ANALYTIC NUMBER THEORY

THOMAS F. BLOOM

These are lecture notes for the Part III lecture course given in Lent Term 2019. They are meant to be a faithful copy of the material given in lectures, with some supplementary footnotes and historical notes. The lectures themselves are the guide for what material is examinable, and any additional material in these printed notes will be marked as non-examinable. In the case of any doubt, ask the lecturer.

As these are informal lecture notes, I have not given proper references. My main source when compiling these notes, and the recommended textbook for the course, is Multiplicative Number Theory by Montgomery and Vaughan.

If you have any questions, concerns, corrections, please contact the lecturer at `tb634@cam.ac.uk`.

What is analytic number theory?

Analytic number theory is the study of the integers using techniques from analysis, both real and complex.

At first sight this may seem paradoxical – how can the continuous methods of analysis be useful for studying discrete objects? It is remarkable that not only can they be useful, they are often the most successful techniques that we have. In this course we will cover a variety of different methods, and see what they lead to.

We will first give some examples of the kinds of questions that analytic number theory tries to answer. These are usually quantitative questions (e.g. 'how many', or 'how large') asked about simple number theoretic objects, particularly prime numbers.

(1) How many prime numbers are there? One of the first mathematical results was Euclid's theorem that there are infinitely many prime numbers. We can ask for a more precise counting result: if $\pi(x)$ denotes the number of primes p satisfying $1 \leq p \leq x$ then can we give an asymptotic formula for $\pi(x)$? The Prime Number Theorem, one of the great accomplishments of analytic number theory, gives such an asymptotic:

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

This formula, first conjectured (independently) by Legendre and Gauss in the 18th century, was proved (independently) by Hadamard and de la Vallée Poussin in 1896, using complex analysis.

(2) How many twin primes are there? That is, primes p such that $p+2$ is also prime. Unlike the primes themselves, it is not even known whether there are infinitely many such primes, let alone what the precise count is. We can make a reasonable guess, however – the prime number theorem roughly says that the 'probability' that a number $1 \leq n \leq x$ is prime is about $1/\log x$,

so the number of pairs $n, n+2 \leq x$ which are both prime should be about

$$\pi_2(x) \approx \frac{x}{(\log x)^2}.$$

Of course, we cannot prove that this is correct – in this course, however, using sieve methods, we will prove that the implicit upper bound is correct, giving evidence that this is the right guess.

(3) Given $(a,q) = 1$, how many primes are there congruent to a modulo q? Dirichlet was the first to prove that there are infinitely many, in 1837. Once again, we ask the more refined question as to how many such primes are in the interval $[1,x]$. Since all such primes must be in one of the $\phi(q)$ many residue classes modulo q that are coprime to q, we might guess that the primes are evenly distributed amongst them, so that

$$\pi(x;a,q) \approx \frac{1}{\phi(q)}\frac{x}{\log x}.$$

This has been shown to be true for small q by Siegel and Walfisz, and we will prove this in the final section of the course.

Outline of the course

The course will be divided into four roughly equal parts:

(1) Elementary Techniques
(2) Sieve Methods
(3) Riemann's zeta function and the prime number theorem
(4) Primes in arithmetic progressions

The first two parts will use only elementary methods – that is, just real analysis, and no complex analysis. Complex analysis will be used often in the last two parts.

CHAPTER 1

Elementary Techniques

Review of asymptotic notation. We write $f(x) = O(g(x))$ if there exists some constant $C > 0$ such that $|f(x)| \leq C\,|g(x)|$ for all sufficiently large x. We will also use the Vinogradov notation $f \ll g$ to denote the same thing (so that $f = O(g)$ and $f \ll g$ are equivalent).

We write $f(x) = o(g(x))$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$. We write $f \sim g$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$. Observe that

$$f \sim g \text{ if and only if } f = (1 + o(1))g.$$

1. Arithmetic functions

An arithmetic function is simply a function on the natural numbers[1], $f : \mathbb{N} \to \mathbb{R}$. An arithmetic function is multiplicative if

$$f(nm) = f(n)f(m) \text{ whenever } (n, m) = 1,$$

and is completely multiplicative if $f(nm) = f(n)f(m)$ for all $n, m \in \mathbb{N}$.

An important operation on the space of arithmetic functions is that of multiplicative convolution:

$$f \star g(n) = \sum_{ab=n} f(a)g(b).$$

If f and g are both multiplicative functions, then so too is $f \star g$. The most obvious arithmetic function is the constant function:

$$\mathbf{1}(n) = 1.$$

We recall the definition of the Möbius function,

$$\mu(n) = \begin{cases} (-1)^k & \text{if } n = p_1 \cdots p_k \text{ where } p_i \text{ are distinct primes and} \\ 0 & \text{otherwise (i.e. if } n \text{ is divisible by a square).} \end{cases}$$

A fundamental relationship is that of Möbius inversion, which says that the Möbius function acts as an inverse to multiplicative convolution:

$$\mathbf{1} \star f = g \text{ if and only if } \mu \star g = f.$$

A great deal of analytic number theory is concerned with a deep study of the distribution of the prime numbers. For this the 'correct' way to count primes is not, as one might expect, the indicator function

$$1_{\mathbb{P}}(n) = \begin{cases} 1 & \text{if } n \text{ is prime, and} \\ 0 & \text{otherwise,} \end{cases}$$

[1]For the purposes of this course, 0 is not a natural number.

but instead the von Mangoldt function, which firstly also counts prime powers p^k, but also counts them not with weight 1, but with weight $\log p$ instead:

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

The main reason that this function is much easier to work with than $1_{\mathbb{P}}$ directly, is the following identity.

Lemma 1.

$$\mathbf{1} \star \Lambda(n) = \log n \text{ and } \log \star \mu(n) = \Lambda(n).$$

Proof. The second identity follows from the first by Möbius inversion. To establish the first, if we let $n = p_1^{k_1} \cdots p_r^{k_r}$, then

$$\begin{aligned} \mathbf{1} \star \Lambda(n) &= \sum_{i=1}^{r} \sum_{j=1}^{k_i} \log p_i \\ &= \sum_{i=1}^{r} \log p_i^{k_i} \\ &= \log n. \end{aligned}$$

□

2. Summation

A major theme of analytic number theory is understanding the basic arithmetic functions, particularly how large they are on average, which means understanding $\sum_{n \leq x} f(n)$. For example, if f is the indicator function of primes, then this summatory function is precisely the prime counting function $\pi(n)$.

We say that f has average order g if

$$\sum_{n \leq x} f(n) \sim x g(x).$$

One of the most useful tools in dealing with summations is partial summation, which is a discrete analogue of integrating by parts.

Theorem 1 (Partial summation)**.** *If a_n is any sequence of complex numbers and $f : \mathbb{R}_+ \to \mathbb{R}$ is such that f' is continuous then*

$$\sum_{1 \leq n \leq x} a_n f(n) = A(x) f(x) - \int_1^x A(t) f'(t) \, \mathrm{d}t,$$

where $A(x) = \sum_{1 \leq n \leq x} a_n$.

Proof. Let $N = \lfloor x \rfloor$. Using $a_n = A(n) - A(n-1)$

$$\begin{aligned} \sum_{1 \leq n \leq N} a_n f(n) &= \sum_{n=1}^{N} f(n)(A(n) - A(n-1)) \\ &= f(N)A(N) - \sum_{n=1}^{N-1} A(n)(f(n+1) - f(n)). \end{aligned}$$

We now observe that

$$\int_n^{n+1} f'(x)\,\mathrm{d}x = f(n+1) - f(n),$$

and so, since $A(x)$ is constant for $x \in [n, n+1)$,

$$\sum_{1\le n\le N} a_n f(n) = f(N)A(N) - \sum_{n=1}^{N-1} \int_n^{n+1} A(x)f'(x)\,\mathrm{d}x,$$

and the result follows since if $N \le x < N+1$ then

$$A(x)f(x) = A(N)f(x) = A(N)\left(f(N) + \int_N^x f'(x)\,\mathrm{d}x\right).$$

□

This is extremely useful even when the coefficients a_n are identically 1, when $A(x) = \lfloor x \rfloor = x + O(1)$.

Lemma 2.

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

where $\gamma = 0.577\cdots$ is a constant, known as Euler's constant.

Proof. By partial summation

$$\begin{aligned}\sum_{n\le x} \frac{1}{n} &= \frac{\lfloor x \rfloor}{x} + \int_1^x \frac{\lfloor t \rfloor}{t^2}\,\mathrm{d}t \\ &= 1 + \int_1^x \frac{1}{t}\,\mathrm{d}t + \int_1^\infty \frac{\{t\}}{t^2}\,\mathrm{d}t - \int_x^\infty \frac{\{t\}}{t^2}\,\mathrm{d}t + O(1/x) \\ &= \log x + \left(1 + \int_1^\infty \frac{\{t\}}{t^2}\,\mathrm{d}t\right) + O(1/x).\end{aligned}$$

It remains to note that the second term is a constant, since the integral converges. □

It is remarkable how little we understand about Euler's constant – it is not even known whether it is irrational or not.

Lemma 3.

$$\sum_{1\le n\le x} \log n = x\log x - x + O(\log x).$$

Proof. By partial summation

$$\begin{aligned}\sum_{n\le x} \log n &= \lfloor x \rfloor \log x - \int_1^x \frac{\lfloor t \rfloor}{t}\,\mathrm{d}t \\ &= x\log x - x + O(\log x).\end{aligned}$$

□

3. Divisor function

We now turn our attention to number theory proper, and examine one of those most important arithmetic functions: the divisor function[2]

$$\tau(n) = \mathbf{1} \star \mathbf{1}(n) = \sum_{ab=n} 1 = \sum_{d|n} 1.$$

We first find its average order.

Theorem 2.

$$\sum_{n\leq x} \tau(n) = x\log x + (2\gamma - 1)x + O(x^{1/2}).$$

In particular, the average order of $\tau(n)$ is $\log n$.

Proof. A first attempt might go as follows:

$$\begin{aligned}\sum_{n\leq x} \tau(n) &= \sum_{ab\leq x} 1\\ &= \sum_{a\leq x} \left\lfloor \frac{x}{a}\right\rfloor\\ &= x\sum_{a\leq x}\frac{1}{a} + O(x)\\ &= x\log x + \gamma x + O(x).\end{aligned}$$

The problem is that the second term γx is lost in the error term $O(x)$. To improve the error term we use what is known as the hyperbola method, which is the observation that when summing over pairs (a,b) such that $ab\leq x$ we can express this as the sum over pairs where $a\leq x^{1/2}$ and where $b\leq x^{1/2}$, and then subtract the contribution where $\max(a,b)\leq x^{1/2}$.

$$\begin{aligned}\sum_{ab\leq x} 1 &= \sum_{a\leq x^{1/2}} \left\lfloor \frac{x}{a}\right\rfloor + \sum_{b\leq x^{1/2}} \left\lfloor \frac{x}{b}\right\rfloor - \sum_{a,b\leq x^{1/2}} 1\\ &= 2x\sum_{a\leq x^{1/2}}\frac{1}{a} - \lfloor x^{1/2}\rfloor^2 + O(x^{1/2})\\ &= x\log x + (2\gamma-1)x + O(x^{1/2}).\end{aligned}$$

□

It is a deep and difficult problem to improve the error term here – the truth is probably $O(x^{1/4+\epsilon})$, but this is an open problem, and the best known is $O(x^{0.3149\cdots})$.

We have just shown that the 'average' number of divisors of n is $\log n$. The worst case behaviour can differ dramatically from this average behaviour, however.

Theorem 3. *For any $n\geq 1$,*

$$\tau(n) \leq n^{O(1/\log\log n)}.$$

In particular, for any $\epsilon > 0$, $\tau(n) = O_\epsilon(n^\epsilon)$.

[2]Alternative notation used in some places is $d(n)$ or $\sigma_0(n)$.

Proof. Let $0 < \epsilon < 1/2$ be something to be chosen later. Let $n = p_1^{k_1} \cdots p_r^{k_r}$, so that $\tau(n) = (k_1 + 1) \cdots (k_r + 1)$, and hence

$$\frac{\tau(n)}{n^\epsilon} = \prod_{i=1}^{r} \frac{k_i + 1}{p_i^{k_i \epsilon}}.$$

If $p > 2^{1/\epsilon}$ then

$$\frac{k+1}{p^{k\epsilon}} \leq \frac{k+1}{2^k} \leq 1$$

and if $p \leq 2^{1/\epsilon}$ then, using the inequality $x + 1/2 \leq 2^x$, valid for all $x \geq 0$, and the fact that $p \geq 2$,

$$\frac{k+1}{p^{k\epsilon}} \leq \frac{1}{\epsilon}.$$

Trivially bounding the number of primes $p \leq 2^{1/\epsilon}$ by $2^{1/\epsilon}$, it follows that

$$\tau(n) \leq n^\epsilon \epsilon^{-2^{1/\epsilon}}.$$

This already shows the inequality $\tau(n) \ll_\epsilon n^\epsilon$ for any fixed $0 < \epsilon < 1/2$. To be more precise, we choose $\epsilon = c/\log\log n$ for some small enough constant $c > 0$, from which the first result follows. □

4. Estimates on prime numbers

The prime number theorem is the statement that

$$\pi(x) \sim \frac{x}{\log x},$$

or, equivalently (we will justify this equivalence soon),

$$\psi(x) = \sum_{n \leq x} \Lambda(n) \sim x.$$

The proof of this is surprisingly involved, and we will return to it later in the course when we examine the Riemann zeta function.

It is much easier to show, if not an asymptotic formula, at least that this is the correct rate of growth of the function. This was proved in 1850 by Chebyshev.

Theorem 4 (Chebyshev)**.**

$$\psi(x) \asymp x.$$

Proof. We will first prove the lower bound. This relies on the observation that, for any $x \geq 1$, $\lfloor x \rfloor \leq 2\lfloor x/2 \rfloor + 1$.

$$\begin{aligned}
\psi(x) &= \sum_{n \leq x} \Lambda(n) \\
&\geq \sum_{n \leq x} \Lambda(n) \left(\left\lfloor \frac{x}{n} \right\rfloor - 2 \left\lfloor \frac{x}{2n} \right\rfloor \right) \\
&= \sum_{nm \leq x} \Lambda(n) - 2 \sum_{nm \leq x/2} \Lambda(n) \\
&= \sum_{n \leq x} \log n - 2 \sum_{n \leq x/2} \log n.
\end{aligned}$$

By Lemma 3,

$$\psi(x) \geq x\log x - x + O(\log x) - 2\left(\frac{x}{2}\log(x/2) - \frac{x}{2} + O(\log x)\right) = (\log 2)x + O(\log x).$$

It follows that, for any $c > 0$ and x sufficiently large, $\psi(x) \geq (\log 2 - c)x$, and hence $\psi(x) \gg x$.

