EFFECTIVE EQUIDISTRIBUTION AND THE SATO-TATE LAW FOR FAMILIES OF ELLIPTIC CURVES

STEVEN J. MILLER AND M. RAM MURTY

ABSTRACT. Extending recent work of others, we provide effective bounds on the family of all elliptic curves and one-parameter families of elliptic curves modulo p (for p prime tending to infinity) obeying the Sato-Tate Law. We present two methods of proof. Both use the framework of Murty-Sinha [MS]; the first involves only knowledge of the moments of the Fourier coefficients of the L-functions and combinatorics, and saves a logarithm, while the second requires a Sato-Tate law. Our purpose is to illustrate how the caliber of the result depends on the error terms of the inputs and what combinatorics must be done.

1. INTRODUCTION

Recently M. Ram Murty and K. Sinha [MS] proved effective equidistribution results showing the eigenvalues of Hecke operators on the space $S(N,k)$ of cusp forms of weight k and level N agree with the Sato-Tate distribution. Our goal here is to use their framework to prove similar results for families of elliptic curves. We shall do this for the family of all elliptic curves and for one-parameter families of elliptic curves.

We first review notation and previous results. Let $E : y^2 = x^3 + Ax + B$ with $A, B \in \mathbb{Z}$ be an elliptic curve over $\mathbb{Q}$ with associated L-function

$$L(E,s) \;=\; \sum_{n=1}^{\infty} \frac{a_E(n)}{n^s} \;=\; \prod_p \left(1 - \frac{a_E(p)}{p^s} + \frac{\chi_0(p)}{p^{2s-1}}\right)^{-1}, \tag{1.1}$$

2000 *Mathematics Subject Classification.* 11H05 (primary) 11K38, 14H52, 11M41 (secondary).

Key words and phrases. Sato-Tate, Elliptic Curves, Erdös-Turan, Effective Equidistribution.

The authors would like to thank Andrew Granville for pointing out the implicit normalization in Birch's work, Igor Shparlinski for discussions on previous results as well as sharing his recent work, and Frederick Strauch for conversations on a hypergeometric proof of Lemma A.3; the first named author would also like to thank Cameron and Kayla Miller for quietly sleeping on him while many of the calculations were done. Much of this paper was written when the authors attended the Graduate Workshop on L-functions and Random Matrix Theory at Utah Valley University in 2009, and it is a pleasure to thank the organizers. The first named author was partly supported by NSF grants DMS0600848 and DMS0970067. The second named author was partially supported by an NSERC Discovery grant.

where $\Delta = -16(4A^3 + 27B^2)$ is the discriminant of E, χ_0 is the principal character modulo Δ, and

$$\begin{aligned} a_E(p) &= p - \#\{(x,y) \in (\mathbb{Z}/p\mathbb{Z})^2 : y^2 \equiv x^3 + Ax + B \bmod p\} \\ &= -\sum_{x \bmod p} \left(\frac{x^3 + Ax + B}{p}\right). \end{aligned} \tag{1.2}$$

By Hasse's bound we know $|a_E(p)| \le 2\sqrt{p}$, so we may write $a_E(p) = 2\sqrt{p}\cos\theta_E(p)$, where we may choose $\theta_E(p) \in [0,\pi]$. See [Sil1, Sil2, ST] for more details and proofs of all the needed properties of elliptic curves.

How the $a_E(p)$'s vary is of great interest. One reason for this is that they encode local data (the number of solutions modulo p), and are then combined to build the L-function, whose properties give global information about E. For example, the Birch and Swinnerton-Dyer conjecture [BS-D1, BS-D2] states the order of the group of rational solutions of E equals the order of vanishing of $L(E,s)$ at the central point. While we are far from being able to prove this, the evidence for the conjecture is compelling, especially in the case of complex multiplication and rank at most 1 [Bro, CW, GKZ, GZ, Kol1, Kol2, Ru]. In addition there is much suggestive numerical evidence for the conjecture; for example, for elliptic curves with modest geometric rank r, numerical approximations of the first $r-1$ Taylor coefficients are consistent with these coefficients vanishing (see for instance the families studied in [Fe1, Fe2, Mil3]).

If E has complex multiplication[1] then $a_E(p) = 0$ for half the primes; i.e., $\theta_E(p) = \pi/2$. The remaining angles $\theta_E(p)$ are uniformly distributed in $[0,\pi]$ (this follows from [Deu, He1, He2]).

If E does not have complex multiplication, which is the case for most elliptic curves, then Sato and Tate [Ta] conjectured that as we vary p, the distribution of the $\theta_E(p)$'s converges to $2\sin^2\theta d\theta/\pi$. More precisely, for any interval $I \subset [0,\pi]$ we have

$$\lim_{x\to\infty} \frac{\#\{p : p \le x : \theta_E(p) \in I\}}{\#\{p : p \le x\}} = \int_I \frac{2\sin^2\theta d\theta}{\pi}; \tag{1.3}$$

we call $2\sin^2\theta d\theta/\pi$ the Sato-Tate measure, and denote it by $\mu_{\rm ST}$. By recent results of Clozel, Harris, Shepherd-Barron and Taylor [CHT, HS-BT, Tay], this is now known for all such E that have multiplicative reduction at some prime; see also [BZ] for results on the error terms when $|I|$ is small (these results are not for an individual curve, but rather averaged over the family of all elliptic curves) and [B-LGG, B-LGHT] for generalizations to other families of L-functions.

Instead of fixing an elliptic curve and letting the prime vary, we can instead fix a prime p and study the distribution of $\theta_E(p)$ as we vary E. Before describing our results, we briefly summarize related results in the literature concerning Sato-Tate behavior in

[1] This means the endomorphism ring is larger than the integers. For example, $y^2 = x^3 - x$ has complex multiplication, as can be seen by sending $(x,y) \to (-x, iy)$. Note $a_E(p) = 0$ if $p \equiv 3 \mod 4$ (this can be seen from the definition of $a_E(p)$ as a sum of Legendre symbols, sending $x \to -x$).

families. Serre [Ser] considered a similar question, not for elliptic curves, but rather for $S(N,k)$, the space of cusp forms of weight k on $\Gamma_0(N)$. He proved that for even k with $N+k\to\infty$ the eigenvalues of the normalized p^{th} Hecke operators are equidistributed in $[-2,2]$ with respect to the measure

$$\mu_p \;=\; \frac{p+1}{\pi}\frac{\sqrt{1-x^2/4}\,dx}{(p^{1/2}+p^{-1/2})^2-x^2}; \tag{1.4}$$

changing variables by setting $x=2\cos\theta$ this is equivalent to the measure $\widetilde{\mu}_p$ on $[0,\pi]$ given by

$$\widetilde{\mu}_p \;=\; \frac{2(p+1)}{\pi}\frac{\sin^2\theta d\theta}{(p^{1/2}+p^{-1/2})^2-4\cos^2\theta}. \tag{1.5}$$

Note that as $p\to\infty$, $\widetilde{\mu}_p\to\mu_{\rm ST}$; for p large these two measures assign almost the same probability to an interval I, differing by $O(1/p)$. See [CDF, Sar] for other families with a similar distribution.

Serre's theorem was ineffective, and has recently been improved by M. R. Murty and K. Sinha [MS]. They show that if $\{a_n(p)/p^{(k-1)/2}\}_{1\le i\le \#S(N,k)}$ denote the normalized eigenvalues of the Hecke operator T_p on $S(N,k)$, then

$$\frac{\#\{1\le n\le N: a_n(p)/p^{(k-1)/2}\in I\}}{\#S(N,k)} \;=\; \int_I \mu_p + O\left(\frac{\log p}{\log kN}\right), \tag{1.6}$$

where $\#S(N,k)$ is the number of cusp forms of weight k and level N, and if $N\ge 61$ then by Corollary 15 of [MS] we have

$$\frac{3}{200}\frac{(N)}{} \;\le\; \#S(N,k) \;\le\; \frac{(N)}{12}+1, \tag{1.7}$$

where $(N) \;=\; N\prod_{p|N}\left(1+\frac{1}{p}\right)$. This effective version of equidistribution allows Murty and Sinha to derive many results, such as

- an effectively computable constant B_d such that if $J_0(N)$ (the Jacobian of the modular curve $X_0(N)$) is isogenous to a product of $\mathbb{Q}$-simple abelian varieties of dimensions at most d, then $N\le B_d$;
- the multiplicity of any given eigenvalue of the Hecke operators is $\ll \frac{s(N,k)\log p}{\log kN}$.

The purpose of this paper is to expand the techniques in [MS] to families of elliptic curves. Unlike [MS, Ser], we cannot keep the prime fixed throughout the argument, as there are only finitely many distinct reductions of elliptic curves modulo p. Instead we fix a prime and study the angles $\theta_E(p)$ for one of the two families below, and then send $p\to\infty$. We study

(1) The family of all elliptic curves modulo p for $p\ge 5$. We may write these curves in Weierstrass form as $y^2=x^3-ax-b$ with $a,b\in\mathbb{Z}/p\mathbb{Z}$ and $4a^3\ne 27b^2$. The

number of pairs (a, b) satisfying these conditions[2] is $p(p-1)$.

