A PROBABILISTIC MODEL FOR THE DISTRIBUTION OF RANKS OF ELLIPTIC CURVES OVER \( \mathbb{Q} \)

ÁLVARO LOZANO-ROBLEDÒ

Abstract. In this article, we propose a new probabilistic model for the distribution of ranks of elliptic curves in families of fixed Selmer rank, and compare the predictions of our model with previous results, and with the databases of curves over the rationals that we have at our disposal. In addition, we document a phenomenon we refer to as Selmer bias that seems to play an important role in the data and in our models.

1. Introduction

Let \( E/\mathbb{Q} \) be an elliptic curve. The Mordell–Weil theorem states that the group \( E(\mathbb{Q}) \) of rational points on \( E \) is finitely generated and, therefore, we have an isomorphism

\[
E(\mathbb{Q}) \cong E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^{R_E},
\]

where \( E(\mathbb{Q})_{\text{tors}} \) is the (finite) subgroup of points of finite order, and \( R_E = \text{rank}(E(\mathbb{Q})) \geq 0 \) is the rank of the elliptic curve. The torsion subgroups that arise over \( \mathbb{Q} \) are well understood: Mazur’s theorem settles what groups are possible ([21], [22]), the parametrization of the corresponding modular curves are known ([20]), and we know the distribution of elliptic curves with a prescribed torsion subgroup ([15]) as a function of the height of the curve. However, the distribution of ranks of elliptic curves is unknown. Several conjectures can be found in the literature (e.g., on the average rank, see [24]), and also some heuristic models ([29], [23]), but the basic questions about the distribution of the ranks remain unanswered. For instance, it is not known whether the rank can be arbitrarily large (currently, the largest rank known is 28, due to Noam Elkies - see [11] for Elkies’ example, and other current records).

In this article, we propose a new probabilistic model for the distribution of ranks of elliptic curves (in families of fixed 2-Selmer rank) and explore its possible consequences. The model itself is built on a probability space of test elliptic curves and test Selmer elements in the spirit of Cramér’s model for the prime numbers (see [6], [13]). As such, our model is a collection \( T \) of all possible sequences of (finite) sets of test elliptic curves of each height (with certain growth conditions as the height grows). The sequence of ordinary elliptic curves \( \mathcal{E} \) over \( \mathbb{Q} \) belongs to this class, and we make predictions about \( \mathcal{E} \) from the asymptotic average behavior from sequences in \( T \) under the assumption of certain probabilistic hypotheses (see Sections 1.3, 5, and 7 for more details). We use the largest database of elliptic curves at our disposal ([1]), which we will refer to as the BHKSSW database in order to test our model and to make predictions. We concentrate on elliptic curves over \( \mathbb{Q} \) because there are no analogous databases for any other number field \( K \) or function field \( \mathbb{F}(T) \) to test the model, but the same ideas would apply more generally for \( p \)-Selmer groups of abelian varieties over \( K \) or

\[\text{2010 Mathematics Subject Classification. Primary: 11G05, Secondary: 14H52.}\]
The set of all test elliptic curves will be denoted by $\tilde{T}$, and we define the conditional probability $\text{Prob}(E \in \tilde{R}_r^X | E \in S_{r+2j}^X)$ as 0 if $S_{r+2j}^X$ is empty, and by $\# R_r^X \cap S_{r+2j}^X / \# S_{r+2j}^X$ otherwise. In Section 3 we will discuss the known results about the number of elliptic curves up to height $X$.

### 1.2. Notation and setup for test elliptic curves

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

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

The set of all test elliptic curves will be denoted by $\tilde{E}$, those test curves with height $X$ will be $\tilde{E}_X$, and those test curves with height $X$ and Selmer rank $n$ will be denoted by $\tilde{S}_n^X$. We define $\tilde{R}_r^X$ similarly. We let $T$ be a space of sequences of (finite) subsets of $\tilde{E}_X$ with certain growth conditions, defined as follows:

$$T = \left\{ (\tilde{T}_X)_{X \geq 1} : \tilde{T}^X \subseteq \tilde{E}^X, \sum_{N=1}^{X} \# \tilde{T}^N = \kappa X^{5/6} + O(X^{1/2}) \right\},$$

where $\kappa = 2^{4/3} \cdot (\zeta(10) \cdot 3^{3/2})^{-1}$. To each elliptic curve $E$ we can associate a test elliptic curve (Remark 5.2) and the sequence $E = (E^X)$ of ordinary elliptic curves belongs to $T$. Thus, the goal is to predict the behaviour of $E$ from the average asymptotic behaviour of sequences in $T$.

We also need to introduce counting notation for test elliptic curves: if $I$ is a finite interval in $[1, \infty)$, and $\tilde{T} \in T$, we will write $\tilde{T}(I) = \bigcup_{X \in I} \tilde{T}^X$, and $\tilde{T}S_n(I) = \bigcup_{X \in I} \tilde{T}^X \cap \tilde{S}_n^X$. Finally, we
1.3. Probability spaces. In Sections 5 and 7, and for fixed $n \geq 0$ and $X \geq 1$, we state two probabilistic hypotheses, $H_A$ and $H_B$ stated in Hypotheses 5.6 and 7.7 respectively. These hypotheses make $\tilde{E}^X$ and $\tilde{S}_n^X$ into probability spaces:

(H$_A$) **Hypothesis A:** Informally, the probability of drawing a test elliptic curve of Selmer rank $n$ out of the bin $\tilde{E}^X$ is given by a function $\theta_n(X)$. Formally, the function $Y_{\text{Sel},n,X} : \tilde{E}^X \rightarrow \{0,1\}$ such that $Y_{\text{Sel},n,X}(E) = 1$ if $E \in \tilde{S}_n^X$, and $Y_{\text{Sel},n,X}(E) = 0$ otherwise, is a random variable with Bernoulli distribution $B(1, \theta_n(X))$, where $\theta_n(X)$ is a function that depends on $n$ and $X$. In particular, this implies that the expected value $E(Y_{\text{Sel},n,X})$ is $\text{Prob}(E \in \tilde{S}_n^X) = \theta_n(X)$.

(H$_B$) **Hypothesis B:** Let $E \in \tilde{S}_n^X$ be chosen at random. Informally, the probability that the $i$-th coordinate of $\text{Sel}_2(E) = (s_{E,1}, \ldots, s_{E,[n/2]})$ is a MW element is given by a function $\rho_n(X)$ (that does not depend on $i$ or $E$). Formally, for each $1 \leq i \leq [n/2]$, the function $Y_i : \tilde{S}_n^X \rightarrow \{0,1\}$ that takes the value 1 whenever $s_{E,i}$ is a MW element, and 0 otherwise, is a random variable with Bernoulli distribution $B(1, \rho_n(X))$, where $\rho_n(X)$ is a function that depends on $n$ and $X$, but not on $i$ (however, the variables $Y_i$ are not independent in general). From the distribution of the variables $Y_i$ we shall recover the conditional probability $\text{Prob}(E \in \tilde{R}_r^X \mid E \in \tilde{S}_n^X)$ for any $0 \leq r \leq n$ with $n \equiv r \mod 2$ (see Corollary 8.8).

After taking all the available data under consideration (mainly [1]), we formulate a refinement of the model which specifies the shape of $\theta_n(X)$ and $\rho_n(X)$ up to some constants (which are Hypotheses 6.3 and 8.13):

(H$_C$) **Hypothesis C:** Assume $H_A$ and $H_B$. Then, there are constants $C_n$, $D_n$, $e_n$, $f_n$, for each $n \geq 1$, such that

$$\theta_n(X) = \frac{s_n}{1 + C_n X^{-e_n}}, \quad \text{and} \quad \rho_n(X) = \frac{D_n}{X^{f_n}},$$

where the limit values $s_n$ of $\theta_n(X)$ are those given by a conjecture of Poonen and Rains, and all constants are positive except $C_1 < 0$.

The data suggest that for the family of all elliptic curves over $\mathbb{Q}$ the values of the constants of Hypothesis C, for $n = 1, \ldots, 5$, are as given in Tables 5 and 10, and the limit values $s_n$ are discussed in Section 6 (as in [24]). We have also investigated the suitability of the model in the subfamily of curves with $j = 1728$ (see Remark 8.18).

1.4. Summary of results. In our results, we give the expected value and asymptotic behavior of a random sequence $\tilde{T}$ in $T$ under the probabilistic hypotheses $A$, $B$, and $C$. In our main Theorems 9.1 and 10.2, under the assumption of $H_A$ and $H_B$, we provide formulas for $\pi_{\tilde{T},r \cap \tilde{S}_n}(X)$, i.e., the number of test elliptic curves in $\tilde{T}$ of rank $r$ and Selmer rank $n$ up to height $X$, and also for the contribution to the average rank coming from test elliptic curves of Selmer rank $n$. 

\[
\pi_{T}(I) = \#\tilde{T}(I) = \sum_{X \in I} \#\tilde{T}^X \quad \text{and} \quad \pi_{\tilde{T},S_n}(I) = \#\tilde{T}S_n(I) = \sum_{X \in I} \#(\tilde{T}^X \cap \tilde{S}_n^X).
\]
Theorem 1.1 (also Theorem 9.1). Let \( \mathcal{T} \in \mathbf{T} \) be arbitrary, and let \( X, r \geq 0, j \geq 0 \) be fixed, such that \( n(j) = r + 2j \geq 1 \). If we assume Hypothesis C, then

\[
\pi_{\mathcal{T}, r, n(j)}(X) = \frac{5\kappa}{6} \left( \left[ \frac{r}{2} \right] + j \right) \int_0^X \frac{\theta_n(H)}{H^{1/6}} \cdot \mathbb{E}_{\lfloor \frac{r}{2} \rfloor + j}^{n(j)}(H) \, dH + O(X^{1/2})
\]

\[
= \frac{5\kappa}{6} \left( \left[ \frac{r}{2} \right] + j \right) \int_0^X \frac{s_n(j) \cdot \mathbb{E}_{\lfloor \frac{r}{2} \rfloor + j}^{n(j)}(H)}{(1 + C_{n(j)}H^{-e_n(j)}) \cdot H^{1/6}} \, dH + O(X^{1/2}),
\]

where \( \kappa = 2^{4/3} \cdot (\zeta(10) \cdot 3^{3/2})^{-1} \), and \( \mathbb{E}_{\lfloor \frac{r}{2} \rfloor + j}^{n(j)}(H) \) is the expected value defined in Remark 8.6.

In Corollary 9.4 we specialize the formulas of \( \pi_{\mathcal{T}, r, n(j)}(X) \) for \( 0 \leq r \leq n \leq 5 \) (see also Table 13). Using our formulas, we have computed approximations of \( \pi_{\mathcal{T}, r}(X) \) for \( 1 \leq r \leq 5 \) in the range \([0, 2.7 \cdot 10^{10}]\), and plotted them in Figures 16 and 17. The error in our approximations is less than 0.7% in this range (see Table 14).

Our second theorem gives formulas for the contribution to the average rank of test elliptic curves coming from curves of each Selmer rank \( n \geq 1 \). Then, the contributions are added up to estimate the behavior of the average rank.

Theorem 1.2 (also Theorem 10.2 and Corollary 10.4). Let \( \mathcal{T} \in \mathbf{T} \) be arbitrary. Assume \( H_A \) and \( H_B \), and let \( n \geq 1 \) be fixed. Then, the expected value of \( \text{AvgRank}_{\mathcal{T}, S_n}(X) = \frac{1}{\pi_{\mathcal{T}}(X)} \cdot \sum_{E \in \mathcal{T}, S_n} \text{rank}(E) \) is given by

\[
\frac{5\kappa}{6\pi_{\mathcal{T}}(X)} \cdot \int_1^X \frac{\theta_n(H)}{H^{1/6}} \left( (n \mod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H) \right) \, dH + \theta_n(X) \cdot O(X^{-1/3}).
\]

Moreover, the error in approximating \( \text{AvgRank}_{\mathcal{T}, S_n}(X) \) by its expected value is given by

\[
\sqrt{\frac{5\kappa}{6\pi_{\mathcal{T}}(X)}} \cdot \int_1^X \frac{\theta_n(H)}{H^{1/6}} \left( \rho_n(H)(1 - \rho_n(H)) + \left( \left\lfloor \frac{n}{2} \right\rfloor - 1 \right) C_{1, 1}^n(H) \right) \, dH + O(X^{-7/6}),
\]

where \( C_{1, 1}^n(X) \) is the covariance function defined in Proposition 8.2. Further, there are constants \( \tau_n \) such that the expected value of \( \text{AvgRank}_{\mathcal{T}}(X) = \sum_{n=1}^{\infty} \text{AvgRank}_{\mathcal{T}, S_n}(X) \) is given by

\[
\sum_{n=1}^{\infty} s_n \cdot \left( \frac{\tau_n}{X^{5/6}} + \sum_{m=0}^{\infty} \frac{(n \mod 2)(-C_n)^m}{1 - (6/5)\text{me}_n} + f_n^{-\frac{2}{3}} \frac{|D_n(-C_n)^m|}{1 - (6/5)(f_n + \text{me}_n)} X^{-\text{me}_n} \right) + O(X^{-1/3}).
\]

In particular,

\[
\lim_{X \to \infty} \text{AvgRank}_{\mathcal{T}}(X) = \sum_{k=0}^{\infty} s_{2k+1} = \frac{1}{2},
\]

with standard error going to 0 as \( X \to \infty \).

In particular, Theorem 1.2 says that our assumptions imply the so-called “50% – 50% conjecture” (see Conjecture 10.1) and, moreover, it predicts not only the 1/2 limit of the average rank, but also a rate of convergence to said limit. We have used our formulas to compute an approximation of \( \text{AvgRank}_{\mathcal{T}}(X) \approx \sum_{n=1}^{5} \text{AvgRank}_{S_n}(X) \) in the range \([0, 2.7 \cdot 10^{10}]\) and plotted it in Figure 18. The error in our approximation of \( \text{AvgRank}_{\mathcal{T}}(2.7 \cdot 10^{10}) \) is 0.0523% of the actual value (see Remark 10.5).
In Table 1, and under the assumption of Hypothesis C, we have computed approximate values of \(\text{AvgRank}_{\widetilde{T}}(X) \approx \sum_{n=1}^{5} \text{AvgRank}_{\widetilde{T}S_n}(X)\) using numerical integration of the formulas of Theorem 1.2.

Our Hypothesis A also implies a formula for the average 2-Selmer rank of a test elliptic curve.

**Theorem 1.3** (Also Prop. 6.8). Let \(\widetilde{T} \in T\) be arbitrary. Let \(\text{AvgSelRank}(X)\) be defined by

\[
\text{AvgSelRank}(X) = \frac{1}{\pi_{\widetilde{T}}(X)} \sum_{E \in \widetilde{T}(X)} \text{selrank}(E).
\]

If we assume \(H_A\) and we assume that \(0 \leq \theta_n(X) \leq s_n\) for all \(n \geq 2\) and all \(X > 0\), then the expected value of the average Selmer rank is given by

\[
\mathbb{E}(\text{AvgSelRank}(X)) = \frac{5/6}{X^{5/6}} \int_{1}^{X} \sum_{n \geq 1} n \cdot \theta_n(H) \frac{n \cdot \theta_n(H)}{H^{1/6}} dH + O \left(X^{-1/3}\right).
\]

Moreover, \(\lim_{X \to \infty} \mathbb{E}(\text{AvgSelRank}(X)) = \sum_{n \geq 1} n \cdot s_n = 1.26449978 \ldots\)

| \(X\)  | \(\sum_{n=1}^{5} \text{AvgRank}_{\widetilde{T}S_n}(X)\) | \(X\)  | \(\sum_{n=1}^{5} \text{AvgRank}_{\widetilde{T}S_n}(X)\) |
|-------|---------------------------------|-------|---------------------------------|
| \(10^{10}\) | 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 1.** Approximate values of \(\sum_{n=1}^{5} \text{AvgRank}_{\widetilde{T}S_n}(X)\) obtained using numerical integration of the formulas of Theorem 10.2. The integration was done with SageMath, which reports an absolute error in the numerical integration less than \(4 \cdot 10^{-7}\) in all cases. By Theorem 10.2, the limit should be \(s_1 + s_3 + s_5 \approx 0.49999965\).

Finally, a question on Selmer rank bias arises in our work:

**Question 1.4.** Does the expected value of \(Y_{\text{Hasse},n,X}(s_E)\) depend on \(n\)? In other words, does the probability that \(s_E \in \text{Sel}_2(E/Q)\) is globally solvable depend on \(n = \text{selrank}(E(Q))\)?

The answer, surprisingly, seems to be that the probability does depend not only on the parity of \(n\), but also on the value of \(n\) itself (see Fig. 9). For instance, the data suggest that an element of \(\text{Sel}_2(E/Q)\) is significantly more likely to be globally solvable for \(n = 5\) than for \(n = 3\). However, the probabilities for \(n = 2\) and \(n = 4\) are quite similar in the height interval \([0, 2.7 \cdot 10^{10}]\) (but they do not behave identically).
Remark 1.5. In this article we work with elliptic curves over $\mathbb{Q}$ and 2-Selmer groups because the database we have to test our models ([1]) only contains 2-Selmer information. However, the same probabilistic model could be derived for $p$-Selmer groups over a global field $K$.

Acknowledgements. The author would like to thank Jennifer Balakrishnan, Iddo Ben-Ari, Keith Conrad, Harris Daniels, Wei Ho, Jennifer Park, Ari Shnidman, Drew Sutherland, and John Voight for their helpful conversations, comments, and suggestions.

2. Notation and Probability

In Table 2 we include a glossary of notation defined throughout the paper, together with a reference. We also recall here a few definitions of probability concepts for the convenience of the reader. We say that a random variable $Y$ follows a Bernoulli distribution $B(1, p)$, or $Y \sim B(1, p)$, if $Y$ takes the value 1 with success probability of $p$ and the value 0 with probability $1 - p$. The binomial distribution $B(n, p)$ is the discrete probability distribution of the number of successes in a sequence of $n$ independent yes/no experiments, each of which yields success with probability $p$. The expected value and variance of a discrete random variable $Y$ that takes values $y_1, \ldots, y_k$ with probability $p_1, \ldots, p_k$ are defined respectively by

$$
E(Y) = \sum_{i=1}^{k} y_i \cdot p_i, \quad \text{Var}(Y) = E(Y^2) - (E(Y))^2.
$$

The covariance of two random variables $V, W$ is given by

$$
\text{Cov}(V, W) = E(VW) - E(V) \cdot E(W).
$$

If $\text{Cov}(V, W) = 0$, then we say that $V$ and $W$ are uncorrelated random variables. If $V$ and $W$ are independent random variables, then $E(VW) = E(V)E(W)$ and, in particular, $\text{Cov}(V, W) = 0$. Also, we note here that if $a$ and $b$ are constants, then

$$
\text{Var}(aV + bW) = a^2 \text{Var}(V) + b^2 \text{Var}(W) + 2ab \text{Cov}(V, W).
$$

Finally, the standard error of the mean (SEM) of random variables $Y_1, \ldots, Y_m$ is an estimator for the accuracy of the approximation of $\frac{1}{m} \sum Y_i$ by $\frac{1}{m} \sum E(Y_i)$, and it is defined as the square root of the variance of the mean of the variables. In other words, the standard error is given by

$$
\sqrt{\text{Var} \left( \frac{1}{m} \sum_{i=1}^{m} Y_i \right)}.
$$

If $Y_1, \ldots, Y_m$ are $m$ independent random variables following the same distribution with mean $\mu$ and standard deviation $\sigma$, then $
\text{SEM}(Y_1, \ldots, Y_m) = \left( \frac{1}{m^2} \sum \text{Var}(Y_i) \right)^{1/2} = (\text{Var}(Y_i)/m)^{1/2} \sigma/\sqrt{m}.
$

3. The number of elliptic curves with (naive) height $\leq X$

Let $E/\mathbb{Q}$ be an elliptic curve. We shall write each elliptic curve in a short Weierstrass model of the form $y^2 = x^3 + Ax + B$ with $A, B \in \mathbb{Z}$ and $0 \neq 4A^3 + 27B^2$ such that $\Delta_E$ is minimal in absolute value (minimal among all short Weierstrass models isomorphic to $E$ over $\mathbb{Q}$). In other words, we will be working with the set of elliptic curves

$$
\mathcal{E} = \{ E_{A, B} : y^2 = x^3 + Ax + B \mid A, B \in \mathbb{Z}, 4A^3 + 27B^2 \neq 0, \text{ and if } d^4 | A, \ d^6 | B, \text{ then } d = \pm 1 \}.
$$
DISTRIBUTION OF RANKS OF ELLIPTIC CURVES

\[ \mathcal{E} \] Set of elliptic curves over \( \mathbb{Q} \) up to isomorphism \( \S 3 \)

\[ \text{selrank}(E(\mathbb{Q})) \] 2-Selmer rank, equal to \( \dim_{\mathbb{F}_2} \text{Sel}_2(E/\mathbb{Q}) - \dim_{\mathbb{F}_2}(E[2]) \) \( \S 1, 6 \)

\[ \mathcal{S}_n \] For \( n \geq 0 \), curves \( E \in \mathcal{E} \) with \( \text{selrank}(E(\mathbb{Q})) = n \) \( \S 6 \)

\[ \mathcal{R}_r \] For \( r \geq 0 \), curves \( E \in \mathcal{E} \) with \( \text{rank}(E(\mathbb{Q})) = r \) \( \S 9 \)

\( \tilde{E}, \tilde{\mathcal{S}}_n, \tilde{\mathcal{R}}_r \) Test elliptic curves \( (\tilde{E}) \), of Selmer rank \( n \) \( (\tilde{\mathcal{S}}_n) \), of MW rank \( r \) \( (\tilde{\mathcal{R}}_r) \) \( \S 5 \)

\[ \text{ht}(E) \] The naive height of an elliptic curve \( \S 3 \)

\[ C(X) \] For \( X \geq 0 \), curves in \( C = \mathcal{E}, \mathcal{R}_r, \) or \( \mathcal{S}_n \), with (naive) height \( \leq X \) \( \S 3, 6, 9 \)

\[ C(I) \] For an interval \( I \), curves in \( C \) with height in \( I \) \( \S 3, 6, 9 \)

\[ C_X \] For \( X \geq 0 \), curves in \( C \) with height exactly \( X \) \( \S 3, 6, 9 \)

\[ \pi_{C}(X) \] For a set \( C \subseteq \mathcal{E} \), the size of \( C \cap \mathcal{E}(X) \), where \( \mathcal{C} = \mathcal{E}, \mathcal{R}_r, \) or \( \mathcal{S}_n \) \( \S 3, 6, 9 \)

\[ \pi_{C}(I) \] For a set \( C \subseteq \mathcal{E} \) and an interval \( I \), the size of \( C \cap \mathcal{E}(I) \) \( \S 3, 6, 8 \)

\[ \kappa \] Constant equal to \( 2^{4/3} \cdot (\zeta(10) \cdot 3^{3/2})^{-1} \approx 0.484462004349 \ldots \) Thm. 3.1

\[ s_n \] \( \lim_{X \to \infty} \pi_{\mathcal{S}_n}(X)/\pi_{\mathcal{E}}(X) \), given by a conjectural formula by \([24]\) \( \S 6 \)

\[ B(m, p) \] Binomial distribution with \( m \) experiments and probability \( p \) \( \S 6 \)

\[ Y_{\text{Sel}, n, X}(E/\mathbb{Q}) \] Random variable with value 1 if \( \text{selrank}(E(\mathbb{Q})) = n \), and 0 otherwise Hyp. 5.6

\[ \theta_n(X) \] The function giving the expected value of \( Y_{\text{Sel}, n, X}(E/\mathbb{Q}) \) Hyp. 5.6

\[ \theta_n(X, N) \] Moving ratio defined by \( \pi_{\mathcal{S}_n}((X, X + N))/\pi_{\mathcal{E}}((X, X + N)) \) Cor. 6.2

\[ Y_{\text{Hasse}, n, X}(s_E) \] Random variable with value 1 if \( s_E \equiv 0 \in \text{III}(E/\mathbb{Q}) \), and 0 otherwise Hyp. 7.7

\[ \rho_n(X) \] The function giving the expected value of \( Y_{\text{Hasse}, n, X}(s_E) \) Hyp. 7.7

\[ \rho_n(X, N) \] Moving ratio approximating \( \rho_n(X) \) Def. 8.11

\[ C_{n,t}^s(X) \] Covariance function of a certain products of random variables Prop. 8.2

\[ E_{n,t}^s(X) \] Expected value of a certain product of random variables Rem. 8.6

Table 2. Notation defined and used throughout the paper.

\[
\text{ht}(E_{A,B}) = \max\{4|A|^3, 27B^2\},
\]

as used in [1], [4], and [23]. The BHKSSW database ([1]) contains data for all 238,764,310 elliptic curves up to height 26,998,673,868 \( \approx 2.7 \cdot 10^{10} \). While working on this project, we have gathered
data for the curves \( y^2 = x^3 + Ax \), for all fourth-power-free integers \( A \in [1, 10^6] \), that is, about a million curves with \( j = 1728 \), up to height \( 4 \cdot 10^{18} \).

