+\varepsilon+it,\chi)|}$ for $\varepsilon \geqslant 0$. In addition, let $f_\varepsilon(n) = \chi(n)n^{-\varepsilon-it}$ and call $F_{y,\varepsilon}(s)$ the associated sifted Dirichlet series (that equals $L_y(s + \varepsilon + it, \chi)$).

First, we prove a generalization of (22.16): we claim that

$$\sum_{u<p\leqslant v} \frac{\mathrm{Re}(f_\varepsilon(p))}{p} = O(1) \quad \text{when} \quad [u, v] \subseteq [Y_\varepsilon, +\infty). \tag{22.17}$$

(Taking $\varepsilon = 0$ in (22.17) recovers (22.16).) We may assume that $Y_\varepsilon < \infty$; otherwise, (22.17) is vacuous. For $\sigma > 1$, Theorem 22.1 implies that

$$F_{y,\varepsilon}(\sigma) = F_{y,\varepsilon}(1) + \int_1^\sigma F'_{y,\varepsilon}(\alpha)\mathrm{d}\alpha = F_{y,\varepsilon}(1) + O((\sigma - 1)\log y). \tag{22.18}$$

Hence, if C is a large enough constant (independently of any parameter), we have that $|F_{y,\varepsilon}(\sigma)| \asymp |F_{y,\varepsilon}(1)|$ when $0 < \sigma - 1 \leqslant |F_{y,\varepsilon}(1)|/(C\log y)$. Since $|F_{y,\varepsilon}(1)|/\log y = 1/\log Y_\varepsilon$, we infer that $|F_{y,\varepsilon}(1 + 1/\log u, \chi)| \asymp |F_{y,\varepsilon}(1 +$

[1] We are assuming we have no knowledge about the zeroes of $L(s, \chi)$.

$1/\log v, \chi)|$ for $v \geqslant u \geqslant Y_\varepsilon^C$. Taking logarithms and applying Lemma 22.3 twice, we arrive at the estimate

$$\sum_{y<p\leqslant u} \frac{\mathrm{Re}(f_\varepsilon(p))}{p} = \sum_{y<p\leqslant v} \frac{\mathrm{Re}(f_\varepsilon(p))}{p} + O(1).$$

This completes the proof of (22.17) when $[u, v] \subseteq [Y_\varepsilon^C, \infty)$. For the remaining range, we simply note that $\sum_{Y_\varepsilon<p\leqslant Y_\varepsilon^C} 1/p = O(1)$.

Next, we prove (22.15). Fix, for the moment, $\varepsilon > 0$. (We will eventually let $\varepsilon \to 0^+$.) We then know that $Y_\varepsilon < \infty$ from the Euler product representation of $L(s, \chi)$. Now, for any $x \geqslant Y_\varepsilon$, we have that

$$\sum_{y<p\leqslant Y_\varepsilon} \frac{\mathrm{Re}(f_\varepsilon(p))}{p} = \sum_{y<p\leqslant x} \frac{\mathrm{Re}(f_\varepsilon(p))}{p} + O(1) = \log \Big|F_{y,\varepsilon}\Big(1 + \frac{1}{\log x}\Big)\Big| + O(1),$$

where the first equality follows by (22.17) and the second one from Lemma 22.3. Letting $x \to \infty$, we deduce that

$$\sum_{y<p\leqslant Y_\varepsilon} \frac{\mathrm{Re}(f_\varepsilon(p))}{p} = \log |F_{y,\varepsilon}(1)| + O(1) = -\sum_{y<p\leqslant Y_\varepsilon} \frac{1}{p} + O(1),$$

where the second equality follows from the definition of Y_ε and Mertens' second estimate (Theorem 3.4(b)). We conclude that

$$\sum_{y<p\leqslant Y_\varepsilon} \frac{1 + \mathrm{Re}(f_\varepsilon(p))}{p} \leqslant O(1). \tag{22.19}$$

Now fix $[u, v] \subset [y, Y)$. If ε is small enough, then $[u, v] \subset [y, Y_\varepsilon)$. In addition, all summands in (22.19) are non-negative, because $1 + \mathrm{Re}(z) \geqslant 0$ when $|z| \leqslant 1$. In conclusion,

$$0 \leqslant \sum_{u<p\leqslant v} \frac{1 + \mathrm{Re}(f_\varepsilon(p))}{p} \leqslant O(1)$$

for all ε that are sufficiently small. Letting $\varepsilon \to 0^+$ completes the proof of (22.15) and hence of the theorem. □

We now combine the above result with the ideas used in the proof of Theorem 12.3. Note that the result we obtain below can be combined with (22.5) to yield a result that is almost as strong as Theorem 12.3.

Theorem 22.6. *Consider a Dirichlet character χ mod q, and real numbers t and $y \geqslant \max\{q(|t| + 2), 10\}$.*

(a) (i) *If either χ is not real or $|t| \geqslant 1/\log y$, then $|L_y(1 + it, \chi)| \asymp 1$.*
(ii) *If χ is real and non-principal and $|t| \leqslant 1/\log y$, then*

$$|L_y(1 + it, \chi)| \asymp \max\{L_y(1, \chi), |t| \log y\}.$$

(b) *Assume that χ is real and non-principal, and let $\chi' \pmod{q'}$ be another real, non-principal character that is not induced by the same primitive character as χ. If $L_y(1,\chi) \geqslant L_y(1,\chi')$ for some $y \geqslant \max\{q,q'\}$, then $L_y(1,\chi) \asymp 1$.*

Proof. (a-i) Our assumptions on χ, y and t imply that $y \geqslant q_{\chi,t}$ and $y \geqslant q_{\chi^2,t}$. Now, let Y be as in Theorem 22.5. Since we already know that $|L_y(1,\chi)| \ll 1$ from Theorem 22.1, it suffices to prove that $|L_y(1,\chi)| \gg 1$ or, equivalently, that $\log Y \ll \log y$. Theorem 22.5 implies that $\mathbb{D}(\chi(n), \mu(n)n^{it}; y, Y) \ll 1$. Applying Minkowski's inequality as in the proof of Theorem 12.3 (or simply noticing that $|z^2 - w^2| \leqslant 2|z+w|$ when $|z|, |w| \leqslant 1$), we have

$$\mathbb{D}(\chi^2(n), n^{2it}; y, Y) \leqslant 2\mathbb{D}(\chi(n), \mu(n)n^{it}; y, Y) \ll 1.$$

On the other hand, Corollary 22.4 and relation (22.14) imply that

$$\mathbb{D}(\chi^2(n), n^{2it}; y, Y)^2 \geqslant \log(\log Y/\log y) - O(1).$$

Comparing the above estimates, we find that $\log Y \ll \log y$, as needed.

(a-ii) Let $z = e^{1/|t|}$, so that $z \geqslant y \geqslant \max\{q(|t|+2), 10\}$ and $|t| \geqslant 1/\log z$. Hence, $|L_z(1+it,\chi)| \asymp 1$ by part (a-i) applied with z in place of y. Together with Theorem 22.5, this implies that

$$\sum_{z<p\leqslant x} \frac{\mathrm{Re}(\chi(p)p^{-it})}{p} = O(1) \quad (x \geqslant z).$$

We combine the above estimate with Lemma 22.3 to find that

$$\begin{aligned} \log L_y(1+it,\chi) &= \lim_{x\to\infty} \log L_y(1+1/\log x + it, \chi) \\ &= \sum_{y<p\leqslant z} \frac{\chi(p)}{p^{1+it}} + O(1). \end{aligned} \tag{22.20}$$

We have $|p^{-it} - 1| = |\int_0^t (\log p) p^{iu} \mathrm{d}u| \leqslant |t| \log p$. Hence,

$$\log L_y(1+it,\chi) = \sum_{y<p\leqslant z} \frac{\chi(p)}{p} + O(1).$$

Let Y be as in Theorem 22.5 with $t = 0$. Then

$$\log L_y(1+it,\chi) = \sum_{y<p\leqslant \min\{Y,z\}} \frac{\chi(p)}{p} + O(1) = -\sum_{y<p\leqslant \min\{Y,z\}} \frac{1}{p} + O(1),$$

where we first applied (22.16), followed by an application of (22.15) (both with $t = 0$). Since $\log Y \asymp (\log y)/L_y(1,\chi) \gg \log y$, Mertens's second estimate completes the proof of part (a-ii).

(b) Let $Y = \max\{y, y^{1/|L_y(1,\chi)|}\}$ and $Y' = \max\{y, y^{1/|L_y(1,\chi')|}\}$, as in the proof of Theorem 22.5. Since $L_y(1,\chi) \geqslant L_y(1,\chi')$, we must have that $Y \leqslant Y'$. Hence, Theorem 22.5 implies that $\mathbb{D}(\psi, \mu; y, Y) \ll 1$ for $\psi \in \{\chi, \chi'\}$.

Using Minkowski's inequality, we infer that $\mathbb{D}(\chi, \chi'; y, Y) \ll 1$. On the other hand, Corollary 22.4 and relation (22.14) yield that

$$\mathbb{D}(\chi, \chi'; y, Y)^2 \geqslant \log(\log Y/\log y) - O(1),$$

where we used that $\chi\chi'$ is a non-principal mod $[q, q'] \leqslant y^2$, which follows by our assumption that χ and χ' are induced by different primitive characters. Hence, $\log Y \ll \log y$, which completes the proof of the theorem. □

Exercises

Exercise 22.1. Let f and F be as in Lemma 22.3. Fix $t \in \mathbb{R}$ and $y \geqslant 2$, and assume that the function $\sigma \to F(\sigma+it)$ is continuously differentiable for $\sigma \geqslant 1-1/\log y$, as well as that $F_y^{(j)}(\sigma+it, f)| \ll (\log y)^j$ uniformly for $j \in \{0,1\}$ and $\sigma \geqslant 1-1/\log y$. Let $Y = y^{1/\min\{1,|F_y(1+it)|\}}$.

(a) Prove that

$$\begin{cases} \mathbb{D}(f(n), \mu(n)n^{it}; u, v) = O(1) & \text{when} \quad [u,v] \subseteq [y, Y), \\ \sum_{u<p\leqslant v} \mathrm{Re}(f(p)p^{-it})/p = O(1) & \text{when} \quad [u,v] \subseteq [Y, +\infty). \end{cases}$$

(b) Prove that there is an absolute constant $c > 0$ such that

$$|L_y(\sigma+it, \chi)| \asymp \begin{cases} \log y/\log Y & \text{if } 1 - c/\log Y \leqslant \sigma \leqslant 1 + 1/\log Y, \\ (\sigma-1)\log y & \text{if } 1 + 1/\log Y \leqslant \sigma \leqslant 1 + 1/\log y, \\ 1 & \text{if } \sigma \geqslant 1 + 1/\log y. \end{cases}$$

Exercise 22.2.* Assume there are constants $\varepsilon > 0$, $L \geqslant 2$ and q_0 such that

$$\pi(x; q, a) \leqslant \frac{(2-\varepsilon)x}{\varphi(q)\log x} \qquad (x \geqslant q^L,\ q \geqslant q_0).$$

Show that there is some $c = c(q_0, L, \varepsilon) > 0$ such that, for all moduli $q \geqslant 1$, the function $\prod_{\chi \,(\mathrm{mod}\, q)} L(s, \chi)$ has no zeroes with $\sigma \geqslant 1 - c/\log(q(|t|+2))$ (i.e., there are no Landau-Siegel zeroes). [*Hint:* Given a real, non-principal character $\chi \,(\mathrm{mod}\, q)$, prove a lower bound for the sum $\sum_{q^L<p\leqslant x}(1+\chi(p))/p$.]

Exercise 22.3.* Let χ be a real, non-principal Dirichlet character mod q. Theorem 22.6 proves that if $L_q(1, \chi) \gg 1$, then $L(s, \chi)$ does not have a Landau-Siegel zero. This exercise shows that the converse is also true. Moreover, it establishes a precise connection between the location of a potential Landau-Siegel zero and the size of $L_q(1, \chi)$.

Throughout, $q_1 = \max\{q, 10\}^{100}$, $x \geqslant y \geqslant q_1$, $u = \log x/\log y$, $1 - 1/\log y \leqslant \sigma < 1$ and $V(y) = \prod_{p\leqslant y}(1-1/p)$.

(a) Let $f_\sigma(x) = (x^{1-\sigma}-1)/(1-\sigma) = \int_1^x w^{-\sigma}\mathrm{d}w$. Show that there is a constant $\gamma_{\sigma,y} = O(\log y)$ such that

$$\sum_{\substack{n\leqslant x \\ P^-(n)>y}} \frac{1}{n^\sigma} = \big(f_\sigma(x) + \gamma_{\sigma,y}\big)V(y) + O(e^{-100u}).$$

(b) For $A' \geqslant A \geqslant y$ and $x \geqslant y$, show that

$$\sum_{\substack{A<a\leqslant A'\\ P^-(a)>y}} \frac{\chi(a)f_\sigma(x/a)}{a^\sigma} = \int_1^x \sum_{\substack{A<a\leqslant \min\{A',x/w\}\\ P^-(a)>y}} \frac{\chi(a)}{a^\sigma}\cdot\frac{\mathrm{d}w}{w^\sigma} \ll \frac{x^{1-\sigma}\log(xA)}{A^{100/\log y}}.$$

Conclude that

$$\sum_{\substack{n\leqslant x\\ P^-(n)>y}} \frac{(1*\chi)(n)}{n^\sigma} = \Big\{\frac{x^{1-\sigma}}{1-\sigma}L_y(1,\chi) - \Big(\frac{1}{1-\sigma}-\gamma_{\sigma,y}\Big)L_y(\sigma,\chi)\Big\}V(y) + O(e^{-u}).$$

[*Hint:* Recall Lemma 22.2.]

(c) Show that there is an absolute constant $c > 0$ such that if $L_y(\sigma,\chi) \geqslant 0$ for some $\sigma \in [1 - c/\log y, 1)$, then $L_y(1,\chi) \gg (1-\sigma)\log y$. [*Hint:* Examine the sign of $(1/(\sigma-1) - \gamma_{\sigma,y})L_y(\sigma,\chi)$.]

(d) Show that if $L(\sigma,\chi) \neq 0$ for $\sigma \in [1 - c/\log q_1, 1]$, then $L_q(1,\chi) \gg 1$.

(e) If there is $\beta \in [1 - c/\log q_1, 1]$ such that $L(\beta,\chi) = 0$, then show that $1 - \beta \asymp L_q(1,\chi)/\log q$. [*Hint:* To prove the upper bound on $1-\beta$, use part (c). To prove the lower bound, use the Fundamental Theorem of Calculus.]

Part 5

Bilinear methods

Chapter 23

Vinogradov's method

The parity barrier of sieve methods prevents us from getting tight bounds on $\sum_{p\leqslant x} a_p$ under the mere assumption of Axioms 1–3 for the sequence $\mathcal{A} = (a_n)_{n=1}^{\infty}$. In 1934, I. M. Vinogradov[1] developed a new method for estimating $\sum_{p\leqslant x} a_p$ when $\mathcal{A}$ satisfies certain additional hypotheses.

To simplify the exposition of Vinogradov's idea, let us assume that $|a_n| \leqslant 1$ for all n. We then have

$$\sum_{p\leqslant x} a_p = \sum_{\substack{n\leqslant x\\ P^-(n)>\sqrt{x}}} a_n + O(\sqrt{x}).$$

Applying a variant of Buchstab's identity (19.1) to the right-hand side yields that

$$\sum_{p\leqslant x} a_p = \sum_{\substack{n\leqslant x\\ P^-(n)>x^\varepsilon}} a_n - \sum_{x^\varepsilon<p\leqslant x} \sum_{\substack{n\leqslant x\\ P^-(n)=p}} a_n + O(\sqrt{x}), \tag{23.1}$$

where $\varepsilon > 0$ is at our disposal. If we assume that the sequence $\mathcal{A}$ satisfies a suitable version of Axioms 1–3, the first sum on the right-hand side of (23.1) can be estimated accurately using the Fundamental Lemma of Sieve Theory (Theorem 18.11) for small enough values of ε. Thus, it remains to handle the double sum over p and n.

Writing $n = pm$, we find that

$$B := \sum_{x^\varepsilon<p\leqslant x} \sum_{\substack{n\leqslant x\\ P^-(n)=p}} a_n = \sum\sum_{\substack{x^\varepsilon<p\leqslant x\\ mp\leqslant x,\, P^-(m)\geqslant p}} a_{mp}.$$

[1]Not to be confused with A. I. Vinogradov from the Bombieri-Vinogradov theorem.

The right-hand side closely resembles a *bilinear sum*

$$\sum_{k=1}^{K}\sum_{\ell=1}^{L} a_{k\ell}x_k y_\ell \tag{23.2}$$

for appropriate coefficients x_k and y_ℓ. There is a small technicality: the variables p and m are weakly tangled via the relations $pm \leqslant x$ and $P^-(m) \geqslant p$. We can easily decouple them though: we (roughly) have

$$B \approx \sum_{x^\varepsilon < 2^j \leqslant x^{1/2}} B_j \qquad \text{with} \qquad B_j = \sum_{\substack{2^{j-1} < p \leqslant 2^j \\ m \leqslant x/2^j,\, P^-(m) \geqslant 2^j}}\sum a_{mp}, \tag{23.3}$$

so that B is a sum of $O(\log x)$ bilinear sums (B_j is of the form (23.2) with $K = x/2^j$, $L = 2^j$, $x_k = 1_{P^-(k)\geqslant 2^j}$ and $y_\ell = 1_{\ell \text{ is prime}} 1_{\ell \in (2^{j-1}, 2^j]}$).

Vinogradov's groundbreaking idea is that, for certain special sequences $\mathcal{A}$, we can obtain strong estimates for the bilinear sum (23.2) no matter what the coefficients x_k and y_ℓ are, as long as they are of controlled size (e.g. if $|x_k|, |y_\ell| \leqslant 1$ for all k, ℓ) and as long as both K and L are large, so that we have genuine bilinearity.[2] We may thus forget the precise definition of x_k and y_ℓ. If this alleged bilinear estimate (which we can think of as "Axiom 4" of sieve theory) is available in a large enough region of K and L so that both terms on the right-hand side of (23.1) can be handled (the first one by Axioms 1–3 and the second one by Axiom 4), we can break the parity barrier and extract primes from the sequence $(a_n)_{n=1}^\infty$.

We will explain Vinogradov's method more rigorously in the subsequent sections. But first let us note that Axiom 3 of sieve methods can also be thought of as an estimate for a bilinear sum of the form (23.2), but with $y_\ell = 1$ for all ℓ. Indeed, if $(a_n)_{n=1}^\infty \subset [1, L]$ and we assume Axiom 1, then

$$\sum_{k=1}^{K}\sum_{\ell=1}^{L} a_{k\ell}x_k = \sum_{k=1}^{K} x_k A_k = X\sum_{k=1}^{K}\frac{x_k\nu(k)}{k} + \sum_{k=1}^{K} x_k r_k,$$

where A_k is defined by (18.2). If we assume that $|x_k| \leqslant 1$ and that Axiom 3 holds with level of distribution $D \geqslant K$, then we can obtain a strong estimate for $\sum_{k\leqslant K} x_k r_k$. Conversely, if we can estimate this sum for any choice of x_k, we can also estimate it when x_k is the sign of r_k, which brings us right back to Axiom 3.

In conclusion, we may think of Axiom 3 as a bilinear estimate with the coefficients y_ℓ being smooth functions of ℓ. This point of view will be important in the next section.

[2] If, for instance, $K = 1$, then the expression in (23.2) becomes a sum over a single variable. We want to avoid such degenerate situations.

Two types of functions

Various technicalities in Vinogradov's method are simplified if instead of the sum $\sum_{p\leqslant x} a_p$ we work with $\sum_{n\leqslant x} a_n\Lambda(n)$. Indeed, the combinatorial identity $\Lambda = \mu * \log$ readily implies that

$$\sum_{n\leqslant x} a_n\Lambda(n) = \sum_{k\ell\leqslant x} a_{k\ell}\mu(k)\log\ell. \tag{23.4}$$

We thus see right away that $\sum_{n\leqslant x} a_n\Lambda(n)$ has some sort of bilinear structure. To bring the right-hand side of (23.4) into the form (23.2), we localize k into a dyadic interval $(2^{j-1}, 2^j]$, so that $\ell \leqslant x/2^{j-1}$. As we briefly mentioned before, the method of bilinear sums is efficient only when both k and ℓ are "long variables", that is to say, when 2^j and $x/2^j$ are both large (say when $D \leqslant 2^j \leqslant x/D$). On the other hand, when $2^j \leqslant D$, we can take advantage of the fact that the long variable ℓ is weighted with the smooth function log. Hence, this part of the sum can be handled too, provided that we have at our disposal an appropriate version of Axiom 3, as per the discussion in the end of the previous section. It remains to handle the summands with $x/D < 2^j \leqslant x$. If we can rewrite this part of the sum as a linear combination of sums that fit into one of the two above categories (i.e., a combination of some bilinear sums, and of some other ones with at least one smooth variable), we will have completed the estimation of $\sum_{n\leqslant x} a_n\Lambda(n)$.

This brings us to the heart of Vinogradov's method: given $x \geqslant 1$, we seek an identity of the form

$$\Lambda(n) = \sum_{1\leqslant j\leqslant J} (f_j * g_j)(n) + R(n) \qquad \text{for } n \leqslant x, \tag{23.5}$$

where the function R is a negligible "remainder term" in the sense that $\sum_{n\leqslant x} |a_n R(n)|$ is small compared to $\sum_{n\leqslant x} |a_n|$, and for each j the summands $f_j * g_j$ fall into one of the following two categories:

I) $\operatorname{supp}(f_j) \subseteq [1, y_j]$ for some y_j that is small compared to x and $g_j \in C^\infty(\mathbb{R}_{\geqslant 1})$. We then call $f_j * g_j$ a *quasi-smooth* or *type I function* and refer to the sum

$$\sum_{n\leqslant x} a_n(f_j * g_j)(n) = \sum_{k\leqslant y_j} f_j(k) \sum_{\ell\leqslant x/k} a_{k\ell} g_j(\ell)$$

as a *quasi-smooth*, *quasi-linear* or *type I sum*.

II) $\operatorname{supp}(f_j) \subseteq [1, y_j]$ and $\operatorname{supp}(g_j) \subseteq [1, z_j]$, where $D_j \leqslant y_j, z_j \leqslant x/D_j$ for some large D_j. We then call $f_j * g_j$ a *type II function* and its average

$$\sum_{n\leqslant x} a_n(f_j * g_j)(n) = \sum_{k\leqslant y_j,\, \ell\leqslant z_j,\, k\ell\leqslant x} \sum a_{k\ell} f_j(k) g_j(\ell)$$

a *bilinear* or *type II sum*.

Decomposing von Mangoldt's function

Vaughan's identity. One of the simplest and most useful ways to arrive at an identity of the form (23.5) was discovered by Vaughan. Given an arithmetic function f and a parameter V, we write

$$f_{\leqslant V}(n) := 1_{n\leqslant V} \cdot f(n) \qquad \text{and} \qquad f_{>V}(n) := 1_{n>V} \cdot f(n). \tag{23.6}$$

With the above notation, the identity $\Lambda = \mu * \log$ can be written as

$$\Lambda = \mu_{\leqslant V} * \log \, + \, \mu_{>V} * \log . \tag{23.7}$$

The first term on the right-hand side of (23.7) is of type I. But the second term is neither of type I nor of type II. To proceed, we replace $\mu_{>V}$ by $\mu_{\leqslant V}$ using Möbius inversion: we have

$$\mu_{>V} * 1 = \delta \, - \, \mu_{\leqslant V} * 1, \tag{23.8}$$

where we recall the notation $\delta(n) = 1_{n=1}$ from Chapter 3. As preparation for inserting (23.8) into (23.7), we write the latter formula as

$$\Lambda = \mu_{\leqslant V} * \log \, + \, \mu_{>V} * 1 * \Lambda.$$

Because Λ has unrestricted support, we first split it as $\Lambda = \Lambda_{\leqslant U} + \Lambda_{>U}$, where U is some parameter, and then apply (23.8) only to the part of Λ supported on $[1, U]$. We conclude that

$$\Lambda = \mu_{\leqslant V} * \log \, + \, \mu_{>V} * 1 * \Lambda_{>U} \, + \, (\delta \, - \, \mu_{\leqslant V} * 1) * \Lambda_{\leqslant U}.$$

We have thus proven *Vaughan's identity*:

Lemma 23.1. *For any $U, V \geqslant 1$, we have*

$$\Lambda = \mu_{\leqslant V} * \log \, - \, (\Lambda_{\leqslant U} * \mu_{\leqslant V}) * 1 \, + \, (\Lambda_{>U} * 1) * \mu_{>V} \, + \, \Lambda_{\leqslant U}. \tag{23.9}$$

The function $\Lambda_{\leqslant U}$ is supported on small integers and hence contributes a negligible amount to averages of Λ. The function $\mu_{\leqslant V} * \log$ is a quasi-smooth convolution: the first factor is a bounded function supported on integers $\leqslant V$. Similarly, the function $(\Lambda_{\leqslant U} * \mu_{\leqslant V}) * 1$ is also a quasi-smooth convolution, with the factor $\Lambda_{\leqslant U} * \mu_{\leqslant V}$ being supported on $[1, UV]$ and satisfying the pointwise bound $|\Lambda_{\leqslant U} * \mu_{\leqslant V}| \leqslant \Lambda * 1 = \log$. We denote the total contribution to Λ of these two type I functions by

$$\Lambda^\sharp := \mu_{\leqslant V} * \log \, - \, (\Lambda_{\leqslant U} * \mu_{\leqslant V}) * 1. \tag{23.10}$$

Finally, the function

$$\Lambda^\flat := (\Lambda_{>U} * 1) * \mu_{>V} \tag{23.11}$$

is of type II: its first factor is supported on integers $> U$ and its second one on integers $> V$.

A very useful feature of $\Lambda^\flat$ is that one of its factors is the Möbius function that is completely aperiodic (see Corollary 13.4 and Exercise 23.4). As a

result, $\Lambda^\flat$ typically contributes to the error term in the estimation of the sum $\sum_{n\leqslant x} a_n\Lambda(n)$, so that the main term comes from $\Lambda^\sharp$. We thus think of $\Lambda^\sharp$ as the "structured" part of Λ. It resembles a sieve-type weight and we need a suitable version of Axiom 3 to estimate its averages. On the other hand, we think of $\Lambda^\flat$ as an "unstructured/random" error term, and we usually treat it using bilinear methods.

Remark 23.2. By definition, we have

$$\Lambda^\flat(n) = \sum_{\substack{k\ell=n\\ k>U,\,\ell>V}}\sum (\Lambda_{>U} * 1)(k)\mu(\ell).$$

When $n \leqslant x$, we have $U < k = n/\ell \leqslant x/V$. However, we often need better control of the support of the variables k and ℓ. To achieve this goal, we cover the interval $(U, x/V]$ by dyadic intervals $(2^{j-1}, 2^j]$, where $2^j \in (U, 2x/V]$. If $k \in (2^{j-1}, 2^j]$, we also have that $\ell = n/k \leqslant x/2^{j-1}$. This leads us to the more accurate decomposition

$$\Lambda^\flat(n) = \sum_{U<2^j\leqslant 2x/V} (f_j * g_j)(n) \quad \text{for} \quad n \leqslant x, \tag{23.12}$$

where $f_j(k) = (\Lambda_{>U} * 1)(k)1_{2^{j-1}<k\leqslant 2^j}$ and $g_j(\ell) = \mu(\ell)1_{V<\ell\leqslant x/2^{j-1}}$. □

Presieving Λ. In many occasions, it is advantageous to use a variant of Vaughan's identity whose summands enjoy slightly different properties. A simple way of obtaining such a variant is by *presieving* Λ. Indeed, since primes do not have small prime factors, we write

$$\Lambda(n) = \Lambda(n)\cdot 1_{P^-(n)>y} + \Lambda(n)\cdot 1_{P^-(n)\leqslant y}.$$

We expect $\Lambda(n)\cdot 1_{P^-(n)\leqslant y}$ to be small on average because it is supported on prime powers p^m with $p \leqslant y$. Next, we decompose the function $\Lambda(n)\cdot 1_{P^-(n)>y}$ by first replacing Λ by $\mu * \log$. This yields the identity

$$\Lambda(n)1_{P^-(n)>y} = \sum_{\substack{k\ell=n\\ P^-(k\ell)>y}}\sum \mu(k)\log\ell. \tag{23.13}$$

The fact that $\log(1) = 0$ means that the above sum is supported on integers $\ell > 1$. Since we also know that $P^-(\ell) > y$, we must have $\ell > y$. We thus see that we automatically have a long ℓ variable weighted with the smooth function log times the indicator function of integers free of prime factors $\leqslant y$. Even though the latter is not a smooth function, it is quasi-smooth when y is small enough. The reason is that Theorem 19.1 allows us to approximate the function $n \to 1_{P^-(n)>y} = 1_{(n,P(y))=1}$ by convolutions $\lambda^\pm * 1$, where $\lambda^\pm$ take values in $[-1, 1]$ and have small support. Hence, for all practical purposes, we may think of the function $\ell \to 1_{P^-(\ell)>y}\log\ell$ as a quasi-smooth function.

Motivated by the above discussion, we split the right-hand side of (23.13) according to the size of k, which leads us to the following decomposition:

$$\Lambda = \Lambda^{\sharp}_{\text{sieve}} + \Lambda^{\flat}_{\text{sieve}} + R_{\text{sieve}}, \tag{23.14}$$

where

$$\Lambda^{\sharp}_{\text{sieve}}(n) = \sum_{\substack{k\ell=n,\, k\leqslant D\\ P^-(k\ell)>y}}\sum \mu(k)\log\ell, \tag{23.15}$$

$$\Lambda^{\flat}_{\text{sieve}}(n) = \sum_{\substack{k\ell=n,\, k>D,\, \ell>y\\ P^-(k\ell)>y}}\sum \mu(k)\log\ell \tag{23.16}$$

and $R_{\text{sieve}}(n) = 1_{P^-(n)\leqslant y}\Lambda(n)$. Note that $\Lambda^{\sharp}_{\text{sieve}}$ is essentially of type I, $\Lambda^{\flat}_{\text{sieve}}$ is of type II and R_{sieve} is of negligible size on average, since

$$\sum_{n\leqslant x} R_{\text{sieve}}(n) = \sum_{p\leqslant y,\, p^m\leqslant x}\sum \log p \leqslant \sum_{p\leqslant y}\log x \leqslant y\log x. \tag{23.17}$$

A choice of y and D that works for many applications is

$$y = \exp\{(\log x)^{\theta_1}\} \quad\text{and}\quad D = \exp\{(\log x)^{\theta_2}\}, \tag{23.18}$$

where $0 < \theta_1 < \theta_2 < 1$ can be chosen freely.

The main advantage of (23.14) compared to Vaughan's identity is that the functions $\Lambda^{\sharp}_{\text{sieve}}$ and $\Lambda^{\flat}_{\text{sieve}}$ are presieved with all primes $\leqslant y$. This rather technical feature of (23.14) plays a key role in the proof of Linnik's theorem in Chapter 27. We will also see in Exercise 26.4 how it leads to a better version of the Bombieri-Vinogradov theorem.

A secondary advantage of (23.14) versus Vaughan's identity is that its "main term" $\Lambda^{\sharp}_{\text{sieve}}$ consists of a single type I function. This fact makes various calculations easier and will come into play in Chapter 24.

On the other hand, Vaughan's identity offers much more freedom in the choice of the parameters U and V. Therefore, we have more control over the support of the functions appearing in the type I and type II sums, which is very important in certain applications. In contrast, the parameters y and D in (23.14) must be chosen carefully so that we have enough room to apply the Fundamental Lemma of Sieve Theory. In particular, y must be $x^{o(1)}$.

Remark 23.3. It is possible to create a new combinatorial decomposition of Λ that combines the best attributes of Vaughan's identity and of (23.14). This is done by presieving Vaughan's identity, that is to say, by multiplying all summands of (23.9) with the function $n \to 1_{P^-(n)>y}$. □

There are a lot more combinatorial decompositions of von Mangoldt's function than the ones we discussed above. A formula of particular importance is Heath-Brown's identity, given in Exercise 23.5 below. It is not

exactly of the form (23.5). Hence working with it is a bit more complicated, the task being understanding how to rearrange its terms and bring it to the form (23.5). However, Heath-Brown's identity has the important feature that all of the long functions appearing in it are smooth.

A further analysis of the subject of combinatorial decompositions of Λ can be found in [114, Chapter 13] or [59, Chapter 17]. Finally, a more sieve-theoretic approach to Vinogradov's method that is more in line with the discussion in the introduction of this chapter is presented in Harman's book on *prime-detecting sieves* [96].

The additive Fourier transform of the primes

To exemplify Vinogradov's method, we employ it to study a concrete and rather important example: the exponential sum

$$\sum_{p\leqslant x} e(\alpha p).$$

This sum is intimately related with the additive properties of primes and we will use it in the next chapter to study ternary arithmetic progressions in the primes. To get an idea of its size, we begin by studying it assuming the Generalized Riemann Hypothesis. This will serve as a guide for what kind of bounds to look for when we estimate it later via Vinogradov's method.

First, let us consider the special case when α is a rational number, say $\alpha = a/q$ with $(a,q)=1$. Then

$$\begin{aligned}\sum_{p\leqslant x} e(ap/q) &= \sum_{b\in(\mathbb{Z}/q\mathbb{Z})^*} e(ab/q)\pi(x;q,b) + \sum_{p\leqslant x,\, p\mid q} e(ap/q)\\ &= \frac{\operatorname{li}(x)}{\varphi(q)} \sum_{b\in(\mathbb{Z}/q\mathbb{Z})^*} e(ab/q) + O(\sqrt{x}q\log(qx))\end{aligned} \tag{23.19}$$

by Exercise 11.2 and partial summation. Making the change of variables $n \equiv ab \pmod q$, we see that the sum over b is the Gauss sum of the principal character mod q, which equals $\mu(q)$ (see Exercises 10.1 and 10.5). Therefore

$$\sum_{p\leqslant x} e(pa/q) = \frac{\mu(q)}{\varphi(q)} \cdot \operatorname{li}(x) + O(\sqrt{x}q\log(qx)).$$

To estimate $\sum_{p\leqslant x} e(p\alpha)$ for irrational α, we find a good rational approximation to it using the following classical result.

Lemma 23.4 (Dirichlet's approximation theorem)**.** *Let $\alpha \in \mathbb{R}$ and $Q \geqslant 1$. There is a reduced fraction a/q with $q \leqslant Q$ and*

$$\left|\alpha - \frac{a}{q}\right| < \frac{1}{qQ}.$$

Proof. Consider the $\lfloor Q \rfloor + 1$ numbers αq with $0 \leqslant q \leqslant Q$. We reduce them mod 1 to place them in the interval $[0,1)$. By the pigeonhole principle, there must exist $0 \leqslant q_1 < q_2 \leqslant Q$ such that $\|\alpha q_2 - \alpha q_1\| \leqslant 1/(\lfloor Q \rfloor + 1) < 1/Q$. We then take $q' = q_2 - q_1$ and a' to be the unique integer in $[\alpha q' - 1/2, \alpha q' + 1/2)$, so that $1 \leqslant q' \leqslant Q$ and $|\alpha q' - a'| = \|\alpha q'\| < 1/Q$. Letting a/q be the fraction a'/q' in reduced form completes the proof. □

Fix Q and a/q as in Lemma 23.4. If we write $\alpha = \beta + a/q$, then

$$\begin{aligned}\sum_{p\leqslant x} e(\alpha p) &= \int_{2^-}^{x} e(\beta y)\mathrm{d}\sum_{p\leqslant y} e(ap/q)\\ &= \frac{\mu(q)}{\varphi(q)}\int_2^x \frac{e(\beta y)}{\log y}\mathrm{d}y + O\big((1+|\beta|x)\sqrt{x}q\log(qx)\big)\end{aligned} \tag{23.20}$$

by partial summation. Since $|\beta| \leqslant 1/(qQ)$, taking $Q = \sqrt{x}(\log x)^3$ yields

$$\sum_{p\leqslant x} e(p\alpha) = \frac{\mu(q)}{\varphi(q)}\int_2^x \frac{e(\beta y)}{\log y}\mathrm{d}y + O\Big(\frac{x}{(\log x)^2} + \sqrt{x}q\log x\Big). \tag{23.21}$$

In particular, we see that if α is close to a rational number of denominator $q \in [(\log x)^2, \sqrt{x}/(\log x)^3]$, then there is significant cancellation among the numbers $e(\alpha p)$ with $p \leqslant x$, which makes $\sum_{p\leqslant x} e(\alpha p)$ smaller than $\pi(x)$.

The above calculation is a manifestation of an important principle stemming from the *Hardy-Littlewood circle method* that we will study in detail in Chapter 24: the Fourier transform

$$\sum_{n\leqslant x} c_n e(n\alpha) \tag{23.22}$$

of various interesting arithmetic sequences $(c_n)_{n\leqslant x}$ is big when α lies close to a rational number of small denominator, and it is small otherwise. The rough heuristic to explain this dichotomy is that when α is far from any fraction of small denominator, the sequence $(e(n\alpha))_{n\leqslant x}$ lacks any meaningful arithmetic structure, so that it cannot correlate with any "reasonably regular" sequence $(c_n)_{n\leqslant x}$.

A central problem in analytic number theory is to establish strong estimates for the exponential sum $\sum_{n\leqslant x} c_n e(n\alpha)$: an asymptotic formula when α is close to a fraction of small denominator, and a non-trivial upper bound otherwise. In particular, we would like to do so when c_n is the indicator function of the primes without appealing to the unproven Generalized Riemann Hypothesis.

Type I exponential sums

In view of the decomposition of Λ into type I and type II functions, the estimation of $\sum_{p\leqslant x} e(\alpha p)$ boils down to the estimation of $\sum_{n\leqslant x}(f*g)(n)e(\alpha n)$,

when $f * g$ is a function of type I or II. We begin by studying the first category of functions.

Let us begin by handling the simplest non-trivial type I function: the constant function 1. Arguing as in (10.12), we have

$$\Big|\sum_{n\leqslant x} e(\alpha n)\Big| = \Big|e(\alpha)\cdot\frac{1-e(\alpha\lfloor x\rfloor)}{1-e(\alpha)}\Big| \leqslant \frac{1}{2\|\alpha\|}, \tag{23.23}$$

where we recall that $\|\alpha\|$ denotes the distance of α from the nearest integer. We thus immediately see that, as long as $\|\alpha\| = o(1/x)$, the sum $\sum_{n\leqslant x} e(\alpha n)$ is small compared to the trivial bound

$$\Big|\sum_{n\leqslant x} e(\alpha n)\Big| \leqslant \sum_{n\leqslant x} 1 \leqslant x. \tag{23.24}$$

Using partial summation, we may easily pass from (23.23) and (23.24) to an estimate for the Fourier transform of the function $\log^v$, where v is any fixed positive real number. Indeed, we have

$$\begin{aligned}\sum_{n\leqslant x}(\log n)^v e(n\alpha) &= \int_{1^-}^{x}(\log t)^v \mathrm{d}\sum_{n\leqslant t} e(n\alpha)\\ &\ll (\log x)^v\cdot\min\{x, \|\alpha\|^{-1}\}\end{aligned} \tag{23.25}$$

uniformly for $x \geqslant 1$ and $v \geqslant 0$. Similar estimates are true if we replace $\log^v$ by a more general smooth function but we will not need them.

The above observations and the simplest version of Dirichlet's hyperbola method allow us to establish non-trivial estimates for general exponential sums of type I when α is close to a fraction a/q of large denominator (say, with $q \geqslant (\log x)^A$ for some large A). The notation $\|f\|_\infty$ in the statement of Theorem 23.5 below stands for the supremum norm of f. Finally, its proof features an important concept in the study of exponential sums: we say that a set of real numbers $\{\alpha_1, \dots, \alpha_r\}$ is *δ-spaced mod* 1 if

$$\|\alpha_i - \alpha_j\| \geqslant \delta \quad \text{whenever} \quad i \neq j. \tag{23.26}$$

Theorem 23.5. *Let $f : \mathbb{N} \to \mathbb{C}$ be supported on $[1, y]$, $v \geqslant 0$, $x \geqslant 2$, $\alpha \in \mathbb{R}$ and a/q be a reduced fraction with $|\alpha - a/q| \leqslant 1/q^2$. Then*

$$\sum_{n\leqslant x}(f * \log^v)(n)e(n\alpha) \ll \Big(y + \frac{x}{q} + q\Big)(\log x)^{v+1}\|f\|_\infty. \tag{23.27}$$

Proof. If $q = 1$, $q > x$ or $y > x$, we simply note that $|(f * \log^v)(n)| \leqslant \|f\|_\infty \tau(n)(\log n)^v$ and use Theorem 3.3. Assume now that $2 \leqslant q \leqslant x$ and

$y \leqslant x$. Opening the convolution and applying (23.25) yields

$$\sum_{n\leqslant x}(f*\log^v)(n)e(n\alpha) = \sum_{k\leqslant y} f(k) \sum_{\ell\leqslant x/k} (\log \ell)^v e(\ell\cdot k\alpha)$$

$$\ll (\log x)^v \|f\|_\infty \sum_{k\leqslant y} \min\big\{x/k, 1/\|k\alpha\|\big\}. \tag{23.28}$$

We cover the last sum by subsums of length $\tilde{q} := \lfloor q/2\rfloor$ defined by

$$S_m := \sum_{m\tilde{q}<k\leqslant (m+1)\tilde{q}} \min\big\{x/k, 1/\|k\alpha\|\big\}.$$

Since $(a,q)=1$, the numbers ka/q with $m\tilde{q} < k \leqslant (m+1)\tilde{q}$ are all distinct mod 1. Hence, $\|k_1a/q - k_2a/q\| \geqslant 1/q$ whenever $m\tilde{q} < k_1 < k_2 \leqslant (m+1)\tilde{q}$. On the other hand, if we write $\alpha = a/q + \beta$, then $|k_1\beta - k_2\beta| \leqslant \tilde{q}|\beta| \leqslant (q/2)/q^2 = 1/(2q)$. As a consequence, we find that the numbers $k\alpha$ with $m\tilde{q} < k \leqslant (m+1)\tilde{q}$ are $(2q)^{-1}$-spaced mod 1. We index them as $\alpha_1, \dots, \alpha_{\tilde{q}}$ in a way that $\|\alpha_1\| \leqslant \cdots \leqslant \|\alpha_{\tilde{q}}\|$. For each integer $j \in [1,\tilde{q}]$, the interval $(-\frac{j-1}{4q}, \frac{j-1}{4q})$ can contain at most $j-1$ of the reductions mod 1 of the numbers $\alpha_1, \dots, \alpha_{\tilde{q}}$. Hence, we must have that $\|\alpha_j\| \geqslant (j-1)/(4q)$ for $j = 1, \dots, \tilde{q}$.

When $m \geqslant 1$, the above discussion and the fact that $x/k < x/(m\tilde{q})$ whenever $k > m\tilde{q}$ yield the inequality

$$S_m \leqslant \frac{x}{m\tilde{q}} + \sum_{2\leqslant j\leqslant q/2} \frac{4q}{j-1} \ll \frac{x}{mq} + q\log q. \tag{23.29}$$

However, when $m = 0$, we cannot use the above argument as it currently stands because we do not have a good bound for the summand of S_0 corresponding to the integer k with $k\alpha = \alpha_1$. Note though that if $1 \leqslant k \leqslant q/2$, then $|k\beta| \leqslant (q/2)/q^2 = 1/(2q)$ and $\|ka/q\| \geqslant 1/q$. Therefore, $\|k\alpha\| \geqslant 1/(2q)$ for all $k \in \mathbb{Z} \cap [1, q/2]$. In particular, $\|\alpha_1\| \geqslant 1/(2q)$ when $m = 0$, and thus

$$S_0 \leqslant 2q + \sum_{2\leqslant j\leqslant q/2} \frac{4q}{j-1} \ll q\log q. \tag{23.30}$$

Combining (23.29) with (23.30), and noticing that there are $\leqslant y/\tilde{q} \ll y/q$ integers $m \in [1, y/\tilde{q}]$ allows us to estimate the expression in (23.28) and complete the proof of the theorem. □

Type II exponential sums

Let us now consider the exponential sum $\sum_{n\leqslant x}(f*g)(n)e(\alpha n)$ for a type II function $f*g$. For concreteness, we assume momentarily that $\mathrm{supp}(f) \subseteq$

$[1, y]$ and $\mathrm{supp}(g) \subseteq [1, z]$ with $y = x^\theta$ and $z = x^{1-\theta}$ for some $\theta \in (0, 1)$. We then find that

$$\sum_{n \leqslant x} (f * g)(n) e(\alpha n) = \sum_{\substack{k \leqslant y,\, \ell \leqslant z \\ k\ell \leqslant x}} \sum f(k) g(\ell) e(\alpha k \ell). \tag{23.31}$$

The advantage of this formula is that it transforms the Fourier transform of $f * g$ into a double sum that we can interpret as *an average of many sums*. For instance, we may arrange the summation as

$$\sum_{n \leqslant x} (f * g)(n) e(\alpha n) = \sum_{k \leqslant y} f(k) \sum_{\ell \leqslant \min\{z, x/k\}} g(\ell) e(\alpha k \ell). \tag{23.32}$$

In practice, we do not know much about the function g, so that for a given k we cannot hope to do much better than the trivial upper bound

$$\Big| \sum_{\ell \leqslant \min\{z, x/k\}} g(\ell) e(\alpha k \ell) \Big| \leqslant \sum_{\ell \leqslant \min\{z, x/k\}} |g(\ell)|. \tag{23.33}$$

(Consider for instance the case when $g(\ell) = e(-\alpha\ell)$ and $k = 1$.) However, it turns out that (23.33) can be improved for *most* k, something that we can take advantage of since we are averaging over many values of k.

We begin by noticing that the sum in the left-hand side of (23.33) can be interpreted as the Hermitian inner product over $\mathbb{C}$ of the vectors

$$\vec{g} = (g(\ell))_{\ell=1}^d \quad \text{and} \quad \vec{v}_k = (1_{k\ell \leqslant x} \cdot e(-k\ell\alpha))_{\ell=1}^d,$$

where $d = \lfloor z \rfloor$. The key observation is that if $\alpha \approx a/q$ with large q, then the vectors $\vec{v}_k$ are approximately orthogonal to each other, so that the fixed vector $\vec{g}$ cannot correlate strongly with many of them. Consequently, we expect that the trivial bound (23.33) can be improved significantly for most values of k.

To see the claim that the vectors $\vec{v}_k$ are mutually quasi-orthogonal, note that relation (23.23) implies the estimate

$$\langle \vec{v}_{k_1}, \vec{v}_{k_2} \rangle = \sum_{\ell \leqslant \min\{z, x/k_1, x/k_2\}} e(-k_1\ell\alpha)\overline{e(-k_2\ell\alpha)} \ll \frac{1}{\|(k_2 - k_1)\alpha\|}. \tag{23.34}$$

Generalizing the argument used to prove Theorem 23.5, we will show that if α is far from fractions of small denominator, the quantity $\|(k_2 - k_1)\alpha\|$ is away from 0 for most pairs (k_1, k_2) with $k_1 \neq k_2$, so that $\langle \vec{v}_{k_1}, \vec{v}_{k_2} \rangle$ is small.

The above ideas will be vastly generalized in Chapter 25, where we study bounds for general bilinear sums $\sum_{m=1}^M \sum_{n=1}^N a_{m,n} x_m y_n$. We will prove there that there is some Δ that depends at most on the coefficients $a_{m,n}$ such that

$$\Big| \sum_{m=1}^M \sum_{n=1}^N a_{m,n} x_m y_n \Big| \leqslant \Delta \cdot \Big(\sum_{m=1}^M |x_m|^2 \Big)^{1/2} \Big(\sum_{n=1}^N |y_n|^2 \Big)^{1/2}.$$

For now, we use this circle of ideas to derive a strong bound for the Fourier transform of type II functions. The notation $\|f\|_2$ in the statement of Theorem 23.6 stands for the ℓ^2-norm of f, that is to say, $\|f\|_2^2 = \sum_{n\geqslant 1} |f(n)|^2$.

Theorem 23.6. *Let $f, g : \mathbb{N} \to \mathbb{C}$ be two arithmetic functions such that $\mathrm{supp}(f) \subseteq [1, y]$ and $\mathrm{supp}(g) \subseteq [1, z]$. In addition, consider $\alpha \in \mathbb{R}$ and a reduced fraction a/q such that $|\alpha - a/q| \leqslant 1/q^2$. For all $x \geqslant 1$, we have*

$$\sum_{n\leqslant x}(f*g)(n)e(\alpha n) \ll \Big(q + y + z + \frac{yz}{q}\Big)^{1/2}\sqrt{\log(2q)}\cdot\|f\|_2\|g\|_2.$$

In particular, if $yz \leqslant 2x$ and $|f|, |g| \leqslant 1$, so that $\|f\|_2 \leqslant \sqrt{y}$ and $\|g\|_2 \leqslant \sqrt{z}$, then

$$\sum_{n\leqslant x}(f*g)(n)e(n\alpha) \ll \Big(\frac{x}{\sqrt{q}} + \frac{x}{\sqrt{y}} + \frac{x}{\sqrt{z}} + \sqrt{xq}\Big)\sqrt{\log(2q)}.$$

Proof. Let S be the sum we want to bound, which we arrange in the "dual"[3] form to (23.32)

$$S = \sum_{\ell\leqslant z} g(\ell) \sum_{k\leqslant y,\, k\ell\leqslant z} f(k)e(k\ell\alpha).$$

We use the Cauchy-Schwarz inequality to remove the unknown function g:

$$|S|^2 \leqslant \|g\|_2^2 \sum_{\ell\leqslant z}\Big|\sum_{k\leqslant y,\, k\ell\leqslant x} f(k)e(k\ell\alpha)\Big|^2.$$

As a result, the variable ℓ is now weighted with the smooth function 1. Opening the square via the identity $|z|^2 = z\bar{z}$ yields that

$$|S|^2 \leqslant \|g\|_2^2 \sum_{\substack{\ell\leqslant z,\ k_1,k_2\leqslant y\\ k_1\ell,\, k_2\ell\leqslant x}}\sum\sum f(k_1)\bar{f}(k_2)e((k_1-k_2)\ell\alpha)$$

$$= \|g\|_2^2 \sum_{k_1,\,k_2\leqslant y} f(k_1)\bar{f}(k_2) \sum_{\ell\leqslant \min\{z, x/k_1, x/k_2\}} e(\ell\cdot(k_1-k_2)\alpha). \tag{23.35}$$

We bound the innermost sum of (23.35) using (23.34) to find that

$$|S|^2 \ll \|g\|_2^2 \sum_{k_1,\,k_2\leqslant y}\sum |f(k_1)f(k_2)|\cdot\min\Big\{z, \frac{1}{\|(k_2-k_1)\alpha\|}\Big\}.$$

To remove one of the unknown factors $f(k_j)$, we use the inequality $|zw| \leqslant (|z|^2 + |w|^2)/2$. This implies that

$$|S|^2 \ll \|g\|_2^2 \sum_{j\in\{1,2\}}\sum_{k_1,\,k_2\leqslant y}\sum |f(k_j)|^2 \min\Big\{z, \frac{1}{\|(k_2-k_1)\alpha\|}\Big\}.$$

[3]This terminology will be explained in Chapter 25.