(2) One-parameter families over $\mathbb{Q}(T)$: let $A(T), B(T) \in \mathbb{Z}[T]$ and consider the family $y^2 = x^3 + A(T)x + B(T)$ with non-constant $j(T)$.[3] We specialize T to be a $t \in \mathbb{Z}/p\mathbb{Z}$. The cardinality of the family is $p + O_{A,B}(1)$ (we lose a few values when we specialize as we require the reduced curves to be elliptic curves modulo p), where the error is a function of the discriminant of the family.

Notations:

- We let $\mathcal{F}_p$ denote either family, and write V_p for its cardinality (which is $p(p-1)$ in the first case and $p + O(1)$ in the second).
- While we may denote the angles by $\theta_E(p)$, $\theta_{a,b}(p)$ or $\theta_t(p)$, as p is fixed for notational convenience and to unify the presentation we shall denote these by θ_n, with $1 \le n \le V_p$.
- We let $e(x) = e^{2\pi i x}$.

Normalizations:

- For the family of all elliptic curves, we may match the elliptic curves in pairs (E, E') such that $\theta_{E'}(p) = \pi - \theta_E(p)$ (and each curve is in exactly one pair); see Remark 1.1 for a proof. Thus, if we let $x_n = \theta_n(p)/\pi$, we see that the set $\{2x_n\}_{n \le V_p}$ is symmetric about π. This will be very important later, as it means $\sum_{n \le V_p} \sin(2\pi m x_n) = 0$ for any integer m.
- For a one-parameter family of elliptic curves, in general we cannot match the elliptic curves in pairs, and thus the set $\{2\theta_t(p)\}$ is not typically symmetric about π; see Remark 1.2 for some results about biases in the $\theta_t(p)$'s. This leads to some complications in proving equidistribution, as certain sine terms no longer vanish. To overcome this, following other researchers we consider the technically easier situation where for each elliptic curve we include both $\theta_t(p)$ and $2\pi - \theta_t(p)$. To unify the presentation, instead of normalizing these angles by dividing by 2π (to obtain a distribution supported on $[0, 1]$), we first study the angles modulo π and then divide by π. We thus consider the normalized angles $x_t = \theta_t(p)/\pi$ and $x_{t+V_p} = 1 - \theta_t(p)/\pi$ for $1 \le t \le V_p$. Thus we study $2V_p$

[2] If $a = 0$ then the only b which is eliminated is $b = 0$. If a is a non-zero perfect square there are two b that fail, while if a is not a square than no b fail. Thus the number of bad pairs of (a, b) is p.

[3] Up to constants, $j(T)$ is $A(T)^3/(4A(T)^3 + 27B(T)^2)$.

normalized angles in $[0,1]$, unlike the case of all elliptic curves where we had V_p angles.

- We set $\widetilde{V}_p = V_p$ for the family of all elliptic curves, and $2V_p$ for a one-parameter family of elliptic curves. We study the distribution of the normalized angles $\{x_n\}_{1\le n\le \widetilde{V}_p}$.

Remark 1.1. *To see that we may match the angles as claimed for the family of all elliptic curves, consider the elliptic curve $y^2 = x^3 - ax - b$ with $4a^3 \neq 27b^2$. Let c be any non-residue modulo p, and consider the curve $y^2 = x^3 - ac^2x - bc^3$. Using the Legendre sum expressions for $a_E(p)$ and $a_{E'}(p)$, using the automorphism $x \to cx$ we see the second equals $\left(\frac{c}{p}\right)$ times the first; as we have chosen c to be a non-residue, this means $2\sqrt{p}\cos(\theta_{E'}(p)) = -2\sqrt{p}\cos(\theta_E(p))$, or $\theta_{E'}(p) = \pi - \theta_E(p)$ as claimed.*

Remark 1.2. *If the one-parameter family of elliptic curves has rank r over $\mathbb{Q}(T)$ and satisfies Tate's conjecture (see* [Ta, RS]*), then Rosen and Silverman* [RS] *prove a conjecture of Nagao* [Na]*, which states*

$$\lim_{X\to\infty} -\frac{1}{X}\sum_{p\le X}\frac{A_1(p)\log p}{p} \;=\; r \tag{1.8}$$

where $A_1(p) := \sum_{t \bmod p} a_t(p)$. Tate's conjecture is known for rational surfaces.[4] *This bias has been used by S. Arms, Á. Lozano-Robledo and S. J. Miller* [AL-RM] *to construct one-parameter families with moderate rank by finding families where $A(p)$ is essentially $-rp$. As there are about p curves modulo p, this represents a bias of about $-r$ on average per curve; as each $a_t(p)$ is of order $\sqrt{p}$, we see in the limit that this bias should be quite small per curve (though significant enough to lead to rank, it gives a lower order contribution to the distribution for each prime, and will be dwarfed by our other errors).*

Our goal is to prove effective theorems on the rate of convergence as $p \to \infty$ to the Sato-Tate measure, which requires us to obtain effective estimates for

$$\left|\#\{n \le \widetilde{V}_p : \theta_n \in I\} - \mu_{\rm ST}(I)\widetilde{V}_p\right|. \tag{1.9}$$

Here $\mu_{\rm ST}$ is the Sato-Tate measure on $[0,\pi]$ given by

$$\mu_{\rm ST}(T) \;=\; \int_I \frac{2}{\pi}\sin^2 t dt \quad I \subset [0,\pi], \tag{1.10}$$

[4] An elliptic surface $y^2 = x^3 + A(T)x + B(T)$ is rational if and only if one of the following is true: (1) $0 < \max\{3\deg A, 2\deg B\} < 12$; (2) $3\deg A = 2\deg B = 12$ and $\mathrm{ord}_{t=0}t^{12}\Delta(t^{-1}) = 0$.

and for $n \le V_p$, $2\sqrt{p}\cos(\theta_n)$ is the number of solutions modulo p of the elliptic curve $E_n : y^2 = x^3 + a_n x + b_n$. Equivalently, using the normalization $x_n = \theta_n/\pi$ to obtain a distribution on $[0,1]$, the Sato-Tate measure become

$$\mu_{\rm st}(I) \ = \ \int_I 2\sin^2(\pi x)dx, \quad I \subset [0,1]. \tag{1.11}$$

For a sequence of numbers x_n modulo 1, a measure μ and an interval $I \subset [0,1]$, let

$$\begin{aligned} N_I(\widetilde{V}_p) &= \#\{n \le \widetilde{V}_p : x_n \in I\} \\ \mu(I) &= \int_I \mu(t)dt. \end{aligned} \tag{1.12}$$

The discrepancy $D_{I,\widetilde{V}_p}(\mu)$ is

$$D_{I,\widetilde{V}_p}(\mu) \ = \ \left|N_I(\widetilde{V}_p) - \widetilde{V}_p\mu(I)\right|, \tag{1.13}$$

and trivially $D_{I,\widetilde{V}_p}(\mu) \le \widetilde{V}_p$. The goal is to obtain the best possible estimate for how rapidly $D_{I,\widetilde{V}_p}(\mu)/\widetilde{V}_p$ tends to 0.

Previous work has obtained a power savings in convergence to Sato-Tate for two-parameter families of elliptic curves (such as the entire family of all elliptic curves, or parametrizations such as $y^2 = x^3 + f(a)x + g(b)$ with a and b varying in appropriate ranges); see the papers by Banks and Shparlinski [BS, Sh1, Sh2] for saving $\widetilde{V}_p^{1/4}$ in Sato-Tate convergence. The key step in these arguments is

$$\frac{1}{(p-1)^2} \sum_{\substack{a,b \bmod p \\ 4a^3+27b^2 \not\equiv 0 \bmod p}} \frac{\sin((k+1)\theta_{a,b}(p))}{\sin(\theta_{a,b}(p)} \ \ll \ kp^{-1/2}, \quad k \ = \ 1,2,\ldots; \tag{1.14}$$

see Theorem 13.5.3 from [Ka] for a proof. One can obtain new and similar results for one-parameter families of elliptic curves by appealing to a result of Michel [Mic], which we do in §4. Our main results are the following.

Theorem 1.3 (Family of all curves)**.** *For the family of all elliptic curves modulo p, as $p \to \infty$ we have*

$$D_{I,\widetilde{V}_p}(\mu_{\rm st}) \ \le \ C\frac{\widetilde{V}_p}{\log \widetilde{V}_p} \tag{1.15}$$

for some computable C. Note that in this family, $\widetilde{V}_p = V_p$ and for each curve we include one normalized angle, $x_n = \theta_n/\pi \in [0,1]$.

Theorem 1.4 (One-parameter family of elliptic curves)**.** *For a one-parameter family of elliptic curves over $\mathbb{Q}(T)$ with non-constant j-invariant, we have*

$$D_{I,\widetilde{V}_p}(\mu_{\rm st}) \ \le \ C\widetilde{V}_p^{3/4} \tag{1.16}$$

for some computable C. Note that in this family, $\widetilde{V}_p = 2V_p$ and for each curve we include two normalized angles, $x_n = \theta_n/\pi$ and $x_{n+V_p} = 1 - \theta_n/\pi$, with $\theta_n \in [0,\pi]$.

