Connections Between the Riemann Hypothesis and the Sato-Tate Conjecture

Christopher J. Swierczewski
Advisor: William Stein

6 June 2008

Acknowledgments

I'm greatly thankful to William Stein, of course, not only for his patience and support while completing my thesis work, but especially for opening up so many opportunities for me to contribute to current research related to the topic of this paper. Thank you for including me in your correspondences with the "big players" in number theory.

Thank you to Jim Morrow for his assistance in understanding the convergence and analyticity of Dirichlet series. To Barry Mazur for asking difficult questions and acknowledging my computations. To Robert Miller for the opportunity to test my knowledge of the subject over lunch and a cup of coffee and for explaining the partial proof of Sato-Tate by Richard Taylor. Finally, to Tom Boothby for academic and emotional support and for keeping me on track. A special thanks to Megan Karalus for patiently listening to my ramblings, checking for spelling errors, and improving the flow of this thesis.

All calculations and were performed using Sage: Open Source Mathematics Software (http://www.sagemath.org) on a server funded by the National Science Foundation under Grant No. 0555776.

Contents

1 Introduction **4**
- 1.1 Preliminaries . . . 5
  - 1.1.1 Isogenies and Complex Multiplication . . . 6

2 The Sato-Tate Conjecture **8**
  - 2.0.2 Hasse's Bound on $a_E(p)$ . . . 8
  - 2.0.3 A Certain Distribution . . . 9
- 2.1 Elliptic Curves and l-adic Representations . . . 13
  - 2.1.1 Galois Representations: An Algebraic Number Theoretic Approach . . . 14
  - 2.1.2 A Formal Definition of $a_E(p)$ . . . 15
- 2.2 A Statement of Equidistribution . . . 16
  - 2.2.1 Application to the Sato-Tate Conjecture . . . 19

3 The Extended Sato-Tate Conjecture and the Main Theorem **20**
- 3.1 Discrepancy of Sequences . . . 20
- 3.2 The Extended Sato-Tate Conjecture . . . 22
- 3.3 The Generalized Riemann Hypothesis . . . 24
- 3.4 The Main Theorem . . . 27
  - 3.4.1 Dirichlet Series . . . 28
  - 3.4.2 The Main Theorem . . . 30

4 Conclusion **32**
- 4.1 A Closer Look at the Series $\sum a_p/\sqrt{p}$ . . . 32

5 Appendix **35**

1 Introduction

The Generalized Riemann Hypothesis has been one of the holy grails of the number theory community and, for that matter, the mathematical community as a whole for approximately 120 years. Although the generalized version is not exactly the statement with the $1,000,000 bounty, provided by the Clay Mathematics Institute, it is nonetheless an important statement about the behavior of a class of complex-valued functions. In particular, from every elliptic curve we can construct a corresponing "elliptic curve L-function." These elliptic curves L-functions form a very important subset of the set of all L-functions.

Consider the plot in Figure 1.1. The blue line is a plot of the partial sums of a certain sequence associated with an elliptic curve. Would it be difficult to believe that if the blue line coninues to stay bounded above in the limit by the red line, then the Generalized Riemann Hypothesis for this elliptic curve holds true?

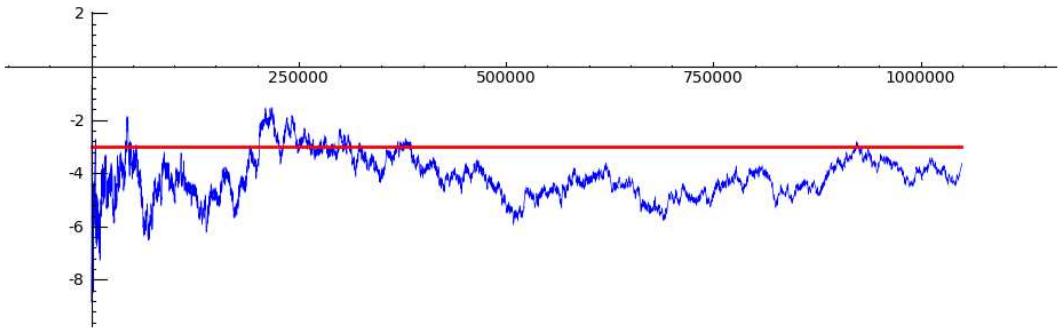

Figure 1.1: A plot suggesting that GRH is true for a particular elliptic curve.

The fact that such a simple picture is evidence for the truth of this elusive hypothesis is deeply connected to some recent developments on the Sato-Tate conjecture—a statement about the convergence of values associated with an elliptic curve. In particular, it was a theorem of Akiyama and Tanigawa in 1999 that an extended version of the conjecture implies GRH for elliptic curve L-functions. [1]

The components to this argument draw from several subjects, including algebra, number theory, analysis, topology, and even statistics. In this thesis, our primary goals are to

(i) present equivalent statements to the Sato-Tate Conjecture in increasing depth and precision using various tools in analysis,

(ii) state the "extended version" of the Sato-Tate Conjecture and present computational support for its truth,

(iii) and develop the necessary background for providing the details to Akiyama and Tanigawa's proof of the main theorem

In this section, we present some preliminary knowledge about elliptic curve— the fundamental construction in this thesis. In Section 2, we state several forms of the Sato-Tate Conjecture, beginning with simple, visual statements of convergence and working our way up to more technical,

advanced statements. In Section 3, we state the Extended Sato-Tate Conjecture and follow the same "simple to advanced" development in the previous section. After introducing related topics, we state and prove the main theorem of this thesis: that the Extended Sato-Tate Conjecture implies the Generalized Riemann Hypothesis.

1.1 Preliminaries

In this section we will give a brief introduction to the theory of elliptic curves. For a more thorough introduction, see [12]. An advanced approach can be found in [11] and [3]. Our first definition concerns the source of what's to come in the rest of this thesis.

Definition 1.1. *An* **elliptic curve** *E over a field K, denoted E/K, is a projective variety defined by the* **Weierstrass equation***:*

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$$

where $a_1, a_2, a_3, a_4, a_6 \in K$. We denote the set of points/solutions over K by

$$E(K) := \{(x, y) \in K^2 : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6\} \cup \{\infty\}$$

plus an additional "point at infinity".

When $\text{char}(K) \neq 2, 3$, we can simplify the equation of an elliptic curve to the form

$$y^2 = x^3 + ax + b \tag{1.1}$$

for $a, b \in K$. Even though many of the statements concerning the Sato-Tate Conjecture can be generalized to elliptic curves over arbitrary number fields, throughout this paper we will restrict our attention to elliptic curves over the rationals $K = \mathbb{Q}$ for simplicity, but mostly for familiarity.

The set of K-rational points, plus an additional "point at infinity" $\mathcal{O}$ acting as the identity element, form an Abelian group structure under the secant intersection operation defined as follows: let $P = (x_0, y_0)$ and $Q = (x_1, y_1)$ be rational points on an elliptic curve E. Since E is a cubic, a line intersecting these two points will have a third point of intersection. (If we pick a line parallel to the y-axis, the third point is considered to be the aforementioned point at infinity. An involved computation shows that this point is also rational. We define the point $P+Q$ by a certain transformation of the third point of intersection. In the case when $a_1 = a_3 = 0$, this transformation is a simple reflection over the x-axis. (This transformation is done to preserve associativity. With the addition of the point at infinity, this is a well-defined group operation.

We denote the group of K-rational solutions to E by $E(K)$. Depending on the curve, this group can be finite or infinite. It is a result of Mordell, in the case of $K = \mathbb{Q}$, and later Weil that for any number field K, $E(K)$ is finitely generated. This gives the famous structure theorem for elliptic curves.

Theorem 1.2. *Let K be an algebraic number field. Then the group of K-rational points $E(K)$ is a finitely generated Abelian group. In particular.*

$$E(\mathbb{Q}) \cong \mathbb{Z}^r \oplus E(\mathbb{Q})_{tor}.$$

The value r is referred to as the **rank** *of the elliptic curve E.*

In 1987, Mazur classified all possible structures for $E(\mathbb{Q})_{\text{tor}}$ in a very deep and difficult paper found in [5].

Theorem 1.3. *The following finite groups consist of all possible structures for the group of torsion points $E(\mathbb{Q})_{tor}$:*

$$\mathbb{Z}/n\mathbb{Z}, \quad \text{where } 1 \leq n \leq 10 \text{ or } n = 12$$

or

$$\mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2n\mathbb{Z}, \quad \text{where } 1 \leq n \leq 4.$$

In the general case, Merel proved the following theorem.

Theorem 1.4. *Let d be a positive integer. Then, over all algebraic number fields K with $[K : \mathbb{Q}] \leq d$, there are only finitely many possibilities for $E(K)_{tor}$.*

1.1.1 Isogenies and Complex Multiplication

We now turn to mappings between elliptic curves.

Definition 1.5. (Isogeny) *An* **isogeny** *is a morphism of curves that preserves the basepoint $\mathcal{O}$. In this case, a morphism preserving the identity $\mathcal{O}$ on the group of rational points $E(\mathbb{Q})$.*

One can define the multiplication by m map $m : E(\mathbb{Q}) \to E(\mathbb{Q})$ by

$$P \mapsto mP = P + \cdots + P, \quad \text{for all } P \in E(\mathbb{Q}) \text{ and } m \geq 0\ . \tag{1.2}$$

As a morphism of curves, the map fixes the basepoint $\mathcal{O}$ and is therefore an isogeny. It is also surjective. (See [11] or [3].) Furthermore, we can consider the kernel, denoted $E[m]$, which consists of the points in $E(\bar{\mathbb{Q}})$ of order m.

Proposition 1.6. *Over $\mathbb{Q}$, and in general, a number field K,*

$$E[m] \cong \mathbb{Z}/m\mathbb{Z} \oplus \mathbb{Z}/m\mathbb{Z}$$

We need not restrict out attention to finite extensions of $\mathbb{Q}$. In fact, much information can be derived about elliptic curves if we look at the set of solutions over $\mathbb{C}$. The complex points $E(\mathbb{C})$ also form an Abelian group. Now, for *any* Abelian group, one can always consider the "multiplication by n homomorphism"

$$E(\mathbb{C}) \to E(\mathbb{C}), \quad P \mapsto nP$$

This map is an endomorphism of E. As an Abelian variety, we can examine the endomorphism ring $\text{End}_{\mathbb{C}}(E)$ of the elliptic curve. For some elliptic curves, the multiplication by n maps for all $n > 0$ consist of all possible endomorphisms. For others, however, there is a different story.

Definition 1.7. *An elliptic curve E is said to have* **complex multiplication** *if there exists an endomorphism of E that is not a multiplication by n map. That is, if the containment $\text{End}_{\mathbb{C}}(E) \supset \mathbb{Z}$ is strict.*

A different kind of map that we can consider, which is especially important in the following section, is the "reduction modulo p" map. Let $p \in \mathbb{Z}$ be a prime. In the case when $\mathrm{char}(K) \neq 2, 3$, there is a natural reduction map

$$E(\mathbb{Q}) \to \tilde{E}(\mathbb{F}_p) \tag{1.3}$$

$$y^2 = x^3 + ax + b \mapsto y^2 = x^3 + \bar{a}x + \bar{b}, \quad \bar{a}, \bar{b} \in \mathbb{F}_p\,. \tag{1.4}$$

In the $\mathrm{char}(K) = 2, 3$ case, one must use the full Weierstrass equation given in Definition 1.1. One must consider issues of singularity, of course, for the reduction of a non-singular curve over $\mathbb{Q}$ may result in a singular curve over $\mathbb{F}_p$. One can computationally show that if $p \mid \Delta$, the discriminant of E, then this singularity over $\mathbb{F}_p$ occurs. When E is a minimal Weierstrass equation, (see [12], [3], or [11]) we can make the following definition.

Definition 1.8. *Let $E/\mathbb{Q}$ be an elliptic curve and $\tilde{E}/\mathbb{F}_p$ be its reduction modulo p. If $\tilde{E}/\mathbb{F}_p$ is nonsingular, then E has* **good reduction modulo** *p. Otherwise, we say that it has* **bad reduction**.

Since there are only finitely many primes dividing Δ in this case, there can only be finitely many primes at which E has bad reduction.

2 The Sato-Tate Conjecture

The Sato-Tate Conjecture is simply a statement about the distribution of a sequences of integers associated with an elliptic curves. Not too much is known about the history of its formulation. However, we do know that it was originally authored by Mikio Sato and John Tate. Although published in 1963, they independently formulated the conjecture during the 1960's. Tate received word in 1963 that Sato had been performing computations on elliptic curves and stated a conjecture that quickly become known as the Sato-Tate Conjecture.

There is, in fact, a proof of the conjecture for a large class of elliptic curves due to Richard Taylor. [13] In 2006, he showed that the conjecture holds as long as E is an elliptic curve over a totally real field with multiplicative reduction at some prime. Equivalently, when $K/\mathbb{Q}$ is a finite extension generated by one root of a polynomial $f(x) \in \mathbb{Z}[x]$ where all the roots of $f(x)$ are real. Since then, he has extended his proof to include an even larger class of elliptic curves. However, for elliptic curves over any Galois extension of $\mathbb{Q}$, it still remains an open question.

In this section, we will explore the breath and depth of the Sato-Tate Conjecture. We will begin an elementary definition of a sequence of integers associated with each elliptic curve and prove Hasse's bound on that sequence. With this definition, we will then make a crucial observation about the distribution of these integers. This observation gives us an elementary statement of the Sato-Tate Conjecture. We will then provide equivalent, more refined statements of the conjecture using the language of the equidistribution of sequences.

