# A heuristic for the distribution of point counts for random curves over a finite field

Jeffrey D. Achter, Daniel Erman, Kiran S. Kedlaya,
Melanie Matchett Wood, and David Zureick-Brown

October 27, 2014

**Abstract**

How many rational points are there on a random algebraic curve of large genus $g$ over a given finite field $\mathbb{F}_q$? We propose a heuristic for this question motivated by a (now proven) conjecture of Mumford on the cohomology of moduli spaces of curves; this heuristic suggests a Poisson distribution with mean $q + 1 + 1/(q - 1)$. We prove a weaker version of this statement in which $g$ and $q$ tend to infinity, with $q$ much larger than $g$.

## 1 Introduction

The purpose of this paper is to propose a heuristic answer to the following question: what is the distribution of the number of rational points on a random algebraic curve over a fixed finite field $\mathbb{F}_q$ as the genus goes to infinity? This is a question that can be translated into a question about the number of $\mathbb{F}_q$-points of the moduli space $M_{g,n}$ of curves of genus $g$ with $n$ marked points. Our fundamental heuristic assumption is that, in the Grothendieck-Lefschetz trace formula to count $\mathbb{F}_q$-points on $M_{g,n}$, only the tautological classes contribute to the main term in the limit; we prove that this assumption implies the distribution of points on a random curves goes to a Poisson distribution with mean $q + 1 + 1/(q - 1)$. Moreover, one can make a more precise statement in a certain limit where $q$ and $g$ both tend to infinity, but $q$ grows significantly faster than $g$. These predictions and results are in the spirit of the work of Ellenberg–Venkatesh–Westerland [1] on the relationship between stable homology of Hurwitz spaces and Cohen-Lenstra heuristics; they are also in a sense reciprocal to the work of Faber–Pandharipande [2], in which point counts on $M_{g,n}$ for small $g$ are used to study the tautological classes.

Before making these statements more precise, we describe some similar questions which have been studied and indicate how this question differs somewhat from these. The distribution of the number of rational points on a random (smooth, projective, geometrically irreducible) algebraic curve of a given class

over a given finite[1] field has become a fundamental theme in the nascent field of *arithmetic statistics*. Some examples of classes for which this topic has been studied previously include hyperelliptic curves [4], cyclic trigonal curves [5], non-cyclic trigonal curves [6], cyclic $p$-gonal curves [7, 8], superelliptic and cyclic $m$-gonal curves [9], abelian covers of the line [10], Artin-Schreier curves [11, 12, 13], smooth plane curves [14], complete intersections in a fixed projective space [15], and curves in a fixed Hirzebruch surface [16]. In each of these cases, every curve $C$ in the family maps to a fixed base space $\phi\colon C \to P$ and the (asymptotic) distribution of points on a random $C$ is given by a sum of independent bounded random variables associated to the rational points of the base space. For each $p \in P(\mathbb{F}_q)$, the associated random variable is the number of rational points in $\phi^{-1}(p)$.

Of course, the most natural and interesting family of smooth, projective curves is the family of all such curves, but proving a result about the distribution of points in this family seems currently out of reach. The class of arbitrary curves differs from the previously mentioned classes in several important ways. One is that the number of rational points on the varying curve is not *a priori* bounded. A prior example sharing this property is that of [17], who considered curves lying in a sequence of surfaces with unbounded point counts. In this case, the average number of points on the curves is unbounded, so one is forced to renormalize to get a limiting distribution with finite mean, which turns out to be Gaussian.

The second distinctive feature of the class of all curves, which separates it from both [17] and most of the preceding examples, is that the moduli space is not rational or even unirational. That is, an arbitrary curve cannot be specified uniformly in terms of a collection of parameters. This makes even the "denominator" in the question, the total number of curves over $\mathbb{F}_q$ of a fixed genus, extremely difficult to understand. (See [18] for an upper bound.)

Finally, the lack of nontrivial maps from curves in the family to a fixed space means there is no way to make sensible probabilistic models which split the point count into a sum of independent random variables.

Let us now make things more precise for the class of curves. Let $M_g$ denote the fine moduli space of curves of genus $g$ in the sense of Deligne and Mumford [19]; it is an object in the category of algebraic stacks over $\operatorname{Spec}(\mathbb{Z})$. The set $|M_g(\mathbb{F}_q)|$ of (isomorphism classes of) $\mathbb{F}_q$-rational points of $M_g$ may then be identified with the set of isomorphism classes of smooth, projective, geometrically connected curves of genus $g$ over $\mathbb{F}_q$. For $x \in |M_g(\mathbb{F}_q)|$, let $C_x$ be any curve in the isomorphism class corresponding to $x$ and let $\operatorname{Aut}(C_x)$ be the group of automorphisms of $C_x$ as a curve over $\mathbb{F}_q$ (not over an algebraic closure over $\mathbb{F}_q$). We equip $|M_g(\mathbb{F}_q)|$ with the probability measure in which each point $x$ is weighted proportionally to $1/\#\operatorname{Aut}(C_x)$. This is well-understood to be the most natural way to count objects with automorphisms, and matches the weighting of points in the Lefschetz trace formula for Deligne-Mumford stacks given by Behrend

[1] The corresponding question over a number field is also central in arithmetic statistics, but has a rather different flavor. See [3] for a comprehensive survey.

[20].

