# The Sato-Tate conjecture for elliptic and hyperelliptic curves

Kiran S. Kedlaya

Department of Mathematics, Massachusetts Institute of Technology; kedlaya@mit.edu
Department of Mathematics, University of California, San Diego; kedlaya@ucsd.edu

New Zealand Mathematical Society colloquium
University of Auckland, December 6, 2011

Joint work with Francesc Fité, Víctor Rotger, Andrew V. Sutherland; see arXiv:1110.6638. See also Serre, *Lectures on* $N_X(p)$ (AK Peters, 2011).
For slides, see http://math.mit.edu/~kedlaya/papers/talks.shtml.

Supported by NSF, DARPA, MIT, UCSD.

# Contents

# Contents

# Counting roots of polynomials

Let $f(x) \in \mathbb{Z}[x]$ be a squarefree polynomial of degree $d > 0$ whose coefficients have no common divisor greater than 1. For each prime number $p$, define

$$N_f(p) = \#\{x \in \{0, \dots, p-1\} : f(x) \equiv 0 \pmod{p}\}.$$

This is an integer in the range $\{0, \dots, d\}$. But how often does each value occur?

# Example: quadratic polynomials

Take $f(x) = ax^2 + bx + c$. Let $\Delta = b^2 - 4ac$ be the discriminant. Then

$$N_f(p) = \begin{cases} 0 & \text{if } \Delta \text{ is not a quadratic residue modulo } p \\ 1 & \text{if } \Delta \text{ is congruent to 0 modulo } p \\ 2 & \text{if } \Delta \text{ is a nonzero quadratic residue modulo } p. \end{cases}$$

For a "randomly chosen" prime number $p$, $N_f(p)$ takes the values 0 and 2 each with probability $\frac{1}{2}$ (unless $\Delta$ is a square).

In this case, one can even give an explicit formula for $N_f(p)$ using the law of quadratic reciprocity. For example, in case $\Delta = 5$, one has (for $p > 2$)

$$N_f(p) = \begin{cases} 0 & \text{if } p \equiv 2, 3 \pmod 5 \\ 1 & \text{if } p = 5 \\ 2 & \text{if } p \equiv 1, 4 \pmod 5. \end{cases}$$

This state of affairs is not typical!

# Example: quadratic polynomials

Take $f(x) = ax^2 + bx + c$. Let $\Delta = b^2 - 4ac$ be the discriminant. Then

$$N_f(p) = \begin{cases} 0 & \text{if } \Delta \text{ is not a quadratic residue modulo } p \\ 1 & \text{if } \Delta \text{ is congruent to 0 modulo } p \\ 2 & \text{if } \Delta \text{ is a nonzero quadratic residue modulo } p. \end{cases}$$

For a "randomly chosen" prime number $p$, $N_f(p)$ takes the values 0 and 2 each with probability $\frac{1}{2}$ (unless $\Delta$ is a square).

In this case, one can even give an explicit formula for $N_f(p)$ using the law of quadratic reciprocity. For example, in case $\Delta = 5$, one has (for $p > 2$)

$$N_f(p) = \begin{cases} 0 & \text{if } p \equiv 2, 3 \pmod 5 \\ 1 & \text{if } p = 5 \\ 2 & \text{if } p \equiv 1, 4 \pmod 5. \end{cases}$$

This state of affairs is not typical!

# The Chebotarëv density theorem

For general $f$, one cannot to find explicit formulas for $N_f(p)$, but one can still determine their average distribution as follows.

Let $\alpha_1, \dots, \alpha_d$ be the roots of $f$ in an algebraic closure of $\mathbb{Q}$. Let $G$ be the Galois group of the number field generated by these roots, acting on $\alpha_1, \dots, \alpha_d$ by permutations. Let $c_i$ denote the probability that a random element of $G$ has exactly $i$ fixed points.

**Theorem (Chebotarëv, early 1920s)**

*For $i = 0, \dots, d$, we have*

$$\lim_{N \to \infty} \frac{\#\{p \text{ prime}, p \leq N, N_f(p) = i\}}{\#\{p \text{ prime}, p \leq N\}} = c_i.$$

In other words, $N_f(p) = i$ with probability $c_i$.

# The Chebotarëv density theorem

For general $f$, one cannot to find explicit formulas for $N_f(p)$, but one can still determine their average distribution as follows.

Let $\alpha_1, \dots, \alpha_d$ be the roots of $f$ in an algebraic closure of $\mathbb{Q}$. Let $G$ be the Galois group of the number field generated by these roots, acting on $\alpha_1, \dots, \alpha_d$ by permutations. Let $c_i$ denote the probability that a random element of $G$ has exactly $i$ fixed points.

