LTCC Advanced Course: Arithmetic Statistics

Christopher Keyes

Updated March 11, 2025

Course description

Arithmetic statistics encompasses a broad range of quantitative problems involving the distribution of number theoretic objects. This course is an introduction to the area through the problem of counting number fields by discriminant. After reviewing some preliminaries from algebraic number theory, we will explore how to count extensions of small degree by discriminant with prescribed Galois group, with the goal of covering in detail Davenport and Heilbronn's famous results on S_3-cubic extensions. We will then touch on some of Bhargava's more recent innovations in this area and their applications, and get a sense of the current state of the art for higher degrees.

Related reading

The central topic of counting cubic extensions follow Davenport and Heilbronn's original paper [DH71] and Bhargava, Shankar, and Tsimerman's more recent paper [BST13]. Other authors have written and published good notes on these topics, including Achenjang [Ach] and Landesman [Lan]. There are excellent notes from the Arizona Winter School on related topics, including (but not limited to) Bhargava's 2009 notes on parametrizing rings [Bha09] and Wood's 2014 notes on arithmetic statistics [Woo14]. (Better still, you can watch these lecture recordings online.)

1 Introduction

Arithmetic statistics is composed of two words, which loosely have the following meanings.

- **Arithmetic** (adj): relating to number theory.
- **Statistics** (noun): the study of collecting and analyzing quantitative data.

Put together, **arithmetic statistics** encompasses questions about number theoretic objects — think primes, number fields, elliptic curves — that are quantitative in nature — think probability, averages, or distributions. Let's start with some examples.

1.1 Primes

Several classical questions about the primes can framed as statistical.

Question 1.1. *How many primes are there?*

Answer 1.2 (Easy)**.** There are infinitely many primes. The proof is an exercise. :)

To give a more precise answer that sheds some light on how the primes are distributed, we need a way to count primes.

Definition 1.3. Define the prime counting function

$$\pi(X) = \#\{1 \leq p \leq X : p \text{ is prime}\}.$$

Here, the choice of counting function is rather obvious. (Can you think of a different one?) Still, we have made a choice, and it's worth thinking about why it's a good one, and what other options are out there.

- For all positive real X, $\pi(X)$ is finite. This is a key property of a counting function.
- It is easy to count all the natural numbers in $[1, X]$, to then compare with $\pi(X)$. This will not be true for all such problems; indeed, a great deal of current research goes into just counting things like number fields or elliptic curves, before counting those which have certain properties.
- We could have chosen $\pi'(X)$ to count the primes in $[-X, X]$. This satisfies both bullet points above. Of course, we have $\pi'(X) = 2\pi(X)$, so we lose nothing by specializing to positive numbers. (One could argue that sticking to positive is better; this hints at how one might want to count primes in a number field...)

Let us briefly recall some asymptotic notation which we will need throughout the course.

Definition 1.4 (Asymptotic notation)**.** Assume all functions are real valued.

- $f(X) = O(g(X))$ means there is a constant c and a number N such that

  $$f(X) \leq cg(X) \text{ for all } X \geq N.$$

- $f(X) \ll g(X)$ is another common notation for $f(X) = O(g(X))$.
- $f(X) \sim g(X)$ means

  $$\lim_{X\to\infty} \frac{f(X)}{g(X)} = 1.$$

  This is quite a strong statement.
- $f(X) = o(g(X))$ means

  $$\lim_{X\to\infty} \frac{f(X)}{g(X)} = 0,$$

  i.e. that f grows slower than g asymptotically. In particular, if $f(X) = o(1)$ then $f(X) \to 0$.

With a counting function in hand, we can give a better answer to Question 1.1.

Theorem 1.5 (Prime number theorem, ∼1896)**.** *We have*

$$\pi(X) \sim \frac{X}{\log X}.$$

This long sought after theorem answers other statistical questions about the prime numbers. In particular, since $\pi(X) = o(X)$, we can see that the primes make up only "0%" of the natural numbers.

Example 1.6 (Naive attempt at proving PNT)**.** Let $\pi(X, X')$ denote the number of primes in $[X, X']$. To prove the prime number theorem, it is good enough to show $\pi(\sqrt{X}, X) \sim \frac{X}{\log X}$.

A number $n \in [\sqrt{X}, X]$ is prime if and only if for all primes $p \leq \sqrt{X}$ we have $p \nmid n$. One might be led to *guess* that the "probability" of $n \in [\sqrt{X}, X]$ being prime is equal to $\prod_{p\leq\sqrt{X}}(1-\frac{1}{p})$. This relies on a naive heuristic assumption that these divisibility conditions are independent, so the product agrees with the density of primes in $[\sqrt{X}, X]$.

However, Mertens' product theorem (which *predates* the prime number theorem) states that

$$\prod_{p \le X} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log X},$$

where $\gamma \approx 0.577$ is the Euler–Mascheroni constant. Thus our product grows like

$$\prod_{p \le \sqrt{X}} \left(1 - \frac{1}{p}\right) \sim \frac{e^{-\gamma}}{\log \sqrt{X}} = \frac{2e^{-\gamma}}{\log X},$$

and $2e^{-\gamma} \approx 1.123$.

Thus our naive independence heurstic suggests $\pi(\sqrt{X}, X) \sim 1.123X/\log X$, which of course is the wrong answer! This tells us that the distribution of primes in intervals like $[\sqrt{X}, X]$ is not quite so nice, so we should be careful about making such independence assumptions.

With something of a handle on counting primes, we'd like to understand how they are distributed, in various ways. One such way is to look at arithmetic progressions. Let a, b be coprime positive integers and define

$$\pi(X; a, b) = \{1 \le p \le X : p \text{ is prime and } p \equiv a \pmod{b}\}.$$

Theorem 1.7. *We have*

$$\pi(X; a, b) \sim \frac{\pi(X)}{\varphi(b)}$$

where φ is Euler's phi function.

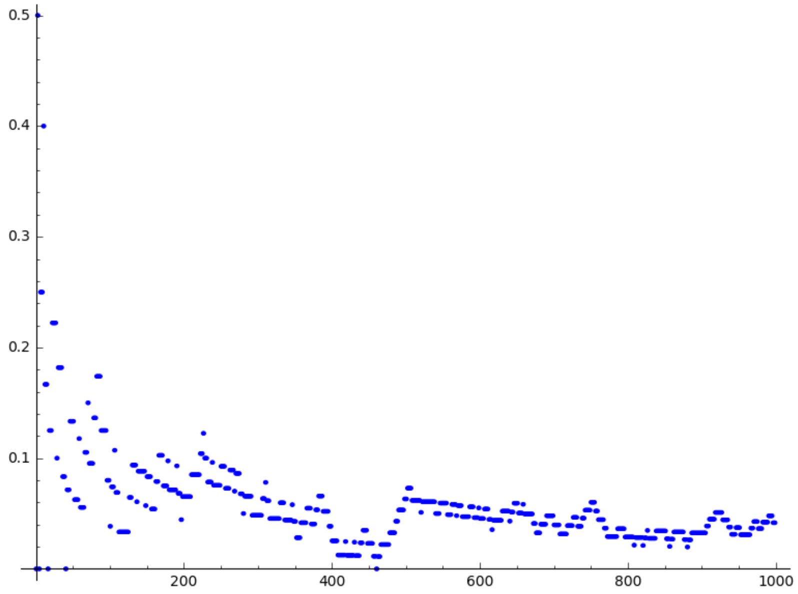
While we won't make this rigorous, this theorem may be interpreted as saying that the primes are *equidistributed* among residue classes of units modulo b. As a concrete example, if $b = 4$, then a(n odd) prime p is either 1 or 3 modulo 4, and each of these happens 50% of the time.

Remark 1.8 (Prime number races)**.** With that said, there are *biases* in the distribution of primes in arithmetic progressions. Let's continue looking at primes modulo 4.

If we plot

$$\frac{\pi(X; 3, 4) - \pi(X; 1, 4)}{\pi(X)}$$

for $X \le 1000$ we obtain the following.



Notice that this function is nonnegative; up to $X = 1000$, we never have *more* primes that are 1 modulo 4 than 3 modulo 4.

We have to go all the way to $X = 26861$ to find the first instance of this function dipping negative, and in fact it does so infinitely many times. However, it can be made precise that for 100% of X, the function is nonnegative. If you like this kind of thing, mess around with the computer (see the exercises), then search for "prime number races" or see this survey article [GM04].

1.2 Number fields

Let $\mathbb{Q}$ denote the field of rational numbers.

Definition 1.9. A **number field** is a finite extension $K/\mathbb{Q}$. Its **degree**, $[K : \mathbb{Q}]$, is equal to its dimension as a $\mathbb{Q}$-vector space.

Question 1.10. *For a given degree d, how many number fields are there?*

Like with the prime numbers, it's not so difficult to show that there are infinitely many number fields $K/\mathbb{Q}$ of degree $d > 1$. Again, we need to agree on a way to count them.

Definition 1.11 (Discriminant)**.** Let $\mathcal{O}_K$ be the ring of integers of K. Write $\mathcal{O}_K = \mathbb{Z}[\beta_i]$ as a free $\mathbb{Z}$-module, and let $\iota_j : K \to \overline{\mathbb{Q}}$ be the d embeddings of K into the algebraic closure. The **discriminant** of $K/\mathbb{Q}$ is

$$\mathrm{Disc}(K/\mathbb{Q}) = \det\left(\iota_j(\beta_i)\right)^2.$$

Another common equivalent definition uses the trace:

$$\mathrm{Disc}(K/\mathbb{Q}) = \det\left(\mathrm{Tr}(\beta_i\beta_j)\right).$$

The discriminant is a reasonable way to count number fields:

- It is actually a measure of size! Minkowski theory relates the discriminant to the volume of a fundamental domain for $\mathcal{O}_K$, viewed as a lattice in Minkowski space; see e.g. [Neu99, I.5].
- The number of $K/\mathbb{Q}$ of degree d with discriminant $|\mathrm{Disc}(K/\mathbb{Q})| \leq X$ is finite. This is known as Hermite's theorem, and its proof also uses Minkowski theory; see e.g. [Neu99, III.2].

- It is *isomorphism invariant*. This means that if we want to count *isomorphism classes* of number fields (which we will), we can still use it.
- The discriminant controls ramification: a prime p is ramified in K if and only if $p \mid \operatorname{Disc}(K/\mathbb{Q})$.

Remark 1.12. There are other invariants used to count number fields. One is called the **conductor**, a different measure of ramification for number fields. We won't talk about it further, but there is significant debate among researchers over what is the "right" invariant for these counting problems.

Definition 1.13. Let $N_n(X)$ denote the number of isomorphism classes of degree n extensions $K/\mathbb{Q}$ with discriminant $\operatorname{Disc}(K/\mathbb{Q}) \leq X$.

We can now state some results for low degrees. We will return to the $n = 2, 3$ cases in particular later in this course.

Theorem 1.14. *For $2 \leq n \leq 5$ there exists a constant c_n such that*

$$N_n(X) \sim c_n X.$$

Degree 2 is classical, closely related to the problem of counting squarefree numbers. The degree 3 case is due to Davenport and Heilbronn [DH71], and is much more involved. We will spend a good chunk of this course setting up and working through their original paper. The $n = 4, 5$ cases are celebrated work of Bhargava [Bha05, Bha10]. It is conjectured that similar results hold for all n.

Conjecture 1.15. *For all $n \geq 2$, there exists a constant c_n such that*

$$N_n(X) \sim c_n X.$$

We can also count number fields subject to certain conditions, much in the same way that we counted primes in arithmetic progressions earlier. A natural choice of condition is to specify the *Galois group*.

Definition 1.16 (G-extensions)**.** For $K/\mathbb{Q}$ of degree d, let $\widetilde{K}$ denote its Galois closure, i.e. the smallest Galois extension of $\mathbb{Q}$ containing K. If $\operatorname{Gal}(\widetilde{K}/\mathbb{Q}) \simeq G$, then we say $K/\mathbb{Q}$ is a **degree n G-extension** of $\mathbb{Q}$.

We can also define a counting function $N_n(X; G)$, which counts the number of such extensions with discriminant at most X.

Warning. When people say "count number fields with Galois group G," they (typically) don't actually mean that they are counting Galois extensions $K/\mathbb{Q}$ with $\operatorname{Gal}(K/\mathbb{Q}) = G$. They almost certainly mean counting G-extensions in the sense defined above.

We can now expand on Theorem 1.14.

