(I.C) THE DISTRIBUTION OF PRIMES

In the last section we showed — via a Euclid-inspired, algebraic argument — that there are infinitely many primes of the form $p = 4n - 1$ (i.e. $4n + 3$). In fact, this is true for primes of the form $4n + 1$ as well, and the ratio of primes of these two forms less than $N$ tends to 1 as $N \to \infty$. We say that the primes are distributed “asymptotically equally” between $\{4n + 1 \mid n \in \mathbb{N}\}$ and $\{4n - 1 \mid n \in \mathbb{N}\}$.

More generally, taking $\mathbb{P} \subset \mathbb{N}$ to denote the primes,

$$\mathbb{N}_{a,b} := \{a + nb \mid n \in \mathbb{Z}\} \cap \mathbb{N},$$

and

$$\mathbb{P}_{a,b} := \mathbb{N}_{a,b} \cap \mathbb{P},$$

there is the famous theorem on primes in arithmetic progressions:

**Theorem 1** (Dirichlet, 1837). Given $a, b \in \mathbb{N}$ such that $(a, b) = 1$, the set $\mathbb{P}_{a,b}$ is infinite. Moreover, for each fixed $b$, the primes are distributed asymptotically equally between the $\{\mathbb{P}_{a,b} \}_{0 < a < b}$.

In the early 20th century, people began to notice that the $\mathbb{N}_{a,b}$ contained consecutive sequences of primes, e.g.

- $\mathbb{N}_{3,4} \supset \{3, 7, 11\}$ [length 3]
- $\mathbb{N}_{7,30} \supset \{7, 37, 67, 97, 127, 157\}$ [length 6] (1909)
- $\mathbb{N}_{199,210} \supset \{199, 409, \ldots\}$ [length 10] (1910)

A sequence of length 11 wasn’t found until 1999; the longest known today has length 26 (and begins with a 16-digit number). In light of this, the theoretical result is impressive:

**Theorem 2** (Green and Tao, 2004). Given any $k$, there exist $a$ and $b$ such that $k$ consecutive elements of $\mathbb{N}_{a,b}$ are prime.
One question which may bug you (for instance, in relation to the sequences (1)) is:

- How do you know if a number $N$ is prime?

Naively, it’s enough to check that no number $\leq \sqrt{N}$ divides $N$, but we will find better methods later. A second question is:

- How can one construct primes?

There is no nice answer here — no known function which produces distinct primes (and only primes).\(^1\)

There are many other longstanding riddles regarding the primes: for example,

**Conjecture 3** (Goldbach). *Any even number is the sum of two primes.*

This is known up to 18 digits but not proved in general. (A famous result of Vinogradov from the 1930s says that any sufficiently large odd number is the sum of 3 primes.) Alternatively, one might try to go further than Theorem 2 and ask whether, given any $k$ and $b$ (with $b$ even), there exist infinitely many sequences

$$\{m + b, m + 2b, \ldots, m + kb\}$$

consisting entirely of primes. Taking $k = 2$ yields the venerable

**Conjecture 4** (de Polignac, 1849). *Given any even $b \in \mathbb{N}$, there exist infinitely many pairs $p, q \in \mathbb{P}$ with $p - q = b$.*

The case $b = 2$ is known as the **twin prime conjecture**. A spectacular and unexpected recent advance is:

**Theorem 5** (Zhang, 2013). *Conjecture 4 holds for some $b < 70,000,000$.*

Recent work has brought this upper bound down to 246, but for the moment, the twin prime conjecture remains open.

\(^1\)In the exercises, you will verify that no polynomial function can possibly do this.
Dirichlet’s L-functions. We now turn to the idea behind the proof of Dirichlet’s theorem (in the special case $b = 4$), starting with Euler’s analytic approach to the infinitude of primes. What follows is far from being rigorous.

Let $s > 1$, and consider the “Euler product”

$$\prod_{p \text{ prime}} \frac{1}{1 - p^{-s}} = \prod_{p \text{ prime}} \left(1 + p^{-s} + p^{-2s} + p^{-3s} + \cdots\right)$$

where we have expanded each factor $(1 - p^{-s})^{-1}$ as a geometric series on the right. If we formally expand the right-hand product, then by the Fundamental Theorem of Arithmetic, each $n^{-s} = (p_1^{a_1} \cdots p_k^{a_k})^{-s} = p_1^{-a_1} \cdots p_k^{-a_k}$ occurs exactly once in the result, yielding

$$= 1 + 2^{-s} + 3^{-s} + 4^{-s} + 5^{-s} + 6^{-s} + \cdots$$

$$= \sum_{n \geq 1} n^{-s} =: \zeta(s),$$

the Riemann zeta function.

