概均質ベクトル空間論の発展（第30回整数論サマースクール報告集、写真なし）

谷口, 隆 ; 杉山, 和成 ; 石塚, 裕大 ; 佐藤, 文広 ; 都築, 正男 ; Thorne, Frank ; 鈴木, 雄太 ; 伊吹山, 知義 ; 鈴木, 美裕 ; 佐野, 薫 ; 山本, 修司

(Citation)
第30回整数論サマースクール報告集「概均質ベクトル空間論の発展」:1-421

(Issue Date)
2024-01-31

(Resource Type)
conference proceedings

(Version)
Version of Record

(JaLCDOI)
https://doi.org/10.24546/0100486229

(URL)
https://hdl.handle.net/20.500.14094/0100486229

Counting cubic fields using Shintani's zeta function

Frank Thorne (University of South Carolina)

Abstract

From analytic properties of zeta functions, we can effectively count integral orbits of prehomogneous vector spaces via Landau's theorem. As a typical example, I will explain how to count cubic fields by their discriminants, using the zeta functions studied by Shintani. In particular, I will:

(1) Explain the Davenport-Heilbronn correspondence for maximality of cubic rings, and how this can be incorporated into Shintani's zeta functions;

(2) Give an overview of Landau's method, and explain how the existence of zeta functions leads to arithmetic density results;

(3) Describe equidistribution of this maximality condition in terms of exponential sums, and give an overview of how these sums can be computed;

(4) Explain how to put these pieces together with a simple sieve, and thereby count cubic fields.

This note are based on a presentation given (remotely) to the Summer School on Prehomogeneous Vector Spaces, organized by Yasuhiro Ishitsuka, Kazunari Sugiyama, and Takashi Taniguchi in Kobe, September 2023.

Note. In some places, there is some textual overlap with some of the author and his collaborators' cited papers.

1 Introduction

Here is a typical example of a result in arithmetic statistics. Let $N_d(X)$ count the number of number fields K, of degree d over $\mathbb{Q}$, and with $|\mathrm{Disc}(K)| < X$.

Then the **Davenport-Heilbronn theorem** [DH71] asserts that

$$N_3(X) \sim \frac{1}{3\zeta(3)} X. \tag{1.1}$$

A succession of papers [Bel99, BBP10, BST13, TT13b, BTT], written (in various permutations) by Karim Belabas, Manjul Bhargava, Carl Pomerance, Arul Shankar, Takashi Taniguchi, Jacob Tsimerman, and the present author, have resulted in progressively stronger results, with a negative secondary term and a power saving error term. The strongest result to date is:

Theorem 1.1. We have

$$N_3(X) = \frac{1}{3\zeta(3)} X + (1 + \sqrt{3}) \frac{4\zeta(1/3)}{5\Gamma(2/3)^3 \zeta(5/3)} X^{5/6} + O(X^{2/3} (\log X)^{2.09}). \tag{1.2}$$

The existence of the (negative) secondary term above had been conjectured by Roberts [Rob01] and Datskovsky and Wright (implicitly in [DW88]). Previously, Shintani [Shi75] had obtained an analogous result for cubic *orders*.

We will give an overview of two questions here:

(1) How can one obtain the asymptotic density result (1.1) at all?
(2) How can one obtain the sharpest error terms possible?

Here we will give an overview of a method using *Shintani zeta functions* and *Landau's method*. This is not the only known method, and indeed we recommend Bhargava, Shankar, and Tsimerman's work [BST13] for an account using *Bhargava's averaging method*, which also obtains the secondary term in (1.2). (See also Y. Suzuki's article [鈴木雄] giving an overview of this method in this volume.)

Bhargava's method applies in significant generality – see Bhargava and Shankar [BS15] for a non-prehomogeneous example (one among many!)

The method discussed here is the most complicated of the known methods – at least, if one is not prepared to black box a lot of relevant background. However, when all the necessary background ingredients are in place, these are the sharpest tools available, leading to the strongest possible error terms.

At the end, we recommend some further reading for the interested reader.

2 Correspondence for cubic rings

Most approaches to the Davenport-Heilbronn theorem proceed by relating cubic fields to **binary cubic forms**, via the correspondences of **Levi** and **Delone–Faddeev**, and **Davenport-Heilbronn**. See [BST13], among other sources, for a more thorough treatment.

First, we recall the relevant definitions. The lattice of *integral binary cubic forms* is defined by

$$V(\mathbb{Z}) := \{au^3 + bu^2v + cuv^2 + dv^3 \ : a, b, c, d \in \mathbb{Z}\}, \tag{2.1}$$

and the *discriminant* of $f(u, v) = au^3 + bu^2v + cuv^2 + dv^3 \in V(\mathbb{Z})$ is given by the equation

$$\mathrm{Disc}(f) = b^2c^2 - 4ac^3 - 4b^3d - 27a^2d^2 + 18abcd. \tag{2.2}$$

Then $\mathrm{Disc}(f)$ is nonzero if and only if $f(u, v)$ factors into distinct linear factors over $\mathbb{C}$.[*1]

The group $\mathrm{GL}_2(\mathbb{Z})$ acts on $V(\mathbb{Z})$ by

$$(\gamma \cdot f)(u, v) = \frac{1}{\det \gamma} f((u, v) \cdot \gamma). \tag{2.3}$$

A cubic form f is *irreducible* if $f(u, v)$ is irreducible as a polynomial over $\mathbb{Q}$, and *nondegenerate* if $\mathrm{Disc}(f) \neq 0$.

One also considers the action (2.3) over other rings such as $\mathbb{R}$, $\mathbb{C}$, or $\mathbb{Z}/n\mathbb{Z}$. Over $\mathbb{R}$ this action has two nondegenerate orbits, corresponding to the sign of $\mathrm{Disc}(f)$. Over $\mathbb{C}$ all of the nondegenerate forms are in the same orbit, making this V a *prehomogeneous vector space.*

The correspondence of Levi [Lev14] and Delone–Faddeev [DF64], as further extended by Gan, Gross, and Savin [GGS02] to include the degenerate case, is as follows:

[*1] If one factor is a scalar multiple of another, then we do not consider these linear factors to be distinct.

