L-polynomial distributions of genus 2 curves

Andrew V. Sutherland

Massachusetts Institute of Technology

May 25, 2010

joint work with Kiran Kedlaya

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

Distributions of Frobenius traces

Let $E/\mathbb{Q}$ be an elliptic curve (non-singular).
Let $t_p = \#E(\mathbb{F}_p) - p + 1$ denote the trace of Frobenius.

Consider the distribution of

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

as $p \leqslant N$ varies over primes of good reduction.

What happens as $N \to \infty$?

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

Trace distributions in genus 1

1. Typical case (no CM)

For any elliptic curve without CM, the limiting distribution is the semicircular distribution [Sato-Tate conjecture].[a]

[a]Proven (for almost all curves) by Clozel, Harris, Shepherd-Baron, and Taylor.

2. Exceptional cases (CM)

All elliptic curves with CM have the same limiting distribution [classical].

Zeta functions and L-polynomials

For a smooth projective curve $C/\mathbb{Q}$ and a good prime p define

$$Z(C/\mathbb{F}_p;T) = \exp\left(\sum_{k=1}^{\infty} N_k T^k/k\right),$$

where $N_k = \#C/\mathbb{F}_{p^k}$. This is a rational function of the form

$$Z(C/\mathbb{F}_p;T) = \frac{L_p(T)}{(1-T)(1-pT)},$$

where $L_p(T)$ is an integer polynomial of degree $2g$. For $g=2$:

$$L_p(T) = p^2T^4 + c_1pT^3 + c_2pT^2 + c_1T + 1.$$

Unitarized L-polynomials

The polynomial

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

has coefficients that satisfy $a_i = a_{2g-i}$ and $|a_i| \leqslant \binom{2g}{i}$.

Given a curve C, we may consider the distribution of $a_1, a_2, \ldots, a_g$, taken over primes $p \leqslant N$ of good reduction, as $N \to \infty$.

This talk focuses on the distribution of a_1 and a_2 in genus 2.

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

The Katz-Sarnak random matrix model

$\bar{L}_p$(T) is a real reciprocal polynomial whose roots lie on the unit circle.

Every such polynomial arises as the characteristic polynomial $\chi(T)$ of a unitary symplectic matrix in $\mathbb{C}^{2g\times 2g}$.

Conjecture 1

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

For $g = 2$, a curve is "typical" if and only if $\mathrm{End}(J(C)) \simeq \mathbb{Z}$ (no CM).

This conjecture has been proven "on average" for universal families of hyperelliptic curves, including all genus 2 curves, by Katz and Sarnak.

The Haar measure on $USp(2g)$

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

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

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.

Research Program

We wish to understand $\bar{L}_p$-distributions in genus 2, both the typical situation, and all the exceptional cases.

This presents three challenges:

- Data collection — Fast $\bar{L}_p$ computations
- Distinguishing distributions — Moment sequences
- Theoretical model — Subgroups of $USp(4)$

Collecting data

There are four ways to compute $\bar{L}_p$ in genus 2:

1. point counting: $\tilde{O}(p^2)$.
2. group computation: $\tilde{O}(p^{3/4})$.
3. p-adic methods: $\tilde{O}(p^{1/2})$.
4. ℓ-adic methods: $\tilde{O}(1)$.

For most of the feasible range of $p \leqslant N$, we found (2) to be the fastest.

For smaller p we can assist by point counting over $\mathbb{F}_p$ (but not $\mathbb{F}_{p^2}$).
For larger p we can assist with ℓ-adic information for $\ell = 2, 3$.

Computing L-series of hyperelliptic curves, ANTS VIII, 2008, KS.

Performance comparison

$p \approx 2^k$	points+group	group	p-adic
2^{14}	**0.22**	0.55	4
2^{15}	**0.34**	0.88	6
2^{16}	**0.56**	1.33	8
2^{17}	**0.98**	2.21	11
2^{18}	**1.82**	3.42	17
2^{19}	**3.44**	5.87	27
2^{20}	**7.98**	10.1	40
2^{21}	18.9	**17.9**	66
2^{22}	52	**35**	104
2^{23}		**54**	176
2^{24}		**104**	288
2^{25}		**173**	494
2^{26}		**306**	871
2^{27}		**505**	1532

Time to compute $L_p(T)$ in CPU milliseconds on a 2.5 GHz AMD Athlon

Time to compute $\bar{L}_p$ for $p \leqslant N$

N	2 cores	16 cores
2^{16}	1	< 1
2^{17}	4	2
2^{18}	12	3
2^{19}	40	7
2^{20}	2:32	24
2^{21}	10:46	1:38
2^{22}	40:20	5:38
2^{23}	2:23:56	19:04
2^{24}	8:00:09	1:16:47
2^{25}	26:51:27	3:24:40
2^{26}		11:07:28
2^{27}		36:48:52

Characterizing distributions

The *moment sequence* of a random variable X is

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

For suitably bounded X, the moment sequence $M[X]$ is well defined 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 $N \to \infty$.

