# VaNTAGe

a virtual math seminar on open conjectures in

number theory and arithmetic geometry

# The Distribution of Ranks of Elliptic Curves and the Minimalist Conjecture

Reconciling Conjectures and Data

Álvaro Lozano-Robledo

University of Connecticut

# Goal

Our goal is to understand the possible structures of $E(\mathbb{Q})$, for an elliptic curve $E/\mathbb{Q}$.

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^{R_{E/\mathbb{Q}}}$$


Donald Anderson, first poster child.

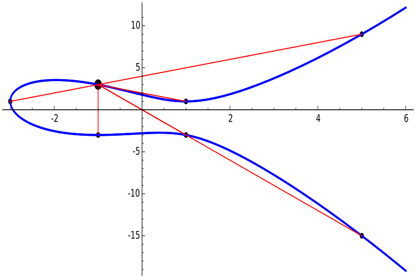
The torsion subgroups over $\mathbb{Q}$ are the "poster child" of what an arithmetic group should be like. Torsion subgroups are:

- Computable
- Classified
- Parametrized in families
- Statistically understood

# Goal

Our goal is to understand the possible structures of $E(\mathbb{Q})$, for an elliptic curve $E/\mathbb{Q}$.

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^{R_{E/\mathbb{Q}}}$$



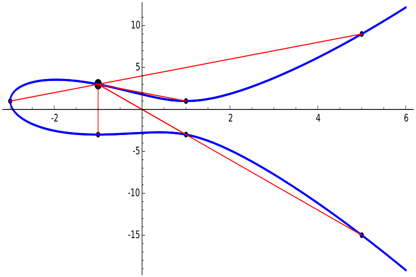

The curve $E : y^2 + xy + y = x^3 + x^2 - 4x + 5$ (42.a5) has torsion subgroup $\langle(-1, 3)\rangle \cong \mathbb{Z}/8\mathbb{Z}$.

- **Computable**
  - Nagell–Lutz theorem.
  - Division polynomials.

# Goal

Our goal is to understand the possible structures of $E(\mathbb{Q})$, for an elliptic curve $E/\mathbb{Q}$.

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^{R_{E/\mathbb{Q}}}$$




The curve $E : y^2 + xy + y = x^3 + x^2 - 4x + 5$ (`42.a5`) has torsion subgroup $\langle(-1, 3)\rangle \cong \mathbb{Z}/8\mathbb{Z}$.

- **Classified**
  - Mazur's theorem:

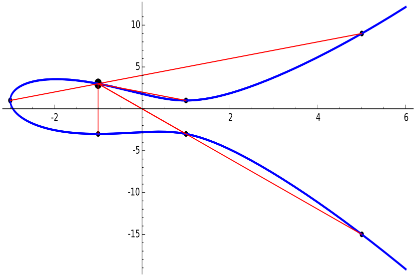
$$E(\mathbb{Q})_{\text{tors}} \simeq \begin{cases} \mathbb{Z}/M\mathbb{Z}, \text{ or} \\ \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}/2N\mathbb{Z} \end{cases}$$

where $1 \leq M \leq 10$ or $M = 12$, and $1 \leq N \leq 4$.

# Goal

Our goal is to understand the possible structures of $E(\mathbb{Q})$, for an elliptic curve $E/\mathbb{Q}$.

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^{R_{E/\mathbb{Q}}}$$




The curve $E : y^2 + xy + y = x^3 + x^2 - 4x + 5$ (42.a5) has torsion subgroup $\langle(-1, 3)\rangle \cong \mathbb{Z}/8\mathbb{Z}$.

- **Parametrized in families**
  - Kubert et al.:

e.g.,
elliptic curves with $\mathbb{Z}/8\mathbb{Z}$ tors.:

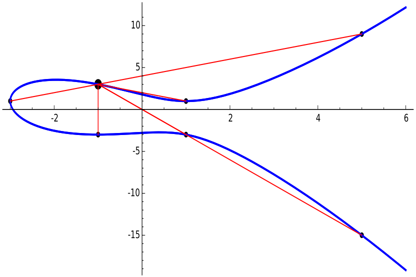
$$E : y^2+(1-a)xy-by^2 = x^3-bx^2$$

with $b = (2t-1)(t-1)$ and $a = b/t$, for any $t \neq 0, 1/2, 1$.

# Goal

Our goal is to understand the possible structures of $E(\mathbb{Q})$, for an elliptic curve $E/\mathbb{Q}$.

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^{R_{E/\mathbb{Q}}}$$




The curve $E' : y^2 + \frac{1}{3}xy - \frac{2}{9}y = x^3 - \frac{2}{9}x^2$ ($\cong_{\mathbb{Q}}$ `42.a5`) has torsion subgroup $\langle(-1,3)\rangle \cong \mathbb{Z}/8\mathbb{Z}$.

- **Parametrized in families**
  - Kubert et al.:

  e.g.,
  elliptic curves with $\mathbb{Z}/8\mathbb{Z}$ tors.:

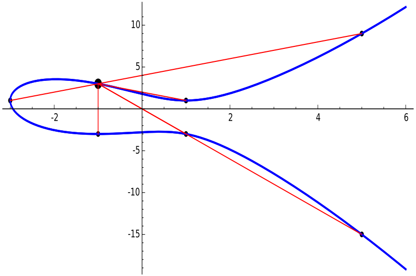
  $E : y^2 + (1-a)xy - by^2 = x^3 - bx^2$

  with $b = (2t-1)(t-1)$ and $a = b/t$, for any $t \neq 0, 1/2, 1$.

# Goal

Our goal is to understand the possible structures of $E(\mathbb{Q})$, for an elliptic curve $E/\mathbb{Q}$.

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^{R_{E/\mathbb{Q}}}$$




The curve $E : y^2 + xy + y = x^3 + x^2 - 4x + 5$ (42.a5) has torsion subgroup $\langle(-1,3)\rangle \cong \mathbb{Z}/8\mathbb{Z}$.

- **Statistically understood**
  - Harron–Snowden (2013):

Let $N_G(X)$ be the number of elliptic curves $E/\mathbb{Q}$ with (naive) height $\leq X$ and $E(\mathbb{Q})_{\text{tors}} \cong G$. Then, there are positive constants $C_1, C_2, d(G)$ such that

$$C_1 X^{d(G)} \leq N_G(X) \leq C_2 X^{d(G)}.$$

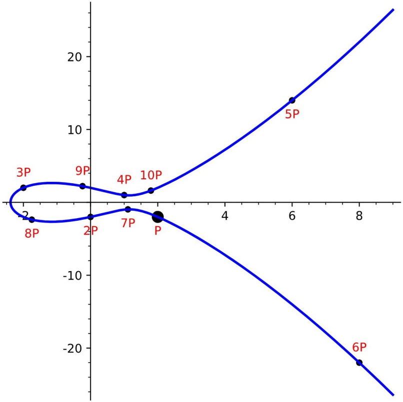
E.g., $d(\{0\}) = 5/6$ and $d(\mathbb{Z}/8\mathbb{Z}) = 1/12$.

* Also see recent similar work by Boggess–Sankar counting elliptic curves with $N$-isogenies!

# Goal

Our goal is to understand the possible structures of $E(\mathbb{Q})$, for an elliptic curve $E/\mathbb{Q}$.

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^{R_{E/\mathbb{Q}}}$$




The curve $E : y^2 = x^3 - 4x + 4$ (`88.a1`) has trivial torsion subgroup and rank 1, with $E(\mathbb{Q}) = \langle (2, -2) \rangle \cong \mathbb{Z}$.

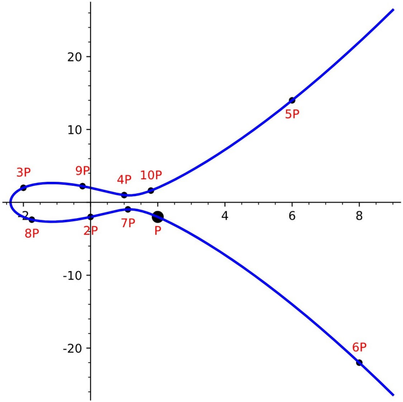
**How about the rank?**

- Computable?
- Classified?
- Parametrized in families?
- Statistically understood?

# Goal

Our goal is to understand the possible structures of $E(\mathbb{Q})$, for an elliptic curve $E/\mathbb{Q}$.

$$E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^{R_{E/\mathbb{Q}}}$$




The curve $E : y^2 = x^3 - 4x + 4$ (`88.a1`) has trivial torsion subgroup and rank 1, with $E(\mathbb{Q}) = \langle(2, -2)\rangle \cong \mathbb{Z}$.

**How about the rank?**

- ~~Computable?~~ Maybe
- ~~Classified?~~ No
- ~~Parametrized in families?~~ No
- ~~Statistically understood?~~ No

# Is the Rank Computable?

- **Analytically?** Yes$^*$, if we assume B–S-D, the rank is the order of vanishing of $L(E, s)$ at $s = 1$. (*Computing values requires $\approx \sqrt{N_E}$ Fourier coefficients, and issues certifying zeroes numerically.)
- **Algebraically?** Yes$^*$, if we assume $Ш(E/\mathbb{Q})[p^\infty]$ is finite, for some prime $p$. (*Computing the rank may involve computing models for high $p$-descendants.)

  Recall:

  $$0 \longrightarrow E(\mathbb{Q})/2E(\mathbb{Q}) \longrightarrow \mathrm{Sel}_2(E/\mathbb{Q}) \longrightarrow Ш(E/\mathbb{Q})[2] \longrightarrow 0,$$

  where $\mathrm{Sel}_n(E/\mathbb{Q})$ is a finite, computable, cohomological group defined by finitely many local conditions.

# Is the Rank Computable?

- **Analytically?** Yes$^*$, if we assume B–S-D, the rank is the order of vanishing of $L(E, s)$ at $s = 1$. ($^*$Computing values requires $\approx \sqrt{N_E}$ Fourier coefficients, and issues certifying zeroes numerically.)

- **Algebraically?** Yes$^*$, if we assume $\text{Ш}(E/\mathbb{Q})[p^\infty]$ is finite, for some prime $p$. ($^*$Computing the rank may involve computing models for high $p$-descendants.)

  Recall:

  $$0 \longrightarrow E(\mathbb{Q})/4E(\mathbb{Q}) \longrightarrow \mathrm{Sel}_4(E/\mathbb{Q}) \longrightarrow \text{Ш}(E/\mathbb{Q})[4] \longrightarrow 0,$$

  where $\mathrm{Sel}_n(E/\mathbb{Q})$ is a finite, computable, cohomological group defined by finitely many local conditions.

# Is the Rank Computable?

- **Analytically?** Yes$^*$, if we assume B–S-D, the rank is the order of vanishing of $L(E, s)$ at $s = 1$. ($^*$Computing values requires $\approx \sqrt{N_E}$ Fourier coefficients, and issues certifying zeroes numerically.)

- **Algebraically?** Yes$^*$, if we assume $Ш(E/\mathbb{Q})[p^\infty]$ is finite, for some prime $p$. ($^*$Computing the rank may involve computing models for high $p$-descendants.)

  Recall:

  $$0 \longrightarrow E(\mathbb{Q})/8E(\mathbb{Q}) \longrightarrow \mathrm{Sel}_8(E/\mathbb{Q}) \longrightarrow Ш(E/\mathbb{Q})[8] \longrightarrow 0,$$

  where $\mathrm{Sel}_n(E/\mathbb{Q})$ is a finite, computable, cohomological group defined by finitely many local conditions.

# Is the Rank Computable?

- **Analytically?** Yes$^*$, if we assume B–S-D, the rank is the order of vanishing of $L(E, s)$ at $s = 1$. ($^*$Computing values requires $\approx \sqrt{N_E}$ Fourier coefficients, and issues certifying zeroes numerically.)

- **Algebraically?** Yes$^*$, if we assume $Ш(E/\mathbb{Q})[p^\infty]$ is finite, for some prime $p$. ($^*$Computing the rank may involve computing models for high $p$-descendants.)

  Recall:

  $$0 \longrightarrow E(\mathbb{Q})/2^m E(\mathbb{Q}) \longrightarrow \mathrm{Sel}_{2^m}(E/\mathbb{Q}) \longrightarrow Ш(E/\mathbb{Q})[2^m] \longrightarrow 0,$$

  where $\mathrm{Sel}_n(E/\mathbb{Q})$ is a finite, computable, cohomological group defined by finitely many local conditions.

# Is the Rank Computable?

- **Analytically?** Yes$^*$, if we assume B–S-D, the rank is the order of vanishing of $L(E, s)$ at $s = 1$. ($^*$Computing values requires $\approx \sqrt{N_E}$ Fourier coefficients, and issues certifying zeroes numerically.)

- **Algebraically?** Yes$^*$, if we assume $Ш(E/\mathbb{Q})[p^\infty]$ is finite, for some prime $p$. ($^*$Computing the rank may involve computing models for high $p$-descendants.)

  Recall:

  $$0 \longrightarrow E(\mathbb{Q})/2^m E(\mathbb{Q}) \longrightarrow \mathrm{Sel}_{2^m}(E/\mathbb{Q}) \longrightarrow Ш(E/\mathbb{Q})[2^\infty] \longrightarrow 0,$$

  where $\mathrm{Sel}_n(E/\mathbb{Q})$ is a finite, computable, cohomological group defined by finitely many local conditions.

# Is the Rank Computable?

- **Analytically?** Yes$^*$, if we assume B–S-D, the rank is the order of vanishing of $L(E, s)$ at $s = 1$. ($^*$Computing values requires $\approx \sqrt{N_E}$ Fourier coefficients, and issues certifying zeroes numerically.)

