SATO-TATE DISTRIBUTIONS

ANDREW V. SUTHERLAND

ABSTRACT. These lecture notes are from a course on Sato–Tate distributions delivered at the 2016 Arizona Winter School at the Southwest Center for Arithmetic Geometry. We explore the relationship between Galois representations, motivic L-functions, and Sato-Tate groups, focusing primarily on the case of abelian varieties over number fields.

1. AN INTRODUCTION TO SATO-TATE DISTRIBUTIONS

Before discussing the Sato-Tate conjecture and Sato-Tate distributions in general, let us start in the more familiar setting of Artin motives (otherwise known as the Galois theory of number fields).

1.1. **A first example.** Let $f \in \mathbf{Z}[x]$ be a squarefree polynomial of degree d; for example, we may take $f(x) = x^3 - x + 1$. Since f has integer coefficients, we can reduce them modulo any prime p to obtain a polynomial f_p with coefficients in the finite field $\mathbf{Z}/p\mathbf{Z} \simeq \mathbf{F}_p$. For each prime p define

$$N_f(p) := \#\{x \in \mathbf{F}_p : f_p(x) = 0\},$$

which we note is an integer between 0 and d. We would like to understand how $N_f(p)$ varies with p. The table below shows the values of $N_f(p)$ when $f(x) = x^3 - x + 1$ for $p < 60$:

p:	2	3	5	7	11	13	17	19	23	29	31	37	41	43	47	53	59
$N_f(p)$	0	0	1	1	1	0	1	1	2	0	0	1	0	1	0	1	3

There does not appear to be any obvious pattern (and we should know not to expect one, the Galois group lurking behind the scenes is nonabelian). The prime $p = 23$ is exceptional because it divides $\mathrm{disc}(f)$, which means that $f_{23}(x)$ has a double root. As we are interested in the distribution of $N_f(p)$ as p tends to infinity, we are happy to ignore such primes, which are necessarily finite in number.

Looking at such a small dataset does not tell us much, so let us increase the bound B on the primes p that we are considering and count how often we see $N_f(p) = 0, 1, 2, 3$. Define

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

for $i = 0, 1, 2, 3$. We may then compute the following statistics:

The author was supported by NSF grants DMS-1115455 and DMS-1522526.

B	$c_0(B)$	$c_1(B)$	$c_2(B)$	$c_3(B)$
10^3	0.323353	0.520958	0.005988	0.155689
10^4	0.331433	0.510586	0.000814	0.157980
10^5	0.333646	0.502867	0.000104	0.163487
10^6	0.333185	0.500783	0.000013	0.166032
10^7	0.333360	0.500266	0.000002	0.166373
10^8	0.333337	0.500058	0.000000	0.166605
10^9	0.333328	0.500016	0.000000	0.166656
10^{10}	0.333334	0.500003	0.000000	0.166663
10^{11}	0.333333	0.500001	0.000000	0.166666
10^{12}	0.333333	0.500000	0.000000	0.166666

Based on these statistics we may conjecture that the limiting values of $c_i(B)$ as $B \to \infty$ are

$$c_0 = 1/3, \quad c_1 = 1/2, \quad c_2 = 0, \quad c_3 = 1/6.$$

There is of course a natural motivation for this conjecture (which is in fact a theorem), one that would allows us to correctly predict the asymptotic ratios c_i without needing to compute any statistics. Let us fix an algebraic closure $\overline{\mathbf{Q}}$ of $\mathbf{Q}$. The absolute Galois group $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts on the roots of $f(x)$ by permuting them. This allows us to define the *Galois representation* (a continuous homomorphism)

$$\rho_f : \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_d(\mathbf{C}),$$

whose image is a subgroup of the permutation matrices in $\mathrm{O}_d(\mathbf{C}) \subseteq \mathrm{GL}_d(\mathbf{C})$; here O_d denotes the orthogonal group (we could replace $\mathbf{C}$ with any field of characteristic zero). Note that $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ and $\mathrm{GL}_d(\mathbf{C})$ are topological groups (the former has the Krull topology), and homomorphisms of topological groups are understood to be continuous. In order to associate a permutation of the roots of $f(x)$ to a matrix in $\mathrm{GL}_d(\mathbf{C})$ we need to fix an ordering of the roots; this amounts to choosing a basis for the vector space $\mathbf{C}^d$, which means that our representation ρ_f is really defined only up to conjugacy.

The value ρ_f takes on $\sigma \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ depends only on the restriction of σ to the splitting field L of f, so we could restrict our attention to $\mathrm{Gal}(L/\mathbf{Q})$. This makes ρ_f an *Artin representation*: a continuous representation $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_d(\mathbf{C})$ that factors through a finite quotient (by an open subgroup). But in the more general settings we wish to consider this may not be true, and even when it is, we may not know L; it is thus more convenient to work with $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$.

To facilitate this, we associate to each prime p an *absolute Frobenius element*

$$\mathrm{Frob}_p \in \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$$

which may be defined as follows. Fix an embedding $\overline{\mathbf{Q}}$ in $\overline{\mathbf{Q}}_p$ and use the valuation ideal $\mathfrak{P}$ of $\overline{\mathbf{Q}}_p$ (the maximal ideal of its ring of integers) to define a compatible system of primes $\mathfrak{q}_L := \mathfrak{P} \cap L$, where L ranges over all finite extensions of $\mathbf{Q}$. For each prime $\mathfrak{q}_L$, let $D_{\mathfrak{q}_L} \subseteq \mathrm{Gal}(L/\mathbf{Q})$, denote its decomposition group, $I_{\mathfrak{q}_L} \subseteq D_{\mathfrak{q}_L}$ its inertia group, and $\mathbf{F}_{\mathfrak{q}_L} := \mathbf{Z}_L/\mathfrak{q}_L$ its residue field, where $\mathbf{Z}_L$ denotes the ring of integers of L. Taking the inverse limit of the exact sequences

$$1 \to I_{\mathfrak{q}_L} \to D_{\mathfrak{q}_L} \to \mathrm{Gal}(\mathbf{F}_{\mathfrak{q}_L}/\mathbf{F}_p) \to 1$$

over finite extensions $L/\mathbf{Q}$ gives an exact sequence of profinite groups

$$1 \to I_p \to D_p \to \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p) \to 1.$$

We now pick $\mathrm{Frob}_p \in D_p \subseteq \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ to be any lift of the Frobenius automorphism $x \to x^p$ in $\mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$. Note that we have made two arbitrary choices in our definition of Frob_p, we chose an element in the inverse image of the Frobenius automorphism under $D_p \to \mathrm{Gal}(\overline{\mathbf{F}}_p/\mathbf{F}_p)$, and we picked an embedding of $\overline{\mathbf{Q}}$ into $\overline{\mathbf{Q}}_p$, so Frob_p is in no way canonical, but it certainly exists. Its key property is that if $L/\mathbf{Q}$ is a finite Galois extension in which p is unramified, then the conjugacy class $\mathrm{conj}_L(\mathrm{Frob}_\mathfrak{p})$ in $\mathrm{Gal}(L/\mathbf{Q})$ of the restriction of $\mathrm{Frob}_\mathfrak{p} : \overline{\mathbf{Q}} \to \overline{\mathbf{Q}}$ to L is uniquely determined, independent of the choices we made. One can think of $\mathrm{Frob}_\mathfrak{p}$ as defining a map $L \to \mathrm{conj}_L(\mathrm{Frob}_\mathfrak{p})$ that assigns to each finite Galois extension $L/\mathbf{Q}$ the conjugacy class of $\mathrm{Gal}(L/\mathbf{Q})$ corresponding to the Frobenius automorphism when p is unramified in L. Everything we have said applies *mutatis mutandi* if we replace $\mathbf{Q}$ by a number field K: put $\overline{K} := \overline{\mathbf{Q}}$, replace p by a prime $\mathfrak{p}$ of K (by which we mean a nonzero prime ideal of $\mathbf{Z}_K$), and replace $\mathbf{F}_p$ by the residue field $\mathbf{F}_\mathfrak{p} := \mathbf{Z}_K/\mathfrak{p}$.

We now make the following observation: for any prime p that does not divide $\mathrm{disc}\, f$ we have

$$N_f(p) = \mathrm{tr}\, \rho_f(\mathrm{Frob}_p). \tag{1}$$

This follows from the fact that the trace of a permutation matrix counts its fixed points. Since p is unramified, the inertia group $I_p \subseteq \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ acts trivially on the roots of $f(x)$, and the action of Frob_p on the roots of $f(x)$ coincides (up to conjugation) with the action of the Frobenius automorphism $x \to x^p$ on the roots of $f_p(x)$, both of which are described by the permutation matrix $\rho(\mathrm{Frob}_p)$. The Chebotarev density theorem implies that we can compute c_i by applying (1) and simply counting the number of matrices in $\rho_f(\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}))$ that have trace i; it is enough to determine the trace and size of each conjugacy class.

Theorem 1.1. CHEBOTAREV DENSITY THEOREM *Let L/K be a finite Galois extension of number fields with Galois group $G := \mathrm{Gal}(L/K)$. For every subset C of G stable under conjugation we have*

$$\lim_{B\to\infty} \frac{\#\{N(\mathfrak{p}) \le B : \mathrm{conj}_L(\mathrm{Frob}_\mathfrak{p}) \subseteq C\}}{\#\{N(\mathfrak{p}) \le B\}} = \frac{\#C}{\#G},$$

where $\mathfrak{p}$ ranges over primes of K and $N(\mathfrak{p}) := \#\mathbf{F}_\mathfrak{p}$ is the cardinality of the residue field $\mathbf{F}_\mathfrak{p} := \mathbf{Z}_K/\mathfrak{p}$.

Proof. See Lecture 2. □

Remark 1.2. One may take C to be a single conjugacy class (the general result follows easily from this case). The asymptotic ratio that appears in the theorem depends only on degree-1 primes (those with prime residue field), since these make up all but a negligible proportion of the primes $\mathfrak{p}$ for which $N(\mathfrak{p}) \le B$ (this follows from earlier density results that are easy to prove). In our statement of the theorem we do not exclude primes of K that are ramified in L because they are finite in number and no matter what value $\mathrm{conj}_L(\mathrm{Frob}_\mathfrak{p})$ takes on these primes it will not change the limiting ratio.

In our example with $f(x) = x^3 - x + 1$, one finds that $G_f := \rho_f(\overline{\mathbf{Q}}/\mathbf{Q})$ is isomorphic to S_3 (the Galois group of its splitting field), and its three conjugacy classes, represented by the matrices

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

have traces 0, 1, 3 and sizes 2, 3, 1, respectively. It follows that

$$c_0 = 1/3, \quad c_1 = 1/2, \quad c_2 = 0, \quad c_3 = 1/6,$$

just as we conjectured.

If we endow the group G_f with the discrete topology it becomes a compact group, and therefore has a *Haar measure* μ that is uniquely determined once we normalize it so that $\mu(G_f) = 1$ (which we always do). Recall that the Haar measure of a compact group G is a translation-invariant Radon measure (so $\mu(gS) = \mu(Sg) = \mu(S)$ for any measurable set S and $g \in G$), and is unique up to a scaling.[1] For finite groups the Haar measure μ is just the normalized counting measure. We can compute the expected value of trace (and many other statistical quantities of interest) by integrating against the Haar measure, which in this case amounts to summing over the finite group G_f:

$$\mathrm{E}[\mathrm{tr}] = \int_{G_f} \mathrm{tr}\,\mu = \frac{1}{\#G_f}\sum_{g\in G_f}\mathrm{tr}(g) = \sum_{i=0}^{d} c_i i.$$

The Chebotarev density theorem implies that this is also the average value of $N_f(p)$, that is,

$$\lim_{B\to\infty}\frac{\sum_{p\le B} N_f(p)}{\sum_{p\le B} 1} = \mathrm{E}[\mathrm{tr}].$$

This average is 1 in our example, because $f(x)$ is irreducible; see Exercise 1.1.

The quantities c_i define a probability distribution on the set $\{\mathrm{tr}(g) : g \in G_f\}$ traces that we can also view as a probability distribution on the set $\{N_f(p) : p \text{ prime}\}$. Picking a random prime p in some large interval $[1, B]$ and computing $N_f(p)$ is the same thing as picking a random matrix g in H_f and computing $\mathrm{tr}(g)$. More precisely, the sequence $(N_f(p))_p$ indexed by primes p is *equidistributed* with respect to the pushforward of the Haar measure μ under the trace map. We will discuss the notion of equidistribution more generally in Lecture 2.

1.2. **Moment sequences.** There is another way to characterize the probability distribution on $\mathrm{tr}(g)$ given by the c_i; we can compute its *moment sequence*:

$$\mathrm{M}[\mathrm{tr}] := (\mathrm{E}[\mathrm{tr}^n])_{n\ge 0},$$

where

$$\mathrm{E}[\mathrm{tr}^n] = \int_{G_f} \mathrm{tr}^n\,\mu.$$

It might seem silly to include the zeroth moment $\mathrm{E}[\mathrm{tr}^0] = \mathrm{E}[1] = 1$, but in Lecture 4 we will see that this convention is useful. In our example we have the moment sequence

$$\mathrm{M}[\mathrm{tr}] = (1, 1, 2, 5, 14, 41, \ldots, \tfrac{1}{2}(3^{n-1}+1), \ldots).$$

The sequence $\mathrm{M}[\mathrm{tr}]$ uniquely determines[2] the distributions of traces and thus captures all the information encoded in the c_i. It may not seem very useful to replace a finite set of rational numbers with an infinite sequence of integers, but when dealing with continuous probability distributions (which we are forced to do as soon as we leave our weight zero setting), it is a convenient tool.

If we pick another cubic polynomial $f \in \mathbf{Z}[x]$, we will typically obtain the same result as we did in our example; when ordered by height almost all cubic polynomials f have Galois group $G_f \simeq S_3$. But there are exceptions: if f is not irreducible over $\mathbf{Q}$ then G_f will certainly be isomorphic to a proper subgroup of S_3, and this also occurs when the splitting field of f is a cyclic cubic extension (this happens

[1] For locally compact groups G one distinguishes left and right Haar measures, but the two coincide when G is compact; see [21] for more background on Haar measures.

[2] Not all moment sequences uniquely determine an underlying probability distribution, but all the moment sequence we shall consider do (they satisfy *Carleman's condition* [50, p. 126], for example).

whenever the discriminant of $f(x)$ is a square; take $f(x) = x^3 - 3x - 1$ for example). Up to conjugacy there are four subgroups of S_3, and each corresponds to a different distribution of $N_f(p)$, as shown in the table below:

$f(x)$	G_f	c_0	c_1	c_2	c_3	M[tr]
$x^3 - x$	1	0	0	0	1	$(1,3,9,27,81,\ldots)$
$x^3 + x$	C_2	0	1/2	0	1/2	$(1,2,5,14,41,\ldots)$
$x^3 - 3x - 1$	C_3	2/3	0	0	1/3	$(1,1,3,19,27,\ldots)$
$x^3 - x + 1$	S_3	1/3	1/2	0	1/6	$(1,1,2,\ 5,14,\ldots)$

One can do the same thing with polynomials of degree $d > 3$. For $d \le 19$ the results are exhaustive: for every transitive subgroup G of S_d the database of Klüners and Malle [49] contains at least one polynomial $f \in \mathbf{Z}[x]$ with $G_f \simeq G$ (including all 1954 transitive subgroups of S_{16}). The non-transitive cases can be constructed as products (of groups and of polynomials) of transitive cases of lower degree. It is an open question whether this can be done for all d (even in principle). This amounts to a strong form of the inverse Galois problem over $\mathbf{Q}$; we are asking not only whether every finite group can be realized as a Galois group over $\mathbf{Q}$, but whether every permutation group of degree d can be realized as the Galois group of the splitting field of a polynomial of degree d.

1.3. **Zeta functions.** For polynomials f of degree $d = 3$ there is a one-to-one correspondence between subgroups of S_d and distributions of $N_f(p)$. This is not true for $d \ge 4$. For example, the polynomials $f(x) = x^4 - x^3 + x^2 - x + 1$ with $G_f \simeq C_4$ and $g(x) = x^4 - x^2 + 1$ with $G_g \simeq C_2 \times C_2$ both have $c_0 = 3/4$, $c_1 = c_2 = c_3 = 0$, and $c_4 = 1/4$, corresponding to the moment sequence $\mathrm{M}[\mathrm{tr}] = (1,1,4,16,64,\ldots)$.

We can distinguish these cases if, in addition to considering the distribution of $N_f(p)$, we also consider the distribution of

$$N_f(p^r) := \#\{x \in \mathbf{F}_{p^r} : f_p(x) = 0\}$$

for integers $r \ge 1$. In our quartic example we have $N_g(p^2) = 4$ for almost all p, whereas $N_f(p^2)$ is 4 or 2 depending on whether p is a square modulo 5 or not. In terms of the matrix group G_f we have

$$N_f(p^r) = \mathrm{tr}\big(\rho_f(\mathrm{Frob}_p)^r\big) \tag{2}$$

for all primes p that do not divide $\mathrm{disc}\, f$. To see this, note that the permutation matrix $\rho_f(\mathrm{Frob}_p)^r$ corresponds to the permutation of the roots of $f_p(x)$ given by the rth power of the Frobenius automorphism $x \mapsto x^p$. Its fixed points are precisely the roots of $f_p(x)$ that lie in $\mathbf{F}_{p^r}$; taking the trace counts these roots, which is, by definition $N_f(p^r)$.

This naturally leads to the definition of the local *zeta function* of f at p:

$$Z_{f_p}(T) := \exp\left(\sum_{r=1}^{\infty} N_f(p^r)\frac{T^r}{r}\right), \tag{3}$$

which can be viewed as a generating function for the sequence $(N_f(p), N_f(p^2), N_f(p^3), \ldots)$. This particular form of generating function may seem strange when first encountered, but it has some very nice properties. For example, if $f, g \in \mathbf{Z}[x]$ are squarefree polynomials with no common factor, then their product fg is also square free, and for all $p \nmid \mathrm{disc}(fg)$ we have

$$Z_{f_p g_p} = Z_{f_p} Z_{g_p}.$$

Remark 1.3. The identity (2) can be viewed as a special case of the Grothendieck-Lefschetz trace formula. It allows us to express the zeta function $Z_{f_p}(T)$ as a sum over powers of the traces of the image of Frob_p under the Galois representation ρ_f. In general one considers the trace of the Frobenius endomorphism acting on étale cohomology, but in dimension zero the only relevant cohomology is H^0.

While defined as a power series, in fact $Z_{f_p}(T)$ is a rational function of the form

$$Z_{f_p}(T) = \frac{1}{L_p(T)}$$

where $L_p(T)$ is an integer polynomial whose roots lie on the unit circle. This can be viewed as a consequence of the Weil conjectures in dimension zero,[3] but in fact it follows directly from (2). Indeed, for any matrix $A \in \mathrm{GL}_d(\mathbf{C})$ we have the identity

$$\exp\left(\sum_{r=1}^{\infty} \mathrm{tr}(A^r)\frac{T^r}{r}\right) = \det(1-AT)^{-1}, \tag{4}$$

which can be proved by expressing the coefficients on both sides as symmetric functions in the eigenvalues of A; see Exercise 1.2. Applying (2) and (4) to the definition of $Z_{f_p}(T)$ in (3) yields

$$Z_{f_p}(T) = \frac{1}{\det(1-\rho_f(\mathrm{Frob}_p)T)},$$

thus

$$L_p(T) = \det(1-\rho_f(\mathrm{Frob}_p)T).$$

The polynomial $L_p(T)$ is precisely the polynomial that appears in the Euler factor at p of the (partial) *Artin L-function* $L(\rho_f, s)$ for the representation ρ_f:

$$L(\rho_f, s) := \prod_p L_p(p^{-s})^{-1},$$

at least for primes p that do not divide $\mathrm{disc}(f)$; for the definition of the Euler factors at ramified primes (and the Gamma factors at archimedean places), see [58, Ch. 2]. We shall not be concerned with the Euler factors at ramified primes, other than to note that they are all holomorphic and nonvanishing. We should note that the L-function $L(\rho_f, s)$ is not primitive, because ρ_f is not irreducible; one can always remove at least a factor of $\zeta(s)$ (the Riemann zeta function).

Returning to our interest in equidistribution, the Haar measure μ on $G_f = \rho_f(\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}))$ allows us to determine the distribution of L-polynomials $L_p(T)$ that we see as p varies. Each polynomial $L_p(T)$ is the *reciprocal polynomial* (obtained by reversing the coefficients) of the characteristic polynomial of $\rho_f(\mathrm{Frob}_p)$. If we fix a polynomial $P(T)$ of degree $d = \deg f$, and pick a prime p at random from some large interval, the probability that $L_p(T) = P(T)$ is equal to the probability that the reciprocal polynomial $T^d P(1/T)$ is the characteristic polynomial of a random element of G_f (this probability will be zero unless $P(T)$ has a particular form; see Exercise 1.3).

Remark 1.4. For $d \le 5$ the distribution of characteristic polynomials uniquely determines each subgroup of S_d (up to conjugacy). This is not true for $d \ge 6$, and for $d \ge 8$ one can find non-isomorphic subgroups of S_d with the same distribution of characteristic polynomials; the transitive permutation groups 8T10 and 8T11 which arise for $x^8 - 13x^6 + 44x^4 - 17x^2 + 1$ and $x^8 - x^5 - 2x^4 + 4x^2 + x + 1$ (respectively) are an example.

[3] Provided one accounts for the fact that $f(x) = 0$ does not define an irreducible variety unless $\deg(f) = 1$; in this case $N_f(p^r) = 1$ and $L_p(T) = 1 - T$, which is consistent with the usual formulation of the Weil conjectures (see Theorem 1.7).

1.4. **Computing zeta functions in dimension zero.** Let us now briefly address the practical question of how one goes about efficiently computing the zeta function $Z_{f_p}(T)$, which amounts to computing the integer polynomial $L_p(T)$. It suffices to compute the integers $N_f(p^r)$ for $r \le d$, which is equivalent to determining the degrees of the irreducible polynomials appearing in the factorization of $f_p(x)$ in $\mathbf{F}_p[x]$; these determine the cycle type, and therefore the conjugacy class, of the permutation of the roots of $f_p(x)$ induced by the action of the Frobenius automorphism $x \mapsto x^p$, which in turn determines the characteristic polynomial of $\rho_f(\mathrm{Frob}_p)$ and the L-polynomial $L_p(T) = \det(1 - \rho_f(\mathrm{Frob}_p)T)$; see Exercise 1.3. To determine the factorization pattern of $f_p(x)$ we may apply the following algorithm.

Algorithm 1.5. Given a polynomial $g \in \mathbf{F}_p[x]$ of degree d, compute the number n_i of degree i factors of g in $\mathbf{F}_q[x]$ for $1 \le i \le d$ as follows:

1. Let $g_1(x)$ be $g(x)$ made monic.
2. For i from 1 to d:
   a. If $\deg(g_i) < i$ then put $n_j := 0$ for $i \le j \le d$ and proceed to step 3.
   b. Using binary exponentiation in the ring $\mathbf{F}_p[x]/(g_i)$, compute $r_i(x) := x^{p^i} \bmod g_i(x)$.
   c. Compute $h_i(x) := \gcd(g_i, r_i(x) - x) = \gcd(g_i(x), x^{p^i} - x)$ using the Euclidean algorithm.
   d. Compute $n_i := \deg(h_i)/i$ and $g_{i+1} := g_i/h_i$ using exact division.
3. Output $n_1, \dots, n_d$.

Algorithm 1.5 makes repeated use of the fact that the polynomial

$$x^{p^i} - x = \prod_{a \in \mathbf{F}_{p^i}} (x - a)$$

is equal to the product of all the monic degree i polynomials in $\mathbf{F}_p[x]$. By removing irreducible factors from $g(x)$ in increasing order by degree we ensure that every degree i factor of $g_i(x)$ is irreducible. Using fast algorithms for integer and polynomial arithmetic and the fast Euclidean algorithm (see [28, §8-11], for example), one can show that this algorithm uses $O((d \log p)^{2+o(1)})$ bit operations, a running time that is quasi-quadratic in the $O(d \log p)$ bit-size of its input $g \in \mathbf{F}_p[x]$. In practical terms, it is extremely efficient. For example, the table of $c_i(B)$ values for our example polynomial $f(x) = x^3 - x + 1$ took less than two minutes to create using the `smalljac` software library [47, 81], which includes an efficient implementation of basic finite field arithmetic. The NTL [76] and FLINT [32, 33] libraries provide similar (and more comprehensive) functionality; the FLINT library is included in Sage [65].