2.0.2 Hasse's Bound on $a_E(p)$

Let E be an elliptic curve without complex multiplication. The primary ingredient in the conjecture of Sato and Tate are the integers

$$a_E(p) := |p + 1 - \#E(\mathbb{F}_p)|. \tag{2.1}$$

defined for each rational prime p. When there is no confusion on the choice of elliptic curve, we will simply write $a_p = a_E(p)$. These a_p terms can be thought of as an error term associated with the order of the group of points of order p on E. That is, it measures how far off the group of points on an elliptic curve modulo p is from an "ideal" group of solutions.

One hopes that error terms have a bound and indeed Figure 2.1 suggests that they do. In 1934, Hasse proved a bound on each a_p in terms of p. When dealing with points on an elliptic curve modulo p, in particular, it is natural to consider the Frobenius endomorphism, $\psi_p : (x, y) \mapsto (x^p, y^p)$. This fundamental map satisfies the properties:

Lemma 2.1. *The degree map* $\deg : \mathrm{End}(E) \to \mathbb{Z}$ *satisfies the relationship*

$$|\deg(\phi - \psi) - \deg(\psi) - \deg(\phi)| \leq 2\sqrt{\deg(\psi)\deg(\phi)}$$

for all $\phi, \psi \in \mathrm{End}(E)$.

Proof. The proof of a generalization of this lemma for positive definite quadratic forms that map arbitrary Abelian groups A to $\mathbb{Z}$ is given in [11]. □

Theorem 2.2. (Hasse's Bound) *Let E be an elliptic curve. Then,*

$$|a_E(p)| \leq 2\sqrt{p}$$

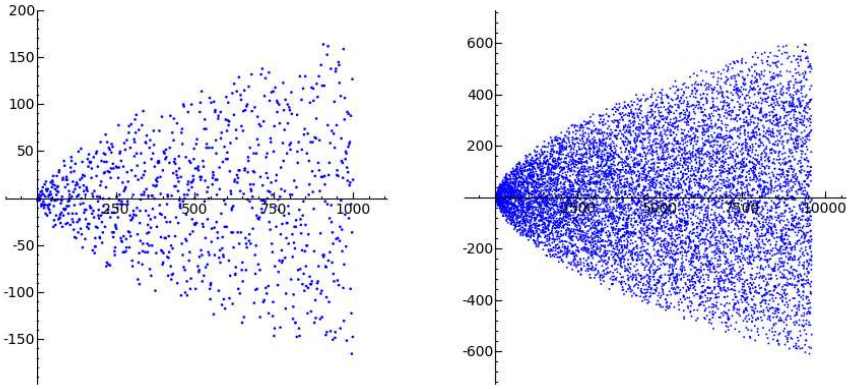

Figure 2.1: Plots of $a_E(p)$ for $p <= 10^3, 10^5$ where $E : y^2 + y = x^3 - x^2 - 10x - 20$.

Proof. Consider the Frobenius map $\psi : E \to E$ defined

$$(x, y) \mapsto (x^p, y^p).$$

Since the Galois group $G_{\bar{\mathbb{F}}_p / \mathbb{F}_p}$ is (topologically) generated by the p-th power map on $\bar{\mathbb{F}}_p$, we see that for a point $P \in E(\bar{\mathbb{F}}_p)$,

$$P \in E(\mathbb{F}_p) \quad \iff \quad \psi(P) = P.$$

That is, $\psi(P) - P = 0$ for all points $P \in E(\mathbb{F}_p)$. Since, ψ is a homomorphism of E, we can write this as

$$E(K) = \ker(1 - \psi).$$

Therefore, $\#E(\mathbb{F}_p) = \#\ker(1 - \psi)$. It's a standard result from Silverman [11] that $1 - \psi$ is a separable elliptic curve endomorphism, when $E/\mathbb{F}_p$ is treated as an algebraic extension, and hence $\#E(\mathbb{F}_p) = \deg(1 - \psi)$. Finally, applying Lemma 2.1 gives us the desired result. □

We see now why a_p is considered to be an error term on the number of solutions to an elliptic curve reduced modulo p. It measures the degree to which the endomorphism $1 - \psi$ on $E(\mathbb{F}_p)$ fails to be an isomorphism. Equivalently, by looking at the kernel of this map, it measures how many points on $E(\mathbb{F}_p)$ are fixed by the Frobenius endomorphism.

2.0.3 A Certain Distribution

Given the sequence (a_p) along with the bound $|a_p| \leq 2\sqrt{p}$, we can create the sequence $(a_p/(2\sqrt{p})) \subset [-1, 1]$. We can then ask, *how is the sequence distributed within the interval* $[-1, 1]$*?* Figure 2.2 displays the distribution of the sequences $(a_p/2\sqrt{p})_{p<X} = (a_E(p)/(2\sqrt{p}))_{p<X}$ for various elliptic curves E and bounds X on the number of terms in the sequence. The frequency histograms suggest that these sequences assume a "squashed" semi-circle distribution—a peculiarity since most naturally-occurring data sets tend to conform to either a normal or uniform distribution.

Simply put, this is indeed the Sato-Tate conjecture!

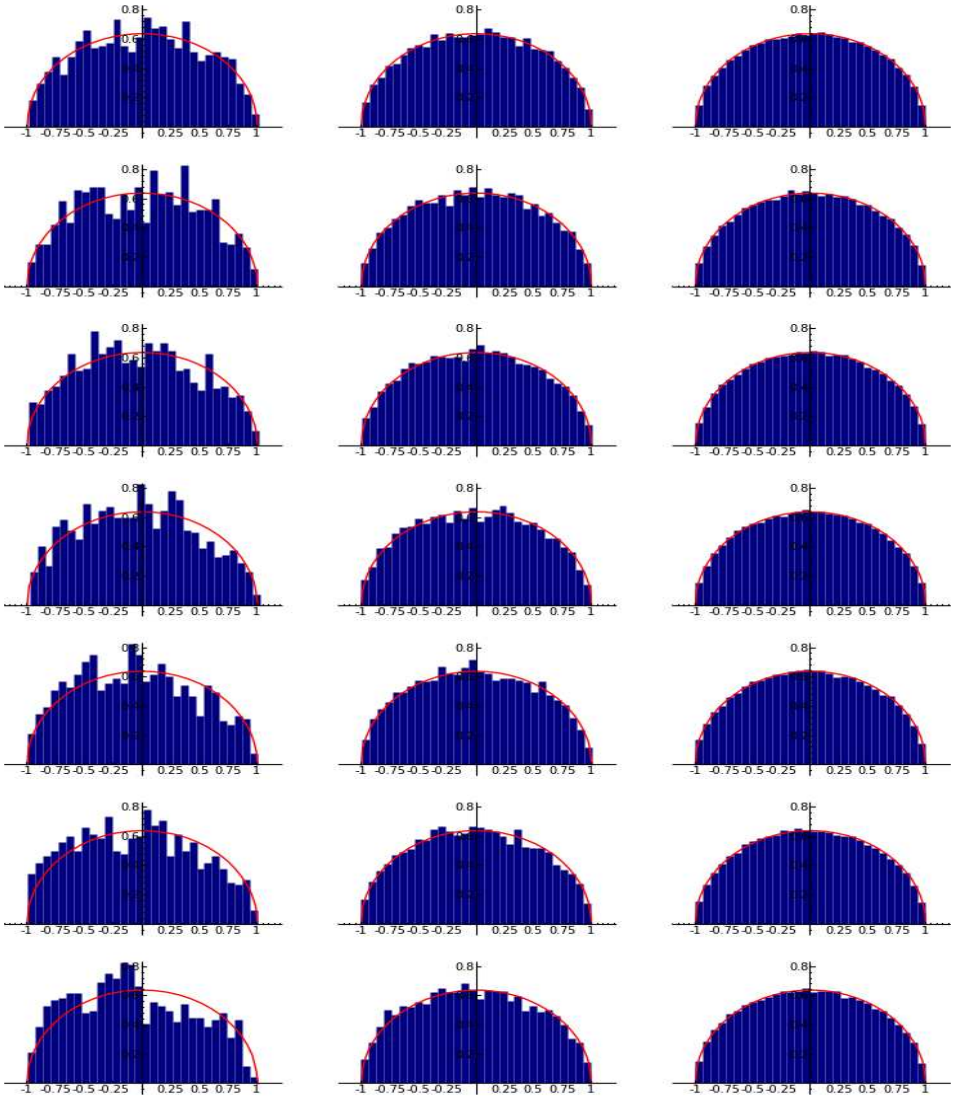

Figure 2.2: Each row features a series of frequency histograms of the sequence $(a_E(p_n)/(2\sqrt{p_n}))_{n \leq N}$ for elliptic curves E of ranks 0 through 6, respectively. Within each row, each plot corresponds to the distribution of the first $N = 10^3, 10^4$, and 10^5 elements of the sequence. Note how the data skews slightly to the left on elliptic curves of high rank, suggesting that for small values of p, there are fewer than ideal number of solutions on $E(\mathbb{F}_p)$.

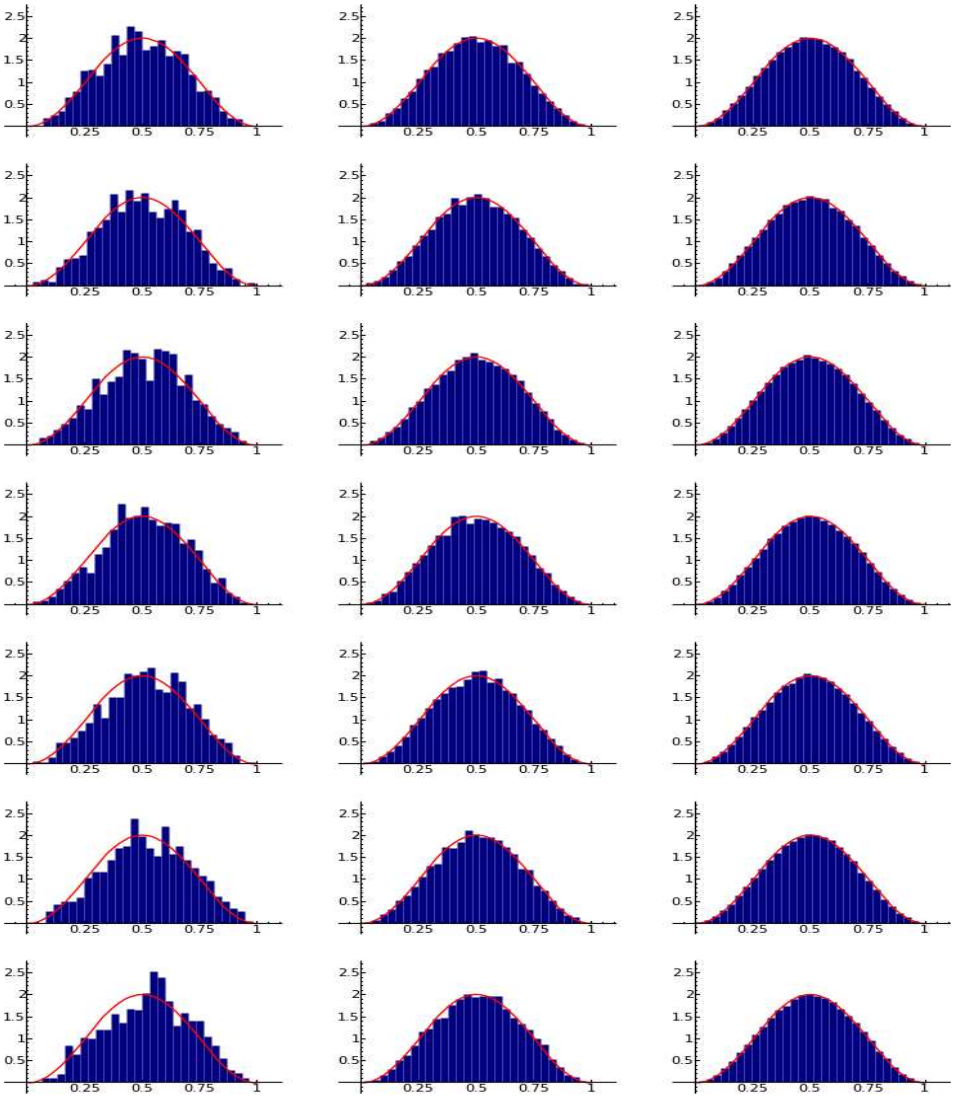

Figure 2.3: The corresponding frequency histograms from 2.2 with the normalized sequences $(x_n)_{n=1}^{N}$ for $N = 10^3, 10^4$, and 10^5. Note that the data for elliptic curves of higher rank skews slightly to the right under this transformation.