For each positive real number \( X \), we define \( \mathcal{E}(X) = \{ E \in \mathcal{E} : \text{ht}(E) \leq X \} \), and \( \pi_{\mathcal{E}}(X) = \#\mathcal{E}(X) \). Similarly, if \( 0 \leq X_1 \leq X_2 \), we shall write \( \mathcal{E}([X_1, X_2]) \) for the set \( \{ E \in \mathcal{E} : X_1 \leq \text{ht}(E) \leq X_2 \} \) and \( \pi_{\mathcal{E}}([X_1, X_2]) = \#\mathcal{E}([X_1, X_2]) \) for its size (in particular, \( \mathcal{E}^X = \mathcal{E}([X, X]) \) denotes the elliptic curves of height exactly \( X \), a set that can be empty depending on the value of \( X \)). We adapt a result of Brumer ([4]) that estimates the value of \( \pi_{\mathcal{E}}(X) \) to our choice of height function.

**Theorem 3.1 ([4, Lemma 4.3]).** The number of elliptic curves of height up to \( X \) satisfies \( \pi_{\mathcal{E}}(X) = \kappa X^{5/6} + O(X^{1/2}) \) where the constant \( \kappa = 2^{4/3} \cdot (\zeta(10) \cdot 3^{3/2})^{-1} \approx 0.484462004349 \).

**Remark 3.2.** Using the BHKSSW database, we have calculated the values of \( \pi_{\mathcal{E}}(X) \) up to \( 2.7 \cdot 10^{10} \) in \( 0.25 \cdot 10^9 \) intervals. We have found (using SageMath, [28]) the best-fit model of the form \( C \cdot X^{5/6} \) for these data points, and found that the best constant is \( C \approx 0.48447036 \) in agreement with Brumer’s constant (\( C \) and \( \kappa \) differ by \( 8.35 \cdot 10^{-6} \)).

![Graph](image.png)

**Figure 1.** Values of \( \pi_{\mathcal{E}}(X) \) from the BHKSSW database (blue dots), and the function \( 0.48447036 \cdot X^{5/6} \) (in red).

**Remark 3.3.** According to Theorem 3.1, the number of curves in the height interval \( (X, X + N] \) is, approximately,

\[
\pi_{\mathcal{E}}((X, X + N]) = \pi_{\mathcal{E}}(X + N) - \pi_{\mathcal{E}}(X) \\
\approx \kappa \cdot ((X + N)^{5/6} - X^{5/6}) \\
= \frac{5\kappa}{6} \int_X^{X+N} \frac{1}{H^{1/6}} dH \approx \frac{5\kappa}{6} \cdot \frac{N}{X^{1/6}},
\]

\( \kappa \) is the constant from Theorem 3.1.
where $5\kappa/6 \approx 0.403718336957$, and the last approximation is valid for large $X$ such that $X \gg N \geq 0$. However, the error in this approximation is still of the order $O(X^{1/2})$, so the error can be quite large. For instance,

$$
\pi([2 \cdot 10^{10}, 2 \cdot 10^{10} + 0.25 \cdot 10^9]) = 1,955,593 \approx 1,937,225.394 ... = \frac{5\kappa}{6} \cdot \frac{0.25 \cdot 10^9 + 1}{(2 \cdot 10^{10})^{1/6}}.
$$

$$
\pi([2.5 \cdot 10^{10}, 2.5 \cdot 10^{10} + 0.25 \cdot 10^9]) = 1,852,352 \approx 1,866,502.107 ... = \frac{5\kappa}{6} \cdot \frac{0.25 \cdot 10^9 + 1}{(2.5 \cdot 10^{10})^{1/6}}.
$$

Nonetheless, we shall prove below (Corollary 3.4) that the approximation $\pi_{\xi}((X, X + N]) \approx \frac{5\kappa}{6} \cdot \frac{N}{X^{1/6}}$ works on average with error going to zero as $X$ goes to infinity (as long as $X > N^2$).

We also point out here that if we want $\pi_{\xi}((X, X + N])$ to be approximately constant as $X \to \infty$, then we need $N = N(X) \asymp C \cdot X^{1/6}$. For instance, if we want $\pi_{\xi}((X, X + N(X)] \approx 10^i$, then we should have $N = N(X) = (6 \cdot 10^i/5\kappa) \cdot X^{1/6}$, where $6/5\kappa \approx 2.476974436029$.

**Corollary 3.4.** Let $N \geq 1$ be fixed, and suppose $X > N^2$. Let $f(X)$ be a function such that $f(X) = \kappa X^{5/6} + O(X^{1/2})$. Then, on average,

$$
f(X + N) - f(X) \approx \int_X^{X+N} \frac{5\kappa/6}{H^{1/6}} \, dH + O \left( \frac{N}{X^{1/2}} \right) \approx \frac{5\kappa}{6} \cdot \frac{N}{(X + N)^{1/6}} + O \left( \frac{N}{X^{1/2}} \right),
$$

in the sense that if $X_i = i \cdot N$ for $i = 0, \ldots, \lfloor X/N \rfloor$, then

$$
\frac{1}{\lfloor X/N \rfloor} \sum_{i=0}^{\lfloor X/N \rfloor-1} \left( f(X_{i+1}) - f(X_i) - \int_{X_i}^{X_{i+1}} \frac{5\kappa/6}{H^{1/6}} \, dH \right) = O \left( \frac{N}{X^{1/2}} \right)
$$

and

$$
\left| \frac{1}{\lfloor X/N \rfloor} \sum_{i=0}^{\lfloor X/N \rfloor-1} \left( f(X_{i+1}) - f(X_i) - \frac{5\kappa}{6} \cdot \frac{N}{X_{i+1}} \right) \right| = O \left( \frac{N}{X^{1/2}} \right).
$$

In particular, $f(X) - f(X - 1) \approx (5\kappa/6) X^{1/6} + O(X^{-1/2})$, on average.

**Proof.** Let $N \geq 1$ be fixed, let $X > N^2$ be fixed, and let us define $X_i = i \cdot N$ for $i = 0, \ldots, \lfloor X/N \rfloor$. Then

$$
\frac{1}{\lfloor X/N \rfloor} \sum_{i=0}^{\lfloor X/N \rfloor-1} \left( f(X_{i+1}) - f(X_i) - \int_{X_i}^{X_{i+1}} \frac{5\kappa/6}{H^{1/6}} \, dH \right)
$$

$$
= \frac{1}{\lfloor X/N \rfloor} \left( f([X/N]N) - f(0) - \int_0^{[X/N]N} \frac{5\kappa/6}{H^{1/6}} \, dH \right)
$$

$$
= \frac{1}{\lfloor X/N \rfloor} \left( f([X/N]N) - \kappa \cdot ([X/N]N)^{5/6} \right)
$$

$$
= \frac{1}{\lfloor X/N \rfloor} \cdot O(([X/N]N)^{1/2}) = O \left( \frac{X^{1/2}}{X/N} \right) = O \left( \frac{N}{X^{1/2}} \right),
$$

DISTRIBUTION OF RANKS OF ELLIPTIC CURVES 9
by Theorem 3.1. Now it follows that
\[
\sum_{i=0}^{\left\lfloor X/N \right\rfloor - 1} \left( f(X_{i+1}) - f(X_i) - \int_{X_i}^{X_{i+1}} \frac{5\kappa/6}{H^{1/6}} dH + \int_{X_i}^{X_{i+1}} \frac{5\kappa/6}{H^{1/6}} dH - \frac{5\kappa}{6} \cdot \frac{N}{X_{i+1}^{1/6}} \right) + \frac{1}{\left\lfloor X/N \right\rfloor} \cdot \sum_{i=0}^{\left\lfloor X/N \right\rfloor - 1} \left( \int_{X_i}^{X_{i+1}} \frac{5\kappa/6}{H^{1/6}} dH - \frac{5\kappa}{6} \cdot \frac{N}{X_{i+1}^{1/6}} \right) = O \left( \frac{N}{X^{1/2}} \right) + O \left( \frac{N}{X^{1/2}} \right) + O \left( \frac{N^2}{X} \right) = O \left( \frac{N}{X^{1/2}} \right),
\]
where we have used the fact that \( X > N^2 \), and therefore \( 1 \geq (N^2/X)^{1/2} > N^2/X > 0 \), and if \( h(x) \) is decreasing, then
\[
0 \leq \int_a^b h(x) dx - (b - a)h(b) \leq (b - a)(h(a) - h(b)),
\]
with \( b - a = X_{i+1} - X_i = N \) and \( h(x) = (5\kappa/6)/x^{1/6} \).
\( \square \)

**Remark 3.5.** For each \( n \geq r \geq 0 \) and \( X \geq N \geq 0 \), we would like to be able to give estimates of the quantities \( \pi_G((X, X + N)), \pi_{S_n}((X, X + N)) \) and \( \pi_{R_n}((X, X + N)) \). However, Corollary 3.4 says that this is not possible, and the best we hope for are results on average. From now on, if \( f = f(X, N) \) and \( g = g(X, N) \) are functions, then we say that \( f \approx g \) on average, for \( X \gg N \), if

\[
\frac{1}{\left\lfloor X/N \right\rfloor} \cdot \sum_{i=0}^{\left\lfloor X/N \right\rfloor - 1} f(X_i, N) - g(X_i, N)
\]
goes to zero as \( X \to \infty \), where \( X_i = i \cdot N \) for \( i = 0, \ldots, \left\lfloor X/N \right\rfloor \). The condition \( X \gg N \) is to be specified every time, but in general we will assume \( X > N^2 \) as in Corollary 3.4. If we want to be more specific about the error terms, we shall say \( f \approx g + O(h(X, N)) \) if the quantity in Eq. (1) is \( O(h(X, N)) \).

We finish this section by stating an immediate consequence of the definition of \( f \approx g \) on average.

**Lemma 3.6.** Let \( f = f(X, N) \) and \( g = g(X, N) \) be functions such that \( f \approx g \) on average for \( X \gg N \geq 0 \), and let \( a(X) \) be a function such that \( |a(X)| \leq 1 \) for all \( X \geq x_0 \), for some \( x_0 \in \mathbb{R} \). Then, \( a(X) \cdot f(X, N) \approx a(X) \cdot g(X, N) \) on average. More generally, suppose \( f \approx g + O(h) \) for some other functions \( h = h(X, N) \).
Tate-Shafarevich group of interest is the conditional probability \([27], [18]\). The value of the constant as the 2-Selmer rank (or, simply, Selmer rank) of \(E\) is also integrable with \(a = O(b(X))\). Then,
\[
a(X) \cdot f(X, N) \approx \int_X^{X+N} a(H) \cdot \hat{g}(H) \, dH + b(X) \cdot O(h(X)) \approx \int_X^{X+N} a(H) \cdot \hat{g}(H) \, dH + O(b(h))
\]
on average.

4. Selmer groups

Let \(\text{Sel}_2(E/Q)\) be the 2-Selmer group of \(E/Q\) and let \(\text{III}(E/Q)[2]\) be the 2-torsion subgroup of the Tate-Shafarevich group of \(E/Q\) (as defined in \([26]\), Chapter X) which fit in a short exact sequence
\[
0 \longrightarrow E(Q)/2E(Q) \overset{\delta_E}{\longrightarrow} \text{Sel}_2(E/Q) \longrightarrow \text{III}(E/Q)[2] \longrightarrow 0
\]
As in \([16]\), we shall refer to the quantity
\[
\text{rank}_{\mathbb{Z}/2\mathbb{Z}}(\text{Sel}_2(E/Q)/(E(Q)_{\text{tors}}/2E(Q)_{\text{tors}})) = \text{rank}_{\mathbb{Z}/2\mathbb{Z}}(\text{Sel}_2(E/Q)) - \text{rank}_{\mathbb{Z}/2\mathbb{Z}}(E(Q)[2])
\]
as the 2-Selmer rank (or, simply, Selmer rank) of \(E/Q\), and will denote it by \(\text{selrank}(E(Q))\). We note here that the exact sequence above implies that \(\text{rank}(E(Q)) \leq \text{selrank}(E(Q))\) for all elliptic curves. We define
\[
\mathcal{S}_n = \{E \in \mathcal{E} : \text{selrank}(E(Q)) = n\},
\]
and we will denote by \(\mathcal{S}_n(X)\) those curves in \(\mathcal{S}_n\) of height up to \(X\), and \(\pi_{\mathcal{S}_n}(X) = \#\mathcal{S}_n(X)\). Imitating the notation in the previous section, we shall also write \(\mathcal{S}_n([X_1, X_2])\) and \(\pi_{\mathcal{S}_n}([X_1, X_2])\) when referring to curves in \(\mathcal{S}_n\) in the height interval \([X_1, X_2]\), and will abbreviate \(\mathcal{S}_n^X = \pi_{\mathcal{S}_n}([X, X])\). Poonen and Rains ([24]) have conjectured a value for the limit \(s_n = \lim_{X \to \infty} \pi_{\mathcal{S}_n}(X)/\pi_\mathcal{E}(X)\), namely
\[
s_n = \text{Prob}(\text{selrank}(E(Q)) = n) = \left( \prod_{j \geq 0} \frac{1}{1 + 2^{-j}} \right) \cdot \left( \prod_{k=1}^{n} \frac{2}{2^k - 1} \right),
\]
and, in fact, they conjecture a similar distribution for \(p\)-Selmer groups of rank \(n\), and any prime \(p\). This probability has been shown to hold for quadratic twists of certain elliptic curves (see \([16], [17], [27]\), and \([18]\)). The value of the constant \(s_0 = \prod_{j \geq 0} (1 + 2^{-j})^{-1}\) is approximately 0.20971122, and we have included approximations of \(s_n\) for \(n = 1, \ldots, 6\) for future reference in Table 1.

| \(s_0\)   | \(s_1\)   | \(s_2\)   | \(s_3\)   | \(s_4\)   | \(s_5\)   | \(s_6\)   |
|---------|---------|---------|---------|---------|---------|---------|
| 0.20971122 | 0.41942244 | 0.27961496 | 0.07988998 | 0.01065199 | 0.00068722 | 0.00002181 |

Table 3. Values of \(s_n = \text{Prob}(\text{selrank}(E(Q)) = n)\)

For our purposes, we are interested in the behavior of the function \(\mathcal{S}_n(X)\), but we are even more interested in the conditional probability
\[
\text{pSel}_n(X) = \text{Prob}(\text{selrank}(E(Q)) = n \mid \text{ht}(E) = X) = \#\mathcal{S}_n^X / \#\mathcal{E}^X,
\]
when \( \#E^X \neq 0 \). In other words, we would like to know the probability that a curve \( E \) of height \( X \) has Selmer rank \( n \). We will model this probability with a random model, described in the following section.

5. Random model, following Cramér, part 1

In this section we define a space \( \tilde{E} \) of “test elliptic curves” and “test Selmer elements”, which will become a probability space when taking into account our probabilistic hypotheses \( H_A \) and \( H_B \) (and their refinement \( H_C \)).

**Definition 5.1.** A test elliptic curve is a triple \( E = (X, n, \text{Sel}_2) \) consisting of:

- a positive integer \( X \geq 1 \), the height of \( E \), also denoted \( X = \text{ht}(E) \),
- a non-negative integer \( n \), the Selmer rank of \( E \), also denoted \( n = \text{selrank}(E) \), and
- a vector \( \text{Sel}_2(E) = (s_{E,1}, s_{E,2}, \ldots, s_{E,\lceil \frac{n}{2} \rceil}) \) of \( \lceil \frac{n}{2} \rceil = (n - (n \mod 2))/2 \) test Selmer elements.

Each test Selmer element is either a MW element, or a III element.

The set of all test elliptic curves will be denoted by \( \tilde{E} \), and the subset of those test curves with height \( X \) will be denoted by \( \tilde{E}^X \).

It follows from the definition of the space \( \tilde{E} \) of test elliptic curves that \( \tilde{E} = \bigcup_{X \geq 1} \tilde{E}^X \).

**Remark 5.2.** If \( E/\mathbb{Q} \) is an elliptic curve, then we can associate a test elliptic curve \( (X, n, \text{Sel}_2) \) to \( E \) as follows. Clearly, \( X = \text{ht}(E) \) is the naive height of \( E \), and the non-negative integer \( n = \text{selrank}(E) \) is the 2-Selmer rank of \( E \) (defined as in Section 4). Let \( \text{Sel}_2(E/\mathbb{Q}) \) be the 2-Selmer group of \( E/\mathbb{Q} \). Then, \( \text{Sel}_2(E/\mathbb{Q}) \cong (\mathbb{Z}/2\mathbb{Z})^{n+1} \), where \( t = \text{rank}_{\mathbb{Z}/2\mathbb{Z}}(E(\mathbb{Q})[2]) \). Further, if we assume the finiteness of \( \text{III}(E/\mathbb{Q}) \), then \( \text{rank}_{\mathbb{Z}/2\mathbb{Z}}(\text{III}(E/\mathbb{Q})[2]) = 2s \) is even, and therefore \( n = R_{E/\mathbb{Q}} + 2s \), where \( R_{E/\mathbb{Q}} \) is the \( \mathbb{Z} \)-rank of \( E(\mathbb{Q}) \). In particular, \( n \equiv R_{E/\mathbb{Q}} \mod 2 \). It follows that if \( n \) is odd, then \( 1 \leq R_{E/\mathbb{Q}} \leq n \) and \( R_{E/\mathbb{Q}} \) is odd, so there is always an element of \( \text{Sel}_2(E/\mathbb{Q})/(E(\mathbb{Q})_{\text{tors}}/2E(\mathbb{Q})_{\text{tors}}) \) that comes from a point of infinite order from the Mordell–Weil group. Hence, when \( n \) is odd, we are interested in the other \( n - 1 \) generators of the Selmer group, to see if they come from the Mordell–Weil group, or generate non-trivial Sha elements. Moreover, there are \( 2m = n - 1 - (2s) \) generators of the Selmer group that come from the Mordell–Weil group. Thus, we define the set of symbols \( \text{Sel}_2 \) by

\[
\begin{cases} 
\text{s elements of the form III and } m = (n - 2s)/2 \text{ elements of the form MW} & \text{if } n \text{ is even,} \\
\text{s elements of the form III and } m = (n - 1 - 2s)/2 \text{ elements of the form MW} & \text{if } n \text{ is odd.}
\end{cases}
\]

so that \( n - 1 = 2(m + s) \). We note here that \( m = R_{E/\mathbb{Q}}/2 \) when \( R_{E/\mathbb{Q}} \) is even, and \( m = (R_{E/\mathbb{Q}} - 1)/2 \) if the rank is odd. We will come back and explain in more detail why \( \text{Sel}_2(E) \) should have \( \lfloor \frac{n}{2} \rfloor \) elements in Section 7.

**Example 5.3.** Let \( E/\mathbb{Q} \) be the elliptic curve \( y^2 = x^3 + 2993x \), with height \( 4 \cdot 2993^3 = 107245762628 \). A 2-descent shows that the Selmer group \( \text{Sel}_2(E/\mathbb{Q}) \cong (\mathbb{Z}/2\mathbb{Z})^{5} \). Since \( E(\mathbb{Q})_{\text{tors}} \cong \mathbb{Z}/2\mathbb{Z} \), it follows that \( \text{selrank}(E) = 4 \). Further, a 4-descent (using Magma) shows that \( E(\mathbb{Q}) \cong \mathbb{Z}/2\mathbb{Z} \oplus (\mathbb{Z}/2\mathbb{Z})^{2} \), and \( \text{III}(E(\mathbb{Q})[2]) \cong (\mathbb{Z}/2\mathbb{Z})^{2} \). Hence, this elliptic curve could be represented as a test elliptic curve by the triple \( (107245762628, 4, (\text{MW, III})) \).

Similarly, the curve \( E' : y^2 = x^3 - 1679x \) has Selmer rank 3, Mordell–Weil rank 1, and so it would correspond to the triple \( (18932679356, 3, (\text{III})) \).
Lemma 5.7. Let $E$ be a probability measure over $\bigcup_{X \geq 1} \mathcal{E}_X$, defining a probability measure $P$.

Definition 5.4. We let $T$ be a space of sequences of subsets of $\mathcal{E}_X$, defined as follows:

$$T = \left\{ (\mathcal{T}_X)_{X \geq 1} : \mathcal{T}_X \subseteq \mathcal{E}_X, \sum_{N=1}^{N} \#T_N = \kappa X^{5/6} + O(X^{1/2}) \right\}.$$ 

If $I$ is a finite interval in $[1, \infty)$, and $\mathcal{T} \in T$, we will write $\mathcal{T}(I) = \bigcup_{X \in I} \mathcal{T}_X$, and $\mathcal{T}S_n(I) = \bigcup_{X \in I} \mathcal{T}_X \cap S_n^X$. Finally, we define

$$\pi_T(I) = \#T(I) = \sum_{X \in I} \#T_X \quad \text{and} \quad \pi_{TS_n}(I) = \#T\mathcal{S}_n(I) = \sum_{X \in I} \#(\mathcal{T}_X \cap S_n^X).$$

Remark 5.5. The sequence of test elliptic curves $(\mathcal{E}_X)_{X \geq 1}$ associated to ordinary elliptic curves over $\mathbb{Q}$ (as in Remark 5.2), belongs to $\mathcal{T}$, by Theorem 3.1. Thus, the goal is to predict the behaviour of $(\mathcal{E}_X)_{X \geq 1}$ from the average asymptotic behaviour of sequences in $\mathcal{T}$.

In the next definition, we fix a Selmer rank $n$, and we make $\mathcal{E}_X$ into a probability space by defining a probability measure $P_n^X$ for each $X \geq 1$.

Hypothesis 5.6 (Hypothesis A, or $H_A$). Let $n \geq 0$, and $X \geq 1$ be fixed. Let $\theta_n(X)$ be a function $[1, \infty) \to (0, 1)$ such that $\lim_{X \to \infty} \theta_n(X) = s_n$. We define a probability space $(\mathcal{E}_X, F_n^X, P_n^X)$ by defining a probability measure $P_n^X$ as follows:

- $\mathcal{E}_X$ is an infinite discrete space of test elliptic curves of height $X$, and $S_n^X$ is the (infinite) subset of test elliptic curves $E$ of height $X$ with $\text{selrank}(E) = n$.
- $F_n^X = \{ \emptyset, S_n^X, \mathcal{E}_X \setminus S_n^X, \mathcal{E}_X \}$.
- $P_n^X \left( S_n^X \right) = \theta_n(X)$, and $P_n^X \left( \mathcal{E}_X \setminus S_n^X \right) = 1 - \theta_n(X)$.

If $X_1, \ldots, X_m$ are natural numbers, then we endow $\prod_{i=1}^{m} \mathcal{E}_{X_i}$ with the product measure

$$\prod_{i=1}^{m} \left( \mathcal{E}_{X_i}, F_{n_i}^{X_i}, P_{n_i}^{X_i} \right).$$

Lemma 5.7. Let $(\mathcal{E}_X, F_n^X, P_n^X)$ be the probability space defined by Hypothesis A, and let $Y_{\text{Sel}, n, X} : \mathcal{E}_X \to \{0, 1\}$ be the function that takes values

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

Then

1. $Y_{\text{Sel}, n, X}$ is a random variable that follows a Bernoulli distribution with probability $\theta_n(X)$, such that $\lim_{X \to \infty} \theta_n(X) = s_n$.
2. For each $i = 1, \ldots, m$, let $E_i$ be a test elliptic curves of height $X_i \geq 1$, chosen at random. Then, the events

$$Y_{\text{Sel}, n, X_1}(E_1) = 1, \ldots, Y_{\text{Sel}, n, X_m}(E_m) = 1$$

are mutually independent.
Proof. For (1), from the definitions, if \( E \in \tilde{E}^X \), then the probability that \( Y_{\text{Sel},n,X}(E) = 1 \) is given by
\[
\Pr(Y_{\text{Sel},n,X}(E) = 1) = P_n^X(\{ E \in \tilde{E}^X : Y_{\text{Sel},n,X}(E) = 1 \}) = P_n^X(\tilde{S}_n^X) = \theta_n(X),
\]
and similarly,
\[
\Pr(Y_{\text{Sel},n,X}(E) = 0) = P_n^X(\{ E \in \tilde{E}^X : Y_{\text{Sel},n,X}(E) = 0 \}) = P_n^X(\tilde{E}^X \setminus \tilde{S}_n^X) = 1 - \theta_n(X).
\]
Thus, \( Y_{\text{Sel},n,X}(E) \) follows a Bernoulli distribution \( B(1, \theta_n(X)) \).

