

Orderings of weakly correlated random variables, and prime number races with many contestants

Adam J. Harper^{1,2} · Youness Lamzouri³

Received: 9 October 2015 / Revised: 17 August 2017 / Published online: 18 September 2017
© Springer-Verlag GmbH Germany 2017

Abstract We investigate the race between prime numbers in many residue classes modulo q , assuming the standard conjectures GRH and LI. Among our results we exhibit, for the first time, n -way prime number races modulo q where the biases do not dissolve when $n, q \rightarrow \infty$. We also study the leaders in the prime number race, obtaining asymptotic formulae for logarithmic densities when the number of competitors can be as large as a power of q , whereas previous methods could only allow a power of $\log q$. The proofs use harmonic analysis related to the Hardy–Littlewood circle method to control the average size of correlations in prime number races. They also use various probabilistic tools, including an exchangeable pairs version of Stein’s method, normal comparison tools, and conditioning arguments. In the process we derive some general results about orderings of weakly correlated random variables, which may be of independent interest.

Adam J. Harper is supported by a research fellowship at Jesus College, Cambridge; and, when this work was started, by a postdoctoral fellowship from the Centre de Recherches Mathématiques, Montréal. Youness Lamzouri is partially supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada.

✉ Youness Lamzouri
lamzouri@mathstat.yorku.ca

Adam J. Harper
A.Harper@warwick.ac.uk

¹ Jesus College, Cambridge CB5 8BL, England, UK

² Present Address: Mathematics Institute, Zeeman Building, University of Warwick, Coventry CV4 7AL, England, UK

³ Department of Mathematics and Statistics, York University, 4700 Keele Street, Toronto, ON M3J 1P3, Canada

Mathematics Subject Classification Primary 11N13; Secondary 11N69 · 11M26

1 Introduction

In an 1853 letter to Fuss, Chebyshev noted that on a fine scale there seem to be more primes congruent to 3 than to 1 modulo 4. This observation led to the birth of comparative prime number theory, which investigates the discrepancies in the distribution of prime numbers. A central problem is the so-called “Shanks–Rényi prime number race” which is described by Knapowski and Turán [14–21]: let $q \geq 3$ and $2 \leq n \leq \varphi(q)$ be positive integers, (where the Euler function $\varphi(q)$ denotes the number of residue classes mod q that are coprime to q), and denote by $\mathcal{A}_n(q)$ the set of ordered n -tuples (a_1, a_2, \dots, a_n) of distinct residue classes that are coprime to q . For $(a_1, a_2, \dots, a_n) \in \mathcal{A}_n(q)$ consider a game with n players called “ a_1 ” through to “ a_n ”, where at time x , the player a_j has a score of $\pi(x; q, a_j)$ (where $\pi(x; q, a)$ denotes the number of primes $p \leq x$ with $p \equiv a \pmod q$). Among the questions that Knapowski and Turán asked in [14–21] are the following:

- Q1. Will each player take the lead for infinitely many integers x ?
- Q2. Will all $n!$ orderings of the players occur for infinitely many integers x ?

It is generally believed that the answer to the stronger question Q2 (and thus to Q1) is yes for all q and all $(a_1, a_2, \dots, a_n) \in \mathcal{A}_n(q)$. An old result of Littlewood [25] shows that this is indeed true when $(q, a_1, a_2) = (4, 1, 3)$ and $(q, a_1, a_2) = (3, 1, 2)$. Since then, this problem has been extensively studied by various authors, including Knapowski and Turán [14–21], Kaczorowski [11–13], Feuerverger and Martin [3], Ford and Konyagin [6], Ford, Konyagin and Lamzouri [8], Fiorilli and Martin [5], Fiorilli [4], and Lamzouri [22,23]. For a complete history as well as recent developments, see the expository papers of Granville and Martin [9], Ford and Konyagin [7], and Martin and Scarfy [26] (which includes a very comprehensive list of references).

Assuming the Generalized Riemann hypothesis GRH and the Linear Independence hypothesis LI (which is the assumption that the nonnegative imaginary parts of the nontrivial zeros of Dirichlet L -functions attached to primitive characters are linearly independent over \mathbb{Q}), Rubinstein and Sarnak [31] affirmatively answered questions Q1 and Q2. In fact, under these hypotheses, they established the stronger result that for any $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$, the set of real numbers $x \geq 2$ such that

$$\pi(x; q, a_1) > \pi(x; q, a_2) > \dots > \pi(x; q, a_n), \tag{1.1}$$

has a positive logarithmic density, which we denote by $\delta(q; a_1, \dots, a_n)$. (Recall that the logarithmic density of a subset S of \mathbb{R} is defined as

$$\lim_{x \rightarrow \infty} \frac{1}{\log x} \int_{t \in S \cap [2,x]} \frac{dt}{t},$$

provided that this limit exists.) This density can be regarded as the “probability” that for each $1 \leq j \leq n$, the player a_j is at the j th position in the prime race.

Among their results on question Q2, Rubinstein and Sarnak [31] showed that for n fixed,

$$\lim_{q \rightarrow \infty} \max_{(a_1, \dots, a_n) \in \mathcal{A}_n(q)} \left| \delta(q; a_1, \dots, a_n) - \frac{1}{n!} \right| = 0. \quad (1.2)$$

Feuerverger and Martin [3] raised the question of having a uniform version of this statement, in which the number of contestants $n \rightarrow \infty$ as $q \rightarrow \infty$. In response to this, Lamzouri [22] established that for any integer n such that $2 \leq n \leq \sqrt{\log q}$ we have

$$\delta(q; a_1, \dots, a_n) = \frac{1}{n!} \left(1 + O\left(\frac{n^2}{\log q}\right) \right),$$

uniformly for all n -tuples $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$. Feuerverger and Martin [3] also asked whether for n sufficiently large in terms of q the asymptotic formula $\delta(q; a_1, \dots, a_n) \sim 1/n!$ might become false. A few years ago, Ford and Lamzouri (unpublished) developed a heuristic argument which suggests that there should be a transition in the behaviour of the densities when $n = (\log q)^{1+o(1)}$. More specifically, they formulated the following conjecture.

Conjecture 1.1 (Ford and Lamzouri) *Let $\varepsilon > 0$ be small and q be sufficiently large.*

- (1) *If $2 \leq n \leq (\log q)^{1-\varepsilon}$, then uniformly for all n -tuples $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$ we have $\delta(q; a_1, \dots, a_n) \sim 1/n!$ as $q \rightarrow \infty$.*
- (2) *If $(\log q)^{1+\varepsilon} \leq n \leq \varphi(q)$, then there exist n -tuples $(a_1, \dots, a_n), (b_1, \dots, b_n) \in \mathcal{A}_n(q)$ for which $n! \cdot \delta(q; a_1, \dots, a_n) \rightarrow 0$ and $n! \cdot \delta(q; b_1, \dots, b_n) \rightarrow \infty$ as $q \rightarrow \infty$.*

Our first result establishes part (1) of this conjecture.

Theorem 1.2 *Assume GRH and LI. Let $2 \leq n \leq \log q / (\log \log q)^4$ be a positive integer. Then, uniformly for all n -tuples $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$ we have*

$$\delta(q; a_1, \dots, a_n) = \frac{1}{n!} \left(1 + O\left(\frac{n(\log n)^4}{\log q}\right) \right).$$

To be clear, when we say that this estimate holds “uniformly for all n -tuples”, we mean that it holds for all n, q and (a_1, \dots, a_n) permitted in the statement, with an absolute implicit constant in the “big Oh” term. To be even more explicit, we mean that

$$\max_{(a_1, \dots, a_n) \in \mathcal{A}_n(q)} |n! \delta(q; a_1, \dots, a_n) - 1| \leq C \frac{n(\log n)^4}{\log q},$$

for all permitted n, q and for some absolute positive constant C . The same explanation applies to the statements of our subsequent theorems.

The second part of Conjecture 1.1 implies, in particular, that the asymptotic formula $\delta(q; a_1, \dots, a_n) \sim 1/n!$ need not hold for all n -tuples $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$, if n lies in the range $(\log q)^{1+\varepsilon} \leq n \leq \varphi(q)$. We believe this to be true because we encounter various error terms in our arguments, from different sources, of size about $1/q$. Thus we believe that when the target density $1/n!$ becomes much smaller than any negative power of q , which happens when $n \geq (\log q)^{1+\varepsilon}$, it is no longer reasonable to expect

all of the $\delta(q; a_1, \dots, a_n)$ to be close to $1/n!$. Our next result shows that this is indeed the case in the smaller range $\varphi(q)^\varepsilon < n \leq \varphi(q)$. Thus we are able to exhibit, for the first time, n -way prime number races for which the biases do not dissolve when $q \rightarrow \infty$, confirming the prediction of Feuerverger and Martin [3].

Theorem 1.3 *Assume GRH and LI. Let $\varepsilon > 0$ and let q be sufficiently large in terms of ε . For every integer $\varphi(q)^\varepsilon \leq n \leq \varphi(q)$ there exists an n -tuple $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$ such that*

$$\delta(q; a_1, \dots, a_n) < (1 - c_\varepsilon) \frac{1}{n!},$$

for some positive constant c_ε which depends only on ε .

We next consider a stronger form of question Q1, concerning the leader in a prime number race with many contestants. By the work of Rubinstein and Sarnak, it follows that for any $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$, the set of real numbers $x \geq 2$ such that

$$\pi(x; q, a_1) > \max_{2 \leq j \leq n} \pi(x; q, a_j),$$

has a positive logarithmic density which we denote by $\delta_1(q; a_1, \dots, a_n)$. Kaczorowski [11] has considered this leadership question (in the special case $a_1 = 1$), and obtained some positive lower density results assuming only GRH rather than LI. One can ask the following natural quantitative question:

Q3. Will each of the players a_1, \dots, a_n have an ‘‘equal chance’’ $1/n$ of leading the race, when $q \rightarrow \infty$?

It follows from Theorem 1.2 that the answer to this question is yes, if the number of contestants n lies in the range $2 \leq n = o(\log q / (\log \log q)^4)$. Using a different approach we extend this significantly, showing that n can be as large as a small power of q .

Theorem 1.4 *Assume GRH and LI. Let $2 \leq n \leq \varphi(q)^{1/32}$ be an integer. Then, uniformly for all n -tuples $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$ we have*

$$\delta_1(q; a_1, \dots, a_n) = \frac{1}{n} \left(1 + O \left(\frac{n^4}{\varphi(q)^{1/8}} + \frac{1}{(n \log q)^{12/25}} \right) \right).$$

The key ingredient in the proof of this theorem is the following probabilistic result, which may be of independent interest. It investigates the probability that a given random variable is the leader among weakly correlated Gaussian random variables.

Theorem 1.5 *Suppose $\varepsilon > 0$ is sufficiently small, and n is sufficiently large. Let X_1, \dots, X_n be mean zero, variance one, jointly normal random variables, and write $r_{i,j} := \mathbb{E}X_i X_j$, and suppose that $|r_{i,j}| \leq \varepsilon$ whenever $i \neq j$. Then*

$$\left| \mathbb{P} \left(X_1 > \max_{2 \leq i \leq n} X_i \right) - 1/n \right| \ll n^{-100} + n^{-1.99} \sum_{2 \leq i \leq n} |r_{1,i}| + n^{-2.99} \sum_{2 \leq i < j \leq n} |r_{i,j}|.$$

Note that the probability would be exactly $1/n$ if the X_i were independent of one another, by symmetry.

We also consider a variant of question Q3 concerning the ordering of the first k contestants in a prime race with many competitors. To this end, for each integer $1 \leq k \leq n$, we define $\delta_k(q; a_1, \dots, a_n)$ to be the logarithmic density of the set of real numbers $x \geq 2$ such that

$$\pi(x; q, a_1) > \pi(x; q, a_2) > \dots > \pi(x; q, a_k) > \max_{k+1 \leq j \leq n} \pi(x; q, a_j).$$

Note that the cases $k = n - 1$ and $k = n$ both correspond to the full ordering (1.1), and hence $\delta_{n-1}(q; a_1, \dots, a_n) = \delta_n(q; a_1, \dots, a_n) = \delta(q; a_1, \dots, a_n)$.

It follows from Theorem 1.2 that in the range $1 \leq k \leq n - 1 \leq \log q / (\log \log q)^4$, we have

$$\delta_k(q; a_1, \dots, a_n) = \frac{(n-k)!}{n!} \left(1 + O\left(\frac{n(\log n)^4}{\log q}\right) \right).$$

Now $\frac{(n-k)!}{n!} = n^{-k(1+o(1))}$ as $n \rightarrow \infty$, so the heuristic discussed following Theorem 1.2 leads us to expect that $\delta_k(q; a_1, \dots, a_n) \sim \frac{(n-k)!}{n!}$ even for very large n , provided roughly that $k \ll (\log q) / \log n$. We will show that the asymptotic does hold on almost this entire range of k , for all n such that $(\log n) / \log q \rightarrow 0$ as $q \rightarrow \infty$. Moreover, unlike the case $k = 1$ (where Theorem 1.4 permits n to be a small fixed power of q), we show that the condition $(\log n) / \log q \rightarrow 0$ is necessary to guarantee an asymptotic for any $k \geq 2$.

Theorem 1.6 *Assume GRH and LI. Let $2 \leq k \leq n - 2$ be integers, and suppose $(\log q) / \log n$ is large enough and $k(\log k)^{10} \leq (\log q) / \log n$. Then, uniformly for all n -tuples $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$ we have*

$$\delta_k(q; a_1, \dots, a_n) = \frac{(n-k)!}{n!} \left(1 + O\left(k(\log k)^6 \frac{\log n}{\log q} + \frac{1}{n \log^{1/10} q}\right) \right).$$

Theorem 1.7 *Assume GRH and LI. Let $\varepsilon > 0$ and let q be sufficiently large in terms of ε . Let $k \geq 2$ be fixed and let n be an integer in the range $\varphi(q)^\varepsilon \leq n < \varphi(q)^{1/41}$. Then there exists an n -tuple $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$ such that*

$$\delta_k(q; a_1, \dots, a_n) < (1 - c_\varepsilon) \frac{(n-k)!}{n!},$$

for some positive constant c_ε that depends only on ε .

Theorem 1.3 can be deduced from Theorem 1.7 as follows. Take $k = 2$, let $\varepsilon > 0$ be suitably small, and suppose first that $\varphi(q)^\varepsilon \leq n < \varphi(q)^{1/41}$ is an integer. Then it follows from Theorem 1.7 that there exists an n -tuple $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$ such that

$$\delta_2(q; a_1, \dots, a_n) < (1 - c_\varepsilon) \frac{(n-2)!}{n!} \quad (1.3)$$

for some positive constant c_ε . Let S_{n-2} be the symmetric group on $n - 2$ elements. Since the logarithmic density of the set of real numbers $x \geq 2$ for which $\pi(x; q, a) = \pi(x; q, b)$ is 0 [which follows from Eq. (2.3) below], we get

$$\delta_2(q; a_1, \dots, a_n) = \sum_{\sigma \in S_{n-2}} \delta(q; a_1, a_2, a_{\sigma(3)}, \dots, a_{\sigma(n)}).$$

Thus by (1.3), there exists $\sigma \in S_{n-2}$ for which

$$\delta(q; a_1, a_2, a_{\sigma(3)}, \dots, a_{\sigma(n)}) < (1 - c_\varepsilon) \frac{1}{n!},$$

completing the proof of Theorem 1.3 provided $n < \varphi(q)^{1/41}$.

However, if n is larger then we can set $m := \lfloor \varphi(q)^{1/41} \rfloor - 1$, and note that by the previous discussion there exists an m -tuple $(a_1, \dots, a_m) \in \mathcal{A}_m(q)$ for which

$$\delta(q; a_1, \dots, a_m) < (1 - c) \frac{1}{m!}.$$

Then if we choose any other coprime residues $a_{m+1}, \dots, a_n \pmod q$, we have

$$\delta(q; a_1, \dots, a_n) = \sum_{\substack{\sigma \in S_n: \\ \sigma^{-1}(1) > \sigma^{-1}(2) > \dots > \sigma^{-1}(m)}} \delta(q; a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(n)}).$$

There are $n!/m!$ terms in the sum, so it follows that for at least one permutation σ we must have $\delta(q; a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(n)}) < (1 - c) \frac{1}{n!}$, as claimed. \square

Let us remark on the fact that all the results we have described and stated, from Rubinstein and Sarnak [31] onwards, relate to limiting logarithmic densities such as $\delta(q; a_1, \dots, a_n)$ and $\delta_k(q; a_1, \dots, a_n)$. It would be nice to have results where x varies at the same time as q, k and n , rather than being sent to infinity first. As can be seen in Sect. 2, a key impediment to producing such results is that the formulation of the LI hypothesis is also rather qualitative and infinitary, stating that there are no linear relations amongst ordinates of zeros. To handle x in a quantitative manner, this hypothesis would need to be reworked as a more quantitative statement, that there are no *approximate* linear relations amongst ordinates of zeros up to a certain height (where the height and the notion of approximate relation would depend on x). It would certainly be interesting to try working out such an approach, but here we follow the simpler course of previous workers, and restrict attention to varying q, k and n when x has been sent to infinity.

Next we shall try to indicate the main ideas in the proofs of our theorems.

