NOTES 1: ELEMENTARY MULTIPLICATIVE NUMBER THEORY

TERENCE TAO

In analytic number theory, an *arithmetic function* is simply a function $f : \mathbf{N} \to \mathbf{C}$ from the natural numbers $\mathbf{N} = \{1, 2, 3, \dots\}$ to the real or complex numbers. (One occasionally also considers arithmetic functions taking values in more general rings than $\mathbf{R}$ or $\mathbf{C}$, as in *this previous blog post*, but we will restrict attention here to the classical situation of real or complex arithmetic functions.) Experience has shown that a particularly tractable and relevant class of arithmetic functions for analytic number theory are the *multiplicative functions*, which are arithmetic functions $f : \mathbf{N} \to \mathbf{C}$ with the additional property that

$$f(nm) = f(n)f(m) \tag{1}$$

whenever $n, m \in \mathbf{N}$ are coprime. (One also considers arithmetic functions, such as the logarithm function $L(n) := \log n$ or the *von Mangoldt function*, that are not genuinely multiplicative, but interact closely with multiplicative functions, and can be viewed as "derived" versions of multiplicative functions; *see this previous post.*) A typical example of a multiplicative function is the *divisor function*

$$\tau(n) := \sum_{d|n} 1 \tag{2}$$

that counts the number of divisors of a natural number n. (The divisor function $n \mapsto \tau(n)$ is also denoted $n \mapsto d(n)$ in the literature.) The study of asymptotic behaviour of multiplicative functions (and their relatives) is known as *multiplicative number theory*, and is a basic cornerstone of modern analytic number theory.

There are various approaches to multiplicative number theory, each of which focuses on different asymptotic statistics of arithmetic functions f. In *elementary multiplicative number theory*, which is the focus of this set of notes, particular emphasis is given on the following two statistics of a given arithmetic function $f : \mathbf{N} \to \mathbf{C}$:

1. The *summatory functions*

$$\sum_{n \leq x} f(n)$$

of an arithmetic function f, as well as the associated natural density

$$\lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x} f(n)$$

(if it exists).

2. The *logarithmic sums*

$$\sum_{n\leq x} \frac{f(n)}{n}$$

of an arithmetic function f, as well as the associated *logarithmic density*

$$\lim_{x\to\infty} \frac{1}{\log x} \sum_{n\leq x} \frac{f(n)}{n}$$

(if it exists).

Here, we are normalising the arithmetic function f being studied to be of roughly unit size up to logarithms, obeying bounds such as $f(n) = O(1)$, $f(n) = O(\log^{O(1)} n)$, or at worst

$$f(n) = O(n^{o(1)}). \tag{3}$$

A classical case of interest is when f is an indicator function $f = 1_A$ of some set A of natural numbers, in which case we also refer to the natural or logarithmic density of f as the natural or logarithmic density of A respectively. However, in analytic number theory it is usually more convenient to replace such indicator functions with other related functions that have better multiplicative properties. For instance, the indicator function $1_{\mathcal{P}}$ of the primes is often replaced with the von Mangoldt function Λ.

Typically, the logarithmic sums are relatively easy to control, but the summatory functions require more effort in order to obtain satisfactory estimates; see Exercise 7 below.

If an arithmetic function f is multiplicative (or closely related to a multiplicative function), then there is an important further statistic on an arithmetic function f beyond the summatory function and the logarithmic sum, namely the *Dirichlet series*

$$\mathcal{D}f(s) := \sum_{n=1}^{\infty} \frac{f(n)}{n^s} \tag{4}$$

for various real or complex numbers s. Under the hypothesis (3), this series is absolutely convergent for real numbers $s > 1$, or more generally for complex numbers s with $\mathrm{Re}(s) > 1$. As we will see below the fold, when f is multiplicative then the Dirichlet series enjoys an important *Euler product factorisation* which has many consequences for analytic number theory.

In the elementary approach to multiplicative number theory presented in this set of notes, we consider Dirichlet series only for real numbers $s > 1$ (and focusing particularly on the asymptotic behaviour as $s \to 1^+$); in later notes we will focus instead on the important *complex-analytic approach* to multiplicative number theory, in which the Dirichlet series (4) play a central role, and are defined not only for complex numbers with large real part, but are often extended analytically or meromorphically to the rest of the complex plane as well.

Remark 1. *The elementary and complex-analytic approaches to multiplicative number theory are the two classical approaches to the subject. One could also consider a*

more "Fourier-analytic" approach, in which one studies convolution-type statistics such as

$$\sum_n \frac{f(n)}{n} G(t - \log n) \tag{5}$$

as $t \to \infty$ for various cutoff functions $G : \mathbf{R} \to \mathbf{C}$, such as smooth, compactly supported functions. See this previous blog post *for an instance of such an approach. Another related approach is the "pretentious" approach to multiplicative number theory currently being developed by Granville-Soundararajan and their collaborators. We will occasionally make reference to these more modern approaches in these notes, but will primarily focus on the classical approaches.*

To reverse the process and derive control on summatory functions or logarithmic sums starting from control of Dirichlet series is trickier, and usually requires one to allow s to be complex-valued rather than real-valued if one wants to obtain really accurate estimates; we will return to this point in subsequent notes. However, there is a cheap way to get *upper bounds* on such sums, known as *Rankin's trick*, which we will discuss later in these notes.

The basic strategy of elementary multiplicative theory is to first gather useful estimates on the statistics of "smooth" or "non-oscillatory" functions, such as the constant function $n \mapsto 1$, the harmonic function $n \mapsto \frac{1}{n}$, or the logarithm function $n \mapsto \log n$; one also considers the statistics of periodic functions such as *Dirichlet characters*. These functions can be understood without any multiplicative number theory, using basic tools from real analysis such as the (quantitative version of the) *integral test* or *summation by parts*. Once one understands the statistics of these basic functions, one can then move on to statistics of more arithmetically interesting functions, such as the divisor function (2) or the *von Mangoldt function* Λ that we will discuss below. A key tool to relate these functions to each other is that of *Dirichlet convolution*, which is an operation that interacts well with summatory functions, logarithmic sums, and particularly well with Dirichlet series.

This is only an introduction to elementary multiplicative number theory techniques. More in-depth treatments may be found in *this text of Montgomery-Vaughan*, or *this text of Bateman-Diamond*.

1. Summing monotone functions

The most fundamental estimate regarding the equidistribution of the natural numbers is the trivial bound

$$\sum_{n \leq x} 1 = x + O(1) \tag{6}$$

for any $x > 0$, which reflects the evenly spaced nature of the natural numbers. One also has the variant

$$\sum_{n \leq x} 1 \leq x, \tag{7}$$

also valid for any $x > 0$. But note that if the summation is not over the natural numbers, but is also allowed to contain $n = 0$, then the sum

$$\sum_{0 \leq n \leq x} 1 \quad \text{or} \quad \sum_{\{n \in \mathbf{Z}: |n| \leq x\}} 1$$

is no longer $O(x)$ when x is small, and one should instead revert to (6).

We have the following generalisation of (6) to summation of monotone functions:

Lemma 2. *(Quantitative integral test) Let $y < x$ be real numbers, and let $f : [y, x] \to \mathbf{R}$ be a monotone function. Then*

$$\sum_{\{n \in \mathbf{Z}: y \leq n \leq x\}} f(n) = \int_y^x f(t)\ dt + O(|f(x)| + |f(y)|).$$

Note that monotone functions are automatically Riemann integrable and Lebesgue integrable, so there is no difficulty in defining the integral appearing above.

Proof. By replacing f with $-f$ if necessary, we may assume that f is non-decreasing. By rounding up y and rounding down x, we may assume that x, y are integers. We have

$$\begin{aligned} \sum_{\{n \in \mathbf{Z}: y \leq n \leq x\}} f(n) &= \int_y^x f(\lfloor t \rfloor)\ dt + f(x) \\ &\leq \int_y^x f(t)\ dt + f(x) \end{aligned}$$

and similarly

$$\begin{aligned} \sum_{\{n \in \mathbf{Z}: y \leq n \leq x\}} f(n) &= \int_y^x f(\lceil t \rceil)\ dt + f(y) \\ &\geq \int_y^x f(t)\ dt + f(y) \end{aligned}$$

and the claim follows. □

Thus, for instance, one has

$$\sum_{n \leq x} \log n = x \log x - x + O(\log(2 + x)) \tag{8}$$

for any $x > 0$ (a weak form of *Stirling's formula*, discussed in *this previous blog post*), and more generally one has

$$\sum_{n \leq x} \log^k n = x P_k(\log x) + O_k(\log^k(2 + x)) \tag{9}$$

for all $x > 0$ and some polynomial $P_k(t)$ with leading term t^k. (The remaining terms in $P_k(t)$ may be computed explicitly, but for our purposes it will not be essential to know what they are.)

Remark 3. *If x, y are not required to be integers, then one cannot improve substantially on the size of the error term $O(f(x)+f(y))$, as one can see by considering what happens if x or y transitions from being infinitesimally smaller than an integer to infinitesimally larger than that integer. But if x, y are integer and one assumes more differentiability on f, one can get more precise control on the error term in Lemma 2 using* the Euler-Maclaurin formula*; see e.g. Exercise 11 below. However, we will usually not need these more refined estimates here. In any event, one can get even better control on the error term if one works with smoother sums such as (5) with G smooth, thanks to tools such as the* Poisson summation formula*. See* this previous blog post *for some related discussion.*

In the converse direction, if f is highly oscillatory then there is usually no simple relationship between the sum $\sum_{\{n\in\mathbf{Z}:y\leq n\leq x\}} f(n)$ and the integral $\int_y^x f(t)\,dt$; consider for instance the example $f(t) := \cos(2\pi nt)$, in which the sum grows linearly in $x-y$ but the integral stays bounded.

Exercise 4. *For non-negative natural numbers $k, l \geq 0$, show that*

$$\sum_{n\leq x} \log^k(n) \log^l\left(\frac{x}{n}\right) = xP_{k,l}(\log x) + O_{k,l}(\log^{k+l}(2+x)) \tag{10}$$

for all $x \geq 0$ and some polynomial $P_{k,l}(t)$ with leading term $l!t^k$.

Lemma 2 combines well with the following basic lemma.

Lemma 5. *(Cauchy sequences converge) Let $f : \mathbf{N} \to \mathbf{C}$, $F : \mathbf{R}^+ \to \mathbf{C}$ and $g : \mathbf{R}^+ \to \mathbf{R}^+$ be functions such that $g(x) \to 0$ as $x \to \infty$. Then the following are equivalent:*

(i) *One has*

$$\sum_{y\leq n<x} f(n) = F(x) - F(y) + O(g(x) + g(y))$$

for all $1 \leq y < x$.

(ii) *There exists a constant $c \in \mathbf{C}$ such that*

$$\sum_{n<x} f(n) = c + F(x) + O(g(x))$$

for all $x \geq 1$. In particular, $c = -F(1) + O(g(1))$.

The quantity c in (ii) is unique; it is also real-valued if f, F are real-valued.

If in addition $F(x) \to 0$ as $x \to \infty$, then when (ii) holds, $\sum_{n=1}^\infty f(n)$ converges conditionally to c.

Exercise 6. *Prove Lemma 5.*

We now give some basic applications of these lemmas. If $s > 0$ is a real number not equal to 1, then from Lemma 2 we have

$$\sum_{y\leq n\leq x} \frac{1}{n^s} = \frac{y^{1-s}}{s-1} - \frac{x^{1-s}}{s-1} + O\left(\frac{1}{y^s}\right)$$

for $1 \leq y < x$, and hence by Lemma 5, there is a real number $\zeta(s)$ with

$$\zeta(s) = \frac{1}{s-1} + O(1) \tag{11}$$

such that

$$\sum_{n \leq x} \frac{1}{n^s} = \zeta(s) - \frac{x^{1-s}}{s-1} + O\left(\frac{1}{x^s}\right) \tag{12}$$

for all $x \geq 1$. In the case $s > 1$, we conclude that the sum $\sum_n \frac{1}{n^s}$ is absolutely convergent, and

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

however for $s < 1$ this identity is technically not true, since the sum on the right-hand side is now divergent using the usual conventions for infinite sums. (The identity can however be recovered by using more general interpretations of infinite sums; see *this previous blog post*). The function ζ is of course the famous *Riemann zeta function*.

For the $s = 1$ case, we again see from Lemma 2 that

$$\sum_{y \leq n \leq x} \frac{1}{n} = \log x - \log y + O\left(\frac{1}{y}\right)$$

for $1 \leq y \leq x$, and thus there exists a real number γ such that

$$\sum_{n \leq x} \frac{1}{n} = \log x + \gamma + O\left(\frac{1}{x}\right) \tag{13}$$

for $x \geq 1$. The constant γ is known as *Euler's constant* (or the *Euler-Mascheroni constant*), and has a value of $\gamma = 0.57721566\ldots$.

Exercise 7. *Let f be an arithmetic function. If the natural density*

$$\lim_{x \to \infty} \frac{1}{x} \sum_{n \leq x} f(n)$$

of f exists and is equal to some complex number α, show that the logarithmic density

$$\lim_{x \to \infty} \frac{1}{\log x} \sum_{n \leq x} \frac{f(n)}{n}$$

also exists and is equal to α. Hint: first establish the identity

$$\sum_{n \leq x} \frac{f(n)}{n} = \frac{1}{x} \sum_{n \leq x} f(n) + \int_1^x \frac{1}{t} \sum_{n \leq t} f(n) \frac{dt}{t}$$

An important counterexample to the converse claim is given in Exercise 11 below.

Exercise 8. *Let f be an arithmetic function obeying the crude bound (3), and let α be a complex number. If the logarithmic density of f exists and is equal to α, show that $(s-1)\mathcal{D}f(s) \to \alpha$ as $s \to 1^+$, or in other words that*

$$\mathcal{D}f(s) = \frac{\alpha}{s-1} + o\left(\frac{1}{s-1}\right)$$

as $s \to 1^+$. Hint: $\frac{1}{n^s} = \frac{1}{n}\int_n^\infty \frac{s-1}{x^s}\ dx$.

Exercise 9.

(i) *For any integer $k \geq 0$, show that*

$$\sum_{n \leq x} \frac{\log^k n}{n} = Q_k(\log x) + O_k\left(\frac{\log^k(2+x)}{x}\right) \tag{14}$$

for all $x \geq 1$, where $Q_k(t)$ is a polynomial with leading term $\frac{1}{k+1}t^{k+1}$ (in fact $Q_k(t)$ is exactly equal to $\frac{1}{k+1}t^{k+1}$!).

(ii) *More generally, for any integers $k, l \geq 0$, show that*

$$\sum_{n \leq x} \frac{\log^k(n)\log^l\left(\frac{x}{n}\right)}{n} = Q_{k,l}(\log x) + O_{k,l}\left(\frac{\log^{k+l}(2+x)}{x}\right) \tag{15}$$

for all $x \geq 1$, where $Q_{k,l}(t)$ is a polynomial with leading term $\frac{k!l!}{(k+l+1)!}t^{k+l+1}$ (again, $Q_{k,l}(t)$ is in fact exactly equal to $\frac{k!l!}{(k+l+1)!}t^{k+l+1}$).