Let $C_g$ be the (random) curve associated to a random $x \in |M_g(\mathbb{F}_q)|$ drawn according to the above probability measure. For each $g$, $\#C_g(\mathbb{F}_q)$ is a random variable taking values in the nonnegative integers, and we are interested in the limiting behavior of the distributions of these random variables as $g \to \infty$. We prove that a heuristic assumption about the cohomology of $M_{g,n}$ (Heuristic 2) implies that these distributions converge to a Poisson distribution with mean $q+1+1/(q-1) = q+1+q^{-1}+q^{-2}+\cdots$; more precisely, we show that Heuristic 2 implies the following predictions.

**Conjecture 1.** *Put $\lambda := \lambda(q) = q + 1 + 1/(q-1)$.*

*a. For all nonnegative integers $n$,*

$$\lim_{g\to\infty} \operatorname{Prob}(\#C(\mathbb{F}_q) = n : C \in M_g(\mathbb{F}_q)) = \frac{\lambda^n e^{-\lambda}}{n!}.$$

*b. For all positive integers $n$,*

$$\lim_{g\to\infty} \mathbb{E}(\#C(\mathbb{F}_q)^n : C \in M_g(\mathbb{F}_q)) = \sum_{i=1}^{n} \left\{ {n \atop i} \right\} \lambda^i,$$

*where $\left\{ {n \atop i} \right\}$ denotes a Stirling number of the second kind (i.e., the number of unordered partitions of $\{1, \ldots, n\}$ into $i$ disjoint sets).*

Note that part (b) implies part (a): the moment sequence of the Poisson distribution has exponential growth and thus determines the distribution uniquely [21, Theorem 30.1], and for such a limiting distribution convergence at the level of moments implies convergence at the level of distributions [21, Theorem 30.2]. If we let $(X)_n := X(X-1)\cdots(X-n+1)$, then the falling moments

$$\lim_{g\to\infty} \mathbb{E}((\#C(\mathbb{F}_q))_n : C \in M_g(\mathbb{F}_q)) = \lambda^n \tag{1}$$

(for all positive integers $n$) are equivalent to the standard moments in (b) above.

Let $M_{g,n}$ denote the moduli space of curves of genus $g$ with $n$ distinct marked points. Each element $x \in M_{g,n}(\mathbb{F}_q)$ now corresponds to a tuple $(C_x, P_1, \ldots, P_n)$ where $C_x$ is as before and $P_1, \ldots, P_n$ are distinct elements of $C_x(\mathbb{F}_q)$. We equip the points of $M_{g,n}(\mathbb{F}_q)$ with the weights where $x$ has weight $1/\operatorname{Aut}(C_x, P_1, \ldots, P_n)$ (i.e., we only consider automorphisms of $C_x$ fixing $P_1, \ldots, P_n$). By an easy orbit counting argument,

$$\mathbb{E}((\#C(\mathbb{F}_q))_n : C \in M_g(\mathbb{F}_q)) = \frac{\#M_{g,n}(\mathbb{F}_q)}{\#M_g(\mathbb{F}_q)}$$

(where # denotes weighted count). Thus, we may rewrite Conjecture 1 as the statement

$$\lim_{g\to\infty} \frac{\#M_{g,n}(\mathbb{F}_q)}{\#M_g(\mathbb{F}_q)} = \lambda^n. \tag{2}$$

Let us now make explicit how we would like to study $\#M_{g,n}(\mathbb{F}_q)/\#M_g(\mathbb{F}_q)$ using the Grothendieck-Lefschetz-Behrend trace formula. For a smooth Deligne-Mumford stack $X$ over $\mathbb{F}_q$, the trace formula asserts that for any prime $\ell$ not dividing $q$,

$$\#X(\mathbb{F}_q) = \sum_{i=0}^{2\dim(X)} (-1)^i \mathrm{Trace}(\mathrm{Frob}, H^i_{c,\mathrm{et}}(X, \mathbb{Q}_\ell))$$

where $H^i_{c,\mathrm{et}}(X, \mathbb{Q}_\ell)$ denotes compactly supported étale cohomology and Frob is the geometric Frobenius automorphism. By Deligne's proof of the Riemann hypothesis for algebraic varieties, each eigenvalue $\alpha$ of Frob on $H^i_{c,\mathrm{et}}(X, \mathbb{Q}_\ell)$ is an algebraic integer with the property that for some $w \in \{0, \ldots, i\}$ (called the *weight* of $\alpha$), the conjugates of $\alpha$ in $\mathbb{C}$ all have absolute value $q^{w/2}$.

This suggests that one should be able to estimate $\#M_{g,n}(\mathbb{F}_q)$, and hence the ratio $\#M_{g,n}(\mathbb{F}_q)/\#M_g(\mathbb{F}_q)$, by computing the action of geometric Frobenius on the highest-degree cohomology groups and burying the other contributions to the trace formula in an error term. Moreover, the highest degree cohomology groups with their Frobenius action are known exactly (see Theorem 8): they are spanned by so-called tautological classes (see below). Unfortunately, this approach does not lead to any provable estimates for fixed $q$ because the Betti numbers of $M_{g,n}$ grow superexponentially in $g$ (e.g., see [22] for the calculation of the Euler characteristic). Thus, even though terms from lower degree cohomology groups contribute to the Grothendieck-Lefschetz sum with smaller weight, there are so many of them that they cannot *a priori* be treated as negligible compared to the top-degree contributions.

