UNIVERSITY OF CALIFORNIA
SANTA CRUZ

AN ELLIPTIC ANALOGUE OF LANDAU'S FIRST PROBLEM

A thesis submitted in partial satisfaction of the
requirements for the degree of

MASTER OF ARTS

in

MATHEMATICS

by

Nora G. Jones

June 2024

The Thesis of Nora G. Jones
is approved:

Professor Martin Weissman, Chair

Professor Pedro Morales-Alazman

Professor Jiayin Pan

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

An Elliptic Analogue of Landau's First Problem

by

Nora G. Jones

Edmund Landau in 1912 questioned whether there are an infinite number of primes p of the form n^2+1. While this problem was and remains beyond the capabilities of modern mathematics, people have made dedicated efforts to studying distributions of these primes and elliptic analogues of similar number theoretic questions. Here, we look at the elliptic curve analogue of this problem and study the frequency that the number of solutions to an elliptic curve, over a finite field with prime order, is a perfect square. We examine some of the anomalies in the distribution of square solutions, investigate whether this distribution is similar to that conjectured of squares of the form $p-1$, and translate previous conjectures on the frequencies of fixed N_p values to our square problem.

Acknowledgments

I would like to thank my advisor, Professor Martin Weissman for helping me throughout this entire process. Without his incredible help and guidance, as well as the idea for the topic, this paper would not have been possible. I would like to thank Professor Pedro Morales and Professor Jiayin Pan for their help and serving on my reading committee. I would like to thank Deewang Bhamidipati for always taking his own time to talk to me about math and answer any question that I had. I would like to thank Jadyn Breland, for without him I may not have majored in mathematics in my undergrad. Lastly, I thank my wonderful family and friends for their countless support throughout my mathematical journey.

Introduction

Number theoretic questions, specifically those about prime numbers, can be some of the most challenging to prove. Particularly fascinating are the four unsolved problems posed by Edmund Landau in 1912 including the Goldbach Conjecture, the Twin Prime Conjecture, the Legendre Conjecture, and his first problem concerning the existence of infinite near-square primes [8]. He claimed at the time that we have not advanced mathematically enough to prove or disprove any of these questions, and this remains true today. However, because the cardinality of the multiplicative group $\#G_m(\mathbb{F}_p) = p-1$, we may study how the size of G_m behaves as an equivalent statement to Landau's first problem. The rational solutions to elliptic curves have fruitful group structure and thus are often studied as an analogue to these complex number theoretic problems.

We focus our attention to the elliptic curve analogue of Landau's first problem which questioned the existence of infinite near-square primes. We begin by providing a detailed background on elliptic curves including their definition, group structure, relevance in number theory, methods for computing their order, and introduce relevant conjectures on the distributions of their cardinality. We then briefly explore progress

made toward understanding Landau's first problem and then discuss the primary subject of this paper: the elliptic curve analogue of this question, how frequently are the number of solutions to a given elliptic curve over finite fields of varying prime order a perfect square? We analyze how this frequency compares to the frequency of primes p of the form $n^2 + 1$.

In the final section of the paper, we detail the methods and computations used for our study and analyze irregularities in our data compared to our prediction of the behavior of the group orders. We end by computing and predicting constants for the distribution of square N_p values for the set of elliptic curves.

Chapter 1

Preliminaries

We begin by defining an elliptic curve over an arbitrary field and define its properties. These preliminary definitions can be found in [15].

1.1 Elliptic Curves Over General Fields

Definition 1.1.1 (Weierstrass Cubic Curve)**.** Let $x, y \in K$ where K is a field. A *Weierstrass cubic curve*, E, is the graph of a cubic equation of the form

$$y^2 = x^3 + ax + b$$

where a, b are constants from a field K.

Definition 1.1.2 (Singular)**.** A curve E is called *singular* if it has repeated roots, i.e. solutions where $y = 0$. Otherwise, the curve is known as *non-singular*. A curve is *singular* if and only if the discriminant of the Weierstrass cubic curve,

$$\Delta = 4a^3 - 27b^2$$

is non-zero. A singular point of a curve E is a point at which E crosses itself as a node or forms a cusp. E is non-differentiable at its singular point. For cubic curves, those that are singular tend to behave differently than non-singular curves and are therefore excluded from the definition of an elliptic curve.

Definition 1.1.3 (Weierstrass Elliptic Curve)**.** E is a *Weierstrass Elliptic Curve* (or simply *Elliptic Curve*) if it is a non-singular Weierstrass cubic curve.

We denote *the set of solutions to an elliptic curve* E over a field K as $E(K)$ where

$$E(K) = \{\infty\} \cup \left\{(x, y) \in K \times K \mid y^2 = x^3 + ax + b\right\}.$$

∞ is the *point of infinity* which is defined to be the point where the line at infinity in the projective plane intersects the elliptic curve.

1.1.1 The Group Law

The points on an elliptic curve satisfy the axioms of a group. Addition of points is defined geometrically: let P_1, P_2 be two points on an elliptic curve E. A line drawn between these two points will intersect the curve at exactly one more point, P_3'. The point $P_1 + P_2 = P_3$ such that if $P_3' = (x_3, y_3)$, $P_3 = (x_3, -y_3)$. Algebraically, we define addition as follows:

Definition 1.1.4 (The Group Law)**.** Given two points $P_1 = (x_1, y_1), P_2 = (x_2, y_2)$ on an elliptic curve E, we obtain another point on the curve, $P_3 = (x_3, y_3)$. We can define an addition of two solutions to an elliptic curve by the following process: For an elliptic

curve E defined by $y^2 = x^3 + ax + b$ and $P_1, P_2 \neq \infty$, and P_3 defined as below. Then

1. If $x_1 \neq x_2$, $x_3 = m^2 - x_1 - x_2$, $y_3 = m(x_1 - x_2) - y_1$, for $m = \dfrac{y_2 - y_1}{x_2 - x_1}$.

2. If $x_1 = x_2$ but $y_1 \neq y_2$, then $P_1 + P_2 = \infty$.

3. If $P_1 = P_2$ and $y_1 \neq 0$, then $x_3 = m^2 - 2x_1$, $y_3 = m(x_1 - x_3) - y_1$, where $\dfrac{3x_1^2 + a}{2y_1^1}$.

4. If $P_1 = P_2$ and $y_1 = 0$, then $P_1 + P_2 = \infty$. For all points P on E, define $P + \infty = P$.

The solutions P_1, P_2 exist over a field K for $a, b \in K$, therefore P_3 exists over K as defined above and thus $E(K)$ is closed under addition. The solutions $P_i = (x_i, y_i)$ on an elliptic curve E with addition defined as above form an abelian group where ∞ is the identity element. This is known as the *Group Law* of an elliptic curve. One may verify these axioms themselves, or otherwise a proof of this result can be found in [15].

1.2 Elliptic Curves Over Finite Fields

This paper's question concerns itself with our field restricted to one of finite order. Let E be an elliptic curve with integer coefficients; then we may view E as an elliptic curve over any field. Specifically, if we view E as an elliptic curve over a finite field $\mathbb{F}_q$, then then there exists only a finite number of pairs that satisfy the equation of the elliptic curve and thus, the group $E(\mathbb{F}_q)$ has finite order and what will be closely examined in this paper.

The cardinality of $E(\mathbb{F}_q)$ (denoted by $\#E(\mathbb{F}_q)$ or N_q) varies with q as we look at each $\mathbb{F}_q$, and is exactly equal to $|q + 1 - a_q|$ where a_q is an error term dependent

on q. More specifically, as proven by Helmut Hasse in 1933 [4], the error value $|a_q|$ is bounded by $2\sqrt{q}$ which will be discussed more extensively in section 1.2.2. We later restrict viewing E over fields of prime order, denoted $\mathbb{F}_p$ as is necessary for the analogue of Landau's question. We begin this chapter with some definitions that can be found in [10].