For the upper bound, we do something very similar, except we note that for $x \in [1,2)$ we have equality $\lfloor x \rfloor = 2\lfloor x/2 \rfloor + 1$. Furthermore, for any $x \geq 1$, we have the lower bound $\lfloor x \rfloor \geq 2\lfloor x/2 \rfloor$. It follows that

$$\begin{aligned}\sum_{x/2<n\leq x}\Lambda(n) &= \sum_{x/2<n\leq x}\Lambda(n)\left(\left\lfloor\frac{x}{n}\right\rfloor - 2\left\lfloor\frac{x}{2n}\right\rfloor\right)\\ &\leq \sum_{n\leq x}\Lambda(n)\left(\left\lfloor\frac{x}{n}\right\rfloor - 2\left\lfloor\frac{x}{2n}\right\rfloor\right)\\ &= (\log 2)x + O(\log x)\end{aligned}$$

by the above calculation. The left hand side is $\psi(x) - \psi(x/2)$, and so we have shown that

$$\psi(x) - \psi(x/2) \leq (\log 2)x + O(\log x).$$

Using the fact that $\psi(x) = 0$ for any $x \leq 1$,

$$\psi(x) = \sum_{k=0}^{\lceil \log_2 x\rceil}(\psi(x/2^k) - \psi(x/2^{k+1})) \leq (2\log 2)x + O((\log x)^2),$$

and hence $\psi(x) \ll x$ as required. □

Chebyshev's estimate is the first non-trivial quantitative information we have about the primes, and leads to a host of other facts about the primes – rather surprisingly, not just big-oh behaviour, but precise asymptotic results.

Lemma 4.

$$\pi(x) = \frac{\psi(x)}{\log x} + O\left(\frac{x}{(\log x)^2}\right).$$

In particular, $\pi(x) \asymp x/\log x$, and $\pi(x) \sim x/\log x$ if and only if $\psi(x) \sim x$.

Proof. We first remove the contribution from prime powers by noting that, if $\theta(x) = \sum_{p\leq x}\log p$, then

$$\psi(x) - \theta(x) \leq \sum_{k\geq 2}\sum_{p\leq x^{1/k}}\log p \ll \sum_{k=2}^{\lceil \log x\rceil} x^{1/k} \ll x^{1/2}\log x.$$

It follows that $\theta(x) = \psi(x) + O(x^{1/2}\log x)$. We apply partial summation with $a_n = \Lambda(n)$ if n is prime, and 0 otherwise, and $f(n) = \frac{1}{\log n}$. This gives

$$\pi(x) = \sum_{p\leq x}1 = \frac{\theta(x)}{\log x} + \int_2^x \frac{\theta(t)}{t(\log t)^2}\,\mathrm{d}t = \frac{\theta(x)}{\log x} + O\left(\frac{x}{(\log x)^2}\right).$$

□

Lemma 5.

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

Proof. Recalling that $\log = \mathbf{1} \star \Lambda$, and using Lemma 3,

$$\begin{aligned} x \log x + O(x) &= \sum_{n \leq x} \log n \\ &= \sum_{ab \leq x} \Lambda(b) \\ &= x \sum_{b \leq x} \frac{\Lambda(b)}{b} + O(\psi(x)). \end{aligned}$$

Using Chebyshev's estimate, this proves that

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

It remains to deal with the contribution from prime powers $p^k \leq x$ for $k \geq 2$, which we bound trivially by

$$\sum_{p \leq x^{1/2}} \log p \sum_{k \geq 2} \frac{1}{p^k} = \sum_{p \leq x^{1/2}} \log p \frac{1}{p^2 - p} \ll 1.$$

□

Lemma 6.

$$\sum_{p \leq x} \frac{1}{p} = \log \log x + b + O(1/\log x),$$

where b is some constant.

Proof. Let $A(x) = \sum_{p \leq x} (\log p)/p = \log x + R(x)$, say, where $R(x) = O(1)$. By partial summation

$$\begin{aligned} \sum_{p \leq x} \frac{1}{p} &= \frac{A(x)}{\log x} + \int_2^x \frac{A(t)}{t(\log t)^2} \, \mathrm{d}t \\ &= 1 + O(1/\log x) + \int_2^x \frac{1}{t \log t} \, \mathrm{d}t + \int_2^x \frac{R(t)}{t(\log t)^2} \, \mathrm{d}t \\ &= \log \log x + 1 - \log \log 2 + \int_2^\infty \frac{R(t)}{t(\log t)^2} \, \mathrm{d}t + O(1/\log x). \end{aligned}$$

□

Lemma 7.

$$\prod_{p \leq x} \left(1 - \frac{1}{p}\right)^{-1} = c \log x + O(1)$$

where $c > 1$ is some constant.

Proof. We use $\log(1-t) = -\sum_{k=1}^{\infty} \frac{t^k}{k}$ to deduce that

$$\begin{aligned}\log\left(\prod_{p\leq x}\left(1-\frac{1}{p}\right)^{-1}\right) &= -\sum_{p\leq x}\log(1-1/p)\\ &= \sum_{k=1}^{\infty}\sum_{p\leq x}\frac{1}{kp^k}\\ &= \sum_{p\leq x}\frac{1}{p} + \sum_{k\geq 2}\sum_{p\leq x}\frac{1}{kp^k}\\ &= \log\log x + b' + O(1/\log x).\end{aligned}$$

The result follows since $e^x = 1 + O(x)$ for $|x| \leq 1$. □

It is a little tricky to determine what the constant c in Lemma 7 actually is – it turns out to be $e^{\gamma} \approx 1.78\cdots$. We can use this fact to point out why the naive probabilistic heuristic can be misleading (and hopefully give some idea why the prime number theorem itself, unlike these simple asymptotics, is hard to prove).

As a heuristic, we might guess that the probability that a given prime number p divides a randomly chosen n is $1/p$. Furtermore, we expect that these probabilities should be independent for distinct primes p. Using the fact that $n \geq 3$ is prime if and only if $p \nmid n$ for all $2 \leq p \leq n^{1/2}$, we might guess that

$$1_{n \text{ is prime}} \approx \mathbb{P}(p \nmid n \text{ for all } 2 \leq p \leq n^{1/2}) \approx \prod_{p\leq n^{1/2}}\left(1-\frac{1}{p}\right) \approx 2e^{-\gamma}/\log n.$$

This would in turn suggest that

$$\pi(x) = \sum_{n\leq x} 1_{n \text{ is prime}} \approx 2e^{-\gamma}\sum_{n\leq x}\frac{1}{\log n} \approx 2e^{-\gamma}\frac{x}{\log x}.$$

But since $2e^{-\gamma} = 1.12\cdots$, this contradicts the prime number theorem! This shows that, while heuristically thinking about discrete concepts in terms of 'probability' can lead to roughly the right order of magnitude, one must take care not to take the constants obtained too seriously!

Indeed, we can use the elementary estimates already obtained to show, not the prime number theorem itself, but at least the fact that *if* $\frac{\pi(x)\log x}{x}$ *converges to a limit at all*, then this limit must be 1, and hence the prime number theorem is true. The hard part is showing that the limit exists.

Theorem 5 (Chebyshev)**.** *If* $\pi(x) \sim c\frac{x}{\log x}$ *then* $c = 1$.

Proof. By partial summation,

$$\sum_{p\leq x}\frac{1}{p} = \frac{\pi(x)}{x} + \int_1^x \frac{\pi(t)}{t^2}\,\mathrm{d}t.$$

The first term is trivially $O(1)$. If $\pi(x) = c(1+R(x))\frac{x}{\log x}$, where $R(x) \to 0$ as $x \to \infty$, then

$$\sum_{p\leq x}\frac{1}{p} = c\int_1^x \frac{1+R(t)}{t\log t}\,\mathrm{d}t + O(1) = c(1+o(1))\log\log x + O(1).$$

By Lemma 6 the left hand side is $(1+o(1))\log\log x$, and hence $c = 1$. □

5. Selberg's elementary approach

Let

$$\Lambda_2(n) = \mu \star (\log)^2 = \sum_{ab=n} \mu(a)(\log b)^2.$$

Why have we defined such a function? We will prove another useful identity for Λ_2, which should also explain this peculiar choice of notation.

Lemma 8.

$$\Lambda_2(n) = \Lambda(n)\log n + \Lambda \star \Lambda(n).$$

In particular, $0 \leq \Lambda_2(n) \leq (\log n)^2$*, and if* $\Lambda_2(n) \neq 0$ *then* n *has at most two distinct prime divisors.*

Proof. We will prove this identity using Möbius inversion, so that it is enough to show that

$$\sum_{d|n} (\Lambda(d)\log d + \Lambda \star \Lambda(d)) = (\log n)^2.$$

To this end, we write the left hand side as

$$\begin{aligned}
\sum_{d|n} \Lambda(d)\log d + \sum_{d|n}\sum_{ab=d} \Lambda(a)\Lambda(b) &= \sum_{d|n} \Lambda(d)\log d + \sum_{a|n} \Lambda(a) \sum_{b|\frac{n}{a}} \Lambda(b) \\
&= \sum_{d|n} \Lambda(d)\log d + \sum_{a|n} \Lambda(a)\log(n/a) \\
&= \log n \sum_{d|n} \Lambda(d) \\
&= (\log n)^2
\end{aligned}$$

This concludes the proof of the identity. From this identity it is obvious that $\Lambda_2(n) \geq 0$, since both terms on the right hand side are, for all $n \geq 1$. Furthermore, if n has at least three distinct prime divisors then the first term is zero, as is the second, since in any decomposition $ab = n$ at least one of a or b has at least two distinct prime divisors, whence $\Lambda(a)\Lambda(b) = 0$.

Finally, since $\sum_{d|n} \Lambda_2(d) = (\log n)^2$, and $\Lambda_2(d) \geq 0$ for all d, it follows that $\Lambda_2(n) \leq (\log n)^2$ for all n. □

Theorem 6.

$$\sum_{n\leq x} \Lambda_2(n) = 2x\log x + O(x).$$

Proof. We have

$$\sum_{n\leq x} \Lambda_2(n) = \sum_{a\leq x} \mu(a) \sum_{b\leq x/a} (\log b)^2.$$

By partial summation

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

To handle this, we first establish by partial summation that

$$\sum_{n\leq x}\frac{\tau(n)}{n}=\frac{1}{2}(\log x)^2+c\log x+c'+O(x^{-1/2})$$
$$=\frac{1}{2}(\log x)^2+\frac{c_1}{x}\sum_{n\leq x}\tau(n)+c_2+O(x^{-1/2}).$$

It follows that the main sum is

$$=2\sum_{a\leq x}\mu(a)\left(\sum_{b\leq x/a}\tau(b)\frac{x}{ab}-c_1\sum_{b\leq x/a}\tau(b)-c_2\frac{x}{a}+O((x/a)^{-1/2})\right).$$

The error term here is

$$\ll x^{-1/2}\sum_{a\leq x}a^{1/2}\ll x.$$

The third term is

$$\ll\sum_{a\leq x}\mu(a)\frac{x}{a}=\sum_{a\leq x}\mu(a)\left\lfloor\frac{x}{a}\right\rfloor+O(x)=\sum_{a\leq x}\mu(a)\sum_{b\leq x/a}1+O(x)=\sum_{n\leq x}\mu\star 1(n)+O(x)\ll x.$$

The second term is

$$\ll\sum_{a\leq x}\mu(a)\sum_{bc\leq x/a}1=\sum_{b\leq x}\sum_{d\leq x/b}1\star\mu(d)\ll x.$$

Finally, the first term is

$$2x\sum_{n\leq x}\frac{\mu\star\tau(n)}{n}=2x\sum_{n\leq x}\frac{1}{n}=2x\log x+O(x).$$

□

A fourteen point plan to proving the prime number theorem (NON-EXAMINABLE)

In this section we give the main points on how to go from Selberg's identity to a full elementary proof of the prime number theorem. You are encouraged to try and follow this outline yourself and try to prove each of the points below, thereby producing a full proof. This will hopefully instill a proper respect for the power of elementary methods, and how involved they can get. Everything in this section is non-examinable, however.

We start by introducing some convenient notation. Let

$$r(x)=\frac{\psi(x)}{x}-1.$$

The prime number theorem is equivalent to $\lim_{x\to\infty}|r(x)|=0$.

(1) Deduce from Selberg's identity that

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

(2) By considering this identity with x replaced by x/m, show that

$$\left|r(x)(\log x)^2\right|\leq\sum_{n\leq x}\frac{\Lambda_2(n)}{n}\left|r(x/n)\right|+O(\log x).$$

(3) Show that
$$\sum_{n\le x}\Lambda_2(n) = 2\int_1^{\lfloor x\rfloor}\log t\,\mathrm{d}t + O(x).$$
(4) Show that
$$\sum_{n\le x}\frac{\Lambda_2(n)}{n}\,|r(x/n)| = 2\sum_{2\le n\le x}|r(x/n)|\,\frac{1}{n}\int_{n-1}^{n}\log t\,\mathrm{d}t + O(\log x).$$
(5) Show that
$$\sum_{2\le n\le x}|r(x/n)|\,\frac{1}{n}\int_{n-1}^{n}\log t\,\mathrm{d}t = \int_1^x \frac{1}{t\log t}\,|r(x/t)|\,\mathrm{d}t + O(\log x).$$
(6) Deduce that
$$\sum_{n\le x}\frac{\Lambda_2(n)}{n}\,|r(x/n)| = 2\int_1^x \frac{1}{t\log t}\,|r(x/t)|\,\mathrm{d}t + O(\log x).$$
(7) Now let $V(u) = r(e^u)$. Show that
$$u^2\,|V(u)| \le 2\int_0^u\int_0^v |V(t)|\,\mathrm{d}t\,\mathrm{d}v + O(u).$$
(8) Let $\alpha = \limsup|V(u)|$ (so that the prime number theorem is equivalent to $\alpha = 0$). Show that
$$\alpha \le \limsup\frac{1}{u}\int_0^u |V(t)|\,\mathrm{d}t.$$
(9) Show that, for every $u < v$,
$$\left|\int_u^v V(t)\,\mathrm{d}t\right| \ll 1.$$
(10) Show that if $u > 0$ and $V(u) = 0$ then
$$\int_0^\alpha |V(u+t)|\,\mathrm{d}t \le \frac{\alpha^2}{2} + O(1/u).$$
(11) Let $\delta > \alpha$ and consider the behaviour of $V(t)$ for $t \in [u, u+\delta-\alpha]$. Show that either $V(t) = 0$ for some t in this interval, or else $V(t)$ changes sign at most once.
(12) If $V(t) = 0$ for some $t \in [u, u+\delta-\alpha]$ show that
$$\int_u^{u+\delta} |V(t)|\,\mathrm{d}t \le \alpha(\delta - \alpha/2) + o(1).$$
(13) If $V(t)$ changes sign just once in $[u, u+\delta-\alpha]$, show that
$$\int_u^{u+\delta} |V(t)|\,\mathrm{d}t < \alpha^2 + O(1).$$
(14) If $\alpha > 0$, by choosing δ suitably, show that
$$\limsup\frac{1}{u}\int_0^u |V(t)|\,\mathrm{d}t < \alpha.$$
In particular, this contradicts (8) above, and hence $\alpha = 0$ and we have proved the prime number theorem.