Stronger results than Theorem 1.3 are known; as remarked above, convergence to Sato-Tate with an error of size $\widetilde{V}_p^{3/4}$ instead of $\widetilde{V}_p/\log\widetilde{V}_p$ is obtained in [BS, Sh1, Sh2]. We present these weaker arguments to highlight how one may attack these problems possessing only knowledge of the moments, and not the functions of the angles, in the hope that these arguments might be of use to other researchers attacking similar questions where we only have formulas for the moments of the coefficients. We will thus illustrate the effectiveness (in both senses of the word) of the techniques in [MS], as well as illustrate the loss of information that comes from having to trivially bound certain combinatorial sums. As we have not found similar effective results in the literature for one-parameter families, in order to get the best possible results we do not use formulas for the moments but rather estimates for the analogue of (1.14). It is worth remarking that we can recover the results of [BS, Sh1, Sh2] by our generalization of [MS] provided we also use (1.14) (see [Ka]) instead of results from Birch [Bi] on moments; this shows the value of the formulation in [MS].

We summarize the key ingredients of the proofs, and discuss why the second result has a much better error term than the first. Similar to [MS], both theorems follow from an analysis of $\sum_{n\le\widetilde{V}_p} e(mx_n)$ (we use $x_n = \theta_n/\pi$ in order to have a distribution supported on $[0,1]$). For the family of all elliptic curves, after some algebra we see this is equivalent to understanding $\sum_{n\le\widetilde{V}_p}\cos(2m\theta_n)$; using a combinatorial identity (see [Mil4]) this is equivalent to a linear combination of sums of the form $\sum_{n\le\widetilde{V}_p}(\cos\theta_n)^{2r}$. These sums are essentially the $2r^{\text{th}}$ moments of the Fourier coefficients of the family of all elliptic curves modulo p. Birch [Bi] evaluated these, and showed the answers are the Catalan numbers[5] plus lower order terms. Our equidistribution result then follows from a combinatorial identity of a sum of weighted Catalan numbers; our error term is poor due to the necessity of losing cancelation in bounding the contribution from the sums of the error terms.

The proof of Theorem 1.4 is easier, as now instead of inputting results on the moments we instead use a result of Michel [Mic] for the sum over the family of $\mathrm{sym}_k(\theta_n) = \sin((k+1)\theta_n)/\sin\theta_n$. This is easily related to our quantity of interest, $\cos(2m\theta_n)$, through identities of Chebyshev polynomials:

$$\cos(2m\theta_n) \;=\; \frac{1}{2}\mathrm{sym}_{2m}(\theta_n) - \frac{1}{2}\mathrm{sym}_{2m-2}(\theta_n). \tag{1.17}$$

The advantage of having a formula for the quantity we want and not a related quantity is that we avoid trivially estimating the errors in the combinatorial sums. These calculations increased the size of the error significantly, and this is why Theorem 1.4 is stronger than Theorem 1.3, though the error term in Theorem 1.3 is comparable to the

[5]The Catalan numbers are the moments of the semi-circle distribution, which is related to the Sato-Tate distribution through a simple change of variables.

error terms of the equivalent quantities in [MS] for the family of cuspidal newforms. Michel proves his result by using a cohomological interpretation, and this results in the error term being $p^{-1/2}$ smaller than the main term; it is this savings in the quantity we are directly interested in that leads to the superior error estimates.

The paper is organized as follows. After reviewing the needed results from Murty-Sinha [MS] in §2, we prove Theorem 1.3 in §3 and Theorem 1.4 in §4. For completeness the needed combinatorial identities are proved in Appendix A, and in Appendix B we correct some errors in explicit formulas for moments in Birch's paper [Bi] (where he neglected to mention that his sums are normalized by dividing by $p-1$).

2. Effective Equidistribution Preliminaries

We quickly review some needed results from Murty-Sinha [MS]; while our setting is similar to the problems they investigated, there are slight differences which require generalizations of some of their results. Assume $\mu = F(-x)dx$ with

$$F(x) \ = \ \sum_{m=-\infty}^{\infty} c_m e(mx) \tag{2.1}$$

where $e(z) = \exp(2\pi i z)$. Theorem 8 from [MS] is

Theorem 2.1. *Let $\{x_n\}$ be a sequence of real numbers in $[0,1]$ and let the notation be as above. Assume for each m that*

$$\lim_{\widetilde{V}_p \to \infty} \frac{1}{\widetilde{V}_p} \sum_{n \le \widetilde{V}_p} e(mx_n) \ = \ c_m \quad \text{and} \quad \sum_{m=-\infty}^{\infty} |c_m| \ < \ \infty. \tag{2.2}$$

Let $||\mu|| = \sup_{x\in[0,1]} |F(x)|$ with $\mu = F(-x)dx$. Then the discrepancy satisfies

$$\begin{aligned} D_{I,\widetilde{V}_p}(\mu) \ &\le \ \frac{\widetilde{V}_p||\mu||}{M+1} \\ &+ \sum_{1\le m\le M} \left(\frac{1}{M+1} + \min\left(b-a, \frac{1}{\pi|m|}\right)\right) \left|\sum_{n=1}^{\widetilde{V}_p} e(mx_n) - \widetilde{V}_p c_m\right| \end{aligned} \tag{2.3}$$

for any natural numbers $\widetilde{V}_p$ and M.

Unfortunately, Theorem 2.1 is not directly applicable in our case. The reason is that there we have a limit as $\widetilde{V}_p \to \infty$ in the definition of the c_m, where for us we fix a prime p and have $\widetilde{V}_p = p(p-1)$ for the family of all elliptic curves curves modulo p, or $p+O(1)$ for a one-parameter family. Analyzing the proof of Theorem 8 from [MS], however, we see that the claim holds for *any* sequence c_m (obviously if $\widetilde{V}_p^{-1}\sum_{n\le\widetilde{V}_p} e(mx_n)$ is not close to c_m then the discrepancy is large). We thus obtain

Theorem 2.2. *Let $\{x_n\}$ be a sequence of real numbers in $[0,1]$ and let the notation be as above. Let $\{c_m\}$ be a sequence of numbers such that $\sum_{m=-\infty}^{\infty}|c_m| < \infty$ (we will take $c_0 = 1$, $c_{\pm 1} = -1/2$ and all other c_m's equal to zero). Let $||\mu|| = \sup_{x\in[0,1]}|F(x)|$ with $\mu = F(-x)dx$. Then the discrepancy satisfies*

$$D_{I,\widetilde{V}_p}(\mu) \;\leq\; \frac{\widetilde{V}_p||\mu||}{M+1} + \sum_{1\leq m\leq M}\left(\frac{1}{M+1} + \min\left(b-a, \frac{1}{\pi|m|}\right)\right)\left|\sum_{n=1}^{\widetilde{V}_p} e(mx_n) - \widetilde{V}_p c_m\right| \tag{2.4}$$

for any natural numbers $\widetilde{V}_p$ and M.

To simplify applying the results from [MS], we study the normalized angles x_n. Under our normalization, the Sato-Tate measure becomes

$$\mu_{\rm st}(I) \;=\; \int_I 2\sin^2(\pi x)dx, \quad I\subset[0,1]. \tag{2.5}$$

The Fourier coefficients of $\mu_{\rm st}$ are readily calculated.

Lemma 2.3. *Let $\mu_{\rm st} = F(-x)dx$ be the normalized Sato-Tate distribution on $[0,1]$ with density $2\sin^2(\pi x)$. We have*

$$F(x) \;=\; 1 - \frac{1}{2}\left(e(x) + e(-x)\right), \tag{2.6}$$

which implies that the Fourier coefficients are $c_0 = 1$, $c_{\pm 1} = -1/2$ and $c_m = 0$ for $|m|\geq 2$.

Proof. The proof is immediate from the expansion of F as a sum of exponentials, which follows from the identities $\cos(2\theta) = 1 - 2\sin^2(\theta)$ and $e(\theta) = \cos(2\pi\theta) + i\sin(2\pi\theta)$. □

3. Proof of Effective Equidistribution for All Curves

We use Birch's [Bi] results on the moments of the family of all elliptic curves modulo p (there are some typos in his explicit formulas; we correct these in Appendix B); unfortunately, these are results for quantities such as $(2\sqrt{p}\cos\theta_n)^{2R}$, and the quantity which naturally arises in our investigation is $e(mx_n)$ (with x_n running over the normalized angles $\theta_{a,b}(p)/\pi$), specifically

$$\left|\sum_{n=1}^{\widetilde{V}_p} e(mx_n) - \widetilde{V}_p c_m\right|. \tag{3.1}$$

By applying some combinatorial identities we are able to rewrite our sum in terms of the moments, which allows us to use Birch's results. The point of this section is not to

obtain the best possible error term (which following [BS, Sh1, Sh2] could be obtained by replacing Birch's bounds with (1.14)) but rather to highlight how one may generalize and apply the framework from [MS].