**Theorem (Chebotarëv, early 1920s)**

*For $i = 0, \dots, d$, we have*

$$\lim_{N \to \infty} \frac{\#\{p \text{ prime}, p \leq N, N_f(p) = i\}}{\#\{p \text{ prime}, p \leq N\}} = c_i.$$

In other words, $N_f(p) = i$ with probability $c_i$.

# The Chebotarëv density theorem

For general $f$, one cannot to find explicit formulas for $N_f(p)$, but one can still determine their average distribution as follows.

Let $\alpha_1, \dots, \alpha_d$ be the roots of $f$ in an algebraic closure of $\mathbb{Q}$. Let $G$ be the Galois group of the number field generated by these roots, acting on $\alpha_1, \dots, \alpha_d$ by permutations. Let $c_i$ denote the probability that a random element of $G$ has exactly $i$ fixed points.

**Theorem (Chebotarëv, early 1920s)**

*For $i = 0, \dots, d$, we have*

$$\lim_{N \to \infty} \frac{\#\{p \text{ prime}, p \leq N, N_f(p) = i\}}{\#\{p \text{ prime}, p \leq N\}} = c_i.$$

In other words, $N_f(p) = i$ with probability $c_i$.

# Prime powers

For $q$ a prime power, we may also define

$$N_f(q) = \#\{x \in \mathbb{F}_q : f(x) = 0\};$$

this agrees with the previous definition when $q$ is prime.

**Theorem**

*For any $c_1, c_2, \cdots \in \{0, \ldots, d\}$, the probability that $N_f(p) = c_1, N_f(p^2) = c_2, \ldots$ equals the probability that a random element $g$ of $G$ has the property that $g$ has $c_1$ fixed points, $g^2$ has $c_2$ fixed points, and so on.*

Again, the probability is defined by counting the proportion of good primes in the range $p \leq N$ and taking the limit as $N \to \infty$.

# Contents

# Counting solutions of polynomial equations

For $f_1, \ldots, f_m \in \mathbb{Z}[x_1, \ldots, x_n]$, we consider the function on prime numbers:

$$p \mapsto N_X(p) := \#\{\mathbf{x} = (x_1, \ldots, x_n) \in \{0, \ldots, p-1\}^n : f_1(\mathbf{x}) \equiv \cdots \equiv f_m(\mathbf{x}) \equiv 0 \pmod{p}\}.$$

If we rewrite this as

$$N_X(p) = \#\{\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_p^n : f_1(\mathbf{x}) = \cdots = f_m(\mathbf{x}) = 0\},$$

we may use the same formula to define $N_X(q)$ when $q$ is a prime power.

In the notation, $X$ denotes the *affine algebraic variety* (or better, the *affine scheme*) defined by $f_1, \ldots, f_m$. One can also define $N_X(q)$ when $X$ is a more general algebraic variety (or better, a scheme of finite type over $\mathbb{Z}$). For example, for $X = \mathbb{P}^n$, the projective space of dimension $n$,

$$N_{\mathbb{P}^n}(q) = \frac{q^{n+1} - 1}{q - 1} = 1 + q + \cdots + q^n.$$

# Counting solutions of polynomial equations

For $f_1, \ldots, f_m \in \mathbb{Z}[x_1, \ldots, x_n]$, we consider the function on prime numbers:

$$p \mapsto N_X(p) := \#\{\mathbf{x} = (x_1, \ldots, x_n) \in \{0, \ldots, p-1\}^n : f_1(\mathbf{x}) \equiv \cdots \equiv f_m(\mathbf{x}) \equiv 0 \pmod{p}\}.$$

If we rewrite this as

$$N_X(p) = \#\{\mathbf{x} = (x_1, \ldots, x_n) \in \mathbb{F}_p^n : f_1(\mathbf{x}) = \cdots = f_m(\mathbf{x}) = 0\},$$

we may use the same formula to define $N_X(q)$ when $q$ is a prime power.

In the notation, $X$ denotes the *affine algebraic variety* (or better, the *affine scheme*) defined by $f_1, \ldots, f_m$. One can also define $N_X(q)$ when $X$ is a more general algebraic variety (or better, a scheme of finite type over $\mathbb{Z}$). For example, for $X = \mathbb{P}^n$, the projective space of dimension $n$,

$$N_{\mathbb{P}^n}(q) = \frac{q^{n+1} - 1}{q - 1} = 1 + q + \cdots + q^n.$$

# First-order behavior

In most cases, one cannot hope to write $N_X(p)$ as an explicit function of $p$. Instead, we will look for statistical properties of $N_X(p)$. The first-order behavior is explained by the following result.