The theorem will then follow if we can prove the following estimate:

$$\sum_{k\leqslant y}\min\Big\{z,\frac{1}{\|k\alpha+\eta\|}\Big\}\ll\Big(q+y+z+\frac{yz}{q}\Big)\log(2q), \tag{23.36}$$

uniformly for all $\eta\in\mathbb{R}$. This will be demonstrated by adapting the argument of the proof of Theorem 23.5.

If $q=1$, (23.36) follows by majorizing all summands by z. Let us consider now the more interesting case when $q\geqslant 2$. We let $\tilde{q}=\lfloor q/2\rfloor$ and break the interval $[1,y]$ into subintervals of length $\tilde{q}$ to find that

$$\sum_{k\leqslant y}\min\Big\{z,\frac{1}{\|k\alpha+\eta\|}\Big\}\leqslant\sum_{m=0}^{\lfloor y/\tilde{q}\rfloor}\sum_{k=m\tilde{q}+1}^{(m+1)\tilde{q}}\min\Big\{z,\frac{1}{\|k\alpha+\eta\|}\Big\}. \tag{23.37}$$

Fix $m\in\mathbb{Z}_{\geqslant 0}$. Arguing as in the proof of Theorem 23.5, we find that the numbers $k\alpha+\eta$ are $(2q)^{-1}$-spaced mod 1 when $m\tilde{q}<k\leqslant(m+1)\tilde{q}$. Hence, a straightforward adaptation of the proof of (23.29) implies that

$$\sum_{m\tilde{q}<k\leqslant(m+1)\tilde{q}}\min\Big\{z,\frac{1}{\|k\alpha+\eta\|}\Big\}\leqslant z+\sum_{2\leqslant j\leqslant q/2}\frac{4q}{j-1}\ll z+q\log q.$$

Inserting this bound into (23.37) completes the proof of (23.36), and hence of the theorem. □

Remark 23.7. Remarkably, the estimate for $\sum_{n\leqslant x}(f*g)e(\alpha n)$ supplied by Theorem 23.6 is essentially sharp. For simplicity, we consider only the case when $y\geqslant z$, since the other one is symmetric.

Indeed, let $x\geqslant yz\geqslant x/2$ and choose $f(k)$ to be the complex conjugate of $\sum_{\ell\leqslant z}g(\ell)e(\alpha k\ell)$. We then have

$$\sum_{n\leqslant x}(f*g)e(\alpha n)=\sum_{k\leqslant y}\Big|\sum_{\ell\leqslant z}g(\ell)e(\alpha k\ell)\Big|^2=\|f\|_2^2.$$

If we now let $\{g(\ell)\}_{\ell\leqslant z}$ be a sequence of independent random variables with $\mathbb{P}(g(\ell)=1)=\mathbb{P}(g(\ell)=-1)=1/2$, we find that

$$\mathbb{E}\Big[\sum_{k\leqslant y}\Big|\sum_{\ell\leqslant z}g(\ell)e(k\ell\alpha)\Big|^2\Big]=\sum_{k\leqslant y}\sum_{\ell\leqslant z}1\asymp x.$$

In particular, there must exist a choice of $g(\ell)$ such that $\sum_{n\leqslant x}(f*g)e(\alpha n)=\|f\|_2^2\gg x$. Since $\|g\|_2^2=\lfloor z\rfloor$, we infer that

$$\sum_{n\leqslant x}(f*g)e(\alpha n)=\|f\|_2^2\gg\sqrt{x}\|f\|_2\asymp\sqrt{y}\cdot\|f\|_2\|g\|_2.$$

By swapping the roles of f and g, we can also find choices of them such that $|\sum_{n\leqslant x}(f*g)e(\alpha n)|\gg\sqrt{z}\cdot\|f\|_2\|g\|_2$.

Finally, let us consider the case when $\alpha = a/q$, $f(k) = e(-ak/q)$ for $k \leqslant y$, and $g(\ell) = 1_{\ell \equiv 1 \,(\mathrm{mod}\, q)}$ for $\ell \leqslant z$. We then have

$$\sum_{n \leqslant x} (f * g)e(\alpha n) = \sum_{\substack{k \leqslant y,\, \ell \leqslant z \\ \ell \equiv 1 \,(\mathrm{mod}\, q)}} \sum 1 \asymp y \cdot (z/q + 1) \asymp \sqrt{yz/q + y} \cdot \|f\|_2 \|g\|_2.$$

To conclude, a general estimate for $\sum_{n \leqslant x} (f*g)e(\alpha n)$ can never be better than $\max\{y, z, yz/q\}^{1/2} \|f\|_2 \|g\|_2$, and Theorem 23.6 comes remarkably close to this bound. □

The additive Fourier transform of the primes: Encore

We shall now apply the methods we have developed to establish Vinogradov's famous estimate.

Theorem 23.8. *Let $\alpha \in \mathbb{R}$ and consider a reduced fraction a/q such that $|\alpha - a/q| \leqslant 1/q^2$. For all $x \geqslant 2$, we have*

$$\sum_{n \leqslant x} \Lambda(n) e(n\alpha) \ll \Big(\frac{x}{\sqrt{q}} + x^{4/5} + \sqrt{xq} \Big) (\log x)^{5/2}.$$

Proof. We may assume that $q \leqslant x$; otherwise, the theorem follows by bounding all summands by $\log x$.

Let us decompose Λ using Vaughan's identity. First, we deal with $\Lambda^\sharp$. We apply Theorem 23.5 twice, once to the convolution $\mu_{\leqslant V} * \log$ (so $v = 1$ and $f = \mu_{\leqslant V}$ here, with $y = V$ and $\|f\|_\infty = 1$) and once to $(\mu_{\leqslant V} * \Lambda_{\leqslant U}) * 1$ (so $v = 0$ and $f = \mu_{\leqslant V} * \Lambda_{\leqslant U}$ here, with $y = UV$ and $|f| \leqslant 1 * \Lambda = \log$, whence $\|f\|_\infty \leqslant \log(UV)$). We thus conclude that

$$\sum_{n \leqslant x} \Lambda^\sharp(n) e(n\alpha) \ll \Big(UV + \frac{x}{q} + q \Big) \log^2(xUV). \tag{23.38}$$

Next, we deal with $\Lambda^\flat$. We rewrite this function using (23.12) and apply Theorem 23.6 to each summand $f_j * g_j$ of that identity. Since $q \leqslant x$, $\|f_j\|_2^2 \leqslant 2^j \log^2 x$ and $\|g_j\|_2^2 \leqslant x/2^{j-1}$, we find that

$$\sum_{n \leqslant x} \Lambda^\flat(n) e(n\alpha) \ll \sum_{U < 2^j \leqslant 2x/V} \Big(\frac{x}{\sqrt{q}} + \sqrt{2^j x} + \frac{x}{2^{j/2}} + \sqrt{xq} \Big) (\log x)^{3/2}.$$

We note that $\sqrt{2^j x} \ll x/\sqrt{V}$ and $x/2^{j/2} \ll x/\sqrt{U}$. Applying these bounds to each of the $O(\log x)$ choices of j yields the estimate

$$\sum_{n \leqslant x} \Lambda^\flat(n) e(n\alpha) \ll \Big(\frac{x}{\sqrt{q}} + \frac{x}{\sqrt{U}} + \frac{x}{\sqrt{V}} + \sqrt{xq} \Big) (\log x)^{5/2}. \tag{23.39}$$

Since we also have that $|\sum_{n\leqslant x}\Lambda_{\leqslant U}(n)e(n\alpha)|\leqslant\sum_{n\leqslant U}\Lambda(n)\ll U$ and $q\leqslant\sqrt{xq}$ by our assumption that $q\leqslant x$, Vaughan's identity in combination with (23.38) and (23.39) implies that

$$\sum_{n\leqslant x}\Lambda(n)e(n\alpha)\ll\big(UV+x/\sqrt{q}+x/\sqrt{U}+x/\sqrt{V}+\sqrt{xq}\big)(\log x)^{5/2}.$$

Taking $U=V=x^{2/5}$ to optimize the above bound completes the proof. □

Theorem 23.8 confirms the prediction we made using the Generalized Riemann Hypothesis that the exponential sum $\sum_{n\leqslant x}\Lambda(n)e(\alpha n)$ can only be large when α is close to a rational number with small denominator. Indeed, if $|\alpha-a/q|\leqslant 1/q^2$ with $(\log x)^A\leqslant q\leqslant x/(\log x)^A$, we find that

$$\sum_{n\leqslant x}\Lambda(n)e(\alpha n)\ll_A x/(\log x)^{(A-5)/2}. \tag{23.40}$$

We will demonstrate the utility of this key estimate in the next chapter.

Conclusion

Vinogradov's method allows us to deal with very general sums of the form

$$\sum_{n\leqslant x}a_n\Lambda(n), \tag{23.41}$$

where $(a_n)_{n=1}^{\infty}$ is some interesting sequence. To estimate (23.41), we first use various combinatorial ideas such as convolution identities and Buchstab iterations to obtain an appropriate decomposition of Λ of the form (23.5). We then handle quasi-smooth sums, namely sums of the form $\sum_{n\leqslant x}a_n(f*g)(n)$ with f "small" and g smooth, using a mix of tools such as the summation formulas of Poisson and of Euler-Maclaurin, L-functions, sieves and estimates for exponential sums (e.g. the Pólya-Vinogradov inequality and other more advanced results beyond the scope of this book). Finally, we estimate bilinear sums, namely sums of the form $\sum_{n\leqslant x}a_n(f*g)(n)$ with f and g both supported on large integers, by employing methods arising from the theory of bilinear forms that we will fully develop in Chapter 25 (with the Cauchy-Schwarz inequality playing a central role), coupled with various exponential sum estimates. We thus see that this approach to the distribution of primes utilizes the full toolset we have at our disposal.

Exercises

Exercise 23.1. Consider $\alpha\in\mathbb{R}$ and a reduced fraction a/q such that $|\alpha-a/q|\leqslant 1/q^2$. Prove that

$$\sum_{n\leqslant x}\tau(n)e(\alpha n)\ll(\sqrt{x}+q+x/q)\log x\qquad(x\geqslant 2).$$

Exercise 23.2. Let $v \geqslant 0$, let f be an arithmetic function supported in $[1, y]$, and χ be a non-principal Dirichlet character mod q. For $x \geqslant 2$, prove that

$$\sum_{n \leqslant x} (f * \log^v)(n)\chi(n) \ll \sqrt{q}(\log q)(\log x)^v \sum_{k \leqslant y} |f(k)|.$$

Exercise 23.3. Let $r, s > 1$ be such that $1/r + 1/s = 1$. Assuming the set-up of Theorem 23.5, prove that

$$\sum_{n \leqslant x} (f * \log^v)(n)e(n\alpha) \ll_r \big(y^{1/r}q^{1/s} + q + x/q\big)(\log x)^v \|f\|_s,$$

where $\|f\|_s = (\sum_{k=1}^{\infty} |f(k)|^s)^{1/s}$.

Exercise 23.4.*

(a) For any $U, V \geqslant 1$, prove that

$$\mu = -\mu_{\leqslant U} * \mu_{\leqslant V} * 1 \ + \ \mu_{>U} * \mu_{>V} * 1 \ + \ \mu_{\leqslant U} \ + \ \mu_{\leqslant V}.$$

(b) Let $\alpha \in \mathbb{R}$, and let a/q be a reduced fraction with $|\alpha - a/q| \leqslant 1/q^2$. For every fixed $\varepsilon > 0$, show that

$$\sum_{n \leqslant x} \mu(n)e(n\alpha) \ll_\varepsilon \big(x/\sqrt{q} + \sqrt{xq}\big)(\log x)^3 + x^{4/5+\varepsilon} \qquad (x \geqslant 3).$$

[*Hint:* Select $U = V = \min\{x^{2/5}, q, x/q\}$ in part (a).]

(c) (Davenport) Fix $A \geqslant 1$. Prove that

$$\sum_{n \leqslant x} \mu(n)e(n\alpha) \ll_A x/(\log x)^A \quad (x \geqslant 2, \ \alpha \in \mathbb{R}).$$

Exercise 23.5 (Heath-Brown's identity)**.** Let $k \in \mathbb{N}$, $x \geqslant 1$ and $V \geqslant x^{1/k}$. For $n \leqslant x$, show that

$$\Lambda(n) = \sum_{j=1}^{k} (-1)^{j-1} \binom{k}{j} (\log * \underbrace{1 * \cdots * 1}_{j-1 \text{ times}} * \underbrace{\mu_{\leqslant V} * \cdots * \mu_{\leqslant V}}_{j \text{ times}})(n).$$

[*Hint:* Let $f = \mu_{\leqslant V} * 1$ and $g = \mu_{>V} * 1$. On the one hand, we have $\Lambda * \underbrace{g * \cdots * g}_{k \text{ times}} = 0$ on $\mathbb{N}_{\leqslant x}$. On the other hand, $g = \delta - f$ with $\delta(n) = 1_{n=1}$.]

Chapter 24

Ternary arithmetic progressions

Vinogradov's method allows us to advance significantly our understanding of additive patterns among the primes. We exemplify this principle here by proving the existence of infinitely many ternary arithmetic progressions $a, a+d, a+2d$ all of whose elements are prime numbers. The same ideas can also be used to study the ternary Goldbach conjecture (stating that every odd integer $\geqslant 7$ is the sum of three primes), as well as other similar "ternary additive problems" (see Exercise 24.1, as well as Chapter 26 of Davenport's book [31]). On the contrary, binary additive problems, such as the twin prime conjecture or the binary Goldbach conjecture (every even integer $\geqslant 4$ is the sum of two primes), are generally out of the reach of the currently available methods. We give a brief explanation of the added difficulties when dealing with binary problems in the last section of this chapter.

We now state the main result of this chapter.

Theorem 24.1. *Fix $A > 0$. For $x \geqslant 2$, we have*

$$\sum_{\substack{n_1,n_2,n_3\leqslant x\\ n_2-n_1=n_3-n_2}}\sum\sum \Lambda(n_1)\Lambda(n_2)\Lambda(n_3) = x^2\prod_{p\geqslant 3}\Big(1-\frac{1}{(p-1)^2}\Big)+O_A\Big(\frac{x}{(\log x)^A}\Big).$$

Remark 24.2. Green and Tao [77] proved that, for any given k, there are infinitely many k-step arithmetic progressions $a,\ a+d,\ a+2d,\ \dots,$ $a+(k-1)d$ all of whose elements are prime numbers. Their proof uses techniques related to the celebrated theorem of Szemerédi that lie beyond the scope of this book. This theorem states that if a set of integers $\mathcal{A}$ has positive lower density, in the sense that $\#\mathcal{A}\cap[1,x]\gg x$ for infinitely many

x, then $\mathcal{A}$ contains arbitrarily long arithmetic progressions. Even though the primes do not have positive lower density, Green and Tao established a suitable *transference principle* that allowed them to pass from Szemerédi's theorem to the case when $\mathcal{A}$ is the set of primes. An important step in doing so is the construction of a "sieve majorant" for the indicator function of the primes. □

The Hardy-Littlewood circle method

In a series of papers, Hardy and Littlewood introduced a general technique that gave them access to a wide array of additive problems. They called their approach the *circle method* for reasons that will become apparent shortly. Their ideas were further developed by I. M. Vinogradov and led to Theorem 24.1. We describe them in the context of counting ternary arithmetic progressions.

The starting point of the circle method is the orthogonality relation

$$1_{n=0} = \int_0^1 e(\alpha n)\mathrm{d}\alpha, \tag{24.1}$$

valid for any integer n. Using it with $n = n_1 + n_3 - 2n_2$ allows us to re-express the indicator function of the event that the integers n_1, n_2, n_3 are in arithmetic progression: $1_{n_2-n_1=n_3-n_1} = \int_0^1 e(\alpha n_1)e(\alpha n_3)e(-2\alpha n_2)\mathrm{d}\alpha$. Multiplying both sides by $\Lambda(n_1)\Lambda(n_2)\Lambda(n_3)$ and summing over $n_1, n_2, n_3 \leqslant x$ yields the formula

$$\sum_{\substack{n_1,n_2,n_3\leqslant x\\ n_2-n_1=n_3-n_2}}\sum\sum \Lambda(n_1)\Lambda(n_2)\Lambda(n_3) = \int_0^1 S(x;\alpha)^2 S(x;-2\alpha)\mathrm{d}\alpha, \tag{24.2}$$

where $S(x;\alpha)$ is the additive Fourier transform of Λ, that is to say,

$$S(x;\alpha) := \sum_{n\leqslant x} \Lambda(n)e(\alpha n).$$

The name "circle method" comes from interpreting the expression in (24.2) as an integral over $\mathbb{R}/\mathbb{Z}$ (which is a circle from a geometric point of view).

Formula (24.2) is what we would call a "gambit" in chess. The gain it offers is that it transforms the unknown expression on its left side into a new expression that we can hope to estimate using our knowledge for the sum $S(x;\alpha)$ from the previous chapter. However, to achieve this transformation, we had to make a sacrifice: the trivial bound for the expression on the right-hand side of (24.2) is $\int_0^1 (\sum_{n\leqslant x}\Lambda(n))^3\mathrm{d}\alpha \asymp x^3$. This means that we must somehow recover the loss of a factor of $1/x$. There are two key ideas that

will allow us to compensate for this loss:

- There is a set $\mathfrak{M}$ of $\alpha \in [0,1]$ (called the set of *major arcs*) that "dominate" the integral in (24.2). This set has Lebesgue measure $(\log x)^{O(1)}/x$, and the sum $S(x;\alpha)$ has size $x/(\log x)^{O(1)}$ on it. It thus contributes $\approx x^3 \cdot 1/x = x^2$ to the integral of (24.2), which is the size of the expected main term.
- The size of $S(x;\alpha)$ for a "generic" $\alpha \in [0,1]$ is $x^{1/2+o(1)}$. This follows from Parseval's identity:[1]

$$\int_0^1 |S(x;\alpha)|^2 \mathrm{d}\alpha = \sum_{n\leqslant x} \Lambda(n)^2 \asymp x\log x. \tag{24.3}$$

It turns out that the major arcs consist of those numbers that are close to rationals of small denominator. More precisely, we may take

$$\mathfrak{M} = \bigcup_{\substack{1\leqslant q\leqslant \mathcal{L},\, 0\leqslant a\leqslant q\\ (a,q)=1}} \bigcup [a/q - \mathcal{L}/x, a/q + \mathcal{L}/x] \tag{24.4}$$

with $\mathcal{L} = (\log x)^{2A+7}$. We also define the set of *minor arcs* $\mathfrak{m} := [0,1] \setminus \mathfrak{M}$. The motivation for this terminology is the fact that $S(x;2\alpha) \ll x/(\log x)^{A+1}$ when $\alpha \in \mathfrak{m}$. Indeed, for any such α, Lemma 23.4 implies the existence of a reduced fraction a/q with $q \leqslant x/\mathcal{L}$ and $|\alpha - a/q| \leqslant \mathcal{L}/(qx) \leqslant 1/q^2$. Since $\alpha \notin \mathfrak{M}$, we must have $q > \mathcal{L}$. Hence, the claimed bound on $S(x;2\alpha)$ follows by (23.40). Together with (24.3), this implies that

$$\int_{\mathfrak{m}} |S(x;\alpha)^2 S(x;-2\alpha)| \mathrm{d}\alpha \ll \int_0^1 |S(x;\alpha)|^2 \frac{x}{(\log x)^{A+1}} \mathrm{d}\alpha \ll \frac{x^2}{(\log x)^A}.$$

Thus most of the contribution to the right side of (24.2) comes from $\alpha \in \mathfrak{M}$.

It remains to estimate $S(x;\alpha)$ when $\alpha \in \mathfrak{M}$. Consider first the case when $\alpha = a/q$. Then, we may argue as in (23.19) to prove that

$$\sum_{n\leqslant x} \Lambda(n) e(an/q) = \frac{\mu(q)}{\varphi(q)}\, x + O_A\big(xe^{-c\sqrt{\log x}}\big) \tag{24.5}$$

for some $c > 0$; indeed, here $q \leqslant (\log x)^{2A+7}$, so the Siegel-Walfisz theorem is applicable and we do not need to appeal to the unproven Generalized Riemann Hypothesis. Finally, if $\alpha = \beta + a/q$ with $|\beta| \leqslant \mathcal{L}/x$, we may adapt the proof of (23.21) to pass from (24.5) to an estimate for $S(x;\alpha)$.

Putting together the above estimates leads to a proof of Theorem 24.1. The details we have omitted are presented in Chapter 26 of Davenport's book [31] (the treatment there concerns the ternary Goldbach conjecture, but it can be easily adapted to our setting). Here, we will give a different

[1] Relation (24.3) admits a short self-contained proof: we use the identity $|z|^2 = z\bar{z}$ to write $|S(x;\alpha)|^2$ as a double sum over $n_1, n_2 \leqslant x$, and then note that $\int_0^1 e(\alpha(n_1 - n_2))\mathrm{d}\alpha = 1_{n_1=n_2}$.

way of calculating the main term of Theorem 24.1. Before proceeding, let us pause for a moment to observe the amazing complementarity of the available estimates: Vinogradov's method can handle $S(x;\alpha)$ when $|\alpha - a/q| \leqslant 1/q^2$ with $q \geqslant (\log x)^{2A+7}$, and stops working well for smaller q. However, the remaining range of q is precisely what is covered by the best-known version of the Prime Number Theorem for arithmetic progressions. This allows us to handle the full range of integration and prove Theorem 24.1 unconditionally.

Making the entire circle a minor arc

Given arithmetic functions f_1, f_2, f_3, we define the weighted count of ternary arithmetic progressions

$$\mathcal{T}(f_1, f_2, f_3; x) = \sum\sum\sum_{\substack{n_1,n_2,n_3\leqslant x\\ n_1+n_3=2n_2}} f_1(n_1)f_2(n_2)f_3(n_3).$$

Our goal is to estimate $\mathcal{T}(\Lambda, \Lambda, \Lambda; x)$. We will extract the main term to this quantity by decomposing Λ.

Rather than employing Vaughan's identity, it is more convenient to use (23.14). This leads us to the formula

$$\mathcal{T}(\Lambda, \Lambda, \Lambda; x) = \sum\sum\sum_{f_1,f_2,f_3\in\{\Lambda^\sharp_{\mathrm{sieve}},E\}} \mathcal{T}(f_1, f_2, f_3; x) \tag{24.6}$$

with $E = \Lambda^\flat_{\mathrm{sieve}} + R_{\mathrm{sieve}}$. If y and D are chosen as in (23.18), we will show that any summand involving at least one term $f_j = E$ is negligible. The key to proving this is the *aperiodicity* of the function $\Lambda^\flat_{\mathrm{sieve}}$ that manifests itself in Theorem 24.3 below. A close examination of its proof reveals another instance of the complementarity of Vinogradov's method and of the Siegel-Walfisz theorem to which we alluded above (though it is subtler now, with the Siegel-Walfisz theorem hiding within the proof of Corollary 13.4).

Theorem 24.3. *Consider $x, y, D \geqslant 2$ satisfying* (23.18) *with $0 < \theta_1 < \theta_2 < 1$, and fix $A \geqslant 1$. If $\Lambda^\flat_{\mathrm{sieve}}$ is defined by* (23.16)*, then for all $\alpha \in \mathbb{R}$ we have*

$$\sum_{n\leqslant x} \Lambda^\flat_{\mathrm{sieve}}(n)e(\alpha n) \ll_{\theta_1,\theta_2,A} \frac{x}{(\log x)^A}.$$

Proof. All implied constants might depend on θ_1, θ_2 and A. Applying Lemma 23.4 with $Q = x/(\log x)^{2A+5}$, we may find a reduced fraction a/q such that $1 \leqslant q \leqslant Q$ and $|\alpha - a/q| \leqslant 1/(qQ)$. We use a different argument according to whether $q \leqslant (\log x)^{2A+5}$ or not.

First, we study the case when $(\log x)^{2A+5} \leqslant q \leqslant x/(\log x)^{2A+5}$. We begin by writing $\Lambda^\flat_{\mathrm{sieve}} = \tilde{f} * \tilde{g}$ with $\tilde{f}(k) = 1_{k>D,\, P^-(k)>y}\, \mu(k)$ and $\tilde{g}(\ell) = 1_{\ell>y,\, P^-(\ell)>y} \log \ell$. We then argue as in the proof of (23.39): we localize

dyadically the factors of the convolution $\tilde{f} * \tilde{g}$ as in (23.12) and then apply Theorem 23.6 to each summand separately. This yields the estimate

$$\sum_{n\leqslant x} \Lambda^\flat_{\mathrm{sieve}}(n)e(n\alpha) \ll \Big(\frac{x}{\sqrt{q}} + \frac{x}{\sqrt{y}} + \frac{x}{\sqrt{D}} + \sqrt{xq}\Big)(\log x)^{5/2}.$$

The right-hand is $\ll x/(\log x)^A$ by our choice of q, y and D, thus completing the proof of the theorem in this case.

Finally, we consider the case when $q \leqslant (\log x)^{2A+5}$. We begin by studying the case $\alpha = a/q$. For each $w \in [\sqrt{x}, x]$, we have

$$\begin{aligned}\sum_{n\leqslant w} \Lambda^\flat_{\mathrm{sieve}}(n)e(na/q) &= \sum_{b\in\mathbb{Z}/q\mathbb{Z}} e(ba/q) \sum_{\substack{n\leqslant w\\ n\equiv b\,(\mathrm{mod}\,q)}} \Lambda^\flat_{\mathrm{sieve}}(n)\\ &= \sum_{b\in\mathbb{Z}/q\mathbb{Z}} e(ba/q) \sum_{\substack{y<\ell\leqslant w/D\\ P^-(\ell)>y}} \log\ell \sum_{\substack{D<k\leqslant w/\ell,\, P^-(k)>y\\ k\ell\equiv b\,(\mathrm{mod}\,q)}} \mu(k). \qquad (24.7)\end{aligned}$$

Fix ℓ and b momentarily, and let $d = (\ell, q)$. The congruence $k\ell \equiv b\,(\mathrm{mod}\,q)$ is equivalent to having $d|b$ and $k \equiv \ell' b/d\,(\mathrm{mod}\,q/d)$, where ℓ' denotes the inverse of ℓ/d mod q/d. Since $w/\ell \geqslant D = \exp\{(\log x)^{\theta_2}\} = \exp\{(\log y)^{\theta_2/\theta_1}\}$ and $q \leqslant (\log x)^{2A+5} = (\log D)^{(2A+5)/\theta_2}$, Corollary 13.4 applied with $m = \prod_{p\leqslant y} p$ implies that the innermost sum of (24.7) is $\ll (w/\ell)/(\log x)^{5A+12}$. Thus,

$$\sum_{n\leqslant w} \Lambda^\flat_{\mathrm{sieve}}(n)e(na/q) \ll \sum_{b\in\mathbb{Z}/q\mathbb{Z}} \sum_{\ell\leqslant w} \log\ell \cdot \frac{w/\ell}{(\log x)^{5A+12}} \ll \frac{w}{(\log x)^{3A+5}}$$

uniformly for $w \in [\sqrt{x}, x]$ and $q \leqslant (\log x)^{2A+5}$. To pass to an estimate for $\sum_{n\leqslant x} \Lambda^\flat_{\mathrm{sieve}}(n)e(n\alpha)$, we write $\alpha = a/q + \beta$, so that $|\beta| \leqslant (\log x)^{2A+5}/x$. Using partial summation similarly to (23.20) implies that

$$\sum_{\sqrt{x}<n\leqslant x} \Lambda^\flat_{\mathrm{sieve}}(n)e(n\alpha) = \int_{\sqrt{x}}^x e(\beta w)\mathrm{d}\sum_{n\leqslant w} \Lambda^\flat_{\mathrm{sieve}}(n)e(na/q) \ll \frac{x}{(\log x)^A}.$$

Since we also have the trivial bound $\sum_{n\leqslant\sqrt{x}} \Lambda^\flat_{\mathrm{sieve}}(n)e(n\alpha) \ll \sqrt{x}\log^2 x$ by noticing that $|\Lambda^\flat_{\mathrm{sieve}}| \leqslant \log$, the theorem follows. □

Let us now prove our claim that any summand on the right-hand side of (24.6) with $f_j = E$ for some j is negligible. For concreteness, assume that $f_3 = E$; the other cases follow similarly. Arguing as in (24.2), we have

$$\mathcal{T}(f_1, f_2, f_3; x) = \int_0^1 S_{f_1}(x;\alpha)S_{f_2}(x;-2\alpha)S_{f_3}(x;\alpha)\mathrm{d}\alpha,$$

where $S_f(x;\alpha) = \sum_{n\leqslant x} f(n)e(\alpha n)$ denotes the additive Fourier transform of the arithmetic function f. Since we have assumed that $f_3 = E$, Theorem 24.3 implies that $S_{f_3}(x;\alpha) = O_A(x/(\log x)^{A+2})$ for *all* $\alpha \in \mathbb{R}$ (i.e., there

are no "major arcs" anymore). Consequently,

$$\begin{aligned}\mathcal{T}(f_1,f_2,f_3;x) &\ll \frac{x}{(\log x)^{A+2}}\int_0^1 |S_{f_1}(x;\alpha)S_{f_2}(x;-2\alpha)|\mathrm{d}\alpha\\ &\leqslant \frac{x}{(\log x)^{A+2}}\Big(\int_0^1 |S_{f_1}(x;\alpha)|^2\mathrm{d}\alpha\int_0^1 |S_{f_2}(x;-2\alpha)|^2\mathrm{d}\alpha\Big)^{1/2}\end{aligned}$$

from the the Cauchy-Schwarz inequality. Parseval's identity implies that $\int_0^1 |S_f(x;k\alpha)|^2\mathrm{d}\alpha = \sum_{n\leqslant x}|f(n)|^2$ for any $k\in\mathbb{Z}\setminus\{0\}$ (see (24.3)). When $f\in\{\Lambda^\sharp_{\mathrm{sieve}},E\}$, we have $|f|\leqslant 2\log$ and thus $\sum_{n\leqslant x}|f(n)|^2\ll x(\log x)^2$. To conclude, we have proved that

$$\mathcal{T}(f_1,f_2,f_3;x)\ll_A x^2/(\log x)^A \qquad \text{when there is } f_j=E.$$

This reduces Theorem 24.1 to proving the following estimate.

Proposition 24.4. *Assuming the set-up of Theorem* 24.3*, and with* $\Lambda^\sharp_{\mathrm{sieve}}$ *defined by* (23.15)*, we have*

$$\mathcal{T}(\Lambda^\sharp_{\mathrm{sieve}},\Lambda^\sharp_{\mathrm{sieve}},\Lambda^\sharp_{\mathrm{sieve}};x) = x^2\prod_{p\geqslant 3}\Big(1-\frac{1}{(p-1)^2}\Big)+O_A\Big(\frac{x^2}{(\log x)^A}\Big).$$

Proof. As we will see, the proof boils down to a lattice-point counting estimate. For technical reasons to be explained later, we break the summation into short intervals. To this end, set $\eta = 1/(\log x)^{A+5}$ and let J be the largest integer such that $(1-\eta)^J\geqslant 1/\sqrt{x}$. Since $|\Lambda^\sharp_{\mathrm{sieve}}|\leqslant\log x$, we have

$$\mathcal{T}(\Lambda^\sharp_{\mathrm{sieve}},\Lambda^\sharp_{\mathrm{sieve}},\Lambda^\sharp_{\mathrm{sieve}};x) = \sum\sum\sum_{\substack{x(1-\eta)^{J+1}<n_1,n_3\leqslant x\\ 2n_2=n_1+n_3}}\prod_{1\leqslant j\leqslant 3}\Lambda^\sharp_{\mathrm{sieve}}(n_j)+O(x^{1.6}).$$

If we cover the range of n_1 and n_3 by intervals of the form $(x(1-\eta)^{j+1},x(1-\eta)^j]$, where $j\in\mathbb{Z}\cap[0,J]$, then the theorem is reduced to proving that

$$\sum\sum\sum_{\substack{x_j(1-\eta)<n_j\leqslant x_j\ (j=1,3)\\ 2n_2=n_1+n_3}}\prod_{1\leqslant j\leqslant 3}\Lambda^\sharp_{\mathrm{sieve}}(n_j) = \rho\eta^2x_1x_3(1+O(1/(\log x)^A)) \tag{24.8}$$

for $x_1,x_3\in[\sqrt{x},x]$, where $\rho=\prod_{3\leqslant p\leqslant y}(1-1/(p-1)^2)$.

Fix x_1,x_3 as above and let $T(x_1,x_3)$ denote the sum in (24.8). We have

$$T(x_1,x_3) = \sum\sum\sum_{k_j\leqslant D,\,P^-(k_j)>y\ \forall j}\mu(k_1)\mu(k_2)\mu(k_3)L(k_1,k_2,k_3), \tag{24.9}$$

where

$$L(k_1,k_2,k_3) := \sum\sum\sum_{\substack{x_j(1-\eta)/k_j<\ell_j\leqslant x/k_j\ (j=1,3)\\ 2k_2\ell_2=k_1\ell_1+k_3\ell_3,\,P^-(\ell_j)>y\ \forall j}}(\log\ell_1)(\log\ell_2)(\log\ell_3).$$

We fix $k_1, k_2, k_3 \leqslant D$ free of prime factors $\leqslant y$ and estimate $L(k_1, k_2, k_3)$. Since the variables k_j are weighted with the Möbius function, it suffices to consider the case when they are all square-free.

We rewrite $L(k_1, k_2, k_3)$ in the notation of Chapter 18. First of all, we remove the logarithmic weights from the variables ℓ_j using the fact that we have restricted them into short intervals. Indeed, the conditions $x_j(1-\eta)/k_j < \ell_j \leqslant x/k_j$ for $j = 1, 3$ imply that $x_2(1-\eta)/k_2 < \ell_2 \leqslant x_2/k_2$ with $x_2 = (x_1 + x_3)/2$. We thus have $\log \ell_j = \log(x_j/k_j) + O(\eta) = (1 + O(\eta/\log x)) \log(x_j/k_j)$ for each j. As a consequence,

$$L(k_1, k_2, k_3) = (1 + O(\eta/\log x)) S(\mathcal{W}, \mathcal{P}) \prod_{1 \leqslant j \leqslant 3} \log(x_j/k_j),$$

where $\mathcal{P} = \{3 \leqslant p \leqslant y\}$ and $\mathcal{W} = (w_n)_{n=1}^{\infty}$ is the sequence of weights

$$w_n = 1_{2 \nmid n} \cdot \#\Big\{ (\ell_1, \ell_2, \ell_3) \in \mathbb{Z}^3 : \begin{array}{l} \ell_1\ell_2\ell_3 = n,\ 2k_2\ell_2 = k_1\ell_1 + k_3\ell_3, \\ x_j(1-\eta)/k_j < \ell_j \leqslant x/k_j \quad (j = 1, 3) \end{array} \Big\}$$

To estimate $S(\mathcal{W}, \mathcal{P})$, we apply Theorem 18.11(a). We must first check Axioms 1–3.

Given $d | \mathcal{P}$, we set $W_d = \sum_{n \equiv 0 \,(\mathrm{mod}\, d)} w_n$. Then,

$$W_d = \#\Big\{ (\ell_1, \ell_2, \ell_3) \in \mathbb{Z}^3 : \begin{array}{l} 2 \nmid \ell_1\ell_2\ell_3,\ d | \ell_1\ell_2\ell_3,\ 2k_2\ell_2 = k_1\ell_1 + k_3\ell_3, \\ x_j(1-\eta)/k_j < \ell_j \leqslant x/k_j \quad (j = 1, 3) \end{array} \Big\}.$$

To estimate W_d, we first split the range of the pairs (ℓ_1, ℓ_3) according to their reduction mod 4, mod k_2 and mod d (which are mutually coprime integers). The permissible reductions lie in the sets

$$\mathcal{A} = \{ (a_1, a_3) \in (\mathbb{Z}/4\mathbb{Z})^2 : k_1a_1 + k_3a_3 \equiv 2 \,(\mathrm{mod}\, 4),\ a_1a_3 \equiv 1 \,(\mathrm{mod}\, 2) \},$$
$$\mathcal{B} = \{ (b_1, b_3) \in (\mathbb{Z}/k_2\mathbb{Z})^2 : k_1b_1 + k_3b_3 \equiv 0 \,(\mathrm{mod}\, k_2) \},$$
$$\mathcal{C} = \{ (c_1, c_3) \in (\mathbb{Z}/d\mathbb{Z})^2 : c_1c_3(k_1c_1 + k_3c_3) \equiv 0 \,(\mathrm{mod}\, d) \}.$$

Given $(a_1, a_3) \in \mathcal{A}$, $(b_1, b_3) \in \mathcal{B}$, $(c_1, c_3) \in \mathcal{C}$ and $(\ell_1, \ell_3) \in \mathbb{Z}^2$ such that

$$\ell_j \equiv a_j \,(\mathrm{mod}\, 4), \quad \ell_j \equiv b_j \,(\mathrm{mod}\, k_2), \quad \ell_j \equiv c_j \,(\mathrm{mod}\, d) \quad (j = 1, 3), \tag{24.10}$$

the equation $2k_2\ell_2 = k_1\ell_1 + k_3\ell_3$ has a unique solution ℓ_2 which is necessarily an odd integer. In addition, the number of pairs (ℓ_1, ℓ_3) that satisfy (24.10) and the inequalities $x_j(1-\eta) < \ell_j \leqslant x_j$ for $j = 1, 3$ equals

$$\prod_{j \in \{1,3\}} \Big(\frac{\eta x_j}{4k_jk_2d} + O(1) \Big) = \frac{\eta^2 x_1 x_3}{16k_1k_2^2k_3d^2} + O\Big(\frac{\eta x}{k_2 \min\{k_1, k_3\} d} + 1 \Big).$$

Therefore,

$$W_d = \Big(\frac{\eta^2 x_1 x_3}{16k_1k_2^2k_3d^2} + O\Big(\frac{\eta x}{k_2 \min\{k_1, k_3\} d} + 1 \Big) \Big) \cdot |\mathcal{A}| \cdot |\mathcal{B}| \cdot |\mathcal{C}|.$$

We easily see that $|\mathcal{A}| = 2$. In addition, since d is square-free and coprime to k_1k_3, the Chinese Remainder Theorem implies that $|\mathcal{C}| = \prod_{p|d}(3p-2)$. By a similar argument, $|\mathcal{B}| = \prod_{p|k_2,\, p\nmid(k_1,k_3)} p \prod_{p|(k_1,k_2,k_3)} p^2 = k_2 \cdot (k_1,k_2,k_3)$, since k_2 is square-free. Putting everything together, we conclude that

$$W_d = \frac{\eta^2 x_1 x_3}{8d^2} \cdot \frac{(k_1,k_2,k_3)}{k_1k_2k_3} \prod_{p|d}(3p-2) + O\big(3^{\omega(d)}(\eta x + k_2^2 d)\big).$$

Therefore, Axiom 1 holds with $X = \frac{\eta^2 x_1 x_3}{8} \cdot \frac{(k_1,k_2,k_3)}{k_1k_2k_3}$, $\nu(d) = \prod_{p|d}(3-2/p)$ and $r_d \ll 3^{\omega(d)}(\eta x + k_2^2 d)$. In addition, Axiom 2 holds with $\kappa = 3$, and Axiom 3 with $D = x^{1/2}$, $A = \kappa + 1$ and $m = 1$, since $k_1,\ k_2,\ k_3 \leqslant \exp\{(\log x)^{\theta_2}\} = x^{o(1)}$ here. Hence, Theorem 18.11(a) implies that

$$L(k_1,k_2,k_3) = \Big(1 + O_A\Big(\frac{\eta}{\log x}\Big)\Big) \cdot \lambda \eta^2 x_1 x_3 \cdot \frac{(k_1,k_2,k_3)}{k_1k_2k_3} \prod_{j\in\{1,2,3\}} \log(x_j/k_j),$$

where $\lambda = 8^{-1}\prod_{3\leqslant p\leqslant y}(1-1/p)(1-2/p)$. Together with (24.9), this gives

$$T(x_1,x_3) = \lambda\eta^2 x_1x_3M + O_A(\eta^3 x_1x_3(\log x)^5), \tag{24.11}$$

where

$$M := \sum\sum\sum_{k_j\leqslant D,\, P^-(k_j)>y\,\forall j} \frac{(k_1,k_2,k_3)}{k_1k_2k_3} \prod_{j\in\{1,2,3\}} \mu(k_j)\log(x_j/k_j)$$

and in the calculation of the error term of (24.11) we used the bound $\sum_{k_1,k_2,k_3\leqslant D} \frac{(k_1,k_2,k_3)}{k_1k_2k_3} \ll (\log x)^3$, which can be seen by letting $g = (k_1,k_2,k_3)$ and $k_j = gn_j$. Since $\eta = 1/(\log x)^{A+5}$, (24.11) reduces (24.8) to proving that

$$M = \prod_{p\leqslant y}(1-1/p)^{-3} + O_A(1/(\log x)^A). \tag{24.12}$$

We claim that we may replace (k_1,k_2,k_3) by 1 in all summands of M at the cost of a small error term. Indeed, if $g = (k_1,k_2,k_3) > 1$, then $g > y$ since $P^-(g) > y$. Hence, if we write $k_j = gn_j$, then $(g-1)/(k_1k_2k_3) \leqslant 1_{g>y}/(g^2n_1n_2n_3)$. Since $\sum_{g>y} 1/g^2 \ll 1/y$, we find that

$$M = \sum\sum\sum_{k_j\leqslant D,\, P^-(k_j)>y\,\forall j} \prod_{j=1}^{3} \frac{\mu(k_j)\log(x_j/k_j)}{k_j} + O((\log x)^6/y). \tag{24.13}$$

If we let $M_j = \sum_{k_j\leqslant D,\, P^-(k_j)>y} k_j^{-1}\mu(k_j)\log(x_j/k_j)$, then the main term of (24.13) factors as $M_1M_2M_3$. By Corollary 13.4, we have the estimate $\sum_{k\leqslant w,\, P^-(k)>y}\mu(k) \ll w/(\log w)^{A+2}$ for $w \geqslant D$. Hence, partial summation implies that

$$M_j = (\log x_j)\cdot(1/\zeta_y)(1) + (1/\zeta_y)'(1) + O_A(1/(\log x)^A)$$

with $\zeta_y(s) = \sum_{P^-(k)>y} k^{-s}$ defined as in Chapter 22. Since $\zeta_y(s) \sim (s-1)^{-1}\prod_{p\leqslant y}(1-1/p)$ and $\zeta_y'(s) \sim (s-1)^{-2}\prod_{p\leqslant y}(1-1/p)$ as $s \to 1^+$, we infer that $(1/\zeta_y)(1) = 0$ and $(1/\zeta_y)'(1) = \prod_{p\leqslant y}(1-1/p)^{-1}$. Relation (24.12) then follows, thus completing the proof of (24.8), and hence of the theorem. □

Binary additive problems

It is natural to wonder whether the circle method can be used to approach binary additive problems such as the twin prime conjecture. For this problem too we have a formula analogous to (24.2):

$$\sum_{n\leqslant x}\Lambda(n)\Lambda(n+2) = \int_0^1 |S(x;\alpha)|^2 e(-2\alpha)\mathrm{d}\alpha. \tag{24.14}$$

We still expect the main term to come from the major arcs. As a matter of fact, it can be shown rigorously that if we define $\mathfrak{M}$ by (24.4), then

$$\int_{\mathfrak{M}} |S(x;\alpha)|^2 e(-2\alpha)\mathrm{d}\alpha \sim c_2 x,$$

where $c_2 = 2\prod_{p\geqslant 3}(1-1/(p-1)^2)$ is the twin prime constant. However, it is not known how to show that the minor arcs contribute a negligible amount. Indeed, as we discussed before, Parseval's identity (24.3) implies that $S(x;\alpha) = x^{1/2+o(1)}$ for a "generic" $\alpha \in [0,1]$. Hence, no matter how we choose $\mathfrak{m}$, we cannot expect to have a better bound than

$$\int_{\mathfrak{m}} |S(x;\alpha)|^2 \mathrm{d}\alpha \ll x^{1+o(1)},$$

which is of comparable size to the expected main term. This means that in order to prove the twin prime conjecture using the circle method, we must exploit cancellation between the various parts of the integral in (24.14) coming from the factor $e(-2\alpha)$.

Exercises

Exercise 24.1. If N is an odd integer, prove that

$$\sum\sum\sum_{n_1+n_2+n_3=N}\Lambda(n_1)\Lambda(n_2)\Lambda(n_3) = G_N N^2 + O_A(N^2/(\log N)^A)$$

for each fixed $A \geqslant 1$, where $G_N = \prod_{p\mid N}(1-1/(p-1)^2)\prod_{p\nmid N}(1+1/(p-1)^3)$.

Exercise 24.2. If $\Lambda^\sharp_{\mathrm{sieve}}$ is as in Proposition 24.4, prove that

$$\sum_{n\leqslant x}\Lambda^\sharp_{\mathrm{sieve}}(n)\Lambda^\sharp_{\mathrm{sieve}}(n+2) = c_2 x + O_A(x/(\log x)^A)$$

for every fixed $A \geqslant 1$, where $c_2 = 2\prod_{p\geqslant 3}(1-1/(p-1)^2)$.

Chapter 25

Bilinear forms and the large sieve

Let us suppose we are given a sequence of complex numbers $(c_n)_{n\in\mathbb{Z}}$ and we wish to study its distribution among the different arithmetic progressions mod q. To be more precise, our aim is to determine the behavior of the sum

$$\sum_{H<n\leqslant H+N} c_n 1_{n\equiv j \,(\mathrm{mod}\, q)}$$

when j varies over $\mathbb{Z}/q\mathbb{Z}$, with $N \geqslant 1$ and H being two given integers. To this end, we consider the additive and multiplicative Fourier transform mod q of the above sum (viewed as a function of j). These are given by

$$S^+(a/q) = \sum_{H<n\leqslant H+N} c_n e(an/q) \quad \text{and} \quad S^\times(\chi) = \sum_{H<n\leqslant H+N} c_n \chi(n),$$

with a running over $\mathbb{Z}/q\mathbb{Z}$ and χ over all Dirichlet characters mod q.

For a general sequence $(c_n)_{n\in\mathbb{Z}}$, we cannot obtain a non-trivial pointwise bound for $S^+(a/q)$. Indeed, if $c_n = e(-na/q)$ for all n, then $S^+(a/q) = N$. A similar obstruction holds for $S^\times(\chi)$, by taking $c_n = \overline{\chi}(n)$ for some Dirichlet character χ. However, we will prove non-trivial bounds on $S^+(a/q)$ and on $S^\times(\chi)$ when we average over many a and q, or many χ.

With the above goal in mind, we consider the sums

$$\sum_{q\leqslant Q} \sum_{a\in(\mathbb{Z}/q\mathbb{Z})^*} |S^+(a/q)|^2 \quad \text{and} \quad \sum_{q\leqslant Q}{\sum_{\chi \,(\mathrm{mod}\, q)}}^* \frac{q}{\varphi(q)} |S^\times(\chi)|^2, \tag{25.1}$$

where the notation $\sum^*$ means that the last sum runs over primitive characters. A few remarks are in order about the shape of these sums:

- The factor $q/\varphi(q)$ is of order 1 most of the time, so it can be ignored. We include it for normalization purposes that we will explain later.
- We sum only over reduced residues $a \pmod q$ because we want the fractions a/q to be distinct mod 1. To see why working with the full sum $T^+ = \sum_{q\leqslant Q}\sum_{a\in\mathbb{Z}/q\mathbb{Z}} |S^+(a/q)|^2$ is problematic, consider the case when $c_n = e(-n/3)$, so that $S^+(1/3) = N$. The sum T^+ contains $\lfloor Q/3\rfloor$ copies of $|S^+(1/3)|^2$; one for each $q \leqslant Q$ that is a multiple of 3. In particular, $T^+ \geqslant \lfloor Q/3\rfloor \cdot N^2$. We thus see that the fact that $(c_n)_{H<n\leqslant H+N}$ is not well distributed with respect to a single modulus causes T^+ to be very large. For a similar reason, we work exclusively with primitive characters instead of working with the full sum $\sum_{q\leqslant Q}\sum_{\chi \,(\mathrm{mod}\, q)}(q/\varphi(q))|S^\times(\chi)|^2$.
- We consider a second moment of $S^+(a/q)$ and of $S^\times(\chi)$ (i.e., an average of squares) because this gives us access to L^2-techniques coming from the theory of bilinear forms. We develop this theory in the next section and use it to study the sums of (25.1) in the subsequent section.

Bilinear forms

An $M \times N$ *bilinear form* over $\mathbb{C}$ is a function $\psi : \mathbb{C}^M \times \mathbb{C}^N \to \mathbb{C}$ that is linear in both coordinates. This is equivalent to the existence of certain coefficients $a_{m,n} \in \mathbb{C}$ such that

$$\psi(\vec{x},\vec{y}) = \sum_{m=1}^{M}\sum_{n=1}^{N} a_{m,n}x_m y_n \tag{25.2}$$

for all $\vec{x} = (x_1,\dots,x_M) \in \mathbb{C}^M$ and all $\vec{y} = (y_1,\dots,y_N) \in \mathbb{C}^N$.

Before we go into more abstract matters, let us discuss a few examples that illustrate the central role of bilinear forms in analytic number theory.

Example 25.1. (a) If we take $a_{m,n} = 1_{mn\leqslant x}\cdot e(mn\alpha)$, then $\psi(\vec{x},\vec{y})$ is equal to $\sum_{n\leqslant x}(f*g)(n)e(n\alpha)$, where $f(m) = 1_{[1,M]}(m)x_m$ and $g(n) = 1_{[1,N]}(n)y_n$ (see relation (23.31)).

