GAUSSIAN DISTRIBUTION FOR THE DIVISOR FUNCTION AND HECKE EIGENVALUES IN ARITHMETIC PROGRESSIONS

ÉTIENNE FOUVRY, SATADAL GANGULY, EMMANUEL KOWALSKI, AND PHILIPPE MICHEL

Abstract. We show that, in a restricted range, the divisor function of integers in residue classes modulo a prime follows a Gaussian distribution, and a similar result for Hecke eigenvalues of classical holomorphic cusp forms. Furthermore, we obtain the joint distribution of these arithmetic functions in two related residue classes. These results follow from asymptotic evaluations of the relevant moments, and depend crucially on results on the independence of monodromy groups related to products of Kloosterman sums.

1. Introduction

The distribution of arithmetic functions in arithmetic progressions is one of the cornerstones of modern analytic number theory, with a particular focus on issues surrounding uniformity with respect to the modulus (see [7] for a recent survey). Besides the case of primes in arithmetic progressions, much interest has been devoted to the divisor function $d(n)$ and higher-divisor functions, in particular because – in some precise sense – a good understanding of a few of these is equivalent to knowledge about the primes themselves (see, e.g., [6, Théorème 4]).

The consideration of the second moment for primes $p \leq X$ in arithmetic progressions to moduli $q \leq Q \leq X/(\log X)^A$ leads to the Barban-Davenport-Halberstam theorem (see, e.g., [13, Th. 17.2]), which has been refined to an asymptotic formula for $Q = X$ by Montgomery [21]. Similarly, Motohashi [22] evaluated asymptotically the variance of the divisor function $d(n)$ for $n \leq X$ in arithmetic progressions modulo $q \leq X$.

We will show that one can determine an asymptotic distribution for the divisor function $d(n)$ for $n \leq X$ in arithmetic progressions modulo a single prime p, provided however that X is a bit smaller than p^2.

Theorem 1.1 (Central Limit Theorem for the divisor function)**.** *Let w be a non-zero real-valued smooth function on $\mathbb{R}$ with compact support in $]0, +\infty[$ and with L^2 norm $\|w\|$. For a prime p, let*

$$S_d(X,p,a) = \sum_{\substack{n\geq 1\\ n\equiv a \bmod p}} d(n) w\Bigl(\frac{n}{X}\Bigr),$$

and

$$\begin{aligned} M_d(X,p) &= \frac{1}{p}\sum_{n\geq 1} d(n) w\Bigl(\frac{n}{X}\Bigr) - \frac{1}{p^2}\int_0^{+\infty} (\log x + 2\gamma - 2\log p) w\Bigl(\frac{x}{X}\Bigr) dx \qquad (1.1)\\ &= \frac{1}{p}\sum_{n\geq 1} d(n) w\Bigl(\frac{n}{X}\Bigr) + O\Bigl(\frac{1}{p^2}X(\log X)\Bigr), \end{aligned}$$

where γ is the Euler constant. For $a \in \mathbb{F}_p^\times$, let

$$E_d(X,p,a) = \frac{S_d(X,p,a) - M_d(X,p)}{(X/p)^{1/2}}.$$

Date: September 9, 2013.

2010 *Mathematics Subject Classification.* 11F11, 11F30, 11T23, 11L05, 60F05.

Key words and phrases. Divisor function, Hecke eigenvalues, Fourier coefficients of modular forms, arithmetic progressions, central limit theorem, Kloosterman sums, monodromy group, Sato-Tate equidistribution.

Ph. M. was partially supported by the SNF (grant 200021-137488) and the ERC (Advanced Research Grant 228304); É. F. thanks ETH Zürich, EPF Lausanne and the Institut Universitaire de France for financial support; S.G. thanks EPF Lausanne for financial support.

Let $\Phi(x) \geq 1$ be any real-valued function, such that

$$\Phi(x) \longrightarrow +\infty \text{ as } x \to +\infty, \qquad \Phi(x) = O_\epsilon(x^\epsilon),$$

for any $\epsilon > 0$ and $x \geq 1$. For any prime p, let $X = p^2/\Phi(p)$. Then as $p \to +\infty$ over prime values, the random variables

$$a \mapsto \frac{E_d(X,p,a)}{\|w\|\sqrt{\pi^{-2}(\log\Phi(p))^3}}$$

on $\mathbb{F}_p^\times$, with the uniform probability on $\mathbb{F}_p^\times$, converge in distribution to a standard Gaussian with mean 0 and variance 1, i.e., for any real numbers $\alpha < \beta$, we have

$$\frac{1}{p-1}\Big|\Big\{a \in \mathbb{F}_p^\times \ \mid\ \alpha \leq \frac{E_d(X,p,a)}{\|w\|\sqrt{\pi^{-2}(\log\Phi(p))^3}} \leq \beta\Big\}\Big| \underset{p\to\infty}{\longrightarrow} \frac{1}{\sqrt{2\pi}}\int_\alpha^\beta e^{-t^2/2}dt.$$

In fact, our results are more general, in three directions: (1) we will consider, in addition to the divisor function, the Fourier coefficients of any classical primitive holomorphic modular form f of level 1 (e.g., the Ramanujan τ function); (2) we will compute the moments of the corresponding random variables and, for a fixed moment, obtain a meaningful asymptotic in a wider range of X and p; (3) we will also consider the joint distribution of

$$a \mapsto (E_d(X,p,a), E_d(X,p,\gamma(a)))$$

when γ is a fixed projective linear transformation (e.g., $\gamma(a) = a+1$, $\gamma(a) = 2a$, $\gamma(a) = -a$, $\gamma(a) = 1/a$, which illustrate various interesting phenomena.) For all these results, the crucial ingredients are the Voronoi summation formula, and the Riemann Hypothesis over finite fields, in the form of results of independence of monodromy groups of sheaves related to Kloosterman sums.

We now introduce the notation to handle these more general problems. As in the statement above, we fix a non-zero smooth function $w : \mathbb{R} \to \mathbb{R}$, with compact support in $[w_0, w_1]$ with $0 < w_0 < w_1 < +\infty$. For any modulus $c \geq 1$, let

$$S_d(X,c,a) = \sum_{\substack{n\geq 1\\ n\equiv a \bmod c}} d(n)w\Big(\frac{n}{X}\Big).$$

This sum has, asymptotically, a natural main term (see, e.g., [18]) which we denote by $M_d(X,c)$, and which coincides with $M_d(X,p)$ when $c = p$ is prime (see (2.8) below). The number of terms in $S_d(X,c,a)$ is $\approx X/c$ and the *square root cancellation philosophy* suggests that its difference with the main term should be of size

$$S_d(X,c,a) - M_d(X,c) \ll (X/c)^{\frac{1}{2}}X^\epsilon, \tag{1.2}$$

as long as X/c gets large. Thus the map

$$\mathrm{Z} :\ a \in (\mathbb{Z}/c\mathbb{Z})^\times \mapsto E_d(X,c,a) = \frac{S_d(X,c,a) - M_d(X,c)}{(X/c)^{1/2}}.$$

is a natural normalized error term that we wish to study as a random variable on $(\mathbb{Z}/c\mathbb{Z})^\times$ equipped with the uniform probability measure (here and below, we sometimes omit the dependency on p and X to lighten the notation Z).

Similarly, consider a primitive (Hecke eigenform) holomorphic cusp form f of even weight k and level 1 (these restrictions are mainly imposed for simplicity of exposition). We write

$$f(z) = \sum_{n\geq 1}\rho_f(n)n^{(k-1)/2}e(nz)$$

its Fourier expansion at infinity, so that $\rho_f(1) = 1$ and $\rho_f(n)$ is the eigenvalue of the Hecke operator $T(n)$ (suitably normalized). We let

$$S_f(X,c,a) = \sum_{n\equiv a (\mathrm{mod}\ c)}\rho_f(n)w\Big(\frac{n}{X}\Big), \quad M_f(X,c) = \frac{1}{c}\sum_{n\geq 1}\rho_f(n)w\Big(\frac{n}{X}\Big),$$

$$E_f(X,c,a) = \frac{S_f(X,c,a) - M_f(X,c)}{(X/c)^{1/2}},$$

for $c \geq 1$ and any integer a. Note that, in this case, the integral representation

$$M_f(X,c) = \frac{1}{c} \times \frac{1}{2\pi i} \int_{2-i\infty}^{2+i\infty} \hat{w}(s)\, X^s\, L(s,f) ds$$

in terms of the Mellin transform $\hat{w}$ of w shows that the main term is very small, namely

$$M_f(X,c) \ll_{f,A} c^{-1} X^{-A} \tag{1.3}$$

for every positive A, uniformly for $c \geq 1$ and $X \geq 1$.

We will study the distribution of

$$a \mapsto E_f(X,p,a), \qquad a \mapsto E_d(X,p,a)$$

for p prime using the method of moments. Thus, for any integer $\kappa \geq 1$, we define

$$\mathcal{M}_\star(X,c\,;\kappa) = \frac{1}{c} \sum_{\substack{a \bmod c \\ (a,c)=1}} E_\star(X,c,a)^\kappa, \qquad \star = d \text{ or } f. \tag{1.4}$$

The first moment is very easy to estimate, and besides Motohashi's work (which considers the average of $\mathcal{M}_d(X,c;2)$ over $c \leq X$), the second moment has recently been discussed by Blomer [2], Lü [19] and Lau–Zhao [18]. In particular, Lau and Zhao obtained an asymptotic formula in the range $X^{1/2} < c < X$ (see (1.10) below; note that the range $c < X^{1/2}$ seems to be much more delicate.)

We will evaluate *any* moment, in a suitable range. Precisely, in §3 we will prove:

Theorem 1.2. *Let the notation be as above, with $\star = d$, the divisor function, or $\star = f$, f a Hecke form of weight k and level 1. Let p be a prime number. Then, for every integer $\kappa \geq 1$, for every positive δ, for every positive ϵ, for every X satisfying*

$$2 \leq X^{1/2} \leq p < X^{1-\delta}, \tag{1.5}$$

we have the equality

$$\mathcal{M}_\star(X,p\,;\kappa) = C_\star(\kappa) + O\Big(p^{-1/2+\epsilon}\Big(\frac{p^2}{X}\Big)^{\kappa/2} + \Big(\frac{X}{p^2}\Big)^{1/2+\epsilon}\Big), \tag{1.6}$$

where the implied constant depends on $(\delta,\epsilon,\kappa,f,w)$, and the constant $C_\star(\kappa)$ is given by

$$C_\star(\kappa) = c_{\star,w}^{\kappa/2}\, m_\kappa, \tag{1.7}$$

with

$$m_\kappa = \begin{cases} 0 & \text{if } \kappa \text{ is odd,} \\ \dfrac{\kappa!}{2^{\kappa/2}(\kappa/2)!} & \text{if } \kappa \text{ is even,} \end{cases} \tag{1.8}$$

and

$$c_{f,w} = \|w\|^2 \|f\|^2 \frac{(4\pi)^k}{\Gamma(k)}, \qquad c_{d,w} = P_w\Big(\log \frac{p^2}{X}\Big), \tag{1.9}$$

for some polynomial $P_w(T) \in \mathbb{R}[T]$, depending only on w, of degree 3 with leading term $\pi^{-2}\|w\|^2 T^3$. Here, for a cusp form f, the L^2-norm of f is computed with respect to the probability measure

$$\frac{3}{\pi} \frac{dxdy}{y^2}$$

on $\mathrm{SL}_2(\mathbb{Z})\backslash\mathbb{H}$, and the L^2-norm of w is computed with respect to the Lebesgue measure on $\mathbb{R}$.

Remark 1.3. In the case $\kappa = 2$, and in the range $X^{1/2} \leq c \leq X$, Lau and Zhao [18, Theorem 1 (2)] have obtained

$$\frac{1}{c} \sum_{a=1}^{c} \Big| \frac{c^{1/2}}{X^{1/2}} \sum_{\substack{n \equiv a (\mathrm{mod}\ c) \\ 1 \leq n \leq X}} \rho_f(n) \Big|^2 = c_f + O\Big(\Big(\frac{c}{X}\Big)^{\frac{1}{6}} d(c) + \Big(\frac{X}{c^2}\Big)^{\frac{1}{4}} \sum_{\ell | c} \frac{\varphi(\ell)}{\ell}\Big), \tag{1.10}$$

for any modulus $c \geq 1$ (not only primes), and a similar result for the divisor function.

We will make further comments on this result after the proof, in Section 3.5. Since m_κ is the κ-th moment of a Gaussian random variable with mean 0 and variance 1, we obtain the following, which implies Theorem 1.1 in the case $\star = d$:

Corollary 1.4 (Central limit theorem)**.** *Let $\Phi(x) \geq 1$ be any real-valued function, such that*

$$\Phi(x) \longrightarrow +\infty \text{ as } x \to +\infty, \qquad \Phi(x) = O_\epsilon(x^\epsilon),$$

for any $\epsilon > 0$, uniformly for $x \geq 1$. For any prime p, let $X = p^2/\Phi(p)$. Then as $p \to +\infty$ over prime values, the random variables

$$a \mapsto \frac{E_\star(X,p,a)}{\sqrt{c_{\star,w}}}$$

on $\mathbb{F}_p^\times$ converge in distribution to a standard Gaussian with mean 0 and variance 1.

