LECTURE NOTES 2 FOR CAMBRIDGE PART III COURSE ON "PROBABILISTIC NUMBER THEORY", MICHAELMAS 2015

ADAM J HARPER

Abstract. These are rough notes covering the second block of lectures in the "Probabilistic Number Theory" course, on probabilistic heuristics. We will develop Cramér's probabilistic model for the primes, and use it to precisely conjecture the distribution of primes in short intervals. Then we will study the Buchstab function and prove Maier's irregularity theorem, which shows that the Cramér model for short intervals cannot be completely correct. We also develop some basic theory of the Riemann zeta function, and discuss Möbius randomness and heuristics for the Riemann Hypothesis.

(No originality is claimed for any of the contents of these notes, which borrow from the books of Montgomery and Vaughan [1] and Titchmarsh [2] as well as from numerous original research papers.)

7. The Borel–Cantelli lemmas

In number theoretic problems, we often try to prove that a certain property is satisfied for all values of some parameter. For example, we might try to prove that a function $f(n)$ obeys a certain estimate for all $n \in \mathbb{N}$. In view of this, when we develop probabilistic heuristics we need to make sure that they really imply (on the probabilistic side) that the properties we are interested in hold for all values of the parameter.

In this section we will prove two basic probabilistic lemmas that are relevant for this.

Lemma 7.1 (First Borel–Cantelli Lemma). *Let $(A_n)_{n\in\mathbb{N}}$ be some sequence of events (that are measurable with respect to a probability measure $\mathbb{P}$), and suppose that $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ converges. Then*

$$\mathbb{P}(\textit{infinitely many of the } A_n \textit{ occur}) =: \mathbb{P}(A_n \textit{ i.o.}) = 0.$$

Proof of Lemma 7.1. Let $N \in \mathbb{N}$ be arbitrary but fixed. If infinitely many of the A_n occur, then certainly at least one A_n with $n \geq N$ must occur, so we see

$$\mathbb{P}(A_n \text{ i.o.}) \leq \mathbb{P}(A_n \text{ holds for some } n \geq N) \leq \sum_{n \geq N} \mathbb{P}(A_n),$$

by the union bound. Since $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ is convergent, the right hand side tends to zero as $N \to 0$ whilst the left hand side is independent of N. Thus we must have $\mathbb{P}(A_n \text{ i.o.}) = 0$. □

Date: 23rd November 2014.

Remark 7.2. Remember that an event that happens with probability zero can still occur. For example, suppose that $\mathbb{P}$ were the measure corresponding to a random variable X unformly distributed on $[0, 1]$, and that $A_n := \{0 \leq X \leq 1/n^2\}$. Then it is obvious that if $X = 0$, all of the events A_n do occur.

The converse of the First Borel–Cantelli Lemma is false (as can be seen by adapting the example in the above remark), but one can obtain a converse by introducing an extra independence assumption.

Lemma 7.3 (Second Borel–Cantelli Lemma). *Let* $(A_n)_{n \in \mathbb{N}}$ *be a sequence of* independent *events (that are measurable with respect to a probability measure* $\mathbb{P}$*), and suppose that* $\sum_{n=1}^{\infty} \mathbb{P}(A_n)$ *diverges. Then*

$$\mathbb{P}(\textit{infinitely many of the } A_n \textit{ occur}) =: \mathbb{P}(A_n \textit{ i.o.}) = 1.$$

Proof of Lemma 7.3. Let $N \leq M$ be arbitrary fixed natural numbers, and observe that

$$\mathbb{P}(A_n \text{ does not occur for all } N \leq n \leq M) = \prod_{N \leq n \leq M} (1 - \mathbb{P}(A_n)),$$

since the A_n are independent. For any number $0 \leq p \leq 1$ we have $0 \leq 1 - p \leq \exp\{-p\}$, and therefore

$$\mathbb{P}(A_n \text{ does not occur for all } N \leq n \leq M) \leq \exp\{-\sum_{N \leq n \leq M} \mathbb{P}(A_n)\}.$$

In particular, if we fix N and let $M \to \infty$ then, since $\sum_n \mathbb{P}(A_n)$ diverges, we must have

$$\mathbb{P}(A_n \text{ does not occur for all } n \geq N) = 0.$$

Finally, using the union bound we have

$$\mathbb{P}(A_n \text{ does not occur for all } n \geq N, \text{ for some } N) \leq \sum_{N=1}^{\infty} \mathbb{P}(A_n \text{ does not occur for all } n \geq N) = 0,$$

and so the complementary event, namely that A_n occurs infinitely often, must have probability 1. □

In the next section we will apply the Borel–Cantelli lemmas in a number-theoretic situation, but it is worthwhile to get an idea of their importance in a pure probabilistic situation also. Let $S_n := \sum_{i=1}^{n} X_i$ be a simple random walk, so that X_i are independent random variables taking values ±1 with probability 1/2 each. Notice that S_n has mean zero and variance n. One can show (Hoeffding's Inequality) that for any $t \geq 0$,

$$\mathbb{P}(|S_n| \geq t\sqrt{n}) \leq 2e^{-t^2/2}.$$

In particular, for any function $t = t(n)$ that tends to infinity with n we will have

$$\mathbb{P}(|S_n| \geq t(n)\sqrt{n}) \to 0 \quad \text{as } n \to \infty,$$

but what does this actually imply about the fluctuations of the random walk S_n as n varies? For any $\delta > 0$ we have

$$\mathbb{P}(|S_n| \geq \sqrt{(2+\delta)n\log n}) \leq 2e^{-(2+\delta)(\log n)/2} = \frac{2}{n^{1+\delta/2}}, \quad \text{and} \quad \sum_{n=1}^{\infty} \frac{1}{n^{1+\delta/2}} < \infty,$$

so the first Borel–Cantelli Lemma implies that with probability 1, the event $|S_n| \geq \sqrt{(2+\delta)n\log n}$ happens only finitely many times. In other words, the fluctuations of S_n are (almost surely) of size *at most* $\sqrt{(2+o(1))n\log n}$, but is this the real truth (remember that the converse of the first Borel–Cantelli lemma is not true)? There seems to be something a bit wasteful about our argument, because we have summed over all n, but if we know that $|S_n| \leq \sqrt{(2+\delta)n\log n}$ for some given n then we automatically know that the same must be approximately true for all nearby n. The truth turns out to be that with probability 1,

$$\limsup \frac{S_n}{\sqrt{2n\log\log n}} = 1,$$

which is Kolmogorov's *Law of the Iterated Logarithm.* One gets this by applying the Borel–Cantelli lemmas at an appropriately chosen subsequence of values n, and it shows the importance of understanding exactly what one is trying to prove "for all n".

8. Cramér's model for primes

In Fact 1 from Chapter 0 (and on the first problem sheet), we saw Chebychev's order estimate $\pi(x) \asymp \frac{x}{\log x}$ for the prime counting function ($x \geq 2$). In a high point of nineteenth century mathematics, Hadamard and de la Vallée Poussin proved (independently) in 1896 that actually

$$\pi(x) \sim \int_2^x \frac{dt}{\log t} \sim \frac{x}{\log x} \quad \text{as } x \to \infty,$$

which is called the *Prime Number Theorem.* This says that a proportion $\sim 1/\log x$ of the numbers around x are prime. The best result known today, due to Vinogradov and Korobov in 1958, is that

$$\pi(x) = \int_2^x \frac{dt}{\log t} + O(xe^{-c(\log^{3/5} x)/(\log\log x)^{1/5}}).$$

In a 1936 paper on the distribution of primes in short intervals, Cramér proposed modelling the primes by a sequence of independent random variables U_n, as follows:

Model 8.1 (Cramér's model). *Let $U_1 = 0$, let $U_2 = 1$, and for $n \geq 3$ let U_n be a sequence of independent Bernoulli random variables taking value 1 with probability $1/\log n$, and taking value 0 otherwise.*

Then for "properly formulated" questions about primes, if a statement is true with probability 1 for the sequence (U_n) then "something similar" may hold for the primes themselves.

Remark 8.2. The deliberately vague terms "properly formulated" and "something similar" in Model 8.1 are a bit disturbing, but it seems difficult to say precisely what they should mean— it is a question of judgement in any given situation (as it would be when deciding if a physical model were reasonable or not). For example, in Cramér's model the values U_{2n} are not distinguished in any way from the values U_{2n+1}, whereas for the actual primes we know the distribution between odd and even numbers is extremely different, so a properly formulated question would have to be one where this distinction is unimportant. The phrase "something similar" means that one should consider inserting some kind of safety factor when translating the probabilistic answers into conjectures, to obtain conjectures that seem more certain.

Remark 8.3. Cramér did not state his model in the general but vague way that we have in Model 8.1. Instead he worked it out concretely in the special case of primes in short intervals.

Theorem 8.4 (Cramér's model for short intervals)**.** *Let U_n be a sequence of independent random variables as in the Cramér model. Define a random sequence $(P_n)_{n\in\mathbb{N}}$ by setting $P_n := \min\{m : \sum_{i=1}^{m} U_i = n\}$.*

Then with probability 1,

$$\limsup_{n\to\infty} \frac{P_{n+1}-P_n}{\log^2 P_n} = 1.$$

Proof of Theorem 8.4. We shall prove the theorem in two parts. Firstly we shall let $\delta > 0$ be small but fixed, and prove that

$$\mathbb{P}(P_{n+1}-P_n > (1+\delta)\log^2 P_n \text{ i.o.}) = 0,$$

which will establish that $\mathbb{P}(\limsup_{n\to\infty} \frac{P_{n+1}-P_n}{\log^2 P_n} \leq 1) = 1$.