(b) If $\{\alpha_1,\dots,\alpha_M\}$ denotes the set of reduced fractions a/q with $1 \leqslant a \leqslant q \leqslant Q$ (often called the *Farey fractions* of order $\leqslant Q$), H is some fixed integer, and we set $a_{m,n} = e(\alpha_m(n+H))$, then

$$\psi(\vec{x},\vec{y}) = \sum_{m=1}^{M} x_m \sum_{n=1}^{N} y_n e((n+H)\alpha_m).$$

In particular, when $x_m = \sum_{n=1}^N \bar{y}_n e(-(n+H)\alpha_m)$ for all m, we have

$$\psi(\vec{x},\vec{y}) = \sum_{m=1}^M \Big| \sum_{n=1}^N y_n e((n+H)\alpha_m)\Big|^2.$$

Letting $y_n = c_{n+H}$, we see that $\psi(\vec{x},\vec{y})$ is equal to the first average of (25.1). Hence, pointwise bounds on $\psi(\vec{x},\vec{y})$ supply information about the distribution of $(c_n)_{H<n\leqslant H+N}$ in arithmetic progressions.

(c) We can state both of the above examples in unified notation: let $H \in \mathbb{Z}$, $M, N \in \mathbb{Z}_{\geqslant 1}$ and $\{\alpha_1, \dots, \alpha_M\}$ be a set of real numbers. In addition, for each $m \in [1, M] \cap \mathbb{Z}$, let I_m be a subinterval of $[1, N]$. We then consider the bilinear form with $a_{m,n} = e(\alpha_m(n+H)) \cdot 1_{I_m}(n)$.

We immediately see that the bilinear form of part (b) can be recast in this notation. On the other hand, if $\alpha_m = m\alpha$ and $I_m = [1, x/m] \cap [1, N]$, then we recover the bilinear form of part (a). □

Example 25.2. Let $\mathcal{C} = \{\chi_1 \,(\mathrm{mod}\, q_1), \dots, \chi_M \,(\mathrm{mod}\, q_M)\}$ be a set of Dirichlet characters, and consider the bilinear form with coefficients $a_{m,n} = \chi_m(n+H)\sqrt{q_m/\varphi(q_m)}$. (The factor $\sqrt{q_m/\varphi(q_m)}$ normalizes the vectors $(a_{m,n})_{n=1}^N$ to all have roughly the same ℓ^2-norm.) Arguing as in Example 25.1(b), we see that this form is related to the second average of (25.1) (with $\mathcal{C}$ being the set of primitive Dirichlet characters of conductor $\leqslant Q$). □

Example 25.3. Bilinear forms can be generalized to multilinear forms in a straightforward way. For instance, given $k \in \mathbb{Z}$, let $a_{m,n,q} = 1_{(q,k)=1} \cdot (1_{mn\equiv k \,(\mathrm{mod}\, q)} - 1_{(mn,q)=1}/\varphi(q))$ and consider the $M \times N \times Q$ trilinear form

$$\begin{aligned}\psi(\vec{x},\vec{y},\vec{z}) &= \sum_{m\leqslant M}\sum_{n\leqslant N}\sum_{q\leqslant Q} a_{m,n,q}x_m y_n z_q \\ &= \sum_{\substack{q\leqslant Q\\(q,k)=1}} z_q\Bigg(\sum_{\substack{m\leqslant M,\, n\leqslant N\\ mn\equiv k\,(\mathrm{mod}\, q)}}\sum x_m y_n - \frac{1}{\varphi(q)}\sum_{\substack{m\leqslant M,\, n\leqslant N\\(mn,q)=1}} x_m y_n\Bigg).\end{aligned}$$

This expression controls the distribution in certain arithmetic progressions of the function $f * g$, where $f(m) = 1_{[1,M]}(m)\cdot x_m$ and $g(n) = 1_{[1,N]}(n)\cdot y_n$. □

The norm of a bilinear form. The power of the method of bilinear forms lies in its ability to produce pointwise bounds for $\psi(\vec{x},\vec{y})$ valid for general vectors $\vec{x}$ and $\vec{y}$. The key notion for doing so is the *norm* of ψ. It is defined to be the smallest positive real number $\|\psi\|$ such that

$$|\psi(\vec{x},\vec{y})| \leqslant \|\psi\| \cdot \|\vec{x}\|_2\|\vec{y}\|_2 \quad \text{for all } \vec{x}\in\mathbb{C}^M,\ \vec{y}\in\mathbb{C}^N, \tag{25.3}$$

where $\|\cdot\|_2$ denotes the usual Euclidean norm, defined by $\|\vec{v}\|_2^2 = v_1^2+\cdots+v_d^2$ for a vector $\vec{v} = (v_1, \dots, v_d) \in \mathbb{C}^d$. The emphasis should be put here on the assumption that (25.3) is true for *all* vectors $\vec{x} \in \mathbb{C}^M$ and $\vec{y} \in \mathbb{C}^N$.

The existence of $\|\psi\|$ is easy to establish: if $\vec{x} = \vec{0}$ or $\vec{y} = \vec{0}$, then (25.3) is trivially true. Otherwise, we set $\vec{v} = \vec{x}/\|\vec{x}\|_2$ and $\vec{w} = \vec{y}/\|\vec{y}\|_2$, so that $\vec{v} \in S^M$ and $\vec{w} \in S^N$, where S^d denotes the d-dimensional complex unit sphere. The bilinearity of ψ renders (25.3) equivalent to the inequality $|\psi(\vec{v}, \vec{w})| \leqslant \|\psi\|$ for all $\vec{v} \in S^M$, $\vec{w} \in S^N$. We thus find that $\|\psi\|$ is equal to $\max\{|\psi(\vec{v}, \vec{w})| : \vec{v} \in S^M, \vec{w} \in S^N\}$, which exists by the compactness of S^M and S^N, and by the continuity of ψ.

Example 25.4. In Theorem 23.6, we saw an instance of (25.3). Indeed, we can reinterpret Theorem 23.6 as stating that the norm of the bilinear form ψ of Example 25.1(a) is $\ll (q + M + N + MN/q)^{1/2}(\log(2q))^{1/2}$, where a/q is a reduced fraction such that $|\alpha - a/q| \leqslant 1/q^2$. In addition, the discussion in Remark 23.7 implies that $\|\psi\| \gg (M+N+MN/q)^{1/2}$ when $\alpha = a/q$. □

A spectral interpretation of the norm. To gain some intuition about what the norm of a bilinear form measures, we study the special case when $M = N$ and the coefficients of ψ form a Hermitian matrix (i.e., $a_{m,n} = \overline{a}_{n,m}$). We then know from the Spectral Theorem for Hermitian matrices that $\mathbb{C}^N$ admits an orthonormal basis of eigenvectors of A, say $\vec{\varepsilon}_1, \dots, \vec{\varepsilon}_N$. Let $\lambda_1, \dots, \lambda_N$ be the corresponding eigenvalues, which are real numbers. We claim that

$$\|\psi\| = \max\{|\lambda_1|, \dots, |\lambda_N|\}. \tag{25.4}$$

To see this identity, it is convenient to work with the Hermitian analogue of ψ, defined by

$$\phi(\vec{x}, \vec{y}) = \sum_{1 \leqslant m,n \leqslant N} \sum a_{m,n} \overline{x}_m y_n.$$

A straightforward computation reveals that $\phi(\vec{\varepsilon}_m, \vec{\varepsilon}_n) = 1_{m=n} \cdot \lambda_n$. Hence, if we express $\vec{x}$ and $\vec{y}$ with respect to the basis $\{\vec{\varepsilon}_1, \dots, \vec{\varepsilon}_N\}$, say $\vec{x} = \sum_{n=1}^N s_n \vec{\varepsilon}_n$ and $\vec{y} = \sum_{n=1}^N t_n \vec{\varepsilon}_n$, then

$$\phi(\vec{x}, \vec{y}) = \sum_{1 \leqslant m,n \leqslant N} \sum \overline{s}_m t_n \phi(\vec{\varepsilon}_m, \vec{\varepsilon}_n) = \sum_{1 \leqslant n \leqslant N} \lambda_n \overline{s}_n t_n.$$

If $L = \max\{|\lambda_1|, \dots, |\lambda_N|\}$, then

$$|\phi(\vec{x}, \vec{y})| \leqslant L \sum_{1 \leqslant n \leqslant N} |\overline{s}_n t_n|.$$

We then apply the Cauchy-Schwarz inequality to the sum on the right side. Since $\sum_{n=1}^N |s_n|^2 = \|\vec{x}\|_2^2$ and $\sum_{n=1}^N |t_n|^2 = \|\vec{y}\|_2^2$ from the orthonormality of the vectors $\vec{\varepsilon}_1, \dots, \vec{\varepsilon}_N$, we conclude that $|\phi(\vec{x}, \vec{y})| \leqslant L \cdot \|\vec{x}\|_2 \|\vec{y}\|_2$ for all $\vec{x}, \vec{y} \in \mathbb{C}^N$, whence $\|\psi\| \leqslant L$.

On the other hand, we have $|\phi(\vec{\varepsilon}_n, \vec{\varepsilon}_n)| \leqslant \|\psi\|$ for all n, as it can be seen by the definition of $\|\psi\|$ and the fact that $\|\vec{\varepsilon}_n\|_2 = 1$. Since $\phi(\vec{\varepsilon}_n, \vec{\varepsilon}_n) = \lambda_n$, we infer that $\|\psi\| \geqslant |\lambda_n|$ for all n, thus completing the proof of (25.4).

In conclusion, we may think of the norm of ψ as something like the size of the large eigenvalue of a matrix with number-theoretic properties.

Approximate orthogonality. In practice, it is hard to get a handle on the eigenvalues of the matrix of coefficients of ψ. Instead, we develop a different method that has as its starting point the obvious relations

$$\psi(\vec{x},\vec{y}) = \sum_{m=1}^{M} x_m \sum_{n=1}^{N} a_{m,n} y_n \quad \text{and} \quad \psi(\vec{x},\vec{y}) = \sum_{n=1}^{N} y_n \sum_{m=1}^{M} a_{m,n} x_m, \tag{25.5}$$

and which does not require M to equal N.

As we briefly explained in Chapter 23, the importance of the above relations is that they express $\psi(\vec{x},\vec{y})$ as an *average of averages*. Applying the Cauchy-Schwarz inequality to the first identity of (25.5), we find that

$$|\psi(\vec{x},\vec{y})|^2 \leqslant \|\vec{x}\|_2^2 \sum_{m=1}^{M} \Big| \sum_{n=1}^{N} a_{m,n} y_n \Big|^2. \tag{25.6}$$

This maneuver allows us to eliminate the unknown coefficients x_m and smoothen out the variable m, which is now weighted with the function 1. Similarly, if we apply the Cauchy-Schwarz inequality to the second identity of (25.5), we obtain the bound

$$|\psi(\vec{x},\vec{y})|^2 \leqslant \|\vec{y}\|_2^2 \sum_{n=1}^{N} \Big| \sum_{m=1}^{N} a_{m,n} x_m \Big|^2. \tag{25.7}$$

We open the square in (25.7) using the identity $|z|^2 = \overline{z} \cdot z$ to deduce that

$$\begin{aligned} |\psi(\vec{x},\vec{y})|^2 &\leqslant \|\vec{y}\|_2^2 \sum_{n=1}^{N} \sum_{m_1=1}^{M} \sum_{m_2=1}^{M} \overline{a}_{m_1,n} \overline{x}_{m_1} a_{m_2,n} x_{m_2} \\ &= \|\vec{y}\|_2^2 \sum_{m_1=1}^{M} \sum_{m_2=1}^{M} \overline{x}_{m_1} x_{m_2} \sum_{n=1}^{N} \overline{a}_{m_1,n} a_{m_2,n}. \end{aligned} \tag{25.8}$$

In particular, if the sum $\sum_{n=1}^{N} \overline{a}_{m_1,n} a_{m_2,n}$ is "small" whenever $m_1 \neq m_2$, we hope to obtain a non-trivial bound for the sum in (25.8), and hence for $\|\psi\|$.

There is a more conceptual way to interpret what we did above. We can write the first identity of (25.5) as

$$\psi(\vec{x},\vec{y}) = \sum_{m=1}^{M} x_m \cdot \langle \vec{y}, \vec{v}_m \rangle, \qquad \text{where} \quad \vec{v}_m = (\overline{a}_{m,1}, \dots, \overline{a}_{m,N}), \tag{25.9}$$

and the second identity of (25.5) as

$$\psi(\vec{x},\vec{y}) = \sum_{n=1}^{N} y_n \cdot \langle \vec{x}, \vec{w}_n \rangle, \qquad \text{where} \quad \vec{w}_n = (\overline{a}_{1,n}, \dots, \overline{a}_{M,n}). \tag{25.10}$$

Similarly, (25.8) becomes

$$|\psi(\vec{x},\vec{y})|^2 \leqslant \|\vec{y}\|_2^2 \sum_{m_1=1}^{M}\sum_{m_2=1}^{M} \overline{x}_{m_1}x_{m_2}\cdot\langle v_{m_1},v_{m_2}\rangle. \tag{25.11}$$

Hence, we see that $\|\psi\|$ should be small when the inner products $\langle v_{m_1},v_{m_2}\rangle$ are small whenever $m_1\neq m_2$, that is to say, when the vectors $\vec{v}_m$ are approximately orthogonal to each other. The relevance of this property can be seen more easily via relation (25.9): a *fixed* vector $\vec{y}$ cannot correlate strongly with *many* of the approximately orthogonal vectors $\vec{v}_m$. Hence, most of the inner products $\langle\vec{y},\vec{v}_m\rangle$ should be small compared to the trivial bound $\|\vec{y}\|_2\|\vec{v}_m\|_2$ coming from the Cauchy-Schwarz inequality. This should yield a non-trivial bound on the sum of (25.9).

Remark 25.5. It must be stressed that in order to exploit the approximate orthogonality of the vectors $\vec{v}_m$, we have to start with (25.10) that involves the vectors w_n rather than with (25.9). □

In order to get a sense of what to expect as a bound on $\|\psi\|$ when the vectors $\vec{v}_m$ are approximately orthogonal to each other, we study the ideal case when they are truly orthogonal to each other. Essentially, the lemma we prove below constitutes a generalization of (25.4) to non-square matrices.

Lemma 25.6. *Let $\psi,\vec{v}_m$ be as above, and set $L=\max\{\|\vec{v}_1\|_2,\dots,\|\vec{v}_M\|_2\}$. Then $\|\psi\|\geqslant L$. If, in addition, we assume that the vectors $v_1,\dots,v_M$ are orthogonal to each other, then $\|\psi\|=L$.*

Proof. Let $\vec{e}_1,\dots,\vec{e}_M$ denote the vectors of the standard basis of $\mathbb{C}^M$. For any $m\leqslant M$, we have $\psi(\vec{e}_m,\vec{v}_m)=\|\vec{v}_m\|_2^2$. On the other hand, (25.3) implies that $\psi(\vec{e}_m,\vec{v}_m)\leqslant\|\psi\|\cdot\|\vec{v}_m\|$. We thus conclude that $\|\psi\|\geqslant\|\vec{v}_m\|_2$. Since m can be chosen arbitrarily, the first part of the lemma follows.

For the second part, note that the set $\{\vec{v}_1/\|\vec{v}_1\|_2,\dots,\vec{v}_M/\|\vec{v}_M\|_2\}$ is orthonormal when the vectors $\vec{v}_1,\dots,\vec{v}_M$ are mutually orthogonal. Combining (25.6) with Bessel's inequality, we deduce that

$$\begin{aligned}|\psi(\vec{x},\vec{y})|^2 \leqslant \|\vec{x}\|_2^2\sum_{m=1}^{M}|\langle\vec{y},\vec{v}_m\rangle|^2 &\leqslant L^2\|\vec{x}\|_2^2\sum_{m=1}^{M}\Big|\Big\langle\vec{y},\frac{\vec{v}_m}{\|\vec{v}_m\|_2}\Big\rangle\Big|^2\\ &\leqslant L^2\|\vec{x}\|_2^2\|\vec{y}\|_2^2.\end{aligned}$$

In particular, $\|\psi\|\leqslant L$, which completes the proof. □

Remark 25.7. Working with the vectors $\vec{w}_n$, and with relations (25.7) and (25.10), we can show that $\|\psi\|\geqslant\max\{\|\vec{w}_1\|_2,\dots,\|\vec{w}_N\|_2\}$. Moreover, this lower bound is sharp when the vectors $\vec{w}_n$ are mutually orthogonal. □

We now return to the more general case when the vectors $\vec{v}_m$ are approximately orthogonal to each other. In Theorem 23.6, we saw one example of how to exploit the mutual quasi-orthogonality of the vectors $\vec{v}_m$. Below, we give another one that deals with Example 25.2 when $\mathcal{C} = \{\chi \,(\mathrm{mod}\, q)\}$.

Example 25.8. Given $q, N \in \mathbb{N}$, let ψ denote the bilinear form with coefficients $a_{m,n} = \chi_m(n)\sqrt{q/\varphi(q)}$, where χ_m ranges over all Dirichlet characters mod q and $n \in \mathbb{Z} \cap [1, N]$. We wish to bound the norm of ψ.

Note that

$$|\langle \vec{v}_{m_1}, \vec{v}_{m_2}\rangle| = \frac{q}{\varphi(q)} \cdot \begin{cases} O(\sqrt{q}\log q) & \text{if } m_1 \neq m_2, \\ N\varphi(q)/q + O(2^{\omega(q)}) & \text{otherwise,} \end{cases} \tag{25.12}$$

with the first case following by the Pólya-Vinogradov inequality and the second one by Theorem 2.1. Inserting (25.12) into (25.11) implies that

$$|\psi(\vec{x}, \vec{y})|^2 \leqslant \|\vec{y}\|_2^2 \Big(\sum_{m=1}^{M} |x_m|^2 \cdot N + \sum_{m_1=1}^{M} \sum_{m_2=1}^{M} |x_{m_1} x_{m_2}| \cdot O\Big(\frac{q^{3/2}\log q}{\varphi(q)}\Big)\Big).$$

Using the inequality $|zw| \leqslant (|z|^2 + |w|^2)/2$ and the symmetry of the above double sum over m_1 and m_2, we find that

$$\sum_{m_1=1}^{M} \sum_{m_2=1}^{M} |x_{m_1} x_{m_2}| \leqslant \sum_{m=1}^{M} |x_m|^2 \sum_{m'=1}^{M} 1 = M\|\vec{x}\|_2^2 = \varphi(q)\|\vec{x}\|_2^2.$$

As a consequence,

$$|\psi(\vec{x}, \vec{y})|^2 \leqslant (N + O(q^{3/2}\log q))\|\vec{x}\|_2^2 \|\vec{y}\|_2^2 \tag{25.13}$$

for all $\vec{x} \in \mathbb{C}^M$ and $\vec{y} \in \mathbb{C}^N$, that is to say, $\|\psi\|^2 \leqslant N + O(q^{3/2}\log q)$. Comparing this upper bound to the lower bound in Lemma 25.6 yields, we see that it is sharp when $N \gg q^{3/2}\log q$.

The bilinear form ψ of the present example has the special feature that we can easily control the inner products of the dual vectors $\vec{w}_n$ too:

$$\langle w_{n_1}, w_{n_2}\rangle = \frac{q}{\varphi(q)} \sum_{\chi \,(\mathrm{mod}\, q)} \overline{\chi}(n_1)\chi(n_2) = q \cdot 1_{(n_1 n_2, q)=1} \cdot 1_{n_1 \equiv n_2 \,(\mathrm{mod}\, q)}.$$

To exploit this identity, we start with (25.9). The analogue of (25.11) in this setting is

$$\begin{aligned} |\psi(\vec{x}, \vec{y})|^2 &\leqslant \|\vec{x}\|_2^2 \sum_{n_1=1}^{N} \sum_{n_2=1}^{N} \overline{y}_{n_1} y_{n_2} \cdot \langle w_{m_1}, w_{m_2}\rangle \\ &= \|\vec{x}\|_2^2 \cdot q \sum_{n_1=1}^{N} \sum_{n_2=1}^{N} \overline{y}_{n_1} y_{n_2} \cdot 1_{(n_1 n_2, q)=1} \cdot 1_{n_1 \equiv n_2 \,(\mathrm{mod}\, q)}. \end{aligned}$$

Since $|y_{n_1}y_{n_2}| \leqslant (|y_{n_1}|^2+|y_{n_2}|^2)/2$ and $\#\{\, n \leqslant N : n \equiv a \,(\mathrm{mod}\, q)\,\} \leqslant 1+N/q$ for all $a \in \mathbb{Z}$, we conclude that

$$|\psi(\vec{x},\vec{y})|^2 \leqslant (N+q)\|\vec{x}\|_2^2\|\vec{y}\|_2^2 \tag{25.14}$$

for all $\vec{x} \in \mathbb{C}^M$ and $\vec{y} \in \mathbb{C}^N$. Hence, $\|\psi\|^2 \leqslant N+q$, which improves upon (25.13) and is sharp in view of Lemma 25.6. □

Duality. The discussion of the above example brings forward an important notion that underlies the theory of bilinear forms. Consider the linear operators $T : \mathbb{C}^M \to \mathbb{C}^N$ and $T^* : \mathbb{C}^N \to \mathbb{C}^M$, defined by $T(\vec{x}) = (\sum_{m=1}^M a_{m,n}x_m)_{n=1}^N$ and $T^*(\vec{y}) = (\sum_{n=1}^N \overline{a}_{m,n}y_n)_{m=1}^M$. We then have

$$\psi(\vec{x},\vec{y}^*) = \langle \vec{x}, T^*(\vec{y})\rangle = \langle T(\vec{x}), \vec{y}\rangle, \tag{25.15}$$

where $\vec{y}^* = (\overline{y}_1,\dots,\overline{y}_n)$. This relation is an equivalent way of writing (25.5) (with the coordinates of $\vec{y}$ conjugated), and it implies that T^* is the adjoint operator of T.

Now, note that applying the Cauchy-Schwarz inequality to the third expression of (25.15) yields the upper bound $|\psi(\vec{x},\vec{y}^*)| \leqslant \|T(\vec{x})\|_2\|\vec{y}\|_2$. Moreover, this is an equality when $\vec{y} = T(\vec{x})$. Consequently, $\|\psi\|$ is equal to the smallest positive number B such that $\|T(\vec{x})\|_2 \leqslant B\cdot\|\vec{x}\|_2$. This number B is precisely the norm of the operator T, which we denote by $\|T\|$.

We have thus proven that $\|\psi\| = \|T\|$. Similarly, working with the second expression of (25.15), we can show that $\|\psi\| = \|T^*\|$. In particular, we see that T and T^* have the same norm, a well-known fact from functional analysis called the *duality principle*[167, Proposition 5.4, p. 183].

To sum up, we have proved the following result.

Theorem 25.9. *Let $(a_{m,n})_{m\leqslant M,\, n\leqslant N}$ be some complex coefficients and let $\Delta \geqslant 0$. The following statements are equivalent:*

(a) $\big|\sum_{m=1}^M\sum_{n=1}^N a_{m,n}x_my_n\big| \leqslant \Delta\|\vec{x}\|_2\|\vec{y}\|_2$ *for all $\vec{x} \in \mathbb{C}^M$, $\vec{y} \in \mathbb{C}^N$.*

(b) $\sum_{n=1}^N \big|\sum_{m=1}^M a_{m,n}x_m\big|^2 \leqslant \Delta^2\|\vec{x}\|_2^2$ *for all $\vec{x} \in \mathbb{C}^M$.*

(c) $\sum_{m=1}^M \big|\sum_{n=1}^N a_{m,n}y_n\big|^2 \leqslant \Delta^2\|\vec{y}\|_2^2$ *for all $\vec{y} \in \mathbb{C}^N$.*

Remark 25.10. Assume that $|a_{m,n}| = 1$ for all m,n. In view of Lemma 25.6 and Remark 25.7, we have that $\|\psi\| \geqslant \max\{\sqrt{M},\sqrt{N}\}$. Assume for simplicity that $N \leqslant M$. Hence, the best possible result we can hope for is of the form $\|\psi\| \ll \sqrt{M}$. In view of Theorem 25.9, this would imply that

$$\frac{1}{M}\sum_{m=1}^M\Big|\sum_{n=1}^N y_n a_{m,n}\Big|^2 \ll N \qquad \text{whenever} \qquad \max_{1\leqslant n\leqslant N}|y_n| \leqslant 1, \tag{25.16}$$

since $\|\vec{y}\|_2 \leqslant \sqrt{N}$ when $|y_n| \leqslant 1$ for all n. If we let $S_m = \sum_{n=1}^{N} y_n a_{m,n}$, we can interpret (25.16) as saying that $S_m \ll N^{1/2}$ on average over $m \leqslant M$. Notice that this is approximately the square root of the trivial pointwise bound $|S_m| \leqslant N$. We then say that S_m *exhibits square-root cancellation* on average over $m \leqslant M$. Consequently, in the special case when $y_n = 1_P(n)$ and $a_{m,n} = \chi_m(n)$, with χ_m running over an appropriate family of Dirichlet characters, relation (25.16) can be interpreted as an averaged version of the Generalized Riemann Hypothesis (see Exercises 8.6, 11.2 and 11.3(b)). □

The large sieve

One of the most important applications of the theory of bilinear forms is the *large sieve*. It was first discovered by Linnik [131] while studying sieve problems with unbounded sifting dimension (hence the name "large sieve"). We will present these arithmetic applications in the next section. First, we develop the large sieve it in its abstract form, as an inequality for additive and multiplicative characters.

Our goal is to bound the norms of the bilinear forms associated to the sums in (25.1). We begin by studying the first of these forms. We consider a slightly more general set-up: let $\{\alpha_1, \dots, \alpha_R\}$ be a set of relation numbers. We want to obtain bounds for the sum

$$\sum_{1 \leqslant r \leqslant R} |S^+(\alpha_r)|^2, \tag{25.17}$$

where we have extended S^+ to all real numbers by letting

$$S^+(\alpha) = \sum_{H < n \leqslant H+N} c_n e(n\alpha).$$

In view of Theorem 25.9, the underlying bilinear form has coefficients $a_{r,n} = e(n\alpha_r)$ with $r \in \{1, \dots, R\}$ and $n \in \{H+1, \dots, H+N\}$. We then consider the vectors

$$\vec{v}_r = (e(-n\alpha_r))_{H < n \leqslant H+N}$$

and note that

$$\langle \vec{v}_r, \vec{v}_s \rangle = \sum_{H < n \leqslant H+N} e(n(\alpha_s - \alpha_r)) \ll \frac{1}{\|\alpha_s - \alpha_r\|} \tag{25.18}$$

by arguing as in (23.34). Hence, if the set $\{\alpha_1, \dots, \alpha_R\}$ is δ-spaced mod 1 (recall that this means that $\|\alpha_r - \alpha_s\| \geqslant \delta$ when $r \neq s$) with $\delta^{-1} = o(N)$, the vectors $\vec{v}_1, \dots, \vec{v}_R$ are approximately orthogonal to each other.

The first average of (25.1) is of the form (25.17) with the set of points α_r being the Farey fractions of order $\leqslant Q$, which we denote by $\mathcal{F}_Q$. If a/q

and a'/q' are distinct elements of $\mathcal{F}_Q$ written in lowest terms, then

$$\Big|\frac{a}{q}-\frac{a'}{q'}+n\Big| = \frac{|aq'-aq'+nqq'|}{qq'} \geqslant \frac{1}{qq'} \geqslant \frac{1}{Q^2} \tag{25.19}$$

for every integer n, that is to say, the set $\mathcal{F}_Q$ is Q^{-2}-spaced. Hence, we can bound $\sum_{a/q\in\mathcal{F}_Q}|S^+(a/q)|^2$ by working with the more general sum of (25.17). For the latter, we prove the following fundamental theorem.

Theorem 25.11 (The additive large sieve inequality)**.** *Let $\{\alpha_1,\dots,\alpha_R\}$ be a set of real numbers that are δ-spaced* mod 1*, $H\in\mathbb{Z}$, $N\in\mathbb{Z}_{\geqslant 1}$ and $\vec{c}=(c_{H+1},\dots,c_{H+N})\in\mathbb{C}^N$. We then have*

$$\sum_{r=1}^{R}\Big|\sum_{H<n\leqslant H+N} c_n e(n\alpha_r)\Big|^2 \ll (N+\delta^{-1})\|\vec{c}\|_2^2. \tag{25.20}$$

Proof. Given any $H'\in\mathbb{Z}$, we have

$$\Big|\sum_{H<n\leqslant H+N} c_n e(n\alpha)\Big| = \Big|\sum_{H'<n\leqslant N+H'} c_{n-H'+H} e(n\alpha)\Big|,$$

by letting $n=m-H'+H$ in the first sum and then noticing that $e(n\alpha)=e((H-H')\alpha)\cdot e(m\alpha)$. It thus suffices to prove the theorem for a special choice of H, and it will automatically follow for all other values of H. We shall take $H=-\lfloor N/2\rfloor-1$ that (almost) symmetrizes the range of n.

Instead of proving (25.20), we use Theorem 25.9 which tells us that it suffices to prove the dual inequality

$$\sum_{H<n\leqslant H+N}\Big|\sum_{r=1}^{R} b_r e(n\alpha_r)\Big|^2 \ll (N+\delta^{-1})\|\vec{b}\|_2^2 \tag{25.21}$$

for all $\vec{b}=(b_1,\dots,b_R)\in\mathbb{C}^R$. We will open the square and take advantage of the mutual quasi-orthogonality of the vectors $(e(n\alpha_r))_{H<n\leqslant H+N}$, but first we smoothen the n variable a bit further. This will improve the quality of the bound we obtain in terms of δ (see Remark 25.12).

Since we have assumed that $H=-\lfloor N/2\rfloor-1$, we have that $|n|\leqslant N/2$ in the left-hand side of (25.21). Therefore,

$$\sum_{H<n\leqslant H+N}\Big|\sum_{r=1}^{R} b_r e(n\alpha_r)\Big|^2 \leqslant 2\sum_{|n|\leqslant N}(1-|n|/N)\Big|\sum_{r=1}^{R} b_r e(n\alpha_r)\Big|^2.$$

Expanding the square on the right side as in (25.8), and bringing the sum over n inside, we find that

$$\sum_{H<n\leqslant H+N}\Big|\sum_{r=1}^{R} b_r e(n\alpha_r)\Big|^2 \leqslant 2\sum_{r=1}^{R}\sum_{s=1}^{R} b_r\overline{b}_s\sum_{|n|\leqslant N}(1-|n|/N)e(n(\alpha_r-\alpha_s)).$$

The innermost sum is the Fejér kernel F_N evaluated at $\alpha_r - \alpha_s$. Indeed, recall the well-known identities

$$F_N(\alpha) = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{|m|\leqslant n} e(m\alpha) = \sum_{|n|\leqslant N}(1-|n|/N)e(n\alpha) = \frac{1}{N}\left(\frac{\sin(N\pi\alpha)}{\sin(\pi\alpha)}\right)^2.$$

In particular, we have

$$|F_N(\alpha)| \leqslant \min\Big\{F_N(0), \frac{1}{N\sin^2(\pi\alpha)}\Big\} \leqslant \frac{N}{\max\{1, 2N\|\alpha\|\}^2}, \tag{25.22}$$

since $\sin(\pi x) \geqslant 2x$ for $x \in [0, 1/2]$. Together with the inequality $2|b_r b_s| \leqslant |b_r|^2 + |b_s|^2$ and the symmetry of the summation in r and s, this implies that

$$\begin{aligned}\sum_{H<n\leqslant H+N}\Big|\sum_{r=1}^{R} b_r e(n\alpha_r)\Big|^2 &\leqslant N\sum_{r=1}^{R}\sum_{s=1}^{R}\frac{|b_r|^2+|b_s|^2}{\max\{1, 2N\|\alpha_r-\alpha_s\|\}^2}\\ &= \sum_{r=1}^{R}|b_r|^2\sum_{s=1}^{R}\frac{2N}{\max\{1, 2N\|\alpha_r-\alpha_s\|\}^2}. \end{aligned}\tag{25.23}$$

It remains to bound the innermost sum over s.

Let $J = \lfloor 1/(2N\delta)\rfloor$ and $r \in \{1, \dots, R\}$. There is an ordering $\alpha_{s_1}, \dots, \alpha_{s_R}$ with $s_1 = r$ such that the sequence $\{\|\alpha_{s_j} - \alpha_r\|\}_{j=1}^R$ is increasing. Since the points α_s are δ-spaced mod 1, we have $\|\alpha_{s_{2j}} - \alpha_r\| \geqslant j\delta$ when $1 \leqslant j \leqslant R/2$, as well as $\|\alpha_{s_{2j+1}} - \alpha_r\| \geqslant j\delta$ when $1 \leqslant j \leqslant (R-1)/2$. Consequently,

$$\begin{aligned}\sum_{s=1}^{R}\frac{2N}{\max\{1, 2N\|\alpha_r-\alpha_s\|\}^2} &\leqslant 2N + \sum_{1\leqslant j\leqslant R/2}\frac{4N}{\max\{1, 2N\delta j\}^2}\\ &\leqslant 2N + 4NJ + \sum_{j\geqslant J+1}\frac{1}{N\delta^2 j^2}\\ &\leqslant 2N + 4NJ + \frac{1}{N\delta^2}\cdot\min\{1/J, \pi^2/6\}, \end{aligned}\tag{25.24}$$

because $\sum_{j\geqslant J+1} j^{-2} \leqslant \int_J^\infty x^{-2}\mathrm{d}x = 1/J$ and $\sum_{j\geqslant 1} j^{-2} = \pi^2/6$.

If $N > 1/(2\delta)$, then $J = 0$, so the expression in (25.24) is $\ll N$; on the other hand, if $N \leqslant 1/(2\delta)$, then $J \asymp 1/(N\delta) \gg 1$, so the expression in (25.24) is $\asymp \delta^{-1}$. In any case,

$$\sum_{s=1}^{R}\frac{2N}{\max\{1, 2N\|\alpha_r-\alpha_s\|\}^2} \ll N + \delta^{-1}.$$

Inserting the above estimate into (25.23) completes the proof of (25.21), and hence of the theorem. □

Remark 25.12. Had we not smoothened the sum over n in (25.21), we would have had to use (25.18) instead of (25.22) and, hence, to replace the

innermost sum of (25.23) by $\sum_{s\leqslant R}\|\alpha_r-\alpha_s\|^{-1}$. This last sum is $\ll \delta^{-1}\log R$ (and this estimate is best possible for general δ-spaced points α_j). We would have thus proven (25.20) with $N+\delta^{-1}\log R$ in place of $N+\delta^{-1}$. □

Remark 25.13. Montgomery-Vaughan [146] and Selberg [163] proved independently that the implicit constant in Theorem 25.11 can be taken to be 1. They also showed that $N+1/\delta$ can be replaced by $N+1/\delta-1$ [114, Theorem 7.7]. As Lemma 25.6 reveals, this is essentially best possible. □

As a direct corollary of Theorem 25.11 and of relation (25.19), we obtain the following result.

Theorem 25.14 (The additive large sieve inequality, II)**.** *Let $Q\geqslant 1$, $H\in\mathbb{Z}$, $N\in\mathbb{Z}_{\geqslant 1}$ and $\vec{c}=(c_{H+1},\dots,c_{H+N})\in\mathbb{C}^N$. We then have*

$$\sum_{q\leqslant Q}\sum_{a\in(\mathbb{Z}/q\mathbb{Z})^*}\Big|\sum_{H<n\leqslant H+N}c_n e(na/q)\Big|^2\ll (N+Q^2)\|\vec{c}\|_2^2.$$

We now turn to the second expression of (25.1).

Theorem 25.15 (The multiplicative large sieve inequality)**.** *Let $Q\geqslant 1$, $H\in\mathbb{Z}$, $N\in\mathbb{Z}_{\geqslant 1}$ and $\vec{c}=(c_{H+1},\dots,c_{H+N})\in\mathbb{C}^N$. We then have*

$$\sum_{q\leqslant Q}\sideset{}{^*}\sum_{\chi\,(\mathrm{mod}\,q)}\frac{q}{\varphi(q)}\Big|\sum_{H<n\leqslant H+N}c_n\chi(n)\Big|^2\ll (N+Q^2)\|\vec{c}\|_2^2,$$

where the notation $\sum^*$ means that the sum runs over primitive characters.

Proof. The associated bilinear form has coefficients $a_{m,n}=\chi_m(n)$, where χ_m ranges over all primitive Dirichlet characters of conductor $\leqslant Q$, and $n\in\mathbb{Z}\cap(H,H+N]$. However, instead of working with this form, we will show that

$$\frac{q}{\varphi(q)}\sideset{}{^*}\sum_{\chi\,(\mathrm{mod}\,q)}|S^\times(\chi)|^2\leqslant\sum_{a\in(\mathbb{Z}/q\mathbb{Z})^*}|S^+(a/q)|^2 \tag{25.25}$$

for all q, and then invoke Theorem 25.11, where the functions S^+ and $S^\times$ are defined as in the beginning of this chapter.

When χ is a primitive character mod q, Theorem 10.3 implies that

$$\begin{aligned}S^\times(\chi)&=\sum_{H<n\leqslant H+N}c_n\cdot\frac{1}{\mathcal{G}(\overline{\chi})}\sum_{a\in(\mathbb{Z}/q\mathbb{Z})^*}\overline{\chi}(a)e(na/q)\\&=\frac{1}{\mathcal{G}(\overline{\chi})}\sum_{a\in(\mathbb{Z}/q\mathbb{Z})^*}\overline{\chi}(a)S^+(a/q).\end{aligned}$$

Since $|\mathcal{G}(\overline{\chi})| = \sqrt{q}$ for a primitive character $\chi \,(\mathrm{mod}\, q)$ by Theorem 10.4, we find that

$$\begin{aligned}\sum_{\chi \,(\mathrm{mod}\, q)}^{*} |S^{\times}(\chi)|^2 &= \frac{1}{q} \sum_{\chi \,(\mathrm{mod}\, q)}^{*} \Big| \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \overline{\chi}(a) S^{+}(a/q) \Big|^2 \\ &\leqslant \frac{1}{q} \sum_{\chi \,(\mathrm{mod}\, q)} \Big| \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \overline{\chi}(a) S^{+}(a/q) \Big|^2 \\ &= \frac{\varphi(q)}{q} \sum_{a \in (\mathbb{Z}/q\mathbb{Z})^*} |S^{+}(a/q)|^2, \qquad (25.26)\end{aligned}$$

with the last relation following from Parseval's identity (10.6) for multiplicative characters. This proves (25.25), thus completing the proof of the theorem. □

The arithmetic form of the large sieve

The use of the term "sieve" in this chapter is not at all evident, since what we have talked about so far bears no resemblance to the sieve theory we developed in Part 4. However, it is possible to use the large sieve inequality to deduce a rather strong sieve upper bound, thus justifying the terminology "large sieve".

The set-up is slightly different compared to the one we saw in Part 4 of the book. To motivate it, consider a polynomial $F(x) \in \mathbb{Z}[x]$ and the sets $\mathcal{A} = \{ F(n) : n \leqslant x \}$ and $\mathcal{P} = \{p \leqslant y\}$. Then, $S(\mathcal{A}, \mathcal{P})$ counts integers $n \leqslant x$ such that $p \nmid F(n)$ for all $p \leqslant y$. Equivalently, if we let

$$R_p = \{ m \in \mathbb{Z}/p\mathbb{Z} : F(m) \equiv 0 \,(\mathrm{mod}\, p) \},$$

then $S(\mathcal{A}, \mathcal{P})$ counts integers $n \leqslant x$ such that $n \notin R_p \,(\mathrm{mod}\, p)$ for all $p \leqslant y$. (We then say that n "avoids" the set R_p for all $p \leqslant y$.) Notice that the set $\mathcal{A}$ satisfies Axiom 1 with $\nu(p) = |R_p|$. In particular, the function φ^* we saw in Chapter 21 and in Theorem 21.1 is given by

$$\varphi^*(n) = n \prod_{p \mid n} (1 - |R_p|/p).$$

With these remarks in mind, we now state the main result of this section.

Theorem 25.16 (The arithmetic large sieve inequality)**.** *Let $y \geqslant 1$, $H \in \mathbb{Z}$ and $N \in \mathbb{Z}_{\geqslant 1}$. In addition, for each prime $p \leqslant y$, let $R_p \subseteq \mathbb{Z}/p\mathbb{Z}$. If*

$$\mathcal{N} \subseteq \{ H < n \leqslant H + N : n \notin R_p \,(\mathrm{mod}\, p) \text{ for all } p \leqslant y \},$$

then

$$\#\mathcal{N} \ll (N + y^2) \Big/ \sum_{m \leqslant y} \mu^2(m) f(m), \qquad \text{where} \quad f(m) = \prod_{p \mid m} \frac{|R_p|}{p - |R_p|}.$$

Proof. The theorem is trivial if $R_p = \mathbb{Z}/p\mathbb{Z}$ for some $p \leqslant y$, because $\mathcal{N} = \emptyset$ and $f(p) = \infty$ in this case. So let us assume that $R_p \neq \mathbb{Z}/p\mathbb{Z}$ for all $p \leqslant y$.

The main idea is that if $\mathcal{N}$ avoids many residue classes modulo each prime $p \leqslant y$, it must be unevenly distributed among residue classes of arithmetic progressions of moduli $\leqslant y$. The large sieve inequality implies that this can only happen if $\mathcal{N}$ is very sparse.

More concretely, let $(c_n)_{n\in\mathbb{Z}}$ be a sequence supported on $\mathcal{N}$, and fix for the moment a prime $p \leqslant y$. To study the distribution of $(c_n)_{n\in\mathbb{Z}}$ mod p, we equip $\mathbb{Z}/p\mathbb{Z}$ with the uniform counting measure (that is to say, $\mathbb{P}(A) = |A|/p$ for $A \subseteq \mathbb{Z}/p\mathbb{Z}$) and consider the random variable $X : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ defined by

$$X(a) = \sum_{n\equiv a \,(\mathrm{mod}\, p)} c_n.$$

We will study the variance of X. On the one hand, $\mathbb{V}[X]$ must be large because $\mathcal{N}$ avoids the set R_p, and thus $X(a) = 0$ whenever $a \in R_p$. On the other hand, we will express $\mathbb{V}[X]$ in terms of the additive Fourier transform of X, which will allow us to study it using the large sieve.

We start by recalling that

$$\mathbb{V}[X] = \mathbb{E}[|X - \mathbb{E}[X]|^2] = \mathbb{E}[|X|^2] - |\mathbb{E}[X]|^2.$$

To take advantage of the fact that $X(a) = 0$ when $a \in R_p$, we use the Cauchy-Schwarz inequality: we have that

$$|\mathbb{E}[X]|^2 = \Big|\frac{1}{p} \sum_{a\in(\mathbb{Z}/p\mathbb{Z})\setminus R_p} X(a)\Big|^2 \leqslant \frac{p - |R_p|}{p} \cdot \mathbb{E}[|X|^2].$$

Since $\mathbb{E}[X] = \sum_{n\in\mathbb{Z}} c_n/p$, we conclude that

$$\mathbb{V}[X] \geqslant \Big(\frac{p}{p - |R_p|} - 1\Big)|\mathbb{E}[X]|^2 = \frac{f(p)}{p^2}\Big|\sum_{n\in\mathcal{N}} c_n\Big|^2. \tag{25.27}$$

On the other hand, we can write $\mathbb{V}[X]$ in terms of the Fourier series

$$S(\alpha) := \sum_{n\in\mathbb{Z}} c_n e(\alpha n).$$

Indeed, we have $X(a) = p^{-1}\sum_{b\,(\mathrm{mod}\,p)} e(-ab/p)S(b/p)$, whence

$$\mathbb{E}[|X|^2] = \frac{1}{p^3} \sum_{a\,(\mathrm{mod}\,p)} \Big|\sum_{b\,(\mathrm{mod}\,p)} e(-ab/p)S(b/p)\Big|^2 = \frac{1}{p^2} \sum_{b\,(\mathrm{mod}\,p)} |S(b/p)|^2$$

by Parseval's identity (10.6) for additive characters. Since we also have that $\mathbb{E}[X] = S(0)/p$, we infer that

$$\mathbb{V}[X] = \frac{1}{p^2} \sum_{b\in(\mathbb{Z}/p\mathbb{Z})^*} |S(b/p)|^2.$$

Comparing the above identity with (25.27), we conclude that

$$\sum_{a\in(\mathbb{Z}/p\mathbb{Z})^*}\Big|\sum_{n\in\mathbb{Z}}c_n e(an/p)\Big|^2 \geqslant f(p)\Big|\sum_{n\in\mathbb{Z}}c_n\Big|^2. \tag{25.28}$$

Notice that we have not assumed anything about the coefficients c_n, other than that they are supported on $\mathcal{N}$.

More generally, we claim that

$$\sum_{a\,(\mathrm{mod}\,q)}^{*}\Big|\sum_{n\in\mathbb{Z}}c_n e(an/q)\Big|^2 \geqslant f(q)\Big|\sum_{n\in\mathbb{Z}}c_n\Big|^2 \tag{25.29}$$

for all $q|P(y)=\prod_{p\leqslant y}p$ and all sequences $(c_n)_{n\in\mathbb{Z}}\subset\mathbb{C}$ supported on $\mathcal{N}$, where $\sum^*$ denotes here a sum running over integers coprime to q. We prove (25.29) by induction on $\omega(q)$. When $q=1$, it holds trivially; when $\omega(q)=1$, it follows by (25.28). Finally, assume that (25.29) holds for all $q|P(y)$ with $\omega(q)\leqslant j$, where j is some positive integer. Let $q|P(y)$ with $\omega(q)=j+1$.

We may write $q=q_1q_2$ with $\omega(q_1),\omega(q_2)\leqslant j$, so that (25.29) holds for the moduli q_1 and q_2. Note that $(q_1,q_2)=1$ by the fact that q is square-free. Consequently, when a_1 ranges over $(\mathbb{Z}/q_1\mathbb{Z})^*$ and a_2 ranges over $(\mathbb{Z}/q_2\mathbb{Z})^*$, then $a_1q_2+a_2q_1$ ranges over $(\mathbb{Z}/q\mathbb{Z})^*$. We thus find that

$$\sum_{a\,(\mathrm{mod}\,q)}^{*}\Big|\sum_{n\in\mathbb{Z}}c_n e\Big(\frac{an}{q}\Big)\Big|^2=\sum_{a_1\,(\mathrm{mod}\,q_1)}^{*}\sum_{a_2\,(\mathrm{mod}\,q_2)}^{*}\Big|\sum_{n\in\mathbb{Z}}c_n e\Big(\frac{a_1n}{q_1}+\frac{a_2n}{q_2}\Big)\Big|^2.$$

We apply (25.29) with q_2 in place of q, and with $c_n e(a_1n/q_1)$ in place of c_n. Hence,

$$\sum_{a_2\,(\mathrm{mod}\,q_2)}^{*}\Big|\sum_{n\in\mathbb{Z}}c_n e\Big(\frac{a_1n}{q_1}+\frac{a_2n}{q_2}\Big)\Big|^2\geqslant f(q_2)\Big|\sum_{n\in\mathbb{Z}}c_n e\Big(\frac{a_1n}{q_1}\Big)\Big|^2.$$

Summing the above inequality over $a_1\in(\mathbb{Z}/q_1\mathbb{Z})^*$ and applying again (25.29), this time with q_1 in place of q, we deduce that

$$\sum_{a\,(\mathrm{mod}\,q)}^{*}\Big|\sum_{n\in\mathbb{Z}}c_n e\Big(\frac{an}{q}\Big)\Big|^2\geqslant f(q_1)f(q_2)\Big|\sum_{n\in\mathbb{Z}}c_n\Big|^2.$$

Since f is a multiplicative function, relation (25.29) follows. This completes the inductive step, and hence the proof of (25.29).

Finally, applying (25.29) with $c_n=1_{n\in\mathcal{N}}$, and summing it over all square-free $q\leqslant y$, we find that

$$|\mathcal{N}|^2\sum_{q\leqslant y}\mu^2(q)f(q)\leqslant\sum_{q\leqslant y}\sum_{a\,(\mathrm{mod}\,q)}^{*}\Big|\sum_{n\in\mathcal{N}}e(na/q)\Big|^2.$$

The right-hand side is $\ll(N+y^2)|\mathcal{N}|$ by Theorem 25.14, thus completing the proof of the theorem. □

Theorem 25.16 is of comparable strength to Theorem 21.1. As a matter of fact, if we use the improved large sieve inequality mentioned in Remark 25.13, we can obtain a new proof of Theorem 21.4 (see also [146] for a further improvement of this result). However, the true strength of the large sieve is revealed when the sets R_p have unbounded cardinality on average, in which case the sifting dimension is also unbounded. We illustrate this point by studying Vinogradov's *least quadratic nonresidue problem*.

Given a prime p, we let n_p denote the *least quadratic nonresidue*, that is to say, the smallest integer $n \geqslant 1$ for which $(n|p) = -1$. We know from elementary number theory that for half of the integers $n \in [1, p-1]$ we have $(n|p) = -1$. In fact, since the Legendre symbol $(\cdot|p)$ is a non-principal Dirichlet character mod p, we have

$$\sum_{n \leqslant N} \left(\frac{n}{p}\right) = O(\sqrt{p}\log p)$$

by the Pólya-Vinogradov inequality. In particular, $n_p = O(\sqrt{p}\log p)$. Exercise 25.6 establishes the improved bound

$$n_p \leqslant p^{1/2\sqrt{e}+o(1)} \qquad (p \to \infty). \tag{25.30}$$

Vinogradov conjectured that the stronger estimate

$$n_p = O_\varepsilon(p^\varepsilon)$$

is true for each fixed $\varepsilon > 0$. We use the large sieve to show that Vinogradov's conjecture holds for the vast majority of primes.

Theorem 25.17. *Fix $\varepsilon > 0$. For $x \geqslant 3$, we have*

$$\#\{\, p \leqslant x : n_p \geqslant p^\varepsilon \,\} \ll_\varepsilon \log\log x.$$

Proof. For every $y \geqslant 1$, we will show that

$$\#\{\, p \leqslant y : n_p \geqslant y^\varepsilon \,\} = O_\varepsilon(1). \tag{25.31}$$

The theorem then follows by setting $y_j = e^{e^j}$ and noticing that

$$\#\{\, p \leqslant x : n_p \geqslant p^\varepsilon \,\} \leqslant O(1) + \sum_{j \leqslant \log\log x} \#\{\, y_{j-1} < p \leqslant y_j : n_p \geqslant y_j^{\varepsilon/e} \,\}.$$

To show (25.31), let

$$\mathcal{N} = \{\, m \leqslant y^2 : P^+(m) \leqslant y^\varepsilon \,\}.$$

On the one hand, $|\mathcal{N}| \gg_\varepsilon y^2$ by Theorem 14.5. On the other hand, we can use the large sieve to bound $|\mathcal{N}|$: for each prime p with $n_p > y^\varepsilon$, we let

$$R_p = \{\, a \,(\mathrm{mod}\, p) : (a|p) \in \{0, -1\}\,\};$$

otherwise, we let $R_p = \emptyset$. We claim that $\mathcal{N}$ avoids the sets R_p.

