The Sato–Tate Conjecture

Nir Elber

Fall 2024

Abstract

This is an expository note for two talks on the Sato–Tate conjecture. In the first talk, we will state the Sato–Tate conjecture for abelian varieties. In the second talk, we will discuss some results related to Sato–Tate groups of the so-called Fermat curves $x^m + y^m + z^m = 0$.

Contents

1 Stating the Sato–Tate Conjecture

In this first talk, we will work our way towards stating the Sato–Tate conjecture for abelian varieties. We will gradually increase the amount of assumed background.

1.1 Quadratic Equations

In order to get a dishonest taste for the sort of results we are after, we will begin with the case of quadratic equations. Fix a monic quadratic polynomial $f(x) = x^2 + ax + b$ with nonzero discriminant $a^2 - 4b$, and consider the curve

$$C\colon y^2 = f(x).$$

Technically speaking, we understand C_f to cut out a smooth projective curve in $\mathbb{P}^2_{\mathbb{Q}}$ given by homogenizing.

Example 1. For this section, it will be enough to follow the example $C\colon y^2 = x^2 + 1$ around. The polynomial $x^2 + 1$ has discriminant -4.

Because this is a talk about arithmetic geometry, we are interested in the number of points on C_f. Studying points over $\mathbb{Q}$ or $\mathbb{Z}$ is rather difficult (and will continue to get harder when we change the curve later), so we will content ourselves with studying the number of points over finite fields. Even though our curve is a priori defined over $\mathbb{Q}$, we see that $f(x) \in \mathbb{Z}[x]$ allows us to give C_f a model over $\mathbb{Z}$ given by the same equation. In this way, we can formally make sense of $C(\mathbb{F}_q)$ for any finite field $\mathbb{F}_q$.

Remark 2. In practice, one can think about $C(\mathbb{F}_q)$ as the number of pairs $(x, y) \in \mathbb{F}_q^2$ satisfying the equation given by C. The above process of finding an integral model should be thought of as some formal general procedure.

Remark 3. Our choice of integral model may not always make sense at all primes p. Notably, we would like for C_f to define a smooth curve over $\mathbb{F}_p$, but this requires $a^2 - 4b \neq 0$ (for example). (For odd primes p, smoothness is equivalent to $a^2 - 4b \neq 0$.) Over $\mathbb{Q}$, this was a hypothesis, but in general, $a^2 - 4b$ can vanish $\pmod p$ for some primes p. For example, $y^2 = x^2 + 1$ fails to be smooth over $\mathbb{F}_2$.

Let us begin with an expectation for $C(\mathbb{F}_q)$. For each $x \in \mathbb{F}_q$, it will be difficult to control the value of $f(x)$. However, about half of $\mathbb{F}_q$ has two square roots, and about half of $\mathbb{F}_q$ has no square roots, so it seems reasonably to expect that each $f(x) \in \mathbb{F}_q$ has on average one square root. Thus, we may expect that

$$C(\mathbb{F}_q) \approx q + 1,$$

where we have added 1 to include the point(s) at infinity. Let's see what we get.

Example 4. We continue with the curve $C\colon y^2 = x^2 + 1$. A computer program gives the following output.

p	3	5	7	11	13	17	19	23	29	31	37	41	43	47
$\#C(\mathbb{F}_p)$	4	6	8	12	14	18	20	24	30	32	38	42	44	48
$p+1$	4	6	8	12	14	18	20	24	30	32	38	42	44	48

Huh, it looks like our guess was pretty spot-on.

Exercise 5. For a curve $C\colon y^2 = x^2 + d$ for some $d \in \mathbb{Z}$, show that $\#C(\mathbb{F}_q) = q + 1$ for any finite field $\mathbb{F}_q$. You may find it helpful to use the substitution $(s, d) = (x + y, x - y)$.

Remark 6. Intuitively, what is going on here is that quadratics cut out genus 0 curves, which must all be isomorphic to $\mathbb{P}^1$.

1.2 Two Elliptic Curves

Let's move on to a more nontrivial example. Most of this first talk will be interested in elliptic curves, for which we pick up the following concrete definition.

Definition 7 (elliptic curve)**.** Fix a field K of characteristic not equal to 2 or 3. Then an *elliptic curve* is a curve of the form

$$E\colon y^2 = x^3 + ax + b,$$

where $a, b \in K$, and the discriminant $-4a^3 - 27b^2$ is nonzero. As usual, E is understood to cut out a smooth projective curve in $\mathbb{P}^2_K$ given by homogeneizing to $Y^2Z = X^3 + aXZ^2 + bZ^3$, so there is one point $[0 : 1 : 0]$ at infinity.

For today, all of our elliptic curves E will be defined over $\mathbb{Q}$ and pretty good integral models.

