Biases in the Shanks–Rényi Prime Number Race

Andrey Feuerverger and Greg Martin

CONTENTS

Rubinstein and Sarnak investigated systems of inequalities of the form $\pi(x;q,a_1) > \cdots > \pi(x;q,a_r)$, where $\pi(x;q,b)$ denotes the number of primes up to x that are congruent to b mod q. They showed, under standard hypotheses on the zeros of Dirichlet L-functions mod q, that the set of positive real numbers x for which these inequalities hold has positive (logarithmic) density $\delta_{q;a_1,\dots,a_r} > 0$. They also discovered the surprising fact that a certain distribution associated with these densities is not symmetric under permutations of the residue classes a_j in general, even if the a_j are all squares or all nonsquares mod q (a condition necessary to avoid obvious biases of the type first observed by Chebyshev). This asymmetry suggests, contrary to prior expectations, that the densities $\delta_{q;a_1,\dots,a_r}$ themselves vary under permutations of the a_j.

Here we derive (under the hypotheses used by Rubinstein and Sarnak) a general formula for the densities $\delta_{q;a_1,\dots,a_r}$, and we use this formula to calculate many of these densities when $q \le 12$ and $r \le 4$. For the special moduli $q = 8$ and $q = 12$, and for $\{a_1, a_2, a_3\}$ a permutation of the nonsquares $\{3, 5, 7\}$ mod 8 and $\{5, 7, 11\}$ mod 12, respectively, we rigorously bound the error in our calculations, thus verifying that these densities are indeed asymmetric under permutation of the a_j. We also determine several situations in which the densities $\delta_{q;a_1,\dots,a_r}$ remain unchanged under certain permutations of the a_j, and some situations in which they are provably different.

1. INTRODUCTION AND SUMMARY

In 1853 Chebyshev remarked that there are more primes congruent to 3 than to 1 modulo 4, and since that time considerable efforts have been expended in attempts to determine in what sense this remark is true. It follows from the prime number theorem for arithmetic progressions (see [Davenport 1980], for instance) that, asymptotically, half of all primes are congruent to 3 mod 4 and half are congruent to 1 mod 4, so that Chebyshev's observation cannot be interpreted in that sense. However, when we compute the numbers of primes up to x that are congruent to 3 mod 4 and to 1 mod 4, we find that for most

Key words: Chebyshev's bias, comparative prime number theory, primes in arithmetic progressions, Shanks–Rényi race.

values of x, the primes congruent to 3 are more numerous than those congruent to 1. Similar "biases" have also been observed, notably by Shanks [1959], for moduli q other than 4; in particular, the numbers of primes in nonsquare residue classes modulo q tend to exceed the numbers of primes in square residue classes. We refer to inequities of this type as "Chebyshev biases".

These observations lead naturally to the study of inequalities of the type

$$\pi(x;q,a_1) > \pi(x;q,a_2) > \cdots > \pi(x;q,a_r), \qquad (1\text{–}1)$$

where $\pi(x;q,a)$ denotes the number of primes $p \le x$ such that $p \equiv a \bmod q$. Littlewood [1914] showed (unconditionally) that the inequalities

$$\pi(x;3,1) > \pi(x;3,2) \quad \text{and} \quad \pi(x;4,1) > \pi(x;4,3),$$

as well as the opposite inequalities, each hold for infinitely many integer values of x. A number of additional results on single inequalities of this type were subsequently derived under certain hypotheses by Knapowski and Turán in a series of papers beginning with [Knapowski and Turán 1962], and Kaczorowski wrote several papers concerning the multiple inequalities (1–1), the most recent of which is [Kaczorowski 1996].

A major advance was made recently by Rubinstein and Sarnak [1994] who showed (conditionally) that for any modulus q and for any distinct reduced residues $a_1, \dots, a_r \bmod q$ (i.e., integers relatively prime to q), the system of inequalities (1–1) holds for infinitely many integers x. More precisely, they worked under the assumption of the Generalized Riemann Hypothesis for Dirichlet L-functions, which we shall abbreviate as GRH, and an additional assumption (their "Grand Simplicity Hypothesis"), which we abbreviate as LI:

Hypothesis LI. *The nonnegative imaginary parts of the nontrivial zeros of Dirichlet L-functions corresponding to primitive characters are linearly independent over the rationals.*

Rubinstein and Sarnak studied the quantities

$$\delta_{q;a_1,\dots,a_r},$$

defined as the logarithmic density of the set of positive real numbers x for which the inequalities (1–1) hold. (The logarithmic density $\delta(\Lambda)$ of any subset Λ of the real numbers is defined as

$$\delta(\Lambda) = \lim_{x\to\infty} \frac{1}{\log x} \int_{\Lambda\cap[2,x]} \frac{dt}{t},$$

provided that this limit exists. Suffice it to say here that logarithmic densities are more appropriate for these problems than ordinary densities; in this paper, by "density" we shall always mean logarithmic density.)

Under the hypotheses above, Rubinstein and Sarnak proved that the densities $\delta_{q;a_1,\dots,a_r}$ exist and are positive for any integer $q \ge 2$ and for any distinct reduced residues $a_1, \dots, a_r \bmod q$. They obtained, for several small moduli q, numerical values for the density of those x for which the primes up to x that are quadratic nonresidues mod q outnumber those that are quadratic residues. Rubinstein and Sarnak also proved that $\delta_{q;a,a'} = \delta_{q;a',a} = \frac{1}{2}$ if a and a' are both squares or both nonsquares mod q, and otherwise $\delta_{q;a,a'}$ is greater than or less than $\frac{1}{2}$ according to whether a or a' is the nonsquare mod q, thus bearing out the biases of the type observed by Chebyshev.

It was generally suspected for $r > 2$ as well that whenever the a_j are all squares or all nonsquares modulo q, the densities $\delta_{q;a_1,\dots,a_r}$ are invariant under permutations of the a_j (and thus equal to $1/r!$). However, Rubinstein and Sarnak showed that certain distributions $\mu_{q;a_1,\dots,a_r}$ on $\mathbb{R}^r$ that are associated naturally with the densities $\delta_{q;a_1,\dots,a_r}$ are not symmetric under permutations of the a_j when $r \ge 3$, except in the special case when $r = 3$ and there exists $\rho \not\equiv 1 \bmod q$ with $\rho^3 \equiv 1 \bmod q$ such that $a_2 \equiv a_1\rho \bmod q$ and $a_3 \equiv a_1\rho^2 \bmod q$. (Note that since $\rho \equiv \rho^4 \bmod q$ is a square, it follows that such $\{a_1, a_2, a_3\}$ are all squares or all nonsquares mod q.) This result suggests, but does not imply, that the $\delta_{q;a_1,\dots,a_r}$ are generally asymmetric under permutation of the a_j.

In this paper, we rigorously establish a number of asymmetries of this type, under the two aforementioned hypotheses. Triples of nonsquares and triples of squares occur for the moduli $q = 7$ and $q = 9$, but these triples fall under the special case that has just been mentioned. Thus the smallest moduli for which such asymmetries of the $\delta_{q;a_1,a_2,a_3}$ could arise are $q = 8$ and $q = 12$, each of which has three nonsquares (and a single square), and $q = 11$, which has five squares and five nonsquares. Our main theorem provides results for the cases $q = 8$ and $q = 12$ showing that these asymmetries do in fact exist:

Theorem 1. *Assume GRH and LI. Let $\delta_{q;a_1,\dots,a_r}$ denote the (logarithmic) density of the set of positive real numbers x for which the system of inequalities (1–1) holds. Then*

$$\delta_{8;3,5,7} = \delta_{8;7,5,3} = 0.1928013 \pm 0.000001,$$
$$\delta_{8;3,7,5} = \delta_{8;5,7,3} = 0.1664263 \pm 0.000001,$$
$$\delta_{8;5,3,7} = \delta_{8;7,3,5} = 0.1407724 \pm 0.000001,$$

and

$$\delta_{12;5,7,11} = \delta_{12;11,7,5} = 0.1984521 \pm 0.000001,$$
$$\delta_{12;5,11,7} = \delta_{12;7,11,5} = 0.1215630 \pm 0.000001,$$
$$\delta_{12;7,5,11} = \delta_{12;11,5,7} = 0.1799849 \pm 0.000001,$$

where the indicated error bounds are rigorous.

The pairwise equalities among the δ's in Theorem 1 are not numerical coincidences, but are provably exact. In fact there are several situations in which we can establish symmetries of this sort. To state these results, we first need to define

$$c(q,a) = -1 + \#\{1 \le b \le q : b^2 \equiv a \bmod q\} \qquad (1–2)$$

for coprime integers a and q. Note that when q is an odd prime, $c(q,a)$ simply equals the Legendre symbol $\left(\frac{a}{q}\right)$. Note further that $c(q,a)$ can take only two possible values for a given q: certainly $c(q,a) = -1$ for every nonsquare a mod q, while $c(q,a) = c(q,1)$ for every square a mod q. We can interpret $c(q,1)$ as the ratio of the number of invertible nonsquares to the number of invertible squares mod q.

We may now state our results concerning symmetries:

Theorem 2. *Assume GRH and LI. Let $q, r \ge 2$ be integers and let $a_1,\dots,a_r$ be distinct reduced residue classes mod q.*

(a) *Letting a_j^{-1} denote the multiplicative inverse of a_j modulo q, we have $\delta_{q;a_1,\dots,a_r} = \delta_{q;a_1^{-1},\dots,a_r^{-1}}$.*
(b) *If b is a reduced residue class modulo q such that $c(q,a_j) = c(q,ba_j)$ for each $1 \le j \le r$, then $\delta_{q;a_1,\dots,a_r} = \delta_{q;ba_1,\dots,ba_r}$. In particular, this holds if b is a square modulo q.*
(c) *If the a_j are all squares modulo q and b is any reduced residue class modulo q, then $\delta_{q;a_1,\dots,a_r} = \delta_{q;ba_1,\dots,ba_r}$.*
(d) *If the a_j are either all squares modulo q or all nonsquares modulo q, then $\delta_{q;a_1,\dots,a_r} = \delta_{q;a_r,\dots,a_1}$.*
(e) *If b is a reduced residue class modulo q such that $c(q,a_j) \ne c(q,ba_j)$ for each $1 \le j \le r$, then $\delta_{q;a_1,\dots,a_r} = \delta_{q;ba_r,\dots,ba_1}$. In particular, this holds if q is an odd prime power or twice an odd prime power and b is any nonsquare modulo q.*

The pairwise equalities in Theorem 1 are special cases of part (d) of Theorem 2, which generalizes the previously mentioned result of Rubinstein and Sarnak that $\delta_{q;a,a'} = \delta_{q;a',a}$ if a and a' are either both squares or both nonsquares modulo q. Their other symmetry result, that $\delta_{q;a_1,a_2,a_3}$ is invariant under permutations of the a_j when there exists $\rho \not\equiv 1 \pmod q$ with $\rho^3 \equiv 1 \pmod q$ such that $a_2 \equiv a_1\rho \pmod q$ and $a_3 \equiv a_1\rho^2 \pmod q$, is also a consequence of Theorem 2 (specifically parts (b) and (d), the former applied with $b = \rho$ and $b = \rho^2$).

To complement Theorem 2, we can also establish several inequalities concerning the densities δ:

Theorem 3. *Assume GRH and LI. Let $q \ge 2$ be an integer, let N and N' be distinct (invertible) nonsquares mod q, and let S and S' be distinct (invertible) squares mod q. Then:*

(a) $\delta_{q;N,N',S} > \delta_{q;S,N',N}$;
(b) $\delta_{q;N,S,S'} > \delta_{q;S',S,N}$;
(c) $\delta_{q;N,S,N'} > \delta_{q;N',S,N}$ *if and only if* $\delta_{q;N,S} > \delta_{q;N',S}$;
(d) $\delta_{q;S,N,S'} > \delta_{q;S',N,S}$ *if and only if* $\delta_{q;S,N} > \delta_{q;S',N}$.

Parts (c) and (d) of Theorem 3 are further examples that the predisposition towards some orderings of $\{\pi(x;q,a_1),\dots,\pi(x;q,a_r)\}$ over others cannot be explained solely in terms of the Chebyshev bias that encourages nonsquares to run ahead of squares in the prime number race. (See also the discussion of "bias factors" in Section 6.)

The most general result in this paper is an explicit formula for an arbitrary density $\delta_{q;a_1,\dots,a_r}$. Because of the amount of notation involved, we have deferred the statement of this result (Theorem 4) to Section 2E. We have used this general formula to calculate the densities given in Theorem 1, and also a number of the $\delta_{q;a_1,\dots,a_r}$ in many interesting cases involving $q \le 12$ and $r \le 4$. For instance, we verify that for $q = 11$, which is the smallest interesting case not covered by Theorem 1, there are again asymmetries in races among triples of squares and among triples of nonsquares. In these additional computations we have not undertaken to rigorously bound the error terms; nevertheless we believe, from numerical considerations, that the results given in Section 4 are accurate to the number of decimal places indicated.

We shall assume GRH and LI throughout this paper. In Section 2 we provide our main analysis leading to Theorem 4, the general formula for $\delta_{q;a_1,\dots,a_r}$. The rigorous bounding of the error terms incurred during the calculation of the densities in Theorem 1 is carried out in Section 3. Details of the computations and the additional numerical results are collected together in Section 4. The proofs of Theorems 2 and 3 are given in Section 5, while in Section 6 we provide concluding remarks, noting some possible directions for further work.

2. ANALYTIC DETERMINATION OF THE DENSITIES $\delta_{q;a_1,\dots,a_r}$

The goal for this section of the paper is to derive Theorem 4 (see Section 2E), a general formula for the densities $\delta_{q;a_1,\dots,a_r}$. We begin by developing some notation and citing the relevant results of Rubinstein and Sarnak in Section 2A. In Section 2B we investigate the function $\hat{\rho}_{q;a_1,\dots,a_r}$ which will figure prominently in the arguments that follow, while in Section 2C we establish some facts about Cauchy principal values of multidimensional integrals; these sections are technical rather than conceptual in nature, and the reader may wish to examine these only briefly on the first reading. Because the general formula given in Theorem 4 and the arguments leading to it are somewhat involved, in Section 2D we first detail the derivation of this formula for the special cases $\delta_{8;a,b,c}$ and $\delta_{12;a,b,c}$ occurring in Theorem 1; the derivation of the formula in the general case is then carried out in Section 2E. We assume the hypotheses GRH and LI throughout.

2A. Notation and Background Results

We begin by establishing the notation necessary for discussing the results of Rubinstein and Sarnak. For any coprime integers q and a and any real number $x \ge 1$, define

$$E(x;q,a) = \frac{\log x}{\sqrt{x}}\big(\varphi(q)\pi(x;q,a) - \pi(x)\big), \qquad (2\text{–}1)$$

so that $E(x;q,a)$ is an error term for the number of primes congruent to $a \bmod q$, normalized so as to vary roughly boundedly as x varies. Since the inequalities $\pi(x;q,a_1) > \cdots > \pi(x;q,a_r)$ hold if and only if $E(x;q,a_1) > \cdots > E(x;q,a_r)$, we wish to study how often the vector

$$E_{q;a_1,\dots,a_r}(x) = \big(E(x;q,a_1),\dots,E(x;q,a_r)\big) \qquad (2\text{–}2)$$

lies in the region $\{(x_1,\dots,x_r) \in \mathbb{R}^r : x_1 > \cdots > x_r\}$. Notice that if $r = \varphi(q)$ then the a_j form a complete set of reduced residues mod q, in which case we see from equation (2–1) that

$$E(x;q,a_1) + \cdots + E(x;q,a_r) = -\frac{\log x}{\sqrt{x}}\varphi(q)\omega(q), \qquad (2\text{–}3)$$

where $\omega(q)$ denotes the number of distinct prime factors of q.

Rubinstein and Sarnak showed, assuming GRH, that the function $E_{q;a_1,\dots,a_r}(x)$ has a limiting distribution $\mu_{q;a_1,\dots,a_r}$, in the sense that

$$\lim_{X\to\infty} \frac{1}{\log X}\int_2^X f(E_{q;a_1,\dots,a_r}(x))\,\frac{dx}{x} = \int\!\cdots\!\int_{\mathbb{R}^r} f(x_1,\dots,x_r)\,d\mu_{q;a_1,\dots,a_r} \qquad (2\text{–}4)$$

for all bounded, continuous functions f on $\mathbb{R}^r$. Under the further assumption of LI, they showed that the distribution $\mu_{q;a_1,\dots,a_r}$ is absolutely continuous with respect to the ordinary Lebesgue measure on $\mathbb{R}^r$. (The exception is the case $r = \varphi(q)$, when equation (2–3) implies that the distribution $\mu_{q;a_1,\dots,a_r}$ is supported on the hyperplane $x_1+\cdots+x_r = 0$; in this case, $\mu_{q;a_1,\dots,a_r}$ is absolutely continuous with respect to Lebesgue measure on this hyperplane.) Consequently, the equation (2–4) holds when f is the characteristic function of any reasonable subset of $\mathbb{R}^r$ (specifically, a measurable subset whose boundary has Lebesgue measure zero in $\mathbb{R}^r$). In particular, it follows from the definition of $\delta_{q;a_1,\dots,a_r}$ that

$$\begin{aligned}\delta_{q;a_1,\dots,a_r} &= \delta\big(\{x \in \mathbb{R} : \pi(x,q,a_1) > \cdots > \pi(x,q,a_r)\}\big)\\ &= \mu_{q;a_1,\dots,a_r}\big(\{x \in \mathbb{R}^r : x_1 > \cdots > x_r\}\big)\\ &= \int\!\cdots\!\int_{x_1>\cdots>x_r} d\mu_{q;a_1,\dots,a_r}\,. \end{aligned} \qquad (2\text{–}5)$$

Another consequence of the absolute continuity of $\mu_{q;a_1,\dots,a_r}$ is that the set of positive real numbers x for which $\pi(x;q,a) = \pi(x;q,a')$ has density zero when a and a' are distinct reduced residues; indeed this is even true of the larger set

$$\{x : |\pi(x;q,a) - \pi(x;q,a')| < \Phi(x)\}$$

for any function Φ such that

$$\lim_{x\to\infty} \frac{\Phi(x)}{\sqrt{x}/\log x} = 0.$$

Next we develop the notation needed to write down Rubinstein and Sarnak's seminal formula for the Fourier transform $\hat{\mu}_{q;a_1,\dots,a_r}$ of the distribution $\mu_{q;a_1,\dots,a_r}$. In this paper we use the normalization

$$\hat{f}(\xi_1,\dots,\xi_n) = \int\cdots\int e^{-i(\xi_1x_1+\cdots+\xi_nx_n)} \times f(x_1,\dots,x_n)\,dx_1\dots dx_n \qquad (2\text{–}6)$$

for the Fourier transform of an integrable function f on $\mathbb{R}^n$, so that the Fourier inversion formula (assuming that $\hat{f}$ is itself integrable) is

$$f(x_1,\dots,x_n) = (2\pi)^{-n}\int\cdots\int e^{i(\xi_1x_1+\cdots+\xi_nx_n)} \times \hat{f}(\xi_1,\dots,\xi_n)\,d\xi_1\dots d\xi_n\,.$$

Likewise we write

$$\hat{\mu}(\xi_1,\dots,\xi_n) = \int\cdots\int e^{-i(\xi_1x_1+\cdots+\xi_nx_n)}\,d\mu$$

for the Fourier transform of a finite measure μ on $\mathbb{R}^r$, so that the Fourier inversion formula (assuming that $\hat{\mu}$ is integrable with respect to Lebesgue measure) is

$$d\mu = (2\pi)^{-n}\biggl(\int\cdots\int e^{i(\xi_1x_1+\cdots+\xi_nx_n)} \times \hat{\mu}(\xi_1,\dots,\xi_n)\,d\xi_1\dots d\xi_n\biggr)\,dx_1\dots dx_n. \qquad (2\text{–}7)$$

To write down the Fourier transform $\hat{\mu}_{q;a_1,\dots,a_r}$, we recall the standard Bessel function of order zero,

$$J_0(z) = \sum_{m=0}^{\infty} \frac{(-1)^m(z/2)^{2m}}{(m!)^2} = 1 - \frac{z^2}{4} + \frac{z^4}{64} - \cdots, \qquad (2\text{–}8)$$

and then set

$$F(z,\chi) = \prod_{\substack{\gamma>0 \\ L(\frac12+i\gamma,\chi)=0}} J_0(\alpha_\gamma z) \qquad (2\text{–}9)$$

in terms of the Dirichlet L-function $L(s,\chi)$ corresponding to the Dirichlet character χ, where we have defined

$$\alpha_\gamma = \frac{2}{\sqrt{\frac14+\gamma^2}}\,. \qquad (2\text{–}10)$$

(Since we are assuming GRH, the product in equation (2–9) is indexed by all the nontrivial zeros of $L(s,\chi)$ in the upper half-plane.) For later use in numerical approximations of $F(z,\chi)$ we also define the truncated version

$$F_T(z,\chi) = \biggl(\prod_{\substack{0<\gamma<T \\ L(\frac12+i\gamma,\chi)=0}} J_0(\alpha_\gamma z)\biggr)(1+b_1z^2) \qquad (2\text{–}11)$$

for any positive real number T, where

$$b_1 = b_1(T,\chi) = -\sum_{\gamma\ge T} \frac{1}{\frac14+\gamma^2}\,. \qquad (2\text{–}12)$$

The polynomial factor in the definition (2–11) of F_T is motivated by the fact that, in view of the power series expansion (2–8) of J_0, b_1 is the coefficient of z^2 in the power series expansion of $\prod_{\gamma>T} J_0(\alpha_\gamma z)$.

With this notation in place, we can now give Rubinstein and Sarnak's formula [1994, equation 1.2] for the Fourier transform $\hat{\mu}_{q;a_1,\dots,a_r}$ of the distribution $\mu_{q;a_1,\dots,a_r}$. They showed, assuming GRH and LI, that

$$\hat{\mu}_{q;a_1,\dots,a_r}(\xi_1,\dots,\xi_r) = \exp\biggl(i\sum_{j=1}^r c(q,a_j)\xi_j\biggr) \times \prod_{\substack{\chi \bmod q \\ \chi\ne\chi_0}} F\biggl(\biggl|\sum_{j=1}^r \chi(a_j)\xi_j\biggr|,\chi\biggr), \qquad (2\text{–}13)$$

where $c(q,a)$ was defined in equation (1–2). This result will be used extensively in the sequel.

Since $J_0(0)=1$ we clearly have

$$F(0,\chi) = F_T(0,\chi) = 1$$

for any character χ. It is known (see for instance the arguments in [Davenport 1980, Chapters 15–16]) that for a fixed character χ, the number of zeros of $L(s,\chi)$ with imaginary part between 0 and T has order of magnitude $T\log T$. From this it can be shown that the product (2–9) defining $F(z,\chi)$ converges uniformly on bounded subsets of the complex plane, and hence F is an entire function. For later use we will need bounds for the decay rate of $F(x,\chi)$

and its derivatives $F^{(N)}(x,\chi)$ on the real axis; this is the subject of the following lemma.

