EXPERIMENTAL MATHEMATICS
https://doi.org/10.1080/10586458.2026.2671126

Heuristics for (Ir)Reducibility of p-rank Strata of the Moduli Space of Hyperelliptic Curves

Thomas Bouchet[a], Erik Davis[b], Steven R. Groen[c], Zachary Porat[d], and Benjamin York[e]

[a]Laboratoire J.A. Dieudonné, Université Côte d'Azur, Nice, France; [b]Department of Mathematics, Texas A&M University, College Station, TX, USA; [c]Department of Mathematics, Lehigh University, Bethlehem, PA, USA; [d]Department of Mathematics and Computer Science, Wesleyan University, Middletown, CT, USA; [e]Department of Mathematics, University of Connecticut, Storrs, CT, USA

ABSTRACT

Let $\mathcal{H}_g$ denote the coarse moduli space of smooth hyperelliptic curves of genus g in characteristic $p \geq 3$, and let $\mathcal{H}_g^f$ denote the p-rank f stratum of $\mathcal{H}_g$ for $0 \leq f \leq g$. Achter and Pries note that determining the number of irreducible components of $\mathcal{H}_g^f$ would lead to several intriguing corollaries. In this paper, we present a computational approach for estimating the number of irreducible components in various p-rank strata. Our strategy involves sampling curves over finite fields and calculating their p-ranks. From the data gathered, we conjecture that the non-ordinary locus is geometrically irreducible for all genera $g > 1$. The data also leads us to conjecture that the moduli space $\mathcal{H}_g^{g-2}$ is irreducible and suggests that $\mathcal{H}_g^f$ is irreducible for all $1 \leq f \leq g$. We conclude with a brief discussion on $\mathcal{H}_g^0$.

KEYWORDS
arithmetic statistics; hyperelliptic curves; moduli space; positive characteristic; p-rank strata; non-ordinary locus

2020 MATHEMATICS SUBJECT CLASSIFICATION
14Q05 (primary); 14H10, 14G15, 14G17 (secondary)

1. Introduction

Throughout this paper, we consider hyperelliptic curves defined over a perfect field k of characteristic $p \geq 3$. Given a smooth curve C/k of genus g_C, we define the p-**rank** of C to be the integer f_C such that $\mathrm{Jac}(C)[p](\bar{k}) \cong (\mathbb{Z}/p\mathbb{Z})^{f_C}$. By [1], the p-rank of C equals the rank of the g^{th} semilinear power of the Frobenius operator on $\mathrm{H}^1(C, \mathcal{O}_C)$. In particular, we have $0 \leq f_C \leq g_C$.

Given a prime number $p \geq 3$, let $\mathcal{H}_g/\mathbb{F}_p$ be the coarse moduli space of smooth hyperelliptic curves of genus g, up to geometric isomorphism. For any $0 \leq f \leq g$, let $\mathcal{H}_g^f$ denote the p-**rank** f **stratum** of $\mathcal{H}_g$, i.e. the coarse moduli space of smooth hyperelliptic curves of genus g and p-rank at most f. The curves whose p-rank is less than g form the **non-ordinary locus** $\mathcal{H}_g^{g-1}$. In this paper, we are interested in the geometry of $\mathcal{H}_g^f$, in particular its number of irreducible components.

Let $\overline{\mathcal{H}}_g^f$ denote the p-rank f stratum of $\overline{\mathcal{H}}_g$, the closure of $\mathcal{H}_g$ inside $\overline{\mathcal{M}}_g$. The full moduli space $\overline{\mathcal{H}}_g = \overline{\mathcal{H}}_g^g$ is known to be irreducible of dimension $2g-1$. Moreover, [2, Proposition 2.1] shows that the p-rank strata $\overline{\mathcal{H}}_g^f$ are pure of codimension $g-f$ and [3, Lemma 3.2] explains that every component of $\overline{\mathcal{H}}_g^f$ intersects $\mathcal{H}_g^f$. Combining these facts, one sees that for any $0 \leq f \leq g$, the space $\overline{\mathcal{H}}_g^f$ is nonempty of dimension $g+f-1$ and the generic point of any component corresponds to a smooth hyperelliptic curve of genus g with p-rank f. Little is known about the number of geometrically irreducible components of $\overline{\mathcal{H}}_g^f$ for $f < g$ in general.

Let $\mathcal{A}_g^f$ be the coarse moduli space of principally polarized abelian varieties of dimension g and p-rank at most f, such that the Torelli morphism injects $\mathcal{H}_g^f$ into $\mathcal{A}_g^f$. By [4, Corollary 1.5], we know that the non-ordinary locus $\mathcal{A}_g^{g-1}$ is geometrically irreducible whenever $g > 1$. Furthermore, by [5, Remark 4.7], we know that $\mathcal{A}_g^f$ is geometrically irreducible for all $g \geq 3$ and all $0 \leq f \leq g$. (More precisely, $\mathcal{A}_g^f$ is the closure of a non-supersingular Newton stratum and it is shown in [5, Remark 4.7] that every non-supersingular Newton stratum is irreducible.) However, much less is known about the p-rank strata of $\mathcal{H}_g$.

In the $g = 1$ case, we have $\mathcal{H}_1 = \mathcal{M}_{1,1} \cong \mathcal{A}_1$ via the Torelli morphism. The p-rank 0 stratum $\mathcal{H}_1^0$ has dimension zero and the Deuring-Eichler mass formula (see [6, Theorem 4.1(c)])) gives the exact number of points of $\mathcal{H}_1^0$, which is roughly $\frac{p-1}{12}$. The $g = 2$ case is also well understood. Oort shows in [4, Corollary 1.5] that $\mathcal{A}_2^1$ is geometrically irreducible, and the geometric irreducibility of $\mathcal{H}_2^1$ follows from the fact that the generic point of $\mathcal{A}_2^1$ corresponds to the Jacobian of a smooth hyperelliptic curve. Finally, the number of geometric components of $\mathcal{H}_2^0$, given in [7, Remarks 2.16 and 2.17], is shown to go to infinity as p increases.

CONTACT Zachary Porat zporat@wesleyan.edu Department of Mathematics and Computer Science, Wesleyan University, Middletown, CT, USA.

For $g \geq 3$ and $f < g$, the number of geometric components of $\mathcal{H}_g^f$ is less accessible, since it is no longer true that the generic point of $\mathcal{A}_g^f$ corresponds to the Jacobian of a hyperelliptic curve. It is observed in [3, Section 3.7] that an understanding of the number of components of $\overline{\mathcal{H}}_g^f$ has interesting applications. More precisely, if $\overline{\mathcal{H}}_g^f$ has only one component, then in particular every component contains the moduli point of a chain of elliptic curves clutched together. (We say two curves are clutched together if a marked point of each is identified in an ordinary double point.) As a corollary, the closure of any component of $\mathcal{H}_g^f$ would intersect the space $\Delta_i[\mathcal{H}_g]^f$, consisting of singular curves that are obtained by clutching at a ramified point two hyperelliptic curves whose respective genera are i and $g-i$ and whose p-ranks add up to f.

In this paper, we investigate this open question from a computational perspective. More precisely, we gather data by sampling curves over finite fields and computing their p-ranks. Using the Lang-Weil bounds described in [8], these data strongly suggest conjectures about the number of components of $\mathcal{H}_g^f$.

Let $c_{g,f,p}$ be the number of geometrically irreducible components of $\mathcal{H}_g^f$ and let $c_{g,f,p,r}$ be the number of geometrically irreducible components of $\mathcal{H}_g^f$ that can be defined over $\mathbb{F}_{p^r}$. Recall that $\overline{\mathcal{H}}_g^f$ is pure of codimension $g-f$ and $\mathcal{H}_g^f$ is dense in $\overline{\mathcal{H}}_g^f$. Thus, if we take a sample S consisting of $\overline{\mathbb{F}}_p$-isomorphism classes of hyperelliptic curves of genus g over $\mathbb{F}_{p^r}$, then the Lang-Weil bounds prescribe the heuristic

$$\frac{\#\{[C] \in S \mid f_C \leq f\}}{\#S} \approx \frac{c_{g,f,p,r}}{p^{r(g-f)}}. \tag{1}$$

Here $[C]$ is the $\overline{\mathbb{F}}_p$-isomorphism class of a curve $C/\mathbb{F}_{p^r}$. This allows us to estimate $c_{g,f,p,r}$, which in turn provides evidence for the number of geometric components $c_{g,f,p}$ when r is varied. See Section 2.3 for a complete discussion of how we derive this heuristic.

