involve
a journal of mathematics

The number of rational points of hyperelliptic curves over subsets of finite fields

Kristina Nelson, József Solymosi, Foster Tom and Ching Wong

msp

2019 vol. 12, no. 5

INVOLVE 12:5 (2019)
dx.doi.org/10.2140/involve.2019.12.755

The number of rational points of hyperelliptic curves over subsets of finite fields

Kristina Nelson, József Solymosi, Foster Tom and Ching Wong

(Communicated by Kenneth S. Berenhaut)

We prove two related concentration inequalities concerning the number of rational points of hyperelliptic curves over subsets of a finite field. In particular, we investigate the probability of a large discrepancy between the numbers of quadratic residues and nonresidues in the image of such subsets over uniformly random hyperelliptic curves of given degrees. We find a constant probability of such a high difference and show the existence of sets with an exceptionally large discrepancy.

1. Introduction

Let q be a prime power and let $\mathbb{F}_q$ be the finite field with q elements. A curve $E : y^2 = f(x)$ (together with a point at infinity $\mathcal{O}$) is called an *elliptic curve* over $\mathbb{F}_q$ if $f(x) \in \mathbb{F}_q[x]$ is a cubic polynomial having distinct roots in the algebraic closure $\overline{\mathbb{F}}_q$ of $\mathbb{F}_q$. The set of *rational points* of E in $\mathbb{F}_q$ is

$$E(\mathbb{F}_q) = \{(x, y) \in \mathbb{F}_q \times \mathbb{F}_q : y^2 = f(x)\} \cup \{\mathcal{O}\}.$$

Suppose that q is odd. Using the fact that there are $(q-1)/2$ invertible quadratic residues and $(q-1)/2$ nonresidues in $\mathbb{F}_q$, one can approximate the size of $E(\mathbb{F}_q)$ as follows. For each $x \in \mathbb{F}_q$, the probability of $f(x)$ being a nonzero square in $\mathbb{F}_q$, and hence contributing two points to $E(\mathbb{F}_q)$, is about $\frac{1}{2}$. With probability about $\frac{1}{2}$ there is no point in $E(\mathbb{F}_q)$ having the first coordinate $x \in \mathbb{F}_q$. Therefore, $\#E(\mathbb{F}_q)$ is expected to be close to $q+1$. Indeed, Hasse [1936] proved that the error in this estimate is at most $2\sqrt{q}$:

$$|\#E(\mathbb{F}_q) - (q+1)| \le 2\sqrt{q}.$$

Knowledge of $\#E(\mathbb{F}_q)$ is crucial in elliptic curve cryptography (ECC), which is considered to be more efficient than the classical cryptosystems, like RSA [Rivest

MSC2010: 68Q87, 68R05.
Keywords: hyperelliptic curves, finite fields.
Solymosi was supported by NSERC and the Hungarian National Research Development and Innovation Fund K 119528.

et al. 1978]. The security of ECC depends on the difficulty of solving the elliptic curve discrete logarithm problem (ECDLP). The best known algorithm to solve ECDLP in finite fields is Pollard's rho algorithm [1975], which requires $O(\sqrt{p})$ time complexity, where p is the prime factor of q. However, some well-studied classes of elliptic curves are not good candidates for ECC. For instance, if the number of rational points of an elliptic curve E in $\mathbb{F}_p$ is exactly p, where p is a prime, then the running time of solving the ECDLP is $O(\log p)$; see [Semaev 1998]. Using verifiably random elliptic curves in ECC improves security since randomly generated curves are unlikely to be part of a weak class. Hyperelliptic curves can also be used in cryptography; see [Cohen et al. 2006] for more details. However, the verifiability of random hyperelliptic curves is much harder; see [Hess et al. 2001; Satoh 2009].

In this paper, we investigate the behaviour of random hyperelliptic curves over subsets S of $\mathbb{F}_q$. We are interested in the hyperelliptic curves $E : y^2 = f(x)$ where $f(x)$ is a polynomial in $\mathbb{F}_q[x]$ of degree $4k-1$ ($k \geq 1$) having distinct roots in $\overline{\mathbb{F}}_q$. Denote by $E(\mathbb{F}_q, S)$ the rational points of E in $\mathbb{F}_q$ where the x-coordinate is in S; i.e.,

$$E(\mathbb{F}_q, S) = \{(x, y) \in S \times \mathbb{F}_q : y^2 = f(x)\}.$$

We remark that the point at infinity $\mathcal{O}$ is not included in $E(\mathbb{F}_q, S)$. The approximation we have described for $\#E(\mathbb{F}_q)$ suggests that the expected value of $\#E(\mathbb{F}_q, S)$ is about $\#S$. For random hyperelliptic curves E over $\mathbb{F}_q$, the probability that the error $|\#E(\mathbb{F}_q, S) - \#S|$ is small has been extensively studied; see [Pelekis and Ramon 2017; Schmidt et al. 1995] for example.