Conjecture 2.3. (Sato-Tate Conjecture) *Let E be an elliptic curve without complex multiplication. Then the frequency distribution of the sequence $(a_E(p)/(2\sqrt{p}))$ on $[-1,1]$ converges to the scaled semi-circle distribution $(2/\pi)\sqrt{1-x^2}$. That is, for all subintervals $[a,b] \subset [-1,1]$,*

$$\lim_{X\to\infty} \frac{\#\{\frac{a_p}{2\sqrt{p}} \in [a,b] : p < X\}}{\pi(X)} \to \frac{2}{\pi}\int_a^b \sqrt{1-x^2}dx$$

We will now create a bounded sequence using these a_p so we can begin to analyze the distribution of the values. Define θ_p by

$$a_p = 2\sqrt{p}\cos(\theta_p) \quad \Rightarrow \quad \theta_p = \arccos\left(\frac{a_p}{2\sqrt{p}}\right) \tag{2.2}$$

By this definition, one should obseve that θ_p is equal to an angle of a root of the polynomial equation $x^2 - a_p x + p$. Now, by Hasse's bound, $\theta_p \in (0,\pi)$. If we let p_n denote the n-th prime number and normalize the sequence of θ_{p_n}, we have the real sequence $x_n \in (0,1)$ given by

$$x_n = \frac{\theta_{p_n}}{\pi}, \qquad \theta_{p_n} = \arccos\left(\frac{a_{p_n}}{2\sqrt{p_n}}\right) \tag{2.3}$$

Since the sequence (x_n) is bounded, we can look at the distribution of the values within the interval $[0,1]$—just as we did with the sequence $(a_p/2\sqrt{p})$. Indeed, a similar picture appears within frequency histogram. We see that the distribution of the data converges to the function $2\sin^2(\pi x)$ over the interval $[0,1]$. Conjecture 2.4 is a mathematical embodiment of this observation.

Conjecture 2.4. (Sato-Tate Conjecture with x_n) *Let E be an elliptic curve without complex multiplication. Then*

$$\lim_{N\to\infty} \frac{\#\{x_n \in [\alpha,\beta) : n \le N\}}{N} \to 2\int_\alpha^\beta \sin^2(\pi x)dx$$

Furthermore, this conjecture is equivalent to the statement of Conjecture 2.3.

Equivalence to Conjecture 2.3. The equivalence is a result of a simple transformation. By the relations in in Equation 2.3,

$$\frac{a_{p_n}}{2\sqrt{p_n}} \mapsto \theta_{p_n} \mapsto x_n$$
$$\frac{2}{\pi}\sqrt{1-x^2} \mapsto \frac{2}{\pi}\sin^2(x) \mapsto 2\sin^2(\pi x).$$

By this transformation of distributions, this proves that Conjectures 2.3 and 2.4 are equivalent. □

During the rest of this section, we will develop the mathematics behind this statement of distribution. In Section 2.1 we will provide more detail on the nature of the a_p's—where they come from and their deeper involvement in the theory of elliptic curves. In Section 2.2 we will develop some tools that can describe the convergence stated by the Sato-Tate Conjecture with much more precision.

2.1 Elliptic Curves and l-adic Representations

In Section 2 we gave a straightforward definition of of the integers $a_E(p)$,

$$a_E(p) := |p + 1 - \#E(\mathbb{F}_p)|,$$

the error terms of the number of solutions to an elliptic curve E modulo p for some prime. Although one should appreciate simple definitions, it's necessary to understand as much as you can about these elusive error terms when attempting a proof for the Sato-Tate Conjecture. In this section, we briefly explore how these $a_E(p)$ naturally arise from looking at endomorphisms of $E[m]$ for $m > 0$. The background presented in this section is not necessary for the proof of the main theorem. Nonetheless, investigating these error terms improves our understanding and reveals their deeper algebraic nature.

We begin with a natural construction of elliptic curve group representations. Then, using some algebraic number theory, we define the Frobenius elements from which these $a_E(p)$ are derived and conclude with the primary relationship between the relevant endomorphisms of $E[m]$ and $a_E(p)$. See [9] and [10] for a thorough treatment of these topics.

Consider the group of points $E[m]$ of order m defined earlier in Section 1.1. Since it consists of solutions to the elliptic curve equation, it lies in $E(\bar{\mathbb{Q}})$, where $\bar{\mathbb{Q}} \subset \mathbb{C}$ consists of all algebraic numbers. With this perspective in mind, the group $\mathrm{Gal}(\bar{\mathbb{Q}}/\,\mathbb{Q})$ acts on $E[m]$ by

$$P = (x, y) \mapsto (\sigma(x), \sigma(y)), \quad \sigma \in \mathrm{Gal}(\bar{\mathbb{Q}}/\,\mathbb{Q})$$

and, in fact, $E[m]$ is stable under this action. Therefore, the action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\,\mathbb{Q})$ on $E[m]$ is described by the group homomorphism

$$\bar{\rho}_{E,m} : \mathrm{Gal}(\bar{\mathbb{Q}}/\,\mathbb{Q}) \to \mathrm{Aut}(E[m]) \tag{2.4}$$

which is in fact a group representation since

$$\mathrm{Aut}(E[m]) \cong \mathrm{GL}_2(\mathbb{Z}/m\,\mathbb{Z}),$$

the set of all invertible 2×2 matrices with entries in $\mathbb{Z}/m\,\mathbb{Z}$. Simply put, we can think of each automorphism of the points of order m on E as having both a well-defined determinant and trace in $\mathbb{Z}/m\,\mathbb{Z}$.

Using Galois theory, we see that the kernel of this map corresponds to a finite Galois extension K_m of $\mathbb{Q}$ in $\bar{\mathbb{Q}}$. We can explicitly construct this extension by adjoining the points in $E[m]$. We therefore have group homomorphism

$$\rho_{E,m} : \mathrm{Gal}(K_m/\,\mathbb{Q}) \to \mathrm{Aut}(E[m]) \cong \mathrm{GL}_2(\mathbb{Z}/m\,\mathbb{Z}). \tag{2.5}$$

That is,

Proposition 2.5. *For an elliptic curve E and a positive integer m, there exists a finite Galois extension $K_m/\,\mathbb{Q}$ whose Galois group G_m is a subgroup of the group $\mathrm{GL}_2(\mathbb{Z}/m\,\mathbb{Z}) \cong \mathrm{Aut}(E[m])$.*

We can ask, "what does the image of $G_m = \mathrm{Gal}(K_m/\,\mathbb{Q})$ look like under $\rho_{E,m}$"? There are some issues in the complex multiplication case which imply that the image of G_m is much smaller than $\mathrm{GL}_2(\mathbb{Z}/m\,\mathbb{Z})$. Details aside we focus our attention only on elliptic curves E without complex multiplication in which case we have the following result of Serre.

Proposition 2.6. *For all but finitely many primes p,*

$$G_p \cong \mathrm{GL}_2(\mathbb{Z}/p\mathbb{Z}).$$

The key property of the extension $K_m/\mathbb{Q}$ is that the discriminant of the extension, meaning the discriminant of the polynomial whose roots generate this extension, is divisible only by the prime number that divide either m or the discriminant Δ of E. Whenever this occurs, we can use algebraic number theory to construct a class of Frobenius automorphisms in G_m from which we can derive a relationship between $\rho_{E,m}$ and $a_E(p)$.

2.1.1 Galois Representations: An Algebraic Number Theoretic Approach

In this section we will outline the structure and some properties of G_m that give rise to the error terms $a_E(p)$. As in the previous section, we refer primarily to [9]. For simplicity, we consider an arbitrary finite Galois extension $K/\mathbb{Q}$ and the its of integers $\mathcal{O}_K$. It can easily be shown that $\mathcal{O}_K$ is invariant under the canonical action of $\mathrm{Gal}(K/\mathbb{Q})$ and consequently that the set of prime ideals of $\mathcal{O}_K$ are permuted under the action. With this in mind, we make the following definition.

Definition 2.7. *Let K be a number field and $\mathcal{O}_K$ its ring of integers. The* **decomposition group** *of a prime ideal $\mathfrak{p} \subset \mathcal{O}_K$ is the set*

$$D_{\mathfrak{p}} = \{\sigma \in \mathrm{Gal}(K/\mathbb{Q}) : \sigma(\mathfrak{p}) = \mathfrak{p}\}.$$

Now take a moment to consider the finite field $\mathbb{F}_{\mathfrak{p}} := \mathcal{O}_K/\mathfrak{p}$. $\mathbb{F}_{\mathfrak{p}}$ is a finite Galois extension of $\mathbb{F}_p$ whose Galois group is generated by the Frobenius automorphism

$$\phi_p : x \mapsto x^p$$

of $\mathbb{F}_{\mathfrak{p}}$. It turns out that if one associates a given $\delta \in D_{\mathfrak{p}}$ to the automorphism of $\mathbb{F}_{\mathfrak{p}}$ induced by δ, then one derives a natural "reduction map" $D_{\mathfrak{p}} \to \mathrm{Gal}(\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p)$.

Returning to the extension $K_m/\mathbb{Q}$, a key observation at the end of the previous section implies that this map is a bijection, similar to that in Proposition 2.6.

Proposition 2.8. *If $\mathfrak{p} \subset \mathcal{O}_{K_m}$ and $p \nmid \Delta, m$ then $D_{\mathfrak{p}} \cong \mathrm{Gal}(\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p)$.*

Therefore, whenever p has this "good reduction", there is a unique automorphism $\sigma_{\mathfrak{p}} \in D_{\mathfrak{p}}$ of $\mathrm{Gal}(K/\mathbb{Q})$ whose image in $\mathrm{Gal}(\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p)$ is the Frobenius automorphism ϕ_p. The map σ_p is referred to as the **Frobenius automorphism for $\mathfrak{p}$**. Since the prime ideals of $\mathcal{O}_{K_m}$ are all conjugate, these Frobenius automorphisms are subsequently conjugate in $\mathrm{Gal}(K_m/\mathbb{Q})$. Hence, we can consider the conjugacy class itself as an automorphism in $\mathrm{Gal}(K_m/\mathbb{Q})$ in terms of p.

Definition 2.9. *The* **Frobenius automorphism** $\sigma_p \in \mathrm{Gal}(K_m/\mathbb{Q})$*, which is only well-defined up to conjugation of the lifts of p in $\mathcal{O}_{K_m}$, is the automorphism whose image under the map $D_{\mathfrak{p}} \to \mathrm{Gal}(\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p)$ is the Frobenius automorphism ϕ_p.*

In other words, we can think of σ_p as a representative for any and all of the Frobenius automorphisms $\sigma_{\mathfrak{p}}$ where $\mathfrak{p}$ is any lift of p in $\mathcal{O}_{K_m}$. In the next section, extend all of the notions discussed thus far into the case $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$.

2.1.2 A Formal Definition of $a_E(p)$

With some care and precision, one can perform the same algebraic analysis over the group $G = \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$. The benefit is that we will be able to construct an element much like σ_p where the trace is exactly a_p—as opposed to the reduction $a_p \bmod m$. See [3] for more information on the constructions that follow. We begin with looking at representations G, just as in finite extension case. Fix a prime l. Given the structure $E[l^n] \cong \mathbb{Z}/l^n\mathbb{Z} \oplus \mathbb{Z}/l^n\mathbb{Z}$, we can extend the notion of the induced Frobenius automorphism on $E[l^n]$ to a projective limit.

Definition 2.10. (Tate Module) *The* l**-adic Tate module of** E *is the group*

$$\mathrm{Tate}_{E,l} = \varprojlim E[l^n] \cong \mathbb{Z}_l \oplus \mathbb{Z}_l$$

where the limit is taken over n *and* $\mathbb{Z}_l$ *are the* l*-adic numbers.*

As seen earlier, for $m = l^n$ the group $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ acts on $E[l^n]$. Furthermore, the action can be shown to commute with the multiplication-by-l maps used to form the projective limit defining the Tate module of E. Therefore, G also acts on $\mathrm{Tate}_{E,l}$. Under the right topology (when we define a notion of continuity) we can finally define a group representation analogue to the one given in Definition 2.4.

Definition 2.11. (l-adic Representation) *The* l**-adic representation of** $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ **on** E *is the group homomorphism*

$$\rho_{E,l}\,\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q}) \to \mathrm{Aut}(\mathrm{Tate}_{E,l})$$

giving the action of $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ *of* $\mathrm{Tate}_{E,l}$.

Now, the analogue algebraic approach: let p be a rational prime and $\mathfrak{p}$ a prime of $\bar{\mathbb{Q}}$ lying over p. We can construct a similar residue field $\mathbb{F}_\mathfrak{p}$, a decomposition group $D_\mathfrak{p}$, and a surjective map between $D_\mathfrak{p}$ and the Galois group $\mathrm{Gal}(\mathbb{F}_\mathfrak{p}/\mathbb{F}_p)$. The Frobenius automorphism $\phi_p : x \mapsto x^p$ lies within this group and, similar to the definition of $\sigma_\mathfrak{p}$, we look at the preimage of ϕ_p under the above map. The corresponding conjugacy class leads to the following symmetric definition in the $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ case,

Definition 2.12. *Let* $\mathfrak{p}$ *be any lift of* p *in* $\mathbb{Q}$. *The symbol* Frob_p*, called the* **Frobenius element for** p **in** $\mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$*, denotes any preimage of* $\phi_p \in \mathrm{Gal}(\mathbb{F}_\mathfrak{p}/\mathbb{F}_p)$ *in* $D_\mathfrak{p}$.

Even though the Frobenius element is now even more ill-defined—it is only defined up to conjugation of the lifts, $\mathfrak{p}$, and the kernel of the map $D_\mathfrak{p} \to \mathrm{Gal}(\mathbb{F}_\mathfrak{p}/\mathbb{F}_p)$, called the inertia subgroup $I_\mathfrak{p}$ of $D_\mathfrak{p}$ is very large—the Frobenius element $\mathrm{Frob}_p \in G$ still has a well-defined trace and determinant since conjugate matrices share the same trace and determinant. Interestingly enough, it turns out that these two matrix operations produce the long-sought $a_E(p)$.