Definition 1.2.1 (Good Reduction)**.** Fix an elliptic curve, E with integer coefficients. A prime p is said to be of *good reduction* if E when reduced modulo p and viewed as an elliptic curve over $\mathbb{F}_p$) remains non-singular. Otherwise, p is a prime of *bad reduction.*

Definition 1.2.2 (Multiplication by m)**.** The *multiplication by m map*, $[m]$ is an endomorphism of elliptic curves defined such that for

$$m \in \mathbb{Z}, \quad [m] : E \longrightarrow E, \quad mP \longmapsto P + P + ... + P \ (m - \text{times}).$$

Definition 1.2.3 (m-Torsion Subgroup)**.** $E[m]$, the *m-torsion subgroup of E*, is the set of points in $E(\overline{\mathbb{Q}})$ that have order m:

$$E[m] = \{P \in E \mid [m]P = \infty\}.$$

Most of the time, these multiplication by m maps will be the only elements of the endomorphism ring of the complex points on E, $E(\mathbb{C})$. However, occasionally the endomorphism ring will be larger than $\mathbb{Z}$ which forces extra symmetries upon E.

Definition 1.2.4 (Complex Multiplication)**.** An elliptic curve E has complex multiplication (commonly denoted as CM) if there exists an endomorphism $\varphi\colon E \to E$ which is not a multiplication by m map. We illustrate this definition with an example.

Example 1.2.1. The elliptic curve, E, defined by $y^2 = x^3 + x$ has complex multiplication map $\varphi(x, y) = (-x, iy)$ as

$$(iy)^2 = -y^2 = -x^3 - x = (-x^3) + (-x)$$

which is an endomorphism that is not multiplication by m.

We will later see examples of how elliptic curves with complex multiplication behave differently than those without complex multiplication.

1.2.1 Galois Representations & The Tate Module

We proceed by defining the Tate Module, which is critical in our understanding of the behavior of N_p. These definitions may be found in chapter 3.7 of [10]. Fix an elliptic curve, E with coefficients in $\mathbb{Z}$.

Definition 1.2.5 (mod-m Galois Representation)**.** The elements of the Galois group, $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, act on $E[m]$ by automorphisms giving a homomorphism

$$\rho_{E,m} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}(E[m]).$$

When viewing E over $\mathbb{F}_p$, then $E[m] \cong (\mathbb{Z}/m\mathbb{Z}) \times (\mathbb{Z}/m\mathbb{Z})$. We obtain a 2-dimensional representation denoted $\bar{\rho}_{E,m}$ such that

$$\bar{\rho}_{E,m} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL_2(\mathbb{Z}/m\mathbb{Z}).$$

Definition 1.2.6. [mod-ℓ Galois Representation] Let ℓ, be a prime. Then we may modify $\bar{\rho}$ to obtain a representation such that

$$\bar{\rho}_{E,\ell} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \to GL_2(\mathbb{Z}/\ell\mathbb{Z}) \cong GL_2(\mathbb{F}_\ell).$$

Definition 1.2.7. Let E be an elliptic curve without complex multiplication. The map $\rho_{E,\ell}$ is *maximal* if $\mathrm{im}(\rho_{E,\ell}) = GL_2(\mathbb{F}_\ell)$.

By Serre's open image theorem [9], the representation $\rho_{E,\ell}$, where E does not have complex multiplication, is maximal for all but finitely many primes ℓ where the image is a finite index subgroup of $GL_2(\mathbb{F}_\ell)$. It has been conjectured in [12] that there are a fixed set of subgroups that $\mathrm{im}(\rho_{E,\ell})$ is either equal or conjugate to.

Definition 1.2.8. Let E be an elliptic curve and ℓ^n a prime-power. The *(ℓ-adic) Tate module* of E, denoted by $T_\ell(E)$ is defined by the group

$$T_\ell(E) = \varprojlim_n E[\ell^n]$$

where the inverse limit respects the canonical maps $E[\ell^{n+1}] \xrightarrow{[\ell]} E[\ell^n]$. Each $E[\ell^n]$ is a $\mathbb{Z}/\ell^n\mathbb{Z}$-module. Therefore by taking the inverse limit, the Tate module $T_\ell(E)$ is in fact a $\mathbb{Z}_\ell$-module.

Now, $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ also acts on the Tate module as the $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$-action on each $E[\ell^n]$ commutes with the multiplication by ℓ maps in the inverse limit. By choosing a $\mathbb{Z}_\ell$-basis for $T_\ell(E)$, we obtain a 2-dimensional representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ as follows.

Definition 1.2.9. We define the ℓ-adic representation of $\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ on E as the map

$$\rho_{E,\ell} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \longrightarrow \mathrm{Aut}(T_\ell(E)) \longrightarrow GL_2(\mathbb{Z}_\ell).$$

Definition 1.2.10. $\mathbb{Z}_\ell$ surjects onto $\mathbb{Z}/\ell\mathbb{Z} \cong \mathbb{F}_\ell$. Therefore, there exists a representation

$$\overline{\rho}_{E,\ell} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(\mathbb{F}_\ell).$$

For our purposes as we are only considering fields of prime order (as opposed to prime powers), we most of the time need only to consider the map $\bar{\rho}$ defined in 1.2.6 as ℓ is always prime.

Definition 1.2.11 (Frobenius Endomorphism)**.** If $E(K)$ defines an elliptic curve, we can apply the qth power map to $E(K)$ and define, the *Frobenius morphism* by

$$\varphi_q : E(K) \longrightarrow E^q(K), \quad \varphi_q(x, y) = (x^q, y^q).$$

Now let $K = \mathbb{F}_p$. We see that

$$\varphi_p : E(\mathbb{F}_p) \longrightarrow E(\mathbb{F}_p).$$

This map is called the *Frobenius endomorphism* of E. For all primes p, these maps lift to form the set of Frobenius elements $\text{Frob}_p \in \text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ which is dense in $\text{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$ by the Frobenius Density Theorem.

1.2.2 The Hasse Bound & Computing N_p

As introduced above, the cardinality of $E(\mathbb{F}_q)$ (denoted by $\#E(\mathbb{F}_q)$ or N_q) varies for each q, but is approximately equal to $q + 1$ minus an error value. More specifically, by Hasse [4]

$$\#E(\mathbb{F}_q) = q + 1 - a_q.$$

This error value a_q is also known as the *Frobenius Trace* and is bounded by $2\sqrt{q}$. This bound on a_q, known as the "Hasse Bound", was conjectured by Emil Artin, and proved by Helmut Hasse in 1933. We provide some definitions in order to prove this theorem which can be found in [10] and [15].

Theorem 1.2.1 (Hasse's Theorem on Elliptic Curves)**.** Let E be an elliptic curve over $\mathbb{F}_q$. Then, $\#E(\mathbb{F}_q)$ satisfies the bound

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

where the error-value for $E(\mathbb{F}_q)$, $a_q = q + 1 - \#E(\mathbb{F}_q)$. When q is restricted to a prime p, we are left with a strict inequality such that

$$-2\sqrt{p} < a_p < 2\sqrt{p}.$$

Theorem 1.2.2. Fix an elliptic curve E and let p, ℓ be primes of good reduction. Then

$$\mathrm{tr}(\bar{\rho}_{E,\ell}(\mathrm{Frob}_p)) = a_p \quad \text{and} \quad \det(\bar{\rho}_{E,\ell}(\mathrm{Frob}_p)) = p.$$

1.2.3 Algorithms for Computing N_p

There are many algorithms for computing the order of $E(\mathbb{F}_q)$, the least complex method being the use of Legendre symbols. More details from these definitions can be found in [15].

