GAUSSIAN POINT COUNT STATISTICS FOR FAMILIES OF CURVES OVER A FIXED FINITE FIELD

PÄR KURLBERG AND IGOR WIGMAN

ABSTRACT. We produce a collection of families of curves, whose point count statistics over $\mathbb{F}_p$ becomes Gaussian for p *fixed*. In particular, the *average* number of $\mathbb{F}_p$-points on curves in these families tends to infinity.

1. INTRODUCTION

The purpose of this note is to exhibit a collection of families of smooth curves whose normalized limiting point count statistics, over a *fixed* finite field $\mathbb{F}_p$, have a Gaussian distribution. Given a finite family $\mathcal{F}_i$ of smooth curves defined over $\mathbb{F}_p$, let $M_i := \frac{1}{|\mathcal{F}_i|}\sum_{C\in\mathcal{F}_i}|C(\mathbb{F}_p)|$ be the average number of $\mathbb{F}_p$-points on the curves $C\in\mathcal{F}_i$ and let $V_i := \frac{1}{|\mathcal{F}_i|}\sum_{C\in\mathcal{F}_i}\left(|C(\mathbb{F}_p)|-M_i\right)^2$ be the variance of the fluctuations in these point counts. Here, and in what follows, $C(\mathbb{F}_p)$ will denote the set of $\mathbb{F}_p$-points on C, and $|C(\mathbb{F}_p)|$ its cardinality. We can now formulate the main result of this note.

Theorem 1. *There exists a sequence of families $\{\mathcal{F}_i\}_{i=1}^{\infty}$ of smooth curves defined over $\mathbb{F}_p$ with the following properties: $|\mathcal{F}_i|, M_i, V_i$ all tend to infinity, and, for all compact intervals I,*

$$\frac{1}{|\mathcal{F}_i|}\left|\left\{C\in\mathcal{F}_i : \frac{|C(\mathbb{F}_p)|-M_i}{V_i^{1/2}}\in I\right\}\right| = \frac{1}{\sqrt{2\pi}}\int_I e^{-x^2/2}dx + o(1),$$

as $i\to\infty$.

To obtain such a sequence we intersect projective surfaces $X_i\subset\mathbb{P}^{n_i}$, chosen so that $|X_i(\mathbb{F}_p)|$ tends to infinity as $i\to\infty$, with families of large degree hypersurfaces. More precisely, let $S_i(d)$ be the set of degree d homogeneous polynomials in n_i+1 variables with coefficients in $\mathbb{F}_p$. For

Date: July 8, 2010.

P.K. was partially supported by grants from the Knut and Alice Wallenberg foundation, the Royal Swedish Academy of Sciences, and the Swedish Research Council. I.W. was supported by grant KAW 2005.0098 from the Knut and Alice Wallenberg Foundation.

$f \in S_i(d)$, let $H_f \subset \mathbb{P}^{n_i}$ be the hypersurface defined by the zero set of f. Intersecting X_i with H_f we generically obtain a (possibly singular) curve $C_i(f) := X_i \cap H_f$. Letting

$$\tilde{S}_i(d) := \{f \in S_i(d) : X_i \cap H_f \text{ is smooth}\},$$

we obtain a family of *smooth* curves

$$\mathcal{F}_i(d) := \{C_i(f) : f \in \tilde{S}_i(d)\}.$$

Our main technical result (in essence a slightly more explicit version of Poonen's [16, Theorem 1.2]) asserts that the distribution of point counts for curves in this family, for d large, is binomial — we can think of it as the number of successes when an unfair coin is tossed $|X_i(\mathbb{F}_p)|$ times.

Proposition 2. *Let $X_i \subset \mathbb{P}^{n_i}$ be a smooth projective surface defined over $\mathbb{F}_p$, and let $\mathcal{F}_i(d)$ be defined as above. Then, as $d \to \infty$,*

$$(1)\quad \frac{|\{C \in \mathcal{F}_i(d) : |C(\mathbb{F}_p)| = s\}|}{|\mathcal{F}_i(d)|}$$
$$= \binom{|X_i(\mathbb{F}_p)|}{s} \left(\frac{p+1}{p^2+p+1}\right)^s \left(1 - \frac{p+1}{p^2+p+1}\right)^{|X_i(\mathbb{F}_p)|-s} \cdot (1+o(1))$$

uniformly for $0 \le s \le |X_i(\mathbb{F}_p)|$. In particular, the average point count of a curve $C \in \mathcal{F}_i(d)$ equals $|X_i(\mathbb{F}_p)| \cdot \frac{(p+1)}{p^2+p+1} \cdot (1+o(1))$ as $d \to \infty$.