Indeed, let p be a prime and $n \in \mathcal{N}$. If $n_p \leqslant y^\varepsilon$, then $R_p = \emptyset$, so we naturally have $n \notin R_p \,(\mathrm{mod}\, p)$. Assume now that $n_p > y^\varepsilon$. Since $P^+(m) \leqslant y^\varepsilon < n_p$, we have $(p'|p) = 1$ for all $p'|n$. The multiplicativity of the Legendre symbol thus implies that $(n|p) = 1$, that is to say, $n \notin R_p \,(\mathrm{mod}\, p)$, as claimed.

From the above discussion, we may apply Theorem 25.16 to find that

$$|\mathcal{N}| \ll y^2 \Big/ \sum_{m \leqslant y^2} \mu^2(m) f(m),$$

where $f(m) = \prod_{p|m} |R_p|/(p - |R_p|)$. Since $|\mathcal{N}| \gg_\varepsilon y^2$, we conclude that

$$\sum_{m \leqslant N} \mu^2(m) f(m) = O_\varepsilon(1).$$

But note that if $p \geqslant 3$ is a prime with $n_p > y^\varepsilon$, then $f(p) = (p+1)/(p-1)$ by the definition of R_p. In particular,

$$\sum_{m \leqslant N} \mu^2(m) f(m) \geqslant \#\{\, p \leqslant y : n_p > y^\varepsilon \,\}.$$

This proves that (25.31) holds, thus completing the proof of the theorem. □

Exercises

Exercise 25.1. Let $\mathcal{C} = \{\chi_1, \dots, \chi_M\}$ be a set of Dirichlet characters, where χ_m is a character to the modulus q_m.

(a) We say that $\mathcal{C}$ is *reduced* if $\chi_{m_1}\overline{\chi}_{m_2}$ is non-principal when $m_1 \neq m_2$. Show that the following sets of characters are reduced: (i) any subset of $\{\chi \,(\mathrm{mod}\, q)\}$; (ii) any set of primitive Dirichlet characters.

(b) Assume that $\mathcal{C}$ is reduced and let

$$Q = \max\{\, [q_{m_1}, q_{m_2}] : 1 \leqslant m_1, m_2 \leqslant M,\ m_1 \neq m_2 \,\}.$$

Prove that

$$\sum_{\chi \in \mathcal{C}} \Big| \sum_{H < n \leqslant H+N} c_n \chi(n) \Big|^2 \leqslant (N + O(M\sqrt{Q}\log Q)) \|\vec{c}\|_2^2$$

for all $N \in \mathbb{Z}_{\geqslant 1}$, $H \in \mathbb{Z}$ and $\vec{c} = (c_{H+1}, \dots, c_{H+N}) \in \mathbb{C}^N$.

Exercise 25.2. Let $\chi \,(\mathrm{mod}\, q)$ be a fixed non-principal Dirichlet character, and let $f, g : \mathbb{N} \to \{\, z \in \mathbb{C} : |z| \leqslant 1 \,\}$ be supported on $[1, M]$ and $[1, N]$, respectively. Explain why the method of bilinear forms *cannot* yield a general non-trivial bound for the sum $\sum_{n \leqslant x} (f * g)(n)\chi(n)$.

Exercise 25.3. Adapt the proof of Theorem 25.11 to show that the factor $\sqrt{\log(2q)}$ can be removed from the statement of Theorem 23.6. Conclude that the exponent of $\log x$ in Theorem 23.8 can be improved from $5/2$ to 2.

Exercise 25.4. Let $N, J \in \mathbb{Z}_{\geqslant 1}$, and set $N_1 = \lfloor N/2 \rfloor$. Define the function $g : \mathbb{Z} \to [0,1]$ by letting $g(n) = 1$ when $|n| \leqslant N_1$, $g(n) = 1 - (|n| - N_1)/J$ when $N_1 < |n| \leqslant N_1 + J$, and $g(n) = 0$ when $|n| > N_1 + J$.

(a) Prove that $1_{|n| \leqslant N/2} \leqslant g(n)$ for all $n \in \mathbb{Z}$ and

$$\sum_{n \in \mathbb{Z}} g(n) e(n\alpha) = J^{-1} \cdot \big[(N_1 + J) F_{N_1+J}(\alpha) - N_1 F_{N_1}(\alpha) \big],$$

where F_k denotes the Fejér kernel.

(b) Choose an appropriate J to prove that the left-hand side of Theorem 25.11 is $\leqslant (N + 1 + \pi\sqrt{6}\delta^{-1}/3) \|\vec{c}\|_2^2$.

Exercise 25.5 (Gallagher)**.**

(a) Given $f \in C^1([a,b])$ and $c \in [a,b]$, prove that

$$|f(c)| \leqslant \frac{1}{b-a} \int_a^b |f(x)| \mathrm{d}x + \frac{\max\{c-a, b-c\}}{b-a} \int_a^b |f'(x)| \mathrm{d}x.$$

[*Hint:* Note that $\int_a^c (f(x) - f(c)) \mathrm{d}x = \int_a^c (a - x) f'(x) \mathrm{d}x$ and $\int_c^b (f(x) - f(c)) \mathrm{d}x = \int_c^b (b - x) f'(x) \mathrm{d}x$.]

(b) If $S(\alpha) = \sum_{H < n \leqslant H+N} c_n e(n\alpha)$ and the points $\alpha_1, \dots, \alpha_R$ are δ-spaced mod 1, prove that

$$\sum_{r=1}^{R} |S(\alpha_r)|^2 \leqslant \frac{1}{\delta} \int_0^1 |S(\alpha)|^2 \mathrm{d}\alpha + \int_0^1 |S(\alpha) S'(\alpha)| \mathrm{d}\alpha.$$

(c) Prove that the left-hand side of 25.11 is $\leqslant (\pi N + \delta^{-1}) \|\vec{c}\|_2^2$.

Exercise 25.6. Let p be an odd prime and let n_p be the least quadratic non-residue mod p.

(a) Prove that if $P^+(m) < n_p$, then $(m|p) = 1$.

(b) Deduce Vinogradov's bound $n_p \leqslant p^{1/2\sqrt{e} + o(1)}$. [*Hint:* For all $x \geqslant 1$, show that $\sum_{m \leqslant x,\, P^+(m) < n_p} (m|p) \leqslant O(\sqrt{p} \log p) + \sum_{m \leqslant x,\, P^+(m) \geqslant n_p} 1$.]

Exercise 25.7. Assume the Generalized Riemann Hypothesis.

(a) Let χ be a non-principal character mod q. In addition, let ϕ be a smooth function supported on $[1,2]$ with Mellin transform Φ. Show that

$$\sum_{n \geqslant 1} \Lambda(n) \chi(n) \phi(n/x) \ll x^{1/2} \log q \qquad (x \geqslant 1).$$

[*Hint:* Exercise 11.4.]

(b) Prove that $n_p \ll (\log p)^2$ for all p.

(c) Use the Chinese Remainder Theorem to prove that if x is large enough, then there is a prime $p \in [x/2, x]$ such that $(q|p) = 1$ for all primes $q \leqslant 0.49 \log x$. In particular, there is an infinite sequence of primes p with $n_p \geqslant 0.49 \log p$.

Chapter 26

The Bombieri-Vinogradov theorem

Having developed the large sieve, we present here one of its most important applications: a proof of the celebrated Bombieri-Vinogradov theorem (Theorem 18.9). Let us recall its statement: for each fixed $A \geqslant 0$ we have

$$\sum_{q \leqslant Q} \max_{y \leqslant x} \max_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \left| \pi(y; q, a) - \frac{\mathrm{li}(y)}{\varphi(q)} \right| \ll_A \frac{x}{(\log x)^{A+1}} \tag{26.1}$$

uniformly for $x \geqslant 2$ and $1 \leqslant Q \leqslant \sqrt{x}/(\log x)^{A+3}$.

This result is often called "the Riemann Hypothesis on average". Indeed, the Generalized Riemann Hypothesis implies that

$$\max_{y \leqslant x} \max_{a \in (\mathbb{Z}/q\mathbb{Z})^*} \left| \pi(y; q, a) - \frac{\mathrm{li}(y)}{\varphi(q)} \right| \ll \sqrt{x} \log(qx) \tag{26.2}$$

for all $q \in \mathbb{N}$ and all $x \geqslant 2$ (see Exercise 11.2). Conversely, Theorem 6.1 and Exercise 11.3(b) show that knowing (26.2) for all $x \geqslant 2$ and all $q \in \mathbb{N}$ implies the Generalized Riemann Hypothesis. Now, observe that (26.2) implies (26.1) with $Q = \sqrt{x}/(\log x)^{A+2}$, which is bigger only by a factor of $\log x$ than the largest value of Q furnished by the Bombieri-Vinogradov theorem. In comparison, note that the Siegel-Walfisz theorem (Theorem 12.1) provides a much poorer range of validity of (26.1), allowing us to establish it only for $Q \leqslant (\log x)^C$, where C is arbitrarily large but nevertheless fixed.

We thus see that if we need to estimate $\pi(x; q, a) - \mathrm{li}(x)/\varphi(q)$ on average over q, the Bombieri-Vinogradov is just as good as the unproven Generalized Riemann Hypothesis. And having access to such strong averaged estimates is of crucial importance in sieve theory (see the discussion in Example 18.8).

Preliminaries

To prove the Bombieri-Vinogradov theorem, we will decompose von Mangoldt's function into type I and type II functions using Vaughan's identity. We will then examine the distribution of each type in arithmetic progressions employing different arguments. But first we must perform some preparatory steps.

Using a more natural main term. Throughout this chapter, we will employ the notation

$$\Delta_f(x;q,a) = \sum_{\substack{n\leqslant x\\ n\equiv a\,(\mathrm{mod}\,q)}} f(n) - \frac{1}{\varphi(q)} \sum_{\substack{n\leqslant x\\ (n,q)=1}} f(n),$$

where f is an arithmetic function, $x \in \mathbb{R}_{\geqslant 1}$, $q \in \mathbb{N}$ and $a \in (\mathbb{Z}/q\mathbb{Z})^*$. This quantity will be small if f is well-distributed among reduced arithmetic progressions. Moreover, it admits a convenient representation in terms of Dirichlet characters: we have

$$\Delta_f(y;q,a) = \frac{1}{\varphi(q)} \sum_{\substack{\chi\,(\mathrm{mod}\,q)\\ \chi\neq\chi_0}} \overline{\chi}(a) \sum_{n\leqslant y} f(n)\chi(n). \tag{26.3}$$

Now, let 1_P denote the indicator function of primes. We want to replace $\mathrm{li}(y)$ in (26.1) with $\sum_{p\leqslant y,\, p\nmid q} 1$, so that we can express the left-hand side of (26.1) using the quantity Δ_{1_P} to which we can apply (26.3). We have

$$\sum_{p\leqslant y,\, p\nmid q} 1 = \pi(y) + O(\log q) = \mathrm{li}(y) + O\big(ye^{-c\sqrt{\log y}} + \log q\big)$$

for $y \geqslant 2$ and $q \in \mathbb{N}$, where c is the constant from Theorem 8.1 (the Prime Number Theorem). This reduces (26.5) to proving that

$$\sum_{q\leqslant Q} \max_{y\leqslant x} \max_{a\in(\mathbb{Z}/q\mathbb{Z})^*} |\Delta_{1_P}(x;q,a)| \ll_A \frac{x}{(\log x)^{A+1}}$$

uniformly for $x \geqslant 2$ and $1 \leqslant Q \leqslant \sqrt{x}/(\log x)^{A+3}$.

Switching to von Mangoldt's function. The next step, in preparation for the application of Vaughan's identity, is to switch from 1_P to Λ. This is accomplished by the following estimate.

Lemma 26.1. *For $x \geqslant 2$, $q \in \mathbb{N}$ and $a \in (\mathbb{Z}/q\mathbb{Z})^*$, we have*

$$\max_{y\leqslant x} |\Delta_{1_P}(y;q,a)| \ll \frac{1}{\log x}\Big(\max_{\sqrt{x}\leqslant y\leqslant x} |\Delta_\Lambda(y;q,a)| + \sqrt{x}\Big).$$

Proof. If $y \leqslant \sqrt{x}$, we have $|\Delta_{1_P}(y;q,a)| \ll \sqrt{x}/\log x$. Assume now that $y \in [\sqrt{x}, x]$. Chebyshev's estimate (Theorem 2.4) implies that

$$\sum_{p\leqslant\sqrt{x}} 1 + \sum_{k\geqslant 2}\sum_{p^k\leqslant x} 1 = \sum_{p\leqslant\sqrt{x}}\sum_{1\leqslant k\leqslant \frac{\log x}{\log p}} 1 \leqslant \sum_{p\leqslant\sqrt{x}} \frac{\log x}{\log p} \ll \frac{\sqrt{x}}{\log x}. \tag{26.4}$$

Hence, if $y \in [\sqrt{x}, x]$, we have

$$\Delta_{1_P}(y;q,a) = \sum_{\substack{\sqrt{x}<n\leqslant y\\ n\equiv a \,(\mathrm{mod}\, q)}} \frac{\Lambda(n)}{\log n} - \frac{1}{\varphi(q)} \sum_{\substack{\sqrt{x}<n\leqslant y\\ (n,q)=1}} \frac{\Lambda(n)}{\log n} + O\Big(\frac{\sqrt{x}}{\log x}\Big).$$

Employing partial summation, we find that

$$\begin{aligned}\Delta_{1_P}(y;q,a) &= \int_{\sqrt{x}}^{y} \frac{1}{\log t} \mathrm{d}\Delta_\Lambda(t;q,a) + O\Big(\frac{\sqrt{x}}{\log x}\Big)\\ &= \frac{\Delta_\Lambda(t;q,a)}{\log t}\Big|_{t=\sqrt{x}}^{y} + \int_{\sqrt{x}}^{y} \frac{\Delta_\Lambda(t;q,a)}{t\log^2 t}\mathrm{d}t + O\Big(\frac{\sqrt{x}}{\log x}\Big).\end{aligned}$$

Applying the triangle inequality and then bounding $|\Delta_\Lambda(t;q,a)|$ by its maximum value over $t \in [\sqrt{x}, x]$ completes the proof of the lemma. □

Lemma 26.1 reduces the Bombieri-Vinogradov theorem to showing that

$$\sum_{q\leqslant Q} \max_{\sqrt{x}\leqslant y\leqslant x} \max_{a\in(\mathbb{Z}/q\mathbb{Z})^*} |\Delta_\Lambda(y;q,a)| \ll_A \frac{x}{(\log x)^A} \tag{26.5}$$

uniformly for $x \geqslant 2$ and $1 \leqslant Q \leqslant \sqrt{x}/(\log x)^{A+3}$.

Applying Vaughan's identity. The key to proving (26.5) is a combinatorial decomposition of von Mangoldt's function in terms of type I and type II functions. We perform this decomposition by appealing to Vaughan's identity (Lemma 23.1): we have $\Lambda = \Lambda^\sharp + \Lambda^\flat + \Lambda_{\leqslant U}$ for some $U, V \in [1, x]$ to be chosen later, where we recall that

$$\Lambda^\sharp = \mu_{\leqslant V} * \log\, -\, (\Lambda_{\leqslant U} * \mu_{\leqslant V}) * 1 \quad \text{and} \quad \Lambda^\flat = (\Lambda_{>U} * 1) * \mu_{>V}.$$

As we will see, we may work with any choice of U and V satisfying the conditions

$$UV \leqslant \sqrt{x} \quad \text{and} \quad U, V \geqslant e^{\sqrt{\log x}}. \tag{26.6}$$

The contribution of the term $\Lambda_{\leqslant U}$ is bounded trivially: we simply note that $\sum_{n\leqslant U}\Lambda(n) \ll U$, whence $\Delta_{\Lambda_{\leqslant U}}(y;q,a) \ll U \leqslant \sqrt{x}$ for all $q, a \in \mathbb{N}$ and all $y \leqslant x$. As a consequence,

$$\sum_{q\leqslant Q} \max_{y\leqslant x} \max_{a\in(\mathbb{Z}/q\mathbb{Z})^*} |\Delta_{\Lambda_{\leqslant U}}(y;q,a)| \ll Q\sqrt{x} \leqslant \frac{x}{(\log x)^{A+2}} \tag{26.7}$$

for $x \geqslant 2$ and $1 \leqslant Q \leqslant \sqrt{x}/(\log x)^{A+3}$.

It remains to study the distribution of $\Lambda^\sharp$ and $\Lambda^\flat$ in reduced arithmetic progressions. We start with the former.

Type I functions in arithmetic progressions

The study of type I functions in reduced arithmetic progressions is relatively easy.[1] We have the following general result, proven by a straightforward application of the simplest version of Dirichlet's hyperbola method.

Theorem 26.2. *Let $v \geqslant 0$, and let f be an arithmetic function supported on $[1, y]$. For $x \geqslant 2$, $q \in \mathbb{N}$ and $a \in (\mathbb{Z}/q\mathbb{Z})^*$, we have*

$$|\Delta_{f*\log^v}(x; q, a)| \leqslant 2(\log x)^v \sum_{k\leqslant y} |f(k)|.$$

Proof. Given $k \in (\mathbb{Z}/q\mathbb{Z})^*$, we write $\overline{k}$ for its multiplicative inverse mod q. Notice that if $n \equiv a \,(\mathrm{mod}\, q)$, then $(n, q) = 1$. Therefore

$$\Delta_{f*\log^v}(x; q, a) = \sum_{\substack{k\leqslant y\\(k,q)=1}} f(k)\Bigg(\sum_{\substack{\ell\leqslant x/k\\ \ell\equiv a\bar{k}\,(\mathrm{mod}\, q)}} (\log \ell)^v - \frac{1}{\varphi(k)} \sum_{\substack{\ell\leqslant x/k\\(\ell,q)=1}} (\log \ell)^v \Bigg).$$

The expression inside the parentheses equals $\Delta_{\log^v}(x/k; q, a\bar{k})$. Hence, the theorem is reduced to proving that

$$|\Delta_{\log^v}(t; q, j)| \leqslant 2(\log t)^v \qquad (t \geqslant 1,\ j \in (\mathbb{Z}/q\mathbb{Z})^*). \tag{26.8}$$

We first prove (26.8) when $v = 0$. We start by noticing that

$$\Delta_1(t; q, j) = \frac{1}{\varphi(q)} \sum_{j'\in(\mathbb{Z}/q\mathbb{Z})^*} \Bigg(\sum_{\substack{n\leqslant t\\ n\equiv j\,(\mathrm{mod}\, q)}} 1 - \sum_{\substack{n\leqslant t\\ n\equiv j'\,(\mathrm{mod}\, q)}} 1 \Bigg).$$

Now, fix $b \in \mathbb{Z}$. Since each string of q consecutive integers contains exactly one integer in the class $b \,(\mathrm{mod}\, q)$, the number of $n \in \mathbb{Z} \cap [1, t]$ in the class $b \,(\mathrm{mod}\, q)$ is either $\lfloor t/q \rfloor$ or $\lfloor t/q \rfloor + 1$. We deduce that $|\Delta_1(t; q, j)| \leqslant 1$, which proves (a stronger former of) (26.8) when $v = 0$.

Finally, when $v > 0$, we use partial summation to find that

$$\Delta_{\log^v}(t; q, j) = \int_1^t (\log s)^v \mathrm{d}\Delta_1(s; q, j).$$

Integrating by parts and using the already proven fact that $|\Delta_1(s; q, j)| \leqslant 1$ yields (26.8) in this case too, thus completing the proof of the theorem. □

[1] The distribution of functions of multiplicative nature can be significantly more complicated over non-reduced progressions (see Exercise 26.1). Of course, focusing on reduced progressions is sufficient for applications to the theory of prime numbers.

As an immediate corollary, we have an estimate for the distribution of $\Lambda^\sharp$ in arithmetic progressions, which is the "structured" part of Λ.

Corollary 26.3. *For $U, V \geqslant 1$, $x \geqslant 2$, $q \in \mathbb{N}$ and $a \in (\mathbb{Z}/q\mathbb{Z})^*$, we have*

$$\max_{y \leqslant x} \max_{a \in (\mathbb{Z}/q\mathbb{Z})^*} |\Delta_{\Lambda^\sharp}(y; q, a)| \ll UV \log x.$$

The above result readily implies the estimate

$$\sum_{q \leqslant Q} \max_{y \leqslant x} \max_{a \in (\mathbb{Z}/q\mathbb{Z})^*} |\Delta_{\Lambda^\sharp}(y; q, a)| \ll QUV \log x. \tag{26.9}$$

If $Q \leqslant \sqrt{x}/(\log x)^{A+3}$ and $UV \leqslant \sqrt{x}$ (as we assumed in (26.6)), then the right-hand side of (26.9) is $\leqslant Q\sqrt{x}(\log x) \leqslant x/(\log x)^{A+2}$. Together with (26.7), this reduces the Bombieri-Vinogradov theorem to proving that

$$\sum_{q \leqslant Q} \max_{\sqrt{x} \leqslant y \leqslant x} \max_{a \in (\mathbb{Z}/q\mathbb{Z})^*} |\Delta_{\Lambda^\flat}(x; q, a)| \ll \frac{x}{(\log x)^A} \tag{26.10}$$

uniformly for $x \geqslant 2$ and $1 \leqslant Q \leqslant \sqrt{x}/(\log x)^{A+3}$.

Type II functions in arithmetic progressions

The main result we will use to study $\Lambda^\flat$ in arithmetic progressions to large moduli is Theorem 25.15 (the multiplicative large sieve inequality). It turns out that this result can only handle the contribution of Dirichlet characters of large conductor to the sum in (26.10). To deal with characters of small conductor, we need the following estimate.

Theorem 26.4. *Fix $A, C \geqslant 1$. If $x \geqslant 3$, $U \in [1, x]$, $V \in [e^{\sqrt{\log x}}, x]$, $r \in \mathbb{N}_{\leqslant x}$ and χ is a character of modulus $q \leqslant (\log x)^C$, then*

$$\max_{\sqrt{x} \leqslant y \leqslant x} \Big| \sum_{n \leqslant y} \Lambda^\flat(n)\chi(n)1_{(n,r)=1} \Big| \ll_{A,C} \frac{x}{(\log x)^A}.$$

Proof. The result is proven by a modification of the second part of the proof of Theorem 24.3. First, we open the convolution $\Lambda^\flat(n) = \sum_{k\ell=n}(1 * \Lambda_{>U})(k)\mu_{>V}(\ell)$. Then, we fix the congruence class of $\ell \pmod q$ and use Corollary 13.4. We leave the details as an exercise. □

Remark 26.5. In Chapter 24, we noticed how the bilinear methods are perfectly complemented by the Siegel-Walfisz theorem, thus yielding results such as Theorem 24.3 that cover all possible values of α. We see this complementarity manifesting itself again in the proof of the Bombieri-Vinogradov theorem: bilinear methods will handle characters of large conductor, but we have to resort to the Siegel-Walfisz theorem (via an application of Corollary 13.4) to handle characters of small conductor.

Finally, as a more technical remark, we mention that it is possible to prove (a version of) (23.12) for non-principal characters by a direct appeal to the Siegel-Walfisz theorem, thus circumventing the use of Corollary 13.4. Indeed, when $(a,q)=1$, $r\in\mathbb{N}_{\leqslant x}$ and $UV\leqslant x/e^{\sqrt{\log x}}$, we claim that

$$\max_{\sqrt{x}\leqslant y\leqslant x}|\Delta_{\Lambda_r^\flat}(y;q,a)|\ll_{A,C} x/(\log x)^A,$$

where $\Lambda_r^\flat(n)=\Lambda^\flat(n)1_{(n,r)=1}$. This estimate is good enough for the purpose of establishing the Bombieri-Vinogradov theorem. To prove it, we start by writing $\Lambda^\flat=\Lambda-\Lambda^\sharp-\Lambda_{\leqslant U}$. After multiplying this identity with the indicator function $n\to 1_{(n,r)=1}$, we apply the Siegel-Walfisz theorem (Theorem 12.1) to the first function on the right side and Theorem 26.2 to the second one. The details are left as an exercise. □

Reduction to character sums. We now show how to use Theorem 26.4 to reduce the proof of (26.10) to a certain large sieve estimate involving sums of $\Lambda^\flat$ twisted by Dirichlet characters. It is convenient to introduce some notation for these character sums: for the rest of this chapter, we let

$$S_r(x,\chi)=\max_{\sqrt{x}\leqslant y\leqslant x}\Big|\sum_{n\leqslant y}\Lambda^\flat(n)\chi(n)1_{(n,r)=1}\Big|.$$

By (26.3) with $f=\Lambda^\flat$ and the triangle inequality, we find that

$$\max_{\sqrt{x}\leqslant y\leqslant x}\max_{a\in(\mathbb{Z}/q\mathbb{Z})^*}|\Delta_{\Lambda^\flat}(x;q,a)|\leqslant\frac{1}{\varphi(q)}\sum_{\substack{\chi\,(\mathrm{mod}\,q)\\ \chi\neq\chi_0}}S_1(x,\chi).$$

Now, let $\xi\,(\mathrm{mod}\,d)$ be the primitive character inducing χ, so that d is the conductor of χ. We then have that $d|q$ and $\chi(n)=1_{(n,q)=1}\xi(n)=1_{(n,q/d)=1}\xi(n)$, where the second equality follows by noticing that the coprimality of n and d is encoded in the definition of $\xi(n)$. Hence,

$$\max_{\sqrt{x}\leqslant y\leqslant x}\max_{a\in(\mathbb{Z}/q\mathbb{Z})^*}|\Delta_{\Lambda^\flat}(x;q,a)|\leqslant\frac{1}{\varphi(q)}\sum_{d|q,\,d>1}\sideset{}{^*}\sum_{\xi\,(\mathrm{mod}\,d)}S_{q/d}(x,\xi).$$

Fix $C\geqslant 0$ to be chosen later. When $d\leqslant(\log x)^C$, we use Theorem 26.4 with $A+2C+1$ in place of A (recall that we have assumed (26.6)). There are $\leqslant(\log x)^{2C}$ pairs (d,ξ) with $d\leqslant(\log x)^C$. Consequently,

$$\max_{\substack{\sqrt{x}\leqslant y\leqslant x\\ a\in(\mathbb{Z}/q\mathbb{Z})^*}}|\Delta_{\Lambda^\flat}(x;q,a)|\leqslant\sum_{\substack{d|q,\ d>(\log x)^C}}\sideset{}{^*}\sum_{\xi\,(\mathrm{mod}\,d)}\frac{S_{q/d}(x,\xi)}{\varphi(q)}+O\Big(\frac{x}{\varphi(q)(\log x)^{A+1}}\Big).$$

Since $\varphi(q) \geqslant \varphi(d)\varphi(r)$ when $q = dr$, we infer that

$$\sum_{q\leqslant Q} \max_{\substack{\sqrt{x}\leqslant y\leqslant x\\ a\in(\mathbb{Z}/q\mathbb{Z})^*}} |\Delta_{\Lambda^\flat}(x;q,a)| \leqslant \sum_{r\leqslant Q} \frac{T_r(x,Q)}{\varphi(r)} + O(x/(\log x)^A), \tag{26.11}$$

where

$$T_r(x,Q) := \sum_{(\log x)^C < d \leqslant Q/r} \frac{1}{\varphi(d)} \sideset{}{^*}\sum_{\xi \,(\mathrm{mod}\, d)} S_r(x,\xi).$$

A large sieve inequality. The next step is to bound $T_r(x,Q)$ using the large sieve in the form of Theorem 25.15. However, this result establishes a bound for the mean square (also called the "second moment") of character sums $\sum_{n\leqslant x} c_n\chi(n)$ when we average over primitive characters χ of conductor $\leqslant Q$. To prove the Bombieri-Vinogradov theorem, we need to estimate the first moment of $\sum_{n\leqslant x} \Lambda^\flat(n)\chi(n)1_{(n,r)=1}$. It turns out that the bilinear structure of $\Lambda^\flat$ allows us to use the Cauchy-Schwarz inequality and pass from the first moment of $\sum_{n\leqslant x} \Lambda^\flat(n)\chi(n)1_{(n,r)=1}$ to a product of two second moments of two other character sums. As a matter of fact, we have the following estimate for general type II functions.

Theorem 26.6. *Let f and g be two arithmetic functions supported on $[1,M]$ and $[1,N]$, respectively. For $x, Q \geqslant 1$ we have*

$$\sum_{q\leqslant Q} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \frac{q}{\varphi(q)} \max_{y\leqslant x} \Big| \sum_{n\leqslant y} (f*g)(n)\chi(n)\Big| \\ \ll (\sqrt{MN} + \sqrt{M}Q + \sqrt{N}Q + Q^2)(\log x)\, \|f\|_2\, \|g\|_2.$$

Proof. Since we are only considering integers $n \leqslant x$, we may assume that $x \geqslant M, N$. Indeed, if for example $x > M$, then we replace f by $f \cdot 1_{[1,x]}$.

Let S be the sum in the statement of the theorem, which we write

$$S = \sum_{q\leqslant Q} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \frac{q}{\varphi(q)} \max_{y\leqslant x} \Big| \sum_{m\leqslant M} \sum_{n\leqslant N} f(m)\chi(m)g(n)\chi(n)1_{mn\leqslant y}\Big|.$$

We want to apply the Cauchy-Schwarz inequality to S to separate the variables m, n and pass to a product of two second moments, so that we can apply Theorem 25.15 to each variable separately. However, there are two technical obstacles. First, the variables m and n are tangled in the indicator function $1_{mn\leqslant y}$; second, we have to take the maximum over $y \leqslant x$. We take care of both of these issues simultaneously by an application of Perron's inversion formula.

As a preparatory step, note that $mn \leqslant y$ if and only if $mn \leqslant \lfloor y\rfloor + 1/2$. Hence, in the definition of S we may replace $\max_{y\leqslant x}$ by $\max_{y=k+1/2,\, k\in\mathbb{N},\, k\leqslant x}$.

Now, let $y = k+1/2$ for some integer $k \in [1, x]$. Lemma 7.1 with $\alpha = 1/\log x$, $T = x^2$ and mn/y in place of y implies that

$$1_{mn\leqslant y} = \frac{1}{2\pi i}\int_{\substack{\mathrm{Re}(s)=\alpha\\ |\mathrm{Im}(s)|\leqslant x^2}} \frac{(y/mn)^s}{s}\mathrm{d}s + O\Big(\frac{(y/mn)^\alpha}{x^2|\log(y/mn)|}\Big).$$

We have $|y - mn| \geqslant 1/2$ for all integers m, n, by our assumption on y. Therefore, $|\log(y/mn)| \gg 1/y \gg 1/x$. Moreover, $(y/mn)^\alpha \ll 1$ for $y \leqslant x + 1/2$ and m, $n \geqslant 1$. We thus conclude that

$$\sum_{m\leqslant M}\sum_{n\leqslant N} f(m)\chi(m)g(n)\chi(n)1_{mn\leqslant y} = \frac{1}{2\pi}\int_{-x^2}^{x^2} F_t(\chi)G_t(\chi)\frac{y^{\alpha+it}}{\alpha+it}\mathrm{d}t + O\Big(x^{-1}\sum_{m\leqslant M}|f(m)|\sum_{n\leqslant N}|g(n)|\Big), \tag{26.12}$$

where

$$F_t(\chi) = \sum_{m\leqslant M}\frac{f(m)\chi(m)}{m^{\alpha+it}} \qquad \text{and} \qquad G_t(\chi) = \sum_{n\leqslant N}\frac{g(n)\chi(n)}{n^{\alpha+it}}.$$

The Cauchy-Schwarz inequality and our assumption that $M, N \leqslant x$ imply that

$$\sum_{m\leqslant M}|f(m)| \leqslant x^{1/2}\|f\|_2 \quad \text{and} \quad \sum_{n\leqslant N}|g(n)| \leqslant x^{1/2}\|g\|_2.$$

In the main term of (26.12), we note that $|y^{\alpha+it}| \ll 1$ for $y \leqslant x + 1/2$, as well as that $|\alpha + it| \asymp \max\{\alpha, |t|\}$. Therefore,

$$\sum_{m\leqslant M}\sum_{n\leqslant N} f(m)\chi(m)g(n)\chi(n)1_{mn\leqslant y} \ll \int_{-x^2}^{x^2}\frac{|F_t(\chi)G_t(\chi)|}{\max\{\alpha,|t|\}}\mathrm{d}t + \|f\|_2\|g\|_2.$$

The right-hand side no longer depends on y. Consequently,

$$S \ll \int_{-x^2}^{x^2}\Big(\sum_{q\leqslant Q}\sideset{}{^*}\sum_{\chi\,(\mathrm{mod}\,q)}\frac{q}{\varphi(q)}\cdot|F_t(\chi)|\cdot|G_t(\chi)|\Big)\frac{\mathrm{d}t}{\max\{\alpha,|t|\}} + Q^2\|f\|_2\|g\|_2.$$

By the Cauchy-Schwarz inequality and two applications of Theorem 25.15, one where we take $c_n = f(n)/n^{\alpha+it}$ for $n \in [1, M]$, and another one with $c_n = g(n)/n^{\alpha+it}$ for $n \in [1, N]$, we conclude that

$$\sum_{q\leqslant Q}\sideset{}{^*}\sum_{\chi\,(\mathrm{mod}\,q)}\frac{q}{\varphi(q)}\cdot|F_t(\chi)|\cdot|G_t(\chi)| \ll \sqrt{(M+Q^2)(N+Q^2)}\|f\|_2\|g\|_2.$$

Since $\sqrt{(M+Q^2)(N+Q^2)} \asymp \sqrt{MN} + \sqrt{M}Q + \sqrt{N}Q + Q^2$ and

$$\int_{-x^2}^{x^2}\frac{\mathrm{d}t}{\max\{\alpha,|t|\}} = \int_{|t|\leqslant\alpha}\frac{\mathrm{d}t}{\alpha} + \int_{\alpha\leqslant|t|\leqslant x^2}\frac{\mathrm{d}t}{|t|} \ll \log x,$$

the theorem has been established. □

Corollary 26.7. *For $x, Q \geqslant 2$, $U, V \in [1, x]$ and $r \in \mathbb{N}$, we have*

$$\sum_{q\leqslant Q}\sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \frac{q}{\varphi(q)}\, S_r(y,\chi) \ll \Big(x + \frac{Qx}{\sqrt{U}} + \frac{Qx}{\sqrt{V}} + Q^2\sqrt{x}\Big)(\log x)^3.$$

Proof. We begin by writing $\Lambda^\flat$ using (23.12), which we also multiply with the indicator function of integers coprime to r:

$$\Lambda^\flat(n)1_{(n,r)=1} = \sum_{U<2^j\leqslant 2x/V} (\alpha_j * \beta_j)(n) \quad \text{for} \quad n \leqslant x,$$

where we have set $\alpha_j(k) = (\Lambda_{>U} * 1)(k)1_{2^{j-1}<k\leqslant 2^j}1_{(k,r)=1}$ and $\beta_j(\ell) = \mu_{>V}(\ell)1_{\ell\leqslant x/2^{j-1}}1_{(\ell,r)=1}$. Therefore, Theorem 26.6 with $f = \alpha_j$, $M = 2^j$, $g = \beta_j$ and $N = x/2^{j-1}$ implies that

$$\begin{aligned}\sum_{q\leqslant Q}\frac{q}{\varphi(q)}\sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} \max_{y\leqslant x}\Big|\sum_{n\leqslant y}(\alpha_j * \beta_j)(n)\chi(n)\Big| \\ \ll (\sqrt{x} + 2^{j/2}Q + \sqrt{x}2^{-j/2}Q + Q^2)(\log x)\|\alpha_j\|_2\|\beta_j\|_2 \\ \ll (x + 2^{j/2}\sqrt{x}Q + x2^{-j/2}Q + \sqrt{x}Q^2)(\log x)^2,\end{aligned}$$

where we bounded $\|\alpha_j\|_2\|\beta_j\|_2$ by $O(\sqrt{x}\log x)$ using the inequalities $|\alpha_j| \leqslant \log$ and $|\beta_j| \leqslant 1$. Summing the above estimate over $2^j \in [U, 2x/V]$ (there are $\ll \log x$ such choices for j) completes the proof of the corollary. □

The easiest way to pass from the above estimate to a bound for $T_r(x; Q)$ is to use a dyadic decomposition trick: we have

$$T_r(x,Q) \leqslant \sum_{(\log x)^C<2^j\leqslant 2Q/r} \frac{1}{2^{j-1}} \sum_{2^{j-1}<d\leqslant 2^j} \frac{d}{\varphi(d)} \sideset{}{^*}\sum_{\xi \,(\mathrm{mod}\, d)} S_r(x,\xi).$$

Corollary 26.7 then implies that

$$\begin{aligned}T_r(x,Q) &\ll \sum_{(\log x)^C<2^j\leqslant 2Q/r} \frac{1}{2^j}\Big(x + \frac{2^j x}{\sqrt{U}} + \frac{2^j x}{\sqrt{V}} + 4^j\sqrt{x}\Big)(\log x)^3 \\ &\ll \frac{x}{(\log x)^{C-3}} + \frac{x(\log x)^4}{\sqrt{U}} + \frac{x(\log x)^4}{\sqrt{V}} + \frac{Q\sqrt{x}(\log x)^3}{r}, \qquad (26.13)\end{aligned}$$

where we used the estimates

$$\sum_{2^j>(\log x)^C} 2^{-j} \ll (\log x)^{-C}, \qquad \sum_{1\leqslant 2^j\leqslant 2Q/r} 1 \ll \log x, \qquad \sum_{1\leqslant 2^j\leqslant 2Q/r} 2^j \ll Q/r.$$

Inserting (26.13) into (26.11) and executing the summation over r yields

$$\sum_{q\leqslant Q}\max_{\substack{\sqrt{x}\leqslant y\leqslant x\\ a\in(\mathbb{Z}/q\mathbb{Z})^*}} |\Delta_{\Lambda^\flat}(y;q,a)| \ll \frac{x}{(\log x)^{C-4}} + \frac{x(\log x)^5}{\sqrt{\min\{U,V\}}} + Q\sqrt{x}(\log x)^3.$$

We take $C = A+4$ and $U = V = e^{\sqrt{\log x}}$ (which satisfy (26.6), at least when x is large enough). This completes the proof of (26.10), and hence of the Bombieri-Vinogradov theorem.

Exercises

Exercise 26.1. Let $q \in \mathbb{N}$ and $a \in \mathbb{Z}$.

(a) If $(a,q) = 1$, prove that $\Delta_\tau(x;q,a) \ll \sqrt{x}$ for $x \geqslant 1$, as well as

$$\sum_{\substack{n\leqslant x\\ n\equiv a \,(\mathrm{mod}\, q)}} \tau(n) \sim \frac{\varphi(q)}{q^2}\cdot x\log x \qquad (x\to\infty).$$

(b) Let $d = (a,q)$, $r = q/d$ and $c = \tau(d)\prod_{p\mid q,\, p\nmid r}(1-1/[(p-1)(v_p(d)+1)])$, where $v_p(d)$ denotes the p-adic valuation of d. Prove that

$$\sum_{\substack{n\leqslant x\\ n\equiv a \,(\mathrm{mod}\, q)}} \tau(n) \sim c\cdot\frac{\varphi(q)}{q^2}\cdot x\log x \quad (x\to\infty).$$

[*Hint*: For each Dirichlet character $\chi \,(\mathrm{mod}\, q)$, evaluate the partial sums of $m \to \chi(m)\tau(dm)/\tau(d)$ by appealing to Theorem 13.2.]

Exercise 26.2. Fix $k \in \mathbb{N}$ and $A > 0$. Show there is $B = B(A,k)$ such that

$$\sum_{q\leqslant Q}\tau_k(q)\max_{y\leqslant x}\max_{a\in(\mathbb{Z}/q\mathbb{Z})^*}\left|\pi(y;q,a)-\frac{\mathrm{li}(y)}{\varphi(q)}\right| \ll_{k,A} \frac{x}{(\log x)^A}$$

for $x \geqslant 2$ and $1 \leqslant Q \leqslant \sqrt{x}/(\log x)^B$. [*Hint:* Use the Brun-Titchmarsch and the Cauchy-Schwarz inequalities to remove the weight $\tau_k(q)$.]

Exercise 26.3.* Let $\mathcal{Q}_{A,B}$ denote the set of integers $q \geqslant 3$ such that

$$\max_{a\in(\mathbb{Z}/q\mathbb{Z})^*}\left|\pi(x;q,a)-\frac{\mathrm{li}(x)}{\varphi(q)}\right| \leqslant \frac{\mathrm{li}(x)}{\varphi(q)(\log x)^A} \qquad \text{for all } x \geqslant q^2(\log q)^B.$$

For each $A > 0$, show that there is $B = B(A)$ such that

$$\#\mathcal{Q}_{A,B}\cap[1,Q] = Q + O_A(Q/(\log Q)^A).$$

Exercise 26.4.* Fix $A \geqslant 1$ and $\varepsilon > 0$. Using the combinatorial decomposition (23.14) in place of Vaughan's identity, show that (26.1) holds uniformly for $x \geqslant 2$ and $1 \leqslant Q \leqslant \sqrt{x}/(\log x)^{A+2+\varepsilon}$. [*Hint:* In the proof of the analogue of Theorem 26.7, with α_j, β_j defined appropriately, show that $\|\alpha_j\|_2\|\beta_j\|_2 \ll x\log x/\log y$.]

Exercise 26.5.* Prove the Bombieri-Vinogradov theorem by decomposing von Mangoldt's function using Heath-Brown's identity from Exercise 23.5.

Chapter 27

The least prime in an arithmetic progression

We have proved that all reduced arithmetic progressions of a given modulus q get their fair share of primes. Since this is an asymptotic result, a fundamental question is how far do we have to go to see the primes becoming equidistributed among the different progressions mod q. A simpler version of this question is how far do we have to go to locate the *first prime* p in the reduced residue class $a \pmod q$. We denote this prime by $P(q,a)$.

The Siegel-Walfisz theorem tells us that $\pi(x;q,a) \sim \mathrm{li}(x)/\varphi(q)$ as soon as $x \gg_\varepsilon \exp\{q^\varepsilon\}$, so that $P(q,a) \ll_\varepsilon \exp\{q^\varepsilon\}$. However, we expect that $\pi(x;q,a) \sim \mathrm{li}(x)/\varphi(q)$ in the much wider range $x \geqslant q^{1+\varepsilon}$ (see Exercise 17.6), which would imply that $P(q,a) \ll_\varepsilon q^{1+\varepsilon}$. If we knew that $\prod_{\chi \,(\mathrm{mod}\, q)} L(s,\chi)$ has no zeroes in the strip $\mathrm{Re}(s) > 1-\delta$, we would immediately deduce that $P(q,a) \ll_\varepsilon q^{1/\delta+\varepsilon}$ (for instance, see Exercise 11.2). Remarkably, Linnik proved that such a strong bound on $P(q,a)$ holds unconditionally.

Theorem 27.1 (Linnik)**.** *There is an absolute and effectively computable constant $L \geqslant 1$ such that $P(q,a) \leqslant q^L$ for all $q \geqslant 3$ and $a \in (\mathbb{Z}/q\mathbb{Z})^*$.*

Linnik's original proof relies on three ingredients:

1) the classical zero-free region given in Theorem 12.3;
2) a *log-free zero-density estimate* which, among others, implies that, for each fixed $C > 0$, the product $\prod_{\chi \,(\mathrm{mod}\, q)} L(s,\chi)$ has $O(1)$ zeroes in the region $\{\, s \in \mathbb{C} : \sigma \geqslant 1 - C/\log(qT),\ |t| \leqslant T \,\}$;
3) the *Deuring-Heilbronn phenomenon*, stating that the classical zero-free region can be enlarged when it contains an exceptional zero.

A proof along these lines is presented in [10] and [114]. We will present here an alternative proof. The three above ingredients are replaced by:

1′) the results of Chapter 22 that build on sieve methods and the theory of *pretentious multiplicative functions*;

2′) the *pretentious large sieve*, as developed by Granville, Harper and Soundararajan [66] building on ideas of Halász [84] and Elliott [40];

3′) an argument of Friedlander-Iwaniec [59, Chapter 24] that allows us to count primes when there is an exceptional zero using a *zero-dimensional sieve*.

The pretentious large sieve

Consider the bilinear form that has coefficients $a_{\chi,n} = \chi(n)\sqrt{q/\varphi(q)}$, where χ runs over all Dirichlet characters mod q and $n \in \mathbb{Z} \cap [1, N]$. In Example 25.8, we proved that this bilinear form has norm $\leqslant N + q$. Using Theorem 25.9, we deduce the bound

$$\sum_{\chi \,(\mathrm{mod}\, q)} \Big| \sum_{n \leqslant N} c_n \chi(n) \Big|^2 \leqslant \frac{\varphi(q)}{q}(N+q)\|\vec{c}\|_2^2 \tag{27.1}$$

for all $\vec{c} \in \mathbb{C}^N$. If $|c_n| \leqslant 1$ and $N \geqslant q$, the right-hand side is $\ll N^2\varphi(q)/q$. This bound could be as big as one term on the left-hand side if, say, $c_n = \overline{\chi}(n)$. However, we should expect that the sequence $(c_n)_{n=1}^N$ can correlate strongly with only a few Dirichlet characters by the approximate orthogonality of the latter. Hence, if $\chi_1, \ldots, \chi_r$ are the characters correlating the most with the sequence $(c_n)_{n \leqslant N}$, it is reasonable to guess that

$$\sum_{\chi \neq \{\chi_1, \ldots, \chi_r\}} \Big| \sum_{n \leqslant N} c_n \chi(n) \Big|^2 = o(N^2\varphi(q)/q).$$

The pretentious large sieve proves such an estimate when $c_n = f(n)$ with f multiplicative and bounded. Rather than presenting it in full generality, we develop it in a rather special case that sidesteps many of the technicalities of the general case (see Lemma 27.6(b) and Remark 27.7 below).

To simplify various details, we count the primes with a logarithmic weight. As we will see shortly, this move allows us to bypass the use of Perron's inversion formula and instead relate prime sums to L-functions using the simple idea behind the proof of Lemma 22.3.

By orthogonality, we have that

$$\sum_{\substack{y < p \leqslant z \\ p \equiv a \,(\mathrm{mod}\, q)}} \frac{1}{p} = \frac{1}{\varphi(q)} \sum_{\chi \,(\mathrm{mod}\, q)} \overline{\chi}(a) \sum_{y < p \leqslant z} \frac{\chi(p)}{p}.$$

We expect the principal contribution to come from the character $\chi = \chi_0$. However, in Chapters 12 and 22 we saw that there is potentially an additional exceptional character whose contribution we cannot control. If we exclude these two characters, we can show that the total contribution of the remaining characters is small.

Theorem 27.2. *Let $q \geqslant 3$. There is a real, non-principal Dirichlet character $\chi_1 \pmod q$ such that for all $z \geqslant y \geqslant q^3$ and all $a \in (\mathbb{Z}/q\mathbb{Z})^*$ we have*

$$\sum_{\substack{y<p\leqslant z\\ p\equiv a \,(\mathrm{mod}\, q)}} \frac{1}{p} = \frac{1}{\varphi(q)}\Big(\sum_{y<p\leqslant z} \frac{1+\chi_1(ap)}{p} + O(1)\Big).$$

Remark 27.3. In fact, if $\mathcal{R}_q$ denotes the set of real, non-principal Dirichlet characters mod q, we will take χ_1 such that $L_q(1,\chi_1) = \min_{\chi\in\mathcal{R}_q} L_q(1,\chi)$. This choice is motivated by Theorem 22.6(b). □

The first step in the proof of Theorem 27.2 is to exploit the type I/II structure of von Mangoldt's function and pass to a second moment estimate to which we can use the method of bilinear forms, much like we did in the proof of the Bombieri-Vinogradov theorem. Due to the presence of logarithmic weights, it suffices to use (23.14) with $D = 1$: we have

$$\Lambda(n)1_{P^-(n)>y} = \Lambda^\sharp_{\mathrm{sieve}}(n) + \Lambda^\flat_{\mathrm{sieve}}(n), \tag{27.2}$$

where

$$\Lambda^\sharp_{\mathrm{sieve}}(n) = 1_{P^-(n)>y}\log n \quad\text{and}\quad \Lambda^\flat_{\mathrm{sieve}}(n) = \sum_{\substack{k\ell=n,\, k>y,\, \ell>y\\ P^-(k\ell)>y}}\sum \mu(k)\log\ell.$$

If χ is a Dirichlet character and we let $L_y(s,\chi)$ be defined as in Chapter 22 (see (22.2)), then the Dirichlet series of $\Lambda^\flat_{\mathrm{sieve}}\chi$ factors as

$$\sum_{n=1}^\infty \frac{\Lambda^\flat_{\mathrm{sieve}}(n)\chi(n)}{n^s} = -L_y'(s,\chi)\big(L_y^{-1}(s,\chi)-1\big). \tag{27.3}$$

Using the above observations and Lemma 22.3, we prove the following preliminary result.

