Arithmetic Statistics

Lectures by: John Cremona
Notes by: Ross Paterson

These notes were taken live during lectures at the CMI-HIMR Computational Number Theory summer school held at the University of Bristol in June 2019. In particular, any mistakes are the fault of the transcriber and not of the lecturer. Remarks in red were not written on the board, and were often added later by the transcriber.

Lecture List

Contents

Lecture 1: Introduction

This will be entirely non-computational, but perhaps some of the motivation is partly computational.

1 Introduction to Local and Global Densities

Question 1. *What is the "probability" of ...*

- *A random plane cubic having*
  - *a rational point?*
  - *a p-adic point?*
  - *p-adic points for all p i.e. is everywhere locally soluble (ELS)?*
- *A random p-adic integer being a square?*
- *A random quadric in n-variables being isotropic?*
- *A random elliptic curve /$\mathbb{Q}$ being semistable?*

More generally, how can we go from knowing the local probabilities to the global probabilities?

Example 1. *What is the probability that a random positive integer is squarefree? Well we can reduce to a local question:*

$$n \text{ is squarefree } \iff \forall p,\ \mathrm{ord}_p(n) \leq 1.$$

The density of $n \in \mathbb{Z}$ for which $\mathrm{ord}_p(n) \leq 1$ is $1 - \frac{1}{p^2}$. Since the condition is determined by $n \mod p^2$, and of these residue classes we know that $p^2 - 1$ pass and 1 fails and so the density should be given by $\frac{p^2-1}{p^2} = 1 - \frac{1}{p^2}$. (Note that we only need a fixed finite precision for p-adic numbers to satisfy such kinds of criteria. This will not always be the case.) One expects that n is squarefree with probability $\prod_p (1 - \frac{1}{p^2}) = \zeta(2)^{-1} = \frac{6}{\pi^2}$ since we expect that each prime should constitute an independent event. It is in fact true that

$$\lim_{X \to \infty} \{n \in \mathbb{Z} : 0 < n \leq X,\ n \text{ squarefree}\} = \frac{6}{\pi^2}.$$

There will be a local step, involving p-adic arithmetic, algebraic geometry (of curves for example) and questions of real solubility.

The definition of global densities involves some choice of real probability distribution; e.g. what is the probability that a random real quadratic $ax^2 + bx + c \in \mathbb{R}[x]$ has a real root? (Note that this is determined by $b^2 - 4ac \geq 0$ which is a homogeneous condition, i.e. scaling a,b,c to $\lambda a, \lambda b, \lambda c$ does not change the outcome so this is a question which can be lifted into projective space.)

$$\mathrm{vol}\left(\left\{(a,b,c) \in [-1,1]^3 : b^2 \geq 4ac\right\}\right)/8$$

The numerator is $(41 + 6\log 2)/9$; If instead we use $ax^2 + 2bx + c$ we get $7/9$, so to my mind this is the better question since it has the nicer answer.

2 Local and Global Densities

Sample space will be $\mathbb{Z}^N$ for some $N \geq 1$; e.g.

- $N = 3$ for random quadratics in one variable,
- $N = 6$ for random conics $aX^2 + bY^2 + cZ^2 + dXY + eYX + fXZ$,

- $N = 5$ for Weierstrass equations

$$(a_1, a_2, a_3, a_4, a_6) \leftrightarrow Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6$$

Note that $\mathbb{Z} \subseteq \mathbb{Z}_p$ and $\mathbb{Z}_p$ is a compact abelian group so has a unique haar measure with total measure 1, so it's our only choice for a local measure. ($\mathbb{R}$ is only locally compact, so we have some choice when looking for a well behaved *finite* measure.)

Definition 2.1 (Local Density)**.** *The p-adic measure on $\mathbb{Z}_p$ has $\mu(\mathbb{Z}_p) = 1$, $\mu(a + p^k\mathbb{Z}_p) = \frac{1}{p^k}$. Moreover, $|p|_p = \frac{1}{p}$ so $x \mapsto xp$ scales measures by $\frac{1}{p}$.*

*If $Y \subseteq \mathbb{Z}_p^N$ then the **density** of Y is $\mu(Y)$.*

Example 2. *1. If Y is defined by congruence conditions mod p^k for some fixed k then $\mu(Y)$ can be found by counting: if m of the p^{kN} residue classes in $\mathbb{Z}_p^N/p^K\mathbb{Z}_p^N$ lie in Y then*

$$\mu(Y) = \frac{m}{p^{kN}}$$

2. p-adic squares:

Theorem 2.2. *The density of squares in $\mathbb{Z}_p$ is* $\begin{cases} \frac{p}{2(p+1)} & p \geq 3 \\ \frac{1}{6} & p = 2 \end{cases}$

proof for $p \neq 2$: Ignore $a = 0$ (single element has density 0). Write $a = up^k$ for $u \in \mathbb{Z}_p^\times$ and $k \geq 0$. So $a \in \mathbb{Z}_p^2$ if and only if $2 \mid k$ and $u \in (\mathbb{Z}_p^\times)^2$. When is $u \in (\mathbb{Z}_p^\times)^2$? This is a mod p condition, and $\frac{p-1}{2}$ classes mod p are squares.

