FROBENIUS DISTRIBUTIONS
(SATO–TATE DISTRIBUTIONS)

COURSE: KIRAN KEDLAYA AND ANDREW SUTHERLAND
NOTES: ROSS PATERSON

DISCLAIMER. These notes were taken live during lectures at the Spring School on Arithmetic Statistics held at CIRM from 8^{th}–12^{th} May 2023. Any mistakes are the fault of the transcriber and not of the lecturer, they have not been proofread in any meaningful way.

LECTURE 1 (KEDLAYA)

For $f \in \mathbb{Z}[X]$ squarefree of degree d, define

$$N_f(p) = \#\{x \in \mathbb{F}_p : f(x) \equiv 1 \mod p\}.$$

Note that clearly $0 \leq N_f(p) \leq d$.

Example 1 ([Sut19, §1.1]).

Definition 2. Let

$$c_i(B) := \frac{\#\{p \leq B \ : \ N_f(p) = i\}}{\#\{p \leq B\}}$$

Claim: We can describe limiting values of $c_i(B)$ for all i.

Let $L = \mathbb{Q}(\alpha_1, \ldots, \alpha_d)$ be the splitting field of f over $\mathbb{Q}$, where $f = \prod_{i=1}^d (x - \alpha_i)$, and $G = \mathrm{Gal}(L/\mathbb{Q})$ which acts transitively on this set of roots. Then we let

$$\rho : G \to \mathrm{GL}_d(\mathbb{C})$$

be the associated permutation representation. For p prime we have an exact sequence given as follows. Choose a prime $\mathfrak{p} \mid p$ of $\mathcal{O}_L$, and let:

- $D_\mathfrak{p}$ be the associated decomposition group (i.e. the stabiliser of $\mathfrak{p}$ under the action of G on the set of primes above p);
- $I_\mathfrak{p}$ be the inertia subgroup of $D_\mathfrak{p}$.

Then we have

$$1 \longrightarrow I_\mathfrak{p} \longrightarrow D_\mathfrak{p} \longrightarrow \mathrm{Gal}\left(\frac{\mathcal{O}_L}{\mathfrak{p}}/\mathbb{F}_p\right) \longrightarrow 1.$$

Note that $\mathrm{Gal}\left(\frac{\mathcal{O}_L}{\mathfrak{p}}/\mathbb{F}_p\right)$ has a canonical generator, $x \mapsto x^p$, and so we denote by $\mathrm{Frob}_\mathfrak{p}$ a choice of lift of this in $D_\mathfrak{p}$. if p is unramified (which is true of all but finitely many p) then $\mathrm{Frob}_\mathfrak{p}$ is a well defined element of $D_\mathfrak{p}$. As $\mathfrak{p}$ varies amongst primes above p, $\mathrm{Frob}_\mathfrak{p}$ traces out a conjugacy class in G, which we denote by Frob_p.

For unramified p, $N_f(p)$ is counting fixed points of $\mathrm{Frob}_\mathfrak{p}$ on $\{\alpha_1, \ldots, \alpha_d\}$. That is,

$$N_f(p) = \mathrm{tr}(\rho(\mathrm{Frob}_\mathfrak{p})).$$

Applying Chebotaryov density theorem, we see that the conjugacy class of Frob_p is uniformly distributed in the set of conjugacy classes of G, denoted $\mathrm{conj}(G)$, with respect to the measure which weights a conjugacy class C proportionately to its size $\#C$.

Example 3. For $f(x) = x^3 - x + 1$ we have $G = S_3$, so

$$\lim_{B\to\infty} c_i(B) = \begin{cases} \frac{2}{6} & \text{if } i = 0 \\ \frac{3}{6} & \text{if } i = 1 \\ \frac{1}{6} & \text{if } i = 3. \end{cases}$$

Aside. If G is abelian then $L \subseteq \mathbb{Q}(\zeta_n)$ for some n, and then $\mathrm{Frob}_{\mathfrak{p}}$ is determined by $p \mod n$.

Think now of G as a discrete topological group, note that this means that it is compact (also Hausdorff). Any compact topological group has a unique left- and right- invariant probability measure in the Radon sense (i.e. continuous functions $G \to \mathbb{R}$ can be integrated) known as the Haar measure, which we denote by μ_G.

I can then take the pushforward measure on the set of conjugacy classes of G. That is, I evaluate the functional on class functions.

Example 4. Consider $\mathrm{SU}(2) = \left\{ A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{C}) \ : \ ad - bc = 1,\ A^{-1} = A^* = \overline{A}^T \right\}.$ Via the trace map, we have a bijection between the conjugacy classes of $\mathrm{conj}(\mathrm{SU}(2))$ and the set $[-2, 2]$.

Definition 5. Let X be some probability space, and $t : X \to \mathbb{R}$ be a random variable. Then the moment sequence of t is $(\mathbb{E}(t^n))_{n\in\mathbb{Z}_{\geq 0}}$, where we always take $t^0 = 1$.

Here is a comment that we won't expand on for now.

