# Densities in certain three-way prime number races

Jiawei Lin and Greg Martin

*Abstract.* Let $a_1$, $a_2$, and $a_3$ be distinct reduced residues modulo $q$ satisfying the congruences $a_1^2 \equiv a_2^2 \equiv a_3^2 \pmod q$. We conditionally derive an asymptotic formula, with an error term that has a power savings in $q$, for the logarithmic density of the set of real numbers $x$ for which $\pi(x; q, a_1) > \pi(x; q, a_2) > \pi(x; q, a_3)$. The relationship among the $a_i$ allows us to normalize the error terms for the $\pi(x; q, a_i)$ in an atypical way that creates mutual independence among their distributions, and also allows for a proof technique that uses only elementary tools from probability.

## 1 Introduction

A major topic in comparative prime number theory is the study of the prime number races among distinct reduced residues $a_1, \ldots, a_r \pmod q$. More precisely, one studies the set of positive real numbers $x$ for which the inequalities $\pi(x; q, a_1) > \cdots > \pi(x; q, a_r)$ hold, where as usual $\pi(x; q, a)$ denotes the number of primes up to $x$ that are congruent to $a$ modulo $q$. It was shown by Rubinstein and Sarnak [11] that the logarithmic density of this set,

$$(1.1) \qquad \delta_{q;a_1,\ldots,a_r} = \lim_{x \to \infty} \frac{1}{\log x} \int\limits_{\substack{2 \le t \le x \\ \pi(t;q,a_1) > \cdots > \pi(t;q,a_r)}} \frac{dt}{t},$$

exists and is strictly between 0 and 1, under two assumptions:

- GRH: the generalized Riemann hypothesis, asserting that all nontrivial zeros $\rho = \beta + i\gamma$ of Dirichlet $L$-functions $L(s, \chi)$ satisfy $\beta = \frac{1}{2}$;
- LI: a linear independence hypothesis, asserting that the multiset $\{\gamma \ge 0 :$ there exists $\chi \pmod q$ such that $L(\frac{1}{2} + i\gamma, \chi) = 0\}$ is linearly independent over the rational numbers.

For a fixed number of contestants $r$, the logarithmic densities $\delta_{q;a_1,\cdots,a_r}$ approach $\frac{1}{r!}$ uniformly as $q \to \infty$. Several authors (the articles [3, 5, 8, 9] are most closely related to the present work) have given asymptotic formulas for the difference for various numbers of contestants (including results when the number of contestants can grow with $q$), and others have provided generalizations to number fields, function fields, and elliptic curves, as well as to the counting functions of integers with a fixed number of prime factors in arithmetic progressions.

In this paper, we investigate a special class of three-way prime number races, where the residues involved satisfy the congruence

$$(1.2) \qquad a_1^2 \equiv a_2^2 \equiv a_3^2 \ (\mathrm{mod}\ q).$$

We use elementary ideas from probability, and an approach involving an unusual normalization, to establish an asymptotic formula for the corresponding density $\delta_{q;a_1,a_2,a_3}$ with a very good error term. To state our theorem, we must first define some notation.

**Definition 1.1** For any Dirichlet character $\chi \ (\mathrm{mod}\ q)$, define

$$b_+(\chi) = \sum_{\substack{\gamma > 0 \\ L(1/2+i\gamma,\chi)=0}} \frac{1}{\frac{1}{4} + \gamma^2} \quad \text{and} \quad b(\chi) = \sum_{\substack{\gamma \in \mathbb{R} \\ L(1/2+i\gamma,\chi)=0}} \frac{1}{\frac{1}{4} + \gamma^2}.$$

Notice that $b(\chi) = b_+(\chi) + b_+(\overline{\chi})$ by the functional equation for Dirichlet $L$-functions, assuming that $L(\frac{1}{2}, \chi) \neq 0$ (which is a consequence of LI). Notice also that if $\chi$ is induced by another character $\chi^*$ then $b_+(\chi) = b_+(\chi^*)$ and similarly for $b(\chi)$.

**Definition 1.2** Define the following sets of characters (mod q):

$$\begin{aligned}
H_0 &= \{\chi \ (\mathrm{mod}\ q): \chi(a_1) = \chi(a_2) = \chi(a_3)\}, \\
H_1 &= \{\chi \ (\mathrm{mod}\ q): \chi(a_2) = \chi(a_3) = -\chi(a_1)\}, \\
H_2 &= \{\chi \ (\mathrm{mod}\ q): \chi(a_1) = \chi(a_3) = -\chi(a_2)\}, \\
H_3 &= \{\chi \ (\mathrm{mod}\ q): \chi(a_1) = \chi(a_2) = -\chi(a_3)\}.
\end{aligned}$$

**Remark 1.3** All these sets have the property that $\chi \in H_i$ if and only if $\overline{\chi} \in H_i$. It is easy to verify that $H_0$ is a subgroup of the group of Dirichlet characters (mod $q$) and that $H_1$, $H_2$, and $H_3$, if nonempty, are cosets of that subgroup.

Furthermore, under the assumption (1.2), we show in Lemma 3.3 below that $H_0$ is an index-4 subgroup of the group of characters (mod $q$) and that $H_1$, $H_2$, and $H_3$ are all its cosets, so that every character (mod $q$) is in exactly one of $H_0$, $H_1$, $H_2$, or $H_3$.

**Definition 1.4** With $b_+(\chi)$ as in Definition 1.1, for $i \in \{0, 1, 2, 3\}$ define

$$V(q) = 2 \sum_{\substack{\chi \ (\mathrm{mod}\ q) \\ \chi \neq \chi_0}} b_+(\chi), \quad V_i = 32 \sum_{\chi \in H_i} b_+(\chi), \quad \text{and} \quad \eta_i = \frac{V_i}{4V(q)} - 1.$$

With this notation in place, we may now state the main theorem of this paper.

**Theorem 1.5** *Assume GRH and LI. If $a_1$, $a_2$, and $a_3$ are distinct reduced residues modulo q satisfying $a_1^2 \equiv a_2^2 \equiv a_3^2 \ (\mathrm{mod}\ q)$, then*

$$\delta_{q;a_1,a_2,a_3} = \frac{1}{2\pi} \arctan \frac{\sqrt{V_1 V_2 + V_1 V_3 + V_2 V_3}}{V_2} + O_\varepsilon(q^{-1/2+\varepsilon}).$$

*Moreover, if $a_1$, $a_2$, and $a_3$ are all quadratic residues, or all quadratic nonresidues (mod q), then the error term can be improved to $O(1/\phi(q) \log q)$.*

As it happens, most of this paper is concerned with proving the second assertion (with the additional hypothesis on the quadratic nature of $a_1$, $a_2$, and $a_3$), after which we derive the first assertion (with its weaker error term) from it.

Note that if $V_1$, $V_2$, and $V_3$ are all quite close to one another (as we shall show is the case), then the argument of arctan in Theorem 1.5 is approximately $\sqrt{3}$, so that the main term is approximately $\frac{1}{6}$ as expected. Stating the theorem with this main term of a perhaps unforeseen shape allows the error term to remain quite small. However, we can derive a simpler asymptotic formula from this theorem if we are less concerned with the quality of the error term:

**Corollary 1.6** *Assume GRH and LI. If $a_1$, $a_2$, and $a_3$ are distinct reduced residues modulo $q$ satisfying $a_1^2 \equiv a_2^2 \equiv a_3^2 \pmod{q}$, then*

$$\delta_{q;a_1,a_2,a_3} = \frac{1}{6} + \frac{\eta_1 + \eta_3}{8\pi\sqrt{3}} - \frac{\eta_2}{4\pi\sqrt{3}} + O\left( \frac{(\log\log q)^2}{(\log q)^2} \right).$$

This version of the result recovers a special case of a theorem of Lamzouri [9], with a somewhat simpler proof; see the end of Section 8 for the details of the comparison.

We have four motivations for presenting Theorem 1.5 and its proof. First, the theorem has a better error term than has been recorded in the literature for any prime number race with three or more competitors; indeed, for such races, it is rare to see a savings of a power of $q$ at all. Second, our proof of Theorem 1.5 involves an unusual normalization (see Definition 3.5 below) of the error terms for the $\pi(x; q, a_i)$, one that allows us to treat the three error terms connected to this race as random variables that are in fact independent, which we hope might inspire similar constructions in other settings. Third, much of the recent progress on prime number races has invoked powerful machinery from probability; we wanted to give an application in this subject where more elementary methods suffice. Finally, we were motivated by generalizing the discussion of the second author from [10], which essentially treats the two smallest cases $q = 8$ and $q = 12$ of Theorem 1.5 numerically, but with a heuristic analysis that anticipates the methods herein.

That being said, the methods from the current literature in comparative prime number theory are capable of treating much more general circumstances, and also, if viewed from a suitable perspective, of providing formulas with error terms nearly as strong as that of Theorem 1.5. See Section 9 (and also the end of Section 3) for further discussion about this wider context.

The rest of the paper is organized as follows. We quote results from the literature in Section 2 concerning the limiting logarithmic distributions of error terms for prime counting functions and the random variables that model them. It is in Section 3 that we define the atypical normalization of these error terms that allows us to treat them independently, and calculate their variances. Using known facts about Bessel functions, we exhibit in Section 4, the characteristic function of our random variables and derive some power series representations of them. In Sections 5 and 6, we establish pointwise bounds between these characteristic functions and the characteristic function of normal variables with the same mean and variance, as well as between the second derivatives of these characteristic functions. This information allows us to compare the density functions and eventually the probabilities themselves of these

two types of random variables in Section 7, at which point we prove Theorem 1.5 and Corollary 1.6 in Section 8. We conclude with some discussion of the relationship between our results and existing results, and espouse a viewpoint on how such results should be conceived, in Section 9.

## 2 Background information

The foundation of the method we use has appeared many times, certainly stimulated in this generation by [11]. It will be most convenient for us to quote several definitions and results from the work by Fiorilli and the second author [3], starting with the traditional normalization of the error term for prime counting functions in arithmetic progressions.

**Definition 2.1**   For any reduced residue $a \pmod q$, define
$$E(x;q,a) = \frac{\phi(q)\pi(x;q,a) - \pi(x)}{\sqrt{x}/\log x},$$
where $\pi(x) = \pi(x;1,1)$ as usual.

The following explicit formula for $E(x;q,a)$ is [11, Lemma 2.1], simplified slightly by the assumption of GRH.

**Lemma 2.2**   *Assume GRH. For any reduced residue $a \pmod q$*
$$E(x;q,a) = -c_q(a) + \sum_{\substack{\chi \pmod q \\ \chi \neq \chi_0}} \overline{\chi}(a)E(x,\chi) + o(1)$$
*as $x \to \infty$; here*

(2.1)
$$c_q(a) = -1 + \#\{b \pmod q : b^2 \equiv a \pmod q\}$$

*and, for any Dirichlet character $\chi$,*
$$E(x,\chi) = \sum_{\substack{\gamma \in \mathbb{R} \\ L(1/2+i\gamma,\chi)=0}} \frac{x^{i\gamma}}{1/2 + i\gamma}$$

*(which converges conditionally when interpreted as the limit of $\sum_{|\gamma|<T}$ as $T$ tends to infinity).*

It is convenient to be able to interpret the distribution of values of $E(x;q,a)$ in terms of certain random variables.

**Definition 2.3**   For any Dirichlet character $\chi \pmod q$, define the random variable
$$Z_\chi = \sum_{\substack{\gamma>0 \\ L(1/2+i\gamma,\chi)=0}} \frac{Z_\gamma}{\sqrt{\frac{1}{4} + \gamma^2}},$$
where $Z_\gamma$ are independently uniformly distributed on the unit circle in $\mathbb{C}$. We also use the notation $X_\gamma = \Re Z_\gamma$ and $X_\chi = \Re Z_\chi$, so that the $X_\gamma$ also form an independent

collection of random variables, as do the $X_\chi$, assuming that $L(s, \chi)$ have no zeros in common (which is a consequence of LI).

It is known that vector-valued relatives of $E(x; q, a)$ have limiting logarithmic distributions that can be expressed in terms of these random variables; the following proposition is [3, Proposition 2.3].

**Proposition 2.4**    *Assume LI. Let $\{c_\chi \colon \chi \pmod q\}$ be a collection of $\mathbb{C}^r$-vectors, indexed by the Dirichlet characters $\pmod q$, satisfying $c_{\overline{\chi}} = \overline{c_\chi}$. The limiting logarithmic distribution of any $\mathbb{R}^r$-valued function of the form*

$$\sum_{\chi \pmod q} c_\chi E(x, \chi) + o(1)$$

*is the same as the distribution of the random variable*

$$2\Re \sum_{\chi \pmod q} c_\chi Z_\chi.$$

## 3 Special three-way races and error terms with atypical normalizations

In this section, we set out some notation that will be used throughout the main part of this paper (from this point through Section 8). In particular, the assumptions on $a_1$, $a_2$, and $a_3$ in the first definition will be in force in these sections without explicit mention, as our main result is concerned only with these special three-way prime number races.

**Definition 3.1**    Let $a_1$, $a_2$, and $a_3$ denote distinct reduced residues $\pmod q$ such that

$$a_1^2 \equiv a_2^2 \equiv a_3^2 \pmod q.$$

We assume, through the middle of Section 8, that $a_1$, $a_2$, and $a_3$ are either all quadratic residues or all quadratic nonresidues $\pmod q$. (Later in Section 8 we will discuss how this assumption can be removed to establish Theorem 1.5 in its entirety.)

We will often use $i, j, k$ as indices that denote a generic permutation $(i, j, k)$ of $(1, 2, 3)$. For example, we define

$$a_0 \equiv a_i a_j a_k^{-1} \pmod q,$$

which is independent of the permutation $(i, j, k)$.

**Remark 3.2**    It is easy to show that most integers, $q$, possess three distinct reduced residues $a_1$, $a_2$, and $a_3$ such that the congruences $a_1^2 \equiv a_2^2 \equiv a_3^2 \pmod q$ are satisfied—indeed, the integers that do not are precisely the integers with primitive roots. It is also straightforward to show that one almost always can choose these reduced residues so that $a_1$, $a_2$, and $a_3$ are all quadratic nonresidues; for example, such a choice is possible whenever $q$ has at least three distinct odd prime factors.

In the special situation described in Definition 3.1, the sets $H_i$ from Definition 1.2 have a tidy relationship with one another.