Lemma 27.4. *Let $x \geqslant y \geqslant q^3$. In addition, let χ_1 be any real non-principal character* mod q*, and set $\mathcal{C}_q = \{\chi \pmod q : \chi \neq \chi_0, \chi_1\}$ and*

$$S_q = -\int_{1+1/\log z}^{1+1/\log y} \sum_{\chi\in\mathcal{C}_q} \overline{\chi}(a)L_y'(\sigma,\chi)\big(L_y^{-1}(\sigma,\chi)-1\big)\mathrm{d}\sigma.$$

Then

$$\sum_{\substack{y<p\leqslant z\\ p\equiv a \,(\mathrm{mod}\, q)}} \frac{1}{p} = \frac{1}{\varphi(q)}\Big(\sum_{y<p\leqslant z} \frac{1+\chi_1(ap)}{p} + S_q + O(1)\Big).$$

Proof. We may assume that $q \geqslant 10$. Set

$$\delta(n) = 1_{n\equiv a \,(\mathrm{mod}\, q)} - \frac{1}{\varphi(q)} \sum_{\chi\in\{\chi_0,\chi_1\}} \overline{\chi}(a)\chi(n) = \frac{1}{\varphi(q)} \sum_{\chi\in\mathcal{C}_q} \overline{\chi}(a)\chi(n), \tag{27.4}$$

so that our goal is to estimate the sum

$$\sum_{\substack{y<p\leqslant z\\ p\equiv a\,(\mathrm{mod}\, q)}} \frac{1}{p} - \frac{1}{\varphi(q)} \sum_{y<p\leqslant z} \frac{1+\chi_1(ap)}{p} = \sum_{y<p\leqslant z} \frac{\delta(p)}{p}.$$

First of all, with the notational convention that $\Lambda(1)/\log(1) = 0$, we claim that

$$\sum_{\substack{y<p\leqslant w\\ p\equiv a\,(\mathrm{mod}\, q)}} \frac{1}{p} = \sum_{\substack{P^-(n)>y\\ n\equiv a\,(\mathrm{mod}\, q)}} \frac{\Lambda(n)/\log n}{n^{1+1/\log w}} + O(1/\varphi(q)) \qquad (w \geqslant y).$$

Indeed, this follows from a simple adaptation of the proof of Lemma 22.3. The two needed estimates are

$$\sum_{\substack{p>w\\ p\equiv a\,(\mathrm{mod}\, q)}} \frac{1}{p^{1+1/\log w}} \ll \frac{1}{\varphi(q)} \quad \text{and} \quad \sum_{\substack{y<p\leqslant w\\ p\equiv a\,(\mathrm{mod}\, q)}} \frac{\log p}{p} \ll \frac{\log w}{\varphi(q)}$$

for $w \geqslant y$, which are both corollaries of the Brun-Titchmarsch inequality (Theorem 20.1) and partial summation, since we have assumed that $y \geqslant q^3$.

In addition, for any character $\chi \,(\mathrm{mod}\, q)$, Lemma 22.3 implies that

$$\sum_{y<p\leqslant w} \frac{\chi(p)}{p} = \sum_{P^-(n)>y} \frac{\chi(n)\Lambda(n)/\log n}{n^{1+1/\log w}} + O(1) \qquad (w \geqslant y).$$

We thus find that

$$\sum_{y<p\leqslant w} \frac{\delta(p)}{p} = \sum_{P^-(n)>y} \frac{\delta(n)\Lambda(n)}{n^{1+1/\log w}\log n} + O(1/\varphi(q)) \qquad (w \geqslant y). \tag{27.5}$$

Next, we rewrite $\Lambda(n)1_{P^-(n)>y}$ using (27.2) and show that the contribution of $\Lambda^\sharp_{\mathrm{sieve}}$ to the right side of (27.5) is negligible. Indeed, Theorem 18.11(a) and our assumption that $y \geqslant q^3$ imply that

$$\sum_{\substack{n\leqslant x,\, P^-(n)>y\\ n\equiv b\,(\mathrm{mod}\, q)}} 1 = \frac{x\prod_{p\leqslant y}(1-1/p)}{\varphi(q)} + O\left(\frac{x^{1-1/\log y}}{\varphi(q)\log y}\right) \quad (x \geqslant y,\ b \in (\mathbb{Z}/q\mathbb{Z})^*).$$

Note that δ is a q-periodic function supported on integers coprime to q. In addition, $\sum_{n\in\mathbb{Z}/q\mathbb{Z}} \delta(n) = 0$ and $\sum_{n\in\mathbb{Z}/q\mathbb{Z}} |\delta(n)| \leqslant 3$. As a consequence,

$$\sum_{n\leqslant x} \frac{\delta(n)\Lambda^\sharp_{\mathrm{sieve}}(n)}{\log n} = \sum_{1<n\leqslant x} \delta(n)1_{P^-(n)>y} \ll \frac{x^{1-1/\log y}}{\varphi(q)\log y}$$

for $x \geqslant y$. Together with partial summation, this implies that

$$\sum_{n=1}^{\infty} \frac{\delta(n)\Lambda_{\mathrm{sieve}}^{\sharp}(n)}{n^{\sigma}\log n} \ll \frac{1}{\varphi(q)} \qquad (\sigma \geqslant 1).$$

Combining the above estimate with (27.5) and (27.2), we conclude that

$$\sum_{y<p\leqslant w} \frac{\delta(p)}{p} = \sum_{n=1}^{\infty} \frac{\delta(n)\Lambda_{\mathrm{sieve}}^{\flat}(n)}{n^{1+1/\log w}\log n} + O(1/\varphi(q)) \qquad (w \geqslant y). \tag{27.6}$$

We want to express the right-hand side of (27.6) as an integral. We do this using a trick: we may trivially arrange the summation as $\sum_{y<p\leqslant z} = \sum_{y<p\leqslant z} - \sum_{y<p\leqslant y}$. Hence, applying (27.6) with $w \in \{y, z\}$ implies that

$$\begin{aligned}\sum_{y<p\leqslant z} \frac{\delta(p)}{p} &= \sum_{n=1}^{\infty} \frac{\delta(n)\Lambda_{\mathrm{sieve}}^{\flat}(n)}{n^{1+1/\log z}\log n} - \sum_{P^-(n)>y} \frac{\delta(n)\Lambda_{\mathrm{sieve}}^{\flat}(n)}{n^{1+1/\log y}\log n} + O(1/\varphi(q)) \\ &= -\int_{1+1/\log z}^{1+1/\log y} \sum_{n=1}^{\infty} \frac{\Lambda_{\mathrm{sieve}}^{\flat}(n)\delta(n)}{n^{\sigma}} \mathrm{d}\sigma + O(1/\varphi(q)),\end{aligned}$$

by the Fundamental Theorem of Integral Calculus. Using (27.4) to rewrite $\delta(n)$ in the integrand in terms of characters $\chi \in \mathcal{C}_q$, and then employing (27.3) to factor the Dirichlet series of $\Lambda_{\mathrm{sieve}}^{\flat}\chi$ completes the proof. □

The next natural step is to apply the Cauchy-Schwarz inequality to the sum S_q of Lemma 27.4 and use bounds like (27.1). But first, we exploit the fact that we are summing over the restricted set of characters $\mathcal{C}_q$.

Lemma 27.5. *Let* $\mathcal{C}_q = \{\chi \,(\mathrm{mod}\, q) : \chi \neq \chi_0, \chi_1\}$ *with* χ_1 *defined as in Remark* 27.3. *Then* $|L_y(\sigma,\chi)| \asymp 1$ *for all* $\chi \in \mathcal{C}_q$, $y \geqslant q$ *and* $\sigma \geqslant 1$.

Proof. If χ is a complex character, then $|L_q(1,\chi)| \asymp 1$ by Theorem 22.6(a). On the other hand, if $\chi \in \mathcal{C}_q$ is real, then $L_q(1,\chi) \geqslant L_q(1,\chi_1)$ by the choice of χ_1, and thus $L_q(1,\chi) \asymp 1$ by Theorem 22.6(b). In all cases, we have $|L_q(1,\chi)| \asymp 1$. Combining this fact with Theorem 22.5 (applied with $y = q$ and $t = 0$) yields that $\sum_{u<p\leqslant v} \chi(p)/p = O(1)$ for all $v \geqslant u \geqslant q$. Finally, if we insert this estimate into Lemma 22.3, we infer that $|L_y(\sigma,\chi)| \asymp 1$ for all $y \geqslant q$ and all $\sigma \geqslant 1$, as needed. □

By the above lemma, we have $L_y^{-1}(\sigma,\chi) - 1 \ll 1$ for all $\chi \in \mathcal{C}_q$. However, we do not want to remove the factor $L_y^{-1}(\sigma,\chi) - 1$ completely from the sum S_q of Lemma 27.4 because this will destroy its bilinear structure. Instead, we note that $L_y^{-1}(s,\chi) = F_y(s,\chi)^2$, where

$$F_y(s,\chi) = \prod_{p>y} \left(1 - \frac{\chi(p)}{p^s}\right)^{1/2} = \sum_{P^-(n)>y} \frac{\tau_{-1/2}(n)\chi(n)}{n^s}$$

with τ_κ being defined by (13.3). Since $|z^2-1| \ll |z-1|$ for $|z| \ll 1$, we have

$$S_q \ll \int_{1+1/\log z}^{1+1/\log y} \sum_{\chi \in \mathcal{C}_q} |L_y'(\sigma,\chi)| \cdot |F_y(\sigma,\chi)-1| \mathrm{d}\sigma$$
$$\leqslant \int_{1+1/\log z}^{1+1/\log y} \Big(\sum_{\chi \in \mathcal{C}_q} |L_y'(\sigma,\chi)|^2 \sum_{\chi \in \mathcal{C}_q} |F_y(\sigma,\chi)-1|^2 \Big)^{1/2} \mathrm{d}\sigma.$$

Noticing that $|\tau_{-1/2}| \leqslant \tau_{1/2}$, Theorem 27.2 follows from the following estimate.

Lemma 27.6. *Let $\kappa > 0$, $y \geqslant q^3 > 1$ and $1 < \sigma \leqslant 1 + 1/\log y$.*

(a) *Let f be an arithmetic function with $|f| \leqslant \tau_\kappa$. We have*

$$\sum_{\chi \,(\mathrm{mod}\, q)} \Big| \sum_{n>1} \frac{1_{P^-(n)>y} f(n)\chi(n)}{n^\sigma} \Big|^2 \ll_\kappa \frac{1}{[(\sigma-1)(\log y)]^{2\kappa}}.$$

(b) *In addition, we have*

$$\sum_{\substack{\chi \,(\mathrm{mod}\, q) \\ \chi \neq \chi_0}} |L_y'(\sigma,\chi)|^2 \ll (\log y)^2.$$

Remark 27.7. Part (b) exemplifies the idea of the pretentious large sieve: we exclude the principal character from the summation because we know that $L_y'(\sigma,\chi_0) \to \infty$ when $\sigma \to 1^+$, whereas we know that we can control the size of $L_y'(\sigma,\chi)$ for $\chi \neq \chi_0$ using ideas from Chapter 22. □

Proof. (a) Let $f_y(n) = 1_{P^-(n)>y} f(n)$. By the orthogonality of Dirichlet characters, we have

$$\sum_{\chi \,(\mathrm{mod}\, q)} \Big| \sum_{n>1} \frac{f_y(n)\chi(n)}{n^\sigma} \Big|^2 = \varphi(q) \sum_{\substack{n_1>1 \\ (n_1,q)=1}} \frac{f_y(n_1)}{n_1^\sigma} \sum_{\substack{n_2>1 \\ n_2 \equiv n_1 \,(\mathrm{mod}\, q)}} \frac{\bar{f}_y(n_2)}{n_2^\sigma}. \tag{27.7}$$

Since f_y is supported on integers free of primes $\leqslant y$, we may assume that the above sum runs over integers $n_1, n_2 > y$.

For $a \in (\mathbb{Z}/q\mathbb{Z})^*$ and $x \geqslant y \geqslant q^3$, Theorem 20.3 and our assumption that $|f| \leqslant \tau_\kappa$ with $\kappa > 0$ yield the bound

$$\sum_{\substack{n \leqslant x \\ n \equiv a \,(\mathrm{mod}\, q)}} |f_y(n)| \ll \frac{x}{q} \exp\Big\{ \sum_{y<p\leqslant x} \frac{\kappa}{p} - \sum_{\substack{p \leqslant x \\ p \nmid q}} \frac{1}{p} \Big\} \ll \frac{x}{\varphi(q)(\log x)^{1-\kappa}(\log y)^\kappa}.$$

Using this estimate and partial summation, we find that

$$\sum_{\substack{n_2>1 \\ n_2 \equiv a \,(\mathrm{mod}\, q)}} \frac{|f_y(n_2)|}{n_2^\sigma} \ll \frac{1}{\varphi(q)[(\sigma-1)\log y]^\kappa}$$

uniformly for $a \in (\mathbb{Z}/q\mathbb{Z})^*$ and $\sigma \in (1, 1 + 1/\log y]$. Together with (27.7), this completes the proof of part (a).

(b) To prove the second part of the lemma, we combine the proof of part (a) and of Theorem 22.1. We begin by splitting $L'_y(\sigma, \chi)$ as

$$L'_y(\sigma,\chi) = -\sum_{j\geqslant 1} \sum_{\substack{y^j<n\leqslant y^{j+1}\\ P^-(n)>y}} \frac{\chi(n)\log n}{n^\sigma}.$$

Fix j for the moment and let $\lambda_j^\pm$ be as in Theorem 19.1 with $D = y^{j/3}$ and $\mathcal{P} = \{p \leqslant y\}$. Moreover, set

$$\delta_j(n) = (\lambda_j^+ * 1)(n) - 1_{P^-(n)>y}, \quad \text{so that} \quad 0 \leqslant \delta_j \leqslant (\lambda_j^+ - \lambda_j^-) * 1.$$

Arguing as in the proof of (22.8), we find that

$$\begin{aligned}\sum_{y^j<n\leqslant y^{j+1}} &\frac{\chi(n)(\lambda_j^+ * 1)(n)\log n}{n^\sigma}\\ &= \sum_{d\leqslant y^{j/3}} \frac{\lambda_j^+(d)\chi(d)}{d^\sigma} \sum_{y^j/d<m\leqslant y^{j+1}/d} \frac{\chi(m)\log(dm)}{m^\sigma}\\ &\ll \sum_{d\leqslant y^{j/3}} \frac{1}{d^\sigma}\cdot\frac{q\log(y^j)}{(y^j/d)^\sigma} \leqslant \frac{jq\log y}{y^{2j/3}} \leqslant \frac{j\log q}{q^{2j-1}}\end{aligned}$$

for $y \geqslant q^3$ and $\chi \neq \chi_0$. Consequently,

$$-L'_y(\sigma,\chi) = O(1/\sqrt{q}) + \sum_{j\geqslant 1} \sum_{y^j<n\leqslant y^{j+1}} \frac{\chi(n)\delta_j(n)\log n}{n^\sigma}.$$

We then apply the Cauchy-Schwarz inequality twice to find that

$$\begin{aligned}|L'_y(\sigma,\chi)|^2 &\leqslant O(1/q) + 2\Big|\sum_{j\geqslant 1}\frac{1}{j}\cdot j \sum_{y^j<n\leqslant y^{j+1}} \frac{\chi(n)\delta_j(n)\log n}{n^\sigma}\Big|^2\\ &\ll \frac{1}{q} + \sum_{j\geqslant 1} j^2\Big|\sum_{y^j<n\leqslant y^{j+1}} \frac{\chi(n)\delta_j(n)\log n}{n^\sigma}\Big|^2. \qquad (27.8)\end{aligned}$$

Summing over all non-principal characters $\chi \,(\mathrm{mod}\, q)$, we infer the bound

$$\begin{aligned}\sum_{\substack{\chi\,(\mathrm{mod}\,q)\\ \chi\neq\chi_0}} |L'_y(\sigma,\chi)|^2 &\ll 1 + \sum_{j=1}^\infty j^2 \sum_{\chi\,(\mathrm{mod}\,q)} \Big|\sum_{y^j<n\leqslant y^{j+1}} \frac{\chi(n)\delta_j(n)\log n}{n^\sigma}\Big|^2\\ &=: 1 + \sum_{j=1}^\infty j^2 S_j. \qquad (27.9)\end{aligned}$$

Using the orthogonality of Dirichlet characters as in (27.7) and the inequality $0 \leqslant \delta_j(n) \log n \ll \delta_j(n) \log(y^j)$ for $n \leqslant y^{j+1}$, we deduce that

$$S_j \ll \varphi(q) j^2 (\log y)^2 \sum_{\substack{y^j < n_1 \leqslant y^{j+1} \\ (n_1,q)=1}} \frac{\delta_j(n_1)}{n_1^\sigma} \sum_{\substack{y^j < n_2 \leqslant y^{j+1} \\ n_2 \equiv n_1 \,(\mathrm{mod}\, q)}} \frac{\delta_j(n_2)}{n_2^\sigma}.$$

Now, since $0 \leqslant \delta_j \leqslant \lambda_j^+ - \lambda_j^-$, Theorem 19.1 implies that

$$0 \leqslant \sum_{\substack{n \leqslant x \\ n \equiv a \,(\mathrm{mod}\, q)}} \delta_j(n) \leqslant \sum_{\substack{d \leqslant y^{j/3} \\ (d,q)=1}} (\lambda_j^+ - \lambda_j^-)(d) \cdot \Big(\frac{x}{dq} + O(1) \Big) \ll \frac{x e^{-j}}{\varphi(q) \log y}$$

for $x \geqslant y^j$ and $a \in (\mathbb{Z}/q\mathbb{Z})^*$, where we used our assumption that $y \geqslant q^3$. Together with partial summation, this yields the bound $S_j \ll j^4 e^{-2j} (\log y)^2$. Inserting this estimate into (27.9) completes the proof of the lemma. □

Endgame

Having shown Theorem 27.2, we now pass to the proof of Linnik's theorem. Let us recall that we must prove that $P(q,a) \leqslant q^L$ if L is large enough. We may assume that $q \geqslant 10$. Throughout, χ_1 is as in Theorem 27.2.

The two following cases are easy to handle.

Case 1: $\chi_1(a) = -1$. Then, Theorem 27.2 and Corollary 22.4 imply that

$$\sum_{\substack{y < p \leqslant z \\ p \equiv a \,(\mathrm{mod}\, q)}} \frac{1}{p} = \frac{1}{\varphi(q)} \Big(\sum_{y < p \leqslant z} \frac{1 - \chi_1(p)}{p} + O(1) \Big) \geqslant \frac{1}{\varphi(q)} \Big(\sum_{y < p \leqslant z} \frac{1}{p} - O(1) \Big)$$

for all $z \geqslant y \geqslant q^3$. We take $z = q^L$ and $y = q^3$ to find that the right-hand side of the above inequality is $\geqslant (\log L - O(1))/\varphi(q)$. Hence, if L is large enough, we immediately deduce that $P(q,a) \leqslant z = q^L$.

Case 2: $\chi_1(a) = 1$ and $L_q(1,\chi_1) \geqslant L^{-0.99}$. In this case, we let $y = q^{L^{0.99}}$ and $z = q^L$. Theorem 22.5 then implies that $\sum_{y < p \leqslant z} \chi_1(p)/p = O(1)$. Together with Theorem 27.2, this yields the estimate

$$\sum_{\substack{y < p \leqslant z \\ p \equiv a \,(\mathrm{mod}\, q)}} \frac{1}{p} = \frac{1}{\varphi(q)} \Big(\sum_{y < p \leqslant z} \frac{1}{p} + O(1) \Big) = \frac{\log L + O(1)}{100\, \varphi(q)}.$$

If L is large enough, the above expression is positive. Hence, $P(a,q) \leqslant z = q^L$, as needed.

The last case remaining is thus:

Case 3: $\chi_1(a) = 1$ and $L_q(1,\chi_1) \leqslant L^{-0.99}$. Under the second assumption, Exercise 22.3 implies that $L(s,\chi_1)$ has a zero $\beta_1 \geqslant 1 - O(L^{-0.99}/\log q)$. We

do not need this fact, but it is useful to keep it in mind. We will use sieve methods to detect primes in the arithmetic progression $a \,(\mathrm{mod}\, q)$. The key observation is that if n is square-free, then $(1 * \chi_1)(n) = 0$, unless n has no prime factors p with $\chi_1(p) = -1$. Now, (22.15) implies that $\chi_1(p) = -1$ for the vast majority of primes $p \in [q, q^{1/L_q(1,\chi_1)}]$. This means that weighing n with the function $(1 * \chi_1)(n)$ presieves it with most of the large primes. Hence, if we let

$$S(x, y; q, a) = \sum_{\substack{1<n\leqslant x,\, P^-(n)>y \\ n\equiv a \,(\mathrm{mod}\, q)}} (1 * \chi_1)(n)$$

and we choose x and y appropriately in terms of q, we expect that

$$\sum_{\substack{\sqrt{x}<p\leqslant x \\ p\equiv a \,(\mathrm{mod}\, q)}} (1 + \chi_1(p)) = S(x, \sqrt{x}; q, a) \approx S(x, y; q, a) \tag{27.10}$$

with y much smaller than x. On the other hand, the Fundamental Lemma of Sieve Theory is very efficient at estimating $S(x, y; q, a)$ for small y, as the following result demonstrates.

Proposition 27.8. *Assume the above notation. In addition, let $q \geqslant 10$, $b \in (\mathbb{Z}/q\mathbb{Z})^*$, $x \geqslant q^{100}$, $y \in [q, x]$ and $u = \log x/\log y$. Then*

$$S(x, y; q, b) = \big(1 + \chi_1(b) + O(x^{-1/\log y})\big)\frac{xL_y(1, \chi_1)}{\varphi(q)} \prod_{p\leqslant y}\left(1 - \frac{1}{p}\right).$$

Proof. Note that $S(x, y; q, b) = S(\mathcal{A}, \mathcal{P})$, where $\mathcal{A} = (a_n)_{n=1}^{\infty}$ with

$$a_n = (1 * \chi_1)(n) \cdot 1_{n\leqslant x,\, n\equiv b \,(\mathrm{mod}\, q)}$$

and $\mathcal{P} = \{\, p \leqslant y : p \nmid q \,\}$. We will apply Theorem 18.11. We must first check Axioms 1–3.

Fix, for the moment, $d|\mathcal{P}$ such that $d \leqslant \sqrt{x}$. We have

$$A_d = \sum_{\substack{n\leqslant x,\, d|n \\ n\equiv b \,(\mathrm{mod}\, q)}} (1 * \chi_1)(n) = \sum_{\substack{k\ell\leqslant x,\, d|k\ell \\ k\ell\equiv b \,(\mathrm{mod}\, q)}} \chi_1(k).$$

Notice that $\chi_1(\ell) = \chi_1(b)\chi_1(k)$ in the above sum. We then split A_d into three subsums: in the first one $k < \ell$, in the second one $\ell < k$, and in the third one $k = \ell$. Since $\chi_1(\ell) = \chi_1(b)\chi_1(k)$, the second subsum equals $\chi_1(b)$ times the first subsum. Moreover, the third subsum is $O(\sqrt{x})$, since it has $\leqslant \sqrt{x}$ terms all of magnitude $\leqslant 1$. Hence,

$$A_d = (1 + \chi_1(b)) \sum_{k<\sqrt{x}} \chi_1(k) \sum_{\substack{k<\ell\leqslant x/k,\, \ell\equiv \bar{k}b \,(\mathrm{mod}\, q) \\ \ell\equiv 0 \,(\mathrm{mod}\, d/(d,k))}} 1 + O(\sqrt{x}),$$

where $\bar{k}$ denotes the multiplicative inverse of $k \pmod q$. Since $d|\mathcal{P}$, we have $(d,q)=1$. Hence, the sum over ℓ equals $(x/k-k)/(dq/(d,k))+O(1)$. We thus infer that

$$A_d = (1+\chi_1(b)) \sum_{k<\sqrt{x}} \chi_1(k) \cdot \frac{x/k-k}{dq} \cdot (d,k) + O(\sqrt{x}).$$

Using the identity $(d,k)=\sum_{m|d,k}\varphi(m)$, and letting $k=mr$, we find that

$$A_d = \frac{1+\chi_1(b)}{dq} \sum_{m|d} \varphi(m)\chi_1(m) \sum_{r<\sqrt{x}/m} \chi_1(r)\Big(\frac{x}{mr}-mr\Big) + O(\sqrt{x}).$$

The Pólya-Vinogradov inequality (Theorem 10.6) and partial summation yield that

$$\sum_{r<\sqrt{x}/m} \chi_1(r)mr \ll \sqrt{xq}\log q \qquad \text{and} \qquad \sum_{r>\sqrt{x}/m} \chi_1(r)\cdot\frac{x}{mr} \ll \sqrt{xq}\log q.$$

Consequently,

$$A_d = \frac{(1+\chi_1(b))xL(1,\chi_1)}{dq} \sum_{m|d} \frac{\varphi(m)\chi_1(m)}{m} + O(\sqrt{xq}\log q).$$

We thus conclude that the pair $(\mathcal{A},\mathcal{P})$ satisfies Axiom 1 with $X=(1+\chi_1(b))xL(1,\chi_1)/q$, $\nu(d)=\sum_{m|d}\varphi(m)\chi_1(m)/m$ and $r_d=O(\sqrt{xq}\log q)$. Axiom 2 also holds with $\kappa=2$. Finally, since $xL(1,\chi_1)/q \gg x/(q^{3/2}\log^2 q)$ by Theorem 12.8 and we have assumed that $x \geqslant q^{100}$ and $y \geqslant 10$, we have $\sum_{d|\mathcal{P},\, d\leqslant x^{1/100}} |r_d| \ll x^{1-1/\log y}L(1,\chi_1)/(q\log^2 y)$. Noticing that $1-\nu(p)/p = (1-\chi_1(p)/p)(1-1/p)$ for $p \nmid q$, the lemma follows from Theorem 18.11(a). □

We also need a stronger version of the first part of Theorem 22.5.

Lemma 27.9. *Let $\chi \pmod q$ be a real, non-principal character. Then*

$$\sum_{y<p\leqslant z} \frac{1+\chi(p)}{p} \ll \frac{\log z}{\log Q} + y^{-1/(100\log q)} \quad \textit{for} \quad q \leqslant y \leqslant z \leqslant Q := q^{1/L_q(1,\chi)}.$$

Proof. We may assume that $q \geqslant 10$, as well as that $y \geqslant q_1 := q^{100}$, since $\sum_{q<p\leqslant q_1} 1/p = O(1)$. The non-negativity of $1*\chi$ implies that

$$\sum_{y<p\leqslant z} \frac{1+\chi(p)}{p} \leqslant \sum_{\substack{y<n\leqslant z \\ P^-(n)>q_1}} \frac{(1*\chi)(n)}{n}. \tag{27.11}$$

If we let $\alpha = L_{q_1}(1,\chi)\prod_{p\leqslant q_1}(1-1/p) \asymp 1/\log Q$, we have

$$\sum_{n\leqslant x,\, P^-(n)>q_1} (1*\chi)(n) = \alpha x + O\Big(\frac{x^{1-1/\log q_1}}{\log q}\Big) \tag{27.12}$$

for $x \in [y, z]$. Indeed, this follows by breaking the summation according to the congruence class of $n \,(\mathrm{mod}\, q)$ and by applying Proposition 27.8 to each subsum with χ in place of χ_1, while noticing that $L_{q_1}(1, \chi) \ll 1$ from Theorem 22.1. Inserting (27.12) into (27.11) via partial summation completes the proof. □

Proof of Theorem 27.1 in Case 3. Our starting point is the first equality in (27.10). We take $x = q^{L^{0.49}}$ and $y = q^{50 \log L}$. Since $1/L_q(1, \chi_1) \geqslant L^{0.99}$, Lemma 27.9 implies that

$$\sum_{y < p \leqslant x} \frac{1 + \chi_1(p)}{p} \ll \frac{1}{\sqrt{L}}, \tag{27.13}$$

that is to say, $\chi_1(p) = -1$ for most $p \in (y, x]$. We then use a variation of Buchstab's identity (19.12) to find that

$$\begin{aligned} S(x, \sqrt{x}; q, a) &= S(x, y; q, a) - \sum_{y < p \leqslant \sqrt{x}} \sum_{\substack{n \leqslant x,\, P^-(n) = p \\ n \equiv a \,(\mathrm{mod}\, q)}} (1 * \chi_1)(n) \\ &= S(x, y; q, a) - \sum_{m \geqslant 1,\, y < p \leqslant \sqrt{x}} \sum (1 * \chi_1)(p^m) S(x/p^m, p; q, \bar{p}^m a), \end{aligned}$$

where $\bar{b}$ denotes the inverse of $b \,(\mathrm{mod}\, q)$. When $m \geqslant 2$, we use the trivial bound $S(x/p^m, p; q, \bar{p}^j a) \leqslant \sum_{n \leqslant x/p^m} \tau(n) \ll x(\log x)/p^m$. For the summands with $m = 1$, we note that $(1 * \chi_1)(p) = 2 \cdot 1_{\chi_1(p) = 1}$. Consequently,

$$S(x, \sqrt{x}; q, a) = S(x, y; q, a) - 2 \sum_{\substack{y < p \leqslant \sqrt{x} \\ \chi_1(p) = 1}} S(x/p, p; q, \bar{p} a) + O\Big(\frac{x \log x}{y}\Big).$$

Assuming that L is large enough, we may apply Proposition 27.8 to all terms of the right side. This yields the estimate

$$\begin{aligned} S(x, \sqrt{x}; q, a) = \Bigg(2 - 4 \sum_{\substack{y < p \leqslant \sqrt{x} \\ \chi_1(p) = 1}} \frac{1}{p} + O(1/L)\Bigg) \frac{x L_y(1, \chi_1)}{\varphi(q)} \prod_{p \leqslant y} \Big(1 - \frac{1}{p}\Big) \\ + O(x(\log x)/y). \end{aligned}$$

Employing (27.13) and the lower bound $L(1, \chi) \gg 1/(\sqrt{q} \log^2 q)$ from Theorem 12.8, we conclude that

$$S(x, \sqrt{x}; q, a) = \big\{2 + O\big(L^{-1/2} + x^{-1/2}\big)\big\} \frac{x L_y(1, \chi_1)}{\varphi(q)} \prod_{p \leqslant y} \Big(1 - \frac{1}{p}\Big).$$

If we take L and x large enough, we have $S(x, \sqrt{x}; q, a) > 0$. This completes the proof of Theorem 27.1 in Case 3 too. □

Exercises

Exercise 27.1.* Assume the set-up of Lemma 27.6(a) and fix $r \in \mathbb{R}_{\geqslant 0}$.

(a) For all $T \in \mathbb{R}$, prove that

$$\sum_{\chi \,(\mathrm{mod}\, q)} \int_T^{T+1} \Big| \sum_{\substack{n>1 \\ P^-(n)>y}} \frac{f(n)(\log n)^r \chi(n)}{n^{\sigma+it}} \Big|^2 \mathrm{d}t \ll_{\kappa,r} \frac{(\sigma-1)^{1-2r}}{[(\sigma-1)(\log y)]^{2\kappa}}.$$

[*Hint:* First, reduce to the case $T = -1/2$. Then, for any smooth function $g : \mathbb{R} \to \mathbb{R}_{\geqslant 0}$ majorizing $1_{[-1/2,1/2]}$, show that the above sum is

$$\leqslant \varphi(q) \sum_{\substack{n_1,n_2>1,\, P^-(n_1n_2)>y \\ n_1 \equiv n_2 \,(\mathrm{mod}\, q)}} \frac{\tau_\kappa(n_1)\tau_\kappa(n_2)(\log n_1)^r(\log n_2)^r |\widehat{g}(\log(n_1/n_2))|}{(n_1n_2)^\sigma}.$$

To estimate this new sum, show that $\widehat{g}(\xi) \ll 1/(1+|\xi|^{r+\kappa+2})$, and then split the range of n_1, n_2 into intervals of the form $(e^j, e^{j+1}]$.]

(b) Deduce that

$$\sum_{\chi \,(\mathrm{mod}\, q)} \int_{-\infty}^{\infty} \Big| \sum_{\substack{n>1 \\ P^-(n)>y}} \frac{f(n)\chi(n)(\log n)^r}{n^{\sigma+it}} \Big|^2 \frac{\mathrm{d}t}{1+t^2} \ll_{\kappa,r} \frac{(\sigma-1)^{1-2r}}{[(\sigma-1)(\log y)]^{2\kappa}}.$$

Exercise 27.2.* For each $q \geqslant 3$, show that there is a real, non-principal Dirichlet character $\chi_1 \,(\mathrm{mod}\, q)$ such that for any fixed smooth and compactly supported function $g : \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ and any fixed $\varepsilon > 0$ we have

$$\sum_{n \equiv a \,(\mathrm{mod}\, q)} \Lambda(n) g(n/x) = \frac{x}{\varphi(q)} \int_0^\infty g(t)\mathrm{d}t + \frac{\chi_1(a)}{\varphi(q)} \sum_{n \geqslant 1} \chi(n)\Lambda(n) g(n/x) + O_{g,\varepsilon}(xu^{-1+\varepsilon}/\varphi(q))$$

uniformly for $x = q^u$ with $u \geqslant 1$. [*Hint:* The case $u \leqslant 100$ is trivial. For the case when $u \geqslant 100$, decompose Λ using (23.14) with $D = 1$ and $y = \max\{q^3, (\log x)^{100}\}$. To control the contribution of $\Lambda^\sharp_{\mathrm{sieve}}$, use the Fundamental Lemma of Sieve Theory. To control the contribution of $\Lambda^\flat_{\mathrm{sieve}}$, use Mellin inversion to find that

$$\sum_{\chi \in \mathcal{C}_q} \Big| \sum_{n \geqslant 1} \Lambda^\flat_{\mathrm{sieve}}(n) g(n/x) \Big| \ll_g x \sum_{\chi \in \mathcal{C}_q} \int_{\substack{\sigma=\alpha \\ t \in \mathbb{R}}} \frac{|(1 - L_y^{-1}(s,\chi))L_y'(s,\chi)|}{1+t^{100}} \mathrm{d}t,$$

where $\mathcal{C}_q = \{\, \chi \,(\mathrm{mod}\, q) : \chi \neq \chi_0, \chi_1 \,\}$ and $\alpha = 1 + 1/\log x$. After applying Cauchy-Schwarz, the integrals must be split into two ranges: when $|t| \leqslant y$, the bound $L_y^{-1}(s,\chi) \ll 1$, Exercise 27.1 and a suitable adaptation of Lemma 27.6(b) can be used. When $|t| > y \geqslant (\log x)^{100}$, Exercise 27.1 suffices.]

Part 6

Local aspects of the distribution of primes

Chapter 28

Small gaps between primes

The Prime Number Theorem establishes important global aspects of the distribution of primes but it does not reveal much about their statistical properties at a microscopic scale. We dedicate this last part of the book to the study of the local behavior of the sequence of primes. Firstly, we study how close successive members of this sequence can get.

There are $\sim x/\log x$ primes $\leqslant x$, so the average gap between them is $\sim \log x$. On the other hand, the twin prime conjecture predicts that the gap equals 2 infinitely often. Given that this conjecture is out of reach, we set a more modest goal: if $p_1 < p_2 < p_3 < \cdots$ are the primes in increasing order, we want to show that $\liminf_{n\to\infty}(p_{n+1} - p_n) < \infty$. Note that if this is true, then we immediately deduce the existence of some $s \in \mathbb{N}$ such that there are infinitely many primes p with $p + 2s$ also being prime. Hence, there is at least one even number satisfying Polignac's conjecture.

Remarkably, an even stronger result can be proved.

Theorem 28.1. *For each $m \in \mathbb{N}$, we have*

$$\liminf_{n\to\infty}(p_{n+m} - p_n) \ll e^{4m}m^5.$$

The case $m = 1$ of Theorem 28.1 is due to Zhang [188], whereas the case $m > 1$ was proven independently by Maynard [138] and Tao [171]. Granville's article [65, Section 1.4] contains an extended account of the fascinating developments that led to this major breakthrough.

The main goal of this chapter is to give a proof of Theorem 28.1.

The GPY sieve

The basic strategy to detect small gaps between primes is due to Goldston, Pintz and Yıldırım. They used their method, now called the *GPY sieve* after them, to prove that the normalized gap $(p_{n+1} - p_n)/\log n$ becomes arbitrarily small infinitely often. The main idea is to find weights $w_n \geqslant 0$ such that

$$\sum_{N \leqslant n \leqslant 2N} w_n \Big(\sum_{1 \leqslant s \leqslant H} 1_P(n+s) - m \Big) > 0 \tag{28.1}$$

for H that is as small as possible, where 1_P denotes the indicator function of the set of primes as usually. Indeed, if this is the case, then there must exist some $n \in [N, 2N]$ for which at least $m+1$ of the "shifts" $n+1, \ldots, n+H$ are primes.

To conceptualize the above task, it is helpful to assume a more probabilistic point of view. The weights w_n naturally induce a probability measure on $\mathbb{Z} \cap [N, 2N]$ via the relation

$$\mathbb{P}_{[N,2N]}(m) := \frac{w_m}{\sum_{N \leqslant n \leqslant 2N} w_n}.$$

In this notation, (28.1) becomes

$$\mathbb{E}_{N \leqslant n \leqslant 2N} \Big[\sum_{1 \leqslant s \leqslant H} 1_P(n+s) \Big] > m. \tag{28.2}$$

Hence, our goal is to find a probability measure on $\mathbb{Z} \cap [N, 2N]$ that is sufficiently concentrated on integers n for which many of the shifts $n+1$, $n+2, \ldots, n+H$ are primes.

As we saw in the discussion of Cramér's model in the end of Chapter 17, the numbers $n+1, n+2, \ldots, n+H$ have strong multiplicative dependencies stemming from their reduction modulo small primes. To this end, we consider integers $1 \leqslant s_1 < s_2 < \cdots < s_k \leqslant H$ forming an admissible[1] k-tuple $(s_1, \ldots, s_k)$ and aim to show that

$$\mathbb{E}_{N \leqslant n \leqslant 2N} \Big[\sum_{1 \leqslant j \leqslant k} 1_P(n+s_j) \Big] > m. \tag{28.3}$$

The weights w_n must be chosen in a way that achieves simultaneously two things: (i) they correlate strongly enough with the indicator function of the event that many of the shifts $n+s_1, \ldots, n+s_k$ are prime; (ii) they allow the estimation of the left-hand side of (28.3) unconditionally. Condition (i) rules out choices such as $w_n = 1$, and condition (ii) rules out choices such as $w_n = \prod_{j=1}^{k} 1_P(n+s_j)$. Instead, we use sieve theory to "interpolate" between these two extremal examples.

[1]Recall that this means that, for each prime p, the reductions $s_j \pmod p$ do not cover $\mathbb{Z}/p\mathbb{Z}$.

The Maynard-Tao weights

The original choice of w_n by Goldston, Pintz and Yıldırım was to consider the Selberg-type sieve weights

$$w_n^{\mathrm{GPY}} = \Big(\sum_{d|Q(n)} \lambda(d)\Big)^2 \qquad \text{with} \qquad Q(n) := \prod_{j=1}^{k}(n+s_j) \tag{28.4}$$

and λ an arithmetic function to be determined. However, Maynard and Tao discovered that it is much more efficient to work with a multidimensional version of the above weights: given $\lambda : \mathbb{N}^k \to \mathbb{R}$, they defined

$$w_n^{\mathrm{MT}} = \Big(\sum_{d_j|n+s_j\,\forall j} \lambda(d_1,\dots,d_k)\Big)^2. \tag{28.5}$$

For both of the above choices, the left-hand side of (28.3) can be computed under rather general assumptions on λ. As in the study of Selberg's sieve, the goal is then to optimize the choice of λ.

Various technical details are simplified if we "presieve" the support of the weights w_n with all primes $\leqslant y$. There are two main ways of accomplishing this. The first one is to restrict the support of w_n to integers $n \equiv a \,(\mathrm{mod}\, P(y))$ for some slowly growing y and an appropriate congruence class $a \,(\mathrm{mod}\, P(y))$. This is the approach taken in [138,170]. The second one, which we opt for here, is to modify slightly the weights of (28.5) by applying what is called a "preliminary sieve". By this we mean that the small prime factors of $Q(n)$ will be handled separately, using a simpler sieve.

To define the weights w_n we will use, we introduce the parameters

$$D = N^{1/4}e^{-\sqrt{\log N}}, \qquad y = \exp\{(\log\log N)^2\}, \qquad Y = \exp\{(\log\log N)^3\}.$$

We then set

$$w_n = \Big(\sum_{m|Q(n)} \mu^+(m)\Big)\Big(\sum_{d_j|n+s_j\ \forall j} \lambda(d_1,\dots,d_k)\Big)^2, \tag{28.6}$$

where:

- μ^+ is the sieve weight[2] λ^+ constructed in Theorem 19.1 with $\kappa = k$, $\mathcal{P} = \{p \leqslant y\}$ and $u = \log\log N$. In particular, $|\mu^+| \leqslant 1$ and μ^+ is supported on $\{\, d \leqslant Y : d|P(y) \,\}$.
- $\lambda : \mathbb{N}^k \to \mathbb{R}$ is a uniformly bounded function supported on

$$\mathscr{D} := \big\{(d_1,\dots,d_k) \in \mathbb{N}^k \ :\ d_1\cdots d_k \leqslant D, \quad P^-(d_j) > y \quad (1 \leqslant j \leqslant k)\big\}.$$

[2] We use the letter μ^+ instead of λ^+ to avoid confusion with the function λ.

Remark 28.2. What is important in the definition of the parameters D, Y, y is that y and $N^{1/4}/(DY^2)$ are both bigger than any fixed power of $\log N$, D grows polynomially in N and $\log Y/\log y$ is larger than $\frac{\log\log N}{\log\log\log N}$ by a factor going to infinity. A good exercise is to check that any such choice of D, Y, y is sufficient for the proof of Theorem 28.1 to go through. □

Calculations

Assuming that the weights w_n are given by (28.6), our task is to estimate the quantity

$$\mathbb{E}_{N\leqslant n\leqslant 2N}[1_P(n+s_\ell)] = \frac{\sum_{N-s_\ell\leqslant p\leqslant 2N-s_\ell} w_{p-s_\ell}}{\sum_{N\leqslant n\leqslant 2N} w_n} \tag{28.7}$$

for each $\ell = 1, \dots, k$. All implicit constants in this section might depend on k, the choice of the k-tuple $(s_1, \dots, s_k)$ and the supremum norm $B := \|\lambda\|_\infty$. We will also make use of the following notation:

$$\nu(d) = \#\{\, n \in \mathbb{Z}/d\mathbb{Z} : Q(m) \equiv 0 \,(\mathrm{mod}\, d)\,\} \qquad \text{and} \qquad V = \prod_{p\leqslant y}\Big(1-\frac{\nu(p)}{p}\Big).$$

Lemma 28.3. *Assume the above set-up and define*

$$\xi(a_1, \dots, a_k) = \sum_{(m_1,\dots,m_k)\in\mathscr{D}} \frac{\lambda(a_1m_1, \dots, a_km_k)}{m_1\cdots m_k}.$$

For any fixed $A > 0$, we have

$$\sum_{N\leqslant n\leqslant 2N} w_n = VN \sum_{(a_1,\dots,a_k)\in\mathscr{D}} \frac{\xi(a_1, \dots, a_k)^2}{a_1\cdots a_k} + O_A(N/(\log N)^A).$$

Proof. For brevity, we write $\mathbf{d}$ to denote the k-tuple of integers $(d_1, \dots, d_k)$. If $\mathbf{d}, \mathbf{e} \in \mathscr{D}$ are such that $d_i, e_i | n + s_i$ for each i, then $(d_ie, d_je_j) | s_i - s_j$ for $i \neq j$. Since the numbers d_ie_i and d_je_j have no prime factors $\leqslant y$, they must be coprime as soon as $y \geqslant s_k - s_1 \geqslant |s_i - s_j|$, which we assume from now on. Consequently,

$$\sum_{N\leqslant n\leqslant 2N} w_n = \sum_m \mu^+(m) \sum_{\substack{\mathbf{d},\mathbf{e}\in\mathscr{D}\\ (d_ie_i,d_je_j)=1\ \forall i\neq j}} \lambda(\mathbf{d})\lambda(\mathbf{e}) \sum_{\substack{N\leqslant n\leqslant 2N,\ m|Q(n)\\ [d_j,e_j]|n+s_j\ \forall j}} 1.$$

By assumption, μ^+ is supported on integers $m | P(y) = \prod_{p\leqslant y} p$, whereas λ is supported on tuples $(d_1, \dots, d_k)$ with $(d_j, P(y)) = 1$ for all j. Since we also know that $(d_ie_i, d_je_j) = 1$ for $i \neq j$, the Chinese Remainder Theorem implies that there are precisely $\nu(m)$ values of n modulo $m\prod_{j=1}^k [d_j, e_j]$ such

that $m|Q(n)$ and $[d_j, e_j]|n + s_j$ for $j = 1, \dots, k$. Therefore,

$$\sum_{\substack{N\leqslant n\leqslant 2N,\ m|Q(n)\\ [d_j,e_j]|n+s_j\ \forall j}} 1 = \frac{\nu(m)N}{m\prod_{j=1}^{j}[d_j,e_j]} + O(\nu(m)). \tag{28.8}$$

We thus arrive at the estimate

$$\sum_{N\leqslant n\leqslant 2N} w_n = V^+N \sum_{\substack{\mathbf{d},\mathbf{e}\in\mathscr{D}\\ (d_ie_i,d_je_j)=1\ \forall i\neq j}} \frac{\lambda(\mathbf{d})\lambda(\mathbf{e})}{\prod_{j=1}^{k}[d_j,e_j]} + O(R),$$

where

$$V^+ = \sum_m \frac{\mu^+(m)\nu(m)}{m} \quad \text{and} \quad R = \sum_m \sum_{\mathbf{d},\mathbf{e}\in\mathscr{D}} |\mu^+(m)\lambda(\mathbf{d})\lambda(\mathbf{e})|\nu(m).$$

We have $R = O(N^{2/3})$. Indeed, to see this, we use that $\nu(m) \leqslant \tau_k(m)$ for square-free m, $|\mu^+| \leqslant 1$, $\|\lambda\|_\infty = B = O(1)$, μ^+ is supported on $[1, Y]$ and λ is supported on tuples $(d_1, \dots, d_k)$ with $d_1 \cdots d_k \leqslant D \leqslant N^{1/4}$.

In addition, we have $V^+ = V(1 + O_A(1/(\log N)^{A+3k}))$ by Theorem 19.1, as well as

$$\sum_{\mathbf{d},\mathbf{e}\in\mathscr{D}} \frac{|\lambda(\mathbf{d})\lambda(\mathbf{e})|}{\prod_{j=1}^{k}[d_j,e_j]} \leqslant B^2 \sum_{m_1\cdots m_k\leqslant\sqrt{N}} \frac{\tau_3(m_1)\cdots\tau_3(m_k)}{m_1\cdots m_k} \ll (\log N)^{3k}, \tag{28.9}$$

where we set $m_j = [d_j, e_j]$ and used the fact that the equation $m_j = [d_j, e_j]$ has $\prod_{p^\nu\| m_j}(2\nu + 1) \leqslant \tau_3(m_j)$ solutions.

Putting everything together, we conclude that

$$\sum_{N\leqslant n\leqslant 2N} w_n = VN \sum_{\substack{\mathbf{d},\mathbf{e}\in\mathscr{D}\\ (d_ie_i,d_je_j)=1\ \forall i\neq j}} \frac{\lambda(\mathbf{d})\lambda(\mathbf{e})}{\prod_{j=1}^{k}[d_j,e_j]} + O_A(N/(\log N)^A).$$

Next, we remove the conditions that $(d_ie_i, d_je_j) = 1$ for $i \neq j$. Since $\mathbf{d}, \mathbf{e} \in \mathscr{D}$, we have $(d_ie_i, d_je_j, P(y)) = 1$. Hence, if $(d_ie_i, d_je_j) > 1$ for some $i \neq j$, there must exist a prime $p > y$ dividing $[d_i, e_i]$ and $[d_j, e_j]$. Setting $m_r = [d_r, e_r]$ for $r = 1, \dots, k$, we conclude that

$$\sum_{\substack{\mathbf{d},\mathbf{e}\in\mathscr{D}\\ (d_ie_i,d_je_j)>1}} \frac{|\lambda(\mathbf{d})\lambda(\mathbf{e})|}{\prod_{j=1}^{k}[d_j,e_j]} \leqslant B^2 \sum_{p>y} \sum_{\substack{m_r\leqslant D^2\ \forall r\\ p|m_i,m_j}} \frac{\tau_3(m_1)\cdots\tau_3(m_k)}{m_1\cdots m_k} \ll \frac{(\log N)^{3k}}{y}.$$

Hence, we have arrived at the estimate

$$\sum_{N\leqslant n\leqslant 2N} w_n = VN \sum_{\mathbf{d},\mathbf{e}\in\mathscr{D}} \frac{\lambda(\mathbf{d})\lambda(\mathbf{e})}{\prod_{j=1}^{k}[d_j,e_j]} + O_A(N/(\log N)^A). \tag{28.10}$$

The next step is to rewrite the terms $1/[d_j, e_j]$. To do so, we use that

$$\frac{1}{[d,e]} = \frac{(d,e)}{de} = \frac{1}{de} \sum_{a|(d,e)} \varphi(a),$$

just like we did when we studied Selberg's sieve. We thus deduce that

$$\sum_{N \leqslant n \leqslant 2N} w_n = VN \sum_{\mathbf{a} \in \mathscr{D}} \frac{\varphi(a_1) \cdots \varphi(a_k)}{a_1^2 \cdots a_k^2} \cdot \xi(\mathbf{a})^2 + O_A(N/(\log N)^A).$$

Finally, we remove the factors $\varphi(a_j)/a_j = \prod_{p|a_j}(1 - 1/p)$. Note that $\omega(a_j) \ll \log a_j \leqslant \log N$, as well as $(a_j, P(y)) = 1$ for all j. Consequently,

$$1 \geqslant \varphi(a)/a \geqslant (1 - 1/y)^{O(\log N)} \geqslant 1 - O((\log N)/y) \tag{28.11}$$

by our choice of y. Bounding the total contribution of the error terms using (28.9) comletes the proof of the lemma. □

Lemma 28.4. *Assume the above set-up and define*

$$\zeta_\ell(a_1, \dots, a_k) = 1_{a_\ell = 1} \sum_{\substack{(m_1, \dots, m_k) \in \mathscr{D} \\ m_\ell = 1}} \frac{\lambda(a_1 m_1, \dots, a_k m_k)}{m_1 \cdots m_k}.$$

If we let $X = \int_N^{2N} \mathrm{d}t/\log t$, *then for any fixed* $A > 0$ *we have*

$$\sum_{N - s_\ell \leqslant p \leqslant 2N - s_\ell} w_{p - s_\ell} = \frac{VX}{\prod_{p \leqslant y}(1 - 1/p)} \sum_{(a_1, \dots, a_k) \in \mathscr{D}} \frac{\zeta_\ell(a_1, \dots, a_k)^2}{a_1 \cdots a_k} + O_A(N/(\log N)^A).$$

Proof. To simplify the notation, we consider the case $\ell = 1$; the proof of the other cases follows *mutatis mutandis*.

