MATH 250: THE DISTRIBUTION OF PRIMES

ROBERT J. LEMKE OLIVER

7. The prime number theorem in arithmetic progressions

In this lecture, we continue our discussion of primes in arithmetic progressions. Let $\gcd(a,q)=1$, and define

$$\pi(x;q,a) := \#\{p<x : p\equiv a \pmod q\}.$$

In light of our proof of Dirichlet's theorem, which established

$$\lim_{s\to 1^+} \frac{\sum_{p\equiv a(\mathrm{mod}\, q)} p^{-s}}{\sum_p p^{-s}} = \frac{1}{\phi(q)}, \tag{7.1}$$

it is reasonable to expect that primes $p\equiv a \pmod q$ have density $1/\phi(q)$ within the set of all primes, i.e.

$$\lim_{x\to\infty} \frac{\pi(x;q,a)}{\pi(x)} = \frac{1}{\phi(q)} \quad \text{or} \quad \pi(x;q,a)\sim \frac{\mathrm{Li}(x)}{\phi(q)}. \tag{7.2}$$

Using (7.1), it is possible to show that if the limit in (7.2) exists, then it must be equal to $1/\phi(q)$. By itself, however, (7.1) is not quite enough to establish (7.2). For example, (7.1) is consistent with a weird scenario in which no primes congruent to $a \pmod q$ start with the digit 9, a bizarre situation (7.2) can rule out. Nevertheless, combining the ideas behind the proof of Dirichlet's theorem with those of the prime number theorem, it is possible to establish (7.2) without too much additional work. We perform a brief survey of the approach, leaving most details unexplored as they are almost directly analogous to those of the prime number theorem. It is to be understood, however, that this is part of a beautiful theory that students interested in analytic number theory should be encouraged to explore.

In particular, if we define

$$\psi(x;q,a) = \sum_{\substack{n\le x\\ n\equiv a(\mathrm{mod}\, q)}} \Lambda(n)$$

then Lemma 6.5 implies

$$\begin{aligned}\psi(x;q,a) &= \sum_{n\le x}\Lambda(n)\cdot\frac{1}{\phi(q)}\sum_{\chi(\mathrm{mod}\, q)}\overline{\chi}(a)\chi(n)\\ &= \frac{1}{\phi(q)}\sum_{\chi(\mathrm{mod}\, q)}\overline{\chi}(a)\psi(x;\chi),\end{aligned} \tag{7.3}$$

say, where

$$\psi(x;\chi)=\sum_{n\le x}\chi(n)\Lambda(n).$$

Date: October 22, 2017.

Moreover, by Perron's formula, in analogy with Theorem 3.10 we find for each $\chi \pmod q$

$$\psi(x;\chi) - \frac{1}{2}\Lambda(x)\chi(x) = \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} -\frac{L'(s,\chi)}{L(s,\chi)}x^s\frac{ds}{s}.$$

This suggests that there should be an explicit formula for $\psi(x;\chi)$, and proceeding along the lines of the proof of Theorem 4.4, we obtain the following.

Theorem 7.1 (Explicit formula for $\psi(x;\chi)$)**.** *For a character $\chi \pmod q$, let*

$$\delta(\chi) = \begin{cases} 1, & \text{if } \chi = \chi_0 \text{ is trivial, and} \\ 0, & \text{if } \chi \neq \chi_0. \end{cases}$$

Then for $x \geq 1$, we have

$$\psi(x;\chi) - \frac{1}{2}\Lambda(x)\chi(x) = \delta(\chi)x - \sum_{\rho_\chi \neq 0}\frac{x^{\rho_\chi}}{\rho_\chi} - \mathrm{Res}_{s=0}\left[\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}\right],$$

where the sum runs over the zeros ρ_χ of $L(s,\chi)$.

Remark. We saw in the first exercise of the previous section that $L(0,\chi) = 0$ whenever χ is a non-trivial character such that $\chi(-1) = 1$. Thus, the residue at $s = 0$ may be more complicated than it is for $\zeta(s)$.