Given Proposition 2, we can easily obtain Theorem 1 by the central limit theorem type argument for coin flip models (cf. Section 2) *provided* we can find a sequence of surfaces $\{X_i\}_{i\ge 1}$ such that $|X_i(\mathbb{F}_p)| \to \infty$. Any such sequence suffices, but for concreteness we will use Ihara's construction [10] (independently discovered by Tsfasman, Vlăduţ, and Zink [18]), of families of *curves* with many points over $\mathbb{F}_{p^2}$, and a (Weil) restriction of scalars argument will then produce a projectively embedded $\mathbb{F}_p$-surface having many points — see Section 2.2 for more details.

Remark: Since the relation $|C(\mathbb{F}_p)| = p + 1 - \mathrm{Tr}(\mathrm{Frob}\,|H^1_c(\overline{C}, \overline{\mathbb{Q}}_l))$ holds between the number of $\mathbb{F}_p$-points on a curve C and the trace of Frobenius (see e.g [11, Ch. 11]), we have in fact exhibited families of smooth curves for which the average of the trace of Frobenius exhibits a strong negative bias. (Note that the average trace of Frobenius should be zero according to random matrix theory predictions; cf. Section 1.1.1.)

1.1. **Background and discussion.**

1.1.1. *Gaussian point count statistics in other models.* The number of $\mathbb{F}_p$-points on a smooth curve C of genus g, defined over a finite field $\mathbb{F}_p$, can be written as

$$|C(\mathbb{F}_p)| = p + 1 - \sum_{i=1}^{2g} \alpha_i$$

where $\alpha_i \in \bar{\mathbb{Q}}$ are the eigenvalues of the Frobenious action on a certain cohomology group. By Weil's proof of the Riemann hypothesis for curves, $|\alpha_i| = \sqrt{p}$ for $1 \leq i \leq 2g$. Hence $|C(\mathbb{F}_p)| = p+1-p^{1/2} \cdot \mathrm{Tr}(U_C)$, where $U_C \in U(2g)$ is a unitary $2g \times 2g$ matrix, unique up to conjugacy. If we let C range over a family $\mathcal{F}(\mathbb{F}_p)$ of smooth curves defined over $\mathbb{F}_p$ (e.g., the family of hyperelliptic curves $\mathcal{F} = \{C_t\}_{t \in \mathbb{F}_p : f(t) \neq 0}$ where $C_t = \{(x, y) : y^2 = f(x)(x - t)\}$ and $f \in \mathbb{F}_p[x]$ is a square free polynomial of degree $2g$) it is natural to study the distribution of the fluctations of the points counts by looking at the normalized fluctuations, i.e., the quantity $(|C(\mathbb{F}_p)| - p - 1)/\sqrt{p} = -\mathrm{Tr}(U_C)$.

By Deligne's equidistribution theorem, the distribution of the conjugacy classes of the U_C's, as $p \to \infty$, are given by random matrix theory when the family has "large monodromy". For example, Katz and Sarnak [13] has shown that for the family of hyperelliptics given above, the limiting distribution on the U_C-conjugacy classes, as $p \to \infty$, is given by the Haar measure on $USp(2g)$, the group of unitary symplectic $2g \times 2g$ matrices. On the other hand, in the limit $g \to \infty$, Diaconis and Shahshahani has shown [6] that the limiting distribution of $\{\mathrm{Tr}(U)\}_{U \in USp(2g)}$, as $g \to \infty$, is a *Gaussian* with mean zero and variance one[1]. Thus, if a collection of families of curves have large monodromies, then the (normalized) point count fluctuations are Gaussian in the double limit $\lim_{g\to\infty} \lim_{p\to\infty}$.

