A SURVEY OF LOWER ORDER BIASES IN MOMENT EXPANSIONS OF ONE PARAMETER FAMILIES OF ELLIPTIC CURVES

TIMOTHY CHEEK, PICO GILMAN, KAREEM JABER, STEVEN J. MILLER, VISMAY SHARAN, AND MARIE-HÉLÈNE TOMÉ

Abstract. For a fixed elliptic curve E without complex multiplication, $a_p := p + 1 - \#E(\mathbb{F}_p)$ is $O(\sqrt{p})$ and $a_p/2\sqrt{p}$ converges to a semicircular distribution. Michel proved that for a one-parameter family of elliptic curves $y^2 = x^3 + A(T)x + B(T)$ with $A(T), B(T) \in \mathbb{Z}[T]$ and non-constant j-invariant, the second moment of $a_p(t)$ is $p^2 + O(p^{3/2})$. The size and sign of the lower order terms have applications to the distribution of zeros near the central point of Hasse-Weil L-functions and the Birch and Swinnerton-Dyer conjecture. S. J. Miller conjectured that the highest order term of the lower order terms of the second moment that does not average to zero is on average negative. Previous work on the conjecture has been restricted to a small set of highly nongeneric families where the resulting sums can be done in closed form, and thus may not be representative of the typical family. We create a database and a framework to quickly and systematically investigate biases in the second moment of any one-parameter family. When looking at families which have so far been beyond current theory, we find several potential violations of the conjecture for $p \leq 250,000$ and discuss new conjectures motivated by the data.

1. Introduction

We assume the reader is familiar with elliptic curves (see, for example, [Sil94; Sil09]). Consider $\mathcal{E} \to \mathbb{P}^1$, a non-split elliptic surface over $\mathbb{Q}$

$$\mathcal{E}\colon\ y^2 \ = \ x^3 + A(T)x + B(T) \tag{1.1}$$

with $4A(T)^3 + 27B(T)^2 \neq 0$. Its j-invariant is given by

$$j(T) \ := \ 1728\frac{4A(T)^3}{4A(T)^3 + 27B(T)^2}. \tag{1.2}$$

and almost all specializations $T = t \in \mathbb{Z}$ result in an elliptic curve E_t,

$$E_t \ := \ y^2 \ = \ x^3 + A(t)x + B(t). \tag{1.3}$$

We say the set of all such E_t forms a *one-parameter family* of elliptic curves over $\mathbb{Q}$.

The n^{th} moment of the Frobenius trace of a one parameter family is defined for each prime p as

$$\mathcal{A}_{n,\mathcal{E}}(p) \ := \ \sum_{t=0}^{p-1} a_t(p)^n \ = \ \sum_{t=0}^{p-1}\left(\sum_{x=0}^{p-1} -\frac{x^3 + A(T)x + B(T)}{p}\right)^n,$$

as the $a_t(p)$'s can be written as a sum of Legendre symbols. Note that we do not normalize by $1/p$, so that this sum is always an integer. The moments of a one-parameter family encode arithmetic properties of the family. For example, Rosen and Silverman [RS98] proved under the assumption that Tate's conjecture holds for the surface[1] that a conjecture of Nagao [Nag97] is true:

$$\lim_{X\to\infty} \frac{1}{X} \sum_{p \leq X} \frac{\mathcal{A}_{1,\mathcal{E}}(p)\log p}{p} \ = \ -\text{rank}\mathcal{E}(\mathbb{Q}(T)), \tag{1.4}$$

[1] This is known for rational surfaces. 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 $\text{ord}_{T=0} T^{12}\Delta(T^{-1}) = 0$.

where the rank of a family is defined to be the minimum rank r that appears infinitely often among the curves E_t in the one-parameter family. This result thusly relates the rank of a surface and its first moments, and comes from a negative bias in the values of the $a_t(p)$'s.