Definition 1.2.12 (Legendre Symbol)**.** Given an odd prime p, the Legendre Symbol, $\left(\frac{x}{p}\right)$ is defined as follows:

$$\left(\frac{x}{p}\right) = \begin{cases} +1 & \text{if } t^2 \equiv x \pmod{p} \text{ has a solution for t} \not\equiv 0 \pmod{p}, \\ -1 & \text{if } t^2 \equiv x \pmod{p} \text{ has no solution,} \\ 0 & \text{if } x \equiv 0 \pmod{p} \end{cases}$$

Theorem 1.2.3. Let E be defined by $y^2 = x^3 + ax + b$ over a field $\mathbb{F}_p$ for some prime p. Then

$$N_p = p + 1 + \sum_{x \in \mathbb{F}_p} \Big(\frac{x^3 + ax + b}{p}\Big).$$

Proof. For a field $\mathbb{F}_p$ where p is an odd prime, we have that $t^2 = x$ has 0, 1, or 2 solutions. Therefore, there are two points (x_0, y) such that $x_0^3 + ax_0 + b = z$ if z is a non-zero square in $\mathbb{F}_p$, one point if $x_0^3 + ax_0 + b = 0$, and zero points if z is not a square. Thus, the number of points such that (x_0, y) is a solution for a given $x_0 \in \mathbb{F}_p$ is equal to $1 + \Big(\frac{x_0^3 + ax_0 + b}{p}\Big)$. The number of solutions N_p summing over all $x \in \mathbb{F}_p$, adjoin the point of infinity, is

$$\begin{aligned} N_p &= 1 + \sum_{x \in \mathbb{F}_p} \Big(1 + \Big(\frac{x^3 + ax + b}{p}\Big)\Big) \\ &= p + 1 + \sum_{x \in \mathbb{F}_p} \Big(\frac{x^3 + ax + b}{p}\Big). \end{aligned}$$

□

Following directly from Hasse's Theorem, we have that

$$\left| \sum_{x \in \mathbb{F}_p} \Big(\frac{x^3 + ax + b}{p}\Big) \right| \leqslant 2\sqrt{p}$$

and that

$$a_p = -\sum_{x \in \mathbb{F}_p} \Big(\frac{x^3 + ax + b}{p}\Big) \quad \text{for} \quad -2\sqrt{p} \leqslant a_p \leqslant 2\sqrt{p}.$$

The primary algorithm used to gather evidence for the question at hand involves computing a_p values and using those to find each N_p value.

For large primes, it is computationally quicker to first compute the a_p values and then compute N_p using those values. Brute force computations of these a_p values

take $O(p^{1+o(1)})$ time which is efficient enough for primes, $p < 10000$ [2]. However for larger primes, more efficient algorithms need to be implemented. Shanks's "Baby Step, Giant Step" is an algorithm for computing the order, h, of a cyclic group G. We can modify this algorithm for the specific case of solutions to elliptic curves E over $\mathbb{F}_q$ as found in [2].

Algorithm 1.2.1 (Baby Step, Giant Step (Modified))**.** Begin with a cyclic group, G of order h with identity 1, and an inequality $B/2 < C \leqslant h \leqslant B$. Let $q = p^n$ for some prime, p. The generalized form of this algorithm can be found in [2]. In our case, G is $E(\mathbb{F}_q)$ of order N_q with the identity $1 = \infty$. Hasse's theorem gives us a bound on N_q values such that

$$q + 1 - 2\sqrt{q} \leqslant N_q \leqslant q + 1 + 2\sqrt{q}.$$

Therefore, with $C = q+1-2\sqrt{q}$ and $B = q+1+2\sqrt{q}$, Shanks's algorithm modified for the group $E(\mathbb{F}_q)$ by [15] proceeds as follows:

1. Begin by computing $Q = (p+1)P$.

2. Let $m = \lceil q^{1/4} \rceil$. Compute the points jP for $0 \leqslant j \leqslant m$.

3. Then compute the points $Q + k(2mP)$ for $-m \leqslant k \leqslant m$ and stop when $Q + k(2mP) = \pm jP$ from step (2).

4. We then can conclude that $(q+1+2mk \mp j)P = \infty$. Then let $M = q+1+2mk \mp j$.

5. Factor M such that $p_1, \cdots, p_r$ are the unique prime factors of M.

6. Then calculate $(M/p_i)P$ for $1 \leqslant i \leqslant r$. If $(M/p_i)P = \infty$ for some i, then $M \leftarrow (M/p_i)$ and return to step (5). If $(M/p_i)P \neq \infty$, for all i, then the point M has order M.

7. To find N_q, repeat steps (1)-(6) for random points $(x, y) \in E(\mathbb{F}_q)$ until the greatest common multiple of the orders divides only one integer N such that $q+1-2\sqrt{q} \leqslant N \leqslant q+1+2\sqrt{q}$. Then $N = N_q$.

This version of the algorithm combines elements from Shanks's original Baby Step, Giant Step algorithm with modifications introduced by Jean-François Mestre to compute the values a_p. A specific algorithm for computing a_p values can be found in [2].

Another algorithm for computing the number of solutions to an elliptic curve was published by René Schoof in 1985 and allows for significantly faster computations for very large primes p. However, this is not immediately relevant to the computations in this paper and thus will not be looked at into depth. Details on Schoof's Algorithm may be found in [15].

1.2.4 Sato-Tate Conjecture

The Sato-Tate conjecture, posed independently by Mikio Sato and John Tate around 1960 and proved in 2008 in a series of three articles [13][1][3], is an important statistical assertion on a family of elliptic curves $E(\mathbb{F}_p)$ over functionally all primes p. As described in [7], let E be an elliptic curve with integer coefficients so it can be reduced

mod p for each $\mathbb{F}_p$. Write the Hasse bound as

$$N_p/p = 1 + O(1/\sqrt{p}).$$

Then the conjecture describes the distribution of the term O as $p \to \infty$ for a curve without complex multiplication.

Theorem 1.2.4 (Sato-Tate Conjecture)**.** Let θ_p be a solution to the equation

$$p + 1 - N_p = 2\sqrt{p}\cos\theta_p$$

for $0 \leqslant \theta_p \leqslant \pi$. Then for all fixed angles α, β such that $0 \leqslant \alpha \leqslant \beta \leqslant \pi$,

$$\lim_{X\to\infty} \frac{\#\{p \leqslant X : \alpha \leqslant \theta_p \leqslant \beta\}}{\pi(X)} = \frac{2}{\pi}\int_\alpha^\beta \sin^2 t\, dt.$$

The number of solutions N_p may now be described by the formula

$$N_p = p + 1 + 2\sqrt{p}\cos\theta_p$$

where the elliptic curve angles, θ_p lie within the interval $[0, \pi]$ via a $\sin^2\theta$ distribution.

We can see this theorem demonstrated with the following example:

Example 1.2.2. Let E be an elliptic curve defined by $y^2 = x^3 + 2x + 6$. We can see by plotting values of θ_p, for primes up to 1,000,000, against the frequency of N_p, $\theta_p \in [0, \pi]$ and θ_p is distributed via a $\sin^2\theta$ function (overlaid in red on the plot below).

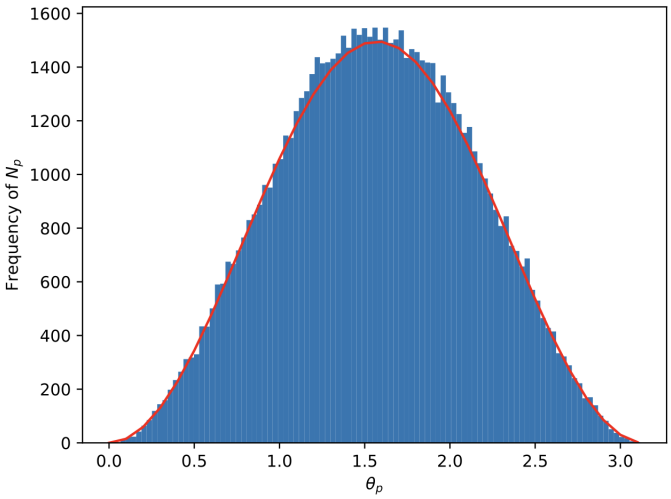

Figure 1.1: Graph of Sato-Tate Distribution of θ_p for the curve: $y^2 = x^3 + 2x + 6$

As mentioned previously, the distribution of angles θ_p is completely different for a curve with complex multiplication.

Example 1.2.3. Let E be an elliptic curve defined by $y^2 = x^3 + x$. The distribution of θ_p for primes up to 1,000,000, in this figure below is clearly not that of a $\sin^2\theta$ function.