Theorem 2.13. *Let* $\mathrm{Frob}_p \in \mathrm{Gal}(\bar{\mathbb{Q}}/\mathbb{Q})$ *be the Frobenius element given in Definition 2.12. Then*

$$\begin{aligned} \det\left(\rho_{E,l}(\mathrm{Frob}_p)\right) &= p \\ \mathrm{tr}\left(\rho_{E,l}(\mathrm{Frob}_p)\right) &= a_p \end{aligned}$$

where l *is determined by the* l*-adic representation* ρ_E *given in Definition 2.11 and where the error term is the usual* $a_p = p + 1 - \#E(\mathbb{F}_p)$.

Note that although entries of the image $\rho_{E,l}(\text{Frob}_p)$ lie in $\mathbb{Z}_l$, the determinant and trace indeed lie in $\mathbb{Z}$. Therefore, this is exactly how the group representation $\rho_{E,l}$ encapsulates information about the a_p for primes of good reduction which do not divide n. Richard Taylor uses these facts, along with many others, about a_p in his partial proof of the Sato-Tate Conjecture which one can read about in Mazur [6].

2.2 A Statement of Equidistribution

Even without examining the source of the $a_E(p)$ in too much detail, one can say that the Sato-Tate Conjecture is a statement about the distribution of the normalized error terms $a_p/2\sqrt{p}$ or, equivalently, the normalized angles $x_n = \theta_{p_n}/\pi$. There is a relationship between the number of x_n that lie in each subinterval of $[0,1]$ and the function $g(x) = 2\sin^2(\pi x)$. The theory of equidistribution, developed by Niederreiter, is a valuable tool in describing such a relationship. See [7] and [8] for more on the topic.

We begin with the basic definition.

Definition 2.14. *The sequence (x_n) is said to have the* **asymptotic distribution function** mod1 $g(x)$ *(abbreviated a.d.f.* mod1*) if*

$$\lim_{N\to\infty} \frac{A([0,x);N;(x_n))}{N} = g(x) \qquad \textit{for all } x \in [0,1]$$

where

$$A([0,x);N;(x_n)) = \#\{x_n - \lfloor x_n \rfloor \in [0,x) : n \leq N\}.$$

(x_n) *is simply said to be* **equidistributed** mod1 *if it has the asymptotic distribution function $g(x) = x$.*

There is a natural interpretation of this notion. In the case when $g(x) = x$, and $(x_n) \subset [0,1]$, the sequence is equidistributed when each subinterval of $[0,1]$ gets its "fair share" of x_n's as $N \to \infty$. In general, each subinterval gets a share of the elements of the sequence weighted by the distribution function g. So we can think of g as a "weight function" on the sequence (x_n).

(Example) One example of an equidistributed sequence mod1 is the sequence

$$(n\sqrt{2})_{n=1}^{N}, \text{ where } a \text{ is irrational.} \tag{2.6}$$

It is the statement of the aptly-named Equidistribution Theorem that this sequence is indeed equidistributed mod1—that is, it has $g(x) = x$ as its asymptotic distribution function. See Figure 2.4 for evidence that this is indeed true. The frequency histogram of the fractional parts of the sequence converges to the uniform distribution.

(Non-Example) The sequence

$$(\log(n))_{n=1}^{N} \tag{2.7}$$

is not equidistributed mod1 as $N \to \infty$. See 2.5 for plots supporting this claim. The frequency histogram does not tend toward a uniform distribution.

When interpreting equidistribution as assignment of weights to elements of a sequence within an interval, one can see immediate applications to the theory of integration. Thus, we can think of asymptotic distribution functions as functions that help measure the mean value of a function over a sequence (x_n). The following theorem describes this relationship.

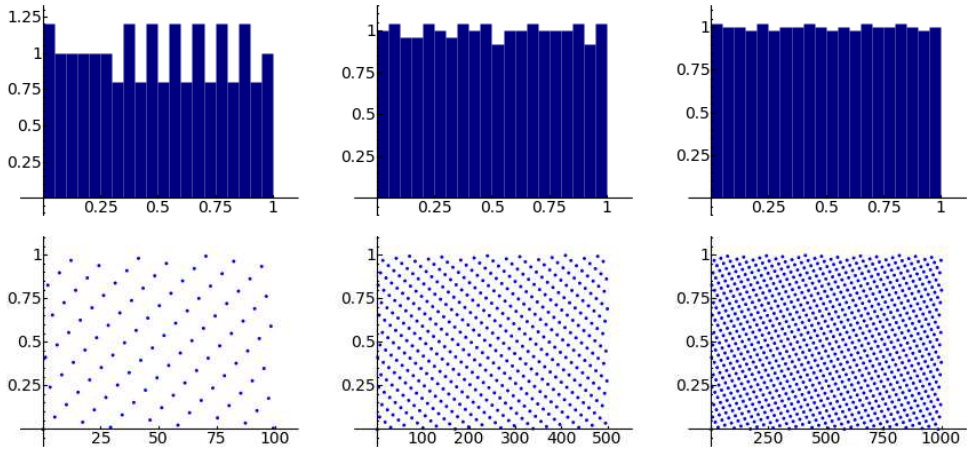

Figure 2.4: An example of an equidistributed sequence mod1: $(n\sqrt{2})_{n=1}^{N}$. The first sequence of plots is a frequency histogram for the first 100, 500, and 1000 elements of the sequence. The second sequence of plots are graphs of the fractional part of the n-th element of the sequence for the first 100, 500, and 1000 elements of the sequence.

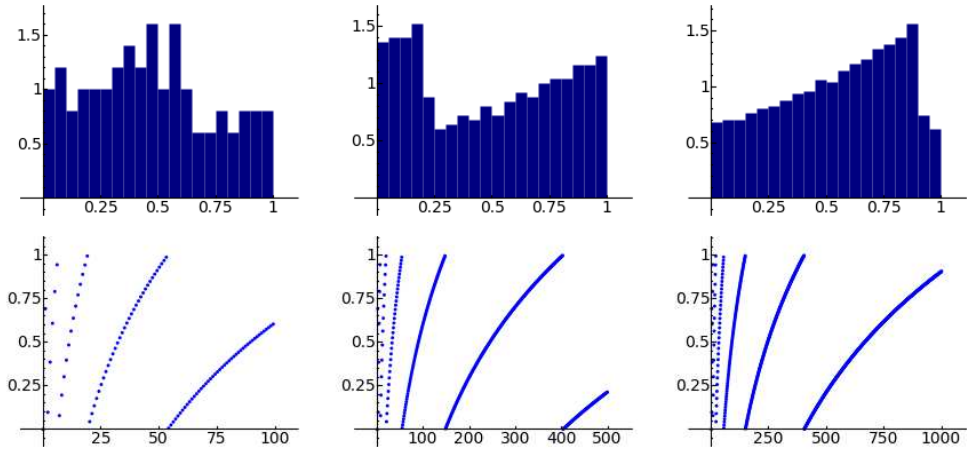

Figure 2.5: A non-example of an equidistributed sequence mod1: $(\log(n))_{n=1}^{N}$. The first sequence of plots is a frequency histogram for the first 100, 500, and 1000 elements of the sequence. The second sequence of plots are graphs of the fractional part of the n-th element of the sequence for the first 100, 500, and 1000 elements of the sequence.

Theorem 2.15. *A sequence (x_n) has the continuous a.d.f.* $\mathrm{mod}1g(x)$ *if and only if for every real-valued continuous function $f : [0.1] \to [0, 1]$,*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(\{x_n\}) = \int_0^1 f(x)dg(x).$$

Proof. (See p.54 [4])
($\Leftarrow$)
By the definition of the Riemann-Stieltjes integral,

$$\int_0^1 f(x)dg(x) = \lim_{P} \sum_{x_i \in P} f(c_i)(g(x_{i+1}) - g(x_i))$$

where the limit is taken over the mesh of the partition P of $[0, 1]$ and $c_i \in [x_i, x_{i+1}]$. So suppose (x_n) has the a.d.f. $\mathrm{mod}1g(x)$. Define the sequence of functions (F_N) on $\mathbb{R}$ by

$$F_N(x) = \begin{cases} \frac{A([0,x);N;x_n)}{N} & \text{for } x \in [0, 1] \\ 0 & \text{for } x < 0 \\ 1 & \text{for } x > 1 \end{cases}$$

Since each F_N is nondecreasing and left continuous on $\mathbb{R}$ with $\lim_{x\to-\infty} F_N(x) = 0$ and $\lim_{x\to\infty} F_N(x) = 1$, they are probability distribution functions. ($\Rightarrow$)
Suppose the equation holds for every real-valued continuous function on $[0, 1]$. Let $[a, b)$ be an arbitrary subinterval of I. For all $\epsilon > 0$, $\exists g_1, g_2 : [0, 1] \to [0, 1]$ continuous such that $g_1(x) \le g(x) \le g_2(x)$ for $x \in [0, 1]$ and that $\int_0^1 (g_2(x) - g_1(x))dx \le \epsilon$. Then, by the definition of the Riemann Stieltjes integral, we have

$$\begin{aligned} g(b) - g(a) - \epsilon &\le \int_0^1 g_2(x)dx - \epsilon \le \int_0^1 g_1(x)dx = \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g_1(\{x_n\}) \\ &\le \liminf_{N\to\infty} \frac{A([a, b); N; x_n)}{N} \le \limsup_{N\to\infty} \frac{A([a, b); N; x_n)}{N} \\ &\le \lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} g_2(\{x_n\}) = \int_0^1 g_2(x)dx \le \int_0^1 g_1(x)dx + \epsilon \\ &\le g(b) - g(a) + \epsilon \end{aligned}$$

Since $\epsilon > 0$ is arbitrarily small, this implies

$$g(b) - g(a) = \lim_{N\to\infty} \frac{A([a, b); N; x_n)}{N}$$

for all subintervals $[a, b) \subset [0, 1]$, thus finishing the proof. □

More information on how equidistributed sequences are related to integral approximations can be found in [7] and [8].

2.2.1 Application to the Sato-Tate Conjecture

Recall the defintion of the Sato-Tate sequence:

$$x_n = \theta_{p_n}/\pi, \quad a_{p_n} = 2\sqrt{p}\cos(\theta_{p_n}) \tag{2.8}$$

Figure 2.3 suggests the distribution of the values in $(x_n)_{n=1}^N$ over the interval $[0,1]$ approaches the function $g(x) = 2\sin(\pi x)$ as $N \to \infty$.

In light of the definition of an equidistributed sequence, we can reformulate the Sato-Tate Conjecture as follows:

Conjecture 2.16. (Sato-Tate Conjecture with a.d.f.'s) *Let E be an elliptic curve without complex multiplication. Then the Sato-Tate Conjecture is the statement that (x_n) defined in Equation 2.3 has the asymptotic distribution function*

$$ST(x) = x - \frac{\sin(2\pi x)}{2\pi} \tag{2.9}$$

referred to as the "Sato-Tate measure".

Throughout this section, we've been working with elliptic curves over the rationals. As mentioned in the introduction, though, much of this theory can be generalized to number fields. For more information, see Mazur [6].

3 The Extended Sato-Tate Conjecture and the Main Theorem

When one has a convergent sequence, a natural question to ask is "How fast does the convergence occur?" In 1996, Shigeki Akiyama and Yoshido Tanigawa submitted a paper on numerically calculating values of elliptic curve L-functions in the critical strip. [1] They addressed this question applied to the convergence in the Sato-Tate conjecture by using the language of discrepancies of sequences. Independently, in 2007, Barry Mazur, William Stein, and Chris Swierczewski asked the same question and found similar results in the language of L_2 and L_∞ norms of the difference of the area of sections the $(a_p/(2\sqrt{p}))$ histograms and the corresponding area of the conjectured semi-circle distribution.

In this section, we begin with a discussion on discrepancies. Using those tools, we formulate a statement about the rate of convergence of the Sato-Tate conjecture which we shall call the "Extended" Sato-Tate Conjecture. After a brief introduction to the Generalized Riemann Hypothesis, we present a fascinating relationship between the Extended Sato-Tate Conjecture and the Generalized Riemann Hypothesis; namely, that the former implies the latter! We present the proof of this extraordinary fact and provide some computational evidence for the truth of the Extended Sato-Tate Conjecture.

3.1 Discrepancy of Sequences

In Section 2.2, we studied equidistribution with respect to a distribution function, g. With the same philosophy in mind when we ask "how fast does a sequence converge", we can ask "how close is a sequence to a distribution function". The tool that helps answer that question is the notion of the discrepancy of a sequence. In short, the discrepancy measures how well a sequence is uniformly distributed with respect to a distribution function, that is, how much that sequence deviates from an ideal distribution.

In this section we introduce discrepancies and some of the properties that will be applicable to answering these questions, beginning with the fundamental definition:

Definition 3.1. (Discrepancy with Respect to a Distribution Function) *Let g be a non-decreasing function on $[0,1]$ with $g(0) = 0$ and $g(1) = 1$ and let (x_n) be a sequence in $\mathbb{R}$. The* **discrepancy of a finite sequence** $(x_n)_{n=1}^N$ **with respect to** *g, $D_N^{(g)}(x_n)$, is defined*

$$D_N^{(g)}(x_n) = \sup_{0\le\alpha\le 1} \left| \frac{A([0,\alpha); N; (x_n))}{N} - g(\alpha) \right|. \tag{3.1}$$

When $N \to \infty$, we call $D^{(g)}(x_n)$ the **discrepancy of the sequence** (x_n) **with respect to** *g.*

When $g(x) = x$, we simply say $D_N^{(g)}(x_n) = D_N(x_n)$ is the **discrepancy of the sequence** (x_n). The notion of discrepancy was first developed by Niederreiter and has applications in Quasi-Monte Carlo Methods of integration and pseudo-random number generation. (See [7] and [8].)