It is natural to ask if biases also exist in higher moments, and if these biases are also arithmetically significant. We first describe previous work, which was restricted to special non-generic families where the resulting Legendre sums can be computed in closed form. While there are simple, closed form expressions for $\sum_{u \bmod p} \left(\frac{f(u)}{p}\right)$ when f is linear or quadratic, for a typical degree three or higher f there are no such formulas (which is why it is so challenging to work with elliptic curves, as the $a_t(p)$'s are cubic Legendre sums). Thus previous work has focused on low moments with one-parameter families with $A(T), B(T)$ of small degree (one changes the order of summation, and for specially constructed families the result after summing over t is a tractable sum of at most degree two Legendre symbols), and thus the evidence may be misleading. Therefore, we adopt a different approach and instead systematically explore many families for lots of primes. To this end, we build a large database of Frobenius traces for primes up to 250,000; given more computing power, the framework we establish can easily be used to generate a larger database, and investigate more families and higher moments. Note this suffices to investigate any family, as for each t we just need to find $x^3 + A(t)x + B(t) \bmod p$; rather than doing many similar computations for different families, we compute all possible values first, and can then easily look at any one parameter family. Thus there is an enormous start-up cost in creating the database, but once constructed it leads to a tremendous decrease in run-time to analyze a given family.

To state earlier work and make our investigations precise, we recall the following asymptotic second moment expansion.

Theorem 1.1 (Michel [Mic95])**.** *Let $\mathcal{E}$ be an elliptic surface with nonconstant j-invariant. Then the second moment is of the form*

$$\mathcal{A}_{2,\mathcal{E}}(p) \;=\; p^2 + O\left(p^{3/2}\right). \tag{1.5}$$

In his Ph.D. thesis, motivated by biases in the first moment expansion, Miller [Mil02] formulated the Bias Conjecture for the second moment.

Bias Conjecture: *The largest lower order term in the second moment expansion of a one-parameter family of elliptic curves with non-constant $j(T)$ which does not average to 0 is on average negative.*

Results confirming the Bias Conjecture in narrow cases have been obtained in [Asa+23; Bat+24; KN21; KN22; Mac+16; Mil02; Mil04; Mil05]. For instance, see [Bat+24] for a table of families with closed form first and second moment expressions as obtained by [Asa+23; Mil02; Mil05]. As remarked, the majority of these calculations rely on linear and quadratic Legendre symbol identities.

To this end, we computationally investigate one-parameter families defined by higher degree polynomials $A(T)$ and $B(T)$ where Legendre symbol calculations are intractable and to which the results of [KN21; KN22] on cubic pencils do not apply. Our numerical work seeks to evaluate whether we expect the Bias Conjecture to hold by investigating a more generic swath of families. To isolate the lower order bias, we compute the normalized second moment

$$\mathcal{B}_{2,\mathcal{E}}(p) \;=\; \frac{\mathcal{A}_{2,\mathcal{E}}(p) - p^2}{p^{3/2}} \tag{1.6}$$

and take a running average over primes p. We expect our database to have many applications to computing other quantities related to the traces of Frobenius of a one-parameter family, and can be used to attempt to formulate analogous Bias Conjectures for the higher moments $n \geq 3$.

We take advantage of two automorphisms and Legendre symbol identities to reduce the number of a_p we need to compute and store to

$$\#a_p \text{ we need to compute and store } = \begin{cases} 2p+2 & \text{if } \ p \equiv 1 \bmod 4 \\ 2p & \text{if } \ p \equiv 3 \bmod 4 \end{cases} + \begin{cases} 6 & \text{if } \ p \equiv 1 \bmod 3 \\ 0 & \text{otherwise.} \end{cases}$$

While we expect our database to have many applications, we were motivated by calculating the second moment of one-parameter families of elliptic curves. We showcase how our database allows us to systematically computationally investigate the Bias Conjecture in one-parameter families for which obtaining a closed form of the second moment through Legendre symbols is intractable. To isolate and study the bias in the lower order terms, we compute the normalized second moment $\mathcal{B}_{2,\mathcal{E}}(p)$ for each prime p smaller than $250,000$ and we compute the running average

$$\frac{1}{\pi(P)-1} \sum_{2<p\leq P} \mathcal{B}_{2,\mathcal{E}}(p). \tag{1.7}$$

We also introduce the log-weighted running average of the normalized second moment:

$$\frac{1}{N_w(P)} \sum_{2<p\leq P} \mathcal{B}_{2,\mathcal{E}}(p) \log p, \tag{1.8}$$

where $N_w(P) \coloneqq \sum_{2<p\leq P} \log p \sim P$. This is desirable since as can be seen in the graphs that follow, the running average for smaller primes is not representative of the long-term behavior.