As far as we know, this is the first result of this type. We will prove this in Section 3, and give further comments, in Section 3.6.

Remark 1.5. It is natural to ask if a corresponding property holds for Maass forms. This is indeed the case, and indeed the result can be extended to cusp forms on GL_N for all $N \geq 3$, see [17].

Among the other natural generalizations of Corollary 1.4, we consider next the following one: given a map $a \mapsto \gamma(a)$ on $\mathbb{F}_p^\times$, what is the asymptotic joint distribution of

$$a \mapsto (E_\star(X,p,a), E_\star(X,p,\gamma(a))) \text{ ?}$$

We study this when γ is given by a fractional linear transformation. Precisely, let

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{GL}_2(\mathbb{Q}) \cap M_2(\mathbb{Z}) \tag{1.11}$$

be a fixed invertible matrix with integral coefficients. For $p \nmid \det \gamma$, the matrix γ has a canonical reduction modulo p in $\mathrm{PGL}_2(\mathbb{F}_p)$, which we denote by $\pi_p(\gamma)$. In the usual manner, γ (or $\pi_p(\gamma)$) defines a fractional linear transformation on $\mathbb{P}^1_{\mathbb{F}_p}$ by

$$z \in \mathbb{P}^1_{\mathbb{F}_p} \mapsto \gamma \cdot z = \frac{az+b}{cz+d}.$$

By Corollary 1.4, we know that, in the range of validity of this result, both

$$\mathrm{Z}\,:\, a \mapsto \frac{E_\star(X,p,a)}{\sqrt{c_{\star,w}}} \text{ and } \mathrm{Z}\circ\gamma\,:\, a \mapsto \frac{E_\star(X,p,\gamma\cdot a)}{\sqrt{c_{\star,w}}}, \tag{1.12}$$

seen as random variables defined on the set

$$\{a \in \mathbb{F}_p \mid a, \gamma\cdot a \neq 0, \infty\}$$

converge to the normal law. We then wish to know the asymptotic joint distribution of the vector $(\mathrm{Z}, \mathrm{Z}\circ\gamma)$, and we study this issue, as before, using moments.

For κ and λ positive integers, let

$$\mathcal{M}_\star(X,p\,;\kappa,\lambda\,;\gamma) := \frac{1}{p} \sum_{\substack{a\in\mathbb{F}_p \\ a,\ \gamma\cdot a \neq 0,\infty}} E_\star(X,p,a)^\kappa\, E_\star(X,p,\gamma\cdot a)^\lambda, \tag{1.13}$$

be the *mixed moment of order* (κ, λ).

In analogy with Theorem 1.2, we will estimate these moments in §4. To state the result, we note that if γ is diagonal, there is a unique triple of integers $(\alpha_\gamma, \gamma_1, \gamma_2)$, such that we have the *canonical form*

$$\gamma = \alpha_\gamma \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix}, \ \gamma_1 \geq 1 \text{ and } (\gamma_1,\gamma_2) = 1. \tag{1.14}$$

We further introduce the arithmetic functions

$$\boldsymbol{\rho}_{a,f} = \prod_{p^\alpha \| a} \Big(\rho_f(p^\alpha) - \frac{\rho_f(p)\rho_f(p^{\alpha-1})}{p+1}\Big), \qquad \boldsymbol{\rho}_{a,d} = \prod_{p^\alpha \| a} \Big(d(p^\alpha) - \frac{d(p)d(p^{\alpha-1})}{p+1}\Big), \tag{1.15}$$

for $a \geq 1$, and $\boldsymbol{\rho}_{a,f} = 0$ for $a < 0$, $\boldsymbol{\rho}_{a,d} = \boldsymbol{\rho}_{-a,d}$ for $a < 0$. For $\star = f$, we also define the constant

$$c_f = \|f\|^2 (4\pi)^k \Gamma(k)^{-1}. \tag{1.16}$$

Our result is:

Theorem 1.6. *Let γ be defined by* (1.11)*.*

(1) *For every integers κ and λ, for every δ and $\epsilon > 0$, for every prime $p \geq p_0(\gamma)$ and X satisfying* (1.5)*, there exists $C_\star(\kappa, \lambda, \gamma)$ such that*

$$\mathcal{M}_\star(X, p\,; \kappa, \lambda\,; \gamma) = C_\star(\kappa, \lambda, \gamma) + O\Big(p^{-\frac{1}{2}+\epsilon}\Big(\frac{p^2}{X}\Big)^{(\kappa+\lambda)/2} + \Big(\frac{X}{p^2}\Big)^{1/2+\epsilon}\Big). \tag{1.17}$$

(2) *If γ is non-diagonal, then*

$$C_\star(\kappa, \lambda, \gamma) = C_\star(\kappa) C_\star(\lambda). \tag{1.18}$$

(3) *If γ is diagonal, and written in the canonical form* (1.14)*, then*

$$C_\star(\kappa, \lambda, \gamma) = \begin{cases} 0 \text{ if } \kappa + \lambda \text{ is odd,} \\ \displaystyle\sum_{\substack{0 \leq \nu \leq \min(\kappa,\lambda) \\ \nu \equiv \kappa \equiv \lambda \bmod 2}} \nu! \binom{\kappa}{\nu} \binom{\lambda}{\nu} m_{\kappa-\nu} m_{\lambda-\nu} \left(c_{\star,w}\right)^{\frac{\kappa+\lambda}{2}-\nu} \left(\tilde{c}_{\star,w,\gamma}\right)^{\nu}, \text{ otherwise,} \end{cases} \tag{1.19}$$

where

$$\tilde{c}_{f,w,\gamma} = c_f \boldsymbol{\rho}_{\gamma_1\gamma_2,f} \Big(\int_{-\infty}^{\infty} w(\gamma_1 t) w(\gamma_2 t) dt\Big),$$

and for $\star = d$, we have

$$\tilde{c}_{d,w,\gamma} = P_{\gamma_1\gamma_2,w}\Big(\log \frac{p^2}{X}\Big)$$

for some polynomial $P_{\gamma_1\gamma_2,w}(T) \in \mathbb{R}[T]$, of degree ≤ 3 and with coefficient of T^3 given by

$$\frac{1}{\pi^2} \boldsymbol{\rho}_{\gamma_1\gamma_2,d} \Big(\int_{-\infty}^{\infty} w(\gamma_1 t) w(\gamma_2 t) dt\Big) T^3.$$

In (1.17)*, the implied constant depends at most on $(\gamma, \delta, \varepsilon, \kappa, \lambda)$, and in* (1.19)*, we make the convention that $0^\nu = 1$ if $\nu = 0$.*

Of course, if γ is the identity, we recover Theorem 1.2. More generally, we can now determine the joint asymptotic distribution of $(\mathrm{Z}, \mathrm{Z} \circ \gamma)$ in the same range as Corollary 1.4.

Recall that a pair (X, Y) of random variables is a *Gaussian vector* if and only if, for every complex numbers α and β, the random variable $\alpha\mathrm{X} + \beta\mathrm{Y}$ has a Gaussian distribution (see, e.g., [14, pp. 121–124]). If (X, Y) is a Gaussian vector, its covariance matrix $\mathrm{cov}(\mathrm{X}, \mathrm{Y})$ is defined by

$$\mathrm{cov}(\mathrm{X}, \mathrm{Y}) = \begin{pmatrix} \mathbb{E}(\mathrm{X}^2) - \mathbb{E}(\mathrm{X})^2 & \mathbb{E}(\mathrm{XY}) - \mathbb{E}(\mathrm{X})\mathbb{E}(\mathrm{Y}) \\ \mathbb{E}(\mathrm{XY}) - \mathbb{E}(\mathrm{X})\mathbb{E}(\mathrm{Y}) & \mathbb{E}(\mathrm{Y}^2) - \mathbb{E}(\mathrm{Y})^2 \end{pmatrix}, \tag{1.20}$$

where $\mathbb{E}$ denotes the expectation of a random variable. Recall also that a Gaussian vector (X, Y) has independent components if and only if $\mathbb{E}(\mathrm{XY}) = \mathbb{E}(\mathrm{X})\mathbb{E}(\mathrm{Y})$, i.e., if the covariance matrix is diagonal (see [14, Theorem 16.4] for instance).

Corollary 1.7. *Let Φ be a function as in Corollary 1.4, and let $X = p^2/\Phi(p)$. Then, for $\star = f$ or d, as p tends to infinity, the random vector $(\mathrm{Z}, \mathrm{Z} \circ \gamma)$ converges in distribution to a centered Gaussian vector with covariance matrix*

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad \textit{if } \gamma \textit{ is not diagonal.} \tag{1.21}$$

$$\begin{pmatrix} 1 & G_{\star,\gamma,w} \\ G_{\star,\gamma,w} & 1 \end{pmatrix}, \qquad \textit{if } \gamma \textit{ is diagonal,} \tag{1.22}$$

where the covariance $G_{\star,\gamma,w}$ is given by

$$G_{\star,\gamma,w} = \frac{\rho_{\gamma_1\gamma_2,\star}}{\|w\|^2} \int_{\mathbb{R}} w(\gamma_1 t) w(\gamma_2 t) dt.$$

Thus from Corollary 1.7 (noting that $\boldsymbol{\rho}_{a,d} \neq 0$ for any integer $a \neq 0$), we get a criterion for asymptotic independence of $(\mathrm{Z}, \mathrm{Z} \circ \gamma)$:

Corollary 1.8. *We adopt the notations and hypotheses of Corollary 1.7. Then as p tends to ∞, the random variables Z and $\mathrm{Z} \circ \gamma$ tend to independent Gaussian random variables, if and only if one of the following conditions holds:*

(1) *If γ is not a diagonal matrix, i.e., $a \mapsto \gamma \cdot a$ is not a homothety,*

(2) *If γ is a diagonal matrix and*

$$\int_{-\infty}^{\infty} w(\gamma_1 t) w(\gamma_2 t)\, dt = 0,$$

(3) *If $\star = f$, γ is a diagonal matrix in the from* (1.14)*, and there exists a prime p and $\alpha \geq 1$ such that $p^{\alpha} \| \gamma_2 \gamma_1$ and such that*

$$(p+1)\rho_f(p^{\alpha}) = \rho_f(p)\rho_f(p^{\alpha-1}).$$

Remark 1.9. (1) Corollary 1.8 shows for instance that, for $p \to \infty$, the random variables $a \mapsto E_{\star}(p^2/\Phi(p), p, a)$ and $a \mapsto E_{\star}(p^2/\Phi(p), p, \gamma \cdot a)$ converge to independent Gaussian variables, if γ is one of the following functions

$$\gamma \cdot a = a + 1, \qquad \gamma \cdot a = -a, \qquad \gamma \cdot a = 1/a.$$

The case of $\gamma \cdot a = 2a$ is more delicate, since it depends on the value of the integral $\int_0^{+\infty} w(t)w(2t)\, dt$. For instance, this integral is zero when one has the inequalities $w_0 < w_1 < 2w_0 < 2w_1$, where as before $\mathrm{supp}(w) \subset [w_0, w_1]$. The possible dependency here reflects the obvious fact that if $n \equiv a \bmod p$ and $d \mid n$, then $2n \equiv 2a \bmod p$ and $d \mid 2n$.

(2) We do not know if any primitive Hecke form f of level 1 exists for which Condition (3) in this last corollary holds for some p^{α}! Certainly the "easiest" way it could apply would be if, for some p, we had $\rho_f(p) = 0$, but the existence of a primitive cusp form of level 1 and a prime p with $\rho_f(p) = 0$ seems doubtful (e.g., a conjecture of Maeda suggests that the characteristic polynomials of the Hecke operators $T(p)$ in level 1 are irreducible.) On the other hand, if we extend the result to forms of fixed level $N \geq 1$, it is possible to have $\rho_f(p) = 0$ for some p (e.g., for weight $k = 2$ and f corresponding to an elliptic curve.)

1.1. Sketch of the proof.

We will sketch the proof in the case of cusp forms, which is technically a bit simpler, though we present the actual proofs in a unified manner. For Theorem 1.2, the crucial starting point is the Voronoi summation formula, as in [2, 18], which expresses $E_f(X, c, a)$ for any $c \geq 1$ in terms of sums weighted by some smooth function of the Fourier coefficients $\rho_f(n)$ twisted by Kloosterman sums $S(a, n; c)$. One then sees that the main contribution to this sum comes from the n of size roughly $Y = c^2/X$ (see Proposition 2.1).

Considering the κ-th moment, we obtain therefore an average over a mod p of a product of κ Kloosterman sums $S(a, n_i; p)$, where all variables n_i are of size approximately p^2/X. The sum over $a \in \mathbb{F}_p^{\times}$, when the variables n_i are fixed, can be evaluated using deep results on the independence of Kloosterman sheaves (see Proposition 3.2). This allows us to gain a factor $p^{1/2}$ compared with a direct application of the Weil bound for Kloosterman sums, except for special, well-understood, configurations of the n_i modulo p. These configurations lead, by combinatorial arguments, to the Gaussian main term of Theorem 1.2. (Note that we can take no advantage of the summation over the variables n_i, which turn out to have a short range in the cases where our result is non-trivial, see Section 3.5.)