The work of Rubinstein and Sarnak [31] showing the existence of the logarithmic densities $\delta(q; a_1, \dots, a_n)$ (assuming GRH and LI) in fact shows that

$$\delta(q; a_1, \dots, a_n) = \mathbb{P}(X(q, a_1) > X(q, a_2) > \dots > X(q, a_n)), \tag{1.4}$$

where each $X(q, a)$ is a sum of the same independent random variables twisted by certain arithmetic coefficients depending on q and a . Using a quantitative multivariate form of the central limit theorem (which we extract from Stein's method, replacing direct characteristic function calculations in the previous literature), one can replace the $X(q, a)$ by jointly Gaussian random variables with the same means and covariances. Since the behaviour of Gaussians is entirely determined by those means and covariances, our task is then to obtain as much information as possible about them (on the number theory side), and deduce the best results we can on the ordering probabilities (on the probabilistic side).

Theorem 1.2 uses a relatively naive probabilistic treatment, namely a direct estimation of the relevant multivariate Gaussian density. The improvement over the result of Lamzouri [22] comes from substantially improved estimates for the average size of the covariances feeding into that density (they are typically small, so we are close to a standard multivariate Gaussian). These estimates rely on a harmonic analysis lemma related to the Hardy–Littlewood method, which we use to deduce that the differences $a_i - a_j$ cannot too often be divisible by large divisors of q .

All of our other theorems exploit more sophisticated tools for comparing a multivariate Gaussian distribution with the standard multivariate Gaussian, such as the famous comparison lemmas of Slepian (see e.g. Piterbarg [29]) and Li and Shao [24]. However, none of these tools seem directly able to prove our theorems, because the probabilities of the events we are interested in are rather small and the bounds we have on the off-diagonal covariances are comparatively large. For example, in the case of Theorem 1.4 we need to show that

$$\delta_1(q; a_1, \dots, a_n) = \mathbb{P}\left(X(q, a_1) > \max_{2 \leq j \leq n} X(q, a_j)\right) = \frac{1}{n}(1 + o(1)),$$

where potentially $1/n$ is as small as $1/\varphi(q)^{1/32} = 1/q^{1/32+o(1)}$, but where the largest off-diagonal covariances of the $X(q, a)$ (when they are normalised to have variance 1) are $\asymp 1/\log q$. To address this problem, we observe that if Z_1, Z_2, \dots, Z_n are independent standard normal random variables, then with very high probability we have

$$\max_{2 \leq j \leq n} Z_j = \sqrt{(2 - o(1)) \log n}, \quad \text{and} \quad \mathbb{P}(Z_1 > \sqrt{(2 - o(1)) \log n}) = \frac{1}{n^{1+o(1)}}.$$

In other words, *most of the size $1/n$ of $\mathbb{P}(Z_1 > \max_{2 \leq j \leq n} Z_j)$ is determined just by the probability of Z_1 being large enough to possibly be the leader*. This means that we can “factor out” most of the small size of our target probability $1/n$ by first conditioning on $X(q, a_1)$ being roughly large enough to be in the lead, leaving a more achievable error bound to be obtained from comparison inequalities. Since the $X(q, a_i)$ are not really independent of one another, this conditioning step itself requires some work and the use of Slepian's Lemma. Theorem 1.5 is a general probabilistic statement that we will prove using these methods.

Our result on the first k places in the race, Theorem 1.6, is proved by combining the direct density arguments of Theorem 1.2 with the conditioning arguments of

Theorem 1.4. However, since the size of our target probability is now much smaller compared with n the latter part of the argument becomes more challenging, and in fact will be the hardest element of this paper. In particular, we need to prove a modified normal comparison lemma incorporating within it an application of Slepian’s lemma, and when applying this we exploit the remarkable known fact that if the correlations of the random variables $X(q, a)$ have positive sign, then they are very small.

Finally, the negative result Theorem 1.7 is proved by noting that in an event

$$X(q, a_1) > X(q, a_2) > \dots > X(q, a_k) > \max_{k+1 \leq j \leq n} X(q, a_j),$$

with very high probability the random variables $X(q, a_1), \dots, X(q, a_k)$ will have (normalised) size $\sqrt{(2 - o(1)) \log n}$, and so in the relevant probability integral there will be terms in the exponential of size $\asymp \log n$. In particular, if there is a correlation between the random variables $X(q, a_i)$ of size about $1/\log q$, this will appear in the density in the exponential and will noticeably distort the multivariate normal probability if $(\log n)/\log q$ isn’t small. So by choosing the tuple (a_1, \dots, a_n) so that there is a correlation of size $\asymp 1/\log q$, which is (the maximum) possible, we obtain an ordering probability that does not converge to uniformity.

We end by explaining the organisation of the rest of the paper. In Sect. 2 we explicitly state the correspondence between logarithmic densities in prime races and orderings of suitable random variables. In Sect. 3 we prove average estimates for the covariances of those random variables. Section 4 contains various probabilistic tools tailored to our needs, as well as the proof of Theorem 1.5 and the new probabilistic preparations for Theorem 1.6. Sections 5 and 6 are relatively short, and contain the deductions of Theorems 1.2 and 1.4. Finally, Sect. 7 contains the somewhat difficult proof of Theorem 1.6, and Sect. 8 has the proof of our negative result, Theorem 1.7.

We have tried to use notation that will not cause confusion between readers from a more number theoretic or a more probabilistic background, but two brief remarks might be in order. Firstly, we use Vinogradov’s notation \ll , which has the same meaning as the “big Oh” notation (thus $x \ll x/10$, for example). In particular, \ll does *not* mean “much less than”. Secondly, if the implicit constant in a statement depends on some ambient parameter, we may adorn the notation with that parameter to reflect the dependence (e.g. we might write $f(x) = O_\epsilon(x^\epsilon)$, meaning that $|f(x)| \leq C(\epsilon)x^\epsilon$ for some $C(\epsilon)$).

2 Logarithmic densities of prime races and corresponding random variables

In order to investigate the distribution of prime numbers in arithmetic progressions, a standard tool in analytic number theory is to connect the prime counting functions $\pi(x; q, a)$ to zeros of Dirichlet L -functions, via the so-called “explicit formulas”. We shall briefly summarise some relevant aspects of this theory here, and the reader may consult Davenport’s book [2] for many more details.

As usual, we denote by Λ the von Mangoldt function, defined by $\Lambda(n) = \log p$ if $n = p^k$ for some prime p and positive integer k , and $\Lambda(n) = 0$ otherwise. For a Dirichlet character $\chi \pmod{q}$ we define

$$\psi(x, \chi) := \sum_{n \leq x} \chi(n) \Lambda(n).$$

Let χ_0 denote the principal character modulo q . For $\chi \neq \chi_0 \pmod{q}$ we have the following classical explicit formula (see for example Eq. (7) of Chapter 19 of [2]):

$$\psi(x, \chi) = - \sum_{|\operatorname{Im}(\rho_\chi)| \leq T} \frac{x^{\rho_\chi}}{\rho_\chi} + O_q \left(\frac{x \log^2(xT)}{T} + \log x \right), \quad (2.1)$$

where ρ_χ runs over the non-trivial zeros of the Dirichlet L -function $L(s, \chi)$. Now if a is a reduced residue modulo q , by the orthogonality of characters we deduce from (2.1) that

$$\begin{aligned} \sum_{\substack{n \leq x \\ n \equiv a \pmod{q}}} \Lambda(n) &= \frac{1}{\varphi(q)} \sum_{\chi \pmod{q}} \overline{\chi(a)} \psi(x, \chi) \\ &= \frac{\psi(x, \chi_0)}{\varphi(q)} - \frac{1}{\varphi(q)} \sum_{\substack{\chi \neq \chi_0 \\ \chi \pmod{q}}} \overline{\chi(a)} \sum_{|\operatorname{Im}(\rho_\chi)| \leq T} \frac{x^{\rho_\chi}}{\rho_\chi} \\ &\quad + O_q \left(\frac{x \log^2(xT)}{T} + \log x \right). \end{aligned}$$

If we assume the GRH, (which implies that $\Re \rho_\chi = 1/2$ for all the terms in the sum over zeros), then a summation by parts argument leads to (see Eqs. (2.3) and (2.5) of [31])

$$\begin{aligned} \pi(x; q, a) &= \frac{\pi(x)}{\varphi(q)} - \frac{C_q(a)}{\varphi(q)} \frac{\sqrt{x}}{\log x} - \frac{\sqrt{x}}{\varphi(q) \log x} \sum_{\substack{\chi \neq \chi_0 \\ \chi \pmod{q}}} \overline{\chi(a)} \sum_{|\gamma_\chi| \leq T} \frac{x^{i\gamma_\chi}}{1/2 + i\gamma_\chi} \\ &\quad + O_q \left(\frac{x \log^2(xT)}{T \log x} + \frac{\sqrt{x}}{\log^2 x} \right), \end{aligned}$$

where $\pi(x)$ denotes the total number of primes less than x , and γ_χ runs over the imaginary parts of the nontrivial zeros of $L(s, \chi)$, and

$$C_q(a) := -1 + |\{b \pmod{q} : b^2 \equiv a \pmod{q}\}|.$$

It is worth remarking here that the factor \sqrt{x} in the second and third terms in the preceding display comes from different sources. In the second term (which is the origin of Chebyshev's bias for small q), \sqrt{x} reflects the count of prime squares less than x , weighted by the von Mangoldt function. In the third term, $\sqrt{x} = x^{\Re \rho_\chi}$ emerges from the sum over zeros as a consequence of GRH.

Thus, if we define

$$E(x; q, a) := \frac{\log x}{\sqrt{x}} (\varphi(q)\pi(x; q, a) - \pi(x)),$$

then we have

$$\begin{aligned} E(x; q, a) &= -C_q(a) - \sum_{\substack{\chi \neq \chi_0 \\ \chi \pmod q}} \overline{\chi(a)} \sum_{|\gamma_\chi| \leq T} \frac{e^{i\gamma_\chi \log x}}{1/2 + i\gamma_\chi} \\ &\quad + O_q \left(\frac{\sqrt{x} \log^2(xT)}{T} + \frac{1}{\log x} \right) \\ &= -C_q(a) - 2\operatorname{Re} \left(\sum_{\substack{\chi \neq \chi_0 \\ \chi \pmod q}} \overline{\chi(a)} \sum_{0 < \gamma_\chi \leq T} \frac{e^{i\gamma_\chi \log x}}{1/2 + i\gamma_\chi} \right) \\ &\quad + O_q \left(\frac{\sqrt{x} \log^2(xT)}{T} + \frac{1}{\log x} \right). \end{aligned} \tag{2.2}$$

Here the second equality follows because if ρ_χ is a zero of $L(s, \chi)$, then its conjugate $\overline{\rho_\chi}$ is a zero of $L(s, \overline{\chi})$. The normalization of $E(x; q, a)$ is such that, under our assumption of GRH, $E(x; q, a)$ varies roughly boundedly as x varies.

If we assume LI then all of the positive values of γ_χ are linearly independent over \mathbb{Q} , and in particular are distinct. Let $\Gamma = \bigcup_{\chi \neq \chi_0 \pmod q} \{\gamma_\chi > 0\}$. Then assuming LI, it follows from Kronecker’s Theorem that for any finite subset $\{\gamma_1, \dots, \gamma_N\} \subset \Gamma$, the set of values $\{(y\gamma_1, \dots, y\gamma_N), y \in \mathbb{R}\}$ is *uniformly distributed* in \mathbb{U}^N , where \mathbb{U} is the unit circle. Combining this fact with the explicit formula (2.2), Rubinstein and Sarnak [31] proved (under GRH and LI) that if one looks at logarithmic densities, then $E(x; q, a)$ has a limiting distribution as $x \rightarrow \infty$. In fact they established a stronger result, which we shall now describe. Let a_1, \dots, a_n be distinct reduced residues modulo q , and define

$$E_{q; a_1, \dots, a_n}(x) := \left(E(x; q, a_1), \dots, E(x; q, a_n) \right).$$

Then, notice that

$$\pi(x; q, a_1) > \pi(x; q, a_2) > \dots > \pi(x; q, a_n) \iff E(x; q, a_1) > E(x; q, a_2) > \dots > E(x; q, a_n).$$

Assuming GRH and LI, Rubinstein and Sarnak [31] proved that for any Lebesgue measurable set $S \subset \mathbb{R}^n$ whose boundary has measure zero, we have

$$\delta_{q; a_1, \dots, a_n}(S) := \lim_{X \rightarrow \infty} \frac{1}{\log X} \int_{\substack{x \in [2, X] \\ E_{q; a_1, \dots, a_n}(x) \in S}} \frac{dx}{x} = \int_S d\mu_{q; a_1, \dots, a_n} \tag{2.3}$$

where $\mu_{q; a_1, \dots, a_n}$ is a certain probability measure on \mathbb{R}^n .

Indeed, under the same assumptions they obtained the following formula for the Fourier transform of $\mu_{q;a_1,\dots,a_n}$:

$$\hat{\mu}_{q;a_1,\dots,a_n}(\xi_1, \dots, \xi_n) = \exp\left(i \sum_{j=1}^n C_q(a_j) \xi_j\right) \prod_{\substack{\chi \neq \chi_0 \\ \chi \pmod{q}}} \prod_{\gamma_\chi > 0} J_0\left(\frac{2|\sum_{j=1}^n \chi(a_j) \xi_j|}{\sqrt{\frac{1}{4} + \gamma_\chi^2}}\right) \quad (2.4)$$

for $(\xi_1, \dots, \xi_n) \in \mathbb{R}^n$, where $J_0(z) = \sum_{m=0}^{\infty} (-1)^m (z/2)^{2m} / (m!)^2$ is the Bessel function of order 0. Since $J_0(z)$ is precisely the characteristic function of (the real part of) a random variable that is uniformly distributed on \mathbb{U} , one can deduce from (2.4) that $\mu_{q;a_1,\dots,a_n}$ is the probability measure corresponding to the random vector $(X(q, a_1), \dots, X(q, a_n))$, where

$$X(q, a) := -C_q(a) + \sum_{\substack{\chi \neq \chi_0 \\ \chi \pmod{q}}} \operatorname{Re} \left(2\chi(a) \sum_{\gamma_\chi > 0} \frac{U(\gamma_\chi)}{\sqrt{\frac{1}{4} + \gamma_\chi^2}} \right),$$

and $\{U(\gamma_\chi)\}_{\gamma_\chi \in \Gamma}$ is a sequence of independent random variables uniformly distributed on \mathbb{U} . Note that for $(a, q) = 1$ the function $C_q(a)$ takes only two values: $C_q(a) = -1$ if a is a non-square modulo q , and $C_q(a) = C_q(1)$ if a is a square modulo q . An elementary argument shows that $C_q(a) < d(q) \ll_\epsilon q^\epsilon$ for any $\epsilon > 0$, where $d(q) = \sum_{m|q} 1$ is the usual divisor function. Thus it will turn out that the shifts $C_q(a)$ can essentially be ignored when $q \rightarrow \infty$.

Let $\operatorname{Cov}_{q;a_1,\dots,a_n}$ be the covariance matrix of $(X(q, a_1), \dots, X(q, a_n))$. Then a straightforward computation (see also Lemma 2.1 of [22], for example) shows that

$$\operatorname{Cov}_{q;a_1,\dots,a_n}(i, j) = \begin{cases} \operatorname{Var}(q) & \text{if } i = j \\ B_q(a_i, a_j) & \text{if } i \neq j, \end{cases}$$

where

$$\operatorname{Var}(q) := 2 \sum_{\substack{\chi \neq \chi_0 \\ \chi \pmod{q}}} \sum_{\gamma_\chi > 0} \frac{1}{\frac{1}{4} + \gamma_\chi^2}, \quad \text{and } B_q(a, b) := \sum_{\substack{\chi \neq \chi_0 \\ \chi \pmod{q}}} \sum_{\gamma_\chi > 0} \frac{\chi\left(\frac{b}{a}\right) + \chi\left(\frac{a}{b}\right)}{\frac{1}{4} + \gamma_\chi^2}.$$

We end this section by recording several basic estimates for the quantities $\operatorname{Var}(q)$ and $B_q(a, b)$ that will be useful in our subsequent work.

Lemma 2.1 *Assume GRH. Then for any non-principal character $\chi \pmod{q}$,*

$$\sum_{\gamma_\chi > 0} \frac{1}{\frac{1}{4} + \gamma_\chi^2} \ll \log q. \quad (2.5)$$

Moreover, we have

$$\text{Var}(q) \sim \varphi(q) \log q \quad \text{as } q \rightarrow \infty. \tag{2.6}$$

Proof These estimates follow from Lemma 3.1 of [23], for example. □

Lemma 2.2 *Assume GRH. For all distinct reduced residues $a, b \pmod q$, we have*

$$B_q(a, b) \ll \varphi(q). \tag{2.7}$$

Moreover, if a, b are distinct residue classes such that $1 \leq |a| < |b| \leq q/2$ and $|b|/|a|$ is not a prime power, then we have

$$B_q(a, b) \ll |b|(\log q)^2. \tag{2.8}$$

On the other hand we have

$$B_q(1, -1) = -(\log 2)\varphi(q) + O\left((\log q)^2\right). \tag{2.9}$$

Finally, if a, b are distinct residue classes and $B_q(a, b) \geq 0$ then

$$B_q(a, b) \ll \log q. \tag{2.10}$$

Proof The first bound (2.7) corresponds to Corollary 5.4 of [23]. The estimates (2.8) and (2.9) follow from Proposition 6.1 of [23]. The fact (2.10) that positive correlations are always very small is noted in Remark 5.1 of [23], for example. □

3 An average result for the sums of the covariances $B_q(a_i, a_j)$

3.1 A double average

In view of Lemma 2.2, all of the non-diagonal covariances in our prime number race satisfy (when normalised by the variance $\text{Var}(q)$)

$$\frac{|B_q(a_i, a_j)|}{\text{Var}(q)} \ll \frac{1}{\log q}, \quad i \neq j.$$

This bound is useful, but to obtain strong results we will need to exploit the fact that, if we are looking at many residue classes a_1, \dots, a_n , the covariances will on average be much smaller. This was established in Theorem 5 of Lamzouri [23] when averaging over *all* pairs of distinct reduced residue classes, but we will need a strong result when averaging only over a subset, which requires quite different methods.