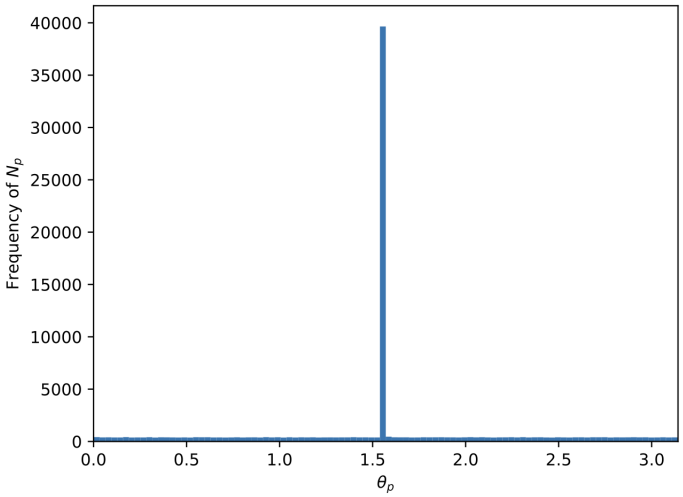

Figure 1.2: Distribution of θ_p for the curve: $y^2 = x^3 + x$

1.2.5 Lang-Trotter Conjecture

In 1976, Serge Lang and Hale Trotter formed a conjecture predicting the asymptotic number of primes with a fixed a_p value in [5]. The theorem is stated as follows:

Theorem 1.2.5 (Lang-Trotter Conjecture)**.** Let E be and elliptic curve over $\mathbb{Q}$ and a_p defined as above for primes p of good reduction. Fix an integer a and define

$$\pi_{E,a}(X) = \#\{p : p \leqslant X \text{ and } a_p(E) = a\}.$$

Then there is a constant dependent on E and a, $C_{E,a}$ such that

$$\pi_{E,a}(X) = C_{E,a}\frac{\sqrt{X}}{\log X}.$$

Let m be an integer and ℓ a prime. By Lang-Trotter, this constant (for all but finitely many primes ℓ) is described to be

$$C_{E,a} = \frac{2}{\pi} \cdot m \cdot \frac{\left|\bar{\rho}_{E,m}(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}))_a\right|}{\left|\bar{\rho}_{E,m}(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}))\right|} \prod_{\ell \nmid m} \frac{\left|\bar{\rho}_{E,\ell}(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}))_a\right|}{\left|\bar{\rho}_{E,\ell}(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}))\right|},$$

where $\bar{\rho}_{E,m}(\mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}))_a$ are the matrices with trace a.

Chapter 2

The Elliptic Analogue of Landau's First Problem

2.1 Landau's Original Question

At the 1912 International Congress of Mathematicians, Edmund Landau presented a survey of the history of studies on prime numbers and the Riemann-zeta function as well as four problems that he claimed to be unsolvable with the current mathematical knowledge of the time. Today, all four problems have yet to be solved. As originally presented [8], the problems are:

1. Does the function $u^2 + 1$ represent infinitely many primes for integer values of u?

2. (The Goldbach Conjecture) Does the equality $m = p + p'$ have for any even $m > 2$ a solution?

3. (The Twin Prime Conjecture) Does the equality $2 = p - p'$ have infinitely many

solutions in primes?

4. (The Legendre Conjecture) Does there always exist at least one prime between consecutive perfect squares?

We will focus on the first statement questioning the existence of primes of the form $n^2 + 1$. In 1923, Hardy and Littlewood conjectured that the number of primes of the form $n^2 + 1 \leqslant X$ is asymptotically equal to

$$\prod_{p>2} \Big(1 - \frac{1}{p-1}\Big) \frac{\sqrt{X}}{\log X}.$$

This has yet to be proven, but there has been evidence gathered for primes up to at least 10^{20} [16] and appears to be supported. In our case, because we are looking at a relatively small number of primes (up to $1,000,000$), we may calculate directly how many of these primes are of the form $n^2 + 1$ and compare this to the elliptic analogue of this question.

2.2 The Elliptic Analogue

For the multiplicative group G_m, $\#G_m(\mathbb{F}_p) = p - 1$ hence, asking Landau's question is equivalent to asking the question of whether there are an infinite number of primes p such that $\#G_m(\mathbb{F}_p)$ is square. As discussed extensively above, the group of points on an elliptic curves has useful structure and are often used as an analogue to study number theoretic problems such as this one. The elliptic analogue of Landau's First problem is as follows: Define an elliptic curve, E with integer coefficients, where

$E(\mathbb{F}_p)$ is the number of solutions to E over each $\mathbb{F}_p$. How often is $\#E(\mathbb{F}_p) = N_p$ a perfect square? We hypothesize that this frequency will mimic the frequency that $p-1$ is a square as described above. We will use this as a baseline to compare our results.

2.3 Methods

The computations below were performed using SageMath 10.0 [14]. We begin by describing the process for determining the frequency of square N_p values.

Let a_p be defined as above. For a fixed elliptic curve E, the SAGE function `.aplist(X)` generates a list of the a_p values for primes p up to an integer X using algorithms described in section 1.2.3. Recall by Hasse's Theorem, we know the values a_p will always lie between the bounds

$$-2\sqrt{p} < a_p < 2\sqrt{p}$$

or equivalently

$$-1 < \frac{a_p}{2\sqrt{p}} < 1.$$

Example 2.3.1. Consider the curve E defined by $y^2 = x^3 + 2x + 6$. We may plot a histogram of the frequency values of $\frac{a_p}{2\sqrt{p}}$.

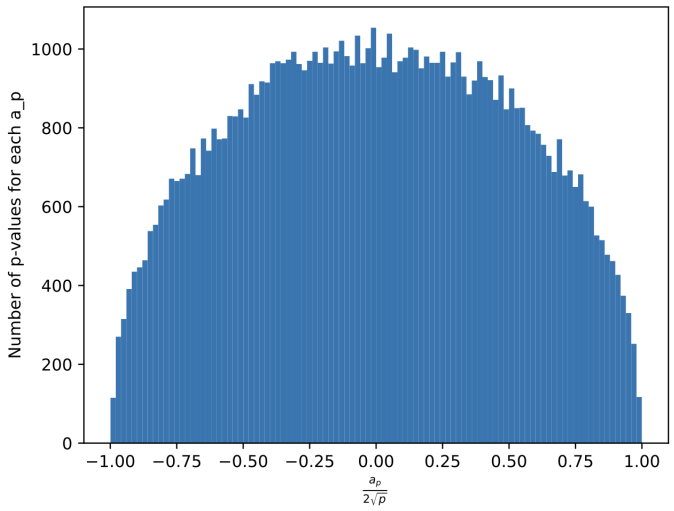

Figure 2.1: Histogram Plot $\frac{a_p}{2\sqrt{p}}$ for the curve: $y^2 = x^3 + 2x + 6$

We can see that the values for $\frac{a_p}{2\sqrt{p}}$ lie within $(-1, 1)$, therefore $a_p \in (-2\sqrt{p}, 2\sqrt{p})$, verifying Hasse's Theorem.

Previously, we described how elliptic curves with complex multiplication have extra symmetries that force them to behave differently. Below is an example plot affirming the Hasse bound for a curve that has complex multiplication.

Example 2.3.2. Let E be a curve defined by $y^2 = x^3 + x$. We can graph a histogram plotting again $\frac{a_p}{2\sqrt{p}}$ against the frequency of each a_p value.

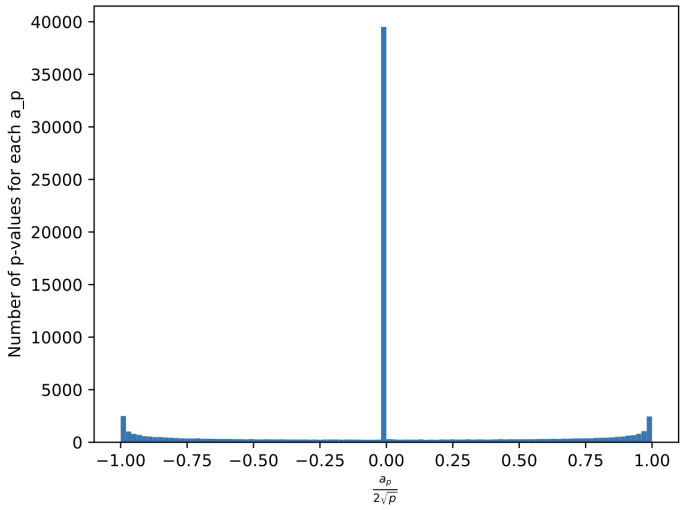

Figure 2.2: Histogram Plot $\frac{a_p}{2\sqrt{p}}$ for the curve: $y^2 = x^3 + x$

Although the elliptic curve still satisfies the Hasse bound, it does not have the same even distribution pattern as the example without complex multiplication (2.3.1) does.

