hexagon:2611.00852v1

Second Moments of Hecke Traces and Vertical Sato–Tate

Zhongpai Gao

21 November 2026

Abstract

We relate sums of squared Hecke traces to the distribution of Frobenius traces of elliptic curves over $\mathbf{F}_p$. For every fixed $1/6 \le a < 1/2$, the sum of squared normalized traces at level one up to weight $K = p^a$ is asymptotic to $K^2/48$ on a density-one set of primes. In the same range, fixed-level limits hold on sets of relative density one in the relevant prime classes, with constants that factor over prime powers. We compute the mixed moments between newform spaces at different levels and between character spaces at odd prime level.

The connection comes from tossing a coin whose bias is determined by the trace of a uniformly chosen nonsingular Weierstrass equation. The Sato–Tate predictor's excess log loss tends to zero exactly when $T = o(p)$. For $T = \lfloor \lambda K^2 \rfloor$ with fixed $\lambda > 0$ it is asymptotic to $T/(96p)$ on the level-one prime set. When $T/p \to c > 0$, we compute the limiting losses along prime subsequences whose residues stabilize modulo each fixed integer. The two nonzero classes modulo 3 can give different limits. At level one we compute the leading variances across prime intervals for the normalized trace sum and the normalized chi-square divergence.

1 Introduction

For a prime p, let $\tau_k(p)$ be the trace of T_p on $S_k(\mathrm{SL}_2(\mathbf{Z}))$ divided by $p^{(k-1)/2}$. We prove that for each fixed $1/6 \le a < 1/2$ there is a density-one set of primes on which, with $K = p^a$,

$$\sum_{\substack{4 \le k \le K \\ k \text{ even}}} |\tau_k(p)|^2 \sim \frac{K^2}{48}.$$

The sum has the same leading term as the total dimension, since $\dim S_k(\mathrm{SL}_2(\mathbf{Z})) = k/12 + O(1)$ at even weights. At fixed level the full and newform moments are products of local factors that can depend on the residue of p. Distinct squarefree newform levels have zero limiting mixed moment. At odd prime level we also separate the character components.

Jung and Sardari [18, Theorem 1.3] bound the trace square sum on newspaces of fixed squarefree level $N > 1$ at $k \asymp \sqrt{p}$. At that scale Bober et al. [4, Theorem 1.1] and Kuan and Lesesvre [26, Theorem 1] study signed first moments under GRH, the latter with Petersson weights. Petrow [33, Theorem 1.1] bounds individual traces for $k \ge p^{3/10+\eta}$ at fixed level. At fixed weight, averaging over primes gives the diagonal dimension by Rankin–Selberg orthogonality [19, Theorems 2.5–2.6]. In a residue class the prime average can pair forms related by a Dirichlet twist. Our asymptotic concerns a sharp cutoff $K = p^a = o(\sqrt{p})$ at individual primes in sets of relative density one, with explicit mixed constants at every fixed level.

Binomial smoothing supplies the weight cutoff. For a uniform nonsingular short Weierstrass equation E over $\mathbf{F}_p$, toss a coin T times with bias $f_t = 1/2 + t/(4\sqrt{p})$, where $t = p + 1 - \#E(\mathbf{F}_p)$. Write P_T for the count law and Q_T for its analogue under the Sato–Tate prior $\mathrm{Beta}(3/2, 3/2)$.

The expected extra log loss is $D_{\mathrm{KL}}(P_T\|Q_T)$. The chi-square divergence sums squared Chebyshev moments with weights that decay past $\sqrt{T}$. Kaplan and Petrow's formula [20] identifies these moments with normalized Hecke traces. Varying T recovers the sharp sum.

We call $1/2+(4\sqrt{p})^{-1}\mathbf{Z}$ the *trace lattice*. In the interior the sampling error $T^{-1/2}$ reaches its spacing at $T\asymp p$, where the limiting losses reflect the residues of p. At shorter lengths the main term comes from squared Fourier coefficients. Averaging over primes controls the products at distinct frequencies. At $a=1/6$ the proof also uses the quadratic functional equation and additive reciprocity. The short-interval estimates of David, Koukoulopoulos and Smith [9, Theorem 1.3] control the likelihood ratios.

For the short-length results, put $K=p^a$ and $T=\lfloor \lambda K^2\rfloor$ with fixed $1/6\le a<1/2$ and $\lambda>0$.

Result	**Conclusion**	**Prime coverage**
Level-one traces	$\sum_{4\le k\le K,\ k\ \text{even}}\lvert\tau_k\rvert^2\sim K^2/48$	Density one
Fixed-level traces	Local products in units of $K^2/48$	Relative density one
Level-one variance	Explicit multiples of $X^{a-1/2}\log X$ for $a\ge 3/8$	Average over $X<p\le 2X$
Short-length losses	$p\chi^2/T\to 1/48$, $pD_{\mathrm{KL}}/T\to 1/96$ at level one. Local multiples for fixed trace classes.	Relative density one
Loss threshold	$D_{\mathrm{KL}}\to 0$ exactly when $T=o(p)$	All primes
Critical profiles	Explicit χ^2 and KL limits when $T/p\to c>0$	Fixed limiting residues

Section 5 derives the Fourier profile at the threshold. Section 6 uses it to prove the trace asymptotics. The supplement lists the proof dependencies.

The trace asymptotics and leading variance are unconditional. The second terms for the prime means use Alpöge's Dirichlet theorem, released in Lean on 20 November 2026 and accepted by independent kernels. A full written proof by Alpöge has yet to appear [2, 39]. Its quadratic case also gives the sharper bounds in Section 7 and S1.

2 Prediction loss for binary mixtures

For word laws A and B, the chain rule gives the expected excess log loss of using B as $D_{\mathrm{KL}}(A\|B)=\mathbb{E}_A\log(A/B)$, with natural logarithms. We also use $\chi^2(A\|B)=\mathbb{E}_B[(A/B-1)^2]$. Draw $\theta\in[0,\pi]$ once, then draw bits independently with bias $f=(1+\cos\theta)/2$. Write $A_\nu^{(T)}$ for the word law with $f\sim\nu$. The likelihood ratio depends only on the count of ones, so either divergence can be computed from the count laws.

The reference $B_{a,b}^{(T)}$ uses a Beta(a,b) prior ($a,b>0$) and predicts a one with probability $(k+a)/(j+a+b)$ after k ones in j bits. The choices $a=b=1/2$ and $a=b=3/2$ give the uniform and Sato–Tate angle laws, respectively. The first is the Krichevsky–Trofimov predictor [24].

2.1 A coefficient formula for prediction

Proposition 2.1. *Let P_m be the orthonormal polynomials for the Beta(a,b) law of f, with $P_0=1$. With*

$$\kappa_m(T)=\prod_{j=0}^{m-1}\frac{T-j}{T+a+b+j},$$

one has

$$\chi^2(A_\nu^{(T)}\|B_{a,b}^{(T)})=\sum_{m=1}^{T}\kappa_m(T)\left(\int P_m\,d\nu\right)^2.$$

For $a = b = 1/2$, $P_m = \sqrt{2}\cos(m\theta)$ for $m \geq 1$. For $a = b = 3/2$, $P_m = U_m(\cos\theta)$, where U_m is the Chebyshev polynomial of the second kind. In both cases $\kappa_m(T) \leq e^{-m(m+a+b-1)/(2T+2)}$.

Proof. Under the beta-binomial reference, the conditional expectations from count to bias and from bias to count are adjoints. Binomial factorial moments and the Beta$(a+k, b+T-k)$ posterior moments show that both maps preserve polynomial degree. By orthogonality, each map therefore sends a degree-m orthonormal polynomial to a scalar multiple of its counterpart. From the leading coefficients, the composition multiplies by $T(T-1)\cdots(T-m+1)/\prod_{j<m}(T+a+b+j) = \kappa_m(T)$. In particular, the coefficient of the actual count density relative to the reference in the degree-m basis is $\sqrt{\kappa_m(T)}\int P_m\,d\nu$, up to the choice of signs. Parseval gives the identity. The two special polynomial families follow by the change $f = (1+\cos\theta)/2$. For the bound, use

$$\frac{T-j}{T+a+b+j} \leq \exp\left(-\frac{2j+a+b}{2T+2}\right), \qquad j < m \leq T,$$

for $a + b \in \{1, 3\}$, and multiply. □

The beta-binomial diagonalization is classical [11, Propositions 4.1 and 4.2]. Its weights decay rapidly beyond degree $\sqrt{T}$. Jensen's inequality gives $D_{\mathrm{KL}} \leq \log(1+\chi^2)$.

Lemma 2.2 (Finite support)**.** *For $a = b \in \{1/2, 3/2\}$, if ν has at most N atoms, then*

$$D_{\mathrm{KL}}(A_\nu^{(T)} \| B_{a,b}^{(T)}) \geq \tfrac{1}{2}\log(T+1) - \log N - O(1).$$

Proof. The divergence of the count mixture is the mean of the divergences of its individual binomial laws minus the mutual information between the count and the atom. The latter is at most $\log N$. For a binomial count k of bias f, put $v = Tf(1-f)$. Uniformly in f, its entropy $H(k)$ is $\frac{1}{2}\log(1+v)+O(1)$. The upper bound follows by adding an independent uniform variable of length one and applying the Gaussian entropy bound [7]. For the lower bound, Fourier inversion and $|1-f+fe^{it}|^T \leq e^{-2v\sin^2(t/2)}$ give a maximum count mass $O((1+v)^{-1/2})$.

Write $Q_T(k)$ for the reference count mass. For $a = b = 3/2$, $Q_T(k) \ll (T+1)^{-1}$ since the Beta density is bounded. For $a = b = 1/2$, the gamma ratio formula gives $Q_T(k) \ll ((k+1)(T-k+1))^{-1/2}$. The identity $\mathbb{E}[1/(k+1)] = (1-(1-f)^{T+1})/((T+1)f)$, interpreted at $f = 0$ by continuity, gives $\mathbb{E}[1/(k+1)] \leq 2/(1+Tf)$. Jensen's inequality applied to $k+1$ and to its reciprocal gives $\mathbb{E}\log(k+1) = \log(1+Tf)+O(1)$, with the reflected estimate for $T-k$. Since $(1+Tf)(1+T(1-f)) = T+1+Tv$, in both cases $\mathbb{E}[-\log Q_T(k)] - H(k) \geq \frac{1}{2}\log(T+1) - O(1)$. □

3 Elliptic-curve coins

Let P_T be the count law from a uniform nonsingular equation $y^2 = x^3 + ax + b$ over $\mathbf{F}_p$, $p \geq 5$, with bias f_t as in the introduction. Its Sato–Tate reference Q_T is beta-binomial. For $1 \leq T \leq X$ and $\varepsilon > 0$ put

$$R_\varepsilon(T, X) = \frac{T}{X}\log^3(2X/T) + X^{-1/3+\varepsilon}\max(T, \sqrt{X})^{1/4}.$$

Theorem 3.1. *For integers $1 \leq T \leq p$ and every $\varepsilon > 0$,*

$$D_{\mathrm{KL}}(P_T \| Q_T) \leq \chi^2(P_T \| Q_T) \ll_\varepsilon R_\varepsilon(T, p).$$

As $p \to \infty$, the KL divergence tends to zero if and only if $T = o(p)$. For $T \geq p$ it is $\frac{1}{2}\log(T/p)+O(1)$, uniformly.

The short-interval estimate of [9, Theorem 1.3] gives vanishing loss for $T \le p^{1-\eta}$ with fixed $\eta > 0$, by Section 6.2 and monotonicity. For a quadratic field K with character χ_K, the factorization $\zeta_K(s) = \zeta(s)L(s,\chi_K)$ transfers Alpöge's critical-line theorem [2, 39] to ζ_K. Thus Ma [29, Theorem 1.5] gives histogram convergence for trace lengths $f(p) = o(\sqrt{p})$ with $f(p) \ge \exp(C(\log\log p)^2)$ eventually for every fixed $C > 0$. To reach all $T = o(p)$ we estimate the Fourier modes of the trace weights after binomial sampling.

3.1 Class numbers on the trace lattice

Let $D_t = 4p - t^2 > 0$ and $\Delta_p = (4\sqrt{p})^{-1}$. We use the Hurwitz class number $H(D)$ with the normalization $H(3) = 1/3$, $H(4) = 1/2$. By Deuring's theorem, the $\mathbf{F}_p$-isomorphism classes with trace t, each weighted by $2/\#\operatorname{Aut}_{\mathbf{F}_p}(E)$, have total weight $H(D_t)$ [8, Theorem 14.18][28, (1.5)–(1.6)]. A class contains $(p-1)/\#\operatorname{Aut}_{\mathbf{F}_p}(E)$ equations, so $(p-1)H(D_t)/2$ equations have trace t. Dividing by $p(p-1)$ gives the bias mass $\Delta_p\sigma(f_t)F(t)$, where $\sigma(f) = (8/\pi)\sqrt{f(1-f)}$ and $F(t) = \pi H(D_t)/\sqrt{D_t}$ is the correction to the sampled Sato–Tate mass.

Write $-D_t = d_0 r^2$ with d_0 fundamental and $\chi = \chi_{d_0}$. Let $\rho_t(n)$ count roots of $x^2 - tx + p$ modulo n and set $b_t = \lambda_{\mathrm{L}} * \rho_t$, with λ_{L} the Liouville function. The change $x \mapsto t - 2x$ identifies the roots modulo n with the classes $b \bmod 2n$ satisfying $b^2 \equiv -D_t \pmod{4n}$, including at 2. Zagier's formula [40, Proposition 3] gives

$$b_t = c_t * \chi, \qquad c_t(n) = \sum_{\substack{ge^2 = n \\ ge \mid r}} \mu(g)\chi(g)e, \qquad F(t) = \sum_{n \ge 1} b_t(n)/n.$$

The last series converges in ordinary order because c_t has finite support and χ is nonprincipal. Its value at $s = 1$ is the class-number formula for the orders of discriminant dividing $-D_t$ by a square [8, Theorem 7.24]. We use $\sum |c_t(n)|/n \le \sum_{m \mid r} \tau(m)/m \le \sigma_{-1}(r)^2$ and write $F_N(t) = \sum_{n \le N} b_t(n)/n$.

Lemma 3.2 (Truncating the class-number weights)**.** *For $1 \le N \le p^{13/30}$ and every $\varepsilon > 0$,*

$$\sum_{|t| < 2\sqrt{p}} |F(t) - F_N(t)|^2 \ll_\varepsilon \sqrt{p}\left(\frac{\log^4(2N)}{N} + p^{-1/15+\varepsilon}\right).$$

Proof. Fourier coefficients of the truncated series. For $0 \le A < B < p$, expand $\sum_{A < n \le B} b_t(n)/n$ into $e(at/d) = \exp(2\pi i at/d)$ at reduced fractions a/d with $d \le B$. Every root modulo $n < p$ is a unit. Fourier inversion of $t = x + p/x \bmod n$ gives the Kloosterman sum $S(a, ap; n)$, where $S(m, n; d) = \sum^*_{x \bmod d} e((mx + nx^{-1})/d)$ and the star restricts the sum to units. Reducing the frequency to denominator d contributes a factor $\varphi(n)/\varphi(d)$, by the surjection of unit groups. After convolution the coefficient is

$$\frac{S(a, ap; d)}{\varphi(d)} \sum_{A < n \le B} \frac{h_d(n)}{n}, \qquad h_d(n) = \sum_{\substack{m \mid n \\ d \mid m}} \lambda_{\mathrm{L}}(n/m)\frac{\varphi(m)}{m}.$$

The sign of the Fourier phase is immaterial since $S(-a, -ap; n) = S(a, ap; n)$. The coefficient at $d = 1$ uses the zero frequency.

Multiplying the local factors gives the Dirichlet series $\varphi(d)d^{-1-s}\zeta(2s)\zeta(s+1)^{-1}\prod_{\ell \mid d}(1-\ell^{-1-s})^{-1}$ for h_d. Expanding the three factors after d^{-1-s} into a square, a divisor m with weight $\mu(m)/m$, and an integer j supported on primes dividing d with weight $1/j$ gives

$$\sum_{n > A} \frac{|h_d(n)|}{n} \ll \frac{\varphi(d)}{d^2}\min(1, \sqrt{d/A}),$$

with the minimum interpreted as 1 at $A = 0$. For the tail, sum $s^{-2}m^{-2}j^{-2}$ over $s^2mj > A/d$, first in s. In particular the constant coefficient of F_N is $1+O(N^{-1/2})$. Writing $\tau(d)$ for the divisor function, the Weil bound [16, Corollary 11.12] shows that the coefficient of $e(at/d)$ is $O(\tau(d)d^{-3/2}\min(1, \sqrt{d/A}))$.

The terms with $N < n \le p^{13/30}$. Take $A = N$ and $B = p^{13/30}$. In the mean square over the trace interval the diagonal is $\ll \sqrt{p}\,\log^4(2N)/N$, by summing $\tau(d)^2d^{-2}\min(1, d/N)$. For distinct reduced fractions the geometric sum costs $O(\|a/d - a'/d'\|^{-1})$, where $\|x\|$ is the distance to the nearest integer. Writing $g = \gcd(d, d')$, the map $(a, a') \mapsto ad' - a'd \bmod dd'$ on all residue pairs has image the multiples of g and fibres of size g. Omitting zero, the inverse distances sum to $O(dd'\log(2dd'))$. Hence the off-diagonal contribution is $O(B\log^3(2B))$, using $\sum_{d\le B}\tau(d)d^{-1/2} \ll \sqrt{B}\log(2B)$.

The tail beyond $p^{13/30}$. Partial summation in Burgess's estimate with exponent parameter 3 [5] gives $\sum_{n>y}\chi(n)/n \ll_\varepsilon p^{1/9+\varepsilon}y^{-1/3}$. This also holds for $0 < y < 1$ by $L(1,\chi) \ll \log p$. Apply it to each term of $b_t = c_t * \chi$ at $y = B/m$. Then

$$|F(t) - F_B(t)| \ll_\varepsilon p^{1/9+\varepsilon}B^{-1/3}\sum_m |c_t(m)|m^{-2/3} \ll_\varepsilon p^{1/9+2\varepsilon}B^{-1/3},$$

since the sum is at most $\sum_{ge|r} g^{-2/3}e^{-1/3} \ll_\varepsilon p^\varepsilon$. The square of the far-tail estimate and the normalized off-diagonal bound are both $O_\varepsilon(p^{-1/15+\varepsilon})$, after decreasing ε in the intermediate estimates. □

3.2 From trace weights to prediction loss

Write $b_{T,f}(k)$ for the binomial mass at k and use K for the random count. For weights u_t, put $P^u(k) = \sum_t \Delta_p\sigma(f_t)u_t b_{T,f_t}(k)$ and $\|A\|_Q^2 = \sum_k |A(k)|^2/Q_T(k)$. Thus $P_T = P^F$, and P^1 is the reference prior sampled on the trace lattice before normalization. The posterior under the continuous reference is $\mathrm{Beta}(k + 3/2, T - k + 3/2)$. Write $m_{T,k}$ for its maximum density. Stirling's bounds at the mode give $m_{T,k} \ll T/\sqrt{1 + \min(k, T-k)}$. The beta-binomial formula gives $Q_T(k) \ll (T+1)^{-2}\sqrt{(k+1)(T-k+1)}$, so $\sum_k Q_T(k)m_{T,k}^2 \ll T$.

The grid has step Δ_p and end gaps at most Δ_p. The Riemann-sum error of a unimodal density is $O(\Delta_p)$ times its maximum, so $P^1(k)/Q_T(k) = 1 + O(\Delta_p m_{T,k})$ and $\|P^1 - Q_T\|_Q^2 \ll T/p$. Cauchy–Schwarz at each count, followed by summing over counts, gives

$$\|P^u\|_Q^2 \ll \Delta_p\sum_t \sigma(f_t)|u_t|^2(1 + \Delta_p\,\mathbb{E}_{\mathrm{Bin}(T,f_t)}m_{T,K}).$$

The binomial reciprocal moment bounds used in Lemma 2.2 give $\mathbb{E}m_{T,K} \ll \sqrt{T}(f_t^{-1/2} + (1-f_t)^{-1/2})$. The density $\sigma(f_t)$ cancels these singularities, giving a norm bound for errors in the trace weights.

$$\|P^u\|_Q^2 \ll \Delta_p(1 + \Delta_p\sqrt{T})\sum_t |u_t|^2. \tag{1}$$

The norm bound treats all trace modes alike. Binomial sampling gives extra decay for the nonconstant modes.

Lemma 3.3 (Binomial smoothing)**.** *Let $V(t) = \sum_\alpha \gamma_\alpha e(\alpha t)$, with nonzero reduced frequencies $\alpha = a/d \bmod 1$, $d \le B \ll p$ and $|\gamma_\alpha| \ll \tau(d)d^{-3/2}$. Suppose that $|V(t)| \ll_\varepsilon p^\varepsilon$ on the trace lattice. Then for $\sqrt{p}(\log p)^8 \le T \le p$,*

$$\|P^V\|_Q^2 \ll_\varepsilon \frac{T}{p}\log^3(2p/T) + p^\varepsilon\left(\frac{B\sqrt{T}}{p} + \frac{1}{T}\right). \tag{2}$$

Proof. Put $\theta_t = \arccos(t/(2\sqrt{p}))$, $h = (2T+2)^{-1}$ and $G(x) = \sum_{j\in\mathbf{Z}} e^{-hj^2/2}e^{ijx}$. The coefficient identity in Proposition 2.1, with the degree-zero term included for signed measures, and $\sigma(f_t)U_m(\cos\theta_t) = (4/\pi)\sin((m+1)\theta_t)$ give by Parseval

$$\|P^V\|_Q^2 \ll p^{-1}\int_{-\pi}^{\pi}\left|\sum_t V(t)G(\phi-\theta_t)\right|^2 d\phi.$$

Here G is a Gaussian summed over translates by 2π. The quantity $w_\phi = 2\sqrt{p}\sin\phi$ is the absolute derivative of the trace with respect to the angle. It vanishes at the endpoints.

Near the endpoints. An angular interval of radius z about $\phi\in[0,\pi]$ contains $O(1+w_\phi z+\sqrt{p}\,z^2)$ traces. Thus the sum of the kernels is $O(\sqrt{T})$ where $w_\phi < \sqrt{T}$ and outside $[0,\pi]$. Up to an exponentially small Gaussian error only traces within angular distance $(\log p)/\sqrt{T}$ of these regions contribute. They lie in two intervals of length $O(T/\sqrt{p})$. On any interval of length L, the geometric sum estimate used in Lemma 3.2 gives $\sum_t |V(t)|^2 \ll L+1+B\log^3(2B)$, using the coefficient bound with no truncation factor. By Cauchy–Schwarz against the positive kernels and $\int G = 2\pi$, the edge contribution is $O((T/p)^{3/2} + B\sqrt{T}\,p^{-1}\log^3(2B))$.

Away from the endpoints. In the remaining region set $s = 2\sqrt{p}\cos\phi$ and $w(s) = \sqrt{4p-s^2} \geq \sqrt{T}$. For $|t-s| \leq (\log p)w(s)/\sqrt{T}$, Taylor expansion replaces $G(\phi-\theta_t)$ by $(2\pi/h)^{1/2}\exp(-(t-s)^2/(2hw(s)^2))$ with relative error $O((\log p)^3\sqrt{p}/(\sqrt{T}\,w(s)))$. At the boundary of this range, $|\theta_t-\phi| \geq (1-o(1))(\log p)/\sqrt{T}$. Monotonicity of θ_t makes the remaining Gaussian tails negligible, and the approximating kernels sum to $O(w(s))$. With the bound for V on the trace lattice, the squared replacement error after division by p is $O_\varepsilon(p^\varepsilon/T)$. The trace interval ends are at distance $\gg w(s)^2/\sqrt{p}$. Using $\sup_{\mathbf{Z}}|V| \leq \sum_\alpha|\gamma_\alpha| \ll \sqrt{B}\log(2B)$, extension of the approximating sum to all integers adds only Gaussian tails. Poisson summation gives

$$2\pi w(s)\sum_\beta \gamma_{\beta \bmod 1} e(\beta s)\exp(-2\pi^2 h w(s)^2\beta^2),$$

where β runs over all integer translates of the frequencies.

Squaring the Fourier sum. Split each half of the s-interval into dyadic ranges $w(s)\asymp W$ and put $H = W/\sqrt{T} \geq 1$. Such an interval has length $O(W^2/\sqrt{p})$. Since $d\phi = |ds|/w(s)$, the diagonal contribution on it is bounded by

$$\frac{W^3}{p^{3/2}}\sum_{d\leq B}\frac{\tau(d)^2}{d^3}\sum_{a\neq 0} e^{-cH^2a^2/d^2} \ll \frac{TW}{p^{3/2}}\log^3(2H).$$

For $d \geq H$ the inner sum is $O(d/H)$ and for $d < H$ it is $O((d/H)^3)$. Summing in W gives the first term in (2).

For distinct β,β', integration by parts on the interval bounds the corresponding integral, before division by p, by

$$\frac{CW}{|\beta-\beta'|}\,e^{-cH^2(\beta^2+\beta'^2)}.$$

The amplitude $w\exp(-2\pi^2hw^2(\beta^2+\beta'^2))$ has at most one turning point along the interval. Its total variation and the two endpoint values are bounded by a constant times its maximum. For dyadic denominator ranges $d\asymp D$, $d'\asymp E$, where by symmetry we may take $D\leq E$, the β' are spaced by $\gg E^{-2}$. Their nonzero distances from a fixed β are at least $\gg (DE)^{-1}$. Therefore

$$\sum_{\beta'\neq\beta}\frac{e^{-cH^2\beta'^2}}{|\beta-\beta'|} \ll E^2\log(2E), \qquad \sum_{\beta:\ d\asymp D} e^{-cH^2\beta^2} \ll D^2/H.$$

In the first sum use spacing for distances at most one and the Gaussian for larger distances. The second follows by summing over nonzero integer numerators. Including the Fourier coefficients, the mixed contribution is $\ll_\varepsilon p^\varepsilon W\sqrt{DE}/(pH)$. Summing the dyadic ranges gives $O_\varepsilon(p^\varepsilon B\sqrt{T}/p)$, proving (2). □

Put $T_0 = \lceil \sqrt{p}(\log p)^8 \rceil$.

Proof of Theorem 3.1. Short observations. To balance truncation and smoothing, choose a smooth cutoff ψ equal to one on $[0,1]$ and zero on $[2,\infty)$, with $0 \le \psi \le 1$, and set $F_{B,\psi}(t) = \sum_n b_t(n)n^{-1}\psi(n/B)$. Write $C_t(s) = \sum_m c_t(m)m^{-s}$ and let $\widehat{\psi}$ be the Mellin transform. In the integral for $F_{B,\psi}$, shift the line in $L(1+z,\chi)C_t(1+z)B^z\widehat{\psi}(z)$ to $\Re z = -1/2$. The formula $\widehat{\psi}(z) = -z^{-1}\int_1^2 \psi'(x)x^z\,dx$ gives rapid decay and a simple pole at zero of residue one. The residue of the integrand is $F(t)$. On the new line $|C_t(1/2+iv)| \le \sum_{ge|r} g^{-1/2} \ll_\varepsilon p^\varepsilon$. The hybrid Weyl bound [34, Theorem 1.1] therefore gives

$$|F(t) - F_{B,\psi}(t)| \ll_\varepsilon p^{1/6+\varepsilon}B^{-1/2}.$$

By (1) the squared norm of the omitted tail is $O_\varepsilon(p^{1/3+\varepsilon}/B)$. The finite Fourier expansion in Lemma 3.2 applies with the cutoff inserted in the h_d sum. For $2B < p$ this gives the same coefficient bounds up to denominator $2B$ and a bounded constant term. On the trace lattice the cutoff weight is $O_\varepsilon(p^\varepsilon)$, since $b_t = c_t * \chi$ gives $\sum_{n\le 2B}|b_t(n)|/n \ll \log(2B)\sum_{ge|r}(ge)^{-1}$. Thus its nonconstant part satisfies (2). The two errors that depend on the cutoff have sizes $p^{1/3+\varepsilon}/B$ and $p^\varepsilon B\sqrt{T}/p$. Taking $B = p^{2/3}T^{-1/4}$ balances them. Together with $\|P^1 - Q_T\|_Q^2 \ll T/p$ this gives

$$\chi^2(P_T\|Q_T) \ll_\varepsilon \frac{T}{p}\log^3(2p/T) + p^{-1/3+\varepsilon}T^{1/4}.$$

The squared norm is minimized by subtracting the measure's mass times Q_T, so centering at Q_T preserves this bound.

For $T < T_0$ apply the bound at T_0. Keeping T of the T_0 flips contracts the squared norm by conditional Cauchy–Schwarz. The bound at T_0 is $O_\varepsilon(p^{-5/24+\varepsilon})$, absorbing powers of $\log p$. This covers the shorter lengths in $R_\varepsilon(T,p)$, after enlarging the constants for the finitely many primes with $T_0 > p$.

Long observations. Since $F_1 = 1$, Lemma 3.2 gives $\Delta_p\sum_t F(t)^2 \ll 1$. Applying (1) to F yields $\sum_k P_T(k)^2/Q_T(k) \ll 1 + \sqrt{T/p}$. Jensen's inequality gives the logarithmic upper bound for $T \ge p$. Its matching lower bound is Lemma 2.2, since there are at most $4\sqrt{p}+1$ biases.

The threshold. The short-length bound tends to zero for $T = o(p)$. For $T \ge cp$ with fixed $c > 0$, detect the lattice by testing the count N_T with $\cos(\xi(N_T/T - 1/2))$, where $\xi = 2\pi/\Delta_p$. The binomial characteristic function, expanded at ξ/T, gives the complex expectation

$$e^{i\xi(f-1/2)}\exp\left(-\frac{\xi^2}{2T}f(1-f)\right) + O_c(p^{-1/2})$$

uniformly in f. Indeed $T(\xi/T)^3 = O_c(p^{-1/2})$. The first factor is 1 on the trace lattice, so the actual expectation of the cosine is at least $e^{-8\pi^2/c} + o(1)$. Against the reference, integrate the oscillation by parts. The product of $\sigma(f)$ and the damping factor has variation $O(1)$, so the reference expectation is $O(p^{-1/2})$. Pinsker's inequality for the bounded test gives a positive lower limit for the KL divergence for every $T \ge cp$. □

To apply the bound to restricted families, we allow omission of a fixed finite set J of primes. Write $F^{(J)}$ for the series for F with terms divisible by any prime in J omitted, and $F^{(\ell)}$ for $J = \{\ell\}$.

Lemma 3.4 (Restricted trace weights)**.** *Let A be a trace congruence class modulo a fixed integer M, and put $u = F^{(J)}1_A$. For $1 \le T \le p$,*

$$\|P^u - \gamma_B Q_T\|_Q^2 \ll_{M,J,\varepsilon} R_\varepsilon(T,p). \tag{3}$$

Here γ_B is the constant Fourier coefficient of the restricted weight after applying the smooth cutoff at $B = p^{2/3}\max(T,T_0)^{-1/4}$.

Proof. Write $U = 1_A F_{B,\psi}^{(J)}$ for the restricted cutoff weight. In the Mellin integral, restrict C_t to integers prime to J and multiply $L(s,\chi)$ by $\prod_{\ell\in J}(1-\chi(\ell)\ell^{-s})$. The extra factors are bounded on $\Re s = 1/2$, so $\|P^{u-U}\|_Q^2 \ll_{J,\varepsilon} p^{1/3+\varepsilon}/B$ as before. The same bounds for the cutoff on the trace lattice hold because $\sum_{n\le 2B}|b_t(n)|/n$ also bounds the restricted sum.

In the Fourier coefficients, the omission restricts the h_d sum to integers prime to J. Multiplication by 1_A shifts the frequencies by j/M. If a reduced denominator d becomes d', then $d \mid Md'$ and $d' \mid Md$, so $\tau(d)d^{-3/2} \ll_M \tau(d')d'^{-3/2}$. The constant coefficient γ_B is bounded because it only uses denominators dividing M. Decompose the centered measure as

$$P^u - \gamma_B Q_T = P^{u-U} + P^{U-\gamma_B} + \gamma_B(P^1 - Q_T).$$

Lemma 3.3 applies to the second term with denominator bound $2MB$. The last term has squared norm $O_M(T/p)$. With the chosen B these estimates give (3) for $T \ge T_0$, and for $T < T_0$ we keep the cutoff and contract from T_0. As in the proof of Theorem 3.1, this costs $O_{M,J,\varepsilon}(p^{-5/24+\varepsilon})$ after absorbing logarithms. □

4 Congruence classes and modular families

Smoothing gives prediction bounds under trace and torsion restrictions. The trace formula turns these into Hecke bounds.

4.1 Trace congruence classes

Fix $M \ge 1$ and a residue t_0 modulo M, and condition the curve on $t \equiv t_0 \pmod M$. For $t_0 = p+1$ these are the curves whose group order is divisible by M. Write P_T^M for the resulting count law. The biases now lie on a lattice of step $M\Delta_p$.

Proposition 4.1. *Let $p \ge p_0(M)$. The class has probability $C/M + o(1)$ as $p \to \infty$, where $C \ge (2/3)^{\omega(M)}$ depends only on M, t_0 and p modulo M, and $\omega(M)$ counts the distinct prime divisors of M. For $1 \le T \le p$, $\chi^2(P_T^M\|Q_T) \ll_{M,\varepsilon} R_\varepsilon(T,p)$. For $T \ge p$, $D_{\mathrm{KL}}(P_T^M\|Q_T) = \frac{1}{2}\log(T/p) + O_M(1)$. The logarithmic loss tends to zero if and only if $T = o(p)$.*

Proof. Let $\mathcal{A}$ be the set of traces in the class and $Z = \sum_{t\in\mathcal{A}}\Delta_p\sigma(f_t)F(t)$ its probability, so $P_T^M = P^{F1_\mathcal{A}}/Z$. In (3), the constant γ_B is $1/M$ times the sum of the modes of $F_{B,\psi}$ that are constant on $\mathcal{A}$, evaluated there. Therefore $|Z - \gamma_B|^2 \ll_{M,\varepsilon} R_\varepsilon(T,p)$. Let C/M be the constant obtained by completing the Fourier coefficients. Lemma 3.2 gives $|\gamma_B - C/M| \ll_M B^{-1/2}$. For $d \mid M$, the factor of $h_d(n)$ at primes outside M is that of $h_1(n)$. Its Euler product at $s = 1$ is one, by $\zeta(2s)/\zeta(s+1)$. The Chinese remainder theorem on the remaining moduli gives

$$C = \prod_{\ell\mid M}\left((1+\ell^{-1})^{-1}\sum_{j\ge 0}\frac{\rho_{t_0}(\ell^{\min(j,v_\ell(M))})}{\ell^j}\right).$$

Indeed the trace class averages the local root series, whose Liouville factor is $(1+\ell^{-1})^{-1}$. For $j > v = v_\ell(M)$, each unit root modulo ℓ^v has ℓ^{j-v} lifts x, each requiring the single trace $x+p/x \bmod \ell^j$. The conditional mean is therefore $\rho_{t_0}(\ell^v)$. The $j=0$ term gives $C \ge (2/3)^{\omega(M)}$, and the limited moduli give the claimed dependence on p.