Since $w_n = n^{o(1)}$ from the divisor bound (see Exercise 2.9(f)), we have

$$\sum_{N - s_1 \leqslant n \leqslant 2N - s_1} w_{p - s_1} = \sum_{N \leqslant p \leqslant 2N} w_{p - s_1} + O(N^{o(1)}). \tag{28.12}$$

If W denotes the sum on the right side of (28.12), then

$$W = \sum_m \mu^+(m) \sum_{\mathbf{d}, \mathbf{e} \in \mathscr{D}} \lambda(\mathbf{d}) \lambda(\mathbf{e}) \sum_{\substack{N \leqslant p \leqslant 2N,\ m | Q(p - s_1) \\ [d_j, e_j] | p - s_1 + s_j\ \forall j}} 1.$$

As in Lemma 28.3, we can only have $d_j, e_j | p - s_1 + s_j$ for all j when $(d_i e_i, d_j e_j) = 1$ for all $i \neq j$. Notice though that there is something special that takes place when $j = 1$: we then have $d_1, e_1 | p - s_1 + s_1 = p$. Since a prime number p has only trivial factors and $d_1, e_1 \leqslant N^{1/4} < p$, we conclude

that $d_1 = e_1 = 1$. Similarly, if $m|Q(p-s_1) = \prod_{j=1}^k (p+s_j-s_1)$, then $m|Q^*(p)$, where

$$Q^*(x) = \prod_{j=2}^k (x+s_j-s_1).$$

As a consequence,

$$W = \sum_m \mu^+(m) \sum_{\substack{\mathbf{d},\mathbf{e}\in\mathscr{D},\ d_1=e_1=1\\ (d_ie_i,d_je_j)=1\ \forall i\neq j}} \lambda(\mathbf{d})\lambda(\mathbf{e}) \sum_{\substack{N<p\leqslant 2N,\ m|Q^*(p)\\ [d_j,e_j]|p+s_j-s_1\ \forall j\geqslant 2}} 1. \tag{28.13}$$

To evaluate the innermost sum, we adapt the argument leading to (28.8). If $q = m\prod_{j=2}^k [d_j,e_j]$, then the Chinese Remainder Theorem implies that the number of $x \pmod q$ such that $m|Q^*(x)$ and $[d_j,e_j]|x+s_j-s_1$ equals $\#\{x \pmod m : m|Q^*(x)\}$. However, since p is prime, we must only count solutions that are reduced residues mod q. Whenever $x \equiv s_1-s_j \pmod{[d_j,e_j]}$, we also have $(x,[d_j,e_j]) = 1$ because $(d_je_j, P(y)) = 1$ and $y > |s_1-s_j|$ for each j. In conclusion, the number of reduced solutions mod q is

$$\nu^*(m) := \#\{\, x \in (\mathbb{Z}/m\mathbb{Z})^* : m|Q^*(x)\,\}.$$

Hence, the innermost sum in (28.13) equals

$$\frac{\nu^*(m)}{\varphi(q)}X + O(\nu^*(m)E(N,q)), \tag{28.14}$$

where

$$E(N,q) := \max_{(a,q)=1} \big|\pi(2N;q,a) - \pi(N;q,a) - X/\varphi(q)\big|.$$

The modulus q here is an integer

$$\leqslant Q := YD^2 = N^{1/2}\exp\big((\log\log N)^3 - 2\sqrt{\log N}\,\big)$$

by our assumptions on the support of μ^+ and of λ. Moreover, if we are given such an integer q, there are $\leqslant \tau_{k-1}(q)\tau_3(q)$ ways to write it in the form $m\prod_{j=2}^k [d_j,e_j]$ with m, d_2, …, d_k, e_2, …, e_k as in the right-hand side of (28.13). Since we also have $\nu^*(m) \leqslant (k-1)^{\omega(q)}$, we arrive at the formula

$$W = XV^* \sum_{\substack{\mathbf{d},\mathbf{e}\in\mathscr{D},\ d_1=e_1=1\\ (d_ie_i,d_je_j)=1\ \forall i\neq j}} \frac{\lambda(\mathbf{d})\lambda(\mathbf{e})}{\prod_{j=2}^k \varphi([d_j,e_j])} + O(R^*)$$

with

$$V^* = \sum_m \frac{\mu^+(m)\nu^*(m)}{\varphi(m)} \qquad \text{and} \qquad R^* = \sum_{q\leqslant Q} \tau_{k-1}(q)^2\tau_3(q)E(N,q).$$

We use the Bombieri-Vinogradov theorem (see also Exercise 26.2) to find that $R^* = O_A(N/(\log N)^A)$. In addition, we note that $\nu^*(p) = p-1$ and use

Theorem 19.1 to find that $V^* = [1 + O_A(1/(\log N)^{A+3k})]V/\prod_{p\leqslant y}(1-1/p)$. As a consequence,

$$W = \frac{VX}{\prod_{p\leqslant y}(1-1/p)} \sum_{\substack{\mathbf{d},\mathbf{e}\in\mathscr{D},\ d_1=e_1=1\\ (d_ie_i,d_je_j)=1\ \forall i\neq j}} \frac{\lambda(\mathbf{d})\lambda(\mathbf{e})}{\prod_{j=2}^k \varphi([d_j,e_j])} + O_A\Big(\frac{N}{(\log N)^A}\Big),$$

where the error term from the estimation of V^* was handled using (28.9).

Next, as in the proof of Lemma 28.3, we may remove the conditions $(d_ie_i, d_je_j) = 1$ when $i \neq j$ at the cost of an error of size $\ll N(\log N)^{3k}/y$. Finally, we may replace $\varphi([d_j, e_j])$ by $[d_j, e_j]$ using (28.11) at the cost of an error term of size $N(\log N)^{O(1)}/y$. Hence, we arrive at the formula

$$W = \frac{VX}{\prod_{p\leqslant y}(1-1/p)} \sum_{\substack{\mathbf{d},\mathbf{e}\in\mathscr{D}\\ d_1=e_1=1}} \frac{\lambda(\mathbf{d})\lambda(\mathbf{e})}{\prod_{j=2}^k [d_j,e_j]} + O_A\Big(\frac{N}{(\log N)^A}\Big),$$

which is analogous to (28.10). It is now straightforward to adapt the argument from the proof of Lemma 28.3 that estimates the right-hand side of (28.10), and to complete the proof of the lemma. □

A change of variables à la Selberg

Motivated by the theory of the Selberg sieve, we will switch from the function λ to the function ξ defined in Lemma 28.3. We must write ζ_ℓ in terms of this new function.

Lemma 28.5. *For each $(a_1, \dots, a_k) \in \mathscr{D}$ and each $\ell \in \{1, \dots, k\}$, we have*

$$\zeta_\ell(a_1,\dots,a_k) = 1_{a_\ell=1} \sum_b \frac{\mu(b)\xi(a_1,\dots,a_{\ell-1},b,a_{\ell+1},\dots,a_k)}{b}.$$

Proof. To ease the notation, we demonstrate the calculation when $\ell = 1$. As in the proof of Theorem 21.1 (see the argument leading to relation (21.5)), we have the inversion formula

$$\lambda(d_1,\dots,d_k) = 1_{(d_1,\dots,d_k)\in\mathscr{D}} \sum_{b_1,\dots,b_k} \frac{\mu(b_1)\cdots\mu(b_k)\xi\big(b_1d_1,\dots,b_kd_k\big)}{b_1\cdots b_k}. \tag{28.15}$$

Consequently, if $(a_1, \dots, a_k) \in \mathscr{D}$ with $a_1 = 1$, then

$$\begin{aligned}\zeta_1(a_1,a_2,\dots,a_k) &= \sum_{d_2,\dots,d_k} \frac{\lambda(1,a_2d_2,\dots,a_kd_k)}{d_2\cdots d_k}\\ &= \sum_{d_2,\dots,d_k}\sum_{b_1,\dots,b_k} \frac{\mu(b_1)\cdots\mu(b_k)\xi(b_1,a_2b_2d_2,\dots,a_kb_kd_k)}{d_2\cdots d_kb_1\cdots b_k}.\end{aligned}$$

Making the change of variables $m_j = b_j d_j$ for all $j > 1$ implies that

$$\zeta_1(a_1, a_2, \dots, a_k) = \sum_{b_1, m_2, \dots, m_k} \frac{\mu(b_1)\xi(b_1, a_2 m_2, \dots, a_k m_k)}{b_1 m_2 \cdots m_k} \prod_{j=2}^{k} \sum_{b_j d_j = m_j} \mu(b_j).$$

The innermost sum vanishes unless $m_j = 1$, thus completing the proof. □

Choosing the function ξ

Motivated by Lemma 28.5, we set

$$\xi(a_1, \dots, a_k) := \frac{1_{P^-(a_1 \cdots a_k) > y} (-1)^{\Omega(a_1 \cdots a_k)}}{(\log D)^k \prod_{p \leqslant y} (1 - 1/p)^k} \cdot f\Big(\frac{\log a_1}{\log D}, \dots, \frac{\log a_k}{\log D}\Big),$$

where f is a smooth function supported on the simplex

$$\Delta_k := \{(x_1, \dots, x_k) \in [0,1]^k : x_1 + \cdots + x_k \leqslant 1\},$$

and the factor $(-1)^{\Omega(a_1 \cdots a_k)}$ is introduced to annihilate the sign changes caused by $\mu(b)$ in the expression for ζ_ℓ in Lemma 28.5. Lastly, the denominator $(\log D)^k \prod_{p \leqslant y} (1 - 1/p)^k$ is introduced for normalization purposes, so that $\|\lambda\|_\infty = O(1)$ by (28.15), as needed. With this choice of ξ, we have the following result.

Lemma 28.6. *Let $\ell \in \{1, \dots, k\}$, and set*

$$I_\ell(f) = \int_{\mathbb{R}^{k-1}} \bigg(\int_{\mathbb{R}} f(x_1, \dots, x_k) \mathrm{d}x_\ell \bigg)^2 \mathrm{d}x_1 \cdots \mathrm{d}x_{\ell-1} \mathrm{d}x_{\ell+1} \cdots \mathrm{d}x_k$$

and

$$J(f) = \int_{\mathbb{R}^k} f(x_1, \dots, x_k)^2 \mathrm{d}x_1 \cdots \mathrm{d}x_k.$$

If ξ and f are as above, and we assume that $J(f) \gg 1$, then

$$\mathbb{E}_{N \leqslant n \leqslant 2N} \big[1_P(n + s_\ell)\big] = \frac{I_\ell(f)}{4J(f)} + O\big(1/\sqrt{\log N}\,\big).$$

The proof of Lemma 28.6 rests on the following result, which we will eventually apply with $w = y$ so that $u \asymp (\log N)/(\log\log N)^2$.

Lemma 28.7. *If $r \in \mathbb{Z}_{\geqslant 1}$, $g : \mathbb{R}^r \to \mathbb{R}$ is a smooth function supported on Δ_r and $u, w \geqslant 2$ are such that $D = w^u$, then*

$$\sum_{\substack{P^-(n_j) > w \\ 1 \leqslant j \leqslant r}} \frac{g(\frac{\log n_1}{\log D}, \dots, \frac{\log n_r}{\log D})}{n_1 \cdots n_r} = (\log D)^r \prod_{p \leqslant w} \Big(1 - \frac{1}{p}\Big)^r \bigg(\int_{\Delta_r} g + O\Big(\frac{\log u}{u}\Big)\bigg),$$

where the implied constant depends at most on r, and on the supremum norm of g and of its partial derivatives.

Proof. All implied constants might depend on g and on r as described in the statement of the lemma. Throughout the proof, we set

$$z := w^{\log u} = D^{\frac{\log u}{u}} \qquad \text{and} \qquad G(x_1, \dots, x_r) := \frac{g(\frac{\log x_1}{\log D}, \dots, \frac{\log x_r}{\log D})}{x_1 \cdots x_r}.$$

Note that

$$G(x_1, \dots, x_r) \ll \frac{1}{x_1 \cdots x_r}, \qquad \frac{\partial G}{\partial x_j}(x_1, \dots, x_r) \ll \frac{1}{x_j} \cdot \frac{1}{x_1 \cdots x_r}. \tag{28.16}$$

We will often denote the r-tuple $(x_1, \dots, x_r)$ by the bold letter $\mathbf{x}$.

Given a parameter $X \geqslant 1$, let

$$\mathcal{D}_X := \{(x_1, \dots, x_r) \in [X, +\infty)^r : x_1 \cdots x_r \leqslant D\}.$$

We first show we may restrict our attention to tuples $(n_1, \dots, n_r) \in \mathcal{D}_z$. Indeed, for each fixed $j \in \{1, \dots, r\}$, the contribution to the sum of the statement of the lemma of those summands with $n_j \leqslant z$ is

$$\ll \sum_{\substack{n_j \leqslant z,\ n_i \leqslant D\ \forall i \neq j \\ P^-(n_1 \cdots n_r) > w}} \cdots \sum \frac{1}{n_1 \cdots n_r} \ll (\log D)^{r-1} (\log z) \prod_{p \leqslant w} \left(1 - \frac{1}{p}\right)^r,$$

which is of admissible size by our choice of z.

Next, we treat the part of the sum over $(n_1, \dots, n_r) \in \mathcal{D}_z$ by splitting it into small rectangles. We set $z_0 = z$ and, having defined z_a, we set $z_{a+1} = z_a + \sqrt{z_a}$. Moreover, let $\mathscr{I}$ denote the set of rectangles of the form $I = \prod_{j=1}^r (x_j, x_j + \sqrt{x_j}]$, where $x_1, \dots, x_r \in \{z_0, z_1, \dots\}$. Finally, we write $\mathcal{D}'_z$ for the union of all such rectangles that lie entirely within $\mathcal{D}_z$, that is to say, $\mathcal{D}'_z$ is the union of I of the above form for which $\prod_{j=1}^r (x_j + \sqrt{x_j}) \leqslant D$. In particular, if $(x_1, \dots, x_r) \in \mathcal{D}_z \setminus \mathcal{D}'_z$, then $D/(1 + z^{-1/2})^r \leqslant x_1 \cdots x_r \leqslant D$. Combining this with the first bound of (28.16), we find that

$$\sum_{\substack{\mathbf{n} \in \mathcal{D}_z \\ P^-(n_1 \cdots n_r) > w}} G(\mathbf{n}) = \sum_{\substack{\mathbf{n} \in \mathcal{D}'_z \\ P^-(n_1 \cdots n_r) > w}} G(\mathbf{n}) + O(N/D), \tag{28.17}$$

where $N := \#\{\, (n_1, \dots, n_r) \in \mathbb{N}^r : n_1 \cdots n_r \in [D/(1+z^{-1/2})^r, D]\,\}$. Exercise 3.10 implies that $N \ll D(\log D)^{r-1}/\sqrt{z} + D^{1-1/r}$, so that the error term in (28.17) is of admissible size.

Now, fix $I = \prod_{j=1}^r (x_j, x_j + \sqrt{x_j}] \subseteq \mathcal{D}_z$. Since I has volume $\sqrt{x_1 \cdots x_r}$, we have

$$G(\mathbf{n}) = \frac{1}{\sqrt{x_1 \cdots x_r}} \int_I G(\mathbf{n}) \mathrm{d}\mathbf{t}.$$

When $\mathbf{n}, \mathbf{t} \in I$, the Mean Value Theorem and the second bound of (28.16) imply that $G(\mathbf{n}) - G(\mathbf{t}) = O(z^{-1/2}/(t_1 \cdots t_r))$, since $1/(1+1/\sqrt{z}) \leqslant t_j/n_j \leqslant$

$1 + 1/\sqrt{z}$ for all j. We thus conclude that

$$G(\mathbf{n}) = \frac{1}{\sqrt{x_1 \cdots x_r}} \int_I \Big(G(\mathbf{t}) + \frac{O(z^{-1/2})}{t_1 \cdots t_r} \Big) \mathrm{d}\mathbf{t}$$

for all $\mathbf{n} \in I$. In addition, applying Theorem 18.11(a) r times, we find that

$$\begin{aligned} &\#\{\, (n_1, \dots, n_r) \in I \cap \mathbb{Z}^r : P^-(n_1 \cdots n_r) > w \,\} \\ &\qquad = (1 + O(1/u)) \sqrt{x_1 \cdots x_r} \prod_{p \leqslant w} \Big(1 - \frac{1}{p} \Big)^r, \end{aligned}$$

since $x_j \geqslant z = w^{\log u}$ for each j. Putting the above estimates together yields the formula

$$\sum_{\substack{\mathbf{n} \in I \\ P^-(n_1 \cdots n_r) > w}} G(\mathbf{n}) = \prod_{p \leqslant w} \Big(1 - \frac{1}{p} \Big)^r \int_I \Big(G(\mathbf{t}) + \frac{O(1/u)}{t_1 \cdots t_r} \Big) \mathrm{d}\mathbf{t}.$$

Finally, we sum the above estimate over all rectangles $I \in \mathscr{I}$; they are all subsets of $\mathcal{D}_z$ by definition. The lemma then follows by combining the resulting formula with (28.17) and the fact that

$$\int_{\mathcal{E}} G(t_1, \dots, t_r) \mathrm{d}t_1 \cdots \mathrm{d}t_r \ll \int_{\mathcal{E}} \frac{\mathrm{d}t_1 \cdots \mathrm{d}t_r}{t_1 \cdots t_r} \ll (\log D)^{r-1} \log z,$$

where $\mathcal{E} = (\mathcal{D}_1 \setminus \mathcal{D}_z) \cup (\mathcal{D}_z \setminus \mathcal{D}'_z)$. □

Proof of Lemma 28.6. To ease the notation, we give the proof when $\ell = 1$; the other cases are similar. Let us begin by recalling relation (28.7). Together with Lemmas 28.3 and 28.4, it implies that

$$\mathbb{E}_{N \leqslant n \leqslant 2N} \big[1_P(n + s_1) \big] = \frac{VXS_1 / \prod_{p \leqslant y} (1 - 1/p) + O_A(N/(\log N)^{A+1})}{VNT + O_A(N/(\log N)^A)}$$

for any A, where

$$S_1 = \sum_{\substack{(a_1, \dots, a_k) \in \mathscr{D} \\ a_1 = 1}} \frac{\zeta_1(a_1, \dots, a_k)^2}{a_1 \cdots a_k} \qquad \text{and} \qquad T = \sum_{(a_1, \dots, a_k) \in \mathscr{D}} \frac{\xi(a_1, \dots, a_k)^2}{a_1 \cdots a_k}.$$

The choice of ξ and Lemma 28.7 (applied with $g = f^2$, $r = k$, $w = y$ and $u \asymp (\log N)/(\log\log N)^2$) implies that

$$T = L^{-k} \big(J(f) + O(1/\sqrt{\log N}) \big), \tag{28.18}$$

where

$$L := (\log D) \prod_{p \leqslant y} (1 - 1/p) \asymp \frac{\log N}{\log y}.$$

In addition, note that $V \gg 1/(\log y)^k$. Therefore, if take $A = k + 2$, then $O(N/(\log N)^A) = O(VNL^{-k}/\log N)$. Since we also have that $X/N =$

$1/\log N + O(1/\log^2 N)$ and $\log N = 4\log D + O(\sqrt{\log N})$, the lemma will follow as long as we can show that

$$S_1 = L^{1-k}\big(I_1(f) + O(1/\sqrt{\log N})\big). \tag{28.19}$$

For any $\mathbf{a} = (a_1, \dots, a_k) \in \mathscr{D}$ with $a_1 = 1$, Lemma 28.5 and our choice of ξ imply that

$$\zeta_1(\mathbf{a}) = \frac{(-1)^{\Omega(a_2\cdots a_k)}}{L^k} \sum_{P^-(b)>y} \frac{\mu^2(b) f(\frac{\log b}{\log D}, \frac{\log a_2}{\log D}, \dots, \frac{\log a_k}{\log D})}{b}.$$

We remove the weight $\mu^2(b)$, by noticing that if $\mu^2(b) = 0$ and $P^-(b) > y$, then there is a prime $p > y$ such that $p^2|b$. Hence, the total error produced by replacing $\mu^2(b)$ with 1 is $\ll L^{-k}/y$. To the rest of the sum, we apply Lemma 28.7 with $r = 1$, $w = y$ and $u \asymp (\log N)/(\log\log N)^2$. This yields the estimate

$$\zeta_1(\mathbf{a}) = \frac{(-1)^{\Omega(a_2\cdots a_k)}}{L^{k-1}} \bigg(\int_{\mathbb{R}} f\Big(x_1, \frac{\log a_2}{\log D}, \dots, \frac{\log a_k}{\log D}\Big) \mathrm{d}x_1 + O\big(1/\sqrt{\log N}\,\big) \bigg).$$

Therefore

$$S_1 = L^{2-2k} \sum_{P^-(a_2\cdots a_k)>y} \frac{\big(\int_{\mathbb{R}} f\big(x_1, \frac{\log a_2}{\log D}, \dots, \frac{\log a_k}{\log D}\big)\mathrm{d}x_1\big)^2}{a_2\cdots a_k} + O\bigg(\frac{L^{1-k}}{\sqrt{\log N}}\bigg),$$

where we used an upper bound sieve to control the contribution of the remainder terms. Finally, we apply again Lemma 28.6, this time with $r = k-1$, to deduce (28.19). This completes the proof of the lemma. □

Optimizing the function f

In view of Lemma 28.6, our goal is to choose f supported on Δ_k and maximizing the ratio

$$\rho_k(f) := \sum_{\ell=1}^{k} \frac{I_\ell(f)}{J(f)}.$$

If we can show that $\rho_k(f) > 4m$ for k large enough in terms of m, we automatically conclude that $\liminf_{n\to\infty}(p_{n+m} - p_n) \leqslant s_k - s_1 < \infty$.

As a warm-up exercise, we study $\rho_k(f)$ using calculus of variations. Note that we may drop the assumption that f is smooth, since the integral of any measurable function over a compact region can be approximated arbitrarily closely by integrals of smooth functions. Consider the linear operator

$$(\mathcal{L}_k f)(x_1, \dots, x_k) := \sum_{\ell=1}^{k} \int_{\mathbb{R}} f(x_1, \dots, x_{\ell-1}, t, x_{\ell+1}, \dots, x_k)\mathrm{d}t,$$

acting on $C_c(\mathbb{R})$, the space of compactly supported, continuous functions $f : \mathbb{R}^k \to \mathbb{R}$. Letting $\langle f, g\rangle = \int_{\mathbb{R}^k} fg$, we find that

$$\sum_{\ell=1}^{k} I_\ell(f) = \langle \mathcal{L}_k f, f\rangle,$$

whereas $J(f) = \langle f, f\rangle$. If, now, f is a maximizer of the function $\rho_k(\cdot)$ over all f supported on Δ_k, then the function $\varepsilon \to \rho_k(f + \varepsilon g)$ has a maximum at $\varepsilon = 0$ for any continuous $g : \mathbb{R}^k \to \mathbb{R}$ supported on Δ_k. So its derivative at $\varepsilon = 0$ must vanish, which implies that

$$\langle \mathcal{L}_k f, g\rangle + \langle \mathcal{L}_k g, f\rangle = 2\rho_k(f)\langle f, g\rangle.$$

It is easy to see that $\mathcal{L}_k$ is a self-adjoint operator, so we find that

$$\langle \mathcal{L}_k f, g\rangle = \rho_k(f)\langle f, g\rangle. \tag{28.20}$$

Lastly, a standard continuity argument allows us to extend (28.20) to all bounded measurable functions g that are supported on Δ_k.

We apply (28.20) for a special choice of g. Let $(B_n)_{n=1}^{\infty}$ be a shrinking family of cubes centered at a given point $(x_1, \dots, x_k)$ in the interior of the simplex Δ_k, and take $g = 1_{B_n}/\operatorname{Vol}(B_n)$ with $n \to \infty$. Applying (28.20) to this family of functions g, we deduce that $(\mathcal{L}_k f)(x_1, \dots, x_k) = \rho(f) f(x_1, \dots, x_k)$. Since $(x_1, \dots, x_k)$ is arbitrary and f is continuous, this implies that $(\mathcal{L}_k f)|_{\Delta_k} = \rho(f) \cdot f$. In particular, f is an eigenfunction of the operator $\widetilde{\mathcal{L}}_k g := (\mathcal{L}_k g)|_{\Delta_k}$ that acts on continuous functions $g : \Delta_k \to \mathbb{R}$. The corresponding eigenvalue is $\rho_k(f)$.

Now, note that if f is an eigenfunction of the operator $\widetilde{\mathcal{L}}_k$ of eigenvalue $\rho_k(f)$, so is its symmetric version

$$\widetilde{f}(x_1, \dots, x_k) := \sum_{\sigma \in S_k} f(x_{\sigma(1)}, \dots, x_{\sigma(k)}).$$

In light of this observation, we may restrict our attention to symmetric functions f, in which case

$$\rho_k(f) = kI_1(f)/J(f). \tag{28.21}$$

To this end, we define

$$R_k := \sup\{\, \rho_k(f) : f : \Delta_k \to \mathbb{R},\ f \text{ symmetric and continuous}\,\}.$$

An asymptotic estimation for R_k is given in Proposition 28.8 below. In addition, explicit bounds on R_k can be found [138, 151].

Proposition 28.8. *For large integers k, we have that*

$$\log k - 4\log\log k + O(1) \leqslant R_k \leqslant \log k + \log\log k + O(1).$$

Proof. For the lower bound, we consider functions of the form

$$f(x_1, \dots, x_k) = 1_{(x_1,\dots,x_k)\in\Delta_k} \cdot g(kx_1) \cdots g(kx_k),$$

where $g : \mathbb{R} \to \mathbb{R}_{\geqslant 0}$ is a function supported on the interval $[0, \delta k]$ with $\delta \in (0, 1)$ to be chosen later, and such that $\int_0^\infty g(t)^2 \mathrm{d}t = 1$. Then

$$J(f) = \int_{\mathbb{R}^k} f(x_1, \dots, x_k)^2 \mathrm{d}x_1 \cdots \mathrm{d}x_k \leqslant \Big(\int_{\mathbb{R}} g(kx)^2 \mathrm{d}x \Big)^k = \frac{1}{k^k}.$$

Together with (28.21) and the change of variables $t_j = kx_j$, this implies that

$$\begin{aligned}
\rho_k(f) &\geqslant \int_{\mathbb{R}^{k-1}} g(t_2)^2 \cdots g(t_k)^2 \Big(\int_0^{k-(t_2+\cdots+t_k)} g(t_1)\mathrm{d}t_1 \Big)^2 \mathrm{d}t_2 \cdots \mathrm{d}t_k \\
&\geqslant \Big(\int_0^\infty g(t)\mathrm{d}t \Big)^2 \int_{t_2+\cdots+t_k \leqslant (1-\delta)k} g(t_2)^2 \cdots g(t_k)^2 \mathrm{d}t_2 \cdots \mathrm{d}t_k \\
&= \Big(\int_0^\infty g(t)\mathrm{d}t \Big)^2 \mathbb{P}\big(X_2 + \cdots + X_k \leqslant (1-\delta)k\big).
\end{aligned}$$

where $X_2, \dots, X_k$ are independent random variables with density function g^2. Let

$$\mu = \mathbb{E}[X_2] = \int_0^\infty t g(t)^2 \mathrm{d}t$$

and $Y_i = X_i - \mu$, $2 \leqslant i \leqslant k$, so that $Y_2, \dots, Y_k$ are mean-zero independent random variables that are identically distributed.

If we assume that $\delta < 1 - \mu$, then

$$\begin{aligned}
\mathbb{P}\big(X_2 + \cdots + X_k > (1-\delta)k\big) &\leqslant \mathbb{P}\big(Y_2 + \cdots + Y_k > (1-\delta-\mu)k\big) \\
&\leqslant \frac{\mathbb{V}[Y_2 + \cdots + Y_k]}{(1-\delta-\mu)^2 k^2} \leqslant \frac{\mathbb{V}[Y_2]}{(1-\delta-\mu)^2 k}
\end{aligned}$$

by Chebyshev's inequality and the independence of the Y_i's. Furthermore,

$$\mathbb{V}[Y_2] \leqslant \mathbb{E}[X_2^2] = \int_0^\infty t^2 g(t)^2 \mathrm{d}t \leqslant \delta k \int_0^\infty t g(t)^2 \mathrm{d}t = \delta k \mu$$

by our assumption that g is supported on $[0, \delta k]$. In conclusion, we have

$$R_k \geqslant \rho_k(f) \geqslant \Big(\int_0^\infty g(t)\mathrm{d}t \Big)^2 \Big(1 - \frac{\delta\mu}{(1-\delta-\mu)^2} \Big) \tag{28.22}$$

for any measurable function $g \geqslant 0$ supported on $[0, \delta k]$ with $\int_{\mathbb{R}} g^2 = 1$ and $\mu = \int u g(u)^2 \mathrm{d}t < 1 - \delta$. We choose

$$g(t) = c \cdot \frac{1_{[0,\delta k]}(t)}{1 + At},$$

where the parameters δ, c, A will be determined shortly.

First of all, note that the hypothesis that $\int_{\mathbb{R}} g^2 = 1$ implies that

$$c^{-2} = \int_0^{\delta k} \frac{\mathrm{d}t}{(1+At)^2} = \frac{1}{A \cdot (1 + 1/(A\delta k))}.$$

Hence

$$\mu = \int_0^{\delta k} \frac{c^2 t}{(1+At)^2}\mathrm{d}t = \frac{1+1/(A\delta k)}{A}\Big(\log(1+A\delta k) - 1 + \frac{1}{1+A\delta k}\Big).$$

To force μ to be close to 1 and δ to be smaller than $1-\mu$, we take $A = \log k$ and $\delta = 1/(\log k)^3$, so that

$$\begin{aligned}\mu &= \frac{1}{\log k}\big(\log(k/(\log k)^2) + O(1)\big)\\ &= 1 - \frac{2\log\log k}{\log k} + O(1/\log k)\\ &\leqslant 1 - \delta - 1/\log k\end{aligned}$$

for k large enough. Since $\int_0^\infty g(t)\mathrm{d}t = c\log(1+A\delta k)/A$, the above inequality and (28.22) imply the lower bound

$$\begin{aligned}R_k &\geqslant \frac{c^2\log^2(1+A\delta k)}{A^2}(1 - 1/\log k)\\ &= \frac{\log^2(k/(\log k)^2)}{\log k}(1 - 1/\log k) + O(1)\\ &= \log k - 4\log\log k + O(1),\end{aligned}$$

as claimed.

Finally, we prove the upper bound on R_k. Let f be a symmetric, measurable function supported on Δ_k. Motivated by the shape of f yielding our lower bound on R_k, we use the Cauchy-Schwarz inequality in the following fashion:

$$\begin{aligned}\Big(\int_{\mathbb{R}} f(\mathbf{x})\mathrm{d}x_1\Big)^2 = \Big(\int_0^1 f(\mathbf{x})\mathrm{d}x_1\Big)^2 &\leqslant \int_0^1 (1+kAx_1)f(\mathbf{x})^2\mathrm{d}x_1 \int_0^1 \frac{\mathrm{d}x_1}{1+kAx_1}\\ &= \frac{\log(1+kA)}{kA}\int_0^1 (1+kAx_1)f(\mathbf{x})^2\mathrm{d}x_1.\end{aligned}$$

Therefore

$$I_1(f) \leqslant \frac{\log(1+kA)}{kA}\int_0^1 (1+kAx_1)f(\mathbf{x})^2\mathrm{d}\mathbf{x}.$$

By symmetry,

$$\rho_k(f)\int_{\mathbb{R}^k} f(\mathbf{x})^2\mathrm{d}\mathbf{x} = \sum_{\ell=1}^k I_\ell(f) \leqslant \frac{\log(1+kA)}{kA}\sum_{\ell=1}^k \int_0^1 (1+kAx_\ell)f(\mathbf{x})^2\mathrm{d}\mathbf{x}.$$

Since $x_1 + \cdots + x_k \leqslant 1$ in the support of f, we conclude that $\rho_k(f) \leqslant (1+kA)\log(1+kA)/(kA)$. Taking $A = \log k$ completes the proof. □

We may now complete the proof of the main result of this chapter.

Proof of Theorem 28.1. Combining Lemma 28.6 and Proposition 28.8, we find that there is a choice of weights w_n such that

$$\mathbb{E}_{N\leqslant n\leqslant 2N}\Big(\sum_{1\leqslant \ell\leqslant k} 1_P(n+s_\ell)\Big) \geqslant \frac{\log k}{4} - \log\log k + O(1).$$

We take $k = \lceil Cm^4e^{4m}\rceil$ for a large enough constant C so that the right-hand side becomes $> m$. In particular, there must exist $n \in [N, 2N]$ such that $n + s_j$ is prime for at least $m + 1$ values of j. Since N can be taken to be arbitrarily large, we conclude that $\liminf_{n\to\infty}(p_{n+m} - p_n) \leqslant s_k - s_1$. We take s_j to be the jth prime that is $> k$. We may easily check that the tuple $(s_1, \dots, s_k)$ is admissible. Since $s_k \lesssim k\log k \ll e^{4m}m^5$ by the Prime Number Theorem, we have completed the proof of Theorem 28.1. □

Exercises

Exercise 28.1. (a) Fix coprime $a, q \in \mathbb{N}$ and let $p'_1 < p'_2 < \cdots$ be the sequence of primes $\equiv a \pmod q$. Show that $\lim_{n\to\infty}(p'_{n+1} - p'_n) < \infty$.

(b) Let $q_1 < q_2 < \cdots$ be an infinite sequence of primes. Find necessary conditions so that $\liminf_{n\to\infty}(q_{n+1} - q_n) < \infty$.

Exercise 28.2. Let $\mathcal{S}$ be the set of integers $s \geqslant 1$ for which there are infinitely many primes p such that $p + 2s$ is also prime. For every x that is sufficiently large, show that $\#\mathcal{S} \cap [1, x] \gg x$. [*Hint:* Show that there is some H such that $\mathcal{S} \cap (m, m + H] \neq \emptyset$ for all $m \geqslant 1$.]

Exercise 28.3. (a) If $f(x_1, \dots, x_k) = F(x_1 + \cdots + x_k)1_{x_1,\dots,x_k\geqslant 0}$ with $F : [0, 1] \to \mathbb{R}$ continuous,[3] then show that

$$\rho_k(f) = \tilde\rho_k(F) := k(k-1)\cdot \frac{\int_0^1 u^{k-2}(\int_u^1 F)^2\mathrm{d}u}{\int_0^1 u^{k-1}F(u)^2\mathrm{d}u}. \tag{28.23}$$

(b) (Goldston-Pintz-Yıldırım [63]) Show that $\sup_F \tilde\rho_k(F) \geqslant 4 - o_{k\to\infty}(1)$. [*Hint:* Take $F(u) = (1 - u)^m$.]

(c) Assuming the Bombieri-Vinogradov theorem (Theorem 18.9) holds for $Q \leqslant x^\theta$ with $\theta > 1/2$, deduce that $\liminf_{n\to\infty}(p_{n+1} - p_n) < \infty$.

(d) (Soundararajan) Integrate by parts to show that

$$\tilde\rho_k(F) = 2k\cdot \frac{\int_0^1 u^{k-1}F(u)(\int_u^1 F)\mathrm{d}u}{\int_0^1 u^{k-1}F(u)^2\mathrm{d}u} \leqslant 2k\cdot\Big(\frac{\int_0^1 u^{k-1}(\int_u^1 F)^2\mathrm{d}u}{\int_0^1 u^{k-1}F(u)^2\mathrm{d}u}\Big)^{1/2}.$$

Generalize this argument to conclude that $\tilde\rho_k(F) \leqslant 4k/\prod_{j=1}^\infty (k+2^j-2)^{1/2^j} < 4$ for all $F \neq 0$.

[3] This corresponds essentially to the definition (28.4) of the GPY weights. There, one must also assume that F is smooth enough, but this is not needed when optimizing the quantity $\rho_k(f)$.

Remark 28.9. Exercise 28.3(d) proves that the original GPY weights cannot prove that $\lim_{n\to\infty}(p_{n+1}-p_n) < \infty$ without access to an improved version of the Bombieri-Vinogradov theorem. Zhang's breakthrough was to supply this necessary improvement. In contrast, in the more general Maynard-Tao weights the quantity $\rho_k(f)$ can become arbitrarily large when $k \to \infty$, thus sidestepping the need for an improved Bombieri-Vinogradov theorem. □

Exercise 28.4 (Conrey)**:** Define $\tilde{\rho}(F)$ by (28.23) and let $\lambda = k(k-1)/\sup_F \tilde{\rho}_k(F)$.

(a) If there is $G \in C([0,1])$ such that $\tilde{\rho}(G) = k(k-1)/\lambda$, then show that $G(x) \geqslant 0$ for all $x \in [0,1]$. In addition, use calculus of variations to show that G must satisfy the integral equation

$$x^{k-1}G(x) = \lambda \int_0^x u^{k-2} \int_u^1 G(t)\mathrm{d}t\mathrm{d}u \qquad (0 \leqslant x \leqslant 1). \tag{28.24}$$

(b) Conversely, let G be a continuous and non-negative function satisfying (28.24) and whose set of roots has null measure. Show that $\tilde{\rho}_k(G) = k(k-1)/\lambda$ and that $\tilde{\rho}_k(F) \leqslant \tilde{\rho}_k(G)$ for any continuous $F : [0,1] \to \mathbb{R}$. [*Hint:* For the second part, use the Cauchy-Schwarz inequality.]

(c) Show that any continuous solution to (28.24) must be smooth on $[0,1]$ and satisfy the differential equation $xG''(x) + kG'(x) + \lambda G(x) = 0$. In addition, a solution to this differential equation that is analytic around 0 must be a multiple of the function $x \to A(\lambda x)$, where

$$A(x) = \sum_{r\geqslant 0} \frac{(-x)^r}{r!(r+k-1)!} = \frac{J_{k-1}(2\sqrt{x})}{x^{(k-1)/2}}$$

with J_{k-1} denoting the $(k-1)$th Bessel function of the first kind (see [183]).

(d) Let $A_n(x) = \sum_{r=0}^n (-x)^r/[r!(r+k-1)!]$. Show that a solution to (28.24) normalized so that $G(0) = 1/(k-1)!$ must satisfy

$$\lambda \int_0^1 G(t)\mathrm{d}t = \frac{1}{(k-2)!}. \tag{28.25}$$

In addition, for $x \in [0,1]$ and $n \in \mathbb{Z}_{\geqslant 0}$, we must have

$$G(x) - A_n(\lambda x) = -\lambda \int_0^1 u^{k-2} \int_0^{ux} \big[G(t) - A_{n-1}(\lambda t)\big]\mathrm{d}t\mathrm{d}u.$$

(e) Show that if there is a continuous and non-negative solution to (28.24) with $\lambda > 0$, then $A_{2n+1}(\lambda x) \leqslant G(x) \leqslant A_{2n}(\lambda x)$ for $x \in [0,1]$ and $n \in \mathbb{Z}_{\geqslant 0}$, and thus $G(x) = A(\lambda x)$ for $x \in [0,1]$.

(f) When $G(x) = A(\lambda x)$, show that (28.25) is equivalent to $J_{k-2}(\sqrt{2\lambda}) = 0$.

Remark 28.10. When $m \in \mathbb{N}$, it is known that J_m has infinitely many positive real zeros. If z_m denotes the smallest such zero, we also know that $z_1 < z_2 < \cdots$ and $z_m > m + \pi - 1/2$ [36,110]. Hence, if $\lambda = z_{k-2}^2/4$, we infer that $A(\lambda x) > 0$ for $x \in [0,1)$. In particular, (28.24) has a non-negative and continuous solution G for which $\sup_F \tilde{\rho}_k(F) = \tilde{\rho}_k(G) = 4k(k-1)/z_{k-2}^2 < 4$. Thus, we recover the conclusion of Exercise 28.3(d). □

Chapter 29

Large gaps between primes

In the previous chapter, we demonstrated gaps in the sequence of primes $p_1 < p_2 < \cdots$ that are much smaller than the expected size of $p_{n+1} - p_n$, which is $\log n$. We now turn to the opposite question: does the gap $p_{n+1} - p_n$ get large compared to $\log n$? The answer should be affirmative. To see why, we turn again to Cramér's model.

Recall that $(X_k)_{k=1}^{\infty}$ is a sequence of independent Bernoulli random variables such that $X_1 = 0$, $X_2 = 1$, and with $\mathbb{P}(X_k = 1) = 1/\log k$ for $k \geqslant 3$. Let P_n be the random variable that equals the nth smallest index k such that $X_k = 1$ (that is to say, P_n models the nth smallest prime number).

Proposition 29.1. *With probability* 1*, we have*

$$\limsup_{n\to\infty} \frac{P_{n+1} - P_n}{(\log n)^2} = 1.$$

The above result is a simple consequence of the Borel-Cantelli lemma from probability theory [7, Theorems 4.3 and 4.4].

Lemma 29.2 (Borel-Cantelli)**.** *Let $E_1, E_2, \ldots$ be some events in a probability space, and let E be the event that infinitely many of the E_j's occur.*

(a) *If $\sum_{j\geqslant 1} \mathbb{P}(E_j) < \infty$, then $\mathbb{P}(E) = 0$.*

(b) *If the events $E_1, E_2, \ldots$ are mutually independent and $\sum_{j\geqslant 1} \mathbb{P}(E_j) = \infty$, then $\mathbb{P}(E) = 1$.*

Proof of Proposition 29.1. For each $k \in \mathbb{N}$, $r \geqslant 0$ and $\lambda > 0$, let $E_k(r, \lambda)$ be the event that $X_j = 1$ for at most r integers $j \in (k, k + \lambda \log^2 k]$. We will use the Borel-Cantelli lemma to prove two key facts about these events:

Claim 1. If $\lambda > 1$ is fixed, then with probability 1 at most finitely many of the events $E_k(1, \lambda)$ occur.

Claim 2. If $\lambda < 1$ is fixed, then with probability 1 infinitely many of the events $E_k(0, \lambda)$ occur.

We leave it as an exercise on Cramér's model to verify how these two claims can be combined to complete the proof of the proposition.

Now, let us prove Claims 1 and 2. If we let $J_k = \mathbb{Z} \cap (k, k + \lambda \log^2 k]$, then the independence of the X_j's implies that

$$\mathbb{P}(E_k(0, \lambda)) = \prod_{j \in J_k} \Big(1 - \frac{1}{\log j}\Big) = k^{-\lambda + o(1)}$$

and

$$\mathbb{P}(E_k(1, \lambda) \setminus E_k(0, \lambda)) \leqslant \sum_{i \in J_k} \frac{1}{\log i} \prod_{j \in J_k \setminus \{i\}} \Big(1 - \frac{1}{\log j}\Big) = k^{-\lambda + o(1)}$$

as $k \to \infty$. In particular, Claim 1 follows immediately from Lemma 29.2(a).

Finally, let us fix $\lambda < 1$ and prove Claim 2. If $k_j = \lfloor j \log^3 j \rfloor$, then we may easily check that the events $E_{k_j}(0, \lambda)$ are mutually independent for large enough j, as well as that $\sum_{j \geqslant 1} \mathbb{P}(E_{k_j}(0, \lambda)) = \infty$. Thus, Claim 2 follows from Lemma 29.2(b). □

Proposition 29.1 leads us to guess that

$$\limsup_{n \to \infty} \frac{p_{n+1} - p_n}{(\log n)^2} = 1.$$

However, Granville's refinement of Cramér's model suggests the lower bound

$$\limsup_{n \to \infty} \frac{p_{n+1} - p_n}{(\log n)^2} \geqslant 2e^{-\gamma} = 1.12291\ldots$$

(see Exercise 29.2). It is not clear what the true value of this lim sup is (though see Exercise 30.1). In this chapter, we will prove a weaker result. A simple corollary of it is that the normalized gap $(p_{n+1} - p_n)/\log n$ can get arbitrarily large.

Theorem 29.3. *We have that*

$$\limsup_{n \to \infty} \frac{p_{n+1} - p_n}{L(n)} = \infty \quad \textit{with} \quad L(n) = \frac{\log n \, \log_2 n \, \log_4 n}{(\log_3 n)^2},$$

where $\log_j$ *denotes the* j*th iteration of the logarithmic function.*

This theorem was proven independently by Ford, Green, Konyagin and Tao [46], and by Maynard [139]. Here, we follow Maynard's argument which is more in tune with the ideas we have developed thus far. Ford, Green, Konyagin and Tao used a different technique, building on the work of Green and Tao on long arithmetic progressions in the sequence of primes [77].

The Erdős-Rankin construction

All constructions of long strings of composite numbers are based on the concept of a *covering system of congruences*. We say that the system of congruences $\{a_j \pmod{q_j}\}_{j=1}^{k}$ *covers* the set of integers $\mathcal{N}$ if for each $n \in \mathcal{N}$ there is some j such that $n \equiv a_j \pmod{q_j}$.

Recall the notation $P(z) = \prod_{p \leqslant z} p$. Our strategy for proving Theorem 29.3 is based on the following simple lemma.

Lemma 29.4. *Let $H \geqslant 1$ and $z \geqslant 2$. Assume that there is a system of congruences $\{a_p \pmod{p}\}_{p \leqslant z}$ that covers $\mathbb{Z} \cap [1, H]$. Then there exists some $n \in (P(z), 2P(z)]$ for which there are no primes in $(n, n + H]$.*

Proof. Let n be the unique integer in $(P(z), 2P(z)]$ satisfying the congruences $n \equiv -a_p \pmod{p}$ for all primes $p \leqslant z$. If $h \in [1, H] \cap \mathbb{Z}$, then $h \equiv a_p \pmod{p}$ for some $p \leqslant z$, that is to say, there exists a prime $p \leqslant z$ that divides $n + h$. In particular, $n + h$ cannot be a prime number. □

In preparation for our proof of Theorem 29.3, we first show the weaker result due to Rankin that

$$\limsup_{n \to \infty} \frac{p_{n+1} - p_n}{L(n)} > 0. \tag{29.1}$$

Even though the gap between (29.1) and Theorem 29.3 seems small, making this leap was a long-standing open problem due to Erdős.[1]

To prove (29.1), we fix a small constant $c > 0$ to be chosen later. In addition, we let $z \geqslant 2$ be a parameter tending to infinity, $X = P(z)$ and

$$H = \frac{cz \log z \log_3 z}{(\log_2 z)^2}, \qquad \text{whence} \qquad z \sim \frac{H(\log_2 H)^2}{c \log H \log_3 H}. \tag{29.2}$$

We will show that we can pick congruence classes $a_p \pmod{p}$ with $p \leqslant z$ covering the integers $\leqslant H$. Assuming that this is indeed possible, we can then apply Lemma 29.4 to find that $\max_{X < p_n \leqslant 2X}(p_{n+1} - p_n) \geqslant H$. Since $\log X \sim z$ by the Prime Number Theorem, we deduce that

[1] Paul Erdős had a legendary knack for asking very hard questions with deceptively simple statements. Occasionally, he would offer a monetary award for their solution. The award he offered for proving Theorem 29.3 was $10,000, the largest "Erdős prize" ever.

$$\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{L(n)} \geqslant c, \tag{29.3}$$

which establishes (29.1).

Let us now explain how to construct the classes $a_p \pmod p$. We will select them in three stages, determined by two parameters y and Y to be chosen so that $(\log H)^3 \leqslant y \leqslant Y \leqslant H/2$. Throughout, the letters p and q denote prime numbers.

Stage 1: Intermediate primes. When $p \in (y, Y]$, we select $a_p = 0$. This choice is the key to the success of the Erdős-Rankin method because it leaves uncovered few integers $\leqslant H$. Specifically, if $\mathcal{N}$ is the set of integers $n \leqslant H$ not covered by the classes $0 \pmod p$ for $p \in (y, Y]$, then either n is y-smooth, or $n = mq$ with $q > Y$ prime and $m \leqslant H/Y$. Writing $y = H^{1/u}$ and applying Theorems 16.4 and 20.1, we find that

$$\begin{aligned} |\mathcal{N}| \leqslant \Psi(H, y) + \sum_{m\leqslant H/Y} \pi(H/m) &\ll \frac{H}{u^u} + \sum_{m\leqslant H/Y} \frac{H}{m\log Y} \\ &\ll \frac{H}{u^u} + \frac{H\log(H/Y)}{\log Y}. \end{aligned}$$

Stage 2: Small primes. For the primes $p \leqslant y$, we select the progressions $a_p \pmod p$ "greedily".

We begin by letting $a_2 \pmod 2$ be any class $a \pmod 2$ maximizing the quantity $\#\{\, n \leqslant \mathcal{N} : n \equiv a \pmod 2 \,\}$. Having chosen a_2, we set $\mathcal{N}_2 = \{\, n \in \mathcal{N} : n \not\equiv a_2 \pmod 2 \,\}$ and note that $|\mathcal{N}_2| \leqslant |\mathcal{N}|/2$.

Next, we let $a_3 \pmod 3$ be any class $a \pmod 3$ maximizing the quantity $\#\{\, n \in \mathcal{N}_2 : n \equiv a \pmod 3 \,\}$. If we set $\mathcal{N}_3 = \{\, n \in \mathcal{N}_2 : n \not\equiv a_3 \pmod 3 \,\}$, then $|\mathcal{N}_3| \leqslant (1 - 1/3)|\mathcal{N}_2|$.

Continuing this way, we find that there are progressions $a_p \pmod p$ indexed by the primes $p \leqslant y$, such that the set

$$\mathcal{N}' := \{\, n \in \mathcal{N} : \exists p \leqslant y \text{ for which } n \not\equiv a_p \pmod p \,\}$$

has cardinality

$$|\mathcal{N}'| \leqslant |\mathcal{N}| \cdot \prod_{p\leqslant y} \Big(1 - \frac{1}{p}\Big) \ll \frac{H}{u^u \log y} + \frac{H\log(H/y)}{\log y \log Y}. \tag{29.4}$$

Stage 3: Large primes. Due to the nature of the second stage, we have completely lost control of the residual set $\mathcal{N}'$. Hence, we will apply a trivial argument in the third and final stage of the argument. For it to work, we

must guarantee that the remaining primes (i.e., the primes in $(Y, z]$) are more than the number of integers in $\mathcal{N}'$. We thus choose

$$y = H^{\log_3 H/(3\log_2 H)} \quad \text{and} \quad Y = \frac{z\log_3 z}{\log_2 z} \sim \frac{H\log_2 H}{c\log H}. \tag{29.5}$$

Indeed, with this choice of parameters, we have $u^u \log y = (\log H)^{4+o(1)}$, as well as that $H\log(H/Y)/(\log y \log Y) \asymp H(\log_2 H)^2/(\log H \log_3 H)$. Applying (29.4), and taking c small enough and z large enough in (29.2), implies that $|\mathcal{N}'| \leqslant z/(3\log z) < \#\{Y < p \leqslant z\}$.

Since there are more primes than integers left to cover, we can easily complete the proof: if $q_1, \dots, q_k$ are the primes in $(Y, z]$ and $n_1, \dots, n_\ell$ are the integers in $\mathcal{N}'$, we let $a_{q_j} = n_j$ for each $j \leqslant \ell$, and choose a_{q_j} for $\ell < j \leqslant k$ arbitrarily. This concludes the construction of the claimed covering system of congruences, and thus the proof of (29.1).