For p fixed and $g \to \infty$ it is less clear what to expect regarding the distribution of the U_C-conjugacy classes and their traces. For instance, random matrix theory (RMT) is clearly not an appropriate model since the inequality $0 \leq |C(\mathbb{F}_p)| = p + 1 - \sqrt{p} \cdot \mathrm{Tr}(U_C)$ does not hold for all $U_C \in USp(2g)$ for g sufficiently large[2]. Similarly, if the curves in the families can be embedded into $\mathbb{P}^n(\mathbb{F}_p)$ for n fixed, the bounds

[1]It is rather remarkable that even though $\mathrm{Tr}(U)$ is a sum of $2g$ complex numbers on the unit circle, the variance does not scale as $\sqrt{g}$, but is in fact identically equal to one.

[2]However, it is worth noting that certain statistics of the eigenvalues are consistent with RMT, e.g., the fluctuations of the number of eigenvalues in random short intervals (cf. [7]). Moreover, in [17] Rudnick found that the one-level density,

$0 \le |C(\mathbb{F}_p)| = p+1-\sqrt{p}\cdot \mathrm{Tr}(U_C) \le |\mathbb{P}^n(\mathbb{F}_p)|$ rules out a Gaussian. In fact the normalized distribution cannot even have continuous support since $|C_f(\mathbb{F}_p)|$ is integer valued and the variance is bounded.

To get some insight into the large genus limit while keeping the ground field fixed, Kurlberg and Rudnick studied (cf. [14]) families of hyperelliptic curves of the form $C_f : y^2 = f(x)$, where f ranges over monic square free polynomials of degree d. They found that the fluctuations in this family, as $d \to \infty$, has the same distribution as $\sum_{i=1}^{p} x_i$, where $x_1, \dots, x_p$ are independent random variables taking the values $0, -1, 1$ with probabilities $1/(p+1), 1/2(1+p^{-1}), 1/2(1+p^{-1})$, respectively. Moreover, the moments for the normalized point count distribution were shown to be Gaussian as long as both $p, d \to \infty$, i.e., any joint limit rather than letting p tend to infinity first (as described above.)

In [3], Bucur, David, Feigon and Lalín generalized this to cyclic l-fold covers of $\mathbb{P}^1$; here the distribution of the fluctuations are given by $2\,\mathrm{Re}(\sum_{i=1}^{p} x_i)$ where $x_1, \dots, x_p$ are independent random variables taking the values $0, \exp(2\pi i/l), \dots, \exp(2\pi i(l-1)/l)$ with probabilities $2/(p+2), p/(3(p+2)), \dots, p/(3(p+2))$, respectively. Further, in [2], they studied the family of smooth curves in the projective plane cut out by degree d homogenous polynomials, and found that the distribution of the point count statistics for this family, as $d \to \infty$, is the same as that of $\sum_{i=1}^{p^2+p+1} x_i$, where $x_1, \dots, x_{p^2+p+1}$ are independent random variables taking the values $0, 1$ with probabilities $p^2/(p^2+p+1), (p+1)/(p^2+p+1)$, respectively. In both [3, 2], Gaussian moments were obtained when p tends to infinity; in the first case along any limit $p, d \to \infty$, whereas the assumption $d \gg p^{1+\epsilon}$ is needed in the second case.

Using slightly different methods, namely character sum estimates, Xiong [19] recently obtained similar results for families of curves of the form $y^l = f(x)$, where f ranges either over all l-th power free polynomials of degree d, or over all monic irreducible polynomials of degree d (as $d \to \infty$.)

We finally note that in [15] (unpublished), Larsen obtained Gaussian moments for a smooth family of hyperelliptic curves of the form $Y^2 = \prod_{i=1}^{n}(X - a_i)$, where $a_1, \dots, a_n \in \mathbb{F}_q$, ranges over distinct elements.

1.1.2. *Remarks on vanishing probabilities.* Given a point $P \in X_i(\mathbb{F}_p)$, the probability of a polynomial $f \in S_i(d)$ vanishing at P is $1/p$, so one might expect that the average number of $\mathbb{F}_p$-points on $X_i \cap H_f$ should equal $|X_i(\mathbb{F}_p)|/p$. However, as we have seen, this prediction is

a local statistics, was in agreement with RMT (even though averages of traces of small powers was not.)

not quite correct — by conditioning on f so that $X_i \cap H_f$ is smooth, the probability of f vanishing at P turns out to be slightly smaller than expected, and is given by $\frac{p+1}{p^2+p+1}$ (rather than by $1/p$). A similar phenomenon was already observed in [2] for smooth plane curves: the probability of point $P \in \mathbb{P}^2(\mathbb{F}_p)$ belonging to a smooth curve given by the zero set of a random homogenous polynomial (of large degree) is $(p+1)/(p^2+p+1)$, rather than $1/p$.