The mass Z is independent of T, so taking $T = \lceil p^{9/10} \rceil$ gives $Z = C/M + o(1)$. For the chi-square bound, center (3) at ZQ_T and divide by Z^2. Applying (1) to $F1_{\mathcal{A}}$ gives $\sum_k P_T^M(k)^2/Q_T(k) \ll_M 1+\sqrt{T/p}$, hence the upper bound for $T \ge p$ by Jensen's inequality. With at most $4\sqrt{p}/M+1$ biases, Lemma 2.2 gives $\frac{1}{2}\log(TM^2/p) - O(1)$ for $p \ge M^2$. Finally, for $T \ge cp$ test the count with $\cos(\xi(N_T/T - 1/2) - 2\pi t_0/M)$ and $\xi = 2\pi/(M\Delta_p)$. On the class $\xi(f_t - 1/2) = 2\pi t/M$ is constant modulo 2π. The cosine test in Theorem 3.1 then gives a positive lower limit for the loss. □

4.2 The modular family and Hecke traces

Fix an odd prime $\ell \ne p$. For $\ell \ge 5$, let $P_T^{Y_1}$ be the count law obtained by sampling uniformly in $Y_1(\ell)(\mathbf{F}_p)$ and using the trace of the corresponding elliptic curve for the bias. Write $N_\ell(p) = \#Y_1(\ell)(\mathbf{F}_p)$. For $\ell = 3$ give each isomorphism class of pairs (E,P) weight $1/|\operatorname{Aut}(E,P)|$ and let $N_3(p)$ be their total weight. We use $P_T^{Y_1}$ only when this family is nonempty.

Proposition 4.2 (Prediction on $Y_1(\ell)$)**.** *For every fixed odd prime ℓ and all sufficiently large $p \ne \ell$,*

$$\chi^2(P_T^{Y_1} \| Q_T) \ll_{\ell,\varepsilon} R_\varepsilon(T,p) \qquad (1 \le T \le p).$$

As $p \to \infty$, the loss on $Y_1(\ell)$ tends to zero exactly when $T = o(p)$, and for $T \ge p$ it is $\frac{1}{2}\log(T/p) + O_\ell(1)$.

Proof. Let C_ℓ be the event $\ell \mid \#E(\mathbf{F}_p)$ and D_ℓ the event that all ℓ-torsion is rational. The number of points of order ℓ is $(\ell-1)(1_{C_\ell} + \ell 1_{D_\ell})$. A nontrivial automorphism α of E fixes only $\ker(1-\alpha)$, of order at most four, so for $\ell \ge 5$ the automorphisms act freely on the points of order ℓ and every pair (E,P) has trivial automorphism group. For $\ell = 3$ the law $P_T^{Y_1}$ is by definition the normalized law with the weights $1/|\operatorname{Aut}(E,P)|$. By orbit-stabilizer, the normalized $Y_1(\ell)$ trace law is the Weierstrass law reweighted by $1_{C_\ell} + \ell 1_{D_\ell}$. Indeed the weight of a Weierstrass class is $1/(p|\operatorname{Aut}(E)|)$. If $Z = \Pr(C_\ell) + \ell \Pr(D_\ell)$, then $N_\ell(p) = p(\ell-1)Z$. Proposition 4.1 gives $Z \gg_\ell 1$, and $Z \le 1+\ell$. When $p \not\equiv 1 \pmod{\ell}$, the event D_ℓ is empty and the congruence-class estimate applies directly.

When all ℓ-torsion is rational. For $p \equiv 1 \pmod{\ell}$, the group order is divisible by ℓ on $A = \{t : t \equiv p+1 \bmod \ell\}$. Rationality of all the ℓ-torsion can occur only on $B = \{t : t \equiv p+1 \bmod \ell^2\}$. The mass of curves with all ℓ-torsion rational at trace t is $H(D_t/\ell^2)/(2p)$ on B by [20, Lemma 4]. These traces are ordinary since $t \equiv 2 \pmod{\ell}$ and the only supersingular trace over $\mathbf{F}_p$ is zero. For $t \in B$ compare the root counts of $x^2 - tx + p$ with those of $y^2 - ((t-2)/\ell)y + (p+1-t)/\ell^2$, whose discriminant is $-D_t/\ell^2$. The substitution $x = 1 + \ell y$ identifies the root counts at moduli prime to ℓ. At ℓ the old count is one, and at ℓ^{j+2} it is ℓ times the new count at ℓ^j. The Liouville convolution multiplies each local root series at $s=1$ by $(1+1/\ell)^{-1}$. Hence the old and new local factors satisfy $L_\ell = 1 + L'_\ell/\ell$. Zagier's identity holds for every nonzero discriminant, so $F'(t) = \pi H(D_t/\ell^2)/\sqrt{D_t/\ell^2}$ is the value at $s=1$ of the series for the new quadratic. Since $b_t = c_t * \chi$ with c_t multiplicative, $F = L_\ell F^{(\ell)}$ and $F' = L'_\ell F^{(\ell)}$ with the same $F^{(\ell)}$. Therefore $F - F^{(\ell)} = F'/\ell = \pi H(D_t/\ell^2)/\sqrt{D_t}$ on B. The resulting unnormalized Y_1 weight is

$$U(t) = F(t)1_A(t) + \ell\big(F(t) - F^{(\ell)}(t)\big)1_B(t). \tag{4}$$

For each of the three terms in (4), (3) bounds the squared norm of the associated count measure after subtracting its mass times Q_T. Combine the three bounds by the triangle inequality and

normalize by Z. The local factors are positive and $0 \le F^{(\ell)} \le F$ on B, so $0 \le U \le (1+\ell)F$. Equation (1) gives the upper bound for long observations. The finite-support bound and the lattice cosine test give the lower estimates just as for P_T. □

For $k \ge 2$ set

$$\tau_k = p^{-(k-1)/2}\,\mathrm{tr}(T_p \mid S_k(SL_2(\mathbf{Z}))), \qquad \tau_k^{(\ell)} = p^{-(k-1)/2}\,\mathrm{tr}(T_p \mid S_k(\Gamma_1(\ell))).$$

Corollary 4.3. *For $1 \le T \le p$,*

$$\sum_{\substack{4\le k\le T+2\\ k \text{ even}}} \kappa_{k-2}(T)\big(\tau_k + p^{-(k-1)/2}\big)^2 = p\,\chi^2(P_T\|Q_T) \ll_\varepsilon pR_\varepsilon(T,p).$$

For every p with $N_\ell(p) > 0$ and all $T \ge 1$, the prime-level identity is

$$\sum_{k=3}^{T+2} \kappa_{k-2}(T)\big(\tau_k^{(\ell)} + (\ell-1)c_k p^{-(k-1)/2}\big)^2 = \frac{N_\ell(p)^2}{p}\,\chi^2(P_T^{Y_1}\|Q_T),$$

where $c_k = (1 + 1_{p\equiv 1\,(\ell)} + (-1)^k 1_{p\equiv -1\,(\ell)})/2$. In particular, for $2 \le K \le \sqrt{p}$,

$$\sum_{\substack{4\le k\le K\\ k \text{ even}}} |\tau_k|^2,\ \sum_{3\le k\le K} |\tau_k^{(\ell)}|^2 \ll_{\ell,\varepsilon} K^2\log^3(2p/K^2) + p^{2/3+\varepsilon}\max(K,p^{1/4})^{1/2}. \tag{5}$$

The constant for the level-one sum is independent of ℓ.

Proof. Level one. The Birch–Ihara formula [20, Theorems 2 and 3] gives the degree-$(k-2)$ Chebyshev moment $-p^{-1/2}(\tau_k + p^{-(k-1)/2})$ for even $k \ge 4$. The odd moments vanish by quadratic twisting, so Proposition 2.1 gives the first identity.

The prime-level trace identity. Let C_ℓ and D_ℓ be the divisibility and full-torsion events in the proof of Proposition 4.2. Use [20, Theorem 3]. In its notation, when $p \equiv 1 \pmod\ell$, multiplying the unnormalized C_ℓ and D_ℓ moments by p gives $T_{\ell,1}/(\ell-1) - \ell T_{\ell,\ell}$ and $T_{\ell,\ell}$, respectively, with $T_{a,b} = T_{a,b}(p,1)$. Thus the full-level trace cancels in the moment with weight $1_{C_\ell} + \ell 1_{D_\ell}$. For $p \not\equiv 1 \pmod\ell$ the C_ℓ moment alone gives $T_{\ell,1}/(\ell-1)$. In both cases, for $k \ge 3$, the prime-index formula for $T_{\ell,1}$ is $-\mathrm{tr}(T_p \mid S_k(\Gamma_1(\ell))) - (\ell-1)c_k$. Dividing by the mass and replacing $U_{k-2}(t,p)$ by $p^{(k-2)/2}U_{k-2}(t/(2\sqrt{p}))$ gives the normalized Chebyshev moment $-\sqrt{p}\,(\tau_k^{(\ell)} + (\ell-1)c_k p^{-(k-1)/2})/N_\ell(p)$. Apply Proposition 2.1.

Bounding the sum up to weight K. Take $T = \lfloor K^2 \rfloor$ in the two trace identities. For the weights retained in (5), $\kappa_{k-2}(T) \gg 1$. Apply Theorem 3.1 and Proposition 4.2 on the right sides of the identities. Removing the explicit corrections by the triangle inequality costs $O_\ell(1)$ in the squared bound. □

Corollary 4.4 (Hecke trace bounds)**.** *For $K = p^a$ with $4/15 < a < 1/2$, (5) saves a power of p over the bounds from Deligne, Petrow and the angular large sieve. At $K = p^{2/5}$ it gives $O_{\ell,\varepsilon}(p^{13/15+\varepsilon})$, compared with $O_\ell(p)$ from the large sieve, the strongest of these three benchmarks at that weight.*

Proof. The dimension bound and Deligne give $O_\ell(K^3)$. For any fixed small $\eta > 0$, Petrow [33, Theorems 1.1–1.2] gives $|\tau_k|, |\tau_k^{(\ell)}| \ll_{\ell,\eta} p^{1/44+\eta}k^{61/66}$ for $k \ge p^{3/10+\eta}$. Sum its square and use Deligne below that cutoff to obtain $O_{\ell,\varepsilon}(p^{1/22+\varepsilon}K^{94/33})$ in the range of Petrow's estimate. The trace angles have spacing $\gg p^{-1/2}$ and $\sum_t F(t)^2 \ll \sqrt{p}$, so the angular large sieve gives $O_\ell(p)$ for $K \le \sqrt{p}$. At prime level, use U from (4), or $F1_A$ if there is no full rational torsion. Comparing exponents in (5) gives the claimed range. The K^2 term with its logarithm dominates for $K \ge p^{4/9+\eta}$ with any fixed $\eta > 0$ and sufficiently small ε. At $K \asymp \sqrt{p}$ both (5) and the large sieve are $O_\ell(p)$. □

A 2-isogeny identifies the Legendre trace law with that of curves equipped with a rational point of order four, so the same smoothing bound applies.

Corollary 4.5 (Legendre family)**.** *For $p \geq 5$ sample λ uniformly in $\mathbf{F}_p \setminus \{0,1\}$ and use the trace of $y^2 = x(x-1)(x-\lambda)$ for the bias. Its count law P_T^{Leg} satisfies $\chi^2(P_T^{\mathrm{Leg}} \| Q_T) \ll_\varepsilon R_\varepsilon(T,p)$ for $1 \leq T \leq p$. The logarithmic loss tends to zero exactly when $T = o(p)$, and equals $\frac{1}{2}\log(T/p) + O(1)$ for $T \geq p$.*

Proof. The universal $Y_1(4)$ curve $y^2 = (x+s)(x^2+x+s)$ has a point $(0,s)$ of order four. Quotienting by the subgroup generated by its double gives the Legendre curve with $\lambda = 1 - 4s$ [22, Section 2]. The isogeny preserves the trace, and the parameters are in bijection. By the Tate normal form, $s \mapsto (E_s, (0,s))$ is a bijection onto the $\mathbf{F}_p$-isomorphism classes of pairs of a curve and a point of order four, and automorphisms act freely on such points. Thus the Legendre law is obtained by weighting each Weierstrass curve by its number of rational points of order four and dividing by the mean $(p-2)/p$. Writing $n = \#E(\mathbf{F}_p)$ and $A_j = \{E[j] \subset E(\mathbf{F}_p)\}$, this number is

$$2(1_{4|n} - 1_{A_2}) + 4\,1_{A_2}1_{8|n} + 8\,1_{A_4}. \tag{6}$$

When the 2-torsion is rational, the 2-part of the group is $\mathbf{Z}/2^a \times \mathbf{Z}/2^b$ with $a \geq b \geq 1$, and $8 \mid n$ exactly when $a \geq 2$. Indeed, the rational 4-torsion is either cyclic, $(\mathbf{Z}/2)^2$, $\mathbf{Z}/4 \times \mathbf{Z}/2$, or $(\mathbf{Z}/4)^2$.

At trace t, the mass of A_2 is $H(D_t/4)/(2p)$ when $4 \mid p+1-t$, and that of A_4 is $H(D_t/16)/(2p)$ when $p \equiv 1 \pmod 4$ and $16 \mid p+1-t$, with zero otherwise [20, Lemmas 3–4]. This includes $t = 0$ for A_2. It occurs only for $p \equiv 3 \pmod 4$, when $H(p) = h_w(-p)$ in their notation. On the first progression the substitution $x = 1 + 2y$ used in (4) gives the local relation $L_0 = 1 + L_1/2$. On the second, the new quadratic is y^2 modulo 2 and one can divide again with $y = 2z$, giving $L_1 = 1 + L_2/2$. Hence the two masses on their respective progressions are $\Delta_p\sigma(f_t)(F - F^{(2)})(t)$ and $\Delta_p\sigma(f_t)(F - \frac{3}{2}F^{(2)})(t)$. The bounds for the restricted weights F and $F^{(2)}$ follow from Lemma 3.4, with $J = \varnothing$ or $J = \{2\}$. Subtract the mass times Q_T from each of these measures and combine their bounds to obtain the short-length estimate. The order-four weight is at most twelve, so (1) gives the upper bound for long observations. The finite-support bound and the lattice test give the lower estimates. □

5 The critical profile

We derive the limiting likelihood ratio at $T/p \to c > 0$ and use it to compute both divergences. We then let $c \downarrow 0$ and extend the fixed-c formulas to the restricted families.

5.1 The profile at the trace-lattice scale

At $T \sim cp$ the sampling error is on the scale of the bias spacing, so neighboring traces contribute to the same count. Take a subsequence of primes on which $p \bmod d$ eventually stabilizes for each fixed d, with unit residue system A. For a reduced fraction $a/d \pmod 1$ put

$$\gamma_A(a/d) = \frac{S(a, aA; d)}{d^2} \prod_{\ell \mid d} (1 - \ell^{-2})^{-1}, \qquad \gamma_A(0) = 1,$$

where aA uses the residue of A modulo d. Extend γ_A to $\mathbf{Q}$ by periodicity, and let μ_A put mass $|\gamma_A(\beta)|^2$ at $|\beta|$ for each nonzero rational β.

Let R record the residues of the integer trace nearest the observed proportion of ones on the bias scale. In the limit, R is uniform on $\widehat{\mathbf{Z}} = \prod_\ell \mathbf{Z}_\ell$. For each prime ℓ let $\rho_{A,R}(\ell^j)$ count the roots of $x^2 - Rx + A$ modulo ℓ^j, with $\rho_{A,R}(1) = 1$. Set

$$L_{\ell,A}(R) = \frac{1}{1+\ell^{-1}} \sum_{j \ge 0} \frac{\rho_{A,R}(\ell^j)}{\ell^j}, \qquad F_A(R) = \prod_\ell L_{\ell,A}(R).$$

The product converges in L^2. Take Y with density σ and U uniform on $[-1/2, 1/2)$, independently of each other and R. The binomial variance of the proportion of ones in squared lattice steps tends to $v = 16Y(1-Y)/c$. Let ϕ_v be the centered normal density of variance v and set

$$B_{A,c}(Y, R, U) = \sum_{j \in \mathbf{Z}} F_A(R + j)\phi_v(U - j), \qquad v = 16Y(1-Y)/c.$$

Theorem 5.1. *Suppose $T/p \to c > 0$ along a sequence of primes whose residues stabilize modulo every fixed integer, and let A be the resulting system of unit residues. Then the expected excess log loss satisfies*

$$D_{\mathrm{KL}}(P_T \| Q_T) \longrightarrow \mathcal{K}_A(c) = \mathbb{E}\big[B_{A,c} \log B_{A,c}\big]. \tag{7}$$

The corresponding chi-square limit is

$$\chi^2(P_T \| Q_T) \longrightarrow \mathcal{C}_A(c) = \int_0^1 \sigma(y) \sum_{\substack{\beta \in \mathbf{Q} \\ \beta \ne 0}} |\gamma_A(\beta)|^2 e^{-64\pi^2 y(1-y)\beta^2/c}\, dy. \tag{8}$$

Both limits are finite and positive. As $c \to \infty$ the limiting logarithmic loss satisfies

$$\mathcal{K}_A(c) = \tfrac{1}{2}\log c - 1 - \tfrac{1}{2}\log(2\pi) + \mathbb{E}[F_A \log F_A] + o(1). \tag{9}$$

If $A \equiv 1 \pmod 3$ and $A' \equiv 2 \pmod 3$, then $\mathcal{C}_A$ and $\mathcal{C}_{A'}$ are distinct. If their components away from 3 coincide, then $\mathcal{K}_{A'}(c) > \mathcal{K}_A(c)$ for all sufficiently large c.

Gekeler [13] derived the local trace weights, also described in [9, Theorem 1.1]. Ma [29, Theorem 1.3, Corollary 1.4 and Lemma 7.1] gives subsequence laws for the isogeny-class size at a uniformly sampled trace. Its local factors depend on (p/ℓ) at odd ℓ and on $p \bmod 8$ at 2. Here binomial sampling combines the joint weights of neighboring traces. Related smoothing appears in channel capacity for equally spaced priors [1, Sections 3.1 and 4] and in Lee's preprint for one curve across primes under symmetric-power GRH [27, Theorem 1.4].

Proof. The limiting weights. Each local series converges uniformly by the quadratic root bound $\rho_{A,R}(\ell^j) \ll_\ell \ell^{j/2}$. For $j \ge 1$ the mean of $\rho_{A,R}(\ell^j)$ is $\varphi(\ell^j)/\ell^j$: each unit x determines the single trace $x + A/x$. Thus $\mathbb{E}L_{\ell,A} = 1$. For a primitive frequency a/ℓ^k, the Fourier coefficient of $\rho_{A,R}(\ell^j)$ is zero when $j < k$ and is $\ell^{-k}S(a, aA; \ell^k)$ when $j \ge k$, by the same trace parametrization. Summing in j gives the coefficient $\ell^{-2k}(1-\ell^{-2})^{-1}S(a, aA; \ell^k)$. Weil's estimate and Parseval give $\mathrm{Var}(L_{\ell,A}) \ll \sum_{k \ge 1}(k+1)^2\ell^{-2k} \ll \ell^{-2}$. The local weights are independent and nonnegative. Their finite products are Cauchy in L^2 because their second moments are products of $1 + \mathrm{Var}(L_{\ell,A})$. For coprime moduli d_1, d_2 one has $a/(d_1d_2) \equiv a\bar{d}_2/d_1 + a\bar{d}_1/d_2 \pmod 1$ and the twisted multiplicativity $S(m, n; d_1d_2) = S(m\bar{d}_2, n\bar{d}_2; d_1)S(m\bar{d}_1, n\bar{d}_1; d_2)$. Hence the local coefficients multiply, and the Fourier coefficients of F_A are γ_A.

The likelihood ratio. For fixed N the truncated weight F_N has period $M = \operatorname{lcm}(1, \dots, N)$ and eventually depends on A alone. By Lemma 3.2 and (1),

$$\|P^F - P^{F_N}\|_Q^2 \ll_{c,\varepsilon} \log^4(2N)/N + p^{-1/15+\varepsilon}.$$

It therefore suffices to compute the limit for F_N, with N fixed, and then let N tend to infinity.

For $y = k/T$ in a fixed compact subinterval of $(0, 1)$, set $u = 4\sqrt{p}(y - \frac{1}{2})$ and $v = 16y(1 - y)/c$. Uniformly in this range,

$$\frac{P^{F_N}(k)}{Q_T(k)} = B_N(y, u) + o(1), \qquad B_N(y, u) = \sum_{m \in \mathbf{Z}} F_N(m)\phi_v(u - m).$$

Indeed, the exact ratio is $\Delta_p \sum_m F_N(m) g_{T,k}(f_m)$ over the trace grid, where $g_{T,k}$ is the beta posterior density. Stirling's formula and a Taylor expansion at y give the Gaussian approximation to $\Delta_p g_{T,k}(y + \Delta_p s)$ for bounded s, with variance $16(p/T)y(1-y)$. For all s in the grid, the ratio $g_{T,k}(y+\Delta_p s)/g_{T,k}(y)$ is bounded by a constant times $\exp(-2T\Delta_p^2 s^2)$. This follows by writing its exponential factor as $\exp(-TD(\mathrm{Ber}(y) \| \mathrm{Ber}(y + \Delta_p s)))$ and applying Pinsker's inequality. The bound allows summation over m and replacement of $16(p/T)$ by $16/c$.

On $\eta \le y \le 1 - \eta$, the reference count mass is $Q_T(k) = (1 + o(1))\sigma(y)/T$. The Riemann sum for $\sigma(y)|B_N(y, u) - 1|^2$ has error $O_{N,c,\eta}(\sqrt{p}/T)$, as $du/dy = 4\sqrt{p}$ and the summand is smooth and periodic in u. Partitioning the integral into periods of length $M/(4\sqrt{p})$ in y and freezing y within each period gives the uniform average over $0 \le u < M$. The same averaging with bounded Lipschitz tests gives convergence in distribution of $P^{F_N}(K)/Q_T(K)$ for $K \sim Q_T$. The limiting variable is $B_N(Y, r_N + U)$, where r_N is any representative of R modulo M. Periodicity makes the choice of representative immaterial. The omitted probability near the endpoints tends to zero as $\eta \to 0$. For convergence of the squared norm, use $|P^{F_N}(k)/Q_T(k)| \ll_N 1 + \Delta_p m_{T,k}$. The bounds for $Q_T(k)$ and $m_{T,k}$ give $O_{N,c}(\sqrt{\eta} + T^{-1/2})$ for the omitted part of the squared norm. The continuous end integral is $O_{N,c}(\sqrt{\eta})$, since $|B_N(y, u)| \ll_N 1 + v^{-1/2}$. We obtain

$$\lim_{p \to \infty} \|P^{F_N} - Q_T\|_Q^2 = \int_0^1 \sigma(y) \, \frac{1}{M} \int_0^M |B_N(y, u) - 1|^2 \, du \, dy.$$

Write $\gamma_{N,A}$ for the Fourier coefficients of F_N on its integer period, extended periodically and set to zero off $M^{-1}\mathbf{Z}$. The Dirichlet series for h_d at $s = 1$ gives $\gamma_{N,A}(a/d) \to \gamma_A(a/d)$. The bound $|\gamma_{N,A}(a/d)| \ll \tau(d)d^{-3/2}$ and $\sum_d \varphi(d)\tau(d)^2 d^{-3} < \infty$ give $F_N \to F_A$ in $L^2(\widehat{\mathbf{Z}})$. Put $Z_v(u) = \sum_j \phi_v(u - j)$. By Cauchy–Schwarz and Haar invariance, smoothing any weight W has squared L^2 norm in (R, U) at most $\|W\|_2^2 \int_{-1/2}^{1/2} Z_v(u)^2 \, du$. The latter integral is $O(1 + v^{-1/2})$, which is integrable in Y. Writing $m = r_N + j$ in $B_N(Y, r_N + U)$ gives convergence in L^2 to $B_{A,c}$. The initial truncation estimate then gives convergence in distribution of $P_T(K)/Q_T(K)$ to $B_{A,c}$. These nonnegative variables have uniformly bounded second moment by (1). Since $x \log x = o(x^2)$ at infinity and extends continuously at zero, expectations converge, proving (7). The limit is finite since $\mathbb{E}B_{A,c}^2 \le \|F_A\|_2^2 \, \mathbb{E} \int Z_v^2 < \infty$. The posterior mean is $(k + 3/2)/(T + 3)$ and differs from k/T by $O(p^{-1/2})$ lattice steps, which does not affect the limit.

The chi-square limit. Poisson summation gives the coefficient of $e(\beta u)$ in B_N as $\gamma_{N,A}(\beta)e^{-2\pi^2 v\beta^2}$. Thus Parseval gives the sum in (8), with $\gamma_{N,A}$ in place of γ_A and the additional zero-mode term $|\gamma_{N,A}(0) - 1|^2$. In particular the nonzero integer frequencies occur. The Gaussian sum over integer translates of each reduced frequency is $O(1+v^{-1/2})$, so the same summable coefficient bound permits passage to the limit. The initial truncation bound proves (8). Integer frequencies ensure its positivity.

The logarithmic loss for large c. For fixed y write $B_{A,c} = Z_v(U)\overline{F}_v(R,U)$, so $\overline{F}_v$ is the average of $F_A(R+j)$ with weights $\phi_v(U-j)/Z_v(U)$. As $\mathbb{E}_R\overline{F}_v = 1$, the entropy splits into $\int Z_v \log Z_v$ and $\int Z_v\, \mathbb{E}_R[\overline{F}_v \log \overline{F}_v]$. The first is positive since the wrapped Gaussian has mean one and a nonzero first Fourier coefficient. The second is nonnegative by Jensen. In particular $\mathcal{K}_A(c) > 0$ for every $c > 0$.

If G_v is a centered normal variable of variance v, then its reduction to $[-1/2, 1/2)$ has density Z_v and its nearest integer has entropy $o(1)$ as $v \to 0$, by Gaussian tails. The entropy chain rule therefore gives $\int Z_v \log Z_v = -\frac{1}{2}\log(2\pi e v) + o(1)$. By Jensen and translation invariance,

$$\int_{-1/2}^{1/2} Z_v(u)\, \mathbb{E}_R|\overline{F}_v(R,u) - F_A(R)|^2\, du \le 4\|F_A\|_2^2 \Pr(|G_v| \ge 1/2).$$

By Jensen, $\mathbb{E}_R\overline{F}_v^2 \le \|F_A\|_2^2$ for every u, so the variables $\overline{F}_v \log \overline{F}_v$ are uniformly integrable. With the L^2 bound just displayed, the second term tends to $\mathbb{E}[F_A \log F_A]$. Both errors vanish uniformly for $v \le 4/c$ as $c \to \infty$. The beta law gives $\mathbb{E}\log(Y(1-Y)) = 1 - 4\log 2$. Averaging proves (9).

Dependence on the prime residues. For $1/2 < s < 1$, applying the Mellin transform in $x = 1/c$ gives

$$\int_0^\infty \mathcal{C}_A(1/x) x^{s-1}\, dx = \frac{\Gamma(s)}{(64\pi^2)^s} \left(\int_0^1 \frac{\sigma(y)}{(y(1-y))^s}\, dy \right) \int_0^\infty r^{-2s}\, d\mu_A(r).$$

The spectral integral converges: near zero it is bounded by a constant times $\sum_d \tau(d)^2 d^{2s-3}$, and at infinity the integer translates are summable. Equality of two profiles would give equality of the last integrals on the whole strip, by analytic continuation. On a vertical line $s = s_0 + it$, these are Fourier transforms at frequency $2t$ of the finite measures $r^{-2s_0}\, d\mu_A(r)$ in the variable $\log r$. Uniqueness would force equality of the spectral measures. But $S(1, A; 3)$ is -1 for $A \equiv 1$ and 2 for $A \equiv 2$, so the atom at $r = 1/3$ has mass $1/32$ and $1/8$, respectively.

For the logarithmic loss, independence gives $\mathbb{E}[F_A \log F_A] = \sum_\ell \mathbb{E}[L_{\ell,A} \log L_{\ell,A}]$, first for finite products and then by L^2 convergence. For A and A' agreeing away from 3, the difference is thus the difference of their local entropies at 3. For either residue system $3/4 \le L_3 \le 3/2$. Indeed, if h is the valuation of $R^2 - 4A$, the number of roots modulo 3^j is $3^{\lfloor j/2 \rfloor}$ through $j = h$. Beyond h the counts vanish unless $h = 2e$ and the remaining unit is a square. In that case the count $2 \cdot 3^e$ gives the same weighted tail as for discriminant zero, which attains $3/2$.

For $A' \equiv 2 \pmod 3$, the factor takes only the endpoints, with mean one and entropy $\frac{1}{2}\log(9/8)$. For $A \equiv 1 \pmod 3$, the valuation equals one on $4/9$ of the residues modulo 9, where $L_{3,A} = 1$. The chord of $x \log x$ between $3/4$ and $3/2$ thus gives a gap

$$\lim_{c\to\infty} (\mathcal{K}_{A'}(c) - \mathcal{K}_A(c)) \ge \tfrac{2}{9}\log(9/8) > 0.$$

Dirichlet's theorem realizes both compatible residue systems by successively choosing primes in their residue classes modulo $j!$. □

5.2 The short-length limit of the Fourier profile

For each fixed $0 < y < 1$, the Gaussian in (8) suppresses frequencies larger than order $\sqrt{c}$. To obtain the leading term as $c \downarrow 0$ we estimate the mass of μ_A near zero. Choosing A_ℓ independently and uniformly in $\mathbf{Z}_\ell^\times$ gives Haar probability on $\widehat{\mathbf{Z}}^\times$.

Theorem 5.2. *For Haar-almost every $A \in \widehat{\mathbf{Z}}^\times$,*

$$\mu_A((0,x)) \sim \zeta(2) x^2 \quad (x \downarrow 0), \qquad \mathcal{C}_A(c) \sim \frac{c}{48} \quad (c \downarrow 0).$$

Proof. Local moments. For $(a,d)=1$ put $Y_{a,d}=d^3|\gamma_A(a/d)|^2$. The change of variable in the Kloosterman sum gives $S(a,aA;d)=S(1,a^2A;d)$. The Chinese remainder theorem factors $Y_{a,d}$ over the primes of d, with unit twists in the numerators. Its expectation is a multiplicative function $g(d)$ independent of a. At a prime ℓ, additive orthogonality gives

$$g(\ell)=(1-\ell^{-2})^{-2}\left(1-\frac{1}{\ell(\ell-1)}\right),\qquad g(\ell^k)=(1-\ell^{-2})^{-2}\quad(k\ge 2).$$

Indeed, summing $|S(1,b;\ell^k)|^2$ over all b gives $\ell^k\varphi(\ell^k)$. At $k=1$ the term $b=0$ contributes one. For $k\ge 2$ the terms with $\ell\mid b$ vanish by summing over the translates of a unit u by multiples of ℓ^{k-1}. Thus g is bounded and $g*\mu$ is absolutely summable.

We also need the mean at one prime with A_ℓ held fixed. Define $w_A(d)=d^2\sum^*_{a \bmod d}|\gamma_A(a/d)|^2$, which is multiplicative by the same factorization. Then

$$\sum_{k\ge 1} w_A(\ell^k)\ell^{-k}=\frac{\ell}{\ell^2-1}. \tag{10}$$

To see this, let N_k count unit pairs (u,v) modulo ℓ^k with $(u-v)(uv-A_\ell)\equiv 0 \pmod{\ell^k}$. Opening the square gives $\ell^kN_k-\ell^{k+1}N_{k-1}$ for the sum of $|S(a,aA;\ell^k)|^2$ over unit a when $k\ge 2$, and gives $\ell N_1-(\ell-1)^2$ when $k=1$. With normalized Haar measure on $\mathbf{Z}_\ell^2$, $\sum_{k\ge1}N_k\ell^{-2k}$ is the integral of $v_\ell(u-v)+v_\ell(uv-A_\ell)$ over the unit pairs. For each fixed unit u, either valuation is at least $j\ge1$ on a set of v of measure ℓ^{-j}. This gives $\sum_{k\ge1}N_k\ell^{-2k}=2/\ell$. Summing the square identities with weight ℓ^{-3k} and multiplying by $(1-\ell^{-2})^{-2}$ proves (10).

The mean with small-prime residues fixed. Write

$$Z_x=\frac{\mu_A((0,x))}{x^2}=\frac{2}{x^2}\sum_{a\ge1}\sum_{\substack{d>a/x\\(a,d)=1}}\frac{Y_{a,d}}{d^3}.$$

When d is of order $1/x$, only a bounded number of numerators contribute. We use the mean over denominators for each such numerator. Let $\mathbb{E}_L$ denote conditional expectation with the coordinates at primes $\ell\le L$ held fixed, and put $P=\prod_{\ell\le L}\ell$. Write $d=rm$, with r composed of primes dividing P and $(m,P)=1$. For fixed a and r with $(r,a)=1$, the conditional mean of $Y_{a,rm}$ is $g(m)$ times $Y_{a\bar m,r}$, where $\bar m$ is taken modulo r. The function $g(m)1_{(m,aP)=1}$ has mean

$$M(aP)=\prod_{\ell\mid aP}(1-\ell^{-1})\prod_{\ell\nmid aP}\left((1-\ell^{-1})\sum_{k\ge0}\frac{g(\ell^k)}{\ell^k}\right).$$

In each unit class modulo r, its sum up to D is $DM(aP)/\varphi(r)+O(\tau(aP))$. For this estimate expand $g=1*(g*\mu)$, then use inclusion–exclusion at the primes dividing aP that do not divide r. Absolute summability of $g*\mu$ controls the sum of the errors.

Partial summation above $m=a/(xr)$ and averaging $Y_{a\bar m,r}$ over the unit classes gives the main term $M(aP)w_A(r)/(a^2\varphi(r))$ in $\mathbb{E}_LZ_x$. For $r\le x^{-1/8}$ the sum of the errors is

$$\ll x\,\tau(P)\sum_{r\le x^{-1/8}}r\tau(r)^2\sum_{a\ge1}\frac{\tau(a)}{a^3}.$$

Here the coefficient bound for a unit numerator is $Y_{a,r}\ll\tau(r)^2$. For larger r, both the sum and its main term are bounded by a constant times $\sum_{r>x^{-1/8},\,r\mid P^\infty}\tau(r)^2/r$. This is at most $x^{1/16}\exp(O(\sqrt{L}))$,

by inserting $r^{1/2}$ and summing the product over $\ell \mid P$. Taking $L = \frac{1}{10}\log(1/x)$, both errors are $O(x^{1/32})$.