Theorem 2.1. There is a canonical, discriminant-preserving bijection between the set of $\mathrm{GL}_2(\mathbb{Z})$-orbits on $V(\mathbb{Z})$ and the set of isomorphism classes of cubic rings. Under this correspondence, irreducible cubic forms correspond to orders in cubic fields, and if a cubic form f corresponds to a cubic ring R, then $\mathrm{Stab}_{\mathrm{GL}_2(\mathbb{Z})}(f)$ is isomorphic to $\mathrm{Aut}(R)$.

To count cubic fields, we count their maximal orders. A cubic ring is the maximal order in a cubic field if and only if it is:

- **nondegenerate**: its discriminant is not zero.
- **an integral domain**, true if and only if f is irreducible.
- **maximal**.

Our counting methods will naturally exclude the degenerate rings, so we can ignore those. We will end up counting all of the maximal rings, including the reducible ones, but the reducible rings correspond to quadratic fields and so are easily counted.

The maximality condition is the most subtle. It turns out that maximality can be checked locally: a cubic ring R is maximal over $\mathbb{Z}$ if and only if $R \otimes_{\mathbb{Z}} \mathbb{Z}_p$ is maximal over $\mathbb{Z}_p$ for all primes p. We call the latter condition *maximality at* p, and the **Davenport–Heilbronn maximality condition** translates it into the language of binary cubic forms:

Proposition 2.2 ([DH71])**.** Under the Levi–Delone–Faddeev correspondence, a cubic ring R is maximal at p if and only if any corresponding cubic form f belongs to the set $U_p \subseteq V(\mathbb{Z})$ for all p, defined by the following conditions:

- the cubic form f is not a multiple of p; and
- there is no $\mathrm{GL}_2(\mathbb{Z})$-transformation of $f(u,v) = au^3 + bu^2v + cuv^2 + dv^3$ such that a is a multiple of p^2 and b is a multiple of p.

See, e.g. [BST13] for a complete proof. To give the idea of a proof, consider the Delone-Faddeev correspondence over $\mathbb{Q}$ rather than over $\mathbb{Z}$. The form f will be nonmaximal at p if and only if there is some other integral binary cubic form f', $\mathrm{GL}_2(\mathbb{Z})$-inequivalent but $\mathrm{GL}_2(\mathbb{Q})$-equivalent to f, and with $\mathrm{Disc}(f') = p^{-k}\mathrm{Disc}(f)$ for some positive integer k. For example, f' can be chosen to be the maximal order in the cubic field (or algebra) corresponding to the order corresponding to f.

We can now easily see half of the if and only if: that if f is any form satisfying the negation of the Davenport-Heilbronn conditions, there exists such a f'. In the first case, $f = pf'$ for some g, and since scalar matrices act by the same scalars, there is our f'.

In the second case, note that

$$\begin{pmatrix} p^{-1} & 0 \\ 0 & 1 \end{pmatrix} \cdot f(u,v) = pf(u/p, v).$$

Then, after suitable $\mathrm{GL}_2(\mathbb{Z})$-transformation, our condition guarantees precisely that $pf(u/p, v)$ has integral coefficients. (And that it has last coefficient divisible by p, but we will not need this.)

We now give a *very brief* introduction to Shintani's zeta function theory [Shi72]. Define the *Shintani zeta functions*

$$\xi^{\pm}(s) = \sum_n a^{\pm}(n) n^{-s} := \sum_{\substack{x \in \mathrm{GL}_2(\mathbb{Z}) \backslash V(\mathbb{Z}) \\ \pm\mathrm{Disc}(x) > 0}} \frac{1}{|\mathrm{Stab}(x)|} |\mathrm{Disc}(x)|^{-s}, \tag{2.4}$$

and note that by the Delone-Faddeev correspondence we have

$$\xi^{\pm}(s) = \sum_{\pm\mathrm{Disc}(R) > 0} \frac{1}{|\,\mathrm{Aut}(R)|} |\mathrm{Disc}(R)|^{-s}, \tag{2.5}$$

where the sum is over cubic rings.

Shintani's work establishes that these are 'nice zeta functions', enjoying analytic continuation to $\mathbb{C}$ apart from simple poles at $s = 1$ and $s = 5/6$ (whose residues Shintani computed), and satisfying a functional equation. By Landau's method, which we will discuss shortly, it is known how to compute the partial sums

$$\sum_{0 < \pm\mathrm{Disc}(R) < X} \frac{1}{|\,\mathrm{Aut}(R)|}. \tag{2.6}$$

Moreover, the factor of $\frac{1}{|\operatorname{Aut}(R)|}$ is well understood (see Proposition 5.1), and can be removed without too much effort.

If our goal is to count cubic *fields*, we must handle the three conditions described previously: nondegeneracy, irreducibility, and maximality.

Nondegeneracy is automatic, as rings with discriminant zero are excluded from the zeta function. Irreducibility will be handled at the end: ignoring this condition, our cubic field count will also include a count of algebras of the form $K \times \mathbb{Q}$, where K is quadratic with $|\mathrm{Disc}(K)| < X$ – and these algebras are easily counted, so their counting function can be subtracted.

The nonmaximality condition is the serious one. Although we cannot adapt our zeta function to count only maximal rings, we may handle finitely many of the maximality conditions at once, and then run a sieve. By work of F. Sato [Sat89] and Datskovsky-Wright [Wri85, DW86], Shintani's zeta function may be extended to define

$$\xi^{\pm}(s, \Phi_m) = \sum_n a^{\pm}(\Phi_m, n) n^{-s} := \sum_{\substack{x \in \mathrm{GL}_2(\mathbb{Z}) \backslash V(\mathbb{Z}) \\ \pm \mathrm{Disc}(x) > 0}} \frac{1}{|\mathrm{Stab}(x)|} \Phi_m(x) |\mathrm{Disc}(x)|^{-s}, \tag{2.7}$$

where Φ_m is any $\mathrm{GL}_2(\mathbb{Z}/m\mathbb{Z})$-invariant function, lifted to a function on $V(\mathbb{Z})$, which we think of as describing some 'local condition' $\pmod{m}$. In particular, we will be interested in the case where m is the square of a product of distinct primes, and Φ_m is the function $\pmod{m}$ which, by Proposition 2.2, detects nonmaximality at *each* prime divisor of m.

