COMPOSITIO MATHEMATICA

Kloosterman paths and the shape of exponential sums

Emmanuel Kowalski and William F. Sawin

Compositio Math. **152** (2016), 1489–1516.

doi:10.1112/S0010437X16007351

Compositio Math. 152 (2016) 1489–1516
doi:10.1112/S0010437X16007351

Kloosterman paths and the shape of exponential sums

Emmanuel Kowalski and William F. Sawin

Dedicated to the memory of Marc Yor
'L'avenir est au hasard' (Jacques Brel)

Abstract

We consider the distribution of the polygonal paths joining partial sums of classical Kloosterman sums $\mathrm{Kl}_p(a)$, as a varies over $\mathbf{F}_p^\times$ and as p tends to infinity. Using independence of Kloosterman sheaves, we prove convergence in the sense of finite distributions to a specific random Fourier series. We also consider Birch sums, for which we can establish convergence in law in the space of continuous functions. We then derive some applications.

1. Introduction

For a prime number p and $a \in \mathbf{F}_p$, we denote by

$$\mathrm{Kl}_p(a) = \frac{1}{\sqrt{p}} \sum_{x=1}^{p-1} \psi_p(ax + \bar{x})$$

the normalized classical Kloosterman sum, where $\psi_p(z) = e(z/p) = e^{2i\pi z/p}$ is the standard additive character modulo p and $\bar{x}$ denotes the inverse of x modulo p.

Motivated partly by curiosity, arising to a large extent from staring at the corresponding plots (see [Kow15]), we consider in this paper the geometric properties of the *Kloosterman paths* in the complex plane. These are defined as follows: for each prime p and integer $a \in \mathbf{F}_p^\times$, we first let $\gamma_p(a)$ denote the polygonal path obtained by concatenating the closed segments $[z_j, z_{j+1}]$ joining the successive partial sums

$$z_j = \frac{1}{\sqrt{p}} \sum_{1 \leqslant x \leqslant j} \psi_p(ax + \bar{x}), \quad z_{j+1} = \frac{1}{\sqrt{p}} \sum_{1 \leqslant x \leqslant j+1} \psi_p(ax + \bar{x})$$

for $0 \leqslant j \leqslant p-2$. We then define a continuous map

$$t \mapsto K_p(t, a)$$

Received 1 April 2015, accepted in final form 13 November 2015, published online 15 April 2016.
2010 Mathematics Subject Classification 11T23, 11L05, 14F20, 60F17, 60G17 (primary), 60G50 (secondary).
Keywords: Kloosterman sums, Kloosterman sheaves, Riemann Hypothesis over finite fields, random Fourier series, short exponential sums, probability in Banach spaces.

E.K. was supported partly by a DFG-SNF lead agency program grant (grant 200021L_153647) this material is based upon work supported by the National Science Foundation Graduate Research Fellowship under grant no. DGE-1148900.

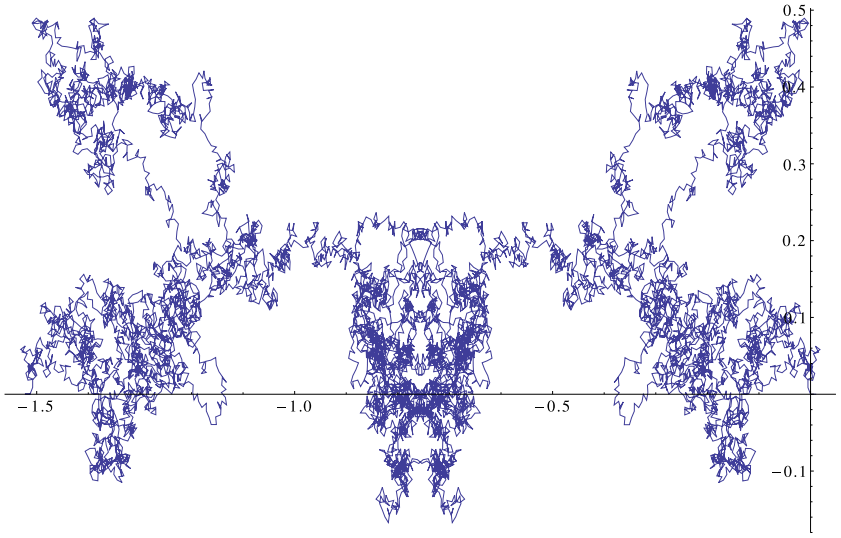

Figure 1. Plot of $t \mapsto K_{10\,007}(t, 1)$.

for $t \in [0, 1]$ by parameterizing the path $\gamma_p(a)$, each segment $[z_j, z_{j+1}]$ being parameterized linearly by $t \in [j/(p-1), (j+1)/(p-1)]$. Figure 1 shows the plot of $t \mapsto K_{10\,007}(t, 1)$, which should explain clearly the meaning of the definition.

We view each $a \mapsto K_p(t, a)$ as a random variable on the finite probability space

$$(\mathbf{F}_p^\times, \text{ uniform probability measure}),$$

and frequently write simply $K_p(t)$ for this random variable. Thus, $(K_p(t))_{t\in[0,1]}$ is a (simple) stochastic process.

We will use the computation of monodromy groups of Kloosterman sheaves to find the limiting distribution of $(K_p(t))$ as $p \to +\infty$ in the sense of convergence of finite distributions.

Theorem 1.1. *Let $(\mathrm{ST}_h)_{n\in\mathbf{Z}}$ denote independent identically distributed random variables with distribution equal for all $h \in \mathbf{Z}$ to the Sato–Tate measure*

$$\mu_{ST} = \frac{1}{\pi}\sqrt{1-(x/2)^2}\,dx$$

on $[-2, 2]$.

(1) *The random series*

$$\mathrm{K}(t) = \sum_{h\in\mathbf{Z}} \frac{e^{2\pi iht} - 1}{2\pi ih}\mathrm{ST}_h \tag{1.1}$$

converges almost surely and in law, taking symmetric partial sums, where the term $h = 0$ is interpreted as $t\mathrm{ST}_0$. Its limit, as a random function, is almost surely continuous. In addition, it is almost surely nowhere differentiable. Moreover, for any $t \in [0, 1]$, we have

$$\mathbf{E}(\mathrm{K}(t)) = 0, \quad \mathbf{V}(\mathrm{K}(t)) = t. \tag{1.2}$$

(2) *The sequence of processes $(K_p(t))_{t\in[0,1]}$ converges to the process $(\mathrm{K}(t))_{t\in[0,1]}$ in the sense of convergence of finite distributions, i.e. for every $k \geqslant 1$, for every k-tuple $0 \leqslant t_1 < \cdots < t_k \leqslant 1$, the vectors*

$$(K_p(t_1), \ldots, K_p(t_k))$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

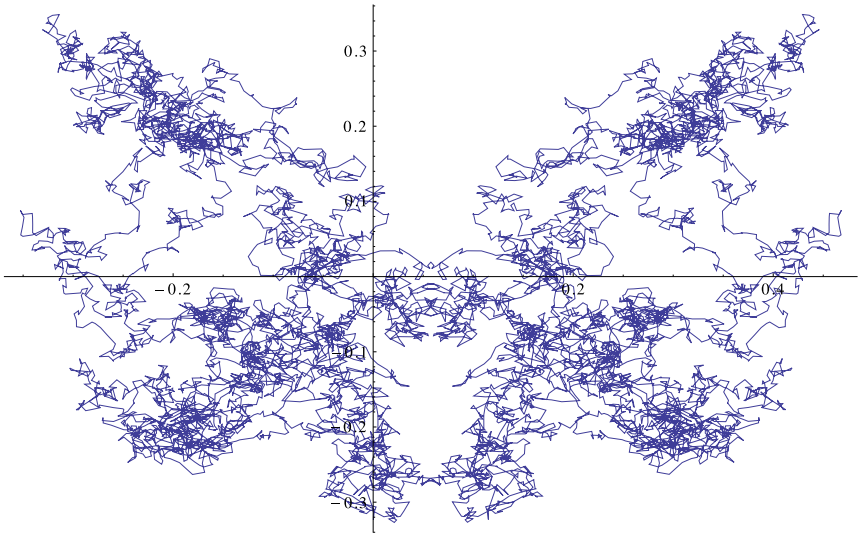

FIGURE 2. A sample of the random Fourier series $\mathrm{K}(t)$.

converge in law, as $p \to +\infty$, *to*

$$(\mathrm{K}(t_1), \dots, \mathrm{K}(t_k)).$$

Remark 1.2. (1) Lehmer [Leh76] and Loxton [Lox83, Lox85] considered the 'graphs' of various exponential sums, which are the analogues of the paths $t \mapsto K_p(t, a)$, but not necessarily over finite fields (see for instance the pictures in [Leh76, p. 127] and [Lox83, pp. 154–155]). Other studies of this type are due to Dekking and Mendès France [DMF81] and Deshouillers [Des85].

In particular, in [Lox85, p. 16], Loxton mentions briefly that the paths of Kloosterman sums 'seem to be absolutely chaotic'. Our result indicates one precise way in which this is true (or false).

(2) Figure 2 shows a sample simulation of the process $(\mathrm{K}(t))_{t\in[0,1]}$ with $N = 10\,000$ steps, obtained as follows: values at j/N are simulated for $0 \leqslant j \leqslant 9999$, by summing the partial sum of the random series between -5000 and 5000 (using samples of a Sato–Tate distribution), and then the corresponding points are interpolated linearly as in the Kloosterman paths.

Intuitively, the statement of Theorem 1.1 is not quite satisfactory, because we may wish to have convergence in law of the processes as random elements in the space $C([0,1])$ of continuous functions from $[0,1]$ to $\mathbf{C}$. We will see that some highly natural conjectures concerning short exponential sums lead to this conclusion (see § 3). Moreover, we can show unconditionally such a stronger convergence in law in two cases:

(1) if we consider the family of Kloosterman paths also on average over all additive characters $x \mapsto \psi_p(\alpha x)$ for $\alpha \in \mathbf{F}_p^\times$;
(2) for the partial sums of the family of cubic exponential sums (sometimes known as Birch sums).

We will also see that this stronger result extends (with possibly different limiting distribution) to a number of other cases.

We introduce some notation for this purpose. For each prime p, we assume that we are given a probability space Ω_p and a family $\mathcal{X}_p = (\xi_p(x))_{x\in\mathbf{F}_p}$ of complex-valued random variables on Ω_p.

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

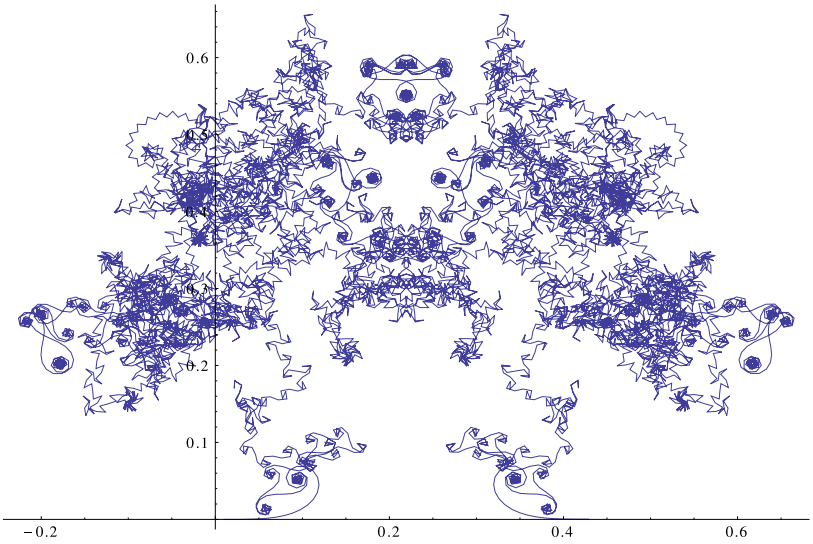

Figure 3. Path of Birch sum $\mathrm{Bi}_{10\,007}(1)$.