On the other hand, it is easy to see that there exist many hyperelliptic curves of any (positive) even degree so that the error $|\#E(\mathbb{F}_p, S) - \#S|$ is very large. Indeed, the error is about $\#S$ when $f(x)$ is the square of any nonconstant polynomial in $\mathbb{F}_q[x]$ for any $S \subset \mathbb{F}_p$.

However, an error bound is not obvious in the case of hyperelliptic curves of odd degree, which we study in the probabilistic setting. Equivalently, we examine the difference between the numbers of quadratic residues and nonresidues in the image multiset $f(S)$. Using $4k$-wise independence, we show that all subsets S of $\mathbb{F}_q$ behave similarly, in the sense that the interested discrepancy is proportional to $\sqrt{\#S}$ and has a positive probability which depends only on the degree of the curve.

Theorem 1. *Given a positive integer k and $\varepsilon > 0$, there exist $\delta > 0$ and a threshold N such that the following holds: for every odd prime power $q > N$, if a curve $E : y^2 = f(x)$ is chosen uniformly at random among all hyperelliptic curves of degree $4k-1$ over $\mathbb{F}_q$, then with a probability at least $(4\pi^{3/2}/e^3)2^{-2k} - \varepsilon$, we have*

$$|\#E(\mathbb{F}_q, S) - \#S| > \delta\sqrt{\#S}$$

for any set $S \subset \mathbb{F}_q$ with $\#S \geq N$.

Theorem 2. *Given a positive integer k, there exist a threshold N and $\varepsilon > 0$ such that the following holds: for every odd prime power $q > N$, if a curve $E : y^2 = f(x)$ is chosen uniformly at random among all hyperelliptic curves of degree $4k-1$ over $\mathbb{F}_q$, then with a probability at least ε, we have*

$$|\#E(\mathbb{F}_q, S) - \#S| > 0.8577\sqrt{k}\sqrt{\#S}$$

for any set $S \subset \mathbb{F}_q$ with $\#S \geq N$.

These two theorems imply that one can expect large deviation of magnitude $\sqrt{\#S}$. In the last section, we show that for small sets S of prime fields $\mathbb{F}_p$, the error is often much larger.

2. Preliminaries

Throughout this section, let q be an odd prime power and let n, k be positive integers such that $4k < n \leq q$. Suppose $S = \{s_1, \ldots, s_n\} \subset \mathbb{F}_q$, and

$$f(x) = \sum_{j=0}^{4k-1} a_j x^j \in \mathbb{F}_q[x]$$

is chosen uniformly at random.

We denote by $\#QR$, $\#NR$ and $\#R$ the numbers of $s_i \in S$ such that $f(s_i)$ is an invertible quadratic residue, a quadratic nonresidue and zero in $\mathbb{F}_q$, respectively. Then, $n = \#QR + \#NR + \#R$. It follows that, provided the curve $E : y^2 = f(x)$ forms a hyperelliptic curve of degree $4k-1$ over $\mathbb{F}_q$, the discrepancy we are interested in is

$$|\#E(\mathbb{F}_q, S) - n| = |2\#QR + \#R - n| = |\#QR - \#NR|. \tag{1}$$

This suggests we look at the random variables

$$X_i = \left(\frac{f(s_i)}{q}\right),$$

where $\left(\frac{a}{q}\right)$ is the Legendre symbol defined as

$$\left(\frac{a}{q}\right) = \begin{cases} 0 & \text{if } a \text{ is the zero in } \mathbb{F}_q, \\ 1 & \text{if } a \text{ is a nonzero square in } \mathbb{F}_q, \\ -1 & \text{otherwise.} \end{cases}$$

We note that among all polynomials $f(x) \in \mathbb{F}_q[x]$ of degree at most 3, only a small fraction fail to form elliptic curves. Indeed, the exceptions, where $f(x)$ has degree strictly less than 3 or has multiple roots, contribute $q^3 + q^2(q-1)$ of all the q^4 polynomials considered. When q is large, such exceptions are negligible. This situation generalizes to hyperelliptic curves.