The study of mixed moments (see Theorem 1.6) has a lot of similarities. The only significant difference lies in the study of the independence of Klosterman sheaves, when some of them are twisted by the rational transformation γ. However, Proposition 3.2 is general enough to show that these sheaves are dependent if and only if we are in the "obvious" cases. The main terms then require some computations of integrals using properties of the Bessel transforms.

1.2. Possible extensions.

A Gaussian law similarly appears if one studies the random variable $a \mapsto E_{\star}(X, p, P(a))$, where P is a non–constant fixed polynomial with integer coefficients. The fact that P is not necessarily a bijection on $\mathbb{F}_p$ does not affect the Gaussian behavior. The proof of this extension requires a suitable generalization of Proposition 3.2.

It also seems that the present method can be extended to the study of the distribution of sums of the shape

$$a \mapsto S_\star(X,p,K_a) = \sum_{n\geq 1} \tau_\star(n) K_a(n) w\Big(\frac{n}{X}\Big)$$

where $\tau_\star(\cdot)$ is either $d(\cdot)$ or $\rho_f(\cdot)$, and $K_a(n) = K(an)$ for a fairly general trace function K as in [8]. The shape of the analogue of Theorem 1.2 would then depend on the nature of the geometric monodromy group of a suitable “Bessel transform” of the sheaves underlying $K(\cdot)$.

Another natural extension, which we are currently considering, is that of coefficients of cusp forms on higher-rank groups, and of higher divisor functions.

1.3. Notations.

We use synonymously the notation $f(x) \ll g(x)$ for $x \in X$ and $f = O(g)$ for $x \in X$. We denote $e(z) = e^{2i\pi z}$ for $z \in \mathbb{C}$. For $c \geq 1$ and a, b integers, or congruence classes modulo c, the Kloosterman sum $S(a,b;c)$ is defined by

$$S(a,b;c) = \sum_{\substack{x \bmod c\\ (x,c)=1}} e\Big(\frac{ax+b\bar{x}}{c}\Big)$$

where $\bar{x}$ is the inverse of x modulo c. The normalized Kloosterman sum is defined by

$$\mathrm{Kl}_2(a,b;c) = \frac{S(a,b;c)}{c^{1/2}},$$

and for $(a,b,c) = 1$ it satisfies the Weil bound

$$|\mathrm{Kl}_2(a,b;c)| \leq d(c). \tag{1.23}$$

To lighten notations, we define

$$\mathrm{Kl}_2(a;c) := \mathrm{Kl}_2(a,1;c),$$

and recall the equality $\mathrm{Kl}_2(a,b;c) = \mathrm{Kl}_2(ab;c)$, whenever $(b,c) = 1$.

We will use the Bessel functions J_{k-1}, where $k \geq 2$ is an integer, Y_0 and K_0; precise definitions can be found for instance in [11, App. B.4] and in [25].

Acknowledgements. We thank G. Ricotta for pointing out a computational error in the first draft of this paper.

2. Preliminaries

We gather in this section some facts we will need during the proof of the main results. The reader may wish to skip to Section 3 and refer to the results when they are needed.

We begin with the Voronoi formula in the form we need:

Proposition 2.1 (Voronoi summation). *Let $\star = f$, for a cusp form f of level 1 and weight k, or $\star = d$. Let c be any positive integer, with c prime if $\star = d$. Then for any $X \geq 1$ and for any integer a, we have the equality*

$$E_\star(X,c,a) = \frac{X^{1/2}}{c} \sum_{\substack{c_1 | c\\ c_1 > 1}} \Big(\frac{c}{c_1}\Big)^{1/2} \sum_{n \neq 0} \tau_\star(n) W_\star\Big(\frac{nX}{c_1^2}\Big) \mathrm{Kl}_2(a,n;c_1), \tag{2.1}$$

where n runs on the right over non-zero integers in $\mathbb{Z}$ and

$$\tau_f(n) = \begin{cases} \rho_f(n) & \text{if } n \geq 1,\\ 0 & \text{otherwise,}\end{cases} \tag{2.2}$$

$$\tau_d(n) = d(|n|), \tag{2.3}$$

and

$$W_f(y) = 2\pi i^k \int_0^\infty w(u) J_{k-1}(4\pi\sqrt{uy})du \quad \textit{for } y > 0, \tag{2.4}$$

$$W_f(y) = 0, \quad \textit{for } y < 0,$$

$$W_d(y) = -2\pi \int_0^\infty w(u) Y_0(4\pi\sqrt{uy})du, \quad \textit{for } y > 0, \tag{2.5}$$

$$W_d(y) = 4 \int_0^\infty w(u) K_0(4\pi\sqrt{u|y|})du, \quad \textit{for } y < 0. \tag{2.6}$$

In particular, if $c = p$, a prime, we have

$$E_\star(X, p, a) = \Big(\frac{X}{p^2}\Big)^{1/2} \sum_{n\neq 0} \tau_\star(n) W_\star\Big(\frac{nX}{p^2}\Big) \mathrm{Kl}_2(a, n; p). \tag{2.7}$$

For the proof we recall the standard Voronoi summation formula (see, e.g., [13, p. 83] for $\star = f$ and [13, (4.49)] for $\star = d$, which we rewrite as a single sum over positive and negative integers instead of two sums).

Lemma 2.2. *Let c be a positive integer and a an integer coprime to c.*

(1) *For any smooth function w compactly supported on $]0, \infty[$, we have*

$$\sum_{n\geq 1} \rho_f(n) w(n) e\Big(\frac{an}{c}\Big) = \frac{1}{c} \sum_{n\geq 1} \rho_f(n) W_f\Big(\frac{n}{c^2}\Big) e\Big(-\frac{n\bar{a}}{c}\Big),$$

if f is a cusp form of level 1 and weight k.

(2) *For any smooth function w compactly supported on $]0, \infty[$, we have*

$$\sum_{n\geq 1} d(n) w(n) e\Big(\frac{an}{c}\Big) = \frac{1}{c} \int_0^{+\infty} (\log x + 2\gamma - 2\log c) w(x) dx + \frac{1}{c} \sum_{n\neq 0} d(|n|) W_d\Big(\frac{n}{c^2}\Big) e\Big(-\frac{\bar{a}n}{c}\Big). \tag{2.8}$$

Proof of Proposition 2.1. We consider the case of $\star = f$, the divisor function being handled similarly (it is easier since c is prime; the definition (1.1) of the main term is designed to cancel out the first main term in (2.8)). Using orthogonality of additive characters, and separating the contribution of the trivial character from the others, we write

$$\begin{aligned} S_f(X, c, a) &= \frac{1}{c} \sum_{b=0}^{c-1} e\Big(-\frac{ab}{c}\Big) \sum_{n\geq 1} \rho_f(n) w\Big(\frac{n}{X}\Big) e\Big(\frac{bn}{c}\Big) \\ &= \frac{1}{c} \sum_{n\geq 1} \rho_f(n) w\Big(\frac{n}{X}\Big) + \frac{1}{c} \sum_{1\leq b\leq c-1} e\Big(-\frac{ab}{c}\Big) \sum_{n\geq 1} \rho_f(n) w\Big(\frac{n}{X}\Big) e\Big(\frac{bn}{c}\Big), \end{aligned}$$

which yields the expression

$$E_f(X, c, a) = \frac{1}{(cX)^{1/2}} \sum_{1\leq b\leq c-1} e\Big(-\frac{ab}{c}\Big) \sum_{n\geq 1} \rho_f(n) w\Big(\frac{n}{X}\Big) e\Big(\frac{bn}{c}\Big).$$

We split the second according to the value of the g.c.d $d = (b, c)$, writing

$$d = (b, c), b = db_1, c = dc_1,$$

and note that

$$1 < c_1 \leq c, \quad 1 \leq b_1 < c_1.$$

We then get

$$\begin{aligned} E_f(X, c, a) &= \frac{1}{(cX)^{1/2}} \sum_{d|c} \sum_{\substack{1\leq b<c \\ (b,c)=d}} e\Big(-\frac{ab}{c}\Big) \sum_{n\geq 1} \rho_f(n) w\Big(\frac{n}{X}\Big) e\Big(\frac{bn}{c}\Big) \\ &= \frac{1}{(cX)^{1/2}} \sum_{\substack{c_1|c \\ c_1>1}} \sum_{\substack{1\leq b_1<c_1 \\ (b_1,c_1)=1}} e\Big(-\frac{ab_1}{c_1}\Big) \sum_{n\geq 1} \rho_f(n) w\Big(\frac{n}{X}\Big) e\Big(\frac{b_1 n}{c_1}\Big). \end{aligned}$$

We can now apply Lemma 2.2 since $(b_1, c_1) = 1$, and we get

$$\sum_{n\geq 1} \rho_f(n) w\Big(\frac{n}{X}\Big) e\Big(\frac{b_1 n}{c_1}\Big) = \frac{X}{c_1} \sum_{n\geq 1} \rho_f(n) W\Big(\frac{nX}{c_1^2}\Big) e\Big(-\frac{\bar{b}_1 n}{c_1}\Big).$$

The proposition now follows since the terms with $n < 0$ are identically zero for this case. □

We will need some basic information on the behavior of the Bessel transforms $W_\star(y)$.

Proposition 2.3. *Let w be a smooth function with support included in $]0, +\infty[$. Let $W_\star(y)$ be one of the Bessel transforms of w as defined in Proposition 2.1, for some integer $k \geq 2$ in the case $\star = f$ of weight k.*

(1) *The function $W_\star$ is smooth on $\mathbb{R}^\times$, and for every $A \geq 0$ and $j \geq 0$, we have*

$$y^j W_\star^{(j)}(y) \ll_{A,j} \min\big(1 + \big|\log|y|\big|, |y|^{-A}\big), \tag{2.9}$$

for $y \neq 0$.

(2) *We have*

$$\|W_\star\| = \|w\|, \tag{2.10}$$

where the L^2-norm of $W_\star$ and w are computed in $L^2(\mathbb{R}^\times)$ with respect to Lebesgue measure.

(3) *More generally, for any two non-zero real numbers m and n, we have*

$$\int_{-\infty}^{\infty} W_\star(mt) W_\star(nt) dt = \int_{-\infty}^{\infty} w(mt) w(nt) dt.$$

Proof. (1) (Compare, e.g., with [2, p. 280], [18, Lemma 3.1]) We begin with the case $j = 0$. For y small, we use the bounds

$$J_{k-1}(x) \ll_k 1, \quad Y_0(x) \ll 1 + |\log x|, \quad K_0(x) \ll 1 + |\log x|$$

for $0 < x \leq 1$ which immediately imply

$$W_\star(y) \ll 1 + \big|\log|y|\big| \tag{2.11}$$

in all cases.

To deal with the case where $|y| \geq 1$, we first make the change of variable

$$v = 4\pi\sqrt{u|y|}$$

in the integrals (2.4) (resp. (2.5), (2.6)), so that we always get

$$W_\star(y) = \frac{1}{|y|} \int_0^\infty w\Big(\frac{v^2}{16\pi^2 y^2}\Big) v B_0(v) dv,$$

where $B_0 = cJ_{k-1}$, 0, cY_0 or cK_0, for some fixed multiplicative constant $c \in \mathbb{C}$.

We denote $\alpha = (16\pi^2 y^2)^{-1}$. To exploit conveniently the oscillations of the Bessel functions B_0 we integrate by parts, using the relations (see [10, 8.472.3, 8.486.14])

$$(x^{\nu+1} Z_{\nu+1}(x))' = \epsilon x^{\nu+1} Z_\nu(x), \tag{2.12}$$

where

$$\epsilon = \begin{cases} +1 & \text{if } Z_\nu = J_\nu \text{ or } Y_\nu, \\ \\ -1 & \text{if } Z_\nu = K_\nu. \end{cases}$$

For $\star = f$, remembering that w vanishes at 0 and ∞, we obtain, for instance, the equality

$$W_f(y) = -\frac{c}{|y|} \int_0^\infty \Big(2\alpha v^2 w'(\alpha v^2) + (1-k) w(\alpha v^2)\Big) J_k(v) dv \quad (y > 0).$$

By iterating $\ell \geq 1$ times, and then arguing similarly for $\star = d$, we see that there exist coefficients $\xi_{\ell,\nu}$ such that

$$W_\star(y) = \frac{1}{|y|} \int_0^\infty \Big(\sum_{\nu=0}^{\ell} \xi_{\ell,\nu}\ (\alpha v^2)^\nu w^{(\nu)}(\alpha v^2)\Big) v^{-\ell+1} B_\ell(v) dv, \tag{2.13}$$

where $B_\ell = J_{k-1+\ell}$, 0, Y_ℓ or K_ℓ corresponding to the different cases $\star = f$ or $\star = d$, $y > 0$ or $y < 0$.

Since w has compact support in $[w_0, w_1]$, the above integral can be restricted to the interval

$$\mathcal{I} := \big[(w_0/\alpha)^{1/2}, (w_1/\alpha)^{1/2}\big],$$

and using the estimates[1]

$$J_{k-1+\ell}(v) \ll_\ell v^{-1/2}, \qquad Y_\ell(v) \ll_\ell v^{-1/2}, \qquad K_\ell(v) \ll_\ell v^{-1/2}$$

for $v \geq 1$, we obtain the inequality

$$W_\star(y) \ll |y|^{-1} \int_{\mathcal{I}} v^{-\ell+\frac{1}{2}} dv \ll |y|^{-1-\ell/2+3/2} \tag{2.14}$$

for $|y| \geq 1$. Since $\ell \geq 0$ is arbitrary, this gives the result for $j = 0$.

