MATH 250: THE DISTRIBUTION OF PRIMES

ROBERT J. LEMKE OLIVER

1. Euler's work on primes

For $s \in \mathbb{R}$, define the function $\zeta(s)$ by

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

which converges if $s > 1$ and diverges if $s \leq 1$. In fact, though we will not exploit this fact just yet, we may also make sense of $\zeta(s)$ for $s \in \mathbb{C}$ provided that $\Re(s) > 1$. The connection between $\zeta(s)$ and prime numbers is the following.

Lemma 1.1 (Euler product for $\zeta(s)$)**.** *For $\Re(s) > 1$, we have*

$$\zeta(s) = \prod_{p \text{ prime}} \left(1 - \frac{1}{p^s}\right)^{-1}. \tag{1.1}$$

Proof. Geometric series + unique factorization. □

Notation. Going forward, whenever we index a summation or product by a variable p, we will implicitly assume it is to be taken only over prime values of p.

This factorization affords another proof that there are infinitely many primes.

Proposition 1.2. *There are infinitely many primes.*

Proof. Were there only finitely many primes, expanding the right-hand side of (1.1) as a geometric series would result in a series that converges absolutely at $s = 1$. However, the left-hand side diverges, therefore the right-hand side must as well, so there must be infintely many primes. □

In fact, the proof of Proposition 1.2 yields rather more information:

Theorem 1.3 ("The primes are not too sparse.")**.** *The series $\sum_p \frac{1}{p}$ diverges.*

Proof. The proof of Proposition 1.2 shows that the infinite product $\prod_p (1 - 1/p)^{-1}$ diverges. Thus,

$$\lim_{N \to \infty} \prod_{p \leq N} \left(1 - \frac{1}{p}\right)^{-1} = +\infty,$$

and applying log to both sides, we find

$$\lim_{N \to \infty} \sum_{p \leq N} -\log\left(1 - \frac{1}{p}\right) = +\infty$$

Date: September 6, 2017.

as well. Next, recall the Taylor series for $-\log(1-x)$,

$$-\log(1-x) = x + \frac{x^2}{2} + \frac{x^3}{3} + \dots,$$

which is absolutely convergent if $|x| < 1$. We therefore find

$$\begin{aligned}\sum_{p\le N} -\log\left(1-\frac{1}{p}\right) &= \sum_{p\le N}\frac{1}{p} + \sum_{p\le N}\frac{1}{2p^2} + \sum_{p\le N}\frac{1}{3p^3} + \dots \\ &= \sum_{p\le N}\frac{1}{p} + \sum_{k=2}^{\infty}\sum_{p\le N}\frac{1}{kp^k}.\end{aligned}$$

We know that the left-hand side above diverges as $N \to \infty$, so if we can show that the contribution on the right-hand side from the terms involving p^k with $k \ge 2$ is convergent, the theorem will follow. Naturally, it is now our goal to establish this convergence. We use a straightforward comparison test, noting that the summands are positive.

As the Taylor series for $-\log(1-x)$ is absolutely convergent for $|x| < 1$ and $1/p \le 1/2$ for any prime p, we have

$$\begin{aligned}\sum_{k=2}^{\infty}\sum_{p\le N}\frac{1}{kp^k} &= \sum_{p\le N}\sum_{k=2}^{\infty}\frac{1}{kp^k} \\ &\le \sum_{p\le N}\sum_{k=2}^{\infty}\frac{1}{2p^k} \\ &= \frac{1}{2}\sum_{p\le N}\frac{p^{-2}}{1-1/p} \\ &\le \sum_{p\le N}\frac{1}{p^2},\end{aligned}$$

where we have used the formula for the sum of a geometric series and the simple inequality $(1-1/p)^{-1} \le 2$ for all primes p. The limit as $N \to \infty$ of the term above is bounded by $\zeta(2)$, so it is convergent. Thus, we have

$$\sum_{p\le N}\frac{1}{p} \ge \sum_{p\le N} -\log\left(1-\frac{1}{p}\right) - \zeta(2),$$