In order to gather data on $c_{g,f,p,r}$, we use two different methods for generating isomorphism classes of hyperelliptic curves. One method is based on a family of pairwise non-isomorphic hyperelliptic curves over $\mathbb{F}_q$, which is relatively efficient compared to methods based on the invariants of the curves. The second method is based on an adaptation of [9], in which hyperelliptic curves over finite fields are enumerated. To make this second method more computationally viable, we restrict our attention to the special case where all branch points are defined over $\mathbb{F}_{p^r}$. The p-rank of a curve C is computed efficiently as the rank of the g^{th} semilinear power of a Hasse-Witt matrix of C, that is a matrix representing the semilinear action of Frobenius on $\mathrm{H}^1(C, \mathcal{O}_C)$ (see Section 2.2).

Since the sample size is limited by computational capacity, the estimate from Equation (1) is most precise for the non-ordinary locus $\mathcal{H}_g^{g-1}$. We use a large range of characteristics p, vary the exponent r, and run computations for genera g with $3 \leq g \leq 20$. This leads to the following conjecture, extending the known fact that $\mathcal{H}_2^1$ is geometrically irreducible.

Conjecture A (Conjecture 4.3). *For $g \geq 3$ and $p \geq 3$, the non-ordinary locus $\mathcal{H}_g^{g-1}$ is geometrically irreducible.*

Our methodology allows us to determine heuristics for lower p-rank strata as well. Since curves with smaller p-rank are rarer, the evidence is generally less strong than for the non-ordinary locus. Nevertheless, the data allows us to make the following conjecture for the moduli space $\mathcal{H}_g^{g-2}$.

Conjecture B (Conjecture 5.1). *For $g \geq 3$ and $p \geq 3$, the moduli space $\mathcal{H}_g^{g-2}$ is geometrically irreducible.*

Additionally, our data suggests that in general, this phenomenon remains the same for other p-rank strata. That is, the moduli space $\mathcal{H}_g^f$ is geometrically irreducible for $1 \leq f \leq g$ and $p \geq 3$. While the data is leading in this direction, we have insufficient evidence to make a concrete conjecture. Instead, we pose an open question to the community.

Question C (Question 5.2). *For $1 \leq f \leq g$ and $p \geq 3$, is $\mathcal{H}_g^f$ always geometrically irreducible?*

In contrast, our data on the p-rank 0 locus $\mathcal{H}_3^0$ suggests a different trend, namely that the number of components of $\mathcal{H}_3^0$ grows with p. This leads us to make a conjecture.

Conjecture D (Conjecture 5.3). *The number of components of $\mathcal{H}_3^0$ goes to infinity as p grows.*

Furthermore, since the observed behavior for $\mathcal{H}_3^0$ extends the established behavior of $\mathcal{H}_1^0$ and $\mathcal{H}_2^0$, we pose the following question.

Question E (Question 5.4). *For $g \geq 3$, does the number of geometric components of $\mathcal{H}_g^0$ go to infinity as p grows?*

Should Conjectures 4.3, 5.1 and 5.3 hold and Questions 5.2 and 5.4 be answered in the affirmative, we would have a complete understanding of the number of components of $\mathcal{H}_g^f$ for any g and f, generalizing the results in the literature for the specific cases when $g \leq 2$ or when $f = g$.

The overall structure of the paper is as follows. We start in Section 2 with a theoretical background on hyperelliptic curves, Hasse-Witt matrices and Lang-Weil bounds. In Section 3, the two methods for generating pairwise non-isomorphic hyperelliptic curves over $\mathbb{F}_q$ are outlined. Then we turn our attention to analyzing data for the non-ordinary locus $\mathcal{H}_g^{g-1}$ in Section 4, presenting strong evidence that this locus is geometrically irreducible for genera g with $3 \leq g \leq 20$ and $p \geq 3$. We conclude in Section 5 with a discussion of heuristics for other p-rank strata of $\mathcal{H}_g$.

Our work makes extensive use of the computer algebra system Magma [10]. The scripts written for and data collected throughout this project can be found in the GitHub repository [11].

2. Background

2.1. Hyperelliptic curves over $\mathbb{F}_q$

In this section, we will let p denote an odd prime, $r \geq 1$ be an integer, and $q = p^r$. For our purposes, a curve is a geometrically connected algebraic variety of dimension one. In particular, we are interested in hyperelliptic curves, which we define as follows.

Definition 2.1. Let k be a field not of characteristic 2. A **hyperelliptic curve** over k is a smooth, geometrically irreducible curve C/k of genus at least 2 that admits a double cover to $\mathbb{P}^1$.

Let $C/\mathbb{F}_q$ be a hyperelliptic curve defined over $\mathbb{F}_q$. Let g_C denote the genus of C. Then C has an affine model of the form $y^2 = f(x)$ where $f(x) \in \mathbb{F}_q[x]$ is a separable polynomial of degree $2g_C + 1$ or $2g_C + 2$.

In this article, we often identify an isomorphism between two hyperelliptic curves given by affine models with the induced automorphism of $\mathbb{P}^1$, and a scalar specifying the lift of this automorphism. More precisely, we use the following proposition.

Proposition 2.2. *([12, Prop. 1.10]). Let C_1 and C_2 be two hyperelliptic curves defined over k, where k is a field not of characteristic 2. Furthermore, let $y^2 = f_i(x)$ be an affine model of C_i. We assume that C_1 and C_2 are k-isomorphic. Then there exist $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(k)$ and $e \in k^\times$ such that the isomorphism is given by*

$$(x, y) \mapsto \left(\frac{ax + b}{cx + d}, \frac{ey}{(cx + d)^{g+1}} \right).$$

The pair (M, e) is unique up to replacement by $(\lambda M, \lambda^{g+1} e)$ for $\lambda \in k^\times$. In addition, the composition of (M, e) and (M', e') is given by $(M'M, e'e)$.

Given a hyperelliptic curve $C/\mathbb{F}_q$ with genus g_C, we let $\mathrm{Jac}(C)$ denote the Jacobian of C, an abelian variety of dimension g_C. A major invariant of interest is the p-rank f_C of C, that is the integer $0 \leq f_C \leq g_C$ such that $\#\mathrm{Jac}(C)[p](\overline{\mathbb{F}}_p) = p^{f_C}$, where $\mathrm{Jac}(C)[p]$ is the p-torsion group scheme of $\mathrm{Jac}(C)$. We say that a curve with p-rank $f_C = g_C$ is ordinary, and non-ordinary otherwise.

Fix an integer $g \geq 2$ and a prime $p \geq 3$. We denote by $\mathcal{H}_g/\mathbb{F}_p$ the coarse moduli space of smooth hyperelliptic curves of genus g up to geometric isomorphism. It is known that $\mathcal{H}_g$ is a geometrically irreducible quasi-projective scheme for any choice of g. For $0 \leq f \leq g$, we denote by $\mathcal{H}_g^f$ the coarse moduli space of smooth hyperelliptic curves of genus g and p-rank at most f, again up to geometric isomorphism. (Note that $\mathcal{H}_g^g = \mathcal{H}_g$.) In the case of $f = g - 1$, we refer to $\mathcal{H}_g^{g-1}$ as the non-ordinary locus of $\mathcal{H}_g$.

Our goal is to collect data on the p-ranks of hyperelliptic curves over $\mathbb{F}_q$ up to geometric isomorphism. In turn, this data allows us to draw conclusions primarily about the geometric structure of $\mathcal{H}_g^{g-1}$. Our method also enables us to investigate the spaces $\mathcal{H}_g^f$ for $f < g - 1$. While [2, Theorem 2.3] shows that $\overline{\mathcal{H}}_g^f$ is pure of codimension $g - f$ and that $\mathcal{H}_g^f$ sits densely inside $\overline{\mathcal{H}}_g^f$, we are particularly interested in determining the number of geometrically irreducible components of $\mathcal{H}_g^f$.

2.2. The Hasse-Witt matrix

The first step toward our goal involves sampling hyperelliptic curves over $\mathbb{F}_q$, which we detail in Section 3. Once we have a sample of hyperelliptic curves, we then need to calculate the p-rank for each curve. To compute the p-rank of a specific curve C, we can study a Hasse-Witt matrix of C.

Suppose that $C/\mathbb{F}_q$ is a hyperelliptic curve of genus g over $\mathbb{F}_q$, where $q = p^r$. A Hasse-Witt matrix of C is a $g \times g$-matrix corresponding to the action of the Frobenius operator on a given basis of $\mathrm{H}^1(C, \mathcal{O}_C)$. Let A denote a Hasse-Witt matrix of C, and let σ denote the action of the p^{th} power map on A, i.e. $A^\sigma = (a_{i,j}^p)$. It follows from the main result of [1] that the p-rank of C is equal to the rank of the g^{th} semilinear power of A, which can be written as

$$A^{(g)} = AA^\sigma \cdots A^{\sigma^{g-1}}.$$

Note that when C is defined over $\mathbb{F}_p$, we have $A^{(g)} = A^g$. To compute a Hasse-Witt matrix in practice, we use the following proposition.