2. Potential positive bias families

We conducted an exhaustive family search for a potential positive bias family by looking at one parameter families defined by all possible combinations of polynomials $A(T), B(T)$ of degree ≤ 5 with coefficients in $\{0,1\}$. Obviously this is a very small subset of elliptic curves with defining polynomials of low degree, but we cannot find closed form expressions for most of the families, and thus we hope that the elements here are more representative of a generic family. We computed the second moment for each of the resulting families for primes up to $1,000$ and noted those families whose running average of the normalized second moment was positive more than 95% of the time, and then explored these families for primes up to 250,000. [2] Our numerical investigations generate hypotheses and directions for future investigation but cannot prove or disprove the Bias Conjecture. In this sense, our work is very similar to recent progress on murmurations of elliptic curves, where machine learning has led to incredible predictions and discovered new phenomena to study; see for example [Cow23; He+24; HLO22b; HLO22a; HLO23; Zub23; Boo+24]

It is instructive to compare the graphs of the running averages of the families our family search found to the family $y^2 = x^3+x+T^3$ for which [Bat+24] proved that the second moment of this family has positive bias for half of the primes. While the graph of the running averages for $y^2 = x^3+x+T^3$ is negative most of the time but seems to tend to zero at $250,000$, the graphs in Figures 1 and 2 remain positive up to $250,000$ and seem to stabilize for primes beyond $100,000$.

3. Distribution of the Normalized Second Moment $\mathcal{B}_{2,\mathcal{E}}(p)$

Motivated by the Sato-Tate Conjecture, we study the distribution of the normalized second moment of a one-parameter family. It is natural to ask what determines the variance of the distribution of the normalized second moment of a one-parameter family. Based on our numerical computations, we formulate the following conjecture.

[2] Future work could focus on developing better measures of whether a family should be suspected of having positive bias and hence investigated further. For example, excluding primes $p \leq P_{\min}$.

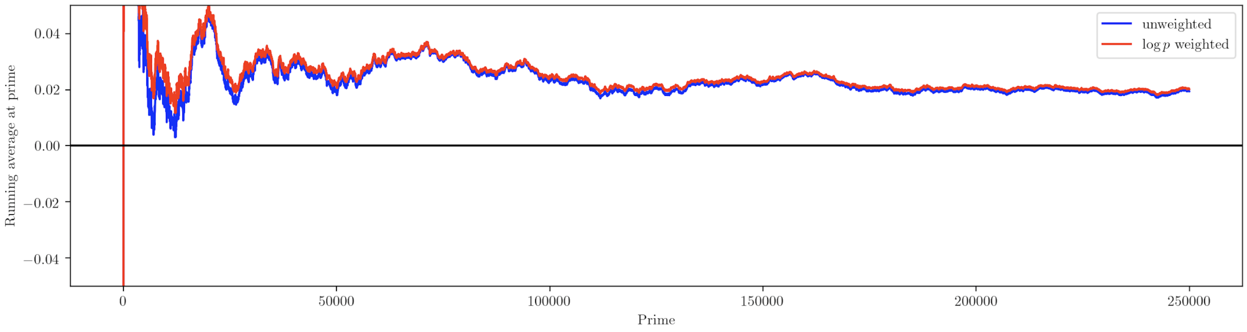


FIGURE 1. Running averages of $y^2 = x^3 + (T^5 + T^3 + T)x + T^3 + T^2 + T + 1$.

