

TOPICS IN NUMBER THEORY (PART B)

BEN GREEN

CONTENTS

1. Distribution of primes and Mertens' estimates	3
2. Arithmetic functions	7
3. The Riemann ζ -function	15
4. Dirichlet characters and L -functions.	18
5. Binary quadratic forms	25
6. Sums of four squares	32
7. Irrationals and transcendental numbers, approximation	34
8. Continued fractions	37
9. Continued fractions and approximation by rationals	46
Appendix A. Results from real analysis	47
Appendix B. Results from complex analysis	49
Appendix C. Real quadratic characters	49
References	51

Overview. Notes for Part B Topics in Number Theory, a new course being taught for the first time in 2026/7. The synopsis below is the same as the one advertised, up to ordering, except for items explicitly added **in red**. These were added during the process of writing the course.

Synopsis.

Elementary distribution of primes. Upper and lower bounds on $\pi(X)$ (Chebyshev bounds). **Mertens' estimates** [2 lectures]

Arithmetic functions and their distribution. **Möbius inversion and Dirichlet convolution.** The functions τ (number of divisors), ω (number of prime factors) and ϕ (Euler's totient function). Mean value of τ and the hyperbola method. Hardy-Ramanujan law and normal order of ω and τ . Divisor bound, mean value of ϕ and the probability that random integers are coprime. Introduction to Dirichlet series. [3 lectures]

The Riemann ζ -function and Euler's proof of infinitely many primes. Analytic continuation to $\operatorname{Re} s > 0$ and statement of the Riemann Hypothesis. Discussion of the prime number theorem and consequences of the Riemann Hypothesis. Dirichlet characters and L -functions. Dirichlet's proof that

there are infinitely many primes $a \pmod q$ when a, q are coprime, assuming $L(1, \chi) \neq 0$. Proof that $L(1, \chi) \neq 0$ in special cases. [3 lectures]

Binary quadratic forms and sums of squares. Basic reduction theory of positive definite binary forms, class number. Review of quadratic reciprocity. Representation of primes by positive definite binary quadratic forms. Brief discussion on [the Kronecker symbol, quadratic characters and the class number formula](#). Lagrange's theorem on sums of four squares. [4 lectures]

Irrationality, approximation and continued fractions. Continued fractions of rationals and Euclid's algorithm. Dirichlet's theorem. Best approximations. Continued fractions of quadratic irrationals. Pell's equation. Irrationality of e . Construction of a transcendental number. [4 lectures]

Reading list. Full notes will be provided for the course but there are several excellent books which provide extra discussion and context. Perhaps the most relevant are

Davenport, H. *The higher arithmetic*, 8th edition, Cambridge University Press 2008.

Hardy, G. H. and Wright, E. M. *An introduction to the theory of numbers*, 6th edition, Oxford University Press 2008.

Further reading. The following further texts or parts of text are at an appropriate level for the course and provide further context.

A nice introduction to probabilistic number theory is the recent book

Kowalski, E. *An introduction to probabilistic number theory*, Cambridge Studies in Advanced Mathematics **192**, Cambridge University Press 2021.

Another recommended book is

Erdős, P. and Surányi, J., *Topics in the theory of numbers*, Undergraduate Texts in Mathematics, Springer-Verlag 2003.

Another highly-recommended text is Serre's classic

Serre, J.-P. *A course in arithmetic*, Graduate Texts in Mathematics **7**, Springer-Verlag 1973

Montgomery and Vaughan's book *Multiplicative Number Theory I. Classical Theory*. Cambridge Advanced Studies in Math **97** is a useful text for some of the details.

Sheet X. Sheet X is an exercise sheet which goes over the details of various items we will use without proof in the course, for instance basic analytic estimates for sums. These are all the sort of things at the level of prelims or Part A courses. To follow the present course, all you need is the statements of these results, but it should be helpful to refresh your memory of the proofs.

Notation. If m, n are two positive integers, then I like to write (m, n) for the highest common factor of m and n . This is pretty standard notation,

and is almost ubiquitous in research papers. Some authors write $\text{hcf}(m, n)$ or $\text{gcd}(m, n)$; you may find this in past exam papers.

Although we will not use it in this course, I write $[m, n]$ for the lowest common multiple of m and n . Some authors use $\text{lcm}(m, n)$ or, worse still, $\text{l.c.m.}(m, n)$.

The notation $\lfloor x \rfloor$ ‘the floor of x ’ means the greatest integer less than or equal to x . For example, $\lfloor \pi \rfloor = 3$.

Throughout the course we will use number 1 to mean the indicator of some event, for example $1_{n \leq N}$ equals 1 if $n \leq N$ and 0 otherwise.

In analytic number theory, $\log x$ always means natural log and we never write $\ln x$.

An important part of this course will be introducing so-called *asymptotic notation*: the big-oh notation $O(\cdot)$ and similar things. This has its own section, Section 1.2 below.

I use the abbreviations LHS and RHS for left-hand side and right-hand side respectively.

Use of AI tools. I have written these notes myself using published resources and books, as I always have done. Completed versions of each section were put through an LLM to pick up typos and minor inaccuracies, which were then corrected. Suggestions by the LLM to change content or exposition were ignored.

1. DISTRIBUTION OF PRIMES AND MERTENS’ ESTIMATES

Our first topic of discussion is the distribution of prime numbers. The main highlights of this section are bounds for $\pi(X)$, the number of primes less than or equal to X . We will also prove two estimates of Mertens, the second of which gives an asymptotic formula for $\sum_{p \leq X} 1/p$. (Whenever we write $\sum_p$ it will mean a sum over primes.)

1.1. Chebyshev bounds. Our first main result is that the number of primes $\leq X$ is on the order of $X/\log X$.

Theorem 1.1. *There are absolute constants $0 < c_1 < c_2$ such that*

$$c_1 \frac{X}{\log X} \leq \pi(X) \leq c_2 \frac{X}{\log X}$$

for all $X \geq 2$.

Proof. We begin with the lower bound. We look at the prime factorisation of $\binom{2n}{n}$, which we write as

$$\binom{2n}{n} = \prod_{p \leq 2n} p^{w_p(n)}.$$

It follows from Legendre's formula for the number of powers of p dividing $m!$ (see Sheet X) and the definition of binomial coefficients that

$$w_p(n) = \sum_{i=1}^{\infty} \left\lfloor \frac{2n}{p^i} \right\rfloor - 2 \left\lfloor \frac{n}{p^i} \right\rfloor. \quad (1.1)$$

Now each term in (1.1) is at most 1, since $\lfloor 2x \rfloor - 2\lfloor x \rfloor \in \{0, 1\}$ for all real x . Moreover, the terms can only be non-zero for $i \leq \log(2n)/\log p$. It follows that

$$p^{w_p(n)} \leq p^{\log(2n)/\log p} = 2n.$$

Taking products over p (noting that all primes dividing $\binom{2n}{n}$ are $\leq 2n$) we obtain

$$\binom{2n}{n} \leq (2n)^{\pi(2n)}. \quad (1.2)$$

Now $\binom{2n}{n}$ is almost as big as 4^n ; note that $\sum_{r=0}^{2n} \binom{2n}{r} = 4^n$ and, in this sum, the middle term $\binom{2n}{n}$ is the largest one. It follows from this observation and (1.2) that, for n sufficiently large,

$$2^n < \frac{4^n}{2n+1} \leq \binom{2n}{n} \leq (2n)^{\pi(2n)}.$$

Taking logs gives

$$\pi(2n) \geq \frac{n \log 2}{\log(2n)}.$$

This is the lower bound in [Theorem 1.1](#) in the case $X = 2n$ an even integer (with $c_1 = \frac{1}{2} \log 2$). The general case follows easily by considering the greatest even integer $\leq X$, and reducing c_1 slightly. We leave the details as an exercise.