**Lemma 3.3**    *The sets $H_0$, $H_1$, $H_2$, and $H_3$ partition the group of Dirichlet characters $\pmod q$ into four subsets each of cardinality $\frac{1}{4}\phi(q)$.*

**Proof**  The assumption $a_1^2 \equiv a_2^2 \equiv a_3^2 \pmod{q}$ implies that $\chi(a_1^2) = \chi(a_2^2) = \chi(a_3^2)$, or equivalently $\chi^2(a_1) = \chi^2(a_2) = \chi^2(a_3)$, for every Dirichlet character $\chi \pmod{q}$. Therefore, if $a_1^2 \equiv a_2^2 \equiv a_3^2 \pmod{q}$, then for every $\chi \pmod{q}$ each of $\chi(a_1)$, $\chi(a_2)$, and $\chi(a_3)$ must be a square root of the common value $\chi^2(a_i)$. Since there are only two such square roots, at least two of the character values must be equal, and the third value (if not equal to the other two) is the negative of the others. In particular, the sets $H_0$, $H_1$, $H_2$, and $H_3$ partition the group of characters $\pmod{q}$. The fact that they have equal cardinalities, which must necessarily be $\frac{1}{4}\phi(q)$, now follows from the observation made in Remark 1.3 that $H_1$, $H_2$, and $H_3$ are all cosets of the subgroup $H_0$.  ∎

We are also able to simplify certain combinations of character values in this special situation.

**Lemma 3.4**  *For any permutation* $(i, j, k)$ *of* $(1, 2, 3)$,

$$\overline{\chi}(a_i) + \overline{\chi}(a_0) - \overline{\chi}(a_j) - \overline{\chi}(a_k) = \begin{cases} 4\overline{\chi}(a_i), & \text{if } \chi \in H_i, \\ 0, & \text{otherwise,} \end{cases}$$

*where* $H_i$ *is the set of characters from* Definition 1.2.

**Proof**  It is immediate from Definitions 1.2 and 3.1 that $\chi(a_0) = \chi(a_i)$ if $\chi \in H_0 \cup H_i$ and that $\chi(a_0) = -\chi(a_i)$ if $\chi \in H_j \cup H_k$, and then that

$$\overline{\chi}(a_i) + \overline{\chi}(a_0) - \overline{\chi}(a_j) - \overline{\chi}(a_k)$$
$$= \begin{cases} \overline{\chi}(a_i) + \overline{\chi}(a_i) - \overline{\chi}(a_i) - \overline{\chi}(a_i) = 0, & \text{if } \chi \in H_0, \\ \overline{\chi}(a_i) + \overline{\chi}(a_i) + \overline{\chi}(a_i) + \overline{\chi}(a_i) = 4\overline{\chi}(a_i), & \text{if } \chi \in H_i, \\ \overline{\chi}(a_i) - \overline{\chi}(a_i) + \overline{\chi}(a_i) - \overline{\chi}(a_i) = 0, & \text{if } \chi \in H_j, \\ \overline{\chi}(a_i) - \overline{\chi}(a_i) - \overline{\chi}(a_i) + \overline{\chi}(a_i) = 0, & \text{if } \chi \in H_k. \end{cases} \quad ∎$$

At this point, we introduce an unusual normalization, tailored to this special situation, of the error terms $E(x; q, a)$ from Definition 2.1.

**Definition 3.5**  For any permutation $(i, j, k)$ of $(1, 2, 3)$, define

$$E^*(x; q, a_i) = \alpha + E(x; q, a_i) + E(x; q, a_0) - E(x; q, a_j) - E(x; q, a_k),$$

where $\alpha = c_q(a_0) - c_q(a_i)$ (which, by assumption, is independent of $i \in \{1, 2, 3\}$). Note that

$$E(x; q, a_i) - \tfrac{1}{2}E^*(x; q, a_i) = \tfrac{1}{2}\big(E(x; q, a_i) + E(x; q, a_j) + E(x; q, a_k) - E(x; q, a_0) - \alpha\big)$$

is independent of the permutation $(i, j, k)$, so that the ordering of the $E^*$ terms is always the same as the ordering of the $E$ terms. In particular, $E^*(x; q, a_i) > E^*(x; q, a_j) > E^*(x; q, a_k)$ if and only if $\pi(x; q, a_i) > \pi(x; q, a_j) > \pi(x; q, a_k)$, so that equation (1.1) becomes

$$(3.1) \qquad \delta_{q; a_1, a_2, a_3} = \lim_{x \to \infty} \frac{1}{\log x} \int_{\substack{2 \le t \le x \\ E^*(t; q, a_1) > E^*(t; q, a_2) > E^*(t; q, a_3)}} \frac{dt}{t}.$$

We can immediately start to see the benefit of this atypical normalization, in that the explicit formulas for the $E(x; q, a_i)$ involve disjoint sets of Dirichlet characters.

**Lemma 3.6** *Assume GRH. For $i \in \{1, 2, 3\}$, we have $E^*(x; q, a_i) = 4 \sum_{\chi \in H_i} \overline{\chi}(a_i) E(x, \chi) + o(1)$.*

**Proof** By Lemma 2.2,

$$
E^*(x; q, a_i) = \left( \alpha + E(x; q, a_0) \right) + E(x; q, a_i) - E(x; q, a_j) - E(x; q, a_k)
$$

$$
(3.2) \qquad = -c_q(a_i) - c_q(a_i) + c_q(a_j) + c_q(a_k)
$$

$$
+ \sum_{\substack{\chi \pmod q \\ \chi \neq \chi_0}} \left( \overline{\chi}(a_i) + \overline{\chi}(a_0) - \overline{\chi}(a_j) - \overline{\chi}(a_k) \right) E(x, \chi) + o(1).
$$

The assumption that $a_1$, $a_2$, and $a_3$ are either all quadratic residues or all quadratic nonresidues (mod $q$) means that the four quantities on line (3.2) are all equal and thus cancel one another. The lemma now follows from Lemma 3.4. ∎

We remark that Lemma 3.6 is the only place in our argument where we use the standing assumption that $a_1$, $a_2$, and $a_3$ are either all quadratic residues or all quadratic nonresidues (mod $q$); in particular, we use this assumption to ensure that $E^*(x; q, a_i)$ have no constant term, so that their limiting logarithmic distributions will have mean 0. In Section 8, we show how we can derive the general form of Theorem 1.5 from the version that requires this assumption.

At this point, we are ready to introduce certain random variables that model, in their distributions, the normalized error terms $E^*(x; q, a_i)$.

**Definition 3.7** For $i \in \{1, 2, 3\}$, define the random variable

$$
X_i = 8 \sum_{\chi \in H_i} X_\chi,
$$

where $X_\chi$ is as in Definition 2.3; note that the disjointness of $H_1$, $H_2$, and $H_3$ and the independence of the $X_\chi$ imply that $X_1$, $X_2$, and $X_3$ are mutually independent. Furthermore, for any permutation $(i, j, k)$ of $(1, 2, 3)$, define the vector-valued random variable

$$
X_{i,j,k} = (X_i, X_j, X_k).
$$

**Lemma 3.8** *For $i \in \{1, 2, 3\}$, the variance of $X_i$ is the quantity $V_i$ from* Definition 1.4.

**Proof** By Definitions 3.7 and 2.3,

$$
\sigma^2(X_i) = \sigma^2 \left( 8 \sum_{\chi \in H_i} X_\chi \right) = 64 \sigma^2 \left( \sum_{\chi \in H_i} \sum_{\substack{\gamma > 0 \\ L(1/2 + i\gamma, \chi) = 0}} \frac{X_\gamma}{\sqrt{\frac{1}{4} + \gamma^2}} \right).
$$

Since $X_\gamma$ are independent by assumption,

$$\sigma^2(X_i) = 64 \sum_{\chi \in H_i} \sum_{\substack{\gamma > 0 \\ L(1/2 + i\gamma, \chi) = 0}} \frac{\sigma^2(X_\gamma)}{\frac{1}{4} + \gamma^2}$$

$$= 32 \sum_{\chi \in H_i} \sum_{\substack{\gamma > 0 \\ L(1/2 + i\gamma, \chi) = 0}} \frac{1}{\frac{1}{4} + \gamma^2} = 32 \sum_{\chi \in H_i} b_+(\chi) = V_i$$

as claimed. ∎

   Our final proposition of the section records the fact that these random variables truly are substitute objects of study for the normalized prime-counting error terms. Recall the logarithmic density $\delta_{q;a_i,a_j,a_k}$ from Definition 3.1.

**Proposition 3.9** *Assume GRH and LI. For any permutation* $(i, j, k)$ *of* $(1, 2, 3)$, *the limiting logarithmic distribution of the vector-valued function* $\big(E^*(x; q, a_i), E^*(x; q, a_j), E^*(x; q, a_k)\big)$ *is the same as the distribution of the random variable* $X_{i,j,k}$; *in particular,*

$$\delta_{q;a_1,a_2,a_3} = \Pr(X_1 > X_2 > X_3).$$

**Proof**   By Lemma 3.6,

$$\big(E^*(x; q, a_i), E^*(x; q, a_j), E^*(x; q, a_k)\big)$$

$$= 4\Bigg( \sum_{\chi \in H_i} (\overline{\chi}(a_i), 0, 0) E(x, \chi) + \sum_{\chi \in H_j} (0, \overline{\chi}(a_j), 0) E(x, \chi)$$

$$+ \sum_{\chi \in H_k} (0, 0, \overline{\chi}(a_k)) E(x, \chi) \Bigg) + o(1),$$

whose limiting logarithmic distribution, by Proposition 2.4, is the same as the distribution of

$$2\mathfrak{R}\Bigg( 4\Bigg( \sum_{\chi \in H_i} (\overline{\chi}(a_i), 0, 0) Z_\chi + \sum_{\chi \in H_j} (0, \overline{\chi}(a_j), 0) Z_\chi + \sum_{\chi \in H_k} (0, 0, \overline{\chi}(a_k)) Z_\chi \Bigg) \Bigg).$$

Since $Z_\chi$ is uniformly distributed on the unit circle in $\mathbb{C}$, and $\overline{\chi}(a_i)$ is a point on the unit circle, we have simply $\overline{\chi}(a_i) Z_\chi = Z_\chi$, and similarly with $i$ replaced by $j$ or $k$. Therefore,

$$2\mathfrak{R}\Bigg( 4\Bigg( \sum_{\chi \in H_i} (\overline{\chi}(a_i), 0, 0) Z_\chi + \sum_{\chi \in H_j} (0, \overline{\chi}(a_j), 0) Z_\chi + \sum_{\chi \in H_k} (0, 0, \overline{\chi}(a_k)) Z_\chi \Bigg) \Bigg)$$

$$= 8\mathfrak{R}\Bigg( \sum_{\chi \in H_i} (Z_\chi, 0, 0) + \sum_{\chi \in H_j} (0, Z_\chi, 0) + \sum_{\chi \in H_k} (0, 0, Z_\chi) \Bigg)$$

$$= 8\mathfrak{R}\Bigg( \sum_{\chi \in H_i} Z_\chi, \sum_{\chi \in H_j} Z_\chi, \sum_{\chi \in H_k} Z_\chi \Bigg) = (X_i, X_j, X_k)$$

as claimed. The final assertion follows from equation (3.1) and the definition of a limiting logarithmic distribution. ∎

**Remark 3.10** From Definition 3.7, we note that we only need to assume GRH and LI for Dirichlet $L$-functions corresponding to the subset $H_1 \cup H_2 \cup H_3$ of Dirichlet characters (mod $q$).

We have shown that the standing assumptions from Definition 3.1 imply that atypical normalizations of the error terms for prime counting functions can be made independent of one another. Our proof used the fact that the sets from Definition 1.2 comprised all Dirichlet characters (mod $q$); said another way, every $\chi$ (mod $q$) takes at most two distinct values on $a_1$, $a_2$, and $a_3$. It turns out that a property of this type is more fundamental to our method than the original congruence assumption, an idea which we now take a slight detour to explore.

**Definition 3.11** Given distinct reduced residues $a_1, \ldots, a_r$ (mod $q$), we call a Dirichlet character $\chi$ (mod $q$) *almost unanimous on* $\{a_1, \ldots, a_r\}$ if there exists an index $1 \le k_\chi \le r$ and a complex number $\omega_\chi$ such that $\chi(a_1) = \cdots = \chi(a_{k_\chi - 1}) = \chi(a_{k_\chi + 1}) = \cdots = \chi(a_r) = \omega_\chi$.

Note that this definition includes the possibility that $\chi(a_{k_\chi})$ is also equal to $\omega_\chi$ (in which case $k_\chi$ can take any value in $\{1, \ldots, r\}$); for example, the principal character $\chi_0$ is always almost unanimous on any set of reduced residues. Note also that if $\chi$ is almost unanimous on $\{a_1, \ldots, a_r\}$, then so is $\overline{\chi}$, and $k_{\overline{\chi}} = k_\chi$ and $\omega_{\overline{\chi}} = \overline{\omega_\chi}$.

Furthermore, we call the set $\{a_1, \ldots, a_r\}$ itself *almost unanimous* if every Dirichlet character (mod $q$) is almost unanimous on $\{a_1, \ldots, a_r\}$.

If a set $\{a_1, \ldots, a_r\}$ of distinct reduced residues is almost unanimous, then one can create an atypical normalization by subtracting the quantity $\sum_{\chi \,(\mathrm{mod}\, q)} \omega_\chi E(x, \chi)$ from each error term $E(x; q, a)$. Note that this quantity is real-valued since $\omega_{\overline{\chi}} = \overline{\omega_\chi}$, and, therefore, subtracting it from every $E(x; q, a)$ is order-preserving. The resulting differences have the form

$$E(x; q, a_j) - \sum_{\chi \,(\mathrm{mod}\, q)} \omega_\chi E(x, \chi) = -c_q(a_j) + \sum_{\substack{\chi \,(\mathrm{mod}\, q) \\ k_\chi = j}} (\overline{\chi}(a) - \omega_\chi) E(x, \chi);$$

in particular, the sets of characters appearing in the sums for different values of $j$ are disjoint. As a result, assuming GRH and LI, the random variables modeling this atypically normalized error term will be mutually independent.

An examination of Definition 3.5 reveals that the quantities $E^*(x; q, a_i)$, up to constant factors, are precisely the result of applying this construction (the supplemental residue $a_0$, while making the definition concise, is not crucial to the construction). In principle, then, this process of atypically normalizing the error terms for any almost unanimous set of residues would result in independent error terms.

The unfortunate news, however, is that there are no almost unanimous sets of $r \ge 3$ residues other than the ones described in Definition 3.1. (All sets of 1 or 2 residue classes are trivially almost unanimous, and there are many ways to normalize two of these error terms to create independent functions—including simply replacing

$E(x; q, a)$ and $E(x; q, b)$ with $E(x; q, a) - E(x; q, b)$ and 0, which is common practice.) The following two lemmas justify this anticlimactic assertion.