Correlation Estimate 1 *Assume GRH. Let q be large and let $r, s \geq 1$. For any collections a_1, \dots, a_r and b_1, \dots, b_s of distinct reduced residue classes modulo q , we have*

$$\sum_{1 \leq j \leq r} \sum_{\substack{1 \leq k \leq s, \\ b_k \neq a_j}} \frac{|B_q(a_j, b_k)|}{\text{Var}(q)} \ll \frac{\sqrt{rs} \log^2(2rs)}{\log q}.$$

In particular, we have

$$\sum_{\substack{1 \leq k \leq s, \\ b_k \neq a_1}} \frac{|B_q(a_1, b_k)|}{\text{Var}(q)} \ll \frac{\sqrt{s} \log^2(2s)}{\log q}, \quad \text{and} \quad \sum_{\substack{1 \leq j, k \leq r, \\ j \neq k}} \frac{|B_q(a_j, a_k)|}{\text{Var}(q)} \ll \frac{r \log^2(2r)}{\log q}.$$

Note that this estimate saves roughly a factor of \sqrt{rs} as compared with a trivial treatment using the pointwise bound $\frac{|B_q(a_j, b_k)|}{\text{Var}(q)} \ll 1/\log q$.

To prove Correlation Estimate 1 we shall need two lemmas. The first is the following, which will reduce the problem to upper bounding some easier sums.

Lemma 3.1 *Assume GRH. In the setting of Correlation Estimate 1, for any $1 \leq j \leq r$ we have*

$$\sum_{\substack{1 \leq k \leq s, \\ b_k \neq a_j}} \frac{|B_q(a_j, b_k)|}{\varphi(q)} \ll \log^2(2s) + \sum_{\substack{1 \leq k \leq s, \\ b_k \neq a_j}} \frac{\Lambda(q/(q, a_j - b_k))}{\varphi(q/(q, a_j - b_k))},$$

where $\Lambda(n)$ denotes the von Mangoldt function.

Proof of Lemma 3.1 Let $x = (q \log q)^2$, and for simplicity of writing set $a = a_j$. Then it follows from Proposition 5.1 of [22] that

$$\begin{aligned} \sum_{\substack{1 \leq k \leq s, \\ b_k \neq a}} |B_q(a, b_k)| &\ll \varphi(q) \left(1 + \sum_{\substack{1 \leq k \leq s, \\ b_k \neq a}} \left(\frac{\Lambda\left(\frac{q}{(q, a - b_k)}\right)}{\varphi\left(\frac{q}{(q, a - b_k)}\right)} + M_1(q; a, b_k) \right. \right. \\ &\quad \left. \left. + M_1(q; b_k, a) + M_2(q; a, b_k) + M_2(q; b_k, a) \right) \right) + s \log q, \end{aligned} \tag{3.1}$$

where

$$\begin{aligned} M_1(q; a, d) &= \sum_{\substack{n \leq 2x \log x \\ an \equiv d \pmod{q}}} \frac{\Lambda(n)}{n} e^{-n/x}, \quad \text{and} \\ M_2(q; a, d) &= \sum_{p^v \parallel q} \sum_{\substack{1 \leq e \leq 2 \log x \\ ap^e \equiv d \pmod{q/p^v}}} \frac{\log p}{p^{e+v-1}(p-1)}, \end{aligned}$$

and $p^v \parallel q$ denotes that p^v is the largest power of p that divides q .

Now it follows from Lemma 5.3 of [22] that

$$M_1(q; a, b_k) = \frac{\Lambda(n_k)}{n_k} + O\left(\frac{\log^2 q}{q}\right) \leq \frac{\log(n_k)}{n_k} + O\left(\frac{\log^2 q}{q}\right),$$

where n_k is the least positive residue of $b_k a^{-1}$ modulo q . Since $(\log n)/n$ is a decreasing function for $n \geq 3$, we deduce that

$$\begin{aligned} \sum_{\substack{1 \leq k \leq s, \\ b_k \neq a}} M_1(q; a, b_k) &\ll \sum_{\substack{1 \leq k \leq s, \\ b_k \neq a}} \frac{\log(n_k)}{n_k} + \frac{s \log^2 q}{q} \ll 1 + \sum_{k \leq s} \frac{\log k}{k} + \frac{s \log^2 q}{q} \\ &\ll \log^2(2s) + \frac{s \log^2 q}{q}. \end{aligned}$$

A similar bound holds for $\sum M_1(q; b_k, a)$.

Next we bound the sum $\sum (M_2(q; a, b_k) + M_2(q; b_k, a))$. We have

$$\begin{aligned} \sum_{k=1}^s M_2(q; a, b_k) &= \sum_{p^v \parallel q} \sum_{e \leq 2 \log x} \frac{\log p}{p^{e+v-1}(p-1)} \sum_{\substack{1 \leq k \leq s \\ ap^e \equiv b_k \pmod{q/p^v}}} 1 \\ &\leq \sum_{p^v \parallel q} \sum_{e \leq 2 \log x} \frac{\log p}{p^{e+v-1}(p-1)} \min\{p^v, s\}, \end{aligned}$$

since the b_k are distinct modulo q . Splitting the outer sum over the primes p dividing q into the cases $p \leq s$ and $p > s$, we find the above is

$$\ll \sum_{p \leq s} \sum_{e=1}^{\infty} \frac{\log p}{p^{e-1}(p-1)} + s \sum_{p > s} \sum_{e=1}^{\infty} \frac{\log p}{p^e(p-1)} \ll \sum_{p \leq s} \frac{\log p}{p} + s \sum_{p > s} \frac{\log p}{p^2} \ll \log(2s).$$

A similar bound holds for $\sum_{k=1}^s M_2(q; b_k, a)$. Putting everything together (and remembering that $s \leq q$, so $(s \log^2 q)/q \ll \log^2(2s)$) completes the proof of Lemma 3.1. □

To control the sum on the right hand side in Lemma 3.1 we shall deploy the following harmonic analysis lemma. This will be the really new aspect of our analysis of the covariances.

Lemma 3.2 (Following pp 305–307 of Bourgain [1], 1989) *Let x be large and let $Q \geq 1$. Define*

$$G(\theta) := \sum_{q \leq Q} \frac{\Lambda(q)}{q} \sum_{a=0}^{q-1} \mathbf{I}_{\|\theta - a/q\| \leq 1/x},$$

where $\|\cdot\|$ denotes distance to the nearest integer, and \mathbf{I} denotes the indicator function.

Then if $\theta_1, \dots, \theta_R$ and ϕ_1, \dots, ϕ_S are any real numbers that are $1/x$ -spaced (i.e. such that $\|\theta_{r_1} - \theta_{r_2}\| \geq 1/x$ when $r_1 \neq r_2$, and such that $\|\phi_{s_1} - \phi_{s_2}\| \geq 1/x$ when $s_1 \neq s_2$), we have

$$\sum_{\substack{1 \leq r \leq R, \\ 1 \leq s \leq S}} G(\theta_r - \phi_s) \ll \sqrt{RS} \log^2(2QRS) + \frac{RSQ}{x}.$$

Lemma 3.2 encodes the fact that rationals a/q are well spaced, so it is impossible for lots of the points $\theta_r - \phi_s$ to be very close to lots of rationals. The proof uses additive characters, the problem being analytically nice because we are looking at pairwise differences, which corresponds to a convolution on the harmonic analysis side.

Proof of Lemma 3.2 Lemma 3.2 is inspired by a result in work of Bourgain [1] (in a substantially different context), but differs in that it is concerned with two sets of points θ_r and ϕ_s rather than one, and it involves a weight $\Lambda(q)/q$ rather than $1/q$ (where $\Lambda(q)$ denotes the von Mangoldt function). It turns out that the former adaptation is easy, and the latter one simplifies and strengthens the argument, since all work with divisor functions becomes much easier. For the sake of completeness, we provide a fairly complete proof here.

Recall that we are given two sets $\theta_1, \dots, \theta_R$ and ϕ_1, \dots, ϕ_S of $1/x$ -spaced real numbers. Corresponding to these, let $I_\theta, I_\phi : \mathbb{R}/\mathbb{Z} \rightarrow [0, 10]$ be bounded variation continuous functions on the real numbers mod 1, that satisfy

$$I_\theta(t) \geq \sum_{1 \leq r \leq R} \mathbf{1}_{\|t - \theta_r\| \leq 1/x} \quad \text{and} \quad I_\phi(t) \geq \sum_{1 \leq s \leq S} \mathbf{1}_{\|t - (-\phi_s)\| \leq 1/x},$$

(note the negative signs attached to the ϕ_s here), and also

$$\int_0^1 I_\theta(t) dt \leq \frac{1000R}{x}, \quad \int_0^1 I_\phi(t) dt \leq \frac{1000S}{x}, \quad \text{supp}(\hat{I}_\theta), \text{supp}(\hat{I}_\phi) \subseteq [-x, x],$$

where $\hat{I}_\theta(k) := \int_0^1 I_\theta(t) e^{-2\pi i t k} dt$ and $\hat{I}_\phi(k)$ (for $k \in \mathbb{Z}$) denote the Fourier transforms of I_θ, I_ϕ . It is a standard fact that one can construct such functions I , as a sum of Beurling–Selberg type smooth functions that approximate $\mathbf{1}_{\|t - \theta_r\| \leq 1/x}$ and $\mathbf{1}_{\|t - (-\phi_s)\| \leq 1/x}$, and each have integral at most $1000/x$ and Fourier transform supported in $[-x, x]$ (see e.g. Chapter 1.2 of Montgomery's book [27] for details about Beurling–Selberg functions). Then we see immediately that the convolution

$$\begin{aligned} (I_\theta * I_\phi)(t) &:= \int_0^1 I_\theta(u) I_\phi(t - u) du \geq \sum_{\substack{1 \leq r \leq R, \\ 1 \leq s \leq S}} \int_0^1 \mathbf{1}_{\|u - \theta_r\| \leq \frac{1}{x}} \mathbf{1}_{\|t - u + \phi_s\| \leq \frac{1}{x}} du \\ &\geq \frac{1}{x} \sum_{\substack{1 \leq r \leq R, \\ 1 \leq s \leq S}} \mathbf{1}_{\|t - (\theta_r - \phi_s)\| \leq \frac{1}{x}}. \end{aligned}$$

Consequently, in Lemma 3.2 we have

$$\sum_{\substack{1 \leq r \leq R, \\ 1 \leq s \leq S}} G(\theta_r - \phi_s) = \sum_{q \leq Q} \frac{\Lambda(q)}{q} \sum_{a=0}^{q-1} \sum_{\substack{1 \leq r \leq R, \\ 1 \leq s \leq S}} \mathbf{1}_{\|(\theta_r - \phi_s) - a/q\| \leq \frac{1}{x}}$$

$$\leq x \sum_{q \leq Q} \frac{\Lambda(q)}{q} \sum_{a=0}^{q-1} (I_\theta * I_\phi)(a/q).$$

On writing $(I_\theta * I_\phi)(a/q)$ in terms of its Fourier coefficients, and using that the geometric progression $\sum_{a=0}^{q-1} e^{2\pi ika/q}$ vanishes unless q divides k (in which case the sum equals q), we find

$$\sum_{a=0}^{q-1} (I_\theta * I_\phi)(a/q) = \sum_{a=0}^{q-1} \sum_{k=-\infty}^{\infty} (\widehat{I_\theta * I_\phi})(k) e^{2\pi ika/q} = q \sum_{\substack{k=-\infty, \\ q|k}}^{\infty} (\widehat{I_\theta * I_\phi})(k).$$

Using also the fact that $(\widehat{I_\theta * I_\phi})(k) = \widehat{I_\theta}(k)\widehat{I_\phi}(k)$, we conclude overall that

$$\sum_{\substack{1 \leq r \leq R, \\ 1 \leq s \leq S}} G(\theta_r - \phi_s) \leq x \sum_{q \leq Q} \Lambda(q) \sum_{\substack{k=-\infty, \\ q|k}}^{\infty} \widehat{I_\theta}(k)\widehat{I_\phi}(k) \leq x \sum_{k=-\infty}^{\infty} \sum_{\substack{q \leq Q, \\ q|k}} \Lambda(q) |\widehat{I_\theta}(k)| |\widehat{I_\phi}(k)|.$$

To finish, we define $\mathcal{B} := \{-x \leq k \leq x : k \neq 0, \sum_{q \leq Q, q|k} \Lambda(q) \geq \log^2(2QRS)\}$, and note that since $\widehat{I_\theta}, \widehat{I_\phi}$ are supposed to vanish outside the interval $[-x, x]$, and since we always have $\sum_{q \leq Q} \Lambda(q) \leq \sum_{q \leq Q} \Lambda(q) \ll Q$ and also $|\widehat{I_\theta}(k)| \leq \int_0^1 I_\theta(t) dt \leq \frac{1000R}{x}$ and $|\widehat{I_\phi}(k)| \leq \frac{1000S}{x}$, the right hand side above is

$$\ll \log^2(2QRS)x \sum_{-x \leq k \leq x} |\widehat{I_\theta}(k)| |\widehat{I_\phi}(k)| + Qx |\widehat{I_\theta}(0)| |\widehat{I_\phi}(0)| + Qx \sum_{k \in \mathcal{B}} |\widehat{I_\theta}(k)| |\widehat{I_\phi}(k)|$$

$$\ll \log^2(2QRS)x \sqrt{\sum_k |\widehat{I_\theta}(k)|^2} \sqrt{\sum_k |\widehat{I_\phi}(k)|^2} + \frac{RSQ}{x} + \frac{RSQ}{x} \#\mathcal{B}.$$

The second term above is acceptable for Lemma 3.2. By Parseval’s Identity we have

$$\sqrt{\sum_k |\widehat{I_\theta}(k)|^2} \sqrt{\sum_k |\widehat{I_\phi}(k)|^2} \ll \sqrt{\int_0^1 I_\theta(t)^2 dt} \sqrt{\int_0^1 I_\phi(t)^2 dt} \ll \frac{\sqrt{RS}}{x},$$

and so the first term above is also acceptable.

Turning to the third term, since we always have $\sum_{q \leq Q} \Lambda(q) \leq \sum_{q|k} \Lambda(q) = \log |k|$ provided $k \neq 0$, there can be no elements of \mathcal{B} having modulus smaller than $\exp\{\log^2(2QRS)\}$. Moreover, if $\exp\{\log^2(2QRS)\} \leq k \leq x$ and $\sum_{q \leq Q} \Lambda(q) \geq \log^2(2QRS)$ then in fact we must be able to write $k = nm$, where n is Q smooth (i.e. all the prime factors of n are at most Q) and also $n \geq \exp\{\log^2(2QRS)\}$. Therefore, picking up a factor of 2 from counting positive and negative k , we have the upper bound

$$\#\mathcal{B} \leq 2 \sum_{\substack{\exp\{\log^2(2QRS)\} \leq n \leq x, \\ n \text{ is } Q \text{ smooth}}} \sum_{m \leq x/n} 1 \leq 2 \sum_{\substack{\exp\{\log^2(2QRS)\} \leq n \leq x, \\ n \text{ is } Q \text{ smooth}}} \frac{x}{n}.$$

And for any $\exp\{\log^2(2QRS)\} \leq N \leq x$, standard upper bounds for the counting function of smooth numbers (see e.g. Theorem 7.6 of Montgomery and Vaughan [28]) imply that the dyadic sum

$$\begin{aligned} \sum_{\substack{N \leq n \leq 2N, \\ n \text{ is } Q \text{ smooth}}} \frac{1}{n} &\leq \frac{1}{N} \sum_{\substack{N \leq n \leq 2N, \\ n \text{ is } Q \text{ smooth}}} 1 \ll (\log 2Q) \exp \left\{ -\frac{\log N}{\log 2Q} \log \left(\frac{\log N}{\log 2Q} \right) \right\} \\ &\ll \frac{1}{(QRS)^{10} \log^2 N}, \end{aligned}$$

say. So splitting up the full sum by taking $N = 2^j \exp\{\log^2(2QRS)\}$ for $j \geq 0$, we obtain that $\#\mathcal{B} \ll x/(QRS)^{10}$. It follows that $\frac{RSQ}{x} \#\mathcal{B} \ll 1/(QRS)^9 \leq 1$, which is certainly an acceptable contribution for Lemma 3.2. \square

Proof of Correlation Estimate 1 We may suppose without loss of generality that $s \geq r$. In view of Lemma 3.1, and the fact that $\text{Var}(q) \sim \varphi(q) \log q$ that is contained in Lemma 2.1, to prove Correlation Estimate 1 it will certainly suffice to prove that

$$\sum_{1 \leq j \leq r} \sum_{\substack{1 \leq k \leq s, \\ b_k \neq a_j}} \frac{\Lambda(q/(q, a_j - b_k))}{\varphi(q/(q, a_j - b_k))} \ll \sqrt{rs} \log^2(2rs).$$

Rewriting a little, the double sum is

$$\leq \sum_{n|q} \frac{\Lambda(n)}{\varphi(n)} \sum_{1 \leq j \leq r} \sum_{\substack{1 \leq k \leq s, \\ b_k \neq a_j}} \mathbf{1}_{a_j \equiv b_k \pmod{q/n}} \ll \sum_{n|q} \frac{\Lambda(n)}{n} \sum_{1 \leq j \leq r} \sum_{\substack{1 \leq k \leq s, \\ b_k \neq a_j}} \mathbf{1}_{a_j \equiv b_k \pmod{q/n}},$$

where $\mathbf{1}$ denotes the indicator function, and where we used the fact that $\varphi(n) \asymp n$ if n is a prime power (and so $\Lambda(n) \neq 0$). We also observe that we cannot have $a_j \equiv b_k$ modulo q/n for two values n that are powers of different primes, since in that

case we would have $a_j \equiv b_k$ modulo q , which is false by assumption. Therefore the contribution to the sum from those $n > rs$ is trivially

$$\ll \sum_{1 \leq j \leq r} \sum_{\substack{1 \leq k \leq s, \\ b_k \neq a_j}} \frac{\log(2rs)}{rs} \leq \log(2rs),$$

which is acceptable.