In particular, suppose g is the asymptotic distribution function for (x_n). By definition, the term within the supremum is equal to zero for all values of α. Therefore, we can think of the discrepancy of a sequence as a measure of how close the function g is to the corresponding a.d.f.. We formulate this natural connection between the discrepancy and these cumulative distribution functions.

Theorem 3.2. *The sequence (x_n) is equidistributed with respect to a continuous asymptotic distribution function g if and only if $\lim_{N\to\infty} D_N^{(g)}(x_n) = 0$.*

Proof. ($\Leftarrow$)
Suppose the supremum is achieved at α. If the supremum given Equation 3.1 approaches zero in the limit, then certainly, for all $N > 0$ there exists an $\epsilon > 0$ such that for all $\beta \in [0, 1]$,

$$\left|\frac{A([0,\beta); N; (x_n))}{N} - g(\beta)\right| \leq \left|\frac{A([0,\alpha); N; (x_n))}{N} - g(\alpha)\right| \leq \epsilon$$

Hence, for all $\beta \in [0, 1]$,

$$\lim_{N\to\infty} \frac{A([0,\beta); N; (x_n))}{N} = g(\beta).$$

($\Rightarrow$)
Let $m \geq 2$ and denote the subinterval $I_k \subset [0, 1]$ defined by $I_k = [k/m, (k+1)/m]$ for $0 \leq k \leq m-1$. Since (x_n) is equidistributed with respect to the a.d.f. g, for all m there exists an N_0 such that for $N \geq N_0$ and for every $k = 0, \ldots, m-1$ we have

$$\frac{1}{m}\left(1 - \frac{1}{m}\right) \leq \frac{A(I_k; N; (x_n))}{N} \leq \frac{1}{m}\left(1 + \frac{1}{m}\right) \tag{3.2}$$

Now consider an arbitrary subinterval $J = [\alpha, \beta) \subset [0, 1]$. There exist intervals J_1, J_2 which are finite unions of the intervals I_k such that $J_1 \subset J \subset J_2$, $g(\lambda(J)) - g(\lambda(J_1)) < 2/m$, and $g(\lambda(J_2)) - g(\lambda(J))$ where λ is the Lebesgue measure since g is a continuous function on $[0, 1]$ From Equation 3.2, we get that for all $N \geq N_0$,

$$\begin{aligned} g(\lambda(J_1))\left(1 - \frac{1}{m}\right) \leq \frac{A(J_1; N; (x_n)}{N} \leq \frac{A(J; N; (x_n)}{N} \leq \\ \leq \frac{A(J_2; N; (x_n)}{N} \leq g(\lambda(J_1))\left(1 + \frac{1}{m}\right). \end{aligned}$$

Consequently,

$$\left(g(\lambda(J)) - \frac{2}{m}\right)\left(1 - \frac{1}{m}\right) < \frac{A(J; N; (x_n))}{N} < \left(g(\lambda(J)) + \frac{2}{m}\right)\left(1 + \frac{1}{m}\right)$$

and, since $g(\lambda(J)) \leq 1$,

$$-\frac{3}{m} - \frac{2}{m^2} < \frac{A(J; N; (x_n))}{N} - g(\lambda(J)) < \frac{3}{m} + \frac{2}{m^2} \tag{3.3}$$

for all $N \geq N_0$. Since the bounds in Equation 3.3 are independent of J, we arrive at $D_N^{(g)}(x_n) \leq (3/m) + (2/m^2)$ for all $N \geq N_0$. But m was chosen arbitrarily and independent of N. Hence,

$$D_N^{(g)}(x_n) \to 0 \text{ as } N \to \infty$$

□

One can prove the following corollary relating the discrepancy to integral approximations using Theorem 3.2 and 2.15. However, when we discuss discrepancies in Section 3.1 and Koksma's lemma in Section 3.4, this interesting result will become immediate.

Corollary 3.3. $\lim_{N\to\infty} D_N^{(g)}(x_n) = 0$ *if and only if for every real-valued continuous function* f *on* $[0,1]$*,*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} f(\{x_n\}) = \int_0^1 f(x)dg(x).$$

In the next section, we apply the theory of discrepancy to the Sato-Tate Conjecture, giving us a more compact formulation of the statement than in Conjectures 2.4 and 2.16.

3.2 The Extended Sato-Tate Conjecture

Throughout this paper, we have mentioned an extended version of the Sato-Tate Conjecture and its connection with the Generalized Riemann Hypothesis. In this section, we will first restate the Sato-Tate Conjecture in terms of discrepancies. This will allow us to concretely and naturally ask more detailed questions. Recall the Sato-Tate Conjecture: Let E be an elliptic curve over $\mathbb{Q}$ without complex multiplication. Then for all $0 \le \alpha \le \beta \le 1$

$$\lim_{N\to\infty} \frac{\#\{x_n \in [\alpha,\beta) : n \le N\}}{N} = 2\int_\alpha^\beta \sin^2(\pi t)dt.$$

Richard Taylor's partial proof [13] and the calculations shown in Figures 2.2 and 2.3, and the appendix, give suggestive evidence that a general proof of the conjecture exists. In fact, it is indeed true for elliptic curves over $\mathbb{Q}$. Therefore, it seems justified to ask more detailed questions about the conjecture. In particular, about the rate of this convergence.

In the previous section, we proved an intimate relationship between asymptotic distribution functions and the discrepancy. In particular, Theorem 3.2 can be used to reformulate the Sato-Tate Conjecture in terms of discrepancies.

Conjecture 3.4. (Sato-Tate Conjecture with Discrepancy) *Let* E *be an elliptic curve without complex multiplication. Then*

$$\lim_{N\to\infty} D_N^{(ST)}(x_n) = 0 \tag{3.4}$$

Proof of equivalence to Conjecture 2.16. By Theorem 3.2, Equation 3.4 holds if and only if (x_n) is equidistributed with respect to the a.d.f. ST. This is precisely Conjecture 2.16. □

Given this formulation of the Sato-Tate Conjecture in terms of a convergent sequence of discrepancies, we have a starting point for analyzing the corresponding rate of convergence. Before performing any mathematical analysis at all, we can plot the discrepancy of the Sato-Tate sequence $(x_n)_{n=1}^N$ for large values of N. Figure 3.1 features $D_N^{(ST)}(x_n)$ for $N < 10^6$ for the elliptic curves given in Table 3.1.

As the plots suggest, $D_N^{(ST)}(x_n)$ approaches zero on the order of $O(1/N^k)$ for some $k > 0$. We can estimate the value of k for each elliptic curve by the identity

$$D_N^{(ST)}(x_n) = O\left(\frac{1}{N^k}\right) \quad \Leftrightarrow \quad \frac{\log D_N^{(ST)}(x_n)}{\log N} = O(k+\epsilon)$$

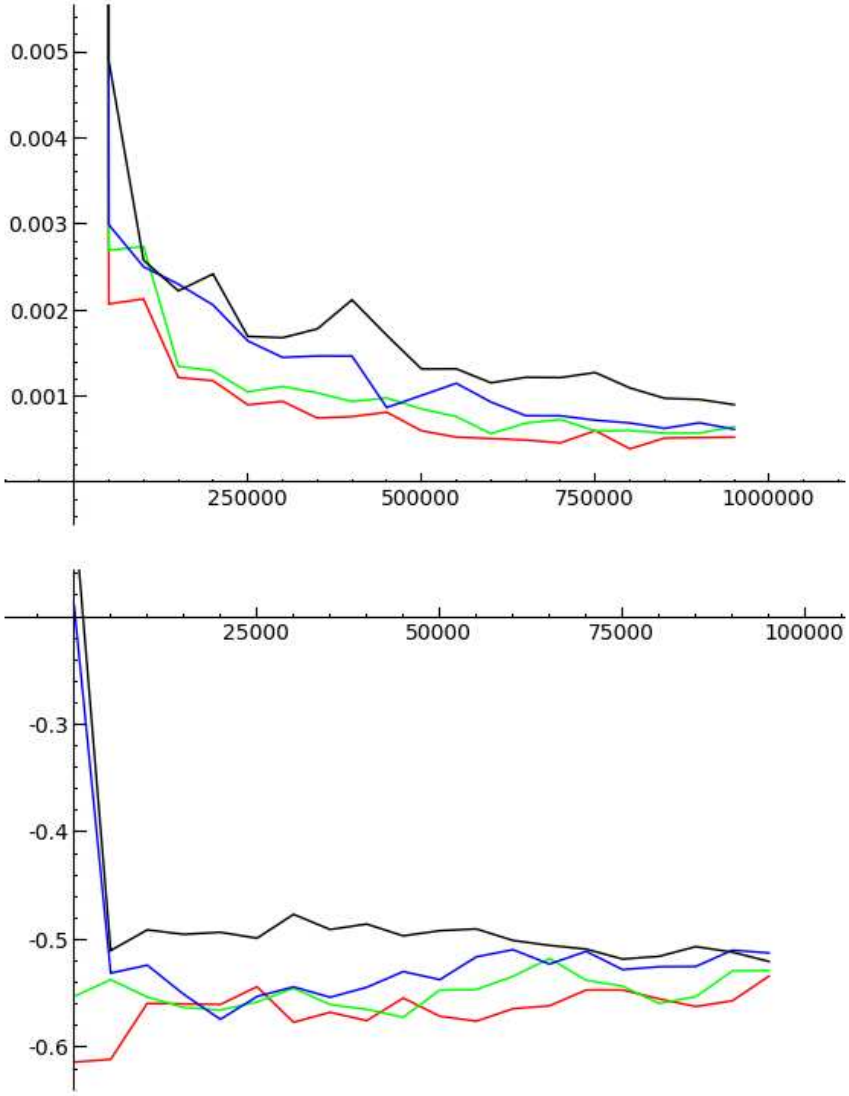

Figure 3.1: The discrepancy $D_N^{(ST)}(x_n)$ and the function $\log(D_N^{(ST)}(x_n))/\log(x)$ for the elliptic curves $E_0, \ldots, E_3$, defined in Table 3.1, in red, green, blue, and black; respectively. Notice the possible bound on the second set of graphs by the line $y = -1/2$.

E_0:	$y^2 + y = x^3 - x^2 - 10x - 20$
E_1:	$y^2 + y = x^3 - x$
E_2:	$y^2 + y = x^3 + x^2 - 2x$
E_3:	$y^2 + y = x^3 - 7x + 6$

Table 3.1: We shall use these four curves for many of the proceeding plots. Each E_r is the unique curve, up to isogeny, of algebraic rank r of lowest conductor.

for some $\epsilon > 0$. The second plot in Figure 3.1 presents these "log-discrepancies" for the same elliptic curves $E_0, \ldots, E_3$. It suggests that the log-discrepancy is bounded by $k = -1/2$. Calculations performed on other elliptic curves suggest that this bound holds for all elliptic curves without complex multiplication. This statement is exactly the one we wish to make on the rate of convergence occurring in the Sato-Tate Conjecture. Formally, we suggest

Conjecture 3.5. (The Extended Sato-Tate Conjecture) *For all $\epsilon > 0$,*

$$D_N^{(ST)}(x_n) = O(N^{-1/2+\epsilon}).$$

Additional computational support for the formulation of this conjecture can be found in the Appendix . Before discussing this conjecture's role in the main theorem, we first give a brief introduction to the theory of elliptic curve L-functions and the Generalized Riemann Hypothesis.

3.3 The Generalized Riemann Hypothesis

Since the Generalized Riemann Hypothesis is a statement about the zeros of elliptic curve L-functions, we will take the time in this section to highlight the relevant properties of these important gadgets. At the end of the section, we will give two statements for the Generalized Riemann Hypothesis— the classical statement in terms of the location of the zeroes of the L-function, and one in terms of the absence of zeroes on a certain subset of the complex plane. It turns out that under our approach, proving the main theorem of this paper will be much easier with the latter statement than the well-known former.

Our story begins with the classical Riemann zeta function.

Definition 3.6. (The Riemann Zeta Function) *For all* $\mathrm{Re}(s) > 1$*, the* **Riemann zeta function** *is the complex-valued function defined by the Dirichlet series*

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \tag{3.5}$$

By the comparison test and the identity $|\zeta(\sigma + it)| \leq \zeta(\sigma)|$ this series converges absolutely for $\mathrm{Re}(s) > 1$ and uniformly for $\mathrm{Re}(s) \geq 1 + \epsilon$ for all $\epsilon > 0$. There is a well known identity given in the theorem

Theorem 3.7. (Euler Product of $\zeta(s)$) *For* $\mathrm{Re}(s) > 1$*,*

$$\zeta(s) = \prod_p \frac{1}{1 - p^{-s}} \tag{3.6}$$

over all primes p.