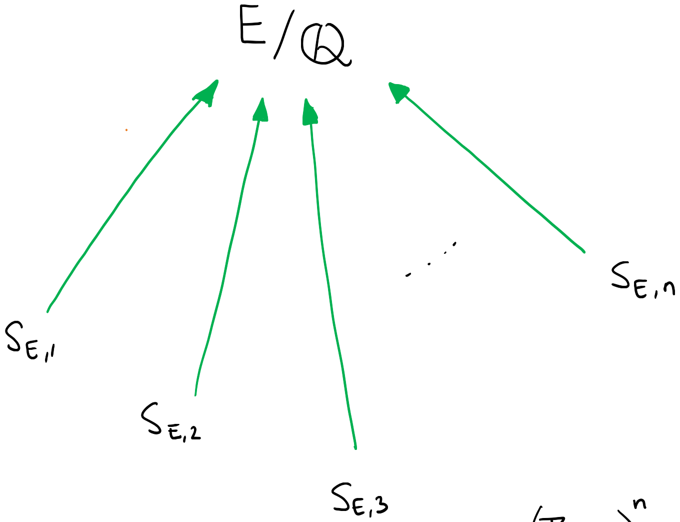
- **Algebraically?** Yes$^*$, if we assume $\text{Ш}(E/\mathbb{Q})[p^\infty]$ is finite, for some prime $p$. ($^*$Computing the rank may involve computing models for high $p$-descendants.)

  Recall:

  $$0 \longrightarrow E(\mathbb{Q})/2^{m+1}E(\mathbb{Q}) \longrightarrow \mathrm{Sel}_{2^{m+1}}(E/\mathbb{Q}) \longrightarrow \text{Ш}(E/\mathbb{Q})[2^\infty] \longrightarrow 0,$$

  where $\mathrm{Sel}_n(E/\mathbb{Q})$ is a finite, computable, cohomological group defined by finitely many local conditions.

# 2-Descent




where $\mathrm{Sel}_2(E/\mathbb{Q}) = \langle S_{E,1}, S_{E,2}, \ldots, S_{E,n} \rangle \cong \left(\mathbb{Z}/2\mathbb{Z}\right)^n$

# 2-Descent: Example

$$E/\mathbb{Q}$$

$$y^2 = x^3 - 105196x - 12970320$$

$S_{E,1}$

$$y^2 = 289x^4 + 1020x^2 + 1156$$

$S_{E,2}$

$$y^2 = 4x^4 - 2244x^2 + 289$$

$S_{E,3}$

$$y^2 = 289x^4 + 408x^3 + 1428x^2 + 816x + 1156$$

$S_{E,4}$

$$y^2 = 8x^4 + 306x^2 + 2312$$

$S_{E,5}$

$$y^2 = -8x^4 + 306x^2 - 2312$$

# 2-Descent: Example

$$E/\mathbb{Q}$$

$$y^2 = x^3 - 105196x - 12970320$$

$(0,34) \in S_{E,1}$

$$y^2 = 289x^4 + 1020x^2 + 1156$$

$(0,17) \in S_{E,2}$

$$y^2 = 4x^4 - 2244x^2 + 289$$

$(0,34) \in S_{E,3}$

$$y^2 = 289x^4 + 408x^3 + 1428x^2 + 816x + 1156$$

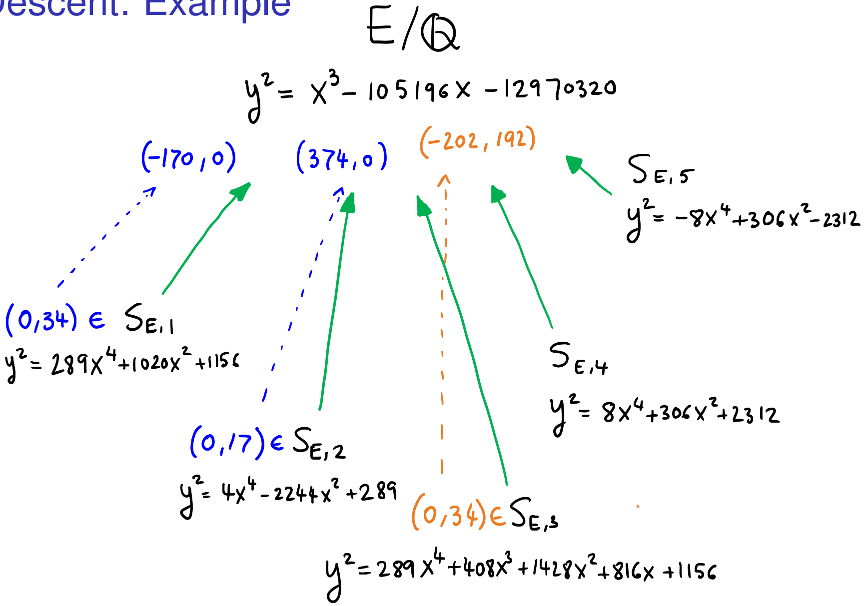
$S_{E,4}$

$$y^2 = 8x^4 + 306x^2 + 2312$$

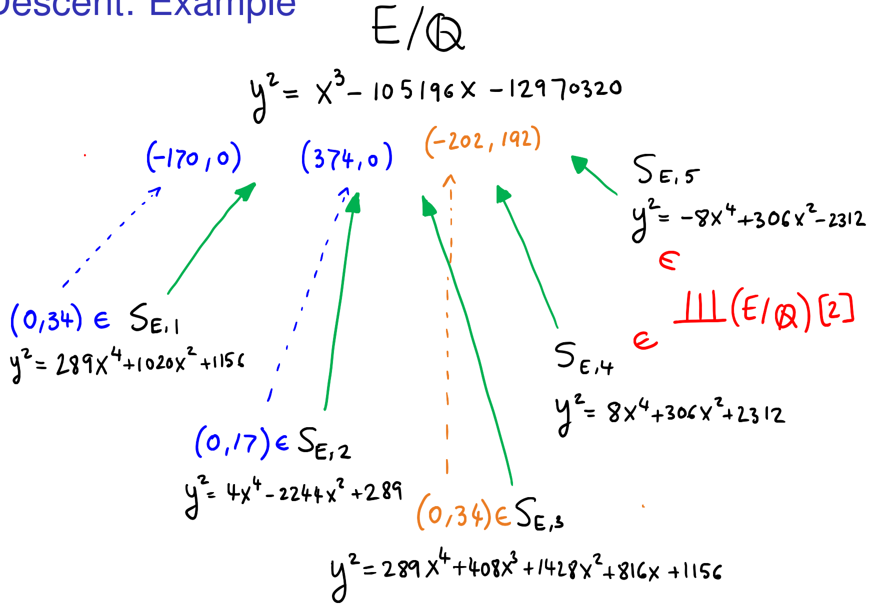
$S_{E,5}$

$$y^2 = -8x^4 + 306x^2 - 2312$$

# 2-Descent: Example



# 2-Descent: Example



# 2-Descent: Example

$E/\mathbb{Q}$

$$y^2 = x^3 - 105196x - 12970320$$

$(-170, 0)$ $(374, 0)$ $(-202, 192)$

$(0,34) \in S_{E,1}$

$$y^2 = 289x^4 + 1020x^2 + 1156$$

$(0,17) \in S_{E,2}$

$$y^2 = 4x^4 - 2244x^2 + 289$$

$(0,34) \in S_{E,3}$

$$y^2 = 289x^4 + 408x^3 + 1428x^2 + 816x + 1156$$

$S_{E,4}$

$$y^2 = 8x^4 + 306x^2 + 2312$$

$S_{E,5}$

$$y^2 = -8x^4 + 306x^2 - 2312$$

$S_{E,4}, S_{E,5} \in Ш(E/\mathbb{Q})[2]$

$$Sel_2(E/\mathbb{Q}) \cong (\mathbb{Z}/2\mathbb{Z})^2 \oplus \mathbb{Z}/2\mathbb{Z} \oplus (\mathbb{Z}/2\mathbb{Z})^2$$

TORSION · MW RANK · SHA

# Rank Statistics?

The average is the simplest statistic... so what is the “average rank”?

- We will consider elliptic curves (up to isomorphism over $\mathbb{Q}$) given by a minimal short Weierstrass model over $\mathbb{Z}$, that is,

$$\mathcal{E} = \{E/\mathbb{Q} : y^2 = x^3 + Ax + B, \text{ with } A, B \in \mathbb{Z}\},$$

  with $4A^3 + 27B^2 \neq 0$, and such that if $d^4|A$ and $d^6|B$, then $d = \pm 1$.
- The naive height of $E \in \mathcal{E}$ is defined by

$$\mathrm{ht}(E) = \max\{4|A|^3, 27B^2\}.$$

- $\mathcal{E}(X) = \{E \in \mathcal{E} : \mathrm{ht}(E) \leq X\}$, all elliptic curves up to height $X$.
- $\pi_{\mathcal{E}}(X) = \#\mathcal{E}(X)$.

# Rank Statistics?

The average is the simplest statistic... so what is the average rank?

- $\mathcal{E} = \{E/\mathbb{Q} : y^2 = x^3 + Ax + B, \text{ with } A, B \in \mathbb{Z}\}$.
- $\text{ht}(E) = \max\{4|A|^3, 27B^2\}$.
- $\mathcal{E}(X) = \{E \in \mathcal{E} : \text{ht}(E) \leq X\}$.
- $\pi_{\mathcal{E}}(X) = \#\mathcal{E}(X)$.

We would like to understand the behavior of

$$\text{AveRank}_{\mathcal{E}}(X) = \frac{\sum_{E \in \mathcal{E}(X)} \text{rank}(E(\mathbb{Q}))}{\pi_{\mathcal{E}}(X)}.$$

# Rank Statistics?

We would like to understand the behavior of

$$\mathrm{AveRank}_{\mathcal{E}}(X) = \frac{\sum_{E \in \mathcal{E}(X)} \mathrm{rank}(E(\mathbb{Q}))}{\pi_{\mathcal{E}}(X)}.$$

1. B–S-D together with the functional equation of $L(E, s)$ implies that the rank *parity* is dictated by the sign of the functional equation (root number). * Progress on B–S-D/parity! (e.g., Kolyvagin, Nekovář, Dokchitser-Dokchitser, etc.).
2. Root numbers are believed to be nicely distributed (because they are defined by local conditions). * Progress! (e.g., Helfgott, Rohrlich).
3. This leads to the Parity Principle: 50% of elliptic curves have odd rank, and 50% of elliptic curves have even rank.
4. In addition, the Minimalist Principle proclaims that there are as few rational points on elliptic curves as is possible, given the constraint of the parity principle.

# Rank Statistics?

We would like to understand the behavior of

$$\mathrm{AveRank}_{\mathcal{E}}(X) = \frac{\sum_{E \in \mathcal{E}(X)} \mathrm{rank}(E(\mathbb{Q}))}{\pi_{\mathcal{E}}(X)}.$$

The Parity Principle plus the Minimalist Principle put together:

**The Minimalist Conjecture**

Fix a global field $k$. Asymtotically, 50% of elliptic curves over $k$ have rank 0, and 50% have rank 1. Moreover, the average rank is $1/2$, that is

$$\mathrm{AveRank}_{\mathcal{E}}(X) = \frac{\sum_{E \in \mathcal{E}(X)} \mathrm{rank}(E(k))}{\pi_{\mathcal{E}}(X)} \longrightarrow \frac{1}{2} \quad \text{as } X \to \infty.$$

**Theorem** (Bhargava, Dokchitser$^2$, Shankar, Skinner, Urban, Zhang,...)

*For $k = \mathbb{Q}$, we have* $0.261 < \lim^{*}_{X\to\infty} \mathrm{AveRank}_{\mathcal{E}}(X) < 0.885.$

$^*$ The limit is not known to exist! The bounds are for the liminf and limsup.

# Rank Statistics?

The Minimalist Conjecture was first proposed over families of quadratic twists. For $E : y^2 = x^3 + Ax + B$, let $E^d : y^2 = x^3 + Ad^2x + Bd^3$.

**Goldfeld's Conjecture ($\sim$ 79)**

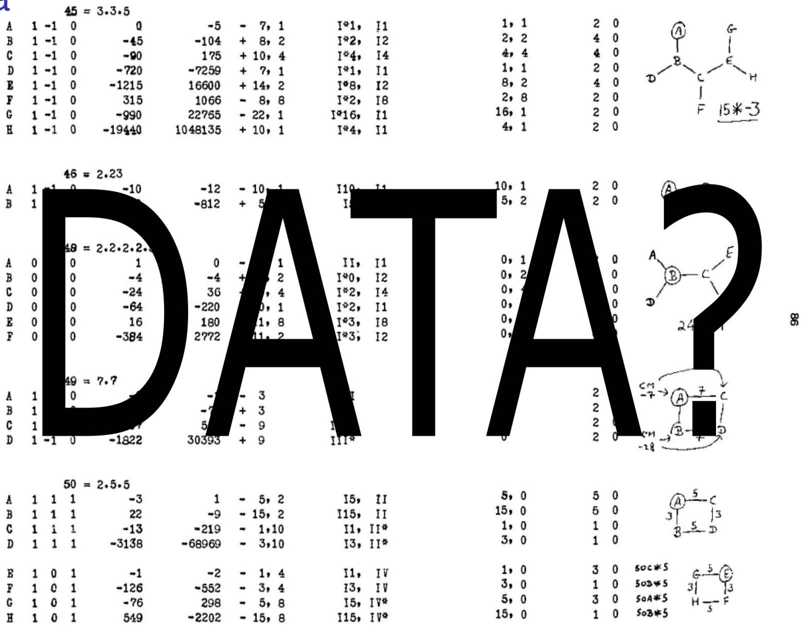
Let $E/\mathbb{Q}$ be an elliptic curve. Then:

$$\lim_{X\to\infty} \frac{\sum_{|d|<X} \operatorname{rank}(E^d(\mathbb{Q}))}{\#\{d : \text{ fund. disc. with } |d| < X\}} = \frac{1}{2}.$$

Progress!

- Support from random matrix theory and work over function fields (Keating-Snaith, Conrey-Keating-Rubinstein-Snaith, Katz-Sarnak), including heuristics on twists of even parity and positive rank.
- Smith has proved that B–S-D implies Goldfeld in many cases.
- Kriz has shown Goldfeld holds for the congruent number family.

