Analytic Number Theory

Robin Truax

January 18, 2023

Contents

Thanks to Lynnelle Ye for teaching Math 155, from which these notes are derived, Keith Conrad for his analytic number theory notes, and fellow classmates for tips and corrections when I needed them.

1 An Introduction to Estimation

1.1 Review of Asymptotic Notation

Definition 1.1 (Big-O Notation)**.** We say that $f(x) = O(g(x))$ or $f(x) \ll g(x)$ if there exists a constant $C > 0$ such that $|f(x)| \leq Cg(x)$ for all x from 1 to infinity. Also, $f(x) \asymp g(x)$ means that $f(x) \ll g(x)$ and $f(x) \gg g(x)$.

Definition 1.2 (Little-O Notation)**.** We say that $f(x) = o(g(x))$ if for all $\varepsilon > 0$, there exists N such that $|f(x)| \leq \epsilon g(x)$ for all $x \geq N$.

Definition 1.3 (Asymptotic Equivalence)**.** We say that $f \sim g$ when $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 1$.

Notice that $f \sim g$ implies $f \asymp g$, but the converse is not true in general (consider, for example, $f(x) = x$ and $g(x) = 2x$).

1.2 Partial Summation

It will be convenient soon to have the following definitions:

Definition 1.4 (Floor Function)**.** For $x \in \mathbb{R}$, let $\lfloor x \rfloor$ be the greatest integer less than or equal to x (this is called the *integer part* of x). Also let $\{x\} = x - \lfloor x \rfloor$ (this is called the *fractional part* of x).

Let us begin with an elementary result to illustrate the general theory of asymptotic estimation:

Proposition 1.5. *There exists a constant C such that*

$$\sum_{n\leq N} \frac{1}{n} = \log(N) + C + O\left(\frac{1}{N}\right)$$

The rough idea here is that $\log(n)$ is the antiderivative of $\frac{1}{n}$, and the summation on the left looks like a Riemann sum of $\frac{1}{n}$. However, to make this more precise, we are going to develop and use the theory of partial summation developed by Abel. The idea of partial summation is as follows:

Suppose that $a(n)$ is any function and define $A(x) = \sum_{n\leq x} a(n)$. Furthermore suppose that we have an estimate for $A(x)$, but we would like to find a related sum, such as

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

for some function $f(n)$. Now, the final case is the most general (indeed it subsumes the first two cases), so we will focus on it. We begin with some easy algebraic manipulation:

$$\begin{aligned}\sum_{n=1}^{N} a(n)f(n) &= \sum_{n=1}^{N} f(n)(A(n) - A(n-1)) = \sum_{n=1}^{N} f(n)A(n) - \sum_{n=1}^{N} f(n)A(n-1)\\ &= \sum_{n=1}^{N} f(n)A(n) - \sum_{n=0}^{N-1} f(n+1)A(n) = A(N)f(N) - A(0)f(1) - \sum_{n=1}^{N-1} A(n)(f(n+1) - f(n)).\end{aligned}$$

Recall that we have an estimate for $A(x)$, so this sum has a good chance to be calculable. Now, a better way to think of this strategy (which allows us to use the power of integration) is to think of this as integration by parts. The key step here is the second equality, which is made precise with the Riemann-Stieltjes integral:

$$\sum_{n=1}^{N} a(n)f(n) = \sum_{n=1}^{N} f(n)(A(n) - A(n-1)) = \int_{1^-}^{N^+} f(t)d(A(t)) = f(t)A(t)\Big|_{1^-}^{N^+} - \int_{1^-}^{N^+} A(t)d(f(t))$$

Now, we are equipped to prove the result displayed above:

Proof. Suppose that, borrowing the notation above, $a(n) = 1$ and $f(n) = \frac{1}{n}$. Then $A(x) = \lfloor x \rfloor$, so that

$$\sum_{n \leq N} \frac{1}{n} = \sum_{n=1}^{N} a(n)f(n) = \int_{1^-}^{N^+} \frac{1}{t} d(\lfloor t \rfloor) = \frac{\lfloor t \rfloor}{t}\bigg|_{1^-}^{N^+} - \int_{1^-}^{N^+} \lfloor t \rfloor \, d\left(\frac{1}{t}\right) = 1 - 0 + \int_{1^-}^{N^+} \frac{t - \{t\}}{t^2} dt.$$

Then,

$$\int_{1^-}^{N^+} \frac{t - \{t\}}{t^2} dt = \int_1^N \frac{1}{t} dt - \int_1^N \frac{\{t\}}{t^2} dt = \log(N) - \int_1^N \frac{\{t\}}{t^2} dt.$$

Now,

$$\int_1^N \frac{\{t\}}{t^2} dt = \int_1^\infty \frac{\{t\}}{t^2} - \int_N^\infty \frac{\{t\}}{t^2}.$$

The former part is a small constant C' bounded above by $\int_1^\infty \frac{1}{t^2} dt < \infty$, and the latter part is

$$\int_N^\infty \frac{\{t\}}{t^2} = O\left(\int_N^\infty \frac{1}{t^2} dt\right) = O\left(\frac{1}{N}\right).$$

In summary, it follows that $\sum_{n \leq N} \frac{1}{n} = 1 - C' + \log(N) + O\left(\frac{1}{N}\right) = \log(N) + C + O\left(\frac{1}{N}\right)$. □

Let us do another example, to do with finding an asymptotic expression for the factorial of N. It is plainly obvious that $c^N \ll N! \ll N^N$ for any constant c. Let us find a more precise approximation:

Proposition 1.6 (Stirling's Formula)**.**

$$N! \asymp \sqrt{N} \left(\frac{N}{e}\right)^N.$$

Proof. The tactic is to estimate $\log(N!) = \sum_{n \leq N} \log(n)$. Again, we will use $a(n) = 1$; then, $f(n) = \log(n)$. As before, $A(x) = \lfloor x \rfloor$, so that

$$\sum_{n \leq N} \log(n) = \sum_{n=1}^{N} a(n)f(n) = \int_{1^-}^{N^+} \log(t) d \lfloor t \rfloor = \lfloor t \rfloor \log(t)\bigg|_{1^-}^{N^+} - \int_{1^-}^{N^+} \frac{\lfloor t \rfloor}{t} dt = N \log N - N + 1 + \int_1^N \frac{\{t\}}{t} dt$$

where the final equality follows from expressing $\lfloor t \rfloor = t - \{t\}$ and simplifying as possible. Then,

$$\int_1^N \frac{\{t\}}{t} dt = \int_1^N \frac{1/2}{t} dt + \int_1^N \frac{\{t\} - 1/2}{t} dt = \frac{1}{2} \log N + \int_1^N \frac{\{t\} - 1/2}{t} dt.$$

Define, for simplicity, $B(y) = \{y\} - \frac{1}{2}$. Then, rearranging and using integration by parts:

$$\int_1^N \frac{B(t)}{t} dt = \int_1^N \frac{1}{t} d\left(\int_1^t B(y) dy\right) = \frac{1}{t} \int_1^t B(y) dy \bigg|_1^N + \int_1^N \int_1^t B(y) dy \frac{dt}{t^2} = \int_1^N \int_1^t B(y) dy \frac{dt}{t^2}$$

as the first part vanishes. Furthermore,

$$\int_1^t B(y) dy = \int_{\lfloor t \rfloor}^t B(y) dy = \int_0^{\{t\}} B(y) dy = \int_0^{\{t\}} \left(y - \frac{1}{2}\right) dy = \frac{\{t\}^2 - \{t\}}{2}$$

It is not hard to see that this is bounded between 0 and $-\frac{1}{8}$ for any t. Hence, by using the same tactic of going up to infinity and subtracting off the tail, we may show that

$$\int_1^N \frac{\{t\} - 1/2}{t} dt = \int_1^N \int_1^t B(y) dy \frac{dt}{t^2} = \int_1^N \frac{\{t\}^2 - \{t\}}{2t^2} dt = C' + O\left(\frac{1}{N}\right)$$

for a constant C'. Summing everything and collapsing all constant terms into a single constant C yields

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

Hence, by exponentiating, we find that

$$N! \asymp \sqrt{N}\left(\frac{N}{e}\right)^N .$$

□

Note that the expression for $\sum_{n\leq N}\log(n) = N\log(N) - N + \frac{1}{2}\log(N) + C + O\left(\frac{1}{N}\right)$ is actually slightly more precise, in the sense that it gives us more information than the final expression. Also note that a more precise version of this statement with the correct constant is as follows (we do not prove this):

$$N! \sim \sqrt{2\pi N}\left(\frac{N}{e}\right)^N .$$

1.3 Euler-Maclaurin Summation

The Euler-Maclaurin summation formula is a general method for making partial summation results more precise. The theory is very dense, and we will not use it often, but for completeness we include it.

Definition 1.7 (Bernoulli Polynomials)**.** For $k \in \mathbb{Z}_{\geq 0}$, let the Bernoulli polynomials $B_k(x)$ be defined inductively by $B_0(x) = 1$,

$$\frac{d}{dx}B_k(x) = kB_{k-1}(x)$$

and

$$\int_0^1 B_k(x)dx = 0.$$

Proposition 1.8 (Computing Small Bernoulli Polynomials)**.** *$B_1(x) = x - \frac{1}{2}$, $B_2(x) = x^2 - x + \frac{1}{6}$, and $B_3(x) = x^3 - \frac{3}{2}x + \frac{1}{2}x$.*

Proof.

1. $B_1(x) = x - \frac{1}{2}$ as $\frac{d}{dx}\left(x - \frac{1}{2}\right) = 1 = 1 \cdot B_0(x)$ and $\int_0^1 \left(x - \frac{1}{2}\right)dx = 0$.
2. $B_2(x) = x^2 - x + \frac{1}{6}$ as $\frac{d}{dx}\left(x^2 - x + \frac{1}{6}\right) = 2x - 1 = 2 \cdot B_1(x)$ and $\int_0^1 \left(x^2 - x + \frac{1}{6}\right)dx = 0$.
3. $B_3(x) = x^3 - \frac{3}{2}x + \frac{1}{2}x$ as $\frac{d}{dx}\left(x^3 - \frac{3}{2}x + \frac{1}{2}x\right) = 3x^2 - 3x + \frac{1}{2} = 3 \cdot B_2(x)$ and $\int_0^1 \left(x^3 - \frac{3}{2}x + \frac{1}{2}x\right)dx = 0$.

□

Theorem 1.9 (Euler-Maclaurin Summation Formula)**.** *For all $K \geq 1$,*

$$\sum_{a<n\leq b} f(n) = \int_a^b f(x)dx + \sum_{k=1}^{K}\frac{(-1)^k}{k!}(B_k(\{b\})f^{(k-1)}(b) - B_k(\{a\})f^{(k-1)}(a)) - \frac{(-1)^K}{K!}\int_a^b B_K(\{x\})f^{(K)}(x)dx.$$

We begin with a lemma that allows us to complete the inductive step:

Lemma 1.10. *For any function f,*

$$\frac{(-1)^K}{K!}(B_K(\{b\})f^{(k-1)}(b) - B_K(\{a\})f^{(k-1)}(a)) - \\ \frac{(-1)^K}{K!}\int_a^b B_K(\{x\})f^{(K)}(x)dx + \frac{(-1)^{K-1}}{(K-1)!}\int_a^b B_{K-1}(\{x\})f^{(K-1)}(x)dx = 0$$

Proof. By partial integration,

$$\begin{aligned}\int_a^b B_K(\{x\})f^{(k)}(x)dx &= B_K(\{b\})f^{(k-1)}(b) - B_K(\{a\})f^{(k-1)}(a) - \int_a^b B_K'(\{x\})f^{(k-1)}(x)dx \\ &= B_K(\{b\})f^{(k-1)}(b) - B_K(\{a\})f^{(k-1)}(a) - K\int_a^b B_{K-1}(\{x\})f^{(k-1)}(x)dx\end{aligned}$$

Then, by multiplying throughout by $\frac{(-1)^K}{K!}$ and rearranging everything onto one side, we obtain

$$\frac{(-1)^K}{K!}(B_K(\{b\})f^{(k-1)}(b) - B_K(\{a\})f^{(k-1)}(a)) -$$
$$\frac{(-1)^K}{K!}\int_a^b B_K(\{x\})f^{(K)}(x)dx + \frac{(-1)^{K-1}}{(K-1)!}\int_a^b B_{K-1}(\{x\})f^{(K-1)}(x)dx = 0$$

which is the desired result. □

Now, we can begin the main proof:

Proof. First, notice that if c is an integer less than a, b, then $\sum_{a<n\leq b} f(n) = \sum_{n=c}^{\lfloor b \rfloor} f(n) - \sum_{n=c}^{\lfloor a \rfloor} f(n)$. Then, we will apply partial summation to decompose the first half of this expression:

$$\sum_{n=c}^{\lfloor b \rfloor} f(n) = f(t)\lfloor t \rfloor \Big|_{c^-}^{b^+} - \int_c^b \lfloor t \rfloor f'(t)dt = f(b)\lfloor b \rfloor - f(c)\lfloor c-1 \rfloor - \int_c^b \lfloor t \rfloor f'(t)dt.$$

Then, repeating this decomposition for the second half and subtracting, we find that

$$\sum_{a<n\leq b} f(n) = \sum_{n=c}^{\lfloor b \rfloor} f(n) - \sum_{n=c}^{\lfloor a \rfloor} f(n) = f(b)\lfloor b \rfloor - f(a)\lfloor a \rfloor - \int_a^b \lfloor t \rfloor f'(t)dt.$$

Then,

$$\int_a^b \lfloor t \rfloor f'(t)dt = \int_a^b tf'(t)dt - \int_a^b \{t\}f'(t)dt = bf(b) - af(a) - \int_a^b f(t)dt - \int_a^b \{t\}f'(t)dt.$$

Hence we have

$$\sum_{a<n\leq b} f(n) = \{a\}f(a) - \{b\}f(b) + \int_a^b f(t)dt + \int_a^b \{t\}f'(t)dt.$$

Now, we add $0 = \frac{1}{2}f(b) - \frac{1}{2}f(a) - \frac{1}{2}f(b) + \frac{1}{2}f(a) = \frac{1}{2}f(b) - \frac{1}{2}f(a) - \int_a^b \frac{1}{2}f'(t)dt$ to the above equation, getting

$$\begin{aligned}\sum_{a<n\leq b} f(n) &= \left(\{a\} - \frac{1}{2}\right)f(a) - \left(\{b\} - \frac{1}{2}\right)f(b) + \int_a^b f(t)dt + \int_a^b \left(\{t\} - \frac{1}{2}\right)f'(t)dt \\ &= \int_a^b f(x)dx + B_1(\{a\})f(a) - B_1(\{b\})f(b) + \int_a^b B_1(\{x\})f'(x)dx\end{aligned}$$

which is precisely the case of the Euler-Maclaurin summation formula for $K = 1$; that is, the base case.

Now we can perform the inductive step. Suppose the result holds for some $K-1$; that is,

$$\begin{aligned}\sum_{a<n\leq b} f(n) = \int_a^b f(x)dx &+ \sum_{k=1}^{K-1} \frac{(-1)^k}{k!}(B_k(\{b\})f^{(k-1)}(b) - B_k(\{a\})f^{(k-1)}(a)) \\ &- \frac{(-1)^{K-1}}{(K-1)!}\int_a^b B_{K-1}(\{x\})f^{(K-1)}(x)dx\end{aligned}$$

Then, adding 0 to both sides using the above lemma, we get that

$$\sum_{a<n\leq b} f(n) = \int_a^b f(x)dx + \sum_{k=1}^{K-1} \frac{(-1)^k}{k!}(B_k(\{b\})f^{(k-1)}(b) - B_k(\{a\})f^{(k-1)}(a))$$

$$-\frac{(-1)^{K-1}}{(K-1)!}\int_a^b B_{K-1}(\{x\})f^{(K-1)}(x)dx + \frac{(-1)^K}{K!}(B_K(\{b\})f^{(k-1)}(b) - B_K(\{a\})f^{(k-1)}(a)) -$$

$$\frac{(-1)^K}{K!}\int_a^b B_K(\{x\})f^{(K)}(x)dx + \frac{(-1)^{K-1}}{(K-1)!}\int_a^b B_{K-1}(\{x\})f^{(K-1)}(x)dx$$

$$= \int_a^b f(x)dx + \sum_{k=1}^{K} \frac{(-1)^k}{k!}(B_k(\{b\})f^{(k-1)}(b) - B_k(\{a\})f^{(k-1)}(a)) - \frac{(-1)^K}{K!}\int_a^b B_K(\{x\})f^{(K)}(x)dx,$$

which completes the inductive step. □

As an application of the above work, we give a more precise version of Stirling's approximation.

Proposition 1.11.

$$\sum_{n\leq N} \log n = N\log N - N + \frac{1}{2}\log N + C_1 + \frac{1}{12N} + O\left(\frac{1}{N^2}\right)$$

for some constant C_1.

Proof. Suppose that $f(n) = \log(n)$, so that $f'(n) = \frac{1}{n}$, $f''(n) = -\frac{1}{n^2}$, and $f^{(3)}(n) = \frac{2}{n^3}$. Then, in light of the above formula, $\sum_{n\leq N}\log(n) = \sum_{1<n\leq N}\log(n)$ is equal to the following in the case $K = 3$:

$$\int_1^N \log(x)dx + \sum_{k=1}^{3}\frac{(-1)^k}{k!}(B_k(\{N\})f^{(k-1)}(N) - B_k(\{1\})f^{(k-1)}(1)) + \frac{1}{6}\int_1^N \frac{2B_3(\{x\})}{x^3}dx$$

First, let us compute

$$\int_1^N \log(x)dx = (x\log(x) - x)\Big|_1^N = N\log(N) - N + 1.$$

Next, let us bound $\frac{1}{6}\int_1^N \frac{2B_3(\{x\})}{x^3}dx = \frac{1}{3}\int_1^N \frac{B_3(\{x\})}{x^3}dx$. For this, notice that obviously $B_3(\{x\}) \leq 1 + \frac{3}{2} + \frac{1}{2} = 3$, so $\int_1^\infty \frac{B_3(\{x\})}{x^3}dx$ converges to a constant C'. Then,

$$\int_1^N \frac{B_3(\{x\})}{x^3}dx = C' - \int_N^\infty \frac{B_3(\{x\})}{x^3}dx$$

and $\int_N^\infty \frac{B_3(\{x\})}{x^3}dx \leq \int_N^\infty \frac{3}{x^3}dx = \frac{-3}{2x^2}\Big|_N^\infty = \frac{3}{2N^2}$, whence $\int_1^N \frac{B_3(\{x\})}{x^3}dx = C' + O(\frac{1}{N^2})$ for some constant C'. Therefore, only the middle sum needs to be determined:

$$\sum_{k=1}^{3}\frac{(-1)^k}{k!}(B_k(\{N\})f^{(k-1)}(N) - B_k(\{1\})f^{(k-1)}(1)).$$

Let us simplify all three terms separately:

$$k = 1 \text{ term}: \quad (-1)(B_1(\{N\})f(N) - B_1(\{1\})f(1)) = (-1)((-1/2)\log(N) - (-1/2)0) = \frac{1}{2}\log N$$

$$k = 2 \text{ term}: \quad \frac{1}{2}(B_2(\{N\})f'(N) - B_2(\{N\})f'(1)) = \frac{1}{2}\left(\frac{1}{6}\cdot\frac{1}{N} - \frac{1}{6}\cdot\frac{1}{1}\right) = \frac{1}{12N} - \frac{1}{6}$$

$$k = 3 \text{ term}: \quad \frac{-1}{6}(B_3(\{N\})f''(N) - B_3(\{1\})f''(1)) = \frac{-1}{6}(0 - 0) = 0$$

Hence the middle sum is equal to

$$\frac{1}{2}\log N + \frac{1}{12N} - \frac{1}{6}.$$

Therefore, combining everything, we find that

$$\sum_{n\leq N}\log(n) = N\log(N) - N + \frac{1}{2}\log(N) + C + \frac{1}{12N} + O\left(\frac{1}{N^2}\right).$$

□

1.4 Estimating the Reciprocal of the Primes

In this section, we are going to work towards estimating

$$\sum_{\substack{p\leq N\\ p\text{ prime}}}\frac{1}{p}$$

Let us try to guess the answer using the prime number theorem (which we have not proven yet, and will prove much later in the course). Now, the prime number theorem states that $\pi(x)\sim\frac{x}{\log x}$, where $\pi(x)$ counts the number of primes less than x. Yet this implies that the nth prime is about $n\log n$ (verifying this is left as an exercise for the reader), so the above sum is about $\sum_{n\leq N}\frac{1}{n\log n}\sim\int_1^N\frac{1}{t\log t}dt\sim\log\log N$. However, without the prime number theorem, we will need to develop a few tools to prove this result.

Definition 1.12 (von Mangoldt function)**.** Define *the von Mangoldt function* to be $\Lambda:\mathbb{N}\to\mathbb{R}$ given by

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

It is not difficult to see that $\log n = \sum_{d|n}\Lambda(d)$.

Definition 1.13 (Second Chebyshev Function)**.** The *second Chebyshev function* is defined to be

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

This function is extremely important, and indeed its asymptotic behavior is related to the prime number theorem. Precisely, we will see later that the statement $\psi(x)\sim x$ is equivalent to the Prime Number Theorem. We will not be able to do this for a while, but we are able to show that $\psi(x)\asymp x$.

For this, we first need a combinatorial lemma:

Lemma 1.14.

$$\frac{4^N}{2N+1}\leq\binom{2N}{N}\leq 4^N$$

Proof. This follows from the fact that

$$\sum_{i=0}^{2N}\binom{2N}{i} = \sum_{i=0}^{2N}\binom{2N}{i}1^i1^{2N-i} = (1+1)^{2N} = 4^N,$$

so clearly $\binom{2N}{N}\leq 4^N$. Furthermore, since the middle binomial coefficient is the largest (which one can see by noticing that to get to binomial coefficients on each side, one multiplies by something less than 1), we must have that $\binom{2N}{N}$ is at least the average of $\binom{2N}{i}$ over all $0\leq i\leq 2N$. By the above sum, this average is $\frac{4^N}{2N+1}$. Hence $\frac{4^N}{2N+1}\leq\binom{2N}{N}$, so we are done. □

Theorem 1.15 (Chebyshev Bounds)**.** *There exist constants c and C such that $cx \leq \psi(x) \leq Cx$ for all x sufficiently large. More precisely, we have the following asymptotic behavior:*

$$(\log 2 + o(1))x \leq \psi(x) \leq (2\log 2 + o(1))x.$$

Proof. First, notice that

$$\log N! = \sum_{n \leq N} \log n = \sum_{n\leq N}\sum_{d|n} \Lambda(d) = \sum_{d \leq N}\sum_{\substack{n\leq N \\ n\in d\mathbb{Z}}} \Lambda(d) = \sum_{d\leq N}\left\lfloor \frac{N}{d}\right\rfloor \Lambda(d).$$

Now, $\log(\binom{2N}{N}) = \log((2N)!) - 2\log(N!)$ whence

$$\begin{aligned}\log\left(\binom{2N}{N}\right) &= \sum_{d\leq 2N}\Lambda(d)\left\lfloor\frac{2N}{d}\right\rfloor - 2\sum_{d\leq N}\Lambda(d)\left\lfloor\frac{N}{d}\right\rfloor \\ &= \sum_{d\leq 2N}\Lambda(d)\left\lfloor\frac{2N}{d}\right\rfloor - 2\sum_{d\leq N}\Lambda(d)\left\lfloor\frac{N}{d}\right\rfloor = \sum_{d\leq 2N}\Lambda(d)\left(\left\lfloor\frac{2N}{d}\right\rfloor - 2\left\lfloor\frac{N}{d}\right\rfloor\right).\end{aligned}$$

Now, $\left\lfloor\frac{2N}{d}\right\rfloor - 2\left\lfloor\frac{N}{d}\right\rfloor$ is either 0 or 1 for all d. Hence, we can achieve an upper bound for the sum by assuming that it is 1. Namely,

$$\log\left(\frac{4^N}{2N+1}\right) \leq \log\left(\binom{2N}{N}\right) \leq \sum_{d\leq 2N}\Lambda(d) = \psi(2N) \Rightarrow 2N\log(2) - \log(2N+1) \leq \psi(2N)$$

whence by replacing $2N$ with x we find that $x\log(2) - \log(x+1) \leq \psi(x)$ which, since $\frac{\log(x+1)}{x} = o(1)$, yields that $x(\log 2 + o(1)) \leq \psi(x)$, as desired. Therefore, we have the lower bound!

Now, the upper bound is slightly more complicated, but it is ultimately not too bad. The key step here is that when $N < d \leq 2N$, $\left\lfloor\frac{2N}{d}\right\rfloor - 2\left\lfloor\frac{N}{d}\right\rfloor = 1$. Hence

$$\log(4^N) \geq \log\binom{2N}{N} \geq \sum_{N<d\leq 2N}\Lambda(d) = \psi(2N) - \psi(N).$$

In short, we have $\psi(2N) - \psi(N) \leq 2N\log(2)$. Yet this implies that:

$$\begin{aligned}\psi(N) - \psi(N/2) &\leq N(\log 2 + o(1)) \\ \psi(N/2) - \psi(N/4) &\leq \frac{1}{2}N(\log 2 + o(1)) \\ &\vdots\end{aligned}$$

Here, the $o(1)$ terms are added because we accumulate small errors when N is not divisible by 2^k. Yet these errors are ultimately unimportant, because there are only $\log_2(N)$ of them. In other words, if we sum all these errors up, the sum telescopes and we are left with $\psi(N) \leq 2N(\log 2 + o(1)) = (2\log 2 + o(1))N$, which is the desired upper bound when we replace N by the variable x. Therefore we are done. □

Next, we show the following theorem:

Proposition 1.16.

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

Proof. First, let us see why the two sums are equal. The difference between these two sums is

$$\sum_{\substack{p \text{ prime} \\ k\geq 2 \\ p^k\leq N}}\frac{\log p}{p^k} \leq \sum_{p \text{ prime}}\sum_{k\geq 2}\frac{\log p}{p^k} = \sum_{p \text{ prime}}\frac{\log p}{p^2}\cdot\frac{1}{1-1/p} = \sum_{p \text{ prime}}\frac{\log p}{p(p-1)}.$$

Yet the latter expression is bounded by a constant:

$$\sum_{p \text{ prime}} \frac{\log p}{p(p-1)} = O\left(\sum_{p \text{ prime}} \frac{\log p}{p^2}\right) = O\left(\sum_{p \text{ prime}} \frac{1}{p^{1.5}}\right) = O\left(\sum_{n \text{ prime}} \frac{1}{n^{1.5}}\right) = O\left(\int_1^\infty t^{-1.5} dt\right) = O(1).$$

Hence it suffices to show the second equality. For this, recall from our work with Stirling's approximation that $\log(N!) = N\log N + O(N)$. On the other hand, in the preceding problem, we established that

$$\log(N!) = \sum_{d\le N} \Lambda(d) \left\lfloor \frac{N}{d} \right\rfloor = \sum_{d\le N} \Lambda(d) \left(\frac{N}{d} + O(1)\right) = N \sum_{d\le N} \frac{\Lambda(d)}{d} + O(\sum_{d\le N} \Lambda(d)).$$

Yet the result of the previous problem means precisely that the additional part $O(\sum_{d\le N} \Lambda(d))$ is $O(N)$. Hence we have $N \log N + O(N) = N \sum_{d\le N} \frac{\Lambda(d)}{d} + O(N)$. Dividing by N and isolating the sum, we get that $\sum_{d\le N} \frac{\Lambda(d)}{d} = \log N + O(1)$, which is the desired result. □

Now we have the tools to compute the sum of the reciprocals of the primes.

Theorem 1.17.

$$\sum_{p\le N} \frac{1}{p} = \log\log N + C + O\left(\frac{1}{\log(N)}\right)$$

Proof. We know from the preceding theorem that $A(x) = \sum_{p\le x} \frac{\log p}{p} = \log x + E(x)$ for some function $E(x) = O(1)$. Then, using partial summation with $a(n) = \frac{\log n}{n}$ when n is prime and 0 otherwise (so $A(x) = \sum_{p\le x} \frac{\log p}{p}$) and $f(n) = \frac{1}{\log n}$,

$$\sum_{p\le N} \frac{1}{p} = \int_{2^-}^{N^+} \frac{1}{\log t} d(A(t)) = \frac{A(N)}{\log N} - \int_2^N A(t) d\left(\frac{1}{\log t}\right) = 1 + O\left(\frac{1}{\log N}\right) + \int_2^N \frac{A(t)}{t(\log t)^2} dt.$$

It remains only to evaluate the integral which is the final term of the right-hand side. Indeed,

$$\int_2^N \frac{A(t)}{t(\log t)^2} dt = \int_2^N \frac{\log t}{t\log(t)^2} dt + \int_2^N \frac{E(t)}{t(\log t)^2} dt.$$

The former part is $\log\log N + \log\log 2$. For the latter, we use the usual trick: $\int_2^\infty \frac{1}{t(\log t)^2} dt$ converges, so

$$\int_2^N \frac{E(t)}{t(\log t)^2} dt = \int_2^\infty \frac{E(t)}{t(\log t)^2} dt - \int_N^\infty \frac{E(t)}{t(\log t)^2} dt = C' + O\left(\frac{1}{\log(N)}\right)$$

for some constant C'. Then, combining everything, we get the desired result. □

2 Riemann's ζ-Function and Dirichlet Series

2.1 The Riemann ζ-Function

Definition 2.1 (Riemann ζ-Function)**.** The *Riemann ζ-function* is equal to $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$ for all $s \in \mathbb{C}$ such that the sum converges. Precisely, this sum converges (indeed, it converges absolutely) iff $s \in \mathbb{C}$ has $\Re(s) > 1$ (that is, the real part of s is greater than 1). One first verifies this for real s, and then notices that changing s by an imaginary quantity does not change the magnitude of $\frac{1}{n^s}$.

Lemma 2.2. *The function* $\sin(z)$ *has no non-real roots.*

Proof. Recall that $\sin(z) = \frac{e^{iz}-e^{-iz}}{2i}$; hence $\sin(z) = 0$ if and only if $e^{iz} = e^{-iz}$; that is, if $e^{2iz} = 1$. Now write $z = x + iy$ for real x, y, so that we have $e^{2i(x+iy)} = 1$. But this means precisely that $e^{2ix-2y} = e^{-2y}e^{2ix} = 1$. Taking magnitudes, we find that $|e^{-2y}||e^{2ix}| = 1$ which implies $|e^{-2y}| = 1$ which implies $y = 0$, as desired. □

Proposition 2.3. $\zeta(2) = \sum_{n=1}^{\infty} \frac{1}{n^2}$.

Proof. We recount Euler's original (very non-rigorous) proof; it can be made rigorous using complex analysis (see Stein and Shakarchi, Chapter 5), but we do not do that here. Recall that $\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \cdots$ for all $x \in \mathbb{C}$. The real roots of this function are $0, \pm\pi, \pm 2\pi, \ldots$, and indeed it is easy to check that all of these roots are simple. Namely, recall that if a root r of an analytic function f has multiplicity greater than 1, then f' has a root at r. Yet $\sin'(x) = \cos(x)$, and $\cos(x)$ has no roots at multiples of π. Indeed, these are all the roots of $\sin(x)$, as Lemma 2.2 shows.

Hence the roots of $\frac{\sin x}{x}$ are (including multiplicity), $n\pi$ for $n \neq 0$. Now, if $p(x)$ is a polynomial with constant coefficient 1 and roots $r_1, \ldots, r_d$, $p(x) = \left(1 - \frac{x}{d_1}\right) \cdots \left(1 - \frac{x}{r_d}\right)$. The idea is that we do the same thing with $\sin(x)$; it has an everywhere-converging expansion (which can be proven with complex analysis)

$$\begin{aligned} \frac{\sin x}{x} &= \left(1 - \frac{x}{\pi}\right)\left(1 + \frac{x}{\pi}\right)\left(1 - \frac{x}{2\pi}\right)\left(1 + \frac{x}{2\pi}\right)\cdots \\ &= \prod_{n \geq 1}\left(1 - \frac{x^2}{n^2\pi^2}\right). \end{aligned}$$

Then the coefficient of x^2 in the infinite product is equal to $-\frac{1}{1^2\pi^2} - \frac{1}{2^2\pi^2} - \frac{1}{9\pi^2} = -\frac{1}{\pi^2}\sum_{n=1}^{\infty}\frac{1}{n^2}$. On the other hand, the x^2 coefficient of $\frac{\sin x}{x}$ (by looking at the Taylor expansion of $\sin(x)$ and cancelling) is plainly $-\frac{1}{6}$. Therefore, we can conclude that

$$-\frac{1}{6} = -\frac{1}{\pi^2}\zeta(2) \Rightarrow \frac{\pi^2}{6} = \zeta(2).$$

□

2.2 Infinite Products

In this section, we develop some formal logic around infinite products which is helpful for finding "Euler products" for various arithmetical functions (this concept will be revisited, so don't worry if it seems opaque).

