Math 259: Introduction to Analytic Number Theory

How many points can a curve of genus g have over $\mathbf{F}_q$?

Let k be a finite field of q elements, and C/k a (smooth projective) curve of genus $g = g(C)$. Let $K = k(C)$ be its function field. A "prime" (a.k.a. "place", "valuation") p of K is a Galois orbit of $\bar{k}$-rational points of C. If that orbit has size $d = d_p$ (the "degree" of p) then we are dealing with d_p conjugate points defined over the q^{d_p}-element field (and no smaller field intermediate between it and k), which is the residue field of p. The *zeta function* $\zeta_K(s) = \zeta_C(s)$ of this field or curve may be defined as the Euler product

$$\zeta_C(s) = \prod_p \frac{1}{1-(q^{d_p})^{-s}} = \prod_p \frac{1}{1-Z^d}$$

extending over all primes p, where $Z = q^{-s}$. Then

$$\log \zeta_C(s) = \sum_p \sum_{m=1}^{\infty} Z^{d_p m}/m = \sum_{n=1}^{\infty} Z^n \sum_{d_p|n} \frac{d_p}{n} = \sum_{n=1}^{\infty} \Big(\sum_{d_p|n} d_p \Big) \frac{Z^n}{n}.$$

But the inner sum is just the number $N_n = N_n(C)$ of points of C rational over the field of q^n elements. Note[1] that $N_n \ll_C q^n$, whence the sum and thus the Euler product converge for $|Z| < 1/q$, i.e., for $\sigma > 1$.

As in the number-field case, ζ_C satisfies a functional equation relating its values at s and $1-s$:

$$\zeta_C(1-s) = q^{(2-2g)(\frac{1}{2}-s)} \zeta_C(s) = (qZ^2)^{1-g} \zeta_C(s);$$

equivalently,

$$\xi_C(s) := q^{(1-g)(s-\frac{1}{2})} \zeta_C(s)$$

is invariant under $s \leftrightarrow 1-s$. Moreover, $\zeta_C(s)$ is of the form

$$\zeta_C(s) = P(Z)/(1-Z)(1-qZ)$$

for some polynomial P of degree $2g$ with $P(0) = 1$. It then follows from the functional equation that $P(1/qZ) = P(Z)/(qZ^2)^g$, which is to say that we can factor $P(Z)$ as

$$P(Z) = \prod_{j=1}^{g} (1-\lambda_j Z)(1-\lambda_{g+j} Z)$$

for some complex numbers $\lambda_1, \ldots, \lambda_{2g}$ such that

$$\lambda_j \lambda_{g+j} = q$$

[1]For instance, let $f : C \to \mathbf{P}^1$ be any nonconstant function; then

$$N_n(C) \le (\deg f) N_n(\mathbf{P}^1) = (\deg f)(q^n + 1) \ll q^n.$$

for $j = 1, \ldots, g$. Comparing this with our formula for N_n we find

$$N_n = q^n + 1 - \sum_{j=1}^{2g} \lambda_j^n.$$

(Fortunately this agrees with our known formula for N_n when $g = 0$.) The analogue of the Dirichlet class number formula is the fact that "Jacobian" $J_C(k)$ of C over k has size

$$P(1) = \prod_j (1 - \lambda_j),$$

which is essentially the residue of $\zeta_C(s)$ at its pole $s = 1$.

So far all this can be proved by more-or-less elementary means, and even extends to varieties over k of any dimension [Dwork 1960]. A much harder, but known, result is that the Riemann hypothesis holds: $P(q^{-s})$ can vanish only for s such that $\sigma = 1/2$, i.e., $|Z| = q^{-1/2}$; thus all the λ_j have absolute value $q^{1/2}$, and $\lambda_{g+j} = \bar{\lambda}_j$. This theorem of Weil, and its generalization by Deligne to varieties of arbitrary dimension over finite fields, is at least to some tastes the strongest evidence so far for the truth of the original Riemann hypothesis and its various generalizations.