Analogous to Corollary 4.5, we deduce from Theorem 7.1 that the primes inherently must be somewhat erratically distributed among the classes $\pmod q$.

Corollary 7.2 ("The primes $\pmod q$ are complicated")**.** *Each $L(s,\chi)$ has infinitely many zeros.*

Proof. The left-hand side of the explicit formula is discontinuous, so the right-hand side must be as well. □

Additionally, combining Theorem 7.1 with (7.3) yields an explicit formula for $\psi(x;q,a)$.

Corollary 7.3. *Let $\gcd(a,q) = 1$. Then*

$$\psi(x;q,a) - \frac{1}{2}\Lambda_{a(\mathrm{mod}\,q)}(x) = \frac{x}{\phi(q)} - \frac{1}{\phi(q)}\sum_{\chi(\mathrm{mod}\,q)}\overline{\chi}(a)\sum_{\rho_\chi\neq 0}\frac{x^{\rho_\chi}}{\rho_\chi} - \sum_{\chi(\mathrm{mod}\,q)}\overline{\chi}(a)\mathrm{Res}_{s=0}\left[\frac{L'(s,\chi)}{L(s,\chi)}\frac{x^s}{s}\right],$$

where $\Lambda_{a(\mathrm{mod}\,q)}(x) = \Lambda(x)$ if x is an integer congruent to $a \pmod q$ and is 0 otherwise.

We are thus motivated to understand the zeros of the Dirichlet L-functions $L(s,\chi)$. As with the zeros of $\zeta(s)$, the major input is to show that $L(s,\chi)$ satisfies a suitable functional equation. To do this cleanly, we require a notion of what it means for a character to be primitive.

If the homomorphism $\chi\colon \mathbb{Z}/q\mathbb{Z}^\times \to \mathbb{C}^\times$ factors through $\mathbb{Z}/q_1\mathbb{Z}^\times$ for some proper divisor q_1 of q, i.e. there is some character $\chi_1 \pmod{q_1}$ such that $\chi(n) = \chi_1(n)$ whenever $(n,q) = (n,q_1) = 1$, then we say that χ is *induced* by χ_1. If there is no such character χ_1, then we say that χ is primitive $\pmod q$.

Example. The character $\chi_4 \pmod 4$ from the previous section is a primitive character $\pmod 4$. However, χ_4 also defines a valid Dirichlet character $\pmod 8$, which is not primitive, even though it is identical as a function. Similarly, the character $\chi' \pmod{12}$ defined by $\chi'(n) =$

$\chi_4(n)$ if $\gcd(3,n)=1$ and $\chi'(n)=0$ if $\gcd(3,n)>1$ is a valid Dirichlet character $(\mathrm{mod}\, 12)$, but it is not primitive $(\mathrm{mod}\, 12)$. Notice, however, that

$$L(s,\chi') = \prod_p \left(1-\frac{\chi'(p)}{p^s}\right)^{-1} = \prod_{p\neq 3}\left(1-\frac{\chi_4(p)}{p^s}\right)^{-1} = \left(1+\frac{1}{3^s}\right)L(s,\chi_4),$$

so that all of the interesting analytic properties of $L(s,\chi')$ are controlled by those of $L(s,\chi_4)$.

Given a primitive character χ, we define the *Gauss sum*

$$\tau(\chi) = \sum_{a(\mathrm{mod}\, q)} \chi(a)e^{2\pi i a/q},$$

which measures the correlation between the character χ of the multiplicative group $\mathbb{Z}/q\mathbb{Z}^\times$ and the character $e^{2\pi i\cdot/q}$ of the additive group $\mathbb{Z}/q\mathbb{Z}$. It is a classical exercise to show that $|\tau(\chi)| = q^{1/2}$.

