Bull. Aust. Math. Soc. **100** (2019), 27–33
doi:10.1017/S0004972718001648

THE ERROR TERM IN THE SATO–TATE THEOREM OF BIRCH

M. RAM MURTY and NEHA PRABHU[✉]

(Received 19 September 2018; accepted 14 November 2018; first published online 8 February 2019)

Abstract

We establish an error term in the Sato–Tate theorem of Birch. That is, for p prime, $q = p^r$ and an elliptic curve $E : y^2 = x^3 + ax + b$, we show that

$$\#\{(a,b) \in \mathbb{F}_q^2 : \theta_{a,b} \in I\} = \mu_{ST}(I)q^2 + O_r(q^{7/4})$$

for any interval $I \subseteq [0,\pi]$, where the quantity $\theta_{a,b}$ is defined by $2\sqrt{q}\cos\theta_{a,b} = q + 1 - E(\mathbb{F}_q)$ and $\mu_{ST}(I)$ denotes the Sato–Tate measure of the interval I.

2010 *Mathematics subject classification*: primary 11G05; secondary 11K38.

Keywords and phrases: Sato–Tate theorem, elliptic curves.

1. Introduction

In 1968, Birch [3] proved that the Sato–Tate conjecture holds for the family of elliptic curves

$$y^2 = x^3 + ax + b \,(\mathrm{mod}\ p)$$

as (a,b) varies over elements in $(\mathbb{Z}/p\mathbb{Z})^2$ such that $4a^3 + 27b^2 \not\equiv 0 \,(\mathrm{mod}\ p)$, where p is a fixed prime. Let $(\frac{\cdot}{p})$ denote the Legendre symbol. More precisely, he proved that

$$\sum_{a,b=0}^{p-1}\left|\sum_{x=0}^{p-1}\left(\frac{x^3+ax+b}{p}\right)\right|^{2R} \sim \frac{1}{R+1}\binom{2R}{R}p^{R+2} \quad \text{as } p \to \infty.$$

These are the moments predicted by the Sato–Tate conjecture and by standard probability theory one can deduce the relevant distribution. This moment calculation then implies the Sato–Tate distribution for the angles $\theta_{a,b}$, where we write

$$a_p(a,b) = -\sum_{x=0}^{p-1}\left(\frac{x^3+ax+b}{p}\right) = 2\sqrt{p}\cos\theta_{a,b}.$$

Here, $E(\mathbb{F}_p)$ denotes the number of $\mathbb{F}_p$-points (including the point at infinity) on the elliptic curve $y^2 = x^3 + ax + b$ and $a_p(a,b) = p + 1 - E(\mathbb{F}_p)$. There are two key

ingredients in Birch's proof. The first is Deuring's theorem [8] that there are a total of $H(t^2-4p)$ isomorphism classes of elliptic curves over $\mathbb{F}_p$ with $p+1-t$ points, where $H(N)$ denotes the Hurwitz–Kronecker class number. The second ingredient is the Eichler–Selberg trace formula (see [12, Appendix]), which gives the trace of the nth Hecke operator acting on the space of holomorphic cusp forms of even weight $k \geq 4$ with respect to the full modular group.

A natural question that arises from Birch's paper concerns the order of the error term in the Sato–Tate distribution. More precisely, fix an interval $I = [\alpha, \beta] \subseteq [0,1]$. We want to count the number

$$N_I(p) := \#\left\{(a,b) \in (\mathbb{Z}/p\mathbb{Z})^2 : p \nmid \Delta(a,b) \frac{\theta_{a,b}}{\pi} \in I\right\}$$

and give a quantitative estimate for

$$|N_I(p) - \mu_{ST}(I)p^2|, \tag{1.1}$$

where the Sato–Tate measure of the interval I is given by $\mu_{ST}(I) = 2\int_I \sin^2(\pi\theta)\, d\theta$.