To bound the contribution to the sum from those $n \leq rs$, we apply Lemma 3.2 with the choices $Q = rs$, $x = \max\{q, (rs)^2\}$, $\theta_j := a_j/q$ and $\phi_k := b_k/q$ for all j, k . Thus all the points θ_j and ϕ_k are indeed $1/x$ -spaced, and $a_j \equiv b_k$ modulo q/n if and only if $\theta_j - \phi_k = u/n$ for some integer u . We deduce that

$$\begin{aligned} \sum_{n \leq rs, n|q} \frac{\Lambda(n)}{n} \sum_{1 \leq j \leq r} \sum_{\substack{1 \leq k \leq s, \\ b_k \neq a_j}} \mathbf{1}_{a_j \equiv b_k \pmod{q/n}} &\leq \sum_{\substack{1 \leq j \leq r, n \leq rs \\ 1 \leq k \leq s}} \sum_{u=0}^{n-1} \frac{\Lambda(n)}{n} \mathbf{1}_{\|(\theta_j - \phi_k) - u/n\| \leq 1/x} \\ &\ll \sqrt{rs} \log^2(2rs) + 1, \end{aligned}$$

which is enough to prove Correlation Estimate 1. □

3.2 Some matrix estimates involving the covariances

We will need some information about the determinant and inverse of a covariance matrix that is close to the identity matrix. Let $\mathcal{M}_n(\epsilon)$ denote the set of all $n \times n$ symmetric matrices whose diagonal entries are 1, and whose off-diagonal entries have absolute value at most ϵ .

Lemma 3.3 *If $\epsilon \leq 1/2n$ then for any $A = (a_{j,k}) \in \mathcal{M}_n(\epsilon)$ we have*

$$\det(A) = 1 + O\left(\epsilon \sum_{\substack{1 \leq j, k \leq n, \\ j \neq k}} |a_{j,k}|\right).$$

In addition, if $\epsilon \leq 1/2n$ then A is invertible, and if we let $\tilde{a}_{j,k}$ denote the entries of the inverse matrix A^{-1} then we have

$$\tilde{a}_{j,k} = \begin{cases} 1 + O\left(\epsilon \sum_{\substack{1 \leq l, m \leq n, \\ l \neq m}} |a_{l,m}|\right) & \text{if } j = k, \\ O\left(|a_{j,k}| + \sum_{i \neq j, k} |a_{j,i}| |a_{i,k}| + \epsilon^2 \sum_{\substack{1 \leq l, m \leq n, \\ l \neq m}} |a_{l,m}|\right) & \text{if } j \neq k. \end{cases}$$

Lemma 3.3 extends Lemmas 4.1 and 4.2 of Lamzouri [22], which would give roughly the same result if all the off-diagonal entries $a_{j,k}$ had size about ϵ , but are weaker if the $a_{j,k}$ are on average smaller (as will later be the case for us when we take ϵ of size about $1/\log q$).

Proof of Lemma 3.3 We have

$$\det(A) = 1 + \sum_{\substack{\sigma \in S_n, \\ \sigma \neq 1}} \operatorname{sgn}(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)},$$

where S_n denotes the symmetric group on n elements. We divide the sum according to the number t of points that are *not* fixed by a permutation σ . Thus the only term with $t = 0$ is the identity permutation, which we removed from the sum; there are no terms with $t = 1$; and the contribution from $t = 2$ has size at most

$$\sum_{1 \leq j \leq n} \sum_{\substack{1 \leq k \leq n, \\ k \neq j}} |a_{j,k}|^2 \leq \epsilon \sum_{\substack{1 \leq j, k \leq n, \\ j \neq k}} |a_{j,k}|.$$

For any $3 \leq t \leq n$, by averaging the total contribution is at most

$$\frac{1}{t} \sum_{1 \leq j \leq n} \sum_{\substack{\sigma \in S_n, \sigma(j) \neq j, \\ \sigma \text{ has } t \text{ non-fixed points}}} |a_{j,\sigma(j)}| \epsilon^{t-1} = \frac{1}{t} \sum_{\substack{1 \leq j, k \leq n, \\ j \neq k}} |a_{j,k}| \epsilon^{t-1} \sum_{\substack{\sigma \in S_n, \sigma(j)=k, \\ \sigma \text{ has } t \text{ non-fixed points}}} 1.$$

In the inner sum we have $\binom{n-2}{t-2}$ choices of points that are not fixed (in addition to j and k), and for any such choice there are at most $(t-1)!$ ways to construct σ such that $\sigma(j) = k$. Thus the total contribution from any $3 \leq t \leq n$ is

$$\leq \sum_{\substack{1 \leq j, k \leq n, \\ j \neq k}} |a_{j,k}| \epsilon^{t-1} \frac{1}{t} \binom{n-2}{t-2} (t-1)! \leq \sum_{\substack{1 \leq j, k \leq n, \\ j \neq k}} |a_{j,k}| \epsilon (\epsilon n)^{t-2},$$

and our claim about $\det(A)$ follows on using the assumption $\epsilon \leq 1/2n$ and summing over t . In particular, we see that $\det(A) \geq 1/2$ for all $A \in \mathcal{M}_n(\epsilon)$, so A is invertible.

Next we need to prove the claims about $\tilde{a}_{j,k}$. If $j = k$ then we have

$$\tilde{a}_{j,j} = \frac{\det(A_{j,j})}{\det(A)},$$

where $A_{j,j}$ denotes the matrix A with the j th row and column removed. In particular we have $A_{j,j} \in \mathcal{M}_{n-1}(\epsilon)$, and so the claim about $\tilde{a}_{j,j}$ follows from our determinant results.

In the off-diagonal case we have

$$|\tilde{a}_{j,k}| = \left| \frac{\det(A_{k,j})}{\det(A)} \right| = O(|\det(A_{k,j})|) = O\left(\sum_{\sigma \in \mathcal{B}_{k,j}} \prod_{i \neq k} |a_{i,\sigma(i)}| \right),$$

where now $\mathcal{B}_{k,j}$ denotes the set of all bijections from $\{1, 2, \dots, n\} \setminus \{k\}$ to $\{1, 2, \dots, n\} \setminus \{j\}$. We again divide the sum according to the number t of points that are *not* fixed

by σ , noting that since $k \neq j$ the point j is necessarily always a non-fixed point, and the point k will always be the image of a non-fixed point. Thus there are no terms with $t = 0$, and the contribution from $t = 1$ is simply $|a_{j,k}|$. When $t = 2$ the contribution is

$$\sum_{i \neq j,k} |a_{j,i}| |a_{i,k}|.$$

For any $3 \leq t \leq n - 1$, each σ must have at least $t - 2$ non-fixed points that are different from j , and whose image is different from k , so by averaging the total contribution is at most

$$\begin{aligned} & \frac{1}{t-2} \sum_{\substack{1 \leq i \leq n, \\ i \neq k, j}} \sum_{\substack{\sigma \in \mathcal{B}_{k,j}, \sigma(i) \neq i, j, k, \\ \sigma \text{ has } t \text{ non-fixed points}}} |a_{i,\sigma(i)}| \epsilon^{t-1} \\ &= \frac{1}{t-2} \sum_{i \neq k, j} \sum_{h \neq i, j, k} |a_{i,h}| \epsilon^{t-1} \sum_{\substack{\sigma \in \mathcal{B}_{k,j}, \sigma(i)=h, \\ \sigma \text{ has } t \text{ non-fixed points}}} 1. \end{aligned}$$

In the inner sum we have $\binom{n-4}{t-3}$ choices of points that are not fixed (in addition to j, i and h), and for any such choice there are at most $(t - 1)!$ ways to construct σ such that $\sigma(i) = h$. Thus the total contribution from any $3 \leq t \leq n - 1$ is

$$\leq \sum_{\substack{1 \leq i, h \leq n, \\ i \neq h}} |a_{i,h}| \epsilon^{t-1} \frac{1}{t-2} \binom{n-4}{t-3} (t-1)! \leq \sum_{\substack{1 \leq i, h \leq n, \\ i \neq h}} |a_{i,h}| \epsilon^2 (\epsilon n)^{t-3} (t-1).$$

Proposition 1 follows on using the assumption $\epsilon \leq 1/2n$ and summing over t . □

4 Probabilistic tools and results

4.1 Passing to the Gaussian case

As described in Sect. 2, the prime number race between residue classes a modulo q is associated with random variables of the general shape

$$W_a := \sum_{i=1}^m \Re(c_i(a) V_i),$$

where $(V_i)_{1 \leq i \leq m}$ is a sequence of independent, mean zero, complex valued random variables (depending on q), and where $c_i(a) \in \mathbb{C}$ are deterministic coefficients.

In order to access the tools associated with Gaussian random processes, we would like a multivariate normal approximation (i.e. multivariate central limit theorem) for the n -dimensional random vector $W = (W_{a_j})_{1 \leq j \leq n}$. This will allow us to replace the

W_{a_j} by Gaussian random variables with the same means and covariances. We need an explicit bound for the error arising in the normal approximation, which in particular makes clear the dependence on n . There are not too many such results in the literature, and we will deduce a suitable result from the work of Reinert and Röllin [30].

Lemma 4.1 (Following Theorem 2.1 of Reinert and Röllin [30], 2009) *Let the situation be as described above, let \mathcal{A} be a finite set, and let $Z = (Z_a)_{a \in \mathcal{A}}$ denote a multivariate normal random vector with the same mean vector and covariance matrix as $W = (W_a)_{a \in \mathcal{A}}$. Assume that $\mathbb{E}|V_i|^4 \leq K^4/m^2$ for all i , for some $K \geq 1$.*

Then for any three times differentiable function $h : \mathbb{R}^{\#\mathcal{A}} \rightarrow \mathbb{R}$ we have

$$|\mathbb{E}h(W) - \mathbb{E}h(Z)| \ll \frac{|h|_2 K^2}{m} \sum_{a,b \in \mathcal{A}} \sqrt{\sum_{i=1}^m |c_i(a)|^2 |c_i(b)|^2} + \frac{|h|_3 K^3}{m^{3/2}} \sum_{i=1}^m \left(\sum_{a \in \mathcal{A}} |c_i(a)| \right)^3,$$

where $|h|_2 := \sup_{a,b \in \mathcal{A}} \|\frac{\partial^2}{\partial x_a \partial x_b} h\|_\infty$ and $|h|_3 := \sup_{a,b,c \in \mathcal{A}} \|\frac{\partial^3}{\partial x_a \partial x_b \partial x_c} h\|_\infty$.

Reinert and Röllin's work develops a multivariate version of Stein's method of exchangeable pairs, and applies in far more general situations than above. In Appendix A we very briefly indicate how to deduce Lemma 4.1 from Theorem 2.1 of Reinert and Röllin [30] (this deduction being, by now, a fairly standard calculation).

For the special case of prime number races, we centre and normalize the random variables $X(q, a_j)$ from Sect. 2 by setting

$$Y_j := \frac{X(q, a_j) + C_q(a_j)}{\sqrt{\text{Var}(q)}} = \frac{1}{\sqrt{\text{Var}(q)}} \sum_{\substack{\chi \bmod q, \\ \chi \neq \chi_0}} \Re \left(2\chi(a_j) \sum_{\gamma_\chi > 0} \frac{U(\gamma_\chi)}{\sqrt{\frac{1}{4} + \gamma_\chi^2}} \right), \quad 1 \leq j \leq n,$$

where the $U(\gamma_\chi)$ are independent random variables distributed uniformly on the unit circle. Then Y_1, \dots, Y_n have mean zero and variance 1. Moreover, we have

$$\mathbb{E}Y_i Y_j = \frac{B_q(a_i, a_j)}{\text{Var}(q)} \ll \frac{1}{\log q}, \quad i \neq j,$$

by (2.6) and (2.7). In this case, we obtain the following corollary.

Lemma 4.2 *Let $Z = (Z_j)_{1 \leq j \leq n}$ denote a multivariate normal random vector whose components have mean zero, variance one, and correlations*

$$\mathbb{E}Z_i Z_j := \mathbb{E}Y_i Y_j = \frac{B_q(a_i, a_j)}{\text{Var}(q)}.$$

Then for any three times differentiable function $h : \mathbb{R}^n \rightarrow \mathbb{R}$ we have

$$|\mathbb{E}h(Y) - \mathbb{E}h(Z)| \ll \frac{n^2 |h|_2 + n^3 |h|_3}{\sqrt{\varphi(q)}}.$$

Proof of Lemma 4.2 We apply Lemma 4.1 with the sum over $1 \leq i \leq m$ replaced by a sum over characters $\chi \neq \chi_0 \pmod q$ (so $m = \varphi(q) - 1$), and with $c_i(a_j)$ replaced by $\chi(a_j)$ and with V_i replaced by

$$V_\chi := \frac{2}{\sqrt{\text{Var}(q)}} \sum_{\gamma_\chi > 0} \frac{U(\gamma_\chi)}{\sqrt{\frac{1}{4} + \gamma_\chi^2}}.$$

The V_χ are indeed independent, mean zero random variables. Moreover, observe that $\mathbb{E}U(\gamma_{\chi_1})U(\gamma_{\chi_2})\overline{U(\gamma_{\chi_3})}U(\gamma_{\chi_4})$ vanishes unless $\{\chi_1, \chi_2\} = \{\chi_3, \chi_4\}$, and in that case it is $\ll 1$. Therefore we have

$$\mathbb{E}|V_\chi|^4 \ll \frac{1}{\text{Var}(q)^2} \left(\sum_{\gamma_\chi > 0} \frac{1}{\frac{1}{4} + \gamma_\chi^2} \right)^2 \ll \frac{\log^2 q}{\text{Var}(q)^2} \ll \frac{1}{\varphi(q)^2},$$

where the final inequalities follow from Lemma 2.1. So we have checked that Lemma 4.1 is applicable with K an absolute constant.

Using the facts that $|c_i(a_j)| = |\chi(a_j)| \leq 1$ and $\#\mathcal{A} = n$ and $m = \varphi(q) - 1$, Lemma 4.2 now follows immediately from Lemma 4.1. □

The above still isn't quite what we need, since we are interested in the probabilities of certain orderings of the Y_j [or really of the $X(q, a_j)$], and these correspond to the expectations of indicator functions that are not three times differentiable. So we need to approximate indicator functions by smooth functions h , with some control on the resulting error in the probabilities. There is a substantial literature that attempts to do this as efficiently as possible, but for us a simple approach will suffice.

Lemma 4.3 *Let $Y = (Y_j)_{1 \leq j \leq n}$ and $Z = (Z_j)_{1 \leq j \leq n}$ be as in Lemma 4.2. Let (S) be any subset of $\{1, \dots, n\} \times \{1, \dots, n\}$ not including any diagonal pairs (i, i) , and let S be the subset of \mathbb{R}^n defined by*

$$S := \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_i \geq x_j \ \forall (i, j) \in (S)\}.$$

Finally, let $\delta > 0$ be a small parameter.

Then

$$|\mathbb{P}(Y \in S) - \mathbb{P}(Z \in S)| \ll \frac{n^6}{\delta^3 \sqrt{\varphi(q)}} + n^2 \delta.$$

Proof of Lemma 4.3 Let $\phi : \mathbb{R} \rightarrow \mathbb{R}$ be a three times differentiable function such that

$$\phi(x) = \begin{cases} 1 & \text{if } x \geq \delta \\ \in [0, 1] & \text{if } 0 < x < \delta \\ 0 & \text{if } x \leq 0, \end{cases}$$

and $\phi^{(d)}(x) \ll (1/\delta)^d$ for $1 \leq d \leq 3$ (where $\phi^{(d)}$ denotes the d th derivative of ϕ). Note that such ϕ exists since the interval on which ϕ changes from 0 to 1 has length δ .