We turn now to the upper bound. At the cost of increasing c_2 by a factor of 2, it is enough to handle the case $X = 2^k$, k a positive integer (since $t/\log t$ is an increasing function of t). We again consider the binomial coefficient $\binom{2n}{n}$, noting now that if n is an integer and if $n < p \leq 2n$ then p divides $\binom{2n}{n}$ precisely once. Therefore

$$n^{\pi(2n) - \pi(n)} \leq \prod_{n < p \leq 2n} p \leq \binom{2n}{n} \leq 4^n,$$

and so, taking logs,

$$\pi(2n) - \pi(n) \leq \frac{n \log 4}{\log n}.$$

Applying this with $n = 2, 2^2, 2^3, \dots, 2^{k-1}$ and summing gives

$$\pi(2^k) \leq 2^k \left(\frac{1}{k-1} + \frac{1}{2(k-2)} + \frac{1}{2^2(k-3)} + \dots + \frac{1}{2^{k-2}} \right).$$

We claim that the bracketed expression is bounded by C/k for some absolute constant C . There are many ways to see this. For instance, the ratio between

the terms $1/2^{i-1}(k-i)$ and $1/2^i(k-i-1)$ is $2(k-i-1)/(k-i)$, which is $\geq 4/3$ for $i \leq k-3$, so

$$\sum_{i=2}^{k-1} \frac{1}{2^{i-1}(k-i)} \leq \frac{1}{k-1} \left(1 + \frac{3}{4} + \left(\frac{3}{4}\right)^2 + \dots \right) + \frac{1}{2^{k-2}},$$

which implies the stated bound that $\pi(2^k) \leq C2^k/k$ for some C . $\square$

The famous *prime number theorem* is the assertion that one can take both c_1 and c_2 arbitrarily close to 1 for large X , namely

$$\lim_{X \rightarrow \infty} \frac{\pi(X)}{X/\log X} = 1 \quad (1.3)$$

This is quite a difficult theorem and we will not prove it in this course, though we will make some further comments later on.

1.2. Asymptotic notation. Before continuing, we introduce some notation which is commonplace throughout the more analytic part of number theory, and which we will use quite extensively in the course. This is the use of ‘asymptotic’ notations $\ll, \gg$ and O , which are used to describe the rough size of error terms. The statements $X = O(Y)$, $X \ll Y$ and $Y \gg X$ all mean the same thing, namely that for some absolute constant C we have $|X| \leq CY$. The main point of the notation is that we do not specify the constant C . Usually, this is because we do not need to for applications. It might be tedious to produce an explicit C that works, or occasionally impossible. Different instances of the notation may imply different values of C .

One often sees the $O()$ notation used by itself for error terms. An example is that for $X \geq 1$ we have

$$\sum_{n \leq X} \log n = X \log X + O(X), \quad (1.4)$$

which is (a weak) form of Stirling’s formula, which we will use below. What this means is that

$$\left| \sum_{n \leq X} \log n - X \log X \right| \leq CX$$

for some absolute constant C , and for all $X \geq 1$. For the proof of (1.4), see Sheet X.

1.3. Mertens estimates. We now turn to two estimates of Mertens, which give interesting information about the distribution of primes. Here, and throughout the rest of the course, $\sum_p$ means a sum over primes, whilst $\prod_p$ denotes a product over primes. (We will discuss infinite products a little more when we get to them.)

Proposition 1.2. *We have Mertens' first estimate*

$$\sum_{p \leq X} \frac{\log p}{p} = \log X + O(1) \quad (1.5)$$

*and Mertens' second estimate*¹

$$\sum_{p \leq X} \frac{1}{p} = \log \log X + O(1). \quad (1.6)$$

Remark. It is at first sight surprising that it is not too hard to prove these two statements, whereas it is much more difficult (the prime number theorem) to get an asymptotic for $\sum_{p \leq X} 1$. What makes this so is that the ‘weights’ $\log p/p$ and $1/p$ in (1.5) and (1.6) are much ‘smoother’ than the weight $1_{p \leq X}$.

Proof. We first prove (1.5). It suffices to handle the case X an integer. The idea is to look at the prime factorization of $X!$. By Legendre’s formula (see Sheet X), the power of p dividing $X!$ is $\sum_{i \geq 1} \lfloor \frac{X}{p^i} \rfloor$. Thus, taking logs,

$$\log X! = \sum_{p \leq X} \log p \left(\lfloor \frac{X}{p} \rfloor + \lfloor \frac{X}{p^2} \rfloor + \dots \right) \quad (1.7)$$

By Stirling’s formula (in the weak form (1.4)) we have

$$\log X! = \sum_{n \leq X} \log n = X \log X + O(X). \quad (1.8)$$

Turning to the right-hand side of (1.7), we first look at the contribution from the terms $\lfloor \frac{X}{p^i} \rfloor$ with $i \geq 2$. This contribution is bounded by

$$X \sum_{p \leq X} (\log p) \left(\frac{1}{p^2} + \frac{1}{p^3} + \dots \right) \leq 2X \sum_{p \leq X} \frac{\log p}{p^2} \ll X;$$

in fact the sum over *integers* $\sum_{n \leq X} \frac{\log n}{n^2}$ is convergent (see Sheet X). Now we look at the contribution to the RHS of (1.7) from the terms $\lfloor X/p \rfloor$. We have $\lfloor X/p \rfloor = \frac{X}{p} + O(1)$, and the contribution from the first of these terms is precisely $X \sum_{p \leq X} \frac{\log p}{p}$, that is to say X times the sum we are interested in. The contribution from the $O(1)$ terms is

$$\ll \sum_{p \leq X} \log p \leq \pi(X) \log X \ll X,$$

where for the second bound we used the upper Chebyshev estimate (that is, the upper bound in Theorem 1.1). Putting all this together gives that the RHS of (1.7) is

$$X \sum_{p \leq X} \frac{\log p}{p} + O(X).$$

¹This is actually a weak form of the result usually known as Mertens’ second estimate, which is that $\sum_{p \leq X} \frac{1}{p} = \log \log X + B + O(\frac{1}{\log X})$ for some constant B . See, for instance, Montgomery and Vaughan Theorem 2.7 (d).

Since the LHS of (1.7) is given as in (1.8), comparing the two gives the desired bound (1.5).