We easily expand this calculation of a_p values to a larger amount of elliptic curves by cycling through tuple values of coefficients (a, b) of the curve defined $E : y^2 = x^3 + ax + b$. We let the (a, b) values vary such that $(a, b) \in [1, 50] \times [1, 50]$, generating a list of 2500 elliptic curves, none of which have complex multiplication. We compute the N_p values for each curve for primes up to X, for X values 10,000, 100,000, and 1,000,000.

2.4 Preliminary Observations

Depicted below are three histograms graphing the number of square N_p values for primes up to 10,000, 100,000, and 1,000,000 for all of the elliptic curves. Along with the histogram, in red we have depicted the mean value of the N_p squares amongst all of the curves. In yellow, we have the the value of the classical Landau problem: for $X \in \mathbb{Z}^+$, how many primes $p < X$ are such that $p - 1$ is a perfect square?

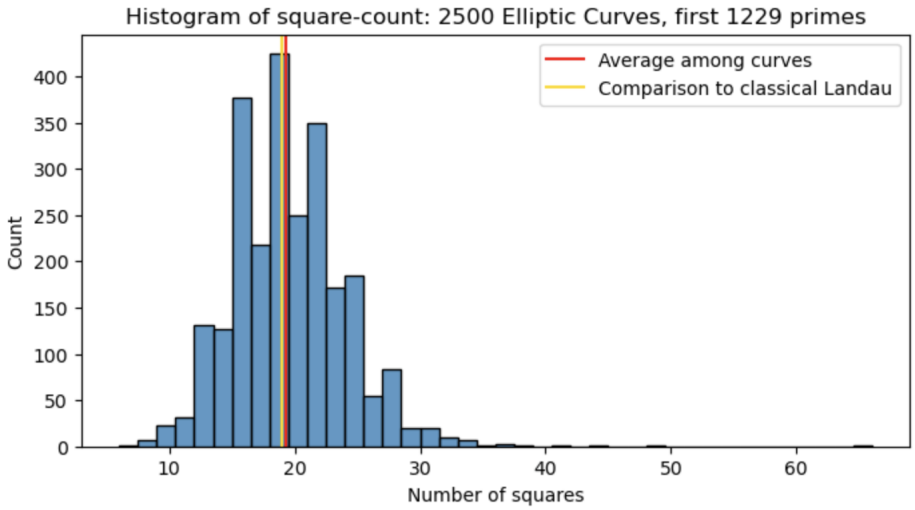

Figure 2.3: N_p Square Count: Primes up to 10,000

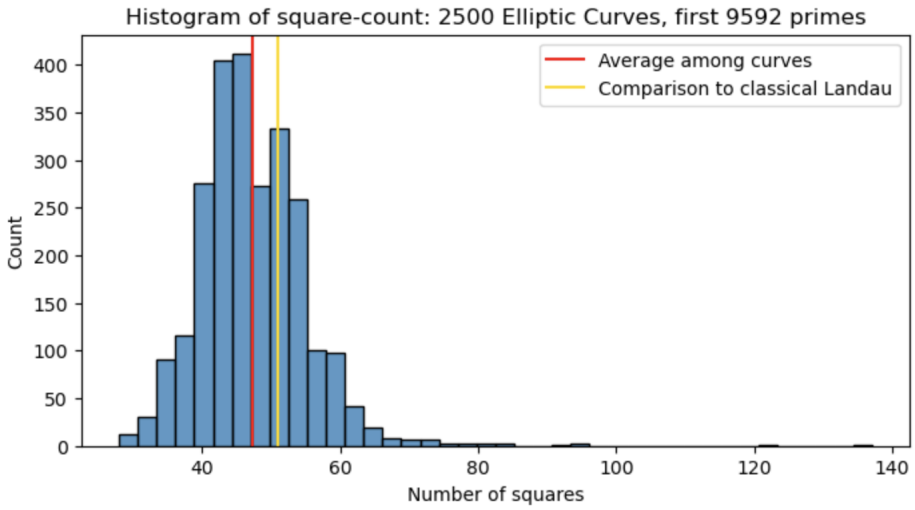

Figure 2.4: N_p Square Count: Primes up to 100,000

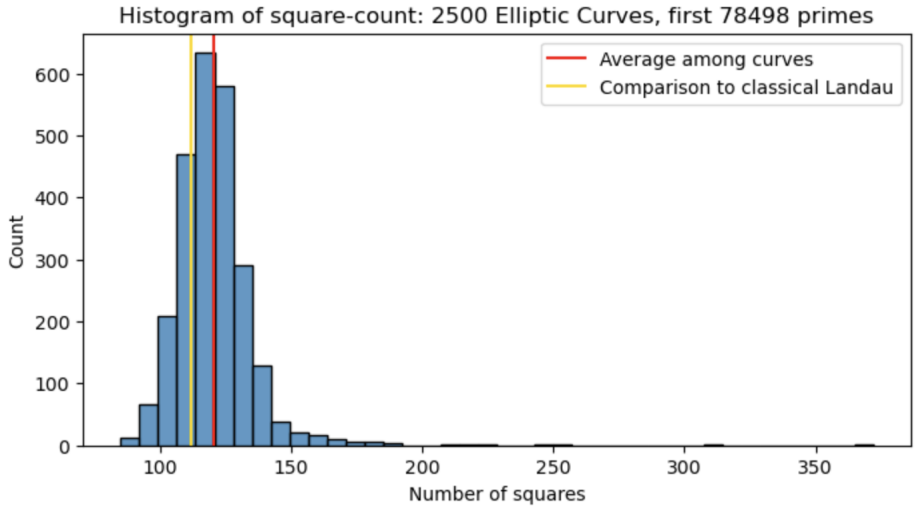

Figure 2.5: N_p Square Count: Primes up to 1,000,000

We initially observe that all three histograms are right-skewed with several outliers. In all three graphs, we observe the most extremal curve, consistent for all data

sets, had coefficients $(6, 7)$ with N_p values as follows:

Number of N_p squares for $(6, 7)$			
	$p < 10,000$	$p < 100,000$	$p < 1,000,000$
$(6, 7)$	66	137	372
avg. N_p squares amongst all curves	19.32	47.38	120.49

Figure 2.6: N_p square count for extremal curve $(6, 7)$

The curve with coefficients $(6, 7)$ had on average 3.09 times as many squares when compared to the average number of N_p squares over all 2500 elliptic curves. The minimum values are not consistent across all three curves. Below is a table illustrating the average values of primes of the form $n^2 + 1$ compared to the N_p squares for primes up to 10,000, 100,000, and 1,000,000.

Comparing frequency of the N_p and $p - 1$ squares		
X	avg. N_p squares amongst all curves	$p - 1$ squares
$10,000$	19.32	19
$100,000$	47.38	51
$1,000,000$	120.49	112

Figure 2.7: Average N_p counts for all curves compared to classical Landau.

2.5 The Reducible Outliers

These observations lead to questions on what causes a higher probability that an elliptic curve will have an N_p value that is a square. We began by examining the curve with coefficient $(6, 7)$. This elliptic curve, in fact is reducible as well as several more of the outlier curves in the data set. By reducible, we mean for E, the polynomial $x^3 + ax + b$ can be factored over $\mathbb{Q}$. Upon removing the 77 reducible curves from the data set, the histogram appears as follows in Figure 3.