Indeed, for any given n, by definition of P_n we have

$$\mathbb{P}(P_{n+1}-P_n > (1+\delta)\log^2 P_n) = \mathbb{P}(U_i = 0 \ \forall P_n + 1 \leq i \leq P_n + (1+\delta)\log^2 P_n),$$

and since the U_i are independent of one another (and in particular independent of the value of P_n) we see

$$\begin{aligned}
\mathbb{P}(P_{n+1} - P_n > (1+\delta)\log^2 P_n) &= \sum_{m=n+1}^{\infty} \mathbb{P}(P_n = m)\mathbb{P}(U_i = 0 \ \forall m+1 \leq i \leq m + (1+\delta)\log^2 m) \\
&= \sum_{m=n+1}^{\infty} \mathbb{P}(P_n = m) \prod_{m+1 \leq i \leq m+(1+\delta)\log^2 m} \left(1 - \frac{1}{\log i}\right) \\
&\leq \sum_{m=n+1}^{\infty} \mathbb{P}(P_n = m) \exp\{- \sum_{m+1 \leq i \leq m+(1+\delta)\log^2 m} \frac{1}{\log i}\}.
\end{aligned}$$

Now if $i \leq m + (1+\delta)\log^2 m$ then we have

$$\frac{1}{\log i} \geq \frac{1}{\log(m + (1+\delta)\log^2 m)} = \frac{1}{\log m + O((\log^2 m)/m)} = \frac{1}{\log m} + O(\frac{1}{m}),$$

and so

$$\begin{aligned}
\mathbb{P}(P_{n+1} - P_n > (1+\delta)\log^2 P_n) &\leq \sum_{m=n+1}^{\infty} \mathbb{P}(P_n = m) \exp\{-\frac{(1+\delta)\log^2 m + O(1)}{\log m} + O\left(\frac{\log^2 m}{m}\right)\} \\
&\ll \sum_{m=n+1}^{\infty} \mathbb{P}(P_n = m) \frac{1}{m^{1+\delta}} \leq \frac{1}{n^{1+\delta}}.
\end{aligned}$$

Since $\sum_n \frac{1}{n^{1+\delta}}$ is a convergent series, our first statement follows from the first Borel–Cantelli lemma (Lemma 7.1).

Now let $\delta > 0$ be small but fixed, and define a deterministic increasing sequence x_n by setting $x_1 = 2$, and $x_{n+1} = x_n + (1-\delta)\log^2(x_n)$, and define events

$$A_n := \{U_i = 0 \ \forall x_n < i < x_{n+1}\}.$$

Notice that if the event A_n occurs, and if P_m is the largest member of the sequence (P_n) that is smaller than x_n, then (since log is an increasing function) we must have $P_{m+1} - P_m \geq (1-\delta)\log^2(P_m)$. We shall prove that $\mathbb{P}(A_n \text{ i.o.}) = 1$, which will establish that $\mathbb{P}(\limsup_{n\to\infty} \frac{P_{n+1}-P_n}{\log^2 P_n} \geq 1) = 1$.

In fact, very similar calculations as before show that

$$\begin{aligned}
\mathbb{P}(A_n) = \prod_{x_n < i < x_{n+1}} \left(1 - \frac{1}{\log i}\right) &= \exp\{\sum_{x_n < i < x_{n+1}} \log(1 - 1/\log i)\} \\
&= \exp\{\sum_{x_n < i < x_{n+1}} \left(-\frac{1}{\log i} + O\left(\frac{1}{\log^2 i}\right)\right)\} \\
&\gg \frac{1}{x_n^{1-\delta}} \gg \frac{1}{(n\log^2(2n))^{1-\delta}}.
\end{aligned}$$

The events A_n depend on the behaviour of disjoint subsets of the independent random variables U_i, so they are independent events. Thus the second Borel–Cantelli lemma (Lemma 7.3) is applicable, and since $\sum_n \frac{1}{(n\log^2(2n))^{1-\delta}}$ is a divergent series the fact that $\mathbb{P}(A_n \text{ i.o.}) = 1$ immediately follows. □

Translating the conclusion of Theorem 8.4 into a conjecture about the primes, we obtain the following.

Conjecture 8.5 (Cramér's Conjecture, 1936). *Let* $2 = p_1 < p_2 < ...$ *denote the sequence of primes taken in increasing order. Then*

- *(strong conjecture)* $\limsup_{n\to\infty} \frac{p_{n+1}-p_n}{\log^2 p_n} = \limsup_{n\to\infty} \frac{p_{n+1}-p_n}{\log^2 n} = 1.$
- *(weak conjecture)* $p_{n+1} - p_n \ll \log^2 n.$

Cramér did not explicitly make either the weak or the strong conjecture, saying only that "we may take [Theorem 8.4] as a suggestion that, for the particular sequence of ordinary prime numbers p_n, some similar relation may hold". However, either statement is traditionally called Cramér's Conjecture.

Using the Vinogradov–Korobov estimate for $\pi(x)$, one can deduce that

$$\pi(x+y) - \pi(x) \sim \frac{y}{\log x} \quad \forall x e^{-c(\log^{3/5} x)/(\log\log x)^{1/5}} \leq y \leq x,$$

for a suitable constant $c > 0$. In particular, we must have $p_{n+1}-p_n \ll ne^{-c(\log^{3/5} n)/(\log\log n)^{1/5}}$. But more sophisticated tools are available for studying primes in intervals than simply taking the difference of estimates for $\pi(x)$, and so it is actually known that $p_{n+1} - p_n \ll n^{0.525+o(1)}$ unconditionally, and $p_{n+1} - p_n \ll n^{1/2+o(1)}$ if the Riemann Hypothesis is true (in fact Cramér is responsible for the sharpest known result like this). But all of these bounds are far weaker than Cramér's Conjecture.

In 1943, Selberg showed that if the Riemann Hypothesis is true then for any function $y(x) \leq x$ satisfying $y(x)/\log^2 x \to \infty$ as $x \to \infty$, we have $\pi(x+y)-\pi(x) = (1+o(1))\frac{y}{\log x}$ for "most" values of x (meaning all apart from a proportion $o(1)$ of x-values). This seems like strong support for Cramér's Conjecture, although we should note that asking for something to happen for most x is really a different problem than asking it for all x, and Cramér's model would actually give a bit different prediction in this different problem.

9. Maier's Irregularity Result

In this section we will show that, in fact, the Cramér model does *not* entirely accurately predict the distribution of primes in short intervals. Maier proved this in 1985, and at the time it was a rather surprising and influential discovery.

Definition 9.1. Let $z \geq 2$. A number n is said to be *z-sieved*, or *z-rough*, if

$$p \mid n \Rightarrow p \geq z.$$

We let $\Phi(x,z) := \#\{n \leq x : n \text{ is } z\text{-sieved}\}$.

As we saw when computing probabilities in Chapter 1, heuristic arguments would suggest that $\Phi(x,z) \approx x\prod_{p<z}\left(1-\frac{1}{p}\right)$. To obtain Maier's irregularity result for primes, a key step will be to show that $\Phi(x,z)$ need *not* always satisfy the heuristic estimate.

We begin by showing that $\Phi(x,z)$ can be well approximated by a certain continuous function.

Definition 9.2. We define the *Buchstab function* $w(u)$ by setting $w(u) = 1/u$ when $1 \leq u \leq 2$, and

$$\frac{d}{du}(uw(u)) = w(u-1) \quad \forall u > 2.$$

Equivalently (Quick Exercise), we have $w(u) = \frac{1}{u}\left(\int_1^{u-1} w(v)dv + 1\right)$ for all $u \geq 2$.

Theorem 9.3 (Buchstab, 1937). *Fix $U \geq 1$. Then for any $z \geq 2$, and any $1 \leq u \leq U$, we have*

$$\Phi(z^u, z) = \frac{z^u w(u)}{\log z} - \frac{z}{\log z} + O_U\left(\frac{z^u}{\log^2 z}\right) = \frac{z^u}{\log z}\left(w(u) - \frac{1}{z^{u-1}} + O_U\left(\frac{1}{\log z}\right)\right).$$

In particular, for any fixed $u > 1$ we have

$$\Phi(z^u, z) \sim \frac{z^u w(u)}{\log z} \quad \textit{as } z \to \infty.$$

To help when proving the theorem, we shall first obtain some basic but more explicit information about the Buchstab function $w(u)$.

Lemma 9.4. *For any $u \geq 1$, the Buchstab function $w(u)$ satisfies $1/2 \leq w(u) \leq 1$. Moreover, for any $u \geq 2$ it satisfies $|w'(u)| \leq 1/(2u)$.*

Proof of Lemma 9.4. When $1 \leq u \leq 2$ the function $w(u) = 1/u$ trivially satisfies $1/2 \leq w(u) \leq 1$, and if we know the estimate for all $u \leq U$, for some $U \geq 2$, then we can obtain it inductively for all $u \leq U+1$ using the expression $w(u) = \frac{1}{u}\left(\int_1^{u-1} w(v)dv + 1\right)$.

Moreover, the chain rule implies that $w(u-1) = \frac{d}{du}(uw(u)) = w(u) + uw'(u)$ when $u > 2$, and so $|w'(u)| = |\frac{w(u-1)-w(u)}{u}| \leq 1/(2u)$. □

Proof of Theorem 9.3. Slightly abusing the notation in the statement of the theorem, we shall prove by induction that for any $U \in \mathbb{N}$, the statement holds for all $U \leq u < U+1$ and all $z \geq 2$.

