THE NUMBER OF RATIONAL POINTS OF HYPERELLIPIC CURVES OVER
SUBSETS OF FINITE FIELDS

KRISTINA NELSON
Department of Mathematics,
University of California, Berkeley,
Berkeley, CA 94720-3840

JÓZSEF SOLYMOSI*, FOSTER TOM, CHING WONG
Department of Mathematics,
University of British Columbia,
Vancouver, BC, Canada, V6T 1Z2

Abstract. We prove two related concentration inequalities concerning the number of rational points 
of hyperelliptic curves over subsets of a prime field. In particular, we investigate the probability of a 
large discrepancy between the numbers of quadratic residues and non-residues 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 $p$ be a prime and let $\mathbb{F}_p$ be the finite field with $p$ elements. A curve $E : y^2 = f(x)$ (together with 
a point of infinity $O$) is called an elliptic curve over $\mathbb{F}_p$ if $f(x)$ is a cubic polynomial having distinct 
roots in the algebraic closure $\overline{\mathbb{F}_p}$ of $\mathbb{F}_p$. The set of rational points of $E$ in $\mathbb{F}_p$ is

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

One can approximate the size of $E(\mathbb{F}_p)$ as follows. For each $x \in \mathbb{F}_p$, the probability of $f(x)$ being a 
non-zero square in $\mathbb{F}_p$, and hence contributing 2 points to $E(\mathbb{F}_p)$, is about $1/2$. With about probability 
$1/2$ there is no point in $E(\mathbb{F}_p)$ having the first coordinate $x \in \mathbb{F}_p$. Therefore, $\#E(\mathbb{F}_p)$ is expected to 
be close to $p + 1$. Indeed, Hasse [2] proved that the error in this estimate is at most $2\sqrt{p}$:

$$\left| \#E(\mathbb{F}_p) - (p + 1) \right| \leq 2\sqrt{p}.$$ 

Knowledge of $\#E(\mathbb{F}_p)$ is crucial in elliptic curve cryptography (ECC), which is considered to be 
more efficient than the classical cryptosystems, like RSA [6]. The security of ECC depends on the dif-
ficulty of solving the Elliptic Curve Discrete Logarithm Problem (ECDLP). The best known algorithm 
to solve ECDLP in finite fields is Pollard’s Rho Algorithm [5], which requires $O(\sqrt{p})$ time complexity. 
However, some well studied elliptic curves or elliptic curves of certain forms 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 $p$, then 
the running time of solving the ECDLP is $O(\log p)$, see [5]. Using verifiably random elliptic curves 
in ECC can ensure higher security. Hyperelliptic curves can also be used in cryptography, see [1] for 
more details; however, the verifiability of random hyperelliptic curves is much harder, see [5, 7].

*The second author was supported by NSERC and Hungarian National Research Development and Innovation Fund 
K 119528.
In this paper, we investigate the behaviour of random hyperelliptic curves over subsets $S$ of $\mathbb{F}_p$. We are interested in the hyperelliptic curves $E : y^2 = f(x)$ where $f(x)$ is a polynomial of degree $4k - 1$ ($k \geq 1$) having distinct roots in $\mathbb{F}_p$. Denote by $E(\mathbb{F}_p, S)$ the rational points of $E$ in $\mathbb{F}_p$ where the $x$-coordinate is in $S$, i.e.

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

We remark that the point of infinity $\mathcal{O}$ is not included in $E(\mathbb{F}_p, S)$. The approximation we have described for $\#E(\mathbb{F}_p)$ also suggests that the expected value of $\#E(\mathbb{F}_p, S)$ is about $\#S$. For random hyperelliptic curves $E$ over $\mathbb{F}_p$, the probability that the error $|\#E(\mathbb{F}_p, S) - \#S|$ is small has been extensively studied, see [4], [9] for example. We will instead focus on the probability of having large error over a given subset $S \subseteq \mathbb{F}_p$. Equivalently, we examine the difference between the numbers of quadratic residues and non-residues in the image multiset $f(S)$. Using $4k$-wise independence, we find a lower bound on the probability that the error is larger than $\delta \sqrt{\#S}$, as well as a lower bound on the error where the probability is positive.

