How many points on a random curve over a finite field?

Kiran S. Kedlaya

Department of Mathematics, University of California, San Diego
kedlaya@ucsd.edu
http://kskedlaya.org/slides/

Number fields and function fields:
coalescences, contrasts and emerging applications
The Royal Society at Chicheley Hall, May 29, 2014

Based on work at the workshop "Arithmetic statistics over number fields and function fields" (American Institute of Mathematics, January 2014).
Collaborators: J. Achter, D. Erman, C. Hall, M. Wood, D. Zureick-Brown.

Additional support from NSF (grant DMS-1101343), UCSD (Warschawski chair).

The zeta function on a curve over a finite field

Let C be a (smooth, projective, geometrically irreducible) curve of genus g over a finite field $\mathbb{F}_q$. The zeta function of its function field has the form

$$\zeta_C(s) = \frac{P(q^{-s})}{(1-q^{-s})(1-q^{1-s})}$$

where $P(T)$ is an integer polynomial of degree $2g$ factoring over $\mathbb{C}$ as $(1 - q^{1/2}\alpha_1 T)\cdots(1 - q^{1/2}\alpha_{2g} T)$ with $|\alpha_i| = 1$ and $\alpha_{g+i} = \overline{\alpha_i}$. Also,

$$\#C(\mathbb{F}_{q^n}) = q^n + 1 - q^{n/2}(\alpha_1^n + \cdots + \alpha_{2g}^n) \qquad (n = 1, 2, \ldots).$$

We are interested in the statistical distribution of $\#C(\mathbb{F}_{q^n})$ as C varies in various large families. Before focusing on the particular question at hand in this talk, let us recall some other types of questions of this form.

The large q limit

Suppose first that C is sampled randomly from a geometric family (e.g., hyperelliptic curves of a given genus), but we consider statistics in the limit as $q \to \infty$. Using étale cohomology (work of Deligne, Katz, Sarnak), one can read off statistical properties of $\#C(\mathbb{F}_{q^n})$ from the geometry of the corresponding moduli space.

The polynomial $(T - \alpha_1) \cdots (T - \alpha_{2g})$ behaves like the characteristic polynomial of a random matrix in a certain compact Lie group (related to the *geometric monodromy group* of the family). For instance, in the family of hyperelliptic curves of genus g (or any larger family) this group is the unitary symplectic group $\mathrm{USp}(2g)$.

Arithmetic families

Suppose next that C is obtained from a fixed curve over a number field by reduction modulo a varying prime ideal. One can make predictions about $\#C(\mathbb{F}_{q^n})$ using analysis formally similar to the large q limit, again in terms of a certain compact Lie group depending on C (related to the *Mumford-Tate group*). However, these predictions are only unconditional in a few cases where one has analytic continuation of certain L-functions (e.g., elliptic curves over $\mathbb{Q}$ by Taylor et al.).

For a fixed genus g, only finitely many distinct distributions can occur. These are classified only for $g = 1$ (3 cases: CM over the base field, CM over an extension field, no CM) and $g = 2$ (52 cases: see Fité-K-Rotger-Sutherland).

Hyperelliptic curves over a fixed finite field

From now on, we fix q. To get questions about infinite sets, we must consider families in which g varies. The natural probability distributions are now *discrete*, so it appears (at first) that random matrix theory does not have anything to say.

A typical example is hyperelliptic curves of arbitrary genus with q odd (Kurlberg-Rudnick). In this case, for each n, $\#C(\mathbb{F}_{q^n})$ behaves as a sum of independent random variables corresponding to the points of $\mathbb{P}^1(\mathbb{F}_{q^n})$. For instance, $\#C(\mathbb{F}_q)$ behaves like the sum of $q+1$ iid random variables taking the values $0, 1, 2$ with respective probabilities

$$\frac{q}{2(q+1)}, \frac{1}{q+1}, \frac{q}{2(q+1)}.$$

These are the probabilities that a particular value of a squarefree polynomial is a nonsquare, zero, or a square.

More results over a fixed finite field

Similar results have been obtained for a large number of families: cyclic p-gonal curves (Bucur-David-Feigon-Lalín), plane curves (BDFL), trigonal curves (Wood, Xiong), complete intersections in projective space (Bucur-K), Artin-Schreier curves (BDFL), etc.

However, all of these examples share some features which are not entirely typical.

- The relevant moduli spaces are rational, so one can describe the family in terms of parameters.
- The number of $\mathbb{F}_{q^n}$-rational points is bounded uniformly in g.