The main terms over a and r form an Euler product. Whether $\ell \le L$ or $\ell > L$, its factor is

$$(1-\ell^{-1})\left(1+\frac{1}{\ell^2-1}\right)+\frac{\ell}{\ell^2-1} = (1-\ell^{-2})^{-1}.$$

At a prime held fixed this uses (10). At the other primes one uses its expectation. Consequently $\mathbb{E}_L Z_x = \zeta(2) + O(x^{1/32})$ uniformly in the fixed coordinates.

Fluctuations from the remaining primes. In the fourth moment of $Z_x - \mathbb{E}_L Z_x$, consider four denominators $d_1, \dots, d_4$. If one denominator shares no prime $\ell > L$ with the other three, its centered factor is independent of theirs once the small-prime residues are fixed. Its conditional mean is zero, so the term vanishes. The full Haar expectation of the absolute value of the product of the four centered weights is bounded by

$$\ll \prod_{i<j} \tau\big((d_i, d_j)\big)^2.$$

To obtain this bound, expand the four differences. At each prime use the mean of the factor with greatest exponent and bound the others by Weil's bound. The mean of $|S|^2/\ell^k$ is at most one, and its maximum is at most $(k+1)^2$. Factors already averaged by $\mathbb{E}_L$ satisfy the same bounds. The bound is used only after the outer expectation over the small-prime residues; it need not hold conditionally on them. The normalizing factors $(1-\ell^{-2})^{-2}$ have bounded product.

For the remaining tuples choose at most four distinct primes $\ell_s > L$ and sets $I_s \subset \{1,2,3,4\}$ such that $|I_s| \ge 2$, their union is all four indices, and $\ell_s \mid d_i$ for $i \in I_s$. Expand $\tau(n)^2 = \sum_{h\mid n}(\tau^2 * \mu)(h)$, whose coefficients at ℓ^k are $2k+1$ for $k \ge 1$. If D_i is the least common multiple of the divisibility conditions imposed on d_i, then

$$\sum_{\substack{d_i > a_i/x \\ D_i \mid d_i}} d_i^{-3} \ll \frac{x^2}{D_i a_i^2}.$$

After dividing by x^8 and summing the a_i, the remaining divisor sums are an Euler product. At a prime let k be the largest exponent in the six divisors for the pairs $i<j$. If $k \ge 1$, the exponents in D_i sum to at least $2k$. At a chosen prime with set I the sum is at least $|I| + 2(k-1)$. Summing the polynomial divisor weights therefore gives $1 + O(\ell^{-2})$ at unchosen primes and $O(\ell^{-|I|})$ at chosen primes. A single chosen prime must divide all four denominators, while two or more give at least two factors $\sum_{\ell > L} \ell^{-2}$. Thus

$$\mathbb{E}\big|Z_x - \mathbb{E}_L Z_x\big|^4 \ll L^{-2}.$$

At $x_n = \exp(-n^{3/4})$ the fourth-moment bounds are summable, so Markov's inequality and Borel–Cantelli give $Z_{x_n} \to \zeta(2)$ almost surely. Since $x_{n+1}/x_n \to 1$ and $\mu_A((0,x))$ is increasing in x, the same limit holds as $x \downarrow 0$.

From the frequency mass to the profile. Put $b(y) = 64\pi^2 y(1-y)$. Integration by parts in (8), followed by $z = b(y)x^2/c$, gives

$$\frac{\mathcal{C}_A(c)}{c} = \int_0^1 \frac{\sigma(y)}{b(y)} \int_0^\infty z e^{-z} Z_{\sqrt{cz/b(y)}}\, dz\, dy.$$

For a residue system satisfying the first limit, Z_x is bounded for all $x > 0$. For large x this follows from the periodicity of γ_A and its square-summable coefficients modulo one. Dominated convergence gives $\zeta(2)/(8\pi^2) = 1/48$, since $\int_0^1 \sigma(y)/(y(1-y))\, dy = 8$. □

The preprint of Jiang and Lau [17, Corollary 4.1] gives the upper bound $\mathcal{C}_A(c) \ll c$ for every residue system. With $z = 1$ and $\nu = 2$, their corollary bounds $\sum_{d \le D} |S(1, b; d)|^2/d \ll D$ uniformly over all nonzero integers b. For each D, choose a nonzero representative of $a^2 A$ modulo $\operatorname{lcm}(1, \dots, D)$. Partial summation then bounds the contribution from a fixed numerator a by $O(x^2/a^2)$. Summing over a gives $\mu_A((0, x)) \ll x^2$, and the preceding integral gives $\mathcal{C}_A(c) \ll c$. The theorems below do not use this bound.

5.3 Profiles for restricted families

For a trace class $t \equiv t_0 \pmod M$, replace each factor $L_{\ell,A}$ for $\ell \mid M$ in F_A by its restriction to $R \equiv t_0 \pmod{\ell^{v_\ell(M)}}$, divided by its mass. For $Y_1(\ell)$, replace just $L_{\ell,A}$ by the normalization of

$$V_A(R) = L_{\ell,A}(R) 1_{R \equiv A+1\,(\ell)} + 1_{A \equiv 1\,(\ell)}\, \ell\big(L_{\ell,A}(R) - 1\big) 1_{R \equiv A+1\,(\ell^2)}. \tag{11}$$

Write $F_A^\flat$ for the resulting weight, $\gamma_A^\flat$ for its Fourier coefficients, and $B_{A,c}^\flat$ for the Gaussian sum $B_{A,c}$ with $F_A^\flat$ in place of F_A. In either case write $P_T^\flat$ for the corresponding count law.

Proposition 5.3. *Suppose $T/p \to c > 0$ along a sequence of primes whose residues stabilize modulo every fixed integer, with limiting system A. Then*

$$D_{\mathrm{KL}}(P_T^\flat \| Q_T) \longrightarrow \mathbb{E}\big[B_{A,c}^\flat \log B_{A,c}^\flat\big], \qquad \chi^2(P_T^\flat \| Q_T) \longrightarrow \int_0^1 \sigma(y) \sum_{\beta \neq 0} |\gamma_A^\flat(\beta)|^2 e^{-64\pi^2 y(1-y)\beta^2/c}\, dy.$$

Both limits are finite and positive, and as $c \to \infty$ the first is $\frac{1}{2}\log c - 1 - \frac{1}{2}\log(2\pi) + \mathbb{E}[F_A^\flat \log F_A^\flat] + o(1)$.

Proof. For P_T^M the weight is $F 1_{\mathcal{A}}/Z$, and for $P_T^{Y_1}$ it is U/Z with U as in (4). Truncating F and $F^{(\ell)}$ at N gives periodic weights that stabilize for fixed N. Their squared tail bound from Lemma 3.2 persists under fixed congruence restrictions and omission of ℓ. The masses converge to the local means by Propositions 4.1 and 4.2 and the cutoff estimate (3). In the modular case (4) gives exactly (11). Thus the normalized cutoffs converge to $F_A^\flat$ in L^2. The likelihood-ratio and Parseval arguments of Theorem 5.1 apply to these cutoffs, with their endpoint contributions bounded by (1) and the beta-binomial estimate used there. The full ratio is nonnegative and its bounded second moment yields uniform integrability of $x \log x$. Normalization gives mean one for $F_A^\flat$, so the same wrapped Gaussian proves positivity and the large-c expansion. □

6 Short observations and Hecke trace asymptotics

A trace weight g is a fixed finite real linear combination of $F^{(J)} 1_{\mathcal{A}}$ as in Lemma 3.4. The coefficients may depend on a fixed residue class of p. Choose M divisible by the selector periods, the moduli specifying the coefficients, and all primes in the omitted sets J. Write P^g for the corresponding unnormalized count measure. The pairing $\mathcal{B}_r(g, h)$ is defined in Section 6.1.

Theorem 6.1 (Short-length limits for trace weights)**.** *Fix $0 < \delta \le 2/3$, a unit class $r \bmod M$, and a finite list of trace weights. Write Z_g for the mass of P^g and $m_{g,p}(j)$ for the degree-j Chebyshev moment of the unnormalized bias measure with weight g. There is a set of relative density one among primes in the class on which, for every pair g, h from the list and every fixed $\lambda > 0$, with $T = \lfloor \lambda p^{1-\delta} \rfloor$ and $K = p^{(1-\delta)/2}$,*

$$\frac{p}{T} \langle P^g - Z_g Q_T, P^h - Z_h Q_T \rangle_Q \longrightarrow \mathcal{B}_r(g, h)/48, \tag{12}$$

$$\frac{p}{K^2} \sum_{1 \le j \le K-2} m_{g,p}(j) m_{h,p}(j) \longrightarrow \mathcal{B}_r(g, h)/48. \tag{13}$$

Here $\langle P,R\rangle_Q=\sum_k P(k)R(k)/Q_T(k)$. *For fixed* $\lambda>0$ *and* $0<\eta\le 1$, *count the primes* $p\ge 5$ *in* $(X,2X]$ *in the designated class with* $T\ge 1$ *on which the first expression differs from its limit by more than* η. *Their number is*

$$O_{g,h,M,\delta,\lambda}\big(\pi(X)/(\eta^2\log X)\big).$$

Write $\mathcal{P}_\delta$ for the set with $g=h=F$ and $M=1$. The two limits are then the normalized chi-square divergence and, by Corollary 4.3, the level-one trace sum. The common set can depend on δ, the weights and r mod M. The dyadic bound is for each fixed λ and pair.

Proof of Theorem 6.1. We truncate the class-number weights and control the products at distinct frequencies. Section 6.1 averages the Fourier diagonal over primes and recovers the sharp cutoff.

Truncating the class-number weights. Let g_B be the smooth cutoff at B. Average its error by grouping the Mellin integrals by fundamental discriminant. Each d with $|d|\asymp D$ occurs for at most $O(X/\sqrt{D})$ pairs (p,t) in $(X,2X]$, since $4p=t^2+|d|r_t^2$. Heath–Brown's fourth moment [14, Theorem 2] and Cauchy–Schwarz bound the sum of $|L(1/2+iv,\chi_d)|^2$ over these discriminants by $O_\varepsilon(D^{1+\varepsilon}(1+|v|)^2)$. Multiplying by the multiplicity and the grid step $\Delta_p\asymp X^{-1/2}$, summing in D and integrating with weight $(1+|v|)^{-10}$ gives a total $O_\varepsilon(X^{1+\varepsilon})$. On shifting the Mellin integral to $\Re s=1/2$, Cauchy–Schwarz supplies B^{-1} in the squared error. The divisor factors cost p^ε and the omitted Euler factors are bounded. Markov's inequality and (1) give

$$\Delta_p\sum_{t^2<4p}|g(t)-g_B(t)|^2+\|P^g-P^{g_B}\|_Q^2\ll_{g,\varepsilon}p^\varepsilon/B \tag{14}$$

for all $B\ge 1$ and $1\le T\le p$ on the same set of primes, outside $O_{g,\varepsilon}(X^{1-\varepsilon/4})$ exceptions for $0<\varepsilon<1$. See Lemma S5.1 for the integral used in this bound.

Lemma 6.2 (Uniform approximation at short lengths)**.** *Let* A_p *have the residue of* p *at each prime other than* p, *and* p*-adic coordinate* 1. *For a weight* g *as above, let* Z_p *be the mass of* P^g. *Write* $\mathcal{C}^g_{A_p}(c)$ *for the nonzero-frequency integral of Theorem 5.1 with the Fourier coefficients of the corresponding weight on* $\widehat{\mathbf{Z}}$. *Fix* $0<\delta_0<1/2$. *Outside* $O_{g,\delta_0}(X^{1-\xi})$ *primes in each* $(X,2X]$, *uniformly for* $p^{-\delta_0}\le c=T/p\le 1/2$,

$$\|P^g-Z_pQ_T\|_Q^2=\mathcal{C}^g_{A_p}(c)+O_{g,\delta_0}(c^{3/2}+cp^{-\nu}) \tag{15}$$

for some $\xi,\nu>0$. *For fixed* $\lambda>0$, $1/6\le a\le 1/4$, *and every* $J>0$, *at* $T=\lfloor\lambda p^{2a}\rfloor$ *the same formula holds with* $c=\lambda p^{2a-1}$ *and* $p^{-\nu}$ *replaced by* $(\log X)^{-J}$, *outside* $O_{g,a,\lambda,J}(X(\log X)^{-J})$ *primes. In both cases the mass* Z_p *differs from the complete constant Fourier coefficient by a power-saving error.*

Proof. From moments to an angular integral. Write $U=T$ in the first range and $U=\lambda p^{2a}$ at a fixed weight exponent, and put $c=U/p$. Set $w=2\sqrt{p}\sin\phi$, $s=2\sqrt{p}\cos\phi$, $\theta_t=\arccos(t/(2\sqrt{p}))$ and $G(z)=\sum_{j\in\mathbf{Z}}e^{-j^2/(2U)}e^{ijz}$. After expanding the trace selectors, the cutoff Fourier coefficients are sums of $C_dS(m,mp;d)$ at $\beta=m/d+\alpha$, where $(m,d)=1$, α is a fixed rational shift and $|C_d|\ll_g d^{-2}$ by Lemma 3.2. With B fixed on $(X,2X]$, each C_d is constant in the given prime class. Write $\gamma_B(\beta)$ for the combined coefficient and Z_B for the mass of P^{g_B}. The odd kernel gives the angular sum

$$\mathcal{D}(\phi)=\sum_t g_B(t)\big(G(\phi-\theta_t)-G(\phi+\theta_t)\big)-2\pi\gamma_B(0)we^{-1/(2U)}.$$

Its coefficient of $\sin((m+1)\phi)$ is $4\pi\sqrt{p}\,e^{-(m+1)^2/(2U)}(m_{g_B,p}(m)-1_{\{m=0\}}\gamma_B(0))$. Sine Parseval gives (S2), with first-sine term $e^{-1/U}|Z_B-\gamma_B(0)|^2$. In the interior, linearizing the trace-to-angle map and

using Poisson summation give the mode $2\pi we(\beta s)e^{-2\pi^2w^2\beta^2/U}$. The equal-frequency terms give the truncated profile, since $w^2 = 16py(1-y)$ for $y = (1+\cos\phi)/2$ and $w^2|d\phi|/(2\pi p) = \sigma(y)|dy|$.

Longer observations. Take $\rho = (1/2-\delta_0)/2$, $B = X^{1/2-\rho}$ and split the Fourier sum at denominator $X^{1/4}$. Integration by parts bounds the low mixed terms, and the additive large sieve bounds the high squared norm on average. Their relative cost is $O(X^{\varepsilon}X^{1/4}/\sqrt{T})$, while the class-number tail costs $O(X^{\varepsilon}/(Bc))$. Both are $O(X^{-\rho+\varepsilon})$. The normalized squared Taylor and mass errors are $O(X^{\varepsilon}((Tc)^{-1} + B/(pc)))$. The edge contributes $O(c^{3/2} + cX^{-\nu})$ in squared norm. The uniform version and its dyadic summation are in Lemma S5.2.

Shorter observations. For $1/6 < a \le 1/4$, use $B = X^{3/4-a/2}$ and $D_+ = X^{1/2-a+\omega}$, with $\omega > 0$ small. For $\mu = (3a-1/2)/2$ we have $Bc \asymp X^{\mu}$ and $B/X^{1/2+a} = X^{-\mu}$. The odd kernel gives the edge bound. On rescaling the trace by $L = X^{1/2-2a}$ its Fourier transform decays with exponent $3/2$. The large sieve on moving intervals of length $\asymp \sqrt{X}L$ gives the relative squared error $X^{O(\nu+\varepsilon)}(X^{-a} + D/X^{1/2+a})$ at denominator scale D. In the interior the Taylor parameter is $\sqrt{p}/(w\sqrt{U}) \ll X^{-\nu}$ for $w \ge X^{1/2-a+\nu}$. The continuous integral $2\pi we^{-1/(2U)}$ cancels the zero mode. These estimates give the approximation in Lemma S5.3,

$$\chi_X(\phi)\, 2\pi w \sum_{\beta\neq 0} \gamma_B^+(\beta) e(\beta s) e^{-2\pi^2w^2\beta^2/U}, \qquad \chi_X(\phi) = \chi(X^{\omega}w/\sqrt{X}).$$

Here χ is a smooth cutoff from zero to one on $[1,2]$ and γ_B^+ restricts the original denominators to $d \le D_+$. For the first sine projection use $\sqrt{X}|\beta| \gg X^{a-\omega}$. Both squared errors are $O(c(\log X)^{-J})$ off $O(X(\log X)^{-J})$ primes.

Nearby frequencies. Oscillation in the trace removes well-separated pairs. For the rest, open the Kloosterman sums and average over integers. Put $b_0 = 1/2 - a$, $v = p/X$ and $u = s/\sqrt{X}$. Each pair has amplitude proportional to $\chi_X^2\sqrt{4v-u^2}/v$ times a Gaussian with exponent $-C_\lambda X^{2b_0}(4v - u^2)(\beta^2+\beta'^2)/v^{2a}$. Each derivative costs $O(X^{2\omega})$ times a weaker Gaussian. Gaussian decay and integration by parts in u allow us to impose

$$|\beta|, |\beta'| \le X^{-b_0+\omega+\varepsilon}, \qquad 0 < |\beta-\beta'| \le X^{-1/2+2\omega+\varepsilon},$$

with arbitrary power savings, also for one v derivative. Write $\beta = A/(Md)$, $\beta' = A'/(Me)$ with M containing the fixed shifts. Then $0 < |A|, |A'| \ll_g J = X^{2\omega+\varepsilon}$, $de \gg_g P = X^{2b_0-2\omega-2\varepsilon}$ and $0 < |h| = |Ae - A'd| \ll_g H = X^{2b_0-1/2+4\omega+\varepsilon}$. For $g_1 = (d,e) \asymp G$ there are $O_g(J^2D_+H/G)$ pairs: $g_1 \mid h$, and $A, A', d/g_1, h/g_1$ determine e/g_1.

If $(d,e) = 1$, the unshifted numerators give $h_0 = me - m'd$ and

$$S(m, mn; d)\overline{S(m', m'n; e)} = S(h_0^2, n; de).$$

Indeed $r = xe - x'd \bmod de$ runs through the units, and its inverse phase has numerator $(m^2e^2 + m'^2d^2)\bar{r} \equiv h_0^2\bar{r} \bmod de$. In general, frequencies lie on the grid of period $l = [d,e]$ with at most g_1 preimages, so the coefficients are $O_g(g_1(de)^{-2})$. Call the sum $F(n,v)$, with phases at n and amplitudes at v.

For fixed v, use a smooth nonnegative majorant of $1_{[1,2]}(n/X)$. Poisson summation bounds the kernel by $O_A((1+X\|\xi-\xi'\|)^{-A})$. Differences between grids of periods l, l' run through the grid of period $[l,l']$ with multiplicity (l,l'). Off zero the integral test gives $O(ll'/X)$. Since $g_1 l = de$, the bound per two pairs is $O_g(1/(Xde\,d'e'))$, giving $O_g((J^2D_+H)^2/(XP^2))$.

At a coincident frequency of reduced denominator q we have $l/q \mid g_1$: the numerator is a unit where the exponents of d, e differ. For $g_1 \asymp G$, choosing $l/q, d, e, A, A'$ gives at most $O_{g,\varepsilon}(J^2GX^{\varepsilon})$ pairs per frequency. Each has squared coefficient sum $O_g(GP^{-3})$. Cauchy–Schwarz and $G \ll H$ give $O_{g,\varepsilon}(X^{\varepsilon}J^4D_+H^2P^{-3})$ after the logarithmic cost of summing the blocks.

One v derivative costs $X^{4\omega}$ in these bounds. The inequality $\sup|f|^2 \ll \|f\|_2^2 + \|f\|_2\|f'\|_2$ in v and Cauchy–Schwarz in n cost $X^{2\omega}$ in the mean square and let us set $v = p/X$. The resulting bound is $O_{g,\lambda}(X^{-1-2a+O(\omega+\varepsilon)} + X^{a-3/2+O(\omega+\varepsilon)})$. Relative to c^2 the exponents are -4μ and -2μ before losses. Choose ω, ε small in terms of μ to get $O(X^{-\mu})$. Markov bounds the primes with a cross term above $c(\log X)^{-J}$ by $O(X^{1-\mu}(\log X)^{2J})$. At equal *trace* frequencies we add the coefficients before squaring in the profile.

The endpoint $a = 1/6$. Now $c \asymp X^{-2/3}$ and the preceding mean-square bounds reach the scale c^2. In Lemma S5.4 the quadratic functional equation converts the class-number remainder at $B = X^{2/3-\rho}$ into a dual sum of length $O(X^{1/3+\rho+\varepsilon})$. Removing square factors supported on the omitted primes reduces the conductor. The cutoff condition $\widehat{\psi}(-1/2) = 0$ cancels the pole in the zero mode. In squared angular norm the edge costs $O(X^{-7/6+7\nu+3\rho+\varepsilon})$ and the nonzero dual modes cost $O(X^{-1+\rho+\varepsilon})$ on average by the large sieve. Both are smaller than c by a power. The averaged tail estimate controls $g - g_{B_1}$ for $B_1 = X^{2/3+\rho}$. The angular estimates in the proof of S5.3 control the difference of the two cutoffs for $d \le 2B$. Subtracting this part bounds the block $d > 2B$ with its original phases.

Nearby frequencies have denominators of size $X^{1/3+O(\omega+\varepsilon)}$ and determinants $|h|, |h'| \ll X^{1/6+O(\omega+\varepsilon)}$. At coincident integer frequencies Parseval gives $O(X^{-3/2+O(\omega+\varepsilon)})$. For nonzero Poisson modes, $O(1/(Xd_1d_2d_3d_4))$ per quadruple gives a power saving on exceptional shared factors and discarded numerators. After separating shared primes, reciprocity gives a phase proportional to $h^2P_2/P_1 + h'^2P_1/P_2$, where $P_1 = d_1d_2$ and $P_2 = d_3d_4$. The modulus is a squarefree factor j_0 of the Poisson numerator, of size $X^{1/3}$ up to small powers. With h, h', d_3, d_4 fixed, the phase has two simple poles in d_1. Weil's bound and completion cost $O_\varepsilon(X^\varepsilon(1 + D_+/j_0)\sqrt{j_0})$. The determinants, second pair and numerator cost X, and the four coefficients supply $X^{-8/3}$, up to small powers. Thus $X^{1+1/6-8/3}$ saves a power over $c^2 \asymp X^{-4/3}$. The same bound for one v derivative lets us set $v = p/X$. Markov gives $O(cX^{-\xi})$ outside $O(X^{1-\xi})$ primes for some $\xi > 0$.

Recovering the profile. The truncated diagonal is $O_g(c\log^3 X)$. In the short range the omitted angular region costs $O_g(cX^{-\omega}\log^3 X + c^{3/2})$ and the diagonal above D_+ costs $O_{g,\varepsilon}(X^\varepsilon\sqrt{c}/D_+)$. Above the endpoint, the coefficient error in Lemma S5.2 is $O_g(B^{-1}\log^4 B)$ in squared diagonal norm, so Bc gives a power margin. Use Cauchy–Schwarz for overlaps and approximation errors, then remove the first sine using its mass estimate. At $a = 1/6$, S5.4 restores the full heat norm using the dual remainder and the coefficient comparison below D_+. For $m + 1 \le \sqrt{U}\log X$, the moment multiplier $\kappa_m(T)$ differs relatively from $e^{-(m+1)^2/U}$ by $O(U^{-1/2}\log^3 X)$, including the effect of $T = \lfloor U \rfloor$. The remaining degrees are negligible by Gaussian decay and $|m_{g_B,p}(m)| + |m_{g,p}(m)| \ll_{g,\varepsilon} X^\varepsilon(m+1)$. Thus Proposition 2.1 gives the centered count norm. Above the endpoint, projection and (14) replace g_B by g with a power saving. The same tail bound controls the change in mass at all lengths, including the endpoint, and the constant coefficient changes by $O_g(B^{-1/2})$. □

6.1 From the Fourier profile to sharp trace sums

By the Chinese remainder theorem a trace weight is a finite sum of products with local factors $w_{\ell,A} = a_\ell L_{\ell,A} + b_\ell$ at finitely many primes and $L_{\ell,A}$ elsewhere. Here a_ℓ, b_ℓ are real and periodic modulo ℓ^{v_ℓ} with $v_\ell \ge 1$ at every modified prime. Conversely, expanding a product gives a combination of $F^{(J)}1_{\mathcal{A}}$: a term without $L_{\ell,A}$ omits that prime from the class-number series.

Define the local pairing by

$$\mathcal{B}_{\ell,A}(w,z) = (1-\ell^{-2})\left(\frac{\ell}{\ell+1}\widehat{w}(0)\overline{\widehat{z}(0)} + \sum_{j\ge1}\ell^j \sideset{}{^*}\sum_{b \bmod \ell^j} \widehat{w}(b/\ell^j)\overline{\widehat{z}(b/\ell^j)}\right). \tag{16}$$

It is real for real weights. Equation (10) gives $\mathcal{B}_{\ell,A}(L_{\ell,A}, L_{\ell,A}) = 1$. For products take the product of the local pairings over $\ell \mid M$, and extend bilinearly to sums. This defines $\mathcal{B}_r(g,h)$ for $A \equiv r \bmod M$, since the local computation below uses only the residue of A modulo the common period.

Computing the local pairing. Fix ℓ and write $q = \ell^v$, $v \geq 1$, for the common period. For $j \leq v$ the coefficients are obtained from the class means of $L_{\ell,A}$. The mean on $R \bmod q$ is

$$\frac{1}{1+\ell^{-1}} \left(\sum_{i=0}^{v-1} \frac{\rho_i(R)}{\ell^i} + \frac{\rho_v(R)}{q(1-\ell^{-1})} \right),$$

where $\rho_i(R)$ counts roots modulo ℓ^i. Each lift of a root above level v determines one lift of the trace, giving this mean. For $j > v$, put $f(u) = u + A/u$. The root series gives

$$\widehat{w}(b/\ell^j) = \frac{\ell^{-2j}}{1-\ell^{-2}} \sideset{}{^*}\sum_{u \bmod \ell^j} a_\ell(f(u)) e(-bf(u)/\ell^j) \qquad (\ell \nmid b).$$

Indeed only the terms of the root series at levels j and above contribute. For $j \geq 2v$, split u modulo $\ell^{\lceil j/2 \rceil}$ and sum the lifts. The multiplier is constant on the lifts and the phase forces $u^2 \equiv A \bmod \ell^{\lfloor j/2 \rfloor}$. There are at most four unit roots at this modulus, including at 2. Thus the coefficient is $O_{w,\ell}(\ell^{-3j/2})$.

For $z = a'_\ell L_{\ell,A} + b'_\ell$, let N_j be the collision sum over unit pairs with $f(u) \equiv f(u') \pmod{\ell^j}$, with weight $a_\ell(f(u))a'_\ell(f(u'))$. Additive orthogonality on primitive numerators gives the part above level v of the frequency sum inside (16) as

$$\frac{(1-\ell^{-1}) \sum_{j>v} N_j/\ell^{2j} - N_v/(\ell q^2)}{(1-\ell^{-2})^2}. \tag{17}$$

The sum over j integrates $v_\ell(f(u)-f(u')) - v$ over the matching classes modulo q, with their weights and Haar measure on $\mathbf{Z}_\ell^2$. Since $f(u) - f(u') = (u-u')(uu'-A)/(uu')$, the involution $u' \mapsto A/u'$ makes the two valuations have the same integral, preserving the classes of the traces and their weights. The sum is thus q^{-2} times the finite sum over matching unit pairs modulo q of their weight times

$$2\min\{v, v_\ell(u-u')\} - v + \frac{2\,1_{u=u' \,(\bmod q)}}{\ell-1}. \tag{18}$$

On a diagonal pair of residue classes the extra valuation has mean $1/(\ell-1)$. All these calculations use only $A \bmod q$.

Proof of Theorem 6.1. Averaging over residue systems. The short-length constant comes from the Fourier mass near zero. For two product weights, replace the squared coefficient in the proof of Theorem 5.2 by the product of the two coefficients, conjugating the second. The positive and negative frequencies combine to give a real profile. Include all modified primes among the primes $\ell \leq L$ held fixed. At a modified prime the term with denominator exponent zero is $(\ell/(\ell+1))\widehat{w}(0)\overline{\widehat{z}(0)}$, by summing the powers of ℓ in the numerator. At exponent $j \geq 1$ the average over unit numerators gives the corresponding summand of (16). The Euler product becomes $\zeta(2)\mathcal{B}_r(g,h)$ on the class $A \equiv r \bmod M$. The uniform bounds of the local coefficients preserve the conditional mean error $O_{g,h,M}(x^{1/32})$. In the fourth moment the finitely many modified factors are uniformly bounded in their exponents, so the vanishing condition and divisor counts at the other primes give the same $O_{g,h,M}(L^{-2})$ estimate, also conditional on the designated class.

The resulting Z_x differs from $\zeta(2)\mathcal{B}_r(g,h)$ in conditional L^2 by $O_{g,h,M}((\log(1/x))^{-1/2})$ for $0 < x \le 1/2$. Its norm is bounded for larger x by periodicity and square-summability. In the Gaussian integral of Theorem 5.2, the region $b(y) \ge c^{1/4}$ and $z \le c^{-1/4}$ gives $\sqrt{cz/b(y)} \le c^{1/4}$. Its complement has integral weight $O(c^{1/8})$. Minkowski's inequality yields

$$\mathbb{E}\left[\left|\frac{\mathcal{C}_A^{g,h}(c)}{c} - \frac{\mathcal{B}_r(g,h)}{48}\right|^2 \;\middle|\; A \equiv r \bmod M\right] \ll_{g,h,M} \frac{1}{\log(1/c)} \tag{19}$$

for $0 < c \le 1/2$, with $\mathcal{C}_A^{g,h}$ the bilinear profile of Lemma 6.2. Extend by bilinearity to sums.

Averaging over primes. The transfer to primes uses Dirichlet characters. Group $\mathcal{C}_A^{g,h}(c)/c$ into terms $H_d(A,c)$ by reduced denominator d. They depend on $A \bmod \operatorname{lcm}(M,d)$. Integrating the Gaussian against σ gives $O(\min(1, c^{3/2}|\beta|^{-3}))$ at $\beta \ne 0$. The coefficient bounds then give

$$|H_d(A,c)| + |c\partial_c H_d(A,c)| \ll_{g,h,M} \tau(d)^2 \min\{\sqrt{c}, (d^2\sqrt{c})^{-1}\}. \tag{20}$$

Use $c(p) = \lambda p^{-\delta}$ on $(X, 2X]$ and truncate at $D = \lfloor X^{5/12} \rfloor$. Since $\delta \le 2/3$, $D\sqrt{c(p)} \gg_\lambda X^{1/12}$ and $Q = MD^2 \ll_M X^{5/6}$. The tail $O_{g,h,M}(\log^3(2D)/(D\sqrt{c(p)}))$ saves a power, as does the resulting error in the square since the full sum is $O_{g,h,M}(\log^3(2/c(p)))$.

Expand the truncated centered square, multiply by the indicator of $r \bmod M$, and group the terms by modulus b. For a product the modulus is $\operatorname{lcm}(M,d,e)$. Write $V_b(a;c)$ for the resulting function on the units modulo b, with $b \le Q = MD^2$. From (20), both H_d and $c\partial_c H_d$ are $O_{g,h,M}(\tau(d)^2/d)$. There are at most $\tau(b)^2$ pairs and $de \ge b/M$. The linear and constant terms obey the same bound, so

$$\|V_b\|_\infty + \|c\partial_c V_b\|_\infty \ll_{g,h,M} \tau(b)^6/b.$$

Expand on the units in Dirichlet characters. Parseval gives the same bound for the Euclidean norm of the coefficients. Since $p > Q$ for sufficiently large X, each character in the expansion can be replaced on primes by the primitive character inducing it. For primitive characters of conductor f, the norm of the combined coefficient vector is at most

$$\ll_{g,h,M} \sum_{\substack{b \le Q \\ f \mid b}} \frac{\tau(b)^6}{b} \ll_{g,h,M} \frac{\tau(f)^6}{f}(\log X)^C$$

for an absolute C. The same estimate holds after $c\partial_c$.

The principal character gives the exact Haar mean of the class-weighted square. For conductors f in a dyadic block $R \le f < 2R$, the squared coefficient norm is $O_{g,h,M}(R^{-1}(\log X)^{C'})$ by the usual mean bound for divisor powers. The multiplicative large sieve [16, §7.5] applied to the prime indicator on any subinterval of $(X, 2X]$, followed by Cauchy–Schwarz, bounds the contribution by $O_{g,h,M}((\log X)^{C''}(X/\sqrt{R} + \sqrt{XR}))$. For $f \le (\log X)^A$, Siegel–Walfisz [16, Corollary 5.29] gives arbitrary logarithmic savings for the nonprincipal characters. Take A sufficiently large and sum the dyadic blocks. With $Q \ll_M X^{5/6}$, the total is $O_{g,h,M,J}(X/(\log X)^J)$ for any J. This estimate also allows the variation of $c(p)$. Apply the fixed-parameter bound first at $c(X)$ and then, for each $X \le y \le 2X$, to the derivative on $(y, 2X]$, and integrate with $|c'(y)|/c(y)$. Its integral is bounded. Dividing by the prime count in the specified class, (19) gives a mean squared error $O_{g,h,M,\delta,\lambda}(1/\log X)$.

Apply Lemma 6.2 to $g, h, g+h$ and polarize, choosing $\delta < \delta_0 < 1/2$ in its first range. On the retained primes the left side of (12) differs from the normalized profile at $c(p)$ by $O((\log X)^{-2})$,

including the floor in T by (20). Chebyshev's inequality and the exceptional sets of the lemma give the count for $\eta > (\log X)^{-1/2}$. For smaller η the count is trivial.

A common prime set. It suffices to control diagonal norms for the given weights and all their pairwise sums. Enumerate the positive rational multipliers and let B_j be the bad set for the first j multipliers at tolerance $1/j$. These sets increase and have relative density zero. Choose increasing X_j so that B_j has proportion at most $1/j$ up to every $X \ge X_j$ in the class, and remove $B_j \cap [X_j, X_{j+1})$. All removals up to X in the jth interval lie in B_j. The remaining set has relative density one. The diagonal squared norm increases with T by contraction under thinning, including for signed weights. Bracketing by rational multipliers and then polarizing gives (12) for every $\lambda > 0$ on this set.