**Theorem (Lang-Weil, middle 1950s)**

*Let $d$ be the relative dimension of $X$ over $\mathbb{Z}$. (If $X$ is defined by an irredundant system of equations, then $d = n - m$.) Then there exist a polynomial $g \in \mathbb{Z}[x]$ and a constant $c > 0$ such that for all prime powers $q$,*

$$|N_X(q) - N_g(q)q^d| \leq cq^{d-1/2}.$$

From now on, we mostly consider cases in which $X$ is *geometrically irreducible* (i.e., $X$ has only only irreducible component when viewed as an algebraic variety over $\mathbb{C}$). For such $X$, we may take $g = x$, so $N_g(q) = 1$.

# Generalization to number fields

Let $K$ be a *number field*, i.e., a finite extension of the field $\mathbb{Q}$. Let $\mathfrak{o}_K$ be the ring of algebraic integers in $K$. For instance, one could take $K = \mathbb{Q}(i)$, the field of Gaussian numbers, in which case $\mathfrak{o}_K = \mathbb{Z}[i]$.

For $X$ a scheme of finite type over $\mathfrak{o}_K$, we may define $N_X(\mathfrak{p}^e)$ for each maximal ideal $\mathfrak{p}$ of $\mathfrak{o}_K$ and each positive integer $e$ (as the number of points of $X$ defined over the degree $e$ field extension of $\mathfrak{o}_K/\mathfrak{p}$), and ask similar questions. This will be important later: some phenomena will not be visible if we only consider $K = \mathbb{Q}$.

Note: the maximal ideals $\mathfrak{p}$ of absolute degree 1 have density 1 among all maximal ideals. That is, for averaging purposes, we need only consider those $\mathfrak{p}$ for which the residue field is a prime field $\mathbb{F}_p$. We'll then write $N_X(p^e)$ instead of $N_X(\mathfrak{p}^e)$.

# Generalization to number fields

Let $K$ be a *number field*, i.e., a finite extension of the field $\mathbb{Q}$. Let $\mathfrak{o}_K$ be the ring of algebraic integers in $K$. For instance, one could take $K = \mathbb{Q}(i)$, the field of Gaussian numbers, in which case $\mathfrak{o}_K = \mathbb{Z}[i]$.

For $X$ a scheme of finite type over $\mathfrak{o}_K$, we may define $N_X(\mathfrak{p}^e)$ for each maximal ideal $\mathfrak{p}$ of $\mathfrak{o}_K$ and each positive integer $e$ (as the number of points of $X$ defined over the degree $e$ field extension of $\mathfrak{o}_K/\mathfrak{p}$), and ask similar questions. This will be important later: some phenomena will not be visible if we only consider $K = \mathbb{Q}$.

Note: the maximal ideals $\mathfrak{p}$ of absolute degree 1 have density 1 among all maximal ideals. That is, for averaging purposes, we need only consider those $\mathfrak{p}$ for which the residue field is a prime field $\mathbb{F}_p$. We'll then write $N_X(p^e)$ instead of $N_X(\mathfrak{p}^e)$.

# Generalization to number fields

Let $K$ be a *number field*, i.e., a finite extension of the field $\mathbb{Q}$. Let $\mathfrak{o}_K$ be the ring of algebraic integers in $K$. For instance, one could take $K = \mathbb{Q}(i)$, the field of Gaussian numbers, in which case $\mathfrak{o}_K = \mathbb{Z}[i]$.

For $X$ a scheme of finite type over $\mathfrak{o}_K$, we may define $N_X(\mathfrak{p}^e)$ for each maximal ideal $\mathfrak{p}$ of $\mathfrak{o}_K$ and each positive integer $e$ (as the number of points of $X$ defined over the degree $e$ field extension of $\mathfrak{o}_K/\mathfrak{p}$), and ask similar questions. This will be important later: some phenomena will not be visible if we only consider $K = \mathbb{Q}$.

Note: the maximal ideals $\mathfrak{p}$ of absolute degree 1 have density 1 among all maximal ideals. That is, for averaging purposes, we need only consider those $\mathfrak{p}$ for which the residue field is a prime field $\mathbb{F}_p$. We'll then write $N_X(p^e)$ instead of $N_X(\mathfrak{p}^e)$.

