Dirichlet's theorem on primes in an arithmetic progression

Jeremy Avigad
(joint work with Rebecca Morris)

Department of Philosophy and
Department of Mathematical Sciences
Carnegie Mellon University

July, 2013

Dirichlet's theorem

Theorem

Let m and k be relatively prime. Then the arithmetic progression $m, m+k, m+2k, \ldots$ contains infinitely many primes.

For example, there are no primes in the sequence

$$6, 15, 24, 33, 42, 51, \ldots .$$

There are infinitely many primes in the sequence

$$5, 14, 23, 32, 41, 50, \ldots$$

Dirichlet's theorem

Legendre assumed this in 1798, in giving a purported proof of the law of quadratic reciprocity.

Gauss pointed out this gap, and presented two proofs of quadratic reciprocity in his *Disquisitiones Arithmeticae* of 1801.

He ultimately published six proofs of quadratic reciprocity, and left two more proofs in his *Nachlass*. But he never proved the theorem on primes in an arithmetic progression.

Dirichlet's 1837 proof is notable for the sophisticated use of analytic methods to prove a number-theoretic statement.

Euler's theorem

The starting point for Dirichlet's proof:

Theorem

The series $\sum_q \frac{1}{q}$ diverges, where q ranges over the prime numbers.

In particular, there are infinitely many primes.

Euler's proof

The Euler product formula: for $s > 1$,

$$\begin{aligned}\sum_n \frac{1}{n^s} &= \prod_q (1 + \frac{1}{q^s} + \frac{1}{q^{2s}} + \ldots) \\ &= \prod_q \left(1 - \frac{1}{q^s}\right)^{-1}\end{aligned}$$

Take logarithms of both sides:

$$\begin{aligned}\log \sum_n \frac{1}{n^s} &= -\sum_q \log(1 - \frac{1}{q^s}) \\ &= \sum_q \frac{1}{q^s} + O(1).\end{aligned}$$

As $s \to 1$, LHS $\to \infty$, so $\sum_q \frac{1}{q^s} \to \infty$.

Dirichlet's idea

Fix m and k relatively prime, and try to show

$$\sum_{q \equiv m \text{ mod } k} \frac{1}{q}$$

diverges.

Write this as

$$\sum_{q} \frac{1_{m,k}(q)}{q}$$

where

$$1_{m,k}(q) = \begin{cases} 1 & \text{if } q \equiv m \text{ mod } k \\ 0 & \text{otherwise} \end{cases}$$

and try to repeat the Euler argument.

Dirichlet's idea

The sticking point: the Euler product formula

$$\sum_n \frac{\chi(n)}{n^s} = \prod_q \left(1 - \frac{\chi(q)}{q^s}\right)^{-1}$$

only holds if χ is *completely multiplicative*:

$$\chi(nn') = \chi(n)\chi(n')$$

The solution: decompose

$$1_{m,k}(n) = a_1\chi_1(n) + \ldots + a_u\chi_u(n),$$

a linear combination of such functions.

Group characters

If G is a finite abelian group, χ is a *character on G* if it is a homomorphism from G to the nonzero complex numbers, i.e.

$$\chi(g_1 g_2) = \chi(g_1)\chi(g_2)$$

for every g_1 and g_2 in G.

There is always a trivial character, $\chi_0(g) \equiv 1$.

The set of characters on G forms a group $\hat{G}$ with pointwise multiplication and identity χ_0. In fact, $|G| = |\hat{G}|$.

Group characters

The following two “orthogonality” relations hold:

$$\sum_{g\in G}\chi(g)=\begin{cases}|G| & \text{if } \chi=\chi_0\\ 0 & \text{otherwise}\end{cases}$$

and

$$\sum_{\chi\in\hat{G}}\chi(g)=\begin{cases}|G| & \text{if } g=1\\ 0 & \text{otherwise}\end{cases}$$

This makes it possible to do “finite Fourier analysis”: if $\hat{f}(\chi)=\sum_g f(g)\chi(g)$, then $f=\frac{1}{|G|}\sum_\chi \hat{f}(\chi)\chi$.