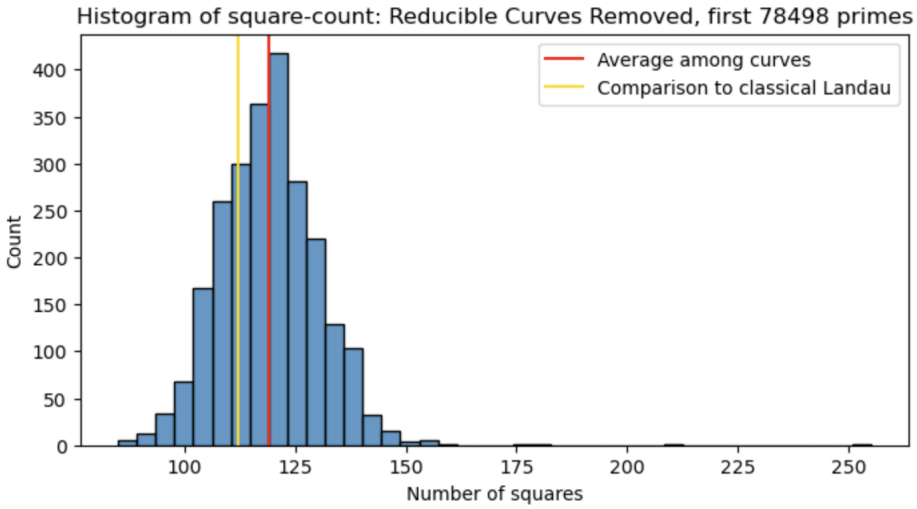

Figure 2.8: N_p Square Count: Primes up to 1,000,000

We observe that the distribution is less right-skewed than the original data set, however there still remains a right tail of extremal values that the reducible case clearly did not resolve.

For each reducible elliptic curve in the data set, we observed that for each

mod-ℓ representation, the image of $\rho_{E,\ell}$ was maximal except for $\ell = 2$. This can be explained by the restrictions that a reducible elliptic curve puts on the points in $E[2]$. Let $E : y^2 = x^3 + ax + b$. There are two cases such that E is reducible:

(i) E is partially reducible where $E : y^2 = (x^2 + Ax + B)(x + C)$.

(ii) E is fully reducible where $E : y^2 = (x + A)(x + B)(x + C)$.

Let E be fully reducible. Then the points $(A, 0), (B, 0), (C, 0) \in E(\mathbb{Q})$ and as by the rational root theorem, if A is a rational root of E, then it is in fact a root over the integers and is a factor of b. These points are in fact all of the 2-torsion points of E [11] and $E(\mathbb{Q})[2] \subset E(\overline{\mathbb{Q}})[2]$. Now observe the 2-adic Galois representation of E,

$$\bar{\rho}_{2,E} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(\mathbb{F}_2).$$

But $GL_2(\mathbb{F}_2) = \mathrm{Aut}(E(\overline{\mathbb{Q}})[2])$, so $\bar{\rho}_{2,E}$ acts trivially on the 2-torsion points of $E(\mathbb{Q})$ and thus $\mathrm{im}(\bar{\rho}_{2,E}) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ which is not maximal.

Now let E be partially reducible. Then the point $(C, 0) \in E(\mathbb{Q})$ and is a 2-torsion point of E. Because $(C, 0)$ is the only point of order 2, $\bar{\rho}_{2,E}$ must map to a subgroup containing one point of order 2 in $GL_2(\mathbb{F}_2)$. There is only one such point in $GL_2(\mathbb{F}_2)$ and so $\mathrm{im}(\bar{\rho}_{2,E}) = \{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\}$ which again, is not maximal. Thus, a reducible elliptic curve will never have a maximal image in the mod-2 Galois representation. This condition affects the distribution of square N_p values as will be described later.

2.6 Congruences of the Other Outliers

After removing the reducible curves from the data set, the furthest outlier curve had coefficients $(21, 26)$. Upon some examinations of the congruences of the N_p values for the curve (i.e. how the N_p values behave modulo different integers), we found that this curve has no N_p values that are equal to 2 (mod 3) for all primes up to $1,000,00$. In fact, all of the curves that lied in the right tail of the of the histogram had the same congruence issue with no N_p values equalling 2 (mod 3). This congruence does not always affect the squares as there are curves with this condition that does not appear to have an abnormal amount of these N_p squares. Below is a table detailing the coefficients whose elliptic curves had no N_p values equalling 2 (mod 3).

Coefficient	N_p **Square Count**
(21, 26)	255
(24, 2)	209
(15, 23)	180
(45, 18)	176
(9, 2)	155
(12, 8)	153
(36, 16)	151
(27, 27)	146
(12, 37)	138
(39, 23)	136
(15, 11)	128
(9, 18)	121
mean	120.49

Table 2.1: N_p Square Count for elliptic curves with $N_p \neq 2(\text{mod } 3)$

Upon removing these curves from the data set, we obtain the following graph:

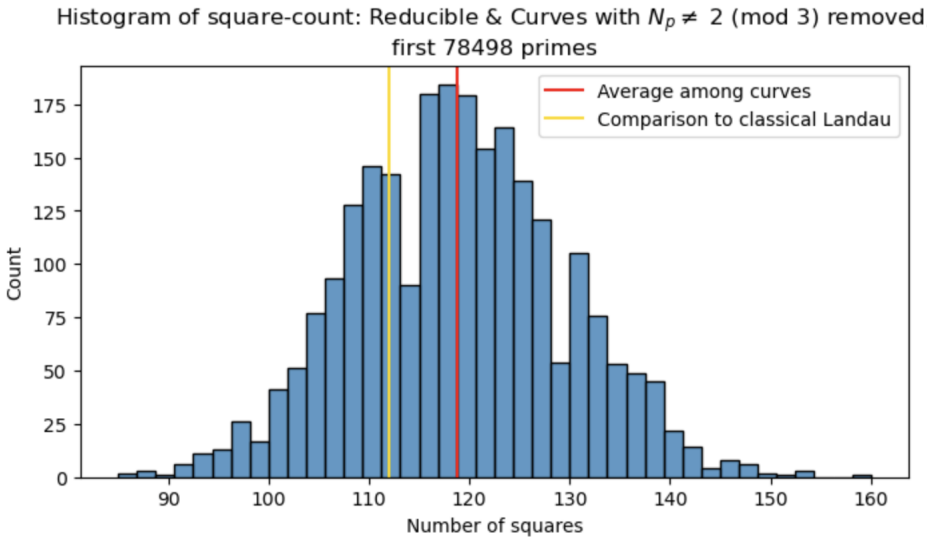

Figure 2.9: N_p Square Count: Primes up to 1,000,000

The squares modulo 3 are 0 and 1, therefore if N_p is always square modulo 3 (or for all but one prime), this increases the likely hood of a square N_p over $\mathbb{Z}$. An immediate question following this observation is why these curves have these odd condition regarding the values of $N_p \pmod 3$. Reducing the value of $N_p \pmod \ell$ determines the mod-ℓ Galois representation we observe.

Recall in 1.2.6, for prime ℓ of good reduction, there exists a mod-ℓ Galois representation

$$\overline{\rho}_{E,\ell} : \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \longrightarrow GL_2(\mathbb{F}_\ell).$$

Reducing the value of $N_p \pmod \ell$ determines the mod-ℓ Galois representation we observe.

Examining the mod-ℓ Galois representation of $(21, 26)$, $\bar{\rho}$ has a maximal image

except for when $\ell = 3$, the prime at which there are odd congruence occurrences. According to LMFDB [6, Elliptic Curve 432.b3], $\mathrm{im}(\bar{\rho}_3)$ is equal to the subgroup generated by the matrices

$$\begin{bmatrix} 1 & 0 \\ 0 & 2 \end{bmatrix}, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

or equivalently when reduced modulo 3 the group of matrices,

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

which are in fact the the Klein-four group. The $\mathrm{im}(\rho_{E,\ell}(\mathrm{Frob}_p)) \subset \mathrm{im}(\rho_{E,\ell})$, thus $\bar{\rho}_\ell(\mathrm{Frob}) \in \begin{bmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{bmatrix}$. Recall that $a_p = \mathrm{tr}(\rho_3(\mathrm{Frob}_p))$ and $p = \det(\rho_3(\mathrm{Frob}_p))$. We have that $N_p = p + 1 - a_p$, thus

$$N_p = \det(\rho_3(\mathrm{Frob}_p)) + 1 - \mathrm{tr}(\rho_3(\mathrm{Frob}_p))$$

for all primes p. The traces of the Klein-four matrices are $\{0, -2, 0, 2\}$ and the determinants are $\{-1, 1, -1, 1\}$. Thus when reduced modulo 3, $N_p((21, 26)) = \{0, 1\}$ for all p, forcing N_p to always be square mod 3, increasing the likely hood of a square elsewhere mod p.

