# Computing $L$-series of hyperelliptic curves and distributions of Frobenius eigenvalues

Kiran S. Kedlaya

Massachusetts Institute of Technology

July 3, 2009

joint work with Andrew Sutherland

http://arxiv.org/abs/0801.2278
http://arxiv.org/abs/0803.4462

# $L$-series of curves

Let $C$ be a (smooth, projective, geometrically irreducible) curve of genus $g$ over a number field $K$. The $L$-series associated to $C$ is the Dirichlet series given by the Euler product

$$L(C, s) = \prod_{\mathfrak{p}} L_{\mathfrak{p}}(C, \mathrm{Norm}(\mathfrak{p})^{-s})^{-1}$$

in which $L_{\mathfrak{p}}(C, T)$ is a polynomial to be described shortly.

The product converges absolutely for $\mathrm{Real}(s) > \frac{3}{2}$, conjecturally with analytic continuation to $\mathbb{C}$ and functional equation $s \mapsto 2 - s$ with a specified gamma factor. This is only known in general for $g = 1, K = \mathbb{Q}$.

Motivating problem: compute enough terms of $L(C, s)$ to compute (conditionally) $L^{(i)}(C, 1)$ to high precision. (See Dokchitser, Computing special values of motivic $L$-functions.)

# $L$-polynomials

For $\mathfrak{p}$ a prime of good reduction for $C$, with residue field $\mathbb{F}_q$, the $L$-polynomial $L_{\mathfrak{p}}(C, T)$ is equal to $(1 - T)(1 - qT)$ times the zeta function

$$\zeta(C_{\mathbb{F}_q}, T) = \prod_{x \in C_{\mathbb{F}_q}} (1 - T^{[\kappa(x):\mathbb{F}_q]})^{-1},$$

where $x$ runs over closed points of the scheme $C_{\mathbb{F}_q}$. We also have

$$\zeta(C_{\mathbb{F}_q}, T) = \exp\left(\sum_{n=1}^{\infty} \frac{T^n}{n} \#C(\mathbb{F}_{q^n})\right).$$

For $\mathfrak{p}$ of bad reduction, $L_{\mathfrak{p}}(C, T)$ can be read off from a minimal regular model of $C$ at $\mathfrak{p}$. There are only a finite number of such $\mathfrak{p}$; I'll mostly ignore them. Instead, we ask how to efficiently compute $L_{\mathfrak{p}}(C, T)$ for all $\mathfrak{p}$ of good reduction with $\mathrm{Norm}(\mathfrak{p}) \leq N$.

# Distribution of Frobenius eigenvalues

The polynomial $L_{\mathfrak{p}}(C, T) \in \mathbb{Z}[T]$ has degree $2g$ and constant coefficient 1, and satisfies

$$L_{\mathfrak{p}}(C, T) = q^g T^{2g} L_{\mathfrak{p}}\left(C, \frac{1}{qT}\right).$$

Moreover, the roots of $L_{\mathfrak{p}}(C, T)$ in $\mathbb{C}$ have norm $q^{-1/2}$. Multiplying by $q^{1/2}$ gives the *normalized roots*, which lie on the unit circle.

Motivating problem: compute enough $L_{\mathfrak{p}}(C, T)$ to detect the limiting joint distribution of the normalized roots. Only finitely many possibilities; more on them later.

# Computing $L_{\mathfrak{p}}(C, T)$ for $g = 1$

For $C$ of the form $y^2 = x^3 + Ax + B$, one way to compute $L_{\mathfrak{p}}(C, T)$ is to enumerate $C(\mathbb{F}_q)$. This requires $\tilde{O}(q)$ steps.

For all but the smallest $q$, it is faster to examine the group $C(\mathbb{F}_q)$ using Shanks's baby-step-giant-step method. Since we know $|q + 1 - C(\mathbb{F}_q)| \leq 2q^{1/2}$, this requires $\tilde{O}(q^{1/4})$ steps.

