

Biases in Second Moments of Elliptic Curves

Zoë Batterman (zxba2020@mymail.pomona.edu)
Aditya Jambhale (aj644@cam.ac.uk)

(joint with Akash L. Narayanan, Kishan Sharma, Andrew Yang, and Chris Yao)

Advisor: Steven J. Miller
SMALL REU at Williams College

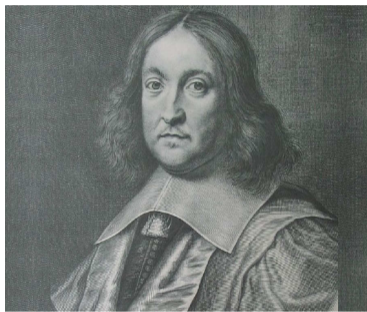
2023 Young Mathematicians Conference at Ohio State University

August 16, 2023

- Why Elliptic Curves?
- Definitions
- Bias conjecture
- Explicit formulae
- Counterexample

Modularity theorem for semistable elliptic curves (Andrew Wiles, 1995).

Andrew Wiles proved that elliptic curves over the field of rational numbers \mathbb{Q} are related to modular forms.



Corollary. (Fermat's Last Theorem, 1637)

No three positive integers a , b , and c can satisfy the equation

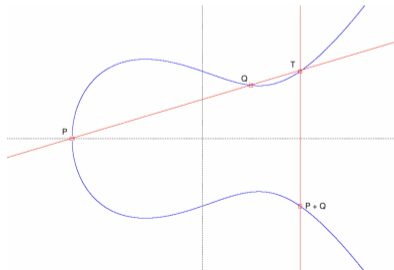
$$a^n + b^n = c^n, \quad n \in \mathbb{N}_{\geq 3}.$$

An **elliptic curve** \mathcal{E} is a non-singular curve of genus 1 of the form $y^2 = x^3 + Ax + B$ where $A, B \in \mathbb{C}$. We may consider the set $\mathcal{E}(\mathbb{Q})$ of rational solutions of \mathcal{E} plus the point at infinity $O_{\mathcal{E}}$.

Theorem. (Mordell-Weil, 1922)

Let P, Q , and $P * Q$ be points on \mathcal{E} which lie on a line. Then the binary operation $P \cdot Q = (P * Q) * O_{\mathcal{E}}$ turns $(\mathcal{E}(\mathbb{Q}), \cdot)$ into a finitely generated abelian group. In particular,

$$\mathcal{E}(\mathbb{Q}) \cong \mathcal{E}(\mathbb{Q})_{\text{torsion}} \oplus \mathbb{Z}^{\text{rank}}$$



Motivating question

Can we find an elliptic curve of large rank?

Also known as twenty septendecillion sixty-seven sexdecillion seven hundred sixty-two quindecillion four hundred fifteen quattuordecillion five hundred seventy-five tredecillion five hundred twenty-six duodecillion five hundred eighty-five undecillion thirty-three decillion two hundred eight nonillion two hundred nine octillion three hundred thirty-eight septillion five hundred forty-two sextillion seven hundred fifty quintillion nine hundred thirty quadrillion two hundred thirty trillion three hundred twelve billion one hundred seventy-eight million nine hundred fifty-six thousand five hundred two

Motivating question

Can we find an elliptic curve of large rank?

In 2006, Noam Elkies set the record by finding an elliptic curve of rank at least 28:

$$y^2 + xy + y = x^3 - x^2$$

– 20067762415575526585033208209338542750930230312178956502x

+ 34481611795030556467032985690390720374855944359319180361266008296291939448732243429

Also known as twenty septendecillion sixty-seven sexdecillion seven hundred sixty-two quindecillion four hundred fifteen quattuordecillion five hundred seventy-five tredecillion five hundred twenty-six duodecillion five hundred eighty-five undecillion thirty-three decillion two hundred eight nonillion two hundred nine octillion three hundred thirty-eight septillion five hundred forty-two sextillion seven hundred fifty quintillion nine hundred thirty quadrillion two hundred thirty trillion three hundred twelve billion one hundred seventy-eight million nine hundred fifty-six thousand five hundred two

- Why Elliptic Curves?
- Definitions
- Bias conjecture
- Explicit formulae
- Counterexample

A *one-parameter family* of elliptic curves is given by

$$\mathcal{E} : y^2 = x^3 + A(T)x + B(T),$$

where $A(T)$, $B(T)$ are polynomials in $\mathbb{Z}[T]$.