Proposition 2.3. *([13], clarified by [14]). Let $C/\mathbb{F}_q$ be a hyperelliptic curve with affine model $y^2 = f(x)$, and define the constants $c_m \in \mathbb{F}_q$ by*

$$f(x)^{\frac{p-1}{2}} = \sum_m c_m x^m.$$

Then the $g \times g$-matrix $A = (a_{i,j})$ defined by $a_{i,j} = c_{jp-i}$ for all $0 \leq i, j \leq g$ is a Hasse-Witt matrix of C.

Proof. Let B be a $g \times g$-matrix corresponding to the action of the Verschiebung operator on a basis of $\mathrm{H}^0(C, \Omega^1_C)$, commonly called a Cartier-Manin matrix of C. If we choose the basis $\{x^{i-1}\frac{dx}{y} : 1 \leq i \leq g\}$ for $\mathrm{H}^0(C, \Omega^1_C)$, then we can compute a Cartier-Manin matrix $B = (b_{i,j})$ by writing

$$f(x)^{\frac{p-1}{2}} = \sum_m c_m x^m$$

and taking $b_{i,j} = c_{ip-j}^{1/p}$, as initially described in [13] and later clarified in [14, Section 3.1]. Since B is a Cartier-Manin matrix of C, $A = (B^\sigma)^T = (c_{jp-i})_{1 \leq i,j \leq g}$ is a Hasse-Witt matrix of C (see [14, Section 2.4]). □

2.3. Lang-Weil bounds

Having computed p-ranks for the curves in our sample, we next need to turn this data about p-ranks into information about the geometry of $\mathcal{H}_g^f$. In order to obtain heuristics for the number of components of $\mathcal{H}_g^f$, we make use of the Lang-Weil bounds.

Theorem 2.4. *([8, Cor. 2]). Let $\mathbb{F}_q$ be a finite field and let $V \subseteq \mathbb{P}^n$ be a geometrically irreducible variety of dimension d defined over $\mathbb{F}_q$. Denote by $V(\mathbb{F}_q)$ the set of $\mathbb{F}_q$-rational points of V. Then we have*

$$\#V(\mathbb{F}_{q^r}) = q^{rd} + O\left(q^{r(d-\frac{1}{2})}\right),$$

as $r \geq 1$ grows.

As a consequence, if $V/\mathbb{F}_q$ has $c_{V,\mathbb{F}_{q^r}}$ top-dimensional geometrically irreducible components that can be defined over $\mathbb{F}_{q^r}$, then we have

$$\#V(\mathbb{F}_{q^r}) = c_{V,\mathbb{F}_{q^r}} q^{rd} + O\left(q^{r(d-\frac{1}{2})}\right),$$

as $r \geq 1$ grows.

For the purposes of this paper, one may take $V = \overline{\mathcal{H}}_g^f$, which is defined over $\mathbb{F}_p$. Let $c_{g,f,p,r}$ be the number of irreducible components of $\mathcal{H}_g^f \times \mathrm{Spec}(\mathbb{F}_{p^r})$ that are geometrically irreducible, and let $c_{g,f,p}$ be the number of irreducible components of $\mathcal{H}_g^f \times \mathrm{Spec}(\overline{\mathbb{F}}_p)$. The fact that $\mathcal{H}_g^f$ is dense in $\overline{\mathcal{H}}_g^f$ (see [3, Lemma 3.2]) implies that one may interchange $\mathcal{H}_g^f$ with $\overline{\mathcal{H}}_g^f$ in the definitions of $c_{g,f,p,r}$ and $c_{g,f,p}$. Then since $\overline{\mathcal{H}}_g^f$ is pure of dimension $g + f - 1$, as $r \geq 1$ grows, we have

$$\#\overline{\mathcal{H}}_g^f(\mathbb{F}_{p^r}) = c_{g,f,p,r} p^{r(g+f-1)} + O\left(p^{r(g+f-1-\frac{1}{2})}\right).$$

Note that $\mathcal{H}_g^f$ is dense inside every component of $\overline{\mathcal{H}}_g^f$, so we find the same number of components if we restrict our search to smooth hyperelliptic curves. Recall also that the full space $\overline{\mathcal{H}}_g$ is geometrically irreducible. Given a sufficiently large sample S from $\mathcal{H}_g(\mathbb{F}_{p^r})$, it follows that

$$\frac{\#\left(S \cap \mathcal{H}_g^f\right)}{\#S} \approx \frac{\overline{\mathcal{H}}_g^f(\mathbb{F}_{p^r})}{\overline{\mathcal{H}}_g(\mathbb{F}_{p^r})} = \frac{c_{g,f,p,r} p^{r(g+f-1)} + O\left(p^{r(g+f-1-\frac{1}{2})}\right)}{p^{r(2g-1)} + O\left(p^{r(2g-1-\frac{1}{2})}\right)} \approx \frac{c_{g,f,p,r}}{p^{r(g-f)}}.$$

This heuristic immediately results in an approximation of $c_{g,f,p,r}$:

$$c_{g,f,p,r} \approx \frac{p^{r(g-f)} \#(S \cap \mathcal{H}_g^f)}{\#S}. \tag{2}$$

Fixing p, we note that the approximation becomes more accurate as the sample size grows and as r increases (i.e. as $q = p^r$ grows), since the relative Lang-Weil error is $O(q^{-1/2})$. Finally, we have

$$\max_{r \leq R} c_{g,f,p,r} = c_{g,f,p} \tag{3}$$

for R sufficiently large, since every geometric component can be defined over a finite field. Thus, this sampling method gives rather strong evidence for the number $c_{g,f,p}$, which is our primary interest.

Remark 2.5. We note that $c_{g,f,p,r}$ need not be equal to the number of components of $\mathcal{H}_g^f \times \mathrm{Spec}(\mathbb{F}_{p^r})$, since $\mathcal{H}_g^f \times \mathrm{Spec}(\mathbb{F}_{p^r})$ may have irreducible components that are not geometrically irreducible. Over an algebraic closure, such a component splits into multiple components that are Galois conjugate over $\mathbb{F}_{p^r}$. For the purposes of this paper, we are primarily interested in $c_{g,f,p}$ and Equation (3) shows that $c_{g,f,p}$ can be studied by studying $c_{g,f,p,r}$ for varying r. Finally, we observe that the number of components of $\mathcal{H}_g^f \times \mathrm{Spec}(\mathbb{F}_{p^r})$ is at least $c_{g,f,p,r}$ and at most $c_{g,f,p}$.

3. Computational methods

Let $g \geq 2$, p an odd prime, $r \geq 1$, and $q = p^r$. In the following sections, we detail the methods used to get our heuristics. We also explain how our sampling methods avoid introducing biases in the p-rank data.

3.1. Family method

For this section, we assume that p does not divide $2g+1$ (otherwise, we use the alternate method described in Section 3.2). Our first method for generating samples of hyperelliptic curves in $\mathcal{H}_g \times \mathrm{Spec}(\mathbb{F}_q)$ involves working with a quasi-affine subscheme $Y_g \subset \mathcal{H}_{g,1}$ whose $\mathbb{F}_q$-points are pairs (C, ∞), where C is a hyperelliptic curve admitting an affine model over $\mathbb{F}_q$ of the special form

$$y^2 = x^{2g+1} + \alpha_1 x^{2g-1} + \alpha_2 x^{2g-2} + \cdots, \tag{4}$$

where $\alpha_1 = \alpha_2 \neq 0$, and $\infty \in C$. Forgetting the marked point at ∞ yields a quasi-finite map $Y_g \to \mathcal{H}_g$, whose fibers can be easily computed algorithmically (see Proposition 3.4). In addition, since Y_g and $\mathcal{H}_g$ both have dimension $2g-1$ and $\mathcal{H}_g$ is irreducible, the image of Y_g under this forgetful map is Zariski-dense in $\mathcal{H}_g$. As a consequence, it is not contained in a closed proper subscheme of $\mathcal{H}_g$. Moreover, appealing to Proposition 2.3, we do not expect that imposing the condition $\alpha_1 = \alpha_2 \neq 0$ in Equation (4) has any effect on the p-rank of the curve. In conclusion, we do not expect that this sampling method introduces any bias. We then devise the following sampling algorithm:

(1) Sample a random subset S of $\mathcal{F}_{g,q} := Y_g(\mathbb{F}_q)$.
(2) For each $C \in S$, compute the subset $L_C \subset S$ consisting of curves that are geometrically isomorphic to C, and select the minimum of L_C, for a chosen ordering on $\mathbb{F}_q[x]$.
(3) Remove duplicates that appear.