The sharp cutoff. For a weight g, consider the positive measure

$$\nu_p = \frac{p}{K^2} \sum_{j \ge 1} m_{g,p}(j)^2 \delta_{j^2/K^2}.$$

Here δ_x is unit mass at x. By Proposition 2.1, the sum against $\kappa_j(T)$ is $(p/K^2)\|P^g - Z_g Q_T\|_Q^2$, which tends to $\lambda \mathcal{B}_r(g,g)/48$. For the Sato–Tate parameters the product gives, for $j = O(K)$, $\log \kappa_j(T) = -j(j+2)/T + O((j+1)^3/T^2)$. Thus at $u = j^2/K^2$ the weights converge uniformly on compact intervals to $e^{-u/\lambda}$. There they are bounded below, so the masses are bounded. For $j \le T$ and $T \le s \le 2T$ each term of the logarithmic derivative of $\kappa_j(s)$ is $(2r+3)/((s-r)(s+r+3)) \ge (2r+3)/(12T^2)$, where $0 \le r < j$. Integrating gives $\kappa_j(T) \le e^{-j^2/(12T)} \kappa_j(2T)$. The weights vanish for $j > T$, and $2T \le \lfloor 2\lambda K^2 \rfloor$. Hence the sum at 2λ bounds the weighted tail $u > R$ by $O(e^{-R/(12\lambda)})$.

Test along a vaguely convergent subsequence with continuous cutoffs on compact intervals, then let R tend to infinity using the tail bound. Each limit measure has Laplace transform $\lambda \mathcal{B}_r(g,g)/48$. Multiplying by e^{-u} and mapping u to $z = e^{-u}$, the choices $\lambda = 1/(n+1)$ give the moments of $(\mathcal{B}_r(g,g)/48)\, dz$ on $[0,1]$. Polynomials determine this measure. Hence ν_p converges on compact intervals to $(\mathcal{B}_r(g,g)/48)\, du$. This measure has no atom at 1, so the cutoff $(K-2)^2/K^2 \to 1$ gives the diagonal limit. Polarization proves (13). □

6.2 The log loss at short lengths

To recover the log loss from the chi-square divergence, we use a quadratic expansion on counts where the likelihood ratio is close to one and bound the remaining counts separately.

Proposition 6.3. *Fix $0 < \eta < 1/2$ and $L > 0$. Uniformly for $Lp^{1/3} \le T \le p^{1-\eta}$, as $p \to \infty$,*

$$D_{\mathrm{KL}}(P_T \| Q_T) = \big(\tfrac{1}{2} + o(1)\big) \chi^2(P_T \| Q_T) + o(T/p).$$

Corollary 6.4. *On the set $\mathcal{P}_\delta$ of Theorem 6.1, for every fixed $\lambda > 0$ and $T = \lfloor \lambda p^{1-\delta} \rfloor$,*

$$\frac{p}{T} D_{\mathrm{KL}}(P_T \| Q_T) \longrightarrow \frac{1}{96}.$$

Proof of Proposition 6.3. Put $u_k = P_T(k)/Q_T(k)$ and $\Psi(u) = u \log u - u + 1$, so that $D_{\mathrm{KL}}(P_T \| Q_T) = \sum_k Q_T(k) \Psi(u_k)$. One has $0 \le \Psi(u) \le (u-1)^2$, and $\Psi(u) = \frac{1}{2}(u-1)^2(1 + O(|u-1|))$ for $|u-1| \le 1/2$. For $T \le p^{9/20}$ every count admits a one-sided interval of length $\ell = p^{-1/2+\varepsilon}$ from $x = k/T$ toward the farther endpoint y. On that interval $f = (1-u)x + uy$ with $u \le 2\ell$, so $D(x\|f) \le -\log(1-u) \ll \ell$ for Bernoulli laws. The binomial likelihood is at least $m \exp(-CT\ell)$ there, where m is its maximum. Integrate the masses of its superlevel intervals as below. For heights above $m \exp(-CT\ell)$ use the mass at that height as an upper bound. The short-interval theorem then gives $u_k = 1 + O_J((\log p)^{-J} + T\ell)$ uniformly in k. Take $\varepsilon < 1/20$ and use the expansion of Ψ to finish this case.

For $T > p^{9/20}$ put $c = T/p$, choose $\varepsilon > 0$ small in terms of η, and put $\vartheta = p^{3\varepsilon}c + p^{-1/2+2\varepsilon}$. Call k a bulk count when $\hat{k}(1-\hat{k}) \geq \vartheta$ for $\hat{k} = k/T$, and an edge count otherwise.

For a bulk count the function $f \mapsto b_{T,f}(k)$ is unimodal with maximum m at $\hat{k}$. Its superlevel sets I_s are intervals, and $P_T(k)$ and $Q_T(k)$ are the integrals over $0 < s < m$ of the masses of I_s under the trace law and under the Sato–Tate law. Since the relative entropy of two Bernoulli laws of parameters $\hat{k}$ and f is at most $(f-\hat{k})^2/(f(1-f))$, the interval $|f-\hat{k}| \leq \frac{1}{2}p^{-1/2+\varepsilon}$ lies in the superlevel set of height $me^{-Cp^{-1+2\varepsilon}T/(\hat{k}(1-\hat{k}))}$. The bulk threshold keeps the interval away from the endpoints. Let s_* be the largest s for which I_s has length at least $p^{-1/2+\varepsilon}$. Then $m - s_* \ll mp^{-1+2\varepsilon}T/(\hat{k}(1-\hat{k})) \leq mp^{-\varepsilon}$. For $s \leq s_*$ the two masses of I_s agree up to a factor $1 + O_J((\log p)^{-J})$ by [9, Theorem 1.3]. For $s > s_*$ each is at most its value at s_*, and $Q_T(k)$ is at least s_* times the Sato–Tate mass of I_{s_*}. Hence $|u_k - 1| \ll_J (\log p)^{-J} + p^{-\varepsilon}$ on the bulk.

For the edge, $\Psi(u) \leq (u-1)^2 \leq 2u^2 + 2$. So the edge counts contribute to either divergence at most twice the sum of $P_T(k)^2/Q_T(k)$ plus twice the sum of $Q_T(k)$ over the edge. The second sum is $O(\vartheta^{3/2} + T^{-3/2})$, the last term accounting for the counts 0 and T. By Cauchy–Schwarz at each count,

$$P_T(k)^2 \leq P^{F^2}(k)P^1(k), \qquad P^1(k) \ll Q_T(k)\big(1 + \Delta_p m_{T,k}\big).$$

One has $F(t) \ll \log p\,(\log\log p)^2$, by $L(1,\chi) \ll \log p$ and the bound $\sigma_{-1}(r)^2$ for the finite factor. This gives

$$\sum_{\text{edge}} \frac{P_T(k)^2}{Q_T(k)} \ll (\log p)^3 \sum_{\text{edge}} Q_T(k)\big(1 + \Delta_p^2 m_{T,k}^2\big) \ll (\log p)^3\big(\vartheta^{3/2} + T^{-3/2} + c\sqrt{\vartheta} + T^{1/2}p^{-1}\big),$$

by the bounds for $Q_T(k)$ and $m_{T,k}$. After division by c the terms save powers: $\vartheta^{3/2}/c \ll_L p^{9\varepsilon/2-\eta/2} + p^{-1/5+3\varepsilon}$ and $pT^{-5/2} \ll_L p^{-1/8}$. On the bulk, $\Psi(u_k) = \frac{1}{2}(u_k-1)^2(1+o(1))$ uniformly. Summing proves the proposition. □

For the corollary, fix δ, λ as in Theorem 6.1. The length $\lfloor \lambda p^{1-\delta} \rfloor$ belongs to the range of the proposition with $\eta = \delta/2$ and $L = \lambda/2$ for large p. Combining the proposition with $p\chi^2/T \to 1/48$ on $\mathcal{P}_\delta$ gives the stated limit.

6.3 Prime level by character

The trace-weight pairing gives both the second moment in each nebentypus and the mixed moments between characters. For a Dirichlet character χ mod ℓ, write $\tau_{\chi,k}(p)$ for the trace of T_p on $S_k(\Gamma_0(\ell), \chi)$ multiplied by $p^{-(k-1)/2}$. Set it to zero if $\chi(-1) \neq (-1)^k$. These are the character components of the trace on $\Gamma_1(\ell)$.

Theorem 6.5 (Prime-level character traces)**.** *Fix an odd prime ℓ, a unit class r mod ℓ and $0 < \delta \leq 2/3$. There is a relative-density-one set of primes $p \equiv r$ mod ℓ on which, for $K = p^{(1-\delta)/2}$ and all characters χ, ψ mod ℓ,*

$$\frac{48}{K^2} \sum_{3 \leq k \leq K} \tau_{\chi,k}(p)\overline{\tau_{\psi,k}(p)} \longrightarrow (\ell+1)\big(1_{\chi=\psi} + \chi(r)1_{\psi=\bar{\chi}}\big) - (\ell-1)1_{\chi=\psi=1}. \tag{21}$$

The diagonal is $\ell + 3$ for the trivial character, $(\ell+1)(1+\chi(r))$ for the quadratic character, and $\ell + 1$ for a nonreal character.

Corollary 6.6. *On the set of Theorem 6.5, for every fixed $\lambda > 0$ and $T = \lfloor \lambda p^{1-\delta} \rfloor$,*

$$\frac{48}{K^2} \sum_{3 \leq k \leq K} |\tau_k^{(\ell)}|^2, \qquad \frac{48p}{T}\chi^2(P_T^{Y_1} \| Q_T) \quad \longrightarrow \quad (\ell-1)\big(\ell + (\ell+1)1_{r=1}\big).$$

Proof of Theorem 6.5 and Corollary 6.6. We compute the pairing for fixed Frobenius eigenvalues modulo ℓ, then take character sums. Put $L = L_{\ell,A}$ and $n = \ell - 1$ at determinant A. For $a \in \mathbf{F}_\ell^\times$ choose a unit lift and put

$$W_a(R) = n\left(L(R)1_{R\equiv a+A/a\,(\ell)} + \ell(L(R)-1)1_{A\equiv a^2\,(\ell)}1_{a^2-aR+A\equiv 0\,(\ell^2)}\right).$$

The last condition, on the repeated-root class, is independent of the lift. Let g_χ have local factor $n^{-1}\sum_a \chi(a)W_a$ at ℓ and the usual factors elsewhere. By [20, Theorem 10, (43)–(44)], an elliptic order with index prime to ℓ contributes the sum of $\chi(a)$ over roots modulo ℓ. When ℓ divides the index, the repeated root lifts modulo ℓ^2 and receives the additional factor ℓ. Restricting to these orders removes the first term of the local class-number sum, giving $L - 1$ as in Supplement S6. The hyperbolic term gives, in matching parity,

$$\tau_{\chi,k}(p) = -\sqrt{p}\, m_{g_\chi,p}(k-2) - (1+\chi(p))p^{-(k-1)/2}.$$

In the other parity both the trace and the moment vanish, since $W_a(-R) = W_{-a}(R)$. The squared corrections sum to $O_\ell(p^{-2})$.

Scaling the root variable by a reduces the mean and diagonal pairing of W_a to those of W_1 at determinant A/a^2. If $a^2 \not\equiv r \bmod \ell$, the roots are simple and W_a is ℓ times its trace-class indicator. Otherwise, on classes modulo ℓ^2, the mean of W_a is n times the indicator modulo ℓ plus ℓ times the indicator modulo ℓ^2. The root-series means of Section 6.1 give mean one in both cases and, for a repeated root, coefficient magnitudes 1 and $1/\ell$ at denominators ℓ and ℓ^2. In (17) the ℓ lifts all give the same trace modulo ℓ^2, with coefficient $\ell^2 - 1$ for L. Using $\sum_{u,u'} \min(2, v_\ell(u-u')) = \ell^2 + \ell$ in (18) gives higher-denominator energy ℓ. Thus $\mathcal{B}_{\ell,A}(W_a, W_a)$ is $n\ell$ in the simple case and $n(2\ell+1)$ in the repeated case.

The weights satisfy $W_{r/a} = W_a$. Reindexing gives $g_\chi = \chi(r)g_{\bar\chi}$, and the trace corrections satisfy the same relation. Thus $\tau_{\chi,k}\overline{\tau_{\bar\chi,k}} = \chi(p)|\tau_{\chi,k}|^2$ exactly. If b is neither a nor r/a, the weights have disjoint trace classes modulo ℓ. In that case the cross Fourier sum is -1 at denominator ℓ and zero at ℓ^j for $j \ge 2$, by the Ramanujan sum on the difference of the trace classes. Inserting these values into (16) gives the single formula

$$\mathcal{B}_{\ell,A}(W_a, W_b) = n\big((\ell+1)(1_{b=a} + 1_{b=r/a}) - 1\big). \tag{22}$$

The involution term in the character sum is $\sum_a \chi(a)\overline{\psi(r/a)} = n\chi(r)1_{\psi=\bar\chi}$, which gives the phase in (21). Apply Theorem 6.1 to the real and imaginary parts of all the weights, with $M = \ell^2$, and extend the pairing by sesquilinearity. The constants depend only on r mod ℓ, so the finitely many prime classes modulo ℓ^2 give the asserted set.

Summing over both characters gives the trace limit in the corollary. For the prediction limit, $\sum_\chi g_\chi$ has local factor W_1 by character orthogonality, so the global weight is $(\ell-1)U$ from (4). Its mass tends to one. The smoothed limit in Theorem 6.1 then gives the chi-square limit. □

In these units the dimension is $\ell^2 - 1$ [37, Section 2]. The two level-one copies count twice in the trivial-sector dimension and four times in the second moment, adding two to the difference. The remaining correction is $(\ell+1)\sum_{\chi\neq 1}\chi(r)$. For nonsquare r the quadratic trace vanishes and the mixed terms sum to zero. For square $r \neq 1$ the mixed sum is $-2(\ell+1)$. At fixed weight a Rankin–Selberg prime average counts common automorphic constituents after twisting. Of the twists by characters modulo ℓ, the only nontrivial level-preserving one is the contragredient twist by $\bar\chi$ for $\chi \neq 1$ [3, Sections 1 and 3].

6.4 Variance and exceptional primes

At level one, the variances of both statistics in Theorem 6.1, over all primes $X < p \le 2X$, are asymptotic to explicit multiples of $X^{-\delta/2}\log X$ for every fixed $0 < \delta \le 1/4$. The arithmetic factor is the same for the two statistics. The kernel in the second factor depends on the cutoff. See S8 for $p\chi^2/T$ and S9 for the sharp trace sum at $K = p^{(1-\delta)/2}$.

Corollary 6.7. *Fix $0 < \delta \le 2/3$ and $\lambda > 0$, and let $T = \lfloor \lambda p^{1-\delta} \rfloor$. Uniformly for $X \ge 2$ and $0 < \eta \le 1$, counting only primes $p \ge 5$ for which $T \ge 1$,*

$$\#\Big\{X < p \le 2X : \Big|\frac{p}{T}\chi^2(P_T\|Q_T) - \frac{1}{48}\Big| > \eta\Big\} \ll_{\delta,\lambda} \frac{\pi(X)}{\eta^2 \log X}.$$

The same bound holds for $p\, D_{\mathrm{KL}}(P_T\|Q_T)/T$ and $1/96$.

Proof. Apply Theorem 6.1 with $g = h = F$. On the retained primes $p\chi^2/T \ll \log^3(2p)$ by the profile estimates in its proof. In Proposition 6.3 the bulk Taylor error can be taken as $O((\log p)^{-4})$ relative to χ^2, and the edge error is $O((T/p)p^{-\kappa})$ for some $\kappa > 0$, by its proof. Thus $p\, D_{\mathrm{KL}}/T = \frac{1}{2}p\chi^2/T + O(1/\log p)$ there. The profile estimates remove at most $O(X(\log X)^{-3})$ primes, giving the same bound for log loss. □

6.5 Short intervals and log loss for restricted families

Lemma 6.8 (Sato–Tate on short intervals)**.** *Let g be a weight from Theorem 6.1, with $g(t) \ge 0$ on the trace grid for the primes in the designated class. Assume its complete constant coefficient Z_r is positive. The normalized bias measure μ_p^g satisfies*

$$\mu_p^g(I) = \big(1 + O_{g,M,\varepsilon,J}((\log p)^{-J})\big) \int_I \sigma(f)\, df$$

for every $\varepsilon > 0$, $J \ge 1$, and $I \subset [0,1]$ of length at least $p^{-1/2+\varepsilon}$.

Proof. Expand the weight into terms $F^{(J_0)}1_{\mathcal{A}}$ and treat each trace progression $t = t_0 + Na$ as a separate Euler product average. In the local formula of [9, Theorem 3.2 and Corollary 3.3] use trace t and determinant p. The mass is $\Delta_p\sigma(f_t)\prod_s f_s(t,p,1)$ with the primes in J_0 omitted. The factor at p is $1 + O(p^{-1})$ uniformly and may be dropped at that relative cost.

Apply [9, Theorem 4.2] to the single integer variable a, with $d = k = 1$, analytic scale $\asymp \sqrt{p}$, and polynomial $(t_0 + Na)^2 - 4p$. Its content is bounded by N^2. Its discriminant is nonzero and its values are negative and congruent to 0 or 1 modulo 4. At retained primes $s \ne p$ take the normalized matrix counts of trace t and determinant p as periodic approximants. At omitted primes both the factor and its approximants are one. Theorem 3.2 of that paper supplies the local bounds. The retained local means are at least $(1 + s^{-1})^{-1}$ by the root series, and outside N the means are one, since t runs through all residue classes. The primes dividing $2pN\prod_{s\in J_0} s$ may be included in its exceptional integer. The product of the means together with the progression factor $1/N$ is the complete constant coefficient of the term.

For $s_I = |I|$ and $H = \sup_I \sigma$, concavity gives $\int_I \sigma \ge s_I H/2$. On every class of $a \bmod b$ the Riemann sum of the weights differs from $(Nb)^{-1}\int_I \sigma$ by $O_N(H/\sqrt{p})$, by bounded variation. Thus the total weight is $\asymp_N s_I H$, and condition (4.6) of that theorem holds with this weight as $\widetilde{W}$. The sum of discrepancies is $O_N(QH/\sqrt{p})$ for $Q = \exp((\log\log(2p))^2)$. The square-class error is $O_N(H/\sqrt{p})$: a fixed square class gives a fixed imaginary quadratic field containing $(t + \sqrt{t^2 - 4p})/2$, an algebraic integer of norm p. There are at most two ideals of norm p, each with at most six generators. Dividing

these errors by the total weight saves a power and gives an error $O((\log p)^{-J})\int_I \sigma$ for the term. Summing the finitely many terms gives main mass $Z_r\int_I \sigma$. Evaluating at $I=[0,1]$ and dividing by the actual mass gives the result. □

Proposition 6.9. *For g as in Lemma 6.8 let $P_T^g = P^g/Z_g$. Fix $0<\eta<1/2$, $L>0$ and $J\ge 1$. For some $\kappa=\kappa(\eta)>0$, uniformly for $Lp^{1/3}\le T\le p^{1-\eta}$,*

$$D_{\mathrm{KL}}(P_T^g\|Q_T) = \tfrac{1}{2}\chi^2(P_T^g\|Q_T) + O_{g,M,\eta,J,L}\left((\log p)^{-J}\chi^2(P_T^g\|Q_T) + \frac{T}{p}\,p^{-\kappa}\right).$$

In particular, on the sets in Theorem 6.1, the limit of $p\,D_{\mathrm{KL}}(P_T^g\|Q_T)/T$ is $\mathcal{B}_r(g,g)/(96Z_r^2)$.

Proof. Lemma 6.8 and the same superlevel argument give likelihood ratios $1+O_J((\log p)^{-J})$ on every count for $T\le p^{9/20}$, and on the bulk for larger T. For larger T the edge estimate follows from $P_T^g \ll_{g,M} P_T$. Indeed $L_{\ell,p}\ge(1+\ell^{-1})^{-1}$ by its root series, so $|F^{(J_0)}|\ll_{J_0} F$. The same ϑ therefore gives a power saving after division by $c=T/p$. The bulk Taylor expansion of $u\log u-u+1$ gives the factor $1/2$ with relative error $O_J((\log p)^{-J})$. For the limit, take $\eta=\delta/2$ and $L=\lambda/2$ and use (12) on the diagonal divided by Z_g^2, where $Z_g\to Z_r$ by Lemma 6.2. □

Corollary 6.10. *On the sets of primes in Corollary 6.6, for every fixed $\lambda>0$ and $T=\lfloor\lambda p^{1-\delta}\rfloor$,*

$$\frac{p}{T}D_{\mathrm{KL}}(P_T^{Y_1}\|Q_T)\longrightarrow\begin{cases}\ell(\ell-1)/96, & p\not\equiv 0,1 \pmod{\ell},\\ (2\ell+1)(\ell-1)/96, & p\equiv 1 \pmod{\ell}.\end{cases}$$

Proof. The length T lies in the range of Proposition 6.9 with $\eta=\delta/2$ and $L=\lambda/2$ for large p. Combine it with the chi-square limits of Corollary 6.6. □

6.6 The constants for arbitrary trace classes

For a fixed modulus $M\ge 2$, take $p\equiv r \pmod M$ with r a unit and restrict the trace to $t\equiv t_0 \pmod M$. The pairing reduces to counting pairs of roots at each $q=\ell^v\|M$. Let $\mathcal{R}_\ell$ be the roots of u^2-t_0u+r modulo q, let $h=|\mathcal{R}_\ell|$, and let ρ_i count the roots modulo ℓ^i, with $\rho_0=1$. Set

$$B_\ell=\frac{1}{q(1+\ell^{-1})}\left(\sum_{i=0}^{v-1}\frac{\rho_i}{\ell^i}+\frac{h}{q(1-\ell^{-1})}\right),\qquad S_\ell=\sum_{u,w\in\mathcal{R}_\ell}\min\{v,v_\ell(u-w)\}.$$

Here the summand for $u=w$ is v. The local factor is

$$A_\ell = q\varphi(q)+\frac{(1-\ell^{-1})(2S_\ell-vh^2)+(2h-h^2)/\ell}{(1-\ell^{-2})q^2B_\ell^2}. \tag{23}$$

Theorem 6.11. *Fix M,r,t_0 as above and $0<\delta\le 2/3$. There is a relative-density-one set of primes in the class r on which every fixed $\lambda>0$ and $T=\lfloor\lambda p^{1-\delta}\rfloor$ give*

$$\frac{p}{T}\chi^2(P_T^M\|Q_T)\longrightarrow\frac{1}{48}\prod_{\ell\mid M}A_\ell,\qquad \frac{p}{T}D_{\mathrm{KL}}(P_T^M\|Q_T)\longrightarrow\frac{1}{96}\prod_{\ell\mid M}A_\ell.$$

If each local weight is constant on its trace class, the product is $M\varphi(M)$. This includes odd ℓ with $v_\ell(t_0^2-4r)<v$ and $\ell=2$ with odd t_0. At $v=1$ with a repeated root, (23) gives $\ell(\ell-1)+1-\ell^{-2}$, including $11/4$ for even traces modulo 2. For example, $t_0\equiv 2$ and $r\equiv 1 \bmod 9$ give $A_3=54+216/121$.

Proof. In (16) use $W = L_{\ell,A} 1_{R \equiv t_0\,(q)}/B_\ell$. The class mean formula gives the stated $B_\ell > 0$. Because W has mass one on a single class, every coefficient of denominator at most q has modulus one. Those terms, including the constant term, contribute $q\varphi(q)$ after multiplication by $1 - \ell^{-2}$. For the higher terms the unit pairs in (18) range over $\mathcal{R}_\ell^2$. Their sum is $2S_\ell - vh^2 + 2h/(\ell - 1)$, and $N_v = h^2/B_\ell^2$. Formula (17) gives the correction in (23). The product of the local pairings gives the normalized chi-square limit by Theorem 6.1, and Proposition 6.9 gives the KL limit. □

6.7 Traces on $\Gamma_0(N)$

The local pairing also gives the full and new trace moments at arbitrary fixed levels. For $p \nmid N$ and even $k \geq 4$, let $\tau_{N,k}^0(p)$ and $\tau_{N,k}^{\mathrm{new}}(p)$ be the full and new T_p traces at level N, multiplied by $p^{-(k-1)/2}$. For a prime ℓ, write $F_e(\ell, r)$ and $B_{ef}(\ell, r)$ for the full diagonal and new mixed factors computed in S6. Let G_N be the full trace weight defined there and $P_T^N = P^{G_N}/Z_{G_N}$.

Theorem 6.12 (Mixed Hecke traces at fixed levels)**.** *Fix $1/6 \leq a < 1/2$ and a finite list of levels with least common multiple M. In every unit class $r \bmod M$ there is a common relative-density-one prime set on which, with $K = p^a$,*

$$\frac{48}{K^2} \sum_{\substack{4 \leq k \leq K \\ k \text{ even}}} |\tau_{N,k}^0(p)|^2 \longrightarrow \prod_{\ell \mid M} F_{v_\ell(N)}(\ell, r),$$

$$\frac{48}{K^2} \sum_{\substack{4 \leq k \leq K \\ k \text{ even}}} \tau_{D,k}^{\mathrm{new}}(p) \overline{\tau_{E,k}^{\mathrm{new}}(p)} \longrightarrow \prod_{\ell \mid M} B_{v_\ell(D), v_\ell(E)}(\ell, r)$$

for levels N, D, E in the list. At odd primes only the Legendre symbols matter. At 2 at most the residue of $p \bmod 8$ is needed. For each fixed $\lambda > 0$ and $T = \lfloor \lambda K^2 \rfloor$, $p\, D_{\mathrm{KL}}(P_T^N \| Q_T)/T$ tends to the full product divided by 96.

Proof. Section S6 uses the lifted-root trace formula to express G_N and the new weight as products of local factors $aL_{\ell,p} + b$ with fixed periods dividing M. The errors in the normalized traces have total square $O_M(p^{-3})$ and can be removed by Cauchy–Schwarz. Equation (13) with $\delta = 1 - 2a$ gives both products, throughout $1/6 \leq a < 1/2$. The full weight is nonnegative with limiting mass one, so Proposition 6.9 gives the log loss. □

Section S6 classifies the vanishing local factors in the mixed product. For squarefree levels the products are $1_{D=E}\varphi(D)$ for the newspaces and $\prod_{\ell \mid N}(\ell + 3)$ for the full space in every class. The new diagonal then agrees with the leading total dimension over even weights [30, Theorem 1].

Corollary 6.13. *For each fixed $\lambda > 0$, Corollary 6.7 applies to the laws and limits in Corollary 6.6 and Theorem 6.11 within each fixed prime class. The implied constant may also depend on ℓ or M.*

Proof. Apply Theorem 6.1 and normalize using the mass bound of Lemma 6.2. The lemma gives $p\chi^2/T \ll \log^3 p$, so Proposition 6.9 with $J = 4$ gives the KL rate, including the removed primes as in Corollary 6.7. □

Corollary 6.14 (Prime-square level)**.** *In the units of Theorem 6.12, the full and new second moments at level ℓ^2 have constants*

$$\ell + 3 + s(\ell + 1)^2, \qquad \ell + s(\ell^2 - 2\ell - 1), \qquad s = \#\{u \bmod \ell : u^2 \equiv p \bmod \ell\}.$$

For odd ℓ, $s = 2$ in the square class and $s = 0$ in the nonsquare class. At level 4 the constants are 14 and 1 for all odd primes in the set of the theorem.

7 Estimates from the Dirichlet critical-line theorem

The trace asymptotics and leading variance use unconditional estimates. The quadratic case of Alpöge's Dirichlet critical-line theorem [2, 39] improves the Mellin truncation bound for quadratic characters of conductor at most $4p$. For $1 \le T \le X$ put

$$R_\varepsilon^*(T,X) = \frac{T}{X}\log^3(2X/T) + X^{-1/2+\varepsilon}\max(T,\sqrt{X})^{1/4}.$$

Proposition 7.1. *Let $p \ge 5$ be prime.*

1. *Theorem 3.1, Propositions 4.1 and 4.2, and Corollaries 4.3 and 4.5 hold with R_ε^* in place of R_ε.*
2. *For $2 \le K \le \sqrt{p}$ the two sums in (5) are $O_{\ell,\varepsilon}(K^2\log^3(2p/K^2) + p^{1/2+\varepsilon}\max(K,p^{1/4})^{1/2})$. For $K = p^a$ this saves a power of p over Deligne's bound, Petrow's bound and the angular large sieve when $5/24 < a < 1/2$, and the term $K^2\log^3$ dominates for $a > 1/3$.*

Proof. Alpöge's theorem and [16, Corollary 5.20] give $L(1/2+iv,\chi) \ll_\varepsilon (p(1+|v|))^\varepsilon$. The Mellin shift of Theorem 3.1 gives squared error p^ε/B, also with bounded omitted factors. Balance it with $p^\varepsilon B\sqrt{T}/p$ from Lemma 3.3 at $B = p^{1/2}T^{-1/4}$ and thin from T_0. Corollary 4.3 gives the trace remainder. It saves a power over K^3 for $a > 5/24$, over Petrow's bound in its range, and over K^2 for $a > 1/3$. □

Corollary S9.2 uses the full Dirichlet theorem to average the nonprincipal character terms over primes. It gives second terms for the sharp trace mean at $0 < \delta < 1/6$ and the count mean at $0 < \delta < 1/5$.

8 Open questions

Do the short-length limits hold for every prime sequence, and does $\mathcal{C}_A(c)/c \to 1/48$ as $c \downarrow 0$ for every fixed A? How far can the trace asymptotic, variance and mean formulas extend?

AI Declaration. OpenAI Codex and Anthropic Claude assisted with parts of the mathematical argument, including the derivation and checking of estimates in S5.4 and S8–S9. They also assisted with numerical checks and with drafting and revising the text. The author is responsible for the proofs and results.

Supplementary material

Main trace theorem	S5 gives the analytic estimates. S7 lists their inputs.
Fixed-level traces	S6 computes the local constants.
Variance and mean	S8 proves the variance for smooth cutoffs. S9 handles sharp cutoffs and second terms for the mean.
Related results	S1 treats extension fields. S2 studies other angle laws. S3 lists open questions and S4 gives numerical checks.

Section, result and equation numbers with the prefix S refer to the supplementary material. We use the article's notation Q_T for the beta-binomial law, $\|\cdot\|_Q$ for the weighted norm, and P^u for the unnormalized count measure. For a field of size q we use the same bias grid with step $\Delta_q = (4\sqrt{q})^{-1}$ and write

$$R_\varepsilon(T,q) = \frac{T}{q}\log^3(2q/T) + q^{-1/3+\varepsilon}\max(T,\sqrt{q})^{1/4}.$$

S1 Prediction over prime-power fields

Let $q = p^n$ with $p \geq 5$ and $n \geq 2$. Choose a nonsingular short Weierstrass equation uniformly over $\mathbf{F}_q$. The isomorphism classes then have weights proportional to $1/\#\operatorname{Aut}_{\mathbf{F}_q}(E)$. Let P_T be the count law for bias $f_t = 1/2 + t/(4\sqrt{q})$, where $t = q+1-\#E(\mathbf{F}_q)$, and P_T^{o} its restriction to ordinary curves $(p \nmid t)$, normalized to have mass one.

Lemma S1.1 (Restricted trace weights in fixed characteristic)**.** *Let $q = p^n$ with fixed p, let J be a fixed finite set of primes containing p, and let A be a fixed trace class modulo M. For $u = F^{(J)}1_A$ on $|t| < 2\sqrt{q}$ and $1 \leq T \leq q$,*

$$\|P^u - \gamma_B Q_T\|_Q^2 \ll_{M,J,\varepsilon} R_\varepsilon(T,q).$$

Here $F^{(J)}$ omits terms divisible by any prime of J, $B = q^{2/3}\max(T, \lceil\sqrt{q}(\log q)^8\rceil)^{-1/4}$ and γ_B is the constant Fourier coefficient of the restricted smooth cutoff.

Proof. Apply Lemma 3.4 with grid step Δ_q. Omitting the factors at J leaves moduli prime to q, so all the roots used in the expansion are units. The Fourier shifts from 1_A change denominators by factors bounded in terms of M. The conductor in the Mellin shift is at most $4q$, and the grid and smoothing estimates apply to the interior lattice, whether or not $2\sqrt{q}$ is integral. □

Theorem S1.2. *With absolute implied constants, for $1 \leq T \leq q$,*

$$D_{\mathrm{KL}}(P_T\|Q_T) \ll \sqrt{T/q}\,\log^4(2q/T) + q^{-1/100}\log^7 q,$$

and for $T \geq q$,

$$D_{\mathrm{KL}}(P_T\|Q_T) = \tfrac{1}{2}\log(T/q) + O\big(1 + \tfrac{p}{q}\log T\big).$$

As $q \to \infty$ through such prime powers, the divergence tends to zero if and only if $T = o(q)$. Moreover, for $1 \leq T \leq q$ and every $\varepsilon > 0$,

$$\chi^2(P_T^{\mathrm{o}}\|Q_T) \ll_\varepsilon R_\varepsilon(T,q) + q^\varepsilon/p.$$

For each fixed p the term q^ε/p can be omitted, with an implied constant depending on p and ε. In both cases $D_{\mathrm{KL}}(P_T\|Q_T)$ satisfies the same bound with an additional $O(q^{-1/2}\log(q+1))$.