Exercise 10. *Show rigorously that for any non-negative integer $k \geq 0$ and real $s > 1$, the Riemann zeta function ζ is k-times differentiable at s and one has*

$$\begin{aligned} \zeta^{(k)}(s) &= (-1)^k \sum_{n=1}^\infty \frac{\log^k n}{n^s} \\ &= (-1)^k \frac{k!}{(s-1)^{k+1}} + O_k(1). \end{aligned} \tag{16}$$

(There are several ways to justify the term-by-term differentiation in the first equation; one way is to instead establish a term-by-term integration formula and then apply the fundamental theorem of calculus. Another is to use Taylor series with remainder to control the error in Newton quotients. A third approach is to use complex analysis. You may find Exercise 11 below to be useful for some of these approaches.)

Exercise 11. *Let $y < x$ be real numbers, and let $f : [y, x] \to \mathbf{C}$ be a continuously differentiable function. Show that*

$$\sum_{\{n \in \mathbf{Z} : y \leq n \leq x\}} f(n) = \int_y^x f(t)\ dt + O\left(\int_y^x |f'(t)|\ dt + |f(y)|\right).$$

(Hint: compare $f(n)$ with $\int_n^{n+1} f(t)\ dt$. One may first wish to consider the case when x, y are integers, and deal with the roundoff errors when this is not the case later.) Conclude that if t is a non-zero number, that the function $n \mapsto n^{it}$ has

logarithmic density zero, but does not have a natural density. Hint: for the latter claim, argue by contradiction and consider sums of the form $\sum_{x \leq n \leq (1+\varepsilon)x} n^{it}$.

Remark 12. *The above exercise demonstrates that logarithmic density is a cruder statistic than natural density, as it completely ignores oscillations by* n^{it}*, whereas natural density is very sensitive to such oscillations. As such, controlling natural density of some arithmetic functions (such as the von Mangoldt function* Λ*) often boils down to determining to what extent such functions "conspire", "correlate", or "pretend to be"* n^{it} *for various real numbers* t*. This fact is a little tricky to discern in the elementary approach to multiplicative number theory, but is more readily apparent in the complex analytic approach, which we will discuss in later notes.*

2. The Euler Product and Rankin's Trick

The *fundamental theorem of arithmetic* tells us that every natural number n is uniquely expressible as a prime factorisation

$$n = p_1^{a_1} \dots p_k^{a_k}. \tag{17}$$

A very useful way to encode this fact analytically (in a "*generating function*" form) is as follows. Recall from the introduction that an arithmetic function $f : \mathbf{N} \to \mathbf{C}$ is *multiplicative* if one has $f(1) = 1$ and $f(nm) = f(n)f(m)$ whenever n, m are coprime. If f is a multiplicative function, then we have

$$f(n) = f(p_1^{a_1}) f(p_2^{a_2}) \dots f(p_k^{a_k}),$$

and hence

$$\sum_n f(n) = \prod_p (1 + f(p) + f(p^2) + f(p^3) + \dots) \tag{18}$$

at least when f is non-negative. If f is complex-valued and the product

$$\prod_p (1 + |f(p)| + |f(p^2)| + |f(p^3)| + \dots)$$

is finite, then an application of dominated convergence shows that (18) holds in this case also, with both sides of the equation being absolutely convergent. Multiplying f by the multiplicative function $\frac{1}{n^s}$, we conclude in particular the *Euler product identity*

$$\mathcal{D}f(s) = \prod_p \left(1 + \frac{f(p)}{p^s} + \frac{f(p^2)}{p^{2s}} + \frac{f(p^3)}{p^{3s}} + \dots\right) \tag{19}$$

whenever the right-hand side is absolutely convergent in the sense that

$$\prod_p \left(1 + \frac{|f(p)|}{|p^s|} + \frac{|f(p^2)|}{|p^{2s}|} + \frac{|f(p^3)|}{|p^{3s}|}| + \dots\right)$$

is finite, or equivalently that

$$\sum_p \frac{|f(p)|}{|p^s|} + \frac{|f(p^2)|}{|p^{2s}|} + \frac{|f(p^3)|}{|p^{3s}|} + \dots$$

is finite. Observe that if f obeys the crude bound (3), then this absolute convergence is obtained whenever $\mathrm{Re}(s) > 1$. Thus for instance one has the Euler product identity for the Riemann zeta function $\zeta(s)$,

$$\begin{aligned}\zeta(s) &= \prod_p (1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \dots) \\ &= \prod_p \left(1 - \frac{1}{p^s}\right)^{-1},\end{aligned} \tag{20}$$

whenever $s > 1$. From (11) we therefore have

$$\prod_p \left(1 - \frac{1}{p^s}\right)^{-1} = \zeta(s) = \frac{1}{s-1} + O(1) \tag{21}$$

Taking logarithms, we thus have

$$-\sum_p \log\left(1 - \frac{1}{p^s}\right) = \log \zeta(s) = \log \frac{1}{s-1} + O(s-1) \tag{22}$$

when $s > 1$ (note that $\zeta(s)$ is clearly at least one). The logarithm here on the left-hand side is not so desirable, but we may remove it by using the Taylor expansion

$$-\log\left(1 - \frac{1}{p^s}\right) = \frac{1}{p^s} + O\left(\frac{1}{p^{2s}}\right)$$

and noting that $\sum_p \frac{1}{p^{2s}} \leq \sum_n \frac{1}{n^2}$ is absolutely convergent, to conclude that

$$\sum_p \frac{1}{p^s} = \log \frac{1}{s-1} + O(1) \tag{23}$$

when $s > 1$ is bounded. Taking $s \to 1$, and using monotone convergence, we recover Euler's theorem $\sum_p \frac{1}{p} = +\infty$, but (23) gives more precise information. For instance, we have the following bound, already anticipated by Euler:

Theorem 13. *(Cheap second Mertens' theorem) We have*

$$\sum_{p \leq x} \frac{1}{p} \ll \log\log x$$

for $x \geq 10$.

We will improve this bound later in these notes.

Proof. We use a device known as *Rankin's trick* to compare a Dirichlet series with a natural or logarithmic mean. Namely, observe that

$$\frac{1}{n} 1_{n \leq x} \ll \frac{1}{n^{1+1/\log x}} \tag{24}$$

for all $x > 1$ and all natural numbers n (in fact the implied constant can be given to be e), and hence

$$\sum_{p \leq x} \frac{1}{p} \ll \sum_p \frac{1}{p^{1+1/\log x}}.$$

The claim now follows from (23). □

Remark 14. *The same Rankin trick (24), when used to bound the harmonic series* $\sum_{n\leq x}\frac{1}{n}$*, gives an upper bound of* $e\log x+O(1)$*, which is inferior to (13) by a factor of about* e*. In general, one expects Rankin's trick to lose a constant factor when estimating non-negative logarithmic sums.*

Rankin's trick can also be used to cheaply bound other means of multiplicative functions, as follows.

Theorem 15. *(Cheap upper bound on multiplicative functions) Let* $k\geq 0$ *be a real number, and let* f *be a multiplicative function such that*

$$|f(p)|\leq k+O_k\left(\frac{1}{p}\right) \tag{25}$$

for all primes p*, and*

$$|f(p^j)|\ll_k j^{O_k(1)}$$

for all primes p *and* $j\geq 1$*. Then we have*

$$\sum_{n\leq x}\frac{|f(n)|}{n}\ll_k \log^k x$$

for $x\geq 2$*.*

Proof. For brevity we allow all implied constants here to depend on k. By replacing f by $|f|$, we may assume that f is non-negative. From Rankin's trick (24) we have

$$\sum_{n\leq x}\frac{f(n)}{n}\ll\sum_{n=1}^{\infty}\frac{f(n)}{n^{1+1/\log x}}.$$

Using the Euler product (19), we conclude that

$$\sum_{n\leq x}\frac{f(n)}{n}\ll\prod_p\left(1+\frac{f(p)}{p^{1+1/\log x}}+\frac{f(p^2)}{p^{2+2/\log x}}+\dots\right).$$

We can crudely bound

$$\sum_{j=2}^{\infty}\frac{f(p^j)}{p^{j+j/\log x}}\ll\sum_{j=2}^{\infty}\frac{j^{O(1)}}{p^j}\ll\frac{1}{p^2}$$

and hence by (25)

$$\sum_{n\leq x}\frac{f(n)}{n}\ll\prod_p\left(1+\frac{k}{p^{1+1/\log x}}+O\left(\frac{1}{p^2}\right)\right).$$

By Taylor expansion we have

$$\left(1+\frac{k}{p^{1+1/\log x}}+O\left(\frac{1}{p^2}\right)\right)\leq\exp\left(O\left(\frac{1}{p^2}\right)\right)\left(1-\frac{1}{p^{1+1/\log x}}\right)^{-k}$$

and hence by (21)

$$\sum_{n \leq x} \frac{f(n)}{n} \ll \log^k x \exp\left(\sum_p O\left(\frac{1}{p^2}\right)\right).$$

But $\sum_p \frac{1}{p^2} \leq \sum_n \frac{1}{n^2} \ll 1$, and the claim follows. □

Comparing this bound with (14), we are led to the guess that for f as in the above theorem, $f(n)$ should behave roughly like $\log^{k-1} n$ on the average. In the next section we will see some arguments that can make this more precise.

We can get more mileage out of (22) by differentiating in s. Formally differentiating in s, we arrive at the identity

$$\sum_p \frac{\log p}{p^s - 1} = -\frac{\zeta'(s)}{\zeta(s)} \tag{26}$$

for all $s > 1$.

Exercise 16. *Derive (26) rigorously. (Hint: For fixed $s > 1$ and small $\varepsilon > 0$, expand $\log(1 - \frac{1}{p^{s+\varepsilon}})$ using Taylor series with remainder.)*

Using the geometric series expansion

$$\frac{\log p}{p^s - 1} = \frac{\log p}{p^s} + \frac{\log p}{p^{2s}} + \frac{\log p}{p^{3s}} + \dots$$

and introducing the *von Mangoldt function* $\Lambda : \mathbf{N} \to \mathbf{R}$, defined by setting $\Lambda(p^j) := \log p$ whenever p is a prime and j is a natural number, and $\Lambda(n) = 0$ for all other n, we thus obtain the important identity

$$\sum_n \frac{\Lambda(n)}{n^s} = -\frac{\zeta'(s)}{\zeta(s)} \tag{27}$$

for the Dirichlet series of Λ, and for all $s > 1$. (In fact this identity will hold for a wider range of s, but we will address this in later notes.) For comparison, a Taylor expansion of (22) gives the closely related identity

$$\sum_n \frac{\Lambda(n)}{n^s \log n} = \log \zeta(s). \tag{28}$$

Indeed, (27) is essentially the derivative of (28) in s.

From (11), (16), (27) we have

$$\sum_n \frac{\Lambda(n)}{n^s} = \frac{1}{s-1} + O(1) \tag{29}$$

for $s > 1$ (note the claim is trivial if $s - 1$ is large).

Exercise 17. *(Cheap first Mertens' theorem) Show that*

$$\sum_{p \leq x} \frac{\log p}{p} \leq \sum_{n \leq x} \frac{\Lambda(n)}{n} \ll \log x$$

for all $x \geq 2$. (Hint: use Rankin's trick.)

One can try to use Rankin's trick to bound summatory functions by using the very crude bound

$$1_{n\leq x} \leq \frac{x}{n}, \tag{30}$$

but this is wasteful (often losing a factor of about $\log x$). For instance, for k, f as in Theorem 15, one could bound the summatory function $\sum_{n\leq x} f(n)$ by using

$$|\sum_{n\leq x} f(n)| \leq \sum_{n\leq x} \frac{x}{n}|f(n)|$$

and then using that Theorem, but this only gives the crude upper bound

$$\sum_{n\leq x} f(n) \ll_k x \log^k x$$

for $x \geq 2$, which turns out to be off from the truth by a factor of $\log x$ (as can be seen for instance in the simple example $k = 0$, $f(n) = 1$). The bound (30) also only gives the trivial upper bound of $O(x \log x)$ for $\sum_{n\leq x} \Lambda(n) \ll x \log x$.

The problem is that (30) is very inefficient when n is much smaller than x. To do better, there is a standard rearrangement trick that moves much of the "mass" of a summatory function $\sum_{n\leq x} f(n)$ to smaller values of n, effectively increasing the cutoff $1_{n\leq x}$ and so replacing (30) with a less inefficient comparison. To describe this rearrangement trick, we return to the fundamental theorem of arithmetic (17) and take logarithms to obtain

$$\log n = a_1 \log p_1 + \cdots + a_k \log p_k$$

whenever n is a natural number with prime factorisation (17). Using the von Mangoldt function defined previously, we can thus encode the fundamental theorem of arithmetic neatly in the important identity

$$\log n = \sum_{d|n} \Lambda(d). \tag{31}$$

For $n \leq x$, we rearrange this identity as

$$1 = \frac{1}{\log x} \left(\log \frac{x}{n} + \sum_{d|n} \Lambda(d) \right)$$

which allows us to rewrite an arbitrary summatory function $\sum_{n\leq x} f(n)$ for some $x > 1$ as

$$\sum_{n\leq x} f(n) = \frac{1}{\log x} \left(\sum_{n\leq x} f(n) \log \frac{x}{n} + \sum_{n\leq x} f(n) \left(\sum_{d|n} \Lambda(d) \right) \right).$$

For the latter sum, we write $n = dm$ and interchange summations to obtain (after replacing m with n) the rearrangement identity

$$\sum_{n\leq x} f(n) = \frac{1}{\log x} \sum_{n\leq x} \left(f(n) \log \frac{x}{n} + \sum_{d\leq x/n} f(dn)\Lambda(d) \right). \tag{32}$$

The expression in parentheses can be viewed as a weighted version of $f(n)$, with the weight tending to be larger for small n than for large n. Because of this reweighting, if one applies Rankin's trick to bound the sum on the right-hand side, one can often obtain an upper bound on $\sum_{n\leq x} f(n)$ that recovers the loss of $\log x$ that would occur if one applied Rankin's trick directly.

To use (32) effectively, we need the following basic upper bound on the von Mangoldt function, which unfortunately does not seem to be provable by the techniques outlined above, but can be established fairly quickly by alternate means:

Proposition 18. *(Chebyshev upper bound) We have*

$$\sum_{n\leq x} \Lambda(n) \ll x$$

for all $x \geq 0$. In particular, specialising n to primes we have

$$\sum_{p\leq x} \log p \ll x.$$

Proof. We use the following slick (but perhaps somewhat unmotivated) argument. By telescoping series it suffices to establish the bound

$$\sum_{N<n\leq 2N} \Lambda(n) \ll N$$

for all natural numbers N. We first consider the contribution of those n that are powers p^j of a prime p for some $j \geq 2$. One has $p \ll \sqrt{N}$, so there are at most $O(\sqrt{N})$ such primes, and each prime contributes at most $O(\log 2N)$ to the above sum; since $\sqrt{N}\log 2N = O(N)$, we may discard this contribution and reduce to showing that

$$\sum_{N<p\leq 2N} \log p \ll N.$$

We will achieve this through an inspection of the binomial coefficient

$$\binom{2N}{N} = \frac{(2N)!}{(N!)^2}.$$

Observe that if p is a prime with $N < p \leq 2N$, then p divides $(2N)!$ but not $(N!)^2$, and so divides $\binom{2N}{N}$. Taking logarithms, we conclude that

$$\sum_{N<p\leq 2N} \log p \leq \log \binom{2N}{N}.$$

From the binomial theorem we have $\binom{2N}{N} \leq 2^{2N}$, and the claim follows. □

One can also prove this proposition by several other means, for instance through Möbius inversion; see Exercise 60 below. Unfortunately, the rearrangement identity (32) does not help directly with establishing this proposition, as Λ is supported on numbers with too few prime factors for the rearrangement to have any non-trivial effect.