Theorem

If X is a coefficient of the characteristic polynomial of a random matrix in a compact subgroup of $GL_n(\mathbb{C})$, then $M[X]$ is an integer sequence.

The typical trace moment sequence in genus 1

Using the measure μ 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

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

which is sequence A138349, and in genus 3 we have

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

which is sequence A138540.

In genus g, the nth moment of the trace is the number of returning walks of length n on $\mathbb{Z}^g$ with $x_1 \geqslant x_2 \geqslant \cdots \geqslant x_g \geqslant 0$ [Grabiner-Magyar].

The exceptional trace moment sequence 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.

An exceptional trace moment sequence in Genus 2

For a hyperelliptic curve whose Jacobian is isogenous to 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

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

we obtain $A138551$,

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

The second moment already differs from the standard sequence, and the fourth moment differs greatly (11 versus 3).

Sieving for exceptional curves

We surveyed the $\bar{L}_p$-distributions of 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 $|c_i| \leqslant 64$ and $|b_i| \leqslant 16$, over 10^{10} curves.

We initially computed $\bar{L}_p$ for $p \leqslant N \approx 2^{12}$.

We then filtered out "unexceptional" curves (over 99% of them), extended the computation using $N = 2^{16}$, and filtered again.

We were left with about 30,000 non-isomorphic "exceptional" curves, with what appeared to be about 20 different distributions.

Representative examples were then extended to $N = 2^{26}$.

Survey highlights

Some provisional observations:

- The moment statistics always appear to converge to integers.
- At least 20 apparently distinct $\bar{L}_p$-distributions were found. This exceeds the possibilities for $\mathrm{End}(J(C))$ and $\mathrm{Aut}(C)$.
- The same $\bar{L}_p$-distribution can arise for split and simple Jacobians.
- There appear to be at least 9 distinct possibilities for the density $z(C)$ of zero traces. Several exceptional cases have $z(C) = 0$.
- The a_1 distribution appears to determine the a_2 distribution.

#	$z(C)$	M_2	M_4	M_6	M_8	$f(x)$
1	0	1	3	14	84	$x^5 + x + 1$
2	0	2	10	70	588^*	$x^5 - 2x^4 + x^3 + 2x - 4$
3	0	2	11	90	888^*	$x^5 + 20x^4 - 26x^3 + 20x^2 + x$
4	0	2	12	110	1203^*	$x^5 + 4x^4 + 3x^3 - x^2 - x$
5	0	4	32	320	3581^*	$x^5 + 7x^3 + 32x^2 + 45x + 50$
6	1/6	2	12	100	979^*	$x^5 - 5x^3 - 5x^2 - x$
7	1/4	2	12	100	1008^*	$x^5 + 2x^4 + 2x^2 - x$
8	1/4	2	12	110	1257^*	$x^5 - 4x^4 - 2x^3 - 4x^2 + x$
9	1/2	1	5	35	293^*	$x^5 - 2x^4 + 11x^3 + 4x^2 + 4x$
10	1/2	1	6	55	601^*	$x^5 - 2x^4 - 3x^3 + 2x^2 + 8x$
11	1/2	2	16	160	1789^*	$x^5 + x^3 + x$
12	1/2	2	18	220	3005^*	$x^5 - 3x^4 + 19x^3 + 4x^2 + 56x - 12$
13	1/2	4	48	640	8949^*	$x^6 + 1$
14	7/12	1	6	50	489^*	$x^5 - 4x^4 - 3x^3 - 7x^2 - 2x - 3$
15	7/12	2	18	200	2446^*	$x^6 + 2$
16	5/8	1	6	50	502^*	$x^5 + x^3 + 2x$
17	5/8	2	18	200	2515^*	$x^5 - 10x^4 + 50x^2 - 25x$
18	3/4	1	8	80	894^*	$x^5 - 2x^3 - x$
19	3/4	1	9	100	1222^*	$x^5 - 1$
20	3/4	1	9	110	1501^*	$11x^6 + 11x^3 - 4$
21	3/4	2	24	320	4474^*	$x^5 + x$
22	13/16	1	9	100	1254^*	$x^5 + 3x$
23	7/8	1	12	160	2237^*	$x^5 + 2x$

Random matrix subgroup model

Conjecture 1

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

Conjecture 2

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

Equality holds if and only if C has large Galois image.