**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 prime $p > N$, if a curve $E : y^2 = f(x)$ is chosen uniformly at random among all degree $4k - 1$ hyperelliptic curves over $\mathbb{F}_p$, then with probability at least $2^{1/2 - 2k} - \varepsilon$, we have

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

for any set $S \subseteq \mathbb{F}_p$ 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 prime $p > N$, if a curve $E : y^2 = f(x)$ is chosen uniformly at random among all degree $4k - 1$ hyperelliptic curves over $\mathbb{F}_p$, then with probability at least $\varepsilon$, we have

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

for any set $S \subseteq \mathbb{F}_p$ with $\#S \geq N$.

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

2. **Preliminaries**

Throughout this section, let $p \geq 5$ be a prime and let $n, k$ be positive integers such that $4k < n \leq p$. Suppose $S = \{s_1, \ldots, s_n\} \subseteq \mathbb{F}_p$, and

$$f(x) = \sum_{j=0}^{4k-1} a_j x^{2^j} \in \mathbb{F}_p[x]$$

is chosen uniformly at random. We are interested in the normalized sum $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$ of the random variables $X_i = (\frac{f(s_i)}{p})$, where $(\frac{a}{p}) \in \{0, 1, -1\}$ is the Legendre symbol. The desired probability will first be estimated in terms of the moments

$$E_k := E \left( \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \right)^k \right),$$

which can be easily computed to the main order as follows.

**Lemma 3.** For each $k \geq 1$, we have

$$E_{2k} = \frac{(2k)!}{2^{2k} k!} + O \left( \frac{1}{n} \right), \quad \text{as} \quad n \to \infty.$$ 

In particular, for each fixed $k$, $E_{2k}$ is bounded uniformly in $n \geq 1$. As a consequence,

$$\sqrt{\frac{2k}{e}} + O \left( \frac{1}{n} \right) \leq \sqrt{\mathbb{E}_{2k}} \leq \sqrt{\frac{2k}{\pi}} + O \left( \frac{1}{n} \right), \quad \text{as} \quad n \to \infty,$$  

(1)
for all $k \geq 1$, and

\begin{equation}
\frac{E_{2k}^2}{E_{4k}} = \frac{(2k)!^3}{(k!)^2(4k)!} + O\left(\frac{1}{n}\right), \quad \text{as} \quad n \to \infty.
\end{equation}

Proof. The proof is by direct computations and the multinomial theorem. Since $f$ is a random polynomial of degree at most $4k-1$, the random variables $X_i$ exhibit $4k$-wise independence. We also use that the mean of $X_i$ is 0 and the even powers of $X_i$ are 1 except when $f(s_i) = 0$. As an illustration, we find the exact formula for $E_6$ as follows.

\begin{align*}
E_6 &= E\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i\right)^6 \\
&= \frac{1}{n^3} \left(\sum_{i=1}^{n} E(X_i^6) + \frac{6!}{4!2!} \sum_{i \neq j} E(X_i^4 X_j^2) + \frac{6!}{2!2!2!} \sum_{i < j < k} E(E_i^2 E_j^2 E_k^2)\right) \\
&= \frac{1}{n^3} \left(n \left(1 - \frac{1}{p}\right) + 15n(n-1) \left(1 - \frac{1}{p}\right)^2 + 90 \frac{n}{3} \left(1 - \frac{1}{p}\right)^3\right) \\
&= 15 \left(1 - \frac{1}{p}\right)^3 - \frac{15}{n} \left(1 - \frac{1}{p}\right)^2 \left(2 - \frac{3}{p}\right) + \frac{1}{n^2} \left(1 - \frac{1}{p}\right) \left(16 - \frac{45}{p} + \frac{30}{p^2}\right)
\end{align*}

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

Lemma 4. For every $k \geq 1$, there is a constant $c_k$ depending only on $k$ such that all except at most a fraction $c_k/p$ of polynomials of degree at most $4k-1$ define a (hyper)elliptic curve of degree $4k-1$.

With this observation, the following more accurate estimates imply our main theorems, as shown in the next section.