# Contents

# Elliptic curves

An *elliptic curve* over a field $K$ is a smooth proper algebraic curve over $K$ of genus 1. If $\mathrm{char}(K) \neq 2$, these are the projective algebraic curves defined by equations of the form

$$y^2 = P(x)$$

where $P \in K[x]$ has degree 3 and has no repeated factors.

There is a natural *group structure* on the set of $K$-rational points on an elliptic curve over $K$, under which any three collinear points sum to 0. This makes elliptic curves (particularly over finite fields) useful not only in number theory, but also in cryptography!

# Elliptic curves

An *elliptic curve* over a field $K$ is a smooth proper algebraic curve over $K$ of genus 1. If $\mathrm{char}(K) \neq 2$, these are the projective algebraic curves defined by equations of the form

$$y^2 = P(x)$$

where $P \in K[x]$ has degree 3 and has no repeated factors.

There is a natural *group structure* on the set of $K$-rational points on an elliptic curve over $K$, under which any three collinear points sum to 0. This makes elliptic curves (particularly over finite fields) useful not only in number theory, but also in cryptography!

# The Hasse bound

Given an elliptic curve $E$ over a number field $K$, by clearing denominators we get a scheme $X$ over $\mathfrak{o}_K$. We can study the function $N_X(p^e)$, ignoring the finitely many primes modulo which the defining equation of $X$ does not reduce to a smooth equation.

**Theorem (Hasse, 1930s)**

*For any positive integer $e$, $|p^e + 1 - N_X(p^e)| \leq 2p^{e/2}$.*

# The Hasse bound

Given an elliptic curve $E$ over a number field $K$, by clearing denominators we get a scheme $X$ over $\mathfrak{o}_K$. We can study the function $N_X(p^e)$, ignoring the finitely many primes modulo which the defining equation of $X$ does not reduce to a smooth equation.

**Theorem (Hasse, 1930s)**

*For any positive integer $e$, $|p^e + 1 - N_X(p^e)| \leq 2p^{e/2}$.*

# A key quantity

By Hasse's theorem, there is a unique way to write

$$N_X(p) = p + 1 - p^{1/2}(\alpha_p + \beta_p)$$

with $\alpha_p, \beta_p \in \mathbb{C}$, $\mathrm{Im}(\alpha_p) \geq 0$, $|\alpha_p| = |\beta_p| = 1$, and $\alpha_p \beta_p = 1$.

**Theorem (Weil, 1940s)**

*For each positive integer e,*

$$N_X(p^e) = p^e + 1 - p^{e/2}(\alpha_p^e + \beta_p^e).$$

We may thus focus our attention on the function $p \mapsto N_X(p)$, or equivalently on the function $p \mapsto \alpha_p$. Since these are not limited to finitely many values, in order to talk about their average behavior, we must use the language of *equidistribution with respect to a measure*.

# Measures and equidistribution

For $S$ a compact topological space, let $\mathrm{Cont}(S, \mathbb{R})$ be the Banach space of continuous functions $S \to \mathbb{R}$. A *Radon measure* on $S$ is a continuous linear function $\mu : \mathrm{Cont}(S, \mathbb{R}) \to \mathbb{R}$ which is *positive* ($f \geq 0 \Rightarrow \mu(f) \geq 0$) and of *mass 1* ($f = 1 \Rightarrow \mu(f) = 1$).

For $S = [a, b] \subseteq \mathbb{R}$, we only consider measures which are sums of:

- a *continuous part*, given by $f \mapsto \int_S fg$ for some measurable function $g : S \to \mathbb{R}$ (the *density function*); and
- a *discrete part*, given by $f \mapsto c_1 f(x_1) + \cdots + c_n f(x_n)$ for some $c_1, \ldots, c_n \in \mathbb{R}$ and some $x_1, \ldots, x_n \in S$.

A sequence $x_1, x_2, \cdots \in S$ is *$\mu$-equidistributed* if for all $f \in \mathrm{Cont}(S, \mathbb{R})$,

$$\lim_{n \to \infty} \frac{1}{n}(f(x_1) + \cdots + f(x_n)) = \mu(f).$$

This is akin to *ergodicity* (time average = space average).

# Measures and equidistribution

For $S$ a compact topological space, let $\mathrm{Cont}(S, \mathbb{R})$ be the Banach space of continuous functions $S \to \mathbb{R}$. A *Radon measure* on $S$ is a continuous linear function $\mu : \mathrm{Cont}(S, \mathbb{R}) \to \mathbb{R}$ which is *positive* ($f \geq 0 \Rightarrow \mu(f) \geq 0$) and of *mass 1* ($f = 1 \Rightarrow \mu(f) = 1$).