1.1.3. *Remarks on families of curves with many points.* In order to obtain Gaussian (normalized) point count statistics, it is essential that there is no a priori upper bound on the number of $\mathbb{F}_p$-points on the curves; in particular, families of plane curves, families with bounded genus, or families with bounded gonality cannot be used. In fact, something even stronger is needed: since point counts are integer valued, the variance must grow to infinity for the normalized distribution to have continuous support. Thus, since the limiting distribution must be symmetric around the mean, together with the fact that the number of points on a curve is non-negative, the *average* number of $\mathbb{F}_p$-points of a curve in the family also must tend to infinity.

A natural candidate for families of curves with unbounded point counts over $\mathbb{F}_p$ is $M_g(\mathbb{F}_p)$, the set of isomorphism classes of genus g curves. However, *if* these point counts can be modeled by random matrix theory (e.g., as in the case of g fixed and $p \to \infty$ as shown in [13]), the average number of points would be $p+1$ since the average trace of Frobenius equals zero by random matrix theory predictions; hence it is unclear whether using $M_g(\mathbb{F}_p)$, $g \to \infty$ as a collection of families would work. (Also see [1, 12] for explicit results on average point counts, as $p \to \infty$, for curves in various families.)

Another possibility to avoid a priori upper bounds might be to consider smooth curves given by intersecting $n-1$ generic hypersurfaces in $\mathbb{P}^n$. However, the average number of points on curves in this family turns out to be bounded; Bucur and Kedlaya recently [4] showed that it is slightly *less* than $p+1$. In particular, the average trace of Frobenius for curves in this family is *not* equal to zero.

1.1.4. *Acknowledgments:* The authors would like to thank Alina Bucur, Chantal David, Nicholas Katz, and Zeév Rudnick for helpful discussions.

2. Proof of Theorem 1

2.1. **Normal distribution from coin flip model.** We first recall some facts about the binomial distribution. For $i \geq 1$, let B_i be the

binomial random variable counting the successful tosses among

$$n_i := |X_i(\mathbb{F}_p)| \tag{2}$$

tosses of an unfair coin, whose probability of success equals

$$r := \frac{p+1}{p^2+p+1}. \tag{3}$$

Let M_i' and V_i' be the expected value and the variance of the number B_i of successful tosses respectively. That is, $M_i' := \mathbb{E}[B_i] = n_i \cdot r$ and $V_i' := Var(B_i) = n_i \cdot r(1-r)$. It is a classical result in probability (see e.g. [8], chapter 7) that B_i, suitably normalized, tends to the standard Gaussian, provided that $n_i \to \infty$ (note that the probability of a success stays constant). Namely, $\frac{B_i - M_i'}{\sqrt{V_i'}} \to \mathcal{N}(0,1)$ in the sense that for any compact interval $I \subseteq \mathbb{R}$,

$$\text{Prob}\left(\frac{B_i - M_i'}{\sqrt{V_i'}} \in I\right) \to \frac{1}{\sqrt{2\pi}} \int_I e^{-x^2/2} dx. \tag{4}$$

We are now in position to prove Theorem 1.

Proof of Theorem 1 assuming Proposition 2. Recall the definitions of n_i and r in Equations (2) and (3) respectively. For the coin flip model the probability of precisely k successes is given by

$$\text{Prob}\left(B_i = k\right) = \binom{n_i}{k} r^k \left(1-r\right)^{n_i - k}$$

whereas, by Proposition 2, $\text{Prob}\left(|C(\mathbb{F}_p)| = k\right)$ for our family of curves equals

$$\binom{n_i}{k} r^k \left(1-r\right)^{n_i - k} \cdot (1 + o(1))$$

with constant involved in the 'o'-notation being uniform in k. In particular, we find that

$$V_i = \frac{1}{|\mathcal{F}_i|} \sum_{C \in \mathcal{F}_i} \left(|C(\mathbb{F}_p)| - M_i\right)^2 = V_i' \cdot (1 + o(1)).$$

Further,

$$\text{Prob}\left(\frac{|C(\mathbb{F}_p)| - M_i}{V_i} \in I\right) = \sum_{\substack{k \in \mathbb{Z} \\ k \in M_i + I \cdot V_i \cap [0, n_i]}} \binom{n_i}{k} r^k \left(1-r\right)^{n_i - k} \cdot (1 + o(1))$$

so by comparing with the corresponding sum in the coin flip model and using the classical (4), together with

$$\sum_{k=0}^{n_i} \binom{n_i}{k} r^k (1-r)^{n_i-k} = 1$$

to control the error term, we obtain the statement of Theorem 1.