Proof. For ordinary t, the number of equations is $(q-1)H(4q-t^2)/2$ [20, Lemma 2]. The supersingular probability η is at most $(p+23)/(12q)$, since there are at most $(p-1)/12+2$ supersingular j-invariants [38, Theorem V.4.1] and $q-1$ equations per j [20, Lemma 2]. The beta-binomial bound $Q_T(k) \gg (T+1)^{-3/2}$ and convexity give

$$D_{\mathrm{KL}}(P_T\|Q_T) \le \log(1+\chi^2(P_T^{\mathrm{o}}\|Q_T)) + \eta\big(\tfrac{3}{2}\log(T+1)+O(1)\big).$$

Short observations. Set $F(t) = \pi H(4q-t^2)/\sqrt{4q-t^2}$ and take the coefficients of Lemma 3.2 with q in place of p. The ordinary masses are $\Delta_q\sigma(f_t)F(t)/(1-\eta)$. For $p \nmid t$, the two simple roots modulo p of x^2-tx+q lift to all powers of p, so $b_t(p^j)=1$. Define $\tilde{b}_t(n) = b_t(n_0)$ for all traces, where n_0 is the prime-to-p part of n, and let $\tilde{F}_N(t) = \sum_{n\le N}\tilde{b}_t(n)/n$. For a reduced a/d with $p\nmid d$, the coefficient of $e(at/d)$ in $\sum_{A<n\le B}\tilde{b}_t(n)/n$ is

$$\frac{S(a,aq;d)}{\varphi(d)}\sum_{j\ge 0}p^{-j}\sum_{\substack{A/p^j<m\le B/p^j\\ p\nmid m}}\frac{h_d(m)}{m} \ll \tau(d)d^{-3/2}\min(1,\sqrt{d/A}).$$

Here the minimum is 1 for $A=0$. The estimate uses the coefficient tail of Lemma 3.2 and $\sum_j p^{-j/2} \ll 1$. The constant coefficient of $\tilde{F}_N$ is $(1-1/p)^{-1}+O(N^{-1/2})$. In the proof of Lemma 3.2 the middle range may be summed over all traces, and the Burgess estimate on ordinary traces has conductor at most $4q$. This gives, for $1 \le N \le q^{1/100}$,

$$\sum_{\substack{|t|<2\sqrt{q}\\ p\nmid t}}|F(t)-\tilde{F}_N(t)|^2 \ll \sqrt{q}\left(\frac{\log^4(2N)}{N}+q^{-1/100}\log^7 q\right).$$

For either the ordinary cutoff $U=\tilde{F}_N$ or a smooth cutoff used below, separate the terms by writing

$$F1_{p\nmid t} = (F-U)1_{p\nmid t} + U - U1_{p\mid t}.$$

For $U=\tilde{F}_N$, (1) bounds the first term. On the full grid and on the subgrid $p\mathbf{Z}$, unimodal Riemann sums give the constant terms in the proportions 1 and $1/p$, with errors bounded by $O(\Delta_q)$ times the posterior maximum. Their difference gives $Q_T + O(N^{-1/2})Q_T$ up to these grid errors. Abel summation bounds a nonzero mode's contribution to count k by $O(\Delta_q Q_T(k) m_{T,k}/\|a/d\|)$, with ap/d in place of a/d on $p\mathbf{Z}$. Multiplication by p permutes the reduced numerators. Summing the coefficient bound and $\sum^*_{a \bmod d}\|a/d\|^{-1} \ll d\log(2d)$ therefore costs $O((T/q)N\log^4(2N))$ in squared norm on both grids. Take $N = \lfloor \min(\sqrt{q/T}, q^{1/100})\rfloor$ and center at $(1-\eta)Q_T$ before normalizing. This gives the claimed short-length bound for $\chi^2(P_T^{\mathrm{o}}\|Q_T)$, and $p/q \le q^{-1/2}$ makes the supersingular contribution $O(q^{-1/2}\log(q+1))$ for $T \le q$.

Long observations. For $T \ge q$, (1) gives $1+\chi^2(P_T^{\mathrm{o}}\|Q_T) \ll 1+\sqrt{T/q}$, while Lemma 2.2 uses at most $4\sqrt{q}+1$ biases for the lower bound. When n is even, the two endpoints have combined mass $(p-1)/(12q)$ [20, Lemma 2]. Their biases are 0 and 1, with KL divergence $\frac{3}{2}\log(T+1)+O(1)$ from Q_T. The mixture argument of Lemma 2.2 then strengthens the lower bound to $\frac{1}{2}\log(T/q)+\frac{p-1}{12q}\log T - O(1)$. The endpoints also contribute to the squared norm at order $(p/q)^2T^{3/2}$. For any fixed $c>0$, the cosine test of Theorem 3.1 detects the integer trace grid when $T \ge cq$, proving necessity of $T=o(q)$.

Sharper estimates for ordinary curves. Now use $U=\sum_n \tilde{b}_t(n)n^{-1}\psi(n/B)$ with the smooth cutoff of Theorem 3.1. On ordinary traces $\tilde{b}_t = b_t = c_t * \chi$, with χ of conductor at most $4q$. The Mellin shift and hybrid Weyl bound give $|F-U| \ll_\varepsilon q^{1/6+\varepsilon}B^{-1/2}$, which costs $O_\varepsilon(q^{1/3+\varepsilon}/B)$ in squared norm. The Fourier formula with the cutoff inserted gives coefficients $O(\tau(d)d^{-3/2})$ for $p\nmid d \le 2B$.

On the interior grid $U = O_\varepsilon(q^\varepsilon)$, since $4q - t^2 \neq 0$ and the sum of $|b_t(m)|/m$ over $m \le 2B$ prime to p has that bound. Thus (2) applies with q in place of p, and (1) bounds the term $U1_{p|t}$ in squared norm by $O_\varepsilon(q^\varepsilon \Delta_q(\sqrt{q}/p + 1))$. Take $B = q^{2/3}T^{-1/4}$, center at the mass of the ordinary law, and normalize by $1 - \eta$. Below the range of (2), contract by keeping only T flips as in Theorem 3.1.

For fixed p, apply Lemma S1.1 with $J = \{p\}$ to the nonzero trace classes modulo p, since $F1_{p\nmid t} = (1 - 1/p)^{-1}F^{(p)}1_{p\nmid t}$. Convexity then gives the two asserted KL bounds. □

Corollary S1.3 (Legendre family over $\mathbf{F}_q$)**.** *Fix a prime $p \ge 5$. For $q = p^n$ sample λ uniformly in $\mathbf{F}_q \setminus \{0, 1\}$ and use the trace of $y^2 = x(x-1)(x-\lambda)$ over $\mathbf{F}_q$ for the bias $1/2 + t/(4\sqrt{q})$. The count law of the ordinary parameters has chi-square divergence $O_{p,\varepsilon}(R_\varepsilon(T, q))$ from Q_T for $1 \le T \le q$. As $n \to \infty$ the logarithmic loss of the family tends to zero if and only if $T = o(q)$, and for $T \ge q$ it is $\frac{1}{2}\log(T/q) + O_p(1 + (p/q)\log T)$.*

Proof. The prime-field case is Corollary 4.5. The $Y_1(4)$ isogeny remains defined over $\mathbf{F}_q$, so (6) counts rational points of order four there. At ordinary traces the formulas for full rational 2- and 4-torsion [20, Lemmas 2–4] and the local identities at 2 express the weights in terms of F and $F^{(2)}$. Remove the factor at p as in Theorem S1.2, then apply Lemma S1.1 with $J = \{p\}$ or $\{2, p\}$ and moduli dividing $16p$. The supersingular parameters are roots of a polynomial of degree $(p-1)/2$ [38, Theorem V.4.1], so the ordinary probability is bounded below and the additional KL cost is $O((p/q)\log(T+1))$. This yields the short-length result.

Each curve has at most twelve points of order four, so the ordinary Legendre count law is bounded by a fixed multiple of the full ordinary law. Its squared norm for $T \ge q$ is $O_p(1 + \sqrt{T/q})$. For the lower bounds, all traces lie in one class modulo 4. The support has at most $\sqrt{q} + 1$ points and the cosine test at frequency $2\pi\sqrt{q}$ detects the lattice for $T \ge cq$. □

S1.1 Bounds from the Dirichlet critical-line theorem

The quadratic case of Alpöge's theorem [2, 39] allows us to replace $R_\varepsilon(T, q)$ in Theorem S1.2 and Corollary S1.3 by

$$R^*_\varepsilon(T, q) = \frac{T}{q}\log^3(2q/T) + q^{-1/2+\varepsilon}\max(T, \sqrt{q})^{1/4}.$$

Apply the balance and thinning in Proposition 7.1 with q in place of p. The Mellin shift gives the same squared tail $O_\varepsilon(q^\varepsilon/B)$ with fixed Euler factors omitted. Add the varying-characteristic and supersingular terms of Theorem S1.2.

S2 Hecke angles, Kloosterman angles and general families

Proposition 2.1 bounds loss using moments up to order $\sqrt{T}$, and Lemma 2.2 gives lower bounds from the support size.

S2.1 Gaussian Hecke angles

Let $\xi((\alpha)) = (\alpha/|\alpha|)^4$ over $\mathbb{Q}(i)$, and draw a prime ideal of norm at most X with weight $w_\mathfrak{p} = \log N\mathfrak{p}/\Theta(X)$, where $\Theta(X) = \sum_{N\mathfrak{p}\le X}\log N\mathfrak{p}$. Use the bias $(1 + \Re\xi(\mathfrak{p}))/2$. Write $A_X^{(T)}$ for this word law. The reference is $B_{\mathrm{KT}}^{(T)} = B_{1/2,1/2}^{(T)}$. The moments of the Hecke-angle law in Proposition 2.1 are $\sqrt{2}\Re M_m(X)$, where $M_m(X) = \sum_\mathfrak{p} w_\mathfrak{p}\xi(\mathfrak{p})^m$. Write $\log^+ x = \max(0, \log x)$.

Theorem S2.1. *For sufficiently large X, unconditionally,*

$$\tfrac{1}{2}\log^+\big(T(\log X)^2/X^2\big) - O(1) \le D_{\mathrm{KL}}(A_X^{(T)}\|B_{\mathrm{KT}}^{(T)}) \le \tfrac{1}{2}\log^+(T/X^2) + \log\log X + O(1).$$

Thus for $T \ge X^2$ the loss is $\frac{1}{2}\log(T(\log X)^2/X^2) + O(1)$. If $1/2 < \beta < 1$ and all $L(s,\xi^m)$, $m \ge 1$, are zero-free in $\Re s > \beta$, then

$$D_{\mathrm{KL}}(A_X^{(T)}\|B_{\mathrm{KT}}^{(T)}) \ll_\beta \sqrt{T}\, X^{2\beta-2}\log^4(X(2+T)).$$

In particular the loss tends to zero for $T \le X^{4(1-\beta)-\varepsilon}$ with fixed $\varepsilon > 0$.

Proof. Inert and ramified prime ideals have total weight $O(X^{-1/2})$. For distinct split prime ideals, unique factorization prevents their generators from having $\alpha\bar{\gamma}$ equal to a unit times a rational integer. After rotation by a unit, its argument lies in $[-\pi/4, \pi/4]$ and $|\Im(\alpha\bar{\gamma})| \ge 1$. Since $|\alpha\bar{\gamma}| \le X$, the angles $4\arg\alpha$ are spaced $\gg 1/X$. The dual large sieve [31] bounds the squared moment sum over an interval of L modes by $(L+X)\sum_{\mathfrak{p}} w_{\mathfrak{p}}^2 \ll (L+X)\log X/X$. Apply this on blocks of length $\lceil\sqrt{T+1}\rceil$ in Proposition 2.1. Including the inert and ramified ideals gives $\chi^2 \ll \log X(1+\sqrt{T}/X)$. Take $D_{\mathrm{KL}} \le \log(1+\chi^2)$. The support bound uses $O(X/\log X)$ prime ideals.

For the zero-free estimate the finite conductor is fixed and the archimedean parameter is $O(m)$, giving $O(\log(2+m+|t|))$ zeros per unit height interval. In the explicit formula [36, pp. 282–285] use a half-integer cutoff and height $H \asymp X$ at distance $\gg 1/\log(X+m+2)$ from the zero ordinates. Moving to $(1-\beta)/2$ costs $O_\beta((X^{(1-\beta)/2} + X/H)\log^2(X+m+2))$ on the vertical and horizontal contours. The zeros contribute $O_\beta(X^\beta \log^2(X+m+2))$ and the prime powers $O(\sqrt{X}\log^2 X)$. Thus $M_m(X) \ll_\beta X^{\beta-1}\log^2(X(2+m))$. Sum the squares with the weights of Proposition 2.1. □

Proposition S2.2. *For every $\varepsilon > 0$, the Gaussian Hecke loss tends to zero unconditionally and uniformly for $1 \le T \le X^{6/5-\varepsilon}$.*

Proof. Choose $0 < \tau < 3/5$ and $0 < c < 1$ with $2c\tau > 6/5 - \varepsilon$. Add a nonnegative smooth compactly supported radial weight $\Phi(N\mathfrak{p}/Y)$ to the moment, denoting the normalized result by $M_m^\Phi(Y)$. For a fixed smooth even angular test g of compact support with $\widehat{g}(0) > 0$, Parseval and the sector variance bound [15, Corollary 5] give $\sum_{m\ne 0}|\widehat{g}(m/Y^\tau)|^2|M_m^\Phi(Y)|^2 = o(1)$. Thus the unweighted square sum up to δY^τ is $o(1)$ for some fixed $\delta > 0$. In Proposition 2.1 the remaining modes cost $O(T\exp(-CY^{2\tau}/T))$ for some $C > 0$, using $|M_m^\Phi| \le 1$. The smoothed divergence is $o(1)$ uniformly for $Y \ge X^c$.

For the interval $(Y/2, Y]$ take $\Phi \ge 1$ on $[1/2, 1]$ and $\int \Phi \le (1+\eta)/2$. The prime ideal theorem bounds the smooth to sharp mass ratio by $1 + \eta + o(1)$. This also bounds the ratio of the sharp likelihood to the smooth likelihood, so the squared norm of the sharp likelihood has upper limit at most $(1+\eta)^2$. Let $\eta \to 0$ and combine the intervals down to $Y_0 \asymp X^c$ by convexity. The primes below Y_0 have mass $O(X^{c-1})$ and divergence $O(\log X)$ by the large sieve bound since $\sqrt{T} = o(Y_0)$. □

S2.2 Kloosterman angles

For an odd prime p, choose c uniformly in $\mathbf{F}_p^\times$, and let $\theta_c \in [0, \pi]$ satisfy

$$\sum_{x\in\mathbf{F}_p^\times} e_p(x + c/x) = 2\sqrt{p}\cos\theta_c.$$

Here $e_p(t) = \exp(2\pi i t/p)$ for an integer representative of $t \in \mathbf{F}_p$. Use bias $(1+\cos\theta_c)/2$ and write $A_p^{(T)}$ for its word law. The reference is $B_{\mathrm{ST}}^{(T)} = B_{3/2,3/2}^{(T)}$.

Lemma S2.3 (One bias against the Sato–Tate reference)**.** *If $D_T(f)$ is the KL divergence of T flips of bias f from $B_{\mathrm{ST}}^{(T)}$, then uniformly for $f \in [0,1]$,*

$$D_T(f) = \tfrac{5}{2}\log(T+1) - \log\big((1+Tf)(1+T(1-f))\big) + O(1).$$

Proof. For the reference count law, the gamma ratios give

$$Q_T(k) \asymp (T+1)^{-2}\sqrt{(k+1)(T-k+1)}.$$

Apply the two expected logarithm estimates and the entropy estimate in the proof of Lemma 2.2. The identity $(1+Tf)(1+T(1-f)) = T+1+Tv$, where $v = Tf(1-f)$, gives the stated form. □

Theorem S2.4. *For $T \geq 1$, with absolute constants,*

$$D_{\mathrm{KL}}(A_p^{(T)} \| B_{\mathrm{ST}}^{(T)}) \leq \chi^2(A_p^{(T)} \| B_{\mathrm{ST}}^{(T)}) \ll (T+1)^{3/2}/p,$$
$$D_{\mathrm{KL}}(A_p^{(T)} \| B_{\mathrm{ST}}^{(T)}) \geq \tfrac{1}{2}\log(T+1) - \log(p-1) - O(1).$$

For $p^{2/3} \leq T \leq e^{\sqrt{p}}$,

$$D_{\mathrm{KL}}(A_p^{(T)} \| B_{\mathrm{ST}}^{(T)}) \leq \tfrac{1}{2}\log(T/p^{2/3}) + O(1).$$

In particular, the loss tends to zero for $T = o(p^{2/3})$ and diverges if $T/p^2 \to \infty$.

Proof. The mth moment is $(p-1)^{-1}\sum_c U_m(\cos\theta_c)$. Katz bounds the sum by $(m+1)\sqrt{p}/2$ [21, Example 13.6]. Indeed, it is, up to the factor $\pm p^{-m/2}$, the trace sum of $\mathrm{Sym}^m \mathrm{Kl}_2$ on $\mathbb{G}_m$. For $m \geq 1$, the SL_2 monodromy gives $H_c^2 = 0$ [21, Theorem 11.1]. The space H_c^1 has dimension at most $(m+1)/2$ [12, Theorem 0.2] and weights at most $m+1$ [10]. Proposition 2.1 sums the squared moment bound, and Lemma 2.2 gives the lower bound since the support has at most $p-1$ points.

For the intermediate range put $S = \lfloor p^{2/3} \rfloor$. The divergence χ^2 at length S is bounded, so the true count mass $a_S(k)$ satisfies $a_S(k) \leq b_S(k) + \sqrt{b_S(k)\chi^2}$ for the reference mass $b_S(k)$. At 0 and S, $b_S(k) = O(S^{-3/2})$. Therefore $a_S(k) = O(p^{-1/2})$ at these counts, and $(1-1/S)^S \geq 1/4$ bounds the bias mass of $\min(f, 1-f) < 1/S$ by $\eta = O(p^{-1/2})$. Outside this set both ratios $(1+Tf)/(1+Sf)$ and $(1+T(1-f))/(1+S(1-f))$ are at least $T/(2S)$. Everywhere they are at least one. Lemma S2.3 now gives $\mathbb{E}(D_T(f) - D_S(f)) \leq \frac{1}{2}\log(T/S) + O(1) + 2\eta\log(T/S)$. The mixture loss at T is $\mathbb{E}D_T(f) - I(f; W_T)$, where W_T is the length-T word. Since the information is nondecreasing in T and the loss at S is bounded, the same upper bound holds for the loss at T. Its last term is bounded for $T \leq e^{\sqrt{p}}$. □

The normalized Gauss sums of nontrivial characters give the same vanishing-loss range $T = o(p^{2/3})$ against the uniform-angle reference. Orthogonality gives hyper-Kloosterman sums with moments $O(m/\sqrt{p})$ [21, Scholium 4.0.1 and Theorem 4.1.1]. The Wasserstein estimates of Kowalski–Untrau and Rahimi [25, 35] also imply total-variation convergence in this range, using affinity $\cos^T((\theta-\phi)/2)$ for the product laws at angles θ, ϕ. Proposition 2.1 controls logarithmic loss.

S2.3 One-parameter families of elliptic curves

Let $\mathbf{F}_q$ have characteristic at least 5. Let U be an open subset of $\mathbb{P}^1$ over $\mathbf{F}_q$ whose complement has r geometric points. Let $E \to U$ be an elliptic curve with nonconstant j-invariant. For $n \geq 1$ choose x uniformly in $U(\mathbf{F}_{q^n})$, put $a_x = q^n + 1 - \#E_x(\mathbf{F}_{q^n})$, and use the bias $1/2 + a_x/(4q^{n/2})$. Write $A_{E,n}^{(T)}$ for the word law.

Proposition S2.5. *If $q^n \geq 2r$ then, with absolute constants,*

$$\chi^2(A_{E,n}^{(T)} \| B_{\mathrm{ST}}^{(T)}) \ll (r-2)^2 (T+1)^{3/2} q^{-n},$$
$$D_{\mathrm{KL}}(A_{E,n}^{(T)} \| B_{\mathrm{ST}}^{(T)}) \geq \tfrac{1}{2} \log(T+1) - \tfrac{n}{2} \log q - O(1).$$

For fixed r, the moment bound gives vanishing loss at $T = o(q^{2n/3})$, and the trace lattice gives divergence when $T/q^n \to \infty$. For Legendre, the class-number calculation in Corollary S1.3 closes the gap in fixed characteristic.

Proof. The fibre cohomology sheaf $\mathcal{F}$ has rank two and weight one. Nonconstant j gives an affine base and geometric monodromy SL_2 [10, Lemme 3.5.5]. Semistable reduction after an extension of degree dividing 24 gives tameness at the r missing points. In particular $r \geq 3$, since the tame fundamental group of the affine line is trivial and that of $\mathbb{G}_m$ is abelian. For $m \geq 1$, $\mathrm{Sym}^m \mathcal{F}$ is geometrically nontrivial and irreducible. The Euler–Poincaré formula [21, 2.3.1] gives $\dim H_c^1 = (m+1)(r-2)$, with the other compactly supported cohomology zero. Its weights are at most $m+1$ [10, Théorème 3.3.1], so the trace formula gives

$$\Big| \sum_{x \in U(\mathbf{F}_{q^n})} U_m(a_x/(2q^{n/2})) \Big| \leq (m+1)(r-2) q^{n/2}.$$

Divide by $|U(\mathbf{F}_{q^n})| \geq q^n/2$ and apply Proposition 2.1. The lower bound follows from Lemma 2.2 because there are at most $4q^{n/2}+1$ integer traces. □

S3 Open questions for the supplementary families

The short-length constant for ordinary curves over $\mathbf{F}_{p^n}$, $n \geq 3$, and its dependence on p and n remain open. Stronger sector variance bounds could improve the Gaussian Hecke range. For Kloosterman angles and one-parameter families, weighted square sums of Frobenius moments through degree $\sqrt{T}$ could narrow the unresolved intervals $[p^{2/3}, p^2]$ and $[q^{2n/3}, q^n]$, respectively.

S4 Numerical checks

The ancillary file `paper_checks.json` records prediction losses and checks of local constants. The variance samples are in `variance_check.json`. The shortest run has cutoffs $7, 9, 12$ and therefore reaches at most the first nonzero level-one cusp-form space. In samples near 10^5 and 10^6, comparable lengths T give comparable variances even when T/p differs by a factor of ten, leaving the leading term in Theorem S8.1 unresolved at these sizes.

S5 Estimates for the short-length profile

We use the notation of Lemma 6.2. Let g be a trace weight and g_B its smooth cutoff. Write γ_B for the cutoff Fourier coefficients, extended with period one, and Z_p, Z_B for the masses of P^g, P^{g_B}.

Lemma S5.1 (The averaged class-number tail)**.** *For $0 < \varepsilon < 1$, outside $O_{g,\varepsilon}(X^{1-\varepsilon/4})$ primes in $(X, 2X]$, the bound*

$$\Delta_p \sum_{t^2 < 4p} |g(t) - g_B(t)|^2 + \|P^g - P^{g_B}\|_Q^2 \ll_{g,\varepsilon} p^{\varepsilon}/B \tag{S1}$$

holds for all $B \geq 1$ and $1 \leq T \leq p$.

Proof. With $-D_t = d_t r_t^2$, the discriminant count and Heath–Brown bound in Section 6 of the article give $\sum_{X<p\le 2X} A(p) \ll_\varepsilon X^{1+\varepsilon}$ for

$$A(p) = \Delta_p \sum_{t^2<4p} \int_{\mathbf{R}} |L(1/2+iv,\chi_{d_t})|^2 (1+|v|)^{-10}\, dv.$$

Using exponent $\varepsilon/4$ and Markov gives $A(p) \le p^{\varepsilon/2}$ outside $O_\varepsilon(X^{1-\varepsilon/4})$ primes. Shift the cutoff integral to $\Re s = 1/2$ as in Section 3, crossing the full weight at $s=1$. On the new line the finite divisor factor is $\ll_\varepsilon \sum_{ij|r_t} i^{-1/2} \ll_\varepsilon p^{\varepsilon/4}$, and the fixed omitted factors are bounded. The rapid Mellin decay and Cauchy–Schwarz give squared grid error $O_{g,\varepsilon}(B^{-1}p^{\varepsilon/2}A(p))$. The grid inequality of Section 3.2 transfers it to the count norm, since $1+\Delta_p\sqrt{T} = O(1)$ for $T \le p$. □

The angular integral. Set $h = 1/U$, $G_h(z) = \sum_j e^{-hj^2/2} e^{ijz}$, $\theta_t = \arccos(t/(2\sqrt{p}))$ and $w_\phi = 2\sqrt{p}\sin\phi$. For the angular sum $\mathcal{D}$ in the proof of Lemma 6.2, use the norm $\|f\|_{\mathrm{ang}}^2 = (8\pi^3 p)^{-1}\int_0^\pi |f(\phi)|^2\, d\phi$. The sine coefficient calculation there gives

$$\|\mathcal{D}\|_{\mathrm{ang}}^2 = \sum_{m\ge 1} e^{-h(m+1)^2} m_{g_B,p}(m)^2 + e^{-h}|Z_B - \gamma_B(0)|^2. \tag{S2}$$

Lemma S5.2 (Lengths above the endpoint)**.** *Fix $0<\delta_0<1/2$. For some $\xi,\nu>0$, outside $O_{g,\delta_0}(X^{1-\xi})$ primes in $(X,2X]$, uniformly for $p^{-\delta_0} \le c = T/p \le 1/2$,*

$$\|P^g - Z_p Q_T\|_Q^2 = \mathcal{C}_{A_p}^g(c) + O_{g,\delta_0}(c^{3/2} + cp^{-\nu}).$$

On these primes Z_p differs from the complete constant Fourier coefficient by $O_{g,\delta_0}(p^{-\nu})$.

Proof. On $X < p \le 2X$, put

$$D_0 = X^{1/4}, \qquad B = X^{(1/2+\delta_0)/2}, \qquad \rho = (1/2-\delta_0)/2 > 0.$$

Take ε small relative to ρ. The squared tail (S1) divided by c is $O(X^{-\rho+\varepsilon})$. For the mass, sum the Abel bound $O((\sqrt{p}\|\beta\|)^{-1})$ at nonintegral frequencies against γ_B, and add the $O_g(p^{-1/2})$ Riemann-sum error at zero. This gives $|Z_B - \gamma_B(0)| \ll_{g,\varepsilon} p^{-1/2}B^{1/2+\varepsilon}$, whose square saves a power since $B/(pc) \ll X^{-3\rho}$. Only fixed denominators shift to zero, where the coefficient tail is $O_g(B^{-1/2})$. Equation (S1) gives the claim for Z_p.

Use $h = 1/T$. On $w_\phi < \sqrt{T}$ only traces in two intervals of length $O(T/\sqrt{p})$ contribute appreciably, since T^2/p grows by a power. The edge argument of Lemma 3.3 bounds the squared norm by $O(\sqrt{T}/p)$ times the local sum of $|g_B(t)|^2$; the subtracted sine costs $O_g(c^{3/2})$. For $c \ge p^{-\rho}$, Lemma 3.3 with the constant included gives $O_{g,\varepsilon}(c^{3/2} + p^\varepsilon B\sqrt{T}/p)$, with $B/\sqrt{T} \ll X^{-\rho/2}$. For $c < p^{-\rho}$, use $|g_B(t)| \ll_{g,\varepsilon} p^\varepsilon$ to get $O_{g,\varepsilon}(p^\varepsilon(c^{3/2} + \sqrt{T}/p))$. Thus the edge costs $O_{g,\delta_0}(c^{3/2} + cp^{-\nu})$.

On $w = w_\phi \ge \sqrt{T}$, the Taylor and Poisson calculation in the article gives, with $s = 2\sqrt{p}\cos\phi$,

$$\mathcal{D}(\phi) = 2\pi w \sum_{\beta\ne 0} \gamma_B(\beta) e(\beta s) e^{-2\pi^2 w^2\beta^2/T} + E(\phi), \qquad p^{-1}\int |E|^2 \ll_{g,\varepsilon} p^\varepsilon/T.$$

The reflection is exponentially small and the zero mode cancels up to $O_g(w/T)$. The squared error saves a power relative to c since $(Tc)^{-1} \le p^{2\delta_0-1}$.

Averaging the large denominators. Expand the selectors into fixed rational shifts α. Lemma 3.2 gives the coefficients $C_d S(a,ap;d)$ of $\beta = a/d+\alpha$, $(a,d)=1$, with $|C_d| \ll_g d^{-2}$ and $d \le 2B$.

The scalars C_d are fixed on the prime block. Split by d before shifting. The low frequencies have reduced denominators $O_g(D_0)$, and Lemma 3.3 bounds their mixed term by $O_{g,\varepsilon}(p^\varepsilon D_0\sqrt{T}/p)$, with $D_0/\sqrt{T} \ll X^{-\rho}$.

Fix dyadic bands $T \asymp U$ and $w \asymp W$ meeting $w \ge \sqrt{T}$. Put $H = W/\sqrt{U}$, so $1/2 \le H \ll X^{\delta_0/2}$ and $d/H \gg 1$ for $d > D_0$. Write the Gaussian as $e^{-vH^2\beta^2}$ with v in a fixed compact positive interval. Open $S(a, an; d) = \sum^*_{b \bmod d} e((a^2\bar{b} + bn)/d)$. The coefficient at frequency b/d in n is

$$C_d e(\alpha s) \sum_{\substack{a\in\mathbf{Z},\ (a,d)=1\\ a/d+\alpha\neq 0}} e(a^2\bar{b}/d)e(as/d)e^{-vH^2(a/d+\alpha)^2}.$$

After removing $C_d e(\alpha s)$, Parseval bounds the square mean over a period by $O(d/H)$, also for the v derivative. Since $d \le 2B \ll \sqrt{X}$, the additive large sieve [32, Theorem 1] and integration over $|s| \le 3\sqrt{X}$ bound the integer sum of the high squared integrals by

$$\ll_{g,\varepsilon} (X + B^2)\sqrt{X} \sum_{d>D_0} d^{-4} d\frac{d}{H} \ll_{g,\varepsilon} \frac{X^{3/2}}{HD_0}.$$

The derivative bound and the fundamental theorem of calculus allow the supremum over v. The angular factor is $O(W/X)$ since $w^2 d\phi = w|ds|$. Restricting to the target prime class, summing the W bands and dividing by $X(U/X)$ gives $O_{g,\varepsilon}(X^{-\rho+\varepsilon})$. Sum over shifts by the triangle inequality. Markov over the $O(\log X)$ length bands gives $O_{g,\delta_0}(cp^{-\nu})$ outside a power-saving exceptional set.

Comparison with the diagonal. For $d > D_0$, the shifted Gaussian sum is $O(d\sqrt{T}/w)$ since $w/\sqrt{T} \ll X^{\delta_0/2}$. Weil's bound gives high diagonal $O_{g,\varepsilon}(p^\varepsilon\sqrt{c}/D_0)$ after integration. Cauchy–Schwarz also controls coincidences between low and high modes of different shifts. The low diagonal is $O_g(c\log^3(2/c))$ by Lemma 3.3. Combining the bounds identifies the bulk heat norm with the diagonal up to $O(cp^{-\nu})$. The coefficient square sum modulo one is bounded, so the translate sum $O(1 + \sqrt{c/(y(1-y))})$ at $y = (1+\cos\phi)/2$ bounds the diagonal edge by $O_g(c^{3/2})$. The cutoff-to-complete squared coefficient error is $O_g(B^{-1}\log^4(2B))$ by Weil and the h_d tail. The same bound holds with fixed shifts and after angular integration, so Cauchy–Schwarz saves a power since $(Bc)^{-1} \ll X^{-\rho}$. Here $w^2 = 16py(1-y)$ and $w^2|d\phi|/(2\pi p) = \sigma(y)|dy|$ give the diagonal $\mathcal{C}^g_{A_p}(c)$.

The multiplier comparison in the proof of Lemma 6.2 has relative error $O(T^{-1/2}\log^3 p)$ against the heat bound $O_g(c\log^3(2p))$. The remaining terms are negligible by Gaussian decay and $|m_{g_B,p}(m)| \ll_{g,\varepsilon} p^\varepsilon(m+1)$. Project off the first sine in (S2) and use (S1) and Cauchy–Schwarz to restore the full centered count measure.

□

Lemma S5.3 (Reduction at short weight cutoffs)**.** *Fix $\lambda > 0$, $1/6 < a \le 1/4$ and a sufficiently small $\omega > 0$. Put $U = \lambda p^{2a}$, $c = U/p$, $B = X^{3/4-a/2}$ and $D_+ = X^{1/2-a+\omega}$. In* (S2) *use $h = 1/U$. Write V_+ for the nonzero-frequency Gaussian sum*

$$V_+(\phi) = 2\pi w \sum_{\beta\neq 0} \gamma_B^+(\beta)e(\beta s)e^{-2\pi^2 w^2\beta^2/U}, \qquad w = 2\sqrt{p}\sin\phi, \quad s = 2\sqrt{p}\cos\phi.$$

Here γ_B^+ includes only the terms with original denominator $d \le D_+$ before the shifts by the trace selectors. Let $\chi_X(\phi) = \chi(X^\omega w/\sqrt{X})$, where χ is smooth with $0 \le \chi \le 1$, equal to zero up to one and to one from two onwards. For every fixed $J > 0$, outside $O_{g,\lambda,a,\omega,J}(X(\log X)^{-J})$ primes in $(X, 2X]$,

$$\|\mathcal{D} - \chi_X V_+\|^2_{\mathrm{ang}} + |Z_B - \gamma_B(0)|^2 \ll_{g,\lambda,a,\omega,J} c(\log X)^{-J}.$$

The same assertion holds at $a = 1/6$ with $B = X^{2/3-\rho}$ for any sufficiently small fixed $\rho > 0$.

Proof. The angular endpoints. Take $W_0 = X^{1/2-a+\nu}$ with small $\nu > 0$, and use the finitely many shifted components $\beta = m/d + \alpha$ of S5.2. Near $\phi = 0$ with $w < W_0$, put $z = 2\sqrt{p}$, $L_p = \sqrt{p}/U$, $\eta = \sqrt{U}\phi \ll X^{\nu}$ and $y = (z-t)/L_p$. The odd kernel equals $\sqrt{2\pi U}\, F_{\eta,U}(y)$ up to exponentially small tails, where

$$F_{\eta,U}(y) = \zeta(y/U)\big(e^{-(Q_U(y)-\eta)^2/2} - e^{-(Q_U(y)+\eta)^2/2}\big), \qquad Q_U(y) = \sqrt{U}\arccos(1 - y/(2U)).$$

Extend by zero off $[0, U/2]$ and use a smooth ζ equal to one up to $1/4$ and zero from $1/2$ onwards. Then $Q_U(y) = \sqrt{y}\, q(y/U)$ with q smooth and positive there. The other edge uses $-t$ and $\pi - \phi$.

The Fourier transforms of $y^r(U\partial_U)^i F_{\eta,U}$ for $r, i \in \{0,1\}$ are $O((1+\eta)^C(1+|\xi|)^{-3/2})$. Oddness in Q_U makes the derivatives of orders $j \le 2$ on $0 < y \le 1$ of size $O((1+\eta)^C y^{1/2-j})$. For $y \ge 1$ their L^1 norms grow polynomially in $1+\eta$ by Gaussian decay. Split at $|\xi|^{-1}$ and integrate twice by parts beyond it. Poisson summation therefore gives the multiplier $\sqrt{2\pi U} L_p e(\beta z)\widehat{F}_{\eta,U}(L_p\beta)$. The continuous integral at $\beta = 0$ is $2\pi w e^{-1/(2U)}$ up to the omitted exponential tails, so the zero mode cancels.

Put $L = X^{1/2-2a}$ and $N \asymp \sqrt{X}L$. An interval $z = z_0 + vL$, $0 \le v \le 1$, corresponds to $O(N)$ consecutive integers, and there are $O(X/N)$ such intervals. Hold v and η fixed in the kernel and vary the integer n in $S(m, mn; d)$. For $D \le d < 2D$ and $R \le 1 + L|\beta| < 2R$ with R dyadic, write $A_m = e(\beta z)\widehat{F}_{\eta,U}(L_p\beta)$. Its bound is $O((1+\eta)^C R^{-3/2})$, and differentiation in v costs at most $O(R)$, since U and L_p have logarithmic derivatives $O(X^{-2a})$ in v.