and taking the limit as $N \to \infty$ yields the result. □

Remark. Theorem 1.3 provides much more information than Euclid's proof of the infinitude of the primes. For example, it shows that at least "typically," the n-th prime p_n must satisfy $p_n \le n^2$, since the series $\sum 1/n^2$ converges. In fact, in the same wishy-washy, "typical" sense, we must have $p_n \le n^{1+\epsilon}$ for any $\epsilon > 0$, or even $p_n \le n(\log n)^{1+\epsilon}$, which is very close to what is provided by the prime number theorem, that $p_n \sim n \log n$. More specifically, using techniques to be developed later in this lecture, Theorem 1.3 establishes for example that for any $\epsilon > 0$

$$\limsup_{x\to\infty} \frac{\pi(x)}{x/(\log x)^{1+\epsilon}} = +\infty. \tag{1.2}$$

To contrast this with Euclid's proof, set $a_1 = 2$, and for $n \geq 1$ define $a_{n+1} = a_1 \dots a_n + 1$. Notice that for $n \geq 2$, $a_{n+1} = (a_1 \dots a_{n-1})a_n + 1 = (a_n - 1)a_n + 1 = a_n^2 - a_n + 1$. In particular, a_{n+1} is close to a_n^2, and certainly $a_{n+1} \leq a_n^2$. Just using this inequality, it follows that $a_n \leq 2^{2^{n-1}}$, so that the same bound holds for p_n, namely $p_n \leq 2^{2^{n-1}}$. This holds unconditionally, and not in any wishy-washy sense, but all we may deduce from this is that

$$\liminf_{x\to\infty} \frac{\pi(x)}{\log\log x} > 0,$$

which is markedly worse than (1.2).

Surprisingly, even though Theorem 1.3 is about showing that the primes are not too sparse, very similar ideas show that the primes cannot be too common.

Theorem 1.4 ("The primes are not too common."). $\lim_{x\to\infty} \frac{\pi(x)}{x} = 0.$

Proof. As a first and quite crude approximation, note that every prime apart from 2 is odd, and that there are $\lfloor \frac{x-1}{2} \rfloor$ odd numbers $\leq x$. Thus,

$$\lim_{x\to\infty} \frac{\pi(x)}{x} \leq \lim_{x\to\infty} \frac{1 + \lfloor (x-1)/2 \rfloor}{x} = \frac{1}{2}.$$

To do better than this, we naturally want to incorporate more divisibility conditions other than just those coming from two. However, in carrying this out, it becomes rather painful rather quickly to come up with exact expressions akin to the $\lfloor \frac{x-1}{2} \rfloor$ count of odd numbers up to x. (Write out the corresponding expression for numbers that are relatively prime to 6. Do you have the stamina to do it for 30? For $2 \cdot 3 \cdot 5 \dots p$?) We therefore forgo our fetish for exact formulae; instead, as with something like the prime number theorem, we write down an approximate formula and keep track of how close it is.

Let $N \geq 2$ be a fixed integer (i.e., independent of x) and set $P(N) = \prod_{p \leq N} p$. The proportion of integers $n \leq x$ that have no prime factor $\leq N$ is roughly $\prod_{p\leq N}(1 - 1/p)$. In fact, this proportion is exact if x is an integral multiple of $P(N)$, i.e.

$$\#\{n < x : \gcd(n, P(N)) = 1\} \;=\; x \prod_{p \leq N} \left(1 - \frac{1}{p}\right),$$

if $x = kP(N)$ for some $k \in \mathbb{Z}$. Since this count is exact at multiples of $P(N)$, between multiples of $P(N)$ the count could be off by at most the difference between the two endpoint values, which we crudely bound by $P(N)$. Thus, for any x,

$$\#\{n < x : \gcd(n, P(N)) = 1\} \;=\; x \prod_{p \leq N} \left(1 - \frac{1}{p}\right) + O(P(N)). \tag{1.3}$$