We then define the associated path process $K_p^{\mathcal{X}}(t)$ for $t \in [0,1]$ by parameterizing by $t \in [0,1]$ the polygonal path joining the successive partial sums

$$\omega \mapsto \frac{1}{\sqrt{p}} \sum_{0 \leqslant x \leqslant j-1} \xi_p(x,\omega)$$

for $\omega \in \Omega_p$ and $0 \leqslant j \leqslant p-1$, which we interpolate between j/p and $(j+1)/p$.

We can also have a family $\mathcal{X}_p = (\xi_p(x))_{x \in \mathbf{F}_p^\times}$ parameterized by $\mathbf{F}_p^\times$ (as in the case of Theorem 1.1), and then we define $K_p^{\mathcal{X}}(t)$ for $t \in [0,1]$ as in that case, interpolating between $j/(p-1)$ and $(j+1)/(p-1)$.

We view these processes as $C([0,1])$-valued random processes. We will study them in a special case, and mention possible generalizations at the end of the paper.

Theorem 1.3. *For $\Omega_p = \mathbf{F}_p^\times$ with the uniform probability measure and*

$$\xi_p(x,a) = \psi_p(x^3 + ax)$$

for $a \in \mathbf{F}_p^\times$ and $x \in \mathbf{F}_p$, the processes $(K_p^{\mathcal{X}}(t))$ converge to the process $(\mathrm{K}(t))_{t \in [0,1]}$ in the sense of convergence in law in $C([0,1])$.

Remark 1.4. In Figure 3, we plot the function $t \mapsto K_{10\,007}^{\mathcal{X}}(t,1)$ for this choice of $\mathcal{X}$.

The complete character sums in this case are given by

$$\mathrm{Bi}_p(a) = \frac{1}{\sqrt{p}} \sum_{x \in \mathbf{F}_p} \psi_p(ax + x^3).$$

These sums were considered by Birch [Bir68, § 3], who conjectured that they are Sato–Tate distributed (on average over a). This statement was first proved by Livné [Liv87]. However, his result (and its proof) would not suffice for our applications, because the argument lacks the group-theoretic interpretation of the Sato–Tate distribution. We rely crucially on Deligne's

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

equidistribution theorem that provides it (as a very special case), as well as on algebraic properties of the monodromy group.

Looking at the graph for the Birch sums, one notices not only the general similarity with the graphs of Kloosterman paths (and of the random Fourier series $\mathrm{K}(t)$), but also clear local differences. These are intuitively explained by the fact that the function $x \mapsto \psi_p(ax+x^3)$ is the restriction of a smooth function defined on $\mathbf{R}$, and that in very short ranges (much shorter than $0 \leqslant x \leqslant pt$ for fixed t), the sum is well approximated by oscillatory integrals which exhibit the type of spirals and curlicues visible in the picture (see the pictures in [Lox83] or [Des85]). It would be quite interesting to find a probabilistic statement that reflects this difference between Kloosterman sums and Birch sums.

For Kloosterman sums, as we mentioned, we can also average over the additive character to get convergence in law, as follows.

THEOREM 1.5. *For p prime, $(\alpha, a) \in \mathbf{F}_p^\times \times \mathbf{F}_p^\times$, let*

$$t \mapsto \mathcal{K}_p(t;\alpha,a)$$

denote the continuous function interpolating the partial sums

$$\frac{1}{\sqrt{p}} \sum_{1 \leqslant x \leqslant j} \psi_p(\alpha(ax+\bar{x})).$$

The processes $(\mathcal{K}_p(t))$ on the probability spaces

$$(\mathbf{F}_p^\times \times \mathbf{F}_p^\times, \textit{uniform probability})$$

converge to the process $(\mathrm{K}(t))_{t\in[0,1]}$ in the sense of convergence in law in $C([0,1])$.

We recall (see e.g. [Bil99, ch. 1, § 1] or [RY99, Definition 0.5.5]) that the convergence in law to K of any sequence of $C([0,1])$-valued processes $(L_p(t))_{t\in[0,1]}$ in $C([0,1])$ means that for any map

$$\varphi : C([0,1]) \longrightarrow \mathbf{C},$$

which is continuous and bounded on $C([0,1])$, with respect to the topology of uniform convergence, we have

$$\lim_{p\to+\infty} \mathbf{E}(\varphi(L_p)) = \mathbf{E}(\varphi(\mathrm{K})).$$

This condition is stronger than the convergence of finite-dimensional distributions. Indeed, given the convergence of finite-dimensional distributions of $(L_p(t))$ to those of $(\mathrm{K}(t))$, one knows that convergence in law is equivalent to the weak-compactness property known as *tightness* (this is Prokhorov's theorem; see e.g. [Bil99, Theorem 7.1]).

As an application of Theorems 1.3 and 1.5, we will obtain fairly sharp bounds for the probability of large values of partial sums of the corresponding families of exponential sums, as follows.

THEOREM 1.6. *There exists $c > 0$ such that we have*

$$c^{-1}\exp(-\exp(cA)) \leqslant \lim_{p\to+\infty} \frac{1}{p-1}\left|\left\{a \in \mathbf{F}_p^\times \mid \max_{0\leqslant j\leqslant p-1} \frac{1}{\sqrt{p}}\left|\sum_{0\leqslant x\leqslant j} \psi_p(ax+x^3)\right| \geqslant A\right\}\right| \leqslant c\exp(-\exp(c^{-1}A))$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

as well as

$$\begin{aligned} & c^{-1}\exp(-\exp(cA)) \\ & \leqslant \lim_{p\to+\infty}\frac{1}{(p-1)^2}\left|\left\{(\alpha,a)\in\mathbf{F}_p^\times\times\mathbf{F}_p^\times \mid \max_{1\leqslant j\leqslant p-1}\frac{1}{\sqrt{p}}\left|\sum_{1\leqslant x\leqslant j}\psi_p(\alpha(ax+\bar{x}))\right|\geqslant A\right\}\right| \\ & \leqslant c\exp(-\exp(c^{-1}A)) \end{aligned}$$

for all $A>0$.

As we will clearly see, convergence of finite distributions and tightness are valid in great generality for many processes corresponding to partial sums of one-variable exponential sums over finite fields. They follow naturally from two important properties.

(1) Computation and properties of the *monodromy group* of certain families of exponential sums (these are the Kloosterman sums for Theorems 1.1 and 1.5, and the Birch sums for Theorem 1.3, but other cases lead to slightly different sums). This is used to prove convergence of finite distributions.

(2) Existence of estimates (in a suitable averaged form) of *short* sums of the original summands, over (arbitrarily located) intervals of length y close to $\sqrt{p}$ in logarithmic scale; this is used to prove tightness. Such sums are at the edge of the so-called *Pólya–Vinogradov range*, which refers to y a bit larger than $\sqrt{p}\log p$, and which can be treated in considerable generality.

The first ingredient is exceptionally deep: it involves all of Deligne's work on the Riemann Hypothesis over finite fields [Del80], as well as many additional algebraic and geometric results. The second ingredient is also very delicate, and is not currently known in great generality, although one can certainly conjecture that it should be true under very general conditions. (It is, at the current time, not known for Kloosterman sums when averaging only over a, which is the reason why Theorem 1.5 requires additional averaging over the additive character.)

It is a very appealing and striking feature of this work that it shows how these two arithmetic aspects of exponential sums are unified to contribute to a single clear conclusion. Since both properties are very important in many applications in different ways, this is quite an interesting phenomenon. In fact, some sporadic relations between these two types of properties have already appeared (e.g. in the proof of Burgess's bounds for short character sums, see e.g. [IK04, Theorem 12.6], or in more recent work of Fouvry and Michel [FM98] on exponential sums over primes). None is, as far as we know, as direct and clearly focused as the phenomenon that we present.

The outline of the remainder of the paper is as follows: in § 2, we present the proof of Theorem 1.1, and more generally of convergence of finite distributions for the situation of Theorem 1.3 (and potentially many other cases, possibly for a different random Fourier series than $\mathrm{K}(t)$). In § 3, we address the additional condition of tightness, and relate it to short exponential sums. In § 4, we give applications, especially proving Theorem 1.6. Finally, in § 5, we make a few remarks concerning other potential cases of convergence in law of paths of exponential sums, as well as concerning a few earlier works that have some similarity with this paper (besides those of Loxton already mentioned).

Notation

For a prime p and a function $\varphi:\mathbf{F}_p\longrightarrow\mathbf{C}$, we denote by

$$\hat{\varphi}(h)=\frac{1}{\sqrt{p}}\sum_{x\in\mathbf{F}_p}\varphi(x)\psi_p(hx)$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

the unitarily normalized Fourier transform modulo p. We then have the inversion formula

$$\varphi(x) = \frac{1}{\sqrt{p}} \sum_{h \in \mathbf{F}_p} \hat{\varphi}(x)\psi_p(-hx)$$

and the Plancherel identity

$$\sum_{x \in \mathbf{F}_p} \varphi_1(x)\overline{\varphi_2(x)} = \sum_{h \in \mathbf{F}_p} \hat{\varphi}_1(h)\overline{\hat{\varphi}_2(h)}.$$

For any probability space $(\Omega, \Sigma, \mathbf{P})$, we denote by $\mathbf{P}(A)$ the probability of some event A and, for a $\mathbf{C}$-valued random variable X defined on Ω, we denote by $\mathbf{E}(X)$ the expectation and by $\mathbf{V}(X) = \mathbf{E}(|X - \mathbf{E}(X)|^2)$ the variance of X, when they exist. We sometimes use different probability spaces, but keep the same notation for all expectations and probabilities.

For $\sigma > 0$, a σ-subgaussian (real-valued) random variable N is a random variable such that

$$\mathbf{E}(e^{\lambda N}) \leqslant e^{\sigma^2\lambda^2/2}$$

for all $\lambda \in \mathbf{R}$. We then have

$$\mathbf{E}(|N|^k) \leqslant c_k\sigma^k$$

for any $k \geqslant 0$, where $c_k = k2^{k/2}\Gamma(k/2)$; in particular, $c_4 = 16$.

For $\sigma > 0$, it will be convenient to say that a complex-valued random variable $N = R + iI$ is σ-subgaussian, with R and I real-valued, if R and I are σ-subgaussian (we make no assumption on the independence or not of R and I). We then have

$$\mathbf{E}(|N|^k) \leqslant c'_k\sigma^k \tag{1.3}$$

for some $c'_k > 0$. Here one can for instance take $c'_4 = 256$. If N_1, N_2 are σ_i-subgaussian and independent (real- or complex-valued), then $N_1 + N_2$ is $\sqrt{\sigma_1^2 + \sigma_2^2}$-subgaussian.

We will write $\|\varphi\|_\infty$ for the supremum norm of a continuous function φ on $[0, 1]$; to avoid confusion, if X is a random variable defined on a space Ω, we will write $\|X\|_{L^\infty(\Omega)}$ for the (essential) supremum of X.

2. Proof of convergence of finite distributions

We begin by studying the random series $\mathrm{K}(t)$. This will prove the first part of Theorem 1.1, and some additional properties are relevant to the remainder of the proof of Theorem 1.1.

PROPOSITION 2.1. (1) *For any fixed $t \in [0, 1]$, the symmetric partial sums*

$$\mathrm{K}_m(t) = \sum_{|h|<m/2} \frac{e^{2i\pi ht} - 1}{2i\pi h}\mathrm{ST}_h$$

of the random series defining $\mathrm{K}(t)$ converge to $\mathrm{K}(t)$ in law, almost surely, and in every space $L^q(\Omega)$ for $1 \leqslant q < +\infty$, where Ω is the probability space on which the Sato–Tate variables ST_h are defined. In fact, we have

$$\|\mathrm{K}_m(t)\|_{L^\infty(\Omega)} \ll (\log m) \tag{2.1}$$

and

$$\mathbf{E}(|\mathrm{K}(t) - \mathrm{K}_m(t)|) \ll m^{-1/2} \tag{2.2}$$

for $m \geqslant 1$, where the implied constants are absolute.