Lemma 2.1. *Given a modulus $q \ge 2$ and a nonnegative integer N, there exist positive constants β_1 and β_2 such that*

$$|F^{(N)}(x,\chi)| \le \beta_1 e^{-\beta_2|x|}$$

for all real numbers x.

Warning. In the next three sections, the constants β_1 and β_2 will not necessarily have the same values at different occurrences; each statement should be interpreted as holding for some suitable positive values of β_1 and β_2.

Proof. In this proof we will use the symbol γ, with or without subscript, exclusively to denote a positive imaginary part of a nontrivial zero of $L(s,\chi)$. We also use Γ to denote an ordered N-tuple $(\gamma_1,\dots,\gamma_N)$, and we let $m_\Gamma(\gamma)$ denote the number (possibly zero) of coordinates of Γ that equal γ. When convenient we can also assume that $x>1$, since F is an even, smooth function. From the definition of $F(z,\chi)$ in equation (2–9), an N-fold application of the product rule gives us the expression

$$\begin{aligned} F^{(N)}(x,\chi) &= \sum_{\Gamma=(\gamma_1,\dots,\gamma_N)} \alpha_{\gamma_1}\dots\alpha_{\gamma_N}\prod_\gamma J_0^{(m_\Gamma(\gamma))}(\alpha_\gamma x) \\ &= \sum_\Gamma \Phi(x,\Gamma)F(x,\chi,\Gamma) \qquad (2–14)\end{aligned}$$

for the N-th derivative of $F(x,\chi)$, where we have set

$$\Phi(x,\Gamma) = \prod_{\gamma\in\Gamma}\alpha_\gamma^{m_\Gamma(\gamma)}J_0^{(m_\Gamma(\gamma))}(\alpha_\gamma x) \qquad (2–15)$$

and

$$F(x,\chi,\Gamma) = \prod_{\gamma\notin\Gamma} J_0(\alpha_\gamma x).$$

We can show that $F(x,\chi)$ decays rapidly on the real axis by using the standard bound [Rubinstein and Sarnak 1994, equation 4.5]

$$|J_0(x)| \le \min\left\{1, \sqrt{\frac{2}{\pi|x|}}\right\}$$

for the Bessel function on the real axis. This bound implies that

$$\begin{aligned} |F(x,\chi)| &\le \prod_\gamma \min\left\{1,\sqrt{\frac{2}{\pi|\alpha_\gamma x|}}\right\} \\ &\le \prod_{j=1}^J \sqrt{\frac{2}{\pi|\alpha_{\gamma_j}x|}} \\ &= (\pi|x|)^{-J/2}\prod_{j=1}^J\left(\tfrac14+\gamma_j^2\right)^{1/4} \qquad (2–16)\end{aligned}$$

for any positive integer J, where the γ_j have been indexed in increasing order. Choose $J=J(x)$ to be the number of zeros of $L(s,\chi)$ up to height $x/2$. For any $0<\gamma\le x/2$, it is easily verified that the factor $(\pi|x|)^{-1/2}(\frac14+\gamma^2)^{1/4}$ does not exceed $\frac12$. Therefore the upper bound (2–16) implies that $|F(x,\chi)|\le 2^{-J}$. Since the order of magnitude of J is $x\log x$, this argument shows that as x tends to infinity, $|F(x,\chi)|$ decreases at least as fast as a function of the form $c^{x\log x}$ for some constant c depending on χ.

The same conclusion holds for $F(x,\chi,\Gamma)$, since removing the indices j in equation (2–16) for which $\gamma_j\in\Gamma$ changes J by at most N and thus does not affect the order of magnitude of J. Certainly then there exist positive constants β_1 and β_2 (depending only on N and χ) such that $|F(x,\chi,\Gamma)|\le\beta_1e^{-\beta_2|x|}$ for all real numbers x. Since this implies from equation (2–14) that

$$|F(x,\chi)| \le \beta_1 e^{-\beta_2|x|}\sum_\Gamma\big|\Phi(x,\Gamma)\big|, \qquad (2–17)$$

the lemma will be established (possibly with different values of β_1 and β_2) if we can show that this last sum is bounded by some polynomial function of $|x|$.

To this end, we employ the crude bounds $|J_0'(t)|\le t/2$ and $|J_0^{(n)}(t)|\le 1$ for the derivatives of the Bessel function, which follow easily from the integral representation

$$J_0(t) = \frac{2}{\pi}\int_0^{\pi/2}\cos(t\sin\theta)\,d\theta.$$

Again supposing that $x>1$, the definition (2–15) of $\Phi(x,\Gamma)$ leads to the bound

$$|\Phi(x,\Gamma)| \le \Bigg(\prod_{\substack{\gamma\in\Gamma\\ m_\Gamma(\gamma)=1}}\alpha_\gamma^2|x|\Bigg)\Bigg(\prod_{\substack{\gamma\in\Gamma\\ m_\Gamma(\gamma)>1}}\alpha_\gamma^{m_\Gamma(\gamma)}\Bigg).$$

It follows that

$$\sum_\Gamma |\Phi(x,\Gamma)| \le |x|^N \sum_\Gamma \alpha_\gamma^{\max\{m_\Gamma(\gamma),2\}} \le |x|^N N! \prod_\gamma (1+\alpha_\gamma^2+\alpha_\gamma^3+\cdots+\alpha_\gamma^N).$$

Since the j-th constant α_γ has order of magnitude

$$\frac{1}{\gamma_j} \sim \frac{\log j}{j},$$

this last product converges to some constant depending only on χ. Combining this bound with the inequality (2–17) establishes the lemma. □

Of course it also follows from the first line of equation (2–16) that $|F(x,\chi)|$ is bounded above by 1 on the real axis.

In Sections 3A and 3E we will need to make use of the fact that $\mu_{q;a_1,\dots,a_r}$ can also be thought of as the joint distribution of a certain set of r real-valued random variables, and it is convenient to exhibit these random variables explicitly at this time. For given values of q, r, and $a_1,\dots,a_r$, define the vector

$$b_{q;a_1,\dots,a_r} = -\big(c(q,a_1),\dots,c(q,a_r)\big).$$

Next, for any character $\chi \bmod q$, define both the vector

$$v_{q;a_1,\dots,a_r}(\chi) = \big(\chi(a_1),\dots,\chi(a_r)\big)$$

and the random variable

$$X(\chi) = \sum_{\substack{\gamma>0 \\ L(\frac12+i\gamma,\chi)=0}} \alpha_\gamma \sin(2\pi U_\gamma), \tag{2–18}$$

where the α_γ are as in (2–10) and the U_γ are independent random variables uniformly distributed on $[0,1]$. Note that by the hypothesis LI, the γ's corresponding to different L-functions are distinct, so that a given U_γ appears in the definition of only one of the $X(\chi)$; consequently the random variables $\{X(\chi)\}$ are mutually independent. Then Rubinstein and Sarnak showed that the distribution $\mu_{q;a_1,\dots,a_r}$ is in fact the same as the probability measure corresponding to the random vector

$$b_{q;a_1,\dots,a_r} + \sum_{\substack{\chi \bmod q \\ \chi\ne\chi_0}} X(\chi) v_{q;a_1,\dots,a_r}(\chi). \tag{2–19}$$

2B. The Function $\hat\rho_{q;a_1,\dots,a_r}$

In this section we introduce the function $\hat\rho_{q;a_1,\dots,a_r} : \mathbb{R}^{r-1} \to \mathbb{C}$, which we define by the formula

$$\hat\rho_{q;a_1,\dots,a_r}(\eta_1,\dots,\eta_{r-1}) = \hat\mu_{q;a_1,\dots,a_r}(\eta_1,\eta_2-\eta_1,\dots,\eta_{r-1}-\eta_{r-2},-\eta_{r-1}), \tag{2–20}$$

so that

$$\hat\rho_{q;a_1,\dots,a_r}(\eta_1,\dots,\eta_{r-1}) = \exp\bigg(\sum_{j=1}^{r-1}\big(c(q,a_j)-c(q,a_{j+1})\big)\eta_j\bigg) \times \prod_{\substack{\chi \bmod q \\ \chi\ne\chi_0}} F\bigg(\bigg|\sum_{j=1}^{r-1}\big(\chi(a_j)-\chi(a_{j+1})\big)\eta_j\bigg|,\chi\bigg) \tag{2–21}$$

from the formula (2–13) for $\hat\mu_{q;a_1,\dots,a_r}$. We will see in Sections 2D and 2E that $\hat\rho_{q;a_1,\dots,a_r}$ is the Fourier transform of a certain measure $\rho_{q;a_1,\dots,a_r}$ on $\mathbb{R}^{r-1}$ associated with $\mu_{q;a_1,\dots,a_r}$. We remark that in the special case where the a_j are all squares or all nonsquares, we have $c(q,a_1)=\cdots=c(q,a_r)$ and so the exponential term in the formula (2–21) is identically 1, so that $\hat\rho_{q;a_1,\dots,a_r}$ is real-valued and symmetric with respect to reflection through the origin.

The function $\hat\rho_{q;a_1,\dots,a_r}$ will feature significantly in the remainder of this paper, and it will be important to establish some of its smoothness and decay properties. To avoid frequent repetition of the same properties, we shall say that a function f on $\mathbb{R}^n$ is *well-behaved* if it has continuous derivatives of all orders and if there exist positive constants β_1 and β_2 such that, for every subset $\{j_1,\dots,j_k\}$ of $\{1,\dots,n\}$, the mixed partial derivative $\frac{\partial^k f}{\partial x_{j_1}\dots\partial x_{j_k}}$ satisfies the inequality

$$\bigg|\frac{\partial^k f}{\partial x_{j_1}\dots\partial x_{j_k}}(x_1,\dots,x_n)\bigg| \le \beta_1 e^{-\beta_2\|x\|}, \tag{2–22}$$

where $\|x\| = \|(x_1,\dots,x_n)\| = \sqrt{x_1^2+\cdots+x_n^2}$ is the Euclidean norm of x. This criterion must also be satisfied for the empty subset of $\{1,\dots,n\}$, so that the actual values of f must also be bounded by the right-hand side of (2–22). Certainly any well-behaved function is integrable as well. We remark that all of the functions shown to be well-behaved below in fact satisfy an inequality analogous to (2–22) for partial derivatives of all orders; however our proof of Lemma 2.4 below requires this assumption only on the mixed linear partial derivatives.

It is easily seen that finite sums and products of well-behaved functions are again well-behaved. If f and g are well-behaved functions on $\mathbb{R}^m$ and $\mathbb{R}^n$, respectively, then fg is a well-behaved function on $\mathbb{R}^{m+n}$; conversely, the restriction of a well-behaved function on $\mathbb{R}^n$ to any subspace defined by setting certain variables equal to zero is a well-behaved function on that subspace. Also, if $L : \mathbb{R}^m \to \mathbb{R}^n$ is an injective linear map and f is a well-behaved function on $\mathbb{R}^n$, then the composite function $f \circ L$ is a well-behaved function on $\mathbb{R}^m$: the partial derivatives of $f \circ L$ will just be linear combinations of the partial derivatives of f, and the fact that L is injective means that $\|L(x)\|$ is bounded below by a constant multiple of $\|x\|$, so that the estimate (2–22) for f on $\mathbb{R}^n$ can be converted to a similar estimate for $f \circ L$ on $\mathbb{R}^m$.

The following two lemmas establish the important fact that the functions $\hat{\rho}_{q;a_1,\ldots,a_r}$ are well-behaved.

Lemma 2.2. *For every subset* $\{j_1, \ldots, j_k\} \subset \{1, \ldots, r\}$,

$$\frac{\partial^k}{\partial x_{j_1} \ldots \partial x_{j_k}} F\left(\left| \sum_{j=1}^{r} \chi(a_j) x_j \right|, \chi \right) \le \beta_1 \exp\left(-\beta_2 \left| \sum_{j=1}^{r} \chi(a_j) x_j \right| \right) \tag{2–23}$$

for some positive constants β_1 *and* β_2.

Proof. Define

$$H(x_1, \ldots, x_r; \chi) = F\left(\left| \sum_{j=1}^{r} \chi(a_j) x_j \right|, \chi \right).$$

The argument of F on the right-hand side of this definition involves a modulus and hence implicitly a square root, which could potentially cause discontinuities in the derivatives of H when this argument equals zero; however, the Bessel function J_0 is even, whence the function $F(x, \chi)$ involves only even powers of x in its power series expansion about the origin. Consequently, H has continuous derivatives of all orders. Note also that it suffices to establish the upper bound (2–23) when

$$\left| \sum_{j=1}^{r} \chi(a_j) x_j \right| > 1,$$

since the bound on the complementary set follows immediately from continuity, with some value of β_1.

If we write $\tilde{F}(x, \chi) = F(\sqrt{|x|}, \chi)$, then it is easy to check by induction that the n-th derivative of $\tilde{F}$ equals

$$\tilde{F}^{(n)}(x, \chi) = \sum_{k=1}^{n} \alpha_{n,k} F^{(k)}\big(\sqrt{|x|}, \chi\big) |x|^{-n+j/2}$$

for some constants $\alpha_{n,k}$. In particular, when $|x| > 1$ we see from Lemma 2.1 that

$$|\tilde{F}^{(n)}(x, \chi)| \le \beta_1 e^{-\beta_2 \sqrt{|x|}} \tag{2–24}$$

for some positive constants β_1 and β_2.

In this notation we have

$$H(x_1, \ldots, x_r; \chi) = \tilde{F}\left(\left(\operatorname{Re} \sum_{j=1}^{r} \chi(a_j) x_j \right)^2 + \left(\operatorname{Im} \sum_{j=1}^{r} \chi(a_j) x_j \right)^2, \chi \right).$$

Suppressing the details, we note that the mixed partial derivative

$$\frac{\partial^k H}{\partial x_{j_1} \ldots \partial x_{j_k}}$$

can be computed using the product rule as a combination of three types of expressions: derivatives of $\tilde{F}$ evaluated at

$$\left| \sum_{j=1}^{r} \chi(a_j) x_j \right|^2,$$

linear factors of the form $2\operatorname{Re}(\bar{\chi}(a_k) \sum_{j=1}^{r} \chi(a_j) x_j)$, and constants of the form $2\operatorname{Re}(\bar{\chi}(a_k)\chi(a_{k'}))$. From equation (2–24), the expressions of the first type can be bounded above by $\beta_1 \exp(-\beta_2 |\sum_{j=1}^{r} \chi(a_j) x_j|)$, while the expressions of the other types grow only as fast as a polynomial in $|\sum_{j=1}^{r} \chi(a_j) x_j|$. This establishes the lemma for suitable positive values of β_1 and β_2. □

Lemma 2.3. *The function* $\hat{\rho}_{q;a_1,\ldots,a_r}$ *is well-behaved for any integers* $q, r \ge 2$ *and any distinct reduced residues* $\{a_1, \ldots, a_r\}$.

Proof. From (2–13), the function $\hat{\mu}_{q;a_1,\ldots,a_r}$ certainly has continuous derivatives of all orders (see the proof of Lemma 2.2), and thus the same is true of $\hat{\rho}_{q;a_1,\ldots,a_r}$. We begin by examining the behavior of the mixed partial derivatives of the function $\hat{\mu}_{q;a_1,\ldots,a_r}$. Let $S = \{j_1, \ldots, j_k\}$ be a subset of indices from the set $\{1, \ldots, r\}$, and let $\frac{\partial^k}{\partial x_S}$ denote the result of taking the partial x_j-derivatives for every j in S. The product rule applied to the formula (2–13) for $\hat\mu_{q;a_1,\dots,a_r}$ yields

$$\frac{\partial^k}{\partial x_S}\hat\mu_{q;a_1,\dots,a_r}(\xi) = \sum_{S_0,\{S_\chi\}} \left(\frac{\partial^k}{\partial x_{S_0}} \exp\left(i\sum_{j=1}^r c(q,a_j)\xi_j \right) \right.$$
$$\left. \times \prod_{\substack{\chi \bmod q\\ \chi\ne\chi_0}} \frac{\partial^k}{\partial x_{S_\chi}} F\left(\left| \sum_{j=1}^r \chi(a_j)\xi_j \right|, \chi \right) \right), \qquad (2\text{–}25)$$

where the outer summation is taken over the finitely many partitions of the index set S into

$$S_0 \cup \left(\textstyle\bigcup_{\chi\ne\chi_0} S_\chi \right).$$

Each mixed partial derivative appearing in the exponential term is bounded, while from Lemma 2.2 each mixed partial derivative of $F(|\sum_{j=1}^r \chi(a_j)\xi_j|,\chi)$ is exponentially decaying as a function of its argument. We conclude from equation (2–25) that there exist positive constants β_1 and β_2 such that

$$\left| \frac{\partial^k}{\partial x_S}\hat\mu_{q;a_1,\dots,a_r}(\xi) \right| \le \beta_1 \prod_{\substack{\chi \bmod q\\ \chi\ne\chi_0}} \exp\left(-\beta_2 \left| \sum_{j=1}^r \chi(a_r)\xi_j \right| \right)$$
$$= \beta_1 e^{-\beta_2 Q(\xi)^{1/2}}, \qquad (2\text{–}26)$$

where we have defined

$$Q(\xi) = Q_{q;a_1,\dots,a_r}(\xi) = \left(\sum_{\substack{\chi \bmod q\\ \chi\ne\chi_0}} \left| \sum_{j=1}^r \chi(a_j)\xi_j \right| \right)^2.$$

We thus seek a lower bound on $Q(\xi)$.

We may certainly write

$$Q(\xi) \ge \sum_{\substack{\chi \bmod q\\ \chi\ne\chi_0}} \left| \sum_{j=1}^r \chi(a_j)\xi_j \right|^2$$
$$= \sum_{\chi \bmod q} \left| \sum_{j=1}^r \chi(a_j)\xi_j \right|^2 - \left(\sum_{j=1}^r \xi_j \right)^2.$$

Now

$$\sum_{\chi \bmod q} \left| \sum_{j=1}^r \chi(a_j)\xi_j \right|^2 = \sum_{i=1}^r \sum_{j=1}^r \xi_i\xi_j \sum_{\chi \bmod q} \chi(a_i)\overline{\chi(a_j)}$$
$$= \varphi(q) \sum_{j=1}^r \xi_j^2$$

by the orthogonality of the characters χ. Therefore

$$Q(\xi) \ge \varphi(q) \sum_{j=1}^r \xi_j^2 - \left(\sum_{j=1}^r \xi_j \right)^2. \qquad (2\text{–}27)$$

We assume for now that r is strictly less than $\varphi(q)$, commenting at the end of the proof on the slight differences in the case $r = \varphi(q)$. The quadratic form on the right-hand side of the inequality (2–27) turns out to be positive definite when $r < \varphi(q)$, and so we can write

$$Q(\xi) \ge \varphi(q)\lambda_r \|\xi\|^2, \qquad (2\text{–}28)$$

where λ_r is the smallest eigenvalue of that quadratic form. From the inequalities (2–26) and (2–28), it follows that

$$\left| \frac{\partial^k}{\partial x_S}\hat\mu_{q;a_1,\dots,a_r}(\xi) \right| \le \beta_1 e^{-\beta_2\|\xi\|}$$

for some different positive constants β_1 and β_2. Since the index set $S \subset \{1,\dots,r\}$ was arbitrary, this shows that the function $\hat\mu_{q;a_1,\dots,a_r}$ is well-behaved.

Furthermore, from its definition (2–20) the function $\hat\rho_{q;a_1,\dots,a_r}$ is simply the composition of $\hat\mu_{q;a_1,\dots,a_r}$ with the injective linear transformation

$$(\eta_1,\dots,\eta_{r-1}) \mapsto (\eta_1, \eta_2-\eta_1, \dots, \eta_{r-1}-\eta_{r-2}, -\eta_{r-1})$$

from $\mathbb{R}^{r-1}$ to $\mathbb{R}^r$. As mentioned before, this implies that $\hat\rho_{q;a_1,\dots,a_r}$ is itself a well-behaved function.

When $r = \varphi(q)$, the function $\hat\mu_{q;a_1,\dots,a_r}$ is invariant under translation in the direction of the vector $(1,\dots,1)$, and so it is not well-behaved even though it has the required decay properties on the hyperplane orthogonal to $(1,\dots,1)$ (one can check that the quadratic form on the right-hand side of the inequality (2–27) is positive semi-definite when $r = \varphi(q)$, with its zero set being the multiples of the $(1,\dots,1)$ vector). However, the image of the linear transformation

$$(\eta_1,\dots,\eta_{r-1}) \mapsto (\eta_1, \eta_2-\eta_1, \dots, \eta_{r-1}-\eta_{r-2}, -\eta_{r-1})$$

lies within this hyperplane, so we can still deduce that $\hat\rho_{q;a_1,\dots,a_r}$ is well-behaved even when $r = \varphi(q)$. This establishes the lemma. □

Of course we also have the trivial bound

$$|\hat\rho_{q;a_1,\dots,a_r}| \le 1.$$

Lemma 2.3 implies in particular that $\hat{\rho}_{q;a_1,\dots,a_r}$ is integrable, and consequently the Fourier inversion formula (2–7) is valid for $\rho_{q;a_1,\dots,a_r}$, becoming

$$d\rho_{q;a_1,\dots,a_r} = (2\pi)^{-r}\biggl(\int\cdots\int e^{i(\xi_1x_1+\cdots+\xi_rx_r)} \times \hat{\rho}_{q;a_1,\dots,a_r}(\xi_1,\dots,\xi_r)\,d\xi_1\dots d\xi_r\biggr)\,dx_1\dots dx_r. \tag{2–29}$$

2C. Multidimensional Cauchy Principal Values

In one dimension, the Cauchy principal value

$$\text{P.V.}\int_{-\infty}^{\infty}\frac{f(x)}{x}\,dx = \lim_{\varepsilon\to 0}\int_{|x|>\varepsilon}\frac{f(x)}{x}\,dx$$

is a familiar object. For our purposes it will be necessary to make use of the multidimensional analogue

$$\text{P.V.}\int\cdots\int\frac{f(x_1,\dots,x_n)}{x_1\dots x_n}\,dx_1\dots dx_n = \lim_{\varepsilon\to 0}\int\cdots\int_{\min\{|x_1|,\dots,|x_n|\}>\varepsilon}\frac{f(x_1,\dots,x_n)}{x_1\dots x_n}\,dx_1\dots dx_n; \tag{2–30}$$

in particular, we would like to know that this limit exists. The purpose of this section is to establish the existence of these multidimensional Cauchy principal values for well-behaved functions, a class which by Lemma 2.3 includes the functions $\hat{\rho}_{q;a_1,\dots,a_r}$ discussed in the previous section. We remark that while the lemmas in this section could certainly be obtained under somewhat weaker hypotheses, they suffice for our purposes as stated.