A more efficient covering system

We now turn to the proof of Theorem 29.3. As before, we wish to find a system of congruences $\{a_p \pmod p\}_{p\leqslant z}$ that covers $[1, H]$, but with c being arbitrarily large. In view of (29.3), this suffices to prove Theorem 29.3.

To find this more efficient covering system, we will improve upon Stage 3. Specifically, we will show that it is possible to choose the a_p's for $p \in (z/2, z]$ in a way that each congruence class covers many elements of the residual set $\mathcal{N}'$ and not just one.

Throughout, y and Y are defined by (29.5). However, the first stage of the selection of the covering system has an extra auxiliary part that deals with very small primes and helps simplify the situation in the last two stages.

Stage 1a: Intermediate primes. We again choose $a_p = 0$ for the primes $p \in (y, Y]$. We are then left with integers $n \leqslant H$ such that either n is y-smooth or $n = qm$ with $q > Y$ prime and $m \leqslant H/Y \leqslant \log z$.

Stage 1b: Very small primes. We also select $a_p = 0$ when $p \leqslant \log z$. The effect of this auxiliary stage is a simplification of the residual set, which now equals $\{\, n \leqslant H : p|n \ \Rightarrow\ \log z < p \leqslant y \,\} \cup \{Y < q \leqslant H\}$. Its first component has small size by Theorem 16.4. We will cover it trivially at the end of Stage 3. We thus focus on the set of primes in $(Y, H]$.

Stage 2: Small primes. Next, we choose a_p for $p \in (\log z, y]$. In the previous section, we selected these congruence classes greedily. Here, we simply take $a_p = 1$. This has essentially the same effect as choosing the a_p's

greedily. The reason is that $q-1$ looks a lot like a "random" integer, so the chance that it has no prime factors in

$$\mathcal{P} := \{\log z < p \leqslant y\}$$

is about $\prod_{p\in\mathcal{P}}(1-1/p)$. The advantage of having $a_p = 1$ for all $p \in \mathcal{P}$ is that it allows for an explicit description of the residual set: indeed, after Stage 2, we are left with certain y-smooth integers $n \leqslant H$ and with the set of primes $\{\, Y < q \leqslant H : (q-1, \mathcal{P}) = 1 \,\}$. We must cover them using congruence classes $a_p \pmod p$ with $p \in (Y, z]$.

Stage 3a: Large primes. This stage will be the most delicate and will take most of the remaining chapter to be completed. We summarize it in the following proposition whose proof is postponed till the next section.

Proposition 29.5. *Fix $\varepsilon > 0$. Let z, H and y be as above. There is a choice of congruence classes $a_p \pmod p$, $p \in (z/2, z]$, covering $\geqslant (100-\varepsilon)\%$ of the set of primes $\mathcal{Q} := \{\, z < q \leqslant H : (q-1, \mathcal{P}) = 1 \,\}$.*

Assuming the above result for now, let us see how to use it to complete the proof of Theorem 29.3.

Stage 3b: Large primes – cleaning up. Let $\mathcal{Q}$ be the set of primes defined in Proposition 29.5. A simple application of Theorem 18.11(a) (with the required level of distribution supplied by the Bombieri-Vinogradov theorem) implies that

$$|\mathcal{Q}| \sim \frac{H}{\log H} \prod_{p\in\mathcal{P}} \left(1 - \frac{1}{p-1}\right) \sim \frac{3cz}{\log z}.$$

Now, from the above discussion, we know there are congruence classes $a_p \pmod p$ with $p \in [1, Y] \cup (z/2, z]$ that cover all of $[1, H] \cap \mathbb{Z}$, except perhaps the set $\mathcal{R} := \{\, n \leqslant H : P^+(n) \leqslant y \,\} \cup \mathcal{Q}_1 \cup \mathcal{Q}_2$, where $\mathcal{Q}_1 = \{\, q \leqslant z : (q-1, \mathcal{P}) = 1 \,\}$ and $\mathcal{Q}_2 \subseteq \mathcal{Q}$ has $\leqslant |\mathcal{Q}|/(4c) \sim z/(4\log z)$ elements. Since $\Psi(H, y) \leqslant H/(\log H)^{3+o(1)}$ by Theorem 16.4, and $|\mathcal{Q}_1| = o(z/\log z)$ by Theorem 18.11(a) (see also Exercise 20.3(a)), we conclude that $|\mathcal{R}| \leqslant z/(3\log z) < \#\{Y < p \leqslant z/2\}$ as long as z is large enough. Hence, arguing as in Stage 3 of the previous section, we may trivially cover $\mathcal{R}$ using residues $a_p \pmod p$ with $p \in (Y, z/2]$. This completes the proof of Theorem 29.3.

Random covering systems of congruences

We now turn to the proof of Proposition 29.5. The main idea is to construct for each $p \in (z/2, z]$ a probability measure δ_p on $\mathbb{Z}/p\mathbb{Z}$ such that

$$\sum_{z/2<p\leqslant z} \delta_p(q) \geqslant -\log(\varepsilon/100) \quad \text{for all primes } q \in \mathcal{Q}. \tag{29.6}$$

Before explaining how to do this, let us see how such a construction yields Proposition 29.5.

Naturally, the measures δ_p with $p \in (z/2, z]$ induce a product measure δ on the space $G := \prod_{z/2<p\leqslant z} \mathbb{Z}/p\mathbb{Z}$ by taking

$$\delta(\mathbf{a}) := \prod_{z/2<p\leqslant z} \delta_p(a_p)$$

for each $\mathbf{a} = (a_p)_{z/2<p\leqslant z} \in G$. Given any $q \in \mathcal{Q}$, the probability that it is not covered by a random tuple $\mathbf{a} \in G$ equals $\prod_{z/2<p\leqslant z}(1-\delta_p(q))$. Thus

$$\mathbb{E}_{\mathbf{a}\in G}\Big[\#\Big\{ q \leqslant Q : q \notin \bigcup_{z/2<p\leqslant z} a_p \,(\mathrm{mod}\, p) \Big\}\Big] = \sum_{q\leqslant Q} \prod_{z/2<p\leqslant z} (1-\delta_p(q)).$$

Hence, (29.6) and the inequality $1-x \leqslant e^{-x}$ imply that the right-hand side is $\leqslant \varepsilon|\mathcal{Q}|/100$. In particular, there must exist a choice of $\mathbf{a} \in G$ such that the number of $q \in \mathcal{Q}$ not covered by $\mathbf{a}$ is $\leqslant \varepsilon|\mathcal{Q}|/100$, which is precisely what we claimed in Proposition 29.5.

To construct measures δ_p satisfying (29.6), we go back to the ideas of Chapter 28: we set

$$s_1 = 0 \qquad \text{and} \qquad s_j = p_{\pi(C_k)+j-1} \prod_{p\leqslant C_k} p \qquad \text{for } j = 2, \dots, k,$$

where C_k is an auxiliary integer that will be taken to be large enough and, as usual, p_n denotes the nth prime. In particular, if $C_k > k$, then the k-tuple $(s_1, \dots, s_k)$ is admissible. We also set

$$\nu(d) = \#\{\, n \,(\mathrm{mod}\, d) : (n-s_1)\cdots(n-s_k) \equiv 0 \,(\mathrm{mod}\, d)\,\}.$$

Next, we consider two upper bound sieve weights. We let μ_1^+ be the function λ^+ supplied by Theorem 19.1 when applied with $\kappa = k$, set of primes $\{p \leqslant \log z\}$ and level of distribution $D_1 = y^{\log_3 H}$. We also let μ_2^+ be the function λ^+ from Theorem 19.1 applied with $\kappa = 2k$, set of primes $\mathcal{P} = \{\log z < p \leqslant y\}$ and level of distribution $D_2 = y^{\log_3 H}$.

In addition, we let $\lambda : \mathbb{N}^k \to \mathbb{R}$ be the function constructed in Chapter 28 with z in place of N. In particular, λ is supported on the set

$$\mathscr{D} := \{\, \mathbf{d} \in \mathbb{N}^k : d_1 \cdots d_k \leqslant D,\, P^-(d_j) > y \ \forall j \,\} \quad \text{with} \quad D = z^{1/4} e^{-\sqrt{\log z}}$$

and it has bounded supremum norm. Note that $\log D/\log y \asymp \log_2 H/\log_3 H$. Hence, if ξ is as in Lemma 28.3, ζ_ℓ is as in Lemma 28.4, and $I_\ell(f)$ and $J(f)$ are as is in Lemma 28.6, then λ satisfies the asymptotic estimates

$$\sum_{(a_1,\dots,a_k)\in\mathscr{D}} \frac{\xi(a_1,\dots,a_k)^2}{a_1\cdots a_k} = \frac{J(f) + O((\log_3 H)^2/\log_2 H)}{(\log D)^k \prod_{p\leqslant y}(1-1/p)^k} \tag{29.7}$$

and

$$\sum_{\substack{(a_1,\dots,a_k)\in\mathscr{D}\\ a_\ell=1}} \frac{\zeta_\ell(a_1,\dots,a_k)^2}{a_1\cdots a_k} = \frac{I_\ell(f)+O((\log_3 H)^2/\log_2 H)}{(\log D)^{k-1}\prod_{p\leqslant y}(1-1/p)^{k-1}} \tag{29.8}$$

for $\ell = 1,\dots,k$ by adapting the proof of Lemma 28.6 (the only difference is that we must apply Lemma 28.7 with $w = y$ and $u \asymp \log_2 H/\log_3 H$).

Finally, we let

$$Q(n,n') = \prod_{1\leqslant j\leqslant k}(n - s_j n'),$$

which is the homogeneous version of the polynomial $\prod_{j=1}^k (x-s_j)$, and we introduce the sieve weights

$$w_{n,p} = \Big(\sum_{a\mid Q(n,p)} \mu_1^+(a)\Big)\Big(\sum_{b\mid Q(n,p)Q(n-1,p)} \mu_2^+(b)\Big)\Big(\sum_{\substack{d_j\mid n-ps_j,\, p\nmid d_j\\ 1\leqslant j\leqslant k}} \lambda(\mathbf{d})\Big)^2.$$

The probability measure δ_p is then defined by

$$\delta_p(a) := \frac{\sum_{n\leqslant 2H,\, n\equiv a\,(\mathrm{mod}\,p)} w_{n,p}}{\sum_{n\leqslant 2H} w_{n,p}}$$

for each $a \in \mathbb{Z}/p\mathbb{Z}$. It is designed to be biased towards the progressions $a\,(\mathrm{mod}\,p)$ containing many elements of $\mathcal{Q}$. In particular, note that if $n = q + ps_\ell$ with $q \in \mathcal{Q}$, then the sum over $(d_1,\dots,d_k)$ in the definition of λ can be restricted to those k-tuples with $d_\ell = 1$, analogously to the situation in Lemma 28.4.

Because of this nature of the weights δ_p, we ignore all summands but those of the form $q + ps_\ell$ in the numerator. This yields the inequality $\delta_p(q) \geqslant \sum_{\ell=1}^k w_{q+ps_\ell}/\sum_{n\leqslant 2H} w_{n,p}$, whence

$$\sum_{z/2<p\leqslant z} \delta_p(q) \geqslant \sum_{\ell=1}^k \frac{\sum_{z/2<p\leqslant z} w_{q+ps_\ell,p}}{\sum_{n\leqslant 2H} w_{n,p}}. \tag{29.9}$$

Calculations

The next step is to estimate the right-hand side of (29.9) by adapting the methods of Chapter 28. This is accomplished in Lemmas 29.6 and 29.7 below. A very good (albeit tough) exercise on the methods on Chapter 28 is to demonstrate these lemmas without consulting their proofs.

All implied constants from now on might depend on k.

Lemma 29.6. *Assume the above notation and let $p_0 \in (z/2, z]$. Then*

$$\sum_{n \leqslant 2H} w_{n,p_0} = \frac{2VH \cdot \big(J(f) + O(1/\sqrt{\log_2 H})\big)}{(\log D)^k \prod_{p \leqslant y}(1-1/p)^k},$$

where $V = (\log_2 z/\log y)^{2k} \prod_{p \leqslant \log z}(1-\nu(p)/p)$.

Proof. Let $\mathbf{d}, \mathbf{e} \in \mathscr{D}$ be such that $d_j, e_j | n + p_0 s_j$ and $p_0 \nmid d_j e_j$ for each j. Arguing as in the beginning of the proof of Lemma 28.3, we find that $(d_i e_i, d_j e_j) = 1$ for all $i \neq j$, provided that $y \geqslant s_k - s_1 = s_k$. Therefore

$$\sum_{n \leqslant 2H} w_{n,p_0} = \sum_{\substack{a,b,\mathbf{d},\mathbf{e} \\ (d_i e_i, p_0 d_j e_j)=1 \, \forall i \neq j}} \mu_1^+(a)\mu_2^+(b)\lambda(\mathbf{d})\lambda(\mathbf{e}) N(a,b,\mathbf{d},\mathbf{e}),$$

where $N(a,b,\mathbf{d},\mathbf{e})$ denotes the cardinality of integers $n \in [1, 2H]$ such that $a | Q(n,p_0)$, $b | Q(n,p_0)Q(n-1,p_0)$ and $[d_j, e_j] | n - p_0 s_j$ for all j.

By our assumptions on the support of μ_1^+, μ_2^+ and λ, the numbers a, b and $[d_1, e_1], \dots, [d_k, e_k]$ can be assumed to be mutually coprime. The Chinese Remainder Theorem then implies that

$$N(a,b,\mathbf{d},\mathbf{e}) = 2H \cdot \frac{\nu_1(a)\nu_2(b)}{ab \prod_{j=1}^k [d_j, e_j]} + O(\nu_1(a)\nu_2(b)),$$

where $\nu_1(a)$ counts the number of solutions $n \in \mathbb{Z}/a\mathbb{Z}$ to the congruence $Q(n,p_0) \equiv 0 \,(\mathrm{mod}\, a)$ and, similarly, $\nu_2(b)$ counts the number of solutions $n \in \mathbb{Z}/b\mathbb{Z}$ to the congruence $Q(n,p_0)Q(n-1,p_0) \equiv 0 \,(\mathrm{mod}\, b)$. (In particular, these functions might depend on p_0.) Consequently,

$$\sum_{n \leqslant 2H} w_{n,p_0} = V_1 V_2 H S + O(H^{1/10}),$$

where

$$V_j = \sum_d \frac{\mu_j^+(d)\nu_j(d)}{d} \qquad \text{and} \qquad S = \sum_{\substack{\mathbf{d},\mathbf{e} \\ (d_i e_i, p_0 d_j e_j)=1 \, \forall i \neq j}} \frac{\lambda(\mathbf{d})\lambda(\mathbf{e})}{\prod_{j=1}^k [d_j, e_j]}.$$

We now estimate S. Firstly, we remove the condition from its summands that $p_0 \nmid d_1 e_1 \cdots d_k e_k$. The error produced is

$$\leqslant \sum_{j=1}^k \sum_{\substack{\mathbf{d},\mathbf{e} \in \mathscr{D} \\ p_0 | d_j e_j}} \frac{|\lambda(\mathbf{d})\lambda(\mathbf{e})|}{\prod_{j=1}^k [d_j, e_j]} \ll \frac{(\log H)^{3k}}{p_0} \asymp \frac{(\log H)^{3k}}{z}.$$

The remaining sum over $\mathbf{d}$ and $\mathbf{e}$ is estimated as in Lemma 28.3, as well as using (29.7). In conclusion,

$$S = \sum_{\mathbf{a}\in\mathscr{D}} \frac{\xi(\mathbf{a})^2}{a_1\cdots a_k} + O\biggl(\frac{(\log H)^{3k}}{y}\biggr) = \frac{J(f) + O((\log_3 H)^2/\log_2 H)}{(\log D)^k \prod_{p\leqslant y}(1-1/p)^k}.$$

Finally, we estimate V_1 and V_2 using Theorem 19.1. If we let $\varepsilon = 1/\log_2 H$, then we have

$$V_1 = (1+O(\varepsilon)) \prod_{p\leqslant \log z} \biggl(1-\frac{\nu_1(p)}{p}\biggr), \quad V_2 = (1+O(\varepsilon)) \prod_{\log z<p\leqslant y} \biggl(1-\frac{\nu_2(p)}{p}\biggr).$$

Since $p \neq p_0$ when $p \leqslant \log z$, we have $\nu_1(p) = \nu(p)$. On the other hand, we have $\nu_2(p) = 2k$, unless $p|p_0(s_i-s_j)(p_0(s_i-s_j)+1)$ for some $i \neq j$. There are $O(\log z/\log_2 z)$ such p by Exercise 2.8(c). Therefore,

$$V_2 = (1+O(\varepsilon))(1+O(1/\log z))^{O(\log z/\log_2 z)} \prod_{\log z<p\leqslant y} (1-2k/p)$$

$$= (1+O(\varepsilon))(\log_2 z/\log y)^{2k}, \tag{29.10}$$

where we used Mertens' third estimate (Theorem 3.4(c)). This completes the proof of the lemma. □

Lemma 29.7. *For each $\ell \in \{1,\dots,k\}$ and each $q_0 \in \mathcal{Q}$, we have*

$$\sum_{z/2<p\leqslant z} w_{q_0+ps_\ell,p} = \frac{Vz\log y}{8\log_2 z} \cdot \frac{I_\ell(f) + O(1/\sqrt{\log_2 H})}{(\log D)^k \prod_{p\leqslant y}(1-1/p)^k}$$

with V as in Lemma 29.6.

Proof. To ease the notation, we consider the case $\ell = 1$, in which case $s_1 = 0$. The other values of ℓ are treated similarly.

Since $q_0 > z$, any integer $d_1 \leqslant D$ that divides q_0 must equal 1. Hence,

$$w_{q_0,p} = \biggl(\sum_{a|G_0(p)} \mu_1^+(a)\biggr)\biggl(\sum_{b|G_0(p)G_1(p)} \mu_2^+(b)\biggr)\biggl(\sum_{\substack{d_j|q_0-s_jp,\, p\nmid d_j\\ d_1=1}} \lambda(\mathbf{d})\biggr)^2,$$

where we have set

$$G_h(n) = \prod_{2\leqslant j\leqslant k} (q_0 - h - s_j n).$$

As in the proof of Lemma 29.6, we have

$$\sum_{z/2<p\leqslant z} w_{q_0,p} = \sum_{\substack{a,b,\mathbf{d},\mathbf{e},\ d_1=e_1=1\\ (d_ie_i,d_je_j)=1\ \forall i\neq j}} \mu_1^+(a)\mu_2^+(b)\lambda(\mathbf{d})\lambda(\mathbf{e}) \cdot P(a,b,\mathbf{d},\mathbf{e}),$$

where $P(a,b,\mathbf{d},\mathbf{e})$ counts the number of primes $p \in (z/2,z]$ such that $p \nmid d_1e_1\cdots d_ke_k$, $a|G_0(p)$, $b|G_0(p)G_1(p)$ and $[d_j,e_j]|q_0-s_jp$ for all $j \geqslant 2$. The

number of primes p dividing one of $d_1,\ e_1,\ \dots,\ d_k,\ e_k$ is $O(\log z)$. Hence, adapting the argument leading to (28.14) implies that

$$P(a,b,\mathbf{d},\mathbf{e}) = \frac{\nu_1^*(a)\nu_2^*(b)X}{\varphi(r)} + O(\nu_1^*(a)\nu_2^*(b)E(z;r) + \log z),$$

where $r = ab\prod_{j=2}^k [d_j, e_j]$, the functions $\nu_1^*(m)$ and $\nu_2^*(m)$ count the number of solutions $t \in (\mathbb{Z}/m\mathbb{Z})^*$ to the congruences $G_0(t) \equiv 0 \,(\mathrm{mod}\, m)$ and $G_0(t)G_1(t) \equiv 0 \,(\mathrm{mod}\, m)$, respectively,

$$X = \mathrm{li}(z) - \mathrm{li}(z/2) = z/(2\log x) + O(z/(\log z)^2),$$

and

$$E(z;r) = \max_{t\in(\mathbb{Z}/r\mathbb{Z})^*} \big|\pi(z;r,t) - \pi(z/2;r,t) - X/\varphi(r)\big|.$$

Together with the Bombieri-Vinogradov theorem, this implies that

$$\sum_{z/2<p\leqslant z} w_{q+ps_\ell,p} = V_1^* V_2^* XT + O_A(z/(\log z)^A)$$

for any fixed $A > 0$, where

$$V_j^* = \sum_m \frac{\mu_j^+(m)\nu_j^*(m)}{\varphi(m)} \qquad \text{and} \qquad T = \sum_{\substack{\mathbf{d},\mathbf{e}\in\mathscr{D},\ d_1=e_1=1\\ (d_ie_i,d_je_j)=1\ \forall i\neq j}} \frac{\lambda(\mathbf{d})\lambda(\mathbf{e})}{\prod_{j=2}^k \varphi([d_j,e_j])}.$$

Next, we estimate the sum over $\mathbf{d}$ and $\mathbf{e}$ exactly as in the proof of Lemma 28.4, and then use (29.8). This yields that

$$T = \sum_{\substack{\mathbf{a}\in\mathscr{D}\\ a_1=1}} \frac{\zeta_1(\mathbf{a})^2}{a_1\cdots a_k} + O\bigg(\frac{(\log H)^{3k}}{y}\bigg) = \frac{I_1(f) + O((\log_3 H)^2/\log_2 H)}{(\log D)^{k-1}\prod_{p\leqslant y}(1-1/p)^{k-1}}.$$

As a consequence,

$$\sum_{z/2<p\leqslant z} w_{q+ps_\ell,p} = zV_1^*V_2^* \cdot \frac{I_1(f) + O((\log_3 H)^2/\log_2 H)}{8(\log D)^k\prod_{p\leqslant y}(1-1/p)^{k-1}}.$$

Finally, we apply Theorem 19.1 to deduce that

$$V_1^* = (1+O(\varepsilon))\prod_{p\leqslant \log z}\bigg(1-\frac{\nu_1^*(p)}{p-1}\bigg),\ \ V_2^* = (1+O(\varepsilon))\prod_{\log z<p\leqslant y}\bigg(1-\frac{\nu_2^*(p)}{p-1}\bigg),$$

where $\varepsilon = 1/\log_2 H$. We note that $\nu_1^*(p) = \nu(p) - 1$ for all $p \leqslant \log z < q_0$. Indeed, $\nu_1^*(p)$ equals $\#\{\, n \in (\mathbb{Z}/p\mathbb{Z})^* : \exists s_j \not\equiv 0 \,(\mathrm{mod}\, p),\ s_j n \equiv q_0 \,(\mathrm{mod}\, p)\,\}$, which also equals $\nu(p) - 1$ because $s_1 = 0$. We thus conclude that

$$V_1^* = \prod_{p\leqslant \log z}\bigg(1-\frac{\nu(p)}{p}\bigg)\bigg(1-\frac{1}{p}\bigg)^{-1}.$$

Moreover, we have $\nu_2^*(p) = 2k-2$, unless $p|q_0(q_0-1)s_1\cdots s_k$ or $p|(s_i - s_j)(q_0(s_i-s_j)+s_j)$ for some $i \neq j$. Arguing as in (29.10), we find that

$$V_2^* = (1+O(1/\log_2 H))\Big(\frac{\log_2 z}{\log y}\Big)^{2k-1} \prod_{\log z < p \leqslant y} \Big(1-\frac{1}{p}\Big)^{-1}.$$

This completes the proof of the lemma. □

We are now ready to prove (29.6). We estimate the right-hand side of (29.9) using Lemmas 29.6 and 29.7. Assuming that $J(f) \gg 1$, we infer that

$$\sum_{z/2<p\leqslant z} \delta_p(q) \geqslant (1+o(1))\frac{z\log y}{16H\log_2 z}\sum_{\ell=1}^{k}\frac{I_\ell(f)}{J(f)} \sim \frac{1}{48c}\sum_{\ell=1}^{k}\frac{I_\ell(f)}{J(f)}$$

for all $q \in \mathcal{Q}$ as $z \to \infty$. Choosing the function f as in Chapter 28, we find that the sum over ℓ is $\geqslant \log k - 4\log\log k + O(1)$. Taking k to be large enough in terms of c and ε completes the proof of (29.6), and hence of Theorem 29.3.

Exercises

Exercise 29.1. For each fixed $\lambda > 0$, it is believed that

$$\#\{\, p_n \leqslant x : p_{n+1} - p_n > \lambda \log x \,\} \sim \frac{e^{-\lambda}x}{\log x} \qquad (x \to \infty).$$

Use Cramér's model to give evidence in support of this conjecture.

Exercise 29.2 (Granville [64])**.** Let $(Y_n)_{n=1}^{\infty}$ be the Cramér-Granville model of parameter y, as defined in (17.13). Let $E_k(\lambda)$ be the event that $Y_j = 0$ for all integers $j \in (k, k+\lambda\log^2 k]$. In addition, let $M = P(y)^{\lfloor \log\log P(y)\rfloor}$.

(a) Let $\lambda > 0$ be fixed. Prove that

$$\mathbb{P}(E_k(\lambda)) = k^{-\lambda e^\gamma/2+o(1)}$$

for $k = P(y), 2P(y), \dots, MP(y)$ as $y \to \infty$.

(b) If $0 < \lambda < 2e^{-\gamma}$, show that the probability that none of the events of part (a) occurs is $o(1)$ as $y \to \infty$.

Chapter 30

Irregularities in the distribution of primes

So far we concentrated our efforts on proving that the primes behave in the "expected way". In this last chapter, we will show that their distribution has subtle irregularities that can be seen when zooming in on certain short intervals.

As we discussed in Chapter 29, we expect that $p_{n+1} - p_n = O(\log^2 n)$. Therefore, it seems reasonable to expect that the interval $(x, x+y]$ contains the expected number of primes, namely $\sim y/\log x$, as soon as $y \geqslant (\log x)^{2+\varepsilon}$. Assuming the validity of the Riemann Hypothesis, Selberg [161] proved in 1943 that this is indeed true for *almost all* x.

Theorem 30.1 (Selberg)**.** *Assume that the Riemann Hypothesis is true. Fix $\varepsilon > 0$ and $\delta > 0$. For all but $o_{X\to\infty}(X)$ integers $x \in [2, X]$, we have*

$$(1-\varepsilon)\frac{y}{\log x} \leqslant \pi(x+y) - \pi(x) \leqslant (1+\varepsilon)\frac{y}{\log x} \quad \textit{with} \quad y = (\log x)^{2+\delta}.$$

However, in 1985 Maier [134] arrived at the groundbreaking conclusion that the asymptotic formula $\pi(x+y) - \pi(x) \sim y/\log x$ fails infinitely often for y a fixed but arbitrarily large power of $\log x$.

Theorem 30.2 (Maier)**.** *For every fixed $C > 1$, we have that*

$$\liminf_{x\to\infty} \frac{\pi(x + (\log x)^C) - \pi(x)}{(\log x)^{C-1}} < 1 < \limsup_{x\to\infty} \frac{\pi(x + (\log x)^C) - \pi(x)}{(\log x)^{C-1}}.$$

The goal of this chapter is to establish Maier's theorem. On the contrary, we will not show Selberg's theorem because its proof lies beyond the scope of this book.

Maier matrices

The starting point for proving Theorem 30.2 is the observation we made in Chapter 17 that Cramér's model has to be adjusted by presieving the integers under consideration with small primes. Indeed, if $(\log x)^2 < h \leqslant x$, then the more accurate Cramér-Granville model suggests that[1]

$$\#\{x < p \leqslant x+h\} \sim \frac{\#\{\, x < n \leqslant x+h : (n, P(w)) = 1 \,\}}{\log x} \cdot \prod_{p \leqslant w} \Big(1 - \frac{1}{p}\Big)$$
$$\sim \frac{e^\gamma \log w}{\log x} \cdot \#\{\, x < n \leqslant x+h : (n, P(w)) = 1 \,\}$$

with w a slowly growing function of x. We strategically choose $x = P(w)$. In this case, we have $w \sim \log x$. In addition, the Chinese Remainder Theorem implies that the number of integers in $[x+1, x+h]$ that are coprime to $P(w)$ is exactly equal to the number of integers in $[1, h]$ that are coprime to $P(w)$. In particular, if $h = (\log x)^u$ with u, then

$$\#\{\, x < n \leqslant x+h : (n, P(w)) = 1 \,\} \sim \frac{h \cdot B(u)}{\log w}$$

as $x \to \infty$, where B is Buchstab's function. Putting everything together, we arrive at the guess that

$$\#\{x < p \leqslant x+h\} \sim e^\gamma B(u) \cdot \frac{h}{\log x}.$$

However, as we saw in Exercise 14.11, the difference $B(u) - e^{-\gamma}$ changes signs infinitely often (but with the amplitude of its oscillations tending to 0). Hence, we may choose u arbitrarily large such that the number of primes in $(x, x+h]$ is a bit larger than expected. Similarly, we can also find arbitrarily large u for which the number of primes in $(x, x+h]$ is smaller than expected.

In order to make the above heuristic rigorous, Maier averaged over many intervals $(x, x+h]$ with x a multiple of $P(w)$. It is convenient to display these intervals in the form of a matrix. To this end, given positive integers k, ℓ, q and h, with $\ell \geqslant 2k$ and $h \leqslant q$, we define the *Maier matrix*

$$\mathcal{M}(k, \ell; q, h) := \begin{pmatrix} 1+(k+1)q & 2+(k+1)q & \cdots & h+(k+1)q \\ 1+(k+2)q & 2+(k+2)q & \cdots & h+(k+2)q \\ \vdots & \vdots & & \vdots \\ 1+\ell q & 2+\ell q & \cdots & h+\ell q \end{pmatrix}.$$

We will eventually take $q = P(w)$ for some convenient choice of w.

[1] We have presented matters in reverse chronological order. Granville's modification of Cramér's model was inspired by Maier's work on irregularities in the distribution of primes.

Note that the ith row of $\mathcal{M}(k,\ell;q,h)$ contains all integers in the short interval $((k+i)q, h+(k+i)q]$, whereas its jth column contains all integers in the arithmetic progression

$$\{\, n \equiv j \,(\mathrm{mod}\, q) : j + kq < n \leqslant j + \ell q \,\}. \tag{30.1}$$

Now, if each arithmetic progression of the form (30.1) contains the expected number of primes, and we write $p \in \mathcal{M}(k,\ell;q,h)$ to denote that p appears among the entries of $\mathcal{M}(k,\ell;q,h)$, then

$$\#\{p \in \mathcal{M}(k,\ell;q,h)\} \sim \sum_{\substack{1\leqslant j\leqslant h\\(j,q)=1}} \frac{1}{\varphi(q)} \int_{j+kq}^{j+\ell q} \frac{\mathrm{d}t}{\log t} \sim \frac{(\ell-k)q}{\varphi(q)\log(\ell q)} \sum_{\substack{1\leqslant j\leqslant h\\(j,q)=1}} 1.$$

On the other hand, if all short intervals $(mq, h+mq]$ with $m \in (k,\ell]$ contain the expected proportion of primes, then

$$\#\{p \in \mathcal{M}(k,\ell;q,h)\} \sim \sum_{k<m\leqslant\ell} \frac{h}{\log(mq)} \sim \frac{(\ell-k)h}{\log(q\ell)}.$$

Comparing the above estimates, we see that if we can find a sequence of q and h going to infinity in a way that

$$\frac{\#\{\, 1 \leqslant j \leqslant h : (j,q) = 1 \,\}}{hq/\varphi(q)} \geqslant c > 1, \tag{30.2}$$

then we obtain a contradiction. A similar conclusion holds if the left-hand side can be made $\leqslant c' < 1$ for an infinite sequence of q and h.

Calibrating the parameters

As we explained above, we will let $q = P(w)$, so that $w \sim \log q$ and $q/\phi(q) \sim e^\gamma \log w$, as well as $h = w^u$ with u to be chosen later in a way that the difference $B(u) - e^{-\gamma}$ has a predetermined sign. In order to deduce Theorem 30.2, we need to be able to show that each arithmetic progression of the form (30.1) contains the expected number of primes. We will take $k = q^2$ and $\ell = q^L$ for some large L. Theorem 12.1 cannot be used in this range. However, we can apply Theorem 27.1 (so we will eventually weigh primes logarithmically). Firstly, we claim that we may choose w in such a way that rules out the existence of the exceptional character χ_1 for an infinite sequence of moduli of the form $q = P(w)$.

Lemma 30.3. *There are infinitely many $w \in \mathbb{N}$ such that if $q = P(w)$, then*

$$\sum_{\substack{y<p\leqslant z\\ p\equiv a\,(\mathrm{mod}\, q)}} \frac{1}{p} = \frac{\log(\log z/\log y) + O(1)}{\varphi(q)}$$

uniformly for $z \geqslant y \geqslant q^3$ and $a \in (\mathbb{Z}/q\mathbb{Z})^*$.

Proof. For each $q \geqslant 3$, we let $\mathcal{R}_q$ denote the set of real non-principal characters. In view of Theorem 27.1, it suffices to show we can choose infinitely many values of $q = P(w)$ such that

$$\sum_{y<p\leqslant z} \frac{\chi(p)}{p} = O(1) \quad (z \geqslant y \geqslant q,\ \chi \in \mathcal{R}_q). \tag{30.3}$$

Recall the definition of the sifted L-function $L_y(s,\chi)$ from Chapter 22. If $Q_\chi = q^{1/\min\{1,L_q(1,\chi)\}}$, then Theorem 22.5 implies that

$$\sum_{y<p\leqslant z} \frac{\chi(p)}{p} = O(1) \quad (z \geqslant y \geqslant Q_\chi,\ \chi \in \mathcal{R}_q). \tag{30.4}$$

Now, let $w \in \mathbb{N}$. We will show that either (30.3) holds for $q = P(w)$, or we can find $w' \geqslant w$ such that (30.3) holds for $q' = P(w')$.

Fix an auxiliary large constant M to be chosen later, and let $q = P(w)$. Firstly, we consider the case when $L_q(1,\chi) \geqslant 1/M^2$ for all $\chi \in \mathcal{R}_q$. Since $\sum_{q<p\leqslant q^{M^2}} 1/p = O(\log M)$, (30.3) follows from (30.4) in this case, with the implicit constant depending on M.

Assume now there is some $\chi_1 \in \mathcal{R}_q$ such that $L_q(1,\chi_1) \leqslant 1/M^2$. We will show that if M is sufficiently large, then we may construct another modulus $q' = P(w')$ satisfying (30.3). Precisely, we take $w' = M^{-1}\log Q_{\chi_1}$, so that $\log q' \sim M^{-1}\log Q_{\chi_1}$. Note that $w' < \infty$ by Theorem 12.8.

Consider $\chi \in \mathcal{R}_{q'}$. If χ is induced by χ_1, then the fact that $\log q' \sim M^{-1}\log Q_{\chi_1}$ and (30.4) imply that

$$\sum_{y<p\leqslant z} \frac{\chi(p)}{p} = \sum_{\max\{Q_{\chi_1},y\}<p\leqslant z} \frac{\chi_1(p)}{p} + O(\log M) = O_M(1) \qquad (z \geqslant y \geqslant q').$$

Assume now that χ is not induced by χ_1. We then know from Theorem 22.6(b) that there is an absolute constant $c > 0$ such that

$$\max\{L_{q'}(1,\chi), L_{q'}(1,\chi_1)\} \geqslant c.$$

On the other hand, we claim that if M is large enough, then $L_{q'}(1,\chi_1) < c$.

Indeed, for all $\sigma = 1 + 1/\log x$ with $x \geqslant Q_{\chi_1}$, we have

$$\log L_{q'}(\sigma,\chi) = \sum_{q'<p\leqslant x} \frac{\chi_1(p)}{p} + O(1)$$

by Lemma 22.3. Applying (22.16) from Theorem 22.5, followed by (22.15), we can rewrite the right side of the above formula as

$$\log L_{q'}(\sigma,\chi) = \sum_{q'<p\leqslant Q_{\chi_1}} \frac{\chi_1(p)}{p} + O(1) = -\sum_{q'<p\leqslant Q_{\chi_1}} \frac{1}{p} + O(1).$$

Hence, Mertens estimate implies that $\log L_{q'}(\sigma,\chi) = -\log M + O(1)$. Letting $x \to \infty$ implies that $L_{q'}(1,\chi_1) \ll 1/M$, so that $L_{q'}(1,\chi_1) < c$ by choosing a sufficiently large M. Hence, for this choice of M, we have $L_{q'}(1,\chi) \geqslant c > 0$. Combining this relation with (30.4) proves that (30.3) holds for χ in this case too. This completes the proof of the lemma. □

Proof of Theorem 30.2. Let $q = P(w)$ be as in Lemma 30.3. In addition, let $h = w^u$, and let L be a large constant to be chosen later. For brevity, we write $\mathcal{M}$ to denote the Maier matrix $\mathcal{M}(q^2, q^L; q, h)$.

One the one hand, Lemma 30.3 implies that

$$\sum_{p\in\mathcal{M}} \frac{1}{p} = \sum_{\substack{1\leqslant j\leqslant h\\ (j,P(w))=1}} \sum_{\substack{j+q^3<p\leqslant j+q^{L+1}\\ p\equiv j \,(\mathrm{mod}\, q)}} \frac{1}{p} = \sum_{\substack{1\leqslant j\leqslant h\\ (j,P(w))=1}} \frac{\log L + O(1)}{\varphi(q)}.$$

Hence, applying Theorem 14.4 and noticing that $q/\varphi(q) = e^\gamma \log w + O(1)$ by Mertens's third estimate (Theorem 3.4(c)), we conclude that

$$\sum_{p\in\mathcal{M}} \frac{1}{p} = e^\gamma B(u) \cdot \frac{h}{q} \cdot (1 + O_u(1/\log w)) \cdot (\log L + O(1)). \tag{30.5}$$

On the other hand, for each fixed u and for $w \to \infty$, we have

$$\sum_{p\in\mathcal{M}} \frac{1}{p} = \sum_{q^2<m\leqslant q^L} \sum_{mq<p\leqslant h+mq} \frac{1}{p} \sim \sum_{q^2<m\leqslant q^L} \frac{\pi(h+mq)-\pi(mq)}{mq}, \tag{30.6}$$

since $q/h \to \infty$ when $w \to \infty$.

Now, we select u such that $e^\gamma B(u) > 1$ and set $\delta = e^\gamma B(u) - 1 > 0$. We have

$$\sum_{q^2<m\leqslant q^L} \frac{1}{m\log(qm)} = \log L + O(1)$$

by partial summation. Together with (30.5), this implies that there are constants $L_0 = L_0(\delta)$ and $w_0 = w_0(\delta, u)$ such that

$$\sum_{p\in\mathcal{M}} \frac{1}{p} \geqslant (1+\delta/2) \sum_{q^2<m\leqslant q^L} \frac{h/\log(mq)}{mq} \tag{30.7}$$

whenever $L \geqslant L_0$ and $w \geqslant w_0$. From now on, we suppose that $L = L_0$, so that L is fixed in terms of δ.

Comparing (30.7) with (30.6), and assuming w is large enough in terms of δ and u, we find that there must exist some $m \in \mathbb{Z} \cap (q^2, q^L]$ such that

$$\pi(h+mq) - \pi(mq) \geqslant (1+\delta/3) \cdot \frac{h}{\log(mq)}. \tag{30.8}$$

Recall that $h = w^u$, so that $h \sim (\log q)^u \asymp_{u,\delta} (\log(mq))^u$ for all $m \in (q^2, q^L]$. Hence, if $1 < C \leqslant u-1$ and we let w be large enough in terms of δ

and u, then (30.8) and the pigeonhole principle imply the existence of some $x \in [mq, h+mq]$ such that $\pi(x+(\log x)^C) - \pi(x) \geqslant (1+\delta/4)(\log x)^{C-1}$.

The above discussion proves the rightmost inequality in Theorem 30.2 for $1 < C \leqslant u-1$. Since u can be taken to be arbitrarily large by Exercise 14.11(e), we have established that for each fixed $C > 1$, the lim sup of $(\pi(x+(\log x)^{C-1}) - \pi(x))/(\log x)^{C-1}$ is > 1 as $x \to \infty$.

An obvious modification of the above argument, where we work with a sequence of u for which $B(u) < e^{-\gamma}$, proves that the leftmost inequality in the statement of Theorem 30.2 is also true for all $C > 1$. This concludes the proof of Maier's theorem. □

Exercises

Exercise 30.1 (Banks-Ford-Tao [6]). For $y \geqslant 3$ and $u > 2$, let

$$\beta^+(y,u) = \max_{s\in\mathbb{N}} \frac{\#\{\, s < n \leqslant s + y^u : (n, P(y)) = 1 \,\}}{y^u/\log y}$$

and

$$\beta^-(y,u) = \min_{s\in\mathbb{N}} \frac{\#\{\, s < n \leqslant s + y^u : (n, P(y)) = 1 \,\}}{y^u/\log y}.$$

(a) Let $2 < v \leqslant u$. Show that $\beta^+(y,u) \leqslant (1+o_{y\to\infty}(1))\beta^+(y,v)$ and $\beta^-(y,u) \geqslant (1+o_{y\to\infty}(1))\beta^-(y,v)$. [*Hint:* Use the pigeonhole principle.]

(b) Give a heuristic argument that justifies the following claims:
(i) $\max_{p_n\leqslant x}(p_{n+1}-p_n) \sim (\log x)^2/\big[e^\gamma \cdot \beta^-(\log x, 2)\big]$ as $x \to \infty$.
(ii) Fix $u > 2$. If $X \to \infty$, then

$$\max_{X^{1/\log\log X}\leqslant x\leqslant X} \frac{\pi(x+(\log x)^u) - \pi(x)}{(\log x)^{u-1}} \sim e^\gamma \beta^+(\log X, u) \tag{30.9}$$

and

$$\min_{X^{1/\log\log X}\leqslant x\leqslant X} \frac{\pi(x+(\log x)^u) - \pi(x)}{(\log x)^{u-1}} \sim e^\gamma \beta^-(\log X, u). \tag{30.10}$$

(c) Assume that for all $u > 2$ and $\varepsilon > 0$, there are $y_0 \geqslant 3$ and $\delta > 0$ such that $|\beta^\pm(y',u') - \beta^\pm(y,u)| \leqslant \varepsilon$ when $|\log(y'/y)| \leqslant \delta$, $y \geqslant y_0$ and $|u'-u| \leqslant \delta$.

Prove that the ratio of the left over the right side of (30.9) is $\geqslant 1+o(1)$, and that the ratio of the left over the right side of (30.10) is $\leqslant 1+o(1)$.

Appendices

Appendix A

The Riemann-Stieltjes integral

The Riemann-Stieltjes integral is a generalization of the Riemann integral that is very useful in analytic number theory, because it allows us to transform discrete sums into integrals and thus easily manipulate them using our intuition from integral calculus. We present here the basic definitions and properties of this theory following the treatment in [3, Chapter 7]. The basic theory is also presented in [158, Chapter 6] and [146, Appendix A].

Consider two functions $f, \alpha : [a, b] \to \mathbb{R}$, a partition $\mathcal{P} = \{x_0, x_1, \ldots, x_n\}$ of $[a, b]$, and a selection of points $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_n\}$ with $\xi_j \in [x_{j-1}, x_j]$ for each j. We then define the *Riemann-Stieltjes sum* of f with respect to α, $\mathcal{P}$ and $\boldsymbol{\xi}$ by

$$S(f, \alpha; \mathcal{P}, \boldsymbol{\xi}) = \sum_{j=1}^{n} f(\xi_j) \cdot (\alpha(x_j) - \alpha(x_{j-1})).$$

Assume there is a real number I with the property that, given any $\varepsilon > 0$, there is a partition $\mathcal{P}_\varepsilon$ such that

$$|S(f, \alpha; \mathcal{P}, \boldsymbol{\xi}) - I| < \varepsilon$$

whenever $\mathcal{P}$ is a refinement of $\mathcal{P}_\varepsilon$ (i.e., $\mathcal{P} \supseteq \mathcal{P}_\varepsilon$). We then say that f is *integrable* with respect to α (over $[a, b]$) and write symbolically $f \in \mathcal{R}(\alpha)$. The number I is called the *Riemann-Stieltjes integral* of f with respect to α and it is denoted by

$$I = \int_a^b f \mathrm{d}\alpha = \int_a^b f(x) \mathrm{d}\alpha(x).$$

The following theorem establishes the needed properties of the Riemann-Stieltjes integral for the purposes of this book. Its proof is contained in [3] (see Theorems 7.2, 7.3, 7.27, 7.6, 7.8 and 7.11 there, respectively).

Theorem A.1. *Let $f, g, \alpha, \beta : [a, b] \to \mathbb{R}$ and $\lambda, \mu \in \mathbb{R}$.*

(a) *If $f, g \in \mathcal{R}(\alpha)$, then $\lambda f + \mu g \in \mathcal{R}(\alpha)$. We further have*

$$\int_a^b (\lambda f + \mu g)\mathrm{d}\alpha = \lambda \int_a^b f\mathrm{d}\alpha + \mu \int_a^b g\mathrm{d}\alpha.$$

(b) *If $f \in \mathcal{R}(\alpha) \cap \mathcal{R}(\beta)$, then $f \in \mathcal{R}(\lambda\alpha + \mu\beta)$. We further have*

$$\int_a^b f\mathrm{d}(\lambda\alpha + \mu\beta) = \lambda \int_a^b f\mathrm{d}\alpha + \mu \int_a^b f\mathrm{d}\beta.$$

(c) *If f is continuous and α is of bounded variation, then $f \in \mathcal{R}(\alpha)$.*

(d) *If $f \in \mathcal{R}(\alpha)$, then $\alpha \in \mathcal{R}(f)$ and*

$$\int_a^b f\mathrm{d}\alpha = f(x)\alpha(x)\Big|_{x=a}^b - \int_a^b \alpha\mathrm{d}f.$$

(e) *If $f \in \mathcal{R}(\alpha)$ and α is continuously differentiable on $[a, b]$, then the Riemann integral $\int_a^b f(x)\alpha'(x)\mathrm{d}x$ exists and we have*

$$\int_a^b f(x)\mathrm{d}\alpha(x) = \int_a^b f(x)\alpha'(x)\mathrm{d}x.$$

(f) *Assume that α is a step function whose only discontinuities are at the finitely many points $x_1, \ldots, x_n \in [a, b]$, with corresponding jumps $\Delta\alpha_j := \alpha(x_j^+) - \alpha(x_j^-)$.*

Assume further that, at each point x_j, at least one of f and α is continuous from the right, and at least one of them is continuous from the left.

Then $f \in \mathcal{R}(\alpha)$ and we have

$$\int_a^b f\mathrm{d}\alpha = \sum_{j=1}^n f(x_j)\Delta\alpha_j.$$

Appendix B

The Fourier and the Mellin transforms

We write $L^1(\mathbb{R})$ for the space of Lebesgue integrable functions $f : \mathbb{R} \to \mathbb{C}$. Given such a function, we define its *Fourier transform* $\widehat{f} : \mathbb{R} \to \mathbb{C}$ via the formula

$$\widehat{f}(\xi) := \int_{-\infty}^{\infty} f(x) e^{-2\pi i \xi x} \mathrm{d}x. \tag{B.1}$$

We then have the Fourier inversion formula [45, (7.16), p. 218].

Theorem B.1. *If f is continuous and such that $f, \widehat{f} \in L^1(\mathbb{R})$, then*

$$f(x) = \int_{-\infty}^{\infty} \widehat{f}(\xi) e^{2\pi i x \xi} \mathrm{d}\xi.$$

The condition that $\widehat{f} \in L^1(\mathbb{R})$ is not always easy to verify. The simplest way to guarantee it is by assuming that f is smooth enough. Indeed, if the derivatives $f, f', \dots, f^{(j)}$ exist, are in $L^1(\mathbb{R})$ and tend to 0 at $\pm\infty$, then integrating by parts j times in (B.1) yields that

$$\widehat{f}(\xi) = \frac{1}{(2\pi i \xi)^j} \int_{-\infty}^{\infty} f^{(j)}(x) e^{-2\pi i \xi x} \mathrm{d}x \ll_{j,f} \frac{1}{|\xi|^j}. \tag{B.2}$$

In particular, if we can take $j = 2$, then $\widehat{f} \in L^1(\mathbb{R})$, so the hypotheses of Theorem B.1 are met.

Sometimes, we want to know that the Fourier inversion formula holds under even weaker conditions. Such a result is provided by the following theorem [45, Theorem 7.6, p. 220].

Theorem B.2. *Let $f : \mathbb{R} \to \mathbb{C}$ be piecewise continuously differentiable*[1] *and Lebesgue integrable over $\mathbb{R}$. For each $x \in \mathbb{R}$, we have*

$$\frac{f(x^+) + f(x^-)}{2} = \lim_{R\to\infty} \int_{-R}^{R} \widehat{f}(\xi) e^{2\pi i x\xi} \mathrm{d}\xi.$$

Finally, a very useful property of the Fourier transform is the *Poisson summation formula*. This formula states that

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \widehat{f}(n) \tag{B.3}$$

for all "nice" functions $f : \mathbb{R} \to \mathbb{C}$. There are various ways to define what we mean by "nice". An easy way is to assume that f is continuously differentiable twice and that we have $f^{(j)}(x) \ll 1/x^2$ for $j \in \{1, 2\}$ and $|x| \geqslant 1$, that is to say, f, f' and f'' decay at infinity at least as fast as the inverse of a quadratic polynomial. We then use relation (B.2) to obtain the bound $\widehat{f}(\xi) \ll 1/\xi^2$ for $|\xi| \geqslant 1$. In particular, both sides of (B.3) are well-defined.

In order to prove (B.3), we define $g(x) = \sum_{n\in\mathbb{Z}} f(x + n)$, which is a 1-periodic function in $C^2(\mathbb{R})$. Thus, Theorem 2.1 in [45, p. 35] implies that

$$g(x) = \sum_{m\in\mathbb{Z}} c_m e^{2\pi i m x},$$

where $c_m = \int_0^1 g(x) e^{-2\pi i m x} \mathrm{d}x$. We then note that

$$c_m = \int_0^1 \sum_{n\in\mathbb{Z}} f(x + n) e^{-2\pi i m x} \mathrm{d}x = \sum_{n\in\mathbb{Z}} \int_0^1 f(x + n) e^{-2\pi i m x} \mathrm{d}x$$

by Lebesgue's Dominated Convergence Theorem, since $f(x+n) \ll 1/(1+n^2)$ for all $x \in [0, 1]$ and all $n \in \mathbb{Z}$. Setting $y = x + n$ and noticing that $e^{2\pi i m n} = 1$, we conclude that

$$c_m = \sum_{n\in\mathbb{Z}} \int_n^{n+1} f(y) e^{-2\pi i m y} \mathrm{d}y = \widehat{f}(m).$$

Consequently

$$\sum_{n\in\mathbb{Z}} f(x + n) = g(x) = \sum_{m\in\mathbb{Z}} \widehat{f}(m) e^{2\pi i m x}.$$

Taking $x = 0$ proves (B.3). To sum up, we have shown the following result.