Remark 19. *Chebyshev also established the matching lower bound* $\sum_{n \leq x} \Lambda(n) \gg x$ *for* $x \geq 2$*; we will establish this bound (by a slightly different method) in Exercise 35. Both of Chebyshev's bounds will later be superseded by the* prime number theorem $\sum_{n \leq x} \Lambda(n) = (1 + o(1))x$*, discussed in later notes.*

Exercise 20. *Show that the number of primes up to* x *is* $O(\frac{x}{\log x})$ *for any* $x \geq 2$*.*

We now give an illustration of (32):

Proposition 21. *(Cheap upper bound for summatory functions) With* k, f *as in Theorem 15, we have*

$$\sum_{n \leq x} f(n) \ll_k x \log^{k-1} x$$

for $x \geq 2$*.*

Proof. By replacing f by $|f|$, we may assume f to be non-negative. By (32), it suffices to establish the bounds

$$\sum_{n \leq x} f(n) \log\left(\frac{x}{n}\right) \ll_k x \log^k x$$

and

$$\sum_{n \leq x} \sum_{d \leq x/n} f(dn)\Lambda(d) \ll_k x \log^k x. \tag{33}$$

The first estimate follows directly from Theorem 15 after using the inequality $\log \frac{x}{n} \leq \frac{x}{n}$, so we turn to the latter. We first consider the contribution when d, n are coprime. In that case, we may factor $f(dn)$ as $f(d)f(n)$, so the contribution of this case to (33) is bounded by

$$\sum_{n \leq x} f(n) \sum_{d \leq x/n} f(d)\Lambda(d).$$

To control the sum $\sum_{d \leq x/n} f(d)\Lambda(d)$, we observe that each prime $\sqrt{x/n} < p \leq x/n$ contributes at most $O_k(\log p)$ to this sum, while the primes $p \leq \sqrt{x/n}$ contribute $O(\log^{O_k(1)} x/n)$ through all the powers of p less than x/n. By Proposition 18, we thus have

$$\sum_{d \leq x/n} f(d)\Lambda(d) \ll_k \frac{x}{n} + \sqrt{\frac{x}{n}} \log^{O_k(1)} \frac{x}{n} \ll_k \frac{x}{n}$$

and so the contribution of the coprime case is acceptable by Theorem 15.

It remains to deal with the case when d, n are not coprime; thus $d = p^j$ and $n = p^l m$ for some prime p, some $j, l \geq 1$, and some $m \leq x/p^{j+l}$. We can then bound this contribution to (33) by

$$\sum_p \sum_{j,l \geq 1} \sum_{m \leq x/p^{j+l}} f(p^{j+l}m) \log p.$$

We have $f(p^{j+l}m) \ll_k (j+l)^{O_k(1)} f(m)$, so by Theorem 15 we may bound this expression by

$$\ll_k \sum_p \log p \sum_{j,l \geq 1} (j+l)^{O_k(1)} \frac{x}{p^{j+l}} \log^k x.$$

Performing the j, l sums, this is

$$\ll_k \left(\sum_p \frac{\log p}{p^2} \right) x \log^k x;$$

since the summation is convergent, the claim follows. □

Remark 22. *We have seen that the von Mangoldt function Λ obeys at least three useful identities: (27), (28) and (31). (Not surprisingly, these identities are closely related to each other: (27) is basically the derivative of (28) and the Dirichlet series transform of (31).) It is because of these identities (and further related identities which we will see in later posts) that the von Mangoldt function is an extremely convenient proxy for the prime numbers in analytic number theory. Indeed, many question about primes in analytic number theory are most naturally tackled by converting them to a statement about the von Mangoldt function, by adding or subtracting the contribution of prime powers p^j for $j > 2$, which are typically of significantly lower order), in order to access identities such as (27), (28) or (31).*

3. The Number of Divisors

We now consider the asymptotics of the *divisor function*

$$\tau(n) := \sum_{d|n} 1$$

that counts the number of divisors of a given natural number n. Informally: how many divisors does one expect a typical large number n to have? Remarkably, the answer turns out to depend on exactly on what one means by "typical".

The first basic observation to make is that τ is a multiplicative function: $\tau(nm) = \tau(n)\tau(m)$ whenever n, m are coprime, since every divisor of nm can be uniquely expressed as the product of a divisor of n and a divisor of m. The same argument has an important generalisation: if $f, g : \mathbf{N} \to \mathbf{C}$ are multiplicative functions, then the *Dirichlet convolution* $f * g : \mathbf{N} \to \mathbf{C}$, defined by

$$(f * g)(n) := \sum_{d|n} f(d) g\left(\frac{n}{d}\right)$$

is also multiplicative. The multiplicativity of the divisor function then follows from the identity

$$\tau = 1 * 1,$$

since the constant function $n \mapsto 1$ is clearly multiplicative.

We begin with a crude bound:

Lemma 23. *(Divisor bound) We have $\tau(n) = n^{o(1)}$ as $n \to \infty$. Equivalently, we have $\tau(n) \ll n^\varepsilon$ for any fixed $\varepsilon > 0$.*

Proof. Let $\varepsilon > 0$ be fixed. Observe that for any prime power p^j, we have $\tau(p^j) = j+1$. In particular, we see that $\tau(p^j) \leq p^{\varepsilon j}$ whenever p is sufficiently large depending on ε, and j is a natural number. For the remaining values of p, we have $\tau(p^j) \ll p^{\varepsilon j}$. By the multiplicativity of τ, we thus have $\tau(n) \ll n^\varepsilon$ as required. □

Exercise 24. *If f, g are arithmetic functions obeying (3), show that the Dirichlet convolution $f * g$ also obeys (3). Then establish the fundamental relationship*

$$\mathcal{D}(f * g)(s) = \mathcal{D}f(s)\mathcal{D}g(s) \tag{34}$$

relating the Dirichlet series of $f, g, f * g$ for all $s > 1$. (Compare with analogous identities for the Fourier transform or Laplace transform of ordinary convolutions.) Use this and (31) to obtain an alternate proof of (27). We will make heavier use of (34) in later notes.

Exercise 25. *Obtain the sharper bound $\tau(n) \leq n^{O(1/\log\log n)}$ for all $n \geq 10$. (Hint: first establish that $\tau(n) \leq \exp(\exp(O(1/\varepsilon)))n^{\varepsilon}$ for any $\varepsilon > 0$ by performing the proof of Lemma 23 more carefully, then optimise in ε.) It was in fact established by Wigert that one in fact has*

$$\tau(n) \leq n^{\frac{\log 2 + o(1)}{\log\log n}}$$

as $n \to \infty$, and that the constant $\log 2$ cannot be replaced by any smaller constant.

Now we consider the mean value of τ. From the $k = 2$ case of Theorem 15 we have

$$\sum_{n \leq x} \frac{\tau(n)}{n} \ll \log^2 x$$

and from Proposition 21 we have

$$\sum_{n \leq x} \tau(n) \ll x \log x,$$

which suggest that the mean value of $\tau(n)$ for $n \leq x$ is comparable to $\log x$. We can verify this by using the general identity

$$\begin{aligned}\sum_{n \leq x} f * g(n) &= \sum_{n \leq x} \sum_{d|n} f(d) g(\frac{n}{d}) \\ &= \sum_{d,m: dm \leq x} f(d) g(m)\end{aligned}$$

and thus

$$\sum_{n \leq x} f * g(n) = \sum_{d} f(d) \left(\sum_{m \leq x/d} g(m) \right). \tag{35}$$

A similar argument gives the variant

$$\sum_{n \leq x} \frac{f * g(n)}{n} = \sum_{d \leq x} \frac{f(d)}{d} \left(\sum_{m \leq x/d} \frac{g(m)}{m} \right). \tag{36}$$

We can use this to control the summatory function and logarithmic sum of the divisor function. For instance, by applying (35) with $f = g = 1$ followed by (6), we

have

$$\begin{aligned}\sum_{n\leq x}\tau(n) &= \sum_{d\leq x}(\sum_{m\leq x/d}1)\\ &= \sum_{d\leq x}(\frac{x}{d}+O(1))\end{aligned}$$

and hence by (13)

$$\sum_{n\leq x}\tau(n) = x\log x + O(x). \tag{37}$$

(We will improve the control on the error term here later in this section.) Similarly, from (36) followed by (13) we have

$$\begin{aligned}\sum_{n\leq x}\frac{\tau(n)}{n} &= \sum_{d\leq x}\frac{1}{d}(\sum_{m\leq x/d}\frac{1}{m})\\ &= \sum_{d\leq x}\frac{1}{d}\left(\log\frac{x}{d}+\gamma+O\left(\frac{d}{x}\right)\right)\end{aligned}$$

and thus by (14) and a brief calculation

$$\sum_{n\leq x}\frac{\tau(n)}{n} = P(\log x)+O(1) \tag{38}$$

for some quadratic polynomial P with leading term $\frac{1}{2}t^2$. Comparing these bounds with (9) and (8), we see this is indeed compatible with the heuristic that $\tau(n)$ behaves like $\log n$ on the average.

Remark 26. *One can justify the heuristic $\tau(n)\approx\log n$ by the following non-rigorous probabilistic argument: for a typical large number n, each number $d\leq n$ would be expected to divide n with probability about $\frac{1}{d}$; since $\sum_{d\leq n}\frac{1}{d}\sim\log n$, the "expected value" of $\tau(n)$ should thus be about $\log n$. We will continue this heuristic argument later in this set of notes.*

Even though $\tau(n)$ behaves like $\log n$ on the average, it can fluctuate to be much larger or much smaller than this value. For instance, $\tau(n)$ equals 2 when n is prime, and when n is odd, $\tau(2n)$ is twice as large as $\tau(n)$, even though $\log(2n)$ and $\log n$ are roughly the same size. A further hint of the irregularity of distribution can be seen by considering the k^{th} *moments* $\sum_{n\leq x}\frac{\tau^k(n)}{n}$ and $\sum_{n\leq x}\tau^k(n)$ of n for natural numbers $k\geq 0$. The function τ^k is multiplicative with $\tau^k(p)=2^k$ for every prime p. From Theorem 15 and Proposition 21 we have the upper bounds

$$\sum_{n\leq x}\frac{\tau^k(n)}{n}\ll_k \log^{2^k}x$$

and

$$\sum_{n\leq x}\tau^k(n)\ll_k x\log^{2^k-1}x \tag{39}$$

for all $x \geq 2$, suggesting that $\tau^k(n)$ behaves like $\log^{2^k-1} n$ on average. This may seem at first glance to be incompatible with the heuristic that $\tau(n)$ behaves like $\log n$ on the average, but what is happening here is that $\tau(n)$ can occasionally be much larger than $\log n$, which does not significantly affect the mean of $\tau(n)$, but does affect higher moments with $k > 1$. (Continuing Remark 26, the issue here is that the events "d divides n" can be highly correlated with each other as d varies, due to common factors between different choices of d.) In fact, the *typical* value (in the sense of median, rather than mean) of $\tau(n)$ is not $\log n$ or $\log^{2^k-1} n$, but is in fact $\log^{\log 2+o(1)} n$. See Section 5 below.

We can be more precise on the mean behaviour of τ^k, by establishing the following variant of Theorem 15.

Theorem 27. *(Mean values of divisor-type multiplicative functions) Let $k \geq 0$ be a natural number, and let f be a multiplicative function obeying the estimates*

$$f(p) = k + O_k\left(\frac{1}{p}\right) \tag{40}$$

for all primes p, and

$$f(p^j) = O_k(j^{O_k(1)})$$

for all primes p and all $j \geq 1$. Let $\mathfrak{S} = \mathfrak{S}_{k,f}$ denote the singular series

$$\mathfrak{S} := \prod_p \left(1 - \frac{1}{p}\right)^k \left(1 + \frac{f(p)}{p} + \frac{f(p^2)}{p^2} + \dots\right). \tag{41}$$

Note from Taylor expansion that

$$(1 - \frac{1}{p})^k (1 + \frac{f(p)}{p} + \frac{f(p^2)}{p^2} + \dots) = 1 + O_k(\frac{1}{p^2})$$

so $\mathfrak{S}$ is finite (though it may vanish, if one of its factors vanishes), with $\mathfrak{S} = O_k(1)$.

(i) *If $k \geq 0$, then one has*

$$\mathcal{D}f(s) = \frac{\mathfrak{S}}{(s-1)^k} + O_k(1 + \frac{1}{(s-1)^{k-1}})$$

for all $s > 1$.

(ii) *If $k \geq 0$, we have*

$$\sum_{n \leq x} \frac{f(n)}{n} = \frac{1}{k!}\mathfrak{S} \log^k x + O_k(\log^{k-1}(2+x))$$

for $x \geq 1$.

(iii) *If $k \geq 1$, one has*

$$\frac{1}{x}\sum_{n \leq x} f(n) = \frac{1}{(k-1)!}\mathfrak{S} \log^{k-1} x + O_k(\log^{k-2}(2+x))$$

for $x \geq 1$.

Note that the $k = 2$ case of this proposition gives (a slightly weaker form of) the above estimates for τ, since in that case $\mathfrak{S} = 1$. Comparing with (9), (14), (16) we see that f behaves approximately like $\frac{1}{(k-1)!}\mathfrak{S}\log^k x$ on the average for $k \geq 1$. The hypotheses on this theorem may be relaxed somewhat; see Exercise 29 below.

Proof. For brevity we omit the dependence of implicit constants on k. We begin with (i). If $s - 1 \gg 1$, then from (19) and crude bounds we have $\mathcal{D}f(s) = O(1)$, which suffices, so we will assume that s is sufficiently close to 1. From (19) we have

$$\mathcal{D}f(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-k} \gamma_p(s)$$

where $\gamma_p(s) := (1 - \frac{1}{p^s})^k \sum_{j=0}^\infty \frac{f(p^j)}{p^{sj}}$. For s close to 1, we see from applying the chain rule and Taylor expansion that

$$\frac{d}{ds}\gamma_p(s) = O\left(\frac{\log p}{p^2}\right)$$

and hence

$$\gamma_p(s) = \left(\gamma_p(1) + \left((s-1)\frac{\log p}{p^2}\right)\right)$$

where

$$\gamma_p(1) := \left(1 - \frac{1}{p}\right)^k \left(1 + \frac{f(p)}{p} + \frac{f(p^2)}{p^2} + \dots\right) = 1 + O\left(\frac{1}{p^2}\right).$$

Since $\sum_p \frac{\log p}{p^2} \leq \sum_n \frac{\log n}{n^2}$ is convergent, we conclude that

$$\mathcal{D}f(s) = \left(\prod_p (1 - \frac{1}{p^s})^{-k}\right)(\mathfrak{S} + O(s-1)),$$

and the claim follows from (21).

To prove (ii) we induct on k. First we establish the base case $k = 0$ of (ii). From (19) we have

$$\sum_n \frac{f(n)}{n} = \mathfrak{S}$$

so it suffices to show that

$$\sum_{n>x} \frac{f(n)}{n} = O(\frac{1}{\log x}).$$

But from (19) we see that

$$\sum_n \frac{|f(n)|}{n^{2/3}} = O(1)$$

(for instance), and the claim follows.