We begin with a technical lemma:

Lemma 2.4. *For $0 < \varepsilon < 1$ (e.g. $\varepsilon = 1/2$), there are constants c_ε, $C_\varepsilon > 0$ so that $c_\varepsilon|x| \leq |\log(1+x)| \leq C_\varepsilon|x|$ when $|x| \leq 1 - \varepsilon$ (here, we allow x to be complex).*

Proof. Fix $\varepsilon \in (0, 1)$. Then, when $|x| \leq 1 - \varepsilon$,

$$|\log(1+x)| \leq \sum_{m\geq 1}\left|(-1)^{m-1}\frac{x^m}{m}\right| = |x|\sum_{m\geq 1}\left|\frac{x^{m-1}}{m}\right| \leq |x|\sum_{m\geq 1}(1-\varepsilon)^{m-1} = \frac{|x|}{\varepsilon}.$$

On the other hand, when $0 \leq x \leq 1-\varepsilon$,

$$
\begin{aligned}
|\log(1+x)| &\geq \left|\sum_{m\geq 1}(-1)^{m-1}\frac{x^m}{m}\right| \geq |x| - \left|\sum_{m\geq 2}(-1)^{m-1}\frac{x^m}{m}\right| = |x| - |x|\left|\sum_{m\geq 2}(-1)^{m-1}\frac{x^{m-1}}{m}\right| \\
&\geq |x| - |x|\left|\sum_{m\geq 2}(-1)^{m-1}x^{m-1}\right| \geq |x| - |x|\left|\sum_{m\geq 2}(\varepsilon-1)^{m-1}\right| = |x| - |x| \cdot \left|\frac{\varepsilon-1}{1-(\varepsilon-1)}\right| \\
&= |x| - |x| \cdot \frac{1-\varepsilon}{2-\varepsilon} = |x| \cdot \frac{(2-\varepsilon)-(1-\varepsilon)}{(2-\varepsilon)} = |x| \cdot \frac{1}{2-\varepsilon}.
\end{aligned}
$$

Hence $c_\varepsilon = \frac{1}{2-\varepsilon}$ and $C_\varepsilon = \frac{1}{\varepsilon}$ suffice. □

Next, we provide a helpful criterion for the converging of infinite products:

Lemma 2.5. *Suppose $\{a_n\}$ is a real sequence for which $\sum a_n$ is absolutely convergent. If $a_n \neq -1$ for all n then $\prod(1+a_n) = \lim_{N\to\infty}\prod_{n=1}^{N}(1+a_n)$ converges to a nonzero value unaffected by rearrangement of $\{a_n\}$.*

Proof. First, I claim that $\sum_{n=1}^{\infty}\log(1+a_n)$ converges absolutely. To see why, notice that for sufficiently large N, there exists ε such that $|a_n| < 1-\varepsilon$ for all $n > N$, and then by Lemma 2.4

$$c_\varepsilon \sum_{n=N}^{\infty}|a_n| \leq \sum_{n=N}^{\infty}|\log(1+a_n)| \leq C_\varepsilon \sum_{n=N}^{\infty}|a_n|$$

and therefore, since it is sandwiched between two finite quantities (recall that $\sum a_n$ is absolutely convergent), $\sum_{n=N}^{\infty}|\log(1+a_n)|$ converges. Hence $\sum_{n=1}^{\infty}|\log(1+a_n)|$ converges and therefore $\sum_{n=1}^{\infty}\log(1+a_n)$ converges absolutely. In particular, this implies that $\sum_{n=1}^{\infty}\log(1+a_n)$ converges to a value unaffected by rearrangement of $\{a_n\}$. But then

$$e^{\sum_{n=1}^{\infty}\log(1+a_n)} = e^{\lim_{N\to\infty}\sum_{n=1}^{N}\log(1+a_n)} = \lim_{N\to\infty} e^{\sum_{n=1}^{N}\log(1+a_n)} = \lim_{N\to\infty}\prod_{n=1}^{N} e^{\log(1+a_n)} = \lim_{N\to\infty}\prod_{n=1}^{N}(1+a_n)$$

where the second equality follows by continuity of $x \mapsto e^x$. This means precisely that $\lim_{N\to\infty}\prod_{n=1}^{N}(1+a_n)$ converges, and it converges to a nonzero value since it is expressed as e^C for some constant C (since the exponential is everywhere nonvanishing). The fact that this value is unaffected by the rearrangement of $\{a_n\}$ follows immediately from the fact that it is equal to the exponential of $\sum_{n=1}^{\infty}\log(1+a_n)$, which as we noted earlier, is independent of the order of the $\{a_n\}$. □

Now, let Ω denote $\{z \in \mathbb{C} \mid \Re(z) > 0\}$, the open right half-plane. Then, we can define the principal branch of the logarithm on Ω, to be denoted Log, by the usual methods. Then, the same result holds:

Lemma 2.6. *Suppose $\{z_n\}$ is a sequence of complex numbers in Ω such that $\sum z_n$ is absolutely convergent. Then $\prod(1+z_n) = \lim_{N\to\infty}\prod_{n=1}^{N}(1+z_n)$ converges to a nonzero value unaffected by rearrangement of $\{z_n\}$.*

Proof. First notice that Lemma 2.4 holds for all complex numbers z such that $|z| < 1-\varepsilon$. Therefore, by exactly the same proof as in the above lemma, $\sum_{n=1}^{\infty}\mathrm{Log}(1+z_n)$ converges absolutely. Now, for any n, $e^{\mathrm{Log}(1+z_n)} = 1+z_n$ whence $\prod_{n=1}^{N}(1+z_n) = \prod_{n=1}^{N}e^{\mathrm{Log}(1+z_n)} = e^{\sum_{n=1}^{N}\mathrm{Log}(1+z_n)}$. By continuity of $z \mapsto e^z$,

$$\lim_{N\to\infty}\prod_{n=1}^{N}(1+z_n) = \lim_{N\to\infty}e^{\sum_{n=1}^{N}\mathrm{Log}(1+z_n)} = e^{\lim_{N\to\infty}\sum_{n=1}^{N}\mathrm{Log}(1+z_n)}$$

and since $\lim_{N\to\infty}\sum_{n=1}^{N}\mathrm{Log}(1+z_n)$ exists and is independent of the order of the $\{z_n\}$, we may conclude that $\lim_{N\to\infty}\prod_{n=1}^{N}(1+z_n)$ exists, is independent of the order of the $\{z_n\}$, and is nonzero (since the exponential is everywhere nonvanishing on the complex plane). Therefore, we are done. □

Note that in the last proof, we had to use a slightly different method since Log does not, in general, take finite products to finite sums. However, the exponential does indeed take finite sums to finite products. This latter result is not quite a generalization, because we do not allow $\{a_n\}$ to be a negative real number for any n in the latter proof but we do in the former, but it is a helpful pseudo-generalization.

Proposition 2.7 (Euler's Product Formula)**.**

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \prod_{p \text{ prime}} \left(1 + \frac{1}{p^s} + \frac{1}{p^{2s}} + \cdots\right) = \prod_{p \text{ prime}} \left(\frac{1}{1 - \frac{1}{p^s}}\right).$$

Proof. The intuitive understanding of this is given by prime factorization; it is possible to make this precise by a routine application of the above theorems. □

2.3 Dirichlet Convolution and Dirichlet Series

In this section, we define "Dirichlet series", which are in some sense a generalization of the Riemann ζ-function.

Definition 2.8 (Arithmetical Function)**.** An arithmetical function is a function $f : \mathbb{Z}^+ \to \mathbb{C}$.

Definition 2.9 (Multiplicative)**.** An arithmetical function f is said to be *multiplicative* if whenever m and n are coprime, $f(mn) = f(m)f(n)$. Furthermore, f is said to be *totally* or *completely multiplicative* if $f(mn) = f(m)f(n)$ for all positive integers m and n (not just pairs of coprime positive integers).

Definition 2.10 (Dirichlet Convolution)**.** Given two functions f and g, their *Dirichlet convolution* $f \star g$ is defined to be the function $(f \star g)(n) = \sum_{d|n} f(d) g\left(\frac{n}{d}\right)$.

Definition 2.11 (Basic Arithmetical Functions)**.** There are three functions from which many other Dirichlet convolutions are often built up:

1. δ denotes the arithmetical function such that $\delta(1) = 1$ and $\delta(n) = 0$ for $n > 1$. This function is important because $\delta \star f = f \star \delta = f$, so δ serves the role of "multiplicative identity" in a sense which will be made precise in the next section.

2. 1 denotes the constant function at 1; the arithmetical function $1(n) = 1$.

3. id denotes the identity function on $\mathbb{Z}^+$; the arithmetical function $\mathrm{id}(n) = 1$.

Following is a "sanity-check" that Dirichlet convolution is a "nice" operation (it is associative):

Lemma 2.12. *Dirichlet convolution is associative:* $f \star (g \star h) = (f \star g) \star h$.

Proof. The key to simplifying this result is to avoid indexing over divisors d of n and working with $\frac{n}{d}$, and instead index over pairs (d_1, d_2) of positive integers whose product is n. Namely, fix an integer n. Then,

$$\begin{aligned}(f \star (g \star h))(n) &= \sum_{\substack{d_1, d_2 \in \mathbb{Z}^+ \\ d_1 d_2 = n}} f(d_1)(g \star h)(d_2) = \sum_{\substack{d_1, d_2 \in \mathbb{Z}^+ \\ d_1 d_2 = n}} f(d_1) \sum_{\substack{d_3, d_4 \in \mathbb{Z}^+ \\ d_3 d_4 = d_2}} g(d_3)h(d_4) = \sum_{\substack{d_1, d_2, d_3 \in \mathbb{Z}^+ \\ d_1 d_2 d_3 = n}} f(d_1)g(d_2)h(d_3) \\ &= \sum_{\substack{d_1, d_2 \in \mathbb{Z}^+ \\ d_1 d_2 = n}} \sum_{\substack{d_3, d_4 \in \mathbb{Z}^+ \\ d_3 d_4 = d_1}} \left(f(d_3)g(d_4)\right) h(d_2) = \sum_{\substack{d_1, d_2 \in \mathbb{Z}^+ \\ d_1 d_2 = n}} (f \star g)(d_1)h(d_2) = ((f \star g) \star h)(n).\end{aligned}$$

□

Obviously, the operation of Dirichlet convolution is also commutative.

Definition 2.13 (Divisor Counting-Function)**.** Let $d(n)$ be the number of divisors of n; that is, $d(n) = \sum_{d|n} 1$. Then, we also have that $d(n) = \sum_{d|n} 1 \cdot 1 = (1 \star 1)(n)$, so $d = 1 \star 1$. Furthermore, it is easy to see that if $n = p_1^{e_1} \cdots p_k^{e_k}$, then $d(n) = (e_1 + 1) \cdots + (e_k + 1)$; from this it is easy to see that the divisor function is multiplicative.

Now, here we have an example of a function which is obtained by the Dirichlet convolution of two multiplicative functions being multiplicative. A natural question to ask is whether or not this always happens: fortunately, the answer is yes! One of the nicest properties of Dirichlet convolution is that it preserves multiplicativity, as the following theorem demonstrates.

Theorem 2.14. *If f and g are multiplicative arithmetical functions, then $(f \star g)$ is also multiplicative.*

Proof. Suppose m and n are coprime, and consider the following:

$$(f \star g)(mn) = \sum_{d|mn} f(d) g\left(\frac{mn}{d}\right) = \sum_{\substack{d_1|m \\ d_2|n}} f(d_1 d_2) g\left(\frac{m}{d_1} \cdot \frac{n}{d_2}\right) = \sum_{\substack{d_1|m \\ d_2|n}} f(d_1) f(d_2) g\left(\frac{m}{d_1}\right) g\left(\frac{n}{d_2}\right)$$

where, for the second equality, we notice that since m and n are coprime the divisors of mn can be split up into divisors of m and divisors of n, and for the third equality we are using the fact that f and g are multiplicative. Now, we will factor the final sum given, to see that

$$\sum_{\substack{d_1|m \\ d_2|n}} f(d_1) f(d_2) g\left(\frac{m}{d_1}\right) g\left(\frac{n}{d_2}\right) = \left(\sum_{d_1|m} f(d_1) g\left(\frac{m}{d_1}\right)\right)\left(\sum_{d_2|n} f(d_2) g\left(\frac{n}{d_2}\right)\right) = (f \star g)(m) \cdot (f \star g)(n),$$

which is the desired result. □

Following is a example showing that Dirichlet convolution does not preserve complete multiplicativity, though it might seem like a natural notion:

Example 2.15. Let $d(n)$ be the number of divisors of n. Then, $d(n) = \sum_{d|n} 1 \cdot 1 = (1 \star 1)(n)$, so $d = 1 \star 1$ is the convolution of the completely multiplicative function 1 with itself. Yet d is not completely multiplicative: $d(2) = 2$ while $d(4) = 3$, so $d(2)d(2) \neq d(2 \cdot 2)$. This verifies that the convolution of two completely multiplicative functions is not necessarily completely multiplicative.

This is a useful tool for showing that functions are multiplicative:

Definition 2.16 (σ-function)**.** The σ-function $\sigma(n)$ is defined to be the sum of all of the divisors of n:

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

Proposition 2.17. *The σ-function is multiplicative.*

Proof. Notice that if $\mathrm{id} : \mathbb{Z}^+ \to \mathbb{C}$ is the identity function $\mathrm{id}(n) = n$ and $1 : \mathbb{Z}^+ \to \mathbb{C}$ is the trivial function $1(n) = 1$, then both id and 1 are totally multiplicative (and hence multiplicative) functions. But also

$$\sigma = \mathrm{id} \star 1$$

so because the Dirichlet convolution of multiplicative functions is multiplicative, σ is multiplicative. □

Here is another example where proving multiplicativity is important, but not actually the end goal:

Lemma 2.18. *$\sum_{d|n} \varphi(d) = n$; in other words, $\varphi \star 1 = \mathrm{id}$.*

Proof. Notice that $\sum_{d|n} \varphi(d)$ is equal to the Dirichlet convolution $\varphi \star 1$, where $1 : n \to 1$ is the trivial arithmetical function. Since φ and 1 are both multiplicative, by Theorem 2.14 $\sum_{d|n} \varphi(d)$ is multiplicative. Hence it suffices to show that

$$\sum_{d|p^k} \varphi(d) = p^k$$

for every prime power p^k (since then by multiplicativity the result will follow). Yet this is simple:

$$\sum_{d|p^k} \varphi(d) = \varphi(1) + \varphi(p) + \varphi(p^2) + \cdots + \varphi(p^k) = 1 + (p-1) + (p^2 - p) + \cdots + (p^k - p^{k-1}) = p^k$$

with the final equality following by telescoping. □

Next, let us introduce a function whose Dirichlet convolutions are extremely important.

Definition 2.19 (Möbius Function)**.** Define $\mu(n)$ to be the following arithmetical function:

$$\mu(n) = \begin{cases} (-1)^k & n = p_1 \cdots p_k \text{ for distinct primes } p_1, \ldots, p_k \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 2.20 (Möbius Inversion Formula)**.** *If* 1 *is the constant arithmetical function* $1(n) = 1$*, then* $(1 \star \mu) = \delta$*. More generally, if* $F = 1 \star f = f \star 1$*, then* $f = F \star \mu = \mu \star F$*.*

Proof. We begin with the first part. Now, $(\mu \star 1)(1) = \delta(1) = 1$. Then, for any $n > 1$, write $n = p_1^{e_1} \cdots p_k^{e_k}$. Then $\mu(d)$ (for $d \mid n$) is only nonzero if d is the product of distinct primes among $p_1, \ldots, p_k$. Hence

$$\sum_{d|n} \mu(d) = \sum_{S \subseteq \{1,\ldots,k\}} \mu\left(\prod_{s \in S} p_s\right) = \sum_{S \subseteq \{1,\ldots,k\}} (-1)^{|S|} = \sum_{i=0}^{k} \binom{k}{i} (-1)^i = (1-1)^k = 0$$

where the third equality is grouping together terms corresponding to subsets of the same size and the fourth equality is simply the Binomial Theorem. Hence $(\mu * 1)(n) = \delta(n)$ for any $n > 1$, so in general $\mu * 1 = \delta$.

Since $f \star \delta = f$, the conclusion is immediate:

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

□

Definition 2.21. If f and g are arithmetical functions satisfying $g = f \star 1$ and $f = \mu \star g$, then f and g are said to be *Möbius transforms* of one another. In particular, f is multiplicative if and only if g is multiplicative (by the fact that the Dirichlet convolution of multiplicative functions is multiplicative, Theorem 2.14).

For example, the Euler φ-function and the identity map $n \mapsto n$ are Möbius transforms of one another. If we demonstrated this fact without relying on the multiplicativity of φ, it would provide an alternate proof of the fact that φ is multiplicative.

Definition 2.22 (Dirichlet Series)**.** The *Dirichlet series associated to* f is $F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}$.

The relationship between Dirichlet convolution and Dirichlet series is given by the following proposition:

Proposition 2.23. *If* $F(s)$ *and* $G(s)$ *are the Dirichlet series associated to* f *and* g *respectively, then*

$$F(s) \cdot G(s) = \sum_{n=1} \frac{(f \star g)(n)}{n^s}.$$

Example 2.24. Let 1 be the constant function 1. Then the Dirichlet series associated to 1 is the Riemann ζ-function. Importantly, the Dirichlet series $D(s)$ associated to the divisor-counting function d is the square of the Riemann ζ-function; $D(s) = \zeta(s)^2$.

Example 2.25. The Dirichlet series $\sum_{n=1}^{\infty} \frac{\varphi(n)}{n^s}$ is equal to $\frac{\zeta(s-1)}{\zeta(s)}$ for all s such that $\Re(s) > 2$. To see why, recall Lemma 2.18, which implies that $\varphi \star 1 = \text{id}$, and therefore that

$$\left(\sum_{n=1}^{\infty} \frac{\varphi(n)}{n^s}\right)\left(\sum_{n=1}^{\infty} \frac{1}{n^s}\right) = \left(\sum_{n=1}^{\infty} \frac{(\varphi \star 1)(n)}{n^s}\right) = \left(\sum_{n=1}^{\infty} \frac{n}{n^s}\right) \Rightarrow \left(\sum_{n=1}^{\infty} \frac{\varphi(n)}{n^s}\right) \zeta(s) = \zeta(s-1)$$

as long as everything converges, which it does iff $\Re(s) > 2$.

Definition 2.26 (Ring of Dirichlet Series)**.** The set of all Dirichlet series of arithmetical functions on $\mathbb{Z}$ forms a commutative ring, with addition and multiplication defined by

$$\left(\sum_{n=1}^{\infty} \frac{f(n)}{n^s}\right) + \left(\sum_{n=1}^{\infty} \frac{g(n)}{n^s}\right) = \left(\sum_{n=1}^{\infty} \frac{f(n)+g(n)}{n^s}\right) \quad \left(\sum_{n=1}^{\infty} \frac{f(n)}{n^s}\right)\left(\sum_{n=1}^{\infty} \frac{g(n)}{n^s}\right) = \left(\sum_{n=1}^{\infty} \frac{(f \star g)(n)}{n^s}\right).$$

The only hard part to verify is associativity of the product, but this is proven in the previous section (see Lemma 2.12). The multiplicative identity of this ring is the Dirichlet series of $\delta(n)$.

Proposition 2.27. *The Dirichlet series of μ is the multiplicative inverse of the Riemann ζ-function.*

Proof. This can either be derived from Proposition 2.20 or directly by using the Euler product formula. For the latter, notice that by the Euler product formula $\zeta^{-1}(s) = \prod_p \left(1 - \frac{1}{p^s}\right)$, and by expanding it is not hard to see that we get the Dirichlet series of μ. □

One interesting idea about Dirichlet series is that Dirichlet series of multiplicative functions always have (at least formally) Euler products: if f is multiplicative, then

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

at least formally; we need to check that it converges, of course.

3 More Advanced Approximations

3.1 Counting With The Riemann ζ-Function

Proposition 3.1. *The number of squarefree integers between $1, \ldots, x$ is $\frac{6x}{\pi^2} + O(\sqrt{x})$. Therefore, in the sense of natural density, $\frac{6}{\pi^2}$ of integers are squarefree.*

Proof. Let $1_{\text{sqf}}(n)$ be 1 if n is squarefree and 0 otherwise. Then our goal is to compute $\sum_{n \le x} 1_{\text{sqf}}(n)$. Now, any positive integer has a unique decomposition $n = a^2 b$, where a is a positive integer and b is squarefree. Then $1_{\text{sqf}}(n) = \delta(a) = \sum_{d \mid a} \mu(d) = \sum_{d^2 \mid n} \mu(d)$. Hence

$$\sum_{n \le x} 1_{\text{sqf}}(n) = \sum_{n \le x} \sum_{d^2 \mid n} \mu(d) = \sum_{d \le \sqrt{x}} \mu(d) \sum_{\substack{d^2 \mid n \\ n \le x}} 1 = \sum_{d \le \sqrt{x}} \mu(d) \left\lfloor \frac{x}{d^2} \right\rfloor = \sum_{d \le \sqrt{x}} \mu(d) \left(\frac{x}{d^2} + O(1) \right)$$

$$= x \sum_{d \le \sqrt{x}} \frac{\mu(d)}{d^2} + O\left(\sum_{d \le \sqrt{x}} \mu(d) \right) = x \sum_{d \le \sqrt{x}} \frac{\mu(d)}{d^2} + O\left(\sqrt{x}\right).$$

Next, we analyze $\sum_{d \le \sqrt{x}} \frac{\mu(d)}{d^2}$. The key is that $\sum_{d=1}^{\infty} \frac{\mu(d)}{d^2} = \frac{1}{\zeta(2)} = \frac{6}{\pi^2}$. Hence

$$x \sum_{d \le \sqrt{x}} \frac{\mu(d)}{d^2} = \frac{x}{\zeta(2)} - x \sum_{d > \sqrt{x}} \frac{\mu(d)}{d^2}$$

and of course $\sum_{d > \sqrt{x}} \frac{\mu(d)}{d^2} = O(\frac{1}{\sqrt{x}})$. Putting everything together, we are done. □

One can also informally guess this answer by noticing that the probability that $p^2 \nmid n$ is $1 - \frac{1}{p^2}$ for some "uniformly chosen n". Therefore the probability that some uniformly chosen n is squarefree is approximately $\prod_p \left(1 - \frac{1}{p^2}\right) = \frac{1}{\zeta(2)}$. Of course, this is not a precise argument, but it is an interesting heuristic. The theory of approximations teaches us that these heuristic arguments can be incredibly helpful in illustrating the correct answer even before we know how to formally justify them. For historical proof, one only needs to look at Euler, widely considered one of the greatest mathematicians of all time – and a proponent of (to put it kindly) less-than-rigorous ideas.

Proposition 3.2. *The average value of the Euler φ-function is $\frac{x}{2\zeta(2)} + O(\log(x))$. That is,*

$$\sum_{n \le x} \varphi(n) = \frac{x^2}{2\zeta(2)} + O(x \log x).$$

Proof. First, notice that $\varphi \star 1 = \text{id}$ (Lemma 2.18) implies $\varphi = \mu \star \text{id}$ whence $\varphi(n) = \sum_{d \mid n} \mu(d) \frac{n}{d}$. Factoring out a copy of n, we arrive at the formula $\frac{\varphi(n)}{n} = \sum_{d \mid n} \frac{\mu(d)}{d}$. Now, we will first compute $\sum_{n \le x} \frac{\varphi(n)}{n}$, and then we will remove the factor of $\frac{1}{n}$ using partial summation. For this, we see that

$$\sum_{n \le x} \frac{\varphi(n)}{n} = \sum_{n \le x} \sum_{d \mid n} \frac{\mu(d)}{d} = \sum_{d \le x} \frac{\mu(d)}{d} \sum_{\substack{n \le x \\ d \le x}} 1 = \sum_{d \le x} \frac{\mu(d)}{d} \left\lfloor \frac{x}{d} \right\rfloor = \sum_{d \le x} \frac{\mu(d)}{d} \left(\frac{x}{d} + O(1) \right).$$

Now, expanding this becomes $x \sum_{d \le x} \frac{\mu(d)}{d^2} + O(\sum_{d \le x} \frac{1}{d}) = x \left(\frac{1}{\zeta(2)} + O(\frac{1}{x}) \right) + O(\log x)$, where the simplification of $\sum_{d \le x} \frac{\mu(d)}{d^2}$ is essentially what we did in the previous result but with different bounds. Of course, collapsing error terms all of this is equal to $\sum_{n \le x} \frac{\varphi(n)}{n} = \frac{x}{\zeta(2)} + O(\log x)$. Now that we have this estimate, by taking $a(n) = \frac{\varphi(n)}{n}$ and $f(n) = n$, we apply the formula of partial summation to get

$$\begin{aligned}\sum_{n\le x}\varphi(n) &= \int_{1^-}^{x} td\left(\sum_{n\le t}\frac{\varphi(n)}{n}\right) = x\cdot\sum_{n\le x}\frac{\varphi(n)}{n} - \int_{1}^{x}\sum_{n\le t}\frac{\varphi(n)}{n}dt \\ &= x\left(\frac{x}{\zeta(2)} + O(\log x)\right) - \int_{1}^{x}\left(\frac{t}{\zeta(2)} + O(\log(t)\right) dt \\ &= \frac{x^2}{\zeta(2)} + O(x\log x) - \frac{x^2}{2\zeta(2)} + O(x\log x) = \frac{x^2}{2\zeta(2)} + O(x\log x).\end{aligned}$$

□

Corollary 3.2.1. *Now, $\sum_{n\le x}\varphi(n) = |\{(m,n) \mid m,n\le x \text{ such that } (m,n)=1\}$; that is, the sum counts the number of unordered pairs of coprime positive integers at most x. Therefore, the number of unordered pairs of coprime positive integers at most x is $\frac{x^2}{2\zeta(2)} + O(x\log x)$.*

This proportion can also be informally verified by a heuristic argument like the previous one; finding it is left as an exercise for the reader.

3.2 More Estimations to Do With Divisors

Finally, we offer some bounds on the divisor-counting function, showing that divisors tend to be quite rare. Now, there is an easy first bound: $d(n) = \sum_{ab=n} 1 \le 2\sum_{a\le\sqrt{n}} 1 \le 2\sqrt{n}$.

Proposition 3.3. *$d(n) \ll_\varepsilon n^\varepsilon$ for all $\varepsilon > 0$, where $\ll_\varepsilon$ serves to show that the constant C such that $d(n) \le Cn^\varepsilon$ depends on ε.*

Proof. Suppose that $n = p_1^{e_1}\cdots p_k^{e_k}$. Then, via the formula in the definition of the divisor-counting function,

$$\frac{d(n)}{n^\varepsilon} = \prod_{j=1}^{k}\frac{e_j+1}{p_j^{\varepsilon e_j}}$$

Now, if I fix p and ε, then the expression $t_{p,\varepsilon}(e) = \frac{e+1}{p^{\varepsilon e}}$ is a differentiable function of e with $t_{p,\varepsilon}(0) = 1$ and

$$t'_{p,\varepsilon}(e) = \frac{p^{\varepsilon e} - (e+1)\varepsilon\log(p)p^{\varepsilon e}}{p^{2\varepsilon e}} = \frac{1-(e+1)\varepsilon\log(p)}{p^{\varepsilon e}}$$

Notice that for sufficiently large p, this eventually is negative for all e, and therefore for sufficiently large p (say all primes larger than some prime P), the largest value that $t_{p,\varepsilon}$ takes is at 0, where it is equal to 1. Yet then,

$$\frac{d(n)}{n^\varepsilon} = \prod_{j=1}^{k} t_{p_j,\varepsilon}(e_j) \le \prod_{j=1}^{k}\max_{e\in\mathbb{R}^{\ge 0}}\{t_{p_j,\varepsilon}(e)\} \le \prod_{\text{prime } p}\max_{e\in\mathbb{R}^{\ge 0}}\{t_{p,\varepsilon}(e)\} = \prod_{\substack{\text{prime } p\\ p\le P}}\max_{e\in\mathbb{R}^{\ge 0}}\{t_{p_j,\varepsilon}(e)\}$$

where the second equality follows because for all $p > P$, the maximum is 1 (as we mentioned earlier). Yet then the right-hand side is a finite product, and therefore yields a finite constant C_ε. The result follows. □

This is a very strong result, but it is also helpful to have a sense of the average value of the divisor function:

Proposition 3.4. *$\sum_{n\le x} d(n) = x\log x + O(x)$; that is, the average value of the divisor function between 1 and x is asymptotically $\log x + O(1)$.*

Proof. The proof follows the usual sum-switching

$$\sum_{n\le x}\sum_{d|n} 1 = \sum_{d\le x}\sum_{\substack{n\le x\\ d|n}} 1 = \sum_{d\le x}\lfloor x\rfloor d = \sum_{d\le x}\left(\frac{x}{d} + O(1)\right).$$

Now, of course, by Proposition 1.5, this is equal to $x\left(\log x + \gamma + O\left(\frac{1}{x}\right)\right) + O(x) = x\log x + O(x)$, where γ is the Euler-Mascheroni constant (which is the correct constant for that proposition). □

We will now prove a more complicated theorem:

Proposition 3.5. $\sum_{n\le x} d(n) = x\log x + (2\gamma - 1)x + O(\sqrt{x})$.

Proof. The proof is by the "hyperbola method". The tactic is used when one can express something as the convolution of two things. That is,

$$\sum_{n\le x} d(n) = \sum_{n\le x}\sum_{ab=n} 1 = \sum_{\substack{a,b\\ ab\le x}} 1$$

But this latter question can be understood as the number of lattice points between the hyperbola $ab = x$ and the a and b-axis. Therefore, fix a point (A, B) on the hyperbola $ab = x$. Then there are three types of points: (1) points (a, b) such that $a \le A$, (2) points (a, b) such that $b \le B$, and (3) points in (1) and (2). Then, to count the points, we can count the number of points in the first category, add the number of points in the second category, and subtract the number of points in the third.

Now, to approximate the first category, notice that it is equal to

$$\sum_{a\le A}\sum_{b\le \frac{x}{a}} 1 = \sum_{a\le A}\left(\frac{x}{a} + O(1)\right) = x\sum_{a\le A}\frac{1}{a} + O(A) = x\left(\log A + \gamma + O\left(\frac{1}{A}\right)\right) + O(A)$$

which, by expanding and using the fact that $\frac{x}{A} = B$, yields that the first category has $x\log A + x\gamma + O(A+B)$ points. Now, to approximate the second category, we notice that everything is symmetric, so the second category has $x\log B + x\gamma + O(A+B)$ points. Now, of course the third category has $(A+O(1))(B+O(1)) = AB + O(A+B) = x + O(A+B)$ points. Summing everything, and then choose our point (A, B) to be $(\sqrt{x}, \sqrt{x})$, we get the desired result. □

It is indeed known that the error is $O(x^{131/416+\varepsilon})$ for any $\varepsilon > 0$, greater than $O(x^{1/4})$,. Furthermore, it is conjectured that the error is $O(x^{1/4+\varepsilon})$ for any $\varepsilon > 0$.

The hyperbola method is a more general version of this tactic: one estimates

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

by splitting it up sums for $a \le A$, $b \le B$, and both (where A, B are such that $AB = x$). The above proof was the special case where f and g were the constant function 1, as $1 \star 1 = d$.

Definition 3.6 (ω and Ω)**.** Suppose that $n = p_1^{e_1}\cdots p_k^{e_k}$. Then $\omega(n) = k$ is the number of prime divisors of n, and $\Omega(n) = e_1 + \cdots + e_k$ is the number of prime divisors of n counting multiplicity.

Proposition 3.7. $\sum_{n\le x}\omega(n) = x\log\log x + O(x)$. *On average, therefore, the value of* $\omega(n)$ *between* 1 *and* x *is* $\log\log x + O(1)$; $\omega(n)$ *is usually around* $\log\log n$.

Proof. This is a simple proof:

$$\sum_{n\le x}\omega(n) = \sum_{n\le x}\sum_{p\mid n} 1 = \sum_{p\le x}\sum_{\substack{n\le x\\ p\mid n}} 1 = \sum_{p\le x}\left(\frac{x}{p} + O(1)\right) = x\sum_{p\le x}\frac{1}{p} + O\left(\sum_{p\le x} 1\right).$$

where, by using our asymptotic expression for the reciprocal of the primes, we find that this is equal to $x(\log\log x + O(1)) + O(x) = x\log\log x + O(x)$. □

3.3 From Averages to Distributions, Exceptional Values

Recall the following definition.

Definition 3.8 (Variance)**.** The *variance* of a set of numbers $\{s_1, \ldots, s_N\}$ with mean $\overline{s}$ is $\frac{1}{N}\sum_{n=1}^{N}(s_i - \overline{s})^2$. The square root of the variance is the standard deviation.

Lemma 3.9. *$\sum_{n\le x}\omega(n)^2 = x(\log\log x)^2 + O(x\log\log x)$.*

Proof.

$$\sum_{n\le x}\omega(n)^2 = \sum_{n\le x}\left(\sum_{p|n}1\right)^2 = \sum_{n\le x}\sum_{p|n}\sum_{q|n}1 = \sum_{p,q\le x}\sum_{\substack{n\le x\\ p,q|n}}1.$$

Now, if $p \neq q$, then $\sum_{\substack{n\le x\\ p,q|n}}1 = \left\lfloor\frac{x}{pq}\right\rfloor = \frac{x}{pq} + O(1)$. On the other hand, if $p = q$, then $\sum_{\substack{n\le x\\ p,q|n}}1 = \left\lfloor\frac{x}{p}\right\rfloor = \frac{x}{p} + O(1)$. Therefore, we need to compute

$$\sum_{\substack{pq\le x\\ p\neq q}}\left(\frac{x}{pq} + O(1)\right) + \sum_{p\le x}\left(\frac{x}{p} + O(1)\right) = \sum_{\substack{pq\le x\\ p\neq q}}\left(\frac{x}{pq} + O(1)\right) + x\log\log x + O(x).$$

Now, we seek to compute

$$\sum_{\substack{pq\le x\\ p\neq q}}\left(\frac{x}{pq} + O(1)\right) = \sum_{pq\le x}\left(\frac{x}{pq} + O(1)\right) - \sum_{\substack{p=q\\ p^2\le x}}\left(\frac{x}{p^2} + O(1)\right) = \sum_{pq\le x}\left(\frac{x}{pq}\right) + O(x) - O(x) = \sum_{pq\le x}\left(\frac{x}{pq}\right) + O(x).$$

Therefore,

$$\sum_{n\le x}\omega(n)^2 = \sum_{pq\le x}\left(\frac{x}{pq}\right) + x\log\log x + O(x).$$

Now,

$$\sum_{pq\le x}\frac{1}{pq} \le \sum_{p\le x}\sum_{q\le x}\frac{1}{pq} = \left(\sum_{p\le x}\frac{1}{p}\right)^2 = (\log\log x + O(1))^2 = (\log\log x)^2 + O(\log\log x)$$

On the other hand,

$$\sum_{pq\le x}\frac{1}{pq} \ge \left(\sum_{p\le\sqrt{x}}\frac{1}{p}\right)^2 = \left(\log\log\sqrt{x} + O(1)\right)^2 = (\log(\frac{1}{2}\log(x)) + O(1))^2 = (\log\log x + \log(1/2) + O(1))^2$$
$$= (\log\log x + O(1))^2 = (\log\log x)^2 + O(\log\log x).$$

Therefore, $\sum_{pq\le x}\left(\frac{x}{pq}\right) = x(\log\log x)^2 + O(x\log\log x)$. Hence we may conclude that $\sum_{n\le x}\omega(n)^2 = x(\log\log x)^2 + O(x\log\log x)$. □

Proposition 3.10. *The variance of $\{\omega(1), \ldots, \omega(x)\}$ is $O(\log\log x)$. Hence the standard deviation of $\{\omega(1), \ldots, \omega(x)\}$ is $O(\sqrt{\log\log x})$.*

Proof. Define $E(x)$ to be the error term function such that $x\log\log x + xE(x) = \sum_{n\le x}\omega(n)$. Then, $S = \sum_{n\le x}\left(\omega(n) - \log\log x - E(x)\right)^2 = \sum_{n\le x}\omega(n)^2 - 2\log\log x\sum_{n\le x}\omega(n) + \sum_{n\le x}(\log\log x)^2 + \text{(error)}$ where $\text{(error)} = O(x\log\log x)$.

Now, we can evaluate the latter two non-error terms by our work earlier or trivially, so we have

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

Then, by applying the preceding lemma, we find that $S = O(x\log\log x)$, so $\frac{S}{x} = O(\log\log x)$; this is the variance of $\{\omega(1), \ldots, \omega(x)\}$ by definition, so we are done. □

Proposition 3.11. *Let $\mathcal{E}_{A,x} = \{n \leq x \mid |\omega(n) - \log\log x| \geq A\sqrt{\log\log x}\}$. Then, $|\mathcal{E}_{A,x}| \ll \frac{x}{A^2}$.*

Proof. This is a routine computation. Namely, $x \log\log x \gg \sum_{n \leq x}(\omega(n) - \log\log x)^2 \geq \sum_{n \leq \mathcal{E}_{A,x}} \left(A\sqrt{\log\log x}\right)^2 = |\mathcal{E}_{A,x}| A^2 \log\log x$. Therefore, $|\mathcal{E}_{A,x}| \ll \frac{x}{A^2}$. □

Therefore, in particular, if A is a function of x heading to infinity as $x \to \infty$, then the proportion of exceptional values goes to 0. As we saw before, the average value of $d(n)$ over all n less than or equal to x is $\log x + O(1)$. Now, plainly $2^{\omega(n)} \leq d(n) \leq 2^{\Omega(n)}$ either by counting divisors or by noticing that $2 \leq e + 1 \leq 2^e$ for all non-negative integers e and then applying the formula for $\omega(n)$, $d(n)$, and $\Omega(n)$ using the prime factorization of n. Therefore, by our above work, for almost all n we have

$$2^{\log\log n + O(1)} \leq d(n) \leq 2^{\log\log n + O(1)} \Rightarrow e^{\log 2 \log\log n + O(1)} \leq d(n) \leq e^{\log 2 \log\log n + O(1)}.$$

Therefore, $d(n)$ is, for almost all n, $d(n) = O(\log n^{\log 2}) = o(\log n)$. Therefore, the mean of $d(n)$ is skewed by a small number of large values (the variance is very high). Later, we will indeed show that $\frac{1}{x}\sum_{n \leq x}(d(n) - \log x)^2 = O((\log x)^3)$.

The following result can be used to show that $\Omega(n)$ has the same statistics as $\omega(n)$:

Theorem 3.12.