Dirichlet's theorem

Fix k, and "lift" the characters on $(\mathbb{Z}/k\mathbb{Z})^*$ to functions on $\mathbb{N}$.

Define

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

Euler product expansion:

$$L(s,\chi) = \prod_{q} \left(1 - \frac{\chi(q)}{q^s}\right)^{-1} = \prod_{q \nmid k} \left(1 - \frac{\chi(p)}{p^s}\right)^{-1}$$

This converges when $Re(s) > 1$.

Dirichlet's theorem

Taking logarithms of both sides yields

$$\log L(s, \chi) = \sum_{q \nmid k} \frac{\chi(q)}{q^s} + O(1).$$

Multiply both sides by $\overline{\chi(m)}$ and sum over χ.

$$\sum_{\chi} \overline{\chi(m)} \log L(s, \chi) = \sum_{\chi} \sum_{q \nmid k} \overline{\chi(m)} \frac{\chi(q)}{q^s} + O(1).$$

Using the orthogonality relations,

$$\sum_{\chi} \overline{\chi(m)} \log L(s, \chi) = \varphi(m) \sum_{q \equiv m \bmod k} \frac{1}{q^s} + O(1).$$

Dirichlet's theorem

Let $s \to 1$ from above.

$$\sum_{\chi} \overline{\chi(m)} \log L(s, \chi) = \varphi(m) \sum_{q \equiv m \bmod k} \frac{1}{q^s} + O(1).$$

Divide the characters into three types:

1. The trivial character, χ_0.
2. The nontrivial real-valued characters.
3. The (properly) complex characters.

Show:

- $L(s, \chi_0)$ has a simple pole at $s = 1$.
- For $\chi \neq \chi_0$, $L(s, \chi)$ has a nonzero limit at $s \to 1$.

This yields the result.

Functions as objects

Nineteenth century methodological changes:

1. Unification / generalization of the function concept
2. Liberalization of the function concept
3. Extensionalization of the function concept
4. Reification of the function concept

Functions as objects

“Reification of the function concept” is vague. Roughly, I mean treating them on par with objects like the natural numbers.

Some aspects:

1. Treating functions extensionally, independent of representations.
2. Sending functions as arguments to other functions, $F(f)$.
3. Forming sets of functions, groups of functions, spaces of functions.
4. Quantifying over functions (in definitions, in theorems).
5. Summing over functions.

Functions as objects

In the modern presentation of Dirichlet's theorem:

- The notion of a character is defined.
- One determines some of their properties.
- Characters appear as arguments to other functions ($L(s, \chi)$).
- One sums over sets of characters ($\sum_{\chi} \ldots$), without having representations for any particular one.
- One carries out proofs (in fact, one has to!) without making reference to any particular representation.
- One characterizes sets of characters extensionally.
- Characters bear a group structure.

These are the main points of contrast with the historical sources.

Outline of this talk

- Contemporary proofs of Dirichlet's theorem
- Functions as objects
- Dirichlet's proof
- Subsequent presentations, from Dedekind to Landau
- Reflections on mathematical method

Dirichlet 1837

Dirichlet did not introduce a notation for characters. Rather, he used explicit expressions.

In the case where the common difference is a prime, p:

- Let c be a primitive element modulo p.
- For every n coprime to p, let γ_n be such that $c^{\gamma_n} \equiv n \bmod p$.
- Characters χ correspond to $p-1$st roots of unity ω, where $\chi(n) = \omega^{\gamma_n}$.
- Dirichlet writes ω^{γ_n} where we would write $\chi(n)$.

Pick a generator Ω of the $p-1$st roots of unity, $\{\Omega^0, \ldots, \Omega^{p-2}\}$.