Firstly, when $U = 1$ and $1 \leq u < 2$ we have

$$\Phi(z^u, z) = 1 + \pi(z^u) - \pi(z),$$

since a number $n \leq z^u$ has all its prime factors $\geq z$ only if it is a prime $\geq z$, or it is 1 (which has no prime factors). By a version of the Prime Number Theorem, it follows that

$$\Phi(z^u, z) = 1 + \frac{z^u}{\log(z^u)} + O\left(\frac{z^u}{(\log(z^u))^2}\right) - \frac{z}{\log z} + O\left(\frac{z}{(\log z)^2}\right),$$

which is acceptable for the theorem (since $w(u) = 1/u$ when $1 \leq u < 2$).

For the inductive step, going from U to $U+1$, we shall make use of the following *Buchstab identity*:

$$\Phi(x, z) = 1 + \sum_{\substack{z \leq p \leq x, \\ p \text{ prime}}} \#\{n \leq x : p \text{ is the smallest prime factor of } n\} = 1 + \sum_{z \leq p \leq x} \Phi(x/p, p).$$

If $U + 1 \leq u < U + 2$ then the Buchstab identity implies that

$$\Phi(z^u, z) - \Phi(z^u, z^{u/(U+1)}) = \sum_{z \leq p < z^{u/(U+1)}} \Phi(z^u/p, p).$$

However, if $z \leq p$ then we have $z^u/p \leq z^{u-1} \leq p^{u-1} < p^{U+1}$, so the inductive hypothesis applies and gives that

$$\Phi(z^u/p, p) = \frac{z^u w(u_p)}{p \log p} - \frac{p}{\log p} + O\left(\frac{z^u}{p \log^2 p}\right), \quad \text{where } u_p := \frac{\log(z^u/p)}{\log p} = \frac{\log(z^u)}{\log p} - 1.$$

Using the estimates of Chebychev and Mertens (Facts 1 and 2 from Chapter 0), and the fact that $U + 1 \geq 2$, we have

$$\begin{aligned} \sum_{z \leq p < z^{u/(U+1)}} \left(-\frac{p}{\log p} + O\left(\frac{z^u}{p \log^2 p}\right)\right) &\ll \frac{z^{u/(U+1)}}{\log z} \pi(z^{u/(U+1)}) + \frac{z^u}{\log^2 z} \sum_{z \leq p < z^{u/(U+1)}} \frac{1}{p} \\ &\ll \frac{z^{2u/(U+1)}}{\log^2 z} + \frac{z^u}{\log^2 z} \ll \frac{z^u}{\log^2 z}, \end{aligned}$$

which is acceptable as an error term in the theorem.

We can also write

$$\sum_{z \leq p < z^{u/(U+1)}} \frac{w(u_p)}{p \log p} = \int_z^{z^{u/(U+1)}} \frac{w(\frac{\log(z^u)}{\log t} - 1)}{t \log t} d\pi(t) = \int_z^{z^{u/(U+1)}} \frac{w(\frac{\log(z^u)}{\log t} - 1)}{t \log t} d\left(\int_2^t \frac{ds}{\log s} + R(t)\right),$$

and a form of the Prime Number Theorem implies that $R(t) := \pi(t) - \int_2^t \frac{ds}{\log s} = O(t/\log^2 t)$. Therefore, using integration by parts and the facts that $|w(\cdot)| \leq 1$ and $\frac{d}{dt} w(\frac{\log(z^u)}{\log t} - 1) \ll \frac{\log(z^u)}{t \log^2 t} |w'(\frac{\log(z^u)}{\log t} - 1)| \ll \frac{1}{t \log t}$ (by Lemma 9.4), we see

$$\begin{aligned} \int_z^{z^{u/(U+1)}} \frac{w(\frac{\log(z^u)}{\log t} - 1)}{t \log t} dR(t) &= \int_z^{z^{u/(U+1)}} \frac{w(\frac{\log(z^u)}{\log t} - 1)}{t \log t} \frac{dR(t)}{dt} dt \\ \ll \frac{1}{\log^3 z} &+ \left| \int_z^{z^{u/(U+1)}} \frac{d}{dt} \left(\frac{w(\frac{\log(z^u)}{\log t} - 1)}{t \log t} \right) R(t) dt \right| \\ \ll \frac{1}{\log^3 z} &+ \int_z^{z^{u/(U+1)}} \frac{1}{t^2 \log t} |R(t)| dt \ll \frac{1}{\log^3 z} + \int_z^{z^{u/(U+1)}} \frac{1}{t \log^3 t} dt \ll \frac{1}{\log^2 z}, \end{aligned}$$

which gives an acceptable overall contribution $\ll \frac{z^u}{\log^2 z}$.

To summarise, we have shown thus far that

$$\Phi(z^u, z) = \Phi(z^u, z^{u/(U+1)}) + z^u \int_z^{z^{u/(U+1)}} \frac{w(\frac{\log(z^u)}{\log t} - 1)}{t \log t} d\left(\int_2^t \frac{ds}{\log s}\right) + O\left(\frac{z^u}{\log^2 z}\right).$$

Finally, we see the integral here is

$$\int_z^{z^{u/(U+1)}} \frac{w(\frac{\log(z^u)}{\log t} - 1)}{t \log t} \frac{d}{dt}\left(\int_2^t \frac{ds}{\log s}\right) dt = \int_z^{z^{u/(U+1)}} \frac{w(\frac{\log(z^u)}{\log t} - 1)}{t \log^2 t} dt = \frac{1}{\log(z^u)} \int_U^{u-1} w(v) dv,$$

the final equality following from the substitution $v = \frac{\log(z^u)}{\log t} - 1$. From the definition of the Buchstab function, we have $\frac{1}{\log(z^u)} \int_U^{u-1} w(v)dv = \frac{uw(u) - (U+1)w(U+1)}{\log(z^u)}$, and so

$$\begin{aligned}\Phi(z^u, z) &= \Phi(z^u, z^{u/(U+1)}) + z^u \frac{uw(u) - (U+1)w(U+1)}{\log(z^u)} + O\left(\frac{z^u}{\log^2 z}\right) \\ &= \frac{z^u w(u)}{\log z} + \Phi(z^u, z^{u/(U+1)}) - z^u \frac{(U+1)w(U+1)}{\log(z^u)} + O\left(\frac{z^u}{\log^2 z}\right).\end{aligned}$$

Moreover we can apply the inductive hypothesis to $\Phi(z^u, z^{u/(U+1)})$, obtaining that $\Phi(z^u, z^{u/(U+1)}) = \frac{z^u w(U+1)}{\log(z^{u/(U+1)})} - \frac{z^{u/(U+1)}}{\log(z^{u/(U+1)})} + O\left(\frac{z^u}{\log^2 z}\right) = \frac{z^u (U+1) w(U+1)}{\log(z^u)} + O\left(\frac{z^u}{\log^2 z}\right)$. Inserting this in the previous display, and noting that $-\frac{z}{\log z} = O(\frac{z^u}{\log^2 z})$ when $u \geq U+1 \geq 2$, gives the estimate for $\Phi(z^u, z)$ claimed in the theorem. □

Remark 9.5. Note that once one has the idea to attempt an inductive proof using Buchstab's identity, the above argument would allow one to guess the recursive definition $w(u) = \frac{1}{u}\left(\int_1^{u-1} w(v)dv + 1\right)$ of the Buchstab function if it were not given, since we must define it to have the proper cancellation with $\Phi(z^u, z^{u/(U+1)})$ at the final step.

Now that we are equipped with Theorem 9.3, we revisit the heuristic $\Phi(x, y) \approx x \prod_{p<y} \left(1 - \frac{1}{p}\right)$. In view of Fact 3 from Chapter 0, we have

$$\prod_{p<y} \left(1 - \frac{1}{p}\right) = \frac{e^{-c_2}}{\log y}\left(1 + O\left(\frac{1}{\log y}\right)\right), \quad y \geq 2,$$

where c_2 is a constant. (In fact c_2 is equal to Euler's constant $\gamma := \lim_{N\to\infty} \left(\sum_{n\leq N} \frac{1}{n} - \log N\right) \approx 0.577$). Comparing with the theorem, we see that for any fixed $u > 1$ and for large z the heuristic for $\Phi(z^u, z)$ will be quite accurate if $w(u)$ is equal to $e^{-\gamma}$, and will be off by a constant multiplicative factor if $\omega(u)$ is not equal (or very close) to $e^{-\gamma}$. *So to prove an irregularity result for $\Phi(z^u, z)$ (i.e. to prove the heuristic is not always very accurate), we just need to prove that $w(u)$ is not always very close to $e^{-\gamma}$.*

Lemma 9.6. *The limit $w := \lim_{u\to\infty} w(u)$ of the Buchstab function exists, and we have*

$$|w(u) - w| \ll \frac{1}{\Gamma(u+1)} \quad \forall u \geq 1,$$

where $\Gamma(t) := \int_0^\infty e^{-x} x^{t-1} dx = e^{t\log t + O(t)}$ *for all* $t \geq 1$ *(say).*