3 Landau's method

Landau's method belongs with the following simple observation, known as **Perron's formula**:

Proposition 3.1. Let $c > 0$ be a real number. We have

$$\frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} x^s \frac{ds}{s} = \begin{cases} 1, & \text{if } x > 1, \\ 0, & \text{if } 0 < x < 1. \end{cases} \tag{3.1}$$

The integral is $\frac{1}{2}$ if $x = 1$, but we can avoid caring about this. The proof is a straightforward exercise in complex analysis. Roughly: if $x > 1$, shift the contour all the way to the right, and the integrand decays exponentially as one does this; if $x < 1$, shift the contour all the way to the left, this time picking up the residue from a pole at $s = 0$.

We quickly obtain the following:

Proposition 3.2. Let $L(s) := \sum_{n\geq 1} a(n)n^{-s}$ be a Dirichlet series. Then, we have

$$\sum_{n<X} a(n) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} L(s)X^s \frac{ds}{s}, \tag{3.2}$$

valid for any c for which the Dirichlet series defining $L(s)$ converges absolutely.

To prove this, replace $L(s)$ by its definition, switch the order of summation and integration, and use the previous proposition. (One must worry about convergence, but there are no problems here. We assume that X is not an integer to avoid an additional $\frac{1}{2}a(X)$ term on the left.)

How might one use this? For example, let $d(n)$ denote the *divisor function* of n, counting the number of positive prime divisors. Then it is easily proved that

$$\sum_n d(n)n^{-s} = \zeta(s)^2,$$

and hence it follows that

$$\sum_{n<X} d(n) = \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} \zeta(s)^2 X^s \frac{ds}{s}. \tag{3.3}$$

To bound this, we work *formally* and leave any convergence questions for later. Shift the contour to the line $\Re(s) = \frac{1}{2}$, passing a double pole at $s = 1$, and computing the residue we conclude that

$$\sum_{n<X} a(n) = X \log X + (2\gamma - 1)X + E(X), \tag{3.4}$$

for an error term E satisfying

$$|E(X)| \leq X^{1/2} \int_{1/2-i\infty}^{1/2+i\infty} |\zeta(s)|^2 \left|\frac{ds}{s}\right|. \tag{3.5}$$

Since the integrand is independent of X, as long as it is absolutely convergent we have $E(X) \ll X^{1/2}$.

We notice two things about this computation:

(1) This relies on analytic continuation of $\zeta(s)$ to $\Re(s) = 1/2$, where it is not defined by its Dirichlet series. At the heart of the usual proof of analytic continuation is an application of Poisson summation.
(2) Unfortunately, the convergence issues are serious here. The integral in (3.5) does *not* converge, and the above cannot be rigorously justified. So, one must refine the method so as to allow a rigorous proof.

Before describing Landau's method, which gives a way of handling the convergence issues, we present a sample result. The following variation is due to Lowry-Duda, Taniguchi, and the author [LDTT22]:

Theorem 3.3. Let $\phi(s) = \sum_n a(n)\lambda_n^{-s}$ be a zeta function with nonnegative coefficients, absolutely convergent for $\Re(s) > 1$, enjoying an analytic continuation to $\mathbb{C}$ which is holomorphic away from a simple pole at $s = 1$, and with a 'well behaved' functional equation of degree d relating $\phi(s)$ to $\widehat{\phi}(1-s)$ for a 'dual zeta function' $\widehat{\phi}(s) = \sum_n b(n)\mu_n^{-s}$.

Then, for $X \geq 2$ we have

$$\sum_{\lambda_n < X} a(n) = \mathrm{Res}_{s=1}\big(\phi(s)\big)X + O\Big(X^{\frac{d-1}{d+1}}\delta_1^{\frac{d-1}{d+1}}\widehat{\delta_1}^{\frac{2}{d+1}} + \widehat{\delta_1}\log(X)\Big), \tag{3.6}$$

provided that the error term is bounded by the main term, and where

$$\delta_1 = \mathrm{Res}_{s=1}\big(\phi(s)\big),$$
$$\widehat{\delta_1} = \sup_Z \frac{1}{Z}\sum_{\mu_n < Z} |b(n)|.$$

The implied constant depends on the functional equation, but does not depend further on $\phi(s)$ or the $a(n)$.

The statement generalizes further, for example to allow a double pole as in $\zeta(s)^2$ and (3.5). Unfortunately the most general version of this result, although not so complicated theoretically, is rather complicated to state.

Note that the error term may be written simply as $O(X^{\frac{d-1}{d+1}})$, where the implied constant depends on ϕ. In typical examples, ϕ is fixed and X is allowed to increase, and the precise ϕ-dependence is not of much interest. In this case, our work reduces to results obtained by Landau [Lan12, Lan15] and Chandrasekharan and Narasimhan [CN62]. The novelty of [LDTT22] was to carefully track the ϕ-dependence.

We will apply this to the Shintani zeta functions of the form defined in (2.7), and the uniformity will allow us to track the dependence on Φ_m. From the standpoint of Landau's method, the zeta function dual to $\xi^{\pm}(s, \Phi_m)$ is $m^{4s}\xi^{\pm}(s, \widehat{\Phi_m})$, where we define $\widehat{\Phi_m} : V^*(\mathbb{Z}/m\mathbb{Z}) \to \mathbb{C}$ by the usual Fourier transform formula

$$\widehat{\Phi_m}(x) := \frac{1}{m^4} \sum_{y \in V(\mathbb{Z}/m\mathbb{Z})} \Phi_m(y) \exp\left(2\pi i \cdot \frac{[x, y]}{m}\right), \tag{3.7}$$

and lift $\widehat{\Phi_m}$ to a function on $V^*(\mathbb{Z})$, which we identify with $V(\mathbb{Z})$. (There are some mild technicalities in this identification, which we sweep under the rug.)