CHAPTER 2

Sieve methods

The idea of sieve theory is to start with some set of integers (usually an interval) and remove, or 'sift out', those integers divisible by some set of primes. The classic example is the Sieve of Eratosthenes: beginning with the set of integers $\{1, \dots, n\}$, and removing all integers divisible by some prime $p \leq n^{1/2}$, we are left with only primes. By counting how many removals took place, we aim to find estimates for the count of what is left.

6. Preliminaries

We first introduce some important notation – we do this in some generality, for part of the virtue of sieve theory is that the same tools can be applied to study many different problems. Our basic data will be:

- a finite set A, the set which is to be sifted;
- a set of primes P, which we will sift by (usually this is just the set of all primes);
- a sifting limit z, which is an upper bound on how large the primes we sieve by;
- the sifting function
$$S(A, P; z) = \sum_{n \in A} 1_{(n, P(z))=1},$$
where $P(z) = \prod_{p \in P, p<z} p$, which counts those integers in A left after the sieve;
- we are able to count using sieve theory only if we are able to count A under various divisibility conditions - if
$$A_d = \{n \in A : d \mid n\}$$
then we suppose that
$$|A_d| = \frac{f(d)}{d} X + R_d,$$
where $f(d) \geq 0$ is some multiplicative function and R_d is thought of as an error term. In general, we will only need this estimate for squarefree d. By convention we suppose that $f(p) = 0$ if $p \notin P$;
- finally, the expected density is defined as
$$W(z) = W_P(z) = \prod_{\substack{p \in P \\ p<z}} \left(1 - \frac{f(p)}{p}\right).$$

We note that
$$|A| = |A_1| = \frac{f(1)}{1} X + R_1 = X + R_1,$$

so that X is an approximation to the size of our original set. We demonstrate this with two examples.

Example 1: The most basic example is sifting an interval by all primes. Here $A = \{x < n \leq x + y\}$, so that

$$|A_d| = \left\lfloor \frac{x+y}{d} \right\rfloor - \left\lfloor \frac{x}{d} \right\rfloor = \frac{y}{d} + O(1)$$

for all d. Thus we let $f \equiv 1$ and then this gives us an error bound of $R_d \ll 1$, and $X = y$ is a smooth count of how many integers are in A. The sifting function is

$$S(A, P; z) = \{x < n \leq x + y : p \mid n \implies p \geq z\}.$$

For example, if $x = 0$ and $z = y^{1/2}$, then

$$S(A, P; z) = \pi(y) - \pi(y^{1/2}) + 1.$$

Example 2: We can also use sieve methods to count the number of twin primes. For this we let $A = \{n(n+2) : 1 \leq n \leq x\}$ and P be the set of all primes except 2. We will show later that if d is odd and square-free then

$$|A_d| = \frac{2^{\omega(d)}}{d} x + O(2^{\omega(d)}),$$

so that we can take f to be the multiplicative function defined by $f(2) = 0$ and $f(p) = 2$ for all odd primes. The sifting function can be used to count the number of twin primes, since if either n or $n + 2$ is composite then it must be divisible by a prime at most $x^{1/2}$, and so

$$S(A, P; z) = \pi_2(x) + O(x^{1/2}).$$

The most basic sieve is the sieve of Eratosthenes which counts the sifting function by an inclusion-exclusion argument. That is, we first subtract from the count all integers divisible by a single sifted prime, then add back all those divisible by two primes, and so on. In analytic language this can be cleanly expressed using the identity

$$\sum_{d \mid n} \mu(d) = \begin{cases} 1 & \text{if } n = 1 \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 7 (Sieve of Eratosthenes-Legendre)**.**

$$S(A, P; z) = XW(z) + O\left(\sum_{d \mid P(z)} |R_d| \right).$$

Proof.

$$\begin{aligned}S(A,P;z) &= \sum_{n\in A}\sum_{d|(n,P(z))}\mu(d)\\ &= \sum_{d|P(z)}\mu(d)\sum_{n\in A}1_{d|n}\\ &= \sum_{d|P(z)}\mu(d)\,|A_d|\\ &= X\sum_{d|P(z)}\frac{\mu(d)f(d)}{d}+O\left(\sum_{d|P(z)}|R_d|\right)\\ &= X\prod_{\substack{p\in P\\ p<z}}\left(1-\frac{f(p)}{p}\right)+O\left(\sum_{d|P(z)}|R_d|\right).\end{aligned}$$

□

We can apply this to the interval as in Example 1 to get an upper bound for the number of primes in an interval. If we take $P=\prod_{p\leq z}p$ then

$$W(z)=\prod_{p\leq z}\left(1-\frac{1}{p}\right)\asymp 1/\log z$$

by Lemma 7. Choosing $z=\log y$ yields the following, which is striking in that the upper bound is completely independent of x.

Corollary 1. *For any $x,y\geq 1$,*

$$\pi(x+y)-\pi(x)\ll\frac{y}{\log\log y}.$$

Proof. Let $A=\{x<n\leq x+y\}$ and P be the set of all primes. Then

$$W(z)=\prod_{p<z}\left(1-\frac{1}{p}\right)\ll\frac{1}{\log z}$$

and $R_d=O(1)$. It follows that

$$\sum_{d|P(z)}|R(d)|\ll\tau(P(z))\ll 2^z$$

and so

$$|\{x<n\leq x+y:\text{if } p\mid n \text{ then } p\geq z\}|\ll\frac{y}{\log z}+2^z.$$

Choosing $z=\log y$ means that the right hand side is $\ll y/\log\log y$. We are done noting that any primes in $(x,x+y]$ must survive this sieving process (unless they are $\ll\log y$ themselves, which introduces an error of only $O(\log y)$). □

7. Selberg's sieve

The main weakness of the sieve of Eratosthenes is the large error term, which forces us to take z quite small. This is because we have to sift by all d dividing $P(z)$, the number of which grow exponentially with z. We can instead consider what happens if we only sieve up to some limit, so only considering the effect of

those $d < D$. We must then abandon any hope of getting an exact formula, but we can hope that we're still counting the main terms, and thus might get useful upper and lower bounds for the sifting function – and now with much smaller error terms.

One of the most successful sieves was developed by Selberg. We will return to lower bound sieves later on, and for now just focus on upper bounds. The key input in the sieve of Eratosthenes was the replacing of

$$1_{(n,P(z))=1} \quad \text{by} \quad \sum_{d|(n,P(z))} \mu(d).$$

If we are content with an upper bound then we'd be happy instead replacing it with any function $F(n)$ such that $F(1) \geq 1$ and $F(n) \geq 0$ for all n. The starting point for Selberg's sieve is the trivial observation that any sequence of real numbers $\lambda_d \in \mathbb{R}$ such that $\lambda_1 = 1$ can lead to such a function, since

$$\left(\sum_{d|n} \lambda_d\right)^2 \geq \begin{cases} 1 & \text{if } n = 1 \text{ and} \\ 0 & \text{if } n > 1. \end{cases}$$

We have complete freedom to choose λ_d to be whatever weights we wish (possibly depending on A and P) to make the right-hand side as small as possible, limited only by the restriction $\lambda_1 = 1$.

Before giving Selberg's sieve, we first make the assumption that

$$0 \leq f(p) < p \quad \text{for all } p,$$

and also that $f(p) \neq 0$ for $p \in P$. This allows us to make the useful definition of a multiplicative function g such that for all primes p,

$$g(p) = \left(1 - \frac{f(p)}{p}\right)^{-1} - 1 = \frac{f(p)}{p - f(p)},$$

extending this definition to all square-free d multiplicatively. Note that

$$\sum_{d|P(z)} g(d) = \prod_{\substack{p \in P \\ p<z}} (1 + g(p)) = \frac{1}{W(z)}.$$

Theorem 8 (Selberg's sieve)**.**

$$S(A, P; z) \leq \frac{X}{G(t, z)} + \sum_{\substack{d|P(z) \\ d<t^2}} 3^{\omega(d)} |R_d|$$

for all $t \geq 1$, where

$$G(t, z) = \sum_{\substack{d|P(z) \\ d<t}} g(d).$$

Proof. Let $\lambda_d \in \mathbb{R}$ be some sequence to be chosen later, with the only restriction that $\lambda_1 = 1$. By the observation above, using the notation $[d, e]$ for the least

common multiple of d and e,

$$\begin{aligned}S(A,P;z) &= \sum_{n\in A} 1_{(n,P(z))=1}\\ &\leq \sum_{n\in A}\left(\sum_{d|(n,P(z))}\lambda_d\right)^2\\ &= \sum_{d,e|P(z)}\lambda_d\lambda_e\sum_{n\in A}1_{[d,e]|n}\\ &= XV+R,\end{aligned}$$

say, where

$$V=\sum_{d,e|P(z)}\lambda_d\lambda_e\frac{f([d,e])}{[d,e]}$$

and

$$R=\sum_{d,e|P(z)}\lambda_d\lambda_e R_{[d,e]}.$$

We will first examine the main term V. If we write $[d,e]=abc$ where $c=(d,e)$ and $d=ac$ and $e=bc$, so that $(a,b)=(b,c)=(a,c)=1$, then using the fact that f is multiplicative,

$$\begin{aligned}V &= \sum_{c|P(z)}\frac{f(c)}{c}\sum_{\substack{ab|P(z)/c\\(a,b)=1}}\frac{f(a)f(b)}{ab}\lambda_{ac}\lambda_{bc}\\ &= \sum_{c|P(z)}\frac{f(c)}{c}\sum_{ab|P(z)/c}\frac{f(a)f(b)}{ab}\lambda_{ac}\lambda_{bc}\sum_{d|(a,b)}\mu(d)\\ &= \sum_{c|P(z)}\frac{f(c)}{c}\sum_{d|P(z)}\mu(d)\left(\sum_{d|m|P(z)/c}\frac{f(m)}{m}\lambda_{cm}\right)^2\\ &= \sum_{d|P(z)}\mu(d)\sum_{c|P(z)/d}\frac{c}{f(c)}\left(\sum_{cd|n|P(z)}\frac{f(n)}{n}\lambda_n\right)^2\\ &= \sum_{m|P(z)}\left(\sum_{c|m}\mu(m/c)\frac{c}{f(c)}\right)\left(\sum_{m|n|P(z)}\frac{f(n)}{n}\lambda_n\right)^2.\end{aligned}$$

For primes $p\in P$, we note that

$$1+\frac{1}{g(p)}=1+\frac{p-f(p)}{f(p)}=\frac{p}{f(p)}.$$

In particular, the identity

$$\sum_{d|n}\frac{1}{g(d)}=\frac{n}{f(n)}$$

holds for all primes $n \in P$ and thus by multiplicativity also for all $n \mid P(z)$. It follows by Möbius inversion that

$$\sum_{c|m} \mu(m/c)\frac{c}{f(c)} = \frac{1}{g(m)}.$$

It's convenient to introduce the sifting limit t at this point, so we will assume that λ_d is chosen so that $\lambda_d = 0$ whenever $d \geq t$. Then

$$\begin{aligned} V &= \sum_{m|P(z)} \frac{1}{g(m)} \left(\sum_{m|n|P(z)} \frac{f(n)}{n} \lambda_n \right)^2 \\ &= \sum_{\substack{k|P(z) \\ k<t}} \frac{y_k^2}{g(k)}, \end{aligned}$$

say. Since we are trying to produce an upper bound here, we want to choose λ_d (and hence y_k) such that V is as small as possible.

What is the relationship between λ_d and y_k? First note that $y_k = 0$ for $k \geq t$. For fixed d,

$$\begin{aligned} \sum_{k|P(z)} \mu(k) y_k 1_{d|k} &= \sum_{k|P(z)} \mu(k) \sum_{e|P(z)} \frac{f(e)\lambda_e}{e} 1_{k|e} 1_{d|k} \\ &= \sum_{e|P(z)} \frac{f(e)\lambda_e}{e} 1_{d|e} \sum_{k|e/d} \mu(k) \\ &= \mu(d)\frac{f(d)\lambda_d}{d}. \end{aligned}$$

In particular,

$$\sum_{k|P(z)} \mu(k) y_k = \lambda_1 = 1.$$

Since

$$1 = \sum_{\substack{k|P(z) \\ k<t}} \mu(k) y_k = \sum_{\substack{k|P(z) \\ k<t}} \mu(k) g(k)^{1/2} \cdot \frac{y_k}{g(k)^{1/2}},$$

by the Cauchy-Schwarz inequality, it follows that

$$1 \leq \left(\sum_{k|P(z)} \frac{y_k^2}{g(k)} \right) \left(\sum_{\substack{k|P(z) \\ k<t}} g(k) \right) = G(t,z)V.$$

In particular, $V \geq G(t,z)^{-1}$. Furthermore, we can achieve equality if there is some constant c such that

$$y_k = c\mu(k)g(k).$$

Using $\lambda_1 = 1$ as above again, we must have $c = G(t,z)^{-1}$. This choice determines all y_k (for $k < t$, for $k \geq t$ we choose $y_k = 0$), and hence all λ_d. The only thing left to do is to control the error term. We will show that $|\lambda_d| \leq 1$. The exact expression for λ_d for $d < t$ is

$$\lambda_d = \mu(d)\frac{g(d)d}{f(d)}\frac{G_d}{G}$$

where $G = G(t,z)$ and

$$G_d = \sum_{\substack{e|P(z)\\ e<t/d\\ (e,d)=1}} g(e).$$

We now note that

$$\begin{aligned} G &= \sum_{\substack{e|P(z)\\ e<t}} g(e)\\ &= \sum_{k|d} \sum_{\substack{e|P(z)\\ e<t\\ (e,d)=k}} g(e)\\ &= \sum_{k|d} g(k) \sum_{\substack{m|P(z)\\ m<t/k\\ (m,d)=1}} g(m)\\ &\geq G_d \sum_{k|d} g(k)\\ &= G_d \frac{g(d)d}{f(d)}. \end{aligned}$$

It follows that $|\lambda_d| \leq 1$. Finally,

$$R \leq \sum_{d,e|P(z)} \left|\lambda_d \lambda_e R_{[d,e]}\right| \leq \sum_{\substack{n|P(z)\\ n<t^2}} |R_n| \sum_{[d,e]=n} 1,$$

and

$$\sum_{[d,e]=n} 1 \leq 3^{\omega(n)}.$$

□

We first give an application to the problem considered in the previous section – deriving an upper bound for $\pi(x+y) - \pi(x)$ uniform in x.