For (2), consider \( (E_1, \ldots, E_m) \in \prod_{i=1}^m \tilde{E}^{X_i} \). By Hypothesis A, the probability measure on the product space is the product measure \( \prod_{i=1}^m (\tilde{E}^{X_i}, \mathcal{F}^{X_i}, P_n^{X_i}) \). Consider the events
\[
A_i = \{(E_1, \ldots, E_m) : E_i \in \tilde{S}_n^{X_i}\}
\]
for \( i = 1, \ldots, m \). Then,
\[
\Pr(A_1) \cdots \Pr(A_m) = P(A_1) \cdots P(A_m) = \prod_{i=1}^m \left( P_n^{X_i}(\tilde{S}_n^{X_i}) \cdot \prod_{j \neq i} P_n^{X_j}(\tilde{E}^{X_j}) \right) = \prod_{i=1}^m \theta_n(X_i).
\]
On the other hand \( A_1 \cap \cdots \cap A_m = \{(E_1, \ldots, E_m) : E_i \in \tilde{S}_n^X \text{ for all } i = 1, \ldots, m\} \) so
\[
\Pr(A_1 \cap \cdots \cap A_m) = P(A_1 \cap \cdots \cap A_m) = P_n^{X_1}(\tilde{S}_n^{X_1}) \cdots P_n^{X_m}(\tilde{S}_n^{X_m}) = \prod_{i=1}^m \theta_n(X_i).
\]
Thus, \( \Pr(A_1) \cdots \Pr(A_m) = \Pr(A_1 \cap \cdots \cap A_m) \) as desired. \( \square \)

Remark 5.8. Suppose \( E/\mathbb{Q} \) and \( E'/\mathbb{Q} \) are two non-isomorphic elliptic curves. If \( E \) and \( E' \) happen to be in the same isogeny class, then their Selmer ranks will not be independent events. However, by a theorem of Kenku ([19]), an elliptic curve \( E/\mathbb{Q} \) is isogenous to at most 8 non-isomorphic elliptic curves over \( \mathbb{Q} \). Thus, we may disregard the possibility of isogenous curves, as it would only contribute a negligible error that would vanish as we increase the sample size.

6. The number of curves with Selmer rank \( n \) up to height \( X \)

In this section we prove several consequences of the probabilistic model defined in Section 5, and we investigate the properties of \( \theta_n(X) \).

Corollary 6.1. Assume \( H_A \), and let \( \mathcal{E} = \{E_1, \ldots, E_m\} \) be a set of test elliptic curves in \( \tilde{E}^X \) chosen at random. Then, the number of curves in \( \mathcal{E} \) of Selmer rank \( n \) follows a binomial distribution \( B(m, \theta_n(X)) \). In particular the expected value of \( \#(\mathcal{E} \cap \tilde{S}_n)/\#\mathcal{E} \) is \( \theta_n(X) \) with standard error \( \sqrt{\frac{1}{m} \theta_n(X)(1 - \theta_n(X))} \). More generally, if \( \{E_1, \ldots, E_m\} \) are test elliptic curves in \( \tilde{E} \), with \( E_i \) of height \( X_i \) for \( i = 1, \ldots, m \), and chosen at random, then
\[
\mathbb{E}(\#(\mathcal{E} \cap \tilde{S}_n)/\#\mathcal{E}) = \frac{1}{m} \sum_{i=1}^m \theta_n(X_i)
\]
with standard error \( \sqrt{\frac{1}{m^2} \sum \theta_n(X_i)(1 - \theta_n(X_i))} \).
Proof. Let us assume $H_A$ and let us first show the most general case. Let $E_1, \ldots, E_m$ be test elliptic curves in $\mathcal{E}$ of height $X_1, \ldots, X_n$, respectively, chosen at random. In particular, by $H_A$, each random variable $Y_{\text{Sel},n,X_i}(E_i) \sim B(1, \theta_i(X_i))$, and since the curves are chosen at random, $H_A$ says that the events $Y_{\text{Sel},n,X_i}(E_i) = 1$ are mutually independent. Then, the number of elements in $\# \mathcal{E} \cap \mathcal{S}_n$ can be expressed as

$$t = \# \mathcal{E} \cap \mathcal{S}_n = \sum_{i=1}^m Y_{\text{Sel},n,X_i}(E_i).$$

It follows that the expected value of $t$ is

$$E(t) = E\left(\sum_{i=1}^m Y_{\text{Sel},n,X_i}(E_i)\right) = \sum_{i=1}^m E(Y_{\text{Sel},n,X_i}(E_i)) = \sum_{i=1}^m \theta_i(X_i),$$

and so the expected value of $\# \mathcal{E} \cap \mathcal{S}_n/\# \mathcal{E}$ is $1/m \sum \theta_i(X_i)$. The standard error of the approximation of $t/m$ by $1/m \sum \theta_i(X_i)$ is given by the square root of the variance of $t/m$. We compute

$$\text{Var}\left(\frac{1}{m} \sum_{i=1}^m Y_{\text{Sel},n,X_i}(E_i)\right) = \frac{1}{m^2} \sum_{i=1}^m \text{Var}(Y_{\text{Sel},n,X_i}(E_i)) = \frac{1}{m^2} \sum_{i=1}^m \theta_i(X_i)(1 - \theta_i(X_i)),$$

were we have used the fact that $Y_{\text{Sel},n,X_i}(E_i)$ are independent, which implies they are uncorrelated, and therefore the covariance terms vanish. Thus, the standard error is $\sqrt{\frac{1}{m^2} \sum \theta_i(X_i)(1 - \theta_i(X_i))}$ as claimed.

Now, if $X = X_1 = \ldots = X_m$, then $t = \# \mathcal{E} \cap \mathcal{S}_n = \sum_{i=1}^m Y_{\text{Sel},n,X}(E_i)$ follows a binomial $B(m, \theta(X))$, with mean $m \cdot \theta(X)$ and variance $\frac{1}{m} \theta(X)(1 - \theta(X))$, so the expected value of $t/m$ is $\theta(X)$ with standard error $\sqrt{\frac{1}{m^2} \theta(X)(1 - \theta(X))}$, as desired. \hfill \square

**Corollary 6.2.** If we assume $H_A$, then $\sum_{n=0}^\infty \theta(X) = 1$. Moreover, if $T \in \mathcal{T}$ and we define

$$\theta_n(X, N) = \frac{\pi_T \mathcal{S}_n((X, X + N])}{\pi_T((X, X + N])},$$

for each $X \geq 1$ and $N \geq 1$, then

1. The expected value of $\pi_T \mathcal{S}_n((X, X + N])$ is given by the formula

$$E(\pi_T \mathcal{S}_n((X, X + N])) \approx \frac{5N}{6} \int_X^{X+N} \frac{\theta_n(H)}{H^{1/6}} dH + O\left(\frac{N}{X^{1/2}}\right),$$

on average (see Remark 3.5).

2. For $X > N^2 \geq 0$, we have that the expected value of $\theta_n(X, N)$ is, on average, $\theta_n(X) + O(X^{-1/3})$, with a standard error given on average by

$$\sqrt{\frac{6 \cdot (X + N)^{1/6} \cdot \theta_n(X)(1 - \theta_n(X))}{5N^2N}} + O\left(\frac{1}{NX^{1/6}}\right).$$

3. Let $N = N(X)$ be a function of $X$. Then, $\lim_{X \to \infty} \theta_n(X, N(X)) = s_n$ as long as the growth condition $\lim_{X \to \infty} X^{1/6}/N(X) = 0$ is satisfied.
Proof. Let \( E = (X, n, \text{Sel}_2) \) be a fixed test elliptic curve of height \( X \). Since \( \text{selrank}(E) \) takes precisely one value (a non-negative number \( n \geq 0 \)), it follows that
\[
\sum_{n=0}^{\infty} \theta_n(X) = \sum_{n \geq 0} \Pr(Y_{\text{Sel},n,X}(E) = 1) = 1,
\]
by the laws of probability. For part (1) of the statement, let \( \tilde{T} \in T \) be arbitrary (as in Definition 5.4). We note that
\[
\pi_{\tilde{T}S_n}((X, X + N]) = \sum_{E \in \tilde{T}((X,X+N])} Y_{\text{Sel},n,X}(E).
\]
and therefore we may use Corollary 6.1 to obtain the expected value.
\[
\mathbb{E}(\pi_{\tilde{T}S_n}((X, X + N])) = \sum_{E \in \tilde{T}((X,X+N])} \theta_n(\text{ht}(E)) = \frac{X+N}{H=X+1} \sum_{E \in \tilde{T}([H,H])} \theta_n(H) = \sum_{H=X+1}^{X+N} \pi_{\tilde{T}}([H,H]) \cdot \theta_n(H)
\]
Since \( \tilde{T} \in T \), by definition, we have \( \pi_{\tilde{T}}([H,H]) \approx \frac{5\kappa}{6} \cdot \int_{H-1}^{H} \frac{1}{T^{1/6}} \,dT + O(H^{-1/2}) \) on average (see Remark 3.5 for the definition of the term “on average” in this context). Since \( |\theta_n(X)| \leq 1 \) for all \( X \), Lemma 3.6 implies that \( \pi_{\tilde{T}}([H,H]) \theta_n(H) \approx \frac{5\kappa}{6} \cdot \int_{H-1}^{H} \frac{\theta_n(T)}{T^{1/6}} \,dT + O(H^{-1/2}) \) on average. Thus,
\[
\mathbb{E}(\pi_{\tilde{T}S_n}((X, X + N])) \approx \frac{5\kappa}{6} \cdot \int_{X}^{X+N} \theta_n(H) \,dH + \theta_n(X) \cdot O\left(\frac{N}{X^{1/2}}\right),
\]
on average. For part (2), we use instead the second part of Corollary 3.4 which says \( \pi_{\tilde{T}}([H,H]) \approx \frac{(5\kappa/6)}{H^{1/6}} + O(H^{-1/2}) \), on average. Thus, \( \mathbb{E}(\pi_{\tilde{T}S_n}((X, X + N])) \) is, on average, given by
\[
\approx \frac{5\kappa}{6} \cdot \frac{N}{(X + N)^{1/6}} \cdot \theta_n(X) + O\left(\frac{N}{X^{1/2}}\right) \approx \pi_{\tilde{T}}((X, X + N]) \cdot \theta_n(X) + O\left(\frac{N}{X^{1/2}}\right).
\]
Thus, \( \mathbb{E}(\theta_n(X, N)) \approx \theta_n(X) + O\left(\frac{X + N)^{1/6}}{X^{1/2}}\right) = \theta_n(X) + O\left(X^{-1/3}\right) \) as claimed, since we are assuming that \( X > N^2 \). Similarly, the variance is given by
\[
\text{Var}(\theta_n((X, X + N])) = \frac{1}{(\pi_{\tilde{T}}((X, X + N]))^2} \sum_{E \in \tilde{T}((X,X+N])} \theta_n(\text{ht}(E))(1 - \theta_n(\text{ht}(E)))
\]
\[
= \frac{1}{(\pi_{\tilde{T}}((X, X + N]))^2} \sum_{H=X+1}^{X+N} \sum_{E \in \tilde{T}([H,H])} \theta_n(H)(1 - \theta_n(H))
\]
\[
\approx \frac{5\kappa/6}{(\pi_{\tilde{T}}((X, X + N]))^2} \left( \int_{X}^{X+N} \theta_n(H)(1 - \theta_n(H)) \,dH + O\left(\frac{N}{X^{1/2}}\right) \right).
\]
We provide the best-fit values of $C_n$ for the data of \( \pi_{SN}(X) \) in the interval \( [2.6975 \cdot 10^{10}, 2.7 \cdot 10^{10}] \), together with the values of \( s_n \).

Finally, we have tried to model the graphs of \( \theta_n(X,N) \) using simple rational functions (and assuming the conjectural limit values given by Poonen–Rains [24]), and we have found using SageMath best-fit models for the data of \( \theta_n(X,N) \) of the form

\[
\theta_n(X,N) \approx \frac{s_n}{1 + C_n X^{-e_n}}.
\]

We provide the best-fit values of \( C_n \) and \( e_n \) in Table 5.

6.1. Testing Hypothesis A. In order to test Hypothesis A, we shall use the sequence \( (E^X)_{X \geq 1} \) as a representative of \( T \) (see Remarks 5.2 and 5.5). The BHKSSW data (Section 3, [1]) is thus used to estimate the moving ratios \( \theta_n(X,N) \) of Corollary 6.2. We have plotted approximate values of \( \theta_n(X,0.025 \cdot 10^9) \) for \( n = 1, \ldots, 5 \) using the BHKSSW database, and the graphs can be found in Figure 2.

In Table 4 we record the last values of \( \theta_n(X,0.025 \cdot 10^9) \) that appear in the graphs (which correspond to \( X \approx 2.6975 \cdot 10^{10} \)). We also record the values of \( \pi_{SN} \) in \( [2.6975 \cdot 10^{10}, 2.7 \cdot 10^{10}] \). The total number of elliptic curves in the same interval is 182,823.

| \( n \) | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| \( \pi_{SN}(2.6975 \cdot 10^{10}, 2.7 \cdot 10^{10}) \) | 80,996 | 47,427 | 10,556 | 821 | 29 |
| \( \theta_n(2.6975 \cdot 10^{10}, 0.025 \cdot 10^9) \) | 0.44083621 | 0.26278969 | 0.05835152 | 0.00463836 | 0.00008751 |
| \( s_n \) | 0.41942244 | 0.27961496 | 0.07988998 | 0.01065199 | 0.00068722 |

Table 4. The number of curves of Selmer rank \( 1 \leq n \leq 5 \), and the values of \( \theta_n(X,N) \) in the interval \( [2.6975 \cdot 10^{10}, 2.7 \cdot 10^{10}] \), together with the values of \( s_n \).
Figure 2. Graphs of the moving ratios $\theta_n(X, 0.025\cdot10^9)$ for $n = 1$ (blue), 2 (green), 3 (red), 4 (gray), 5 (purple).

| $n$ | $C_n$  | $e_n$      |
|-----|-------|------------|
| 1   | -0.40116957 | 0.08540201 |
| 2   | 1.41108621  | 0.12348659 |
| 3   | 11.18222736 | 0.14061542 |
| 4   | 179.71749981| 0.20339670 |
| 5   | 95474.85098037| 0.39937065 |

Table 5. The parameters of the best-fit models $\theta_n(X, N) \approx s_n/(1 + C_nX^{-e_n})$.

The models constructed above are remarkably good approximations of the values of $\theta_n(X, 0.025\cdot10^9)$, at least up to height $2.7\cdot10^{10}$. See Figures 3 and 4. Thus, we refine our Cramér-like model of Section 5 by specifying $\theta_n(X)$ up to two constants $C_n$ and $e_n$.

**Hypothesis 6.3** (Hypothesis $H'_A$). Hypothesis $H_A$ holds and, for each $n \geq 1$, there are constants $C_n$ and $e_n$ such that $\theta_n(X) = s_n/(1 + C_nX^{-e_n})$. Moreover, if $n = 1, \ldots, 5$, then the approximate values of $C_n$ and $e_n$ are as in Table 5.

**Remark 6.4.** Let us assume Hypothesis $H'_A$, and let us use Corollary 6.2 to estimate the error in the approximation $\theta_n(X) \approx \theta_n(X, N)$. The error should be given by the expression

$$
\text{err}_n(X, N) = \sqrt{\frac{6 \cdot X^{1/6} \cdot \theta_n(X)(1 - \theta_n(X))}{5N\kappa}} = \sqrt{\frac{6 \cdot X^{1/6}}{5N\kappa}} \cdot \frac{s_n}{1 + C_nX^{-e_n}} \left(1 - \frac{s_n}{1 + C_nX^{-e_n}}\right).
$$
In Table 6 we include the values of $\theta_n(X,N)$ based on the BHKSSW data, our model of $\theta_n(X)$ with the constants from Table 5, the error $|\theta_n(X,N) - \theta_n(X)|$, and the predicted standard error $\text{err}_n(X,N)$, for $X = 2.6975 \cdot 10^{10}$ and $N = 0.025 \cdot 10^9$.

| $n$   | 1               | 2               | 3               | 4               | 5               |
|-------|-----------------|-----------------|-----------------|-----------------|-----------------|
| $\theta_n(2.6975 \cdot 10^{10}, 0.025 \cdot 10^9)$ | 0.44083621      | 0.26278969      | 0.05835152      | 0.00463836      | 0.00008751      |
| $\theta_n(2.6975 \cdot 10^{10})$ | 0.44223400      | 0.26066727      | 0.05781814      | 0.00451697      | 0.00009141      |
| $|\text{Error}|$ | 0.00139779      | 0.00212241      | 0.00053337      | 0.00012138      | 0.00000390      |
| $\text{err}_n(2.6975 \cdot 10^{10}, 0.025 \cdot 10^9)$ | 0.000115688     | 0.000102258     | 0.000054367     | 0.000015619     | 0.00002227      |

Table 6. Values of: $\theta_n(X,N)$, our model of $\theta_n(X)$, the error $|\theta_n(X,N) - \theta_n(X)|$, and the predicted standard error $\text{err}_n(X,N)$, for $X = 2.675 \cdot 10^{10}$ and $N = 0.025 \cdot 10^9$.

**Remark 6.5.** The BHKSSW database ([1]) also includes small databases of random samples of elliptic curves at larger heights. In particular, for each $k \in [11, 16]$, they calculated the Selmer rank and rank of a set $L_k$ consisting of about 100,000 curves from a uniform distribution of all curves in the height range $[10^k, 2 \cdot 10^k]$. We have tested $H'_A$ on these sets $\mathcal{E}_k$ of curves of large height. In Table 7, we include the value of the moving ratio for $\mathcal{E}_{16} \cap \mathcal{S}_n$, the value of $\theta_n(10^{16})$, the error, and the predicted error $\text{err}_n$. The predicted error for $n = 5$ is too large (similarly for $n = 4$ to a lesser degree), so the sample is just too small to provide significant evidence. Otherwise, the data for $n = 1, 2, 3$ shows that Hypothesis $H'_A$ seems to be a very good match for the data, even for large heights.

| $n$   | 1               | 2               | 3               | 4               | 5               |
|-------|-----------------|-----------------|-----------------|-----------------|-----------------|
| $\#\mathcal{E}_{16} \cap \mathcal{S}_n$ | 42,631          | 27,543          | 7327            | 836             | 38              |
| $\#\mathcal{E}_{16} \cap \mathcal{S}_n / \#\mathcal{E}_{16}$ | 0.42636116      | 0.27546305      | 0.07327879      | 0.00836100      | 0.00038004      |
| $\theta_n(10^{16})$ | 0.42678631      | 0.27550444      | 0.07516196      | 0.00968314      | 0.00066148      |
| $|\text{Error}|$ | 0.00042515      | 0.00004138      | 0.00188317      | 0.00132213      | 0.00028144      |
| $\text{err}_n(\mathcal{E}_{16} \cap \mathcal{S}_n)$ | 0.00239552      | 0.00269200      | 0.00308012      | 0.00338681      | 0.00275683      |

Table 7. Values of: the moving ratio $\theta_n$, our model of $\theta_n(X)$, the error, and the predicted standard error $\text{err}_n(X,N)$, for the database $\mathcal{E}_{16}$ of height $X \approx 10^{16}$.

**Example 6.6.** By Corollary 6.1, Hypothesis $A$ implies that if $E_1, \ldots, E_m$ are test elliptic curves with height $\approx X$, then the number of test curves $E_i$ of Selmer rank $n$ would follow a binomial distribution $B(m, \theta_n(X))$. We have tested this against the BHKSSW database and the data and
Figure 3. Graphs of the moving ratios $\theta_n(X, 0.025 \cdot 10^9)$ for $n = 1$ (blue), 2 (green), 3 (red), and the corresponding models of the form $s_n/(1 + C_n X^{-\epsilon_n})$ (in red).

Figure 4. Graphs of the moving ratios $\theta_n(X, 0.025 \cdot 10^9)$ for $n = 4$ (gray), 5 (purple), and the corresponding models of the form $s_n/(1 + C_n X^{-\epsilon_n})$ (in red, and blue).
have always found the result to be in nice agreement with the predictions. For instance, let $T$ be the first 100,000 elliptic curves with height $\geq 9 \cdot 10^9$. We let $m = 100$, and pick 100 curves at random in $T$, and repeat this process 10000 times. For a fixed $n$ and for each of the 10000 trials, the distribution of the number $0 \leq t \leq 100$ of curves with Selmer rank $n$ would follow a binomial $B(100, \theta_n(X))$, where $X$ is in the interval $[9000573228, 9012972924]$. We use our models $\theta_n(X) = s_n/(1 + C_n X^{-\epsilon_n})$ in order to approximate values, for $n = 1, \ldots, 4$. We obtain:

$$\theta_1(X) \approx 0.444608, \ \theta_2(X) \approx 0.258128, \ \theta_3(X) \approx 0.055273, \ \theta_4(X) \approx 0.003948$$

for any $X$ in the given interval. If our event of picking 100 curves follows a binomial $B(100, \theta_n(X))$, then it must be approximately a normal $N(100 \cdot \theta_n(X), 100 \cdot \theta_n(X) \cdot (1 - \theta_n(X)))$, where $100 \cdot \theta_n(X)$ and $100 \cdot \theta_n(X) \cdot (1 - \theta_n(X))$ are, respectively, the mean and the variance of the binomial distribution. We have plotted the result of the 10000 experiments in Figure 5, together with the normal distributions predicted by $H'_A$.

Now, we can put our results together to estimate the number of curves of Selmer rank $1, \ldots, 5$ up to height $X$.

**Proposition 6.7.** If we assume $H_A$, and $\tilde{T} \in \mathbf{T}$ is arbitrary, then:

1. The expected value of $\pi_{\tilde{T}S_n}(X)$ is given by

$$\mathbb{E}(\pi_{\tilde{T}S_n}(X)) = \frac{5\kappa}{6} \int_0^X \frac{\theta_n(H)}{H^{1/6}} \, dH + \theta_n(X) \cdot O(X^{1/2}),$$

where $\kappa = 2^{4/3} \cdot (\zeta(10) \cdot 3^{3/2})^{-1}$. If in addition we assume Hypothesis 6.3 ($H'_A$), then

$$\pi_{\tilde{T}S_n}(X) \approx \frac{5\kappa s_n}{6} \int_0^X \frac{1}{H^{1/6}(1 + C_n H^{-\epsilon_n})} \, dH + O(X^{1/2}),$$

where the constants $s_n$, $C_n$, and $\epsilon_n$ are given in Tables 3 and 5 for $1 \leq n \leq 5$.