Proof of Lemma 9.6. Recall that when $u > 2$ we have $w(u-1) = \frac{d}{du}(uw(u)) = w(u) + uw'(u)$, which we can rewrite as $w'(u) = \frac{w(u-1)-w(u)}{u}$. On the other hand, we have $w(u-1) - w(u) = -\int_{u-1}^{u} w'(t)dt$, and so we have

$$|w'(u)| \leq \frac{1}{u} \max_{u-1\leq t\leq u} |w'(t)| \quad \forall u > 2.$$

This seems to suggest that $|w'(u)|$ tends to zero rapidly with u. Indeed, if we let $M(u) := \max_{t\geq u} |w'(t)|$ then inductively we have

$$M(u) \leq \frac{M(u-1)}{u} \leq \frac{M(u-2)}{u(u-1)} \leq ... \ll \frac{1}{\Gamma(u+1)}.$$

(Here we apply the fact that $\Gamma(u+1) = u\Gamma(u)$.) It follows that $\lim_{u\rightarrow\infty} w(u) = w(2) + \int_2^\infty w'(t)dt$ certainly exists, and

$$|w(u) - w| \leq \int_u^\infty |w'(t)|dt \ll \int_u^\infty \frac{dt}{\Gamma(t+1)} \ll \frac{1}{\Gamma(u+1)} \quad \forall u \geq 2.$$

(The corresponding statement when $1 \leq u < 2$ is trivial, since both sides are of order 1.) □

If the limit w from Lemma 9.6 were not equal to $e^{-\gamma}$, we could now immediately conclude an irregularity result for z-sieved numbers, but unfortunately for us here (although unsurprisingly, and fortunately for much of the rest of number theory!) it turns out that $w = e^{-\gamma}$. Nevertheless, it turns out that we can show that $w(u)$ oscillates infinitely often around its limit w (though with the amplitude of the oscillations decreasing rapidly with u, as shown by Lemma 9.6), and therefore we have an irregularity result for z-sieved numbers, both “from above” and “from below”, for arbitrarily large fixed values of u.

Lemma 9.7 (Buchstab oscillations)**.** *There exists a sequence of arbitrarily large values of* $u \in \mathbb{R}$ *for which the Buchstab function* $w(u) < w$*, and there exists a sequence of arbitrarily large values for which* $w(u) > w$*.*

Proof of Lemma 9.7. We will only prove the first statement, since the proof of the second is exactly similar.

Suppose, for a contradiction, that for all sufficiently large u we had $w(u) \geq w$. Then we could define $u_0 := \sup\{u \geq 1 : w(u) < w\}$ (with the convention, should it be needed, that the supremum of the empty set is 1), so that $w(u) - w \geq 0$ for all $u \geq u_0$. However, recall that we have $uw(u) = \int_1^{u-1} w(v)dv + 1$ for all $u \geq 2$, and therefore we have

$$u(w(u) - w) = \int_1^{u-1} (w(v) - w)dv + (1 - 2w) \quad \forall u \geq 2.$$

In particular, if $u_0 + 1 \leq u_1 \leq u_2$ then we have

$$u_2(w(u_2) - w) - u_1(w(u_1) - w) = \int_{u_1-1}^{u_2-1} (w(v) - w)dv \geq 0,$$

so $u(w(u) - w) \geq 0$ would be a non-decreasing function on the interval $[u_0 + 1, \infty)$.

But Lemma 9.6 implies that $u(w(u) - w) \to 0$ as $u \to \infty$, so this situation could only be possible if $w(u) = w$ for all $u \geq u_0 + 1$. In this case we would have $w = \frac{d}{du}(uw(u)) = w(u-1)$ for all $u \geq u_0 + 1$, and so $w(u) = w$ for all $u \geq u_0$, and so inductively we see we would actually have $w(u) = w$ for *all* $u \geq 1$. However this is false, since $w(u) = 1/u$ is non-constant on the interval $[1, 2]$, for example. □

Corollary 9.8 (Irregularity Theorem for sieved numbers)**.** *There exists a sequence of arbitrarily large values of $u \in \mathbb{R}$, and of small real numbers $\kappa(u) > 0$, for which*

$$\Phi(z^u, z) \geq (1 + \kappa(u))z^u \prod_{p<z} \left(1 - \frac{1}{p}\right) \quad \forall \textit{ large } z.$$

There also exists a sequence of arbitrarily large values of $u \in \mathbb{R}$, and of small real numbers $\kappa(u) > 0$, for which

$$\Phi(z^u, z) \leq (1 - \kappa(u))z^u \prod_{p<z} \left(1 - \frac{1}{p}\right) \quad \forall \textit{ large } z.$$

Proof of Corollary 9.8. Combining our asymptotic for sieved numbers (Theorem 9.3) with the Buchstab oscillations lemma, we find there exists a sequence of arbitrarily large u for which

$$\Phi(z^u, z) \geq \frac{z^u}{\log z}(1 + \kappa(u))w \quad \forall \text{ large } z,$$

and the same for the $\leq (1 - \kappa(u))$ statement. Assuming the Facts that $\prod_{p<z}\left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log z}$ as $z \to \infty$, and that the limit w of the Buchstab function is $e^{-\gamma}$ (whose proof is omitted), the Corollary follows.

((Note that if w were not equal to $e^{-\gamma}$ we would obtain an irregularity theorem for *all* large u, rather than just a special sequence, but only “from above” or “from below” depending on whether w were larger or smaller than $e^{-\gamma}$.)) □

Remark 9.9. The ultimate source of the oscillations in Lemma 9.7 is the fact that the Buchstab function is non-constant when $1 \leq u \leq 2$, in other words the fact that the heuristic $\Phi(x, z) \approx x \prod_{p<z}\left(1 - \frac{1}{p}\right)$ typically fails when $\sqrt{x} \leq z \leq x$. *This is just a restatement of the fact that divisibility by different primes larger than $\sqrt{x}$ are obviously not “almost independent”, as we explored in Chapter 1.*

Now we are ready to upgrade Corollary 9.8 into the promised irregularity result for the primes.

Theorem 9.10 (Maier's Irregularity Theorem, 1985)**.** *Fix* $\lambda > 1$. *Then*

$$\limsup_{x\to\infty} \frac{\pi(x+\log^{\lambda} x) - \pi(x)}{\log^{\lambda-1} x} > 1, \quad \textit{and} \quad \liminf_{x\to\infty} \frac{\pi(x+\log^{\lambda} x) - \pi(x)}{\log^{\lambda-1} x} < 1.$$

Since the density of primes around x is $\sim 1/\log x$, the expected number of primes between x and $x + \log^{\lambda} x$ is $\log^{\lambda-1} x$. The Cramér model would have predicted that, once $\lambda > 2$, both the limsup and the liminf here should equal 1.

The proof of Theorem 9.10 is an averaging argument that is often called the *Maier matrix method.* The overall idea is simply that if we know that the z-sieved numbers on a certain range are irregularly distributed, and if we know that the quantity of primes in a set depends on the number of z-sieved numbers in a predictable way (which is true in various situations), then the number of primes in the set must also have some irregularities. The Maier matrix argument comes in because the set we deal with is a union of arithmetic progressions, assembled as the columns of a matrix.

Proof of Theorem 9.10. We will only prove that $\limsup_{x\to\infty} \frac{\pi(x+\log^{\lambda} x)-\pi(x)}{\log^{\lambda-1} x} > 1$, since the proof of the liminf statement is exactly similar. In view of Corollary 9.8, we may fix a value $u > \lambda$ for which

$$\Phi(z^u, z) \geq (1+\kappa(u)) z^u \prod_{p<z}\left(1 - \frac{1}{p}\right) \quad \forall \text{ large } z.$$

We will also use the deep Fact that if $\pi(x; q, a) := \#\{p \leq x : p \text{ prime}, \ p \equiv a \bmod q\}$, then there exists a sequence of arbitrarily large values of z for which

$$\pi(x; \prod_{p<z} p, a) = \frac{1}{\prod_{p<z} p\left(1-\frac{1}{p}\right)} \left(\int_2^x \frac{dt}{\log t}\right)\left(1 + O(e^{-c\min\{z, (\log x)/z\}})\right) \ \ \forall x \geq \left(\prod_{p<z} p\right)^3, \ (a, \prod_{p<z} p) = 1$$

(This Fact says that the primes are roughly equidistributed among all the $\prod_{p<z} p\left(1-\frac{1}{p}\right)$ coprime arithmetic progressions modulo $\prod_{p<z} p$.)