Using the discrepancy estimate of Niederreiter [19], Banks and Shparlinski [2] noted that

$$N_I(p) - \mu_{ST}(I)p^2 = O(p^{7/4}) \tag{1.2}$$

would follow from the work of Katz [10] extended to their setting. This extension is not routine and appears in the work of Michel [14], where he deals with the case of one-parameter families of elliptic curves. There is also a related paper of Fisher [9]. The work of Michel [14] relies heavily on Weil II (see Deligne [7]). From the perspective of classical analytic number theory, Weil II and its cohomological mysteries present formidable prerequisites that often represent a 'black box' whose pronouncements must be accepted on faith. On the other hand, using the moment estimates in Birch's proof, Miller and Ram Murty [15] also estimated the quantity (1.1), but were only able to achieve a logarithmic saving over the trivial estimate of p^2. The goal of this paper is to show that the estimate (1.2) can be deduced using classical techniques, from just the Ramanujan–Petersson conjecture, now a theorem due to Deligne (which is implied by Weil I [5, 6]), as well as the two key ingredients of Birch [3] mentioned earlier. The result is, in fact, true in the more general case of elliptic curves over a finite field $\mathbb{F}_q = \mathbb{F}_{p^r}$. Let

$$N_I(q) := \#\left\{(a,b) \in \mathbb{F}_q^2 : \Delta(a,b) \neq 0, \frac{\theta_{a,b}}{\pi} \in I\right\},$$

where we now have $a_q(a,b) = q + 1 - E(\mathbb{F}_q) = 2\sqrt{q}\cos\theta_{a,b}$. We prove the following theorem.

Theorem 1.1. *Assume the notation above. Then*

$$N_I(q) - \mu_{ST}(I)q^2 = O_r(q^{7/4}).$$

We note that the result of Banks and Shparlinski (and that in Theorem 1.1) gives a true error term only when the size of the interval I is greater that $p^{-1/4+\varepsilon}$. This was improved (on average) by Baier and Zhao in [1] and recently by David *et al.* [4], where the effective version of Birch's theorem is shown to hold for intervals I of length greater than $p^{-1/2+\varepsilon}$, although in these cases the saving is only a power of a logarithm over the main term.

2. Preliminaries

2.1. Isomorphism classes of elliptic curves. We briefly discuss the ingredients from the theory of counting elliptic curves, which will be needed for the proof. For more details, see [13] or [20].

For $p \neq 2, 3$, consider the elliptic curve over $\mathbb{F}_q$ in Weierstrass form

$$E_{a,b} : y^2 = x^3 + ax + b.$$

Analogous to the case of $\mathbb{F}_p$, the number of $\mathbb{F}_q$-points on E, given by $E(\mathbb{F}_q)$, is $q + 1 - a_q(a, b)$. We have Hasse's bound $|a_q(a, b)| \leq 2\sqrt{q}$. Two curves $E = E_{a,b}$ and $E' = E_{a',b'}$ over $\mathbb{F}_q$ are isomorphic if there is an element $u \in \mathbb{F}_q^*$ such that $a' = u^4 a$ and $b' = u^6 b$. An automorphism of E is an isomorphism from E to E. Clearly, isomorphism of elliptic curves is an equivalence relation and the size of the equivalence class of E is $\#\mathbb{F}_q^*/\#\mathrm{Aut}\, E$.

For $t^2 < 4q$, Deuring [8] essentially showed that the number of isomorphism classes of elliptic curves E with $q + 1 - t$ points, weighted by $\#(\mathrm{Aut}\, E)^{-1}$, is given by $H(t^2 - 4q)$, where $H(N)$ is the Hurwitz–Kronecker class number (see [13] for a detailed description of these numbers). Thus, for $t^2 < 4q$, the total number of curves E over $\mathbb{F}_q$ with $q + 1 - t$ points is $(q - 1)H(t^2 - 4q)$.

2.2. Chebyshev polynomials. The Chebyshev polynomials $U_n(x)$ of the second kind, for integers $n \geq 0$, are defined recursively in the following way:

$$\begin{aligned} U_0(x) &= 1, \\ U_1(x) &= 2x, \\ U_n(x) &= 2xU_{n-1}(x) - U_{n-2}(x), \quad n \geq 2. \end{aligned}$$

If $x = \cos\theta$, then the polynomials can be written explicitly as

$$U_n(x) = \frac{\sin(n+1)\theta}{\sin\theta}.$$

In our application of these polynomials, $x = \cos\theta_{a,b}$. It is not hard to see that

$$\frac{\rho^{k-1} - \bar{\rho}^{k-1}}{\rho - \bar{\rho}} = q^{(k-2)/2} U_{k-2}\Big(\frac{t}{2\sqrt{q}}\Big),$$

where ρ and $\bar{\rho}$ are the roots of the equation $y^2 - ty + q = 0$. Observe that

$$2\cos(n\theta) = \frac{\sin(n+1)\theta}{\sin\theta} - \frac{\sin(n-1)\theta}{\sin\theta} = U_n(\cos\theta) - U_{n-2}(\cos\theta),$$

a fact that will be needed later.