Example 8. For any lattice $\Lambda \subseteq \mathbb{C}$, it turns out that one can realize $\mathbb{C}/\Lambda$ as an elliptic curve over $\mathbb{C}$.

Example 9. We will follow around the two elliptic curves

$$E_1 : y^2 = x^3 + 1 \qquad \text{and} \qquad E_2 : y^2 = x^3 + x + 1$$

for this section. Though they look similar, these two curves have very different behavior!

As in the previous section, we note that one can frequently define a notion of $E(\mathbb{F}_q)$. Namely, the cubic equation defining E will only have finitely many denominators, and the discriminant of the cubic equation will only have finitely many prime factors; away from these primes, the equation defining E will define a perfectly reasonable elliptic curve over $\mathbb{F}_{p^r}$ for any $r \geq 1$.

Example 10. We discuss models for our elliptic curves E_1 and E_2. Note that their defining equations have no denominators.

- We see that $E_1 : y^2 = x^3 + 1$ has discriminat -27, so we get an elliptic curve over $\mathbb{F}_{p^r}$ for any $p \neq 3$.
- Similarly, the curve $E_2 : y^2 = x^3 + x + 1$ has discriminant -31, so we get an elliptic curve over $\mathbb{F}_{p^r}$ for any $p \neq 31$.

Once again, for any value of x, there is no reason to expect $x^3 + ax + b$ to be a square or not, so a reasonable expectation is for

$$E(\mathbb{F}_q) \approx q + 1.$$

Let's see what we get.

Example 11. We continue with the two elliptic curves $E_1 : y^2 = x^3 + 1$ and $E_2 : y^2 = x^3 + x + 1$.

p	5	7	11	13	17	19	23	29	37	41	43	47
$\#E_1(\mathbb{F}_p)$	6	12	12	12	18	12	24	30	48	42	36	48
$\#E_2(\mathbb{F}_p)$	9	5	14	18	18	21	28	36	48	35	34	60
$p+1$	6	8	12	14	18	20	24	30	38	42	44	48

Our guess seems to be pretty close but not quite spot-on. Let's examine the error.

Example 12. We continue with the two elliptic curves $E_1 : y^2 = x^3 + 1$ and $E_2 : y^2 = x^3 + x + 1$.

p	5	7	11	13	17	19	23	29	37	41	43	47
$\#E_1(\mathbb{F}_p) - (p+1)$	0	4	0	-2	0	-8	0	0	10	0	-8	0
$\#E_2(\mathbb{F}_p) - (p+1)$	3	-3	2	4	0	1	4	6	10	-7	-10	12

The error seems to be small, but perhaps it is difficult to quantify. We now state a theorem.

Theorem 13 (Hasse–Weil)**.** Fix an elliptic curve E defined over a finite field $\mathbb{F}_q$. Then

$$|\#E(\mathbb{F}_q) - (q+1)| \leq 2\sqrt{q}.$$

We will not prove this, but we will state a more general version later. For now, we will content ourselves with the following example.

Example 14. We continue with the two elliptic curves $E_1 : y^2 = x^3 + 1$ and $E_2 : y^2 = x^3 + x + 1$.

p	5	7	11	13	17	19	23	29	37	41	43	47
$E_1(\mathbb{F}_p) - (p+1)$	0	4	0	-2	0	-8	0	0	10	0	-8	0
$E_2(\mathbb{F}_p) - (p+1)$	3	-3	2	4	0	1	4	6	10	-7	-10	12
$\lfloor 2\sqrt{p} \rfloor$	4	5	6	7	8	8	9	10	12	12	13	13

The bound seems to hold and even come close to equality quite frequently! This motivates us to define

$$a_q(E) := \frac{\#E(\mathbb{F}_q) - (q+1)}{\sqrt{q}} \in [-2, 2].$$

Here is the table again.

Example 15. We continue with the two elliptic curves $E_1 \colon y^2 = x^3 + 1$ and $E_2 \colon y^2 = x^3 + x + 1$.

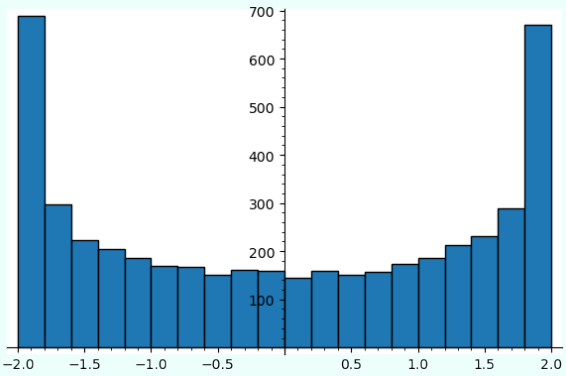
p	5	7	11	13	17	19	23	29	37	41	43	47
$a_p(E_1)$	0.00	1.51	0.00	−0.56	0.00	−1.84	0.00	0.00	1.64	0.00	−1.22	0.00
$a_p(E_2)$	1.34	−1.13	0.60	1.11	0.00	0.23	0.83	1.11	1.64	−1.09	−1.53	1.75

These numbers appear sufficiently random (aside from maybe the large number of 0s), so we can state a heuristic.