We first set some notation. Let $\sigma_k(T_p)$ denote the trace of the Hecke operator T_p acting on the space of cusp forms of dimension $-2k$ on the full modular group. We have $\sigma_{k+1}(T_p) = O(p^{k+c+\epsilon})$, where from [Sel] we see we may take $c = 3/4$ (there is no need to use the optimal c, as our final result, namely (3.17), will yield the same order of magnitude result for $c = 3/4$ or $c = 0$). Let $\mathcal{M}_p(2R)$ denote the $2R^{\text{th}}$ moment of $2\cos(\theta_n) = 2\cos(\pi x_n)$ (as we are concerned with the normalized values, we use slightly different notation than in [Bi]):

$$\mathcal{M}_p(2R) \;=\; \frac{1}{\widetilde{V}_p}\sum_{n=1}^{\widetilde{V}_p}\left(2\cos(\pi x_n)\right)^{2R}. \tag{3.2}$$

Lemma 3.1 (Birch)**.** *Notation as above, we have*

$$\mathcal{M}_p(2R) \;=\; \frac{1}{R+1}\binom{2R}{R} + O\left(2^{2R}\widetilde{V}_p^{-\frac{1-c-\epsilon}{2}}\right); \tag{3.3}$$

we may take $c = 3/4$ and thus there is a power saving.[6]

Proof. The result follows from dividing the equation for $S_R^*(p)$ on the bottom of page 59 of [Bi] by p^R, as we are looking at the moments of the normalized Fourier coefficients of the elliptic curves, and then using the bound $\sigma_{k+1}(T_p) = O(p^{k+c+\epsilon})$, with $c = 3/4$ admissible by [Sel]. Recall $\widetilde{V}_p = p(p-1)$ is the cardinality of the family. We have

$$\begin{aligned}\mathcal{M}_p(2R) &= \frac{1}{R+1}\binom{2R}{R}\frac{p(p-1)}{\widetilde{V}_p} \\ &\quad + O\left(\sum_{k=1}^{R}\frac{2k+1}{R+k+1}\binom{2R}{R+k}\frac{p^{1+c+\epsilon}}{\widetilde{V}_p} + \frac{p}{p^R\widetilde{V}_p}\right) \\ &= \frac{1}{R+1}\binom{2R}{R} + O\left(2^{2R}\widetilde{V}_p^{-\frac{1-c-\epsilon}{2}}\right)\end{aligned} \tag{3.4}$$

since $\widetilde{V}_p = p(p-1)$. □

A simple argument (see Remark 1.1) shows that the normalized angles are symmetric about $1/2$. This implies

$$\sum_{n=1}^{\widetilde{V}_p} e(mx_n) \;=\; \sum_{n=1}^{\widetilde{V}_p}\cos(2\pi m x_n) + i\sum_{n=1}^{\widetilde{V}_p}\sin(2\pi m x_n) \;=\; \sum_{n=1}^{\widetilde{V}_p}\cos(2m\theta_n), \tag{3.5}$$

[6]Note $\frac{1}{R+1}\binom{2R}{R}$ is the R^{th} Catalan number. The Catalan numbers are the moments of the semi-circle distribution, which is related to the Sato-Tate distribution by a simple change of variables.

where the sine piece does not contribute as the angles are symmetric about $1/2$, and we are denoting the $\widetilde{V}_p$ non-normalized angles by θ_n.

Thus it suffices to show we have a power saving in

$$\left|\sum_{n=1}^{\widetilde{V}_p}\cos(2m\theta_n) - \widetilde{V}_p c_m\right|. \tag{3.6}$$

By symmetry, it suffices to consider $m \geq 0$.

Lemma 3.2. *Let $c_0 = 1$, $c_{\pm 1} = -1/2$ and $c_m = 0$ otherwise. There is some $c < 1$ such that*

$$\left|\sum_{n=1}^{\widetilde{V}_p}\cos(2m\theta_n) - \widetilde{V}_p c_m\right| \ll \left(m^2 2^{3m}\widetilde{V}_p^{-\frac{1-c-\epsilon}{2}}\right); \tag{3.7}$$

by the work of Selberg [Sel] *we may take $c = 3/4$.*

Proof. The case $m = 0$ is trivial. For $m = 1$ we use the trigonometric identity $\cos(2\theta_n) = 2\cos^2(\theta_n) - 1$. As $c_{\pm 1} = -1/2$ we have

$$\begin{aligned}\sum_{n=1}^{\widetilde{V}_p}\cos(2\theta_n) - \frac{\widetilde{V}_p}{2} &= \sum_{n=1}^{\widetilde{V}_p}\left[\left(2\cos^2\theta_n - 1\right) + \frac{1}{2}\right] \\ &= \frac{1}{2}\sum_{n=1}^{\widetilde{V}_p}\left((2\cos\theta_n)^2 - 1\right) \\ &= \frac{1}{2}\sum_{n=1}^{\widetilde{V}_p}\left(\frac{(2\sqrt{p}\cos\theta_n)^2}{p} - 1\right). \end{aligned} \tag{3.8}$$

Note the sum of $(2\sqrt{p}\cos\theta_n)^2$ is the second moment of the number of solutions modulo p. From [Bi] we have that this is $p + O(1)$; the explicit formula given in [Bi] for the second moment is wrong; see Appendix B for the correct statement. Substituting yields

$$\left|\sum_{n=1}^{\widetilde{V}_p}\cos(2\theta_n) - \frac{\widetilde{V}_p}{2}\right| \ll O(1). \tag{3.9}$$

The proof is completed by showing that $\sum_{n=1}^{\widetilde{V}_p}\cos(2m\theta_n) = O_m(\widetilde{V}_p^{1/2})$ provided $2 \leq m \leq M$. In order to obtain the best possible results, it is important to understand the implied constants, as M will have to grow with $\widetilde{V}_p$ (which is of size p^2). While it is possible to analyze this sum for any m by brute force, we must have M growing with p, and thus we need an argument that works in general. As $c_{\pm 1} \neq 0$ but $c_m = 0$ for $|m| \geq 2$, we expect (and we will see) that the argument below does break down when $|m| = 1$.

There are many possible combinatorial identities we can use to express $\cos(2m\theta_n)$ in terms of powers of $\cos(\theta_n)$. We use the following (for a proof, see Definition 2 and equation (3.1) of [Mil4]):

$$2\cos(2m\theta_n) \;=\; \sum_{r=0}^{m} c_{2m,2r}(2\cos\theta_n)^{2r}, \tag{3.10}$$

where $c_{2r} = (2r)!/2$, $c_{0,0} = 0$, $c_{2m,0} = (-1)^m 2$ for $m \geq 1$, and for $1 \leq r \leq m$ set

$$c_{2m,2r} \;=\; \frac{(-1)^{r+m}}{c_{2r}} \prod_{j=0}^{r-1}(m^2 - j^2) \;=\; \frac{(-1)^{m+r}}{c_{2r}} \frac{m\cdot(m+r-1)!}{(m-r)!}. \tag{3.11}$$

We now sum (3.10) over n and divide by $\widetilde{V}_p$, the cardinality of the family. In the argument below, at one point we replace 2^{2r} in an error term with $2012\frac{1}{r+1}\binom{2r}{r}\cdot m^2$; this allows us to pull the r^{th} Catalan number, $\frac{1}{r+1}\binom{2r}{r}$, out of the error term.[7] Using Lemma 3.1 we find

$$\begin{aligned}
\frac{1}{\widetilde{V}_p}\sum_{n=1}^{\widetilde{V}_p} 2\cos(2m\theta_n) \;&=\; \sum_{r=0}^{m} c_{2m,2r}\frac{1}{\widetilde{V}_p}\sum_{n=1}^{\widetilde{V}_p}(2\cos\theta_n)^{2r} \\
&= \sum_{r=0}^{m}\left(\frac{1}{r+1}\binom{2r}{r} + O\left(2^{2r}\widetilde{V}_p^{-\frac{1-c-\epsilon}{2}}\right)\right) c_{2m,2r} \\
&= \sum_{r=0}^{m}\left(\frac{1}{r+1}\frac{(2r)!}{r!r!}\frac{(-1)^{m+r}2}{(2r)!}\frac{m\cdot(m+r)!}{(m-r)!\cdot(m+r)}\right) \\
&\quad\cdot\left(1 + O\left(m^2\widetilde{V}_p^{-\frac{1-c-\epsilon}{2}}\right)\right) \\
&= (-1)^m 2m\sum_{r=0}^{m}\left((-1)^r\frac{m!}{r!(m-r)!}\frac{(m+r)!}{m!r!}\frac{1}{(r+1)(m+r)}\right) \\
&\quad\cdot\left(1 + O\left(m^2\widetilde{V}_p^{-\frac{1-c-\epsilon}{2}}\right)\right) \\
&= (-1)^m 2m\sum_{r=0}^{m}\left((-1)^r\binom{m}{r}\binom{m+r}{r}\frac{1}{(r+1)(m+r)}\right) \\
&\quad\cdot\left(1 + O\left(m^2\widetilde{V}_p^{-\frac{1-c-\epsilon}{2}}\right)\right).
\end{aligned} \tag{3.12}$$

We first bound the error term. For our range of r, $\binom{m+r}{r} \leq \binom{2m}{m} \leq 2^{2m}$. The sum of $\binom{m}{r}$ over r is 2^m, and we get to divide by at least $m + r \geq m$. Thus the error term is

[7] The reason this is valid is that the largest binomial coefficient is the middle (or the middle two when the upper argument is odd). Thus $2^{2r} = (1+1)^{2r} \leq (2r+1)\binom{2r}{r} \leq 2(m+1)\binom{2r}{r}$ (as $m \leq r$), and the claim follows from $\frac{2012m^2}{r+1} \geq 2(m+1)$ for $m \geq 2$ and $0 \leq r \leq m$.

bounded by

$$O\left(m^2 2^{3m} \widetilde{V}_p^{-\frac{1-c-\epsilon}{2}}\right). \tag{3.13}$$

We now turn to the main term. It it just $(-1)^m 2m$ times the sum in Lemma A.3, which is shown in that lemma to equal 0 for any $|m| \geq 2$. □

Remark 3.3. *Without Lemma A.3, our combinatorial expansion would be useless. We thus give several proofs in the appendix (including a brute force, hypergeometric and an application of Zeilberger's Fast Algorithm).*

Remark 3.4. *It is possible to get a better estimate for the error term by a more detailed analysis of $\sum_{r\leq m}\binom{m}{r}\binom{m+r}{r}$; however, the improved estimates only change the constants in the discrepancy estimates, and not the savings. This is because this sum is at least as large as the term when $r \approx m/2$, and this term contributes something of the order $3^{3m/2}/m$ by Stirling's formula. We will see that any error term of size 3^{am} for a fixed a gives roughly the same value for the best cutoff choice for M, differing only by constants. Thus we do not bother giving a more detailed analysis to optimize the error here.*

We now prove the first of our two main theorems.