2.3. Beurling–Selberg polynomials. The Beurling–Selberg polynomials have been frequently used to obtain effective results on equidistribution and, in this note, we use this method to study the quantity $N_I(q)$. We give a brief introduction to these polynomials (see [16, Ch. 1] for a detailed exposition). Let $\chi_I(x)$ denote the characteristic function of the interval $I = [\alpha, \beta] \subseteq [0, 1]$ and let $M \geq 1$ be an integer. One can construct trigonometric polynomials $S^-_{M,I}(x)$ and $S^+_{M,I}(x)$ of degree less than or equal to M, respectively called the minorant and majorant Beurling–Selberg polynomials for the interval I, such that:

(a) for all $x \in \mathbb{R}$, $\quad S^-_{M,I}(x) \leq \chi_I(x) \leq S^+_{M,I}(x)$;

(b)
$$\widehat{S}^{\pm}_{M,I}(0) = \int_0^1 S^{\pm}_{M,I}(x)\,dx = \beta - \alpha \pm \frac{1}{M+1};$$

(c) for $0 < |m| \leq M$,
$$\left|\widehat{S}^{\pm}_{M,I}(m) - \widehat{\chi}_I(m)\right| \leq \frac{1}{M+1}.$$

Henceforth, we will suppress the I in the subscript of the Fourier coefficients, with the understanding that the definition of these approximating polynomials depends on the interval I.

For $I = [\alpha, \beta]$, we will also use the following estimates, which follow from properties (b) and (c) listed above:
$$\widehat{S}^{\pm}_{M}(0) = (\beta - \alpha) \pm \frac{1}{M+1} \tag{2.1}$$
and, for $m > 0$,
$$\widehat{S}^{\pm}_{M}(m) + \widehat{S}^{\pm}_{M}(-m) = \frac{\sin 2\pi m\beta - \sin 2\pi m\alpha}{m\pi} + O\Big(\frac{1}{M+1}\Big). \tag{2.2}$$

These polynomials were used in [18] to study the 'vertical' Sato–Tate distribution in the case of modular forms.

3. Proof of the main theorem

We consider the angles $\{\theta_{a,b}/\pi, -\theta_{a,b}/\pi\}$ and count when they occur in $I' = I/\pi$. Approximating using Beurling–Selberg polynomials,

$$\begin{aligned} N_I(q) &= \sum_{\substack{a,b\in\mathbb{F}_q \\ \Delta(a,b)\neq 0}} \chi_{I'}\Big(\frac{\theta_{a,b}}{\pi}\Big) \leq \sum_{\substack{a,b\in\mathbb{F}_q \\ \Delta(a,b)\neq 0}} \Big(\sum_{|m|\leq M} \widehat{S}^{+}_{M}(m) e\Big(m\frac{\theta_{a,b}}{\pi}\Big) + \sum_{|m|\leq M} \widehat{S}^{+}_{M}(m) e\Big(-m\frac{\theta_{a,b}}{\pi}\Big)\Big) \\ &= \sum_{\substack{a,b\in\mathbb{F}_q \\ \Delta(a,b)\neq 0}} \sum_{|m|\leq M} \widehat{S}^{+}_{M}(m) 2\cos(2m\theta_{a,b}) \\ &= 2\widehat{S}^{+}_{M}(0) \sum_{\substack{a,b\in\mathbb{F}_q \\ \Delta(a,b)\neq 0}} 1 + \sum_{0<|m|\leq M} \widehat{S}^{+}_{M}(m) \sum_{\substack{a,b\in\mathbb{F}_q \\ \Delta(a,b)\neq 0}} [U_{2m}(\cos\theta_{a,b}) - U_{2m-2}(\cos\theta_{a,b})], \end{aligned}$$

where, as noted in Section 2.2,

$$U_r(\cos\theta_{a,b}) = \frac{\sin(r+1)\theta_{a,b}}{\sin\theta_{a,b}}$$

denotes the rth Chebyshev polynomial of the second kind evaluated at $\cos(\theta_{a,b})$. Note that $U_0(x) = 1$. Therefore,

$$\begin{aligned} N_I(q) \le q^2(2\widehat{S}^+_M(0) - (\widehat{S}^\pm_M(1) + \widehat{S}^\pm_M(-1))) + (\widehat{S}^\pm_M(1) + \widehat{S}^\pm_M(-1)) \sum_{\substack{a,b\in\mathbb{F}_q \\ \Delta(a,b)\neq 0}} U_2(\cos\theta_{a,b}) \\ + \sum_{2\le m\le M} (\widehat{S}^\pm_M(m) + \widehat{S}^\pm_M(-m)) \sum_{\substack{a,b\in\mathbb{F}_q \\ \Delta(a,b)\neq 0}} [U_{2m}(\cos\theta_{a,b}) - U_{2m-2}(\cos\theta_{a,b})]. \end{aligned} \tag{3.1}$$