Over the rational integers $\mathbb{Z}$, every non-zero ideal is principal and thus of the form $\mathfrak{a} = (n)$ for $n \geq 1$. So if we take the absolute norm N of an ideal of $\mathbb{Z}$, which is simply the number of all residue classes of $\mathbb{Z}$ modulo $\mathfrak{a}$, we have $N\mathfrak{a} = n$ and $\mathfrak{a}$ is a prime ideal if and only if $n = p$ where $p \in \mathbb{Z}$ is prime. From these remarks, we can generalize the zeta function to a number field K using the ideals in the ring of integers $\mathcal{O}_K$ with the absolute norm

$$N\mathfrak{a} = \#(\mathcal{O}_K/\mathfrak{a}), \quad \mathfrak{a} \text{ an ideal of } \mathcal{O}_K,$$

giving us the definition,

Definition 3.8. *For all* $\mathrm{Re}(s) > 1$*, the Dirichlet series expansion of the* **Dedekind zeta function** *for the number field* K *is*

$$\zeta_K(s) = \sum_{\mathfrak{a} \neq 0} \frac{1}{N\mathfrak{a}^s}$$

When $[K : \mathbb{Q}] = n$, there are at most n ideals $\mathfrak{a}$ of $\mathcal{O}_K$ such that $N\mathfrak{a}$ is equal to some prime p. This observation shows why we can take the same right half-plane of convergence as $\zeta(s)$. Now, these zeta functions contain information about the distribution of the prime ideals in $\mathcal{O}_K$. We can add further information by associating with a zeta function a corresponding L-function. We define

Definition 3.9. *The Dirichlet series expansion of the* **Dirichlet** L**-function** *associated with a character* χ *mod* m *is*

$$L(\chi, s) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}$$

with Euler product given by

$$L(\chi, s) = \prod_p \frac{1}{1 - \chi(p)p^{-s}}.$$

where both the Dirichlet series and the Euler product formulations converge absolutely for $\mathrm{Re}(s) > 1$

Let E be an elliptic curve. We can apply this theory to the extension $E/\mathbb{Q}$ akin to the theory discussed in Section 2.1. From the analysis in Section 2.1, we can try to determine the characters of the representation. We omit the details here, but the $a_E(p)$ are involved. Therefore we can associate to E the "generating function"

$$\tilde{L}(E, s) = \prod_{p \nmid \Delta} \frac{1}{1 - a_p p^{-s} + p^{1-2s}} \tag{3.7}$$

where $a_p = p + 1 - \#E(\mathbb{F}_p)$ and Δ is the discriminant of the elliptic curve. In other words, we only consider the Euler product over all primes of good reduction. Details aside, this product arises from applying the theory discussed around Definitions 3.8 and 3.9. However, for the purposes of this discussion, we need only to think of $\tilde{L}(E, s)$ as a function of a complex variable and an elliptic curve.

We will state without proof that $\tilde{L}(E, s)$ defines a complex analytic function on the right half-plane $\mathrm{Re}(s) > 3/2$. A theorem due to Wiles et. al. [14] shows that $\tilde{L}(E, s)$ can be analytically continued to an entire function. For more on the function $\tilde{L}(E, s)$, see [2].

In complex analysis, one approach to understanding a function is to start by locating the zeroes and poles. One strategy in finding the zeroes of $\tilde{L}(E, s)$ is to examine $\log \tilde{L}(E, s)$ instead. By the

properties of the complex logarithm, if the latter function is analytic in some region $S \subset \mathbb{C}$, then $\log \tilde{L}(E,s)$ has no poles in S. Hence, $L(E,s)$ has no zeroes in S. We will use the following lemma to show that $L(E,s)$ has no zeroes in the right half plane $\mathrm{Re}(s) > 3/2$ as well as in the main theorem of this paper.

Lemma 3.10. *Let E be an elliptic curve. Then,*

$$\log \tilde{L}(E,s) = \sum_{p \nmid \Delta} a_p p^{-s} + O\left(\sum_{p \nmid \Delta} p^{1-2\sigma}\right)$$

where the error term $O(\sum p^{1-2\sigma})$ is a complex analytic function on the right half plane $\mathrm{Re}(s) > 1$.

Proof. Using the Taylor expansion of $\log(1-z)$ for $|z| < 1$, we can write

$$\begin{aligned}
\log \tilde{L}(E,s) &= \sum_{p \nmid \Delta} -\log\left(1 - a_p p^{-s} + p^{1/2-s}\right) \\
&= \sum_{p \nmid \Delta} \sum_{n=1}^{\infty} \frac{(a_p p^{-s} - p^{1/2-s})^n}{n} \\
&= \sum_{p \nmid \Delta} \left(a_p p^{-s} - p^{1/2-s} + \sum_{n=2}^{\infty} \frac{(a_p p^{-s} - p^{1/2-s})^n}{n}\right)
\end{aligned}$$

Note that by Hasse's bound on the a_p, for all $\mathrm{Re}(s) > 1$,

$$\left|a_p p^{-s} - p^{1/2-s}\right| = \left|\frac{a_p - \sqrt{p}}{p^s}\right| \leq \left|\frac{\sqrt{p}}{p^s}\right| \leq \left|\frac{1}{\sqrt{p}}\right| \leq 1$$

Thus, the convergence of the Taylor expansion for $\log(1-z_p)$, where $z_p = a_p p^{-s} - p^{1/2-s}$, occurs for all p prime.

In the outer sum, every summand after the first involves a term of the form $p^{1/2-s}$ with coefficients $(a_p p^{-s})^{n-k}(p^{1/2-s})^{k-1}/n$. We can collect these terms under the inner sum and write

$$\log \tilde{L}(E,s) = \sum_{p \nmid \Delta} a_p p^{-s} + O\left(\sum_{n=1}^{\infty} p^{1/2-s}\right)$$

Finally, by comparison with the series $\sum n^{-(1+\epsilon)}$ for all $\epsilon > 0$, the series in the error term converges absolutely and, by the Weierstrass M-test, converges uniformly. Hence, the error term $\left(\sum p^{1/2-s}\right)$ is analytic on $\mathrm{Re}(s) > 1$. □

If we make the right half plane just a bit smaller, in particular, if we look far enough away from the critical strip, the function $\log \tilde{L}(E,s)$ itself is analytic. For all $\mathrm{Re}(s) > 3/2$, the series $\sum a_p p^{-s}$ converges absolutely by the comparison test and Hasse's bound, thus proving the lemma

Lemma 3.11. $\log \tilde{L}(E,s)$ *is analytic on the right half-plane* $\mathrm{Re}(s) > 3/2$.

which immediately implies

Theorem 3.12. *$\tilde{L}(E,s)$ has no zeroes on the right half-plane* $\mathrm{Re}(s) > 3/2$.

The proof of Theorem 3.12 from Lemma 3.10 isn't terribly complicated and only requires some analytical trickery. However, many statements to be made about $\tilde{L}(E,s)$ are of extraordinary depth, such as the recent entire function extension theorem of Wiles. Even in general, one must be careful when speaking about the analyticity of complex infinite series. Despite the precautions, we have the following incredible statement about the nature of the L-function for elliptic curves.

Proposition 3.13. (Generalized Riemann Hypothesis) *Let E be an elliptic curve. The following are equivalent statements of the Generalized Riemann Hypothesis:*

1. *$\tilde{L}(E,s)$ has no zeroes on the right half-plane* $\mathrm{Re}(s) > 1$.
2. *The zeroes of $\tilde{L}(E,s)$ all lie on the critical line* $\mathrm{Re}(s) = 1/2$.

Proof. The proof of the equivalence of these two statements follows from a functional equation for $\tilde{L}(E,s)$ that gives an explicit relationship between $\tilde{L}(E,s)$ and $\tilde{L}(E,2-s)$. See [3, p.317] for a description of the functional equation and its implications for the L-function. □

When we arrive at the main theorem, we will derive the Generalized Riemann Hypothesis from the Extended Sato-Tate Conjecture by using the first formulation stated in Proposition 3.13. In the next section, we will start on the path toward this proof using what we have discussed thus far.

3.4 The Main Theorem

We will now begin assembling the pieces to prove the main theorem of this paper. A lemma due to Koksma gives a formula for the rate of convergence of the Riemann sum of a function over a partition $(x_n)_{n=1\ldots N}$ of the interval $[0,1]$ to its integral in terms of the discrepancy $D_N^{(g)}(x_n)$ for some distribution function g. Along with several statements about the convergence of Dirichlet series, we will uses these facts for the proof of the main theorem of this paper.

We begin with the statement and proof of Koksma's Lemma.

Lemma 3.14. (Koksma's Lemma) *Let f be a real valued function on $[0,1]$. Suppose that f has bounded variation. Let g be a real-valued continuous strictly increasing function on $[0,1]$ for which $g(0) = 0$ and $g(1) = 1$. Then we have*

$$\left| \frac{1}{N} \sum_{n=1}^{N} f(x_n) - \int_0^1 f(t) dg(t) \right| \leq D_N^{(g)}(x_n) V(f)$$

for any sequence of real numbers (x_n) in $[0,1]$. Here, $V(f)$ is the total variation of f in $[0,1]$ and $D_N = D_N^{(x)}$ is the discrepancy function given in Definition 3.1.

Proof. If f is of bounded variation and continuous on I, then Koksma's inequality can be proven very quickly using integration by parts. [7, p.964] Define

$$R_N^{(g)}(t) := \frac{A([0,t); N; (x_n))}{N} - g(t)$$

where $A([\alpha,\beta);N;(x_n))$ is the counting function of the sequence (x_n) defined in Section 2.2. Then, using integration by parts,

$$\begin{aligned}\int_0^1 R_n^{(g)}(t)df(t) &= \frac{1}{N}\sum_{n=1}^{N}\int_0^1 \chi_{[0,t)}(x_n)df(t) - \int_0^1 g(t)df(t)\\ &= \frac{1}{N}\sum_{n=1}^{N}(f(1)-f(x_n)) - g(1)f(1) + \int_0^1 f(t)dg(t)\\ &= -\frac{1}{N}\sum_{n=1}^{N} f(x_n) + \int_0^1 f(t)dg(t),\end{aligned}$$

where χ_S is the characteristic function on $S \subset [0,1]$, and so

$$\begin{aligned}\left|\frac{1}{N}\sum_{n=1}^{N} f(x_n) - \int_0^1 f(t)dg(t)\right| &= \left|\int_0^1 R_N^{(g)}(t)df(t)\right|\\ &\le V(f)\sup_{0\le t\le 1}\left|R_N^{(g)}(t)\right|\\ &= V(f)D_N^{(g)}(x_n)\end{aligned}$$

For a proof not assuming the continuity of f, see [4, p.143] □

An immediate corollary from Koksma's lemma is that if g is the asymptotic distribution function to a sequence (x_n), Theorem 3.2 implies that the average of f over the sequence closely approximates the integral of f with respect to g. That is, the sequence (x_n) forms a partition fine enough to approximate the corresponding integral. More about how discrepancies and asymptotic distribution functions can be used to approximate integrals can be found in [7] and [8].

Corollary 3.15. *If a sequence (x_n) has the a.d.f. g and if f has bounded variation, then*

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} f(x_n) = \int_0^1 f(t)dg(t).$$

3.4.1 Dirichlet Series

As shown in Lemma 3.10, we can derive properties of elliptic curve L-functions by looking at corresponding Dirichlet series. The following propositions about Dirichlet series will be incorporated in the proof of the main theorem. We begin with the following lemma.

Lemma 3.16. *If $s = \sigma + it$ and $\sigma \neq 0$ then for all $n \in \mathbb{Z}, n > 0$,*

$$\left|\frac{1}{n^s} - \frac{1}{(n+1)^s}\right| \le \frac{|s|}{\sigma}\left(\frac{1}{n^\sigma} - \frac{1}{(n+1)^\sigma}\right)$$

This lemma can then be used to prove the analyticity of Dirichlet series based on the behavior of the coefficient terms. In particular, we can extend the domain of analyticity. The following theorem does exactly that.

Lemma 3.17. *Suppose* $\left|\sum_{n=1}^{N} a_n\right| = O(N^{1/2+\epsilon})$ *for all* $\epsilon > 0$. *Then the Dirichlet series* $\sum a_n n^{-s}$ *is analytic on* $\mathrm{Re}(s) > 1/2$.

Proof. Let $A_N = \sum_{n=1}^{N} a_n$. Using Abel's summation formula, we can write

$$\sum_{n=1}^{N} \frac{a_n}{n^s} = \frac{A_N}{N^s} + \sum_{k=1}^{N-1} A_k \left(\frac{1}{k^s} - \frac{1}{(k+1)^s}\right).$$

Note that the first term converges.