This algorithm can be adapted to account only for $\mathbb{F}_q$-isomorphisms and quadratic twists (see Remark 3.5), and it is computationally much less costly. This modification is largely negligible in practice since, by [15, Theorem 3.7], the generic point of any component of $\mathcal{H}_g^f$ corresponds to a curve with automorphism group C_2 and such curves can only have one non-trivial twist by [16, Corollary 10].

For small values of g and q, our computational power allows us to take $S = \mathcal{F}_{g,q}$, but generally this is not feasible. The gathered data for this sampling method is analyzed in Section 4.1. The following proposition justifies that it is a natural choice to consider curves admitting a model of the form given in Equation (4).

Proposition 3.1. *([17, Prop. 1.2]). Let $C/\mathbb{F}_q$ be a hyperelliptic curve of genus g with an $\mathbb{F}_q$-rational branch point. Then C admits a model of the form $y^2 = f(x)$, where $f \in \mathbb{F}_q[x]$ is a monic polynomial of degree $2g+1$. Furthermore, if p does not divide $2g+1$, we can assume that*

$$f(x) = x^{2g+1} + \tilde{f}(x),$$

where $\tilde{f} \in \mathbb{F}_q[x]$ is of degree at most $2g-1$.

Proof. Since C has a rational branch point, it may be sent to infinity by a suitable change of coordinates. Thus, there exists $f \in \mathbb{F}_q[x]$ of degree $2g+1$ such that C is given by an affine model of the form $y^2 = f(x)$. After multiplying x and y by elements in $\mathbb{F}_q^\times$, we may assume that the coefficient of x^{2g+1} in f is 1. Furthermore, since p does not divide $2g+1$, a well-chosen linear change of variables allows us to eliminate the x^{2g} term. □

We define a pointed curve to be a pair (C, P) consisting of a curve C and a marked point P on C. We consider pointed hyperelliptic curves over $\mathbb{F}_q$ with their marked point at infinity. Any isomorphism between two such pointed curves must preserve the point at infinity, and isomorphisms of this form can be explicitly described using the following proposition.

Proposition 3.2. *([17, Prop. 1.2]). Let $(C_1, \infty)/\mathbb{F}_q$ and $(C_2, \infty)/\mathbb{F}_q$ be two pointed hyperelliptic curves defined by*

$$y^2 = x^{2g+1} + f_i(x)$$

where $\deg(f_i) \leq 2g-1$. Then (C_1, ∞) and (C_2, ∞) are geometrically isomorphic if and only if there exists $u \in \overline{\mathbb{F}}_q$ such that

$$f_2(x) = u^{2g+1} f_1\left(\frac{x}{u}\right).$$

To avoid the complication presented by this scaling action, we are led to consider the subset of pointed hyperelliptic curves with marked points at infinity, where each curve has an affine model of the form

$$y^2 = x^{2g+1} + \alpha_1 x^{2g-1} + \alpha_2 x^{2g-2} + \cdots,$$

such that $\alpha_1 = \alpha_2 \neq 0$. We denote by Y_g the quasi-affine $\mathrm{Spec}(\mathbb{F}_p)$-subscheme of $\mathcal{H}_{g,1}$ consisting of curves with a model of the form given in Equation (4). Then we let $\mathcal{F}_{g,q} := Y_g(\mathbb{F}_q)$ be its set of $\mathbb{F}_q$-points.

Corollary 3.3. *With the same notation as above, let $(C_1, \infty), (C_2, \infty) \in \mathcal{F}_{g,q}$ be defined by f_1 and f_2 respectively, such that $f_1 \neq f_2$. Then these pointed hyperelliptic curves are not geometrically isomorphic.*

Now, we wish to study the forgetful map $Y_g \to \mathcal{H}_g$ in order to separate the curves in $\mathcal{F}_{g,q}$ by their geometric isomorphism classes as non-pointed curves.

Proposition 3.4. *Let $(C, \infty) \in \mathcal{F}_{g,q}$. There are at most $2g + 2$ curves $(C', \infty) \in \mathcal{F}_{g,q}$ such that C is geometrically isomorphic to C'. We provide an algorithm to compute the collection of all such curves.*

Proof. Let $(C, \infty), (C', \infty) \in \mathcal{F}_{g,q}$, given by affine equations $y^2 = h(x)$ and $y^2 = h'(x)$, respectively. We show that to any isomorphism $\varphi : C' \to C$, one can canonically associate a branch point of C. This correspondence bounds the number of curves in $\mathcal{F}_{g,q}$ isomorphic to C by $2g + 2$.

Indeed, the isomorphism φ realizes a bijection between the branch points of C' and those of C. Let r be the branch point of C that is the image under φ of the point at ∞ on C'. Now, consider the curve C'' obtained from C by a change of coordinates that sends r to ∞. The curve C'' then admits an affine model of the form $y^2 = \alpha x^{2g+1} + \ldots$, with $\alpha \neq 0$. After scaling x and y by suitable constants, and eliminating the term in x^{2g} (which we can do since p does not divide $2g + 1$ by assumption), we obtain a model of C'' of the form

$$y^2 = x^{2g+1} + f(x),$$

where $f \in k[x]$ has degree at most $2g - 1$.

By construction, the composition of these coordinate changes yields an isomorphism from C' to C'', that preserves the branch point at infinity. Therefore, (C', ∞) and (C'', ∞) are isomorphic as pointed curves, thus by Proposition 3.2, they must differ only by a scaling of x and y. We therefore conclude that for each branch point r of C, there exists at most one curve C_r in $\mathcal{F}_{g,q}$ such that $C_r \simeq C$, so there can only be at most $2g + 2$ of them.

This observation leads to a straightforward procedure for computing all curves in $\mathcal{F}_{g,q}$ that are isomorphic to a given curve $(C, \infty) \in \mathcal{F}_{g,q}$. First, compute the set of roots of the polynomial h defining C via $y^2 = h(x)$. Then, for each root r, determine a change of coordinates that sends r to ∞, and normalize the resulting equation to the form

$$y^2 = x^{2g+1} + f(x),$$

where f is of degree at most $2g - 1$. If the coefficients of either x^{2g-1} or x^{2g-2} in f are zero, the corresponding curve does not belong to $\mathcal{F}_{g,q}$ and hence we ignore the result. Otherwise, use Proposition 3.2 to make the coefficients in x^{2g-1} and x^{2g-2} equal. Finally, return only those curves whose models have coefficients in $\mathbb{F}_q$. □

Remark 3.5. In theory, Proposition 3.4 allows us to distinguish geometric isomorphism classes of curves. In practice, we consider only the $\mathbb{F}_q$-rational branch points of the curve C. This allows us to account for $\mathbb{F}_q$-isomorphisms and quadratic twists, which makes the algorithm more efficient and does not affect the quality of the data by [15, Theorem 3.7] and [16, Corollary 10].

3.2. Galois type method

We now turn our attention to an alternate method of sampling hyperelliptic curves to use in the cases when the family method is not applicable and as reinforcement of the family method results. In [9], Howe describes an algorithm to enumerate all hyperelliptic curves over $\mathbb{F}_q$ up to $\mathbb{F}_q$-isomorphism for a specified genus. He does so by considering each possible Galois type over $\mathbb{F}_q$, where the **Galois type** of a hyperelliptic curve is the sorted tuple of degrees of the irreducible factors of the defining polynomial f in an affine equation $y^2 = f(x)$. We adapted Algorithm 7.1 in [9] to compute the set of geometrically non-isomorphic curves of Galois type $(1, \ldots, 1)$ over $\mathbb{F}_q$, or equivalently, of hyperelliptic curves over $\mathbb{F}_q$ with models of the form $y^2 = f(x)$, where

(1) the polynomial f splits completely over $\mathbb{F}_q$, and
(2) the curves are unique up to geometric isomorphism.

Proposition 3.6 provides that curves of Galois type $(1, \ldots, 1)$ have no twists other than the quadratic twists, which are easy to exclude from our sample. For that reason, we did not need to make many changes to Howe's algorithm to ensure that the curves we consider are geometrically non-isomorphic.

We now give a geometric argument for sampling curves of Galois type $(1, \ldots, 1)$ in order to obtain heuristics for $c_{g,f,p}$. For $g \geq 2$, let X_g be the open subscheme of $(\mathbb{P}^1)^{2g+2}$ whose coordinates are all pairwise distinct. Consider the map $\pi : X_g \to \mathcal{H}_g$ that sends a point $(a_1, \ldots, a_{2g+2})$ to the hyperelliptic curve admitting a double cover to $\mathbb{P}^1$ branched along the divisor $\sum_{i=1}^{2g+2} a_i$. Note that the geometric isomorphism classes of hyperelliptic curves defined over $\mathbb{F}_{p^r}$ that are of Galois type $(1, \ldots, 1)$ correspond precisely to $\pi(X_g(\mathbb{F}_{p^r}))$. Now, since π is surjective on geometric points, i.e. $\pi(X_g(\overline{\mathbb{F}}_p)) = \mathcal{H}_g(\overline{\mathbb{F}}_p)$, one may expect that sampling from $\pi(X_g(\mathbb{F}_{p^r}))$ for varying r gives heuristics that are representative for all of $\mathcal{H}_g$.