Lemma 3. *Let q be a prime power, k be a positive integer and $\mathbb{F}_q[x]_{4k-1}$ be the set of polynomials in $\mathbb{F}_q[x]$ of degree at most $4k-1$. Then at most a $2/q$ fraction of the polynomials in $\mathbb{F}_q[x]_{4k-1}$ fail to define a hyperelliptic curve of degree $4k-1$.*

Proof. A polynomial in $\mathbb{F}_q[x]$ defines a hyperelliptic curve precisely when it is separable, or equivalently when it is square-free because finite fields are perfect. As shown in [Carlitz 1932], the number of monic square-free polynomials in $\mathbb{F}_q[x]$ of degree $4k-1$ is $q^{4k-1}-q^{4k-2}$. Thus, accounting for scaling, there are $(q-1)(q^{4k-1}-q^{4k-2})$ polynomials in $\mathbb{F}_q[x]$ that define a hyperelliptic curve of degree $4k-1$. Therefore, the fraction of those polynomials in $\mathbb{F}_q[x]$ of degree at most $4k-1$ that do not is

$$\frac{q^{4k}-(q-1)(q^{4k-1}-q^{4k-2})}{q^{4k}} = \frac{2q^{4k-1}-q^{4k-2}}{q^{4k}} < \frac{2}{q}. \qquad \square$$

Hence, the probability that, among all hyperelliptic curves of degree $4k-1$ over $\mathbb{F}_q$, the discrepancy (1) is larger than some $\delta\sqrt{n}$ is at least the probability that, among all polynomials of degree at most $4k-1$ over $\mathbb{F}_q$, the absolute value of the sum of the random variables X_i is larger than the same $\delta\sqrt{n}$ minus $2/q$; i.e.,

$$\mathbb{P}(|\#E(\mathbb{F}_q, S)-n| > \delta\sqrt{n}) \geq \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| > \delta\sqrt{n}\right) - \frac{2}{q}. \tag{2}$$

In the next two subsections, we will first estimate the higher moments

$$\mathbb{E}_j := \mathbb{E}\left(\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right)^j\right), \quad \text{where } 1 \leq j \leq 4k,$$

by finding their main order, and then give lower bounds on the interested probabilities involving the random variables X_i.

2.1. ***Estimating $\mathbb{E}_{2k}$ and $\mathbb{E}_{4k}$.*** Since $f(x) \in \mathbb{F}_q[x]$ is a random polynomial of degree at most $4k-1$, the random variables X_i exhibit $4k$-wise independence. Indeed, by solving a system of linear equations, the number of polynomials $f(x)$ in $\mathbb{F}_q[x]$ of degree at most $4k-1$ satisfying

$$f(s_{i_1}) = r_1, \quad f(s_{i_2}) = r_2, \quad \ldots, \quad f(s_{i_\ell}) = r_\ell$$

is exactly $q^{4k-\ell}$, given $\ell \leq 4k$, $r_1, \ldots, r_\ell \in \mathbb{F}_q$ and distinct $i_1, \ldots, i_\ell \in \{1, \ldots, n\}$. Thus,

$$\begin{aligned}\mathbb{E}(X_{i_1}^{h_1} \cdots X_{i_\ell}^{h_\ell}) &= \sum_{r_1,\ldots,r_\ell \in \mathbb{F}_q} \mathbb{P}(f(s_{i_1})=r_1, \ldots, f(s_{i_\ell})=r_\ell)\left(\frac{r_1}{q}\right)^{h_1} \cdots \left(\frac{r_\ell}{q}\right)^{h_\ell} \\ &= \sum_{r_1,\ldots,r_\ell \in \mathbb{F}_q} \frac{q^{4k-\ell}}{q^{4k}}\left(\frac{r_1}{q}\right)^{h_1} \cdots \left(\frac{r_\ell}{q}\right)^{h_\ell}\end{aligned}$$

$$
\begin{aligned}
&= \Big[\sum_{r_1\in\mathbb{F}_q} \frac{1}{q}\Big(\frac{r_1}{q}\Big)^{h_1}\Big]\cdots\Big[\sum_{r_\ell\in\mathbb{F}_q} \frac{1}{q}\Big(\frac{r_\ell}{q}\Big)^{h_\ell}\Big] \\
&= \Big[\sum_{r_1\in\mathbb{F}_q} \mathbb{P}(f(s_{i_1})=r_1)\Big(\frac{r_1}{q}\Big)^{h_1}\Big]\cdots\Big[\sum_{r_\ell\in\mathbb{F}_q} \mathbb{P}(f(s_{i_\ell})=r_\ell)\Big(\frac{r_\ell}{q}\Big)^{h_\ell}\Big] \\
&= \mathbb{E}(X_{i_1}^{h_1})\cdots\mathbb{E}(X_{i_\ell}^{h_\ell}).
\end{aligned}
$$

