Arithmetic Statistics and the Cohen–Lenstra Principle

Robin Ammon

University of Glasgow

23 November 2023

Arithmetic Statistics

What is Arithmetic Statistics?

Arithmetic Statistics: Deal with statistical questions about number-theoretic objects:

- primes
- number fields
- elliptic curves
- quadratic forms
- ...

Questions one can ask:

- How many of these objects are there?
  - How many primes are there? Euclid: ∞ many
- How many up to a given 'height'?
  - What is $|\{\, p \leq x \mid p \text{ prime} \,\}|$? Prime number theorem: $\sim \frac{x}{\log x}$

What is Arithmetic Statistics?

Arithmetic Statistics: Deal with statistical questions about number-theoretic objects: primes, number fields, elliptic curves, ...

Questions one can ask:

- How many of these objects are there?
- How many up to a given 'height'?
- How often does a certain event occur?
  - Q1: How often is a prime the sum of two squares?
  - Q2: When $d > 0$ varies, how often is $x^2 - dy^2 = 4$ soluble over $\mathbb{Z}$?
- How are the objects distributed?
  - Q3: Distribution of ideal class groups
  - Distribution of ranks of elliptic curves
- ...

What is Arithmetic Statistics?

Arithmetic Statistics: Deal with statistical questions about number-theoretic objects: primes, number fields, elliptic curves, ...

Motivation:

- Statistical behaviour reflects properties of the objects
- So: Understand statistics $\Longrightarrow$ understand objects
- Shift of perspective: study all objects at once rather than individually

Note: deal with deterministic objects, so can prove statements

- Theorems are usually very powerful as they contain a lot of information about the objects
- Findind a good conjecture is often hard

Q1: How often is a prime the sum of two squares?

We have

$$2 = 1 + 1, \quad 5 = 1 + 4, \quad 13 = 4 + 9, \quad 17 = 1 + 16, \quad 29 = 4 + 25$$

Suggests: $p = \square + \square$ when $p = 2$ and $p \equiv 1 \bmod 4$.

If p odd, $p = a^2 + b^2$, then wlog a even, b odd, so $p \equiv 1 \bmod 4$.

Converse? Observe: $a^2 + b^2 = |a + bi|^2$ where $a + bi \in \mathbb{Z}[i] \subseteq \mathbb{C}$.

We use the following elementary facts:

(1) If $p \equiv 1 \bmod 4$, then $x^2 + 1$ has a zero mod p.

(2) The ring $\mathbb{Z}[i]$ is a UFD.

Q1: How often is a prime the sum of two squares?

Observe: $a^2 + b^2 = |a + bi|^2$ where $a + bi \in \mathbb{Z}[i] \subseteq \mathbb{C}$.

We use the following elementary facts:

(1) If $p \equiv 1 \bmod 4$, then $x^2 + 1$ has a zero mod p.

(2) The ring $\mathbb{Z}[i]$ is a UFD.

Let $p \equiv 1 \bmod 4$. By (1), there is c s.t.

$$p \mid c^2 + 1 = (c + i)(c - i).$$

But $p \nmid c + i$ and $p \nmid c - i$. So p not a prime element of $\mathbb{Z}[i]$.

By (2) there are non-units $\alpha = a + bi, \beta \in \mathbb{Z}[i]$ s.t. $p = \alpha\beta$

$$\overset{|\cdot|^2}{\Longrightarrow} \quad p^2 = \underbrace{|\alpha|^2}_{\in \mathbb{Z}_{>1}} \cdot \underbrace{|\beta|^2}_{\in \mathbb{Z}_{>1}} \qquad \Longrightarrow \quad p = |\alpha|^2 = a^2 + b^2.$$

Q1: How often is a prime the sum of two squares?

So: $2 \neq p$ sum of two squares $\iff p \equiv 1 \bmod 4$.

Theorem (Primes in arithmetic progressions)

There are infinitely many primes p with $p \equiv 1 \bmod 4$. In fact,

$$\lim_{x \to \infty} \frac{|\{ p \leq x \,|\, p \text{ prime}, \, p \equiv 1 \bmod 4 \}|}{|\{ p \leq x \,|\, p \text{ prime} \}|} = \frac{1}{2}.$$

So half of the primes are a sum of two squares!

Ideal Class Groups

Number Fields and Rings of Integers

Seen: when studying integer solutions of equations like $p = x^2 + y^2$, helpful to work with rings slightly bigger than $\mathbb{Z}$.