Note the factor of m^{4s}! This comes from the functional equation, but if we include it in the definition of the zeta function – as we will do here – then the functional equation will be the same for all Φ_m simultaneously, so that we can use our 'Uniform Landau' Theorem 3.3. We obtain the following:

Theorem 3.4. As described above, given a zeta function pair

$$\xi^{\pm}(s, \Phi_m) := \sum_n a^{\pm}(\Phi_m, n) n^{-s}, \quad \xi^{\pm}(s, \widehat{\Phi_m}) := \sum_n a^{\pm}(\widehat{\Phi_m}, n) n^{-s}.$$

Let

$$\delta_1 = \delta_1(\Phi_m) := \mathrm{Res}_{s=1}\xi^{\pm}(s, \Phi_m) \tag{3.8}$$

and

$$\widehat{\delta_1} = \widehat{\delta_1}(\Phi_m) := m^4 \cdot \sup_N \frac{1}{N} \sum_{\alpha \in \{\pm\}} \sum_{n < N} a^{\alpha}(|\widehat{\Phi_m}|, n), \tag{3.9}$$

Assuming two technical conditions which we omit here, we have

$$\begin{aligned} N^{\pm}(X, \Phi_m) &:= \sum_{n < X} a^{\pm}(\Phi_m, n) \\ &= \sum_{\sigma \in \{1, \frac{5}{6}\}} \frac{X^{\sigma}}{\sigma} \cdot \mathrm{Res}_{s=\sigma}\xi^{\pm}(s, \Phi_m) + O\left(X^{3/5}\delta_1^{3/5}(\widehat{\delta_1})^{2/5}\right). \end{aligned} \tag{3.10}$$

We will end up summing the error term over many different Φ_m. In practice we will have something like $\delta_1(\Phi_m) \ll \frac{1}{m}$ for many Φ_m of interest – and naturally we are happy to see negative exponents of m in the error term! What is clear from this formula is that we want to bound $\widehat{\delta_1}$ – and hence $\widehat{\Phi_m}$ – as much as we can.

Idea of the proof of uniform Landau. To even give a reasonable sketch of the proof would take us too far afield, but we will at least give the idea. As explained above, we have

$$\sum_{n<X} a^{\pm}(\Phi_m, n) = \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} \xi^{\pm}(s, \Phi_m) X^s \frac{ds}{s}$$

and morally the idea is to shift the contour to the left of the critical strip, picking up polar contributions at $s = 1$ and $s = 5/6$, and using the functional equation of the zeta function to estimate and bound the resulting integral.

Unfortunately, the zeta function *grows* along vertical lines to the left of the critical strip. This also implies that known bounds grow *within* the critical strip.

To remedy the convergence issues, we consider the following variation of Perron's formula, based on *Riesz means*, which holds for each positive integer k:

$$\begin{aligned} N(X, \Phi_m, k) &:= \frac{1}{k!} \sum_{n<X} a(\Phi_m, n) \Big(1 - \frac{n}{X}\Big)^k \\ &= \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} \xi(s, \Phi_m) \frac{X^s}{s(s+1)\cdots(s+k)} ds. \end{aligned} \tag{3.11}$$

As the growth of ξ is *polynomial* along vertical lines, by taking k large enough we obtain convergence of the integral if we shift the line to $\Re(s) = c$, for $c \in [0, 1]$, or for c slightly smaller than 0. The above strategy then works as it was earlier described.

We could simply declare victory for a variant of our cubic field counting problem, where we count each field with $|\mathrm{Disc}(K)| < X$ with weight $(1 - |\mathrm{Disc}(K)|/X)^k$. This is an example of a general principle in number theory, that incorporating smooth weights often leads to simpler analysis and/or better error terms. (At heart, the general principle comes from Fourier analysis: the smoother the function, the more rapidly its Fourier transform decays.) See Shankar, Södergren, and Templier [SST23] for an example where the authors 'declare victory' with a different smoothing function, and then obtain striking results on the central values of the associated Dedekind zeta functions.

Instead, we recover the unsmoothed estimate, up to an error term. Define the kth finite difference operator Δ_y^k by

$$\Delta_y^k F(X) = \sum_{\nu=0}^{k} (-1)^{k-\nu} \binom{k}{\nu} F(X + \nu y), \tag{3.12}$$

and then in (3.11) we have

$$N(X, \Phi_m) = y^{-k} \Delta_y^k \big(X^k N(X, \Phi_m, k) \big) + O\Big(\sum_{X \le n \le X + ky} |a(\Phi_m, n)| \Big). \tag{3.13}$$

Provided that one can also analyze the finite differences of the shifted integral in (3.11), we are able to obtain the theorem.

4 Equidistribution and exponential sums

Recall from (3.10) that we can obtain good error terms by bounding the partial sums of $|\widehat{\Phi_m}(x)|$ where by (3.7) we defined

$$\widehat{\Phi_m}(x) := \frac{1}{m^4} \sum_{y \in V(\mathbb{Z}/m\mathbb{Z})} \Phi_m(y) \exp\left(2\pi i \cdot \frac{[x, y]}{m} \right). \tag{4.1}$$

It turned out that Taniguchi and I got very lucky, in that when $\Phi_m = \Psi_{p^2}$ is the characteristic function of those binary cubic forms which are nonmaximal at p, we obtained the following explicit formula in [TT13a]:

Theorem 4.1. The Fourier transform of Ψ_{p^2} is given as follows:

(1) Let $b \in pV(\mathbb{Z}/p^2\mathbb{Z})$. We write $b = pb'$ and regard b' as an element of $V(\mathbb{Z}/p\mathbb{Z})$. Then

$$\widehat{\Psi_{p^2}}(pb') = \begin{cases} p^{-2} + p^{-3} - p^{-5} & b' : \text{of type } (0), \\ p^{-3} - p^{-5} & b' : \text{of type } (1^3), (1^2 1), \\ -p^{-5} & b' : \text{of type } (111), (21), (3). \end{cases}$$

(2) For $b \in V(\mathbb{Z}/p^2\mathbb{Z}) \setminus pV(\mathbb{Z}/p^2\mathbb{Z})$,

$$\widehat{\Psi_{p^2}}(b) = \begin{cases} p^{-3} - p^{-5} & b: \text{ of type } (1^3_{**}), \\ -p^{-5} & b: \text{ of type } (1^3_{*}),\ (1^3_{\max}), \\ 0 & \text{otherwise.} \end{cases}$$

We omit the precise definitions of notation such as (1^3_{**}). What is most important is that the 'otherwise' case is by far the most common, so that we have

$$|\widehat{\Psi_{p^2}}(b)| \approx p^{-7} \tag{4.2}$$

on average. As the error term (3.10) contains sums of $|\widehat{\Psi_{m^2}}(b)|$ over many different b, the power savings here is very visible in our final error terms!