We also note that the random variables X_i only take the values $0, 1, -1$, and so $X_i^{2h-1} = X_i$ and $X_i^{2h} = X_i^2$ for all $h \geq 1$. Also, by convention, $X_i^0 = 0$. Therefore we have

$$
\mathbb{E}(X_i^{2h-1}) = \mathbb{E}(X_i) = \sum_{r\in\mathbb{F}_q} \mathbb{P}(f(s_i)=r)\Big(\frac{r}{q}\Big) = \sum_{r\in\mathbb{F}_q} \frac{1}{q}\Big(\frac{r}{q}\Big) = 0,
$$

$$
\mathbb{E}(X_i^{2h}) = \mathbb{E}(X_i^2) = \sum_{r\in\mathbb{F}_q} \mathbb{P}(f(s_i)=r)\Big(\frac{r}{q}\Big)^2 = \sum_{r\in\mathbb{F}_q} \frac{1}{q}\Big(\frac{r}{q}\Big)^2 = 1-\frac{1}{q}.
$$

To summarize the above two observations, we have the following lemma:

Lemma 4. *Let $\ell \leq 4k$, let $h_1, \ldots, h_\ell$ be positive integers, and let $i_1 \ldots, i_\ell$ be distinct numbers from $\{1, \ldots, n\}$. Then,*

$$
\mathbb{E}(X_{i_1}^{h_1}\cdots X_{i_\ell}^{h_\ell}) = \begin{cases} (1-1/q)^\ell & \text{if } h_1, \ldots, h_\ell \text{ are all even numbers,} \\ 0 & \text{otherwise.} \end{cases}
$$

Before we estimate the general $\mathbb{E}_j$, let us compute $\mathbb{E}_6$ (when $k \geq 2$) as a toy version:

$$
\begin{aligned}
\mathbb{E}_6 &= \mathbb{E}\Big(\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i\Big)^6 \\
&= \frac{1}{n^3}\Big(\sum_{i=1}^n \mathbb{E}(X_i^6) + \frac{6!}{4!\,2!}\sum_{i\neq j}\mathbb{E}(X_i^4X_j^2) + \frac{6!}{2!\,2!\,2!}\sum_{i<j<k}\mathbb{E}(X_i^2X_j^2X_k^2)\Big) \\
&= \frac{1}{n^3}\Big(n\Big(1-\frac{1}{q}\Big) + 15n(n-1)\Big(1-\frac{1}{q}\Big)^2 + 90\binom{n}{3}\Big(1-\frac{1}{q}\Big)^3\Big) \\
&= 15\Big(1-\frac{1}{q}\Big)^3 - \frac{15}{n}\Big(1-\frac{1}{q}\Big)^2\Big(2-\frac{3}{q}\Big) + \frac{1}{n^2}\Big(1-\frac{1}{q}\Big)\Big(16-\frac{45}{q}+\frac{30}{q^2}\Big).
\end{aligned}
$$

We derive in the lemma below how the number 15 in the leading term can be expressed in terms of $j = 6$.

Lemma 5. *For $1 \leq j \leq 4k$, we have*

$$
\mathbb{E}_j = \begin{cases} \dfrac{j!}{2^{j/2}(j/2)!} + O_j\Big(\dfrac{1}{n}\Big) \text{ as } n\to\infty, & \text{if } j \text{ is an even number,} \\ 0 & \text{otherwise.} \end{cases}
$$

Proof. If j is an odd number, then every term in the multinomial expansion has at least one odd index, and hence vanishes by Lemma 4.

Suppose now that j is an even integer. Using the multinomial theorem and Lemma 4, we have

$$\begin{aligned}\mathbb{E}_j &= \frac{1}{n^{j/2}}\mathbb{E}\left(\left(\sum_{i=1}^{n} X_i\right)^j\right) = \frac{1}{n^{j/2}}\mathbb{E}\left(\sum_{h_1+\cdots+h_n=j} \frac{j!}{h_1!\cdots h_n!}\prod_{t=1}^{n} X_t^{h_t}\right)\\ &= \frac{1}{n^{j/2}}\sum_{h_1+\cdots+h_n=j} \frac{j!}{h_1!\cdots h_n!}\mathbb{E}\left(\prod_{t=1}^{n} X_t^{h_t}\right)\\ &= \frac{1}{n^{j/2}}\sum_{\substack{h_1+\cdots+h_n=j\\ h_i \text{ even}}} \frac{j!}{h_1!\cdots h_n!}\left(1-\frac{1}{q}\right)^{\#\{i:h_i>0\}} = \frac{1}{n^{j/2}}\sum_{m=1}^{j/2}\left(1-\frac{1}{q}\right)^m H(j,m),\end{aligned}$$