Despite this imbalance, we can still make a reasonable heuristic about what we expect the asymptotics of the Grothendieck-Lefschetz sum to be. One can classify the Frobenius eigenvalues of $H^*_{c,\mathrm{et}}$ of each weight $w$ as "causal" and "random." The causal eigenvalues are the ones whose presence is compelled by the existence of certain algebraic cycles (in our case, the eigenvalues of the tautological classes); these eigenvalues must be integral powers of $q$. It is plausible to model the random eigenvalues of a given weight $w$ by the eigenvalues of a random unitary[2] matrix times $q^{w/2}$. Let $b_k$ be the number of "random" eigenvalues of weight $2\dim M_{g,n} - k$ (*coweight* $k$). (We have $b_k = 0$ for $k \leq \frac{2g-2}{3}$; see Theorem 8.) For $k > \frac{2g-2}{3}$, if there are few eigenvalues of coweight $k$, e.g. $b_k = o(q^{k/2})$, then the weight $k$ eigenvalues contribute nothing to the Grothendieck-Lefschetz sum in the limit as $g \to \infty$. On the other hand, if there are many eigenvalues of coweight $k$, and we model them with eigenvalues of a large random unitary matrix, we know from a result of Diaconis–Shahshahani [23] that this matrix has bounded trace with high probability. It is thus a sensible heuristic to neglect the contribution of all but the causal eigenvalues. Our neglect of the random Frobenius eigenvalues is also consistent with a commonly held philosophy in the study of moduli spaces, that no natural geometric questions depend on the non-tautological classes (e.g., see [24]).

[2] In middle cohomology, it is more natural to use a random unitary symplectic matrix or a random Hermitian matrix instead, but the same discussion applies to these models.

That this heuristic is sensible relies crucially on the fact that there are no random eigenvalues of large weight, which is a deep fact about the cohomology of moduli spaces of curves conjectured by Mumford and later proved using topological techniques (see Section 2 for references). The compactly supported étale cohomology in high degrees (or equivalently by Poincaré duality and a Betti-étale comparison isomorphism, the Betti cohomology in low degrees; see Theorem 8) is spanned by *tautological* classes, i.e., classes which arise from algebraic cycles produced by canonical morphisms between moduli spaces. The prototypical example of such a class is the first Chern class of the relative dualizing sheaf of the morphism $M_{g,n} \to M_{g,n-1}$ obtained by forgetting one marked point.

We may formalize our heuristic as follows. Write $R^*_{c,\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell)$ for the subspace of $H^i_{c,\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell)$ generated by tautological classes, and $B^*_{c,\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell) := H^*_{c,\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell)/R^*_{c,\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell)$.

**Heuristic 2.** *As $g \to \infty$, only the tautological classes are asymptotically relevant to a Grothendieck-Lefschetz trace formula computation of $\#M_{g,n}(\mathbb{F}_q)$. More precisely,*

$$\lim_{g\to\infty} \frac{\sum_{i=0}^{2\dim M_{g,n} - \frac{2g-2}{3}} (-1)^i \mathrm{Trace}(\mathrm{Frob}, B^i_{c,\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell))}{q^{\dim M_{g,n}}} = 0.$$

It is convenient for our heuristic that the tautological classes are *stable*. This means that for $i \geq \dim(M_{g,n}) - \frac{2g-2}{3}$, the groups $R^i_{c,\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell)$ (and thus the groups $H^i_{c,\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell)$) can be described in a manner independent of $g$, making it particularly nice to take the limit in $g$. Further, the number of lower degree tautological classes is sufficiently bounded that we can ignore their contribution to the Grothendieck-Lefschetz sum.

Our first main result is the following theorem.

**Theorem 3.** *Heuristic 2 implies Conjecture 1.*

Our second main result establishes unconditionally a weaker version of Conjecture 1 in which both $g$ and $q$ tend to infinity; this result lends some credence to Conjecture 1. In particular, since the error term is smaller than $q^{-m}$ for any fixed $m$, this result rules out any alternate conjecture in which each moment is a universal Laurent series in $q^{-1}$.

**Theorem 4.** *For any $K > 144$, any function $q(g) > g^K$, and any nonnegative integers $n$, for $q = q(g)$ we have*

$$\lim_{g\to\infty} \frac{\#M_{g,n}(\mathbb{F}_q)}{\#M_g(\mathbb{F}_q)} = \lambda^n + O(q^{-g/6}).$$

The key to proving Theorem 4 is that, so long as $q \gg g$, the unstable homology is negligible in the Grothendieck-Lefschetz trace computation.

It would be interesting to compute what a heuristic similar to Heuristic 2 suggests about the average number of points on a stable curve of genus $g$, as

$g \to \infty$. Our approach fails to directly yield an answer. In particular, we would seek a computation along the lines of Lemma 10, but this is complicated by the fact that the dimensions of the tautological cohomology $R^i(\overline{M_{g,n}})$ can grow exponentially in $g$, as $g \to \infty$.

In Section 2, we review the topological results showing that the low degree singular cohomology of $M_{g,n}$ is tautological and giving a precise description of the cohomology groups. In Section 3, we translate these results into compactly supported étale cohomology using comparison isomorphisms and determine the effect of Frobenius. In Section 4 we prove Theorem 3. In Section 5, we prove Theorem 4. In Section 6, we outline some thoughts and questions about how a random matrix model might give evidence for or against Conjecture 1.


## Acknowledgements

This project was started at the AIM workshop "Arithmetic statistics over number fields and function fields" (January 27-31, 2014). Achter was supported by Simons Foundation grant 204164 and NSA grant H98230-14-1-0161. Erman was supported by NSF grant DMS-1302057. Kedlaya was supported by NSF grant DMS-1101343; he also thanks MSRI for its hospitality during fall 2014 as supported by NSF grant DMS-0932078. Wood was supported by the American Institute of Mathematics and NSF grants DMS-1147782 and DMS-1301690. Thanks also to Jordan Ellenberg, Aaron Pixton, and Ravi Vakil for helpful discussions.