# Data



# A Brief History of Elliptic Curve Data

- Birch–Kuyk–Swinnerton-Dyer, Antwerp **1972**'s "Numerical Tables on Elliptic Curves," list of all curves with conductor $N_E \leq 200$.
- Brumer–McGuinness, **1990**, found 311,219 curves of prime conductor $N_E \leq 10^8$.
- Cremona, **1997**, found all 782,493 curves up to conductor $N_E \leq 120{,}000$.
- Stein–Watkins, **2002**, found all curves of prime conductor $\leq 10^{10}$, and those with $|\Delta_E| \leq 10^{12}$ and $N_E \leq 10^8$.
- ...

# The Average Rank in the Stein–Watkins Database

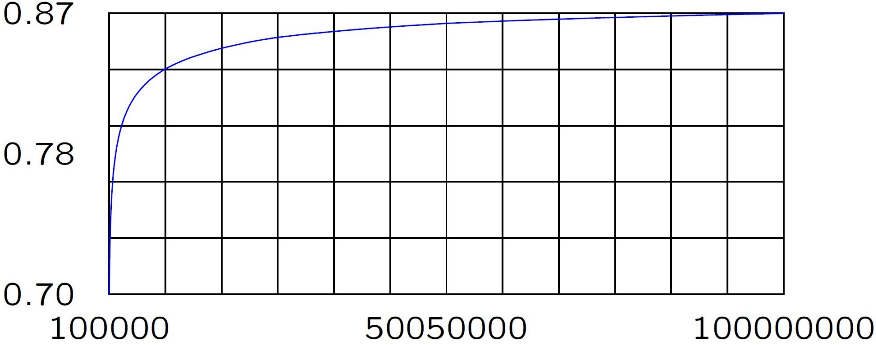
Let

- $\mathcal{E}_{\mathsf{SW}}(X) = \{E \in \mathcal{E} : N_E \leq X \text{ and } |\Delta_E| \leq 10^{12}\}$,
- $\pi_{\mathcal{E},\mathsf{SW}}(X) = \#\mathcal{E}_{\mathsf{SW}}(X)$, and

$$\text{AveRank}_{\mathcal{E},\mathsf{SW}}(X) = \frac{\sum_{E \in \mathcal{E}_{\mathsf{SW}}(X)} \text{rank}(E(\mathbb{Q}))}{\pi_{\mathcal{E},\mathsf{SW}}(X)}.$$

Then...

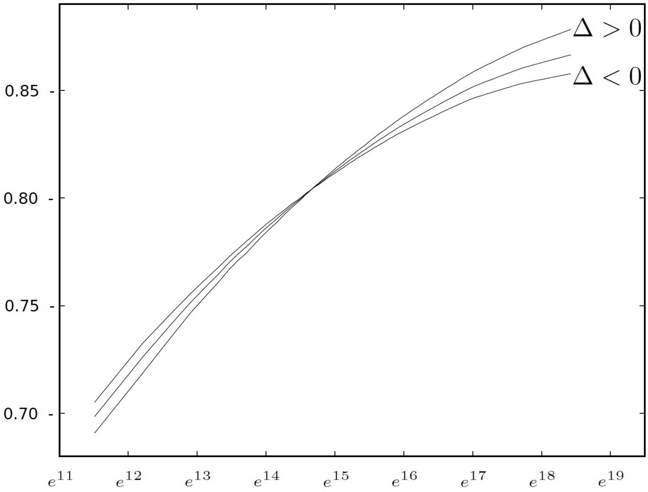
# The Average Rank in the Stein–Watkins Database




S–W average rank as a function of the conductor.
(Note: last value is 0.8664 . . .).

**Source**: “The Average Rank of Elliptic Curves,” Bhargava, **2011**.

# The Average Rank in the Stein–Watkins Database



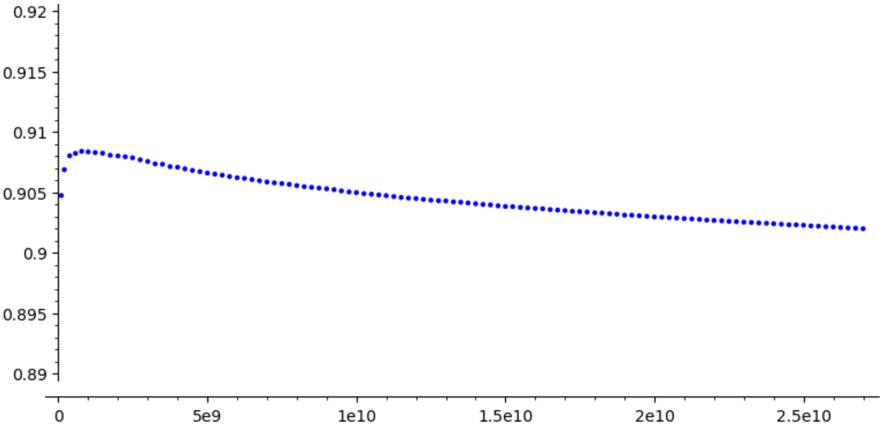

S–W average rank as a function of log of the conductor.

**Source**: "Average ranks of elliptic curves: tension between data and conjecture," Bektemirov, Mazur, Stein, Watkins, **2007**.

# A Brief History of Elliptic Curve Data

- Birch–Kuyk–Swinnerton-Dyer, Antwerp **1972**'s "Numerical Tables on Elliptic Curves," ("presumably complete") list of all curves with conductor $N_E \leq 200$.
- Brumer–McGuinness, **1990**, found 311,219 curves of prime conductor $N_E \leq 10^8$.
- Cremona, **1997**, found all 782,493 curves up to conductor $N_E \leq 120{,}000$. As of **2019**, the database contains all 3,064,705 curves of conductor $N_E \leq 500{,}000$. Data also available at LMFDB.
- Stein–Watkins, **2002**, found all curves of prime conductor $\leq 10^{10}$, and those with $|\Delta_E| \leq 10^{12}$ and $N_E \leq 10^8$.
- The BHKSSW database (Balakrishnan, Ho, Kaplan, Spicer, Stein, Weigandt), **2016**, covers all 238,764,310 elliptic curves up to naive height $26{,}998{,}673{,}868 \approx 2.7 \cdot 10^{10}$.
  - Also six large-height data sets of 100,000 curves with height $\sim 10^k$ for $k = 11, 12, 13, 14, 15, 16$.

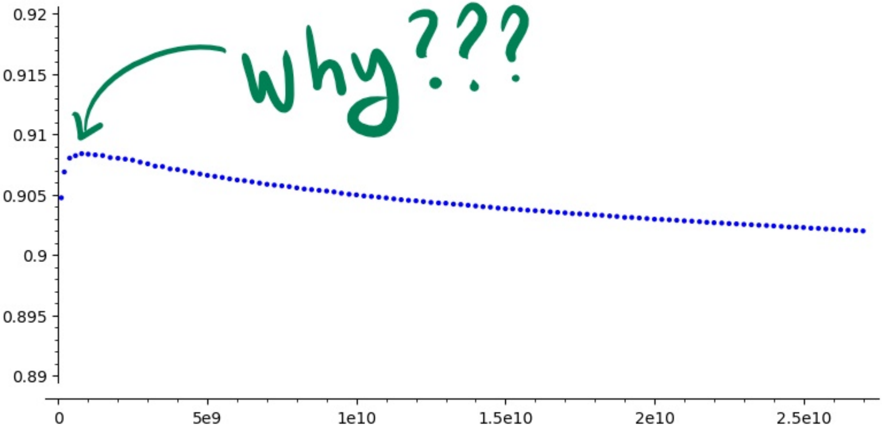
# The Average Rank in the BHKSSW Database




Average rank as a function of the naive height.

Notes: the local max ($\approx 0.90838\ldots$) happens at $X \approx 7.8 \cdot 10^8$. At $X = 2.7 \cdot 10^{10}$ the average rank is $0.90197580\ldots$.

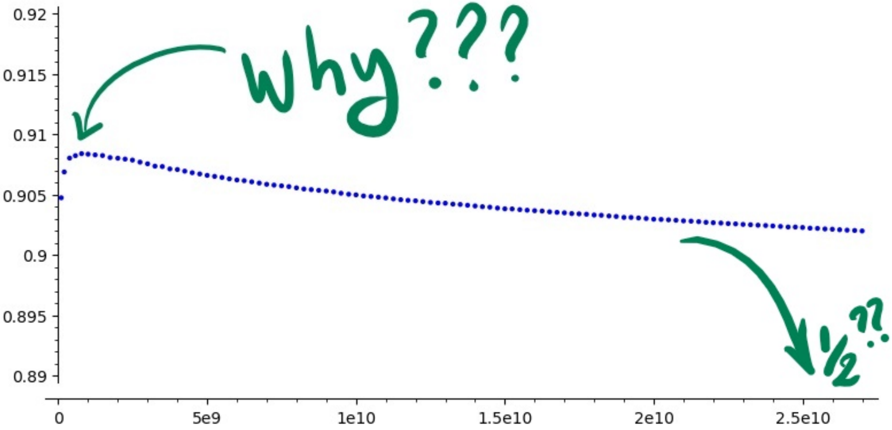
# The Average Rank in the BHKSSW Database




Average rank as a function of the naive height.

Notes: the local max ($\approx 0.90838\ldots$) happens at $X \approx 7.8 \cdot 10^8$. At $X = 2.7 \cdot 10^{10}$ the average rank is $0.90197580\ldots$.

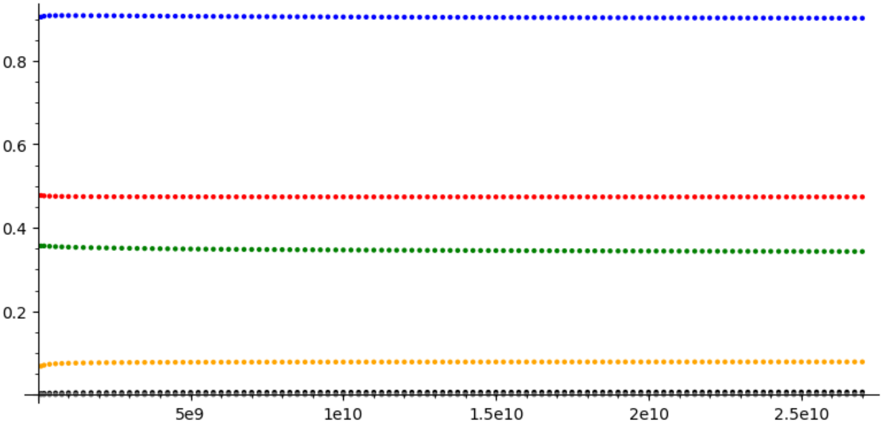
# The Average Rank in the BHKSSW Database




Average rank as a function of the naive height.

Notes: the local max ($\approx 0.90838\ldots$) happens at $X \approx 7.8 \cdot 10^8$. At $X = 2.7 \cdot 10^{10}$ the average rank is $0.90197580\ldots$.

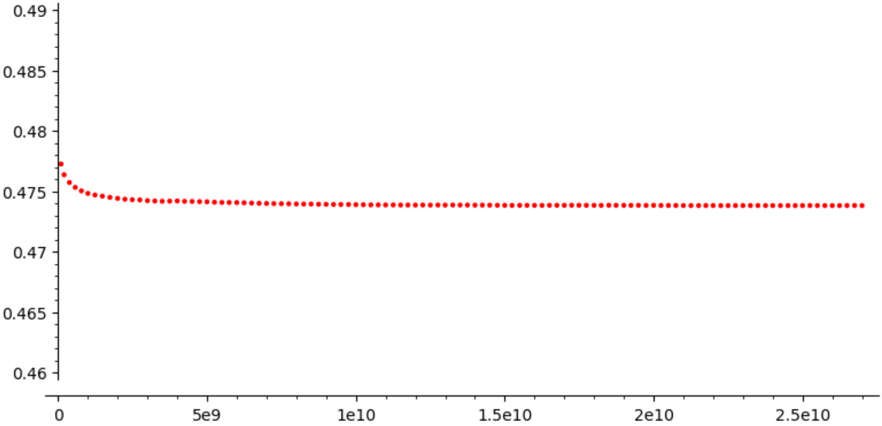
# Signal Analysis: Contributions by Rank




Average rank as a function of the naive height (blue).

Contributions by rank (color): rank 1 (red), rank 2 (green), rank 3 (orange), rank 4 (black), and rank 5 (grey).