where

$$H(j,m) = \sum_{\substack{h_1+\cdots+h_n=j\\ h_i \text{ even}\\ \#\{i:h_i>0\}=m}} \frac{j!}{h_1!\cdots h_n!} = \binom{n}{m}\sum_{\substack{h_1'+\cdots+h_m'=j\\ h_i'>0 \text{ even}}} \frac{j!}{h_1'!\cdots h_m'!}$$

is a polynomial (with integer coefficients) in n of degree m. Therefore, the leading term of $\mathbb{E}_j$ comes from the summand where $m = j/2$. In this case, $h_i' = 2$ for every $1 \le i \le j/2$ and so

$$H(j,\, j/2) = \binom{n}{j/2}\frac{j!}{2^{j/2}}$$

has leading term

$$\frac{j!}{(j/2)!2^{j/2}}n^{j/2}.$$

It follows that

$$\begin{aligned}\mathbb{E}_j &= \frac{1}{n^{j/2}}\left(\left(1-\frac{1}{q}\right)^{j/2}\frac{j!}{(j/2)!\,2^{j/2}}n^{j/2}+\cdots\right)\\ &= \left(1-\frac{1}{q}\right)^{j/2}\frac{j!}{(j/2)!\,2^{j/2}} + O_j\left(\frac{1}{n}\right) = \frac{j!}{(j/2)!\,2^{j/2}} + O_j\left(\frac{1}{n}\right)\end{aligned}$$

as $n \to \infty$. □

In particular, for each fixed k,

$$\mathbb{E}_{2k} = \frac{(2k)!}{2^k k!} + O_k\left(\frac{1}{n}\right)$$

is bounded uniformly in $n \ge 1$. As a consequence, one can have the following estimates, which will be used later in our proof, using Stirling's approximation. For

all fixed $k \geq 1$, we have

$$\sqrt[2k]{\mathbb{E}_{2k}} \geq \sqrt{\frac{2k}{e}} + O_k\Big(\frac{1}{n}\Big) \tag{3}$$

and

$$\frac{\mathbb{E}_{2k}^2}{\mathbb{E}_{4k}} \geq \Big(\frac{\sqrt{2\pi}}{e}\Big)^3 2^{1/2-2k} + O_k\Big(\frac{1}{n}\Big) \tag{4}$$

as $n \to \infty$.

2.2. *Lower bounds for the probabilities.*

Proposition 6. *Under the setting stated in the beginning of this section, we have*

$$\mathbb{P}\bigg(\bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\bigg| > \delta\bigg) \geq \frac{(\mathbb{E}_{2k} - \delta^{2k})^2}{\mathbb{E}_{4k} - 2\delta^{2k}\mathbb{E}_{2k} + \delta^{4k}} \tag{5}$$

for any $0 < \delta < \frac{1}{2}$, *and*

$$\mathbb{P}\bigg(\bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\bigg| \geq \sqrt[2k]{\mathbb{E}_{2k}} - \varepsilon^{1/2-o(1)}\bigg) \geq \varepsilon > 0 \tag{6}$$

as $\varepsilon \to 0$.

Proof. Let $c \geq 1$ be a parameter to be determined. Using the second-moment Markov inequality, one can show that for $0 < \lambda < c^{2k}$,

$$\begin{aligned}
\mathbb{P}\bigg(\bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\bigg| > \sqrt[2k]{c^k - \sqrt{\lambda}}\bigg) &= \mathbb{P}\bigg(\bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\bigg)^{2k} - c^k > -\sqrt{\lambda}\bigg) \\
&\geq \mathbb{P}\bigg(\bigg|\bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\bigg)^{2k} - c^k\bigg| < \sqrt{\lambda}\bigg) \\
&\geq 1 - \frac{1}{\lambda}\mathbb{E}\bigg(\bigg(\bigg(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\bigg)^{2k} - c^k\bigg)^2\bigg) \\
&= 1 - \frac{c^{2k} - 2c^k\mathbb{E}_{2k} + \mathbb{E}_{4k}}{\lambda}.
\end{aligned} \tag{7}$$

To prove (5), we take $\lambda = (c^k - \delta^{2k})^2$, where $\delta > 0$ is small. Maximizing the right-hand side of (7) over c, we see that the maximum is

$$1 - \frac{c^{2k} - 2c^k\mathbb{E}_{2k} + \mathbb{E}_{4k}}{(c^k - \delta^{2k})^2} = \frac{(\mathbb{E}_{2k} - \delta^{2k})^2}{\mathbb{E}_{4k} - 2\delta^{2k}\mathbb{E}_{2k} + \delta^{4k}},$$

when

$$c^k = \frac{\mathbb{E}_{4k} - \delta^{2k}\mathbb{E}_{2k}}{\mathbb{E}_{2k} - \delta^{2k}}.$$