Theorem 7.4 (Functional equation for $L(s,\chi)$)**.** *Let $\chi\,(\mathrm{mod}\, q)$ be a nontrivial primitive character, and set $\mathfrak{a}=0$ if $\chi(-1)=1$ and $\mathfrak{a}=1$ if $\chi(-1)=-1$. Define the completed L-function $\xi(s,\chi)$ by*

$$\xi(s,\chi) = (q/\pi)^{(s+\mathfrak{a})/2}\Gamma\left(\frac{s+\mathfrak{a}}{2}\right)L(s,\chi).$$

Then $\xi(s,\chi)$ is an everywhere holomorphic function satisfying

$$\xi(1-s,\chi) = \frac{i^{\mathfrak{a}}q^{1/2}}{\tau(\overline{\chi})}\xi(s,\overline{\chi}).$$

Corollary 7.5 (Trivial zeros)**.** *If χ is a primitive Dirichlet character $(\mathrm{mod}\, q)$, then $L(s,\chi)\neq 0$ in the region $\Re(s)\geq 1$. In the region $\Re(s)\leq 0$, $L(s,\chi)\neq 0$ except when $\chi(-1)=1$ and $s\leq 0$ is an even integer, and when $\chi(-1)=-1$ and $s\leq 0$ is an odd integer.*

Proof. Non-vanishing in the region $\Re(s)>1$ follows from the absolute convergence of the Euler product for $L(s,\chi)$, while non-vanishing on the line $\Re(s)=1$ is the content of Theorems 6.6 and 6.7. The claim about the region $\Re(s)\leq 0$ follows upon recalling that $\Gamma(s)$ is everywhere non-vanishing and has poles exactly at the non-positive integers. □

As with $\zeta(s)$, beyond what's stated by Corollary 7.5, we expect quite a bit more to be true, namely that all zeros lie on the line of symmetry of the functional equation, $\Re(s)=1/2$.

Conjecture 7.6 (The generalized Riemann hypothesis)**.** *If ρ_χ is a non-trivial zero of $L(s,\chi)$ for some primitive character $\chi\,(\mathrm{mod}\, q)$, then $\Re(\rho_\chi)=1/2$.*

As with the prime number theorem, while GRH provides quite a bit of control over the sum of zeros in the explicit formula, we also need to know that they don't accumulate too rapidly as we progress up the line $\Re(s)=1/2$ vertically.

Theorem 7.7. *Let $\chi\,(\mathrm{mod}\, q)$ be a primitive character, and define*

$$N(T;\chi) := \#\{\rho_\chi \text{ non-trivial} : |\Im(\rho_\chi)|\leq T\}.$$

Then for $T\geq 1$,

$$N(T;\chi) = \frac{T}{\pi}\log\left(\frac{qT}{2\pi e}\right) + O\left(\log qT\right).$$

With all these results in hand, it is now possible to prove the (rather cumbersomely titled) prime number theorem for primes in arithmetic progressions.

Theorem 7.8. *Let* $\gcd(a, q) = 1$. *Assume the generalized Riemann hypothesis. For* $x \geq q$,

$$\psi(x; q, a) = \frac{x}{\phi(q)} + O\left(x^{1/2}(\log x)^2\right),$$

and thereby

$$\pi(x; q, a) = \frac{\mathrm{Li}(x)}{\phi(q)} + O\left(x^{1/2} \log x\right).$$

Remark. The step of passing from $\psi(x; q, a)$ to $\pi(x; q, a)$ is innocuous and is typically breezed past in courses on analytic number theory. However, it turns out to be the source of an interesting phenomenon known as Chebyshev's bias. Chebyshev observed that, even though the primes $(\mathrm{mod}\, 4)$ are evenly distributed (as follows from Theorem 7.8, though in Chebyshev's time this was known only in the form of Dirichlet's theorem), there nevertheless seems to be a decided preference for the class $3\,(\mathrm{mod}\, 4)$ over $1\,(\mathrm{mod}\, 4)$; see Table 1.

x	$\pi(x)$	$\pi(x; 4, 1)$	$\pi(x; 4, 3)$
7919	10^3	495	504
104729	10^4	4984	5015
$1.299 \cdot 10^6$	10^5	49949	50050
$1.548 \cdot 10^7$	10^6	499798	500201
$1.794 \cdot 10^8$	10^7	4999452	5000547
$2.038 \cdot 10^9$	10^8	49998748	50001251
$2.280 \cdot 10^{10}$	10^9	499996831	500003168

TABLE 1. Chebyshev's bias $(\mathrm{mod}\, 4)$.