We can reduce the general case to $j = 0$ using the formulas (see [10, 8.472.2, 8.486.13])

$$xZ'_\nu(x) = \nu Z_\nu(x) - xZ_{\nu+1}(x),$$

from which it follows that

$$y\frac{d}{dy}\Big(\int_0^\infty w(u)Z_\nu(4\pi\sqrt{uy})du\Big) = \frac{\nu}{2}\int_0^\infty w(u)Z_\nu(4\pi\sqrt{uy})du - 2\pi\sqrt{y}\int_0^\infty w(u)\sqrt{u}Z_{\nu+1}(4\pi\sqrt{uy})du.$$

Applying the previous method to the relevant Bessel functions then leads to

$$yW'_\star(y) \ll_{\star,A} \min(1 + |\log|y||, y^{-A})$$

and by induction a similar argument deals with higher derivatives.

(2) In the case $\star = f$, the identity

$$\int_0^{+\infty} W_f(u)^2 du = \int_0^{+\infty} w(u)^2 du = \|w\|^2$$

is a direct consequence of the unitarity of the Hankel transform, i.e., of the Fourier transform for radial functions (see, e.g., [18, Lemma 3.4]). The case $\star = d$ is less classical, although it is formally similar, the hyperbolas $xy = r$ replacing the circles $x^2 + y^2 = r^2$ (see [13, §4.5]). We use a representation-theoretic argument to get a quick proof. The unitary principal series representation $\rho = \pi(0)$ of $\mathrm{PGL}_2(\mathbb{R})$ (in the notation of [4, p. 10]) can be defined by its Kirillov model with respect to the additive character $\psi(x) = e(x)$, which is a unitary representation of $\mathrm{PGL}_2(\mathbb{R})$ on $L^2(\mathbb{R}^\times, |x|^{-1}dx)$. In this model, the unitary operator

$$T = \rho\Big(\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\Big)$$

on $L^2(\mathbb{R}^\times, |x|^{-1}dx)$ is given by

$$T\varphi(x) = \int_{\mathbb{R}^\times} \varphi(t)\mathcal{J}(xt)\frac{dt}{|t|},$$

where $\mathcal{J}$ is the so-called "Bessel function" of ρ (with respect to ψ, see [4, Th. 4.1]). By [4, Prop. 6.1, (ii)] (see also [1, §6, §21]), we have

$$\mathcal{J}(u) = \begin{cases} -2\pi\sqrt{u}Y_0(4\pi\sqrt{u}) & \text{for } u > 0, \\ 4\sqrt{|u|}K_0(4\pi\sqrt{|u|}) & \text{for } u < 0. \end{cases}$$

Hence by (2.5) and (2.6), we see that

$$W_d(y) = |y|^{-1/2}T(\varphi)(y), \quad \text{where} \quad \varphi(x) = \begin{cases} \sqrt{x}w(x) & \text{if } x > 0 \\ 0 & \text{if } x < 0. \end{cases} \tag{2.15}$$

The unitarity of T means that

$$\int_{\mathbb{R}^\times} |T(\varphi)(y)|^2 \frac{dy}{|y|} = \int_{\mathbb{R}^\times} |\varphi(x)|^2 \frac{dx}{|x|},$$

i.e.,

$$\int_{\mathbb{R}^\times} |W_d(y)|^2 dy = \int_0^{+\infty} |w(x)|^2 dx = \|w\|^2.$$

[1] For the last one, one knows in fact that $K_\ell(v)$ decays exponentially fast for $v \to +\infty$.

(3) We consider different cases. If $mn > 0$, changing t to $-t$ allows us to assume that m and n are positive. Then a simple polarization argument from (2.10) shows that

$$\int_{-\infty}^{+\infty} W_\star(mt)W_\star(nt)dt = \int_{-\infty}^{\infty} \mathfrak{w}_m(u)\mathfrak{w}_n(u)du, \tag{2.16}$$

where $u \mapsto \mathfrak{w}_m(u)$ is the function for which the Bessel transform of is $t \mapsto W_\star(mt)$ and similarly for $\mathfrak{w}_n(u)$. But it is immediate that $\mathfrak{w}_m(u) = (1/m)\, w(u/m)$, and therefore (2.16) gives the result.

If $mn < 0$, then since the support of w is contained in $[0, +\infty[$, we have $w(mt)w(nt) = 0$ for all t, hence

$$\int_{\mathbb{R}} w(mt)w(nt)dt = 0,$$

and we must show that the integral of $W_\star(mt)W_\star(nt)$ is also zero. If $\star = f$, a cusp form, this is immediate since $W_f(y) = 0$ for $y < 0$, so that $W_f(mt)W_f(nt) = 0$ for all t.

For $\star = d$, we use representation theory as in (2). With the same notation as used there, and for any real-number $a \neq 0$, we denote

$$U_a = \rho\Bigl(\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}\Bigr)$$

so that, by definition of the Kirillov model (see [4, §4.2, (4.1)]), we have

$$U_a(\varphi)(x) = \varphi(ax)$$

for $\varphi \in L^2(\mathbb{R}^\times, |x|^{-1}dx)$. Observe that, in $\mathrm{PGL}_2(\mathbb{R})$, we have

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -a \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a^{-1} & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

hence

$$T \circ U_a = U_{a^{-1}} \circ T.$$

Using this and the unitarity of T, we deduce that

$$\begin{aligned}\int_{\mathbb{R}} (T\varphi)(ax)\overline{(T\varphi)(bx)}\frac{dx}{|x|} &= \langle U_a(T\varphi), U_b(T\varphi)\rangle = \langle T(U_{a^{-1}}\varphi), T(U_{b^{-1}}\varphi)\rangle \\ &= \langle U_{a^{-1}}\varphi, U_{b^{-1}}\varphi\rangle \\ &= \int_{\mathbb{R}} \varphi\Bigl(\frac{x}{a}\Bigr)\overline{\varphi\Bigl(\frac{x}{b}\Bigr)}\frac{dx}{|x|} = \int_{\mathbb{R}} \varphi(bx)\overline{\varphi(ax)}\frac{dx}{|x|}.\end{aligned}$$

Now, applying (2.15) and the fact that W_d is real-valued, we derive

$$\int_{\mathbb{R}} W_d(ax)W_d(bx)dx = \int_{\mathbb{R}} w(ax)w(bx)dx$$

for all non-zero a and b. □

Remark 2.4. One can also give a direct proof of the last part of this proposition using known properties of Bessel functions: the crucial point is that the function

$$\psi(a, b) = \int_0^\infty Y_0(a\sqrt{y})K_0(b\sqrt{y})\, dy$$

is antisymmetric, which follows from an explicit evaluation using [10, 6.523] and [23, p. 153, 2.34]. Conversely, the results for cusp forms can be proved using representation theory, the discrete series representation of weight k replacing the representation ρ.

Our last preliminary results concern the sums which will give rise to the leading terms in the main results. Recall the definitions and (1.16).

Proposition 2.5. *Let p be a prime number, $\delta > 0$ a parameter and $X \geq 1$ such that*

$$X^{1/2} \leq p \leq X^{1-\delta}.$$

Let $Y = p^2/X$. For $\star \in \{d, f\}$, and for a and b coprime non-zero integers, not necessarily positive, let

$$\mathcal{B}_\star(a,b,Y) = \sum_{\substack{n\neq 0\\ 1\leq |an|,\, |bn|<p/2}} \tau_\star(an)\tau_\star(bn)W_\star\Bigl(\frac{an}{Y}\Bigr)W_\star\Bigl(\frac{bn}{Y}\Bigr).$$

(1) *If $\star = f$, we have*

$$\mathcal{B}_\star(a,b,Y) = c_f\boldsymbol{\rho}_{ab,f}\Bigl(\int_{-\infty}^{\infty} w(at)w(bt)dt\Bigr)Y + O(Y^{1/2+\epsilon})$$

for any $\epsilon > 0$.

(2) *If $\star = d$, there exists a polynomial $P_{ab} \in \mathbb{R}[T]$ of degree at most 3, depending on w, such that*

$$\mathcal{B}_d(a,b,Y) = P_{ab}(\log Y)Y + O(Y^{\frac{1}{2}+\epsilon})$$

for any $\epsilon > 0$, and with coefficient of T^3 given by

$$\frac{1}{\pi^2}\boldsymbol{\rho}_{ab,d}\Bigl(\int_{-\infty}^{\infty} w(at)w(bt)\,dt\Bigr)T^3. \tag{2.17}$$

In both cases, the implied constants depend on $(\delta, \epsilon, \star, a, b)$.

We will use standard complex integration techniques, and first determine the relevant generating series (it is here that it is important that f be a Hecke eigenform.) We denote

$$F_\star(s) = \sum_{n\geq 1} \tau_\star(n)^2 n^{-s},$$

so that

$$F_f(s) = \frac{L(s, f\times f)}{\zeta(2s)}$$

if f is a Hecke eigenform, where $L(s, f\times f)$ is the Rankin-Selberg convolution L-function, and

$$F_d(s) = \frac{\zeta(s)^4}{\zeta(2s)}.$$

In both cases, $F_\star(s)$ extends to a meromorphic function, with polynomial growth in vertical strips, for $\mathrm{Re}(s) > 1/2$. It has only a pole at $s = 1$ in this region (of order 1 if $\star = f$, and order 4 if $\star = d$).

Lemma 2.6. *Let $\star = f$ or d, and a, b be non-zero coprime integers, not necessarily positive. Let*

$$F_{\star,a,b}(s) = \sum_{n\geq 1} \tau_\star(an)\tau_\star(bn)n^{-s}.$$

If $\star = f$ and $ab < 0$, we have $F_{\star,a,b} = 0$. Otherwise, we have

$$F_{\star,a,b}(s) = F_\star(s)\prod_{p^{\nu_p}||ab}\Bigl(\tau_\star(p^{\nu_p}) - \frac{\tau_\star(p)\tau_\star(p^{\nu_p-1})}{p^s+1}\Bigr).$$

In particular, $F_{\star,a,b}$ always extends to a meromorphic function for $\mathrm{Re}(s) > 1/2$, *with polynomial growth in vertical strips.*

Proof. One sees immediately that it is enough to treat the case where a, $b \geq 1$ and $ab \neq 1$. Then the assumption that $(a,b) = 1$ allows us to write

$$F_{\star,a,b}(s) = F_{\star,ab,1}(s)$$

so that we can further reduce to the case where $b = 1$, in which case we write $F_{\star,a,1} = F_{\star,a}$. Now, writing any integer $n \geq 1$ (uniquely) as $n = jm$ where $j \geq 1$ has all prime factors dividing a and $m \geq 1$ is coprime with a, and summing over j first, we get

$$\begin{aligned} F_{\star,a}(s) &= \sum_{1\leq j|a^\infty} \sum_{(m,a)=1} \tau_\star(jm)\tau_\star(ajm)(jm)^{-s} \\ &= \sum_{j|a^\infty} \tau_\star(j)\tau_\star(aj)j^{-s} \sum_{(m,a)=1} \tau_\star(m)^2 m^{-s} \\ &= F_\star(s)\Bigl(\prod_{p|a}\sum_{k\geq 0}\tau_\star(p^k)^2 p^{-ks}\Bigr)^{-1} \sum_{j|a^\infty} \tau_\star(j)\tau_\star(aj)j^{-s}, \end{aligned}$$

by multiplicativity of $\tau_\star$.

Now write

$$a = \prod_{p|a} p^{\nu_p}$$

the factorization of a. Again by multiplicativity, we get

$$\sum_{j|a^\infty} \tau_\star(j)\tau_\star(aj)j^{-s} = \prod_{p|a}\sum_{k\geq 0} \tau_\star(p^k)\tau_\star(p^{k+\nu_p})p^{-ks}.$$

Let

$$G_i = \sum_{k\geq 0} \tau_\star(p^k)\tau_\star(p^{k+i})p^{-ks}$$

for some fixed prime p and integer $i \geq 0$. For $i \geq 1$ and $k \geq 1$, we have

$$\tau_\star(p^{k+i}) = \tau_\star(p^k)\tau_\star(p^i) - \tau_\star(p^{k-1})\tau_\star(p^{i-1}),$$

and therefore

$$G_i = \tau_\star(p^i)G_0 - p^{-s}\tau_\star(p^{i-1})G_1$$

for $i \geq 1$. In particular, the case $i = 1$ gives

$$(1 + p^{-s})G_1 = \tau_\star(p)G_0,$$

which then implies that

$$G_i = \Bigl(\tau_\star(p^i) - \frac{\tau_\star(p)\tau_\star(p^{i-1})}{p^s + 1}\Bigr)G_0$$

for $i \geq 1$. Now, since $\nu_p \geq 1$ by definition, it follows that

$$F_{\star,a}(s) = F_\star(s)\prod_{p|a}\Bigl(\tau_\star(p^{\nu_p}) - \frac{\tau_\star(p)\tau_\star(p^{\nu_p - 1})}{p^s + 1}\Bigr)$$

as claimed. □

Proof of Proposition 2.5. Using Proposition 2.3, (1), we obtain first

$$\mathcal{B}_\star(a, b, Y) = \mathcal{B}^0_\star(a, b, Y) + \mathcal{B}^0_\star(-a, -b, Y) + O(p^{-1})$$

where, for any coprime integers a and b, we put

$$\mathcal{B}^0_\star(a, b, Y) = \sum_{n\geq 1} \tau_\star(an)\tau_\star(bn)W_\star\Bigl(\frac{an}{Y}\Bigr)W_\star\Bigl(\frac{bn}{Y}\Bigr).$$

We now estimate these sums. Let

$$\varphi_{a,b}(s) = \int_0^\infty W_\star(ax)W_\star(bx)x^{s-1}dx,$$

be the Mellin transform of the function $x \mapsto W_\star(ax)W_\star(bx)$.