**Lemma 3.12** *If $r \geq 4$, then there does not exist an almost unanimous set $a_1, \ldots, a_r$ of distinct reduced residues (mod q).*

**Proof** Suppose, for the sake of contradiction, that $a_1, \ldots, a_r$ is almost unanimous. Since $a_1 \not\equiv a_2 \pmod{q}$, there exists a character $\chi_1$ such that $\chi_1(a_1) \neq \chi_1(a_2)$. By assumption, $\chi_1$ is almost unanimous on $\{a_1, \ldots, a_r\}$; without loss of generality, $\chi_1(a_2) = \cdots = \chi_1(a_r)$. Similarly, there exists a character $\chi_2$ such that $\chi_1(a_{r-1}) \neq \chi_1(a_r)$, and without loss of generality, $\chi_2(a_1) = \cdots = \chi_2(a_{r-1})$. (It is in this second "without loss of generality" step that we use the assumption $r \geq 4$, so that there is no overlap between $\{a_1, a_2\}$ and $\{a_{r-1}, a_r\}$.)

Now set $\chi_3 = \chi_1 \chi_2$. We see immediately that $\chi_3(a_2) = \cdots = \chi_3(a_{r-1})$. However, the known facts $\chi_1(a_1) \neq \chi_1(a_2)$ and $\chi_2(a_1) = \chi_2(a_2)$ imply that $\chi_3(a_1) \neq \chi_3(a_2)$; similarly, $\chi_1(a_{r-1}) = \chi_1(a_r)$ and $\chi_2(a_{r-1}) \neq \chi_2(a_r)$ imply that $\chi_3(a_{r-1}) \neq \chi_3(a_r)$. It follows that $\chi_3$ is not almost unanimous on $\{a_1, \ldots, a_r\}$, contrary to assumption. ∎

**Lemma 3.13** *Let $a_1, a_2, a_3$ be distinct reduced residue classes* (mod *q*)*. Then, $\{a_1, a_2, a_3\}$ is almost unanimous (mod q) if and only if $a_1^2 \equiv a_2^2 \equiv a_3^2 \pmod{q}$.*

**Proof** The proof of Lemma 3.3 shows that if $a_1^2 \equiv a_2^2 \equiv a_3^2 \pmod{q}$ then $\{a_1, a_2, a_3\}$ is almost unanimous (mod *q*).

Conversely, suppose that $\{a_1, a_2, a_3\}$ is almost unanimous (mod *q*). Define

$$G_1 = \{\chi \pmod{q} : \chi(a_2) = \chi(a_3)\}$$
$$G_2 = \{\chi \pmod{q} : \chi(a_1) = \chi(a_3)\}$$
$$G_3 = \{\chi \pmod{q} : \chi(a_1) = \chi(a_2)\},$$

and note that $G_1 \cap G_2 \cap G_3 = H_0$ as in Definition 1.2; moreover, by the definition of almost unanimous, $G = G_1 \cup G_2 \cup G_3$. It is obvious that each $G_i$ is a subgroup of the group $G$ of Dirichlet characters (mod *q*). Furthermore, for any pair of distinct residues (mod *q*), there is always a Dirichlet character (mod *q*) that takes different values on the two residues; in particular, the $G_i$ are proper subgroups.

Scorza (see [14]) proved that a group $G$ is the union of three proper subgroups $G_1$, $G_2$, and $G_3$ if and only if it has a quotient isomorphic to the Klein 4-group $K$ (a result that has been rediscovered more than once—see [4] for example), in which case $G_1$, $G_2$, and $G_3$ are the inverse images of the three two-element subgroups of $K$. In particular, the square of every element of $G$ is in $G_1 \cap G_2 \cap G_3 = H_0$. In our situation, we deduce that $\chi^2(a_1) = \chi^2(a_2) = \chi^2(a_3)$ for every $\chi \pmod{q}$, which implies that $\chi(a_1^2) = \chi(a_2^2) = \chi(a_3^2)$ for every $\chi \pmod{q}$; this situation is possible only if $a_1^2 \equiv a_2^2 \equiv a_3^2 \pmod{q}$. ∎

While there seems to be no direct generalization of our construction, we hope that the ideas described herein might inspire other beneficial atypical normalizations in the future.
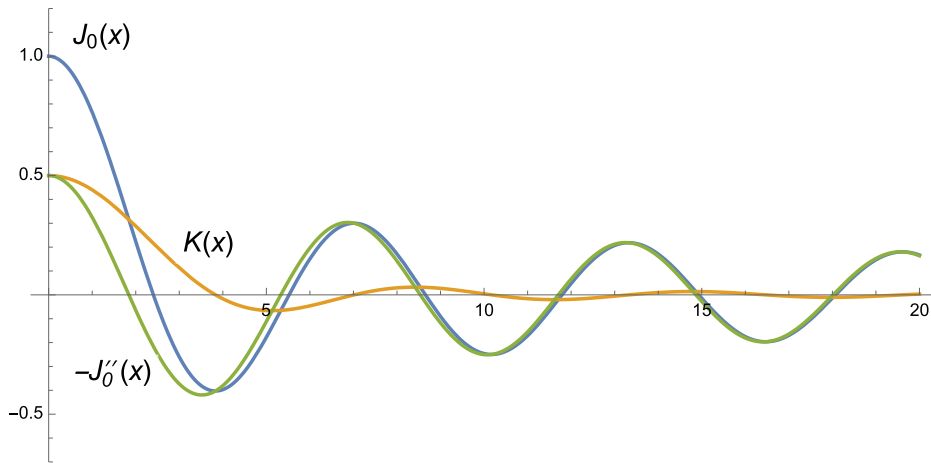
*Figure 1* The Bessel function and its relatives

## 4 Bessel functions and bounds for characteristic functions

In this section, we exhibit exact formulas for the characteristic function of the random variables $X_i$ introduced in the previous section, as well as various estimates and series representations of those characteristic functions that will be needed in our analysis.

***Definition 4.1***   Let $J_0(z)$ be the standard Bessel function of order 0. Let $\lambda_n$ be the coefficients in the power series expansion

$$\log J_0(z) = \sum_{n=0}^{\infty} \lambda_n z^n,$$

which is valid for $|z| \le \frac{12}{5}$ since $J_0$ has no zeros in this disk: this assertion can be verified computationally for real $z$—see the graph of $J_0(x)$ in Figure 1—while Hurwitz [7] proved that $J_0(z)$ has no nonreal zeros (see also [6]).

The following lemma is [3, Lemma 2.8]:

***Lemma 4.2***   *With* $\lambda_n$ *as in* Definition 4.1:
(a) $\lambda_n \ll \left(\frac{5}{12}\right)^n$ *uniformly for* $n \ge 0$;
(b) $\lambda_0 = 0$ *and* $\lambda_{2m-1} = 0$ *for every* $m \ge 1$;
(c) $\lambda_{2m} < 0$ *for every* $m \ge 1$.

It would be advantageous if this Bessel function were decreasing for $x \ge 0$, say, so that we could bound the tail of the function $|J_0(x)|$ simply by $|J_0(\kappa)|$ for any fixed $0 \le \kappa \le x$. Inconveniently, $J_0(x)$ and its derivatives have oscillations in sign; it is the case, however, that their values are contained in a gradually decaying envelope. Consequently, their values near $x = 0$ are indeed their largest, an observation we codify in the following lemma.

***Definition 4.3***   Define $K(x) = -J_0'(x)/x$, with the value $K(0) = \frac{1}{2}$ chosen for continuity. Moreover, define $D(x) = -J_0'(x)/J_0(x) = xK(x)/J_0(x)$.

***Lemma 4.4***   *If $0 \le \kappa \le \frac{1}{2}$, then $J_0(\kappa)$, $K(\kappa)$, $-J''(\kappa)$, and $D(\kappa)$ are all positive; and for all real numbers $x$ with $|x| \ge \kappa$ we have $J_0(\kappa) \ge |J_0(x)|$ and $K(\kappa) \ge |K(x)|$ and $|J_0''(\kappa)| \ge |J_0''(x)|$. In particular,*

$$(4.1) \qquad |K(x)| \le \frac{1}{2} \quad and \quad |J_0''(x)| \le \frac{1}{2} \quad for\ all\ x \in \mathbb{R}.$$

**Proof**   These assertions are clear from the graphs of the functions are Figure 1 (since the first three functions are even, we may restrict attention to $x \ge 0$); a rigorous proof is unenlightening, and we omit most of the details. Derivatives of Bessel functions are related to Bessel functions of higher order, and in particular

$$K(x) = \frac{J_1(x)}{x} \quad and \quad -J_0''(x) = J_0(x) - \frac{J_1(x)}{x} = \frac{J_0(x) - J_2(x)}{2}.$$

Serviceable bounds for these functions can be easily derived from [13, Section VII.3, equation (1)] and the prior equations, showing that the lemma is true for $x \ge 6$, say. The computations establishing the lemma for the remaining range can be done to any desired accuracy by computer. The smallest value of the three functions for $\kappa \in [0, \frac{1}{2}]$ is $-J_0''(\frac{1}{2}) > 0.45$, while the closest any of these functions come to violating the asserted inequality is the local minimum of $-J_0''(x)$ near $x = 3.5$, at which $-J_0''(x) > -0.42$. ∎

The following convergent infinite products of Bessel functions is central in the subject of prime number races.

***Definition 4.5***   For any Dirichlet character $\chi$, define

$$F(z, \chi) = \prod_{\substack{\gamma > 0 \\ L(\frac{1}{2} + i\gamma, \chi) = 0}} J_0\left( \frac{2z}{\sqrt{\frac{1}{4} + \gamma^2}} \right).$$

Then define, for any permutation $(i, j, k)$ of $(1, 2, 3)$,

$$\Phi_i(z) = \prod_{\chi \pmod q} F\big(|\chi(a_i) + \chi(a_0) - \chi(a_j) - \chi(a_k)|z, \chi\big) = \prod_{\chi \in H_i} F(4z, \chi),$$

where the last equality holds by Lemma 3.4. The products defining $F(z, \chi)$ and $\Phi_i(z)$ converge uniformly on bounded subsets of the complex plane (a fact that will follow from the upper bounds we establish below for these functions).

We can immediately see the relevance of this function to the characteristic functions $\hat{X}_i(z)$ of the random variables $X_i$ from Definition 3.7.

***Proposition 4.6***   *For $i \in \{1, 2, 3\}$, we have $\hat{X}_i(z) = \Phi_i(z)$, where $\Phi_i(z)$ is as in Definition 4.5.*

**Proof**   An extremely similar computation is carried out in [3, Proposition 2.13] as well as in other sources; the key observations are that the $X_\gamma$ are independent (so that the characteristic function of the double sum defining $X_i$ is the product of the

individual characteristic functions) and that the characteristic function of $cX_\gamma$ equals $J_0(cz)$ for any constant $c \in \mathbb{R}$.                                                                                       ∎

We proceed as in [3, Propositions 2.10–2.12], writing the power series of the logarithms of these infinite products in terms of the following quantities.

**Definition 4.7**  For $i \in \{1, 2, 3\}$ and any positive integer $m$, define

$$W_i(m) = \frac{8^{2m}|\lambda_{2m}|}{V_i} \sum_{\substack{\chi \in H_i}} \sum_{\substack{\gamma > 0 \\ L(\frac{1}{2}+i\gamma, \chi)=0}} \frac{1}{(\frac{1}{4}+\gamma^2)^m},$$

with $\lambda_n$ as in Definition 4.1.

**Lemma 4.8**  *We have $W_i(1) = \frac{1}{2}$ and $W_i(m) \ll (\frac{20}{3})^{2m}$ for all $m \geq 2$.*

**Proof**  We compute directly from Definition 4.1 that $\lambda_2 = -\frac{1}{4}$, and thus

$$W_i(1) = \frac{8^2|\lambda_2|}{V_i} \cdot \sum_{\substack{\chi \in H_i}} \sum_{\substack{\gamma > 0 \\ L(\frac{1}{2}+i\gamma, \chi)=0}} \frac{1}{\frac{1}{4}+\gamma^2} = \frac{16}{V_i} \cdot \frac{V_i}{32} = \frac{1}{2}$$

by Definition 1.4. On the other hand, since $\frac{1}{4} + \gamma^2 \geq \frac{1}{4}$, by Lemma 4.2, we have

$$W_i(m) = \frac{8^{2m}|\lambda_{2m}|}{V_i} \sum_{\substack{\chi \in H_i}} \sum_{\substack{\gamma > 0 \\ L(\frac{1}{2}+i\gamma, \chi)=0}} \frac{1}{(\frac{1}{4}+\gamma^2)^m}$$

$$\ll \frac{8^{2m}}{V_i}\left(\frac{5}{12}\right)^{2m} \sum_{\substack{\chi \in H_i}} \sum_{\substack{\gamma > 0 \\ L(\frac{1}{2}+i\gamma, \chi)=0}} \frac{4^{m-1}}{\frac{1}{4}+\gamma^2}$$

$$= \frac{1}{V_i}\left(\frac{10}{3}\right)^{2m} 4^{m-1} \cdot \frac{V_i}{32} = \frac{1}{128}\left(\frac{20}{3}\right)^{2m}.$$                                       ∎

The next lemma codifies the standard fact that power series can be estimated by their first terms in compact subsets inside their open disks of convergence.

**Lemma 4.9**  *Let $i \in \{1, 2, 3\}$. For any integers $k \geq 0$ and $d \geq -2k$ and any polynomial $P(x)$,*

$$\sum_{m=k}^{\infty} P(m)W_i(m)z^{2m+d} \ll_{P,k} |z|^{2k+d} \quad \text{uniformly for } |z| \leq \tfrac{1}{10},$$

$$\exp\left(-V_i \sum_{m=k}^{\infty} P(m)W_i(m)z^{2m+d}\right) = 1 + O_{P,k}(V_i|z|^{2k+d}) \quad \text{uniformly for } |z| \leq V_i^{-1/(2k+d)}.$$

**Proof**  By Lemma 4.8, we know that $P(m)W_i(m) \ll |P(m)|(\frac{20}{3})^{2m} \ll_P 7^m$. Thus, when $|z| \leq \frac{1}{10}$, we obtain

$$\sum_{m=k}^{\infty} P(m)W_i(m)z^{2m+d} \ll_P \sum_{m=k}^{\infty} 7^m \left(\frac{1}{10}\right)^{2(m-k)} |z|^{2k+d} = \frac{100}{93}7^k|z|^{2k+d},$$

which suffices for the first bound. The second bound follows from the first because $e^w = 1 + O_{P,k}(|w|)$ uniformly for $|w| \ll_{P,k} 1$. ∎

The final result of this section is the connection between the characteristic functions $\Phi_i(z)$ and the quantities $W_i(m)$.