On a nonempty band the numerators lie in two intervals of total length $O_g(M_0)$, $M_0 = dR/L \gg_g 1$, since $|\beta| \gg_g 1/d$. Open the sum in the integer n. Ramanujan summation gives

$$\sideset{}{^*}\sum_{u \bmod d} \Big| \sum_m A_m e(m^2\bar{u}/d)\Big|^2 \ll_\varepsilon d^\varepsilon(dM_0 + M_0^2)\max_m |A_m|^2.$$

Indeed $|c_d(k)| \le \sum_{r|(d,k)} r$, and for each $r \mid d$ the unit congruence $m'^2 = m^2 \bmod r$ has $O_\varepsilon(d^\varepsilon)$ classes, each with $O_g(1 + M_0/r)$ terms. The additive large sieve for u/d costs $O(N + D^2)$. The supremum in v costs a single R by $\sup|f|^2 \ll \|f\|_2^2 + \|f\|_2\|f'\|_2$. Including $|C_d|^2$, the squared sum on the integer interval is

$$\ll (N + D^2)(1+\eta)^C D^{-1+\varepsilon}\big((LR)^{-1} + L^{-2}\big).$$

For $R \le X^{C_1}$, sum the brackets using $R \gg_g \max(1, L/D)$ to get $O_g(L^{-1}\min(1, D/L)\log X)$, with logarithmic cost by Cauchy–Schwarz. The $R^{-3/2}$ decay and the trivial Kloosterman bound handle $R > X^{C_1}$ for large C_1.

Enlarge the angular integral to a common interval $0 \le \eta \le C_2 X^{\nu}$ before averaging the integers. The squared prefactor in the angular norm, including $d\phi = d\eta/\sqrt{U}$, is $O(U^{-3/2})$. Thus the mean squared edge norm of a denominator block divided by c is

$$\ll X^{1-5a+O(\nu+\varepsilon)}\frac{(1 + D^2/N)\min(1, D/L)}{DL} \ll X^{O(\nu+\varepsilon)}\left(X^{-a} + \frac{D}{X^{1/2+a}}\right).$$

This saves a power on summing $D \le 2B$, since the exponent of B is strictly below $1/2 + a$, with gap ρ at $a = 1/6$. We may restrict the sum to the primes in the designated class. Markov's inequality gives a power-saving exceptional count. The cutoff χ_X is zero on $w < W_0$ for $\omega + \nu < a$.

The interior kernel. For $w \ge W_0$, put $x = (t-s)\sqrt{U}/w$ and $r = \sqrt{p}/(w\sqrt{U})$. On $|x| \le \log X$ the distance to the edge changes relatively by $O(r\log X)$ and $\theta^{(j)}(t) = O_j(p^{(j-1)/2}w^{1-2j})$. The scaled angle is $-x$ to first order, with degree-j coefficient $O_j(r^{j-1})$. Taylor-expanding the

Gaussian in these orders gives polynomial corrections of size $O_A(r^i)$ for $1 \le i < A$ and remainder $O_A(r^A(1+|x|)^{C_A}e^{-x^2/4})$. The reflection and off-window tails are negligible. Since $r \ll X^{-\nu}$ and $w/\sqrt{U}$ grows by a power, the summed remainder is $O_A(wX^{-A\nu}(\log X)^C)$, using the logarithmic bound for g_B on the grid and its polynomial bound on the Fourier extension. The continuous integral cancels the zero modes. Poisson summation at the other frequencies turns the corrections into polynomials in $w\beta/\sqrt{U}$ times the Gaussian, with relative prefactors $O_A(X^{-\nu})$.

Averaging the denominator blocks. Put $H = W/X^a$ on $W \le w < 2W$. At fixed s, open the Kloosterman sum with n in the interval $W^2 \le 4n - s^2 < 4W^2$ of length $N \ll W^2$, or use the full interval $N = X$. For $D \le \sqrt{X}$ the S5.2 large-sieve argument bounds the normalized average by

$$\ll_g \frac{(N+D^2)\sqrt{X}}{X^{1+a}D}\min(1,(D/H)^2).$$

The Gaussian sum over numerators is $O_g((d/H)\min(1,(d/H)^2))$. This holds with a polynomial factor or parameter derivative, since $|\beta| \gg_g 1/d$ for $\beta \ne 0$. The squared periodic integral in s costs at most $O(\sqrt{X})$ times its mean. For $D > \sqrt{X}$, replace the period estimate by the preceding Ramanujan argument. At each s, summing over the square congruence classes gives $O_{g,\varepsilon}(d^{2+\varepsilon}/H)$ since $1 \le d/H \le d$. The same bound applies to the polynomial and derivative weights. Integration over $O(\sqrt{X})$ and the large sieve now give $O_{g,\varepsilon}(D^{1+\varepsilon}/X^{1/2+a})$ after normalization. The angular factor is $O(W/X)$ in both estimates. Taking a supremum over the Gaussian parameter leaves the integer interval fixed.

Take $N = X$ for $d > D_+$. The bound up to $\sqrt{X}$ is $O(X^{1/2-a}/D)$, and above $\sqrt{X}$ the new bound saves a power since $D \le 2B$. With $N \ll W^2$, the bounds on $D \le W$, $W < D \le \sqrt{X}$ and $D > \sqrt{X}$ are respectively $O(W/\sqrt{X})$, $O(X^{-a})$ and $O(B^{1+\varepsilon}/X^{1/2+a})$. This removes the bands $W \ll X^{1/2-\omega}$ and also bounds the polynomial corrections by their squared prefactors $O_A(X^{-2\nu})$ times powers of $\log X$ after normalization by c. These bounds apply to the full sum and to $d \le D_+$. Markov's inequality gives the asserted precision and exceptional count.

The first sine coefficient of $\mathcal{D}$ is a constant multiple of $\sqrt{p}\,e^{-1/(2U)}(Z_B - \gamma_B(0))$. Projecting onto $\sin\phi$ therefore gives the mass assertion once the projection of $\chi_X V_+$ is power-small. Set $v = p/X$ and $u = s/\sqrt{X}$. After division by $\sqrt{p}$, the projection of each frequency is an integral in u with phase $e(\sqrt{X}\beta u)$ and smooth amplitude proportional to $\chi(X^\omega\sqrt{4v-u^2})\sqrt{4v-u^2}/v$ times the Gaussian. It extends by zero and each differentiation in u costs only $O(X^{2\omega})$ times a weaker Gaussian. Since each nonzero $\sqrt{X}|\beta| \gg_g X^{a-\omega}$, repeated integration by parts gives arbitrary power savings, also after summing the frequencies. This proves the lemma. □

Lemma S5.4 (The endpoint for fixed trace weights)**.** *Fix a trace weight g, a unit class $p \bmod M$ as in Section 6 of the article, and $\lambda > 0$. Put $U = \lambda p^{1/3}$ and $c = U/p$. For some $\xi > 0$, outside $O_{g,M,\lambda}(X^{1-\xi})$ primes in the class in $(X, 2X]$,*

$$\sum_{m\ge1} e^{-(m+1)^2/U} m_{g,p}(m)^2 = \mathcal{C}^g_{A_p}(c) + O_{g,M,\lambda}(cX^{-\xi}).$$

On the same primes the mass differs from its complete constant Fourier coefficient by $O_{g,M,\lambda}(X^{-\xi})$.

Proof. The class-number remainder with omitted factors. Take $B = X^{2/3-\rho}$ for a small fixed $\rho > 0$. Add a bump supported on $(1,2)$ to the smooth cutoff ψ to impose $\widehat{\psi}(-1/2) = 0$. By linearity take $g = 1_{\mathcal{A}}F^{(J)}$ and put $P_J = \prod_{\ell\in J}\ell$. For $D_t = |d_t|r_t^2$ let $k_\ell = v_\ell(r_t)$, $\epsilon_\ell = \chi_{d_t}(\ell)$ and $R_0 = \prod_{\ell\in J}\ell^{2k_\ell}$. Dropping the factors at J from C_t leaves C_t^* with $C_t^*(s) = (r_t/\sqrt{R_0})^{1-2s}C_t^*(1-s)$

by divisor substitution. Thus $L_t^* = L(s, \chi_{d_t})C_t^*(s)$ obeys the odd functional equation at conductor $D_t' = D_t/R_0$,

$$(D_t'/\pi)^{(s+1)/2}\Gamma((s+1)/2)L_t^*(s) = (D_t'/\pi)^{(2-s)/2}\Gamma((2-s)/2)L_t^*(1-s).$$

The original omitted series is $L_t^*(s)\prod_{\ell\in J}(1-\epsilon_\ell\ell^{-s})$. Shifting its cutoff integral past the residue $F^{(J)}$ and applying the equation gives

$$F^{(J)}(t) - F_B^{(J)}(t) = (D_t'/\pi)^{-1/2}\sum_{(n,P_J)=1}\sum_v b_t(n)a_{\boldsymbol{\epsilon}}(v)\mathcal{J}(\pi nvB/D_t'),$$

where $\widehat{\mathcal{J}}(s) = -\widehat{\psi}(-s)\Gamma((1+s)/2)/\Gamma(1-s/2)$ and the coefficients in v are given by the absolutely convergent expansion

$$\sum_v a_{\boldsymbol{\epsilon}}(v)v^{-s} = \prod_{\ell\in J}\frac{1-\epsilon_\ell\ell^{s-1}}{1-\epsilon_\ell\ell^{-s}} \qquad (\Re s > 0).$$

Only primes in J occur in v, with exponents at least -1. The absolute Dirichlet series is bounded for each $\Re s = \sigma > 0$. Mellin shifts make $\mathcal{J}$ and its logarithmic derivatives rapidly decreasing at infinity and bounded at zero, so $n \ll_J X^\varepsilon D_t'/B < p$ suffices with arbitrary power savings. On $w < W_0 = X^{1/3+\nu}$, with ν small relative to ρ, only $D_t \ll W_0^2$ contribute appreciably. The quadratic root bound and Liouville convolution give $|b_t(n)| \ll_\varepsilon n^{1/2+\varepsilon}$. Summing the kernel gives remainder $O_{J,\varepsilon}(X^\varepsilon D_t'/B^{3/2})$, hence $O_{J,\varepsilon}(X^\varepsilon W_0^2/B^{3/2})$ there. Since the heat kernels sum to $O(W_0)$ and the angle length is $O(W_0/\sqrt{X})$, the squared edge norm is $O_\varepsilon(X^\varepsilon W_0^7/(B^3X^{3/2})) = O(cX^{-\xi})$.

In the interior, fix a pattern $(k_\ell, \epsilon_\ell)_{\ell\in J}$. For a given p the pattern and $\mathcal{A}$ select $O_{M,J}(1)$ trace classes modulo $R \asymp_{M,J} R_0$. At odd ℓ the test uses $t^2 - 4p$ modulo $\ell^{2k_\ell+2}$ and, for $k_\ell \geq 1$, the at most two roots of $t^2 \equiv 4p \bmod \ell^{2k_\ell}$. At 2 the fundamental discriminant has valuation 0, 2 or 3, and the valuation and odd unit part modulo 8 are determined modulo 2^{2k_2+6}. For $k_2 \geq 2$, there are at most four roots of $(t/2)^2 \equiv p \bmod 2^{2k_2-2}$. Taking the least common multiple with the fixed selector period retains $O_{M,J}(1)$ classes.

Before this selector is applied, expand $b_t(n)$ by Lemma 3.2 of the article with D_t' fixed. The coefficient at m/d, for d prime to J, is $S(m, mp; d)$ times

$$\frac{(D_t'/\pi)^{-1/2}}{\varphi(d)}\sum_{(n,P_J)=1}\sum_v h_d(n)a_{\boldsymbol{\epsilon}}(v)\mathcal{J}(\pi nvB/D_t').$$

Extend the truncated h_d sum back to infinity with negligible error. Summing its square and divisor expansion first in the square variable, and using $\sum_v |a_{\boldsymbol{\epsilon}}(v)|v^{-1/2} \ll_J 1$, bounds the factor and its derivatives in $\log D_t'$ by $O_{A,J}(B^{-1/2}d^{-3/2}(1+dB/D_t')^{-A})$. The growing part of R is supported on J. A zero frequency after shifting by j/R therefore requires $d \mid M$. For each such fixed d, shift the coefficient's Mellin integral to $\Re s = \varepsilon$. The series for h_d has the factor $\zeta(2s)/\zeta(1+s)$. Omitting J and inserting the factors in v leave the other factors regular on $\Re s > 0$. The pole at $1/2$ is canceled by $\widehat{\mathcal{J}}(1/2) = 0$. The resulting bound is $O_{M,J,\varepsilon}((D_t')^{-1/2}X^\varepsilon)$. Each selector coefficient is $O_{M,J}(1/R)$, so the zero frequency together with its Taylor corrections costs $O_{M,J,\varepsilon}(X^{-1+\varepsilon})$ in squared norm.

On a band $w \asymp W \geq W_0$, Taylor expand the smooth conductor factor at $t = s = 2\sqrt{p}\cos\phi$ along with the heat kernel. Here $D_t/w^2 = 1 + O(r\log X)$ on the Gaussian window and $r = \sqrt{p}/(w\sqrt{U}) \ll X^{-\nu}$. The logarithmic derivative bounds give the polynomial Gaussian terms and negligible remainder of S5.3. For the other modes $d \ll X^\varepsilon W^2/(R_0B) \ll \sqrt{X}$, first apply weighted Cauchy–Schwarz over the shifts. The square is bounded by $O_{M,J}(1/R)$ times the sum of the squares at each shift, so the p-dependent selector coefficients have dropped out. At each fixed shift and

smooth parameters $w/W, p/X$, open the Kloosterman sums and use the additive large sieve with cost $O(X)$ in the integer replacing p. The s integral of the squared numerator sum costs $O(\sqrt{X})$ times its period mean, since the shift is common to its terms and $d \le \sqrt{X}$. For $H = W/\sqrt{U}$, the averaged Gaussian count is

$$\frac{1}{R} \sum_{j \bmod R} \sum_{m/d+j/R \neq 0} e^{-CH^2(m/d+j/R)^2} \ll d/H.$$

Indeed the shifted grid is $\operatorname{lcm}(d,R)^{-1}\mathbf{Z}$ with multiplicity (d,R), and the zero term is excluded. Polynomial Gaussian corrections obey the same bound with a smaller $C > 0$. Taking suprema in the two smooth parameters uses the same bounds for their first and mixed derivatives. The angular factor W/X, the prime normalization, and summation in d give squared norm $O_\varepsilon(X^\varepsilon B^{-1} W/(H\sqrt{X})) = O_\lambda(cX^{-\xi})$ on average. This also bounds the Taylor corrections. Sum over the $O_J((\log X)^{|J|})$ patterns and the fixed components of g. Markov's inequality restores the full weight with a power saving off a power-saving exceptional set.

This also controls the larger denominators with their original phases. Take $B_1 = X^{2/3+\rho}$. The positive kernel $G(\phi - \theta_t) + G(\phi + \theta_t)$ has bounded integral and sum $O(\sqrt{X})$ over t, by angle spacing. Cauchy–Schwarz and S5.1 give squared angular error $O(cX^{-\xi})$ at B_1, and thus the same bound for the difference of cutoffs. In its part with $d \le 2B$, the coefficient difference is $O(d^{-2}\min(1, \sqrt{d/B}))$ before the Kloosterman sum. The edge estimate in the proof of S5.3 bounds these blocks, as does its interior estimate for $d > D_+$. For $d \le D_+$ the latter costs $X^{1/3}/d$ after normalization, with an extra factor d/B from the squared coefficient difference. The constant mode leaves only its nonzero integer translates after first-sine projection, as in S5.3. Subtracting this low-denominator part bounds the full block $d > 2B$ of g_{B_1} by $O(cX^{-\xi})$ in squared angular norm off a power-saving exceptional set. This comparison applies to each fixed shifted component, using its original denominator before the shift.

Products at nearby trace frequencies. Use the reduction S5.3 with $D_+ = X^{1/3+\omega}$. Let $A(n,v)$ be the signed off-diagonal with p replaced by n in the phases and p/X by v in the weights. For each fixed choice of four shifts, the restrictions in Lemma 6.2 give $\beta_i = m_i/d_i + \kappa_i/M = A_i/(Md_i)$ with $A_i = Mm_i + \kappa_i d_i$, $d_i = X^{1/3+O(\omega+\varepsilon)}$, $0 < |A_i| \ll_M X^{O(\omega+\varepsilon)}$, and

$$0 < |h|, |h'| \le X^{1/6+O(\omega+\varepsilon)}, \qquad h = A_1 d_2 - A_2 d_1, \quad h' = A_3 d_4 - A_4 d_3.$$

Use denominators prime to the omitted sets. Their scalar coefficients are $d^{-2}\prod_{\ell \mid d}(1-\ell^{-2})^{-1}$ up to relative $O(\sqrt{d/B})$ by Lemma 3.2, since an omitted h_d factor at $s = 1$ equals one. The unsigned bound of Lemma 6.2, $O(X^{-4/3+O(\omega+\varepsilon)})$, absorbs this error, including one v derivative.

At coincident frequencies in n, Parseval on $[d_1, d_2]$ and Weil's bound give squared coefficient mass $O(X^\varepsilon (d_1 d_2)^{-3})$ per pair. At $(d_1, d_2) \asymp G$ there are $O(X^{1/2+O(\omega+\varepsilon)}/G)$ pairs and at most $O(GX^{O(\omega+\varepsilon)})$ contribute at any one frequency, by the count in Lemma 6.2. Cauchy–Schwarz over these pairs gives $O(X^{-3/2+O(\omega+\varepsilon)})$.

For distinct frequencies, open the four Kloosterman sums, with signs $+,-,-,+$, then sum against a smooth majorant of the prime interval by Poisson summation over integers. Let $P = \prod_i d_i$ and $Q = [d_1, d_2, d_3, d_4]$. The nonzero Poisson numerator j has weight $\widehat{\varphi}(Xj/Q)$. Each numerator has at most P/Q preimages. In particular, the absolute bound per quadruple is $O(1/(XP))$, including the four Fourier coefficients, since $\sum_{j \neq 0} |\widehat{\varphi}(Xj/Q)| \ll Q/X$.

Choose small $0 < \omega, \varepsilon \ll \eta \ll \vartheta \ll 1$, with ω, ε small also in terms of ρ, ν. Write $d_i = b_i u_i$, taking b_i to be the full part on primes dividing M or shared by at least two denominators. Discard tuples with any gcd exceeding X^η or any $b_i > X^{5\eta}$. An internal gcd divides the determinant.

For a gcd across pairs, choose the two denominators involved and sum $O(D_+^2/f^2)$ over common divisors $f > X^\eta$. In both counts the partner of a denominator is fixed by its shifted numerators and determinant. The first powers of primes dividing M cost a fixed factor. With all gcds small, $b_i > X^{5\eta}$ forces a squarefull part greater than X^η and hence a square divisor b^2 with $b > X^{\eta/3}$. This costs $O(D_+X^{-\eta/3})$ choices for that denominator. The absolute bound for these quadruples thus saves a power. We can impose just $b_i \le X^{5\eta}$ in the completed sums: any reintroduced tuples with a large gcd are bounded by the same count.

In the fixed b_i tuples every prime outside M occurs at least twice. The u_i are pairwise coprime and prime to MC, where $C = \prod b_i$. Put $L = [b_1, b_2, b_3, b_4]$, $V = \prod u_i$, so $Q = LV$. The weight restricts $|j| \le X^{1/3+O(\omega+\varepsilon)}$ with negligible error. Discard also $|j| < X^{1/3-\vartheta}$, and j having squarefull part or part supported on the primes of $2MChh'u_3u_4\prod_i A_i$ greater than X^ϑ in absolute value. For fixed quadruple the square-divisor count costs a factor $X^{-\vartheta/3}$ in the count of j. For a fixed integer $A = X^{O(1)}$, $\sum_{v|A^\infty,\ v>Z} v^{-1} \ll_\varepsilon Z^{-1/2}X^\varepsilon$, by the Euler product at $1/2$. Together with $Q \ge X^{4/3-O(\eta+\omega+\varepsilon)}$, these bounds give a power saving by the same absolute estimate. Write $|j| = j_0j_1$ on the retained terms, where j_0 is squarefree and prime to that fixed integer, $j_1 \le X^{2\vartheta}$ and $j_0 \ge X^{1/3-3\vartheta}$.

Write $P_1 = d_1d_2, P_2 = d_3d_4$. Fix the opened numerators x_i modulo b_i. Their phases there and the frequency condition modulo L are fixed by $u_i \bmod ML$, since $m_i = (A_i - \kappa_i d_i)/M$. On u_i, the congruence $\pm x_iQ/d_i \equiv j \pmod{u_i}$ fixes x_i, with $(j, V) = 1$, and $m_i \equiv A_i\overline{M} \bmod u_i$. Thus the inverse phase on V is

$$e_V\big(L(h^2P_2^2 + h'^2P_1^2)\overline{jC^2M^2}\big), \qquad e_q(x) = e(x/q).$$

Indeed the lifted ith inverse contribution is $LA_i^2(P/d_i)^2\overline{jC^2M^2}$ modulo V. Completing each determinant square adds a multiple of V. Reciprocity transfers the inverse to modulus $|j|C^2M^2$ and leaves the smooth phase $e((L/(jCM^2))(h^2P_2/P_1 + h'^2P_1/P_2))$.

Fix A_i, h, h', d_3 (hence d_4), the b_i, and j. Then $d_2 = (h + A_2d_1)/A_1$. Fix u_1, u_2 modulo $j_1C^2M^2$ for the small part of the reciprocal phase and the preceding conditions, including $(u_i, MC) = 1$. Expand the Euler products into divisor sums with squarefree coefficients $\prod_{\ell|r}(\ell^2-1)^{-1}$. Truncating at $r \le X^\eta$ costs $O(X^{-\eta+\varepsilon})$ in the unsigned bound. The divisor conditions and integrality restrict d_1 to progressions. The condition $(m_i, b_i) = 1$ is fixed. Impose $(A_i, u_i) = 1$, $(u_1, u_2) = 1$, and $(u_i, u_3u_4) = 1$ for $i = 1, 2$ by Möbius inversion over divisors of A_i, h and u_3u_4, respectively. All steps are prime to j_0: a divisor condition involving a prime of j_0 would force a nonunit u_i and contributes zero. The remaining unit conditions at j_0 are included in the complete sums.

Modulo each prime of j_0 the reciprocal phase is a unit times $h^2P_2/P_1 + h'^2P_1/P_2$. As a function of d_1 it has simple poles at the two distinct roots of $P_1 = d_1(h + A_2d_1)/A_1$. The rational-function Weil bound, also after an additive linear twist, and completion give $O_\varepsilon(X^\varepsilon(1 + D_+/j_0)\sqrt{j_0})$ per progression by [6, Theorem 1.1] and the Chinese remainder theorem. Partial summation costs $X^{O(\vartheta+\omega+\varepsilon)}$: the Gaussian and angular weights depend on the ratios A_i/d_i and $\sqrt{X}h/P_1$, the Poisson weight has bounded variation, and the smooth reciprocal phase has variation $O(X^{O(\omega+\varepsilon)}X^{1/3}/|j|)$. The frequency inequalities cut out a bounded number of intervals along the progression.

The fixed choices of h, h', d_3, j cost $X^{1+O(\omega+\varepsilon)}$. The remaining small moduli and residues cost $X^{O(\eta+\vartheta+\omega+\varepsilon)}$. Including the factor $X^{-8/3+O(\omega+\varepsilon)}$ from the four coefficients, the completed sum is $O(X^{-3/2+O(\eta+\vartheta+\omega+\varepsilon)})$. Since $c^2 \asymp X^{-4/3}$, the exponent margin here is $1/6$ before the small losses. With the indicated parameter order, the completed and discarded contributions are $O(c^2X^{-\xi})$. The same holds for one v derivative after decreasing ξ. The one-variable Sobolev bound in Lemma 6.2 thus allows $v = p/X$ and gives a power saving for the signed off-diagonal outside $O(X^{1-\xi})$ primes.

The diagonal and the full angular norm. Group equal shifted frequencies before squaring.

Lemma 6.2 bounds each component's diagonal, and Cauchy–Schwarz controls their overlaps. The angle and denominator truncations cost $O(cX^{-\xi})$, while the relative error $O(\sqrt{D_+/B})$ below D_+ saves a power against $O(c\log^3 X)$. S5.3 and the remainder comparison restore the full norm after first-sine projection. The cross term with $d > 2B$ also costs $O(cX^{-\xi})$ after decreasing ξ. Now (S2) gives the result. S5.1 and S5.3 bound $Z_g - Z_B$ and $Z_B - \gamma_B(0)$. The remaining constant coefficient tail is $O_{g,M}(B^{-1/2})$ since its denominators are fixed. □

S6 Local trace weights at arbitrary fixed level

The factors of Theorem 6.12 come from a single prime ℓ at a time. For trace R and unit determinant A, let ρ_i be the root count of $x^2 - Rx + A$ modulo ℓ^i. Set $\rho_0 = 1$ and $L = (1+\ell^{-1})^{-1}\sum_{i\ge 0}\ell^{-i}\rho_i$.

S6.1 Full and new trace weights

At level ℓ^e the full local trace weight is

$$G_0 = L, \qquad G_e = \sum_{j\ge 0}\ell^{-j}\rho_{e+j} = \ell^e\left((1+\ell^{-1})L - \sum_{i<e}\ell^{-i}\rho_i\right) \quad (e \ge 1). \tag{S3}$$

G_e is nonnegative and has form $aL + b$ with b of period ℓ^{e-1} for $e \ge 1$. Its mean is one since each unit x mod ℓ^i $(i \ge 1)$ specifies one trace $R = x + A/x$, so $\mathbb{E}\rho_i = 1 - \ell^{-1}$.

For the trace formula take $A = p$ and $R = t$. In [20, Theorem 10, (43)–(44)], write $t^2 - 4p = df^2$ with d fundamental and $\nu = v_\ell(f)$. In an integral basis $1, \omega_\ell$ we have $\pi = (t + \sqrt{t^2-4p})/2 = b + f\omega_\ell$. The local class-number weight for the orders indexed by $m \mid f$ is ℓ^{-j} when $j = v_\ell(m) < \nu$ and $\ell^{-\nu}(1 - \chi_d(\ell)/\ell)^{-1}$ when $j = \nu$.

For $j < e$, $j \le \nu$, the multiplier is ℓ^j times the number of classes modulo ℓ^e lifting to roots modulo ℓ^{e+j}. The polynomial at $b + y$ is $y^2 - f\,\mathrm{Tr}(\omega_\ell)y + f^2\,\mathrm{N}(\omega_\ell)$, so valuations force $\ell^j \mid y$. After division by ℓ^{2j}, the congruence is modulo ℓ^{e-j}. Each class has ℓ^j lifts, giving multiplier ρ_{e+j}.

If $\nu < e$, the last order accounts for $\sum_{j\ge\nu}\ell^{-j}\rho_{e+j}$. For $j \ge \nu$, divide the root count by ℓ^ν and use the polynomial of the maximal-order generator $(\pi - b)/\ell^\nu$ modulo $\ell^{e+j-2\nu}$. The counts are two at each modulus in the split case, zero in the inert case, and one at ℓ followed by zero in the ramified case. In the last case a lift x modulo ℓ^2 would make the generator minus x divisible by ℓ in the maximal order, by the integral trace and norm criterion. This contradicts generation, also at 2. The resulting geometric sum gives the last weight with factor $(1 - \chi_d(\ell)/\ell)^{-1}$.

If $\nu \ge e$, only $b \bmod \ell^e$ lifts modulo ℓ^{2e}, so all orders with $j \ge e$ have multiplier $\ell^{e-1}(\ell+1)$. The same valuation comparison gives $\rho_i = \ell^{\lfloor i/2\rfloor}$ for $i < 2e$. Hence these orders contribute $\ell^{e-1}(\ell+1)(L - \sum_{j<e}\ell^{-j}) = \sum_{j\ge e}\ell^{-j}\rho_{e+j}$. This proves (S3).

The Chinese remainder theorem multiplies these factors at level N. For $p \nmid N$ and even $k \ge 4$, the normalized trace is $-\sqrt{p}$ times the degree-$(k-2)$ moment plus the hyperbolic term $O_N(p^{-(k-1)/2})$. The oldform multiplicity is the divisor count of N/D at level $D \mid N$. Inverting this count by second differences gives

$$W_e = G_e - 2G_{e-1} + G_{e-2}, \qquad G_j = 0 \text{ for } j < 0. \tag{S4}$$

The same hyperbolic bound holds for the new trace. All the weights are even in t.

S6.2 Fourier energies

At denominator ℓ^j with $j \ge 1$, (S3) multiplies the Fourier coefficient of L by

$$v_0(j) = 1, \qquad v_e(j) = (\ell+1)\ell^{\min(j,e)-1} \quad (e \ge 1).$$

For $i < j$ the root count ρ_i has zero coefficient there. For $i \geq j$ the coefficient at numerator c is

$$\ell^{-j} \sum_{u \bmod \ell^j}^{*} \exp\bigl(-2\pi i c(u + A/u)/\ell^j\bigr),$$

independently of i. Summing the geometric series gives the multiplier. At the constant coefficient $v_e(0) = 1$ for $e \geq 0$.

Let c_j be the contribution of denominator ℓ^j to $\mathcal{B}_{\ell,A}(L, L)$ in Section 6, with $c_0 = 1 - \ell^{-1}$, and put $C_j = \sum_{i \geq j} c_i$ for $j \geq 1$. Define $s_h = \#\{u \bmod \ell^h : u^2 \equiv A \bmod \ell^h\}$ and $s_0 = 1$. Then

$$C_1 = \ell^{-1}, \qquad C_j = \frac{\ell^{-j} s_{\lfloor j/2 \rfloor}}{1 - \ell^{-2}} \quad (j \geq 2). \tag{S5}$$

Indeed, let D_j be the Haar measure in $\mathbf{Z}_\ell^2$ of unit pairs (u, v) with $(u - v)(uv - A) \equiv 0 \bmod \ell^j$. Primitive additive orthogonality gives $c_j = (D_j - \ell^{-1} D_{j-1})/(1 - \ell^{-2})$. For fixed unit u put $d = v_\ell(u^2 - A)$. If $j \leq 2d$, the eligible v lie in one class modulo $\ell^{\lceil j/2 \rceil}$. If $j > 2d$, there are two disjoint classes modulo ℓ^{j-d}, centered at u and A/u. With $P_h = \ell^{-h} s_h$ for $h \geq 1$, subtracting the adjacent measures gives, for $h \geq 1$,

$$(1 - \ell^{-2}) c_{2h} = (1 - \ell^{-1}) \ell^{-h} P_h, \qquad (1 - \ell^{-2}) c_{2h+1} = \ell^{-h-1}(P_h - P_{h+1}).$$

These identities telescope to give C_j for $j \geq 2$. The value $C_1 = 1/\ell$ follows from $\mathcal{B}_{\ell,A}(L, L) = 1$. In particular $c_1 = \ell^{-1} - s_1/(\ell^2 - 1)$.

S6.3 Pairings and orthogonality

Write $F_e = \mathcal{B}_{\ell,A}(G_e, G_e)$ and $B_{mn} = \mathcal{B}_{\ell,A}(W_m, W_n)$. Summation by parts of $c_0 + \sum_{j \geq 1} c_j v_e(j)^2$ gives

$$F_0 = 1, \qquad F_e = \ell + 3 + (\ell + 1)^2 \sum_{j=2}^{e} \ell^{j-2} s_{\lfloor j/2 \rfloor} \quad (e \geq 1). \tag{S6}$$

For the new pairings set $b_n = (\ell^2 - 1)\ell^{n-3}$ for $n \geq 3$, and

$$d_0 = 1, \qquad d_1 = \ell - 1, \qquad d_2 = \ell^2 - \ell - 1, \qquad d_n = (\ell - 1) b_n \quad (n \geq 3).$$

The diagonal is

$$\begin{aligned} &B_{00} = 1, \qquad B_{11} = \ell - 1, \qquad B_{22} = \ell + s_1(\ell^2 - 2\ell - 1), \\ &B_{nn} = b_n\bigl(s_{\lfloor (n-1)/2 \rfloor} + (\ell - 2) s_{\lfloor n/2 \rfloor}\bigr) \qquad\qquad (n \geq 3). \end{aligned} \tag{S7}$$

The mixed entries for $m < n$ are

$$B_{mn} = \begin{cases} d_m(s_h - s_{h-1}), & n = 2h,\ h \geq 1, \\ 0, & n \text{ odd}. \end{cases} \tag{S8}$$

For W_m the multiplier at all positive denominator exponents $j \geq m$ is d_m. If $n \geq 3$, the multiplier of W_n vanishes below $n - 1$, is $-b_n$ at $n - 1$, and is $(\ell - 1)b_n$ above. Thus for $m < n$ the pairing is $d_m b_n(\ell C_n - C_{n-1})$. The squared norm of W_n is $b_n^2(C_{n-1} + (\ell^2 - 2\ell)C_n)$. Now use (S5) in both expressions. The constant coefficients of W_0, W_1, W_2 are $1, -1, 0$. At positive exponents W_0, W_1

have constant multipliers $1, \ell - 1$. The multiplier of W_2 is $-\ell$ at exponent one and d_2 thereafter. Inserting c_0, c_1, C_1, C_2 gives the remaining entries.

By character orthogonality, $s_h - s_{h-1}$ sums the quadratic characters of exact conductor ℓ^h at A. For odd ℓ, all s_h with $h \geq 1$ equal $1 + (\frac{A}{\ell})$. The unequal pairs $\{0,2\}, \{1,2\}$ are the only nonzero ones, in both square classes. The diagonal vanishes exactly for $n \geq 3$ in the nonsquare class. At 2 the counts are

$$s_1 = 1, \qquad s_2 = 2\,1_{A\equiv 1\,(4)}, \qquad s_h = 4\,1_{A\equiv 1\,(8)} \quad (h \geq 3).$$

Thus a mixed entry between unequal exponents vanishes except at $n = 4$ or 6 for $n > m$. At $n = 4$ the factor $s_2 - s_1$ is 1 or -1 according to A mod 4. At $n = 6$ the factor $s_3 - s_2$ is 2 at $A \equiv 1$ mod 8, -2 at 5, and zero at $3, 7$. An equal-exponent entry at 2 is positive through $n = 4$, is positive at $n = 5, 6$ exactly for $A \equiv 1$ mod 4, and at $n \geq 7$ exactly for $A \equiv 1$ mod 8. These are all the possible zeros of a factor in the product.