Each specialization of T to an integer t gives an elliptic curve E_t over \mathbb{Q} .

A *one-parameter family* of elliptic curves is given by

$$\mathcal{E} : y^2 = x^3 + A(T)x + B(T),$$

where $A(T)$, $B(T)$ are polynomials in $\mathbb{Z}[T]$.

Each specialization of T to an integer t gives an elliptic curve E_t over \mathbb{Q} .

Moments of a family of elliptic curves

The r^{th} moment (note we do not normalize by $1/p$) is

$$\mathcal{A}_{r,\mathcal{E}}(p) = \sum_{t \in \mathbb{F}_p} a_{E_t}(p)^r,$$

where $a_{E_t}(p) = p + 1 - \#(\text{solutions to } E_t \text{ mod } p)$ is the Frobenius trace of E_t .

The first moment is related to the rank of the elliptic curve family:

$\mathcal{A}_{1,\mathcal{E}}(p)$ and Family Rank (Nagao, Rosen-Silverman, 1998)

Given certain technical assumptions (Tate's Conjecture) hold for \mathcal{E} , then

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

The first moment is related to the rank of the elliptic curve family:

$\mathcal{A}_{1,\mathcal{E}}(p)$ and Family Rank (Nagao, Rosen-Silverman, 1998)

Given certain technical assumptions (Tate's Conjecture) hold for \mathcal{E} , then

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

- By $\sum_{p \leq x} \log p \sim x$, if $\mathcal{A}_{1,\mathcal{E}(t)}(p) = -rp + O(1)$, then $\text{rank } \mathcal{E}(\mathbb{Q}(T)) = r$.

The first moment is related to the rank of the elliptic curve family:

$\mathcal{A}_{1,\mathcal{E}}(p)$ and Family Rank (Nagao, Rosen-Silverman, 1998)

Given certain technical assumptions (Tate's Conjecture) hold for \mathcal{E} , then

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

- By $\sum_{p \leq x} \log p \sim x$, if $\mathcal{A}_{1,\mathcal{E}(t)}(p) = -rp + O(1)$, then $\text{rank } \mathcal{E}(\mathbb{Q}(T)) = r$.
- The “rank” of the family means that except for finitely many t , the elliptic curve E_t has rank greater or equal to r .

- Why Elliptic Curves?
- Definitions
- Bias conjecture
- Explicit formulae
- Counterexample

The $j(T)$ -invariant is $j(T) = 1728 \frac{4A(T)^3}{4A(T)^3 + 27B(T)^2}$.

Second moment asymptotic (Michel, 1995)

For a one-parameter family \mathcal{E} with $j(T)$ -invariant non-constant, the second moment is

$$A_{2,\mathcal{E}} = p^2 + O(p^{3/2}),$$

with lower-order terms of size $p^{3/2}$, p , $p^{1/2}$, and 1.

The $j(T)$ -invariant is $j(T) = 1728 \frac{4A(T)^3}{4A(T)^3 + 27B(T)^2}$.

Second moment asymptotic (Michel, 1995)

For a one-parameter family \mathcal{E} with $j(T)$ -invariant non-constant, the second moment is

$$A_{2,\mathcal{E}} = p^2 + O(p^{3/2}),$$

with lower-order terms of size $p^{3/2}$, p , $p^{1/2}$, and 1.