Lemma 2.4. *Let f be a well-behaved function on $\mathbb{R}^n$ that vanishes whenever any of the first k coordinates $x_1, \dots, x_k$ equals zero. Then the function $f(x_1,\dots,x_n)/x_1\dots x_k$ extends across the coordinate hyperplanes to a continuous integrable function satisfying the upper bound*

$$\left|\frac{f(x_1,\dots,x_n)}{x_1\dots x_k}\right| \le \beta_1 e^{-\beta_2\|x\|} \tag{2–31}$$

for some positive constants β_1 and β_2.

Although this lemma holds in one dimension without any assumptions on the derivatives of f, already in $\mathbb{R}^2$ one can construct an exponentially decaying, smooth (even real-analytic) function $f(x,y)$ that satisfies $f(0,y) = 0$ for all y but for which $f(x,y)/x$ is not integrable.

Proof. That $f(x)/x_1\dots x_k$ extends across the coordinate hyperplanes to a continuous function follows from the fact that f has continuous derivatives of all orders; therefore only the upper bound (2–31) remains to be proved, since integrability is a consequence of this bound. Furthermore, by continuity it suffices to establish this upper bound when none of the variables equals zero. Also, if all of the $|x_j|$ are bounded by 1 then the function $f(x)/x_1\dots x_k$ is uniformly bounded; therefore we may assume (after inflating the constant β_1 if necessary) that there exists an x_j with $|x_j| > 1$.

Permuting the first k variables if necessary, we can choose an integer $1 \le m \le k$ such that

$$0 < |x_1|,\dots,|x_m| \le 1 \quad\text{and}\quad |x_{m+1}|,\dots,|x_k| > 1.$$

Since f vanishes when x_1 equals zero, there exists a number t_1 with $|t_1| \le |x_1|$ such that

$$f(x_1,\dots,x_n) = f(x_1,\dots,x_n) - f(0,x_2,\dots,x_n) = x_1\frac{\partial f}{\partial x_1}(t_1,x_2,\dots,x_n)$$

by the mean value theorem in the variable x_1. Similarly, f vanishes whenever x_2 equals zero, so in particular $\partial f/\partial x_1$ equals zero when $x_2 = 0$. Therefore, there exists a number t_2 with $|t_2| \le |x_2|$ such that

$$\frac{\partial f}{\partial x_1}(t_1,x_2,\dots,x_n) = \frac{\partial f}{\partial x_1}(t_1,x_2,\dots,x_n) - \frac{\partial f}{\partial x_1}(t_1,0,x_3,\dots,x_n) = x_2\frac{\partial^2 f}{\partial x_1\partial x_2}(t_1,t_2,x_3,\dots,x_n)$$

by the mean value theorem in the variable x_2. Continuing in this way, we find numbers t_i with $|t_i| \le |x_i|$ for each $1 \le i \le m$ such that

$$f(x_1,\dots,x_n) = x_1\dots x_m\frac{\partial^m f}{\partial x_1\dots\partial x_m}(t_1,\dots,t_m,x_{m+1},\dots,x_n).$$

It follows immediately that

$$\left|\frac{f(x)}{x_1\dots x_k}\right| \le \left|\frac{\partial^m f}{\partial x_1\dots\partial x_m}(t_1,\dots,t_m,x_{m+1},\dots,x_n)\right| \tag{2–32}$$

since $|x_{m+1}|,\dots,|x_k| > 1$.

Since f is well-behaved, there exist positive constants β_1 and β_2 such that

$$\left|\frac{\partial^m f}{\partial x_1 \dots \partial x_m}(t_1, \dots, t_m, x_{m+1}, \dots, x_n)\right| \le \beta_1 \exp\left(-\beta_2\sqrt{t_1^2 + \dots + t_m^2 + x_{m+1}^2 + \dots + x_n^2}\right). \tag{2–33}$$

But notice that

$$t_1^2 + \dots + t_m^2 + x_{m+1}^2 + \dots + x_n^2 \ge \sum_{\substack{1\le j\le n \\ |x_j|>1}} x_j^2 \ge \frac{\#\{1 \le j \le n : |x_j| > 1\}}{n} \sum_{j=1}^n x_j^2.$$

Since we are working under the assumption that at least one of the $|x_j|$ exceeds 1, we can use this fact in the inequality (2–33) to see that

$$\left|\frac{\partial^m f}{\partial x_1 \dots \partial x_m}(t_1, \dots, t_m, x_{m+1}, \dots, x_n)\right| \le \beta_1 e^{-\beta_2\|x\|/\sqrt{n}}.$$

Combining this bound with the inequality (2–32), this establishes the lemma (upon replacing $\beta_2/\sqrt{n}$ by β_2). □

For the proof of the next lemma, as well as for the formulation of the general formula for $\delta_{q;a_1,\dots,a_r}$ (Theorem 4 in Section 2E), we require the following notation: for a function f on $\mathbb{R}^n$ and a subset B of $\{1, \dots, n\}$, define

$$f(B) = f(B)(\{x_j : j \in B\}) = f(\theta_1, \dots, \theta_n), \tag{2–34}$$

where $\theta_j = x_j$ if $j \in B$, and $\theta_j = 0$ otherwise. For example, if $n = 6$ and $B = \{2, 4, 5\}$ then $f(B)$ is a function of the three variables x_2, x_4, and x_5, namely $f(B) = f(0, x_2, 0, x_4, x_5, 0)$; in general $f(B)$ will be a function on the appropriate $|B|$-dimensional subspace of $\mathbb{R}^n$, where $|B|$ denotes the cardinality of B. In the case $B = \varnothing$ we simply have $f(B) = f(0, \dots, 0)$.

Lemma 2.5. *If f is a well-behaved function on $\mathbb{R}^n$, then*

$$\text{P.V.} \int \cdots \int \frac{f(x_1, \dots, x_n)}{x_1 \dots x_n}\, dx_1 \dots dx_n$$

is well-defined; i.e., the limit in equation (2–30) exists.

Proof. Let $g_1(x)$ be an even, well-behaved function on $\mathbb{R}^1$ with $g_1(0) = 1$ (for instance, we might have in mind $g_1(x) = e^{-x^2}$), and let

$$g(x_1, \dots, x_n) = g_1(x_1) \dots g_1(x_n).$$

Define an operator G on well-behaved functions f by

$$G(f) = G(f)(x_1, \dots, x_n) = \sum_{B\subset\{1,\dots,n\}} (-1)^{n-|B|} f(B) g(\bar{B}) \tag{2–35}$$

in the notation of equation (2–34), where $\bar{B}$ denotes the complement $\{1, \dots, n\} \setminus B$ of B. Since f and g are well-behaved functions, the same is true of $G(f)$.

Consider the term in (2–35) corresponding to some particular proper subset B of $\{1, \dots, n\}$. Provided we choose $l \notin B$, the term $f(B)g(\bar{B})$ can be written as $g(x_l)$ times a function independent of x_l. Thus $f(B)g(\bar{B})$ is an even function of x_l, and hence integrates to zero against any odd function of x_l. In particular,

$$\int_{\min\{|x_1|,\dots,|x_n|\}>\varepsilon} \cdots \int \frac{f(B)g(\bar{B})}{x_1 \dots x_r}\, dx_1 \dots dx_n = 0$$

for any proper subset B of $\{1, \dots, n\}$ and any positive ε. Since the term in the sum (2–35) corresponding to $B = \{1, \dots, n\}$ is simply the function f itself, we see that

$$\int_{\min\{|x_1|,\dots,|x_n|\}>\varepsilon} \cdots \int \frac{f(x_1, \dots, x_n)}{x_1 \dots x_n}\, dx_1 \dots dx_n = \int_{\min\{|x_1|,\dots,|x_n|\}>\varepsilon} \cdots \int \frac{G(f)}{x_1 \dots x_n}\, dx_1 \dots dx_n \tag{2–36}$$

for any $\varepsilon > 0$.

On the other hand, we claim that $G(f)$ evaluates to zero when any of the variables x_l equals zero. To see this, let B be a subset of $\{1, \dots, n\}$ not containing l. When $x_l = 0$ we see that the term $(-1)^{n-|B|} f(B) g(\bar{B})$ corresponding to B in the sum (2–35) reduces to

$$(-1)^{n-|B|} f(B) g(\bar{B} \setminus \{l\}).$$

The term

$$(-1)^{n-|B\cup\{l\}|} f(B \cup \{l\}) g(\overline{B \cup \{l\}}) = (-1)^{n-1-|B|} f(B \cup \{l\}) g(\bar{B} \setminus \{l\})$$

corresponding to $B \cup \{l\}$ reduces to

$$(-1)^{n-1-|B|} f(B) g(\bar{B} \setminus \{l\})$$

when $x_l = 0$. It follows that when $x_l = 0$, the terms in (2–35) will cancel pairwise in the natural pairing between the subsets of $\{1, \dots, n\}$ not containing l and those containing l.

Because of this, Lemma 2.4 tells us that the function $G(f)/x_1 \dots x_n$ is integrable, whence the dominated convergence theorem implies

$$\lim_{\varepsilon \to 0} \underset{\min\{|x_1|,\dots,|x_n|\} > \varepsilon}{\int \cdots \int} \frac{G(f)}{x_1 \dots x_n} \, dx_1 \dots dx_n = \int \cdots \int \frac{G(f)}{x_1 \dots x_n} \, dx_1 \dots dx_n. \quad (2\text{–}37)$$

This together with equation (2–36) shows that the principal value (2–30) exists — in fact it equals the integral on the right-hand side of equation (2–37). □

Lemma 2.6. *If f is a well-behaved function on $\mathbb{R}^n$, then for any $1 \le k \le n$,*

$$\lim_{c \to 0+} c^{n-k} \underset{\mathbb{R}^n}{\int \cdots \int} \frac{f(x_1, \dots, x_n) x_1 \dots x_k}{(c^2 + x_1^2) \dots (c^2 + x_n^2)} \, dx_1 \dots dx_n = \pi^{n-k} \, \mathrm{P.V.} \underset{\mathbb{R}^k}{\int \cdots \int} \frac{f(x_1, \dots, x_k, 0, \dots, 0)}{x_1 \dots x_k} \, dx_1 \dots dx_k. \quad (2\text{–}38)$$

Proof. We proceed along lines similar to the proof of Lemma 2.5. Analogously to the definition (2–35) of the operator $G(f)$, define the operator

$$G_k(f) = G_k(f)(x_1, \dots, x_n) = \sum_{B \subset \{1, \dots, k\}} (-1)^{k-|B|} f\big(B \cup \{k+1, \dots, n\}\big) g(\bar{B}),$$

so that $G_k(f)$ is itself a well-behaved function. The arguments leading to the validity of equation (2–36) in the proof of Lemma 2.5 show that the function $G_k(f) - f$ integrates to 0 against any function that is odd in each of the variables $x_1, \dots, x_k$ separately. In particular,

$$c^{n-k} \int \cdots \int \frac{f(x_1, \dots, x_n) x_1 \dots x_k}{(c^2 + x_1^2) \dots (c^2 + x_n^2)} \, dx_1 \dots dx_n = c^{n-k} \int \cdots \int \frac{G_k(f)(x_1, \dots, x_n) x_1 \dots x_k}{(c^2 + x_1^2) \dots (c^2 + x_n^2)} \, dx_1 \dots dx_n.$$

Making the change of variables $x_j \mapsto c x_j$ for $k < j \le n$ and rearranging terms, we see that

$$c^{n-k} \int \cdots \int \frac{G_k(f)(x_1, \dots, x_n) x_1 \dots x_k}{(c^2 + x_1^2) \dots (c^2 + x_n^2)} \, dx_1 \dots dx_n = \int \cdots \int \frac{\tilde{G}_k(f)(x_1, \dots, x_k, c x_{k+1}, \dots, c x_n) x_1^2 \dots x_k^2}{(c^2 + x_1^2) \dots (c^2 + x_k^2)(1 + x_{k+1}^2) \dots (1 + x_n^2)} \times dx_1 \dots dx_n, \quad (2\text{–}39)$$

where we have defined

$$\tilde{G}_k(f)(x) = \frac{G_k(f)(x)}{x_1 \dots x_k}.$$

As in the proof of Lemma 2.5, we can check that $G_k(f)$ evaluates to zero whenever any of the first k variables equals zero, and thus by Lemma 2.4 the function $\tilde{G}_k(f)$ is continuous and integrable and satisfies an upper bound of the form

$$|\tilde{G}_k(f)(x)| \le \beta_1 e^{-\beta_2 \|x\|} \quad (2\text{–}40)$$

for some positive constants β_1 and β_2.

Now define $S_c(x_1, \dots, x_k)$ to be equal to

$$\beta_1 e^{-\beta_2 c \|x\|}$$

if $\|x\| > 1/\sqrt{c}$ and to

$$\max_{|t_{k+1}|, \dots, |t_n| \le \sqrt{c}} \left| \frac{\tilde{G}_k(f)(x_1, \dots, x_k, t_{k+1}, \dots, t_n)}{(1 + x_{k+1}^2) \dots (1 + x_n^2)} \right|$$

if $\|x\| \le 1/\sqrt{c}$. One can check that the integrand on the right-hand side of equation (2–39) is bounded in absolute value by $S_c(x_1, \dots, x_k)$ when $0 < c < 1$. Moreover, the continuity of $\tilde{G}_k(f)$ implies that S_c is bounded on the set $\{x \in \mathbb{R}^k : \|x\| \le 1/\sqrt{c}\}$, and therefore S_c is integrable. Furthermore, both S_c and the integrand on the right-hand side of equation (2–39) tend pointwise to the function

$$\frac{\tilde{G}_k(f)(x_1, \dots, x_k, 0, \dots, 0)}{(1 + x_{k+1}^2) \dots (1 + x_n^2)}$$

as c tends to zero, and this function is also integrable by the exponential decay (2–40) of $\tilde{G}_k(f)$. Therefore, taking limits on both sides of equation (2–39) and using the generalized dominated convergence theorem, we conclude that

$$\lim_{c \to 0+} c^{n-k} \int \cdots \int \frac{G_k(f)(x_1, \dots, x_n) x_1 \dots x_k}{(c^2 + x_1^2) \dots (c^2 + x_n^2)} \, dx_1 \dots dx_n$$

$$= \int\cdots\int \frac{\tilde{G}_k(f)(x_1,\dots,x_k,0,\dots,0)}{(1+x_{k+1}^2)\dots(1+x_n^2)}\,dx_1\dots dx_n$$
$$= \pi^{n-k}\int\cdots\int \frac{G_k(f)(x_1,\dots,x_k,0,\dots,0)}{x_1\dots x_k}\,dx_1\dots dx_k.$$

But just as in the proof of Lemma 2.5, this last integral equals the principal value of the integral of $f(x_1,\dots,x_k,0,\dots,0)/x_1\dots x_k$, which establishes the lemma. □

Of course, the lemma would also hold if both occurrences of the product $x_1\dots x_k$ in equation (2–38) were replaced by any product $x_{j_1}\dots x_{j_k}$ of k distinct variables (and the variables of integration on the right-hand side adjusted accordingly).

2D. Analysis for the Special Case

In this section we derive analytic expressions for the logarithmic density $\delta_{8;3,5,7}$ of the set

$$\{x\in\mathbb{R} : \pi(x;8,3) > \pi(x;8,5) > \pi(x;8,7)\}$$

and for the other densities in Theorem 1. These formulas are special cases of Theorem 4, which will be established in the next section; however, we present a complete analysis in these special cases to illustrate and motivate the techniques in the general case.

We begin by noting the special case

$$\delta_{8;3,5,7} = \iiint\limits_{x>y>z} d\mu_{8;3,5,7}(x,y,z)$$

of equation (2–5). Making the change of variables $u = x-y$, $v = y-z$, and $w = z$ gives

$$\delta_{8;3,5,7} = \int\limits_{u>0}\int\limits_{v>0}\int\limits_{w\in\mathbb{R}} d\nu_{8;3,5,7}(u,v,w),$$

where the measure $\nu_{q;a_1,\dots,a_r}$ is defined, in obvious notation, by

$$\nu_{8;3,5,7}(u,v,w) = \mu_{8;3,5,7}(u+v+w,\ v+w,\ w), \tag{2–41}$$

or equivalently

$$\mu_{8;3,5,7}(x,y,z) = \nu_{8;3,5,7}(x-y,\ y-z,\ z).$$

Integrating out the w variable, we obtain

$$\delta_{8;3,5,7} = \int\limits_{u>0}\int\limits_{v>0} d\rho_{8;3,5,7}(u,v), \tag{2–42}$$

where we have defined, again in obvious notation,

$$\rho_{8;3,5,7}(u,v) = \int\limits_{w\in\mathbb{R}} d\nu_{8;3,5,7}(u,v,w). \tag{2–43}$$

It is easily checked that the Fourier transform of $\rho_{8;3,5,7}$ is related to that of $\mu_{8;3,5,7}$ via

$$\hat\rho_{8;3,5,7}(\xi,\eta) = \hat\mu_{8;3,5,7}(\xi,\ \eta-\xi,\ -\eta),$$

which is a particular case of equation (2–20).

We can appeal to the formula (2–21) for $\hat\rho_{q;a_1,\dots,a_r}$ to write $\hat\rho_{8;3,5,7}(\xi,\eta)$ explicitly. Recall that a discriminant is an integer congruent to 0 or 1 mod 4, and a fundamental discriminant D is an integer that cannot be written in the form $D = dn^2$ for some discriminant d and integer $n\ge 2$. For any fundamental discriminant D, let χ_D denote the character

$$\chi_D(n) = \Big(\frac{D}{n}\Big)$$

where we use Kronecker's extension of the Legendre symbol [Davenport 1980, Chapter 5]. Then the three nonprincipal characters mod 8 are simply χ_{-8}, χ_{-4}, and χ_8. In this setting, equation (2–21) becomes

$$\hat\rho_{8;3,5,7}(\xi,\eta) = F(|2\xi|,\chi_{-8})F(|2\eta-2\xi|,\chi_{-4})F(|-2\eta|,\chi_8), \tag{2–44}$$

showing that $\hat\rho_{8;3,5,7}$ is real-valued and symmetric with respect to reflection through the origin. The same argument gives formulas for $\delta_{q;a,b,c}$ for any permutation $\{a,b,c\}$ of $\{3,5,7\}$, where the arguments of the $F(\cdot,\chi)$ functions in equation (2–44) simply are permuted accordingly. Since each $F(z,\chi)$ is an even function, we can omit the absolute value signs in these arguments. Similar remarks hold for the modulus 12, where the three nonprincipal characters are (induced by) χ_{-4}, χ_{-3}, and χ_{12}.

Using the monotone convergence theorem and the Fourier inversion formula (2–29), equation (2–42) becomes

$$\delta_{8;3,5,7} = \lim_{c\to 0+}\int\limits_{u>0}\int\limits_{v>0} e^{-c(u+v)}\,d\rho_{8;3,5,7}(u,v)$$
$$= \lim_{c\to 0+}\int\limits_{u>0}\int\limits_{v>0} e^{-c(u+v)}\bigg(\frac{1}{4\pi^2}\iint e^{i(u\xi+v\eta)} \times \hat\rho_{8;3,5,7}(\xi,\eta)\,d\xi\,d\eta\bigg)\,du\,dv.$$

We next use Fubini's theorem to write

$$\begin{aligned}
&\delta_{8;3,5,7}\\
&= \frac{1}{4\pi^2}\lim_{c\to 0+}\iint \hat\rho_{8;3,5,7}(\xi,\eta)\\
&\qquad\qquad \left(\int\limits_{u>0}\int\limits_{v>0} e^{u(-c+i\xi)+v(-c+i\eta)}\,du\,dv\right) d\xi\,d\eta\\
&= \frac{1}{4\pi^2}\lim_{c\to 0+}\iint \frac{\hat\rho_{8;3,5,7}(\xi,\eta)}{(c-i\xi)(c-i\eta)}\,d\xi\,d\eta\\
&= \frac{1}{4\pi^2}\lim_{c\to 0+}\iint \frac{\hat\rho_{8;3,5,7}(\xi,\eta)(c^2+ic(\xi+\eta)-\xi\eta)}{(c^2+\xi^2)(c^2+\eta^2)}\,d\xi\,d\eta\\
&= \frac{1}{4\pi^2}(G_{8;3,5,7}+iH_{8;3,5,7}-I_{8;3,5,7}),
\end{aligned} \tag{2–45}$$

where we have defined

$$G_{8;3,5,7} = \lim_{c\to 0+} c^2 \iint \frac{\hat\rho_{8;3,5,7}(\xi,\eta)}{(c^2+\xi^2)(c^2+\eta^2)}\,d\xi\,d\eta\,, \tag{2–46}$$

$$H_{8;3,5,7} = \lim_{c\to 0+} c \iint \frac{\hat\rho_{8;3,5,7}(\xi,\eta)(\xi+\eta)}{(c^2+\xi^2)(c^2+\eta^2)}\,d\xi\,d\eta\,, \tag{2–47}$$

$$I_{8;3,5,7} = \lim_{c\to 0+} \iint \frac{\hat\rho_{8;3,5,7}(\xi,\eta)\xi\eta}{(c^2+\xi^2)(c^2+\eta^2)}\,d\xi\,d\eta\,. \tag{2–48}$$

In equation (2–46) we make the change of variables $\alpha = \xi/c$ and $\beta = \eta/c$ to obtain

$$\begin{aligned}
G_{8;3,5,7} &= \lim_{c\to 0+}\iint \frac{\hat\rho_{8;3,5,7}(c\alpha,c\beta)}{(1+\alpha^2)(1+\beta^2)}\,d\xi\,d\eta\\
&= \iint \frac{\hat\rho_{8;3,5,7}(0,0)}{(1+\alpha^2)(1+\beta^2)}\,d\alpha\,d\beta = \pi^2,
\end{aligned}$$

where we have again used the dominated convergence theorem together with the trivial bound

$$|\hat\rho_{8;3,5,7}(\xi,\eta)| \le \hat\rho_{8;3,5,7}(0,0) = 1.$$

Next, we note that $H_{8;3,5,7}$ equals zero since the integrand in equation (2–47) is odd under reflection through the origin. Finally, we observe that equation (2–48) may be written as

$$\begin{aligned}
&I_{8;3,5,7} = \lim_{c\to 0+}\\
&\iint \frac{(\hat\rho_{8;3,5,7}(\xi,\eta)-\hat\rho_{8;3,5,7}(\xi,0)\hat\rho_{8;3,5,7}(0,\eta))\,\xi\eta}{(c^2+\xi^2)(c^2+\eta^2)}\,d\xi\,d\eta,
\end{aligned}$$

since the term introduced is odd in either variable separately and so integrates to zero. This is the same as

$$\begin{aligned}
&I_{8;3,5,7}\\
&= \lim_{c\to 0+}\iint \frac{\hat\rho_{8;3,5,7}(\xi,\eta)-\hat\rho_{8;3,5,7}(\xi,0)\hat\rho_{8;3,5,7}(0,\eta)}{\xi\eta}\\
&\qquad\qquad\qquad \times \frac{\xi^2\eta^2}{(c^2+\xi^2)(c^2+\eta^2)}\,d\xi\,d\eta.
\end{aligned}$$

Note that the expression

$$\hat\rho_{8;3,5,7}(\xi,\eta) - \hat\rho_{8;3,5,7}(\xi,0)\hat\rho_{8;3,5,7}(0,\eta)$$

is well-behaved by Lemma 2.3, and since

$$\hat\rho_{8;3,5,7}(0,0) = 1,$$

it evaluates to zero when either ξ or η equals zero. Therefore, the first fraction in the integrand can be extended across the ξ and η axes to a continuous integrable function by Lemma 2.4. We may thus use the dominated convergence theorem to see that

$$\begin{aligned}
&I_{8;3,5,7}\\
&= \iint \frac{\hat\rho_{8;3,5,7}(\xi,\eta)-\hat\rho_{8;3,5,7}(\xi,0)\hat\rho_{8;3,5,7}(0,\eta)}{\xi\eta}\,d\xi\,d\eta.
\end{aligned} \tag{2–49}$$

This integral may be written as the multivariate Cauchy principal value

$$I_{8;3,5,7} = \text{P.V.}\iint \frac{\hat\rho_{8;3,5,7}(\xi,\eta)}{\xi\eta}\,d\xi\,d\eta \tag{2–50}$$

as discussed in Section 2C, since $\hat\rho_{8;3,5,7}(\xi,0)$ and $\hat\rho_{8;3,5,7}(0,\eta)$ are even functions and hence the term omitted in passing from (2–49) to (2–50) is odd in either variable. (Of course, we could have arrived at (2–50) directly from the definition of $I_{8;3,5,7}$ by invoking Lemma 2.6; however, not only is this derivation more concrete, in keeping with the spirit of this section, but we will also need the formula (2–49) during our error analysis in Section 3.)