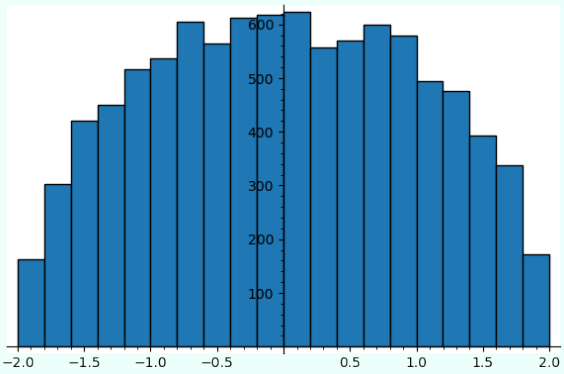
Conjecture 16 (Sato–Tate)**.** Fix an elliptic curve E defined over $\mathbb{Q}$. Then the numbers $a_p(E)$ equidistribute.

This conjecture is correct if we correctly interpret the word "equidistribute." We will spend most of the talk figuring out how to do this. For example, we do not expect the $a_p(E)$s to equidistribute in $[-2, 2]$, as the following examples show.

Example 17. Here is a histogram of the values of $a_p(E_1)$ for $p < 10^5$.



Example 18. Here is a histogram of the values of $a_p(E_2)$ for $p < 10^5$.



These are remarkable histograms! We encourage the reader to investigate other elliptic curves.

Remark 19. It turns out that "most of the time" we will get a semicircle distribution as in Example 18.

A primary goal of the talk is to explain the source of these strange curves.

1.3 The Weil Conjectures

This subsection is included for motivational purposes only and can therefore be skipped without loss of too much continuity.

It will turn out that the values of $a_p(E)$ are controlled by a Galois representation attached to E. The most concrete way to construct this Galois representation is to use the Tate module, which we will do shortly. However, for motivation, we will state the Weil conjectures and explain how they solve our problem.

Theorem 20 (Weil conjectures)**.** Fix a smooth projective variety X over a finite field $\mathbb{F}_q$ of dimension n. Then the formal power series

$$\zeta_X(T) \coloneqq \exp\left(\sum_{r=1}^{\infty} \#X\left(\mathbb{F}_{q^r}\right) \frac{T^r}{r}\right)$$

admits the following desirable properties.

(a) Rationality: one can write

$$\zeta_X(T) = \frac{P_1(T)\cdots P_{2n-1}(T)}{P_2(T)\cdots P_{2n}(T)}$$

for polynomials $P_\bullet(T) \in 1 + T\mathbb{Z}[T]$.

(b) Riemann hypothesis: the roots of the polynomial $P_\bullet(T)$ are complex numbers with roots of magnitude $q^{-\bullet/2}$.

(c) Betti numbers: suppose X is the reduction of a smooth projective variety $\mathcal{X}$ defined over a number ring $\mathcal{O}_K$. Then $\deg P_\bullet = \dim_{\mathbb{C}} H^\bullet(\mathcal{X}(\mathbb{C}), \mathbb{C})$.

We will not bother to explain why $\zeta_X(T)$ is the correct ζ-function to look at (roughly speaking, one is supposed to plug in $T \coloneqq q^{-s}$).

The mention of Betti numbers is rather compelling because it suggests that there ought to be a way to write down a cohomology theory for smooth projective varieties X. Without going into too much detail, let's explain how this is done. It is possible to define a satisfactory cohomology theory called "ℓ-adic cohomology" which takes as input an auxiliary prime ℓ which is nonzer in $\mathbb{F}_q$ and then is able to (functorialy) produce cohomology groups $H^\bullet(X, \mathbb{Q}_\ell)$ which are $\mathbb{Q}_\ell$-vector spaces. For example, in the situation of Theorem 20(c), we find that

$$\dim_{\mathbb{Q}_\ell} H^\bullet(X, \mathbb{Q}_\ell) = \dim_{\mathbb{C}} H^\bullet(\mathcal{X}(\mathbb{C}), \mathbb{C}).$$

Now, for Theorem 20, the main point is to be able to compute $\#X(\mathbb{F}_q)$. Well, the main idea is that $X(\mathbb{F}_q)$ consists of the elements of $X(\overline{\mathbb{F}_q})$ which are fixed by the Frobenius morphism $\mathrm{Frob}_q \colon \overline{\mathbb{F}_q} \to \overline{\mathbb{F}_q}$. Thus, one can use the Lefschetz trace formula on our cohomology theory $H^\bullet(X, \mathbb{Q}_\ell)$ to find

$$\#X(\mathbb{F}_q) = \sum_{i=0}^{2d} (-1)^i \operatorname{tr}\left(\mathrm{Frob}_q | H^i(X, \mathbb{Q}_\ell)\right).$$

Properties of the cohomology theory are then enough to prove the Weil conjectures purely formally.