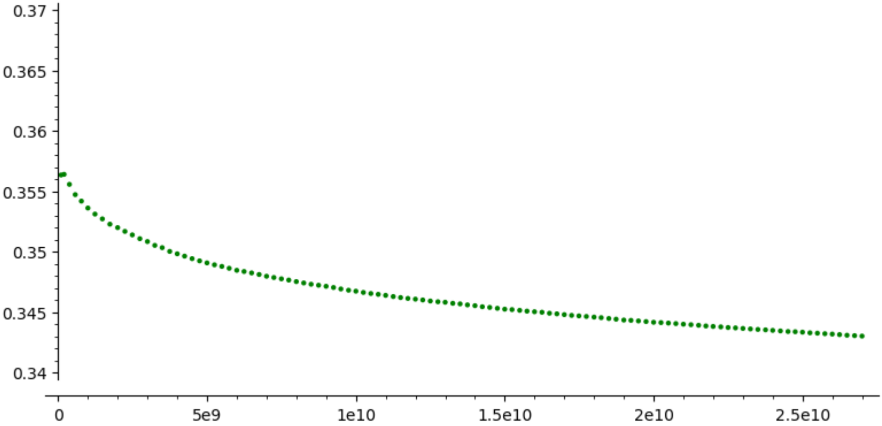
# Signal Analysis: Contributions by Rank




Contribution to the average from rank 1 curves.

$$\text{AveRank}_{\text{rank } 1}(X) = \frac{1}{\pi_{\mathcal{E}}(X)} \cdot \sum_{\substack{E \in \mathcal{E}(X) \\ \text{rank}(E(\mathbb{Q}))=1}} 1.$$

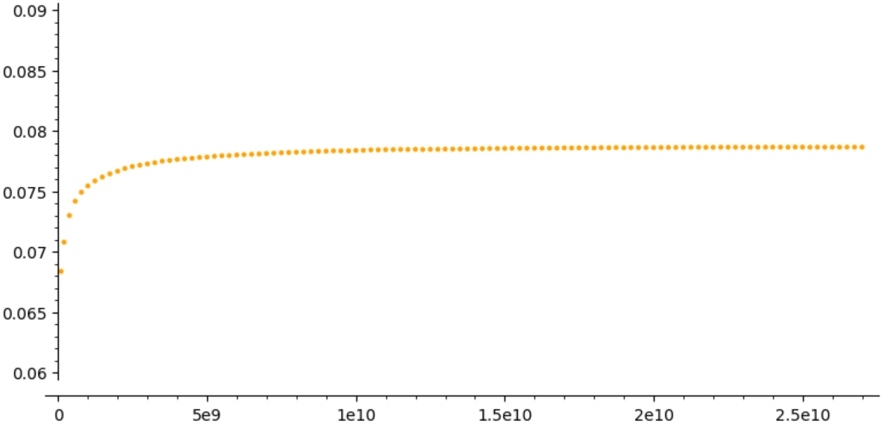
# Signal Analysis: Contributions by Rank




Contribution to the average from rank 2 curves.

$$\text{AveRank}_{\text{rank } 2}(X) = \frac{1}{\pi_{\mathcal{E}}(X)} \cdot \sum_{\substack{E \in \mathcal{E}(X) \\ \text{rank}(E(\mathbb{Q}))=2}} 2.$$

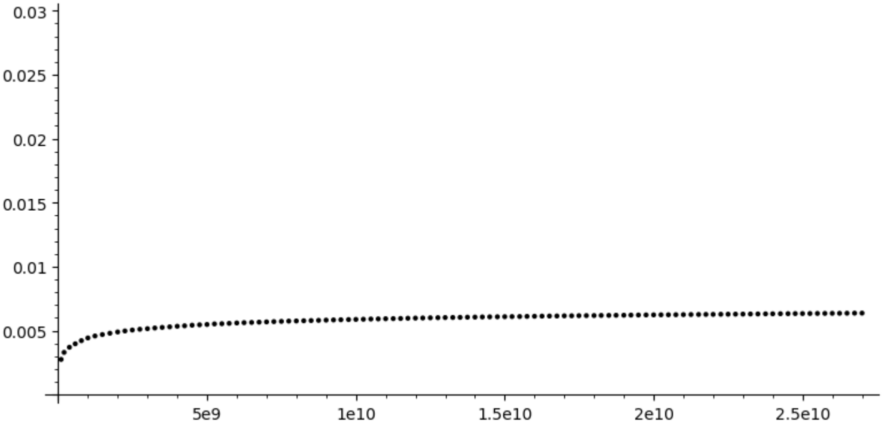
# Signal Analysis: Contributions by Rank




Contribution to the average from rank 3 curves.

$$\text{AveRank}_{\text{rank }3}(X) = \frac{1}{\pi_{\mathcal{E}}(X)} \cdot \sum_{\substack{E \in \mathcal{E}(X) \\ \text{rank}(E(\mathbb{Q}))=3}} 3.$$

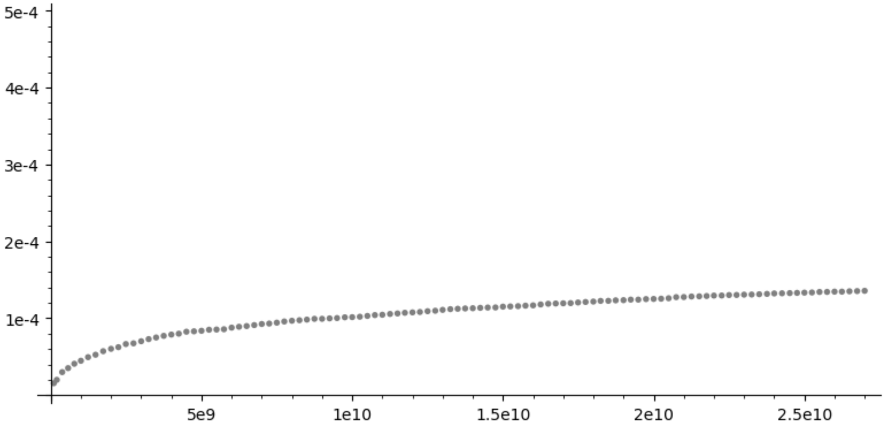
# Signal Analysis: Contributions by Rank




Contribution to the average from rank 4 curves.

$$\text{AveRank}_{\text{rank }4}(X) = \frac{1}{\pi_{\mathcal{E}}(X)} \cdot \sum_{\substack{E \in \mathcal{E}(X) \\ \text{rank}(E(\mathbb{Q}))=4}} 4.$$

# Signal Analysis: Contributions by Rank




Contribution to the average from rank 5 curves.

$$\text{AveRank}_{\text{rank } 5}(X) = \frac{1}{\pi_{\mathcal{E}}(X)} \cdot \sum_{\substack{E \in \mathcal{E}(X) \\ \text{rank}(E(\mathbb{Q}))=5}} 5.$$

# Goals

**Question**

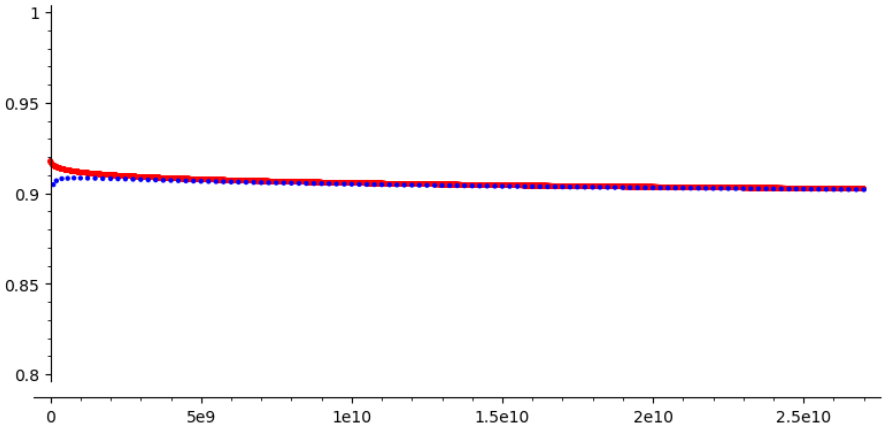
If the Minimalist Conjecture holds, at what naive height $X$ should we expect $\text{AveRank}_{\mathcal{E}}(X) \approx 0.5$?

**GOAL 1:** A (probabilistic) model that explains the graph of the average rank up to height $X$.

**GOAL 2:** A model that explains the proportion of elliptic curves of each rank $r \geq 0$ up to height $X$.

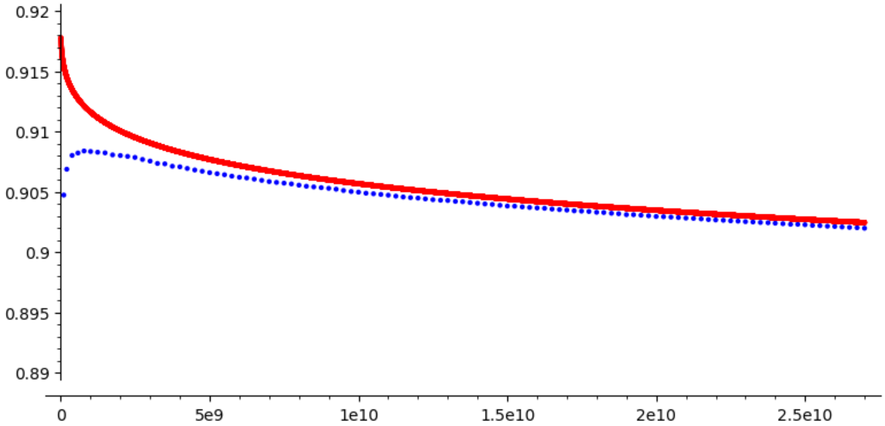
In **2016**, we proposed a probabilistic (Cramér-like) model for ranks.

# Probabilistic Model: Average Rank Predictions



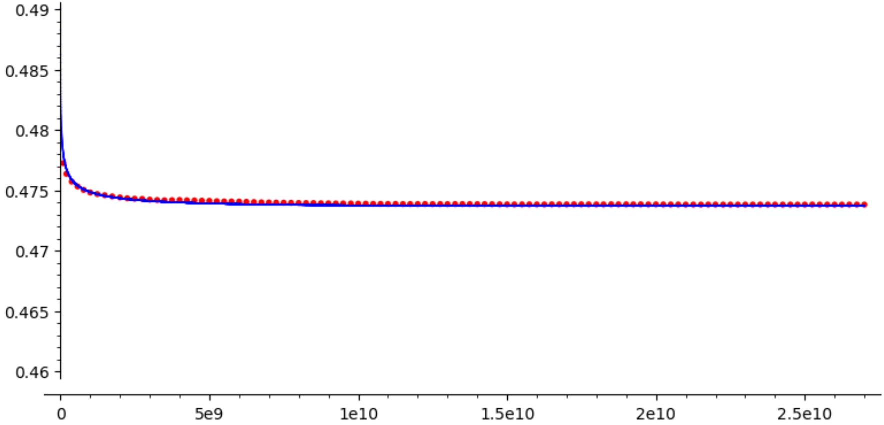

Values of $\text{AveRank}_{\mathcal{E}}(X)$ from the BHKSSW database (blue dots), and the approximations predicted by our model (in red).

# Probabilistic Model: Ave. Rank Predictions (zoom in)



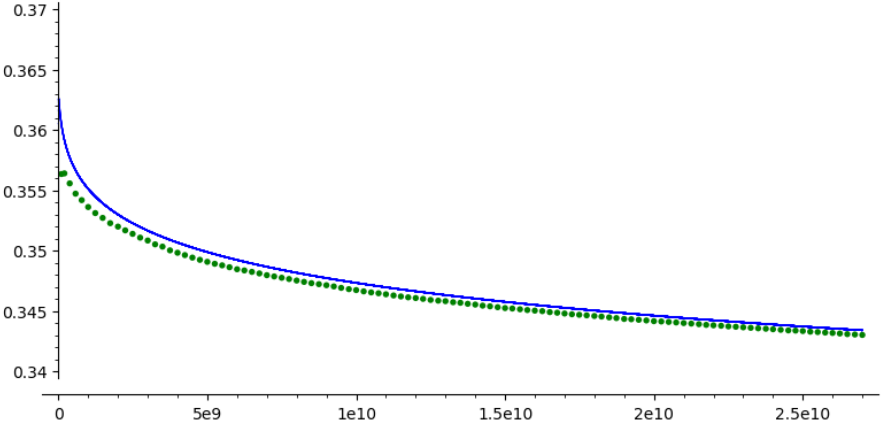

Values of $\mathrm{AveRank}_{\mathcal{E}}(X)$ from the BHKSSW database (blue dots), and the approximations predicted by our model (in red).

# Probabilistic Model: Ave. Rank 1 Predictions



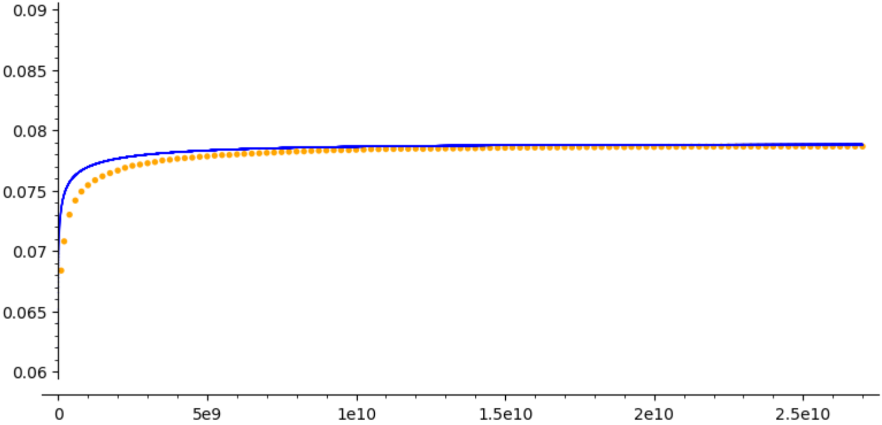

Values of $\text{AveRank}_{\text{rank } 1}(X)$ from the BHKSSW database (blue dots), and the approximations predicted by our model (in red).

# Probabilistic Model: Ave. Rank 2 Predictions



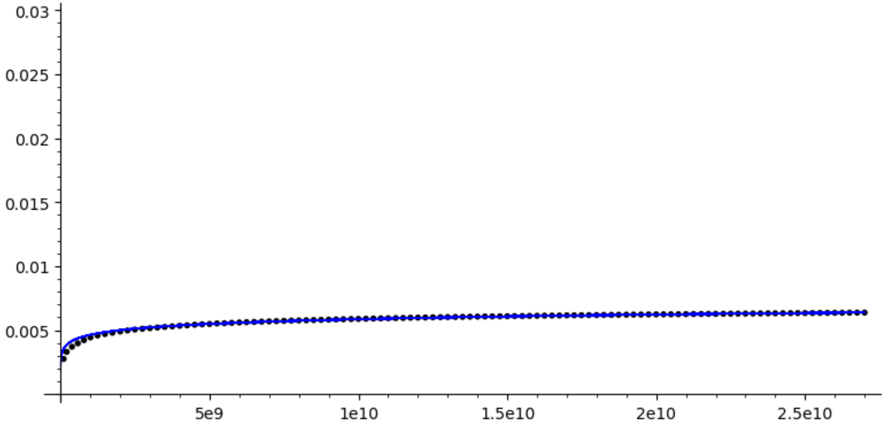

Values of AveRank $_{\text{rank }2}(X)$ from the BHKSSW database (blue dots), and the approximations predicted by our model (in red).