It follows that the right-hand side of (2–45) can be evaluated to give

$$\begin{aligned}
\delta_{8;3,5,7} &= \frac14 - \frac{1}{4\pi^2} I_{8;3,5,7}\\
&= \frac14 - \frac{1}{4\pi^2}\,\text{P.V.}\iint \frac{\hat\rho_{8;3,5,7}(\xi,\eta)}{\xi\eta}\,d\xi\,d\eta\,,
\end{aligned} \tag{2–51}$$

where $\hat\rho_{8;3,5,7}$ is given explicitly in equation (2–44). The identical argument, of course, applies for evaluating $\delta_{8;a,b,c}$ for any permutation $\{a,b,c\}$ of $\{3,5,7\}$ to yield

$$\delta_{8;a,b,c} = \frac{1}{4} - \frac{1}{4\pi^2}\,\mathrm{P.V.}\iint \frac{\hat\rho_{8;a,b,c}(\xi,\eta)}{\xi\eta}\,d\xi\,d\eta, \tag{2–52}$$

where $\rho_{8;a,b,c}$ is defined via obvious analogy to (2–41) and (2–43), and similarly

$$\delta_{12;a,b,c} = \frac{1}{4} - \frac{1}{4\pi^2}\,\mathrm{P.V.}\iint \frac{\hat\rho_{12;a,b,c}(\xi,\eta)}{\xi\eta}\,d\xi\,d\eta, \tag{2–53}$$

for any permutation $\{a,b,c\}$ of $\{5,7,11\}$.

We remark that the numerator of the integrand in (2–49) may be viewed as a "measure of dependence" in the Fourier domain for the bivariate distribution of a random vector (X,Y) in $\mathbb{R}^2$ having density $\rho_{8;3,5,7}$. In fact, the integrand in (2–49) is the Fourier transform of the natural dependence measure based on factorizability of the bivariate cumulative distribution function corresponding to $\rho_{8;3,5,7}$. This interpretation is important in Section 3A, where a random vector (X,Y) of this type is analyzed to yield bounds for the tail of the measure $\rho_{8;3,5,7}$.

2E. Analysis for the General Case

We are now at the point where we have the notation and tools needed for the statement and proof of a general formula for the densities $\delta_{q;a_1,\dots,a_r}$.

Theorem 4. *Assume GRH and LI. Let $q,r\ge 2$ be integers, and let $a_1,\dots,a_r$ be distinct reduced residue classes mod q. Then*

$$\delta_{q;a_1,\dots,a_r} = 2^{-(r-1)}\Biggl(1+\sum_{\substack{B\subset\{1,\dots,r-1\}\\ B\ne\varnothing}}\Bigl(\frac{i}{\pi}\Bigr)^{|B|} \times \mathrm{P.V.}\int\!\cdots\!\int \hat\rho_{q;a_1,\dots,a_r}(B)\prod_{j\in B}\frac{d\eta_j}{\eta_j}\Biggr), \tag{2–54}$$

where $\hat\rho_{q;a_1,\dots,a_r}(B)$ uses the notation of (2–34) *applied to the function*

$$\hat\rho_{q;a_1,\dots,a_r}(\eta_1,\dots,\eta_{r-1}) = \exp\Biggl(\sum_{j=1}^{r-1}(c(q,a_j)-c(q,a_{j+1}))\eta_j\Biggr) \times \prod_{\substack{\chi \bmod q\\ \chi\ne\chi_0}} F\Biggl(\Biggl|\sum_{j=1}^{r-1}(\chi(a_j)-\chi(a_{j-1}))\eta_j\Biggr|,\ \chi\Biggr).$$

Proof. We follow the strategy used for the special cases in Section 2D. For notational simplicity we use the abbreviations

$$\delta=\delta_{q;a_1,\dots,a_r},\qquad \rho=\rho_{q;a_1,\dots,a_r},$$

and so on.

Our starting point is equation (2–5),

$$\delta = \int\!\cdots\!\int_{x_1>x_2>\cdots>x_r} d\mu(x_1,\dots,x_r).$$

We make the change of variables $u_1=x_1-x_2,\ \dots,\ u_{r-1}=x_{r-1}-x_r,\ u_r=x_r$ to obtain

$$\delta = \int\!\cdots\!\int_{\substack{u_1>0,\ \dots,\ u_{r-1}>0\\ u_r\in\mathbb{R}}} d\nu(u_1,\dots,u_r),$$

where the measure ν is defined, in obvious notation, by

$$\nu(u_1,\dots,u_r)=\mu(u_1+\cdots+u_r,\,u_2+\cdots+u_r,\dots,u_r),$$

or equivalently

$$\mu(x_1,\dots,x_r)=\nu(x_1-x_2,\dots,x_{r-1}-x_r,x_r).$$

Integrating out the u_r variable leads to

$$\delta = \int_{u_1>0}\!\cdots\int_{u_{r-1}>0} d\rho(u_1,\dots,u_{r-1}), \tag{2–55}$$

where we have defined (again in obvious notation)

$$\rho(u_1,\dots,u_{r-1}) = \int_{v\in\mathbb{R}} d\nu(u_1,\dots,u_{r-1},v).$$

It is easily checked that the Fourier transform of ρ is related to that of μ by the identity (2–20).

At this point, our goal is to evaluate the integral on the right-hand side of equation (2–55) in terms of the Fourier transform $\hat\rho$ of ρ. The correct final formula could be obtained by writing this as the integral of $d\rho$ against the characteristic function of the region of integration, and using Parseval's identity in the context of the theory of generalized functions; the following analysis derives this final formula rigorously.

Using first the monotone convergence theorem and then the Fourier inversion formula (2–29), equation (2–55) becomes

$$\delta = \lim_{c\to 0+} \int\limits_{u_1>0}\cdots\int\limits_{u_{r-1}>0} e^{-c(u_1+\cdots+u_{r-1})}\,d\rho(u_1,\ldots,u_{r-1})$$

$$= \lim_{c\to 0+} \int\limits_{u_1>0}\cdots\int\limits_{u_{r-1}>0} e^{-c(u_1+\cdots+u_{r-1})} \times \biggl((2\pi)^{-(r-1)}\int\!\cdots\!\int e^{i(u_1\xi_1+\cdots+u_{r-1}\xi_{r-1})} \times \hat\rho(\xi_1,\ldots,\xi_{r-1})\,d\xi_1\ldots d\xi_{r-1}\biggr) \times du_1\ldots du_{r-1}.$$

Then by Fubini's theorem this becomes

$$(2\pi)^{(r-1)}\times\delta = \lim_{c\to 0+}\int\!\cdots\!\int \hat\rho(\xi_1,\ldots,\xi_{r-1}) \times\biggl(\int\limits_{u_1>0}\cdots\int\limits_{u_{r-1}>0} e^{u_1(-c+i\xi_1)+\cdots+u_{r-1}(-c+i\xi_{r-1})} \times du_1\ldots du_{r-1}\biggr)\,d\xi_1\ldots d\xi_{r-1}$$

$$= \lim_{c\to 0+}\int\!\cdots\!\int \frac{\hat\rho(\xi_1,\ldots,\xi_{r-1})}{(c-i\xi_1)\ldots(c-i\xi_{r-1})}\,d\xi_1\ldots d\xi_{r-1}$$

$$= \lim_{c\to 0+}\int\!\cdots\!\int \frac{\hat\rho(\xi_1,\ldots,\xi_{r-1})(c+i\xi_1)\ldots(c+i\xi_{r-1})}{(c^2+\xi_1^2)\ldots(c^2+\xi_{r-1}^2)} \times d\xi_1\ldots d\xi_{r-1},$$

and expanding the product $(c+i\xi_1)\ldots(c+i\xi_{r-1})$ leads to

$$\delta = (2\pi)^{-(r-1)}\sum_{B\subset\{1,2,\ldots,r-1\}} i^{|B|}\,I(B), \tag{2–56}$$

where we have defined

$$I(B) = \lim_{c\to 0+} c^{r-1-|B|} \int\!\cdots\!\int \frac{\hat\rho(\xi_1,\ldots,\xi_{r-1})\bigl(\prod_{j\in B}\xi_j\bigr)}{(c^2+\xi_1^2)\ldots(c^2+\xi_{r-1}^2)}\,d\xi_1\ldots d\xi_{r-1}.$$

Appealing to Lemma 2.6 with $n=r-1$ and $k=|B|$, we see that

$$I(B) = \pi^{r-1-|B|}\,\text{P.V.}\int\!\cdots\!\int \hat\rho(B)\prod_{j\in B}\frac{d\eta_j}{\eta_j},$$

which includes the special case $I(\varnothing)=\pi^{r-1}$. Using this fact in equation (2–56) establishes the theorem. □

The measure $\rho_{q;a_1,\ldots,a_r}$ is the limiting distribution of the vector

$$\frac{\varphi(q)\log x}{\sqrt{x}}\bigl(\pi(x;q,a_1)-\pi(x;q,a_2),\ \ldots,\ \pi(x;q,a_{r-1})-\pi(x;q,a_r)\bigr)$$

in $\mathbb{R}^{r-1}$, so its usefulness to the investigation of those x with $\pi(x;q,a_1)>\cdots>\pi(x;q,a_r)$ is not surprising.

To conclude this section, we consider two special cases of Theorem 4. In the case $r=2$ (in other words, when we are comparing simply a pair a_1,a_2 of residues modulo q) the formula (2–54) reduces to

$$\delta_{q;a_1,a_2} = \frac12\biggl(1+\frac{i}{\pi}\,\text{P.V.}\int\frac{\hat\rho_{q;a_1,a_2}(\eta)d\eta}{\eta}\biggr) = \frac12-\frac{1}{2\pi}\int\frac{\sin(\{c(q,a_1)-c(q,a_2)\}\eta)}{\eta} \times\prod_{\substack{\chi \bmod q\\ \chi\ne\chi_0}} F_\chi\bigl(|\chi(a_1)-\chi(a_2)|\eta\bigr)\,d\eta, \tag{2–57}$$

the corresponding cosine term in the last integral being omitted by virtue of symmetry. When $c(q,a_1)=c(q,a_2)$, the integrand is identically zero and hence $\delta_{q;a_1,a_2}=\frac12$, as was proved by Rubinstein and Sarnak. In fact, our formula (2–57) is analogous to one of theirs [Rubinstein and Sarnak 1994, equation 4.1].

In the case $r=3$, Theorem 4 becomes

$$\delta_{q;a_1,a_2,a_3} = \frac14+\frac{i}{4\pi}\,\text{P.V.}\int\bigl(\hat\rho_{q;a_1,a_2,a_3}(\eta,0)+\hat\rho_{q;a_1,a_2,a_3}(0,\eta)\bigr)\frac{d\eta}{\eta} - \frac{1}{4\pi^2}\,\text{P.V.}\iint\hat\rho_{q;a_1,a_2,a_3}(\eta_1,\eta_2)\,\frac{d\eta_1 d\eta_2}{\eta_1\eta_2}.$$

If the a_j are all squares or all nonsquares, the one-dimensional integral again vanishes due to symmetry, yielding a generalization of the formulas (2–52) and (2–53) of Section 2D.

3. RIGOROUS ERROR BOUNDS

In this section, we describe how the densities in Theorem 1 were calculated and provide a rigorous analysis bounding the error between the calculated and true values.

Suppose that we wish to evaluate $\delta_{8;3,5,7}$. According to equation (2–51), we need only to evaluate

$$\text{P.V.}\iint\frac{\hat\rho_{8;3,5,7}(\xi,\eta)}{\xi\eta}\,d\xi\,d\eta,$$

which in turn by formula (2–44) equals

$$\text{P.V.}\iint \frac{F(2\xi,\chi_{-8})F(2\eta-2\xi,\chi_{-4})F(-2\eta,\chi_8)}{\xi\eta}\,d\xi\,d\eta. \tag{3–1}$$

We shall approximate this integral by sampling the integrand on the (symmetrically offset) grid of points

$$\Big\{\Big(\frac{m\varepsilon}{2},\frac{n\varepsilon}{2}\Big) : \Big|\frac{m\varepsilon}{2}\Big|, \Big|\frac{n\varepsilon}{2}\Big| \le C,\ m,n \text{ odd}\Big\}$$

for some appropriately small $\varepsilon > 0$ and some appropriately large $C > 0$. In fact the quantity we actually compute is $4S_{8;3,5,7}(\varepsilon, C, T)$, where we define

$$S_{8;3,5,7}(\varepsilon,C,T) = \sum_{\substack{|m|,|n|\le 2C/\varepsilon \\ m,n\,\text{odd}}}\sum \frac{F_T(m\varepsilon,\chi_{-8})F_T((n-m)\varepsilon,\chi_{-4})F_T(-n\varepsilon,\chi_8)}{mn}; \tag{3–2}$$

here $F_T(z,\chi)$ is the approximation to $F(z,\chi)$ defined in equation (2–11), and as before χ_D is the character given by the Kronecker symbol

$$\chi_D(n) = \Big(\frac{D}{n}\Big).$$

The quantity $S_{8;3,5,7}(\varepsilon, C, T)$ is a discrete, truncated approximation to the integral (3–1) involving an approximated summand as well. The overall error incurred in evaluating (3–1) by means of (3–2) thus consists of three components: error due to discretizing the integral, error due to truncating the resulting infinite sum, and error due to approximating the summand. In Sections 3A to 3C we obtain rigorous bounds for each of these sources of error, and in Section 3D we combine these bounds to establish Theorem 1. Section 3E provides some technical bounds that are required for our arguments in Section 3A. While in the sections to follow, all of the specific expressions we write down (such as $S_{8;3,5,7}(\varepsilon, C, T)$) are those that arise in the calculation of the single density $\delta_{8;3,5,7}$, the given constants and error bounds were chosen so as to apply also to the analogous quantities arising during the calculation of any of the densities listed in Theorem 1.

3A. Error Due To Discretization

The first step is to discretize the calculation of $I_{8;3,5,7}$ by converting the integral defining $I_{8;3,5,7}$ into a sum; we may bound the error incurred by doing so using the Poisson summation formula, as we now explain. Let $f(\xi,\eta)$ be a continuous, integrable function on $\mathbb{R}^2$ such that both f and $\hat f$ decay rapidly enough near infinity (for instance, exponential decay certainly suffices). Then f satisfies the Poisson summation formula

$$\varepsilon_1\varepsilon_2 \sum_{k=-\infty}^{\infty}\sum_{l=-\infty}^{\infty} f(k\varepsilon_1+\alpha, l\varepsilon_2+\beta) = \sum_{\kappa=-\infty}^{\infty}\sum_{\lambda=-\infty}^{\infty} \hat f\Big(\frac{2\pi\kappa}{\varepsilon_1},\frac{2\pi\lambda}{\varepsilon_2}\Big) e^{2\pi i(\kappa\alpha/\varepsilon_1+\lambda\beta/\varepsilon_2)}$$

(see for instance [Stein and Weiss 1971, Corollary 2.6 of Chapter VII], although we are using a Fourier transform (2–6) with a different choice of constants). In this formula, set $\varepsilon_1 = \varepsilon_2 = \varepsilon$ and $\alpha = \beta = \varepsilon/2$, and make the change of variables $m = 2k+1$ and $n = 2l+1$ on the left-hand side, to obtain

$$\varepsilon^2 \sum_{\substack{m,n\in\mathbb{Z}\\ m,n\text{ odd}}}\sum f\Big(\frac{m\varepsilon}{2},\frac{n\varepsilon}{2}\Big) = \hat f(0,0) + \sum_{\substack{\kappa,\lambda\in\mathbb{Z}\\ (\kappa,\lambda)\ne(0,0)}}\sum \hat f\Big(\frac{2\pi\kappa}{\varepsilon},\frac{2\pi\lambda}{\varepsilon}\Big)(-1)^{\kappa+\lambda}. \tag{3–3}$$

Now let

$$f(\xi,\eta) = \frac{\hat\rho_{8;3,5,7}(\xi,\eta) - \hat\rho_{8;3,5,7}(\xi,0)\hat\rho_{8;3,5,7}(0,\eta)}{\xi\eta},$$

which can be extended continuously over the coordinate axes as was noted in Section 2C. This function f is integrable and has exponential decay near infinity by Lemmas 2.3 and 2.4, and its Fourier transform can be seen to equal

$$\hat f(u,v) = -4\pi^2(\overline{P}(u,v) - \overline{P}_1(u)\overline{P}_2(v)), \tag{3–4}$$

where

$$\overline{P}(u,v) = \int_u^\infty\int_v^\infty d\rho_{8;3,5,7}$$

is the upper cumulative distribution function of the measure $\rho_{8;3,5,7}$ and $\overline{P}_1(u) = \overline{P}(u,-\infty)$ and $\overline{P}_2(v) = \overline{P}(-\infty,v)$ are the corresponding "upper marginals". (Note that $\hat f(u,v)$ is a dependence measure of the type mentioned at the end of Section 2D.) At the end of this section we will show that the function $\hat f$ decays exponentially as well, so that we are justified in applying the form (3–3) of the Poisson summation formula to f.

Now observe from equation (2–49) that

$$I_{8;3,5,7} = \iint f(\xi,\eta)\,d\xi\,d\eta = \hat{f}(0,0),$$

so applying equation (3–3) to f we have

$$I_{8;3,5,7} = \varepsilon^2 \sum_{\substack{m,n\in\mathbb{Z}\\ m,n\,\text{odd}}}\sum \frac{\hat\rho_{8;3,5,7}(\frac{m\varepsilon}{2},\frac{n\varepsilon}{2}) - \hat\rho_{8;3,5,7}(\frac{m\varepsilon}{2},0)\hat\rho_{8;3,5,7}(0,\frac{n\varepsilon}{2})}{(\frac{m\varepsilon}{2})(\frac{n\varepsilon}{2})} + \mathrm{Err}_1, \quad (3\text{–}5)$$

where Err_1, the error due to discretization, is given by

$$\mathrm{Err}_1 = 4\pi^2 \sum_{\substack{\kappa,\lambda\in\mathbb{Z}\\ (\kappa,\lambda)\neq(0,0)}}\sum \Bigl(\overline{P}\Bigl(\frac{2\pi\kappa}{\varepsilon},\frac{2\pi\lambda}{\varepsilon}\Bigr) - \overline{P}_1\Bigl(\frac{2\pi\kappa}{\varepsilon}\Bigr)\overline{P}_2\Bigl(\frac{2\pi\lambda}{\varepsilon}\Bigr)\Bigr)(-1)^{\kappa+\lambda}. \quad (3\text{–}6)$$

Defining

$$\begin{aligned}Q(u,v) = {}&\bigl(\overline{P}(2\pi u,2\pi v) - \overline{P}_1(2\pi u)\overline{P}_2(2\pi v)\bigr) + \\ &\bigl(\overline{P}(-2\pi u,2\pi v) - \overline{P}_1(-2\pi u)\overline{P}_2(2\pi v)\bigr) + \\ &\bigl(\overline{P}(2\pi u,-2\pi v) - \overline{P}_1(2\pi u)\overline{P}_2(-2\pi v)\bigr) + \\ &\bigl(\overline{P}(-2\pi u,-2\pi v) - \overline{P}_1(-2\pi u)\overline{P}_2(-2\pi v)\bigr),\end{aligned}$$

and grouping the terms on the right-hand side of equation (3–6) analogously, we obtain

$$\mathrm{Err}_1 = 4\pi^2\biggl(\sum_{\kappa,\lambda\in\mathbb{Z}^+}\sum(-1)^{\kappa+\lambda}Q\Bigl(\frac{\kappa}{\varepsilon},\frac{\lambda}{\varepsilon}\Bigr) + \frac12\sum_{\kappa\in\mathbb{Z}^+}(-1)^{\kappa}Q\Bigl(\frac{\kappa}{\varepsilon},0\Bigr) + \frac12\sum_{\lambda\in\mathbb{Z}^+}(-1)^{\lambda}Q\Bigl(0,\frac{\lambda}{\varepsilon}\Bigr)\biggr),$$

so that

$$|\mathrm{Err}_1| \le 4\pi^2 \sum_{\substack{\kappa,\lambda\ge0\\ (\kappa,\lambda)\neq(0,0)}}\sum \Bigl|Q\Bigl(\frac{\kappa}{\varepsilon},\frac{\lambda}{\varepsilon}\Bigr)\Bigr|. \quad (3\text{–}7)$$

