THE GEOMETRIC AVERAGE SIZE OF SELMER GROUPS

AARON LANDESMAN

1. Background on Elliptic curves

Theorem 1.1 (Mordell-Weil)**.** *Let E be an elliptic curve over a global field K (such as $\mathbb{Q}$ or $\mathbb{F}_q(t)$). Then the group of K-rational points $E(K)$ is a finitely generated abelian group.*

For E an elliptic curve over K, write $E(K) \simeq \mathbb{Z}^r \oplus T$ for T a finite group. Then, r is the **rank** of E.

Question 1.2 (Motivating Question)**.** What is the average rank of an elliptic curve?

Conjecture 1.3 (Minimalist Conjecture)**.** The average rank of elliptic curves is $1/2$, and moreover, 50% of curves have rank 0 and 50% have rank 1, and 0% have rank more than 1.

We'll explain, why, in a certain sense, 0% of elliptic curves have rank more than 1, in accordance with this minimalist conjecture. The idea is to bound the rank in terms of a related object called the selmer group.

Let $K = \mathbb{F}_q(t)$, and let v index the closed points of $\mathbb{P}^1_{\mathbb{F}_q}$. For E an elliptic curve over K, the multiplication by n exact sequence induces the sequences on étale cohomology

$$\begin{array}{ccccccccc} 0 & \longrightarrow & E(K)/nE(K) & \longrightarrow & H^1(\operatorname{Spec} K, E[n]) & \xrightarrow{\beta} & H^1(\operatorname{Spec} K, E)[n] & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & \searrow^{\alpha} & \downarrow{\scriptstyle \gamma} & & \\ 0 & \to & \prod_{v \in \mathbb{P}^1_{\mathbb{F}_q}} E(K_v)/nE_v(K_v) & \to & \prod_v H^1(\operatorname{Spec} K_v, E_v[n]) & \to & \prod_v H^1(\operatorname{Spec} K_v, E_v)[n] & \to & 0. \end{array}$$

Definition 1.4. The **n-Selmer** group of E is

$$\operatorname{Sel}_n(E) := \ker \alpha$$

Lemma 1.5. *There is an injection $E(K)/nE(K) \to \operatorname{Sel}_n(E)$. In particular, $\operatorname{rk}(E) \leq \operatorname{rk}_{\mathbb{Z}/n} \operatorname{Sel}_n(E)$.*

Proof. $E(K)/nE(K) = \ker \beta \subset \ker \alpha = \operatorname{Sel}_n(E)$. □

2. Average sizes of Selmer groups

We'd next like to make sense of a notion of the average size of Selmer groups. Unfortunately, there are infinitely many elliptic curves. To deal with this problem, we introduce a notion of height. There will be only finitely many elliptic curves of a given height, and so we can compute the average size by computing the average for a given height, and then taking a limit.

Definition 2.1. Say E is in minimal Weierstrass form by

$$y^2z = x^3 + A(s,t)xz^2 + B(s,t)z^3$$

(so char $\mathbb{F}_q > 3$) where there exists d so that $A(s,t)$ and $B(s,t)$ are homogeneous polynomials of degrees $4d$ and $6d$. The **height** of E is

$$h(E) := d.$$

Definition 2.2. The **average size** of the n-Selmer group of height up to d is

$$\text{Average}^{\leq d}(\#\text{Sel}_n\,/\mathbb{F}_q(t)) := \frac{\sum_{E/\mathbb{F}_q(t), h(E)\leq d} \#\text{Sel}_n(E)}{\#\{E/\mathbb{F}_q(t)\colon h(E)\leq d\}},$$

where the sum runs over isomorphism classes of elliptic curves $E/\mathbb{F}_q(t)$, having $h(E) \leq d$. We refer to this informally as "the average size of $\#\text{Sel}_n$ for elliptic curves of height $\leq d$ over $\mathbb{F}_q(t)$."

Conjecture 2.3 (Bhargava–Shankar [BS13a, Conjecture 4] and Poonen–Rains [PR12, Conjecture 1.4(b)])**.** When all elliptic curves are ordered by height,

$$\lim_{q\to\infty}\lim_{d\to\infty} \text{Average}^{\leq d}(\#\text{Sel}_n\,/\mathbb{F}_q(t)) = \sum_{s|n} s.$$

Remark 2.4.

- An analogous statement over $\mathbb{Q}$ was shown for $n = 2, 3, 4, 5$ by Bhargava and Shankar.
- The upper bound was shown for $n = 3$ over $\mathbb{F}_q(t)$ by de Jong.
- This was shown for $n = 2$ more generally over function fields by Ho, Le Hung, and Ngo.