L_m is the L-series corresponding to the root Ω^m. Dirichlet summed over m, rather than χ.

Dirichlet 1837

After demonstrating the Euler product formula,

$$\prod \frac{1}{1-\omega^{\gamma}\frac{1}{q^s}} = \sum \omega^{\gamma}\frac{1}{n^s} = L,$$

Dirichlet wrote:

> *The equation just found represents $p-1$ different equations that result if we put for ω its $p-1$ values. It is known that these $p-1$ different values can be written using powers of the same Ω when it is chosen correctly, to wit:*
>
> $$\Omega^0,\ \Omega^1,\ \Omega^2,\ \ldots,\ \Omega^{p-2}$$
>
> *According to this notation, we will write the different values L of the series or product as:*
>
> $$L_0,\ L_1,\ L_2,\ \ldots,\ L_{p-2}$$

Dirichlet 1837

In the case where the modulus k is not prime:

- Decompose $(\mathbb{Z}/k\mathbb{Z})^*$ into a product of cyclic groups.
- Choose generators for each cyclic group.
- A number n modulo k has indices α_n, β_n, γ_n, $\gamma'_n, \ldots$
- Each character corresponds to a choice of roots of unity, $\theta, \varphi, \omega, \omega', \ldots$
- Dirichlet writes $\theta^\alpha \varphi^\beta \omega^\gamma \omega'^{\gamma'} \ldots$ where we would write $\chi(n)$.

Notice that the dependence on n is left implicit.

Moreover, as before, if we choose appropriate primitive roots of unity, each character is given by a list of indices $a, b, c, c', \ldots$.

Thus Dirichlet wrote $L_{a,b,c,c',\ldots}$ in “a comfortable way” where we would write $L(s, \chi)$.

Dirichlet 1837

Summing over characters: in the case where k is prime, Dirichlet wrote

$$\log L_0 + \Omega^{-\gamma_m} \log L_1 + \Omega^{-2\gamma_m} \log L_2 + \ldots + \Omega^{-(p-2)\gamma_m} \log L_{p-2}$$

where we would write $\sum_\chi \overline{\chi(m)} \log L(s,\chi)$.

For composite k, he wrote

$$\sum \Theta^{-\alpha_m a}\ \Phi^{-\beta_m b} \Omega^{-\gamma_m c} \Omega^{-\gamma_{m'} c'} \ldots \log L_{a,b,c,c' \ldots}$$

where the sum is over all combinations of $a, b, c, c' \ldots$.

Dirichlet 1837

Dirichlet divided the L functions into three classes:

- the one in which all the roots are 1
- the ones in which all the roots are real (± 1)
- those in which at least one of the roots is not real

This is an *intensional* characterization.

Summary:

- Dirichlet did not name or identify "characters."
- The L functions depend on a tuple of natural numbers ($L_{a,b,c,c'\ldots}$ rather than $L(s,\chi)$)
- Instead of summing over L functions, he summed over these tuples.
- The L functions are classified intensionally.

A timeline

- Dirichlet 1837: Dirichlet's original proof
- Dirichlet 1840, 1841: extensions to Gaussian integers, quadratic forms
- Dedekind 1863: presention of Dirichlet's theorem
- Dedekind 1879, Weber 1882: characters on arbitrary abelian groups
- Hadamard 1896: presentation of Dirichlet's theorem and extensions
- de la Vallée Poussin 1897: presentation of Dirichlet's theorem and extensions
- Kronecker (1901, really 1870's and 1880's): constructive, quantitative treatment
- Landau 1909, 1927: presentation of Dirichlet's theorem and extensions

Identifying characters

In 1841, Dirichlet considered expressions

$$\Omega_n = \varphi^{\alpha_n} {\varphi'}^{\alpha'_n} \times \ldots \times \psi^{\beta_n} \chi^{\gamma_n} {\psi'}^{\beta'_n} {\chi'}^{\gamma'_n} \times \ldots \times \theta^{\delta_n} \eta^{\varepsilon_n}$$

analogous to the characters in his 1837 proof.