For $S = [a, b] \subseteq \mathbb{R}$, we only consider measures which are sums of:

- a *continuous part*, given by $f \mapsto \int_S fg$ for some measurable function $g : S \to \mathbb{R}$ (the *density function*); and
- a *discrete part*, given by $f \mapsto c_1 f(x_1) + \cdots + c_n f(x_n)$ for some $c_1, \ldots, c_n \in \mathbb{R}$ and some $x_1, \ldots, x_n \in S$.

A sequence $x_1, x_2, \cdots \in S$ is *$\mu$-equidistributed* if for all $f \in \mathrm{Cont}(S, \mathbb{R})$,

$$\lim_{n \to \infty} \frac{1}{n}(f(x_1) + \cdots + f(x_n)) = \mu(f).$$

This is akin to *ergodicity* (time average = space average).

# Measures and equidistribution

For $S$ a compact topological space, let $\mathrm{Cont}(S, \mathbb{R})$ be the Banach space of continuous functions $S \to \mathbb{R}$. A *Radon measure* on $S$ is a continuous linear function $\mu : \mathrm{Cont}(S, \mathbb{R}) \to \mathbb{R}$ which is *positive* ($f \geq 0 \Rightarrow \mu(f) \geq 0$) and of *mass 1* ($f = 1 \Rightarrow \mu(f) = 1$).

For $S = [a, b] \subseteq \mathbb{R}$, we only consider measures which are sums of:

- a *continuous part*, given by $f \mapsto \int_S fg$ for some measurable function $g : S \to \mathbb{R}$ (the *density function*); and
- a *discrete part*, given by $f \mapsto c_1 f(x_1) + \cdots + c_n f(x_n)$ for some $c_1, \ldots, c_n \in \mathbb{R}$ and some $x_1, \ldots, x_n \in S$.

A sequence $x_1, x_2, \cdots \in S$ is *$\mu$-equidistributed* if for all $f \in \mathrm{Cont}(S, \mathbb{R})$,

$$\lim_{n \to \infty} \frac{1}{n}(f(x_1) + \cdots + f(x_n)) = \mu(f).$$

This is akin to *ergodicity* (time average = space average).

# Complex multiplication

An elliptic curve $E$ over $K$ has *complex multiplication (CM)* if there is an algebraic map $E \to E$ (possibly defined over a field larger than $K$) which corresponds to an endomorphism of the group structure other than multiplication by an integer. For example, the elliptic curve

$$y^2 = x^3 - x$$

admits the automorphism $(x, y) \mapsto (-x, y\sqrt{-1})$ of order 4.

Suppose $E$ has CM. From an explicit formula of Hecke (generalizing an example of Gauss), it follows that if the extra endomorphisms are defined over $K$, then the $\alpha_p$ are equidistributed for the uniform measure on the semicircle

$$\{z \in \mathbb{C} : |z| = 1, \quad \mathrm{Im}(z) \geq 0\}.$$

Otherwise, one must take half of the uniform measure plus half of the discrete measure at $i$. That is, $N_X(p) = p + 1$ with probability $1/2$.

# Complex multiplication

An elliptic curve $E$ over $K$ has *complex multiplication (CM)* if there is an algebraic map $E \to E$ (possibly defined over a field larger than $K$) which corresponds to an endomorphism of the group structure other than multiplication by an integer. For example, the elliptic curve

$$y^2 = x^3 - x$$

admits the automorphism $(x, y) \mapsto (-x, y\sqrt{-1})$ of order 4.

Suppose $E$ has CM. From an explicit formula of Hecke (generalizing an example of Gauss), it follows that if the extra endomorphisms are defined over $K$, then the $\alpha_p$ are equidistributed for the uniform measure on the semicircle

$$\{z \in \mathbb{C} : |z| = 1, \quad \mathrm{Im}(z) \geq 0\}.$$

Otherwise, one must take half of the uniform measure plus half of the discrete measure at $i$. That is, $N_X(p) = p + 1$ with probability $1/2$.

# The Sato-Tate conjecture

**Conjecture (Sato-Tate, early 1960s)**

*If $E$ does not have CM, then the $\mathrm{Re}(\alpha_p)$ are equidistributed with respect to the continuous measure on $[-1,1]$ with density function $\frac{2}{\pi}\sqrt{1-t^2}$.*

In other words, if one picks an elliptic curve and computes a histogram for the values $(N_X(p) - p - 1)/\sqrt{p}$ over a large range of prime ideals, one always observes convergence to one of three limiting shapes!