- We could also look at

  $$N_f(p^k) = \#\left\{x \in \mathbb{F}_{p^k} \ : \ f(x) = 0\right\},$$

  and for fixed p we could package this collection (indexed by k) into a local zeta function (see later in the course). Then

  $$N_f(p^k) = \mathrm{tr}(\rho(\mathrm{Frob}_{\mathfrak{p}}^k))$$

1. Arithmetic Schemes

Definition 6. Let X be a scheme of finite type over $\mathbb{Z}$. For each prime number p we define

$$N_X(p) := \#X(\mathbb{F}_p).$$

Question 7. *How does this depend on p?*

Elliptic Curves. Consider $E \subseteq \mathbb{P}^2_{\mathbb{Z}}$ cut out by the affine model $y^2 = x^3 + Ax + B$. Assume that $x^3 + Ax + B$ is squarefree (so the generic fibre $E_{\mathbb{Q}}$ is an elliptic curve), and that $E_{\mathbb{Q}}$ does not have complex multiplication.

Theorem 8 (Hasse)**.** *For each prime number p, write*

$$\#E(\mathbb{F}_p) = p + 1 - t_p.$$

Then, so long as $E_{\mathbb{F}_p}$ is smooth, $|t_p| \leq 2\sqrt{p}$.

This suggests we should look at $\frac{t_p}{\sqrt{p}} \in [-2, 2]$. Looking at these numbers experimentally, there appears to be a clear pattern in their distribution, which is explained by the following theorem.

Theorem 9. *The values* $\frac{t_p}{\sqrt{p}}$ *are equidistributed for the pushforward of the Haar measure on* $\mathrm{conj}(\mathrm{SU}(2))$ *on* $[-2, 2]$.

Lecture 2 (Sutherland): Equidistribution

Generalities

Let X be a compact Hausdorff topological space, and let $C(X)$ denote the Banach space of continuous functions $f : X \to \mathbb{C}$ under the sup-norm. For $f, g \in C(X)$ which are $\mathbb{R}$-valued with $f(x) \leq g(x)$ for all $x \in Z$ then we will write $f \leq g$. If we write such an inequality then part of the data is the assertion that the functions are real valued.

Definition 10. a (positive normalised Radon) measure is a continuous $\mathbb{C}$-linear

$$\mu : C(X) \to \mathbb{C}$$

such that for all $f \geq 0$, we have $\mu(f) \geq 0$, and moreover $\mu(1_X) = 1$.

Example 11. Note the dirac measure at a point $x \in X$

$$\delta_X : C(X) \to \mathbb{C}$$

given by $f \mapsto f(x)$.

Notation 12. We denote for $f \in C(X)$

$$\int_X f\mu := \mu(f).$$

Given such a measure, we can define a measure of subsets $S \subset C(X)$

$$\mu(S) := \begin{cases} \sup\{\mu(f) \;:\; 0 \leq f \leq 1_S\} & \text{if } S \text{ is open} \\ 1 - \mu(X - S) & \text{if } S \text{ is closed} \\ 0 & \text{if } \forall \varepsilon > 0 \; \exists \text{ open } U \supset S \text{ with } \mu(U) \leq \varepsilon \\ \mu(\overline{S}) = \mu(S^\circ) & \text{if } \mu(\partial S) = 0, \; \partial S = \overline{s} \backslash S^\circ. \end{cases}$$

Definition 13. A sequence $(x_1, x_2, \ldots)$ in X is μ-equidistributed if

$$\mu(f) = \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f(x_i)$$

for all $f \in C(X)$.

Lemma 14. *Let* (f_j) *be a family of functions in* $C(X)$ *whose* $\mathbb{C}$*-span is dense in* $C(X)$*. If* (x_i) *is a sequence in* X *for which* $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f_j(x_i)$ *converges for all* f_j *in our family, then there exists a unique measure* μ *on* X *for which* (x_i) *is* μ*-equidistributed.*

Proof. See [Ser68]. □

Definition 15. $S \subseteq C(X)$ is μ-quarrable if $\mu(\partial S) = 0$.

Proposition 16. *If* (x_i) *is* μ*-equidistributed and* S *is* μ*-quarrable, then*

$$\mu(S) = \lim_{n\to\infty} \frac{\#\{x_i \in S \ : \ i \leq n\}}{n}$$

Proof. Exercise. □

Example 17. $X = [0,1]$, μ the Lebesgue measure, then (x_i) is equidistributed if and only if $\forall 0 \leq a < b \leq 1$,

$$\lim_{n\to\infty} \frac{\#\{x_i \in [a,b] \ : \ i \leq n\}}{n} = \mu([a,b]) = b - a$$

Compact Groups

From now on, $X := \mathrm{conj}(G)$ for some compact group G. The Haar measure on G induces a measure μ on X via

$$\mu(f) := \mu(f \circ \mathrm{conj}).$$

In this setting "equidistributed" means μ-equidistributed with respect to this μ.