Let $Y_k = \{a \in \mathbb{Z}_p : \mathrm{ord}_p(a) = k\}$, $Y_0 = \mathbb{Z}_p^\times$ and $\mu(Y_0) = 1 - \frac{1}{p}$. Also, $Y_{k+1} = pY_k$ for $k \geq 0$ so $\mu(Y_{k+1}) = \frac{1}{p}\mu(Y_k)$. But

$$\bigcup_{k \geq 0} Y_k = \mathbb{Z}_p \setminus \{0\}$$

so that this union has measure 1 and $\mu(Y_k) = \frac{p-1}{p^{k+1}}$ for all $k \geq 0$.

$Y_{even} = \bigcup_{k \in 2\mathbb{Z}_{\geq 0}} Y_K$ has

$$\mu(Y_{even}) = \frac{p-1}{p}(1 + \frac{1}{p^2} + \dots) = \frac{p-1}{p}\frac{1}{1 - 1/p^2} = \frac{p}{p+1}$$

So the density of p-adic squares should be $\frac{1}{2}\frac{p}{p+1}$

Better Alternative: Let $\alpha = \mu(\text{squares})$, then

$$\begin{aligned} \{\text{squares}\} &= \{\text{unit squares}\} \cup \{\text{non-unit squares}\} \\ &= \{\text{unit squares}\} \cup p^2\{\text{squares}\} \end{aligned}$$

so $\alpha = \frac{p-1}{p}\frac{1}{2} + \frac{1}{p^2}\alpha$ so that

$$\alpha = \frac{p}{2(p+1)}.$$

□

Definition 2.3. *Let $\mathscr{D}$ be a probability distribution on $\mathbb{R}^N$, so $\mathscr{D} \geq 0$, $\int_{\mathbb{R}^N} \mathscr{D}(x)dx = 1$, which is "nice" (peicewise smooth and rapidly decaying). Now for $Y \subset \mathbb{Z}^N$ define*

$$\rho^{\mathscr{D}}(Y) = \lim_{X\to\infty} \frac{\sum_{y\in Y} \mathscr{D}(y/X)}{\sum_{y\in\mathbb{Z}^N} \mathscr{D}(y/X)}$$

if the limit exists.

Example 3. $\mathscr{D}(x) = \begin{cases} \frac{1}{2^N} & |x|_\infty \leq 1 \\ 0 & \text{else} \end{cases}$

$\mathscr{D}(y/X) = \begin{cases} \frac{1}{2^N} & |y|_\infty \leq X \\ 0 & \text{else} \end{cases}$

For quadrics in n variables ($N = \frac{n(n+1)}{2}$):

$$\begin{pmatrix}\text{Prob of a random quadric}\\ /\mathbb{R} \text{ wrt } \mathscr{D} \text{ having a root}\end{pmatrix} = \begin{pmatrix}\text{prob of a random quadric}\\ /\mathbb{Z} \text{ wrt } \mathscr{D} \text{ having a real root}\end{pmatrix}$$

and

$$\begin{pmatrix}\text{Prob of a random quadric}\\ /\mathbb{Z}_p \text{ having a } p\text{−adic root}\end{pmatrix} = \begin{pmatrix}\text{prob of a random quadric}\\ /\mathbb{Z} \text{ wrt } \mathscr{D} \text{ having a } p\text{−adic root}\end{pmatrix}$$

Note that this final one is not a typo, $\mathscr{D}$ somehow becomes irrelevant once we move to the p-adic case.

Lecture 2: Quadratics and Quadrics

3 Quadratics and Quadrics

Today we will actually be computing some densities, and it makes sense to start with quadratic equations and quadrics.

Proposition 3.1. *A random monic quadratic $\mathbf{q} : X^2 + bX + c \in \mathbb{Z}_p[X]$ has roots in $\mathbb{Q}_p$ with probability*

$$\frac{p}{2(p+1)}$$

for all p.

(What is nice about this result, is that $p = 2$ need not be dealt with seperately. In fact one can even prove it without considering 2 seperately.)

Proof. Consider $\mathbf{q} \mod p$ the reduction of the quadratic, and let δ be the probability we seek.

- With probability
$$\frac{1}{2}\frac{p(p-1)}{p^2}$$
$\mathbf{q}$ has distinct roots in $\mathbb{F}_p$ (choose 1 root, then the other and account for ordering). These lift to $\mathbb{Q}_p$.

- With probability $\frac{1}{2}\frac{p^2-p}{p^2}$ $\mathbf{q}$ has conjugate roots over $\mathbb{F}_{p^2}$ (choose one root in $\mathbb{F}_{p^2}$, which is not in $\mathbb{F}_p$ and account for ordering): these do not lift to $\mathbb{Q}_p$.
- With probability $\frac{p}{p^2}$, $\mathbf{q}$ has double root in $\mathbb{F}_p$ (choose a single root): Unclear if these lift to $\mathbb{Q}_p$.