Remark 6. The series converges for $s > 1$ in $\mathbb{R}$, by the integral comparison test

$$\sum_{n \geq 2} \frac{1}{n^s} < \int_1^\infty \frac{dx}{x^s} = \left[ \frac{x^{-s+1}}{-s+1} \right]_1^\infty = \frac{1}{s-1},$$

and more generally for $\text{Re}(s) > 1$ in the complex numbers $\mathbb{C}$. One may “analytically continue” it to get an analytic function on $\mathbb{C} \setminus \{1\}$ (with a simple pole at 1).

Now what could some analytic function have to do with the distribution of primes? Quite a bit: to begin with, formally taking the limit of the above as $s \to 1^+$ gives

$$\prod_{p \text{ prime}} \frac{1}{1 - p^{-1}} = \sum_{n \geq 1} \frac{1}{n} = \infty,$$

which “proves” the infinitude of primes.$^3$

---

2 because there exists a unique prime factorization of each $n \in \mathbb{N}$

3 In mathematics, “formally” often means “manipulating symbols”, which is about as far from rigor as one gets (and is great for producing or conveying ideas but also
The idea of Dirichlet’s proof is to refine this observation. Let
\[ \chi_0(n) := \begin{cases} 0, & n \text{ even} \\ 1, & n \text{ odd} \end{cases} \]
and
\[ \chi_1(n) := \begin{cases} 0, & n \text{ even} \\ 1, & n = 4k + 1 \\ -1, & n = 4k + 3 \end{cases} \]
you should check that
(2) \[ \chi_i(mn) = \chi_i(m) \chi_i(n) \]
for \( m, n \in \mathbb{N} \) (and \( i = 0, 1 \)). We now carry out an analogue of the above argument, but in reverse, starting with the Dirichlet series (or \( L \)-function)
\[ L(\chi_i, s) := \sum_{n \geq 1} \frac{\chi_i(n)}{n^s} \]
which using the Fundamental Theorem and (2) becomes
(3) \[ = \prod_{p \text{ prime}} \left( \sum_{k \geq 0} \frac{\chi_i(p^k)}{p^{ks}} \right) = \prod_{p \text{ prime}} \left( \sum_{k \geq 0} \left( \frac{\chi(p)}{p^s} \right)^k \right) \]
\[ = \prod_{p \text{ prime}} \frac{1}{1 - \chi_i(p) \frac{p^s}{s}}. \]
Taking log of (3) yields
\[ \log L(\chi_i, s) = - \sum_{p \text{ prime}} \log \left( 1 - \frac{\chi_i(p)}{p^s} \right), \]
which using \(- \log(1 - x) = \sum_{k \geq 1} \frac{x^k}{k}\) becomes
\[ = \sum_{p \text{ prime}} \sum_{k \geq 1} \frac{\chi_i(p^k)}{kp^{ks}} = \sum_{p \text{ prime}} \frac{\chi_i(p)}{p^s} + \sum_{p \text{ prime}} \sum_{k \geq 2} \frac{\chi_i(p^k)}{kp^{ks}}. \]
\[ =: f_i(s) \]
can be terrifically misleading). For a proof without the quote marks, see the next subsection.
We can bound this last term (for \( i = 0 \) or \( 1 \)) by

\[
|f_i(s)| \leq \sum_{p \text{ prime}} \sum_{k \geq 2} \frac{1}{kp^s} \leq \sum_{p \text{ prime}} \sum_{k \geq 2} \frac{1}{(p^s)^k} = \sum_{p \text{ prime}} \frac{p^{-2s}}{1 - p^{-s}}
\]

\[
\leq 2 \sum_{p \text{ prime}} p^{-2s} \leq 2 \sum_{n \geq 1} n^{-2s}
\]

which for \( s \geq 1 \) is

\[
\leq 2 \sum_{n \geq 1} \frac{1}{n^s} \leq 4.
\]

Finally, using

\[
\frac{\chi_0(n) + \chi_1(n)}{2} = \begin{cases} 
1, & n = 4k + 1 \\
0, & \text{otherwise}
\end{cases}
\]

and

\[
\frac{\chi_0(n) - \chi_1(n)}{2} = \begin{cases} 
1, & n = 4k + 3 \\
0, & \text{otherwise}
\end{cases}
\]

we have

(4) \( \frac{1}{2} \left( \log L(\chi_0, s) + \log L(\chi_1, s) \right) = \frac{f_0 + f_1}{2} + \sum_{p \text{ prime}} \frac{1}{p^s} \) \( \tag{A} \)

and

(5) \( \frac{1}{2} \left( \log L(\chi_0, s) - \log L(\chi_1, s) \right) = \frac{f_0 - f_1}{2} + \sum_{p \text{ prime}} \frac{1}{p^s} \) \( \tag{B} \)

Since \( L(\chi_1, 1) = \sum \frac{\chi_1(n)}{n} \) converges by the alternating series test (with nonzero limit \( \frac{\pi}{4} \)), only the \( \log L(\chi_0, s) \) and \( \sum_p \) terms of (4) and (5) diverge as \( s \to 1^+ \). It follows that (A) and (B) diverge at the same rate. This proves that there are infinitely many primes of the form \( 4k + 1 \), and suggests that they are distributed asymptotically equally to those of the form \( 4k + 3 \).
The infinitude of primes. Finally, we shall describe one way of making Euler’s argument above completely airtight, which has the added bonus of putting a lower bound on partial sums of inverse primes.