Now suppose that $k \geq 1$ and the claim (ii) has already been proven for $k - 1$. We let g be the multiplicative function with

$$g(p^j) := f(p^j) - f(p^{j-1})$$

for primes p and $j \geq 1$, then one easily verifies that $f = 1 * g$. Note that g satisfies the same hypotheses as f, but with k replaced by $k-1$. A brief calculation shows that

$$(1-\frac{1}{p})^{-1}(1+\frac{g(p)}{p}+\frac{g(p^2)}{p^2}+\dots) = 1+\frac{f(p)}{p}+\frac{f(p^2)}{p^2}+\dots$$

and thus $\mathfrak{S}_{k-1,g} = \mathfrak{S}_{k,f} = \mathfrak{S}$. From (36) one has

$$\sum_{n\leq x}\frac{f(n)}{n} = \sum_{d\leq x}\frac{1}{d}\left(\sum_{m\leq x/d}\frac{g(m)}{m}\right)$$

and thus by induction hypothesis

$$\sum_{n\leq x}\frac{f(n)}{n} = \sum_{d\leq x}\frac{1}{d}\left(\frac{1}{(k-1)!}\mathfrak{S}\log^{k-1}(x/d)+O(\log^{k-2}(2+x/d))\right).$$

From Lemma 2 we have

$$\sum_{d\leq x}\frac{\log^{k-1}(x/d)}{d} = \frac{1}{k}\log^k x + O(\log^{k-1}(2+x))$$

and

$$\sum_{d\leq x}\frac{\log^{k-2}(2+x/d)}{d} \ll \log^{k-1}(2+x)$$

and the claim (ii) follows.

Finally, we prove (iii). Let $k \geq 1$; if $k > 1$, we assume inductively that (iii) is established for $k-1$. Let g be as before. From (35) one has

$$\frac{1}{x}\sum_{n\leq x} f(n) = \sum_{d\leq x}\frac{g(d)}{d}\left(\frac{1}{x/d}\sum_{m\leq x/d} 1\right)$$

and thus by (6)

$$\frac{1}{x}\sum_{n\leq x} f(n) = \sum_{d\leq x}\frac{g(d)}{d} + O\left(\frac{1}{x}\sum_{d\leq x}|g(d)|\right).$$

From (ii) we have

$$\sum_{d\leq x}\frac{g(d)}{d} = \frac{1}{(k-1)!}\mathfrak{S}\log^{k-1}(x) + O\left(\log^{k-2}(2+x)\right).$$

The error term $\frac{1}{x}\sum_{d\leq x}|g(d)|$ can be treated by the induction hypothesis when $k > 1$, giving the claim. When $k = 1$, we instead use the Rankin trick and bound

$$\frac{1}{x}\sum_{d\leq x}|g(d)| \leq \frac{1}{x^{1/3}}\sum_n \frac{|g(n)|}{n^{2/3}}$$

and use (19) to bound $\sum_n \frac{|g(n)|}{n^{2/3}} = O(1)$ as before, giving the claim. □

Thus for instance we have

$$\frac{1}{x}\sum_{n\leq x}\tau^k(n) = \frac{1}{(2^k-1)!}\mathfrak{S}_k \log^{2^k-1} x + O_k(\log^{2^k-2}(2+x))$$

for any $k \geq 1$, where $0 < \mathfrak{S}_k < \infty$ is the quantity

$$\mathfrak{S}_k := \prod_p \left(1-\frac{1}{p}\right)^{-2^k}\left(1+\frac{2^k}{p}+\frac{3^k}{p^2}+\dots\right).$$

Among other things, this shows that $\tau(n)$ can be larger than any fixed power of $\log n$ for arbitrarily large n; compare with Lemma 23.

A more accurate description of the distribution of $\tau(n)$ is that $\log\tau(n)$ for $n \leq x$ is asymptotically distributed in the limit $x \to \infty$ like a gaussian random variable with mean and variance comparable to $\log\log x$; see Section 5 below.

Exercise 28. *Let μ^2 be the arithmetic function defined by setting $\mu^2(n) := 1$ when n is* squarefree *(that is, n is not divisible by any perfect square other than 1) and zero otherwise; the reason for the notation μ^2 will be given later. Show that*

$$\frac{1}{x}\sum_{n\leq x}\mu^2(n) = \frac{1}{\zeta(2)} + O\left(\frac{1}{\log x}\right)$$

and

$$\frac{1}{\log x}\sum_{n\leq x}\frac{\mu^2(n)}{n} = \frac{1}{\zeta(2)} + O\left(\frac{1}{\log x}\right)$$

for $x > 2$. Thus, the square-free numbers have natural and logarithmic density $\frac{1}{\zeta(2)}$. It can be shown by a variety of methods that

$$\zeta(2) = \frac{\pi^2}{6},$$

although we will not use this fact here. The error term $O(\frac{1}{\log x})$ may also be improved significantly (as we shall see when we turn to sieve-theoretic methods).

Exercise 29. *The purpose of this exercise is to give an alternate derivation of parts (ii) and (iii) of Theorem 27 that does not rely so strongly on k being an integer; it also allows $f(p)$ to only be close to k/p on average, rather than in the pointwise sense. The arguments here are from Appendix A.2* Friedlander-Iwaniec*, which are in turn based on* this paper of Wirsing*.*

(i) *Let k, f be as in Theorem 27, and define $M_f(x) := \sum_{n\leq x}\frac{f(n)}{n}$. By using (32) with $f(n)$ replaced by $f(n)/n$, establish the integral equation*

$$M_f(x) = \frac{k+1}{\log x}\int_1^x M_f(t)\frac{dt}{t} + O_k(\log^{k-1}(2+x))$$

for all $x \geq 1$, and deduce that

$$M_f(x) = c\log^k x + O_k(\log^{k-1}(2+x))$$

for all $x \geq 1$ and some c independent of x. Then use part (i) of Theorem 27 to deduce that $c = \mathfrak{S}/k!$, thus establishing part (ii) of Theorem 27.

(ii) *Assume the following form of the prime number theorem:* $\sum_{n \leq x} \Lambda(n) = x + O(x/\log x)$ *for all* $x \geq 2$. *By using (32) for* $f(n)$, *establish part (iii) of Theorem 27.*

(iii) *Extend Theorem 27 to the case when* $k \geq 0$ *is real rather than integer (replacing factorials by Gamma functions as appropriate), and (40) is replaced by* $\sum_{p \leq x} f(p) \log p = kx + O_k(x/\log x)$ *for all* $x \geq 10$. *For part (ii) of Theorem 27, one can weaken (40) further to*

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

We now present a basic method to improve upon the estimate (37), namely the *Dirichlet hyperbola method*. The point here is that the error term in (6) is much stronger for large x than for small x, so one would like to rearrange the proof of (37) so that (6) is only used in the former case and not the latter. To do this, we decompose

$$1 = 1_{>\sqrt{x}} + 1_{\leq \sqrt{x}}$$

where $1_{>\sqrt{x}}(n) := 1_{n>\sqrt{x}}$ and $1_{\leq \sqrt{x}}(n) := 1_{n \leq \sqrt{x}}$, so that $\tau(n)$ may be decomposed as

$$\tau = 1_{>\sqrt{x}} * 1_{>\sqrt{x}} + 2 \times 1_{>\sqrt{x}} * 1_{\leq \sqrt{x}} + 1_{\leq \sqrt{x}} * 1_{\leq \sqrt{x}}.$$

Actually, it is traditional to rearrange this a bit as the identity

$$\tau = 1_{>\sqrt{x}} * 1_{>\sqrt{x}} + 2 \times 1_{\leq \sqrt{x}} * 1 - 1_{\leq \sqrt{x}} * 1_{\leq \sqrt{x}}. \tag{42}$$

This decomposition is convenient for improving (37) for two reasons. Firstly, $1_{>\sqrt{x}} * 1_{>\sqrt{x}}$ is supported on $(x, +\infty)$ and thus makes no contribution to $\sum_{n \leq x} \tau(n)$. At the other extreme, $1_{\leq \sqrt{x}} * 1_{\leq \sqrt{x}}$ is supported in $[1, x]$ and so the restriction $n \leq x$ may be removed here, simplifying the sum substantially:

$$\sum_{n \leq x} 1_{\leq \sqrt{x}} * 1_{\leq \sqrt{x}}(n) = \sum_n 1_{\leq \sqrt{x}} * 1_{\leq \sqrt{x}}(n) = \left(\sum_n 1_{\leq \sqrt{x}}(n) \right)^2.$$

For the final sum $1_{\leq \sqrt{x}} * 1$, we use (35) and (6) as before, to conclude that

$$\begin{aligned} \sum_{n \leq x} \tau(n) &= 2 \sum_{d \leq \sqrt{x}} \sum_{n \leq x/d} 1 - (\sum_{n \leq \sqrt{x}} 1)^2 \\ &= 2 \sum_{d \leq \sqrt{x}} (\frac{x}{d} + O(1)) - (\sqrt{x} + O(1))^2 \\ &= 2 \sum_{d \leq \sqrt{x}} \frac{x}{d} - x + O(\sqrt{x}). \end{aligned} \tag{43}$$

By (13), we thus obtain an improved version of (37):

$$\sum_{n \leq x} \tau(n) = x \log x + (2\gamma - 1)x + O(\sqrt{x}). \tag{44}$$

Remark 30. *The Dirichlet hyperbola method may be visualised geometrically as follows, in a fashion that explains the terminology "hyperbola method". The sum* $\sum_{n\leq x}\tau(n) = \sum_{d\leq x}\sum_{n\leq x/d} 1$ *can be interpreted as the number of lattice points (that is to say, elements of* $\mathbf{N}^2$*) lying underneath the hyperbola* $\{(a,b) : ab \leq x\}$*. The proof of (37) basically proceeded to count these lattice points by summing up the contribution of each column separately; this was an efficient process for the columns close to the y-axis, but led to relatively large error terms for columns far away from the y-axis. Symmetrically, one could proceed by summing by rows, which is efficient for rows close to the x-axis, but not far from that axis. The hyperbola method splits the difference between these two counting procedures, counting rows within* $\sqrt{x}$ *of the x-axis and columns within* $\sqrt{x}$ *of the y-axis, and then removing one copy of the square* $[0,\sqrt{x}]^2$ *to correct for it being double-counted. The estimation of this lattice point problem can be made more precise still by more sophisticated decompositions and approximations of the hyperbola, but we will not discuss this problem (known as the* Dirichlet divisor problem*) here.*

From an algebraic perspective, it is the identity (42), decomposing τ *into Dirichlet convolutions of expressions with good spatial support properties, that is the key to the successful application of the hyperbola method. In later posts we will encounter more sophisticated identities that decompose various arithmetic functions (such as the von Mangoldt function* Λ*) into similar convolutions of spatially localised expressions. Unfortunately these identities are not as easy to visualise geometrically as the hyperbola method identity, as the corresponding geometric picture often takes place in three or higher dimensions.*

Exercise 31. *Define the third divisor function* τ_3 *by* $\tau_3 := 1 * 1 * 1$*, or equivalently* $\tau_3(n) = \sum_{d_1,d_2,d_3:d_1d_2d_3=n} 1$*. Show that*

$$\sum_{n\leq x}\tau_3(n) = xP(\log x) + O(x^{2/3}\log x)$$

for all $x \geq 2$ *and some polynomial* $P(t)$ *with leading term* $\frac{1}{2}t^2$*. (Note that Theorem 27(iii) only gives* $\frac{1}{2}x\log^2 x + O(x\log x)$*; one needs a three-dimensional version of the hyperbola method to get the better error term. Hint: Decompose* 1 *at the threshold* $x^{1/3}$*. If one is having difficulty figuring out exactly what identity to use, try working backwards, by writing down all the relevant convolutions (e.g.* $1_{\leq x^{1/3}} * 1_{\leq x^{1/3}} * 1$*) that look tractable, and then do some elementary linear algebra to express* τ_3 *in terms of all the expressions that you know how to estimate well.) The* $O(x^{2/3}\log x)$ *term can be improved further (for instance, the logarithmic factor can be removed), but this requires more sophisticated methods than the ones given above.*

Exercise 32. *State and prove an extension of the previous exercise to the* k^{th} *divisor function* $\tau_k = 1^{*k}$ *for* $k \geq 1$*, defined as the Dirichlet convolution* 1^{*k} *of* k *copies of* 1*.*

4. Merten's Theorems

In the previous section, we used identities such as (35) and (36) to obtain control on statistics of Dirichlet convolutions $f * g$ in terms of statistics of the individual

factors f, g. This method is particularly well suited for functions such as the divisor function τ, which can be expressed conveniently in such Dirichlet convolution form.

When dealing with arithmetic functions related to the primes, it turns out that one often has to run this procedure in reverse, for instance trying to control statistics of f given information on $f * g$ and g. This is basically a *deconvolution* problem, which from a Fourier-analytic point of view corresponds to dividing one Fourier transform by another. This can become problematic when the latter Fourier transform vanishes; in our arithmetic context, this corresponds to zeroes of the Dirichlet series $\mathcal{D}g$. As such, we will see the location of the zeroes of Dirichlet series such as the Riemann zeta function play an extremely important role in later posts. However, the elementary approach cannot easily access this information directly. As such, it is somewhat limited in the type of information it can recover about the primes (at least in the more basic formulations of the theory); however, one can still obtain some non-trivial control on the primes by purely elementary methods, as we shall shortly see.

Let us first see where this deconvolution problem is coming from. As discussed in Remark 22, it is convenient to study the primes through the von Mangoldt function. The identity (31) concerning this function can be rewritten as

$$L = \Lambda * 1 \tag{45}$$

where $L : n \mapsto \log n$ is the logarithm function. Thanks to the methods in Section 1, we understand the statistics of L and 1 relatively well. The "deconvolution problem" is then to somehow use this information to control the statistics of Λ. We demonstrate this with the following improvement of Exercise 17, controlling logarithmic sums of the von Mangoldt function:

Theorem 33. *(First Mertens theorem) We have the estimate*

$$\sum_{n \leq x} \frac{\Lambda(n)}{n} = \log x + O(1). \tag{46}$$

for all $x \geq 1$.

Proof. Applying (35) to (45) we have

$$\sum_{n \leq x} \log n = \sum_{d \leq x} \Lambda(d) \sum_{n \leq x/d} 1. \tag{47}$$

If we apply (8) and (6), we conclude that

$$x \log x + O(x) = x \sum_{d \leq x} \frac{\Lambda(d)}{d} + O(\sum_{d \leq x} \Lambda(d)) \tag{48}$$

and so (46) follows from Proposition 18. □

It is easy to convert control of Λ to control on primes:

Corollary 34. *(Alternate form of first Mertens' theorem) We have*

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

for all $x \geq 1$.

Proof. From the definition of the von Mangoldt function, one has

$$\sum_{n \leq x} \frac{\Lambda(n)}{n} = \sum_{p \leq x} \frac{\log p}{p} + \sum_{j=2}^{\infty} \sum_{p \leq x^{1/j}} \frac{\log p}{p^j}.$$

We crudely bound $\frac{\log p}{p^j} \ll \frac{1}{2^{j/2}} \frac{\log p}{p^{3/2}}$ for $j \geq 2$. Since $\sum_n \frac{\log n}{n^{3/2}} = O(1)$, we obtain

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

and the claim follows from (46). □

Exercise 35. *(Chebyshev's theorem) Show that there exists an absolute constant $0 < c < 1$ such that there are $\asymp \frac{x}{\log x}$ primes in $[cx, x]$ for all sufficiently large x. Conclude in particular that $\sum_{n \leq x} \Lambda(n) \asymp x$.*

From this we can obtain a sharper form of Theorem 13, controlling logarithmic sums of the primes:

Theorem 36. *(Second Mertens' theorem) One has*

$$\sum_{p \leq x} \frac{1}{p} = \log\log x + c_1 + o(1)$$

as $x \to \infty$ for some absolute constant c_1. Similarly, one has

$$\sum_{n \leq x} \frac{\Lambda(n)}{n \log n} = \log\log x + c_2 + o(1) \tag{49}$$

as $x \to \infty$ for some absolute constant c_2.