If we consider the full moduli space of curves, both of these features disappear; even making a plausible heuristic model for the distribution of point counts becomes unclear.

Arbitrary curves over a fixed finite field

The purpose of this talk is to propose and justify the following conjecture.

Conjecture (precise version below)

For q fixed, the distribution of $\#C(\mathbb{F}_q)$ as C varies over isomorphism classes of curves is Poisson with mean $\lambda = q + 1 + q^{-1} + q^{-2} + \cdots$.

More precisely, equip the set of isomorphism classes of genus g curves over $\mathbb{F}_q$ with the measure where the class of C has weight proportional to $1/\#\operatorname{Aut}(C)$. Then for each nonnegative integer n, we conjecture that

$$\lim_{g\to\infty} \operatorname{Prob}(\#C(\mathbb{F}_q) = n : g(C) = g) = \frac{\lambda^n e^{-\lambda}}{n!}$$

and that

$$\lim_{g\to\infty} \mathbb{E}((\#C(\mathbb{F}_q))^n : g(C) = g) = \sum_{i=1}^{n} \left\{ {n \atop i} \right\} \lambda^i,$$

where $\left\{ {n \atop i} \right\}$ counts unordered partitions of $\{1, \dots, n\}$ into i disjoint sets.

Justification: cohomology of moduli spaces

For $g, n \geq 0$, let $M_{g,n}$ be the moduli space of genus g curves with n distinct marked points. Our conjecture can now be interpreted as

$$\lim_{g \to \infty} \frac{\# M_{g,n}(\mathbb{F}_q)}{\# M_g(\mathbb{F}_q)} = \lambda^n$$

provided that points are again weighted by an automorphism factor.

This is suggested by the Lefschetz trace formula: one has

$$\# M_{g,n}(\mathbb{F}_q) \approx \sum_{i \text{ small}} q^{h-i} \dim H^{2d-2i}(M_{g,n}) \quad (d = \dim(M_{g,n}) = 3g - 3 + n)$$

because the low degree cohomology of $M_{g,n}$ consists only of cycle classes (namely Mumford's *tautological classes*). The difference between the cohomology rings of $M_{g,n+1}$ and $M_{g,n}$ is one polynomial generator (the Chern class of the tautological line bundle defined by the $(n+1)$-st point).

Back to the large q limit

Using results on the stable cohomology of moduli spaces, one can prove a result in the large q limit.

Theorem

For all nonnegative integers m, n, as $g, q \to \infty$ we have

$$\mathbb{E}((\#C(\mathbb{F}_q))^n : g(C) = g) = \sum_{i=1}^{n} \left\{ {n \atop i} \right\} \lambda^i + O(q^{-m})$$

provided *that q grows sufficiently fast compared to g.*

The intermediate Betti numbers of $M_{g,n}$ grow far too fast to get a meaningful estimate for q fixed. However, the conjecture is still plausible if you believe that the nontautological classes of $M_{g,n}$ behave like a random unitary symplectic matrix, whose trace is then *bounded*.

Random matrices with discrete invariants?

Question

Is there a random matrix model that would reproduce our conjecture?

In the large q limit, $\#C(\mathbb{F}_{q^n})$ relates to random matrices in $\mathrm{USp}(2g)$. In the fixed q case, the numbers

$$\#C(\mathbb{F}_{q^n}) = q^n + 1 - q^{n/2}(\alpha_1^n + \cdots + \alpha_{2g}^n) \qquad (n = 1, 2, \dots).$$

have some extra properties that must be built into the model:

- the number $\#C(\mathbb{F}_{q^n})$ is a nonnegative integer;
- whenever ℓ is prime, $\#C(\mathbb{F}_{q^{n\ell}}) - \#C(\mathbb{F}_{q^n})$ is a nonnegative multiple of ℓ.

Random matrices with discrete invariants?

Consider the joint distribution for the traces of the first n powers of a random matrix in the circular symplectic ensemble. This distribution is continuous, so it is defined by a continuous function on $\mathbb{R}^n$; one can then restrict this function to the discrete set of points corresponding to allowable values of $\#C(\mathbb{F}_q), \ldots, \#C(\mathbb{F}_{q^n})$ to get a distribution on these n-tuples.

Question

Does this predict a distribution of $\#C(\mathbb{F}_q)$ which is again Poisson with mean λ?

What about extension fields?

Question

What is the distribution of $\#C(\mathbb{F}_{q^n})$ as C runs over curves over $\mathbb{F}_q$?

My best guess: a sum of Poisson distributions, one for each divisor of n.