Proposition 18. *A sequence* (x_i) *in* $X = \mathrm{conj}(G)$ *is equidistributed if and only if for every irreducible character* $\chi : G \to \mathbb{C}$

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \chi(x_i) = \mu(\chi)$$

Proof. The Peter–Weyl theorem shows that the irreducible characters χ span a dense subspace of $C(X)$. □

Corollary 19. (x_i) *is equidistributed if and only if for all nontrivial irreducible* χ.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \chi(x_i) = 0$$

Proof. For $\chi = 1$, the $\mu(1) = 1$ is always immediate. For the nontrivial χ,

$$\mu(\chi) = \int_G \chi\mu = \int_G 1 \cdot \chi\mu = 0$$

□

An easy Sato–Tate Result

Let $E/\mathbb{F}_q$ be an elliptic curve with $\#E(\mathbb{F}_q) = q+1-t_q$, where $t_q = \mathrm{tr}(\pi_E) = \alpha+\overline{\alpha}$ for some $\alpha \in \mathbb{C}$ with $|\alpha| = q^{1/2}$. Considering the base change, we have

$$\#E(\mathbb{F}_{q^r}) = q^r + 1 - \mathrm{tr}(\pi_E^r) = q^r + 1 - (\alpha^r + \overline{\alpha}^r).$$

Let $t_{q^r} = q^r + 1 - \#E(\mathbb{F}_{q^r})$.

Proposition 20. *Assume that* E *is ordinary, and let* $x_r := \frac{t_{q^r}}{q^{r/2}}$*. The sequence* (x_r) *is equidistributed in* $[-2,2]$ *with respect to the measure*

$$\mu := \frac{1}{\pi} \frac{dz}{\sqrt{4 - z^2}},$$

where dz *is the Lebesgue measure on* $[-2,2]$.

Proof. Let $U(1) = \{u \in \mathbb{C}^\times : |u| = 1\}$. For $u = e^{i\theta}$, θ is uniformly distributed under the Haar measure for $U(1)$. To compute the pushforward of the Haar measure to $z := 2\cos\theta$

$$dz = 2\sin(\theta)d\theta = \sqrt{4 - z^2}d\theta,$$

consider $\theta \in [0, \pi]$, $\mu = \frac{d\theta}{\pi} = \frac{1}{\pi}\frac{dz}{\sqrt{4-z^2}}$.

Nontrivial irreducible characters $U(1) \to \mathbb{C}^\times$ look like $\phi_a : u \mapsto u^a$ for some $a \in \mathbb{Z}_{\neq 0}$. Then

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \phi_a\left(\frac{\alpha^i}{q^{i/2}}\right) = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} \phi_a\left(\frac{\alpha^i}{q^{i/2}}\right)$$

□

L-Functions

Let us fix a number field K and consider a sequence $x = (x_\mathfrak{p})$ in $X = \mathrm{conj}(G)$ indexed by primes $\mathfrak{p}$ of K. Order these by $N(\mathfrak{p}) := \#\mathcal{O}_K/\mathfrak{p}$.

Definition 21. For each irreducible representation $\rho : G \to \mathrm{GL}_d(\mathbb{C})$ we define

$$L_X(\rho, s) := \prod_{\mathfrak{p}} \det\left(1 - \rho(x_\mathfrak{p})N(\mathfrak{p})^{-s}\right)^{-1},$$

which converges on $\Re(s) > 1$.

Theorem 22. *Suppose for every irreducible representation ρ, the function $L_x(\rho, s)$ is meromorphic on $\Re(s) \geq 1$ with no zeros or poles away from $s = 1$. Then $x = (x_\mathfrak{p})_\mathfrak{p}$ is equidistributed if and only if $L_x(\rho, 1) \notin \{0, \infty\}$ for every irreducible $\rho \neq 1$.*

Proof. See [Ser68], also Fité's notes from 2015 have a very nice exposition. □

Corollary 23. *L/K finite Galois, then $x = \left(\mathrm{conj}(\mathrm{Frob}_\mathfrak{p}|_L)\right)_\mathfrak{p}$ is equidistributed.*

Remark 24. This implies the Chebotaryev density theorem.

Proof. If $\rho = 1$ then $L_x(\rho, s) \approx \zeta_K(s)$ which is holomorphic and nonvanishing on $\Re(s) \geq 1$ except simple pole at $s = 1$ (Hecke).

If $\rho \neq 1$ then $L_X(\rho, s) \approx L(\rho, s)$ the Artin L-function and this is holomorphic and nonvanishing on $\Re(s) \geq 1$ (Artin). □

Sato–Tate for CM Elliptic Curves

Definition 25. A Hecke character is a continuous homomorphism $\psi : \mathbb{A}_K^\times \to \mathbb{C}^\times$ with $K^\times \subseteq \ker(\psi)$.

$$\mathrm{cond}(\psi) := \prod_{\mathfrak{p}} \mathfrak{p}^{e_\mathfrak{p}},$$

where $e_\mathfrak{p}$ is the least nonnegative integer such that

$$1 + \mathfrak{p}^{e_\mathfrak{p}} \subset \mathcal{O}_{K_\mathfrak{p}}^\times \subseteq \ker(\psi).$$

The L-function is

$$L(\psi, s) := \prod_{\mathfrak{p}\nmid\mathrm{cond}(\psi)} \left(1 - \psi(\pi_\mathfrak{p})N(\mathfrak{p})^{-s}\right)^{-1},$$

where $\pi_\mathfrak{p}$ is any choice of uniformizer for $K_\mathfrak{p}$.

Remark 26. We can unitarise a Hecke characer ψ via

$$\psi := \psi / |\psi| .$$

In particular we can always consider them as functions to $U(1)$.

Lemma 27. *For any unitarized Hecke character ψ, the sequence $(\psi(\mathfrak{p}))_{\mathfrak{p}}$ is equidistributed in $U(1)$.*

Proof. As above, irreducible representaitons of $U(1)$ are $\phi_a(u) = u^a$ for $a \in \mathbb{Z}$, and $\psi_a := \phi_a \circ \psi$ is also a unitarized Hecke character.