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

(2) *For any $t \in [0,1]$, the Laplace transform*

$$\mathbf{E}(e^{\lambda \operatorname{Re}(\mathrm{K}(t))+\nu \operatorname{Im}(\mathrm{K}(t))})$$

is well defined for all non-negative (λ, ν). In particular, $\mathrm{K}(t)$ has moments of all orders. Moreover, we have

$$\mathbf{E}(\mathrm{K}(t)) = 0, \quad \mathbf{V}(\mathrm{K}(t)) = t.$$

(3) *The process $(\mathrm{K}(t))_{t\in[0,1]}$ is almost surely continuous and almost surely nowhere differentiable. More precisely, it is almost surely Hölder continuous of all orders $\alpha < 1/2$ on $[0,1]$, and almost surely nowhere Hölder continuous of order $1/2$.*

Note that this indeed contains all statements in Theorem 1.1(1), in particular (1.2).

Proof. The convergence almost surely, hence in law, of the series for any fixed t is an immediate consequence of Kolmogorov's three-series theorem, together with the fact that the Sato–Tate measure has mean 0 and is compactly supported.

The other results, however, are most easily derived as consequences of general facts about random Fourier series, which we quote from the work of Kahane [Kah85].

We can write $\mathrm{K}(t) = t\mathrm{ST}_0 + 4A(2\pi t) + 4iB(2\pi t)$, where A and B are the random Fourier series

$$A(t) = \sum_{h\geqslant 1} \frac{a_h}{\pi h} \cos(ht), \quad B(t) = \sum_{h\geqslant 1} \frac{b_h}{\pi h} \cos(ht - \pi/2),$$

where

$$a_h = \frac{\mathrm{ST}_h - \mathrm{ST}_{-h}}{4}, \quad b_h = \frac{\mathrm{ST}_h + \mathrm{ST}_{-h}}{4}$$

(note that (a_h) and (b_h) are identically distributed since the Sato–Tate law is symmetric).

Both series are of the type considered in [Kah85, chs 5, 7 and 8] and, especially, note that the random variables a_h and b_h are 1-subgaussian (indeed, this is a property of any centred real random variable with absolute value bounded by 1). The existence of the Laplace transforms is then given by [Kah85, § 5.5, Theorem 1], and we see from [Kah85, § 7.4, Theorem 3] that each of $A(t)$ and $B(t)$ (hence also $\mathrm{K}(t)$) is α-Hölder on $[0,1]$ if $\alpha < 1/2$. Furthermore, it follows from [Kah85, § 8.6, Theorem 4] that each of $A(t)$ and $B(t)$ is almost surely nowhere $1/2$-Hölder continuous.

The bound (2.1) is clear since $|\mathrm{ST}_h| \leqslant 2$. Then, for any $q \geqslant 1$, the convergence of the partial sums $\mathrm{K}_m(t)$ in $L^q(\Omega)$ follows from the convergence in $L^1(\Omega)$ implied by (2.2). For the latter, it is enough to note that $\mathrm{K}(t) - \mathrm{K}_m(t)$ is σ_m-subgaussian with

$$\sigma_m^2 = \sum_{|h|\geqslant m/2} \left| \frac{e^{2i\pi ht} - 1}{2i\pi h} \right|^2 \ll \frac{1}{m},$$

together with the property (1.3) of subgaussian random variables.

Finally, from the convergence in $L^1(\Omega)$ we deduce that

$$\mathbf{E}(\mathrm{K}(t)) = \lim_{m\to+\infty} \mathbf{E}(\mathrm{K}_m(t)) = 0$$

since $\mathbf{E}(\mathrm{ST}_h) = 0$ for all h. From the convergence in $L^2(\Omega)$ and the fact that

$$\mathbf{E}(\mathrm{ST}_{h_1}\mathrm{ST}_{h_2}) = \begin{cases} 1 & \text{if } h_1 = h_2, \\ 0 & \text{if } h_1 \neq h_2 \end{cases}$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

(by the independence of the ST_h), we derive

$$\mathbf{V}(\mathrm{K}(t)) = \sum_{h\in\mathbf{Z}} \left| \frac{e^{2i\pi ht}-1}{2i\pi h} \right|^2 = t$$

by the Parseval formula for the 1-periodic function coinciding on $[0,1]$ with the characteristic function of $[0,t]$. □

Remark 2.2. It is interesting to contrast the result with the Fourier series

$$W(t) = \sum_{h\in\mathbf{Z}} \frac{e^{2i\pi ht}-1}{2i\pi h} N_h,$$

where $(N_h)_{h\in\mathbf{Z}}$ is a sequence of independent standard complex *gaussian* random variables, i.e. $G_h = R_h + iI_h$, where R_h and I_h are independent standard real gaussian random variables. Then (this was already known to Paley and Wiener) $W(t)$ is a standard complex Brownian motion (see [Kah85, ch. 16, § 3].

We now begin the proof of the second part of Theorem 1.1. The argument extends immediately to the cases considered in Theorems 1.3 and 1.5, because the main arithmetic property required is also valid then. Thus, we will only comment briefly on this part of Theorems 1.3 and 1.5 after the proof.

The key ingredient, which explains where the Fourier series expansion for the limiting distribution comes from, is the following elementary lemma.

LEMMA 2.3. *Let $p \geqslant 3$ be a prime and $t \in [0,1]$. We have*

$$\frac{1}{\sqrt{p}} \sum_{1\leqslant x\leqslant (p-1)t} \psi_p(ax+\bar{x}) = \frac{1}{\sqrt{p}} \sum_{|h|<p/2} \alpha_p(h;t)\mathrm{Kl}_p(a-h), \tag{2.3}$$

where

$$\alpha_p(h;t) = \frac{1}{\sqrt{p}} \sum_{1\leqslant x\leqslant (p-1)t} \psi_p(hx).$$

Proof. This follows from the discrete Plancherel formula, since the coefficients $h \mapsto \alpha_p(h;t)$ are the discrete Fourier coefficients modulo p of the characteristic function of the interval $1 \leqslant x \leqslant (p-1)t$. □

From the results of Katz [Kat88, Example 13.6], we know that for each fixed h, the Kloosterman sums $(\mathrm{Kl}_p(a-h))_{a\in\mathbf{F}_p}$ are asymptotically Sato–Tate distributed. Moreover, it can be shown that they are asymptotically independent (see Lemma 2.5 below, which implies this). Since moreover $\alpha_p(h;t)/\sqrt{p} \to (e^{2i\pi ht}-1)/2i\pi h$ as $p \to +\infty$ (for fixed h), the expansion (2.3) is very similar to (1.1).

We might continue the proof in this manner, but to have a more quantitative approximation (and because this will be useful in considering tightness), we use instead the method of moments. Since Proposition 2.1 shows that the Laplace transforms of the finite distributions of $(\mathrm{K}(t))$ exist, we can use this method to prove convergence in law of the finite distributions. The next proposition therefore implies Theorem 1.1(2), and concludes the proof of that result.

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

Proposition 2.4. *Let $k \geqslant 1$ be given, and $0 \leqslant t_1 < \cdots < t_k \leqslant 1$ be fixed. Fix also non-negative integers $(n_1, \dots, n_k)$ and $(m_1, \dots, m_k)$. Let*

$$M_p = \frac{1}{p-1} \sum_{a \in \mathbf{F}_p^\times} \prod_{i=1}^{k} K_p(t_i, a)^{n_i} \overline{K_p(t_i, a)}^{m_i}.$$

We have

$$M_p = \mathbf{E}\left(\prod_{i=1}^{k} \mathrm{K}(t_i)^{n_i} \overline{\mathrm{K}(t_i)}^{m_i} \right) + O(p^{-1/2} (\log p)^{m+n})$$

for $p \geqslant 2$, where $n = n_1 + \cdots + n_k$, $m = m_1 + \cdots + m_k$ and the implied constant depends only on m and n.

Since the notation may obscure the essential arithmetical point, the reader is encouraged to first read through the proof under the assumption that $k = 1$.

We fix once and for all the sequence $(\mathrm{ST}_h)_{h \in \mathbf{Z}}$ of independent Sato–Tate random variables used to define the process $\mathrm{K}(t)$.

Proof. First of all, we deal with the linear interpolation involved in the definition of $K_p(t)$. Let

$$\tilde{K}_p(t, a) = \frac{1}{\sqrt{p}} \sum_{1 \leqslant x \leqslant (p-1)t} \psi_p(ax + \bar{x})$$

for p prime, $a \in \mathbf{F}_p^\times$ and $t \in [0, 1]$. This is a discontinuous function of t, and we have

$$|K_p(t, a) - \tilde{K}_p(t, a)| \leqslant \frac{1}{\sqrt{p}} \tag{2.4}$$

for all p, a and t. We now use the formula (2.3) above, which states that

$$\tilde{K}_p(t, a) = \frac{1}{\sqrt{p}} \sum_{|h| < p/2} \alpha_p(h; t) \mathrm{Kl}_p(a - h). \tag{2.5}$$

We observe that it is well known that

$$\sum_{|h| < p/2} |\alpha_p(h; t)| \leqslant \sqrt{p} (\log 3p) \tag{2.6}$$

for all $t \in [0, 1]$, so that in particular we have

$$|\tilde{K}_p(t, a)| \leqslant 2(\log 3p) \tag{2.7}$$

for all p, t and a, by Weil's bound for Kloosterman sums.

We deduce from this that

$$\left| \prod_{i=1}^{k} K_p(t_i, a)^{n_i} \overline{K_p(t_i, a)}^{m_i} - \prod_{i=1}^{k} \tilde{K}_p(t_i, a)^{n_i} \overline{\tilde{K}_p(t_i, a)}^{m_i} \right| \ll p^{-1/2} (\log p)^{m+n},$$

where the implied constant depends only on m and n. Hence, it is enough to prove the moment estimate for

$$\tilde{M}_p = \frac{1}{p-1} \sum_{a \in \mathbf{F}_p^\times} \prod_{i=1}^{k} \tilde{K}_p(t_i, a)^{n_i} \overline{\tilde{K}_p(t_i, a)}^{m_i}.$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

We compute $\tilde{M}_p$ by replacing each $\tilde{K}_p(t_i, a)$ and its conjugate by the formula (2.5) and taking the (n_i)th or (m_i)th power. We obtain

$$\tilde{M}_p = \frac{1}{p^{(m+n)/2}} \frac{1}{p-1} \sum_{a \in \mathbf{F}_p^\times} \sum_{\boldsymbol{h}_1, \dots, \boldsymbol{h}_k} \cdots \sum \alpha_p(\boldsymbol{h}_1; t_1) \cdots \alpha_p(\boldsymbol{h}_k; t_k) \times \prod_{l=1}^{n_1+m_1} \mathrm{Kl}_p(a - h_{1,l}) \cdots \prod_{l=1}^{n_k+m_k} \mathrm{Kl}_p(a - h_{k,l}),$$

where each

$$\boldsymbol{h}_j = (h_{j,1}, \dots, h_{j,n_j}, h_{j,n_j+1}, \dots h_{j,n_j+m_j})$$

ranges over all $(n_j + m_j)$-tuples of integers $h_{j,l}$ in $]-p/2, p/2[$ and

$$\alpha_p(\boldsymbol{h}_j; t_j) = \prod_{l=1}^{n_j} \alpha_p(h_{j,l}; t_j) \prod_{l=1}^{m_j} \overline{\alpha_p(h_{j,n_j+l}; t_j)}.$$

Exchanging the order of the sums, we deduce that

$$\tilde{M}_p = \frac{1}{p^{(m+n)/2}} \sum_{\boldsymbol{h}_1, \dots, \boldsymbol{h}_k} \cdots \sum \prod_{1 \leqslant j \leqslant k} \alpha_p(\boldsymbol{h}_j; t_j) S(\boldsymbol{h}_1, \dots, \boldsymbol{h}_k; p)$$

with

$$S(\boldsymbol{h}_1, \dots, \boldsymbol{h}_k; p) = \frac{1}{p-1} \sum_{a \in \mathbf{F}_p^\times} \prod_{l=1}^{n_1+m_1} \mathrm{Kl}_p(a - h_{1,l}) \cdots \prod_{l=1}^{n_k+m_k} \mathrm{Kl}_p(a - h_{k,l}).$$

The sums $S(\boldsymbol{h}_1, \dots, \boldsymbol{h}_k; p)$ are complete sums of products of Kloosterman sums. The crucial point, which we explain below in Lemma 2.5, is that from Deligne's Riemann Hypothesis over finite fields, the computation of the geometric monodromy group of the Kloosterman sheaf of rank 2 by Katz [Kat88] and the Goursat–Kolchin–Ribet criterion [Kat90, § 1.8], we can derive the estimate

$$S(\boldsymbol{h}_1, \dots, \boldsymbol{h}_k; p) = \mathbf{E}\left(\prod_{l=1}^{n_1+m_1} \mathrm{ST}_{h_{1,l}} \cdots \prod_{l=1}^{n_k+m_k} \mathrm{ST}_{h_{k,l}} \right) + O(p^{-1/2}),$$

where the implied constant depends only on m and n.