**Proposition 4.10** *For* $|z| < \frac{3}{20}$ *and* $i \in \{1, 2, 3\}$, *we have* $\Phi_i(z) = \exp\left(-V_i \sum_{m=1}^{\infty} W_i(m) z^{2m}\right)$.

**Proof** From Definition 4.5,

$$\log \Phi_i(z) = \log\left(\prod_{\chi \in H_i} F(4z, \chi)\right) = \sum_{\chi \in H_i} \sum_{\substack{\gamma > 0 \\ L(\frac{1}{2} + i\gamma, \chi) = 0}} \log J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right).$$

For $|z| < \frac{3}{20}$, the argument of $J_0$ is less than $8 \cdot \frac{3}{20} / \frac{1}{2} = \frac{12}{5}$, and so the power series expansion of $\log J_0$ converges absolutely by Lemma 4.2(a), giving

$$\log \Phi_i(z) = \sum_{\chi \in H_i} \sum_{\substack{\gamma > 0 \\ L(\frac{1}{2} + i\gamma, \chi) = 0}} \sum_{n=0}^{\infty} \lambda_n \left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right)^n.$$

By Lemma 4.2(b)–(c), the $n = 0$ term and the terms with $n$ odd vanish, and we may change $\lambda_n$ for $n$ even to $-|\lambda_{2m}|$. Since the sum over $\gamma$ converges absolutely for $n \geq 2$, we may rearrange terms to obtain

$$\log \Phi_i(z) = -\sum_{m=1}^{\infty} |\lambda_{2m}| (8z)^{2m} \sum_{\chi \in H_i} \sum_{\substack{\gamma > 0 \\ L(\frac{1}{2} + i\gamma, \chi) = 0}} \frac{1}{(\frac{1}{4} + \gamma^2)^m}$$

$$= -V_i \sum_{m=1}^{\infty} z^{2m} |\lambda_{2m}| \frac{8^{2m}}{V_i} \sum_{\chi \in H_i} \sum_{\substack{\gamma > 0 \\ L(\frac{1}{2} + i\gamma, \chi) = 0}} \frac{1}{(\frac{1}{4} + \gamma^2)^m} = -V_i \sum_{m=1}^{\infty} z^{2m} W_i(m)$$

by Definition 4.7. ∎

# 5 Comparison of characteristic functions

The goal of this section is to obtain pointwise bounds for the difference between the characteristic function $\Phi_i(x)$ and the characteristic function $e^{-V_i x^2/2}$ of a normal random variable with mean 0 and variance $V_i$. We begin by establishing the asymptotic sizes of these variances.

**Lemma 5.1** *Assume GRH. If $q \geq 3$, then $b(\chi) = \log q_{\chi^*} + O(\log \log q)$, where $q_{\chi^*}$ is the conductor of $\chi$.*

**Proof** According to [3, Lemma 3.5 and the proof of Proposition 3.6], on GRH we have

$$b(\chi) = \log q_{\chi^*} + 2\Re \frac{L'(1, \chi^*)}{L(1, \chi^*)} + O(1)$$

$$= \log q_{\chi^*} + O(\log \log q_{\chi^*}) + O(1) = \log q_{\chi^*} + O(\log \log q)$$

for $q_{\chi^*} > 1$. Since $b(\chi) = O(1)$ when $q_{\chi^*} = 1$, we conclude that $b(\chi) = \log q_{\chi^*} + O(\log\log q)$ for all characters $\chi \pmod q$. ∎

We can now establish the sizes of the quantities from Definition 1.4.

**Proposition 5.2**  *Assume GRH. We have* $V(q) = \phi(q)\log q + O(\phi(q)\log\log q)$ *and* $V_i = 4\phi(q)\log q + O(\phi(q)\log\log q)$. *In particular,* $\eta_i \ll (\log\log q)/\log q$.

**Proof**  It suffices to prove the two asymptotic formulas, as then the estimate for $\eta_i$ follows directly from Definition 1.4.

If $S$ is any set of characters $\pmod q$ such that $\chi \in S$ if and only if $\overline{\chi} \in S$, then

$$2\sum_{\chi \in S} b_+(\chi) = \sum_{\chi \in S} b_+(\chi) + \sum_{\overline{\chi} \in S} b_+(\overline{\chi}) = \sum_{\chi \in S} b_+(\chi) + \sum_{\chi \in S} b_+(\overline{\chi}) = \sum_{\chi \in S} b(\chi)$$

as noted in Definition 1.1. Then, by Lemma 5.1,

$$2\sum_{\chi \in S} b_+(\chi) = \sum_{\chi \in S} \left( \log q_{\chi^*} + O(\log\log q) \right)$$

$$= \sum_{\chi \in S} \log q - \sum_{\chi \in S}(\log q - \log q_{\chi^*}) + O(\#S \log\log q)$$

$$\text{(5.1)} \qquad = \#S\log q + O\left( \sum_{\chi \in S}(\log q - \log q_{\chi^*}) \right) + O(\phi(q)\log\log q).$$

However, [3, Proposition 3.3 and the proof of Proposition 3.6] implies that

$$\sum_{\chi \pmod q} (\log q - \log q_{\chi^*}) = \phi(q) \sum_{p | q} \frac{\log p}{p-1} \ll \phi(q)\log\log q,$$

and so (since $\log q - \log q_{\chi^*}$ is nonnegative) the first error term can be absorbed into the second.

In particular, combining Definition 1.4 with equation (5.1) yields

$$V(q) = 2 \sum_{\substack{\chi \pmod q \\ \chi \neq \chi_0}} b_+(\chi) = (\phi(q) - 1)\log q + O(\phi(q)\log\log q)$$

$$V_i = 32 \sum_{\chi \in H_i} b_+(\chi) = 16\left( \tfrac{1}{4}\phi(q)\log q + O(\phi(q)\log\log q) \right),$$

since $\#H_i = \tfrac{1}{4}\phi(q)$ by Remark 1.3. ∎

In our proofs, we will need $q$ to be sufficiently large for some of our inequalities to hold; the following quantity $q_0$ will be used through the end of Section 8.

**Definition 5.3**  We define a positive real number $q_0$ as follows. By Proposition 5.2, we know that $V_i \gg \phi(q)\log q$ uniformly for all choices of $a_i, a_j, a_k$ from Definition 3.1. Therefore, we can choose $q_0 > 0$ so that $V_i \geq \max\{2^{20}, \phi(q)\}$ for all $q > q_0$. We will often use (without comment) the specific consequence that $V_i^{-1/4} \leq \frac{1}{32}$ for $q > q_0$.

We proceed now to establish several estimates for $\Phi_i(x)$ valid for various ranges of $x$. The first such formula, for arguments close to 0, is similar to [3, Proposition 2.12].

**Proposition 5.4** *Assume GRH. For $i \in \{1,2,3\}$ and $q > q_0$, we have $\Phi_i(z) = e^{-V_i z^2/2}\left(1 + O\left(V_i|z|^4\right)\right)$ for all complex numbers $z$ with $|z| \le V_i^{-1/4}$.*

**Proof** Since

$$\Phi_i(z) = \exp\left(-V_i \sum_{m=1}^{\infty} W_i(m)z^{2m}\right) = e^{-V_i W_i(1)z^2} \exp\left(-V_i \sum_{m=2}^{\infty} W_i(m)z^{2m}\right)$$

by Proposition 4.10, and $W_i(1) = \frac{1}{2}$ by Lemma 4.8, the estimate $\Phi_i(z) = e^{-V_i z^2/2} \exp(O(V_i|z|^4))$ (which implies the asserted statement) follows immediately from Lemma 4.9. ∎

**Lemma 5.5** *Assume GRH. For $i \in \{1,2,3\}$ and $q > q_0$, we have $|\Phi_i(x)| \ll e^{-V_i^{1/2}/2}$ for all real numbers $x$ with $|x| \ge V_i^{-1/4}$.*

**Proof** From Definition 4.5,

$$\Phi_i(x) = \prod_{\chi \in H_i} \prod_{\substack{\gamma > 0 \\ L(\frac{1}{2}+i\gamma,\chi)=0}} J_0\left(\frac{8x}{\sqrt{\frac{1}{4} + \gamma^2}}\right).$$

In each factor, set $\kappa = 8V_i^{-1/4}/\sqrt{\frac{1}{4} + \gamma^2}$. We have $0 \le \kappa \le 8V_i^{-1/4}/\frac{1}{2} \le 16 \cdot \frac{1}{32} = \frac{1}{2}$ since $q > q_0$, and therefore Lemma 4.4 applies to each factor, yielding

$$|\Phi_i(x)| = \prod_{\chi \in H_i} \prod_{\substack{\gamma > 0 \\ L(\frac{1}{2}+i\gamma,\chi)=0}} \left|J_0\left(\frac{8x}{\sqrt{\frac{1}{4} + \gamma^2}}\right)\right|$$

$$\le \prod_{\chi \in H_i} \prod_{\substack{\gamma > 0 \\ L(\frac{1}{2}+i\gamma,\chi)=0}} J_0\left(\frac{8V_i^{-1/4}}{\sqrt{\frac{1}{4} + \gamma^2}}\right) = \Phi_i\left(V_i^{-1/4}\right)$$

for $|x| \ge V_i^{-1/4}$. The lemma now follows from the estimate $\Phi_i\left(V_i^{-1/4}\right) \ll e^{-V_i^{1/2}/2}$, which is a special case of Proposition 5.4. ∎

The following lemma could be proved directly (derived from [3, Lemma 2.16], for example); however, we will need a more general result later, so it is more efficient to derive this proposition from that later result.

**Lemma 5.6** *Assume GRH. For $i \in \{1,2,3\}$, we have $|\Phi_i(x)| < e^{-\phi(q)|x|/8}$ for all real numbers $x$ with $|x| \ge 50$.*

**Proof** The lemma follows immediately from equation (6.6) and Lemma 6.8 with $A = \emptyset$. ∎

Ultimately, we will want to compare the characteristic function $\Phi_i(x)$ to the characteristic function $e^{-V_i x^2/2}$ of a normal random variable with mean 0 and variance $V_i$. The following proposition summarizes the results of this section in that light.

**Proposition 5.7**    *Assume GRH. For $i \in \{1, 2, 3\}$ and $q > q_0$, and for any real number $x$,*

$$\Phi_i(x) - e^{-V_i x^2/2} \ll \begin{cases} V_i x^4 e^{-V_i x^2/2}, & if\ |x| \leq V_i^{-1/4}, \\ e^{-V_i^{1/2}/2}, & if\ V_i^{-1/4} \leq |x| \leq 50, \\ e^{-\phi(q)|x|/8}, & if\ |x| \geq 50. \end{cases}$$

**Proof**    The first assertion is immediate from Proposition 5.4. For the second and third assertions, we use $|\Phi_i(x) - e^{-V_i x^2/2}| \leq |\Phi_i(x)| + e^{-V_i x^2/2}$, and note that the term $e^{-V_i x^2/2}$ is insignificant compared with the asserted estimates (due to the range of $x$ in the second case and the definition of $q_0$ in the third case). Therefore, the second assertion follows from Lemma 5.5 while the third assertion follows from Lemma 5.6.                                                                                    ∎

# 6  Comparison of second derivatives of characteristic functions

We continue to use the methods of the previous section, now with the goal of providing analogous bounds for $\Phi_i''(x)$ for various ranges of $x$, with an eye toward an eventual comparison with the second derivative of $e^{-V_i x^2/2}$. We are fortunate to have access to several different representations of $\Phi_i''(x)$, as no one of them will be entirely sufficient for our needs. We begin with the following power series representation.

**Lemma 6.1**    *For $i \in \{1, 2, 3\}$ and $|z| < \frac{3}{20}$,*

$$\Phi''(z) = \Phi_i(z) \left\{ \left( V_i \sum_{m=1}^{\infty} 2m W_i(m) z^{2m-1} \right)^2 - V_i \sum_{m=1}^{\infty} 2m(2m-1) W_i(m) z^{2m-2} \right\}.$$

**Proof**    We know that $(e^{h(z)})'' = e^{h(z)}(h'(z)^2 + h''(z))$ for any smooth function $h(z)$. The proposition follows from applying this identity with $h(z)$ equal to the power series in the exponent of the formula for $\Phi_i(z)$ given in Proposition 4.10, which can be differentiated term-by-term on any open set on which it converges.                                                                                    ∎

**Lemma 6.2**    *For $i \in \{1, 2, 3\}$ and $q > q_0$, we have*

$$\Phi''(z) = e^{-V_i z^2/2} \left( V_i^2 z^2 - V_i + O(V_i|z|^2 + V_i^3|z|^6) \right)$$

*for $|z| \leq V_i^{-1/4}$.*

**Proof**    Using Proposition 5.4 followed by Lemma 4.9 twice, we see that for $|z| \leq V_i^{-1/4}$,

$$\Phi_i(z) = e^{-V_i z^2/2} \left( 1 + O(V_i|z|^4) \right)$$

$$\sum_{m=1}^{\infty} 2m W_i(m) z^{2m-1} = z + O(|z|^3)$$

$$\sum_{m=1}^{\infty} 2m(2m-1) W_i(m) z^{2m-2} = 1 + O(|z|^2)$$

since $W_i(1) = \frac{1}{2}$. (The second and third formulas require $|z| \leq \frac{1}{10}$, which is implied by $|z| \leq V_i^{-1/4}$ since $q > q_0$.) Therefore, by Lemma 6.1, for $|z| \leq V_i^{-1/4}$ we have

$$
\Phi_i(z)\left(V_i \sum_{m=1}^{\infty} 2m W_i(m) z^{2m-1}\right)^2 = e^{-V_i z^2/2}\left(1 + O(V_i|z|^4)\right)\left(V_i(z + O(|z|^3))\right)^2
$$

$$
(6.1) \qquad\qquad = e^{-V_i z^2/2}\left(V_i^2 z^2 + O(V_i^2|z|^4 + V_i^3|z|^6)\right)
$$

and ultimately

$$
\begin{aligned}
\Phi''(z) &= e^{-V_i z^2/2}\left(1 + O(V_i|z|^4)\right)\left\{\left(V_i(z + O(|z|^3))\right)^2 - \left(V_i(1 + O(|z|^2))\right)\right\} \\
&= e^{-V_i z^2/2}\left(1 + O(V_i|z|^4)\right)\left(V_i^2 z^2 - V_i + O(V_i|z|^2 + V_i^2|z|^4)\right) \\
&= e^{-V_i z^2/2}\left(V_i^2 z^2 - V_i + O(V_i|z|^2 + V_i^2|z|^4 + V_i^3|z|^6)\right),
\end{aligned}
$$

which implies the statement of the proposition since $V_i^2|z|^4$ is always dominated by one of the other two error terms. ∎

In the proof of Lemma 5.5, we used the fact that $\Phi(x)$ was a simple product of terms all of which were positive, and took their largest values, near the origin. The corresponding expression for $\Phi_i''(x)$ is more complicated, however, and involves functions whose values near the origin have both signs. We, therefore, establish a particular decomposition of $\Phi_i''(x)$ into two pieces, each of which has the unanimity of sign necessary for us to infer from Lemma 4.4 that its largest values are near the origin.