Remark 1.6. Note that Algorithm 1.5 does *not* output the factorization of $g(x)$, just the degrees of its irreducible factors. The algorithm can be extended to a *probabilistic* algorithm that outputs the complete factorization of $g(x)$, see [28, Alg. 14.8], with an expected running time that is quasi-quadratic. However, no *deterministic* polynomial-time algorithm for factoring polynomials over finite fields is known, not even in the case $d = 2$. This is a famous open problem. One approach to solving it is to first prove the generalized Riemann hypothesis (GRH), which would at least address the case $d = 2$ and many others, but it is not known whether the GRH is sufficient to address all cases.[4]

1.5. **Arithmetic schemes.** We now want to generalize our first example. Let us replace the equation $f(x) = 0$ with an *arithmetic scheme* X, a scheme of finite type over $\mathbf{Z}$; the case we have been considering is $X = \mathrm{Spec} A$, where $A = \mathbf{Z}[x]/(f)$. For each prime p the fiber X_p of $X \to \mathrm{Spec}\,\mathbf{Z}$ is a scheme of finite type over $\mathbf{F}_p$, and we let $N_X(p) := X_p(\mathbf{F}_p)$ count its $\mathbf{F}_p$-points; equivalently, we may define $N_X(p)$ as

[4] On the plus side, if you succeed with this first step the Clay institute will help fund the remaining work.

the number of closed points (maximal ideals) of X whose residue field has cardinality p, and similarly define $N_X(q)$ for prime powers $q = p^r$. The local zeta function of X at p is then defined as

$$Z_{X_p}(T) := \exp\left(\sum_{r=1}^{\infty} N_X(p^r)\frac{T^r}{r}\right).$$

These local zeta functions can then be combined to into a single *arithmetic zeta-function*

$$\zeta_X(s) := \prod_p Z_{X_p}(p^{-s}).$$

In our example with $X = \operatorname{Spec}\mathbf{Z}[x]/(f)$, the zeta function $\zeta_X(s)$ coincides with the Artin L-function $L(\rho_f, s) = \prod L_p(s)^{-1}$ up to a finite set of factors at primes p that divide $\operatorname{disc}(f)$.

The definitions above generalize to any number field K: replace $\mathbf{Q}$ by K, replace $\mathbf{Z}$ by $\mathbf{Z}_K$, replace p by a prime $\mathfrak{p}$ of K (nonzero prime ideal of $\mathbf{Z}_K$), replace $\mathbf{F}_p \simeq \mathbf{Z}/p\mathbf{Z}$ by the residue field $\mathbf{F}_\mathfrak{p} := \mathbf{Z}_K/\mathfrak{p}$. When considering questions of equidistribution we order primes $\mathfrak{p}$ by their norm $N(\mathfrak{p}) := \mathbf{F}_\mathfrak{p}$ (we may break ties arbitrarily), so that rather that summing over $p \le B$ we sum over $\mathfrak{p}$ for which $N(\mathfrak{p}) \le B$.

1.6. **A second example.** We now leave the world of Artin motives (motives of weight 0) and consider our first example in weight 1, an elliptic curve $E/\mathbf{Q}$, which is the setting in which the Sato–Tate conjecture was originally formulated. Such a curve can always be written in the form

$$E\colon y^2 = x^3 + Ax + B,$$

with $A, B \in \mathbf{Z}$. This equation is understood to define a smooth projective curve in $\mathbf{P}^2$ (homogenize the equation by introducing a third variable z), which has a single projective point $P_\infty := (0 : 1 : 0)$ at infinity that we take as the identity element of the group law. Recall that an elliptic curve is not just a curve, it comes equipped with a distinguished rational point; after applying a suitable automorphism of $\mathbf{P}^2$ we can always take this to be the point P_∞.

The group operation on E can be defined via the usual chord-and-tangent law (three points on a line sum to zero), which can be used to derive explicit formulas with coefficients in $\mathbf{Q}$, or in terms of the divisor class group $\operatorname{Pic}^0(E)$ (divisors of degree zero modulo principal divisors), in which every divisor class can be uniquely represented by a divisor of the form $P - P_\infty$, where P is a point on the curve. This latter view is more useful in that it easily generalizes to curves of genus $g > 1$, whereas the chord-and-tangent law does not. The Abel–Jacobi map $P \mapsto P - P_\infty$ gives a bijection between points on E and points on $\operatorname{Jac}(E)$ that commutes with the group operation, so the two approaches are isomorphic.

For each prime p that does not divide the discriminant $\Delta := -16(4A^3 + 27B^2)$ we can reduce our equation for E modulo p to obtain an elliptic curve $E_p/\mathbf{F}_p$; in this case we say that p is a *prime of good reduction for* E (or simply a *good prime*). We should note that the discriminant Δ is not necessarily minimal, the curve E may have another model that has good reduction at primes that divide Δ (including the prime 2), but we are happy to ignore any finite set of primes, including all those that divide Δ.[5]

For every prime p of good reduction for E we have

$$N_E(p) := \#E_p(\mathbf{F}_p) = p + 1 - t_p,$$

where the integer t_p satisfies the *Hasse-bound* $|t_p| \le 2\sqrt{p}$. In contrast to our first example, the integers $N_E(p)$ now tend to infinity with p: we have $N_E(p) = p + 1 + O(\sqrt{p})$. In order to study how the error

[5]All elliptic curves over $\mathbf{Q}$ have a global minimal model for which the primes of bad reduction are precisely those that divide the discriminant, but this model is not necessarily of the form $y^2 = x^3 + Ax + B$. Over general number fields global minimal models do not always exist (they do when the class number is one).

term varies with p we want to consider the normalized value

$$x_p := t_p/\sqrt{p} \in [-2,2].$$

We are now in a position to conduct the following experiment: given an elliptic curve $E/\mathbf{F}_p$, compute x_p for all good primes $p \le B$ and see how the x_p are distributed over the real interval $[-2,2]$.

One can see an example for the elliptic curve $E : y^2 = x^3+x+1$ in Figure 1, which shows a histogram with the x-axis ranging over the interval $[-2,2]$. This interval is subdivided into approximately $\sqrt{\pi(B)}$ subintervals, each of which contains a colored bar whose height is proportional to the number of x_p (for $p \le B$) that lie in the subinterval. The gray line shows the height of the uniform distribution for scale (note that the vertical and horizontal scales are not the same, they were chosen judiciously). For $0 \le n \le 10$, the moment statistics

$$M_n := \frac{\sum_{p\le B} x_p^n}{\sum_{p\le B} 1},$$

are shown below the histogram. They appear to be converging to $1, 0, 1, 0, 2, 0, 5, 0, 14, 0, 42$, which is the start of sequence A126120 in the Online Encyclopedia of Integer Sequences (OEIS) [62]).

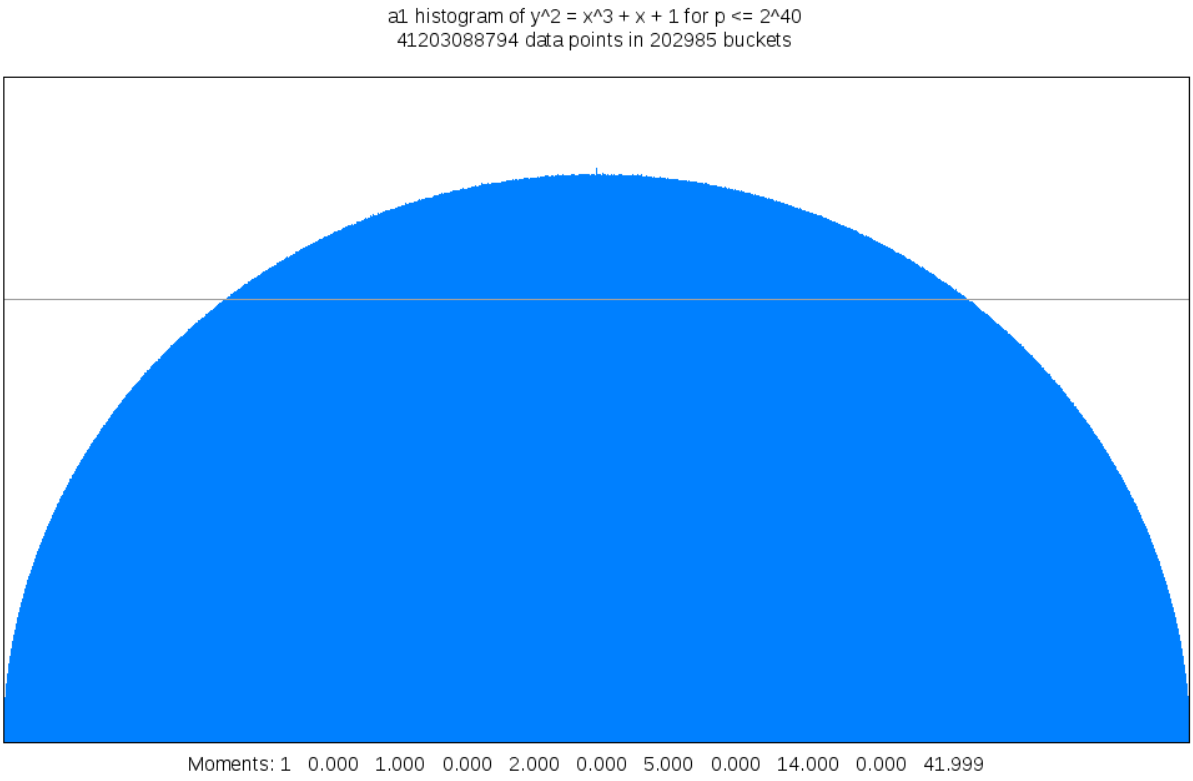


FIGURE 1. Click image to animate (requires Adobe Reader), or visit this web page.

The Sato–Tate conjecture for elliptic curves over $\mathbf{Q}$ (now a theorem) implies that for almost all $E/\mathbf{Q}$, whenever we run this experiment we will see the same asymptotic distribution of Frobenius traces that is visible in the figure above (and the same limiting sequence of moments). In order to make this precise we would like to explain where the conjectured distribution comes from. In our first example we had a compact matrix group G_f associated to the scheme $X = \operatorname{Spec} \mathbf{Z}[x]/(f)$ whose Haar measure

governed the distribution of $N_f(p)$. In fact we showed that more is true: there is a direct relationship between characteristic polynomials of elements of G_f and the L-polynomials $L_p(T)$ that appear in the local zeta functions $Z_{f_p}(T)$.

The same is true here. In order to identify a candidate group G_E whose Haar measure controls the distribution of normalized Frobenius traces x_p we need to look at the local zeta functions $Z_{E_p}(T)$. Let us recall what the Weil conjectures [92] (proved by Deligne [17, 18]) tell us about the zeta function of a variety over a finite field. The case of one-dimensional varieties (curves) was proved by Weil [90], who also proved an analogous result for abelian varieties [91], and this covers all the cases we shall consider in these lectures, but let us state the general result. Recall that for a compact manifold X over $\mathbf{C}$, the *Betti number* b_i is the rank of the singular homology group $H_i(X,\mathbf{Z})$, and the *Euler characteristic* χ of X is defined by $\chi := \sum(-1)^i b_i$.

Theorem 1.7 (Weil Conjectures)**.** *Let X be a geometrically irreducible non-singular projective variety of dimension n defined over a finite field $\mathbf{F}_q$ and define the zeta function*

$$Z_X(T) := \exp\left(\sum_{r=1}^{\infty} N_X(q^r)\frac{T^r}{r}\right),$$

where $N_X(q^r) := \#X(\mathbf{F}_{q^r})$. The following hold:

(1) ***Rationality**: $Z_X(T)$ is a rational function of the form*

$$Z_X(T) = \frac{P_1(T)\cdots P_{2n-1}(T)}{P_0(T)\cdots P_{2n}(T)},$$

with $P_i \in 1 + T\mathbf{Z}[T]$.

(2) ***Functional Equation**: the roots of $P_i(T)$ are the same as the roots of $T^{\deg P_{2n-j}} P_{2n-j}(1/(q^n T))$.*[6]

(3) ***Riemann Hypothesis**: the complex roots of $P_i(T)$ all have absolute value $q^{-i/2}$.*

(4) ***Betti Numbers**: if X is the reduction of a non-singular variety Y defined over a number field $K \subseteq \mathbf{C}$, then the degree of P_i is equal to the Betti number b_i of $Y(\mathbf{C})$.*

The curve E_p is a curve of genus $g = 1$, so we may apply the Weil conjectures in dimension $n = 1$, with Betti numbers $b_0 = b_2 = 1$ and $b_1 = 2g = 2$. This implies that its zeta function has the form

$$Z_{E_p}(T) = \frac{L_p(T)}{(1-T)(1-pT)}, \tag{5}$$

where $L_p \in \mathbf{Z}[T]$ is a polynomial of the form

$$L_p(T) = pT^2 + c_1 T + 1,$$

with $|c_1| \le 2\sqrt{p}$ (by the Riemann Hypothesis). If we expand both sides of (5) as power series in $\mathbf{Z}[[T]]$ we obtain

$$1 + N_E(p)T^2 + \cdots = 1 + (p + 1 + c_1)T + \cdots,$$

so we must have $N_E(p) = p + 1 + c_1$, and therefore

$$c_1 = N_E(p) - p - 1 = -t_p.$$

It follows that the integer $N_E(p)$ determines the zeta function $Z_{E_p}(T)$.

[6] Moreover, one has $Z_X(T) = \pm q^{-n\chi/2} T^{-\chi} Z_X(1/(q^n T))$, where χ is the Euler characteristic of X, which is defined as the intersection number of the diagonal with itself in $X \times X$.

Corresponding to our normalization $x_p = t_p/\sqrt{p}$, we define the *normalized L-polynomial*

$$\bar{L}_p(T) := L_p(T/\sqrt{p}) = T^2 + a_1 T + 1,$$

where $a_1 = c_1/\sqrt{p} = -x_p$ is a real number in the interval $[-2,2]$ and the roots of $\bar{L}_p(T)$ lie on the unit circle. In our first example we obtained the group G_f as a subgroup of permutation matrices in $\mathrm{GL}_d(\mathbf{C})$. Here we want a subgroup of $\mathrm{GL}_2(\mathbf{C})$ whose elements have eigenvalues that are

(i) inverses (by the functional equation);
(ii) on the unit circle (by the Riemann hypothesis).

Constraint (i) makes it clear that every element of G_E should have determinant 1, so $G_E \subseteq \mathrm{SL}_2(\mathbf{C})$. Constraints (i) and (ii) together imply that in fact $G_E \subseteq \mathrm{SU}(2)$. As in the weight zero case, we expect that G_E should in general be as large as possible, that is, $G_E = \mathrm{SU}(2)$.

We now consider what it means for an elliptic curve to be generic.[7] Recall that the endomorphism ring of an elliptic curve E necessarily contains a subring isomorphic to $\mathbf{Z}$, corresponding to the multiplication-by-n maps $P \mapsto nP$. Here

$$nP = P + \cdots + P$$

denotes repeated addition under the group law, and we take the additive inverse if n is negative. For elliptic curves over fields of characteristic zero, this typically accounts for all the endomorphisms, but in special cases the endomorphism ring may be larger, in which case it contains elements that are not multiplication-by-n maps but can be viewed as "multiplication-by-α" maps for some $\alpha \in \mathbf{C}$. One can show that the characteristic polynomials of these extra endomorphisms are necessarily quadratic, with negative discriminants, so such an α necessarily lies in an imaginary quadratic field K, and in fact $\mathrm{End}(E) \otimes_{\mathbf{Z}} \mathbf{Q} \simeq K$. When this happens we say that E has *complex multiplication* (CM) by K (or more precisely, by the order in $\mathbf{Z}_K$ isomorphic to $\mathrm{End}(E)$).

We can now state the Sato-Tate conjecture, as independently formulated in the mid 1960's by Mikio Sato (based on numerical data) and John Tate (as an application of what is now known as the Tate conjecture [84]), and finally proved in the late 2000's [5, 6, 31].

Theorem 1.8 (Sato–Tate conjecture)**.** *Let $E/\mathbf{Q}$ be an elliptic curve without CM. The sequence of normalized Frobenius traces x_p associated to E is equidistributed with respect to the pushforward of the Haar measure on* $\mathrm{SU}(2)$ *under the trace map. In particular, for every subinterval $[a,b]$ of $[-2,2]$ we have*

$$\lim_{B\to\infty} \frac{\#\{p \le B : x_p \in [a,b]\}}{\#\{p \le B\}} = \frac{1}{2\pi}\int_a^b \sqrt{4-t^2}\,dt.$$

We have not defined x_p for primes of bad reduction, but there is no need to do so; this theorem is purely an asymptotic statement. To see where the expression in the integral comes from, we need to understand the Haar measure on $\mathrm{SU}(2)$ and its pushforward onto the set of conjugacy classes $\mathrm{conj}(\mathrm{SU}(2))$ (in fact we only care about the latter). A conjugacy class in $\mathrm{SU}(2)$ can be described by an *eigenangle* $\theta \in [0,\pi]$; its eigenvalues are then $e^{\pm i\theta}$ (a conjugate pair on the unit circle, as required). In terms of eigenangles, the pushforward of the Haar measure to $\mathrm{conj}(\mathrm{SU}(2))$ is given by

$$\mu = \frac{2}{\pi}\sin^2\theta\,d\theta$$

[7] The criterion given here in terms of endomorphism rings suffices for elliptic curves (and curves of genus $g \le 3$ or abelian varieties of dimension $g \le 3$), but in general one wants the Galois image to be as large as possible, which is a strictly stronger condition for $g > 3$. This issue will be discussed further in Lecture 3.

(see Exercise 2.4), and the trace is $t = 2\cos\theta$; from this one can deduce the trace measure $\frac{1}{2\pi}\sqrt{4-t^2}dt$ on $[-2,2]$ that appears in Theorem 1.8. We can also use the Haar measure to compute the nth moment of the trace

$$\mathrm{E}[t^n] = \frac{2}{\pi}\int_0^\pi (2\cos\theta)^n \sin^2\theta\, d\theta = \begin{cases} 0 & \text{if } n \text{ is odd,} \\ \frac{1}{m+1}\binom{2m}{m} & \text{if } n = 2m \text{ is even,} \end{cases} \tag{6}$$

and find that the $2m$th moment is the mth Catalan number.[8]

1.7. **Exercises.**

Exercise 1.1. Let $f \in \mathbf{Z}[x]$ be a nonconstant squarefree polynomial. Prove that the average value of $N_f(p)$ over $p \le B$ converges to the number of irreducible factors of f in $\mathbf{Z}[x]$ as $B \to \infty$.

Exercise 1.2. Prove that the identity in (4) holds for all matrices $A \in \mathrm{GL}_d(\mathbf{C})$.

Exercise 1.3. Let $f_p \in \mathbf{F}_p[x]$ denote a squarefree polynomial of degree $d > 0$ and let $L_p(T)$ denote the denominator of the zeta function $Z_{f_p}(T)$. We know that the roots of $L_p(T)$ lie on the unit circle in the complex plane; show that in fact each is an nth root of unity for some $n \le d$. Then give a one-to-one correspondence between (1) cycle-types of degree-d permutations, (2) possible factorization patterns of f_p in $\mathbf{F}_p[x]$, and (3) the possible polynomials $L_p(T)$. Explain why non-conjugate elements of $\rho_f(\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}))$ may have the same characteristic polynomial (give an explicit example).

Exercise 1.4. Construct a (not necessarily irreducible) quintic polynomial $f \in \mathbf{Z}[x]$ with no roots in $\mathbf{Q}$ for which $f_p(x)$ has a root in $\mathbf{F}_p$ for every prime p (hint: think about what its Galois group must be). Compute $c_0, \dots, c_5$ and G_f.

Exercise 1.5. Let X be the arithmetic scheme $\operatorname{Spec}\mathbf{Z}[x,y]/(f,g)$, where

$$f(x,y) := y^2 - 2x^3 + 2x^2 - 2x - 2, \qquad g(x,y) := 4x^2 - 2xy + y^2 - 2.$$

By computing $Z_{X_p}(T) = L_p(T)^{-1}$ for sufficiently many small primes p, construct a list of the polynomials $L_p \in \mathbf{Z}[T]$ that you believe occur infinitely often, and estimate their relative frequencies. Use this data to derive a candidate for the matrix group $G_X := \rho_X(\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$, where ρ_X is the Galois representation defined by the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on $X(\overline{\mathbf{Q}})$. You may want to use of computer algebra system such as Sage [65] or Magma or [10] to facilitate these calculations.

[8] This gives yet another way to define the Catalan numbers, one that does not appear to among the 214 combinatorial interpretations enumerated in [80].

2. Equidistribution, L-functions, and the Sato-Tate conjecture for elliptic curves

2.1. **Equidistribution.** Let us now formally define the notion of equidistribution, following [68, §1A]. For a compact Hausdorff space X, we use $C(X)$ to denote the Banach space of complex-valued continuous functions $f : X \to \mathbf{C}$ equipped with the sup-norm $\|f\| := \sup_{x\in X} |f(x)|$. The space $C(X)$ is closed under pointwise addition and multiplication and contains all constant functions; it is thus a commutative $\mathbf{C}$-algebra with unit $\mathbb{1}_X$ (the function $x \mapsto 1$).[9] For any $\mathbf{C}$-valued functions f and g (continuous or not), we write $f \le g$ whenever f and g are both $\mathbf{R}$-valued and $f(x) \le g(x)$ for all $x \in X$; in particular, $f \ge 0$ means $\mathrm{im}(f) \subseteq \mathbf{R}_{\ge 0}$. The subset of $\mathbf{R}$-valued functions in $C(X)$ form a distributive lattice under this order relation.

Definition 2.1. A (positive normalized Radon) *measure* on a compact Hausdorff space X is a continuous $\mathbf{C}$-linear map $\mu\colon C(X) \to \mathbf{C}$ that satisfies $\mu(f) \ge 0$ for all $f \ge 0$ and $\mu(\mathbb{1}_X) = 1$.

Example 2.2. For each point $x \in X$ the map $f \mapsto f(x)$ defines the *Dirac measure* δ_x.

The value of μ on $f \in C(X)$ is often denoted using integral notation

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

and we shall use the two interchangeably.[10]

Having defined the measure μ as a function on $C(X)$, we would like to use it to assign values to (at least some) subsets of X. It is tempting to define the measure of a set $S \subseteq X$ as the measure of its indicator function $\mathbb{1}_S$, but in general the function $\mathbb{1}_S$ will not lie in $C(X)$; this occurs if and only if S is both open and closed (which we note applies to $S = X$). Instead, for each open set $S \subseteq X$ we define

$$\mu(S) = \sup\{\mu(f) : 0 \le f \le \mathbb{1}_S,\ f \in C(X)\} \in [0,1],$$

and for each closed set $S \subseteq X$ we define

$$\mu(S) = 1 - \mu(X - S) \in [0,1].$$

If $S \subseteq X$ has the property that for every $\epsilon > 0$ there exists an open set $U \supseteq S$ of measure $\mu(U) \le \epsilon$, then we define $\mu(S) = 0$ and say that S has *measure zero*. If the boundary $\partial S := \overline{S} - S^0$ of a set S has measure zero, then we necessarily have $\mu(S^0) = \mu(\overline{S})$ and define $\mu(S)$ to be this common value; such sets are said to be *μ-quarrable*.

For the purpose of studying equidistribution, we shall restrict our attention to μ-quarrable sets S. This typically does not include all measurable sets in the usual sense, by which we mean elements of the Borel σ-algebra Σ of X generated by the open sets under complements and countable unions and intersections; see Exercise 2.1. However, if we are given a regular Borel measure μ on X of total mass 1, by which we mean a countably additive function $\mu\colon \Sigma \to \mathbf{R}_{\ge 0}$ for which $\mu(S) = \inf\{\mu(U) : S \subseteq U,\ U \text{ open}\}$ and $\mu(X) = 1$, it is easy to check that defining $\mu(f) := \int_X f\,\mu$ for each $f \in C(X)$ yields a measure under Definition 2.1; see [40, §1]).[11] This measure is completely determined by the values μ takes on μ-quarrable sets; see [95]. In particular, the Haar measure of a compact group uniquely determines a measure in the sense of Definition 2.1.

[9] In fact it is a commutative C^*-algebra with complex conjugation as its involution, but we will not make use of this.

[10] Note that this is a definition; taking a measure-theoretic approach avoids the need to develop an integration theory.