We hope that the reader observes a few things from Table 1. First, it is possible to witness the error term $O(x^{1/2} \log x)$ from Theorem 7.8, in that roughly the first half of the digits are correct and roughly the second half are noise. Second, within that noise, we observe a preference toward primes that are $3\,(\mathrm{mod}\, 4)$. This does not hold for all x – for example, $\pi(26861; 4, 1) > \pi(26861; 4, 3)$ – but it is now known that, assuming well-supported but far out of reach conjectures, it is true about 99.59% of x.[1] Obtaining this percentage is beyond the scope of this course, but an answer to the cruder yet also more important question of why there is any preference at all is within reach. Amazingly, it all boils down to the step of passing from $\psi(x; q, a)$ to $\pi(x; q, a)$.

From the explicit formula, it follows that $\psi(x; q, a)$ is unbiased amongst the admissible classes $a\,(\mathrm{mod}\, q)$: the zeros of $L(s, \chi)$ are complex conjugates to those of $L(s, \overline{\chi})$, and all classes $a\,(\mathrm{mod}\, q)$ are treated equally by Corollary 7.3. To pass from $\psi(x; q, a)$ to $\pi(x; q, a)$, we define

$$\theta(x; q, a) = \sum_{\substack{p \leq x: \\ p \equiv a (\mathrm{mod}\, q)}} \log p,$$

[1] This follows from work of Rubinstein and Sarnak in Rubinstein's undergraduate thesis. In the author's opinion, this is among the best undergraduate theses of all time, having given birth to the study of "prime number races," a vast generalization of Chebyshev's bias.

and note that

$$\begin{aligned}\psi(x;q,a)-\theta(x;q,a) &= \sum_{\substack{p^k\le x:\\ p^k\equiv a(\mathrm{mod}\,q)\\ k\ge 2}} \log p\\ &= \sum_{\substack{p\le x^{1/2}:\\ p^2\equiv a(\mathrm{mod}\,q)}} \log p + \sum_{\substack{p^k\le x:\\ p^k\equiv a(\mathrm{mod}\,q)\\ k\ge 3}} \log p\\ &= \sum_{\substack{p\le x^{1/2}:\\ p^2\equiv a(\mathrm{mod}\,q)}} \log p + O(\psi(x^{1/3})).\end{aligned}$$

By the prime number theorem, the error term is $O(x^{1/3})$, which washes out in comparison to the $O(x^{1/2})$ error coming from the zeros of the various $L(s,\chi)$. On the other hand, *the first term will only exist* if the residue class $a\,(\mathrm{mod}\,q)$ is a square, in which case it will be of size about $x^{1/2}$. More specifically, using Theorem 7.8 we find

$$\theta(x;q,a)=\psi(x;q,a)-\frac{\#\{b\,(\mathrm{mod}\,q):b^2\equiv a\,(\mathrm{mod}\,q)\}}{\phi(q)}x^{1/2}+O(x^{1/3}),$$

so that $\theta(x;q,a)$ has a slight negative bias depending on whether $a\,(\mathrm{mod}\,q)$ is a square. However, notice that this bias is of the same order of magnitude as the contribution from a single zero ρ_χ. Since the sum over zeros ρ_χ is not absolutely convergent, for a suitable but rare value of x, the contribution from all the zeros will line up in such a way to overpower the negative bias coming from whether $a\,(\mathrm{mod}\,q)$ is a square. This creates the 0.41% chance that $\theta(x;4,1)>\theta(x;4,3)$. Moreover, the process

$$\pi(x;q,a)=\int_{2^-}^{x}\frac{1}{\log t}\,d\theta(x;q,a)$$

of passing from $\theta(x;q,a)$ to $\pi(x;q,a)$ is unbiased, so exactly the same conclusions hold for $\pi(x;q,a)$ as for $\theta(x;q,a)$.