We will deduce the second Mertens estimate (1.6) from the first one using a technique called partial summation. What makes this work is that $\frac{1}{p}$ is a ‘smoother’ weight than $\frac{\log p}{p}$, which allows us to express $\sum_{p \leq X} \frac{1}{p}$ in terms of an average

$$\int_2^X f(x) \left(\sum_{p \leq x} \frac{\log p}{p} \right) dx, \quad (1.9)$$

for a suitably chosen f . To motivate the choice of f , we swap the order of summation and integration in (1.9) to get

$$\sum_{p \leq X} \frac{\log p}{p} \int_p^X f(x) dx.$$

This suggests the choice $f(x) = \frac{1}{x(\log x)^2}$ (that is, the derivative of $-\frac{1}{\log x}$), since then

$$\int_p^X f(x) dx = \frac{1}{\log p} - \frac{1}{\log X}.$$

With this choice of f , we therefore have

$$\sum_{p \leq X} \frac{1}{p} = \frac{1}{\log X} \sum_{p \leq X} \frac{\log p}{p} + \int_2^X \frac{dx}{x(\log x)^2} \left(\sum_{p \leq x} \frac{\log p}{p} \right). \quad (1.10)$$

Using the first Mertens estimate (1.5), this is

$$O(1) + \int_2^X \frac{dx}{x(\log x)^2} (\log x + O(1)) = O(1) + \int_2^X \frac{dx}{x \log x},$$

using here that the integral $\int_2^X \frac{dx}{x(\log x)^2} = \frac{1}{\log 2} - \frac{1}{\log X} = O(1)$. To complete the proof, we note that $\int_2^X \frac{dx}{x \log x} = \log \log X - \log \log 2$. $\square$

2. ARITHMETIC FUNCTIONS

In this section and the next we study *arithmetic functions*, which are simply functions $f : \mathbf{N} \rightarrow \mathbf{C}$.

2.1. The main functions. Divisor bound. We will particularly focus on functions of particular interest in number theory, and these are

- The divisor function τ , where $\tau(n)$ is defined to be the number of positive integer divisors of n (including 1 and n);
- The number of distinct prime factors of n , denoted by $\omega(n)$;
- The number of prime factors of n counted with multiplicity, denoted by $\Omega(n)$;
- Euler’s ϕ function, where $\phi(n)$ is defined to be the number of integers x with $1 \leq x \leq n$ which are coprime to n , or equivalently the order of the multiplicative group $(\mathbf{Z}/n\mathbf{Z})^\times$;

- The Möbius function $\mu(n)$, defined to be $(-1)^k$ if n is a product of k distinct primes, and 0 otherwise.

Multiplicativity. We say that an arithmetic function f is multiplicative if $f(mn) = f(m)f(n)$ whenever m, n are coprime. (Note carefully that we do not require $f(mn) = f(m)f(n)$ for all m, n ; that property, when it holds, is called *complete multiplicativity*.)

Lemma 2.1. *The functions τ, ϕ, μ are multiplicative.*

Proof. For μ this follows immediately from the definition. That $\phi(mn) = \phi(m)\phi(n)$ when m, n are coprime follows from the Chinese remainder theorem; in fact $(\mathbf{Z}/mn\mathbf{Z})^\times \cong (\mathbf{Z}/m\mathbf{Z})^\times \times (\mathbf{Z}/n\mathbf{Z})^\times$. This was shown in ASO Number Theory. Finally, for τ we note that if m and n are coprime then the divisors of mn are precisely the integers dd' with $d \mid m$ and $d' \mid n$, and that all such integers are distinct. $\square$

A surprisingly powerful fact, which comes up everywhere in number theory, is that the number of divisors of n is rather small. This is called the *divisor bound*.

Proposition 2.2 (Divisor bound). *Let $\varepsilon > 0$. Then there is a constant $C_\varepsilon > 0$ such that*

$$\tau(n) \leq C_\varepsilon n^\varepsilon$$

for all n .

Remark. Quite often one will see this bound written as $\tau(n) \ll_\varepsilon n^\varepsilon$; it is understood here that $\varepsilon > 0$ can be arbitrary. Another notation one will sometimes see is $\tau(n) = n^{o(1)}$.

Proof. We first note a formula for $\tau(n)$ in terms of the prime factorisation $n = p_1^{a_1} \cdots p_k^{a_k}$, namely

$$\tau(n) = (a_1 + 1) \cdots (a_k + 1). \quad (2.1)$$

This follows immediately from the multiplicativity of τ and the fact that $\tau(p^a) = a + 1$, since the divisors of p^a are precisely the elements of the set $\{1, p, p^2, \dots, p^a\}$.

We have $a + 1 \leq e^a \leq p^{a\varepsilon}$ provided $p > e^{1/\varepsilon}$.

Since $\lim_{a \rightarrow \infty} \frac{a+1}{2^{a\varepsilon}} = 0$, there is some constant B (depending on ε) such that $a + 1 \leq B2^{a\varepsilon}$ for all a . We of course have $a + 1 \leq Bp^{a\varepsilon}$ for all $p \geq 2$.

By (2.1), the result now follows with $C_\varepsilon := B^m$, where m is the number of primes $\leq e^{1/\varepsilon}$. $\square$

2.2. Dirichlet convolution and Möbius inversion. If $f, g : \mathbf{N} \rightarrow \mathbf{C}$ are two arithmetic functions then we define

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

This operation is known as *Dirichlet convolution*. It is easy to check that it is commutative (that is, $f \star g = g \star f$) and associative (that is, $f \star (g \star h) = (f \star g) \star h$). It also has an identity, which is the function δ such that $\delta(1) = 1$ and $\delta(n) = 0$ for all other n . For this function, we have $f \star \delta = f$ for any function f .

The following result is known as *Möbius inversion*. Here, 1 denotes the arithmetic function taking the value 1 at all inputs.

Proposition 2.3. *Let $f, g : \mathbf{N} \rightarrow \mathbf{C}$ be two arithmetic functions. Then $g = f \star 1$ if and only if $f = g \star \mu$.*

Proof. Note that $\mu \star 1 = \delta$. This is a routine check: if $n = p_1^{a_1} \dots p_k^{a_k}$ with $k \geq 1$ (and the p_i primes) then

$$\begin{aligned} \sum_{d|n} \mu(d) &= \sum_{\varepsilon_i \in \{0,1\}} \mu(p_1^{\varepsilon_1} \dots p_k^{\varepsilon_k}) \\ &= \sum_{\varepsilon_i \in \{0,1\}} \mu(p_1^{\varepsilon_1}) \dots \mu(p_k^{\varepsilon_k}) = (1-1) \dots (1-1) = 0. \end{aligned}$$

Here we used the multiplicativity of μ . Now if $g = f \star 1$ then

$$g \star \mu = (f \star 1) \star \mu = f \star (1 \star \mu) = f \star \delta = f.$$

Conversely if $f = g \star \mu$ then

$$f \star 1 = (g \star \mu) \star 1 = g \star (\mu \star 1) = g \star \delta = g.$$

This concludes the proof. $\square$

2.3. Dirichlet series. Suppose that $f : \mathbf{N} \rightarrow \mathbf{C}$ is an arithmetic function. Then we define the *Dirichlet series*

$$D_f(s) := \sum_{n=1}^{\infty} f(n)n^{-s}. \quad (2.2)$$

The definition is not really meaningful unless the series is known to converge (though there is a notion of a ‘formal Dirichlet series’). A condition on f which comes up in practice is that, for some $a \in \mathbf{R}$, we have

$$|f(n)| \leq C_\varepsilon n^{a+\varepsilon} \quad (2.3)$$

for all $\varepsilon > 0$ (often this would be written as $|f(n)| \leq n^{a+o(1)}$). For example, the divisor bound implies that the divisor function τ satisfies this condition with $a = 0$, and clearly the constant function 1 satisfies this condition, also with $a = 0$. We have the following analytic result.