Now let D be a large natural number, whose value we will set later in terms of u, and let z be one of the numbers for which we have the above estimate for $\pi(x; \prod_{p<z} p, a)$. We shall consider the matrix $\mathcal{M} = (m_{r,s})$, where

$$m_{r,s} := s + r\prod_{p<z} p, \qquad 1 \leq s \leq z^u, \quad \left(\prod_{p<z} p\right)^{D-1} < r \leq 2\left(\prod_{p<z} p\right)^{D-1}.$$

In a column where $(s, \prod_{p<z} p) > 1$, since all of the entries $m_{r,s}$ are strictly larger than $\prod_{p<z} p$ we see none of them can be prime. Meanwhile, if $(s, \prod_{p<z} p) = 1$ then in the s-column of $\mathcal{M}$ there will be $\pi(s+2\left(\prod_{p<z} p\right)^D; \prod_{p<z} p, s) - \pi(s+\left(\prod_{p<z} p\right)^D; \prod_{p<z} p, s)$ primes. Using the Fact about $\pi(x; \prod_{p<z} p, a)$, and using that $\log\left(\prod_{p<z} p\right)^D = D\sum_{p<z}\log p \asymp$

Dz (which follows from Chebychev's estimate, Fact 1 from Chapter 0), this quantity is

$$\begin{aligned} & \frac{1}{\prod_{p<z} p\left(1-\frac{1}{p}\right)} \left(\int_{s+\left(\prod_{p<z} p\right)^D}^{s+2\left(\prod_{p<z} p\right)^D} \frac{dt}{\log t} \right) \left(1+O(e^{-c\min\{z,D\}})\right) \\ = & \frac{1}{\prod_{p<z} p\left(1-\frac{1}{p}\right)} \frac{\left(\prod_{p<z} p\right)^D}{\log((\prod_{p<z} p)^D)} \left(1+O(e^{-cD})+o(1)\right), \end{aligned}$$

where the $o(1)$ term tends to zero as $z \to \infty$. So the total number of prime entries in the matrix $\mathcal{M}$ is

$$\begin{aligned} & \#\{s \le z^u : (s, \prod_{p<z} p) = 1\} \cdot \frac{1}{\prod_{p<z} p\left(1-\frac{1}{p}\right)} \frac{\left(\prod_{p<z} p\right)^D}{\log((\prod_{p<z} p)^D)} \left(1+O(e^{-cD})+o(1)\right) \\ = & \Phi(z^u, z) \frac{1}{\prod_{p<z} p\left(1-\frac{1}{p}\right)} \frac{\left(\prod_{p<z} p\right)^D}{\log((\prod_{p<z} p)^D)} \left(1+O(e^{-cD})+o(1)\right) \\ \ge & (1+\kappa(u)) z^u \frac{(\prod_{p<z} p)^{D-1}}{\log((\prod_{p<z} p)^D)} (1+O(e^{-cD})+o(1)). \end{aligned}$$

Now averaging over the *rows* of $\mathcal{M}$, of which there are $\left(\prod_{p<z} p\right)^{D-1}$, we find there must exist some row r containing at least

$$\frac{(1+\kappa(u))z^u}{\log((\prod_{p<z} p)^D)} (1+O(e^{-cD})+o(1))$$

primes. But the r-th row of $\mathcal{M}$ is simply the interval of length z^u between $1 + r\prod_{p<z} p$ and $z^u + r\prod_{p<z} p$, where $r \asymp \left(\prod_{p<z} p\right)^{D-1}$. In particular, if D and z are large enough there will be at least

$$\frac{(1+\kappa(u)/2)z^u}{\log(1+r\prod_{p<z} p)}$$

primes in the interval. As noted previously, we have $\log(1+r\prod_{p<z} p) \asymp \log((\prod_{p<z} p)^D) \asymp Dz$, so $z^u \gg \log^u(1+r\prod_{p<z} p) \ge \log^\lambda(1+r\prod_{p<z} p)$. Thus there must exist a subinterval of length $\log^\lambda(1+r\prod_{p<z} p)$ containing at least a factor $(1+\kappa(u)/2)$ more than the expected number of primes. □

Remark 9.11. The reason that the irregularities in Maier's theorem are on a scale of powers of $\log x$ is because the size x of the entries in our arithmetic progressions (at least $\left(\prod_{p<z} p\right)^D$) is exponential in the scale z at which we obtain irregularities for the z-sieved numbers.

Remark 9.12. Note that the sequence of values x along which we show that $\frac{\pi(x+\log^{\lambda} x)-\pi(x)}{\log^{\lambda-1} x}$ is large, is very sparse. This is why Maier's Theorem does not contradict the result of Selberg giving an asymptotic for primes in "most" short intervals of length $\log^{2+o(1)} x$.

Remark 9.13. The deep Fact about $\pi(x;q,a)$ that we used should really be true for all $q \leq x^{1/3}$ (say) and all $(a,q)=1$, and with a much stronger error term. This would follow from the Generalised Riemann Hypothesis. But since we only need it for a special sequence of moduli q, it can be proved unconditionally (using methods originating with Linnik).

10. An introduction to the Riemann zeta function

To finish Chapter 2 we shall think about heuristics for the behaviour of the Riemann zeta function, and more specifically a heuristic in favour of the Riemann Hypothesis. Before we can do this we need to develop a little of the theory of the zeta function (which we shall also use in Chapter 3).

The Riemann zeta function $\zeta(s)$ is a meromorphic function on the entire complex plane, but its definition is not straightforward to explain for all $s \in \mathbb{C}$. We will begin by defining the zeta function when $\Re(s) > 1$. Later we will extend the definition to cover the range $\Re(s) > 0$, which is by far the most important for applications. (Note that, since the half plane $\{\Re(s) > 1\}$ is a set containing a limit point, the Identity Theorem from complex analysis implies there is at most one *analytic continuation* of $\zeta(s)$ to a meromorphic function on $\mathbb{C}$.)

Definition 10.1. For each $s \in \mathbb{C}$ such that $\Re(s) > 1$, the Riemann zeta function $\zeta(s)$ is defined by

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

Note that the series is absolutely convergent.

For each $\Re(s) > 1$, the zeta function is built from some information about every natural number n. Since each n is a (possibly empty) product of primes, in an essentially unique way, one might hope that the values of the zeta function can be related to the behaviour of p^s for primes p only. The following result, the so-called *Euler product* expression for $\zeta(s)$, provides such a connection.

Lemma 10.2 (Euler product expression). *If* $\Re(s) > 1$ *then*

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

where the infinite product is defined to be $\lim_{P\to\infty} \prod_{p \leq P} \left(1 - \frac{1}{p^s}\right)^{-1}$.

Proof of Lemma 10.2. Note that, for any prime p,

$$\left(1 - \frac{1}{p^s}\right)^{-1} = \sum_{k=0}^{\infty} \frac{1}{p^{ks}}.$$

Here the geometric series is absolutely convergent if $\Re(s) > 0$.

Since we can freely multiply out and rearrange the terms in a finite product of absolutely convergent series, and since every integer has a unique prime factorisation up to ordering (the *Fundamental Theorem of Arithmetic*), we have

$$\prod_{p \le P} \left(1 - \frac{1}{p^s}\right)^{-1} = \prod_{p \le P} \sum_{k=0}^{\infty} \frac{1}{p^{ks}} = \sum_{n=1}^{\infty} \frac{c_P(n)}{n^s},$$

where $c_P(n)$ is 1 if all the prime factors of n are $\le P$, and $c_P(n)$ is zero otherwise.

Then we see

$$\left| \zeta(s) - \prod_{p \le P} \left(1 - \frac{1}{p^s}\right)^{-1} \right| = \left| \sum_{\substack{n=1, \\ c_P(n)=0}}^{\infty} \frac{1}{n^s} \right| \le \sum_{\substack{n=1, \\ c_P(n)=0}}^{\infty} \frac{1}{n^{\Re(s)}}.$$

Since we certainly have $c_P(n) = 1$ if $n \le P$, the right hand side is

$$\le \sum_{n=P+1}^{\infty} \frac{1}{n^{\Re(s)}}.$$

If $\Re(s) > 1$ then this tends to zero as $P \to \infty$, as claimed. □

The importance of the zeta function arises from the fact that (at least when $\Re(s) > 1$) it is simultaneously a product over primes, which are the basic object of study in multiplicative number theory, and a sum over integers that can be approximated and manipulated analytically. *One wants to establish similar properties for other $s \in \mathbb{C}$, and to play the properties off against one another to deduce information about the zeta function and the primes.*

By playing with the series definition of $\zeta(s)$, we can obtain a second definition that makes sense whenever $\Re(s) > 0$, except at $s = 1$ (where the function has a simple pole), and agrees with our original definition 10.1 when $\Re(s) > 1$.

Definition 10.3. For each $s \in \mathbb{C}$ such that $\Re(s) > 0$, except for $s = 1$, and for any $x > 0$, the Riemann zeta function is defined by

$$\zeta(s) := \sum_{n \le x} \frac{1}{n^s} + \frac{x^{1-s}}{s-1} + \frac{\{x\}}{x^s} - s \int_x^{\infty} \{w\} \frac{dw}{w^{s+1}},$$

Here $\{w\} := w - \lfloor w \rfloor$ denotes the fractional part of w.

The value of the right hand side is independent of the choice of x.

Proof of well definedness. The idea is simply to approximate the tail of the series $\sum_n \frac{1}{n^s}$ by an integral.

Indeed, if $\Re(s) > 1$ and $x > 0$ then

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} = \sum_{n \leq x} \frac{1}{n^s} + \sum_{n > x} s \int_n^{\infty} \frac{dw}{w^{s+1}} = \sum_{n \leq x} \frac{1}{n^s} + s \int_x^{\infty} \Big(\sum_{x < n \leq w} 1 \Big) \frac{dw}{w^{s+1}}.$$

Now $\sum_{x < n \leq w} 1 = \lfloor w \rfloor - \lfloor x \rfloor = (w - x) - \{w\} + \{x\}$, so we can rewrite the above as

$$\zeta(s) = \sum_{n \leq x} \frac{1}{n^s} + s \int_x^{\infty} (w - x) \frac{dw}{w^{s+1}} + s\{x\} \int_x^{\infty} \frac{dw}{w^{s+1}} - s \int_x^{\infty} \{w\} \frac{dw}{w^{s+1}}.$$

An easy calculation shows that the second and third terms here are $\frac{x^{1-s}}{s-1}$ and $\frac{\{x\}}{x^s}$ respectively, so we have verified that definitions 10.3 and 10.1 agree when $\Re(s) > 1$.