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

Proof. Recall that $\text{ord}_p(n)$ is the largest natural number k such that $p^k \mid n$. Then,

$$\begin{aligned}\sum_{n \leq x}(\Omega(n) - \omega(n))^2 &= \sum_{n \leq x}\left(\sum_{p \mid n}(\text{ord}_p(n) - 1)\right)^2 = \sum_{n \leq x}\sum_{p,q \mid n}(\text{ord}_p(n) - 1)(\text{ord}_q(n) - 1) \\ &= \sum_{p,q \leq x}\sum_{\substack{n \leq x \\ pq \mid n}}(\text{ord}_p(n) - 1)(\text{ord}_q(n) - 1).\end{aligned}$$

Now, we split this sum into two parts: the part where $p = q$, and the part where $p \neq q$. In the former part, we have

$$\sum_{p \leq x}\sum_{\substack{n \leq x \\ p \mid n}}(\text{ord}_p(n) - 1)^2 = \sum_{p \leq x}\left\lfloor \frac{x}{p^2} \right\rfloor + 3\left\lfloor \frac{x}{p^3} \right\rfloor + 5\left\lfloor \frac{x}{p^4} \right\rfloor + \cdots \leq x\sum_{p \leq x}\frac{1}{p^2} + \frac{3}{p^3} + \frac{5}{p^4} + \cdots$$

Now, let $S = \frac{1}{p^2} + \frac{3}{p^3} + \frac{5}{p^4} + \cdots$. Then,

$$S - \frac{S}{p} + \frac{1}{p^2} = 2\left(\frac{1}{p^2} + \frac{1}{p^3} + \cdots\right) \Rightarrow \left(1 - \frac{1}{p}\right)S = \frac{2/p^2}{(1 - 1/p)} - \frac{1}{p^2} \Rightarrow S = \frac{2}{(p-1)^2} - \frac{1}{p(p-1)}$$

which is decreasing and therefore bounded above by its value at 2, which is $\frac{3}{2}$. In other words, this part of the sum is bounded above by $\frac{3}{2}x$.

In the latter part, we have

$$\sum_{\substack{p \neq q \\ p,q \leq x}}\sum_{\substack{pq \mid n \\ n \leq x}}(\text{ord}_p n - 1)(\text{ord}_q n - 1) = \sum_{\substack{p \neq q \\ p,q \leq x}}\sum_{n \leq x/pq}(\text{ord}_p n)(\text{ord}_q n).$$

Yet, for any y, we have

$$\begin{aligned}\sum_{n \leq y}\text{ord}_p n \, \text{ord}_q n &= \sum_{n \leq y}\sum_{\substack{k \geq 1 \\ p^k \mid n}}\sum_{\substack{m \geq 1 \\ q^m \mid n}} 1 = \sum_{n \leq y}\sum_{\substack{(m,k) \geq (1,1) \\ p^k q^m \mid n}} 1 = \sum_{\substack{(m,k) \geq 1,1 \\ p^k q^m \leq y}}\sum_{\substack{n \leq y \\ p^k q^m \mid n}} 1 = \sum_{\substack{(m,k) \geq 1,1 \\ p^k q^m \leq y}}\left\lfloor \frac{y}{p^k q^m} \right\rfloor \\ &\leq y\sum_{\substack{(m,k) \geq 1,1 \\ p^k q^m \leq y}}\frac{1}{p^k q^m} \leq y\left(\sum_{k \geq 1}\frac{1}{p^k}\right)\left(\sum_{m \geq 1}\frac{1}{q^m}\right) \leq \frac{y}{(p-1)(q-1)}.\end{aligned}$$

Therefore,

$$\sum_{\substack{p\neq q\\ p,q\leq x}}\sum_{\substack{pq\mid n\\ n\leq x}}(\operatorname{ord}_p n-1)(\operatorname{ord}_q n-1)=\sum_{\substack{p\neq q\\ p,q\leq x}}\sum_{n\leq x/pq}(\operatorname{ord}_p n)(\operatorname{ord}_q n)=\sum_{\substack{p\neq q\\ p,q\leq x}}\frac{x}{pq(p-1)(q-1)}$$

$$\leq x\sum_{p,q\leq x}\frac{1}{pq(p-1)(q-1)}=x\left(\sum_{p\leq x}\frac{1}{p(p-1)}\right)\left(\sum_{q\leq x}\frac{1}{q(q-1)}\right)$$

$$\leq x\left(\sum_{n=2}^{\infty}\frac{1}{n(n-1)}\right)^2=x$$

since by telescoping $\sum_{n=2}^{\infty}\frac{1}{n(n-1)}=1$.

Therefore,

$$\sum_{n\leq x}(\Omega(n)-\omega(n))^2=\sum_{p,q\leq x}\sum_{\substack{n\leq x\\ pq\mid n}}(\operatorname{ord}_p(n)-1)(\operatorname{ord}_q(n)-1)\leq\frac{3}{2}x+x=\frac{5}{2}x$$

from which we may conclude the result. □

We leave it as an exercise to the reader to use this result to show that $\Omega(n)$ and $\omega(n)$ have the same statistics, but this is not hard (notice that, in particular, this result immediately implies that $\frac{1}{x}\sum_{n\leq x}(\Omega(n)-\omega(n))\ll 1$, since $\Omega(n)-\omega(n)$ is a natural number and therefore at least as small as $(\Omega(n)-\omega(n))^2$).

We conclude with a cute little theorem of Erdös about multiplication tables.

Theorem 3.13 (Erdös)**.** *Let $D(N)$ denote the number of distinct integers in an $N\times N$ multiplication table. That is, $D(N)=|\{n\in\mathbb{Z}^+\mid ab=n,1\leq a\leq N,1\leq b\leq N\}|$. Then $D(N)=o(N^2)$; that is, almost all numbers between 1 and N^2 do not appear in an $N\times N$ multiplication table.*

Proof. Notice that the typical size of $\Omega(n)$ for $n\leq N^2$ is $\log\log(N^2)+\log\log N+\log 2$. On the other hand, suppose that n is an integer appearing in the $N\times N$ multiplication table; say. $n=ab$ where $a,b\leq N$. Then the typical size of $\Omega(n)=\Omega(a)+\Omega(b)$ is $\log\log N+\log\log N=2\log\log N$. Therefore, all but $o(N^2)$ of the numbers in the multiplication table satisfy $\Omega(n)$ being far from $\log\log N$. In light of our earlier results, which show that the number of exceptional values of $\Omega(n)$ less than N^2 is $o(N^2)$, the result is shown. □

4 The Prime-Counting Function

4.1 The Prime Number Theorem

Recall the following definition:

Definition 4.1 (Prime-Counting Function)**.** The *prime-counting function* $\pi(x)$ counts the number of primes less than or equal to x; that is, if $1_{\text{prime}}(n)$ is the characteristic function of the primes ($1_{\text{prime}}(n) = 1$ if n is prime and 0 otherwise), then

$$\pi(x) = |\{p \leq x \mid p \text{ prime}\}| = \sum_{p \leq x} 1 = \sum_{n \leq x} 1_{\text{prime}}(n)$$

Definition 4.2 ($\mathrm{Li}(x)$)**.** Let $\mathrm{li}(x) = \int_2^x \frac{dt}{\log t}$ be the *logarithmic integral.*

The prime number theorem, which we will prove, states that $\pi(x) \sim \frac{x}{\log x}$; this is equivalent to $\mathrm{li}(x) \sim \frac{x}{\log x}$, since $\mathrm{li}(x) = \frac{x}{\log x} + O\left(\frac{x}{\log(x)^2}\right)$. Precisely, we show that $\pi(x) = \mathrm{li}(x) + O(xe^{-C\sqrt{\log x}})$ for some $C > 0$. In general, it is often enough to use the worse but simpler approximation $\pi \sim \frac{x}{\log x}$, but if you do need precision, the approximation with the logarithmic integral is much better.

Now, in fact we expect that the logarithmic integral is an even better approximation than this; the state-of-the-art bound is that $\pi(x) = \mathrm{li}(x) + O(xe^{-C(\log x)^{3/5}})$, but if the Riemann hypothesis is true, then we even have $\pi(x) = \mathrm{li}(x) + O(\sqrt{x} \log x)$, which is much, much better than current bounds.

4.2 Relationships to Other Functions

Recall the following three functions

1. The von Mangoldt function $\Lambda(n)$ (defined in Definition 1.12),
2. The first Chebyshev function $\vartheta(x) = \sum_{p \leq x} \Lambda(p)$, and
3. The second Chebyshev function $\psi(x) = \sum_{n \leq x} \Lambda(n)$.

We have already shown basic bounds to do with some of these functions in Theorem 1.15 and Proposition 1.16. Now, in this section, our goal is to rewrite PNT in terms of $\vartheta(x)$ and $\psi(x)$.

First, notice that the basic connection between $\vartheta(x)$ and $\pi(x)$ is given as follows:

$$\vartheta(x) = \int_{1^-}^{x} \log t d\pi(t) = \pi(x)\log(x) - \int_1^x \pi(t) d\log t = \pi(x)\log(x) - \int_1^x \frac{\pi(t)}{t} dt.$$

Now, the prime number theorem implies that $\pi(x) = \frac{x}{\log x} + O(\frac{x}{\log(x)^2})$. If this holds, we have that

$$\vartheta(x) = \left(\frac{x}{\log x} + O\left(\frac{x}{\log x}\right)^2\right)\log x - O\left(\int_1^x \frac{t}{\log t}\frac{dt}{t}\right) = x + O\left(\frac{x}{\log x}\right) - O\left(\int_1^x \frac{dt}{\log t}\right) = x + O\left(\frac{x}{\log x}\right).$$

That is, a good bound on π can be converted into a good bound on ϑ. In fact, we can be very precise about this relationship – precise enough to get a valid reformulation of the prime number theorem.

Suppose that $E(x)$ is the error function such that $\pi(x) = \mathrm{li}(x) + E(x)$; we make no claims on the size of $E(x)$. Then $\vartheta(x) = \mathrm{li}(x)\log(x) + E(x)\log(x) - \int_1^x \frac{\mathrm{li}(t)}{t} dt - \int_1^x \frac{E(t)}{t} dt$. Now,

$$\begin{aligned}\int_{t=1}^{t=x} \frac{\mathrm{li}(t)}{t} dt &= \int_{t=1}^{t=x} \int_{y=2}^{y=t} \frac{dy}{\log y}\frac{dt}{t} = \int_{y=2}^{y=x} \int_{t=y}^{t=x} \frac{dy}{\log y}\frac{dt}{t} = \int_{y=2}^{y=x} \frac{(\log x - \log y)dy}{\log y} \\ &= \int_{y=2}^{y=x} \frac{\log x}{\log y} dy - (x-2) = \log(x)\,\mathrm{li}(x) - x + 2\end{aligned}$$

Therefore, $\vartheta(x) = x + E_0(x)$ where $E_0(x) = E(x)\log(x) - 2 - \int_1^x \frac{E(t)}{t}dt$. This shows that a bound on $E(x)$ can be used to create a bound on $E_0(x)$, the error function for $\vartheta(x)$.

Indeed, suppose that $\vartheta(x) = x + E_0(x)$. Then $\pi(x) = \int_{2^-}^x \frac{1}{\log t}d\vartheta(t)$ whence

$$\pi(x) = \frac{\vartheta(x)}{\log x} + \int_2^x \frac{\vartheta(t)}{t(\log t)^2}dt = \frac{x}{\log x} + \int_2^x \frac{1}{(\log t)^2}dt + \frac{E_0(x)}{\log x} + \int_2^x \frac{E_0(t)}{t(\log t)^2}dt.$$

Now, notice that via integration by parts,

$$\mathrm{li}(x) = \int_2^x \frac{dt}{\log t} = \frac{t}{\log t}\bigg|_2^x + \int_2^x \frac{1}{(\log t)^2}dt = \frac{x}{\log x} + \int_2^x \frac{1}{(\log t)^2}dt - \frac{2}{\log 2}.$$

That is, $\pi(x) = \mathrm{li}(x) + \frac{2}{\log 2} + \frac{E_0(x)}{\log x} + \int_2^x \frac{E_0(t)}{t(\log t)^2}dt$, so $E(x) = \frac{2}{\log 2} + \frac{E_0(x)}{\log x} + \int_2^x \frac{E_0(t)}{t(\log t)^2}dt$. This shows that a bound on $E_0(x)$ can be used to create a bound on $E(x)$. In conclusion, $\pi(x) \sim \mathrm{li}(x) \Leftrightarrow \vartheta(x) \sim x$. Indeed, more precisely, we can interpolate errors from one function to the other, which will allow us to prove that $\pi(x) = \mathrm{li}(x) + O(xe^{-C\sqrt{\log x}})$. In summary, we have the following result:

Theorem 4.3 (Transferal of Errors Between ϑ and π)**.** *Let $\pi(x) = \mathrm{li}(x) + E(x)$ and $\vartheta(x) = x + E_0(x)$. Then,*

$$E_0(x) = E(x)\log(x) - 2 - \int_1^x \frac{E(t)}{t}dt \qquad E(x) = \frac{2}{\log 2} + \frac{E_0(x)}{\log x} + \int_2^x \frac{E_0(t)}{t(\log t)^2}dt$$

Furthermore, we would like to know that ϑ and ψ are not very different; this will let us transfer bounds between ψ and π as well.

Proposition 4.4.

$$\psi(x) - \vartheta(x) = O\left(x^{1/2}(\log x)^2\right)$$

Proof. It is a matter of direct computation:

$$\psi(x) = \sum_{n\le x}\Lambda(n) = \sum_{p\le x}\log p + \sum_{p^2\le x}\log p + \sum_{p^3\le x}\log p + \cdots + = \vartheta(x) + \vartheta(\sqrt{x}) + \vartheta(\sqrt[3]{x}) + \cdots$$

There are $\log_2(x) + O(1) = O(\log x)$ nonzero terms in this sum after $\vartheta(x)$. Each of these are less than $\sum_{n\le x^{1/2}}\log n \le x^{1/2}\log x$, so the total is bounded above by $x^{1/2}(\log x)^2$. The result follows. □

This shows that $\pi(x) \sim \mathrm{li}(x)$ is equivalent to $\psi(x) \sim x$; the errors transfer too. Indeed, our discussions about the Riemann Hypothesis and bounds on $\pi(x)$ translate to $\psi(x)$. Precisely, the Riemann Hypothesis is equivalent to the proposition that $\pi(x) = \frac{x}{\log x} + O(\sqrt{x}\log x)$ which is equivalent to the proposition that $\psi(x) = x + O(\sqrt{x}(\log x)^2)$.

4.3 Asymptotic Behavior of $\pi(x)$

Now, we might not know that $\psi(x) \sim x$ yet, but we do know that $\psi(x) \asymp x$ from the Chebyshev bounds. Therefore, we can already use our technique of error transferal to show that $\pi(x) \asymp \frac{x}{\log x}$.

Theorem 4.5. *$\pi(x) \asymp \mathrm{li}(x) \asymp \frac{x}{\log x}$; indeed, more precisely, we have the following bounds:*

$$(\log(2) + o(1))\,\mathrm{li}(x) \le \pi(x) \le (2\log 2 + o(1))\,\mathrm{li}(x),$$
$$(\log(2) + o(1))\left(\frac{x}{\log x}\right) \le \pi(x) \le (2\log 2 + o(1))\left(\frac{x}{\log x}\right).$$

Proof. First, the Chebyshev bounds state that $(\log 2 + o(1))x \le \psi(x) \le (\log 4 + o(1))x$. Furthermore, $\psi(x) - \vartheta(x) = O\left(x^{1/2}(\log x)^2\right)$, so also $(\log 2 + o(1))x \le \vartheta(x) \le (\log 4 + o(1))x$. As before, $\pi(x) = \mathrm{li}(x) + E(x)$ and $\vartheta(x) = x + E_0(x)$. Then, via our discussion above,

$$E_0(x) = E(x)\log(x) - 2 - \int_1^x \frac{E(t)}{t}dt \qquad E(x) = \frac{2}{\log 2} + \frac{E_0(x)}{\log x} + \int_2^x \frac{E_0(t)}{t(\log t)^2}dt.$$

Now, $(\log 2 + o(1))x \leq \vartheta(x)$ implies $E_0(x) \geq ((\log 2 - 1) + o(1))x$. Let $c = (\log 2 - 1) + o(1)$. Then,

$$E(x) \geq \frac{2}{\log 2} + \frac{c}{\log x} + \int_x^2 \frac{c}{(\log t)^2} dt = \frac{2}{\log 2} + \frac{cx}{\log x} + c\,\mathrm{li}(x) - \frac{cx}{\log x} - c\,\mathrm{li}(2) - \frac{2c}{\log 2} = c\,\mathrm{li}(x) - C$$

for some constant C. But then, $\pi(x) = \mathrm{li}(x) + E(x) \geq (\log 2 + o(1))\,\mathrm{li}(x) - C$. Similarly, $(\log 4 + o(1))x \geq \vartheta(x)$ implies $(\log 4 + o(1))\,\mathrm{li}(x) \geq \pi(x)$. This demonstrates the first bound; for the second bound, we use the previously discussed fact that $\mathrm{li}(x) - \frac{x}{\log x} = O\left(\frac{x}{(\log x)^2}\right)$, which implies that $\mathrm{li}(x) - \frac{x}{\log x} = o(\mathrm{li}(x)) = o\left(\frac{x}{\log x}\right)$ (and therefore any perturbations caused by swapping them are contained in the $o(1)$ term). □

Corollary 4.5.1. *Let p_n denote the nth prime. Then $p_n \asymp n \log n$; indeed, more precisely,*

$$(\log 2 + o(1))p_n \leq n \log n \leq (\log 4 + o(1))p_n$$

$$\left(\frac{1}{\log 4} + o(1)\right) n \log n \leq p_n \leq \left(\frac{1}{\log 2} + o(1)\right) n \log n.$$

Proof. For the first part, notice that $\pi(p_n) = n$ implies $\pi(p_n) = n \asymp \frac{p_n}{\log p_n}$ whence $p_n \asymp n \log p_n \asymp n \log(n \log p_n) \asymp n \log n + n \log \log p_n \asymp n \log n$. For the next part, we use the following bound:

$$(\log 2 + o(1)) \left(\frac{x}{\log x}\right) \leq \pi(x) \leq (\log 4 + o(1)) \left(\frac{x}{\log x}\right)$$

Now, since $\pi(p_n) = n$, we can argue the following:

$$(\log 2 + o(1)) \left(\frac{p_n}{\log p_n}\right) \leq \pi(p_n) = n \Rightarrow (\log 2 + o(1))p_n \leq n \log p_n.$$

$$(\log 4 + o(1)) \left(\frac{p_n}{\log p_n}\right) \geq \pi(p_n) = n \Rightarrow (\log 4 + o(1))p_n \geq n \log p_n.$$

Then, we can use the first bound for p_n on its own right-hand side:

$$(\log 2 + o(1))p_n \leq n \log p_n \leq n \log \left(\frac{n \log p_n}{\log 2 + o(1)}\right) = n \log n + n \log \log p_n - n \log(\log 2 + o(1)) \sim n \log n.$$

Similarly, we can use the second bound for p_n on its own right-hand side:

$$(\log 4 + o(1))p_n \geq n \log p_n \geq n \log \left(\frac{n \log p_n}{\log 4 + o(1)}\right) = n \log n + n \log \log p_n - n \log(\log 4 + o(1)) \sim n \log n.$$

Then $(\log 2 + o(1))p_n \leq n \log n \leq (\log 4 + o(1))p_n$. Hence

$$\left(\frac{1}{\log 4} + o(1)\right) n \log n \leq p_n \leq \left(\frac{1}{\log 2} + o(1)\right) n \log n.$$

□

Proposition 4.6. *Let $\#n$ denote the primorial of n (the product of all primes up to n), so that $\#p_k$ denotes the product of the first k prime numbers. Then,*

$$\exp(((\log 2)^2 + o(1))k \log k) \leq \#p_k \leq \exp(((\log 4)^2 + o(1))k \log k).$$

Proof. This is easy: $\vartheta(x) = \sum_{p \leq x} \log(p)$ implies that $\exp(\vartheta(p_k)) = \#p_k$. Since $\vartheta \geq (\log 2 + o(1))x$ and $p_k \geq (\log 2 + o(1))k \log k$ (both facts noted and discussed in further detail in the first problem),

$$\#p_k \geq \exp(\vartheta(p_k)) \geq \exp((\log 2 + o(1))x) \geq \exp((\log 2 + o(1))^2 k \log k) = \exp(((\log 2)^2 + o(1))k \log k).$$

Similarly, $\#p_k \leq \exp(((\log 4)^2 + o(1))k \log k)$. □

Corollary 4.6.1.

$$\omega(n) \ll \frac{\log n}{\log\log n}.$$

Proof. Now, since $\omega(n)$ peaks at primorials (that is, $\#p_k$ is the smallest n such that $\omega(n) = k$) and $\frac{\log n}{\log\log n}$ is increasing, it suffices to show that $\omega(n) \ll \frac{\log n}{\log\log n}$ only looking at primorials $\#p_k$. But $\omega(\#p_k) = k$, whereas

$$\frac{\log(\#p_k)}{\log(\log(\#p_k))} \geq \frac{((\log 2)^2 + o(1))k\log k}{\log(((\log 4)^2 + o(1))k\log k)} = \frac{((\log 2)^2 + o(1))k\log k}{\log(k\log k) + O(1)} = \frac{((\log 2)^2 + o(1))k\log k}{\log(k) + \log\log(k) + O(1)}$$

Yet, for sufficiently large k, this is greater than to $\frac{k\log k}{2\log k} = \frac{k}{2}$. Therefore, for sufficiently large k, $\frac{\log(\#p_k)}{\log(\log(\#p_k))} \geq \frac{k}{2}$ whence we may conclude the desired result using the "peaking" reasoning mentioned above:

$$\omega(\#p_k) \ll \frac{\log(\#p_k)}{\log(\log(\#p_k))} \Rightarrow \omega(n) \ll \frac{\log(n)}{\log(\log(n))}$$

□

5 Details of the Riemann ζ-Function

5.1 Log-Derivatives and the Mertens Function

Theorem 5.1 (Jensen-Cahen)**.** *Suppose that s_0 is a complex number such that the partial sums $\sum_{n\le N}\frac{f(n)}{n^{s_0}}$ are bounded. Let $\sigma_0 = \Re(s_0)$.*

1. *On the open half-plane $\sigma > \sigma_0$, $F(s) = \sum_{n=1}^{\infty}\frac{f(n)}{n^s}$ converges, is analytic, and its derivative is equal to $F'(s) = \sum_{n=1}^{\infty}\frac{d}{ds}\left(\frac{f(n)}{n^s}\right) = \sum_{n=1}^{\infty}\frac{-\log n f(n)}{n^s}$ (that is, it can be computed termwise).*

2. *The series $F(s) = \sum_{n=1}^{\infty}\frac{f(n)}{n^s}$ converges absolutely for $\sigma > \sigma_0 + 1$.*

Proof. This is just analysis, so we give a reference: Keith Conrad's Analytic Number Theory notes, 2.6.8. □

As an application of the above result, using $\log n = (\Lambda\star 1)(n)$ and $\Lambda(n) = (\log\star\mu)(n)$, we may conclude that

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

This final expression is called the "log-derivative" of $\zeta(s)$. Let us briefly discuss the log-derivative's relationship with polynomials, as this shines a light on why the zeroes of the ζ-function are so important. Consider the example of a real polynomial P. Say $P(s) = (s-r_1)\cdots(s-r_k)$. Then, $\frac{d}{ds}\log P(s) = \frac{d}{ds}\log(s-r_1)+\cdots+\frac{d}{ds}\log(s-r_k) = \frac{1}{s-r_1}+\cdots+\frac{1}{s-r_k}$. In other words, we can compute the log-derivative using the roots of the function – if this continues to hold for complex-valued functions like the ζ-function, then analyzing the roots of $\zeta(s)$ can help us compute the Dirichlet series of $\Lambda(n)$.

Theorem 5.2. *Let $M(x) = \sum_{n\le x}\mu(n)$ be the* Mertens function*. Then the prime number theorem (in the form $\psi(x)\sim x$) is equivalent to $M(x) = o(x)$.*

Proof. First, recall that $\Lambda = \mu\star\log$. It would be nice to replace log with a function which is more "arithmetic". Recall that $d(n)$ has size around $\log n$, which makes it a good candidate. Precisely, $\sum_{n\le N}\log n = N\log N - N + O(\log N)$ and $\sum_{n\le N} d(n) = N\log N + (2\gamma-1)N + O(\sqrt{N})$. Therefore, $\sum_{n\le N}(\log n - d(n) + 2\gamma) = O(\sqrt{N})$. Define $e(n) = \log n - d(n) + 2\gamma$; then, our above work becomes $\sum_{n\le x} e(n) = O(\sqrt{x})$.

Now,

$$(\mu\star\log)(n) = (\mu\star(d-2\gamma+e))(n) = (\mu\star d)(n) - 2\gamma(\mu\star 1)(n) + (\mu\star e)(n)$$

Of course $\mu\star 1 = \delta$, and $(\mu\star d) = (\mu\star 1\star 1) = (\mu\star 1)\star 1 = \delta\star 1 = 1$, so

$$(\mu\star\log)(n) = 1 - 2\gamma\delta(n) + (\mu\star e)(n).$$

Therefore,

$$\psi(x) = \sum_{n\le x}\Lambda(n) = \sum_{n\le x}(1-2\gamma\delta(n) + (\mu\star e)(n)) = \lfloor x\rfloor - 2\gamma + \sum_{n\le x}(\mu\star e)(n).$$

Now, we show that the following result holds under the assumption $M(x) = o(x)$, which shows that $M(x) = o(x) \Rightarrow \psi(x)\sim x$.

Lemma 5.3. *$\sum_{n\le x}(\mu\star e)(n) = o(x)$.*

Proof. Fix $\varepsilon > 0$. Now,

$$\sum_{n\le x}(\mu\star e)(n) = \sum_{n\le x}\sum_{ab=n}\mu(a)e(b) = \sum_{ab\le x}\mu(a)e(b)$$

which we will approximate using the hyperbola method. Therefore, let A, B be such that $AB = x$.

As usual, there are three sums to approximate. First, the sum of terms where $a \le A$. This is equal to

$$\sum_{a\le A}\mu(a)\sum_{b\le \frac{x}{a}} e(b) = \sum_{a\le A}\mu(a)\sqrt{\frac{x}{a}} \le \sqrt{x}\sum_{a\le A}\frac{1}{\sqrt{a}} \le \sqrt{x}\left(1+\int_1^A \frac{1}{\sqrt{t}}dt\right) = \sqrt{x}\left(2\sqrt{A}-1\right) \le 2\sqrt{xA} = \frac{2x}{\sqrt{B}}.$$

Next, the sum of terms where $b \in B$. This is where we use the assumption $M(x) = o(x)$. Indeed, by assumption there exists A' such that $|\sum_{n\le y}\mu(n)| \le \varepsilon y$ for all $y > A'$. Then, for B such that $BA' < x$, we have $\frac{x}{b} > \frac{x}{B} > A'$ for all $b \le B$, so

$$\left|\sum_{b\le B} e(b)\sum_{a\le \frac{x}{b}}\mu(n)\right| \le \sum_{b\le B}\frac{\varepsilon x|e(b)|}{b} \le \varepsilon x\sum_{b\le B}|e(b)| \le \varepsilon x\left(\sum_{b\le B}\log b+\sum_{b\le B} d(b)\right) \le 3\varepsilon x B\log B$$

where the last inequality is for sufficiently large B. Now, the final set of terms is the set where $a \le A$ and $b \le B$. Yet this simply reduces to $\sum_{b\le B} e(b)\sum_{a\le A}\mu(a) \le C\sqrt{B}\varepsilon A = \frac{C\varepsilon x}{\sqrt{B}}$ as long as $A > A'$, for some constant C.

Now, combining everything, we find that $\sum_{n\le x}(\mu\star e)(n) \le \frac{(2+C\varepsilon)x}{\sqrt{B}} + 3\varepsilon x B\log B$. Then pick $B = \frac{1}{\sqrt{\varepsilon}}$. Then, for sufficiently large x the hypothesis $A > A'$ is always met, and we plug in this value for B to get

$$\sum_{n\le x}(\mu\star e)(n) \le \varepsilon^{1/4}(2+C\varepsilon)x + \frac{3}{2}\varepsilon^{1/2}\log(1/\varepsilon)x = \left(\varepsilon^{1/4}(2+C\varepsilon) + \frac{3}{2}\varepsilon^{1/2}\log(1/\varepsilon)\right)x.$$

for sufficiently large x. Define $f(\varepsilon) := \varepsilon^{1/4}(2+C\varepsilon) + \frac{3}{2}\varepsilon^{1/2}\log(1/\varepsilon)$, and notice that as $\varepsilon \to 0$, $f(\varepsilon) \to 0$. Therefore, for any $\varepsilon' > 0$, $\sum_{n\le x}(\mu\star e)(n)$ is eventually smaller than $\varepsilon' x$ (by picking the right ε to have $f(\varepsilon) < \varepsilon'$). Therefore, $\sum_{n\le x}(\mu\star e)(n) = o(x)$, and the result follows. □

Next, we will prove the other direction. First, recall that our assumption is that $\psi(x) \sim x$; this means precisely that for any $\varepsilon > 0$, there exists z such that $y \ge z$ implies $|\psi(y) - y| \le \varepsilon y$. We will use this fact in a moment. Before that, however, we derive a new identity involving μ and Λ. Indeed, using our formula for differentiation of Dirichlet series,

$$\frac{1}{\zeta(s)} = \sum_{n=1}^{\infty}\frac{\mu(n)}{n^s} \Rightarrow \left(\frac{1}{\zeta(s)}\right)' = \sum_{n=1}^{\infty}\frac{-\mu(n)\log n}{n^s}$$

Yet on the other hand, using the Chain Rule,

$$\left(\frac{1}{\zeta(s)}\right)' = -\frac{\zeta'(s)}{\zeta(s)^2} = \left(-\frac{\zeta'(s)}{\zeta(s)}\right)\left(\frac{1}{\zeta(s)}\right) = \left(\sum_{n=1}^{\infty}\frac{\Lambda(n)}{n^s}\right)\left(\sum_{n=1}^{\infty}\frac{\mu(n)}{n^s}\right).$$

Combining these results, we have that $-\mu(n)\log(n) = (\mu\star\Lambda)(n)$. Now,

$$\sum_{n\le x}\mu(n)\log(n) = -\sum_{ab\le x}\mu(a)\Lambda(b) = -\sum_{a\le x}\mu(a)\sum_{b\le x/a}\Lambda(b) = -\sum_{a\le x}\mu(a)\psi(x/a).$$

We rewrite this latter sum in two parts:

$$\sum_{a\le x}\mu(a)\psi(x/a) = \sum_{a\le x}\mu(a)\left(\psi\left(\frac{x}{a}\right) - \frac{x}{a}\right) - \sum_{a\le x}\mu(a)\frac{x}{a}.$$

Now, $\sum_{a\le x}\mu(a)\frac{x}{a}$ can be computed using the following chain of equalities:

$$1 = \sum_{n\le x}\delta(n) = \sum_{n\le x}\sum_{a|n}\mu(a) = \sum_{a\le x}\sum_{a|n}\mu(a) = \sum_{a\le x}\mu(a)\left\lfloor\frac{x}{a}\right\rfloor = \sum_{a\le x}\mu(a)\left(\frac{x}{a} - O(1)\right) = \sum_{a\le x}\mu(x)\frac{x}{a} - O(x).$$

Therefore, $\sum_{a\le x}\mu(a)\frac{x}{a}=O(x)$ (which implies, as a side note, that $\sum_{a\le x}\frac{\mu(a)}{a}=O(1)$).

On the other hand,

$$\begin{aligned}\left|\sum_{a\le x}\mu(a)\left(\psi\left(\frac{x}{a}\right)-\frac{x}{a}\right)\right| &\le \sum_{a\le x}\left|\mu(a)\psi\left(\frac{x}{a}\right)-\frac{x}{a}\right| \le \sum_{a\le x/z}\left|\psi\left(\frac{x}{a}\right)-\frac{x}{a}\right| + \sum_{x/z<a\le x}\left|\psi\left(\frac{x}{a}\right)-\frac{x}{a}\right| \\ &\le \sum_{a\le x/z}\varepsilon\cdot\frac{x}{a}+O\left(\sum_{x/z\le a\le x}\frac{x}{a}\right)=\varepsilon x\log(x/z)+O(x(\log x-\log(x/z))) \\ &=\varepsilon x\log(x/z)+O(x\log z)=\varepsilon x\log(x)+O(x\log z).\end{aligned}$$

Therefore, combining our work, we find that $\left|\sum_{n\le x}\mu(n)\log n\right|=\varepsilon x\log x+O(x\log z)$. Next, we show

$$\left|\sum_{n\le x}\mu(n)\log n\right|\sim\left|\sum_{n\le x}\mu(n)\log x\right|=\left|M(x)\log x\right|.$$

For this, notice that

$$\sum_{n\le x}\mu(n)\log x=\sum_{n\le x}\mu(n)\log n+\sum_{n\le x}\mu(n)\left(\log x-\log n\right).$$

Then, $\sum_{n\le x}\mu(n)\left(\log x-\log n\right)=O\left(\sum_{n\le x}\left(\log x-\log n\right)\right)=O\left(x\log x-x\log x+O(x)\right)=O(x)$, where the latter part follows from Stirling's approximation. This shows that $|M(x)\log x|=\varepsilon x\log x+O(x\log z)$ for any $\varepsilon>0$, so indeed $|M(x)|=\varepsilon x+O(x\log z/\log x)\le 2\varepsilon x$ for any $\varepsilon>0$. Therefore, if $\psi(x)\sim x$, it follows that $M(x)=o(x)$, so we have completed this direction too. □

Of course, we expect this to be true: we expect the nonzero values of μ to be 1 about half the time, and -1 about half the time, and therefore mostly cancel out as $x\to\infty$. This provides heuristic evidence for the prime number theorem (but, of course, does not suffice as a proof).

Now, using similar techniques, we can actually give an elementary proof of the Prime Number Theorem. This proof was discovered by Erdös and Selberg in the 1940s (to learn more, read the articles on the subject by Goldfeld and Levinson). However, this proof is not instructive; it does not develop useful techniques that are used elsewhere. Therefore, we will instead give a non-elementary proof using complex analysis on the Riemann ζ-function, which is much more fun and also more instructive.

5.2 Meromorphic Continuation of the Riemann ζ-Function

Recall that

$$\zeta(s)=\sum_{n=1}^{\infty}=\prod_{p}\left(1-\frac{1}{p^s}\right)^{-1}.$$

Now, let $s=\sigma+it$. Then $\frac{1}{n^{\sigma+it}}=n^{-\sigma}e^{-it\log n}$, so we can see that the terms spiral around 0 getting closer and closer to 0 for $t\ne 0$. One might hope that for $t\ne 0$, there might be a way to make this conditionally converge, and therefore find an analytic continuation of $\zeta(s)$.

Now, we will indeed find another function defined on a larger region than $\Re(s)>1$ which agrees with $\zeta(s)$ on $\Re(s)>1$. Since analytic continuations are unique if they exist, we are justified in calling this continuation ζ too. We will find this continuation using partial summation. Indeed,

$$\zeta(s)=\int_{1^-}^{\infty}\frac{1}{y^s}d\left\lfloor y\right\rfloor=\left[\frac{\lfloor y\rfloor}{y^s}\right]_{1^-}^{\infty}-\int_{1}^{\infty}\lfloor y\rfloor\,d\left(\frac{1}{y^s}\right)=\left[\frac{\lfloor y\rfloor}{y^s}\right]_{1^-}^{\infty}+s\int_{1}^{\infty}\frac{\lfloor y\rfloor}{y^{s+1}}dy.$$

Now, of course

$$s\int_1^\infty \frac{\lfloor y\rfloor}{y^{s+1}}dy = s\int_1^\infty \frac{1}{y^s}dy - s\int_1^\infty \frac{\{y\}}{y^{s+1}}dy.$$

The first one is equal to $s\cdot\frac{s}{s-1}$ for all $\Re s > 1$. The second one converges for all $\Re s > 0$. Furthermore, $\left[\frac{\lfloor y\rfloor}{y^s}\right]_{1^-}^{\infty} = 0$ for all $\Re s > 1$. Therefore, we find that for all $\Re s > 1$, $\zeta(s) = s\cdot\frac{s}{s-1} - s\int_1^\infty \frac{\{y\}}{y^{s+1}}dy$, but this function converges for all $\Re s > 0$ with the exception of $s \neq 1$. Therefore, we can define this as a meromorphic continuation of the Riemann ζ-function, as desired.