By (2.6), the contribution E_p of the error terms to $\tilde{M}_p$ is bounded by

$$E_p \ll p^{-1/2} \frac{1}{p^{(m+n)/2}} p^{(m+n)/2} (\log 3p)^{m+n} \ll p^{-1/2} (\log 3p)^{m+n},$$

where the implied constant depends only on m and n.

On the other hand, by reverting the computation, we get

$$\tilde{M}_p = \mathbf{E}\left(\prod_{i=1}^{k} X_p(t_i)^{n_i} \overline{X_p(t_i)}^{m_i} \right) + O(p^{-1/2} (\log p)^{m+n}),$$

where the random variables $X_p(t)$ are given by

$$X_p(t) = \sum_{|h| < p/2} \frac{\alpha_p(h; t)}{\sqrt{p}} \mathrm{ST}_h. \tag{2.8}$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

We now denote

$$\beta(h;t) = \frac{e^{2i\pi ht} - 1}{2i\pi h}$$

for $h \in \mathbf{Z}$, with $\beta(0;t) = t$, and we consider the partial sums

$$\mathrm{K}_p(t) = \sum_{|h|<p/2} \beta(h;t)\mathrm{ST}_h$$

of $\mathrm{K}(t)$. From (2.1) and (2.2) in Proposition 2.1, we see that

$$\mathbf{E}\left(\prod_{i=1}^{k} \mathrm{K}_p(t_i)^{n_i}\overline{\mathrm{K}_p(t_i)}^{m_i}\right) = \mathbf{E}\left(\prod_{i=1}^{k} \mathrm{K}(t_i)^{n_i}\overline{\mathrm{K}(t_i)}^{m_i}\right) + O(p^{-1/2}(\log p)^{m+n}),$$

where the implied constant depends only on m and n.

It is therefore enough to prove that

$$\mathbf{E}\left(\prod_{i=1}^{k} X_p(t_i)^{n_i}\overline{X_p(t_i)}^{m_i}\right) = \mathbf{E}\left(\prod_{i=1}^{k} \mathrm{K}_p(t_i)^{n_i}\overline{\mathrm{K}_p(t_i)}^{m_i}\right) + O(p^{-1/2}(\log p)^{m+n})$$

in order to finish the proof of the proposition.

In view of the bound (2.1) and the analogue

$$\|X_p(t)\|_{L^\infty(\Omega)} \ll (\log p),$$

where the implied constant is absolute, it suffices to prove that for any fixed $t \in [0,1]$, we have

$$\mathbf{E}(|X_p(t) - \mathrm{K}_p(t)|^2) \ll p^{-1}.$$

But, since the random variables ST_h are independent with $\mathbf{E}(\mathrm{ST}_h) = 0$ and $\mathbf{E}(|\mathrm{ST}_h|^2) = 1$, we have

$$\mathbf{E}(|X_p(t) - \mathrm{K}_p(t)|^2) = \sum_{|h|<p/2} \left|\frac{\alpha_p(h;t)}{\sqrt{p}} - \beta(h;t)\right|^2.$$

By definition and summing a geometric sum, we get

$$\frac{\alpha_p(h;t)}{\sqrt{p}} = \frac{1}{p}\sum_{1\leqslant x\leqslant (p-1)t} \psi_p(hx) = \frac{\psi_p(h)}{p}\frac{1-\psi_p(h\lfloor (p-1)t\rfloor)}{1-\psi_p(h)},$$

with the convention that

$$\frac{\alpha_p(0;t)}{\sqrt{p}} = \frac{\lfloor (p-1)t\rfloor}{p}.$$

For $h = 0$, we therefore find that

$$\frac{\alpha_p(0;t)}{\sqrt{p}} - \beta(0;t) = \frac{\lfloor (p-1)t\rfloor}{p} - t \ll \frac{1}{p}$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

for all $t \in [0,1]$ and $p \geqslant 3$. For all h such that $1 \leqslant |h| < p/2$, we can then write for instance

$$\begin{aligned}\frac{\alpha_p(h;t)}{\sqrt{p}} - \beta(h;t) = &\frac{(\psi_p(\lfloor (p-1)t\rfloor h)-1) - (e(ht)-1)}{2i\pi h}\\ &- (\psi_p(\lfloor (p-1)t\rfloor h)-1)\left(\frac{1}{2i\pi h} - \frac{1}{p(\psi_p(h)-1)}\right)\\ &+ (\psi_p(h)-1)\frac{\psi_p(\lfloor (p-1)t\rfloor h)-1}{p(\psi_p(h)-1)},\end{aligned}$$

and simple bounds for the three terms show that we also have

$$\frac{\alpha_p(h;t)}{\sqrt{p}} - \beta(h;t) \ll \frac{1}{p}$$

uniformly for all $t \in [0,1]$ and all p.

Squaring and summing over h, it follows therefore that

$$\mathbf{E}(|X_p(t) - \mathrm{K}_p(t)|^2) \ll p^{-1},$$

which gives the desired bound and finishes the proof. □

Here is the crucial arithmetic lemma that we used.

LEMMA 2.5. *With notation as in the proof, we have*

$$S(\boldsymbol{h}_1,\dots,\boldsymbol{h}_k;p) = \mathbf{E}\left(\prod_{l=1}^{n_1+m_1} Z_{h_{1,l}} \cdots \prod_{l=1}^{n_k+m_k} Z_{h_{k,l}}\right) + O(p^{-1/2}),$$

where the $(Z_h)_{h\in\mathbf{F}_p}$ are independent random variables with Sato–Tate distributions, and the implied constant depends only on m and n.

Proof. We can write

$$S(\boldsymbol{h}_1,\dots,\boldsymbol{h}_k;p) = \frac{1}{p-1}\sum_{a\in\mathbf{F}_p^\times}\prod_{\tau\in\mathbf{F}_p}\mathrm{Kl}_p(a+\tau)^{\mu(\tau)},$$

where

$$\mu(\tau) = \sum_{j=1}^{k} |\{1 \leqslant l \leqslant n_j+m_j \mid h_{j,l} = \tau \pmod p\}|,$$

for any $\tau \in \mathbf{F}_p$, is the multiplicity of the factor $\mathrm{Kl}_p(a+\tau)$ among the shifted Kloosterman sums in $S(\boldsymbol{h}_1,\dots,\boldsymbol{h}_k;p)$.

This type of sums of products can be estimated by the Riemann Hypothesis over finite fields, as explained in detail in [FKM15]. Indeed, the result we claim follows from [FKM15, Corollary 1.7, § 3(b)], but we recall the basic argument.

Katz showed [Kat88, Theorem 11.1] that the geometric and arithmetic monodromy groups of the Kloosterman sheaf $\mathcal{F} = \mathcal{K}\ell_2$ are equal and isomorphic to SL_2. Furthermore, if $\tau \neq 0$, there does not exist a rank 1 sheaf $\mathcal{L}$ such that

$$[+\tau]^*\mathcal{K}\ell_2 \simeq \mathcal{K}\ell_2 \otimes \mathcal{L}$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

(most simply seen here because the left-hand side is unramified at 0, while the right-hand side is ramified). Using the Goursat–Kolchin–Ribet criterion [Kat90, § 1.8], it follows that the geometric and arithmetic monodromy groups of

$$\bigoplus_{\tau\in\mathbf{F}_p}[+\tau]^*\mathcal{F}$$

are equal to $\mathrm{SL}_2\times\cdots\times\mathrm{SL}_2$. The Riemann Hypothesis then gives the asymptotic formula

$$S(\boldsymbol{h}_1,\dots,\boldsymbol{h}_k;p)=\prod_{\tau\in\mathbf{F}_p}A(\mu(\tau))+O(p^{-1/2}),$$

where $A(\mu)$, for any integer $\mu\geqslant 0$, denotes the multiplicity of the trivial representation of SU_2 in the μth tensor power of its standard two-dimensional representation, and the implied constant depends only on

$$\sum_{\tau\in\mathbf{F}_p}\mu(\tau)=\sum_i(n_i+m_i)=m+n.$$

However, we have by character theory

$$A(\mu)=\mathbf{E}(\mathrm{ST}^\mu)$$

for any Sato–Tate distributed random variable ST and $\mu\geqslant 0$. Thus, by reversing the computation, we see that

$$\prod_{\tau\in\mathbf{F}_p}A(\mu(\tau))=\mathbf{E}\Bigg(\prod_{l=1}^{n_1+m_1}Z_{h_{1,l}}\cdots\prod_{l=1}^{n_k+m_k}Z_{h_{k,l}}\Bigg),$$

where the Z_h are independent and Sato–Tate distributed. □

Remark 2.6. (1) We emphasize once again that, for our application, it is essential to obtain the correct main term, and not only a criterion for cancellation in these sums. This contrasts with many other applications of such estimates.

(2) Interestingly, similar sums of products of shifted Kloosterman sums also occur in a recent work of Irving [Irv15] concerning the divisor function in arithmetic progressions to smooth moduli; there, however, only the cancellation criterion is required.

We can now see why convergence of finite distributions also holds in the case considered in Theorem 1.3. We have then

$$\frac{1}{\sqrt{p}}\sum_{0\leqslant x\leqslant pt}\psi_p(ax+x^3)=\frac{1}{\sqrt{p}}\sum_{|h|<p/2}\alpha'_p(h;t)\mathrm{Bi}_p(a-h)$$

exactly as in (2.5), where $\alpha'_p(h;t)$ are the discrete Fourier coefficients of the interval $0\leqslant n\leqslant pt$ modulo p and

$$\mathrm{Bi}_p(a)=\frac{1}{\sqrt{p}}\sum_{x\in\mathbf{F}_p}\psi_p(ax+x^3)$$

are the Birch sums. Since Katz also showed that the geometric and arithmetic monodromy groups of the lisse sheaf $\mathcal{G}$ on $\mathbf{A}^1_{\mathbf{F}_p}$ parameterizing these sums (namely, the sheaf-theoretic Fourier transform of the Artin–Schreier sheaf $\mathcal{L}_{\psi_p(x^3)}$) are both equal to SL_2 for $p>7$ (see [Kat87, Theorem 19 and Corollary 20]), the proof of Proposition 2.4 applies essentially verbatim to give

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

convergence of finite distributions. In checking the analogue of Lemma 2.5, one has to check that the sheaf $\mathcal{G}$ is such that there is no geometric isomorphism

$$[+\tau]^*\mathcal{G} \simeq \mathcal{G} \otimes \mathcal{L},$$

where $\mathcal{L}$ is of rank 1 and $\tau \neq 0$. But, indeed, such a sheaf $\mathcal{L}$ would need to be lisse on $\mathbf{A}^1$ (since $\mathcal{G}$ is). Laumon's theory of the Fourier transform shows that $\mathcal{G}$ has unique slope $3/2$ at ∞ (see [Kat90, Theorem 7.4.1(1)]), so the only possibility for $\mathcal{L}$ is that it has slope 0 or 1 at infinity. This means that $\mathcal{L} \simeq \mathcal{L}_{\psi(hX)}$ for some h. Then the condition

$$[+\tau]^*\mathcal{G} \simeq \mathcal{G} \otimes \mathcal{L}_{\psi(hX)}$$

would imply (by taking Fourier transform) that $\mathcal{L}_{\psi(-Y^3-\tau Y)} \simeq \mathcal{L}_{\psi(-(Y+h)^3)}$, which is not the case for $\tau \neq 0$ or $h \neq 0$.