Now, let $h_\delta^+, h_\delta^- : \mathbb{R}^n \rightarrow \mathbb{R}$ be three times differentiable functions defined by

$$h_\delta^-(x_1, \dots, x_n) := \prod_{(i,j) \in (S)} \phi(x_i - x_j), \quad h_\delta^+(x_1, \dots, x_n) := \prod_{(i,j) \in (S)} \phi(x_i - x_j + \delta).$$

By repeated application of the product rule we see

$$\begin{aligned} \frac{\partial^2}{\partial x_a \partial x_b} h_\delta^-(x_1, \dots, x_n) &= \sum_{\substack{(i,j),(k,l) \in (S), \\ (i,j) \neq (k,l)}} \left(\prod_{\substack{(u,v) \in (S), \\ (u,v) \neq (i,j),(k,l)}} \phi(x_u - x_v) \right) \frac{\partial}{\partial x_a} \phi(x_i - x_j) \\ &\quad \times \frac{\partial}{\partial x_b} \phi(x_k - x_l) \\ &+ \sum_{(i,j) \in (S)} \left(\prod_{\substack{(u,v) \in (S), \\ (u,v) \neq (i,j)}} \phi(x_u - x_v) \right) \frac{\partial^2}{\partial x_a \partial x_b} \phi(x_i - x_j). \end{aligned}$$

Here each of the products has absolute value at most 1. The derivatives vanish unless $a \in \{i, j\}$ and $b \in \{k, l\}$, so there are $\ll n^2$ non-vanishing terms in the sums, and each term contributes $\ll 1/\delta^2$ because $\phi^{(d)}(x) \ll (1/\delta)^d$. We conclude from this calculation that $|h_\delta^-|_2 \ll n^2/\delta^2$. An exactly similar argument shows that $|h_\delta^-|_3 \ll n^3/\delta^3$, and the same for h_δ^+ .

Now observe that $\mathbb{E}h_\delta^-(Y) \leq \mathbb{P}(Y \in S) \leq \mathbb{E}h_\delta^+(Y)$. In addition, by Lemma 4.2 we have

$$|\mathbb{E}h_\delta^-(Y) - \mathbb{E}h_\delta^-(Z)| \ll \frac{n^6}{\delta^3 \sqrt{\varphi(q)}} \quad \text{and} \quad |\mathbb{E}h_\delta^+(Y) - \mathbb{E}h_\delta^+(Z)| \ll \frac{n^6}{\delta^3 \sqrt{\varphi(q)}},$$

so to prove Lemma 4.3 it will suffice to show that $|\mathbb{P}(Z \in S) - \mathbb{E}h_\delta^-(Z)| \ll n^2\delta$ and $|\mathbb{P}(Z \in S) - \mathbb{E}h_\delta^+(Z)| \ll n^2\delta$. But this is fairly easy, because we have for example that

$$\begin{aligned} |\mathbb{P}(Z \in S) - \mathbb{E}h_\delta^-(Z)| &\leq \mathbb{P}(|Z_i - Z_j| \leq \delta \text{ for some } (i, j) \in (S)) \\ &\leq \sum_{(i,j) \in (S)} \mathbb{P}(|Z_i - Z_j| \leq \delta). \end{aligned}$$

And each difference $Z_i - Z_j$ is a normal random variable with mean zero and variance

$$\mathbb{E}(Z_i - Z_j)^2 = \mathbb{E}Z_i^2 + \mathbb{E}Z_j^2 - 2\mathbb{E}Z_i Z_j = 2 - 2 \frac{B_q(a_i, a_j)}{\text{Var}(q)} = 2 + O(1/\log q),$$

in view of (2.6) and (2.7) (and our assumption that if $(i, j) \in (S)$ then $i \neq j$, and therefore $a_i \neq a_j$). Such a normal random variable has probability $O(\delta)$ of lying in

any δ -ball, and there are at most n^2 pairs in (S) , so we have the required bound $\ll n^2\delta$ for the right hand side. □

Finally, by making the optimal choice of δ in Lemma 4.3, and correcting for the small shifts $C_q(a_j)/\sqrt{\text{Var}(q)}$ in the definition of Y_j , we obtain the following result that we shall actually use.

Normal Approximation Result 1 *Let $X := \left(\frac{X(q,a_j)}{\sqrt{\text{Var}(q)}}\right)_{1 \leq j \leq n}$, and let $Z = (Z_j)_{1 \leq j \leq n}$ denote a multivariate normal random vector whose components have mean zero, variance one, and correlations $\mathbb{E}Z_i Z_j := \frac{B_q(a_i,a_j)}{\text{Var}(q)}$. Then for any set S as in Lemma 4.3, we have*

$$|\mathbb{P}(X \in S) - \mathbb{P}(Z \in S)| \ll \frac{n^3}{\varphi(q)^{1/8}}.$$

Proof of Normal Approximation Result 1 Choosing $\delta = n/\varphi(q)^{1/8}$ in Lemma 4.3 is optimal and leads to the bound

$$|\mathbb{P}(Y \in S) - \mathbb{P}(Z \in S)| \ll \frac{n^3}{\varphi(q)^{1/8}}.$$

Since we have $X = Y - \left(\frac{C_q(a_j)}{\sqrt{\text{Var}(q)}}\right)_{1 \leq j \leq n}$, and as described in Sect. 2 we always have $C_q(a_j)/\sqrt{\text{Var}(q)} \ll_\epsilon q^\epsilon/\sqrt{\text{Var}(q)} \ll 1/\varphi(q)^{1/2-\epsilon}$ (which is much smaller than $\delta = n/\varphi(q)^{1/8}$), then as well as the inequality $\mathbb{E}h_\delta^-(Y) \leq \mathbb{P}(Y \in S) \leq \mathbb{E}h_\delta^+(Y)$ in the proof of Lemma 4.3 we will actually have $\mathbb{E}h_\delta^-(Y) \leq \mathbb{P}(X \in S) \leq \mathbb{E}h_\delta^+(Y)$, provided the function ϕ there is chosen suitably [e.g. such that $\phi(x) = 0$ if $x \leq \delta/100$, and $\phi(x) = 1$ if $x \geq 99\delta/100$]. So we will have the same bound for $|\mathbb{P}(X \in S) - \mathbb{P}(Z \in S)|$. □

4.2 Normal comparison results

In this subsection we record some celebrated results that let one compare probabilities for multivariate Gaussians with different covariance matrices.

Normal Comparison Result 1 (Slepian’s Lemma, see e.g. Piterbarg [29]) *Let $X = (X_i)_{1 \leq i \leq n}$ and $W = (W_i)_{1 \leq i \leq n}$ be vectors of jointly normal, mean zero, variance one random variables. Suppose that $\mathbb{E}X_i X_j \leq \mathbb{E}W_i W_j$ for all pairs i, j . Then for any real numbers u_1, \dots, u_n we have*

$$\mathbb{P}(X_i \leq u_i \ \forall 1 \leq i \leq n) \leq \mathbb{P}(W_i \leq u_i \ \forall 1 \leq i \leq n).$$

Slepian’s Lemma says that decreasing the correlations of normal random variables makes them stochastically larger. It allows one to replace complicated correlations with simpler ones and it has the great advantage that it is never worse than trivial.

Using Slepian’s Lemma we can prove the following useful bound.

Normal Comparison Result 2 Suppose that $n \geq 2$ and that $\epsilon > 0$ is sufficiently small. Let X_1, \dots, X_n be mean zero, variance one, jointly normal random variables, and suppose that $\mathbb{E}X_i X_j \leq \epsilon$ whenever $i \neq j$. Then for any $A \geq 1$ and any $B > 0$ we have

$$\mathbb{P}\left(\max_{1 \leq i \leq n} X_i \leq A\right) \ll \exp\left\{-\Theta\left(n \frac{e^{-A^2/2+O(\epsilon A^2+AB+B^2)}}{A+B}\right)\right\} + e^{-B^2/\epsilon}.$$

In particular, for any $100\epsilon \leq \delta \leq 1/100$ (say) we have

$$\mathbb{P}\left(\max_{1 \leq i \leq n} X_i \leq \sqrt{(2-\delta)\log n}\right) \ll e^{-\Theta(n^{\delta/20}/\sqrt{\log n})} + n^{-\delta^2/50\epsilon}.$$

Proof of Normal Comparison Result 2 In view of Slepian's Lemma, the probability is at most as large as with the X_i replaced by W_i , where $\mathbb{E}W_i W_j = \epsilon$ whenever $i \neq j$. However, it is well known that one can explicitly construct random variables with this covariance structure, by letting Z_0, Z_1, \dots, Z_n be independent standard normal random variables, and then taking

$$W_i = \sqrt{\epsilon}Z_0 + \sqrt{1-\epsilon}Z_i \quad \forall 1 \leq i \leq n.$$

So conditioning on the value of Z_0 , we deduce that

$$\mathbb{P}\left(\max_{1 \leq i \leq n} X_i \leq A\right) \leq \mathbb{P}\left(\max_{1 \leq i \leq n} W_i \leq A\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(\Phi\left(\frac{A+\sqrt{\epsilon}y}{\sqrt{1-\epsilon}}\right)\right)^n e^{-y^2/2} dy,$$

where Φ denotes the standard normal cumulative distribution function.

Splitting the integral at $y = B\sqrt{2/\epsilon}$, we find it is

$$\leq \left(\Phi\left(\frac{A+B\sqrt{2}}{\sqrt{1-\epsilon}}\right)\right)^n + \int_{B\sqrt{2/\epsilon}}^{\infty} e^{-y^2/2} dy \ll \left(\Phi\left(\frac{A+B\sqrt{2}}{\sqrt{1-\epsilon}}\right)\right)^n + e^{-B^2/\epsilon}. \quad (4.1)$$

Moreover, for any $z \geq 1$ we have

$$\Phi(z) = 1 - \frac{1}{\sqrt{2\pi}} \int_z^{\infty} e^{-y^2/2} dy \leq 1 - \Theta\left(\frac{e^{-z^2/2}}{z}\right),$$

and so (since ϵ is small) we deduce

$$\left(\Phi\left(\frac{A+B\sqrt{2}}{\sqrt{1-\epsilon}}\right)\right)^n \leq \left(1 - \Theta\left(\frac{e^{-(1+2\epsilon)(A+B\sqrt{2})^2/2}}{A+B}\right)\right)^n, \quad (4.2)$$

from which the first bound claimed in Normal Comparison Result 2 follows.

The second bound follows on taking $A = \sqrt{(2 - \delta) \log n}$ and $B = \delta\sqrt{(\log n)/50}$ in (4.1) and (4.2), and noting then that

$$\begin{aligned}
 n \frac{e^{-(1+2\epsilon)(A+B\sqrt{2})^2/2}}{A+B} &\gg n \frac{e^{-(1+2\epsilon)A^2/2-2AB}}{A+B} \gg n \frac{e^{-(1+\delta/50)(1-\delta/2)\log n - (2/5)\delta \log n}}{\sqrt{\log n}} \\
 &\geq \frac{n^{\delta/20}}{\sqrt{\log n}}.
 \end{aligned}$$

□

The above results only give one sided bounds on probabilities, whereas in our theorems we want to show probabilities are equal up to a small error. The following result can supply such estimates in some cases.

Normal Comparison Result 3 (See Theorem 2.1 of Li and Shao [24]) *Let $X = (X_i)_{1 \leq i \leq n}$ and $W = (W_i)_{1 \leq i \leq n}$ be vectors of jointly normal, mean zero, variance one random variables. For each pair i, j define $\rho_{i,j} := \max\{|\mathbb{E}X_i X_j|, |\mathbb{E}W_i W_j|\}$. Then for any real numbers u_1, \dots, u_n we have*

$$\begin{aligned}
 &\mathbb{P}(X_i \leq u_i \ \forall 1 \leq i \leq n) - \mathbb{P}(W_i \leq u_i \ \forall 1 \leq i \leq n) \\
 &\leq \frac{1}{2\pi} \sum_{\substack{1 \leq i < j \leq n, \\ \mathbb{E}X_i X_j > \mathbb{E}W_i W_j}} (\arcsin(\mathbb{E}X_i X_j) - \arcsin(\mathbb{E}W_i W_j)) \exp \left\{ -\frac{(u_i^2 + u_j^2)}{2(1 + \rho_{i,j})} \right\}.
 \end{aligned}$$

Note that by swapping the roles of X and W one can obtain two sided bounds from this result.

A problem with Normal Comparison Result 3 is that the probabilities on the left may themselves be very small, and so the bound on the right may be worse than trivial. We know of no result that can overcome this difficulty in general, and a major issue in proving our theorems will be arranging things so that we only apply Normal Comparison Result 3 to probabilities that are fairly large, as discussed in the introduction.

4.3 A result on leaders

In this subsection we shall establish Theorem 1.5. To this end, we shall first prove the following.

Lemma 4.4 *Let the situation be as in Theorem 1.5. Then for any $x \geq 1$ we have*

$$\begin{aligned}
 |\mathbb{P} \left(\max_{2 \leq i \leq n} X_i \leq x | X_1 = x \right) - \Phi(x)^{n-1}| &\ll x e^{-(x^2/2)(1+O(\epsilon))} \sum_{2 \leq i \leq n} |r_{1,i}| \\
 &+ e^{-x^2(1+O(\epsilon))} \sum_{2 \leq i < j \leq n} |r_{i,j}|.
 \end{aligned}$$

Proof of Lemma 4.4 Consider the transformed random variables

$$V_i := \frac{X_i - r_{1,i}X_1}{\sqrt{1 - r_{1,i}^2}}, \quad 2 \leq i \leq n.$$

It is easy to check that these are all standard normal random variables, and they satisfy

$$\begin{aligned} \mathbb{E}V_iX_1 &= 0 \quad \forall 2 \leq i \leq n, \\ \mathbb{E}V_iV_j &= \frac{r_{i,j} - r_{1,i}r_{1,j}}{\sqrt{(1 - r_{1,i}^2)(1 - r_{1,j}^2)}} = O(|r_{i,j}| + |r_{1,i}||r_{1,j}|) = O(\epsilon) \quad \forall i \neq j. \end{aligned}$$

In particular, since the V_i are uncorrelated with X_1 they are independent from X_1 , and so we see

$$\begin{aligned} \mathbb{P}\left(\max_{2 \leq i \leq n} X_i \leq x | X_1 = x\right) &= \mathbb{P}\left(V_i \leq \frac{x(1 - r_{1,i})}{\sqrt{1 - r_{1,i}^2}} \quad \forall 2 \leq i \leq n\right) \\ &= \mathbb{P}\left(V_i \leq x \sqrt{\frac{1 - r_{1,i}}{1 + r_{1,i}}}\right). \end{aligned}$$

Next, using Normal Comparison Result 3 twice (with X and W swapped the second time) to compare the V_i with independent standard normals we obtain

$$\begin{aligned} &\mathbb{P}\left(V_i \leq x \sqrt{\frac{1 - r_{1,i}}{1 + r_{1,i}}}\right) - \prod_{2 \leq i \leq n} \Phi\left(x \sqrt{\frac{1 - r_{1,i}}{1 + r_{1,i}}}\right) \\ &= O\left(\sum_{2 \leq i < j \leq n} |\mathbb{E}V_iV_j| \exp\left(-x^2 \frac{((1 - r_{1,i})/(1 + r_{1,i})) + ((1 - r_{1,j})/(1 + r_{1,j}))}{2(1 + |\mathbb{E}V_iV_j|)}\right)\right) \\ &= O\left(e^{-x^2(1+O(\epsilon))} \sum_{2 \leq i < j \leq n} (|r_{1,i}||r_{1,j}| + |r_{i,j}|)\right). \end{aligned}$$

Notice here that the contribution from $|r_{i,j}|$ is acceptable for Lemma 4.4, and the contribution from $|r_{1,i}||r_{1,j}|$ may be rewritten as

$$\left(e^{-x^2/2(1+O(\epsilon))} \sum_{2 \leq i \leq n} |r_{1,i}|\right)^2.$$

If this term is smaller than one then it is acceptable because it is smaller than the first error term in Lemma 4.4, and if it is bigger than one then so is the first error term in Lemma 4.4, so the lemma is trivially true.

To finish it will suffice to prove that

$$\left| \prod_{2 \leq i \leq n} \Phi \left(x \sqrt{\frac{1-r_{1,i}}{1+r_{1,i}}} \right) - \Phi(x)^{n-1} \right| \ll x e^{-(x^2/2)(1+O(\epsilon))} \sum_{2 \leq i \leq n} |r_{1,i}|.$$

But this follows simply because we have

$$\begin{aligned} \Phi \left(x \sqrt{\frac{1-r_{1,i}}{1+r_{1,i}}} \right) - \Phi(x) &= \frac{1}{\sqrt{2\pi}} \int_x^{x\sqrt{(1-r_{1,i})/(1+r_{1,i})}} e^{-y^2/2} dy \\ &= O \left(x|r_{1,i}|e^{-(x^2/2)(1+O(\epsilon))} \right) \end{aligned}$$

for each $2 \leq i \leq n$, on noting that $\sqrt{\frac{1-r_{1,i}}{1+r_{1,i}}} = 1 + O(|r_{1,i}|) = 1 + O(\epsilon)$ for all such i . □

Proof of Theorem 1.5 Lemma 4.4 looks very close to Theorem 1.5, but it doesn't immediately yield the theorem because it is quite weak (possibly worse than trivial) unless x is fairly large. (As x becomes smaller we expect the probabilities on the left to become very small, whereas the bound on the right becomes larger.) Fortunately we can deal with the case of small x using Normal Comparison Result 2.