So $\delta = \frac{\frac{1}{2}(p-1)}{p} + \frac{1}{p}\alpha$ for α the probability of solvability given that $\mathbf{q}$ has a double root mod p. WLOG, the double root is at $0 \mod p$ and $b = pb_1$ and $c = pc_1$.

$$\mathbf{q} : X^2 + pb_1X + pc_1 \in \mathbb{Z}_p[X] \qquad \text{Replace } X \text{ by } pX$$

$$\frac{1}{p}\mathbf{q} : pX^2 + pb_1X + c_1 \qquad \begin{cases} \text{with probability } 1 - \frac{1}{p},\ p \nmid c_1 \\ \quad \text{(which has no lift as } p^2 \mid \mathbf{q}(pX)) \\ \text{with probability } \frac{1}{p},\ c_1 = pc_2 \\ \quad \text{(which may have a lift, so we continue.)} \end{cases}$$

$$\frac{1}{p^2}\mathbf{q} : X^2 + b_1X + c_2 \qquad \text{has roots with prob } \delta.$$

Hence $\alpha = (1 - \frac{1}{p}) \cdot 0 + \frac{1}{p}\delta = \frac{\delta}{p}$.
$\delta = \frac{\frac{1}{2}(p-1)}{p} + \frac{1}{p^2}\delta$ and so $\delta = \frac{p}{2(p+1)}$

□

Note that this was uniform in p, and a 2-step recursion.

Proposition 3.2. *A random quadric $q : aX^2 + bXY + cY^2 \in \mathbb{Z}_p[X, Y]$ is **isotropic** over $\mathbb{Q}_p$ with probability $\rho = \frac{1}{2}$.*

Proof. With probability

- $\frac{\frac{1}{2}(p+1)p(p-1)}{p^3}$: the quadric is split, which lift to $\mathbb{Q}_p$. (We are choosing 2 points from $\mathbb{P}^1(\mathbb{F}_p)$ and then a scaling factor in $\mathbb{F}_p^\times$)
- $\frac{\frac{1}{2}(p^2-p)(p-1)}{p^3}$: the quadric has 2 distinct Galois conjugate factors, which does not lift to $\mathbb{Q}_p$ (note that $p^2 - p = \#\left[\mathbb{P}^1(\mathbb{F}_{p^2}) \backslash \mathbb{P}^1(\mathbb{F}_p)\right]$)
- $\frac{(p+1)(p-1)}{p^3}$: the quadric has two copies of the same factor (is a constant multiple of a square quadric), these lift to $\mathbb{Q}_p$ with probability α. (choose a point in $\mathbb{P}^1(\mathbb{F}_p)$ and then a scaling factor in $\mathbb{F}_p^\times$)
- $\frac{1}{p^3}$: the quadric is imprimitive (i.e. is $0 \mod p$), so that

So $\rho = \frac{\frac{1}{2}(p^2-1)}{p^2} \cdot 1 + \frac{(p^2-1)}{p^3} \cdot \alpha + \frac{1}{p^3}\rho$. (Note that we can reduce to the case $X^2 \equiv q \mod p$ so that α is as in Proposition 3.1) Since $\alpha = \frac{1}{p}\frac{p}{2(p+1)}$,

$$\rho\left(1 - \frac{1}{p^3}\right) = \frac{p^2-1}{2p^2} + \frac{p-1}{2p^3} = \frac{1}{2}\left(1 - \frac{1}{p^3}\right)$$

so $\rho = \frac{1}{2}$. □

More generally, let $\rho_n(p)$ be the probability that a random quadric in n variables over $\mathbb{Z}_p$ (i.e. coefficients lie here), is **isotropic** over $\mathbb{Q}_p$.

Theorem 3.3 ([BCF+15]).

n	1	2	3	4	≥ 5
$\rho_n(p)$	*0*	$\frac{1}{2}$	$1-\frac{p}{2(p+1)^2}$	$1-\frac{p^3(p-1)}{4(p+1)^2(p^5-1)}$	*1*

The $n=3$ case tells us the density of plane conics over $\mathbb{Q}_p$ which have $\mathbb{Q}_p$-points. Also, $\rho_3(p)$ is the density of integral $q \in \mathbb{Z}[X,Y,Z]_2$ which define conics with $\mathbb{Q}_p$-points.

Now, $\prod_p(1-\frac{p}{2(p+1)^2}) = 0$ (since $\sum_p \frac{1}{p}$ diverges) and hence the probability that a random conic $/\mathbb{Q}$ is everywhere locally soluble is

$$\leq \prod_{p \leq B}\left(1-\frac{p}{2(p+1)^2}\right) \to 0 \qquad B \to \infty$$

So conics have rational points with probability 0.