Strong and Weak Bias conjecture

- **Weak:** The largest lower term in the second moment expansion which does not average to 0 is on average **negative**.
- **Strong:** The largest lower term in the second moment expansion which does not average to 0 is **negative except for finitely many p**

Relation with Excess Rank

- If we have lower order negative bias, then the bound for the average rank in families increases.

Relation with Excess Rank

- If we have lower order negative bias, then the bound for the average rank in families increases.
- However, lower order negative biases increases bound only by a small amount, which is not enough to explain observed excess rank.

- Why Elliptic Curves?
- Definitions
- Bias conjecture
- Explicit formulae
- Counterexample

For a specialization $E_t : y^2 = x^3 + A(t)x + B(t)$, we may write

$$a_{E_t}(p) = - \sum_{x \in \mathbb{F}_p} \left(\frac{x^3 + A(t)x + B(t)}{p} \right)$$

where $\left(\frac{\cdot}{p}\right)$ is the Legendre symbol mod p given by

$$\left(\frac{x}{p}\right) = \begin{cases} 1 & x \text{ a non-zero square modulo } p, \\ 0 & x \equiv 0 \pmod{p}, \\ -1 & \text{otherwise.} \end{cases}$$

For a specialization $E_t : y^2 = x^3 + A(t)x + B(t)$, we may write

$$a_{E_t}(p) = - \sum_{x \in \mathbb{F}_p} \left(\frac{x^3 + A(t)x + B(t)}{p} \right)$$

where $\left(\frac{\cdot}{p}\right)$ is the Legendre symbol mod p given by

$$\left(\frac{x}{p}\right) = \begin{cases} 1 & x \text{ a non-zero square modulo } p, \\ 0 & x \equiv 0 \pmod{p}, \\ -1 & \text{otherwise.} \end{cases}$$

Observe that $\left(\frac{x}{p}\right) + 1$ is precisely the number of solutions to $x = y^2 \pmod{p}$.

Linear and quadratic Legendre sums

We have the following

$$\sum_{x(p)} \left(\frac{ax + b}{p} \right) = 0 \quad p \nmid a,$$

$$\sum_{x(p)} \left(\frac{ax^2 + bx + c}{p} \right) = \begin{cases} - \left(\frac{a}{p} \right) & p \nmid b^2 - 4ac, \\ (p - 1) \left(\frac{a}{p} \right) & p \mid b^2 - 4ac. \end{cases}$$

Linear and quadratic Legendre sums

We have the following

$$\sum_{x(p)} \left(\frac{ax + b}{p} \right) = 0 \quad p \nmid a,$$

$$\sum_{x(p)} \left(\frac{ax^2 + bx + c}{p} \right) = \begin{cases} -\left(\frac{a}{p}\right) & p \nmid b^2 - 4ac, \\ (p-1) \left(\frac{a}{p}\right) & p \mid b^2 - 4ac. \end{cases}$$

Average values of Legendre symbols

Taking the limit of the average of the Legendre symbol over all primes gives

$$\lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \sum_{p \leq x} \left(\frac{x}{p} \right) = \begin{cases} 1 & x \text{ a non-zero square,} \\ 0 & \text{otherwise.} \end{cases}$$

- The moments become intractable when $A(T)$ and $B(T)$ have high degree.
- For the following special families, the following is known:

Family	$A_{1,\varepsilon}(p)$	$A_{2,\varepsilon}(p)$
$y^2 = x^3 + 2^4(-3)^3(9T + 1)^2$	0	$\begin{cases} 2p^2 - 2p & p \equiv 2 \pmod{3} \\ 0 & p \equiv 1 \pmod{3} \end{cases}$
$y^2 = x^3 \pm 4(4T + 2)x$	0	$\begin{cases} 2p^2 - 2p & p \equiv 1 \pmod{4} \\ 0 & p \equiv 3 \pmod{4} \end{cases}$
$y^2 = x^3 + (T + 1)x^2 + Tx$	0	$p^2 - 2p - 1$
$y^2 = x^3 + x^2 + 2T + 1$	0	$p^2 - 2p - -3$
$y^2 = x^3 + Tx^2 + 1$	$-p$	$p^2 - n_{3,2,p}p - 1 + c_{3/2}(p)$
$y^2 = x^3 - T^2x + T^2$	$-2p$	$p^2 - p - c_1(p) - c_0(p)$
$y^2 = x^3 - T^2x + T^4$	$-2p$	$p^2 - p - c_1(p) - c_0(p)$
$y^2 = x^3 + Tx^2 - (T + 3)x + 1$	$-2c_{p,1;4}p$	$p^2 - 4c_{p,1;6}p - 1$