For example, in the case of a curve C, we find that

$$\#C(\mathbb{F}_q) = q + 1 - \operatorname{tr}\left(\mathrm{Frob}_q | H^1(C, \mathbb{Q}_\ell)\right),$$

so we become interested in the Galois action on some ℓ-adic vector space $H^1(C, \mathbb{Q}_\ell)$. In the following section, we will provide a direct construction for (the dual of) $H^1(E, \mathbb{Q}_\ell)$ for an elliptic curve E.

Exercise 21. Use the above discussion to prove Theorem 13.

1.4 The Tate Module for Elliptic Curves

We now go on the hunt for a Galois representation attached to E, where E continues to be an elliptic curve defined over $\mathbb{Q}$ (with a suitable integral model). It turns out that there basically one way to do this, though the recipe is somewhat roundabout. Something special about E is that it comes with a group law, which means that $E(R)$ is a group (in a functorial way) for all $\mathbb{Q}$-algebras R. For example, Example 8 realized $E(\mathbb{C})$ as $\mathbb{C}/\Lambda$ for some lattice Λ, and the group law on $\mathbb{C}/\Lambda$ is the expected one.

As such, we note that $E(\overline{\mathbb{Q}})$ is a group with an action by $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. One can see this purely by the ambient functoriality; alternatively, one can directly construct this action by noting that $E(\overline{\mathbb{Q}}) \subseteq \mathbb{P}^2_{\mathbb{Q}}(\overline{\mathbb{Q}})$, and of course $\mathbb{P}^2_{\mathbb{Q}}(\overline{\mathbb{Q}})$ has a Galois action by acting on the coordinates:

$$\sigma([X:Y:Z]) \coloneqq [\sigma(X):\sigma(Y):\sigma(Z)]$$

for all $\sigma \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ and $[X:Y:Z] \in \mathbb{P}^2_{\mathbb{Q}}(\overline{\mathbb{Q}})$.

However, in order to do linear algebra, we want our Galois representation to be valued in a vector space. Thus, the Galois action on $E(\overline{\mathbb{Q}})$ will not quite work. As such, we will want the following fact about the group law; throughout, $G[n]$ denotes the n-torsion of a group G.

Proposition 22. Fix an elliptic curve E defined over an algebraically closed field $\overline{K}$. For any n which is nonzero in K, there is a non-canonical isomorphism

$$E(\overline{K})[n] \cong (\mathbb{Z}/n\mathbb{Z})^2.$$

Example 23. For $E = \mathbb{C}/\Lambda$, it is not hard to show that $E[n]$ is a group isomorphic to $(\mathbb{Z}/n\mathbb{Z})^2$. In particular, as abstract groups, one finds that $\mathbb{C} \cong \Lambda \otimes_{\mathbb{Z}} \mathbb{R}$, so

$$E \cong (\mathbb{R}/\mathbb{Z})^2,$$

whose n-torsion has the desired property.

This is slightly better: for any prime ℓ, we see that $E(\overline{\mathbb{Q}})[\ell]$ is a 2-dimensional $\mathbb{F}_\ell$-vector space, and more generally, $E(\overline{\mathbb{Q}})[\ell^\bullet)$ is a free module over $(\mathbb{Z}/\ell^\bullet\mathbb{Z})$ of rank 2. In order to not have to deal with torsion, we take an inverse limit.

Definition 24 (Tate module)**.** Fix an elliptic curve E defined over an algebraically closed field $\overline{K}$. For any prime ℓ nonzero in K, we define the *Tate module* as the inverse limit

$$T_\ell E \coloneqq \varprojlim E\left[\ell^\bullet\right],$$

where the transition maps $E\left[\ell^{\bullet+1}\right] \twoheadrightarrow E\left[\ell^\bullet\right]$ are given by multiplication by ℓ. We also define $V_\ell E \coloneqq T_\ell E \otimes_{\mathbb{Z}} \mathbb{Q}$.

Remark 25. We see that $T_\ell E$ is non-canonically isomorphic to $\mathbb{Z}_\ell^2$, so $V_\ell E$ is non-canonically isomorphic to $\mathbb{Q}_\ell^2$.

Remark 26. Intuitively, one should think about $T_\ell E$ as a complex-analytic version of $H_1(E,\mathbb{Z})$. For example, $(V_\ell E)^\vee$ should be analogous to $H^1(E,\mathbb{Q})$.