If $\psi_a = 1$ then $L(\psi_a, s) \approx \zeta_K(s)$, so all good. If $\psi_a \neq 1$ then $L(\psi_a, s)$ is hoolomorphic and nonvanishing on $\Re(s) \geq 1$. $\square$

Now assume that K is imaginary quadratic, and E/K is an elliptic curve with CM. Then K has a corresponding Hecke character ψ_E for which

$$|\psi_E(\pi_{\mathfrak{p}})| = N(\mathfrak{p})^{1/2},$$

with $t_{\mathfrak{p}} = \mathrm{tr}(\pi_E) = \psi_E(\pi_{\mathfrak{p}}) + \overline{\psi_E(\pi_{\mathfrak{p}})}$. Uniformize to get $x_{\mathfrak{p}} = \psi_E(\pi_{\mathfrak{p}}) + \overline{\psi}(\pi_{\mathfrak{p}}) \in [-2, 2]$.

Proposition 28. *The sequence $(x_{\mathfrak{p}})_{\mathfrak{p}}$ is equidistributed with respect to the measure*

$$\mu_{\mathrm{CM}} = \frac{1}{\pi} \frac{dz}{\sqrt{4 - z^2}}.$$

Proof. μ_{CM} is the pushforward of the Haar measure for $U(1)$ via $u \mapsto u + \overline{u}$ and the proposition then follows from the previous theorem on unitarized Hecke characters. $\square$

Lecture 3 (Kedlaya)

Today will break down as follows:

(I) Sato–Tate conjecture for non-CM elliptic curves,
(II) Ask a similar question for higher-dimensional abelian varieties,
(III) Show some features of the answers.

Sato–Tate for non-CM Elliptic Curves

Let $E/\mathbb{Q}$ be an elliptic curve. For each (all but finitely many) prime number p

$$\#E(\mathbb{F}_p) = p + 1 - t_p$$

where $|t_p| \leq 2\sqrt{p}$. We want to understand how $\frac{t_p}{2\sqrt{p}} \in [-2, 2]$ is distributed.

Let $G = \mathrm{SU}(2)$, and $X = \mathrm{conj}(G) \cong [-2, 2]$ where the isomorphism is via the trace map. Each class has a representative of the form

$$\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}$$

for $0 \leq \theta \leq \pi$. Then $t = \text{trace} = 2\cos(\theta)$. The pushforward of the Haar measure on $\mathrm{SU}(2)$ is

$$\mu = \frac{2}{\pi} \sin^2(\theta) d\theta = \frac{1}{2\pi} \sqrt{4 - t^2} dt.$$

For (all but finitely many) primes p, let $x_p = \frac{t_p}{\sqrt{p}} \in X$.

Theorem 29 (Realisation of Sato–Tate)**.** *If E does not have CM, then the sequence $(x_p)_p$ in X is equidistributed with respect to μ.*

To get started, we follow the model that Drew showed us for the CM case and appeal to L-functions associated to irreducible representations of G. For each nonnegative integer m, we have an irreducible representation

$$\rho_m : G \to \mathrm{GL}_{m+1}(\mathbb{C})$$

given by: for $m = 0$ this is the trivial representation; for $m = 1$ this is the standard representaton of $\mathrm{SU}(2) \subseteq \mathrm{GL}_2(\mathbb{C})$; for general m, $\rho_m = \mathrm{Sym}^m \rho_1$. We build L-functions via

$$L(\rho_m, s) = \prod_p \det\left(1 - \rho_m(x_p) p^{-s}\right)^{-1}$$

which converge for $\Re(s) \gg 0$.
Claim: For each $m > 0$, $L(\rho_m, s)$ extends to a holomorphic function on $\Re(s) \geq 1$ which does not vanish on this region.

Remark 30. For $m = 0$ we have the Riemann ζ-function $\zeta(s) = L(\rho_0, s)$ which has a pole at $s = 1$.

This is a *hard* theorem. To see how hard, look at the case $m = 1$. Then $L(\rho_1, s) = L\left(E, s + \frac{1}{2}\right)$, and the claim here follows from modularity of elliptic curves (a crowning achievement of 20th century mathematics). For $m > 1$ this does not follow immediately from the case $m = 1$, there is extra work which took longer. For the CM case, at this point, we had much more classical work (generalisations of the proof of analytic continuation of Riemann ζ) which handled Hecke L-functions. There is a long story here but we shall leave it for now.

Question 31. *Where does this break down if E has CM?*

Answer 32. In this case, some of the $L(\rho_m, s)$ also have poles at $s = 1$! In this case we get equidistribution for the embedding $\mathrm{U}(1) \to \mathrm{SU}(2)$ given by

$$e^{i\theta} \mapsto \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}.$$

For even m, $\rho_m|_{U(1)}$ contains a copy of the trivial represenation and so we actually get poles in our L-function (using Artin formalism: $L(\rho_1 \oplus \rho_2, s) = L(\rho_1, s) L(\rho_2, s)$, and the trivial representation gives us ζ).

Question 33. *What about an elliptic curve over a number field E/K?*

Answer 34. For CM, same proof holds. For non-CM this is only known when K is totally real or CM. We'd always have one of the cases below.