Now let (X,Y) denote a pair of real-valued random variables whose joint distribution is given by $\rho_{8;3,5,7}$ (these random variables are given explicitly in equation (3–33) below, though their explicit form is not needed here). Then $\overline{P}(u,v) = \Pr(X>u,\,Y>v)$ and hence $\overline{P}_1(u)=\Pr(X>u)$ and $\overline{P}_2(v)=\Pr(Y>v)$. With this interpretation, and using the fact that $\rho_{8;3,5,7}$ is symmetric about the origin, the identity

$$Q(u,v) = \Pr(X>2\pi u, Y>2\pi v) - \Pr(X>2\pi u, Y<-2\pi v) \quad (3\text{–}8)$$

is easily verified. Clearly

$$0 \le \Pr(X>u,\,Y>v) \le \min\{\Pr(X>u),\Pr(Y>v)\}.$$

Moreover, since $\rho_{8;3,5,7}$ is symmetric about the origin, each component X and Y is a symmetric random variable, so that

$$\begin{aligned}0 \le \Pr(X>u, Y<-v) &\le \min\{\Pr(X>u),\ \Pr(Y<-v)\}\\ &= \min\{\Pr(X>u),\ \Pr(Y>v)\}.\end{aligned}$$

It therefore follows from the identity (3–8) that

$$|Q(u,v)| \le \min\{\Pr(X>2\pi u),\ \Pr(Y>2\pi v)\}. \quad (3\text{–}9)$$

In Section 3E we shall establish the bounds

$$\begin{aligned}\Pr(X\ge u) &\le \exp(-0.04(u-3)^2)\\ \Pr(Y\ge u) &\le \exp(-0.04(u-3)^2)\end{aligned} \quad (3\text{–}10)$$

for any $u\ge 3$. Hence by the inequality (3–9),

$$|Q(u,v)| \le \exp\bigl(-0.04(2\pi\max\{u,v\}-3)^2\bigr)$$

if either u or v exceeds 1, so that by equation (3–7),

$$\begin{aligned}|\mathrm{Err}_1| &\le 4\pi^2\biggl(\sum_{\kappa=0}^{\infty}\sum_{\lambda=\max\{\kappa,1\}}^{\infty}\exp\Bigl(-0.04\Bigl(\frac{2\pi\lambda}{\varepsilon}-3\Bigr)^2\Bigr) + \sum_{\lambda=0}^{\infty}\sum_{\kappa=\max\{\lambda,1\}}^{\infty}\exp\Bigl(-0.04\Bigl(\frac{2\pi\kappa}{\varepsilon}-3\Bigr)^2\Bigr)\biggr)\\ &\le 8\pi^2\sum_{\kappa=0}^{\infty}\sum_{\lambda=\max\{\kappa,1\}}^{\infty}\exp\Bigl(-0.04\Bigl(\frac{2\pi\lambda}{\varepsilon}-3\Bigr)^2\Bigr)\end{aligned} \quad (3\text{–}11)$$

if $\varepsilon<1$, say. Now, for any positive integer λ_0,

$$\sum_{\lambda=\lambda_0}^{\infty}\exp\Bigl(-0.04\Bigl(\frac{2\pi\lambda}{\varepsilon}-3\Bigr)^2\Bigr) \le 2\exp\Bigl(-0.04\Bigl(\frac{2\pi\lambda_0}{\varepsilon}-3\Bigr)^2\Bigr),$$

since each term of the sum is at most half of the preceding term. Applying this inequality twice in succession to the bound (3–11) gives

$$\begin{aligned}|\mathrm{Err}_1| &\le 8\pi^2\biggl(2\exp\Bigl(-0.04\Bigl(\frac{2\pi}{\varepsilon}-3\Bigr)^2\Bigr) + 2\sum_{\kappa=1}^{\infty}\exp\Bigl(-0.04\Bigl(\frac{2\pi\kappa}{\varepsilon}-3\Bigr)^2\Bigr)\biggr)\\ &\le 48\pi^2\exp\Bigl(-0.04\Bigl(\frac{2\pi}{\varepsilon}-3\Bigr)^2\Bigr).\end{aligned}$$

We therefore conclude that

$$|\mathrm{Err}_1| < 5\times10^{-12} \quad (3\text{–}12)$$

for any choice of $\varepsilon < \frac{1}{5}$, which is more than adequate for our purposes.

To conclude this section, we return to the matter of showing that the function $\hat{f}$ given in equation (3–4) decays exponentially. In terms of the random variables X and Y, the formula (3–4) becomes

$$\hat{f}(u,v) = -4\pi^2(\Pr(X>u,\, Y>v) - \Pr(X>u)\Pr(Y>v)).$$

By an argument similar to the one used for the function Q, we see that

$$\begin{aligned}|\hat{f}(u,v)| &\le 4\pi^2 \max\{\Pr(X>u,\, Y>v), \Pr(X>u)\Pr(Y>v)\} \\ &\le 4\pi^2 \min\{\Pr(X>u), \Pr(Y>v)\}. \end{aligned} \tag{3–13}$$

On the other hand, elementary considerations yield

$$\begin{aligned}\hat{f}(-u,-v) &= 4\pi^2(\Pr(X>-u, Y>-v) - \Pr(X>-u)\Pr(Y>-v)) \\ &= 4\pi^2\big((1 - \Pr(X\le -u) - \Pr(Y\le -v) \\ &\qquad + \Pr(X\le -u, Y\le -v)) \\ &\qquad -(1-\Pr(X\le -u))(1-\Pr(Y\le -v))\big) \\ &= 4\pi^2(\Pr(X\le -u, Y\le -v) - \Pr(X\le -u)\Pr(Y\le -v)).\end{aligned}$$

By the same argument as in equation (3–13) we get

$$\begin{aligned}|\hat{f}(-u,-v)| &\le 4\pi^2 \min\{\Pr(X\le -u), \Pr(Y\le -v)\} \\ &= 4\pi^2 \min\{\Pr(X\ge u), \Pr(Y\ge v)\},\end{aligned}$$

since X and Y are symmetric. We can therefore apply the bounds (3–10) to conclude that

$$|\hat{f}(-u,-v)| \le 4\pi^2 \exp\left(-0.04(\max\{|u|,|v|\}-3)^2\right)$$

if either $|u|$ or $|v|$ exceeds 3. In particular, the function $\hat{f}$ decays (faster than) exponentially, as claimed.

3B. Error Due to Truncating the Infinite Sums

From equation (3–5) we have

$$\begin{aligned}I_{8;3,5,7} &= 4 \sum\sum_{\substack{m,n\in\mathbb{Z} \\ m,n \text{ odd}}} \left(\frac{\hat\rho(\frac{m\varepsilon}{2}, \frac{n\varepsilon}{2})}{mn} - \frac{\hat\rho(\frac{m\varepsilon}{2},0)\hat\rho(0,\frac{n\varepsilon}{2})}{mn} \right) + \mathrm{Err}_1 \\ &= 4S_{8;3,5,7}(\varepsilon) + \mathrm{Err}_1,\end{aligned} \tag{3–14}$$

where $\rho = \rho_{8;3,5,7}$ and we have defined

$$S_{8;3,5,7}(\varepsilon) = \sum\sum_{\substack{m,n\in\mathbb{Z} \\ m,n \text{ odd}}} \frac{\hat\rho_{8;3,5,7}(\frac{m\varepsilon}{2}, \frac{n\varepsilon}{2})}{mn}. \tag{3–15}$$

(The term that has been omitted in the latter equality in equation (3–14) equals zero, since

$$\hat\rho_{8;3,5,7}\big(\tfrac{m\varepsilon}{2},0\big)\hat\rho_{8;3,5,7}\big(0,\tfrac{n\varepsilon}{2}\big)(mn)^{-1}$$

is odd in either variable separately due to the symmetry of the functions $\hat\rho_{8;3,5,7}(\frac{m\varepsilon}{2},0)$ and $\hat\rho_{8;3,5,7}(0,\frac{n\varepsilon}{2})$ through the origin.) At this point we have accomplished the first step of converting our integral $I_{8;3,5,7}$ into a discrete sum, with a manageable error; the next step is to truncate the ranges of summation so that the resulting sum has only finitely many terms.

From the formula (2–44) for $\hat\rho_{8;3,5,7}$, the definition (3–15) becomes

$$\begin{aligned}S_{8;3,5,7}(\varepsilon) &= \sum\sum_{\substack{m,n\in\mathbb{Z} \\ m,n \text{ odd}}} \frac{F(m\varepsilon,\chi_{-8})F((n-m)\varepsilon,\chi_{-4})F(-n\varepsilon,\chi_8)}{mn} \\ &= S_{8;3,5,7}(\varepsilon, C) + \mathrm{Err}_2,\end{aligned} \tag{3–16}$$

where we have defined the truncated series

$$S_{8;3,5,7}(\varepsilon,C) = \sideset{}{'}\sum\sideset{}{'}\sum_{|m|,|n|\le 2C/\varepsilon} \frac{F(m\varepsilon,\chi_{-8})F((n-m)\varepsilon,\chi_{-4})F(-n\varepsilon,\chi_8)}{mn} \tag{3–17}$$

(the primes indicating that the sums are taken over only odd values of m and n) and the error due to truncation

$$\mathrm{Err}_2 = \sideset{}{'}\sum\sideset{}{'}\sum_{\max\{|m|,|n|\}>2C/\varepsilon} \frac{F(m\varepsilon,\chi_{-8})F((n-m)\varepsilon,\chi_{-4})F(-n\varepsilon,\chi_8)}{mn}.$$

We rewrite this as

$$\begin{aligned}\mathrm{Err}_2 = 2\Bigg(&\sideset{}{'}\sum_{m>2C/\varepsilon} \sideset{}{'}\sum_{n=-m}^{m/2} + \sideset{}{'}\sum_{m>2C/\varepsilon} \sideset{}{'}\sum_{n=m/2}^{m} \\ &+ \sideset{}{'}\sum_{n>2C/\varepsilon} \sideset{}{'}\sum_{m=-n}^{n/2} + \sideset{}{'}\sum_{n>2C/\varepsilon} \sideset{}{'}\sum_{m=n/2}^{n} \Bigg) \\ &\frac{F(m\varepsilon,\chi_{-8})F((n-m)\varepsilon,\chi_{-4})F(-n\varepsilon,\chi_8)}{mn},\end{aligned} \tag{3–18}$$

where the factor of 2 comes from grouping together the terms corresponding to (m,n) and $(-m,-n)$ by the symmetry of the summand through the origin.

To bound Err_2, we will certainly need explicit estimates for the functions $F(x,\chi)$ on the real axis. We recall the upper bound (2–16),

$$|F(x,\chi)| \le (\pi|x|)^{-J/2}\prod_{j=1}^{J}\bigl(\tfrac14+\gamma_j^2\bigr)^{1/4},$$

where J is any positive integer and $0<\gamma_1<\gamma_2<\dots$ are the imaginary parts of the nontrivial zeros of $L(s,\chi)$. Any particular choice of J gives an upper bound of the form

$$|F(x,\chi)| \le d(\chi)|x|^{-e(\chi)} \tag{3–19}$$

for some positive constants $d(\chi)$ and $e(\chi)$. For any fixed x the optimal choice of J is the largest integer such that $(\pi x)^2 > \frac14+\gamma_J^2$; but for our present purposes, we obtain sufficiently good results that are easy to apply uniformly in x by choosing J so that γ_J is just less than 30. Table 1 lists, for each of the five characters χ relevant to the densities mod 8 and mod 12, the values of J chosen and the resulting values of $d(\chi)$ and $e(\chi)$ in the bound (3–19), which we computed from the lists of zeros of the $L(s,\chi)$ supplied to us by R. Rumely.

χ	J	$d(\chi)$	$e(\chi)$
χ_{-8}	56	1.3×10^{32}	28
χ_{8}	56	2.1×10^{32}	28
χ_{-4}	46	8.5×10^{26}	23
χ_{-3}	42	7.5×10^{24}	21
χ_{12}	62	3.0×10^{35}	31

TABLE 1. Allowable constants in the bound (3–19) for $|F(x,\chi)|$.

Since $|F(x,\chi)|$ is also bounded by 1 on the real axis, we can estimate the first double sum in equation (3–18) by

$$\begin{aligned}
&\Biggl|\sideset{}{'}\sum_{m>2C/\varepsilon}\ \sideset{}{'}\sum_{n=-m}^{m/2}\frac{F(m\varepsilon,\chi_{-8})F((n-m)\varepsilon,\chi_{-4})F(-n\varepsilon,\chi_8)}{mn}\Biggr|\\
&\quad\le \sum_{m>2C/\varepsilon}\ \sum_{n=-\infty}^{m/2}\left|\frac{F(m\varepsilon,\chi_{-8})F((n-m)\varepsilon,\chi_{-4})}{m}\right|\\
&\quad\le d(\chi_{-8})d(\chi_{-4})\varepsilon^{-e(\chi_{-8})-e(\chi_{-4})}\\
&\qquad\times\sum_{m>2C/\varepsilon}\ \sum_{n=-\infty}^{m/2} m^{-e(\chi_{-8})-1}(m-n)^{-e(\chi_{-4})}
\end{aligned}\tag{3–20}$$

using the bound (3–19) for χ_{-8} and χ_{-4}.

Now we claim that

$$\begin{aligned}
\sum_{m>M}\ \sum_{n=-\infty}^{m/2} m^{-\alpha}(m-n)^{-\beta} &= \sum_{m>M}\ \sum_{n=m/2}^{\infty} m^{-\alpha}n^{-\beta}\\
&< 2^{\beta-1}M^{1-\alpha-\beta}\left(\frac{2}{\alpha+\beta-1}+\frac{M}{(\alpha+\beta-2)(\beta-1)}\right)
\end{aligned}\tag{3–21}$$

for any real numbers $\alpha,\beta>1$. The equality is clear upon making the change of variables $n\mapsto m-n$, while the inequality follows from the elementary argument

$$\begin{aligned}
&\sum_{m>M}\ \sum_{n=m/2}^{\infty} m^{-\alpha}n^{-\beta}\\
&\quad< \sum_{m>M} m^{-\alpha}\left(\left\lceil\frac{m}{2}\right\rceil^{-\beta}+\int_{\lceil m/2\rceil}^{\infty} t^{-\beta}\,dt\right)\\
&\quad\le \sum_{m>M} m^{-\alpha}\left(\left(\frac{m}{2}\right)^{-\beta}+\frac{1}{\beta-1}\left(\frac{m}{2}\right)^{1-\beta}\right)\\
&\quad< 2^{\beta}\int_M^{\infty} t^{-\alpha-\beta}\,dt+\frac{2^{\beta-1}}{\beta-1}\int_M^{\infty} t^{1-\alpha-\beta}\,dt\\
&\quad= \frac{2^{\beta}}{\alpha+\beta-1}M^{1-\alpha-\beta}+\frac{2^{\beta-1}}{(\beta-1)(\alpha+\beta-2)}M^{2-\alpha-\beta},
\end{aligned}$$

this last expression being equivalent to the right-hand side of (3–21).

For the rest of this section we use the abbreviations $e_D=e(\chi_D)$ and $d_D=d(\chi_D)$. Applying the upper bound (3–21), with $\alpha=e_{-8}+1$ and $\beta=e_{-4}$, to equation (3–20) gives

$$\begin{aligned}
&\Biggl|\sideset{}{'}\sum_{m>2C/\varepsilon}\ \sideset{}{'}\sum_{n=-m}^{m/2}\frac{F(m\varepsilon,\chi_{-8})F((n-m)\varepsilon,\chi_{-4})F(-n\varepsilon,\chi_8)}{mn}\Biggr|\\
&\quad< d_{-8}d_{-4}\varepsilon^{-e_{-8}-e_{-4}}2^{e_{-4}-1}\left(\frac{2C}{\varepsilon}\right)^{-e_{-8}-e_{-4}}\\
&\qquad\times\left(\frac{2}{e_{-8}+e_{-4}}+\frac{2C/\varepsilon}{(e_{-8}+e_{-4}-1)(e_{-4}-1)}\right)\\
&\quad= d_{-8}d_{-4}2^{-e_{-8}}C^{-e_{-8}-e_{-4}}\\
&\qquad\times\left(\frac{1}{e_{-8}+e_{-4}}+\frac{C/\varepsilon}{(e_{-8}+e_{-4}-1)(e_{-4}-1)}\right).
\end{aligned}\tag{3–22}$$

Substituting the appropriate values from Table 1, we find that this last expression is less than 1.85×10^{-7} when $\varepsilon=\frac{1}{20}$ and $C=15$.

The second double sum in equation (3–18) may be similarly bounded as

$$\Bigg| \sum_{m>2C/\varepsilon}' \sum_{n=-m/2}^{m}{}' \frac{F(m\varepsilon,\chi_{-8})F((n-m)\varepsilon,\chi_{-4})F(-n\varepsilon,\chi_8)}{mn} \Bigg| \le \sum_{m>2C/\varepsilon} \sum_{n=m/2}^{\infty} \left| \frac{F(m\varepsilon,\chi_{-8})F(-n\varepsilon,\chi_8)}{m} \right| \le d_{-8}d_8\varepsilon^{-e_{-8}-e_8} \sum_{m>2C/\varepsilon} \sum_{n=m/2}^{\infty} m^{-e_{-8}-1}n^{-e_8}.$$

Applying (3–19) to bound this last expression yields just the right-hand side of equation (3–22) except with χ_{-4} replaced with χ_8; upon substituting values from Table 1 we find that this expression is also less than 1.85×10^{-7} when $\varepsilon = \frac{1}{20}$ and $C = 15$. The third and fourth double sums in (3–18) are treated the same way, and so we conclude from equation (3–18) that

$$|\mathrm{Err}_2| < 8(1.85\times10^{-7}) < 1.5\times10^{-6} \tag{3–23}$$

when $\varepsilon = \frac{1}{20}$ and $C = 15$.

3C. Error Due to Approximating $F(z,\chi)$ by $F_T(z,\chi)$

We have accomplished the second step of approximating the infinite sum $S_{8;3,5,7}(\varepsilon)$ by the finite sum $S_{8;3,5,7}(\varepsilon, C)$; however, this latter sum is still unsuitable for computation, since it involves the functions $F(z,\chi)$ which are infinite products. The last step is to replace the functions $F(z,\chi)$ by their truncated counterparts $F_T(z,\chi)$ defined in equation (2–11).

Recall the definition (2–12) of b_1,

$$b_1 = b_1(T,\chi) = -\sum_{\gamma\ge T} \frac{1}{\frac14+\gamma^2},$$

and put

$$\Delta_T(z,\chi) = \frac{\prod_{\gamma>T} J_0(\alpha_\gamma z)}{1+b_1z^2} - 1. \tag{3–24}$$

From the definitions (2–9) and (2–11) of F and F_T we see that

$$F(z,\chi) = F_T(z,\chi)(1+\Delta_T(z,\chi)).$$

Making this substitution in equation (3–17) for χ_{-8}, χ_{-4}, and χ_8, we then obtain

$$S_{8;3,5,7}(\varepsilon, C) = S_{8;3,5,7}(\varepsilon, C, T) + \mathrm{Err}_3, \tag{3–25}$$

where $S_{8;3,5,7}(\varepsilon, C, T)$ is as defined in (3–2) and

$$\mathrm{Err}_3 = \sum_{|m|,|n|\le 2C/\varepsilon}' \sum{}' \frac{F_T(m\varepsilon,\chi_{-8})F_T((n-m)\varepsilon,\chi_{-4})F_T(-n\varepsilon,\chi_8)}{mn} \times \big((1+\Delta_T(m\varepsilon,\chi_{-8}))(1+\Delta_T((n-m)\varepsilon,\chi_{-4})) \times (1+\Delta_T(-n\varepsilon,\chi_8)) - 1\big). \tag{3–26}$$

Regarding the size of the function Δ_T, Rubinstein and Sarnak [1994, Section 4.3] established that

$$\left| \bigg(\prod_{\gamma>T} J_0(\alpha_\gamma x) \bigg) - (1+b_1x^2) \right| \le \frac{b_1^2x^4}{2(1-|b_1|x^2)}$$

for real numbers x satisfying $|b_1|x^2 < 1$. From the definition (3–24) of Δ_T this immediately yields

$$|\Delta_T(x,\chi)| = \frac{\left|\prod_{\gamma>T} J_0(\alpha_\gamma x) - (1+b_1x^2)\right|}{|1+b_1x^2|} < \frac{b_1^2x^4}{2(1-|b_1|x^2)^2} \qquad \text{if } |b_1|x^2<1. \tag{3–27}$$

The quantities b_1 can be computed if we know all the zeros of $L(s,\chi)$ up to height T, since

$$b_1 = \sum_{0<\gamma<T} \frac{1}{\frac14+\gamma^2} - \sum_{\gamma>0} \frac{1}{\frac14+\gamma^2}$$

and we have the formula [Davenport 1980, p. 83]

$$\sum_{\gamma>0} \frac{1}{\frac14+\gamma^2} = \frac12 \sum_{\gamma} \frac{1}{\frac14+\gamma^2} = \frac12 \log\frac{q}{\pi} - \frac{\gamma_0}{2} - (1+\chi(-1))\frac{\log 2}{2} + \frac{L'(1,\chi)}{L(1,\chi)} \tag{3–28}$$

for a real primitive character χ mod q, where $\gamma_0 = 0.577215\ldots$ is Euler's constant. The values $L(1,\chi)$ can be calculated in closed form by classical formulas [Davenport 1980]; the values $L'(1,\chi)$ can be calculated in closed form using a formula of Selberg and Chowla [1967] for the odd characters and a formula of Deninger [1984] for the even characters. The former formula expresses $L'(1,\chi)$ in terms of the logarithm of the Γ-function, while the latter expresses $L'(1,\chi)$ in terms of a function $R(x)$ defined as

$$R(x) = \Big(\frac{\partial^2\zeta(s,x)}{\partial s^2} \Big)\Big|_{s=0};$$

here $\zeta(s,x)$ is the Hurwitz zeta function, defined when $x>-1$ by $\zeta(s,x) = \sum_{n=1}^{\infty}(n+x)^{-s}$ for $\mathrm{Re}\, s > 1$

and extended by meromorphic continuation to the complex s-plane.

Mathematica can calculate $\log\Gamma(x)$ and $R(x)$ to arbitrary precision, and thus by the formula (3–28) the sums $\sum_{\gamma>0} 1/\big(\frac14+\gamma^2\big)$ can also be so calculated. Table 2 contains the results of such calculations for the five characters relevant to the densities mod 8 and mod 12. (The third column actually lists the values of $L'(1,\chi)-(\gamma_0+\log 2\pi)L(1,\chi)$, to save space in the individual entries.)