The constant c_1 has no particular significance in applications, but the constant c_2 can be usefully computed: see (51) below.

Proof. We just prove the first claim, as the second is similar (and can be deduced from the first by a modification of the argument used to prove Corollary 34). One could proceed here using *summation by parts*, but we will argue using an essentially equivalent method, based on the fundamental theorem of calculus. By Lemma 5, it suffices to show that

$$\sum_{y \leq p \leq x} \frac{1}{p} = \log\log x - \log\log y + o(1)$$

as $y < x$ both go to infinity. From the fundamental theorem of calculus we have

$$\frac{1}{p} = \frac{\log p}{p} \left(\frac{1}{\log x} + \int_y^x 1_{p \leq t} \, \frac{dt}{t \log^2 t} \right)$$

for all $y \leq p \leq x$, and thus

$$\sum_{y \leq p \leq x} \frac{1}{p} = \frac{1}{\log x} \sum_{y \leq p \leq x} \frac{\log p}{p} + \int_y^x \sum_{y \leq p \leq t} \frac{\log p}{p} \frac{dt}{t \log^2 t}.$$

From Corollary 34 one has

$$\sum_{y \leq p \leq t} \frac{\log p}{p} = \log t - \log y + O(1)$$

for all $t \geq y$; since

$$\int_y^x \frac{dt}{t \log^2 t} = \frac{1}{\log y} - \frac{1}{\log x} = o(1)$$

and

$$\int_y^x (\log t - \log y) \frac{dt}{t \log^2 t} = \log\log x - \log\log y + \frac{\log y}{\log x} - 1,$$

the claim follows. □

This leads to a useful general formula for computing various slowly varying sums over primes:

Exercise 37.

(i) *For any fixed $0 < a < b < \infty$, show that*

$$\sum_{x^a \leq p \leq x^b} \frac{1}{p} = \log \frac{b}{a} + o(1)$$

as $x \to \infty$.

(ii) *If $f : (0, +\infty) \to \mathbf{C}$ is a fixed compactly supported, Riemann integrable function, show that*

$$\sum_p \frac{1}{p} f\left(\frac{\log p}{\log x}\right) = \int_0^\infty f(t)\ \frac{dt}{t} + o(1)$$

as $x \to \infty$.

(iii) *If $d \geq 1$ is a fixed natural number and $f : (0, +\infty)^d \to \mathbf{C}$ is a fixed compactly supported, Riemann integrable function, show that*

$$\sum_{p_1, \ldots, p_d} \frac{1}{p_1 \ldots p_d} f\left(\frac{\log p_1}{\log x}, \ldots, \frac{\log p_d}{\log x}\right) = \int_{(0,\infty)^d} f(t_1, \ldots, t_d)\ \frac{dt_1 \ldots dt_d}{t_1 \ldots t_d} + o(1)$$

as $x \to \infty$.

(iv) *Obtain analogues of (i)-(iii) when the sum over primes p are replaced by the sum over integers n, but with factors of $\frac{1}{p}$ replaced by $\frac{\Lambda(n)}{n \log n}$ (with the convention that this expression vanishes at $n = 0$).*

Remark 38. *An alternate way to phrase the above exercise is that the* Radon measures

$$\sum_p \frac{1}{p} \delta_{\frac{\log p}{\log x}}$$

on $(0,+\infty)$ converge in the vague topology *to the absolutely continuous measure $\frac{dt}{t}$ in the limit $x \to \infty$, where δ_t denotes the Dirac probability measure at t. Similarly for the Radon measures*

$$\sum_n \frac{\Lambda(n)}{n \log n} \delta_{\frac{\log n}{\log x}}.$$

To put this another way, the Radon measures $\sum_p \frac{\log p}{p} \delta_{\log p}$ or $\sum_n \frac{\Lambda(n)}{n} \delta_{\log n}$ behave like Lebesgue measure on dyadic intervals such as $[u,(1+\varepsilon)u]$ for fixed $\varepsilon > 0$ and large u. This is weaker than the prime number theorem *that we will prove in later notes, which basically asserts the same statement but on the much smaller intervals $[u, u+\varepsilon]$. (Statements such as the Riemann hypothesis make the same assertion on even finer intervals, such as $[u, u + \exp(-(\frac{1}{2} - \varepsilon)u)]$.) See this* previous blog post *for some examples of this Radon measure perspective, which we will not emphasise in this set of notes.*

Exercise 39. *(Smooth numbers) For any $x, z > 0$, let $S(x,z)$ denote the set of natural numbers less than x which are z-smooth, in the sense that they have no prime factors larger than z.*

(i) *Show that for any fixed $u > 0$, one has*

$$|S(x, x^{1/u})| = (\rho(u) + o(1))x$$

where the Dickman function *$\rho(u)$ is defined by the alternating series*

$$\rho(u) := 1 + \sum_{j=1}^{\infty} \frac{(-1)^j}{j!} \int_{[1,+\infty)^j} 1_{t_1+\cdots+t_j \leq u} \frac{dt_1 \ldots dt_j}{t_1 \ldots t_j}$$

(note that for any given u that only finitely many of the summands are non-zero; one can also view the 1 term as the $j = 0$ term of the summation after carefully working out what zero-dimensional spaces and empty products evaluate to). Hint: Use the inclusion-exclusion principle *to remove the multiples of p from $[1,x]$ for each prime $p > z$. This is a simple example of a* sieve*, which we will study in more detail in later notes.*

(ii) *(Delay-differential equation) Show that ρ is continuous on $(0,+\infty)$, continuously differentiable except at $u = 1$, equals 1 for $0 \leq u \leq 1$ and obeys the equation*

$$\frac{d}{du}\rho(u) = -\frac{1}{u}\rho(u-1)$$

for $u > 1$. Give a heuristic justification for this equation by considering how $S(x, x^{1/u})$ varies with respect to small perturbations of u.

(iii) *(Wirsing integral equation) Show that*

$$u\rho(u) = \int_0^u 1_{[0,1]}(t)\rho(u-t)\, dt$$

for all $u \geq 0$, or equivalently that

$$u\rho = 1_{[0,1]} * \rho.$$

Give a heuristic justification for this equation by starting with a $x^{1/u}$-smooth number $n \in [1,x]$ and considering a factor n/p, where p is a prime factor of n chosen at random with probability $j\frac{\log p}{\log n}$ (if p occurs j times in the prime factorisation of n).

(iv) *(Laplace transform) Show that*

$$\int_0^\infty \rho(u) e^{-su}\ du = \frac{1}{s} \exp(-\int_s^\infty \frac{e^{-t}}{t}\ dt)$$

for all $s > 0$.

Now we incorporate the information on primes coming from Euler products that we established in Section 2. Recall from (11), (28) that

$$\sum_n \frac{\Lambda(n)}{n^s \log n} = -\log(s-1) + O(s-1) \tag{50}$$

for $s > 1$. We can compare this against (49) by a variant of the Rankin trick. Namely, if we apply (50) with $s = 1 + \frac{1}{\log x}$ for some large x, one obtains

$$\sum_n \frac{\Lambda(n)}{n \log n} e^{-\log n/\log x} = \log\log x + o(1)$$

and thus on subtracting (49)

$$\sum_n \frac{\Lambda(n)}{n\log n}\left(e^{-\log n/\log x} - 1_{[0,1]}\left(\frac{\log n}{\log x}\right)\right) = -c_2 + o(1).$$

For any fixed $0 < \varepsilon < 1 < N$, we see from Exercise 37 that

$$\begin{aligned}\sum_{x^\varepsilon \leq n \leq x^N} &\frac{\Lambda(n)}{n\log n}(e^{-\log n/\log x} - 1_{[0,1]}(\frac{\log n}{\log x})) \\ &= \int_\varepsilon^N (e^{-t} - 1_{[0,1]}(t))\frac{dt}{t} + o(1) \\ &= \int_\varepsilon^N \frac{e^{-t}}{t}\ dt - \log\frac{1}{\varepsilon} + o(1).\end{aligned}$$

On the other hand, by using Exercise 37 and dyadic decomposition of $\log n$ we see that the expressions

$$\sum_{n<x^\varepsilon} \frac{\Lambda(n)}{n\log n}\left(e^{-\log n/\log x} - 1_{[0,1]}\left(\frac{\log n}{\log x}\right)\right)$$

and

$$\sum_{n>x^N} \frac{\Lambda(n)}{n\log n}(e^{-\log n/\log x} - 1_{[0,1]}(\frac{\log n}{\log x}))$$

can be made arbitrarily small by making ε sufficiently small, N sufficiently large, and x sufficiently large. Putting all this together, we conclude that

$$-c_2 = \lim_{\varepsilon\to 0} \int_\varepsilon^\infty \frac{e^{-t}}{t} - \log\frac{1}{\varepsilon}.$$

We can compute this limit explicitly:

Lemma 40. *(Exponential integral asymptotics) For sufficiently small ε, one has*

$$\int_\varepsilon^\infty \frac{e^{-t}}{t}\ dt = \log\frac{1}{\varepsilon} - \gamma + O(\varepsilon).$$

Proof. We start by using the identity $\frac{1}{i} = \int_0^1 x^{i-1}\ dx$ to express the harmonic series $H_n := 1 + \frac{1}{2} + \ldots + \frac{1}{n}$ as

$$H_n = \int_0^1 1 + x + \ldots + x^{n-1}\ dx$$

or on summing the geometric series

$$H_n = \int_0^1 \frac{1-x^n}{1-x}\ dx.$$

Since $\int_0^{1-1/n} \frac{1}{1-x} = \log n$, we thus have

$$H_n - \log n = \int_0^1 \frac{1_{[1-1/n,1]}(x) - x^n}{1-x}\ dx;$$

making the change of variables $x = 1 - \frac{t}{n}$, this becomes

$$H_n - \log n = \int_0^n \frac{1_{[0,1]}(t) - (1-\frac{t}{n})^n}{t}\ dt.$$

As $n \to \infty$, $\frac{1_{[0,1]}(t)-(1-\frac{t}{n})^n}{t}$ converges pointwise to $\frac{1_{[0,1]}(t)-e^{-t}}{t}$ and is pointwise dominated by $O(e^{-t})$. Taking limits as $n \to \infty$ using dominated convergence and (13), we conclude that

$$\gamma = \int_0^\infty \frac{1_{[0,1]}(t) - e^{-t}}{t}\ dt.$$

or equivalently

$$\int_0^\infty \frac{e^{-t} - 1_{[0,\varepsilon]}(t)}{t}\ dt = \log\frac{1}{\varepsilon} - \gamma.$$

The claim then follows by bounding the $\int_0^\varepsilon$ portion of the integral on the left-hand side. □

Exercise 41. *Show that*

$$\int_0^\infty e^{-t}\log t\ dt = -\gamma.$$

Conclude that if $\Gamma(s)$ is the Gamma function*, defined for $Re(s) > 0$ by the formula*

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

then one has $\Gamma'(1) = -\gamma$.

From this lemma we see that

$$c_2 = \gamma, \tag{51}$$

thus

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

By arguing as in the proof of Corollary 34, we then have

$$\sum_{p \leq x} \sum_{j=1}^{\infty} \frac{1}{jp^j} = \log\log x + \gamma + o(1)$$

or equivalently by Taylor expansion

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

Taking exponentials, we conclude

Theorem 42. *(Third Mertens' theorem) We have*

$$\prod_{p \leq x} (1 - \frac{1}{p}) = \frac{e^{-\gamma} + o(1)}{\log x}$$

as $x \to \infty$.

Because of this theorem, the factors $e^{\gamma} = 1.78107\ldots$ and $e^{-\gamma} = 0.56145\ldots$ frequently arise in analytic prime number theory. We give two examples of this below.

Exercise 43. *Let ρ be the Dickman function from Exercise 39. Show that*

$$\int_0^{\infty} \rho(u)\ du = e^{\gamma}.$$

Exercise 44. *For any natural number n, define the* Euler totient function *$\phi(n)$ of n to be the number of natural numbers less than n that are coprime to n, or equivalently $\phi(n)$ is the order of the multiplicative group $(\mathbf{Z}/n\mathbf{Z})^\times$ of $\mathbf{Z}/n\mathbf{Z}$.*

(i) *Show that*

$$\frac{n}{\log\log n} \ll \phi(n) \leq n$$

for all $n \geq 100$.

(ii) *Show the more refined lower bound*

$$\phi(n) \geq (e^{-\gamma} - o(1)) \frac{n}{\log\log n}$$

as $n \to \infty$. Show that $e^{-\gamma}$ cannot be replaced here by any larger quantity. (This result is due to Landau.)

5. The number of prime factors (optional)

Given a natural number n, we use $\omega(n)$ to denote the number of prime factors of n (not counting multiplicity), and $\Omega(n)$ to denote the number of prime factors of n (counting multiplicity). Thus we can write ω, Ω as *divisor sums*

$$\omega(n) = \sum_{p|n} 1 \tag{52}$$

and

$$\Omega(n) = \sum_{j=1}^{\infty} \sum_{p:p^j|n} 1.$$

We can ask what the mean values of ω and Ω are. We start with the estimation of

$$\sum_{n\leq x} \omega(n).$$

We can rearrange this sum (cf. (35)) as

$$\sum_{p\leq x} \sum_{m\leq x/p} 1$$

which by (6) is

$$\sum_{p\leq x} \frac{x}{p} + O\left(\sum_{p\leq x} 1\right)$$

so by Theorem 36 (and crudely bounding $O(\sum_{p\leq x} 1)$ by $O(x)$, as we will not need additional accuracy here) gives

$$\sum_{n\leq x} \omega(n) = x \log\log x + O(x) \tag{53}$$

for $x \geq 10$ (say). Thus we see that for $n \leq x$, one expects n to have about $\log\log x$ prime factors on the average.