## 2 Stability and tautological classes: singular cohomology

We begin by recalling some deep results on the stable singular cohomology of moduli spaces. These results are typically stated without marked points; we must add a bit of extra analysis to deal with the markings.

**Theorem 5.** *For any nonnegative integers $g, n, i$ with $i \leq \frac{2g-2}{3}$, there exists an isomorphism $H_i(M_{g,n}, \mathbb{Q}) \to H_i(M_{g+1,n}, \mathbb{Q})$. By the universal coefficient theorem, this gives rise to an isomorphism $H^i(M_{g,n}, \mathbb{Q}) \to H^i(M_{g+1,n}, \mathbb{Q})$.*

*Proof.* This was first proved with a slightly more restrictive bound on $i$ by Harer [25, 26]. The statement as given includes results of several authors; see [27, Theorem 1.1]. □

The proof of this result is ultimately topological: by Teichmüller theory, one may identify $M_{g,n}$ up to homotopy with a classifying space of the mapping class group $\Gamma_{g,n}$ of a compact Riemann surface (without boundary) of genus $g$ with $n$ marked points. One may take a homotopy limit to obtain a group $\Gamma_{\infty,n}$ whose group (co)homology computes the stable (co)homology of $M_{g,n}$.

Let us now momentarily restrict attention to the case $n = 0$. Following Mumford, we define the *tautological ring* to be the graded polynomial ring $R := \mathbb{Q}[\kappa_1, \kappa_2, \dots]$ with $\deg(\kappa_j) = 2j$. We obtain a map from $R$ to the Chow ring of $M_g$ as follows: let $\psi$ be the relative dualizing sheaf of the morphism $M_{g,1} \to M_g$ which forgets the marked point, then let $\kappa_j$ be the pushforward of $\psi^{j+1}$ along $M_{g,1} \to M_g$.

**Theorem 6.** *The induced map $R \to H^*(M_g, \mathbb{Q})$ of graded rings is an isomorphism in degrees up to $\frac{2g-2}{3}$.*

*Proof.* This follows from Theorem 5 plus a theorem of Madsen and Weiss identifying $R$ with the stable cohomology ring [28]. □

We now consider the effect of marked points. Define the tautological ring $R_n = R[\psi_1, \dots, \psi_n]$ with $\deg(\psi_i) = 2$. We obtain a map from $R_n$ to the Chow ring of $M_{g,n}$ as follows: map $\kappa_j$ as before, and map $\psi_i$ to the relative dualizing sheaf of the morphism $M_{g,n} \to M_{g,n-1}$ which forgets the $i$-th marked point.

**Theorem 7.** *The induced map $R_n \to H^*(M_{g,n}, \mathbb{Q})$ of graded rings is an isomorphism in degrees up to $\frac{2g-2}{3}$.*

*Proof.* This follows from the existence of a homotopy equivalence

$$B\Gamma_{\infty,n+1} \sim B\Gamma_{\infty,n} \times \mathbb{CP}^\infty$$

as constructed in [29, Corollary 1.2] (see also [30, Theorem 4.3]). □

# 3 Stability and tautological classes: étale cohomology

We next translate the stability of cohomology from singular cohomology to compactly supported étale cohomology, and determine the effect of Frobenius on the stable cohomology classes, in order to use the Grothendieck-Lefschetz-Behrend trace formula.

Put

$$R_{n,\ell} := R_n \otimes_{\mathbb{Q}} \mathbb{Q}_\ell = \mathbb{Q}_\ell[\psi_1, \dots, \psi_n, \kappa_1, \kappa_2, \dots],$$

again graded by $\deg(\psi_i) = 2$ and $\deg(\kappa_j) = 2j$. Equip $R_{n,\ell}$ with a $\mathbb{Q}_\ell$-linear endomorphism Frob as follows:

$$\begin{cases} \operatorname{Frob} \psi_i & = q\psi_i \\ \operatorname{Frob} \kappa_j & = q^j \kappa_j. \end{cases}$$

Let $R^i_{n,\ell}$ denote the $i$th graded piece of the ring. For each $g, n$, we have a homomorphism of graded rings (with Frob action)

$$R^*_{n,\ell} \to H^*_{\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell) \tag{3}$$

again factoring through the Chow ring.

**Theorem 8.** *For $0 \leq i \leq \frac{2g-2}{3}$, the homomorphism in Equation* (3) *gives an isomorphism of Frobenius modules*

$$R^i_{n,\ell} \cong H^i_{\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell).$$

*Proof.* Let $0 \leq i \leq \frac{2g-2}{3}$. We consider $X = M_{g,n}$ as a scheme over $\operatorname{Spec} \mathbb{Z}_p$ with boundary divisor $D = \overline{M}_{g,n} - M_{g,n}$. Since the tautological classes arise from the Chow ring they are of Tate type, so the map (3) is Frobenius equivariant. We claim that there is a chain of functorial isomorphisms

$$R^i_{n,\ell} \to H^i_{\mathrm{et}}(X_{\mathbb{F}_p}, \mathbb{Q}_\ell) \cong H^i_{\mathrm{et}}(X_{\mathbb{Q}_p}, \mathbb{Q}_\ell) \cong H^i(X_{\mathbb{Q}_p}, \mathbb{Q}_\ell). \tag{4}$$

The first isomorphism follows from [1, Theorem 7.4] (since $\overline{M}_{g,n}$ is smooth and proper over $\mathbb{Z}_p$ and $D$ is a relative simple normal crossings divisor), and the second from [31, Theorem 11.6]. (We note that the two isomorphisms also hold for stacks – for the special case of a global quotient $[Y/G]$, the Hochschild-Serre spectral sequence [32, Theorem 2.20] gives $H^i_{\mathrm{et}}([Y/G]; \mathbb{Q}_\ell) = H^i_{\mathrm{et}}(Y; \mathbb{Q}_\ell)^G$ and the isomorphism follows. By [33], $M_{g,n}$ is a global quotient of a smooth scheme by a finite group.)