For $n=4$ the expression is ugly, we have that a random quadric over $\mathbb{Q}$ is isotropic with probability

$$\rho_4^{GOE}(\infty)\cdot\prod_p \rho_4(p) = \left(\frac{1}{2}+\frac{1}{4\sqrt{2}}+\frac{1}{\pi}\right)\prod_p\left(1-\frac{p^3(p-1)}{4(p+1)^2(p^5-1)}\right) \approx 98.3\%$$

Where note that we have changed distribution to GOE (Gaussian orthogonal ensemble). If we change to the uniform distribution instead of GOE, then the number changes and we can only evaluate numerically.

4 Plane Curves

4.1 Plane Conics

Let $\mathscr{C} : F(X,Y,Z) = 0$ for $F \in \mathbb{Z}[X,Y,Z]_2$ of $\mathbb{Z}_p[X,Y,Z]_2$. How do we find this ρ_3? Well note we have

$$\mathbb{Z}[X,Y,Z]_2 \leftrightarrow \mathbb{Z}^6$$
$$\mathbb{Z}_p[X,Y,Z]_2 \leftrightarrow \mathbb{Z}_p^6$$

Consider how F factors $\mod p$.

1. If $F \mod p$ is smooth then it has $\mathbb{F}_p$-points (Exercise, to do with the Hasse bound) which lift.
2. If $F \equiv L_1L_2 \mod p$ for $L_1 \neq L_2$ then each L_i has smooth points.
3. If $F \equiv \lambda L\overline{L}$ over $\mathbb{F}_{p^2}$ then has only one $\mathbb{F}_p$-point and this point is singular.
4. If $F \equiv \lambda L^2 \mod p$ then there are lots of $\mathbb{F}_p$-points but all are singular
5. $F \equiv 0 \mod p$ with probability $\frac{1}{p^6}$ and so has roots with probability $\frac{1}{p^6}\rho_3(p)$.

Case 3 WLOG $F \equiv$ irreducible binary form in X,Y, say $F \equiv B(X,Y) \mod p$. In particular, the intersection point is $[0:0:1]$. Any point on $\mathscr{C}$ will have $p \mid X$ and $p \mid Y$ so replace $X \mapsto pX$ and $Y \mapsto pY$ and divide by p. So now $F \equiv cZ^2$ (see Figure **??**).

If $p \nmid c$ then $[X:Y:1]$ cannot be a point, giving a contradiction, thus with probability $1-\frac{1}{p}$ there are 0 solutions in this case,

$$\begin{array}{ccccc} & (X^2) & & & (X^2) \\ & \geq 0 & & & \geq 1 \\ & & & X \mapsto pX & \\ & & & Y \mapsto pY & \\ \geq 0 & \geq 1 & & \implies & \geq 1 \quad \geq 1 \\ & & \text{Divide by } p & & \\ (Y^2) \geq 0 & \geq 1 & \geq 1\,(Z^2) & & (Y^2) \geq 1 \quad \geq 1 \quad \geq 0\,(Z^2) \end{array}$$

Figure 1: Valuations of coefficients for Case 3

$$\begin{array}{ccccc} & (X^2) & & & (X^2) \\ & \geq 0 & & & \geq 0 \\ & & & X \mapsto pX & \\ & & & Y \mapsto pY & \\ \geq 0 & \geq 1 & & \implies & \geq 0 \quad \geq 0 \\ & & \text{Divide by } p^2 & & \\ (Y^2) \geq 0 & \geq 1 & \geq 1\,(Z^2) & & (Y^2) \geq 0 \quad \geq 0 \quad \geq 0\,(Z^2) \end{array}$$

Figure 2: Valuations of coefficients for Case 3 if $p \mid c$

$$\begin{array}{ccccc} & (X^2) & & & (X^2) \\ & = 0 & & & = 1 \\ & & & & \\ \geq 1 & \geq 1 & & \implies & \geq 1 \quad \geq 1 \\ & & \text{Divide by } p & & \\ (Y^2) \geq 1 & \geq 1 & \geq 1\,(Z^2) & & (Y^2) \geq 0 \quad \geq 0 \quad \geq 0\,(Z^2) \end{array}$$

Figure 3: Valuations of coefficients for Case 4

If $p \mid c$ then (see Figure **??**) this is just like at the very start but with a side condition, namely $\mathscr{C}(\mathbb{F}_p) \cap \{Z = 0\} = \emptyset$.

Case 4 The idea here is similar, so we cover it even more briefly. WLOG $F \equiv aX^2$, for $p \nmid a$ $p \mid X$ see Figure **??**.

In the imprimitive case, we end up getting figure **??** get as at the start except that $p \nmid$ the coefficient of X^2, i.e. $[1:0:0] \notin \mathscr{C}(\mathbb{F}_p)$. i.e. we obtain a point condition here.

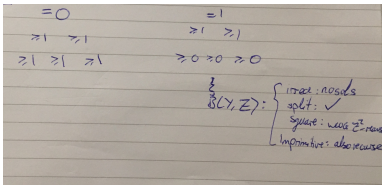

Figure 4:

Lecture 3: Plane Curves

We have some summary remarks from last time.