Now we look at the second moment

$$\sum_{n\leq x} \omega(n)^2.$$

If we expand this out directly, we get

$$\sum_{n\leq x} \sum_{p_1,p_2:p_1,p_2|n} 1$$

which we can rearrange as

$$\sum_{p_1,p_2\leq x} \sum_{m\leq \frac{x}{[p_1,p_2]}} 1$$

where $[p_1, p_2]$ is the least common multiple of p_1, p_2, that is to say $p_1 p_2$ if $p_1 \neq p_2$ and p_1 if $p_1 = p_2$. Here we run into a serious problem, which is that $\frac{x}{[p_1,p_2]}$ can be

significantly less than 1, in which case the estimate

$$\sum_{m \leq \frac{x}{[p_1,p_2]}} 1 = \frac{x}{[p_1,p_2]} + O(1)$$

is horrendously inefficient (the error term is larger than the main term, which is always a bad sign). One could fix this by using the trivial estimate $\sum_{m \leq \frac{x}{[p_1,p_2]}} 1 = 0$ when $[p_1,p_2] > x$. But there is another cheap trick one can use here, due to Turán, coming from the fact that a given natural number $n \leq x$ can have at most $O(1)$ prime factors that are larger than, say, $x^{1/10}$. Thus we can approximate the divisor sum (52) by a *truncated* divisor sum:

$$\omega(n) = \sum_{p|n: p \leq x^{1/10}} 1 + O(1).$$

One can then run the previous argument with the truncated divisor sum and avoid the problem of $\frac{x}{[p_1,p_2]}$ dipping below 1. Indeed, on squaring we see that

$$\omega(n)^2 = \left(\sum_{p|n: p \leq x^{1/10}} 1 \right)^2 + O(\omega(n)) + O(1)$$

and hence (by (53))

$$\sum_{n \leq x} \omega(n)^2 = \sum_{n \leq x} \sum_{p_1,p_2 \leq x^{1/10}: p_1,p_2|n} 1 + O(x \log\log x)$$

for $x \geq 10$. Rearranging and using (6) as before, we obtain

$$\sum_{n \leq x} \omega(n)^2 = \sum_{p_1,p_2 \leq x^{1/10}} \left(\frac{x}{[p_1,p_2]} + O(1) \right) + O(x \log\log x).$$

The contribution of the $O(1)$ error may be crudely bounded by $O(x^{1/10} \times x^{1/10})$, which can easily be absorbed into the error term. The diagonal case $p_1 = p_2$ also contributes $O(x \log\log x)$ thanks to Theorem 36. We conclude that

$$\sum_{n \leq x} \omega(n)^2 = \sum_{p_1,p_2 \leq x^{1/10}} \frac{x}{p_1 p_2} + O(x \log\log x)$$

and thus by a final application of Theorem 36

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

We can combine this with (53) to give a variance bound:

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

To interpret this probabilistically, we see that if we pick $n \leq x$ uniformly at random, then the random quantity $\omega(n)$ will have mean $\log\log x + O(1)$ and standard deviation $O(\sqrt{\log\log x})$. In particular, by Chebyshev's inequality, we expect n to have $\log\log x + O(\sqrt{\log\log x})$ prime factors most of the time.

Remark 45. *Remark 45 The strategy of truncating a divisor sum to obtain better control on error terms (perhaps at the expense of some inefficiency in the main terms) is one of the core techniques in* sieve theory*, which we will discuss in later notes.*

The same estimates are true for Ω:

Exercise 46.

(i) *Show that*

$$\sum_{n\leq x}(\Omega(n)-\omega(n))^k \ll x$$

for $k = 1, 2$, and conclude that

$$\sum_{n\leq x}(\Omega(n)-\log\log x)^2 \ll x\log\log x.$$

(ii) *Show that*

$$\sum_{n\leq x}(\Omega(n)-\omega(n))^k \ll_k x$$

for all natural numbers k.

From the above exercise and Chebyshev's inequality, we now know the typical number of prime factors of a large number, a fact known as the *Hardy-Ramanujan theorem*:

Theorem 47. *(Hardy-Ramanujan theorem) Let x be an asymptotic parameter going to infinity, and let $a(x)$ be any quantity depending on x that goes to infinity as $x \to \infty$. Let n be a natural number selected uniformly at random from $\{n \in \mathbf{N} : n \leq x\}$. Then with probability $1 - o(1)$, n has $\log\log x + O(a(x)\sqrt{\log\log x})$ distinct prime factors, and $O(a(x))$ repeated prime factors (counting multiplicity, thus p^j counts as $j-1$ repeated prime factors). In particular, n has $\log\log x + O(a(x)\sqrt{\log\log x})$ prime factors counting multiplicity.*

This already has a cute consequence with regards to the multiplication table:

Proposition 48. *(Multiplication table bound) At most $o(x^2)$ of the natural numbers up to x^2 can be written in the form ab with $a, b \leq x$ natural numbers.*

In other words, for large N, the $N \times N$ multiplication table only uses a small fraction of the numbers up to N^2. This simple-sounding fact is surprisingly hard to prove if one does not use the simple argument provided below.

Proof. Pick natural numbers $a, b \leq x$ uniformly at random. By the Hardy-Ramanujan theorem, with probability $1 - o(1)$, a and b will each have $(1 + o(1))\log\log x$ prime factors counting multiplicity. Hence with probability $1 - o(1)$, ab will have $(2 + o(1))\log\log x$ prime factors counting multiplicity. But by a further application of the Hardy-Ramanujan theorem, the set of natural numbers up to x^2 with this property has cardinality $o(x^2)$. Thus all but $o(x^2)$ of the products ab with $a, b \leq x$ are contained in a set of cardinality $o(x^2)$, and the claim follows. □

Remark 49. *In fact, the cardinality of the $N \times N$ multiplication table is known to be comparable to*

$$\frac{N^2}{\log^\delta N (\log\log N)^{3/2}}$$

with $\delta := 1 - \frac{1+\log\log 2}{\log 2}$; see this paper of Ford.

Exercise 50. *(Typical number of divisors) Let x be an asymptotic parameter going to infinity. Show that one has $\tau(n) = \log^{\log 2 + o(1)} n = \log^{\log 2 + o(1)} x$ for all but $o(x)$ of the natural numbers n less than x. (Hint: first establish the bounds $2^{\omega(n)} \leq \tau(n) \leq 2^{\Omega(n)}$.)*

In fact, one can describe the distribution of $\omega(n)$ or $\Omega(n)$ more precisely, a fact known as the *Erdös-Kac theorem*:

Exercise 51. *(Erdös-Kac theorem) (This exercise is intended for readers who are familiar with probability theory, and more specifically with the moment method proof of the central limit theorem, as discussed in* this previous set of notes*.) Let x be an asymptotic parameter going to infinity, and let ω' denote the truncated divisor sum*

$$\omega'(n) := \sum_{p|n: p \leq x^{1/(\log\log x)^{1/10}}} 1.$$

Define the quantity

$$\mu := \sum_{p \leq x^{1/(\log\log x)^{1/10}}} \frac{1}{p},$$

thus by Mertens' theorem

$$\mu = \log\log x + o((\log\log x)^{1/2}).$$

(i) *Show that $\omega'(n) = \omega(n) + o((\log\log x)^{1/2})$ for all $n \leq x$.*

(ii) *For any fixed natural number k, show that*

$$\frac{1}{x} \sum_{n \leq x} (\omega'(n) - \mu)^k = (c_k + o(1))(\log\log x)^{k/2},$$

where the quantity c_k is defined to equal zero when k is odd, or

$$c_k := \frac{k!}{(k/2)! 2^{k/2}}$$

when k is even.

(iii) *If n is drawn uniformly at random from the natural numbers up to x, show that the random variable*

$$\frac{\omega(n) - \log\log x}{\sqrt{\log\log x}}$$

converges in distribution to the standard normal distribution $\frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$ in the limit $x \to \infty$. (You will need something like the moment continuity theorem *from Theorem 4* of these notes*.)*

(iv) *Obtain the same claim as (iii), but with $\omega(n)$ replaced by $\Omega(n)$.*

Informally, we thus have the heuristic formula

$$\omega(n), \Omega(n) \approx \log\log x + \sqrt{\log\log x} G$$

for $n \leq x$, where G is distributed approximately according to a standard normal distribution. As in Exercise 50, this leads to a related heuristic formula

$$\tau(n) \approx 2^{\log\log x + \sqrt{\log\log x} G} \tag{54}$$

for the number of divisors. This helps reconcile (though does not fully explain) the discrepancy between the typical (or median) value of $\tau(n)$, which is $\log^{\log 2 + o(1)} x$, and the mean (or higher moments) of $\tau(n)$, which is of the order of $\log x$ or $\log^{2^k - 1} x$, as it suggests that $\tau(n)$ is in fact significantly larger than its median value of $\log^{\log 2 + o(1)} x$ with a relatively high probability. (Unfortunately, though, the heuristic (54) is not very accurate at the very tail end of the distribution when $\tau(n)$ is extremely large, and one cannot recover the correct exponent of the logarithm in (39), for instance, through a naive application of this heuristic.)

Remark 52. *Another rough heuristic to keep in mind is that a "typical" number n less than x would be expected to have $\asymp 1$ divisors in any dyadic block $[\exp(j), \exp(j+1)]$ between 1 and x, and $\asymp 1$ prime divisors in any hyper-dyadic block $[\exp(\exp(j)), \exp(\exp(j+1))]$ between 1 and x. For instance, for any fixed $0 < \alpha < \beta < 1$, one should have $\asymp \log x$ divisors in the interval $[x^\alpha, x^\beta]$, but only $\asymp 1$ prime divisors. Typically, the only prime divisors that occur with multiplicity will be quite small (of size $O(1)$). Note that such heuristics are compatible with the fact that $\tau(n)$ has mean $\sim \log x$ and that $\omega(n)$ and $\Omega(n)$ both have mean $\sim \log\log x$. One can make these heuristics more precise by introducing the Poisson-Dirichlet process, as is done in* this previous blog post*, but we will not do so here. The study of the distribution of factors of "typical" large natural numbers is a topic sometimes referred to as* anatomy of integers*. Interestingly, there is a strong analogy between this problem and the problem of studying the distribution of cycles of "typical" large permutations; see for instance* this article of Granville *for further discussion.*

Exercise 53. *Let k be a natural number. Show that for sufficiently large x, the number of natural numbers up to x that are the products of exactly k distinct primes is $\asymp_k \frac{x}{\log x} (\log\log x)^{k-1}$.*

6. Mobius inversion and the Selberg symmetry formula

In Section 4, we used the identity $L = \Lambda * 1$, together with elementary estimates on L and 1, to deduce various estimates on the von Mangoldt function Λ. Another way to extract information about Λ from this identity is to "deconvolve" or "invert" the operation of convolution to 1. This can be achieved by the basic tool of *Möbius inversion*, which we now discuss. We first observe that the Kronecker delta function $\delta : \mathbf{N} \to \mathbf{R}$, defined by $\delta(n) := 1_{n=1}$, is an identity for Dirichlet convolution, thus

$$f * \delta = \delta * f = f$$

for any arithmetic function f. Since Dirichlet convolution is associative and commutative, this implies that if we can find an arithmetic function μ with the property

that

$$\mu * 1 = \delta, \tag{55}$$

then any formula of the form $f * 1 = F$ may be inverted to the equivalent form $F * \mu = f$, a fact known as the *Möbius inversion formula*. It is then a routine matter to locate such a function μ:

Exercise 54. *Define the* Möbius function $\mu : \mathbf{N} \to \mathbf{R}$ *by setting* $\mu(n) := (-1)^k$ *when* n *is the product of* k *distinct primes for some* $k \geq 1$*, and* $\mu(n) = 0$ *otherwise. Show that* μ *is the unique arithmetic function that obeys (55).*

Observe that μ is a multiplicative function that obeys the trivial bound

$$|\mu(n)| \leq 1 \tag{56}$$

for all n. Furthermore, $\mu^2(n)$ is 1 precisely when n is square-free, and zero otherwise, so the notation here is consistent with that in Exercise 28.

One can express the Möbius inversion formula as the assertion that

$$f(1) = \sum_d \mu(d) \sum_m f(dm) \tag{57}$$

for any compactly supported arithmetic function f. This already reveals that μ must exhibit some cancellation beyond the trivial bound (56):

Lemma 55. *(Weak cancellation for* μ*) For any non-negative integer* k*, we have*

$$\sum_{n \leq x} \frac{\mu(n)}{n} \log^k \left(\frac{x}{n}\right) = Q_k(\log x) + O_k(1) \tag{58}$$

for all x *and some polynomial* $Q_k(t)$ *with leading term* kt^{k-1} *if* $k \geq 1$ *(and* $Q_k = 0$ *if* $k = 0$*).*

Proof. We may of course take $x \geq 1$.

First suppose that $k = 0$. We apply (57) with $f(n) := 1_{n \leq x}$ to conclude that

$$\sum_{d \leq x} \mu(d) \sum_{m \leq x/d} 1 = 1. \tag{59}$$

Since $\sum_{m \leq x/d} 1 = \frac{x}{d} + O(1)$, we conclude from (56) that

$$\sum_{d \leq x} \mu(d) \frac{x}{d} = O(x),$$

and the $k = 0$ case of (58) follows.

Now suppose inductively that $k \geq 1$, and that the claim has already been proven for smaller values of k. We apply (57) with $f(n) := 1_{n \leq x} \frac{\log^{k-1}(\frac{x}{n})}{n}$ to conclude that

$$\sum_{d \leq x} \frac{\mu(d)}{d} \sum_{m \leq x/d} \frac{\log^{k-1} \frac{x}{dm}}{m} = \log^{k-1} x.$$

By (15) we have

$$\sum_{m \leq x/d} \frac{\log^{k-1}\left(\frac{x}{dm}\right)}{m} = P\left(\log \frac{x}{d}\right) + O_k\left(\frac{1+\log^{k-1}(\frac{x}{d})}{x/d}\right)$$

for some polynomial $P(t)$ with leading term $\frac{1}{k}t^k$. Inserting this asymptotic and using the induction hypothesis to handle all the lower order terms of P, and (56) to handle the error term, we conclude that

$$\frac{1}{k}\sum_{d \leq x} \frac{\mu(d)}{d} \log^k \frac{x}{d} = \log^{k-1} x + P(\log x) + O_k(1 + \sum_{d \leq x} \frac{1+\log^{k-1}\left(\frac{x}{d}\right)}{x})$$

for some polynomial P of degree at most $k-2$. By (10) the error term is $O_k(1)$, and the claim follows. □

Exercise 56. *Sharpen the $k=0$ case of the above lemma to $|\sum_{n \leq x} \frac{\mu(n)}{n}| \leq 1$, for any $x > 0$.*

From Möbius inversion we can write Λ in terms of μ:

$$\Lambda = \mu * L. \tag{60}$$

Since $L(nm) = L(n) + L(m)$, we have the general Leibniz-type identity

$$L \times (f * g) = (Lf) * g + f * (Lg). \tag{61}$$

Since $L\delta = 0$, we can obtain the alternative representation

$$\Lambda = -(L\mu) * 1 \tag{62}$$

by multiplying (55) by L then applying (61), (60).

Using these identities and Lemma 55, we can recover many of the estimates in Section 4:

Exercise 57. *(Alternate proof of Chebyshev and Mertens bounds) Use (60) and Lemma 55 to reprove the estimates*

$$\sum_{n \leq x} \Lambda(n) \ll x$$

and

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

If one has a little bit more cancellation in the Möbius function, one can do better:

Theorem 58. *(Equivalent forms of the prime number theorem) The following three statements are logically equivalent:*

(i) *We have $\sum_{n \leq x} \Lambda(n) = x + o(x)$ as $x \to \infty$.*
(ii) *We have $\sum_{n \leq x} \mu(n) = o(x)$ as $x \to \infty$.*
(iii) *We have $\sum_{n \leq x} \frac{\mu(n)}{n} = o(1)$ as $x \to \infty$.*

In later notes we will prove that the claims (i), (ii), (iii) are indeed true; this is the famous *prime number theorem*. This result also illustrates a general principle, that one route to distribution estimates of the primes is via distribution estimates on the Möbius function, which can sometimes be a simpler object to study (for instance, the Möbius function is bounded in magnitude by 1, whereas the von Mangoldt function can grow logarithmically).

Proof. We use some *arguments of Diamond*. We first show that (i) implies (ii). From (62) and Möbius inversion we have

$$L\mu = -\mu * \Lambda. \tag{63}$$

By (35) we thus have

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

For any fixed $\varepsilon > 0$, we may use (i) to write $\sum_{m\leq x/d}\Lambda(m)$ as $x/d + O(\varepsilon x/d)$ when x/d is sufficiently large, and $O(x/d)$ otherwise. From this, (56), and the $k = 0$ case of (58) we thus have

$$\sum_{n\leq x} \mu(n)\log n = O(\varepsilon x \log x)$$

when x is large enough. By (10), (56) we have

$$\sum_{n\leq x} \mu(n)\log \frac{x}{n} = O(x);$$

summing and then dividing by $\log x$, we obtain (ii) since ε is arbitrary.