Corollary 2.5. *Let $P_q^{\leq d}$ denote the proportion of elliptic curve of rank ≥ 2 over $\mathbb{F}_q(t)$ of height up to d. If Conjecture 2.3 were true,*

$$\lim_{d\to\infty} P_q^{\leq d} = 0$$

That is, 100% *of elliptic curves over $\mathbb{F}_q(t)$ have rank at most* 1.

Proof. Suppose some positive proportion ϵ have rank ≥ 2. Since $n^{\mathrm{rk}\, E} \leq \#\mathrm{Sel}_n(E)$, Then,

$$\epsilon n^2 \leq \mathrm{Average}(n^{\mathrm{rk}\, E}) \leq \mathrm{Average}\,(\#\mathrm{Sel}_n(E)) = \sum_{s|n} s.$$

For primes n, this says $\epsilon n^2 \leq n+1$, which is false for sufficiently large n. □

While counting $\lim_{d\to\infty} \mathrm{Average}^{\leq d}(\#\mathrm{Sel}_n\, / \mathbb{F}_q(t))$ may be difficult, there is another way we can try to take this average. Instead of taking the large d limit, we can try to take the large q limit. First, we can take a large q limit. That is, we can try to show

$$\lim_{q\to\infty} \mathrm{Average}^{\leq d}(\#\mathrm{Sel}_n\, / \mathbb{F}_q(t)) = \sum_{s|n} s.$$

Indeed, our main result is

Theorem 2.6. *For $n \geq 1$ and $d \geq 2$,*

$$\lim_{\substack{q\to\infty \\ \gcd(q,2n)=1}} \mathrm{Average}^{\leq d}(\#\mathrm{Sel}_n\, / \mathbb{F}_q(t)) = \sum_{s|n} s.$$

Remark 2.7. In particular, we find

$$\lim_{d\to\infty} \lim_{\substack{q\to\infty \\ \gcd(q,2n)=1}} \mathrm{Average}^{\leq d}(\mathrm{Sel}_n\, / \mathbb{F}_q(t)) = \sum_{s|n} s.$$

This is a version of the Bhargava-Shankar conjecture by with the limits reversed. But switching the limits may be a hard problem!

Corollary 2.8. *If $P_q^{\leq d}$ denotes the proportion of elliptic curve of rank ≥ 2 over $\mathbb{F}_q(t)$ of height up to d, for $d \geq 2$,*

$$\lim_{q\to\infty} P_q^{\leq d} = 0$$

That is, 100% *of elliptic curves of height up to d have rank at most* 1.

Proof. The proof is the same as Corollary 2.8. □

3. Remarks on the main result

Remark 3.1 (Three heuristics for the average size of Selmer groups)**.** When computing the average size of Selmer groups, it is natural to ask if there is some deeper reason for why the average size of n-Selmer groups should be $\sum_{s|n} s$.

(1) In [BS13a, Conjecture 4], the conjecture is based on the fact that the Tamagawa number PGL_s is s, and their average size is revealed to be the sum of the Tamagawa numbers of PGL_s for $s \mid n$. The same heuristic is used in [BS15a, BS15b, BS13b], and [HLHN14].
(2) In this paper, we present another heuristic: the average size of n-Selmer groups is the number of orbits of a certain orthogonal group on a rank $12d-4$ free $\mathbb{Z}/n\mathbb{Z}$ module. Such orbits are in bijection with geometric components (i.e., irreducible components over an algebraic closure) of a moduli space for Selmer elements, which we call the n-Selmer space.

Omit this one? Yet a third heuristic appears in [dJ02] for 3-Selmer groups, in [Vak01] for 2-Selmer groups, and in [dJF11, Theorem 5.4] for n-Selmer groups. These works suggest that the average size of the n-Selmer group should equal the number of balanced (also called rigid) rank m projective bundles over $\mathbb{P}^1$ for $s \mid n$. Indeed, the balanced rank m projective bundles are all of the form $\mathrm{Proj}_{\mathbb{P}^1} \mathrm{Sym}^\bullet(\mathscr{O}^{\oplus a} \oplus \mathscr{O}(-1)^{\oplus m-a})$ for $1 \leq a \leq m$, and so there are m total such bundles. Altogether, there are $\sum_{s|n} s$ such bundles as s ranges over the divisors of n.