For $\mathrm{Re}(s) > 0$, this is, by Proposition 2.3, (1), a holomorphic function which is bounded and which decays quickly in vertical strips. We have the integral representation

$$\mathcal{B}^0_\star(a, b, Y) = \frac{1}{2i\pi}\int_{(2)} F_{\star,a,b}(s)Y^s\varphi_{a,b}(s)ds,$$

and we proceed to shift the contour to $\mathrm{Re}(s) = 1/2 + \epsilon$, for a fixed $\epsilon > 0$. The integral on the line $\mathrm{Re}(s) = 1/2 + \epsilon$ satisfies

$$\frac{1}{2i\pi}\int_{(1/2+\epsilon)} F_{\star,a,b}(s)Y^s\varphi_{a,b}(s)ds \ll Y^{1/2+\epsilon}$$

where the implied constant depends on $(\star, a, b, \epsilon, w)$. On the other hand, the unique singularity that occurs during the shift of contour is the pole at $s = 1$ so that

$$\mathcal{B}^0_\star(a, b, Y) = \mathrm{res}_{s=1}\, F_{\star,a,b}(s)Y^s\varphi_{a,b}(s) + O(Y^{1/2+\epsilon})$$

and hence

$$\mathcal{B}_\star(a, b, Y) = \mathrm{res}_{s=1}\, F_{\star,a,b}(s)Y^s\varphi_{a,b}(s) + \mathrm{res}_{s=1}\, F_{\star,-a,-b}(s)Y^s\varphi_{-a,-b}(s) + O(Y^{1/2+\epsilon}).$$

If $\star = f$, then the two residues vanish if $ab < 0$, while if $ab \geq 1$, one residue is zero and the other is equal to

$$\mathrm{res}_{s=1}\, F_{\star,|a|,|b|}(s)Y^s\varphi_{|a|,|b|}(s) = Y\varphi_{|a|,|b|}(1)\,\mathrm{res}_{s=1}\, F_{f,|a|,|b|}(s).$$

Since

$$\varphi_{a,b}(1) + \varphi_{-a,-b}(1) = \int_{\mathbb{R}} W_f(at)W_f(bt)dt = \int_{\mathbb{R}} w(at)w(bt)dt$$

by Proposition 2.3, (3), and since it is well-known that

$$\mathrm{res}_{s=1}\, F_f(s) = \|f\|^2(4\pi)^k\Gamma(k)^{-1} = c_f,$$

(from Rankin-Selberg theory, see, e.g., [12, (13.52), (13.53)]), we see that Lemma 2.6 gives the result in the case of a cusp form.

On the other hand, if $\star = d$, then by Lemma 2.6 both $F_{d,a,b}$ and $F_{f,-a,-b}$ have a pole of order 4, and they satisfy

$$\mathrm{res}_{s=1}\, F_{d,a,b}(s)Y^s\varphi_{a,b}(s) = YQ_{a,b}(\log Y)$$

where the polynomial $Q_{a,b}$ has degree at most 3 and has coefficient of T^3 given by

$$\frac{1}{6}\frac{1}{\zeta(2)}\boldsymbol{\rho}_{ab,d}\Big(\int_0^{+\infty} W_d(at)W_d(bt)dt\Big)T^3.$$

Hence the sum of both terms has the desired form with $P_{ab} = P_{a,b} = Q_{a,b} + Q_{-a,-b}$, and since

$$\int_0^{+\infty} W_d(at)W_d(bt)dt + \int_0^{+\infty} W_d(-at)W_d(-bt)dt = \int_{\mathbb{R}} w(at)w(bt)dt,$$

again by Proposition 2.3, (3), this concludes the proof. □

3. Proof of Theorem 1.2

3.1. **First step.** Let p be a prime such that the condition (1.5) holds. To shorten the notation, we write

$$Y = p^2/X, \tag{3.1}$$

which is ≥ 1 under our assumption. We also write simply $W = W_\star$ depending on whether we treat the case of cusp forms or of the divisor function.

From (2.7) in Proposition 2.1, we deduce

$$\mathcal{M}_\star(X, p; \kappa) = \frac{1}{pY^{\kappa/2}}\sum_{n_1,\ldots,n_\kappa\neq 0}\cdots\sum \tau_\star(n_1)\cdots\tau_\star(n_\kappa)W\Big(\frac{n_1}{Y}\Big)\cdots W\Big(\frac{n_\kappa}{Y}\Big)$$
$$\times\sum_{1\leq a<p}\mathrm{Kl}_2(an_1; p)\cdots\mathrm{Kl}_2(an_\kappa; p), \tag{3.2}$$

which we write in the form

$$\mathcal{M}_\star(X, p; \kappa) := \frac{1}{pY^{\kappa/2}}\big(\Sigma_1 + \Sigma_2\big) \tag{3.3}$$

where Σ_1 corresponds to the contribution of the $(n_1, \ldots, n_\kappa)$ such that $1 \leq |n_i| < p/2$ for all i and Σ_2 is the complementary contribution of those $(n_1, \ldots, n_\kappa)$ such that $|n_i| \geq p/2$ for one i at least.

3.2. **Study of Σ_2.** We first deal with Σ_2, which is easy. By symmetry, we may restrict to the case where $|n_1| \geq p/2$. By Deligne's bound

$$|\rho_f(n)| \leq d(n) \tag{3.4}$$

(in the case of a Hecke eigenform f) and the Weil bound (1.23) for Kloosterman sums, we have in both cases

$$\Sigma_2 \ll \Big(\sum_{|n_1| \geq p/2} d(|n_1|) \Big| W\Big(\frac{n_1}{Y}\Big) \Big| \Big) \times \Big(\sum_{n \neq 0} d(|n|) \Big| W\Big(\frac{n}{Y}\Big) \Big| \Big)^{\kappa - 1}.$$

Applying (2.9) with $A \geq 3$, we deduce

$$\Sigma_2 \ll \; X^\epsilon \, (Y^A / p^{A-1}) Y^{\kappa - 1}$$

for any $\epsilon > 0$ and hence

$$\Sigma_2 \ll X^\epsilon p \Big(\frac{p}{X}\Big)^A \Big(\frac{p^2}{X}\Big)^{\kappa - 1}.$$

By assumption, we have $p < X^{1-\delta}$, hence taking $A = A(\delta, \kappa)$ sufficiently large we prove the inequality

$$\Sigma_2 \ll X^{-1}, \tag{3.5}$$

which combined with (3.3) is acceptable in view of the error term claimed in (1.6).

3.3. **Study of Σ_1.** The study of Σ_1 is the crux of the matter. To handle precisely the sum of Kloosterman sums over a in (3.2), which is a sum over a finite field, we will use a deep result in algebraic geometry. But first of all, we must prepare the combinatorial configurations of the arguments $n_1, \ldots, n_\kappa$, in order to be able to detect the main term. We shall even put it in a more general setting to cover the proof of Theorem 1.6. The following definition deals with the *decreasing sequence of multiplicities.*

Definition 3.1 (Configuration)**.** *Let p be prime. Let $\boldsymbol{\beta} := (\beta_1, \ldots, \beta_\kappa) \in (\mathrm{PGL}_2(\mathbb{F}_p))^\kappa$ be a κ-tuple of projective linear transformations modulo p. There exist an integer ν satisfying $1 \leq \nu \leq \kappa$, a ν-tuple $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_\nu)$ of positive integers μ_i satisfying*

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_\nu \geq 1 \text{ and } \mu_1 + \cdots + \mu_\nu = \kappa.$$

and ν distinct elements $(\sigma_1, \cdots, \sigma_\nu) \in \big(\mathrm{PGL}_2(\mathbb{F}_p)\big)^\nu$, such that we have

$$\big\{\beta_1, \cdots, \beta_\kappa\big\} = \{\sigma_1, \cdots, \sigma_\nu\},$$

and

$$|\{i \,:\, 1 \leq i \leq \kappa,\, \beta_i = \sigma_j\}| = \mu_j$$

for all j, with $1 \leq j \leq \nu$. The integer ν and the ν-tuple $(\mu_1, \ldots, \mu_\nu)$ are unique, and the latter will be called the configuration *of $\boldsymbol{\beta}$, the integer ν will be called the* length of the configuration *and the entries μ_j its* multiplicities*.*

If all the multiplicities μ_j are even, we will say that $\boldsymbol{\beta}$ has a mirror configuration*. In particular its length μ is even.*

In the next proposition, we will see that the asymptotics for a sum of products of Kloosterman sums shifted by the projective transformations β_i depends only on the configuration of $\boldsymbol{\beta}$, rather than on the precise values of the β_i.

Proposition 3.2. *Let p be a prime. Let $\kappa \geq 1$, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_\kappa) \in \big(\mathrm{PGL}_2(\mathbb{F}_p)\big)^\kappa$ be a κ-tuple of elements of the projective linear group with associated configuration $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_\nu)$.*

Consider the sum

$$\mathfrak{S}(\kappa, \boldsymbol{\beta}, p) = \sum_{\substack{a \bmod p \\ \beta_i \cdot a \neq 0, \infty (1 \leq i \leq \kappa)}} \mathrm{Kl}_2(\beta_1 \cdot a; p) \cdots \mathrm{Kl}_2(\beta_\kappa \cdot a; p).$$

We then have

$$\mathfrak{S}(\kappa, \boldsymbol{\beta}, p) = A(\boldsymbol{\mu}) p + O_\kappa(p^{\frac{1}{2}}), \tag{3.6}$$

where $A(\boldsymbol{\mu})$ is the product of integrals

$$A(\boldsymbol{\mu}) = \Big(\frac{2}{\pi} \int_0^\pi (2\cos\theta)^{\mu_1} \, \sin^2\theta d\theta \Big) \cdots \Big(\frac{2}{\pi} \int_0^\pi (2\cos\theta)^{\mu_\nu} \, \sin^2\theta d\theta \Big).$$

The product $A(\boldsymbol{\mu})$ is an integer, which is positive if and only if $\boldsymbol{\beta}$ is in a mirror configuration and 0 *otherwise, in which case we have*

$$\mathfrak{S}(\kappa,\boldsymbol{\beta},p)=O(p^{\frac{1}{2}}).$$

Finally we have

$$A(2,2,\ldots,2)=1. \tag{3.7}$$

This is a generalization of a result of Fouvry, Michel, Rivat and Sárközy (see [9, Lemma 2.1]), which only dealt with the case where the β_i are all diagonal and distinct modulo p.

Proof. By the definition of the configuration, the sum equals

$$\mathfrak{S}(\kappa,\boldsymbol{\beta},p)=\sum_{\substack{a\in\mathbb{F}_p,\ \sigma_i\cdot a\neq 0,\infty\\(1\leq i\leq\nu)}}\mathrm{Kl}_2(\sigma_1\cdot a;p)^{\mu_1}\cdots\mathrm{Kl}_2(\sigma_\nu\cdot a;p)^{\mu_\nu},$$

where the elements σ_i, $1\leq i\leq\nu$, are distinct in $\mathrm{PGL}_2(\mathbb{F}_p)$.

For $\ell\neq p$, let $\mathcal{K}\ell$ be the (normalized) ℓ-adic Kloosterman sheaf constructed by Deligne and studied by Katz in [15]. This is a lisse $\overline{\mathbb{Q}}_\ell$-sheaf of rank 2 on $\mathbb{G}_{m,\mathbb{F}_p}$, which has trivial determinant. For some isomorphism $\iota:\overline{\mathbb{Q}}_\ell\to\mathbb{C}$, it satisfies

$$\iota(\mathrm{trace}(\mathrm{Frob}_{a,\mathbb{F}_p}|\mathcal{K}\ell))=-\mathrm{Kl}_2(a;p)$$

for any $a\in\mathbb{F}_p^\times$. Moreover, $\mathcal{K}\ell$ is Lie-irreducible, tamely ramified at 0 with a single unipotent Jordan block, and wildly ramified at ∞ with Swan conductor 1 and with a single break at 1/2.

Given $\gamma\in\mathrm{PGL}_2(\mathbb{F}_p)$, let $\gamma^*\mathcal{K}\ell$ be the pullback of $\mathcal{K}\ell$ by the fractional linear transformation $\gamma:\ x\mapsto\gamma\cdot x$; this sheaf is lisse on $\mathbb{P}^1_{\mathbb{F}_p}-\{\gamma^{-1}(\{0,\infty\})\}$ and for any $a\in\mathbb{F}_p$ such that $\gamma\cdot a\neq 0,\infty$, it satisfies

$$\iota(\mathrm{trace}(\mathrm{Frob}_{a,\mathbb{F}_p}|\gamma^*\mathcal{K}\ell))=-\mathrm{Kl}_2(\gamma\cdot a;p).$$

Katz [15] computed the geometric monodromy group of $\mathcal{K}\ell$, and showed that it is equal to SL_2, and coincides with the arithmetic monodromy group of $\mathcal{K}\ell$. The same is therefore true for $\gamma^*\mathcal{K}\ell$.