Theorem 1.17. *In low degrees n and for certain G, we have asymptotics for $N_n(X; G)$.*

$$N_2(X; S_2) \sim \frac{X}{\zeta(2)} \tag{1.1}$$

$$N_3(X, C_3) \sim o(X) \tag{1.2}$$

$$N_3(X, S_3) \sim \frac{X}{3\zeta(3)} \qquad \textit{(Davenport–Heilbronn)} \tag{1.3}$$

$$N_4(X, D_4) \sim c_4(D_4)X \qquad \textit{(Cohen–Diaz y Diaz–Olivier)} \tag{1.4}$$

$$N_4(X, S_4) \sim c_4(S_4)X \qquad \textit{(Bhargava)} \tag{1.5}$$

$$N_5(X, S_5) \sim c_5(S_5)X \qquad \textit{(Bhargava)} \tag{1.6}$$

$$\tag{1.7}$$

More precise conjectures have been made for $N_n(X;G)$ in general.

Conjecture 1.18 (Malle, 2002)**.** *Fix a degree $n \geq 2$ and a transitive subgroup $G \subseteq S_n$. We have*

$$N_n(X;G) \sim c_n(G) X^{a(G)} (\log X)^{b(G)-1},$$

where $a(G)$ and $b(G)$ are invariants related to the group G, viewed as a permutation group on n elements.

Regrettably, the conjecture is false in general, with counterexamples given by Klüners [Klü05]. However, it is known to hold in several cases, including for small degrees as above, for abelian groups G, and for certain G arising from semidirect products of other groups where the conjecture is known. Very recent work of Loughran and Santens suggests an explicit description of $c_n(G)$ and a way to modify the conjecture to account for the known counterexamples; the introduction also lists many references if you are interested in further reading [LS24].

1.2.1 Local behavior

Let $K/\mathbb{Q}$ be *Galois* extension with ring of integers $\mathcal{O}_K$. Recall that $\mathcal{O}_K$ is a Dedekind domain, i.e. its ideals factor uniquely into prime ideals. If p is an integer prime, then

$$p\mathcal{O}_K = \left(\prod_{i=1}^{g} \mathfrak{p}_i \right)^e,$$

where the $\mathfrak{p}_i$ are distinct primes of $\mathcal{O}_K$. Let $\mathbb{F}_\mathfrak{p} = \mathcal{O}_K/\mathfrak{p}$ denote the residue field, which is finite. In the Galois case, we have $\mathbb{F}_{\mathfrak{p}_i} \simeq \mathbb{F}_{\mathfrak{p}_j}$, so we can safely denote them all by $\mathbb{F}_\mathfrak{p}$.

Let's quickly review some terms to describe prime splitting in number fields.

Definition 1.19 (Inertia and ramification)**.** The **inertia degree** of $\mathfrak{p} \mid p$ is the degree of $\mathbb{F}_\mathfrak{p}/\mathbb{F}_p$, denoted f.

The **ramification degree** of $\mathfrak{p} \mid p$ is the exponent e in the factorization. An important fact is

$$[K : \mathbb{Q}] = efg.$$

There are some important special cases:

- if $e = 1$ we say p is **unramified**;
- conversely, if $e > 1$, we say p is **ramified**, and if $g = 1$ then p is **totally ramified**;
- when $p\mathcal{O}_K$ itself is prime, we say p is **inert**, which coincides with $g = 1$, $e = 1$, and $f = [K : \mathbb{Q}]$;
- if p is unramified and $f = 1$ for all i, then we say p is **totally split**.

Definition 1.20 (Decomposition and inertia groups)**.** Let the **decomposition group**, $D_\mathfrak{p} \subset \mathrm{Gal}(K/\mathbb{Q})$ denote the subgroup of $\mathrm{Gal}(K/\mathbb{Q})$ consisting of automorphisms that fix $\mathfrak{p}$.

There is a natural exact sequence

$$0 \to I_\mathfrak{p} \to D_\mathfrak{p} \to \mathrm{Gal}(\mathbb{F}_\mathfrak{p}/\mathbb{F}_p) \to 0$$

where the kernel, $I_\mathfrak{p}$ is called the **inertia group**.

When $\mathfrak{p} \mid p$ is unramified, the inertia group vanishes, so $D_\mathfrak{p}$ is cyclic. In particular, the Frobenius map, which generates $\mathrm{Gal}(\mathbb{F}_\mathfrak{p}/\mathbb{F}_p)$ has a unique preimage in $D_\mathfrak{p}$. Moreover, since each $D_\mathfrak{p}$ is conjugate to $D_{\mathfrak{p}'}$ for $\mathfrak{p}, \mathfrak{p}' \mid p$, the respective Frobenius elements form a *conjugacy class* in $\mathrm{Gal}(K/\mathbb{Q})$.

Definition 1.21 (Frobenius conjugacy class)**.** For a prime p, the **Frobenius class** $\mathrm{Frob}_p \subset \mathrm{Gal}(K/\mathbb{Q})$ is this conjugacy class.

Example 1.22 (Quadratic fields)**.** Suppose $K/\mathbb{Q}$ is a quadratic field. Since $\mathrm{Gal}(K/\mathbb{Q}) \simeq C_2$ is abelian, a conjugacy class corresponds to a group element. If $p\mathcal{O}_K = \mathfrak{p}\overline{\mathfrak{p}}$ is split, Then $D_{\mathfrak{p}}, D_{\overline{\mathfrak{p}}}$ are both trivial, and Frob_p is the identity of C_2.

In the inert case, $\mathfrak{p} = p\mathcal{O}_K$, we have $D_{\mathfrak{p}} \simeq \mathrm{Gal}(\mathbb{F}_{\mathfrak{p}}/\mathbb{F}_p) \simeq C_2$, so its generator must also generate the Galois group.

Let's revisit the prime number theorem, but this time count only those primes with prescribed Frobenius class. Let

$$\pi(X; K, C) = \#\{1 \le p \le X : \mathrm{Frob}_p = C\},$$

for a fixed conjugacy class $C \subset \mathrm{Gal}(K/\mathbb{Q})$. Chebotarev's theorem is powerful generalization of Theorem 1.7.

Theorem 1.23 (Chebotarev Density Theorem)**.** *Let $K/\mathbb{Q}$ be a finite Galois extension. We have*

$$\pi(X; K, C) \sim \frac{\#C}{\#\,\mathrm{Gal}(K/\mathbb{Q})} \cdot \frac{X}{\log X}.$$

That is, the primes with a prescribed Frobenius class have density proportional to the size of Frob_p as a conjugacy class. This is in fact stronger than Theorem 1.7, and proving so is a good exercise.

1.2.2 Class groups

Let's get back to counting number fields, rather than primes, but this time count by their class groups.

Definition 1.24 (Class group)**.** Let $K/\mathbb{Q}$ be a number field. There is an exact sequence

$$1 \to K^\times/\mathcal{O}_K^\times \to \oplus_{\mathfrak{p}\text{ prime}}\mathbb{Z} \to \mathrm{Cl}(K) \to 0,$$

where the left hand map is div, which sends an element $x \in K^\times$ to the primes dividing its numerator minus those dividing its denominator (with multiplicity). The **ideal class group** is the quotient $\mathrm{Cl}(K)$.[1]

The class group measure the failure of $\mathcal{O}_K$ to be a unique factorization domain. Even in the simplest case of $K/\mathbb{Q}$ a quadratic field, they remain somewhat mysterious. Here are some things we know and don't know:

- there are finitely many *imaginary* quadratic fields with trivial $\mathrm{Cl}(K)$;
- in fact, as $\mathrm{Disc}(K/\mathbb{Q})$ grows, so does $\mathrm{Cl}(K)$ for an imaginary quadratic;
- it is predicted, but not known, that there are infinitely many *real* quadratic fields with trivial $\mathrm{Cl}(K)$.

Let's focus on imaginary quadratics $K/\mathbb{Q}$. One sensible question might be "how often is the order of $\mathrm{Cl}(K)$ divisible by a prime p?" Or, "what is the average size of $\mathrm{Cl}(K)[p]$?" The case of $p = 2$ is rather classical; see the exercises.

For odd p, we can make sense of the latter question via the Cohen–Lenstra heuristics.

[1] The middle group, $\oplus_{\mathfrak{p}}\mathbb{Z}$ should really be thought of as the group of fractional ideals in K.

Conjecture 1.25 (Cohen–Lenstra)**.** *Let p be an odd prime. Then the average p-torsion in the class group of an imaginary quadratic field is*

$$\lim_{X\to\infty} \frac{\sum_{K/\mathbb{Q} \textit{ im. quad.}} \# \mathrm{Cl}(K)[p]}{\sum_{K/\mathbb{Q} \textit{ im. quad.}} 1} = 2.$$

This conjecture, and many other statements about average class group behavior, are based on the following heuristic assumption: *class groups behave like random finite abelian groups A, weighted by $\frac{1}{\# \mathrm{Aut}(A)}$.* Conjecture 1.25 boils down to equating the average on the left-hand-side with

$$\frac{\sum_A \frac{\# A[p]}{\# \mathrm{Aut}(A)}}{\sum_A \frac{1}{\# \mathrm{Aut}(A)}},$$

where A runs over all finite abelian p-groups. In their seminal paper, Cohen and Lenstra actually computed this ratio to be 2, along with a number of other averages and probability results for finite abelian groups assuming this distribution.

Remark 1.26. Conjecture 1.25 is only known to hold for $p = 3$, due to Davenport and Heilbronn [DH71, Theorem 3]. We will come back to this later.

1.3 Elliptic curves

Definition 1.27 (Elliptic curve)**.** Let K be a field. An **elliptic curve** E/K is a smooth projective geometrically integral curve of genus 1.

Alternatively, and more explicitly, if $\mathrm{char}(K) \neq 2, 3$, an elliptic curve is isomorphic to a plane cubic curve given by a **short Weierstrass equation**

$$E\colon y^2 = x^3 + ax + b,$$

where $4a^3 + 27b^2 \neq 0$ (for smoothness).

A characteristic feature of an elliptic curve is that the K-points satisfy a *group law*. We write $E(K)$ for this group of points; if E is in short Weierstrass form, the identity is the point at infinity, $[0 : 1 : 0]$ in projective coordinates.

Theorem 1.28 (Mordell–Weil)**.** *Let K be a number field and E/K an elliptic curve. Then $E(K)$ is a finitely generated abelian group,*

$$E(K) \simeq \mathbb{Z}^r + T,$$

for $r \geq 0$ and T a finite abelian group.

The integer r is known as the **rank** of E/K, and T is known as the **torsion subgroup** of $E(K)$.

There are lots of statistical questions we can ask about elliptic curves and their Mordell–Weil groups.

- As $E/\mathbb{Q}$ varies, what groups T arise as torsion subgroups? In the 1970s, Mazur proved a longstanding conjecture that exactly 16 groups T can show up, each one infinitely often. Similar questions have been studied for different base fields.
- As $E/\mathbb{Q}$ varies, what ranks r arise? This is not only an open question, but one that divides our community: some believe that the ranks of $E/\mathbb{Q}$ is bounded, while others believe they are not. The largest known rank is at least 29, for a curve $E/\mathbb{Q}$ discovered in 2024 (last year!) by Elkies and Klagsbrun.

- For a fixed $E/\mathbb{Q}$, how does $E(K)$ vary as K varies? In particular, when is $E(K) \supsetneq E(\mathbb{Q})$? Or, how often does the rank change when going from $E(\mathbb{Q})$ to $E(K)$? These are areas of ongoing research, and can be interpreted as counting number fields $K/\mathbb{Q}$ coming from E in particular ways.

Let's think a little more about ranks. Even though the maximum is not known, we could try to access other statistical statements about them, such as the *average*. As always, we'll need a reasonable way to count elliptic curves.

Definition 1.29 (Naive height)**.** Let $E/\mathbb{Q}$ be an elliptic curve given by

$$E\colon y^2 = x^3 + ax + b.$$

The **naive height** is given by

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

This height satisfies some of the counting properties we like, namely that there are finitely many E with $\mathrm{ht}(E) \leq X$. However, notice that it fails to be an isomorphism invariant.

Example 1.30. Let

$$\begin{aligned} E\colon y^2 &= x^3 + x + 1 \\ E'\colon y^2 &= x^3 + 16x + 64. \end{aligned}$$

There is an isomorphism $E \to E'$ defined on points by $(x, y) \mapsto (4x, 8y)$. The naive heights of E, E', respectively, are 27 and 110592.

While this is a little inconvenient, we can get around it by recognizing that the problem here is that we have a prime p such that $p^4 \mid a$ and $p^6 \mid b$. As long as we impose that there is no such p, then the short Weierstrass form of $E/\mathbb{Q}$, and hence its naive height, are unique.

Remark 1.31 (Other heights)**.** As always, there are other notions of height. Namely, we could've taken the discriminant of $x^3 + ax + b$ or the conductor, an invariant which measures how degenerate the curve becomes when reduced modulo primes p.

There are also theories for counting points on *weighted* spaces, which captures what's going on in the example above. In some sense, these height functions might be considered more natural, since they from the geometry of the space of elliptic curves themselves.