Indeed, we want to find a meromorphic continuation of the Riemann ζ-function defined *everywhere* on the complex plane (except, of course, at the pole $s = 1$). For this, we want to find an expression for

$$\int_1^\infty \frac{\{y\}}{y^{s+1}}dy$$

that converges everywhere. For this, we write

$$\int_1^\infty \frac{\{y\}}{y^{s+1}}dy = \int_1^\infty \frac{\{y\}-\frac{1}{2}}{y^{s+1}}dy + \int_1^\infty \frac{\frac{1}{2}}{y^{s+1}}dy.$$

The latter evaluates to $\frac{1}{2s}$ for all s, and the former is equal to

$$\int_1^\infty \frac{1}{y^{s+1}}d\left(\int_1^y \left(\{t\}-\frac{1}{2}\right)dt\right) = (s+1)\int_1^\infty \frac{\int_1^y\left(\{t\}-\frac{1}{2}\right)dt}{y^{s+2}}dy = (s+1)\int_1^\infty \frac{\{y\}^2-\{y\}}{2y^{s+2}}dy.$$

Putting everything together, this yields a new expression for $\zeta(s)$:

$$\frac{s}{s-1} - s\left((s+1)\int_1^\infty \frac{\{y\}^2-\{y\}}{2y^{s+2}}dy + \frac{1}{2s}\right).$$

This agrees our definition of $\zeta(s)$ on $\Re s > 0$ and is defined for $\Re s > -1$ except at $s = 1$ (notice that there is no pole at $s = 0$ since the s and the $\frac{1}{2s}$ cancel out). We can repeat this procedure of Euler-Maclaurin summation to yield the desired meromorphic continuation of the Riemann ζ-function. Later, we will see a functional equation relating $\zeta(s)$ and $\zeta(s-1)$ that allows us to do this procedure more easily, but since for the prime number theorem we only need $\zeta(s)$ for $\Re s > 0$, we postpone this discussion later.

This functional equation immediately yields that the Riemann ζ-function has zeroes at $-2, -4, -6, \ldots$; these are called the trivial zeroes.

Conjecture 5.1 (Riemann Hypothesis)**.** *The most famous unsolved problem in mathematics, the Riemann hypothesis, states that other than the trivial zeroes, all the zeroes of $\zeta(s)$ are on the line $\Re(s) = \frac{1}{2}$.*

5.3 Bounds and Estimations for the Riemann ζ-Function

First, we will demonstrate a formula for estimating $\zeta(s)$ for $\Re s > 0$.

Theorem 5.4. *For any N and $s \neq 1$, $\zeta(s) = \sum_{n\le N}\frac{1}{n^s} - \frac{N^{1-s}}{s-1} - s\int_N^\infty \frac{\{y\}}{y^{s+1}}dy = \sum_{n\le N}\frac{1}{n^s} - \frac{N^{1-s}}{s-1} - O\left(\frac{1}{N^s}\right)$.*

Proof. Write

$$\begin{aligned}\zeta(s) &= \sum_{n\le N}\frac{1}{n^s} + \int_{N^+}^\infty \frac{1}{y^s}d\lfloor y\rfloor = \sum_{n\le N}\frac{1}{n^s} + \left[\frac{\lfloor y\rfloor}{y^s}\right]_{N^+}^{\infty} + s\int_N^\infty \frac{\lfloor y\rfloor}{y^{s+1}}dy \\ &= \sum_{n\le N}\frac{1}{n^s} - N^{1-s} + s\int_N^\infty \frac{1}{y^s}dy - s\int_N^\infty \frac{\{y\}}{y^{s+1}}dy \\ &= \sum_{n\le N}\frac{1}{n^s} - \frac{N^{1-s}}{s-1} - s\int_N^\infty \frac{\{y\}}{y^{s+1}}dy\end{aligned}$$

Now, $\int_N^\infty \frac{\{y\}}{y^{s+1}}dy = O(\frac{1}{N^s})$. □

Next, we'll find some upper bounds for $\zeta(s)$ (mostly as a function of $\Re s$).

Proposition 5.5. *Let $s = \sigma + it$ for $\sigma > 1$. Then, $|\zeta(\sigma + it)| = O\left(\frac{1}{\sigma - 1}\right)$.*

Proof.

$$|\zeta(s + it)| = \left|\sum_{n=1}^{\infty} \frac{1}{n^{\sigma+it}}\right| \leq \sum_{n=1} \left|\frac{1}{n^{\sigma+it}}\right| = \sum_{n=1}^{\infty} \frac{1}{n^{\sigma}}$$

whence $|\zeta(\sigma + it)| \leq \zeta(\sigma) = O\left(\frac{1}{\sigma-1}\right)$. □

Proposition 5.6. *If $\sigma > 1$, then $|\zeta(\sigma + it)| \geq \frac{\zeta(2\sigma)}{\zeta(\sigma)}$.*

Proof. Recall that $\zeta(s) = \prod_p \left(\frac{1}{1-p^{-s}}\right)$. Therefore,

$$|\zeta(\sigma + it)| = \prod_p \left|\frac{1}{1 - p^{-\sigma - it}}\right| \qquad \text{and} \qquad \frac{\zeta(2\sigma)}{\zeta(\sigma)} = \prod_p \frac{1 - p^{-\sigma}}{1 - p^{-2\sigma}}.$$

To show the described inequality, it therefore suffices to show that term-by-term the second product is bigger. That is, it suffices to show that

$$\left|\frac{1}{1 - p^{-\sigma - it}}\right| \geq \frac{1 - p^{-\sigma}}{1 - p^{-2\sigma}} \Leftrightarrow 1 - p^{-2\sigma} \geq |1 - p^{-\sigma - it}|(1 - p^{-\sigma})$$

Yet of course $|1 - p^{-\sigma - it}| \leq 1 + p^{-\sigma}$ by the triangle inequality, whence indeed we have, as desired,

$$|1 - p^{-\sigma - it}|(1 - p^{-\sigma}) \leq (1 + p^{-\sigma})(1 - p^{-\sigma}) = 1 - p^{-2\sigma}.$$

□

Proposition 5.7. *Let $s = 1 + it$. Then, for all $|t| \geq \frac{1}{2}$, $|\zeta(1 + it)| \leq \log(1 + |t|) + O(1)$.*

Proof. Using our above approximation for $\zeta(s)$, we have

$$\zeta(1 + it) = \left|\sum_{n \leq N} \frac{1}{n^{1+it}} - \frac{N^{-it}}{it} - (1 + it)\int_N^{\infty} \frac{\{y\}}{y^{2+it}} dy\right| \leq \sum_{n \leq N} \frac{1}{n} + \frac{1}{|t|} + (1 + |t|)\int_N^{\infty} \frac{1}{y^2} dy.$$

Now, we have $|t| \geq \frac{1}{2}$, and we assign $N = \lfloor 1 + |t| \rfloor$. Then, plugging this into the above formula, we find that $\zeta(1 + it) \leq \log(1 + |t|) + O(1) + \frac{1 + \lfloor t \rfloor}{N}$, implying that $|\zeta(1 + it)| \leq \log(1 + |t|) + O(1)$ for all $|t| \geq \frac{1}{2}$. □

Proposition 5.8. *Let $s = \sigma + it$ for $0 < \varepsilon < \sigma < 1 - \varepsilon < 1$, and suppose that $|s - 1| \geq \frac{1}{2}$. Then,*

$$|\zeta(\sigma + it)| \ll_{\varepsilon} (1 + |t|)^{1-\sigma}.$$

Proof. Suppose that $|\sigma + it - 1| \geq \frac{1}{2}$. Then,

$$\begin{aligned}
|\zeta(\sigma + it)| &\leq \sum_{n \leq N} \frac{1}{n^{\sigma}} + \frac{N^{1-\sigma}}{|\sigma + it - 1|} + (\sigma + |t|)\int_N^{\sigma} \frac{dy}{y^{1+\sigma}} \\
&\leq \frac{N^{1-\sigma} - 1}{1 - \sigma} - \frac{N^{1-\sigma}}{1/2} + (\sigma + |t|)\frac{N^{-\sigma}}{\sigma}.
\end{aligned}$$

Now, suppose that σ lies between ε and $1 - \varepsilon$ for some $\varepsilon > 0$. Then,

$$\begin{aligned}
|\zeta(\sigma + it)| &\leq \frac{N^{1-\sigma} - 1}{\varepsilon} - \frac{N^{1-\sigma}}{1/2} + (\sigma + |t|)\frac{N^{-\sigma}}{\varepsilon} \\
&= O\left(\frac{1}{\varepsilon}(1 + |t|)^{1-\sigma}\right)
\end{aligned}$$

where the final part comes from assigning $N = \lfloor 1 + |t| \rfloor$. The result then follows. □

Following is the most important bound on $\zeta(s)$, combining all our previous results:

Theorem 5.9. *For all $\sigma > \varepsilon$ such that $|\sigma + it - 1| \geq \frac{1}{2}$,*

$$|\zeta(\sigma+it)| \ll_{\varepsilon} ((1+|t|)^{1-\sigma}+1)\log(1+|t|)+1.$$

Proof. The range $\sigma \in (\varepsilon, 1-\varepsilon)$ is solved by Proposition 5.8, the line $\sigma = 1$ is solved by Proposition 5.7, and the line $\sigma > 1$ is solved by Proposition 5.5. Now, suppose that $\sigma \in (1, 1+\varepsilon]$. Define $\sigma' = \sigma - 1$, so that $s = 1 + \sigma' + it$. Then, for any N and any $s \neq 1$, we have $\zeta(s) = \sum_{n \leq N} \frac{1}{n^s} - \frac{N^{1-s}}{s-1} - O\left(\frac{1}{N^s}\right)$ by Theorem 5.4. Taking absolute values and recalling that $|s-1| \geq \frac{1}{2}$, we have

$$|\zeta(s)| \leq \sum_{n \leq N} \frac{1}{n^{1+\sigma'}} + 2N^{-\sigma'} + O\left(N^{-\sigma}\right) \leq \sum_{n \leq N} \frac{1}{n} + O(N^{-\sigma'}) = \log(N) + O(1)$$

but of course by setting $N = \lfloor 1 + |t| \rfloor$, we have that $|\zeta(s)| \ll_{\varepsilon} \log(1+|t|)$, which suffices.

Therefore, it only remains to handle the case of $\sigma \in [1-\varepsilon, 1)$. Again, define $\sigma' = \sigma - 1$, so that $s = 1 + \sigma' + it$. Similarly, we also apply absolute values to Theorem 5.4 and recall that $|s-1| \geq \frac{1}{2}$ to achieve that

$$|\zeta(s)| \leq \sum_{n \leq N} \frac{1}{n^{1+\sigma'}} + 2N^{-\sigma'} + O\left(N^{-\sigma}\right) \leq N^{-\sigma'} \sum_{n \leq N} \frac{1}{n} + O(N^{-\sigma'}) = N^{-\sigma'} \log(N) + O(1)$$

but of course by setting $N = \lfloor 1 + |t| \rfloor$ and recalling the definition $\sigma' = \sigma - 1$, we have that $|\zeta(s)| \ll_{\varepsilon} (1+|t|)^{1-\sigma} \log(1+|t|)$, as desired. Therefore, we are done. □

5.4 Almost-Periodicity of the Riemann ζ-Function

Lemma 5.10. *Let $\|x\| = \min_{n \in \mathbb{Z}} |x - n|$ denote the distance between x and the nearest integer. Then, given real numbers $\alpha_1, \ldots, \alpha_K$ and any integer $N \geq 1$, there exists n with $1 \leq n \leq N^K$ such that $\|n\alpha_j\| \leq 1/N$ for each $j = 1, \ldots, K$.*

Proof. Suppose that $\alpha_1, \ldots, \alpha_K$ are real numbers, and $N \geq 1$ is an integer. For any integer n, we write $[n] = \{1, \ldots, n\}$ and $[n] - 1 = \{0, \ldots, n-1\}$. The key is to define a function

$$\phi : [N^k] \to ([N]-1)^K \text{ given by } n \mapsto (\lfloor N\{n\alpha_i\} \rfloor)_{i=1}^{K}$$

where $\{r\}$ denotes the fractional part of a real number r. Now, there are two cases:

Case 1: $(0, \ldots, 0)$ **lies within the image of** ϕ**.** Then let n be such that $\phi(n) = (0, \ldots, 0)$. Then n satisfies $1 \leq n \leq N^K$ and is such that $N\{na_j\} < 1$ for each j; this implies that $\{na_j\} < 1/N$ for each $j = 1, \ldots, K$. Yet this plainly implies that $\|na_j\| < 1/N$ for each $j = 1, \ldots, K$. Therefore, we are done.

Case 2: $(0, \ldots, 0)$ **does not lie within the image of** ϕ**.** In this case, ϕ maps $[N^K]$, a set of cardinality N^k, into $([N]-1)^K \setminus \{(0, \ldots, 0)\}$, a set of cardinality $N^K - 1$. Therefore, ϕ cannot be injective, and there exist $1 \leq n_1 < n_2 \leq N$ such that $\phi(n_1) = \phi(n_2)$. Then $N\{n_1\alpha_i\}$ and $N\{n_2\alpha_i\}$ have the same floor, so $|N\{n_2\alpha_i\} - N\{n_1\alpha_i\}| < 1$, so that $|\{n_2\alpha_i\} - \{n_1\alpha_i\}| < 1/N$. Yet this that $\{(n_2 - n_1)\alpha_i\}$ is either less than $1/N$ or more than $1 - 1/N$; either way, if we define $n = n_2 - n_1$, $\|na_j\| \leq 1/N$ for each $j = 1, \ldots, K$. Since we have chosen n in such a way that guarantees $1 \leq n \leq N^K$, we are done. □

Lemma 5.11. *For complex $|x| \leq \frac{1}{2}$, $|1 - e^x| \leq 2|x|$.*

Proof. This is a very coarse bound, and it also follows immediately from the Maclaurin series for e^z:

$$\begin{aligned} |1 - e^x| &= \left| x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots \right| \leq |x| + \left| \frac{x^2}{2!} \right| + \left| \frac{x^3}{3!} \right| + \cdots \leq |x| + |x|^2(1 + |x| + |x|^2 + \cdots) \\ &= |x| + \frac{|x|^2}{1 - |x|} \leq |x| + 2|x|^2 \leq |x| + |x| \leq 2|x|. \end{aligned}$$

□

Theorem 5.12. *Fix $\sigma > 1$ and $\varepsilon > 0$. Then there exists a nonzero real number $T = T(\sigma, \varepsilon)$ such that*

$$|\zeta(\sigma + it) - \zeta(\sigma + it + iT)| \leq \varepsilon$$

for all $t \in \mathbb{R}$. We say that ζ is almost periodic *on the line $\Re(s) = \sigma$, and that T is an ϵ-*almost period *of it.*

Proof. Fix $\sigma > 1$ and $\varepsilon > 0$. Then, for any j and any $s = \sigma + it$,

$$\left|\zeta(s) - \sum_{n \leq j} \frac{1}{n^s}\right| \leq \left|\sum_{n > j} \frac{1}{n^s}\right| \leq \sum_{n > j} \frac{1}{|n^s|} = \sum_{n > j} \frac{1}{n^\sigma} \leq \int_j^\infty \frac{1}{x^\sigma} dx = \left.\frac{x^{1-\sigma}}{1-\sigma}\right|_j^\infty = \frac{j^{1-\sigma}}{\sigma - 1}.$$

The final expression plainly tends to 0 as $j \to \infty$ since $\sigma > 1$. Therefore, there exists some j for which $\left|\zeta(s) - \sum_{n \leq j} \frac{1}{n^s}\right|$ is less than $\frac{\varepsilon}{4}$ for any value of t. Fix this value of j.

Next, fix an integer $N \geq 10$, and consider the real numbers $\frac{\log(n)}{2\pi}$ for $n = 1, \ldots, j$. By Lemma 5.10, there exists some T such that $\|T \log(n)/2\pi\| < 1/N$ for all $n = 1, \ldots, j$. Yet, for such T,

$$\begin{aligned}
\left|\sum_{n \leq j} n^{-(\sigma+it)} - \sum_{n \leq j} n^{-(\sigma+it+iT)}\right| &= \left|\sum_{n \leq j} n^{-(\sigma+it)} \left(1 - n^{-iT}\right)\right| \\
&= \left|\sum_{n \leq j} n^{-(\sigma+it)} \left(1 - e^{-iT\log(n)}\right)\right| \\
&= \sum_{n \leq j} \left|n^{-(\sigma+it)}\right| \left|1 - e^{-iT\log(n)}\right| \\
&= \sum_{n \leq j} \left|n^{-(\sigma+it)}\right| \cdot \frac{4\pi}{N} = \frac{4\pi\zeta(\sigma)}{N}
\end{aligned}$$

where the second-to-last inequality follows from the fact that $\|T\log(n)/2\pi\| < 1/N$ implies that $T\log(n)$ is within $2\pi/N$ of a multiple of 2π, which means that it can be replaced by a real number x satisfying $|x| < 2\pi/N$ without changing the value of the expression, and then applying Lemma 5.11.

But then there exists some N, dependent on σ and ε, such that $\frac{4\pi\zeta(\sigma)}{N} < \frac{\varepsilon}{2}$. Take the associated T; I claim that this is the ε-almost period we are hunting for. Indeed, all that remains is the Triangle Inequality:

$$\begin{aligned}
&|\zeta(\sigma + it) - \zeta(\sigma + it + iT)| \\
\leq &\left|\zeta(\sigma + it) - \sum_{n \leq j} \frac{1}{n^{\sigma+it}}\right| + \left|\sum_{n \leq j} n^{-(\sigma+it)} - \sum_{n \leq j} n^{-(\sigma+it+iT)}\right| + \left|\zeta(\sigma + it + iT) - \sum_{n \leq j} \frac{1}{n^{\sigma+it+iT}}\right| \\
< &\frac{\varepsilon}{4} + \frac{4\pi\zeta(\sigma)}{N} + \frac{\varepsilon}{4} < \frac{\varepsilon}{4} + \frac{\varepsilon}{2} + \frac{\varepsilon}{4} < \varepsilon.
\end{aligned}$$

□

6 Perron's Formula and Applications

6.1 Relating Dirichlet Series and Partial Sums

We would like to relate Dirichlet series $F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}$ of $f(n)$ with the partial sums $A(x) = \sum_{n \le x} f(n)$; in particular, we will use this to prove the Prime Number Theorem by applying it to the special case of $\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n} = -\frac{\zeta'(s)}{\zeta(s)}$, the Dirichlet series of $\Lambda(n)$, and therefore getting information about $\psi(x) = \sum_{n \le x} \Lambda(n)$ (which we have already seen will allow us to deduce the prime number theorem).

Theorem 6.1 (Perron's Formula)**.** *Suppose that $c > 0$, and let $\int_{(c)} f(s)ds$ denote the limit $\int_{c-i\infty}^{c+i\infty} f(s)ds$ of the line integrals $\int_{c-Ti}^{c+Ti} f(s)ds$ along the vertical line $c - Ti$ to $c + Ti$ as $T \to \infty$. Then, if $y > 0$,*

$$\frac{1}{2\pi i} \int_{(c)} \frac{y^s}{s} ds = \begin{cases} 1 & y > 1 \\ \frac{1}{2} & y = 1 \\ 0 & y < 1. \end{cases}$$

Proof. Notice that $\frac{y^s}{s}$ has a unique pole at $s = 0$, and this pole is simple, so we can compute its residue to be $\lim_{s \to 0} y^s = 1$. Now, we will perform casework on the possible values of y.

Case 1: $0 < y < 1$.
Let γ_d be the contour from $d - iT$ to $d + iT$ to $c + iT$ to $c - iT$ to $d - iT$ again (all straight lines, so γ is a rectangle) for $d > c$. Then, since γ_d does not contain any poles of $\frac{y^s}{s}$, $\int_{\gamma_d} \frac{y^s}{s} ds = 0$. Therefore,

$$\int_{c-iT}^{c+iT} \frac{y^s}{s} ds = \int_{c-iT}^{d-iT} \frac{y^s}{s} ds + \int_{d-iT}^{d+iT} \frac{y^s}{s} ds - \int_{c+iT}^{d+iT} \frac{y^s}{s} ds.$$

The absolute value of the first integral on the right-hand side is bounded above by

$$\left| \int_{c-iT}^{d+iT} \frac{y^s}{s} \right| = \le \int_c^d \frac{y^\sigma}{T} d\sigma \le \frac{1}{T} \int_c^\infty y^\sigma \le \frac{y^c}{T|\log y|}.$$

The absolute value of the third integral can be bounded above using exactly the same bound. To bound above the second integral, notice that

$$\left| \int_{d-iT}^{d+iT} \frac{y^s}{s} ds \right| = y^d \left| \int_{d-iT}^{d+iT} \frac{ds}{s} ds \right| = y^d O(\log T) = O(y^d \log T).$$

Then, let $d = T$ (for sufficiently large T so that $d > c$), and we see that the absolute value of the integral on the left-hand side is bounded above as so:

$$\left| \int_{c-iT}^{c+iT} \frac{y^s}{s} ds \right| \le \frac{2y^c}{T|\log y|} + O(y^T \log T)$$

which goes to 0 as $T \to \infty$, implying that $\int_{(c)} \frac{y^s}{s} ds = 0$ in this case, as desired.

Case 2: $y > 1$.
Let γ_d be the contour from $c - iT$ to $c + iT$ to $d + iT$ to $d - iT$ to $c - iT$ for $d < 0 < c$ (we switch the order so that the contour is still counterclockwise). Since γ_d contains the simple pole at $s = 0$,

$$\int_{c-iT}^{c+iT} \frac{y^s}{s} ds + \int_{c+iT}^{d+iT} \frac{y^s}{s} ds + \int_{d+iT}^{d-iT} \frac{y^s}{s} ds + \int_{d-iT}^{c-iT} \frac{y^s}{s} ds = 2\pi i.$$

We can bound each of the three integrals except the first as decreasing functions in $-d$ and T (just as did for the first case), and then by setting $d = -T$ and driving $T \to \infty$, we find that $\int_{(c)} \frac{y^s}{s} ds = 2\pi i$, as desired.

Case 3: $y = 1$.

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{1}{s}ds = \frac{1}{2\pi}\int_{-T}^{T}\frac{dt}{c+iT} = \frac{1}{2\pi}\int_0^T\left(\frac{1}{c+it}+\frac{1}{c-it}\right)dt = \frac{1}{2\pi}\int_0^T\frac{2c}{c^2+t^2}dt = \frac{1}{\pi}\arctan\left(\frac{T}{c}\right)$$

Of course, $\lim_{T\to\infty}\frac{1}{\pi}\arctan\left(\frac{T}{c}\right) = \frac{1}{\pi}\cdot\frac{\pi}{2} = \frac{1}{2}$, whence $\frac{1}{2\pi i}\int_{(c)}\frac{1}{s}ds = \frac{1}{2}$. Hence we are done. □

Indeed, we have the following more precise result containing the explicit bounds:

Proposition 6.2 (Perron's Formula with Error Bounds)**.**

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{y^s}{ds} = \begin{cases} 1+O\left(y^c\min\left\{1,\frac{1}{T|\log y|}\right\}\right) & y>1 \\ \frac{1}{2}+O\left(\frac{c}{\pi T}\right) & y=1 \\ 0+O\left(y^c\min\left\{1,\frac{1}{T|\log y|}\right\}\right) & y<1 \end{cases}$$

Proof. This is, of course, essentially the same as our previous theorem, except now the bounds for $y>1$ and $y<1$ do not blow up near $y=1$. Firstly, let us verify the bound for $y=1$. This is easy:

$$\frac{1}{2\pi}\int_T^\infty\frac{2c}{c^2+t^2}dt < \frac{2}{\pi}\int_T^\infty\frac{dt}{t^2} = \frac{c}{\pi T}.$$

Now, to verify the bounds for $y>1$ and $y<1$, it suffices (in light of the previous theorem) to show that the error terms are bounded above by $O(y^c)$. In the case $y<1$, we take the contour γ which is a line from $c+iT$ to $c-iT$ and then an arc of the circle with center 0 and radius $\sqrt{c^2+T^2}$ from $c-iT$ to $c+iT$ to the right (notice that this is not quite a semicircle). Then, of course,

$$\left|\frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{y^s}{s}\right| < \frac{1}{2\pi}\frac{y^c}{\sqrt{c^2+T^2}}\cdot 2\pi\sqrt{c^2+T^2} = y^c$$

since the length of the curve being integrated over is $2\pi\sqrt{c^2+T^2}$, and the size of the integrand can be bounded above by $\frac{y^c}{\sqrt{c^2+T^2}}$. In the case $y>1$, one uses the contour γ which is a line from from $c-iT$ to $c+iT$ and then an arc of the circle with center 0 and radius $\sqrt{c^2+T^2}$ from $c+iT$ to $c-iT$ passing to the left (notice that this is just more than a semicircle). We then bound it similarly. □

Our plan is to use these results to get a useful expression for $A(x)$ in terms of $F(s)$. The key is to recognize that $n\le x$ if and only if $\frac{x}{n}\ge 1$. Now, Perron's formula almost gives us an indicator of whether or not a value is at most 1; indeed, there is a slight issue at the boundary. To resolve this, we let $\sum_{n\le x}^* f(n)$ denote a sum such that if x is an integer, then the last term $f(x)$ is only counted halfway. Then,

$$\sum_{n\le x}^* f(n) = \sum_{n=1}^\infty f(n)\cdot\frac{1}{2\pi i}\int_{(c)}\frac{(x/n)^s}{s}ds = \frac{1}{2\pi i}\int_{(c)}\sum_{n=1}^\infty\frac{f(n)}{n^s}\cdot\frac{x^s}{s}ds = \frac{1}{2\pi i}\int_{(c)}F(s)\frac{x^s}{s}ds$$

Theorem 6.3 (Perron's Second Formula)**.** *Suppose $F(s)=\sum_{n=1}^\infty\frac{f(n)}{n^s}$ converges absolutely for $\Re(s)>\sigma_0\ge 0$ and $c>\sigma_0$. Then, for all $x>0$,*

$$\sum_{n\le x}^* f(n) = \sum_{n=1}^\infty\frac{f(n)}{2\pi i}\int_{(c)}\frac{(x/n)^s}{s}ds = \frac{1}{2\pi i}\int_{(c)}F(s)\frac{x^s}{s}ds$$

Ironically, we're not actually going to use Perron's Second Formula often; instead, to get the right error bounds, we will use Perron's Formula with Error Bounds directly with the derivation above. This will allow us to estimate $\psi(x)$ using poles of $-\frac{\zeta'(s)}{\zeta(s)}$ (which we will see can be found by looking at zeroes of $\zeta(s)$).

6.2 Examples Using Perron's Formula

Let's start with a dumb example to illustrate the general procedure:

Example 6.4 (A Dumb Example)**.** Suppose that we want to count $\sum_{n\le x} 1$, pretending for the moment that it is not obviously $\lfloor x \rfloor$. Then, using the fact that $\zeta(s) = \sum \frac{1}{n^s}$, we have that

$$\sum_{n\le x}{}^{*} 1 = \sum_{n=1}^{\infty} \frac{1}{2\pi i}\int_{(c)} \frac{(x/n)^s}{s} ds = \frac{1}{2\pi i}\int_{(c)} \zeta(s)\frac{x^s}{s} ds.$$

Now, of course by applying Perron's Formula with Error Bounds to the middle expression, this is equal to

$$\sum_{n\le x}{}^{*} 1 + O\left(\sum_{n=1}^{\infty}\left(\frac{x}{n}\right)^c \min\left(1, \frac{1}{T|\log\frac{x}{n}|}\right)\right)$$

Now, this only makes sense for $c > 1$, but at the same time, we would like to integrate along the line $d - iT$ to $d + iT$ for some $0 < d < 1$, since otherwise the error term we get is unreasonably large. Therefore, we are going to use a contour γ which travels from $c - iT$ to $c + iT$ to $d + iT$ to $d - iT$; this contour encircles the pole of $\zeta(s)$ at $s = 1$ with residue 1, and we can bound the necessary integrals to get the desired result.

In general, when applying Perron's formula, there are three steps: (1) simplify the error term, (2) bound the auxiliary integrals, and (3) choose T to balance (and thereby minimize) the error terms.

Let us proceed with the first step. First, assume that $\{x\} = \frac{1}{2}$ – that is, x is directly between integers; this is a useful tactic which we will do regularly when applying Perron's Formula. This is useful because it allows us to drop the minimum in our error term, leaving us the new error term $\sum_{n=1}^{\infty}(\frac{x}{n})^c\frac{1}{T|\log\frac{x}{n}|}$. Next, we split the terms of this sum into two parts: $n > 0.9x$ or $n > 1.1x$ and $0.9x \le n \le 1.1x$. In the first case, we have

$$\sum_{\substack{n<0.9x\\ n>1.1x}}\left(\frac{x}{n}\right)^c \frac{1}{T\left|\log\frac{x}{n}\right|} \le \sum_{\substack{n<0.9x\\ n>1.1x}}\left(\frac{x}{n}\right)^c \frac{1}{T} \ll \frac{x^c}{T}\sum_{n=1}^{\infty}\frac{1}{n^c} = \frac{x^c\zeta(c)}{T} \ll \left(\frac{x^c}{c-1}\right)\frac{1}{T} \ll \frac{x\log x}{T}$$

where the final step follows by choosing $c = 1 + \frac{1}{\log x}$ and noticing that $x^{\frac{1}{\log x}} = x^{\log_x(e)} = e$ is a constant.

In the second case, write $n = \lfloor x \rfloor + k$ where k is an integer satisfying $|k| \le 0.1x$. Then,

$$\left|\log\frac{x}{n}\right| = \left|\log\left(\frac{\lfloor x\rfloor + \frac{1}{2}}{\lfloor x\rfloor + k}\right)\right| = \left|\log\left(1 - \frac{k-\frac{1}{2}}{\lfloor x\rfloor + k}\right)\right| \asymp \frac{|k-\frac{1}{2}|}{x}$$

Furthemore, for these terms, $\left(\frac{x}{n}\right)^c$ is bounded above and below by constants (so they are asymptotic with 1). This allows us to conclude that

$$\sum_{0.9x\le n\le 1.1x}\left(\frac{x}{n}\right)^c\frac{1}{T\left|\log\frac{x}{n}\right|} \ll \sum_{0.9x\le n\le 1.1x}\frac{1}{T\left|\log\frac{x}{n}\right|} \ll \sum_{0.9x\le n\le 1.1x}\frac{x}{T\left|k-\frac{1}{2}\right|} \ll \frac{x\log x}{T}.$$

Therefore, in summary, the error term is $O\left(\frac{x\log x}{T}\right)$.

Now, the next step is to bound the vertical and horizontal integrals. Recall our key bound on ζ: for any $\varepsilon > 0$, $|\zeta(\sigma + it)| \ll_\varepsilon (1 + |t|)^{1-\sigma}\log(1 + |t|) + 1$ for $\sigma > \varepsilon$. It is then an elementary matter to show that the vertical integral at d is $O\left(x^d T^{1-d}\log T\right)$, and the horizontal integrals are $O\left(\frac{x\log x}{T} + \frac{x^d T^{1-d}\log T}{T}\right)$. Putting everything together, the result is that

$$\sum_n \delta\left(\frac{x}{n}\right) = \sum_{n\le x} 1 + O\left(\frac{x}{\log x}T\right) = x + O\left(\frac{x}{\log x}T + x^d T^{1-d}\log T\right)$$

where the first inequality comes from our work above, and the second inequality comes from integrating the contour. This demonstrates that $\sum_{n\leq x} 1 = x + O\left(\frac{x\log x}{T} x^{\varepsilon} T^{1-\varepsilon}\log T\right)$. Now, to minimize the error, we balance the error terms with a smart choice of T; in this case, the best choice is $T = \sqrt{x}$, yielding the final result $\sum_{n\leq x} 1 = x + O(x^{1/2+\varepsilon})$.

Example 6.5. Let $d_k(n)$ be the kth divisor function $\sum_{n=a_1\cdots a_k} 1$. Notice that $d_k(n) = (1\star 1\star\cdots\star 1)(n)$ (where the right-hand side is the convolution of k copies of 1), so $\zeta(s)^k = \sum_{n=1}\frac{d_k(n)}{n^s}$. In the appendix, we show that $d_k(n) \ll_\varepsilon n^\varepsilon$ for any $\varepsilon > 0$, so this Dirichlet sum indeed converges for all $\Re(s) > 1$.

The goal is to understand $\sum_{n\leq x} d_k(n)$ by using Perron's formula to compute

$$\sum_{n\leq x} d_k(n) = \frac{1}{2\pi i}\int_{c-iT}^{c+iT}\frac{x^s\zeta(s)^k}{s}ds + O\left(\sum_{n=1}^{\infty} d_k(n)\left(\frac{x}{n}\right)^c \frac{1}{T|\log\frac{x}{n}|}\right).$$

where $x = x' + \frac{1}{2}$ for some $x' \in \mathbb{N}$ and $c > 1$. Just as before, the tactic is to compute the integral using the contour $c - iT$ to $c + iT$ to $d + iT$ to $d - iT$ back to $c - iT$, where $d < 1$.

First, we need to compute the residue of $s = 1$ of $\frac{\zeta(s)^k x^s}{s}$. Now, we know the Laurent expansion of $\zeta(s) = \frac{1}{s-1}+\gamma+\cdots+$. Similarly, the Laurent expansion of $\frac{1}{s}$ at $s = 1$ is $\frac{1}{s} = \frac{1}{1+(s-1)} = 1-(s-1)+(s-1)^2-\cdots$. Finally, $x^s = x\cdot x^{s-1} = xe^{(s-1)(\log x)} = x\left(1 + (s-1)\log x + \frac{1}{2!}((s-1)^2\log(x)^2) + \cdots\right)$. Multiplying all these together, we find that the main term is $xP_{k-1}(\log x)$ where P_{k-1} is a polynomial of degree $k - 1$.

Then, we choose $c = 1 + \frac{1}{\log x}$ and $d = \varepsilon$. As usual, there are two cases. When $n < 0.9x$ or $n > 1.1x$

$$\sum_{\substack{n<0.9x\\ n>1.1x}} d_k(n)\left(\frac{x}{n}\right)^c\frac{1}{T\log|\frac{x}{n}|} \ll \frac{x^c}{T}\sum\frac{d_k(n0}{n^c} \ll \frac{x^c}{T}\zeta(c)^k \ll \frac{x^c}{T}\left(\frac{1}{c-1}\right)^k \ll \frac{x(\log x)^k}{T}$$

In the other case, $0.9x \leq n \leq 1.1x$, we have $\sum_{0.9x\leq n\leq 1.1x}$ can be bounded above by $\ll \frac{x^{1+\varepsilon}}{T}$. Then,

$$\left|\int_{d-iT}^{d+iT}\zeta(s)^k\frac{x^s}{s}ds\right| \ll \int_{-T}^{T}|\zeta(\varepsilon+it)|^k\frac{x^\varepsilon}{|\varepsilon+it|}dt \ll_\varepsilon \int_{-T}^{T}(1+|t|)^{k(1-\varepsilon)}\log T\frac{x^\varepsilon}{1+|t|}dt \ll x^\varepsilon(\log T)^k T^{k(1-\varepsilon)-1}$$

The resulting horizontal integrals are similar to the $\zeta(s)$ case (they are very easy). The final error is $\frac{x^{1+\varepsilon}}{T} + x^\varepsilon T^{k(1-\varepsilon)+\varepsilon}$. To minimize this, we let $x = T^{\frac{1}{k+1}}$, so that the final error term is $O(x^{\frac{k}{k+1}+\varepsilon})$. This proves that

$$\sum_{n\leq x} d_k(n) = xP_{k-1}(\log x) + O\left(x^{\frac{k}{k+1}+\varepsilon}\right).$$

Example 6.6. Let $F(s) = \sum_{n=1}^{\infty}\frac{d(n)^2}{n^s} = \prod_p\left(1+\frac{d(p)^2}{p^s}+\frac{d(p^2)^2}{p^{2s}}+\frac{d(p^3)^2}{p^{3s}}\right) = \prod_p\left(1+\frac{2^2}{p^s}+\frac{3^2}{p^{2s}}+\cdots\right)$. This is absolutely convergent for $\Re(s) > 1$ since $d(n) \ll_\varepsilon n^\varepsilon$. Now,

$$(1+\frac{4}{p^s}+\frac{9}{p^{2s}}+\cdots)(1-\frac{4}{p^s}-\cdots) = (1+\frac{c}{p^{2s}}+\cdots+)$$

for all $\Re(s) > 1/2$. But the second is just $\zeta(s)^{-4}$. Thus, if $\prod_p(1+\frac{c}{p^{2s}}+\cdots) = G(s)$, then $F(s) = \zeta(s)^4 G(s)$ where $G(s)$ is absolutely convergent for $\Re(s) > \frac{1}{2}$. Then, $\sum_{n\leq x} d(n)^2 = \frac{1}{2\pi i}\int_{c-iT}^{c+iT}\zeta(s)^4 G(s)\frac{x^s}{s}ds + O\left(\frac{x^{1+\varepsilon}}{T}\right)$ and $c = 1 + \frac{1}{\log x}$. Then we let d be between $\frac{1}{2}$ and 1; we use the usual contour.