By Lemma 3.16 and the bound on A_k,

$$\begin{aligned}\left|\sum_{k=1}^{N} A_k \left(\frac{1}{k^s} - \frac{1}{(k+1)^s}\right)\right| &\leq \sum_{k=1}^{N} |A_k| \left|\frac{1}{k^s} - \frac{1}{(k+1)^s}\right| \\ &\leq \frac{C|s|}{\sigma} \sum_{k=1}^{N} k^{1/2} \left(\frac{1}{k^\sigma} - \frac{1}{(k+1)^\sigma}\right)\end{aligned}$$

for all $\sigma > 1/2$ and some constant $C > 0$. Note that by adding and subtracting a $(k+1)^{1/2}$ term, we can write

$$\begin{aligned}\sum_{k=1}^{N} k^{1/2} \left(\frac{1}{k^\sigma} - \frac{1}{(k+1)^\sigma}\right) &\leq \sum_{k=1}^{N} \frac{1}{k^{\sigma-1/2}} - \frac{\left(k^{1/2} + (k+1)^{1/2} - (k+1)^{1/2}\right)}{(k+1)^\sigma} \\ &\leq \sum_{k=1}^{N} \left(\frac{1}{k^{\sigma-1/2}} - \frac{1}{(k+1)^{\sigma-1/2}}\right) + \frac{\left(-k^{1/2} + (k+1)^{1/2}\right)}{(k+1)^\sigma}\end{aligned}$$

The first summand forms a term of a telescoping series. Thus, we only need to consider the second summand of the above series. Note,

$$\begin{aligned}\sum_{k=1}^{N} \frac{-k^{1/2} + (k+1)^{1/2}}{(k+1)^\sigma} &= \sum_{k=1}^{N} \frac{1}{(k+1)^\sigma \left((k+1)^{1/2} + k^{1/2}\right)} \\ &= \sum_{k=1}^{N} \frac{1}{(k+1)^{\sigma+1/2} + (k+1)^\sigma n^{1/2}} \leq \sum_{k=1}^{N} \frac{1}{2k^{\sigma+1/2}}.\end{aligned}$$

This series converges for all $\sigma > 1/2$. Thus, the series

$$\frac{C|s|}{\sigma} \sum_{k=1}^{N} k^{1/2} \left(\frac{1}{k^\sigma} - \frac{1}{(k+1)^\sigma}\right)$$

converges absolutely on all compact subsets of the right half plane $\mathrm{Re}(s) > 1/2$ as $N \to \infty$. By Morerra's Theorem, the above series is analytic on $\mathrm{Re}(s) > 1/2$ as $N \to \infty$ and therefore so is $\sum_{n=1}^{\infty} a_n n^{-s}$. ∎

3.4.2 The Main Theorem

Without further ado, we present the main theorem of the paper:

Theorem 3.18. (Main Theorem) *Let E be an elliptic curve defined over $\mathbb{Q}$ without complex multiplication. Then the Extended Sato-Tate Conjecture implies the generalized Riemann Hypothesis.*

Proof. Let E be an elliptic curve and consider the corresponding L-function $L(E,s)$. To show that this is analytic on $\mathrm{Re}(s) > 0$, it is enough to show that $\tilde{L}(E,s)$ is analytic on $\mathrm{Re}(s) > 0$. By Lemma 3.10,

$$\log \tilde{L}(E,s) = \sum_{p \nmid \Delta} a_p p^{-s} + O\left(\sum_{p \nmid \Delta} p^{1-2\sigma}\right)$$

where the error term $O(\sum p^{1-2\sigma})$ is analytic on the right half plane $\mathrm{Re}(s) > 1$.

Now, if $\sum_p a_p p^{-s}$ is analytic then certainly $\sum_{p \nmid \Delta} a_p p^{-s}$ is also analytic. Therefore, we can disregard the primes of bad reduction and write

$$\sum_p a_p p^{-s} = 2 \sum_p \cos(\theta_p) p^{1/2-s}.$$

Let $f(t) = \cos(\pi t)$ and $g(t) = ST(t)$. Since f has bounded variation on $[0,1]$ and since g is continuous and strictly increasing over $(0,\pi)$. We can apply Koksma's Lemma 3.14 and, with an appropriate change of variables, arrive at

$$\begin{aligned}\left|\frac{1}{N}\sum_{n=1}^{N} \cos(\theta_{p_n}) - \int_0^{\pi} \cos(t) dST(t)\right| &\leq D_N^{(ST)}(x_n) V(f) \\ &\leq 2 D_N^{(ST)}(x_n)\end{aligned}$$

for all $N \in \mathbb{Z}, N > 0$. Now, $V(f)$ and $\int \cos(t) dST(t)$ are finite. Thus, by the Extended Sato-Tate Conjecture, for all $N \in \mathbb{Z}, N > 0$ and for all $\epsilon > 0$,

$$\left|\frac{1}{N}\sum_{n=1}^{N} \frac{a_p}{\sqrt{p}}\right| = O\left(N^{-1/2+\epsilon}\right) \quad \Rightarrow \quad \left|\sum_{n=1}^{N} \frac{a_p}{\sqrt{p}}\right| = O\left(N^{1/2+\epsilon}\right)$$

If we perform the change of variables, $s - 1/2 \mapsto s$, Lemma 3.17 implies that the series $\sum a_p p^{-s}$ is analytic on the right half plane $\mathrm{Re}(s) > 1$. Therefore, $\tilde{L}(E,s)$, and hence $\tilde{L}(E,s)$, has no zeroes on $\mathrm{Re}(s) > 1$, proving the Generalized Riemann Hypothesis for elliptic curves. □

This amazing result reduces the problem of solving the Generalized Riemann Hypothesis for elliptic curves to solving the Extended Sato-Tate Conjecture. However, this also implies that the Extended Sato-Tate Conjecture is at least equally difficult, if not more difficult, than the 120-year-old problem. In addition to the implication in the main theorem, Akiyama and Tanigawa mention in even more brevity that the converse is true, implying that EST and GRH are, in fact, equivalent statements. Unfortunately, a proof of this claim hasn't been published.

Aesthetically, the Extended Sato-Tate Conjecture appears to be more efficient means of computationally verifying the claim of the Generalized Riemann Hypothesis for elliptic curve L-functions. Imagine trying to computationally show that $L(E, s)$ has no zeroes on the right half-plane $\mathrm{Re}(s) > 1$. Computing $D_N^{(ST)}(x_n)$, though still taxing for very large values of N, is easier to conceptually grasp and, more importantly, is instead a one-dimensional problem. However, much work has gone into developing means of computationally verifying GRH for elliptic curve L-functions. Rubenstein, in particular, developed a sophisticated algorithm for actually proving that the zeroes up to a given bound all lie on the critical line. Nevertheless, one can refer to the appendix for additional plots supporting the claim of the Extended Sato-Tate Conjecture while keeping in mind they also imply the Generalized Riemann Hypothesis for elliptic curves.

4 Conclusion

Given the immensity of the Generalized Riemann Hypothesis, it's always exciting to see how other branches of mathematics relate to the problem. There's no a priori evidence that the Sato-Tate Conjecture, a mere 70-year-old statement, could serve as an approach to proving GRH. However, the analysis of the implications of the main theorem need not stop here. We wrap up this paper by noting a possible connection with another Millennium Prize Problem.

4.1 A Closer Look at the Series $\sum a_p/\sqrt{p}$

Recall in the proof of the main theorem that under the assumption of the Extended Sato-Tate Conjecture, we constructed a series involving the error terms $a_p/(2\sqrt{p})$ along with the following bound:

$$\sum_{n=1}^{N} \frac{a_{p_n}}{\sqrt{p_n}} = O\left(N^{1/2+\epsilon}\right). \tag{4.1}$$

Since applying Lemma 3.17 to this statement immediately produces the desired result, we really only need the truth of this statement in order to prove GRH for elliptic curves, not necessarily the truth of the Extended Sato-Tate Conjecture even though they are obviously related. Therefore, one might be interested in plotting $\sum^{N} a_{p_n}/\sqrt{p_n}$ for various bounds N. Before doing so, we perform some scalings akin to the analysis of the discrepancy function $D_N^{(ST)}(x_n)$ in Section 3.1. By multiplying both sides of Equation 4.1 by $1/\sqrt{N}$, we have

$$\frac{1}{\sqrt{N}} \sum_{n=1}^{N} \frac{a_{p_n}}{\sqrt{p_n}} \approx O\left(C\right). \tag{4.2}$$

Experts currently examining this sum also introduce a "smoothing factor", $\log(N)$. Multiplying by this term doesn't seem to cause the sum in equation 4.2 to diverge—that is, it is only conjectural that the addition of this smoothing factor produces a still convergent series. Given the computational evidence for the truth of Equation 4.1, we therefore expect the convergence

$$\frac{\log(N)}{\sqrt{N}} \sum_{n=1}^{N} \frac{a_{p_n}}{\sqrt{p_n}} \to C. \tag{4.3}$$

See Figure 4.1 for computational evidence for this convergence. The plots perhaps suggest that the series stays within some bound from a respective line.

The actual value of C is conjectured by Sarnak et al. to have a close connection with the Birch and Swinnerton-Dyer Conjecture. One statement of BSD, the more simple of two, declares that the "algebraic rank", that is, if K is a number field then the rank of $E(K)$ as a $\mathbb{Z}$-module is equal to the so-called "analytic rank": the multiplicity of zero of the corresponding L-function at $s = 1$.

Conjecture 4.1. (Sarnak, et. al.) *Let E be an elliptic curve without complex multiplication. Then*

$$\frac{\log(N)}{\sqrt{N}} \sum_{n=1}^{N} \frac{a_{p_n}}{\sqrt{p_n}} \to 1 - 2r$$

where r is the rank of E.

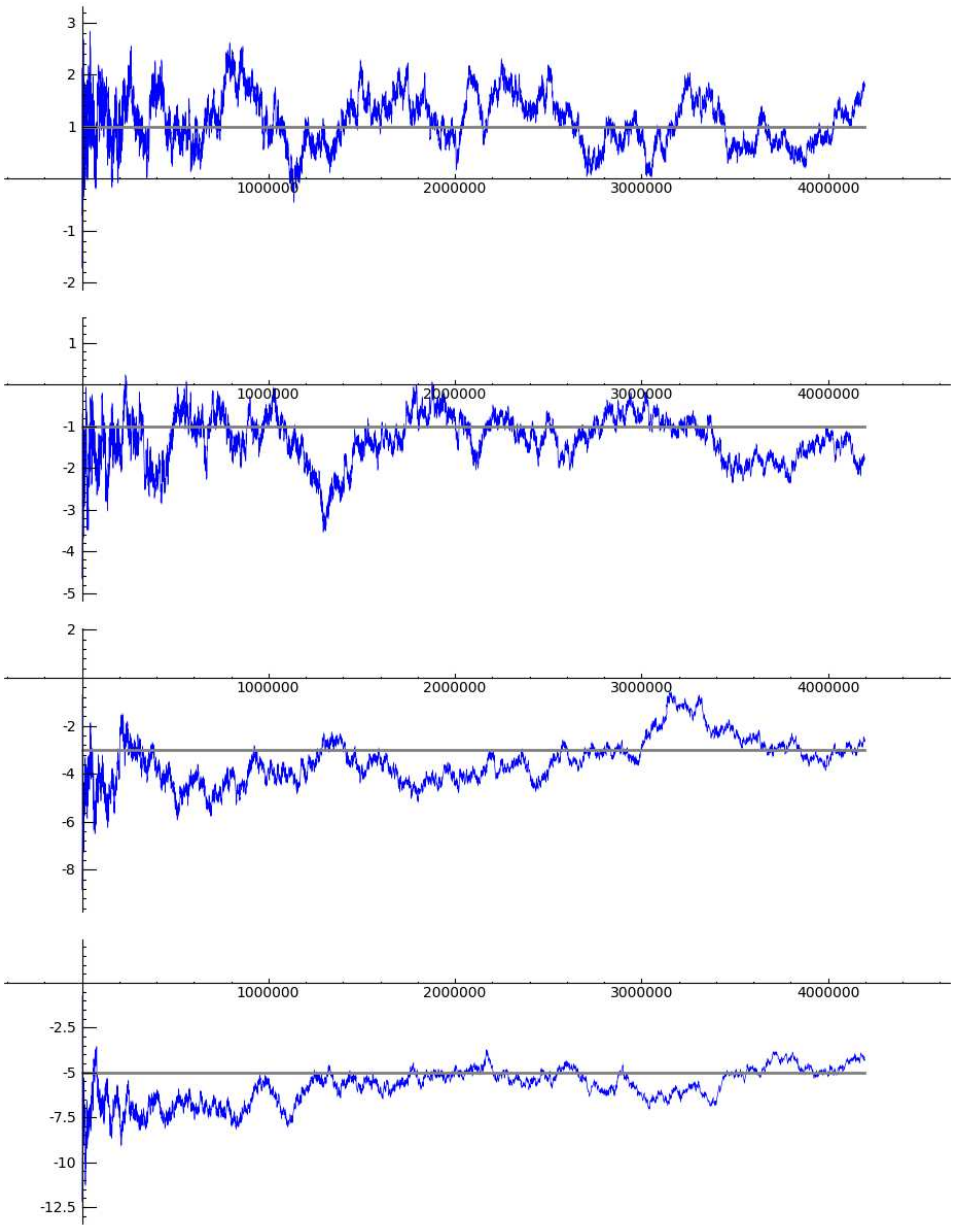

Figure 4.1: $\frac{\log(X)}{\sqrt{X}} \sum_{p<X} \frac{a_p}{\sqrt{p}}$ for the elliptic curves $E_0, \ldots, E_3$, $X < 2^{22}$. Note how each series hovers around the line $1 - 2r$ where $r =$ the rank of E.

So from the Extended Sato-Tate Conjecture, it is perhaps possible to derive the relationship above between an analytic object and the algebraic rank of an elliptic curve E, suggesting a connection between EST and BSD.

It's amazing how one statement, beginning with an observation about frequency histograms, can extend to these two very important problems in number theory: the Generalized Riemann Hypothesis and the Birch and Swinnerton-Dyer Conjecture. With these relationships in mind, the Extended Sato-Tate Conjecture serves as a possible means of proving the Generalized Riemann Hypothesis for elliptic curves much like how the Taniyama-Shimura conjecture was used to prove Fermat's Last Theorem.

5 Appendix

The appendix contains figures containing additional computational verification to claims made by the Sato-Tate Conjecture and the Extended Sato-Tate Conjecture. We present plots of

- the discrepancy function for elliptic curves of ranks 0-6,
- and additional plots of the function $F(X) = \frac{\log(X)}{\sqrt{X}} \sum_{p<X} \frac{a_p}{\sqrt{p}}$ for elliptic curves of ranks 0-6.

References

[1] S. Akiyama and Y. Tanigawa, *Calculation of values of l-functions associated to elliptic curves*, Mathematics of Computation **68** (1999), no. 227, 1201–1231.