Lemma 7. \( e^{x+x^2} \geq \frac{1}{1-x} \) for \( x \in [0, \frac{1}{2}] \). (In particular, \( e^{\frac{1}{p} + \frac{1}{p^2}} \geq \frac{1}{1-\frac{1}{p}} \) for each prime \( p \).)

Proof. It suffices to show that

\[
(1 - x)e^{x+x^2} \geq 1.
\]

The left-hand side of this is 1 at \( x = 0 \) and has derivative \( x(1 - 2x)e^{x+x^2} \geq 0 \) for \( x \in [0, \frac{1}{2}] \). □

Theorem 8. For any real number \( y > 2 \),

\[
\sum_{\substack{p \leq y \\ p \text{ prime}}} \frac{1}{p} > \log(\log y) - 1.
\]

Corollary 9. \( \sum_{p \text{ prime}} \frac{1}{p} \) diverges. (In particular, there are infinitely many primes.)

Proof of Theorem 8. Given \( y > 2 \), set

\[
N_y := \{ n \in \mathbb{N} \mid n = p_1^{a_1} \cdots p_k^{a_k}, \text{ all } p_i \leq y \},
\]
and denote the greatest integer less than or equal to \( y \) by \( \lfloor y \rfloor \). Now using the lemma together with the Fundamental Theorem, we find

\[
\prod_{\substack{p \leq y \\ p \text{ prime}}} e^{\frac{1}{p} + \frac{1}{p^2}} \geq \prod_{\substack{p \leq y \\ p \text{ prime}}} \frac{1}{1 - \frac{1}{p}} \\
= \prod_{\substack{p \leq y \\ p \text{ prime}}} \left(1 + \frac{1}{p} + \frac{1}{p^2} + \frac{1}{p^3} + \cdots\right) \\
= \sum_{n \in \mathbb{N}_y} \frac{1}{n} \\
\geq \sum_{n=1}^{\lfloor y \rfloor} \frac{1}{n} \geq \int_1^{1+\lfloor y \rfloor} \frac{dx}{x} = \log(1 + \lfloor y \rfloor) > \log(y).
\]

Taking log of both sides,

\[
\log \log y < \log \left( \prod_{\substack{p \leq y \\ p \text{ prime}}} e^{\frac{1}{p} + \frac{1}{p^2}} \right) \\
= \sum_{\substack{p \leq y \\ p \text{ prime}}} \left(\frac{1}{p} + \frac{1}{p^2}\right) \\
< \sum_{\substack{p \leq y \\ p \text{ prime}}} \frac{1}{p} + \sum_{n=2}^{\infty} \frac{1}{n^2}.
\]

which by the integral test is

\[
< \sum_{\substack{p \leq y \\ p \text{ prime}}} \frac{1}{p} + \int_1^{\infty} \frac{dx}{x^2} < 1
\]

\[\square\]

In the next section, we will study the function

\[\pi(x) := \text{ number of primes less than or equal to } x\]
on $\mathbb{R}_+ = (0, \infty)$. As a preliminary step, we can push Theorem 8 a bit further to get

**Corollary 10.** For $x > 2$,

$$\frac{\pi(x)}{x} + \int_2^x \frac{\pi(u)}{u^2} du > \log(\log x) - 1.$$  

**Proof.** Write $\pi(u) = \sum_{p \text{ prime}} \chi_{[p, \infty)}(u)$, where for any subset $\mathcal{S} \subset \mathbb{R}$,

$$\chi_{\mathcal{S}}(u) := \begin{cases} 1, & u \in \mathcal{S} \\ 0, & u \notin \mathcal{S} \end{cases}$$

is the characteristic function. Then we have

$$\int_2^x \frac{\pi(u)}{u^2} du = \sum_{p \text{ prime}} \int_2^x \frac{\chi_{[p, \infty)}(u)}{u^2} du$$

$$= \sum_{p \leq x \atop p \text{ prime}} \int_p^x \frac{du}{u^2}$$

$$= \sum_{p \leq x \atop p \text{ prime}} \left[ -\frac{1}{u} \right]_p^x$$

$$= \sum_{p \leq x \atop p \text{ prime}} \frac{1}{p} - \sum_{p \leq x \atop p \text{ prime}} \frac{1}{x}$$

$$= \sum_{p \leq x \atop p \text{ prime}} \frac{1}{p} - \frac{\pi(x)}{x}$$

$$> \log(\log x) - 1 - \frac{\pi(x)}{x},$$

as desired. \qed

\footnote{Note that only finitely many terms of the sum contribute, so switching with the integral is permissible.}