It will be convenient to define the set of ordinates

$$
(6.2) \qquad\qquad U_i = \bigcup_{\chi \in H_i}\{\gamma > 0 : L(1/2 + i\gamma, \chi) = 0\}
$$

that indexes the infinite product that defines $\Phi_i(z)$.

**Lemma 6.3** *For $i \in \{1, 2, 3\}$ and $z \in \mathbb{C}$ we may write*

$$
(6.3) \qquad\qquad \Phi_i''(z) = z^2 \Psi_i(z) + \Theta_i(z),
$$

*where*

(6.4)

$$
\Psi_i(z) = \sum_{\substack{\gamma_1, \gamma_2 \in U_i \\ \gamma_1 \neq \gamma_2}} \frac{64}{\frac{1}{4} + \gamma_1^2} K\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma_1^2}}\right) \frac{64}{\frac{1}{4} + \gamma_2^2} K\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma_2^2}}\right) \prod_{\gamma \in U_i \setminus \{\gamma_1, \gamma_2\}} J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right)
$$

$$
(6.5) \qquad \Theta_i(z) = \sum_{\gamma_1 \in U_i} \frac{64}{\frac{1}{4} + \gamma_1^2} J_0''\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma_1^2}}\right) \prod_{\gamma \in U_i \setminus \{\gamma_1\}} J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right).
$$

**Proof** By Definition 4.5,

$$
(6.6) \qquad \Phi_i(z) = \prod_{\chi \in H_i} \prod_{\substack{\gamma > 0 \\ L(1/2 + i\gamma, \chi) = 0}} J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right) = \prod_{\gamma \in U_i} J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right).
$$

This infinite product of analytic functions converges uniformly in any bounded subset of $\mathbb{C}$ (since $J_0(z) = 1 + O(z^2)$ near $z = 0$ and the series $\sum_{\gamma \in U_i} 1/\gamma^2$ converges). Therefore, we may differentiate the infinite product by applying the product rule twice:

$$\Phi''(z) = \sum_{\gamma_1 \in U_i} \frac{d^2}{dz^2} J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma_1^2}}\right) \prod_{\gamma \in U_i \setminus \{\gamma_1\}} J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right)$$

$$+ \sum_{\substack{\gamma_1, \gamma_2 \in U_i \\ \gamma_1 \neq \gamma_2}} \frac{d}{dz} J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma_1^2}}\right) \frac{d}{dz} J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma_2^2}}\right) \prod_{\gamma \in U_i \setminus \{\gamma_1, \gamma_2\}} J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right)$$

$$= \sum_{\gamma_1 \in U_i} \frac{64}{\frac{1}{4} + \gamma_1^2} J_0''\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma_1^2}}\right) \prod_{\gamma \in U_i \setminus \{\gamma_1\}} J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right)$$

$$+ \sum_{\substack{\gamma_1, \gamma_2 \in U_i \\ \gamma_1 \neq \gamma_2}} \frac{8}{\sqrt{\frac{1}{4} + \gamma_1^2}} J_0'\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma_1^2}}\right) \frac{8}{\sqrt{\frac{1}{4} + \gamma_2^2}} J_0'\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma_2^2}}\right)$$

$$\times \prod_{\gamma \in U_i \setminus \{\gamma_1, \gamma_2\}} J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right).$$

Consulting Definition 4.3 reveals that this last expression is the same as equation (6.3) (the negative signs in the definition of $K(t)$ come in pairs). ∎

To efficiently bound, for small $|z|$, the first component $z^2 \Psi_i(z)$ in the above decomposition of $\Phi''(z)$, we need to first write it in a different form. Recall the function $D(x)$ from Definition 4.3.

**Lemma 6.4** *For $i \in \{1, 2, 3\}$ and complex numbers $z$ satisfying $|z| \leq \frac{3}{20}$,*

(6.7)

$$z^2 \Psi_i(z) = \Phi_i(z) \left\{ \left( \sum_{\gamma \in U_i} \frac{8}{\sqrt{\frac{1}{4} + \gamma^2}} D\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right) \right)^2 - \sum_{\gamma \in U_i} \frac{64}{\frac{1}{4} + \gamma^2} D\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right)^2 \right\}.$$

*In particular, for real numbers $x$ satisfying $|x| \leq \frac{1}{32}$,*

(6.8)
$$x^2 \Psi_i(x) \leq \Phi_i(x) \left\{ \left( \sum_{\gamma \in U_i} \frac{8}{\sqrt{\frac{1}{4} + \gamma^2}} D\left(\frac{8x}{\sqrt{\frac{1}{4} + \gamma^2}}\right) \right)^2 \right\}.$$

**Proof** Again we use $\left(e^{h(z)}\right)'' = e^{h(z)}\left(h'(z)^2 + h''(z)\right)$, this time with $h(z)$ equal to the infinite series in the identity

$$\Phi_i(z) = \exp\left( \sum_{\gamma \in U_i} \log J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right) \right),$$

valid for $|z| \leq \frac{3}{20}$ since the argument of $J_0$ does not vanish there; we obtain

$$\Phi_i''(z) = \Phi_i(z) \left( \sum_{\gamma \in U_i} \frac{8}{\sqrt{\frac{1}{4} + \gamma^2}} \frac{J_0' \left( 8z/\sqrt{\frac{1}{4} + \gamma^2} \right)}{J_0 \left( 8z/\sqrt{\frac{1}{4} + \gamma^2} \right)} \right)^2$$

$$+ \Phi_i(z) \sum_{\gamma \in U_i} \frac{64}{\frac{1}{4} + \gamma^2} \frac{J_0 \left( 8z/\sqrt{\frac{1}{4} + \gamma^2} \right) J_0'' \left( 8z/\sqrt{\frac{1}{4} + \gamma^2} \right) - J_0' \left( 8z/\sqrt{\frac{1}{4} + \gamma^2} \right)^2}{J_0 \left( 8z/\sqrt{\frac{1}{4} + \gamma^2} \right)^2}$$

$$= \Phi_i(z) \left\{ \left( \sum_{\gamma \in U_i} \frac{8}{\sqrt{\frac{1}{4} + \gamma^2}} D\left( \frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}} \right) \right)^2 - \sum_{\gamma \in U_i} \frac{64}{\frac{1}{4} + \gamma^2} D\left( \frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}} \right)^2 \right\}$$

$$(6.9) \qquad + \Phi_i(z) \sum_{\gamma \in U_i} \frac{64}{\frac{1}{4} + \gamma^2} \left( \frac{J_0'' \left( 8z/\sqrt{\frac{1}{4} + \gamma^2} \right)}{J_0 \left( 8z/\sqrt{\frac{1}{4} + \gamma^2} \right)} \right)$$

by Definition 4.3 (the negative signs all occur in pairs). However, using equation (6.6) yields

$$\Phi_i(z) \sum_{\gamma \in U_i} \frac{64}{\frac{1}{4} + \gamma^2} \left( \frac{J_0'' \left( 8z/\sqrt{\frac{1}{4} + \gamma^2} \right)}{J_0 \left( 8z/\sqrt{\frac{1}{4} + \gamma^2} \right)} \right)$$

$$= \prod_{\gamma \in U_i} J_0 \left( \frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}} \right) \sum_{\gamma_1 \in U_i} \frac{64}{\frac{1}{4} + \gamma_1^2} J_0 \left( \frac{8z}{\sqrt{\frac{1}{4} + \gamma_1^2}} \right)^{-1} J_0'' \left( \frac{8z}{\sqrt{\frac{1}{4} + \gamma_1^2}} \right) = \Theta_i(z)$$

by equation (6.5); thus by the identity (6.3), we conclude that the expression on line (6.9) must equal $z^2 \Psi_i(z)$, establishing the first assertion of the lemma.

As for the second assertion, when $z = x$ is a real number satisfying $|x| \leq \frac{1}{32}$, all of the summands in the two series in equation (6.9) are positive by Lemma 4.4, since the argument of $D$ is at most $16|x| \leq \frac{1}{2}$ in absolute value. Notice that the second sum in equation (6.9) consists precisely of the squares of the summands from the first sum; in particular, both sums are positive and the second sum is no larger than the square of the first sum. We may, therefore, ignore the second sum when finding an upper bound, which establishes the second assertion of the lemma. ∎

**Lemma 6.5** *For $i \in \{1, 2, 3\}$ and $q > q_0$, we have $\Psi_i(x) \ll e^{-V_i^{1/2}/2} V_i^{3/2}$ for $|x| \geq V_i^{-1/4}$.*

**Proof** In each factor in equation (6.4), set $\kappa = 8V_i^{-1/4} / \sqrt{\frac{1}{4} + \gamma^2}$. We have $0 \leq \kappa \leq 8V_i^{-1/4} / \frac{1}{2} \leq 16 \cdot \frac{1}{32} = \frac{1}{2}$ since $q > q_0$, and therefore Lemma 4.4 applies to each factor, yielding

$$|\Psi_i(x)| = \sum_{\substack{\gamma_1, \gamma_2 \in U_i \\ \gamma_1 \neq \gamma_2}} \frac{64}{\frac{1}{4} + \gamma_1^2} K\left( \frac{8x}{\sqrt{\frac{1}{4} + \gamma_1^2}} \right) \frac{64}{\frac{1}{4} + \gamma_2^2} K\left( \frac{8x}{\sqrt{\frac{1}{4} + \gamma_2^2}} \right) \prod_{\gamma \in U_i \setminus \{\gamma_1, \gamma_2\}} J_0 \left( \frac{8x}{\sqrt{\frac{1}{4} + \gamma^2}} \right)$$

$$\leq \sum_{\substack{\gamma_1,\gamma_2 \in U_i \\ \gamma_1 \neq \gamma_2}} \frac{64}{\frac{1}{4}+\gamma_1^2} K\left(\frac{8V_i^{-1/4}}{\sqrt{\frac{1}{4}+\gamma_1^2}}\right) \frac{64}{\frac{1}{4}+\gamma_2^2} K\left(\frac{8V_i^{-1/4}}{\sqrt{\frac{1}{4}+\gamma_2^2}}\right) \prod_{\gamma \in U_i \setminus \{\gamma_1,\gamma_2\}} J_0\left(\frac{8V_i^{-1/4}}{\sqrt{\frac{1}{4}+\gamma^2}}\right)$$

$$= \Psi_i(V_i^{-1/4})$$

for $|x| \geq V_i^{-1/4}$. Since $V_i^{-1/4} \leq \frac{1}{32}$ when $q > q_0$, we may apply the upper bound (6.8) to obtain

$$\Psi_i(V_i^{-1/4}) \leq \Phi_i(V_i^{-1/4}) \left( \sum_{\gamma \in U_i} \frac{8}{\sqrt{\frac{1}{4}+\gamma^2}} D\left(\frac{8V_i^{-1/4}}{\sqrt{\frac{1}{4}+\gamma^2}}\right) \right)^2.$$

It follows from Definitions 4.1 and 4.3 that

$$D(t) = -\frac{J_0'(t)}{J_0(t)} = -\frac{d}{dt}\log J_0(t) = \sum_{m=1}^{\infty} |\lambda_{2m}| 2m t^{2m-1}$$

for $|t| \leq \frac{12}{5}$, and so

$$\Psi_i(V_i^{-1/4}) \leq \Phi_i(V_i^{-1/4}) \left( \sum_{\gamma \in U_i} \frac{8}{\sqrt{\frac{1}{4}+\gamma^2}} \sum_{m=1}^{\infty} |\lambda_{2m}| 2m \left(\frac{8V_i^{-1/4}}{\sqrt{\frac{1}{4}+\gamma^2}}\right)^{2m-1} \right)^2$$

$$= \Phi_i(V_i^{-1/4}) \left( \sum_{m=1}^{\infty} |\lambda_{2m}| 2m \cdot 8^{2m} V_i^{-(2m-1)/4} \sum_{\gamma \in U_i} \frac{1}{(\frac{1}{4}+\gamma^2)^m} \right)^2$$

$$= \Phi_i(V_i^{-1/4}) \left( V_i \sum_{m=1}^{\infty} 2m W_i(m) V_i^{-(2m-1)/4} \right)^2 \ll \Phi_i(V_i^{-1/4}) \left( V_i \cdot V_i^{-1/4} \right)^2$$

by Lemma 4.9. The statement of the proposition now follows from Proposition 5.4. ∎

The estimates we have derived for $\Phi_i''(x)$ and $\Psi_i(x)$ for small $|x|$ imply a similar estimate for $\Theta_i(x)$ for small $|x|$; thanks to Lemma 4.4, we can deduce an estimate for $\Theta_i(x)$ for large $|x|$, which we can subsequently use to estimate $\Phi_i''(x)$ itself for larger $|x|$.

