THE SATO-TATE CONJECTURE: INTRODUCTION TO THE PROOF

Michael Harris

Introduction

An elliptic curve is the set E of solutions of a cubic curve in two variables, for example

$$E: \ y^2 + y = x^3 + x,$$

I will only consider elliptic curves with rational coefficients, which after a change of variables can be written

$$y^2 = x^3 + Ax + B$$

These are not all distinct, and one can isolate two invariants: the *discriminant*

$$\Delta = \Delta_E - 16(4A^3 + 27B^2)$$

which is not really an invariant of E, but which has the following property: if $\Delta_E \neq 0$ then E is non-singular, which we always assume. There is also the j-invariant, which really depends on E and not just on the equation:

$$j(E) = 1728 \cdot \frac{4A^3}{4A^3 + 27B^2}$$

which determines E up to isomorphism over an algebraically closed field.

We may assume $A, B \in \mathbb{Z}$. It then makes sense to reduce the equation modulo a prime p and ask how many solutions E has modulo p:

$$N_p = N_p(E) = |E(\bar{\mathbb{F}}_p)|.$$

Suppose for the moment we replace E by a line L, given by a *linear* equation

$$L: \ y = ax + b.$$

Then the number of solutions of L in the plane $\bar{\mathbb{F}}_p^2$ obviously equals p, to which we add 1 for the point at infinity:

It turns out that $p+1$ is in a natural sense the optimal number of points for a curve of any genus (or degree). Skipping over quadric curves, we define an integer $a_p = a_p(E)$, for each prime p, by

$$N_p(E) = p + 1 - a_p(E).$$

We only consider p for which E remains nonsingular modulo p, which is somewhat weaker than the condition that $\Delta_E \neq 0 \pmod{p}$. Such a p is called a *prime of good reduction*. One can consider the beginning of arithmetic algebraic geometry to be Hasse's discovery that

$$|a_p| \leq 2\sqrt{p}$$

for any prime of good reduction.

In other words, $p+1$ is a good approximation to N_p to square-root order. This can be compared to the square-root good approximation to $\pi(x)$, the number of primes less than x:

$$\pi(x) = \int_2^x \frac{dx}{logx} + \text{ Error}(x)$$

where the Riemann hypothesis is the assertion that

$$\text{Error}(x) = O(x^{\frac{1}{2}})$$

and indeed Hasse's theorem was generalized by Weil to a version of the Riemann hypothesis valid for all curves over finite fields.

The next question is whether anything can be said about the behavior of the $a_p(E)$ as p varies. Is $a_p(E)$ more likely to be positive or negative? Is it more likely to cluster around 0 or around $\pm 2\sqrt{p}$? The rough answer is that it is as random as possible, but it is not immediately obvious how to make sense of this. We normalize all the a_p simultaneously to allow them to be compared:

$$a_p^{norm}(E) = \frac{1}{2\sqrt{p}} a_p(E) \in [-1, 1].$$

Thus there is a unique $\theta_p = \theta_p(E) \in [0, \pi]$ such that $a_p^{norm}(E) = cos(\theta_p)$. We ask about the distribution of the a_p^{norm} in $[-1, 1]$, or equivalently of the $\theta_p \in [0, \pi]$. Over forty years ago, Sato and Tate independently formulated the following conjecture:

Sato-Tate Conjecture. *Suppose E has no complex multiplication. Then the $a_p^{norm}(E)$ (resp. the θ_p) are equidistributed in $[-1, 1]$ (resp. $[0, \pi]$) with respect to the probability measure*

$$\frac{2}{\pi}\sqrt{1-t^2}dt \quad (\textit{ resp. } \frac{2}{\pi}\sin^2(\theta)d\theta \ ) \ .$$

Theorem (L. Clozel, MH, N. Shepherd-Barron, R. Taylor). *Suppose $j(E)$ is not an integer. Then the Sato-Tate Conjecture is valid for E.*