Example

- Consider $x^2 - dy^2 = n$ as in Q2. Then

$$n = x^2 - dy^2 = (x + \sqrt{d}y)(x - \sqrt{d}y) \qquad \rightsquigarrow \qquad \mathbb{Z}[\sqrt{d}]$$

- Consider $x^n + y^n = z^n$ as in Fermat's Last Theorem. Then

$$y^n = z^n - x^n = (z - x)(z - \zeta_n x) \cdots (z - \zeta_n^{n-1} x) \qquad \rightsquigarrow \qquad \mathbb{Z}[\zeta_n]$$

Number Fields and Rings of Integers

Seen: when studying integer solutions of equations like $p = x^2 + y^2$, helpful to work with rings slightly bigger than $\mathbb{Z}$.

In general, in a finite extension K of $\mathbb{Q}$ (a number field) consider the ring

$$\mathbb{Z}_K := \text{integral closure of } \mathbb{Z} \text{ in } K$$

called the ring of integers of K.

$$\begin{array}{ccc} \mathbb{Q} & \subseteq & K \\ | & & | \\ \mathbb{Z} & \subseteq & \mathbb{Z}_K \end{array}$$

E.g. for $K = \mathbb{Q}(i)$ have $\mathbb{Z}_K = \mathbb{Z}[i]$.

Number Fields and Rings of Integers

In general, in a finite extension K of $\mathbb{Q}$ (a number field) consider the ring

$$\mathbb{Z}_K := \text{integral closure of } \mathbb{Z} \text{ in } K$$

called the ring of integers of K.

$$\begin{array}{ccc} \mathbb{Q} & \subseteq & K \\ | & & | \\ \mathbb{Z} & \subseteq & \mathbb{Z}_K \end{array}$$

E.g. for $K = \mathbb{Q}(i)$ have $\mathbb{Z}_K = \mathbb{Z}[i]$.
Here usually consider $K = \mathbb{Q}(\sqrt{d})$ with $d < 0$, an *imaginary quadratic field*

Crucial to know: is $\mathbb{Z}_K$ a UFD?

Characterised by ideal class group!

Ideal Class Groups

For $0 \neq I, J \trianglelefteq \mathbb{Z}_K$ define

$$I \sim J \quad :\Longleftrightarrow \quad (a)I = (b)J \text{ for some } a, b \in \mathbb{Z}_K \setminus \{0\}$$

Theorem

$\mathrm{Cl}_K := \{0 \neq I \trianglelefteq \mathbb{Z}_K\} / \sim$ is a finite abelian group, the ideal class group.
It satisfies:

$$\mathbb{Z}_K \text{ UFD} \quad \Longleftrightarrow \quad \mathbb{Z}_K \text{ PID} \quad \Longleftrightarrow \quad \mathrm{Cl}_K = 0.$$

- Cl_K measures failure of unique factorisation in $\mathbb{Z}_K$
- also related to solubility of equations over $\mathbb{Z}$
- central object in number theory
- But: not well understood!

Ideal Class Groups

$\mathrm{Cl}_K := \{0 \neq I \trianglelefteq \mathbb{Z}_K\} / \sim$ is central object, but not well understood.

Cohen, Lenstra: Study the distribution! ($\rightsquigarrow$ Q3)

G finite abelian group. If we pick a random imaginary quadratic field $\mathbb{Q}(\sqrt{d})$, $d < 0$, what is

$$\mathbb{P}(\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})} \cong G)?$$

Computational Data

- $\mathbb{P}(\left|\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}\right| \text{ is divisible by } 3) \approx 0.44$
- $\mathbb{P}(\text{the 3-Sylow subgroup } \mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}[3^\infty] \text{ is cyclic}) \approx 0.98$

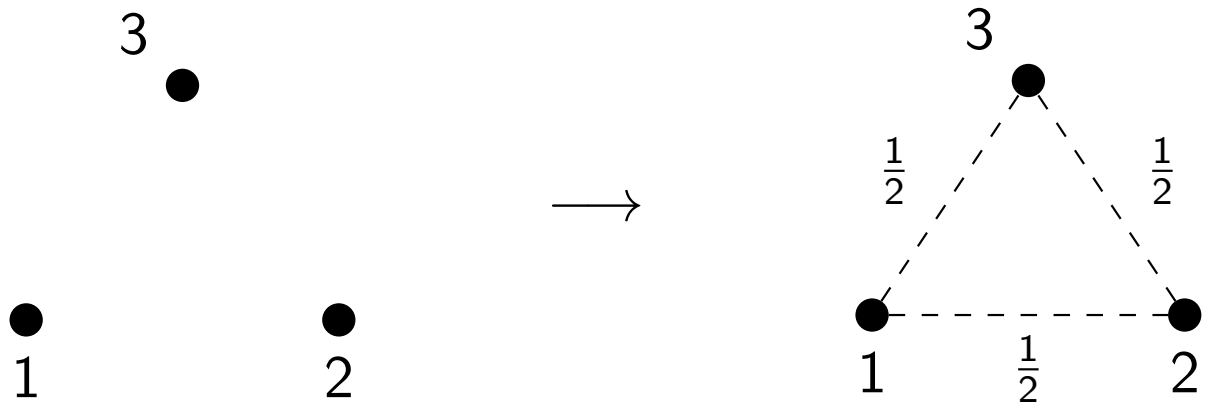
Looks strange! But what would we expect?