Each step in the formation of the tautological classes involves either pushing forward or pulling back cohomology classes, or formation of Chern classes (which by [32, Theorem 10.3] are characterized entirely by certain maps on cohomology). Since each map in Equation (4) is functorial, the tautological classes thus map to tautological classes. The composition is thus the map from Theorem 6; all maps but one are isomorphisms, so the remaining one is also an isomorphism and the claim follows.

□

**Corollary 9.** *For $0 \leq i \leq \frac{2g-2}{3}$, the following is true.*

*a. If $i$ is odd, then $H^{2\dim M_{g,n}-i}_{c,\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell) = 0$.*

*b. If $i$ is even, then $H^{2\dim M_{g,n}-i}_{c,\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell)$ has $\mathbb{Q}_\ell$-dimension equal to that of $R^i_{n,\ell}$, and* Frob *acts on it by multiplication by $q^{\dim M_{g,n}-i/2}$.*

*Proof.* Since $M_{g,n}$ is smooth, we may apply Poincaré duality for étale cohomology to deduce the claim from Theorem 8. (As in the proof of Theorem 8, we deduce duality for $M_{g,n}$ from duality for smooth schemes by a spectral sequence.) □

In our Grothendieck-Lefschetz trace computation, we will handle different

parts of the cohomology of $M_{g,n}$ in different ways. We thus define

$$\begin{aligned}
\mathbf{T}^{\text{stable}}_{g,n,q} &:= \sum_{0\le i\le \lfloor \frac{2g-2}{3}\rfloor} (-1)^i \operatorname{Tr}(\text{Frob}, H^{2\dim M_{g,n}-i}_{c,\text{et}}(M_{g,n},\mathbb{Q}_\ell)) \\
&= \sum_{0\le i\le \lfloor \frac{2g-2}{3}\rfloor} (-1)^i \operatorname{Tr}(\text{Frob}, R^{i}_{n,\ell}) \\
\mathbf{T}^{\text{unstable}}_{g,n,q} &:= \sum_{\lfloor \frac{2g-2}{3}\rfloor < i\le 2\dim M_{g,n}} (-1)^i \operatorname{Tr}(\text{Frob}, R^{2\dim M_{g,n}-i}_{c,\text{et}}(M_{g,n},\mathbb{Q}_\ell)) \\
\mathbf{N}_{g,n,q} &:= \sum_{\lfloor \frac{2g-2}{3}\rfloor < i\le 2\dim M_{g,n}} (-1)^i \operatorname{Tr}(\text{Frob}, B^{2\dim M_{g,n}-i}(M_{g,n},\mathbb{Q}_\ell)).
\end{aligned}$$

Note that, since these account for all of the cohomology of $M_{g,n}$, we have

$$\#M_{g,n}(\mathbb{F}_q) = \mathbf{T}^{\text{stable}}_{g,n,q} + \mathbf{T}^{\text{unstable}}_{g,n,q} + \mathbf{N}_{g,n,q}. \tag{5}$$

# 4 Heuristic 2 yields Conjecture 1

In this section, we prove Theorem 3. We first note that Heuristic 2 is equivalent to the assertion that

$$\lim_{g\to\infty} q^{-\dim M_{g,n}} \mathbf{N}_{g,n,q} = 0. \tag{6}$$

Thus, to prove Theorem 3, we need to understand the limiting behavior of $\mathbf{T}^{\text{stable}}_{g,n,q}$ and $\mathbf{T}^{\text{unstable}}_{g,n,q}$.

Let $R_n$ be the tautological ring as defined in Section 2. Note that the Hilbert series (or Poincaré series) $HS_{R_n}(z) := \sum_{i=0}^{\infty} \dim R_n^{2i} \cdot z^{2i}$ may be rewritten as

$$HS_{R_n}(z) = \prod_{i=1}^{n} \frac{1}{1-z^2} \prod_{j=1}^{\infty} \frac{1}{1-z^{2j}}.$$

**Lemma 10.** *We have the following:*

*a.* $\lim_{g\to\infty} q^{-\dim M_{g,n}} \mathbf{T}^{\text{stable}}_{g,n,q} = HS_{R_n}(q^{-1/2})$;

*b.* $\lim_{g\to\infty} q^{-\dim M_{g,n}} \mathbf{T}^{\text{unstable}}_{g,n,q} = 0$.