E with CM by $M \subset K$	E with CM by $M \not\subseteq K$	E not CM
$U(1)$	$N(U(1))$ (N=normaliser of)	$\mathrm{SU}(2)$

Question 35. *For an abelian variety over a number field, A/K, of dimension g, and $\mathfrak{p} \leq \mathcal{O}_K$ a prime ideal, write $q = \mathcal{O}_K/\mathfrak{p}$. Then*

$$A(\mathbb{F}_{q^k}) = \prod_{i=1}^{2g} (1 - \alpha_{\mathfrak{p},i}^k),$$

with $|\alpha_{\mathfrak{p},i}| = \sqrt{q}$. Can the distribution of $\frac{|\alpha_{\mathfrak{p},i}|}{\sqrt{q}}$ be modelled by $\mathrm{conj}(G)$ *for some compact Lie group G?*

Answer 36. We'll discuss this more tomorrow, now we have some discussion on data for Sato–Tate groups from Sutherland (links: genus 1, genus 2, genus 3)

Lecture 4 (Kedlaya)

2. Sato–Tate for Abelian Varieties

Let A/K be an abelian variety of dimension g over a number field (with a fixed polarisation).

Goal: To define (up to conjugacy):

- a compact Lie group $G = \mathrm{ST}(A) \leq U_{2g}(\mathbb{C})$ called the Sato–Tate group of A;
- for each prime ideal $\mathfrak{p} \leq \mathcal{O}_K$ of good reduction for A, a conjugacy class

$$X_{\mathfrak{p}} \in \mathrm{conj}(\mathrm{ST}(A))$$

  such that the eigenvalues of $X_{\mathfrak{p}}$ coincide with the normalised Frobenius eigenvalues at $\mathfrak{p}$ for the reduction of A over $\mathcal{O}_K/\mathfrak{p}$.

Remark 37. It will turn out later on that for structural reasons our groups $\mathrm{ST}(A)$ will have more restriction: they will be subgroups of $\mathrm{USp}(2g)$.

We will follow Serre (Lectures on $N_X(p)$) in our exposition. See also Banaszak-Kedlaya.

The analogue of the Sato–Tate conjecture for A will say that $\{X_{\mathfrak{p}}\}$ is equidistributed in $\mathrm{conj}(\mathrm{ST}(A))$ for the pushforward of the Haar measure.

Step 1. Pick a prime ℓ (coprime to the degree of our polarisation), and let T_ℓ be the ℓ-adic Tate module. That is:

$$T_\ell := \varprojlim_n A(\overline{\mathbb{Q}})[\ell^n],$$

where the limit is over the natural maps given by multiplication by ℓ on the group schemes

$$A[\ell^{n+1}] \to A[\ell^n].$$

Note that $A(\overline{\mathbb{Q}})[\ell^n] \cong (\mathbb{Z}/\ell^n\mathbb{Z})^{2g}$, and so as abelian groups

$$T_\ell \cong \mathbb{Z}_\ell^{2g}.$$

Moreover T_ℓ is acted on by G_K, it is a Galois module. We define

$$V_\ell = T_\ell \otimes_{\mathbb{Z}} \mathbb{Q}.$$

Step 2. Take the Galois representation

$$\rho_\ell : G_K \to \mathrm{GL}(V_\ell).$$

For a prime $\mathfrak{p}$ of K of good reduction, we have the $\mathrm{Frob}_{\mathfrak{p}} \in G_K$ and $\rho_\ell(\mathrm{Frob}_{\mathfrak{p}})$ has eigenvalues which are the Frobenius eigenvalues of A over $\mathcal{O}_K/\mathfrak{p}$.

Write G_ℓ for the image of ρ_ℓ. This is a compact group inside of $\mathrm{GL}(V_\ell)$.

Remark 38. Note it is the wrong kind of compact group: it is profinite, it's not a compact Lie group like we're looking for! We will do this via algebraic groups.

Step 3. Fix a basis of V_ℓ, so that we have $\rho_\ell : G_K \to \mathrm{GL}_{2g}(\mathbb{Q}_\ell)$, and G_ℓ is inside of this matrix group. We write G_ℓ^{zar} for the Zariski closure of G_ℓ in $\mathrm{GL}_{2g,\mathbb{Q}_\ell}$.

Remark 39. In other words, look at all elements in the function field of $\mathrm{GL}_2 2g, \mathbb{Q}_\ell$ which vanish on G_ℓ, and define G_ℓ^{zar} to be the subvariety of $\mathrm{GL}_{2,\mathbb{Q}_\ell}$ cut out by these algebraic functions.

Note that G_ℓ^{zar} is an algebraic group over $\mathbb{Q}_\ell$ (not just a group of points).
Recall: The Weil pairing is a Symplectic pairing

$$V_\ell \times V_\ell \to \mathbb{Q}_\ell(1),$$

where the right hand side is the rational Tate module of $\mathbb{G}_m$ (that is, by definition $\mathbb{Q}_\ell(1) = \varprojlim_n \mu_{\ell^n}(\overline{\mathbb{Q}}) \otimes_{\mathbb{Z}} \mathbb{Q}$)

This forces $G_\ell^{\mathrm{zar}} \subseteq \mathrm{GSp}_{2g,\mathbb{Q}_\ell}$ (symplectic similitudes), since for $g \in G_K$:

$$\langle g(a), g(b) \rangle = \chi(g) \langle a, b \rangle$$

where χ is the ℓ-adic character (for the representation $\mathbb{Q}(1)$).