He isolated their key properties:

1. $\Omega_{nn'} = \Omega_n \Omega_{n'}$.
2. $\Omega_{n'} = \Omega_n$ whenever $n' \equiv n \pmod{k}$.
3. the first orthogonality lemma
4. the second orthogonality lemma

Identifying characters

In 1863, Dedekind explained that the Euler product formula holds, in general, for multiplicative functions.

He also used χ to denote values of the characters. But he used Dirichlet's explicit representations in the proof itself.

In 1882, Weber gave the general definition of a character of an abelian group, and proved the general properties.

In 1909, Landau emphasized that the four "key properties" of characters are all that is needed in the proof of Dirichlet's theorem.

Summing over characters

Dirichlet originally summed over representing data: $\sum_{a,b,c,c',\dots}$. Dedekind did this as well.

In 1841, Dirichlet introduced a special notation S_Ω for summing over the characters. Later, de la Vallée Poussin used S_χ.

Hadamard numbered the characters $\psi_1, \psi_2, \dots, \psi_M$ and summed over the indices, $\sum_v \dots$. Landau does this too in 1909.

In 1927, Landau wrote $\sum_\chi \dots$.

Characters as arguments

Recall that Dirichlet wrote $L_{a,b,c,c',\dots}$ where we would write $L(s,\chi)$.

Hadamard could write L_v for the series corresponding to ψ_v.

In 1897, de la Vallée Poussin wrote $Z(s,\chi)$. Subsequent authors wrote $L(s,\chi)$.

Classifying characters and L-series

Extensional classification of characters:

- the character with constant value 1
- the (other) real-valued characters
- the (other) complex-valued characters

This yields three classes of L-series $L(s, \chi)$.

Most authors favored an *intensional* classification, in terms of the roots used in the defining expressions.

Some authors gave both.

Objects vs. representations

After deriving a key identity parametrized by the characters, Dirichlet wrote:

> *The general equation, in which the different roots $\theta, \varphi, \omega, \omega', \ldots$ can be combined with one another arbitrarily, clearly contains K-many particular equations.*

In a similar context, Dedekind wrote in 1863:

> *Since these roots can have $a, b, c, c', \ldots$ values, respectively, the form L contains altogether $abcc' \ldots = \varphi(k)$ different particular series...*

Objects vs. representations

Here is Weber in 1882:

> *Each of the formulas ... represents h different formulas, corresponding to the h different characters* $\chi_1, \chi_2, \ldots, \chi_h$.

And de la Vallée Poussin in 1897:

> *... this equation (E) represents in reality* $\varphi(M)$ *distinct ones, which result from exchanging the characters amongst themselves.*

Summary

Over time:

- The notion of a character was defined.
- Authors isolated general properties of characters.
- Authors got used to summing over characters, rather than representing data.
- Authors got used to functional dependences on characters, rather than representing data.
- Authors began to adopt extensional characterizations and classifications of characters.
- The use of explicit symbolic representations for the characters diminished and was ultimately eliminated.

Outline of this talk

- Contemporary proofs of Dirichlet's theorem
- Functions as objects
- Dirichlet's proof
- Subsequent presentations, from Dedekind to Landau
- Reflections on mathematical method

Reflections on mathematical method

Treating functions as objects brings benefits.

- Expressions are simplified.
- Proofs become modular.
- The reader has to keep track of less information when parsing expressions.
- The reader has to keep track of less information when reading a proof.
- The relevant data and relations are made more salient.
- Lemmas and definitions can be reused elsewhere.
- Lemmas and definitions can be modified and adapted.
- Abstraction leads to greater generality.

Reflections on mathematical method

To summarize:

- Dependencies between components are minimized.
- The mathematics become easier to understand.
- It becomes easier to ensure correctness.
- Components are adaptable, reusable, and generalizable.
- Proofs can be modified and varied more easily.