How do random groups behave?

Distribution of Random Algebraic Objects

Distribution of Random Graphs

Given 3 vertices, build a random graph R by independently inserting an edge between two vertices with probability $\frac{1}{2}$

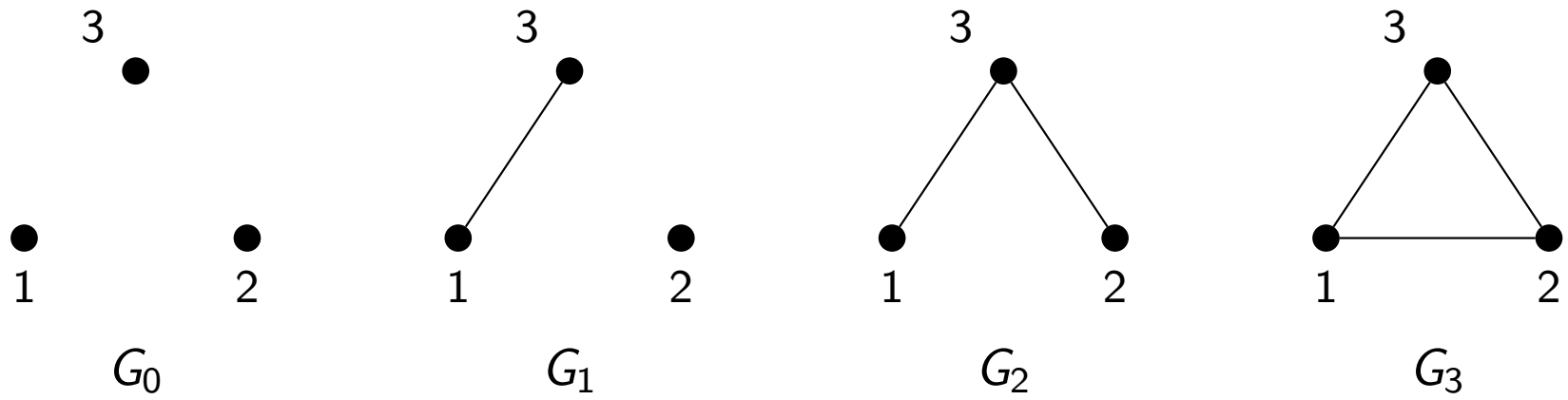


A graph isomorphism between two such graphs is $\sigma \in S_3$ such that vertices i and j are adjacent if and only if $\sigma(i)$ and $\sigma(j)$ are

Distribution of Random Graphs

Given 3 vertices, build a random graph R by independently inserting an edge between two vertices with probability $\frac{1}{2}$

Possible outcomes up to isomorphism:



What is $\mathbb{P}(R \cong G_i)$?

Distribution of Random Graphs

What is $\mathbb{P}(R \cong G_i)$?

- There are $2 \cdot 2 \cdot 2 = 8$ possible outcomes.
- Let $\mathrm{Iso}(G_i) = \{\text{outcomes } \cong G_i\}$, then

$$\mathbb{P}(R \cong G_i) = \frac{|\mathrm{Iso}(G_i)|}{8}.$$

- S_3 operates transitively on $\mathrm{Iso}(G_i)$ and stabiliser of $H \in \mathrm{Iso}(G_i)$ is $\mathrm{Aut}\, H \cong \mathrm{Aut}\, G_i$. Orbit-Stabiliser Thm: $S_3/\mathrm{Aut}(G_i) \overset{1:1}{\longleftrightarrow} \mathrm{Iso}(G_i)$.
- Conclude

$$\mathbb{P}(R \cong G_i) = \frac{6}{8} \cdot \frac{1}{|\mathrm{Aut}(G_i)|}$$

Distribution of Random Groups of Order n

- Generate a random group R of order n by writing down random $n \times n$ multiplication table (repeat if this is not a group structure)
- If G is a group of order n, what is $\mathbb{P}(R \cong G)$?
- Exact same arguments as before yield

$$\mathbb{P}(R \cong G) = \frac{|\{\text{tables isomorphic to } G\}|}{|\{\text{tables that give group structure}\}|} \sim \frac{1}{|\mathrm{Aut}(G)|}$$