[11] Here we are using an integral to define a measure, rather than the other way around.

Definition 2.3. A sequence $(x_1, x_2, x_3, \ldots)$ in X is said to be *equidistributed with respect to* μ, or simply *μ-equidistributed*, if for every $f \in C(X)$ we have

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

Remark 2.4. When we speak of equidistribution, note that we are talking about a *sequence* (x_i) of elements of X in a particular order; it does not make sense to say that a *set* is equidistributed. For example, suppose we took the set of odd primes and arranged them in the sequence $(5, 13, 3, 17, 29, 7, \ldots)$ where we list two primes congruent to 1 modulo 4 followed by one prime congruent to 3 modulo 4. The sequence obtained by reducing this sequence modulo 4 is not equidistributed with respect to the uniform measure on $(\mathbf{Z}/4\mathbf{Z})^\times$, even though the sequence of odd primes in their usual order is (by Dirichlet's theorem on primes in arithmetic progressions). However, local rearrangements that change the index of an element by nor more than a bounded amount do not change its equidistribution properties. This applies to sequences indexed by primes in a number field that are ordered by norm with ties broken arbitrarily (but that this does not hold for function fields).

If (x_i) is a sequence in X, for each $f \in C(X)$ we define the kth-*moment* of the sequence $(f(x_i))$ by

$$\mathrm{M}_k[(f(x_i)] := \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f^k(x_i).$$

If these limits exist for all $k \geq 0$, then we define the *moment sequence*

$$\mathrm{M}[f(x_i)] := (\mathrm{M}_0[(f(x_i)], \mathrm{M}_1[(f(x_i)], \mathrm{M}_2[(f(x_i)], \ldots).$$

If (x_i) is μ-equidistributed, then we have $\mathrm{M}_k[f(x_i)] = \mu(f^k)$ and the moment sequence

$$\mathrm{M}[f(x_i)] = (\mu(f^0), \mu(f^1), \mu(f^2), \ldots) \tag{7}$$

is independent of the sequence (x_i); it depends only on the function f and the measure μ.

Remark 2.5. There is a partial converse that is relevant to some of our applications. To simplify matters, let us restrict our attention to real-valued functions; so for the purposes of this remark, let $C(X)$ denote the Banach algebra of real-valued functions on X and replace $\mathbf{C}$ with $\mathbf{R}$ in Definition 2.1. Let (x_i) be a sequence in X and let $f \in C(X)$. Then $f(X)$ is a compact subset of $\mathbf{R}$, and we may view $(f(x_i))$ as a sequence in $f(X)$. If the moments $\mathrm{M}_k[f(x_i)]$ exist for all $k \geq 0$, then there is a *unique* measure on $f(X)$ with respect to which the sequence $(f(x_i))$ is equidistributed; this follows from the Stone-Weierstrass theorem. If μ is a measure on $C(X)$, we define the pushforward measure $\mu_f(g) := \mu(g \circ f)$ on $C(f(X)$, and we see that the sequence $(f(x_i))$ is μ_f-equidistributed if and only if (7) holds. This gives a necessary (but in general not sufficient condition) for (x_i) to be μ-equidistributed that can be checked by comparing moment sequences. If we have a collection of functions $f_j \in C(X)$ such that the pushforward measures μ_{f_j} uniquely determine μ, we obtain a necessary and sufficient condition involving the moment sequences of the f_j with respect to μ. One can generalize this remark to the complex-valued case using the theory of C^*-algebras.

More generally, we have the following lemma.

Lemma 2.6. *Let (f_j) be a family of functions whose linear combinations are dense in $C(X)$. If (x_i) is a sequence in X for which the limit $\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f_j(x_i)$ converges for every f_j, then there is a unique measure μ on X for which (x_i) is μ-equidistributed.*

Proof. See [68, Lemma A.1, p. I-19]. □

Proposition 2.7. *If (x_i) is a μ-equidistributed sequence in X and S is a μ-quarrable set in X then*

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

Proof. See Exercise 2.2. □

Example 2.8. If $X = [0,1]$ and μ is the Lebesgue measure then a sequence (x_i) is μ-equidistributed if and only if for every $0 \le a < b \le 1$ we have

$$\lim_{n\to\infty} \frac{\#\{x_i \in [a,b] : i \le n\}}{n} = b - a.$$

More generally, if X is a compact subset of $\mathbf{R}^n$ and μ is the normalized Lebesgue measure, then (x_i) is μ-equidistributed if and only if for every μ-quarrable set $S \subseteq X$ we have $\lim_{n\to\infty} \frac{1}{n}\#\{x_i \in S : i \le n\} = \mu(S)$.

2.2. **Equidistribution in compact groups.** We now specialize to the case where $X := \mathrm{conj}(G)$ is the space of conjugacy classes of a compact group G, obtained by taking the quotient of G as a topological space under the equivalence relation defined by conjugacy; let $\pi : G \to X$ denote the quotient map. We then equip X with the pushforward of the Haar measure μ on G (normalized so that $\mu(G) = 1$), which we also denote μ. Explicitly, π induces a contravariant map of Banach spaces

$$\begin{aligned} C(X) &\to C(G) \\ f &\mapsto f \circ \pi, \end{aligned}$$

and the value of μ on $C(X)$ is defined by

$$\mu(f) := \mu(f \circ \pi).$$

We say that a sequence (x_i) in X or a sequence (g_i) in G is *equidistributed* if it is μ-equidistributed (when we speak of equidistribution in a compact group without explicitly mentioning a measure, we always mean the Haar measure).

Proposition 2.9. *Let G be a compact group with Haar measure μ, and let $X := \mathrm{conj}(G)$. A sequence (x_i) in X is μ-equidistributed if and only if for every irreducible character χ of G we have*

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

Proof. As explained in [68, Prop. A.2], this follows from Lemma 2.6 and the Peter-Weyl theorem, since the irreducible characters χ of G generate a dense subset of $C(X)$. □

Corollary 2.10. *Let G be a compact group with Haar measure μ, and let $X := \mathrm{conj}(G)$. A sequence (x_i) in X is μ-equidistributed if and only if for every nontrivial irreducible character χ of G we have*

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

Proof. For the trivial character we have $\mu(1) = \mu(G) = 1$, and for any nontrivial irreducible character χ we must have $\mu(\chi) = \int_G \chi\mu = \int_G 1 \cdot \chi\mu = 0$, by Schur orthogonality; the corollary follows. □

To illustrate these results, let us apply Corollary 2.10 to prove an equidistribution result for elliptic curves over finite fields that will be useful later. We first recall some basic facts. Let E be an elliptic curve over $\mathbf{F}_q$. The *Frobenius endomorphism* π_E is defined by

$$(x : y : z) \mapsto (x^q : y^q : z^q).$$

Like all endomorphisms of elliptic curves, π_E has a characteristic polynomial of the form

$$T^2 - (\operatorname{tr} \pi_E)T + \deg \pi_E$$

that is satisfied by both π_E and its dual $\hat{\pi}_E$, where $\operatorname{tr} \pi_E = \pi_E + \hat{\pi}_E$ and $q = \deg \pi_E = \pi_E \hat{\pi}_E$ are both integers.[12] The set $E(\mathbf{F}_q)$ is, by definition, the subset of $E(\overline{\mathbf{F}}_q)$ fixed by π_E, equivalently, the kernel of the endomorphism $\pi_E - 1$. One can show that $\pi_E - 1$ is a separable, and therefore

$$\#E(\mathbf{F}_q) = \#\ker(\pi_E - 1) = \deg(\pi_E - 1) = (\pi_E - 1)(\hat{\pi}_E - 1) = \hat{\pi}_E \pi_E + 1 - (\hat{\pi}_E + \pi_E) = q + 1 - \operatorname{tr} \pi_E.$$

It follows that $t_q := q + 1 - \#E(\mathbf{F}_q)$ is the *trace of Frobenius* $\operatorname{tr} \pi_E$. As we showed in the previous lecture in the case $q = p$, the zeta function of E can be written as

$$Z_E(T) = \frac{qT^2 - t_q T + 1}{(1 - T)(1 - qT)},$$

where the complex roots of $qT^2 - t_q T + 1$ have absolute value $q^{-1/2}$. This implies that we can write $t_q = \alpha + \bar{\alpha}$ for some $\alpha \in \mathbf{C}$ with $|\alpha| = q^{1/2}$, and we have $\#E(\mathbf{F}_q) = q + 1 - (\alpha + \bar{\alpha})$.

We now observe that for any integer $r \geq 1$, the set $E(\mathbf{F}_{q^r})$ is the subset of $E(\overline{\mathbf{F}}_q)$ fixed by π_E^r, which corresponds to the q^r-power Frobenius automorphism; it follows that

$$\#E(\mathbf{F}_{q^r}) = q^r + 1 - (\alpha^r + \bar{\alpha}^r),$$

and therefore $\alpha^r + \bar{\alpha}^r$ is the trace t_{q^r} of the Frobenius endomorphism of the base change of E to $\mathbf{F}_{q^r}$.

As an application of Corollary 2.10, we now prove the following result, taken from [23, Prop 2.2]. Recall that $E/\mathbf{F}_q$ is said to be *ordinary* if t_q is not zero modulo the characteristic of $\mathbf{F}_q$.

Proposition 2.11. *Let $E/\mathbf{F}_q$ be an ordinary elliptic curve and for integers $r \geq 1$, let $t_{q^r} := q^r + 1 - \#E(\mathbf{F}_{q^r})$ and define*

$$x_r := t_{q^r}/q^{r/2}.$$

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 α be as above, with $|\alpha| = q^{1/2}$ and $\operatorname{tr} \pi_E = \alpha + \bar{\alpha}$. Then $x_r = (\alpha^r + \bar{\alpha}^r)/q^{r/2}$ for all $r \geq 1$. Let $\mathrm{U}(1) := \{u \in \mathbf{C}^\times : u\bar{u} = 1\}$ be the unitary group. For $u = e^{i\theta}$, the Haar measure on $\mathrm{U}(1)$ corresponds to the uniform measure on $\theta \in [-\pi, \pi]$, this follows immediately the translation invariance of the Haar measure. Let us compute the pushforward of the Haar measure of $\mathrm{U}(1)$ to $[-2, 2]$ via the map $u \mapsto z := u + \bar{u} = 2\cos\theta$. We have $dz = 2\sin\theta\, d\theta$, and see that the pushforward is precisely μ.

[12] By the *dual* of an endomorphism of a principally polarized abelian variety we mean the Rosati dual, which for elliptic curves we may identify with the dual isogeny.

The nontrivial irreducible characters $\mathrm{U}(1) \to \mathbf{C}^\times$ are of the form $\phi_a(u) = u^a$ for some nonzero $a \in \mathbf{Z}$. For each such ϕ_a we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{n} \phi_a(\alpha^r/q^{r/2}) = \lim_{n\to\infty} \frac{1}{n} \sum_{r=1}^{n} (\alpha/q^{1/2})^{ra} = \lim_{n\to\infty} \frac{1}{n} \frac{(\alpha/q^{1/2})^{a(n+1)} - (\alpha/q^{1/2})^a}{(\alpha/q^{1/2})^a - 1} = 0.$$

The hypothesis that E is ordinary guarantees that $\alpha/q^{1/2}$ is not a root of unity (see Exercise 2.3), thus $(\alpha/q^{1/2})^a - 1$ is nonzero for all nonzero $a \in \mathbf{Z}$. Corollary 2.10 implies that $(\alpha^r/q^{r/2})$ is equidistributed in $\mathrm{U}(1)$, and therefore (x_r) is μ-equidistributed. □

See [2] for a generalization to smooth projective curves $C/\mathbf{F}_q$ of arbitrary genus $g \geq 1$.

2.3. **Equidistribution for L-functions.** As above, let G be a compact group and let $X := \mathrm{conj}(G)$. Let K be a number field, and let $P := (\mathfrak{p}_1, \mathfrak{p}_2, \mathfrak{p}_3, \ldots)$ be a sequence consisting of all but finitely many primes $\mathfrak{p}$ of K ordered by norm; this means that $N(\mathfrak{p}_i) \leq N(\mathfrak{p}_j)$ for all $i \leq j$. Let $(x_\mathfrak{p})$ be a sequence in X indexed by P, and for each irreducible representation $\rho : G \to \mathrm{GL}_d(\mathbf{C})$, define the L-function

$$L(\rho, s) := \prod_{\mathfrak{p}\in P} \det(1 - \rho(x_\mathfrak{p})N(\mathfrak{p})^{-s})^{-1},$$

for $s \in \mathbf{C}$ with $\mathrm{Re}(s) > 1$.

Theorem 2.12. *Let G and $(x_\mathfrak{p})$ be as above, and suppose $L(\rho, s)$ is meromorphic on $\mathrm{Re}(s) \geq 1$ with no zeros or poles except possibly at $s = 1$, for every irreducible representation ρ of G. The sequence $(x_\mathfrak{p})$ is equidistributed if and only if for each $\rho \neq 1$, the L-function $L(\rho, s)$ extends analytically to a function that is holomorphic and nonvanishing on $\mathrm{Re}(s) \geq 1$.*

Proof. See the corollary to [68, Thm. A.2], or see [23, Thm. 2.3]. □

A notable case in which the hypothesis of Theorem 2.12 is known to hold is when $L(\rho, s)$ corresponds to an Artin L-function. As in Lecture 1, to each prime $\mathfrak{p}$ in K we associate an absolute Frobenius element $\mathrm{Frob}_\mathfrak{p} \in \mathrm{Gal}(\overline{K}/K)$, and for each finite Galois extension L/K we use $\mathrm{conj}_L(\mathrm{Frob}_\mathfrak{p})$ to denote the conjugacy class in $\mathrm{Gal}(L/K)$ of the restriction of $\mathrm{Frob}_\mathfrak{p}$ to L.

Corollary 2.13. *Let L/K be a finite Galois extension with $G := \mathrm{Gal}(L/K)$ and let P be the sequence of unramified primes of K ordered by norm (break ties arbitrarily). The sequence $(\mathrm{conj}_L(\mathrm{Frob}_\mathfrak{p}))_{\mathfrak{p}\in P}$ is equidistributed in $\mathrm{conj}(G)$; in particular, the Chebotarev density theorem (Theorem 1.1) holds.*

Proof. For the trivial representation, the L-function $L(1, s)$ agrees with the Dedekind zeta function $\zeta_K(s)$ up to a finite number of holomorphic nonvanishing factors, and, as originally proved by Hecke, $\zeta_K(s)$ is holomorphic and nonvanishing on $\mathrm{Re}(s) \geq 1$ except for a simple pole at $s = 1$; see [60, Cor. VII.5.11], for example. For every nontrivial irreducible representation $\rho : G \to \mathrm{GL}_d(\mathbf{C})$, the L-function $L(\rho, s)$ agrees with the corresponding Artin L-function for ρ, up to a finite number of holomorphic nonvanishing factors, and, as originally proved by Artin, $L(\rho, s)$ is holomorphic and nonvanishing on $\mathrm{Re}(s) \geq 1$; see [13, p.225], for example. The corollary then follows from Theorem 2.12. □

2.4. **Sato–Tate for CM elliptic curves.** As a second application of Theorem 2.12, let us prove an equidistribution result for CM elliptic curves. To do so we need to introduce Hecke characters, which may be viewed as (quasi-)characters of the idéle class group of a number field.

Definition 2.14. Let K be a number field and let $\mathbf{I}_K$ denote its idéle group. A *Hecke character* is a continuous homomorphism

$$\psi : \mathbf{I}_K \to \mathbf{C}^\times$$

whose kernel contains $K^\times$. The *conductor* of ψ is the $\mathbf{Z}_K$-ideal $\mathfrak{f} := \prod_\mathfrak{p} \mathfrak{p}^{e_\mathfrak{p}}$ in which each $e_\mathfrak{p}$ is the minimal nonnegative integer for which $1 + \hat{\mathfrak{p}}^{e_\mathfrak{p}} \subseteq \mathbf{Z}_{K_\mathfrak{p}}^\times \hookrightarrow \mathbf{I}_K$ lies in the kernel of ψ (all but finitely many $e_\mathfrak{p}$ are necessarily zero because ψ is continuous).

Each Hecke character ψ has an associated *Hecke L-function*

$$L(\psi, s) := \prod_{\mathfrak{p} \nmid \mathfrak{f}} (1 - \psi(\mathfrak{p}) N(\mathfrak{p})^{-s})^{-1},$$

where $\psi(\mathfrak{p}) := \psi(\pi_{\hat{\mathfrak{p}}})$ for any uniformizer $\pi_{\hat{\mathfrak{p}}}$ of $\hat{\mathfrak{p}}$ (we have omitted the gamma factors at archimedean places). We now want to consider the sequence of unitarized values

$$x_\mathfrak{p} := \frac{\psi(\mathfrak{p})}{|\psi(\mathfrak{p})|} \in \mathrm{U}(1)$$

indexed by primes $\mathfrak{p} \nmid \mathfrak{f}$ ordered by norm.

Lemma 2.15. *The sequence $(x_\mathfrak{p})$ is equidistributed in* $\mathrm{U}(1)$.

Proof. As in the proof of Proposition 2.11, the nontrivial irreducible characters of $\mathrm{U}(1)$ are those of the form $\phi_a(z) = z^a$ with $a \in \mathbf{Z}$ nonzero, and in each case the corresponding L-function is a Hecke L-function (if ψ is a Hecke character, so is ψ^a and its unitarized version). If ψ is trivial then, as in the proof of Corollary 2.13, $L(1, s)$ is holomorphic and nonvanishing on $\mathrm{Re}(s) \geq 1$ except for a simple pole at $s = 1$ because the same is true of $\zeta_K(s)$. Hecke proved [39] that when ψ is nontrivial $L(\psi, s)$ is holomorphic and nonvanishing on $\mathrm{Re}(s) \geq 1$, and the lemma then follows from Theorem 2.12. □

As an application of Corollary 2.15, we can now prove the Sato-Tate conjecture for CM elliptic curves. Les us first consider the case where K is an imaginary quadratic field and E/K is an elliptic curve with CM by K (so $K \simeq \mathrm{End}(E) \otimes_\mathbf{Z} \mathbf{Q}$). As explained below, the general case (including $K = \mathbf{Q}$) follows easily.

Let $\mathfrak{f}$ be the *conductor* of E; this is a $\mathbf{Z}_K$-ideal divisible only by the primes of bad reduction for E; see [77, §IV.10]. A classical result of Deuring [77, Thm. II.10.5] implies the existence of a Hecke character ψ_E of K of conductor $\mathfrak{f}$ such that for each prime $\mathfrak{p} \nmid \mathfrak{f}$ we have $|\psi_E(\mathfrak{p})| = N(\mathfrak{p})^{1/2}$ and

$$\psi_E(\mathfrak{p}) + \overline{\psi_E(\mathfrak{p})} = t_\mathfrak{p},$$

where $t_\mathfrak{p} := \mathrm{tr}\, \pi_E = N(\mathfrak{p}) + 1 - \#E_\mathfrak{p}(\mathbf{F}_\mathfrak{p}) \in \mathbf{Z}$ is the trace of Frobenius of the reduction of E modulo $\mathfrak{p}$.

Proposition 2.16. *Let K be an imaginary quadratic field and let E/K be an elliptic curve of conductor $\mathfrak{f}$ with CM by K. Let P be the sequence of primes of K that do not divide $\mathfrak{f}$ ordered by norm (break ties arbitrarily), and for $\mathfrak{p} \in P$ let $x_\mathfrak{p} := t_\mathfrak{p}/N(\mathfrak{p})^{1/2} \in [-2, 2]$ be the normalized Frobenius trace of $E_\mathfrak{p}$. The sequence $(x_\mathfrak{p})$ is equidistributed on $[-2, 2]$ with respect to the measure*

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

Proof. By the previous lemma, the sequence $(\psi_E(\mathfrak{p})/N(\mathfrak{p})^{1/2})_{\mathfrak{p} \in P}$ is equidistributed in $\mathrm{U}(1)$. As shown in the proof of Proposition 2.11, the measure μ_{cm} is the pushforward of the Haar measure on $\mathrm{U}(1)$ to

$[-2,2]$ under the map $u \mapsto u + \bar{u}$. For each $\mathfrak{p} \in P$ the image of $\psi_E(\mathfrak{p})/N(\mathfrak{p})^{1/2}$ under this map is

$$\frac{\psi_E(\mathfrak{p})}{N(\mathfrak{p})^{1/2}} + \frac{\overline{\psi_E(\mathfrak{p})}}{N(\mathfrak{p})^{1/2}} = \frac{t_p}{N(\mathfrak{p})^{1/2}} = x_{\mathfrak{p}}.$$

□

Figure 2 shows a trace histogram for the CM elliptic curve $y^2 = x^3 + 1$ over its CM field $\mathbf{Q}(\sqrt{-3})$.

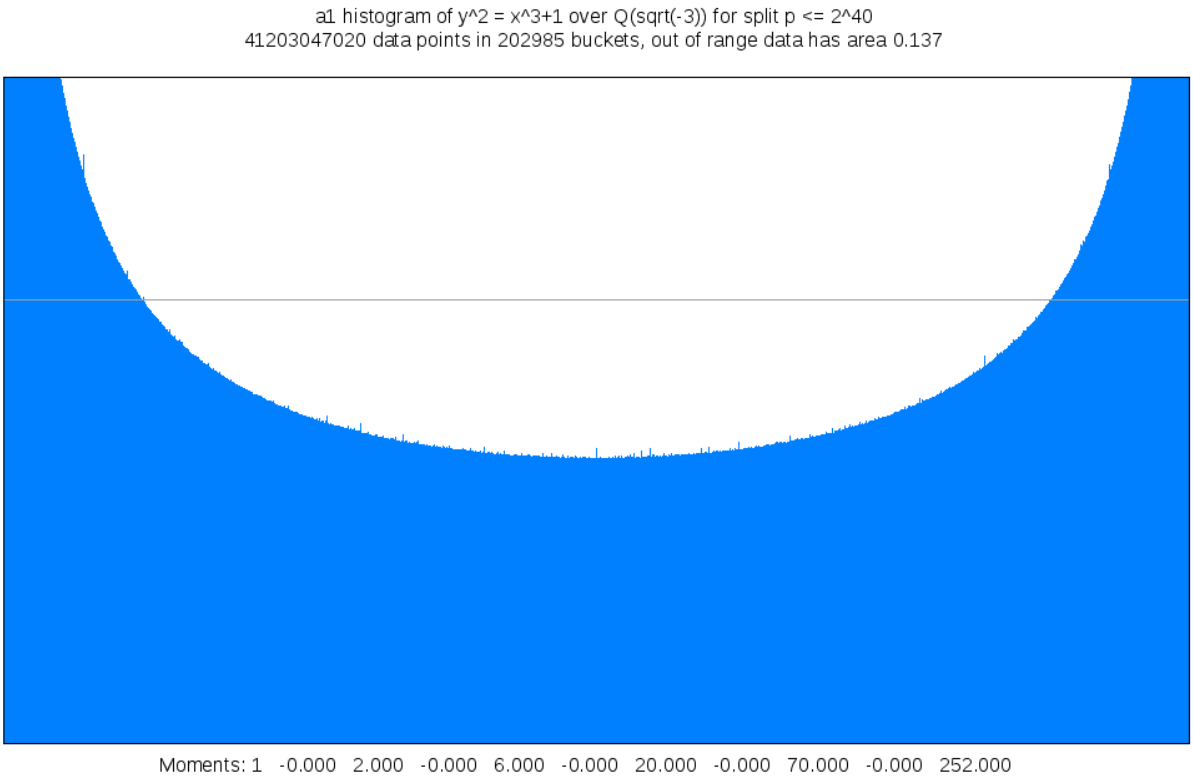


FIGURE 2. Click image to animate (requires Adobe Reader), or visit this web page.

Let us now consider the case of an elliptic curve $E/\mathbf{Q}$ with CM by F. For primes p of good reduction that are inert in F, the endomorphism algebra $\mathrm{End}(E_p)_{\mathbf{Q}} := \mathrm{End}(E_p) \otimes_{\mathbf{Z}} \mathbf{Q}$ of the reduced curve E_p contains two distinct imaginary quadratic fields, one corresponding to the CM field $F \simeq \mathrm{End}(E)_{\mathbf{Q}}$ and the other generated by the Frobenius endomorphism (the two cannot coincide because p is inert in F but the Frobenius endomorphism has norm p in $\mathrm{End}(E_p)_{\mathbf{Q}}$). It follows that $\mathrm{End}(E_p)_{\mathbf{Q}}$ must be a quaternion algebra, E_p is supersingular, and for $p > 3$ we must have $t_p = 0$, since $t_p \equiv 0 \bmod p$ and $|t_p| \le 2\sqrt{p}$; see [78, III,9,V.3] for these and other facts about the endomorphism ring of an elliptic curve.

