

## Introduction

The systematic study of character sums over finite fields may be said to have begun over 200 years ago, with Gauss. The Gauss sums over  $\mathbb{F}_p$  are the sums

$$\sum_{x \in \mathbb{F}_p^\times} \psi(x)\chi(x),$$

for  $\psi$  a nontrivial additive character of  $\mathbb{F}_p$ , e.g.,  $x \mapsto e^{2\pi ix/p}$ , and  $\chi$  a nontrivial multiplicative character of  $\mathbb{F}_p^\times$ . Each has absolute value  $\sqrt{p}$ . In 1926, Kloosterman [**Kloos**] introduced the sums (one for each  $a \in \mathbb{F}_p^\times$ )

$$\sum_{xy=a \text{ in } \mathbb{F}_p} \psi(x+y)$$

which bear his name, in applying the circle method to the problem of four squares. In 1931 Davenport [**Dav**] became interested in (variants of) the following questions: for how many  $x$  in the interval  $[1, p-2]$  are both  $x$  and  $x+1$  squares in  $\mathbb{F}_p$ ? Is the answer approximately  $p/4$  as  $p$  grows? For how many  $x$  in  $[1, p-3]$  are each of  $x, x+1, x+2$  squares in  $\mathbb{F}_p$ ? Is the answer approximately  $p/8$  as  $p$  grows? For a fixed integer  $r \geq 2$ , and a (large) prime  $p$ , for how many  $x$  in  $[1, p-r]$  are each of  $x, x+1, x+2, \dots, x+r-1$  squares in  $\mathbb{F}_p$ . Is the answer approximately  $p/2^r$  as  $p$  grows? These questions led him to the problem of giving good estimates for character sums over the prime field  $\mathbb{F}_p$  of the form

$$\sum_{x \in \mathbb{F}_p} \chi_2(f(x)),$$

where  $\chi_2$  is the quadratic character  $\chi_2(x) := (\frac{x}{p})$ , and where  $f(x) \in \mathbb{F}_p[x]$  is a polynomial with all distinct roots. Such a sum is the “error term” in the approximation of the number of mod  $p$  solutions of the equation

$$y^2 = f(x)$$

by  $p$ , indeed the number of mod  $p$  solutions is exactly equal to

$$p + \sum_{x \in \mathbb{F}_p} \chi_2(f(x)).$$

And, if one replaces the quadratic character by a character  $\chi$  of  $\mathbb{F}_p^\times$  of higher order, say order  $n$ , then one is asking about the number of mod  $p$  solutions of the equation

$$y^n = f(x).$$

This number is exactly equal to

$$p + \sum_{\chi|\chi^n=\mathbf{1},\chi\neq\mathbf{1}} \sum_{x\in\mathbb{F}_p} \chi(f(x)).$$

The “right” bounds for Kloosterman’s sums are

$$\left| \sum_{xy=a \text{ in } \mathbb{F}_p} \psi(x+y) \right| \leq 2\sqrt{p}$$

for  $a \in \mathbb{F}_p^\times$ . For  $f(x) = \sum_{i=0}^d a_i x^i$  squarefree of degree  $d$ , the “right” bounds are

$$\left| \sum_{x\in\mathbb{F}_p} \chi(f(x)) \right| \leq (d-1)\sqrt{p}$$

for  $\chi$  nontrivial and  $\chi^d \neq \mathbf{1}$ , and

$$\left| \chi(a_d) + \sum_{x\in\mathbb{F}_p} \chi(f(x)) \right| \leq (d-2)\sqrt{p}$$

for  $\chi$  nontrivial and  $\chi^d = \mathbf{1}$ . These bounds were foreseen by Hasse [**Ha-Rel**] to follow from the Riemann Hypothesis for curves over finite fields, and were thus established by Weil [**Weil**] in 1948.

Following Weil’s work, it is natural to “normalize” such a sum by dividing it by  $\sqrt{p}$ , and then ask how it varies in an algebro-geometric family. For example, one might ask how the normalized<sup>1</sup> Kloosterman sums

$$-(1/\sqrt{p}) \sum_{xy=a \text{ in } \mathbb{F}_p} \psi(x+y)$$

vary with  $a \in \mathbb{F}_p^\times$ , or how the sums

$$-(1/\sqrt{p}) \sum_{x\in\mathbb{F}_p} \chi_2(f(x))$$

vary as  $f$  runs over all squarefree cubic polynomials in  $\mathbb{F}_p[x]$ . [In this second case, we are looking at the  $\mathbb{F}_p$ -point count for the elliptic curve  $y^2 = f(x)$ .] Both these sorts of normalized sums are real, and lie in the closed interval  $[-2, 2]$ , so each can be written as twice the cosine of a

<sup>1</sup>The reason for introducing the minus sign will become clear later.

unique angle in  $[0, \pi]$ . Thus we define angles  $\theta_{a,p}$ ,  $a \in \mathbb{F}_p^\times$ , and angles  $\theta_{f,p}$ ,  $f$  a squarefree cubic in  $\mathbb{F}_p[x]$ :