Proof of Theorem 1.3. We must determine the optimal M to use in (2.4):

$$\begin{aligned} D_{I,\widetilde{V}_p}(\mu_{\rm st}) &\ll \frac{\widetilde{V}_p}{M+1} + \sum_{1\leq m\leq M}\left(\frac{1}{M+1}+\frac{1}{m}\right)\left(m^2 2^{3m}\widetilde{V}_p^{-\frac{1-c-\epsilon}{2}}\right) \\ &\ll \frac{\widetilde{V}_p}{M} + M 2^{3M}\widetilde{V}_p^{-\frac{1-c-\epsilon}{2}} \end{aligned} \tag{3.14}$$

as $\frac{1}{M+1} \ll \frac{1}{m}$ and $\sum_{m\leq m} 2^{3m} \ll 2^{3M}$. For all $c > 0$ we find the minimum error by setting the two terms equal to each other, which yields

$$\widetilde{V}_p^{\frac{3-c-\epsilon}{2}} \;=\; M^2 2^{3M} \;\ll\; e^{3M}, \tag{3.15}$$

which when equating yields[8]

$$e^{3M} \;\approx\; e^{\frac{3-c-\epsilon}{2}\log \widetilde{V}_p}, \tag{3.16}$$

which implies

$$M \;\approx\; \frac{3-c-\epsilon}{6}\log \widetilde{V}_p. \tag{3.17}$$

We thus see that we may find a constant C such that

$$D_{I,\widetilde{V}_p}(\mu_{\rm st}) \;\leq\; C\frac{\widetilde{V}_p}{\log \widetilde{V}_p}. \tag{3.18}$$

[8]We could obtain a slightly better constant below with a little more work; however, as it will not affect the quality of our result we prefer to give the simpler argument with a slightly worse constant.

□

4. Proof of Effective Equidistribution for One-parameter families

Instead of studying the family of all elliptic curves, we can also investigate one-parameter families over $\mathbb{Q}(T)$. Thus, consider the family $\mathcal{E} : y^2 = x^3 + A(T)x + B(T)$, where $A(T)$ and $B(T)$ are in $\mathbb{Z}(T)$. We assume that $j(T)$ is not constant for the family. Michel [Mic] proved a Sato-Tate law for such families. In particular, he proved

Theorem 4.1 (Michel [Mic])**.** *Consider a one-parameter family of elliptic curves over $\mathbb{Q}(T)$ with non-constant j-invariant. Let c_Δ denote the number of complex zeros of $\Delta(z) = 0$ (where Δ is the discriminant), $\quad_p$ an additive character (and set $\delta_{\quad p} = 0$ if this character is trivial and 1 otherwise), and write $a_{t,p}$ as $2\sqrt{p}\cos\theta_{t,p}$ with $\theta_{t,p} \in [0,\pi]$. Let*

$$\mathrm{sym}_k(\theta) \;=\; \frac{\sin((k+1)\theta)}{\sin\theta}. \tag{4.1}$$

Then

$$\left|\frac{1}{p}\sum_{\substack{t \bmod p\\ \Delta(t)\neq 0}} \mathrm{sym}_k\theta_{t,p}\right| \;\le\; \frac{(k+1)(c_\Delta - \delta_{\quad p} - 1)}{\sqrt{p}}. \tag{4.2}$$

Additionally, we have

$$\left|\frac{1}{p}\sum_{\substack{t \bmod p\\ \Delta(t)\neq 0}} \cos\theta_{t,p}\right| \;\le\; \frac{C}{\sqrt{p}} \tag{4.3}$$

for some C depending on the family. Finally, we may drop the additive character and drop the restriction that $\Delta(t) \neq 0$ at the cost of a bounded number of summands, each of which is at most $(k+1)$,[9] which implies these relations still hold provided we multiply the bounds on the right hand side by some constant C'.

Remark 4.2. *Miller* [Mil2] *showed that the error term in Theorem 4.1 is sharp. Specifically, the second moment of the family $y^2 = x^3 + Tx^2 + 1$ of elliptic curves over $\mathbb{Q}(T)$ for $p > 2$ is*

$$A_2(p) \;:=\; \sum_{t \bmod p} a_t(p)^2 \;=\; p^2 - n_{3,2,p}p - 1 + p\sum_{x \bmod p}\left(\frac{4x^3+1}{p}\right), \tag{4.4}$$

where $n_{3,2,p}$ denotes the number of cube roots of 2 modulo p. For any $[a,b] \subset [-2,2]$ there are infinitely many primes $p \equiv 1 \bmod 3$ such that

$$A_2(p) - \left(p^2 - n_{3,2,p}p - 1\right) \;\in\; [a\cdot p^{3/2}, b\cdot p^{3/2}]. \tag{4.5}$$

[9] This is readily seen by writing $\sin((k+1)\theta) = \sin(\theta)\cos(k\theta) + \cos(\theta)\sin(k\theta)$ and proceeding by induction.

Theorem 4.1 is used by Michel to obtain good estimates for the average rank in these families, as well as (of course) proving Sato-Tate laws. Using our techniques above, we can convert Michel's bounds to a quantified equidistribution law.

We recall the notation for Theorem 1.4. Consider a one-parameter family of elliptic curves over $\mathbb{Q}(T)$ with non-constant $j(T)$. Let there be $V_p = p + O(1)$ reduced curves modulo p, and set $\widetilde{V}_p = 2V_p$. For each curve E_t consider the angles $\theta_{t,p}$ and $\pi - \theta_{t,p}$, with $\theta_{t,p} \in [0,1]$, and the normalized angles $x_n = \theta_{t,p}/\pi$ and $x_{n+V_p} = 1 - \theta_{t,p}/\pi$ (for $1 \le n \le V_p$).

Proof of Theorem 1.4. We must show $D_{I,\widetilde{V}_p}(\mu_{\rm st}) \ll \widetilde{V}_p^{3/4}$ (where $\widetilde{V}_p \approx 2p$). As in the proof of Theorem 1.3, it suffices to show

$$\left|\sum_{t \bmod p} \cos(2m\theta_{t,p}) - c_m p\right|, \tag{4.6}$$

with $c_0 = 1$, $c_1 = -1/2$ and all other $c_m = 0$. This is because we have enlarged our set of normalized angles to be symmetric about 1/2. Thus when we study $e(mx_n) = \cos(2\pi m x_n) + i\sin(2\pi m x_n)$, the sine sum vanishes. We are therefore left with the cosine sum, with the normalized angles x_n and x_{n+V_p} contributing equally. Thus we may replace the sum of the cosine piece over n with a sum over the angles $\theta_{t,p}$, so long as we remember to multiply by 2 when computing the discrepancy later. While we should subtract $c_m V_p$ and not $c_m p$, as $V_p = p + O(1)$ the error in doing this is dwarfed by the error of the piece we are studying.