Very good methods for computing $N_X(p)$ have been developed (partly for applications in computer science); one can thus confirm the conjectured convergence with high numerical accuracy. For visual proof, see

`http://math.mit.edu/~drew/g1SatoTateDistributions.html`.

We will have more to say about experimental metholodogy later.

# The Sato-Tate conjecture

**Conjecture (Sato-Tate, early 1960s)**

*If $E$ does not have CM, then the $\mathrm{Re}(\alpha_p)$ are equidistributed with respect to the continuous measure on $[-1,1]$ with density function $\frac{2}{\pi}\sqrt{1-t^2}$.*

In other words, if one picks an elliptic curve and computes a histogram for the values $(N_X(p) - p - 1)/\sqrt{p}$ over a large range of prime ideals, one always observes convergence to one of three limiting shapes!

Very good methods for computing $N_X(p)$ have been developed (partly for applications in computer science); one can thus confirm the conjectured convergence with high numerical accuracy. For visual proof, see

`http://math.mit.edu/~drew/g1SatoTateDistributions.html`.

We will have more to say about experimental metholodogy later.

# Group-theoretic interpretation

Just like for polynomials in one variable, the measures with respect to which the $\alpha_p$ are equidistributed admit simple group-theoretic descriptions. In all three cases, the pairs $\{\alpha_p, \beta_p\}$ are distributed like the eigenvalues of a matrix chosen uniformly at random in a certain compact Lie group.

- If $E$ does not have CM, the group is $\mathrm{SU}(2)$.
- If $E$ has CM defined over $K$, the group is $\mathrm{SO}(2)$.
- If $E$ has CM not defined over $K$, the group is $N(\mathrm{SO}(2))$, the normalizer of $\mathrm{SO}(2)$ in $\mathrm{SU}(2)$. This group has two connected components, on one of which the eigenvalues are always $i, -i$.

The measure on each group is *Haar measure*, the unique translation-invariant measure.

# Progress on Sato-Tate

**Theorem (Clozel, Harris, Taylor, et al., late 2000s; very hard!)**

*The Sato-Tate conjecture holds for $K = \mathbb{Q}$ (and more generally for $K$ a totally real number field).*

It was shown by Serre (imitating the proof of the prime number theorem) that Sato-Tate would follow from analytic continuation of certain *L-functions* associated to $E$. Some of these were treated by the work of Wiles et al. establishing modularity of elliptic curves (and Fermat's last theorem).

The above theorem relies on rather sophisticated improvements of the method of Wiles, plus related developments in the *Langlands program* linking Galois representations and automorphic forms.

# Progress on Sato-Tate

**Theorem (Clozel, Harris, Taylor, et al., late 2000s; very hard!)**

*The Sato-Tate conjecture holds for $K = \mathbb{Q}$ (and more generally for $K$ a totally real number field).*

It was shown by Serre (imitating the proof of the prime number theorem) that Sato-Tate would follow from analytic continuation of certain *L-functions* associated to $E$. Some of these were treated by the work of Wiles et al. establishing modularity of elliptic curves (and Fermat's last theorem).

The above theorem relies on rather sophisticated improvements of the method of Wiles, plus related developments in the *Langlands program* linking Galois representations and automorphic forms.

# Progress on Sato-Tate

**Theorem (Clozel, Harris, Taylor, et al., late 2000s; very hard!)**

*The Sato-Tate conjecture holds for $K = \mathbb{Q}$ (and more generally for $K$ a totally real number field).*

It was shown by Serre (imitating the proof of the prime number theorem) that Sato-Tate would follow from analytic continuation of certain *L-functions* associated to $E$. Some of these were treated by the work of Wiles et al. establishing modularity of elliptic curves (and Fermat's last theorem).

The above theorem relies on rather sophisticated improvements of the method of Wiles, plus related developments in the *Langlands program* linking Galois representations and automorphic forms.

# Contents

# Hyperelliptic curves

A *hyperelliptic curve* over a field $K$ (again assuming $\mathrm{char}(K) \neq 2$) is a smooth projective algebraic curve $C$ defined by an equation of the form

$$y^2 = P(x)$$

where $P \in K[x]$ has no repeated factors. For $d = \deg(P)$, the *genus* of such a curve is $\lfloor \frac{d-1}{2} \rfloor$; for $K = \mathbb{C}$, this is the number of handles of the Riemann surface corresponding to this curve.