□

2.2. **Surfaces with many points.** A crucial assumption in Theorem 1 is the existence of a sequence of surfaces $\{X_i\}$ whose point counts over $\mathbb{F}_p$ tends to infinity; here we give a concrete example of such a sequence. We begin by recalling the construction of modular curves with many $\mathbb{F}_{p^2}$-points given in [18].

Theorem 3. *If l is a prime greater than p, there exists a smooth complete curve*[3] *$X_0(l)$, defined over $\mathbb{F}_p$, having at least $(p-1)(l+1)/12$ points over $\mathbb{F}_{p^2}$.*

If C is curve of genus $g > 1$, and C is not hyperelliptic, there exists a canonical embedding $\phi : C \to \mathbb{P}^{g-1}$, whose image is a curve of degree $2g-2$ inside $\mathbb{P}^{g-1}$ (cf. [9, Ch. 4.5]). Now, the genus of $X_0(l)$ equals $g_l = [l/12]$, and since a hyperellipic curve has at most $2(p+1)$ points, $X_0(l)$ clearly cannot be hyperelliptic for l sufficiently large. Thus $X_0(l)$ can be canonically embedded into $\mathbb{P}^{g_l-1}$ for l large.

Letting l_i denote the i-th prime we now define a sequence of surfaces $\{X_i\}$ by letting X_i be the $\mathbb{F}_p$-surface obtained by Weil restriction of scalars, from $\mathbb{F}_{p^2}$ to $\mathbb{F}_p$, of $X_0(l_i)$. As remarked above, $X_0(l_i)/\mathbb{F}_{p^2}$ is canonically embedded into $\mathbb{P}^{g_{l_i}-1}_{\mathbb{F}_{p^2}}$, and it is known (e.g., see Lemma 7.5 in Ch. I.7.2 of [5]) that $\mathbb{P}^n_{\mathbb{F}_{p^2}}$ can be projectively embedded in $\mathbb{P}^{(n+1)^2-1}_{\mathbb{F}_p}$, hence each X_i is a projective surface with $|X_i(\mathbb{F}_p)| = |X_0(l_i)(\mathbb{F}_{p^2})| \geq (p-1)(l_i+1)/12$ for i sufficiently large.

3. Proof of Proposition 2

We begin with some notation: given a variety X defined over $\mathbb{F}_q$, let $\zeta_X(s)$ denote the zeta function attached to X, i.e.,

$$\zeta_X(s) := \prod_{\substack{P \in X \\ P \text{ closed}}} (1 - p^{-\deg(P)s})^{-1}.$$

[3] $X_0(l)$ is the projective smooth model of the modular curve $Y_0(l)$, which parametrizes pairs (E, C) where E is an elliptic curve and $C \subset E$ is a cyclic subgroup of order l.

In addition we will use the following standard notation from algebraic geometry (see e.g. [9, Sections I.3, II.2]). Given a variety (or scheme) X, its structure sheaf is denoted by $\mathcal{O}_X$, that is, for any open set $U \subseteq X$, $\mathcal{O}_X(U)$ is the ring of regular functions on U. If P is a point of X, we define the local ring $\mathcal{O}_{X,P}$ to be the germs of regular functions on X near P (i.e. the equivalence classes of functions $f \in \mathcal{O}_X(U)$ for some $P \in U$; we identify f with $g \in \mathcal{O}_X(V)$ if $f = g$ on $U \cap V$). It is known that $\mathcal{O}_{X,P}$ is a local ring; its maximal ideal $\mathfrak{m}_{X,P}$ is the set of germs of regular functions that vanish at P.