At split primes $p = \mathfrak{p}\bar{\mathfrak{p}}$ the reduced curve E_p will be isomorphic to the reduction modulo $\mathfrak{p}$ of its base change to F (both of which will be elliptic curves over $\mathbf{F}_p = \mathbf{F}_{\mathfrak{p}}$), and will have the same trace of Frobenius $t_p = t_{\mathfrak{p}}$. By the Chebotarev density theorem, the split and inert primes both have density $1/2$, and it follows that the sequence of normalized Frobenius traces $x_{\mathfrak{p}} := t_p/\sqrt{p} \in := [-2,2]$ is equidistributed with respect to the measure $\frac{1}{2}\delta_0 + \frac{1}{2}\mu_{\mathrm{cm}}$, where we use the Dirac measure δ_0 to put half the mass at 0 to account for the inert primes. This can be seen in Figure 3, which shows a trace

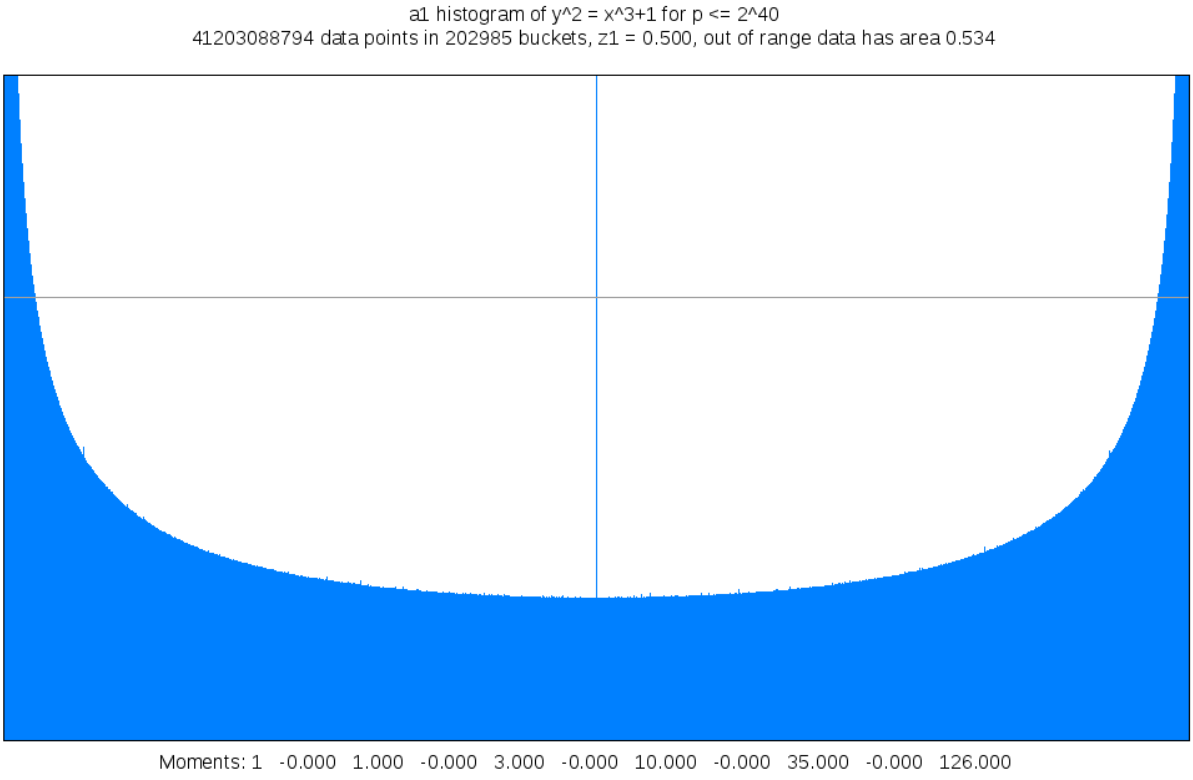


FIGURE 3. Click image to animate (requires Adobe Reader), or visit this web page.

histogram for the CM elliptic curve $y^2 = x^3 + 1$ over $\mathbf{Q}$; the thin spike in the middle of the histogram at zero has area $1/2$ (one can also see that the nontrivial moments are half what they were in Figure 2).

A similar argument applies when E is defined over a number field K that does not contain the CM field F. For the sake of proving an equidistribution result we can restrict our attention to the degree-1 primes $\mathfrak{p}$ of K, those for which $N(\mathfrak{p}) = p$ is prime. Half of these primes $\mathfrak{p}$ will split in the compositum KF, and the subsequence of normalized traces $x_\mathfrak{p}$ at these primes will be equidistributed with respect to the measure μ_{cm}, and half will be inert in KF, in which case $x_\mathfrak{p} = t_\mathfrak{p} = 0$.

2.5. **Sato–Tate for non-CM elliptic curves.** We can now state the Sato-Tate conjecture in the form originally given by Tate, following [68, §1A]. Tate's seminal paper [84] describes what is now known as the *Tate conjecture*, which comes in two conjecturally equivalent forms **T1** and **T2**, the latter of which is stated in terms of L-functions. The Sato-Tate conjecture is obtained by applying **T2** to all powers of a fixed elliptic curve $E/\mathbf{Q}$ (as products of abelian varieties); see [64] for an introduction to the Tate conjecture and an explanation of how the Sato-Tate conjecture fits within it.

Let G be the compact group $\mathrm{SU}(2)$ of 2×2 unitary matrices with determinant 1. The irreducible representations of G are the mth symmetric powers ρ_m of the natural representation ρ_1 of degree 2 given by the inclusion $\mathrm{SU}(2) \subseteq \mathrm{GL}_2(\mathbf{C})$. Each element of $X := \mathrm{conj}(G)$ can be uniquely represented by a matrix of the form

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

where $\theta \in [0,\pi]$ is the eigenangle of the conjugacy class. It follows that each $f \in C(X)$ can be viewed as a continuous function $f(\theta)$ on the compact set $[0,\pi]$. The pushforward of the Haar measure is

$$\mu = \frac{2}{\pi}\sin^2\theta\, d\theta, \tag{8}$$

see Exercise 2.4, which means that for each $f \in C(X)$ we have

$$\mu(f) = \frac{2}{\pi}\int_0^\pi f(\theta)\sin^2\theta\, d\theta.$$

Let $E/\mathbf{Q}$ be an elliptic curve without CM, let $P := (p)$ be the sequence of primes that do not divide the conductor N of E, in order, and for each $p \in P$ let $x_p \in X$ to be the element of X corresponding to the unique $\theta_p \in [0,\pi]$ for which $2\cos\theta_p\sqrt{p} = t_p := p + 1 - \#E_p(\mathbf{F}_p)$ is the trace of Frobenius of the reduced curve E_p.

We are now in the setting of §2.3. We have a compact group $G := \mathrm{SU}(2)$, its space of conjugacy classes $X := \mathrm{conj}(G)$, a number field $K = \mathbf{Q}$, a sequence P containing all but finitely many primes of K ordered by norm, a sequence (x_p) in X indexed by P, and an irreducible representation $\rho_m\colon G \to \mathrm{GL}_{m+1}(\mathbf{C})$, for each $m \geq 1$. The L-function corresponding to ρ_m is given by

$$L(\rho_m, s) := \prod_{p\nmid N}\det(1-\rho_m(x_p)p^{-s})^{-1} = \prod_{p\nmid N}\prod_{k=0}^{m}(1-e^{i(m-2k)\theta_p}p^{-s})^{-1}.$$

For each $p \nmid N$, let α_p and $\bar{\alpha}_p$ be the roots of $T^2 - t_pT + p$, so that $\alpha_p = e^{i\theta_p}p^{1/2}$. If we now define

$$L_m^1(s) := \prod_{p\nmid N}\prod_{r=0}^{m}(1-\alpha_p^{m-r}\bar{\alpha}_p^r p^{-s})^{-1},$$

then for $m \geq 1$ we have

$$L(\rho_m, s) = L_m^1(s - m/2).$$

Tate conjectured in [84] that $L_m^1(s)$ is holomorphic and nonvanishing on $\mathrm{Re}(s) \geq 1+m/2$, which implies that each $L(\rho_m, s)$ is holomorphic and nonvanishing on $\mathrm{Re}(s) \geq 1$. Assuming this is true, Theorem 2.12 implies that the sequence (x_p) is μ-equidistributed, which is equivalent to the Sato-Tate conjecture.

We now recall the *modularity theorem* for elliptic curves over $\mathbf{Q}$, which states that there is a one-to-one correspondence between isogeny classes of elliptic curves $E/\mathbf{Q}$ of conductor N and modular forms

$$f(z) = \sum_{n\geq 1} a_n e^{2\pi i n z} \in S_2(\Gamma_0(N))^{\mathrm{new}} \qquad (a_1 = 1)$$

that are eigenforms for the action of the Hecke algebra on the space $S_2(\Gamma_0(N))$ of cuspforms of weight 2 and level N. We require f to be *new* at level N, meaning that it does not lie in $S_2(\Gamma_0(M))$ for any M properly dividing N. Such modular forms f are called (normalized) *newforms*, of weight 2 and level N. The modularity theorem was proved is the case that N is squarefree by Taylor and Wiles [87, 94], and extended to all elliptic curves over $\mathbf{Q}$ by Breuil, Conrad, Diamond, and Taylor [11].

The modular form f is a simultaneous eigenform for all the Hecke operators T_n, and the normalization $a_1 = 1$ ensures that for each prime $p \nmid N$, the coefficient a_p is the eigenvalue of f for T_p, and moreover, an integer. Under the correspondence given by the modularity theorem, the eigenvalue a_p is equal to the trace of Frobenius t_p of the reduced curve E_p, where E is any representative of the corresponding isogeny class. Here we are using the fact that if E and E' are isogenous elliptic curves over $\mathbf{Q}$, then they have the same conductor N and the same trace of Frobenius t_p at ever $p \nmid N$. This

is actually a part of the Tate conjecture from [84] that was proved by Faltings in [22]. More generally, Faltings proved that isogenous abelian varieties over a number field have the same L-function.

There is an L-function $L(f,s)$ associated to the modular form f, and the modularity theorem guarantees that it coincides with the L-function $L(E,s)$ of E. So not only does $a_p = t_p$ for all $p \nmid N$, the Euler factors at the bad primes $p|N$ also agree. We need not concern ourselves with Euler factors at these primes, other than to note that they are holomorphic and nonvanishing on $\mathrm{Re}(s) \geq 3/2$. After removing the Euler factors at bad primes, the L-functions $L(E,s)$ and $L(f,s)$ both have the form

$$\prod_{p\nmid N}(1-a_p p^{-s}+p^{1-2s})^{-1} = \prod_{p\nmid N}\prod_{r=0}^{1}(1-\alpha_p^{1-r}\bar{\alpha}_p^r p^{-s})^{-1} = L_1^1(s),$$

where α_p and $\bar{\alpha}_p$ are the roots of $T^2 - a_p T + p = T^2 - t_p T + p$.

The L-function $L(f,s)$ is holomorphic and nonvanishing on $\mathrm{Re}(s) \geq 3/2$; see [20, Prop. 5.9.1]. The modularity theorem tells us that the same is true of $L(E,s)$, and therefore of $L_1^1(s)$. Thus the modularity theorem proves that Tate's conjecture regarding $L_m^1(s)$ holds when $m = 1$. To prove the Sato-Tate conjecture we need this for all $m \geq 1$.

Theorem 2.17. *Let $f(z) := \sum_{n\geq 1} a_n e^{2\pi i z n} \in S_2(\Gamma_0(N)^{\mathrm{new}})$ be a normalized newform without CM. For each prime $p \nmid N$ let $\alpha_p, \bar{\alpha}_p$ be the roots of $T^2 - a_p T + p$. Then*

$$\prod_{p\nmid N}\prod_{r=0}^{m}(1-\alpha_p^{m-r}\bar{\alpha}_p^r p^{-s})^{-1} = L_m^1(s)$$

is holomorphic and nonvanishing on $\mathrm{Re}(s) \geq 1 + m/2$.

Proof. Apply [6, Theorem B.2] with weight $k = 2$, trivial nebentypus $\psi = 1$, and trivial character $\chi = 1$ (as noted in [6], this special case was already addressed in [31]). □

Corollary 2.18. *The Sato-Tate conjecture (Theorem 1.8) holds.*

Remark 2.19. The Sato-Tate conjecture is also known to hold for elliptic curves over totally real fields; this was proved for elliptic curves with potentially multiplicative reduction at some prime in [31, 86], and this technical assumption can be removed (see the discussion in the introduction to [5]). The Sato-Tate conjecture remains open for elliptic curves over number fields that are not totally real.

2.6. **Exercises.**

Exercise 2.1. Let X be a compact Hausdorff space. Show that the only sets $S \subseteq X$ that are μ-quarrable for every measure μ on X are the sets that are both open and closed.

Exercise 2.2. Prove Proposition 2.7.

Exercise 2.3. Let E an elliptic curve over $\mathbf{F}_q$ and let α be a root of the characteristic polynomial of the Frobenius endomorphism π_E. Prove that $\alpha/\sqrt{q}$ is a root of unity if and only if E is supersingular.

Exercise 2.4. Show that the set of conjugacy classes of SU(2) is in bijection with the set of eigenangles $\theta \in [0,\pi]$. Then prove that the pushforward of the Haar measure of SU(2) onto $[0,\pi]$ is given by $\mu := \frac{2}{\pi}\sin^2\theta\, d\theta$ (hint: show that SU(2) is isomorphic to the 3-sphere S^3 and use this isomorphism together with the translation invariance of the Haar measure to determine μ)

Exercise 2.5. Compute the trace moment sequence for SU(2) (that is, prove (6)). Embed U(1) in SU(2) via the map $u \mapsto \left(\begin{smallmatrix} u & 0 \\ 0 & \bar{u} \end{smallmatrix}\right)$ and compute its trace moment sequence (compare to Figure 2). Now determine the normalizer $N(\mathrm{U}(1))$ of U(1) in SU(2) and compute its trace moment sequence (compare to Figure 3).

3. Sato-Tate groups

In the previous lecture we showed that there are three distinct Sato-Tate distributions that arise for elliptic curves E over number fields K (only two of which occur when $K = \mathbf{Q}$). All three distributions can be associated to the Haar measure of a compact subgroup $G \subseteq \mathrm{SU}(2)$, in which we embed $\mathrm{U}(1)$ via the map $u \mapsto \left(\begin{smallmatrix} u & 0 \\ 0 & \bar{u} \end{smallmatrix}\right)$. We are interested in the pushforward μ of the Haar measure onto $\mathrm{conj}(G)$, which can be expressed in terms of the eigenangle $\theta \in [0, \pi]$. The three possibilities for G are listed below.

- $\mathrm{U}(1)$: we have $\mu(\theta) = \frac{1}{\pi} d\theta$ and trace moments: $(1, 0, 2, 0, 6, 0, 20, 0, 70, 0, 252, \ldots)$.
  Arises for CM elliptic curves defined over a field that contains the CM field.
- $N(\mathrm{U}(1))$: we have $\mu(\theta) = \frac{1}{2\pi} d\theta + \frac{1}{2}\delta_{\pi/2}$ and trace moments: $(1, 0, 1, 0, 3, 0, 10, 0, 35, 0, 126, \ldots)$.
  Arises for CM elliptic curves defined over a field that does not contain the CM field.
- $\mathrm{SU}(2)$: we have $\mu(\theta) = \frac{2}{\pi} \sin^2\theta \, d\theta$ and trace moments: $(1, 0, 1, 0, 2, 0, 5, 0, 14, 0, 42, \ldots)$.
  Arises for non-CM elliptic curves (conjecturally so when the ground field is not totally real).

We have written μ in terms of θ, but we may view it as a linear function on the Banach space $C(X)$, where we identify $X := \mathrm{conj}(G)$ with $[0, \pi]$, by defining $\mu(f) = \int_0^\pi f(\theta)\mu(\theta)$, as in §2.1. In particular, μ assigns a value to the trace function $\mathrm{tr}\colon X \to [-2, 2]$, where $\mathrm{tr}(\theta) = 2\cos\theta$, and to its powers tr^n, which allows us to compute the trace moment sequence $(\mu(\mathrm{tr}^n))_{n \geq 0}$.

3.1. **The Sato-Tate group of an elliptic curve.** Thus far the link between the elliptic curve E and the compact group G whose Haar measure is claimed (and in many cases proven) to govern the distribution of Frobenius traces has been made via the measure μ. That is, we have an equidistribution claim for the sequence (x_p) of normalized Frobenius traces associated to E that is phrased in terms of μ. We would like to establish a direct relationship between E and G that defines G as an arithmetic invariant of E. We may then state the Sato-Tate conjecture/theorem as an equidistribution claim involving the Haar measure of G, but we would like the definition of G not to depend on whether we can prove this equidistribution claim or not (or whether it is even true, although of course we expect it to be).

In Lecture 1 we considered the Galois representation $\rho_f\colon \mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q}) \to \mathrm{GL}_d(\mathbf{C})$ defined by the action of $\mathrm{Gal}(\overline{\mathbf{Q}}/\mathbf{Q})$ on the roots of a squarefree polynomial $f \in \mathbf{Z}[x]$. We thereby obtained a compact group G_f and a map that sends each prime p of good reduction for f to an element of $\mathrm{conj}(G_f)$ (namely, the map $p \mapsto \rho_f(\mathrm{Frob}_p)$). We were then able to relate the image of p under this map to the quantity $N_f(p)$ of interest, via (1). This construction did not involve any discussion of equidistribution, but we could then prove, via the Chebotarev density theorem, that the conjugacy classes $\rho_f(p)$ are equidistributed with respect to the pushforward of the Haar measure to $\mathrm{conj}(G_f)$.

We take a similar approach here. To each elliptic curve E over a number field K we will associate a compact group G that is constructed via a Galois representation attached to E, equipped with a map that sends each prime $\mathfrak{p}$ of good reduction for E to an element $x_\mathfrak{p}$ of $\mathrm{conj}(G)$ that we can directly relate to the quantity $N_E(\mathfrak{p}) := p + 1 - t_\mathfrak{p}$ whose distribution we wish to study. We may then conjecture (and prove, when E has CM or K is totally real), that the sequence $(x_\mathfrak{p})$ is equidistributed in $X := \mathrm{conj}(G)$ (with respect to the pushforward of the Haar measure of G).

The group G is the *Sato–Tate* group of E, and will be denoted $\mathrm{ST}(E)$. It is a compact subgroup of $\mathrm{SU}(2)$, and our construction will make it easy to show that $\mathrm{ST}(E)$ is always one of the three groups $\mathrm{U}(1)$, $N(\mathrm{U}(1))$, $\mathrm{SU}(2)$ listed above, depending on whether E has CM or not, and if so, whether the CM field is contained in the ground field or not. None of this depends on any equidistribution results. This construction will be our prototype for the definition of the Sato-Tate group of an abelian variety of dimension g, so we will work out the $g = 1$ case in some detail.

In order to associate a Galois representation to E/K, we need a set on which $\mathrm{Gal}(\overline{K}/K)$ can act. For each integer $n \geq 1$, let $E[n] := E(\overline{K})[n]$ denote the n-torsion subgroup of $E(\overline{K})$, a free $\mathbf{Z}/n\mathbf{Z}$-module of rank 2 (see [78, Cor. III.6.4]). The group $\mathrm{Gal}(\overline{K}/K)$ acts on points in $E(\overline{K})$ coordinate-wise, and $E[n]$ is invariant under this action because it is the kernel of the multiplication-by-n map $[n]$, an endomorphism of E that is defined over K; one can concretely define $E[n]$ as the zero locus the n-division polynomials, which have coefficients in K. The action of $\mathrm{Gal}(\overline{K}/K)$ on $E[n]$ induces the *mod-n Galois representation*

$$\mathrm{Gal}(\overline{K}/K) \to \mathrm{Aut}(E[n]) \simeq \mathrm{GL}_2(\mathbf{Z}/n\mathbf{Z}).$$

This Galois representation is insufficient for our purposes, because the image $M_\mathfrak{p}$ of $\mathrm{Frob}_\mathfrak{p}$ in $\mathrm{GL}_2(\mathbf{Z}/n\mathbf{Z})$ does not determine $t_\mathfrak{p}$, we only have $t_\mathfrak{p} \equiv \operatorname{tr} M_\mathfrak{p} \bmod n$; we need to let $\mathrm{Gal}(\overline{K}/K)$ act on a bigger set.

So let us fix a prime ℓ (any prime will do), and consider the inverse system

$$\cdots \xrightarrow{[\ell]} E[\ell^3] \xrightarrow{[\ell]} E[\ell^2] \xrightarrow{[\ell]} E[\ell].$$

The inverse limit

$$T_\ell := \varprojlim_n E[\ell^n]$$

is the ℓ-adic *Tate-module* of E; it is a free $\mathbf{Z}_\ell$-module of rank 2. The group $\mathrm{Gal}(\overline{K}/K)$ acts on T_ℓ via its action on the groups $E[\ell^n]$, and this action is compatible with the multiplication-by-ℓ map $[\ell]$ because this map is defined over K (it can be written as a rational map with coefficients in K). This yields the *ℓ-adic Galois representation*

$$\rho_{E,\ell} \colon \mathrm{Gal}(\overline{K}/K) \to \mathrm{Aut}(T_\ell) \simeq \mathrm{GL}_2(\mathbf{Z}_\ell).$$

The representation $\rho_{E,\ell}$ enjoys the following property: for every prime $\mathfrak{p} \nmid \ell$ of good reduction for E the image of $\mathrm{Frob}_\mathfrak{p}$ is a matrix $M_\mathfrak{p} \in \mathrm{GL}_2(\mathbf{Z}_\ell)$ that has the same characteristic polynomial as the Frobenius endomorphism of $E_\mathfrak{p}$, namely, $T^2 - t_\mathfrak{p} T + N(\mathfrak{p})$, where $t_\mathfrak{p} := \operatorname{tr} \pi_{E_\mathfrak{p}}$. The matrix $M_\mathfrak{p}$ is determined only up to conjugacy; there is ambiguity both in our choice of $\mathrm{Frob}_\mathfrak{p}$ (see §1.1) and in our choice of a basis for T_ℓ giving the isomorphism $\mathrm{Aut}(T_\ell) \simeq \mathrm{GL}_2(\mathbf{Z}_\ell)$. We should thus think of $\rho_{E,\ell}(\mathrm{Frob}_\mathfrak{p})$ as representing a conjugacy class in $\mathrm{GL}_2(\mathbf{Z}_\ell)$.

We prefer to work over a field, so let us define the *rational Tate module*

$$V_\ell := T_\ell \otimes_{\mathbf{Z}} \mathbf{Q},$$

which is a 2-dimensional $\mathbf{Q}_\ell$-vector space on which $\mathrm{Gal}(\overline{K}/K)$ acts (trivially on $\mathbf{Q}$ of course). This allows us to view $\rho_{E,\ell}$ as having image $G_\ell \subseteq \mathrm{GL}_2(\mathbf{Q}_\ell)$. We also prefer to work with an algebraic group, so let us define G_ℓ^{zar} to be the $\mathbf{Q}_\ell$-algebraic group obtained by taking the Zariski closure of G_ℓ in $\mathrm{GL}_2(\mathbf{Q}_\ell)$; the group G_ℓ^{zar} is the *ℓ-adic monodromy group* of E (it may also be denoted G_ℓ^{alg}).