Theorem B.3. *Let $f \in C^2(\mathbb{R})$. Assume further that $f^{(j)}(x) \ll 1/x^2$ for $j \in \{0, 1, 2\}$ and $|x| \geqslant 1$. Then $\widehat{f}(\xi) \ll 1/\xi^2$ for $|\xi| \geqslant 1$ and*

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{n\in\mathbb{Z}} \widehat{f}(n).$$

[1] This means that f and f' are piecewise continuous over $\mathbb{R}$, that is to say, they both have a discrete set of discontinuities (i.e., with no accumulation points) that are all of the first kind (i.e., "jump discontinuities").

The Mellin transform

Given a function $g : \mathbb{R}_{>0} \to \mathbb{C}$, we define its *Mellin transform* to be

$$G(s) = \int_0^\infty g(x) x^{s-1} \mathrm{d}x \tag{B.4}$$

for all $s \in \mathbb{C}$ that this integral converges. An important example of a Mellin transform is the Gamma function that we studied at the end of Chapter 1.

The Mellin transform is a close relative of the Fourier transform. Indeed, if we make a change of variables $x = e^u$, we immediately see that $G(s) = \hat{h}(-s/2\pi i)$ with $h(u) = g(e^u)$. In particular, we have the *Mellin inversion formula* as a consequence of Fourier inversion:

Theorem B.4. *Let g and G be as above, with g piecewise continuously differentiable. Assume further that there are $\alpha_1 < \alpha_2$ such that the function $x \to |g(x)|x^{\sigma-1}$ is in $L^1(\mathbb{R}_{\geqslant 0})$ for all $\sigma \in (\alpha_1, \alpha_2)$.*

Then G is a holomorphic function in the strip $\alpha_1 < \mathrm{Re}(s) < \alpha_2$. In addition, the inversion formula

$$\frac{g(x^+) + g(x^-)}{2} = \frac{1}{2\pi i} \int_{(\alpha)} G(s) x^{-s} \mathrm{d}s := \frac{1}{2\pi i} \lim_{T \to \infty} \int_{\substack{\mathrm{Re}(s)=\alpha \\ |\mathrm{Im}(s)| \leqslant T}} G(s) x^{-s} \mathrm{d}s$$

holds for all $x > 0$ and all $\alpha \in (\alpha_1, \alpha_2)$.

Proof. Let ε be positive and smaller than $(\alpha_2 - \alpha_1)/2$. By the hypotheses of the theorem, the integrals

$$\int_0^1 |g(x)| x^{\alpha_1 + \varepsilon - 1} \mathrm{d}x \qquad \text{and} \qquad \int_1^\infty |g(x)| x^{\alpha_2 - \varepsilon - 1} \mathrm{d}x$$

converge. Hence, the integral defining $G(s)$ converges absolutely and uniformly in the strip $\alpha_1 + \varepsilon \leqslant \mathrm{Re}(s) \leqslant \alpha_2 - \varepsilon$. The holomorphicity of G then follows. For the Mellin inversion formula, note that the function $\xi \to G(\alpha - 2\pi i \xi)$ is the Fourier transform of the function $u \to g(e^u) e^{\alpha u}$. Applying Theorem B.2 completes the proof. □

Appendix C

The method of moments

We prove here a generalized theorem of Theorem 15.2, which is the main probabilistic tool needed in the proof of the Erdős-Kac theorem.

Given a constant $c > 0$ and a random variable X (defined on some ambient probability space), we write $X \in \mathcal{E}(c)$ if $\mathbb{P}(|X| > u) \ll e^{-cu}$ uniformly for $u \geqslant 0$. In addition, we write $X \in \mathcal{E}(\infty)$ if $X \in \mathcal{E}(c)$ for each fixed $c > 0$. Clearly, the standard normal distribution is in the class $\mathcal{E}(\infty)$, as well as any compactly supported distribution.

We will prove the following generalization of Theorem 15.2.

Theorem C.1. *Let X be a random variable in the class $\mathcal{E}(\infty)$, and let $(X_j)_{j=1}^{\infty}$ be a sequence of random variables.*

(a) *Assume that*

$$\lim_{j\to\infty} \mathbb{E}[X_j^k] = \mathbb{E}[X^k] \quad \text{for all } k \in \mathbb{N}. \tag{C.1}$$

Then $(X_j)_{j=1}^{\infty}$ converges in distribution to X.

(b) *Conversely, assume that $(X_j)_{j=1}^{\infty}$ converges in distribution to X. If, in addition, $\sup_{j\geqslant 1} \mathbb{E}[X_j^{2k}] < \infty$ for all $k \in \mathbb{N}$, then* (C.1) *holds.*

Before we embark on the proof of Theorem C.1, we make a few remarks.

Remark C.2. (a) The condition that $X \in \mathcal{E}(c)$ is closely related to having that $\mathbb{E}[|X|^k] \ll k!/c^k$ uniformly for $k \in \mathbb{Z}_{\geqslant 1}$. Indeed, if $X \in \mathcal{E}(c)$, then

$$\mathbb{E}[|X|^k] = \int_0^{\infty} ku^{k-1}\mathbb{P}(|X| > u)\mathrm{d}u \ll k\int_0^{\infty} u^{k-1}e^{-cu}\mathrm{d}u = k!/c^k. \tag{C.2}$$

Conversely, if the moments of X satisfy the uniform bound $\mathbb{E}[|X|^k] \ll k!/c^k$, then Markov's inequality implies that

$$\mathbb{P}(|X| > u) \leqslant u^{-k}\mathbb{E}[|X|^k] \ll (cu)^{-k}k! \asymp \sqrt{k} \cdot e^{-k}(k/cu)^k$$

for all $k \in \mathbb{Z}_{\geqslant 1}$. Taking $k = \lfloor cu \rfloor + 1$ proves that $\mathbb{P}(|X| > u) \ll_c u^{1/2}e^{-cu}$ for all $u \geqslant 1$. In particular, $X \in \mathcal{E}(c')$ for all $c' < c$.

(b) Theorem C.1 holds even if $X \in \mathcal{E}(c)$ with $0 < c < \infty$. (See Theorems 29.3 and 30.1 in Billingsley's book [7].) □

Proof of Theorem C.1. We start with part (b) that is simpler. Convergence in distribution is equivalent to weak convergence, that is to say,

$$\lim_{j\to\infty} \mathbb{E}[f(X_j)] = \mathbb{E}[f(X)] \tag{C.3}$$

for any bounded $f \in C(\mathbb{R})$ [7, Theorem 25.8]. Hence, if we let $\phi(x) = 1$ for $|x| \leqslant 1$, and $\phi(x) = \max\{0, 2 - |x|\}$ for $|x| > 1$, then

$$\lim_{j\to\infty} \mathbb{E}[X_j^k\phi(X_j/M)] = \mathbb{E}[X^k\phi(X/M)]$$

for any $M > 0$. In addition, we have

$$\Big|\mathbb{E}[X_j^k] - \mathbb{E}[X_j^k\phi(X_j/M)]\Big| \leqslant \mathbb{E}[|X_j|^k 1_{|X_j|>M}] \leqslant M^{-k}\mathbb{E}[X_j^{2k}] \ll_k 1/M$$

uniformly in $j \in \mathbb{Z}_{\geqslant 1}$ and $M \geqslant 1$. Similarly, $\mathbb{E}[X^k\phi(X/M)] = \mathbb{E}[X^k] + O_k(1/M)$. We thus infer the validity of (C.1).

We now prove part (a), where we assume that (C.1) holds. It suffices to prove that if f is a smooth function supported on $[a, b]$, then (C.3) holds. We will employ an explicit version of Weierstrass's approximation theorem using Chebyshev polynomials of the first kind, defined by $T_n(\cos\theta) = \cos(n\theta)$. Using the formula $e^{i\theta} = \cos\theta + i\sin\theta$, we may easily deduce that

$$T_n(x) = \sum_{0\leqslant j\leqslant n/2} \binom{n}{2j} x^{n-2j}(x^2 - 1)^j.$$

In particular, we have

$$|T_n(x)| \leqslant 2^n x^n \leqslant 2^n x^{2n+2} \quad \text{for } |x| \geqslant 1. \tag{C.4}$$

Now, fix M to be a large enough parameter so that $[a, b] \subseteq (-M, M)$ and consider the function $\alpha \to f(M\cos(2\pi\alpha))$, which is 1-periodic, even and smooth. We may thus develop it in its Fourier series, say

$$f(M\cos(2\pi\alpha)) = \sum_{n\geqslant 0} a_{n,M}\cos(2\pi n\alpha),$$

where

$$a_{n,M} = (1 + 1_{n>0})\int_0^1 f(M\cos(2\pi\alpha))\cos(2\pi n\alpha)\mathrm{d}\alpha.$$

Integrating by parts twice, we find that

$$a_{n,M} = -\frac{1+1_{n>0}}{n^2}\int_0^1 g_M(\alpha)\cos(2\pi n\alpha)\mathrm{d}\alpha,$$

where $g_M(\alpha) = M^2\sin^2(2\pi\alpha)f''(M\cos(2\pi\alpha)) - M\cos(2\pi\alpha)f'(M\cos(2\pi\alpha))$. Since f has bounded support, g_M is supported on $\alpha \in [0,1]$ such that $\cos(2\pi\alpha) = O_f(1/M)$ (we think of M as big in terms of a and b). This set has measure $O_f(1/M)$. Thus $a_{n,M} = O_f(M/n^2)$. We conclude that

$$f(M\cos(2\pi\alpha)) = \sum_{0\leqslant n\leqslant M/\varepsilon} a_{n,M}\cos(2\pi n\alpha) + O_f(\varepsilon) \tag{C.5}$$

uniformly for $M \geqslant 1$ and $0 < \varepsilon \leqslant 1$.

We use (C.5) to write f in terms of the Chebyshev polynomials. If $|x| \leqslant M$, then $x = M\cos(2\pi\alpha)$ for some $\alpha \in [0,1]$ and thus

$$f(x) = \sum_{0\leqslant n\leqslant M/\varepsilon} a_{n,M}T_n(x/M) + O_f(\varepsilon). \tag{C.6}$$

On the other hand, when $|x| > M$, the left-hand side of (C.6) is 0, whereas the right-hand side is $\ll_f \varepsilon + \sum_{0\leqslant n\leqslant M/\varepsilon}(2x/M)^{2n+2}$ by (C.4) and the trivial bound $a_{n,M} = O_f(1)$. This proves that for all $x \in \mathbb{R}$ we have

$$f(x) = \sum_{0\leqslant n\leqslant M/\varepsilon} a_{n,M}T_n(x/M) + O_f\Big(\varepsilon + \sum_{0\leqslant n\leqslant M/\varepsilon}(2x/M)^{2n+2}\Big).$$

Applying the above formula twice, we deduce that

$$\begin{aligned}\mathbb{E}[f(X_j)] - \mathbb{E}[f(X)] = &\sum_{0\leqslant n\leqslant M/\varepsilon} a_{n,M}\big(\mathbb{E}[T_n(X_j/M)] - \mathbb{E}[T_n(X/M)]\big)\\ &+O_f\Big(\varepsilon + \sum_{0\leqslant n\leqslant M/\varepsilon}(2/M)^{2n+2}\big(\mathbb{E}[X_j^{2n+2}] + \mathbb{E}[X^{2n+2}]\big)\Big).\end{aligned}$$

When $j \to \infty$, the main term goes to 0 by assumption of (C.1). Thus

$$\limsup_{j\to\infty}\big|\mathbb{E}[f(X_j)] - \mathbb{E}[f(X)]\big| \ll_f \varepsilon + \sum_{0\leqslant n\leqslant M/\varepsilon}(2/M)^{2n+2}\mathbb{E}[X^{2n+2}]. \tag{C.7}$$

Let $k = n+1$. Since $X \in \mathcal{E}(4)$, we use (C.2) with $2k$ in place of k to find

$$\sum_{1\leqslant k\leqslant\sqrt{M}}(2/M)^{2k}\mathbb{E}[X^{2k}] \ll \sum_{1\leqslant k\leqslant\sqrt{M}}(k/M)^{2k} \leqslant \sum_{1\leqslant k\leqslant\sqrt{M}} M^{-k} \ll 1/M.$$

For larger k we use that $X \in \mathcal{E}(4\varepsilon^{-3/2})$. Hence,

$$\sum_{\sqrt{M}<k\leqslant M/\varepsilon}(2/M)^k\mathbb{E}[X^{2k}] \ll_\varepsilon \sum_{\sqrt{M}<k\leqslant M/\varepsilon}(\varepsilon^{3/2}k/M)^{2k} \ll \varepsilon^{\sqrt{M}}.$$

Thus, letting $M \to \infty$ in (C.7) yields $\limsup_{j\to\infty}|\mathbb{E}[f(X_j)] - \mathbb{E}[f(X)]| \ll_f \varepsilon$. Finally, we let $\varepsilon \to 0^+$ to deduce (C.3). This completes the proof. □

Bibliography

[1] M. Agrawal, N. Kayal, and N. Saxena, *PRIMES is in P*, Ann. of Math. (2) **160** (2004), no. 2, 781–793, DOI 10.4007/annals.2004.160.781. MR2123939

[2] L. V. Ahlfors, *Complex analysis*: *An introduction to the theory of analytic functions of one complex variable*, 3rd ed., International Series in Pure and Applied Mathematics, McGraw-Hill Book Co., New York, 1978. MR510197

[3] T. M. Apostol, *Mathematical analysis*, 2nd ed., Addison-Wesley Publishing Co., Reading, Mass.-London-Don Mills, Ont., 1974. MR0344384

[4] T. M. Apostol, *Introduction to analytic number theory*, Undergraduate Texts in Mathematics, Springer-Verlag, New York-Heidelberg, 1976. MR0434929

[5] S. Axler, *Linear algebra done right*, 2nd ed., Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1997. MR1482226

[6] W. Banks, K. Ford and T. Tao, *Large prime gaps and probabilistic models.* Preprint (2019), 38 pages, arXiv:1908.08613.

[7] P. Billingsley, *Probability and measure*, 3rd ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1995. A Wiley-Interscience Publication. MR1324786

[8] E. Bombieri, *On the large sieve*, Mathematika **12** (1965), 201–225, DOI 10.1112/S0025579300005313. MR0197425

[9] E. Bombieri, *The asymptotic sieve* (English, with Italian summary), Rend. Accad. Naz. XL (5) **1/2** (1975/76), 243–269 (1977). MR0491570

[10] E. Bombieri, *Le grand crible dans la théorie analytique des nombres* (French, with English summary), Astérisque **18** (1987), 103. MR891718

[11] E. Bombieri, J. B. Friedlander, and H. Iwaniec, *Primes in arithmetic progressions to large moduli*, Acta Math. **156** (1986), no. 3-4, 203–251, DOI 10.1007/BF02399204. MR834613

[12] E. Bombieri, J. B. Friedlander, and H. Iwaniec, *Primes in arithmetic progressions to large moduli. II*, Math. Ann. **277** (1987), no. 3, 361–393, DOI 10.1007/BF01458321. MR891581

[13] E. Bombieri, J. B. Friedlander, and H. Iwaniec, *Primes in arithmetic progressions to large moduli. III*, J. Amer. Math. Soc. **2** (1989), no. 2, 215–224, DOI 10.2307/1990976. MR976723

[14] E. Bombieri and H. Iwaniec, *On the order of* $\zeta(\frac{1}{2} + it)$, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **13** (1986), no. 3, 449–472. MR881101

[15] E. Bombieri and H. Iwaniec, *Some mean-value theorems for exponential sums*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **13** (1986), no. 3, 473–486. MR881102

[16] J. Bourgain, *Decoupling, exponential sums and the Riemann zeta function*, J. Amer. Math. Soc. **30** (2017), no. 1, 205–224, DOI 10.1090/jams/860. MR3556291

[17] N. G. de Bruijn, *On the number of positive integers $\leq x$ and free of prime factors $> y$*, Nederl. Acad. Wetensch. Proc. Ser. A. **54** (1951), 50–60. MR0046375

[18] N. G. de Bruijn, *On the number of positive integers $\leq x$ and free prime factors $> y$. II*, Nederl. Akad. Wetensch. Proc. Ser. A 69=Indag. Math. **28** (1966), 239–247. MR0205945

[19] V. Brun, *Über das Goldbachsche Gesetzund die Anzahlder Primzahlpaare,* Archiv for Math. og Naturvid. **34** (1915), no. 8, 19 pp.

[20] V. Brun, *La série* $1/5 + 1/7 + 1/11 + 1/13 + 1/17 + 1/19 + 1/29 + 1/31 + 1/41 + 1/43 + 1/59 + 1/61 + \cdots$ *où les dénominateurs sont "nombres premiers jumeaus" est convergente ou finie,* Bull. Sci. Math. (2) **43** (1919), 100–104; 124–128.

[21] V. Brun, *Reflections on the sieve of Eratosthenes*, Norske Vid. Selsk. Skr. (Trondheim) **1967** (1967), no. 1, 9. MR0219466

[22] A. A. Buchstab, *Asymptotic estimates of a general number-theoretic function* (Russian), Mat. Sb. (2) **44** (1937), 1239–1246.

[23] A. A. Buchstab, *New improvements in the method of the sieve of Eratosthenes,* Mat. Sb. (N. S.) 4 **46** (1938), 375–387.

[24] Jing-run Chen, *On the representation of a large even integer as the sum of a prime and the product of at most two primes*, Kexue Tongbao (Foreign Lang. Ed.) **17** (1966), 385–386. MR0207668

[25] J. R. Chen, *On the representation of a large even integer as the sum of a prime and the product of at most two primes. II*, Sci. Sinica **21** (1978), no. 4, 421–430. MR511293

[26] A. C. Cojocaru and M. R. Murty, *An introduction to sieve methods and their applications*, London Mathematical Society Student Texts, vol. 66, Cambridge University Press, Cambridge, 2006. MR2200366

[27] H. Cramér, *Some theorems concerning prime numbers*, Arkiv för Mat. Astr. o. Fys. **15** (1920), no. 5, 1–32.

[28] H. Cramér, *On the distribution of primes*, Proc. Camb. Phil. Soc. **20** (1920), 272–280.

[29] H. Cramér, *Prime numbers and probability*, Skand. Mat.-Kongr. **8** (1935), 107–115.

[30] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arith. **2** (1936), 23–46.

[31] H. Davenport, *Multiplicative number theory*, 3rd ed., Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York, 2000. Revised and with a preface by Hugh L. Montgomery. MR1790423

[32] H. Delange, *Sur des formules dues à Atle Selberg* (French), Bull. Sci. Math. (2) **83** (1959), 101–111. MR0113836

[33] H. G. Diamond and H. Halberstam, *A higher-dimensional sieve method*, Cambridge Tracts in Mathematics, vol. 177, Cambridge University Press, Cambridge, 2008. With an appendix ("Procedures for computing sieve functions") by William F. Galway. MR2458547

[34] H. Diamond, H. Halberstam, and H.-E. Richert, *Combinatorial sieves of dimension exceeding one*, J. Number Theory **28** (1988), no. 3, 306–346, DOI 10.1016/0022-314X(88)90046-7. MR932379

[35] K. Dickman, *On the frequency of numbers containing prime factors of a certain relative magnitude,* Ark. Mat. Astr. fys. **22** (1930), 1–14.

[36] Á. Elbert, *Some recent results on the zeros of Bessel functions and orthogonal polynomials*, Proceedings of the Fifth International Symposium on Orthogonal Polynomials, Special Functions and their Applications (Patras, 1999), J. Comput. Appl. Math. **133** (2001), no. 1-2, 65–83, DOI 10.1016/S0377-0427(00)00635-X. MR1858270

[37] H. M. Edwards, *Riemann's zeta function*, Dover Publications, Inc., Mineola, NY, 2001. Reprint of the 1974 original [Academic Press, New York; MR0466039 (57 #5922)]. MR1854455

[38] P. D. T. A. Elliott, *Probabilistic number theory. I*: *Mean-value theorems*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science], vol. 239, Springer-Verlag, New York-Berlin, 1979. MR551361

[39] P. D. T. A. Elliott, *Probabilistic number theory. II*: *Central limit theorems*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 240, Springer-Verlag, Berlin-New York, 1980. MR560507

[40] P. D. T. A. Elliott, *Multiplicative functions on arithmetic progressions. VII. Large moduli*, J. London Math. Soc. (2) **66** (2002), no. 1, 14–28, DOI 10.1112/S0024610702003228. MR1911217

[41] P. D. T. A. Elliott and H. Halberstam, *A conjecture in prime number theory*, Symposia Mathematica, Vol. IV (INDAM, Rome, 1968/69), Academic Press, London, 1970, pp. 59–72. MR0276195

[42] P. Erdös, *The difference of consecutive primes*, Duke Math. J. **6** (1940), 438–441. MR1759

[43] P. Erdös and M. Kac, *The Gaussian law of errors in the theory of additive number theoretic functions*, Amer. J. Math. **62** (1940), 738–742, DOI 10.2307/2371483. MR0002374

[44] L. Euler, *Commentationes Arithmeticae. V. 3* (Latin), Leonhardi Euleri Opera Omnia (1) 4, Orell Füssli, Zurich; B. G. Teubner, Leipzig, 1941. Edited by Rudolf Fueter. MR0006112

[45] G. B. Folland, *Fourier analysis and its applications*, The Wadsworth & Brooks/Cole Mathematics Series, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1992. MR1145236

[46] K. Ford, B. Green, S. Konyagin, and T. Tao, *Large gaps between consecutive prime numbers*, Ann. of Math. (2) **183** (2016), no. 3, 935–974, DOI 10.4007/annals.2016.183.3.4. MR3488740

[47] K. Ford, B. Green, S. Konyagin, J. Maynard, and T. Tao, *Long gaps between primes*, J. Amer. Math. Soc. **31** (2018), no. 1, 65–105, DOI 10.1090/jams/876. MR3718451

[48] K. Ford and H. Halberstam, *The Brun-Hooley sieve*, J. Number Theory **81** (2000), no. 2, 335–350, DOI 10.1006/jnth.1999.2479. MR1752258

[49] É. Fouvry, *Répartition des suites dans les progressions arithmétiques* (French), Acta Arith. **41** (1982), no. 4, 359–382, DOI 10.4064/aa-41-4-359-382. MR677549

[50] É. Fouvry, *Autour du théorème de Bombieri-Vinogradov* (French), Acta Math. **152** (1984), no. 3-4, 219–244, DOI 10.1007/BF02392198. MR741055

[51] E. Fouvry and H. Iwaniec, *On a theorem of Bombieri-Vinogradov type*, Mathematika **27** (1980), no. 2, 135–152 (1981), DOI 10.1112/S0025579300010032. MR610700

[52] E. Fouvry and H. Iwaniec, *Primes in arithmetic progressions*, Acta Arith. **42** (1983), no. 2, 197–218, DOI 10.4064/aa-42-2-197-218. MR719249

[53] J. Friedlander and A. Granville, *Limitations to the equi-distribution of primes. I*, Ann. of Math. (2) **129** (1989), no. 2, 363–382, DOI 10.2307/1971450. MR986796

[54] J. Friedlander and A. Granville, *Limitations to the equi-distribution of primes. III*, Compositio Math. **81** (1992), no. 1, 19–32. MR1145606

[55] J. Friedlander, A. Granville, A. Hildebrand, and H. Maier, *Oscillation theorems for primes in arithmetic progressions and for sifting functions*, J. Amer. Math. Soc. **4** (1991), no. 1, 25–86, DOI 10.2307/2939254. MR1080647

[56] J. Friedlander and H. Iwaniec, *On Bombieri's asymptotic sieve*, Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4) **5** (1978), no. 4, 719–756. MR519891

[57] J. Friedlander and H. Iwaniec, *Asymptotic sieve for primes*, Ann. of Math. (2) **148** (1998), no. 3, 1041–1065, DOI 10.2307/121035. MR1670069

[58] J. Friedlander and H. Iwaniec, *The polynomial $X^2 + Y^4$ captures its primes*, Ann. of Math. (2) **148** (1998), no. 3, 945–1040, DOI 10.2307/121034. MR1670065

[59] J. Friedlander and H. Iwaniec, *Opera de cribro*, American Mathematical Society Colloquium Publications, vol. 57, American Mathematical Society, Providence, RI, 2010. MR2647984

[60] D. M. Goldfeld, *A simple proof of Siegel's theorem*, Proc. Nat. Acad. Sci. U.S.A. **71** (1974), 1055, DOI 10.1073/pnas.71.4.1055. MR0344222

[61] L. Goldmakher, *Multiplicative mimicry and improvements to the Pólya-Vinogradov inequality*, Algebra Number Theory **6** (2012), no. 1, 123–163, DOI 10.2140/ant.2012.6.123. MR2950162

[62] D. A. Goldston, S. W. Graham, J. Pintz, and C. Y. Yıldırım, *Small gaps between primes or almost primes*, Trans. Amer. Math. Soc. **361** (2009), no. 10, 5285–5330, DOI 10.1090/S0002-9947-09-04788-6. MR2515812

[63] D. A. Goldston, J. Pintz, and C. Y. Yıldırım, *Primes in tuples. I*, Ann. of Math. (2) **170** (2009), no. 2, 819–862, DOI 10.4007/annals.2009.170.819. MR2552109

[64] A. Granville, *Harald Cramér and the distribution of prime numbers*, Scand. Actuar. J. **1** (1995), 12–28, DOI 10.1080/03461238.1995.10413946. Harald Cramér Symposium (Stockholm, 1993). MR1349149

[65] A. Granville, *Primes in intervals of bounded length*, Bull. Amer. Math. Soc. (N.S.) **52** (2015), no. 2, 171–222, DOI 10.1090/S0273-0979-2015-01480-1. MR3312631

[66] A. Granville, A. J. Harper, and K. Soundararajan, *Mean values of multiplicative functions over function fields*, Res. Number Theory **1** (2015), Art. 25, 18, DOI 10.1007/s40993-015-0023-5. MR3501009

[67] A. Granville, D. M. Kane, D. Koukoulopoulos, and R. J. Lemke Oliver, *Best possible densities of Dickson m-tuples, as a consequence of Zhang-Maynard-Tao*, Analytic number theory, Springer, Cham, 2015, pp. 133–144. MR3467396

[68] A. Granville and D. Koukoulopoulos, *Beyond the LSD method for the partial sums of multiplicative functions*, Ramanujan J. **49** (2019), no. 2, 287–319, DOI 10.1007/s11139-018-0119-3. MR3949071

[69] A. Granville, D. Koukoulopoulos, and K. Matomäki, *When the sieve works*, Duke Math. J. **164** (2015), no. 10, 1935–1969, DOI 10.1215/00127094-3120891. MR3369306

[70] A. Granville and G. Martin, *Prime number races*, Amer. Math. Monthly **113** (2006), no. 1, 1–33, DOI 10.2307/27641834. MR2202918

[71] A. Granville and K. Soundararajan, *An uncertainty principle for arithmetic sequences*, Ann. of Math. (2) **165** (2007), no. 2, 593–635, DOI 10.4007/annals.2007.165.593. MR2299742

[72] A. Granville and K. Soundararajan, *Large character sums: pretentious characters and the Pólya-Vinogradov theorem*, J. Amer. Math. Soc. **20** (2007), no. 2, 357–384, DOI 10.1090/S0894-0347-06-00536-4. MR2276774

[73] A. Granville and K. Soundararajan, *Sieving and the Erdős-Kac theorem*, Equidistribution in number theory, an introduction, NATO Sci. Ser. II Math. Phys. Chem., vol. 237, Springer, Dordrecht, 2007, pp. 15–27, DOI 10.1007/978-1-4020-5404-4_2. MR2290492

[74] A. Granville and K. Soundararajan, *Pretentious multiplicative functions and an inequality for the zeta-function*, Anatomy of integers, CRM Proc. Lecture Notes, vol. 46, Amer. Math. Soc., Providence, RI, 2008, pp. 191–197. MR2437976

[75] A. Granville and K. Soundararajan, *Multiplicative number theory*, Snowbird MRC notes (unpublished), 2011.

[76] G. Greaves, *Sieves in number theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 43, Springer-Verlag, Berlin, 2001. MR1836967

[77] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), no. 2, 481–547, DOI 10.4007/annals.2008.167.481. MR2415379

[78] B. Green and T. Tao, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), no. 3, 1753–1850, DOI 10.4007/annals.2010.171.1753. MR2680398

[79] B. Green and T. Tao, *The Möbius function is strongly orthogonal to nilsequences*, Ann. of Math. (2) **175** (2012), no. 2, 541–566, DOI 10.4007/annals.2012.175.2.3. MR2877066

[80] B. Green, T. Tao, and T. Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Ann. of Math. (2) **176** (2012), no. 2, 1231–1372, DOI 10.4007/annals.2012.176.2.11. MR2950773

[81] J. Hadamard, *Étude sur les propriétés des fonctions entières et en particulier d'une fonction considŕé par Riemann*, J. Math. Pures Appl. (4) **9** (1893), 171–215.

[82] J. Hadamard, *Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques* (French), Bull. Soc. Math. France **24** (1896), 199–220. MR1504264

[83] J. Hadamard, *Sur la distribution des zéros de la fonction $\zeta(s)$ et ses conséquences arithmétiques* (French), Bull. Soc. Math. France **24** (1896), 199–220. MR1504264

[84] G. Halász, *On the distribution of additive and the mean values of multiplicative arithmetic functions*, Studia Sci. Math. Hungar. **6** (1971), 211–233. MR0319930

[85] D. K. Faddeyev, S. M. Lozinsky, and A. V. Malyshev, *Yuri V. Linnik (1915–1972): a biographical note*, Acta Arith. **27** (1975), 1–2, DOI 10.4064/aa-27-1-1-2. Collection of articles in memory of Juriĭ Vladimirovič Linnik. MR0421941

[86] H. Halberstam and H.-E. Richert, *Sieve methods*, London Mathematical Society Monographs, vol. 4, Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], London-New York, 1974. MR0424730

[87] G. H. Hardy and J. E. Littlewood, *A new solution to Waring's problem*, Q. J. Math. **48** (1919), 272–293.

[88] G. H. Hardy and J. E. Littlewood, *Some problems of "partitio numerorum": I. A new solution to Waring's problem*, Göttingen Nachrichten, 1920, 33–54.

[89] G. H. Hardy and J. E. Littlewood, *Some problems of "partitio numerorum": II. Proof that every large number is the sum of at most 21 biquadrates*, Math. Z. **9** (1921), no. 1-2, 14–27, DOI 10.1007/BF01378332. MR1544448

[90] G. H. Hardy and J. E. Littlewood, *Some problems of 'Partitio numerorum'; III: On the expression of a number as a sum of primes*, Acta Math. **44** (1923), no. 1, 1–70, DOI 10.1007/BF02403921. MR1555183

[91] G. H. Hardy and J. E. Littlewood, *Some problems of 'Partitio Numerorum': IV. The singular series in Waring's Problem and the value of the number $G(k)$*, Math. Z. **12** (1922), no. 1, 161–188, DOI 10.1007/BF01482074. MR1544511

[92] G. H. Hardy and J. E. Littlewood, *Some problems of 'Partitio numerorum': V. A further contribution to the study of Goldbach's problem*, Proc. London Math. Soc. (2) **22** (1924), 46–56.

[93] G. H. Hardy and J. E. Littlewood, *Some problems of 'Partitio numerorum' (VI): Further researches in Waring's Problem*, Math. Z. **23** (1925), no. 1, 1–37, DOI 10.1007/BF01506218. MR1544728

[94] G. H. Hardy and J. E. Littlewood, *Some problems of 'Partitio Numerorum' (VIII): The number Gamma(k) in Waring's Problem*, Proc. London Math. Soc. (2) **28** (1928), no. 7, 518–542, DOI 10.1112/plms/s2-28.1.518. MR1575871

[95] G. H. Hardy and S. Ramanujan, *Proof that almost all numbers n are composed of about $\log\log n$ prime factors [Proc. London Math. Soc. (2) **16** (1917), Records for 14 Dec. 1916]*, Collected papers of Srinivasa Ramanujan, AMS Chelsea Publ., Providence, RI, 2000, pp. 242–243. MR2280875

[96] G. Harman, *Prime-detecting sieves*, London Mathematical Society Monographs Series, vol. 33, Princeton University Press, Princeton, NJ, 2007. MR2331072

[97] D. R. Heath-Brown, *Prime numbers in short intervals and a generalized Vaughan identity*, Canad. J. Math. **34** (1982), no. 6, 1365–1377, DOI 10.4153/CJM-1982-095-9. MR678676

[98] D. R. Heath-Brown, *Primes represented by $x^3 + 2y^3$*, Acta Math. **186** (2001), no. 1, 1–84, DOI 10.1007/BF02392715. MR1828372

[99] D. R. Heath-Brown and X. Li, *Prime values of $a^2 + p^4$*, Invent. Math. **208** (2017), no. 2, 441–499, DOI 10.1007/s00222-016-0694-0. MR3639597

[100] D. R. Heath-Brown and B. Z. Moroz, *Primes represented by binary cubic forms*, Proc. London Math. Soc. (3) **84** (2002), no. 2, 257–288, DOI 10.1112/plms/84.2.257. MR1881392

[101] D. R. Heath-Brown and B. Z. Moroz, *On the representation of primes by cubic polynomials in two variables*, Proc. London Math. Soc. (3) **88** (2004), no. 2, 289–312, DOI 10.1112/S0024611503014497. MR2032509

[102] A. Hildebrand, *Integers free of large prime factors and the Riemann hypothesis*, Mathematika **31** (1984), no. 2, 258–271 (1985), DOI 10.1112/S0025579300012481. MR804201

[103] A. Hildebrand and G. Tenenbaum, *On integers free of large prime factors*, Trans. Amer. Math. Soc. **296** (1986), no. 1, 265–290, DOI 10.2307/2000573. MR837811

[104] A. Hildebrand and G. Tenenbaum, *Integers without large prime factors*, J. Théor. Nombres Bordeaux **5** (1993), no. 2, 411–484. MR1265913

[105] C. Hooley, *On the Brun-Titchmarsh theorem*, J. Reine Angew. Math. **255** (1972), 60–79, DOI 10.1515/crll.1972.255.60. MR0304328

[106] C. Hooley, *On the Brun-Titchmarsh theorem. II*, Proc. London Math. Soc. (3) **30** (1975), 114–128, DOI 10.1112/plms/s3-30.1.114. MR0369296

[107] C. Hooley, *Applications of sieve methods to the theory of numbers*, Cambridge Tracts in Mathematics, vol. 70, Cambridge University Press, Cambridge-New York-Melbourne, 1976. MR0404173

[108] C. Hooley, *On an almost pure sieve*, Acta Arith. **66** (1994), no. 4, 359–368, DOI 10.4064/aa-66-4-359-368. MR1288352

[109] M. N. Huxley, *Area, lattice points, and exponential sums*, London Mathematical Society Monographs. New Series, vol. 13, The Clarendon Press, Oxford University Press, New York, 1996. Oxford Science Publications. MR1420620

[110] E. K. Ifantis and P. D. Siafarikas, *A differential equation for the zeros of Bessel functions*, Applicable Anal. **20** (1985), no. 3-4, 269–281, DOI 10.1080/00036818508839574. MR814954

[111] A. E. Ingham, *The distribution of prime numbers*, Cambridge Mathematical Library, Cambridge University Press, Cambridge, 1990. Reprint of the 1932 original; With a foreword by R. C. Vaughan. MR1074573

[112] H. Iwaniec, *Rosser's sieve*, Acta Arith. **36** (1980), no. 2, 171–202, DOI 10.4064/aa-36-2-171-202. MR581917

[113] H. Iwaniec, *A new form of the error term in the linear sieve*, Acta Arith. **37** (1980), 307–320, DOI 10.4064/aa-37-1-307-320. MR598883

[114] H. Iwaniec and E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, vol. 53, American Mathematical Society, Providence, RI, 2004. MR2061214

[115] W. B. Jurkat and H.-E. Richert, *An improvement of Selberg's sieve method. I*, Acta Arith. **11** (1965), 217–240, DOI 10.4064/aa-11-2-217-240. MR0202680

[116] M. Kac, *Statistical independence in probability, analysis and number theory*, The Carus Mathematical Monographs, No. 12, Published by the Mathematical Association of America. Distributed by John Wiley and Sons, Inc., New York, 1959. MR0110114

[117] A. Khintchine, *Über das Gesetz der großen Zahlen* (German), Math. Ann. **96** (1927), no. 1, 152–168, DOI 10.1007/BF01209158. MR1512310

[118] N. M. Korobov, *Weyl's estimates of sums and the distribution of primes* (Russian), Dokl. Akad. Nauk SSSR **123** (1958), 28–31. MR0103862

[119] D. Koukoulopoulos, *Pretentious multiplicative functions and the prime number theorem for arithmetic progressions*, Compos. Math. **149** (2013), no. 7, 1129–1149, DOI 10.1112/S0010437X12000802. MR3078641

[120] D. Koukoulopoulos, *On multiplicative functions which are small on average*, Geom. Funct. Anal. **23** (2013), no. 5, 1569–1630, DOI 10.1007/s00039-013-0235-6. MR3102913

[121] E. Kowalski, *Gaps between prime numbers and prime numbers in arithmetic progressions, after Y. Zhang and J. Maynard*, Survey (Bourbaki seminar, March 2014).

[122] J. Kubilius, *Probabilistic methods in the theory of numbers*, Translations of Mathematical Monographs, Vol. 11, American Mathematical Society, Providence, R.I., 1964. MR0160745

[123] Y. Lamzouri and A. P. Mangerel, *Large odd order character sums and improvements of the Pólya-Vinogradov inequality*. Preprint (2017), 34 pages, arXiv:1701.01042.

[124] E. Landau, *Über den Zusammenhang einiger neuer Sätze der analytischën Zahlentheorie,* Wiener Sitzungberichte, Math. Klasse **115** (1906), 589–632.

[125] E. Landau, *Neuer Beweis des Primzahlsatzes und Beweis des Primidealsatzes* (German), Math. Ann. **56** (1903), no. 4, 645–670, DOI 10.1007/BF01444310. MR1511191

[126] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen* (German), Teubner, Leipzig-Berlin, 1909.

[127] E. Landau, *Losung des Lehmer'schen Problems* (German), Amer. J. Math. **31** (1909), no. 1, 86–102, DOI 10.2307/2370180. MR1506062

[128] E. Landau, *Über die Wurzeln der Zetafunktion* (German), Math. Z. **20** (1924), no. 1, 98–104, DOI 10.1007/BF01188073. MR1544664

[129] E. Landau, *Über die Einteilung der positiven ganzen Zahlen in vier Klassen nach der Mindestzahl der zu ihrer additiven Zusammensetzung erforderlichen Quadrate*, Arch. Math. Phys. (3) **13** (1908), 305–312; Collected Works, Vol. 4, Essen:Thales Verlag, 1986, pp. 59–66.

[130] A. F. Lavrik, *The approximate functional equation for Dirichlet L-functions* (Russian), Trudy Moskov. Mat. Obšč. **18** (1968), 91–104. MR0236126

[131] U. V. Linnik, *"The large sieve"*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **30** (1941), 292–294. MR0004266

[132] U. V. Linnik, *On the least prime in an arithmetic progression. I. The basic theorem* (English, with Russian summary), Rec. Math. [Mat. Sbornik] N.S. **15(57)** (1944), 139–178. MR0012111

[133] U. V. Linnik, *On the least prime in an arithmetic progression. II. The Deuring-Heilbronn phenomenon* (English, with Russian summary), Rec. Math. [Mat. Sbornik] N.S. **15(57)** (1944), 347–368. MR0012112

[134] H. Maier, *Primes in short intervals*, Michigan Math. J. **32** (1985), no. 2, 221–225, DOI 10.1307/mmj/1029003189. MR783576

[135] H. Maier and C. Pomerance, *Unusually large gaps between consecutive primes*, Trans. Amer. Math. Soc. **322** (1990), no. 1, 201–237, DOI 10.2307/2001529. MR972703

[136] H. von Mangoldt, *Zu Riemanns Abhandlung "Ueber die Anzahl der Primzahlen unter einer gegebenen Grösse"* (German), J. Reine Angew. Math. **114** (1895), 255–305, DOI 10.1515/crll.1895.114.255. MR1580379

[137] D. A. Marcus, *Number fields*, Universitext, Springer, Cham, 2018. Second edition of [MR0457396]; With a foreword by Barry Mazur. MR3822326

[138] J. Maynard, *Small gaps between primes*, Ann. of Math. (2) **181** (2015), no. 1, 383–413, DOI 10.4007/annals.2015.181.1.7. MR3272929

[139] J. Maynard, *Large gaps between primes*, Ann. of Math. (2) **183** (2016), no. 3, 915–933, DOI 10.4007/annals.2016.183.3.3. MR3488739

[140] J. Maynard, *Dense clusters of primes in subsets*, Compos. Math. **152** (2016), no. 7, 1517–1554, DOI 10.1112/S0010437X16007296. MR3530450

[141] J. Maynard, *Primes represented by incomplete norm forms*, Preprint (2015), 56 pages, arXiv: 1507.05080.

[142] F. Mertens, *Ein Beitrag zur analytischen Zahlentheorie* (German), J. Reine Angew. Math. **78** (1874), 46–62, DOI 10.1515/crll.1874.78.46. MR1579612

[143] H. L. Montgomery, *Problems concerning prime numbers*, Mathematical developments arising from Hilbert problems (Proc. Sympos. Pure Math., Northern Illinois Univ., De Kalb, Ill., 1974), Amer. Math. Soc., Providence, R. I., 1976, pp. 307–310. MR0427249

[144] H. L. Montgomery, *Ten lectures on the interface between analytic number theory and harmonic analysis*, CBMS Regional Conference Series in Mathematics, vol. 84, Published for the Conference Board of the Mathematical Sciences, Washington, DC; by the American Mathematical Society, Providence, RI, 1994. MR1297543

[145] H. L. Montgomery and R. C. Vaughan, *The large sieve*, Mathematika **20** (1973), 119–134, DOI 10.1112/S0025579300004708. MR0374060

[146] H. L. Montgomery and R. C. Vaughan, *Multiplicative number theory. I. Classical theory*, Cambridge Studies in Advanced Mathematics, vol. 97, Cambridge University Press, Cambridge, 2007. MR2378655

[147] M. Nair, *On Chebyshev-type inequalities for primes*, Amer. Math. Monthly **89** (1982), no. 2, 126–129, DOI 10.2307/2320934. MR643279

[148] J. Neukirch, *Algebraic number theory*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 322, Springer-Verlag, Berlin, 1999. Translated from the 1992 German original and with a note by Norbert Schappacher; With a foreword by G. Harder. MR1697859

[149] R. E. A. C. Paley, *A theorem on characters*, J. London Math. Soc. **7** (1932), no. 1, 28–32, DOI 10.1112/jlms/s1-7.1.28. MR1574456

[150] J. Pintz, *Very large gaps between consecutive primes*, J. Number Theory **63** (1997), no. 2, 286–301, DOI 10.1006/jnth.1997.2081. MR1443763

[151] D. H. J. Polymath, *Variants of the Selberg sieve, and bounded intervals containing many primes*, Res. Math. Sci. **1** (2014), Art. 12, 83, DOI 10.1186/s40687-014-0012-7. MR3373710

[152] R. A. Rankin, *The difference between consecutive prime numbers*, J. London Math. Soc. **11** (1936), no. 4, 242–245, DOI 10.1112/jlms/s1-13.4.242. MR1574971

[153] R. A. Rankin, *The difference between consecutive prime numbers. V*, Proc. Edinburgh Math. Soc. (2) **13** (1962/1963), 331–332, DOI 10.1017/S0013091500025633. MR0160767

[154] B. Riemann, *Über die Anzahl der Primzahlen unter einer gegebenen Grösse*, Monatsberichte der Berliner Akademie. In Gesammelte Werke, Teubner, Leipzig (1892), Reprinted by Dover, New York (1953). Original manuscript (with English translation). Reprinted in (Borwein et al. 2008) and (Edwards 1974).

[155] B. Riemann, Unpublished papers, Handschriftenabteilung Niedersächische Staatsund Universitätsbibliotek, Göttingen.

[156] G. Rodriquez, *Sul problema dei divisori di Titchmarsh* (Italian, with English summary), Boll. Un. Mat. Ital. (3) **20** (1965), 358–366. MR0197409

[157] M. Rubinstein and P. Sarnak, *Chebyshev's bias*, Experiment. Math. **3** (1994), no. 3, 173–197. MR1329368

[158] W. Rudin, *Principles of mathematical analysis*, 3rd ed., International Series in Pure and Applied Mathematics, McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976. MR0385023

[159] W. Rudin, *Real and complex analysis*, 3rd ed., McGraw-Hill Book Co., New York, 1987. MR924157

[160] L. G. Sathe, *On a problem of Hardy on the distribution of integers having a given number of prime factors. II*, J. Indian Math. Soc. (N.S.) **17** (1953), 83–141. MR0058632

[161] A. Selberg, *On the normal density of primes in small intervals, and the difference between consecutive primes*, Arch. Math. Naturvid. **47** (1943), no. 6, 87–105. MR12624

[162] A. Selberg, *Note on a paper by L. G. Sathe*, J. Indian Math. Soc. (N.S.) **18** (1954), 83–87. MR0067143

[163] A. Selberg, *Collected papers. Vol. II*, Springer-Verlag, Berlin, 1991. With a foreword by K. Chandrasekharan. MR1295844

[164] C. L. Siegel, *Über Riemanns Nachlass zur analytischen Zahlentheorie*, Quellen Studien zur Geschichte der Math. Astron. und Phys. Abt. B: Studien **2** (1932), 45–80.

[165] P. Shiu, *A Brun-Titchmarsh theorem for multiplicative functions*, J. Reine Angew. Math. **313** (1980), 161–170, DOI 10.1515/crll.1980.313.161. MR552470

[166] K. Soundararajan, *Small gaps between prime numbers: the work of Goldston-Pintz-Yıldırım*, Bull. Amer. Math. Soc. (N.S.) **44** (2007), no. 1, 1–18, DOI 10.1090/S0273-0979-06-01142-6. MR2265008

[167] E. M. Stein and R. Shakarchi, *Real analysis: Measure theory, integration, and Hilbert spaces*, Princeton Lectures in Analysis, vol. 3, Princeton University Press, Princeton, NJ, 2005. MR2129625

[168] D. W. Stroock, *Probability theory: An analytic view*, 2nd ed., Cambridge University Press, Cambridge, 2011. MR2760872

[169] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245, DOI 10.4064/aa-27-1-199-245. Collection of articles in memory of Juriĭ Vladimirovič Linnik. MR0369312

[170] T. Tao, *The parity problem is sieve methods*, blog post (2007). URL: `https://terrytao.wordpress.com/2007/06/05/open-question-the-parity-problem-in-sieve-theory/`

[171] T. Tao, *Polymath8b: Bounded intervals with many primes, after Maynard*, blog post (2013). URL: `https://terrytao.wordpress.com/2013/11/19/polymath8b-bounded-intervals-with-many-primes-after-maynard/`

[172] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, 3rd ed., Graduate Studies in Mathematics, vol. 163, American Mathematical Society, Providence, RI, 2015. Translated from the 2008 French edition by Patrick D. F. Ion. MR3363366

[173] E. C. Titchmarsh, *The theory of functions*, Oxford University Press, Oxford, 1958. Reprint of the second (1939) edition. MR3155290

[174] E. C. Titchmarsh, *The theory of the Riemann zeta-function*, 2nd ed., The Clarendon Press, Oxford University Press, New York, 1986. Edited and with a preface by D. R. Heath-Brown. MR882550

[175] C. de la Vallée Poussin, *Recherches analytiques sur la théorie des nombres premiers, I-III*, Ann. Soc. Sci. Bruxelles **20** (1896), 183–256, 281–362, 363–397.

[176] A. I. Vinogradov, *The density hypothesis for Dirichet L-series* (Russian), Izv. Akad. Nauk SSSR Ser. Mat. **29** (1965), 903–934. MR0197414

[177] I. M. Vinogradov, *Representation of an odd number as a sum of three primes*, C. R. Acad. Sci. URSS **15** (1937), 6–7.

[178] I. M. Vinogradov, *Simplest trigonometrical sums with primes*, C. R. (Doklady) Acad. Sci. URSS (N.S.) **23** (1939), 615–617. MR0001763

[179] I. M. Vinogradov, *On the estimations of some simplest trigonometrical sums involving prime numbers* (Russian), Bull. Acad. Sci. URSS. Sér Math. [Izvestia Akad. Nauk SSSR] (1939), 371–398.

[180] I. M. Vinogradov, *The method of trigonometrical sums in the theory of numbers* (Russian), Trav. Inst. Math. Stekloff **23** (1947), 109 pp.

[181] I. M. Vinogradov, *A new estimate of the function* $\zeta(1 + it)$ (Russian), Izv. Akad. Nauk SSSR. Ser. Mat. **22** (1958), 161–164. MR0103861

[182] A. Walfisz, *Weylsche Exponentialsummen in der neueren Zahlentheorie* (German), Mathematische Forschungsberichte, XV, VEB Deutscher Verlag der Wissenschaften, Berlin, 1963. MR0220685

[183] E. W. Weisstein, *Bessel function of the first kind*, from MathWorld–A Wolfram Web Resource. `http://mathworld.wolfram.com/BesselFunctionoftheFirstKind.html`

[184] E. W. Weisstein, *Brun's constant*, from MathWorld–A Wolfram Web Resource. `http://mathworld.wolfram.com/BrunsConstant.html`

[185] E. Westzynthius, *Uber die Verteilung der Zahlen, die zu den n ersten Primzahlen teilerfremd sind*, Comm. Phys. Math. Soc. Sci. Fenn. **25** (1931).

[186] E. Wirsing, *Das asymptotische Verhalten von Summen über multiplikative Funktionen* (German), Math. Ann. **143** (1961), 75–102, DOI 10.1007/BF01351892. MR0131389

[187] E. Wirsing, *Das asymptotische Verhalten von Summen über multiplikative Funktionen. II* (German), Acta Math. Acad. Sci. Hungar. **18** (1967), 411–467, DOI 10.1007/BF02280301. MR0223318

[188] Y. Zhang, *Bounded gaps between primes*, Ann. of Math. (2) **179** (2014), no. 3, 1121–1174, DOI 10.4007/annals.2014.179.3.7. MR3171761

Index