Next we seek to compute the residue at $s = 1$ of $\zeta(s)^4 G(s)\cdot\frac{x^s}{s}$. Now, $\zeta(s)^4 = \left(\frac{1}{s-1}+\gamma+\cdots\right)^4$ and $x^s = xe^{(s-1)(\log x)} = x(1 + (s-1)\log x + \frac{1}{2!}(s-1)^2(\log x)^2 + \cdots)$. Since $G(s)$ is holomorphic at 1, $G(s) =$

$g_0 + g_1(s-1) + g_2(s-1)^2 + \cdots$. Putting everything together, we can show that the residue's principal term is $cx(\log x)^3 + \cdots$ for some constant c. In other words, the residue is $xP_3 \log(x)$ for some polynomial P_3 of degree 3. Then, if we write out all the integrals, everything operates in essentially the same way as usual, and we get $\sum_{n \le x} d(n)^2 = xP_3(\log x) + O\left(x^{5/6+\varepsilon}\right)$. Hence $\sum_{n\le x} d(n)^2 \asymp x(\log x)^3$.

Problem 1. Generalize this to compute $\sum_{n\le x} d_k(n)^2$.

6.3 Counting Finite Abelian Groups

Theorem 6.7. *The number of finite abelian groups of order at most x is asymptotic to $(\zeta(2)\zeta(3)\cdots)x = (2.29485\dots)x$.*

Proof. Let $a(n)$ denote the number of finite abelian groups of order n, and $A(x) = \sum_{n\le x} a(n)$. Via the Structure Theorem for Finitely Generated Abelian Groups implies, any abelian group of order $n = p_1^{e_1} \cdots p_k^{e_k}$ can be written as the product over $i = 1, \dots, k$ of abelian groups order $p_i^{e_i}$. With elementary group theory, it is clear that this implies that $a(n)$ is multiplicative. Our next question, then, is how many abelian groups of order p^e are there? The answer is fairly obvious: it is the number of ways to write e as the sum of positive integers (where order does not matter). In other words, $a(p^e)$ is equal to the number of partitions of e.

Hence, by assigning $z = p^{-s}$ and letting $\operatorname{part}(e)$ be the number of partitions of e, we have the Euler product

$$\sum_{n=1}^{\infty} \frac{a(n)}{n^s} = \prod_p \left(\sum_e \operatorname{part}(e) z^e \right).$$

It is an easy combinatorial result of Young Diagrams that if e is any partition of i and e_i is the number of is in e, then there is an involution on the set of partitions given by $e \leftrightarrow e_1 + 2e_2 + 3e_3 + \cdots + ie_i$. But this implies that

$$\sum_e \operatorname{part}(e) z^e = (1 + z + z^2 + \cdots)(1 + z^2 + z^4 + \cdots)(1 + z^3 + z^6 + \cdots) = \prod_{i=1}^{\infty} \frac{1}{1 - z^i}$$

which converges absolutely whenever $|z| < 1$ (which always holds since p^{-s}). Hence

$$\sum_{n=1}^{\infty} \frac{a(n)}{n^s} = \prod_p \prod_{i=1}^{\infty} \frac{1}{1 - p^{-si}} = \prod_{j=1}^{\infty} \prod_p \left(1 - \frac{1}{p^{js}}\right)^{-1} = \prod_{j=1}^{\infty} \zeta(js).$$

Therefore, by Perron's formula, we have $\sum_{n\le x} a(n) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} \prod_{j=1}^{\infty} \zeta(js) \frac{x^s}{s}$. Then, via the usual techniques, the main term of this expression is given by the pole(s) of the integrand; in this case, in the usual contour (we pick $\frac{1}{2} < d < 1$), the only pole is at $s = 1$. Now, it is easy to compute that the residue of $\prod_{j=1}^{\infty} \zeta(js) \frac{x^s}{s}$ at $s = 1$ is $x\zeta(2)\zeta(3)\cdots = (2.29485\dots)x$, so we are done. □

Corollary 6.7.1. *On average, the number of finite abelian groups of a given order is $(\zeta(2)\zeta(3)\cdots)$.*

By pushing d smaller, we can get more accurate results; however, this makes it necessary to compute more poles, as the infinite product has poles at 1, $\frac{1}{2}$, $\frac{1}{3}$, and so on. Now, the residue at the pole $s = \frac{1}{2}$ is $\sqrt{x}\zeta(\frac{1}{2})\zeta(\frac{3}{2})\zeta(\frac{4}{2})\cdots$. Similarly, the residue at $s = \frac{1}{3}$ is $x^{1/3}\zeta(\frac{1}{3})\zeta(\frac{2}{3})\zeta(\frac{4}{3})\cdots$, etc. This last bound is as far as is known currently, because we do not know if the integrals remain bounded. We do know that the integrals become unbounded at 6 poles, but they could indeed become unbounded sooner.

7 Proof of the Prime Number Theorem

7.1 Introduction

First, let us choose a branch of the logarithm on the complex numbers. To motivate our choice, recall that $\sum_{n=1}^{\infty} \frac{\Lambda(n)}{n^s} = -\frac{\zeta'(s)}{\zeta(s)} = -\frac{d}{ds}\log(\zeta(s))$. Now, we have

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

We will choose the branch of the logarithm with the standard Taylor expansion, so that the expression $\log(1-p^{-s})$ becomes $-\sum_{k\geq 1} \frac{1}{k}(\frac{1}{p^s})^k$. Therefore, under this choice, $\log(\zeta(s)) = \sum_p \sum_k \frac{1}{k}\left(\frac{1}{p^s}\right)^k$. Now, normally we will usually use $-\frac{\zeta'(s)}{\zeta(s)}$ directly but this expression is helpful for showing that $\zeta(1+it) \neq 0$.

Now, notice that $\frac{(fg)'}{fg} = \frac{f'}{f} + \frac{g'}{g}$. Then, suppose that f has a zero of degree n at s_0. Then, $f(s) = (s-s_0)^n g(s)$ where g is nonzero and holomorphic in a neighborhood of 0. Then, $\frac{f'(s)}{f(s)} = \frac{n(s-s_0)^{n-1}}{(s-s_1)^n} + \frac{g'}{g}(s) = \frac{n}{s-s_0} + \frac{g'}{g}(s)$. Therefore, $\frac{f'}{f}$ has a simple pole at s_0 with residue n if and only if f has a simple zero at s_0 with multiplicity n. Similarly, f has a pole at s_0 with multiplicity n if and only if $\frac{f'}{f}$ has a simple pole at s with residue $-n$. For example, ζ has a pole at $s=1$, so $-\frac{\zeta'(s)}{\zeta(s)}$ has a pole at $s=1$ with residue -1.

Now,

$$\psi(x) = \sum_{n\leq x} \Lambda(n) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \left(-\frac{\zeta'(s)}{\zeta(s)}\right)\frac{x^s}{s}dx.$$

The integrand has a pole at $s=0$ with residue $-\frac{\zeta'(0)}{\zeta(0)}$. Now, each zero ρ of $\zeta(s)$ with multiplicity $m(\rho)$ gives the log-derivative a simple pole with residue $-m(\rho)$. If there are not many zeroes, then we might be able to conclude that $\psi(x) = x - \sum_p \frac{x^\rho}{p} - (\zeta'/\zeta)(0)$, which is exactly what we are looking for.

Now, first notice that $\sigma + it$ is zero of $\zeta(s)$ if and only if $\sigma - it$ is a zero of $\zeta(s)$; the reason for this is the Schwarz reflection principle. As we will later show with the functional equation, there are also trivial zeroes at $-2, -4, -6, \ldots$. These trivial zeroes are the only zeroes outside of $0 < \Re(s) < 1$, since $\zeta(s) \neq 0$ for $\Re(s) > 1$ (it suffices to notice that $\zeta(s)$ for $\Re(s) > 1$ has an absolutely convergent Euler product with nonzero terms). The functional equation also implies that zeroes in $0 < \Re(s) < 1$ are symmetric about $\Re(s) = \frac{1}{2}$. Nonetheless, we need more information about the roots of $\zeta(s)$; in particular, we need to show that it has no zeroes on the line $\Re(s) = 1$. This is the content of the next section.

7.2 The Riemann ζ-Function is Nonvanishing on $\Re(s) = 1$

In this section, we follow the original proof by Hadamard and de la Vallée Poussin that the Riemann ζ-function is nonvanishing on $\Re(s) = 1$.

Lemma 7.1. *For all $\sigma > 1$,*

$$\zeta(\sigma)^3|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)| \geq 1$$

for all $t \in \mathbb{R}$.

Proof. By taking the logarithm of both sides, we find that the above inequality is equivalent to

$$3\log\zeta(\sigma) + 4\Re(\log\zeta(\sigma+it)) + \Re(\log\zeta(\sigma+2it)) \geq 0.$$

Now, recall that we choose the specific branch of the logarithm which made $\log(\zeta(s)) = \sum_{p,k}\frac{1}{kp^{ks}}$ true. This implies that the quantity above can be rewritten as

$$\sum_{p,k}\frac{1}{k}\left(\frac{3}{p^{k\sigma}} + 4\Re\left(\frac{1}{p^{k(\sigma+it)}}\right) + \Re\left(\frac{1}{p^{k(\sigma+2it)}}\right)\right) = \sum_{p,k}\frac{1}{kp^{k\sigma}}\left(3 + 4\cos(kt)\log p) + \cos(2tk\log p)\right).$$

Therefore, it suffices to show that for any $\theta \in \mathbb{R}$, $3 + 4\cos(\theta) + \cos(2\theta) \geq 0$. But this is easy:

$$3 + 4\cos(\theta) + \cos(2\theta) = 3 + 4\cos(\theta) + 2\cos(\theta)^2 - 1 = 2(1 + \cos\theta)^2 \geq 0.$$

Therefore we are done. □

Theorem 7.2 (Hadamard and de la Vallée Poussin)**.** *Suppose that $\sigma + it$ is a zero of $\zeta(s)$. Then $\sigma < 1$.*

Proof. Of course, $\zeta(s) \neq 0$ for any $\sigma > 1$ (by the Euler product), so it suffices to show that $\zeta(s) \neq 0$ for $\sigma = 1$. Now, recall that $\zeta(s) = \frac{1}{s-1} + \gamma + \cdots$, so ζ is nonzero for all s sufficiently close to 1. Therefore, we may assume that there exists $\delta > 0$ such that $|t| \geq \delta$.

Under this assumption, assume for the sake of contradiction that $\zeta(1 + it) = 0$. Then, for σ close to 1,

$$\zeta(\sigma + it) = \zeta(\sigma - 1 + 1 + it) = C(\sigma - 1) + \text{higher order terms}$$

for some constant C depending on t. In particular, $|\zeta(\sigma + it)| \leq C(\sigma - 1)$. Furthermore, $|\zeta(\sigma + 2it)| \leq \log(2 + |t|)$, and $\zeta(\sigma) = \frac{1}{\sigma-1} + \gamma + \cdots$. Therefore, applying the previous lemma

$$1 \leq \zeta(\sigma)^3|\zeta(\sigma + it)|^4|\zeta(\sigma + 2it)| \leq \left(\frac{1}{\sigma - 1}\right)^3 (\sigma - 1)^4 C^4 = C^4(\sigma - 1)$$

which is obviously false if σ is sufficiently close to 1. This yields the desired contradiction, so we are done. □

7.3 Further Technical Prerequisites

In this section, we perform some laborious computations which are necessary for the final proof.

Lemma 7.3. *Suppose that $c > 0$, $y > 0$, and $b > 0$. Then,*

$$\frac{1}{2\pi i}\int_{(c)} \frac{y^s}{s+b} ds = \begin{cases} y^{-b} & y > 1 \\ \frac{1}{2} & y = 1 \\ 0 & y < 1 \end{cases} \quad \textit{and} \quad \frac{1}{2\pi i}\int_{(c)} \frac{y^s}{s(s+b)} ds = \begin{cases} \frac{1}{b}\left(1 - y^{-b}\right) & y \geq 1 \\ 0 & y \leq 1. \end{cases}$$

Proof. For the first fact, recall from Perron's Formula that we have

$$\frac{1}{2\pi i}\int_{(c)} \frac{y^s}{s} ds = \begin{cases} 1 & y > 1 \\ \frac{1}{2} & y = 1 \\ 0 & y < 1. \end{cases}$$

Then, if $b > 0$, via the substitution $s + b = t$

$$\frac{1}{2\pi i}\int_{s=c-i\infty}^{s=c+i\infty} \frac{y^s}{s+b} ds = \frac{1}{2\pi i}\int_{t=c+b-i\infty}^{t=c+b+i\infty} \frac{y^{t-b}}{t} dt = y^{-b} \cdot \frac{1}{2\pi i}\int_{t=c+b-i\infty}^{t=c+b+i\infty} \frac{y^t}{t} dt = \begin{cases} y^{-b} & y > 1 \\ \frac{1}{2} & y = 1 \\ 0 & y < 1. \end{cases}$$

For the second fact, we use the identity $\frac{1}{s(s+b)} = \frac{1/b}{s} - \frac{1/b}{s+b}$. This yields the following chain of equalities:

$$\frac{1}{2\pi i}\int_{(c)} \frac{y^s}{s(s+b)} ds = \frac{1}{b} \cdot \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{y^s}{s} ds - \frac{1}{b} \cdot \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{y^s}{s+b} ds = \begin{cases} \frac{1}{b}\left(1 - y^{-b}\right) & y > 1 \\ 0 & y = 1 \\ 0 & y < 1. \end{cases}$$

Yet this is the desired result. □

Corollary 7.3.1. *Suppose that $x > 0$. Then, by assigning $b = 1$ and replacing y with $\frac{x}{n}$, we find that*

$$\frac{1}{2\pi i}\int_{(c)} \left(\frac{x}{n}\right)^s \frac{ds}{s(s+1)} = \begin{cases} \left(1 - \frac{n}{x}\right) & \frac{x}{n} \geq 1 \\ 0 & \frac{x}{n} \leq 1 \end{cases} = \begin{cases} \left(1 - \frac{n}{x}\right) & n \leq x \\ 0 & n \geq x. \end{cases}$$

Lemma 7.4. *For $|t| \geq 1$,*

$$|\zeta'(1+it)| \ll (\log(1+|t|))^2 \sim (\log|t|)^2.$$

Proof. Recall that for $\Re(s) > 0$ such that $s \neq 1$,

$$\zeta(s) = \sum_{n \leq N} \frac{1}{n^s} - \frac{N^{1-s}}{s-1} - s\int_N^\infty \frac{\{y\}}{y^{s+1}}dy.$$

Therefore, by taking derivatives, we find that

$$\zeta'(s) = \sum_{n \leq N} \frac{-\log(n)}{n^s} + \frac{N^{1-s}\log(N)}{s-1} + \frac{N^{1-s}}{(s-1)^2} - \int_N^\infty \frac{\{y\}}{y^{s+1}}dy.$$

Now, choose $s = 1 + it$ for some real t, and take absolute values to obtain

$$\begin{aligned}|\zeta'(s)| &\leq \left|\sum_{n \leq N} \frac{-\log(n)}{n^{1+it}}\right| + \left|\frac{N^{-it}\log(N)}{it}\right| + \left|\frac{N^{-it}}{(it)^2}\right| + \left|\int_N^\infty \frac{1}{y^{2+it}}dy\right| \\ &\leq \log(N)^2 + \frac{\log(N)}{t} + \frac{1}{t^2} + \int_N^\infty \frac{1}{y^2}dy \\ &\leq \log(N)^2 + \frac{\log(N)}{t} + \frac{1}{t^2} + \frac{1}{N}.\end{aligned}$$

Now, recall that $|t| \geq 1$, and define $N = \lfloor 1 + |t| \rfloor$. Then we quickly obtain the desired result:

$$|\zeta'(s)| \leq \log(1+|t|))^2 + \frac{\log(1+|t|)}{t} + \frac{1}{t^s} + \frac{1}{\log(1+|t|)} \ll \log(1+|t|))^2.$$

□

Lemma 7.5. *For all $t > 2$,*

$$|\zeta(1+it)| \gg (\log|t|)^{-8}.$$

Proof. First, recall from Lemma 7.1 that $\zeta(\sigma)^3|\zeta(\sigma+it)|^4|\zeta(\sigma+2it)| \geq 1$, so

$$|\zeta(\sigma+it)| \geq |\zeta(\sigma+2it)|^{-1/4}\zeta(\sigma)^{-3/4} \geq (C\log(1+|t|))^{-1/4}\left(\frac{1}{\sigma-1}\right)^{-3/4} \gg (\sigma-1)^{-3/4}(\log(1+|t|))^{-1/4}.$$

But then $\zeta(1+(\log|t|)^{-10}+it) \gg (\log|t|)^{-7.75}$. But notice that

$$\zeta(1+it) = \zeta\left(1+(\log|t|)^{-10}\right) - \int_0^{(\log|t|)^{-10}} \zeta'(1+\lambda+it)d\lambda.$$

Then, by taking absolute values, we find that

$$|\zeta(1+it)| \geq |\zeta(1+(\log|t|)^{-10}+it)| - \int_0^{(\log|t|)^{-10}} |\zeta'(1+\lambda+it)|d\lambda \gg (\log|t|)^{-7.75} - (\log|t|)^{-8} \sim (\log|t|)^{-7.75}.$$

and we are done as $(\log|t|)^{-7.75} > (\log|t|)^{-8}$ for sufficiently large t. □

Corollary 7.5.1.

$$\left|\frac{\zeta'}{\zeta}(1+it)\right| = \frac{|\zeta'(1+it)|}{|\zeta(1+it)|} \ll (\log(2+|t|))^{12}.$$

7.4 Proving the Prime Number Theorem

We begin with an auxiliary definition:

Definition 7.6. *$\psi_1(x) = \sum_{n\le x} \Lambda(n)(x-n) = \int_0^x \psi(t)dt$.*

Proposition 7.7. *$\psi_1(x) \sim \frac{x^2}{2}$ implies $\psi(x) \sim x$.*

Proof. For $h \le x$, notice that

$$h\psi(x+h) \ge \psi_1(x+h) - \psi_1(x) = \int_x^{x+h} \psi(t)dt \ge h\psi(x).$$

By hypothesis, $\psi_1(x+h)-\psi_1(x) = hx + \frac{h^2}{2} + o(x^2)$. Then $\psi(x) \le x + \frac{h}{2} + o\left(\frac{x^2}{h}\right) \le x + \frac{h}{2} + \frac{\varepsilon x^2}{h}$. Then, by setting $h = \sqrt{\varepsilon}x$, we obtain $\psi(x) \le x + o(x)$. Similarly, we have $\psi(x+h) \ge x + \frac{h}{2} + \frac{\varepsilon x^2}{h}$. But then by replacing $x+h$ with y, we have $\psi(y) \ge y - \frac{h}{2} + \frac{\varepsilon(y-h)^2}{h} = y - \frac{h}{2} + \frac{\varepsilon y^2}{h} - 2\varepsilon y + \varepsilon h$, and then by setting $h = \sqrt{\varepsilon}y$, we obtain $\psi(x) \ge x - o(x)$. Thus $\psi(x) \sim x$. □

Theorem 7.8. *$\psi_1(x) \sim \frac{x^2}{2}$.*

Proof. Now, if $c > 1$, by applying Corollary 7.3.1 and the usual rearrangement of Perron's Formula:

$$\psi_1(x) = x\sum_{n\le x}\Lambda(n)\left(1-\frac{n}{x}\right) = \sum_n \Lambda(n)\left(\frac{1}{2\pi i}\int_{(c)}\left(\frac{x}{n}\right)^s \frac{ds}{s(s+1)}\right)x = x\left(\frac{1}{2\pi i}\int_{(c)} -\frac{\zeta'}{\zeta}(s)x^s\frac{ds}{s(s+1)}\right).$$

Therefore, it suffices to show that $\frac{1}{2\pi i}\int_{(c)} -\frac{\zeta'}{\zeta}(s)x^s\frac{ds}{s(s+1)} \sim \frac{x}{2}$. Notice that $-\frac{\zeta'}{\zeta}(s)x^s\frac{ds}{s(s+1)}$ has a pole at $s=1$ of residue $\frac{x}{2}$, so we expect that this will work. However, to show this rigorously, we will shift the contour. First, define $\delta(T)$ to be such that $\zeta(s)\neq 0$ in the region $\Re(s) > 1-\delta(T)$ and $|\operatorname{Im}(s)| \le T$. This is possible because of (1) Theorem 7.2, (2) the fact that zeroes of meromorphic functions are isolated, and (3) compactness of the interval $[1-iT, 1+iT]$. Then, we define the contour γ to be

$$(1-i\infty)\to(1-iT)\to(1-\delta(T)-iT)\to(1-\delta(T)+iT)\to(1+iT)\to(1+i\infty).$$

The key here is that the only pole of the integrand in the region bounded by the line from $c-i\infty$ to $c+i\infty$ and the contour γ is at $s=1$ (and it has residue $\frac{x}{2}$). This implies that

$$\frac{1}{2\pi i}\int_{(c)} -\frac{\zeta'}{\zeta}(s)x^s\frac{ds}{s(s+1)} = \frac{1}{2\pi i}\int_{\gamma} -\frac{\zeta'}{\zeta}(s)x^s\frac{ds}{s(s+1)} + \frac{x}{2}.$$

Therefore, all that remains is to bound the contour integral. Now, define

$$F(T) = \max_{s\in[1-\delta(T)-iT, 1-\delta(T)+iT]}\left|\frac{\zeta'}{\zeta}(s)\right|$$

Then, the middle vertical piece can be bounded above as follows:

$$\begin{aligned}\left|\int_{1-\delta(T)-iT}^{1-\delta(T)+iT} -\frac{\zeta'}{\zeta}(s)x^s\frac{ds}{s(s+1)}\right| &\le x^{1-\delta(T)}\int_{1-\delta(T)-iT}^{1-\delta(T)+iT}\left|-\frac{\zeta'}{\zeta}(s)\right|\left|\frac{ds}{s(s+1)}\right| \\ &\le F(T)x^{1-\delta(T)}\int_{1-\delta(T)-iT}^{1-\delta(T)+iT}\frac{|ds|}{|s(s+1)|} \\ &\le F(T)x^{1-\delta(T)}2\int_{0.5}^{\infty}\frac{ds}{s(s+1)} \ll x^{1-\delta(T)}.\end{aligned}$$

Similarly, the horizontal pieces can be bounded above as follows:

$$\left|\int_{1-\delta(T)+iT}^{1+iT} -\frac{\zeta'}{\zeta}(s)x^s\frac{ds}{s(s+1)}\right| \ll F(T)\int_{1-\delta(T)}^{1} x^\sigma d\sigma \le F(T)\frac{x}{\log x}.$$

Then, the remaining vertical integrals can be bounded above using Corollary 7.5.1:

$$\left|\int_{1+iT}^{1+i\infty} -\frac{\zeta'}{\zeta}(s)x^s \frac{ds}{s(s+1)}\right| \ll x\int_T^\infty \frac{\left|\frac{\zeta'}{\zeta}(1+it)\right|}{1+t^2}dt \ll x\int_T^\infty \frac{(\log t)^{10}}{t^2}dt \ll x\frac{(\log T)^{10}}{T}$$

Putting everything together, this implies that

$$\frac{1}{2\pi i}\int_{(c)} -\frac{\zeta'}{\zeta}(s)x^s\frac{ds}{s(s+1)} = \frac{x}{2} + O\left(x^{1-\delta(T)}F(T) + \frac{x}{|logx}F(T) + x\frac{(\log T)^{12}}{T}\right).$$

Now fix $\varepsilon > 0$. First, choose T such that $\frac{(\log T)^{10}}{T} < \frac{\varepsilon}{2}$. Then, choose x so large that $\left(\frac{x}{\log x} + x^{1-\delta(T)}\right)F(T) \leq \frac{\varepsilon x}{2}$. Then the error term becomes $O(\varepsilon x)$; since ε was arbitrary, this implies that the error term is $o(x)$, so indeed the integral is asymptotic to $\frac{x}{2}$ and we are done. □

Corollary 7.8.1 (The Prime Number Theorem)**.** $\pi(x) \sim \frac{x}{\log x} \sim \mathrm{li}(x)$.

Proof. This follows from combining Theorem 4.3, Proposition 4.4, Proposition 7.7, and Theorem 7.8. □

7.5 Refining the Proof and Clarifying Error Bounds

Given some additional information, we can provide explicit error bounds for the prime number theorem. This additional information comes in the form of an explicit zero-free region: precisely, suppose that we have shown that $\zeta(\sigma + it) \neq 0$ for $\sigma \geq 1 - \frac{c}{\log|t|+2}$ for some constant c. Also suppose that we have demonstrated that $|\frac{\zeta'}{\zeta}(\sigma+it)| = O((\log|t|)^{10})$ in this entire region. Both of these facts are true, but we will not prove them.

In any case, knowing these facts allows us to choose $\delta(T) = \frac{C}{\log T}$ for some constant C. Then, revisiting our earlier error bounds, we can rewrite them as $x^{1-\frac{C}{\log T}} + \frac{x}{T}(\log T)^{10}$. Now, we want $T = x^{\frac{C}{\log T}}$; solving for T, this yields $T = \exp(\sqrt{\log x})$. Plugging this in gives an error bound of $O(xe^{-c\sqrt{\log x}})$, implying the following:

Theorem 7.9 (Prime Number Theorem with Error Bounds)**.**

$$\psi_1(x) = \frac{x^2}{2} + O(x^2e^{-c\sqrt{\log x}}) \implies \psi(x) + O(xe^{-c\sqrt{\log x}}) \implies \pi(x) = \mathrm{li}(x) + O(x^2e^{-c\sqrt{\log x}}).$$

8 The Functional Equation for the Riemann ζ-Function

In this section, we develop one of the main tools for analyzing the Riemann ζ-function.

8.1 Background on the Gamma Function

Definition 8.1 (Gamma Function)**.** The *Gamma function* is defined by

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

It is straightforward to see that this integral converges (and therefore $\Gamma(s)$ is analytic) for $\Re(s) > 0$. However,

Proposition 8.2. *The Gamma function has a simple pole with residue* 1 *at* $s = 0$.

Proof. This follows from using the Taylor series for e^{-t}, integrating the power series, and then using the Taylor series for $\ln(t)$. □

Proposition 8.3. *The Gamma function satisfies* $\Gamma(s+1) = s\Gamma(s)$.

Proof. This is a routine computation.

$$\Gamma(s+1) = \int_0^\infty t^s d(-e^{-t}) = 0 + \int_0^\infty s t^{s-1} e^{-t} dt = s\Gamma(s).$$

□

Corollary 8.3.1. *For positive integers* s, $\Gamma(s) = (s-1)!$

Proof. Suppose that $s = 1$. Then $\Gamma(s) = \int_0^\infty e^{-t} t^0 dt = \int_0^\infty e^{-t} dt = 1 = 0!$. Now, for the sake of induction, suppose that the result holds for some positive integer s. Then, by the above functional equation, $\Gamma(s+1) = s\Gamma(s) = s(s-1)! = s!$. Therefore, by induction, the result holds. □

This functional equation can be used to give a meromorphic continuation of $\Gamma(s)$ to $\mathbb{C}$. The poles of the Gamma function to the right of the line $\Re(s) = 0$ come from the unique pole in the original definition, $s = 0$.

Corollary 8.3.2. *By the functional equation,* $\Gamma(s)$ *has a simple pole with residue* $\frac{(-1)^n}{n!}$ *at* $-n$ *for each non-negative integer* n.

8.2 Stating the Functional Equation

The functional equation for the Riemann ζ-function is as so:

Theorem 8.4 (Functional Equation for $\zeta(s)$)**.**

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

One way to encode this idea more simply is with the ξ-function:

Definition 8.5 (ξ-function)**.** Let $\xi(s) = s(s-1)\pi^{-s/2}\Gamma(s/2)\zeta(s)$.

Then, the above functional equation simply becomes $\xi(s) = \xi(1-s)$. This functional equation can be used to deduce a few things. First, notice that for $\Re(s) > 0$, $\zeta(s)$ has only one pole at $s = 1$, and this pole is canceled by the term of $(s-1)$ in the definition of $\xi(s)$. Therefore, $\xi(s)$ is analytic on $\Re(s) > 0$. But then the functional equation implies that $\xi(s) = \xi(1-s)$ is analytic on $\Re(s) < 1$; since these regions overlap, this implies that ξ is analytic everywhere. This allows us to deduce the following fact:

Proposition 8.6. $\zeta(-2) = \zeta(-4) = \cdots = 0$.

Proof. Notice that $\Gamma(s/2)$ has poles at $0, -2, -4, -6$, etc. because $\Gamma(s)$ has poles at $0, -1, -2, -3$, etc. Since $\xi(s)$ is analytic, this implies that $s(s-1)\zeta(s)$ must have zeroes at $0, -2, -4, -6$, etc. Yet $s(s-1)$ has zeroes only at 0 and 1, so $\zeta(s)$ must have zeroes at $-2, -4, -6$ and so on, as desired. □

Problem 2. Show that $\zeta(-n) \in \mathbb{Q}$ using our original work on meromorphically continuing ζ. Then, use the functional equation for $\zeta(s)$ to show that $\zeta(2), \zeta(4), \ldots$ are rational multiples of some powwer of π.

8.3 Poisson Summation Formula

Definition 8.7. Recall that if f is a rapidly decreasing function, the *Fourier transform* $\hat{f}$ *of* f is

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} f(x)e^{-2\pi ix\xi}dx.$$

Theorem 8.8 (Poisson Summation Formula)**.** *Suppose f is rapidly decreasing. Then,*

$$\sum_{n\in\mathbb{Z}} f(n) = \sum_{k\in\mathbb{Z}} \hat{f}(k).$$

Proof. Let $F(x) = \sum_{n\in\mathbb{Z}} f(x+n)$; this is absolutely convergent. Furthermore, it is easy to see that $F(x+1) = F(x)$; thus, F is a smooth function $F : \mathbb{R}/\mathbb{Z} \to \mathbb{R}$. The Fourier coefficients of F are then

$$\begin{aligned}\hat{F}(n) &= \int_0^1 F(x)e^{-2\pi inx}dx = \int_0^1 \left(\sum_{n\in\mathbb{Z}} f(x+n)\right) e^{-2\pi ik(x+n)}dx = \sum_{n\in\mathbb{Z}} \int_0^1 f(x+n)e^{-2\pi ik(x+n)}dx \\ &= \sum_{n\in\mathbb{Z}} \int_n^{n+1} f(y)e^{-2\pi iky}dy = \int_{-\infty}^{\infty} f(y)e^{-2\pi iky} = \hat{f}(k).\end{aligned}$$

Yet $F(x) = \sum_{k\in\mathbb{Z}} \hat{F}(k)e^{2\pi ikx}$. Thus $F(x) = \sum_{n\in\mathbb{Z}} f(n+x) = \sum_{k\in\mathbb{Z}} \hat{f}(k)e^{2\pi ikx}$. By plugging in $x = 0$, we achieve the desired result $\sum_{n\in\mathbb{Z}} f(n) = \sum_{k\in\mathbb{Z}} \hat{f}(k)$. □

Example 8.9. In this example, we compute the Fourier transform of a Gaussian $f(x) = e^{-\pi x^2}$. Indeed,

$$\hat{f}(\xi) = \int_{-\infty}^{\infty} e^{-\pi x^2-2\pi ix\xi}dx = \int_{-\infty}^{\infty} e^{-\pi(x+i\xi)^2-\pi\xi^2}dx = e^{-\pi\xi^2}\int_{-\infty+i\xi}^{\infty+i\xi} e^{-\pi z^2}dz = e^{-\pi\xi^2}\left(\int_{-\infty}^{\infty} e^{-\pi z^2}dz\right) = e^{-\pi\xi^2}.$$

8.4 Proving the Functional Equation for $\zeta(s)$)

We begin with a brief technical proposition:

Proposition 8.10. *Define* $\theta(t) = \sum_{n\in\mathbb{Z}} e^{-\pi n^2 t}$

$$\theta(t) = \frac{1}{\sqrt{t}}\theta\left(\frac{1}{t}\right).$$

Proof. First, recall our computation of the Fourier transform of a Gaussian (Example 8.9). Then, substituting $y = \sqrt{t}x$, we see that $f_t(x) = e^{-\pi x^2 t}$ has Fourier transform $\frac{1}{\sqrt{t}}e^{-\pi x^2/t}$. Then, applying the Poisson summation formula yields the desired result:

$$\theta(t) = \sum_{n\in\mathbb{Z}} f_t(n) = \sum_{k\in\mathbb{Z}} \frac{1}{\sqrt{t}}e^{-\pi k^2/t} = \frac{1}{\sqrt{t}}\theta\left(\frac{1}{t}\right).$$

□

Now, we can prove the functional equation:

Theorem 8.11 (Functional Equation for $\zeta(s)$)**.**

$$s(s-1)\pi^{-\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)\zeta(s) = s(s-1)\pi^{-\frac{1-s}{2}}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s).$$

Proof. Suppose that $\Re(s) > 1$. Then, we have

$$\begin{aligned}\pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s) &= \pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\sum_{n=1}^{\infty}\frac{1}{n^s} = \sum_{n=1}^{\infty}\int_0^{\infty} e^{-y}\left(\frac{y}{\pi n^2}\right)^{s/2}\frac{dy}{y} = \sum_{n=1}^{\infty}\int_0^{\infty} e^{-\pi n^2 z}z^{s/2}\frac{dz}{z} \\ &= \int_0^{\infty}\left(\sum_{n=1}^{\infty} e^{-\pi n^2}z\right)z^{s/2}\frac{dz}{z} = \int_0^{\infty}\frac{(\theta(z)-1)}{2}z^{s/2}\frac{dz}{z}.\end{aligned}$$

where the third equality comes from substituting $y = \pi n^2 z$, and the last one comes from noticing that $\omega(z) = \sum_{n=1}^{\infty} e^{-\pi n^2 z} = \frac{\theta(z)-1}{2}$. Now, we split the integral into two parts as so:

$$\int_0^\infty \frac{(\theta(z)-1)}{2} z^{s/2}\frac{dz}{z} = \int_0^1 \frac{(\theta(z)-1)}{2} z^{s/2}\frac{dz}{z} + \int_1^\infty \frac{(\theta(z)-1)}{2} z^{s/2}\frac{dz}{z}.$$

For the first integral, we use the previous proposition and rearrange:

$$\begin{aligned}\int_0^1 \frac{\theta(z)-1}{2} z^{s/2}\frac{dz}{z} &= \int_0^1 \frac{\frac{1}{\sqrt{z}}\theta(\frac{1}{z})-1}{2} z^{s/2}\frac{dz}{z} = \int_0^1 \frac{\frac{1}{\sqrt{z}}(\theta(\frac{1}{z})-1)+\frac{1}{\sqrt{z}}-1}{2} z^{s/2}\frac{dz}{z} \\ &= \int_0^1 \frac{\frac{1}{\sqrt{z}}(\theta(\frac{1}{z})-1)}{2} z^{s/2}\frac{dz}{z} + \int_0^1 \frac{z^{\frac{s-1}{2}}}{2z}dz - \int_0^1 \frac{z^{s/2}}{2z}dz \\ &= \int_0^1 \frac{\left(\theta\left(\frac{1}{z}\right)-1\right)}{2} z^{\frac{s-1}{2}}\frac{dz}{z} + \frac{1}{s-1} - \frac{1}{s} \\ &= \int_0^1 \frac{\left(\theta\left(\frac{1}{z}\right)-1\right)}{2} z^{\frac{s-1}{2}}\frac{dz}{z} + \frac{1}{s(s-1)}.\end{aligned}$$

Next, we apply the substitution $y = \frac{1}{z}$ (which gives $dy = -\frac{1}{z^2}dz$ and therefore $-\frac{1}{y^2}dy = dz$) to achieve

$$\int_0^1 \frac{\left(\theta\left(\frac{1}{z}\right)-1\right)}{2} z^{\frac{s-1}{2}}\frac{dz}{z} = \int_\infty^1 \frac{(\theta(y)-1)}{2} y^{\frac{1-s}{2}} y\left(-\frac{1}{y^2}\right)dy = \int_1^\infty \frac{(\theta(y)-1)}{2} y^{\frac{1-s}{2}}\frac{dy}{y}$$

Next, putting everything together, we find that

$$\pi^{-s/2}\Gamma(s/2)\zeta(s) = \int_0^\infty \frac{(\theta(z)-1)}{2} z^{s/2}\frac{dz}{z} = \frac{1}{s(s-1)} + \int_1^\infty \left(\frac{\theta(y)-1}{2}\right)\left(y^{\frac{s}{2}} + y^{\frac{1-s}{2}}\right)\frac{dy}{y}.$$

Multiplying by $s(s-1)$, we get the following expression, which is holomorphic for all s:

$$s(s-1)\pi^{-s/2}\Gamma(s/2)\zeta(s) = 1 + s(s-1)\int_1^\infty \left(\frac{\theta(y)-1}{2}\right)\left(y^{\frac{s}{2}} + y^{\frac{1-s}{2}}\right)\frac{dy}{y}.$$

Now, notice that the right-hand side does not change when we replace s by $1-s$. Therefore, the left-hand side does not change when replace s by $1-s$ for any $s \in \mathbb{C}$. In other words,

$$s(s-1)\pi^{-\frac{s}{2}}\Gamma\left(\frac{s}{2}\right)\zeta(s) = s(s-1)\pi^{-\frac{1-s}{2}}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s).$$

□

8.5 Computationally Verifying the Riemann Hypothesis

In this section, we discuss the question of computationally finding zeroes of the Riemann ζ-function. Now, notice that ξ and ζ have the same zeroes in the critical strip. Therefore, it suffices to discuss the question of finding zeroes of ξ in the critical strip.