For $C_i(f) = X_i \cap H_f$ to be smooth, H_f must intersect X_i transversally at all points $P \in X_i(\overline{\mathbb{F}_p})$ such that $f(P) = 0$. I.e., if $f(P) = 0$ for $P \in X(\mathbb{F}_{p^k})$ and we write $f|_{X_i}$ in local coordinates x, y (say with P corresponding to $x = y = 0$), we must have $f|_{X_i} = ax + by +$ (higher order terms), where $a, b \in \mathbb{F}_{p^k}$ *and* $(a, b) \neq (0, 0)$. In other words, $f|_{X_i}$ must not have quadratic order of vanishing at any point $P \in X_i(\mathbb{F}_{p^k})$ (for any k), or equivalently, the image of f in $\mathcal{O}_{X,P}/\mathfrak{m}^2_{X,P}$ must be nonzero for all $P \in X_i(\mathbb{F}_{p^k})$ and all k.

Since X is a smooth surface, $|\mathcal{O}_{X,P}/\mathfrak{m}^2_{X,P}| = p^{3k}$ for all $P \in X_i(\mathbb{F}_{p^k})$, so the "probability of smoothness at P" equals $1 - p^{-3k}$; if these conditions are sufficently independent as P varies, the probability of $C_i(f)$ being smooth should be given by $1/\zeta_{X_i}(3)$. Using a sieving argument, Poonen showed that this heuristic indeed gives the correct answer when $d \to \infty$ and f ranges over elements in $S_i(d)$. Further, his "Finite field Bertini with Taylor coefficients" [16, Theorem 1.2] allows for controlling the behaviour of f in the neighborhood of a finite number of points, best formulated in terms of schemes. We shall need the following slightly more explicit version:

Theorem 4. *Let X be a quasiprojective subscheme of $\mathbb{P}^n$ over $\mathbb{F}_q$, and let $S(d)$ be the set of degree d homogenuous polynomials in $n+1$ variables. Let Z be a finite subscheme of $\mathbb{P}^n$, and assume that $U := X - (Z \cap X)$ is smooth of dimension 2. Fix a subset $T \subset H^0(Z, \mathcal{O}_Z)$. Given $f \in S(d)$, let $f|_Z$ be the element of $H^0(Z, \mathcal{O}_Z)$ that on each connected component Z_k equals the restriction of $x_j^{-d} f$ to Z_k, where $j = j(k)$ is the smallest $j \in 0, 1, \dots, n$ such that the coordinate x_j is invertible on Z_k. Then as $d \to \infty$*

$$|\{f \in S(d) : H_f \cap U \text{ is a smooth curve, and } f|_Z \in T\}| = \frac{|S(d)|}{\zeta_U(3)} \cdot \frac{|T|}{|H^0(Z, \mathcal{O}_Z)|} \cdot (1 + o_Z(1)),$$

Proof. We will closely follow the closed point sieve of [16, Section 2]. The case $T = \emptyset$ is trivial, so we may assume that $|T| \geq 1$. Let U_r be

the set of closed points of U of degree $< r$, and define $U_{\geq r}$ similarly. By the proof of [16, Lemma 2.2], for r fixed and d sufficiently large,

$$|\{f \in S(d) : H_f \cap U \text{ is smooth at all } P \in U_r \text{ and } f|_Z \in T\}| = \frac{|S(d)| \cdot |T|}{|H^0(Z, \mathcal{O}_Z)|} \cdot \prod_{P \in U_{<r}} (1 - p^{-3\deg(P)})$$

which in turn equals

$$\frac{|S(d)| \cdot |T|}{|H^0(Z, \mathcal{O}_Z)|} \cdot \frac{(1 + O(p^{-3r}))}{\zeta_U(3)}$$

Letting r grow with d in a suitable fashion, we obtained the claimed main term; to conclude the proof it is enough to bound the number $f \in S(d)$ for which smoothness of $H_f \cap U$ is violated at some point P of degree larger than r.

Medium degree points: By the proof of [16, Lemma 2.4],

$$|\{f \in S(d) : H_f \cap U \text{ is not smooth at some } P \in U_{\geq r} \cap U_{<d/3}\}| \ll_U |S(d)| \cdot p^{-r}.$$

Large degree points: By the proof of [16, Lemma 2.6] (in particular, see Claim 1 and Claim 2), there exists $\tau \in \mathbb{Z}^+$, only depending on U, such that

$$|\{f \in S(d) : H_f \cap U \text{ is not smooth at some } P \in U_{\geq d/3}\}| = |S(d)| \cdot O(dp^{-(d-\tau)/p} + d^2 p^{-\min([d/p]+1, d/3)})$$

which, for d sufficiently large, is

$$\ll |S(d)| p^{-d/(p+2)}$$

Now, taking $r = [d/4]$, we find that the total contribution from medium and large degree primes is $\ll |S(d)| p^{-d/5}$ and the result follows by taking d sufficiently large so that $p^{-d/(p+3)} = o\left(\frac{1}{|H^0(Z,\mathcal{O}_Z)|\zeta_U(3)}\right)$.