$$-(1/\sqrt{p}) \sum_{xy=ain\mathbb{F}_p} \psi(x+y) = 2 \cos \theta_{a,p},$$

$$-(1/\sqrt{p}) \sum_{x \in \mathbb{F}_p} \chi_2(f(x)) = 2 \cos \theta_{f,p}.$$

In both these cases, the Sato-Tate conjecture asserted that, as  $p$  grows, the sets of angles  $\{\theta_{a,p}\}_{a \in \mathbb{F}_p^\times}$  (respectively  $\{\theta_{f,p}\}_{f \in \mathbb{F}_p[x] \text{ squarefree cubic}}$ ) become equidistributed in  $[0, \pi]$  for the measure  $(2/\pi) \sin^2(\theta) d\theta$ . Equivalently, the normalized sums themselves become equidistributed in  $[-2, 2]$  for the “semicircle measure”  $(1/2\pi) \sqrt{4-x^2} dx$ . These Sato-Tate conjectures were shown by Deligne to fall under the umbrella of his general equidistribution theorem, cf. [De-Weil II, 3.5.3 and 3.5.7] and [Ka-GKM, 3.6 and 13.6]. Thus for example one has, for a fixed nontrivial  $\chi$ , and a fixed integer  $d \geq 3$  such that  $\chi^d \neq \mathbf{1}$ , a good understanding of the equidistribution properties of the sums

$$-(1/\sqrt{p}) \sum_{x \in \mathbb{F}_p} \chi(f(x))$$

as  $f$  ranges over various algebro-geometric families of polynomials of degree  $d$ , cf. [Ka-ACT, 5.13].

In this work, we will be interested in questions of the following type: fix a polynomial  $f(x) \in \mathbb{F}_p[x]$ , say squarefree of degree  $d \geq 2$ . For each multiplicative character  $\chi$  with  $\chi^d \neq \mathbf{1}$ , we have the normalized sum

$$-(1/\sqrt{p}) \sum_{x \in \mathbb{F}_p} \chi(f(x)).$$

How are these normalized sums distributed as we **keep  $f$  fixed but vary  $\chi$**  over all multiplicative characters  $\chi$  with  $\chi^d \neq \mathbf{1}$ ? More generally, suppose we are given some suitably algebro-geometric function  $g(x)$ , what can we say about suitable normalizations of the sums

$$\sum_{x \in \mathbb{F}_p} \chi(x)g(x)$$

as  $\chi$  varies? This case includes the sums  $\sum_{x \in \mathbb{F}_p} \chi(f(x))$ , by taking for  $g$  the function  $x \mapsto -1 + \#\{t \in \mathbb{F}_p | f(t) = x\}$ , cf. Remark 17.7.

The earliest example we know in which this sort of question of variable  $\chi$  is addressed is the case in which  $g(x)$  is taken to be  $\psi(x)$ , so

that we are asking about the distribution on the unit circle  $S^1$  of the  $p - 2$  normalized Gauss sums

$$-(1/\sqrt{p}) \sum_{x \in \mathbb{F}_p^\times} \psi(x)\chi(x),$$

as  $\chi$  ranges over the nontrivial multiplicative characters. The answer is that as  $p$  grows, these  $p - 2$  normalized sums become more and more equidistributed for Haar measure of total mass one in  $S^1$ . This results [Ka-SE, 1.3.3.1] from Deligne’s estimate [De-ST, 7.1.3, 7.4] for multi-variable Kloosterman sums. There were later results [Ka-GKM, 9.3, 9.5] about equidistribution of  $r$ -tuples of normalized Gauss sums in  $(S^1)^r$  for any  $r \geq 1$ . The theory we will develop here “explains” these last results in a quite satisfactory way, cf. Corollary 20.2.

Most of our attention is focused on equidistribution results over larger and larger finite extensions of a given finite field. Emanuel Kowalski drew our attention to the interest of having equidistribution results over, say, prime fields  $\mathbb{F}_p$ , that become better and better as  $p$  grows. This question is addressed in Chapter 28, where the problem is to make effective the estimates, already given in the equicharacteristic setting of larger and larger extensions of a given finite field. In Chapter 29, we point out some open questions about “the situation over  $\mathbb{Z}$ ” and give some illustrative examples.

We end this introduction by pointing out two potential ambiguities of notation.

(1) We will deal both with lisse sheaves, usually denoted by calligraphic letters, most commonly  $\mathcal{F}$ , on open sets of  $\mathbb{G}_m$ , and with perverse sheaves, typically denoted by roman letters, most commonly  $N$  and  $M$ , on  $\mathbb{G}_m$ . We will develop a theory of the Tannakian groups  $G_{geom,N}$  and  $G_{arith,N}$  attached to (suitable) perverse sheaves  $N$ . We will also on occasion, especially in Chapters 11 and 12, make use of the “usual” geometric and arithmetic monodromy groups  $G_{geom,\mathcal{F}}$  and  $G_{arith,\mathcal{F}}$  attached to lisse sheaves  $\mathcal{F}$ . The difference in typography, which in turns indicates whether one is dealing with a perverse sheaf or a lisse sheaf, should always make clear which sort of  $G_{geom}$  or  $G_{arith}$  group, the Tannakian one or the “usual” one, is intended.

(2) When we have a lisse sheaf  $\mathcal{F}$  on an open set of  $\mathbb{G}_m$ , we often need to discuss the representation of the inertia group  $I(0)$  at 0 (respectively the representation of the inertia group  $I(\infty)$  at  $\infty$ ) to which  $\mathcal{F}$  gives rise. We will denote these representations  $\mathcal{F}(0)$  and  $\mathcal{F}(\infty)$  respectively. We will also wish to consider Tate twists  $\mathcal{F}(n)$  or  $\mathcal{F}(n/2)$  of  $\mathcal{F}$  by **nonzero** integers  $n$  or half-integers  $n/2$ . We adopt the convention that  $\mathcal{F}(0)$  (or  $\mathcal{F}(\infty)$ ) always means the representation of the

corresponding inertia group, while  $\mathcal{F}(n)$  or  $\mathcal{F}(n/2)$  with  $n$  a nonzero integer always means a Tate twist.