Taniguchi and I developed two methods for proving formulas such as the one above. Our work in [TT13a] was based on work of Mori [Mor10]; see also Hough [Hou20] for a formidable computation of this type. In [TT20], we developed a simpler method which works in many cases.

To give the reader some idea of how such formulas may be proved, we will give a complete proof of a simpler formula, using the method in [TT20].

Proposition 4.2. Let $w_p : V(\mathbb{F}_p) \to \mathbb{C}$ be the counting function of the number of roots of $v \in V(\mathbb{F}_p)$ in $\mathbb{P}^1(\mathbb{F}_p)$. Then, assuming that $p \neq 3$ we have

$$\widehat{w_p}(v) = \begin{cases} 1 + p^{-1} & v = 0, \\ p^{-1} & v \neq 0 \text{ and } v \text{ has a triple root in } \mathbb{P}^1(\mathbb{F}_p), \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

It turns out that this formula is itself of significant interest! But for now we notice the parallel structure: better than square root cancellation in average, with the largest values confined to the most singular orbits.

Here is the proof. First we note that there is a **bilinear form** defined on V, by

$$[x, x'] := aa' + bb'/3 + cc'/3 + dd', \tag{4.4}$$

which satisfies $[gx, g^{-T}x'] = [x, x']$ identically, as can be checked by hand or by more highbrow methods. This formula is formally true over $\mathbb{Z}$ or over $\mathbb{F}_p$ ($p \neq 3$), and we already used it implicitly to identify $V^*(\mathbb{Z})$ with (a sublattice of) $V(\mathbb{Z})$.

Now write $\langle n\rangle := \exp(2\pi i n/p)$, and write Φ_p for the characteristic function of the orbit (1^3): those nonzero elements of $V(\mathbb{F}_p)$ which have a triple root. By Fourier inversion, it suffices to compute the Fourier transform of the right side of (4.3), and thus to compute $\widehat{\Phi_p}$.

Using the facts that (1^3) is a single $\mathrm{GL}_2(\mathbb{F}_p)$-orbit, and that our bilinear form is $\mathrm{SL}_2(\mathbb{F}_p)$-invariant, we compute that

$$\begin{aligned} p^4\widehat{\Phi_p}(y) &= \frac{1}{p^2-p}\sum_{g\in\mathrm{SL}_2(\mathbb{F}_p)}\sum_{t\in\mathbb{F}_p^\times}\langle[g\cdot(t,0,0,0),y]\rangle \\ &= \frac{1}{p^2-p}\sum_{g\in\mathrm{SL}_2(\mathbb{F}_p)}\sum_{t\in\mathbb{F}_p^\times}\langle[(t,0,0,0),g^Ty]\rangle \end{aligned}$$

The inner sum is equal to $p-1$ if $[1:0]\in\mathbb{P}^1(\mathbb{F}_p)$ is a root of g^Ty, and -1 if it is not. Equivalently, the inner sum is equal to $p-1$ if g^Ty is in the subspace $(0,*,*,*)$ defined by $a=0$, and -1 if it is not.

For each root α of y, counted with multiplicity, $[1:0]$ will be a root of g^Ty for $\frac{|\mathrm{SL}_2(\mathbb{F}_p)|}{p+1} = p^2-p$ elements $g\in\mathrm{SL}_2(\mathbb{F}_p)$, so that

$$p^4\widehat{\Phi_p}(y) = \frac{1}{p^2-p}\cdot(p^2-p)\cdot\Big(pw_p(x)-(p+1)\Big).$$

Proposition 4.2 now follows easily.

We can isolate the following principle from the proof. Let $W=(*,0,0,0)$ be the subspace of binary cubic forms which are multiples of u^3; this consists of the zero form and forms in (1^3). We then have $W^\perp=(0,*,*,*)$, and all forms in this subspace have a root. Translating each of these subspaces around by all of SL_2 or GL_2, we obtain 'dual' functions which depend only on the GL_2-orbits of y, and whose Fourier transforms are related by the above formula.

Generalizing this recipe, and following [TT20], suppose we have:

- A prehomogeneous vector space (G,V);
- A finite number of orbits, which we list as $\mathcal{O}_1,\ldots\mathcal{O}_r$. We write $|\mathcal{O}_i|$ for their cardinalities and e_i for their characteristic functions.
- A bilinear form $[-.-]$ satisfying an analogue of the G-invariance property described above.

Then for any subspace W of V, we have

$$\sum_{1\le i\le r} \frac{|\mathcal{O}_i \cap W|}{|\mathcal{O}_i|} \cdot \widehat{e_i} = \frac{|W|}{|V|} \sum_{1\le i\le r} \frac{|\mathcal{O}_i \cap W^\perp|}{|\mathcal{O}_i|} \cdot e_i. \tag{4.5}$$

Then, both sides are functions of the G-*orbits* on V. If we choose enough different W, so that the functions on the left span the r-dimensional vector space of functions of the G-orbits of V, then we obtain all the $\widehat{e_i}$ – and hence an explicit formula for the Fourier transform of any G-invariant function.

5 Putting it all together

We now see how to put all of these ingredients together. For convenience, we will track the error terms a bit less carefully than in (1.2), and obtain an error term of $O(X^{2/3+\epsilon})$.