Proposition 8.12. *For all s such that $\Re(s) = \frac{1}{2}$, $\xi(s) \in \mathbb{R}$.*

Proof. By the Schwarz reflection principle and looking at the integral formula, we find that $\xi(\overline{s}) = \overline{\xi(s)}$ for all s. Therefore,

$$\overline{\xi\left(\frac{1}{2}+it\right)} = \xi\left(\frac{1}{2}-it\right) = \xi\left(1-\left(\frac{1}{2}-it\right)\right) = \xi\left(\frac{1}{2}+it\right).$$

□

Therefore, a lower bound on the number of zeroes on the line $\Re(s) = \frac{1}{2}$ (which is a strict lower bound if there are zeroes of multiplicity greater than 1) is given by the number of sign changes of $\zeta(s)$ as s ranges from $\frac{1}{2}$ to $\frac{1}{2} + iT$.

On the other hand, to count zeroes of $\xi(s)$ in the rectangle 0 to 1 to $1 + iT$ to iT back to 0 (which we denote by γ), we consider the integral $\frac{1}{2\pi i} \int_\gamma \frac{\xi'(s)}{\xi(s)}$. One can find out the number of zeroes in this region (with multiplicity) by computationally estimating this integral (above and below) – since the answer is always an integer, eventually we can be certain of the exact result.

Then, if the number of sign changes is equal to the integral, we can be sure that all the zeroes in the region lie on the line; this allows us to verify the Riemann Hypothesis as far up as we want (we have already checked about ten trillion zeroes).

8.6 Theoretical Evidence for the Riemann Hypothesis

Recall that we write $M(x) = \sum_{n \le x} \mu(n)$, and the PNT is equivalent to $M(x) = o(x)$. Now, we expect that the nonzero values of μ act, in some sense, like they are randomly and uniformly distributed between 1 and -1. If this was the case, then it is easy to show using probability theory that $M(x)$ would be $O(x^{1/2} + \varepsilon)$ for any $\varepsilon > 0$. But this, as we show in the next few results, would imply the Riemann Hypothesis. Therefore, this heuristic argument provides some theoretical evidence for the Riemann Hypothesis.

Lemma 8.13. *For $\Re(s) > 1$, we have*

$$\frac{1}{\zeta(s)} = s \int_1^\infty \frac{M(x)}{x^{s+1}} dx.$$

Proof. But

$$M(n) \left(\frac{1}{(n+1)^s} - \frac{1}{n^s} \right) = -M(n) \int_n^{n+1} \frac{s}{x^{s+1}} dx = -s \int_n^{n+1} \frac{M(x)}{x^{s+1}} dx$$

where the final step holds since $M(x) = M(n)$ if $x \in [n, n+1)$. Thus,

$$\frac{M(N)}{N^s} - \sum_{n=1}^{N-1} M(n) \left(\frac{1}{(n+1)^s} - \frac{1}{n^s} \right) = \frac{M(N)}{N^s} + s \sum_{n=1}^{N-1} \int_n^{n+1} \frac{M(x)}{x^{s+1}} dx = \frac{M(N)}{N^s} + s \int_1^N \frac{M(x)}{x^{s+1}} dx.$$

Therefore, in summary, $\sum_{n=1}^N \frac{\mu(n)}{n^s} = \frac{M(N)}{N^s} + s \int_1^N \frac{M(x)}{x^{s+1}} dx$. But notice that $M(x)$ is certainly at worst $O(x)$, so whenever $\Re(s) > 1$, $\frac{M(N)}{N^s} \to 0$ as $N \to \infty$. Similarly, this implies that $\left| \int_1^N \frac{M(x)}{x^{s+1}} dx \right| \le \int_1^N \left| \frac{M(x)}{x^{s+1}} \right| dx \le \int_1^N \left| \frac{x}{x^{s+1}} \right| dx = \int_1^N \left| \frac{1}{x^s} \right| dx = \int_1^N \frac{1}{x^\sigma} dx$, which converges as $N \to \infty$ whenever $\Re(s) = \sigma > 1$.

Therefore, we may conclude that if $\Re(s) > 1$,

$$\frac{1}{\zeta(s)} = \sum_{n=1}^\infty \frac{\mu(n)}{n^s} = \lim_{N \to \infty} \sum_{n=1}^N \frac{\mu(n)}{n^s} = \lim_{N \to \infty} \left(\frac{M(N)}{N^s} + s \int_1^N \frac{M(x)}{x^{s+1}} dx \right) = s \int_1^\infty \frac{M(x)}{x^{s+1}} dx$$

which converges by our above work. □

Lemma 8.14. *Suppose there is a positive constant θ such that $M(x) = O(x^\theta)$ for $x \ge 1$. Then*

$$\frac{1}{\zeta(s)} = s \int_1^\infty \frac{M(x)}{x^{s+1}} dx.$$

and $\zeta(s) \ne 0$ for $\sigma > \theta$.

Proof. Suppose that there is a positive constant θ such that $M(x) = O(x^\theta)$ for $x \geq 1$. Then, $\frac{M(N)}{N^s} \to 0$ as $N \to \infty$ whenever $\sigma > \theta$. Furthermore, whenever $\sigma > \theta$,

$$\left| s \int_1^\infty \frac{M(x)}{x^{s+1}} dx \right| \leq |s| \int_1^\infty \left| \frac{M(x)}{x^{s+1}} \right| dx \leq C|s| \int_1^\infty \left| \frac{x^\theta}{x^{s+1}} \right| dx = C|s| \int_1^\infty \frac{1}{x^{\sigma-\theta+1}} dx.$$

and $\sigma - \theta + 1 > 1$ whence the integral on the right converges to a finite value. Thus, we can again conclude that whenever $\sigma > \theta$,

$$\sum_{n=1}^\infty \frac{\mu(n)}{n^s} = \lim_{N\to\infty} \sum_{n=1}^N \frac{\mu(n)}{n^s} = \lim_{N\to\infty} \left(\frac{M(N)}{N^s} + s \int_1^N \frac{M(x)}{x^{s+1}} dx \right) = s \int_1^\infty \frac{M(x)}{x^{s+1}} dx.$$

Now, by uniqueness of analytic continuations, the left-hand side is equal to $\frac{1}{\zeta(s)}$ wherever $\frac{1}{\zeta(s)}$ is defined.

Now, this a priori does not show that $\zeta(s) \neq 0$ (since when $\zeta(s) = 0$, $\frac{1}{\zeta(s)}$ is not defined). Instead, the idea is that when $\zeta(s)$ has a zero, $\frac{1}{\zeta(s)}$ is unbounded in any neighborhood of that zero, but if $\Re(s) > \theta$, $s \int_1^\infty \frac{M(x)}{x^{s+1}} dx$ can be bounded above (by taking absolute values and evaluating the integral) for all sufficiently small neighborhoods of s (precisely, neighborhoods which are bounded away from the line $\Re(s) = \theta$). Therefore, $\zeta(s)$ cannot have a zero such that $\Re(s) > \theta$. □

Corollary 8.14.1. *$M(x) = O(x^{1/2+\varepsilon})$ for all $\varepsilon > 0$ would implies the Riemann Hypothesis. The converse is also true but we do not prove it: see Chapter 5.9 of Conrad.*

9 Dirichlet's Theorem on Primes in Arithmetic Progressions

In this section, we prove Dirichlet's Theorem on Primes in Arithmetic Progressions, which states that whenever a and b are coprime, the arithmetic progression $\{a + nb \mid n \in \mathbb{N}\}$ contains infinitely many primes. To do this, we develop the language of Dirichlet characters and L-functions (which are, in some sense, a generalization of the ζ-function). Then, with some complex analysis, we can prove the result.

9.1 Dirichlet Characters

Definition 9.1 (Dirichlet Character)**.** A *Dirichlet character mod n* can be defined in either of the following equivalent ways:

(1) A *Dirichlet character mod n* is a homomorphism $\chi : (\mathbb{Z}/n\mathbb{Z})^\times \to \mathbb{C}^\times$.

(2) A *Dirichlet character mod n* is a completely multiplicative function $\chi : \mathbb{Z} \to \mathbb{C}^\times$ with period n such that $\chi(m) = 0$ if and only if $(m, n) > 1$.

Example 9.2. The function χ_{-4} given by

$$\chi_{-4}(a) = \begin{cases} 1 & a \equiv 1 \bmod 4 \\ -1 & a \equiv 3 \bmod 4 \\ 0 & \text{otherwise} \end{cases}$$

is a Dirichlet character mod 4.

Example 9.3. Suppose that p is a prime. Then it is a fact from elementary number theory that $(\mathbb{Z}/p\mathbb{Z})^\times$ has a generator a_0 (that is, $(\mathbb{Z}/p\mathbb{Z})^\times$ is cyclic). Then, clearly any homomorphism $\chi : (\mathbb{Z}/p\mathbb{Z})^\times \to \mathbb{C}^\times$ is completely determined by $\chi(a_0)$, and $\chi(a_0)^{p-1} = \chi(a_0^{p-1}) = \chi(1) = 1$. Conversely, as long as ζ satisfies $\zeta^{p-1} = 1$, then we can define a homomorphism $\chi : (\mathbb{Z}/p\mathbb{Z})^\times \to \mathbb{C}^\times$ by $\chi(a_0^k) = \zeta^k$. Thus the group of Dirichlet characters mod p is isomorphic to the group of $(p-1)$th roots of unity; this is not a natural isomorphism as choosing this isomorphism requires choosing one of the $\varphi(\varphi(p))$ generators of $(\mathbb{Z}/p\mathbb{Z})^\times$.

More generally, $a^{\varphi(n)} \equiv 1 \bmod n$ (Fermat's Little Theorem) implies that if χ is a Dirichlet character mod n, $\chi(a)$ is a $\varphi(n)$th root of unity.

Example 9.4. The Dirichlet character $(\mathbb{Z}/n\mathbb{Z})^\times \to \mathbb{C}^\times$ given by $a \mapsto 1$ is called the *trivial character* and denoted χ_0; it is a Dirichlet character mod n for every $n \in \mathbb{N}$.

Example 9.5. Given a Dirichlet character χ mod n, the function $\overline{\chi}$ given by $\overline{\chi}(a) = \overline{\chi(a)}$ is a Dirichlet character mod n. Then, if we define $\chi_1\chi_2$ by $(\chi_1\chi_2)(a) = \chi_1(a)\chi_2(a)$, notice that $\chi\overline{\chi} = \chi_0$; thus, $\overline{\chi}$ is the inverse of χ in the group $\mathrm{Hom}((\mathbb{Z}/n\mathbb{Z})^\times, \mathbb{C}^\times)$ of Dirichlet characters.

Proposition 9.6. *The number of Dirichlet characters mod n is $\varphi(n)$.*

Proof. First, recall that φ is multiplicative. Also notice that the number of Dirichlet characters mod n is multiplicative. This follows from (1) the Chinese Remainder Theorem, which states that $(\mathbb{Z}/mn\mathbb{Z}) \simeq (\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/n\mathbb{Z})$ as rings whenever m and n are coprime (so in particular, by taking unit groups, $(\mathbb{Z}/mn\mathbb{Z})^\times \simeq (\mathbb{Z}/m\mathbb{Z})^\times \times (\mathbb{Z}/n\mathbb{Z})^\times$), and (2) the fact that group homomorphisms $G \times H \to K$ are in bijection with pairs of group homomorphisms $G \to K$ and $H \to K$.

Therefore, it suffices to show the result for prime powers. Now, for all odd primes, $(\mathbb{Z}/p^e\mathbb{Z})^\times$ is cyclic, so by the same logic from the earlier example on Dirichlet characters mod p, there are $\varphi(p^e)$ Dirichlet characters mod p^e. Hence it suffices to show that there are $\varphi(2^e)$ Dirichlet characters mod 2^e. Now, the result is obvious for $e = 1$ and $e = 2$. For $e > 2$, recall that $(\mathbb{Z}/2^e\mathbb{Z})^\times \simeq C_2 \times C_{2^{e-2}}$ (another fact from elementary number theory). Then, it is straightforward to count that number of homomorphisms from this latter group into $\mathbb{C}^\times$ is $2 \cdot 2^{e-2} = 2^{e-1} = \varphi(2^e)$. Therefore the result follows. □

Proposition 9.7. *Suppose that $a \neq b \in (\mathbb{Z}/n\mathbb{Z})^\times$. Then there exists a character χ such that $\chi(a) \neq \chi(b)$.*

Proof. It suffices to show that, for any nonidentity element a, there exists χ such that $\chi(a) \neq 1$. To see why, notice that if $a \neq b$, ab^{-1} is a nonidentity element, so by hypothesis there exists χ such that $\chi(ab^{-1}) = \chi(a)\chi(b)^{-1} \neq 1$ which implies $\chi(a) \neq \chi(b)$). But this latter task we can do explicitly; we constructing a Dirichlet character which works if n is a prime power (treating the case where $n = 2^e$ and $n = p^e$ for an odd prime p separately), and then use the fact that Dirichlet characters mod n can be built out of Dirichlet characters mod prime powers using the Chinese Remainder Theorem. The proof is laborious but relatively obvious, so we skip the details. □

Proposition 9.8 (Row Sum)**.** *Suppose that χ is a Dirichlet character mod m. Then,*

$$\sum_{a \in (\mathbb{Z}/m\mathbb{Z})^\times} \chi(a) = \begin{cases} \phi(m) & \chi = \chi_0 \\ 0 & \textit{otherwise.} \end{cases}$$

Proof. The result is obviously true when $\chi = \chi_0$ is trivial. Therefore, we may assume that χ is nontrivial. In particular, there exists b such that $\chi(b) \neq 1$. But then

$$\chi(b)\left(\sum_a \chi(a)\right) = \sum_a \chi(ab) = \sum_a \chi(a) \Rightarrow \sum_a \chi(a) = 0$$

which is the desired result. □

Proposition 9.9 (Column Sum)**.** *Suppose that $a \in (\mathbb{Z}/m\mathbb{Z})^\times$. Then,*

$$\sum_\chi \chi(a) = \begin{cases} \phi(m) & a = 1 \\ 0 & \textit{otherwise.} \end{cases}$$

where the sum is taken over all Dirichlet characters mod m.

Proof. The result is obviously true when $a = 1$. Therefore, assume that $a \neq 1$. Then, by Proposition 9.7, there exists a character ψ such that $\psi(a) \neq 1$. But then,

$$\psi(a)\left(\sum_\chi \chi(a)\right) = \sum_\chi (\psi\chi)(a) = \sum_\chi \chi(a) \Rightarrow \sum_\chi \chi(a) = 0$$

which is the desired result. □

Corollary 9.9.1 (Picking Out Elements)**.** *Suppose that χ is a Dirichlet character mod m. Then,*

$$\frac{1}{\varphi(m)} \sum_\chi \chi(n)\overline{\chi(a)} = \begin{cases} 1 & n \equiv a \\ 0 & \textit{otherwise.} \end{cases}$$

Proof.

$$\frac{1}{\varphi(m)} \sum_\chi \chi(n)\overline{\chi(a)} = \frac{1}{\varphi(m)} \sum_\chi \chi(n)\chi(a^{-1}) = \frac{1}{\varphi(m)} \sum_\chi \chi(na^{-1}) = \begin{cases} 1 & na^{-1} \equiv 1 \\ 0 & \text{otherwise.} \end{cases} = \begin{cases} 1 & n \equiv a \\ 0 & \text{otherwise.} \end{cases}$$

□

9.2 Dirichlet L-Functions

Definition 9.10 (Dirichlet L-function)**.** Suppose that $\chi : \mathbb{Z} \to \mathbb{C}$ is a Dirichlet character mod m. Then the corresponding *Dirichlet L-function* $L(s, \chi)$ is defined by

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

There are a few helpful facts about Dirichlet L-functions which are analogous to the Riemann ζ-function. Firstly, since $|\chi(n)| = 0$ or 1 for all n, by taking absolute values we notice that $L(s, \chi)$ absolutely converges for any χ when $\Re(s) > 1$. Secondly, since χ is bounded and completely multiplicative, we have an extremely simple Euler product:

$$L(s,\chi) = \prod_p \left(1 + \frac{\chi(p)}{p^s} + \frac{\chi(p^s)}{p^{2s}} + \cdots\right) = \prod_p \sum_{k=0}^{\infty} \left(\frac{\chi(p)}{p^s}\right)^k = \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}.$$

However, not all properties of Dirichlet L-functions are analogous to the Riemann ζ-function. Indeed, when χ is not the trivial character mod m, $L(s, \chi)$ converges conditionally for $\Re(s) > 0$.

Proposition 9.11. *Suppose that χ is a nontrivial Dirichlet character mod m. Then $L(s, \chi)$ converges conditionally for $\Re(s) > 0$. In particular, $L(s, \chi)$ has no pole at $s = 1$ whenever χ is nontrivial.*

Proof. Let $A(y) = \sum_{n \le y} \chi(n)$. Then,

$$\sum_{n=1}^{N} \frac{\chi(n)}{n^s} = \int_{1^-}^{N^+} \frac{1}{y^s} dA(y) = \frac{A(y)}{y^s}\bigg|_{1^-}^{N^+} - \int_1^N A(y) \left(\frac{-s}{y^{s+1}}\right) dy$$

Now, if χ is not the trivial character, it follows that $A(y) = O(1)$. Therefore, the first term on the right-hand side goes to 0 as $N \to \infty$ for any $s > 0$. Thus, in this case we have

$$\sum_{n=1}^{N} \frac{\chi(n)}{n^s} = s \int_1^N \frac{A(y)}{y^{s+1}} dy$$

and since $A(y) = O(1)$, this integral converges absolutely as $N \to \infty$ whenever $\Re(s) > 0$. □

However, when χ is the trivial character, then $L(s, \chi)$ behaves very similarly to the Riemann ζ-function. Indeed, if 1_m denotes the trivial character mod m, $L(s, 1_m) = \sum_{\gcd(n,m)=1} \frac{1}{n^s} = \prod_{p \nmid m} \left(1 - \frac{1}{p^s}\right)$. That is, $L(s, 1_m) = \zeta(s) \prod_{p | m} \left(1 - \frac{1}{p^s}\right)$. Therefore, $L(s, 1_m)$ converges in exactly the same places and ways that $\zeta(s)$ does, since it is equal to $\zeta(s)$ times a positive constant.

Next, we will select a preferred branch of the logarithm as we did for the Riemann ζ-function. Indeed, we choose the branch of the logarithm which makes the following reasoning true, just as for the ζ-function:

$$\log L(s, \chi) = \log \prod_p \left(1 - \frac{\chi(p)}{p^s}\right)^{-1} = -\sum_p \log\left(1 - \frac{\chi(p)}{p^s}\right) = \sum_{p,k} \frac{\chi(p)^k}{kp^{ks}}$$

This Dirichlet series is absolutely convergent for $\Re(s) > 1$.

Finally, there is no particular place to put this, so we quickly include it here:

Proposition 9.12.

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

Proof. Trivial, left to the reader. □

Lemma 9.13. *Suppose that $\sum_{n \ge 0} \frac{b_n}{n^s}$ is a Dirichlet series which is absolutely convergent when $\Re(s) > \sigma_0$. Then $\exp\left(\sum_{n \ge 0} \frac{b_n}{n^s}\right)$ is a Dirichlet series $\sum_{n \ge 0} \frac{a_n}{n^s}$ is a Dirichlet series which is absolutely convergent when $\Re(s) > \sigma_0$. Furthermore, if $b_n \ge 0$ for all n, $0 \le b_n \le a_n$ for all n.*

Proof. First, suppose that $\sum_{n\geq 1} b_n/n^s$ converges absolutely when $\Re(s) > \sigma_0$. Then, let b denote the function $b(n) = b_n$. Then, let $b^{\star k} = b \star \cdots \star b$ (where, on the right-hand side, we are taking the Dirichlet convolution of k copies of b). Now, $\exp(\sum_{n\geq 1} b_n/n^s)$ converges absolutely when $\Re(s) > \sigma_0$. But since $\sum_{k\geq 0} \frac{x^k}{k!}$ converges uniformly to $\exp(x)$ on compact subsets, we can write

$$\exp\left(\sum_{n\geq 1} \frac{b_n}{n^s}\right) = \sum_{k\geq 0} \frac{1}{k!}\left(\sum_{n\geq 1} \frac{b_n}{n^s}\right)^k = \sum_{k\geq 0} \frac{1}{k!} \sum_{n\geq 1} \frac{b^{\star k}(n)}{n^s}.$$

and the sum on the right converges absolutely whenever $\Re(s) > \sigma_0$. Yet, by absolute convergence, we can swap the sums and write

$$\sum_{k\geq 0} \frac{1}{k!} \sum_{n\geq 1} \frac{b^{\star k}(n)}{n^s} = \sum_{n\geq 1} \frac{1}{n^s} \sum_{k\geq 0} \frac{b^{\star k}(n)}{k!}$$

as long as $\sum_{k\geq 0} \frac{b^{\star k}(n)}{k!}$ converges absolutely. To see this, notice that if $\max_{x\in\{1,\ldots,n\}} |b(x)| = C$, then $|b^{\star k}(n)| \leq C^k n^k$. But $\sum_{k\geq 0} \frac{C^k n^k}{k!}$ converges by the Ratio Test. Thus the coefficient $\sum_{k\geq 0} \frac{b^{\star k}(n)}{k!}$ converges absolutely, as desired. Therefore, $\exp\left(\sum_{n\geq 1} \frac{b_n}{n^s}\right)$ is equal to the Dirichlet series $\sum_{n\geq 1} \frac{1}{n^s} \sum_{k\geq 0} \frac{b^{\star k}(n)}{k!}$, which is absolutely convergent whenever $\Re(s) > \sigma_0$.

Now, we have $a_n = \sum_{k\geq 0} \frac{b^{\star k}(n)}{k!}$. But it is obvious that if $b(n) \geq 0$ for all n, then $b^{\star k}(n) \geq 0$ for all n and all k, and so in particular a_n is the infinite sum of non-negative terms and is non-negative. Furthermore, since all the terms of $\sum_{k\geq 0} \frac{b^{\star k}(n)}{k!}$ are non-negative, the sum is bounded below by the $k = 1$ term, which is precisely $b(n) = b_n$. Thus $a_n \geq b_n$, as desired. Thus we are done. □

9.3 Dirichlet's Theorem

Lemma 9.14. *Suppose that χ is a nontrivial Dirichlet character mod m. Then $L(1, \chi) \neq 0$.*

Proof. The key is Landau's Theorem (Corollary 3.4.2 in Conrad). Landau's Theorem states that if a Dirichlet series $F(s) = \sum_{n=1}^{\infty} \frac{f(n)}{n^s}$ converges for $\Re(s) > \sigma_0$, has non-negative coefficients (i.e. $f(n) \geq 0$ for all n), and $F(s)$ has an analytic continuation to a larger half-plane $\Re(s) > \sigma_1$ (i.e., $\sigma_1 < \sigma_0$), then the continuation of $F(s)$ equals $\sum_{n=1}^{\infty} \frac{f(n)}{n^s}$ on $\Re(s) > \sigma_1$. In particular, the Dirichlet series $F(s)$ converges on this domain.

Let $F(s) = \zeta(s)^2 L(s, \chi) L(s, \overline{\chi})$. Then if $L(1, \chi) = 0$, the relation $\overline{L(s, \chi)} = L(\overline{s}, \overline{\chi})$ implies $L(1, \overline{\chi}) = 0$. Thus the double pole of $\zeta(s)^2$ cancels out with the double zero of $L(s, \chi)L(s, \overline{\chi})$ at $s = 1$, so $F(s)$ has no pole at 1. Now, by taking the Euler product for each term of $F(s)$, we get the following Euler product for $F(s)$:

$$F(s) = \prod_p \left(1 - \frac{1}{p^s}\right)^{-2} \left(1 - \frac{\chi(p)}{p^s}\right)^{-1} \left(1 - \frac{\overline{\chi}(p)}{p^s}\right)^{-1}$$

Taking the logarithm, we have

$$-2\log\left(1 - \frac{1}{p^s}\right) - \log\left(1 - \frac{\chi(p)}{p^s}\right) - \log\left(1 - \frac{\overline{\chi(p)}}{p^s}\right) = \sum_{k\geq 1} \frac{2 + \chi(p)^k + \overline{\chi}(p)^k}{kp^{ks}} = \sum_{k\geq 1} \frac{2 + 2\Re(\chi(p)^k}{kp^{ks}}$$

and of course this implies that the coefficients of the sum are non-negative. But then, if we take the exponential of this to achieve a Dirichlet series for F, the Lemma above implies that this Dirichlet series has all nonnegative coefficients. Then, since this Dirichlet series converges for all $\Re(s) > 1$ and $F(s)$ has an analytic continuation to $\Re(s) > 0$, this Dirichlet series converges for $\Re(s) > 0$.

Now, for simplicity, write $z = \frac{1}{p^s}$ and $c = \chi(p)$. Then, looking at the Euler factor for p,

$$\begin{aligned}\left(1-\frac{1}{p^s}\right)^{-2}\left(1-\frac{\chi(p)}{p^s}\right)^{-1}\left(1-\frac{\overline{\chi}(p)}{p^s}\right)^{-1} &= \frac{1}{(1-z)^2}\frac{1}{1-cz}\frac{1}{1-\overline{c}z} \\ &= (1+2z+3z^2+\cdots)(1+cz+c^2z^2+\cdots)(1+\overline{c}z+\overline{c}^2z^2+\cdots) \\ &= 1+(2+c+\overline{c})z+(c^2+c\overline{c}+2c+2\overline{c}+3)z^2+\cdots.\end{aligned}$$

Then the coefficient of z^k is the coefficient of $\frac{1}{p^{ks}}$ in $F(s)$. In particular, the coefficient of $\frac{1}{p^{2s}}$ is $(c+\overline{c})^2+2(c+\overline{c})+3-c\overline{c} = (c+\overline{c}+1)^2+2-|c|^2 \geq 2-|c|^2$, which is at least 1. Thus for all $s \in \mathbb{R}_{>0}$, $F(s) \geq \sum_p \frac{1}{p^{2s}}$ (by forgetting every term except the term for p^{2s} and using the bound for the coefficient above). But then, as $s \to \frac{1}{2}$, $\sum_p \frac{1}{p^{2s}}$ goes to infinity. Thus $F(s)$ has a pole at $\frac{1}{2}$, a contradiction with the fact that it is analytic on $\Re(s) > 0$, a contradiction. □

Theorem 9.15 (Dirichlet's Theorem)**.** *Suppose that* $\gcd(a,m)=1$. *Then there are infinitely many primes* p *in the arithmetic progression* $a, a+m, a+2m, \ldots$.

Proof. First, notice that the sum $\sum_{p\equiv a \bmod m}\frac{1}{p^s}$ converges absolutely for $\Re(s) > 1$. Furthermore,

$$\sum_{p\equiv a \bmod m}\frac{1}{p^s} = \frac{1}{\varphi(m)}\sum_p\sum_\chi\frac{\chi(p)\overline{\chi(a)}}{p^s} = \frac{1}{\varphi(m)}\sum_\chi\overline{\chi(a)}\sum_p\left(\frac{\chi(p)}{p^s}\right)$$

Now, suppose that $\chi \neq 1_m$; then, the previous lemma states that $L(1,\chi)$ is a nonzero finite value. Therefore, we can define $\log L(s,\chi)$ in an open ball U around 1. We want to show that this is the same as the branch of the logarithm chosen earlier (so that we can conclude that $\log L(s,\chi) = \sum_{p,k}\frac{\chi(p)^k}{kp^{ks}}$ in a neighborhood of 1). To see why this holds, notice that both the old logarithm and the new logarithm are defined on the intersection of U and the half-plane $\Re(s) > 1$. But the difference between them must necessarily be $2\pi ki$ for some $k \in \mathbb{Z}$. To see why, notice that the exponential of their difference is 1, so their difference lies in $2\pi\mathbb{Z}i$; this is totally disconnected, so because the neighborhood $U \cap \{s \mid \Re(s) > 1\}$ we are considering is connected, their difference must be constant.

But then the new logarithm minus $2\pi ki$ serves to define the unique analytic extension of the old logarithm to U. In summary, we may conclude that $\log L(s,\chi) = \sum_{p,k}\frac{\chi(p)^k}{kp^{ks}} = \sum_p\left(\frac{\chi(p)}{p^s}\right) + \sum_{p,k\geq 2}\frac{\chi(p)^k}{kp^{ks}}$ in an open ball around 1 for any $\chi \neq 1_m$. Yet $\sum_{p,k\geq 2}\frac{\chi(p)^k}{kp^{ks}}$ converges absolutely when $\Re(s) > \frac{1}{2}$, so in particular we can bound it above by a constant c_χ whenever $s \geq \frac{3}{4}$. Therefore, whenever $\Re(s) > 1$,

$$\sum_{p\equiv a \bmod m}\frac{1}{p^s} \geq \frac{1}{\varphi(m)}\sum_\chi\overline{\chi(a)}\left(\log L(s,\chi) - c_\chi\right).$$

whenever $s \geq \frac{3}{4}$. But then since $L(1,\chi)$ is finite and nonzero for all nontrivial characters χ by the above lemma, and since $L(1,1_m) \to \infty$ as $s \to 1^+$, the limit of the sum on the right-hand side is ∞. Thus $\lim_{s\to 1^+}\sum_{p\equiv a \bmod m}\frac{1}{p^s} = \infty$. But if this sum were finite, then the limit would be equal to the finite sum $\sum_{p\equiv a \bmod m}\frac{1}{p}$. Thus the sum over $p \equiv a \bmod m$ must have infinitely many terms, as desired. □

9.4 Natural Density

Corollary 9.15.1. *Suppose that* $d(a,m)=1$. *Then* $\lim_{s\to 1^+}$ *of* $\frac{\sum_{p\equiv a(m)}\frac{1}{p^s}}{\sum_p\frac{1}{p^s}} = \frac{1}{\varphi(m)}$.

Proof. This is very simple: $\sum_{p\equiv a(m)} = \frac{1}{\varphi(m)}\sum_\chi\overline{\chi}(a)\left(\sum_p\frac{\chi(p)}{p^s}\right) = \frac{1}{\varphi(m)}\sum_p\frac{1}{p^s}+O(1)$. □

We hope that the primes are about evenly distributed amongst the reduced residue classes. That is, more formally, we want to show that "$\{p \equiv a(m)\}$ has density $\frac{1}{\varphi(m)}$ in the set of all primes"; more precisely,

$$\lim_{x\to\infty}\frac{\pi(x;a,m)}{\pi(x)} = \frac{1}{\varphi(m)}$$

where $\pi(x\ a, m) = |\{p \leq x \mid p \equiv a(m)\}|$. Let us elaborate and formalize this notion now, and then we will walk through a rough skeleton of the proof.

Definition 9.16. Suppose that $B \subseteq A \subseteq \mathbb{N}$. Then the *natural density of B in A* is defined to be

$$\lim_{x\to\infty} \frac{|\{b \in B \mid b \leq x\}|}{|\{a \in A \mid a \leq x\}|}$$

if this limit exists, and the natural density is said to *not exist* if it does not. By abuse of terminology, we call the natural density of B in $\mathbb{N}$ simply the *natural density of B*.

Example 9.17. The set of primes has natural density 0, since $\frac{x/\log x}{x} = \frac{1}{\log x} \to 0$ as $x \to \infty$.

Besides natural density, there are two other important notions of density:

Definition 9.18 (Logarithmic Density)**.** Define logarithmic density as $\lim_{x\to\infty} \frac{\sum_{p\in\wp, p\leq x} \frac{1}{p}}{\sum_{p\leq x} \frac{1}{p}}$. This matches Dirichlet density exactly.

Definition 9.19 (Dirichlet Density)**.** Suppose that $\wp$ is a subset consisting of prime numbers. Then, the *Dirichlet density of $\wp$ in the primes* is defined to be

$$\lim_{s\to 1^+} \frac{\sum_{p\in\wp} \frac{1}{p^s}}{\sum_p \frac{1}{p^s}}.$$

Proposition 9.20. *Suppose that $\wp$ is a subset of the primes with natural density δ. Then $\wp$ has Dirichlet density δ in the primes.*

Proof. See Conrad's notes on *Analytic Number Theory.* □

Proposition 9.21. *Suppose that $\wp$ is a subset of the primes. Then $\wp$ has Dirichlet density δ in the primes if and only if $\wp$ has logarithmic density δ in the primes.*

Proof. See Conrad's notes on *Analytic Number Theory.* □

Now, in this language, it is indeed a theorem that if $(a, m) \equiv 1$, then the collection of primes congruent to a mod m have natural density $\frac{1}{\varphi(m)}$ in the set of all primes. This is equivalent to stating that $\pi(x\ a, m) \sim \frac{1}{\varphi(m)} \frac{x}{\log x}$, and it follows from the three steps described below:

1. Theorem (Conrad 4.6.1): Suppose that $A \subseteq \mathbb{N}$ and $\wp(A)$ is the subset of primes in A. Suppose

$$F(s) = \sum_{n\in A} \frac{\Lambda(n)}{n^s} = \sum_{p^k\in A} \frac{\log p}{p^{ks}}$$

   has a continuation to $\Re(s) \geq 1$ except for a possible pole at $s = 1$ with residue r. Then the natural density $\wp(A)$ in the set of primes is r. The proof of this fact is essentially an analogue of the prime number theorem (including the error bound, which is $O(x^{1/2+\varepsilon})$ conditional on the Generalized Riemann Hypothesis).