**Lemma 6.6** *For $i \in \{1,2,3\}$ and $q > q_0$, we have $\Theta_i(x) \ll e^{-V_i^{1/2}/2} V_i^{3/2}$ for $|x| \geq V_i^{-1/4}$.*

**Proof** In each factor in equation (6.5), set $\kappa = 8V_i^{-1/4}/\sqrt{\frac{1}{4}+\gamma^2}$. We have $0 \leq \kappa \leq 8V_i^{-1/4}/\frac{1}{2} \leq 16 \cdot \frac{1}{32} = \frac{1}{2}$ since $q > q_0$, and therefore Lemma 4.4 applies to each factor, yielding

$$|\Theta_i(x)| = \sum_{\gamma_1 \in U_i} \frac{64}{\frac{1}{4}+\gamma_1^2} J_0''\left(\frac{8x}{\sqrt{\frac{1}{4}+\gamma_1^2}}\right) \prod_{\gamma \in U_i \setminus \{\gamma_1\}} J_0\left(\frac{8x}{\sqrt{\frac{1}{4}+\gamma^2}}\right)$$

$$\leq \sum_{\gamma_1 \in U_i} \frac{64}{\frac{1}{4}+\gamma_1^2} J_0''\left(\frac{8x}{\sqrt{\frac{1}{4}+\gamma_1^2}}\right) \prod_{\gamma \in U_i \setminus \{\gamma_1\}} J_0\left(\frac{8x}{\sqrt{\frac{1}{4}+\gamma^2}}\right) = \Theta_i(V_i^{-1/4})$$

for $|x| \geq V_i^{-1/4}$. On the other hand, by the identity (6.3) and Lemmas 6.2 and 6.5, we have

$$\Theta_i(V_i^{-1/4}) = \Phi_i''(V_i^{-1/4}) - (V_i^{-1/4})^2 \Psi_i(V_i^{-1/4})$$
$$\ll e^{-V_i^{1/2}/2} V_i^{3/2} + V_i^{-1/2} e^{-V_i^{1/2}/2} V_i^{3/2} \ll e^{-V_i^{1/2}/2} V_i^{3/2}$$

as desired. ∎

**Lemma 6.7**  *For $i \in \{1, 2, 3\}$ and $q > q_0$, we have $\Phi_i''(x) \ll e^{-V_i^{1/2}/2} V_i^{3/2}$ for $V_i^{-1/4} \leq |x| \leq 50$.*

**Proof**  The lemma follows immediately from the identity (6.3) and Lemmas 6.5 and 6.6, since $x \ll 1$ by assumption. ∎

Lastly, we use a standard method to estimate $\Phi_i''(x)$ for the largest values of $|x|$. The proof is complicated only slightly by the fact that the relevant infinite products of Bessel functions are missing a small number of terms after the differentiations; the following lemma provides a serviceable bound that uniformly takes such omitted terms into account.

**Lemma 6.8**  *Fix $i \in \{1, 2, 3\}$. If $A$ is any finite subset of $U_i$, then for $|x| \geq 50$,*

$$\prod_{\gamma \in U_i \setminus A} J_0\left(\frac{8x}{\sqrt{\frac{1}{4} + \gamma^2}}\right) < 2^{\#A} e^{-\phi(q)|x|/8}.$$

**Proof**  Since both sides are even functions of $x$, we may assume that $x \geq 50$. Let $N(T, \chi)$ denote the number of nontrivial zeros of $L(s, \chi)$ having imaginary part between $-T$ and $T$. By [2, Proposition 2.5], for $T \geq 150$,

(6.10)
$$N(T, \chi) \geq \left(\frac{T}{\pi} - 0.399\right) \log \frac{q_{\chi^*} T}{2\pi e} - 5.338 \geq \left(\frac{T}{\pi} - 0.399\right) \log \frac{T}{2\pi e} - 5.338 > \frac{T}{2}$$

where $q_{\chi^*} \geq 1$ is the conductor of $\chi$. From the classical inequality (see [12, Theorem 7.31.2])

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

we see that

$$\left|\prod_{\gamma \in U_i \setminus A} J_0\left(\frac{8x}{\sqrt{\frac{1}{4} + \gamma^2}}\right)\right| \leq \prod_{\substack{\gamma < 3x \\ \gamma \in U_i \setminus A}} \left|J_0\left(\frac{8x}{\sqrt{\frac{1}{4} + \gamma^2}}\right)\right| \leq \prod_{\substack{\gamma < 3x \\ \gamma \in U_i \setminus A}} \frac{(\frac{1}{4} + \gamma^2)^{1/4}}{2\sqrt{\pi x}}.$$

One can easily check that when $|x| \geq 50$ and $|\gamma| < 3x$, the factor $(\frac{1}{4} + \gamma^2)^{1/4}/2\sqrt{\pi x}$ is always less than $\frac{1}{2}$.

If we define $N_+(T, \chi)$ to be the number of nontrivial zeros of $L(s, \chi)$ having imaginary part between 0 and $T$, then $N(T, \chi) = N_+(T, \chi) + N_+(T, \overline{\chi})$ by the functional

equation. Since $\chi \in H_i$ if and only if $\overline{\chi} \in H_i$, the number of factors in the product is

$$\sum_{\chi \in H_i} N_+(3x, \chi) - \#\{A \cup (-3x, 3x)\}$$

$$= \frac{1}{2}\left(\sum_{\chi \in H_i} N_+(3x, \chi) + \sum_{\overline{\chi} \in H_i} N_+(3x, \overline{\chi})\right) - \#\{A \cup (-3x, 3x)\}$$

$$= \frac{1}{2}\sum_{\chi \in H_i} N(3x, \chi) - \#\{A \cup (-3x, 3x)\} \geq \frac{\phi(q)}{8}\frac{3x}{2} - \#A.$$

So

$$\prod_{\substack{\gamma < 3x \\ \gamma \in U_i \setminus A}} \frac{(\frac{1}{4} + \gamma^2)^{1/4}}{2\sqrt{\pi|x|}} \leq 2^{-(3\phi(q)x/16 - \#A)} < 2^{\#A} e^{-\phi(q)x/8},$$

since $\#H_i = \phi(q)/4$ by Remark 1.3. ∎

**Lemma 6.9**   For $i \in \{1, 2, 3\}$, we have $\Phi_i''(x) \ll V_i^2 x^2 e^{-\phi(q)|x|/8}$ for $|x| \geq 50$.

**Proof**   From equations (6.4) and (6.5),

$$\Psi_i(x) = \sum_{\substack{\gamma_1, \gamma_2 \in U_i \\ \gamma_1 \neq \gamma_2}} \frac{64}{\frac{1}{4} + \gamma_1^2} K\left(\frac{8x}{\sqrt{\frac{1}{4} + \gamma_1^2}}\right) \frac{64}{\frac{1}{4} + \gamma_2^2} K\left(\frac{8x}{\sqrt{\frac{1}{4} + \gamma_2^2}}\right) \prod_{\gamma \in U_i \setminus \{\gamma_1, \gamma_2\}} J_0\left(\frac{8x}{\sqrt{\frac{1}{4} + \gamma^2}}\right),$$

$$\Theta_i(x) = \sum_{\gamma_1 \in U_i} \frac{64}{\frac{1}{4} + \gamma_1^2} J_0''\left(\frac{8x}{\sqrt{\frac{1}{4} + \gamma_1^2}}\right) \prod_{\gamma \in U_i \setminus \{\gamma_1\}} J_0\left(\frac{8x}{\sqrt{\frac{1}{4} + \gamma^2}}\right).$$

We apply equation (4.1) and Lemma 6.8 to obtain, for $|x| \geq 50$,

$$\Psi_i(x) \ll \sum_{\substack{\gamma_1, \gamma_2 \in U_i \\ \gamma_1 \neq \gamma_2}} \frac{32}{\frac{1}{4} + \gamma_1^2} \frac{32}{\frac{1}{4} + \gamma_2^2} e^{-\phi(q)|x|/8} \leq \left(\sum_{\gamma_1 \in U_i} \frac{32}{\frac{1}{4} + \gamma_1^2}\right)^2 e^{-\phi(q)|x|/8} = V_i^2 e^{-\phi(q)|x|/8},$$

$$\Theta_i(x) \ll \sum_{\gamma_1 \in U_i} \frac{32}{\frac{1}{4} + \gamma_1^2} e^{-\phi(q)|x|/8} = V_i e^{-\phi(q)|x|/8}.$$

So by equation (6.3),

$$\Phi_i''(x) = x^2 \Psi_i(x) + \Theta_i(x) \ll (V_i^2 x^2 + V_i)e^{-\phi(q)|x|/8},$$

which implies the statement of the lemma since $V_i \ll V_i^2 x^2$ in this range. ∎

**Remark 6.10**   The method of proof of Proposition 6.3 gives the expression

$$\Phi_i'(z) = z \sum_{\gamma_1 \in U_i} \frac{64}{\frac{1}{4} + \gamma_1^2} K\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma_1^2}}\right) \prod_{\gamma \in U_i \setminus \{\gamma_1\}} J_0\left(\frac{8z}{\sqrt{\frac{1}{4} + \gamma^2}}\right)$$

for the first derivative of $\Phi(z)$, from which the estimate $\Phi_i'(x) \ll V_i|x|e^{-\phi(q)|x|/8}$ for $|x| \geq 50$ follows from the method of proof of Lemma 6.9; in particular, $\Phi_i'(x)$ tends to 0 as $|x| \to \infty$, a fact we will need later.

We may now assemble the various bounds for $\Phi_i''(x)$ derived in this section to compare that function to $(V_i^2 z^2 - V_i) e^{-V_i z^2/2}$, which is the second derivative of the characteristic function $e^{-V_i z^2/2}$ of a normal random variable with mean 0 and variance $V_i$.

**Proposition 6.11** *For $i \in \{1, 2, 3\}$ and $q > q_0$, and for any real number $x$,*

$$\Phi_i''(x) - (V_i^2 x^2 - V_i) e^{-V_i x^2/2} \ll \begin{cases} V_i x^2 e^{-V_i x^2/2}, & \text{if } |x| \le V_i^{-1/2}, \\ V_i^3 x^6 e^{-V_i x^2/2}, & \text{if } V_i^{-1/2} \le |x| \le V_i^{-1/4}, \\ V_i^{3/2} e^{-V_i^{1/2}/2}, & \text{if } V_i^{-1/4} \le |x| \le 50, \\ V_i^2 x^2 e^{-\phi(q)|x|/8}, & \text{if } |x| \ge 50. \end{cases}$$

**Proof** The first two assertions are immediate from Lemma 6.2. For the third and fourth assertions, we use

$$|\Phi_i''(x) - (V_i^2 x^2 - V_i) e^{-V_i x^2/2}| \le |\Phi_i''(x)| + (V_i^2 x^2 - V_i) e^{-V_i x^2/2}$$
$$\le |\Phi_i''(x)| + V_i^2 x^2 e^{-V_i x^2/2},$$

and note that the term $V_i^2 x^2 e^{-V_i x^2/2}$ is insignificant compared to the asserted estimates (due to the range of $x$ in the third case and the definition of $q_0$ in the fourth case). Therefore, the third assertion follows from Lemma 6.7 while the fourth assertion follows from Lemma 6.9. ∎

## 7 Comparison of probabilities

We are now able to estimate the difference between probabilities involving the random variables $X_i$ from Definition 3.7 and normal random variables of the same mean and variance. Using the results of the previous sections, we will bound the integrals of their characteristic functions and the second derivatives thereof over $\mathbb{R}$; subsequently we will be able to bound the difference between their density functions themselves. We begin with a quick and standard lemma giving the order of magnitude of even moments of a normal distribution.

**Lemma 7.1** *For any positive constant $C$ and any nonnegative integer $m$,*

$$\int_{-\infty}^{\infty} x^{2m} e^{-Cx^2} \, dx \ll_m \frac{1}{C^{m+1/2}}.$$

**Proof** When $m = 0$, the formula $\int_{-\infty}^{\infty} e^{-Cx^2} \, dx = \sqrt{2\pi/C}$ is well known. For $m \ge 1$, we integrate by parts to obtain

$$\int_{-\infty}^{\infty} x^{2m} e^{-Cx^2/2} \, dx = \int_{-\infty}^{\infty} x^{2m-1} \cdot x e^{-Cx^2/2} \, dx = \int_{-\infty}^{\infty} (2m-1) x^{2m-2} \cdot \frac{2}{C} e^{-Cx^2/2} \, dx,$$

from which the lemma follows by induction on $m$. ∎

**Lemma 7.2**   *Assume GRH. For $i \in \{1, 2, 3\}$ and $q > q_0$, we have*

$$\int_{-\infty}^{\infty} \left| \Phi_i(x) - e^{-V_i x^2/2} \right| dx \ll V_i^{-3/2}$$

$$\int_{-\infty}^{\infty} \left| \Phi_i''(x) - (V_i^2 x^2 - V_i) e^{-V_i x^2/2} \right| dx \ll V_i^{-1/2}.$$

**Proof**   We write

$$\int_{-\infty}^{\infty} \left| \Phi_i(x) - e^{-V_i x^2/2} \right| dx = \int_{|x| \le V_i^{-1/4}} \left| \Phi_i(x) - e^{-V_i x^2/2} \right| dx$$

$$+ \int_{V_i^{-1/4} \le |x| \le 50} \left| \Phi_i(x) - e^{-V_i x^2/2} \right| dx + \int_{|x| \ge 50} \left| \Phi_i(x) - e^{-V_i x^2/2} \right| dx.$$

Using the bounds in Proposition 5.7,

$$\int_{-\infty}^{\infty} \left| \Phi_i(x) - e^{-V_i x^2/2} \right| dx \ll \int_{|x| \le V_i^{-1/4}} V_i x^4 e^{-V_i x^2/2} dx$$

$$+ \int_{V_i^{-1/4} \le |x| \le 50} e^{-V_i^{1/2}/2} dx + \int_{|x| \ge 50} e^{-\phi(q)|x|/8} dx$$

$$\le V_i \int_{-\infty}^{\infty} x^4 e^{-V_i x^2/2} dx + 100 e^{-V_i^{1/2}/2} + 2 \int_{50}^{\infty} e^{-\phi(q)x/8} dx$$

$$\ll V_i \cdot \frac{1}{V_i^{5/2}} + e^{-V_i^{1/2}/2} + \frac{1}{\phi(q)} e^{-\phi(q)50/8} \ll V_i^{-3/2}$$

by Lemma 7.1, where the final simplification uses Proposition 5.2.