The case of $2m = 0$ is trivial. If $2m = 2$, then we are studying $\cos 2\theta_{t,p} = -\frac{1}{2} + \frac{1}{2}\mathrm{sym}_2(\theta)$. By Theorem 4.1, we thus find that

$$\left|\sum_{t \bmod p} \cos(2\theta_{t,p}) + \frac{p}{2}\right| \;=\; \left|\sum_{t \bmod p} \frac{1}{2}\mathrm{sym}_2(\theta)\right| \;\le\; \frac{C}{\sqrt{p}}. \tag{4.7}$$

For higher m, we use Chebyshev polynomials (see [Wi]). The Chebyshev polynomials of the first kind are given by $T_\ell(\cos\theta) = \cos(\ell\theta)$; the Chebyshev polynomials of the second kind are $U_\ell(\cos\theta) = \mathrm{sym}_{\ell+1}(\theta)$. These polynomials are related by

$$T_\ell(\cos\theta) \;=\; \frac{U_\ell(\cos\theta) - U_{\ell-2}(\cos\theta)}{2} \;=\; \frac{\mathrm{sym}_\ell(\theta) - \mathrm{sym}_{\ell-2}(\theta)}{2}; \tag{4.8}$$

we use this with $\ell = 2m \ge 4$. Using Theorem 4.1 we see that for $m \ge 2$,

$$\left|\sum_{t \bmod p} \cos(2m\theta_{t,p})\right| \;\le\; Cm\sqrt{p}. \tag{4.9}$$

From (2.4), the discrepancy satisfies

$$\frac{1}{2}D_{I,\widetilde{V}_p}(\mu_{\rm st}) \le \frac{p||\mu||}{M+1} + \sum_{1\le m\le M}\left(\frac{1}{M+1}+\min\left(b-a,\frac{1}{\pi|m|}\right)\right)\left|\sum_{t=1}^{p}e(mx_n)-c_mp\right|. \tag{4.10}$$

Using our bounds, we have

$$D_{I,p}(\mu) \ll \frac{p||\mu||}{M+1}+\sum_{m=1}^{M}\frac{Cm\sqrt{p}}{m} \ll \frac{p}{M}+M\sqrt{p}. \tag{4.11}$$

The two error terms are of the same order of magnitude when $M^2=\sqrt{p}$, or $M=p^{1/4}$. This leads to

$$D_{I,p}(\mu) \ll p^{3/4}. \tag{4.12}$$

□

Remark 4.3. *Note we could have used the Chebyshev identities to handle the $m=1$ case as well, as in fact we implicitly did when we rewrote $\cos 2\theta$; we prefer to break the analysis into two cases as the $m=1$ case has $c_m\neq 0$.*

Remark 4.4. *Rosen and Silverman* [RS] *proved a conjecture of Nagao* [Na] *relating the distribution of the $a_E(p)$'s and the rank. Unfortunately the known lower order term due to the rank of the family is of size $p^{1/2}$, which is significantly smaller than the error terms of size $p^{3/4}$ analyzed above. As noted in Remark 4.2, the error term is sharp and cannot be improved for all families.*

APPENDIX A. COMBINATORIAL IDENTITIES

We first state some needed properties of the binomial coefficients. For n,r non-negative integers we set $\binom{n}{k}=\frac{n!}{k!(n-k)!}$. We generalize to real n and k a positive integer by setting

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-(k-1))}{k!}, \tag{A.1}$$

which clearly agrees with our original definition for n a positive integer. Finally, we set $\binom{n}{0}=1$ and $\binom{n}{k}=0$ if k is a negative integer.

To prove our main result we need the following two lemmas; we follow the proofs in [Ward].

Lemma A.1 (Vandermonde's Convolution Lemma)**.** *Let r,s be any two real numbers and k,m,n integers. Then*

$$\sum_k \binom{r}{m+k}\binom{s}{n-k} = \binom{r+s}{m+n}. \tag{A.2}$$

Proof. It suffices to prove the claim when r, s are integers. The reason is that both sides are polynomials, and if the polynomials agree for an infinitude of integers then they must be identical. It suffices to consider the special case $m = 0$, in which case we are reduced to showing

$$\binom{r}{k}\binom{s}{n-k} \;=\; \binom{r+s}{n}. \tag{A.3}$$

Consider the polynomial

$$(x+y)^r(x+y)^s \;=\; (x+y)^{r+s}. \tag{A.4}$$

If we use the binomial theorem to expand the left hand side of (A.4), we get the coefficient of the $x^n y^{r+s-n}$ is the left hand side of (A.3), while if we use the binomial theorem to find the coefficient of $x^n y^{r+s-n}$ on the right hand side of (A.4) we get (A.3), which completes the proof. □

Lemma A.2. *Let ℓ, m, s be non-negative integers. Then*

$$\sum_k (-1)^k \binom{\ell}{m+k}\binom{s+k}{n} \;=\; (-1)^{\ell+m}\binom{s-m}{n-\ell}. \tag{A.5}$$

Proof. Using $\binom{a}{b} = \binom{a}{a-b}$, we rewrite $\binom{s+k}{n}$ as $\binom{s+k}{s+k-n}$, and we then rewrite $\binom{s+k}{s+k-n}$ as $(-1)^{s+k-n}\binom{-n-1}{s+k-n}$ by using the extension of the binomial coefficient, where we have pulled out all the negative signs in the numerators. The advantage of this simplification is that the summation index is now only in the denominator; further, the power of -1 is now independent of k. Factoring out the sign, our quantity is equivalent to

$$\begin{aligned}&(-1)^{s-n}\sum_k \binom{\ell}{m+k}\binom{-n-1}{s+k-n}\\ &\;=\; (-1)^{s-n}\sum_k \binom{\ell}{\ell-m-k}\binom{-n-1}{s+k-n},\end{aligned} \tag{A.6}$$

where we again use $\binom{a}{b} = \binom{a}{a-b}$. By Vandermonde's Convolution, this equals $(-1)^{s-n}$ $\binom{\ell-n-1}{\ell-m-n+s}$. Using $\binom{s-m}{\ell-m-n+s} = \binom{s-m}{n-\ell}$ and collecting powers of -1 completes the proof (note $(-1)^{\ell-m} = (-1)^{\ell+m}$). □

Lemma A.3. *Let m be an integer greater than or equal to 1. Then*

$$\sum_{r=0}^{m}(-1)^r\binom{m}{r}\binom{m+r}{r}\frac{1}{(r+1)(m+r)} \;=\; \begin{cases}1/2 & \text{if } m=1\\ 0 & \text{if } m\geq 2.\end{cases} \tag{A.7}$$

Proof. The case $m = 1$ follows by direct evaluation. Consider now $m \geq 2$. We have

$$\begin{aligned} S_m &= \sum_{r=0}^{m}(-1)^r\binom{m}{r}\binom{m+r}{r}\frac{1}{(r+1)(m+r)} \\ &= \sum_{r=0}^{m}(-1)^r\binom{m}{r}\frac{m+1}{m+1}\binom{m+r}{r}\frac{1}{(r+1)(m+r)} \\ &= \sum_{r=0}^{m}(-1)^r\frac{m!(m+1)}{(r+1)\cdot r!m!}\frac{1}{m+1}\frac{(m+r)(m+r-1)!}{r!m\cdot(m-1+r)!}\frac{1}{m+r} \\ &= \sum_{r=0}^{m}(-1)^r\binom{m+1}{r+1}\binom{m-1+r}{r}\frac{1}{m(m+1)} \\ &= \frac{1}{m(m+1)}\sum_{r=0}^{m}(-1)^r\binom{m+1}{r+1}\binom{m-1+r}{m-1}. \end{aligned} \tag{A.8}$$

We change variables and set $u = r + 1$; as r runs from 0 to m, u runs from 1 to $m + 1$. To have a complete sum, we want u to start at 0; thus we add in the $u = 0$ term, which is $\binom{m-2}{m-1}$. As $m \geq 2$, this is 0 from the extension of the binomial coefficient (this is the first of two places where we use $m \geq 2$). Our sum S_m thus equals

$$S_m = -\frac{1}{m(m+1)}\sum_{u=0}^{m+1}(-1)^u\binom{m+1}{u}\binom{m-2+u}{m-1}. \tag{A.9}$$

We now use Lemma A.2 with $k = u$, $m = 0$, $\ell = m+1$, $s = m-2$ and $n = m-1$; note the conditions of that lemma require s to be a non-negative integer, which translates to our $m \geq 2$. We thus find

$$S_m = -\frac{1}{m(m+1)}(-1)^{m+1}\binom{m-2}{-2} = 0, \tag{A.10}$$

which completes the proof. □

We give another proof of Lemma A.3 below using hypergeometric functions; we thank Frederick Strauch for showing us this approach.

Remark A.4. *We present an alternative proof of Lemma A.3 using the hypergeometric function*

$${}_2F_1(a,b,c;z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c-b)}\int_0^1 \frac{t^{b-1}(1-t)^{c-b-1}dt}{(1-tz)^a}. \tag{A.11}$$

The following identity for the normalization constant of the Beta function is crucial in the expansions:

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}. \tag{A.12}$$

We can use the geometric series formula to expand (A.11) *as a power series in z involving Gamma factors. Rewriting $\binom{m}{r}$ as $(-1)^r\binom{r-m-1}{r}$, after some algebra we find*

$$S_m \;=\; \frac{\Gamma(m)\,{}_2F_1(-m,m,2;1)}{\Gamma(2)\Gamma(1+m)} \;=\; \frac{\Gamma(m)}{\Gamma(1+m)\Gamma(2+m)\Gamma(2-m)} \tag{A.13}$$

(our summation over r in the definition of S_m has become the series expansion of ${}_2F_1(-m,m,2;1)$), where the last step uses

$${}_2F_1(a,b,c;1) \;=\; \frac{\Gamma(c)\Gamma(c-a-b)}{\Gamma(c-a)\Gamma(c-b)} \tag{A.14}$$

which follows from the normalization constant of the Beta function. Note that the right hand side of (A.13) *is $1/2$ when $m=1$ and 0 for $m\ge 2$ because for such m, $1/\Gamma(2-m)=0$ due to the pole of $\Gamma(2-m)$.*

Remark A.5. *It is also possible to prove this lemma through symbolic manipulations. Using the results from* [PS, PSR]*, one may input this into a Mathematica package, which outputs a proof.*

APPENDIX B. MOMENTS FOR THE FAMILY OF ALL CURVES

Birch [Bi] claims the following: Let

$$S_R(p) \;=\; \sum_{a \bmod p}\;\sum_{b \bmod p}\left[\sum_{x \bmod p}\left(\frac{x^3-ax-b}{p}\right)\right]^{2R}. \tag{B.1}$$

Then for $p\ge 5$,

$$\begin{aligned} S_1(p) &= p^2 \\ S_2(p) &= 2p^3-3p \\ S_3(p) &= 5p^4-9p^2-5p. \end{aligned} \tag{B.2}$$

There are obviously typos here. We know the Legendre sum is at most $2\sqrt{p}$ in absolute value, thus we expect $S_R(p)$ to be on the order of $p^2\cdot(\sqrt{p})^{2R}=p^{R+2}$; note the powers of p are too low (and they are too high for dividing $S_R(p)$ by the cardinality of the family).