For $q$ very large, one should (after Schoof, Elkies, Atkin) first compute $C(\mathbb{F}_q)$ modulo some small primes $\ell$, by computing the $\ell$-division polynomial (over $K$) and factoring modulo $\mathfrak{p}$. To get up to $q \sim 2^{32}$, this is only relevant for $\ell \leq 5$. (In cryptography, one considers only a single value of $q$, but it could be as big as $2^{160}$ or more.)

# Computing $L_{\mathfrak{p}}(C, T)$ for $g > 1$

To compute $L_{\mathfrak{p}}(C, T)$ by counting points, one must count over $\mathbb{F}_{q^n}$ for $n = 1, \ldots, g$. Better is to balance this against baby-step-giant-step on the Jacobian group $J(C)(\mathbb{F}_q)$.

The analogue of Schoof's algorithm for $g > 1$ is much less tractable (though Gaudry and Schost recently carried it out for $q = 2^{127} - 1$ and $g = 2$).

A better approach uses an explicit $p$-adic Weil cohomology. For $g = 2, K = \mathbb{Q}$, this is better for $q > 2^{32}$; for hyperelliptic curves with $g = 3, K = \mathbb{Q}$, the crossover is closer to $q > 2^{16}$.

# Weil cohomology

For $\mathfrak{p}$ of good reduction, there are several ways to construct a vector space $H^1(C)$ equipped with an endomorphism $F$, such that

$$\#C(\mathbb{F}_{q^n}) = 1 + q^n - \mathrm{Trace}(F^n, H^1(C)).$$

One of these is étale cohomology; this amounts to the dualized Tate module $V_\ell(C)^\vee$. This is hard to write down concretely.

Another is $p$-adic (rigid/crystalline) cohomology, which is much easier to write down concretely (though more difficult to prove theorems about). In fact, as a vector space, it can be identified with the algebraic de Rham cohomology $H^1_{\mathrm{dR}}(C)$; the hard part is to compute the operator $F$. This is usually computed using a method due to Monsky and Washnitzer.

# Example of $p$-adic cohomology: a hyperelliptic curve

Let $C$ be the hyperelliptic curve $y^2 = P(x)$ over $\mathbb{Q}$, with $P$ of odd degree without repeated roots. Then $H^1_{\mathrm{dR}}(C)$ is equal to the algebraic de Rham cohomology of the affine curve, which is freely generated by

$$\frac{x^i\,dx}{y} \qquad (i = 0, \dots, 2g-1).$$

(Remember that $2y\,dy = P'(x)\,dx$.) By writing down relations in cohomology like

$$0 \equiv d\left(\frac{A(x)}{y^s}\right) = \frac{A'(x)\,dx}{y^s} - \frac{sA(x)P'(x)\,dx}{2y^{s+2}}$$

we can explicitly write any 1-form in terms of the basis. Note that none of this refers explictly to a prime $p$.

# Example of $p$-adic cohomology: a Frobenius action

Now pick a prime $p$; we expect to get a matrix describing the Frobenius action, with entries in $\mathbb{Q}_p$ (which we will only compute to some $p$-adic accuracy, enough to determine the zeta function).

One way to compute the Frobenius action is to construct a continuous endomorphism $F$ of a certain weak $p$-adic completion of $\mathbb{Q}[x, y, z]/(y^2 - P(x), yz - 1)$ satisfying $F(x) = x^p$. This amounts to using Newton iteration to solve for

$$F(y) = F(P(x))^{1/2} = y^p \left(1 - pz^{2p} \frac{P(x)^p - P(x^p)}{p}\right)^{1/2}.$$

This requires $\tilde{O}(p)$ steps. For $p$ large, David Harvey has found a method requiring $\tilde{O}(p^{1/2})$ steps; this is what we used.

# Shameless advertisement for $p$-adic cohomology

$p$-adic cohomology can be used effectively for computing zeta functions of high-genus curves and higher-dimensional varieties, which most other methods cannot.

Sample: Magma can compute the zeta function of a genus 50 hyperelliptic curve over $\mathbb{F}_3$.