For the process of Theorem 1.5, proving convergence in finite distributions is only a matter of checking that the convergence in finite distributions for Kloosterman sums holds for any choice of non-trivial additive character modulo p, instead of ψ_p, and this is immediate.

3. Proof of tightness

Now we consider tightness to finish the proof of Theorem 1.3. More generally, we consider a sequence of processes $(K_p^{\mathcal{X}}(t))_{t\in[0,1]}$ constructed as described before Theorem 1.3, with summands $\xi_p(x,\omega)$ defined either for $x \in \mathbf{F}_p$ or $x \in \mathbf{F}_p^\times$.

We will use Kolmogorov's criterion to find a condition that implies tightness.

PROPOSITION 3.1 (Kolmogorov tightness criterion). *Let $(L_p(t))_{t\in[0,1]}$ be a sequence of $C([0,1])$-valued processes such that $L_p(0) = 0$ for all p.*

If there exist constants $\alpha > 0$, $\delta > 0$ and $C \geqslant 0$, such that for any p and any $s < t$ in $[0,1]$, we have

$$\mathbf{E}(|L_p(t) - L_p(s)|^\alpha) \leqslant C|t-s|^{1+\delta}, \tag{3.1}$$

then the sequence $(L_p(t))$ is tight.

This is found in e.g. [RY99, Theorem XIII.1.8]. We then obtain the following criterion for paths of exponential sums.

LEMMA 3.2 (Tightness and short sums). *Assume that $\mathcal{X} = (\xi_p(x))_{x\in\mathbf{F}_p}$ is defined on Ω_p, a finite set with uniform probability measure, and satisfies the following conditions.*

(1) *There exists $H \geqslant 1$ such that we have*

$$|\xi_p(x,\omega)| \leqslant 1, \quad |\hat{\xi}_p(h,\omega)| \leqslant H$$

for all primes p, $x \in \mathbf{F}_p$, $h \in \mathbf{F}_p$ and $\omega \in \Omega_p$, where

$$\widehat{\xi}_p(\omega,h) = \frac{1}{\sqrt{p}} \sum_{x\in\mathbf{F}_p} \xi_p(x,\omega)\psi_p(hx)$$

is the discrete Fourier transform of $x \mapsto \xi_p(x,\omega)$.

(2) *We have*

$$\frac{1}{|\Omega_p|} \sum_{\omega\in\Omega_p} \widehat{\xi}_p(h_1,\omega)\widehat{\xi}_p(h_2,\omega)\overline{\widehat{\xi}_p(h_3,\omega)\widehat{\xi}_p(h_4,\omega)} = \mathbf{E}(\mathrm{ST}_{h_1}\cdots\mathrm{ST}_{h_4}) + O(p^{-1/2}) \tag{3.2}$$

for all primes p and $(h_1,\dots,h_4) \in \mathbf{F}_p^4$.

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

(3) *There exist $\alpha > 0$, $\delta_1 > 0$ and $\delta_2 > 0$ such that, for any prime p and any interval $I \subset \mathbf{F}_p^\times$ of length*

$$p^{1/2-\delta_1} \leqslant |I| \leqslant p^{1/2+\delta_1},$$

we have

$$\frac{1}{|\Omega_p|}\sum_{\omega\in\Omega_p}\left|\frac{1}{\sqrt{p}}\sum_{x\in I}\xi_p(x,\omega)\right|^\alpha \ll p^{-1/2-\delta_2}. \tag{3.3}$$

Then the sequence $(K_p^{\mathcal{X}}(t))_{t\in[0,1]}$ is tight as $C([0,1])$-valued random variables. Moreover, the same holds if the summands $\mathcal{X} = (\xi_p(x))_{x\in\mathbf{F}_p^\times}$ are parameterized by $\mathbf{F}_p^\times$ instead of $\mathbf{F}_p$.

Remark 3.3. Note that (2) is, in practice, a special case of the main estimate of (the analogue of) Lemma 2.5 that is used to prove convergence in finite distributions. Moreover, (1) is a standard condition for the type of exponential sums we consider (typically, bounds on the Fourier transform would already follow from Weil's theory of exponential sums in one variable).

Thus, the practical meaning of this lemma is that, once convergence of finite distributions is known 'for standard reasons', tightness becomes a consequence of the estimate (3.3). The latter concerns the average distribution (over $\omega \in \Omega_p$) of short partial sums of the summands $\xi_p(x,\omega)$, where the length of the sums is close to $p^{1/2}$, but can be a bit smaller.

If, as one certainly expects in many cases, there exists $\eta > 0$ such that we have a uniform non-trivial individual bound

$$\frac{1}{\sqrt{p}}\sum_{x\in I}\xi_p(x,\omega) \ll p^{-\eta}$$

for all $\omega \in \Omega_p$ and all intervals I of length about $p^{1/2}$ (as in the statement of (3)), then taking $\alpha > 0$ large enough yields (3).

This will suffice for Birch sums, but is not known for Kloosterman sums at this time. However, in some cases, one can get average bounds without proving first individual estimates, and an example is given by Theorem 1.5.

We first assume the validity of Lemma 3.2, and use it to prove Theorems 1.3 and 1.5.

Proof of Theorem 1.3. Recall that $\Omega_p = \mathbf{F}_p^\times$ and $\xi_p(x,a) = \psi_p(ax + x^3)$. The first condition of Lemma 3.2 is then clear. The second condition holds by (the analogue for the Birch sums of) Lemma 2.5. For (3), the point is that individual bounds for sums over intervals of polynomials of rather short length are known, from methods such as Weyl differencing, so we can use the argument indicated in the previous remark.

Precisely, by Weyl's method, one gets

$$\sum_{x\in I}\psi_p(ax+x^3) \ll |I|^{1+\varepsilon}\left(\frac{1}{|I|}+\frac{p}{|I|^3}\right)^{1/4}$$

for $1 \leqslant |I| < p$ and for any $\varepsilon > 0$, where the implied constant depends only on ε (see e.g. [IK04, Lemma 20.3]). In particular, if we assume that

$$p^{5/12} \leqslant |I| \leqslant p^{7/12}$$

(for instance), then we have

$$\frac{1}{\sqrt{p}}\sum_{x\in I}\psi_p(ax+x^3) \ll p^{-\eta},$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

where $\eta > 0$ and the implied constant are absolute. For any $\alpha > 0$, it follows that

$$\frac{1}{p-1}\sum_{a\in\mathbf{F}_p^\times}\left|\frac{1}{\sqrt{p}}\sum_{x\in I}\psi_p(ax+x^3)\right|^\alpha \ll p^{-\alpha\eta}$$

and, selecting α large enough, we obtain the desired estimate (3.3). □

Proof of Theorem 1.5. As before, it only remains to prove (3.3) for the process $(\mathcal{K}_p(t))$ (here the summands $\psi_p(\alpha(ax+\bar{x}))$ are parameterized by $x \in \mathbf{F}_p^\times$). We take $\alpha = 4$ and compute the fourth moment (just as Kloosterman did for the full interval to get the first non-trivial bounds for Kloosterman sums).

We have

$$\begin{aligned}\frac{1}{(p-1)^2}&\sum_{(\alpha,a)\in\mathbf{F}_p^\times\times\mathbf{F}_p^\times}\left|\frac{1}{\sqrt{p}}\sum_{x\in I}\psi_p(\alpha(ax+\bar{x}))\right|^4\\ &= \frac{1}{p^2}\sum\sum_{x_1,\dots,x_4\in I}\frac{1}{p-1}\sum_{\alpha\in\mathbf{F}_p^\times}\psi_p\left(\alpha\left(\frac{1}{x_1}+\frac{1}{x_2}-\frac{1}{x_3}-\frac{1}{x_4}\right)\right)\\ &\quad\times\frac{1}{p-1}\sum_{a\in\mathbf{F}_p^\times}\psi_p(\alpha a(x_1+x_2-x_3-x_4)).\end{aligned}$$

By orthogonality of characters, this is equal to

$$\frac{1}{p(p-1)}\sum\sum_{\substack{x_1,\dots,x_4\in I\\ x_1+x_2=x_3+x_4}}\frac{1}{p-1}\sum_{\alpha\in\mathbf{F}_p^\times}\psi_p\left(\alpha\left(\frac{1}{x_1}+\frac{1}{x_2}-\frac{1}{x_3}-\frac{1}{x_4}\right)\right)+O(|I|^4p^{-3})$$

and then to

$$\frac{1}{(p-1)^2}\sum\sum_{\substack{x_1,\dots,x_4\in I\\ x_1+x_2=x_3+x_4\\ x_1^{-1}+x_2^{-1}=x_3^{-1}+x_4^{-1}}}1+O(|I|^4p^{-3}).$$

But, for any fixed x_1 and x_2, provided $x_1+x_2\neq 0$, the equations

$$\begin{cases}x_3+x_4=x_1+x_2,\\ x_3^{-1}+x_4^{-1}=x_1^{-1}+x_2^{-1}\end{cases}$$

have at most two pairs of solutions (x_3,x_4), so that the contribution of these (x_1,x_2) is at most $2|I|^2/(p-1)^2$. On the other hand, if $x_1+x_2=0$, then we also have $x_3+x_4=0$, so that these contribute also at most $|I|^2(p-1)^{-2}$. Hence, we get

$$\frac{1}{(p-1)^2}\sum_{(\alpha,a)\in\mathbf{F}_p^\times\times\mathbf{F}_p^\times}\left|\frac{1}{\sqrt{p}}\sum_{x\in I}\psi_p(\alpha(ax+\bar{x}))\right|^4\ll|I|^2p^{-2}+|I|^4p^{-3}.$$

If we take $|I|$ close to $p^{1/2}$, this is close to p^{-1}, and (3.3) therefore follows easily. □

Remark 3.4. (1) Without the average over α, we obtain

$$\begin{aligned}\frac{1}{p-1}\sum_{a\in\mathbf{F}_p^\times}\left|\frac{1}{\sqrt{p}}\sum_{x\in I}\psi_p(ax+\bar{x})\right|^4 &= \frac{1}{p(p-1)}\sum\sum_{\substack{x_1,\dots,x_4\in I\\ x_1+x_2=x_3+x_4}}\psi_p\left(\frac{1}{x_1}+\frac{1}{x_2}-\frac{1}{x_3}-\frac{1}{x_4}\right)\\ &\quad+O(|I|^4p^{-3}).\end{aligned}$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

Since the number of points of summation is about $|I|^3$ (because I is an interval), this leads to a bound $p^{-1/2}$ when $|I|$ is itself $p^{1/2}$. The difficulty is therefore that we must get some cancellation in the exponential sum over the x_i to avoid the extra average over additive characters.