Now we prove (6). To make

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i\right| > \sqrt[2k]{c^k - \sqrt{\lambda}}\right) \geq \varepsilon,$$

we take

$$\lambda = \frac{c^{2k} - 2c^k\mathbb{E}_{2k} + \mathbb{E}_{4k}}{1-\varepsilon}.$$

Since we require $c^{2k} > \lambda$, it follows that

$$c^{2k} - 2c^k\mathbb{E}_{2k} + \mathbb{E}_{4k} < c^{2k} - c^{2k}\varepsilon,$$

and therefore

$$\eta := \varepsilon c^k < 2\mathbb{E}_{2k} - \frac{\mathbb{E}_{4k}}{c^k} < 2\mathbb{E}_{2k}.$$

To compute the leading terms of $\sqrt[2k]{c^k - \sqrt{\lambda}}$ as $\varepsilon \to 0$, we first use the binomial series to expand the numerator of $\sqrt{\lambda}$ as

$$c^k\sqrt{1 - \left(\frac{2\mathbb{E}_{2k}}{c^k} - \frac{\mathbb{E}_{4k}}{c^{2k}}\right)} = c^k\left(1 - \mathbb{E}_{2k}\frac{1}{c^k} + \frac{\mathbb{E}_{4k} - \mathbb{E}_{2k}^2}{2}\frac{1}{c^{2k}} + O\left(\frac{1}{c^{3k}}\right)\right) \tag{8}$$

as $c \to \infty$. Indeed, the bracket inside the square root in (8) is small in view of Lemma 5. To get $\sqrt{\lambda}$, we multiply (8) by

$$\frac{1}{\sqrt{1-\varepsilon}} = 1 + \tfrac{1}{2}\varepsilon + \tfrac{3}{8}\varepsilon^2 + O(\varepsilon^3).$$

Substituting $c^k = \eta/\varepsilon$, we have

$$\begin{aligned} c^k - \sqrt{\lambda} &= \frac{\eta}{\varepsilon}\left[1 - \left(1 + \frac{1}{2}\varepsilon + \frac{3}{8}\varepsilon^2 + O(\varepsilon^3)\right)\left(1 - \frac{\mathbb{E}_{2k}}{\eta}\varepsilon + \frac{\mathbb{E}_{4k} - \mathbb{E}_{2k}^2}{2\eta^2}\varepsilon^2 + O\left(\frac{\varepsilon^3}{\eta^3}\right)\right)\right] \\ &= \mathbb{E}_{2k} - \frac{1}{2}\eta + \left(\frac{\mathbb{E}_{2k}^2 - \mathbb{E}_{4k}}{2} + \frac{\mathbb{E}_{2k}}{2}\eta - \frac{3}{8}\eta^2\right)\frac{\varepsilon}{\eta} + O\left(\frac{\varepsilon^2}{\eta^2}\right). \end{aligned}$$

We may now take η satisfying $\sqrt{\varepsilon} \ll \eta \ll 1$ so that the terms in the last line are indeed arranged in decreasing order of magnitude. Therefore,

$$\sqrt[2k]{c^k - \sqrt{\lambda}} = \sqrt[2k]{\mathbb{E}_{2k} - \varepsilon^{1/2-o(1)}} = \sqrt[2k]{\mathbb{E}_{2k}} - \varepsilon^{1/2-o(1)}$$

as $\varepsilon \to 0$, establishing (6). □

3. Proofs of the theorems

Proof of Theorem 1. Write $n = \#S$, as in Section 2. Given $\varepsilon > 0$, we choose N large enough so that $2/N < \varepsilon/3$, and the error appearing in (4) has an absolute value less than $\varepsilon/3$.