Indeed, if we choose $\delta = 1/1000$, say, then we clearly have

$$\begin{aligned} &\left| \mathbb{P} \left(X_1 > \max_{2 \leq i \leq n} X_i \right) - \mathbb{P} \left(X_1 > \max_{2 \leq i \leq n} X_i, \text{ and } X_1 > \sqrt{(2-\delta) \log n} \right) \right| \\ &\leq \mathbb{P} \left(\max_{2 \leq i \leq n} X_i \leq \sqrt{(2-\delta) \log n} \right), \end{aligned}$$

and by Normal Comparison Result 2 the right hand side is

$$\ll e^{-\Theta(n^{1/20000}/\sqrt{\log n})} + n^{-1/50000000\epsilon} \ll n^{-100},$$

since ϵ is assumed to be small enough. We would have the same bound if the X_i were replaced by independent normals \tilde{X}_i , and in that case we have $\mathbb{P}(X_1 > \max_{2 \leq i \leq n} \tilde{X}_i) = 1/n$ by symmetry, so to prove Theorem 1.5 it will suffice to show that

$$\begin{aligned} &\left| \mathbb{P} \left(X_1 > \max_{2 \leq i \leq n} X_i, \text{ and } X_1 > \sqrt{(2-\delta) \log n} \right) \right. \\ &\quad \left. - \mathbb{P} \left(\tilde{X}_1 > \max_{2 \leq i \leq n} \tilde{X}_i, \text{ and } \tilde{X}_1 > \sqrt{(2-\delta) \log n} \right) \right| \\ &\ll n^{-1.99} \sum_{2 \leq i \leq n} |r_{1,i}| + n^{-2.99} \sum_{2 \leq i < j \leq n} |r_{i,j}|. \end{aligned}$$

But using Lemma 4.4 we have

$$\begin{aligned}
 & \mathbb{P} \left(X_1 > \max_{2 \leq i \leq n} X_i, \text{ and } X_1 > \sqrt{(2-\delta) \log n} \right) \\
 & \quad - \mathbb{P} \left(\tilde{X}_1 > \max_{2 \leq i \leq n} \tilde{X}_i, \text{ and } \tilde{X}_1 > \sqrt{(2-\delta) \log n} \right) \\
 & = \int_{\sqrt{(2-\delta) \log n}}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \mathbb{P} \left(\max_{2 \leq i \leq n} X_i \leq x \mid X_1 = x \right) dx \\
 & \quad - \int_{\sqrt{(2-\delta) \log n}}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} \Phi(x)^{n-1} dx \\
 & = O \left(n^{-(2-\delta)(1+O(\epsilon))} \sum_{2 \leq i \leq n} |r_{1,i}| + n^{-(3/2)(2-\delta)(1+O(\epsilon))} \sum_{2 \leq i < j \leq n} |r_{i,j}| \right).
 \end{aligned}$$

The Theorem follows on remembering that $\delta = 1/1000$ and ϵ is sufficiently small. \square

4.4 Preparation for the k -contestant theorems

In this subsection we shall develop two lemmas that will be required later for the proof of Theorem 1.6. We present these in a moderate amount of generality, and they might be of wider interest.

We begin with a kind of hybrid normal comparison inequality.

Lemma 4.5 *Let $(W_i)_{i \in \mathcal{I}}$ be a finite collection of jointly standard normal random variables, let $\epsilon \geq \epsilon_1 > 0$ be small, and suppose that*

$$|\mathbb{E}W_i W_j| \leq \epsilon \quad \forall i \neq j, \quad \text{and} \quad \mathbb{E}W_i W_j \leq \epsilon_1 \quad \forall i \neq j.$$

Then if $w \geq 1$, and $w/2 \leq w_i \leq 2w$ for all $i \in \mathcal{I}$, we have

$$\begin{aligned}
 & \left| \mathbb{P}(W_i \leq w_i \quad \forall i \in \mathcal{I}) - \prod_{i \in \mathcal{I}} \Phi(w_i) \right| \\
 & \ll \left(\prod_{i \in \mathcal{I}} \Phi((1+O(\epsilon))w_i) + e^{-\Theta((\epsilon w)^2/(\epsilon_1 + \epsilon^3))} \right) \\
 & \quad \times \sum_{\substack{i, j \in \mathcal{I}, \\ i \neq j}} |\mathbb{E}W_i W_j| e^{-(1/2+O(\epsilon))(w_i^2 + w_j^2)}.
 \end{aligned}$$

The key advantage of Lemma 4.5, as opposed to a result like Normal Comparison Result 3, is the presence of the bracketed prefactor on the right hand side, which can make the inequality much more powerful if $\prod_{i \in \mathcal{I}} \Phi(w_i)$ is small. Notice also that if one

has a stronger upper bound for $\mathbb{E}W_i W_j$ than an absolute value bound (i.e. if ϵ_1 is appreciably smaller than ϵ), then the second term in the bracket is improved. This should not be too surprising, since Slepian’s Lemma (Normal Comparison Result 1) implies that negative correlations make normal random variables stochastically smaller.

Proof of Lemma 4.5 If one looks inside the proofs of the usual normal comparison lemmas (see e.g. Li and Shao’s paper [24]), they yield that

$$\begin{aligned} |\mathbb{P}(W_i \leq w_i \ \forall i \in \mathcal{I}) - \prod_{i \in \mathcal{I}} \Phi(w_i)| &\ll \sum_{\substack{i, j \in \mathcal{I}, \\ i \neq j}} P_{i, j} |\mathbb{E}W_i W_j| e^{-(w_i^2 + w_j^2)/2(1 + |\mathbb{E}W_i W_j|)} \\ &= \sum_{\substack{i, j \in \mathcal{I}, \\ i \neq j}} P_{i, j} |\mathbb{E}W_i W_j| e^{-(1/2 + O(\epsilon))(w_i^2 + w_j^2)}, \end{aligned}$$

where

$$P_{i, j} := \sup_{0 \leq h \leq 1} \mathbb{P}\left(W_k^{(h)} \leq w_k \ \forall k \in \mathcal{I} \mid W_i^{(h)} = w_i, W_j^{(h)} = w_j\right),$$

and the $W_k^{(h)}$ are mean zero, variance one, jointly normal random variables with correlations

$$\mathbb{E}W_k^{(h)} W_l^{(h)} = h \mathbb{E}W_k W_l, \quad k \neq l.$$

(Thus, in particular, when $h = 1$ the $W_k^{(h)}$ are simply the W_k , and when $h = 0$ the $W_k^{(h)}$ are independent standard normal random variables.) We will show that

$$\begin{aligned} \mathbb{P}(W_k \leq w_k \ \forall k \in \mathcal{I} \mid W_i = w_i, W_j = w_j) \\ \ll \prod_{i \in \mathcal{I}} \Phi((1 + O(\epsilon))w_i) + e^{-\Theta((\epsilon w)^2/(\epsilon_1 + \epsilon^3))} \end{aligned}$$

for any $i \neq j$. Exactly the same argument would yield the corresponding estimate for any $0 \leq h \leq 1$ (uniformly over h), thus proving Lemma 4.5.

Indeed, if we define

$$\begin{aligned} V_k := &W_k - \frac{(\mathbb{E}W_k W_i - \mathbb{E}W_k W_j \mathbb{E}W_i W_j)}{1 - (\mathbb{E}W_i W_j)^2} W_i \\ &- \frac{(\mathbb{E}W_k W_j - \mathbb{E}W_k W_i \mathbb{E}W_i W_j)}{1 - (\mathbb{E}W_i W_j)^2} W_j \quad \forall k \in \mathcal{I} \setminus \{i, j\}, \end{aligned}$$

then one can check that $\mathbb{E}V_k W_i = \mathbb{E}V_k W_j = 0$, and so if $k, l \in \mathcal{I} \setminus \{i, j\}$ then

$$\mathbb{E}V_k V_l = \mathbb{E}V_k W_l = \mathbb{E}V_k W_l = \mathbb{E}W_k W_l - \mathbb{E}W_k W_i \mathbb{E}W_i W_l - \mathbb{E}W_k W_j \mathbb{E}W_j W_l + O(\epsilon^3).$$

In particular, it follows that $\mathbb{E}V_k^2 = 1 + O(\epsilon^2)$. One can also check that when $k \neq l$ we have the upper bound

$$\mathbb{E}V_k V_l \leq \epsilon_1 + 2\epsilon\epsilon_1 + O(\epsilon^3) \leq (C/2)(\epsilon_1 + \epsilon^3),$$

for a suitable positive constant C , since $\mathbb{E}W_k W_l \leq \epsilon_1$ and if $-\mathbb{E}W_k W_l \mathbb{E}W_i W_l$ is positive then one of the factors must be positive and one negative, so one has size at most ϵ_1 and the other has size at most ϵ .

Now if we set $\tilde{V}_k := \frac{V_k}{\sqrt{\mathbb{E}V_k^2}}$, then $\mathbb{P}(W_k \leq w_k \forall k \in \mathcal{I} | W_i = w_i, W_j = w_j)$ is

$$\begin{aligned} &= \mathbb{P}\left(V_k \leq w_k - \frac{(\mathbb{E}W_k W_i - \mathbb{E}W_k W_j \mathbb{E}W_i W_j)}{1 - (\mathbb{E}W_i W_j)^2} w_i \right. \\ &\quad \left. - \frac{(\mathbb{E}W_k W_j - \mathbb{E}W_k W_i \mathbb{E}W_i W_j)}{1 - (\mathbb{E}W_i W_j)^2} w_j \forall k \in \mathcal{I} \setminus \{i, j\}\right) \\ &= \mathbb{P}(\tilde{V}_k \leq (1 + O(\epsilon))w_k \forall k \in \mathcal{I} \setminus \{i, j\}), \end{aligned}$$

where the first equality uses the fact that V_k is uncorrelated with, and therefore independent of, W_i and W_j (so we can remove the conditioning), and the second equality uses the fact that $w_i, w_j \asymp w \asymp w_k$.

Now the \tilde{V}_k are jointly standard normal random variables with off-diagonal correlations that are $\leq 1 + O(\epsilon^2)$. $\mathbb{E}V_k V_l \leq C(\epsilon_1 + \epsilon^3)$, so by Slepian's Lemma (Normal Comparison Result 1) the probability is at most as large as if all the off-diagonal correlations were equal to $C(\epsilon_1 + \epsilon^3)$. Arguing as in the proof of Normal Comparison Result 2, it follows that

$$\begin{aligned} &\mathbb{P}(\tilde{V}_k \leq (1 + O(\epsilon))w_k \forall k \in \mathcal{I} \setminus \{i, j\}) \\ &\leq \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \prod_{\substack{k \in \mathcal{I}, \\ k \neq i, j}} \Phi\left(\frac{(1 + O(\epsilon))w_k + \sqrt{C(\epsilon_1 + \epsilon^3)}y}{\sqrt{1 - C(\epsilon_1 + \epsilon^3)}}\right) e^{-y^2/2} dy, \end{aligned}$$

and Lemma 4.5 follows on splitting the integral at $y = \epsilon w / \sqrt{C(\epsilon_1 + \epsilon^3)}$. \square

We finish this section by applying Lemma 4.5 to prove the following general estimate for $\mathbb{P}(\max_{k+1 \leq i \leq n} X_i \leq x_k | X_1 = x_1, \dots, X_k = x_k)$, which is what we shall actually use later when proving Theorem 1.6.

Lemma 4.6 *Let X_1, \dots, X_n be mean zero, variance one, jointly normal random variables, and write $r_{i,j} := \mathbb{E}X_i X_j$. Let $\epsilon \geq \epsilon_1, \epsilon_2 > 0$ be sufficiently small, and suppose that*

$$|r_{i,j}| \leq \epsilon \forall i \neq j, \quad \text{and} \quad r_{i,j} \leq \epsilon_1 \forall i \neq j.$$

Let $1 \leq k \leq n - 2$ be such that ϵk is sufficiently small, and suppose further that

$$\sum_{l=1}^k \sum_{s=1, s \neq l}^k |r_{i,l}| |r_{j,s}| \left(|r_{l,s}| + \sum_{t=1, t \neq l, s}^k |r_{l,t}| |r_{s,t}| + \epsilon^2 \left(\sum_{\substack{1 \leq t, u \leq k, \\ t \neq u}} |r_{t,u}| \right) \right) \leq \epsilon_2$$

for any distinct $k + 1 \leq i, j \leq n$.

Then for any real numbers $x_1, \dots, x_{k-1} \geq x_k \geq 1$ such that $\sum_{i=1}^k x_i^2 \ll 1/\epsilon$, we have

$$\begin{aligned} & \left| \mathbb{P} \left(\max_{k+1 \leq i \leq n} X_i \leq x_k \mid X_1 = x_1, \dots, X_k = x_k \right) - \prod_{i=k+1}^n \Phi(w_i) \right| \\ & \ll e^{-(1+O(\sqrt{\epsilon k}))x_k^2} \left(\prod_{i=k+1}^n \Phi((1+O(\epsilon))w_i) + e^{-\Theta((\epsilon x_k)^2/(\epsilon_1 + \epsilon_2 + \epsilon^3))} \right) \\ & \times \left(\sum_{k+1 \leq i < j \leq n} |r_{i,j}| + \sum_{l=1}^k \left(\sum_{i=k+1}^n |r_{i,l}| \right)^2 \right), \end{aligned} \tag{4.3}$$

for certain numbers w_i that satisfy

$$w_i = (1 + O(\epsilon^2 k))x_k + O \left(\sum_{l=1}^k x_l |r_{i,l}| \right) = (1 + O(\sqrt{\epsilon k}))x_k.$$

The reader may wish to compare Lemma 4.6 with Lemma 4.4. The first bracketed term on the right hand side of Lemma 4.6 will be crucial when we come to apply the lemma to prime number races with k large, since on the relevant range of x_1, \dots, x_k it will turn out that $\prod_{i=k+1}^n \Phi(w_i)$ is rather small (in fact of size roughly e^{-Ck}).

Proof of Lemma 4.6 Note first that provided $\epsilon k \leq 1/2$, Lemma 3.3 implies that the covariance (sub-)matrix $\mathcal{A} := (r_{i,j})_{1 \leq i, j \leq k}$ is invertible. Let $(\tilde{r}_{i,j})_{1 \leq i, j \leq k} := \mathcal{A}^{-1}$ denote the inverse matrix, and for all $k + 1 \leq i \leq n$ and all $1 \leq l \leq k$ set

$$u_{i,l} := \sum_{s=1}^k \tilde{r}_{l,s} r_{i,s}, \quad \text{and} \quad V_i := X_i - \sum_{l=1}^k u_{i,l} X_l.$$

Then the random variables V_i are zero mean, jointly normal random variables, and for any $1 \leq t \leq k$ they satisfy

$$\mathbb{E} V_i X_t = r_{i,t} - \sum_{l=1}^k u_{i,l} r_{l,t} = r_{i,t} - \sum_{1 \leq l, s \leq k} r_{i,s} \tilde{r}_{l,s} r_{l,t} = r_{i,t} - \sum_{1 \leq s \leq k} r_{i,s} \mathbf{1}_{s=t} = 0.$$

In other words, the random variables V_i are uncorrelated with, and therefore independent from, all of X_1, X_2, \dots, X_k . Let us also note that for any $k+1 \leq i \leq n$ and any $1 \leq l \leq k$, using Lemma 3.3 to estimate the $\tilde{r}_{l,s}$ yields that

$$u_{i,l} = \left(1 + O \left(\epsilon \sum_{\substack{1 \leq t, u \leq k, \\ t \neq u}} |r_{t,u}| \right) \right) r_{i,l} + O \left(\sum_{\substack{s=1, \\ s \neq l}}^k |r_{i,s}| \left(|r_{l,s}| + \sum_{t \neq l, s} |r_{l,t}| |r_{s,t}| + \epsilon^2 \sum_{\substack{1 \leq t, u \leq k, \\ t \neq u}} |r_{t,u}| \right) \right), \quad (4.4)$$

and using our assumptions that the off-diagonal covariances are bounded by ϵ , and that ϵk is small, it follows in particular that

$$|u_{i,l}| \ll |r_{i,l}| + \epsilon \sum_{\substack{s=1, \\ s \neq l}}^k |r_{i,s}| \ll \epsilon. \quad (4.5)$$

We also see that

$$\mathbb{E}V_i^2 = \mathbb{E}V_i X_i - 0 = \mathbb{E}V_i X_i = 1 + O \left(\epsilon \sum_{l=1}^k |u_{i,l}| \right) = 1 + O(\epsilon^2 k). \quad (4.6)$$

In view of the above discussion, if we rewrite the event $\max_{k+1 \leq i \leq n} X_i \leq x_k$ in terms of the V_i so that we can remove the conditioning, we find

$$\begin{aligned} & \mathbb{P} \left(\max_{k+1 \leq i \leq n} X_i \leq x_k \mid X_1 = x_1, \dots, X_k = x_k \right) \\ &= \mathbb{P} \left(V_i \leq x_k - \sum_{l=1}^k x_l u_{i,l} \quad \forall k+1 \leq i \leq n \right) \\ &= \mathbb{P} (W_i \leq w_i \quad \forall k+1 \leq i \leq n) \end{aligned}$$

where $w_i := \frac{1}{\sqrt{\mathbb{E}V_i^2}} \left(x_k - \sum_{l=1}^k x_l u_{i,l} \right)$, and $W_i := \frac{V_i}{\sqrt{\mathbb{E}V_i^2}}$ are mean zero, variance one, jointly normal random variables. Proceeding to estimate the off-diagonal correlations, we note first that

$$\mathbb{E}W_i W_j = (1 + O(\epsilon^2 k)) \mathbb{E}V_i V_j = (1 + O(\epsilon^2 k)) \mathbb{E}X_i V_j,$$

and then from (4.4) we deduce that

$$\begin{aligned} \mathbb{E}W_iW_j &= (1 + O(\epsilon^2k))r_{i,j} - \left(1 + O\left(\epsilon^2k + \epsilon \sum_{\substack{1 \leq t, u \leq k, \\ t \neq u}} |r_{t,u}| \right) \right) \sum_{l=1}^k r_{i,l}r_{j,l} \\ &+ O\left(\sum_{l=1}^k \sum_{s=1, s \neq l}^k |r_{i,l}||r_{j,s}| \left(|r_{l,s}| + \sum_{t=1, t \neq l, s}^k |r_{l,t}||r_{s,t}| + \epsilon^2 \left(\sum_{\substack{1 \leq t, u \leq k, \\ t \neq u}} |r_{t,u}| \right) \right) \right). \end{aligned} \tag{4.7}$$

Since $|r_{i,j}| \leq \epsilon$ and $r_{i,j} \leq \epsilon_1$ when $i \neq j$, and since ϵk is small, it follows in particular that whenever $i \neq j$ we have $|\mathbb{E}W_iW_j| = O(\epsilon)$, and

$$\mathbb{E}W_iW_j \leq (1 + O(\epsilon^2k))\epsilon_1 + (1 + O(\epsilon^2k^2))k\epsilon\epsilon_1 + O(\epsilon_2) = O(\epsilon_1 + \epsilon_2).$$

Finally, using that $\sum_{l=1}^k x_l \leq \sqrt{k \sum_{l=1}^k x_l^2} \ll \sqrt{k/\epsilon}$ (which follows from the Cauchy–Schwarz inequality) together with (4.5) we find

$$\begin{aligned} w_i &= (1 + O(\epsilon^2k))x_k + O\left(\sum_{l=1}^k x_l|r_{i,l}| + \sqrt{\epsilon k} \sum_{s=1}^k |r_{i,s}| \right) \\ &= (1 + O(\epsilon^2k))x_k + O\left(\sum_{l=1}^k x_l|r_{i,l}| \right). \end{aligned}$$

Here the final equality follows because the x_l are all ≥ 1 whereas ϵk is small. We also have $\sum_{l=1}^k x_l|r_{i,l}| \leq \epsilon \sum_{l=1}^k x_l \ll \sqrt{\epsilon k}$, and therefore $w_i = (1 + O(\sqrt{\epsilon k}))x_k$, and hence $x_k/2 \leq w_i \leq 2x_k$ for all $k + 1 \leq i \leq n$.