Step 4. Fix an algebraic (**not** topological) embedding

$$i : \mathbb{Q}_\ell \to \mathbb{C},$$

and base extend G_ℓ^{zar} to $\mathbb{C}$ to get $G_{\ell,\mathbb{C}}^{\mathrm{zar}}$. Define also

$$G_{\ell,\mathbb{Q}_\ell}^{1,\mathrm{zar}} = G_{\ell,\mathbb{Q}_\ell}^{\mathrm{zar}} \cap \mathrm{Sp}_{2g,\mathbb{Q}_\ell}$$
$$G_{\ell,\mathbb{C}}^{1,\mathrm{zar}} = G_{\ell,\mathbb{C}}^{\mathrm{zar}} \cap \mathrm{Sp}_{2g,\mathbb{C}}$$

Step 5. Let $\mathrm{ST}(A)$ be a maximal compact subgroup (unique up to conjugacy) of

$$G_\ell^{1,\mathrm{zar}}(\mathbb{C}),$$

as a Lie group (via i). Moreover

$$\mathrm{ST}(A) \subseteq \mathrm{USp}_{2g}.$$

Step 6. We need to move our Frobenius traces across somehow! Note that $\rho_\ell(\mathrm{Frob}_\mathfrak{p}) \in G_\ell \subseteq G_\ell^{\mathrm{zar}}(\mathbb{Q}_\ell) \subseteq G_\ell^{\mathrm{zar}}(\mathbb{C})$. Write $M_\mathfrak{p} \in G_\ell^{\mathrm{zar}}(\mathbb{C})$ for the corresponding element. Moreover write

$$\overline{M_\mathfrak{p}} := N(\mathfrak{p})^{-1} M_\mathfrak{p} \in G_\ell^{1,\mathrm{zar}}(\mathbb{C}).$$

Claim: $\overline{M}_\mathfrak{p}$ is conjugate to some element of $\mathrm{ST}(A)$ (and the resulting class is uniquely determined).
Proof. Tate: $M_\mathfrak{p}$ is semisimple, thus so too is $\overline{M}_\mathfrak{p}$, so in particular $\overline{M}_\mathfrak{p}$ is contained in some compact subgroup of $G_\ell^{1,\mathrm{zar}}(\mathbb{C})$ (take the subgroup it generates and then take its closure, this is necessarily compact). This is contained in a maximal compact subgroup, which must then be conjugate to $\mathrm{ST}(A)$ (using the theory of Lie groups to say that maximal compact subgroups are unique up to conjugacy).

Regarding the Choices we made: We chose ℓ and $i : \mathbb{Q}_\ell \to \mathbb{C}$. We are happy with these not mattering up to conjugacy, since we only expect to define the Sato–Tate group up to conjugacy. It is a theorem that these choices do not affect the construction if $g \leq 3$. In general they do not affect the construction if the Mumford–Tate conjecture (which we won't define) holds for A (Cantoral Farfán–Comellin).

Remarks. There is an exact sequence of groups

$$1 \longrightarrow \mathrm{ST}(A)^\circ \longrightarrow \mathrm{ST}(A) \longrightarrow \pi_0(\mathrm{ST}(A)) \longrightarrow 1$$

where $\mathrm{ST}(A)^\circ$ is the connected component of the identity and depends only on $A_{\overline{\mathbb{Q}}}$. The definition is equivalent to the Mumford–Tate group after base change from $\mathbb{Q}$ to $\mathbb{R}$, and is related to endomorphism algebra of $A_{\overline{\mathbb{Q}}}$. Moreover $\pi_0(\mathrm{ST}(A))$ is canonically isomorphic to $\pi_0(G^{1,\mathrm{zar}}_{\ell,\mathbb{Q}_\ell})$, and moreover if you trace back through the construction you get a canonical isomorphism

$$\pi_0(\mathrm{ST}(A)) \cong \mathrm{Gal}(L/K)$$

for some finite Galois extension L/K. For $g \leq 3$ we get that L is the endomorphism field of A (that is, L is the smallest field such that $\mathrm{End}(A_L) \cong \mathrm{End}(A_{\overline{\mathbb{Q}}})$). In general the endomorphism field is contained in L, so you can bound this by controlling L (so the components of Sato–Tate groups) – see work of Guralnick–Kedlaya.

For $g = 1$, in fact $\mathrm{ST}(A) \in \{U_1, N(U_1), \mathrm{SU}_2, \mathrm{USp}_2\}$.

Theorem 40 (Fité–Kedlaya–Rotger–Sutherland)**.** *For $g = 2$ there are (up to conjugacy) exactly 52 closed subgroups of USp_4 that occur as $\mathrm{ST}(A)$ for some A, K.*

Theorem 41 (Fité–Kedlaya–Sutherland)**.** *For $g = 3$, there are 410 groups.*

3. Mumford–Tate and Hodge Groups

Let A be an abelian variety over a number field K and fix an embedding so that $K \subset \mathbb{C}$. Recall from last time the ℓ-adic monodromy group G^{zar}_ℓ, which is the Zariski closure of $\rho_{A,\ell}(G_K)$ in $\mathrm{GSp}_{2g}(\mathbb{Q}_\ell)$.