Let, for each squarefree q, $\Psi_{q^2} : V(\mathbb{Z}/q^2\mathbb{Z}) \to \mathbb{C}$ denote the characteristic function of those binary cubic forms which are nonmaximal at every prime dividing q. The Levi–Delone–Faddeev correspondence, the Davenport–Heilbronn correspondence, and inclusion-exclusion give

$$N^{\pm}_{\le 3}(X) = \sum_q \mu(q) N^{\pm}(X, \Psi_{q^2}), \tag{5.1}$$

where:

- As above,

$$\begin{aligned} N^{\pm}(X, \Psi_{q^2}) &:= \sum_{n<X} a^{\pm}(\Psi_{q^2}, n), \\ \xi^{\pm}(s, \Psi_{q^2}) &= \sum_n a^{\pm}(\Psi_{q^2}, n) n^{-s} \\ &:= \sum_{\substack{x\in \mathrm{GL}_2(\mathbb{Z})\backslash V(\mathbb{Z}) \\ \pm\mathrm{Disc}(x)>0}} \frac{1}{|\mathrm{Stab}(x)|} \Psi_{q^2}(x) |\mathrm{Disc}(x)|^{-s}. \end{aligned}$$

  Equivalently, $N^{\pm}(X, \Psi_{q^2})$ is the number of binary cubic forms x, weighted by $|\mathrm{Stab}(x)|^{-1}$, with $0 < \pm\mathrm{Disc}(x) < X$, which are nonmaximal at q;

- $N^{\pm}_{\leq 3}(X)$ is the number of maximal cubic rings R with $0 < \pm\mathrm{Disc}(R) < X$, each weighted by $|\operatorname{Aut}(R)|^{-1}$. These are in bijection with the following sets: cubic fields K; algebras $L \times \mathbb{Q}$ where L is a quadratic field; and $\mathbb{Q} \times \mathbb{Q} \times \mathbb{Q}$.

The factor of $|\operatorname{Aut}(R)|$ is a little bit annoying, but not actually difficult to deal with:

Proposition 5.1. Let R be the maximal order in K, $L \times \mathbb{Q}$, or $\mathbb{Q} \times \mathbb{Q} \times \mathbb{Q}$ as above. Then:

- If K is an S_3-cubic field, then $|\operatorname{Aut}(R)| = |\operatorname{Aut}(K)| = 1$.
- If K is an cyclic cubic field, then $|\operatorname{Aut}(R)| = |\operatorname{Aut}(K)| = 3$.
- If R is a maximal order in $L \times \mathbb{Q}$ with L quadratic, then $|\operatorname{Aut}(R)| = |\operatorname{Aut}(L \times \mathbb{Q})| = |\operatorname{Aut}(L)| = 2$.
- If R is the maximal order in $\mathbb{Q} \times \mathbb{Q} \times \mathbb{Q}$, then $|\operatorname{Aut}(R)| = 6$. (There is a unique such example – so the contribution to the asymptotics is negligible.)

As there are only $O(X^{1/2})$ cyclic cubic fields K with $|\mathrm{Disc}(K)| < X$, we thus have

$$N^{\pm}_3(X) = N^{\pm}_{\leq 3}(X) - \frac{1}{2} N^{\pm}_2(X) + O(X^{1/2}) \tag{5.2}$$

where $N^{\pm}_3(X)$ and $N^{\pm}_2(X)$ count the number of cubic and quadratic fields, respectively, counted without weighting, with discriminant bounded by X.

We now split the sum into two parts in accordance with whether $q \leq Q$ or $q > Q$, and apply Landau's method (Theorem 3.4) for the former. We obtain

$$\begin{aligned} N^{\pm}_{\leq 3}(X, \Sigma) &= \sum_{q \leq Q} \mu(q) N^{\pm}(X, \Psi_{q^2}) + \sum_{q > Q} \mu(q) N^{\pm}(X, \Psi_{q^2}) \\ &= \sum_{\sigma \in \{1, \frac{5}{6}\}} \frac{X^\sigma}{\sigma} \sum_{q \leq Q} \mu(q) \cdot \mathrm{Res}_{s=\sigma} \xi^{\pm}(s, \Psi_{q^2}) + O\left(E_2 + E_3\right), \\ &= \sum_{\sigma \in \{1, \frac{5}{6}\}} \frac{X^\sigma}{\sigma} \sum_{q=1}^{\infty} \mu(q) \cdot \mathrm{Res}_{s=\sigma} \xi^{\pm}(s, \Psi_{q^2}) + O\left(E_1 + E_2 + E_3\right), \end{aligned}$$

with

$$\begin{aligned}
E_1 :=& \sum_{\sigma \in \{1, \frac{5}{6}\}} X^{\sigma} \sum_{q > Q} \left| \mathrm{Res}_{s=\sigma} \xi^{\pm}(s, \Psi_{q^2}) \right|, \\
E_2 :=& X^{\frac{3}{5}} \sum_{q \leq Q} \mathrm{Res}_{s=1} \xi^{\pm}(s, \Psi_{q^2})^{\frac{3}{5}} \widehat{\delta}_1(\Psi_{q^2})^{\frac{2}{5}}, \\
E_3 :=& \sum_{q > Q} N^{\pm}(X, \Psi_{q^2}).
\end{aligned}$$

We handle the residues first; we have

$$\mathrm{Res}_{s=1} \xi^{\pm}(s, \Psi_{q^2}) = \alpha^{\pm} \prod_{p \mid q} (p^{-2} + p^{-3} - p^{-5}) + \beta \prod_{p \mid q} (2p^{-2} - p^{-4}), \tag{5.3}$$

where

$$\alpha^{+} = \frac{\pi^2}{72}, \quad \alpha^{-} = \frac{\pi^2}{24}, \quad \beta = \frac{\pi^2}{24}, \tag{5.4}$$

and the residues at $s = 5/6$ are smaller. Roughly, if not *quite* technically, we have $\mathrm{Res}_{s=1} \xi^{\pm}(s, \Psi_q) \asymp \frac{1}{q^2}$. We obtain

$$E_1 \ll \sum_{q > Q} \frac{X}{q^{2-\epsilon}} \ll \frac{X}{Q^{1-\epsilon}}.$$

For E_3, we have a *tail estimate*

$$N^{\pm}(X, \Psi_{q^2}) \ll X / q^{2-\epsilon}, \tag{5.5}$$

and hence we get

$$E_3 \ll \sum_{q > Q} \frac{X}{q^2} \ll \frac{X}{Q^{1-\epsilon}}.$$

The existence of the tail estimate (5.5) is probably the most unique feature of this problem; in analogous situations, very often tail estimates are expected but cannot be proved.