Theorem 1.32 (Bhargava–Shankar [BS15])**.** *We have*

$$\limsup_{X\to\infty} \frac{\sum_{\mathrm{ht}(E)\leq X} \mathrm{rk}(E)}{\sum_{\mathrm{ht}(E)\leq X} 1} \leq 1.5.$$

In other words, the average rank of an elliptic curve over $\mathbb{Q}$, when counted by naive height, is at most 1.5.

This remarkable result both improved upon previous results and removed conditions on the Generalized Riemann Hypothesis and Birch and Swinnerton-Dyer Conjecture. The proof actually involves bounding the average size of the 2-Selmer group of $E/\mathbb{Q}$. To do this, they associate to each element of the 2-Selmer group a binary quartic form (up to some group action), and then count orbits quartic forms. We won't have time to cover this further, but the basic philosophy has a lot in common with our approach to cubic fields.

2 Quadratic fields and squarefree numbers

As something of a warmup, let's carefully prove (1.1),

$$N_2(X) \sim \frac{X}{\zeta(2)}.$$

We will need to borrow some standard facts from analytic number theory, which we will try to keep to a minimum.

2.1 Möbius inversion

We will only need two tools: the Möbius function and Dirichlet convolution.

Definition 2.1 (Möbius function)**.** The **Möbius function**, denoted $\mu(n)$ is defined on $n \in \mathbb{Z}$ by

$$\mu(n) = \begin{cases} (-1)^{\omega(n)} & n \text{ squarefree,} \\ 0 & n \text{ not squarefree,} \end{cases}$$

where $\omega(n)$ is the number of prime factors of n.

Why is this function useful? It tends to come up in inclusion-exclusion arguments involving primes. It also appears in several handy identities that relate arithmetic functions.

The first such formula states that if f, g are arithmetic functions,

$$g(n) = \sum_{d|n} f(d) \implies f(n) = \sum_{d|n} \mu(d) g\left(\frac{n}{d}\right).$$

Definition 2.2. Let f, g be arithmetic functions. The **Dirichlet convolution**, denoted $f * g$, is an arithmetic function given by

$$(f * g)(n) = \sum_{ab=n} f(a)g(b) = \sum_{d|n} f(d)g(n/d).$$

Analytic number theorists love Dirichlet series. Given this fact, you can see why they also love convolution by observing the following identity[2]

$$\left(\sum_n \frac{f(n)}{n^s}\right)\left(\sum_n \frac{g(n)}{n^s}\right) = \left(\sum_n \frac{(f*g)(n)}{n^s}\right). \tag{2.1}$$

Simply expanding the product on the left reveals the identity.

Let's define two auxiliary arithmetic functions.

$$\mathbb{1}(n) = 1 \text{ for all } n, \text{and}$$
$$\varepsilon(n) = \begin{cases} 1 & n = 1, \\ 0 & \text{otherwise.} \end{cases}$$

We claim that

$$\mathbb{1} * \mu = \varepsilon.$$

This is easy to verify by unwinding the definitions. The key point is that $\sum_{d|n} \mu(d) = 0$ unless $n = 1$.

[2] Let's not worry about convergence, though in general one probably should. Just think of this as a formal identification for now.

We'd like to apply (2.1) to this identity. Indeed, the Dirichlet series for $\mathbb{1}$ is the Riemann ζ function! This provides that

$$\sum_n \frac{\mu(n)}{n^s} = \frac{1}{\zeta(s)}. \tag{2.2}$$

This is the bare minimum we need to prove things about squarefree numbers and quadratic fields.

2.2 Squarefree numbers

Let $\mathrm{SF}(X)$ denote the counting function

$$\mathrm{SF}(X) = \#\{1 \leq n \leq X : n \text{ squarefree}\}.$$

Our immediate goal is to prove the following classical result.

Proposition 2.3. *We have*

$$\mathrm{SF}(X) \sim \frac{X}{\zeta(2)} = \frac{6}{\pi^2} X.$$

Proof. Observe that

$$\mathrm{SF}(X) = \sum_{n \leq X} \sum_{d^2 | n} \mu(d),$$

since if n is squarefree, the inner sum is just 1 (coming from $d = 1$), and if not, the inner sum is over all factors of the square part of n, and it vanishes. Switching order of summation and collecting terms, we have

$$\begin{aligned} \sum_{n \leq X} \sum_{d^2 | n} \mu(d) &= \sum_{d \leq \sqrt{X}} \mu(d) \sum_{\substack{n \leq X \\ d^2 | n}} 1 \\ &= \sum_{d \leq \sqrt{X}} \mu(d) \left\lfloor \frac{X}{d^2} \right\rfloor \\ &= X \sum_{d \leq \sqrt{X}} \frac{\mu(d)}{d^2} + O(\sqrt{X}). \end{aligned}$$

We now compare $\sum_{d \leq X} \frac{\mu(d)}{d^2}$ to the Dirichlet series:

$$\sum_{d \leq \sqrt{X}} \frac{\mu(d)}{d^2} = \sum_d \frac{\mu(d)}{d^2} - \sum_{d > \sqrt{X}} \frac{\mu(d)}{d^2}.$$

The first term on the right-hand-side is $\frac{1}{\zeta(2)}$ by (2.2), while the second term is seen to be $O(1/\sqrt{X})$, by estimating $|\mu(d)| \leq 1$ and taking the integral. This gives the result. □

Notice that we actually showed $\mathrm{SF}(X) = X/\zeta(2) + O(\sqrt{X})$, which is stronger than the statement! If you're the sort of person that really cares about sharpening error terms, then arithmetic statistics would make a wonderful playground. In this course, we'll be focusing on the main terms for the most part.

We actually will want to know something stronger than Proposition 2.3. How are the squarefree numbers distributed in residue classes modulo n? For this we define

$$\mathrm{SF}(X; a, b) = \#\{1 \leq n \leq X : n \text{ squarefree and } n \equiv a \pmod{b}\}.$$

Example 2.4 (Squarefrees modulo 4)**.** If we compute the first million integers, we count 607926 are squarefree. Of these, the counts of those which are 1, 2, or 3 modulo 4 are 202636, 202640, and 202650, respectively. These are all pretty close, lending credence to the possibility that $\mathrm{SF}(X; a, 4) \sim X/3\zeta(2)$ for $a \in \{1, 2, 3\}$.

Let's try to prove this for $b = p^2$ by modifying our previous proof of Proposition 2.3.

Proposition 2.5. *Let p be a prime and a such that $a \not\equiv 0 \pmod{p^2}$. Then we have*

$$\mathrm{SF}(X; a, p^2) = \frac{X}{(p^2-1)\zeta(2)} + O(\sqrt{X}).$$

Proof. Following our previous argument, we have

$$\begin{aligned}
\mathrm{SF}(X; a, p^2) &= \sum_{\substack{n \leq X \\ n \equiv a\,(p^2)}} \sum_{d^2 \mid n} \mu(d) \\
&= \sum_{d \leq \sqrt{X}} \mu(d) \sum_{\substack{n \leq X \\ d^2 \mid n \\ n \equiv a\,(p^2)}} 1 \\
&= \frac{X}{p^2} \sum_{\substack{d \leq \sqrt{X} \\ \gcd(d,p)=1}} \frac{\mu(d)}{d^2} + O(\sqrt{X}) \qquad (2.3)
\end{aligned}$$

Note the factor of $1/p^2$ that has appeared. We also see that the sum over d is restricted to d coprime to p; since $a \not\equiv 0 \pmod{p^2}$, those terms with $p \mid d$ contribute nothing.

Continuing with our previous approach, we want to compare $\sum_{\substack{d \leq \sqrt{X} \\ \gcd(d,p)=1}} \frac{\mu(d)}{d^2}$ to $1/\zeta(2)$. For this we see

$$\begin{aligned}
\sum_{d} \frac{\mu(d)}{d^s} &= \sum_{p \nmid d} \frac{\mu(d)}{d^s} + \sum_{p \mid d} \frac{\mu(d)}{d^s} \\
&= \sum_{p \nmid d} \frac{\mu(d)}{d^s} + \sum_{d'} \frac{\mu(pd')}{(pd')^s} \\
&= \sum_{p \nmid d} \frac{\mu(d)}{d^s} + \sum_{p \nmid d'} \frac{\mu(pd')}{(pd')^s} && (\text{if } p \mid d' \text{ then } \mu(pd') = 0) \\
&= \sum_{p \nmid d} \frac{\mu(d)}{d^s} - \frac{1}{p^s} \sum_{p \nmid d'} \frac{\mu(d')}{d'^s} && (\mu(pd') = \mu(p)\mu(d') \text{ for } p \nmid d') \\
&= \left(1 - \frac{1}{p^s}\right) \sum_{p \nmid d} \frac{\mu(d)}{d^s}.
\end{aligned}$$

This last line yields, with $s = 2$,

$$\sum_{p \nmid d} \frac{\mu(d)}{d^2} = \frac{p^2}{p^2-1} \sum_{d} \frac{\mu(d)}{d^2}. \qquad (2.4)$$

Combining (2.3) and (2.4), we have

$$\mathrm{SF}(X; a, p^2) = \frac{X}{p^2} \left(\frac{p^2}{p^2-1} \sum_{d} \frac{\mu(d)}{d^2} - \sum_{\substack{d > \sqrt{X} \\ p \nmid d}} \frac{\mu(d)}{d^2} \right) + O(\sqrt{X}).$$

The sum over $d > \sqrt{X}$ term is estimated the same way as in Proposition 2.3. This leaves the main term of $X/(p^2-1)\zeta(2)$, as desired. □

Proposition 2.5 illustrates an idea that comes back again and again in these sort of counting problems. We often reduce problems we care about to counting (things like) integers, but we need this count to still behave well when we impose extra *local conditions*. In the case of squarefree numbers, our count does indeed behave well: squarefree numbers are equidistributed among residue classes mod p^2.[3]

2.3 Quadratic fields

We now turn our attention to the $n = 2$ case of Theorem 1.17. For this, we make use of the characterization of quadratic fields. This is well known; see Cox's book [Cox22] for a great reference for all things quadratic fields.

Proposition 2.6 (Characterization of quadratic fields)**.** *There is a bijection*

$$\{\textit{Squarefree numbers } B \in \mathbb{Z}\} \to \{\textit{Quadratic extensions } K/\mathbb{Q}\}\,/\textit{iso.}$$

Given a squarefree B, its image is (the class of) the extension $K = \mathbb{Q}(\sqrt{B})$, which has:

- $\mathcal{O}_{\mathbb{Q}(\sqrt{B})} \simeq \begin{cases} \mathbb{Z}\left[\frac{1+\sqrt{B}}{2}\right] & B \equiv 1 \pmod 4, \\ \mathbb{Z}[\sqrt{B}] & B \equiv 2,3 \pmod 4. \end{cases}$
- $\operatorname{Disc}(\mathbb{Q}(\sqrt{B})) = \begin{cases} B & B \equiv 1 \pmod 4, \\ 4B & B \equiv 2,3 \pmod 4. \end{cases}$

Proof of Theorem 1.17 ($n = 2$ case). We have

$$N_2(X) = 2\left(\mathrm{SF}(X;1,4) + \mathrm{SF}(X/4;2,4) + \mathrm{SF}(X/4;3,4)\right),$$

with the factor of 2 accounting for the fact that $\mathrm{SF}(X)$ only counts positive squarefree integers. By Proposition 2.5, this becomes

$$\frac{2X}{\zeta(2)}\left(\frac{1}{3} + \frac{1}{12} + \frac{1}{12}\right) = \frac{X}{\zeta(2)}.$$

Thus despite some discrepancy in the discriminant, we actually get the same answer as counting squarefree numbers. □

3 Parametrizing rings ranks 2 and 3

A great reference for this material is [Bha09].

Definition 3.1 (rank n ring)**.** A **rank n ring** is a ring R which is isomorphic to $\mathbb{Z}^n$ as a $\mathbb{Z}$-module.

Example 3.2 (Quadratic rings)**.** Let $n = 2$. Here are some examples of quadratic rings:

- If $K/\mathbb{Q}$ is a quadratic field, then $\mathcal{O}_K$ is a quadratic ring. As a $\mathbb{Z}$-module, it is $\mathbb{Z} + \omega\mathbb{Z}$, where ω might be $\sqrt{B}$ or $\frac{1+\sqrt{B}}{2}$.
- The ring $\mathbb{Z} \oplus \mathbb{Z}$ is a quadratic ring. However, it is degenerate in some sense, being the direct product of degree 1 rings.

[3] At least, those for which it's possible to be squarefree!

- Fix $K/\mathbb{Q}$ a quadratic field and let $f > 1$ be a positive integer. Write $\mathcal{O}_K = \mathbb{Z} + \omega\mathbb{Z}$ as above and let $\mathcal{O} = \mathbb{Z} + f\omega\mathbb{Z}$ be the order of conductor f in $\mathcal{O}_K$. This is easily checked to be a ring.
- Quadratic rings can have nilpotents, e.g. $\mathbb{Z}[x]/(x^2)$.

Definition 3.3. A rank n ring is **maximal** if it is not contained in another rank n ring.