These are exactly the benefits associated with modularity in software engineering.

Reflections on mathematical method

Why did it take so long to arrive at the contemporary treatment of functions?

Reading mathematics involves a good deal of tacit knowledge.

When I publish a proof, my intention is that you will read it, understand it, and accept it as correct.

Reflections on mathematical method

Concerns raised by any methodological expansion:

- Do concepts and notations come with clear rules of use?
- Are they appropriate to the mathematics?
  - Are they meaningful?
  - Are they useful?
  - Do they answer the questions we have asked?
- Are they reliable?

Changes to the practice have to be accepted by the *community*.

Reflections on mathematical method

Treating sets and functions as objects like numbers was a dramatic change.

It affected fundamental aspects of mathematical language and method.

Mathematical change is best understood in terms of weighing very pragmatic benefits against very pragmatic concerns.

Addendum

Functions in the nineteenth century

Nineteenth century instances of the function concept:

1. Functions defined on the continuum ($\mathbb{R}$ to $\mathbb{R}$, $\mathbb{C}$ to $\mathbb{C}$)
2. Sequences and series ($\mathbb{N}$ to $\mathbb{R}$ or $\mathbb{Q}$)
3. Number theoretic functions ($\mathbb{N}$ to $\mathbb{N}$)
4. Transformations of the plane
5. Permutations of a finite set A (bijections from A to A)
6. Characters ($\mathbb{Z}$ to $\mathbb{C}$, or $(\mathbb{Z}/m\mathbb{Z})^*$ to $\mathbb{C}$)
7. Arbitrary mappings, or correspondences, between domains

Functions in the nineteenth century

Some landmarks:

- In 1850, Eisenstein explicitly introduced the term "zahlentheoretische Funktion."
- Dedekind 1854: "Über die Einführung neuer Funktionen in der Mathematik; Habilitationsvortrag"
- In 1879, in the third edition of the *Vorlesungen*, Dedekind refers to characters on the class groups as functions.
- In 1879, in the *Begriffsschrift*, Frege introduces a very general notion of function.
- In 1888, Dedekind considers arbitrary mappings (*Abbildung*) between domains.

Functions in the nineteenth century

> *Since, with the concept of a function, one moved away from the necessity of having an analytic construction, and began to take its essence to be a tabular collection of values associated to the values of one or several variables, it became possible to take the concept to include functions which, due to conditions of an arithmetic nature, have a determinate sense only when the variables occurring in them have integral values, or only for certain value-combinations arising from the natural number series. For intermediate values, such functions remain indeterminate and arbitrary, or without any meaning.*

(Eisenstein, 1850)

Functions in the nineteenth century

> *. . . the function $\chi(\mathfrak{a})$ also posseses the property that it takes the same value on all ideals $\mathfrak{a}$ belonging to the same class A; this value is therefore appropriately denoted by $\chi(A)$ and is clearly always an hth root of unity. Such functions χ, which in an extended sense can be termed characters, always exist; and indeed it follows easily from the theorems mentioned at the conclusion of §149 that the class number h is also the number of all distinct characters $\chi_1, \chi_2, \ldots, \chi_h$ and that every class A is completely characterized, i.e. is distinguished from all other classes, by the h values $\chi_1(A), \chi_2(A), \ldots, \chi_h(A)$.*

(Dedekind 1879)

Functions in the nineteenth century

> *If, in an expression (whose content need not be a judgeable content), a simple or complex symbol occurs in one or more places, and we think of it as replaceable at all or some of its occurrences by another symbol (but everywhere by the same symbol), then we call the part of the expression that on this occasion appears invariant the function, and the replaceable part its argument.*

> *One sees here particularly clearly that the concept of function in Analysis, which in general I have followed, is far more restricted than that developed here.*

(Frege 1879)