(2) Interestingly, if we interpolate between the partial sums

$$\frac{1}{\sqrt{p}}\sum_{1\leqslant x\leqslant j}\psi_p(x+a\bar{x})$$

(moving the parameter a), although the end point is still the Kloosterman sum $\mathrm{Kl}_p(a)$, the corresponding process *does* converge in $C([0,1])$, but with the slightly different limit

$$\sum_{h\neq -1}\frac{e(ht)-1}{2i\pi h}\mathrm{ST}_h=\mathrm{K}(t)+\frac{e(-t)-1}{2i\pi}\mathrm{ST}_{-1}$$

(the value $h=-1$ is omitted because the relevant analogue of Lemma 2.5 involves a product over h of $\mathrm{Kl}_p(a(h+1))$, which is constant, equal to $1/\sqrt{p}$, for $h=-1$, so that any moment where $h=-1$ appears with positive multiplicity does not contribute to the asymptotic). Here tightness follows because orthogonality gives

$$\frac{1}{p-1}\sum_{a\in\mathbf{F}_p^\times}\left|\frac{1}{\sqrt{p}}\sum_{x\in I}\psi_p(x+a\bar{x})\right|^4=\frac{1}{p(p-1)}\sum_{\substack{x_1,\dots,x_4\in I\\ x_1^{-1}+x_2^{-1}=x_3^{-1}+x_4^{-1}}}\sum\psi_p(x_1+x_2-x_3-x_4)+O(|I|^4p^{-3})$$

and a result of Bourgain and Garaev [BoG14, Theorem 1] shows that the number of points of summation is $\ll|I|^{8/3+\varepsilon}$ for any $\varepsilon>0$, provided $|I|\leqslant p^{3/4}$, which is enough to verify the hypothesis of Lemma 3.2.

Proof of Lemma 3.2. We will verify Kolmogorov's criterion (3.1) for suitable $\alpha>0$ and $\delta>0$. We will deal with the case where $\mathcal{X}=(\xi_p(x))_{x\in\mathbf{F}_p}$ is parameterized by $\mathbf{F}_p$, the other being analogous.

Let $L_p(t)=K_p^{\mathcal{X}}(t)$ for p prime and $t\in[0,1]$. Let also

$$\tilde{L}_p(t,\omega)=\frac{1}{\sqrt{p}}\sum_{0\leqslant x\leqslant pt}\xi_p(x,\omega)$$

be the discontinuous analogue of $L_p(t)$. We first reduce the problem to proving a moment estimate for $\tilde{L}_p(t)-\tilde{L}_p(s)$.

We do this in two steps. First, if $|t-s|<1/p$, then the definition by linear interpolation implies that

$$|L_p(t,a)-L_p(s,a)|\leqslant\sqrt{p}|t-s|\leqslant|t-s|^{1/2}$$

and hence

$$\mathbf{E}(|L_p(t)-L_p(s)|^\alpha)\leqslant|t-s|^{\alpha/2}\tag{3.4}$$

for any $\alpha>0$ and all primes p, which is fine as soon as $\alpha>2$.

Thus, we assume from now on that $|t-s|\geqslant 1/p$. We then use the bound

$$|L_p(t,a)-\tilde{L}_p(t,a)|\leqslant\frac{1}{\sqrt{p}}$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

for all $t \in [0,1]$ and $a \in \mathbf{F}_p^\times$ (as in (2.4)) and deduce that

$$\begin{aligned}\mathbf{E}(|L_p(t)-L_p(s)|^\alpha) &\ll \mathbf{E}(|\tilde{L}_p(t)-\tilde{L}_p(s)|^\alpha)+p^{-\alpha/2}\\ &\ll \mathbf{E}(|\tilde{L}_p(t)-\tilde{L}_p(s)|^\alpha)+|t-s|^{\alpha/2}\end{aligned} \tag{3.5}$$

for any $\alpha \geqslant 1$, where the implied constant depends only on α. This shows that, provided $\alpha > 2$, we obtain (3.1) from the corresponding statement for $\tilde{L}_p(t)$. We now begin to prove this.

We denote by

$$X_p(t) = \sum_{|h|<p/2} \frac{\alpha'_p(h;t)}{\sqrt{p}}\mathrm{ST}_h$$

the analogue of the random variables in (2.8) for intervals $0 \leqslant j \leqslant pt$ instead of $1 \leqslant j \leqslant (p-1)t$. For $s \leqslant t$ in $[0,1]$, we will also denote by I the interval $ps \leqslant x \leqslant pt$, of length $|I| \asymp |t-s|p$ (recall that $|s-t| \geqslant 1/p$ now).

We denote by $\delta_1 > 0$ and $\delta_2 > 0$ the parameters in (3.3). We first make the remark that we may replace δ_1 by any smaller positive number without affecting the validity of (3.3), so that we can assume that $\delta_1 < \delta_2$.

We first claim that

$$\mathbf{E}(|\tilde{L}_p(t)-\tilde{L}_p(s)|^4) = \mathbf{E}(|X_p(t)-X_p(s)|^4) + O(p^{-1/2}(\log p)^4). \tag{3.6}$$

This indeed follows from the assumption (3.2) of Lemma 3.2, by the same method used in the proof of convergence of finite distributions.

Next, we claim that there exists $C \geqslant 0$ such that, for all s, t in $[0,1]$ and all p, we have

$$\mathbf{E}(|X_p(t)-X_p(s)|^4) \leqslant C|t-s|^2 \tag{3.7}$$

(in particular, the sequence $(X_p(t))_{t\in[0,1]}$ is itself tight by Kolmogorov's criterion). Indeed, we can use the fact that

$$X_p(t)-X_p(s) = \sum_{|h|<p/2} \frac{\alpha'_p(h;t)-\alpha'_p(h;s)}{\sqrt{p}}\mathrm{ST}_h$$

is σ_p-subgaussian, where

$$\sigma_p^2 = \frac{1}{p}\sum_{|h|<p/2}|\alpha'_p(h;t)-\alpha'_p(h;s)|^2 = \frac{1}{p}\sum_{ps\leqslant x\leqslant pt} 1 \ll |t-s|$$

(by the discrete Plancherel formula). Since, for a σ-subgaussian variable N, we have the bound $\mathbf{E}(|N|^4) \leqslant 256\sigma^4$ (see (1.3)), the claim follows.

Combining (3.6) and (3.7), we get for any $\eta > 0$ and $\varepsilon > 0$ the bound

$$\mathbf{E}(|\tilde{L}_p(t)-\tilde{L}_p(s)|^4) \ll |t-s|^{1+\eta-\varepsilon}$$

for all s and t such that $|t-s| \geqslant p^{-1/(2(1+\eta))}$, where the implied constant depends only on ε. For suitable $\eta > 0$ and $\varepsilon > 0$, this gives

$$\mathbf{E}(|\tilde{L}_p(t)-\tilde{L}_p(s)|^4) \ll |t-s|^{1+\delta'} \tag{3.8}$$

for all s, t such that $|t-s| \geqslant p^{-1/2+\delta_1}$, where $\delta' > 0$.

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

Next, suppose that $p^{-1} \leqslant |t-s| \leqslant p^{-1/2-\delta_1}$. We then note that the trivial bound

$$|\tilde{L}_p(t,\omega) - \tilde{L}_p(s,\omega)| \leqslant \frac{|I|}{\sqrt{p}} \ll p^{1/2}|t-s| \ll p^{-\delta_1}$$

leads to

$$\mathbf{E}(|\tilde{L}_p(t) - \tilde{L}_p(s)|^\alpha) \ll p^{-\alpha\delta_1} \ll |t-s|^2$$

provided $\alpha > 2\delta_1^{-1}$. Thus, we get

$$\mathbf{E}(|\tilde{L}_p(t) - \tilde{L}_p(s)|^{\alpha_1}) \ll |t-s|^2 \tag{3.9}$$

for $\alpha_1 = 4\delta_1^{-1}$ and $p^{-1} \leqslant |t-s| \leqslant p^{-1/2-\delta_1}$.

Finally, assume that

$$p^{-1/2-\delta_1} \leqslant |s-t| \leqslant p^{-1/2+\delta_1},$$

so that

$$p^{1/2-\delta_1} \ll |I| \ll p^{1/2+\delta_1}.$$

By (3.3), with $\alpha > 0$ as in that bound, we deduce that

$$\mathbf{E}(|\tilde{L}_p(t) - \tilde{L}_p(s)|^\alpha) \ll p^{-1/2-\delta_2} \ll |t-s|^{(1/2+\delta_2)/(1/2+\delta_1)} \leqslant |t-s|^{1+\delta'} \tag{3.10}$$

for some $\delta' > 0$, since we assumed that $\delta_2 > \delta_1$.

We can now combine (3.8)–(3.10). All ranges of $|t-s|$ are covered by the combination of the three bounds, and we have a suitable inequality for each range. The exponents on both sides of the inequalities do not necessarily match, however. But since we have

$$\tilde{L}_p(t,\omega) \ll (\log p)$$

uniformly (by the completion method, as in (2.7), which applies thanks to the assumption (1) of Lemma 3.2 that the Fourier transforms are uniformly bounded by H), we can replace the exponent of $|\tilde{L}_p(t) - \tilde{L}_p(s)|$ by $\max(4, \alpha_1, \alpha) > 0$, which uniformizes the exponent on the left, at the cost of a power of $\log p$. Since $|t-s| \leqslant 1$, we can also replace the exponent of $|t-s|$ by

$$0 < \delta < \min(1, \delta', \delta'')$$

and obtain then

$$\mathbf{E}(|\tilde{L}_p(t) - \tilde{L}_p(s)|^\alpha) \ll (\log p)^C |t-s|^{1+\delta}$$

for some $\alpha > 0$, $\delta > 0$, $C \geqslant 0$ and for all s, t with $1/p \leqslant |t-s| \leqslant 1$. In this range, this means that

$$\mathbf{E}(|\tilde{L}_p(t) - \tilde{L}_p(s)|^\alpha) \ll |t-s|^{1+\delta-\varepsilon}$$

for any $\varepsilon > 0$, where the implied constant depends only on $\varepsilon > 0$. Together with the introductory reduction, this verifies Kolmogorov's criterion. □

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

4. Applications

The fact that, for any fixed $t_0 \in [0,1]$, there is a limiting distribution for $K_p(t_0)$ (or for the Birch process at t_0) is already interesting, although it is only the simplest case of convergence of finite distributions. We can then use known results on sums of independent variables to deduce some interesting properties of the corresponding partial sums. We study here only the tail behaviour of the limiting distribution at t_0, and get the following proposition.

PROPOSITION 4.1. *Let $t_0 \in]0,1[$ be given. There exists a constant $c(t_0) > 0$ such that for any $A > 0$, we have*

$$\frac{1}{c(t_0)}\exp(-\exp(c(t_0)A)) \leqslant \lim_{p\to+\infty}\frac{1}{p-1}|\{a \in \mathbf{F}_p^\times \mid |\mathrm{Re}(K_p(t_0))| \geqslant A\}| \leqslant c(t_0)\exp\left(-\exp\left(\frac{A}{c(t_0)}\right)\right)$$

if $t_0 \neq 1/2$, and

$$\frac{1}{c(t_0)}\exp(-\exp(c(t_0)A)) \leqslant \lim_{p\to+\infty}\frac{1}{p-1}|\{a \in \mathbf{F}_p^\times \mid |\mathrm{Im}(K_p(t_0))| \geqslant A\}| \leqslant c(t_0)\exp\left(-\exp\left(\frac{A}{c(t_0)}\right)\right).$$

In particular, for any fixed $t_0 \in]0,1[$, the partial sums

$$\frac{1}{\sqrt{p}}\sum_{1\leqslant x\leqslant (p-1)t_0}\psi_p(ax+\bar{x})$$

are unbounded as p and a vary.

Remark 4.2. (1) For $t_0 = 1/2$, we have $\mathrm{Re}(K_p(1/2)) = \frac{1}{2}\mathrm{ST}_0$, so that the upper-bound estimate for $|\mathrm{Re}(K_p(1/2))|$ holds trivially, but the lower-bound estimate fails for $A \geqslant 1$.

(2) In fact, since we know the exact distribution of $\mathrm{K}(t_0)$, one could probably improve this result with more work, using moment methods similar to those used by Granville and Soundararajan [GS06].

Proof. We consider the real part, the imaginary part being handled similarly. By convergence of finite distributions, it is equivalent to prove that there exists $c(t_0) > 0$ such that

$$c(t_0)^{-1}\exp(-\exp(c(t_0)A)) \leqslant \mathbf{P}(|\mathrm{Re}(\mathrm{K}(t_0))| \geqslant A) \leqslant c(t_0)\exp(-\exp(c(t_0)^{-1}A))$$

for $t_0 \in]0,1[$, $t_0 \neq 1/2$.