3.1 Quadratic rings

Let's set the stage by classifying quadratic rings. Suppose $R = \mathbb{Z}+\omega\mathbb{Z}$. What could its ring structure look like? We understand the addition structure, since R is a free $\mathbb{Z}$-module. For multiplication, we have

$$(a + b\omega)(c + d\omega) = ac + (ad + bc)\omega + bd\omega^2.$$

The ac and $(ad + bc)\omega$ terms make sense. What about the $bd\omega^2$ term? We know $\omega^2 \in R$, so it must be that

$$\omega^2 = \tau + \theta\omega.$$

The values of $\tau, \theta \in \mathbb{Z}$ control the multiplication in R, and hence the ring structure! Another way to put this is that the matrix representing multiplication by ω,

$$\begin{pmatrix} 0 & \tau \\ 1 & \theta \end{pmatrix},$$

controls the ring R.

Remark 3.4. Note that ω, τ, θ are not unique. For example, $\mathbb{Z}[\omega + 1]$ or $\mathbb{Z}[-\omega]$ give the same ring as $\mathbb{Z}[\omega]$. In fact, if $\omega' = \omega + 1$ for instance, then the associated invariants are $\tau' = \tau + 1$, $\theta' = \theta + 2$. In particular, this allows us to always take $\theta \in \{0, 1\}$.

Proposition 3.5. *The set of quadratic rings $\mathbb{Z}[\omega]$, up to isomorphism, is in bijection with the set of integers $\{D \equiv 0, 1 \pmod 4\}$.*

Proof. The map is the discriminant,

$$\mathrm{Disc}(\mathbb{Z}[\omega]) = \det \begin{pmatrix} \mathrm{Tr}(1) & \mathrm{Tr}(\omega) \\ \mathrm{Tr}(\omega) & \mathrm{Tr}(\omega^2) \end{pmatrix}.$$

By definition, $\mathrm{Tr}(\omega) = \theta$, and by linearity we have $\mathrm{Tr}(\omega^2) = \mathrm{Tr}(\tau) + \mathrm{Tr}(\theta\omega) = 2\tau + \theta^2$, so

$$\mathrm{Disc}(\mathbb{Z}[\omega]) = \det \begin{pmatrix} 2 & \theta \\ \theta & 2\tau + \theta^2 \end{pmatrix} = 4\tau + \theta^2.$$

We'll leave showing that the discriminant is an isomorphism class invariant to the exercises. □

A different approach to counting quadratic fields is to instead count quadratic *rings*, then sieve out those which are not integral domains and those which are not maximal. Indeed, this is the approach we will take for cubic fields. Here's what it might look like in the quadratic case.

Lemma 3.6. *Let R_D be the quadratic ring of discriminant D. In light of Proposition 3.5, we can write*

$$D = \begin{cases} 4D' & D \equiv 0 \pmod 4, \\ D' & D \equiv 1 \pmod 4. \end{cases}$$

Then we can infer the following about R_D from D'.

(i) R_D is an integral domain if and only if D' is not an integer square.

(ii) R_D *is maximal if and only if* D' *is squarefree.*

Proof. See the exercises. □

To count quadratic fields, we need to count *maximal* quadratic rings R_D which are *integral domains.*

- Proposition 3.5 tells us that there are X quadratic rings R_D with $|\operatorname{Disc}(R_D)| \leq X$.
- Lemma 3.6(i) tells us that $O(\sqrt{X})$ of those R_D are not integral domains.
- We can then sieve out those D' which are not squarefree — by repeating the arguments in the previous section — to obtain the count of maximal quadratic rings.

3.2 Cubic rings

We will parameterize cubic rings in terms of **integral binary cubic forms**,

$$f(x,y) = ax^3 + bx^2y + cxy^2 + dy^3$$

where $a,b,c,d \in \mathbb{Z}$. There is a natural twisted action of $\operatorname{GL}_2(\mathbb{Z})$ on the space of such forms, given by

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \cdot f(x,y) = \det \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}^{-1} f(\alpha x + \beta y, \gamma x + \delta y).$$

It turns out that cubic rings are parameterized by orbits of cubic forms under this action, in a way that preserves their discriminant.

Proposition 3.7. *There is a bijection*

$$\{\textit{cubic rings } R\}/\textit{iso} \to \{\textit{binary cubic forms } f\}/\operatorname{GL}_2(\mathbb{Z}).$$

If R_f *denotes a cubic ring mapping to a cubic form* $f(x,y)$*, then moreover we have*

$$\operatorname{Disc}(R_f) = \operatorname{Disc}(f) = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd.$$

Proof. We follow the exposition in [BST13]. Let $R = \mathbb{Z}[\omega, \theta]$. After possibly translating ω, θ by integers, we have $\omega\theta \in \mathbb{Z}$. Write

$$\begin{aligned} \omega\theta &= n, \\ \omega^2 &= m - b\omega + a\theta, \\ \theta^2 &= \ell - d\omega + c\theta, \end{aligned} \tag{3.1}$$

for integers n, m, ℓ, a, b, c, d. We define the correspondence by

$$R \mapsto f(x,y) = ax^3 + bx^2y + cxy^2 + dy^3.$$

To see this map is well defined, let's introduce a more conceptual perspective for where this map comes from. We have $R/\mathbb{Z} \simeq \mathbb{Z}^2$, with basis $\{\omega, \theta\}$. Define a map

$$\begin{aligned} R/\mathbb{Z} &\xrightarrow{\phi} \wedge^2(R/\mathbb{Z}) \\ r &\mapsto r \wedge r^2. \end{aligned}$$

Note that $\wedge^2 R/\mathbb{Z} \simeq \mathbb{Z}$, generated by the symbol $\omega \wedge \theta = -\theta \wedge \omega$. Using our formulae for ω^2 and θ^2 above, an explicit calculation shows

$$\phi(x\omega + y\theta) = f(x,y)(\omega \wedge \theta).$$

If $\{\omega', \theta'\}$ is another basis for $R/\mathbb{Z}$ (we should think of this as coming from an isomorphic cubic ring R') then in $R/\mathbb{Z}$ we have

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} \omega \\ \theta \end{pmatrix} = \begin{pmatrix} \omega' \\ \theta' \end{pmatrix} \text{ for } \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Z}).$$

Moreover, we see that

$$\phi(x\omega' + y\theta') = \phi((x\alpha + y\beta)\omega + (x\gamma + y\delta)\theta) = f(x\alpha + y\beta, x\gamma + y\delta)(\omega \wedge \theta).$$

The extra factor of $\det \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ is added to make this compatible with the natural isomorphism $\wedge^2 R/\mathbb{Z} \to \wedge^2 R/\mathbb{Z}$ coming from the change of basis.

Finally, we construct an inverse to the correspondence. Given a binary cubic form $f(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$, we can construct a candidate $R = \mathbb{Z}[\omega, \theta]$ using the relations above; we just need to come up with n, m, ℓ values. Since we want R to be associative, we need $\omega^2\theta = \omega^2 \cdot \theta = \omega \cdot \omega\theta$. Flushing this through the equations defining multiplication in R, we find $n = -ad$. A similar argument with $\omega\theta^2$ yields $m = -ac$, $\ell = -bd$. If we took f, f' in the same $\mathrm{GL}_2(\mathbb{Z})$-equivalence class, then we would find using the methods above that $R_f/\mathbb{Z} \simeq R_{f'}/\mathbb{Z}$.

The calculation that $\mathrm{Disc}(R_f) = \mathrm{Disc}(f)$ is left as an exercise. □

With this characterization, we would like to read off information about R_f from $f(x, y)$. Namely, we'd like to be able to tell when R_f is an integral domain, and when it is a maximal cubic ring.

Lemma 3.8 (See [BST13, Prop. 11]). *R_f is an integral domain if and only if f is irreducible.*

Proof. If f is reducible, then after a change of coordinates, we may assume $y \mid f$, so $a = 0$. Constructing R_f as in the proof of Proposition 3.7, we have $\omega\theta = n = -ad = 0$, so R_f is not an integral domain.

Conversely, suppose R has zero divisors, $\alpha\beta = 0$ for $\alpha, \beta \neq 0$. Multiplying the characteristic cubic equation for α by β, we have

$$\beta(\alpha^3 + c_1\alpha^2 + c_2\alpha + c_3) = 0 \implies c_3 = 0.$$

Thus $\alpha(\alpha^2 + c_1\alpha + c_2) = 0$. If $\alpha^2 + c_1\alpha + c_2 = 0$, then set $\omega = \alpha$. Otherwise, set $\omega = \alpha^2 + c_1\alpha + c_2$ and observe $\omega^2 = c_2\omega$. In both cases, we can extend $\{1, \omega\}$ to a basis $\{1, \omega, \theta\}$ for R. From our computation of ω^2, we see that the associated form $f(x, y)$ has $a = 0$, and so it is reducible. □

For maximality, we recognize that a cubic ring R is maximal if and only if it is either

- $R = \mathcal{O}_K$ for a cubic number field K,
- $R = \mathcal{O}_K \times \mathbb{Z}$ for a quadratic number field K, or
- $R = \mathbb{Z}^3$.

That is, R is maximal if and only if it is the product of rings of integers of number fields.

Definition 3.9. Let p be a prime. A cubic ring R is **maximal at p** if $R_p = R \otimes \mathbb{Z}_p$ is maximal as a cubic $\mathbb{Z}_p$-algebra.

By the characterization above, R is maximal if and only if it is maximal at p for all primes p. Later, we will need to know how to see whether R_f is maximal at p by looking at f.

Lemma 3.10 (See [BST13, Lemma 13]). *Let R be a cubic ring. R is not maximal at p if there is a basis $\{1, \omega, \theta\}$ such that one of the following holds:*

- $\mathbb{Z}[\omega/p, \theta]$ *is a ring;*
- $\mathbb{Z}[\omega/p, \theta/p]$ *is a ring.*

Proof. Suppose $R \subset R'$ for a cubic ring R' and $p \mid [R' : R]$. After possibly replacing R' by $R' \cap (R \otimes_{\mathbb{Z}} \mathbb{Z}[1/p])$, we may assume that the index $[R' : R]$ is a (nonzero) power of p. This means that as abelian groups, we have $(R'/\mathbb{Z})/(R/\mathbb{Z}) \simeq \mathbb{Z}/p^i\mathbb{Z} \times \mathbb{Z}/p^j\mathbb{Z}$. Equivalently, $R' = \mathbb{Z}[\omega/p^i, \theta/p^j]$ for $R = \mathbb{Z}[\omega, \theta]$. We may also safely assume that $\{1, \omega/p^i, \theta/p^j\}$ is a normalized basis, i.e. $\omega\theta/p^{i+j} \in \mathbb{Z}$.

After possibly swapping basis elements, we may assume $i \geq j$ (and not both zero). If $i = 1$, then we are done. If both $i, j \geq 1$, then we argue that $\mathbb{Z}[\omega/p^{i-1}, \theta/p^{j-1}]$ is a ring. To see this, it suffices to check it contains $(\omega/p^{i-1})^2$ and $(\theta/p^{j-1})^2$. But

$$\left(\frac{\omega}{p^{i-1}}\right)^2 = p^2\left(\frac{\omega}{p^i}\right) = p^2 m - pb\frac{\omega}{p^{i-1}} + pa\frac{\theta}{p^{j-1}} \in \mathbb{Z}[\omega/p^{i-1}, \theta/p^{j-1}],$$

where m, a, b are integers coming from the multiplication structure of R'. A similar calculation shows $(\theta/p^{j-1})^2$ is in the ring. The same argument in the $j = 0$ case shows that $\mathbb{Z}[\omega/p^{i-1}, \theta]$ is a ring.

Iterating this process as necessary, we eventually produce R' of one of the forms desired in the statement of the lemma. □

What does this tell us about $f(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$?

- If $R' = \mathbb{Z}[\omega/p, \theta]$ is a ring, with multiplication given by (3.1) with constants $n', m', \ell', a', b', c', d'$, then for R we have

  $$\omega^2 = p^2(\omega/p)^2 = p^2 m' - pb'\omega + p^2 a'\theta,$$
  $$\theta^2 = \ell - d\omega + c\theta.$$

  That is, we have $p^2 \mid a$ and $p \mid b$.

- If $R' = \mathbb{Z}[\omega/p, \theta/p]$ then a similar argument shows

  $$\omega^2 = p^2(\omega/p)^2 = p^2 m' - pb'\omega + pa'\theta,$$
  $$\theta^2 = p^2(\theta/p)^2 = p^2\ell' - pd'\omega + pc'\theta.$$

  In this case, $p \mid a, b, c, d$.

Definition 3.11. Let $\mathcal{U}_p$ denote the subset of integral binary cubic forms $f(x, y)$ such that there does *not* exist $M \in \mathrm{GL}_2(\mathbb{Z})$ such that $M \cdot f$ satisfies one of these two conditions.