where $c_{p,a;m} = 1$ if $p \equiv a \pmod{m}$ and 0 otherwise; $n_{3,2,p}$ is the number of cubes roots of 2 mod p ; $c_\alpha(p)$ are certain legendre sums multiplied by p .

Example

Consider $\mathcal{F} : y^2 = x^3 - T^2x + T^4$. Then the first moment is

$$\begin{aligned}\mathcal{A}_{1,\mathcal{F}}(p) &= \sum_{T \in \mathbb{F}_p} a_{p,E_t} \\ &= - \sum_{t \in \mathbb{F}_p} \sum_{x \in \mathbb{F}_p} \left(\frac{x^3 - t^2x + t^4}{p} \right).\end{aligned}$$

In general, quartic Legendre sums are intractable.

Consider $\mathcal{F} : y^2 = x^3 - T^2x + T^4$. Then the first moment is

$$\begin{aligned} \mathcal{A}_{1,\mathcal{F}}(p) &= \sum_{T \in \mathbb{F}_p} a_{p,E_t} \\ &= - \sum_{t \in \mathbb{F}_p} \sum_{x \in \mathbb{F}_p} \left(\frac{x^3 - t^2x + t^4}{p} \right). \end{aligned}$$

In general, quartic Legendre sums are intractible.

But we may apply the clever [substitution](#) $x \mapsto tx$ which gives

$$\begin{aligned} &= - \sum_{x \in \mathbb{F}_p} \left(\frac{x^3}{p} \right) - \sum_{t \neq 0(p)} \sum_{x \in \mathbb{F}_p} \left(\frac{t^3x^3 - t^3x + t^4}{p} \right) \\ &= - \sum_{t \neq 0 \in \mathbb{F}_p} \sum_{x \in \mathbb{F}_p} \left(\frac{tx^3 - tx + t^2}{p} \right). \end{aligned}$$

So, we obtained a closed-form expression.

- Why Elliptic Curves?
- Definitions
- Bias conjecture
- Explicit formulae
- Counterexample

- We computationally evaluated second moments of various families of elliptic curves.
- By Michel's theorem, we assume that

$$\mathcal{A}_{2,\varepsilon}(p) = p^2 + \alpha(p)p^{3/2} + \beta(p)p + O(p^{1/2})$$

where $\alpha(p)$ and $\beta(p)$ are $O(1)$. To investigate the $\alpha(p)$ coefficient, we graphed the *bias* of the second moment

- We computationally evaluated second moments of various families of elliptic curves.
- By Michel's theorem, we assume that

$$\mathcal{A}_{2,\varepsilon}(p) = p^2 + \alpha(p)p^{3/2} + \beta(p)p + O(p^{1/2})$$