The algorithm goes as follows:

(1) Take $a_1 = 0, a_2 = \infty$, and loop over values of distinct $0 < a_3 < \ldots < a_{2g+2} \in \mathbb{F}_q^\times$ (for a chosen order on $\mathbb{F}_q^\times$) with $\sum_{i=3}^{2g+2} a_i = 1$. The a_i's correspond to the branch points of a curve C with Galois type $(1, \ldots, 1)$, and for every such curve there are multiple possible values for the a_i's.
(2) To account for the different possible values for the a_i's, consider all matrices in $\mathrm{PGL}_2(\mathbb{F}_q)$ that send two of the a_i's to 0 and ∞, and keep the sum of the rest of the other branch points to 1.
(3) Sort each set of roots obtained, and take the minimal set of roots with respect to the lexicographical order.
   (a) If the minimum appears several times, the corresponding curve has extra automorphisms, and we discard it.
   (b) If the minimum is the original set of roots, then add the curve $y^2 = x\prod_{i=3}^{2g+2}(x - a_i)$ to the list of curves. Note that to ensure geometric isomorphism, the quadratic twist is not added to that list.
   (c) If not, this set of branch points has been added already.
(4) If the desired number of curves has been reached, or all possible values of a_i's have been tested, return the list of curves. Otherwise, continue.

Unlike the algorithm in [9], we exclude hyperelliptic curves with extra automorphisms. We do so for two reasons. First, excluding these curves was less computationally expensive than including them (we can see that a configuration of roots is reached several times without having to loop over all of the possible elements in $\mathrm{PGL}_2(\mathbb{F}_q)$), and only slight modifications to the original code from [9] were necessary to exclude them. Second, the exclusion of curves with extra automorphisms does not result in a diluted data set. Indeed, by [15, Theorem 3.7], the generic point of any geometrically irreducible component of $\mathcal{H}_g^f$ corresponds to a curve whose automorphism group is C_2. Thus, the curves with extra automorphisms do not significantly affect our data, and can be disregarded.

We also note that this method does not produce a random sample of hyperelliptic curves, as we return a sample of size s whose roots are the smallest possible under the ordering on $\mathbb{F}_q$. However, we see no relationship between the size of the roots under this choice of ordering and the p-rank strata of the corresponding hyperelliptic curves (see Proposition 2.3), so we do not believe this biases our data toward particular p-rank strata. See Section 4.2 for the data generated via this method.

The following proposition helps us distinguish the curves of Galois type $(1, \ldots, 1)$ up to geometric isomorphism.

Proposition 3.6. *Let C and C' be two geometrically isomorphic hyperelliptic curves defined over $\mathbb{F}_q$ such that all of their branch points are defined over $\mathbb{F}_q$. Then C and C' are either $\mathbb{F}_q$-isomorphic, or quadratic twists of each other.*

Proof. Let $\psi : C \to C'$ be a (geometric) isomorphism. After choosing affine models

$$C : y^2 = f(x), \qquad C' : y^2 = f'(x),$$

with $f, f' \in \mathbb{F}_q[x]$, we identify ψ with the induced automorphism of $\mathbb{P}^1$, represented by a Möbius transformation $M \in \mathrm{PGL}_2(\overline{\mathbb{F}}_p)$, together with a scalar $e \in \overline{\mathbb{F}}_p^\times$ accounting for the choice of quadratic twist (see Proposition 2.2). Since ψ maps any branch point of C to a branch point of C', M must send the $2g+2$ branch points of C to those of C', seen as subsets of $\mathbb{P}^1(\mathbb{F}_q)$. It is known that a Möbius transformation is uniquely determined by specifying the image of three distinct points; therefore, since the images of $2g + 2 \geq 3$ points lie in $\mathbb{P}^1(\mathbb{F}_q)$, the transformation M has coefficients in $\mathbb{F}_q$. Applying this transformation to the defining equation of C yields the equation

$$e^2 y^2 = f''(x),$$

with f'' having the same roots as the polynomial f'. Thus, the polynomials f' and f'' differ only by a scaling element in $\mathbb{F}_q$, so that $e^2 \in \mathbb{F}_q^\times$, and hence $e \in \mathbb{F}_{q^2}^\times$. It follows that either ψ is a $\mathbb{F}_q$-isomorphism, or that C' is a quadratic twist of C. □

Remark 3.7. Suppose two curves over $\mathbb{F}_q$ of Galois type $(1, \ldots, 1)$ returned by our algorithm are quadratic twists of each other. The proof of Proposition 3.6 shows that their sets of roots must lie in the same $\mathrm{PGL}_2(\mathbb{F}_q)$-orbit. This is not possible, because we only pick one curve for each of these orbits. In fact, we remove the step in the algorithm in [9] that appends quadratic twists to the list.

3.3. Computation of p-ranks

Having collected a sample of hyperelliptic curves, we next need to compute the p-ranks of these curves. In practice, we have implemented the following algorithm to compute the p-ranks of hyperelliptic curves using Hasse-Witt matrices. See Section 2.2 for more details.

INPUT: The polynomial $f(x) \in \mathbb{F}_q[x]$ and prime power $q = p^r$.
OUTPUT: The p -rank of $C : y^2 = f(x)$.

(1) Compute $f(x)^{\frac{p-1}{2}} = \sum_m c_m x^m$ and $g = \lfloor (\deg(f) - 1)/2 \rfloor$.
(2) Define the Hasse-Witt matrix $A = (a_{i,j})$ by setting $a_{i,j} = c_{jp-i}$ for all $0 \leq i, j \leq g$.
(3) If $q = p^1$, return the rank of A^g, otherwise continue to step (4).
(4) If $q = p^r$ for $r > 1$, return the rank of $A^{(g)} = AA^\sigma \cdots A^{\sigma^{g-1}}$.

Table 1. Analysis of the data collected using the family sampling method for $g = 3$.

Field	Prime range	Sample size	Median	Mean	Min.	Max.	Std. dev.
$\mathbb{F}_p$	5–1000	10000000∗	1.012	1.016	0.982	1.092	0.020
$\mathbb{F}_{p^2}$	3–300	10000000	1.006	1.009	0.877	1.161	0.045
$\mathbb{F}_{p^3}$	3–100	100000000	0.995	0.985	0.794	1.092	0.059
$\mathbb{F}_{p^4}$	3–30	100000000	1.002	1.005	0.976	1.054	0.026

Table 2. Analysis of the data collected using the family sampling method for $g = 4$.

Field	Prime Range	Sample Size	Median	Mean	Min.	Max.	Std. Dev.
$\mathbb{F}_p$	5–1000	10000000∗	1.001	1.001	0.979	1.027	0.007
$\mathbb{F}_{p^2}$	5–300	10000000	1.001	0.998	0.831	1.118	0.049
$\mathbb{F}_{p^3}$	5–100	100000000	1.001	0.999	0.912	1.058	0.037
$\mathbb{F}_{p^4}$	5–30	100000000	1.001	0.998	0.954	1.049	0.032

4. Non-ordinary locus

In the following section, we provide an overview of the data collected regarding the number of components of the non-ordinary locus. Sections 4.1 and 4.2 provide an analysis of the data collected using the sampling methods described in Sections 3.1 and 3.2, respectively. Then we present our main conjecture based on these findings in Section 4.3.

The analysis for both sampling methods followed the same general structure. First, a sample of curves over $\mathbb{F}_q$ was generated using one of the two algorithms previously discussed; we will denote the number of curves in this sample by s. Then the p-rank for each curve in the sample was calculated using the method described in Section 2.2. With the p-rank data in hand, we could next count the number of non-ordinary curves, which we will denote by N.

Using the above information, we computed

$$M_{g,p,r} = \frac{N}{s} \cdot p^r$$

for various p and appropriate values of r. (We give explicit ranges for p and values for r in the following subsections.) Recall that our goal is to determine the number of geometrically irreducible components of the non-ordinary locus, $c_{g,g-1,p} \in \mathbb{Z}_{>0}$. As outlined in Section 2.3, we expect the ratio of non-ordinary curves to sample curves to be proportional to $c_{g,g-1,p,r}$, as defined in Equation (2). Therefore, by computing $M_{g,p,r}$, we can approximate $c_{g,g-1,p,r}$, which in turn allows us to find

$$c_{g,g-1,p} = \max_{r \le R} c_{g,g-1,p,r} \approx \max_{r \le R} M_{g,p,r}$$

for sufficiently large R.