This definition, combined with our characterization of maximality and the correspondence yields the following corollary.

Corollary 3.12. *A cubic ring R is maximal if and only if it corresponds to a cubic form f such that $f \in \mathcal{U}_p$ for all primes p.*

Remark 3.13. One can also calculate the discriminant in these two cases. We have

$$p^2 \operatorname{Disc}(\mathbb{Z}[\omega/p, \theta]) = \operatorname{Disc}(\mathbb{Z}[\omega, \theta]),$$
$$p^4 \operatorname{Disc}(\mathbb{Z}[\omega/p, \theta/p]) = \operatorname{Disc}(\mathbb{Z}[\omega, \theta]).$$

If $R \subset R'$ for a maximal cubic ring R', then $\operatorname{Disc}(R) = f^2 \operatorname{Disc}(R')$ for an integer f; this follows from successive applications of Lemma 3.8 and the calculations above.

3.2.1 The Hessian covariant

Let $f(x,y)$ be a cubic form as usual.

Definition 3.14 (Hessian)**.** The **Hessian covariant** of $f(x,y)$ is a quadratic form

$$\begin{aligned} H_f(x,y) &= \frac{1}{4}\det\begin{pmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{xy}(x,y) & f_{yy}(x,y) \end{pmatrix} \\ &= (3ac-b^2)x^2 + (9ad-bc)xy + (3bd-c^2)y^2, \end{aligned}$$

where $f_{xx} = \frac{\partial^2 f}{\partial x^2}$, etc.

It is a straightforward, if somewhat tedious, exercise in the chain rule to check that the action of $\mathrm{GL}_2(\mathbb{Z})$ on cubic forms is compatible with taking Hessians. By this we mean, if $M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}$ then

$$\begin{aligned} H_{M\cdot f}(x,y) &= \frac{1}{4(\det M)^2}\det\begin{pmatrix} \frac{\partial^2}{\partial x^2}(f(M(x,y)^T)) & \frac{\partial^2}{\partial x \partial y}(f(M(x,y)^T)) \\ \frac{\partial^2}{\partial x \partial y}(f(M(x,y)^T)) & \frac{\partial^2}{\partial y^2}(f(M(x,y)^T)) \end{pmatrix} \\ &= \frac{1}{4}\det\begin{pmatrix} f_{xx}(x,y) & f_{xy}(x,y) \\ f_{xy}(x,y) & f_{yy}(x,y) \end{pmatrix} \\ &= M \cdot H_f(x,y). \end{aligned}$$

3.2.2 The quadratic resolvent

Let R be a cubic ring. The discriminant satisfies $\mathrm{Disc}(R) \equiv 0, 1 \pmod 4$ (see Exercise 3.1). By Proposition 3.5, there is a *unique* quadratic ring S of discriminant $\mathrm{Disc}(R)$.

Definition 3.15 (quadratic resolvent)**.** If R is a cubic ring, its **quadratic resolvent** is the unique quadratic ring S such that $\mathrm{Disc}(R) = \mathrm{Disc}(S)$.

Suppose R is in order in K, a cubic field. Recall the discriminant of an element $x \in R$ is given by

$$\mathrm{Disc}(x) = (x-x')^2(x'-x'')^2(x''-x)^2,$$

where x, x', x'' are the conjugates of x in the Galois closure of $K/\mathbb{Q}$. This allows us to define a map

$$\begin{aligned} f\colon R &\to S \\ x &\mapsto \frac{\mathrm{Disc}(x) + \sqrt{\mathrm{Disc}(x)}}{2}. \end{aligned}$$

The discriminant of $f(x) \in S$ is the same as that of $x \in R$. This map can be seen to give the same binary form as in the correspondence of Proposition 3.7, hence the notation "f"; see the exercises if you are interested in this point.

3.2.3 Cubic forms over $\mathbb{R}$

What does this correspondence look like over $\mathbb{R}$? A real cubic form $f \in \mathbb{R}[x,y]$ with no repeated roots has either 1 or 3 real roots. Any form with 1 real root (resp. 3 real roots) can be transformed into another via the action of $\mathrm{GL}_2(\mathbb{R})$.

Thus there are only two nondegenerate real cubic rings: $\mathbb{R} \oplus \mathbb{C}$ and $\mathbb{R} \oplus \mathbb{R} \oplus \mathbb{R}$. The former has automorphism group C_2, while the latter has automorphism group S_3.

3.3 Higher rank

Like the quadratic resolvent of a cubic ring, quartic rings have a cubic resolvent, while quintic rings have a sextic resolvent. These are used to give parametrizations of rings of ranks 4 and 5, in terms of pairs of ternary quadratic forms up to equivalence and quadruples of alternating quinary quadratic forms, respectively.

This gets complicated, but it's really cool, and unsurprisingly is key to the proof of Theorem 1.17 in the quartic and quintic cases. For more, see Bhargava's Arizona Winter School notes [Bha09]. We do not have a general way to parametrize rings of rank 6 or more; if you can think of one, it would make for a very exciting paper!

4 Local densities for cubic forms

In this section, we make a slight departure and prove some density results for cubic forms, namely, the density of $\mathcal{U}_p$. We will use these to understand how to go from counting cubic rings to counting maximal cubic rings. But for now we can keep the exposition moderately self-contained.

Definition 4.1 (p-adic measure)**.** Recall the p-adic integers $\mathbb{Z}_p = \varprojlim \mathbb{Z}/p^n\mathbb{Z}$. Let μ_p denote the natural measure on $\mathbb{Z}_p$ induced by the p-adic metric, normalized so that $\mu_p(\mathbb{Z}_p) = 1$.

If you haven't thought about this much, the important bit is that μ_p is a Haar measure, i.e. it is translation invariant under the additive structure of $\mathbb{Z}_p$. This forces, for example, $\mu_p(p\mathbb{Z}_p) = \frac{1}{p}$. Informally, if we think of picking random elements out of $\mathbb{Z}_p$, each residue class modulo p^n is equally likely to occur, and indeed $\mu_p(a + p^n\mathbb{Z}_p) = \frac{1}{p^n}$. We will frequently abuse this notation by using μ_p to refer to the product measure on $\mathbb{Z}_p^n$, normalized so $\mu_p(\mathbb{Z}_p^n) = 1$.

Consider now binary forms of degree d. We can consider the space of such forms as $\mathbb{Z}_p^{d+1}$, with natural measure μ_p normalized such that $\mu_p(\mathbb{Z}_p^{d+1}) = 1$. We will be interested in how often such f factor in different ways, as measured by μ_p.

Definition 4.2 (Factorization type)**.** Suppose $f(x, y)$ is a nonzero form of degree d over a field. If f factors as $\prod_{i=1}^r f_i^{e_i}$ for distinct irreducible forms f_i of degree d_i, then we say f has **factorization type** $(d_1^{e_1} \dots d_r^{e_r})$.

Let $T_p(d_1^{e_1} \dots d_r^{e_r}) \subset \mathbb{Z}_p^{d+1}$ denote the subset of degree d forms whose reduction to $\mathbb{F}_p$ is nonzero with factorization type $(d_1^{e_1}, \dots, d_r^{e_r})$.

4.1 Quadratic forms

Let $q(x, y) = ax^2 + bxy + cy^2$ be a quadratic form over $\mathbb{Z}_p$. With the shorthand defined above above, we have

$$\begin{aligned}
\mu_p(T_p(1^2)) &= \frac{(p-1)(p+1)}{p^3} \\
\mu_p(T_p(11)) &= \frac{(p-1)(p+1)}{2p^2} \\
\mu_p(T_p(2)) &= \frac{(p-1)^2}{2p^2}
\end{aligned}$$

These add up to $1/p^3$, the measure of those forms which reduce to the zero form modulo p.

In this case, since the subsets of $\mathbb{Z}_p$ we care about are determined by residue classes modulo p, we can do a direct count over $\mathbb{F}_p$ instead. For example, in the 1^2 case, one simply counts how many (nonzero) linear forms $ax + by$ there are, finding $p^2 - 1$, then divide by the total number of forms.

It is somewhat easier to make this systematic by counting $p+1$ linear forms over $\mathbb{F}_p$ *up to scaling*, and then multiplying by the $p-1$ elements of $\mathbb{F}_p^\times$ by which we can scale.

A similar argument works for $T_p(11)$, except now we choose two distinct linear forms. The count of irreducible quadratic forms over $\mathbb{F}_p$ is then deduced by the fact that a nonzero quadratic form is either a product of linear forms or irreducible. Try working this out carefully in the exercises.

Remark 4.3. A related problem is to compute the p-adic measure of the set of quadratic forms in n variables with a p-adic zero. This is not quite the same problem, because we cannot simply discount those forms which reduce to 0 modulo p. Nevertheless, these densities were worked out explicitly only fairly recently [BCF+16]. Similar questions have been asked for arbitrary degree polynomials, or families of varieties and how often they have p-adic points.

4.2 Cubic forms

In the cubic case, we make the following calculations.

Lemma 4.4. *We have*

$$\begin{aligned}
\mu_p(T_p(1^3)) &= \frac{p^2-1}{p^4} \\
\mu_p(T_p(1^21)) &= \frac{p^2-1}{p^3} \\
\mu_p(T_p(111)) &= \frac{p^3-p^2-p+1}{6p^3} \\
\mu_p(T_p(21)) &= \frac{p^3-p^2-p+1}{2p^3} \\
\mu_p(T_p(3)) &= \frac{p^3-p^2-p+1}{3p^3}
\end{aligned}$$

Proof. As in the quadratic case, this amounts to counting nonzero cubic forms over $\mathbb{F}_p$.

There are p^2-1 nonzero linear forms out of p^4 total, yielding the density for $T_p(1^3)$.

For $T_p(1^21)$, we can count ordered pairs of linear forms *up to scaling*, obtaining $p(p+1)$ such pairs. Multiplying by $p-1$ to account for scaling and dividing by p^4 yields the count.

For $T_p(111)$ we count unordered triples of linear forms up to scaling. After accounting for scaling, we obtain $(p-1)\binom{p+1}{3} = (p-1)^2p(p+1)/6$, which agrees with the result upon division by p^4.

For $T_p(21)$ we recall there were $p(p-1)/2$ irreducible quadratic forms over $\mathbb{F}_p$ up to scaling, so counting pairs of quadratic and linear forms yields $p(p-1)(p+1)/2$. Accounting for scaling and dividing by p^4 gives the result.

Finally, we obtain $T_p(3)$ by

$$\mu_p(T_p(3)) = 1 - \frac{1}{p^4} - \mu_p(T_p(21)) - \mu_p(T_p(111)) - \mu_p(T_p(1^21)) - \mu_p(T_p(1^3)),$$

since $1/p^4$ is the density of forms which reduce to 0 modulo p. □

Next we use these densities to compute the density of $\mathcal{U}_p$; recall these are the f's coming from cubic rings maximal at p.

Lemma 4.5. *We have* $\mu_p(\mathcal{U}_p) = \frac{(p^3-1)(p^2-1)}{p^5}$.

Proof. Write

$$\mathcal{U}_p = \coprod_F \mathcal{U}_p \cap T_p(F)$$

where F runs over the factorization types (1^3), (1^21), (111), (21), (3).

Note that $T_p(111), T_p(21), T_p(3) \subset \mathcal{U}_p$, since $f \notin \mathcal{U}_p$ implies that f has a double root modulo p (or is zero mod p). Thus we need to compute $\mu_p(\mathcal{U}_p \cap T_p(1^3))$ and $\mu_p(\mathcal{U}_p \cap T_p(1^21))$.

Since $T_p(F)$ is $\mathrm{GL}_2(\mathbb{Z}_p)$-invariant, we are free to apply a convenient change of variables. In the case of (1^3), we can write $f = ax^3 + bx^2y + cxy^2 + dy^3$ and assume $a, b, c \equiv 0 \pmod p$ with $d \not\equiv 0 \pmod p$ after acting by $\mathrm{GL}_2(\mathbb{Z})$. Such f is in $\mathcal{U}_p$ if and only if $p^2 \nmid a$. This occurs with density $\frac{p-1}{p}$, so we have

$$\mu_p(\mathcal{U}_p \cap T_p(1^3)) = \frac{p-1}{p}\mu_p(T_p(1^3)) = \frac{p^3 - p^2 - p + 1}{p^5}.$$

A similar calculation for the (1^21) case shows

$$\mu_p(\mathcal{U}_p \cap T_p(1^21)) = \frac{p-1}{p}\mu_p(T_p(1^21)) = \frac{p^3 - p^2 - p + 1}{p^4}.$$

Adding these up produces

$$\mu_p(\mathcal{U}_p) = \frac{1}{p^5}(p^2-1)(p-1)(p^2+p+1) = \frac{(p^3-1)(p^2-1)}{p^5},$$

as requested. □

In a precise sense, $\mu_p(\mathcal{U}_p)$ is the *probability* that a cubic ring is maximal at p.