*Proof.* For the first statement, we compute:

$$\begin{aligned}
\lim_{g\to\infty} q^{-\dim M_{g,n}} \mathbf{T}^{\text{stable}}_{g,n,q} &= \lim_{g\to\infty} q^{-\dim M_{g,n}} \sum_{i=0}^{\lfloor \frac{2g-2}{3} \rfloor} (-1)^i \operatorname{Tr}(\text{Frob}, R^{2\dim M_{g,n}-i}_{c,\text{et}}(M_{g,n}, \mathbb{Q}_\ell)) \\
&= \lim_{g\to\infty} q^{-\dim M_{g,n}} \sum_{i=0}^{\lfloor \frac{2g-2}{3} \rfloor} (-1)^i \operatorname{Tr}(\text{Frob}, R^{i}_{n,\ell}) \\
&= \lim_{g\to\infty} \sum_{j=0}^{\lfloor \frac{g-1}{3} \rfloor} q^{-j} \cdot \dim R^{2j}_n \\
&= \sum_{j=0}^{\infty} q^{-j} \cdot \dim R^{2j}_n.
\end{aligned}$$

Using the Hilbert series of $R_n$ we may then rewrite the above sum as

$$\begin{aligned}
&= HS_{R_n}(q^{-1/2}) \\
&= \prod_{i=1}^{n} \frac{1}{1-q^{-1}} \prod_{j=1}^{\infty} \frac{1}{1-q^{-j}}.
\end{aligned}$$

Note that we use the fact (from Theorem 8) that for $0 \le i \le \frac{2g-2}{3}$, we have

$$R^i_{n,\ell} = R^i_n \otimes_{\mathbb{Q}} \mathbb{Q}_\ell \cong R^{2\dim M_{g,n}-i}_{c,\text{et}}(M_{g,n}, \mathbb{Q}_\ell)$$

For the second part, we let $P(z)$ be the generating function for the partition numbers $p(j)$, and let $Q_n(z) = \sum \binom{n+j-1}{j} z^j$ be the generating function whose $j$th coefficient is the number of multisets of size $j$ on $n$ elements. Then

$$HS_{R_n}(z) = Q_n(z^2)P(z^2).$$

In particular

$$\dim R^{2i}_n = \sum_{j=0}^{i} \binom{n+j-i}{j-1} p(i-j) \le \exp(c_n\sqrt{i}). \tag{7}$$

Since $R^*_{c,\text{et}}(M_{g,n}, \mathbb{Q}_\ell)$ is defined in terms of the image of a map from $R_n$ to cohomology, we further obtain

$$\dim R^{2\dim M_{g,n}-2i}_{c,\text{et}}(M_{g,n}, \mathbb{Q}_\ell) \le \exp(c_n\sqrt{i}). \tag{8}$$

Of course, when $i$ is odd this group is zero-dimensional.

We compute

$$\begin{aligned}
\lim_{g\to\infty} q^{-\dim M_{g,n}} \mathbf{T}^{\text{unstable}}_{g,n,q} &= \lim_{g\to\infty} q^{-\dim M_{g,n}} \sum_{i=\lfloor \frac{2g-2}{3}\rfloor+1}^{2\dim M_{g,n}} (-1)^i \operatorname{Tr}(\text{Frob}, R^{2\dim M_{g,n}-i}_{c,\text{et}}(M_{g,n},\mathbb{Q}_\ell)) \\
&\leq \lim_{g\to\infty} \sum_{i=\lfloor \frac{2g-2}{3}\rfloor+1}^{2\dim M_{g,n}} (-1)^i q^{-(i/2)} R_n^{2i} \\
&\leq \lim_{g\to\infty} \sum_{i=\lfloor \frac{2g-2}{3}\rfloor+1}^{2\dim M_{g,n}} (-1)^i q^{-(i/2)} \exp(c_n\sqrt{i}) \\
&= 0.
\end{aligned}$$

□

*Proof of Theorem 3.* By combining Equations (5) and (6) and Lemma 10 we get:

$$\begin{aligned}
\lim_{g\to\infty} \frac{\#M_{g,n}(\mathbb{F}_q)}{\#M_g(\mathbb{F}_q)} &= \lim_{g\to\infty} \frac{\mathbf{T}^{\text{stable}}_{g,n,q} + \mathbf{T}^{\text{unstable}}_{g,n,q} + \mathbf{N}_{g,n,q}}{\mathbf{T}^{\text{stable}}_{g,0,q} + \mathbf{T}^{\text{unstable}}_{g,0,q} + \mathbf{N}_{g,0,q}} \\
&= q^n \frac{HS_{R_n}(q^{-1/2}) + 0 + 0}{HS_R(q^{-1/2}) + 0 + 0} \\
&= q^n \prod_{i=1}^{n} \frac{1}{1-q^{-1}} \\
&= \lambda^n.
\end{aligned}$$

□

# 5 Proof of Theorem 4

In contrast to the previous sections, where $q$ was fixed, in this section we consider a case where $q$ and $g$ both go to infinity. We show that, so long as $q$ goes to infinity much faster than $g$, then we obtain an unconditional version of Conjecture 1.

The following lemma is the key result for this section, as it essentially shows that if $q \gg g$, then the main terms in the Grothendieck-Lefschetz trace computation will come from the stable cohomology range.

**Lemma 11.** *For any $K > 144$ and any nonnegative integer $n$, there exists a constant $K' = K'(n) > 0$ such that if $g > K'(n+1)$ and $q > g^K$, then*

$$\left|\mathbf{T}^{\text{unstable}}_{g,n,q} + \mathbf{N}_{g,n,q}\right| < q^{\dim M_{g,n} - g/6}.$$

*Proof.* We bound the total cohomology of $M_{g,n}$ by

$$\sum_i \dim H^i_{c,\mathrm{et}}(M_{g,n}, \mathbb{Q}_\ell) \leq (2+2g)^n (12g)!.$$

For $n = 0$, see [18, Lemma 5.1]. For general $n$, the bound follows from $n = 0$ by iteratively applying the Serre spectral sequence for $M_{g,i+1}$ over $M_{g,i}$.