Since $\mathbb{E}_{4k} > \mathbb{E}_{2k} \geq \frac{1}{2}$, there exists a small $\delta > 0$ such that

$$\left| \frac{(1-\delta^{2k}/\mathbb{E}_{2k})^2}{1-2\delta^{2k}(\mathbb{E}_{2k}/\mathbb{E}_{4k})+\delta^{4k}(1/\mathbb{E}_{4k})} - 1 \right| < \frac{\varepsilon}{3}\frac{\mathbb{E}_{4k}}{\mathbb{E}_{2k}^2}.$$

Together with (2), (5) and (4), we have

$$\begin{aligned}
\mathbb{P}(|\#E(\mathbb{F}_q, S) - n| > \delta\sqrt{n}) &\geq \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| > \delta\sqrt{n}\right) - \frac{2}{q} \\
&\geq \frac{\mathbb{E}_{2k}^2}{\mathbb{E}_{4k}} \frac{(1-\delta^{2k}/\mathbb{E}_{2k})^2}{1-2\delta^{2k}(\mathbb{E}_{2k}/\mathbb{E}_{4k})+\delta^{4k}(1/\mathbb{E}_{4k})} - \frac{\varepsilon}{3} \\
&\geq \frac{\mathbb{E}_{2k}^2}{\mathbb{E}_{4k}} - \frac{\varepsilon}{3} - \frac{\varepsilon}{3} \geq \left(\frac{\sqrt{2\pi}}{e}\right)^3 2^{1/2-2k} - \varepsilon,
\end{aligned}$$

as desired. □

Proof of Theorem 2. Similarly we write $n = \#S$. Using the estimate (3), we choose N so large and ε so small that the following lower bound implied by (6) is large:

$$\sqrt[2k]{\mathbb{E}_{2k}} - \varepsilon^{1/2-o(1)} > 0.8577\sqrt{k}.$$

Here 0.8577 is a number strictly smaller than $\sqrt{2/e}$. Now, increasing N if necessary, we also have $2/N < \varepsilon/2$. Then, by (2) and (6), we have

$$\begin{aligned}
\mathbb{P}\big(|\#E(\mathbb{F}_q, S) - n| > 0.8577\sqrt{k}\sqrt{n}\big) &\geq \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| > 0.8577\sqrt{k}\sqrt{n}\right) - \frac{2}{q} \\
&\geq \mathbb{P}\left(\left|\sum_{i=1}^{n} X_i\right| > (\sqrt[2k]{\mathbb{E}_{2k}} - \varepsilon^{1/2-o(1)})\sqrt{n}\right) - \frac{\varepsilon}{2} \\
&\geq \frac{\varepsilon}{2}.
\end{aligned}$$

□

4. Sets with exceptionally large discrepancy

So far we have considered sets of arbitrarily large size. We will show, as one may expect, that if n is a constant, then for each prime p large enough, there is a probability $\alpha > 0$ that the error is much larger than $\sqrt{n}$ for $\beta\binom{p}{n}$ of the subsets $S \subset \mathbb{F}_p$ of size n. In particular, for each n, there is a probability 2^{-n-1} that a randomly chosen subset $S \subset \mathbb{F}_p$ of size n has the following property — a randomly chosen monic separable cubic f over $\mathbb{F}_p$ has a probability 2^{-n-1} so that $f(S)$ consists only of nonzero quadratic residues or quadratic nonresidues.

Let $\mathcal{F}$ be the set of monic, separable cubics over $\mathbb{F}_p$. Note that $\#\mathcal{F} = p^3 - p^2$. Let m, n be constants independent of p such that $n - 2m > \sqrt{n}$. We construct a bipartite graph G with $\binom{p}{n}$ "S-vertices" in one partition, each associated with a

set $S \subset \mathbb{F}_p$ of size n, and $p^3 - p^2$ "f-vertices" in the other, each associated with an $f \in \mathcal{F}$. We draw an edge between the vertex corresponding to f and the vertex corresponding to S when

$$\left| \sum_{s_i \in S} \left(\frac{f(s_i)}{p} \right) \right| \geq n - 2m.$$

Fix $f \in \mathcal{F}$, and let $\mathcal{Q} \subset \mathbb{F}_p$ be the set of points mapped by f to a nonzero quadratic residue, and $\mathcal{N} \subset \mathbb{F}_p$ be those points mapped to a nonresidue. Let $p/2 + A_f$ be the size of the larger of these two sets. Then the degree of the vertex associated to f in G is at least

$$\binom{p/2 - A_f}{m} \binom{p/2 + A_f}{n-m}. \tag{9}$$

By Hasse's theorem we have $A_f \leq \sqrt{p}$, and so (9) is bounded below by

$$\binom{p/2 - \sqrt{p}}{m} \binom{p/2 - \sqrt{p}}{n-m} = \binom{p}{n} \left[\binom{n}{m} 2^{-n} + o(1) \right]$$

as $p \to \infty$. Thus the number of edges in our graph, E, is at least

$$\binom{p}{n} \left[\binom{n}{m} 2^{-n} + o(1) \right] (p^3 - p^2).$$

Now if only $\beta \binom{p}{n}$ of the S-vertices achieve degree at least $\alpha(p^3 - p^2)$, then we have

$$E \leq \beta \binom{p}{n} (p^3 - p^2) + \binom{p}{n} (1 - \beta) \alpha (p^3 - p^2),$$

and so

$$\beta \geq \frac{1}{1 - \alpha} \left[\binom{n}{m} 2^{-n} - \alpha + o(1) \right] > 0$$

as $p \to \infty$, provided that $\alpha > 0$ is small enough.