- A lot of counting is necessary, #conics/$\mathbb{F}_p$:
  - L_1L_2 : $\frac{1}{2}(p^2+p+1)(p^2+p)(p-1)$
  - $L\overline{L}$: $\frac{1}{2}\left[(p^4+p^2+1)-(p^2+p+1)\right](p-1)$
  - L^2 : $(p^2+p+1)(p-1)$
  - : 0 : 1

  If we add up and subtract from p^6 we get that the number of smooth conics is $p^2(p-1)(p^3-1)$. The same with 2 side conditions for number of $\mathbb{F}_p$ points with $Z=0$, $Y=Z=0$ etc...
- The proof is just like the algorithm for testing solubility.
  - End up with linear equations for the probabilities of solution in various configurations. Solving gives the solution.
- First step leads to original configuration but with side conditions. Second step leads to original configuration with no side conditions (A miracle!). This happens for:
  - Plane conics
  - Plane cubics
  - n-ary quadrics
  - Hyperelliptic equations of any genus $g \geq 1$ ($g=1$ is degree 4 polynomial) (the authors are not quite finished working out the details.)

4.2 Plane Cubics

Theorem 4.1. *A random plane cubic over $\mathbb{Q}_p$ has a $\mathbb{Q}_p$ rational point with probability*

$$\rho(p) = \frac{p^9 - p^8 + p^6 - p^4 + p^3 + p^2 - 2p + 1}{3(p^2+1)(p^4+1)(p^6+p^3+1)}$$

so $1-\rho(p) \sim \frac{1}{3p^3}$.

Corollary 4.2. *A random plane cubic /$\mathbb{Q}$ is everywhere locally soluble with probability*

$$\prod_p \rho(p) \approx 97.256\%$$

i..e

$$\lim_{B\to\infty} \frac{\#\left\{\text{cubics in } \mathbb{Z}[X,Y,Z]_3 \text{ Such that all 10 coefficients} \leq B \text{ which are ELS}\right\}}{\#\left\{\text{cubics in } \mathbb{Z}[X,Y,Z]_3 \text{ Such that all 10 coefficients} \leq B\right\}}$$

Remark 4.3. *The ∞ place doesn't contribute because real cubics always have a real root. This also means that we do not need to specify our distribution, since it only matters at the ∞ place.*

Sketch proof: Given $F \in \mathbb{Z}_p[X,Y,Z]_3 \leftrightarrow \mathbb{Z}_p^6$, consider $F \mod p \in \mathbb{F}_p[X,Y,Z]_3$. Possible factorisations are:

- L^3 - we know nothing yet!

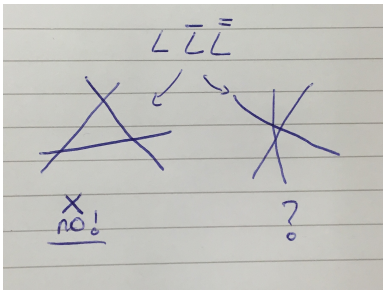

Figure 5: The possible cases for cubics splitting into 3 conjugate lines, along with whether these lift to $\mathbb{Q}_p$.

- $L_1^2 L_2$ We know how to deal with L_1^2 by our work on conics, and then the linear factor is coprime to the L_1^2 so lifts to $\mathbb{Q}_p$,
- $L_1 L_2 L_3$ Three coprime factors will lift to $\mathbb{Q}_p$,
- $L_1 L_2 \overline{L_2}$ We know how to deal with $L_2\overline{L_2}$ by our work on conics, and then the linear factor L_1 is coprime so lifts to $\mathbb{Q}_p$,
- $L\overline{L}\overline{\overline{L}}$ (see figure 5)
- LQ
- Irreducible cubic.

The irreducible case can be smooth (which, using the Hasse bound, can give a point which lifts to $\mathbb{Q}_p$) or 1 singular point. Else this can have a node, which is split or nonsplit, or a cusp.

- If split node, #smooth points $= \#\mathbb{G}_m = p - 1 > 0$
- If nonsplit node, #smooth points $= p + 1 > 0$
- If cusp, #smooth points $= \#\mathbb{G}_a = p > 0$

and so we have points in all of these cases (see your favourite elliptic curves text to see why these are the numbers for the singularity types). □

The reducible counts above add up to

$$(p-1)(p^6 + p^5 + p^4 + p^2 + 1)$$

leaving $p^5(p+1)(p^3-1)$ irreducible cubics. Let $N = p^3(p^3-1)(p^2-1) = \#\,\mathrm{PGL}_3(\mathbb{F}_p)$, so we have $Np^2/(p-1)$ irreducible cubics over $\mathbb{F}_p$. Of these (see exercises for details): N have a node; $N/(p-1)$ have a cusp; Np are smooth. This interesting relationship with PGL_3 is no mere coincidence!