# Probabilistic Model: Ave. Rank 3 Predictions



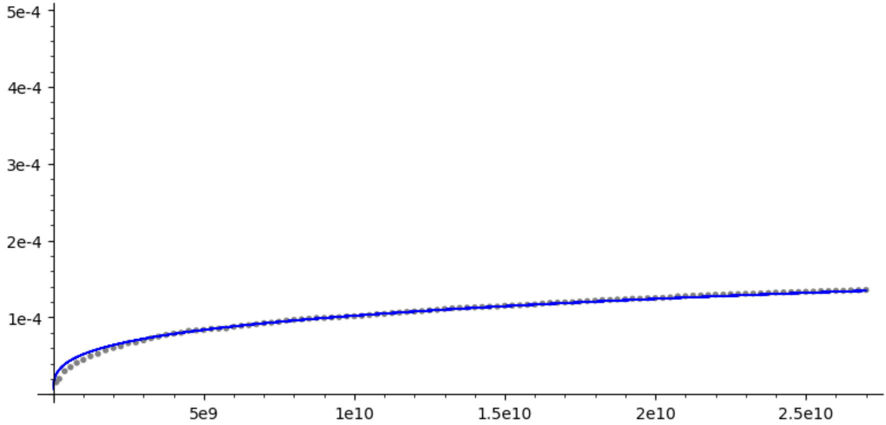

Values of AveRank $_{\text{rank } 3}(X)$ from the BHKSSW database (blue dots), and the approximations predicted by our model (in red).

# Probabilistic Model: Ave. Rank 4 Predictions



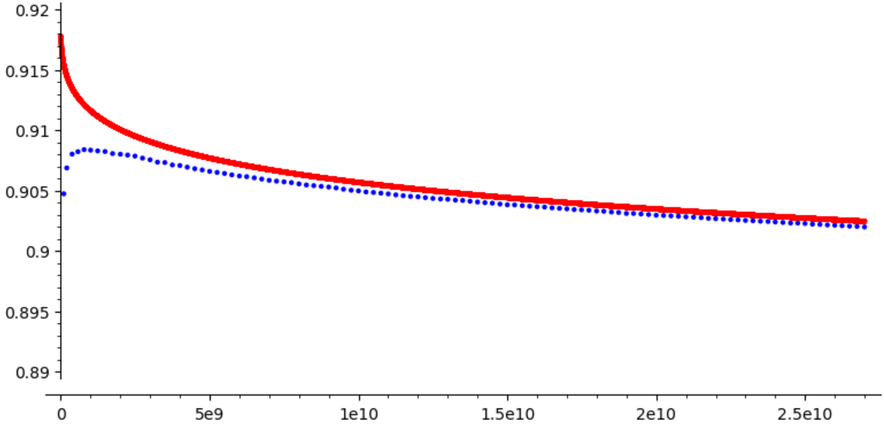

Values of $\text{AveRank}_{\text{rank } 4}(X)$ from the BHKSSW database (blue dots), and the approximations predicted by our model (in red).

# Probabilistic Model: Ave. Rank 5 Predictions



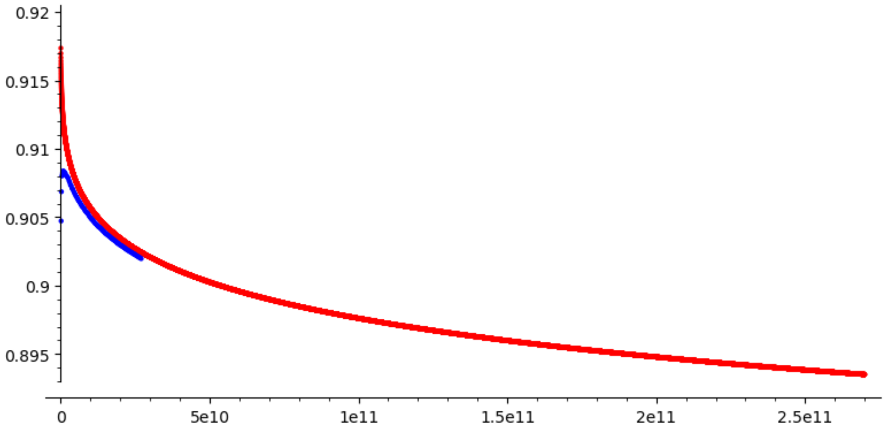

Values of $\text{AveRank}_{\text{rank } 5}(X)$ from the BHKSSW database (blue dots), and the approximations predicted by our model (in red).

# Probabilistic Model: Ave. Rank (zoom in)



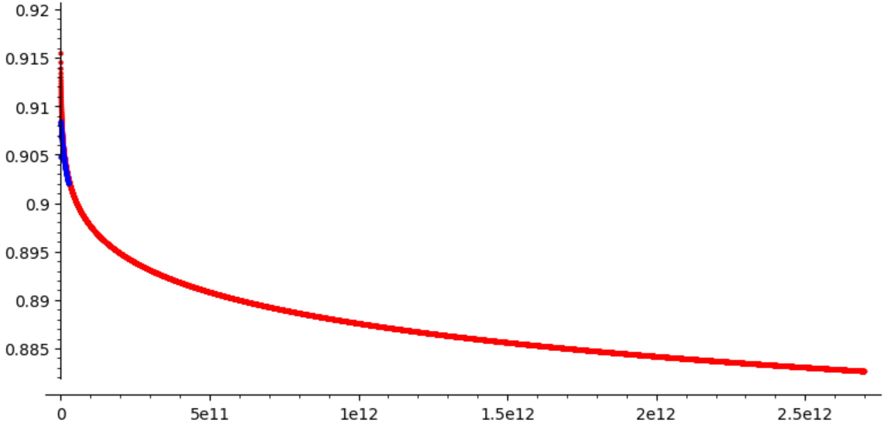

Values of $\text{AveRank}_{\mathcal{E}}(X)$ from the BHKSSW database (blue dots), and the approximations predicted by our model (in red).

# Probabilistic Model: Ave. Rank (zoom out!)



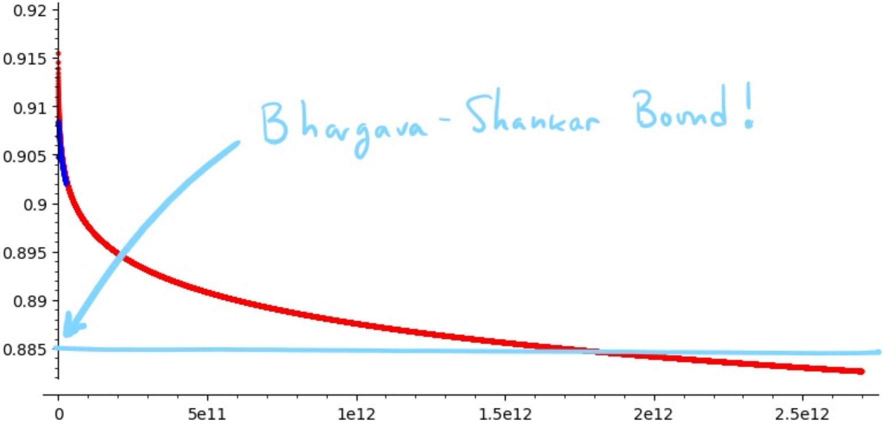

Values of $\text{AveRank}_{\mathcal{E}}(X)$ from the BHKSSW database (blue dots), and the approximations predicted by our model (in red).

# Probabilistic Model: Ave. Rank (zoom out!!)



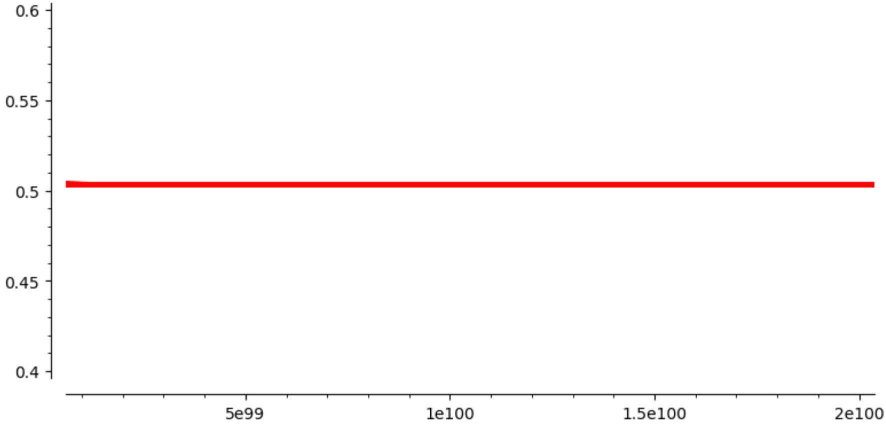

Values of $\text{AveRank}_{\mathcal{E}}(X)$ from the BHKSSW database (blue dots), and the approximations predicted by our model (in red).

# Probabilistic Model: Ave. Rank (zoom out!!)




Values of $\text{AveRank}_{\mathcal{E}}(X)$ from the BHKSSW database (blue dots), and the approximations predicted by our model (in red).

# Probabilistic Model: Ave. Rank (zoom OUT!)




Values of $\mathrm{AveRank}_{\mathcal{E}}(X)$ from the BHKSSW database (blue dots), and the approximations predicted by our model (in red).

| $X$ | AveRank($X$) | $X$ | AveRank($X$) |
|---|---|---|---|
| $10^{10}$ | 0.905665 | $10^{50}$ | 0.548880 |
| $10^{15}$ | 0.846828 | $10^{75}$ | 0.512531 |
| $10^{20}$ | 0.766868 | $10^{100}$ | 0.503256 |
| $10^{30}$ | 0.649901 | $10^{150}$ | 0.500215 |
| $10^{40}$ | 0.585108 | $10^{200}$ | 0.500006 |

Conjectural approximate values of $\text{AveRank}_{\mathcal{E}}(X)$ obtained using our model.

# Probabilistic Model: Rank Predictions

Let $\mathcal{R}_r(X) = \{E \in \mathcal{E}(X) : \text{rank}(E(\mathbb{Q})) = r\}, \quad \pi_{\mathcal{R}_r}(X) = \#\mathcal{R}_r(X)$.

| | $r = 1$ | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $\pi_{\mathcal{R}_r}(2.7 \cdot 10^{10})$ | 113128929 | 40949289 | 6259157 | 380519 | 6481 |
| Approx. value | 113133971 | 41005107 | 6273138 | 381272 | 6438 |
| $\lvert\text{Error}\rvert$ | 5042 | 55818 | 13981 | 753 | 43 |
| Error % | 0.00445 | 0.13631 | 0.22336 | 0.19788 | 0.66347 |
| $\approx s_r \cdot X^{1/2}$ | 68848.72 | 45942.96 | 13112.47 | 1749.97 | 111.73 |

Table: Values of $\pi_{\mathcal{R}_r}(2.7 \cdot 10^{10})$ from the BHKSWW database, the approximate values (rounded to the closest integer) given by the model, the absolute error, the error as a percentage of the actual value of $\pi_{\mathcal{R}_r}$, and the size of the predicted error $s_r \cdot (2.7 \cdot 10^{10})^{1/2}$.

# The Probabilistic Model

# Cramér's random model of the prime numbers

The prime number theorem suggests that an integer $X \geq 3$ is prime with probability $1/\log X$.

**Cramér's model (1936):** Let $B^3, B^4, \ldots, B^X, \ldots$ be bins with red and white balls, one for each integer $X \geq 3$.

1. The chance of drawing a red ball from bin $B^X$ is $1/\log X$.
2. Draw one ball from each bin, and let $q_n \in \mathbb{N}$ be the index of the bin where we got the $n$-th red ball.
3. Let $C$ be the space of all sequences $\{q_n\}_{n\geq 1}$.
4. Then, we predict properties of prime numbers from the asymptotic statistics of $C$.

# How does the probabilistic model work?

Recall the short exact sequence

$$0 \to E(\mathbb{Q})/2E(\mathbb{Q}) \to \mathrm{Sel}_2(E/\mathbb{Q}) \to Ш(E/\mathbb{Q})[2] \to 0.$$

We define the (2-)Selmer rank (or selrank) of $E(\mathbb{Q})$ by

$$\mathrm{selrank}(E(\mathbb{Q})) = \dim_{\mathbb{F}_2} \mathrm{Sel}_2(E/\mathbb{Q}) - \dim_{\mathbb{F}_2} E(\mathbb{Q})[2].$$

Then,

$$\boxed{\mathrm{rank}(E(\mathbb{Q})) \leq \mathrm{selrank}(E(\mathbb{Q})).}$$

We model the distribution of M–W ranks for a fixed selrank $= n$ and a fixed naive height $X$.