Sample: Abbott, K, Roe computed some zeta functions of quartic K3 surfaces over $\mathbb{F}_p$ for $p \leq 23$. (Direct point counts require $\tilde{O}(p^{20})$ steps.) A method of Lauder can probably do even better.

Sample: Alan Lauder computed $L$-functions of elliptic surfaces over $\mathbb{F}_7$ to see whether 100% of them have analytic rank 0 or 1. They do! (In the number field case, a bias towards rank 2 persists as far as has been computed.)

# Eigenvalue distributions for elliptic curves over $\mathbb{Q}$

From now on, we mostly take $K = \mathbb{Q}$ and write $L_p(T)$ for $L_{\mathfrak{p}}(C, T)$.

## Curves with complex multiplication

All elliptic curves with CM have the same limiting distribution. This follows from classical results. (Over a number field $K$, one must distinguish between whether or not the CM is defined over $K$.)

## Conjecture (Sato-Tate)

*For any elliptic curve without CM, the limiting distribution of the normalized trace of Frobenius is the semicircular distribution.*

Proven by Clozel, Harris, Shepherd-Baron, and Taylor (2006), provided $E$ does not have purely additive reduction.

# Unitarized $L$-polynomials

The polynomial

$$\bar{L}_p(T) = L_p(T/\sqrt{p}) = \sum_{i=0}^{2g} a_i T^i$$

is a real symmetric polynomial whose roots lie on the unit circle.

Every such polynomial arises as the characteristic polynomial $\chi(T)$ of some matrix in $USp(2g)$ ($2g \times 2g$ complex matrices that are both unitary and symplectic).

Note that the coefficients satisfy $|a_i| \leq \binom{2g}{i}$.

# The Katz-Sarnak model

## Conjecture (Katz-Sarnak)

*For a typical curve of genus $g$, the distribution of $\bar{L}_p(T)$ converges to the distribution of $\chi(T)$ in $USp(2g)$.*

"Typical" means curves with large Galois image.
For $g = 2$ this is equivalent to $\mathrm{End}(C) \cong \mathbb{Z}$ (i.e. no CM).

This conjecture is known to be true "on average" for universal families of hyperelliptic curves (including all genus 2 curves).

# The Haar measure on $USp(2g)$

Let $e^{\pm i\theta_1}, \ldots, e^{\pm i\theta_g}$ denote the eigenvalues of a random matrix (conjugacy class) in $USp(2g)$. The Weyl integration formula yields the Haar measure

$$\mu = \frac{1}{g!}\left(\prod_{j<k}(2\cos\theta_j - 2\cos\theta_k)\right)^2 \prod_j \left(\frac{2}{\pi}\sin^2\theta_j d\theta_j\right).$$

In genus 1 we have $USp(2) = SU(2)$ and $\mu = \frac{2}{\pi}\sin^2\theta d\theta$, which is the Sato-Tate distribution.

Note that $-a_1 = \sum 2\cos\theta_j$ is the trace.

# Moment sequences

The *moment sequence* of a random variable $X$ is

$$M[X] = (\mathrm{E}[X^0], \mathrm{E}[X^1], \mathrm{E}[X^2], \ldots).$$

Provided $X$ is suitably bounded, $M[X]$ exists and uniquely determines the distribution of $X$.

# Moment sequences

The *moment sequence* of a random variable $X$ is

$$M[X] = (\mathrm{E}[X^0], \mathrm{E}[X^1], \mathrm{E}[X^2], \ldots).$$

Provided $X$ is suitably bounded, $M[X]$ exists and uniquely determines the distribution of $X$.

Given sample values $x_1, \ldots, x_N$ for $X$, the nth *moment statistic* is the mean of $x_i^n$. It converges to $\mathrm{E}[X^n]$ as $N \to \infty$.

# Moment sequences

The *moment sequence* of a random variable $X$ is

$$M[X] = (\mathrm{E}[X^0], \mathrm{E}[X^1], \mathrm{E}[X^2], \ldots).$$

Provided $X$ is suitably bounded, $M[X]$ exists and uniquely determines the distribution of $X$.