For a fixed list of levels, the constants are independent of A at odd primes occurring only to the first power. They depend only on the Legendre symbol at the other odd primes. At 2 they are independent of A when the largest exponent is at most 3, are determined by A mod 4 when it is 4 or 5, and by A mod 8 for larger exponents. The relative-density-one sets in finer classes can be combined into sets in these coarser classes, since the number of finer classes is finite.

S7 Dependencies of the main proof

Main trace theorem. Section 6 of the article gives the proof order. The count identity uses Diaconis–Khare–Saloff-Coste, Propositions 4.1–4.2. Lemmas 3.2–3.4 use the Deuring, Gekeler and Zagier formulas, Burgess with parameter 3, and Petrow–Young, Theorem 1.1. S5.1–S5.3 use Heath–Brown's fourth moment [14, Theorem 2] and the additive large sieve [32, Theorem 1 and Lemma 1]. S5.4 adds the discriminant functional equation and the rational-function Weil bound. Theorem 6.1 uses Siegel–Walfisz and the multiplicative large sieve.

Applications. Proposition 6.3 uses the short-interval bound of David–Koukoulopoulos–Smith [9, Theorem 1.3]. Lemma 6.8 adapts their progression argument for Proposition 6.9 [9, Theorems 3.2 and 4.2, Corollary 3.3, §5.2]. Theorems 6.5, 6.11 and 6.12 use local factors and oldform inversion, with Kaplan–Petrow [20, Theorem 10, (43)–(44)]. Their Lemmas 2–4 and Theorems 2–3 give the weights in Section 4.

Variance and prime means. Theorem S8.1 and Corollary S9.1 use the same fourth moment and sieves, plus the zero-density bound [23, (1.2)]. Corollary S9.2 uses Alpöge's Dirichlet critical-line theorem [2, 39]. Proposition 7.1 and the strengthened S1 bounds use its quadratic case.

S8 Variance across primes at level one

Choose the local unit residues independently and uniformly in the Haar model. We compute its variance by character completion and transfer it to the primes using bounds grouped by conductor divisibility. Put $x_X = \sqrt{\lambda}\, X^{-\delta/2}$.

For B uniform on the units modulo ℓ^k set $Y_{\ell,k}(B) = (1 - \ell^{-2})^{-2} |S(1, B; \ell^k)|^2 / \ell^k$. Define the common arithmetic factor and the kernel integral by

$$\mathfrak{A} = \prod_{\ell} (1 - \ell^{-1})^2 \left(\frac{1}{1 - \ell^{-5}} + \sum_{k \geq 1} \ell^{-k} \mathbb{E}_B[Y_{\ell,k}(B)^2] \right), \qquad I(H) = \int_0^\infty v^4 H(v)^2 \, dv. \tag{S9}$$

Let $W(v) = 2 \int_0^1 \sigma(y) e^{-64\pi^2 y(1-y)v^2} \, dy$ and put $J_\delta = \int_1^2 v^{-\delta/2} \, dv$.

Theorem S8.1. *Fix $\lambda > 0$ and $0 < \delta \le 1/4$. Put $T_p = \lfloor \lambda p^{1-\delta} \rfloor$, $c_p = \lambda p^{-\delta}$ and $x_p = \sqrt{c_p}$, and write $\mathbb{E}_X$ for the uniform average over primes $X < p \le 2X$. For $Y_p = p\chi^2(P_{T_p} \| Q_{T_p})/T_p$ and $F_x(A) = x^{-2}\mathcal{C}_A(x^2)$ one has*

$$\operatorname{Var}_X(Y_p) = \mathbb{E}_X \operatorname{Var}_A(F_{x_p}) + o(x_X \log(2/x_X)) \sim \frac{\delta\sqrt{\lambda}}{4}\, J_\delta\, \mathfrak{A}\, I(W) X^{-\delta/2} \log X.$$

The same holds for $T_p^{-1} \sum_{k \ge 4,\ 2|k} e^{-(k-1)^2/T_p} |\tau_k(p)|^2$.

S8.1 Fourier coefficients in the determinant

The profile of Theorem 5.1 gives

$$F_x(A) = x^{-2} \sum_{a \ge 1} \sum_{(d,a)=1} d^{-3} W(a/(xd)) Y_{a,d}(A), \qquad Y_{a,d} = d^3 |\gamma_A(a/d)|^2.$$

The sum converges uniformly in A by Weil's bound and $\sum_a W(a/(xd)) \ll xd$. The dyadic supremum and variation of $(z\partial_z)^j W$ are $O_j(\min(1, z^{-3}))$: each derivative is bounded by the integral of a weaker Gaussian against $\sigma(y)$. At ℓ^k write the local factor as $h_\ell U_k(B)$, where $h_\ell = (1-\ell^{-2})^{-2}$ and $U_k(B) = |S(1, B; \ell^k)|^2/\ell^k$. Its principal coefficient $g(\ell^k)$ is the one in Theorem 5.2. Thus g is bounded above and below by positive constants and $g = 1 * h$ with $\sum_m |h(m)| < \infty$.

For a nontrivial character χ of conductor ℓ^j, $j \le k$, put $A_{\ell,k}(\chi) = \mathbb{E}_B[h_\ell U_k(B)\overline{\chi(B)}]$. On opening the square, the B-sum requires $x_1 - x_2 = \ell^{k-j} v$ with v a unit modulo ℓ^j and contributes $\ell^{k-j}\chi(-v/(x_1x_2))\tau(\bar\chi)$. Set $x_1 = vu,\ x_2 = v(u - \ell^{k-j})$. There are ℓ^{k-j} lifts of u, and the v-sum gives a second Gauss sum:

$$\frac{A_{\ell,k}(\chi)}{h_\ell} = \frac{\chi(-1)\tau(\bar\chi)^2}{\ell^j \varphi(\ell^j)} \sum_{\substack{u \bmod \ell^j \\ u,\ u-r \text{ units}}} \bar\chi(u(u-r)), \qquad r = \ell^{k-j}.$$

For $j = k$ the sum is $\bar\chi(-1) J(\bar\chi, \bar\chi)$. At odd ℓ, a nonquadratic primitive character has primitive square, so the Jacobi identity gives $|A_{\ell,j}(\chi)| = h_\ell \ell^{j/2}/\varphi(\ell^j)$. For $k > j \ge 2$ the lifts modulo ℓ^{j-1} cancel since $2u - r$ is a unit. For $j = 1 < k$ the sum $\sum_u \bar\chi^2(u)$ vanishes unless χ is quadratic. The quadratic coefficient is $-h_\ell/(\ell-1)$ at $k = 1$ and h_ℓ at $k \ge 2$.

At 2 a nonquadratic primitive character has $j \ge 4$. There are no units $u, u - r$ for $r = 1$. For $4 \mid r$, translating by 2^{j-2} multiplies the summand by $\bar\chi(1 + 2^{j-1}) = -1$. For $r = 2$, put $b = \lfloor j/2 \rfloor$ and split into classes modulo 2^{j-b}. The quadratic term in the lift parameter vanishes modulo 2^j; primitivity leaves at most four classes, determined by $2u - 2 \equiv 0 \bmod 2^b$. Their 2^b lifts give $A_{2,k}(\chi) = O(2^{-j/2})$ at $k = j + 1$. For real dyadic characters use $O((k+1)^2)$ and conductors dividing 8. In fact $\sum_{k \ge 1} A_{\ell,k}(\chi)\ell^{-k} = 0$ for each real nontrivial local character. The unit-numerator average in Theorem 5.2 is constant in the determinant and unit-square twists fix real characters.

S8.2 Bounds for the Fourier energy

Write $F_{x,\le D}$ for the denominator cutoff in the profile and $G_{x,D} = F_{x,\le D} - \mathbb{E}_A F_{x,\le D}$. Set $V(x) = x\log(2/x)$, the scale of the variance. We take $x \asymp x_X$ and $D \le X^{1/4}$, holding cutoffs fixed under differentiation. For every integer $w \ge 1$, every fixed $j \ge 0$ and every $\varepsilon > 0$,

$$\sum_{w | f_\chi} \frac{f_\chi}{\varphi(f_\chi)} |\widehat{(x\partial_x)^j G_{x,D}}(\chi)|^2 \ll_{j,\varepsilon,\lambda,\delta} V(x) w^{-1+\varepsilon}. \tag{S10}$$

Here f_χ is the primitive conductor. For the centered difference of two cutoffs, if $D \geq B$ and $Z = xB \geq 2$, then

$$\sum_{\chi} \frac{f_\chi}{\varphi(f_\chi)} |(x\partial_x)^j (\widehat{G_{x,D}} - G_{x,B})(\chi)|^2 \ll_{j,\lambda,\delta} V(x)/Z. \tag{S11}$$

Split the conductor into q from non-real local characters and s from real nontrivial ones, with at most three choices at 2 and one at each odd prime. The local support gives $d = rm$, $(m, r) = 1$, where $r = \kappa q t$, with $\kappa = 1$ for odd q and 2 otherwise. The integer t has the prime support of s and is divisible by s. The Chinese remainder theorem gives the cofactor twist $\overline{\chi_q}^2(m)$, with unit phases for fixed a, r. The non-real coefficients cost $O(q^{-1/2} a_q)$ with $a_q = q/\varphi(q)$. The product α_t of the real local bounds satisfies

$$\sum_t \frac{\alpha_t}{\sqrt{t}} \ll_\epsilon s^{-1+\epsilon}.$$

The local sum is $O(\ell^{-1}(1 + O(\ell^{-1/2})))$ at each odd prime and bounded at 2.

For a weight of supremum and variation at most W_0 on $[N, 2N)$, the vector of sums against $\bar{\chi}_q^2$, weighted by the non-real coefficients, has norm

$$\ll \sqrt{a_q}\, W_0 \left[\sqrt{N} + 2^{\omega(q)/2} \min(N/\sqrt{q}, \sqrt{q})\right].$$

For short sums count $n^2 = n'^2 \bmod q$. Besides the $O(N)$ diagonal pairs, assign every odd prime power to $n - n'$ or $n + n'$ and, at 2, assign $2^{v_2(q)-1}$ to the factor with larger valuation. The assigned moduli have product at least $q/2$, giving $O(N^2/q)$ pairs per assignment. For long sums each unit class gives the same integral divided by q up to $O(W_0)$. The integral cancels since χ_q^2 is nonprincipal. Apply Parseval to the errors, with at most $2^{\omega(q)+1}$ preimages of a square. Multiplying the orthogonality factor $\sqrt{\varphi(q)}$ by $q^{-1/2} a_q$ gives the norm for $N \geq 1/2$.

For $m \asymp M$, the weight $m^{-3} W(a/(xrm))$ and its variation are $O(M^{-3} \min(1, (xrM/a)^3))$. Expand $g = 1 * h$ and impose coprimality to as by Möbius inversion. The character imposes coprimality to q. Writing $m = un$ gives $N \asymp M/u$ at total cost $\tau(as) \sum_m |h(m)|$. Apply the vector bound and use

$$S(z) = \sum_{a \geq 1} \tau(a) \min(1, (z/a)^3) \ll \begin{cases} z^3 & z \leq 1, \\ z \log(2z) & z \geq 1. \end{cases}$$

Write $v = xr$. The first row term, without $\sqrt{a_q} \alpha_t \tau(s)$, is bounded by

$$x^{-2} r^{-3} \sum_M M^{-5/2} S(vM) \ll \sqrt{x/r} \begin{cases} 1 & v \leq 1, \\ v^{-3/2} \log(2v) & v > 1. \end{cases}$$

The second term, also with $2^{\omega(q)/2}$ omitted, is

$$x^{-2} r^{-3} \sum_M M^{-3} S(vM) \min(M/\sqrt{q}, \sqrt{q}) \ll (r\sqrt{q})^{-1}.$$

All M-sums are dyadic from 1. For the last bound, use $M/\sqrt{q}$ and split at $1/v$. If $vq \leq 1$, splitting at both q and $1/v$ gives the alternative bound $\ll x\sqrt{q}(1 + \log(1/(vq)))$.

Assign the prime powers of w to $w_q \mid q$ or $w_s \mid s$. For $q \asymp U$ restricted by $w_q \mid q$, use

$$\sum a_q^2 \ll_\varepsilon U w_q^{-1+\varepsilon}, \qquad \sum 2^{\omega(q)} a_q^2 \ll_\varepsilon U w_q^{-1+\varepsilon} \log(2U).$$

Write $q = w_q v$, use submultiplicativity and expand a_v^2 over squarefree divisors with local coefficients $O(1/\ell)$. For the second bound use $2^{\omega(v)} \leq \tau(v)$ and sum over multiples of each divisor.

The first squared row contributes $O((x/t)\log(2D))$. Split the second at $q=(xt)^{-1/2}$ to obtain $O((x/t)\log(2/x))$. Both retain $w_q^{-1+\varepsilon}$. Taking the triangle inequality in t before squaring over s bounds the remaining summands by $\tau(s)^2(s/\varphi(s))s^{-2+\varepsilon'}$. Thus $w_s \mid s$ gives $w_s^{-2+\varepsilon}$, and smaller intermediate exponents absorb the $2^{\omega(w)}$ assignments.

For the band $B<rm\le D$, put $v=xr$ and $z_0=v\max(1,B/(2r))$. Nonempty dyadic m-bands have $vM\gg z_0\gg Z$. The numerator sum is $S(vM)\ll vM\log(2vM)$. In the first row the sum improves by $O(z_0^{-3/2}\log(2z_0))$. In the second, using each side of $\min(M/\sqrt{q},\sqrt{q})$ separately gives, before $2^{\omega(q)/2}$,

$$\ll \log(2z_0)\min\bigl((r\sqrt{q}\,z_0)^{-1},x\sqrt{q}/z_0^2\bigr)\ll Z^{-1/2}\min\bigl((r\sqrt{q})^{-1},x\sqrt{q}\bigr).$$

Sum the squares over q, then use the real-support sum as above. This gives (S11) for $q>1$.

For a nonprincipal real character, $g=1*h$ and Möbius inversion give a linear summatory term for $g(m)1_{(m,as)=1}$ with error $O(\tau(as))$. After $d=rm$ the integral cutoff is independent of r, so $\sum_k A_{\ell,k}(\chi_\ell)\ell^{-k}=0$ cancels the integrals, including those with $r>D$. For $rm\le D$, the better of the partial summation error $O(\tau(s)x\log(2+1/(xr)))$ and the bound $O(1/r)$ on the sum and integral gives $O(\tau(s)\sqrt{x/r})$. On a band with $r\le B$, partial summation costs $O(\tau(s)xZ^{-2}\log(2Z))$, summing the variation over $m\gg B/r$. For $r>B$, use $\sum_a|W(a/(xd))|\ll xd$ and integrate m^{-2} over $m>B/r$ to bound the sum and integral by $O((Zr)^{-1})$. Both cases give $O(\tau(s)\sqrt{x/r}\,Z^{-1/2})$, also for derivative weights. The same real-support sum proves (S10) and (S11).

S8.3 The leading Haar variance

Write $F_{H,x}$ for the profile with kernel H in place of W. For the Gaussian and sharp kernels we prove

$$\operatorname{Var}_A F_{H,x}\sim \tfrac12\mathfrak{A}I(H)x\log(1/x). \tag{S12}$$

First take H smooth with compact support in $(0,\infty)$ and put $L=1/x$. Haar Parseval expresses the variance as a sum over nonprincipal characters. The purely real characters contribute $O_H(x)$ by the cancellation of the full local sums above.

Completing the cofactor sums. For $q>1$, χ_q^2 is primitive of conductor $q'=q$ if q is odd and $q/2$ otherwise: squaring preserves the odd exponents and lowers the dyadic exponent $j\ge4$ by one. For a smooth weight at scale N with scaled derivatives $O(W_0)$, Poisson summation has dual length q'/N, prefactor of modulus $N/\sqrt{q'}$, and zero constant term. Apply the same square-congruence count to both sums. For $N\ll q$ this gives

$$\left(\sum_{\chi_q}\left|C_q(\chi_q)\sum_n w(n)\chi_q^2(n)\right|^2\right)^{1/2}\ll\sqrt{a_q}W_0\left(\sqrt{N}+2^{\omega(q)/2}\min(N/\sqrt{q},1)\right),$$

Here $C_q(\chi_q)$ is the product of the non-real coefficients. The dual characters $\overline{\chi_q}^2$ use the count modulo q after separating signs and summing rapidly decaying dyadic tails. Since $2^{\omega(q)/2}\ll q^{1/4}$, the norm is $O(\sqrt{a_q}W_0\sqrt{N})$, with an extra $(q/N)^A$ for $N\ge q$ by rapid decay.

With $d=\kappa qtm$ as above, let $\alpha_s(t)$ be the product of the actual real local coefficients. Up to a phase independent of a,t,m, the full Fourier coefficient is

$$C_q(\chi_q)x^{-2}\sum_{\substack{a,t,m\\(a,qtm)=1,\ (m,qs)=1}}\frac{\alpha_s(t)g(m)}{(\kappa qtm)^3}H\bigl(a/(x\kappa qtm)\bigr)\chi_q^2(a/(tm)).$$

Here $(q,s)=1$, and the unit-square phases are as in the cofactor bound above. For fixed a,t, its divisor expansion again costs $\tau(as)\sum|h|$, with row lengths at most $M \asymp La/(\kappa qt)$. The smooth row bound gives a norm

$$\ll_H \sqrt{a_q}\,|\alpha_s(t)|\tau(as)\sqrt{x/(\kappa qt)}\,a^{-5/2}.$$

Since $q \ll_H La/t$ and $\sum_{q\asymp Y} a_q/q \ll 1$, the real-support bound and $\sum_a \tau(a)a^{-5/2}\sqrt{\log(2a)} < \infty$ allow fixed truncations of a,s,t at arbitrarily small cost in units of $\sqrt{x\log L}$. The same estimate on multiples of 2^J bounds their contribution by $O_H(2^{-J/2}\sqrt{x\log L})$. The ranges $L^{1/2-\eta} < q < L^{1/2+\eta}$ and $L^{1-\eta} < q \ll L$ then have squared norm $O_H(\eta x\log L + x)$, with constants depending on the truncations. For $q \le L^{1/2-\eta}$, truncate the absolutely summable divisor coefficients in the same norm. The remaining rows have $N/q' \gg L^{2\eta}$, giving $o(x)$ in squared norm by completion.

The diagonal in the remaining range. Take $L^{1/2+\eta} \le q \le L^{1-\eta}$ and keep each dyadic local character fixed. Put $Q = q/2^{v_2(q)}$ and $n = tm$. Equal ratios $a/n = a'/n'$ give $a = a'$ and $n = n'$ by coprimality. Then $t = t'$ since its exponents are exactly those of n at the primes of s.

For distinct ratios, average the squared-character phase over the odd components with the normalization $1/\varphi(Q)$. At $\ell^j \| Q$ the factor at a unit ratio z is

$$\begin{cases} 1_{\{z^2=1 \bmod \ell\}} - 2/(\ell-1), & j=1,\\ 1_{\{z^2=1 \bmod \ell^j\}} - \ell^{-1}1_{\{z^2=1 \bmod \ell^{j-1}\}}, & j\ge 2. \end{cases}$$

The non-real coefficient magnitudes are independent of the odd characters. Expanding the product gives divisors $D \mid \mathrm{rad}(Q)$ with coefficients $O_\epsilon(Q^\epsilon/D)$ and congruence $Q/D \mid (an'-a'n)(an'+a'n)$. For fixed a,a', both forms are nonzero of size $O(M)$, $M = L/q$. Each prime power of Q/D divides exactly one of them because the entries are units there and the prime is odd. Each assignment counts $O(M^2D/Q)$ pairs by counting the nonzero multiples in the two forms, which determine n,n'. Including the assignments and divisors gives $O_\epsilon(M^2Q^{-1+\epsilon})$. The cofactor weights before $x^{-2}(\kappa q)^{-3}$ are $O(M^{-3})$. The sum of the squared coefficient magnitudes over all characters modulo Q, using the common magnitude of the admissible ones, is $O_\epsilon(Q^\epsilon)$. Consequently the off-diagonal sum is bounded by $O_\epsilon(\sum_{q\ge L^{1/2+\eta}} q^{-3+\epsilon}) = o(x)$ for sufficiently small ϵ.

The constant. Put $D_\ell = \sum_{k\ge0} g(\ell^k)^2/\ell^k$ and $\rho = \prod_\ell (1-\ell^{-1})D_\ell$. The mean of $g(m)^2 1_{(m,P)=1}$ is $m(P) = \rho\prod_{\ell\mid P} D_\ell^{-1}$ with summatory error $O(\tau(P))$. Expand g^2 as the convolution of 1 with an absolutely summable function and impose coprimality. For each fixed a,s,t in the diagonal,

$$\int_0^\infty m^{-6} H\big(a/(x\kappa qtm)\big)^2\,dm = (x\kappa qt/a)^5 I(H).$$

Partial summation gives an error of $O_\epsilon(x^2q^\epsilon)$ per q after normalization, hence $o(x)$ in this range. With $C(q) = \sum_{\chi_q} |C_q(\chi_q)|^2$, the main term is

$$\frac{xI(H)}{a^5 t}\frac{C(q)}{\kappa q}\,|\alpha_s(t)|^2\,m(aqs)\,1_{\{(a,qs)=(q,s)=1\}}.$$

For fixed a,s,t and fixed real character, the coefficient of $1/q$ has a mean in q. Its multiplicative part is $1+O(1/\ell)$ at large primes and bounded at their powers, so its convolution with μ is absolutely summable with weight $1/q$. Summation against $1/q$ gives this mean times $(1/2-2\eta)\log L + o(\log L)$.

Send η to zero and remove the fixed truncations. The cofactor mean and the mean in q give $(1-\ell^{-1})^2$ at each prime. At denominator exponent zero the numerator powers sum to $(1-\ell^{-5})^{-1}$.

At exponent $k \geq 1$ the numerator is a unit. The cofactor, real support and non-real conductor supply the principal, real nontrivial and non-real coefficients, respectively. Their squared moduli sum to $\mathbb{E}_B[Y_{\ell,k}(B)^2]$ by local Parseval, with weight ℓ^{-k} in each case. For a non-real conductor 2^j, $k = j + 1$ and $1/\kappa = 1/2$ adjusts the weight accordingly. This gives (S9). At odd primes the local formulas give $\mathbb{E}_B[Y_{\ell,1}^2] = 2 + O(\ell^{-1})$, and Weil's bound controls $k \geq 2$ by $O((k+1)^4)$. The factors are therefore $1 + O(\ell^{-2})$, and the product is positive.

For W and $W_\#$, the upper bound for the Fourier energy with $w = 1$ applies uniformly to each rescaled compact piece with bounded supremum and variation, by a polynomial denominator cutoff and the uniform tail bound. The identity $F_{H(\cdot/b),x} = b^2 F_{H,bx}$ gives bounds $O(b^2)$ for $b \leq 1$ and $O(b^{-1/2})$ for $b \geq 1$ on the dyadic pieces at $v \asymp b$, in units of $\sqrt{x \log(1/x)}$. For $bx \geq 1$, use the absolute bound $F_{H,bx} = O((bx)^{-1})$ for a rescaled compact piece. Cut off both tails and smooth the remaining kernel in $\log v$. For $W_\#$ the variation estimate in S9 makes the smoothing error arbitrarily small in the same norm. The integrals $I(H)$ converge in these approximations, proving (S12).

S8.4 The mean of the Haar profile

For transfer to primes, the uniform bound $O(x^{-1}D^{-1}\log^3(2D))$ above denominator D follows from Weil's bound and $\sum_a W(a/(xd)) \ll xd$. For $D = X^{1/4}$ it is $o(\sqrt{V(x)})$.

We also need a mean accurate to $o(\sqrt{V(x)})$ for centering the prime statistic. The Haar mean has the expansion

$$\mathbb{E}_A F_x = \frac{1}{48} - \frac{\sqrt{\pi}}{2304}x + O_\eta(x^{1+2\eta}) \qquad (0 < \eta < 1/4).$$

For $1 < \Re s < 2$ take the Mellin transform

$$\widehat{W}(s) = \frac{8}{\pi}\Gamma(s/2)(64\pi^2)^{-s/2}\mathrm{B}((3-s)/2, (3-s)/2).$$

By Mellin inversion and absolute convergence, the mean is the integral of $x^{s-2}\widehat{W}(s)\zeta(s)\zeta(3-s)\mathcal{G}(s)$ on this line. The coprimality of numerator and denominator gives the Euler factor

$$\mathcal{G}_\ell(s) = (1-A)(1-B)\left((1-A)^{-1} + g(\ell)B + g(\ell^2)B^2/(1-B)\right), \qquad A = \ell^{-s}, \quad B = \ell^{s-3}.$$

Expanding the factor gives

$$1 + (g(\ell) - 1)B - g(\ell)AB + (g(\ell^2) - g(\ell))(1-A)B^2.$$

Both differences of g are $O(\ell^{-2})$. The Euler product is absolutely convergent and bounded on closed vertical substrips of $1 < \Re s < 7/2$. The local mean identity at $s = 2$ $(1 - \ell^{-1})\sum_{k\geq 1} g(\ell^k)\ell^{-k} = \ell/(\ell^2 - 1)$ gives $\mathcal{G}_\ell(2) = 1$. At $s = 3$, $\mathcal{G}_\ell(3) = (1-\ell^{-3})(1-\ell^{-2})^{-2}$. Stirling's formula gives exponential vertical decay of the continued Mellin factor.

Shift to $3 + 2\eta$, subtracting the residues at 2 and 3. At 2 use $\widehat{W}(2) = 1/(8\pi^2)$ and the residue -1 of $\zeta(3-s)$. At 3 the principal part of $\widehat{W}(s)$ is $1/(32\pi^{7/2}(3-s))$. Together with $\zeta(0) = -1/2$ and $\mathcal{G}(3) = \zeta(2)^2/\zeta(3)$, this gives the negative term. The shifted integral is $O_\eta(x^{1+2\eta})$.

S8.5 A heat estimate for both cutoffs

Put $T = \lambda p^{1-\delta}$, $c_p = T/p$, $h = z/T$ and $R = |z|$ on $\Re z = 1$. We seek $o(X^{-\gamma})$, with $\gamma = \delta/4$ for the variance and $\gamma = \delta/2$ for the second mean term. For small $\eta_1 > 0$ set $\kappa = \gamma + \eta_1$, $d_* = \delta/2 + 2\gamma + \eta_1$, $\beta_* = \delta + 2\gamma + 2\eta_1$ and $\omega_* = \delta/2 + 2\kappa + \eta_1$. Use the class-number cutoff $B = X^{\beta_*}$, denominator split $D_0 = X^{d_*}$ and angular cutoff $W_0 = X^{\omega_*}$, with $R \leq 2X^\kappa$. Choose η_1 to preserve the strict margins

at $\eta_1 = 0$ and ε still smaller. At $\delta = 1/4$ take $\gamma = 1/16$, $u = \eta_1 = 1/1024$ and $\vartheta = 1/128$. On each band $w \asymp W$, use $D_0 = X^{1/4-\vartheta}$ if $W \le X^{1/2-\vartheta}$ and $D_0 = X^{1/4+\vartheta}$ otherwise. The remaining parameters stay as above.

With $w = 2\sqrt{p}\sin\phi$ and $s = 2\sqrt{p}\cos\phi$, use $\mathcal{D}_h$ from (S2) with complex h. The bilinear square retains $e^{-h(m+1)^2}$ and the first-sine term $e^{-h}(Z_B - \gamma_B(0))^2$. The absolute square uses $\Re h$. The two transforms to compare are

$$M_p(z) = T^{-1} \sum_{\substack{k\ge 4\\ k \text{ even}}} |\tau_k(p)|^2 e^{-z(k-1)^2/T}, \qquad M_p^{\mathrm{diag}}(z) = c_p^{-1} \int_0^1 \sigma(y) \sum_{\beta\ne 0} |\gamma_{A_p}(\beta)|^2 e^{-zb(y)\beta^2/c_p}\, dy, \tag{S13}$$

Here $b(y) = 64\pi^2 y(1-y)$, so $M_p^{\mathrm{diag}}(1) = F_{x_p}(A_p)$. At $T = K^2$ this transform defines the sharp model in S9.

Edge Poisson summation for the complex Gaussian gives

$$|G_h(\theta)| \le R^{1/2} G_{R^2/T}(\theta)$$

with a positive real Gaussian on the right. For $w < W_0$ the effective trace angles lie within $O(W_0/\sqrt{X})$ of the ends since $R\log X/\sqrt{T} \ll W_0/\sqrt{X}$ with a power saving. This gives $O(1 + W_0^2/\sqrt{X})$ points. The angular count $O(1 + w\omega + \sqrt{X}\,\omega^2)$ within distance ω of ϕ bounds $\sum_t |G_h(\phi \pm \theta_t)|$ by

$$O\big(\sqrt{T}R^{-1/2}(1 + W_0R/\sqrt{T} + \sqrt{X}R^2/T)\big).$$

The last two terms in parentheses have negative exponents $\delta + 3\gamma - 1/2$ and $\delta + 2\gamma - 1/2$ at $\eta_1 = 0$. Cauchy–Schwarz with $F_B(t) \ll_\varepsilon X^\varepsilon$ and angular L^1 kernel norm $O(R^{1/2})$ gives normalized squared edge norm $O(X^\varepsilon(T^{-1/2} + W_0^2/\sqrt{pT}))$. The first-sine subtraction $O(w)$ adds $O(W_0^3/(c_pX^{3/2}))$.

Taylor replacement On $w \ge W_0$, set $q = (t_{\mathrm{tr}} - s)\sqrt{T}/w$ and $\rho = \sqrt{p}/(w\sqrt{T})$. The window $|q| \le R\log X$ stays inside the trace range since $\rho R^2 \log^3 X = o(1)$. There the angle expansion gives $\sqrt{T}(\theta_{t_{\mathrm{tr}}} - \phi) = -q + O(\rho q^2)$ uniformly. The change in the exponent is $O(\rho|q|^3/R)$, so the Gaussian error is bounded by

$$O\big(\sqrt{T}R^{-1/2}\rho|q|^3R^{-1}e^{-q^2/(4R^2)}\big).$$

Monotonicity and the Poisson bound discard the reflected kernel and the angular tails. The width $wR/\sqrt{T}$ is power-smaller than the distance to the trace ends, so the polynomial bound for F_B allows extension to integer traces.

Summing the displayed error with lattice spacing $\sqrt{T}/w$ costs

$$O_\epsilon\big(X^\epsilon(\rho\sqrt{T}R^{3/2} + w\rho R^{5/2})\big).$$

The first term covers $w < \sqrt{T}$. Angle bands $w \asymp W$ have length $O(W/\sqrt{p})$, also when $W \asymp \sqrt{p}$. Their total squared error before dividing by c_p is $O_\epsilon(X^\epsilon(R^3/(W_0\sqrt{X}) + R^5/T))$. Poisson summation gives

$$2\pi w \sum_\beta \gamma_B(\beta) e(\beta s) e^{-2\pi^2 z w^2 \beta^2/T}.$$

The zero-mode error is $O(wR/T)$. Its normalized square integrates to $O(R^2/(T^2c_p))$.

Denominator comparison On a band $w \asymp W$ and $T \asymp U$ with fixed U, put $H = W/\sqrt{U}$. The multiplier has real Gaussian absolute value, and the integration by parts of Lemma 3.3 of the article costs a factor $O(R)$ on mixed terms. The amplitude has variation $O(RWe^{-CH^2(\beta^2+\beta'^2)})$. The reciprocal-distance sum at denominator scale E is $O(E^2 \log X)$ with the Gaussian weight. For distances through $\max(1, H^{-1})$ use the spacing $\gg E^{-2}$ (and the minimum nonzero distance $\gg (DE)^{-1}$ for $D \le E$). Beyond that distance use the Gaussian mass $O(E^2(1+H^{-1}))$. The other frequency count at scale D is $O(D^2/H)$. As in Lemma 3.3, this bounds the normalized mixed terms by $O(X^\epsilon RB/\sqrt{T})$, or $O(X^\epsilon RD_0/\sqrt{T})$ for two low denominators. For real weights the opposite frequencies give the diagonal of (S13).

For a better bilinear bound, replace β' by $-\beta'$ and put $Q = \beta^2 + \beta'^2$, $\Delta = \beta - \beta' \neq 0$. Since $w^2 = 4p - s^2$, the phase is $2\pi\Delta s + As^2$ up to a constant, with $A = 2\pi^2(\Im z)Q/T$. On each monotone band the real amplitude has maximum and variation $O(We^{-CH^2Q})$. If $|\Delta|$ exceeds a sufficiently large constant times $|A|\sqrt{X}$, the first derivative gives $O(1/|\Delta|)$. Otherwise the quadratic integral is $O(|A|^{-1/2})$, by splitting where the first derivative is smaller than $\sqrt{|A|}$. Use $|\Delta|/\sqrt{|A|} \ll \sqrt{|A|X}$ and absorb $H\sqrt{Q}$ into a weaker Gaussian. Thus the mixed bound also holds with R replaced by $1 + \sqrt{R}\sqrt{X}/W$ for the bilinear integral.

The normalized diagonal is $O(X^\epsilon)$. Use Lemma 3.3 for $H \ge 1$, and for $H < 1$ sum the integer translates to bound the diagonal by $O(H^{-1})$ times the squared coefficients modulo one, integrated against $w^2\, d\phi/p$. This also bounds the normalized diagonal lost at $w < W_0$ by $O(X^\epsilon W_0^2/\sqrt{pT})$. Put $L_B = 1 + B/\sqrt{T}$. The real Gaussian has normalized squared norm $O(X^\epsilon L_B)$ by the diagonal and mixed bounds. The edge, Taylor and zero-mode bounds give the same for $\mathcal{D}_{1/T}$, hence for $\mathcal{D}_h$ by sine Parseval. They also give this bound for the full complex Gaussian on $w \ge W_0$. In particular the Taylor and zero-mode cross terms cost $O(X^\epsilon L_B^{1/2})$ times the normalized error norms. The low Gaussian has normalized squared norm $O(X^\epsilon)$, since $RD_0/\sqrt{T} = o(1)$. Subtraction bounds the high part pointwise by $O(X^\epsilon L_B)$ in normalized squared norm.

To use the dual sieve of Lemma S5.2 on the high denominators, hold W, U, s fixed. In the weight $e^{-2\pi^2 zvH^2\beta^2}$, the variable $v = (w/W)^2 U/T$ lies in a fixed compact positive interval. For $T = \lambda p^{1-\delta}$, extend it to integers $n \asymp X$ by

$$v(n,s) = \frac{U}{\lambda W^2}(4n - s^2)n^{\delta-1}.$$

For $|s| \le 3\sqrt{X}$ the derivative $U(\lambda W^2)^{-1}(4\delta n^{\delta-1} + (1-\delta)s^2 n^{\delta-2})$ is positive and comparable to W^{-2} with constants depending on δ. Split the v interval into $O(R)$ pieces of length $O(1/R)$. Each restricts the integers to an interval of length $O_\delta(1 + W^2/R)$ at fixed s. Apply the additive large sieve at a fixed v on each piece. In the parameter rescaled to length one, differentiating the weight costs a polynomial in $H|\beta|$ absorbed by a weaker real Gaussian. The fundamental theorem of calculus and Cauchy–Schwarz give the same bound for the supremum over the piece.