Lemma 4.6. *Let R be a maximal cubic ring corresponding to $f(x,y)$ and suppose p is a prime. Then $f \not\equiv 0 \pmod p$ and the factorization of f over $\mathbb{F}_p$ determines the splitting behavior of p in R. In particular, p is totally ramified in R if and only if $f \in T_p(1^3)$.*

Proof. This is alluded to in [BST13, §4]. We won't make everything precise, but the idea is that our correspondence in Proposition 3.7 in fact extends to cubic A-algebras for any commutative ring A. Even better, it is functorial in that if $A \to A'$ is a ring map, then the correspondence is compatible with the natural maps on cubic A-algebras, $R \mapsto R \otimes_A A'$, and on binary cubic forms induced by $A[x,y] \mapsto A'[x,y]$.

In particular, when $A = \mathbb{Z}$ and $A' = \mathbb{F}_p$, we have that p is totally ramified in R if and only if $R \otimes_{\mathbb{Z}} \mathbb{F}_p \simeq \mathbb{F}_p[t]/(t^3)$. The other splitting types are similar, e.g. p is totally split if $R \otimes_{\mathbb{Z}} \mathbb{F}_p \simeq \mathbb{F}_p^3$.

To see one direction in the totally ramified case, consider the multiplication structure forced on $R \otimes_{\mathbb{Z}} \mathbb{F}_p$ in the case that $f \in T_p(1^3)$. In this situation, we have $a, b, c = 0$ and $d \neq 0$ (all these equalities are in $\mathbb{F}_p$). Following the construction in Proposition 3.7, we see that this forces $n = m = \ell = 0$, $\omega^2 = 0$, and $\theta^2 = d\omega$. Thus $R \otimes_{\mathbb{Z}} \mathbb{F}_p \simeq \mathbb{F}_p[\theta]/(\theta^3)$. Or, if $f \in T_p(1^3)$, we have p is totally ramified in R.

Repeating for the other factorization types, we realize the other splitting types and get the other implication. □

Davenport and Heilbronn defined another quantity, $\mathcal{V}_p$, by

$$\mathcal{V}_p = \{f \in \mathcal{U}_p : f \notin T_p(1^3)\}.$$

This corresponds to the probability that a cubic form f corresponds to a cubic ring R which is maximal at p and *not* totally ramified at p. This will come up later when we study 3-torsion in the class groups of imaginary quadratic fields. We can perfom similar calculations to determine its measure.

Lemma 4.7. *We have $\mu_p(\mathcal{V}_p) = \frac{(p^2-1)^2}{p^4}$.*

5 Davenport and Heilbronn's theorems

Let's remind ourselves what we are doing here and state the theorems we want to prove. Let $N_3(\xi, \eta)$ denote the number of cubic fields $K/\mathbb{Q}$ with discriminant $\xi < \operatorname{Disc} K/\mathbb{Q} < \eta$, counted up to isomorphism as usual.

Theorem 5.1 (Davenport–Heilbronn [DH71, Theorem 1])**.** *We have*

$$N_3(0, X) \sim \frac{X}{12\zeta(3)} \tag{5.1}$$

$$N_3(-X, 0) \sim \frac{X}{4\zeta(3)} \tag{5.2}$$

See also Bhargava, Shankar, and Tsimerman's paper where they give a second main term of order $X^{5/6}$ and a sharper error term [BST13, Theorems 1, 3]. We won't discuss this second main term here.

Definition 5.2. Let $V_{\mathbb{Z}}^{+}$ (resp. $V_{\mathbb{Z}}^{-}$) denote the space of integral binary cubic forms with positive (resp. negative) discriminant.

That is, $V_{\mathbb{Z}}^{\pm} \subset V_{\mathbb{Z}} = \mathbb{Z}^4$. One can similarly define $V_A^{\pm}$ for an arbitrary commutative ring A; we will mostly only be interested in $A = \mathbb{Z}, \mathbb{R}$. Note that we are nearly following the notation in [BST13] here.

Definition 5.3. Let $N(V_{\mathbb{Z}}^{\pm}, X)$ denote the number of *irreducible* $\mathrm{GL}_2(\mathbb{Z})$-orbits $f \in V_{\mathbb{Z}}^{\pm}$ with $|\mathrm{Disc}(f)| < X$.

A key step in the proof is to prove the following theorem, originally due to Davenport.

Theorem 5.4 (Davenport, see [BST13, Theorem 20])**.** *We have*

$$N(V_{\mathbb{Z}}^{+}, X) \sim \frac{\pi^2}{72} X \tag{5.3}$$

$$N(V_{\mathbb{Z}}^{-}, X) \sim \frac{\pi^2}{24} X \tag{5.4}$$

Davenport's method essentially boils down to thinking of integral binary cubic forms as lattice points in the space of *real* binary cubic forms, all up to the action of $\mathrm{GL}_2(\mathbb{Z})$. Then he essentially counts lattice points in a fundamental domain for the action of $\mathrm{GL}_2(\mathbb{Z})$ on $V_{\mathbb{R}}^{\pm}$. In "good" situations, counting lattice points in a region should be about the same as computing the volume of that region. Unfortunately, the fundamental domain in our case isn't so nicely shaped; it has *cusps*, i.e. regions extending off towards infinity that contain lots of lattice points. While Davenport is able to get around this, computing the fundamental volume boils down to a really nasty integral; indeed he made some mistakes in the originally published version.

One of the innovations in [BST13] is to average over several fundamental domains, in some sense, rather than choose only one of them. This essentially allows us to replace our badly behaved fundamental domain with a nicer compact region, where it is easier to count lattice points. This technique has had other exciting applications, including in Theorem 1.32. One thing we have to be careful of here is that we are counting *irreducible* forms, so we will have to take care to check that there aren't too many reducible forms showing up. At this point, one can also show that few cubic rings with automorphism group C_3 show up, allowing our result to be for S_3-cubic fields.

With Theorem 5.4 in hand, we can count irreducible cubic rings by discriminant. In fact, this can be upgraded to allow for (finitely many) congruence conditions to be imposed on the cubic forms. This plays nicely with the count in that we simply multiply by the product of the p-adic measures of the congruence conditions imposed. Together with the local densities we computed in Lemma 4.5, this will allow us to sieve out those cubic rings which are not maximal, thereby proving Theorem 5.1.

5.1 Reduction theory

In this section, we seek to have some understanding of a fundamental domain for the action of $\mathrm{GL}_2(\mathbb{Z})$ on $\mathrm{GL}_2(\mathbb{R})$.

Definition 5.5 (Subgroups of $\mathrm{GL}_2(\mathbb{R})$)**.** We have the following subgroups of $\mathrm{GL}_2(\mathbb{R})$:

$$\begin{aligned} K_1 &= O_2(\mathbb{R}) = \{\text{orthogonal transformations}\}, \\ A_+ &= \left\{ \begin{pmatrix} t^{-1} & 0 \\ 0 & t \end{pmatrix} : t \in \mathbb{R}_+ \right\}, \\ N &= \left\{ \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix} : u \in \mathbb{R} \right\}, \\ \Lambda &= \left\{ \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix} : \lambda > 0 \right\}. \end{aligned}$$

For any $g \in \mathrm{GL}_2(\mathbb{R})$, we have a unique decomposition $g = kan\lambda$ for $k \in K_1,\ a \in A_+,\ n \in N,\ \lambda \in \Lambda$.

We also have a natural Haar measure on $\mathrm{GL}_2(\mathbb{R})$, normalized so that $\mathrm{SO}_2(\mathbb{R}) \subset K_1$ has volume 1, given by $\frac{1}{\lambda t^3} dndtdkd\lambda$. To help see this, one can write $\mathrm{GL}_2(\mathbb{R}) = \mathbb{G}_m \times \mathrm{SL}_2(\mathbb{R})$ and find a translation invariant measure in the appropriate coordinates on each. Try checking this in the exercises, or see [Lan, Lemma 3.22].

Lemma 5.6. *There is a fundamental domain $\mathcal{F}$ for $\mathrm{GL}_2(\mathbb{Z})$ acting on $\mathrm{GL}_2(\mathbb{R})$, given by*

$$\mathcal{F} = \{nak\lambda : n \in N',\ a \in A',\ k \in K,\ \lambda \in \Lambda\},$$

where $K = \mathrm{SO}_2(\mathbb{R})$ is the special orthogonal group,

$$\begin{aligned} N' = N'(a) &= \left\{ \begin{pmatrix} 1 & 0 \\ n & 1 \end{pmatrix} : n \in \nu(a) \right\}, \\ A' &= \left\{ \begin{pmatrix} t^{-1} & 0 \\ 0 & t \end{pmatrix} : t \geq \sqrt[4]{3}\sqrt{2} \right\}, \end{aligned}$$

and $\nu(a)$ is a subinterval of $[-\frac{1}{2}, \frac{1}{2}]$.

Proof. See [Lan, Lemma 3.33] for a proof. □

Recall that $\mathrm{GL}_2(\mathbb{R})$ acts transitively on $V_{\mathbb{R}}^{+}$ with S_3-stabilizers, while it acts transitively on $V_{\mathbb{R}}^{-}$ with C_2-stabilizers. Set $n_+ = 6,\ n_- = 2$ to account for this.

Suppose $v \in V_{\mathbb{R}}^{\pm}$. Then translating by $\mathcal{F}$, we obtain a union of $n_\pm$ fundamental domains for the action of $\mathrm{GL}_2(\mathbb{Z})$ on $V_{\mathbb{R}}^{\pm}$. It is convenient to think of $\mathcal{F}v$ as a multiset: we have $v' \in \mathcal{F}v$ means $v' = gv$ for $g \in \mathcal{F}$, and the multiplicity of v' is the number of $g \in \mathcal{F}$ for which this occurs.

This (almost) means that $n_\pm N(V_{\mathbb{Z}}^{\pm}, X)$ is equal to the number of irreducible integer points in $\mathcal{F}v$ with discriminant less than X (and the proper sign). To make this precise, we must take into account the possible presence of $\mathrm{GL}_2(\mathbb{Z})$-stabilizers of such points; this amounts to counting cubic rings with automorphism group C_3 with weight $\frac{1}{3}$ instead, since with our multiset counting approach, we have overcounted these points. However, these points are rather sparse, and this can be ignored entirely for cubic forms of negative discriminant, so we will mostly gloss over this detail.

5.2 Reducible forms

Before we count irreducible forms, we count reducible forms.

Lemma 5.7. *Fix a compact subspace $B \subset V_{\mathbb{R}}$ containing elements of discriminant at least 1. Suppose $v \in B$. Then*

$$\#\left\{f \in \mathcal{F}v \cap V_{\mathbb{Z}} : |\mathrm{Disc}(v')| < X,\ f \text{ reducible with } a \neq 0\right\} = O_B(X^{3/4+\epsilon}).$$

Proof. Write $f = nak\lambda v$.[4] We have $\mathrm{Disc}(f) = \lambda^4 \mathrm{Disc}(v)$, so $\lambda < X^{1/4}$. (Exercise: work out how $n,\ a,\ k,\ \lambda$ affect the discriminant.)

Let's now see how the coefficients are affected. Since B is compact, the original coefficients of v are all $O_B(1)$. Acting by λ scales them all, while acting by k or n scales them by a bounded amount. Acting by $a(t)$ sends the coefficients $(a,b,c,d) \mapsto (a/t^3, b/t, tc, t^3d)$. Thus we can estimate the sizes

$$a = O(\lambda/t^3),\ b = O(\lambda/t),\ c = O(\lambda t),\ d = O(\lambda t^3).$$

Immediately we have that $abc = O(\lambda^3/t^3)$, so there are $O(X^{3/4})$ $f \in \mathcal{F}v \cap V_{\mathbb{Z}}$ with $|\mathrm{Disc}(f)| < X$ and $d = 0$.

Assume now that $ad \neq 0$. There are $O(X^{3/4})$ choices for (a,b,d). Fixing such a choice, we count the c for which f is reducible. For f to be reducible, it has a linear factor $rx + sy$ for $\gcd(r,s) = 1$, from which it follows that $r \mid a,\ s \mid d$. The divisor function grows like $O(\frac{\log X}{\log\log X})$, so there are at most $O(X^\epsilon)$ divisors of a and d.

Now for each tuple (a,b,d,r,s), there is a unique (real) c determined by solving $f(-s,r) = 0$ for c. Thus there are at most $O(X^{3/4+\epsilon})$ reducible forms with $a \neq 0$. □

In the positive discriminant case, we will also want a handle on those cubic forms which have stabilizers in $\mathrm{GL}_2(\mathbb{Z})$. This is done in [BST13, Lemma 22] using the Hessian covariant. It turns out that

$$\#\left\{f \in \mathcal{F}v \cap V_{\mathbb{Z}} : |\mathrm{Disc}(v')| < X,\ f \text{ has stabilizer } C_3\right\} = O_B(X^{3/4+\epsilon}).$$

For simplicity we will skip this, but content ourselves with knowing that both the reducible and C_3-points are negligible for our purposes.