Background 3.1 (algebraic groups)**.** An affine (equivalently, linear) *algebraic group* over a field k is a group object in the category of (not necessarily irreducible) affine varieties over k. The only projective algebraic groups we shall consider are smooth and connected, hence abelian varieties, so when we use the term algebraic group without qualification, we mean an affine algebraic group.[13] The canonical example is GL_n, which can be defined as an affine variety in $\mathbf{A}^{n^2+1}$ (over any field) by the equation $t \det M = 1$ (here $\det M$ denotes the determinant polynomial in n^2 variables M_{ij}), with morphisms $m \colon \mathrm{GL}_n \times \mathrm{GL}_n \to \mathrm{GL}_n$ and $i \colon \mathrm{GL}_n \to \mathrm{GL}_n$ defined by polynomial maps corresponding to matrix multiplication and inversion (one uses t as the inverse of $\det A$ when defining i). The classical

[13] There are interesting algebraic groups (group schemes of finite type over a field) that are neither affine nor projective (even if we restrict our attention to those that are smooth and connected), but we shall not consider them here.

groups SL_n, Sp_{2n},U_n, SU_n, O_n, SO_n are all affine algebraic groups (assume $\mathrm{char}(k) \neq 2$ for O_n and SO_n), as are the groups $\mathrm{USp}_{2n} := \mathrm{Sp}_{2n} \cap \mathrm{U}_{2n}$ and GSp_{2n} that are of particular interest to us (see below); the $\mathbf{R}$ and $\mathbf{C}$ points of these groups are *Lie groups* (differentiable manifolds with a group structure). If G is an affine algebraic group over k and L/k is a field extension, the Zariski closure of a subgroup $H \subseteq G(L)$ of the L-points of G is an affine variety over L (the minimal one containing H), and it turns out to also be a group under the morphisms m and i defining G; this makes it an algebraic group, even though H need not be. The connected and irreducible components of an algebraic group G coincide, and are necessarily finite in number. The connected component G^0 of the identity is itself an algebraic group, it is a normal subgroup of G, and invariant under base change. For more on algebraic groups see any of the classic texts [9, 41, 79], or for a more modern treatment, check out Milne's notes [53].

Having defined the $\mathbf{Q}_\ell$-algebraic group G_ℓ^{zar}, we now restrict our attention to the subgroup $G_\ell^{1,\mathrm{zar}}$ obtained by imposing the symplectic constraint

$$M^t \Omega M = \Omega, \qquad \Omega := \left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right),$$

which corresponds to putting a *symplectic form* (a nondegenerate bilinear alternating pairing) on the vector space V_ℓ (we could of course choose any Ω that defines such a form). This condition can clearly be expressed by a polynomial (a quadratic form in fact), thus $G_\ell^{1,\mathrm{zar}}$ is an algebraic group over $\mathbf{Q}_\ell$ contained in Sp_2. Now in fact $\mathrm{Sp}_2 = \mathrm{SL}_2$, so we could have just required $\det M = 1$, but this is an accident of low dimension: the inclusion $\mathrm{Sp}_{2n} \subseteq \mathrm{SL}_{2n}$ is strict for all $n > 1$.

Finally, we choose an embedding $\iota\colon \mathbf{Q}_\ell \to \mathbf{C}$ and let $G_{\ell,\iota}^{1,\mathrm{zar}}$ be the $\mathbf{C}$-algebraic group obtained from $G_\ell^{1,\mathrm{zar}}$ by base change to $\mathbf{C}$ (via ι).[14] The group $G_{\ell,\iota}^{1,\mathrm{zar}}(\mathbf{C})$ is a subgroup of $\mathrm{Sp}_2(\mathbf{C})$ that we may view as a Lie group with finitely many connected components. It therefore contains a maximal compact subgroup that is unique up to conjugacy [61, Thm. IV.3.5], and we take this as the *Sato–Tate group* $\mathrm{ST}(E)$ of E (which is thus defined only up to conjugacy). It is a compact subgroup of $\mathrm{USp}(2) = \mathrm{SU}(2)$ (this equality is another accident of low dimension).

For each prime $\mathfrak{p} \nmid \ell$ of good reduction for E, let $M_\mathfrak{p}$ denote the image of $\mathrm{Frob}_\mathfrak{p}$ under the maps

$$\mathrm{Gal}(\overline{K}/K) \xrightarrow{\rho_{E,\ell}} G_\ell \hookrightarrow G_\ell^{\mathrm{zar}}(\mathbf{Q}_\ell) \hookrightarrow G_{\ell,\iota}^{\mathrm{zar}}(\mathbf{C}),$$

where the map in the middle is inclusion and we use the embedding $\iota\colon \mathbf{Q}_\ell \to \mathbf{C}$ to obtain the last map. We now consider the normalized Frobenius image

$$\bar{M}_\mathfrak{p} := N(\mathfrak{p})^{-1/2} M_\mathfrak{p};$$

it is a matrix with trace $t_\mathfrak{p}/N(\mathfrak{p})^{-1/2} \in [-2,2]$ and determinant 1, and its eigenvalues $e^{\pm i\theta_\mathfrak{p}}$ lie on the unit circle.[15] The eigenangle $\theta_\mathfrak{p}$ determines a unique conjugacy class in $\mathrm{ST}(E)$, which we take as $x_\mathfrak{p}$. The characteristic polynomial of $x_\mathfrak{p}$ is the normalized L-polynomial $\bar{L}_\mathfrak{p}(T) := L_p(N(\mathfrak{p})^{-1/2}T)$, where $L_\mathfrak{p}(T)$ is the numerator of the zeta function of $E_\mathfrak{p}$, and $L_\mathfrak{p}(N(\mathfrak{p})^{-s})$ is the Euler factor at $\mathfrak{p}$ in the L-series $L(E,s)$.

The Sato–Tate conjecture then amounts to the statement that the sequence $(x_\mathfrak{p})$ in $X := \mathrm{conj}(\mathrm{ST}(E))$ is equidistributed. Notice that the statement is the same in both the CM and non-CM and CM cases,

[14] Yes the notation $G_{\ell,\iota}^{1,\mathrm{zar}}$ is a bit over the top, but it is the last stepping stone toward the end goal $\mathrm{ST}(E)$, at which point we can forget it; it is useful to let the notation keep track of which step we are on so we don't get lost along the way.

[15] Note that we embed $G_\ell^{\mathrm{zar}}(\mathbf{Q}_\ell)$ in $G_{\ell,\iota}^{\mathrm{zar}}(\mathbf{C})$ before normalizing by $N(\mathfrak{p})^{-1/2}$; as pointed out by Serre [74, p. 131], we want to take the square root in $\mathbf{C}$ where it is unambiguously defined.

but the measure on X is different, because $\mathrm{ST}(E)$ is different. Indeed, there are three possibilities for $\mathrm{ST}(E)$, corresponding to the three distributions that we noted at the start of this lecture.

Theorem 3.2. *Let E be an elliptic curve over a number field K. Up to conjugacy in $\mathrm{SU}(2)$ we have*

$$\mathrm{ST}(E) = \begin{cases} \mathrm{U}(1) & \text{if } E \text{ has CM defined over } K, \\ N(\mathrm{U}(1)) & \text{if } E \text{ has CM not defined over } K, \\ \mathrm{SU}(2) & \text{if } E \text{ does not have CM,} \end{cases}$$

where $\mathrm{U}(1)$ is embedded in $\mathrm{SU}(2)$ via $u \mapsto \left(\begin{smallmatrix} u & 0 \\ 0 & \bar{u} \end{smallmatrix}\right)$.

Proof. If E has CM defined over K then G_ℓ is abelian, because the action of $\mathrm{Gal}(\overline{K}/K)$ on V_ℓ factors through the abelian group $\mathrm{Gal}(L/K)$, where $L := K(E[\ell^\infty])$ is obtained by adjoining the coordinates of the ℓ-power torsion points of E (this follows from [77, Thm. II.2.3]). Therefore G_ℓ lies in a *Cartan subgroup* of $\mathrm{GL}_2(\mathbf{Q}_\ell)$ (a maximal abelian subgroup), which necessarily splits when we pass to $G_{\ell,\iota}^{\mathrm{zar}}(\mathbf{C})$, where it is conjugate to the group of diagonal matrices. This implies that $\mathrm{ST}(E)$ is conjugate to $\mathrm{U}(1)$, the subgroup of diagonal matrices in $\mathrm{SU}(2)$.

If E has CM not defined over K, then G_ℓ lies in the normalizer of a Cartan subgroup of $\mathrm{GL}_2(\mathbf{Q}_\ell)$, but not in the Cartan itself, and $\mathrm{ST}(E)$ is conjugate to the normalizer $N(\mathrm{U}(1))$ of $\mathrm{U}(1)$ in $\mathrm{SU}(2)$; the argument is as above, but now the action of $\mathrm{Gal}(\overline{K}/K)$ factors through $\mathrm{Gal}(FL/K)$, where F is the CM field and $\mathrm{Gal}(FL/K)$ contains the abelian subgroup $\mathrm{Gal}(FL/FK)$ with index 2.

If E does not have CM then Serre's open image theorem (see [68, §IV.3] and [69]) implies that G_ℓ is a finite index subgroup of $\mathrm{GL}_2(\mathbf{Z}_\ell)$; we therefore have $G_\ell^{1,\mathrm{zar}} = \mathrm{SL}_2$, which implies $\mathrm{ST}(E) = \mathrm{SU}(2)$. □

It follows from Theorem 3.2 that (up to conjugacy), the Sato–Tate group $\mathrm{ST}(E)$ does not depend on our choice of the prime ℓ or the embedding $\iota\colon \mathbf{Q}_\ell \to \mathbf{C}$ that we used. We should also note that $\mathrm{ST}(E)$ depends only on the isogeny class of E; this follows from the fact that we used the rational Tate module V_ℓ to define it (indeed, two abelian varieties over a number field are isogenous if and only if their rational Tate modules are isomorphic as Galois modules, by Faltings' isogeny theorem [22], but we are only using the easy direction of this equivalence here).

3.2. **The Sato–Tate group of an abelian variety.** We are now ready to define the Sato–Tate group of an abelian variety over a number field. Recall that an *abelian variety* is a smooth connected projective variety that is also an algebraic group, where the group operations are now given by morphisms of projective varieties; on any affine patch they can be defined by a polynomial map. Remarkably, the fact that abelian varieties are commutative algebraic groups is not part of the definition, it is a consequence; see [52, Cor 1.4]. We also recall that an *isogeny* of abelian varieties is simply an isogeny of algebraic groups, a surjective morphism with finite kernel.

Abelian varieties of dimension g arise naturally as the Jacobian $\mathrm{Jac}(C)$ of a smooth projective curve C/k of genus g. If C has a k-rational point (as when C is an elliptic curve), one can functorially identify $\mathrm{Jac}(C)$ with the *divisor class group* $\mathrm{Pic}^0(C)$, the group of degree-zero divisors modulo principal divisors, but one can unambiguously define $\mathrm{Jac}(C)$ in any case; see [52, Ch. III] for details.

If C is a smooth projective curve over a number field K and $A := \mathrm{Jac}(C)$ is its Jacobian, then for every prime $\mathfrak{p}$ of good reduction for C, the abelian variety A also has good reduction at $\mathfrak{p}$,[16] and the

[16] For $g > 1$ the converse does not hold (in general); this impacts only finitely many primes $\mathfrak{p}$ and will not concern us.

L-polynomial $L_{\mathfrak{p}}(T)$ appearing in the numerator of the zeta function $Z_{C_p}(T)$ is reciprocal to the characteristic polynomial $\chi_p(T)$ of the Frobenius endomorphism $\pi_{A_{\mathfrak{p}}}$ of $A_{\mathfrak{p}}$, which acts on points of A via the $N(\mathfrak{p})$-power Frobenius automorphism (coordinate-wise). In particular, we have the identity

$$L_{\mathfrak{p}}(T) = T^{2g}\chi_{\mathfrak{p}}(T^{-1}), \tag{9}$$

in which both sides are integer polynomials of degree $2g$ whose complex roots have absolute value $N(\mathfrak{p})^{-1/2}$. As with elliptic curves, one can consider the L-function $L(A,s)$ attached to A, which is defined as an Euler product with factors $L_{\mathfrak{p}}(N(\mathfrak{p})^{-s})$ at each prime $\mathfrak{p}$ where A has good reduction.[17] Studying the distribution of the normalized L-polynomials $\bar{L}_{\mathfrak{p}}(T)$ associated to C is thus equivalent to studying the distribution of the normalized characteristic polynomials of $\pi_{A_{\mathfrak{p}}}$, and also equivalent to studying the distribution of the normalized Euler factors of $L(A,s)$.

Remark 3.3. Each of these three perspectives is successively more general than the previous, the last vastly so. There are abelian varieties over K that are not the Jacobian of any curve defined over K, and L-functions that can be written as Euler products over primes of K that are not the L-function of any abelian variety. One can more generally consider the distribution of normalized Euler factors of *motivic L-functions*, which we also expect to be governed by the Haar measure of a Sato-Tate group associated to the underlying motive, as defined in [73, 74]. See [25] for some concrete examples in weight 3.

The recipe for defining the Sato-Tate group $\mathrm{ST}(A)$ of an abelian variety A over a number field K is exactly the same as when $g = 1$; as with elliptic curves, $\mathrm{ST}(A)$ depends only on the isogeny class of A. We proceed as follows:

1. Pick a prime ℓ, define the Tate module $T_\ell := \varprojlim_n A[\ell^n]$, a free $\mathbf{Z}_\ell$-module of rank $2g$, and the rational Tate module $V_\ell := T_\ell \otimes_{\mathbf{Z}} \mathbf{Q}$, a $\mathbf{Q}_\ell$-vector space of dimension $2g$.
2. Use the Galois representation $\rho_{A,\ell}\colon \mathrm{Gal}(\overline{K}/K) \to \mathrm{Aut}(V_\ell) \simeq \mathrm{GL}_{2g}(\mathbf{Q}_\ell)$ to define $G_\ell := \operatorname{im}\rho_{A,\ell}$.
3. Let G_ℓ^{zar} be the Zariski closure of G_ℓ in $\mathrm{GL}_{2g}(\mathbf{Q}_\ell)$ and define $G_\ell^{1,\mathrm{zar}}$ by adding the symplectic constraint $M^t\Omega M = \Omega$, so that $G_\ell^{1,\mathrm{zar}}$ is a $\mathbf{Q}_\ell$-algebraic subgroup of Sp_{2g}.
4. Pick an embedding $\iota\colon \mathbf{Q}_\ell \to \mathbf{C}$ and use it to define $G_{\ell,\iota}^{1,\mathrm{zar}}$ as the base-change of $G_\ell^{1,\mathrm{zar}}$ to $\mathbf{C}$.
5. Define $\mathrm{ST}(A) \subseteq \mathrm{USp}(2g)$ as a maximal compact subgroup of $G_{\ell,\iota}^{1,\mathrm{zar}}(\mathbf{C})$, unique up to conjugacy.
6. For each good prime $\mathfrak{p} \nmid \ell$, let $M_{\mathfrak{p}}$ be the image of $\mathrm{Frob}_{\mathfrak{p}}$ in $G_{\ell,\iota}^{\mathrm{zar}}(\mathbf{C})$ and define $x_{\mathfrak{p}} \in \mathrm{conj}(\mathrm{ST}(A))$ to be the conjugacy class of $\bar{M}_{\mathfrak{p}} := N(\mathfrak{p})^{-1/2}M_{\mathfrak{p},\iota}$ in $\mathrm{ST}(A)$.

Step 6 requires some justification; it is not obvious why $\bar{M}_{\mathfrak{p}}$ should necessarily be conjugate to an element of $\mathrm{ST}(A)$. Here we are relying on two key facts.

First, the image G_ℓ of $\rho_{A,\ell}$ in $\mathrm{GL}_{2g}(\mathbf{Q}_\ell)$ actually lies in $\mathrm{GSp}_{2g}(\mathbf{Q}_\ell)$, the group of *symplectic similitudes*. The algebraic group GSp_{2g} is defined by imposing the constraint

$$M^t\Omega M = \lambda\Omega, \qquad \Omega := \begin{pmatrix} 0 & -I_g \\ I_g & 0 \end{pmatrix},$$

where λ is necessarily an element of the multiplicative group $\mathbf{G}_m := \mathrm{GL}_1$, since M is invertible. The morphism $\mathrm{GSp}_{2g} \to \mathbf{G}_m$ defined by λ is the *similitude character*, and we have an exact sequence of algebraic groups

$$1 \to \mathrm{Sp}_{2g} \hookrightarrow \mathrm{GSp}_{2g} \xrightarrow{\lambda} \mathbf{G}_m \to 1.$$

[17] Determining the Euler factors at bad primes is hard. Practical methods are known only in special cases, such as when A is the Jacobian of a hyperelliptic curve (even in this case there is much room for improvement).

The action of $\mathrm{Gal}(\overline{K}/K)$ on the Tate module is compatible with the Weil pairing, and this forces the image G_ℓ of $\rho_{E,\ell}$ to lie in $\mathrm{GSp}_{2g}(\mathbf{Q}_\ell)$; see Exercise 3.1. By fixing a symplectic basis for V_ℓ in step 1 we can view $\rho_{A,\ell}$ as a continuous homomorphism

$$\rho_{A,\ell}\colon \mathrm{Gal}(\overline{K}/K) \to \mathrm{GSp}_{2g}(\mathbf{Q}_\ell) \subseteq \mathrm{GL}_{2g}(\mathbf{Q}_\ell)$$

For $g = 1$ we have $\mathrm{GL}_2 = \mathrm{GSp}_2$, but for $g > 1$ the algebraic group GSp_{2g} is properly contained in GL_{2g}.

Second, we are relying on the fact that $M_\mathfrak{p}$, and therefore $\bar{M}_\mathfrak{p}$, is *semisimple* (diagonalizable, since we are working over $\mathbf{C}$). This follows from Tate's proof of the Tate conjecture for abelian varieties over finite fields (combine the main theorem and part (a) of Theorem 2 in [85]). The matrix $\bar{M}_\mathfrak{p}$ is thus diagonalizable and has eigenvalues of absolute value 1; it therefore lies in a compact subgroup of $G_{\ell,\iota}^{1,\mathrm{zar}}(\mathbf{C})$ (take the closure of the group it generates). This compact group is necessarily conjugate to a subgroup of the maximal compact subgroup $\mathrm{ST}(A)$, which must contain an element conjugate to $\bar{M}_\mathfrak{p}$.

Remark 3.4. When defining the Sato-Tate group in more general settings one instead uses the semisimple component of the (multiplicative) Jordan decomposition (see [9, Thm. I.4.4]) of $\bar{M}_\mathfrak{p}$ to define $x_\mathfrak{p}$, as in [74, §8.3.3]. This avoids the need to assume the conjectured *semisimplicity of Frobenius*, which is known for abelian varieties but not in general.

Background 3.5 (Weil pairing)**.** If A is an abelian variety over a field k and $A^\vee$ is its *dual abelian variety* (see [52, §I.8]), then for each $n \geq 1$ prime to the characteristic of k, the *Weil pairing* is a nondegenerate bilinear map

$$A[n] \times A^\vee[n] \to \mu_n(\overline{k})$$

that commutes with the action of $\mathrm{Gal}(\overline{k}/k)$; here μ_n denotes the group of nth roots of unity (the algebraic group defined by $x^n = 1$). Letting n vary over powers of a prime $\ell \neq \mathrm{char}(k)$ and taking inverse limits yields a bilinear map on the corresponding Tate modules:

$$e_\ell\colon T_\ell \times T_\ell^\vee \to \mu_{\ell^\infty}(\overline{k}) := \varprojlim_n \mu_{\ell^n}(\overline{k}).$$

If we have a *polarization*, an isogeny $\phi\colon A \to A^\vee$, we can use it define a bilinear pairing

$$\begin{aligned} e_\ell^\phi\colon T_\ell \times T_\ell &\to \mu_{\ell^\infty}(\overline{k}) \\ (x,y) &\mapsto e_\ell(x,\phi(y)) \end{aligned}$$

that is also compatible with the action of $\mathrm{Gal}(\overline{k}/k)$. One can always choose a polarization ϕ so that the pairing e_ℓ^ϕ is nondegenerate and skew symmetric, meaning that $e_\ell^\phi(a,b) = e_\ell^\phi(b,a)^{-1}$ for all $a, b \in T_\ell$; see [52, Prop. I.13.2]. When A is the Jacobian of a curve it is naturally equipped with a *principal polarization* ϕ, an isomorphism $A \xrightarrow{\sim} A^\vee$, for which this automatically holds; in this situation it is common to simply identify e_ℓ with e_ℓ^ϕ without mentioning ϕ explicitly.

3.3. **The identity component of the Sato-Tate group.** There are two algebraic groups that one can associate to an abelian variety A over a number field K that are closely related to its Sato–Tate group, the *Mumford–Tate group* and the *Hodge group*, both of which conjecturally determine the identity component of the Sato–Tate group (provably so whenever the Mumford–Tate conjecture is known to hold for A, which includes all A of dimension $g \leq 3$). In order to define these groups we need to recall some facts about complex abelian varieties and their associated Hodge structures.

Background 3.6 (complex abelian varieties)**.** Let A be an abelian variety of dimension g over $\mathbf{C}$. Then $A(\mathbf{C})$ is a connected compact Lie group and therefore isomorphic to a torus V/Λ, where $V \simeq \mathbf{C}^g$ is a

complex vector space of dimension g and $\Lambda \simeq \mathbf{Z}^{2g}$ is a full lattice in V that we view as a free $\mathbf{Z}$-module; one can obtain Λ as the kernel of the exponential map $\exp\colon T_0(A(\mathbf{C})) \to A(\mathbf{C})$, where $T_0(A(\mathbf{C}))$ denotes the tangent space at the identity. While every complex abelian variety corresponds to a complex torus, the converse is true only when $g = 1$. The complex tori $X := V/\Lambda$ that correspond to abelian varieties are those that admit a *polarization* (or *Riemann form*), a positive definite Hermitian form $H\colon V \times V \to \mathbf{C}$ with $\operatorname{Im} H(\Lambda, \Lambda) = \mathbf{Z}$ (here Im means imaginary part). Given a polarization H on X, the map $v \mapsto H(v, \cdot)$ defines an isogeny to the *dual torus* $X^\vee := V^*/\Lambda^*$, where $V^* := \{f \in \operatorname{Hom}(V^+, \mathbf{C}^+) : f(\alpha v) = \bar{\alpha} f(v)\}$ and $\Lambda^* := \{f \in V^* : \operatorname{Im} f(\Lambda) \subseteq \mathbf{Z}\}$. This isogeny is a polarization of X as an abelian variety; conversely, any polarization on A (one always exists) can be used to define a polarization on the complex torus $A(\mathbf{C})$. One can then show that the map $A \mapsto A(\mathbf{C})$ defines an equivalence of categories between complex abelian varieties and polarizable complex tori. For more background on complex abelian varieties, see the overviews in [52, §1] or [57, §1], or see [7] for a comprehensive treatment.

Let us fix an embedding $K \hookrightarrow \mathbf{C}$ so that we can work with the complex abelian variety $A_{\mathbf{C}}$ (the base change of A to $\mathbf{C}$), and let $\mathbf{C}^g/\Lambda$ be the corresponding complex torus. We may identify Λ with the singular homology group $H_1(A_{\mathbf{C}}, \mathbf{Z})$, and we similarly have $\Lambda_R := \Lambda \otimes_{\mathbf{Z}} R \simeq H_1(A_{\mathbf{C}}, R)$ for any ring R.

The isomorphisms $A_{\mathbf{C}} \simeq \mathbf{C}^g/\Lambda$ and $A_{\mathbf{C}} \simeq \mathbf{R}^{2g}/\Lambda$ of complex and real Lie groups allow us to view

$$\Lambda_{\mathbf{R}} \simeq H_1(A_{\mathbf{C}}, \mathbf{R})$$

as a real vector space of dimension $2g$ equipped with a *complex structure*, by which we mean an $\mathbf{R}$-algebra homomorphism $h\colon \mathbf{C} \to \operatorname{End}(\Lambda_{\mathbf{R}})$. In the language of Hodge theory, this amounts to the statement that (Λ, h) is an *integral Hodge structure* (pure of weight -1).

We can also view h as morphism of $\mathbf{R}$-algebraic groups $h\colon \mathbf{S} \to \mathrm{GL}(\Lambda_{\mathbf{R}})$. Here $\mathbf{S}$ denotes the *Deligne torus*, obtained by viewing $\mathbf{C}^\times$ as an $\mathbf{R}$-algebraic group (this amounts to taking the restriction of scalars of $\mathbf{G}_m := \mathrm{GL}_1$ from $\mathbf{C}$ to $\mathbf{R}$, see Exercise 3.2), and we view $\mathrm{GL}(\Lambda_{\mathbf{R}})$ as an $\mathbf{R}$-algebraic group by taking its Zariski closure in GL_{2g}. The fact that h can be defined over $\mathbf{R}$ follows from the fact that $\mathbf{C}^g/\Lambda$ is a polarizable torus, since it comes from an abelian variety (in general this need not hold). The real Lie group $\mathbf{S}(\mathbf{R}) \simeq \mathbf{C}^\times$, is generated by $\mathbf{R}^\times$ and $\mathrm{U}(1) = \{z \in \mathbf{C}^\times : z\bar{z} = 1\}$, which intersect in $\{\pm 1\}$; taking Zariski closures yields $\mathbf{R}$-algebraic subgroups $\mathbf{G}_m$ and U_1 of $\mathbf{S}$ that intersect in μ_2. Restricting h to $\mathrm{U}_1 \subseteq \mathbf{S}$ yields a map $\mathrm{U}(1) \to \mathrm{GL}(\Lambda_{\mathbf{R}})$ with the following property: the image of each $u \in \mathrm{U}(1)$ is an element of $\mathrm{GL}(\Lambda_{\mathbf{R}})$ with eigenvalues u, u^{-1}, each of multiplicity g; see [7, Prop. 17.1.1]. The image of such a map is known as a *Hodge circle*.