In addition, we note that each cohomology group which arises in the calculation of $\mathbf{T}^{\mathrm{unstable}}_{g,n,q}$ and $\mathbf{N}_{g,n,q}$ is mixed of weight less than $2\dim M_{g,n} - \lfloor \frac{2g-2}{3} \rfloor$. We thus have

$$\left|\mathbf{T}^{\mathrm{unstable}}_{g,n,q} + \mathbf{N}_{g,n,q}\right| < q^{\dim M_{g,n} - \lfloor \frac{g-1}{3} \rfloor} (2g+2)^n (12g)!.$$

To ensure that this is at most $q^{\dim M_{g,n} - g/6}$, it suffices to take $q$ satisfying

$$\left( \lfloor \frac{g-1}{3} \rfloor - \frac{g}{6} \right) \log(q) > n \log(2g+2) + 12g \log(12g),$$

which would in turn follow from

$$\frac{1}{2} \left( \lfloor \frac{g-1}{3} \rfloor - \frac{g}{6} \right) \log(q) > \max\{n \log(2g+2), 12g \log(12g)\}.$$

Since $K > 144$, for any sufficiently small $\epsilon > 0$ we can choose $K'$ such that for $g > K'(n+1)$ and $q > g^K$,

$$\left( \lfloor \frac{g-1}{3} \rfloor - \frac{g}{6} \right) \log(q) > (1-\epsilon) \frac{g}{6}$$

and

$$\log(q) > \frac{144}{1-\epsilon} \log(12g).$$

This proves the claim. □

*Proof of Theorem 4.* We combine Equation (5) and Lemmas 10 and 11 to compute

$$\begin{aligned}
\lim_{g\to\infty} \frac{\#M_{g,n}(\mathbb{F}_q)}{\#M_g(\mathbb{F}_q)} &= \lim_{g\to\infty} \frac{\mathbf{T}^{\mathrm{stable}}_{g,n,q} + \mathbf{T}^{\mathrm{unstable}}_{g,n,q} + \mathbf{N}_{g,n,q}}{\mathbf{T}^{\mathrm{stable}}_{g,0,q} + \mathbf{T}^{\mathrm{unstable}}_{g,0,q} + \mathbf{N}_{g,0,q}} \\
&= q^n \frac{HS_{R_n}(q^{-1/2}) + O(q^{-g/6})}{HS_R(q^{-1/2}) + O(q^{-g/6})} \\
&= \lambda^n + O(q^{-g/6}).
\end{aligned}$$

□

## 6 Connections to random matrix models

Since much previous intuition about the behavior of random curves has come from the world of random matrix models, we would like to close with an invitation to the random matrix theory community to come up with evidence in favor of or opposed to Conjecture 1. Let us say a few words about how this might be possible.

When $q$ is large compared to $g$, one typically models the behavior of the zeta function of a random curve $C$ of genus $g$ over $\mathbb{F}_q$ by positing that the normalized characteristic polynomial of Frobenius behaves like that of a random matrix $M$ in the unitary symplectic group $\mathrm{USp}(2g)$. Equivalently, the sequence of point counts $\{\#C(\mathbb{F}_{q^n})\}_{n=1}^{\infty}$ has the same distribution as $\{q^n + 1 - q^{n/2}\operatorname{Tr}(M^n)\}_{n=1}^{\infty}$.

This model fails to apply in the case of fixed $q$ for three different reasons.

- **Discreteness:** For $n$ a positive integer, because $\#C(\mathbb{F}_{q^n}) \in \mathbb{Z}$, one must insist that $\operatorname{Tr}(M^n) \in q^{-n/2}\mathbb{Z}$.
- **Positivity:** Because $\#C(\mathbb{F}_q) \geq 0$, one must insist that $q+1-q^{1/2}\operatorname{Tr}(M) \geq 0$.
- **More positivity**: For $n_1, n_2$ two positive integers, because $\#C(\mathbb{F}_{q^{n_1 n_2}}) \geq \#C(\mathbb{F}_{q^{n_1}})$, one must insist that $q^{n_1 n_2} + 1 - q^{n_1 n_2/2}\operatorname{Tr}(M^{n_1 n_2}) \geq q^{n_1} + 1 - q^{n_1/2}\operatorname{Tr}(M^{n_1})$.

It seems unlikely that the statistics for such restricted random matrices can be computed in closed form, even in the limit as $g \to \infty$. However, it may be feasible to make numerical experiments for particular values of $q$ and $g$ to see how they compare to the predictions made by Conjecture 1.

## References


[1] Jordan S. Ellenberg, Akshay Venkatesh, and Craig Westerland. Homological stability for Hurwitz spaces and the Cohen-Lenstra conjecture over function fields, 2009, arXiv:0912.0325.

[2] C. Faber and R. Pandharipande. Tautological and non-tautological cohomology of the moduli space of curves. In *Handbook of moduli. Vol. I*, volume 24 of *Adv. Lect. Math. (ALM)*, pages 294–330. Int. Press, Somerville, MA, 2013.

[3] Wei Ho. How many rational points does a random curve have? *Bull. Amer. Math. Soc. (N.S.)*, 51(1):27–52, 2014.

[4] Pär Kurlberg and Zeév Rudnick. The fluctuations in the number of points on a hyperelliptic curve over a finite field. *J. Number Theory*, 129(3):580–587, 2009.

[5] Alina Bucur, Chantal David, Brooke Feigon, and Matilde Lalín. Statistics for traces of cyclic trigonal curves over finite fields. *Int. Math. Res. Not. IMRN*, (5):932–967, 2010.

[6] Melanie Matchett Wood. The distribution of the number of points on trigonal curves over $\mathbb{F}_q$. *Int. Math. Res. Not. IMRN*, (23):5444–5456, 2012.