Corollary 2. *For any $x, y \geq 2$,*

$$\pi(x+y) - \pi(x) \ll \frac{y}{\log y}.$$

Proof. As before, we take $A = \{x < n \leq x+y\}$ and P to be the set of all primes, but now we apply Selberg's sieve. To this end, observe that since $f(p) = 1$ for all

p, we have $g(p) = 1/(p-1)$. We now estimate

$$\begin{aligned} G(z,z) &= \sum_{p_1\cdots p_h<z} \prod_{i=1}^{h} (p_i-1)^{-1} \\ &= \sum_{p_1\cdots p_h<z} \prod_{i=1}^{h} \left(\frac{1}{p_i} + \frac{1}{p_i^2} + \cdots\right) \\ &= \sum_{p_1\cdots p_h<z} \sum_{k_i\geq 1} \frac{1}{p_1^{k_1}\cdots p_h^{k_h}} \\ &\geq \sum_{n<z} \frac{1}{n} \\ &\gg \log z. \end{aligned}$$

Furthermore, since $3^{\omega(d)} \leq \tau(d)^{\frac{\log 3}{\log 2}} \ll_\epsilon d^\epsilon$ for any squarefree d, as in the estimate for divisor function, the error term is $\ll_\epsilon z^{2+\epsilon}$. In total then,

$$S(A,P;z) \ll \frac{y}{\log z} + z^{2+\epsilon},$$

and the proof is complete, choosing $z = y^{1/3}$, say. □

The power of Selberg's sieve is made most evident by the following application, which gives an upper bound for the number of twin primes of the correct order of magnitude.

Theorem 9 (Brun)**.** *Let $\pi_2(x)$ count the number of $n \leq x$ such that both n and $n+2$ are prime. We have*

$$\pi_2(x) \ll \frac{x}{(\log x)^2}.$$

Proof. We apply Selberg's sieve with $A = \{n(n+2) : 1 \leq n \leq x\}$ and P being the set of all primes except 2. We note that

$$|A_d| = \#\{1 \leq n \leq x : d \mid n(n+2)\}.$$

If d is odd and square-free, say $d = p_1\cdots p_r$, then $d \mid n(n+2)$ if and only if $n \equiv 0$ or $-2 \pmod{p_i}$ for all $1 \leq i \leq r$. By the Chinese Remainder Theorem this is true if and only if n lies in one of $2^{\omega(d)}$ many residue classes modulo d. It follows that

$$|A_d| = \frac{2^{\omega(d)}}{d} x + O(2^{\omega(d)}),$$

so that $f(2)=0$ and $f(p)=2$ otherwise. In particular, $g(p)=2/(p-2)\geq 2/(p-1)$, and so $g(d)\geq 2^{\omega(d)}/\phi(d)$ for all odd square-free d. We thus have

$$\begin{aligned}G(z,z)&\geq \sum_{\substack{d<z\\ d \text{ odd and square-free}}}\frac{2^{\omega(d)}}{\phi(d)}\\ &\gg \sum_{p_1\cdots p_h<z}\prod_{i=1}^{h}\frac{2}{p_i-1}\\ &=\sum_{p_1\cdots p_h<z}\prod_{i=1}^{h}\left(\frac{2}{p_i}+\frac{2}{p_i^2}+\cdots\right)\\ &\geq \sum_{n<z}\frac{2^{\omega(n)}}{n}.\end{aligned}$$

We claim that the right hand side is $\gg (\log z)^2$. By partial summation it suffices to show that

$$\sum_{d<z}2^{\omega(d)}\gg z\log z.$$

For this we use the identity $\sum_{d^2m=n}\mu(d)\tau(m)=2^{\omega(n)}$, which is true since both sides are multiplicative and it is easily checked for prime powers. We therefore have, using that $\sum_{m<y}\tau(m)=y\log y+O(y)$,

$$\begin{aligned}\sum_{d<z}2^{\omega(d)}&=\sum_{d^2<z}\mu(d)\sum_{m<z/d^2}\tau(m)\\ &=z\log z\sum_{d^2<z}\frac{\mu(d)}{d^2}-z\sum_{d^2<z}\mu(d)\frac{\log d}{d^2}+O\left(z\sum_{d^2<z}\frac{1}{d^2}\right)\\ &=z\log z\sum_{d}\frac{\mu(d)}{d^2}+O\left(z+z\log z\sum_{d>z^{1/2}}\frac{1}{d^2}\right)\\ &=cz\log z+O(z),\end{aligned}$$

where $c=\sum_{n=1}^{\infty}\frac{\mu(d)}{d^2}>0$ is some constant. This is $\gg z\log z$ as required. Finally, we note that the error term is bounded above by

$$\sum_{d<z^2}6^{\omega(d)}\ll_{\epsilon} z^{2+\epsilon}$$

for any $\epsilon>0$, since $6^{\omega(d)}\ll_{\epsilon} d^{\epsilon}$. It follows that, for any z,

$$S(A,P;z)\ll \frac{x}{(\log z)^2}+z^3,$$

say. If we choose $z=x^{1/4}$, say, then the right hand side is $\ll x/(\log x)^2$, and the left-hand side is at least $\pi_2(x)+O(x^{1/4})$, and the result follows.

□

8. Combinatorial sieve

Lemma 9 (Buchstab formula)**.**

$$S(A,P;z) = |A| - \sum_{p|P(z)} S(A_p, P; p).$$

Proof. The left-hand side counts the number of elements in

$$S = \{n \in A : \text{ if } p \mid n \text{ and } p \in P \text{ then } p \geq z\}.$$

If we let

$$S_p = \{n \in A : p \mid n \text{ and if } q \mid n \text{ and } q \in P \text{ then } q \geq p\}$$

then it is clear that S_p is disjoint from S if $p < z$, and furthermore S_{p_1} and S_{p_2} are disjoint if $p_1 \neq p_2$. Finally, if $n \in A$ then either $n \in S$ or $n \in S_p$ where $p = \ell(n)$, the least prime divisor of n. It follows that

$$A = S \sqcup \bigsqcup_{\substack{p \in P \\ p<z}} S_p,$$

and the result follows by equating the cardinalities of both sides. □

By the same reasoning we can obtain a similar formula for the main term $W(z) = \prod_{p\in P, p<z}(1 - f(p)/p)$.

Lemma 10.

$$W(z) = 1 - \sum_{p|P(z)} \frac{f(p)}{p} W(p).$$

Corollary 3. *For any $r \geq 1$,*

$$S(A,P;z) = \sum_{\substack{d|P(z) \\ \omega(d)<r}} \mu(d)\,|A_d| + (-1)^r \sum_{\substack{d|P(z) \\ \omega(d)=r}} S(A_d, P; \ell(d)),$$

where $\ell(d)$ denotes the least prime divisor of d. Similarly,

$$W(z) = \sum_{\substack{d|P(z) \\ \omega(d)<r}} \mu(d)\frac{f(d)}{d} + (-1)^r \sum_{\substack{d|P(z) \\ \omega(d)=r}} \frac{f(d)}{d} W(\ell(d)).$$

Proof. This follows by iteratively applying Buchstab's identity. We use induction on r, noting that the case $r = 1$ is precisely Buchstab's identity. In general, for fixed $d \mid P(z)$, with $\omega(d) = r$, by Buchstab's identity

$$S(A_d, P; \ell(d)) = |A_d| - \sum_{\substack{p \in P \\ p<\ell(d)}} S(A_{pd}, P; p).$$

Since $p < \ell(d)$, the numbers pd are all square-free and $\ell(pd) = p$. Furthermore, as d ranges over all divisors of $P(z)$ with $\omega(d) = r$, these pd range over all divisors of $P(z)$ with $\omega(d') = r + 1$. The general form follows by induction, since $\mu(d) = (-1)^r$. □

In particular,

$$S(A,P;z) \geq \sum_{\substack{d|P(z) \\ \omega(d)<r}} \mu(d)\,|A_d|$$

if r is odd, and similarly this holds with an upper bound if r is even. We can use this formula to truncate the Sieve of Eratosthenes at any level r, and examine the remainder term. This leads us to Brun's pure sieve, the simplest form of the combinatorial sieve.

Theorem 10 (Brun's pure sieve)**.** *If $r \geq 6|\log W(z)|$ then*

$$S(A,P;z) = XW(z) + O\left(2^{-r}X + \sum_{\substack{d|P(z)\\ d\leq z^r}} |R_d|\right).$$

Proof. We note the trivial bounds

$$0 \leq S(A_d, P; \ell(d)) \leq |A_d| = X\frac{f(d)}{d} + R_d.$$

Therefore, by Corollary 3,

$$S(A,P;z) = X\sum_{\substack{d|P(z)\\ \omega(d)<r}} \mu(d)\frac{f(d)}{d} + O\left(X\sum_{\substack{d|P(z)\\ \omega(d)=r}} \frac{f(d)}{d} + \sum_{\substack{d|P(z)\\ \omega(d)\leq r}} |R_d|\right).$$

Also by Corollary 3,

$$W(z) = \sum_{\substack{d|P(z)\\ \omega(d)<r}} \mu(d)\frac{f(d)}{d} + O\left(\sum_{\substack{d|P(z)\\ \omega(d)=r}} \frac{f(d)}{d}\right),$$

and so

$$S(A,P;z) = XW(z) + O\left(X\sum_{\substack{d|P(z)\\ \omega(d)=r}} \frac{f(d)}{d} + \sum_{\substack{d|P(z)\\ \omega(d)\leq r}} |R_d|\right).$$

We now note that

$$\sum_{\substack{d|P(z)\\ \omega(d)=r}} \frac{f(d)}{d} \leq \frac{\left(\sum_{p|P(z)} \frac{f(p)}{p}\right)^r}{r!} \leq \left(\frac{e\sum_{p|P(z)} \frac{f(p)}{p}}{r}\right)^r.$$

Furthermore,

$$\sum_{p|P(z)} \frac{f(p)}{p} \leq \sum_{p|P(z)} -\log(1-f(p)/p) = -\log W(z).$$

If $-\log W(z) \leq r/2e$, then, the right-hand side above is $\leq 2^{-r}$. The final error term we bound crudely by noting that if $d \mid P(z)$ and $\omega(d) \leq r$ then certainly $d \leq z^r$, since it is the product of at most r primes, all less than z. □

For example, if $W(z) = \prod_{p<z}(1-1/p) \approx 1/\log z$, then we can take $r = O(\log\log z)$, so the level is $e^{\log z \log\log z}$, compared to 2^z from the Sieve of Eratosthenes.

Corollary 4. *For any $z \le \exp\left(o\left(\frac{\log x}{\log\log x}\right)\right)$ such that $z \to \infty$ with x we have the asymptotic formula*

$$|\{1 \le n \le x : \text{if } p \mid n \text{ then } p \ge z\}| \sim e^{-\gamma}\frac{x}{\log z}.$$

Proof. We apply Brun's pure sieve to $A = \{1 \le n \le x\}$ and P the set of all primes. As we have shown in the previous chapter,

$$W(z) = \prod_{p \le z}\left(1 - \frac{1}{p}\right) = (1 + o(1))e^{-\gamma}\frac{1}{\log z}.$$

In particular, if we take x large enough, then $r = 100\lceil \log\log z\rceil$ will satisfy the lower bound in Brun's sieve. For this value of r,

$$2^{-r}x \le 2^{-100\log\log z}x \le \frac{x}{(\log z)^2},$$

say. Furthermore, since $|R_d| \ll 1$, the other error term is

$$\ll \sum_{d \le z^r} 1 \le z^r \le \exp(200\log z\log\log z).$$

If $\log z = o(\log x/\log\log x)$ then, for x sufficiently large, this is at most $x^{1/2}$, say. It follows that

$$S(A, P; z) = (1 + o(1))e^{-\gamma}\frac{x}{\log z} + O\left(\frac{x}{(\log z)^2} + x^{1/2}\right) = (1 + o(1))e^{-\gamma}\frac{x}{\log z},$$

since $z \to \infty$ as $x \to \infty$. □

CHAPTER 3

The Riemann zeta function

In this chapter, we will use (as is traditional for this topic) the letter s to denote a complex variable, and σ and t to denote its real and imaginary parts respectively, so that $s = \sigma + it$. Before we begin, it's worth pausing to explicitly point out what we mean by n^s, where n is a natural number and $s \in \mathbb{C}$. By definition this is

$$n^s = e^{s \log n} = n^\sigma e^{it \log n}.$$

It is easy to check the multiplicative property, that $(nm)^s = n^s m^s$.

A Dirichlet series is an infinite series of the form

$$\sum_{n=1}^{\infty} \frac{a_n}{n^s},$$

for some coefficients $a_n \in \mathbb{C}$.

Lemma 11. *For any sequence a_n there is an abscissa of convergence σ_c such that $\alpha(s)$ converges for all s with $\sigma > \sigma_c$ and for no s with $\sigma < \sigma_c$. If $\sigma > \sigma_c$ then there is a neighbourhood of s in which $\alpha(s)$ converges uniformly. In particular, $\alpha(s)$ is holomorphic at s.*

Proof. It suffices to show that if $\alpha(s)$ converges at $s = s_0$ and we take some s with $\sigma > \sigma_0$ then α converges uniformly in some neighbourhood of s. The lemma then follows by taking $\sigma_c = \inf\{\sigma : \alpha(s) \text{ converges}\}$.

Suppose that $\alpha(s)$ converges at $s = s_0$. If we let $R(u) = \sum_{n>u} a_n n^{-s_0}$ then by partial summation, for any s,

$$\sum_{M<n\leq N} a_n n^{-s} = R(M)M^{s_0-s} - R(N)N^{s_0-s} + (s_0 - s)\int_M^N R(u)u^{s_0-s-1}\,\mathrm{d}u.$$

If $|R(u)| \leq \epsilon$ for all $u \geq M$, and if $\sigma > \sigma_0$, then it follows that

$$\left|\sum_{M<n\leq N} a_n n^{-s}\right| \leq 2\epsilon + \epsilon\,|s - s_0| \int_M^\infty t^{\sigma_0-\sigma-1}\,\mathrm{d}t \leq \left(2 + \frac{|s-s_0|}{\sigma-\sigma_0}\right)\epsilon.$$

There is some neighbourhood of s in which $|s - s_0| \ll \sigma - \sigma_0$, and hence by Cauchy's principle the series converges uniformly at s. □

Lemma 12. *If $\sum a_n n^{-s} = \sum b_n n^{-s}$ for all s in some half-plane $\sigma > \sigma_0$ then $a_n = b_n$ for all n.*

Proof. It suffices to show that if $\sum c_n n^{-s} = 0$ for all s with $\sigma > \sigma_0$ then $c_n = 0$ for all n. Suppose that $c_n = 0$ for all $n < N$. We can write

$$c_N = -\sum_{n>N} c_n (n/N)^{-\sigma}.$$

Since the sum here is convergent, the summands tend to 0, and hence $c_n \ll n^{\sigma_0}$. It follows that that this sum is absolutely convergent for $\sigma > \sigma_0 + 1$. Since each term tends to 0 as $\sigma \to \infty$, and the series is absolutely convergent, the right-hand side tends to 0, and hence $c_N = 0$. □

Lemma 13. *If $\alpha(s)$ and $\beta(s)$ are two Dirichlet series, both absolutely convergent at s, then*

$$\sum_{n=1}^{\infty} \left(\sum_{cd=n} a_c b_d \right) n^{-s}$$

is absolutely convergent and equals $\alpha(s)\beta(s)$.