The Galois action of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on each $E(\overline{\mathbb{Q}})\left[\ell^\bullet\right]$ will now assemble into a Galois action on $T_\ell E$ and thus $V_\ell E$, so we have produced a homomorphism

$$\rho_\ell \colon \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{GL}(V_\ell E).$$

It is rather difficult to compute this Galois representation directly, but it knows quite a bit about the arithmetic of E.

Proposition 27. Fix an elliptic curve E defined over a field K, and let ℓ be a prime nonzero in K. For any endomorphism $\varphi\colon E \to E$, let $P_\ell(T)$ be the monic characteristic polynomial of φ acting on $V_\ell E$. For each $n \in \mathbb{Z}$, the number $P_\ell(n)$ is independent of ℓ and equals the degree of the map $\varphi - [n]$, where $[n]\colon E \to E$ is multiplication by n.

Example 28. Let's compare this with what we expect to happen for $H_1(E, \mathbb{Z})$ when $E = \mathbb{C}/\Lambda$. Then $H_1(E, \mathbb{Z}) = \Lambda$, so the characteristic polynomial P_φ of some endomorphism $\varphi\colon E \to E$ will satisfy

$$\begin{aligned} |P_\varphi(n)| &= |\det(\varphi - [n] | H_1(E, \mathbb{Z}))| \\ &= |\det(\varphi - [n] | \Lambda)| \\ &= \#\left(\frac{\Lambda}{(\varphi - [n])\Lambda}\right) \\ &= \#\left(\frac{(\varphi - [n])^{-1}\Lambda}{\Lambda}\right) \\ &= \#\ker(\varphi - [n]) \\ &= \deg(\varphi - [n]). \end{aligned}$$

Here are two examples.

Corollary 29. Fix an elliptic curve E defined over $\mathbb{Q}$ with a good enough integral model, and let ℓ be a prime. For a prime $p \neq \ell$, let $P_\ell(T)$ be the characteristic polynomial of $\rho_\ell(\mathrm{Frob}_p)$, for any choice of Frob_p in the conjugacy class of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Then

$$P_\ell(0) = p \qquad \text{and} \qquad P_\ell(1) = \#E(\mathbb{F}_p).$$

Proof. Let $F\colon A(\overline{\mathbb{F}_p}) \to A(\overline{\mathbb{F}_p})$ be given by the action of Frob_p. Proposition 27 tells us that $P_\ell(0)$ equals the degree of F, which is p because it corresponds to the degree of some field extension which looks like $K(t^p) \subseteq K(t)$. Similarly, we see $P_\ell(1)$ equals the degree of the map $F - \mathrm{id}_E$, which is the number of points fixed by F, which is the size of $E(\mathbb{F}_p)$. ∎

In particular, in the situation of the corollary, we are able to compute that

$$P_\ell(T) = T^2 - aT + p,$$

where $a = (p+1) - \#E(\mathbb{F}_p)$. Factoring $P_\ell(T) = (T - \alpha_1)(T - \alpha_2)$, we find that

$$\frac{1}{\sqrt{p}} \operatorname{tr} \rho_\ell(\mathrm{Frob}_p) = \frac{\alpha_1 + \alpha_2}{\sqrt{p}} = \frac{(p+1) - \#E(\mathbb{F}_p)}{\sqrt{p}} = a_p(E).$$

Thus, our Galois representation is good enough to understand our desired numbers $a_p(E)$!

1.5 Defining the Sato–Tate Group: Elliptic Curves

We continue with our elliptic curve E defined $\mathbb{Q}$, and we fix our auxiliary prime ℓ. The moral of the story is that we can measure the distribution of $a_p(E)$s via the distribution of $\rho_\ell(\mathrm{Frob}_p)$; we then recover the distribution of the $a_p(E)$s by taking the trace.

Once again, we may guess that the $\rho_\ell(\mathrm{Frob}_p)$s must equidistribute in $\mathrm{GL}(V_\ell E) \cong \mathrm{GL}_2(\mathbb{Q}_\ell)$. However, this cannot be the case because

$$\det \rho_\ell(\mathrm{Frob}_p) = p$$

by Corollary 29. Thus, we would like to rescale $\rho_\ell(\mathrm{Frob}_p)$ to account for this determinant condition. Namely, we would like to replace $\rho_\ell(\mathrm{Frob}_p)$ with $\frac{1}{\sqrt{p}}\rho_\ell(\mathrm{Frob}_p)$, but there is no reasonable way to do this because $\mathbb{Q}_\ell$

may not have the element $1/\sqrt{p}$ for all the primes p we want to look at. And even when it does, there are two reasonable square roots to look at, so it is not obvious which one to choose: a different choice will lead to a different trace!