Similarly, we write

$$\int_{-\infty}^{\infty} \left| \Phi_i''(x) - (V_i^2 x^2 - V_i) e^{-V_i x^2/2} \right| dx$$

$$= \int_{|x| \le V_i^{-1/2}} \left| \Phi_i''(x) - (V_i^2 x^2 - V_i) e^{-V_i x^2/2} \right| dx$$

$$+ \int_{V_i^{-1/2} \le |x| \le V_i^{-1/4}} \left| \Phi_i''(x) - (V_i^2 x^2 - V_i) e^{-V_i x^2/2} \right| dx$$

$$+ \int_{V_i^{-1/4} \le |x| \le 50} \left| \Phi_i''(x) - (V_i^2 x^2 - V_i) e^{-V_i x^2/2} \right| dx$$

$$+ \int_{|x| \ge 50} \left| \Phi_i''(x) - (V_i^2 x^2 - V_i) e^{-V_i x^2/2} \right| dx.$$

Using the bounds in Proposition 6.11,

$$
\int_{-\infty}^{\infty} \left| \Phi_i''(x) - e^{-V_i x^2/2} \right| dx
$$

$$
\ll \int_{|x| \le V_i^{-1/2}} V_i x^2 e^{-V_i x^2/2}\, dx + \int_{V_i^{-1/2} \le |x| \le V_i^{-1/4}} V_i^3 x^6 e^{-V_i x^2/2}\, dx
$$

$$
+ \int_{V_i^{-1/4} \le |x| \le 50} V_i^{3/2} e^{-V_i^{1/2}/2}\, dx + \int_{|x| \ge 50} V_i^2 x^2 e^{-\phi(q)|x|/8}\, dx
$$

$$
\le V_i \int_{-\infty}^{\infty} x^2 e^{-V_i x^2/2}\, dx + V_i^3 \int_{-\infty}^{\infty} x^6 e^{-V_i x^2/2}\, dx
$$

$$
+ 100 V_i^{3/2} e^{-V_i^{1/2}/2} + 2V_i^2 \int_{50}^{\infty} x^2 e^{-\phi(q)x/8}\, dx
$$

$$
\ll V_i \cdot \frac{1}{V_i^{3/2}} + V_i^3 \cdot \frac{1}{V_i^{7/2}} + e^{-V_i^{1/2}/2} + \frac{1}{\phi(q)} e^{-\phi(q)50/8} \ll V_i^{-1/2}
$$

again by Lemma 7.1 and Proposition 5.2 (and a routine calculation to evaluate the final integral exactly). ∎

Let $f_i(t)$ denote the density function of the random variable $X_i$ from Definition 3.7, and let $g_i(t) = (2\pi V_i)^{-1/2} e^{-t^2/2V_i}$ be the density function of a normal random variable with mean 0 and variance $V_i$. We can bound the difference between these two functions by writing them in terms of their characteristic functions.

**Lemma 7.3**  *Assume GRH. For $i \in \{1, 2, 3\}$ and $q > q_0$, we have*

$$
f_i(t) - g_i(t) \ll \min\left\{ V_i^{-3/2}, V_i^{-1/2} t^{-2} \right\}.
$$

**Proof**  We begin with the inverse Fourier transform formula

$$
f_i(t) - g_i(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \left( \Phi_i(x) - e^{-V_i x^2/2} \right) dx.
$$

On the one hand, this integral can be estimated trivially using the first estimate in Lemma 7.2:

$$
f_i(t) - g_i(t) \ll \int_{-\infty}^{\infty} \left| \Phi_i(x) - e^{-V_i x^2/2} \right| dx \ll V_i^{-3/2}.
$$

On the other hand, we can also integrate by parts twice before estimating, since both $\Phi_i(x)$ and $e^{-V_i x^2/2}$ and their first derivatives tend to 0 as $|x| \to \infty$ (see Remark 6.10):

$$
f_i(t) - g_i(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-ixt} \left( \Phi_i(x) - e^{-V_i x^2/2} \right) dx
$$

$$= -\frac{1}{2\pi t^2} \int_{-\infty}^{\infty} e^{-ixt} \left( \Phi_i''(x) - (V_i^2 x^2 - V_i) e^{-V_i x^2/2} \right) dx$$

$$\ll \frac{1}{t^2} \int_{-\infty}^{\infty} \left| \Phi_i''(x) - (V_i^2 x^2 - V_i) e^{-V_i x^2/2} \right| dx \ll \frac{V_i^{-1/2}}{t^2}$$

by the second estimate in Lemma 7.2. ∎

**Lemma 7.4** *Assume GRH. For $i \in \{1, 2, 3\}$ and $q > q_0$, we have*

$$\int_{-\infty}^{\infty} \left| f_i(t) - g_i(t) \right| dt \ll V_i^{-1}.$$

**Proof** By Lemma 7.3,

$$\int_{-\infty}^{\infty} \left| f_i(t) - g_i(t) \right| dt = \int_{|t| \le V_i^{1/2}} \left| f_i(t) - g_i(t) \right| dt + \int_{|t| \ge V_i^{1/2}} \left| f_i(t) - g_i(t) \right| dt$$

$$\ll \int_{|t| \le V_i^{1/2}} V_i^{-3/2} \, dt + \int_{|t| \ge V_i^{1/2}} \frac{V_i^{-1/2}}{t^2} \, dt$$

$$\ll V_i^{-3/2} \cdot V_i^{1/2} + \frac{V_i^{-1/2}}{V_i^{1/2}} \ll V_i^{-1}. \qquad \blacksquare$$

We are now ready to compare the probability that the random variables $X_i$ from Definition 3.7, which are relevant to prime number races, come in a particular order to the probability that normal random variables of the same mean and variance come in a particular order.

**Definition 7.5** For $i \in \{1, 2, 3\}$, let $Y_i$ denote a normal variable with mean $0$ and variance $V_i$, with the convention that $Y_1$, $Y_2$, and $Y_3$ are mutually independent. Note that the density function of $Y_i$ equals $g_i(t)$ as defined before Lemma 7.3.

**Theorem 7.6** *Assume GRH. For any permutation $(i, j, k)$ of $(1, 2, 3)$ and any $q > q_0$,*

$$\Pr(X_i > X_j > X_k) = \Pr\left( Y_i > Y_j > Y_k \right) + O\left( \frac{1}{\phi(q) \log q} \right).$$

**Proof** Given the formulas

$$\Pr(X_i > X_j > X_k) = \iiint_{x > y > z} f_i(x) f_j(y) f_k(z) \, dx \, dy \, dz$$

$$\Pr(Y_i > Y_j > Y_k) = \iiint_{x > y > z} g_i(x) g_j(y) g_k(z) \, dx \, dy \, dz,$$

we have

$$\left| \Pr(X_i > X_j > X_k) - \Pr(Y_i > Y_j > Y_k) \right|$$

$$= \left| \iiint_{x>y>z} \left( f_i(x)f_j(y)f_k(z) - g_i(x)g_j(y)g_k(z) \right) dx\,dy\,dz \right|$$

$$= \left| \iiint_{x>y>z} \left( f_i(x) - g_i(x) \right) f_j(y) f_k(z)\,dx\,dy\,dz \right.$$

$$+ \iiint_{x>y>z} g_i(x)\left( f_j(y) - g_j(y) \right) f_k(z)\,dx\,dy\,dz$$

$$+ \left. \iiint_{x>y>z} g_i(x)g_j(y)\left( f_k(z) - g_k(z) \right) dx\,dy\,dz \right|$$

$$\le \iiint_{\mathbb{R}^3} \left| f_i(x) - g_i(x) \right| f_j(y) f_k(z)\,dx\,dy\,dz$$

$$+ \iiint_{\mathbb{R}^3} g_i(x)\left| f_j(y) - g_j(y) \right| f_k(z)\,dx\,dy\,dz$$

$$+ \iiint_{\mathbb{R}^3} g_i(x)g_j(y)\left| f_k(z) - g_k(z) \right| dx\,dy\,dz$$

$$= \int_{\mathbb{R}} \left| f_i(x) - g_i(x) \right| dx + \int_{\mathbb{R}} \left| f_j(y) - g_j(y) \right| dy + \int_{\mathbb{R}} \left| f_k(z) - g_k(z) \right| dz,$$

since each integral of a probability density function over $\mathbb{R}$ equals 1. It follows from Lemma 7.4 and Proposition 5.2 that

$$\Pr(X_i > X_j > X_k) = \Pr(Y_i > Y_j > Y_k) + O(V_i^{-1} + V_j^{-1} + V_k^{-1})$$

$$= \Pr(Y_i > Y_j > Y_k) + O\left( \frac{1}{\phi(q)\log q} \right)$$

as desired. ∎

## 8 Proof of the main theorem

By this point, we have essentially reduced the problem of asymptotically evaluating the prime-race density $\delta_{q;a_1,a_2,a_3}$ (still under the assumptions from Definition 3.1) purely to a problem in probability. In this section, we complete the proof of Theorem 1.5 (including showing how to derive the first assertion from the second) and Corollary 1.6, with very little input needed from number theory. We begin with a classical (but perhaps not well known) formula for the probability that three normal variables assume a prescribed ordering.

**Lemma 8.1** *Let $N_a$, $N_b$, and $N_c$ denote mutually independent normal random variables with mean 0 and variances a, b, and c, respectively. Then*

$$\Pr(N_a > N_b > N_c) = \frac{1}{2\pi} \arctan \frac{\sqrt{ab + bc + ac}}{b}.$$

**Proof** If $Z_1$ and $Z_2$ are normal random variables with mean 0 and correlation coefficient $\rho$, there is a classical formula (see [1, equation (4)] for example) for the "orthant probability" that both random variables are positive:

$$\Pr(Z_1 > 0 \text{ and } Z_2 > 0) = \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho.$$

We apply this formula with $Z_1 = N_a - N_b$ and $Z_2 = N_b - N_c$, which are indeed normal random variables with mean 0 and correlation coefficient

$$(8.1) \qquad \rho = \frac{\mathbb{E}((N_a - N_b)(N_b - N_c))}{\sqrt{\mathbb{E}((N_a - N_b)^2) \cdot \mathbb{E}((N_b - N_c)^2)}}$$

$$= \frac{-\sigma^2(N_b)}{\sqrt{(\sigma^2(N_a) + \sigma^2(N_b))(\sigma^2(N_b) + \sigma^2(N_c))}} = \frac{-b}{\sqrt{ab + ac + bc + b^2}},$$

so that

$$\Pr(N_a > N_b > N_c) = \Pr(N_a - N_b > 0 \text{ and } N_b - N_c > 0)$$

$$= \frac{1}{4} - \frac{1}{2\pi} \arcsin \left( \frac{b}{\sqrt{ab + ac + bc + b^2}} \right).$$

The lemma now follows from the identities (valid for $0 \le x \le y$)

$$\arcsin \left( \frac{x}{y} \right) = \arctan \left( \frac{x}{\sqrt{y^2 - x^2}} \right) = \frac{\pi}{2} - \arctan \left( \frac{\sqrt{y^2 - x^2}}{x} \right). \qquad \blacksquare$$

At this point, we can complete the proof of an important special case of our main theorem, assuming the restriction from Definition 3.1 that has been in force since that point. Recall the quantity $\delta_{q;a_1,a_2,a_3}$ from Definition 3.1.

**Proof of Theorem 1.5 under the assumption that $a_1, a_2,$ and $a_3$ are either all quadratic residues or all quadratic nonresidues (mod $q$)** We may assume that $q > q_0$ from Definition 5.3, since the asymptotic formula is trivially valid for any bounded range of $q$. We simply combine the three equalities in Proposition 3.9, Theorem 7.6 (using the notation of Definition 7.5), and Lemma 8.1, obtaining

$$\delta_{q;a_1,a_2,a_3} = \Pr(X_1 > X_2 > X_3) = \Pr \left( Y_1 > Y_2 > Y_3 \right) + O\left( \frac{1}{\phi(q) \log q} \right)$$

$$= \frac{1}{2\pi} \arctan \frac{\sqrt{V_1 V_2 + V_2 V_3 + V_1 V_3}}{V_2} + O\left( \frac{1}{\phi(q) \log q} \right)$$

as claimed. $\blacksquare$

It is not difficult to remove the assumption that $a_1, a_2,$ and $a_3$ are either all quadratic residues or all quadratic nonresidues (mod $q$), at least if we allow ourselves the larger error term asserted in Theorem 1.5. Again, all we require is a quick lemma from probability.

**Lemma 8.2**  *Let $Z_1$, $Z_2$, and $Z_3$ be random variables, and let $\mu_1$, $\mu_2$, and $\mu_3$ be real numbers. The event*

(8.2)    *exactly one of $(Z_1 > Z_2 > Z_3)$ and $(Z_1 + \mu_1 > Z_2 + \mu_2 > Z_3 + \mu_3)$ is true*

*is contained in the event*

(8.3)    $$\bigl(|Z_1 - Z_2| \le |\mu_1 - \mu_2|\bigr) \text{ or } \bigl(|Z_2 - Z_3| \le |\mu_2 - \mu_3|\bigr).$$

**Proof**    First observe that

- if $|Z_1 - Z_2| > |\mu_1 - \mu_2|$, then the two inequalities, $Z_1 > Z_2$ and $Z_1 + \mu_1 > Z_2 + \mu_2$, are either both true or both false;
- if $|Z_2 - Z_3| > |\mu_2 - \mu_3|$, then the two inequalities, $Z_2 > Z_3$ and $Z_2 + \mu_2 > Z_3 + \mu_3$, are either both true or both false.

It follows that if both $\bigl(|Z_1 - Z_2| > |\mu_1 - \mu_2|\bigr)$ and $\bigl(|Z_2 - Z_3| > |\mu_2 - \mu_3|\bigr)$ are true, then $(Z_1 > Z_2 > Z_3)$ and $(Z_1 + \mu_1 > Z_2 + \mu_2 > Z_3 + \mu_3)$ are either both true or both false; this implication is the contrapositive of the proposition.  ∎

At this point, we no longer assume that $a_1$, $a_2$, and $a_3$ have the same quadratic nature (mod $q$), although the congruences (1.2) are still in force.