From now on, we assume $C$ is a hyperelliptic curve of genus 2 over a number field $K$ (and again let $X$ be a model over $\mathfrak{o}_K$ obtained by clearing denominators). There is no group structure on $C$, but there is an associated object that does have a group structure (the *Jacobian variety*). This structure is also relevant for cryptography! As a result, information about $N_X(p)$ has some practical value.

# Hyperelliptic curves

A *hyperelliptic curve* over a field $K$ (again assuming $\mathrm{char}(K) \neq 2$) is a smooth projective algebraic curve $C$ defined by an equation of the form

$$y^2 = P(x)$$

where $P \in K[x]$ has no repeated factors. For $d = \deg(P)$, the *genus* of such a curve is $\lfloor \frac{d-1}{2} \rfloor$; for $K = \mathbb{C}$, this is the number of handles of the Riemann surface corresponding to this curve.

From now on, we assume $C$ is a hyperelliptic curve of genus 2 over a number field $K$ (and again let $X$ be a model over $\mathfrak{o}_K$ obtained by clearing denominators). There is no group structure on $C$, but there is an associated object that does have a group structure (the *Jacobian variety*). This structure is also relevant for cryptography! As a result, information about $N_X(p)$ has some practical value.

# Extension fields

Unlike for elliptic curves, for hyperelliptic curves it is not the case that $N_X(p)$ determines $N_X(p^e)$ for all $e$. However, in genus 2, there do exist $\alpha_{p,1}, \dots, \alpha_{p,4} \in \mathbb{C}$ with $|\alpha_{p,i}| = 1$ for which for each positive integer $e$,

$$N_X(p^e) = p^e + 1 - p^{e/2}(\alpha_{p,1}^e + \cdots + \alpha_{p,4}^e).$$

We can also number things so that $\alpha_{p,1}\alpha_{p,3} = \alpha_{p,2}\alpha_{p,4} = 1$. Another interpretation: $N_X(p)$ and $N_X(p^2)$ together determine $N_X(p^e)$ for all $e$.

Upshot: in order to formulate an analogue of the Sato-Tate conjecture for hyperelliptic curves of genus 2, we will need to consider measures on two-dimensional spaces.

# Extension fields

Unlike for elliptic curves, for hyperelliptic curves it is not the case that $N_X(p)$ determines $N_X(p^e)$ for all $e$. However, in genus 2, there do exist $\alpha_{p,1}, \dots, \alpha_{p,4} \in \mathbb{C}$ with $|\alpha_{p,i}| = 1$ for which for each positive integer $e$,

$$N_X(p^e) = p^e + 1 - p^{e/2}(\alpha_{p,1}^e + \cdots + \alpha_{p,4}^e).$$

We can also number things so that $\alpha_{p,1}\alpha_{p,3} = \alpha_{p,2}\alpha_{p,4} = 1$. Another interpretation: $N_X(p)$ and $N_X(p^2)$ together determine $N_X(p^e)$ for all $e$.

Upshot: in order to formulate an analogue of the Sato-Tate conjecture for hyperelliptic curves of genus 2, we will need to consider measures on two-dimensional spaces.

# The generic Sato-Tate conjecture in genus 2

**Conjecture (Katz-Sarnak, 2000s)**

*Suppose that $C$ is generic (i.e., the Jacobian variety has no extra endomorphisms). Then as we vary over prime ideals, the multisets $\{\alpha_{p,1}, \alpha_{p,2}, \alpha_{p,3}, \alpha_{p,4}\}$ are equidistributed for the measure corresponding to the characteristic polynomial of a random (for Haar measure) matrix in the group $\mathrm{USp}(4)$ of $4 \times 4$ unitary symplectic matrices.*

For example, $N_X(p)$ should behave like $p+1$ minus $\sqrt{p}$ times the trace of a random matrix in $\mathrm{USp}(4)$. This is confirmed by experimental evidence:

`http://math.mit.edu/~drew/g2_D1_a1f.gif`

Unfortunately, there is not a single $C$ for which the methods used in genus 1 seem to be able to establish the Katz-Sarnak conjecture! (They do cover many of the exceptional cases described in the next conjecture.)

# The Sato-Tato conjecture in genus 2

Conjecture (Fité, K, Rotger, Sutherland, 2011)

(a) *There exists a closed subgroup $G$ of $\mathrm{USp}(4)$ (called the* **Sato-Tate group** *of $G$) for which the multisets $\{\alpha_{p,1}, \alpha_{p,2}, \alpha_{p,3}, \alpha_{p,4}\}$ are equidistributed for the measure corresponding to the characteristic polynomial of a random (for Haar measure) matrix in $G$.*

(b) *The group $G$ is conjugate to one of 52 possible groups, all of which can occur.*

(c) *If $K = \mathbb{Q}$, then exactly 34 groups can occur.*

The group $G$ can have as many as 48 connected components, e.g., for the curve $y^2 = x^6 - 5x^4 + 10x^3 - 5x^2 + 2x - 1$.