The proof of (5.5) is algebraic rather than analytic, and it is not difficult to prove. One must understand, for each maximal cubic ring R, how many nonmaximal cubic rings it can contain with index q. See, for example, Proposition 29 of [BST13] for an elementary upper bound, which is stated for q prime but which readily generalizes to q squarefree. More precise equalities can also be obtained: see Proposition 33 of [BST13] or Section 2 of [DW88].

The trickiest part of the argument – at least that was not known prior to [TT13b], is to verify that $\widehat{\delta_1}(\Psi_{q^2}) \ll q^{1+\epsilon}$. We will not try to explain the messy details here, but the heart of the argument is that

$$\widehat{\delta_1}(\Psi_{q^2}) \ll q^{8-7+\epsilon}, \tag{5.6}$$

where this -7 is the same -7 as in (4.2). Roughly speaking, heuristically speaking one expects that the -7 of (4.2) *should* lead to a -7 in (5.6), and some of the more technical parts of [BTT] are dedicated to demonstrating that the details can be made to work.[*2]

Given that, we see that

$$E_2 \ll X^{\frac{3}{5}} \sum_{q \leq Q} q^{-6/5+\epsilon} \cdot q^{2/5+\epsilon} \ll X^{\frac{3}{5}} Q^{1/5+\epsilon},$$

and so we obtain a final error term of

$$E_1 + E_2 + E_3 \ll X^{\epsilon}\Big(\frac{X}{Q} + X^{\frac{3}{5}} Q^{1/5} + \frac{X}{Q}\Big).$$

We optimize by taking $Q = X^{1/3}$ and getting an error term of $O(X^{2/3+\epsilon})$.

Further reading

Within this volume, we recommend three other contributions on closely related topics: M. Suzuki's discussion [鈴木美] of Hough's work [Hou19] on the shape of cubic fields; Y. Suzuki's work [鈴木雄] describing Bhargava's averaging method, proving similar results without the use of zeta functions; and Yamamoto's note [山本] describing O'Dorney's work [O'D] on 'algebraic functional equations', which we describe a bit more below.

Some additional references (a far from exhaustive list!) are:

- For the proof of Theorem 1.1, along the lines presented here, see Bhargava, Taniguchi, and the author's work [BTT]. This paper also presents a version

[*2] See [TT13b, Theorem 3.1] for a direct proof of (5.6). In [BTT], (5.6) is proved on average over q, in conjunction with a variant of Landau's method that permits this. This allowed us to simultaneously obtain an analogue to (1.2) for 3-torsion in quadratic fields, where the analogue of (5.6) can't be proved directly but can be proved on average.

of Theorem 1.1 with 'local conditions' – for example, if one wants to count cubic fields where 5 is ramified and 7 is inert.

This 'local conditions' version has various applications to other arithmetic statistics problems; see [BTT] for a summary and further references.

- Taniguchi and the author's previous work [TT13b] also proves a variation, counting cubic fields in arithmetic progressions. Here we were able to demonstrate unexpected biases. For example, in the functions counting cubic fields K with $\mathrm{Disc}(K) \equiv a \pmod 7$, the secondary term is different for every a!
- For an alternative treatment of Theorem 1.1 with a somewhat larger error term, see Bhargava, Shankar, and Tsimerman [BST13]. Their approach avoids the zeta function theory, instead applying Bhargava's averaging method, and is much more self-contained. We also recommend [BST13] for a particularly readable treatment of the Delone-Faddeev and Davenport-Heilbronn correspondences.

  Their methods generalize widely; see, for example, Bhargava and Shankar [BS15] for one of many examples of spectacular results that can be thus obtained.
- Yet another alternative treatment involves *smoothing* the sums; see Shankar, Södergren, and Templier [SST23] for such a variation of Theorem 1.1.
- Analogues of these questions are also interesting in the function field setting, where some algebro-geometric considerations 'explain' the secondary term. See Zhao [Zha13].
- For background on Shintani zeta functions, we recommend Shintani's original paper [Shi72]. For a more comprehensive overview to prehomogeneous vector spaces and their zeta functions, see Kimura's book: [木村 98] in Japanese, or [Kim03] in English translation. See also Sato-Shintani [SS74] for the landmark paper on which much of Kimura's book is based, and Yukie's book [Yuk93] for a research monograph on this more general family of Shintani zeta functions, most notably treating the 'quartic case' of pairs of ternary quadratic forms.

- Finally, the Shintani zeta functions satisfy a stunning – and to the author, totally surprising – "algebraic functional equation", proved via class field theory instead of complex and Fourier analysis. This was conjectured by Ohno [Ohn97] and proved by Nakagawa [Nak98]; see Gao [Gao18] and O'Dorney [O'D17] for further proofs. See also [O'D] for further results by O'Dorney in this vein, or Yamamoto [**?**] for an overview of O'Dorney's work.

Acknowledgments

The author thanks Takashi Taniguchi for helpful comments, for the invitation to participate in the Summer School, and also for the invitation to visit him in Japan, to participate in person in a preliminary summer school earlier in the year.

The author is partially supported by the National Science Foundation under Grant No. DMS-2101874.

参考文献

[BBP10] Karim Belabas, Manjul Bhargava, and Carl Pomerance. Error estimates for the Davenport-Heilbronn theorems. *Duke Math. J.*, 153(1):173–210, 2010.