2.7 The Distribution of Square N_p Values

After removing both the reducible and the elliptic curves, E, such that $\bar{\rho}_{E,3}$ is not maximal, the histogram appears to have a normal distribution. We see that when comparing the average N_p-square count to the number of squares in the classical Landau question, it is skewed to the right by some factor. Utilizing the work of Lang-Trotter,

we propose below a constant $C_{E,a}$ such to describe the distribution of square N_p values. As introduced above in section 1.2.5 the original Lang-Trotter conjecture proposes that the distribution of a_p-values as p varies for a fixed trace follows $\pi_{E,a}(X) = C_{E,a}\dfrac{\sqrt{X}}{\log X}$.

We conjecture that the square counting function will be some factor $F_{\ell,E}$ of $\dfrac{\sqrt{X}}{\log X}$ where $F_{\ell,E}$ is adjusting to the case of a moving target of a square a.

We estimate this constant using the original data set without removing the outliers as seen in figure 2.4. We multiply the N_p square counts of each of the 2500 curves and multiply by a factor of $\dfrac{\log X}{\sqrt{X}}$. It appears that the size of the estimated constant correlates to a small ℓ-adic image for a small prime ℓ which corresponds to a large N_p square count. Below is a graph of the growth of N_p squares for some chosen curves—those with and without maximal image mod ℓ for all ℓ—compared to the growth of $p - 1$ squares.

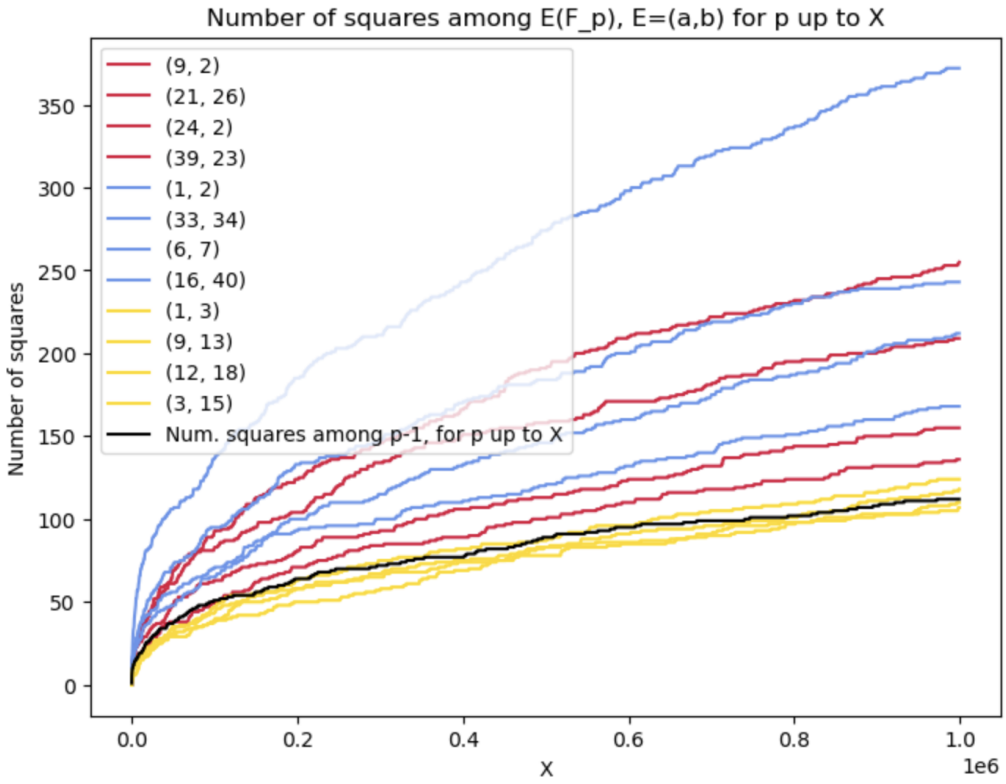

Figure 2.10: Distribution of N_p squares compared to classical Landau. In red are a selection of curves with small 3-adic image, blue are a selection of curves that are reducible, and in yellow are a selection of curves with maximal image at all primes ℓ.

We see in the graph above that the elliptic curves with maximal image have a similar distribution of squares, similar to that of classical Landau. There is more variation in the distribution of the elliptic curves that are reducible (a small 2-adic image) and the elliptic curves with a small 3-adic image. Below is a table describing the estimated constants for the curves in the graph above, as well as their N_p square count for primes up to 1,000,000.

Coefficient	**Estimated Constant**	**Num. N_p squares for p up to1 mil.**	**Primes l where $\mathrm{im}(\rho_{E,\ell})$ is not maximal**
(6, 7)	5.139	372	2
(21, 26)	3.523	255	3
(1, 2)	3.357	243	2
(33, 34)	2.943	213	2
(24, 2)	2.887	209	3
(16, 40)	2.321	168	2
(9, 2)	2.141	155	3
(39, 23)	1.879	136	3
(1, 3)	1.713	124	—
(3, 15)	1.630	118	—
(12, 18)	1.520	110	—
(9, 13)	1.478	107	—
$p-1$	1.547	—	—

Table 2.2: Estimated Constants, N_p Square Counts, and ℓ-adic image for select elliptic curves.

We then chose to examine the exact ℓ-adic images of chosen elliptic curves and compared the estimated constants to those of the same images. Many of the elliptic curves in the graphs below are not in the original data set as there is a lot of variation in the possibilities for the ℓ-adic images. These images are labeled with those used by LMFDB [6]. Below is a graph of elliptic curves with non-maximal, but varying 2-adic images.

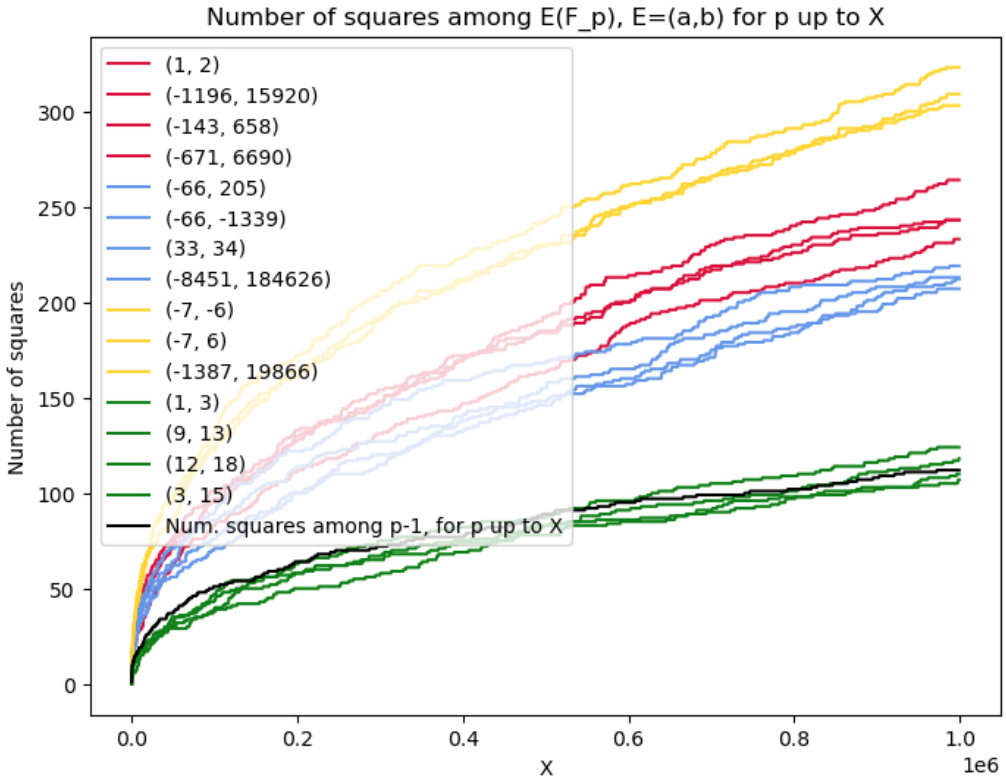

Figure 2.11: Distribution of N_p squares of curves with non-maximal 2-adic image compared to classical Landau. In red are curves with 2-adic image 8.24.0.50 [6, Elliptic Curve 56.a4], in blue are a selection of curves with 2-adic image 4.12.07 [6, Elliptic Curve 360.e4], in yellow are a selection of curves with 2-adic image 8.49.0.39 [6, Elliptic Curve 40.a2], and in green are a selection of curves with maximal image at all primes ℓ.