For each $n \geq 0$, we define $\mathcal{E}^X = \{E \in \mathcal{E} : \mathrm{ht}(E) = X\}$ and

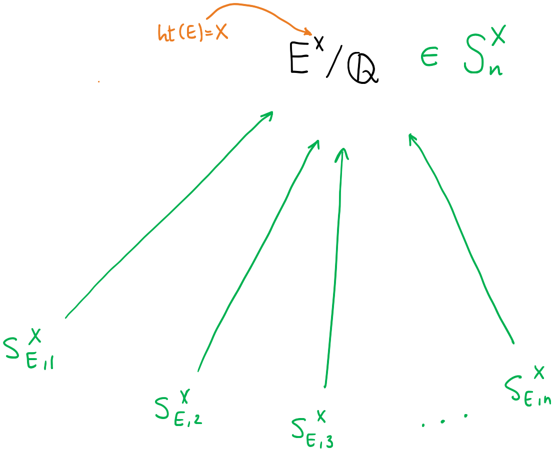
$$\mathcal{S}_n^X = \{E \in \mathcal{E}^X : \mathrm{selrank}(E(\mathbb{Q})) = n\}.$$

# Model

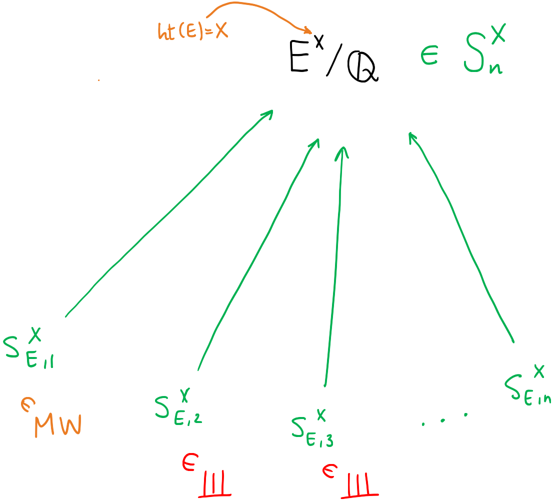
ht(E) = X

$E^X / \mathbb{Q}$

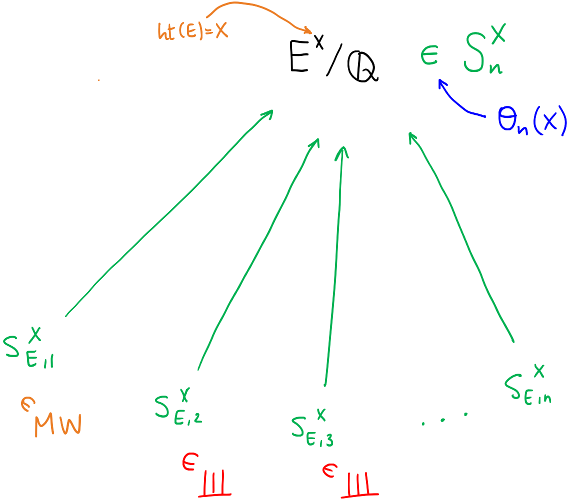
# Model



# Model

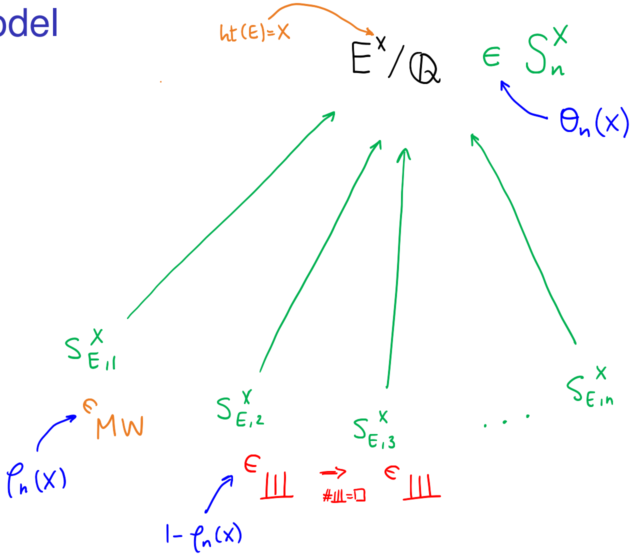


# Model

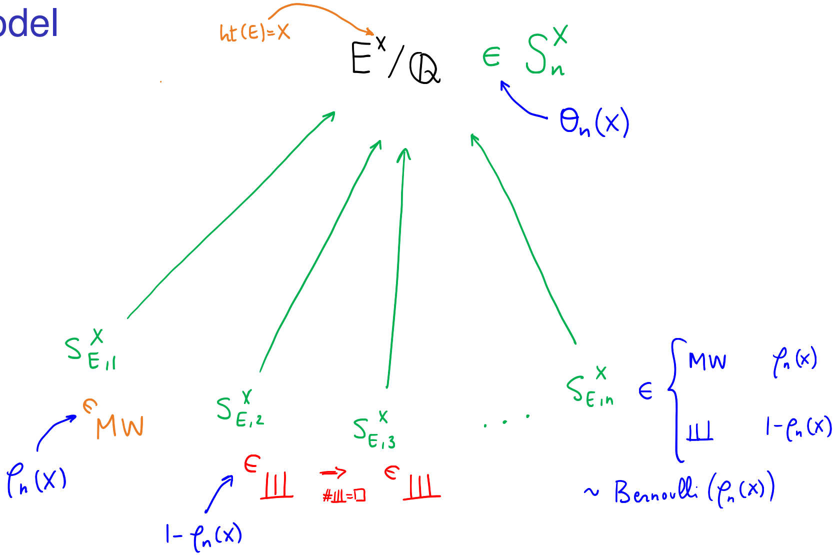


$ht(E)=X$
$E^X/\mathbb{Q} \in S_n^X$
$\theta_n(X)$
$S^X_{E,1} \in MW$
$S^X_{E,2} \in Ш$
$S^X_{E,3} \in Ш$
$\dots$
$S^X_{E,n}$

# Model

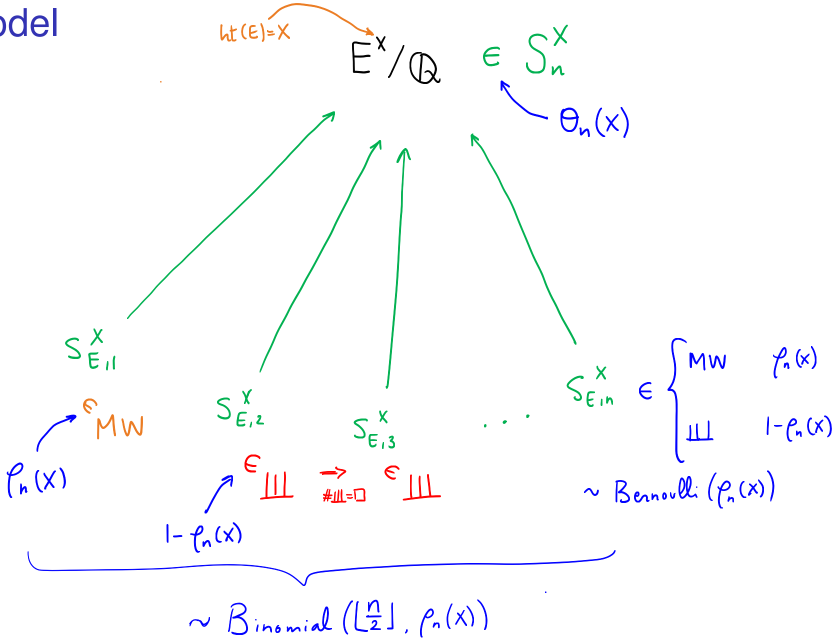


# Model



ht(E)=X
$E^X/\mathbb{Q} \in S_n^X$
$\theta_n(X)$
$S_{E,1}^X$
$\in$ MW
$\rho_n(X)$
$S_{E,2}^X$
$\in$ Ш
$1-\rho_n(X)$
$\Longrightarrow$ #Ш=□
$S_{E,3}^X$
$\in$ Ш
. . .
$S_{E,n}^X \in$
MW $\rho_n(X)$
Ш $1-\rho_n(X)$
~ Bernoulli($\rho_n(X)$)

# Model


ht(E)=X
$E^X/\mathbb{Q} \in S_n^X$
$\theta_n(X)$
$S_{E,1}^X$
$\in$ MW
$\rho_n(X)$
$S_{E,2}^X$
$\in$ Ш
$1-\rho_n(X)$
$\#Ш = \square$
$S_{E,3}^X$
$\in$ Ш
. . .
$S_{E,n}^X \in$
MW $\rho_n(x)$
Ш $1-\rho_n(x)$
$\sim$ Bernoulli$(\rho_n(x))$
$\sim$ Binomial$(\lfloor \frac{n}{2} \rfloor, \rho_n(x))$

# Model

For example,

- If $n = 4$, (the expected value of) the number of elliptic curves of Selmer rank 4, and Mordell–Weil ranks 0, 2, and 4 are given respectively by

$$\begin{aligned} (Ш, Ш, Ш, Ш) &\Rightarrow \#\mathcal{E}^X \cdot \theta_4(X) \cdot (1 - \rho_4(X))^2, \\ (Ш, Ш, MW, MW) &\Rightarrow \#\mathcal{E}^X \cdot \theta_4(X) \cdot 2\rho_4(X)(1 - \rho_4(X)), \\ (MW, MW, MW, MW) &\Rightarrow \#\mathcal{E}^X \cdot \theta_4(X) \cdot \rho_4(X)^2. \end{aligned}$$

- If $n = 5$, the number of elliptic curves of Selmer rank 5, and Mordell–Weil ranks 1, 3, and 5 are given respectively by

$$\begin{aligned} (Ш, Ш, Ш, Ш, MW) &\Rightarrow \#\mathcal{E}^X \cdot \theta_5(X) \cdot (1 - \rho_5(X))^2, \\ (Ш, Ш, MW, MW, MW) &\Rightarrow \#\mathcal{E}^X \cdot \theta_5(X) \cdot 2\rho_5(X)(1 - \rho_5(X)), \\ (MW, MW, MW, MW, MW) &\Rightarrow \#\mathcal{E}^X \cdot \theta_5(X) \cdot \rho_5(X)^2. \end{aligned}$$

* Warning! Events are *not* independent! A covariance factor needs to be calculated to correctly compute the expected values.

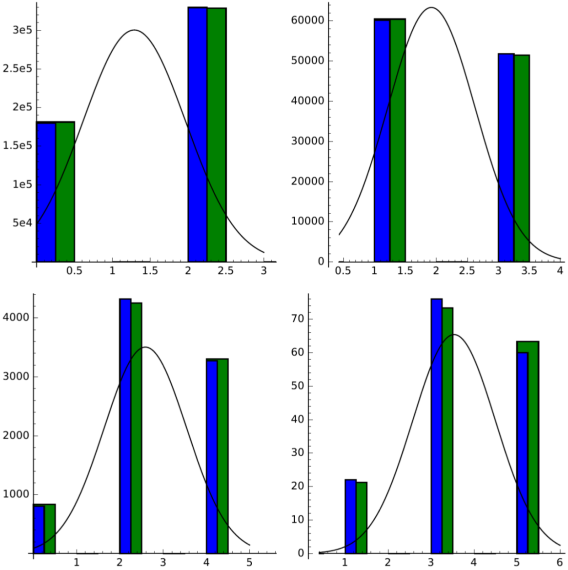



Figure: Distribution of Mordell–Weil ranks (in blue) among elliptic curves in $\mathcal{E}([2 \cdot 10^{10}, 2.025 \cdot 10^{10}])$ by Selmer rank $n = 2, 3, 4, 5$, and compared to the predicted M–W ranks (in green) that we would expect from the models.

# Model

| $n$ | $\pi_{\mathcal{S}_n}(I)$ | M–W ranks observed in $\mathcal{S}_n$ | M–W ranks predicted |
|---|---|---|---|
| 2 | 509,845 | $[180128, 0, 329717, 0, 0, 0]$ | $[181246.58, 0, 328598.41, 0, 0, 0]$ |
| 3 | 111,926 | $[0, 60149, 0, 51777, 0, 0]$ | $[0, 60455.09, 0, 51470.90, 0, 0]$ |
| 4 | 8399 | $[803, 0, 4321, 0, 3275, 0]$ | $[836.68, 0, 4256.52, 0, 3305.78, 0]$ |
| 5 | 158 | $[0, 22, 0, 76, 0, 60]$ | $[0, 21.24, 0, 73.38, 0, 63.36]$ |

Mordell–Weil ranks observed and the ranks predicted by the models in the height interval $I = [2 \cdot 10^{10}, 2.025 \cdot 10^{10}]$.

# Test Elliptic Curves

A **test elliptic curve** is a triple $E = (X, n, \mathrm{Sel}_2)$ consisting of:

- a positive integer $X \geq 1$, the height of $E$, also denoted $X = \mathrm{ht}(E)$,
- a non-negative integer $n$, the Selmer rank of $E$, also denoted $n = \mathrm{selrank}(E)$, and
- a vector $\mathrm{Sel}_2(E) = (s_{E,1}, s_{E,2}, \ldots, s_{E,\lfloor n/2 \rfloor})$ of $\lfloor n/2 \rfloor$ **test Selmer elements.** Each Selmer element is a symbol, which is either a MW, or a Ш symbol.

We define:

- $\widetilde{\mathcal{E}}$, the set of all test elliptic curves,
- $\widetilde{\mathcal{E}}^X$, test elliptic curves with height $X$,
- $\widetilde{\mathcal{S}}_n^X$, test elliptic curves with height $X$ and Selmer rank $n$,
- $\mathrm{rank}(E) = (n \bmod 2) + 2 \cdot \#\{\text{MW elements in } \mathrm{Sel}_2(E)\}$.

# Test Elliptic Curves

To each ordinary elliptic curve we can attach a test elliptic curve.

## Example

Let $E/\mathbb{Q}$ be the elliptic curve $y^2 = x^3 + 2993x$.

- Height is $X = 4 \cdot 2993^3 = 107245762628$.
- A 2-descent shows $\mathrm{Sel}_2(E/\mathbb{Q}) \cong (\mathbb{Z}/2\mathbb{Z})^5$. Since $E(\mathbb{Q})_{\mathrm{tors}} \cong \mathbb{Z}/2\mathbb{Z}$, it follows that $\mathrm{selrank}(E) = 4$.
- A 4-descent shows that $E(\mathbb{Q}) \cong \mathbb{Z}/2\mathbb{Z} \oplus \mathbb{Z}^2$, and $\text{Ш}(E/\mathbb{Q})[2] \cong (\mathbb{Z}/2\mathbb{Z})^2$.

Hence, this elliptic curve would be represented as a test elliptic curve by the triple

$$(107245762628, 4, (\mathrm{MW}, \text{Ш})).$$

# Pieces of the Probabilistic Model

To put our probabilistic model together we need estimates of

- $\#\mathcal{E}^X$ : the number of elliptic curves of height $X$.
- $\#\mathcal{S}_n^X$ : the number of elliptic curves of height $X$, selrank $n$.
- $\theta_n(X) = \#\mathcal{S}_n^X / \#\mathcal{E}^X$ : the proportion of ell. curves of height $X$ and selrank $n$, among all curves of height $X$ (when $\#\mathcal{E}^X \neq 0$).
- $\rho_n(X)$ : the probability that a non-$E(\mathbb{Q})_{\text{tors}}$ Selmer element coming from a selrank-$n$ elliptic curve of height $X$ is a MW element.