For all five of these characters, when we choose $T = 10{,}000$ we find that $|b_1| < 0.000173$. The upper bound (3–27) can then be written more simply as $|\Delta_T(x,\chi)| \le D(x)$ for $|x| < 74$, where we've set

$$D(x) = \frac{1.5\times 10^{-8}x^4}{(1-0.00018x^2)^2}. \tag{3–29}$$

Consequently, the definition (3–26) of Err_3 implies

$$|\mathrm{Err}_3| \le \sum_{|m|,|n|\le 2C/\varepsilon}{}' \sum{}' \left|\frac{F_T(m\varepsilon,\chi_{-8})F_T((n-m)\varepsilon,\chi_{-4})F_T(-n\varepsilon,\chi_8)}{mn}\right| \times \big((1+D(m\varepsilon))(1+D((n-m)\varepsilon))(1+D(-n\varepsilon))-1\big). \tag{3–30}$$

The quantity on the right-hand side of this inequality was computed at the same time as $S_{8;3,5,7}(\varepsilon,C,T)$ was computed, and we obtained the bound

$$|\mathrm{Err}_3| < 5.5\times 10^{-6}. \tag{3–31}$$

3D. Conclusion

From the relationships (3–15), (3–16), and (3–25) among the various intermediate sums $S_{8;3,5,7}$, we have

$$I_{8;3,5,7} = 4S_{8;3,5,7}(\varepsilon,C,T) + \mathrm{Err}_1 + 4\mathrm{Err}_2 + 4\mathrm{Err}_3.$$

Using this identity in equation (2–51) yields

$$\begin{aligned}\delta_{8;3,5,7} &= \frac14 - \frac{1}{4\pi^2} I_{8;3,5,7}\\ &= \frac14 - \frac{1}{4\pi^2}\big(4S_{8;3,5,7}(\varepsilon,C,T)\\ &\qquad + \mathrm{Err}_1 + 4\mathrm{Err}_2 + 4\mathrm{Err}_3\big),\end{aligned}$$

whence it follows that

$$\left|\delta_{8;3,5,7} - \left(\frac14 - \frac{S_{8;3,5,7}(\varepsilon,C,T)}{\pi^2}\right)\right| \le \frac{|\mathrm{Err}_1|}{4\pi^2} + \frac{|\mathrm{Err}_2|+|\mathrm{Err}_3|}{\pi^2}.$$

Thus by the inequalities (3–12), (3–23), and (3–31), we conclude that

$$\left|\delta_{8;3,5,7} - \left(\frac14 - \frac{S_{8;3,5,7}(\varepsilon,C,T)}{\pi^2}\right)\right| < 8\times 10^{-7}$$

when $\varepsilon = \frac{1}{20}$, $C = 15$, and $T = 10{,}000$. Using these values for ε, C, and T, the sum $S_{8;3,5,7}(\varepsilon,C,T)$ was calculated and found to equal $0.5645285\ldots$, and therefore we have rigorously that

$$\delta_{8;3,5,7} = 0.1928013 \pm 9\times 10^{-7},$$

which is slightly stronger than the first assertion of Theorem 1.

The error analysis in Sections 3A–3C can be repeated for each of the densities in Theorem 1; the constants mentioned in the error analysis have been chosen to apply to all of these densities. Therefore, the densities calculated for Theorem 1 are all correct to within the same margin 9×10^{-7}, which is enough to establish the theorem.

3E. Appendix: Probability Bounds

In this section we establish the bounds (3–10) for $\Pr(X\ge u)$ and $\Pr(Y\ge u)$ which were used for the computations in Section 3A.

To do so, we first recall from Section 2A the explicit form of the random variables having the distribution $\mu_{q;a_1,\ldots,a_r}$. Specializing the representation (2–19) to the case $q=8$ and $\{a_1,a_2,a_3\}=\{3,5,7\}$, we find that $\mu_{8;3,5,7}$ is the distribution of the random $\mathbb{R}^3$-vector

$$(1,1,1) + X(\chi_{-8})(1,-1,-1) + X(\chi_{-4})(-1,1,-1) + X(\chi_8)(-1,-1,1).$$

Next, recalling the changes of variables (2–41) and (2–43) that took us from μ to ν and then to ρ, we observe that $\rho_{8;3,5,7}$ is the distribution of of the random $\mathbb{R}^2$-vector

$$X(\chi_{-8})(2,0) + X(\chi_{-4})(-2,2) + X(\chi_8)(0,-2). \tag{3–32}$$

Now define the two real-valued random variables

$$\begin{aligned} X &= 2\sum_{\substack{\gamma>0\\ L(\frac12+i\gamma,\chi_{-8})=0}} \alpha_\gamma \sin(2\pi U_\gamma) - 2\sum_{\substack{\gamma>0\\ L(\frac12+i\gamma,\chi_{-4})=0}} \alpha_\gamma \sin(2\pi U_\gamma),\\ Y &= 2\sum_{\substack{\gamma>0\\ L(\frac12+i\gamma,\chi_{-4})=0}} \alpha_\gamma \sin(2\pi U_\gamma) - 2\sum_{\substack{\gamma>0\\ L(\frac12+i\gamma,\chi_{8})=0}} \alpha_\gamma \sin(2\pi U_\gamma). \end{aligned} \tag{3–33}$$

χ	$L(1,\chi)$	$L'(1,\chi)-(\gamma_0+\log 2\pi)L(1,\chi)$	$\displaystyle\sum_{\gamma>0}\frac{1}{\frac14+\gamma^2}$
χ_{-8}	$\dfrac{\pi}{2\sqrt2}$	$\dfrac{\pi}{2\sqrt2}\log\dfrac{\Gamma(\frac58)\Gamma(\frac78)}{\Gamma(\frac18)\Gamma(\frac38)}$	0.158037
χ_{8}	$\dfrac{\log(1+\sqrt2)}{\sqrt2}$	$\dfrac{1}{2\sqrt2}\big(R(\frac18)-R(\frac38)-R(\frac58)+R(\frac78)\big)$	0.117716
χ_{-4}	$\dfrac{\pi}{4}$	$\dfrac{\pi}{2}\log\dfrac{\Gamma(\frac34)}{\Gamma(\frac14)}$	0.077784
χ_{-3}	$\dfrac{\pi}{3\sqrt3}$	$\dfrac{\pi}{\sqrt3}\log\dfrac{\Gamma(\frac23)}{\Gamma(\frac13)}$	0.056615
χ_{12}	$\dfrac{\log(2+\sqrt3)}{\sqrt3}$	$\dfrac{1}{2\sqrt3}\big(R(\frac1{12})-R(\frac5{12})-R(\frac7{12})+R(\frac{11}{12})\big)$	0.165083

TABLE 2. Values of $L(1,\chi)$, $L'(1,\chi)$, and $\sum_{\gamma>0}1/\big(\frac14+\gamma^2\big)$.

We see from the definition (2–18) of the $X(\chi)$ that the random vector (X,Y) equals the random vector (3–32).

The following lemma gives information about the tails of random variables of this type.

Lemma 3.1. *Let $r_1\ge r_2\ge\dots$ be a sequence of positive real numbers such that $\sum_{k=1}^\infty r_k=\infty$ but $\sum_{k=1}^\infty r_k^2=R<\infty$. Let U_1, U_2, … be independent random variables uniformly distributed on* $[0,1]$, *and define the random variable*

$$W=\sum_{k=1}^\infty r_k\sin(2\pi U_k).$$

Then, for any real number $w\ge 2r_1$,

$$\Pr(W\ge w)\le\exp\Big(\frac{-3(w-2r_1)^2}{16R}\Big).$$

Proof. By [Montgomery 1980, Section 3, Theorem 1], under the assumptions of this lemma, we have

$$\Pr\Big(W\ge 2\sum_{k=1}^K r_k\Big)\le\exp\Big(-\frac34\Big(\sum_{k=1}^K r_k\Big)^2\Big/\sum_{k>K}r_k^2\Big) \tag{3–34}$$

for any integer $K\ge1$. Since the r_k are decreasing and $\sum_{k=1}^\infty r_k=\infty$, it is clear that for any $w\ge 2r_1$ there exists a $K\ge1$ such that

$$\frac w2-r_1\le\sum_{k=1}^K r_k\le\frac w2.$$

With this choice of K, inequality (3–34) simplifies to

$$\begin{aligned}\Pr(W\ge w)&\le\exp\Big(-\frac34\Big(\frac w2-r_1\Big)^2\Big/\sum_{k=K}^\infty r_k^2\Big)\\&\le\exp\Big(\frac{-3(w-2r_1)^2}{16R}\Big).\end{aligned}$$ □

We now apply this lemma to the random variables X and Y defined in equation (3–33). (Note that because each variable U_γ is uniformly distributed on $[0,1]$, we may replace the U_γ in the second sum on each line with $U_\gamma+\frac12$; this has the effect of changing the subtraction signs in the equations (3–33) to addition signs, thus rendering X and Y into the form to which Lemma 3.1 applies.) For the variable X, the sequence corresponding to r_k is

$$\begin{aligned}&\big\{2\alpha_\gamma : L(\tfrac12+i\gamma,\chi_{-8})=0,\ \gamma>0\big\}\\&\qquad\cup\big\{2\alpha_\gamma : L(\tfrac12+i\gamma,\chi_{-4})=0,\ \gamma>0\big\}.\end{aligned}$$

For this sequence, the largest element r_1 is less than 1.5, and the sum R of the squares of the elements does not exceed 4.5. Therefore an application of Lemma 3.1 gives

$$\Pr(X\ge u)\le\exp(-0.04(u-3)^2)$$

for any $u\ge3$. One shows similarly that Y satisfies the same estimate, establishing the upper bounds (3–10). In fact, the constants mentioned above will work for every pair of characters that arises in the computations of $\rho_{8;a_1,a_2,a_3}$, where $\{a_1,a_2,a_3\}$ is a permutation of $\{3,5,7\}$, and in $\rho_{12;a_1,a_2,a_3}$, where $\{a_1,a_2,a_3\}$ is a permutation of $\{5,7,11\}$.

4. COMPUTATIONAL RESULTS

The mathematical and numerical computations described in this paper were implemented on an SGI Challenge computer using Mathematica, which has the capability to perform computations to arbitrary and verifiable precision. A typical quantity to be calculated is the expression $S_{8;3,5,7}(\varepsilon, C, T)$ defined in equation (3–2), which depends on the functions $F_T(z, \chi)$ defined in equation (2–11). In order to compute these functions we needed, for the Dirichlet L-functions corresponding to characters to the moduli $q \le 12$, lists of the zeros whose imaginary parts are bounded by $T = 10{,}000$. These lists of imaginary parts of zeros (accurate to twelve decimal places) were kindly supplied to us by R. Rumely [1993]. For the estimation of Err_3 in Section 3C it was also necessary to compute quantities typified by the right-hand side of equation (3–30), which is no harder than computing $S_{8;3,5,7}(\varepsilon, C, T)$ itself.

In addition to the results reported in Theorem 1, further computations were carried out involving certain cases with $q \le 12$ and $r \le 4$. In these additional results, which are presented below, we report only the numbers of decimal places in which we have some degree of confidence; specifically, we expect the entries to be correct to within one or two units in the last decimal place reported.

Table 3 shows the calculated densities $\delta_{q;a_1,a_2}$ for the two-way races between $\pi(x; q, a_1)$ and $\pi(x; q, a_2)$, for the moduli $q = 3$, 4, and 5. For example, the first line of the table indicates that $\delta_{3;2,1} = 0.9990633$ (rounded to seven decimal places). Throughout this section we use the symbol N to stand for any nonsquare mod q and S to stand for any square mod q (although distinct occurrences of N or S in a single entry stand for distinct residues) to make the Chebyshev biases more clearly evident where appropriate.

q	$a_1 a_2$	$\delta_{q;a_1,a_2}$
3	NS: 21	0.9990633
	SN: 12	0.0009367
4	NS: 31	0.9959280
	SN: 13	0.0040720
5	NS: 21, 24, 31, 34	0.952140
	NN: 23, 32 SS: 14, 41	$\frac{1}{2}$
	SN: 12, 13, 42, 43	0.047860

TABLE 3. Two-way races for the moduli $q = 3, 4, 5$. N and S stand for nonsquare and square (mod q).

Of course, since $\varphi(3) = \varphi(4) = 2$, the two-way races shown are the only possible races for the moduli 3 and 4. The densities for these moduli were calculated by Rubinstein and Sarnak, and our calculations agree with theirs to six decimal places. (Although they were only reported in [Rubinstein and Sarnak 1994] truncated to four decimal places, they had in fact been calculated to higher accuracy.)

For the races modulo 5, it turns out that the densities $\delta_{q;a_1,a_2}$ depend only on whether or not a_1 and a_2 are squares mod 5, due to the symmetry results given in Theorem 2. (In fact this is true for the races between multiple residues mod 5 as well.) For instance, applying Theorem 2(b) with $a_1 = 2$, $a_2 = 1$, and $b = 4$ shows that $\delta_{5;2,1} = \delta_{5;3,4}$; then applying Theorem 2(a) to each of these expressions shows further that $\delta_{5;2,1} = \delta_{5;3,1}$ and $\delta_{5;3,4} = \delta_{5;2,4}$. Since the two nonsquares mod 5 are $\{2, 3\}$ while the two squares are $\{1, 4\}$, these equalities show that all four densities represented by $\delta_{5;N,S}$ are equal, as indicated in Table 3.

The fact that $\delta_{q;N,N} = \delta_{q;S,S} = \frac{1}{2}$, as shown in the penultimate entry of the table, was proved by Rubinstein and Sarnak, and it also follows from our Theorem 2(d). We calculated these densities anyway, and the calculated answers differed from $\frac{1}{2}$ by at most 10^{-16}, which is the default machine precision for our Mathematica calculations. This degree of accuracy is not unexpected in this instance, as the integral in the formula (2–57) is identically zero when a_1 and a_2 are both squares or both nonsquares mod q.

Table 4 provides the calculated densities $\delta_{q;a_1,a_2,a_3}$ for the three-way races modulo 5. Again, in this case the densities only depend on whether a_1, a_2, and a_3

$a_1 a_2 a_3$	$\delta_{5;a_1,a_2,a_3}$
NNS: 231, 234, 321, 324 NSS: 214, 241, 314, 341	0.45678
NSN: 213, 243, 312, 342 SNS: 124, 134, 421, 431	0.03859
SNN: 123, 132, 423, 432 SSN: 142, 143, 412, 413	0.00464

TABLE 4. Three-way races modulo $q = 5$.

are squares mod 5, by the symmetry results (a) and (b) of Theorem 2. In addition, each density matches two different types of permutations: for instance, Theorem 2(e) with $a_1 = 2$, $a_2 = 3$, $a_3 = 1$, and $b = 2$ asserts that $\delta_{5;2,3,1} = \delta_{5;2,1,4}$, as indicated in the first entry of the table.

As mentioned at the beginning of this section, we are confident from numerical considerations that the numbers reported in Table 4 are accurate to the five decimal places given there, with a possible error of one or two units in the fifth decimal place. Thus, for instance, if we choose a particular triple of residues such as $\{1, 2, 3\}$ and add up the densities from Table 4 corresponding to the six permutations of that triple, the result is 1.00002. Moreover, the three ordered triples $\{3, 2, 1\}$, $\{2, 3, 1\}$, and $\{2, 1, 3\}$ are the three permutations in which 2 is ahead of 1, and so we have the identity

$$\delta_{5;2,1} = \delta_{5;3,2,1} + \delta_{5;2,3,1} + \delta_{5;2,1,3}; \qquad (4\text{–}1)$$

compare equation (5–1). Table 3 gives 0.952140 for the left-hand side of this identity, while adding the appropriate entries from Table 4 gives 0.95215 for the right-hand side.

There are two reasons why our calculations of the densities in three-way races for moduli other than 8 and 12 are less accurate than the full six-decimal-place accuracy proven in Theorem 1, both stemming from the fact that there are complex-valued Dirichlet characters associated with the other moduli. First, when we calculate the function $F_T(z, \chi)$ we do so only on a discrete set of points, evenly spaced at intervals of $\varepsilon/2$. These points are the only ones needed to evaluate sums such as $S_{8;3,5,7}(\varepsilon, C, T)$, as we see from the definition (3–2), but for the sums corresponding to other moduli we need to know the value of $F_T(z, \chi)$ at irrational multiples of ε. We estimated this value by interpolating linearly between the two nearest values, and this estimation introduces an additional error into the calculations.

Second, the zeros of L-functions corresponding to complex characters are not symmetric with respect to the real axis, and so the quantity $\sum_{\gamma>0} 1/(\frac14+\gamma^2)$, needed to compute $b_1(T, \chi)$, cannot be evaluated in closed form. Since we can evaluate $b_1(T, \chi) + b_1(T, \bar\chi)$ in closed form, we used half of this quantity in place of both $b_1(T, \chi)$ and $b_1(T, \bar\chi)$; this gives the correct first-order approximation to the tail of $F(z, \chi)F(z, \bar\chi)$, but the absolute error in our calculations can be somewhat higher as a result. For higher moduli, the sheer number of characters will also play a role, as the product of the $\varphi(q) - 1$ functions $F_T(z, \chi)$ required for the evaluation of $\hat\rho_{q;a_1,\dots,a_r}$ will gradually erode the accuracy of the calculated number.

Since there are exactly four reduced residues modulo 5, it is natural to look at the complete four-way race mod 5; Table 5 shows the calculated densities for this four-way race. Here again, the densities only depend on whether a_1, a_2, a_3, and a_4 are squares mod 5, by the symmetry results from parts (a) and (b) of Theorem 2, with the added symmetry between the densities in the third and fourth lines of the table following from Theorem 2(e). Once again we can estimate the accuracy of these densities by comparing the sum of all twenty-four densities to 1, and also by comparing the values here to those in Table 4 using identities such as

$$\delta_{5;1,2,3} = \delta_{5;4,1,2,3} + \delta_{5;1,4,2,3} + \delta_{5;1,2,4,3} + \delta_{5;1,2,3,4}.$$

In all cases, these sums of densities from Table 5 are precise to within a few units in the fifth decimal place.

In the calculation of these four-way densities, the general formula given in Theorem 4 involves a three-dimensional integral which must be computed numerically. Performing this calculation with a reasonable degree of accuracy lies at the limit of the computing capabilities of the method used for the calculations in this paper; in particular, we found it necessary to reduce the value of C and increase the value of ε somewhat to make the computations feasible.

Since the distribution of the primes into residue classes modulo 6 is fully determined by their distribution mod 3, the next modulus of interest is $q = 7$. Table 6 shows the calculated densities $\delta_{7;a_1,a_2}$ for the two-way races modulo 7. Here, for the first time, we see that the density does not depend merely on whether a_1 and a_2 are squares mod 7: the squares mod 7 are $\{1, 2, 4\}$, so the first and third lines of the table list densities of the form $\delta_{7;N,S}$, while the second and fourth lines list densities of the form $\delta_{7;S,N}$. In other words, Chebyshev's bias is not the only factor causing asymmetries in the Shanks–Rényi race games. (For a somewhat more precise discussion of

$a_1a_2a_3a_4$	$\delta_{5;a_1,a_2,a_3}$
$NNSS$: 2314, 2341, 3214, 3241	0.21136
$NSNS$: 2134, 2431, 3124, 3421	0.02985
$NSSN$: 2143, 2413, 3142, 3412	0.00424
$SNNS$: 1234, 1324, 4231, 4321	0.00424
$SNSN$: 1243, 1342, 4213, 4312	0.00028
$SSNN$: 1423, 1432, 4123, 4132	0.00007

TABLE 5. The full four-way race modulo $q = 5$.

a_1a_2	$\delta_{7;a_1,a_2}$
31; 51; 32; 62; 54; 64	0.874349
13; 15; 23; 26; 45; 46	0.125651
61; 52; 34	0.845210
16; 25; 43	0.154790
$12, 21; 14, 41; 24, 42; 35, 53; 36, 63; 56, 65$	$\frac{1}{2}$

TABLE 6. Two-way races modulo $q = 7$.

Chebyshev biases for r-tuples with $r \geq 3$, see the discussion of bias factors in Section 6.) The bottom row of the table again indicates the known fact that all densities of the form $\delta_{7;N,N}$ and $\delta_{7;S,S}$ equal $\frac{1}{2}$.

Table 7 gives the calculated densities for the three-way races modulo 7. Because the number of different values for the densities in Tables 6 and 7 is larger than in the previous cases, we have not organized them strictly by decreasing size, but rather we have grouped together the values corresponding to *isomorphic race games*. We will say that two r-tuples $\{a_1, \ldots, a_r\}$ and $\{b_1, \ldots, b_r\}$ of reduced residue classes mod q have isomorphic race games if there exists a bijection τ from the set $\{1, \ldots, n\}$ to itself such that each residue a_j acts exactly like the corresponding residue $b_{\tau(j)}$, that is, if

$$\delta_{q;a_{\sigma(1)},\ldots,a_{\sigma(r)}} = \delta_{q;b_{\tau(\sigma(1))},\ldots,b_{\tau(\sigma(r))}}$$

for any permutation σ of $\{1, \ldots, n\}$.

For instance, Theorem 2(a) tells us that $\delta_{7;1,2,5} = \delta_{7;1,3,4}$ and similarly for the corresponding permutations of $\{1, 2, 5\}$ and $\{1, 3, 4\}$. Therefore the bijection $\tau : \{1, 2, 5\} \to \{1, 3, 4\}$ given by

$$\tau(a) \equiv a^{-1} \pmod 7$$

shows that these triples have isomorphic race games. Table 7 shows that there are ten triples whose race games are in the isomorphism class determined by $\{1, 2, 5\}$; the six densities for the race games in this class are all distinct. In addition, there are five triples in the isomorphism class of $\{1, 2, 3\}$; the race games in this class have only three distinct densities due to an internal symmetry generated by Theorem 2(a). Finally, the two special triples $\{S, S, S\} = \{1, 2, 4\}$ and $\{N, N, N\} = \{3, 5, 6\}$ each give completely symmetric race games; this is the smallest modulus to which parts (d) and (e) of Theorem 2 can be applied, since three distinct squares or non-squares are needed. The complete symmetry for these two race games was also proven by Rubinstein and Sarnak. We remark that our computations of these densities yielded $\frac{1}{6}$ to five decimal places. We did not proceed further with computations modulo 7, since there is no natural four-way race and races with five or more residues are beyond the present capabilities of our computing set-up.

Table 8 shows the calculated densities for the two-way races modulo 8. Because only one fourth of the residues mod 8 are squares (that is, $c(8, 1) = 3$), in contrast to the lower moduli, there are fewer

$a_1a_2a_3$	$\delta_{7;a_1,a_2,a_3}$
512; 314; 631; 651; 621; 324; 532; 562; 641; 542; 354; 364	0.4038
521; 341; 361; 561; 612; 342; 352; 652; 614; 524; 534; 634	0.3678
251; 431; 316; 516; 162; 432; 325; 625; 164; 254; 543; 643	0.1027
152; 134; 613; 615; 261; 234; 523; 526; 461; 452; 345; 346	0.0736
215; 413; 136; 156; 126; 423; 235; 265; 146; 245; 453; 463	0.0295
125; 143; 163; 165; 216; 243; 253; 256; 416; 425; 435; 436	0.0226
312, 321; 351, 531; 514, 541; 362, 632; 624, 642; 564, 654	0.3943
132, 231; 315, 513; 154, 451; 326, 623; 264, 462; 546, 645	0.0857
123, 213; 135, 153; 145, 415; 236, 263; 246, 426; 456, 465	0.0200
124, 142, 214, 241, 412, 421; 356, 365, 536, 563, 635, 653	$\frac{1}{6}$

TABLE 7. Three-way races modulo $q = 7$.

a_1a_2	$\delta_{8;a_1,a_2}$
31	0.9995688
13	0.0004312
51	0.9973946
15	0.0026054
71	0.9989378
17	0.0010622
35, 53; 37, 73; 57, 75	$\frac{1}{2}$

TABLE 8. Two-way races modulo $q = 8$.

symmetries among the densities. (This is somewhat counterintuitive, since the multiplicative group modulo 8 is highly symmetric.) This higher value of $c(8,1)$ also causes a larger bias towards nonsquares, as can be seen by the fact that the values in Table 8 are more extreme than those in Tables 3 and 6.