Proof. We simply multiply out the product of two series,

$$\left(\sum_n a \frac{a_n}{n^s} \right) \left(\sum_m \frac{b_m}{m^s} \right) = \sum_{n,m} \frac{a_n b_m}{(nm)^s} = \sum_k \left(\sum_{nm=k} a_n b_m \right) k^{-s},$$

which is justified since both series are absolutely convergent. □

Lemma 14. *If f is multiplicative and $\sum |f(n)|\, n^{-\sigma}$ converges then*

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

Proof. By comparison each sum in the product is absolutely convergent. Since a product of finitely many absolutely convergent series can be arbitrarily rearranged,

$$\prod_{p \le y} \left(1 + f(p)p^{-s} + f(p^2)p^{-2s} + \cdots \right) = \sum_{\substack{n \\ p|n \implies p \le y}} f(n) n^{-s}.$$

Therefore the difference between the product here and the Dirichlet series is at most

$$\sum_{n > y} |f(n)|\, n^{-\sigma} \to 0 \text{ as } y \to \infty.$$

□

We now define the Riemann zeta function in the half-plane $\sigma > 1$ by

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

Observe that this series diverges at $s = 1$, and the series actually converges absolutely for $\sigma > 1$. By the above, $\zeta(s)$ defines a holomorphic function in this half-plane. For our applications, we need to extend this definition to be able to talk about $\zeta(s)$ for $\sigma > 0$.

Lemma 15. *For $\sigma > 1$,*

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

Proof. By partial summation, for any x,

$$\sum_{1 \le n \le x} n^{-s} = \frac{\lfloor x \rfloor}{x^s} + s \int_1^x \frac{\lfloor t \rfloor}{t^{s+1}}\, \mathrm{d}t.$$

The integral here is

$$s\int_1^x t^{-s}\,\mathrm{d}t - s\int_1^x \frac{\{t\}}{t^{s+1}}\,\mathrm{d}t = \frac{s}{s-1} - \frac{s}{s-1}x^{1-s} - s\int_1^x \frac{\{t\}}{t^{s+1}}\,\mathrm{d}t.$$

Since $\sigma > 1$, if we take the limit as $x \to \infty$, we have

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

noting that the integral converges. □

The integral here is convergent for any $\sigma > 0$, and therefore the right hand side defines an analytic function for $\sigma > 0$, aside from a simple pole at $s = 1$ with residue 1. We have therefore given an analytic continuation for $\zeta(s)$ up to $\sigma = 0$. We also record here the Euler product for $\zeta(s)$,

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

which is valid for $\sigma > 1$. From this it follows that $\zeta(s) \neq 0$ for $\sigma > 1$. The Euler product leads to the identity

$$\frac{1}{\zeta(s)} = \prod_p \left(1 - \frac{1}{p^s}\right) = \sum_n \frac{\mu(n)}{n^s}.$$

Furthermore, when $\sigma > 1$, the series is absolutely convergent, and so the derivative can be computed summand by summand, leading to

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

From the Euler product we have

$$\log\zeta(s) = -\sum_p \log\left(1 - p^{-s}\right) = -\sum_n \frac{\Lambda(n)}{\log n} n^{-s}.$$

Finally, taking the derivative of this, we obtain the Dirichlet series with $\Lambda(n)$ as coefficients:

$$\frac{\zeta'}{\zeta}(s) = -\sum_n \frac{\Lambda(n)}{n^{-s}}.$$

9. Prime number theorem

By partial summation, in a generalisation of the argument from the previous section, one can show that if

$$\alpha(s) = \sum_n \frac{a_n}{n^s} \text{ and } A(x) = \sum_{n\leq x} a_n$$

then

$$\alpha(s) = s\int_1^\infty \frac{A(t)}{t^{s+1}}\,\mathrm{d}t.$$

That is, $\alpha(s)$ can be expressed as a function of $A(x)$. We are more interested in the converse – for example, if $a_n = \Lambda(n)$, then $\alpha(s) = -\frac{\zeta'}{\zeta}(s)$, which we hope we can understand via analysis, and $A(x) = \sum_{n\leq x}\Lambda(n) = \psi(x)$, the asymptotics of which are the subject of the prime number theorem.

Lemma 16. *If $\sigma_0 > 0$ then*

$$\frac{1}{2\pi i}\int_{\sigma_0-iT}^{\sigma_0+iT}\frac{y^s}{s}\,\mathrm{d}s = \begin{cases}1 & \text{if } y>1 \text{ and}\\ 0 & \text{if } 0<y<1\end{cases} + O(y^{\sigma_0}/T\log y).$$

Proof. Let $y > 1$, $u < 0$, and let C be the contour from $-u - iT$ to $\sigma_0 - iT$ to $\sigma_0 + iT$ to $-u + iT$. The function y^s/s is analytic apart from a simple pole at $s = 0$, where the residue is 1. It follows that

$$\frac{1}{2\pi i}\int_C \frac{y^s}{s}\,\mathrm{d}s = 1.$$

We can bound the contribution from the two unwanted paths by

$$\int_{-u-iT}^{\sigma_0-iT}\frac{y^s}{s}\,\mathrm{d}s = \int_{-u}^{\sigma_0}\frac{y^{\sigma-iT}}{\sigma-iT}\,\mathrm{d}\sigma \ll \frac{1}{T}\int_{-u}^{\sigma_0} y^\sigma\,\mathrm{d}\sigma \le \frac{1}{T}\int_{-\infty}^{\sigma_0} y^\sigma\,\mathrm{d}\sigma = \frac{y^{\sigma_0}}{T\log y}.$$

Note that here we used that $y > 1$ for the final part. The contribution from the left-hand side of the rectangle is

$$\ll \int_{-T}^{T}\frac{y^u}{|u-it|}\,\mathrm{d}t \ll T\frac{y^u}{u} \to 0 \text{ as } u\to-\infty.$$

The case $0 < y < 1$ is similar, but we need to take the mirror image of the contour. □

Theorem 11 (Perron's formula)**.** *Suppose that $\alpha(s)$ is absolutely convergent for $\sigma > \sigma_a$. If $\sigma_0 > \max(0, \sigma_a)$ and $x > 0$ is not an integer then, for any $T \ge 1$,*

$$\sum_{n\le x} a_n = \frac{1}{2\pi i}\int_{\sigma_0-iT}^{\sigma_0+iT}\alpha(s)\frac{x^s}{s}\,\mathrm{d}s + O\left(2^{\sigma_0}\frac{x}{T}\sum_{x/2<n<2x}\frac{|a_n|}{|x-n|} + \frac{x^{\sigma_0}}{T}\sum_n\frac{|a_n|}{n^{\sigma_0}}\right).$$

Proof. Since $\sigma_0 > 0$, we can write

$$1_{n<x} = \frac{1}{2\pi i}\int_{\sigma_0-iT}^{\sigma_0+iT}\frac{(x/n)^s}{s}\,\mathrm{d}s + O\left(\frac{(x/n)^{\sigma_0}}{T\log(x/n)}\right).$$

It follows that, since the series converges absolutely, and so we can interchange it with a finite integration,

$$\sum_{n<x} a_n = \frac{1}{2\pi i}\int_{\sigma_0-iT}^{\sigma_0+iT}\frac{x^s}{s}\alpha(s) + O\left(\frac{x^{\sigma_0}}{T}\sum_n\frac{|a_n|}{n^{\sigma_0}\,|\log(x/n)|}\right).$$

Note that the interchange of integration and summation is valid since both are over finite ranges. To simplify the error term, write $\log(x/n) = |\log(1 + (n-x)/x)|$ and use the fact that $|\log(1+\delta)| \asymp |\delta|$ uniformly for $-1/2 \le \delta \le 1$, so that $|\log(x/n)| \asymp \frac{n-x}{x}$ uniformly for $x/2 < n < 2x$. For other values of n, we have $|\log(x/n)| \gg 1$. The error term is therefore

$$\ll \frac{2^{\sigma_0}x}{T}\sum_{x/2<n<2x}\frac{|a_n|}{|x-n|} + \frac{x^{\sigma_0}}{T}\sum_n\frac{|a_n|}{n^{\sigma_0}}.$$

□

For the proof of the prime number theorem we require the following three facts about zero-free regions of ζ, which we will prove in the next section.

(1) there is a constant c such that, if $\sigma > 1 - c/\log t$ and $|t| \geq 7/8$, then $\zeta(s) \neq 0$ and $\frac{\zeta'}{\zeta}(s) \ll \log(|t|+2)$, and
(2) $\zeta(s) \neq 0$ for $\frac{8}{9} < \sigma$ and $|t| \leq 7/8$, and
(3) $\frac{\zeta'}{\zeta}(s) = \frac{-1}{s-1} + O(1)$ for $5/6 \leq \sigma \leq 2$ and $|t| \leq 7/8$.

Theorem 12 (Prime Number Theorem)**.** *There is $c > 0$ such that*

$$\psi(x) = x + O\left(\frac{x}{\exp(c\sqrt{\log x})}\right).$$

In particular, $\psi(x) \sim x$.

Proof. We can assume that $x = N + 1/2$ for some integer N. By Perron's formula, for any $2 > \sigma_0 > 1$, we have

$$\psi(x) = -\frac{1}{2\pi i}\int_{\sigma_0 - iT}^{\sigma_0+iT} \frac{\zeta'}{\zeta}(s)\frac{x^s}{s}\,\mathrm{d}s + O\left(2^{\sigma_0}\frac{x}{T}\sum_{x/2<n<2x}\frac{\Lambda(n)}{|x-n|} + \frac{x^{\sigma_0}}{T}\sum_n \frac{\Lambda(n)}{n^{\sigma_0}}\right).$$

Let's first examine the error term here. The first summand is

$$\ll \frac{x\log x}{T}\sum_{n=1}^{x}\frac{1}{n} \ll \frac{x(\log x)^2}{T}.$$

The infinite sum in the second summand is $\ll \frac{1}{\sigma_0 - 1}$. Overall, then the error term is

$$\ll \frac{x(\log x)^2}{T} + \frac{x^{\sigma_0}}{T(\sigma_0 - 1)}.$$

Choosing $\sigma_0 = 1 + 1/\log x$ this is $O(x(\log x)^2/T)$. Now let $\sigma_1 = 1 - c/\log T$, where c is such that $\zeta(s) \neq 0$ for $\sigma \geq \sigma_1$, and C be the rectangle contour connecting the two lines. Since $\zeta'/\zeta(s)$ has a simple pole at $s = 1$ with residue -1, and is otherwise analytic (as there are no zeros of ζ inside the contour), we have

$$\frac{1}{2\pi i}\int_C \frac{\zeta'}{\zeta}(s)\frac{x^s}{s}\,\mathrm{d}s = -x.$$

The right-hand side of this contour is $-\psi(x) = O(xT^{-1}(\log x)^2)$ by the above. It remains to bound the contribution from the other sides of the rectangle. To this end, first note that

$$\int_{\sigma_1+iT}^{\sigma_0+iT}\frac{\zeta'}{\zeta}(s)\frac{x^s}{s}\,\mathrm{d}s \ll \frac{x^{\sigma_0}}{T}(\sigma_1 - \sigma_0)\log T,$$

since $|s| \gg T$ on this line, $|x^s| \leq x^{\sigma_0}$, the line has length $\sigma_1 - \sigma_0$, and $\frac{\zeta'}{\zeta}(s) \ll \log t$. Using our choices for σ_0 and σ_1, it follows that the two short sides of the rectangle contribute $O(x/T)$, provided $T \leq x$.

Finally, we turn our attention to the long left-hand side of the rectangle. Away from $t = 0$ (where $|t| \geq 1$, say) we use the bound $\frac{\zeta'}{\zeta} \ll \log|t|$ to bound the contribution along this line by

$$\ll x^{\sigma_1}\log T\int_1^T \frac{1}{t}\,\mathrm{d}t \ll x^{\sigma_1}(\log T)^2.$$

In the interval $-1 < t < 1$ we use the bound $\frac{\zeta'}{\zeta} \ll \frac{1}{|s-1|}$ to bound the contribution to the integral by

$$\ll x^{\sigma_1} \int_{-1}^{1} \frac{1}{|t(t-1)|}\,\mathrm{d}t \ll x^{\sigma_1}.$$

Combining these estimates we have, for any $1 \le T \le x$,

$$\psi(x) = x + O\left(\frac{x}{T}(\log x)^2 + x^{1-c/\log T}(\log T)^2\right).$$

We now make a choice of T to optimise this error term, which is $T = \exp(c\sqrt{\log x})$ for some small constant $c > 0$, and the proof is complete. □

10. Zero-free region

Theorem 13. *If $\sigma > (1+t^2)/2$ then $\zeta(s) \neq 0$. In particular, $\zeta(s) \neq 0$ if $\frac{8}{9} \le \sigma \le 1$ and $|t| \le \frac{7}{8}$. Furthermore,*

$$\zeta(s) = \frac{1}{s-1} + O(1) \text{ and } -\frac{\zeta'}{\zeta}(s) = \frac{1}{s-1} + O(1)$$

uniformly for $\frac{8}{9} \le \sigma \le 2$ and $|t| \le \frac{7}{8}$.

Proof. We recall the identity

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

In particular,

$$\left|\zeta(s) - \frac{s}{s-1}\right| \le \frac{|s|}{\sigma},$$

which proves the first claim and the second. The final claim follows since if $\zeta(s) = (s-1)^{-1} + f(s)$ then

$$\frac{\zeta'}{\zeta}(s) = \frac{-(s-1)^{-2} + f'(s)}{(s-1)^{-1} + f(s)} = \frac{-1}{s-1} + \frac{f(s) + f'(s)(s-1)}{1 + f(s)(s-1)} = \frac{-1}{s-1} + O(1).$$

□

Theorem 14 (Maximum modulus principle)**.** *If U is a bounded connected open set and f is holomorphic on U then $|f|$ attains its maximum on the boundary $\partial U = \overline{U}\backslash U$.*

Lemma 17 (Borel-Carathéodory Lemma)**.** *Let f be holomorphic on $|z| \le R$ such that $f(0) = 0$ and suppose $\Re f(z) \le M$ for all $|z| \le R$. For any $r < R$,*

$$\sup_{|z|\le r}\left(|f(z)|, |f'(z)|\right) \ll_{r,R} M.$$

Proof. Let

$$g(z) = \frac{f(z)}{z(2M - f(z))},$$

so that g is holomorphic for $|z| \le R$. Observe that, using $\Re f(z) \le M$,

$$|f(z)|^2 = \Re(f(z))^2 + \Im(f(z))^2 \le (2M - \Re(f(z)))^2 + \Im(f(z))^2 = |2M - f(z)|^2,$$

and so $|2M - f(z)| \ge |f(z)|$ for $|z| \le R$. In particular, if $|z| = R$ then $|g(z)| \le 1/R$. By the maximum modulus principle, if $|z| = r$, then

$$|g(z)| = \frac{|f(z)|}{r\,|2M - f(z)|} \le \frac{1}{R},$$

and hence

$$R\left|f(z)\right| \leq \left|2Mr - rf(z)\right| \leq 2Mr + r\left|f(z)\right|,$$

or

$$|f(z)| \leq \frac{2r}{R-r}M.$$

This shows that $|f(z)| \ll M$. To deduce the same bound for $f'(z)$, we use Cauchy's formula

$$f'(z) = \frac{1}{2\pi i}\int_{r'} \frac{f(w)}{(w-z)^2}\,\mathrm{d}w,$$

where the integral is taken over some circle of radius $r < r' < R$, say. □

Lemma 18. *Suppose that $f(z)$ is analytic in a domain containing $|z| \leq 1$, that $|f(z)| \leq M$ in this disc, and that $f(0) \neq 0$. Let $0 < r < R < 1$. Then for $|z| \leq r$,*

$$\frac{f'}{f}(z) = \sum_{k=1}^{K} \frac{1}{z - z_k} + O\left(\log \frac{M}{|f(0)|}\right)$$

where the sum is over all zeros z_k of f for which $|z_k| \leq R$.