Lemma 2.4. *Suppose that f is an arithmetic function satisfying (2.3). Then $D_f(s)$ converges to a holomorphic function in the domain $\{s \in \mathbf{C} : \operatorname{Re} s > a + 1\}$.*

Proof. Suppose that $\operatorname{Re} s \geq a + 1 + \delta$. Then for $N > N_0(\delta)$ sufficiently large we have

$$\left| \sum_{n=N+1}^{\infty} f(n)n^{-s} \right| \leq \sum_{n=N+1}^{\infty} n^{a+\delta/2} n^{-a-1-\delta} \ll N^{-\delta/2}.$$

(See Sheet X for the last estimate.) Thus if we set

$$D_f^{(N)}(s) := \sum_{n=1}^N f(n)n^{-s} \quad (2.4)$$

then $D_f^{(N)}(s) \rightarrow D_f(s)$ uniformly in $\operatorname{Re} s \geq a + 1 + \delta$, and so certainly on $\operatorname{Re} s > a + 1 + \delta$. Since a uniform limit of holomorphic functions on a domain is holomorphic on that domain (Proposition B.1), it follows that $D_f(s)$ is holomorphic on $\operatorname{Re} s > a + 1 + \delta$. Since $\delta > 0$ was arbitrary, the result follows. $\square$

We now establish a result linking Dirichlet convolution and Dirichlet series. Formally, the result is that $D_{f_1 \star f_2} = D_{f_1} D_{f_2}$, and the proof is morally

$$\begin{aligned} D_{f_1 \star f_2}(s) &= \sum_n \left(\sum_{n_1 n_2 = n} f_1(n_1) f_2(n_2) \right) n^{-s} \\ &= \left(\sum_{n_1} f_1(n_1) n_1^{-s} \right) \left(\sum_{n_2} f_2(n_2) n_2^{-s} \right) = D_{f_1}(s) D_{f_2}(s). \end{aligned}$$

However, one must worry about convergence. The following lemma handles this more carefully.

Lemma 2.5. *Suppose that f_1, f_2 satisfy (2.3) with exponents $a_1, a_2 \geq 0$ respectively. Then $f := f_1 \star f_2$ satisfies (2.3) with exponent $a_1 + a_2$, and so $D_{f_1}, D_{f_2}, D_{f_1 \star f_2}$ are holomorphic on $\{s \in \mathbf{C} : \operatorname{Re} s > a_1 + a_2 + 1\}$. On this domain, we have $D_{f_1 \star f_2}(s) = D_{f_1}(s) D_{f_2}(s)$.*

Proof. For the proof (which is not examinable) see Sheet 1. $\square$

In the remainder of this section we look in greater detail at the three specific functions ω, τ, ϕ .

2.4. The number of prime factors. In this section we look at $\omega(n)$, the number of distinct prime factors of n . We will study the typical behaviour of $\omega(n)$ for $n \leq N$.

The second Mertens estimate (1.6) quite easily gives the following result on the average behaviour of $\omega(n)$.

Proposition 2.6. *We have*

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

Proof. Given n , write $X_p(n) = 1$ if $p \mid n$, and 0 otherwise. Then

$$\omega(n) = \sum_{p \leq N} X_p(n). \quad (2.5)$$

Since the number of $n \leq N$ which are divisible by p is $\frac{N}{p} + O(1)$, we have using (1.6) that

$$\sum_{n \leq N} \omega(n) = \sum_{p \leq N} \sum_{n \leq N} X_p(n) = \sum_{p \leq N} \left(\frac{N}{p} + O(1) \right) = N(\log \log N + O(1)).$$

This completes the proof. $\square$

We will go further than this and show that $\omega(n)$ is in fact strongly concentrated about $\log \log N$ for $n \leq N$, a result known as the Hardy-Ramanujan Theorem. The key to this is the following second moment estimate.

Proposition 2.7. *We have*

$$\frac{1}{N} \sum_{n \leq N} \omega(n)^2 \leq (\log \log N)^2 + O(\log \log N).$$

Proof. Using (2.5) again, we have

$$\sum_{n \leq N} \omega(n)^2 = \sum_{p, p' \leq N} \#\{n \leq N : p, p' \mid n\}.$$

We split the sum into $p = p'$ and $p \neq p'$ and use that

$$\#\{n \leq N : p \mid n\} \leq \frac{N}{p}$$

and that, when $p \neq p'$,

$$\#\{n \leq N : p, p' \mid n\} \leq \frac{N}{pp'}.$$

It follows that

$$\frac{1}{N} \sum_{n \leq N} \omega(n)^2 \leq \sum_{p \leq N} \frac{1}{p} + \sum_{\substack{p, p' \leq N \\ p \neq p'}} \frac{1}{pp'} = \left(\sum_{p \leq N} \frac{1}{p} \right)^2 - \sum_{p \leq N} \frac{1}{p^2} + \sum_{p \leq N} \frac{1}{p}.$$

The result follows from the second Mertens estimate, (1.6). $\square$

Remarks. This argument is an example of the *second moment method*. It is possible to think of this probabilistically. The prime factors counting function $\omega(n)$ is a sum $\sum_p X_p(n)$. If we imagine that $\mathbf{n}$ is sampled uniformly at random from $\{1, \dots, N\}$, each $X_p(\mathbf{n})$ is a random variable, and $\omega(\mathbf{n})$ is a sum of these. The intuition is then that the $X_p(\mathbf{n})$ are approximately pairwise independent. Proposition 2.7 then amounts to verifying that these variables are indeed close to pairwise independent by computing the variance.

One could run a similar analysis with $\omega(n)$ replaced by $\Omega(n)$, which counts the number of prime factors of n *with multiplicity*. However, we also have the following result which states that these two functions are close on average.

Proposition 2.8. *We have*

$$\frac{1}{N} \sum_{n \leq N} (\Omega(n) - \omega(n)) \leq 4.$$

Proof. The left-hand side is equal to

$$\frac{1}{N} \sum_{n \leq N} \sum_{j \geq 2} \sum_p 1_{p^j | n}.$$

We have $\sum_{n \leq N} 1_{d|n} = \lfloor N/d \rfloor \leq N/d$ for any d , so this is bounded above by

$$\sum_p \sum_{j \geq 2} p^{-j} = \sum_p \frac{1}{p^2} \frac{1}{1 - 1/p} \leq 2 \sum_p \frac{1}{p^2} < 4.$$

□

The following corollary is a statement that ω and Ω are concentrated about their means (much more precise statements are possible).

Corollary 2.9. *Let $\varepsilon > 0$. Then, for N large enough in terms of ε , at least $(1 - \varepsilon)N$ integers $n \leq N$ have $(1 - \varepsilon) \log \log N \leq \omega(n) \leq (1 + \varepsilon) \log \log N$. A similar result holds for $\Omega(n)$.*

Proof. Using [Proposition 2.6](#) and [Proposition 2.7](#) we compute

$$\frac{1}{N} \sum_{n \leq N} (\omega(n) - \log \log N)^2 \ll \log \log N.$$

It follows that, writing

$$S(\varepsilon, N) := \{n \leq N : \omega(n) \notin [(1 - \varepsilon) \log \log N, (1 + \varepsilon) \log \log N]\},$$

we have

$$|S(\varepsilon, N)| (\varepsilon \log \log N)^2 \ll N \log \log N,$$

which means that $|S(\varepsilon, N)| \ll (\varepsilon^2 \log \log N)^{-1} N$. In particular, if N is large enough then $|S(\varepsilon, N)| \leq \varepsilon N$. This establishes the result for ω .