Table 9 shows the densities we calculated for the three-way races modulo 8, including the values for $\delta_{8;N,N,N}$ highlighted in Theorem 1. Since all of the characters mod 8 are real, the additional sources of computational error mentioned in the discussion of Table 4 are not present here, and so we feel justified in reporting these figures to seven decimal places; in fact note that the appropriate three-way densities sum to the two-way densities in Table 8 in a manner analogous to equation (4–1), with the sums all agreeing to within one or two units in the seventh decimal place.

As with the modulus 5, it is natural to look at the complete four-way race modulo 8; Table 10 shows the calculated densities for this four-way race, listed in the lexicographical ordering on the permutations of $\{1,3,5,7\}$. Despite the need to use slightly cruder values of C and ε in the calculations of the three-dimensional integrals that arise in the formulas for these densities, the sum of all 24 densities and numerical checks against Table 9 suggest that these densities are also accurate to within one or two units in the seventh decimal place.

Tables 11 and 12 show the calculated densities for the two-way and three-way races modulo 9. Since the multiplicative group mod 9 is isomorphic to the multiplicative group mod 7 (both are cyclic of order 6), the various symmetries present in Tables 11 and 12 mirror those found in Tables 6 and 7, with the squares mod 9 being $\{1,4,7\}$.

Again, the distribution of the primes into residue classes modulo 10 is determined by their distribution

a_1a_2	$\delta_{9;a_1,a_2}$
21; 51; 24; 84; 57; 87	0.881584
12; 15; 42; 48; 75; 78	0.118416
81; 27; 54	0.864230
18; 72; 45	0.135770
14, 41; 17, 71; 25, 52; 28, 82; 47, 74; 58, 85	$\frac{1}{2}$

TABLE 11. Two-way races modulo $q = 9$.

$a_1a_2a_3$	$\delta_{8;a_1,a_2,a_3}$	$a_1a_2a_3$	$\delta_{8;a_1,a_2,a_3}$	$a_1a_2a_3$	$\delta_{8;a_1,a_2,a_3}$	$a_1a_2a_3$	$\delta_{8;a_1,a_2,a_3}$
531	0.4996015	731	0.4995765	571	0.4990135	357	0.1928013
351	0.4974123	371	0.4989440	751	0.4974474	753	0.1928013
315	0.0025550	317	0.0010483	715	0.0024769	375	0.1664263
513	0.0003808	713	0.0004173	517	0.0009337	573	0.1664263
135	0.0000327	137	0.0000077	175	0.0000757	735	0.1407724
153	0.0000177	173	0.0000062	157	0.0000528	537	0.1407724

TABLE 9. The four three-way races modulo $q = 8$.

$a_1a_2a_3a_4$	$\delta_{8;a_1,a_2,a_3,a_4}$	$a_1a_2a_3a_4$	$\delta_{8;a_1,a_2,a_3,a_4}$	$a_1a_2a_3a_4$	$\delta_{8;a_1,a_2,a_3,a_4}$	$a_1a_2a_3a_4$	$\delta_{8;a_1,a_2,a_3,a_4}$
1357	0.0000014	3157	0.0000500	5137	0.0000027	7135	0.0000261
1375	0.0000029	3175	0.0000696	5173	0.0000023	7153	0.0000154
1537	0.0000007	3517	0.0007972	5317	0.0001315	7315	0.0008983
1573	0.0000006	3571	0.1919526	5371	0.1406374	7351	0.1398456
1735	0.0000023	3715	0.0015371	5713	0.0000848	7513	0.0002910
1753	0.0000009	3751	0.1648170	5731	0.1663386	7531	0.1924939

TABLE 10. The full four-way race modulo $q = 8$.

$a_1a_2a_3$	$\delta_{9;a_1,a_2,a_3}$
514; 217; 821; 851; 841; 247; 524; 584; 871; 574; 257; 287	0.4010
541; 271; 281; 581; 814; 274; 254; 854; 817; 547; 527; 827	0.3814
451; 721; 218; 518; 184; 724; 245; 845; 187; 457; 572; 872	0.0992
154; 127; 812; 815; 481; 427; 542; 548; 781; 754; 275; 278	0.0819
415; 712; 128; 158; 148; 742; 425; 485; 178; 475; 752; 782	0.0194
145; 172; 182; 185; 418; 472; 452; 458; 718; 745; 725; 728	0.0172
214, 241; 517, 571; 251, 521; 284, 824; 847, 874; 587, 857	0.3965
124, 421; 157, 751; 215, 512; 248, 842; 487, 784; 578, 875	0.0885
142, 412; 175, 715; 125, 152; 428, 482; 478, 748; 758, 785	0.0149
147, 174, 417, 471, 714, 741; 258, 285, 528, 582, 825, 852	$\frac{1}{6}$

TABLE 12. Three-way races modulo $q = 9$.

mod 5, so the next modulus of interest is $q = 11$. In Table 13 we show the calculated densities for the two-way races modulo 11, the symbol T representing the residue 10 mod 11.

We do not include the calculations of the three-way races mod 11 for reasons of space. Using Theorem 2 it can be checked that of the 120 distinct (unordered) triples of residues mod 11, the twenty triples of the form $\{ab^{-1}, a, ab\}$ with $b \equiv 3$ or $b \equiv 5 \pmod{11}$, in which ab^{-1}, a, and ab are all nonsquares or all squares mod 11, comprise two isomorphism classes of race games of ten triples each; a race game in either of these isomorphism classes has only two distinct densities, one taken by four permutations of the triple and the other taken by the other two permutations. The twenty triples of the form $\{ab^{-1}, a, ab\}$ with

$$b \equiv 2 \text{ or } b \equiv 7 \pmod{11},$$

in which ab^{-1} and ab have the opposite quadratic character mod 11 from a, also form two isomorphism classes with ten triples in each class; a race game in one of these classes has three distinct densities. Finally, the remaining eighty triples form four isomorphism classes of twenty race games each; a race game in one of these classes has all six densities distinct. There are 34 densities that remain to be calculated after these symmetries from Theorem 2 are taken into account, and the calculations reveal that these 34 densities are indeed distinct.

As mentioned previously, determining the densities in a five-way race game lies beyond the scope of the computing methods used for the calculations in this paper (though this barrier is only technological, as Theorem 4 is valid for arbitrarily large race games). If this barrier were overcome (for example, by recoding in a lower level computing language), the five-way race among the squares mod 11 and the five-way race among the nonsquares mod 11 would be natural and interesting questions to consider, especially in light of the nearly-cyclic behavior of the leaders in these five-way race games reported by Bays and Hudson [1983]. Because of the symmetries of Theorem 2, it turns out that only eight distinct densities would need to be calculated for both of these five-way race games to be completely determined.

Tables 14–16 show the two-way, three-way, and four-way race games modulo 12, using the symbol E to represent the residue 11 mod 12. Since the multiplicative group mod 12 is isomorphic to the multiplicative group mod 8 (both groups being isomorphic to the Klein group of order 4), the various symmetries present in Tables 14–16 mirror those found in Tables 8–10. As with the modulus 8 case,

a_1a_2	$\delta_{11;a_1,a_2}$
71; 81; 23; 25; T3; 64; T4; 75; 69; 89	0.761121
17; 18; 32; 52; 3T; 46; 4T; 57; 96; 98	0.238879
21; 61; 24; 63; 73; 84; 85; T5; 79; T9	0.731135
12; 16; 42; 36; 37; 48; 58; 5T; 97; 9T	0.268865
T1; 29; 83; 74; 65	0.713943
1T; 92; 38; 47; 56	0.286057
NN; SS	$\frac{1}{2}$

TABLE 13. Two-way races modulo $q = 11$. The entry "NN; SS" refers to the forty pairs $\{a_1, a_2\}$ where a_1 and a_2 are either both among the nonsquares $\{2, 6, 7, 8, \mathrm{T}\}$ or both among the squares $\{1, 3, 4, 5, 9\}$ mod 11.

a_1a_2	$\delta_{12;a_1,a_2}$
51	0.9992059
15	0.0007941
71	0.9986061
17	0.0013939
E1	0.9999766
1E	0.0000234
57, 75; 5E, E5; 7E, E7	$\frac{1}{2}$

TABLE 14. Two-way races modulo $q = 12$.

all the characters mod 12 are real-valued, and so we feel justified in reporting seven decimal places of the numbers in these tables.

Notice from Table 15 that the densities $\delta_{12;5,11,1}$ and $\delta_{12;7,11,1}$ only differ by one unit in the sixth decimal place, and that there are several other entries that differ by similarly small amounts owing to their small sizes. Nevertheless, we see no reason to believe that any of the twenty-one densities in Table 15 is equal to any other. Similar remarks hold for the twenty-four densities in Table 16 and for the corresponding Tables 9 and 10 for the race games modulo 8. One observation supporting our view is that whenever the symmetries of Theorem 2 imply that two densities are equal, the computed densities agree to within a few multiples of the default machine precision rather than to only five or six decimal places.

5. EQUALITIES AND INEQUALITIES BETWEEN DENSITIES

We will now establish Theorem 2, concerning symmetries of the densities $\delta_{q;a_1,\dots,a_r}$ under certain permutations of the residue classes $\{a_1,\dots,a_r\}$, and Theorem 3, giving some strict inequalities in the same setting. We first present the proof of Theorem 3 since it is simpler than that of Theorem 2.

Proof of Theorem 3. Let a_1, a_2, and a_3 be distinct reduced residue classes mod q. We begin with the simple observation that if x is a real number such that $\pi(x;q,a_1) > \pi(x;q,a_2)$, then the quantity $\pi(x;q,a_3)$ must either equal one of $\pi(x;q,a_1)$ and $\pi(x;q,a_2)$, lie between them, exceed both, or be exceeded by both. This observation leads to the density identity

$$\delta_{q;a_1,a_2} = \delta_{q;a_3,a_1,a_2} + \delta_{q;a_1,a_3,a_2} + \delta_{q;a_1,a_2,a_3}, \quad (5\text{–}1)$$

since the set of real numbers x such that

$$\pi(x;q,a_3) = \pi(x;q,a_1) \quad \text{or} \quad \pi(x;q,a_3) = \pi(x;q,a_2)$$

has density zero, as mentioned in Section 2A. It follows that

$$\delta_{q;a_1,a_2,a_3} - \delta_{q;a_3,a_2,a_1} = \delta_{q;a_1,a_2} - \delta_{q;a_3,a_2}, \quad (5\text{–}2)$$

by using the appropriate identity of the type (5–1) on both terms on the right-hand side of (5–2) and simplifying.

Now we can use our knowledge of the two-way densities to study the difference on the left-hand

$a_1a_2a_3$	$\delta_{12;a_1,a_2,a_3}$	$a_1a_2a_3$	$\delta_{12;a_1,a_2,a_3}$	$a_1a_2a_3$	$\delta_{12;a_1,a_2,a_3}$	$a_1a_2a_3$	$\delta_{12;a_1,a_2,a_3}$
751	0.4992728	5E1	0.4999772	7E1	0.4999780	57E	0.1984521
571	0.4986582	E51	0.4992062	E71	0.4986066	E75	0.1984521
517	0.0012750	E15	0.0007931	E17	0.0013919	E57	0.1799849
715	0.0006751	51E	0.0000225	71E	0.0000214	75E	0.1799849
157	0.0000668	1E5	0.0000006	1E7	0.0000015	5E7	0.1215630
175	0.0000521	15E	0.0000003	17E	0.0000006	7E5	0.1215630

TABLE 15. The four three-way races modulo $q = 12$.

$a_1a_2a_3a_4$	$\delta_{12;a_1,a_2,a_3,a_4}$	$a_1a_2a_3a_4$	$\delta_{12;a_1,a_2,a_3,a_4}$	$a_1a_2a_3a_4$	$\delta_{12;a_1,a_2,a_3,a_4}$	$a_1a_2a_3a_4$	$\delta_{12;a_1,a_2,a_3,a_4}$
157E	$< 10^{-7}$	517E	0.0000004	715E	0.0000001	E157	0.0000664
15E7	0.0000001	51E7	0.0000010	71E5	0.0000002	E175	0.0000519
175E	$< 10^{-7}$	571E	0.0000152	751E	0.0000059	E517	0.0011332
17E5	$< 10^{-7}$	57E1	0.1984364	75E1	0.1799788	E571	0.1787850
1E57	0.0000002	5E17	0.0001403	7E15	0.0000243	E715	0.0006505
1E75	0.0000001	5E71	0.1214216	7E51	0.1215384	E751	0.1977496

TABLE 16. The full four-way race modulo $q = 12$.

side of (5–2). In particular, if $c(q,a_1) = c(q,a_2)$ then $\delta_{q;a_1,a_2} = \frac{1}{2}$, and hence $\delta_{q;a_1,a_2,a_3} - \delta_{q;a_3,a_2,a_1} = \frac{1}{2} - \delta_{q;a_3,a_2}$, an expression whose sign is known from the work of Rubinstein and Sarnak. More specifically, if N and N' are nonsquares mod q while S is a square mod q, then $\delta_{q;N,N',S} - \delta_{q;S,N',N} = \frac{1}{2} - \delta_{q;S,N'} > 0$; therefore $\delta_{q;N,N',S} > \delta_{q;S,N',N}$, which establishes part (a) of the theorem. Similarly, if N is a nonsquare mod q while S and S' are squares mod q, then $\delta_{q;S',S,N} < \delta_{q;N,S,S'}$, which establishes part (b) of the theorem.

Another application is to the difference $\delta_{q;N,S,N'} - \delta_{q;N',S,N}$ when N and N' are nonsquares mod q while S is a square mod q. In this case equation (5–2) becomes

$$\delta_{q;N,S,N'} - \delta_{q;N',S,N} = \delta_{q;N,S} - \delta_{q;N',S},$$

which immediately implies part (c) of the theorem. The analogous observation about the difference

$$\delta_{q;S,N,S'} - \delta_{q;S',N,S}$$

when S and S' are squares mod q while N is a nonsquare mod q establishes part (d) of the theorem. □

We remark that the identity (5–2), applied when a_1, a_2, and a_3 are all nonsquares mod q, becomes $\delta_{q;a_1,a_2,a_3} - \delta_{q;a_3,a_2,a_1} = 0$; this is another way of seeing that the densities calculated in Theorem 1 are equal in pairs as indicated.

Our next goal is to establish Theorem 2. Before doing so, it will be helpful to recall the relationships between the density $\delta_{q;a_1,\dots,a_r}$ and the measures $\mu_{q;a_1,\dots,a_r}$ and $\rho_{q;a_1,\dots,a_r}$. We begin by recalling from equation (2–5) that

$$\delta_{q;a_1,\dots,a_r} = \int_{x_1>\cdots>x_r}\!\!\cdots\!\int d\mu_{q;a_1,\dots,a_r}\,. \tag{5–3}$$

We remark that if σ is a permutation of the indices $\{1,\dots,r\}$, then we can express the density $\delta_{q;a_{\sigma(1)},\dots,a_{\sigma(r)}}$ in two different ways: we have

$$\delta_{q;a_{\sigma(1)},\dots,a_{\sigma(r)}} = \int_{x_1>\cdots>x_r}\!\!\dots\!\int d\mu_{q;a_{\sigma(1)},\dots,a_{\sigma(r)}}$$

corresponding to the formula (5–3), but we also have the alternate form

$$\delta_{q;a_{\sigma(1)},\dots,a_{\sigma(r)}} = \int_{x_{\sigma(1)}>\cdots>x_{\sigma(r)}}\!\!\dots\!\int d\mu_{q;a_1,\dots,a_r}$$

since $\mu_{q;a_1,\dots,a_r}$ is the limiting distribution of the vector $(E(x;q,a_1),\dots,E(x;q,a_r))$, whose coordinates are ordered by size exactly as the coordinates of the vector $(\pi(x;q,a_1),\dots,\pi(x;q,a_r))$.

If we make the change of variables $u_1 = x_1 - x_2$, $\dots$, $u_{r-1} = x_{r-1} - x_r$, $u_r = x_r$ and integrate out the variable u_r, as in Section 2E, the formula (5–3) becomes

$$\delta_{q;a_1,\dots,a_r} = \int_{u_1>0,\,\dots,\,u_{r-1}>0}\!\!\cdots\!\int d\rho_{q;a_1,\dots,a_r}. \tag{5–4}$$

For the special permutation σ that reverses the set $\{1,\dots,n\}$, we see that

$$\begin{aligned} x_{\sigma(1)} > \cdots > x_{\sigma(r)} &\iff x_r > \cdots > x_1 \\ &\iff u_{r-1} < 0,\ \dots,\ u_1 < 0. \end{aligned}$$

Consequently, we have

$$\delta_{q;a_r,\dots,a_1} = \int_{u_1<0,\,\dots,\,u_{r-1}<0}\!\!\cdots\!\int d\rho_{q;a_1,\dots,a_r} \tag{5–5}$$

as a companion formula to equation (5–4).

As a final prerequisite to the proof of Theorem 2, we recall from equation (2–13) the explicit formula

$$\begin{aligned} &\hat\mu_{q;a_1,\dots,a_r}(\xi_1,\dots,\xi_r) \\ &= \exp\biggl(i\sum_{j=1}^r c(q,a_j)\xi_j\biggr) \prod_{\substack{\chi \bmod q\\ \chi\ne\chi_0}} F\biggl(\biggl|\sum_{j=1}^r \chi(a_j)\xi_j\biggr|,\chi\biggr) \end{aligned} \tag{5–6}$$

for the Fourier transform of $\mu_{q;a_1,\dots,a_r}$, and the related formula (2–21)

$$\begin{aligned} &\hat\rho_{q;a_1,\dots,a_r}(\eta_1,\dots,\eta_{r-1}) \\ &= \exp\biggl(\sum_{j=1}^{r-1}(c(q,a_j) - c(q,a_{j+1}))\eta_j\biggr) \\ &\quad\times \prod_{\substack{\chi \bmod q\\ \chi\ne\chi_0}} F\biggl(\biggl|\sum_{j=1}^{r-1}(\chi(a_j)-\chi(a_{j-1}))\eta_j\biggr|,\ \chi\biggr) \end{aligned} \tag{5–7}$$

for the Fourier transform of $\rho_{q;a_1,\dots,a_r}$.

Proof of Theorem 2. Let a_j^{-1} denote the inverse of a_j mod q. We will show that the Fourier transforms $\hat\mu_{q;a_1,\dots,a_r}$ and $\hat\mu_{q;a_1^{-1},\dots,a_r^{-1}}$ are the same function. This is enough to establish part (a), since the densities $\mu_{q;a_1,\dots,a_r}$ and $\mu_{q;a_1^{-1},\dots,a_r^{-1}}$ will then be identical, which by equation (5–3) will imply

$$\delta_{q;a_1,\dots,a_r} = \underset{x_1>x_2>\cdots>x_r}{\int\cdots\int} d\mu_{q;a_1,\dots,a_r} = \underset{x_1>x_2>\cdots>x_r}{\int\cdots\int} d\mu_{q;a_1^{-1},\dots,a_r^{-1}} = \delta_{q;a_1^{-1},\dots,a_r^{-1}}.$$

We use the formula (5–6) for $\hat\mu_{q;a_1,\dots,a_r}$ and the analogous formula for $\hat\mu_{q;a_1^{-1},\dots,a_j^{-1}}$. Notice that the square roots of a_j^{-1} are precisely the inverses mod q of the square roots of a_j. In particular, $c(q,a_j^{-1}) = c(q,a_j)$, and so the exponential term in (5–6) is unchanged if we replace each a_j by a_j^{-1}. Moreover, we see that for each character χ mod q,

$$\left|\sum_{j=1}^r \chi(a_j^{-1})\xi_j\right| = \left|\sum_{j=1}^r \overline{\chi(a_j)}\xi_j\right| = \left|\sum_{j=1}^r \chi(a_j)\xi_j\right|$$

since the ξ_j are real, so that each term $F(\cdot,\chi)$ in (5–6) is also unchanged by replacing all of the a_j with the a_j^{-1}. This shows that

$$\hat\mu_{q;a_1,\dots,a_r} = \hat\mu_{q;a_1^{-1},\dots,a_r^{-1}},$$

which establishes part (a) of the theorem.

We use a similar strategy to prove part (b). Let b be a reduced residue class mod q such that $c(q,a_j) = c(q,ba_j)$ for each $1\le j\le r$. Because of this hypothesis, the exponential term in the formula (5–6) is unchanged if we replace each a_j by ba_j as above. Moreover, for each character χ mod q,

$$\begin{aligned}\left|\sum_{j=1}^r \chi(ba_j)\xi_j\right| &= \left|\chi(b)\sum_{j=1}^r \chi(a_j)\xi_j\right| \\ &= |\chi(b)|\left|\sum_{j=1}^r \chi(a_j)\xi_j\right| \\ &= \left|\sum_{j=1}^r \chi(a_j)\xi_j\right|, \qquad (5\text{–}8)\end{aligned}$$

so that each term $F(\cdot,\chi)$ in (5–6) is also unchanged by replacing all of the a_j with the ba_j. This shows that $\hat\mu_{q;a_1,\dots,a_r} = \hat\mu_{q;ba_1,\dots,ba_r}$, which establishes part (b) of the theorem.

The proofs of parts (c) and (d) rely on the formula (5–7) for the function $\hat\rho_{q;a_1,\dots,a_r}$. When the a_j are all squares mod q, then the exponential term in (5–7) is identically 1. Moreover, if b is a square mod q then each ba_j is also a square, while if b is a nonsquare mod q then each ba_j is a nonsquare; in either case we have $c(q,ba_1) = \cdots = c(q,ba_r)$, so that the exponential term in the analogous formula to equation (5–7) for $\hat\rho_{q;ba_1,\dots,ba_r}$ is also identically 1. Since the chain of equalities (5–8) again shows that each term $F(\cdot,\chi)$ is unchanged upon replacing the a_j with ba_j, we see that $\hat\rho_{q;a_1,\dots,a_r} = \hat\rho_{q;ba_1,\dots,ba_r}$ and so $\delta_{q;a_1,\dots,a_r} = \delta_{q;ba_1,\dots,ba_r}$ by virtue of equation (5–4), which establishes part (c) of the theorem.