Proof. Without loss of generality, we may suppose that $f(0) = 1$. Furthermore, we can assume that there are no zeros of f in the annulus $r < R - \epsilon \leq |z| \leq R + \epsilon < 1$ for some small enough $\epsilon > 0$. All implicit constants in this proof can depend on r, R, and ϵ. Let

$$g(z) = f(z)\prod_{k=1}^{K} \frac{R^2 - z\overline{z_k}}{R(z - z_k)}.$$

Observe that the kth factor has a pole at z_k, and has modulus 1 on $|z| = R$. It follows that g is an analytic function in $|z| \leq R$, and if $|z| = R$ then $|g(z)| = |f(z)| \leq M$. By the maximum modulus principle,

$$|g(0)| = \prod_{k=1}^{K} \frac{R}{|z_k|} \leq M.$$

It follows that, since each z_k satisfies $|z_k| \leq R - \epsilon$, the number of zeros satisfies $K \ll \log M$. Let $h(z) = \log(g(z)/g(0))$ (which is allowed since $g(z)$ has no zeros in $|z| \leq R$). We have $h(0) = 0$, and

$$\Re h(z) = \log|g(z)| - \log|g(0)| \leq \log M$$

for $|z| \leq R$. By the Borel-Carathéodory lemma,

$$|h'(z)| \ll \log M.$$

Finally,

$$h'(z) = \frac{f'}{f}(z) - \sum_{k=1}^{K} \frac{1}{z - z_k} + \sum_{k=1}^{K} \frac{1}{z - R^2/\overline{z_k}}.$$

If $|z| \leq r$ then $\left|z - R^2/\overline{z_k}\right| \geq \left|R^2/\overline{z_k}\right| - |z| \geq R - r$, and so the final sum is $\ll \log M$. □

Since $|\zeta(3/2 + it)| \gg 1$ and $|\zeta(s + 3/2 + it)| \ll t$ for $|s| \leq 1$ we obtain the following information about the zeta function.

Corollary 5. *If $|t| \geq 7/8$ and $5/6 \leq \sigma \leq 2$ then*

$$\frac{\zeta'}{\zeta}(s) = \sum_{\rho} \frac{1}{s-\rho} + O(\log|t|),$$

where the sum is over all zeros ρ of $\zeta(s)$ in the region $|\rho - (3/2 + it)| \leq 5/6$.

Theorem 15. *There is a constant $c > 0$ such that*

$$\zeta(s) \neq 0 \text{ for } \sigma \geq 1 - \frac{c}{\log t}.$$

Proof. Let $\rho = \sigma + it$ be such that $\zeta(\rho) = 0$, and let $\delta > 0$ be something to be chosen later. By Corollary 5,

$$-\Re\frac{\zeta'}{\zeta}(1+\delta+it) = -\frac{1}{1+\delta-\sigma} - \Re\sum_{\rho' \neq \rho} \frac{1}{1+\delta+it-\rho'} + O(\log t)$$

Since $\Re\rho' \leq 1$ for all zeros ρ', it follows that $\Re(1/(1+\delta+it-\rho')) > 0$, provided $\delta > 0$. In particular,

$$-\Re\frac{\zeta'}{\zeta}(1+\delta+it) \leq -\frac{1}{1+\delta-\sigma} + O(\log t).$$

Similarly,

$$-\Re\frac{\zeta'}{\zeta}(1+\delta+2it) \ll \log t.$$

Finally,

$$-\frac{\zeta'}{\zeta}(1+\delta) = \frac{1}{\delta} + O(1).$$

We now note that

$$\Re\left(-3\frac{\zeta'}{\zeta}(1+\delta) - 4\frac{\zeta'}{\zeta}(1+\delta+it) - \frac{\zeta'}{\zeta}(1+\delta+2it)\right)$$

is

$$\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^{1+\delta}} \left(3 + 4\cos(t\log n) + \cos(2t\log n)\right).$$

Since $3 + 4\cos\theta + \cos 2\theta = 2(1+\cos\theta)^2 \geq 0$, the entire sum is ≥ 0. It follows that

$$\frac{3}{\delta} - \frac{4}{1+\delta-\sigma} + O(\log t) \geq 0.$$

This implies that $1 - \sigma \gg 1/\log t$, choosing $\delta \approx 1/\log t$. □

This result has been improved by Korobov and Vinogradov to $1 - c/(\log t)^{2/3+\epsilon}$.

11. Error terms

Assuming the Riemann hypothesis, the proof above of the prime number theorem gives

$$\psi(x) = x + O(x^{1/2+o(1)}).$$

In fact the error term here can be taken as $x^{1/2}(\log x)^2$, but this more precise form takes some finesse to achieve. It is natural to ask how good an error term we might hope for here. It turns out that, just as the absence of zeros of the zeta function allows us to show the error term is small, their presence allows us to show that error term must be somewhat large.

For this we will need the following lemma of Landau. The real interest of this lemma is that when our integrand is non-negative we obtain not just a half-plane of convergence, but more importantly knowledge about what happens on the boundary line itself. A similar fact holds for Dirichlet series.

Lemma 19 (Landau)**.** *Supppose that A is an integrable function bounded in any finite interval, $A(x) \geq 0$ for all large $x \geq X$, and let*

$$\sigma_c = \inf\left\{\sigma : \int_X^\infty A(x)x^{-\sigma}\,\mathrm{d}x < \infty\right\}.$$

The function

$$F(s) = \int_1^\infty A(x)x^{-s}\,\mathrm{d}x$$

is analytic in $\sigma > \sigma_c$ but not at $s = \sigma_c$.

Proof. Divide the integral in the definition of F to $[1, X]$ and $[X, \infty)$, given a corresponding decomposition into $F = F_1 + F_2$, say. The function F_1 is entire. For $\sigma > \sigma_c$, the integral converges absolutely, and hence F_2 also defines an entire function. Suppose that F_2 is analytic at $s = \sigma_c$. We may expand $F_2(s)$ as a power series at $s = \sigma_c + 1$, so that

$$F_2(s) = \sum_{k=0}^\infty c_k(s - 1 - \sigma_c)^k,$$

where

$$c_k = \frac{F_2^{(k)}(1+\sigma_c)}{k!} = \frac{1}{k!}\int_X^\infty A(x)(-\log x)^k x^{-1-\sigma_c}\,\mathrm{d}x.$$

The radius of convergence of this power series is the distance from $1 + \sigma_c$ to the nearest singularity of $F_2(s)$, and hence by assumption is at least $1 + \delta$ for some $\delta > 0$, say. If we consider $s = \sigma_c - \delta/2$, then

$$F_2(s) = \sum_{k=0}^\infty \frac{(1+\sigma_c - s)^k}{k!}\int_X^\infty A(x)(\log x)^k x^{-1-\sigma_c}\,\mathrm{d}x.$$

This is a convergent series with all non-negative terms, and hence we can interchange the integral and summation, to find

$$F_2(s) = \int_X^\infty A(x)x^{-1}\exp((1+\sigma_c - s)\log x)\,\mathrm{d}x = \int_X^\infty A(x)x^{-s}\,\mathrm{d}x,$$

and so the integral must converge at $s = \sigma_c - \delta/2$, which contradicts the definition of σ_c. □

When discussing lower bounds for error terms, the following notation is useful. We say that $f = \Omega_\pm(g)$ if

$$\limsup_{x\to\infty}\frac{f(x)}{g(x)} \geq c > 0$$

and

$$\liminf_{x\to\infty}\frac{f(x)}{g(x)} \leq -c < 0,$$

for some absolute constant $c > 0$. That is, not only does $f(x)$ exceed (some constant multiple of) $g(x)$ infinitely often, but it does so both positively and negatively.

Theorem 16. *If σ_0 is the supremum of the real parts of the zeros of $\zeta(s)$ then, for any $\sigma < \sigma_0$,*

$$\psi(x) = x + \Omega_\pm(x^\sigma).$$

If there is a zero ρ with $\Re\rho = \sigma_0$, then

$$\psi(x) = x + \Omega_\pm(x^{\sigma_0}).$$

Proof. Suppose that $\psi(x) - x \le cx^\sigma$ for all large enough x, say $x > X$. Following Lemma 19 we will consider the function

$$F(s) = \int_1^\infty (cx^\sigma - \psi(x) + x)x^{-s-1}\,\mathrm{d}x = \frac{c}{s-\sigma} + \frac{\zeta'(s)}{s\zeta(s)} + \frac{1}{s-1}.$$

The right-hand side has a pole at $s = \sigma$, but is analytic for real $s > \sigma$. It follows that in fact the above identity must hold for all s with $\Re s > \sigma$. It follows that there can't be any zeros of $\zeta(s)$ in this region, which is a contradiction if $\sigma < \sigma_0$.

For the second, stronger, conclusion, we need to argue a little more carefully. Suppose that there is a zero $\rho = \sigma_0 + it$. Consider instead

$$F(s) + \frac{e^{i\theta}F(s+it) + e^{-i\theta}F(s-it)}{2} = \int_1^\infty (cx^\sigma - \psi(x) + x)\,(1+\cos(\theta - t\log x))x^{-s-1}\,\mathrm{d}x.$$

The coefficients here are still non-negative real numbers. The left-hand side has a pole at $s = \sigma$ with residue

$$c + \frac{me^{i\theta}/\rho + me^{-i\theta}/\overline{\rho}}{2},$$

where m is the multiplicity of the zero ρ. We have freedom to choose θ to be whatever we like, in particular so that this expression is $c - m/|\rho|$. The lim inf of the right-hand side is $> -\infty$ as s approaches σ from the right along the real axis. We must therefore have

$$c - \frac{m}{|\rho|} \ge 0,$$

and hence $c \ge m/|\rho|$. This establishes the Ω_+ aspect of the theorem. For Ω_- we use the same argument with signs reversed, so that we start with

$$F(s) = \int_1^\infty (-cx^\sigma + \psi(x) - x)x^{-s-1}\,\mathrm{d}s,$$

and so on. □

12. Functional equation

Finally, we establish an important analytic property of the zeta function, which in particular reveals a certain symmetry in the location of zeros.

We first extend the definition of the zeta function to a larger half-plane. Recall that for $\sigma > 0$ we defined

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

We will extend the region where this is valid by integrating by parts. First let $f(x) = \frac{1}{2} - \{x\}$, so that

$$\zeta(s) = \frac{1}{2} + \frac{1}{s-1} + s\int_1^\infty \frac{f(u)}{u^{s+1}}\,\mathrm{d}u.$$

If we let $F(x) = \int_0^x f(u)\,\mathrm{d}u$ then, by integration by parts,

$$\int_1^\infty \frac{f(u)}{u^{s+1}} = [F(u)u^{-s-1}]_0^\infty + (s+1)\int_1^\infty \frac{F(u)}{u^{s+2}}\,\mathrm{d}u.$$

Since $F(x)$ is bounded, the integral here converges for any s with $\sigma > -1$, and hence the left-hand side also converges in this region. We may therefore take

$$\zeta(s) = \frac{1}{2} + \frac{1}{s-1} + s\int_1^\infty \frac{f(u)}{u^{s+1}}\,\mathrm{d}u$$

as the definition of $\zeta(s)$ in the half-plane $\sigma > -1$. If $-1 < \sigma < 0$ then

$$\int_0^1 \frac{f(u)}{u^{s+1}}\,\mathrm{d}u = \frac{1}{2}\int_0^1 \frac{1}{u^{s+1}}\,\mathrm{d}u - \int_0^1 \frac{1}{u^s}\,\mathrm{d}u = -\frac{1}{2s} + \frac{1}{s-1},$$

and so in this strip

$$\zeta(s) = s\int_0^\infty \frac{f(u)}{u^{s+1}}\,\mathrm{d}u.$$

We now note that $f(x)$ is a periodic function, continuous in $(0,1)$, and so it has a Fourier series, which is

$$f(u) = \sum_{n=1}^\infty \frac{\sin(2\pi nu)}{\pi n}.$$

For $-1 < \sigma < 0$ we therefore get

$$\zeta(s) = s\int_0^\infty \frac{1}{u^{s+1}}\sum_{n=1}^\infty \frac{\sin(2\pi nu)}{\pi n}\,\mathrm{d}u = \frac{s}{\pi}\sum_{n=1}^\infty \frac{1}{n}\int_0^\infty \frac{\sin(2\pi nu)}{u^{s+1}}\,\mathrm{d}u.$$

We should justify the interchange of integral and summation here. For this, we note that the series is uniformly convergent almost everywhere, furthermore converges to some bounded value in $(-1/2, 1/2]$ almost everywhere, and hence the interchange of limits is justified by the dominated convergence theorem.

By change of variable, we have

$$\int_0^\infty \frac{\sin(2\pi nu)}{u^{s+1}} = (2\pi n)^s\int_0^\infty \frac{\sin(u)}{u^{s+1}}\,\mathrm{d}u.$$

Furthermore, writing $\sin(u) = \frac{1}{2i}(e^{iu} - e^{-iu})$ and using another change of variable,

$$\int_0^\infty \frac{\sin u}{u^{s+1}}\,\mathrm{d}u = -\sin(\pi s/2)\int_0^\infty t^{-s-1}e^{-t}\,\mathrm{d}t.$$

The integral here is an important one known as the Gamma function. It cannot be simplified further, in general, but is one of the fundamental functions of analysis. Let

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

which converges for $\sigma > 0$, and defines a holomorphic function in this region. Integrating by parts we see that

$$\Gamma(s+1) = s\Gamma(s).$$

This identity, combined with the fact that $\Gamma(1) = 1$, implies that $\Gamma(n) = (n-1)!$ for all integer $n \geq 1$. Furthermore, it allows us to analytically extend $\Gamma(s)$ to a meromorphic function on the entire complex plane with poles at $s = 0, -1, -2, \ldots$.

Combining the above we have shown that, for $-1 < \sigma < 0$,

$$\zeta(s) = \frac{s}{\pi} \sum_{n=1}^{\infty} \frac{(2\pi n)^s}{n} \sin(\pi s/2)\Gamma(-s) = 2^s \pi^{s-1} \sin(\pi s/2)\Gamma(1-s)\zeta(1-s).$$

The right-hand side is actually analytic for any $\sigma < 1$, and hence we can take the right-hand side to be a definition of $\zeta(s)$ in this region. By analytic continuation it follows that this identity must hold for all $s \in \mathbb{C}$. We have proved the following.