[Bel99] Karim Belabas. On the mean 3-rank of quadratic fields. *Compositio Math.*, 118(1):1–9, 1999.

[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.

[BTT] Manjul Bhargava, Takashi Taniguchi, and Frank Thorne. Improved error estimates for the Davenport-Heilbronn theorems. *Math. Ann.*

To appear.

[CN62] K. Chandrasekharan and Raghavan Narasimhan. Functional equations with multiple gamma factors and the average order of arithmetical functions. *Ann. of Math. (2)*, 76:93–136, 1962.

[DF64] B. N. Delone and D. K. Faddeev. *The theory of irrationalities of the third degree.* Translations of Mathematical Monographs, Vol. 10. American Mathematical Society, Providence, R.I., 1964.

[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.

[DW86] Boris Datskovsky and David J. Wright. The adelic zeta function associated to the space of binary cubic forms. II. Local theory. *J. Reine Angew. Math.*, 367:27–75, 1986.

[DW88] Boris Datskovsky and David J. Wright. Density of discriminants of cubic extensions. *J. Reine Angew. Math.*, 386:116–138, 1988.

[Gao18] Xia Gao. On the Ohno-Nakagawa theorem. *J. Number Theory*, 189:186–210, 2018.

[GGS02] Wee Teck Gan, Benedict Gross, and Gordan Savin. Fourier coefficients of modular forms on G_2. *Duke Math. J.*, 115(1):105–169, 2002.

[Hou19] Robert Hough. The shape of cubic fields. *Res. Math. Sci.*, 6(3):Paper No. 25,, 2019.

[Hou20] Robert D. Hough. The local zeta function in enumerating quartic fields. *J. Number Theory*, 210:1–131, 2020.

[Kim03] Tatsuo Kimura. *Introduction to prehomogeneous vector spaces*, volume 215 of *Translations of Mathematical Monographs*. American Mathematical Society, Providence, RI, 2003. Translated from the 1998 Japanese original by Makoto Nagura and Tsuyoshi Niitani and revised by the author.

[Lan12] Edmund Landau. Über die Anzahl der Gitterpunkte in gewissen Bereichen. *Nachrichten von der Gessellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pages 687–770, 1912.

[Lan15] Edmund Landau. Über die Anzahl der Gitterpunkte in gewissen Bereichen. Zweite abhandlung. *Nachrichten von der Gessellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse*, pages 209–243, 1915.

[LDTT22] David Lowry-Duda, Takashi Taniguchi, and Frank Thorne. Uniform bounds for lattice point counting and partial sums of zeta functions. *Math. Z.*, 300(3):2571–2590, 2022.

[Lev14] F. Levi. Kubische Zahlkörper und binäre kubische Formenklassen. *Leipz. Ber.*, 66:26–37, 1914.

[Mor10] Shingo Mori. Orbital Gauss sums associated with the space of binary cubic forms over a finite field. *RIMS Kôkyûroku*, 1715:32–36, 2010.

[Nak98] Jin Nakagawa. On the relations among the class numbers of binary cubic forms. *Invent. Math.*, 134(1):101–138, 1998.

[O'D] Evan O'Dorney. Reflection theorems for number rings generalizing the Ohno-Nakagawa identity. Preprint (2022), available at `https://arxiv.org/abs/2111.09784`.

[O'D17] Evan O'Dorney. On a remarkable identity in class numbers of cubic rings. *J. Number Theory*, 176:302–332, 2017.

[Ohn97] Yasuo Ohno. A conjecture on coincidence among the zeta functions associated with the space of binary cubic forms. *Amer. J. Math.*, 119(5):1083–1094, 1997.

[Rob01] David P. Roberts. Density of cubic field discriminants. *Math. Comp.*, 70(236):1699–1705, 2001.

[Sat89] Fumihiro Satō. On functional equations of zeta distributions. In *Automorphic forms and geometry of arithmetic varieties*, volume 15 of *Adv. Stud. Pure Math.*, pages 465–508. Academic Press, Boston, MA, 1989.

[Shi72] Takuro Shintani. On Dirichlet series whose coefficients are class numbers of integral binary cubic forms. *J. Math. Soc. Japan*, 24:132–188, 1972.

[Shi75] Takuro Shintani. On zeta-functions associated with the vector space of quadratic forms. *J. Fac. Sci. Univ. Tokyo Sect. IA Math.*, 22:25–65, 1975.

[SS74] Mikio Sato and Takuro Shintani. On zeta functions associated with prehomogeneous vector spaces. *Ann. of Math. (2)*, 100:131–170, 1974.

[SST23] Arul Shankar, Anders Södergren, and Nicolas Templier. Central values of zeta functions of non-Galois cubic fields, 2023.

[TT13a] Takashi Taniguchi and Frank Thorne. Orbital L-functions for the space of binary cubic forms. *Canad. J. Math.*, 65(6):1320–1383, 2013.

[TT13b] Takashi Taniguchi and Frank Thorne. Secondary terms in counting functions for cubic fields. *Duke Math. J.*, 162(13):2451–2508, 2013.

[TT20] Takashi Taniguchi and Frank Thorne. Orbital exponential sums for prehomogeneous vector spaces. *Amer. J. Math.*, 142(1):177–213, 2020.

[Wri85] David J. Wright. The adelic zeta function associated to the space of binary cubic forms. I. Global theory. *Math. Ann.*, 270(4):503–534, 1985.

[Yuk93] Akihiko Yukie. *Shintani zeta functions*, volume 183 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, Cambridge, 1993.

[Zha13] Yongqiang Zhao. On sieve methods for varieties over finite fields. 2013. Thesis (Ph.D.)–University of Wisconsin-Madison.

[木村 98] 木村達雄. **概均質ベクトル空間**. 岩波書店, 1998.

[鈴木美] 鈴木美裕. 保型形式と概均質ゼータ関数. **本報告集**, 2023.

[鈴木雄] 鈴木雄太. 整数軌道の数え上げ：数の幾何と平均法. **本報告集**, 2023.

[山本] 山本修司. 大野・中川型鏡映定理と Poitou-Tate 双対性. **本報告集**, 2023.