Assuming $S_R(p)$ is a polynomial in p, from exploring the results for small p we are led to

$$\begin{aligned} S_1(p) &= p^3-p^2 \\ S_2(p) &= 2p^4-2p^3-3p^2+3p \\ S_3(p) &= 5p^5-5p^4-9p^3+4p^2+5p. \end{aligned} \tag{B.3}$$

Note these are exactly the results from Birch multiplied by $p-1$; we thank Andrew Granville for pointing this out to us. In other words, the formulas in Birch are what remains after dividing by the trivial multiplicative factor $p-1$.

Let $S'_R(p)$ denote the same sum as $S_R(p)$, but with the additional restriction that $4a^3 \neq 27b^2$. It is readily seen that $S'_R(p) = S_R(p) + (p-1)$; the reason is that if the discriminant equals zero, then $x^3 - ax - b = (x-c)^2(x-d)$ for some c, d, and the sum of these Legendre symbols over all x modulo p is ± 1 (the sum is the same as $\sum_{x \not\equiv c \bmod p} \left(\frac{x-d}{p}\right) = -\left(\frac{c-d}{p}\right) = \pm 1$). Explicitly, we find

$$\begin{aligned} S_1(p) &= p^3 - p^2 - p + 1 \\ S_2(p) &= 2p^4 - 2p^3 - 3p^2 + 2p + 1 \\ S_3(p) &= 5p^5 - 5p^4 - 9p^3 + 4p^2 + 4p + 1. \end{aligned} \tag{B.4}$$

As the evaluation of these sums is central to this and other investigations, we provide two proofs of the formula for $S_1(p)$ in the hopes that these arguments will be of use to other researchers studying similar questions.

We first give the proof in [Mil1]. We have the following expansion of $\left(\frac{x}{p}\right)$:

$$\left(\frac{x}{p}\right) \;=\; G_p^{-1} \sum_{c=1}^{p} \left(\frac{c}{p}\right) \mathbf{e}\left(\frac{cx}{p}\right), \tag{B.5}$$

where $\mathbf{e}\left(\frac{a}{p}\right) = \exp(2\pi i a/p)$ and $G_p = \sum_{a(p)} \left(\frac{a}{p}\right) \mathbf{e}\left(\frac{a}{p}\right)$, which equals $\sqrt{p}$ for $p \equiv 1(4)$ and $i\sqrt{p}$ for $p \equiv 3(4)$. See, for example, [BEW].

For the curve $y^2 = f_E(x) = x^3 - ax - b$, $a_E(p) = -\sum_{x(p)} \left(\frac{f_E(x)}{p}\right)$. We use (B.5) to rewrite $a_E(p)$ as

$$a_E(p) \;=\; -G_p^{-1} \sum_{x(p)} \sum_{c=1}^{p} \left(\frac{c}{p}\right) \mathbf{e}\left(\frac{c f_E(x)}{p}\right). \tag{B.6}$$

We take the complex conjugate, which on the RHS introduces a minus sign into the exponential and sends G_p to $\overline{G_p}$, and has no effect on the LHS (which is real). The sum becomes

$$\begin{aligned} S &= (G_p \overline{G_p})^{-1} \sum_{a=0}^{p-1} \sum_{b=0}^{p-1} \prod_{i=1}^{2} \sum_{x_i=0}^{p-1} \sum_{c_i=0}^{p-1} \left(\frac{c_i}{p}\right) \mathbf{e}\left(\frac{(-1)^{i+1}(c_i x_i^3 - c_i a x_i - c_i b)}{p}\right) \\ &= \frac{1}{p} \sum_{x_1, c_1 = 0}^{p-1} \sum_{x_2, c_2 = 0}^{p-1} \left(\frac{c_1 c_2}{p}\right) \mathbf{e}\left(\frac{c_1 x_1^3 - c_2 x_2^3}{p}\right) \sum_{a=0}^{p-1} \mathbf{e}\left(\frac{-(c_1 x_1 - c_2 x_2)a}{p}\right) \\ &\quad \cdot \sum_{b=0}^{p-1} \mathbf{e}\left(\frac{-(c_1 - c_2)b}{p}\right). \end{aligned} \tag{B.7}$$

The b-sum vanishes unless $p|(c_1 - c_2)$, which only happens if $c_1 = c_2 = c$. The a-sum vanishes unless $p|(cx_1 - cx_2)$. As $c \not\equiv 0(p)$ (we have the factor $\left(\frac{c}{p}\right)$) this forces $x_1 = x_2 = x$. As c is non-zero, $\left(\frac{c^2}{p}\right) = 1$, the first exponential factor is 1, and the sums

collapse to

$$\begin{aligned} S &= \frac{1}{p}\sum_{c=1}^{p-1}1\sum_{x=0}^{p-1}1\sum_{a=0}^{p-1}1\sum_{b=0}^{p-1}1 \\ &= \frac{1}{p}(p-1)\cdot p\cdot p\cdot p = p^3-p^2. \end{aligned} \tag{B.8}$$

Remark B.1. *We sketch an alternate proof for $S_1(R)$. We have*

$$S_1(R) \;=\; \sum_{a \bmod p}\;\sum_{b \bmod p}\;\sum_{x \bmod p}\;\sum_{y \bmod p}\left(\frac{x^3-ax-b}{p}\right)\left(\frac{y^3-ay-b}{p}\right). \tag{B.9}$$

We use the following result:

$$\begin{aligned} \mathcal{R} &= \sum_{n \bmod p}\left(\frac{n+c_1}{p}\right)\left(\frac{n+c_2}{p}\right) \\ &= \sum_{n \bmod p}\left(\frac{n^2+n(c_2-c_1)}{p}\right) \\ &= \sum_{n \bmod p}\left(\frac{n^2+\alpha n(c_2-c_1)}{p}\right) \end{aligned} \tag{B.10}$$

for any $\alpha \not\equiv 0 \bmod p$. Thus

$$(p-1)\mathcal{R} \;=\; \sum_{\alpha\not\equiv 0 \bmod p}\;\sum_{n \bmod p}\left(\frac{n^2+\alpha n(c_2-c_1)}{p}\right) \;=\; -(p-1), \tag{B.11}$$

so $\mathcal{R} = -1$. Thus

$$\sum_{n \bmod p}\left(\frac{n+c_1}{p}\right)\left(\frac{n+c_2}{p}\right) \;=\; \begin{cases} p-1 & \text{if } c_1 \equiv c_2 \bmod p \\ -1 & \text{otherwise.} \end{cases} \tag{B.12}$$

We rewrite our sum (replacing a with $-a$ and b with $-b$) as

$$S_1(R) \;=\; \sum_{a \bmod p}\;\sum_{x \bmod p}\;\sum_{y \bmod p}\left[\sum_{b \bmod p}\left(\frac{b+(x^3+ax)}{p}\right)\left(\frac{b+(y^3+ay)}{p}\right)\right]. \tag{B.13}$$

When is $x^3+ax \equiv y^3+ay \bmod p$? This is always true if $x=y$ and a is arbitrary, which gives a contribution of $p\cdot p\cdot(p-1)$. If $x\neq y$ (which happens p^2-p times), there is a unique value of a that works, namely $-(x^3-y^3)/(x-y)$. For this special a the contribution is $(p^2-p)\cdot 1\cdot(p-1)$, and for the other a the contribution is $(p^2-p)\cdot(p-1)\cdot(-1)$. Adding yields p^3-p^2.

REFERENCES

[AL-RM] S. Arms, Á. Lozano-Robledo and S. J. Miller, *Constructing one-parameter families of elliptic curves over $\mathbb{Q}(T)$ with moderate rank*, Journal of Number Theory **123** (2007), no. 2, 388–402.

[BZ] S. Baier and L. Zhao, *The Sato-Tate conjecture on average for small angles*, Transactions of the AMS **361** (2009), no. 4, 1811–1832.

[BS] W. D. Banks and I. E. Shparlinski, *Sato-Tate, cyclicity, and divisibility statistics on average for elliptic curves of small height*, Israel J. Math. **173** (2009), 253–277.

[B-LGG] T. Barnet-Lamb, T. Gee and D. Geraghty, *The Sato-Tate conjecture for Hilbert modular forms*, preprint (2009). `http://arxiv.org/abs/0912.1054`.

[B-LGHT] T. Barnet-Lamb, D. Geraghty, M. Harris and R. Taylor, *A family of Calabi-Yau varieties and potential automorphy II.*, preprint (2009). `http://www.math.harvard.edu/ rtaylor/cy2.pdf`.