The theorem $|\lambda_j|^2 = q$ also has numerous applications. For instance, it follows immediately that the number $N_1 = N_1(C)$ of k-rational points on C is approximated by $q + 1$:

$$|N_1 - (q+1)| \leq 2g\sqrt{q}. \tag{1}$$

Equality can hold in this *Weil bound* at least for small g, though already for $g = 1$ there are surprises; for instance for $q = 128$ the bound (1) allows N_1 to be as large as 151 and as small as 107, but in fact the maximum and minimum are 150 and 108. See [Serre 1982–4] for much more about this. We ask however what happens for fixed q as $g \to \infty$: how large can $N_1(C)$ grow as a function of g? this is not only a compelling problem in its own right, but has applications to coding theory and similar combinatorial problems, see for instance [Goppa 1981,3; Tsfasman 1996; Elkies 2001]. We shall see that the bound $N_1 < 2g\sqrt{q} + O_q(1)$ coming from (1) cannot be sharp, and obtain an improved bound, the *Drinfeld-Vlăduţ bound*

$$N_1 < (\sqrt{q} - 1 + o(1))\, g, \tag{2}$$

[DV 1983], that turns out to be best possible for square q [Ihara 1981, TVZ 1982]. Moreover, we shall adapt Weyl's equidistribution argument to obtain the asymptotic distribution of the λ_j on the circle $|\lambda|^2 = q$ for curves attaining that bound.

The key idea is much the same as what we used to prove that $\zeta(1 + it) \neq 0$. To start with, note that if the Weil upper bound $N_1 \leq q + 1 + 2g\sqrt{q}$ is attained then each $\lambda_j = -\sqrt{q}$. This can actually happen: for instance, let $q = q'^2$ and let C be the $(q'+1)$-st *Fermat curve*, i.e., the smooth plane curve $x^{q'+1} + y^{q'+1} + z^{q'+1} = 0$ of degree $q' + 1$ and therefore of genus $(q'^2 - q')/2$. Then C has $q'^3 + 1$ points over k, the maximum allowed by (1) [check this!]. But now consider this curve over the quadratic extension $\mathbf{F}_{q^2}$ of k: we have

$$N_2 = q^2 + 1 - \sum_{j=1}^{2g} \lambda_j^2 = q^2 + 1 - 2gq = q^{3/2} + 1 = N_1,$$

i.e., every point rational over $\mathbf{F}_{q^2}$ is already $\mathbf{F}_q$-rational! [It is an amusing problem to verify this directly, without invoking the Riemann hypothesis for ζ_C.] It follows that if g were any larger than $(q-q')/2$ and all the λ_j were equal to $-q'$ then N_2 would actually be smaller than N_1, which is impossible.

So, we have

$$0 \le N_2 - N_1 = q^2 - q + \sum_{j=1}^{2g}(\lambda_j - \lambda_j^2),$$

and likewise

$$0 \le N_n - N_1 = q^n - q + \sum_{j=1}^{2g}(\lambda_j - \lambda_j^n)$$

for each $n = 2, 3, 4, \ldots$ (We also have inequalities $N_{dn} > N_n$, but these do not help us asymptotically.) How to best combine them? For given q, g this is not an easy problem, but if we fix q and only care about asymptotics as $g \to \infty$ then all we need do is use the inequality

$$0 \le \Big| \sum_{m=1}^{M} (\lambda_j/\sqrt{q})^m \Big|^2 = M + \sum_{m=1}^{M-1} (M-m) q^{-m/2} (\lambda_j^m + \lambda_{j+g}^m)$$

for each M. (This is the positivity of the Fejér kernel $|\sum_{m=1}^{M} z^m|^2$ for $|z| = 1$.) Summing this inequality over $j \le g$ we find

$$\begin{aligned} 0 &\le Mg + \sum_{m=1}^{M-1} (M-m) q^{-m/2} (q^m + 1 - N_m) \\ &\le Mg + \sum_{m=1}^{M-1} (M-m) q^{-m/2} (q^m + 1 - N_1) \\ &= Mg + O_M(1) - \Big(\sum_{m=1}^{M-1} (M-m) q^{-m/2} N_1 \Big). \end{aligned}$$