To fix this problem, we cheat: we choose any embedding $\iota\colon \mathbb{Q}_\ell \hookrightarrow \mathbb{C}$, which exists because these two fields have the same cardinality, and $\mathbb{C}$ is algebraically closed. Then we may hope that the elements

$$\frac{1}{\sqrt{p}}\iota(\rho_\ell(\mathrm{Frob}_p)) \in \mathrm{SL}_2(\mathbb{C})$$

will equidistribute as p varies. However, this still cannot be the case because ρ_ℓ is a continuous map with compact source $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, so these elements should actually live in some compact group. As such, we fix the compact subgroup $\mathrm{SU}_2 \subseteq \mathrm{SL}_2(\mathbb{C})$, and we hope that the elements

$$\frac{1}{\sqrt{p}}\iota(\rho_\ell(\mathrm{Frob}_p)) \in \mathrm{SU}_2$$

will equidistribute. Technically, this does not really make sense because Frob_p was only defined up to conjugacy, so the element $\frac{1}{\sqrt{p}}\iota(\rho_\ell(\mathrm{Frob}_p))$ will also only be defined up to conjugacy, so we are really hoping that the elements

$$\left[\frac{1}{\sqrt{p}}\iota(\rho_\ell(\mathrm{Frob}_p))\right] \in \mathrm{Conj}(\mathrm{SU}_2)$$

will equidistribute.

This is generally true.

Example 30. For the curve $E_2\colon y^2 = x^3 + x + 1$, it will turn out that the conjugacy classes

$$\left[\frac{1}{\sqrt{p}}\iota(\rho_\ell(\mathrm{Frob}_p))\right] \in \mathrm{Conj}(\mathrm{SU}_2)$$

do equidistribute in $n\,\mathrm{Conj}(\mathrm{SU}_2)$. This is the "generic" case for our elliptic curves, and one can show that this equidistribution gives rise to the semicircle distribution seen in Example 18 upon applying the trace.

In fact, one has the following.

Theorem 31 (Sato–Tate, non-CM elliptic curve)**.** Fix an elliptic curve E defined over $\mathbb{Q}$, and let ℓ be an auxiliary prime, and let $\iota\colon \mathbb{Q}_\ell \to \mathbb{C}$ be some embedding. Let ρ_ℓ be the Galois representation given by $V_\ell E$. Assume $\mathrm{End}(E_{\overline{\mathbb{Q}}}) = \mathbb{Z}$. Then the elements

$$\left\{\frac{1}{\sqrt{p}}\iota(\rho_\ell(\mathrm{Frob}_p))\right\}_{p \text{ prime}}$$

equidistribute among the conjugacy classes of SU_2.

Remark 32. Perhaps we ought to explain what it means to equistribute in $\mathrm{Conj}(\mathrm{SU}_2)$. Well, SU_2 is a compact topological group, so it has a Haar measure, and there is a procedure to push this measure forward along the canonical projection.

The difference between Examples 17 and 18 imply that we cannot expect the above discussion to be true for all elliptic curves.

To explain what is going on, note that $E_1\colon y^2 = x^3 + 1$ has a bizarre extra endomorphism $\varphi\colon E_1 \to E_1$ given by

$$\varphi(x, y) \coloneqq (\zeta_3 x, y).$$

This endomorphism is defined over $\mathbb{Q}(\zeta_3)$, which means that whenever $\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\zeta_3))$, we will have

$$\rho_\ell(\varphi) \circ \rho_\ell(\mathrm{Frob}_p) = \rho_\ell(\mathrm{Frob}_p) \circ \rho_\ell(\varphi).$$

Thus, we do not expect our elements to equidistribute in $\overline{\mathrm{SU}_2}$: there should be an extra condition to account for commuting with endomorphisms (possibly defined over an extension of $\mathbb{Q}$).

The general way to account for this is to simply ignore it. We have the following definition.

Definition 33 (ℓ-adic monodromy group)**.** Fix an elliptic curve E defined over $\mathbb{Q}$, and let ℓ be an auxiliary prime. Let G_ℓ be the image of ρ_ℓ, and let G_ℓ^{Zar} be the smallest algebraic subgroup of $\mathrm{GL}_2(\mathbb{Q}_\ell)$ which is defined over $\mathbb{Q}_\ell$ containining G_ℓ. We call G_ℓ^{Zar} the *ℓ-adic monodromy group*.

We can now define the Sato–Tate group $\mathrm{ST}(E)$ by retelling the above story.

Definition 34 (Sato–Tate group)**.** Fix an elliptic curve E defined over $\mathbb{Q}$, and let ℓ be an auxiliary prime. Let $G_\ell^{\mathrm{Zar},1}$ be the subgroup of G_ℓ^{Zar} cut out by the condition that the determinant equals 1. Then the *Sato–Tate group* is a maximal compact subgroup of $G_\ell^{\mathrm{Zar},1}(\mathbb{C})$, where we take $\mathbb{C}$-points via some embedding $\mathbb{Q}_\ell \subseteq \mathbb{C}$.