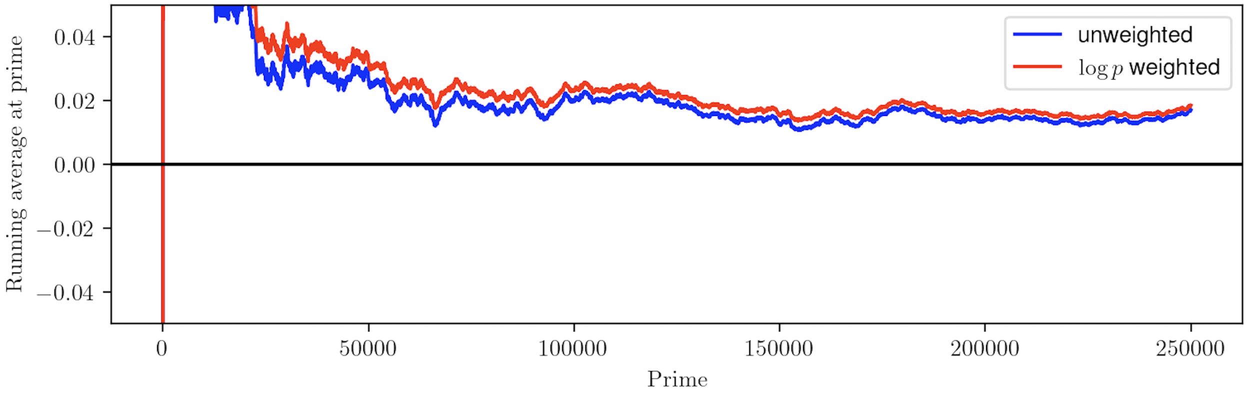
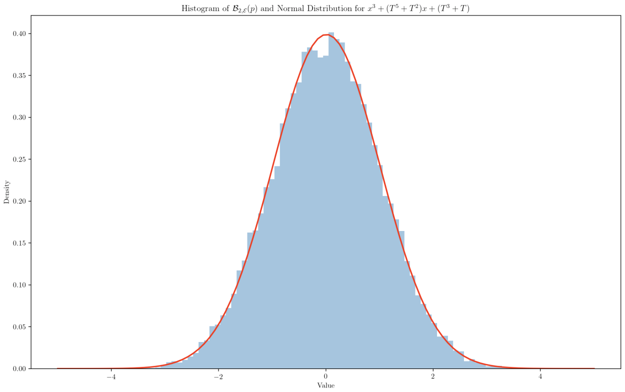


FIGURE 2. Running averages of $y^2 = x^3 + (T^3 + T + 1)x + T^4 + T^2$.

Conjecture 3.1. *The variance of the distribution of $\mathcal{B}_{2,\mathcal{E}}(p)$ always converges to a positive integer.*

Indeed there seems to be a deeper connection between the polynomials $A(T)$ and $B(T)$ and the conjectured integer the variance converges too. Indeed, the one parameter family given by $y^2 = x^3 + A(T^n)x + B(T^n)$ seems to be an integral multiple of $y^2 = x^3 + A(T)x + B(T)$ based on the prime factorizations of n, A, and B.

The following figures have binned the data between the maximum and the minimum over all primes less than $250,000$ and then split into 100 equal sized buckets.



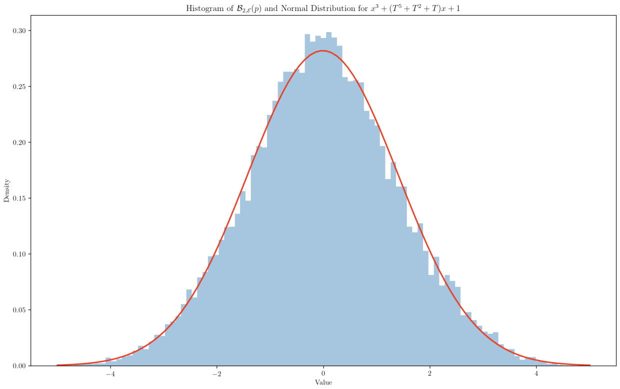


FIGURE 3. A family with variance 1.015

FIGURE 4. A family with variance 1.991.

Our generic case seems to be variance converging to 1 and the distribution of $\mathcal{B}_{2,\mathcal{E}}(p)$ seems to exhibit Gaussian-like behavior. We found one family $\mathcal{E} : y^2 = x^3 + (T^5 + T^2 + 1)x + 1$ whose variance seems to be converging to 2 (see Figure 4).