The (real) Line group $A(\mathbb{C})$ is isomorphic to a complex torus $\mathbb{C}^g/\Lambda$ where $\Lambda \cong H_1(A(\mathbb{C}), \mathbb{Z}) \cong \mathbb{Z}^{2g}$. The $\mathbb{R}$-module $\Lambda_\mathbb{R} := \Lambda \otimes_\mathbb{Z} \mathbb{R} \cong H^1(A(\mathbb{C}), \mathbb{R})$ admits a complex structure $h : \mathbb{C} \to \mathrm{End}(\Lambda_\mathbb{R})$. This makes (Λ, h) an *integral Hodge structure.*

We can also view h as a morphism of $\mathbb{R}$-algebraic groups

$$h : \mathcal{S} \to \mathrm{GL}_{\Lambda_\mathbb{R}},$$

where $\mathcal{S}$ is the Deligne torus.

Definition 42. The Deligne torus is the restriction of scalars of $\mathbb{G}_m$ from $\mathbb{C}$ to $\mathbb{R}$. In other words it is the multiplicative group $\mathbb{C}^\times$ viewes as an $\mathbb{R}$-algebraic group.

The group $S(\mathbb{R}) \cong \mathbb{C}^\times$ contains the circle group, $U(1)$, as a subgroup with $\mathbb{R}$-Zariski closure U_1, and the restriction

$$h : U_1 \to \mathrm{GL}_{\Lambda_\mathbb{R}}.$$

This has the property that $\forall u \in U_1(\mathbb{R}) = U(1)$, $h(u)$ has eigenvalues u, u^{-1} with multiplicity g. The image $h(U(1))$ is a *Hodge circle.*

Now consider $\Lambda_\mathbb{Q} := \Lambda \otimes_\mathbb{Z} \mathbb{Q}$, and the rational Hodge structure $(\Lambda_\mathbb{Q}, h)$.

Definition 43. The Mumford–Tate group of A is $\mathrm{MT}(A)$ is the $\mathbb{Q}$-Zariski closure of $h(\mathcal{S}(\mathbb{R}))$ in $\mathrm{GL}_{\Lambda_\mathbb{R}}(\mathbb{R})$. The Hodge group $\mathrm{Hg}(A)$ is the $\mathbb{Q}$-Zariski closure of $h(U(1))$ in $\mathrm{GL}_{\Lambda_\mathbb{Q}}(\mathbb{R})$.

Note:

- $\mathrm{MT}(A)$ is a $\mathbb{Q}$-algebraic subgroup of GSp_{2g};
- $\mathrm{Hg}(A)$ is a $\mathbb{Q}$-algebraic subgroup of Sp_{2g}.

Moreover $\mathrm{Hg}(A) = \mathrm{MT}(A) \cap \mathrm{Sp}_{2g}$.

Conjecture 44 (Mumford–Tate Conjecture)**.** *The identity component of* G_ℓ^{zar} *is* $\mathrm{MT}(A) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$*. Equivalently, the identity component of* $G_\ell^{\mathrm{zar},1}$ *is* $\mathrm{Hg}(A) \otimes_{\mathbb{Q}} \mathbb{Q}_\ell$*.*

Remark 45. This conjecture is known for $g \leq 3$ by work of Banaszak–Kedlaya and Moonen–Zahrin.

4. Sato–Tate Axioms

Let G be a subgroup of $\mathrm{USp}(2g)$ wit identity component G^0.

Definition 46. A Hodge circle in G is a subgroup H that is the image of a continuous homomorphism $\theta : U(1) \to G$ such that for all $u \in U(1)$ $\theta(u)$ has eigenvalues u, u^{-1} both with multiplicity g. Note that since it is the image of a connected group, $H \subseteq G^0$.

Now we can define the Sato–Tate axioms from a purely Lie-theoretic perspective, not an abelian variety in sight.

Definition 47. We say that G satisfies the Sato–Tate axioms (for abelian varieties of dimension g) if:

1. (Lie) G is a closed subgroup of $\mathrm{USp}(2g)$;
2. (Hodge) The Hodge circles in G generate a dense (nontrivial) subgroup of G^0;
3. (Rationality) For every component C of G, and every irreducible character χ on $\mathrm{GL}_{2g}(\mathbb{C})$,

$$\int_C \chi\mu \in \mathbb{Z}$$

when μ is normalised so that $\mu(1_C) = 1$ (where 1_C is the indicator function for C).

Remark 48. Note that conditions 1 and 2 show that this is an infinite compact subgroup of $\mathrm{USp}(2g)$

Theorem 49 (FKRS '12)**.** *Up to conjugacy in* $\mathrm{USp}(2g)$*, only finitely many* G *satisfy the Sato–Tate axioms.*

Ok that's great, but we should check that these axioms actually hold for the Sato–Tate groups of abelian varieties!

Theorem 50 (FKRS '12)**.** *If the Mumford–Tate conjecture holds for an abelian variety* A *then* $\mathrm{ST}(A)$ *satisfies the Sato–Tate axioms.*

Proposition 51. *Up to conjugacy in* $\mathrm{USp}(2) = \mathrm{SU}(2)$ *there are 3 groups that satisfy the Sato–Tate axioms:*

$$U(1), N(U(1)), \mathrm{SU}(2)$$

where $u \in U(1)$ *is embedded as* $\begin{pmatrix} u & 0 \\ 0 & \overline{u} \end{pmatrix}$.