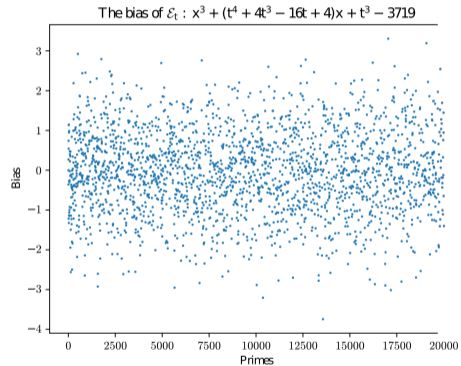
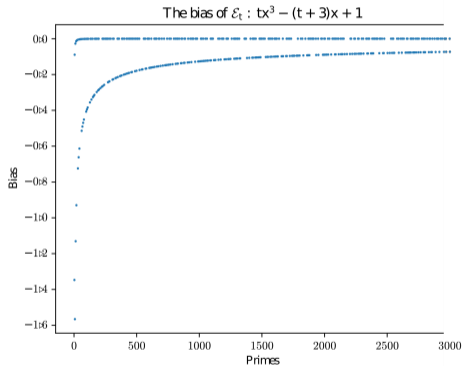
where $\alpha(p)$ and $\beta(p)$ are $O(1)$. To investigate the $\alpha(p)$ coefficient, we graphed the *bias* of the second moment

Bias

We compute the *bias* of $\mathcal{A}_{2,\varepsilon}$ defined by

$$\mathcal{B}_\varepsilon(p) = \frac{\mathcal{A}_{2,\varepsilon} - p^2}{p^{3/2}}.$$

Here are two examples for the graph of the biases, one for a tractable family, and one for not



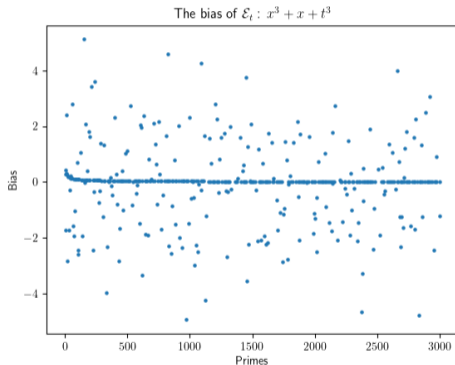
Eventually, we found the family

$$\mathcal{F} : y^2 = x^3 + x + T^3.$$

Eventually, we found the family

$$\mathcal{F} : y^2 = x^3 + x + T^3.$$

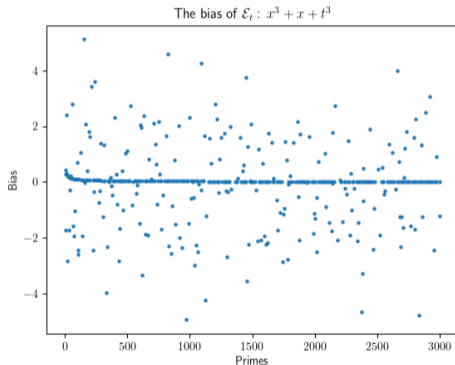
Our reason for suspecting this family was the graph of the bias:



Eventually, we found the family

$$\mathcal{F} : y^2 = x^3 + x + T^3.$$

Our reason for suspecting this family was the graph of the bias:



The graph indicates a clear line where the bias is positive, compared to the graphs in the previous slides.

The Counterexample

Consider the family

$$\mathcal{F}: y^2 = x^3 + x + T^3.$$

Notice for primes p such that $3 \nmid p$, we have $T \mapsto T^3$ a bijection.

Consider the family

$$\mathcal{F}: y^2 = x^3 + x + T^3.$$

Notice for primes p such that $3 \nmid p$, we have $T \mapsto T^3$ a bijection.

So, we sample the simpler elliptic curve family

$$\tilde{\mathcal{F}}: y^2 = x^3 + x + T$$

when $p \equiv 2 \pmod{3}$, which is half of the primes!

Consider the family

$$\mathcal{F}: y^2 = x^3 + x + T^3.$$

Notice for primes p such that $3 \nmid p$, we have $T \mapsto T^3$ a bijection.