2. If $X > N^2 \geq 0$, and we assume Hypothesis 6.3 ($H'_A$), then the expected value, on average, is

$$\mathbb{E}(\pi_{\tilde{T}S_n}([X, X + N])) \approx \frac{5\kappa N \theta_n(X)}{6X^{1/6}} + O\left(\frac{N}{X^{1/2}}\right) \approx \frac{5\kappa s_n N}{6X^{1/6}(1 + C_n X^{-\epsilon_n})} + O\left(\frac{N}{X^{1/2}}\right).$$

**Proof.** If we assume $H_A$, and $\tilde{T} \in \mathbf{T}$ is arbitrary, then the expected value of

$$\pi_{\tilde{T}S_n}(X) = \sum_{H=1}^X \pi_{\tilde{T}S_n}([H, H])$$

is given by $\sum_{H=1}^X \pi_{\tilde{T}}([H, H]) \cdot \theta_n(H)$. Thus, Corollary 3.4 and Lemma 3.6 imply that

$$\mathbb{E}(\pi_{\tilde{T}S_n}(X)) = \sum_{H=1}^X \pi_{\tilde{T}}([H, H]) \cdot \theta_n(H) = \frac{5\kappa}{6} \int_0^X \frac{\theta_n(H)}{H^{1/6}} \, dH + \theta_n(X) \cdot O(X^{1/2})$$

$$\approx \frac{5\kappa s_n}{6} \int_0^X \frac{1}{H^{1/6}(1 + C_n H^{-\epsilon_n})} \, dH + \theta_n(X) \cdot O(X^{1/2}),$$
where the last approximation assumes Hypothesis 6.3. For part (2), if $X > N^2 \geq 0$, then by Corollary 3.4, the expected value on average is given by

$$E(\pi_{TS_n}((X, X + N)]) = \sum_{H=X+1}^{X+N} \pi_p([H, H]) \cdot \theta_n(H) \approx \frac{5\kappa}{6} \sum_{H=X+1}^{X+N} \left( \frac{\theta_n(H)}{H^{1/6}} + O\left(\frac{1}{X^{1/2}}\right) \right)$$

$$\approx \frac{5\kappa N \theta_n(X)}{6X^{1/6}} + O\left(\frac{N}{X^{1/2}}\right)$$

$$\approx \frac{5\kappa s_n N}{6X^{1/6}(1 + C_nX^{-e_n})} + O\left(\frac{N}{X^{1/2}}\right),$$

as claimed. \qed
We have used SageMath to do numerical integration and approximation of the expected values of \( \tilde{\pi}_{\mathcal{T}_{S_n}}(X) \) using the formula of Proposition 6.7, part (1), and we have plotted the graphs against actual data from the BHKSSW database in Figures 6 and 7.

Finally, Proposition 6.7 will allow us to write formulas for the average 2-Selmer rank of a test elliptic curve up to height \( X \). We plot our conjectural formula in Figure 8.

**Proposition 6.8.** Let \( \tilde{T} \in \mathbf{T} \) be arbitrary, and let \( \text{AvgSelRank}_{\tilde{T}}(X) \) be defined by

\[
\text{AvgSelRank}_{\tilde{T}}(X) = \frac{1}{\tilde{\pi}_{\tilde{T}}(X)} \sum_{E \in \tilde{T}(X)} \text{selrank}(E).
\]

If we assume \( H_A \) and we assume that \( 0 \leq \theta_n(X) \leq s_n \) for all \( n \geq 2 \) and all \( X > 0 \), then the expected value of the average Selmer rank is given by

\[
\mathbb{E}(\text{AvgSelRank}_{\tilde{T}}(X)) = \frac{5/6}{X^{5/6}} \int_{0}^{X} \frac{\sum_{n \geq 1} n \cdot \theta_n(H)}{H^{1/6}} dH + O \left( X^{-1/3} \right).
\]

Moreover, \( \lim_{X \to \infty} \mathbb{E}(\text{AvgSelRank}_{\tilde{T}}(X)) = \sum_{n \geq 1} ns_n = 1.26449978 \ldots \)

**Proof.** In order to compute the average Selmer rank, we note that

\[
\text{AvgSelRank}_{\tilde{T}}(X) = \frac{1}{\tilde{\pi}_{\tilde{T}}(X)} \sum_{E \in \tilde{T}(X)} \text{selrank}(E) = \frac{1}{\tilde{\pi}_{\tilde{T}}(X)} \sum_{n \geq 1} \sum_{E \in \mathcal{T}_{S_n}(X)} n = \frac{1}{\tilde{\pi}_{\tilde{T}}(X)} \sum_{n \geq 1} n \cdot \pi_{\mathcal{T}_{S_n}}(X).
\]
Thus, by Prop. 6.7 we have that the expected value of $\text{AvgSelRank}_{\tilde{T}}(X)$ is given by

$$
\mathbb{E}(\text{AvgSelRank}_{\tilde{T}}(X)) = \frac{1}{\pi_{\tilde{T}}(X)} \left( \frac{5\kappa}{6} \int_0^X \sum_{n \geq 1} n \cdot \theta_n(H) \frac{dH}{H^{1/6}} + \left( \sum_{n \geq 1} n \cdot \theta_n(X) \right) \cdot O \left( X^{1/2} \right) \right)
$$

Let us define $t_1 = s_1$ and $t_n = t_1/(2^{n(n-1)/2} - 1)$ for $n \geq 2$. Then, the definition of $s_n$ (in Section 4) implies that $s_n \leq t_n$, and so,

$$
\sum_{n=2}^N n s_n \leq \sum_{n=2}^N nt_n \leq \sum_{n=2}^N \frac{nt_1}{2^{n(n-1)/2} - 1}
$$

for any $N \geq 2$. In particular, $\sum_{n \geq 1} n s_n$ converges. Since we are assuming $0 \leq \theta_n(X) \leq s_n$ for $n \geq 2$, it follows that $\sum_{n \geq 1} n \theta_n(X)$ converges for any $X$, and $\lim_{X \to \infty} \sum_{n \geq 1} n \theta_n(X) = \sum_{n \geq 1} n s_n = 1.26449978\ldots$. Thus, if we use this, and $\pi_{\tilde{T}}(X) = \kappa X^{5/6} + O(X^{1/2})$ from Definition 5.4, we obtain

$$
\mathbb{E}(\text{AvgSelRank}_{\tilde{T}}(X)) = \frac{5/6}{X^{5/6}} \int_0^X \sum_{n \geq 1} n \cdot \theta_n(H) \frac{dH}{H^{1/6}} + O \left( X^{1/3} \right).
$$
Next, we calculate the limit of $E(\text{AvgSelRank}_T(X))$ as $X \to \infty$. Let $\alpha = \sum_{n \geq 1} n s_n$. Then,

$$E(\text{AvgSelRank}_T(X)) = \frac{5/6}{X^{5/6}} \int_0^X \sum_{n \geq 1} \frac{n \cdot \theta_n(H)}{H^{1/6}} dH + O\left(X^{-1/3}\right)$$

$$= \frac{5/6}{X^{5/6}} \left( \int_0^X \frac{\sum_{n \geq 1} n \cdot \theta_n(H) - \alpha}{H^{1/6}} dH + \int_0^X \alpha \frac{H^{1/6}}{H^{1/6}} dH \right) + O\left(X^{-1/3}\right).$$

Now, since $f(X) = \sum_{n \geq 1} n \cdot \theta_n(X) - \alpha$ goes to 0 as $X \to \infty$, it follows that $X^{-5/6} \int_0^X f(H) H^{-1/6} dH$ also vanishes in the limit. Hence,

$$\lim_{X \to \infty} E(\text{AvgSelRank}_T(X)) = \frac{5/6}{X^{5/6}} \int_0^X \frac{\sum_{n \geq 1} n \cdot \theta_n(H)}{H^{1/6}} dH + O\left(X^{-1/3}\right)$$

$$= \lim_{X \to \infty} \frac{5/6}{X^{5/6}} \int_0^X \alpha \frac{H^{1/6}}{H^{1/6}} dH = \alpha = \sum_{n \geq 1} n s_n,$$

as we wanted to prove. 

\[\square\]

**Figure 8.** Values of AvgSelRank($X$) using the BHKSSW database (blue dots), and the corresponding predictions from Proposition 6.8 (red dots).

### 7. Random model, following Cramér, part 2

Let $E/Q$ be an elliptic curve, and let Sel$_2(E/Q)$ and III($E/Q$)[2] be, respectively, the 2-Selmer group of $E/Q$ and the 2-torsion of Sha. We would like to understand how often an element of Sel$_2(E/Q)$ is in the image of $E(Q)/2E(Q)/(E(Q)_{\text{tors}}/2E(Q)_{\text{tors}})$ under the natural injection.
Indeed, it is important to note that if an element of $\text{Sel}_2(E/\mathbb{Q})$ reduces to a non-trivial element in the quotient $\text{Sel}_2(E/\mathbb{Q})/(E(\mathbb{Q})/2E(\mathbb{Q})) \cong \text{III}(E/\mathbb{Q})[2]$. Inspired by the Cohen-Lenstra heuristics for number fields, Delaunay ([9], [10]) has conjectured certain distributions of Tate-Shafarevich groups (see also Section 5 of [23] for a rich account of Delaunay’s conjectures and other related works). As in the case of the results on the density of Selmer ranks discussed in Section 6, Delaunay’s heuristics provide the (conjectural) limit value of the density (i.e., the total probability) of curves with a certain structure of $\text{III}(E/\mathbb{Q})$. However, for our purposes, we are interested in the average size of $\text{III}(E/\mathbb{Q})[2]$ at height $X$, for a curve of fixed Selmer rank $n$. In other words, we are interested in the following conditional probability that measures the failure of the Hasse principle at a given 2-Selmer element of height $X$:

$$\text{pHasse}_n(X) = \text{Prob}(s \in \text{Sel}_2(E/\mathbb{Q}) \text{ is trivial in } \text{III}(E/\mathbb{Q})[2] \mid E \in \mathcal{S}_n \text{ and } \text{ht}(E/\mathbb{Q}) = X).$$

An element $s \in \text{Sel}_2(E/\mathbb{Q})$, in turn, can be visualized as a homogeneous space $H \in \text{WC}(E/\mathbb{Q})$ in the Weil-Châtelet group of $E/\mathbb{Q}$, such that $H$ is locally solvable everywhere, and the quantity $\text{pHasse}_n(X)$ would be realized as the probability of $H(\mathbb{Q})$ having a rational point (see [8]).

At this point, we could measure $\text{pHasse}_n(E/\mathbb{Q})$, the average failure of the Hasse principle for the 2-Selmer elements coming from a fixed elliptic curve $E$ of height $X$, as usual, by

$$\frac{1}{\#\text{III}(E/\mathbb{Q})[2]} = \frac{\#(E(\mathbb{Q})/2E(\mathbb{Q}))}{\#\text{Sel}_2(E/\mathbb{Q})} = \frac{1}{2^{\text{rank}(E(\mathbb{Q}))} - \text{rank}(E(\mathbb{Q}))}. $$

However, this ratio does not capture correctly the probability that a 2-Selmer element is trivial in III. Indeed, it is important to note that if $s, s' \in \text{Sel}_2(E/\mathbb{Q})$ are two distinct elements, then the events $s \equiv 0 \in \text{III}(E/\mathbb{Q})$ and $s' \equiv 0 \in \text{III}(E/\mathbb{Q})$ are in general not independent from a probabilistic point of view. Indeed, if the 2-Selmer rank of $E/\mathbb{Q}$ is $n$, then $\text{Sel}_2(E/\mathbb{Q})$ (modulo 2-torsion contributions) has order $2^n$, but the size of $\text{III}(E/\mathbb{Q})[2]$ is dictated by the classes of $n$ generators $s_1, \ldots, s_n$ of $\text{Sel}_2(E/\mathbb{Q})/(2$-torsion). Thus, a better measure for $\text{pHasse}_n(X)$ may be

$$\frac{\text{rank}_{\mathbb{Z}/2\mathbb{Z}}(E(\mathbb{Q})/2E(\mathbb{Q})) - \text{rank}_{\mathbb{Z}/2\mathbb{Z}} E(\mathbb{Q})[2]}{\text{rank}_{\mathbb{Z}/2\mathbb{Z}}(\text{Sel}_2(E/\mathbb{Q}))/ - \text{rank}_{\mathbb{Z}/2\mathbb{Z}} E(\mathbb{Q})[2]} = \frac{\text{rank}(E(\mathbb{Q}))}{\text{selrank}(E(\mathbb{Q}))}. $$

As it turns out, this ratio is not the correct measure either for odd Selmer rank. If we assume that $\text{III}(E/\mathbb{Q})[2^\infty]$ is finite, then the existence of the Cassels-Tate pairing ([5])

$$\Gamma : \text{III}(E/\mathbb{Q})[2^\infty] \times \text{III}(E/\mathbb{Q})[2^\infty] \to \mathbb{Q}/\mathbb{Z},$$

which is a non-degenerate, alternating, and bilinear, implies that the $F_2$-dimension of $\text{III}(E/\mathbb{Q})[2]$ is always even. It follows that $\text{rank}(E(\mathbb{Q})) \equiv \text{selrank}(E(\mathbb{Q})) \mod 2$. In particular, if $\text{selrank}(E(\mathbb{Q})) = n = 2k$ or $1 + 2k$, then the $F_2$-dimension of $\text{III}(E/\mathbb{Q})[2]$ is in fact dictated by $2k$ classes of $\text{Sel}_2(E/\mathbb{Q})$ (if $n = 1$, then $k = 0$, so we will assume that $n \geq 2$ from now on in this section). Therefore, the correct way to define the failure of the Hasse principle for a given elliptic curve is as follows.
Definition 7.1. Let \( E/\mathbb{Q} \) be an elliptic curve of Selmer rank \( n \geq 2 \). We define the average ratio of failure of the Hasse principle of the 2-Selmer elements of \( E/\mathbb{Q} \) by

\[
pHasse_n(E/\mathbb{Q}) = \begin{cases} 
\frac{\text{rank}(E(\mathbb{Q}))}{\text{selrank}(E(\mathbb{Q}))} & \text{if } n \text{ is even, and} \\
\frac{\text{rank}(E(\mathbb{Q})) - 1}{\text{selrank}(E(\mathbb{Q})) - 1} & \text{if } n \text{ is odd.}
\end{cases}
\]

We note here that, in all cases, we have

\[
pHasse_n(E/\mathbb{Q}) = \frac{\text{rank}(E(\mathbb{Q})) - (n \mod 2)}{n - (n \mod 2)}.
\]

Remark 7.2. The fact that the \( \mathbb{F}_2 \)-dimension of \( \text{III}(E/\mathbb{Q})[2] \) is even implies that \( n = \text{selrank}(E(\mathbb{Q})) \) and \( \text{rank}(E(\mathbb{Q})) \) have the same parity. Thus, the rank of \( E(\mathbb{Q}) \) is determined by \( [n/2] \) pairs of generators \( \{(s_1, \hat{s}_1), (s_2, \hat{s}_2), \ldots, (s_{[n/2]}, \hat{s}_{[n/2]})\} \) of \( \text{Sel}_2(E/\mathbb{Q}) \) such that \( s_i \equiv 0 \in \text{III} \) if and only if \( \hat{s}_i \equiv 0 \in \text{III} \). Indeed, let us assume that \( n \geq 2 \), and first assume that \( n \) is even, \( n = 2k \). Let \( E/\mathbb{Q} \) be an elliptic curve of Selmer rank \( n \), and let \( s_1 \) be an arbitrary element of \( \text{Sel}_2(E/\mathbb{Q})/(2\text{-torsion}) \). We distinguish two cases:

- If \( s_1 \equiv 0 \in \text{III} \), then \( \text{dim}_{\mathbb{F}_2}(\text{III}(E/\mathbb{Q})[2]) \) is now at most \( 2k - 2 \), and this means that there exists a Selmer element \( \hat{s}_1 \), linearly independent from \( s_1 \), such that \( \hat{s}_1 \equiv 0 \in \text{III} \) as well.
- Otherwise, if \( s_1 \) represents a non-trivial element in \( \text{III} \), and if \( \Gamma: \text{III}(E/\mathbb{Q})[2] \times \text{III}(E/\mathbb{Q})[2] \to \mathbb{F}_2 \) is the Cassels-Tate (non-degenerate, alternating, and bilinear) pairing, than we can choose a non-trivial element \( \hat{s}_1 \in \text{III}(E/\mathbb{Q})[2] \) such that \( \Gamma([s_1],[\hat{s}_1]) = 1 \). In particular, \([\hat{s}_1]\) is linearly independent of the class of \( s_1 \) in \( \text{III} \), and therefore if \( \hat{s}_1 \) is now any Selmer element representing the same class \([\hat{s}_1]\) of \( \text{III} \), then \( \hat{s}_1 \) and \( s_1 \) are also linearly independent in \( \text{Sel}_2 \).

In either case, we have found a pair \( (s_1, \hat{s}_1) \) of Selmer elements such that \( s_i \equiv 0 \in \text{III} \) if and only if \( \hat{s}_i \equiv 0 \in \text{III} \). We can continue this process to find pairs \( (s_1, \hat{s}_1), \ldots, (s_k, \hat{s}_k) \). Now let \( n = 1 + 2k \) be odd. The proof is analogous, except that if \( \text{dim}_{\mathbb{F}_2} \text{Sel}_2(E/\mathbb{Q}) \) is odd, then \( \text{dim}_{\mathbb{F}_2}(\text{III}(E/\mathbb{Q})[2]) \) must be even, and so there is automatically a Selmer element \( s_0 \) that is trivial in \( \text{III} \). Now we can proceed as above to find pairs \( (s_1, \hat{s}_1), \ldots, (s_k, \hat{s}_k) \) such that \( s_i \equiv 0 \in \text{III} \) if and only if \( \hat{s}_i \equiv 0 \in \text{III} \).

Thus, it may be best to define

\[
pHasse_n(E/\mathbb{Q}) = \frac{\frac{1}{2}(\text{rank}(E(\mathbb{Q})) - (n \mod 2))}{\frac{1}{2}(\text{selrank}(E(\mathbb{Q})) - (n \mod 2))}
\]

but, of course, the factors of \( \frac{1}{2} \) cancel out and this definition is equivalent to the one given above. This simple remark will be crucial when computing the probability of a given Mordell–Weil rank \( r \) among curves of Selmer rank \( n \) in Theorem 8.3.

Remark 7.3. Let \( E = (X,n,\text{Sel}_2) \) be a test elliptic curve, as in Definition 5.1. The same considerations stated in this section about the parity of \( n \) explain our reasons to define \( \text{Sel}_2 \) as a vector of \( \lfloor \frac{n}{2} \rfloor = (n - (n \mod 2)) / 2 \) symbols \( s_{E,1}, \ldots, s_{E,\lfloor \frac{n}{2} \rfloor} \) in \( \{\text{III},\text{MW}\} \).

Now, we turn our attention back to test elliptic curves and our Cramér-like model. First, we define the (MW) rank of a test elliptic curve.
**Definition 7.4.** We define the rank (or MW rank) of a test elliptic curve \( E = (X, n, \text{Sel}_2) \), as follows:

\[
\text{rank}(E) = (n \mod 2) + 2 \cdot \#\{\text{MW elements in } \text{Sel}_2(E)\}.
\]

**Remark 7.5.** If \( n = 1 \) and \( E = (X, 1, \text{Sel}_2) \) is a test curve in \( \mathcal{S}^X \), then \( \text{Sel}_2(E) \) is empty, and \( \text{rank}(E) = (n \mod 2) = 1 \), so we will concentrate on the case of \( n \geq 2 \) from now on.

**Example 7.6.** Let \( E \) and \( E' \) be the test elliptic curves given by

\[
E = (107245762628, 4, (\text{MW}, \text{III})), \text{ and } E' = (18932679356, 3, (\text{III}))
\]

that appeared in Example 5.3. Then,

\[
\text{rank}(E) = (4 \mod 2) + 2 \cdot 1 = 2, \text{ and } \text{rank}(E') = (3 \mod 2) + 2 \cdot 0 = 1.
\]

We are ready to translate our remarks above into a hypothesis for a probabilistic model of test Selmer elements, and define probability spaces \( (\text{Sel}^X_{2,n}, \mathcal{F}^X_n, P^X_n) \) and \( (\overline{\text{Sel}}^X_{2,n}, \overline{\mathcal{F}}^X_n, P^X_n) \) as follows.

**Hypothesis 7.7** (Hypothesis B, or \( H_B \)). Let \( n \geq 2 \) be fixed, let \( X \geq 1 \), and define

\[
\text{Sel}^X_{2,n} = \bigcup_{E \in \mathcal{S}^X_n} \{ s_{E,i} : 1 \leq i \leq \lfloor n/2 \rfloor \} = \bigcup_{E \in \mathcal{S}^X_n} \{ s_{E,1}, \ldots, s_{E,\lfloor n/2 \rfloor} \}
\]

where the union is over test elliptic curves \( E = (X, n, \text{Sel}_2(E)) \) of fixed height \( X \) and fixed Selmer rank \( n \), and \( \text{Sel}_2 = \{ s_{E,1}, \ldots, s_{E,\lfloor n/2 \rfloor} \} \). If \( m \geq 2 \) is fixed, and \( \overline{X} = (X_1, \ldots, X_m) \) is a vector of arbitrary heights \( X_i \geq 1 \), we let \( \overline{\text{Sel}}^X_{2,n} = \prod_{i=1}^m \text{Sel}^X_{2,n} \).

(a) Let \( \rho_n(X) \) be a function \([1, \infty) \to (0, 1)\) such that \( \lim_{X \to \infty} \rho_n(X) = 0 \). We define a probability space \( (\text{Sel}^X_{2,n}, \mathcal{F}^X_n, P^X_n) \) by defining a probability measure \( P^X_n \) as follows:

- \( \text{MW}^X_n \) is the subset of \( \text{Sel}^X_{2,n} \) of MW symbols, and \( \text{III}^X_n = \text{Sel}^X_{2,n} \setminus \text{MW}^X_n \) is the subset of III symbols,
- \( \mathcal{F}^X_n = \{ \emptyset, \text{MW}^X_n, \text{III}^X_n, \text{Sel}^X_{2,n} \} \),
- \( P^X_n(\text{MW}^X_n) = \rho_n(X) \), and \( P^X_n(\text{III}^X_n) = 1 - \rho_n(X) \).

In other words, \( \mathcal{F}^X_n \) and \( P^X_n \) are chosen so that the random variable \( Y_{\text{Hasse},n,X} : \text{Sel}^X_{2,n} \to \{0, 1\} \) that takes values

\[
Y_{\text{Hasse},n,X}(s) = \begin{cases} 
1 & \text{if } s \in \text{MW}^X_n, \\
0 & \text{if } s \in \text{III}^X_n.
\end{cases}
\]

is \( \mathcal{F}^X_n \)-measurable and \( Y_{\text{Hasse},n,X} \) follows a Bernoulli distribution with probability \( \rho_n(X) \).

(b) Fix \( i \) in the range \( 1 \leq i \leq m \), and let \( Y_i : \text{Sel}^X_{2,n} \to \{0,1\} \) be the random variable defined by \( Y_i(s) = Y_{\text{Hasse},n,X}(s_i) \) for any \( s = (s_1, \ldots, s_m) \in \text{Sel}^X_{2,n} \). Then, we endow \( \text{Sel}^X_{2,n} \) with a \( \sigma \)-algebra \( \mathcal{F}^X_n \) and a probability measure \( P^X_n \) such that:

(b.1) Let \( s = (s_1, \ldots, s_m) \in \text{Sel}^X_{2,n} \) be a vector of test Selmer elements chosen at random. Then, the random variables \( Y_1, \ldots, Y_m \) are mutually independent, and each \( Y_i \) follows a Bernoulli distribution with probability \( \rho_n(X) \).

(b.2) Suppose \( m = \lfloor n/2 \rfloor \), and \( X = X_1 = \ldots = X_m \), so that \( \text{Sel}_2(E) = (s_{E,1}, \ldots, s_{E,\lfloor n/2 \rfloor}) \in \text{Sel}^X_{2,n} \), for any \( E \in \mathcal{S}^X_n \). Let \( 1 \leq k \leq \lfloor n/2 \rfloor \), and \( 1 \leq i_1 < i_2 < \ldots < i_k \leq \lfloor n/2 \rfloor \), and write \( S = \bigcup_{E \in \mathcal{S}^X_n} \text{Sel}_2(E) \). Then,
Definition 7.8. We say that the random variables $Y_1, \ldots, Y_m$ are equicorrelated if

\[ \mathbb{E}(Y_{i_1} Y_{i_1} \cdots Y_{i_k}) \]

only depends on $n, k,$ and $X$, and it is independent of the choice of indices $1 \leq i_1 < \ldots < i_k \leq m$.