I have read that Sato arrived at his conjecture by experiment, whereas Tate formulated his conjecture as part of his general understanding of algebraic cycles in terms of zeroes and poles of L-functions. Serre explained in his 1968 notes on elliptic curves how to derive the Sato-Tate conjecture from the sort of Tauberian theorems that are also used to prove the prime number theorem. The main input is the following. One needs to rewrite the expression for $N_p(E)$:

$$N_p = (1 - \sqrt{p}e^{i\theta_p})(1 - \sqrt{p}e^{-i\theta_p}) = (1 - \alpha_p)(1 - \beta_p)$$

and to define more generally

$$L_p(s, E) = [(1 - \alpha_p p^{-s})(1 - \beta_p p^{-s})]^{-1};\ N_p(E) = L_p(0, E)^{-1}$$

Let S be the set of primes of bad reduction for E, and define

$$L(s, E) = \prod_{p \notin S} L_p(s, E) \times \prod_{p \in S} L_p(s, E)$$

where the factors $L_p(s, E)$ for $p \in S$ are ultimately simpler than for primes of good reduction.

It follows from Hasse's estimate that $L(s, E)$ converges absolutely for $Re(s) > \frac{3}{2}$. Now one can attach a very similar Dirichlet series to a holomorphic modular (cusp) form f of weight 2, and it is known that $L(s, f)$ extends to an entire analytic function that satisfies a functional equation relating $L(s, f)$ to $L(2 - s, f)$, and such that $L(s, f) \neq 0$ for $Re(s) \geq \frac{3}{2}$. The most striking development of number theory in recent years was Wiles' discovery of a technique for proving that any $L(s, E)$ is also an $L(s, f)$. With help from Taylor, Wiles applied this technique to a sufficiently large family of E to prove Fermat's Last Theorem. A few years later, Taylor, together with Breuil, Conrad, and Diamond, proved that every $L(s, E)$ is an $L(s, f)$. In particular,

Theorem [BCDT]. *L(s,E) extends to an entire analytic function with no zeroes on the half-plane $Re(s) \geq \frac{3}{2}$.*

From the data determined by the $N_p(E)$ one can construct an infinite family of L-functions. For each $n \geq 0$, define

$$L_p(s, E, Sym^n) = [\prod_{j=0}^{n} (1 - \alpha_p^j \beta_p^{n-j} p^{-s})]^{-1}$$

if $p \notin S$; there is again a definition for $p \in S$. We define

$$L(s, E, Sym^n) = \prod L_p(s, E, Sym^n),$$

Theorem (Serre). *Suppose E is an elliptic curve and, for all $n > 0$ $L(s, E, Sym^n)$ extends to a meromorphic function which is holomorphic and non-vanishing for $Re(s) \geq 1 + \frac{n}{2}$. Then the Sato-Tate Conjecture holds for E.*

The results of the three papers [CHT], [HST], and [T] together imply that

Theorem (Clozel, MH, Shepherd-Barron, Taylor). *Suppose E is an elliptic curve over $\mathbb{Q}$ with non-integral j-invariant. Then for all $n > 0$, $L(s, E, Sym^n)$ extends to a meromorphic function which is holomorphic and non-vanishing for $Re(s) \geq 1 + \frac{n}{2}$.*

The Langlands conjectures predict that $L(s, E, Sym^n)$ can be associated to a cuspidal automorphic representation of $GL(n+1)_{\mathbb{Q}}$. This would imply that $L(s, E, Sym^n)$ is entire (Godement-Jacquet) and is non-vanishing on the indicated domain. We do not prove this. Instead, we prove that for n odd, $L(s, E, Sym^n)$ is *potentially automorphic*; that is, it is associated to a cuspidal automorphic representation of $GL(n+1)$ over some totally real Galois extension of $\mathbb{Q}$. This argument involves two parts. The first is an extension of Wiles' technique for identifying L-functions of elliptic curves with L-functions of modular forms, and is based essentially on Galois cohomology and an analysis of automorphic representations of different sorts of groups, especially unitary groups. This is begun in [CHT] and completed in [T]. The second is an extension of an idea of Taylor for proving meromorphic continuation of L-functions attached to two-dimensional Galois representations, using weak approximation on moduli spaces. With Shepherd-Barron, Taylor and I found a moduli space that could be used to study n-dimensional Galois representations for any even n; it is a twisted form of the moduli space of certain Calabi-Yau varieties originally studied by Dwork in certain cases, and more generally by physicists interested in mirror symmetry.