Using (2.1) and (2.2),

$$2\widehat{S}^\pm_M(0) - (\widehat{S}^\pm_M(m) + \widehat{S}^\pm_M(-m)) = \mu_{ST}(I) + O\Big(\frac{1}{M+1}\Big), \tag{3.2}$$

where $\mu_{ST}(I)$ denotes the Sato–Tate measure of I. It remains to estimate the sums

$$\sum_{\substack{a,b\in\mathbb{F}_q \\ \Delta(a,b)\neq 0}} \frac{\sin(2m+1)\theta_{a,b}}{\sin\theta_{a,b}}$$

for $m = 1, \dots, M$. If we write $a_q(a,b) = t = 2\sqrt{q}\cos\theta_t$, where $|t| \le 2\sqrt{q}$, then

$$\sum_{\substack{a,b\in\mathbb{F}_q \\ \Delta(a,b)\neq 0}} \frac{\sin(2m+1)\theta_{a,b}}{\sin\theta_{a,b}} = (q-1)\sum_{|t|\le 2\sqrt{q}} H(t^2-4q)\frac{\sin(2m+1)\theta_t}{\sin\theta_t}, \tag{3.3}$$

where we group the curves into isomorphism classes, as discussed in Section 2.1. On the other hand, using the Eichler–Selberg trace formula (see [12] or [11]), for $k \ge 4$,

$$\begin{aligned} \operatorname{Tr} T_k(q) &= \frac{k-1}{12}q^{(k-2)/2}\delta(q,2) - \frac{1}{2}\sum_{|t|\le 2\sqrt{q}} \frac{\rho^{k-1} - \bar\rho^{k-1}}{\rho - \bar\rho} H(t^2-4q) \\ &\quad - \frac{1}{2}\sum_{dd'=q} \min(d,d')^{k-1} \\ &= \frac{k-1}{12}q^{(k-2)/2}\delta(q,2) - \frac{1}{2}\sum_{|t|\le 2\sqrt{q}} q^{(k-2)/2}\frac{\sin(k-1)\theta_t}{\sin\theta_t} H(t^2-4q) \\ &\quad - \frac{1}{2}\sum_{dd'=q} \min(d,d')^{k-1}, \end{aligned}$$

where $\delta(q,2)$ is 1 when q is a square and zero otherwise. Using the Ramanujan–Petersson bound for Hecke eigenvalues,

$$\operatorname{Tr} T_k(q) \ll k q^{(k-1)/2}, \tag{3.4}$$

since the dimension of the space of cusp forms of weight k and full level grows like k. Therefore,

$$\sum_{|t|\le 2\sqrt{q}} H(t^2-4q)\frac{\sin(k-1)\theta_t}{\sin\theta_t} \ll rkq^{1/2}, \tag{3.5}$$

where $q=p^r$. Going back to (3.3), letting $k=2m+2$, we deduce that for $m=1,\dots,M$,

$$\sum_{\substack{a,b\in\mathbb{F}_q\\ \Delta(a,b)\neq 0}} \frac{\sin(2m+1)\theta_{a,b}}{\sin\theta_{a,b}} \ll rmq^{3/2}. \tag{3.6}$$

Applying (3.2), (3.6) and the estimate $(\widehat{S}^{\pm}_M(m)+\widehat{S}^{\pm}_M(-m)) \ll 1/m$ (evident from (2.2)) to (3.1),

$$N_I(q)-q^2\mu_{ST}(I) \ll \frac{q^2}{M}+rMq^{3/2}. \tag{3.7}$$

Finally, letting $M=\lfloor p^{1/4}r^{-1/2}\rfloor$ gives

$$N_I(q)-q^2\mu_{ST}(I) \ll_r q^{7/4}.$$

The lower bound estimation can be obtained in a similar way using $\widehat{S}^{-}_M(m)$.