Now we show that (ii) implies (iii). We start with the identity (59), which we write as

$$\sum_{d\leq x}\mu(d)\lfloor\frac{x}{d}\rfloor = 1.$$

Let $\varepsilon > 0$ be a small fixed quantity. For $\varepsilon x < d \leq x$, $\lfloor\frac{x}{d}\rfloor$ decreases through a fixed set of values, and from (ii) we conclude that

$$\sum_{\varepsilon x < d\leq x}\mu(d)\lfloor\frac{x}{d}\rfloor = o(x).$$

Meanwhile, since $\lfloor\frac{x}{d}\rfloor = \frac{x}{d} + O(1)$, we see from (56) that

$$\sum_{d\leq \varepsilon x}\mu(d)\lfloor\frac{x}{d}\rfloor = x\sum_{d\leq\varepsilon x}\frac{\mu(d)}{d} + O(\varepsilon x).$$

Combining all three inequalities and dividing by x, we conclude that

$$\sum_{d\leq\varepsilon x}\frac{\mu(d)}{d} = O(\varepsilon) + o(1);$$

replacing x by x/ε, then sending ε to zero, we obtain (iii).

To prove that (iii) implies (i), we observe the identity

$$\sum_{n\leq x}\mu(n) = \int_0^x \sum_{t\leq n\leq x} \frac{\mu(n)}{n}\ dt.$$

For any $\varepsilon > 0$, we have from (iii) that $\sum_{t\leq n\leq x}\frac{\mu(n)}{n} = O(\varepsilon)$ if t is sufficiently depending on ε, and $O(1)$ otherwise; thus

$$\sum_{n\leq x}\mu(n) = O(1) + O(\varepsilon x) = O(\varepsilon x)$$

if x is large enough, and the claim follows.

To conclude the theorem, it suffices to show that (ii) and (iii) jointly imply (i). From (60), (35) we have

$$\sum_{n\leq x}\Lambda(n) = \sum_{d\leq x}\mu(d)\sum_{m\leq x/d}\log m.$$

Meanwhile, since $1 = 1*1*\mu = \tau*\mu$, we have from (35) that

$$\sum_{n\leq x} 1 = \sum_{d\leq x}\mu(d)\sum_{m\leq x/d}\tau(m).$$

From (6) we have $\sum_{n\leq x} 1 = x + O(1)$. It will thus suffice to show that

$$\sum_{d\leq x}\mu(d)\sum_{m\leq x/d}(\tau(m) - \log m) = o(x).$$

Let $\varepsilon > 0$ be a small fixed quantity. Arguing as in the implication of (iii) from (ii), we see from (ii) that

$$\sum_{\varepsilon x < d\leq x}\mu(d)\sum_{m\leq x/d}(\tau(m) - \log m) = o(x). \tag{64}$$

Next, we see from (8), (44) that

$$\sum_{m\leq x/d}(\tau(m) - \log m) = 2\gamma\frac{x}{d} + O(\sqrt{x/d}).$$

From Lemma 2 we have

$$\sum_{d\leq \varepsilon x}\sqrt{x/d} \ll \varepsilon^{1/2}x$$

and so from (iii) and (56) we conclude that

$$\sum_{d\leq \varepsilon x}\mu(d)\sum_{m\leq x/d}(\tau(m) - \log m) = O(\varepsilon^{1/2}x) + o(x).$$

Summing this with (64) and then sending ε to zero, we obtain the claim. □

Exercise 59. *(Further reformulations of the prime number theorem) Show that the statements (i)-(iii) in the above theorem are also equivalent to the following statements:*

(iv) *The number of primes less than or equal to x is $(1+o(1))\frac{x}{\log x}$ as $x\to\infty$.*

(v) *The n^{th} prime number p_n is equal to $(1+o(1))n\log n$ as $n\to\infty$.*

Unfortunately it is not so easy to actually obtain the required cancellation of the Möbius function μ, and to obtain the desired asymptotics for Λ. However, one can do better if one works with the higher-order von Mangoldt functions $\Lambda_2, \Lambda_3, \ldots$, defined by setting

$$\Lambda_k := \mu * L^k \tag{65}$$

for all $k \geq 1$. Thus Λ_1 is the usual von Mangoldt function, and from (61), (63) we easily obtain the recursive identity

$$\Lambda_{k+1} = L\Lambda_k + \Lambda_k * \Lambda \tag{66}$$

for $k \geq 1$. Among other things, this implies by induction that the Λ_k are non-negative, and are supported on those natural numbers that have at most k distinct prime factors. We have the following asymptotic for the summatory functions of Λ_k:

Proposition 60. *(Summatory function of Λ_k) For any $k \geq 1$, we have*

$$\sum_{n \leq x} \Lambda_k(n) = xR_k(\log x) + O_k(x)$$

for all $x \geq 1$, and for some polynomial $R_k(t)$ of leading term kt^{k-1}.

For $k = 1$, the error term here is comparable to the main term, and we obtain no improvement over the Chebyshev bound (Proposition 18). However, the estimates here become more useful for $k \geq 2$. For an explicit formula for the polynomials R_k, together with sharper bounds on the error term, see *this paper of Balazard*.

Proof. From (65), (35) we have

$$\sum_{n \leq x} \Lambda_k(n) = \sum_{d \leq x} \mu(d) \sum_{m \leq x/d} \log^k m.$$

By (9) we have

$$\sum_{m \leq x/d} \log^k m = \frac{x}{d} P_k\left(\log \frac{x}{d}\right) + O_k(1 + \log^k \frac{x}{d})$$

for some polynomial $P_k(t)$ with leading term t^k. The claim then follows from (58), using (56), (10) to control the error term. □

The $k = 2$ case of this proposition is known as the *Selberg symmetry formula*:

$$\sum_{n \leq x} \Lambda_2(n) = 2x \log x + O(x). \tag{67}$$

Among other things, this gives an upper bound that comes within a factor of two of the prime number theorem:

Corollary 61. *(Cheap Brun-Titchmarsh theorem) For any $1 \leq y \leq x$, one has*

$$\sum_{y \leq n \leq x} \Lambda(n) \leq 2(x - y) + O\left(\frac{x}{\log x}\right).$$

Using the methods of sieve theory, we will obtain a stronger inequality, known as the *Brun-Titchmarsh inequality*, in later notes. This loss of a factor of two reflects a basic problem in analytic prime number theory known as the *parity problem*: estimates which involve only primes (or more generally, numbers whose number of prime factors has a fixed parity) often lose a factor of two in their upper bounds, and are trivial with regards to lower bounds, unless some non-trivial input about prime numbers is somehow injected into the argument. We will discuss the parity problem in more detail in later notes.

Proof. From the Selberg symmetry formula we have

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

(since $y \log \frac{x}{y} = O(y\frac{x}{y}) = O(x)$). From (66) we have the pointwise bound $\Lambda(n)\log n \leq \Lambda_2(n)$, thus

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

By Proposition 18 and dyadic decomposition we have

$$\sum_{n \leq x} \Lambda(n) \log \frac{x}{n} \ll x,$$

and the claim follows. □

With some additional argument of a "Fourier-analytic" flavour (or using arguments closely related to Fourier analysis, such as Tauberian theorems or Gelfand's theory of Banach algebras), one can use the Selberg symmetry formula to derive the prime number theorem; see for instance *these previous blog posts* for examples of this. However, in this course we will focus instead on the more traditional complex-analytic proof of the prime number theorem, which highlights an important connection between the distribution of the primes and the zeroes of the Riemann zeta function.

Exercise 62. *(Cheap Brun-Titchmarsh, again) Show that for any $1 \leq y \leq x$, the number of primes between y and x is at most*

$$2\frac{x-y}{\log x} + O(\frac{x}{\log^2 x}).$$

Exercise 63. *(Golomb identity) Let $a_1, \ldots, a_k$ be coprime natural numbers. Show that*

$$\Lambda(a_1)\ldots\Lambda(a_k) = \frac{(-1)^k}{k!}(L^k\mu * 1)(a_1 \ldots a_k),$$

an identity of Golomb.

Exercise 64. *(Diamond-Steinig identity) Let $k \geq 1$. Show that $\Lambda_{2k} + \Lambda_k * \Lambda_k$ can be expressed as a linear combination of convolutions of the form $\mu * \ldots * \mu * L^{a_1} * \ldots * L^{a_r}$, where μ appears k times and $a_1, \ldots, a_r$ are non-negative integers with $a_1 + \ldots + a_r = 2k$ and $r \leq k$. Identities of this form are due to* Diamond and Steinig.

7. Dirichlet Characters

Now we consider the following vaguely worded question: how many primes are there in a given congruence class $a\ (q)$? For instance, how many primes are there whose last digit is 7 (i.e. lie in 7 (10))?

If the congruence class $a\ (q)$ is not *primitive*, that is to say that a and q share a common factor, then clearly the answer is either zero or one, with the latter occurring if the greatest common divisor (a,q) of a and q is a prime p which is congruent to a modulo q. So the interesting case is when $a\ (q)$ is primitive, that is to say that it lies in the multiplicative group $(\mathbf{Z}/q\mathbf{Z})^\times$ of primitive congruence classes.

In this case, we have the fundamental *theorem of Dirichlet*:

Theorem 65. *(Dirichlet's theorem, Euclid form) Every primitive congruence class $a\ (q)$ contains infinitely many primes.*

For a small number of primitive congruence classes, such as $1\ (q)$ or $3\ (8)$, it is possible to prove Dirichlet's theorem by mimicking one of the elementary proofs of Euclid's theorem, but we do not know of a general way to do so; see *this paper of Keith Conrad* for some further discussion. For instance, there is no proof known that there are infinitely many primes that end in 7 that does not basically go through most of the machinery of Dirichlet's proof (in particular introducing the notion of a Dirichlet character). Indeed, it looks like the problem of finding a new proof of Dirichlet's theorem is an excellent test case for any proposed alternative approach to studying the primes that does not go through the standard approach of analytic number theory (cf. Remark 2 from *the announcement for this course*).

In fact, Dirichlet's arguments prove the following stronger statement, generalising Euler's theorem (Theorem 2 from *this set of notes*):

Theorem 66. *(Dirichlet's theorem, Euler form) Let $a\ (q)$ be a primitive residue class. Then the sum $\sum_{p=a\ (q)} \frac{1}{p}$ is divergent.*

There is a more quantitative form, analogous to Mertens' theorem:

Theorem 67. *(Dirichlet's theorem, Mertens form) Let $a\ (q)$ be a primitive residue class. Then one has*

$$\sum_{n\leq x} \frac{\Lambda(n)}{n} 1_{n=a\ (q)} = \frac{1}{\phi(q)} \log x + O_q(1)$$

for any $x \geq 1$, where the Euler totient function *$\phi(q)$ is defined as the order of the multiplicative group $(\mathbf{Z}/q\mathbf{Z})^\times$.*

Exercise 68. *Let $a\ (q)$ be a primitive residue class. Use Theorem 67 to show that*

$$\sum_{p=a\ (q): p\leq x} \frac{1}{p} = \frac{1}{\phi(q)} \log\log x + c_{a,q} + o(1)$$

as $x \to \infty$ for some quantity $c_{a,q}$, thus giving Theorem 66. (Hint: adapt the proof of Theorem 36.)

If one tries to adapt one of the above proofs of Mertens' theorem (or Euler's theorem) to this setting, one soon runs into the problem that the function $1_{n=a\ (q)}$ is not multiplicative: $1_{nm=a\ (q)} \neq 1_{n=a\ (q)} 1_{m=a\ (q)}$. To resolve this issue, Dirichlet used some Fourier analysis to express $1_{n=a\ (q)}$ in terms of completely multiplicative functions, known as Dirichlet characters.

We first quickly recall the Fourier analysis of finite abelian groups:

Theorem 69. *(Fourier transform for finite abelian groups) Let G be a finite abelian group (which can be written additively or multiplicatively). Define a* character *on G to be a homomorphism $\chi : G \to S^1$ to the unit circle $\{z \in \mathbf{C} : |z| = 1\}$ of the complex numbers, and let $\hat{G}$ be the set of characters. Then $|\hat{G}| = |G|$. Furthermore, given any function $f : G \to \mathbf{C}$, one has a Fourier decomposition*

$$f(x) = \sum_{\chi \in \hat{G}} \hat{f}(\chi)\chi(x)$$

for all $x \in G$, where the Fourier coefficients $\hat{f}(\chi)$ are given by the formula

$$\hat{f}(\chi) := \frac{1}{|G|} \sum_{x \in G} f(x)\overline{\chi(x)}.$$

Thus for instance one has

$$1_{x=a} = \frac{1}{|G|} \sum_{\chi \in \hat{G}} \chi(x)\overline{\chi(a)} \tag{68}$$

for all $x, a \in G$.

Proof. Let $\ell^2(G)$ be the $|G|$-dimensional complex Hilbert space of functions $f : G \to \mathbf{C}$ with inner product $\langle f, g \rangle := \frac{1}{|G|} \sum_{x \in G} f(x)\overline{g(x)}$. Clearly any character $\chi \in \hat{G}$ is a unit vector in this space. Furthermore, for any two characters $\chi, \chi' \in \hat{G}$, we may shift the x variable by any shift $h \in G$ and conclude that

$$\langle \chi, \chi' \rangle = \chi(h)\overline{\chi'}(h)\langle \chi, \chi' \rangle$$

for any $h \in G$; in particular, we see that if $\chi \neq \chi'$, then $\langle \chi, \chi' \rangle = 0$. Thus $\hat{G}$ is an orthonormal system in $\ell^2(G)$. To complete the proof of the theorem, it thus suffices to show that this orthonormal system is complete, that is to say that the characters span $\ell^2(G)$.

Each shift $h \in G$ generates a unitary shift operator $T_h : \ell^2(G) \to \ell^2(G)$, defined by setting $T_h f(x) := f(hx)$ (if the group G is written multiplicatively). These operators all commute with each other, so by the spectral theorem they may all be simultaneously diagonalised by an orthonormal basis of joint eigenvectors. It is easy to see that these eigenvectors are characters (up to scaling), and so the characters span $\ell^2(G)$ as required. □

See *this previous post* for a more detailed discussion of the Fourier transform on both finite and infinite abelian groups. We remark that an alternate way to establish that the characters of $\ell^2(G)$ span is to use the *classification of finite abelian groups*

to express G as the product of cyclic groups, at which point one can write down the characters explicitly.

Define a *Dirichlet character* of modulus q to be a function $\chi : \mathbf{N} \to \mathbf{C}$ of the form

$$\chi(n) = 1_{(n,q)=1}\tilde{\chi}(n \bmod q)$$

where $\tilde{\chi} : (\mathbf{Z}/q\mathbf{Z})^\times \to \mathbf{C}$ is a character of $(\mathbf{Z}/q\mathbf{Z})^\times$. Thus, for instance, we have the *principal character*

$$\chi_0(n) = 1_{(n,q)} = 1$$

of modulus q. Another important example of a Dirichlet character is the quadratic character χ_p to a prime modulus p, defined by setting $\chi_p(n)$ to be $+1$ when n is a non-zero quadratic residue modulo p, -1 if n is a quadratic residue modulo p, and zero if n is divisible by p. (There are quadratic characters to composite moduli as well, but one needs to define them using the *Kronecker symbol*.) One can also easily verify that the product of two Dirichlet characters is again a Dirichlet character (even if the characters were initially of different modulus).