We make the following:

Claim. For σ_1 and σ_2 distinct elements of $\mathrm{PGL}_2(\mathbb{F}_p)$ and $\mathscr{L}$ any rank one sheaf, lisse on some non-empty open subset of $\mathbb{P}^1_{\mathbb{F}_p}$, the sheaves $\sigma_1^*\mathcal{K}\ell\otimes\mathscr{L}$ and $\sigma_2^*\mathcal{K}\ell$ are not geometrically isomorphic.

Proof. We may assume that $\sigma_1=\mathrm{Id}$ and that $\sigma=\sigma_2$ is not the identity. If σ is a homothety, the claim was proven in [20, Lemme 2.4]. We now reduce to this case. Assume that $\mathcal{K}\ell\otimes\mathscr{L}$ and $\sigma^*\mathcal{K}\ell$ are geometrically isomorphic. Since $\mathscr{L}$ is of rank 1, its only possible breaks at infinity are integral, and hence $\mathcal{K}\ell\otimes\mathscr{L}$ is wildly ramified at ∞. So $\sigma^*\mathcal{K}\ell$ is also wildly ramified at infinity, which means that $\sigma\cdot\infty=\infty$. Furthermore, $\mathcal{K}\ell\otimes\mathscr{L}$ is also ramified at 0, and hence $\sigma^*\mathcal{K}\ell$ must also be ramified, which means $\sigma\cdot 0=0$. But this implies that σ is a homothety, and we apply the result of [20]. □

Since the σ_i, $(i=1,\cdots,\nu)$ are distinct elements in $\mathrm{PGL}_2(\mathbb{F}_p)$, it follows from the Goursat-Kolchin-Ribet criterion (see [16, Prop. 1.8.2]) that the geometric monodromy group of the direct sum

$$\sigma_1^*\mathcal{K}\ell\oplus\cdots\oplus\sigma_\nu^*\mathcal{K}\ell$$

is equal to its arithmetic monodromy group and is the full product group

$$\mathrm{SL}_2\times\cdots\times\mathrm{SL}_2,$$

which indicates an asymptotic independence of the values of the Kloosterman sums $\mathrm{Kl}_2(\sigma_i\cdot a;p)$ as a varies over $\mathbb{F}_p$ such that $\sigma_i\cdot a\neq 0,\infty$, $(i=1,\ldots,\nu)$.

Using Katz's effective form of Deligne's equidistribution theorem ([15, §3.6]), we deduce that

$$\frac{1}{p-1}\sum_{\substack{a\in\mathbb{F}_p,\ \sigma_i.a\neq 0,\infty\\(1\leq i\leq\nu)}}\mathrm{Kl}\left(\beta_1\cdot a,1;p\right)^{\mu_1}\cdots\mathrm{Kl}\left(\beta_\nu\cdot a,1;p\right)^{\mu_\nu}=\prod_{i=1}^{\nu}\mu_{ST}((2\cos(\theta))^{\mu_i})+O_{\mu_1,\cdots,\mu_\nu}(p^{-1/2}),$$

where the implied constant is independent of p and μ_{ST} denotes the Sato-Tate probability measure on $[0,\pi]$, which is given by

$$\mu_{ST}(f(\theta))=\frac{2}{\pi}\int_0^\pi f(\theta)\sin^2\theta d\theta$$

(recall that $[0,\pi]$ is identified with the set of conjugacy classes of the compact group $\mathrm{SU}_2(\mathbb{C})$ via the map

$$g \in \mathrm{SU}_2(\mathbb{C}) \mapsto \mathrm{trace}(g) = 2\cos\theta,$$

and that the Sato-Tate measure is the image of the probability Haar measure of $\mathrm{SU}_2(\mathbb{C})$ under this map.)

It follows by character theory of compact groups that

$$\mathrm{mult}(\mu) = \mu_{ST}((2\cos\theta)^{\mu})$$

is precisely the multiplicity of the trivial representation in the μ-th tensor power $\mathrm{Std}^{\otimes\mu}$ of the standard 2-dimensional representation of $\mathrm{SU}_2(\mathbb{C})$. In particular, $\mathrm{mult}(\mu)$ is a non-negative integer, and it is zero if and only if μ is odd (this is obvious when writing the integrals; representation-theoretically, $\mathrm{mult}(\mu) = 0$ if μ is odd because $\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ acts by multiplication by $(-1)^{\mu}$ on $\mathrm{Std}^{\otimes\mu}$, and $\mathrm{mult}(\mu) \geq 1$ for μ even, because $\mathrm{Std}^{\otimes\mu}$ is self-dual so $\mathrm{mult}(2\mu)$ is the multiplicity of the trivial representation in $\mathrm{End}(\mathrm{Std}^{\otimes\mu})$, and the identity endomorphism gives an invariant subspace; in fact, one can check that $\mathrm{mult}(2\mu) = \binom{2\mu}{\mu}/(\mu+1)$, a Catalan number.)

As a consequence

$$A(\mu_1, \cdots, \mu_\nu) = \prod_{i=1}^{\nu} \mathrm{mult}(\mu_i),$$

is a non-negative integer, and it is non-zero if and only if all the μ_i are even, which corresponds precisely to the mirror configuration. Since $\mathrm{mult}(2) = 1$, we also have $A(2, \cdots, 2) = 1$. □

Remark 3.3. Expanding the Kloosterman sums, we see that $\mathfrak{S}(\kappa, \boldsymbol{\beta}, p)$ is a character sum in $\kappa+1$ variables. The proposition shows that this character sum has square-root cancellation, except if $\boldsymbol{\beta}$ is in mirror configuration. As in [8], we see that the *structure* of $\mathfrak{S}(\kappa, \boldsymbol{\beta}, p)$ (as a sum of products of Kloosterman sums) is crucial to our success, since it reduces the problem to detecting cancellation in the single variable a.

If $\kappa = 2$ and if $\beta_1(a) = b_1 a$ and $\beta_2(a) = b_2 a$ are diagonal, we can use the fact that the Kloosterman sum is the discrete Fourier transform of the function $x \mapsto e(\bar{x}/p)$ (and $0 \mapsto 0$) to get

$$\mathfrak{S}(2, (\beta_1, \beta_2), p) = \sum_{a \in \mathbb{F}_p^\times} \mathrm{Kl}_2(b_1 a; p)\,\mathrm{Kl}_2(b_2 a; p) = \sum_{x \in \mathbb{F}_p^\times} e\Big(\frac{\bar{x}(1 - \bar{b}_1 b_2)}{p}\Big) - \frac{1}{p}$$

by the discrete Plancherel formula. This is essentially a Ramanujan sum, and hence we see that the second moment (as in (1.10)) does not require such delicate considerations. Moreover, because the error term is here $\ll p^{-1}$ (instead of $p^{-1/2}$), the error term for the second moment is better than for the others, which explains the greater range of uniformity in the formula (1.10) of Lau and Zhao. More generally, for $\kappa = 2$ and arbitrary β_1, $\beta_2 \in \mathrm{PGL}_2(\mathbb{F}_p)$, the sum $\mathfrak{S}(2, (\beta_1, \beta_2), p)$ can be identified with a special case of a *correlation sum* as defined in [8, §1.2], for the trace weight $K(n) = e(\bar{n}/p)$. The results of [8, Th. 9.1, §11.1] imply the statement of Proposition 3.2 for $\kappa = 2$.

We can now continue our study of the sum Σ_1 defined in (3.3). Since we have $p \nmid n_i$, we have

$$\mathrm{Kl}_2(an_i; p) = \mathrm{Kl}_2(\beta_i \cdot a; p),$$

where $\beta_i \in \mathrm{PGL}_2(\mathbb{F}_p)$ corresponds to the matrix

$$\begin{pmatrix} n_i & 0 \\ 0 & 1 \end{pmatrix} \pmod{p}.$$

We denote $\boldsymbol{\beta} = (\beta_i, \ldots, \beta_\kappa)$. We also denote by $\boldsymbol{\mu}(\boldsymbol{\beta})$ the configuration of $\boldsymbol{\beta}$. Thus, by Proposition 3.2 and by (2.9), we have the equalities

$$\begin{aligned} \Sigma_1 &= p \sum \cdots \sum_{1 \leq |n_1|, \ldots, |n_\kappa| < p/2} A\big(\boldsymbol{\mu}(\boldsymbol{\beta})\big)\, \tau_\star(n_1) \cdots \tau_\star(n_\kappa) W\Big(\frac{n_1}{Y}\Big) \cdots W\Big(\frac{n_\kappa}{Y}\Big) \\ &\quad + O\Big(p^{\frac{1}{2}} \Big(\sum_{1 \leq |n| < p/2} d(|n|) \Big| W\Big(\frac{n}{Y}\Big)\Big| \Big)^{\kappa} \Big) \\ &= p\Sigma_{1,M} + O\big(p^{\frac{1}{2}+\epsilon} Y^{\kappa}\big), \end{aligned} \tag{3.8}$$

say, for any $\epsilon > 0$.

Collecting (3.3), (3.5) and (3.8), the proof of Theorem 1.2 is already complete when κ is odd, since trivially $\Sigma_{1,M} = 0$ in that case.

3.4. **Study of $\Sigma_{1,M}$ for even κ.** We remark that, by the definition of Σ_1, we have the congruence $n_i \equiv n_j \bmod p$ if and only if $n_i = n_j$. In the summation over $\boldsymbol{n} = (n_1, \ldots, n_\kappa)$ defining $\Sigma_{1,M}$, we can restrict the summation over the set of $\boldsymbol{n}$ such that the associated $\boldsymbol{\beta}$ is in mirror configuration by Proposition 3.2.

We now show that, in fact, the main contribution comes from the $\boldsymbol{n}$ in mirror configuration such that the configuration of the associated $\boldsymbol{\beta}$ is $(2, 2, \ldots, 2)$. It is easy to see that, for the remaining $\boldsymbol{n}$, the associated configuration $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_\nu)$ is such that the length ν is at most $\kappa/2 - 1$ distinct elements, and satisfy $\mu_1 \geq 4$.

The equality (3.7) and some combinatorial considerations lead to the following equality

$$\begin{aligned}
\Sigma_{1,M} &= 3 \cdot 5 \cdots (\kappa - 1)\Big(\sum_{1 \leq |n| < p/2} \tau_\star(n)^2\, W\Big(\frac{n}{Y}\Big)^2\Big)^{\frac{\kappa}{2}} \\
&\quad + O\Bigg(\sum_{1 \leq \nu \leq \frac{\kappa}{2} - 1} \sum_{\substack{\mu_1 \geq \cdots \geq \mu_\nu \geq 2 \\ 2 | \mu_i,\, \mu_1 \geq 4 \\ \mu_1 + \cdots \mu_\nu = \kappa}} \prod_{i=1}^{\nu} \sum_{1 \leq |n| < p/2} d(|n|)^{\mu_i} \Big|W\Big(\frac{n}{Y}\Big)^{\mu_i}\Big|\Bigg) \\
&= m_\kappa \Big(\sum_{1 \leq |n| < p/2} \tau_\star(n)^2\, W\Big(\frac{n}{Y}\Big)^2\Big)^{\frac{\kappa}{2}} + O(Y^{\kappa/2 - 1 + \epsilon}) \qquad (3.9)
\end{aligned}$$

for any $\epsilon > 0$, the error term arising easily from (2.9) (recall that m_κ is given by (1.9) and is the κ-th moment of a standard Gaussian). We therefore see that the proof of Theorem 1.2 is completed by combining (3.3), (3.5), (3.8) and (3.9) together with Proposition 2.5, applied with $a = b = 1$.

3.5. **Further remarks.** We compare here the estimate of Theorem 1.2 with other bounds for the moments which can be derived straightforwardly from earlier results. For simplicity, we restrict our attention to the case of cusp forms.

First, we note that it is fairly easy to deduce from Proposition 2.1 and from Proposition 2.3 that

$$E_f(X, c, a) \ll_f \frac{c}{X^{1/2}}\, d(c)^{5/2}, \qquad (3.10)$$

for any $c \geq 1$, $X \geq c$ and any integer a. When $c \leq X^{2/3}$, this statement is better than the bound

$$E_f(X, c, a) \ll X^{1/2+\epsilon} c^{-1/2}$$

coming from Deligne's estimate for $\rho_f(n)$ (this is very similar to the result first proved by Smith [24, (4)] which has the same range of uniformity; see also the remarks in [2, p. 276] and the work of Duke and Iwaniec [5, Th. 2]). Combining these two bounds in the definition (1.4) of $\mathcal{M}$, we obtain

$$\mathcal{M}_f(X, c; \kappa) \ll_\epsilon X^\epsilon \Big(\frac{c^2}{X}\Big)^{\kappa/2} \min\Big(1, \frac{X^2}{c^3}\Big)^{\kappa/2}.$$

However, for $\kappa \geq 2$, we can also write

$$\mathcal{M}_f(X, c; \kappa) \leq \Big(\max_{a \bmod c} \big|E_f(X, c, a)\big|\Big)^{\kappa - 2} \Big(\frac{1}{c} \sum_{a \bmod c} \big|E_f(X, c, a)\big|^2\Big),$$

and then using the result (1.10) of Lau and Zhao, we deduce a second inequality

$$\mathcal{M}_f(X, c; \kappa) \ll_\epsilon X^\epsilon \Big(\frac{c^2}{X}\Big)^{\kappa/2 - 1}, \qquad (3.11)$$

which holds uniformly for $X^{\frac{1}{2}} \leq c \leq X$. We then see that our result in Theorem 1.2, for $c = p$ a prime, improves (3.11) for

$$X^{\frac{1}{2}} < p < X^{\frac{2}{3}} \text{ and } \kappa \geq 3. \qquad (3.12)$$

We conclude by noting that Theorem 1.2 can be extended without much effort to cusp forms f of arbitrary level and nebentypus, which are not necessarily Hecke forms. On the other hand, it does not seem straightforward to extend the result to an arbitrary composite modulus $c \geq 1$.