□

Remark. *In particular, taking Z to be empty, we obtain*

$$|\tilde{S}(d)| = |\{f \in S(d) : H_f \cap X \text{ is a smooth curve}\}| = \frac{|S(d)|}{\zeta_X(3)} \cdot (1 + o(1)), \tag{5}$$

which can be interpreted as saying that the probability of $H_f \cap X$ being smooth is $1/\zeta_X(3) + o(1)$ for X a surface.

3.1. **Applications to point counting.** To apply Theorem 4, we define a finite subscheme $Z \subset X$ as follows: let $X(\mathbb{F}_p) = \{P_1, \dots, P_t\}$, and for $1 \le i \le t$ let $\mathfrak{m}_i$ be the ideal sheaf of P_i on X. Let Z_i be the closed subscheme of X corresponding to the ideal sheaf $\mathfrak{m}_i^2 \subset \mathcal{O}_X$, and define $Z := \cup_{i=1}^t Z_i$. Note that $H^0(Z, \mathcal{O}_Z) \simeq \prod_{i=1}^t H^0(Z_i, \mathcal{O}_{Z_i})$ and also that $H^0(Z_i, \mathcal{O}_{Z_i}) \simeq \mathcal{O}_{X,P_i}/\mathfrak{m}_{X,P_i}^2$; in particular

$$|H^0(Z_i, \mathcal{O}_{Z_i})| = |\mathcal{O}_{X,P_i}/\mathfrak{m}_{X,P_i}| \cdot |\mathfrak{m}_{X,P_i}/\mathfrak{m}_{X,P_i}^2| = p \cdot p^2 = p^3.$$

Given a collection of s points $W \subset X(\mathbb{F}_p)$, we define a subset $T(W) \subset H^0(Z, \mathcal{O}_Z)$ using the above isomorphisms: let $T(W) := \prod_{i=1}^t T_i$ where $T_i \subset H^0(Z_i, \mathcal{O}_{Z_i}) \simeq \mathcal{O}_{X,P_i}/\mathfrak{m}_{X,P_i}^2$ is given by

$$T_i := \begin{cases} \{\phi \in \mathfrak{m}_{X,P_i}/\mathfrak{m}_{X,P_i}^2 : \phi \neq 0\} & \text{if } P_i \in W \\ \{\phi \in \mathcal{O}_{X,P_i}/\mathfrak{m}_{X,P_i}^2 : \phi \not\equiv 0 \mod \mathfrak{m}_{X,P_i}\} & \text{if } P_i \in X(\mathbb{F}_p) - W \end{cases}$$

Note that $|T_i| = p^2 - 1$ if $P_i \in W$, and $|T_i| = p^3 - p^2$ if $P_i \notin W$, hence $|T(W)| = (p^2-1)^s(p^3-p^2)^{t-s}$. We also note that if $f \in S(d)$ and $f|_Z \in T$, then $(H_f \cap X)(\mathbb{F}_p) = W$, and $H_f \cap X$ is smooth at all $P_i \in W$.

Letting W range over all subsets of $X(\mathbb{F}_p)$ of cardinality s, define a subset $T \subset H^0(Z, \mathcal{O}_Z)$ by

$$T := \cup_{W \subset X(\mathbb{F}_p) : |W| = s} T(W).$$

Then $|T| = \binom{t}{s} \cdot (p^2-1)^s(p^3-p^2)^{t-s}$, and, by construction, $f|_Z \in T$ is equivalent to $X \cap H_f$ being smooth at all points in $X(\mathbb{F}_p)$ and that $|X \cap H_f| = s$.