Given sample values $x_1, \ldots, x_N$ for $X$, the nth *moment statistic* is the mean of $x_i^n$. It converges to $\mathrm{E}[X^n]$ as $N \to \infty$.

If $X$ is a symmetric integer polynomial of the eigenvalues of a random matrix in $USp(2g)$ (e.g. the trace), then $M[X]$ is an *integer* sequence (follows from representation theory). Similarly for any subgroup of $USp(2g)$.

# The typical trace moment sequence in genus 1

Using the measure $\mu$ in genus 1, for $t = -a_1$ we have

$$E[t^n] = \frac{2}{\pi} \int_0^\pi (2\cos\theta)^n \sin^2\theta d\theta.$$

# The typical trace moment sequence in genus 1

Using the measure $\mu$ in genus 1, for $t = -a_1$ we have

$$E[t^n] = \frac{2}{\pi} \int_0^\pi (2\cos\theta)^n \sin^2\theta d\theta.$$

This is zero when $n$ is odd, and for $n = 2m$ we obtain

$$E[t^{2m}] = \frac{1}{2m+1}\binom{2m}{m}.$$

and therefore

$$M[t] = (1, 0, 1, 0, 2, 0, 5, 0, 14, 0, 42, 0, 132, \ldots).$$

This is sequence A126120 in the OEIS.

# The typical trace moment sequence in genus $g > 1$

A similar computation in genus 2 yields

$$M[t] = (1, 0, 1, 0, 3, 0, 14, 0, 84, 0, 594, \ldots),$$

which is sequence A138349, and in genus 3 we have

$$M[t] = (1, 0, 1, 0, 3, 0, 15, 0, 104, 0, 909, \ldots),$$

which is sequence A138540.

The $n$th moment of the trace in genus $g$ is equal to the number of returning lattice paths in $\mathbb{Z}^g$ satisfying $x_1 \geq x_2 \geq \cdots \geq x_g \geq 0$ at every step (a Weyl chamber) [Grabiner-Magyar].

# The trace moment sequence of a CM curve in genus 1

For an elliptic curve with CM we find that

$$E[t^{2m}] = \frac{1}{2}\binom{2m}{m}, \qquad \text{for } m > 0$$

yielding the moment sequence

$$M[t] = (1, 0, 1, 0, 3, 0, 10, 0, 35, 0, 126, 0, \ldots),$$

whose even entries are A008828.

Where does this fit in a random matrix model?

# Exceptional distributions in genus 2

We surveyed the distributions of the genus 2 curves:

$$y^2 = x^5 + c_4x^4 + c_3x^3 + c_2x^2 + c_1x + c_0,$$

$$y^2 = b^6x^6 + b_5x^5 + b_4x^4 + b_3x^3 + b_2x^2 + b_1x + b_0,$$

with integer coefficients $|c_i| \leq 64$ and $|b_i| \leq 16$.
More than $10^{10}$ curves were tested.

We found over 30,000 non-isomorphic curves with exceptional distributions, about 20 distinct shapes.

All apparently converge to integer moment sequences.

# Genus 2 exceptional distributions (one example)

For a hyperelliptic curve whose Jacobian is the direct product of two elliptic curves, we compute $M[t] = M[t_1 + t_2]$ via

$$\mathrm{E}[(t_1 + t_2)^n] = \sum \binom{n}{i} \mathrm{E}[t_1^i]\mathrm{E}[t_2^{n-i}].$$

For example, using

$$M[t_1] = (1, 0, 1, 0, 2, 0, 5, 0, 14, 0, 42, 0, 132, \ldots),$$
$$M[t_2] = (1, 0, 1, 0, 3, 0, 10, 0, 35, 0, 126, 0, 462, \ldots),$$

we obtain $A138551$,

$$M[t] = (1, 0, 2, 0, 11, 0, 90, 0, 889, 0, 9723, \ldots).$$

# Analyzing the data in genus 2

Some survey highlights:

- At least 19 distinct distributions were found. This is exceeds the possibilities for $\mathrm{End}(C)$, $\mathrm{Aut}(C)$, or $\mathrm{MT}(C)$.
- Some obviously correspond to split Jacobians, but many do not. The same distribution can arise for curves with split and simple Jacobians.
- Some have positive zero-trace densities, some do not.
- The $a_1$ distribution appears to determine the $a_2$ distribution.