This all means that Lemma 4.5 is applicable, with ϵ_1 replaced by $O(\epsilon_1 + \epsilon_2)$ and w replaced by x_k . Lemma 4.6 follows from Lemma 4.5 on noting that

$$(1/2 + O(\epsilon))(w_i^2 + w_j^2) = (1 + O(\sqrt{\epsilon k}))x_k^2,$$

and using (4.7) to get

$$\begin{aligned} \sum_{k+1 \leq i < j \leq n} |\mathbb{E}W_iW_j| &\ll \sum_{k+1 \leq i < j \leq n} \left(|r_{i,j}| + \sum_{l=1}^k |r_{i,l}||r_{j,l}| + \epsilon \sum_{l=1}^k \sum_{s=1, s \neq l}^k |r_{i,l}||r_{j,s}| \right) \\ &\ll \sum_{k+1 \leq i < j \leq n} |r_{i,j}| + \sum_{l=1}^k \left(\sum_{i=k+1}^n |r_{i,l}| \right)^2 + \epsilon \left(\sum_{l=1}^k \sum_{i=k+1}^n |r_{i,l}| \right)^2 \\ &\ll \sum_{k+1 \leq i < j \leq n} |r_{i,j}| + \sum_{l=1}^k \left(\sum_{i=k+1}^n |r_{i,l}| \right)^2. \end{aligned}$$

Here the final line uses the Cauchy–Schwarz inequality in the form $\left(\sum_{l=1}^k \sum_{i=k+1}^n |r_{i,l}|\right)^2 \leq k \sum_{l=1}^k \left(\sum_{i=k+1}^n |r_{i,l}|\right)^2$. \square

5 The full prime number race: Proof of Theorem 1.2

Suppose q is large, and let $Z = (Z_j)_{1 \leq j \leq n}$ denote a multivariate normal random vector whose components have mean zero, variance one, and correlations

$$r_{i,j} = \mathbb{E}Z_i Z_j = \frac{B_q(a_i, a_j)}{\text{Var}(q)}.$$

Let $S_1, S_2 \subseteq \{1, 2, \dots, n\}$ be non-empty. Then Correlation Estimate 1 implies that

$$\sum_{\substack{\ell \in S_1, s \in S_2 \\ \ell \neq s}} |r_{\ell,s}| \ll \frac{\sqrt{|S_1| \cdot |S_2|} \log^2(2|S_1| \cdot |S_2|)}{\log q}, \quad \text{and} \quad \sum_{\substack{1 \leq \ell, s \leq n \\ \ell \neq s}} |r_{\ell,s}| \ll \frac{n \log^2(2n)}{\log q}. \quad (5.1)$$

Let $\mathcal{C} = (r_{i,j})_{1 \leq i, j \leq n}$ be the covariance matrix of Z_1, \dots, Z_n . In view of the pointwise bound $|r_{i,j}| \ll 1/\log q$ (for $i \neq j$) from Lemma 2.2, we may apply Lemma 3.3 with $\epsilon \asymp 1/\log q$ provided that n is at most a certain constant times $\log q$, and obtain that

$$\det(\mathcal{C}) = 1 + O\left(\frac{n(\log n)^2}{(\log q)^2}\right), \quad (5.2)$$

and

$$\tilde{r}_{\ell,s} = \begin{cases} 1 + O\left(\frac{n(\log n)^2}{(\log q)^2}\right) & \text{if } \ell = s \\ O\left(|r_{\ell,s}| + \sum_{\substack{1 \leq j \leq n \\ j \neq \ell, s}} |r_{\ell,j}| |r_{s,j}| + \frac{n(\log n)^2}{(\log q)^3}\right) & \text{if } \ell \neq s, \end{cases} \quad (5.3)$$

where $\tilde{r}_{i,j}$ denote the entries of the inverse matrix \mathcal{C}^{-1} .

Now the key ingredient in the proof of Theorem 1.2 is the following proposition, which gives an approximation for the joint density function of Z_1, \dots, Z_n . Let $f(x_1, \dots, x_n)$ be this density, and define the Euclidean norm $\|x\| := (x_1^2 + \dots + x_n^2)^{1/2}$.

Proposition 5.1 *Let $2 \leq n \ll \log q$ be an integer. For any $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, we have*

$$f(x_1, \dots, x_n) = \left(1 + O\left(\frac{n(\log n)^2}{(\log q)^2}\right)\right) \times \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{\|x\|^2}{2} \left(1 + O\left(\frac{(\log n)^4}{\log q} + \frac{n(\log n)^6}{(\log q)^2} + \frac{n^2(\log n)^2}{(\log q)^3}\right)\right)\right).$$

Proof By definition and by (5.2) we have

$$\begin{aligned}
 f(x_1, \dots, x_n) &= \frac{1}{(2\pi)^{n/2} \sqrt{\det(C)}} \exp\left(-\frac{1}{2} \mathbf{x}^T C^{-1} \mathbf{x}\right) \\
 &= \left(1 + O\left(\frac{n(\log n)^2}{(\log q)^2}\right)\right) \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2} \mathbf{x}^T C^{-1} \mathbf{x}\right),
 \end{aligned}
 \tag{5.4}$$

and using (5.3) we obtain

$$\mathbf{x}^T C^{-1} \mathbf{x} = \sum_{1 \leq \ell, s \leq n} x_\ell x_s \tilde{r}_{\ell, s} = \|\mathbf{x}\|^2 \left(1 + O\left(\frac{n(\log n)^2}{(\log q)^2}\right)\right) + \sum_{1 \leq \ell \neq s \leq n} x_\ell x_s \tilde{r}_{\ell, s},
 \tag{5.5}$$

and also that

$$\begin{aligned}
 \sum_{1 \leq \ell \neq s \leq n} x_\ell x_s \tilde{r}_{\ell, s} &\ll \sum_{1 \leq \ell \neq s \leq n} |x_\ell x_s r_{\ell, s}| + \sum_{1 \leq \ell \neq s \leq n} |x_\ell x_s| \sum_{\substack{1 \leq j \leq n \\ j \neq \ell, s}} |r_{\ell, j}| |r_{s, j}| \\
 &\quad + \frac{n(\log n)^2}{(\log q)^3} \sum_{1 \leq \ell \neq s \leq n} |x_\ell x_s|.
 \end{aligned}
 \tag{5.6}$$

Now let $J := \lfloor 2 \log n \rfloor$, and for each $1 \leq j \leq J - 1$ define S_j to be the subset of $\{1, \dots, n\}$ consisting of those ℓ for which $\|\mathbf{x}\|/2^j < |x_\ell| \leq \|\mathbf{x}\|/2^{j-1}$. Also let S_J be the set of those ℓ for which $|x_\ell| \leq \|\mathbf{x}\|/2^{J-1}$. Then note that for all $1 \leq j \leq J$ we have $|S_j| \leq 2^{2j}$, since $|S_j| \leq n \leq 2^{2J}$ and if $j < J$, then $\frac{\|\mathbf{x}\|^2}{2^{2j}} |S_j| \leq x_1^2 + \dots + x_n^2 = \|\mathbf{x}\|^2$. Using (5.1) we deduce that

$$\begin{aligned}
 \sum_{1 \leq \ell \neq s \leq n} |x_\ell x_s r_{\ell, s}| &= \sum_{1 \leq i, j \leq J} \sum_{\substack{\ell \in S_i, s \in S_j \\ \ell \neq s}} |x_\ell x_s r_{\ell, s}| \ll \|\mathbf{x}\|^2 \sum_{1 \leq i, j \leq J} \frac{1}{2^{i+j}} \sum_{\substack{\ell \in S_i, s \in S_j \\ \ell \neq s}} |r_{\ell, s}| \\
 &\ll \frac{(\log n)^4}{\log q} \|\mathbf{x}\|^2.
 \end{aligned}$$

Similarly we have

$$\begin{aligned}
 \sum_{1 \leq \ell \neq s \leq n} |x_\ell x_s| \sum_{\substack{1 \leq j \leq n \\ j \neq \ell, s}} |r_{\ell, j}| |r_{s, j}| &\leq \sum_{1 \leq j \leq n} \left(\sum_{\substack{1 \leq \ell \leq n \\ \ell \neq j}} |x_\ell| |r_{\ell, j}| \right)^2 \\
 &\ll \sum_{1 \leq j \leq n} \left(\sum_{1 \leq i \leq J} \frac{\|\mathbf{x}\|}{2^i} \sum_{\substack{\ell \in S_i, \\ \ell \neq j}} |r_{\ell, j}| \right)^2 \\
 &\ll \frac{n(\log n)^6}{(\log q)^2} \|\mathbf{x}\|^2,
 \end{aligned}$$

since (5.1) implies that $\sum_{\substack{\ell \in S_i, \\ \ell \neq j}} |r_{\ell,j}| \ll (\sqrt{|S_i|} \log^2(2|S_i|)) / \log q \ll 2^i i^2 / \log q$.

Finally, by the Cauchy–Schwarz inequality we have $\sum_{1 \leq \ell \neq s \leq n} |x_\ell x_s| \leq n \|x\|^2$, and hence the contribution of the third term in the right hand side of (5.6) is $\ll n^2 (\log n)^2 \|x\|^2 / (\log q)^3$. Collecting the above estimates completes the proof. \square

Proof of Theorem 1.2 By Proposition 5.1, if $2 \leq n \leq \log q / (\log \log q)^4$ then

$$\begin{aligned} \mathbb{P}(Z_1 > Z_2 > \dots > Z_n) &= \int_{x_1 > \dots > x_n} f(x_1, \dots, x_n) dx_1 \dots dx_n \\ &= \left(1 + O\left(\frac{n(\log n)^2}{(\log q)^2}\right)\right) \frac{1}{(2\pi)^{n/2}} \int_{x_1 > \dots > x_n} \exp\left(-\frac{\|x\|^2}{2} \left(1 + O\left(\frac{(\log n)^4}{\log q}\right)\right)\right) dx_1 \dots dx_n \\ &= \left(1 + O\left(\frac{n(\log n)^2}{(\log q)^2}\right)\right) \frac{1}{n!} \left(\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{t^2}{2} \left(1 + O\left(\frac{(\log n)^4}{\log q}\right)\right)\right) dt\right)^n \\ &= \left(1 + O\left(\frac{n(\log n)^4}{\log q}\right)\right) \frac{1}{n!}, \end{aligned}$$

since the integrand is symmetric in x_1, \dots, x_n .

Combining this with Normal Approximation Result 1, we obtain

$$\begin{aligned} \mathbb{P}(X(q, a_1) > X(q, a_2) > \dots > X(q, a_n)) \\ &= \left(1 + O\left(\frac{n(\log n)^4}{\log q}\right)\right) \frac{1}{n!} + O\left(\frac{n^3}{\varphi(q)^{1/8}}\right). \end{aligned}$$

If $n \leq \log q / (\log \log q)^4$ then $1/n! \geq 1/n^n = 1/q^{o(1)}$, so the second ‘‘big Oh’’ term may be absorbed into the first, and Theorem 1.2 follows. \square

6 The leader: Proof of Theorem 1.4

We may assume that q is sufficiently large, otherwise the theorem is trivial. Then we may assume that $n \geq \log^{0.1} q$, say, otherwise the theorem follows from Theorem 1.2. Let $Z = (Z_j)_{1 \leq j \leq n}$ denote a multivariate normal random vector whose components have mean zero, variance one, and correlations $r_{i,j} = \mathbb{E}Z_i Z_j = B_q(a_i, a_j) / \text{Var}(q)$.

In view of Theorem 1.5 and Correlation Estimate 1, and our assumption that $n \geq \log^{0.1} q$,

$$\begin{aligned} \left| \mathbb{P}(Z_1 > \max_{2 \leq i \leq n} Z_i) - \frac{1}{n} \right| &\ll n^{-100} + n^{-1.99} \sum_{2 \leq i \leq n} |r_{1,i}| + n^{-2.99} \sum_{2 \leq i < j \leq n} |r_{i,j}| \\ &\ll n^{-100} + \frac{\sqrt{n} \log^2 n}{n^{1.99} \log q} + \frac{n \log^2 n}{n^{2.99} \log q} \\ &\ll \frac{\log^2 n}{n^{1.49} \log q} = \frac{1}{n} \cdot \frac{\log^2 n}{n^{0.49} \log q}. \end{aligned}$$

Theorem 1.4 follows by combining this estimate with Normal Approximation Result 1.

7 Ordering the first k contestants: Proof of Theorem 1.6

Throughout this section we let Z_1, \dots, Z_n denote mean zero, variance one, jointly normal random variables corresponding to a prime number race modulo q [i.e. with off-diagonal correlations $r_{i,j} = \mathbb{E}Z_i Z_j = B_q(a_i, a_j)/\text{Var}(q)$], where q is large.

Our key tool in proving Theorem 1.6 will be the version of Lemma 4.6 that arises from specializing to the prime number race situation. We record this now, together with an estimate for the number of large values of the $|r_{i,j}|$ that we shall need when deducing it.

Lemma 7.1 *For any $k + 1 \leq i \leq n$ and any $j \geq 1$ we have*

$$\# \left\{ 1 \leq l \leq k : |r_{i,l}| \geq \frac{1}{2^j \log q} \right\} \ll 2^{2j} j^4.$$

Proof Let \mathcal{S}_j denote the set of $1 \leq l \leq k$ for which $|r_{i,l}| \geq \frac{1}{2^j \log q}$. Using Correlation Estimate 1, we have

$$\sum_{l \in \mathcal{S}_j} |r_{i,l}| \ll \frac{\sqrt{\#\mathcal{S}_j} \log^2(2\#\mathcal{S}_j)}{\log q}.$$

On the other hand, by definition of \mathcal{S}_j the left hand side is $\geq \frac{\#\mathcal{S}_j}{2^j \log q}$, and so the lemma follows by rearranging. □

Lemma 7.2 *Let k be a positive integer such that $k/\log q$ is small enough, and suppose $n \geq k + 2$ is large. Let $1 \leq A \leq 2\sqrt{\log n}$ be real. If $x = (x_1, \dots, x_k)$ is such that $x_1, \dots, x_{k-1} \geq x_k \geq A$ and $\|x\| \leq 10\sqrt{\log q}$, then we have*

$$\begin{aligned} \mathbb{P} \left(\max_{k+1 \leq i \leq n} Z_i \leq x_k \mid Z_1 = x_1, \dots, Z_k = x_k \right) &= \prod_{i=k+1}^n \Phi(w_i) \\ &+ O \left(e^{-A^2 \left(1 + O(\sqrt{k/\log q}) \right)} \frac{n(\log n)^4}{\log q} \left(\prod_{i=k+1}^n \Phi((1 + O(1/\log q))w_i) \right. \right. \\ &\left. \left. + e^{-\Theta(A^2(\log q)/(\log \log q)^9)} \right) \right), \end{aligned}$$

for certain numbers w_i that satisfy

$$w_i = \left(1 + O \left(\frac{k}{(\log q)^2} \right) \right) x_k + O \left(\sum_{l=1}^k x_l |r_{i,l}| \right).$$