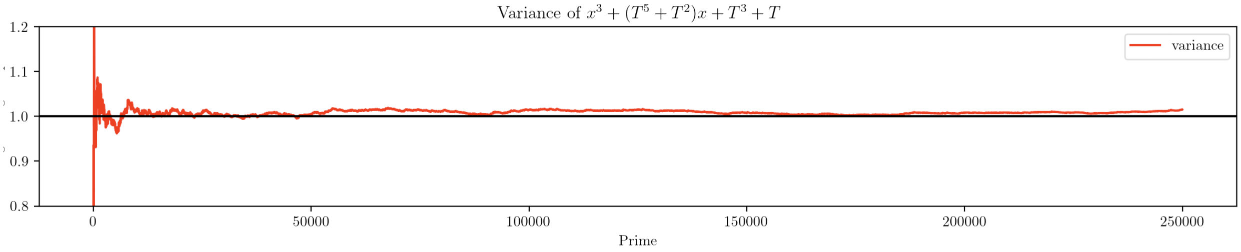


FIGURE 5. The graph of the variance of a family over primes $\leq p$.

Due to Gaussians being unbounded, and thus not the distribution achieved, a natural question is determining what distribution the normalized second moment of a one-parameter family converges to in our seemingly generic case of variance 1 Gaussian-like behavior and what is our generic case, i.e., what restrictions can we impose on our one-parameter family to ensure we are in the generic case?

Our numerics motivate the study of the distribution of the second moment of a one-parameter family. Further, the database allows for efficient computation of higher moments of one-parameter families.

4. Higher moments

Often, calculating higher moments is difficult because it measures a more subtle distribution of a_p values. However, numerically calculating higher moments is no different than the second moment. By Theorem 1 of [Bir68],

$$\mathcal{A}_{2n,\mathcal{E}}(p) \;=\; C_n p^{n+1} + O\left(p^{n+1/2}\right) \tag{4.1}$$

where $C_n = \frac{(2n)!}{n!(n+1)!}$ denotes the n^{th} Catalan number. Like before, we can investigate

$$\mathcal{B}_{2n,\mathcal{E}}(p) \;:=\; \frac{(\mathcal{A}_{2n,\mathcal{E}}(p)/C_n) - p^{n+1}}{p^{n+1/2}}. \tag{4.2}$$

and with our database it is both no harder to do this investigation than the original second moment and we can investigate the first m moments simultaneously.

Acknowledgements

This research was completed at the Williams College SMALL REU Program and was supported by Williams College and the National Science Foundation (grant DMS2241623). The authors are grateful for the support of Duke University, Princeton University, and the University of Michigan. The authors would like to thank Adam Logan for helpful discussions, and the participants of the workshop at Stony Brook.

References

[Asa+23] M. Asada, R. C. Chen, E. Fourakis, Y. H. Kim, A. Kwon, D. J. Lichtman, B. Mackall, S. J. Miller, E. Winsor, K. Winsor, J. Yang, and K. Yang. "Lower-order biases in the second moment of Dirichlet coefficients in families of L-functions". In: *Exp. Math.* 32.3 (2023), pp. 431–456. DOI: 10.1080/10586458.2021.1980453.

[Bat+24] Z. Batterman, A. Jambhale, S. J. Miller, A. L. Narayanan, K. Sharma, A. Yang, and C. Yao. *Applications of Moments of Dirichlet Coefficients in Elliptic Curve Families*. To appear in the ICERM Conference Proceedings for the July 2023 Murmurations Workshop. 2024. arXiv: 2311.17215 [math.NT].

[Bir68] B. J. Birch. "How the Number of Points of An Elliptic Curve Over a Fixed Prime Field Varies". In: *Journal of the London Mathematical Society* s1-43.1 (1968), pp. 57–60. DOI: https://doi.org/10.1112/jlms/s1-43.1.57.

[Boo+24] Andrew R. Booker, Min Lee, David Lowry-Duda, Andrei Seymour-Howell, and Nina Zubrilina. *Murmurations of Maass forms*. 2024. arXiv: 2409.00765 [math.NT]. URL: https://arxiv.org/abs/2409.00765.

[Cow23] Alex Cowan. *Murmurations and explicit formulas*. 2023. arXiv: 2306.10425 [math.NT]. URL: https://arxiv.org/abs/2306.10425.

[He+24] Yang-Hui He, Kyu-Hwan Lee, Thomas Oliver, and Alexey Pozdnyakov. "Murmurations of Elliptic Curves". In: *Experimental Mathematics* (Aug. 2024), pp. 1–13. ISSN: 1944-950X. DOI: 10.1080/10586458.2024.2382361. URL: http://dx.doi.org/10.1080/10586458.2024.2382361.