2. Theorem (Conrad 4.22): Suppose that χ is a nontrivial Dirichlet character mod m. Then $L((1 + it), \chi) \neq 0$ for all t. Proving this amounts to adapting either the proof using Landau's Theorem in Conrad's notes, or the proof in these notes (where we replace $\zeta(\sigma)^3|\zeta(\sigma + it)|^4|\zeta(\sigma + 2it)| \geq 1$ with $\zeta(\sigma)^3|L(\sigma + it, \chi)|^4|L(\sigma + 2it, \chi^2)|$).

3. Finally, we can show the result by applying the first theorem with $A = \{n \in \mathbb{N} \mid n \equiv a \bmod m\}$, and that the residue of F at $s = 1$ is $\frac{1}{\varphi(m)}$.

10 GRH and Primality Testing

In this section, we'll discuss the Generalized Riemann Hypothesis, and an important consequence for theoretical computer science.

10.1 The Generalized Riemann Hypothesis

Like ζ, $L(s,\chi)$ has a functional equation relating it to $L(1-s,\chi)$. This function can be used to show that any zeroes in the half-plane $\Re(s) < 0$ must lie in $\mathbb{R}_{<0}$. These are called the "trivial zeroes" of a Dirichlet L-function, and the Generalized Riemann Hypothesis states that any other must lie on the critical line.

Conjecture 10.1 (Generalized Riemann Hypothesis)**.** *Suppose that χ is a Dirichlet character mod n, and $L(s,\chi)$ is the corresponding Dirichlet L-function. Then $L(s,\chi)$ has no "nontrivial zeroes" outside the line $\Re(s) = \frac{1}{2}$, though it may have "trivial zeroes" on $\mathbb{R}_{<0}$.*

10.2 Consequences of GRH for $(\mathbb{Z}/n\mathbb{Z})^\times$

In this section, we are going to briefly discuss a sketch of the following theorem:

Theorem 10.1. *Assume that the Generalized Riemann Hypothesis is true. Then, if H is a proper subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$, there exists a positive integer x less than $2(\log n)^2$ such that $x \notin H$.*

To prove this, we first need a technical definition:

Definition 10.2 (Primitive Character)**.** Suppose that $\chi : (\mathbb{Z}/m\mathbb{Z})^\times \to \mathbb{C}^\times$ is a Dirichlet character mod m. A character is *primitive* if χ does not factor as a map $(\mathbb{Z}/m\mathbb{Z})^\times \twoheadrightarrow (\mathbb{Z}/d\mathbb{Z})^\times \to C^\times$ for any proper divisor d of m (where the left-hand map is simply the reduction map $a \bmod m \mapsto a \bmod d$).

Furthermore, notice that for any character χ mod m, there exists $d \mid m$ such that χ comes from a primitive character $\chi_* : (\mathbb{Z}/d\mathbb{Z})^\times \to \mathbb{C}^\times$. As functions $\chi, \chi_* : \mathbb{Z} \to \mathbb{C}$, they agree on all a such that $(a,m) = 1$ or $(a,d) > 1$; however, when $(a,d) = 1$ but $(a,m) > 1$, then $\chi(a) = 0$ but $\chi_*(a) \neq 0$. In other words, χ and χ_* are *almost* the same as functions on $\mathbb{Z}$, except that χ might be 0 in more places than χ_* is.

The next step to proving the above theorem is reframing it using the following result:

Proposition 10.3. *The following are equivalent:*

(1) For all $m > 1$, each proper subgroup of $(\mathbb{Z}/m\mathbb{Z})^\times$ omits $n = O((\log m)^2)$.

(2) For all $m > 1$ and each primitive character χ mod m, there exists $n = O((\log m)^2)$ such that $\chi(n) \neq 1$.

Proof.
(1) ⇒ (2): Let χ be a primitive character mod m. Then, consider the kernel $\ker(\chi)$ of χ: this is proper since χ is primitive and therefore nontrivial, so $\ker(\chi)$ omits $n = O((\log m)^2)$. Yet $n \in \ker(\chi)$ means precisely that $\chi(1) = 1$, so the result follows.

(2) ⇒ (1): Suppose that $H \subsetneq (\mathbb{Z}/m\mathbb{Z})^\times$. Then, there exists a nontrivial character χ such that $\chi(H) = 1$. Then, replace χ with the primitive character χ_*. By assumption, there exists $n = O((\log d)^2) = O((\log m)^2)$ such that $\chi_*(n) \neq 1$. Then, $\chi(n)$ is either equal to 0 or $\chi_*(n)$, so in particular $\chi(n) \neq 1$. But then $n \notin H$. □

Finally, we will sketch the proof of the second condition, thereby proving the main theorem of this section. The intuition here is that if χ mod m is primitive and $\chi(n) = 1$ for all $n \leq x$, then $L(s,\chi) = \sum_{n=1}^\infty \frac{\chi(n)}{n^s}$ and $\zeta(s) = \sum_{n=1}^\infty \frac{1}{n^s}$ agree on the first x terms. So if x is large, then these two Dirichlet series should have similar size – but at $s = 1$, the two series have very different behavior, which yields a contradiction when x is large. The below theorem makes this intuitive idea rigorous.

Theorem 10.4. *For all $m > 1$ and each primitive character χ mod m, there exists $n = O((\log m)^2)$ such that $\chi(n) \neq 1$.*

Proof. Recall that if $\psi(x) = \sum_{n \le x} \Lambda(n)$ is the Chebyshev function, then $\psi(x) = \frac{1}{2\pi i}\int_{(c)} \left(-\frac{\zeta'}{\zeta}\right)(s)\frac{x^s}{s}ds$. Generalizing this notion, we define $\psi_\chi(x) = \sum_{n \le x} \chi(n)\Lambda(n)$. Then, $\psi_\chi(x) = \frac{1}{2\pi i}\int_{(c)} \left(-\frac{L'(s,\chi)}{L(s,\chi)}\right)\frac{x^s}{s}ds$.

These two are equal if $\chi(n) = 1$ for all $n \le x$. Then, for a sufficiently large rectangle, these integrals can be approximated using the residues: that is, $\psi(x) \sim \left(x - \sum_\rho \frac{x^\rho}{\rho}\right)$ where the sum is taken over the nontrivial zeroes of ζ. Similarly, $\psi_\chi(x) \sim -\sum_{\rho_\chi} \frac{x^{\rho_\chi}}{\rho_\chi}$ where the sum of taken over the nontrivial zeroes of $L(s,\chi)$. Now, if there are no zeroes except on the critical strip, we might expect the first sum to be something asymptotic to x, whereas the second sum would be asymptotic to $\sqrt{x}$, which would be impossible. However, this is not quite precise, because the sums of the reciprocals of the zeroes diverge (Conrad). Therefore, we need something even more precise.

This additional precision comes from the following formulas: for $b \in (0,1)$,

$$\frac{1}{2\pi i}\int_{(c)} \left(-\frac{\zeta'}{\zeta}(s)\right)\frac{x^s}{(s+b)^2}ds = \sum_{n \le x} \Lambda(n)\left(\frac{n}{x}\right)^b \log\left(\frac{x}{n}\right).$$

$$\frac{1}{2\pi i}\int_{(c)} \left(-\frac{L'}{L}\right)(s,\chi)\frac{x^s}{(s+b)^2}ds = \sum_{n \le x} \chi(n)\Lambda(n)\left(\frac{n}{x}\right)^b \log\left(\frac{x}{n}\right)$$

Then, as usual, we move integral to the left and all the leftover integrals turn out to be small, so the residues are the only parts that matter. Now, $\frac{x}{(1+b)^2}$ is the residue at $s = 1$, and then we have $\frac{x}{(1+b)^2} - \sum_\rho \frac{x^\rho}{(\rho+b)^2} + O(\log(m)) = -\sum_{\rho_\chi} \frac{x^{\rho_\chi}}{(\rho_\chi+b)^2} + O(\log m)$. Thus, we have

$$\frac{x}{(1+b)^2} = \sum_\rho \frac{x^\rho}{(\rho+b)^2} - \sum_{\rho_\chi} \frac{x^{\rho_\chi}}{(\rho_\chi+b)^2} + O(\log m)$$

Taking absolute values and setting $b = \frac{1}{2}$, we find that

$$\frac{4x}{9} \le \sum_\rho \frac{x^{\Re(\rho)}}{|\rho+\frac{1}{2}|^2} + \sum_{\rho_\chi} \frac{x^{\Re(\rho_\chi)}}{|\rho_\chi+\frac{1}{2}|^2} + O(\log m)$$

Then, assuming GRH, we find that this is equal to

$$\frac{4x}{9} \le \sqrt{x}\sum_\rho \frac{1}{|\rho+\frac{1}{2}|^2} + \sqrt{x}\sum_{\rho_\chi} \frac{1}{|\rho_\chi+\frac{1}{2}|^2} + O(\log m)$$

Now, assuming GRH, the first sum is a constant, and the second sum is $O(\log m)$. Putting everything together, we have $\frac{4x}{9} \le O(\sqrt{x}\log m)$; this gives a contradiction unless $x = O((\log m)^2)$. □

Similar proofs can show the following interesting facts:

Proposition 10.5. *Assume that the Generalized Riemann Hypothesis is true. Then, if H is a proper subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$, there exists a positive integer x coprime to n less than $3(\log n)^2$ such that $x \notin H$.*

Corollary 10.5.1. *For any n, $(\mathbb{Z}/n\mathbb{Z})^\times$ is generated by $\{n \in \mathbb{Z}^+ \mid n < 3(\log n)^2\}$.*

10.3 Applications to Primality Testing

The fundamental question of this section is "given n, how can we check if n is prime?" Now, the naive method is to divide n by each positive integer up to $\lfloor\sqrt{n}\rfloor$; if ever the division leaves no remainder, then n is composite (and if it does not, then n is prime). Now, this algorithm is polynomial in n. Using the usual heuristic of theoretical computer scientists "P means efficient", one might believe that this suffices. However, in this case, the size of the input is $\log n$, since n is represented with n bits. Therefore, with respect to the size of the input, this algorithm is actually exponential. An efficient algorithm, then, would be polylogarithmic in n (i.e. polynomial in $\log n$, the size of the input).

Now, the naive method can be sped up slightly using the sieve of Eratosthenes, but the result is still an algorithm which is polynomial in n. Therefore, we need to come up with something more clever. Our first idea is Fermat's Test. Fermat's Test relies on Fermat's Little Theorem, which states that if p is prime, then $a^{p-1} \equiv 1 \bmod p$ for all nonzero a. On the other hand, if $(a, n) > 1$, then $a^{n-1} \not\equiv 1 \bmod n$. Therefore, n is not prime if and only if there exists some a (called a "Fermat witness") such that $a^{n-1} \not\equiv 1 \bmod n$.

Therefore, we have the following idea for the algorithm: pick a positive integer a and compute $a^{n-1} \bmod n$ (which can be done in polylogarithmic time using repeated squaring). If $a^{n-1} \not\equiv 1 \bmod n$, then we have found a Fermat witness, so n is composite. On the other hand, if $a^{n-1} \equiv 1 \bmod n$, then a serves as evidence that n is prime.

The problem with this algorithm comes in two parts:

1. First, there exist positive integers such that there are very few non-coprime numbers less than them.
2. Second, there exist numbers, called *Carmichael numbers*, such that $a^{n-1} \equiv 1 \bmod n$ if $(a, n) = 1$.

Combining this two problems, we see that often it is very hard to find Fermat witnesses, so this algorithm does not provide very strong evidence of primality.

The first problem can be formalized by considering $n = pq$, where p and q are primes. Then,

$$\frac{\varphi(n)}{n} = \frac{pq - 1 - (p-1)(q-1)}{pq} = \frac{p+q-2}{pq} < \frac{1}{p} + \frac{1}{q}$$

so if p and q are similar in size, $\frac{\varphi(n)}{n} \approx \frac{2}{\sqrt{n}}$. This is a vanishingly small proportion: far too small to yield an effective approach. Now, the second problem amounts to finding examples: the smallest example of a Carmichael number is $561 = 3 \cdot 11 \cdot 17$.

Despite these flaws, when n is not a Carmichael number, Fermat witnesses are very common:

Theorem 10.6. *If n is composite, and there exists a Fermat witness a such that $(a, n) = 1$ but $a^{n-1} \not\equiv 1$. Then the proportion of integers which are Fermat witnesses in $\{1, \ldots, n-1\}$ is greater than 50%.*

Proof. First, notice that it suffices to show that the proportion of nonwitnesses in $(\mathbb{Z}/n\mathbb{Z})^\times$ is at least 50% (since then we can boost this percentage to strictly greater than 50% by including numbers not coprime to n). For this, notice that if $a^{n-1} \equiv 1 \bmod n$ and $b^{n-1} \equiv 1 \bmod n$ (i.e. a and b are nonwitnesses), then $(ab)^{n-1} \equiv 1 \bmod n$ and $(a^{-1})^{n-1} \equiv 1 \bmod n$ (thus ab and a^{-1} are nonwitnesses). Since clearly 1 is a nonwitness, the nonwitnesses form a subgroup. By hypothesis, they are not equal to the whole of $(\mathbb{Z}/n\mathbb{Z})^\times$. But then the nonwitnesses must form at most 50% of it by Langrange's Theorem, so the witnesses must form at least 50% of $(\mathbb{Z}/n\mathbb{Z})^\times$, as desired. □

As a side note, we expect this bound to be tight. Indeed, it is tight assuming the following conjecture:

Conjecture 10.2. *There are infinitely many primes p such that $(2p - 1)$ is also prime.*

Assuming this conjecture, we can show that this bound is tight by considering the infinite sequence of numbers of the form $n = p(2p - 1)$ where p and $2p - 1$ both prime. It is easy to see that the number of residues which are witnesses in this case tends to $\frac{1}{2}$ as $p \to \infty$, which implies the bound is sharp.

This inspires the following algorithm: pick $a \in \{1, \ldots, n-1\}$ randomly. Check if $a^{n-1} \equiv 1(n)$: if not, stop, otherwise repeat. After x tries, if the number if neither Carmichael nor prime, the probability that one will not have found a witness is $1 - \frac{1}{2^x}$, which vanishes very quickly.

Nonetheless, due to the existence of Carmichael numbers, we want a better test.

Proposition 10.7 (Miller-Rabin Test). *Suppose that p is an odd prime, and $p-1 = 2^e k$ where k is odd and $e \geq 1$. Then, either each $a \in \{1, \dots, p-1\}$ has $a^k \equiv 1 \bmod p$ or $a^{2^i k} \equiv -1 \bmod p$, some $i \in \{0, \dots, c-1\}$. This is not true for composite n: if a fails the condition for some n, say that a is an* Miller-Rabin witness *to n being composite.*

Proof. Now, $a^{p-1} - 1 \equiv 0 \bmod p$ for each $1, \dots, p-1$. But then

$$\begin{aligned} a^{2^e k} - 1 &= (a^{2^{e-1}k} - 1)(a^{2^{e-1}k} + 1) = (a^{2^{e-2}k} - 1)(a^{2^{e-2}k} + 1)(a^{2^{e-1}k} + 1) = \cdots \\ &= (a^k - 1)(a^k + 1)(a^{2k+1}) \cdots (a^{2^{e-1}k} + 1) \equiv 0 \bmod p \end{aligned}$$

and since $\mathbb{Z}/p\mathbb{Z}$ is a field, this implies that one of these terms, as desired. On the other hand, if n is composite, then any a which is not coprime to n fails the test. □

This is more complicated than Fermat's Test: however, there are far more Miller-Rabin witnesses than Fermat witnesses. The following theorems show that the Miller-Rabin test suffices as a good probabilistic test:

Theorem 10.8. *Suppose that $n > 1$ is odd and composite. Then the Miller-Rabin nonwitnesses coprime to n lie in a proper subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$, so in particular they form strictly less than 50% of the numbers in $\{1, \dots, n-1\}$.*

Proof. This is more complicated; we cite Conrad at this reference. □

Indeed, one can even extend this result to show the following:

Theorem 10.9. *Suppose that $n > 1$ is odd and composite. Then the Miller-Rabin nonwitnesses coprime to n form at most 25% of the numbers in $\{1, \dots, n-1\}$.*

Proof. The same reference contains the proof of this fact. □

As a side note, this latter bound is likely to be sharp, again by anlyzing $n = p(2p-1)$ where both p and $2p-1$ are prime.

Now, this result inspires the following algorithm, called the Miller-Rabin test, to test if n is prime.

1. Check if n is even by looking at the last digit of n.
2. Draw a randomly from $1, \dots, n-1$.
3. Compute $a^k, a^{2k}, a^{4k}, \dots, a^{2^e k}(n)$; this times time $O(\log(n)^3)$ with naive multiplication, but with FFT multiplication it can be sped up to $O(\log(n) \log\log(n))$.
4. If a is a Miller-Rabin witness, stop. Otherwise, return to step 2.

If n is not prime, then the probability of having no witnesses after x repetitions is $1 - \frac{1}{4^x}$. In practice, this algorithm (in particular the version using FFT multiplication) is used, but it would be nice to know that it can be derandomized into a deterministic algorithm.

But our results from the previous section allow us to do that: since, if GRH holds, every proper subgroup $H \subsetneq (\mathbb{Z}/n\mathbb{Z})^\times$ leaves out an integer of size less than $2(\log n)^2$ (and earlier we showed that the nonwitnesses lie in a proper subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$), we can just run the test on $1, 2, \dots, 2(\log n)^2$ to get a deterministic algorithm running in $O(\log(n)^3 \log\log(n))$ time. Now, there exists a deterministic algorithm for primes (AKS) which runs in polylog time unconditionally, but this algorithm is significantly slower than Miller-Rabin assuming GRH, so Miller-Rabin is still what is used in practice.

11 Waring's Problem

In this section, we will discuss Linnik's proof of Waring's Problem:

Problem 3 (Waring's Problem)**.** Given any k, does there exist $g(k)$ such that any positive integer n can be written as the sum of $g(k)$ kth powers.

It has been well known for centuries (via Lagrange, 1770) that any natural number can be expressed as the sum of four perfect squares. A proof of this fact is offered in my notes on Elementary Number Theory. In the same year, Waring conjectured that for any positive integer k, there exists $g(k)$ such that any natural number can be expressed as the sum of $g(k)$ perfect kth powers. Our proof does not offer precise bounds on $g(k)$, though small values are known: $g(3) = 9$, $g(4) = 19$, $g(5) = 37$, and $g(6) = 73$.

This problem is related to the field of "additive number theory", which concerns problems about "sumsets" $A+B$. Indeed, Waring's Problem is precisely a statement about the size of the sumset $N_k + \cdots + N_k$, where N_k is the set of kth powers.

The proof strategy is as follows: first, show that any set A of positive integers with positive Shnrirelman density has a positive integer g_A such that any positive integer can be represented as the sum of g_A elements in A, and then, show that the set of kth powers has positive Shnrirelman density. The former is an elementary couting argument, and the latter uses complicated Fourier analysis.

This section is sourced from this Bachelor's thesis and this paper.

11.1 Shnirelman Density

Definition 11.1. Let $A(n)$ denote the set of positive integers in A not exceeding n. By abuse of notation, we will also use $A(n)$ to denote the cardinality of this set.

Definition 11.2 (Shnirelman Density)**.** The *Shnirelman density* of a set A is $\sigma(A) = \inf_n \frac{A(n)}{n}$.

It follows that the Shnirelman density lies between 0 and 1.

Definition 11.3 (Sumsets)**.** Let A and B be sets of integers. Then the *sumset* $A+B$ is equal to $\{a+b \mid a \in A, b \in B\}$. Similarly, the sumset gA is equal to

$$\underbrace{A + \cdots + A}_{g \text{ times}}.$$

Definition 11.4 (Basis)**.** Suppose that A is a set of positive integers such that hA contains every positive integer. Then A is said to be a *basis of order h*.

Our goal, then, is to show that the set of kth powers is a basis of finite order. In this section, we show that any set with positive Shnirelman density is a basis of finite order.

Lemma 11.5. *Let A and B be sets of integers such that $0 \in A$ and $0 \in B$. If $n \geq 0$ and $A(n) + B(n) \geq n$, then $n \in A + B$.*

Proof. The result follows immediately if $n \in A$ or $n \in B$. Thus, we may assume that $n \notin A$ and $n \notin B$. Then, $A(n-1)+B(n-1) \geq n$, and applying the Pigeonhole Principle to the sets $n-A(n-1)$ and $B(n-1)$ we find that they have nonempty intersection. Yet if $k \in B(n-1) \cap (n - A(n-1))$, it follows that $(n-k)+k = n$ where $n-k \in A$ and $k \in B$, so $n \in A + B$. □

Corollary 11.5.1. *Let A and B be sets of integers such that $0 \in A$ and $0 \in B$. Then $\sigma(A) + \sigma(B) \geq 1$ implies that $n \in A + B$ for every non-negative integer n.*

Lemma 11.6. *Let A and B be sets of itnegers such that $0 \in A$ and $0 \in B$. Then, $\sigma(A + B) \geq \sigma(A) + \sigma(B) - \sigma(A)\sigma(B)$.*

Proof. First, fix n; we will give a lower bound on $(A+B)(n)$. Let $A(n) = \{a_1, \dots, a_k\}$ and $a_0 = 0$. Then, clearly, $a_i, \dots, a_k \in A+B$, which contributes $A(n)$ to $(A+B)(n)$. Furthermore, for any $i < k$ and any $b \in B(a_{i+1} - a_i + 1)$, the elements $a_i + b \in A+B$ lie strictly between a_i and a_{i+1}. This contributes $\sum_{i=1}^{k-1} B(a_{i+1} - a_i - 1)$ to $(A+B)(n)$. Finally, notice that if $b \in B(n - a_k)$, then $a_k + b \in A+B$ lie strictly above a_k. This contributes $B(n-a_k)$ to the sum. Putting this all together, we have

$$\begin{aligned}(A+B)(n) &\geq A(n) + \sum_{i=1}^{k-1} B(a_{i+1} - a_i - 1) + B(n-a_k) \\ &\geq A(n) + \sigma(B)\sum_{i=0}^{k-1}(a_{i+1} - a_i) - \sigma(B)k + \sigma(B)n - \sigma(B)a_k \\ &\geq A(n) + \sigma(B)a_k - \sigma(B)k + \sigma(B)n - \sigma(B)a_k \\ &= A(n) - \sigma(B)k + \sigma(B)n \\ &\geq n\sigma(A) + n\sigma(B) - n\sigma(A)\sigma(B) \\ &= n(\sigma(A) + \sigma(B) - \sigma(B)\sigma(B)).\end{aligned}$$

The result follows. □

Corollary 11.6.1. $1 - \sigma(A+B) \leq (1-\sigma(A))(1-\sigma(B))$

Proof. Rearrangement of the above result. □

Corollary 11.6.2. $1 - \sigma(A_1 + \cdots + A_n) \leq \prod_{i=1}^{n}(1 - \sigma(A_i))$.

Proof. Induction. □

We have finally arrived at the central result of this section.

Theorem 11.7. *Suppose that A satisfies $\sigma(A) > 0$. Then A is a basis of order g for some finite g.*

Proof. Suppose that $\sigma(A) > 0$. Then $1 - \sigma(A_i) < 1$, and there exists some h such that $\prod_{i=1}^{h}(1-\sigma(A)) < \frac{1}{2}$. But then $\sigma(hA) > \frac{1}{2}$ by the preceding result. Then, $\sigma(hA) + \sigma(hA) \geq 1$, so $2hA$ contains every positive integer and $g = 2h$ suffices. □

11.2 Towards Positive Density

Fix $k \geq 2$ and set $g(k) = 8^{k-1}$. Then, let A' be the set of perfect kth powers and define $A = g(k)A'$. Define $r(n)$ to be the number of solutions in non-negative integers to the equation $x_1^k + \cdots + x_{g(k)}^k = n$.

Theorem 11.8. *The set A has positive Shnirelman density.*

Proof.

$$\sum_{n=0}^{N} r(n) = \sum_{\substack{x_1, \dots, x_{g(k)} \geq 0 \\ x_1^k + \cdots x_{g(k)}^k \leq N}} 1 \geq \sum_{0 \leq x_1, \dots, x_{g(k)} \leq (N/g(k))^{1/k}} 1 \geq (N/g(k))^{g(k)/k}.$$

Thus, it follows that $\sum_{n=0}^{N} r(n) \gg_k N^{g(k)/k}$. On the other hand,

$$r(n) = \sum_{\substack{0 \leq x_1, \dots, x_{g(k)} \leq n^{1/k} \\ x_1^k + \cdots + x_{g(k)}^k = n}} 1 = \sum_{0 \leq x_1, \dots, x_{g(k)} \leq n^{1/k}} \int_0^1 e^{2\pi i(x_1^k + \cdots + x_{g(k)}^k - n)\alpha} d\alpha = \int_0^1 \left(\sum_{0 \leq x \leq n^{1/k}} e^{2\pi i x^k \alpha} \right)^{g(k)} e^{-2\pi i n\alpha} d\alpha.$$

Our goal is to show that

$$\int_0^1 \left| \sum_{0 \leq x \leq n^{1/k}} e^{2\pi i x^k \alpha} \right|^{g(k)} d\alpha \ll_k n^{g(k)/k - 1}.$$

or, equivalently, that

$$\int_0^1 \left| \sum_{0 \le x \le N} e^{2\pi i x^k \alpha} \right|^{g(k)} d\alpha \ll_k N^{g(k)-k}.$$

This would demonstrate that $r(n) \ll_k n^{g(k)/k-1}$. But then, $\sum_{n=0}^N r(n) = r(0) + \sum_{n=1}^N r(n) \ll_k 1 + N^{g(k)/k-1} \cdot A(N)$. Applying the earlier result, $N^{g(k)/k} \ll_k \sum_{n=0}^N r(n) \ll_k 1 + N^{g(k)/k-1} \cdot A(N)$ whence $1 \ll_k \frac{1}{N^{g(k)/k}} + \frac{A(N)}{N}$. Thus, there exists a constant $c > 0$ such that $c < \frac{1}{N^{g(k)/k}} + \frac{A(N)}{N}$; in particular, $\frac{A(N)}{N} > \frac{c}{2}$ for all sufficiently large N. The result follows. □

Thus, we have reduced the result to proving the following theorem:

Theorem 11.9. *For $f(x) = x^k - n$,*

$$\int_0^1 \left| \sum_{0 \le x \le N} e^{2\pi i f(x) \alpha} \right|^{g(k)} d\alpha \ll_k N^{g(k)-k}.$$

In fact, we will prove a stronger statement.

11.3 Preliminary Results for the Main Proof

First, we need computations relating to solutions to linear Diophantine equations.

Lemma 11.10. *Let m_1 and m_2 be integers, not both zero, and let $q(n, m_1, m_2)$ be the number of integer solutions to the equation $x_1 m_1 + x_2 m_2 = n$ with $x_1, x_2 \in [-N, N]$. If $g = \gcd(m_1, m_2) \mid n$, then*

$$q(n, m_1, m_2) \le \frac{2N}{\max(|m_1/g, |m_2/g|)} + 1.$$

Proof. First, by dividing through by g, we may assume that $g = 1$. Then, let $x_1 m_1 + x_2 m_2 = 1$ have solution $x_1 = \overline{m}_1$ and $x_2 = \overline{m}_2$. The general solution to $x_1 m_1 + x_2 m_2 = n$ is $x_1 = n\overline{m}_1 + k m_2$ and $x_2 = n\overline{m}_2 - k m_1$. Then, the number of possibilities for k is bounded above by $\frac{2N}{|m_1/g|} + 1$ and $\frac{2N}{|m_1/g|} + 1$, as desired. □

Lemma 11.11. *Let $q(n)$ be the number of integer solutions to the equation $x_1 m_1 + x_2 m_2 = n$ with $x_1, x_2 \in [-N, N]$ and $m_1, m_2 \in [-M, M] \setminus \{0\}$. Then,*

$$q(n) \le \begin{cases} 20MN \sum_{g \mid n} \frac{1}{g} & \textit{if } n \neq 0 \textit{ and } N \ge M \\ 20M^2 N & \textit{if } n = 0. \end{cases}$$

Proof. This is a direct computation using the above result.

$$\begin{aligned} q(n) = 4 \sum_{m_1, m_2 = 1}^{M} q(n, m_1, m_2) &\le 4M^2 + 8N \sum_{\substack{g \mid n \\ g \le M}} \sum_{\substack{1 \le a_1 \le M/g \\ 0 \le a_2 \le M/g}} \frac{1}{\max(|a_1|, |a_2|)} \\ &\le 4M^2 + 8N \sum_{\substack{g \mid n \\ g \le M}} \left(\sum_{1 \le a_1 \le M/g} \sum_{1 \le a_2 < a_1} \frac{1}{a_1} + \sum_{1 \le a_2 \le M/g} \sum_{1 \le a_a < a_2} \frac{1}{a_2} \right) \\ &= 4M^2 + 16N \sum_{\substack{g \mid n \\ g \le M}} \sum_{1 \le a \le M/g} 1 \le 4M^2 + 16NM \sum_{\substack{g \mid n \\ g \le M}} \frac{1}{g}. \end{aligned}$$

Then, splitting into the cases $n = 0$ and $n \neq 0$, the result follows. □

Next, we need a series of results related to variations of the Riemann ζ-function.

Lemma 11.12.

$$\sum_{n\leq N}\left(\sum_{d|n}\frac{1}{d}\right)^2 \leq \frac{5N}{2}\zeta(3)$$

Proof. Now,

$$\left(\sum_{n=1}^{\infty}\frac{1}{n^2}\right)^2 = \sum_{d,e=1}^{\infty}\frac{1}{d^2e^2} = \sum_{\substack{g,a,b=1\\ \gcd(a,b=1}}\frac{1}{g^4a^2b^2} = \left(\sum_{n=1}^{\infty}\frac{1}{n^4}\right)\sum_{\substack{a,b=1\\ \gcd(a,b)=1}}^{\infty}\frac{1}{a^2b^2}.$$

Thus,

$$\sum_{\substack{a,b=1\\ \gcd(a,b)=1}}^{\infty}\frac{1}{a^2b^2} = \frac{\left(\sum_{n=1}^{\infty}\frac{1}{n^2}\right)^2}{\sum_{n=1}^{\infty}\frac{1}{n^4}} = \frac{\pi^4/36}{\pi^4/90} = \frac{5}{2}.$$

Then, $\sum_{n\leq N}\left(\sum_{d|n}\frac{1}{d}\right)^2 = \sum_{n\leq N}\sum_{d,e|n}\frac{1}{de} \leq \sum_{d,e\leq N}\frac{N}{d\operatorname{lcm}(d,e)}$. Let $\gcd(d,e)=g$, $d=ga$, $e=gb$. Then,

$$\sum_{d,e\leq N}\frac{N}{d\operatorname{lcm}(d,e)} = N\sum_{\substack{g,a,b=1\\ \gcd(a,b)=1}}^{N}\frac{1}{g^3a^2b^2} \leq N\sum_{p=1}^{\infty}\frac{1}{p^3}\sum_{\substack{a,b=1\\ \gcd(a,b)=1}}^{\infty}\frac{1}{a^2b^2} = \frac{5N}{2}\sum_{p=1}^{\infty}\frac{1}{p^3}.$$

□

Finally, these are the lemmas which we will directly use in the proof of the result:

Lemma 11.13. *Suppose that $N \geq M$. Then,*

$$\sum_{\substack{m_1,\ldots,m_4=-M\\ m_1,\ldots,m_4\neq 0}}^{M}\ \sum_{\substack{n_1,\ldots,n_4=-N\\ m_1n_1+m_2n_2=m_3n_3+m_4n_4}}^{N} 11 \leq 5250(MN)^3.$$

Proof. First, notice that by applying the previous two lemmas,

$$\begin{aligned}\sum_{\substack{m_1,\ldots,m_4=-M\\ m_1,\ldots,m_4\neq 0}}^{M}\ \sum_{\substack{n_1,\ldots,n_4=-N\\ m_1n_1+m_2n_2=m_3n_3+m_4n_4}}^{N} 1 &= \sum_{n=-2MN}^{2MN} q(n)^2 = q(0)^2 + \sum_{n=1}^{2MN} q(n)^2\\ &= 20^2\left((M^4N^2 + 2M^2N^2\sum_{n=1}^{2MN}\left(\sum_{g|n}\frac{1}{g}\right)^2\right)\\ &\leq 20^2\left(M^4N^2 + 10(MN)^3\zeta(3)\right) \leq 20^2(MN)^3\left(1+10\zeta(3)\right).\end{aligned}$$

But $\left(1+10\sum_{p=1}^{\infty}\frac{1}{p^3}\right) \leq 1 + 10\cdot 1.203 = 13.03$, and $20^2\cdot 13.03 = 5212 \leq 5250$. The result follows. □

Lemma 11.14. *Suppose that $N \geq 2M$. Then,*

$$\sum_{m_1,\ldots,m_4=-M}^{M}\ \sum_{\substack{n_1,\ldots,n_4=-N\\ m_1n_1+m_2n_2=m_3n_3+m_4n_4}}^{N} 1 \leq 162M^4 + 5250(MN)^3.$$

Proof. First, define $Q(n)$ to be the integer solutions to the equation $x_1m_1 + x_2m_2 = n$ with $x_1, x_2 \in [-N, N]$ and $m_1, m_2 \in [-M, M]$. Now, when $n \neq 0$, at least one of m_1 and m_2 must be nonzero; since $N \geq 2M$, the estimate given in Lemma 11.11 is a significant overestimate and one may check that the same bound works for $Q(n)$. On the other hand, in the case $n = 0$, we have the additional term $q(n; 0, 0) = (2M + 1)^2 \leq 9M^2$. Thus, while $q(0)^2$ is bounded by $20^2(MN)^3$, $Q(0)^2$ is bounded by

$$(9M^2 + 20M^2N)^2 \leq 2(9M^2)^2 + 2(20M^2N)^2 = 162M^4 + 2\cdot 20^2M^4N^2 \leq 162M^4 + 20^2(MN)^3.$$

Then, repeating the steps in the previous proof, we obtain the desired estimate. □

11.4 Linnik's Proof of Positive Density

There are two parts to this result, which we prove by induction: a base case, and an inductive step.