Remark 7.9. The equicorrelation condition of $H_B$, part (2), does not add any conditions at all when $n = 2, 3$. When $n = 4, 5$, equicorrelation simply says that $\mathbb{E}(Y_1|S) = \mathbb{E}(Y_2|S)$, where $S = \bigcup_{E \in \mathcal{E}} \text{Sel}_2(E)$. This is already implied by the assumption that $Y_1$ and $Y_2$ follow the same Bernoulli distribution (so in fact $\mathbb{E}(Y_1) = \mathbb{E}(Y_1|S) = \mathbb{E}(Y_2) = \rho_n(X)$). However, the equicorrelation does add new information about the random variables $\{Y_i\}$ for $n \geq 6$. For instance, when $n = 6$, it says that

\[ \mathbb{E}(Y_1 Y_2|S) = \mathbb{E}(Y_1 Y_3|S) = \mathbb{E}(Y_2 Y_3|S) \]

where $\mathbb{E}(Y|S) = \mathbb{E}(Y|S)$. When $n = 8$, it says that

\[ \mathbb{E}(Y_1 Y_2|S) = \mathbb{E}(Y_1 Y_3|S) = \mathbb{E}(Y_1 Y_4|S) = \mathbb{E}(Y_2 Y_3|S) = \mathbb{E}(Y_2 Y_4|S) = \mathbb{E}(Y_3 Y_4|S), \]

and also

\[ \mathbb{E}(Y_1 Y_2 Y_3|S) = \mathbb{E}(Y_1 Y_2 Y_4|S) = \mathbb{E}(Y_1 Y_3 Y_4|S) = \mathbb{E}(Y_2 Y_3 Y_4|S). \]

8. The probability that a 2-Selmer element is globally solvable

The following two results describe the effects of equicorrelation on the covariance of the random variables. We remind the reader that the covariance of two random variables $V, W$ is given by $\text{Cov}(V, W) = \mathbb{E}(VW) - \mathbb{E}(V) \cdot \mathbb{E}(W)$.

Lemma 8.1. Let $Z, Z', W, W'$ be random variables such that $\mathbb{E}(Z) = \mathbb{E}(Z')$, $\mathbb{E}(W) = \mathbb{E}(W')$. Then, $\text{Cov}(Z, W) = \text{Cov}(Z', W')$ if and only if $\mathbb{E}(ZW) = \mathbb{E}(Z'W')$, if and only if $\mathbb{E}((1 - Z)W) = \mathbb{E}((1 - Z')W')$.

Proof. By definition $\mathbb{E}(ZW) = \mathbb{E}(Z)\mathbb{E}(W) + \text{Cov}(Z, W)$. Thus,

\[
\begin{align*}
\mathbb{E}(ZW) - \mathbb{E}(Z'W') &= \mathbb{E}(Z)\mathbb{E}(W) + \text{Cov}(Z, W) - (\mathbb{E}(Z')\mathbb{E}(W') + \text{Cov}(Z', W')) \\
&= \mathbb{E}(Z)\mathbb{E}(W) - \mathbb{E}(Z')\mathbb{E}(W') + \text{Cov}(Z, W) - \text{Cov}(Z', W') \\
&= \text{Cov}(Z, W) - \text{Cov}(Z', W').
\end{align*}
\]

Thus, $\text{Cov}(Z, W) = \text{Cov}(Z', W')$ if and only if $\mathbb{E}(ZW) = \mathbb{E}(Z'W')$. Similarly,

\[
\begin{align*}
\mathbb{E}((1 - Z)W) - \mathbb{E}((1 - Z')W') &= \mathbb{E}(1 - Z)\mathbb{E}(W) + \text{Cov}(1 - Z, W) - (\mathbb{E}(1 - Z')\mathbb{E}(W') + \text{Cov}(1 - Z', W')) \\
&= (1 - \mathbb{E}(Z))\mathbb{E}(W) - (1 - \mathbb{E}(Z'))\mathbb{E}(W') + \text{Cov}(Z, W) + \text{Cov}(Z', W') \\
&= - \text{Cov}(Z, W) + \text{Cov}(Z', W'),
\end{align*}
\]

as claimed, where we have used the fact that $\text{Cov}(a + bX, Y) = b\text{Cov}(X, Y)$, for any constants $a, b$ and random variables $X, Y$. \qed
Proposition 8.2. Assume $H_B$, let $Y_{\text{Hasse},n,X}$ be the random variable defined in $H_B$, let $E \in \tilde{S}_n^X$ be chosen at random, and let $Y_i = Y_{\text{Hasse},n,X}(s_{E,i})$ for $1 \leq i \leq \lfloor n/2 \rfloor$. Let $1 \leq s, t \leq \lfloor n/2 \rfloor$ with $s + t \leq \lfloor n/2 \rfloor$. Then, there is a function $C_{s,t}^n(X)$ such that

$$C_{s,t}^n(X) = \text{Cov}(Y_1, Y_2 \ldots Y_i, S, Y_{k_1}, Y_{k_2} \ldots Y_{k_t} | S)$$

for any sets of indices $1 \leq i_1 < i_2 < \ldots < i_s \leq \lfloor n/2 \rfloor$ and $1 \leq k_1 < k_2 < \ldots < k_t \leq \lfloor n/2 \rfloor$ with $\{i_u\} \cap \{k_v\} = \emptyset$.

Proof. Let $1 \leq i_1 < i_2 < \ldots < i_s \leq m$ and $1 \leq k_1 < k_2 < \ldots < k_t \leq m$ with $\{i_u\} \cap \{k_v\} = \emptyset$, and let $1 \leq i'_1 < i'_2 < \ldots < i'_s \leq m$ and $1 \leq k'_1 < k'_2 < \ldots < k'_t \leq m$ with $\{i'_u\} \cap \{k'_v\} = \emptyset$ be another set of indices. By $H_B$, part (b.2), the random variables are equicorrelated when restricted to $S = \bigcup_{E \in \tilde{S}_n^X} \text{Sel}_2(E)$, i.e., $\text{E}_n^s(X) = \text{E}(Y_1, Y_2 \ldots Y_i | \text{Sel}_2(E)) = \text{E}(Y_{i_1}, Y_{i_2} \ldots Y_{i_s} | \text{Sel}_2(E))$ and similarly $\text{E}_n^t(X) = \text{E}(Y_{k_1}, Y_{k_2} \ldots Y_{k_t} | \text{Sel}_2(E))$ and also

$$\text{E}_{s+t}(X) = \text{E}(Y_1, Y_2 \ldots Y_i, S | \text{Sel}_2(E)) = \text{E}(Y_{i_1}, Y_{i_2} \ldots Y_{i_s}, S | \text{Sel}_2(E))$$

Then, we can apply Lemma 8.1 with $Z = Y_{i_1}, Y_{i_2} \ldots Y_{i_s}$, $W = Y_{k_1}, Y_{k_2} \ldots Y_{k_t}$, and $W' = Y_{k'_1}, Y_{k'_2} \ldots Y_{k'_t}$, to obtain the equality of the covariance terms. Thus, the covariance is independent of the chosen sets of $s$ and $t$ distinct random variables in $\{Y_i\}$, and in fact it only depends on $n$, $s$, $t$, and $X$.

Hypothesis B asserts that $Y_{\text{Hasse},n,X} \sim B(1, \rho(X))$, i.e., $Y_{\text{Hasse},n,X}$ follows a binomial distribution with one trial. Now we want to reconstruct the distribution of the rank of a test curve $E \in \tilde{S}_n^X$ from that of $Y_{\text{Hasse},n,X}$. We remind the reader that if $\overline{X} = (X_1, \ldots, X_n)$ is a vector of heights, then we defined $\text{Sel}_{2,n}^X = \prod_{i=1}^m \text{Sel}_{2,i}^X$.

Theorem 8.3. Let $n \geq 1$ be fixed, assume $H_B$, let $R_n = \{0, 1, \ldots, \lfloor n/2 \rfloor\}$, let $X_i \geq 1$, and let $\overline{X} = (X_1, \ldots, X_{\lfloor n/2 \rfloor})$ be a vector of heights. Let $\text{rank}_{n,X}: \text{Sel}_{2,n}^X \rightarrow R_n$ be the function given by the random variable $Y_1 + \ldots + Y_{\lfloor n/2 \rfloor}$ if $n \geq 2$ (and equal 0 if $n = 1$), where $Y_i(s) = Y_{\text{Hasse},n,X}(s_i)$ for any $s = (s_1, \ldots, s_n) \in \text{Sel}_{2,n}^X$. Then:

1. If $X = X_1 = \ldots = X_{\lfloor n/2 \rfloor}$, $E = (X, n, \text{Sel}_2)$ is a test elliptic curve and $s = \text{Sel}(E)$, then

$$\text{rank}_{n,X}(s) = \frac{\text{rank}(E) - (n \text{ mod } 2)}{2}$$

where $\text{rank}(E)$ appeared in Definition 7.4.

2. If $n \geq 2$, and $S = \bigcup_{E \in \tilde{S}_n^X} \text{Sel}_2(E)$, then the expected value and variance of $\text{rank}_{n,X}$ in $S$ are given by

$$\text{E}(\text{rank}_{n,X} | S) = \lfloor n/2 \rfloor \cdot \rho_n(X), \text{ and}$$

$$\text{Var}(\text{rank}_{n,X} | S) = \lfloor n/2 \rfloor \cdot (\rho_n(X)(1 - \rho_n(X))) + (\lfloor n/2 \rfloor - 1) \cdot C_{1,1}^n(X),$$

where $C_{1,1}^n(X) = \text{Cov}(Y_1, Y_j)$ is the covariance function of any two random variables given by Proposition 8.2.

3. If the events $\{Y_i = 1 | S\}$ were mutually independent (resp. approximately uncorrelated, i.e., if $C_{1,1}^n(X) \approx 0$), then $\text{rank}_{n,X} | S$ follows (resp. approximately) a binomial distribution of the form $B(\lfloor n/2 \rfloor, \rho_n(X))$, with expected value $\lfloor n/2 \rfloor \cdot \rho_n(X)$ and variance $\lfloor n/2 \rfloor \cdot \rho_n(X)(1 - \rho_n(X))$. 

(4) If \( E_i \in \tilde{S}^X_n \) are chosen at random for \( 1 \leq i \leq m \), and \( s_i = \text{Sel}_2(E_i) \), and \( r_i \in R_n \), then the events \( \{\text{rank}_{n,X}(s_i) = r_i\} \) are mutually independent.

**Proof.** For part (1), note that if \( n = 1 \), then \( \text{rank}(E) = 1 \) (see Remark 7.5) and, therefore \( \text{rank}_{n,X}(E) = (\text{rank}(E) - (n \mod 2))/2 = 0 \). For the rest of the proof, let us assume that \( n \geq 2 \). If \( X = X_1 = \ldots = X_{\lfloor n/2 \rfloor} \), \( E = (X, n, \text{Sel}_2) \) is a test elliptic curve and \( s = \text{Sel}_2(E) = (s_{E,1}, \ldots, s_{E,\lfloor n/2 \rfloor}) \), then

\[
\frac{1}{2}(\text{rank}(E) - (n \mod 2)) = \# \{1 \leq i \leq \lfloor n/2 \rfloor : s_{E,i} = \text{MW}\} = \sum_{i=1}^{\lfloor n/2 \rfloor} Y_{\text{Hasse},n,X}(s_{E,i}) = \sum_{i=1}^{\lfloor n/2 \rfloor} Y_i(s) = \text{rank}_{n,X}(s).
\]

For (2), we first compute the expected value of \( \text{rank}_{n,X}(s) \) in \( S \):

\[
\mathbb{E}(\text{rank}_{n,X}(s)|s \in S) = \mathbb{E}\left( \sum_{i=1}^{\lfloor n/2 \rfloor} Y_i(s) | s \in S \right) = \sum_{i=1}^{\lfloor n/2 \rfloor} \mathbb{E}(Y_i(s) | s \in S) = \lfloor n/2 \rfloor \cdot \rho_n(X),
\]

since each \( Y_i \sim Y_{\text{Hasse},n,X} \sim B(1, \rho_n(X)) \) by Hypothesis B. Let us now calculate the variance of \( \text{rank}_{n,X} = \sum Y_i \) in \( S \).

\[
\text{Var}(\text{rank}_{n,X} | S) = \text{Var}\left( \sum_{i=1}^{\lfloor n/2 \rfloor} Y_i | S \right)
\]

\[
= \sum_{i=1}^{\lfloor n/2 \rfloor} \text{Var}(Y_i | S) + 2 \cdot \sum_{1 \leq i < j \leq \lfloor n/2 \rfloor} \text{Cov}(Y_i | S, Y_j | S)
\]

\[
= \sum_{i=1}^{\lfloor n/2 \rfloor} \text{Var}(Y_i) + 2 \cdot \left( \frac{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 1)}{2} \right) \cdot C_{1,1}^n(X)
\]

\[
= [\lfloor n/2 \rfloor] \cdot \rho_n(X)(1 - \rho_n(X)) + [\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 1) - 1] \cdot C_{1,1}^n(X)
\]

\[
= [\lfloor n/2 \rfloor] \cdot \rho_n(X)(1 - \rho_n(X)) + ([\lfloor n/2 \rfloor - 1] - 1 \cdot C_{1,1}^n(X),
\]

where we have used the properties of the variance, the fact that for any \( i \neq j \), we have \( \text{Cov}(Y_i, Y_j) = C_{1,1}^n(X) \) for all \( i \neq j \) by Proposition 8.2, and \( Y_i | S \sim Y_i \sim B(1, \rho_n(X)) \). This proves (2).

In particular, if the random variables \( Y_i \) in \( S \) were independent samples of a Bernoulli distribution (or similarly if \( C_{1,1}^n(X) \approx 0 \)), then \( \text{rank}_{n,X} = \sum Y_i \) would follow a binomial distribution \( B([n/2], \rho_n(X)) \). This proves (3).

For (4), we will show that if \( E \in \tilde{S}^X_n \) and \( E' \in \tilde{S}^{X'}_n \), and \( s = \text{Sel}_2(E) \), \( s' = \text{Sel}_2(E') \), and \( r, r' \in R_n \), then the events \( \{\text{rank}_{n,X}(s) = r\} \) and \( \{\text{rank}_{n,X'}(s') = r'\} \) are independent. We write \( \text{rank}_{n,X}(s) = \sum Y_i \) and \( \text{rank}_{n,X'}(s') = \sum Y'_j \). We claim that the variables \( Y_i \) and \( Y'_j \) are independent, for any choice of \( i \) and \( j \). Indeed, consider \( (s_i, s'_j) \in \text{Sel}_2^n \times \text{Sel}_2^{X'}, \) which is a random element by our random choice of \( E \) and \( E' \). Then, Hypothesis B (part (b.1)) says that the events \( Y_{\text{Hasse},n,X}(s_i) = 1 \) and \( Y_{\text{Hasse},n,X}(s'_j) = 1 \) are independent (that is, \( \mathbb{E}(Y_i Y'_j) = \mathbb{E}(Y_i) \mathbb{E}(Y'_j) \)). Thus, \( Y = \sum Y_i \) and
\[ Y' = \sum Y'_i \] are also independent:

\[
\mathbb{E}(YY') = \mathbb{E}\left( \left( \sum_i Y_i \right) \left( \sum_j Y'_j \right) \right) = \mathbb{E}\left( \sum_{i,j} Y_i Y'_j \right) = \sum_{i,j} \mathbb{E}(Y_i Y'_j) = \sum_{i,j} \mathbb{E}(Y_i) \mathbb{E}(Y'_j) 
\]

\[
= \left( \sum_i \mathbb{E}(Y_i) \right) \left( \sum_j \mathbb{E}(Y'_j) \right) = \left( \mathbb{E} \left( \sum_i Y_i \right) \right) \left( \mathbb{E} \left( \sum_j Y'_j \right) \right) = \mathbb{E}(Y') \mathbb{E}(Y').
\]

This completes the proof of (4) and of the theorem. \( \square \)

Using Theorem 8.3, we shall describe the average rank and distribution of curves by Mordell–Weil rank in a sample set of test curves of Selmer rank \( n \).

**Corollary 8.4.** Let \( E_1, \ldots, E_m \) be test elliptic curves chosen at random of Selmer rank \( n \) and heights \( X_1, \ldots, X_m \). Then, the expected value of the average rank is

\[
\mathbb{E} \left( \frac{1}{m} \sum_{i=1}^m \text{rank}(E_i) \right) = (n \mod 2) + \frac{2[n/2]}{m} \sum_{i=1}^m \rho_n(X_i)
\]

with standard error given by

\[
\frac{1}{m} \sqrt{\left| \frac{n}{2} \right| \sum_{i=1}^m (\rho_n(X_i)(1 - \rho_n(X_i)) + (\left| n/2 \right| - 1)C_{1,1}^n(X_i))}
\]

**Proof:** Let \( E_1, \ldots, E_m \) be as in the statement. Then, Theorem 8.3 gives us the expected value and variance of \( \text{rank}_{n,X_i} \left( \text{Sel}_{2}(E_i) \right) = \text{rank}(E_i) - (n \mod 2)/2 \) and therefore we can compute the expected value.

\[
\mathbb{E} \left( \frac{1}{m} \sum_{i=1}^m \text{rank}(E_i) - (n \mod 2) \right) = \frac{1}{m} \sum_{i=1}^m \mathbb{E} \left( \text{rank}_{n,X_i} \left( \text{Sel}_{2}(E_i) \right) \right) = \frac{1}{m} \sum_{i=1}^m \left| n/2 \right| \rho_n(X_i),
\]

since \( \mathbb{E}(\text{rank}_{n,X_i} | S) = |n/2| \rho_n(X_i) \). Thus,

\[
\mathbb{E} \left( \frac{1}{m} \sum_{i=1}^m \text{rank}(E_i) \right) = (n \mod 2) + \frac{2[n/2]}{m} \sum_{i=1}^m \rho_n(X_i),
\]

as claimed. Next, Theorem 8.3, part (4), shows that the values of the random variables \( Z_i = \text{rank}_{n,X_i} \left( \text{Sel}_{2}(E_i) \right) \) are independent because the test curves \( \{E_i\} \) are chosen at random. In particular, the covariance \( \text{Cov}(Z_i, Z_j) = 0 \) for all \( i \neq j \), and it follows that \( \text{Var}(Z_i + Z_j) = \text{Var}(Z_i) + \text{Var}(Z_j) + 2 \text{Cov}(Z_i, Z_j) = \text{Var}(Z_i) + \text{Var}(Z_j) \). Hence, we can compute the variance as follows:

\[
\text{Var} \left( \frac{1}{m} \sum_{i=1}^m \text{rank}_{n,X_i} \left( \text{Sel}_{2}(E_i) \right) \right) = \frac{1}{m^2} \sum_{i=1}^m \text{Var} \left( \text{rank}_{n,X_i} \left( \text{Sel}_{2}(E_i) \right) \right)
\]

\[
= \frac{1}{m^2} \sum_{i=1}^m \left| n/2 \right| \cdot (\rho_n(X_i)(1 - \rho_n(X_i)) + (\left| n/2 \right| - 1) \cdot C_{1,1}^n(X_i)),
\]

and therefore the standard error is given by

\[
\sqrt{\frac{|n/2|}{m^2} \sum_{i=1}^m (\rho_n(X_i)(1 - \rho_n(X_i)) + (\left| n/2 \right| - 1)C_{1,1}^n(X_i))}
\]
as desired.

Before we go on to describe the probability of a given Mordell–Weil rank, we need a result on equicorrelated random variables.

**Lemma 8.5.** Suppose that the random variables \( \{Y_i\}_{i=1}^n \) are equicorrelated. Then:

1. For any \( 1 \leq m \leq n \), and any indices \( 1 \leq i_1 < \cdots < i_m \leq n \) and \( 1 \leq i'_1 < \cdots < i'_m \leq n \),
   \[
   \mathbb{E}((1 - Y_{i_1}) \cdots (1 - Y_{i_m})) = \mathbb{E}((1 - Y_{i'_1}) \cdots (1 - Y_{i'_m})).
   \]
2. If \( X \) and \( \{Y_i\} \) are all distinct equicorrelated random variables, and \( 1 \leq m \leq n \), then:
   \[
   \text{Cov}(X, (1 - Y_1)(1 - Y_2) \cdots (1 - Y_m)) = \sum_{i=1}^m (-1)^i \binom{m}{i} \text{Cov} \left(X, \prod_{k=1}^i Y_k\right).
   \]

**Proof.** Part (1) can be easily shown via induction on \( m \), where the induction step was essentially proved in Lemma 8.1. For part (2), we note that \( \text{Cov}(X, (1 - Y_1)(1 - Y_2) \cdots (1 - Y_m)) \) equals

\[
= \text{Cov} \left(X, 1 - \left(\sum_i Y_i\right) + \left(\sum_{i \neq j} Y_i Y_j\right) + \cdots + (-1)^m \prod_i Y_i\right)
\]

\[
= -\sum_i \text{Cov}(X, Y_i) + \sum_{i \neq j} \text{Cov}(X, Y_i Y_j) + \cdots + (-1)^m \text{Cov} \left(X, \prod_i Y_i\right)
\]

\[
= \sum_{i=1}^m (-1)^i \binom{m}{i} \text{Cov} \left(X, \prod_{k=1}^i Y_k\right),
\]

where we have used \( \text{Cov}(X, \prod_{s=1}^t Y_{i_s}) = \text{Cov}(X, Y_1 \cdots Y_t) \) for any indices \( 1 \leq i_1 < \cdots < i_t \leq m \) by Proposition 8.2.

**Remark 8.6.** Let us introduce some more notation to simplify our formulas. By Lemmas 8.1 and 8.5, if \( Y_1, \ldots, Y_m \) are distinct equicorrelated random variables, and \( 1 \leq s, t \leq m \), then the value of

\[
\mathbb{E}_{s,t} = \mathbb{E}(Y_1 \cdots Y_s (1 - Y_{s+1}) \cdots (1 - Y_{s+t})),
\]

is independent for any set of \( s + t \) distinct indices \( \{i_k\}_{k=1}^{s+t} \subseteq \{1, \ldots, m\} \). When the random variables \( Y_1, \ldots, Y_{n/2} \) are the ones given by Hypothesis \( H_B \), we will write \( \mathbb{E}_{s,t}^n(X) = \mathbb{E}_{s,t} \), or simply \( \mathbb{E}_{s,t}^n \), to indicate the expected value of a product of random variables as in Equation (3) above (which extends the notation \( \mathbb{E}_k^n(X) = \mathbb{E}(Y_1 \cdots Y_k|S) \) of \( H_B \)). We also write \( \mathbb{E}_{0,0}^1(X) = 1 \). The following lemma gives recursive formulas to compute any expected value \( \mathbb{E}_{s,t}^n \).

**Lemma 8.7.** Let \( Y_1, \ldots, Y_{n/2} \) be the random variables given by Hypothesis \( H_B \) (which are equicorrelated in \( S \)), let \( 0 \leq s, t \leq \lfloor n/2 \rfloor \), and let \( C_{s,t}^n(X) \) be the covariance coefficient of Proposition 8.2. Then, with notation as in Remark 8.6, we have identities

1. \( \mathbb{E}_{1,0}^n = \rho_n(X) \) and \( \mathbb{E}_{0,1}^n = 1 - \mathbb{E}_{1,0}^n = 1 - \rho_n(X) \).
2. If \( s \geq 1 \), then \( \mathbb{E}_{s,0}^n = \mathbb{E}_{s-1,0}^n \cdot \mathbb{E}_{0,0}^n + C_{s-1,1}^n(X) \).
3. If \( t \geq 1 \), then \( \mathbb{E}_{0,t}^n = \mathbb{E}_{0,t-1}^n \cdot \mathbb{E}_{0,0}^n - \sum_{i=1}^{t-1} (-1)^i (t-1-i) C_{1,i}^n(X) \).
4. \( \mathbb{E}_{s,t}^n = \mathbb{E}_{s,0}^n \cdot \mathbb{E}_{0,t}^n + \sum_{i=1}^t (-1)^i \binom{t}{i} C_{s,i}^n(X) \).
Proof. Throughout the proof, we will assume we are computing (conditional) expected values in $S$, so we will omit $S$ from the notation (i.e., $E(Y)$ means $E(Y|S)$ and $\text{Cov}(Y, Z) = \text{Cov}(Y|S, Z|S)$).

1. $E^n_{1,0} = E(Y_1) = \rho_n(X)$ and $E^n_{0,1} = E(1 - Y_1) = 1 - E^n_{1,0}$.
2. If $s \geq 1$, then $E^n_{s,0} = E(Y_1 \cdots Y_{s-1})E(Y_s) + \text{Cov}(Y_1 \cdots Y_{s-1}, Y_s) = E^n_{s-1,0} \cdot E^n_{1,0} + C^n_{s-1,1}(X)$.
3. If $t \geq 1$, then

$$E^n_{0,t} = E((1 - Y_1) \cdots (1 - Y_{t-1}))(1 - Y_t) + \text{Cov}((1 - Y_1) \cdots (1 - Y_{t-1}), 1 - Y_t)$$

and the covariance term equals $-\text{Cov}((1 - Y_1) \cdots (1 - Y_{t-1}), Y_t)$ which in turn is

$$\sum_{i=1}^{t-1} (-1)^i \binom{t-1}{i} \text{Cov}(Y_t, \prod_{k=1}^{i} Y_k) = \sum_{i=1}^{t-1} (-1)^i \binom{t-1}{i} C^n_{t,i}(X)$$

by Lemma 8.5.