# Random matrix subgroup model

## Conjecture

*For a genus $g$ curve $C$, the distribution of $\bar{L}_p(T)$ converges to the distribution of $\chi(T)$ in some infinite compact subgroup $H \subseteq USp(2g)$.*

*Equality holds if and only if $C$ has large Galois image. Serre describes a candidate for $H$ using the motivic Galois group of $C$.*

Serre also shows that this follows from analytic information (analytic continuation and nonvanishing at the end of the critical strip) of the $L$-function *and its symmetric powers*.

# Representations of genus 1 distributions

The Sato-Tate distribution has $H = USp(2g)$, the typical case.

For CM curves, consider the subgroup of $USp(2) = SU(2)$:

$$H = \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \begin{pmatrix} i\cos\theta & i\sin\theta \\ i\sin\theta & -i\cos\theta \end{pmatrix} : \theta \in [0, 2\pi] \right\}.$$

This is a compact group (the normalizer of $SO(2)$ in $SU(2)$).

Its Haar measure yields the desired moment sequence. (Over a field containing the CM, one instead gets just $SO(2)$.)

# Candidate subgroups in genus 2

In genus 2 we have subgroups analogous to the two in genus 1.

Additionally, we consider embeddings of the two genus 1 groups as block diagonal matrices, where we allow "twisting" by $k$th roots of unity that lie in a quadratic extension of $\mathbb{Q}$ (so $k$ is 1,2,3,4, or 6).

This restriction corresponds to the requirement that $L_p(T)$ have integer coefficients (and yields integer moment sequences).

See `http://arxiv.org/abs/0803.4462` for details.

# A conjecturally complete classification in genus 2

This model yields a total of 24 candidates in addition to $USp(4)$ itself. Every distribution found in our survey has a distribution matching one of these candidates.

Initially we found only 19 exceptional distributions, but careful examination of the survey data yielded 3 missing cases.

# A conjecturally complete classification in genus 2

This model yields a total of 24 candidates in addition to $USp(4)$ itself. Every distribution found in our survey has a distribution matching one of these candidates.

Initially we found only 19 exceptional distributions, but careful examination of the survey data yielded 3 missing cases.

One of the remaining 2 candidates was recently ruled out by Serre, who suggests that the other is also similarly obstructed.

# Supporting evidence

In addition to the trace moment sequences, for each candidate subgroup $H \subseteq USp(4)$ we may also consider the component group of $H$ and the dimension of $H$.

Partitioning the $\bar{L}_p(T)$ data according to suitable constraints on $p$ yields the predicted component distributions.

The dimension of $H$ predicts the cardinality of the mod $\ell$ Galois image. For small $\ell$ we estimate this by counting how often the $\ell$-Sylow subgroup of $J(C/\mathbb{F}_p)$ has full rank.

# Open questions

- Consider the zero-trace densities that arise in genus 2. Can one prove that the list

$$0, 1/6, 1/4, 1/2, 7/12, 5/8, 3/4, 13/16, 7/8$$

  is complete in genus 2?

- Is there a lattice path interpretation for each of the identified subgroups in $USp(4)$? (Probably, using Littelmann's work on crystal bases.)

- Can one give a group-theoretic explanation of the list of subgroups we found? (Some ideas from Serre.)

- What happens in genus 3, and beyond?

# Computing *L*-series of hyperelliptic curves and distributions of Frobenius eigenvalues

Kiran S. Kedlaya

Massachusetts Institute of Technology

July 3, 2009

joint work with Kiran Kedlaya

http://arxiv.org/abs/0803.4462