Lemma 11.15 (Base Case)**.** *Let $f(n) = a_2 n^2 + a_1 n$ where a_1, a_2 are integers with $a_2 \neq 0$, $|a_1| \leq c_1 N$, and $|a_2| \leq c_2$. Then, if $C = 162(2c_2 + c_1)^4 + 5250(2c_2 + c_1)^3$,*

$$\int_0^1 \left| \sum_{n=0}^{N} e^{2\pi i f(n)\alpha} \right|^8 d\alpha \leq CN^6.$$

Proof.

$$\int_0^1 \left| \sum_{n=0}^{N} e^{2\pi i f(n)\alpha} \right|^8 d\alpha \leq \int_0^1 \left(\sum_{n=0}^{N} e^{2\pi i f(n)\alpha} \right)^4 \left(\sum_{n=0}^{N} e^{-2\pi i f(n)\alpha} \right)^4 d\alpha$$

$$= \int_0^1 \sum_{n_1,\ldots,n_8=0}^{N} e^{2\pi i (f(n_1)+\cdots+f(n_4)-f(n_5)-\cdots-f(n_8))\alpha} d\alpha = \sum_{\substack{n_1,\ldots,n_8=0 \\ f(n_1)+\cdots+f(n_4)=f(n_5)+\cdots+f(n_8)}}^{N} 1.$$

Now, the equation $f(n_1) + \cdots + f(n_4) = f(n_5) + \cdots + f(n_8)$ can be written as $\sum_{i=1}^4 m_i x_i = 0$ where $m_i = n_i - n_{i+4}$ and $x_i = a_2(n_i + n_{i+4}) + a_1$. Then, $-N \leq m_i \leq N$, while $-N(2c_2 + c_1) \leq x_i \leq N(2c_2 + c_1)$; thus, the final sum above is bounded above by

$$\sum_{m_1,\ldots,m_4=-N}^{N} \sum_{\substack{x_1,\ldots,x_4=-N(2c_2+c_1) \\ m_1x_1+\cdots+m_4x_4=0}}^{N(2c_2+c_1)} 1 = \sum_{m_1,\ldots,m_4=-N}^{N} \sum_{\substack{x_1,\ldots,x_4=-N(2c_2+c_1) \\ m_1x_1+m_2x_2=m_3x_3+m_4x_4=0}}^{N(2c_2+c_1)} 1 \leq 162N^4(2c_1+c_1)^4 + 5250N^6(2c_2+c_1)^3.$$

where the final inequality is given by Lemma 11.13. □

Theorem 11.16 (Full Result)**.** *Let $k \geq 2$ and $f(n) = a_k n^k + \cdots + a_1 n$, where $a_1, \ldots, a_k$ are integers with $a_k \neq 0$ and $|a_j| \leq c_{j,k} N^{k-j}$ for fixed constants $c_{j,k}$. Then,*

$$\int_0^1 \left| \sum_{n=0}^{N} e^{2\pi i \alpha f(n)} \right|^{8^{k-1}} d\alpha \ll_k N^{8^{k-1}-k}.$$

Proof. The case $k = 2$ is given by Lemma 11.15. Then, we will prove the result by induction.

$$\left| \sum_{n=0}^{N} e^{2\pi i \alpha f(n)} \right|^2 = \sum_{m,n=0}^{N} e^{2\pi i (\alpha f(m) - \alpha f(n))} = N + 1 + \sum_{\substack{h=-N \\ h \neq 0}}^{N} b_h$$

where $b_h = \sum_{\substack{m,n=0 \\ m-n=h}}^{N} e^{2\pi i(\alpha f(m) - \alpha f(n))} = \sum_{n=\max(0,-h)}^{\min(N,N-h)} e^{2\pi i \alpha h \phi(n,h)}$ where $\phi(n,h) = \frac{1}{h} f(n+h) - f(n)$ is a degree $k-1$ polynomial in n. Furthermore, the n^{k-1} coefficient of $\phi(n,h)$ is equal to $ka_k \neq 0$. Finally, the n^r coefficient of $\phi(n,h)$ is dominated by N^{k-r-1}; that is, the n^r coefficient of $\phi(n,h)$ is equal to

$$\sum_{j=r+1}^{k} \binom{j}{r} a_j h^{j-r-1} \ll \sum_{j=r+1}^{k} \binom{j}{r} N^{k-j} N^{j-r-1} \ll N^{k-r-1}.$$

Now, by Hölder's Inequality applied to sums,

$$\left| \sum_{n=0}^{N} e^{2\pi i \alpha f(n)} \right|^{2 \cdot 8^{k-2}} \ll N^{8^{k-2}} + \left| \sum_{\substack{h=-N \\ h \neq 0}}^{N} b_h \right|^{8^{k-2}} \ll N^{8^{k-2}} + N^{8^{k-2}-1} \sum_{\substack{h=-N \\ h \neq 0}}^{N} |b_h|^{8^{k-2}}.$$

Then, raising it to a fourth power and integrating over α yields

$$\int_0^1 \left|\sum_{n=0}^{N} e^{2\pi i\alpha f(n)}\right|^{8^{k-1}} d\alpha \ll N^{4\cdot 8^{k-2}} + N^{4\cdot 8^{k-2}-4}\int_0^1 \left(\sum_{\substack{h=-N\\h\neq 0}}^{N} |b_h|^{8^{k-2}}\right)^4 d\alpha.$$

As a function of α, b_h has period $\frac{1}{|h|}$. let $|b_h|^{8^{k-2}}$ have Fourier series $|b_h|^{8^{k-2}} = \sum_{m=-\infty}^{\infty} A(m,h)e^{\alpha hm}$. But notice that, by definition of b_h, these coefficients are eventually 0; that is, $A(m,h)\neq 0$ implies that $|m| \ll \max_{0\leq n\leq N}|\phi(n,h)| \ll N^{k-1}$. Thus, there exists C such that $A(m,h)=0$ for all $m > CN^{k-1}$.

$$\begin{aligned}
A(m,h) = \int_0^1 |b_h|^{8^{k-2}} e^{-2\pi i\alpha hm} d\alpha &= \int_0^{|h|} \left|\sum_{n=\max(0,-h)}^{\min(N,N-h)} e^{2\pi i\,\mathrm{sgn}(h)\beta\phi(n,h)}\right| \frac{e^{-2\pi i\,\mathrm{sgn}(h)\beta m}}{|h|} d\beta \\
&= \int_0^1 \left|\sum_{n=\max(0,-h)}^{\min(N,N-h)} e^{2\pi i\beta\phi(n,h)}\right| e^{-2\pi i\,\mathrm{sgn}(h)\beta m} d\beta \ll N^{8^{k-2}-(k-1)}.
\end{aligned}$$

Then, finally,

$$\begin{aligned}
\int_0^1 \left(\sum_{\substack{h=-N\\h\neq 0}}^{N} |b_h|^{8^{k-2}}\right)^4 d\alpha &= \int_0^1 \left(\sum_{\substack{h=-N\\h\neq 0}}^{N}\sum_{|m|\leq CN^{k-1}} A(m,h)e^{2\pi i\alpha hm}\right)^4 d\alpha \\
&= \sum_{\substack{h_1,\ldots,h_4=-N\\h_1,\ldots,h_4\neq 0}}^{N} \sum_{|m_1|,\ldots,|m_4|\leq CN^{k-1}} \left(\prod_{j=1}^4 A(m_j,h_j)\right)\int_0^1 e^{2\pi i\alpha\sum_{j=1}^4 h_j m_j} d\alpha \\
&\ll \sum_{\substack{h_1,\ldots,h_4=-N\\h_1,\ldots,h_4\neq 0}}^{N} \sum_{\substack{|m_1|,\ldots,|m_4|\leq CN^{k-1}\\h_1m_1+\cdots+h_4m_4=0}} \prod_{j=1}^4 A(m_j,h_j) \\
&\ll N^{4(8^{k-2}-(k-1)} \sum_{\substack{|m_1|,\ldots,|m_4|\leq CN^{k-1}\\h_1m_1+\cdots+h_4m_4=0}} 1 = N^{4\cdot 8^{k-2}-k+4}.
\end{aligned}$$

Then, putting everything together,

$$\int_0^1 \left|\sum_{n=0}^{N} e^{2\pi i\alpha f(n)}\right|^{8^{k-1}} d\alpha \ll N^{4\cdot 8^{k-2}} + N^{4\cdot 8^{k-2}-4}N^{4\cdot 8^{k-2}-k+4} \ll N^{4\cdot 8^{k-2}} + N^{8^{k-1}-k} \ll N^{8^{k-1}-k}.$$

□

Thus, as discussed previously, the result is proven.

12 Appendix of Miscellaneous Results

12.1 Bounding the k-Divisor Function

Recall we proved earlier that the divisor function $d(n) = \sum_{ab=n} 1$ satisfies $d(n) \ll_\varepsilon n^\varepsilon$ for any $\varepsilon > 0$.

Definition 12.1 (k-Divisor Function)**.** The *k-divisor function* is defined by

$$d_k(n) = \sum_{a_1 \cdots a_k = n} 1,$$

that is, it counts the number of ways to write n as a product of k factors. In particular, $d(n) = d_2(n)$.

Theorem 12.2. *$d_k(n) \ll_\varepsilon n^\varepsilon$ for any $\varepsilon > 0$.*

Proof. Suppose that $n = p_1^{e_1} \cdots p_J^{e_J}$. Then, in exactly the same way as we deduced the corresponding result for the usual divisor function, we find that if $f(e)$ is the number of ways to write e as the ordered sum of k non-negative terms, then

$$\frac{d_k(n)}{n^\varepsilon} = \prod_{j=1}^{J} \frac{f(e_j)}{p_j^{\varepsilon e_j}}.$$

It is a fact of elementary combinatorics that $f(e) = \binom{e+k-1}{k-1}$. Now, define the function $t_{p,\varepsilon}(e) = \frac{f(e)}{p^{\varepsilon e}}$. I claim that for sufficiently large p, $\max_{e\in\mathbb{N}}\{t_{p,\varepsilon}(e)\} = 1$. To see why, notice that $t_{p,\varepsilon}(0) = 1$, and for p large enough that $p^\varepsilon > k$, we have

$$p^\varepsilon > k \Rightarrow p^\varepsilon > \frac{e+k}{e+1} \Rightarrow p^\varepsilon \frac{(e+k-1)!}{e!(k-1)!} > \frac{(e+k)!}{(e+1)!(k-1)!} \Rightarrow \binom{e+k-1}{k-1} > \frac{\binom{e+k}{k-1}}{p^\varepsilon} \Rightarrow \frac{\binom{e+k-1}{k-1}}{p^{\varepsilon e}} > \frac{\binom{e+k}{k-1}}{p^{\varepsilon(e+1)}}$$

which means precisely that $t_{p,\varepsilon}(e) > t_{p,\varepsilon}(e+1)$ for all e. Therefore, for such primes p, $t_{p,\varepsilon}(0) = 1$ is the maximal value attained by $t_{p,\varepsilon}(0)$. Now, say P is the largest prime which does not satisfy $p^\varepsilon > k$. Then,

$$\frac{d_k(n)}{n^\varepsilon} = \prod_{j=1}^{J} t_{p_j,\varepsilon}(e_j) \le \prod_{j=1}^{J} \max_{e\in\mathbb{N}}\{t_{p_j,\varepsilon}(e)\} \le \prod_{\text{prime } p} \max_{e\in\mathbb{N}}\{t_{p,\varepsilon}(e)\} = \prod_{\substack{\text{prime } p\\ p\le P}} \max_{e\in\mathbb{N}}\{t_{p_j,\varepsilon}(e)\}$$

where the second equality follows because for all $p > P$, the maximum is 1 (as we mentioned earlier). Yet then the right-hand side is a finite product, and therefore yields a finite constant C_ε. Therefore, $\frac{d_k(n)}{n^\varepsilon} \le C_\varepsilon$ for some constant $C_\varepsilon > 0$, so by definition the result follows. □

12.2 A Stronger Result on the Variance of $\omega(n)$

In this section, we will prove the following theorem.

Theorem 12.3. *Let B be the constant in the asymptotic*

$$\sum_{p\le x} \frac{1}{p} = \log\log x + B + O\left(\frac{1}{\log x}\right).$$

Then, our computation of the variance of $\omega(n)$ can be refined into the following more precise version:

$$\sum_{n\le x} (\omega(n) - \log\log x - B)^2 \sim x \log\log x.$$

For the rest of this section, p and q will always denote primes. We begin with four technical results.

Lemma 12.4.

$$\sum_{\substack{p,q\\ pq\le x}} \frac{1}{pq} = \left(\sum_{p\le x}\frac{1}{p}\right)^2 - 2\sum_{p\le\sqrt{x}}\frac{1}{p}\sum_{x/p<q\le x}\frac{1}{q} - \left(\sum_{\sqrt{x}<p\le x}\frac{1}{p}\right)^2.$$

Proof. First, notice that upon expanding $\left(\sum_{p\leq x}\frac{1}{p}\right)^2$, we get the expression

$$\left(\sum_{p\leq x}\frac{1}{p}\right)^2=\sum_{p,q\leq x}\frac{1}{pq}=\sum_{pq\leq x}\frac{1}{pq}+\sum_{\substack{p,q\leq x\\ pq>x}}\frac{1}{pq}.$$

Now, the set over which the second part $\sum_{\substack{p,q\leq x\\ pq>x}}\frac{1}{pq}$ is indexed is the set of primes p and q such that $p,q\leq x$ and $pq>x$. This can be split up into three categories, each corresponding to a part of this sum:

1. p and q are both larger than $\sqrt{x}$ but at most x; this yields the sum $\left(\sum_{\sqrt{x}<p\leq x}\frac{1}{p}\right)^2$.
2. p is at most $\sqrt{x}$ and q is larger than x/p and at most x; this yields the sum $\sum_{p\leq\sqrt{x}}\sum_{x/p<q\leq x}\frac{1}{pq}$.
3. q is at most $\sqrt{x}$ and p is larger than x/q and at most x; this yields the sum $\sum_{q\leq\sqrt{x}}\sum_{x/q<p\leq x}\frac{1}{pq}$.

Notice that by symmetry, (2) and (3) are equal; that is, $\sum_{\substack{p,q\leq x\\ pq>x}}\frac{1}{pq}=\left(\sum_{\sqrt{x}<p\leq x}\frac{1}{p}\right)^2+2\sum_{p\leq\sqrt{x}}\sum_{x/p<q\leq x}\frac{1}{pq}$. Of course, factoring, we find that this is equal to

$$\sum_{\substack{p,q\leq x\\ pq>x}}\frac{1}{pq}=\left(\sum_{\sqrt{x}<p\leq x}\frac{1}{p}\right)^2+2\sum_{p\leq\sqrt{x}}\frac{1}{p}\sum_{x/p<q\leq x}\frac{1}{q}.$$

Plugging this into our earlier expression, we find that

$$\left(\sum_{p\leq x}\frac{1}{p}\right)^2=\sum_{pq\leq x}\frac{1}{pq}+\left(\sum_{\sqrt{x}<p\leq x}\frac{1}{p}\right)^2+2\sum_{p\leq\sqrt{x}}\frac{1}{p}\sum_{x/p<q\leq x}\frac{1}{q}.$$

Therefore, by rearranging, we have that

$$\sum_{\substack{p,q\\ pq\leq x}}\frac{1}{pq}=\left(\sum_{p\leq x}\frac{1}{p}\right)^2-2\sum_{p\leq\sqrt{x}}\frac{1}{p}\sum_{x/p<q\leq x}\frac{1}{q}-\left(\sum_{\sqrt{x}<p\leq x}\frac{1}{p}\right)^2.$$

□

Lemma 12.5. *For $p\leq\sqrt{x}$,*

$$\sum_{x/p<q\leq x}\frac{1}{q}=O\left(\frac{\log p}{\log x}\right).$$

Proof. Recall that $\sum_{q\leq x}\frac{1}{q}=\log\log x+C+O\left(\frac{1}{\log N}\right)$. Therefore,

$$\sum_{x/p<q\leq x}\frac{1}{q}=\log\log x-\log\log\left(\frac{x}{p}\right)+O\left(\frac{1}{\log x}\right)-O\left(\frac{1}{\log(x/p)}\right)$$

$$\log\log x-\log\log\left(\frac{x}{p}\right)=\log\left(\frac{\log x}{\log x-\log p}\right)=-\log\left(\frac{\log x-\log p}{\log x}\right)=-\log\left(1-\frac{\log p}{\log x}\right).$$

Yet the Taylor series of $\log(1-x)=-x-\frac{x^2}{2}-\frac{x^3}{3}-\cdots$, so we can write this as

$$-\log\left(1-\frac{\log p}{\log x}\right)=\sum_{n=1}^{\infty}\frac{(\log p/\log x)^n}{n}\leq\sum_{n=1}^{\infty}\left(\frac{\log p}{\log x}\right)^n=\frac{\log p/\log x}{1-\log p/\log x}.$$

Yet $p \leq \sqrt{x}$ implies that $\log p \leq \frac{1}{2}\log x$, whence $\log p/\log x \leq \frac{1}{2}$, so indeed

$$\log\log x - \log\log\left(\frac{x}{p}\right) = -\log\left(1 - \frac{\log p}{\log x}\right) \leq \frac{\log p/\log x}{1 - \log p/\log x} \leq 2\frac{\log p}{\log x} = O\left(\frac{\log p}{\log x}\right).$$

Furthermore, $\frac{1}{\log(x/p)} \leq \frac{1}{\log(\sqrt{x})} = \frac{2}{\log x}$. Therefore, $O\left(\frac{1}{\log(x/p)}\right) = O\left(\frac{1}{\log x}\right)$. Combining everything,

$$\sum_{x/p<q\leq x}\frac{1}{q} = O\left(\frac{\log p}{\log x}\right) + O\left(\frac{1}{\log x}\right) = O\left(\frac{\log p}{\log x}\right).$$

□

Lemma 12.6.

$$\sum_{\substack{p,q\\ pq\leq x}}\frac{1}{pq} = \left(\sum_{p\leq x}\frac{1}{p}\right)^2 - O(1).$$

Proof. We apply the result of the second technical lemma to the equation yielded by the first technical lemma in two places: the second part of the second term, and the third term with $p = \sqrt{x}$. This yields

$$\sum_{\substack{p,q\\ pq\leq x}}\frac{1}{pq} = \left(\sum_{p\leq x}\frac{1}{p}\right)^2 - 2\sum_{p\leq\sqrt{x}}\frac{1}{p}\cdot O\left(\frac{\log p}{\log x}\right) - O\left(\frac{\log(\sqrt{x})}{\log(x)}\right)^2 = \left(\sum_{p\leq x}\frac{1}{p}\right)^2 - 2\sum_{p\leq\sqrt{x}}\frac{1}{p}\cdot O\left(\frac{\log p}{\log x}\right) - O(1).$$

Now,

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

Substituting this result back in, we find that

$$\sum_{\substack{p,q\\ pq\leq x}}\frac{1}{pq} = \left(\sum_{p\leq x}\frac{1}{p}\right)^2 - 2O(1) - O(1) = \left(\sum_{p\leq x}\frac{1}{p}\right)^2 - O(1).$$

□

Next, we offer a more precise computation of $\sum_{n\leq x}\omega(n)^2$.

Lemma 12.7. *Let B be the constant in the asymptotic*

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

Then, $\sum_{n\leq x}\omega(n)^2 = x(\log\log x)^2 + (2B+1)x(\log\log x) + O(x)$.

Proof.

$$\sum_{n\leq x}\omega(n)^2 = \sum_{n\leq x}\left(\sum_{p|n}1\right)^2 = \sum_{n\leq x}\sum_{p|n}\sum_{q|n}1 = \sum_{p,q\leq x}\sum_{\substack{n\leq x\\ p,q|n}}1.$$

Now, if $p \neq q$, then $\sum_{\substack{n\leq x\\ p,q|n}}1 = \left\lfloor\frac{x}{pq}\right\rfloor = \frac{x}{pq} + O(1)$. On the other hand, if $p = q$, then $\sum_{\substack{n\leq x\\ p,q|n}}1 = \left\lfloor\frac{x}{p}\right\rfloor = \frac{x}{p} + O(1)$. Therefore, we need to compute

$$\sum_{\substack{pq\leq x\\ p\neq q}}\left(\frac{x}{pq} + O(1)\right) + \sum_{p\leq x}\left(\frac{x}{p} + O(1)\right) = \sum_{\substack{pq\leq x\\ p\neq q}}\left(\frac{x}{pq} + O(1)\right) + x\log\log x + O(x).$$

Now, we seek to compute

$$\sum_{\substack{pq\leq x\\ p\neq q}}\left(\frac{x}{pq}+O(1)\right)=\sum_{pq\leq x}\left(\frac{x}{pq}+O(1)\right)-\sum_{\substack{p=q\\ p^2\leq x}}\left(\frac{x}{p^2}+O(1)\right)=\sum_{pq\leq x}\left(\frac{x}{pq}\right)+O(x)-O(x)=\sum_{pq\leq x}\left(\frac{x}{pq}\right)+O(x).$$

Therefore,

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

Now, by the third technical lemma, the asymptotic formula for the reciprocal of the primes, and obvious expansion and simplification,

$$\sum_{pq\leq x}\frac{1}{pq}=\left(\sum_{p\leq x}\frac{1}{p}\right)^2+O(1)=\left(\log\log x+B+O\left(\frac{1}{\log x}\right)\right)^2+O(1)$$
$$=(\log\log x)^2+2B(\log\log x)+O(1).$$

Therefore, by multiplying by x and combining this with the earlier formula $\sum_{n\leq x}\omega(n)^2=\sum_{pq\leq x}\left(\frac{x}{pq}\right)+x\log\log x+O(x)$,

$$\sum_{n\leq x}\omega(n)^2=x(\log\log x)^2+(2B+1)x(\log\log x)+O(x).$$

□

Theorem 12.8. *Let B be the constant in the asymptotic*

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

Then, our computation of the variance of $\omega(n)$ can be refined into the following more precise version:

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

Proof. It follows from the Chebyshev bounds and our work in Section 3.1 that $\pi(x)=O\left(\frac{x}{\log x}\right)$ (this is weaker than the prime number theorem, as we are not claiming that $\pi(x)\sim O\left(\frac{x}{\log x}\right)$). Therefore, as a preliminary result, we notice that

$$\sum_{n\leq x}\omega(n)=\sum_{n\leq x}\sum_{p|n}1=\sum_{p\leq x}\sum_{\substack{n\leq x\\ p|n}}1=\sum_{p\leq x}\left(\frac{x}{p}+O(1)\right)=x\sum_{p\leq x}\frac{1}{p}+O\left(\sum_{p\leq x}1\right)$$
$$=x\left(\log\log x+B+O\left(\frac{1}{\log x}\right)\right)+O\left(\frac{x}{\log x}\right)=x\log\log x+Bx+O\left(\frac{x}{\log x}\right).$$

Now, in the actual proof, our first step is the following expansion:

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

Now, this simplifies to the following

$$\begin{aligned}
&\sum_{n\le x}(\omega(n)-\log\log x-B)^2\\
&=\sum_{n\le x}\omega(n)^2+x(\log\log x)^2+O(x)+2Bx(\log\log x)-2Bx(\log\log x)-O(x)-\sum_{n\le x}2\log\log x\omega(n)\\
&=\sum_{n\le x}\omega(n)^2+x(\log\log x)^2+O(x)-\sum_{n\le x}2\log\log x\omega(n)\\
&=\sum_{n\le x}\omega(n)^2+x(\log\log x)^2+O(x)-2(\log\log x)\left(x\log\log x+Bx+O\left(\frac{x}{\log x}\right)\right)\\
&=\sum_{n\le x}\omega(n)^2-x(\log\log x)^2-2Bx(\log\log x)+O(x)
\end{aligned}$$

Applying the fourth technical lemma to our above work, we see that

$$\sum_{n\le x}(\omega(n)-\log\log x-B)^2=\sum_{n\le x}\omega(n)^2-x(\log\log x)^2-2Bx(\log\log x)+O(x)=x\log\log(x)+O(x),$$

which yields the desired result. □

12.3 A Uniform Lower Bound on $\varphi(n)$

The key is the following bound, which is Corollary 4.1.8 of Keith Conrad's notes *Analytic Number Theory*:

Proposition 12.9. *For $x\ge 2$, $\prod_{p\le x}\left(1-\frac{1}{p}\right)^{-1}=e^\gamma(\log x)+O(1)$.*

Then, to establish the desired uniform lower bound, we follow the guidance of Exercise 4.1.12. Taking the reciprocal of the above estimate yields

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

Now, recall that the difference between $\frac{1}{f(x)+C}$ and $\frac{1}{f(x)}$ is on the order of $\frac{1}{f(x)^2}$, because the derivative of $\frac{1}{x}$ is on the order of $\frac{1}{x^2}$. If we sought precision including constants, we could use the more precise approximation $\frac{1}{f(x)}-\frac{1}{f(x)+C}\asymp\frac{C}{f(x)^2}$ for large $f(x)$, but for our purposes this constant disappears as we will encase this in $O(\cdot)$. In any case, this allows us to conclude that, as desired

$$\frac{1}{e^\gamma\log(x)+O(1)}=\frac{1}{e^\gamma\log(x)}+O\left(\frac{1}{e^\gamma\log(x)}\right)=\frac{e^{-\gamma}}{\log(x)}+O\left(\frac{1}{\log(x)^2}\right).$$

First, recall that $\varphi(n)=n\prod_{p|n}(1-1/p)\ge n\prod_{p\le n}(1-1/p)$ (since obviously the set of primes dividing n is contained in the set of primes at most n). Hence by the above bound, we can get a uniform lower bound:

$$\varphi(n)\ge n\prod_{p\le n}(1-1/p)=\frac{ne^{-\gamma}}{\log n}+O\left(\frac{n}{\log(n)^2}\right)$$

12.4 Assorted Identities for Dirichlet L-Functions

Proposition 12.10. *$\sum_{n\le 1}\chi(n)d(n)/n^s=L(s,\chi)^2$ for $\Re(s)>1$.*

Proof. To show the result for $\Re(s)>1$, it suffices to show that the Euler factors of both sides at any prime p are equal. First, notice that the Euler factor for p of the right-hand side is $\left(1-\frac{\chi(p)}{p^s}\right)^{-2}$. But of course, we can expand this as

$$\left(1+\frac{\chi(p)}{p^s}+\frac{\chi(p)^2}{p^{2s}}+\cdots\right)^2=1+\frac{2\chi(p)}{p^s}+\frac{3\chi(p)^2}{p^{2s}}+\cdots=1+\frac{\chi(p)d(p)}{p^s}+\frac{\chi(p^2)d(p^2)}{p^{2s}}+\cdots$$

which is precisely the Euler factor for p for the left-hand side. Thus the result follows. □

Proposition 12.11. $\sum_{n\leq 1}\chi(n)\sigma_k(n)/n^s = L(s,\chi)L(s-k,\chi)$ *for* $\Re(s) > k+1$.

Proof. To show the result for $\Re(s) > k+1$, it suffices to show that the Euler factors of both sides at any prime p are equal. First, notice that the Euler factor for p of the right-hand side is $\left(1-\frac{\chi(p)}{p^s}\right)^{-1}\left(1-\frac{\chi(p)}{p^{s-k}}\right)^{-1}$. But of course, we can expand this as

$$\left(1+\frac{\chi(p)}{p^s}+\frac{\chi(p)^2}{p^{2s}}+\cdots\right)\left(1+\frac{\chi(p)}{p^{s-k}}+\frac{\chi(p)^2}{p^{2(s-k)}}+\cdots\right) = \left(1+\frac{\chi(p)}{p^s}+\frac{\chi(p)^2}{p^{2s}}+\cdots\right)\left(1+\frac{\chi(p)p^k}{p^s}+\frac{\chi(p)^2p^{2k}}{p^{2s}}+\cdots\right).$$

Of course, by expanding, we can rewrite this as

$$1+\frac{\chi(p)(1+p^k)}{p^s}+\frac{\chi(p)^2(1+p^k+p^{2k})}{p^{2s}}+\cdots = 1+\frac{\chi(p)\sigma_k(p)}{p^s}+\frac{\chi(p^2)\sigma_k(p^2)}{p^{2s}}+\cdots$$

But this is precisely the Euler factor for p for the left-hand side. Thus we are done. □

Proposition 12.12. $\sum_{n\geq 1}\chi(n)\varphi(n)/n^s = L(s-1,\chi)/L(s,\chi)$ *for* $\Re(s) > 1$.

Proof. To show the result for $\Re(s) > 2$, it suffices to show that the Euler factors of both sides at any prime p are equal. First, notice that the Euler for factor p of the right-hand side is

$$\begin{aligned}\left(1-\frac{\chi(p)}{p^s}\right)\left(1-\frac{\chi(p)}{p^{s-1}}\right)^{-1} &= \left(1-\frac{\chi(p)}{p^s}\right)\left(1+\frac{\chi(p)}{p^{s-1}}+\frac{\chi(p)^2}{p^{2(s-1)}}+\cdots\right)\\ &= \left(1-\frac{\chi(p)}{p^s}\right)\left(1+\frac{p\chi(p)}{p^s}+\frac{p^2\chi(p)^2}{p^{2s}}+\cdots\right)\\ &= 1+\frac{p\chi(p)}{p^s}+\frac{p^2\chi(p)^2}{p^{2s}}+\cdots-\frac{\chi(p)}{p^s}-\frac{p\chi(p)^2}{p^{2s}}-\cdots\\ &= 1+\frac{(p-1)\chi(p)}{p^s}+\frac{(p^2-p)\chi(p)^2}{p^{2s}}+\cdots\\ &= 1+\frac{\varphi(p)\chi(p)}{p^s}+\frac{\varphi(p^2)\chi(p^2)}{p^{2s}}+\cdots\end{aligned}$$

which is exactly the Euler factor for p of the left-hand side. Thus the result follows. □

12.5 An Alternative Proof That $L(1,\chi)\neq 0$

Lemma 12.13. $1-x^d = \prod_{\omega^d=1}(\omega x-1)$.

Proof. First, notice that $x^d-1=\prod_{\omega^d=1}(x-\omega)$ since both sides have precisely the same roots (up to multiplicity) and are monic. Then, notice that the product of all dth roots of unity is -1. To see why, notice that each non-real dth root of unity ω cancels with its unique inverse $\overline{\omega}$, which is also a dth root of unity, so

$$\prod_{\omega^d=1}\omega = \prod_{\substack{\omega^d=1\\ \omega\in\mathbb{R}}}\omega = 1\cdot(-1) = -1.$$

But then,

$$1-x^d = -1(x^d-1) = \prod_{\omega^d=1}\omega\prod_{\omega^d=1}(x-\omega) = \prod_{\omega^d=1}\overline{\omega}\prod_{\omega^d=1}(x-\omega) = \prod_{\omega^d=1}\overline{\omega}(x-\omega) = \prod_{\omega^d=1}(\overline{\omega}x-1) = \prod_{\omega^d=1}(\omega x-1).$$

□

Lemma 12.14. *For real* $s>1$, $\prod_\chi L(s,\chi)\geq 1$.

Proof. Recall that, for each Dirichlet character χ mod m,

$$L(s,\chi)=\prod_p\left(1+\frac{\chi(p)}{p^s}+\frac{\chi(p^s)}{p^{2s}}+\cdots\right)=\prod_p\sum_{k=0}^{\infty}\left(\frac{\chi(p)}{p^s}\right)^k=\prod_p\left(1-\frac{\chi(p)}{p^s}\right)^{-1}.$$

Therefore,

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

where the swap is performed using absolute convergence. Therefore, to show that the final product is greater than or equal to 1, it suffices to show that for each p, $\prod_\chi\left(1-\frac{\chi(p)}{p^s}\right)\leq 1$. Now, there are two cases. The first is the easy one: $p\mid m$. In this case, $\chi(p)=0$ for each χ, so the product above is equal to 1, as desired. The second case, then, is $p\nmid m$. In this case, suppose that p has order $d\mid\varphi(m)$ in $(\mathbb{Z}/m\mathbb{Z})^\times$. Then $\chi(p)$ is equal to each dth root of unity $\frac{\varphi(m)}{d}$ times as we scan over all $\varphi(m)$ characters mod m. That is,

$$\prod_\chi\left(1-\frac{\chi(p)}{p^s}\right)=\left(\prod_{\omega^d=1}\left(1-\frac{\omega}{p^s}\right)\right)^{\frac{\varphi(m)}{d}}=\left(1-p^{sd}\right)^{\frac{\varphi(m)}{d}}.$$

In particular, when $s>1$, then $\left(1-p^{sd}\right)<1$ whence $\left(1-p^{sd}\right)^{\frac{\varphi(m)}{d}}<1$, as desired. □

Lemma 12.15. *Suppose that χ is a nontrivial nonquadratic character. Then $L(1,\chi)\neq 0$.*

Proof. Let $n(\chi)$ denote the order of vanishing of $\prod_\chi L(s,\chi)$ at $s=1$. Notice that the order of vanishing of $\prod_\chi L(s,\chi)$ is equal to the sum of the orders of vanishing of each of the χ. Yet since $\prod_\chi L(s,\chi)\geq 1$ whenever $s>1$, and $\prod_\chi L(s,\chi)$ is meromorphic, $\prod_\chi L(1,\chi)$ cannot vanish. Therefore, the sum of the orders of vanishing of each of the χ is at least 0; that is,

$$0\geq\sum_\chi n(\chi)=n(1_m)+\sum_{\chi\neq 1_m}n(\chi)\Rightarrow 1\geq\sum_{\chi\neq 1_m}n(\chi).$$

since $n(1_m)=-1$. Then, notice that since $n(\chi)\geq 0$ for all $\chi\neq 1_m$, it must be the case that $n(\chi)=0$ for all but possibly one character χ' for which $n(\chi')$ is either 0 or 1. Then recall, as discussed in class, that $\overline{L(s,\chi')}=L(\overline{s},\overline{\chi'})$. In particular, if $L(1,\chi')=0$, then $L(1,\overline{\chi'})=0$. Therefore, if $n(\chi')=1$, then $n(\overline{\chi'})=1$. But this is impossible unless $\chi'=\overline{\chi'}$ (as then $\sum_{\chi\neq 1_m}n(\chi)\geq 2$, a contradiction), which implies that χ' is quadratic. Thus when χ is nonquadratic and nontrivial, $L(1,\chi)\neq 0$. □

Lemma 12.16. *Suppose that χ is a nontrivial quadratic Dirichlet character mod m. Then $L(1,\chi)\neq 0$.*

Proof. The key is Landau's Theorem (Corollary 3.4.2 in Conrad). Landau's Theorem states that if a Dirichlet series $F(s)=\sum_{n=1}^{\infty}\frac{f(n)}{n^s}$ converges for $\Re(s)>\sigma_0$, has non-negative coefficients (i.e. $f(n)\geq 0$ for all n), and $F(s)$ has an analytic continuation to a larger half-plane $\Re(s)>\sigma_1$ (i.e., $\sigma_1<\sigma_0$), then the continuation of $F(s)$ equals $\sum_{n=1}^{\infty}\frac{f(n)}{n^s}$ on $\Re(s)>\sigma_1$. In particular, the Dirichlet series $F(s)$ converges on this domain.

Let $F(s)=\zeta(s)L(s,\chi)$. Then if $L(1,\chi)=0$, the pole of $\zeta(s)$ cancels out with the zero of $L(s,\chi)$ at $s=1$, so $F(s)$ has no pole at 1. In particular, $F(s)$ has an analytic continuation to $\Re(s)>0$ (by using the continuation of the ζ-function). Now, taking Euler products for $L(s,\chi)$ and $\zeta(s)$, we get an Euler product for $F(s)$:

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

Taking the logarithm, we have

$$-\log\left(1-\frac{1}{p^s}\right)-\log\left(1-\frac{\chi(p)}{p^s}\right)=\sum_{k\geq 1}\frac{1+\chi(p)^k}{kp^{ks}}$$

which implies that the coefficients of the sum are non-negative, as $\chi(p)$ is either -1, 0, or 1. But then, if we take the exponential of this to achieve a Dirichlet series for F, Problem 3 above implies that the resulting Dirichlet series has all nonnegative coefficients. Then, since this Dirichlet series converges for all $\Re(s) > 1$ and $F(s)$ has an analytic continuation to $\Re(s) > 0$, this Dirichlet series converges for $\Re(s) > 0$.

Now, for simplicity, write $z = \frac{1}{p^s}$ and $c = \chi(p)$. Then, looking at the Euler factor for p,

$$\begin{aligned}\left(1 - \frac{1}{p^s}\right)^{-1}\left(1 - \frac{\chi(p)}{p^s}\right)^{-1} &= \frac{1}{1-z}\frac{1}{1-cz} = (1 + z + z^2 + \cdots)(1 + cz + c^2z^2 + \cdots) \\ &= 1 + (1+c)z + (1+c+c^2)z^2 + \cdots .\end{aligned}$$

Then the coefficient of z^k is the coefficient of $\frac{1}{p^{ks}}$ in $F(s)$. In particular, the coefficient of $\frac{1}{p^{2s}}$ is $1 + c + c^2$. Since $c = \chi(p)$ is either -1, 0, or 1, by inspection it follows that $1 + c + c^2 \geq 1$. Thus for all $s \in \mathbb{R}_{>0}$, $F(s) \geq \sum_p \frac{1}{p^{2s}}$ (by forgetting every term except the term for p^{2s} and using the bound for the coefficient above). But as $s \to \frac{1}{2}$, $\sum_p \frac{1}{p^{2s}}$ goes to infinity. Thus $F(s)$ has a pole at $\frac{1}{2}$, a contradiction with the fact that it is analytic on $\Re(s) > 0$, a contradiction. □