Thus

$$N_1 < \frac{g}{\sum_{m=1}^{M-1} (1 - \frac{m}{M}) q^{-m/2}} + O_M(1).$$

For each $\epsilon > 0$, the sum can be brought within ϵ of

$$\sum_{m=1}^{\infty} q^{-m/2} = 1/(\sqrt{q} - 1)$$

by taking M large enough. We thus have for each $\delta > 0$

$$N_1 < (\sqrt{q} - 1 + \delta)g + O_\delta(1),$$

from which (2) follows.

What is required for asymptotic equality as C ranges over a sequence of curves with $g \to \infty$? Let $\lambda_j = q^{1/2} e(x_j)$ for $x_j \in \mathbf{R}/\mathbf{Z}$ with $x_{j+g} = -x_j$. Then

$$N_n = -q^{n/2} \sum_{j=1}^{2g} e(nx_j) + q^n + 1.$$

Since $N_n \geq N_1$ is used for each n, we must have $N_n = N_1 + o_n(g)$, and thus

$$\sum_{j=1}^{2g} e(nx_j) = q^{(1-n)/2} \sum_{j=1}^{2g} e(x_j) + o_n(g).$$

Moreover

$$\sum_{j=1}^{2g} e(x_j) = -(1 - q^{-1/2})g + o(g).$$

Adapting the Weyl equidistribution argument (see especially Exercise 2 of the Weyl handout), we conclude that the x_j approach the distribution whose n-th Fourier moment ($n \neq 0$) is $-(1 - q^{-1/2})/2q^{(|n|-1)/2}$, that is, $\delta_q(x)\,dx$ where the density δ_q is

$$1 - (1 - q^{-1/2}) \sum_{n=1}^{\infty} q^{(1-n)/2} \, \frac{e(nx) + e(-nx)}{2}.$$

Since

$$(1 - q^{-1/2}) \sum_{n=1}^{\infty} q^{(1-n)/2} = 1,$$

this density is nonnegative, so it can be attained and (2) is asymptotically the best inequality that can be obtained from $N_n \geq N_1$. In fact it is known [Ihara 1981, TVZ 1982] that when q is a square[2] there are curves with arbitrarily large g for which $N_1 \geq (\sqrt{q} - 1)\, g$; our proof of (2) gives the asymptotic distribution of λ_j on the circle $|\lambda|^2 = q$ for any such sequence. It also lets us compute the size $\#J = \prod_{j=1}^{2g} (1 - \lambda_j)$ of the Jacobian in a logarithmic asymptotic sense:

$$g^{-1} \log \#J \to \log q + \int_0^1 \log |1 - q^{-1/2} e(x)| \, \delta_q(x) \, dx. \tag{3}$$

The integral can be evaluated explicitly using the Taylor expansion of $\log(1-z)$ (see the Exercises). Such formulas are needed to determine the asymptotic performance of families of codes or lattices constructed as in [Tsfasman 1996] from the curves of [Ihara 1981, TVZ 1982].

[2] When q is not a square, $\limsup_{g \to \infty} N_1(C)/g(C)$ is known to be positive (see for instance [Serre 1982–1984]), but its value is still a great mystery even for $q = 2$.

Remark

The only families of curves known to attain the Drinfeld-Vlăduţ bound consist of modular curves of various kinds. Explicit formulas for some such families can be found in [TV 1991, GS 1995] (Drinfeld modular curves), [Elkies 1998, 1998a] (elliptic and Shimura modular curves), and elsewhere.

Exercises

1. Verify that if q' is a prime power then the Fermat curve of degree $q'+1$ has ${q'}^3+1$ rational points over the field of ${q'}^2$ elements.