To derive the result for Ω , observe that from [Proposition 2.8](#) it follows that if we write

$$T(\varepsilon, N) := \{n \leq N : \Omega(n) - \omega(n) \geq \varepsilon \log \log N\}$$

then

$$|T(\varepsilon, N)| \leq \frac{4}{\varepsilon \log \log N} N \leq \varepsilon N,$$

if N is big enough in terms of ε . Finally, note that if $n \notin S(\varepsilon/2, N) \cup T(\varepsilon/2, N)$ (a set of size $\leq \varepsilon N$) then $(1 - \varepsilon) \log \log N \leq \Omega(n) \leq (1 + \varepsilon) \log \log N$. □

Here is a fun application of the above bound for $\Omega(n)$. It concerns the $N \times N$ multiplication table (where the 10×10 or 12×12 table is the one you learned at school).

Theorem 2.10. *Most integers $\leq N^2$ do not lie in the $N \times N$ multiplication table.*

Proof. (Sketch) A typical integer $n \leq N^2$ has $\Omega(n) \approx \log \log N^2 \approx \log \log N$. A typical element $n' = n_1 n_2$ in the multiplication table has $\Omega(n') = \Omega(n_1) + \Omega(n_2) \approx 2 \log \log N$. $\square$

We also have the following result on the typical size (or ‘normal order’) of the divisor function.

Theorem 2.11. *Let $\varepsilon > 0$. Then, if N is sufficiently large in terms of ε , we have*

$$(\log N)^{\log 2 - \varepsilon} \leq \tau(n) \leq (\log N)^{\log 2 + \varepsilon}$$

for at least $(1 - \varepsilon)N$ values of $n \leq N$.

Proof. This follows quickly from [Corollary 2.9](#) and the inequalities

$$2^{\omega(n)} \leq \tau(n) \leq 2^{\Omega(n)},$$

which are immediate from [\(2.1\)](#) and the fact that $2 \leq a + 1 \leq 2^a$ for all integers $a \geq 1$. $\square$

Remark. Sometimes one sees this result written in the following shorthand: $\tau(n) = (\log n)^{\log 2 + o(1)}$ almost surely. This means precisely the statement of [Theorem 2.11](#).

2.5. More on the divisor function. We have already established the *divisor bound*, which is the statement that $\tau(n) \ll_{\varepsilon} n^{\varepsilon}$. In [Theorem 2.11](#) we established the typical size of $\tau(n)$. In this section we study the *average* size of $\tau(n)$ which, surprisingly at first sight, is much larger than the typical size of $\tau(n)$. This is not a contradiction: the bulk of the contribution to the average value $\frac{1}{N} \sum_{n \leq N} \tau(n)$ turns out to be from (exceptional) integers n with more than the typical number of prime factors.

We will need Euler’s constant γ , defined by

$$\gamma := \lim_{Y \rightarrow \infty} \sum_{n \leq Y} \frac{1}{n} - \log Y.$$

This constant exists and has the value $\approx 0.577 \dots$. Moreover, we have

$$\sum_{a \leq Y} \frac{1}{a} = \log Y + \gamma + O\left(\frac{1}{Y}\right) \quad (2.6)$$

For these two facts, see Sheet X.

Theorem 2.12. *We have*

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

In particular, the average size of $\tau(n)$ on $[1, N]$ is essentially $\log N$.

Proof. The argument is known as Dirichlet's hyperbola method. Since $\tau(n)$ is the number of pairs (a, b) of positive integers with $ab = n$, we have²

$$\sum_{n \leq N} \tau(n) = \sum_{a, b} 1_{ab \leq N}.$$

We divide the sum up into (1) $a, b \leq \sqrt{N}$, (2) $a > \sqrt{N}$ and (3) $b > \sqrt{N}$, noting that these possibilities cover everything and are mutually exclusive.

On the range (1), we always have $ab \leq N$ and so the sum is simply the number of pairs (a, b) with $a \leq \sqrt{N}, b \leq \sqrt{N}$, which is $(\sqrt{N} + O(1))^2 = N + O(\sqrt{N})$.

The contributions from (2) and (3) are the same, and each is

$$\sum_{a > \sqrt{N}} \sum_{b \leq N/a} 1 = \sum_{b \leq \sqrt{N}} \sum_{\sqrt{N} < a \leq N/b} 1 = \sum_{b \leq \sqrt{N}} \sum_{a \leq N/b} 1 - N + O(\sqrt{N}).$$

Since

$$\sum_{a \leq N/b} 1 = \frac{N}{b} + O(1),$$

this is

$$N \sum_{b \leq \sqrt{N}} \frac{1}{b} - N + O(\sqrt{N}).$$

We therefore see that the contribution from (2) (and also from (3)) is

$$\frac{1}{2} N \log N + (\gamma - 1)N + O(\sqrt{N}).$$

Adding the contributions from (1), (2) and (3) gives the result. $\square$

2.6. Euler's ϕ function. The main theorem of this section is the following.

Theorem 2.13. *We have*

$$\sum_{n \leq N} \phi(n) = \frac{3}{\pi^2} N^2 + O(N \log N).$$

Consequently, the probability that two random integers are coprime is $6/\pi^2$.

²Here $1_{ab \leq N}$ denotes the indicator of the event $ab \leq N$, that is to say it equals 1 if $ab \leq N$ and 0 otherwise.

Remark. Let us explain the second statement (both what it means, and why it follows from the first). One mathematically precise version of the claim is that if one picks two integers $a, b \in \{1, \dots, N\}$ uniformly at random, then the probability they are coprime tends to $\frac{6}{\pi^2}$ as $N \rightarrow \infty$. But $\sum_{n \leq N} \phi(n)$ is exactly the number of pairs (a, b) of integers with $1 \leq a \leq b \leq N$ and a and b coprime.

Proof. Write F for the arithmetic function $F(n) = n$. We first note that $1 \star \phi = F$, that is to say $\sum_{d|n} \phi(n/d) = n$; this is by dividing up the integers $x \leq n$ according to the highest common factor $d = (n, x)$.

By Möbius inversion, it follows that $\phi = \mu \star F$, that is to say

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

Summing over $n \leq N$ and swapping the order of summation yields

$$\sum_{n \leq N} \phi(n) = \sum_{d \leq N} \frac{\mu(d)}{d} \sum_{\substack{n \leq N \\ d|n}} n = \sum_{d \leq N} \mu(d) \sum_{k \leq N/d} k.$$

Since

$$\sum_{k \leq N/d} k = \frac{1}{2} \lfloor \frac{N}{d} \rfloor (\lfloor \frac{N}{d} \rfloor + 1) = \frac{N^2}{2d^2} + O(\frac{N}{d})$$

(for $d \leq N$) we have

$$\sum_{n \leq N} \phi(n) = \frac{1}{2} N^2 \sum_{d \leq N} \frac{\mu(d)}{d^2} + O(N \sum_{d \leq N} \frac{1}{d}).$$

The error term here is $\ll N \log N$ which is acceptable. The error incurred by completing the sum $\sum_{d \leq N} \frac{\mu(d)}{d^2}$ to infinity is $\ll N^2 \sum_{d > N} \frac{1}{d^2} \ll N$ (see Sheet X). Thus it suffices to evaluate $\sum_d \frac{\mu(d)}{d^2} = D_\mu(2)$ (the Dirichlet series of μ at 2, see (2.2)).

However we have $\mu \star 1 = \delta$ and so, by Lemma 2.5, $D_\mu(2) D_1(2) = D_\delta(2) = 1$. It follows that

$$D_\mu(2) = D_1(2)^{-1} = \left(\sum_{n=1}^{\infty} n^{-2} \right)^{-1} = 6/\pi^2.$$

The result follows. □

3. THE RIEMANN ζ -FUNCTION

In this section we introduce the famous Riemann ζ -function (zeta function) and explore its basic properties as well as the connection to primes.