We now want to consider the *rational Hodge structure* $(\Lambda_{\mathbf{Q}}, h)$.

Definition 3.7. The *Mumford–Tate group* $\mathrm{MT}(A)$ as the smallest $\mathbf{Q}$-algebraic group G in $\mathrm{GL}(\Lambda_{\mathbf{Q}})$ for which $h(\mathbf{S}) \subseteq G(\mathbf{R})$; equivalently, it is the $\mathbf{Q}$-Zariski closure of $h(\mathbf{S})$ in $\mathrm{GL}(\Lambda_{\mathbf{R}})$. The *Hodge group* $\mathrm{Hg}(A)$ is similarly defined as the $\mathbf{Q}$-Zariski closure of $h(\mathrm{U}_1)$ in $\mathrm{GL}(\Lambda_{\mathbf{R}})$.

As defined above, the Mumford–Tate group $\mathrm{MT}(A)$ is a $\mathbf{Q}$-algebraic subgroup of GL_{2g}. But the complex torus $\mathbf{C}^g/\Lambda$ is polarizable, which means that we can put a symplectic form on Λ_R that is compatible with h, and this implies that in fact $\mathrm{MT}(A)$ is a $\mathbf{Q}$-algebraic subgroup of GSp_{2g}. Similarly, the Hodge group $\mathrm{Hg}(A)$ is a $\mathbf{Q}$-algebraic subgroup of Sp_{2g}, and in fact $\mathrm{Hg}(A) = \mathrm{MT}(A) \cap \mathrm{Sp}_{2g}$; this is sometimes used as an alternative definition of $\mathrm{Hg}(A)$. Much of the interest in the Hodge group arises from the fact that it gives us an isomorphism of $\mathbf{Q}$-algebras

$$\operatorname{End}(A_{\mathbf{C}})_{\mathbf{Q}} \simeq \operatorname{End}(\Lambda_{\mathbf{Q}})^{\mathrm{Hg}(A)},$$

where $\mathrm{End}(A_{\mathbf{C}})_{\mathbf{Q}} := \mathrm{End}(A_{\mathbf{C}}) \otimes_{\mathbf{Z}} \mathbf{Q}$ and $\mathrm{Hg}(A)$ acts on $\mathrm{End}(\Lambda_{\mathbf{Q}})$ by conjugation; see [7, Prop. 17.3.4]. To see why this isomorphism is useful, let us note one application.

Theorem 3.8. *For an abelian variety A of dimension g over a number field K, the Hodge group $\mathrm{Hg}(A)$ is commutative if and only if the endomorphism algebra $\mathrm{End}(A_{\overline{K}})_{\mathbf{Q}}$ contains a commutative semisimple $\mathbf{Q}$-algebra of dimension $2g$.*

Proof. See [7, Prop. 17.3.5]. □

Note that when $g = 1$ the abelian varieties A that satisfy the two equivalent properties of Theorem 3.8 are CM elliptic curves. For $g \geq 1$, such abelian varieties are said to be of *CM-type*. For abelian varieties of general type one has the opposite extreme: $\mathrm{End}(A_{\overline{K}})_{\mathbf{Q}} = \mathbf{Q}$ and $\mathrm{Hg}(A) = \mathrm{Sp}_{2g}$; see [7, Prop. 17.4.2].

In the previous section we defined two $\mathbf{Q}_\ell$-algebraic groups $G_\ell^{\mathrm{zar}} \subseteq \mathrm{GSp}_{2g}$ and $G_\ell^{1,\mathrm{zar}} \subseteq \mathrm{Sp}_{2g}$ associated to A. It is reasonable to ask how they are related to the $\mathbf{Q}$-algebraic groups $\mathrm{MT}(A)$ and $\mathrm{Hg}(A)$. Unlike the groups G_ℓ^{zar} and $G_\ell^{1,\mathrm{zar}}$, the algebraic groups $\mathrm{MT}(A)$ and $\mathrm{Hg}(A)$ are necessarily connected (by construction).[18] Deligne proved that the identity component of G_ℓ^{zar} is always a subgroup of $\mathrm{MT}(A) \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$, equivalently, that the identity component of $G_\ell^{1,\mathrm{zar}}$ is a subgroup of $\mathrm{Hg}(A) \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$); see [19]. It is conjectured that these inclusions are in fact equalities.

Conjecture 3.9 (Mumford–Tate Conjecture)**.** *The identity component of G_ℓ^{zar} is equal to $\mathrm{MT}(A) \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$; equivalently, the identity component of $G_\ell^{1,\mathrm{zar}}$ is equal to $\mathrm{Hg}(A) \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$.*

This conjecture is known to hold for abelian varieties of dimension $g \leq 3$; see [3, Th. 6.11], where it is shown that this follows from [55]. When it holds, the Mumford–Tate group (and the Hodge group) uniquely determines the identity component of the Sato–Tate group, up to conjugation in $\mathrm{USp}(2g)$; see [24, Lemma 2.8]. Neither the Mumford–Tate group nor the Hodge group tell us anything about the component groups of G_ℓ^{zar}, $G_\ell^{1,\mathrm{zar}}$, $\mathrm{ST}(A)$ (the three are isomorphic, see [74, §8.3.4]), but there is a closely related $\mathbf{Q}$-algebraic group that conjecturally does.

Conjecture 3.10 (Algebraic Sato–Tate Conjecture)**.** *There exists a $\mathbf{Q}$-algebraic subgroup $\mathrm{AST}(A)$, of Sp_{2g} such that $G_\ell^{1,\mathrm{zar}} = \mathrm{AST}(A) \otimes_{\mathbf{Q}} \mathbf{Q}_\ell$.*

Banaszak and Kedlaya have shown that this conjecture holds for $g \leq 3$; additionally, they are able to give a very explicit description of $\mathrm{AST}(A)$ using *twisted Lefschetz groups*; see [3].

3.4. **The component group of the Sato-Tate group.** We have seen that the Mumford–Tate group conjecturally determines the identity component $\mathrm{ST}(A)^0$ of the Sato–Tate group $\mathrm{ST}(A)$ of an abelian variety A over a number field K (provably so in dimension $g \leq 3$). The identity component $\mathrm{ST}(A)^0$ is a normal finite index subgroup of $\mathrm{ST}(A)$, and we now want to consider the component group $\mathrm{ST}(A)/\mathrm{ST}(A)^0$. As above, for any field extension L/K, we use A_L to denote the base change of A to L.

Theorem 3.11. *Let A be an abelian variety over a number field K. There is a unique finite Galois extension L/K with the property that $\mathrm{ST}(A_L)$ is connected and $\mathrm{Gal}(L/K) \simeq \mathrm{ST}(A)/\mathrm{ST}(A)^0$. The extension L/K is unramified outside the primes of bad reduction for A, and for every subextension F/K of L/K we have $\mathrm{Gal}(L/F) \simeq \mathrm{ST}(A_F)/\mathrm{ST}(A_F)^0$.*

[18] This is true more generally for all motives of odd weight. For motives of even weight the situation is more delicate; complications arise from the fact that we are then working with orthogonal groups rather than symplectic groups; see [3, 4].

Proof. As explained in [74, §8.3.4], the component groups of G_ℓ^{zar} and $\mathrm{ST}(A)$ are isomorphic. Let Γ be the Galois group of the maximal subextension K_{S_ℓ} of $\mathrm{Gal}(\overline{K}/K)$ that is unramified away from the set S_ℓ consisting of the primes of bad reduction for A and the primes of K lying above ℓ. The ℓ-adic Galois representation $\rho_{A,\ell}\colon \mathrm{Gal}(\overline{K}/K) \to \mathrm{Aut}(V_\ell)$ induces a continuous surjective homomorphism

$$\Gamma \to G_\ell^{\mathrm{zar}}/(G_\ell^{\mathrm{zar}})^0$$

whose kernel is a normal open subgroup Γ_0 of Γ. The corresponding fixed field L is a finite Galois extension of K, and it is the minimal Galois extension of K for which $\mathrm{ST}(A_L)$ is connected. It is clearly uniquely determined and unramified outside S_ℓ, and we have isomorphisms

$$\mathrm{Gal}(L/K) \simeq \Gamma/\Gamma_0 \simeq G_\ell^{\mathrm{zar}}/(G_\ell^{\mathrm{zar}})^0 \simeq \mathrm{ST}(A)/\mathrm{ST}(A)^0.$$

As shown by Serre [72], the component group of G_ℓ^{zar} (and therefore $\mathrm{ST}(A)$) is independent of ℓ, and the above argument applies to any choice of ℓ. Thus L/K can be ramified only at primes of bad reduction for A. For any subextension F/K of L/K, replacing A by A_F in the argument above yields the same field L, with $\mathrm{Gal}(L/F) \simeq \mathrm{ST}(A_F)/\mathrm{ST}(A_F)^0$. □

3.5. **Exercises.**

Exercise 3.1. Let A be an abelian variety of dimension g over a number field K. Show that one can choose a basis for $V_\ell = T_\ell \otimes_{\mathbf{Z}} \mathbf{Q}$ so that the matrix M describing the action of any $\sigma \in \mathrm{Gal}(\overline{K}/K)$ on V_ℓ satisfies $M^t \Omega M = \lambda\Omega$ for some $\lambda \in \mathbf{Q}_\ell^\times$, where $\Omega := \left(\begin{smallmatrix} 0 & -I \\ I & 0 \end{smallmatrix}\right)$. Conclude that the image of the corresponding Galois representation lies in $\mathrm{GSp}_{2g}(\mathbf{Q}_\ell)$ and describe the map $\mathrm{Gal}(\overline{K}/K) \to \mathbf{Q}_\ell^\times$ induced by the similitude character λ.

Exercise 3.2. Write down an explicit description of the Deligne torus $\mathbf{S}$ as an $\mathbf{R}$-algebraic group in $\mathbf{A}^4$ (give equations that define it as an affine variety and polynomial maps for the group operations) and express the $\mathbf{R}$-algebraic groups $\mathbf{G}_m$ and U_1 as subgroups of $\mathbf{S}$ that intersect in μ_2. Then prove that $\mathbf{S}(\mathbf{R})$ and $\mathbf{C}^\times$ are isomorphic as real Lie groups (give explicit maps in both directions).

Exercise 3.3. Let L/K be a finite separable field extension of degree d, say $L = K(\alpha)$. Given an affine L-variety Y defined by polynomials $P_k \in L[y_1, \dots, y_n]$, we can construct an affine K-variety $\mathrm{Res}_{L/K}(Y)$ by writing each $y_i = \sum_{j=0}^{d-1} x_{ij}\alpha^j$ in terms of the K-basis $\{1, \alpha, \dots, \alpha^{d-1}\}$ for L and using the minimal polynomial of α to replace each $P_k(y_1, \dots, y_n)$ by a polynomial in $K[x_{11}, \dots, x_{1d}, \dots, x_{n1} \dots, x_{nd}]$. The K-variety $\mathrm{Res}_{L/K}(Y)$ is the *Weil restriction* (or *restriction of scalars*) of Y. The map $Y \to \mathrm{Res}_{L/K}(Y)$ defines a functor from the category of affine L-varieties to the category of affine K-varieties that restricts to a functor of affine algebraic groups; it is adjoint to the base-change functor $X \to X_L$, also known as *extension of scalars*. Show that the $\mathbf{R}$-algebraic group $\mathbf{S}$ defined in 3.2 is the Weil restriction of the $\mathbf{C}$-algebraic group $\mathbf{G}_m$.

4. Sato–Tate axioms and Galois endomorphism types

4.1. **Sato–Tate axioms.** In [74, §8.2], Serre lists a set of axioms that any Sato–Tate group is expected to satisfy. Serre considers Sato–Tate groups in a more general context than we do here, so we will state the axioms as they apply to Sato–Tate groups of abelian varieties. As in §3.3, for a Lie group G we define a *Hodge circle* to be a subgroup H of G that is the image of a continuous homomorphism $\theta\colon \mathrm{U}(1) \to G^0$ whose elements $\theta(u)$ have eigenvalues u and u^{-1} with multiplicity g (note that H necessarily lies in the identity component G^0 of G).

Definition 4.1. A group G satisfies the *Sato–Tate axioms* (for abelian varieties of dimension g) if and only if the following hold:

(ST1) (Lie condition) G is a closed subgroup of $\mathrm{USp}(2g)$.

(ST2) (Hodge condition) The Hodge circles in G generate a dense nontrivial subgroup of G^0.[19]

(ST3) (rationality condition) For each component H of G and irreducible character χ of $\mathrm{GL}_{2g}(\mathbf{C})$, we have $\int_H \chi\mu \in \mathbf{Z}$, where μ is the Haar measure on G normalized so that $\mu(\mathbb{1}_H) = 1$.

Remark 4.2. Definition 4.1 generalizes easily to self-dual motives with rational coefficients. Given an integer weight $w \geq 0$ and Hodge numbers $h^{p,q} \in \mathbf{Z}_{\geq 0}$ indexed by $p, q \in \mathbf{Z}_{\geq 0}$ with $p + q = w$ such that $h^{p,q} = h^{q,p}$ when w is odd, let $d := \sum h^{p,q}$. For abelian varieties we have $w = 1$ and $h^{1,0} = h^{0,1} = g$. In axiom (ST1) we require G to be a closed subgroup of $\mathrm{USp}(d)$ (resp. $\mathrm{O}(d)$) when w is odd (resp. even), and in axiom (ST2) we require elements $\theta(u)$ of a Hodge circle to have eigenvalues u^{p-q} with multiplicity $h^{p,q}$; axiom (ST3) is unchanged.

Axiom (ST1) implies that G is a compact Lie group, and (ST2) rules out finite groups, since G must contain at least one Hodge circle and therefore contains a subgroup isomorphic to $\mathrm{U}(1)$.[20] When G is connected, (ST3) holds automatically and only (ST1) and (ST2) need to be checked; this is an easy application of representation theory, see [48, Prop. 2]. Axiom (ST3) also plays no role when $g = 1$ (see the proof of Proposition 4.4 below), but for $g > 1$ it is crucial. When $g = 2$, for example, for every integer $n \geq 1$ we can diagonally embed $\mathrm{U}(1) \times \mathrm{U}(1)[n]$ in $\mathrm{USp}(4)$ to get infinitely many non-conjugate closed groups $G \subseteq \mathrm{USp}(4)$ whose identity component is a Hodge circle. All these groups G satisfy (ST1) and (ST2), but only finitely many satisfy (ST3). Indeed, if we take χ and let C be a component on which the projection to $\mathrm{U}(1)[n]$ has order n, we have

$$\int_C \chi\mu = \zeta_n + \bar{\zeta}_n \in \mathbf{Z}$$

only for $n \in \{2, 3, 4, 6\}$. More generally. we have the following theorem.

Theorem 4.3. *Up to conjugacy, for any fixed dimension $g \geq 1$ the number of subgroups of $\mathrm{USp}(2g)$ that satisfy the Sato–Tate axioms is finite.*

Proof. See [24, Rem. 3.3] □

Theorem 4.3 motivates the following *classification problem*: given an integer $g \geq 1$, determine the subgroups of $\mathrm{USp}(2g)$ that satisfy the Sato–Tate axioms. The case $g = 1$ is easy.

Proposition 4.4. *For $g = 1$ the three groups $\mathrm{U}(1)$, $N(\mathrm{U}(1)$ and $\mathrm{SU}(2)$ listed in Theorem 3.2 are the only groups that satisfy the Sato–Tate axioms (up to conjugacy).*

[19] The statement of (ST2) in [24] inadvertently omits the requirement that the Hodge circles generate a dense subgroup.

[20] Except in weight 0 where Hodge circles become trivial and the Hodge condition actually forces G to be finite.

Proof. Suppose G satisfies the Sato–Tate axioms. Then G^0 contains a conjugate of $\mathrm{U}(1)$ embedded in $\mathrm{USp}(2)$ via $u \mapsto \left(\begin{smallmatrix} u & 0 \\ 0 & \bar{u} \end{smallmatrix}\right)$, as in Theorem 3.2, and it must be a compact connected Lie group. The only compact connected Lie groups in $\mathrm{USp}(2) = \mathrm{SU}(2)$ are $\mathrm{U}(1)$ and $\mathrm{SU}(2)$ itself (this follows from the classification of compact connected Lie groups but is easy to see directly). Thus either $G^0 = \mathrm{SU}(2)$, in which case $G = \mathrm{SU}(2)$, or G^0 is conjugate to $\mathrm{U}(1)$ and must be a normal subgroup of G (the identity component of a compact Lie group is always a normal subgroup of finite index). The group $\mathrm{U}(1)$ has index 2 in its normalizer, so $\mathrm{U}(1)$ and $N(\mathrm{U}(1))$ are the only possibilities for G when $G^0 = \mathrm{U}(1)$. □

Corollary 4.5. *For $g = 1$ a group G satisfies the Sato–Tate axioms if and only if it is the Sato–Tate group of an elliptic curve over a number field.*

The classification problem for $g = 2$ is more difficult, but it has been solved.

Theorem 4.6. *Up to conjugacy in* $\mathrm{USp}(4)$ *there are* 55 *groups that satisfy the Sato–Tate axioms for $g = 2$. Of these* 55*, the following* 6 *are connected:*

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

were $\mathrm{U}(1)_2$ *denotes* $\mathrm{U}(1) = \left\{\left(\begin{smallmatrix} u & 0 \\ 0 & \bar{u} \end{smallmatrix}\right) : u \in \mathbf{C}^\times\right\}$ *diagonally embedded in* $\mathrm{USp}(4)$*, and similarly for* $\mathrm{SU}(2)_2$.

Proof. See [24, Thm. 3.4], which also gives an explicit description of the 55 groups. □

Remark 4.7. Those familiar with the classification of connected compact Lie groups may notice that the group $\mathrm{U}(2)$, which can be embedded in $\mathrm{USp}(4)$, is missing from Theorem 4.6. This is because it fails to satisfy the Hodge condition (ST2); it contains subgroups isomorphic to $\mathrm{U}(1)$, but there is no way to embed $\mathrm{U}(1) \hookrightarrow \mathrm{U}(2) \hookrightarrow \mathrm{USp}(4)$ and get eigenvalues u and u^{-1} with multiplicity 2; see [25, Rem. 2.3]. However, for motives of weight 3 and Hodge numbers $h^{3,0} = h^{2,1} = h^{1,2} = h^{0,3} = 1$ the modified Hodge condition noted in Remark 4.2 is satisfied by a subgroup of $\mathrm{USp}(4)$ isomorphic to $\mathrm{U}(2)$; see [25] for details, including two examples of motives with Sato-Tate group $\mathrm{U}(2)$.

Corollary 4.5 does not hold for $g = 2$.

Theorem 4.8. *Of the* 55 *groups appearing in Theorem 4.6, only* 52 *arise as the Sato–Tate group of an abelian surface over a number field. Of these,* 34 *arise for abelian surfaces defined over* $\mathbf{Q}$.

Proof. See [24, Thm. 1.5]. □

The three subgroups of $\mathrm{USp}(4)$ that satisfy the Sato–Tate axioms but are not the Sato–Tate group of any abelian surface over a number field are the normalizer of $\mathrm{U}(1)\times\mathrm{U}(1)$ in $\mathrm{USp}(4)$, whose component group is the dihedral group of order 8, and two of its subgroups, one of index 2 and one of index 4. The proof that these three groups do not occur is obtained by establishing a bijection between *Galois endomorphism types* (as defined in the next section) and Sato–Tate groups and showing that there are only 52 Galois endomorphism types of abelian surfaces. Explicit examples of genus 2 curves whose Jacobians realize these 52 possibilities can be find in [24, Table 11], and animated histograms of their Sato–Tate distributions can be found at

http://math.mit.edu/~drew/g2SatoTateDistributions.html

The classification problem for $g = 3$ remains open, but the connected cases have been provisionally determined (see Table 2 in the next section). Before leaving the discussion of the Sato–Tate axioms, it is reasonable to ask whether Sato–Tate groups necessarily satisfy them. Of course we expect this to be the case, but it is difficult to prove in general. However, it can be proved to hold in all cases where the Mumford–Tate conjecture is known, including all cases with $g \le 3$.

Proposition 4.9. *Let A be an abelian variety of dimension g over a number field K for which the Mumford–Tate conjecture holds. Then* $\mathrm{ST}(A)$ *satisfies the Sato–Tate axioms.*

Proof. See [24, Prop. 3.2]. □

4.2. **Galois endomorphism types.** In Lecture 3 we saw that, at least when the Mumford–Tate conjecture is known, the identity component of the Sato-Tate group can be related to an $\mathbf{R}$-algebra (the $\mathbf{R}$-algebra $\mathrm{End}(\Lambda_{\mathbf{R}})$ that we used to define the Mumford–Tate group), and the component group of the Sato–Tate group can be related to a Galois group (the group $\mathrm{Gal}(L/K)$ given by Theorem 3.11). We now want to associate to each abelian variety A an $\mathbf{R}$-algebra equipped with a Galois action that will allow us to completely determine the Sato-Tate group $\mathrm{ST}(A)$ in many cases.

We will work in the abstract category $\mathscr{C}$ whose objects are pairs (G, E) of a finite group G and an $\mathbf{R}$-algebra E equipped with an $\mathbf{R}$-linear action of G, and whose morphisms $\Phi\colon (G,E) \to (G',E')$ are pairs (ϕ_G, ϕ_E), where $\phi_G\colon G \to G'$ is a morphism of groups, and $\phi_E\colon E \to E'$ is an equivariant morphism of $\mathbf{R}$-algebras, meaning that

$$\phi_E(e^g) = \phi_E(e)^{\phi_G(g)} \qquad \text{for all } g \in G \text{ and } e \in E. \tag{10}$$

To each abelian variety A/K we now associate an isomorphism class $[G,E]$ in $\mathscr{C}$ as follows. The minimal extension L/K for which $\mathrm{End}(A_L) = \mathrm{End}(A_{\overline{K}})$ is a finite Galois extension of K; we shall take G to be $\mathrm{Gal}(L/K)$ and E to be the real endomorphism algebra $\mathrm{End}(A_L)_{\mathbf{R}} \coloneqq \mathrm{End}(A_L) \otimes_{\mathbf{Z}} \mathbf{R}$. The Galois group $\mathrm{Gal}(L/K)$ acts on $\mathrm{End}(A_L)$ via its action on the coefficients of the rational maps defining each element of $\mathrm{End}(A_K)$, and this induces an $\mathbf{R}$-linear action of $\mathrm{Gal}(L/K)$ on $\mathrm{End}(A_L)_{\mathbf{R}}$ via composition with the natural map $\mathrm{End}(A_L) \to \mathrm{End}(A_L)_{\mathbf{R}}$. The pair $(\mathrm{Gal}(L/K), \mathrm{End}(A_L)_{\mathbf{R}})$ is thus an object of $\mathscr{C}$.

Definition 4.10. The *Galois endomorphism type* $\mathrm{GT}(A)$ of an abelian variety A/K is the isomorphism class of the pair $(\mathrm{Gal}(L/K), \mathrm{End}(A_L)_{\mathbf{R}})$ in the category $\mathscr{C}$, where L is the minimal extension of K for which $\mathrm{End}(A_L) = \mathrm{End}(A_{\overline{K}})$.

Example 4.11. Let E be an elliptic curve over a number field K. If E does not have CM, or if it has CM defined over K, then its endomorphisms are all defined over $L = K$; otherwise, its endomorphisms are all defined over its CM field L, an imaginary quadratic extension of K. The real endomorphism algebra $\mathrm{End}(E_L)_{\mathbf{R}}$ is isomorphic to $\mathbf{R}$ when E does not have CM, and isomorphic to $\mathbf{C}$ when E does have CM. We therefore have

$$\mathrm{GT}(E) = \begin{cases} [\mathrm{C}_1, \mathbf{C}] & \text{if } E \text{ has CM defined over } K \\ [\mathrm{C}_2, \mathbf{C}] & \text{if } E \text{ has CM not defined over } K \\ [\mathrm{C}_1, \mathbf{R}] & \text{if } E \text{ does not have CM} \end{cases}$$

Here C_n denotes the cyclic group of order n; in the case $[\mathrm{C}_2, \mathbf{C}]$ the action of C_2 on $\mathbf{C}$ corresponds to complex conjugation.