3.6. **Proof of Corollary 1.4.** Corollary 1.4 is an easy consequence of the fact that convergence to a Gaussian can be detected by convergence of the moments to the Gaussian moments (see, e.g., [3, Th. 8.48, Prop. 8.49]). For p prime, let

$$X = p^2/\Phi(p), \qquad \Phi(p) \to +\infty, \qquad \Phi(p) \ll p^{\epsilon}.$$

Denoting

$$\mathrm{M}_{\star}(X, p; \kappa) = \frac{1}{p} \sum_{a \in \mathbb{F}_p^{\times}} \Big(\frac{E_{\star}(X, p, a)}{\sqrt{c_{\star, w}}} \Big)^{\kappa},$$

we see from Theorem 1.2 that for any $\epsilon > 0$, we have

$$\mathrm{M}_{\star}(X, p; \kappa) = m_{\kappa} + O\Big(\Phi(p)^{-1/2+\epsilon} + p^{-\frac{1}{2}+\epsilon} \Phi(p)^{\kappa/2} \Big) \longrightarrow m_{\kappa}$$

as $p \to +\infty$. Since this holds for any fixed integer $\kappa \geq 1$, this finishes the proof.

Remark 3.4. (1) If $X = p^{2-\delta}$ for some fixed $\delta > 0$, we can not prove the Central Limit Theorem, but nevertheless, we still deduce that the κ-moments converge to Gaussian moments when

$$1 \leq \kappa \leq \Big\lfloor \frac{1}{\delta} \Big\rfloor.$$

(2) In this result, the Gaussian moments arise in Proposition 3.2, and in fact the combinatorics of the computation is the same as in a standard case of the Central Limit Theorem, namely the convergence in distribution to a standard Gaussian of a sequence

$$\mathrm{Y}_n = \frac{2\cos(\mathrm{X}_1) + \cdots + 2\cos(\mathrm{X}_n)}{\sqrt{n}}$$

where the (X_i) are independent random variables (defined on some probability space) distributed on $[0, \pi]$ according to the Sato-Tate measure.

(3) It is natural to expect that an asymptotic formula

$$\mathcal{M}_{\star}(X, p; \kappa) \sim C_{\star}(\kappa), \tag{3.13}$$

should be true uniformly for any even κ, and

$$X^{\frac{1}{2}+\epsilon} \leq p \leq X^{1-\delta},$$

for some fixed δ $(0 < \delta < 1/2)$, which (with a corresponding upper-bound for the odd moments) would extend Corollary 1.4 to this range. This conjecture is true for $\kappa = 2$ (by (1.10)), and is in agreement with the square root cancellation philosophy (1.2).

Another partial indication in favor of this conjecture is that a lower bound of that size holds: considering $\star = f$ for simplicity, and taking $\kappa \geq 2$ even, we have

$$\mathcal{M}_f(X, p; 2) \leq \big(\mathcal{M}_f(X, p; \kappa) \big)^{\frac{2}{\kappa}} \cdot \Big(\frac{1}{p} \sum_{1 \leq a \leq p} 1 \Big)^{1 - \frac{2}{\kappa}},$$

and, by combining this with (1.10), we obtain the lower bound

$$\mathcal{M}_f(X, p; \kappa) \gg 1$$

uniformly for $X^{\frac{1}{2}} \leq c \leq X^{1-\delta}$.

4. Proof of Theorem 1.6

The proof of this Theorem has many similarities with the proof of Theorem 1.2, particularly in the computation of the error terms. We will mainly concentrate on the study of the main term of the mixed moment $\mathcal{M}_\star(X,p;\kappa,\lambda;\gamma)$.

We suppose that (1.5) is satisfied and that p is sufficiently large in terms of γ. We start from the definition (1.13) and apply the same computations leading to (3.2), (3.3) and (3.5) to write the equality

$$\mathcal{M}_\star(X,p;\kappa,\lambda;\gamma)=\frac{1}{pY^{\frac{\kappa+\lambda}{2}}}\big(\Sigma_3+O_\delta(X^{-1})\big) \tag{4.1}$$

where

$$\begin{aligned}\Sigma_3=&\sum_{1\leq |m_1|,\dots,|m_\kappa|<p/2}\cdots\sum\tau_\star(m_1)\cdots\tau_\star(m_\kappa)W\Big(\frac{m_1}{Y}\Big)\cdots W\Big(\frac{m_\kappa}{Y}\Big)\\ &\times\sum_{1\leq |n_1|,\dots,|n_\lambda|<p/2}\cdots\sum\tau_\star(n_1)\cdots\tau_\star(n_\lambda)W\Big(\frac{n_1}{Y}\Big)\cdots W\Big(\frac{n_\lambda}{Y}\Big)\\ &\times\sum_{\substack{1\leq a<p\\ a,\gamma\cdot a\neq 0,\infty}}\mathrm{Kl}_2(m_1a;p)\cdots\mathrm{Kl}_2(m_\kappa a;p)\,\mathrm{Kl}_2(n_1(\gamma\cdot a);p)\cdots\mathrm{Kl}_2(n_\lambda(\gamma\cdot a);p).\end{aligned} \tag{4.2}$$

Since p divides none of the m_i or n_j, we see that the inner sum over a is equal to $\mathfrak{S}(\kappa+\lambda,\boldsymbol{\beta},p)$, as defined in Proposition 3.2, where

$$\boldsymbol{\beta}=\big(h_{m_1},\dots,h_{m_\kappa},h_{n_1}\circ\gamma,\dots,h_{n_\lambda}\circ\gamma\big), \tag{4.3}$$

and h_m denotes the homothety

$$h_m=\begin{pmatrix}m&0\\0&1\end{pmatrix}\in\mathrm{PGL}_2(\mathbb{F}_p).$$

To apply Proposition 3.2, we have to understand which $\boldsymbol{\beta}$ are in mirror configuration, in the sense of Definition 3.1. This depends on whether γ is diagonal or not.

4.1. **When γ is not diagonal.** If γ is not a diagonal matrix, then

$$h_{m_i}\neq h_{n_j}\circ\gamma$$

for any $i=1,\dots,\kappa$ and for any $j=1,\dots,\lambda$. Hence, in that case, the configuration of $\boldsymbol{\beta}$ defined by (4.3) has (before ordering the elements by decreasing order) the shape

$$(\boldsymbol{\mu},\boldsymbol{\mu}')=(\mu_1,\cdots,\mu_\nu,\mu'_1,\cdots,\mu'_{\nu'})$$

where

$$\boldsymbol{\mu}=(\mu_1,\cdots,\mu_\nu),\qquad \boldsymbol{\mu}'=(\mu'_1,\cdots,\mu'_{\nu'})$$

are the configurations of

$$\big(h_{m_1},\dots,h_{m_\kappa}\big),\qquad\big(h_{n_1}\circ\gamma,\dots,h_{n_\lambda}\circ\gamma\big),$$

respectively. It follows from Proposition 3.2 that

$$\begin{aligned}\mathfrak{S}(\kappa+\lambda,\boldsymbol{\beta},p)&=\sum_{\substack{1\leq a<p\\ a,\gamma\cdot a\neq 0,\infty}}\mathrm{Kl}_2(m_1a;p)\cdots\mathrm{Kl}_2(m_\kappa a;p)\,\mathrm{Kl}_2(n_1(\gamma\cdot a);p)\cdots\mathrm{Kl}_2(n_\lambda(\gamma\cdot a);p)\\ &=A(\boldsymbol{\mu})A(\boldsymbol{\mu}')\,p+O_{\kappa,\lambda}(p^{\frac12}).\end{aligned} \tag{4.4}$$

Hence by (4.1), (4.2) and (4.4) and by computations similar to those we did in §3.3, we deduce the equality

$$\mathcal{M}_\star(X,p;\kappa,\lambda;\gamma)=Y^{-\frac{\kappa+\lambda}{2}}\Big(\Sigma_{3,M}(\kappa)\Sigma_{3,M}(\lambda)+O(p^{-\frac12+\epsilon}Y^{\kappa+\lambda})\Big)+O(p^{-1}), \tag{4.5}$$

with

$$\begin{aligned}\Sigma_{3,M}(\kappa)&=\sum_{1\leq |m_1|,\dots,|m_\kappa|<p/2}\cdots\sum\tau_\star(m_1)\cdots\tau_\star(m_\kappa)W\Big(\frac{m_1}{Y}\Big)\cdots W\Big(\frac{m_\kappa}{Y}\Big)A(\boldsymbol{\mu}),\\ \Sigma_{3,M}(\lambda)&=\sum_{1\leq |n_1|,\dots,|n_\lambda|<p/2}\cdots\sum\tau_\star(n_1)\cdots\tau_\star(n_\lambda)W\Big(\frac{n_1}{Y}\Big)\cdots W\Big(\frac{n_\lambda}{Y}\Big)A(\boldsymbol{\mu}').\end{aligned}$$

If κ or λ is odd, the product $A(\boldsymbol{\mu})A(\boldsymbol{\mu}')$ is zero, hence (1.18) follows in that case. If κ and λ are both even, then as in (3.9), we prove that the largest contribution comes from the case where $\boldsymbol{\mu} = (2, \dots, 2)$ and $\boldsymbol{\mu}' = (2, \dots, 2)$. Hence, by a computation similar to (3.9) and (1.6), we get the equality

$$\Sigma_{3,M}(\kappa) = \Big\{ C_\star(\kappa) + O\big(Y^{-\frac{1}{2}+\epsilon}\big) \Big\} Y^{\frac{\kappa}{2}},$$

and a similar one for $\Sigma_{3,M}(\lambda)$. Hence, by (4.5), we complete the proof of (1.18).

4.2. **When γ is diagonal.** We then write γ in the canonical form (1.14) and we suppose that

$$p > \max(|\gamma_1|, |\gamma_2|).$$

Then, by making the change of variable $a = \gamma_2 a'$, we find that the sum over a of normalized Kloosterman sums appearing in the last line of (4.2) is equal to $\mathfrak{S}(\kappa + \lambda, \boldsymbol{\beta}, p)$ as defined in Proposition 3.2, with

$$\boldsymbol{\beta} = \big(h_{\gamma_2 m_1}, \dots, h_{\gamma_2 m_\kappa}, h_{\gamma_1 n_1}, \dots, h_{\gamma_1 n_\lambda}\big). \tag{4.6}$$

If the configuration of $\boldsymbol{\beta}$ is not a mirror configuration, we have

$$\mathfrak{S}(\kappa + \lambda, \boldsymbol{\beta}, p) = O(p^{\frac{1}{2}}).$$

In particular, if $\kappa \not\equiv \lambda \bmod 2$, we deduce by (4.2), (2.9) and by similar treatment of the error terms as above, that

$$\Sigma_3 \ll p^{\frac{1}{2}+\epsilon} Y^{\kappa+\lambda}. \tag{4.7}$$

Combining this with (4.1) we complete the proof of (1.19) when κ and λ have opposite parity.

Now assume that κ and λ have same parity. The combinatorics involved is then more delicate than in §4.1, because me must take into account the cases of *crossed mirror configurations*, namely situations when some of the $\gamma_2 m_i$ are equal to some of the $\gamma_1 n_j$.

To be precise, we can decompose Σ_3 (see (4.2)) into

$$\Sigma_3 = B^{\mathrm{nm}} + B^{\mathrm{m}}_0 + \sum_{\substack{0 \le \nu \le \min(\kappa, \lambda) \\ \nu \equiv \kappa \equiv \lambda \bmod 2}} B^{\mathrm{m}}(\nu), \tag{4.8}$$

where

- B^{nm} corresponds to the contribution of the $(\gamma_2 m_1, \dots, \gamma_2 m_\kappa, \gamma_1 n_1, \dots, \gamma_1 n_\lambda)$ which are not in mirror configuration,
- B^{m}_0 corresponds to the contribution of the $(\gamma_2 m_1, \dots, \gamma_2 m_\kappa, \gamma_1 n_1, \dots, \gamma_1 n_\lambda)$ which are in mirror configuration, but that configuration is not $(2, \dots, 2)$,
- $B^{\mathrm{m}}(\nu)$ corresponds to the contribution of the $(\gamma_2 m_1, \dots, \gamma_2 m_\kappa, \gamma_1 n_1, \dots, \gamma_1 n_\lambda)$ which have a mirror configuration equal to $(2, \dots, 2)$, and where exactly ν of the $\gamma_2 m_i$ $(1 \le i \le \kappa)$ are equal to ν of the $n_\nu n_j$ $(1 \le j \le \lambda)$.