# $\#\mathcal{E}^X$

## Theorem (Brumer)

*The number of elliptic curves of height up to X satisfies*

$$\left|\pi_{\mathcal{E}}(X) - \frac{2^{4/3}X^{5/6}}{3^{3/2}\zeta(10)}\right| \le \frac{2X^{1/2}}{3^{3/2}\zeta(6)} + O(X^{1/3}),$$

*for any* $\varepsilon > 0$*. In particular,* $\pi_{\mathcal{E}}(X) = \kappa X^{5/6} + O(X^{1/2})$ *where the constant* $\kappa = 2^{4/3} \cdot (\zeta(10) \cdot 3^{3/2})^{-1} \approx 0.484462004349\ldots$

Thus (on average), we have

$$\pi_{\mathcal{E}}((X, X+N]) \approx \frac{5\kappa}{6}\int_X^{X+N} \frac{1}{H^{1/6}}\,dH + O\left(\frac{N}{X^{1/2}}\right).$$

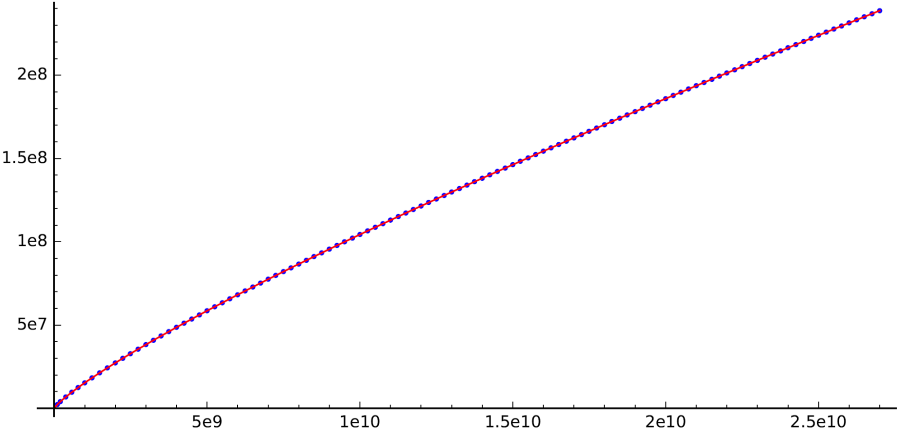



Figure: Values of $\pi_{\mathcal{E}}(X)$ from the BHKSSW database (blue dots), and the function $0.4844620043 \cdot X^{5/6}$ (in red).

# $\#\mathcal{S}_n^X$ : Elliptic Curves of Selmer Rank $n$, Height $X$

Following work on quadratic twists by Heath-Brown, Monsky, Kane, and Swinnerton-Dyer:

**Conjecture (Poonen–Rains, for $p = 2$)**

$$\begin{aligned} s_n &= \text{Prob}(\text{selrank}(E(\mathbb{Q})) = n) = \lim_{X\to\infty} \frac{\pi_{\mathcal{S}_n}(X)}{\pi_{\mathcal{E}}(X)} \\ &= \left(\prod_{j\geq 0} \frac{1}{1+2^{-j}}\right) \cdot \left(\prod_{k=1}^{n} \frac{2}{2^k - 1}\right). \end{aligned}$$

| $s_0$ | $s_1$ | $s_2$ | $s_3$ | $s_4$ | $s_5$ |
|---|---|---|---|---|---|
| 0.209711 | 0.419422 | 0.279614 | 0.079889 | 0.010651 | 0.000687 |

Let $\widetilde{\mathcal{S}}_n$ be the subset of test elliptic curves of selrank $n$, and $\widetilde{\mathcal{S}}_n^X \subseteq \widetilde{\mathcal{E}}^X$.

**In our model:** The probability of picking a test curve in $\widetilde{\mathcal{S}}_n^X$ out of $\widetilde{\mathcal{E}}^X$ is given by $\theta_n(X)$, where $\theta_n(X)$ is a function such that $\lim_{X\to\infty} \theta_n(X) = s_n$.

### Hypothesis A ($H_A$)

*Let $n \geq 0$, let $X \geq 0$, and let $Y_{\mathrm{Sel},n,X} : \widetilde{\mathcal{E}}^X \to \{0,1\}$ be the function that takes values*

$$Y_{\mathrm{Sel},n,X}(E) = \begin{cases} 1 & \text{if } \mathrm{selrank}(E) = n, \\ 0 & \text{otherwise.} \end{cases}$$

*Then, $Y_{\mathrm{Sel},n,X}(E)$ is a random variable that follows a Bernoulli distribution with probability $\theta_n(X)$, such that $\lim_{X\to\infty} \theta_n(X) = s_n$.*

# Proportion of Curves of Each Selrank at Each Height

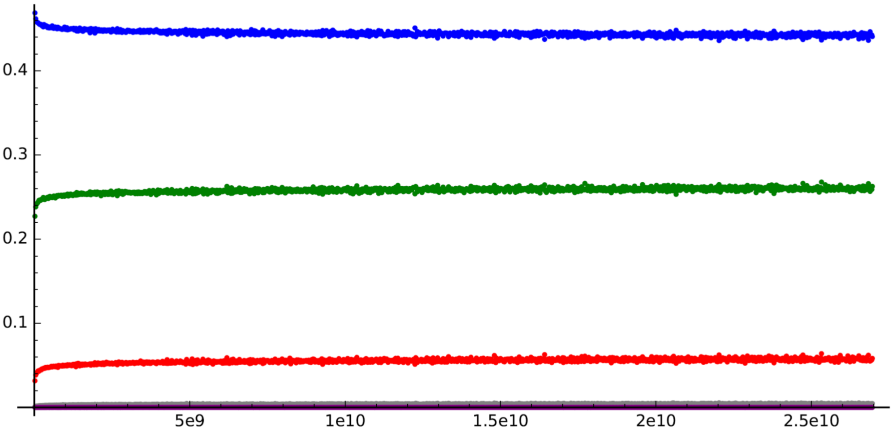



Figure: Graphs of the ratios $\theta_n(X)$ for $n = 1$ (blue), 2 (green), 3 (red), 4 (gray), 5 (purple), based on the BHKSSW data.

## Refined Hypothesis A ($H_A$)

*Let $n \geq 0$, let $X \geq 0$, and let $Y_{\mathrm{Sel},n,X} : \widetilde{\mathcal{E}}^X \to \{0,1\}$ be the function that takes values*

$$Y_{\mathrm{Sel},n,X}(E) = \begin{cases} 1 & \textit{if } \mathrm{selrank}(E) = n, \\ 0 & \textit{otherwise.} \end{cases}$$

*Then, $Y_{\mathrm{Sel},n,X}(E)$ is a random variable that follows a Bernoulli distribution with probability $\theta_n(X)$, where*

$$\theta_n(X) = \frac{s_n}{1 + C_n X^{-e_n}},$$

*for some constants $C_n$ and $e_n$.*

| $n$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $C_n$ | $-0.401169$ | 1.411086 | 11.182227 | 179.717499 | 95474.850980 |
| $e_n$ | 0.085402 | 0.123486 | 0.140615 | 0.203396 | 0.399370 |

Table: The coefficients of the best-fit regression $\theta_n(X) \approx s_n/(1 + C_n X^{-e_n})$.

# Proportion of Curves of Each Selrank and Model

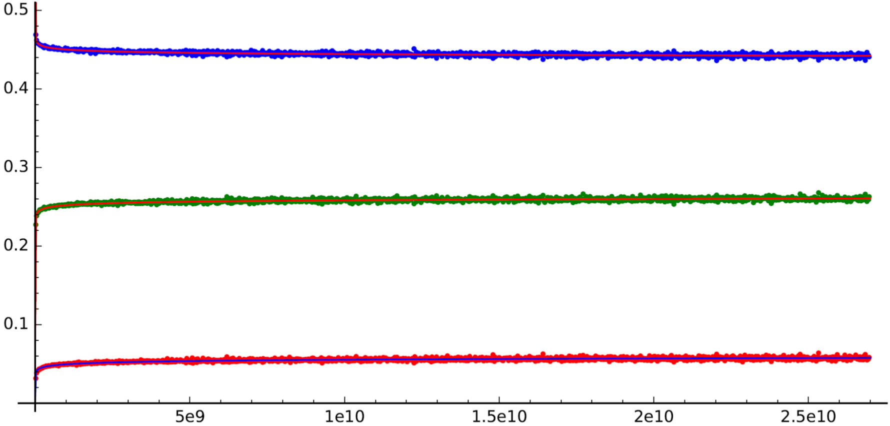



Figure: Graphs of the ratios $\theta_n(X)$ for $n = 1$ (blue), 2 (green), 3 (red), and the corresponding models of the form $s_n/(1 + C_n X^{-e_n})$.

# Proportion of Curves of Each Selrank and Model

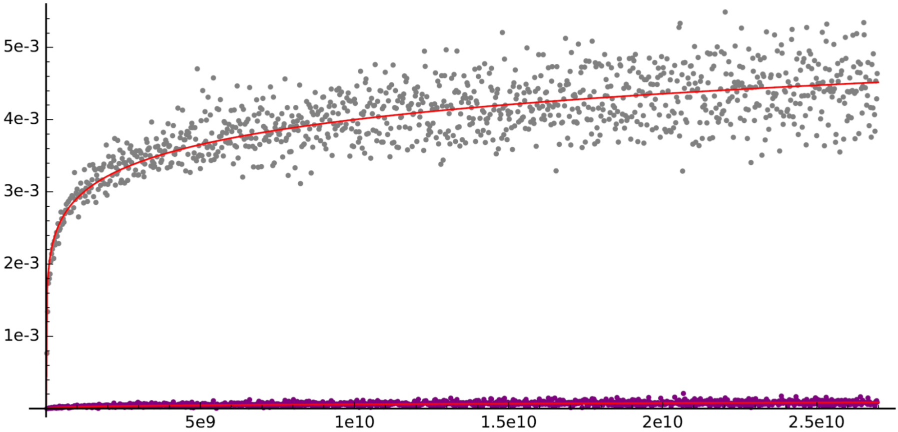



Figure: Graphs of the ratios $\theta_n(X)$ for $n = 4$ (gray), 5 (purple), and the corresponding models of the form $s_n/(1 + C_n X^{-e_n})$, in red.

# Noise in the Distribution of Selranks

## Corollary

*Assume $H_A$, and let $\mathfrak{E} = \{E_1, \ldots, E_m\} \subseteq \widetilde{\mathcal{E}}^X$ be a sample of $m$ test elliptic curves with height $X$ chosen independently. Then, the number of curves in $\mathfrak{E}$ of selrank $n$ follows a binomial distribution $B(m, \theta_n(X))$. In particular*

$$\mathbb{E}\left(\#(\mathfrak{E} \cap \widetilde{\mathcal{S}}_n)/\#\mathfrak{E}\right) = \theta_n(X)$$

*with standard error $\sqrt{\frac{1}{m}\theta_n(X)(1 - \theta_n(X))}$.*

Note: for large values of $m$, the binomial is well approximated by a gaussian distribution.

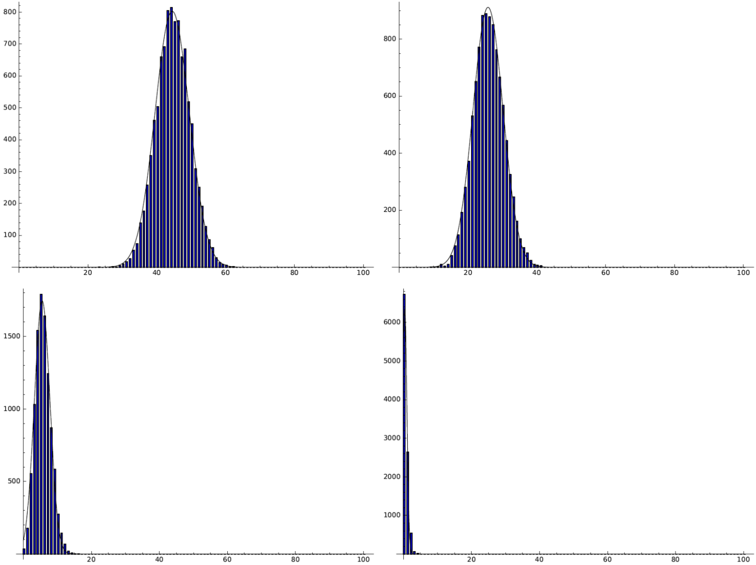



Figure: Histogram of the distribution of 10000 experiments of picking 100 elliptic curves of height $\approx 9 \cdot 10^9$, at random, and counting the number of Selmer ranks equal to $n = 1, 2, 3, 4$, and the normal dist. predicted by $H_A$.

## Proposition

*Assume $H_A$. Then, the expected value of $\pi_{\widetilde{\mathcal{S}}_n}(X)$ is given by*

$$\mathbb{E}(\pi_{\widetilde{\mathcal{S}}_n}(X)) = \frac{5\kappa}{6}\int_1^X \frac{\theta_n(H)}{H^{1/6}}\,dH + O\left(X^{1/2}\right),$$

*where $\kappa = 2^{4/3}\cdot(\zeta(10)\cdot 3^{3/2})^{-1}$. If in addition we assume the refined version of $H_A$, then*

$$\mathbb{E}(\pi_{\widetilde{\mathcal{S}}_n}(X)) = \frac{5\kappa s_n}{6}\int_1^X \frac{1}{H^{1/6}(1+C_nH^{-e_n})}\,dH + O\left(X^{1/2}\right).$$

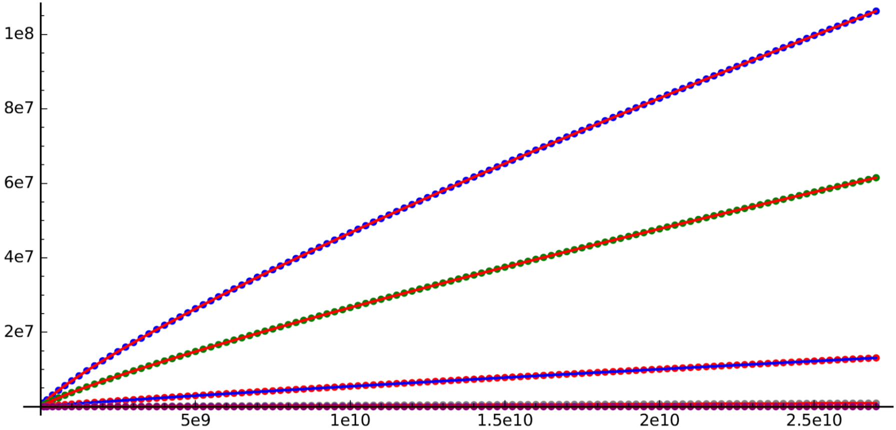



Figure: Values of $\pi_{\mathcal{S}_n}(X)$ using the BHKSSW database are represented by dots for $n = 1$ (blue), 2 (green), 3 (red), and the corresponding predictions from $H_A$ (curves in red, except for $n = 3$ in blue).