We then calculated several statistics on the data sets consisting of values of $M_{g,p,r}$ for every prime p in a certain range of primes. We determined the median and mean of these data sets to estimate the number of geometrically irreducible components. We also found the minimum, maximum, and standard deviation of the data sets to emphasize the uniformity of the data.

4.1. Family method data analysis

This section details the data collected using the sampling method described in Section 3.1. Table 1 gives the data for the genus $g = 3$ case and Table 2 for the $g = 4$ case. We were also able to use this method to collect data for the cases of $5 \le g \le 20$. The data collected for these genera indicate similar phenomena to those observed in the $g = 3$ and $g = 4$ cases, so we use these small genera as examples. The complete data set can be found in [11].

Let $F_{g,q}$ denote the set of $\mathbb{F}_q$-isomorphism classes of hyperelliptic curves C such that $(C, \infty) \in \mathcal{F}_{g,q}$ (see Section 3.1 for notation). We note that for $r = 1$ and genus $g \in \{3, 4, 5\}$, we were able to collect data for all isomorphism classes in $F_{g,p}$ for a small number of primes p, specifically in the cases of

- $g = 3$, for $p = 5$ and $11 \le p \le 47$,
- $g = 4$, for $5 \le p \le 17$, and
- $g = 5$, for $3 \le p \le 7$.

The sample sizes in these cases are noted with an asterisk (∗), as the number of curves sampled exceeded the sample size listed for certain p. In particular,

- for $g = 3$, $\#F_{3,p} > s = 10\,000\,000$ for $29 \le p \le 47$;
- for $g = 4$, $\#F_{4,p} > s = 10\,000\,000$ for $11 \le p \le 17$; and

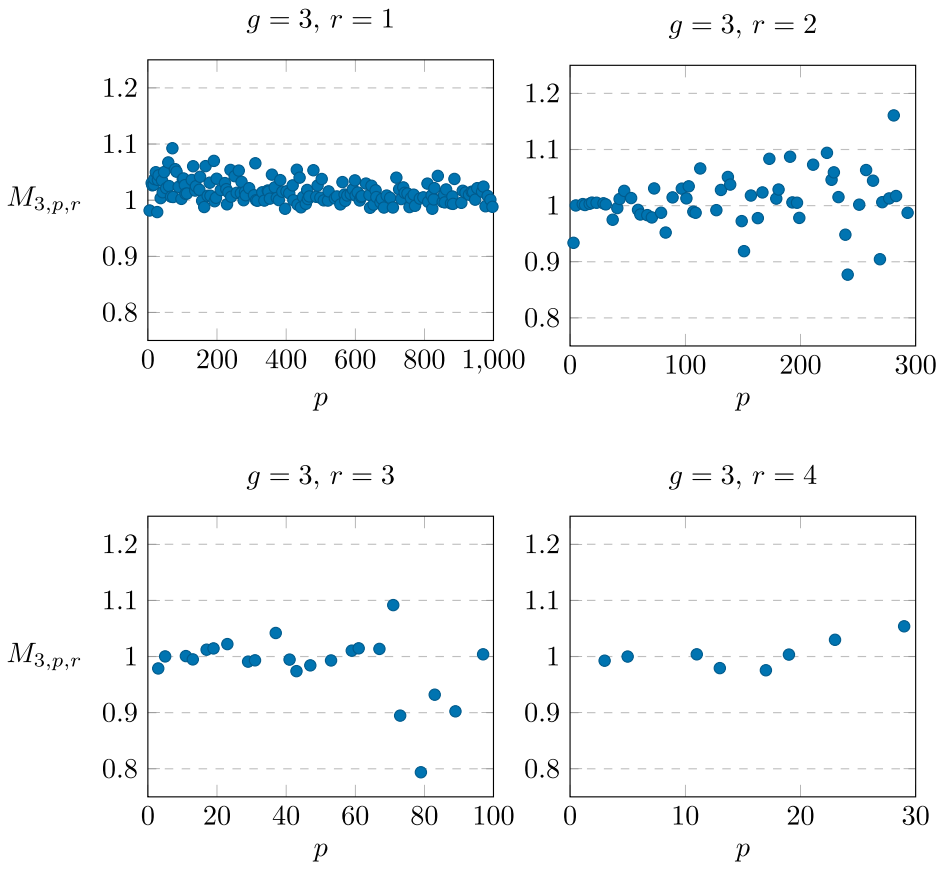

Figure 1. Values of $M_{3,p,r}$ for each prime p.

- for $g = 5$, $\#F_{5,p} > s = 10\,000\,000$ for $p = 7$.

Note that for small p, the number of curves in the sample cannot always reach the intended sample size, e.g. for $g = 4$ and $p = 5$, $\#F_{4,5} = 38\,556$. Therefore, s is the upper bound of the sample size, used for the majority of the primes p, except in the aforementioned cases.

To further support our data analysis, Figure 1 illustrates the values of $M_{3,p,r}$ that we calculated for each prime p and the appropriate value of r. We note that for the majority of data points, the percent error, i.e. the difference between the calculated M-value and the expected value of $c_{g,g-1,p} = 1$, is less than 5%.

Remark 4.1. To corroborate the data we saw for small p when $r \leq 4$, we computed samples of curves for small values of p and large values of r when $g \in \{3, 4, 5\}$, and computed their p-ranks.

For ease of notation, define

$$r_p := \left\lfloor \frac{\log(100000)}{\log(p)} \right\rfloor = \begin{cases} 10 & \text{if } p = 3 \\ 7 & \text{if } p = 5 \\ 5 & \text{if } p = 7 \end{cases}$$

to be the maximum value of r such that $p^r \leq 100\,000$. (This upper bound was chosen as, in practice, we were able to find a sufficient number of non-ordinary curves to make reasonable conclusions when p^r was below this threshold.)

For the cases of

- $g = 3$, $p \in \{3, 5\}$, and $2 \leq r \leq r_p$,
- $g = 4$, $p \in \{5, 7\}$, and $2 \leq r \leq r_p$, and
- $g = 5$, $p \in \{3, 5, 7\}$, and $3 \leq r \leq r_p$.

we computed the p-ranks of samples of up to $10,000,000$ curves, and found similar statistics for $M_{g,p,r}$ as those shared above. Data for these computations can be found in [11].

Table 3. Analysis of the data collected using the Galois type sampling method for $g = 3$.

Field	Prime range	Sample size	Median	Mean	Min.	Max.	Std. dev.
$\mathbb{F}_{p^2}$	17–300	10000000	1.006	1.006	0.858	1.174	0.056
$\mathbb{F}_{p^3}$	7–100	100000000	1.000	0.999	0.945	1.059	0.028
$\mathbb{F}_{p^4}$	5–30	100000000	0.998	1.002	0.982	1.033	0.019

Table 4. Analysis of the data collected using the Galois type sampling method for $g = 4$.

Field	Prime Range	Sample Size	Median	Mean	Min.	Max.	Std. dev.
$\mathbb{F}_{p^2}$	11–300	10000000	1.008	1.009	0.896	1.228	0.053
$\mathbb{F}_{p^3}$	7–100	100000000	1.000	0.997	0.941	1.060	0.026
$\mathbb{F}_{p^4}$	5–30	100000000	1.003	0.998	0.960	1.022	0.019

4.2. Galois type method data analysis

Tables 3 and 4 communicate the data collected using the Galois type method discussed in Section 3.2 for the cases $g = 3$ and $g = 4$, respectively. We also used this method in the $g = 5$ case and found similar statistics; the data can be found in [11]. The Galois type method is more computationally expensive than the family method, so while this method could be extended to higher genera, we primarily used it to reinforce the results from the family method and expect similar behavior in higher genera.

We note that for this data, the lower bound of the prime range is the first prime at which the full sample size is reached. Therefore, every sample has exactly the size listed. As in the previous case, we report the median, mean, minimum, maximum, and standard deviation of the data sets consisting of values of $M_{g,p,r}$ for every prime p in the prime range for a given field of definition. We also remind the reader that this method does not produce a random sample of hyperelliptic curves, as we return a sample whose roots are the smallest possible under the ordering on $\mathbb{F}_q$. As discussed in Section 3.2, we do not believe this biases our data.

Remark 4.2. When the Galois type method is applied over $\mathbb{F}_p$, the results exhibit unexpected behavior. When considering curves defined over $\mathbb{F}_p$, we checked samples of sizes up to `10,000,000` when $g = 3$ and up to `1,000,000` when $g \in \{4, 5\}$ for the prime ranges

- $17 \leq p < 1000$ when $g = 3$,
- $17 \leq p < 660$ when $g = 4$, and
- $17 \leq p < 410$ when $g = 5$.