3.1. The ζ -function: basic definitions and facts. Let $s \in \mathbf{C}$ have $\operatorname{Re} s > 1$. Then we define

$$\zeta(s) = \sum_{n=1}^{\infty} n^{-s}. \quad (3.1)$$

This is the same thing as the Dirichlet series for the constant function 1, and so by [Lemma 2.4](#) it converges in $\operatorname{Re} s > 1$ and defines a holomorphic function there.

Euler product. Unique factorization into primes convinces us of the identity

$$\zeta(s) = 1 + \frac{1}{2^s} + \frac{1}{3^s} + \cdots = \left(1 + \frac{1}{2^s} + \frac{1}{2^{2s}} + \cdots\right) \left(1 + \frac{1}{3^s} + \frac{1}{3^{2s}} + \cdots\right) \cdots,$$

where there is one product for each prime $p = 2, 3, 5, \dots$. Summing each of the geometric series suggests then that

$$\zeta(s) = \prod_p (1 - p^{-s})^{-1}, \quad (3.2)$$

where the infinite product means $\lim_{N \rightarrow \infty} \prod_{p \leq N} (1 - p^{-s})^{-1}$. Let us pause to justify this formally. Let $s, \operatorname{Re} s > 1$, be fixed. Suppose $\operatorname{Re} s = 1 + \delta$. Fix an N , and suppose that $m > \log_2 N$. Then the product

$$P(m, N) := \prod_{p \leq N} \left(\sum_{j=0}^m p^{-js} \right)$$

may be expanded, and it equals $1 + 2^{-s} + \cdots + N^{-s}$ plus a number of further, distinct, terms n^{-s} with $n > N$. Thus

$$|P(m, N) - \zeta(s)| \leq \sum_{n=N+1}^{\infty} |n^{-s}| \ll N^{-\delta}$$

uniformly in m . Letting $m \rightarrow \infty$, we thus obtain

$$\left| \prod_{p \leq N} (1 - p^{-s})^{-1} - \zeta(s) \right| \ll N^{-\delta}.$$

Finally, letting $N \rightarrow \infty$, we confirm the equality [\(3.2\)](#).

Equation [\(3.2\)](#) is called the Euler product for ζ and it is valid for $\operatorname{Re} s > 1$.

An amusing consequence of the above is yet another proof (our third) that there are infinitely many primes. Indeed, suppose not. Then, letting $s \rightarrow 1^+$ in [\(3.2\)](#) we see that $\lim_{s \rightarrow 1^+} \zeta(s) < \infty$. However this is not the case, since

$$\lim_{s \rightarrow 1^+} \sum_{n=1}^N n^{-s} = \sum_{n=1}^N n^{-1},$$

and the harmonic series on the right diverges.

By Möbius inversion, we have $1 \star \mu = \delta$. By [Lemma 2.5](#), it follows that

$$1 = D_\delta(s) = D_{1 \star \mu}(s) = D_1(s)D_\mu(s) = \zeta(s) \left(\sum_{n=1}^{\infty} \mu(n)n^{-s} \right)$$

for $\operatorname{Re} s > 1$ (the required analytic condition [\(2.3\)](#) is trivial to verify in this case, with $a = 0$). An immediate consequence of this is the following nonvanishing result.

Corollary 3.1. *We have $\zeta(s) \neq 0$ when $\operatorname{Re} s > 1$.*

3.2. Analytic continuation. The real power of the ζ -function comes from considering $\zeta(s)$ for $\operatorname{Re} s \leq 1$. The definition [\(3.1\)](#) only makes sense for $\operatorname{Re} s > 1$, but we have the following remarkable fact.

Theorem 3.2. *There is a meromorphic³ continuation of $\zeta(s)$ to the whole complex plane $\mathbf{C}$, holomorphic except for a simple pole at $s = 1$.*

By the identity principle in complex analysis, this meromorphic continuation is unique. We will not prove this theorem in full in this course, but we will show how to meromorphically continue $\zeta(s)$ to the domain $\operatorname{Re} s > 0$.

Proposition 3.3. *The Riemann zeta function $\zeta(s)$ has a (unique) meromorphic continuation to the right half-plane $\operatorname{Re}(s) > 0$, holomorphic except for a simple pole at $s = 1$.*

Proof. For $\operatorname{Re} s > 1$ we have

$$\begin{aligned} \zeta(s) &= \sum_{n=1}^{\infty} n^{-s} = \sum_{n=1}^{\infty} n(n^{-s} - (n+1)^{-s}) \\ &= s \sum_{n=1}^{\infty} n \int_n^{n+1} x^{-s-1} dx \\ &= s \int_1^{\infty} [x] x^{-s-1} dx \\ &= \frac{s}{s-1} - s \int_1^{\infty} \{x\} x^{-s-1} dx. \end{aligned}$$

(Here, $\{x\} := x - [x]$.) The integral here defines a holomorphic function on $\operatorname{Re} s > 0$ by differentiating under the integral. The rigorous justification of this last step is not examinable in this course but may be found in [Proposition A.3](#). $\square$

With this in hand, we can state the famous Riemann hypothesis (RH).

Conjecture 3.4. *The only zeros of $\zeta(s)$ with $0 < \operatorname{Re} s < 1$ are those with $\operatorname{Re} s = \frac{1}{2}$.*

³You may wish to revisit A2.2 Complex Analysis when studying this section, though we will not need any detailed proofs from that course. Meromorphic means holomorphic except for a discrete set of isolated poles.

Only much weaker results are known unconditionally. Even the fact that $\zeta(s) \neq 0$ for $\operatorname{Re} s = 1$ is not trivial. This fact is essentially equivalent to the prime number theorem (1.3). The proof of the equivalence involves a lot of mathematics and in fact is most of the content of the course C3.8: Analytic Number Theory.

The Riemann Hypothesis gives much more precise information about the distribution of primes.

4. DIRICHLET CHARACTERS AND L -FUNCTIONS.

In this section we introduce Dirichlet characters and their Dirichlet series, which are called Dirichlet L -functions. We show how these may be used to prove the following celebrated theorem of Dirichlet.

Theorem 4.1. *Let $q \in \mathbf{N}$ and suppose that a is coprime to q . Then there are infinitely many primes p with $p \equiv a \pmod{q}$.*

4.1. Characters on finite abelian groups. An important object of study in this section will be *Dirichlet characters*. A Dirichlet character to modulus q is the lift to $\mathbf{Z}$ of a character on $(\mathbf{Z}/q\mathbf{Z})^*$, the multiplicative group of $\mathbf{Z}/q\mathbf{Z}$. Before saying exactly what this means, we give the basic theory of characters on finite abelian groups G .

Definition 4.2. Let G be a finite abelian group. Then a *character* on G is a homomorphism $\psi : G \rightarrow \mathbf{C}^* = \mathbf{C} \setminus \{0\}$.

Write e_G for the identity in G , and use addition for the group operation on G . If n is the order of G and $g \in G$ then $\psi(g)^n = \psi/ng) = \psi(e_G) = 1$, so ψ in fact takes values in the n th roots of unity. Here are the basic results on characters.