Moreover, since we always have $0 \leq \{\cdot\} < 1$, every term on the right hand side in Definition 10.3 continues to converge and define a holomorphic function for all $\Re(s) > 0$ (for any choice of x), apart from the term $\frac{x^{1-s}}{s-1}$ which has a simple pole at $s = 1$. Therefore, by the uniqueness of analytic continuation, since the right hand side takes the same value for any x whenever $\Re(s) > 1$, it must take the same value for any x on the whole range $\Re(s) > 0$. □

11. Perron's inversion formula

In classical analytic number theory, one considers generating series of the form $\sum_{n=1}^{\infty} \frac{a_n}{n^s}$ (called *Dirichlet series*, of which the Riemann zeta function is the prototypical example) and tries to study them and recover information about the coefficients a_n, or related objects. One can think of the Dirichlet series $\sum_{n=1}^{\infty} \frac{a_n}{n^{\sigma+it}}$ as a kind of Fourier transform of the sequence (a_n), in which the oscillating terms $n^{-it} = e^{-it \log n}$ are analogous to the exponential phases $e^{2\pi i n t}$ in a Fourier series, and the terms $n^{-\sigma}$ are present to make everything converge. With this in mind, one could reasonably hope to formulate a procedure like Fourier inversion, allowing one to recover information about $\sum_{n \leq x} a_n$ by summing/integrating the series $\sum_{n=1}^{\infty} \frac{a_n}{n^{\sigma+it}}$ over a suitable range of the "frequency" variable t. This is the procedure that we shall now develop.

Lemma 11.1. *Let $y, c, T > 0$, and define*

$$\delta(y) := \begin{cases} 0 & \text{if } 0 < y < 1 \\ 1/2 & \text{if } y = 1 \\ 1 & \text{if } y > 1. \end{cases}$$

Then

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

Proof of Lemma 11.1. The obvious approach is to use Cauchy's Residue Theorem, and evaluate the integral by deforming the line of integration in a suitable way.

For example, if $0 < y < 1$ then the integrand $\frac{y^s}{s}$ tends to zero as $\Re(s) \to \infty$ (in a uniform way, independently of $\Im(s)$), and is holomorphic on the line of integration and to the right of it (the only pole being to the left, at $s = 0$). Thus Cauchy's Residue Theorem implies that

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT} y^s \frac{ds}{s} = -\frac{1}{2\pi i}\int_{c+iT}^{\infty+iT} y^s \frac{ds}{s} + \frac{1}{2\pi i}\int_{c-iT}^{\infty-iT} y^s \frac{ds}{s},$$

and we certainly have $\left|\int_{c+iT}^{\infty+iT} y^s \frac{ds}{s}\right| \leq \frac{1}{T}\int_c^\infty y^\sigma d\sigma = \frac{y^c}{T|\log y|}$. On the other hand, Cauchy's Residue Theorem also implies that

$$\frac{1}{2\pi i}\int_{c-iT}^{c+iT} y^s \frac{ds}{s} = -\frac{1}{2\pi i}\int_{\Gamma(c,T)} y^s \frac{ds}{s},$$

where $\Gamma(c,T)$ is the arc of the circle centred at the origin, with radius $|c+iT| = \sqrt{c^2+T^2}$, that runs from $c+iT$ to $c-iT$ on the right. And we have $\left|\int_{\Gamma(c,T)} y^s \frac{ds}{s}\right| \leq \frac{y^c}{\sqrt{c^2+T^2}}\int_{\Gamma(c,T)} |ds| \leq \pi y^c$.

Similarly, if $y > 1$ then the integrand $\frac{y^s}{s}$ tends to zero as $\Re(s) \to -\infty$, so one can apply Cauchy's Residue Theorem with the contour shifted to the left instead of the right. This time the contour encloses the pole at $s = 0$, which contributes its residue of 1 to the value of the integral.

Finally, if $y = 1$ then the integral is quite easy to estimate directly (by real variable methods). □

Note in particular that we have $\delta(y) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} y^s \frac{ds}{s}$, (where the infinite integral is interpreted as $\lim_{T\to\infty}\int_{c-iT}^{c+iT} y^s \frac{ds}{s}$), and also that for any $x > 0$ and any $n \in \mathbb{N}$ we have

$$\left|\delta(x/n) - \frac{1}{2\pi i}\int_{c-iT}^{c+iT} \frac{x^s}{n^s}\frac{ds}{s}\right| < \begin{cases} \frac{x^c}{n^c}\min\{1, \frac{1}{T|\log(x/n)|}\} & \text{if } n \neq x \\ \min\{1, \frac{c}{T}\} & \text{if } n = x. \end{cases}$$

Lemma 11.2 (Truncated Perron formula)**.** *Let $x, c, T > 0$, and suppose that $\sum_{n=1}^\infty \frac{|a_n|}{n^c}$ is convergent. Then*

$$\sideset{}{'}\sum_{n\leq x} a_n = \frac{1}{2\pi i}\int_{c-iT}^{c+iT}\left(\sum_{n=1}^\infty \frac{a_n}{n^s}\right) x^s \frac{ds}{s} + O\left(x^c \sum_{n=1}^\infty \frac{|a_n|}{n^c}\min\{1, \frac{1}{T|\log(x/n)|}\}\right),$$

where $\sum'_{n\leq x}$ denotes that if x is an integer, then the final summand a_x is replaced by $(1/2)a_x$.

Proof of Lemma 11.2. We have

$$\sideset{}{'}\sum_{n\leq x} a_n = \sum_{n=1}^\infty a_n \delta(x/n) = \sum_{n=1}^\infty a_n \frac{1}{2\pi i}\int_{c-iT}^{c+iT} \frac{x^s}{n^s}\frac{ds}{s} + O\left(x^c \sum_{n=1}^\infty \frac{|a_n|}{n^c}\min\{1, \frac{1}{T|\log(x/n)|}\}\right),$$

in view of our previous observation and the triangle inequality.

Finally, our assumption that $\sum_{n=1}^{\infty} \frac{|a_n|}{n^c}$ converges implies that both $\sum_{n=1}^{\infty} \int_{c-iT}^{c+iT} \left|a_n \frac{x^s}{n^s}\frac{1}{s}\right| |ds|$ and $\int_{c-iT}^{c+iT} \sum_{n=1}^{\infty} \left|a_n \frac{x^s}{n^s}\frac{1}{s}\right| |ds|$ are convergent, and therefore we may swap the summation and integration and deduce that

$$\sum_{n=1}^{\infty} a_n \frac{1}{2\pi i} \int_{c-iT}^{c+iT} \frac{x^s}{n^s}\frac{ds}{s} = \frac{1}{2\pi i} \int_{c-iT}^{c+iT} \left(\sum_{n=1}^{\infty} \frac{a_n}{n^s} \right) x^s \frac{ds}{s}.$$

□

Remark 11.3. In fact, as we have stated Lemma 11.2 it remains valid with $\sum_{n \le x}' a_n$ replaced by $\sum_{n \le x} a_n$, since if x is an integer then the "big Oh" term in the lemma is at least as large as $|a_x|$.

Lemma 11.2 provides our desired relationship between a Dirichlet series and the counting function of its coefficients. Note that if T is chosen larger, meaning that we input information about the Dirichlet series at a wider range of "frequencies" $c + it$, then the "big Oh" error term becomes smaller.

12. The Riemann Hypothesis

In 1859, in his only paper on analytic number theory, Riemann extended the definition of the zeta function to all $s \in \mathbb{C}$ (with only a simple pole at $s = 1$), gave several proofs of a *functional equation* establishing some symmetry of $\zeta(s)$ about the *critical line* $\Re(s) = 1/2$, and proposed how to use the zeta function and Perron's inversion formula to prove results like the Prime Number Theorem (although it took until 1896 for Hadamard and de la Vallée Poussin to actually achieve this). He also said that the following statement about the zeros of $\zeta(s)$ (that is, the values s for which $\zeta(s) = 0$) seems "very probable":

Conjecture 12.1 (Riemann Hypothesis)**.** *There are no zeros s of the Riemann zeta function with $\Re(s) > 1/2$.*

Because of the functional equation, this is equivalent to saying that all zeros of the zeta function in the strip $0 \le \Re(s) \le 1$ satisfy $\Re(s) = 1/2$. The *Riemann Hypothesis* is now an extremely famous conjecture, and in this section we will think about why it might be true, and why it is interesting.

It isn't obvious at first sight that the zeros of $\zeta(s)$ have any significance at all, so we begin by noting that there are certainly no zeros satisfying $\Re(s) > 1$.

Lemma 12.2. *When $\Re(s) > 1$ we have*

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

where the Möbius function $\mu(n)$ is zero if n has any non-trivial square divisors, and $\mu(n) := (-1)^{\omega(n)}$ otherwise.

In particular, when $\Re(s) > 1$ we have $\zeta(s) \neq 0$.

Proof of Lemma 12.2. The Euler product expression for the zeta function (Lemma 10.2) implies that

$$\zeta(s) \cdot \prod_p \left(1 - \frac{1}{p^s}\right) = 1, \quad \Re(s) > 1.$$

By expanding this product, we find it is equal to $\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}$, an absolutely convergent series. In particular, it is some finite complex number whose multiplicative inverse is $\zeta(s)$, so we cannot have $\zeta(s) = 0$. □

Note that the real reason that $\zeta(s) \neq 0$ when $\Re(s) > 1$ is because the zeta function encodes information about the primes via the Euler product. It turns out that if $\zeta(s) \neq 0$ on a wider range of s, then this implies non-trivial things about the primes, which is the most important reason that the Riemann Hypothesis is an interesting conjecture. But why should we believe that it is a true conjecture?