4. $E^n_{s,t} = E^n_{s,0} \cdot E^n_{0,t} + \text{Cov}(Y_1 \cdots Y_s, 1 - Y_t)$ and, by Lemma 8.5, the covariance term equals $\sum_{i=1}^{t} (-1)^i \binom{t}{i} C^n_{s,i}(X)$ as claimed.

\[\square\]

Corollary 8.8. Let us assume $H_B$. Let $n \geq 2$ be fixed, let $0 \leq j \leq \lfloor n/2 \rfloor$, and let $\bar{X} = (X_1, \ldots, X_{\lfloor n/2 \rfloor})$ be a vector of heights $X_i \geq 1$. Then:

1. The random variable $Y_{rk=n-2j} : \text{Sel}_{2,n} \to \{0, 1\}$ given by

$$Y_{rk=n-2j}(s) = \begin{cases} 1 & \text{if } \text{rank}_{n,X}(s) = n - 2j, \\ 0 & \text{otherwise}, \end{cases}$$

is given by

$$Y_{rk=n-2j} = \sum_{1 \leq k_1 < \cdots < k_{m(j)} \leq \lfloor n/2 \rfloor} Y_{k_1} \cdot Y_{k_2} \cdots Y_{k_{m(j)}} \cdot \frac{\prod_{i=1}^{\lfloor n/2 \rfloor} (1 - Y_i)}{(1 - Y_{k_1}) \cdots (1 - Y_{k_{m(j)}})}$$

where $m(j) = \lfloor n/2 \rfloor - j$, and the random variables $Y_i(s) = Y_{\text{Hasse},n,X}(s_i)$ are as given by Hypothesis B, such that $\text{rank}_{n,X} = \sum Y_i$.

2. Suppose that $X = X_1 = \ldots X_{\lfloor n/2 \rfloor}$. The expected value of $Y_{rk=n-2j}$ is given by

$$E(Y_{rk=n-2j}) = \left( \begin{array}{c} \lfloor n/2 \rfloor \\ j \end{array} \right) \rho_n(X)^{\lfloor n/2 \rfloor - j}(1 - \rho_n(X))^j.$$ 

3. Suppose that $X = X_1 = \ldots X_{\lfloor n/2 \rfloor}$. The expected value of $Y_{rk=n-2j}$ in $S = \bigcup_{E \in \mathcal{S}_X} \text{Sel}_2(E)$, using the notation of Remark 8.6 and Lemma 8.7, is given by

$$E(Y_{rk=n-2j}|S) = \left( \begin{array}{c} \lfloor n/2 \rfloor \\ j \end{array} \right) \cdot E^n_{m(j),j}(X),$$

where $E^n_{m(j),j}(X)$ can be calculated recursively using the formulae of Lemma 8.7.

Proof. Let $\{Y_i\}$ be the random variables given by Hypothesis B, such that $\text{rank}_{n,X} = \sum Y_i$. It follows that $\text{rank}_{n,X} = n - 2j$ if and only if there are exactly $m(j)$ coordinates of $s = (s_1, \ldots, s_{\lfloor n/2 \rfloor})$ that are a MW symbol, if and only if there are exactly $m(j)$ indices $1 \leq k_1 < \cdots < k_{m(j)} \leq \lfloor n/2 \rfloor$
such that $Y_{k_1} = \cdots = Y_{k_{m(j)}} = 1$ and $Y_i = 0$ for all other indices. If we fix one such $m(j)$-tuple of indices, then this occurs exactly when the random variable

$$Y_1 \cdot Y_2 \cdots Y_{m(j)} \cdot \frac{\prod_{i=1}^{[n/2]} (1 - Y_i)}{(1 - Y_1)(1 - Y_2) \cdots (1 - Y_{k_{m(j)}})}$$

takes the value 1. Finally, adding over all the possible $m(j)$-tuples $(k_1, \ldots, k_{m(j)})$, we obtain the random variable equal to $Y_{rk=n-2}$, as in the statement.

For the second part of the statement, Hypothesis B says that the random variables $Y_i$ are mutually independent (in particular, $\mathbb{E}(Y_i Y_j) = \mathbb{E}(Y_i) \mathbb{E}(Y_j)$ for any $i \neq j$). Then,

$$\mathbb{E}(Y_{rk=n-2j}) \simeq \sum_{1 \leq k_1 < \cdots < k_{m(j)} \leq [n/2]} \mathbb{E}(Y_{k_1}) \cdots \mathbb{E}(Y_{k_{m(j)}}) \cdot \frac{\prod_{i=1}^{[n/2]} (1 - \mathbb{E}(Y_i))}{(1 - \mathbb{E}(Y_{k_1})) \cdots (1 - \mathbb{E}(Y_{k_{m(j)}}))}$$

$$= \binom{[n/2]}{m(j)} \rho_n(X)^m j (1 - \rho_n(X))^j = \binom{[n/2]}{j} \rho_n(X)^m (1 - \rho_n(X))^j,$$

as claimed, where we have used the facts that (a) if $\text{Cov}(Y, Y') = 0$ (or $\simeq 0$), then $\mathbb{E}(Y Y') = \mathbb{E}(Y) \cdot \mathbb{E}(Y')$ and $\mathbb{E}(1 - Y) = 1 - \mathbb{E}(Y)$, and (b) that $Y_i = Y_{\text{Hasse}, n, X}(s_i)$ and therefore $\mathbb{E}(Y_i) = \rho_n(X)$. For (3), we can calculate the expected value $\mathbb{E}(Y_{rk=n-2j}|S)$ as follows:

$$\mathbb{E}(Y_{rk=n-2j}|S) = \mathbb{E} \left( \sum_{1 \leq k_1 < \cdots < k_{m(j)} \leq [n/2]} Y_{k_1} \cdot Y_{k_2} \cdots Y_{k_{m(j)}} \cdot \frac{\prod_{i=1}^{[n/2]} (1 - Y_i)}{(1 - Y_{k_1})(1 - Y_{k_2}) \cdots (1 - Y_{k_{m(j)}})} | S \right)$$

$$= \sum_{1 \leq k_1 < \cdots < k_{m(j)} \leq [n/2]} \mathbb{E} \left( Y_{k_1} \cdot Y_{k_2} \cdots Y_{k_{m(j)}} \cdot \frac{\prod_{i=1}^{[n/2]} (1 - Y_i)}{(1 - Y_{k_1})(1 - Y_{k_2}) \cdots (1 - Y_{k_{m(j)}})} | S \right)$$

$$= \sum_{1 \leq k_1 < \cdots < k_{m(j)} \leq [n/2]} \mathbb{E}^{m(j)} n j (X),$$

where we have used equicorrelation of random variables $Y_i$ in $S$ for the equality of the expected value of the product of any $m(j)$ random variables, and parts (1) and (2) of Lemma 8.5.

Let us simplify the formulas of Corollary 8.8 for $n = 1, \ldots, 5$.

**Corollary 8.9.** For every $r \geq 0$, we define

$$\tilde{R}_r^X = \{ E = (X, n, \text{Sel}_2) \in \tilde{E}^X : \text{rank}(E) = r \}.$$

If we assume $H_B$, then the probabilities

$$p_n(r) = \text{Prob}(E \in \tilde{R}_r^X | E \in \tilde{S}_n^X) = \mathbb{E}(Y_{rk=r}|S)$$

for $n = 1, \ldots, 5$ and $0 \leq r \leq n$ are given by the formulas in Table 8.

**Proof.** If $E = (X, n, \text{Sel}_2)$ is a test elliptic curve with Selmer rank $n = 1$, then $\text{Sel}_2 = ()$ is empty, and $\text{rank}(E) = 0$ by definition. Thus, $p_1(0) = 0$ and $p_1(1) = 1$. When $n = 2$ or 3, then $\text{Sel}_2 = (s_{E,1})$, so there is a unique random variable $Y_1(\text{Sel}_2(E)) = Y_{\text{Hasse}, n, X}(s_{E,1})$, with mean $\rho_n(X)$, such that
The formulas for the expected value of higher moments of the random variables $Y_i$, $i \leq r \leq n$, it follows that $p_n(r) = 0$ and $p_n(2) = 1 - \rho_n(X)$, and $p_n(r) = 0$ for $r \neq n - 2, n$.

Finally, if $n = 4, 5$, then $\text{rank}_{n, X} = Y_1 + Y_2$, and Corollary 8.8 says that

$$Y_{rk=n} = Y_1Y_2, \quad Y_{rk=n-2} = Y_1(1 - Y_2) + (1 - Y_1)Y_2, \quad Y_{rk=n-4} = (1 - Y_1)(1 - Y_2),$$

with expected value, respectively, given by

\[
\begin{align*}
\mathbb{E}(Y_{rk=n}) &= \mathbb{E}(Y_1)\mathbb{E}(Y_2) + \text{Cov}(Y_1, Y_2) = \rho_n(X)^2 + C_{1,1}^n(X), \\
\mathbb{E}(Y_{rk=n-2}) &= \mathbb{E}(Y_1)(1 - \mathbb{E}(Y_2)) - \text{Cov}(Y_1, Y_2) + (1 - \mathbb{E}(Y_1))\mathbb{E}(Y_2) - \mathbb{E}(Y_1, Y_2) \\
&= 2\rho_n(X)(1 - \rho_n(X)) - 2C_{1,1}^n(X), \\
\mathbb{E}(Y_{rk=n-4}) &= (1 - \mathbb{E}(Y_1))(1 - \mathbb{E}(Y_2)) + \text{Cov}(Y_1, Y_2) = (1 - \rho_n(X))^2 + C_{1,1}^n(X),
\end{align*}
\]

where we have used the equality $\mathbb{E}(Y_1Y_2) = \mathbb{E}(Y_1)\mathbb{E}(Y_2) + \text{Cov}(Y_1, Y_2)$ and the fact that the covariance satisfies $\text{Cov}(a + bY_i, c + dY_j) = bd \text{Cov}(Y_i, Y_j)$ for constants $a, b, c, d$.

**Remark 8.10.** The formulas for the expected value of $Y_{rk=n-2j}$ for $n \geq 6$, unfortunately, cannot be written just in terms of $\mathbb{E}(Y_i)$ and $C_{1,1}^n(X) = \text{Cov}(Y_i, Y_j)$ for $i \neq j$. One needs to know other higher moments of the random variables $Y_i|S$. For instance, let $n = 6$. Then,

\[
\begin{align*}
\mathbb{E}(Y_{rk=6}) &= \mathbb{E}(Y_1Y_2Y_3) = \mathbb{E}(Y_1Y_2)\mathbb{E}(Y_3) + \text{Cov}(Y_1Y_2, Y_3) \\
&= \mathbb{E}(Y_1)\mathbb{E}(Y_2)\mathbb{E}(Y_3) + \text{Cov}(Y_1, Y_2)\mathbb{E}(Y_3) + \text{Cov}(Y_1Y_2, Y_3) \\
&= \rho_6(X)^3 + C_{1,1}^6(X)\rho_6(X) + C_{2,1}^6(X).
\end{align*}
\]

The formulae for $\mathbb{E}(Y_{rk=6-2j})$ can be written in terms of the functions $\rho_6(X)$, $C_{1,1}^6$, and $C_{2,1}^6$. For example,

\[
\begin{align*}
\mathbb{E}(Y_{rk=4}) &= \mathbb{E}(Y_1Y_2(1 - Y_3) + \mathbb{E}(Y_1(1 - Y_2)Y_3) + \mathbb{E}((1 - Y_1)Y_2Y_3) = 3 \cdot \mathbb{E}(Y_1Y_2(1 - Y_3)) \\
&= 3(\mathbb{E}(Y_1)\mathbb{E}(Y_2) + \text{Cov}(Y_1Y_2, 1 - Y_3)) \\
&= 3(\mathbb{E}(Y_1Y_2)(1 - \mathbb{E}(Y_3)) - \text{Cov}(Y_1Y_2, Y_3)) \\
&= 3(\mathbb{E}(Y_1)\mathbb{E}(Y_2) + \text{Cov}(Y_1, Y_2))(1 - \mathbb{E}(Y_3) - \text{Cov}(Y_1Y_2, Y_3)) \\
&= 3(\rho_6(X)^2(1 - \rho_6(X)) + C_{1,1}^6(X)(1 - \rho_6(X)) - C_{2,1}^6(X)).
\end{align*}
\]

### Table 8.
Values of $p_n(r) = \text{Prob}(E \in \tilde{R}_r^X \mid E \in \tilde{S}_n^X)$ for $n = 2, \ldots, 5$ and $0 \leq r \leq n$.

| $p_n(r)$ | 2 | 3 | 4 | 5 |
|----------|---|---|---|---|
| $r = 0$  | $1 - \rho_2(X)$ | 0  | $(1 - \rho_4(X))^2 + C_{1,1}^4(X)$ | 0  |
| 1        | 0  | $1 - \rho_3(X)$ | 0  | $(1 - \rho_5(X))^2 + C_{1,1}^5(X)$ |
| 2        | $\rho_2(X)$ | 0  | $2\rho_4(X)(1 - \rho_4(X)) - 2C_{1,1}^4(X)$ | 0  |
| 3        | $\rho_3(X)$ | 0  | $2\rho_5(X)(1 - \rho_5(X)) - 2C_{1,1}^5(X)$ | 0  |
| 4        | 0  | $\rho_4(X)^2 + C_{1,1}^4(X)$ | 0  | $\rho_5(X)^2 + C_{1,1}^5(X)$ |

Note that $p_1(0) = 0$ and $p_1(1) = 1$. 

\[\text{rank}_{1, X} = Y_1.\]
8.1. **Testing Hypothesis B.** As we did for Hypothesis A, in order to test Hypothesis B, we shall use the sequence \((E^X)_{X \geq 1}\) of ordinary elliptic curves as a representative of \(T\) (see Remarks 5.2 and 5.5). In order to estimate the values of \(\rho_n(X)\), we define the following moving ratio measuring the failure of the Hasse principle for 2-Selmer elements coming from elliptic curves of Selmer rank \(n\) and up to height \(X\).

**Definition 8.11.** Let \(\tilde{T} \in T\) be an arbitrary sequence. For each \(n \geq 2\), and \(X \geq 0\), we define the average failure of the Hasse principle for test Selmer elements in the height interval \((X, X + N]\) by

\[
\rho_n(X, N) = \frac{\sum_{E \in \tilde{T}S_n((X, X + N])} (\text{rank}(E) - (n \text{ mod } 2))}{\sum_{E \in \tilde{T}S_n((X, X + N])} (\text{selrank}(E) - (n \text{ mod } 2))}
\]

\[
= \frac{\sum_{E \in \tilde{T}S_n((X, X + N])} (\text{rank}(E) - (n \text{ mod } 2))}{(n - (n \text{ mod } 2)) \cdot \pi_{\tilde{T}S_n((X, X + N])}}.
\]

**Corollary 8.12.** Let \(\tilde{T} \in T\) be an arbitrary sequence. If we assume \(H_A\) and \(H_B\), and \(X > N^2 \geq 0\), then the expected value of \(\rho_n(X, N)\) is given by \(\rho_n(X) + O(X^{-1/3})\) on average, with a standard error

\[
\approx \sqrt{\frac{6X^{1/6}[n/2](\rho_n(X)(1 - \rho_n(X)) + ([n/2] - 1)C_{1,1}^n(X))}{5\kappa N \theta_n(X)} + O\left(\frac{1}{N^{X^{1/6}}}\right)}
\]

\[
\approx \sqrt{\frac{6X^{1/6}(1 + C_n X^{-\epsilon_n})[n/2](\rho_n(X)(1 - \rho_n(X)) + ([n/2] - 1)C_{1,1}^n(X))}{5\kappa sN}} + O\left(\frac{1}{N^{X^{1/6}}}\right)
\]

where \(C_{1,1}^n(X) = 0\) for \(n = 2, 3\), and the last approximation assumes Hypothesis 6.3.

**Proof.** By Corollary 8.4, the expected value of

\[
\sum_{E \in \tilde{T}S_n((X, X + N])} \frac{\text{rank}(E) - (n \text{ mod } 2)}{2}
\]

is given by

\[
\sum_{E \in \tilde{T}S_n((X, X + N])} [n/2] \rho_n(\text{ht}(E)) = [n/2] \sum_{H = X + 1}^{X + N} \sum_{E \in \tilde{T}S_n^H} \rho_n(H) = [n/2] \sum_{H = X + 1}^{X + N} \pi_{\tilde{T}S_n}([H, H]) \cdot \rho_n(H).
\]

By Proposition 6.7, the expected value of \(\pi_{\tilde{T}S_n}([H, H])\) \(\approx \int_{H - 1}^{H} \frac{5\kappa n (T)}{6T^{1/6}} dT + O\left(\frac{1}{H^{1/2}}\right)\) on average, and since the limit of \(\rho_n(H) = 0\) by \(H_B\), then Lemma 3.6 says that

\[
[n/2] \sum_{H = X + 1}^{X + N} \pi_{\tilde{T}S_n}([H, H]) \cdot \rho_n(H) \approx \frac{5\kappa [n/2]}{6} \int_{X}^{X + N} \frac{\theta_n(T) \rho_n(T)}{T^{1/6}} dT + O\left(\frac{N \rho_n(H)}{H^{1/2}}\right)
\]

on average, which in turn (as in Corollary 3.4) says that, for \(X > N^2 \geq 0\), we have

\[
\text{E}\left(\sum_{E \in \tilde{T}S_n((X, X + N])} \frac{\text{rank}(E) - (n \text{ mod } 2)}{2}\right) \approx \frac{5\kappa [n/2]N \theta_n(X) \rho_n(X)}{6X^{1/6}} + O\left(\frac{N \rho_n(X)}{X^{1/2}}\right).
\]
By Proposition 6.7, we have that \( E(\pi_{TS_n}((X, X + N))) \approx 5\kappa N/(6X^{1/6}) + O(N/X^{1/2}) \), and therefore, the expected value of

\[
\rho_n(X, N) = \frac{2}{(n - (n \mod 2))\pi_{TS_n}((X, X + N))} \sum_{E \in TS_n((X, X + N))} \text{rank}(E) - (n \mod 2)
\]

is given by

\[
\rho_n(X) + O\left(\frac{X^{1/6}\rho_n(X)}{X^{1/2}}\right) = \rho_n(X) + O\left(\frac{\rho_n(X)}{X^{1/3}}\right),
\]
on average, where we have used the simple fact that \( 2[n/2] = n - (n \mod 2) \). Since \( \lim_{X \to \infty} \rho_n(X) = 0 \), we can simplify the error term to \( O(X^{-1/3}) \).

The standard error can be deduced from the formula of Corollary 8.4 for \( \pi_{TS_n}((X, X + N)) \) curves, the number (on average) of curves of each Selmer rank from Prop. 6.7, and Lemma 3.6, and it is given on average by

\[
\begin{aligned}
&= \frac{1}{\pi_{TS_n}((X, X + N))} \sqrt{\frac{X + N}{H_{X+1}}} \sum_{H=X+1}^{[n/2]} \pi_{TS_n}([H, H]) \cdot (\rho_n(H)(1 - \rho_n(H)) + ([n/2] - 1)C_{1,1}^n(H)) \\
&\approx \frac{1}{\pi_{TS_n}((X, X + N))} \sqrt{\frac{6X^{1/6}[n/2]}{5\kappa N\theta_n(X)}} \cdot (\rho_n(X)(1 - \rho_n(X)) + ([n/2] - 1)C_{1,1}^n(X)) + O\left(\frac{1}{NX^{1/3}}\right) \\
&\approx \sqrt{\frac{6X^{1/6}[n/2]}{5\kappa N\theta_n(X)}} \cdot (\rho_n(X)(1 - \rho_n(X)) + ([n/2] - 1)C_{1,1}^n(X)) + O\left(\frac{1}{NX^{1/6}}\right) \\
&\approx \sqrt{\frac{6X^{1/6}(1 + C_nX^{-\kappa_n})[n/2]}{5\kappa_nN}} \cdot (\rho_n(X)(1 - \rho_n(X)) + ([n/2] - 1)C_{1,1}^n(X)) + O\left(\frac{1}{NX^{1/6}}\right)
\end{aligned}
\]
as claimed, where in the approximations we assumed \( H_A \) and we have used the results of Proposition 6.7, part (2).

We have used the BHKSSW data to estimate probability function \( \rho_n(X) \) using the moving ratios \( \rho_n(X, N) \) of Corollary 8.12. We have plotted values of \( \rho_n(X, 0.25 \cdot 10^9) \) for \( n = 2, \ldots, 5 \) using the BHKSSW database, and the graphs can be found in Figure 9.

In Table 9 we record the last values of \( \rho_n(X, 0.25 \cdot 10^9) \) that appear in the graphs (which correspond to \( X \approx 2.675 \cdot 10^{10} \)). We also record the values of \( \pi_{S_n} \) in \([2.675 \cdot 10^{10}, 2.7 \cdot 10^{10}]\). The total number of elliptic curves in the same interval is 1,828,235.

Finally, we have found (using SageMath) best-fit models for the data of \( \rho_n(X, N) \) of the form

\[
\rho_n(X, N) \approx \frac{D_n}{X^{f_n}}.
\]
and we provide the values of \( D_n \) and \( f_n \) in Table 10. We have compared the models with the data in Figure 10.
Figure 9. Graphs of the moving ratios $\rho_n(X, 0.25 \cdot 10^9)$ for $n = 2$ (green), 3 (red), 4 (gray stars), 5 (purple).

Table 9. The number of curves of Selmer rank $2 \leq n \leq 5$, and the values of $\rho_n(X, N)$ in the interval $[2.675 \cdot 10^{10}, 2.7 \cdot 10^{10}]$.

| $n$ | 2       | 3       | 4       | 5       |
|-----|---------|---------|---------|---------|
| $\pi_{S_n}([2.675 \cdot 10^{10}, 2.7 \cdot 10^{10}])$ | 476,579 | 104,922 | 7945    | 152     |
| $\rho_n(2.675 \cdot 10^{10}, 0.25 \cdot 10^9)$ | 0.63989181 | 0.45496654 | 0.63857772 | 0.63486842 |

Table 10. The coefficients of the best-fit models $\rho_n(X, N) \approx D_n/X^{f_n}$.

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

Hypothesis 8.13 (Hypothesis $H_B'$). Hypothesis $H_B$ holds and, for every $n \geq 2$, there are constants $D_n$ and $f_n$ such that $\rho_n(X) \approx \frac{D_n}{X^{f_n}}$. Moreover, for $n = 2, \ldots, 5$ the values of $D_n$ and $f_n$ are approximately as given in Table 10.
Figure 10. Graphs of the moving ratios $\rho_n(X, 0.025\cdot 10^9)$ for $n = 2$ (green), 3 (red), 4 (gray stars), 5 (purple), and the corresponding models of the form $D_n/XF_n$ (in blue for $n = 2, 3$ and red for $n = 4, 5$).

Remark 8.14. Before we can discuss the error in the approximation $\rho_n(X, N) \approx \rho_n(X)$ we need to estimate the covariance functions $C_{n,s,t}(X)$. This can be done via the formulas for the expected value of $Y_{rk=n-2j}$ given by Corollary 8.8 and, for $n = 1, 2, 3, 4, 5$, the simplified formulas given by Corollary 8.9. The first thing to note is that for $n = 1, 2, 3$, we have $C_{n,s,t}(X) = 0$ for all possible values of $s, t$ since there is either none ($n = 1$) or only one random variable $Y_1$ that intervenes ($n = 2, 3$). For $n = 4$ and 5 there are two random variables $Y_1$ and $Y_2$ and

$$C_{1,1}^n(X) = \mathbb{E}(Y_{rk=n}) - \mathbb{E}(Y_1)\mathbb{E}(Y_2) = \mathbb{E}(Y_{rk=n}) - \rho_n(X)^2.$$  

In Figures 11 and 12 we have plotted approximate covariance values of $C_{1,1}^4(X)$ and $C_{1,1}^5(X)$, respectively, using sample height intervals $[X, X + 0.25\cdot 10^9]$, together with the best linear fits for the data which are given by

$$-0.02677186 + (4.06113344\cdot 10^{-14})x \quad \text{and} \quad -0.01328180 + (1.28980002\cdot 10^{-12})x,$$

respectively. In particular, we observe that $|C_{1,1}^4(X) - (-0.025)| \lesssim 0.015$ and $|C_{1,1}^5(X) - 0| \lesssim 0.1$. Thus, below, we will approximate $C_{1,1}^4(X) \approx -0.025$ and $C_{1,1}^5(X) \approx 0$.