The three Galois endomorphism types listed in Example 4.11 correspond to the three Sato-Tate groups listed in Theorem 3.2. Under this correspondence, the real endomorphism algebra $\mathrm{End}(E_L)_{\mathbf{R}}$ determines the identity component $\mathrm{ST}(E)^0$ (up to conjugacy), and the Galois group $\mathrm{Gal}(L/K)$ is isomorphic to the component group $\mathrm{ST}(E)/\mathrm{ST}(E)^0$. Moreover, the field L is precisely the field L given by Theorem 3.11.

Theorem 4.12. *Let A be an abelian variety A of dimension $g \le 3$ defined over a number field K and let L be the minimal field for which* $\mathrm{End}(A_L) = \mathrm{End}(A_{\overline{K}})$. *The conjugacy class of the Sato-Tate group* $\mathrm{ST}(A)$ *determines the Galois endomorphism type* $\mathrm{GT}(A)$*; moreover, the conjugacy class of the identity component*

$\mathrm{ST}(A)^0$ *determines the isomorphism class of* $\mathrm{End}(A_L)_{\mathbf{R}}$ *and* $\mathrm{ST}(A)/\mathrm{ST}(A)^0 \simeq \mathrm{Gal}(L/K)$. *For* $g \le 2$ *the converse holds: the Galois endomorphism type* $\mathrm{GT}(A)$ *determines the Sato–Tate group* $\mathrm{ST}(A)$ *up to conjugacy.*

Proof. See Proposition 2.19 and Theorem 1.4 in [24]. □

It is expected that in fact the Sato–Tate group always determines the Galois endomorphism type, and that the converse holds for $g \le 3$. For $g = 3$ we at least know that the real endomorphism algebra $\mathrm{End}(A_L)_{\mathbf{R}}$ determines the identity component $\mathrm{ST}(A)^0$ and that $\mathrm{Gal}(L/K) \simeq \mathrm{ST}(A)/\mathrm{ST}(A)^0$. At first glance it might seem that this should determine $\mathrm{ST}(A)$, but it does not, not even when $g = 2$. One needs to also understand how $\mathrm{Gal}(L/K)$ acts on $\mathrm{End}(A_L)_{\mathbf{R}}$ and relate this to the Sato-Tate group $\mathrm{ST}(A)$. In [24] this is accomplished for $g = 2$ by looking at the lattice of $\mathbf{R}$-subalgebras of $\mathrm{End}(A_L)_{\mathbf{R}}$ fixed by subgroups of $\mathrm{Gal}(L/K)$ and showing that this is enough to uniquely determine $\mathrm{ST}(A)$; see [24, Thm. 4.3]. To apply the same approach when $g = 3$ requires a more detailed classification of the Galois endomorphism types and Sato–Tate groups that can arise in this case than is currently available.

For $g = 4$ the Galois endomorphism type does not always determine the Sato–Tate group. This follows from an exceptional example constructed by Mumford in [56], in which he proves the existence of an abelian four-fold A for which $\mathrm{End}(A_{\overline{K}}) = \mathbf{Z}$ but $\mathrm{MT}(A) \ne \mathrm{GSp}_8$. The fact that $\mathrm{MT}(A)$ is properly contained in GSp_8 implies that $\mathrm{ST}(A)$ must be properly contained in $\mathrm{USp}(8)$ (this does not depend on the Mumford–Tate conjecture, here we are only using the inclusion that was proved by Deligne). On the other hand, for an abelian variety of general type one has $\mathrm{End}(A_{\overline{K}}) = \mathbf{Z}$ and $\mathrm{ST}(A) = \mathrm{USp}(2g)$; see [30, 96] for an explicit criterion that applies to almost all Jacobians of hyperelliptic curves.

For $g > 4$ one can construct exceptional examples as a product of an abelian variety with one of Mumford's exceptional four-folds, so in general the Galois endomorphism type cannot determine the Sato–Tate group for any $g \ge 4$. However, such examples will not be simple and will have $\mathrm{End}(A) \ne \mathbf{Z}$. In [71] Serre proves an analog of his open image theorem for elliptic curves that applies to abelian varieties of dimension $g = 2, 6$ and g odd. For these values of g, if $\mathrm{End}(A_{\overline{K}}) = \mathbf{Z}$ then $\mathrm{ST}(A) = \mathrm{USp}(2g)$ and no direct analog of Mumford's construction exists.

Remark 4.13. For $g \le 3$, the field L in Theorem 3.11 (the minimal extension for which $\mathrm{ST}(A_L)$ is connected) is the same as the field L in Theorem 4.12 (the minimal extension for which $\mathrm{End}(A_L) = \mathrm{End}(A_{\overline{K}})$). In any case, the former always contains the latter: if $\mathrm{ST}(A_L)$ is connected then we necessarily have $\mathrm{End}(A_{\overline{K}}) = \mathrm{End}(A_L)$. This can be seen as a consequence of Bogomolov's theorem [8], which states that G_ℓ is open in $G_\ell^{\mathrm{zar}}(\mathbf{Q}_\ell)$, and Faltings' theorem [22] that $\mathrm{End}(A)_{\mathbf{Q}_\ell} \simeq \mathrm{End}(V_\ell(A))^{G_\ell}$. If $\mathrm{ST}(A)$ (and therefore G_ℓ^{zar}), is connected, then $\mathrm{End}(A)$ is invariant under base change (nowâĂć apply this to $A = A_L$).

A complete list of the 52 Galois endomorphism types and corresponding Sato-Tate groups that arise when $g = 2$ can be found in [24, Theorem 4.3] and [24, Table 9]. Here we list only the 6 connected cases, which are determined by the correspondence between $\mathrm{End}(A_K)_{\mathbf{R}}$ and $\mathrm{ST}(A)^0$; see Table 1, which also lists the types of abelian surfaces for which they arise. It is worth noting that the Sato–Tate group is in some respects a rather coarse arithmetic invariant; for example, it cannot distinguish a product of non-CM elliptic curves from a geometrically simple abelian surface with real multiplication (RM).

On the other hand, the Haar measures on these 52 Sato–Tate groups all give rise to distinct distributions of characteristic polynomials, which, under the Sato–Tate conjecture, match the distribution of normalized L-polynomials of the abelian variety, and there are some rather fine distinctions among these that the Sato–Tate group detects. We should note that there are only 36 distinct trace distributions among the 52 groups, one needs to look at both the linear and quadratic coefficients of the characteristic

geometric type of abelian surface	$\mathrm{End}(A_{\overline{K}})_{\mathbf{R}}$	$\mathrm{ST}(A)^0$
square of CM elliptic curve	$\mathrm{M}_2(\mathbf{C})$	$\mathrm{U}(1)_2$
QM abelian surface square of non-CM elliptic curve	$\mathrm{M}_2(\mathbf{R})$	$\mathrm{SU}(2)_2$
CM abelian surface product of CM elliptic curves	$\mathbf{C} \times \mathbf{C}$	$\mathrm{U}(1) \times \mathrm{U}(1)$
product of CM and non-CM elliptic curves	$\mathbf{C} \times \mathbf{R}$	$\mathrm{U}(1) \times \mathrm{SU}(2)$
RM abelian surface product of non-CM elliptic curves	$\mathbf{R} \times \mathbf{R}$	$\mathrm{SU}(2) \times \mathrm{SU}(2)$
abelian surface of general type	$\mathbf{R}$	$\mathrm{USp}(4)$

TABLE 1. Real endomorphism algebras and Sato–Tate identity components for abelian surfaces

geometric type of abelian three-fold	$\mathrm{End}(A_K)_{\mathbf{R}}$	$\mathrm{ST}(A)^0$
cube of a CM EC	$\mathrm{M}_3(\mathbf{C})$	$\mathrm{U}(1)_3$
cube of a non-CM EC	$\mathrm{M}_3(\mathbf{R})$	$\mathrm{SU}(2)_3$
product of CM EC and square of CM EC	$\mathbf{C} \times \mathrm{M}_2(\mathbf{C})$	$\mathrm{U}(1) \times \mathrm{U}(1)_2$
product of CM EC and QM abelian surface product of CM EC and square of non-CM EC	$\mathbf{C} \times \mathrm{M}_2(\mathbf{R})$	$\mathrm{U}(1) \times \mathrm{SU}(2)_2$
product of non-CM EC and square of CM EC	$\mathbf{R} \times \mathrm{M}_2(\mathbf{C})$	$\mathrm{SU}(2) \times \mathrm{U}(1)_2$
product of non-CM EC and QM abelian surface product of non-CM EC and square of non-CM EC	$\mathbf{R} \times \mathrm{M}_2(\mathbf{R})$	$\mathrm{SU}(2) \times \mathrm{SU}(2)_2$
CM abelian threefold product of CM EC and CM abelian surface product of three CM ECs	$\mathbf{C} \times \mathbf{C} \times \mathbf{C}$	$\mathrm{U}(1) \times \mathrm{U}(1) \times \mathrm{U}(1)$
product of non-CM EC and CM abelian surface product of non-CM EC and two CM ECs	$\mathbf{C} \times \mathbf{C} \times \mathbf{R}$	$\mathrm{U}(1) \times \mathrm{U}(1) \times \mathrm{SU}(2)$
product of CM EC and RM abelian surface product of CM EC and two non-CM ECs	$\mathbf{C} \times \mathbf{R} \times \mathbf{R}$	$\mathrm{U}(1) \times \mathrm{SU}(2) \times \mathrm{SU}(2)$
RM abelian threefold product of non-CM EC and RM abelian surface product of 3 non-CM ECs	$\mathbf{R} \times \mathbf{R} \times \mathbf{R}$	$\mathrm{SU}(2) \times \mathrm{SU}(2) \times \mathrm{SU}(2)$
product of CM EC and abelian surface	$\mathbf{C} \times \mathbf{R}$	$\mathrm{U}(1) \times \mathrm{USp}(4)$
product of non-CM EC and abelian surface	$\mathbf{R} \times \mathbf{R}$	$\mathrm{SU}(2) \times \mathrm{USp}(4)$
quadratic CM abelian threefold	$\mathbf{C}$	$\mathrm{U}(3)$
generic abelian threefold	$\mathbf{R}$	$\mathrm{USp}(6)$

TABLE 2. Real endomorphism algebras and Sato–Tate identity components for abelian threefolds

polynomials to distinguish them. We should also note that it is entirely possible for two non-conjugate Sato–Tate groups to be isomorphic as abstract groups yet give rise to distinct trace distributions.

Example 4.14. Consider the hyperelliptic curves

$$C_1\colon y^2 = x^6 + 3x^5 + 15x^4 - 20x^3 + 60x^2 - 60x + 28,$$
$$C_2\colon y^2 = x^6 + 6x^5 - 15x^4 + 20x^3 - 15x^2 + 6x - 1,$$

and let $A_1 := \mathrm{Jac}(C_1)$ and $A_2 := \mathrm{Jac}(C_2)$ denote their Jacobians. Over $\overline{\mathbf{Q}}$ both A_1 and A_2 are isogenous to the square of the elliptic curve $y^2 = x^3 + 1$, which has CM by $\mathbf{Q}(\sqrt{-3})$. We necessarily have $\mathrm{ST}(A_1)^0 = \mathrm{ST}(A_2)^0 = \mathrm{U}(1)_2$, and the component groups are both isomorphic to the dihedral group of order 12. However, their Sato–Tate groups are different: in terms of the labels used in [24], we have $\mathrm{ST}(A_1) = D_{6,1}$, while $\mathrm{ST}(A_2) = D_{6,2}$ (see [24, §3.4] for explicit descriptions of these groups in terms of generators). And their normalized trace distributions are quite different. For C_1 the density of zero traces is $3/4$, whereas for C_2 it is $7/12$ (these ratios represent the proportion of Sato–Tate group components on which the trace is identically zero), and their normalized trace moment sequences are $(1, 0, 1, 0, 9, 0, 110, 0, 1505, 0, 21546, \ldots)$ and $(1, 0, 2, 0, 18, 0, 200, 0, 2450, 0, 31752, \ldots)$, respectively. These differences are not conjectural, the Sato-Tate conjecture for these two curves was proved in [26].

4.3. **Sato–Tate measures.** Once we know the Sato–Tate group $\mathrm{ST}(A)$ of an abelian variety A, we are in a position to compute various statistic related to the distribution of its conjugacy classes, such as the moments of characteristic polynomial coefficients (or any other conjugacy class invariant). We can then test the Sato–Tate conjecture by comparing these to corresponding statistics obtained by computing normalized L-polynomials $\bar{L}_{\mathfrak{p}}(T)$ for all primes $\mathfrak{p}$ of good reduction for A up to some norm bound B.

The first step is to determine the Haar measure on $\mathrm{ST}(A)^0$. For $g = 1$ there are only two possibilities: either $\mathrm{ST}(A)^0 = \mathrm{U}(1)$ or $\mathrm{ST}(A)^0 = \mathrm{SU}(2)$, where, as usual we embed $\mathrm{U}(1)$ in $\mathrm{SU}(2)$ via $u \mapsto \left(\begin{smallmatrix} u & 0 \\ 0 & \bar{u} \end{smallmatrix}\right)$. In terms of the eigenangle θ, the pushforward measure on $\mathrm{conj}(\mathrm{ST}(A)^0)$ is one of

$$\mu_{\mathrm{U}(1)} := \tfrac{1}{\pi} d\theta,$$
$$\mu_{\mathrm{SU}(2)} := \tfrac{2}{\pi} \sin^2\theta \, d\theta,$$

with $0 \le \theta \le \pi$. This also addresses two of the possibilities for $\mathrm{ST}(A)^0$ that arise when $g = 2$, the groups $\mathrm{U}(1)_2$ and $\mathrm{SU}(1)_2$ listed in the first two rows of Table 1; these denote two identical copies of $\mathrm{U}(1)$ and $\mathrm{SU}(2)$ diagonally embedded in $\mathrm{USp}(4)$. When expressed in terms of the eigenangle θ, the measure $\mu_{\mathrm{U}(1)_2}$ is exactly the same as $\mu_{\mathrm{U}(1)}$ (and similarly for $\mu_{\mathrm{SU}(2)_2}$), but note that we will get a different distribution on characteristic polynomials (which now have degree 4 rather than degree 2), because each eigenvalue now occurs with multiplicity 2; in particular, the trace becomes $4\cos\theta$ rather than $2\cos\theta$.

For the groups $\mathrm{ST}(A)^0$ that appear in the next three rows of Table 1, the measure on $\mathrm{conj}(\mathrm{ST}(A)^0)$ is a product of measures that we already know:

$$\mu_{\mathrm{U}(1)\times\mathrm{U}(1)} := \tfrac{1}{\pi^2} d\theta_1 \, d\theta_2,$$
$$\mu_{\mathrm{U}(1)\times\mathrm{SU}(2)} := \tfrac{2}{\pi^2} \sin^2\theta_2 \, d\theta_1 \, d\theta_2,$$
$$\mu_{\mathrm{SU}(2)\times\mathrm{SU}(2)} := \tfrac{4}{\pi^2} \sin^2\theta_1 \sin^2\theta_2 \, d\theta_1 \, d\theta_2.$$

To obtain the measure for the generic case $\mathrm{ST}(A) = \mathrm{ST}(A)^0 = \mathrm{USp}(4)$, we use the Weyl integration formula for $\mathrm{USp}(2g)$ (which includes the case $\mathrm{SU}(2) = \mathrm{USp}(2)$ that we already know):

$$\mu_{\mathrm{USp}(2g)} := \frac{1}{g!}\left(\prod_{1\le j<k\le g}\left(2\cos\theta_j - 2\cos\theta_k\right)^2\right)\prod_{1\le j\le g}\left(\tfrac{2}{\pi}\sin^2\theta_j\, d\theta_j\right), \tag{11}$$

with $0 \le \theta_j \le \pi$, see [93, Thm. 7.8B] or [44, §5.0.4]. This covers all the $g = 2$ cases, and (by taking appropriate products) all the $g = 3$ cases listed in Table 2 except for U(3), where we need the Weyl integration formula for $\mathrm{U}(g)$:

$$\mu_{\mathrm{U}(g)} := \frac{1}{g!}\left(\prod_{1\le j<k\le g}\left|e^{i\theta_j} - e^{i\theta_k}\right|\right)\prod_{1\le j\le g}\tfrac{1}{2\pi}d\theta_j, \tag{12}$$

with $0 \le \theta_j \le 2\pi$ (note the 2π); see [93, Thm. 7.4B] or [44, §5.0.3].

With the measure $\mu_{\mathrm{ST}(A)^0}$ in hand, for any continuous class function f on $\mathrm{ST}(A)$, we can compute

$$\mu_{\mathrm{ST}(A)}(f) := \int_{\mathrm{ST}(A)} f(g)\mu_{\mathrm{ST}(A)}(g) = \sum_g \int_{\mathrm{ST}(A)^0} f(gh)\mu_{\mathrm{ST}(A)^0}(h),$$

as a finite sum over a set of left coset representatives $g\,\mathrm{ST}(A)^0$ of $\mathrm{ST}(A)/\mathrm{ST}(A)^0$; see [24, §5.1.1] for details and explicit results in the case $g = 2$.

4.4. **Trace moment sequences.** As a particular application of our work in the previous section where we determined Haar measures for various Sato–Tate groups, let us consider the problem of computing the trace moment sequence of a connected Sato–Tate group; so assume $\mathrm{ST}(A) = \mathrm{ST}(A)^0$. For each integer $n \ge 0$ we wish to compute the nth moment

$$\mathrm{E}_{\mathrm{ST}(A)}[\mathrm{tr}^n] = \int_0^\pi \cdots \int_0^\pi \left(\sum_{j=1}^g 2\cos\theta_j\right)^n \mu_{\mathrm{ST}(A)}(\theta_1,\dots,\theta_g).$$

We have already done this computation for the groups U(1) and SU(2) that arise in dimension $g = 1$. For U(1) we have

$$\mathrm{E}_{\mathrm{U}(1)}[\mathrm{tr}^n] = \frac{1}{\pi}\int_0^\pi (2\cos\theta)^n\, d\theta = b_n := \binom{n}{n/2},$$

where we adopt the convention that $\binom{n}{n/2} = 0$ when n is odd, and for SU(2) we have

$$\mathrm{E}_{\mathrm{SU}(2)}[\mathrm{tr}^n] = \frac{2}{\pi}\int_0^\pi (2\cos\theta)^n \sin^2\theta\, d\theta = c_n := \frac{2}{n+2}\binom{n}{n/2}.$$

We thus obtain the moment sequences

$$\mathrm{M}_{\mathrm{U}(1)}[\mathrm{tr}^n] = (1, 0, 2, 0, 6, 0, 20, 0, 70, 0, 252, \dots),$$
$$\mathrm{M}_{\mathrm{SU}(2)}[\mathrm{tr}^n] = (1, 0, 1, 0, 2, 0, 5, 0, 14, 0, 42, \dots).$$

For $g = 2$ we observe that for 5 of the 6 connected Sato–Tate groups listed in Table 1 we can derive their trace moment sequences directly from the trace moment sequences for U(1) and SU(2); no integration is required. For $\mathrm{U}(1)_2$ and $\mathrm{SU}(2)_2$ we simply have

$$E_{\mathrm{U}(1)_2}[\mathrm{tr}^n] = E_{\mathrm{U}(1)}[2^n\,\mathrm{tr}^n] = 2^n b_n,$$
$$E_{\mathrm{SU}(2)_2}[\mathrm{tr}^n] = E_{\mathrm{SU}(2)}[2^n\,\mathrm{tr}^n] = 2^n c_n,$$

and for $\mathrm{U}(1)\times\mathrm{U}(1)$, $\mathrm{U}(1)\times\mathrm{SU}(2)$, $\mathrm{SU}(2)\times\mathrm{SU}(2)$ we take binomial convolutions to obtain[21]

$$E_{\mathrm{U}(1)\times\mathrm{U}(1)}[\mathrm{tr}^n] = \sum_{r=0}^{n}\binom{n}{r}E_{\mathrm{U}(1)}[\mathrm{tr}^r]E_{\mathrm{U}(1)}[\mathrm{tr}^{n-r}] = \sum_{r=0}^{n}\binom{n}{r}b_r b_{n-r} = b_n^2, \tag{13}$$

$$E_{\mathrm{U}(1)\times\mathrm{SU}(2)}[\mathrm{tr}^n] = \sum_{r=0}^{n}\binom{n}{r}E_{\mathrm{U}(1)}[\mathrm{tr}^r]E_{\mathrm{SU}(2)}[\mathrm{tr}^{n-r}] = \sum_{r=0}^{n}\binom{n}{r}b_r c_{n-r} = \tfrac{1}{2}c_n b_{n+2}, \tag{14}$$

$$E_{\mathrm{SU}(2)\times\mathrm{SU}(2)}[\mathrm{tr}^n] = \sum_{r=0}^{n}\binom{n}{r}E_{\mathrm{SU}(2)}[\mathrm{tr}^r]E_{\mathrm{SU}(2)}[\mathrm{tr}^{n-r}] = \sum_{r=0}^{n}\binom{n}{r}c_r c_{n-r} = c_n c_{n+2}. \tag{15}$$

For the generic case $\mathrm{USp}(4)$ we apply (11) with $g=2$ to obtain

$$E_{\mathrm{USp}(4)}[\mathrm{tr}^n] = \tfrac{2^{n+3}}{\pi^2}\int_0^\pi\int_0^\pi(\cos\theta_1+\cos\theta_2)^n(\cos\theta_1-\cos\theta_2)^2\sin^2\theta_1\sin^2\theta_2\,d\theta_1 d\theta_2 = c_n c_{n+4} - c_{n+2}^2.$$

Here we have applied the general determinantal formula from [48, Thm. 1] that allows one to compute the moment generating function of the kth eigenvalue power-sum in $\mathrm{USp}(2g)$. Recall that the *moment generating function* of a moment sequence $(m_0, m_1, m_2, \ldots)$ is the exponential generating function

$$\mathscr{M}(z) := \sum_{n=0}^{\infty} m_n \frac{z^n}{n!}.$$

One uses exponential generating functions so that products of moment generating functions correspond to binomial convolutions of moment sequences; this means that if $\mathscr{M}_1(z)$ and $\mathscr{M}_2(z)$ are the moment generating functions of two independent random variable X_1 and X_2, then the moment generating function of X_1+X_2 is simply $\mathscr{M}_1(z)\mathscr{M}_2(z)$.

The determinantal formula for the first eigenvalue power-sum (the trace) is simply

$$\mathscr{M}_{\mathrm{USp}(2g)}[\mathrm{tr}] = \det_{g\times g}\left(\mathscr{C}^{i+j-2}\right)_{ij},$$

where $\mathscr{C}^m$ is the moment generating function defined by

$$\mathscr{C}^m(z) := \sum_{r=0}^{m}\binom{n}{r}\left(\mathscr{B}_{2r-n} - \mathscr{B}_{2r-n+2}\right), \qquad \mathscr{B}_s(z) := \sum_{n=0}^{\infty}\frac{z^{2n+s}}{s!(n+s)!}.$$

The function $\mathscr{B}_s(z)$ is related to a hyperbolic Bessel function of the first kind; see [48, p. 13] for details.