Proposition 12.3 (Möbius cancellation implies RH, Littlewood, 1912)**.** *Suppose that for all small $\epsilon > 0$, we have $\sum_{n \leq x} \mu(n) \ll_\epsilon x^{1/2+\epsilon}$. Then the Riemann Hypothesis is true.*

Proof of Proposition 12.3. We will show that if we have the estimate $\sum_{n \leq x} \mu(n) \ll_\epsilon x^{1/2+\epsilon}$, then the Dirichlet series $\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}$ converges and defines a holomorphic function whenever $\Re(s) > 1/2$. If we can show this then we will be done, because Lemma 12.2 implies that

$$\zeta(s) \cdot \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} - 1 = 0 \quad \forall \Re(s) > 1,$$

and so the Identity Theorem from complex analysis will imply that

$$\zeta(s) \cdot \sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} - 1 = 0 \quad \forall \Re(s) > 1/2,$$

and so as before we must have $\zeta(s) \neq 0$.

For any large $N \in \mathbb{N}$, we have

$$\sum_{n=1}^{N} \frac{\mu(n)}{n^s} = \sum_{n=1}^{N} \mu(n) \left(s \int_n^N \frac{dw}{w^{s+1}} + \frac{1}{N^s}\right) = s \int_1^N \frac{\sum_{n \leq w} \mu(n)}{w^{s+1}} dw + \frac{\sum_{n \leq N} \mu(n)}{N^s}.$$

But if $\Re(s) = 1/2 + 2\epsilon$, then

$$\left| \frac{\sum_{n \leq w} \mu(n)}{w^{s+1}} \right| = \frac{|\sum_{n \leq w} \mu(n)|}{w^{3/2+2\epsilon}} \ll_\epsilon \frac{1}{w^{1+\epsilon}},$$

by our assumption about $\sum_{n \leq w} \mu(n)$. Thus the integral in the previous display is absolutely convergent as $N \to \infty$, and the term $\frac{\sum_{n \leq N} \mu(n)}{N^s}$ tends to zero. Moreover the convergence is uniform on any *compact* subset of $\{\Re(s) > 1/2\}$ (since in such a subset the real part of s will be bounded some fixed distance away from 1/2, and $|s|$ will be bounded), so $\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}$ converges to a holomorphic function. □

There are actually many statements whose truth implies the Riemann Hypothesis, so Proposition 12.3 doesn't in itself say that sums of the Möbius function are especially interesting. But it turns out that the Riemann Hypothesis is not only implied by cancellation in $\sum_{n \leq x} \mu(n)$, but is equivalent to it.

Proposition 12.4 (RH implies Möbius cancellation, Littlewood, 1912)**.** *If the Riemann Hypothesis is true, then for all small $\epsilon > 0$ we have $\sum_{n \leq x} \mu(n) \ll_{\epsilon} x^{1/2+\epsilon}$.*

Proof of Proposition 12.4. In view of the truncated Perron formula (Lemma 11.2 and the subsequent Remark), for any $x, T > 0$ and any $c > 1$ we have

$$\begin{aligned}\sum_{n \leq x} \mu(n) &= \frac{1}{2\pi i} \int_{c-iT}^{c+iT} \left(\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s}\right) x^s \frac{ds}{s} + O\left(x^c \sum_{n=1}^{\infty} \frac{|\mu(n)|}{n^c} \min\{1, \frac{1}{T|\log(x/n)|}\}\right) \\ &= \frac{1}{2\pi i} \int_{c-iT}^{c+iT} \frac{1}{\zeta(s)} x^s \frac{ds}{s} + O\left(x^c \sum_{n=1}^{\infty} \frac{1}{n^c} \min\{1, \frac{1}{T|\log(x/n)|}\}\right).\end{aligned}$$

Here we used Lemma 12.2 to write $\sum_{n=1}^{\infty} \frac{\mu(n)}{n^s} = \frac{1}{\zeta(s)}$ when $\Re(s) = c > 1$. We shall assume henceforth that x is large (this being the case we are interested in) and that $T \geq 2$ and $1 < c \leq 2$ (as we shall eventually choose them).

If the Riemann Hypothesis is true, then the integrand $\frac{1}{\zeta(s)} \frac{x^s}{s}$ has no poles in the half-plane $\Re(s) > 1/2$, so Cauchy's Residue Theorem implies that

$$\begin{aligned}\int_{c-iT}^{c+iT} \frac{1}{\zeta(s)} x^s \frac{ds}{s} &= \int_{1/2+\epsilon/2-iT}^{1/2+\epsilon/2+iT} \frac{1}{\zeta(s)} x^s \frac{ds}{s} + \int_{1/2+\epsilon/2+iT}^{c+iT} \frac{1}{\zeta(s)} x^s \frac{ds}{s} - \int_{1/2+\epsilon/2-iT}^{c-iT} \frac{1}{\zeta(s)} x^s \frac{ds}{s} \\ &\ll x^{1/2+\epsilon/2} \log(2+T) \max_{|t| \leq T} \frac{1}{|\zeta(1/2+\epsilon/2+it)|} \\ &\quad + \frac{x^c}{\log x} \frac{1}{T} \max_{\substack{1/2+\epsilon/2 \leq \sigma \leq c, \\ t = \pm T}} \frac{1}{|\zeta(\sigma+it)|}.\end{aligned}$$

To estimate the terms involving $1/\zeta$ we shall invoke the following lemma, whose proof we shall sketch later.

Lemma 12.5. *If the Riemann Hypothesis is true, then for any $\delta, \epsilon > 0$ we have*

$$-\delta \log|t| \leq \log|\zeta(\sigma+it)| = \Re \log \zeta(\sigma+it) \leq \delta \log|t| \quad \forall \sigma \geq 1/2+\epsilon, \ |t| \gg_{\epsilon,\delta} 1.$$

If we apply Lemma 12.5 with the choice $\delta = \epsilon/100$ (and with ϵ replaced by $\epsilon/2$), we deduce that

$$\max_{|t| \leq T} \frac{1}{|\zeta(1/2+\epsilon/2+it)|} \ll_{\epsilon} T^{\epsilon/100}, \quad \text{and} \quad \max_{\substack{1/2+\epsilon/2 \leq \sigma \leq c, \\ t = \pm T}} \frac{1}{|\zeta(\sigma+it)|} \ll_{\epsilon} T^{\epsilon/100}.$$

To control the "big Oh" term from the truncated Perron formula, we will take a fairly crude approach and just note that if $|n-x| \geq \frac{x}{\sqrt{T}}$ then $|\log(x/n)| = |\log(1+\frac{x-n}{n})| \gg |\log(1+\frac{1}{\sqrt{T}})| \gg \frac{1}{\sqrt{T}}$. Therefore we have

$$x^c \sum_{n=1}^{\infty} \frac{1}{n^c} \min\{1, \frac{1}{T|\log(x/n)|}\} \ll x^c \sum_{|n-x|<\frac{x}{\sqrt{T}}} \frac{1}{n^c} + \frac{x^c}{\sqrt{T}} \sum_{|n-x| \geq \frac{x}{\sqrt{T}}} \frac{1}{n^c} \ll \sum_{|n-x|<\frac{x}{\sqrt{T}}} 1 + \frac{x^c}{\sqrt{T}} \sum_{n=1}^{\infty} \frac{1}{n^c},$$

on noting that if $|n-x| < \frac{x}{\sqrt{T}}$ then $n \asymp x$, so $\frac{1}{n^c} \asymp \frac{1}{x^c}$. The first sum here is clearly $\ll 1 + \frac{x}{\sqrt{T}}$, whilst the second term is $\ll \frac{x^c}{\sqrt{T}} \int_1^{\infty} \frac{1}{y^c} dy \ll \frac{x^c}{(c-1)\sqrt{T}}$.

Collecting everything together, we have shown that

$$\sum_{n \leq x} \mu(n) \ll_{\epsilon} x^{1/2+\epsilon/2} \log(2+T) T^{\epsilon/100} + \frac{x^c}{T^{1-\epsilon/100} \log x} + \frac{x}{\sqrt{T}} + \frac{x^c}{(c-1)\sqrt{T}}.$$

The standard choice here is to take $c = 1 + \frac{1}{\log x}$, since this satisfies our condition that $c > 1$ and yet we still have $x^c = x x^{1/\log x} \ll x$. If we make this choice, then our bound becomes

$$\sum_{n \leq x} \mu(n) \ll_{\epsilon} x^{1/2+\epsilon/2} \log(2+T) T^{\epsilon/100} + \frac{x \log x}{\sqrt{T}}.$$

Finally, Proposition 12.4 follows if we choose $T = x$, say. □

At this point we know (apart from the need to prove Lemma 12.5) that the truth of the Riemann Hypothesis is equivalent to having the bounds $\sum_{n \leq x} \mu(n) \ll_{\epsilon} x^{1/2+\epsilon}$ for sums of the Möbius function. We don't know how to prove such strong bounds, but there are various heuristic reasons, of different levels of sophistication, that we might believe they should hold.