**Proof of Theorem 1.5 in the general case**    When we do not assume that $a_1$, $a_2$, and $a_3$ are either all quadratic residues or all quadratic nonresidues (mod $q$), we may still use the random variables $X_i$ and $Y_i$ from Definitions 3.7 and 7.5. However, we cannot rely on full cancellation of the constants in Lemma 3.6, and so Proposition 3.9 must be modified: the distribution of the vector-valued function $\bigl(E^*(x; q, a_1), E^*(x; q, a_2), E^*(x; q, a_3)\bigr)$ is the same as the distribution of the random variable $(\mu_1, \mu_2, \mu_3) + X_{1,2,3}$, where $(\mu_1, \mu_2, \mu_3)$ is defined to be the vector

$$\bigl(c_q(a_2) + c_q(a_3) - 2c_q(a_1), c_q(a_1) + c_q(a_3) - 2c_q(a_2), c_q(a_1) + c_q(a_2) - 2c_q(a_3)\bigr).$$

Consequently, the density we want to evaluate now takes the form

$$\delta_{q; a_1, a_2, a_3} = \Pr\bigl(X_1 + \mu_1 > X_2 + \mu_2 > X_3 + \mu_3\bigr)$$
$$= \Pr\bigl(Y_1 + \mu_1 > Y_2 + \mu_2 > Y_3 + \mu_3\bigr) + O\left(\frac{1}{\phi(q)\log q}\right)$$

by the proof of Theorem 7.6. We deduce from Lemma 8.2 that

$$\bigl|\Pr\bigl(Y_1 + \mu_1 > Y_2 + \mu_2 > Y_3 + \mu_3\bigr) - \Pr\bigl(Y_1 > Y_2 > Y_3\bigr)\bigr|$$
$$\le \Pr\bigl(|Y_1 - Y_2| \le |\mu_1 - \mu_2|\bigr) + \Pr\bigl(|Y_2 - Y_3| \le |\mu_2 - \mu_3|\bigr).$$

Since $Y_i$ are mutually independent, $Y_1 - Y_2$ is a normal random variable with variance $V_1 + V_2$, and hence its density function is bounded pointwise by the constant $1/\sqrt{2\pi(V_1 + V_2)}$; the analogous bound applies to the density function of $Y_2 - Y_3$. In particular,

$$\bigl|\Pr\bigl(Y_1 + \mu_1 > Y_2 + \mu_2 > Y_3 + \mu_3\bigr) - \Pr\bigl(Y_1 > Y_2 > Y_3\bigr)\bigr|$$
$$\le \frac{2|\mu_1 - \mu_2|}{\sqrt{2\pi(V_1 + V_2)}} + \frac{2|\mu_2 - \mu_3|}{\sqrt{2\pi(V_2 + V_3)}}.$$

However, it is a standard fact that $c_q(a) \ll_\varepsilon q^\varepsilon$ (see [3, Definitions 1.2 and 2.4]). Since each $V_i \gg 1/\sqrt{\phi(q)\log q}$ by Proposition 5.2, we deduce that

$$\left| \Pr(Y_1 + \mu_1 > Y_2 + \mu_2 > Y_3 + \mu_3) - \Pr(Y_1 > Y_2 > Y_3) \right| \ll_\varepsilon \frac{q^\varepsilon}{\sqrt{\phi(q)\log q}} \ll_\varepsilon q^{-1/2+\varepsilon}.$$

In conclusion,

$$\begin{aligned}
\delta_{q;a_1,a_2,a_3} &= \Pr(Y_1 + \mu_1 > Y_2 + \mu_2 > Y_3 + \mu_3) + O\left(\frac{1}{\phi(q)\log q}\right) \\
&= \Pr(Y_1 > Y_2 > Y_3) + O_\varepsilon\left(q^{-1/2+\varepsilon} + \frac{1}{\phi(q)\log q}\right) \\
&= \frac{1}{2\pi}\arctan\frac{\sqrt{V_1 V_2 + V_1 V_3 + V_2 V_3}}{V_2} + O_\varepsilon(q^{-1/2+\varepsilon})
\end{aligned}$$

by Lemma 8.1, as claimed.                                                                                        ∎

**Proof**    Since $V_i = 4V(q)(1+\eta_i)$ by Definition 1.4, we may restate Theorem 1.5 as

$$\Pr(X_1 > X_2 > X_3)$$
$$= \frac{1}{2\pi}\arctan\frac{\sqrt{(1+\eta_1)(1+\eta_2) + (1+\eta_1)(1+\eta_3) + (1+\eta_2)(1+\eta_3)}}{(1+\eta_2)} + O_\varepsilon(q^{-1/2+\varepsilon}).$$

It is an easy calculus exercise to compute the linear approximation at the origin to the twice-differentiable function above, obtaining

$$\Pr(X_1 > X_2 > X_3) = \frac{1}{6} + \frac{\eta_1}{8\pi\sqrt{3}} - \frac{\eta_2}{4\pi\sqrt{3}} + \frac{\eta_3}{8\pi\sqrt{3}} + O(\eta_1^2 + \eta_2^2 + \eta_3^2) + O_\varepsilon(q^{-1/2+\varepsilon}).$$

The corollary now follows from the estimate $\eta_i \ll (\log\log q)/\log q$ given in Proposition 5.2.                                                                                        ∎

It turns out that while it is possible for the $\eta_i$ to be as large as $\Omega((\log\log q)/\log q)$, they are usually rather smaller; in such a situation, the error term given in Corollary 1.6 can be reduced considerably. Indeed, an asymptotic formula for $\delta_{q;a_1,a_2,a_3}$ was given by Lamzouri in a form where the secondary main terms and error terms had a more explicit dependence on arithmetic quantities like the $V_i$, including the one we define now.

**Definition 8.3**    For any reduced residue classes $a$ and $b$ modulo $q$, define

$$B_q(a,b) = \sum_{\substack{\chi \,(\mathrm{mod}\,q) \\ \chi \neq \chi_0}} \left(\chi(ab^{-1}) + \chi(ba^{-1})\right)b_+(\chi).$$

(Note, for example, that $B_q(a,a) = V(q)$ from Definition 1.4.)

The following result, which is [9, Corollary 2.3] translated into our notation, applies to *any* distinct reduced residues $a_1$, $a_2$, and $a_3$ (mod $q$).

**Theorem 8.4** (Lamzouri)   *We have*

$$\delta_{q;a_1,a_2,a_3} = \frac{1}{6} + \frac{1}{4\sqrt{\pi}} \frac{c_q(a_3) - c_q(a_1)}{\sqrt{V(q)}}$$

$$+ \frac{1}{4\pi\sqrt{3}} \frac{B_q(a_1, a_2) + B_q(a_2, a_3) - 2B_q(a_1, a_3)}{V(q)}$$

(8.4)
$$+ O\left( \frac{c_q(1)^2}{V(q)} + \frac{\max_{1 \le i < j \le 3} |B_q(a_i, a_j)|^2}{V(q)^2} \right).$$

One motivation for stating this result is to calculate the secondary main terms in our special case $a_1^2 \equiv a_2^2 \equiv a_3^2 \pmod{q}$ and $c_q(a_1) = c_q(a_2) = c_q(a_3)$ in the notation of equation (2.1). Under these assumptions,

$$B_q(a_i, a_j) = 2\left( \sum_{\substack{\chi \pmod{q} \\ \chi(a_i) = \chi(a_j)}} b_+(\chi) - \sum_{\substack{\chi \pmod{q} \\ \chi(a_i) = -\chi(a_j)}} b_+(\chi) - b_+(\chi_0) \right)$$

$$= \frac{V_0 + V_k - V_i - V_j}{16} - 2b_+(\chi_0)$$

in the notation of Definition 1.4 (and thus, from Proposition 5.2, we have $B_q(a_i, a_j)^2/V(q)^2 \ll (\log\log q)^2/(\log q)^2$). In particular,

$$B_q(a_1, a_2) + B_q(a_2, a_3) - 2B_q(a_1, a_3) = \frac{V_1 - 2V_2 + V_3}{8}.$$

Using this identity in equation (8.4), and making the change of variables $V_i = 4V(q)(\eta_i + 1)$ from Definition 1.4, reveals that the secondary main terms in (8.4) are exactly equal to those in Corollary 1.6. (As we see, Lamzouri's result gives yet another secondary main term in the case where $c_q(a_1) \ne c_q(a_3)$, while our method simply gives an error term of that order of magnitude.)

## 9 Discussion

We have already discussed, at the end of Section 3, the algebraic aspects of our special three-way races and the prospects for generalizing our method in that regard. In this final section, we make some additional remarks about the analytic aspects of this paper, under the continuing assumptions of GRH and LI.

The moral we hope to emphasize is that, *to give asymptotic formulas for prime number race densities $\delta_{q;a_1,\ldots,a_r}$ from Definition 1.1 whose error terms are small, one should always utilize a main term that is an "ordering probability" for a multivariate normal distribution.* If $Z = (Z_1, \ldots, Z_r)$ is a normal random variable in $\mathbb{R}^r$ whose covariance matrix is identical to the matrix of covariances $B_q(a_i, a_j)$ (from Definition 8.3) corresponding to our prime number race, then the difference between $\delta_{q;a_1,\ldots,a_r}$ and $\Pr(Z_1 > \cdots > Z_r)$ will decay like a negative power of $q$, as in Theorem 1.5. Therefore, it is best, we claim, to primarily evaluate $\delta_{q;a_1,\ldots,a_r}$ as an ordering probability of this type; further asymptotic evaluations can then be made for the ordering probability itself, an object that is purely probabilistic.

This idea is certainly well established in this topic on a heuristic level, being motivated, for example, by the central limit results established by Rubinstein and Sarnak [11]. Using an ordering probability as the main term is implicit in the work of Lamzouri [8, 9] (though the comparisons there were carried out mainly on the characteristic function side) and more explicitly in the work of Harper and Lamzouri [5, Section 4.1].

While we wanted to describe the construction of our atypical normalizations $E^*(x; q, a)$ of the error terms in prime counting functions for arithmetic progressions, and how they resulted in independent random variables that allowed for a much more elementary analysis, that independence is not necessary to our moral, as it happens. The multivariate normal random variable $Z$ alluded to above is permitted to have correlations among the coordinates, and even to have nonzero means in each coordinate (arising, in our case, from Chebyshev's bias against quadratic residues). Because the matrix of covariances $B_q(a_i, a_j)$ turns out to be close to the identity matrix times $V(q)$ (the quantity from Definition 1.4), and the vector of means turns out to be small compared to $\sqrt{V(q)}$, it is possible to produce asymptotics for these ordering probabilities in terms of the means and covariances, as was done in [9].

Moreover, in the case where the residues $a_j$ are all quadratic residues or all quadratic nonresidues (so that the means of the individual limiting distributions are all equal), these ordering probabilities can actually be evaluated in closed form for $r \le 4$, because they are equivalent (by the method of proof of Lemma 8.1) to orthant probabilities in at most three dimensions. (The cases $r \le 2$ are trivial because zero- and one-dimensional orthant probabilities are trivial under the assumption of equal means.) Recalling Definitions 1.1, 1.4, and 8.3, we define the further notation

$$\rho_{12} = \frac{-V(q) + B_q(a_1, a_2) - B_q(a_1, a_3) + B_q(a_2, a_3)}{2\sqrt{(V(q) - B_q(a_1, a_2))(V(q) - B_q(a_2, a_3))}}$$

$$\rho_{13} = \frac{B_q(a_1, a_3) - B_q(a_2, a_3) - B_q(a_1, a_4) + B_q(a_2, a_4)}{2\sqrt{(V(q) - B_q(a_1, a_2))(V(q) - B_q(a_3, a_4))}}$$

$$\rho_{23} = \frac{-V(q) + B_q(a_2, a_3) - B_q(a_2, a_4) + B_q(a_3, a_4)}{2\sqrt{(V(q) - B_q(a_2, a_3))(V(q) - B_q(a_3, a_4))}}$$

Then, for *any* distinct reduced residues $a_1$, $a_2$, $a_3$, and $a_4$ (mod $q$), known formulas for orthant probabilities [1, equations (4) and (5)] imply that the asymptotic formulas

$$\delta_{q;a_1,a_2,a_3} \sim \frac{1}{4} + \frac{1}{2\pi} \arcsin \rho_{12}$$

$$\delta_{q;a_1,a_2,a_3,a_4} \sim \frac{1}{8} + \frac{1}{4\pi} \left( \arcsin \rho_{12} + \arcsin \rho_{13} + \arcsin \rho_{23} \right)$$

hold up to a negative power of $q$. As a reality check, if we exploit the fact that the $B_q(a_i, a_j)$ are negligible in size compared to $V(q)$, then $\rho_{12} \sim \rho_{23} \sim -\frac{1}{2}$ and $\rho_{13} = o(1)$, which leads to the evaluations $\delta_{q;a_1,a_2,a_3} \sim \frac{1}{6}$ and $\delta_{q;a_1,a_2,a_3,a_4} \sim \frac{1}{24}$ as expected from the central limit theorems of [11].

In this paradigm, we have approximated our number-theoretic limiting logarithmic distributions by normal distributions with the same mean and variance. Of

course, the higher (even central) moments of the two distributions will not match in general, which is a source of error when passing from one measure to the other. It is worth mentioning that for two-way races, asymptotic formulas for the densities exist [3, Theorem 1.1] that incorporate the contribution from higher moments and, correspondingly, have error terms that can be made as small as an arbitrary power of $q$. It would be an interesting project to attempt to produce analogous formulas for prime number races with three or more contestants.

Nevertheless, we hope the viewpoint that prime race densities are best approximated explicitly by ordering probabilities for multivariate normal random variables has some illuminating benefit to practitioners of comparative prime number theory.

## References

[1] R. H. Bacon, *Approximations to multivariate normal orthant probabilities*. Ann. Math. Stat. **34**(1963), no. 1, 191–198.

[2] M. A. Bennett, G. Martin, K. O'Bryant, and A. Rechnitzer, *Explicit bounds for primes in arithmetic progressions*. Ill. J. Math. **62**(2018), nos. 1–4, 427–532.

[3] D. Fiorilli and G. Martin, *Inequities in the Shanks–Rényi prime number race: an asymptotic formula for the densities*. J. Reine. Angew. Math. **2013**(2013), no. 676, 121–212.

[4] S. Haber and A. Rosenfeld, *Groups as unions of proper subgroups*. Am. Math. Mon. **66**(1959), 491–494.

[5] A. J. Harper and Y. Lamzouri, *Orderings of weakly correlated random variables, and prime number races with many contestants*. Probab. Theory Related Fields **170**(2018), nos. 3–4, 961–1010.

[6] E. Hille and G. Szegő, *On the complex zeros of the Bessel functions*. Bull. Am. Math. Soc. **49**(1943), 605–610.

[7] A. Hurwitz, *Ueber die Nullstellen der Bessel'schen Function*. Math. Ann. **33**(1888), no. 2, 246–266.

[8] Y. Lamzouri, *The Shanks-Rényi prime number race with many contestants*. Math. Res. Lett. **19** (2012), no. 3, 649–666.

[9] Y. Lamzouri, *Prime number races with three or more competitors*. Math. Ann. **356**(2013), no. 3, 1117–1162.

[10] G. Martin, *Asymmetries in the Shanks-Rényi prime number race*. Number theory for the millennium, II (Urbana, IL, 2000), A. K. Peters, Natick, MA, 2002, pp. 403–415.

[11] M. Rubinstein and P. Sarnak, *Chebyshev's bias*. Exp. Math. **3**(1994), no. 3, 173–197.

[12] G. Szegő, *Orthogonal polynomials. Vol. XXII*. 4th ed., Amer. Math. Soc. Colloquium Publications, Providence, RI, 1975

[13] G. N. Watson, *A treatise on the theory of Bessel functions*. Cambridge Mathematical Library, Cambridge University Press, Cambridge, MA, 1995, Reprint of the second (1944) edition.

[14] G. Zappa, *The papers of Gaetano Scorza on group theory*. Atti Accad. Naz. Lincei, Cl. Sci. Fis. Mat. Nat. IX. Ser. Rend. Lincei Mat. Appl. **2**(1991), 95–101.

*Department of Mathematics, University of British Columbia Room 121, 1984 Mathematics Road, Vancouver, BC V6T 1Z2, Canada*
*e-mail*: jiawei.lin@alumni.ubc.ca  gerg@math.ubc.ca