Theorem 17 (Functional equation)**.** *The zeta function $\zeta(s)$ can be extended a function meromorphic on the whole complex plane, and for all s satisfies the identity*

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

Many interesting facts can be deduced from this identity. We will first use it to study the possible poles of $\zeta(s)$. We know that $\zeta(s)$ has a simple pole at $s = 1$, and nowhere else for $\sigma > -1$. Suppose that ζ has a pole at s for $\sigma < 0$. Then so too does $\Gamma(1-s)\zeta(1-s)$, but both $\Gamma(s)$ and $\zeta(s)$ are holomorphic for all s with $\Re s > 1$, which is a contradiction. It follows that $\zeta(s)$ only has one pole in $\mathbb{C}$, which is a simple pole at $s = 1$.

We will now consider the zeros of $\zeta(s)$. Suppose that $\zeta(s) = 0$ and $\sigma < 0$. It follows that

$$\sin(\pi s/2)\Gamma(1-s)\zeta(1-s) = 0.$$

Again, neither $\Gamma(1-s)$ nor $\zeta(1-s)$ can be zero or a pole, and so $\sin(\pi s/2)$, which means s must be an even integer. These are called the trivial zeros of $\zeta(s)$, located at $s = -2, -4, -6, \ldots$. Since there are no zeros with $\sigma \geq 1$, there are no other zeros with $\sigma \leq 0$.

Aside from the trivial zeros, then, all zeros of ζ must lie in the critical strip $0 < \sigma < 1$. Furthermore, since the other factors in the functional equation are entire and non-zero in this strip, this implies that if ρ is a zero in the critical strip, then so too is $1-\rho$. There is therefore a symmetry around the critical line $\sigma = 1/2$. The Riemann hypothesis is motivated in part by the belief that this symmetry should collapse so that all the zeros are located exactly on this line.

Using this symmetry and the results of the previous section we obtain the following equivalence between the Riemann hypothesis and the error term in the prime number theorem.

Theorem 18. *The Riemann hypothesis is equivalent to the statement that*

$$\psi(x) = x + O\left(x^{1/2+o(1)}\right).$$

Proof. If the Riemann hypothesis is true then the contour integration proof of the prime number theorem can be improved to show that $\psi(x) = x + O(x^{1/2+o(1)})$. For the converse, we use Theorem 16. If the Riemann hypothesis fails, there is some zero ρ with real part $0 < \sigma < 1$ such that $\sigma \neq 1/2$. Since $1-\rho$ is also a zero, we may assume without loss of generality that $\sigma > 1/2$. If we take some $1/2 < \sigma' < \sigma$ then Theorem 16 shows

$$\psi(x) = x + \Omega_{\pm}(x^{\sigma'}),$$

which would contradict $\psi(x) = x + O(x^{1/2+o(1)})$ for large enough x (large enough so that the $1/2 + o(1)$ exponent is less than σ'), which concludes the proof. □

If we also assume the existence of at least one zero on the critical line (there are infinitely many, and the first is at $1/2 + (14.1\ldots)it$), then we have the following error term bound.

Theorem 19.

$$\psi(x) = x + \Omega_{\pm}(x^{1/2}).$$

Proof. If the Riemann hypothesis is true then we use the existence of a zero on the critical line and the second part of Theorem 16. If the Riemann hypothesis is false then an even stronger statement is true by the first part of Theorem 16. □

CHAPTER 4

Primes in arithmetic progression

13. Dirichlet characters and L-functions

A Dirichlet character modulo q is a totally multiplicative function $\chi : \mathbb{N} \to \mathbb{C}$ such that $\chi(n) = 0$ if $(n, q) \neq 1$ and χ has period q. Another way to think of them is as group homomorphisms from $(\mathbb{Z}/q\mathbb{Z})^\times \to \mathbb{C}$, extended to a function on the whole of $\mathbb{N}$ in the natural way.

For every q there is a trivial character, known as the principal character, denoted by χ_0, which is equal to 1 for all n with $(n, q) = 1$. Since $(\mathbb{Z}/q\mathbb{Z})^\times$ is a finite abelian group, the set of characters is also a finite abelian group, isomorphic to $(\mathbb{Z}/q\mathbb{Z})^\times$. In particular, there are exactly $\phi(q)$ many Dirichlet characters modulo q.

These are multiplicative characters, in contrast to the additive characters $n \mapsto e^{2\pi itn}$ used in Fourier analysis. As such, they are ideal for use in multiplicative number theory. Their most fundamental property is their orthogonality.

Lemma 20. *If χ is a Dirichlet character modulo q then*

$$\sum_{\substack{1 \leq n \leq q \\ (q,n)=1}} \chi(n) = \begin{cases} \phi(q) & \text{if } \chi = \chi_0 \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, for any $n \in \mathbb{N}$,

$$\sum_{\chi} \chi(n) = \begin{cases} \phi(q) & \text{if } n \equiv 1 \pmod{q} \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

In the statement of this lemma, and throughout this chapter unless specified otherwise, $\sum_\chi$ means the sum is taken over all Dirichlet characters modulo q.

Proof. Let S be the first sum. If $\chi = \chi_0$ then the claim is trivial. If $\chi \neq \chi_0$ there exists some a with $(a, q) = 1$ such that $\chi(a) \neq 1, 0$. It follows that

$$\chi(a)S = \sum_{\substack{1 \leq n \leq q \\ (q,n)=1}} \chi(an) = \sum_{\substack{1 \leq m \leq q \\ (q,m)=1}} \chi(m) = S,$$

since $n \mapsto an$ is a permutation on the reduced residue classes modulo q, and hence $S = 0$. The proof of the second claim is similar. □

Our main interest lies in using Dirichlet characters to detect when a number lies in a certain residue class modulo q, for which we note that, if $(a, q) = 1$, then

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

where $\overline{a}$ denotes the multiplicative inverse of a modulo q. Thus, for example, if we want to count primes $\equiv a \pmod q$, then it is natural to consider the sum

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

The innermost sum here we can study using analytic machinery, just as we have done for $\psi(x)$ previously. We now need to consider a more general type of zeta function.

The Dirichlet L-function of χ modulo q is the Dirichlet series

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

This is certainly absolutely convergent, and hence holomorphic, for $\sigma > 1$. By partial summation, however, we can do better, if the character is not principal.

Lemma 21. *If χ is a non-principal character then*

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

converges for all $\sigma > 0$.

Proof. By partial summation, for any x, if $F(x) = \sum_{1\leq n\leq x}\chi(n)$

$$\sum_{1\leq n\leq x}\frac{\chi(n)}{n^s} = \frac{F(x)}{x^s} - s\int_1^x \frac{F(t)}{t^{s+1}}\,\mathrm{d}t.$$

Since χ is periodic and $\sum_{1\leq n\leq q}\chi(n) = 0$, we know $F(x)$ is bounded, and hence the left-hand side has a limit as $x\to\infty$ provided $\sigma > 0$. □

In particular, unlike the Riemann zeta function, there is no pole at $s = 1$. This has surprisingly deep implications, as we will see later on.

Since χ is a totally multiplicative function, $L(s,\chi)$ can be written as an Euler product,

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

for $\sigma > 1$. In particular, note that

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

so that the L-function of the principal character is closely related to the zeta function. In particular, $L(s,\chi_0)$ is analytic for $\sigma > 0$ except for a simple pole at $s = 1$ with residue $\phi(q)/q$.

Just as with the Riemann zeta function, we can take logarithms of this Euler product and then differentiate, which yields

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

which is valid for $\sigma > 1$. In the next section we will use this to prove Dirichlet's theorem.

14. Dirichlet's theorem

One of our goals is to count the number of primes which are congruent to $a \pmod q$ for some fixed residue class with $(a,q)=1$. The first result to establish is that there are in fact infinitely many such primes. This was first established by Dirichlet using the machinery of L-functions.

To this end, note that by the discussion in the previous section, for $\sigma > 1$

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

The summand from $\chi = \chi_0$ contributes

$$\frac{1}{\phi(q)(s-1)} + O_q(1),$$

since $L(s,\chi_0)$ has a simple pole at $s=1$. It follows that

$$\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} 1_{n\equiv a \pmod q} = \frac{1}{\phi(q)(s-1)} + O_q(1) + O_q\left(\sum_{\chi \neq \chi_0} \frac{L'}{L}(s,\chi) \right).$$

To show that there are infinitely many primes $\equiv a \pmod q$ it would suffice to show that the final error term remains bounded as $s \to 1$, which would imply that

$$\sum_{p \equiv a \pmod q} \frac{\log p}{p} = \infty.$$

Since $L(s,\chi)$ is analytic for $\sigma > 0$, L'/L is analytic except possibly for zeros of L. To prove Dirichlet's theorem, then, it suffices to prove the following.

Theorem 20. *If $\chi \neq \chi_0$ then $L(1,\chi) \neq 0$.*

Proof. We will first introduce some convenient terminology: a character is quadratic if $\chi^2 = \chi_0$ but $\chi \neq \chi_0$, so that it takes only values $-1, 0, 1$. Otherwise, χ takes on some non-real values, and it is called complex. We will prove the theorem first for complex characters.

For $\sigma > 1$,

$$\begin{aligned} \prod_{\chi} L(s,\chi) &= \exp\left(\sum_{\chi} \sum_{n=2}^{\infty} \frac{\Lambda(n)}{\log n} \chi(n) n^{-s} \right) \\ &= \exp\left(\phi(q) \sum_{\substack{n \geq 2 \\ n \equiv 1 \pmod q}} \frac{\Lambda(n)}{\log n} n^{-s} \right). \end{aligned}$$

If $s = \sigma > 1$ is real then the sum above is a non-negative real number, and so

$$\prod_{\chi} L(\sigma,\chi) \geq 1$$

for all $\sigma > 1$. Since $L(s,\chi_0)$ has a pole at $s=1$ it follows that $L(1,\chi) = 0$ can hold for at most one $\chi \neq \chi_0$, for otherwise the product would tend to 0 as $\sigma \to 1$. The theorem now follows immediately for complex χ, for if χ is a complex character then so is $\overline{\chi}$, and since $\overline{L(s,\chi)} = L(\overline{s},\overline{\chi})$, and so $L(1,\overline{\chi}) = 0$ if and only if $L(1,\chi) = 0$.

If χ is complex then $\chi \neq \overline{\chi}$, which contradicts the fact that $L(1,\chi) = 0$ for at most one character.

The case when χ is quadratic is harder, because there is no natural other character to pair it with. Instead, we will pair it with $\zeta(s)$ itself. Suppose that $L(1,\chi) = 0$. Then $\zeta(s)L(s,\chi)$ is analytic for $\sigma > 0$, and for $\sigma > 1$ this is a Dirichlet series with coefficients

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

Clearly r is multiplicative, and furthermore $r(n) \geq 0$ for all n, which is easily checked by verifying it for prime powers, since

$$r(p^k) = \sum_{0 \leq j \leq k} \chi(p)^j.$$

If $\chi(p) = 1$ then this is $k+1$, if $\chi(p) = -1$ this is 0 or 1, depending on whether k is odd or even, and if $\chi(p) = 0$ then this is 1. It follows that, not only is $r(n) \geq 0$, but also $r(n^2) \geq 1$ for all n.

We have a Dirichlet series with non-negative coefficients, so we will now apply Landau's lemma for Dirichlet series, stated below. It follows that the series $\sum_n r(n)n^{-s}$ must converge for $\sigma > 0$, but it can't converge for $s = 1/2$, where

$$\sum_{n=1}^{\infty} r(n)n^{-1/2} \geq \sum_{m=1}^{\infty} \frac{1}{m} = \infty.$$

□

Lemma 22 (Landau)**.** *If $L_f(s)$ is a Dirichlet series with non-negative coefficients $f(n) \geq 0$ then, if*

$$\sigma_0 = \inf\left\{\sigma : \sum_{n=1}^{\infty} \frac{f(n)}{n^{\sigma}} < \infty\right\},$$

$L_f(s)$ is analytic for all $\sigma > \sigma_0$ but has a pole at $s = \sigma_0$.

15. Zero-free region

Our main goal is to go further than Dirichlet's theorem and establish a precise quantitative result concerning the number of primes in arithmetic progressions. Using the relationship

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

we will use counter integration, as in the proof of the prime number theorem. For this we first need to establish a zero-free region for $L(s,\chi)$. Much of the argument is similar to that we used for $\zeta(s)$, but there is a surprising difficulty owing to the lack of a pole at $s = 1$.

We introduce the convenient notation of τ for $|t| + 4$, where t as usual denotes the imaginary part of s.

Lemma 23. *If $\chi \neq \chi_0$ is a character modulo q and $5/6 \leq \sigma \leq 2$ then*

$$\frac{L'}{L}(s,\chi) = \sum_{\rho} \frac{1}{s-\rho} + O(\log q\tau),$$

where the sum is over all zeros ρ of $L(s,\chi)$ such that $|\rho - (3/2 + it)| \leq 5/6$.

Proof. This follows from Lemma 18 with $f(s) = L(s + 3/2 + it, \chi)$, $R = 5/6$ and $r = 2/3$. We first establish a lower bound for $f(0)$ using the Euler product, so

$$|f(0)| = \prod_p \left|1 - \frac{\chi(p)}{p^{3/2+it}}\right| \geq \prod_p \left(1 + \frac{1}{p^{3/2}}\right)^{-1} \gg 1.$$

We also require an upper bound for $f(s)$ for $|s| \leq 1$. For this, by partial summation, we have, for $\sigma > 0$, the identity

$$L(s, \chi) = s \int_1^\infty \frac{F(t)}{t^{s+1}} \, \mathrm{d}t,$$

where $F(t) = \sum_{1 \leq n \leq t} \chi(n)$. Observing that $|F(t)| \leq q$, by periodicity of q, we deduce that

$$|L(s, \chi)| \ll |s| \, q \int_1^\infty \frac{1}{t^{\sigma+1}} \, \mathrm{d}t,$$

and hence $|f(z)| \ll q\tau$ for $|z| \leq 1$. □

We will also need a similar lemma for the principal character.

Lemma 24. *If χ_0 is the principal character modulo q and $5/6 \leq \sigma \leq 2$ then*

$$\frac{L'}{L}(s, \chi_0) = -\frac{1}{s-1} + \sum_\rho \frac{1}{s-\rho} + O(\log q\tau)$$

where the sum is over all zeros ρ of $L(s, \chi)$ such that $|\rho - (3/2 + it)| \leq 5/6$.

Proof. Comparing Euler products, we see that for $\sigma > 0$,

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

Taking logarithmic derivatives it follows that

$$\frac{L'}{L}(s, \chi_0) = \frac{\zeta'}{\zeta}(s) + \sum_{p|q} \frac{\log p}{p^s - 1}.$$

When $\sigma \geq 5/6$ the sum over p is $\ll \log q$, since that is a trivial bound for the number of primes dividing q, and each summand is $\ll 1$. The lemma now follows from Theorem 13 (when s is close to 1) and Corollary 5. □

Theorem 21. *Let χ be a non-quadratic Dirichlet character modulo q. There is an absolute constant $c > 0$ such that*

$$L(s, \chi) \neq 0 \text{ for } \sigma > 1 - \frac{c}{\log(q\tau)}.$$

The overally strategy is similar to that we used for the zeta function - cleverly choose a linear combination of L'/L at $1 + \delta$, $1 + \delta + it$, and $1 + \delta + 2it$ to create a Dirichlet series which is always non-negative, and then derive a contradiction if σ is close to 1 by taking $\delta \to 0$. There is some complication, however, introduced by the presence of the character χ. Surprisingly, we need to use not only the L-series for χ, but also that for χ_0 and χ^2 in the proof.