5.3 Averaging

Let $C \geq 1$ be a constant and $B = B(C) \subset V_{\mathbb{R}}$ be the bounded region given by

$$B(C) = \left\{v = (a,b,c,d) \in V_{\mathbb{R}} : 3a^2 + b^2 + c^2 + 3d^2 \leq C,\ |\mathrm{Disc}(v)| \geq 1\right\}.$$

This looks like some sort of 3-dimensional ellipsoid. Moreover, it is closed under the $\mathrm{SO}_2(\mathbb{R})$-action, a fact which can be verified by a lengthy computation (see the exercises).

Let $v \in V_{\mathbb{R}}^{\pm}$ and recall from §5.1 that the number of irreducible integral forms of discriminant at most X is

$$N(V_{\mathbb{Z}}^{\pm}, X) = \frac{1}{n_\pm}\#\left\{f \in \mathcal{F}v \cap V_{\mathbb{Z}}^{\mathrm{irr}} : |\mathrm{Disc}(f)| < X\right\},$$

with the latter counted as a multiset — the points with automorphism group C_3 are counted with weight 1/3.

The starting point v was arbitrary; if we change it, we still compute $N(V_{\mathbb{Z}}^{\pm}, X)$. Better yet, what if we took an appropriate *average* as v varied in some nice region, like $B(C)$? To make this precise, we use the measure $\frac{dv}{|\mathrm{Disc}(v)|^{-1}}$, which is $\mathrm{GL}_2(\mathbb{R})$-invariant. Now we see

$$N(V_{\mathbb{Z}}^{\pm}, X) = \frac{\int_{v \in B \cap V_{\mathbb{R}}^{\pm}} \#\left\{f \in \mathcal{F}v \cap V_{\mathbb{Z}}^{\mathrm{irr}} : |\mathrm{Disc}(f)| < X\right\} |\mathrm{Disc}(v)|^{-1}\, dv}{n_\pm \int_{v \in B \cap V_{\mathbb{R}}^{\pm}} |\mathrm{Disc}(v)|^{-1}\, dv}.$$

Note that this denominator, whatever it is, is some positive constant, being the volume of the compact region $B \cap V_{\mathbb{R}}^{\pm}$ (with respect to the chosen measure).

We now want to exchange our integral over $V_{\mathbb{R}}^{\pm}$ for one over $\mathrm{GL}_2(\mathbb{R})$.

[4] Here $a \in A'$ is a matrix, but we also use a to denote the x^3 coefficient of f.

Proposition 5.8. *Fix $v_\pm \in V_{\mathbb{R}}^\pm$ and $H^\pm \subset \mathrm{GL}_2(\mathbb{R})$ so that $H^\pm \cdot v_\pm = B \cap V_{\mathbb{R}}^\pm$. Then we have*

$$\begin{aligned}
N(V_{\mathbb{Z}}^\pm, X) &= \frac{\int_{v \in B \cap V_{\mathbb{R}}^\pm} \#\left\{x \in \mathcal{F}v \cap V_{\mathbb{Z}}^{\mathrm{irr}} : |\mathrm{Disc}(x)| < X\right\} |\mathrm{Disc}(v)|^{-1}\, dv}{n_\pm \int_{v \in B \cap V_{\mathbb{R}}^\pm} |\mathrm{Disc}(v)|^{-1}\, dv} \\
&= \sum_{\substack{x \in (V_{\mathbb{Z}}^\pm)^{\mathrm{irr}} \\ |\mathrm{Disc}(x)| < X}} \frac{\int_{v \in B \cap V_{\mathbb{R}}^\pm} \#\left\{g \in \mathcal{F} : x = gv\right\} |\mathrm{Disc}(v)|^{-1}\, dv}{n_\pm \int_{v \in B \cap V_{\mathbb{R}}^\pm} |\mathrm{Disc}(v)|^{-1}\, dv} \\
&= \frac{2\pi}{n_\pm} \sum_{\substack{x \in (V_{\mathbb{Z}}^\pm)^{\mathrm{irr}} \\ |\mathrm{Disc}(x)| < X}} \frac{\int_{h \in H^\pm} \#\left\{g \in \mathcal{F} : x = ghv_\pm\right\} dh}{\int_{v \in B \cap V_{\mathbb{R}}^\pm} |\mathrm{Disc}(v)|^{-1}\, dv} \\
&= \frac{2\pi}{n_\pm} \frac{\int_{g \in N'(a)A'\Lambda} \#\{x \in (V_{\mathbb{Z}}^\pm)^{\mathrm{irr}} \cap B(n,t,\lambda,X)\} t^{-3}\lambda^{-1} dn dt d\lambda}{\int_{v \in B \cap V_{\mathbb{R}}^\pm} |\mathrm{Disc}(v)|^{-1}\, dv}
\end{aligned}$$

Proof. The second equality follows from how we are counting our multisets. Each irreducible $x \in V_{\mathbb{Z}}^\pm$ appears in the count once for every $g \in \mathcal{F}$ for which $gv = x$. Thus we can pull out these (finitely many) x's and sum over them to obtain the second equality.

The third equality is where we actually do something! Since $\mathrm{GL}_2(\mathbb{R})$ is an $n_\pm$-fold cover of $V_{\mathbb{R}}^\pm$, we can go from integrating over $V_{\mathbb{R}}^\pm$ to $\mathrm{GL}_2(\mathbb{R})v_\pm$ at the expense of a factor of $n_\pm$. We can then go from integrating over the orbit to integrating over $\mathrm{GL}_2(\mathbb{R})$ itself, via change of variables. The appropriate Jacobian calculation for this change introduces the $|\mathrm{Disc}(v)|^{-1}$ factor in going from the (k, t, n, λ) coordinates for $\mathrm{GL}_2(\mathbb{R})$ to the (a, b, c, d) coordinates of $V_{\mathbb{R}}$. (In particular, this calculation shows $|\mathrm{Disc}(v)|^{-1}\, dv$ is $\mathrm{GL}_2(\mathbb{R})$-invariant on $V_{\mathbb{R}}$; see [BST13, Proposition 23] for more explanation.) This discussion allows us to replace our integral over $v \in B \cap V_{\mathbb{R}}^\pm$ by the one over $h \in H^\pm$.

For the last equality, we first use the translation invariance of our measure on $\mathrm{GL}_2(\mathbb{R})$ to integrate instead over $\mathbb{F}$, replacing $\int_{h \in H^\pm} \#\left\{g \in \mathcal{F} : x = ghv_\pm\right\} dh$ by

$$\int_{g \in \mathcal{F}} \#\{h \in H^\pm : x = ghv_\pm\} dg.$$

Then we bring the sum over x back inside the integral, obtaining in the numerator

$$\int_{g \in \mathcal{F}} \#\{x \in (V_{\mathbb{Z}}^\pm)^{\mathrm{irr}} \cap gB : |\mathrm{Disc}(x)| < X\} dg$$

Finally, we rewrite everything in terms of the coordinates of $\mathcal{F}$; we can drop the $k \in \mathrm{SO}_2$ part because B is SO_2-invariant. Write $B(n, t, \lambda, X)$ for $na\lambda B \cap \{v \in V_{\mathbb{R}}^\pm : |\mathrm{Disc}(v)| < X\}$, the translate of B by an element of $\mathcal{F}$, with bounded discriminant. This gives the final equality. □

5.4 Counting lattice points

We have finally reduced the challenge of proving Theorem 5.4 to counting (irreducible) lattice points in $B(n, t, \lambda, X)$. For this, we use Davenport's lemma.

Lemma 5.9 (Davenport)**.** *Let $\mathcal{R}$ be a bounded semi-algebraic set in $\mathbb{R}^n$. Then the number of lattice points in $\mathcal{R}$ is*

$$\mathrm{Vol}(\mathcal{R}) + O(\max\{\mathrm{Vol}(\overline{\mathcal{R}}), 1\}),$$

where $\overline{\mathcal{R}}$ runs over the projections of $\mathcal{R}$ onto subspaces of $\mathbb{R}^n$ obtained by setting one or more coordinates to zero.

Davenport's lemma makes precise something which is hopefully fairly intuitive.

Example 5.10 (circle)**.** Let $\mathcal{R} \subset \mathbb{R}^2$ be a circle of radius $r > 1$, so $\mathrm{Vol}(\mathcal{R}) = \pi r^2$. Both coordinate projections have volume $2r$, equal to the diameter of the circle. Thus Lemma 5.9 implies that the number of lattice points is $\pi r^2 + O(r)$, agreeing with a classical result of Gauss.

Problems arise when dealing with regions who grow too quickly in one direction than the others; this leads to the main term getting absorbed into the error term. This is what we want to be careful of.

Lemma 5.11. *The number of lattice points in $B(n,t,\lambda,X)$ with $a \neq 0$ is 0 if $C\lambda/t^3 < 1$ and*

$$\mathrm{Vol}(B(n,t,\lambda,X)) + O(\max\{C^3t^3\lambda^3, 1\})$$

otherwise.

Proof. From our description of $B = B(C)$, we have that the a-coordinate is bounded above by C. After translating by the appropriate matrices, this implies that the a-coordinate of a form in $B(n,t,\lambda,X)$ is bounded above by $C\lambda/t^3$, hence if this quantity is less than 1, there are no such forms.

Otherwise, we have $t \geq \sqrt[4]{3}/\sqrt{2}$, and $\lambda \geq t^3/C = (\sqrt[4]{3}/\sqrt{2})^3/C$. Projecting onto the $a = 0$ coordinate subspace, we check that the b, c, and d coordinates are $O(C\lambda/t)$, $O(C\lambda t)$, and $O(C\lambda t^3$, respectively, so the volume of this projection is $O(C^3\lambda^3t^3)$. Similar arguments for the other projections yield the result, so we may apply Davenport's lemma. □

Let's observe that

$$\mathrm{Vol}(B(n,t,\lambda,X)) = O(C^4\lambda^4). \tag{5.5}$$

This follows from a similar calculationto the proof above: each of the (a,b,c,d) coordinates is $O(C\lambda/t^3)$, $O(C\lambda/t)$, $O(C\lambda t)$, $O(C\lambda t^3)$, respectively. We will use this shortly.

Next, define

$$\mathcal{R}_X(v) = \{w \in \mathcal{F}v : |\mathrm{Disc}(w)| < X\},$$

the (multi)set of forms in the translate of $\mathcal{F}$ by v with discriminant at most X. Our next immediate goal is to relate our count from Proposition 5.8 to the volume of $\mathcal{R}_X(v)$.

Set $M_\pm = \frac{2\pi}{n_\pm \int_{v \in B \cap V_{\mathbb{R}}^{\pm}} |\mathrm{Disc}(v)|^{-1} dv}$. Starting from Proposition 5.8, we have

$$\begin{aligned} N(V_{\mathbb{Z}}^{\pm}, X) &= \frac{1}{M_\pm} \int_{g \in N'(a)A'\Lambda} \#\{x \in (V_{\mathbb{Z}}^{\pm})^{\mathrm{irr}} \cap B(n,t,\lambda,X)\} t^{-3}\lambda^{-1} dndtd\lambda \\ &= \frac{1}{M_\pm} \int_{\lambda = (\sqrt[4]{3}/\sqrt{2})^3/C}^{X^{1/4}} \int_{t=\sqrt[4]{3}/\sqrt{2}}^{C^{1/3}\lambda^{1/3}} \int_{N'(t)} \#\{x \in (V_{\mathbb{Z}}^{\pm})^{\mathrm{irr}} \cap B(n,t,\lambda,X)\} t^{-3}\lambda^{-1} dndtd\lambda \\ &= \frac{1}{M_\pm} \int_{\lambda = (\sqrt[4]{3}/\sqrt{2})^3/C}^{X^{1/4}} \int_{t=\sqrt[4]{3}/\sqrt{2}}^{C^{1/3}\lambda^{1/3}} \int_{N'(t)} \mathrm{Vol}(B(n,t,\lambda,X)) + O(\max\{C^3t^3\lambda^3 1\}) t^{-3}\lambda^{-1} dndtd\lambda + O(X^{3/4}). \end{aligned} \tag{5.6}$$

We obtain the second equality by recognizing that $\lambda < X^{1/4}$ and the integrand is zero whenever $C\lambda/t^3 < 1$. By Lemma 5.7, including the irreducible elements in our count introduces an error of $O(X^{3/4})$. The next equality follows from Lemma 5.11. We now look to compute/estimate this integral.