And here is the theorem, due to Richard Taylor and many other people.

Theorem 35 (Sato–Tate, elliptic curve)**.** Fix an elliptic curve E defined over $\mathbb{Q}$, and let ℓ be an auxiliary prime, and let $\iota\colon \mathbb{Q}_\ell \to \mathbb{C}$ be some embedding. Then the elements

$$\left\{ \frac{1}{\sqrt{p}} \iota(\rho_\ell(\mathrm{Frob}_p)) \right\}_{p \text{ prime}}$$

equidistribute among the conjugacy classes of $\mathrm{ST}(E)$.

Remark 36. It is not totally clear why $\frac{1}{\sqrt{p}}\iota(\rho_\ell(\mathrm{Frob}_p))$ would even be conjugate to any element in $\mathrm{ST}(E)$. Roughly speaking, we are combining two properties.

1. It turns out that $\rho_\ell(\mathrm{Frob}_p)$ is diagonalizable over an algebraic closure. Thus, our element (with eigenvalues of absolute value 1) it certainly contained in some compact subgroup.
2. Over $\mathbb{C}$, it turns out that any compact subgroup is conjugate to a subgroup of a given maximal compact subgroup. Thus, we can conjugate our element into $\mathrm{ST}(E)$.

Example 37. For E_1, one shows that $\mathrm{ST}(E_1)$ is a normalizer of U_1 diagonally embedded in SU_2. Roughly speaking, the point is that the field $\mathbb{Q}(\zeta_3)$ has an action on $V_\ell E_1$ which commutes with the Galois action restricted to $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}(\zeta_3))$, which causes the Galois action to diagonalize.

1.6 The Sato–Tate Conjecture: Abelian Varieties

In order to wrap up our story, we note that our discussion of Sato–Tate groups for elliptic curves generalizes to abelian varieties without too much effort.

Definition 38 (abelian variety)**.** Fix a field K. Then an *abelian variety* A over K is a smooth projective geometrically integral group variety over K.

Here, being a group variety means that $A(R)$ is a group (in a functorial way) for all K-alegbras R, as we had for elliptic curves.

Example 39. Any product of elliptic curves continues to be an abelian variety. The category of abelian varieties is in fact abelian, so we can also take kernels and cokernels.

Example 40. Fix a nonnegative integer g. Then for some lattices $\Lambda \subseteq \mathbb{C}^g$, it turns out that the quotient $\mathbb{C}^g/\Lambda$ is a projective variety over $\mathbb{C}$, so this is an abelian variety over $\mathbb{C}$ with the obvious group law.

Let's make explicit how the discussion of the previous two subsections generalizes to abelian varieties. Here is the generalization of Proposition 22.

Proposition 41. Fix an abelian vareity A over a field K of dimension g. For any n which is nonzero in K, there is a non-canonical isomorphism

$$E(\overline{K})[n] \cong (\mathbb{Z}/n\mathbb{Z})^{2g}.$$

This allows us to define the Tate module as before.

Definition 42 (Tate module)**.** Fix an abelian variety A defined over a field K. For any prime ℓ nonzero in K, we define the *Tate module* as the inverse limit

$$T_\ell A \coloneqq \varprojlim A\left[\ell^\bullet\right].$$

We also define $V_\ell A \coloneqq T_\ell A \otimes_{\mathbb{Z}} \mathbb{Q}$. As before, we find that $T_\ell A$ is a free module over $\mathbb{Z}_\ell$ of rank $2g$.

Thus, as before, we get an ℓ-adic Galois representation

$$\rho_\ell \colon \operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \operatorname{GL}(V_\ell A).$$

We can also generalize some of our discussion about the characteristic polynomial. Here is the generalization of Corollary 29.

Proposition 43. Fix an abelian variety A defined over $\mathbb{Q}$ with a good enough integral model, and let ℓ be a prime. For a prime $p \neq \ell$, let $P_\ell(T)$ be the characteristic polynomial of $\rho_\ell(\operatorname{Frob}_p)$, for any choice of Frob_p in the conjugacy class of $\operatorname{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$. Then $P_\ell(T)$ is a polynomial independent of ℓ, and

$$P_\ell(0) = p^g \qquad \text{and} \qquad P_\ell(1) = \#A(\mathbb{F}_p).$$

Thus, our discussion defining $\operatorname{ST}(E)$ can be generalized to build $\operatorname{ST}(A)$, as follows.