Theorem 4.4. *The number of smooth plane cubic curves over $\mathbb{F}_p$ is $p \cdot \#\,\mathrm{PGL}_3(\mathbb{F}_p)$. Moreover for each value $j_0 \in \mathbb{F}_p$, the number whose Jacobian has j invariant j_0 is $\#\,\mathrm{PGL}_3(\mathbb{F}_p) = N$.*

Proof. Every smooth plane cubic curve C is a genus 1 curve, and so its Jacobian $E = \operatorname{Jac} C$ is an elliptic curve. First we fix E, an elliptic curve over $\mathbb{F}_p$, and count the C with $\operatorname{Jac} C \cong E$. Given E, we can construct C's by embedding our elliptic curve $E \to \mathbb{P}^2$ and taking $C = \operatorname{im}(E)$. Such an embedding determines a degree 3 divisor class on E. Choose a divisor D of degree 3, could take $3 \cdot O$, on E. Let $\mathcal{L}(D)$ be the associated Reimann-Roch space, $\dim \mathcal{L}(D) = 3$ so choose a basis $\langle u, v, w \rangle$. Map

$$\begin{aligned} E &\to \mathbb{P}^2 \\ P &\mapsto [u(P) : v(P) : w(P)] \in \mathbb{P}^2 \end{aligned}$$

Let C be the image. It is cut out by some $F \in \mathbb{F}_p[X, Y, Z]_3$ such that $F(u, v, w) = 0$ identically. Replacing D by $D' \sim D$, say $D' = D - \operatorname{Div}(f)$ then $\mathcal{L}(D') = f \cdot \mathcal{L}(D)$. Thus $\langle fu, fv, fw \rangle$ make a basis here, which gives the same embedding in $\mathbb{P}^2$ as we are in homogeneous coordinates.

Now, $D \sim (P) + 2 \cdot (O_E)$ uniquely, with $P \in E(\mathbb{F}_p)$ so the number of divisor classes of degree 3 is $n_E = \#E(\mathbb{F}_p)$. $\mathcal{L}(D)$ is 3 dimensional so has N bases up to scaling. So we have $n_E N$ constructions of C. Counting repeats, say we have 2 embeddings $\alpha : E \to \mathbb{P}^2$ and $\beta : E \to \mathbb{P}^2$ with the same image C. Then $\beta^{-1}\alpha$ is an automorphism of E as a curve. So the number of such automorphisms is

$$\#\operatorname{Aut}(E) \cdot n_E$$

(these are automorphisms as a variety, so as every such automorphism is a group automorphism with a translation we are taking n_E to count the translations). Hence

$$\#\{C \mid \operatorname{Jac} C = E\} = \frac{N}{\#\operatorname{Aut}(E)}.$$

For fixed j_0 there are in general several different ($\mathbb{F}_p$-)isomorphism classes of E. In the exercises you will show that that

$$\sum_{E : j(E) = j_0} \frac{1}{\#\operatorname{Aut}(E)} = 1.$$

(e.g. for $j \neq 0, 1728$, there are 2 isomorphism classes, and $\operatorname{Aut}(E) = 2$, a more surprising example would be $E/\mathbb{F}_2$ and $j = 0 = 1728$ where we have 3-classes of E, with $\#\operatorname{Aut}(E) = 2, 4, 4$). Hence

$$\#\{C : j(\operatorname{Jac}(C)) = j_0\} = \sum_{E : j(E) = j_0} \frac{N}{\#\operatorname{Aut}(E)} = N$$

so $\#C = N \cdot p$. □

Lecture 4: The Ekedahl Sieve

5 The Ekedahl Sieve

This last lecture we discuss the Ekedahl Sieve, named after a paper from 1991 [Eke91] of Ekedahl. Everything so far has been about finding local densities. Today we look at how to put these together to find global results. In many problems we have

- A local density $\rho(p)$ for each prime p and $\rho(\infty)$ for $\mathbb{R}$.

When does $\prod_{p\leq\infty}\rho(p)$ make sense and represent a corresponding global or (everywhere-local) density? References:

- Ekedahl (1991) [Eke91],
- Poonen-Voloch "Random Diophantine equations" (2003) [PV04]– mainly about hypersurfaces,
- Poonen-Stoll "A local-to-global principle for densities" (1999) [PS99b].

The approach we will take is a bit closer to Poonen-Stoll than Ekedahl. Recall we had an ambient dimension $d \geq 1$ of our problem (e.g. for plane conics this is $d = 6$).

$$\mathcal{U} = (U_p)_p \qquad U_p \subseteq \mathbb{Z}_p^d \quad \text{(e.g. if } d = 1 \text{ we could have } U_p = p^2\mathbb{Z}_p)$$

Let $s_p = \mu_p(U_p)$, and assume that $\mu(\partial U_p) = 0$. For each $M > 0$ define

$$S_M(\mathcal{U}) = \left\{\underline{a} \in \mathbb{Z}^d \mid \underline{a} \in U_p \quad \text{for some } p > M\right\}$$

and let

$$\rho_M(\mathcal{U}) = \overline{\rho}(S_M(\mathcal{U})) = \lim \sup_{X\to\infty} \frac{1}{(2X)^d} \#\left\{\underline{a} \in S_M(\mathcal{U}) : |a_i| \leq X \quad \forall i\right\}$$

so we are counting in a box and taking a limit in the box size, as suggested in Lecture 1.