We begin with the upper bound. We will use the martingale method explained by Ledoux and Talagrand [LT91, § 1.3], but there are other options (e.g. the work of Montgomery and Odlyzko [MO88, Theorem 2] or probabilistic methods similar to those in the later Proposition 4.4, along the lines of Montgomery-Smith's work for Rademacher series [MS90]).

We write

$$\mathrm{Re}(\mathrm{K}(t_0)) = \sum_{n\geqslant 0} X_n,$$

where

$$X_0 = t_0\mathrm{ST}_0 \quad \text{and} \quad X_n = \frac{\sin 2\pi n t_0}{2\pi n}(\mathrm{ST}_n + \mathrm{ST}_{-n})$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

for $n \geqslant 1$. The random variables $(X_n)_{n\geqslant 0}$ are independent, bounded, integrable with expectation zero. In particular, for any $N \geqslant 1$, the sum

$$\sum_{0\leqslant n\leqslant N} X_n$$

is an example of a *sum of martingale differences* as described in [LT91, p. 31] (with $d_i = X_i$), with expectation 0. By [LT91, Lemma 1.8], we have

$$\mathbf{P}\left(\left|\sum_{0\leqslant n\leqslant N} X_n\right| > A\right) \leqslant 16\exp\left(-\exp\left(\frac{A}{4a_N}\right)\right)$$

for any $A > 0$, where

$$a_N = \max_{0\leqslant i\leqslant N}(i+1)\|X_i\|_{L^\infty(\Omega)}.$$

For all $N \geqslant 1$, we have

$$\begin{aligned} a_N \leqslant a &= \sup_{n\geqslant 0}(n+1)\|X_n\|_{L^\infty(\Omega)} \\ &= \max\left(2|t_0|, \sup_{n\geqslant 1}\frac{2n}{\pi n}|\sin(2\pi n t_0)|\right) \\ &= \max\left(2|t_0|, \frac{2}{\pi}\sup_{n\geqslant 1}|\sin(2\pi n t_0)|\right) \leqslant \max\left(2|t_0|, \frac{2}{\pi}\right) \end{aligned}$$

and, in fact, if t_0 is irrational, then $a = \max(2|t_0|, 2/\pi)$ (and otherwise it can be computed quite easily as a function of the denominator of t_0).

Thus, for any $N \geqslant 1$ and $A > 0$, we get

$$\mathbf{P}\left(\left|\sum_{0\leqslant n\leqslant N} X_n\right| > A\right) \leqslant 16\exp\left(-\exp\left(\frac{A}{4a}\right)\right).$$

We can now easily let $N \to +\infty$: fix $A > 0$ and let $\varepsilon > 0$; there exists $N \geqslant 1$ such that

$$\mathbf{P}\left(\left|\sum_{n>N} X_n\right| > \varepsilon\right) \leqslant \varepsilon$$

(convergence in probability of the partial sums of $\mathrm{K}(t_0)$, which follows from Proposition 2.1) and thus

$$\mathbf{P}\left(\left|\sum_{n\geqslant 0} X_n\right| > A\right) \leqslant \varepsilon + \mathbf{P}\left(\left|\sum_{0\leqslant n\leqslant N} X_n\right| > A-\varepsilon\right) \leqslant \varepsilon + 16\exp\left(-\exp\left(\frac{A-\varepsilon}{4a}\right)\right).$$

Letting $\varepsilon \to 0$ gives the desired upper bound for the real part, in a rather precise and explicit form. The lower bound can here be derived elementarily. Write

$$X_n = r_n Y_n$$

with

$$\begin{aligned} r_0 = t_0, \quad r_n &= \frac{\sin 2\pi n t_0}{2\pi n} \quad \text{for } n \geqslant 1, \\ Y_0 = \mathrm{ST}_0, \quad Y_n &= \mathrm{ST}_n + \mathrm{ST}_{-n} \quad \text{for } n \geqslant 1. \end{aligned}$$

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

Using symmetry and independence of the variables (Y_n), we have

$$\mathbf{P}(\mathrm{Re}(\mathrm{K}(t_0)) > A) \geqslant \frac{1}{4}\prod_{1\leqslant n\leqslant N}\mathbf{P}(|Y_n|\geqslant 1) = \frac{1}{4}u^{-N-1},$$

where $u = \mathbf{P}(Y_1 \geqslant 1) = \mathbf{P}(Y_1 \leqslant -1) > 0$ for any N such that

$$\sum_{1\leqslant n\leqslant N}|r_n| > A.$$

One can find such an N with $N \ll \exp(\delta_3 A)$ for some $\delta_3 > 0$, where δ_3 and the implied constant depend on t_0 (e.g. again equidistribution of nt_0 in $\mathbf{R}/\mathbf{Z}$, if t_0 is irrational, or by periodicity, using the fact that t_0 is assumed to be $\neq 1/2$). This gives the desired lower bound.

As already mentioned, the imaginary part is handled similarly; note that there is no exception similar to $t_0 = 1/2$ because $(\cos(2\pi n t_0) - 1)/(2\pi n)$ vanishes for all $n \neq 0$ only if $t_0 \in \{0, 1\}$. □

We now consider Theorem 1.6, which is an example of an application requiring convergence in law in $C([0,1])$. Since the norm map $\varphi \mapsto \|\varphi\|_\infty$ is continuous on $C([0,1])$, it follows formally from the definition of convergence in law that the random variables $(\alpha, a) \mapsto \|\mathcal{K}_t(t;\alpha,a)\|_\infty$ (respectively $a \mapsto \|K_p^{\mathfrak{X}}(a)\|_\infty$ for Birch sums) converge in law to the random variable $\|\mathrm{K}\|_\infty$ as $p \to +\infty$. (Recall that the L^∞-norm refers to the space $C([0,1])$, and not to the space $L^\infty(\Omega)$.)

Moreover, since the maximum of the modulus along a segment in $\mathbf{C}$ is achieved at one of the end points, we have

$$\|\mathcal{K}_p(\cdot;\alpha,a)\|_\infty = \max_{1\leqslant j\leqslant p-1}\frac{1}{\sqrt{p}}\left|\sum_{1\leqslant x\leqslant j}\psi_p(\alpha(ax+\bar{x}))\right|,$$

respectively

$$\|K_p^{\mathfrak{X}}(\cdot;a)\|_\infty = \max_{0\leqslant j\leqslant p-1}\frac{1}{\sqrt{p}}\left|\sum_{0\leqslant x\leqslant j}\psi_p(ax+x^3)\right|.$$

Hence, defining μ as the distribution of $\|\mathrm{K}\|_\infty$, Theorem 1.3 gives the following proposition.

PROPOSITION 4.3. *There exists a probability measure μ on $[0,+\infty[$ such that for any bounded continuous function f on $[0,+\infty[$, we have*

$$\lim_{p\to+\infty}\frac{1}{p-1}\sum_{a\in\mathbf{F}_p^\times} f\left(\max_{0\leqslant j\leqslant p-1}\frac{1}{\sqrt{p}}\left|\sum_{0\leqslant x\leqslant j}\psi_p(ax+x^3)\right|\right) = \int f(x)\,d\mu(x)$$

and

$$\lim_{p\to+\infty}\frac{1}{(p-1)^2}\sum\sum_{(\alpha,a)\in\mathbf{F}_p^\times\times\mathbf{F}_p^\times} f\left(\max_{0\leqslant j\leqslant p-1}\frac{1}{\sqrt{p}}\left|\sum_{0\leqslant x\leqslant j}\psi_p(\alpha(ax+\bar{x}))\right|\right) = \int f(x)\,d\mu(x).$$

We expect of course that the same holds for Kloosterman sums without the average over α.

Using this result, we can now prove Theorem 1.6, by getting suitable tail bounds for the limiting distribution of K. More precisely, Theorem 1.6 follows from the next proposition.

PROPOSITION 4.4. *There exists $c > 0$ such that*

$$c^{-1}\exp(-\exp(cA)) \leqslant \mathbf{P}(\|\mathrm{K}\|_\infty \geqslant A) \leqslant c\exp(-\exp(c^{-1}A))$$

for any $A > 0$.

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

Proof. The lower bound is an immediate consequence of the lower bound in Proposition 4.1 (for $t_0 = 1/2$ and the imaginary part, say).

For the upper bound, we will apply some results of probability theory in the Banach space $C_{\mathbf{R}}([0,1])$ of real-valued continuous functions on $[0,1]$, dealing separately with the real and imaginary parts of K.

We view Re(K) as the sum

$$\mathrm{Re(K)} = X_0\varphi_0 + \sum_{n\geqslant 1} X_n\varphi_n,$$

where $(X_n)_{n\geqslant 0}$ are independent, symmetric, real-valued random variables with

$$|X_0| \leqslant 2, \quad |X_n| \leqslant 4 \quad \text{for } n \geqslant 1$$

and

$$\varphi_0(x) = x, \quad \varphi_n(x) = \frac{\sin(2\pi nx)}{2\pi n} \quad \text{for } n \geqslant 1$$

are vectors in $C_{\mathbf{R}}([0,1])$.

By a result of Talagrand [Tal95, Remarks after Theorem 13.2, (13.12)] (or almost equivalently an adaptation of the main theorem of [DMS93] to the variables $(X_n)_{n\geqslant 1}$ instead of Rademacher variables, replacing the crucial theorem of Talagrand [DMS93, Theorem A] used in its proof by [Tal95, Theorem 13.2]), we have

$$\begin{aligned}\mathbf{P}(\|\mathrm{Re(K)}\|_\infty > m + 4N(t)) &= \mathbf{P}\left(\left\|\sum_{n\geqslant 0} X_n\varphi_n\right\|_\infty > m + 4N(t)\right)\\ &\leqslant 4\exp(-t^2/16)\end{aligned}$$

for any $t > 0$, where we have denoted by m a median of the random variable $\|\mathrm{Re(K)}\|_\infty$ and the function $N(t)$ is the function denoted $K_{1,2}^w((x_n), t)$ in [DMS93] (or $\kappa(t)$ in [Tal95, p. 199]) for the sequence $x_n = \varphi_n$ (the factor 4 in front of $N(t)$ is due to the assumption in [Tal95] that the random variables are bounded by 1).

We deduce that

$$\mathbf{P}(\|\mathrm{Re(K)}\|_\infty > A) \leqslant \mathbf{P}(\|\mathrm{Re(K)}\|_\infty > m + 4N(t))$$

for any $t > 0$ such that $m + 4N(t) \leqslant A$.

Now we claim that there exists $c > 0$ such that

$$N(t) \leqslant c\log(ct) \tag{4.1}$$

for all $t \geqslant 1$. This is proved below and, assuming this property, we take

$$t = \frac{1}{c}\exp\left(\frac{A-m}{4c}\right).$$

We may clearly assume that A is large enough so that $t \geqslant 1$ (the desired estimate being trivial otherwise). Then $t \geqslant 1$ also satisfies $m + N(t) \leqslant A$, and we deduce that

$$\mathbf{P}(\|\mathrm{Re(K)}\|_\infty > A) \leqslant 4\exp(-t^2/16),$$

which has the required form for the real part of K. A similar argument applies to the imaginary part, finishing the proof.

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

Now we establish the property (4.1). By definition (see [DMS93, p. 2046]), we have

$$N(t) = \sup_{\|\lambda\| \leqslant 1} N_\lambda(t),$$

where λ runs over elements of the dual space of the real Banach space $C_{\mathbf{R}}([0,1])$ with Banach norm $\|\lambda\| \leqslant 1$, and

$$N_\lambda(t) = \inf\{\|x_1\|_1 + t\|x_2\|_2 \mid (\lambda(\varphi_n)) = x_1 + x_2\},$$

where $x_i \in \ell_i$, and $\|x_i\|_i$ is the ℓ_i-norm in the Banach space ℓ_i of ith power summable sequences of real numbers.