When $p \equiv 34$, we found an abundance of non-ordinary curves whose branch points are defined over $\mathbb{F}_p$. In this case, the values of $M_{g,p,1}$ sit between 1.5 and 4, but are not particularly close to any integer. However, we found a lack of non-ordinary curves when $p \equiv 14$. In these cases $M_{g,p,1}$ is often close to 0, if not equal to 0.

We note that this behavior disappears when the Galois type method is applied over $\mathbb{F}_{p^r}$ for $r > 1$, as in the data above. Additionally, when we instead consider all the Galois types over $\mathbb{F}_p$ for $3 \leq p \leq 47$, there is again no difference between $p \equiv 14$ and $p \equiv 34$; in both cases, the data suggest irreducibility of the non-ordinary locus.

We have no explanation presently as to why this phenomenon occurs. Moreover, our queries failed to produce any hyperelliptic curves of p-rank 0 whose branch points were all defined over $\mathbb{F}_p$ when $p \equiv 14$ and $g \in \{3, 5\}$, even after sampling up to `1,000,000,000` curves for primes $5 \leq p \leq 200$ when $g = 3$ and $5 \leq p \leq 80$ when $g = 5$.

4.3. Main conjecture

Recall that $c_{g,f,p,r}$ is defined as the number of irreducible components of $\mathcal{H}_g^f \times \mathrm{Spec}(\mathbb{F}_{p^r})$ that are also geometrically irreducible. The data presented in Section 4.1 and [11] provides strong evidence that $c_{g,g-1,p,r} = 1$ in the following cases:

- $3 \leq g \leq 5$, $r = 1$ and $5 \leq p < 1000$;
- $3 \leq g \leq 5$, $r = 2$ and $3 \leq p < 300$;
- $3 \leq g \leq 5$, $r = 3$ and $3 \leq p < 100$;
- $3 \leq g \leq 5$, $r = 4$ and $3 \leq p < 30$;
- $6 \leq g \leq 20$, $r = 1$ and $5 \leq p < 500$;
- $6 \leq g \leq 20$, $r = 2$ and $3 \leq p < 100$.
- $3 \leq g \leq 5$, $3 \leq p \leq 7$ and $r \leq \lfloor \log(100000)/\log(p) \rfloor$.

Additionally, the data presented in Section 4.2 and [11] provides strong evidence that $c_{g,g-1,p,r} = 1$ in the following cases:

- $3 \leq g \leq 5$, $r = 2$ and $17 \leq p < 300$;

Table 5. Statistics for data sets of $M_{g,g-2,p,1}$ for $3 \leq g \leq 20$ collected using the family method.

Genus	Prime range	Sample size	Median	Mean	Min.	Max.	Std. dev.
3	5–1000	10000000*	0.992	0.974	0.294	1.643	0.195
4	5–1000	10000000*	1.014	0.999	0.381	1.604	0.182
5	3–1000	10000000*	1.010	0.992	0.432	1.704	0.175
6	5–500	10000000	1.036	1.031	0.219	2.062	0.287
7	7–500	10000000	1.030	1.012	0.202	1.688	0.244
8	3–500	10000000	1.009	1.035	0.425	1.930	0.272
9	3–500	10000000	1.046	1.066	0.303	2.313	0.304
10	5–500	10000000	1.007	0.962	0.429	1.867	0.270
11	3–500	10000000	1.037	0.990	0.322	1.913	0.266
12	3–500	10000000	1.005	1.020	0.294	2.062	0.256
13	5–500	10000000	0.987	1.030	0.213	2.118	0.313
14	3–500	10000000	1.044	1.018	0.000	1.756	0.276
15	3–500	10000000	1.043	1.051	0.269	1.745	0.251
16	5–500	10000000	1.034	1.004	0.144	1.766	0.308
17	3–500	10000000	0.998	0.965	0.229	1.809	0.266
18	3–500	10000000	0.991	0.999	0.201	1.816	0.235
19	5–500	10000000	0.997	0.996	0.229	1.913	0.250
20	3–500	10000000	1.042	1.020	0.218	2.372	0.274

- $3 \leq g \leq 5$, $r = 3$ and $7 \leq p < 100$;
- $3 \leq g \leq 5$, $r = 4$ and $5 \leq p < 30$.

Note that when p^r is small, there do not exist enough curves over $\mathbb{F}_{p^r}$ to give a reliable estimate. On the other hand, when p^r is large compared to the sample size, the non-ordinary curves are too rare to give a reliable estimate. If $c_{g,g-1,p,r}$ would grow with g, p or r, one would expect that such a phenomenon should be visible in the samples presented in this section.

We suspect that the behavior does not change when $g > 20$, when $r > 4$, or when $p > 1000$. For comparison, the reducibility of $\mathcal{H}_1^0$ becomes apparent when $p^r = 5^2$ (see [6, Theorem 4.1(c)]) and the reducibility of $\mathcal{H}_2^0$ becomes apparent already when $p^r = 7^2$ (see [7, Remark 2.17]). In summary, given the behavior for smaller genera, we are confident in making the following conjecture.

Conjecture 4.3. *For $g \geq 3$ and $p \geq 3$, the non-ordinary locus $\mathcal{H}_g^{g-1}$ is geometrically irreducible.*

Recall that $\mathcal{A}_2^1$ is geometrically irreducible by [4, Corollary 1.5], and the geometric irreducibility of $\mathcal{H}_2^1$ follows from the fact that the generic point of $\mathcal{A}_2^1$ corresponds to the Jacobian of a smooth hyperelliptic curve. Thus, this conjecture is a generalization of the known result for $g = 2$.

5. Other *p*-rank strata

5.1. *p*-rank at least 1

A major benefit of our computations is that we not only count curves of p-rank $g-1$, but of all p-ranks $f \leq g-1$. Consequently, we were able to extend our study to other p-rank strata. The method described in Section 3.1 was also applied to gather evidence for the number of components of $\mathcal{H}_g^{g-2}$ for $3 \leq g \leq 20$. This yielded heuristics that are very similar to the heuristics in Section 4.1. Additionally, the Galois type method described in Section 3.2 was applied for $g = 3$. Again by [15, Theorem 3.7], we are able to exclude curves with extra automorphisms without affecting the quality of the data.

To count the number of curves with p-rank $f \leq g-2$, we employed a similar method to that described at the start of Section 4. By abuse of notation, we identify S with the corresponding subset of $\mathcal{H}_g(\mathbb{F}_{p^r})$. Since $\mathcal{H}_g^{g-2}$ has codimension 2, for each sample $S \subseteq \mathcal{H}_g(\mathbb{F}_{p^r})$, we computed the ratio

$$M_{g,g-2,p,r} = \frac{\#(S \cap \mathcal{H}_g^{g-2})}{\#S} p^{2r}.$$

We then calculated the median, mean, minimum, maximum, and standard deviation on the data sets consisting of values of $M_{g,g-2,p,r}$ for every prime p in a certain prime range for a given field of definition.

Table 5 illustrates the above statistics for $M_{g,g-2,p,1}$ for $3 \leq g \leq 20$ collected using the family method. As mentioned above, we also computed similar statistics using the Galois type method. These statistics corroborated the results determined using the family method and so are not included in the table; they can be found in full in [11]. Similar to Tables 1 and 2, we note that since $r = 1$, we were able to collect data for the entire family of curves for a small number of primes p. The sample sizes in these cases are once again noted with an asterisk (∗) to indicate that for certain p, the entire family exceeded the sample size listed.

Table 5 suggests that $\mathcal{H}_g^{g-2}$ is geometrically irreducible for the considered values of g and p. Additional data for $r = 2$ is available in [11] and exhibits similar behavior to the $r = 1$ case for the smaller prime ranges we could compute. We note that the minima are

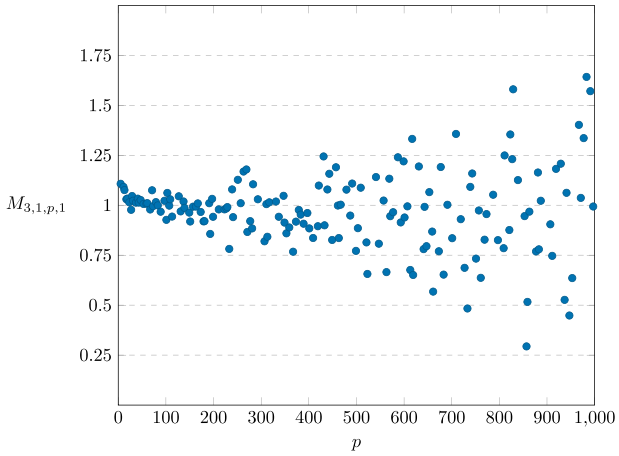

Figure 2. Values of $M_{3,1,p,1}$ for each prime p collected using the family method illustrating how the spread widens as p grows.