Definition 5.1. *$\mathcal{U} = (U_p)$ is **admissible** if $\lim_{M\to\infty}\rho_M(\mathcal{U}) = 0$.*

Lemma 5.2. *If $\mathcal{U} = (U_p)$ is admissible, and $\mathcal{U}' = (U_p)$ with either:*

1. *$U_p = U_p'$ for all $p > p_0$.*
2. *$U_p' \subseteq U_p$ for all $p \geq p_0$*

Then $\mathcal{U}'$ is admissible.

Example 4. *$U_p = p^2\mathbb{Z}$ is admissible. This is because*

$$S_M(\mathcal{U}) = \left\{a \in \mathbb{Z} \mid p^2 \mid a \quad \exists p > M\right\} = \bigcup_{p>M} p^2\mathbb{Z}$$

Then $\rho_M(\mathcal{U}) \leq \sum_{p>M}\rho(p^2\mathbb{Z}) = \sum_{p>M}\frac{1}{p^2} \to 0$ as $M \to \infty$

It seems like these admissible sets should be very small, but somehow they have to fill all the gaps and cannot be too small, as seen in the following example.

Example 5. *$U_p = \{n\}$ where $p = p_n$ is the nth prime. e.g. $U_2 = \{1\}$, $U_3 = \{2\}$, $U_5 = \{3\}, \ldots$. Then*

$$\mu_p(U_p) = 0 \quad \forall p$$

$S_M(\mathcal{U}) = \{n : p_n > M\} = \mathbb{Z}_{>0}\backslash\{\text{finite set}\}$. *Thus*

$$\overline{\rho}(S_M(\mathcal{U})) = \frac{1}{2} \qquad \forall M$$

$$\Rightarrow \lim_{M\to\infty}\rho_M(U) = \frac{1}{2} \neq 0.$$

Thus we do not have an admissible collection.

Proposition 5.3 ([PS99b], after Ekedahl)**.** *Let $f, g \in \mathbb{Z}[x_1, \ldots, x_d]$ be coprime, and let $\mathcal{U} = (U_p)$ where*

$$U_p = \left\{ \underline{a} \in \mathbb{Z}_p^d \mid f(\underline{a}) = g(\underline{a}) = 0 \mod p \right\}$$

Then $\mathcal{U}$ is admissible.

Theorem 5.4 ([PS99b])**.** *Let $\mathcal{U} = (U_p)$ be admissible, and $s_p = \mu_p(U_p)$ with $\mu_p(\partial U_p) = 0$. Then*

1. *$\sum_p s_p$ converges,*
2. *$\rho\left(\left\{\underline{a} \in \mathbb{Z}^d \mid \underline{a} \notin U_p \quad \forall p\right\}\right) = \prod_p (1 - s_p)$*
3. *More generally, if S is any finite set of primes,*

$$\rho\left(\left\{\underline{a} \in \mathbb{Z}^d \mid \underline{a} \in U_p \iff p \in S\right\}\right) = \prod_{p \in S} s_p \prod_{p \notin S} (1 - s_p)$$

Example 6. *$U_p = p^2\mathbb{Z}$ and $s_p = \frac{1}{p^2}$ then ρ*(squarefree integers) $= \zeta(2)^{-1}$.

Example 7. *Let*

$$U_p = \left\{ \underline{a} \in \mathbb{Z}_p^{10} \mid \text{ the plane conic defined by } \underline{a} \text{ as coefficients does not have a point over } \mathbb{Q}_p \right\}.$$

So $\mu(U_p) \sim \frac{1}{3p^3}$ (last lecture). By Theorem 3.6 of [PV04] we get that the density of ELS plane cubics is

$$\prod_p (1 - \mu(U_p))$$

5.1 Weierstrass Equations

This work is joint with M. Sadek, the paper [CS19] is unfinished at present. Let $d = 5$ and write

$$Y^2 + a_1XY + a_3Y = X^3 + a_2X^2 + a_4X + a_6$$

For $R = \mathbb{Z}, \mathbb{Z}_p$ etc we identify R^5 with the set of Weierstrass equations over R.

$\mathbf{R} = \mathbb{F}_\mathbf{p}$: The number of Weierstrass equations is p^5. How many are singular? The answer turns out to be p^4, and it goes like this: If such a curve is singular there is a unique singular point, which is **not** the point at infinity i.e. is affine. WLOG it is the point $(0, 0)$. Note that $(0, 0)$ is on the curve if and only if $a_6 = 0$ and is singular if and only if $a_3 = a_4 = 0$. So we have

$$Y^2 + a_1XY = X^3 + a_2X^2$$

$$\left(p^2 \text{ such equations}\right) \times \left(p^2 \text{ such possible singular points}\right) = p^4$$

Hence the number of smooth equations is $p^5 - p^4$. Density of

$$\left\{\text{Smooth reduction } W_{\underline{a}} \mid \underline{a} \in \mathbb{Z}_p^5\right\} = 1 - \frac{1}{p}$$

and $\mu_p\left(\left\{\text{bad reduction } W_{\underline{a}}\right\}\right) = \frac{1}{p}$.