Proposition 5. Under the above setting, for any $0 < \delta < 1/2$, we have

\begin{equation}
P \left(\left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i\right| > \delta\right) \geq \frac{(E_{2k} - \delta^{2k})^2}{E_{4k} - 2\delta^{2k}E_{2k} + \delta^{4k}}.
\end{equation}

We also have

\begin{equation}
P \left(\left|\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i\right| \geq \sqrt{E_{2k}} - \varepsilon^{\frac{1}{2} - o(1)}\right) \geq \varepsilon > 0,
\end{equation}

as $\varepsilon \to 0$. 

3
Proof. Let \( c \geq 1 \) be a parameter to be determined. Using Markov’s inequality, one can show that for \( 0 < \lambda < c^{2k} \),

\[
\mathbb{P} \left( \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \right| > 2^{k} \sqrt{c^k - \sqrt{\lambda}} \right) = \mathbb{P} \left( \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \right| > c^k - \sqrt{\lambda} \right) \\
\geq \mathbb{P} \left( \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \right| < c^k \right) \\
\geq 1 - \frac{1}{\lambda} \mathbb{E} \left( \left( \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i \right)^2 \right)^{2} \right) \\
= 1 - \frac{c^{2k} - 2c^k \mathbb{E}_{2k} + \mathbb{E}_{4k}}{\lambda}.
\]

(5)

To prove (4), we take \( \lambda = (c^k - \delta^{2k})^2 \), where \( \delta > 0 \) is small. Maximizing the right hand side of (5) over \( c \), we see that the maximum is

\[
1 - \frac{c^{2k} - 2c^k \mathbb{E}_{2k} + \mathbb{E}_{4k}}{(c^k - \delta^{2k})^2} = \frac{(\mathbb{E}_{2k} - \delta^{2k})^2}{\mathbb{E}_{4k} - 2\delta^{2k} \mathbb{E}_{2k} + \delta^{4k}}, \quad \text{when} \quad c^k = \frac{\mathbb{E}_{4k} - \delta^{2k} \mathbb{E}_{2k}}{\mathbb{E}_{2k} - \delta^{2k}}.
\]

Now we are going to prove (4). To make

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

we will take

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

Since \( c^{2k} > \lambda \), it follows that

\[
\eta := c^k \varepsilon < 2\mathbb{E}_{2k}.
\]

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

\[
c^k \sqrt{1 - \frac{2\mathbb{E}_{2k} - \mathbb{E}_{4k}}{c^{2k}}} = c^k \left( 1 - \mathbb{E}_{2k} \frac{1}{c^k} + \frac{\mathbb{E}_{4k} - \mathbb{E}_{2k}^2}{2 c^{2k}} + O \left( \frac{1}{c^{3k}} \right) \right),
\]

as \( c \to \infty \). Indeed, the bracket inside the square root is small in view of Lemma 3. In order to multiply this to

\[
\frac{1}{\sqrt{1 - \varepsilon}} = 1 + \frac{1}{2} \varepsilon + \frac{3}{8} \varepsilon^2 + O(\varepsilon^3),
\]

we have \( c^k = \eta \varepsilon \) and thus

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

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

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

as \( \varepsilon \to 0 \), establishing (4). \( \square \)
3. Proofs of the Theorems

Proof of Theorem 1. Write \( n = \#S \). By Lemma 1 we choose \( N \) large enough so that \( c_k/N < \varepsilon \) and the error appearing in (2) is less than \( \varepsilon \).

Since \( E_{4k} > E_{2k} \geq 1/2 \), there exists a small \( \delta > 0 \) such that

\[
\left| \frac{(1 - \frac{\delta^2 k}{E_{4k}})^2}{1 - 2\delta^2 k \frac{E_{2k}}{E_{4k}} + \delta^4 k \frac{1}{E_{4k}}} - 1 \right| < \varepsilon \frac{E_{4k}}{E_{2k}}.
\]