Proof. We first note that, comparing Euler products,

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

and so if $L(s,\chi_0)=0$ and $\sigma>0$ then $\zeta(s)=0$, and so in this case the theorem follows from the zero-free region we have established for $\zeta(s)$. We will thus henceforth suppose that χ is a complex character.

Let $\rho=\sigma+it$ be such that $L(\rho,\chi)=0$, and let $\delta>0$ be some parameter to be chosen later. From the Euler product we know that $\sigma\leq 1$. By Lemma 23, as in the proof for the zero-free region for ζ, we have

$$-\Re\frac{L'}{L}(1+\delta+it,\chi)\leq -\frac{1}{1+\delta-\sigma}+O(\log q\tau)$$

and

$$\Re\frac{L'}{L}(1+\delta+2it,\chi^2)\ll \log(q\tau).$$

This last step is where we crucially use the fact that χ is not quadratic, and so χ^2 is not the principal character, since Lemma 23 is not applicable to principal characters.

We also have that, by Lemma 24,

$$-\Re\frac{L'}{L}(1+\delta,\chi_0)=\frac{1}{\delta}+O(\log q).$$

Taking a linear combination of these three inequalities, as in the zero-free region for the zeta function, we deduce that

$$\Re\left(-3\frac{L'}{L}(1+\delta,\chi_0)-4\frac{L'}{L}(1+\delta+it,\chi)-\frac{L'}{L}(1+\delta+2it),\chi^2)\right)$$
$$\leq \frac{3}{\delta}-\frac{4}{1+\delta-\sigma}+O(\log q\tau).$$

The reason for this choice of linear combination, as well as the choice for the three Dirichlet characters, becomes clear when we write the left-hand side as a Dirichlet series,

$$\sum_{\substack{n\geq 1\\(n,q)=1}}\frac{\Lambda(n)}{n^{1+\delta}}\Re(3+4\chi(n)n^{-it}+\chi(n)^2n^{-2it}).$$

If $\chi(n)n^{-it}=e^{i\theta}$ then the real part in this is precisely $3+4\cos\theta+\cos(2\theta)\geq 0$. In particular, this is a series with non-negative summands, and hence

$$\frac{3}{\delta}-\frac{4}{1+\delta-\sigma}+O(\log q\tau)\geq 0.$$

We have a contradiction if $\delta=c_1/\log q\tau$ and $\sigma\geq 1-c_2/\log q\tau$ for suitably chosen constants $c_1,c_2>0$, and the proof is complete. □

We have proved a good zero-free region when χ is not a quadratic character. This case is much harder (as you might guess from the earlier difficulty showing even that $L(1,\chi)\neq 0$ when χ is quadratic), and we can show much less.

Theorem 22. *Let χ be a quadratic character modulo q. There exists a constant $c>0$ such that $L(s,\chi)$ has*

1. *no zeros in the region $\sigma>1-c/\log q\tau$ and $t\neq 0$, and*
2. *at most one real zero $1-c/\log q<\rho<1$.*

We cannot rule out the existence of a real zero of $L(s,\chi)$ very close to 1 when χ is quadratic. These are called **exceptional zeros** (and similarly χ is called an **exceptional character** and q an **exceptional modulus**).

Proof. Suppose that $\rho = \sigma + it$ is a zero of $L(s,\chi)$ and that $t \neq 0$. Let $\delta > 0$ be a parameter to be chosen later. As before,

$$-\Re\frac{L'}{L}(1+\delta+it,\chi) \leq -\frac{1}{1+\delta-\sigma} + O(\log q\tau)$$

and

$$-\Re\frac{L'}{L}(1+\delta,\chi_0) \leq \frac{1}{\delta} + O(\log q\tau).$$

The key difference is now that $\chi^2 = \chi_0$, and so we have the additional term $1/s - 1$ in the expansion of L'/L. When $\tau \geq C(1-\sigma)$ for some suitably large absolute constant $C > 0$, this works in our favour. In this case, we have

$$-\Re\frac{L'}{L}(1+\delta+2it,\chi^2) \leq \Re\frac{1}{\delta+2it} + O(\log q\tau) \leq \frac{\delta}{\delta^2+4t^2} + O(\log q\tau).$$

Taking a linear combination and using non-negativity of the Dirichlet series as before implies that

$$0 \leq \frac{3}{\delta} - \frac{4}{1+\delta-\sigma} + \frac{\delta}{\delta^2+4t^2} + O(\log q\tau).$$

If $\sigma = 1$ this is a contradiction as $\delta \to 0$. Otherwise, we can choose $\delta = 1-\sigma$, and again this a contradiction unless $\sigma \leq 1 - c_1/\log q\tau$ for some constant $c_1 > 0$. Observe the importance of $t \neq 0$ here in ensuring that the third summand is $\leq (1-\epsilon)\frac{1}{\delta}$ for some small $\epsilon > 0$.

For small values of τ we require a different argument, and will no longer compare $1+\delta$, $1+\delta+it$ and $1+\delta+2it$. Instead, we will just use $1+\delta$ and $1+\delta+it$, and use additionally the observation that since $\overline{L(\rho,\chi)} = L(\overline{\rho},\chi)$ (since χ is quadratic), if ρ is a zero then $\overline{\rho}$ is also. Since $t \neq 0$ these are two distinct zeros of L, and so

$$-\Re\frac{L'}{L}(1+\delta+it,\chi) \leq -\Re\left(\frac{1}{1+\delta-\rho} + \frac{1}{1+\delta-\overline{\rho}}\right) + O(\log q\tau).$$

The right-hand side is

$$\frac{-2(1+\delta-\sigma)}{(1+\delta-\sigma)^2+t^2} + O(\log q\tau).$$

It follows that

$$-\Re\left(\frac{L'}{L}(1+\delta,\chi_0) + \frac{L'}{L}(1+\delta+it),\chi)\right) \leq \frac{1}{\delta} + \frac{-2(1+\delta-\sigma)}{(1+\delta-\sigma)^2+t^2} + O(\log q\tau).$$

The left-hand side is

$$\sum_{\substack{n\geq 1\\(n,q)=1}} \frac{\Lambda(n)}{n^{1+\delta}}\Re\left(1+\chi(n)n^{it}\right).$$

Again, if $\chi(n)n^{it} = e^{i\theta}$ then the real part here is $1+\cos\theta$, which is obviously ≥ 0. Once again, we obtain a contradiction choosing $\delta = c_1(1-\sigma)$ if $\sigma \geq 1 - c_2/\log q\tau$ for suitable $c_1, c_2 > 0$.

It remains to consider the case of real zeros. The previous strategy no longer works, since ρ and $\overline{\rho}$ are not distinct zeros. But this idea does allow us to rule out the existence of more than one such real zero. □

The previous theorem shows that for a fixed quadratic character modulo q, there is at most one exceptional zero. We can say a little more, and show that in fact amongst all possible characters for fixed q, there is at most one exceptional zero (and so we can justly talk of 'the' exceptional zero of q).

Lemma 25. *If χ_1 and χ_2 are distinct quadratic characters modulo q then $L(s,\chi_1)L(s,\chi_2)$ has at most one real zero β with $1-c/\log q<\beta<1$.*

Proof. Say β_i is a real zero of $L(s,\chi_i)$ for $i=1,2$. Without loss of generality, $5/6\leq\beta_1\leq\beta_2<1$. Let $\delta>0$ be some parameter to be chosen later. We have

$$-\Re\frac{L'}{L}(1+\delta,\chi_i)\leq-\frac{1}{1+\delta-\beta_i}+O(\log q)$$

and

$$-\Re\frac{L'}{L}(1+\delta,\chi_1\chi_2)\leq O(\log q),$$

using the fact that $\chi_1\chi_2\neq\chi_0$. Finally, we have

$$-\Re\frac{L'}{L}(1+\delta,\chi_0)\leq\frac{1}{\delta}+O(\log q).$$

It follows that

$$-\Re\left(\frac{L'}{L}(1+\delta,\chi_0)+\frac{L'}{L}(1+\delta,\chi_1)+\frac{L'}{L}(1+\delta,\chi_2)+\frac{L'}{L}(1+\delta,\chi_1\chi_2)\right)$$
$$\leq\frac{1}{\delta}-\frac{2}{1+\delta-\beta_1}+O(\log q).$$

The left-hand side is the Dirichlet series

$$\sum_{\substack{n\geq1\\(n,q)=1}}\frac{\Lambda(n)}{n^{1+\delta}}\Re(1+\chi_1(n)+\chi_2(n)+\chi_1\chi_2(n)).$$

The term inside the brackets is $(1+\chi_1(n))(1+\chi_2(n))\geq0$, since both χ_1 and χ_2 take only the values ±1. If we choose $\delta=c_1(1-\beta_1)$ for some constant $c_1>0$, then $\beta_1\leq1-c/\log q$ for some $c>0$ as required.

□

16. Prime number theorem for arithmetic progressions

Recall that in proving Dirichlet's theorem we used the identity

$$\sum_{\substack{n\leq x\\ n\equiv a \pmod q}}\Lambda(n)=\frac{1}{\phi(q)}\sum_{\chi}\overline{\chi(a)}\sum_{n\leq x}\Lambda(n)\chi(n).$$

We therefore need to prove asymptotic formulas for $\psi(x,\chi)=\sum_{n\leq x}\Lambda(n)\chi(n)$, which we will now do using Perron's formula.

Theorem 23. *If $q\leq\exp(O(\sqrt{\log x}))$ then*

(1)
$$\psi(x,\chi_0)=x+O\left(x\exp(-c\sqrt{\log x})\right).$$

(2) *If $\chi\neq\chi_0$ and χ has no exceptional zero then*
$$\psi(x,\chi)=O\left(x\exp(-c\sqrt{\log x})\right).$$

(3) *If $\chi \neq \chi_0$ and χ has an exceptional zero at β then*

$$\psi(x,\chi) = -\frac{x^\beta}{\beta} + O\left(x\exp(-c\sqrt{\log x})\right).$$

Proof. By Perron's formula, Theorem 11, if $1 < \sigma_0 < 2$ and x is not an integer then, for any $T \geq 1$,

$$\psi(x,\chi) = -\frac{1}{2\pi i}\int_{\sigma_0-iT}^{\sigma_0+iT} \frac{L'}{L}(s,\chi)\frac{x^s}{s}\,\mathrm{d}s + O\left(\frac{x}{T}\sum_{x/2<n<2x}\frac{\Lambda(n)}{|x-n|} + \frac{x^{\sigma_0}}{T}\sum_n \frac{\Lambda(n)}{n^{\sigma_0}}\right).$$

By the same analysis as for $\zeta(s)$, the error term is $O(x(\log x)^2/T)$ if we choose $\sigma_0 = 1 + 1/\log x$. We extend the contour integral to the rectangular contour with corners at $\sigma_0 \pm iT$ and $\sigma_1 \pm iT$ where $\sigma_1 < 1$ is chosen to avoid any (non-exceptional zeros) on or inside the contour. The error terms from the short sides and the long left-hand side contribute a total of

$$O\left(\frac{x(\log x)^2}{T} + x^{1-q\sigma_1}\right) = O\left(x\exp(-c\sqrt{\log x})\right)$$

if we choose $\sigma_1 = 1 - c_1/\log qT$ and $T = \exp(O\left(\sqrt{\log x}\right))$.

It remains to note that the integral around the integral contour is x if $\chi = \chi_0$ (from the simple pole at $s = 1$), is 0 if $\chi \neq \chi_0$ has no exceptional zeros (since there are no zeros or poles of $L(s,\chi)$ inside the contour), and x^β/β if χ has an exceptional zero at β. □

Using the previous identity, we can immediately deduce the following prime number theorem for arithmetic progressions.

Corollary 6. *Let $(a,q) = 1$ and $q \leq \exp(O(\sqrt{\log x}))$. If q has no exceptional zero then*

$$\psi(x;q,a) = \frac{x}{\phi(q)} + O(x\exp(-c\sqrt{\log x})).$$

If q has an exceptional zero at β and χ_1 then

$$\psi(x;q,a) = \frac{x}{\phi(q)} - \frac{\chi_1(a)x^\beta}{\phi(q)\beta} + O(x\exp(-c\sqrt{\log x})).$$

17. Siegel-Walfisz theorem

The prime number theorem we established in the previous section is quite frustrating in that we have two different results, depending on the existence of an exceptional zero. One may ask whether, if q is small enough, this obstruction can be overcome.

The answer is yes, and is the following.

Theorem 24 (Siegel-Walfisz)**.** *For all $A > 0$, if $(a,q) = 1$ and $q \leq (\log x)^A$ then*

$$\psi(x;q,a) = \frac{x}{\phi(q)} + O_A(x\exp(-c\sqrt{\log x})).$$

This follows immediately from the following similar expression for $\psi(x,\chi)$.

Theorem 25. *If $q \leq (\log x)^A$ and x is large enough (depending only on A) and if $\chi \neq \chi_0$ then*

$$\psi(x,\chi) = O_A(\exp(-c\sqrt{\log x})).$$

In particular, note that this bound holds regardless of whether or not q has an exceptional zero at χ. This in turn follows from the following result of Siegel.

Theorem 26. *For all $\epsilon > 0$ there exists C_ϵ such that if χ is a quadratic character modulo q and β is a real zero then*

$$\beta < 1 - C_\epsilon q^{-\epsilon}.$$

Proof. Omitted; the curious student may find a proof in many standard texts on analytic number theory, such as Davenport's Multiplicative Number Theory or Montgomery and Vaughan's Multiplicative Number Theory. □

A curious feature of Siegel's result is that the constant C_ϵ is **ineffective** – that is, we are not using the notation C_ϵ to hide some constant that we can't be bothered to work out exactly, but there is no way to find from the proof how the constant depends on ϵ at all. In turn, this means that the constant in the Siegel-Walfisz theorem is also ineffective.

Proof that Theorem 26 implies Theorem 25. If χ has no exceptional zero then we have already proved this. Suppose that χ has an exceptional zero at β. By Theorem 26, for any $\epsilon > 0$, we know $\beta < 1 - C_\epsilon q^{-\epsilon}$. It follows that

$$\psi(x,\chi) = O\left(\frac{x^\beta}{\beta} + x\exp(-c\sqrt{\log x})\right) \quad = xO\left(\exp(-C_\epsilon q^{-\epsilon}\log x) + \exp(-c\sqrt{\log x})\right).$$

Using the bound $q \leq (\log x)^A$ the error term here is $O(\exp(-c_\epsilon\sqrt{\log x}))$ as required, if we choose $\epsilon = 1/3A$, say. □