2. What is the best upper bound that can be obtained on N_1 using only the inequality $N_1 \leq N_2$? Prove that the inequalities $N_1 \leq N_n$ $(n=3,4,\ldots)$ further improve this bound if and only if $g > (q^2-q)/\sqrt{2q}$. [It is known that if $q = 2^{2e+1}$ for some integer $e \geq 0$ then there exists a curve of genus $(q^2-q)/\sqrt{2q} = 2^{3e+1}-2^e$ with $N_1 = N_2 = N_3 = q^2+1$. For instance, for $e=0$ this is the elliptic curve with affine equation $y^2+y = x^3+x$ over the 2-element field.]

3. Compute $\delta_q(x)$ and the integral (3) in closed form. Generalize to obtain, for each $s \in \mathbf{C}$ of real part $> 1/2$, a closed form for $\lim_{g\to\infty} g^{-1} \log\bigl((1-q^{1-s})\zeta_C(s)\bigr)$ as C ranges over a family of curves over $\mathbf{F}_{q^2}$ with $N_1(C)/g(C) \to \sqrt{q}-1$. (For the answer and an application to error-correcting codes, see [Elkies 2001], already cited in the Exercises for Euler products.)

References

[DV 1983] Drinfeld, V.G., Vlăduţ, S.: Number of points of an algebraic curve, *Func. Anal.* **17** (1983), 53–54.

[Dwork 1960] Dwork, B.M.: On the rationality of the zeta function of an algebraic variety. *Amer. J. Math.* **82** (1960), 631–648.

[Elkies 1998] Elkies, N.D.: Explicit modular towers. Pages 23–32 in *Proceedings of the Thirty-Fifth Annual Allerton Conference on Communication, Control and Computing*, Univ. of Illinois at Urbana-Champaign, 1998.

[Elkies 1998a] Elkies, N.D.: Explicit modular towers. Pages 1–47 in *Proceedings of ANTS-3* (Lecture Notes in Computer Science 1423), Berlin: Springer, 1998. `math.NT/0005160` at `arXiv.org`.

[Elkies 2001] Excellent nonlinear codes from modular curves, pages 200–208 in *STOC'01: Proceedings of the 33rd Annual ACM Symposium on Theory of Computing, Hersonissos, Crete, Greece.* Isomorphic with `math.NT/0104115` at `arXiv.org`.

[GS 1995] Garcia, A., Stichtenoth, H.: A tower of Artin-Schreier extensions of function fields attaining the Drinfeld-Vladut bound. *Invent. Math.* **121** (1995), 211–233.

[Goppa 1981,3] Goppa, V.D.: Codes on algebraic curves, *Soviet Math. Dokl.* **24** (1981), 170–172; Algebraico-geometric codes, *Math. USSR Izvestiya* **24** (1983), 75–91.

[Ihara 1981] Ihara, Y.: Some remarks on the number of rational points of algebraic curves over finite fields. *J. Fac. Sci. Tokyo* **28** (1981), 721–724.

[Serre 1982–4] Serre, J.-P.: Sur le nombre des points rationnels d'une courbe algébrique sur un corps fini; Nombres de points des courbes algébriques sur $\mathbf{F}_q$; Résumé des cours de 1983–1984: reprinted as ##128,129,132 in his *Collected Works III* [O 9.86.1 (III) / QA3.S47]

[Tsfasman 1996] Tsfasman, M.A.: Algebraic Geometry Lattices and Codes, pages 385–390 in the proceedings of ANTS-II (second Algorithmic Number Theory Symposium), ed. H. Cohen, *Lecture Notes in Computer Science* **1122** [QA75.L4 #1122 in the McKay Applied Science Library].

[TV 1991] Tsfasman, M.A., Vlăduţ: *Algebraic-Geometric Codes.* Dordrecht: Kluwer, 1991.

[TVZ 1982] Tsfasman, M.A., Vlăduţ, S.G., Zink, T.: Modular curves, Shimura curves and Goppa codes better than the Varshamov-Gilbert bound. *Math. Nachr.* **109** (1982), 21–28.