Remark 3.2 (Heuristics for distributions of Selmer groups)**.** In addition to predictions for average sizes of Selmer groups, there are also predictions for the higher moments and distributions of Selmer groups, such as in [PR12] and [BKL+15]. In fact, we can prove the geometric analog:

Theorem 3.3 (Feng-L)**.** *Let n be a squarefree positive integer and $m \geq 1$. For $d \geq \max(2, \frac{m+4}{6})$,*

(3.1)
$$\lim_{\substack{q\to\infty \\ \gcd(q,2n)=1}} \mathrm{Average}^{\leq d}((\#\mathrm{Sel}_n)^m/\mathbb{F}_q(t)) = \prod_{\ell \text{ prime} \mid n} (1+\ell)(1+\ell^2)\cdots(1+\ell^m).$$

Similarly to the $m = 1$ case, for us, this average size is realized as the number of orbits of an orthogonal group $\mathrm{O}(q)$ with underlying vector space V acting diagonally on V^m.

Remark 3.4 (Average sizes, with a twist!)**.** Let $\mathbb{F}_q$ be a finite field of characteristic more than 3. Recall that if E is an elliptic curve over $\mathbb{F}_q(t)$ defined by $y^2z = x^3 + a_4(s,t)xz^2 + a_6(s,t)z^3$, one can define the quadratic twist family of degree d as those elliptic curves of the form $f(s,t)y^2z = x^3 + a_4(s,t)xz^2 + a_6(s,t)z^3$, for $f(s,t) \in k[s,t]$ varying over square-free homogeneous polynomial of degree d. This is a family over an open subscheme of affine space parameterized by the coefficients of $f(s,t)$.

A similar result was worked out for certain quadratic twist families by Niudun Wang and Sun Woo Park, using a similar method (though the relevant

monodromy result they use already holds in characteristic p, whereas in my case, the result only was previously known in characteristic 0).

4. Sketch of the proof in the essential case n is prime

(1) Algebraic geometry: For k a finite field, Construct a space $\mathrm{Sel}^d_{n,k}$ parameterizing pairs (E, X), where E is an elliptic curve over $k(t)$ and X is approximately (but not exactly) an n-Selmer element of E. Letting $\mathscr{W}^d_k$ denote a parameter space for Weierstrass equations of elliptic curves $E/k(t)$ of height d, there is quasifinite étale map $\mathrm{Sel}^d_{n,k} \to \mathscr{W}^d_k$ sending $(E, X) \mapsto [E]$. There is a bijection between $\mathrm{Sel}_n(E)$ and k points of the fiber $[E] \times_{\mathscr{W}^d_k} \mathrm{Sel}^d_{n,k}$.

(2) Number theory: Computing the average size of n-Selmer groups in the large q limit is reduced to computing the ratio $\frac{\#\mathrm{Sel}^d_{n,k}(k')}{\#\mathscr{W}^d_k(k')}$ for sufficiently large finite extensions k' of k.

Theorem 4.1 (Lang-Weil)**.** *For X a finite type space over $\mathbb{F}_p$ with r geometrically irreducible components,* $\lim_{q\to\infty} X(\mathbb{F}_q) = rq^{\dim X} + O(q^{\dim X - 1/2})$.

By the Lang-Weil estimate, since $\mathscr{W}^d_k$ is geometrically irreducible, this ratio is the number of components of $\mathrm{Sel}^d_{n,k}$.

(3) Topology: [Draw a picture] It remains to compute the number of components. We do this by a monodromy argument. To compute the number of components of $\mathrm{Sel}^d_{n,k}$, we show that, over a dense open $\mathscr{W}^{\circ d}_k \subset \mathscr{W}^d_k$, $\mathrm{Sel}^d_{n,k}$ is a finite étale cover, with geometric fibers $(\mathbb{Z}/n\mathbb{Z})^{12d-4}$. Hence, we obtain a monodromy representation (or Galois representation)

$$\rho^d_k(n) : \pi_1^{\text{ét}}(\mathscr{W}^{\circ d}_k) \to \mathrm{GL}(V^d_{n,k}).$$

Because this respects Poincaré duality, it lands in the orthogonal group $\mathrm{O}(Q)$. As n is prime, we want to show this has $\sum_{s|n} s = n + 1$ components. The components are identified with the number of orbits of $\rho^d_k(n)$ on the underlying vector space $(\mathbb{Z}/n\mathbb{Z})^{12d-4}$.

[Draw a picture]

If the monodromy were the full orthogonal group, Chevalley's theorem says there are $n + 1$ orbits: the 0 orbit and the n level sets of the quadratic form. In fact, some group theory shows index 2 subgroups of the orthogonal group have the same orbits. So, we have reduced to showing:

Theorem 4.2. *The image* $\operatorname{im}\rho_k^d(n) \subset \mathrm{O}(Q)$ *is contains an index* 2 *subgroup in* $\mathrm{O}(Q)$.