Relating this back to $\pi(x)$, any prime p is either at most N or relatively prime to N. There are $\pi(N)$ of the first type, while those of the second are included in the count (1.3). Thus,

$$\pi(x) \leq \pi(N) + x \prod_{p \leq N} \left(1 - \frac{1}{p}\right) + O(P(N)).$$

Since N was fixed, this shows that

$$\lim_{x\to\infty} \frac{\pi(x)}{x} \leq \prod_{p\leq N} \left(1 - \frac{1}{p}\right) + \lim_{x\to\infty} \frac{\pi(N) + O(P(N))}{x} = \prod_{p\leq N} \left(1 - \frac{1}{p}\right). \tag{1.4}$$

From before, we know that

$$\lim_{N\to\infty}\prod_{p\le N}\left(1-\frac{1}{p}\right)^{-1}=+\infty,$$

whence

$$\lim_{N\to\infty}\prod_{p\le N}\left(1-\frac{1}{p}\right)=0.$$

Inserting this into (1.4) yields the theorem. □

Remark. It is tempting to use the ideas in the proof of Theorem 1.4 to attack a version of the prime number theorem. In particular, every integer $n \le x$ greater than 1 is either prime or admits a prime factor $p \le x^{1/2}$. This gives a quantitative version of the *Sieve of Eratosthenes*,

$$\pi(x)=\pi(x^{1/2})-1+\#\{n\le x:\gcd(n,P(x^{1/2}))=1\}. \tag{1.5}$$

One might then hope that the formula in (1.3) holds with some controlled error, something like

$$\#\{n\le x:\gcd(n,P(x^{1/2}))=1\}=x\prod_{p\le x^{1/2}}\left(1-\frac{1}{p}\right)+(\text{small error}). \tag{1.6}$$

However, it turns out an expression like this cannot hold with any plausible notion of "small!"

Theorem 1.5 (Mertens' theorem)**.** *For any $z \ge 2$, we have*

$$\prod_{p\le z}\left(1-\frac{1}{p}\right)\sim\frac{e^{-\gamma}}{\log z},$$

where $\gamma = 0.577\ldots$ is the Euler-Mascheroni constant.

We will give a post hoc proof of Theorem 1.5 in a later lecture after we've proved the prime number theorem (even though Theorem 1.5 predates the prime number theorem), but for right now, let us be content with observing that it implies

$$\prod_{p\le x^{1/2}}\left(1-\frac{1}{p}\right)\sim\frac{2e^{-\gamma}}{\log x}.$$

Thus, if (1.6) held, then (1.5) would yield

$$\pi(x)\sim 2e^{-\gamma}\frac{x}{\log x}.$$

This contradicts the prime number theorem, since $e^{-\gamma}=0.561\cdots\ne 1/2$, so something must be awry. The only room to give is in our assumption of (1.6), so it must be the "small error" in (1.6) is not so small.

Despite this being a failed and fundamentally flawed attempt at proving the prime number theorem, there is merit in the approach. First, taking $N = \log x$ in the proof of Theorem 1.4 would yield

$$\limsup_{x\to\infty}\frac{\pi(x)}{x/\log\log x}<\infty,$$

and this choice of N is relatively easy to justify. In fact, with more effort than we care to put in here, a careful treatment of the error term in (1.3) justifies a choice of N that shows

$$\limsup_{x\to\infty} \frac{\pi(x)}{x \log\log x/\log x} < \infty,$$

which, combined with (1.2), is starting to zero in on the prime number theorem. Second, there are problems like the twin prime conjecture where we don't know whether there are infintely many such primes, let alone whether an analogue of the prime number theorem holds. However, expanding on the ideas above, it is possible to obtain upper bounds on the number of such primes. This is the basis of the field called 'sieve theory.' The first person to develop these ideas was Viggo Brun. A very pleasing consequence of Brun's work is the statement that the sum of the reciprocals of the twin primes converges. (In light of the discussion around Theorem 1.3, this amounts to showing that there are not "too many" twin primes.) We unfortunately won't spend very much time on Brun's work in this course, though we will spend some time discussing the so-called Selberg sieve when it comes time to prove the Maynard-Tao theorem.