[HLO22a] Yang-Hui He, Kyu-Hwan Lee, and Thomas Oliver. "Machine-learning number fields". In: *International Press* 2 (2022), pp. 49–66. DOI: https://dx.doi.org/10.4310/MCGD.2022.v2.n1.a2. URL: https://archive.intlpress.com/site/pub/pages/journals/items/mcgd/content/vols/0002/0001/a002/index.php?mode=ns.

[HLO22b] Yang-Hui He, Kyu-Hwan Lee, and Thomas Oliver. "Machine-learning the Sato–Tate conjecture". In: *Journal of Symbolic Computation* 111 (2022), pp. 61–72. ISSN: 0747-7171. DOI: https://doi.org/10.1016/j.jsc.2021.11.002. URL: https://www.sciencedirect.com/science/article/pii/S0747717121000729.

[HLO23] Yang-Hui He, Kyu-Hwan Lee, and Thomas Oliver. "Machine learning invariants of arithmetic curves". In: *Journal of Symbolic Computation* 115 (2023), pp. 478–491. ISSN: 0747-7171. DOI: https://doi.org/10.1016/j.jsc.2022.08.017. URL: https://www.sciencedirect.com/science/article/pii/S0747717122000839.

[KN21] M. Kazalicki and B. Naskrecki. *Second moments and the bias conjecture for the family of cubic pencils*. 2021. arXiv: 2012.11306 [math.NT].

[KN22] M. Kazalicki and B. Naskrecki. "Diophantine triples and K3 surfaces". In: *J. Number Theory* 236 (2022), pp. 41–70. ISSN: 0022-314X. DOI: 10.1016/j.jnt.2021.07.009.

[Mac+16] B. Mackall, S. J. Miller, C. Repti, and K. Winsor. "Lower-order biases in elliptic curve Fourier coefficients in families, in: Frobenius distributions: Lang-Trotter and Sato-Tate conjectures". In: *Contemp. Math., Amer. math. Soc.* 663 (2016), pp. 223–238.

[Mic95] P. Michel. "Rang moyen de famille de courbes elliptiques et lois de Sato-Tate". In: *Monatshefte für Mathematik* 120 (1995), pp. 127–136. DOI: 10.1007/BF01585913.

[Mil02] S. J. Miller. "1- and 2-level densities for families of elliptic curves: evidence for the underlying group symmetries". PhD thesis. Princeton University, 2002.

[Mil04] S. J. Miller. "1- and 2-level densities for families of elliptic curves: evidence for the underlying group symmetries". In: *Compositio Mathematica* 140 (2004), pp. 952–992.

[Mil05] S. J. Miller. "Variation in the number of points on elliptic curves and applications to excess rank". In: *C. R. Math. Acad. Sci. Soc. R. Can.* 27.4 (2005), pp. 111–120.

[Nag97] K. Nagao. "$\mathbb{Q}(T)$-rank of elliptic curves and certain limit coming from the local points". In: *Manuscripta Math* 92 (1997), pp. 13–32. DOI: doi:10.1007/BF02678178.

[RS98] M. Rosen and J. Silverman. "On The Rank Of An Elliptic Surface". In: *Inventiones Mathematicae* 133 (May 1998), pp. 43–67. DOI: 10.1007/s002220050238.

[Sil09] J. H. Silverman. *The Arithmetic of Elliptic Curves (Graduate Texts in Mathematics)*. Springer-Verlag New York, Incorporated, 2009.

[Sil94] J. H. Silverman. *Advanced Topics in the Arithmetic of Elliptic Curves (Graduate Texts in Mathematics)*. Springer, 1994.

[Zub23] Nina Zubrilina. *Murmurations*. 2023. arXiv: 2310.07681 [math.NT]. URL: https://arxiv.org/abs/2310.07681.

Department of Mathematics, University of Michigan
Email address: timcheek@umich.edu

Department of Mathematics, University of California Santa Barbara
Email address: picogilman@gmail.com

Department of Mathematics, Princeton University
Email address: kj5388@princeton.edu

Department of Mathematics, Williams College
Email address: sjm1@williams.edu

Department of Mathematics, Yale University
Email address: vismay.sharan@yale.edu

Department of Mathematics, Duke University
Email address: mariehelene.tome@duke.edu