Over $k = \mathbb{C}$, it is shown in [dJF11] (by looking at small loops around a very singular elliptic surface, and examining a corresponding dynikin diagram) that $\operatorname{im}\rho_{\mathbb{C}}^d(n)$ has index at most 2 in $\mathrm{O}(Q)$, when $d \geq 2$. Fortunately, this index 2 subgroup still has $n+1$ orbits, and so the space has $n+1$ components, over $\mathbb{C}$.

(4) Transferring to positive characteristic: Since we are interested in components over $\mathbb{F}_p$, we need to show the components do not "come together" when one passes to finite fields. That is, we want to show the monodromy over $\overline{\mathbb{F}}_p$ agrees with that over $\mathbb{C}$. The theory of the tame fundamental group shows this is this case if, when one compactifies the map $\mathrm{Sel}^{\circ d}_{n,k} \to \mathscr{W}^{\circ d}_k$, the ramification orders along divisors in the complement of $\mathscr{W}^{\circ d}_k$ are prime to p. There is only one divisor in the complment, corresponding to where two singular fibers come together (the elliptic curve has one place of type I_2 reduction). We want to check travelling around this divisor has order 2. [Draw a picture] One can prove this by describing the effect of travelling once around in terms of combinatorial monodromy data.

Remark 4.3. If one could better understand the topology of the Selmer space, one might be able to compute the average size in the more conventional large d limit as well. The idea for doing this is to use the Lefschetz trace formula (in place of the Lang Weil bounds). If one could show these spaces have homological stability, one could then deduce the result in the large q limit.

These Selmer spaces admit several equivalent descriptions, which suggest they likely have homological stability in d (we showed their H^0 groups stabilize at $d = 2$.

One description is as follows. Let $\mathscr{X}$ denote the moduli stack of genus 1 curves with a degree n divisor. Then, the space $\mathrm{Sel}^d_{n,k}$ is the space of maps from $\mathbb{P}^1 \to \mathscr{X}$ of degree d. There are many results on homological stability in the degree d for maps from $\mathbb{P}^1$ to a space X, e.g., one of the earliest is due to Segal, who showed this for X a projective space.

Another description is as a Hurwitz space of $\mathrm{ASL}_2(\mathbb{Z}/n\mathbb{Z})$ covers $\mathbb{P}^1$, with branching along the singular fibers. If the branch points were unconstrained, this would satisfy homological stability by a result of Ellenberg-Venkatesh-Westerland, but unfortunately the branch points have to lie in a certain special position (but it is essentially a pullback of a Hurwitz space to this special locus).

REFERENCES

[BKL+15] Manjul Bhargava, Daniel M. Kane, Hendrik W. Lenstra, Jr., Bjorn Poonen, and Eric Rains. Modeling the distribution of ranks, Selmer groups, and Shafarevich-Tate groups of elliptic curves. *Camb. J. Math.*, 3(3):275–321, 2015.

[BS13a] Manjul Bhargava and Arul Shankar. The average number of elements in the 4-selmer groups of elliptic curves is 7. *arXiv preprint arXiv:1312.7333v1*, 2013.

[BS13b] Manjul Bhargava and Arul Shankar. The average size of the 5-selmer group of elliptic curves is 6, and the average rank is less than 1. *arXiv preprint arXiv:1312.7859v1*, 2013.

[BS15a] Manjul Bhargava and Arul Shankar. Binary quartic forms having bounded invariants, and the boundedness of the average rank of elliptic curves. *Ann. of Math. (2)*, 181(1):191–242, 2015.

[BS15b] Manjul Bhargava and Arul Shankar. Ternary cubic forms having bounded invariants, and the existence of a positive proportion of elliptic curves having rank 0. *Ann. of Math. (2)*, 181(2):587–621, 2015.

[dJ02] A. J. de Jong. Counting elliptic surfaces over finite fields. *Mosc. Math. J.*, 2(2):281–311, 2002. Dedicated to Yuri I. Manin on the occasion of his 65th birthday.

[dJF11] A. J. de Jong and Robert Friedman. On the geometry of principal homogeneous spaces. *Amer. J. Math.*, 133(3):753–796, 2011.

[HLHN14] Q. P. H`ô, V. B. Lê Hùng, and B. C. Ngô. Average size of 2-Selmer groups of elliptic curves over function fields. *Math. Res. Lett.*, 21(6):1305–1339, 2014.

[PR12] Bjorn Poonen and Eric Rains. Random maximal isotropic subspaces and Selmer groups. *J. Amer. Math. Soc.*, 25(1):245–269, 2012.

[Vak01] Ravi Vakil. Twelve points on the projective line, branched covers, and rational elliptic fibrations. *Math. Ann.*, 320(1):33–54, 2001.