[2] The Birch and a computational approach Swinnerton-Dyer conjecture, *Stein, w.a.*, http://www.wstein.org/papers/ (1991).

[3] D. Husemoller, *Elliptic curves*, Springer-Verlag, New York, 2004.

[4] L. Kuipers and H. Niederreiter, *Uniform distribution of sequences*, John Wiley and Sons, New York, 1974.

[5] B. Mazur, *Rational isogenies of prime degree*, Inventiones Math. **44** (1978), 129–162.

[6] ———, *Finding meaning in error terms*, Bulletin of the American Mathematical Society **45** (2008), 185–228.

[7] H. Niederreiter, *Quasi-monte carlo methods and pseudo-random numbers*, Bulletin of the American Mathematical Society **84** (1978), no. 6.

[8] ———, *Random number generation and quasi-monte carlo methods*, Society for Industrial and Applied Mathematics, Philadelphia, 1992.

[9] K.A. Ribet, *Galois represenations and modular forms*, Bulletin of the American Mathematical Society **32** (1995), 53–79.

[10] J. P. Serre, *Abelian l-adic representations*, W. A. Benjamin Inc., 1968.

[11] J. Silverman, *The arithmetic of elliptic curves*, Graduate Texts in Mathematics, Springer-Verlag, New York, 1986.

[12] J. Silverman and J. Tate, *Rational points on elliptic curves*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 1992.

[13] R. Taylor, *Automoprhy for some l-adic lifts of automorphic mod l galois representations. ii*, (pre-publication) (2006).

[14] A.J. Wiles, *Modular elliptic curves and fermat's last theorem*, Annals of Mathematics **141** (1995), no. 3, 443–551.

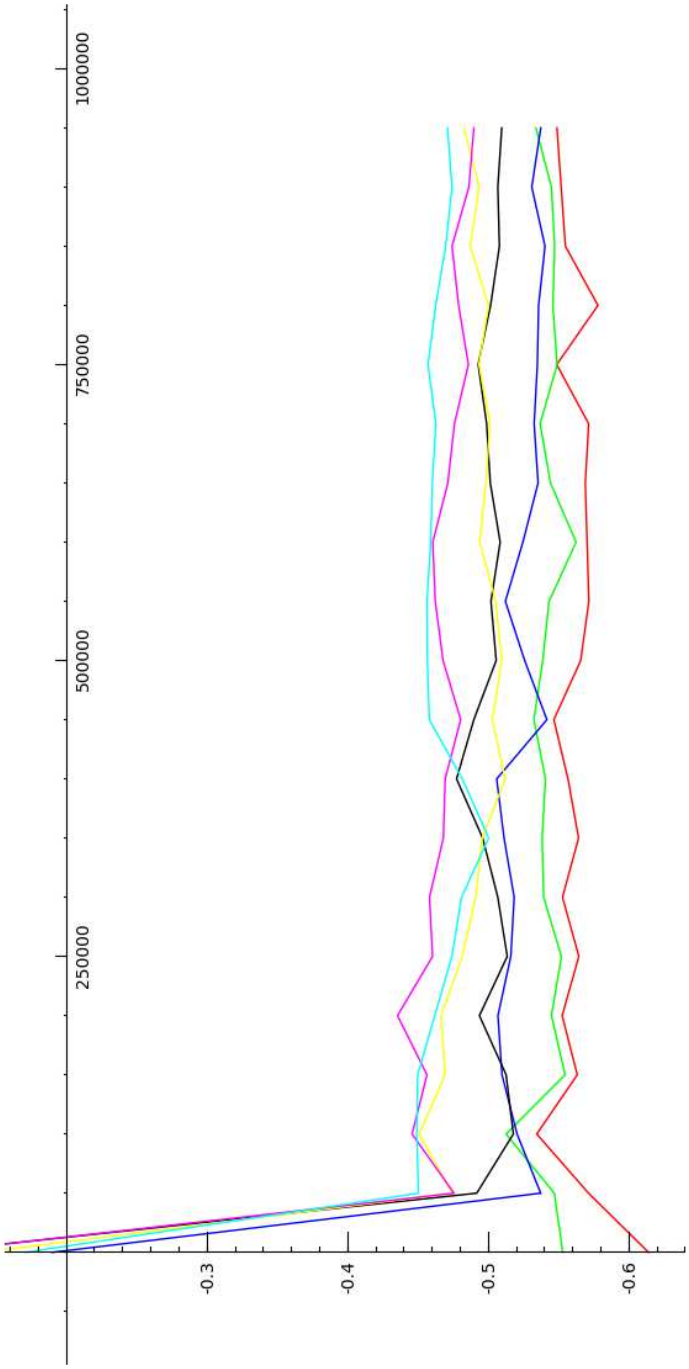

Figure 5.1: $D_N^{(ST)}(x_n)$ for elliptic curves of ranks 0-6, each with smallest conductor in their rank. $N < 10^6$.

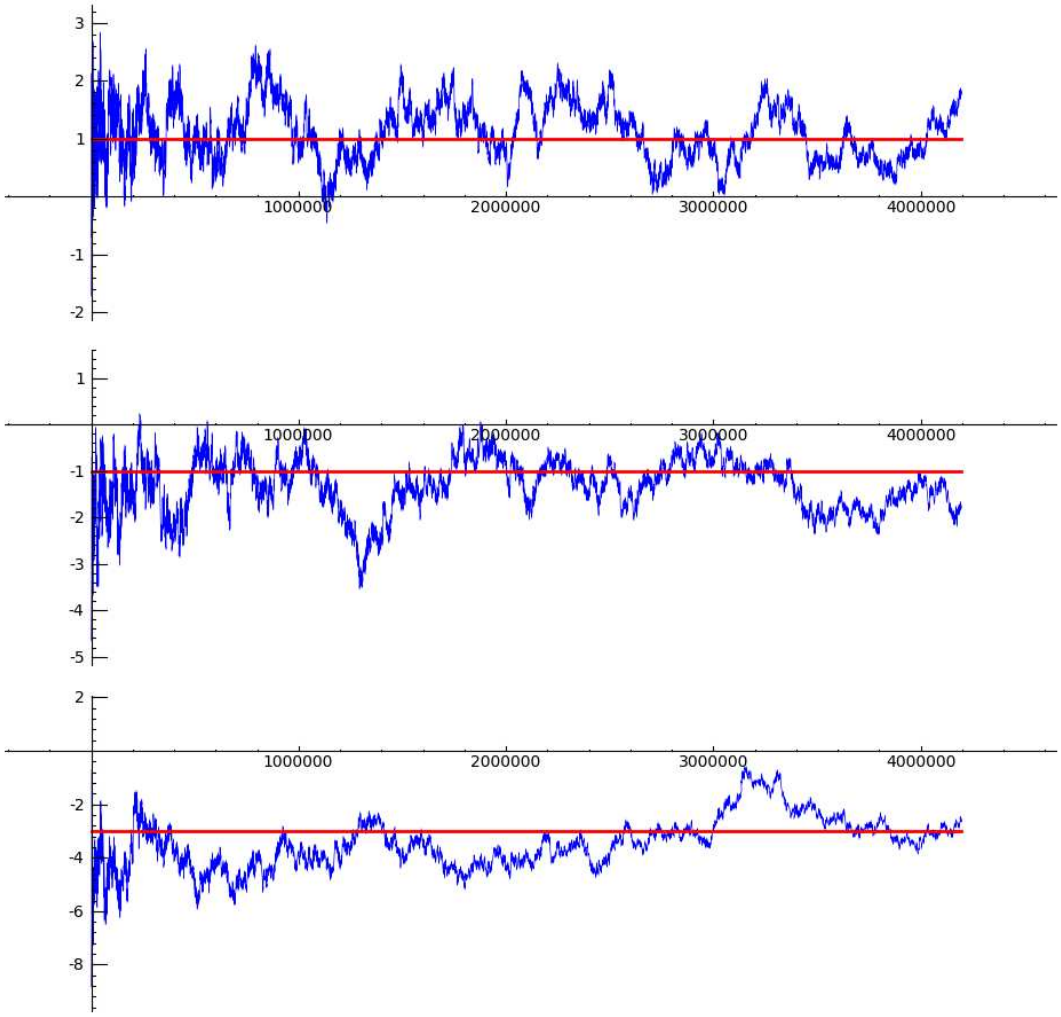

Figure 5.2: $\frac{\log(X)}{\sqrt{X}} \sum_{p<X} \frac{a_p}{\sqrt{p}}$ for elliptic curves of ranks 0, 1, and 2; each of smallest conductor in their rank.

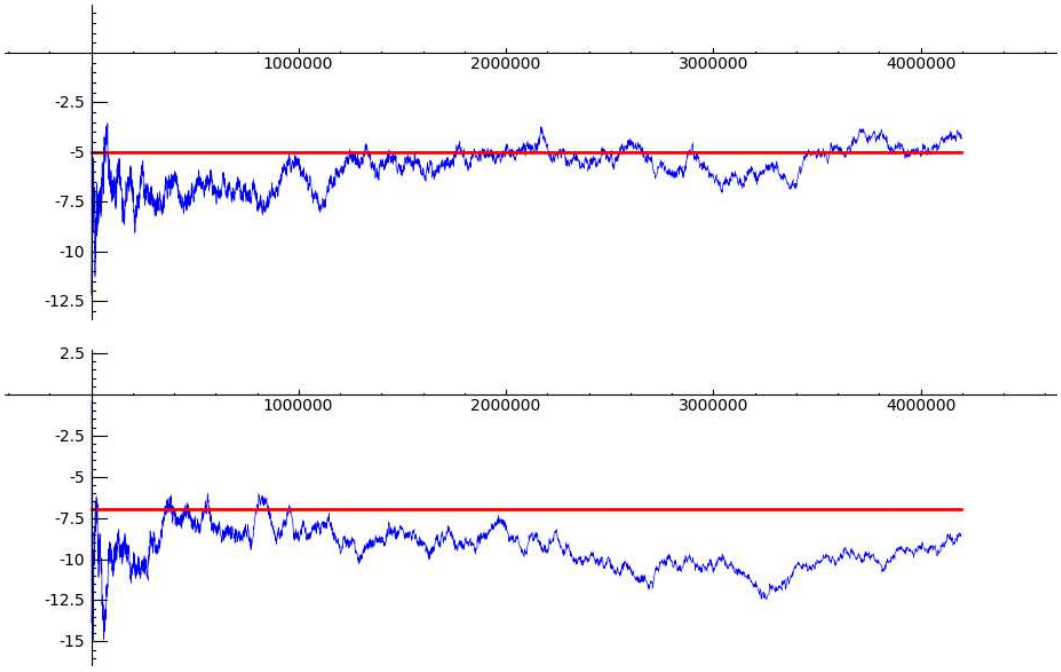

Figure 5.3: $\frac{\log(X)}{\sqrt{X}} \sum_{p<X} \frac{a_p}{\sqrt{p}}$ for elliptic curves of ranks 3 and 4, each of smallest conductor in their rank.

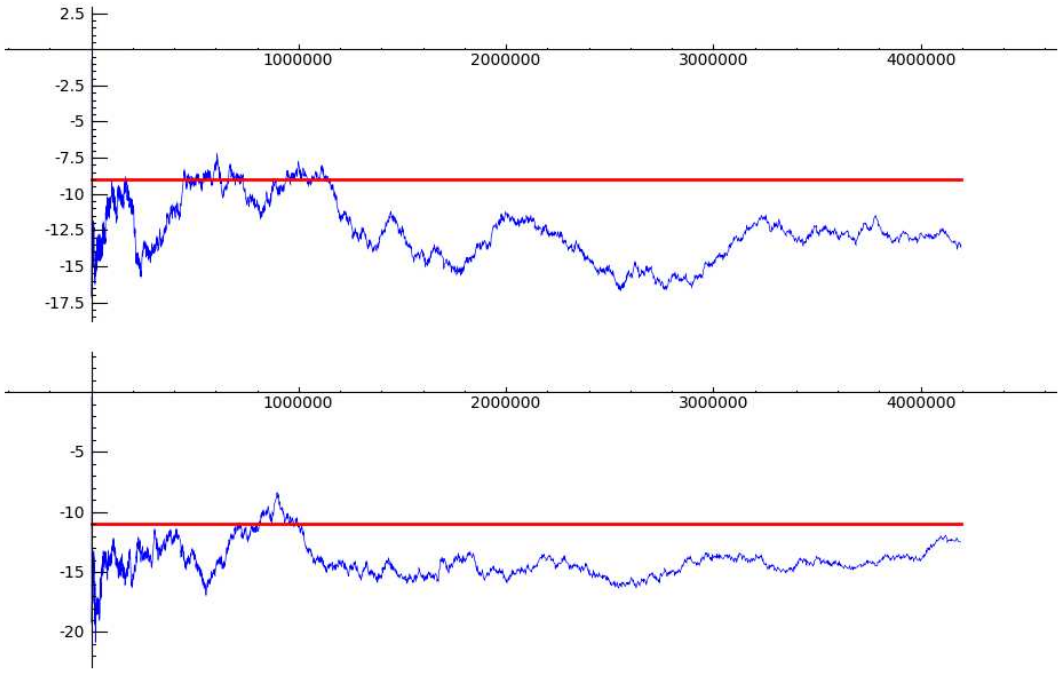

Figure 5.4: $\frac{\log(X)}{\sqrt{X}} \sum_{p<X} \frac{a_p}{\sqrt{p}}$ for elliptic curves of ranks 5 and 6, each of smallest conductor in their rank.