Model 12.6 (Random walk model)**.** *Since $\mu(n)$ takes the values ± 1, and (it turns out) takes each about half of the time (except on the well-understood sequence of n having a non-trivial square divisor, where $\mu(n) = 0$), the most naive model would be to assume that the successive values of $\mu(n)$ (on squarefree values of n) behave like independent random signs, and so $\sum_{n \leq x} \mu(n)$ behaves like a simple random walk.*

If this is true then the bound $\sum_{n \leq x} \mu(n) \ll_{\epsilon} x^{1/2+\epsilon}$ should certainly hold, and in fact the Law of the Iterated Logarithm (as discussed in section 7) would suggest that the fluctuations should be of order $\sqrt{x \log\log x}$ (and not more). One probably wouldn't accept the random walk model, which reflects no arithmetic information, to a very fine level of precision (like the Iterated Logarithm statement), but as a rough guide it is widely accepted. Recently Sarnak has formulated a sequence of precise but general conjectures suggesting there should be lots of cancellation in many sums of the form $\sum_{n \leq x} \mu(n) g(n)$.

Model 12.7 (Random multiplicative model, Wintner, 1940)**.** *To introduce some arithmetic structure into the random walk model, Wintner proposed modelling $\mu(n)$ by a random sequence $f(n)$, where the values $(f(p))_{p \text{ prime}}$ are independent random signs, and for general n we have $f(n) := \prod_{p|n} f(p)$ if n is squarefree, and $f(n) := 0$ if n is not squarefree. The Möbius function is the special case where $f(p) = -1$ for all primes p.*

The random multiplicative model is much less understood as a probabilistic object than the simple random walk, but as Wintner proved it certainly also satisfies $\sum_{n \leq x} f(n) \ll_{\epsilon} x^{1/2+\epsilon}$ with probability 1. It is not clear whether it is a really good model for the Möbius function, but it does seem to model at least some interesting arithmetic phenomena quite well.

Model 12.8 (Random matrix model, Montgomery, 1972)**.** *A popular model for the Riemann zeta function on the critical line $\Re(s) = 1/2$ is the characteristic polynomial of a suitable random matrix, with the eigenvalues of the random matrix then supposed to correspond to the zeros of the zeta function.*

This kind of model isn't directly a model for the Möbius function, but it leads to various conjectures (e.g. for the moments $\int_0^T |\zeta(1/2+it)|^{2k} dt$ of the zeta function, by Keating and Snaith) and thereby, indirectly, to conjectures about the Möbius function that we can compare with the previous two models. Gonek has conjectured, based on these kinds of ideas, that $\sum_{n \leq x} \mu(n)$ should have fluctuations of order $\sqrt{x}(\log\log\log x)^{5/4}$, and not more.

Finally we shall sketch the proof of Lemma 12.5.

Sketch proof of Lemma 12.5. We assume for simplicity that $1/2 + \epsilon \leq \sigma \leq 1$, but the case $\sigma > 1$ can be treated with a similar (or easier) argument. The proof will assume two complex analysis facts, which are both variants of the maximum modulus principle.

The first, the *Borel–Carathéodory theorem*, asserts that one can bound the modulus of a holomorphic function at a point given a bound for its real part on a surrounding disc: if $h(z)$ is holomorphic on the disc $|z| \leq R$, and if $h(0) = 0$, and if $r < R$, then

$$\max_{|z| \leq r} |h(z)| \leq \frac{2r}{R - r} \max_{|z| \leq R} \Re h(z).$$

We apply this to $h(z) := \log \zeta(2 + it + z) - \log \zeta(2 + it)$, with $R = 3/2 - \epsilon/2$. Since we assume the Riemann Hypothesis, $\zeta(2+it+z)$ does not vanish inside the disc $|z| \leq R$, so $\log \zeta(2 + it + z)$ is holomorphic in the disc. By construction we also have $h(0) = 0$. To apply the theorem we need an upper bound for $\Re h(z) = \log|\zeta(2+it+z)| - \log|\zeta(2+it)|$, and since $|\zeta(2+it)| = \prod_p |1 - \frac{1}{p^{2+it}}|^{-1}$ it is easy to see that $\log|\zeta(2+it)| = O(1)$, so we really just need an upper bound for $\max_{|z| \leq 3/2-\epsilon/2} \log|\zeta(2+it+z)|$.

Our second definition of the zeta function (Definition 10.3) implies that, if $\Re(s) \geq 1/2$, if $|\Im(s)| \geq 2$, and if $x \geq 10$ (say) is a parameter, then

$$|\zeta(s)| = \left| \sum_{n \leq x} \frac{1}{n^s} + \frac{x^{1-s}}{s-1} + \frac{\{x\}}{x^s} - s\int_x^\infty \{w\} \frac{dw}{w^{s+1}} \right| \ll \sum_{n \leq x} \frac{1}{n^{\Re(s)}} + x^{1-\Re(s)} + |s| \int_x^\infty \frac{dw}{w^{\Re(s)+1}}$$
$$\ll (1 + x^{1-\Re(s)}) \log x + \frac{|s|}{x^{\Re(s)}}.$$

(Here we used the fact that $\sum_{n \leq x} \frac{1}{n^\sigma} \ll \frac{x^{1-\sigma}}{1-\sigma} \ll x^{1-\sigma}\log x$ if $\sigma \leq 1 - 1/\log x$, and $\sum_{n \leq x} \frac{1}{n^\sigma} \ll \sum_{n \leq x} \frac{1}{n} \ll \log x$ if $\sigma > 1 - 1/\log x$.) In particular, if we just choose $x = 10$ we obtain the crude bound $|\zeta(s)| \ll |s|$ on our range of s, so for $|z| \leq 3/2 - \epsilon/2$ we have $|\zeta(2+it+z)| \ll |t|$, and $\log|\zeta(2+it+z)| \leq \log|t| + O(1)$. (*Note this is only an upper bound, since we haven't yet excluded the possibility that $\zeta(2+it+z)$ could be very close to zero. But the Borel–Carathéodory theorem only requires an upper bound.*) Applying the Borel–Carathéodory theorem with $r = 3/2 - \epsilon$, it follows that

$$\max_{|z| \leq r} |\log\zeta(2+it+z)| = \max_{|z| \leq r} |\log\zeta(2+it+z) - \log\zeta(2+it)| + O(1) \ll \frac{\log|t|}{\epsilon},$$

so in particular

$$|\log\zeta(\sigma+it)| \ll \frac{\log|t|}{\epsilon} \quad \forall \sigma \geq 1/2 + \epsilon, \ |t| \gg 1.$$

This bound is not yet good enough for Lemma 12.5, but we can refine it using our second complex analysis fact. *Hadamard's three-circles theorem* asserts that if $0 < R_1 < R_2$, and if $f(z)$ is a holomorphic function on the annulus $R_1 \leq |z| \leq R_2$, then for any $R_1 \leq r \leq R_2$ we have

$$\max_{|z|=r} |f(z)| \leq \left(\max_{|z|=R_1} |f(z)| \right)^{\frac{\log(R_2/r)}{\log(R_2/R_1)}} \left(\max_{|z|=R_2} |f(z)| \right)^{\frac{\log(r/R_1)}{\log(R_2/R_1)}}.$$

We apply this to the function $f(z) := \log\zeta(1.1+it+z)$, say, and with $R_1 = 0.05$ and $R_2 = 0.6 - \epsilon/2$. The Riemann Hypothesis implies that $\zeta(1.1+it+z)$ does not vanish in the disc $|z| \leq R_2$, so $f(z)$ is holomorphic there. Our previous calculations show that $\max_{|z|=R_2} |f(z)| \ll \frac{\log|t|}{\epsilon}$, whilst the Euler product expression directly implies that $\max_{|z|=R_1} |f(z)| = \max_{|z|=R_1} |\sum_p \log(1 - \frac{1}{p^{1.1+it+z}})| \ll \sum_p \frac{1}{p^{1.05}} \ll 1$. If $1/2 + \epsilon \leq \sigma \leq 1$ then we can take $R_1 < r = 1.1 - \sigma \leq R_2 - \epsilon/2$, and conclude

$$|\log\zeta(\sigma+it)| \leq \max_{|z|=r} |f(z)| \ll \left(\frac{\log|t|}{\epsilon} \right)^{\frac{\log(r/R_1)}{\log(R_2/R_1)}} \ll \left(\frac{\log|t|}{\epsilon} \right)^{1-c\epsilon},$$

for a certain small constant $c > 0$. In particular, for any $\delta > 0$, if $|t|$ is large enough depending on ϵ and δ we will have

$$|\log \zeta(\sigma + it)| \leq \delta \log |t|,$$

which gives Lemma 12.5. □

Remark 12.9. Many of the proofs in this section consist mostly of complex analysis "trickery", such as the Identity Theorem, Borel–Carathéodory theorem, and Hadamard's three-circles theorem. But the key input into the proofs, in which it was important we were working with the zeta function rather than some general function, was the Euler product expression; a rough size estimate deduced from Definition 10.3; and the assumption of the Riemann Hypothesis so we could work with $\log \zeta(s)$ as a holomorphic function. In general, complex analysis provides a language and toolbox for manipulating data about the zeta function (and, through it, about the primes), but we always need estimates for arithmetic things (like primes in the Euler product) as raw data to prove anything interesting.

References

[1] H. L. Montgomery, R. C. Vaughan. *Multiplicative Number Theory I: Classical Theory.* First edition, published by Cambridge University Press. 2007

[2] E. C. Titchmarsh. *The Theory of the Riemann Zeta-function.* Second edition, revised by D. R. Heath-Brown, published by Oxford University Press. 1986

Jesus College, Cambridge, CB5 8BL
E-mail address: A.J.Harper@dpmms.cam.ac.uk