As suggested by a result of Holmstedt (see [Hol70, Theorem 4.1] and [MS90, p. 518]), which gives a two-sided equivalent to $N_\lambda(t)$, we define

$$x_{1,n} = \begin{cases} \lambda(\varphi_n) & \text{if } 0 \leqslant n \leqslant t^2, \\ 0 & \text{if } n > t^2 \end{cases}$$

(which therefore determines x_2). We then have

$$N_\lambda(t) \leqslant \|x_1\|_1 + t\|x_2\|_2.$$

Since the linear form λ has norm $\leqslant 1$, we have

$$|\lambda(\varphi_0)| \leqslant \|\varphi_0\|_\infty = 1$$

and

$$|\lambda(\varphi_n)| = \left|\frac{1}{2\pi n}\lambda(s_n)\right| \leqslant \frac{1}{2\pi n}$$

(where $s_n(x) = \sin(2\pi n x)$) for $n \geqslant 1$. Thus, we get

$$N_\lambda(t) \leqslant 1 + \sum_{1 \leqslant n \leqslant t^2} \frac{1}{2\pi n} + t\left(\sum_{n > t^2} \frac{1}{4\pi^2 n^2}\right)^{1/2} \leqslant c \log ct$$

for some $c > 0$ and all $t \geqslant 1$, as claimed. □

5. Final remarks

5.1 Variants

It is clear that the general setting admits many variations. When looking at other families of one-variable exponential sums over $\mathbf{F}_p$, a number of complications may arise. For instance, in many situations, the analogue of Lemma 2.5 will have to take into account the inter-dependences of the monodromy groups of the analogues of the shifted Kloosterman sheaves, and of course the estimates of short sums necessary for tightness are not always known.

One can also consider families of exponential sums parameterized by multiplicative characters. In that case, we need to exploit Katz's recent definition of an analogue of the monodromy group for Mellin transforms over finite fields [Kat12], in order to have statements similar to Lemma 2.5. But tightness is then sometimes easier to prove, because of the small 'multiplicative energy' of intervals.

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

Yet another variation would involve re-parameterizing the order of summation in Kloosterman (or Birch) sums. The most natural way to do this is to pick a primitive root $g \in \mathbf{F}_p^\times$, and to consider the continuous path interpolating between the partial sums

$$\frac{1}{\sqrt{p}} \sum_{0 \leqslant m \leqslant (p-2)t} \psi_p(ag^m + g^{-m}),$$

which has the same start and end points as $K_p(t)$.

In all of these situations, one can hope to have similar results as those in this paper. We expect to come back to such situations in a later work.

On the other hand, answering the following other very natural question seems well out of reach of current methods.

Problem 5.1. Consider the random variables

$$\tilde{K}_X(t) \,:\, p \mapsto K_p(t,1)$$

on the finite probability space

$$(\{p \text{ prime } \leqslant X\}, \text{uniform measure}).$$

Does $(\tilde{K}_X(t))_{t\in[0,1]}$ also converge in law to $(\mathrm{K}(t))_{t\in[0,1]}$?

This is the analogue for Kloosterman paths of the famous horizontal Sato–Tate conjecture for Kloosterman sums, which remains completely open.

Convergence in finite distributions for this sequence of random variables would follow from the general horizontal equidistribution conjecture of [Saw16], by the same argument as the proof of Theorem 1.1.

5.2 Similar works

We conclude with a brief mention of some related works.

(1) From the probabilistic point of view, it is worth observing that convergence in law of processes (in $C([0,1])$) related to number-theoretic quantities has already been discovered in a few cases. The best-known example is probably Billingsley's generalization of the Erdős–Kac theorem, which gives convergence to Brownian motion of suitable normalized counts of primes dividing integers in varying intervals (see [Bil99, ch. 4, § 17]). One can also mention Bagchi's probabilistic interpretation of Voronin's universality theorem (where convergence in law happens in a space of holomorphic functions; see [Bag81, § 0.2]).

(2) The papers on paths of exponential sums of Lehmer and Loxton [Leh76, Lox83] consider rather different situations, where quite precise asymptotic evaluation is possible, e.g. in terms of Fresnel integrals for incomplete quadratic Gauss sums. In [Lox85], Loxton considers (roughly) sums of $\psi_p(g(x))$, where g ranges uniformly over polynomials of some degree d over $\mathbf{F}_p$, and obtains some limit theorems which are however of a rather different kind to ours.

(3) Another important case is that of the classical character sums

$$S(\chi; N) = \sum_{1 \leqslant n \leqslant N} \chi(n)$$

for χ a non-trivial Dirichlet character modulo $q \geqslant 1$ and $1 \leqslant N \leqslant q$. These have been studied extensively from different perspectives, going back at least as far as Littlewood and Paley. The

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press

papers of Granville and Soundararajan [GS07] and of Bober and Goldmakher [BG13] consider (among other things) the distribution of

$$M(\chi) = \max_{1 \leqslant N \leqslant q} |S(\chi; N)|,$$

the largest modulus of these sums as N varies (compare for instance [BG13, Theorem 1.3] and Theorem 1.6). More recently, Bober, Goldmakher, Granville and Koukoulopoulos [BGGK14] have proved the analogue of the main results of this paper for these sums. The limiting process is very different, however, as it has to take into account the multiplicative structure of the Fourier coefficients.

(4) In another direction, there have been many very precise studies of quadratic Weyl sums; Cellarosi and Marklof [CM15] have recently proved convergence results in $C([0, 1])$, again with a very different limiting process, and we refer to their paper for references to earlier results.

ACKNOWLEDGEMENTS

Thanks to J. Bober, L. Goldmakher, A. Granville and D. Koukoulopoulos for comments concerning their work [BGGK14]. Thanks also to J. Marklof for interesting discussions concerning his work with Cellarosi [CM15] as well as other works concerning limit theorems for quadratic (and other) Weyl sums. We also thank the referees for useful comments.

REFERENCES

Bag81 B. Bagchi, *Statistical behaviour and universality properties of the Riemann zeta function and other allied Dirichlet series*, PhD thesis, Indian Statistical Institute, Kolkata (1981), available at library.isical.ac.in/jspui/handle/10263/4256.

Bil99 P. Billingsley, *Convergence of probability measures*, second edition (Wiley, 1999).

Bir68 B. J. Birch, *How the number of points of an elliptic curve over a fixed prime field varies*, J. Lond. Math. Soc. (2) **43** (1968), 57–60.

BG13 J. W. Bober and L. Goldmakher, *The distribution of the maximum of character sums*, Mathematika **59** (2013), 427–442.

BGGK14 J. W. Bober, L. Goldmakher, A. Granville and D. Koukoulopoulos, *The frequency and the structure of large character sums*, Preprint (2014), arXiv:1410.8189.

BoG14 J. Bourgain and M. Z. Garaev, *Sumsets of reciprocals in prime fields and multilinear Kloosterman sums*, Izv. Math. **78** (2014), 656–707.

CM15 F. Cellarosi and J. Marklof, *Quadratic Weyl sums, automorphic functions and invariance principles*, Preprint (2015), arXiv:1501.07661.

DMF81 F. M. Dekking and M. Mendès France, *Uniform distribution modulo one: a geometrical viewpoint*, J. reine angew. Math. **329** (1981), 143–153.

Del80 P. Deligne, *La conjecture de Weil, II*, Publ. Math. Inst. Hautes Études Sci. **52** (1980), 137–252.

Des85 J.-M. Deshouillers, *Geometric aspects of Weyl sums*, in *Elementary and analytic theory of numbers*, Banach Center Publications, vol. 17 (PWN-Polish Scientific Publisher, Warsaw, 1985), 75–82.

DMS93 S. J. Dilworth and S. J. Montgomery-Smith, *The distribution of vector-valued Rademacher series*, Ann. Probab. **21** (1993), 2046–2052.

FKM15 É. Fouvry, E. Kowalski and Ph. Michel, *A study in sums of products*, Philos. Trans. R. Soc. Lond. A **373** (2015), 20140309.

FM98 É. Fouvry and Ph. Michel, *Sur certaines sommes d'exponentielles sur les nombres premiers*, Ann. Sci. Éc. Norm. Supér. (4) **31** (1998), 93–130.

GS06 A. Granville and K. Soundararajan, *Extreme values of $\zeta(1+it)$*, in *The Riemann zeta function and related themes: papers in honor of Professor K. Ramachandra*, Ramanujan Mathematical Society Lecture Notes Series, vol. 2 (Ramanujan Mathematical Society, India, 2006), 65–80.

GS07 A. Granville and K. Soundararajan, *Large character sums: pretentious characters and the Pólya–Vinogradov theorem*, J. Amer. Math. Soc. **20** (2007), 357–384.

Hol70 T. Holmstedt, *Interpolation of quasi-normed spaces*, Math. Scand. **26** (1970), 177–199.

Irv15 A. J. Irving, *The divisor function in arithmetic progressions to smooth moduli*, Int. Math. Res. Not. IMRN **2015** (2015), 6675–6698, doi:10.1093/imrn/rnu149.

IK04 H. Iwaniec and E. Kowalski, *Analytic number theory*, Colloquium Publications, vol. 53 (American Mathematical Society, Providence, RI, 2004).

Kah85 J.-P. Kahane, *Some random series of functions*, Cambridge Studies in Pure Mathematics, vol. 5 (Cambridge University Press, 1985).

Kat87 N. M. Katz, *On the monodromy attached to certain families of exponential sums*, Duke Math. J. **54** (1987), 41–56.

Kat88 N. M. Katz, *Gauss sums, Kloosterman sums and monodromy groups*, Annals of Mathematical Studies, vol. 116 (Princeton University Press, 1988).

Kat90 N. M. Katz, *Exponential sums and differential equations*, Annals of Mathematical Studies, vol. 124 (Princeton University Press, 1990).

Kat12 N. M. Katz, *Convolution and equidistribution: Sato–Tate theorems for finite field Mellin transforms*, Annals of Mathematical Studies, vol. 180 (Princeton University Press, 2012).

Kow15 E. Kowalski, *The Kloostermania page*, blogs.ethz.ch/kowalski/the-kloostermania-page/.

LT91 M. Ledoux and M. Talagrand, *Probability in Banach spaces: isoperimetry and processes*, Ergeb. Math. Grenzgeb. (3), vol. 23 (Springer, 1991).

Leh76 D. H. Lehmer, *Incomplete Gauss sums*, Mathematika **23** (1976), 125–135.

Liv87 R. Livné, *The average distribution of cubic exponential sums*, J. reine angew. Math. **375–376** (1987), 362–379.

Lox83 J. H. Loxton, *The graphs of exponential sums*, Mathematika **30** (1983), 153–163.

Lox85 J. H. Loxton, *The distribution of exponential sums*, Mathematika **32** (1985), 16–25.

MO88 H. Montgomery and A. Odlyzko, *Large deviations of sums of independent random variables*, Acta Arith. **49** (1988), 427–434.

MS90 S. J. Montgomery-Smith, *The distribution of Rademacher sums*, Proc. Amer. Math. Soc. **109** (1990), 517–522.

RY99 D. Revuz and M. Yor, *Continuous Martingales and Brownian motion*, third edition (Springer, Berlin, 1999).

Saw16 W. F. Sawin, *A Tannakian category of arithmetic exponential sums*, PhD thesis, Princeton University (2016).

Tal95 M. Talagrand, *Concentration of measure and isoperimetric inequalities in product spaces*, Publ. Math. Inst. Hautes Études Sci. **81** (1995), 73–205.

Emmanuel Kowalski kowalski@math.ethz.ch
ETH Zürich – D-MATH, Rämistrasse 101, CH-8092 Zürich, Switzerland

William F. Sawin wsawin@math.princeton.edu
Princeton University, Fine Hall, Washington Road, NJ, USA

https://doi.org/10.1112/S0010437X16007351 Published online by Cambridge University Press