4. Concluding remarks

It is interesting to consider to what extent our error term is best possible. For example, for $q=p$, the sum in (3.4) is essentially $p^{-(k-1)/2}\operatorname{Tr}T_k(p)$. If we accept the prediction about the Sato–Tate distribution corresponding to distinct Hecke eigenforms as discussed in [17], then we can arrange all the Fourier coefficients appearing in $\operatorname{Tr}T_k(p)$ to be arbitrarily close to $2p^{(k-1)/2}$ simultaneously, for infinitely many primes p. Thus, the estimate in (3.5) cannot be improved for all primes p. This does not however say anything about the combined error term in (3.7). Thus, the question of the optimal error term becomes an intriguing problem for future research.

Acknowledgements

The authors would like to thank Igor Shparlinski for his helpful comments and the anonymous referee for suggestions that improved the exposition of this article.

References

[1] S. Baier and L. Zhao, 'The Sato–Tate conjecture on average for small angles', *Trans. Amer. Math. Soc.* **361**(4) (2009), 1811–1832.
[2] 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.
[3] B. J. Birch, 'How the number of points of an elliptic curve over a fixed prime field varies', *J. Lond. Math. Soc.* **43** (1968), 57–60.
[4] C. David, D. Koukoulopoulos and E. Smith, 'Sums of Euler products and statistics on elliptic curves', *Math. Ann.* **368** (2017), 685–752.

[5] P. Deligne, 'Formes modulaires et représentations l-adiques', in: *Séminaire Bourbaki, Vol. 1968/69, Exp. No. 355*, Lecture Notes in Mathematics, 175 (Springer, Berlin, 1971), 139–172.
[6] P. Deligne, 'La conjecture de Weil. I', *Publ. Math. Inst. Hautes Études Sci.* **43** (1974), 273–307.
[7] P. Deligne, 'La conjecture de Weil. II', *Publ. Math. Inst. Hautes Études Sci.* **52** (1980), 137–252.
[8] M. Deuring, 'Die Typen der Multiplikatorenringe elliptischer Funktionenkörper', *Abh. Math. Semin. Univ. Hambg.* **14** (1941), 197–272.
[9] B. Fisher, 'Equidistribution theorems', in: *Columbia University Number Theory Seminar, New York, 1992*, Astérisque, 228 (Société Mathématique de France, Paris, 1995), 69–79.
[10] N. M. Katz, *Gauss Sums, Kloosterman Sums, and Monodromy Groups* (Princeton University Press, Princeton, NJ, 1988).
[11] A. Knightly and C. Li, *Traces of Hecke Operators*, Mathematical Surveys and Monographs, 133 (American Mathematical Society, Providence, RI, 2006).
[12] S. Lang, *Introduction to Modular Forms*, Fundamental Principles of Mathematical Sciences, 222 (Springer, Berlin, 1995), with appendixes by D. Zagier and Walter Feit, corrected reprint of the 1976 original.
[13] H. W. Lenstra Jr, 'Factoring integers with elliptic curves', *Ann. of Math. (2)* **126**(3) (1987), 649–673.
[14] P. Michel, 'Rang moyen de familles de courbes elliptiques et lois de Sato–Tate', *Monatsh. Math.* **120** (1995), 127–136.
[15] S. J. Miller and M. Ram Murty, 'Effective equidistribution and the Sato–Tate law for families of elliptic curves', *J. Number Theory* **131** (2011), 25–44.
[16] H. L. Montgomery, *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*, CBMS Regional Conference Series in Mathematics, 84 (American Mathematical Society, Providence, RI, 1994).
[17] M. Ram Murty, 'Oscillations of Fourier coefficients of modular forms', *Math. Ann.* **262** (1983), 431–446.
[18] M. Ram Murty and K. Sinha, 'Effective equidistribution of eigenvalues of Hecke operators', *J. Number Theory* **129** (2009), 681–714.
[19] H. Niederreiter, 'The distribution of values of Kloosterman sums', *Arch. Math.* **56** (1991), 270–277.
[20] R. Schoof, 'Nonsingular plane cubic curves over finite fields', *J. Combin. Theory Ser.* A **46**(2) (1987), 183–211.

M. RAM MURTY, Queen's University,
Kingston, Ontario K7L 3N6, Canada
e-mail: murty@queensu.ca

NEHA PRABHU, Queen's University,
Kingston, Ontario K7L 3N6, Canada
e-mail: neha.prabhu@queensu.ca