So, we sample the simpler elliptic curve family

$$\tilde{\mathcal{F}}: y^2 = x^3 + x + T$$

when $p \equiv 2 \pmod{3}$, which is half of the primes! This immediately gives us that for such primes,

$$\begin{aligned} \mathcal{A}_{2,\mathcal{F}}(p) &= \sum_{t \in \mathbb{F}_p} \sum_{x,y \in \mathbb{F}_p} \left(\frac{x^3 + x + t^3}{p} \right) \left(\frac{y^3 + y + t^3}{p} \right) \\ &= \sum_{t \in \mathbb{F}_p} \sum_{x,y \in \mathbb{F}_p} \left(\frac{x^3 + x + t}{p} \right) \left(\frac{y^3 + y + t}{p} \right) \\ &= \mathcal{A}_{2,\tilde{\mathcal{F}}}(p) = p^2 - \left(\frac{-3}{p} \right) p = p^2 + p. \end{aligned}$$

Consider the family

$$\mathcal{F}: y^2 = x^3 + x + T^3.$$

Notice for primes p such that $3 \nmid p$, we have $T \mapsto T^3$ a bijection.

So, we sample the simpler elliptic curve family

$$\tilde{\mathcal{F}}: y^2 = x^3 + x + T$$

when $p \equiv 2 \pmod{3}$, which is half of the primes! This immediately gives us that for such primes,

$$\begin{aligned} \mathcal{A}_{2,\mathcal{F}}(p) &= \sum_{t \in \mathbb{F}_p} \sum_{x,y \in \mathbb{F}_p} \left(\frac{x^3 + x + t^3}{p} \right) \left(\frac{y^3 + y + t^3}{p} \right) \\ &= \sum_{t \in \mathbb{F}_p} \sum_{x,y \in \mathbb{F}_p} \left(\frac{x^3 + x + t}{p} \right) \left(\frac{y^3 + y + t}{p} \right) \\ &= \mathcal{A}_{2,\tilde{\mathcal{F}}}(p) = p^2 - \left(\frac{-3}{p} \right) p = p^2 + p. \end{aligned}$$

Bias Revisited

We graph the *bias* of $\mathcal{A}_{2,\varepsilon}$, for calculated values, defined by

$$\mathcal{B}_\varepsilon(p) = \frac{\mathcal{A}_{2,\varepsilon} - p^2}{p^{3/2}}.$$

Recall by Michel's theorem, we have

$$\mathcal{A}_{2,\varepsilon}(p) = p^2 + \alpha(p)p^{3/2} + \beta(p)p + O(p^{1/2})$$

where $\alpha(p)$ and $\beta(p)$ are $O(1)$. To disprove the *weak bias conjecture*, we do two things:

Bias Revisited

We graph the *bias* of $\mathcal{A}_{2,\varepsilon}$, for calculated values, defined by

$$\mathcal{B}_\varepsilon(p) = \frac{\mathcal{A}_{2,\varepsilon} - p^2}{p^{3/2}}.$$

Recall by Michel's theorem, we have

$$\mathcal{A}_{2,\varepsilon}(p) = p^2 + \alpha(p)p^{3/2} + \beta(p)p + O(p^{1/2})$$

where $\alpha(p)$ and $\beta(p)$ are $O(1)$. To disprove the *weak bias conjecture*, we do two things:

- Show that $\alpha(p)$ averages to 0, i.e.

$$\lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \sum_{p \leq x} \alpha(p) = 0.$$

Bias Revisited

We graph the *bias* of $\mathcal{A}_{2,\varepsilon}$, for calculated values, defined by

$$\mathcal{B}_\varepsilon(p) = \frac{\mathcal{A}_{2,\varepsilon} - p^2}{p^{3/2}}.$$

Recall by Michel's theorem, we have

$$\mathcal{A}_{2,\varepsilon}(p) = p^2 + \alpha(p)p^{3/2} + \beta(p)p + O(p^{1/2})$$

where $\alpha(p)$ and $\beta(p)$ are $O(1)$. To disprove the *weak bias conjecture*, we do two things:

- Show that $\alpha(p)$ averages to 0, i.e.

$$\lim_{x \rightarrow \infty} \frac{1}{\pi(x)} \sum_{p \leq x} \alpha(p) = 0.$$

- Show that $\beta(p)$ averages to a positive number.