1. Let G_ℓ be the image of ρ_ℓ in $\operatorname{GL}(V_\ell A)$, and let $G_\ell^{\operatorname{Zar}}$ be the smallest algebraic subgroup defined over $\mathbb{Q}_\ell$ containing G_ℓ.
2. By choosing a Weil pairing on $T_\ell A$ (which we will not define), it turns out that there is a non-degenerate symplectic form on $V_\ell A$, and $\rho_\ell(\operatorname{Frob}_p)$ acts with multiplier p with respect to this form. Thus, we let $G_\ell^{\operatorname{Zar},1}$ be the subgroup of $G_\ell^{\operatorname{Zar}}$ cut out by preserving the corresponding symplectic form.
3. Choose some embedding $\iota \colon \mathbb{Q}_\ell \to \mathbb{C}$. We now let $\operatorname{ST}(A)$ be a maximal compact subgroup of the complex algebraic group $\iota\left(G_\ell^{\operatorname{Zar},1}\right)$.

And here is the conjecture.

Conjecture 44 (Sato–Tate)**.** Fix an abelian variety A defined over $\mathbb{Q}$, and let ℓ be an auxiliary prime, and let $\iota \colon \mathbb{Q}_\ell \to \mathbb{C}$ be some embedding. Let ρ_ℓ be the Galois representation given by $V_\ell A$. Then the elements

$$\left\{ \frac{1}{\sqrt{p}} \iota(\rho_\ell(\operatorname{Frob}_p)) \right\}_{p \text{ prime}}$$

equidistribute among the conjugacy classes of $\operatorname{ST}(A)$.

Remark 36 once again explains why our elements have a conjugacy class in $\operatorname{ST}(A)$. Also, note that this is a conjecture unlike Theorem 35!

1.7 Jacobian Varieties

In order to provide some examples of abelian varieties, we introduce Jacobian varieties into our story. In order to avoid doing any difficult algebraic geometry or complex geometry, we will define these by universal property.

Proposition 45. Fix a curve C defined over a field K equipped with a point $x \in C(K)$. Then there exists an abelian variety $\mathrm{Jac}(C)$ equipped with a map $C \hookrightarrow \mathrm{Jac}(C)$ sending $x \mapsto 0$ and satisfying the following the corresponding universal property: for any abelian variety A and morphism $\varphi\colon C \to A$ sending $\varphi(x) = 0$, there exists a unique map $\widetilde{\varphi}\colon \mathrm{Jac}(C) \to A$ making the following diagram commute.

$$\begin{array}{ccc} C & \hookrightarrow & \mathrm{Jac}(C) \\ & \searrow^{\varphi} & \downarrow{\scriptstyle \widetilde{\varphi}} \\ & & A \end{array}$$

Definition 46 (Jacobian)**.** Fix a curve C defined over a field K equipped with a point $x \in C(K)$. Then the abelian variety $\mathrm{Jac}(C)$ produced by Proposition 45 is called the *Jacobian* of C.

Example 47. For an elliptic curve E, we see that $\mathrm{Jac}(E) = E$ because E is already an abelian variety.

Example 48. Let's give the construction for $\mathrm{Jac}(C)$ when our curve C is defined over $\mathbb{C}$. One can embed the first homology $H_1(C, \mathbb{Z})$ into the complex vector space $H^0(C, \Omega^1_C)^\vee$ by taking $[\gamma] \in H_1(C, \mathbb{Z})$ to the functional

$$\omega \mapsto \int_\gamma \omega.$$

Then $\mathrm{Jac}(C) = H^0(C, \Omega^1_C)^\vee / H_1(C, \mathbb{Z})$.

Remark 49. It is not obvious from the definition, but it turns out that the dimension of $\mathrm{Jac}(C)$ equals the genus of the curve C.

Remark 50. Motivated by section 1.3, the importance of the Jacobian arises from the fact that the map $C \to \mathrm{Jac}(C)$ induces an isomorphism

$$H^1(\mathrm{Jac}(C), \mathbb{Q}_\ell) \to H^1(C, \mathbb{Q}_\ell)$$

on first cohomology. In particular, we can hope to read off properties of the curve (related to its cohomology) from the Jacobian. The analogous statement for $\mathbb{C}$ is simply that $H_1(C, \mathbb{Z}) \to H_1(\mathrm{Jac}(C), \mathbb{Z})$ is an isomorphism, which is clear from our construction.

The following result follows from the discussion in section 1.3 and the above remark.

Proposition 51. Fix a curve C defined over $\mathbb{Q}$ with a good enough integral model, and set $A := \mathrm{Jac}(C)$. For any prime p and an auxiliary prime $\ell \neq p$, we have

$$\#C(\mathbb{F}_p) - (p+1) = \mathrm{tr}\left(\rho_\ell(\mathrm{Frob}_p) | V_\ell A\right)$$

Thus, we expect computable (but perhaps complicated) equidistribution results for general curves to follow from Conjecture 44. Here is one such histogram.

Example 52. Consider the genus-2 curve $C\colon y^2 = x^5 + x + 1$. Here is a histogram for the values of $a_p(C) \coloneqq (\#C(\mathbb{F}_p) - (p+1))/\sqrt{p}$ for $p < 10^5$.