For part (d) we begin with the formula (5–5) for $\delta_{q;a_r,\dots,a_1}$. As noted above, the exponential term of $\hat\rho_{q;a_1,\dots,a_r}$ is identically 1 when the a_j are all squares mod q, so that $\hat\rho_{q;a_1,\dots,a_r}$ will be real valued. Since $\rho_{q;a_1,\dots,a_r}$ is real-valued as well, we conclude that $\rho_{q;a_1,\dots,a_r}$ is symmetric through the origin. Hence making the change of variables $u_j\mapsto -u_j$ for each $1\le j\le r$ in equation (5–5), we obtain

$$\delta_{q;a_r,\dots,a_1} = \underset{u_1>0,\dots,u_{r-1}>0}{\int\cdots\int} d\rho_{q;a_1,\dots,a_r} = \delta_{q;a_1,\dots,a_r},$$

which establishes part (d) of the theorem.

To establish part (e), we first consider the relationship between $\hat\rho_{q;a_1,\dots,a_r}$ and $\hat\rho_{q;ba_1,\dots,ba_r}$ (note that the residue classes ba_j have not yet been reversed in the second subscript). Again, equation (5–8) shows that replacing each a_j with ba_j does not change the terms of the form $F(\cdot,\chi)$, and so we only need to consider the exponential term. Because the quantity $c(q,a)$ can only take the two values -1 and $c(q,1)$, we see that if $c(q,a')\ne c(q,a)$ then $c(q,a') = c(q,1)-1-c(q,a)$. It follows, under our hypothesis that $c(q,ba_j)\ne c(q,a_j)$ for each $1\le j\le r$, that we also have

$$c(q,ba_{j+1}) - c(q,ba_j) = -(c(q,a_{j+1}) - c(q,a_j)),$$

and so the imaginary expression in the exponential term in equation (5–6) is negated upon replacing each a_j by ba_j. The end result is that

$$\hat\rho_{q;ba_1,\dots,ba_r} = \overline{\hat\rho_{q;a_1,\dots,a_r}},$$

which implies that when the measure $\rho_{q;ba_1,\dots,ba_r}$ is reflected through the origin, the resulting measure is identical to $\rho_{q;a_1,\dots,a_r}$.

Since we can express

$$\delta_{q;ba_r,\dots,ba_1} = \underset{u_1<0,\,\dots,\,u_{r-1}<0}{\int\cdots\int} \rho_{q;ba_1,\dots,ba_r}$$

as in equation (5–5), we can make the change of variables $u_j \mapsto -u_j$ for each $1 \le j \le r-1$ to see that

$$\delta_{q;ba_r,\dots,ba_1} = \int\!\!\cdots\!\!\int_{u_1>0,\,\dots,\,u_{r-1}>0} \rho_{q;a_1,\dots,a_r} = \delta_{q;a_1,\dots,a_r}.$$

This establishes the final assertion of the theorem. $\square$

6. REMARKS, QUESTIONS, AND OPEN PROBLEMS

In this final section, we collect together several observations, unanswered questions, and conjectures concerning the results of this paper.

Systems of Inequalities with One Equality

Since we know that $\delta_{q;a,b}$ and $\delta_{q;b,a}$ are both positive (assuming GRH and LI), each inequality $\pi(x;q,a) > \pi(x;q,b)$ and $\pi(x;q,b) > \pi(x;q,a)$ has arbitrarily large solutions, and therefore $\pi(x;q,a) = \pi(x;q,b)$ for infinitely many integers x. However, knowing that $\delta_{q;a,b,c}$ and $\delta_{q;b,a,c}$ are both positive — i.e., that each string of inequalities

$$\pi(x;q,a) > \pi(x;q,b) > \pi(x;q,c)$$

and

$$\pi(x;q,b) > \pi(x;q,a) > \pi(x;q,c)$$

has arbitrarily large solutions — does not imply that there are necessarily any solutions to $\pi(x;q,a) = \pi(x;q,b) > \pi(x;q,c)$. Undoubtably, the equality $\pi(x;q,a) = \pi(x;q,b)$ should hold infinitely often both when their common value exceeds $\pi(x;q,c)$ and when their value is exceeded by $\pi(x;q,c)$. We conjecture more generally that for any given integer $1 \le j \le r$ and reduced residue classes $a_1, \dots, a_r$ and a'_j mod q, the conditions

$$\begin{array}{c} \pi(x;q,a_1) > \cdots > \pi(x;q,a_j) > \cdots > \pi(x;q,a_r) \\ \| \\ \pi(x;q,a'_j) \end{array}$$

should be satisfied for infinitely many integers x.

Multiple Equalities

Another question along these lines involves solutions to

$$\pi(x;q,a_1) = \pi(x;q,a_2) = \cdots = \pi(x;q,a_r) \qquad (6\text{–}1)$$

when $r \ge 3$. If we consider the vectors

$$\begin{aligned} V_{q;a_1,\dots,a_r}(n) = \big(&\pi(p_n;q,a_1) - \pi(p_n;q,a_2), \\ &\pi(p_n;q,a_2) - \pi(p_n;q,a_3),\ \dots, \\ &\pi(p_n;q,a_{r-1}) - \pi(p_n;q,a_r)\big), \end{aligned} \qquad (6\text{–}2)$$

where p_n denotes the n-th prime, then the sequence of vectors $\{V_{q;a_1,\dots,a_r}(n)\}$ might reasonably be expected to resemble a random walk on $\mathbb{Z}^{r-1}$, where the possible steps at each stage are $(1,0,\dots,0)$, $(-1,1,0,\dots,0)$, …, $(0,\dots,0,-1,1)$, and $(0,\dots,0,-1)$ and are chosen with roughly equal probabilities. (Even though the Chebyshev bias will cause a drift in the mean behavior of the vectors (6–2), this drift has the same order of magnitude as the standard deviation of the random walk).

Since random walks on $\mathbb{Z}^n$ return to any point infinitely often with probability 1 when $n = 1$ or 2 but fail to do so with probability 1 when $n \ge 3$ [Polya 1921], this heuristic leads to the prediction that the system of equalities (6–1) has infinitely many solutions when $r \le 3$ but only finitely many solutions for $r \ge 4$. Similar reasoning suggests that any pair of equalities

$$\pi(x;q,a_1) = \pi(x;q,a_2), \quad \pi(x;q,a_3) = \pi(x;q,a_4)$$

with $a_1, \dots, a_4$ distinct should simultaneously hold for arbitrarily large values of x, but three or more equalities will hold simultaneously only finitely many times. Further, we might expect that the conditions

$$\begin{array}{c} \pi(x;q,a_1) > \cdots > \pi(x;q,a_j) > \cdots > \pi(x;q,a_r) \\ \| \\ \pi(x;q,a'_j) \\ \| \\ \pi(x;q,a''_j) \end{array}$$

and

$$\begin{array}{c} \pi(x;q,a'_i) \\ \| \\ \pi(x;q,a_1) > \cdots > \pi(x;q,a_i) > \cdots \\ > \pi(x;q,a_j) > \cdots > \pi(x;q,a_r) \\ \| \\ \pi(x;q,a'_j) \end{array}$$

should hold for infinitely many integers x, but that analogous conditions involving three or more equalities would not.

Bias Factors

To try to quantify the Chebyshev biases for r-tuples of reduced residue classes a_j mod q for all $r \geq 2$, let us define the "bias factor" $\beta_{q;a_1,\dots,a_r}$ to be the difference between the number of nonsquares preceding squares among the a_j and the number of squares preceding nonsquares:

$$\begin{aligned}\beta_{q;a_1,\dots,a_r} &= \#\{i<j : a_i \neq \Box,\, a_j = \Box\} \\ &\qquad - \#\{i<j : a_i = \Box,\, a_j \neq \Box\} \\ &= \sum_{1\leq i<j\leq r}\frac{c(q,a_j)-c(q,a_i)}{c(q,1)+1} \\ &= \frac{1}{c(q,1)+1}\sum_{1\leq j\leq r}(2j-r-1)c(q,a_j).\end{aligned} \tag{6–3}$$

For instance, when $r=2$ the possible bias factors are $\beta_{q;N,S}=1$, $\beta_{q;N,N}=\beta_{q;S,S}=0$, and $\beta_{q;S,N}=-1$. Rubinstein and Sarnak proved that the sign of $\delta_{q;a,b}-\frac{1}{2}$ equals the sign of $\beta_{q;a,b}$ in this notation, thereby showing that

$$\beta_{q;a,b} > \beta_{q;a',b'} \implies \delta_{q;a,b} > \delta_{q;a',b'}.$$

The converse to this statement is false: the first two lines of Table 6 show that $\delta_{q;a,b}$ and $\delta_{q;a',b'}$ can be different even when $\beta_{q;a,b}=\beta_{q;a',b'}$, for instance.

We might hope that the bias factors $\beta_{q;a_1,\dots,a_r}$ would provide some information about the relative sizes of the $\delta_{q;a_1,\dots,a_r}$, perhaps in the form of the implication

$$\beta_{q;a_1,\dots,a_r} > \beta_{q;b_1,\dots,b_r} \implies \delta_{q;a_1,\dots,a_r} > \delta_{q;b_1,\dots,b_r} \tag{6–4}$$

for any fixed r. In this regard, it is worth remarking that all of the symmetries in Theorem 2 are equalities between two r-tuples of residues with equal bias factors. Examining the densities computed in Section 4, we observe that the implication (6–4) holds most of the time, but we do note the following two anomalies:

- $\beta_{8;5,1,3,7}=\beta_{8;5,1,7,3}=-1>-3=\beta_{8;1,3,7,5}$, but it appears from Table 10 that $\delta_{8;1,3,7,5}$ slightly exceeds both $\delta_{8;5,1,3,7}$ and $\delta_{8;5,1,7,3}$;
- $\beta_{12;7,1,11,5}=\beta_{12;7,1,5,11}=-1>-3=\beta_{12;1,11,5,7}$, but it appears from Table 16 that $\delta_{12;1,11,5,7}$ is slightly greater than both $\delta_{12;7,1,11,5}$ and $\delta_{12;7,1,5,11}$.

It would therefore be of interest, in connection with determining whether the implication (6–4) is always valid, to compute more precisely the densities just mentioned in order to verify the apparent inequalities.

Unfortunately, the computation of the densities to arbitrary precision is not simply a matter of reducing ε and increasing C and letting a bigger computer run for a longer period of time. The major source of error in these computations is the effect of truncating the infinite product defining the functions $F(z,\chi)$ to form the approximations $F_T(z,\chi)$ (see Section 3C); to decrease this error it would be necessary to compute zeros of the relevant L-functions to a height greater than 10,000, and perhaps to greater precision than twelve decimal places as well.

It is certainly conceivable that some definition of bias factor different from (6–3) might be better suited to the role of $\beta_{q;a_1,\dots,a_r}$, although it is hard to imagine what natural definition would be able to explain the apparent anomalies noted above. It might also be the case that the implication (6–4) is valid in more limited settings — for instance, when we restrict to r-tuples $\{a_1,\dots,a_r\}$ and $\{b_1,\dots,b_r\}$ where exactly half of the a_j are nonsquares and half squares, and similarly for the b_j.

Convergence to Unbiased Distribution

Rubinstein and Sarnak [1994, Theorem 1.5] proved that for a fixed integer $r\geq 2$,

$$\Big(\max_{a_1,\dots,a_r}|r!\,\delta_{q;a_1,\dots,a_r}-1|\Big)\to 0 \tag{6–5}$$

as q tends to infinity (where the maximum is taken over all r-tuples of distinct reduced residue classes mod q), so that biases of any sort become less and less evident with increasing moduli. Thus although the biases in the two-way races mod 8 and mod 12 are more pronounced than those in the two-way races mod 4, 5, and 7 owing to the larger values of $c(8,1)=c(12,1)=3$, these sorts of extreme biases will not continue (even with a sequence of moduli such as $q_n=4p_2p_3\dots p_n$, say, which satisfies $c(q_n,1)=2^n-1$).

On the other hand, it might happen that an extremely negatively biased density such as

$$\delta_{q;S_1,\dots,S_n,N_1,\dots,N_n}$$

might tend to zero much more rapidly than $1/(2n)!$ as n increases, while an extremely positively biased density such as $\delta_{q;N_1,\dots,N_n,S_1,\dots,S_n}$ might behave more like $1/(n!)^2$. In general, one could investigate the uniformity of the statement (6–5), i.e., attempt to show that the statement holds uniformly for all $r \le r_0$ for some integer-valued function $r_0 = r_0(q)$ satisfying $2 \le r_0 \le \varphi(q)$. For instance, is it the case that

$$\begin{aligned} \limsup_{q\to\infty} \Big(\max_{a_1,\dots,a_{r_0}} r_0!\, \delta_{q;a_1,\dots,a_{r_0}} \Big) &= \infty, \\ \liminf_{q\to\infty} \Big(\min_{a_1,\dots,a_{r_0}} r_0!\, \delta_{q;a_1,\dots,a_{r_0}} \Big) &= 0 \end{aligned} \tag{6–6}$$

if $r_0 = r_0(q)$ grows sufficiently quickly, and if so, how quickly must r_0 grow with q for these phenomena to emerge? We certainly conjecture that

$$\lim_{q\to\infty} \Big(\max_{a_1,\dots,a_{r_0}} \delta_{q;a_1,\dots,a_{r_0}} \Big) = 0 \tag{6–7}$$

for any arbitrary function $r_0 = r_0(q)$ tending to infinity with q, but at this point it seems nontrivial to prove this modest result even for $r_0 = \varphi(q)$ itself.

Race-game Symmetries, Isomorphisms, and Order Equivalences

Another question of interest is whether there exist more symmetry results of the type arising in Theorem 2. Reviewing the proof of Theorem 2, we see that all of the symmetries therein are consequences of provable equalities between two distributions of the type $\mu_{q;a_1,\dots,a_r}$ or $\rho_{q;a_1,\dots,a_r}$ (possibly after reflecting one of the distributions through the origin). We can then ask

(1) whether there exist any equalities between these distributions other than those used in the proof of Theorem 2;

(2) whether there can be numerical "coincidences" between two densities even though their underlying distributions are not related.

An answer to question (1) might be forthcoming from a careful analysis of the Fourier transforms $\hat\rho_{q;a_1,\dots,a_r}$ of the distributions $\rho_{q;a_1,\dots,a_r}$. As for question (2), it seems reasonable to believe the phenomenon addressed therein can never occur, but proving such a claim seems very difficult.

In support of the possibility that Theorem 2 accounts for all numerical equalities between the densities $\delta_{q;a_1,\dots,a_r}$, we remark that among the densities computed in Section 4, each time a symmetry from Theorem 2 was applicable the corresponding computed densities were equal to within a small multiple of the machine precision. Conversely, all such numerical equalities observed among the computed densities are accounted for by the symmetries already asserted in Theorem 2.

Symmetries among individual densities $\delta_{q;a_1,\dots,a_r}$ are of course closely related to isomorphisms between complete race games of r-tuples. Theorem 2 implies that the following bijections between r-tuples induce isomorphisms of race games:

- the map $\tau(a_j) \equiv a_j^{-1} \pmod q$ between the r-tuples $\{a_1,\dots,a_r\}$ and $\{a_1^{-1},\dots,a_r^{-1}\}$;
- the map $\tau(a_j) \equiv ba_j \pmod q$ between the r-tuples $\{a_1,\dots,a_r\}$ and $\{ba_1,\dots,ba_r\}$, if either $c(q,a_j) = c(q,1)$ for each $1 \le j \le r$ or $c(q,a_j) = c(q,ba_j)$ for each $1 \le j \le r$;
- the map $\tau(a_j) \equiv ba_{r+1-j} \pmod q$ between the r-tuples $\{a_1,\dots,a_r\}$ and $\{ba_r,\dots,ba_1\}$, if $c(q,a_j) \ne c(q,ba_j)$ for each $1 \le j \le r$;
- either bijection $\tau : \{a,b\} \to \{a',b'\}$, if $c(q,a) = c(q,b)$ and $c(q',a') = c(q',b')$;
- any bijection $\tau : \{a,b,c\} \to \{a',b',c'\}$, if there exists $\rho \not\equiv 1 \pmod q$ with $\rho^3 \equiv 1 \pmod q$ such that $b \equiv a\rho \pmod q$ and $c \equiv a\rho^2 \pmod q$ and an analogous $\rho' \pmod{q'}$.

(Our definition of isomorphic race games required that the r-tuples consist of reduced residues to the same modulus, but the definition has an obvious extension to two r-tuples of residues to different moduli which encompasses the last two isomorphisms.) We conjecture that any isomorphism between two race games is induced by a composition of bijections from this list; in particular, the only isomorphisms between race games of distinct moduli are those race games with complete internal symmetry, which were determined by Rubinstein and Sarnak.

A weaker relationship than isomorphic race games is *order-equivalent race games*, where there exists a bijection τ on the set $\{1,\dots,n\}$ such that

$$\delta_{q;a_{\sigma(1)},\dots,a_{\sigma(r)}} > \delta_{q;a_{\sigma'(1)},\dots,a_{\sigma'(r)}} \iff \delta_{q';b_{\tau(\sigma(1))},\dots,b_{\tau(\sigma(r))}} > \delta_{q';b_{\tau(\sigma'(1))},\dots,b_{\tau(\sigma'(r))}} \tag{6–8}$$

for any two permutations σ, σ' of $\{1,\dots,n\}$. Order-equivalent race games seem common for small values of r. For instance, any two race games both of the form $\{N, S\}$ are order-equivalent by Rubinstein and Sarnak's results. The tables in Section 4 indicate many three-way race games that are order-equivalent. The triples $\{N, N', 1\}$ mod 7 with $NN' \not\equiv -1$ mod 7, the triples $\{N, N', 1\}$ mod 8, the triples $\{N, N', 1\}$ mod 9 with $NN' \not\equiv -1$ mod 9, and the triples $\{N, N', 1\}$ mod 12 are all order-equivalent to one another. Also, the triples $\{N, N^{-1}, S\}$ mod 5, the triples $\{N, N^{-1}, S\}$ mod 7, and the triples $\{N, N^{-1}, S\}$ mod 9 are all order-equivalent as well (but note that these are not order-equivalent to the triples $\{N, N, N\}$ mod 8 and mod 12).

We remark that, in view of the values in Tables 10 and 16, the bijection

$$\tau(1) = 1, \quad \tau(3) = 11, \quad \tau(5) = 7, \quad \tau(7) = 5$$

is quite close to inducing an order-equivalence between the full four-way race games modulo 8 and 12, respectively (in the sense that the values in these tables would only have to be modified by at most 6×10^{-5} in order for the condition (6–8) to always hold). It would certainly be interesting to try to establish (or even classify) order-equivalent race games, especially for larger values of r and between r-tuples to different moduli.

Another Problem of Knapowski and Turán

Knapowski and Turán [1962] posed a number of problems in comparative prime number theory, several of which have been answered in [Rubinstein and Sarnak 1994] and in this paper. In their Problem 9 they ask whether, for any r-tuple $a_1, \dots, a_r$ of reduced residue classes mod q, the inequalities

$$\pi(x; q, a_i) > \frac{\mathrm{li}(x)}{\varphi(q)}, \quad \text{for } i = 1, \dots, r, \tag{6–9}$$

simultaneously hold for arbitrarily large values of x. Each individual inequality is unbiased if a_j is a nonsquare mod q and biased negatively if a_j is a square mod q. We remark here that if we apply the method of Rubinstein and Sarnak to the error term

$$E_1(x; q, a) = \frac{\log x}{\sqrt{x}} \bigl(\varphi(q)\pi(x; q, a) - \mathrm{li}(x)\bigr),$$

which has been centered in a slightly different way than in the definition (2–1) of $E(x; q, a)$, we can see that this question of Knapowski and Turán is answered in the affirmative, and in fact the set of real numbers x satisfying the inequalities (6–9) has positive density as well.

ACKNOWLEDGEMENTS

We thank Robert Rumely, who supplied us with the results of his calculations of the zeros of the various $L(s, \chi)$ used in this paper, and also Kenneth Williams for pointing out the closed-form expressions for $L'(1, \chi)$ cited herein. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

[Bays and Hudson 1983] C. Bays and R. H. Hudson, "The cyclic behavior of primes in the arithmetic progressions modulo 11", *J. Reine Angew. Math.* **339** (1983), 215–220.

[Davenport 1980] H. Davenport, *Multiplicative number theory*, 2nd ed., Graduate Texts in Mathematics **74**, Springer, New York, 1980.

[Deninger 1984] C. Deninger, "On the analogue of the formula of Chowla and Selberg for real quadratic fields", *J. Reine Angew. Math.* **351** (1984), 171–191.

[Kaczorowski 1996] J. Kaczorowski, "On the Shanks–Rényi race problem", *Acta Arith.* **74**:1 (1996), 31–46.

[Knapowski and Turán 1962] S. Knapowski and P. Turán, "Comparative prime-number theory, I: Introduction", *Acta Math. Acad. Sci. Hungar.* **13** (1962), 299–314.

[Littlewood 1914] J. E. Littlewood, "Distribution des nombres premiers", *C. R. Acad. Sci. Paris* **158** (1914), 1869–1872.

[Montgomery 1980] H. L. Montgomery, "The zeta function and prime numbers", pp. 1–31 in *Proceedings of the Queen's Number Theory Conference* (Kingston, ON, 1979), edited by P. Ribenboim, Queen's Papers in Pure and Appl. Math. **54**, Queen's Univ., Kingston, Ont., 1980.

[Polya 1921] G. Polya, "Über eine Aufgabe der Wahrscheinlichkeitsrechnung betreffend die Irrfahrt im Strassennetz", *Math. Ann.* **84** (1921), 149–160.

[Rubinstein and Sarnak 1994] M. Rubinstein and P. Sarnak, "Chebyshev's bias", *Experiment. Math.* **3**:3 (1994), 173–197.

[Rumely 1993] R. Rumely, "Numerical computations concerning the ERH", *Math. Comp.* **61**:203 (1993), 415–440, S17–S23.

[Selberg and Chowla 1967] A. Selberg and S. Chowla, "On Epstein's zeta-function", *J. Reine Angew. Math.* **227** (1967), 86–110.

[Shanks 1959] D. Shanks, "Quadratic residues and the distribution of primes", *Math. Tables Aids Comput.* **13** (1959), 272–284.

[Stein and Weiss 1971] E. M. Stein and G. Weiss, *Introduction to Fourier analysis on Euclidean spaces*, Princeton Math. Ser. **32**, Princeton Univ. Press, 1971.

Andrey Feuerverger, University of Toronto, Department of Statistics, Toronto, ON M5S 3G3, Canada (andrey@utstat.toronto.edu)

Greg Martin, University of Toronto, Department of Mathematics, Toronto, ON M5S 3G3, Canada (gerg@math.toronto.edu)

Received November 8, 1999; accepted February 22, 2000