Below is a graph of elliptic curves with non-maximal, but varying 3-adic images.

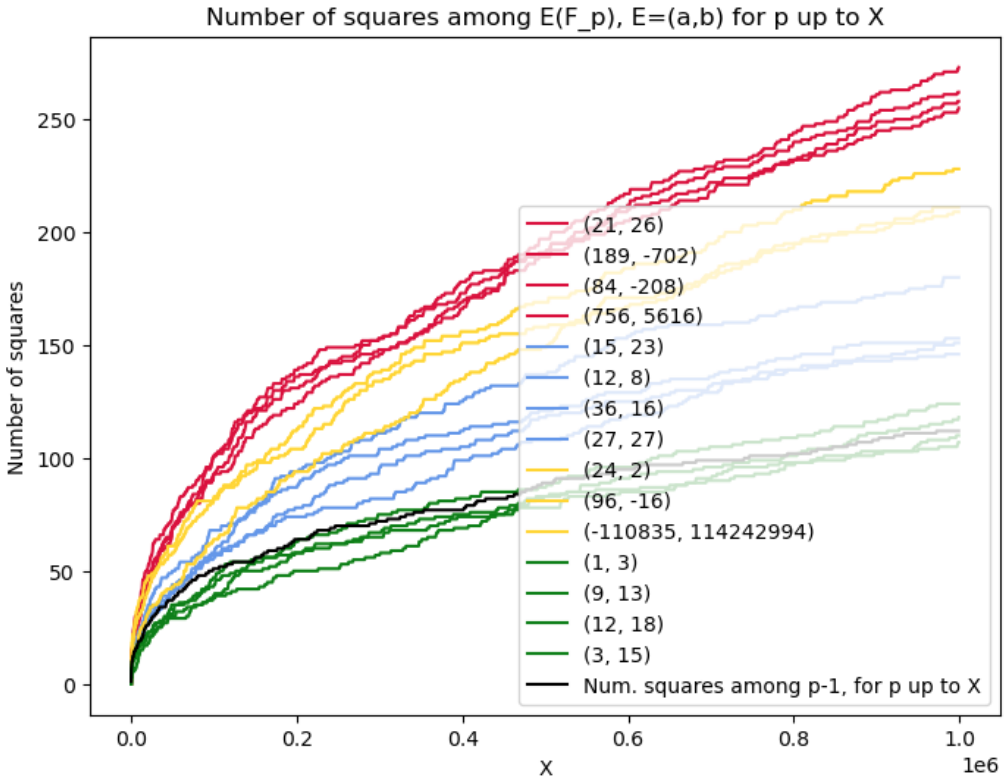

Figure 2.12: Distribution of N_p squares of curves with non-maximal 3-adic image compared to classical Landau. In red are curves with 3-adic image 9.36.0.1 [6, Elliptic Curve 432.b3], in blue are a selection of curves with 3-adic image 3.4.0.1 [6, Elliptic Curve 9072.m2], in yellow are a selection of curves with 3-adic image 27.36.0.1 [6, Elliptic Curve 10944.ck3], and in green are a selection of curves with maximal image at all primes ℓ.

With these two graphs, observe that elliptic curves with identical ℓ-adic images have extremely similar distribution of squares.

Explicitly computing the constant is the immediate next question to consider when thinking about this analogue, however out of the scope of this paper at this time.

Chapter 3

Conclusion

Studying the behavior of the number of solutions to an elliptic curve $E(\mathbb{F}_p)$ are a fascinating way to explore near-impossible number theory problems. In this thesis, we studied the elliptic analogue of Landau's Near-Square prime conjecture and specifically, the frequency of a square N_p for a selection of 2500 elliptic curves, for primes up to $1,000,000$. We chose to closer examine the behavior of elliptic curves with an excess of square N_p values to understand what may cause this excess. By doing this, we observed two conditions—irreducibility and a small ℓ-adic image—that influence the number of square N_p values an elliptic curve may have.

The next immediate question that arises (as introduced in 2.7) is to find a probabilistic constant that describes the behavior of square numbers of solutions for each elliptic curve. It appears that these distributions follow that of some Hardy-Littlewood constant and are dependent on the ℓ-adic images of the Galois representations.

Bibliography

[1] Laurent Clozel, Michael Harris, and Richard Taylor. Automorphy for some l-adic lifts of automorphic mod l Galois representations. *Publications mathématiques de l'IHÉS*, 108(1):1–181, November 2008.

[2] Henri. Cohen. *A Course in Computational Algebraic Number Theory*. Graduate Texts in Mathematics, 138. Springer Berlin Heidelberg, Berlin, Heidelberg, 1st ed. 1993. edition, 1993.

[3] Michael Harris, Nick Shepherd-Barron, and Richard Taylor. A family of Calabi-Yau varieties and potential automorphy. *Annals of Mathematics*, 171(2):779–813, March 2010.

[4] Helmut Hasse. Zur theorie der abstrakten elliptischen funktionenkörper iii. die struktur des meromorphismenrings. die riemannsche vermutung. *Journal für die reine und angewandte Mathematik*, 1936(175):193–208, 1936.

[5] Serge Lang and Hale Freeman Trotter. *Frobenius Distributions in GL2-Extensions: Distribution of Frobenius Automorphisms in GL2-Extensions of the Rational Num-*

bers, volume 504 of *Lecture Notes in Mathematics*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006.

[6] The LMFDB Collaboration. The L-functions and modular forms database. `https://www.lmfdb.org`, 2024. [Online; accessed 21 May 2024].

[7] M. Ram Murty and V. Kumar Murty. The Sato–Tate conjecture and generalizations. pages 639–646.

[8] János Pintz. Landau's problems on primes. *Journal de Theorie des Nombres de Bordeaux*, 21(2):357–404, 2009.

[9] Jean-Pierre Serre. Properiétés galoisiennes des points d'ordre fini des courbes elliptiques. *Inventiones mathematicae*, 15(4):259–331, 1971.

[10] Joseph H. Silverman. *The Arithmetic of Elliptic Curves*. Graduate Texts in Mathematics, 106. Springer New York, New York, NY, 2nd ed. 2009. edition, 2009.

[11] Joseph H. Silverman and John Tate. *Rational Points on Elliptic Curves*. Undergraduate Texts in Mathematics. Springer-Verlag, New York, 1992.

[12] Andrew V. Sutherland. Computing images of galois representations attached to elliptic curves. *Forum of Mathematics, Sigma*, 4, 2016.

[13] Richard Taylor. Automorphy for some l-adic lifts of automorphic mod l galois representations. ii. *Publications mathématiques de l'IHÉS*, 108(1):183–239, Nov 2008.

[14] The Sage Developers. *SageMath, the Sage Mathematics Software System (Version 10.0)*, 2023. `https://www.sagemath.org`.

[15] Lawrence C. Washington. *Elliptic Curves: Number Theory and Cryptography*. Discrete Mathematics and its Applications. Chapman & Hall/CRC, Boca Raton, 2003.

[16] Marek Wolf. Some conjectures on primes of the form $m^2 + 1$. *Journal of Combinatorics and Number Theory*, 5(2):103–, 2013.