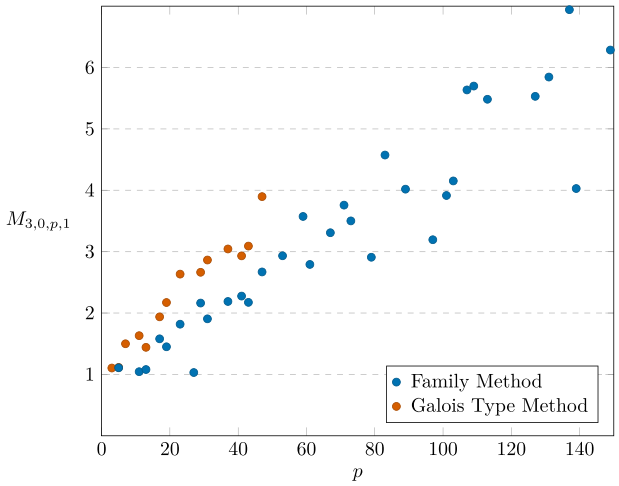

Figure 3. Values of $M_{3,0,p,1}$ for each prime p, collected using both sampling methods.

smaller and the maxima are larger in this data than in the non-ordinary data. We attribute this to the rarity of curves with desired p-rank.

Figure 2 demonstrates how the error of $M_{3,1,p,1}$ behaves. The spread of the values of $M_{3,1,p,1}$ widens as p grows, simply because the curves of p-rank at most $g-2$ become increasingly rare compared to the size of the dataset. For example, in the $g=3$ case, there are 1001 curves with p-rank $f \leq 1$ when $p=103$, but only 10 such curves when $p=883$. Given this explanation along with the data available, we feel confident making the following conjecture.

Conjecture 5.1. *For $g \geq 3$ and $p \geq 3$, the moduli space $\mathcal{H}_g^{g-2}$ is geometrically irreducible.*

Since curves of even smaller p-rank are very rare, it is difficult to obtain reliable heuristics for the number of components of $\mathcal{H}_g^f$ when $f < g-2$. For $3 \leq p < 100$, our data indicates that $c_{4,1,p,1} = c_{5,2,p,1} = 1$, though it is computationally infeasible to obtain statistically significant evidence. The following question aims to generalize both known theoretical results and our computational heuristics.

Question 5.2. *For $1 \leq f \leq g$ and $p \geq 3$, is $\mathcal{H}_g^f$ always geometrically irreducible?*

The answer to this question is known to be affirmative when $f=g$ or when $(f,g)=(1,2)$. The heuristics we provide in this paper form strong evidence that the answer is also affirmative when $g-f$ is small.

5.2. p-Rank 0

We conclude with a brief discussion about the p-rank 0 locus. Recall, in the $g=1$ case, the Deuring-Eichler mass formula gives the exact number of points of $\mathcal{H}_1^0$, which is roughly $\frac{p-1}{12}$. That is, the number of points grows with p. Moreover, in the $g=2$ case, [7, Remarks 2.16 and 2.17] show that the number of geometric components of $\mathcal{H}_2^0$ goes to infinity as p increases. Figure 3 illustrates our

computational findings, showing the values of $M_{3,0,p,1}$ for each prime p, collected using both sampling methods. The trend of the data indicates that the number components of $\mathcal{H}_3^0 \times \mathrm{Spec}(\overline{\mathbb{F}}_p)$ is also growing with p.

While the values of $M_{3,0,p,1}$ are growing with p, curiously, they are not concentrated around integers. We note that a similar phenomenon occurs when running experiments on genus 2 curves, possible thanks to [9]. That is, we once again see non-integral values when $r = 1$ for those curves. However, since there are fewer curves than in higher genera, it is possible to enumerate them over $\mathbb{F}_{p^2}$ for small primes p. What results are values of $M_{2,0,p,2}$ that are much closer to integers. We have no explanation as to why this phenomenon occurs when $r = 1$, but given that we see similar behavior in smaller genera, we feel confident in the validity of our techniques. In summary, our data warrants the following conjecture.

Conjecture 5.3. *The number of geometric components of $\mathcal{H}_3^0$ goes to infinity as p grows.*

As previously noted, this conjecture generalizes the theoretically known behavior for $\mathcal{H}_1^0$ and $\mathcal{H}_2^0$. An interesting question to ask is whether this phenomenon persists for $g \geq 4$; however, we do not currently know how to further the investigation with our present methods and implementations, as the computations become prohibitive in the ranges we would need to test.

Question 5.4. *For $g \geq 4$, does the number of geometric components of $\mathcal{H}_g^0$ go to infinity as p grows?*

Acknowledgments

This project originated at the 2024 Arizona Winter School on Abelian Varieties. The authors would like to thank the organizers of the Arizona Winter School for providing the conditions for this project to begin. In particular, the authors are deeply grateful to Rachel Pries for suggesting this line of inquiry. We further thank Shiruo Wang for his contributions to the early stages of this project. The authors would like to thank Everett Howe for his insights and support of the project. We would finally like to thank our reviewers for their invaluable comments.

Author contributions

CRediT: **Thomas Bouchet**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing; **Erik Davis**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing; **Steven R. Groen**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing; **Zachary Porat**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing; **Benjamin York**: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Writing – original draft, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

[1] Oda, T. (1969). The first de Rham cohomology group and Dieudonné modules. *Ann. Sci. École Norm. Sup.* 2(1): 63–135. DOI: 10.24033/asens.1175
[2] Glass, D, Pries, R. (2005). Hyperelliptic curves with prescribed p-torsion. *Manuscripta Math.* 117(3): 299–317. DOI: 10.1007/s00229-005-0559-0
[3] Achter, J. D., Pries, R. (2011). The p-rank strata of the moduli space of hyperelliptic curves. *Adv. Math.* 227(5): 1846–1872. DOI: 10.1016/j.aim.2011.04.004
[4] Oort, F. (2001). A stratification of a moduli space of abelian varieties. *Mod. Abelian Variet.* 195: 345–416. DOI: 10.1007/978-3-0348-8303-0$_1$3
[5] Chai, C.-L. (2005). Monodromy of Hecke-invariant subvarieties. *Pure Appl. Math. Q.* 1(2): 291–303. DOI: 10.4310/PAMQ.2005.v1.n2.a4
[6] Silverman, J. H. (2009). The arithmetic of elliptic curves. 2nd ed. In: *Graduate Texts in Mathematics*. New York, NY: Springer.
[7] Ibukiyama, T., Katsura, T, Oort, F. (1986). Supersingular curves of genus two and class numbers. *Compositio. Math.* 57(2): 127–152.
[8] Lang, S, Weil, A. (1954). Number of points of varieties in finite fields. *Amer. J. Math.* 76(4): 819–827. DOI: 10.2307/2372655
[9] Howe, E. W. (2025). Enumerating hyperelliptic curves over finite fields in quasilinear time. *Res. number Theory.* 11(1), 26. DOI: 10.1007/s40993-024-00594-7
[10] Bosma W., Cannon J. J., Fieker C., Steel A. (eds.). (2010). *Handbook of Magma functions,* Version 2.28-14, Edition 2.16, 5017. https://magma.maths.usyd.edu.au/magma/handbook/.
[11] Bouchet, T., Davis, E., Groen, S. R., Porat, Z, York, B. (2026). GitHub repository for related Magma scripts and resulting data sets. Available at: https://github.com/Thittho/Non-ordinary-locus.
[12] Lercier, R, Ritzenthaler, C. (2012). Hyperelliptic curves and their invariants: Geometric, arithmetic and algorithmic aspects. *J. Algebra* 372: 595–636. DOI: 10.1016/j.jalgebra.2012.07.054
[13] Yui, N. (1978). On the Jacobian varieties of hyperelliptic curves over fields of characteristic p > 2. *J. Algebra* 52(2):378–410. DOI: 10.1016/0021-8693(78)90247-8

[14] Achter, J. D., Howe, E. W. (2019). Hasse-Witt and Cartier-Manin matrices: A warning and a request. *Arithmetic Geometry: Computation and Applications* 722: 1–18. DOI: 10.1090/conm/722/14534
[15] Achter, J. D., Glass, D, Pries, R. (2008). Curves of given p-rank with trivial automorphism group. *Michigan Math. J.* 56(3): 583–592. DOI: 10.1307/mmj/1231770361
[16] Meagher, S, Top, J. (2010). Twists of genus three curves over finite fields. *Finite Fields Appl.* 16(5): 347–368. DOI: 10.1016/j.ffa.2010.06.001
[17] Lockhart, P. (1994). On the discriminant of a hyperelliptic curve. *Trans. Amer. Math. Soc.* 342(2): 729–752. DOI: 10.2307/2154650