If singular at $(0, 0) \mod p$ then the type of singularity depends on how

$$Y^2 + a_1XY - a_2X^2 = X^3$$

factors $\mod p$. Hence the curve has 3 types of bad reduction:

- Split multiplicative reduction with relative probability $\frac{p-1}{2p}$ to the fact that we assumed $(0,0)$ is singular. Hence the absolute probability is $\frac{p-1}{2p^2}$
- Nonsplit multiplicative reduction with relative probability $\frac{p-1}{2p}$ and absolute probability $\frac{p-1}{2p^2}$
- Additive reduction with relative probability $\frac{1}{p}$ and absolute probability $\frac{1}{p^2}$

since the probability of the singular point was $\frac{1}{p}$. Let

$$U_p = \left\{\underline{a} \in \mathbb{Z}_p^5 \mid W_{\underline{a}} \text{ has additive reduction}\right\}$$

So $\mu_p(U_p) = \frac{1}{p^2}$. Note that we don't know if this is admissible so this doesn't give us global density just yet. For $p \neq 2,3$ the condition that $W_{\underline{a}}$ has additive reduction is that $c_4(\underline{a}) \equiv c_6(\underline{a}) \equiv 0 \mod p$, and $c_4, c_6 \in \mathbb{Z}[X_1, X_2, X_3, X_4, X_6]$ are coprime. Now if we apply the result of Proposition 5.3 to all $p \neq 2,3$ and then use Lemma 5.2 we see that $\mathcal{U} = \{U_p\}$ is admissible. Hence we have:

Theorem 5.5. *A random integral Weierstrass equation represents the global minimal model of a semistable elliptic curve with probability*

$$\prod_p \left(1 - \frac{1}{p^2}\right) = \zeta(2)^{-1}.$$

Remark 5.6. *Note that:*

1. *Semistable is equivalent to not additive at any prime p.*
2. *Rescaling $a_i \mapsto p^i a_i$ gives non-minimal models. We have a fact (this will appear in [CS19]) that: A Weierstrass equation is minimal at p with local density $1 - \frac{1}{p^{10}}$.*

Hence we have:

Corollary 5.7. *A random elliptic curve over $\mathbb{Q}$ is semistable with probability*

$$\frac{\zeta(10)}{\zeta(2)} \approx 60\%$$

5.2 Squarefree Discriminants

We will use a bit of notation about Kodaira symbols. Locally at p we have either:

- Type I_0 $(\operatorname{ord} p(\Delta) = 0)$ is $(1 - \frac{1}{p})$
- Type I_1 $(\operatorname{ord} p(\Delta) = 1)$ is $(\frac{(p-1)^2}{p^3})$

$\mu(I_{m+1}) = \frac{1}{p}\mu(I_m)$.

$$\sum_{m \geq 1} \mu(I_m) = \mu(\text{mult. reduction}) = \frac{1}{p} - \frac{1}{p^2}$$

Hence locally

$$\begin{aligned}\mu(\Delta \text{ squarefree}) &= 1 - \frac{1}{p} + \frac{(p-1)^2}{p^3} \\ &= 1 - \frac{2}{p^2} + \frac{1}{p^3}\end{aligned}$$

Theorem 5.8 ([CS19]). *A random Weierstrass equation over $\mathbb{Z}$ has squarefree discriminant with probability*

$$\prod_p \left(1 - \frac{2}{p^2} + \frac{1}{p^3}\right)$$

References

[BCF+15] Manjul Bhargava, John E Cremona, Tom Fisher, Nick G Jones, and Jonathan P Keating. What is the probability that a random integral quadratic form in n variables has an integral zero? *International Mathematics Research Notices*, 2016(12):3828–3848, 2015.

[BCF16] Manjul Bhargava, John Cremona, and Tom Fisher. The proportion of plane cubic curves over $\mathbb{Q}$ that everywhere locally have a point. *International Journal of Number Theory*, 12(04):1077–1092, 2016.

[CS19] John Cremona and Mohammad Sadek. Densities for elliptic curves. *In Progress*, 2019.

[Eke91] Torsten Ekedahl. An infinite version of the chinese remainder theorem. *Comment. Math. Univ. St. Paul*, 40(1):53–59, 1991.

[PS99a] Bjorn Poonen and Michael Stoll. The cassels-tate pairing on polarized abelian varieties. *Annals of Mathematics*, 150(3):1109–1149, 1999.

[PS99b] Bjorn Poonen and Michael Stoll. A local-global principle for densities. In *Topics in number theory*, pages 241–244. Springer, 1999.

[PV04] Bjorn Poonen and José Felipe Voloch. Random diophantine equations. In *Arithmetic of higher-dimensional algebraic varieties*, pages 175–184. Springer, 2004.