For the first term involving the volume, we have

$$\begin{aligned}
&\frac{1}{M_\pm}\int_{\lambda=(\sqrt[4]{3}/\sqrt{2})^3/C}^{X^{1/4}}\int_{t=\sqrt[4]{3}/\sqrt{2}}^{C^{1/3}\lambda^{1/3}}\int_{N'(t)}\mathrm{Vol}(B(n,t,\lambda,X))t^{-3}\lambda^{-1}dndtd\lambda\\
&\quad=\frac{1}{M_\pm}\int_{(\sqrt[4]{3}/\sqrt{2})^3/C}^{X^{1/4}}\int_{\sqrt[4]{3}/\sqrt{2}}^{\infty}\int_{N'(t)}\mathrm{Vol}(B)t^{-3}\lambda^{-1}dndtd\lambda-\frac{1}{M_\pm}\int_{(\sqrt[4]{3}/\sqrt{2})^3/C}^{X^{1/4}}\int_{C^{1/3}\lambda^{1/3}}^{\infty}\int_{N'(t)}\mathrm{Vol}(B)t^{-3}\lambda^{-1}dndtd\lambda\\
&\quad=\frac{1}{M_\pm 2\pi}\int_{v\in B\cap V_{\mathbb{R}}^{\pm}}\mathrm{Vol}(\mathcal{R}_X(v))\left|\mathrm{Disc}(v)\right|^{-1}dv+O\left(\int_{(\sqrt[4]{3}/\sqrt{2})^3/C}^{X^{1/4}}\int_{C^{1/3}\lambda^{1/3}}^{\infty}\int_{N'(t)}C^4t^{-3}\lambda^3dndtd\lambda\right)\\
&\quad=\frac{\mathrm{Vol}(\mathcal{R}_X(v))}{n_\pm}+O\left(X^{5/6}\right).
\end{aligned}$$

Breaking things down:

- in the first equality, we let t go to infinity and subtract off the excess;
- for the first term in the second equality, we use the same change of variables as we did in Proposition 5.8 to go from integrating over volumes of translates as n,t,λ vary, to integrating over the volumes of $\mathcal{R}_X(v)$'s as v varies in B;
- for the first term in the final equality, we recognize that the volume of $\mathcal{R}_X(v)$ is independent of v, so pulling out this factor leaves a copy of $2\pi M_\pm/n_\pm$;
- for the second term, we use our estimate on $\mathrm{Vol}(B(n,t,\lambda,X))$ to obtain the second equality and integrate.

For the second term, we have

$$O\left(\int_{\lambda=(\sqrt[4]{3}/\sqrt{2})^3/C}^{X^{1/4}}\int_{t=\sqrt[4]{3}/\sqrt{2}}^{C^{1/3}\lambda^{1/3}}\int_{N'(t)}C^3\lambda^2dndtd\lambda\right)=O(X^{5/6})$$

also. Now we have

$$N(V_{\mathbb{Z}}^{\pm},X)=\frac{\mathrm{Vol}(\mathcal{R}_X(v_\pm))}{n_\pm}+O(X^{5/6}).$$

It remains to compute the fundamental volume. This is done by again going from an integral over $V_{\mathbb{R}}$ to one over $\mathrm{GL}_2(\mathbb{R})$ and using the explicit fundamental domain. Since $\mathcal{R}_X(v)$ is a multiset, with each point having multiplicity $n_\pm$ (up to an error of $o(X)$ coming from points with different multiplicities), we have

$$\begin{aligned}
\frac{\mathrm{Vol}(\mathcal{R}_X(v))}{n_\pm}&=\int_{v\in V_{\mathbb{R}}^{\pm}}1_{\mathcal{R}_X(v)}dv\\
&=\frac{2\pi}{n_\pm}\int_0^{X^{1/4}}\lambda^4d\lambda\int_{h\in N'(a)A'K}dh\\
&=\frac{2\pi}{n_\pm}\frac{X}{4}\frac{\pi}{6}=\frac{\pi^2}{12n_\pm}X.
\end{aligned}$$

When going from the volume of $\mathcal{R}_X(v_\pm)$ to the integral over $\mathcal{F}$, we introduce a factor of λ^4 accounting for the discriminant. The volume of $N'(a)A'K$ is $\frac{\pi}{6}$, a classical computation; see [BST13, §5.4]. See also [Lan, §3.7] for an alternative approach using p-adic integrals.

Taken all together, this discussion completes the proof of Theorem 5.4.

Remark 5.12 (Congruence conditions)**.** Suppose $S \subset V_{\mathbb{Z}}^{\pm}$ is a subset cut out by *finitely* many congruence conditions. Then we may define $N(S, X)$ analogously to $N(V_{\mathbb{Z}}^{\pm}, X)$ to be the number of irreducible orbits *in* S with discriminant at most X. Then we have

$$\lim_{X\to\infty} \frac{N(S,X)}{N(V_{\mathbb{Z}}^{\pm},X)} = \prod_p \mu_p(S).$$

That is, $N(S, X)$ grows proportionally to $N(V_{\mathbb{Z}}^{\pm}, X)$, with the proportion given by product of p-adic densities of S.[5]

This is obtained by essentially repeating the proof of Theorem 5.4, inserting the congruence conditions at the appropriate steps and ensuring the error terms are unaffected; see [BST13, Theorem 26].

5.5 Sieving

We begin with a definition.

Definition 5.13 (content)**.** The **content** of an integral binary cubic form $f(x, y) = ax^3 + bx^2y + cxy^2 + dy^3$ is $\mathrm{ct}(f) = \gcd(a, b, c, d)$.

The **content** of a cubic ring R, denoted $\mathrm{ct}(R)$, is the maximal integer n such that $R(f) = \mathbb{Z}+nR'$ for a cubic ring R'.

We say a form or ring is **primitive at p** if p does not divide its content, and simply **primitive** if its content is 1.

It follows from the correspondence between cubic rings and cubic forms that $\mathrm{ct}(R_f) = \mathrm{ct}(f)$.

Next we recognize a fact about subrings of R_f.

Lemma 5.14. *The number of index p subrings of R_f is equal to the number of zeros of f modulo p.*

Proof. Given a zero of $f(x, y)$ modulo p, after acting by linear change of coordinates, we may assume it is at $(0, 1)$, i.e. that $d \equiv 0 \pmod p$. If $R = \mathbb{Z}[\omega, \theta]$ is our normalized basis corresponding to f, then one easily sees that $\mathbb{Z}[p\omega, \theta]$ is a subring of index p.

On the other hand, given a subring $R' \subset R$ of index p, one can find a basis $\{1, \omega, \theta\}$ for R such that $R' = \mathbb{Z}[p\omega, \theta]$. Then we deduce $p \mid d$ in the corresponding form f, yielding a root. □

Definition 5.15. Let $\mathcal{W}_p \subset V_{\mathbb{Z}}$ denote those forms f which correspond to cubic rings *not* maximal at p. Thus we have $V_{\mathbb{Z}} = \mathcal{U}_p \coprod \mathcal{W}_p$.

We can uniformly bound the number of cubic forms/rings which are not maximal at p.

Lemma 5.16. *We have $N(\mathcal{W}_p, X) = O(X/p^2)$ where the implied constant is independent of p.*

Proof. Suppose $R = \mathbb{Z}[\omega, \theta] \in \mathcal{W}_p$ with $|\mathrm{Disc}(R)| < X$. Then by Lemma 3.10, either $R' = \mathbb{Z}[\omega/p, \theta]$ or $R'' = \mathbb{Z}[\omega/p, \theta/p]$ is a ring.

Suppose we are in the former case, with R' a ring. If R' is primitive at p, then $\mathrm{Disc}(R) = \mathrm{Disc}(R')p^2$, so there are at most $O(X/p^2)$ such R', by Theorem 5.4. By the previous lemma, there are at most 3 R per R', so we lose nothing by counting the R' instead.

If R' is not primitive at p, then there is a cubic ring S with $R' = \mathbb{Z} + pS$, and $\mathrm{Disc}(S) = \mathrm{Disc}(R)/p^6 < X/p^6$. Thus there are $O(X/p^6)$ choices of S, hence R'. Now we have $p+1$ possible R per R', but this is still far less than $O(X/p^2)$.

If we are in the latter case, R'' is a ring. Here we have $R = \mathbb{Z}+pR''$ and $\mathrm{Disc}(R'') = \mathrm{Disc}(R)/p^4 < X/p^4$. Again, there are $O(X/p^4)$ such R''.

Putting this all together, we have that there are $O(X/p^2)$ cubic rings R with discriminant at most X which are not maximal at p. □

[5] Since S is defined by finitely many congruence conditions, for $p \gg 0$ we have $\mu_p(S) = 1$, hence the product is actually finite.

At last, we have our tools to prove Theorem 5.1.

Proof of Theorem 5.1. Let $Y > 0$. We can bound

$$N_3(0, X) \leq N(\cap_{p<Y}\mathcal{U}_p \cap V_{\mathbb{Z}}^+, X)$$

so an application of Theorem 5.4 with congruence conditions yields

$$\lim_{X\to\infty} \frac{N_3(0,X)}{X} \leq \lim_{X\to\infty} \lim_{Y\to\infty} \frac{\pi^2}{72} \prod_{p<Y} \left((1 - \frac{1}{p^2})(1 - \frac{1}{p^3}) \right)$$

Letting $Y \to \infty$ gives

$$\lim_{X\to\infty} \frac{N_3(0,X)}{X} \leq \frac{1}{12\zeta(3)}.$$

For a lower bound, note that we have

$$\cap_{p<Y}\mathcal{U}_p \subset (\cap_p \mathcal{U}) \coprod (\cup_{p\geq Y} \mathcal{W}_p).$$

Thus we have

$$\begin{aligned}\lim_{X\to\infty} \frac{N_3(0,X)}{X} &\geq \lim_{X\to\infty} \lim_{Y\to\infty} \frac{N(\cap_{p<Y}\mathcal{U}_p, X)}{X} - \frac{N(\cup_{p\geq Y}\mathcal{W}_p)}{X} \\ &= \lim_{X\to\infty} \lim_{Y\to\infty} \frac{\pi^2}{72} \prod_{p<Y} \left((1 - \frac{1}{p^2})(1 - \frac{1}{p^3}) \right) - \sum_{p\geq Y} O\left(\frac{1}{p^2}\right).\end{aligned}$$

The error term is $o(1)$, being the tail of an absolutely convergent series. Therefore

$$\lim_{X\to\infty} \frac{N_3(0,X)}{X} \geq \frac{1}{12\zeta(3)}.$$

Putting these inequalities together yields

$$N_3(0, X) \sim \frac{X}{12\zeta(3)}.$$

The same argument for negative discriminants gives

$$N_3(-X, 0) \sim \frac{X}{4\zeta(3)},$$

completing the proof. □

References

[Ach] Niven Achenjang. Counting cubic number fields. Available at https://www.mit.edu/~NivenT/assets/pdf/Counting_Cubic_Number_Fields.pdf.

[BCF+16] 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? *Int. Math. Res. Not. IMRN*, (12):3828–3848, 2016.

[Bha05] Manjul Bhargava. The density of discriminants of quartic rings and fields. *Ann. of Math. (2)*, 162(2):1031–1063, 2005.

[Bha09] Manjul Bhargava. On the classification of rings of “small” rank. Arizona Winter School notes, https://swc-math.github.io/aws/2009/09BhargavaNotes.pdf, 2009.

[Bha10] Manjul Bhargava. The density of discriminants of quintic rings and fields. *Ann. of Math. (2)*, 172(3):1559–1591, 2010.

[BS15] 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.

[BST13] Manjul Bhargava, Arul Shankar, and Jacob Tsimerman. On the Davenport-Heilbronn theorems and second order terms. *Invent. Math.*, 193(2):439–499, 2013.

[Cox22] David Cox. *Primes of the Form x^2+ny^2: Fermat, Class Field Theory, and Complex Multiplication*, volume 387 of *American Mathematical Society Chelsea Publishing*. American Mathematical Society, Providence, RI, 2022.

[DH71] H. Davenport and H. Heilbronn. On the density of discriminants of cubic fields. II. *Proc. Roy. Soc. London Ser. A*, 322(1551):405–420, 1971.

[GM04] Andrew Granville and Greg Martin. Prime number races. Available at https://arxiv.org/abs/math/0408319, 2004.

[Klü05] Jürgen Klüuners. A counterexample to Malle’s conjecture on the asymptotics of discriminants. *C. R. Math. Acad. Sci. Paris*, 340(6):411–414, 2005.

[Lan] Aaron Landesman. Notes on counting extensions of degrees 2 and 3, following Bhargava. Available at https://people.math.harvard.edu/~landesman/assets/bhargavology-seminar-notes.pdf.

[LS24] Daniel Loughran and Tim Santens. Malle’s conjecture and brauer groups of stacks, 2024.

[Neu99] Jürgen Neukirch. *Algebraic number theory*, volume 322 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999. Translated from the 1992 German original and with a note by Norbert Schappacher, With a foreword by G. Harder.

[Woo14] Melanie Wood. Asymptotics for number fields and class groups. Arizona Winter School notes, https://swc-math.github.io/aws/2014/2014WoodNotes.pdf, 2014.