Remark 8.15. Let us assume Hypothesis 8.13, and let us use Corollary 8.12 to estimate the error in the approximation $\rho_n(X) \approx \rho_n(X, N)$. The error should be given by the expression

$$\text{err}_{1,n}(X, N) = \sqrt{\frac{|n/2| \rho_n(X)(1 - \rho_n(X)) + (|n/2| - 1)C_{1,1}^n(X)}{\pi S_n((X, X + N))}}$$
Figure 11. Approximate values of $C_{1,1}^{4}(X)$ using sample height intervals $[X, X + 0.25 \cdot 10^9]$ to estimate $E(Y_{i,k=4}) - \rho_4(X)^2$ using the family of elliptic curves. The best-fit line is given by $-0.02677186 + (4.06113344 \cdot 10^{-14})x$.

or by the expression

$$
\text{err}_{2,n}(X, N) = \sqrt{6X^{1/6}(1 + C_n X^{-\varepsilon_n})[n/2](\rho_n(X)(1 - \rho_n(X)) + ([n/2] - 1)C_{1,1}^{n}(X))}
$$

if we assume $H_A$ and Hypothesis 6.3 also. Using our calculations of Remark 8.14, we will take $C_{1,1}^{n}(X) = 0$ for $n = 2, 3$, and $C_{1,1}^{4}(X) = -0.025$, and $C_{1,1}^{5}(X) = 0$. In Table 11 we include the values of: $\rho_n(X, N)$, our model of $\rho_n(X)$, the error of the model $|\rho_n(X, N) - \rho_n(X)|$, and the predicted standard errors $\text{err}_{n,n}(X, N)$, for $i = 1, 2$, and $X = 2.675 \cdot 10^{10}$, with $N = 0.25 \cdot 10^9$.

**Remark 8.16.** As we can see from the errors in Table 11, we seem to have insufficient data for $n = 5$, so our models of $\rho_5(X)$ are not as accurate as we would wish.

**Remark 8.17.** As we have mentioned earlier in Remark 6.5 the BHKSSW database ([1]) also includes small databases of random samples of elliptic curves at larger heights. In order to test $H_B$ and Hypothesis 8.13, we have calculated the average Hasse ratio for the curves in $E_k$ (with notation as in Remark 6.5), and have plotted the ratios together with our models for $\rho_n(X)$, in Figure 13 (note: the $x$-axis is in logarithmic scale). We have also computed the predicted errors (a calculation similar to that carried out in Table 11) and the predictions seem to match the data in large heights, as well.

**Remark 8.18.** It would be interesting to compute the ratio $\rho_n(X)$ in families of quadratic twists. However, such families are very “thin” in the family of all elliptic curves, and the convergence of the Hasse ratios to $\rho_n(X)$ would be unreliable. In order to provide some data in this direction,
Figure 12. Approximate values of $C_{1,1}^5(X)$ using sample height intervals $[X, X + 0.25 \cdot 10^9]$ to estimate $E(Y_{ik=5}) - \rho_5(X)^2$ using the family of elliptic curves. The best-fit line is given by $-0.01328180 + (1.28980002 \cdot 10^{-12})x$.

Table 11. Values of: $\rho_n(X, N)$, our model of $\rho_n(X)$, the error $|\rho_n(X, N) - \rho_n(X)|$, and the two predicted standard errors $err_{i,n}(X, N)$, for $i = 1, 2$, and $X = 2.675 \cdot 10^{10}$, $N = 0.25 \cdot 10^9$.

we have calculated the Selmer rank and Mordell–Weil rank in a family of twists (quadratic and quartic) of $y^2 = x^3 + x$. More precisely, we consider the curves $E_A : y^2 = x^3 + Ax$, with fourth-power-free $1 \leq A \leq 10^6$ (curves up to height $4 \cdot 10^{18}$). Then, we have calculated the moving ratios $\rho_n$ in slices of 10,000 curves, and graphed them against the models of Hypothesis 8.13. See Figure 14. Note, however, that we do not expect the exact same behavior in this family, since
Figure 13. Graphs of the moving ratios $\rho_n(X,N)$ for the curves of large height in the database BHKSSW, 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 $n = 4, 5$). The $x$-axis is in logarithmic scale.

$j(E_A/Q) = 1728$ is fixed, and therefore it is a family of twists (quadratic and quartic). It is likely that if $H_B$ holds, then a similar condition is true for $j = 1728$ up to a constant. That is, we may have $\rho_{n,1728}(X) \approx C_{1728} \cdot \rho_n(X)$, where $C_{1728}$ is a fixed constant. At any rate, the family of curves with $j = 1728$ is very sparse within the family of all curves, and the data only indicates some consistency with our expectations.

Example 8.19. Theorem 8.3, assuming $H_B$, provides the expected value and variance for the rank of an elliptic curve $E/Q$ of Selmer rank $n$ and height $X$. More precisely, in Corollary 8.8 and 8.9, we give formulas for the probabilities for each rank. Now that we have models for $\rho_n(X)$ and $C_{n,1}(X)$ (as in Remark 8.14), we can look at the distribution of ranks in intervals. Let us consider, for instance, the curves $E(I)$ in the height interval $I = [20 \cdot 10^9, 20.25 \cdot 10^9]$ in the BHKSSW database. For each $n = 2, 3, 4, 5$ we have created histograms using the number of curves of Selmer rank $n$ and Mordell–Weil rank $0 \leq n$ (in blue bars), and also created histograms with the number of M–W ranks that we would expect from Corollary 8.9 (in green bars). The resulting histograms can be found in Figure 15 (together with the graph of the normal distribution that would approximate the binomial $B([n/2], \rho_n(X))$. We have also included the raw data of ranks observed and ranks predicted in Table 12.
Graphs of the moving ratios $\rho_n$ in the family $y^2 = x^3 + Ax$ for $n = 2$ (green), 3 (red), 4 (gray stars), 5 (purple), compared to the models of the form $D_n/X^f_n$ (in blue for $n = 2, 3$ and red for $n = 4, 5$).

Table 12. Mordell–Weil ranks observed in the interval height interval $[2 \cdot 10^{10}, 2.025 \cdot 10^{10}]$ and the ranks predicted by the distribution of Theorem 8.3.

| $n$ | $\pi_{S_n}([2 \cdot 10^{10}, 2.025 \cdot 10^{10}])$ | M–W ranks observed in $S_n$ | 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] |

9. Predicting the number of curves with a given rank up to height $X$

Let $X, r \geq 0$ be fixed. We denote the set of elliptic curves of height \( \leq X \) and Mordell–Weil rank $r$ by

\[
\mathcal{R}_r(X) = \{ E \in \mathcal{E}(X) : \text{rank}(E(\mathbb{Q})) = r \},
\]

and we write $\pi_{\mathcal{R}_r}(X) = \# \mathcal{R}_r(X)$. We refer the reader to Sections 3.3 and 3.4 of [23] for a summary of conjectures about $\pi_{\mathcal{R}_r}(X)$, but we point out two in particular:
Figure 15. Histogram (in blue) of the distribution of Mordell–Weil ranks among elliptic curves in $E([2 \cdot 10^{10}, 2.025 \cdot 10^{10}])$ by Selmer rank $n = 2, 3, 4, 5$, and compared to the histogram (in green) of the M–W ranks that we would expect from Theorem 8.3. The graph is that of the normal distribution that best approximates the binomial.

- Watkins ([29]; see also [2] for an expository paper) has conjectured that there is a constant $c$ such that

$$\sum_{k=1}^{\infty} \pi_{R_{2k}}(X) = (c + o(1))X^{10/24}(\log X)^{3/8}.$$
• Park, Poonen, Voight, and Wood ([23]) have developed a heuristic that predicts:
  (1) All but finitely many elliptic curves satisfy \( \text{rank}(E(\mathbb{Q})) \leq 21 \).
  (2) For \( 1 \leq r \leq 20 \), we have \( \sum_{k=r}^{\infty} \pi_{\mathcal{R}}(X) = X^{(21-r)/24+o(1)} \).
  (3) \( \sum_{k=21}^{\infty} \pi_{\mathcal{R}}(X) \leq X^{o(1)} \).

In this section, we denote the set of test elliptic curves of height \( \leq X \) and rank \( r \) by
\[
\tilde{\mathcal{R}}_r(X) = \{ E \in \tilde{\mathcal{E}}(X) : \text{rank}(E) = r \},
\]
and if \( \tilde{T} \in \mathbf{T} \), we write \( \tilde{T}\tilde{\mathcal{R}}_r(X) \) for the subset of test elliptic curves in \( \tilde{T} \) of rank \( r \) and height \( \leq X \). Then, we write \( \pi_{\tilde{T}\tilde{\mathcal{R}}_r}(X) = \# \tilde{T}\tilde{\mathcal{R}}_r(X) \). We shall assume hypotheses \( H_A \) and \( H_B \), and derive the expected value of \( \pi_{\tilde{T}\tilde{\mathcal{R}}_r}(X) \) that follows from the probability distributions we have studied in previous sections. We shall study \( \pi_{\tilde{T}\tilde{\mathcal{R}}_r}(X) \) as the sum of the contributions of rank \( r \) coming from each Selmer rank \( n = r + 2j \). That is, we shall approximate \( \pi_{\tilde{T}\tilde{\mathcal{R}}_r}(X) \) by approximating each term in the infinite sum
\[
\pi_{\tilde{T}\tilde{\mathcal{R}}_r}(X) = \sum_{j=0}^{\infty} \pi_{\tilde{T}\tilde{\mathcal{R}}_r \cap \tilde{S}_{n(j)}(X)}.
\]
Thus, for fixed \( r \geq 0 \), we first give the expected value of \( \pi_{\tilde{T}\tilde{\mathcal{R}}_r \cap \tilde{S}_{n(j)}}(X) \) for each \( j \geq 0 \).

**Theorem 9.1.** Let \( X, r \geq 0, j \geq 0 \) be fixed, such that \( n(j) = r + 2j \geq 2 \). Let \( \tilde{T} \in \mathbf{T} \) be arbitrary. If we assume \( H_A \) and \( H_B \), then the expected value of \( \pi_{\tilde{T}\tilde{\mathcal{R}}_r \cap \tilde{S}_{n(j)}}(X) \) is given by
\[
\mathbb{E} \left( \pi_{\tilde{T}\tilde{\mathcal{R}}_r \cap \tilde{S}_{n(j)}}(X) \right) = \frac{5\kappa}{6} \left( \left\lceil \frac{r}{2} \right\rceil + j \right) \int_{0}^{X} \frac{\theta_{n(j)}(H)}{H^{1/6}} \cdot \mathbb{E}_{\left\lceil \frac{r}{2} \right\rceil,j}(H) dH + \theta_{n(j)}(X) \cdot O \left( X^{1/2} \right),
\]
where \( \mathbb{E}_{\left\lceil \frac{r}{2} \right\rceil,j}(H) \) is the expected value defined in Remark 8.6. Further, if we assume Hypothesis 6.3, then
\[
\pi_{\tilde{T}\tilde{\mathcal{R}}_r \cap \tilde{S}_{n(j)}}(X) \approx \frac{5\kappa}{6} \left( \left\lceil \frac{r}{2} \right\rceil + j \right) \int_{0}^{X} \frac{s_{n(j)} \cdot \mathbb{E}_{\left\lceil \frac{r}{2} \right\rceil,j}(H)}{(1 + C_{n(j)} H^{-e_{n(j)}}) \cdot H^{1/6}} dH + \theta_{n(j)}(X) \cdot O \left( X^{1/2} \right).
\]

**Proof.** Let us write \( n(j) = r + 2j \). Thus, \( \left\lceil \frac{n(j)}{2} \right\rceil = \left\lceil \frac{r}{2} \right\rceil + j \). We compute the expected value of \( \pi_{\tilde{T}\tilde{\mathcal{R}}_r}(X) \) as follows:
\[
\mathbb{E} \left( \pi_{\tilde{T}\tilde{\mathcal{R}}_r \cap \tilde{S}_{n(j)}}(X) \right) = \mathbb{E} \left( \# \{ E \in \tilde{T}\tilde{S}_{n(j)}(X) : \text{rank}(E) = r \} \right)
= \sum_{T=1}^{X} \mathbb{E} \left( \# \{ E \in \tilde{T}\tilde{S}_{n(j)}([T,T]) : \text{rank}(E) = r \} \right)
= \sum_{T=1}^{X} \mathbb{E} \left( \pi_{\tilde{T}\tilde{S}_{n(j)}([T,T])} \right) \cdot \text{Prob} \left( \text{rank}(E) = r \mid E \in \tilde{S}_{n(j)}([X,X]) \right),
\]
by the basic properties of the expected value. If we assume \( H_A \) and \( H_B \) and use Corollary 6.2 for the value of \( \pi_{\tilde{T}\tilde{S}_{n(j)}([H,H])} \) (on average) and Corollary 8.8 for the probability of rank \( r \) in \( \tilde{S}_{n(j)}([X,X]) \),
we obtain the following formula:

$$
\mathbb{E}\left(\pi_{\mathcal{F}_r \cap \mathcal{S}_{n(j)}}(X)\right) = \sum_{T=0}^{X-1} \left(\frac{5\kappa}{6} \int_{T}^{T+1} \frac{\theta_{n(j)}(H)}{H^{1/6}} \cdot \left[\frac{n(j)}{2}\right] \cdot \mathbb{E}_{[\frac{T}{2}]}^{n(j)}(H) \, dH + \theta_{n(j)}(T) \cdot O\left(\frac{1}{T^{1/2}}\right)\right)
$$

$$
= \frac{5\kappa}{6} \left[\frac{n(j)}{2}\right] \int_{0}^{X} \frac{\theta_{n(j)}(H)}{H^{1/6}} \cdot \mathbb{E}_{[\frac{T}{2}]}^{n(j)}(H) \, dH + \theta_{n(j)}(X) \cdot O\left(X^{1/2}\right).
$$

If we further assume Hypothesis 6.3, then

$$
\pi_{\mathcal{F}_r \cap \mathcal{S}_{n(j)}}(X) \approx \frac{5\kappa}{6} \left[\frac{n(j)}{2}\right] \int_{0}^{X} \frac{s_{n(j)} \cdot \mathbb{E}_{[\frac{T}{2}]}^{n(j)}(H)}{1 + C_{n(j)}H^{-\varepsilon_{n(j)}}} \cdot \mathbb{E}_{[\frac{T}{2}]}^{n(j)}(H) \, dH + \left(\sum_{j=0}^{\infty} \theta_{r+2j}(X)\right) \cdot O\left(X^{1/2}\right),
$$

as claimed.

\[ \Box \]

If we now use the formula

$$
\pi_{\mathcal{F}_r}(X) = \sum_{j=0}^{\infty} \pi_{\mathcal{F}_r \cap \mathcal{S}_{r+2j}}(X)
$$

and the fact that \(\sum_{n=0}^{\infty} \theta_n(X) = 1\) (from Corollary 6.2), we obtain the following result.

**Corollary 9.2.** Let \(X, r \geq 0\) be fixed, and let \(\mathcal{F} \subseteq \mathcal{T}\) be arbitrary. If we assume \(H_A\) and \(H_B\), then the expected value of \(\pi_{\mathcal{F}_r}(X)\) is given by the formula

$$
\mathbb{E}\left(\pi_{\mathcal{F}_r}(X)\right) = \frac{5\kappa}{6} \sum_{j=0}^{\infty} \left[\frac{n(j)}{2}\right] \int_{0}^{X} \frac{s_{n(j)} \cdot \mathbb{E}_{[\frac{T}{2}]}^{n(j)}(H)}{1 + C_{n(j)}H^{-\varepsilon_{n(j)}}} \cdot \mathbb{E}_{[\frac{T}{2}]}^{n(j)}(H) \, dH + \left(\sum_{j=0}^{\infty} \theta_{r+2j}(X)\right) \cdot O\left(X^{1/2}\right),
$$

where the error term satisfies \(\left(\sum_{j=0}^{\infty} \theta_{r+2j}(X)\right) \cdot O\left(X^{1/2}\right) = O(X^{1/2}),\) and \(\mathbb{E}_{[\frac{T}{2}]}^{n(j)}(H)\) is the expected value defined in Remark 8.6.

**Remark 9.3.** If we assume \(H_A, H_B,\) and Hypotheses 6.3 and 8.13, and in addition (for the sake of simplicity) we assume that the random variables \(Y_1, \ldots, Y_{[n(j)/2]}\) are independent in \(S = \bigcup \text{Sel}_2(E),\) then we would have

$$
\pi_{\mathcal{F}_r}(X) \approx \frac{5\kappa}{6} \sum_{j=0}^{\infty} \left[\frac{n(j)}{2}\right] \int_{0}^{X} \frac{s_{n(j)} \cdot (D_{n(j)})^{\frac{n(j)}{2}} \cdot (H^{n(j)} - D_{n(j)})^{\frac{n(j)}{2}}}{1 + C_{n(j)}H^{-\varepsilon_{n(j)}}} \cdot \mathbb{E}_{[\frac{T}{2}]}^{n(j)}(H) \, dH.
$$

If we simplify this expression further by just retaining the highest order term (and for now assume \(r \geq 2\)), we obtain:

$$
\pi_{\mathcal{F}_r}(X) \approx \frac{5\kappa}{6} \sum_{j=0}^{\infty} \left[\frac{n(j)}{2}\right] \cdot s_{n(j)} \cdot (D_{n(j)})^{\frac{n(j)}{2}} \int_{0}^{X} \frac{1}{H^{1/6 + [\frac{T}{2}] \cdot f_{n(j)}}} \, dH
$$

$$
\approx \frac{5\kappa}{6} \sum_{j=0}^{\infty} \left[\frac{n(j)}{2}\right] \cdot s_{n(j)} \cdot (D_{n(j)})^{\frac{n(j)}{2}} \cdot \frac{X^{5/6 - [\frac{n(j)}{2}] \cdot f_{n(j)}}}{5/6 - [\frac{n(j)}{2}] \cdot f_{n(j)}}.
$$

In particular, if there is \(j \geq 0\) such that \([\frac{n(j)}{2}] \cdot f_{n(j)} < 5/6\), then there are infinitely many (test) elliptic curves with rank \(r\) (and Selmer rank \(n(j))\).
In our next result, we use Theorem 9.1 to write formulas for the contribution in rank \( r = 1, \ldots, 5 \) coming from Selmer ranks \( n = 1, \ldots, 5 \).

**Corollary 9.4.** If we assume \( H_A, H_B, \) and Hypotheses 6.3 and 8.13, then the formulas in Corollary 8.9 imply approximations of \( \pi_{T_{\mathcal{R}_r \cap \mathcal{S}_n}}(X) \) as given in Table 13, for \( 1 \leq r \leq n \leq 5 \) and \( r \equiv n \mod 2 \).

**Remark 9.5.** Using the formulas given by Corollary 9.4 and Table 14, we can give approximations of \( \pi_{\mathcal{R}_r}(X) \). For instance,

\[
\begin{align*}
\pi_{\mathcal{R}_1}(X) &\approx \pi_{\mathcal{R}_1 \cap \mathcal{S}_1}(X) + \pi_{\mathcal{R}_1 \cap \mathcal{S}_3}(X) + \pi_{\mathcal{R}_1 \cap \mathcal{S}_5}(X), \\
\pi_{\mathcal{R}_2}(X) &\approx \pi_{\mathcal{R}_2 \cap \mathcal{S}_2}(X) + \pi_{\mathcal{R}_2 \cap \mathcal{S}_4}(X), \\
\pi_{\mathcal{R}_3}(X) &\approx \pi_{\mathcal{R}_3 \cap \mathcal{S}_3}(X) + \pi_{\mathcal{R}_3 \cap \mathcal{S}_5}(X), \\
\pi_{\mathcal{R}_4}(X) &\approx \pi_{\mathcal{R}_4 \cap \mathcal{S}_4}(X), \\
\pi_{\mathcal{R}_5}(X) &\approx \pi_{\mathcal{R}_5 \cap \mathcal{S}_5}(X).
\end{align*}
\]

We have used SageMath to numerically integrate and compute said approximations, and we have graphed the results in Figures 16 (for \( r = 1, 2, 3 \)) and 17 (for \( r = 4, 5 \)). In Table 14 we have included the values of \( \pi_{\mathcal{R}_r}(2.7 \cdot 10^{10}) \) according to the data, the values of our approximation, the error, and the relative error (as a percentage of the actual value), and also \( s_r \cdot (2.7 \cdot 10^{10})^{1/2} \), which is, approximately, the size of the error as expected from Corollary 9.2.

![Figure 16. Values of \( \pi_{\mathcal{R}_r}(X) \) from the BHKSSW database (blue dots) for \( r = 1, 2, 3 \), and the approximations given in Remark 9.5 (in red).](image-url)
Corollary 9.6. Let $\tilde{T} \in T$ be arbitrary. If we assume $H_A$, $H_B$, and Hypotheses 6.3 and 8.13, then there are explicit computable positive constants $\lambda_r$ and $h_r$, for $n = 1, 2, 3$ such that

$$
\mathbb{E}(\pi_{\tilde{T}R_1 \cap \tilde{S}_1}(X)) = \lambda_1 + \kappa s_1 X^{5/6} \cdot \sum_{m=0}^{\infty} \frac{(-C_1)^m}{1 - (6/5) \cdot me_1} X^{-me_1} + O(X^{1/2}),
$$

$$
\mathbb{E}(\pi_{\tilde{T}R_2 \cap \tilde{S}_2}(X)) = \lambda_2 + \kappa s_2 D_2 X^{5/6 - f_2} \cdot \sum_{m=0}^{\infty} \frac{(-C_2)^m}{1 - (6/5) \cdot (f_2 + me_2)} X^{-me_2} + O(X^{1/2}),
$$

$$
\mathbb{E}(\pi_{\tilde{T}R_3 \cap \tilde{S}_3}(X)) = \lambda_3 + \kappa s_3 D_3 X^{5/6 - f_3} \cdot \sum_{m=0}^{\infty} \frac{(-C_3)^m}{1 - (6/5) \cdot (f_3 + me_3)} X^{-me_3} + O(X^{1/2}),
$$

for any $X \geq h_r$.

Proof. For each $r = 1, 2, 3$, let $h_r > 0$ be the smallest natural number such that $|C_r h_r^{-e_r}| < 1$. Then,

$$
\pi_{\tilde{T}R_r \cap \tilde{S}_r}(X) = \pi_{\tilde{T}R_r \cap \tilde{S}_r}(h_r) + \pi_{\tilde{T}R_r \cap \tilde{S}_r}([h_r, X])
$$

and, by Corollary 9.2 we have

$$
\mathbb{E}(\pi_{\tilde{T}R_r \cap \tilde{S}_r}([h_0, X])) = \frac{5\kappa}{6} \left( \left\lfloor \frac{7}{2} \right\rfloor \right) \int_{h_0}^{X} \frac{s_r \cdot \mathbb{E}_{[\frac{7}{2}],0}(H)}{1 + C_r H^{-e_r}} \cdot H^{1/6} dH + O(X^{1/2}).
$$

Further, since $|C_r h_r^{-e_r}| < 1$, we can write

$$
\frac{1}{1 + C_r H^{-e_r}} = \sum_{m=0}^{\infty} (-C_r)^m H^{-me_r}.
$$
\[
\begin{aligned}
\pi_{\tilde{\mathcal{R}}_1\cap \tilde{S}_1}(X) & \approx \frac{5\kappa}{6} \int_0^X \frac{\theta_1(H)}{H^{1/6}} \, dH \\
\pi_{\tilde{\mathcal{R}}_1\cap \tilde{S}_3}(X) & \approx \frac{5\kappa}{6} \int_0^X \frac{\theta_3(H) \cdot (1 - \rho_3(H))}{H^{1/6}} \, dH \\
\pi_{\tilde{\mathcal{R}}_1\cap \tilde{S}_5}(X) & \approx \frac{5\kappa}{6} \int_0^X \frac{\theta_5(H) \cdot (1 - \rho_5(H))^2}{H^{1/6}} \, dH \\
\pi_{\tilde{\mathcal{R}}_2\cap \tilde{S}_2}(X) & \approx \frac{5\kappa}{6} \int_0^X \frac{\theta_2(H) \cdot \rho_2(H)}{H^{1/6}} \, dH \\
\pi_{\tilde{\mathcal{R}}_2\cap \tilde{S}_4}(X) & \approx \frac{10\kappa}{6} \int_0^X \frac{\theta_4(H) \cdot (\rho_4(H)(1 - \rho_4(H)) - C_{1,1}^4(X))}{H^{1/6}} \, dH \\
\pi_{\tilde{\mathcal{R}}_3\cap \tilde{S}_3}(X) & \approx \frac{5\kappa}{6} \int_0^X \frac{\theta_3(H) \cdot \rho_3(H)}{H^{1/6}} \, dH \\
\pi_{\tilde{\mathcal{R}}_3\cap \tilde{S}_5}(X) & \approx \frac{10\kappa}{6} \int_0^X \frac{\theta_5(H) \cdot (\rho_5(H)(1 - \rho_5(H)) - C_{1,1}^5(X))}{H^{1/6}} \, dH \\
\pi_{\tilde{\mathcal{R}}_4\cap \tilde{S}_4}(X) & \approx \frac{5\kappa}{6} \int_0^X \frac{\theta_4(H) \cdot (\rho_4(H)^2 + C_{1,1}^4(X))}{H^{1/6}} \, dH \\
\pi_{\tilde{\mathcal{R}}_5\cap \tilde{S}_5}(X) & \approx \frac{5\kappa}{6} \int_0^X \frac{\theta_5(H) \cdot (\rho_5(H)^2 + C_{1,1}^5(X))}{H^{1/6}} \, dH \\
\end{aligned}
\]

Table 13. Approximate values of \(\pi_{\tilde{\mathcal{R}}_r\cap \tilde{S}_n}(X)\) for \(1 \leq r \leq n \leq 5\) and \(r \equiv n \mod 2\).