Integrate over $|s| \le 3\sqrt{X}$. Since $2B \ll \sqrt{X}$ and $d > D_0 \gg H$, Parseval as in Lemma S5.2 costs $O(d/H)$ per denominator, including $H < 1$. The lengths in the large sieve sum to $O_\delta(W^2 + R(1+B^2))$. The angular factor W/p and the prime average therefore give for the normalized high squared norm

$$\mathbb{E}_X \frac{\|\text{high}\|^2_{\text{ang, band}}}{c_p} \ll X^\varepsilon \frac{W^2 + R(1+B^2)}{XD_0\sqrt{c_p}}.$$

The low–high product in prime L^2 costs the square root of this bound times $O(X^\epsilon)$. For the high square add $L_B^{1/2}$ using its pointwise bound. The high diagonal costs $O(X^\varepsilon/(D_0\sqrt{c_p}))$.

At $\delta = 1/4$ the high mean is $O(X^{-1/8-\vartheta+\varepsilon})$ in both ranges of W. In the lower range use $R(1+B^2)/X \ll X^{-3/16+5u} \ll X^{-2\vartheta}$. Since $L_B \ll X^{2u}$, the high terms cost $O(X^{-1/16-\vartheta/2+u+\varepsilon})$ in

prime L^2. For the low mixed terms, use R in the lower range and $1+\sqrt{R}\sqrt{X}/W$ in the upper range. The exponents are $-1/16-\vartheta+u$ and at most $-3/32+2\vartheta+u/2$, respectively. Both are strictly below $-1/16$. Summing the low norms, high norms and angular errors bounds the normalized squared norm of $\mathcal{D}_h$ in prime L^2 by $O(X^{\varepsilon})$. In the open ranges this follows from the pointwise bound.

Tail, normalization and the final error For $u \geq 0$, $|e^{-zu}| = e^{-u}$ on $\Re z = 1$. Thus the squared diagonal error $O(B^{-1}\log^4 X)$ and first-sine squared error $O(X^{\epsilon}B/p)$ from Lemma S5.2 apply. After removing the first sine, the real heat kernel bound $O(\sqrt{X})$ for the trace sum gives normalized squared tail $O(X^{-1/2}\sum_t |F(t)-F_B(t)|^2/c_p)$ by Cauchy–Schwarz. The Mellin estimate in Lemma S5.1 gives $O_{\varepsilon}(X^{\varepsilon}A(p)/(Bc_p))$. Jensen and the same discriminant multiplicity and fourth moment give $\mathbb{E}_X A(p)^2 \ll_{\varepsilon} X^{\varepsilon}$. With the L^2 bound for the truncated squared norm, the tail cross term costs $O_{\varepsilon}(X^{\varepsilon}(Bc_p)^{-1/2})$. Corollary 4.3 and $|\tau_k(p)| \ll k$ give $M_p(z)$ with hyperbolic error $O(p^{-3/2}/T)$.

The table bounds the exponents in the L^2 error at $\eta_1 = 0$, apart from X^{ε}.

	exponent at $\eta_1 = 0$
high denominators, with cross terms	$-\gamma$
mixed low denominators	$-1/2+\delta+3\gamma$
class-number and Fourier tails	$-\gamma$
Taylor, first term	$-1/4+\delta/4+\gamma/2$
Taylor, second term	$-1/2+\delta+5\gamma/2$
edge and discarded diagonal	$\max(-1/2+\delta/2,\ -1+3\delta/2+4\gamma,\ -3/2+5\delta/2+6\gamma)$
first-sine mass and zero mode	$\max(-1+2\delta+2\gamma,\ -1+3\delta/2+\gamma)$

The positive increments in d_* and β_* make the first and third rows strictly smaller than $-\gamma$. Every other row has a strict margin for $\gamma = \delta/4$ at $\delta < 1/4$, and for $\gamma = \delta/2$ at $\delta < 1/6$. In both cases the limiting inequality is $\delta + 4\gamma < 1/2$ from the low mixed term. At $\delta = 1/4$ the band estimates above replace the first two rows. There $(Bc_p)^{-1/2} \asymp X^{-1/16-u}$, and the remaining angular and mass errors retain strict margins even with the factor $L_B^{1/2}$. Indeed, their exponents at $u = 0$ are at most $-3/32$, and they increase by at most $10u$. Thus in all these ranges, for some $\xi > 0$,

$$\sup_{|\Im z| \leq X^{\kappa},\ \Re z = 1} \|M_p(z) - M_p^{\mathrm{diag}}(z)\|_2 \ll_{\delta,\lambda} X^{-\gamma-\xi}. \tag{S14}$$

For the estimate at $z = 1$ set $R = 1$ and $\kappa = 0$ throughout. With $\gamma = \delta/2$, the low mixed term has exponent $-1/2+2\delta$ before the small increments, and $B/\sqrt{T} = o(1)$. Both requirements and the other strict margins hold for $0 < \delta < 1/5$.

S8.6 Comparison with the count statistic

For $0 < \delta < 1/4$ the same real-parameter bounds also give

$$\|Y_p - F_{x_p}(A_p)\|_{L^2(\mathbb{E}_X)} \ll_{\delta,\lambda,\varepsilon} X^{\varepsilon}\left(X^{-\delta/2} + X^{-1/8+\delta/4} + X^{-3/16+\delta/2}\right). \tag{S15}$$

Set $R = 1$, $W_0 = \sqrt{T}$, $D_0 = X^{1/4}$ and $B = X^{3/8}$ in the heat estimates. The edge contributes the first term, and the high denominators and the tail contribute $(D_0\sqrt{c_p})^{-1/2}$ and $(Bc_p)^{-1/2}$, up to X^{ε}. The low mixed term $D_0/\sqrt{T}$, the Taylor error after Cauchy–Schwarz and normalization, and the first-sine and zero-mode errors are smaller.

Corollary 4.3 and Lemma S5.2 give relative error $O(T^{-1/2}\log^3 X)$ between the binomial and heat multipliers up to degree $\sqrt{T}\log X$. Rounding T costs less, and $|\tau_k(p)| \ll k$ controls the

remaining tail. The same comparison works for the Gaussian trace sum with T_p in place of T. Both L^2 errors are $O_\varepsilon(X^\varepsilon T^{-1/2})$: (S14) bounds $\|M_p(1)\|_2$ by $O_\varepsilon(X^\varepsilon)$ since Weil's inequality gives $F_{x_p}(A_p) \ll \log^3(2/x_p)$, splitting at $d = 1/x_p$ with numerator weight sum $O(\min((x_p d)^3, x_p d))$. Thus the count error is $o(X^{-\delta/4})$ for $\delta < 1/4$ by (S15), and at $\delta = 1/4$ by (S14) at $z = 1$. For $\delta < 1/5$ the real estimate with $\gamma = \delta/2$ gives $o(X^{-\delta/2})$.

S8.7 A prime mean-square bound

For primitive characters of conductor $f \le X^\theta$ with fixed $0 < \theta < 1/2$, put $D = X^\theta$. Then

$$\mathbb{E}_X \left| \sum_\chi a_\chi \chi(p) \right|^2 \ll_\theta \sum_\chi \frac{f_\chi}{\varphi(f_\chi)} |a_\chi|^2. \tag{S16}$$

Take $R = X^\beta$ with $0 < 2\beta < 1 - 2\theta$ and majorize the prime indicator on $(X, 2X]$ by $(\sum_{d|n} \lambda_d)^2$, with $\lambda_d = \mu(d) \log(R/d)/\log R$ for $d \le R$ and zero otherwise. Set $n = [d, e]n'$ in the expansion. For $r = [d, e]$, characters with $(r, f) > 1$ vanish and the others acquire phases $\chi(r)$. Gauss expansion gives additive frequencies spaced at least D^{-2} apart and diagonal $(\varphi(f)/f) \sum^*_{\chi \bmod f} |a_\chi|^2$ by orthogonality. The Montgomery–Vaughan mean-square estimate gives X/r times the diagonal with absolute error $O(D^2 \sum_\chi |a_\chi|^2)$, allowing signed sieve coefficients at total error $O(D^2 R^2 \sum_\chi |a_\chi|^2)$.

For each fixed f, the diagonal sieve factor is

$$\sum_{(de,f)=1} \frac{\lambda_d \lambda_e}{[d, e]} = \sum_{(\ell,f)=1} \varphi(\ell) \left(\sum_{\ell | d,\ (d,f)=1} \frac{\lambda_d}{d} \right)^2 \ll \left(\frac{f}{\varphi(f)} \right)^2 / \log R.$$

Only squarefree ℓ contribute. With $d = \ell u$ and $m = \ell f$, use

$$\sum_{\substack{u \le Y \\ (u,m)=1}} \frac{\mu(u)}{u} \log(Y/u) \ll m/\varphi(m).$$

For $m = 1$, use the harmonic convolution $\sum_{u \le Y} \mu(u) H_{\lfloor Y/u \rfloor}/u = 1$. Replacing the logarithm by $H_{\lfloor Y/u \rfloor} - \gamma$ costs $O(1)$. The identity $\sum_{u \le Y} \mu(u) \lfloor Y/u \rfloor = 1$ also gives $\sum_{u \le Y} \mu(u)/u = O(1)$. Convolve with the indicator of integers supported on primes of m for the coprime version. Hence the inner sum is $O((f/\varphi(f))/(\varphi(\ell) \log R))$. Sum $\varphi(\ell)^{-1}$ to get $O(\log(2R))$, multiply by the diagonal factor $\varphi(f)/f$ and divide by the prime count. Since $D^2 R^2 \le X^{1-\xi}$ for some $\xi > 0$, the sieve error saves a power, proving (S16).

S8.8 Averaging the centered moments over primes

Put $B = X^{\delta/2+1/1000}$, $D = X^{1/4}$ and $G_x = G_{x,B}$. Equations (S16) and (S11) bound the centered band $B < d \le D$ and its first logarithmic derivative by $O(\sqrt{V(x_X)/(x_X B)})$ in prime L^2. For $x = x_p$, apply Sobolev on $x = x_X y^{-\delta/2}$ with $1 \le y \le 2$ before averaging. The Weil tail above D is uniformly $o(\sqrt{V(x_X)})$. The count comparison in S8.6 and the Haar mean give

$$\|Y_p - 1/48 - G_{x_p}(A_p)\|_{L^2(\mathbb{E}_X)} = o(\sqrt{V(x_X)}). \tag{S17}$$

The same tail bounds in Haar L^2 give $\mathbb{E}_A G_x^2 = \operatorname{Var}_A F_x + o(V(x_X))$ uniformly for $x \asymp x_X$.

The conductor-f Fourier coefficients of G_x and G_x^2, and each fixed logarithmic derivative, are at most

$$O_\varepsilon\big(V_* f^{-1/2+\varepsilon}\big), \qquad V_* = \sqrt{V(x_X)} \quad \text{or} \quad V(x_X), \tag{S18}$$

respectively. For G_x this is (S10). For the square, a pair with product ψ has conductors whose least common multiple is divisible by f. Assign each exact prime power of f to an input conductor it divides. One character determines the other. Cauchy–Schwarz and (S10) bound the convolution for each assignment, with a smaller intermediate ε absorbing the $2^{\omega(f)}$ choices. Derivatives follow by the product rule. All output conductors satisfy $f \le B^2 \le X^{0.252}$.

For each nonprincipal coefficient C_ψ, apply the explicit formula to $\Lambda(n)\psi(n)$ with weight $w_0(n/X)C_\psi(x_X(n/X)^{-\delta/2})$, where w_0 is smooth and compactly supported in $(0,\infty)$. By (S18) its Mellin transform is $O_{w_0,N,\varepsilon}(V_* f^{-1/2+\varepsilon}(1+|\Im s|)^{-N})$. After division by X and summation over characters, prime powers cost $O_\varepsilon(V_* X^{-1/2}(B^2)^{3/2+\varepsilon}\log^C X) = o(V_*)$. The ordinary zero count gives the same bound for $\Re s < 1/2$. The trivial zero at zero and the integral at $\Re s = -1/2$ cost less by the functional equation's bound $O(\log(f(2+|\Im s|)))$ for the logarithmic derivative on the shifted contour.

For the remaining zeros, the hybrid density bound with exponent $5/2$, recorded by Koukoulopoulos [23, (1.2)], and the ordinary zero count give

$$\ll H^{O(1)} \log^C(2QH) \min(Q^2, Q^{5u})$$

zeros in total for $f \asymp Q = X^h$, height at most $H \ge 2$, and real part at least $1-u \ge 1/2$. Weighting by $f^{-1/2+\varepsilon}X^{-u}$ gives the exponent, apart from εh,

$$-h/2 + \min(2h, 5hu) - u \le \max(-h/2, 3h/2 - 2/5) \le -h/12 \qquad (0 < h \le 0.252).$$

Partition u into intervals of length $1/\log X$. The factors Q^{5u} and X^{-u} vary by bounded factors on each interval. The rapid height decay sums the height bands. For $Q \ge (\log X)^C$ with C large this bound, summed over Q, is $o(V_*)$, taking $\varepsilon < 1/24$. For the smaller conductors the Siegel–Walfisz theorem and partial summation, with the coefficient and derivative bounds (S18), give $o(V_*)$ after summation. Thus the nonprincipal terms in both moments have size $o(V_*)$. The conductor decay covers even a real zero close to 1.

With weights $w_0(p/X)\log p/X$, the first moment is $o(\sqrt{V(x_X)})$ and the second agrees with the weighted Haar variance to $o(V(x_X))$. Bracket the dyadic indicator by fixed nonnegative smooth weights whose integrals differ by at most ϵ. The uniform Haar bound and the prime number theorem bound the second moment on their difference by $O(\epsilon V(x_X)) + o(V(x_X))$. Cauchy–Schwarz bounds the first moment there by $O(\epsilon\sqrt{V(x_X)}) + o(\sqrt{V(x_X)})$. Let $X \to \infty$ and then $\epsilon \to 0$. The prime number theorem replaces $\log p/X$ uniformly by the mass of a prime in $\mathbb{E}_X$, with the second moment controlling the first. Equation (S17) then transfers the variance to Y_p by Cauchy–Schwarz. Since $\mathbb{E}_X[x_p \log(1/x_p)] \sim (\delta\sqrt{\lambda}/2)J_\delta X^{-\delta/2}\log X$, the constant follows from (S12). The Gaussian trace sum differs from Y_p by $o(\sqrt{V(x_X)})$ in prime L^2 and therefore has the same variance asymptotic.

S9 Sharp-cutoff variance and second terms for the mean

Fix $0 < \delta \le 1/4$ and set $K = p^{(1-\delta)/2}$, $c_p = K^2/p$, $x_p = \sqrt{c_p}$ and $x_X = X^{-\delta/2}$. For the normalized level-one traces put

$$S_p = \frac{1}{K^2} \sum_{\substack{4 \le k \le K \\ k \text{ even}}} |\tau_k(p)|^2.$$

With $b(y)$ from (S13), set $W_\#(v) = 2\int_0^1 \sigma(y) 1_{\{b(y)v^2 \le 1\}}\, dy$ and $F_x^\# = F_{W_\#,x}$ in the notation of S8. The heat estimate from Section S8.5 and Fourier inversion recover the sharp cutoff.

Corollary S9.1. *For every fixed $0 < \delta \le 1/4$ and some $\xi_\delta > 0$,*

$$\|S_p - F^{\#}_{x_p}(A_p)\|_{L^2(\mathbb{E}_X)} \ll_\delta X^{-\delta/4-\xi_\delta}. \tag{S19}$$

Consequently

$$\operatorname{Var}_X S_p = \mathbb{E}_X \operatorname{Var}_A F^{\#}_{x_p}(A) + o\big(x_X \log(2/x_X)\big) \sim \frac{\delta}{4} J_\delta \,\mathfrak{A}\, I(W_\#) X^{-\delta/2} \log X. \tag{S20}$$

S9.1 Recovering the sharp weight cutoff

Let ν_p put mass $|\tau_k(p)|^2/K^2$ at $u = (k-1)^2/K^2$ and let μ_p have Laplace transform $M_p^{\mathrm{diag}}(z)$ at $T = K^2$. The real heat bound controls the masses of the positive tilted measures $e^{-u}\,d\nu_p$ and $e^{-u}\,d\mu_p$ in prime L^2 by $O_\varepsilon(X^\varepsilon)$, and hence their Fourier transforms.

For $0 < \eta < 1/4$, the boundary kernel is $W_\#(v/\sqrt{1+\eta}) - W_\#(v/\sqrt{1-\eta})$. Solving the quadratic inequality in y shows that the kernel vanishes below $v_0\sqrt{1-\eta}$, where $v_0 = 1/(4\pi)$. It is $O(\sqrt{\eta})$ for $|v - v_0| \ll \eta$ and $O(\eta/\sqrt{h})$ at bounded v with $|v - v_0| \asymp h \gg \eta$. At large v, the intervals have length $O(\eta/v^2)$ and density $O(1/v)$.

For fixed numerator a and $x = x_p$, a bounded band of width h in $v = a/(xd)$ contains $O((a/x)h+1)$ denominators $d \asymp a/x$. The dyadic bands contribute $O(\eta a/x + \sqrt{\eta})$. Weight by $x^{-2}d^{-3+\varepsilon}$ and sum over a to obtain $O_\varepsilon(x^{-\varepsilon}(\eta + x\sqrt{\eta}))$. The large-$v$ sum with kernel $O(\eta(xd/a)^3)$ over $d \ll a/x$ and a is $O_\varepsilon(x^{-\varepsilon}\eta)$. Thus $\mu_p([1-\eta, 1+\eta]) \ll_\varepsilon X^\varepsilon(\eta + x_p\sqrt{\eta})$.

Choose $\gamma < b_* < \kappa$ and $\eta = X^{-b_*}$. The displacement between the threshold $(1-1/K)^2$ and 1 is $o(\eta)$. Bracket both indicators on $u \ge 0$ by smooth $\varphi_\pm$ agreeing with them off $[1-2\eta, 1+2\eta]$ and extend smoothly to uniformly compact support. Test the e^{-u}-tilted measures against $e^u\varphi_\pm(u)$. Bounded supremum and first derivative L^1 norms, and higher derivative norms $O_j(\eta^{1-j})$, give Fourier L^1 norm $O(\log(2/\eta))$ and tail norm $O_j((\eta X^\kappa)^{1-j})$ beyond X^κ. Fourier inversion and Minkowski use (S14) inside that range and the mass bound outside, giving a power saving over $X^{-\gamma}$. The model boundary mass has the same power margin. Positivity and the brackets give $\|S_p - F^{\#}_{x_p}(A_p)\|_2 \ll X^{-\gamma-\xi'}$ for some $\xi' > 0$. Taking $\gamma = \delta/4$ proves (S19).

S9.2 Variance of the sharp profile

The sharp kernel is nonincreasing and $O(\min(1, v^{-3}))$. Its dyadic supremum and variation bounds give (S12) with $H = W_\#$. Its integral can also be evaluated. Under $\sigma(y)\,dy$ the variable $q = 2\sqrt{y(1-y)}$ has density $4q^2/(\pi\sqrt{1-q^2})$. Integrate v^4 up to $1/(4\pi\max(q, q'))$ for two independent copies, and set $q = \sin\theta$ on the larger variable. This gives

$$I(W_\#) = \frac{1}{80\pi^7}\int_0^{\pi/2} (\theta - \sin\theta\cos\theta)\csc^3\theta\,d\theta = \frac{G + 1/2}{80\pi^7}, \tag{S21}$$

where G is Catalan's constant. For the last equality, integration by parts reduces the integral to $1 - \int_0^{\pi/2}\sin^2\theta\,\log\tan(\theta/2)\,d\theta$. The integral of $\log\tan(\theta/2)$ is $-2G$, and its integral against $\cos 2\theta$ is -1 by integration by parts. The Mellin transform of the sharp kernel is $\widehat{W}_\#(s) = \frac{2}{s}\int_0^1 \sigma(y)b(y)^{-s/2}\,dy$ for $0 < \Re s < 3$. In particular $\widehat{W}_\#(2) = 1/(8\pi^2)$, as for W. The mean calculation below gives $\mathbb{E}_A F_x^{\#} = 1/48 + O(x)$.

The prime moment proof uses logarithmic derivatives. To obtain them, smooth $W_\#$ in $\log v$. Its only derivative singularity is at $v_0 = 1/(4\pi)$. For $u = \log(v/v_0) > 0$,

$$\frac{d}{du}W_\#(v_0e^u) = -\frac{8}{\pi}\frac{e^{-3u}}{\sqrt{1-e^{-2u}}},$$

and the derivative vanishes for $u < 0$. For $|h| < 1/4$, the supremum and dyadic variation of $W_\#(ve^h) - W_\#(v)$ are $O(\sqrt{|h|}\min(1, v^{-3}))$. Integrate the $O(u^{-1/2})$ derivative up to $2|h|$, and use $O(|h|u^{-3/2})$ for its difference beyond that point. Away from the singularity the derivatives decay exponentially at infinity.

Let ρ_{η_0} be a positive smooth bump of mass one supported on $|h| \le \eta_0 = (\log X)^{-C_0}$. Set

$$W_{\eta_0}(v) = \frac{\int W_\#(ve^h)\rho_{\eta_0}(h)\,dh}{\int e^{-2h}\rho_{\eta_0}(h)\,dh}.$$

This preserves the Mellin value at 2. The dyadic supremum and variation of the difference are $O(\sqrt{\eta_0}\min(1, v^{-3}))$. Differentiation of the bump costs a fixed power of $\log X$.

The estimates (S10) and (S11) use only the supremum and variation bounds. The centered prime-moment comparison of S8 therefore applies for the same B, D: the tail estimates save powers, the conductor decay absorbs any logarithmic derivative loss above a sufficiently large power of $\log X$, and Siegel–Walfisz absorbs it below that cutoff. The smoothed Haar mean is $1/48 + O(x)$ as it averages the Haar means at xe^{-h} with normalized weights $e^{-2h}\rho_{\eta_0}(h)$.

The translation and Fourier energy bounds give Haar squared norm $O(\eta_0 x\log(2/x))$ for the centered difference. In both profiles the supremum of the tail above $D = X^{1/4}$ is $O(x^{-1}D^{-1}\log^3 X) = o(\sqrt{V(x_X)})$. Freeze x at the left endpoint of each interval of length $\asymp X^{1-b_*}$, with $\delta/2 < b_* < 1/2$. The translation bound and the factor x^{-2} give error $O(X^{-b_*/2}\log^C X) = o(\sqrt{V(x_X)})$. Haar Parseval bounds the truncated coefficient square sum by $O(\eta_0 x_X\log(2/x_X) + X^{-\epsilon_1}x_X)$ for some $\epsilon_1 > 0$, using the power saving in the tail. The dual multiplicative large sieve costs $O(X^{1-b_*} + D^2)$ per interval. Since $D^2 < X^{1-b_*}$, restricting the result to primes costs $O(\log X)$ relative to the Haar square sum. With $C_0 > 3$, the centered difference is $o(\sqrt{V(x_X)})$ in prime L^2. The mean difference is $O(x) = o(\sqrt{V(x_X)})$. The variance comparison passes to the sharp profile and then to S_p by (S19), proving (S20).

S9.3 Second terms for the mean

The sharp Haar mean has the expansion

$$\mathbb{E}_A F_x^\#(A) = \frac{1}{48} - \frac{x}{1728} + o(x). \tag{S22}$$

Put $P(z) = z^{-3}W_\#(1/z)$, so $P(0) = 1/(24\pi^4)$ by the density calculation. The square-root cusp leaves P absolutely continuous with integrable derivative. As $j \downarrow 0$,

$$\sum_{d\ge1} P(jd) = j^{-1}\int_0^\infty P(z)\,dz - \tfrac{1}{2}P(0) + o(1).$$

The remainder is the integral of P' against a bounded, mean-zero sawtooth of period j. It is uniformly bounded and tends to zero by L^1 approximation by step functions. Dropping coprimality and replacing $Y_{a,d}$ by 1 gives

$$F^0(x) = \sum_{a\ge1}\frac{x}{a^3}\sum_{d\ge1}P(xd/a) = \zeta(2)\int_0^\infty P(z)\,dz - \tfrac{1}{2}x\zeta(3)P(0) + o(x).$$

The residual after subtracting the constant is $O(x)$ for all $x > 0$.

To restore coprimality and $g(d)$, expand the Euler product of the Gaussian mean as $\mathcal{G}(s) = \sum_{u,v} c_{u,v}u^{-s}v^{s-3}$. The local expansion there converges absolutely with the weights at $s = 2$ and

$s = 3$. Multiplying the numerator and denominator series gives $\mathbb{E}_A F_x^{\#} = \sum_{u,v} c_{u,v} u^{-2} v^{-1} F^0(xv/u)$. Dominated convergence at the linear term uses the convergence at $s = 3$. The two coefficients are therefore multiplied by $\mathcal{G}(2) = 1$ and $\mathcal{G}(3) = \zeta(2)^2/\zeta(3)$, respectively. Since $\int P = \widehat{W}_{\#}(2) = 1/(8\pi^2)$, this gives (S22).

Corollary S9.2 (Second terms for the prime means)**.** *With the normalization of Corollary S9.1, for* $0 < \delta < 1/6$*,*

$$\mathbb{E}_X S_p = \frac{1}{48} - \frac{J_\delta}{1728} X^{-\delta/2} + o(X^{-\delta/2}).$$

With the normalization of Theorem S8.1, for $0 < \delta < 1/5$ *and fixed* $\lambda > 0$*,*

$$\mathbb{E}_X Y_p = \frac{1}{48} - \frac{\sqrt{\pi\lambda}\, J_\delta}{2304} X^{-\delta/2} + o(X^{-\delta/2}).$$

Proof. Take $\gamma = \delta/2$ in the sharp cutoff recovery and the real heat comparison. Both errors are $o(x_X)$ in prime L^2 in the stated ranges. Truncate at $D = X^\theta$ with $\delta < \theta < 1/4$, at cost $O(x_X^{-1} D^{-1} \log^3 X) = o(x_X)$. At denominator d the nonprincipal coefficients have ℓ^1 norm and total variation as x ranges over the prime interval bounded by

$$\ll \tau(d)^2 \sqrt{d}\, \min(x_X, (d^2 x_X)^{-1}).$$

Indeed the local product $Y_{a,d}$ is $O(\tau(d)^2)$ by Weil's bound, so Parseval bounds its absolute character sum by $O(\sqrt{d}\,\tau(d)^2)$. For each a, the kernel is monotone in x, with supremum $O(\min(1, (x_X d/a)^3))$. Summing these bounds with the factor $x^{-2} d^{-3}$ gives the assertion, also for variation. Summing over d costs $O(x_X^{-1/2} \log^C X)$.

For $p > X > D$, a character modulo $d \le D$ agrees at p with its primitive inducing character of conductor $f \mid d$. Alpöge's Dirichlet theorem [2, 39] puts the zeros in the explicit formula for a nonprincipal primitive character on $\Re s = 1/2$. The zero count gives $\sum_{n \le u} \Lambda(n)\chi(n) \ll X^{1/2} \log^2(fX)$ for $X \le u \le 2X$. Remove prime powers and use partial summation to bound unweighted prime sums on subintervals by $O(X^{1/2} \log^C X)$. With the coefficient variation above the prime average error is $O(X^{-1/2} x_X^{-1/2} \log^C X) = o(x_X)$ for $\delta < 2/3$. The principal coefficients give the Haar means. The prime number theorem averages $p^{-\delta/2}$ to $J_\delta X^{-\delta/2}(1 + o(1))$. □

Nonprincipal coefficients can have the size of the second term. For an odd prime ℓ and fixed $t > 0$, the quadratic character χ_ℓ satisfies $\ell^2 \widehat{F}_{H,t/\ell^2}(\chi_\ell) \to \mathbb{E}_A F_{H,t}(A) - 1/48$. Indeed the local exponents 1 and 2 give the means at scales t/ℓ and t, multiplied by $-h_\ell/(\ell(\ell-1))$ and h_ℓ/ℓ^2, respectively, with ℓ omitted from the numerator and cofactor. At scale $y = t/\ell$ the omissions cost $O(y^{-1}\ell^{-2} + \ell^{-2})$, and at scale t they tend to zero by dominated convergence. The higher exponents $k \ge 3$ contribute $O_t(\ell^{-4})$ by the bound $O(1/y)$ for the mean at scale $y = t\ell^{k-2}$.

The article's AI Declaration covers the arguments, numerical checks and writing in this supplement.

References

[1] M. C. Abbott and B. B. Machta, A scaling law from discrete to continuous solutions of channel capacity problems in the low-noise limit, *J. Stat. Phys.* **176** (2019), 214–227.

[2] L. Alpöge, announcement of a formal proof of the Riemann hypothesis and of its analogue for Dirichlet L-functions, 20 November 2026, `https://alpo.ge/remark.pdf`.

[3] A. O. L. Atkin and W.-C. W. Li, Twists of newforms and pseudo-eigenvalues of W-operators, *Invent. Math.* **48** (1978), 221–243.

[4] J. Bober, A. R. Booker, M. Lee and D. Lowry-Duda, Murmurations of modular forms in the weight aspect, *Algebra Number Theory* **20** (2026), 299–332.

[5] D. A. Burgess, The character sum estimate with $r = 3$, *J. London Math. Soc.* (2) **33** (1986), 219–226.

[6] T. Cochrane and C. Pinner, Using Stepanov's method for exponential sums involving rational functions, *J. Number Theory* **116** (2006), 270–292.

[7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd ed., Wiley, 2006.

[8] D. A. Cox, *Primes of the Form* $x^2 + ny^2$, 2nd ed., Wiley, 2013.

[9] C. David, D. Koukoulopoulos and E. Smith, Sums of Euler products and statistics of elliptic curves, *Math. Ann.* **368** (2017), 685–752.

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

[11] P. Diaconis, K. Khare and L. Saloff-Coste, Gibbs sampling, exponential families and orthogonal polynomials, *Statistical Science* **23** (2008), 151–178.

[12] L. Fu and D. Wan, L-functions for symmetric products of Kloosterman sums, *J. reine angew. Math.* **589** (2005), 79–103.

[13] E.-U. Gekeler, Frobenius distributions of elliptic curves over finite prime fields, *Int. Math. Res. Not.* (2003), no. 37, 1999–2018.

[14] D. R. Heath-Brown, A mean value estimate for real character sums, *Acta Arith.* **72** (1995), 235–275. doi:10.4064/aa-72-3-235-275.

[15] B. Huang, J. Liu and Z. Rudnick, Gaussian primes in almost all narrow sectors, *Acta Arith.* **193** (2020), 183–192.

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

[17] Y. Jiang and Y.-K. Lau, Mixed moments of arithmetic functions and Kloosterman sums, preprint, arXiv:2610.10372 (2026).

[18] J. Jung and N. Talebizadeh Sardari, Asymptotic trace formula for the Hecke operators, with an appendix by S. Marshall, *Math. Ann.* **378** (2020), 513–557.

[19] I. Kaneko and J. Thorner, Highly uniform prime number theorems, *Ann. Inst. Fourier* **75** (2025), 1901–1923.

[20] N. Kaplan and I. Petrow, Elliptic curves over a finite field and the trace formula, *Proc. Lond. Math. Soc.* **115** (2017), 1317–1372.

[21] N. M. Katz, *Gauss Sums, Kloosterman Sums, and Monodromy Groups*, Annals of Mathematics Studies 116, Princeton University Press, 1988.

[22] N. M. Katz, $2, 3, 5$, Legendre: $\pm$ trace ratios in families of elliptic curves, *Experimental Mathematics* **19** (2010), 267–277.

[23] D. Koukoulopoulos, Primes in short arithmetic progressions, *Int. J. Number Theory* **11** (2015), 1499–1521. doi:10.1142/S1793042115400035.

[24] R. E. Krichevsky and V. K. Trofimov, The performance of universal encoding, *IEEE Trans. Inform. Theory* **27** (1981), 199–207.

[25] E. Kowalski and T. Untrau, Wasserstein metrics and quantitative equidistribution of exponential sums over finite fields, arXiv:2505.22059.

[26] C. I. Kuan and D. Lesesvre, Murmurations using Petersson trace formula, arXiv:2507.11418.

[27] B. Lee, Quantitative Sato–Tate in near-square-root prime intervals under symmetric-power GRH, preprint, version 3.0, 2026. doi:10.5281/zenodo.22112048.

[28] H. W. Lenstra, Jr., Factoring integers with elliptic curves, *Ann. of Math.* (2) **126** (1987), 649–673.

[29] Z. Y. Ma, Refinements on vertical Sato–Tate, *Acta Arith.* **216** (2024), 63–97.

[30] G. Martin, Dimensions of the spaces of cusp forms and newforms on $\Gamma_0(N)$ and $\Gamma_1(N)$, *J. Number Theory* **112** (2005), 298–331.

[31] H. L. Montgomery, The analytic principle of the large sieve, *Bull. Amer. Math. Soc.* **84** (1978), 547–567.

[32] H. L. Montgomery and R. C. Vaughan, The large sieve, *Mathematika* **20** (1973), 119–134.

[33] I. Petrow, Bounds for traces of Hecke operators and applications to modular and elliptic curves over a finite field, *Algebra Number Theory* **12** (2018), 2471–2498.

[34] I. Petrow and M. P. Young, The fourth moment of Dirichlet L-functions along a coset and the Weyl bound, *Duke Math. J.* **172** (2023), 1879–1960.

[35] A. M. Rahimi, Wasserstein rates of all orders for the Sato–Tate law of Kloosterman sums, arXiv:2610.16169v2.

[36] C. S. Rajan, Distribution of values of Hecke characters of infinite order, *Acta Arith.* **85** (1998), 279–291.

[37] E. Ross, Newspaces with nebentypus: an explicit dimension formula and classification of trivial newspaces, *J. Number Theory* **278** (2026), 317–352.

[38] J. H. Silverman, *The Arithmetic of Elliptic Curves*, 2nd ed., Graduate Texts in Mathematics 106, Springer, 2009.

[39] A. V. Sutherland, RH Lean proof: verification files, `https://math.mit.edu/~drew/grh/`, as updated 21 November 2026.

[40] D. Zagier, Modular forms whose Fourier coefficients involve zeta-functions of quadratic fields, in *Modular Functions of One Variable VI*, Lecture Notes in Math. 627, Springer, 1977, 105–169.