Subgroups representing genus 1 $\bar{L}_p$-distributions

In the typical case H is the group $G_1 = USp(2g) = SU(2)$.

For CM curves, we let H be the subgroup $G_2 \subset USp(2)$ defined by

$$G_2 = \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \begin{pmatrix} i\cos\theta & i\sin\theta \\ i\sin\theta & -i\cos\theta \end{pmatrix} : \theta \in [0, 2\pi] \right\}.$$

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

The Haar measure on G_2 yields the desired moment sequence

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

and the correct zero trace density $z(H) = 1/2$.

Candidate subgroups H in genus 2

We can immediately identify four candidates for H:

$$USp(4), \qquad G_1 \times G_1, \qquad G_1 \times G_2, \qquad G_2 \times G_2.$$

Additionally, we define subgroups H_i^k for $i = 1, 2$ and $k = 1, 2, 3, 4, 6$, in which G_i is diagonally embedded with a copy of itself that has been "twisted" by a kth root of unity (the restriction on k is necessary).

Finally, for any of the groups H above, we may consider the group $J(H)$ obtained by including the matrix

$$J = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix}.$$

Not all of these groups yields distinct distributions, but 24 of them do. There is also an index 2 subgroup K of $J(G_2 \times G_2)$.

Candidate subgroups H of $USp(4)$

#	H	d	$c(H)$	$z(H)$	M_2	M_4	M_6	M_8	M_{10}
1	$USp(4)$	10	1	0	1	3	14	84	594
2	$G_1 \times G_1$	6	1	0	2	10	70	588	5544
3	$G_1 \times G_2$	4	2	0	2	11	90	889	9723
4	H_1^3	3	3	0	2	12	110	1204	14364
5	H_1	3	1	0	4	32	320	3584	43008
6	H_1^6	3	6	1/6	2	12	100	980	10584
7	H_1^4	3	4	1/4	2	12	100	1008	11424
8	$G_2 \times G_2$	2	4	1/4	2	12	110	1260	16002
9	$J(G_1 \times G_1)$	6	2	1/2	1	5	35	294	2772
10	$J(H_1^3)$	3	6	1/2	1	6	55	602	7182
11	H_1^-	3	2	1/2	2	16	160	1792	21504
12	H_2^3	1	6	1/2	2	18	220	3010	43092
13	H_2	1	2	1/2	4	48	640	8960	129024
14	$J(H_1^6)$	3	12	7/12	1	6	50	490	5292
15	H_2^6	1	12	7/12	2	18	200	2450	31752
16	$J(H_1^4)$	3	8	5/8	1	6	50	504	5712
17	H_2^4	1	8	5/8	2	18	200	2520	34272
18	$J(H_1^-)$	3	4	3/4	1	8	80	896	10752
19	K	2	4	3/4	1	9	100	1225	15876
20	$J(H_2^3)$	1	12	3/4	1	9	110	1505	21546
21	H_2^-	1	4	3/4	2	24	320	4480	64512
22	$J(H_2^4)$	1	16	13/16	1	9	100	1260	17136
23	$J(H_2^-)$	1	8	7/8	1	12	160	2240	32256
*	$J(G_2 \times G_2)$	2	8	5/8	1	6	55	630	8001
*	$J(H_2^6)$	1	24	19/24	1	9	100	1225	15876

A conjecturally complete classification in genus 2

Every distribution found in our survey (and in the literature) 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.

This left only $J(G_2 \times G_2)$ and $J(H_2^6)$ unaccounted for.

$J(G_2 \times G_2)$ has now been ruled out by Serre.
A similar (but more difficult) argument may apply to $J(H_2^6)$.

Further supporting evidence

For each candidate subgroup $H \subseteq USp(4)$ we may consider the component group of H and the dimension $d(H)$.

In many cases, we can partition the $\bar{L}_p$ data via constraints on p.
In every such case this yields the predicted component distributions.

The mod ℓ Galois image of C should have size $\approx \ell^d$, where $d = d(H)$.
The ℓ-Sylow subgroup of $J(C/\mathbb{F}_p)$ then has full rank for a set of primes of density ℓ^{-d}. This has been confirmed for small d and ℓ.

Open questions

- Can one prove that the list

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

  of values for $z(C)$ is complete in genus 2?

- Is their a lattice path interpretation for each of the identified subgroups in $USp(4)$?

- What happens in genus 3?

L-polynomial distributions of genus 2 curves

Andrew V. Sutherland

Massachusetts Institute of Technology

May 25, 2010

joint work with Kiran Kedlaya

http://arxiv.org/abs/0803.4462