A more precise formulation of the conjecture includes a candidate for $G$ for which (b) and (c) are provably true; its construction is related to that of the *Mumford-Tate group*. (This depends on joint work with Banaszak.)

# The Sato-Tato conjecture in genus 2

**Conjecture (Fité, K, Rotger, Sutherland, 2011)**

(a) *There exists a closed subgroup $G$ of $\mathrm{USp}(4)$ (called the* **Sato-Tate group** *of $G$) for which the multisets $\{\alpha_{p,1}, \alpha_{p,2}, \alpha_{p,3}, \alpha_{p,4}\}$ are equidistributed for the measure corresponding to the characteristic polynomial of a random (for Haar measure) matrix in $G$.*

(b) *The group $G$ is conjugate to one of 52 possible groups, all of which can occur.*

(c) *If $K = \mathbb{Q}$, then exactly 34 groups can occur.*

The group $G$ can have as many as 48 connected components, e.g., for the curve $y^2 = x^6 - 5x^4 + 10x^3 - 5x^2 + 2x - 1$.

A more precise formulation of the conjecture includes a candidate for $G$ for which (b) and (c) are provably true; its construction is related to that of the *Mumford-Tate group*. (This depends on joint work with Banaszak.)

# Testing methodology

As noted earlier, one can prove the previous conjecture in some special cases, but not in any generic case (when $G = \mathrm{USp}(4)$). We are thus left to ask whether one can test the conjecture numerically. This is really two questions.

1. For a given $C$, how can one compute the $\alpha_{p,i}$ for many $p$ (say, all $p \leq 2^{30}$?)
2. Given such data, how can one evaluate the proposed equidistribution property?

# Testing methodology: computing the $\alpha_{p,i}$

Some available techniques for computing the $\alpha_{p,i}$ include the following. (Costing estimates are from K-Sutherland, 2008.)

- Compute $N_X(p), N_X(p^2)$ by actually counting solutions. This is best for $p$ small (say, less than $2^{16}$).
- Identify the group structure of the points of the Jacobian variety (as in genus 1). This works well in the range we consider (roughly $2^{16} < p < 2^{40}$).
- Use a trace formula in $p$-adic cohomology to compute a matrix whose characteristic polynomial has the $\alpha_{p,i}$ as roots. This is optimal in a somewhat larger range than we consider (roughly $2^{40} < p < 2^{64}$).
- Use Pila's generalization of Schoof's algorithm: compute $N_X(p), N_X(p^2)$ modulo $\ell$ for many small primes $\ell$. This is best for very large $p$, e.g., in cryptography applications where one takes a single $p$ of size about $2^{128}$.

# Testing methodology: evaluating the numerical fit

The easiest way to numerically test equidistribution is to consider one function $f$ of the $\alpha_{p,i}$ at a time. One can do this both visually, by making a histogram plot, and by computing *moment statistics*, i.e., the expected values $\mathbf{E}(f^n)$ for $n = 1, 2, \ldots$. The latter are forced to be integers, as they compute characters of certain virtual representations of $G$.

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

One can also make histogram plots on a suitable two-dimensional space, e.g., the first and second elementary symmetric functions of the $\alpha_{p,i}$.

`https://hensel.mit.edu:8000/home/pub/8/`

# Contents

# Curves of higher genus

The numerical methods for computing $N_X(p^e)$ extend to hyperelliptic curves of higher genus (with some loss of efficiency). Similar methods can also be developed for curves which are not hyperelliptic, e.g., those defined by nonsingular quartic polynomials in the plane (these being of genus 3).

In genus 4, the Sato-Tate group becomes a somewhat subtler invariant of the curve, as it is not determined by the endomorphisms of the Jacobian (as discovered by Mumford). Also, there is a possible distinction between Sato-Tate groups that can arise from curves and from abelian varieties. For instance, it is not yet known whether Mumford's exotic examples, which are constructed as abelian varieties, can ever occur as Jacobians of curves. This can be tested numerically!

# More general algebraic varieties

A fairly general version of the Sato-Tate conjecture has been formulated by Serre. It would be interesting to collect numerical data in some simple higher-dimensional cases, e.g., K3 surfaces. Some numerical methodology is suitable for this (especially $p$-adic cohomology, as in ongoing joint work with David Harvey).