The same computation as for (4.7) gives the relation

$$B^{\mathrm{nm}} \ll p^{\frac{1}{2}+\epsilon} Y^{\kappa+\lambda},$$

which, when combined with (4.1), fits with the error term in (1.17).

We can also estimate B^{m}_0 by following the same technique which led to the error term in (3.9), and obtain

$$B^{\mathrm{m}}_0 \ll p Y^{\frac{\kappa+\lambda}{2}-1+\epsilon},$$

which, by (4.1) is absorbed by the error term in (1.17).

The case of $B^{\mathrm{m}}(\nu)$ is more delicate to treat. For the terms in that sum, exactly ν of the $\gamma_2 m_i$ $(1 \le i \le \kappa)$ are equal to ν of the $\gamma_1 n_j$ $(1 \le j \le \lambda)$, and the remaining $\gamma_2 m_i$ (resp. $\gamma_1 n_j$) are in configuration $(2, \dots, 2)$. The condition $\gamma_2 m_i = \gamma_1 n_j$ can be parametrized by $m_i = \gamma_1 t$ and $n_j = \gamma_2 t$ where t is a non-zero integer.

Appealing to Proposition 3.2, and applying some combinatorial considerations, we deduce the formula

$$\begin{aligned} B^{\mathrm{m}}(\nu) = p\,\nu!\binom{\kappa}{\nu}\binom{\lambda}{\nu}\Big(&\sum_{1\leq|\gamma_1 t|,\,|\gamma_2 t|<p/2}\tau_\star(\gamma_1 t)\tau_\star(\gamma_2 t)W\Big(\frac{\gamma_1 t}{Y}\Big)W\Big(\frac{\gamma_2 t}{Y}\Big)\Big)^{\nu}\\ &\times\big(1\cdot 3\cdots(\kappa-\nu-1)\big)\Big(\sum_{1\leq|m|<p/2}\tau_\star^2(m)W^2\Big(\frac{m}{Y}\Big)\Big)^{\frac{\kappa-\nu}{2}}\\ &\times\big(1\cdot 3\cdots(\lambda-\nu-1)\big)\Big(\sum_{1\leq|n|<p/2}\tau_\star^2(n)W^2\Big(\frac{n}{Y}\Big)\Big)^{\frac{\lambda-\nu}{2}}+O\big(p^{\frac12+\epsilon}Y^{\frac{\kappa+\lambda}{2}}\big). \end{aligned} \tag{4.9}$$

In this expression, the first term corresponds to the choice and to the contribution of the ν integers m_i and ν integers n_j which satisfy the condition $\gamma_2 m_i = \gamma_1 n_j$. The second factor corresponds to the contribution of the $\kappa-\nu$ remaining m_i which are in configuration $(2,\ldots,2)$ between themselves, and the third factor to the $\lambda-\nu$ remaining n_j in configuration $(2,\ldots,2)$ between themselves. Finally, the error term comes from the error term in (3.6).

Using the arithmetic sums $\mathcal{B}_\star(m,n,Y)$ defined in Proposition 2.5, we can thus summarize (4.9) in the form

$$B^{\mathrm{m}}(\nu) = p\,\frac{\kappa!\,\lambda!}{\nu!\,2^{\frac{\kappa+\lambda}{2}-\nu}\,((\kappa-\nu)/2)!\,((\lambda-\nu)/2)!}\mathcal{B}_\star(1,1,Y)^{\frac{\kappa+\lambda}{2}-\nu}\,\mathcal{B}_\star(\gamma_1,\gamma_2,Y)^{\nu}+O\big(p^{\frac12+\epsilon}Y^{\frac{\kappa+\lambda}{2}}\big). \tag{4.10}$$

We now obtain (1.19) by combining (3.1), (4.1), (4.8), (4.10) and Proposition 2.5.

5. Proof of Corollary 1.7

We now deduce Corollary 1.7 from Theorem 1.6. The probabilistic tool is the following standard lemma:

Lemma 5.1. *Let* $(\mathrm{X}_n,\mathrm{Y}_n)$ *be a sequence of real-valued random variables. Let* Q *be a positive definite symmetric* 2×2 *matrix. Suppose that, for any integers* λ, $\kappa\geq 0$, *we have*

$$\mathbb{E}(\mathrm{X}_n^{\kappa}\mathrm{Y}_n^{\lambda})\longrightarrow m_{\kappa,\lambda}(Q)$$

as $n\to+\infty$, *where* $m_{\kappa,\lambda}(Q)=\mathbb{E}(\mathrm{A}^{\kappa}\mathrm{B}^{\lambda})$ *for some centered gaussian vector* (A,B) *with covariance matrix* Q. *Then* $(\mathrm{X}_n,\mathrm{Y}_n)$ *converges in law to* (A,B).

This follows from the case of individual sequences using the characterization of the Gaussian vector (A,B) by its linear combinations $\alpha\mathrm{A}+\beta\mathrm{B}$ being Gaussian.

We apply this lemma to the sequence $(\mathrm{Z}_p,\mathrm{Z}_p\circ\gamma)$ for p prime, as in the statement of Corollary 1.7. Note that if $\star=d$, the main term $C_d(\kappa,\lambda,\gamma)$ still depends on p (because of the polynomials of $(\log p^2/X)$ which it involves). However, under the assumptions of Corollary 1.7 on X and p, we see that in all cases, for fixed $\kappa\geq 0$ and $\lambda\geq 0$, the limit

$$L_{\kappa,\lambda}=\lim_{p\to+\infty}\frac{C_\star(\kappa,\lambda,\gamma)}{(c_{\star,w})^{(\kappa+\lambda)/2}}$$

exists, and that

$$\lim_{p\to+\infty}\mathbb{E}(\mathrm{Z}_p^{\kappa}(\mathrm{Z}_p\circ\gamma)^{\lambda})=L_{\kappa,\lambda}. \tag{5.1}$$

If γ is not diagonal, we get by (1.18) and (1.7) that $L_{\kappa,\lambda}=m_\kappa m_\lambda$ which coincides obviously with the mixed moment $\mathbb{E}(\mathrm{A}^{\kappa}\mathrm{B}^{\lambda})$ where (A,B) are independent centered Gaussian variables with variance 1, so we obtain Corollary 1.7 in that case.

If γ is diagonal, we must check that $L_{\kappa,\lambda}$ corresponds to the mixed moments of a gaussian vector (A,B) with covariance matrix given by (1.22). For this purpose, we use the formula (1.19) and note that

$$\frac{(c_{\star,w})^{\frac{\kappa+\lambda}{2}-\nu}\,(\tilde{c}_{\star,w,\gamma})^{\nu}}{c_{\star,w}^{(\kappa+\lambda)/2}}=\Big(\frac{\tilde{c}_{\star,w,\gamma}}{c_{\star,w}}\Big)^{\nu}\to(G_{\star,\gamma,w})^{\nu}$$

as $p \to +\infty$, with notation as in (1.19) and Corollary 1.7. Thus, abbreviating $G = G_{\star,\gamma,w}$, we compute the 2-variable exponential generating series of $L_{\kappa,\lambda}$ by writing

$$\begin{aligned}\sum_{\kappa,\lambda\geq 0}\frac{1}{\kappa!\lambda!}L_{\kappa,\lambda}U^{\kappa}V^{\lambda} &= \sum_{\kappa,\lambda\geq 0}\frac{1}{\kappa!\lambda!}U^{\kappa}V^{\lambda}\sum_{\substack{0\leq\nu\leq\min(\kappa,\lambda)\\ \nu\equiv\kappa\equiv\lambda \bmod 2}}\nu!\binom{\kappa}{\nu}\binom{\lambda}{\nu}m_{\kappa-\nu}\,m_{\lambda-\nu}G^{\nu}\\ &= \sum_{\nu\geq 0}\nu!G^{\nu}\sum_{k,l\geq 0}\frac{U^{\nu+2k}V^{\nu+2l}}{(\nu+2k)!(\nu+2l)!}\binom{\nu+2k}{\nu}\binom{\nu+2l}{\nu}m_{2k}m_{2l}\\ &= \sum_{\nu\geq 0}\frac{G^{\nu}(UV)^{\nu}}{\nu!}\sum_{k\geq 0}\frac{m_{2k}U^{2k}}{(2k)!}\sum_{l\geq 0}\frac{m_{2l}V^{2l}}{(2l)!} = \exp\Big(\frac{U^2}{2}+GUV+\frac{V^2}{2}\Big).\end{aligned}$$

Since this is well-known to be the exponential generating series of the moments of the Gaussian vector with covariance matrix (1.22), we obtain the desired convergence in law.

References

[1] E.M. Baruch and Z. Mao, *Bessel identities in the Waldspurger correspondence over the real numbers*, Israel J. Math. 145 (2005), 1–81.

[2] V. Blomer, *The average value of divisor sums in arithmetic progressions*, Q. J. Math. 59 (2008) no 3, 275–286.

[3] L. Breiman, *Probability*, Addison Wesley, 2000.

[4] J.W. Cogdell and I. Piatetski-Shapiro, *The arithmetic and spectral analysis of Poincaré series*, Perspectives in Math. 13, Academic Press, 1990.

[5] W.D. Duke and H. Iwaniec, *Estimates for coefficients of L-functions, I*, in "Automorphic forms and analytic number theory" (Montreal, PQ, 1989), 43–47, Univ. Montreal, Montreal, QC, 1990.

[6] É. Fouvry, *Autour du théorème de Bombieri–Vinogradov*, Acta Math. 152 (1984), no. 3-4, 219–244.

[7] É. Fouvry, *Cinquante ans de théorie analytique des nombres*, in "Development of mathematics 1950–2000", 485–514, Birkhäuser, Basel, 2000.

[8] É. Fouvry, E. Kowalski, and P. Michel, *Algebraic twists of modular forms and Hecke orbits*, preprint `arXiv:1207.0617`

[9] É. Fouvry, P. Michel, J. Rivat and A. Sárközy, *On the pseudorandomness of the signs of Kloosterman sums*, Journal of the Australian Mathematical Society, Volume 77, December 2004, 425–436.

[10] I.S. Gradshteyn and I.M. Ryzhkik, *Tables of integrals, series and products*, 5th ed., Academic Press (1994).

[11] H. Iwaniec, *Introduction to the spectral theory of automorphic forms*, Biblioteca de la Revista Matemática Iberoamericana, Madrid, 1995.

[12] H. Iwaniec, *Topics in classical automorphic forms*, Grad. Studies in Math. 17, American Mathematical Society, 1997.

[13] H. Iwaniec and E. Kowalski,*Analytic Number Theory*, American Mathematical Society Colloquium Publications, 53. American Mathematical Society, Providence, RI, 2004.

[14] J. Jacod and P. Protter, *Probability Essentials*, Universitext, Springer, 2000.

[15] N.M. Katz, *Gauss sums, Kloosterman sums, and monodromy groups*, Annals of Math. Studies, 116. Princeton University Press, Princeton, NJ, 1988.

[16] N.M. Katz, *Exponential sums and differential equations*, Annals of Math. Studies 124, Princeton Univ. Press, Princeton, NJ, 1990.

[17] E. Kowalski and G. Ricotta, *Fourier coefficients of $GL(N)$–automorphic forms in arithmetic progressions*, preprint (2013).

[18] Y–K. Lau and L. Zhao, *On the variance of Hecke eigenvalues in arithmetic progressions*, J. Number Theory 132 (2012), 869–887.

[19] G. Lü, *The average value of Fourier coefficients of cusp forms in arithmetic progressions*, J. Number Theory 129 (2009), 488-494.

[20] P. Michel, *Autour de la conjecture de Sato-Tate pour les sommes de Kloosterman. I*, Invent. math., 121 (1995), 61–78.

[21] H. Montgomery, *Primes in arithmetic progressions*, Michigan Math. J. 17 (1970), 33–39.

[22] Y. Motohashi, *On the distribution of the divisor function in arithmetic progressions*, Acta Arith. 22 (1973), 175–199.

[23] F. Oberhettinger, *Tables of Bessel transforms*, Springer Verlag 1972.

[24] R.A. Smith, *Fourier coefficients of modular forms over arithmetic progressions. I, II. With remarks by M. R. Murty*, C. R. Math. Rep. Acad. Sci. Canada 15 (1993), no. 2–3, 85–90, 91–98.

[25] G.N. Watson, *A treatise on the theory of Bessel functions*, Merchant Books, 2008.

Université Paris Sud, Laboratoire de Mathématique, Campus d'Orsay, 91405 Orsay Cedex, France
E-mail address: etienne.fouvry@math.u-psud.fr

Indian Statistical Institute, Kolkata, India
E-mail address: satadalganguly@gmail.com

ETH Zürich – D-MATH, Rämistrasse 101, CH-8092 Zürich, Switzerland
E-mail address: kowalski@math.ethz.ch

EPFL/SB/IMB/TAN, Station 8, CH-1015 Lausanne, Switzerland
E-mail address: philippe.michel@epfl.ch