Distribution of Random Groups from Random Matrices

- Observe: If $M \in \mathrm{Mat}_{n\times n}(\mathbb{Z}_p)$ has full rank, then $\operatorname{cok} M = \mathbb{Z}_p^n / \operatorname{im} M$ is a finite abelian p-group
- $\operatorname{cok} M$: group with n generators and n relations given by cols of M
- Can show: There is a probability measure on $\mathrm{Mat}_{n\times n}(\mathbb{Z}_p)$ induced by the *Haar measure*

Theorem (Friedman, Washington, 1989)

p prime, G finite abelian p-group. For each n pick a random matrix $R_n \in \mathrm{Mat}_{n\times n}(\mathbb{Z}_p)$ wrt. Haar measure. Then

$$\lim_{n\to\infty} \mathbb{P}(\operatorname{cok} R_n \cong G) = c_p \cdot \frac{1}{|\operatorname{Aut} G|}.$$

The Cohen–Lenstra Principle

The Cohen–Lenstra Principle

Cohen–Lenstra Principle

The probability that a random algebraic object is isomorphic to a given object G of the same kind is proportional to $\frac{1}{|\mathrm{Aut}\, G|}$.

Cohen, Lenstra: For $d < 0$, p-Sylow subgroup $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}[p^\infty]$ behaves like random finite abelian p-group!

Conjecture (Cohen, Lenstra, 1983)

p odd prime, G finite abelian p-group. Then for random $d < 0$,

$$\mathbb{P}(\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}[p^\infty] \cong G) = c_p \cdot \frac{1}{|\mathrm{Aut}\, G|}$$

Big open problem in number theory!

The Cohen–Lenstra Principle

Conjecture (Cohen, Lenstra, 1983)

p odd prime, G finite abelian p-group. Then for random $d < 0$,

$$\mathbb{P}(\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}[p^\infty] \cong G) = c_p \cdot \frac{1}{|\mathrm{Aut}\, G|}$$

Example

Data: $\mathbb{P}(\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}[3^\infty]$ is cyclic$) \approx 0.98$. On the other hand,

$$\sum_{G \text{ cyclic 3-gp.}/\cong} c_3 \cdot \frac{1}{|\mathrm{Aut}\, G|} = 0.9802\ldots$$

The Cohen–Lenstra Principle

Cohen–Lenstra Principle

The probability that a random algebraic object is isomorphic to a given object G of the same kind is proportional to $\frac{1}{|\mathrm{Aut}\, G|}$.

General philosophy in arithmetic statistics:
Suppose we are interested in objects X_i, $i \in I$.
Idea: statistical behaviour reflects properties of the X_i.

- If the X_i do *not* behave like random objects of some kind, they must have some extra structure
- If the X_i behave like random objects of some kind, have "understood" their structure

This new perspective has been introduced by Cohen and Lenstra.
Has inspired a lot of work in arithmetic statistics and beyond!

The Cohen–Lenstra Principle

General philosophy in arithmetic statistics: When interested in X_i, $i \in I$:

- If the X_i do *not* behave like random objects of some kind, they must have some extra structure
- If the X_i behave like random objects of some kind, have "understood" their structure

Example

- Have seen: for $d < 0$, $\left|\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}\right|$ does not behave like a random number. Reason: it is the size of a group.
- Cohen and Lenstra's conjecture asserts that for $d < 0$ and p odd, only structure on $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}[p^\infty]$ is abelian group structure.

Other Number Fields

Conjecture (Cohen, Lenstra, 1983)

p odd prime, G finite abelian p-group. Then for random $d < 0$,

$$\mathbb{P}(\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}[p^\infty] \cong G) = c_p \cdot \frac{1}{|\mathrm{Aut}\, G|}$$

What about other number fields?

- Similar conjectures by Cohen–Martinet and Bartel–Lenstra
- In general: there is some extra structure
- Can "add" extra structure to Cl_K and this bigger structure satisfies C–L principle!