By Theorem 4 and (5), together with

$$\zeta_X(s) = \zeta_U(s) \prod_{P \in X(\mathbb{F}_p)} (1 - p^{-s})^{-1} = \zeta_U(s)(1-p^{-s})^{-t}$$

we find that for $d \to \infty$,

$$\begin{aligned} &\frac{|\{f \in S(d) : |X(\mathbb{F}_p) \cap H_f| = s \text{ and } X \cap H_f \text{ is smooth }\}|}{|\{f \in S(d) : X \cap H_f \text{ is smooth }\}|} \\ &= \frac{|S(d)|/\zeta_U(3) \cdot |T|/|H^0(Z, \mathcal{O}(Z))|}{|S(d)|/\zeta_X(3)} \cdot (1 + o(1)) \\ &= \frac{|T|/|H^0(Z, \mathcal{O}(Z))|}{(1-p^{-3})^t} \cdot (1 + o(1)), \end{aligned}$$

which, since $|T| = \binom{t}{s} \cdot (p^2-1)^s(p^3-p^2)^{t-s}$ and $|H^0(Z, \mathcal{O}(Z))| = p^{3t}$, equals

$$\frac{\binom{t}{s}(p^2-1)^s(p^3-p^2)^{t-s}}{(p^3-1)^t} \cdot (1 + o(1))$$

$$= \binom{t}{s} \left(\frac{p+1}{p^2+p+1}\right)^s \left(\frac{p^2}{p^2+p+1}\right)^{t-s} \cdot (1+o(1))$$

and hence concluding the proof of Proposition 2.

References

[1] B. W. Brock and A. Granville. More points than expected on curves over finite field extensions. *Finite Fields Appl.*, 7(1):70–91, 2001. Dedicated to Professor Chao Ko on the occasion of his 90th birthday.

[2] A. Bucur, C. David, B. Feigon, and M. Lalín. The fluctuations in the number of points of smooth plane curves over finite fields. *Preprint, arXiv:0912.4761v1.*

[3] A. Bucur, C. David, B. Feigon, and M. Lalín. Statistics for traces of cyclic trigonal curves over finite fields. *IMRN, to appear.*

[4] A. Bucur and K. S. Kedlaya. The probability that a complete intersection is smooth. *Preprint, arXiv:1003.5222.*

[5] H. Cohen, G. Frey, R. Avanzi, C. Doche, T. Lange, K. Nguyen, and F. Vercauteren, editors. *Handbook of elliptic and hyperelliptic curve cryptography*. Discrete Mathematics and its Applications (Boca Raton). Chapman & Hall/CRC, Boca Raton, FL, 2006.

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

[7] D. Faifman and Z. Rudnick. Statistics of the zeros of zeta functions in families of hyperelliptic curves over a finite field. *Compositio Math., to appear.*

[8] W. Feller. An introduction to probability theory and its applications, vol. 1, 1958, Wiley.

[9] R. Hartshorne. *Algebraic geometry*. Springer-Verlag, New York, 1977. Graduate Texts in Mathematics, No. 52.

[10] Y. Ihara. Some remarks on the number of rational points of algebraic curves over finite fields. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.*, 28(3):721–724 (1982), 1981.

[11] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.

[12] N. M. Katz. Frobenius-Schur indicator and the ubiquity of Brock-Granville quadratic excess. *Finite Fields Appl.*, 7(1):45–69, 2001. Dedicated to Professor Chao Ko on the occasion of his 90th birthday.

[13] N. M. Katz and P. Sarnak. *Random matrices, Frobenius eigenvalues, and monodromy*, volume 45 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 1999.

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

[15] M. Larsen. The normal distribution as a limit of generalized sato-tate measures. *Preprint*, arXiv:0810.2012.

[16] B. Poonen. Bertini theorems over finite fields. *Ann. of Math. (2)*, 160(3):1099–1127, 2004.

[17] Z. Rudnick. Traces of high powers of the Frobenius class in the hyperelliptic ensemble. *Acta Arith.*, to appear.

[18] M. A. Tsfasman, S. G. Vlăduţ, and T. Zink. Modular curves, Shimura curves, and Goppa codes, better than Varshamov-Gilbert bound. *Math. Nachr.*, 109:21–28, 1982.
[19] M. Xiong. The fluctuations in the number of points on a family of curves over a finite field. *J. Théor. Nombres Bordeaux*, to appear.

URL: `www.math.kth.se/~kurlberg`

Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden
E-mail address: `kurlberg@math.kth.se`

URL: `www.math.kth.se/~wigman`

Department of Mathematics, Royal Institute of Technology, SE-100 44 Stockholm, Sweden
E-mail address: `wigman@kth.se`