A technical remark: we consider two Dirichlet characters χ, χ' to be equal if they are equal as functions, that is to say that $\chi(n) = \chi'(n)$ for all n. As such, it is possible for a Dirichlet character to have multiple moduli: if χ is a character of modulus q, it is also a character of modulus kq for any natural number k. As such, it is slightly inaccurate to talk about "the" modulus of a Dirichlet character (though one could always work with the *minimal* modulus of a Dirichlet character, if desired), but this ambiguity will not cause much difficulty in practice.

Dirichlet characters χ of modulus q are completely multiplicative (thus $\chi(nm) = \chi(n)\chi(m)$ for all $n, m \in \mathbf{N}$, not necessarily coprime) and periodic of period q (thus $\chi(n+q) = \chi(n)$ for all $n \in \mathbf{Z}$). From Theorem 3 we see that there are exactly $\phi(q)$ Dirichlet characters of modulus q, and from (68) one has the Fourier inversion formula

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

From Mertens' theorem we have

$$\sum_{n \leq x} \frac{\Lambda(n)}{n} \chi_0(n) = \log x + O_q(1),$$

since the contribution of those n for which n is not coprime to q is easily seen to be $O_q(1)$ (n must be a power of a prime dividing q in these cases). Thus, Theorem 67 follows from

Theorem 70. *(Dirichlet's theorem, character form) Let χ be a non-principal Dirichlet character of modulus q. Then*

$$\sum_{n \leq x} \frac{\Lambda(n)\chi(n)}{n} = O_\chi(1)$$

for any $x \geq 1$.

To prove this theorem, we use the "deconvolution" strategy. Observe that for any completely multiplicative function χ, one has

$$\chi(f * g) = (\chi f) * (\chi g)$$

for any arithmetic functions f, g. In particular, from (45) one has

$$L\chi = (\Lambda\chi) * \chi. \tag{69}$$

Theorem 70 is seeking control on the logarithmic sums of $\Lambda\chi$, so it is natural to first control the logarithmic sums of $L\chi$ and χ. To do this we use a somewhat crude lemma (cf. Lemma 2):

Lemma 71. *Let χ be a non-principal character of modulus q, and let f be an arithmetic function that is monotone on an interval $[x, y]$. Then*

$$\sum_{x \leq n < y} f(n)\chi(n) = O(q(|f(x)| + |f(y)|)).$$

Proof. Without loss of generality we may assume that f is monotone non-increasing. By rounding x up and y down to the nearest multiple of q, we may assume that x, y are multiples of q, then the left-hand side may be split into the sum of expressions of the form $\sum_{jq \leq n < (j+1)q} f(n)\chi(n)$. As χ is non-principal, it is orthogonal to the principal character, and in particular $\sum_{jq \leq n < (j+1)q} \chi(n) = 0$. Thus we may write $\sum_{jq \leq n < (j+1)q} f(n)\chi(n)$ as $\sum_{jq \leq n < (j+1)q} (f(n) - f(jq))\chi(n)$, which by the trivial bound $|\chi(n)| \leq 1$ and monotonicity may be bounded by $O(q(f(jq) - f((j+1)q)))$. The claim then follows from telescoping series. □

From this lemma we obtain the crude upper bounds

$$\sum_{n \leq x} \frac{\chi(n) \log n}{n} = O_q(1) \tag{70}$$

and

$$\sum_{x \leq n < y} \frac{\chi(n)}{n} = O_q\left(\frac{1}{x}\right) \tag{71}$$

for any $1 \leq x \leq y$. (Strictly speaking, the function $x \mapsto \frac{\log x}{x}$ is not monotone decreasing for $x \leq e$, but clearly we may just delete this portion of the sum from (70) without significantly affecting the estimate.) By Lemma 5 we thus have

$$\sum_{n \leq x} \frac{\chi(n)}{n} = L(1, \chi) + O_q\left(\frac{1}{x}\right) \tag{72}$$

for all $x \geq 1$ and some complex number $L(1, \chi)$.

Exercise 72. *(Continuity of L-function at 1) Let χ be a non-principal character. For any $s > 1$, define the* Dirichlet L-function $L(s, \chi)$ *by the formula*

$$L(s, \chi) := \sum_n \frac{\chi(n)}{n^s}.$$

Show that $L(s,\chi) = L(1,\chi) + O_q(s-1)$ for any $s > 1$. In particular, the Dirichlet L-function extends continuously to $s = 1$. (In later notes we will extend this function to a much larger domain.)

We will shortly prove the following fundamental fact:

Theorem 73. *(Non-vanishing) One has $L(1,\chi) \neq 0$ for any non-principal character χ.*

Let us assume this theorem for now and conclude the proof of Theorem 70. Starting with the identity (69) and using (36), we see that

$$\sum_{n \leq x} \frac{\chi(n)\log n}{n} = \sum_{d \leq x} \frac{\Lambda(d)\chi(d)}{d} \left(\sum_{m \leq x/d} \frac{\chi(m)}{m} \right).$$

Inserting (70), (72) and using the trivial bound $|\chi(d)| \leq 1$ to control error terms, we conclude that

$$L(1,\chi) \sum_{d \leq x} \frac{\Lambda(d)\chi(d)}{d} = O_q(1) + O_q\left(\sum_{d \leq x} \frac{\Lambda(d)}{d} \frac{1}{x/d} \right),$$

and Theorem 70 follows by dividing by $L(1,\chi)$ and using Proposition 18.

Remark 74. *It is important to observe that this argument is potentially* ineffective*: the implied constant in Theorem 70 will depend on what upper bound one can obtain for the quantity $\frac{1}{|L(1,\chi)|}$. Theorem 73 ensures that this quantity is finite, but does not directly supply a bound for it, and so we cannot explicitly (or* effectively*) describe what the implied constant is as a computable function of q, at least if one only uses Theorem 73 as a "black box". It is thus of interest to strengthen Theorem 73 by obtaining effective lower bounds on $|L(1,\chi)|$ for various characters χ. This can be done in some cases (particularly if χ is not real-valued), but to get a good effective bound for all characters χ is a surprisingly difficult problem, essentially the* Siegel zero problem*; we will return to this issue in later notes.*

Exercise 75. *Show that*

$$\sum_{n \leq x} \Lambda_k(n)\chi(n) = O_{k,\chi}(x)$$

for any $k \geq 1$ and any non-principal character χ. (You will not need to know the non-vanishing of $L(1,\chi)$ to establish this.) Conclude that

$$\sum_{n \leq x : n = a \ (q)} \Lambda_k(n) = \frac{1}{\phi(q)} x R_k(\log x) + O_{k,q}(x)$$

for any $k \geq 1$ and any primitive residue class a (q), where R_k is the polynomial in Proposition 60. Deduce in particular the cheap Brun-Titchmarsh bound

$$\sum_{y \leq n \leq x : n = a \ (q)} \Lambda(n) \leq \frac{2}{\phi(q)}(x-y) + O_q\left(\frac{x}{\log x} \right)$$

for any $1 \leq y \leq x$ and primitive residue class a (q).

It remains to prove the non-vanishing of $L(1,\chi)$. Here we encounter a curious repulsion phenomenon (a special case of the *Deuring-Heilbronn repulsion phenonemon*): the vanishing of $L(1,\chi)$ for one character χ prevents (or "repels") the vanishing of $L(1,\chi')$ for another character χ'. More precisely, we have

Proposition 76. *Let $q \geq 1$. Then there is at most one non-principal Dirichlet character χ of modulus q for which $L(1,\chi) = 0$.*

Proof. Let $\chi_0, \chi_1, \ldots, \chi_{\phi(q)-1}$ denote all the Dirichlet characters of modulus q, including the principal character χ_0. The idea is to exploit a certain positivity when all the characters are combined together, which will be incompatible with two or more of the $L(1,\chi_i)$ vanishing.

There are a number of ways to see the positivity, but we will start with the Euler product identity

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

from (22). We can "twist" this identity by replacing $\frac{1}{p^s}$ by $\frac{\chi(p)}{p^s}$ for any Dirichlet character χ, which by the complete multiplicativity of χ gives

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

for any $s > 1$, where we allow for logarithms to be ambiguous up to multiples of $2\pi i$. By Taylor expansion, we thus have

$$\sum_n \frac{\Lambda(n)\chi(n)}{n^s \log n} = \log L(s,\chi)$$

for $s > 1$ (cf. (28)). Summing this for $\chi = \chi_0, \ldots, \chi_{\phi(q)-1}$, we have

$$\sum_n \frac{\Lambda(n) \sum_{j=0}^{\phi(q)-1} \chi_j(n)}{n^s \log n} = \sum_{j=0}^{\phi(q)-1} \log L(s,\chi_j).$$

From (68) we see that $\sum_{j=0}^{\phi(q)-1} \chi_j(n) = \phi(q) 1_{n=1\ (q)}$. In particular, the left-hand side is non-negative. Exponentiating, we conclude the lower bound

$$\prod_{j=0}^{\phi(q)-1} L(s,\chi_j) \geq 1. \tag{73}$$

Now we let $s \to 1^+$. For non-principal characters χ_j, we see from Exercise 72 that $L(s,\chi_j)$ stays bounded as $s \to 1^+$, and decays like $O_q(s-1)$ if $L(1,\chi_j)$ vanishes. For the principal character χ_0, we will just use the crude upper bound $|L(s,\chi_0)| \leq \zeta(s)$. By (11), we conclude that if two or more $L(1,\chi_j)$ are vanishing, then the product $\prod_{j=0}^{\phi(q)-1} L(s,\chi_j)$ will go to zero as $s \to 1$, contradicting (73), and the claim follows. □

Call a Dirichlet character χ *real* if it only takes real values, and *complex* if it is not real. For instance, the character of modulus 5 that takes the values i on

2 (5), -1 on 4 (5), $-i$ on 3 (5), 1 on 1 (5), and 0 on 0 (5) is a complex character. The above theorem, together with conjugation symmetry, quickly disposes of the complex characters χ, as such characters can be "repelled" by their complex conjugates:

Corollary 77. *Let χ be a complex character. Then $L(1,\chi) \neq 0$.*

Proof. If χ is a complex character of some modulus q, then its complex conjugate $\overline{\chi}$ is a different complex character with the same modulus q, and $L(1,\overline{\chi}) = \overline{L(1,\chi)}$. If $L(1,\chi)$ vanishes, we therefore have at least two non-principal characters of modulus q whose L-function vanishes at 1, contradicting Theorem 73. □

This only leaves the case of real non-principal characters χ to deal with. These characters are also known as *quadratic characters*, as χ^2 is the principal character; they are also connected to quadratic number fields, as we will discuss in a subsequent post. In this case, we cannot exploit the repulsion phenomenon, as we now only have one character for which $L(1,\chi)$ vanishes. On the other hand, for quadratic characters we have a much simpler positivity property, namely that

$$1 + \chi(n) \geq 0 \tag{74}$$

for all natural numbers n. Actually, it is convenient to use a variant of this positivity property, namely that

$$1 * \chi(n) \geq 0, \tag{75}$$

which can be proven first by working in the case that n is a power of a prime p and using (74), and then using multiplicativity to handle the general case. Crucially, we can do a little better than this: we can improve (75) to

$$1 * \chi(n) \geq 1 \tag{76}$$

whenever n is a perfect square. Again, this can be verified by first working in the case when n is an even prime power.

It is now natural to consider sums such as $\sum_{n\leq x} \frac{1*\chi(n)}{n^s}$ to exploit this positivity. It turns out that the best choice of s to use here is $s = 1/2$, that is to say to control the sum

$$\sum_{n\leq x} \frac{1 * \chi(n)}{\sqrt{n}}. \tag{77}$$

On the one hand, from positivity on the squares (76), we can bound this sum by

$$\sum_{m\leq\sqrt{x}} \frac{1}{\sqrt{m^2}} \gg \log x$$

for $x \geq 2$ (say), thanks to (13). On the other hand, we can expand (77) using the Dirichlet hyperbola method (cf. (43)) as

$$\sum_{d\leq\sqrt{x}} \frac{\chi(d)}{\sqrt{d}} \sum_{m\leq x/d} \frac{1}{\sqrt{m}} + \sum_{m\leq\sqrt{x}} \frac{1}{\sqrt{m}} \sum_{\sqrt{x}<d\leq x/m} \frac{\chi(d)}{\sqrt{d}}.$$

From (12) one has

$$\sum_{m \leq x/d} \frac{1}{\sqrt{m}} = 2\sqrt{x/d} + \zeta(1/2) + O(\sqrt{d/x})$$

while from Lemma 71 we have

$$\sum_{\sqrt{x} < d \leq x/m} \frac{\chi(d)}{\sqrt{d}} = O_q(\frac{1}{x^{1/4}})$$

and so (using the trivial bound $|\chi(d)| \leq 1$ to control error terms) the previous expression can be rewritten as

$$2\sqrt{x} \sum_{d \leq \sqrt{x}} \frac{\chi(d)}{d} + \zeta(1/2) \sum_{d \leq \sqrt{x}} \frac{\chi(d)}{\sqrt{d}} + O(1) + O_q(\frac{1}{\sqrt{x}} \sum_{m \leq \sqrt{x}} \frac{1}{m^{1/2} x^{1/4}}).$$

The final error is $O_q(1)$. From Lemma 71 we have $\sum_{d \leq x} \frac{\chi(d)}{\sqrt{d}} = O_q(1)$. Inserting (72), we conclude that

$$\log x \ll 2x^{1/2} L(1,\chi) + O_q(1). \tag{78}$$

If $L(1,\chi) = 0$ vanishes, then this leads to a contradiction if x is large enough. This concludes the proof of Theorem 73, and hence Dirichlet's theorem.

Remark 78. *The inequality (78) in fact shows that $L(1,\chi)$ is positive for every real character χ. In fact, with the assistance of some algebraic number theory, one can show the* class number formula *which asserts (roughly speaking) that $L(1,\chi)$ is proportional to the class number of a certain quadratic number field. This will be discussed in a subsequent post.*

Exercise 79. *By using an effective version of the above arguments, establish the lower bound*

$$|L(1,\chi)| \gg \exp(-q^{O(1)}) \tag{79}$$

for all non-principal characters χ of modulus q (both real and complex).

Remark 80. *The bound (79) is very poor and can be improved. For instance, the class number formula alluded to in the previous remark gives the effective bound $L(1,\chi) \gg 1/\sqrt{q}$ for real non-principal characters. In later notes we will also establish* Siegel's theorem*, which gives an* ineffective *bound of the form $L(1,\chi) \gg_\varepsilon q^{-\varepsilon}$ for such characters, and any $\varepsilon > 0$.*

Exercise 81. *Let χ be a non-principal character. Show that the sum $\sum_p \frac{\chi(p)}{p}$ is conditionally convergent. Then show that the product $\prod_p (1 - \frac{\chi(p)}{p})^{-1}$ is conditionally convergent to $L(1,\chi)$.*

Exercise 82. *Show that*

$$\sum_{n \leq x} \phi(n) = \frac{1}{2\zeta(2)} x^2 + O(x^2/\log x)$$

for any $x > 2$. Conclude that the proportion of pairs (n, m) of natural numbers in the square $\{(n, m) \in \mathbf{N}^2 : n, m \leq x\}$ which are coprime converges to $\frac{1}{\zeta(2)} = \frac{6}{\pi^2}$ as $x \to \infty$.