Ray Class Groups

Ray Class Groups

K number field. For $m \in \mathbb{Z}_{>0}$ have ray class group $\mathrm{Cl}_K(m)$

- $\mathrm{Cl}_K(m)$ finite abelian group with $\mathrm{Cl}_K(1) = \mathrm{Cl}_K$, so generalises Cl_K
- Have exact sequence

$$0 \longrightarrow \ker \varphi \longrightarrow \mathrm{Cl}_K(m) \xrightarrow{\varphi} \mathrm{Cl}_K \longrightarrow 0$$

Fix m and p. Let G finite abelian p-group. For random $d < 0$, what is

$$\mathbb{P}(\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}(m)[p^\infty] \cong G)?$$

Varma (2016): If $m \neq 1$, then $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}(m)[p^\infty]$ behaves differently than $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}[p^\infty]$, not like random group!

Distribution of Ray Class Groups

Varma (2016): If $m \neq 1$, then $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}(m)[p^\infty]$ behaves differently than $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}[p^\infty]$, not like random group!

Why? Extra structure?

Yes! Given by ray class sequence

$$0 \longrightarrow (\ker \varphi)[p^\infty] \longrightarrow \mathrm{Cl}_K(m)[p^\infty] \xrightarrow{\varphi} \mathrm{Cl}_K[p^\infty] \longrightarrow 0$$

Idea: Instead of $\mathrm{Cl}_K(m)[p^\infty]$ consider the whole sequence!

- Pagano–Sofos (2017) conjecture: Ray class sequence is random, i.e. has distribution $\sim 1/|\mathrm{Aut}|$
- This conjecture implies Varma's results!

Distribution of Ray Class Groups

What about other number fields?

Here, ray class sequence

$$0 \longrightarrow (\ker \varphi)[p^\infty] \longrightarrow \mathrm{Cl}_K(m)[p^\infty] \xrightarrow{\varphi} \mathrm{Cl}_K[p^\infty] \longrightarrow 0$$

is not random, has extra structure!

- Solution: "add" extra structure to get slightly bigger sequence
- Problem: Bigger sequence has infinitely many automorphisms
- Solution: Can develop theory to make sense of $1/|\mathrm{Aut}\, X|$ even if $|\mathrm{Aut}\, X| = \infty$
- Can then formulate conjecture for distribution of bigger ray class sequences. It implies conjectures on Cl_K!

A Nice Corollary on Q2

Recall: class groups related to solubility of equations over $\mathbb{Z}$.

Eisenstein (1844): Q2: Among the $d \in \mathbb{Z}_{>0}$ squarefree with $d \equiv 5 \bmod 8$, for how many can you solve the equation

$$x^2 - dy^2 = 4$$

in *odd* integers x, y?

Corollary

Assume the conjecture on the distribution of ray class sequences holds. Then the proportion of such d is $\frac{1}{3}$.

Summary

- Arithmetic statistics: understand number-theoretic objects via statistical behaviour
- Key part: Cohen–Lenstra principle: random $= 1/|\mathrm{Aut}|$

- Ideal class group Cl_K is central object in number theory
  - measures failure of unique factorisation in $\mathbb{Z}_K$
  - related to finding integer solutions to equations

- Cohen–Lenstra: conjectures about distribution of Cl_K

- Extended C–L conjecture to ray class groups $\mathrm{Cl}_K(m)$
  - Also predicts how often $x^2 - dy^2 = 4$ soluble

Why Sylow Sugroups?

Why consider $\mathrm{Cl}_K[p^\infty]$ and not Cl_K?

Theorem (Cohen, Lenstra, 1983)

$$\sum_{M \text{ fin. ab. } p\text{-gp.}/\cong} \frac{1}{|\mathrm{Aut}\, M|} =: c_p < \infty$$

In contrast:

$$\sum_{M \text{ fin. ab. gp.}/\cong} \frac{1}{|\mathrm{Aut}\, M|} \geq \sum_{p \text{ prime}} \frac{1}{|\mathrm{Aut}(\mathbb{Z}/p)|} = \sum_{p \text{ prime}} \frac{1}{p-1} \geq \sum_{p \text{ prime}} \frac{1}{p} = \infty$$

What about $p = 2$?

Can show: $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}[2^\infty]$ for $d < 0$ does not behave like a random finite abelian 2-group.

Why? Extra structure?

Yes: The structure of 2-torsion $\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})}[2]$ is known.

Theorem (Smith, 2022)

G finite abelian 2-group. Then for random $d < 0$,

$$\mathbb{P}((2\,\mathrm{Cl}_{\mathbb{Q}(\sqrt{d})})[2^\infty] \cong G) = c_2 \cdot \frac{1}{|\mathrm{Aut}\, G|}$$