[BEW] B. Berndt, R. Evans and K. Williams, *Gauss and Jacobi Sums*, Canadian Mathematical Society Series of Monographs and Advanced Texts, vol. **21**, Wiley-Interscience Publications, John Wiley & Sons, Inc., New York, 1998.

[Bi] B. Birch, *How the number of points of an elliptic curve over a fixed prime field varies*, J. London Math. Soc. **43**, 1968, $57 - 60$.

[BS-D1] B. Birch and H. Swinnerton-Dyer, *Notes on elliptic curves. I*, J. reine angew. Math. **212**, 1963, $7 - 25$.

[BS-D2] B. Birch and H. Swinnerton-Dyer, *Notes on elliptic curves. II*, J. reine angew. Math. **218**, 1965, $79 - 108$.

[BCDT] C. Breuil, B. Conrad, F. Diamond and R. Taylor, *On the modularity of elliptic curves over **Q**: wild 3-adic exercises*, J. Amer. Math. Soc. **14**, no. 4, 2001, $843 - 939$.

[Bro] M. L. Brown, Heegner modules and elliptic curves, Lecture Notes In Mathematics, vol. 1849, Springer-Verlag, 2004.

[CHT] L. Clozel, M. Harris and R. Taylor, *Automorphy for some ℓ-adic lifts of automorphic mod ℓ Galois representations*, Publications Mathématiques de L'IHÉS **108** (2008), no. 1, 1–181.

[CW] J. Coates and A. Wiles, *On the conjecture of Birch and Swinnerton-Dyer*, Invent. Math. **39** (1977), no. 3, 223–251.

[CDF] J. Brian Conrey, W. Duke and D. Farmer, *The distribution of the eigenvalues of Hecke operators*, Acta Arithmetica **78** (1997), no. 4, 405–409.

[De] P. Deligne, *La conjecture de Weil. II* Inst. Hautes Études Sci. Publ. Math. **52**, 1980, $137 - 252$.

[Deu] M. Deuring, *Die Typen der Multiplikatorenringe elliptischer Funktionenkörper*, Abh. Math. Sem. Hansischen Univ. **14** (1941), 197–272.

[Di] F. Diamond, *On deformation rings and Hecke rings*, Ann. Math. **144**, 1996, $137 - 166$.

[Fe1] S. Fermigier, *Zéros des fonctions L de courbes elliptiques*, Exper. Math. **1**, 1992, $167 - 173$.

[Fe2] S. Fermigier, *Étude expérimentale du rang de familles de courbes elliptiques sur $\mathbb{Q}$*, Exper. Math. **5**, 1996, $119 - 130$.

[GKZ] B. H. Gross, W. Kohnen and D. B. Zagier, *Heegner points and derivatives of L-series. II*, Mathematische Annalen **278** (1987), no. 1Ű4, 497–562.

[GZ] B. H. Gross and D. B. Zagier, *Heegner points and derivatives of L-series*, Inventiones Mathematicae **84** (1986), no. 2, 225–320.

[HS-BT] M. Harris, N. Shepherd-Barron and R. Taylor, *A family of Calabi-Yau varieties and potential automorphy*, to appear in the Annals of Math.

[He1] M. Hecke, *Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen, I*, Math. Z. **1** (1918), 357–376.

[He2] M. Hecke, *Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen, II*, Math. Z. **6** (1920), 11–51.

[Ka] N. Katz, *Gauss Sums, Kloosterman Sums, and Monodromy Groups*, Princeton University Press, Princeton, NJ 1988.

[Kol1] V. A. Kolyvagin, *The Mordell-Weil and Shafarevich-Tate groups for Weil elliptic curves*, Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 6, 1154–1180, 1327; translation in Math. USSR-Izv. **33** (1989), no. 3, 473–499

[Kol2] V. A. Kolyvagin, *Finiteness of $E(Q)$ and $\mathrm{Shah}(E,Q)$ for a subclass of Weil curves*, Izv. Akad. Nauk SSSR Ser. Mat. **52** (1988), no. 3, 522–540, 670–671; translation in Math. USSR-Izv. **32** (1989), no. 3, 523–541.

[Mic] P. Michel, *Rang moyen de familles de courbes elliptiques et lois de Sato-Tate*, Monat. Math. **120**, $1995, 127-136$.

[Mil1] S. J. Miller, 1- *and* 2*-Level Densities for Families of Elliptic Curves: Evidence for the Underlying Group Symmetries*, Princeton University, Ph. D. thesis, 2002. `http://www.williams.edu/go/math/sjmiller/public_html/math/thesis/thesis.html`

[Mil2] S. J. Miller, *Variation in the number of points on elliptic curves and applications to excess rank*, C. R. Math. Rep. Acad. Sci. Canada **27** (2005), no. 4, 111–120.

[Mil3] S. J. Miller, *Investigations of zeros near the central point of elliptic curve L-functions*, Experimental Mathematics **15** (2006), no. 3, 257–279.

[Mil4] S. J. Miller, *An identity for sums of polylogarithm functions*, Integers: Electronic Journal Of Combinatorial Number Theory **8** (2008), #A15.

[MS] M. Ram Murty and K. Sinha, *Effective equidistribution of eigenvalues of Hecke operators*, Journal of Number Theory **129** (2009) 681–714.

[Na] K. Nagao, *$\mathbb{Q}(t)$-rank of elliptic curves and certain limit coming from the local points*, Manuscr. Math. **92**, $1997, 13-32$.

[PS] P. Paule and M. Schorn, *A Mathematica Version of Zeilberger's Algorithm for Proving Binomial Coefficient Identities*, J. Symbolic Computation **11** (1994), 1-25.

[PSR] P. Paule, M. Schorn and A. Riese, *An Implementation Of Zeilberger's Fast Algorithm*, `http://www.risc.uni-linz.ac.at/research/combinat/software/PauleSchorn/index.php`

[RS] M. Rosen and J. Silverman, *On the rank of an elliptic surface*, Invent. Math. **133** (1998), 43–67.

[Ru] K. Rubin, *The one-variable main conjecture for elliptic curves with complex multiplication, L-functions and arithmetic* (Durham, 1989), in London Math. Soc. Lecture Note Series **153**, Cambridge Univ. Press, Cambridge, 1991, pages 353–371.

[Sar] P. Sarnak, *Statistical properties of eigenvalues of the Hecke operators*, in Analytic Number Theory and Diophantine Problems (Stillwater, OK, 1984), Progr. Math. 70, Birkhäuser, Boston, 1987, 321–331.

[Sel] A. Selberg, *On the estimation of Fourier Coefficients of Modular Forms*, Proc. Amer. Math. Soc., Symposia in Pure Math. VIII: Theory of Numbers (Pasedena, 1963), 1–15.

[Ser] J.-P. Serre, *Répartition Asymptotique des Valuers Propres de l'Operateur de Hecke T_p*, J. Amer. Math. Soc. **10** (1997), no. 1, 75–102.

[Sh1] I. E. Shparlinski, *On the Lang-Trotter and Sato-Tate Conjectures on Average for Some Families of Elliptic Curves*, preprint.

[Sh2] I. E. Shparlinski, *On the Sato-Tate Conjecture on Average for Some Families of Elliptic Curves*, preprint.

[Sil1] J. Silverman, *The Arithmetic of Elliptic Curves*, Graduate Texts in Mathematics **106**, Springer-Verlag, Berlin - New York, 1986.

[Sil2] J. Silverman, *Advanced Topics in the Arithmetic of Elliptic Curves*, Graduate Texts in Mathematics **151**, Springer-Verlag, Berlin - New York, 1994.

[ST] J. Silverman and J. Tate, *Rational Points on Elliptic Curves*, Springer-Verlag, New York, 1992.

[Ta] J. T. Tate, *Algebraic cycles and poles of zeta-functions*, in Arithmetical Algebraic geometry (Proc. Purdue Conf. 1963), O. F. G. Schilling (ed.), Harper & Row, 1965, pp. 93–110.

[Tay] R. Taylor, *Automorphy for some ℓ-adic lifts of automorphic mod ℓ Galois representations. II*, Publications Mathématiques de L'IHÉS **108** (2008), no. 1, 183–239.

[TW] R. Taylor and A. Wiles, *Ring-theoretic properties of certain Hecke algebras*, Ann. Math. **141**, 1995, $553 - 572$.

[Ward] K. J. Ward, *Series Sums of Binomial Coefficients*, webpage: `http://www.trans4mind.com/personal_development/mathematics/series/summingBinomialCoefficients.htm`

[Wa] L. Washington, *Class numbers of the simplest cubic fields*, Math. Comp. **48**, number 177, 1987, $371 - 384$.

[Wi] Wikipedia, *Chebyshev Polynomials*, June 21, 2009. `http://en.wikipedia.org/wiki/Chebyshev_polynomials`.

E-mail address: `Steven.J.Miller@williams.edu`

DEPARTMENT OF MATHEMATICS AND STATISTICS, WILLIAMS COLLEGE, WILLIAMSTOWN, MA 01267

E-mail address: `murty@mast.queensu.ca`

DEPARTMENT OF MATHEMATICS, QUEEN'S UNIVERSITY, KINGSTON, ONTARIO, K7L 3N6, CANADA