References

[Carlitz 1932] L. Carlitz, "The arithmetic of polynomials in a Galois field", *Amer. J. Math.* **54**:1 (1932), 39–50. MR Zbl

[Cohen et al. 2006] H. Cohen, G. Frey, R. Avanzi, C. Doche, T. Lange, K. Nguyen, and F. Vercauteren (editors), *Handbook of elliptic and hyperelliptic curve cryptography*, Chapman & Hall, Boca Raton, FL, 2006. MR Zbl

[Hasse 1936] H. Hasse, "Zur Theorie der abstrakten elliptischen Funktionenkörper, III: Die Struktur des Meromorphismenrings, die Riemannsche Vermutung", *J. Reine Angew. Math.* **175** (1936), 193–208. MR Zbl

[Hess et al. 2001] F. Hess, G. Seroussi, and N. P. Smart, "Two topics in hyperelliptic cryptography", pp. 181–189 in *Selected areas in cryptography* (Toronto, 2001), edited by S. Vaudenay and A. M. Youssef, Lecture Notes in Comput. Sci. **2259**, Springer, 2001. MR Zbl

[Pelekis and Ramon 2017] C. Pelekis and J. Ramon, "Hoeffding's inequality for sums of dependent random variables", *Mediterr. J. Math.* **14**:6 (2017), art. id. 243. MR Zbl

[Pollard 1975] J. M. Pollard, "A Monte Carlo method for factorization", *Nordisk Tidskr. Informationsbehandling* **15**:3 (1975), 331–334. MR Zbl

[Rivest et al. 1978] R. L. Rivest, A. Shamir, and L. Adleman, "A method for obtaining digital signatures and public-key cryptosystems", *Comm. ACM* **21**:2 (1978), 120–126. MR Zbl

[Satoh 2009] T. Satoh, "Generating genus two hyperelliptic curves over large characteristic finite fields", pp. 536–553 in *Advances in cryptology: EUROCRYPT 2009* (Cologne, 2009), edited by A. Joux, Lecture Notes in Comput. Sci. **5479**, Springer, 2009. MR Zbl

[Schmidt et al. 1995] J. P. Schmidt, A. Siegel, and A. Srinivasan, "Chernoff–Hoeffding bounds for applications with limited independence", *SIAM J. Discrete Math.* **8**:2 (1995), 223–250. MR Zbl

[Semaev 1998] I. A. Semaev, "Evaluation of discrete logarithms in a group of p-torsion points of an elliptic curve in characteristic p", *Math. Comp.* **67**:221 (1998), 353–356. MR Zbl

Received: 2018-01-19 Revised: 2018-06-21 Accepted: 2018-07-28

krisn@math.berkeley.edu *Department of Mathematics, University of California, Berekeley, CA, United States*

solymosi@math.ubc.ca *Department of Mathematics, University of British Columbia, Vancouver, BC, Canada*

foster@math.ubc.ca *Department of Mathematics, University of British Columbia, Vancouver, BC, Canada*

ching@math.ubc.ca *Department of Mathematics, University of British Columbia, Vancouver, BC, Canada*

involve

msp.org/involve

INVOLVE YOUR STUDENTS IN RESEARCH

Involve showcases and encourages high-quality mathematical research involving students from all academic levels. The editorial board consists of mathematical scientists committed to nurturing student participation in research. Bridging the gap between the extremes of purely undergraduate research journals and mainstream research journals, *Involve* provides a venue to mathematicians wishing to encourage the creative involvement of students.

See inside back cover or msp.org/involve for submission instructions. The subscription price for 2019 is US $195/year for the electronic version, and $260/year (+$35, if shipping outside the US) for print and electronic. Subscriptions, requests for back issues and changes of subscriber address should be sent to MSP.

Involve (ISSN 1944-4184 electronic, 1944-4176 printed) at Mathematical Sciences Publishers, 798 Evans Hall #3840, c/o University of California, Berkeley, CA 94720-3840, is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

Involve peer review and production are managed by EditFlow® from Mathematical Sciences Publishers.

PUBLISHED BY

mathematical sciences publishers

nonprofit scientific publishing

http://msp.org/

involve

2019 vol. 12 no. 5

1944-4176(2019)12:5;1-0