For the connected Sato–Tate groups that arise in dimension $g=2$ we thus obtain the following moment sequences:

$$\begin{aligned}
\mathrm{M}_{\mathrm{U}(1)_2}[\mathrm{tr}] &= (1,\,0,\,8,\,0,\,96,\,0,\,1280,\,0,\,17920,\,0,\,258048,\,\ldots),\\
\mathrm{M}_{\mathrm{SU}(2)_2}[\mathrm{tr}] &= (1,\,0,\,4,\,0,\,32,\,0,\,320,\,0,\,3584,\,0,\,43008,\,\ldots),\\
\mathrm{M}_{\mathrm{U}(1)\times\mathrm{U}(1)}[\mathrm{tr}] &= (1,\,0,\,4,\,0,\,36,\,0,\,400,\,0,\,4900,\,0,\,63504,\,\ldots),\\
\mathrm{M}_{\mathrm{U}(1)\times\mathrm{SU}(2)}[\mathrm{tr}] &= (1,\,0,\,3,\,0,\,20,\,0,\,175,\,0,\,1764,\,0,\,19404,\,\ldots),\\
\mathrm{M}_{\mathrm{SU}(2)\times\mathrm{SU}(2)}[\mathrm{tr}] &= (1,\,0,\,2,\,0,\,10,\,0,\,70,\,0,\,588,\,0,\,5544,\,\ldots),\\
\mathrm{M}_{\mathrm{USp}(4)}[\mathrm{tr}] &= (1,\,0,\,1,\,0,\,3,\,0,\,14,\,0,\,84,\,0,\,594,\,\ldots).
\end{aligned}$$

Recall that for $g=1$ the trace moment sequence $(1,0,1,0,2,0,5,0,14,0,42,\ldots)$ of the generic Sato–Tate group $\mathrm{SU}(2)$ corresponds to the sequence of Catalan numbers with 0's inserted at the odd moments. There is a standard combinatorial interpretation of this sequence: the nth moment counts

[21] It is at this point we see the utility of including zeroth moments in our moment sequences.

the number of returning walks of length n on a 1-dimensional integer lattice that stay to the right of the origin (there are no such walks when n is odd, hence the odd moments are zero).

This combinatorial interpretation generalizes to higher genus. For $g = 2$ the trace moment sequence for the generic Sato–Tate group USp(4) counts returning walks on a 2-dimensional integer lattice that satisfy $x_1 \geq x_2 \geq 0$ (so now there are 3 walks of length 4, not just 2). In general, for any $g \geq 1$ the trace moment sequence for the generic Sato–Tate group $\mathrm{USp}(2g)$ counts returning walks on a g-dimensional integer lattice that satisfy $x_1 \geq \ldots \geq x_g \geq 0$; this follows from a general result of Grabiner and Magyar [29] that relates the decomposition of tensor powers of certain representations of classical Lie groups to lattice paths that are constrained to lie in the closure of the fundamental *Weyl chamber* of the corresponding Lie algebra (which can be defined as an intersection of hyperplanes orthogonal to elements of a basis for the root system).

This has combinatorial fact has an interesting asymptotic consequence. For any integers $g' \geq g > 0$, the moment sequences $\mathrm{M}_{\mathrm{USp}(2g')}[\mathrm{tr}]$ and $\mathrm{M}_{\mathrm{USp}(2g)}[\mathrm{tr}]$ must agree up to the $2g$th moment; see Exercise 4.3. Thus the moments sequences $\mathrm{M}_{\mathrm{USp}(2g)}[\mathrm{tr}]$ converge to a limiting sequence as $g \to \infty$:

$$\begin{aligned}
\mathrm{M}_{\mathrm{USp}(2)}[\mathrm{tr}] &= (1, 0, 1, 0, 2, 0, 5, 0, 14, 0, 42, \ldots),\\
\mathrm{M}_{\mathrm{USp}(4)}[\mathrm{tr}] &= (1, 0, 1, 0, 3, 0, 14, 0, 84, 0, 594, \ldots),\\
\mathrm{M}_{\mathrm{USp}(6)}[\mathrm{tr}] &= (1, 0, 1, 0, 3, 0, 15, 0, 104, 0, 909, \ldots),\\
\mathrm{M}_{\mathrm{USp}(8)}[\mathrm{tr}] &= (1, 0, 1, 0, 3, 0, 15, 0, 105, 0, 944 \ldots).\\
&\vdots\\
\mathrm{M}_{\mathrm{USp}(\infty)}[\mathrm{tr}] &= (1, 0, 1, 0, 3, 0, 15, 0, 105, 0, 945, \ldots).
\end{aligned}$$

The limiting sequence $\mathrm{M}_{\mathrm{USp}(\infty)}[\mathrm{tr}]$ is precisely the moment sequence of the standard normal distribution (mean 0 and variance 1); the nth moment is zero if n is odd, for even n it is simply

$$(n-1)!! := n(n-2)(n-4)\cdots 3 \cdot 1.$$

On the next page we show the a_1-distributions for $g = 1, 2, 3, 4$, normalized to the same scale, which illustrates convergence to the Gaussian.

4.5. **Exercises.**

Exercise 4.1. Give combinatorial proofs of the identities used in (13), (14), (15).

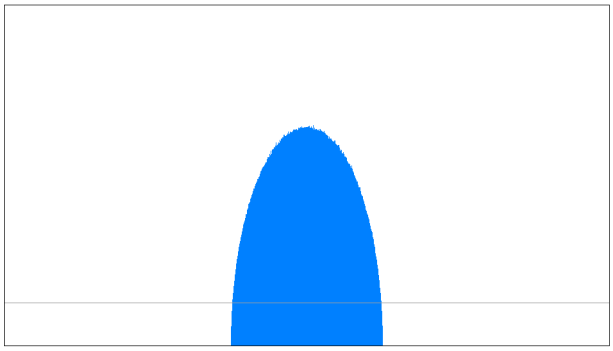
Exercise 4.2. Using the combinatorial interpretation of the generic trace moment sequence $\mathrm{M}_{\mathrm{USp}(2g)}[\mathrm{tr}]$, prove that for $g' > g$ the moment sequences $\mathrm{M}_{\mathrm{USp}(2g')}[\mathrm{tr}]$ and $\mathrm{M}_{\mathrm{USp}(2g)}[\mathrm{tr}]$ agree up to the $2g$th moment but disagree at the $(2g+2)$th moment.

Exercise 4.3. Characterize each of the 6 trace moment sequences that arise for connected Sato–Tate groups in dimension $g = 2$ by showing that each sequence counts returning walks on an 2-dimensional integer lattice that are constrained to a certain region of the plane.

Exercise 4.4. Similarly characterize the 15 trace moment sequences that arise for connected Sato–Tate groups in dimension $g = 3$ in terms of returning walks on a 3-dimensional integer lattice.

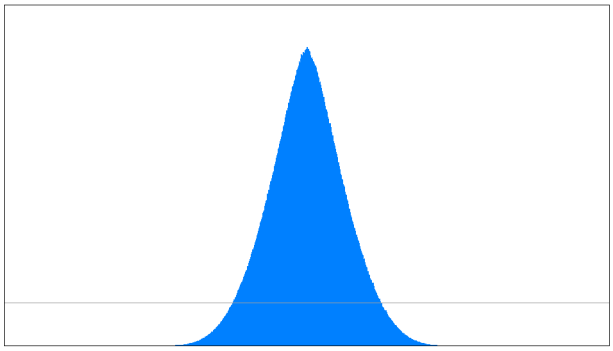
Exercise 4.5. For each of the 5 non-generic connected Sato–Tate groups that arise in dimension $g = 2$ compute the moment sequence for a_2, the quadratic coefficient of the characteristic polynomial.

a1 histogram of y^2 = x^3-x+1 for p <= 2^28
14630841 data points in 3825 buckets



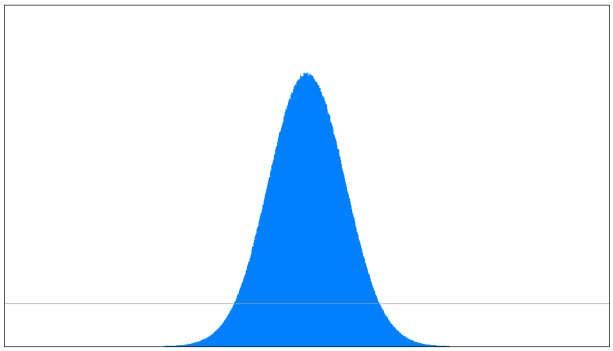
Moments: 1 0.000 1.000 0.000 2.000 0.001 5.001 0.003 14.005 0.011 42.018

a1 histogram of y^2 = x^5 - x + 1 for p <= 2^28
14630838 data points in 3825 buckets



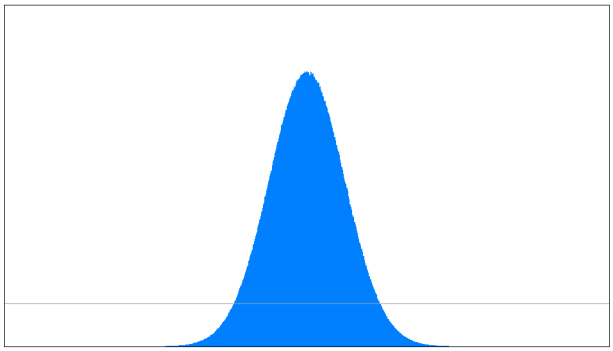
Moments: 1 0.000 1.000 0.000 2.997 0.004 13.979 0.042 83.833 0.458 592.709

a1 histogram of y^2 = x^7 - x + 1 for p <= 2^28
14630837 data points in 3825 buckets



Moments: 1 0.000 1.000 -0.001 3.000 -0.014 15.004 -0.193 103.941 -2.832 905.800

a1 histogram of y^2 = x^9 - x + 1 for p <= 2^28
14630832 data points in 3825 buckets



Moments: 1 0.000 1.000 0.001 3.000 0.006 15.022 0.023 105.431 -0.446 952.758

References

[1] J. Achter and J. Holden, *Notes on an analogue of the Fontaine-Mazur conjecture*, Journal de Théorie des Nombres de Bordeaux **15** (2003), 627–637.

[2] O. Ahmadi and I. E. Shparlinski, *On the distribution of the number of points on algebraic curves in extensions of finite fields*, Mathematical Research Letters **17** (2012), 689–699.

[3] G. Banaszak and K. S. Kedlaya, *An algebraic Sato-Tate group and Sato-Tate conjecture*, Indiana University Mathematics Journal **64** (2015), 245–274.

[4] G. Banaszak and K. S. Kedlaya, *Motivic Serre group, algebraic Sato-Tate group and Sato-Tate conjecture*, in *Frobenius Distributions: Lang–Trotter and Sato–Tate Conjectures*, Contemporary Mathematics **663** (2016), AMS, 11–44.

[5] T. Barnet-Lamb, D. Geraghty, and T. Gee, *The Sato-Tate conjecture for Hilbert modular forms*, Journal of the American Mathematical Society **24** (2011), 411–469.

[6] T. Barnet-Lamb, D. Geraghty, M. Harris, and R. Taylor, *A family of Calabi-Yau varieties and potential automorphy II*, Publications of the Research Institute for Mathematical Sciences **47** (2011), 29–98.

[7] C. Birkenhake and H. Lange, *Complex abelian varieties*, Springer, 2004.

[8] F.A. Bogomolov, *Sur l'algébricité des représentations l-adiques*, Comptes Rendus Acad. Sci. Paris **290** (1980), 701–703.

[9] A. Borel, *Linear algebraic groups*, second edition, 1991.

[10] Wieb Bosma, John Cannon, and Catherine Playoust, *The Magma algebra system I: The user language*, Journal of Symbolic Computation **24** (1997), 235–265.

[11] C. Breuil, B. Conrad, F. Diamond, and R. Taylor, *On the modularity of elliptic curves over* $\mathbf{Q}$*: wild 3-adic exercises*, Journal of the American Mathematical Society **14** (2001), 843–939.

[12] A. Bucur and K. S. Kedlaya, *An application of the effective Sato-Tate conjecture*, in *Frobenius Distributions: Lang–Trotter and Sato–Tate Conjectures*, Contemporary Mathematics **663** (2016), AMS, 45–56.

[13] J. W. S. Cassels and A. Fröhlich, *Algebraic number theory*, second edition, London Mathematical Society, 2010.

[14] W. Castryck, A. Folsom, H. Hubrechts, and A. V. Sutherland, *The probability that the number of points on the Jacobian of a genus 2 curve is prime*, Proceedings of the London Mathematical Society **104** (2012), 1235–1270.

[15] L. Clozel, M. Harris, and R. Taylor, *Automorphy for some ℓ-adic lifts of automorphic mod-ℓ Galois representations*, Publ. Math. IHES **108** (2008), 1–181.

[16] A. C. Cojocaru, R. Davis, A. Silverberg, and K.Ė. Stange, *Arithmetic properties of the Frobenius traces defined by a rational abelian variety*, with two appendices by J-P. Serre, arXiv:1504.00902, 2015.

[17] P. Deligne, *La conjecture de Weil: I*, Publ. Math. IHES **43** (1974), 273–307.

[18] P. Deligne, *La conjecture de Weil: II*, Publ. Math. IHES **52** (1980), 173–252.

[19] P. Deligne, *Hodge cycles on abelian varieties (notes by J.S. Milne)*, Lecture Notes in Mathematics **900** (1982), 9–100.

[20] F. Diamond and J. Shurman, *A first course in modular forms*, Springer, 2005.

[21] J. Diestel and A. Spalsbury, *The joys of Haar measure*, Graduate Studies in Mathematics **150**, AMS, 2014.

[22] G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern*, Inventiones Mathematicae **73** (1983), 349–366.

[23] F. Fité, *Equidistribution, L-functions, and Sato–Tate groups*, Contemporary Mathematics **649** (2015), 63–88.

[24] F. Fité, K. S. Kedlaya, V. Rotger, and A. V. Sutherland, *Sato-Tate distributions and Galois endomorphism modules in genus 2*, Compositio Mathematica **148** (2012), 1390–1442.

[25] F. Fité, K. S. Kedlaya, and A. V. Sutherland, *Sato-Tate groups of some weight 3 motives*, in *Frobenius Distributions: Lang–Trotter and Sato–Tate Conjectures*, Contemporary Mathematics **663**, AMS, 57–102.

[26] F. Fité and A. V. Sutherland, *Sato-Tate distributions of twists of $y^2 = x^5 - x$ and $y^2 = x^6 + 1$*, Algebra and Number Theory **8** (2014), 543–585.

[27] F. Fité and A. V. Sutherland, *Sato-Tate groups of $y^2 = x^8 + c$ and $y^2 = x^7 - cx$*, in *Frobenius Distributions: Lang–Trotter and Sato–Tate Conjectures*, Contemporary Mathematics **663**, AMS, 103–126.

[28] Joachim von zur Gathen and Jürgen Gerhard, *Modern computer algebra*, third edition, Cambridge University Press, 2013.

[29] D. J. Grabiner and P. Magyar, *Random walks in Weyl chambers and the decomposition of tensor powers*, Journal of Algebraic Combinatorics **2** (1993), 239–260.

[30] C. Hall, *An open–image theorem for a general class of abelian varieties, with an appendix by E. Kowalski*, Bulletin of the London Mathematical Society **43** (2011), 703–711.

[31] M. Harris, N. Shepherd-Barron, and R. Taylor, *A family of Calabi-Yau varieties and potential automorphy*, Annals of Mathematics **171** (2010), 779–813.

[32] W. B. Hart, *Fast Library for Number Theory: An introduction*, in *Proceedings of the Third International Congress on Mathematical Software (ICMS 2010)*, LNCS 6327, Springer, 2010, 88–91.
[33] W. B. Hart, F. Johansson and S. Pancratz, *Fast Library for Number Theory*, version 2.5.2, `http://flintlib.org`, 2015.
[34] D. Harvey, *Kedlaya's algorithm in larger characteristic*, International Mathematics Research Notices **2007**.
[35] D. Harvey, *Counting points on hyperelliptic curves in average polynomial time*, Annals of Mathematics **179** (2014), 783–803.
[36] D. Harvey, *Computing zeta functions of arithmetic schemes*, Proceedings of the London Mathematical Society **111** (2015), 1379–1401.
[37] D. Harvey and A. V. Sutherland, *Computing Hasse-Witt matrices of hyperelliptic curves in average polynomial time*, in *Algorithmic Number Theory 11th International Symposium (ANTS XI)*, LMS Journal of Computation and Mathematics **17** (2014), 257–273.
[38] D. Harvey and A. V. Sutherland, *Computing Hasse-Witt matrices of hyperelliptic curves in average polynomial time II*, in *Frobenius Distributions: Lang–Trotter and Sato–Tate Conjectures*, Contemporary Mathematics **663**, AMS, 127–148.
[39] E. Hecke, *Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen. Zweite Mitteilung*, Mathematische Zeitschrift **6** (1920), 11–51.
[40] H. Heyer, *Probability measures on locally compact groups*, Springer, 1977.
[41] J. E. Humphreys, *Linear algebraic groups*, Springer, 1975.
[42] C. Johansson, *On the Sato-Tate conjecture for non-generic abelian surfaces*, with an appendix by Francesc Fité, Transactions of the AMS, to appear.
[43] I. Kaplansky, *Lattices of continuous functions*, Bulletin of the AMS **6** (1947), 617–623.
[44] N. M. Katz and P. Sarnak, *Random matrices, Frobenius eigenvalues, and monodromy*, Colloquium Publications **45**, AMS, 1999.
[45] K. S. Kedlaya, *Counting points on hyperelliptic curves using Monsky-Washnitzer cohomology*, Journal of the Ramanujan Mathematical Society **16** (2001), 323–338.
[46] K. S. Kedlaya, *Computing zeta functions via p-adic cohomology*, in *Algorithmic Number Theory 6th International Symposium (ANTS VI)*, Lecture Notes in Computer Science **3076**, Springer 2004, 1–17.
[47] K. S. Kedlaya and A. V. Sutherland, *Computing L-series of hyperelliptic curves*, in *Algorithmic Number Theory 8th International Symposium (ANTS VIII)*, Lecture Notes in Computer Science **5011**, Springer, 2008, 312–326.
[48] K. S. Kedlaya and A. V. Sutherland, *Hyperelliptic curves, L-polynomials, and random matrices*, in *Arithmetic Geometry, Cryptography, and Coding Theory (AGCCT-11)*, Contemporary Mathematics **487**, American Mathematical Society, 2000, 119–162.
[49] J. Klüners and G. Malle, *A database for field extensions of the rationals*, LMS Journal of Computation and Mathematics **4** (2001), 182–196.
[50] P. Koosis, *The logarithmic integral I*, Cambridge University Press, 1998.
[51] S. Lang and H. Trotter, *Frobenius distributions in GL_2-extensions*, Lecture Notes in Mathematics **504** (1976), Springer.
[52] J. S. Milne, *Abelian varieties*, v2.00, 2008.
[53] J. S. Milne, *Algebraic groups: An introduction to the theory of algebraic group schemes over fields*, v2.00, 2015.
[54] B. Moonen, *An introduction to Mumford–Tate groups*, Monte Verita lecture notes, available at `http://www.math.ru.nl/~bmoonen/Lecturenotes/MTGps.pdf`, 2004.
[55] B. Moonen and Yu.G. Zarhin, *Hodge classes on abelian varieties of low dimension*, Mathematische Annalen **315** (1999), 711–733.
[56] D. Mumford, *A note on Shimura's paper "Discontinuous subgroups and abelian varieties"*, Mathematische Annalen **181** (1969), 345–351.
[57] D. Mumford, *Abelian varieties*, second edition, Tata Institute of Fundamental Research Studies in Mathematics, 1974.
[58] M. R. Murty and V. K. Murty, *Non-vanishing of L-functions and applications*, Modern Birkhäuser Classics, 1997.
[59] V. K. Murty, *Explicit formulae and the Lang-Trotter conjecture*, Rocky Mountain Journal of Mathematics **15** (1985), 535–551.
[60] J. Neukirch, *Algebraic number theory*, Springer, 1999.
[61] A. L. Onishchik and A. L. Vinberg (eds.), *Lie groups and Lie algebras III: Structure of Lie groups and Lie algebras*, Springer, 1994.
[62] OEIS Foundation Inc., *The On-Line Encyclopedia of Integer Sequences*, online database at `http://oeis.org`, 2016.

[63] J. Pila, *Frobenius maps of abelian varieties and finding roots of unity in finite fields*, Mathematics of Computation **55** (1990), 745–763.
[64] D. Ramakrishnan, *Remarks on the Tate Conjecture for beginners*, notes from the AIM Tate Conjecture Workshop, available at `http://www.aimath.org/WWN/tateconjecture/tateconjecture.pdf`, 2007.
[65] The Sage Developers, *Sage Mathematics Software*, Version 7.0, available at `http://www.sagemath.org`, 2016.
[66] R. Schoof, *Elliptic curves over finite fields and the computation of square roots mod p*, Mathematics of Computation **44** (1995), 483–494.
[67] R. Schoof, *Counting points on elliptic curves over finite fields*, Journal de Théorie des Nombres de Bordeaux **7** (1995), 219–254.
[68] J.-P. Serre, *Abelian ℓ-adic representations and elliptic curves*, Research Notes in Mathematics **7**, A.K. Peters, 1998.
[69] J.-P. Serre, *Propriétés galoisiennes des points d'ordre fini des courbes elliptiques*, Inventiones Mathematicae **15** (1972), 259–331.
[70] J.-P. Serre, *Résumé des cours de 1985-1986, Annuaire du Collège de France, 1986, 95–99*; in *Oeuvres – Collected Papers, Volume IV*, Springer, 2003, 33–37.
[71] J.-P. Serre, *Lettre à Marie-France Vignéras du 10/2/1986*, in *Oeuvres – Collected Papers, Volume IV*, Springer, 2003, 38–55.
[72] J.-P. Serre, *Lettres à Ken Ribet du 1/1/1981 et du 29/1/1981*, in *Oeuvres – Collected Papers, Volume IV*, Springer, 2003, 1–20.
[73] J.-P. Serre, *Propriétés conjecturales des groups de Galois motiviques et des représentations ℓ-adiques*, in *Motives*, AMS Proceedings of Symposia in Pure Mathematics **55** (1994), 377–400.
[74] J.-P. Serre, *Lectures on $N_X(p)$*, Research Notes in Mathematics **11**, CRC Press, 2012.
[75] Y.-D. Shieh, *Arithmetic aspects of point counting and Frobenius distributions*, Ph.D. thesis, Université d'Aix-Marseille, 2015.
[76] V. Shoup, *NTL: A Library for doing Number Theory*, version 9.6.4, available at `http://www.shoup.net/ntl/`, 2016.
[77] J. H. Silverman, *Advanced topics in the arithmetic of elliptic curves*, Springer, 1994.
[78] J. H. Silverman, *The arithmetic of elliptic curves*, second edition, Springer, 2009.
[79] T. A. Springer, *Linear algebraic groups*, second edition, Modern Birkhäuser Classics, 1998.
[80] R. Stanley, *Catalan numbers*, Cambridge University Press, 2015.
[81] A. V. Sutherland, `smalljac`, version 5.0, available at `http://math.mit.edu/~drew`, 2016.
[82] A. V. Sutherland, *Order computations in generic groups*, PhD thesis, Massachusetts Institute of Technology, 2007.
[83] A. V. Sutherland, *Structure computation and discrete logarithms in finite abelian p-groups*, Mathematics of Computation **80** (2011), 477–500.
[84] J. Tate, *Algebraic cycles and poles of zeta functions*, in *Arithmetical Algebraic Geometry (Proc. Conf. Purdue Univ., 1963)*, Harper & Row, New York, 1965.
[85] J. Tate, *Endomorphisms of abelian varieties over finite fields*, Inventiones Mathematicae **2** (1966), 134–144.
[86] R. Taylor, *Automorphy for some ℓ-adic lifts of automorphic mod ℓ Galois representations II*, Publ. Math. IHES **108** (2008) 183–239.
[87] R. Taylor and A. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Annals of Mathematics **141** (1995), 553–572.
[88] J. Thorner, *The error term in the Sato-Tate conjecture*, Archiv der Mathematik **103** (2014), 147–156.
[89] P. van Wamelen, *On the CM character of the curves $y^2 = x^q - 1$*, Journal of Number Theory **64** (1997), 59–83.
[90] A. Weil, *Sur les courbes algébriques et les variétés qui s'en déduisent*, Publ. Inst. Math. Univ. Strasbourg **7** (1945).
[91] A. Weil, *Variétés abéliennes et courbes algébriques*, Publ. Inst. Math. Univ. Strasbourg **8** (1946).
[92] A. Weil, *Numbers of solutions of equations in finite fields*, Bulletin of the AMS **55** (1949), 497–508.
[93] H. Weyl, *The classical groups: their invariants and representations*, Princeton University Press, 1966.
[94] A. Wiles, *Modular elliptic curves and Fermat's last theorem*, Annals of Mathematics **141** (1995), 443–551.
[95] A. Wulfsohn, *A note on the vague topology for measures*, Mathematical Proceedings of the Cambridge. Philosophical Society **58** (1962), 421–422.
[96] Yu. G. Zarhin, *Hyperelliptic Jacobians without complex multiplication*, Mathematical Research Letters **7** (2000), 123–132.