For any \(H \geq h_r\). Now, \(E_{\tilde{\mathcal{S}}_r}^1(H) = 1\) for \(r = 1\), and by Corollary 8.9, we have \(E_{\tilde{\mathcal{S}}_r}^1(H) = \rho_r(X)\) for \(r = 2, 3\). Further, assuming \(H_B\) we have \(\rho_n(X) = D_n/X^{f_n}\). Putting everything together we obtain,
| \( r \) | 1     | 2     | 3     | 4     | 5     |
|-------|-------|-------|-------|-------|-------|
| \( \pi_{R_5} (2.7 \cdot 10^{10}) \) | 113128929 | 40949289 | 6259157 | 380519 | 6481 |
| Approximate value | 113133971 | 41005107 | 6273138 | 381272 | 6438 |
| \(|\text{Error}|\) | 5042 | 55818 | 13981 | 753 | 43 |
| Error % | 0.004456 | 0.136310 | 0.223368 | 0.197887 | 0.663477 |

\[
\text{Predicted error} \approx s_r \cdot X^{1/2}
\]

for instance, the following approximation formula for \( E_1 \)

\[
\begin{align*}
\pi_{\tilde{T}_{R_1 \cap S_1}} (h_1) + \pi_{\tilde{T}_{R_1 \cap S_1}} ([h_1, X]) &= \pi_{\tilde{T}_{R_1 \cap S_1}} (h_1) + \frac{5}{6} \int_{h_1}^{X} \frac{\theta_1 (H)}{H^{1/6}} \, dH + O(X^{1/2}) \\
&= \pi_{\tilde{T}_{R_1 \cap S_1}} (h_1) + \frac{5 \kappa s_1}{6} \int_{h_1}^{X} \sum_{m=0}^{\infty} (-C_r)^m H^{-\frac{1}{6} - me_1} \, dH + O(X^{1/2}) \\
&= \pi_{\tilde{T}_{R_1 \cap S_1}} (h_1) - \left( \kappa s_1 h_1^{5/6} \sum_{m=0}^{\infty} \frac{(-C_1)^m}{1 - (6/5) \cdot me_1} h_1^{-me_1} \right) + \kappa s_1 X^{5/6} \sum_{m=0}^{\infty} \frac{(-C_1)^m}{1 - (6/5) \cdot me_1} X^{-me_1} \\
&= \lambda_1 + \kappa s_1 X^{5/6} \sum_{m=0}^{\infty} \frac{(-C_1)^m}{1 - (6/5) \cdot me_1} X^{-me_1} + O(X^{1/2}),
\end{align*}
\]

with \( \lambda_1 = \pi_{\tilde{T}_{R_1 \cap S_1}} (h_0) - \left( \kappa s_1 h_1^{5/6} \sum_{m=0}^{\infty} \frac{(-C_1)^m}{1 - (6/5) \cdot me_1} h_1^{-me_1} \right) \), and we derive formulas for \( r = 2 \) and \( r = 3 \) in a similar manner.

10. Predicting the average rank

In this section we shall estimate the average rank of all elliptic curves of height \( \leq X \):

\[
\text{AvgRank}_E (X) = \frac{\sum_{E \in \mathcal{E}(X)} \text{rank}(E(\mathbb{Q}))}{\pi_\mathcal{E}(X)}
\]

We quote here the average rank conjecture as in [24] (see [12] for Goldfeld’s version for quadratic twists).
**Conjecture 10.1.** 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 $1/2$.

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

$$\text{AvgRank}_{\tilde{T}S_n}(X) = \frac{\sum_{E \in \tilde{T}S_n(X)} \text{rank}(E)}{\pi_{\tilde{T}}(X)}$$

and later we will put them together to estimate the total average rank.

**Theorem 10.2.** Let $\tilde{T} \in T$ be arbitrary. Assume $H_A$ and $H_B$, and let $n \geq 1$ be fixed. Then, the expected value of $\text{AvgRank}_{\tilde{T}S_n}(X)$ is given by

$$\frac{5\kappa}{6\pi_{\tilde{T}}(X)} \cdot \int_{1}^{X} \frac{\theta_n(H)}{H^{1/6}} \left( (n \mod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H) \right) \, dH + \theta_n(X) \cdot O(X^{-1/3}).$$

Moreover, the error in approximating $\text{AvgRank}_{\tilde{T}S_n}(X)$ by its expected value is given by

$$\sqrt{\frac{5\kappa[n/2]}{6\pi_{\tilde{T}}(X)^2}} \int_{0}^{X} \frac{\theta_n(H)}{H^{1/6}} (\rho_n(H)(1 - \rho_n(H)) + (|n/2| - 1)C_{1,1}(H)) \, dH + O(X^{-7/6}).$$

**Proof.** We compute the expected value of the average rank in the sequence $\tilde{T}$ as follows:

$$\mathbb{E}(\text{AvgRank}_{\tilde{T}S_n}(X)) = \mathbb{E} \left( \frac{\sum_{E \in \tilde{T}S_n(X)} \text{rank}(E)}{\pi_{\tilde{T}}(X)} \right) = \frac{1}{\pi_{\til{T}}(X)} \mathbb{E} \left( \sum_{E \in \til{T}S_n(X)} \text{rank}(E) \right)$$

by Corollary 8.4. In particular, Definition 5.4, Corollary 3.4, and $H_A$ imply

$$= \frac{1}{\pi_{\til{T}}(X)} \cdot \sum_{H=1}^{X} \sum_{E \in \til{T}S_n([H,H])} (n \mod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H)$$

$$= \frac{1}{\pi_{\til{T}}(X)} \cdot \sum_{H=1}^{X} \pi_{\til{T}S_n}([H,H]) \cdot \left( (n \mod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H) \right)$$

$$= \frac{1}{\pi_{\til{T}}(X)} \cdot \left( \frac{5\kappa}{6} \int_{1}^{X} \frac{\theta_n(H)}{H^{1/6}} \left( (n \mod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H) \right) \, dH + \theta_n(X) \cdot O(X^{1/2}) \right)$$

$$= \frac{5\kappa}{6\pi_{\til{T}}(X)} \cdot \int_{1}^{X} \frac{\theta_n(H)}{H^{1/6}} \left( (n \mod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H) \right) \, dH + \theta_n(X) \cdot O(X^{-1/3}),$$
where we have used the fact that \( \tilde{T} \in T \) for the estimate \( \pi_{\tilde{T}}(X) = O(X^{5/6}) \). Moreover, by Corollary 8.4, the standard error in the approximation of the average by the expected value is given by

\[
\frac{1}{\pi_{\tilde{T}}(X)} \sqrt{\frac{n}{2}} \sum_{E \in \tilde{T} S_n(X)} \rho_n(\text{ht}(E))(1 - \rho_n(\text{ht}(E))) + (\lfloor n/2 \rfloor - 1)C_{1,1}(\text{ht}(E))
\]

\[
= \frac{1}{\pi_{\tilde{T}}(X)} \sqrt{\frac{5\kappa(n/2)}{6}} \int_1^X \frac{\theta_n(H)}{H^{1/6}} (\rho_n(H)(1 - \rho_n(H)) + (\lfloor n/2 \rfloor - 1)C_{1,1}(H)) \, dH + O(X^{1/2}),
\]

\[
= \sqrt{\frac{5\kappa(n/2)}{6\pi_{\tilde{T}}(X)^2}} \int_1^X \frac{\theta_n(H)}{H^{1/6}} (\rho_n(H)(1 - \rho_n(H)) + (\lfloor n/2 \rfloor - 1)C_{1,1}(H)) \, dH + O(X^{-7/6}).
\]

\[\square\]

**Remark 10.3.** Let \( h_n \) be the smallest positive integer such that \( |C_n h_n^{-e_n}| < 1 \). If we assume Hypotheses 6.3 and 8.13, then \( \text{AvgRank}_{\tilde{T} S_n}(X) \) is approximately given by

\[
\approx \frac{5\kappa}{6\pi_{\tilde{T}}(X)} \cdot \int_1^X \frac{\theta_n(H)}{H^{1/6}} ((n \mod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \rho_n(H)) \, dH
\]

\[
\approx \frac{(5/6)\kappa s_n}{\pi_{\tilde{T}}(X)} \cdot \int_1^X \frac{1}{H^{1/6}(1 + C_n H^{-e_n})} \left( (n \mod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \frac{D_n}{H^{f_n}} \right) \, dH
\]

\[
\approx \frac{(5/6)\kappa s_n}{\pi_{\tilde{T}}(X)} \cdot \left( \mu_n + \int_{h_n}^\infty \sum_{m=0}^\infty (-C_n)^m H^{-1/6-m e_n} \left( (n \mod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \frac{D_n}{H^{f_n}} \right) \, dH \right)
\]

where \( \mu_n = \int_{h_n}^{\infty} \sum_{m=0}^\infty (-C_n)^m H^{-1/6-m e_n} \left( (n \mod 2) + 2 \left\lfloor \frac{n}{2} \right\rfloor \frac{D_n}{H^{f_n}} \right) \, dH \). Thus,

\[
\approx \frac{(5/6)\kappa s_n}{\pi_{\tilde{T}}(X)} \cdot \left( \mu_n - \pi \sum_{m=0}^\infty \frac{(-C_n)^m}{5/6-m e_n} (h_n)^{5/6-m e_n} - 2 \left\lfloor \frac{n}{2} \right\rfloor \sum_{m=0}^\infty \frac{D_n (-C_n)^m}{5/6-f_n-m e_n} (h_n)^{5/6-f_n-m e_n} \right)
\]

\[
+ \frac{(5/6)\kappa s_n}{\pi_{\tilde{T}}(X)} \cdot \left( \pi \sum_{m=0}^\infty \frac{(-C_n)^m}{5/6-m e_n} X^{5/6-m e_n} + 2 \left\lfloor \frac{n}{2} \right\rfloor \sum_{m=0}^\infty \frac{D_n (-C_n)^m}{5/6-f_n-m e_n} X^{5/6-f_n-m e_n} \right),
\]

where we have abbreviated \( \pi = (n \mod 2) \), and below we shall write \( \tau_n \) for the contents inside the first parenthesis, i.e., \( \tau_n = \mu_n - \pi \sum_{m=0}^\infty \cdots - 2 \left\lfloor \frac{n}{2} \right\rfloor \sum_{m=0}^\infty \cdots \)

\[
\approx \frac{\kappa s_n X^{5/6}}{\pi_{\tilde{T}}(X)} \cdot \left( \frac{\tau_n}{X^{5/6}} + \sum_{m=0}^\infty \frac{(n \mod 2) (-C_n)^m}{1-(6/5)m e_n} X^{-f_n} 2 \left\lfloor \frac{n}{2} \right\rfloor \frac{D_n (-C_n)^m}{1-(6/5)(f_n+m e_n)} X^{-m e_n} \right)
\]

\[
\approx \frac{s_n}{X^{5/6}} + \sum_{m=0}^\infty \left( (n \mod 2) (-C_n)^m 1-(6/5)m e_n + X^{-f_n} 2 \left\lfloor \frac{n}{2} \right\rfloor \frac{D_n (-C_n)^m}{1-(6/5)(f_n+m e_n)} X^{-m e_n} \right).
\]

Hence, we obtain the following result about the average rank of (test) elliptic curves.
Finally, we point out that, by Proposition 2.6 of [24], the values $s_n$ such that the expected value of $\text{AvgRank}_{\tilde{T}S_n}(X)$ is given by

$$= \sum_{n=1}^{\infty} \text{AvgRank}_{\tilde{T}S_n}(X) \approx \sum_{n=1}^{\infty} s_n \cdot \left( \frac{\tau_n}{X^{5/6}} + \sum_{m=0}^{\infty} \left( \frac{(n \mod 2)(-C_n)^m}{1 - (6/5)me_n} + X^{-f_n} \frac{2\left(\frac{n}{2}\right)}{1 - (6/5)(f_n + me_n)} \right) X^{-me_n} \right),$$

with standard error $\leq \frac{\sum_{n=2}^{\infty} \sqrt{n/2} \cdot (\left\lfloor n/2 \right\rfloor - 3/4) \cdot s_n}{\sqrt{\kappa}X^{5/12}}$. In particular,

$$\lim_{X \to \infty} \text{AvgRank}_{\tilde{T}S_n}(X) = \sum_{k=0}^{\infty} s_{2k+1} = \frac{1}{2},$$

with standard error going to 0 as $X \to \infty$.

**Proof.** The approximation of the average rank is an immediate consequence of our approximation of the contribution to the average rank coming from each Selmer rank $n$ given in Remark 10.3. From the approximation, it follows that

$$\lim_{X \to \infty} \text{AvgRank}_{\tilde{T}S_n}(X) \approx \sum_{n=1}^{\infty} s_n \cdot (n \mod 2) = \sum_{k=0}^{\infty} s_{2k+1}.$$

Finally, we point out that, by Proposition 2.6 of [24], the values $s_n$ have a generating function

$$\sum_{n \geq 0} s_n z^n = \prod_{i=0}^{\infty} \frac{1 + 2^{-i}z}{1 + 2^{-i}}.$$

In particular, for $z = 1$ we obtain that $\sum_n s_n = 1$, for $z = -1$ we obtain that $\sum_n (-1)^n s_n = 0$, and therefore $\sum_k s_{2k+1} = \sum_{n \equiv 1 \mod 2} s_n = \frac{1}{2}(\sum_n s_n - \sum_n (-1)^n s_n) = \frac{1}{2}$. Let us now estimate the error in the approximation of $\text{AvgRank}_{\tilde{T}S_n}(X)$ using Corollary 8.4:

$$\frac{1}{\pi_{\tilde{T}}(X)} \sqrt{n/2} \sum_{E \in \tilde{T}S(X)} \rho_n(\text{ht}(E))(1 - \rho_n(\text{ht}(E))) + ([n/2] - 1)C_{1,1}^n(\text{ht}(E)) \approx \frac{1}{\pi_{\tilde{T}}(X)} \sqrt{5 \kappa n/2} \int_1^X \frac{\theta_n(H)}{H^{1/6}} (\rho_n(H)(1 - \rho_n(H))) + ([n/2] - 1)C_{1,1}^n(H) \, dH.$$

Recall that the covariance coefficient $C_{1,1}^n(X)$ is given by $\mathbb{E}(Y_1 Y_2) - \mathbb{E}(Y_1)\mathbb{E}(Y_2)$, and since the random variables $Y_i$ take only the values 0, 1, we have $0 \leq \mathbb{E}(Y_i) = \rho_n(X) \leq 1$. In particular, $|C_{1,1}^n(X)| \leq 1$. The approximation, it follows that
Also, notice that \( y(1 - y) \) in the interval \([0, 1]\) obtains the maximum value of \(1/4\) at \(y = 1/2\). Thus,

\[
\approx \frac{1}{\pi \varphi(X)} \sqrt{\frac{5 \kappa [n/2] s_n}{6} \int_1^X \frac{1}{H^{1/6}(1 + C_n H^{-e_n})} \left( \frac{D_n}{H^{1/6}} \left( 1 - \frac{D_n}{H^{1/6}} \right) + (\lfloor n/2 \rfloor - 1) C_{1,1}^n(H) \right) dH.}
\]

\[
\leq \frac{1}{\pi \varphi(X)} \sqrt{\frac{5 \kappa [n/2] s_n}{6} \int_1^X \frac{1}{\sqrt{1/4 + (\lfloor n/2 \rfloor - 1) H^{1/6}}} dH}
\]

\[
\leq \frac{1}{\pi \varphi(X)} \sqrt{\kappa [n/2] s_n (\lfloor n/2 \rfloor - 3/4) X^{5/6}}
\]

\[
= \frac{\sqrt{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 3/4) s_n}}{\sqrt{X^{5/12}}}
\]

Thus, the standard error in the approximation of \(\text{AvgRank}_\varphi(X)\) is bounded by

\[
\sum_{n=2}^\infty \frac{\sqrt{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 3/4) s_n}}{\sqrt{X^{5/12}}}
\]

It remains to show that \(\sum_{n=2}^\infty \sqrt{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 3/4) s_n}\) is convergent. Let us define \(t_1 = s_1\) and

\[
t_n = \frac{t_1}{2^{n(n-1)/2} - 1}
\]

for \(n \geq 2\). Then, the definition of \(s_n\) implies that \(s_n \leq t_n\), and therefore,

\[
\sum_{n=2}^N \sqrt{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 3/4) s_n} \leq \sum_{n=2}^N \frac{n}{2} \sqrt{s_n} \leq \sum_{n=2}^N \frac{n}{2} \sqrt{t_n} \leq \sum_{n=2}^N \frac{n}{2} \sqrt{\frac{\sqrt{t_1}}{4}} \leq \sum_{n=2}^N \frac{\sqrt{s_1} \cdot n}{2^{n(n-1)/2}}
\]

for any \(N\), and therefore \(\sum_{n=2}^\infty \sqrt{\lfloor n/2 \rfloor (\lfloor n/2 \rfloor - 3/4) s_n}\) is convergent. Thus, the error goes to 0 as \(X \to \infty\), as desired.

**Remark 10.5.** Using SageMath, in Figure 18 we have plotted values of \(\text{AvgRank}_\varphi(X)\) from the BHKSSW database, and (via numerical integration) the sum of the approximations given in Theorem 10.2 of \(\text{AvgRank}_{S_n}(X)\) for \(n = 1, \ldots, 5\). According to the database, we have

\[
\text{AvgRank}_\varphi(2.7 \cdot 10^{10}) = 0.90197580
\]

while our approximation gives 0.90244770. Thus, the absolute error is 0.00047189, which represents a 0.0523% of the true value.

**References**

[1] J. S. Balakrishnan, W. Ho, N. Kaplan, S. Spicer, W. Stein, J. Weigandt, Databases of elliptic curves ordered by height and distributions of Selmer groups and ranks, to appear in ANTS XII. 1, 1.3, 1.5, 3, 6.1, 6.5, 8.17

[2] B. Bektemirov, B. Mazur, W. Stein, M. Watkins, Average ranks of elliptic curves: tension between data and conjecture, Bull. (New Series) Amer. Math. Soc., Vol. 44, No. 2, April 2007, 233-254. 9

[3] W. Bosma, J. Cannon, and C. Playoust, The Magma algebra system. I. The user language, J. Symbolic Comput., 24 (1997), 235–265.

[4] A. Brumer, The average rank of elliptic curves I, Invent. Math. 109 (1992), no. 3, 445-472. 3, 3.1

[5] J. W. S. Cassels, Arithmetic on curves of genus 1. IV. Proof of the Hauptvermutung, Journal für die reine und angewandte Mathematik, 211: 95-112. 7
Figure 18. Values of $\text{AvgRank}_E(X)$ from the BHKSSW database (blue dots), and the approximation given in Corollary 10.4 (in red).

[6] H. Cramér, *On the order of magnitude of the difference between consecutive primes*, Acta Arith. 2, 23-46.

[7] J. E. Cremona, *Elliptic curve data*, database available at [http://homepages.warwick.ac.uk/~masgaj/ftp/data/](http://homepages.warwick.ac.uk/~masgaj/ftp/data/).

[8] J. E. Cremona, B. Mazur, *Visualizing Elements in the Shafarevich-Tate Group*, Experiment. Math. 9:1 (2000), 13-28.

[9] C. Delaunay, *Heuristics on Tate-Shafarevitch groups of elliptic curves defined over $\mathbb{Q}$*, Experiment. Math. 10 (2001), no. 2, 191-196.

[10] C. Delaunay, *Moments of the orders of Tate-Shafarevich groups*, Int. J. Number Theory 1 (2005), no. 2, 243-264.

[11] A. Dujella, Website: [https://web.math.pmf.unizg.hr/~duje/tors/tors.html](https://web.math.pmf.unizg.hr/~duje/tors/tors.html), *High rank elliptic curves with prescribed torsion*.

[12] D. Goldfeld, *Conjectures on elliptic curves over quadratic fields*, Number theory, Carbondale 1979 (Proc. Southern Illinois Conf., Southern Illinois Univ., Carbondale, Ill., 1979), Lecture Notes in Math., vol. 751, Springer, Berlin, 1979, pp. 108-118.

[13] A. Granville, *Harald Cramér and the Distribution of Prime Numbers*, Scand. Actuarial J. 1995; 1: 12-28.

[14] G.H. Hardy, E. M. Wright, *An Introduction to the Theory of Numbers*, 5th ed. Oxford, England: Clarendon Press, 1979.

[15] R. Harron, A. Snowden, *Counting elliptic curves with prescribed torsion*, to appear in Journal für die reine und angewandte Mathematik.

[16] D. R. Heath-Brown, *The size of Selmer groups for the congruent number problem*, Invent. Math. 111 (1993), no. 1, 171-195.

[17] D. R. Heath-Brown, *The size of Selmer groups for the congruent number problem. II*, Invent. Math. 118 (1994), no. 2, 331-370.

[18] D. M. Kane, *On the ranks of the 2-Selmer groups of twists of a given elliptic curve*, Algebra Number Theory 7 (2013), 1253-1279.

[19] M. A. Kenku, *On the number of $\mathbb{Q}$-isomorphism classes of elliptic curves in each $\mathbb{Q}$-isogeny class*, J. Number Theory 15 (1982), 199-202.

[20] D. S. Kubert, *Universal bounds on the torsion of elliptic curves*, Compositio Math. 38 (1979), no. 1, 121-128.
[21] B. Mazur, *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. 47 (1977), 33–186. 1
[22] B. Mazur, *Rational isogenies of prime degree*, Invent. Math. 44 (1978), 129–162. 1
[23] J. Park, B. Poonen, J. Voight, M. M. Wood, *A heuristic for boundedness of ranks of elliptic curves*, submitted. 1, 3, 7, 9
[24] B. Poonen, *Average rank of elliptic curves [after Manjul Bhargava and Arul Shankar]*, Séminaire Bourbaki, Janvier 2012 / revised June 13, 64ème année, 2011-2012, no. 1049. 1, 1.3, ??, 4, 6.1, 10, 10
[25] B. Poonen, E. Rains, *Random maximal isotropic subspaces and Selmer groups*, J. Amer. Math. Soc. 25 (2012), 245-269.
[26] J. H. Silverman, *The arithmetic of elliptic curves*, Springer-Verlag, 2nd Edition, New York, 2009. 4
[27] P. Swinnerton-Dyer, *The effect of twisting on the 2-Selmer group*, Math. Proc. Cambridge Philos. Soc. 145 (2008), no. 3, 513-526. 4
[28] The SageMath Developers, Sage Mathematics Software (Version 6.9), 2015, http://www.sagemath.org 3.2
[29] M. Watkins, *Some heuristics about elliptic curves*, Experiment. Math. 17 (2008), no. 1, 105-125. 1, 9

Department of Mathematics, University of Connecticut, Storrs, CT 06269, USA
E-mail address: alvaro.lozano-robledo@uconn.edu
URL: http://alozano.clas.uconn.edu/