Proof. $U(1)$ and $\mathrm{SU}(2)$ are the only connected compact Lie groups in $\mathrm{USp}(2) = \mathrm{SU}(2)$ (follows from the classification of such groups). If $G^0 = U(1)$ then $G \subseteq N(U(1))$ (because G^0 is normal in G), but $[N(U(1)) : U(1)] = 2$, so $U(1)$ or $N(U(1))$ are the only options when $G^0 = U(1)$. □

Moreover, these all arise as $\mathrm{ST}(E)$ for some E/K.

Theorem 52. *Up to conjugacy in* USp(4) *there are* 55 *groups that satisfy the Sato–Tate axioms, of which* 6 *are connected (i.e. have $G^0 - G$). The connected ones are*

$$U(1)_2, \mathrm{SU}(2)_2, U(1)\times U(1), U(1)\times \mathrm{SU}(2), \mathrm{SU}(2)\times\mathrm{SU}(2), \text{ and } \mathrm{USp}(4).$$

Remark 53. Why not $U(2)$? Because no continuous homomorphisms $U(1) \to U(2) \subseteq \mathrm{USp}(4)$ satisfies the Hodge axiom Definition 47.

Remark 54. Only 52 of the 55 groups above arise as $\mathrm{ST}(A)$ for some abelian surface A/K. The 3 missing gorups all lie between $U(1)\times U(1)$ and $N(U(1)\times U(1))$.

There is a similar (with larger numbers) theorem for $g = 3$, with a similar story. In particular that there are groups that cannot be realised.

5. Galois Endomorphism Types

Let H be a finite group and B be an $\mathbb{R}$–module equipped with an $\mathbb{R}$-linear H action (i.e. an $\mathbb{R}[H]$-module).

Definition 55. We consider two such pairs, (H, B) and (H', B') to be *equivalent* if there exists an isomorphism $\phi_H : H \cong H'$ and an isomorphism of $\mathbb{R}$-algebras $\phi_B : B \cong B'$ such that

$$\phi_B(b^h) = \Phi_B(b)^{\phi_H(h)}.$$

Let $[H, B]$ denote the equivalece class of (H, B).

Definition 56. The endomorphism field of an abelian variety over a number field A/K to be the minimal L/K such that $\mathrm{End}(A_L) = \mathrm{End}(A_{\overline{K}})$. The Galois group $\mathrm{Gal}(L/K)$ acts $\mathbb{R}$-linearly on $\mathrm{End}(A_L)\otimes_{\mathbb{Z}}\mathbb{R}$. The *Galois endomorphism type* $\mathrm{GT}(A)$ of A is the class

$$\mathrm{GT}(A) = [\mathrm{Gal}(L/K), \mathrm{End}(A_L)\otimes_{\mathbb{Z}}\mathbb{R}].$$

Example 57. If E is an elliptic curve over K, then

- If E has CM over K then $\mathrm{GT}(E) = [C_1, \mathbb{C}]$ and $\mathrm{ST}(A) = U(1)$;
- If E has CM over a quadratic extension L/K then $\mathrm{GT}(E) = [C_2, \mathbb{C}]$ and $\mathrm{ST}(A) = N(U(1))$;
- If E is non-CM then $\mathrm{GT}(E) = [C_1, \mathbb{R}]$ and $\mathrm{ST}(A) = \mathrm{SU}(2)$.

Theorem 58. *For $g \le 3$ there is a one to one correspondence between* GT*'s and* ST*'s, in which*

$$\mathrm{Gal}(L/K) \cong \mathrm{ST}(A)/\mathrm{ST}^0(A)$$

and

$$\mathrm{ST}^0(A) \leftrightarrow \mathrm{End}(A_L)\otimes\mathbb{R}.$$

Remark 59. Not true for $g = 4$.

For Abelian surfaces the classification result is is as follows.

Abelian Surface	$\mathrm{End}(A_L) \otimes \mathbb{R}$	$\mathrm{ST}(A)$
Square of CM E	$M_2(\mathbb{C})$	$U(1)_2$
QM abelian surface or square of non-CM E	$M_2(\mathbb{R})$	$\mathrm{SU}(2)_2$
CM abelian surface or product of non-isogenous CM E	$\mathbb{C} \times \mathbb{C}$	$U(1) \times U(1)$
Product of CM and non-CM E	$\mathbb{C} \times \mathbb{R}$	$U(1) \times \mathrm{SU}(2)$
RM abelian surface or product of non-isogenoous non-CM E	$\mathbb{R} \times \mathbb{R}$	$\mathrm{SU}(2) \times \mathrm{SU}(2)$
Generic abelian surface	$\mathbb{R}$	$\mathrm{USp}(4)$

References

[Ser68] J.-P. Serre, *Abelian l-adic representations and elliptic curves*, W. A. Benjamin, Inc., New York-Amsterdam, 1968. McGill University lecture notes written with the collaboration of Willem Kuyk and John Labute. MR0263823 ↑1, 1

[Sut19] A. Sutherland, *Sato-Tate distributions*, Contemp. Math. **740** (2019), 197–248. ↑1