Now by Proposition 5,

\[
P \left( \left| \sum_{i=1}^{n} X_i \right| > \delta \sqrt{n} \right) \geq \frac{P^2}{E_{4k}} \frac{(1 - \frac{\delta^2 k}{E_{4k}})^2}{1 - 2\delta^2 k \frac{E_{2k}}{E_{4k}} + \delta^4 k \frac{1}{E_{4k}}} \geq \frac{P^2}{E_{4k}} - \varepsilon \geq 2^{1/2 - 2k - 2\varepsilon}.
\]

□

Proof of Theorem 2. Similarly we write \( n = \#S \). We choose \( N \) so large and \( \varepsilon \) so small that the following hold simultaneously:

1. \( \frac{\sqrt{n}}{\sqrt{E_{2k}} - \frac{1}{\varepsilon}} > 0.8577 \sqrt{k} \)

(here 0.8577 is a number strictly smaller than \( \sqrt{2/\varepsilon} \))

2. \( c_k/N < \varepsilon/2 \).

Indeed, one can first fulfill the first condition using (1) and then increase \( N \) if necessary so that the second also holds. Then

\[
P \left( \left| \sum_{i=1}^{n} X_i \right| > 0.8577 \sqrt{k} \sqrt{n} \right) \geq \varepsilon.
\]

So

\[
P \left( \left| \#E(F_p, S) - \#S \right| \right) \geq \varepsilon - \frac{\varepsilon}{2} = \frac{\varepsilon}{2}.
\]

□

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 \( \beta(p) \) of the subsets \( S \subset F_p \) of size \( n \). In particular, for each \( n \), there is a probability \( 2^{-\sqrt{n}} \) that a randomly chosen subset \( S \subset F_p \) of size \( n \) has the following property — a randomly chosen monic separable cubic \( f \) over \( F_p \) has a probability \( 2^{-n-1} \) so that \( f(S) \) consists only of non-zero quadratic residues or quadratic non-residues.

Let \( \mathcal{F} \) be the set of monic, separable cubics over \( F_p \). 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 \( S \subset F_p \) 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 \( Q \subset F_p \) be the set of points mapped by \( f \) to a non-zero quadratic residue, and \( N \subset F_p \) be those points mapped to a non-residue. 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

\[
(\frac{p}{2} - A_f) (\frac{p}{2} + A_f).
\]

(6)
By Hasse’s theorem we have $A_f \leq \sqrt{p}$, and so (6) is bounded below by

$$\left( \frac{p}{2} - \sqrt{p} \right) \left( \frac{n}{m} \right) \left( \frac{p}{n-m} \right) = \left( \frac{p}{n} \right) \left( \left( \frac{n}{m} \right) 2^{-n} + o(1) \right),$$

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

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

Now if only $\beta \left( \frac{p}{n} \right)$ of the $S$-vertices achieve degree $\geq \alpha (p^3 - p^2)$, then we have

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

and so

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

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

**References**

[1] H. Cohen, G. Frey, R. Avanzi, Ch. Doche, T. Lange, K. Nguyen and F. Vercauteren (Eds.) *Handbook of Elliptic and Hyperelliptic Curve Cryptography*, Discrete Mathematics and its Applications (Boca Raton), Chapman & Hall/CRC (2006).

[2] H. Hasse, *Zur Theorie der abstrakten elliptischen Funktionenkörper I. Die Struktur der Gruppe der Divisorenklassen endlicher Ordnung*, Journal für die reine und angewandte Mathematik 175 (1936) 55-62.

[3] F. Hess, G. Seroussi and N.P. Smart, *Two topics in hyperelliptic cryptography*, in: Selected Areas in Cryptography, Lecture Notes in Comput. Sci. 2259 Springer, Berlin (2001) 181–189.

[4] C. Pelekis and J. Ramon, *Hoeffding’s inequality for sums of weakly dependent random variables*, Preprint. Arxiv: 1507.06871v1

[5] J.M. Pollard, *A Monte Carlo method for factorization*, Nordisk Tidskr. Informationsbehandling (BIT) 15 (3) (1995) 331–334.

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

[7] T. Satoh, *Generating genus two hyperelliptic curves over large characteristic finite fields*, in: Advances in cryptology–Eurocrypt, Lecture Notes in Comput. Sci. 5479, Springer, Berlin (2009) 536–553.

[8] 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.

[9] 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.