Proposition 4.3. *Let G be a finite abelian group. Then the characters on G form a group $\widehat{G}$ under pointwise multiplication of the same order as G (and isomorphic to it) with identity element ψ_0 , the trivial character which takes the value 1 everywhere. Moreover we have*

$$\sum_{g \in G} \psi(g) = 0 \tag{4.1}$$

unless $\psi = \psi_0$ and

$$\sum_{\psi \in \widehat{G}} \psi(g) = 0 \tag{4.2}$$

unless $g = e_G$.

Proof. (Sketch) For the proof we write G with the group operation being addition, which is standard when dealing with abstract abelian groups. Later, when we specialise to $G = (\mathbf{Z}/q\mathbf{Z})^*$, we will use multiplication.

That the characters form a group is an easy exercise in the definition. To show that $|\widehat{G}| = |G|$ (and to establish that $\widehat{\widehat{G}} \cong G$, though we will not actually need this fact) it suffices to handle the case $G = \mathbf{Z}/N\mathbf{Z}$ (with

the group operation being addition!), since every abelian group is a direct product of such groups (in fact with $N = p^j$). On this group, each character is completely determined by the value of $\psi(1)$, which must be an N th root of unity $e^{2\pi ir/N}$ (the whole character is then $\psi(x) = e^{2\pi irx/N}$) and we can see by inspection that $G \cong \widehat{G}$.

To prove (4.1), pick some g_0 such that $\psi(g_0) \neq 1$. Then, since $g + g_0$ ranges over G as g does,

$$\sum_{g \in G} \psi(g) = \sum_{g \in G} \psi(g + g_0) = \psi(g_0) \sum_{g \in G} \psi(g),$$

from which the result follows.

The proof of (4.2) is very similar. Pick some ψ_1 such that $\psi_1(g) \neq 1$. It is not obvious that such a ψ_1 exists. However, it is true when $G = \mathbf{Z}/N\mathbf{Z}$ (by inspection using the above description of characters on $\mathbf{Z}/N\mathbf{Z}$) and then it follows for arbitrary G by writing G as a direct product of such cyclic groups.

Since $\psi\psi_1$ ranges over $\widehat{G}$ as ψ does,

$$\sum_{\psi \in \widehat{G}} \psi(g) = \sum_{\psi \in \widehat{G}} (\psi\psi_1)(g) = \psi_1(g) \sum_{\psi \in \widehat{G}} \psi(g),$$

and the result follows. $\square$

Remark. For students taking B2.1 Representation Theory, this proposition is basically the abelian case of a large part of that course.

4.2. Dirichlet characters. We now specialise to $G = (\mathbf{Z}/q\mathbf{Z})^*$ which, as was shown in ASO Number Theory, is an abelian group of order $\phi(q)$. We write the group operation as multiplication (juxtaposition). Let $\psi : (\mathbf{Z}/q\mathbf{Z})^* \rightarrow \mathbf{C}^*$ be a character, thus $\psi(ab) = \psi(a)\psi(b)$ for all $a, b \in (\mathbf{Z}/q\mathbf{Z})^*$. Then we can ‘lift’ ψ to a function $\chi : \mathbf{Z} \rightarrow \mathbf{C}$ by defining

$$\chi(n) = \begin{cases} \psi(n \pmod{q}) & \text{if } (n, q) = 1; \\ 0 & \text{if } (n, q) \neq 1. \end{cases}$$

Such a function χ is called a *Dirichlet character* to modulus q .

Equivalently, a Dirichlet character to modulus q is a map $\chi : \mathbf{Z} \rightarrow \mathbf{C}$ such that

- $\chi(n + q) = \chi(n)$ for all n ;
- $\chi(n) = 0$ unless $(n, q) = 1$;
- $|\chi(n)| = 1$ if $(n, q) = 1$, and
- $\chi(ab) = \chi(a)\chi(b)$ whenever $(a, q) = (b, q) = 1$.

Note that $\chi(ab) = \chi(a)\chi(b)$ actually holds for all a, b .

Since the Dirichlet characters are in 1-1 correspondence with characters $\psi : (\mathbf{Z}/q\mathbf{Z})^* \rightarrow \mathbf{C}^*$, the number of Dirichlet characters to modulus q is $\phi(q)$.

Example. The constant function 1 is a Dirichlet character to modulus 1 (in fact the only one).

The function χ defined by $\chi(n) = 0$ if n is even and $\chi(n) = 1$ if n is odd is a Dirichlet character to modulus 2, the lift of the trivial character from $(\mathbf{Z}/2\mathbf{Z})^*$ (which is the group of order 1) to $\mathbf{C}^*$.

A slightly less trivial example is the Dirichlet character χ_4 defined by $\chi_4(n) = 0$ if n is even and $\chi_4(1) = 1$, $\chi_4(3) = -1$, and χ_4 periodic with period 4, which is the lift of the nontrivial character on $(\mathbf{Z}/4\mathbf{Z})^*$, a group of order 2.

Given a modulus q , there is one Dirichlet character χ_0 which has a special status, namely the *principal* character, which is the lift of the trivial character ψ_0 on $(\mathbf{Z}/q\mathbf{Z})^*$. Thus it is defined by

$$\chi_0(n) = \begin{cases} 1 & (n, q) = 1; \\ 0 & (n, q) \neq 1. \end{cases}$$

We record the following facts, which are immediate consequences of [Proposition 4.3](#) translated to this setting.

Lemma 4.4. *Let χ be a nonprincipal Dirichlet character to modulus q . Then*

$$\sum_{i=1}^q \chi(n+i) = 0 \tag{4.3}$$

for any integer n . We also have

$$\sum_{\chi} \chi(n) = \begin{cases} \phi(q) & \text{if } n \equiv 1 \pmod{q} \\ 0 & \text{otherwise,} \end{cases} \tag{4.4}$$

where the sum is over all the $\phi(q)$ Dirichlet characters to modulus q .

Proof. [\(4.3\)](#) follows from [\(4.1\)](#) and [\(4.4\)](#) follows from [\(4.2\)](#). $\square$

We can now define the Dirichlet L -function $L(s, \chi)$, which is a generalisation of the Riemann ζ -function.

Definition 4.5. Let χ be a Dirichlet character. Then for $\operatorname{Re} s > 1$ we define

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

Like the ζ -function, Dirichlet L -functions have an Euler product. This (and the proof) is a direct generalisation.

$$L(s, \chi) = \prod_p (1 - \chi(p) p^{-s})^{-1}. \tag{4.5}$$

As for the ζ -function, this is well-defined and holomorphic on $\operatorname{Re} s > 1$. For principal characters, $L(s, \chi_0)$ is in fact very closely related to the ζ -function.

Proposition 4.6. *Let χ_0 be the principal Dirichlet character to modulus q . Let $\operatorname{Re} s > 1$. Then*

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

Proof. This is immediate from the Euler product. $\square$

Since $\prod_{p|q} (1 - p^{-s})$ is an entire function which does not vanish at 1, it follows immediately from the preceding section that $L(s, \chi_0)$ extends to a meromorphic function on $\operatorname{Re} s > 0$ with a simple pole at $s = 1$.

Proposition 4.7. *Suppose that χ is a nonprincipal character to some modulus q . Then the series $\sum_{n=1}^{\infty} \chi(n)n^{-s}$ is (conditionally) convergent for $\operatorname{Re} s > 0$, and it defines a holomorphic function in this domain. We denote this $L(s, \chi)$.*

Proof. The key fact we will use is that

$$\left| \sum_{n \leq x} \chi(n) \right| \leq q \quad (4.6)$$

for all x , which follows immediately from the fact that χ sums to zero over any interval of length q (that is, (4.3)).