By the prime number theorem, one shows

$$\frac{1}{\pi(x)} \sum_{p \leq x} \alpha(p) = \frac{1}{\pi(x)} \sum_{p \leq x} \mathcal{B}_\varepsilon(p) + O(x^{-1/2} \log x)$$

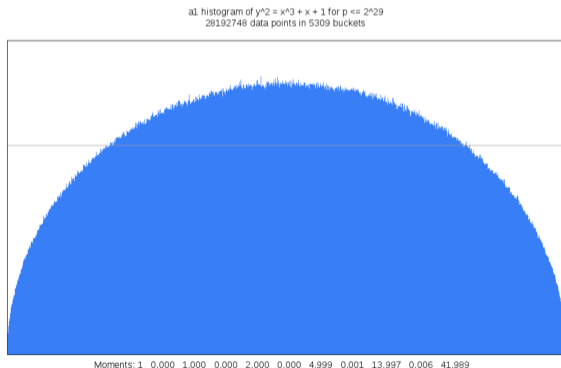
Problem: The constant in the big O term might dominate.

By the prime number theorem, one shows

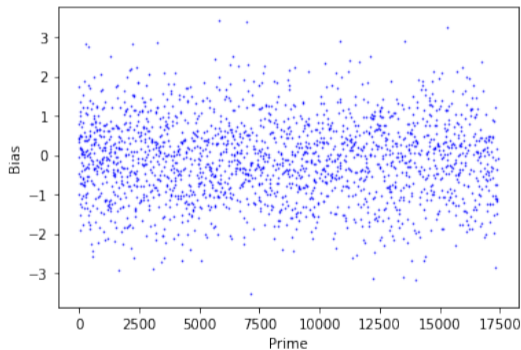
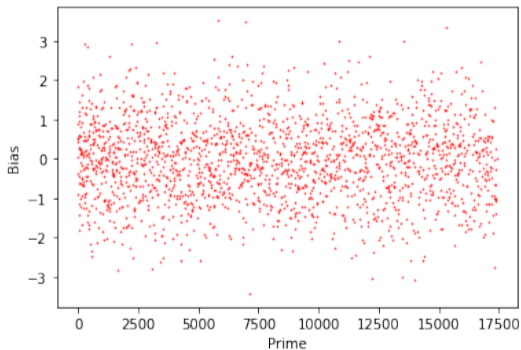
$$\frac{1}{\pi(x)} \sum_{p \leq x} \alpha(p) = \frac{1}{\pi(x)} \sum_{p \leq x} \mathcal{B}_{\mathcal{E}}(p) + O(x^{-1/2} \log x)$$

Problem: The constant in the big O term might dominate.

Solution: Randomly simulate elliptic moments using the *Sato-Tate distribution*.



The following are two graphs which randomly simulate the bias.
One graph has coefficient $\alpha(p) = -0.1$ and the other has $\alpha(p) = 0$.
Can you guess which is which?



Taking the running average of the biases, it is clear there is a bias:

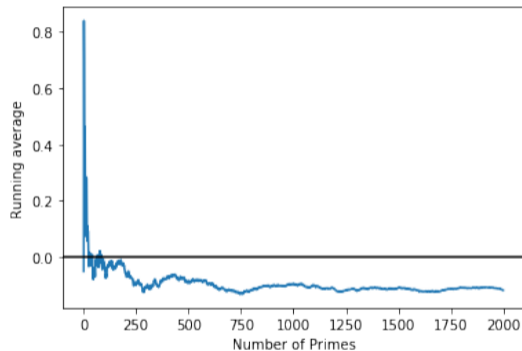
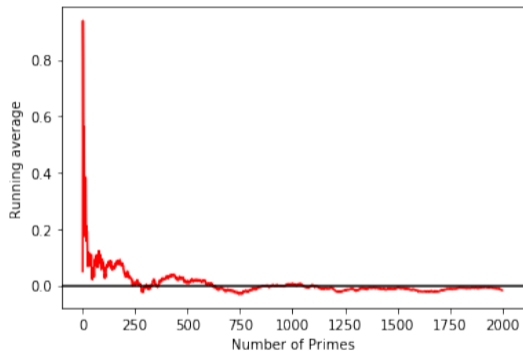
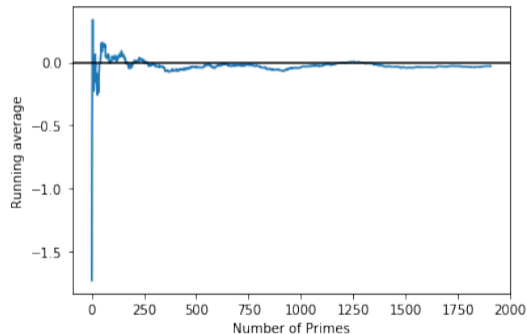
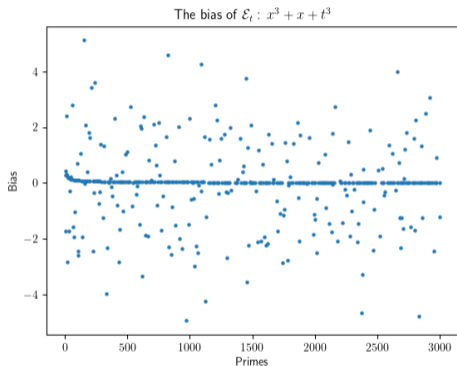


Figure: Unbiased Running Averages (Red) versus Biased Running Averages (Blue) for a random simulation

Doing the same with our family of interest, that is, $y^2 = x^3 + x + t^3$, we get



So we have strong computational evidence the largest term averages to 0.

Special thanks to Professor Steven J. Miller and the Churchill Foundation.

Thanks to our SMALL 2023 faculty, research assistants, and peers for their support.

Thanks to the National Science Foundation for making SMALL 2023 possible.

The conjectured first moment of $y^2 = x^3 + x + t^3$

The first moment $\mathcal{A}_{1,p}$ satisfies

$$|\mathcal{A}_{1,p}| = \begin{cases} 4p & p \text{ is of the form } a^2 + 36b^2, \\ 0 & \text{otherwise.} \end{cases}$$

The conjectured first moment of $y^2 = x^3 + x + t^3$

The first moment $\mathcal{A}_{1,p}$ satisfies

$$|\mathcal{A}_{1,p}| = \begin{cases} 4p & p \text{ is of the form } a^2 + 36b^2, \\ 0 & \text{otherwise.} \end{cases}$$

For a prime $p \not\equiv 1(12)$, the Chinese remainder theorem in conjunction with the changes of variable

$$t \mapsto tx, \quad \text{and} \quad t \mapsto t^3 \quad \implies \quad \mathcal{A}_{1,p} = 0.$$

The conjectured first moment of $y^2 = x^3 + x + t^3$

The first moment $\mathcal{A}_{1,p}$ satisfies

$$|\mathcal{A}_{1,p}| = \begin{cases} 4p & p \text{ is of the form } a^2 + 36b^2, \\ 0 & \text{otherwise.} \end{cases}$$

For a prime $p \not\equiv 1(12)$, the Chinese remainder theorem in conjunction with the changes of variable

$$t \mapsto tx, \quad \text{and} \quad t \mapsto t^3 \implies \mathcal{A}_{1,p} = 0.$$

Using binary quadratic forms, $\mathcal{A}_{1,p} \neq 0$ forces p to be of the form

$$p = a^2 + 36b^2 \quad \text{or} \quad p = 4a^2 + 9b^2.$$

We computationally found $|\mathcal{A}_{1,p}| = 4p$ in the former and $\mathcal{A}_{1,p} = 0$ in the latter.