# $\rho_n(X)$ : Height $X$, Selrank $n$, MW vs Ш

**In our model:** The probability that a test Selmer element $s_E \in \mathrm{Sel}_2(E)$, for $E \in \widetilde{\mathcal{S}}^X$, is a MW element, is given by $\rho_n(X)$, where $\rho_n(X)$ is a function such that $\lim_{X\to\infty} \rho_n(X) = 0$.

## Hypothesis B ($H_B$)

*For each* $1 \le i \le \lfloor n/2 \rfloor$*, let* $Y_i : \widetilde{\mathcal{S}}_n^X \to \{0,1\}$ *be the function that takes values*

$$Y_i(E) = \begin{cases} 1 & \text{if } s_{E,i} \text{ is a MW element,} \\ 0 & \text{otherwise,} \end{cases}$$

*where* $\mathrm{Sel}_2(E) = (s_{E,1}, \ldots, s_{E,\lfloor n/2 \rfloor})$*. Then,* $Y_i$ *is a random variable that follows a Bernoulli distribution with probability* $\rho_n(X)$*, and* $\lim\limits_{X\to\infty} \rho_n(X) = 0$*.*

[*] *There is an additional "equicorrelation" condition. Warning! The variables* $Y_i$ *and* $Y_j$ *are not necessarily independent.*

Recall: $\mathrm{rank}(E) = (n \bmod 2) + 2 \cdot \#\{\text{MW elements in } \mathrm{Sel}_2(E)\}$.

## Corollary

*Let $E_1, \ldots, E_m$ be distinct (non-isomorphic) test elliptic curves chosen independently of Selmer rank n and heights $X_1, \ldots, X_m$. Then, the expected value of the average rank is*

$$\mathbb{E}\left(\frac{1}{m}\sum_{i=1}^{m}\mathrm{rank}(E_i)\right) = (n \bmod 2) + \frac{2\lfloor n/2 \rfloor}{m}\sum_{i=1}^{m}\rho_n(X_i)$$

*with standard error given by*

$$\frac{1}{m}\sqrt{\lfloor n/2 \rfloor \sum_{i=1}^{m}(\rho_n(X_i)(1-\rho_n(X_i)) + (\lfloor n/2 \rfloor - 1)C_{1,1}^{n}(X_i))},$$

*where $C_{1,1}^{n}(X_i)$ is a certain covariance factor.*

# $\rho_n(X)$ : Height $X$, Selrank $n$, MW vs Ш

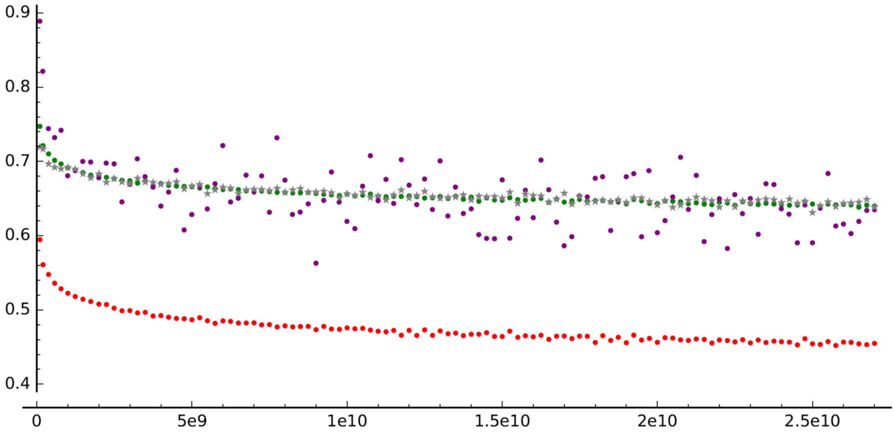



Figure: Graphs of the ratios $\rho_n(X)$ for $n = 2$ (green), 3 (red), 4 (gray stars), 5 (purple), based on the BHKSSW data.

## Refined Hypothesis B

*Hypothesis $H_B$ holds and, for every $n \geq 2$, there are constants $D_n$ and $f_n$ such that*

$$\rho_n(X) = \frac{D_n}{X^{f_n}}.$$

| $n$ | 2 | 3 | 4 | 5 |
|---|---|---|---|---|
| $D_n$ | 1.12465347 | 1.30937016 | 1.07928016 | 1.79161787 |
| $f_n$ | 0.02344245 | 0.04412662 | 0.02158211 | 0.04383626 |

Table: The coefficients of the best-fit linear regression $\rho_n(X) \approx D_n/X^{f_n}$.

# $\rho_n(X)$ : Height $X$, Selrank $n$, MW vs Ш

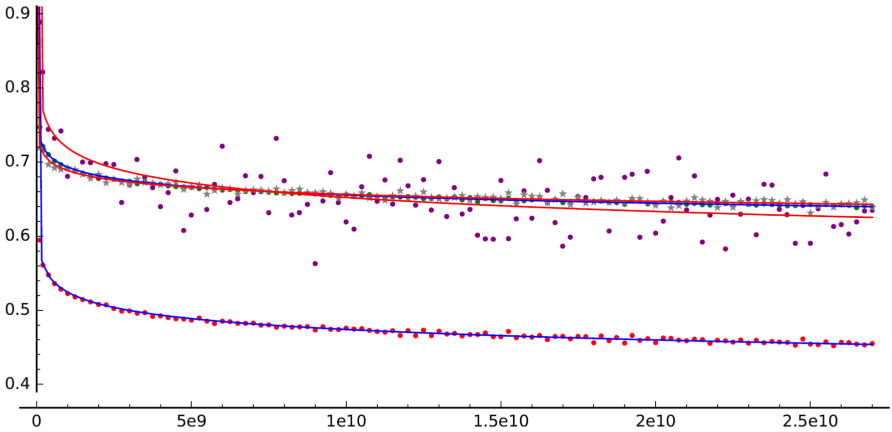



Figure: Graphs of the ratios $\rho_n(X)$ for $n = 2$ (green), 3 (red), 4 (gray stars), 5 (purple), and the corresponding models of the form $D_n/X^{f_n}$ (in blue for $n = 2, 3$ and red for $n = 4, 5$).

Here we consider the average rank contributions from the subsets of elliptic curves of each Selmer rank $n \geq 1$:

$$\mathrm{AveRank}_{\mathcal{S}_n}(X) = \frac{\sum_{E \in \mathcal{S}_n(X)} \mathrm{rank}(E(\mathbb{Q}))}{\pi_{\mathcal{E}}(X)}.$$

## Theorem

*Assume $H_A$ and $H_B$, and let $n \geq 1$ be fixed. Then, the expected value of $\mathrm{AveRank}_{\widetilde{\mathcal{S}}_n}(X)$ is given by*

$$\frac{5\kappa}{6\pi_{\mathcal{E}}(X)} \cdot \int_0^X \frac{\theta_n(H)}{H^{1/6}} \left( (n \bmod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H) \right) dH + O(X^{-1/3}).$$

*Moreover, the error in approximating $\mathrm{AveRank}_{\widetilde{\mathcal{S}}_n}(X)$ by its expected value is approximately given by*

$$\sqrt{\frac{5\kappa\lfloor n/2 \rfloor}{6\pi_{\mathcal{E}}(X)^2} \int_0^X \frac{\theta_n(H)}{H^{1/6}} (\rho_n(H)(1 - \rho_n(H)) + (\lfloor n/2 \rfloor - 1) C_{1,1}^n(H)) \, dH + O(X^{-7/6})}$$

## Corollary

*If we assume the refined versions of $H_A$ and $H_B$, then there are constants $\tau_n$ such that the expected value of $\mathrm{AveRank}_{\widetilde{\mathcal{E}}}(X)$ is given by*

$$\mathbb{E}\left(\sum_{n=1}^{\infty} \mathrm{AveRank}_{\widetilde{\mathcal{S}}_n}(X)\right) =$$

$$\sum_{n=1}^{\infty} s_n \cdot \left( \frac{\tau_n}{X^{5/6}} + \sum_{m=0}^{\infty} \left( \frac{(n \bmod 2)(-C_n)^m}{1-(6/5)me_n} + X^{-f_n} \frac{2\left\lfloor \frac{n}{2} \right\rfloor D_n(-C_n)^m}{1-(6/5)(f_n+me_n)} \right) X^{-me_n} \right)$$

$+O(X^{-1/3})$ *with standard error* $\leq \dfrac{\sum_{n=2}^{\infty} \sqrt{\lfloor n/2 \rfloor \cdot (\lfloor n/2 \rfloor - 3/4) \cdot s_n}}{\sqrt{\kappa} X^{5/12}}$.

*In particular,*

$$\lim_{X\to\infty} \mathrm{AveRank}_{\widetilde{\mathcal{E}}}(X) = \sum_{n=1}^{\infty} s_n \cdot (n \bmod 2) = \sum_{k=0}^{\infty} s_{2k+1} = \frac{1}{2},$$

*in the sense that the expected value goes to* $1/2$ *with standard error going to* $0$ *as* $X \to \infty$.

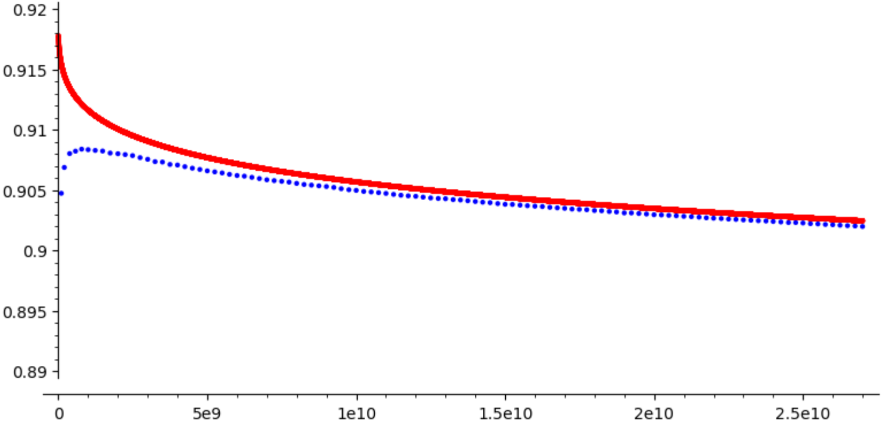



Figure: Values of $\text{AveRank}_{\mathcal{E}}(X)$ from the BHKSSW database (blue dots), and numerical integration of the approximation given in Corollary (in red).

According to the database, we have $\text{AveRank}_{\mathcal{E}}(2.7 \cdot 10^{10}) = 0.90197580$ while our approximation gives 0.90244770. Thus, the absolute error is 0.00047189 (note $(2.7 \cdot 10^{10})^{-1/3} \approx 0.0003$), which is a 0.0523% of the value.

| $X$ | $\sum_{n=1}^{5} \text{AveRank}_{\mathcal{S}_n}(X)$ | $X$ | $\sum_{n=1}^{5} \text{AveRank}_{\mathcal{S}_n}(X)$ |
|---|---|---|---|
| $10^{10}$ | 0.905665 | $10^{50}$ | 0.548880 |
| $10^{15}$ | 0.846828 | $10^{75}$ | 0.512531 |
| $10^{20}$ | 0.766868 | $10^{100}$ | 0.503256 |
| $10^{30}$ | 0.649901 | $10^{150}$ | 0.500215 |
| $10^{40}$ | 0.585108 | $10^{200}$ | 0.500006 |

Table: Conjectural approximate values of $\sum_{n=1}^{5} \text{AveRank}_{\mathcal{S}_n}(X)$ obtained using numerical integration of the formulas. The integration was done with SageMath, which reports an absolute error in the numerical integration less than $4 \cdot 10^{-7}$ in all cases. The limit should be $s_1 + s_3 + s_5 = 0.49999965\ldots$.

# Thank You

**References:**


1. Balakrishnan, Jennifer; Ho, Wei; Kaplan, Nathan; Spicer, Simon; Stein, William; Weigandt, James, *Databases of elliptic curves ordered by height and distributions of Selmer groups and ranks.* LMS J. Comput. Math. 19 (2016), suppl. A, 351-370.
2. Bektemirov, Baur; Mazur, Barry; Stein, William; Watkins, Mark, *Average ranks of elliptic curves: tension between data and conjecture.* Bull. Amer. Math. Soc. (N.S.) 44 (2007), no. 2, 233-254.
3. Bhargava, Manjul; Shankar, Arul, *The average size of the 5-Selmer group of elliptic curves is 6, and the average rank is less than 1*, arXiv:1312.7859.
4. Lozano-Robledo, Álvaro, *A probabilistic model for the distribution of ranks of elliptic curves over* $\mathbb{Q}$, arXiv:1611.01999 (to appear in JNT).
5. Rubin, Karl; Silverberg, Alice, *Ranks of elliptic curves*. Bull. Amer. Math. Soc. (N.S.) 39 (2002), no. 4, 455-474.
6. Silverberg, Alice, *Ranks "Cheat Sheet"*, WIN 2: Research Directions in Number Theory, Contemporary Mathematics 606, AMS-CRM (2013).