Now, using

$$N^{-s} - n^{-s} = s \int_N^n x^{-s-1} dx$$

we have we have

$$\sum_{N < n \leq N'} \chi(n)n^{-s} = N^{-s} \sum_{N < n \leq N'} \chi(n) - s \int_N^{N'} x^{-s-1} \left(\sum_{x < n \leq N'} \chi(n) \right) dx.$$

Therefore, using (4.6), if $\operatorname{Re} s = \sigma$ we have

$$\begin{aligned} \left| \sum_{N < n \leq N'} \chi(n)n^{-s} \right| &\leq qN^{-\sigma} + |s|q \int_N^{N'} x^{-\sigma-1} dx \\ &= qN^{-\sigma} + \frac{|s|q}{\sigma} (N^{-\sigma} - (N')^{-\sigma}) \ll q \left(1 + \frac{|s|}{\sigma}\right) N^{-\sigma}. \end{aligned}$$

It follows that $(\sum_{n \leq N} \chi(n)n^{-s})_N$ is uniformly Cauchy in any domain $\{s \in \mathbf{C} : \operatorname{Re} s > s_0, |s| \leq X\}$ with $s_0 > 0$ and finite X . This implies the result using standard facts from complex analysis. $\square$

Remarks. We caution that we have not shown the Euler product relation (4.5) holds other than for $\operatorname{Re} s > 1$. The above proof is an instance of ‘partial summation’ or ‘Abel summation’ but my personal preference is to proceed from first principles rather than quote a general result.

Suppose that χ is a nonprincipal Dirichlet character. The value $L(1, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n}$ (which is finite) is arithmetically significant. One famous evaluation is

$$L(1, \chi_4) = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots = \frac{\pi}{4}. \quad (4.7)$$

We have the following result, which we will not prove fully in lectures, though a complete proof will be given across lectures and the example sheets.

Proposition 4.8. *Let χ be a nonprincipal Dirichlet character. Then $L(1, \chi) \neq 0$.*

This result can be proved by hand in particular cases, as you are encouraged to do for characters of moduli 3 and 5 on the second example sheet.

4.3. Primes in arithmetic progressions. Dirichlet L -functions, and in particular their non-vanishing at 1, can be used to show the infinitude of primes in fixed arithmetic progressions.

Theorem 4.9. *Assume that $L(1, \chi) \neq 0$ for all nonprincipal χ to modulus q . Then, for any a coprime to q , we have*

$$\sum_{p \equiv a \pmod{q}} \frac{1}{p} = \infty.$$

In particular, there are infinitely many primes congruent to $a \pmod{q}$.

The first step in the proof is to express the condition $1_{n \equiv a \pmod{q}}$ in terms of characters.

Lemma 4.10. *Suppose that $(a, q) = 1$. Then we have*

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

Proof. If n is not coprime to q , both sides are zero. By abuse of notation, consider χ as a character on $(\mathbf{Z}/q\mathbf{Z})^*$. Then for $x \in (\mathbf{Z}/q\mathbf{Z})^*$ we have $\overline{\chi(a)}\chi(x) = \chi(xa^{-1})$. Now we use (4.4). $\square$

Let $s > 1$ be real. It follows using Lemma 4.10 that

$$\sum_{p \equiv a \pmod{q}} p^{-s} = \frac{1}{\phi(q)} \sum_{\chi} \overline{\chi(a)} \sum_{p: p \nmid q} \chi(p) p^{-s}.$$

If $\chi = \chi_0$, the principal character mod q , and if $p \nmid q$ then $\chi_0(p) = 1$ and the inner sum is $\sum_{p: p \nmid q} p^{-s}$. We have $\lim_{s \rightarrow 1+} \sum_p p^{-s} = \infty$ by the second Mertens' estimate.

It therefore suffices to prove the following.

Lemma 4.11. *Suppose that χ is a nonprincipal character mod q and that $L(1, \chi) \neq 0$. Then $\lim_{s \rightarrow 1+} \sum_p \chi(p) p^{-s}$ exists and is finite.*

We first sketch an incorrect ‘proof’ which nonetheless serves to convey the main idea. We ‘have’

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

and so

$$\log L(1, \chi) = - \sum_p \log(1 - \chi(p)p^{-1}) = \sum_p \frac{\chi(p)}{p} + \sum_p \frac{\chi(p)^2}{2p^2} + \dots$$

Since $L(1, \chi) \neq 0$, $\log L(1, \chi)$ is finite. Since all sums $\sum_p \frac{\chi(p)^2}{p^2}, \dots$ are finite, $\sum_p \frac{\chi(p)}{p}$ is finite.

There are two important issues with this attempted proof, namely

- We have not shown the validity of the Euler product expansion at $s = 1$, and
- Complex numbers do not have a unique logarithm; it is well-defined only up to integer multiples of $2\pi i$.

Another issue (more easily circumvented) is that there are infinitely many sums $\sum_p \frac{\chi(p)^j}{p^j}$, $j = 2, 3, \dots$, that we need to consider. We now give an actual proof, motivated by the above flawed sketch.

Proof. First note that, since $L(s, \chi)$ is holomorphic on $\operatorname{Re} s > 0$ and $L(1, \chi) \neq 0$, there is a holomorphic branch of logarithm $\log L(s, \chi)$ on the ball $B_\varepsilon(1)$. (See [Proposition B.2](#) for a reminder of this fact from complex analysis.)

Now for $s > 1$ we have the Euler product

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

Consider the function

$$F(s, \chi) := \sum_{j=1}^{\infty} \frac{1}{j} \sum_p \chi(p)^j p^{-js}.$$

The purpose of this definition is that $F(s, \chi)$ is a logarithm of $L(s, \chi)$ on $\operatorname{Re} s > 1$, due to the series expansion $\log(1 - z) = -\sum_{j=1}^{\infty} \frac{1}{j} z^j$ (see [Proposition B.3](#).) Since complex logarithms are unique up to integer multiples of $2\pi i$, it follows that

$$F(s, \chi) = \log L(s, \chi) + 2\pi i n(s)$$

on the intersection $\mathcal{D} := B_\varepsilon(1) \cap \{s \in \mathbf{C} : \operatorname{Re} s > 1\}$, where $n : \mathcal{D} \rightarrow \mathbf{Z}$ is an integer-valued function. Since both $F(s, \chi)$ and $\log L(s, \chi)$ are continuous on the connected domain $\mathcal{D}$, $n(s)$ must be constant on $\mathcal{D}$. Thus

$$F(s, \chi) = \log L(s, \chi) + C_\chi$$

for $s \in \mathcal{D}$, and so

$$\lim_{s \rightarrow 1^+} F(s, \chi) = \log L(1, \chi) + C_\chi$$

is finite. Now observe that

$$F(s, \chi) = \sum_p \chi(p) p^{-s} + F_*(s, \chi),$$

where

$$F_*(s, \chi) = \sum_{j=2}^{\infty} \frac{1}{j} \sum_p \chi(p)^j p^{-sj}.$$

To conclude the proof, it suffices to check that $\lim_{s \rightarrow 1^+} F_*(s, \chi)$ exists and is finite. This follows from dominated convergence and the bound

$$\sum_{j=2}^{\infty} \frac{1}{j} \sum_p p^{-j} < \sum_p \sum_{j=2}^{\infty} p^{-j} \leq 2 \sum_p p^{-2} < 4.$$

(This final bound is exactly the same as the one in [Proposition 2.8](#).) □