[7] Alina Bucur, Chantal David, Brooke Feigon, and Matilde Lalín. Biased statistics for traces of cyclic $p$-fold covers over finite fields. In *WIN—women in numbers*, volume 60 of *Fields Inst. Commun.*, pages 121–143. Amer. Math. Soc., Providence, RI, 2011.

[8] Maosheng Xiong. The fluctuations in the number of points on a family of curves over a finite field. *J. Théor. Nombres Bordeaux*, 22(3):755–769, 2010.

[9] GilYoung Cheong, Melanie Matchett Wood, and Azeem Zaman. The distribution of points on superelliptic curves over finite fields, 2012, arXiv:1210.0456. to appear in *Proceedings of the American Mathematical Society*.

[10] Maosheng Xiong. Distribution of zeta zeroes for abelian covers of algebraic curves over a finite field, 2013, arXiv:1301.7124.

[11] Alina Bucur, Chantal David, Brooke Feigon, Matilde Lalín, and Kaneenika Sinha. Distribution of zeta zeroes of Artin-Schreier covers. *Math. Res. Lett.*, 19(6):1329–1356, 2012.

[12] Alexei Entin. On the distribution of zeroes of Artin-Schreier L-functions. *Geom. Funct. Anal.*, 22(5):1322–1360, 2012.

[13] Alina Bucur, Chantal David, Brooke Feigon, and Matilde Lalin. Statistics for ordinary artin-schreier covers and other $p$-rank strata, 2013, arXiv:1304.7876.

[14] Alina Bucur, Chantal David, Brooke Feigon, and Matilde Lalín. Fluctuations in the number of points on smooth plane curves over finite fields. *J. Number Theory*, 130(11):2528–2541, 2010.

[15] Alina Bucur and Kiran S. Kedlaya. The probability that a complete intersection is smooth. *J. Théor. Nombres Bordeaux*, 24(3):541–556, 2012.

[16] Daniel Erman and Melanie Matchett Wood. Semiample bertini theorems over finite fields, 2012, arXiv:1209.5266.

[17] Pär Kurlberg and Igor Wigman. Gaussian point count statistics for families of curves over a fixed finite field. *Int. Math. Res. Not. IMRN*, (10):2217–2229, 2011.

[18] Counting the number of curves over a finite field. available at `http://www.math.columbia.edu/~dejong/`.

[19] P. Deligne and D. Mumford. The irreducibility of the space of curves of given genus. *Inst. Hautes Études Sci. Publ. Math.*, (36):75–109, 1969.

[20] Kai A. Behrend. The Lefschetz trace formula for algebraic stacks. *Invent. Math.*, 112(1):127–149, 1993.

[21] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, second edition, 1986.

[22] J. Harer and D. Zagier. The Euler characteristic of the moduli space of curves. *Invent. Math.*, 85(3):457–485, 1986.

[23] Persi Diaconis and Mehrdad Shahshahani. On the eigenvalues of random matrices. *J. Appl. Probab.*, 31A:49–62, 1994. Studies in applied probability.

[24] Ravi Vakil. The moduli space of curves and its tautological ring. *Notices Amer. Math. Soc.*, 50(6):647–658, 2003.

[25] John L. Harer. Stability of the homology of the mapping class groups of orientable surfaces. *Ann. of Math. (2)*, 121(2):215–249, 1985.

[26] John L. Harer. Stability of the homology of the moduli spaces of Riemann surfaces with spin structure. *Math. Ann.*, 287(2):323–334, 1990.

[27] Nathalie Wahl. Homological stability for mapping class groups of surfaces. In *Handbook of moduli. Vol. III*, volume 26 of *Adv. Lect. Math. (ALM)*, pages 547–583. Int. Press, Somerville, MA, 2013.

[28] Ib Madsen and Michael Weiss. The stable moduli space of Riemann surfaces: Mumford's conjecture. *Ann. of Math. (2)*, 165(3):843–941, 2007.

[29] Carl-Friedrich Bödigheimer and Ulrike Tillmann. Stripping and splitting decorated mapping class groups. In Jaume Aguadé, Charles Broto, and Carles Casacuberta, editors, *Cohomological methods in homotopy theory*, volume 196 of *Progr. Math.*, pages 47–57. Birkhäuser, Basel, 2001.

[30] Ulrike Tillmann. Mumford's conjecture—a topological outlook. In *Handbook of moduli. Vol. III*, volume 26 of *Adv. Lect. Math. (ALM)*, pages 399–429. Int. Press, Somerville, MA, 2013.

[31] Eberhard Freitag and Reinhardt Kiehl. *Étale cohomology and the Weil conjecture*, volume 13 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1988. Translated from the German by Betty S. Waterhouse and William C. Waterhouse, With an historical introduction by J. A. Dieudonné.

[32] James S. Milne. *Étale cohomology*, volume 33 of *Princeton Mathematical Series*. Princeton University Press, Princeton, N.J., 1980.

[33] M. Boggi and M. Pikaart. Galois covers of moduli of curves. *Compositio Math.*, 120(2):171–191, 2000.

[34] Enrico Arbarello, Maurizio Cornalba, and Pillip A. Griffiths. *Geometry of algebraic curves. Volume II*, volume 268 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer, Heidelberg, 2011. With a contribution by Joseph Daniel Harris.

[35] Richard Hain and Eduard Looijenga. Mapping class groups and moduli spaces of curves. In *Algebraic geometry—Santa Cruz 1995*, volume 62 of *Proc. Sympos. Pure Math.*, pages 97–142. Amer. Math. Soc., Providence, RI, 1997.