Proof We want to apply Lemma 4.6, and Lemma 2.2 shows that we may do so with $\epsilon \asymp 1/\log q$. We also have $B_q(a, b) \leq C \log q$ (when $a \neq b$) by (2.10), and therefore

we may take $\epsilon_1 \asymp 1/\varphi(q)$, which is very small. It is more difficult to determine a permissible value of ϵ_2 , but we may do so using Lemma 7.1 together with Correlation Estimate 1. Indeed, we can write $\sum_{l=1}^k = \sum_{a \leq 2 \log k} \sum_{l \in S_a}$, where $S_a := \{1 \leq l \leq k : \frac{1}{2^a \log q} < |r_{i,l}| \leq \frac{1}{2^{a-1} \log q}\}$, and write $\sum_{s=1}^k = \sum_{b \leq 2 \log k} \sum_{s \in T_b}$, where $T_b := \{1 \leq s \leq k : \frac{1}{2^b \log q} < |r_{j,s}| \leq \frac{1}{2^{b-1} \log q}\}$ (with suitable adjustments to the definitions of $S_1, S_{\lfloor 2 \log k \rfloor}, T_1, T_{\lfloor 2 \log k \rfloor}$ to ensure that all indices l, s are included). Then Lemma 7.1 implies that $\#S_a \ll 2^{2a} a^4$ and $\#T_b \ll 2^{2b} b^4$, and using Correlation Estimate 1 to bound all the sums of $|r_{l,s}|$ over $l \in S_a$ and $s \in T_b$, we obtain that

$$\sum_{l=1}^k \sum_{s=1, s \neq l}^k |r_{i,l}| |r_{j,s}| |r_{l,s}| \ll \log^2(2k) \frac{\log^4(2k)}{\log^2 q} \frac{\log^2(2k)}{\log q} = \frac{\log^8(2k)}{\log^3 q} \ll \frac{(\log \log q)^8}{\log^3 q}.$$

By dividing up the sum over t according to the size of $|r_{l,t}|$, one can similarly show that

$$\sum_{t=1, t \neq l, s}^k |r_{l,t}| |r_{s,t}| \ll \log(2k) \frac{\log^2(2k)}{\log q} \frac{\log^2(2k)}{\log q} \ll \frac{(\log \log q)^5}{\log^2 q}.$$

Putting these estimates together with the standard bounds $\sum_{l=1}^k |r_{i,l}| \ll \frac{\sqrt{k} \log^2(2k)}{\log q}$ and $\sum_{\substack{1 \leq l, s \leq k, \\ l \neq s}} |r_{l,s}| \ll \frac{k \log^2(2k)}{\log q}$, (coming from Correlation Estimate 1), one checks that it is permissible to take

$$\epsilon_2 \asymp \frac{(\log \log q)^9}{\log^3 q},$$

whereupon Lemma 4.6 implies that

$$\begin{aligned} & \mathbb{P} \left(\max_{k+1 \leq i \leq n} Z_i \leq x_k \mid Z_1 = x_1, \dots, Z_k = x_k \right) - \prod_{i=k+1}^n \Phi(w_i) \\ & \ll e^{-(1+O(\sqrt{k/\log q}))x_k^2} \left(\prod_{i=k+1}^n \Phi(\tilde{w}_i) + e^{-\Theta(A^2(\log q)/(\log \log q)^9)} \right) \\ & \quad \times \left(\sum_{k+1 \leq i < j \leq n} |r_{i,j}| + \sum_{l=1}^k \left(\sum_{i=k+1}^n |r_{i,l}| \right)^2 \right), \end{aligned}$$

where $\tilde{w}_i = (1 + O(1/\log q))w_i$.

Finally, the stated form of the result follows by using Correlation Estimate 1 again to show that the second bracket on the right hand side is

$$\ll \frac{n \log^2 n}{\log q} + \frac{nk \log^4 n}{\log^2 q} \ll \frac{n \log^4 n}{\log q}.$$

□

We have now collected all the necessary ingredients for the proof of Theorem 1.6.

Proof of Theorem 1.6 Let $c > 0$ be a small constant to be fixed later, and suppose first that $cn/\log n < k \leq n$. Then our condition $k \log^{10} k \leq (\log q)/\log n$ implies that $n \ll_c (\log q)/(\log \log q)^{10}$, in which case Theorem 1.2 is applicable and implies that

$$\begin{aligned} \delta_k(q; a_1, \dots, a_n) &= \frac{(n - k)!}{n!} \left(1 + O\left(\frac{n(\log n)^4}{\log q}\right) \right) \\ &= \frac{(n - k)!}{n!} \left(1 + O\left(\frac{k(\log k)^4 \log n}{\log q}\right) \right). \end{aligned}$$

This result is already acceptable for Theorem 1.6, so we may assume henceforth that we are in the other case where $2 \leq k \leq cn/\log n$. We may also assume throughout that $n \geq \log^{0.1} q$, because otherwise the result again follows from Theorem 1.2 regardless of the value of k . Since we assume that q is large, we may assume in particular that n is large.

First, it follows from Normal Approximation Result 1 together with the discussion in Sect. 2 that

$$|\delta_k(q; a_1, \dots, a_n) - \mathbb{P}(Z_1 > Z_2 > \dots > Z_k > \max_{k+1 \leq i \leq n} Z_i)| \ll \frac{n^3}{\varphi(q)^{1/8}}.$$

This error is completely negligible for Theorem 1.6, since we assume $k \log^{10} k \leq (\log q)/\log n$ and $(\log q)/\log n$ is large.

Let $1 \leq A \leq 2\sqrt{\log n}$ be a real parameter to be chosen later. Then we have

$$\begin{aligned} &\mathbb{P}\left(Z_1 > Z_2 > \dots > Z_k > \max_{k+1 \leq i \leq n} Z_i\right) \\ &= \mathbb{P}(Z_1 > \dots > Z_k > \max_{k+1 \leq i \leq n} Z_i, \text{ and } Z_k > A) + O\left(\mathbb{P}\left(\max_{k+1 \leq i \leq n} Z_i \leq A\right)\right), \end{aligned}$$

and using Normal Comparison Result 2 with $\epsilon \asymp 1/\log q$ we get

$$\mathbb{P}\left(\max_{k+1 \leq i \leq n} Z_i \leq A\right) \ll \exp\left\{-\Theta\left(n \frac{e^{-A^2/2 + O(A^2/\log q + AB + B^2)}}{A + B}\right)\right\} + e^{-\Theta(B^2 \log q)},$$

for any $B > 0$. Simply choosing $B = 1$, and using that $1 \leq A \leq 2\sqrt{\log n} \leq 2\sqrt{\log q}$ and that $k \log^{10} k \leq (\log q)/\log n$, we deduce

$$\mathbb{P}\left(\max_{k+1 \leq i \leq n} Z_i \leq A\right) \ll \exp\left\{-\Theta\left(n \frac{e^{-A^2/2 + O(A)}}{A}\right)\right\} + n^{-2k}. \tag{7.1}$$

Ultimately we will choose A so this whole thing is $\ll n^{-2k} \ll \frac{(n-k)!}{n!} \frac{1}{n \log^{0.1} q}$, which will be an acceptable error term.

Next, we have

$$\begin{aligned} & \mathbb{P}(Z_1 > Z_2 > \cdots > Z_k > \max_{k+1 \leq i \leq n} Z_i, \text{ and } Z_k > A) \\ &= \int_{x_1 > \cdots > x_k > A} f(x_1, \dots, x_k) \cdot \mathbb{P}\left(\max_{k+1 \leq i \leq n} Z_i < x_k \mid Z_1 = x_1, \dots, Z_k = x_k\right) \\ & \quad \times dx_1 \cdots dx_k \\ &= \int_{\substack{x_1 > \cdots > x_k > A, \\ \|x\| \leq 3\sqrt{k \log n}}} f(x_1, \dots, x_k) \cdot \mathbb{P}\left(\max_{k+1 \leq i \leq n} Z_i < x_k \mid Z_1 = x_1, \dots, Z_k = x_k\right) \\ & \quad \times dx_1 \cdots dx_k + O(n^{-2k}), \end{aligned} \quad (7.2)$$

where as before f denotes the joint density of Z_1, \dots, Z_k , and where the second equality follows because by Proposition 5.1 (and the fact that $k \ll \log q$) we always have $f(x_1, \dots, x_k) \ll \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{\|x\|^2}{4}\right)$. Moreover, we note that when $\|x\| \leq 3\sqrt{k \log n}$, Proposition 5.1 implies

$$f(x_1, \dots, x_k) = \left(1 + O\left(k \log^4 k \frac{\log n}{\log q}\right)\right) \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{\|x\|^2}{2}\right). \quad (7.3)$$

Now let us note that $\Phi(y \pm \epsilon) = (1 + O(\epsilon e^{-(y+O(\epsilon))^2/2}))\Phi(y)$, for any $y, \epsilon > 0$. In view of this, and the crude bound $\sum_{l=1}^k x_l |r_{i,l}| \ll (1/\log q)\sqrt{k} \sqrt{\sum_{l=1}^k x_l^2} \ll \sqrt{k/\log q}$, the main term $\prod_{i=k+1}^n \Phi(w_i)$ in our Lemma 7.2 estimate for $\mathbb{P}\left(\max_{k+1 \leq i \leq n} Z_i < x_k \mid Z_1 = x_1, \dots, Z_k = x_k\right)$ equals

$$\begin{aligned} & \prod_{i=k+1}^n \Phi\left((1 + O(1/\log q))x_k + O\left(\sum_{l=1}^k x_l |r_{i,l}|\right)\right) \\ &= \prod_{i=k+1}^n \Phi\left((1 + O(1/\log q))x_k\right) \\ & \quad \prod_{\substack{i=k+1, \\ \sum_{l=1}^k x_l |r_{i,l}| \geq \frac{x_k}{\log q}}} \left(1 + O\left(e^{-(1/2 + O(\sqrt{k/\log q}))x_k^2} \sum_{l=1}^k x_l |r_{i,l}|\right)\right) \\ &= \Phi\left((1 + O(1/\log q))x_k\right)^{n-k} \\ & \quad \times \exp\left\{O\left(e^{-\frac{A^2}{2}\left(1 + O(\sqrt{k/\log q})\right)} \sum_{\substack{i=k+1, \\ \sum_{l=1}^k x_l |r_{i,l}| \geq \frac{x_k}{\log q}}} \sum_{l=1}^k x_l |r_{i,l}|\right)\right\}. \end{aligned} \quad (7.4)$$

Now for any given $x_1 > x_2 > \dots > x_k > A$ satisfying $\|x\| \leq 3\sqrt{k \log n}$, and any non-empty subset $\mathcal{S} \subseteq \{k + 1, k + 2, \dots, n\}$, we can divide the sum $\sum_{l=1}^k x_l |r_{i,l}|$ into $O(\log(2k))$ pieces depending on the size of x_l (on dyadic ranges), and apply Correlation Estimate 1 (similarly as in the proof of Proposition 5.1) to deduce that

$$\sum_{i \in \mathcal{S}} \sum_{l=1}^k x_l |r_{i,l}| \ll \log(2k) \|x\| \sqrt{\#\mathcal{S}} \frac{\log^2(2k\#\mathcal{S})}{\log q} \ll \sqrt{\#\mathcal{S}} k \log n \frac{\log^3(2k\#\mathcal{S})}{\log q}.$$

But if $\mathcal{S} := \{k + 1 \leq i \leq n : \sum_{l=1}^k x_l |r_{i,l}| \geq \frac{x_k}{\log q}\}$ then the left hand side must also be $\geq \#\mathcal{S} \frac{x_k}{\log q} \gg \#\mathcal{S} \frac{1}{\log q}$, so we deduce that $\#\mathcal{S} \ll k(\log n) \log^6(2k \log n) \ll k(\log n)(\log \log q)^6$. Substituting back this size estimate for \mathcal{S} implies that

$$\sum_{\substack{i=k+1, \\ \sum_{l=1}^k x_l |r_{i,l}| \geq \frac{x_k}{\log q}}^n} \sum_{l=1}^k x_l |r_{i,l}| \ll \frac{k(\log n)(\log \log q)^6}{\log q}.$$

Collecting the above estimates shows that

$$\prod_{i=k+1}^n \Phi(w_i) = \Phi((1 + O(1/\log q))x_k)^{n-k} \times \exp \left\{ O \left(e^{-\frac{A^2}{2} (1 + O(\sqrt{k/\log q}))} \frac{k(\log n)(\log \log q)^6}{\log q} \right) \right\},$$

and clearly we have the same estimate for $\prod_{i=k+1}^n \Phi((1 + O(1/\log q))w_i)$, which appears in the error term of Lemma 7.2.

At this point we choose A such that

$$e^{0.51A^2} = \frac{n}{k \log n}. \tag{7.5}$$

Note that since we assume that $2 \leq k \leq cn/\log n$, this choice of A will satisfy $1 \leq A \leq 2\sqrt{\log n}$ provided c is fixed small enough. Notice also that if $n \geq \log^2 q$, then $k \log n \ll \log q \leq \sqrt{n}$ and so $A \gg \sqrt{\log n}$. Substituting our choice of A back into (7.1) yields

$$\mathbb{P} \left(\max_{k+1 \leq i \leq n} Z_i \leq A \right) \ll \exp \left\{ -\Theta \left(\frac{e^{0.01A^2 + O(A)}}{A} k \log n \right) \right\} + n^{-2k} \ll n^{-2k}, \tag{7.6}$$

provided c was chosen small so A is large enough. With this choice of A we also have

$$e^{-A^2 \left(1 + O(\sqrt{k/\log q})\right)} \frac{n(\log n)^4}{\log q} = e^{-A^2 \left(0.49 + O(\sqrt{k/\log q})\right)} \frac{k(\log n)^5}{\log q} \ll \frac{k(\log k)^4 \log n}{\log q},$$

and similarly (distinguishing cases according as $k \geq \sqrt{n} \geq \log^{0.05} q$ or not) that

$$e^{-\frac{A^2}{2} \left(1 + O(\sqrt{k/\log q})\right)} \frac{k(\log n)(\log \log q)^6}{\log q} \ll \frac{k(\log n)(\log k)^6}{\log q} \ll 1.$$

Substituting into Lemma 7.2, and using our previous computations of $\prod_{i=k+1}^n \Phi(w_i)$ and our assumption that $k \ll (\log q)/(\log \log q)^{10}$, yields that

$$\begin{aligned} & \mathbb{P} \left(\max_{k+1 \leq i \leq n} Z_i \leq x_k \mid Z_1 = x_1, \dots, Z_k = x_k \right) \\ &= \Phi \left(\left(1 + O\left(\frac{1}{\log q}\right)\right) x_k \right)^{n-k} \left(1 + O\left(\frac{k \log n \log^6 k}{\log q}\right) \right) + O(n^{-2k}). \end{aligned} \quad (7.7)$$

Now, in view of the estimates (7.3) and (7.7), the main term on the right hand side of (7.2) equals

$$\begin{aligned} & \int_{\substack{x_1 > \dots > x_k > A, \\ \|x\| \leq 3\sqrt{k \log n}}} \frac{1}{(2\pi)^{k/2}} e^{-\|x\|^2/2} \Phi \left(\left(1 + O(1/\log q)\right) x_k \right)^{n-k} dx_1 \dots dx_k \\ &= \int_{x_1 > \dots > x_k > A} \frac{1}{(2\pi)^{k/2}} e^{-\|x\|^2/2} \Phi \left(\left(1 + O(1/\log q)\right) x_k \right)^{n-k} dx_1 \dots dx_k + O(n^{-2k}). \end{aligned}$$

Thus Theorem 1.6 will certainly follow if we show that

$$\begin{aligned} & \int_{x_1 > \dots > x_k > A} \frac{1}{(2\pi)^{k/2}} e^{-\|x\|^2/2} \Phi \left(\left(1 + O(1/\log q)\right) x_k \right)^{n-k} dx_1 \dots dx_k \\ &= \left(1 + O\left(\frac{k \log n}{\log q}\right) \right) \frac{(n-k)!}{n!}. \end{aligned} \quad (7.8)$$

We shall only prove an upper bound, since an exactly similar argument would give a matching lower bound. Indeed, for a certain absolute constant $C > 0$ the integral is

$$\begin{aligned} & \leq \int_{x_1 > \dots > x_k > A} \frac{1}{(2\pi)^{k/2}} e^{-\|x\|^2/2} \Phi \left(\left(1 + C/\log q\right) x_k \right)^{n-k} dx_1 \dots dx_k \\ &= \int_{x_k > A} \frac{e^{-x_k^2/2}}{\sqrt{2\pi}} \Phi \left(\left(1 + C/\log q\right) x_k \right)^{n-k} \frac{1}{(k-1)!} \left(\int_{x > x_k} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx \right)^{k-1} dx_k, \end{aligned}$$

by symmetry of the integration variables x_1, \dots, x_{k-1} . Making a substitution shows this is

$$\begin{aligned} &= \left(1 + O\left(\frac{1}{\log q}\right)\right) \int_{x_k > \left(1 + \frac{C}{\log q}\right)A} \frac{e^{-(1+O(1/\log q))x_k^2/2}}{\sqrt{2\pi}} \Phi(x_k)^{n-k} \\ &\quad \times \frac{1}{(k-1)!} \left(\int_{x > (1+C/\log q)^{-1}x_k} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx\right)^{k-1} dx_k \\ &= \left(1 + O\left(\frac{1}{\log q}\right)\right) \int_{x_k > \left(1 + \frac{C}{\log q}\right)A} \frac{e^{-x_k^2/2}}{\sqrt{2\pi}} \Phi(x_k)^{n-k} \cdot \frac{e^{O(kx_k^2/\log q)}}{(k-1)!} \\ &\quad \times \left(\int_{x > x_k} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx\right)^{k-1} dx_k, \end{aligned}$$

and here the term $e^{O(kx_k^2/\log q)}$ is $e^{O(k \log n/\log q)} = 1 + O\left(\frac{k \log n}{\log q}\right)$ provided $x_k \leq 3\sqrt{\log n}$, say. Moreover (and as seen before), the contribution to the integral from the complementary range $x_k > 3\sqrt{\log n}$ is $\ll n^{-2k}$, since in this case we have $\|x\| \geq 3\sqrt{k \log n}$. Hence, our integral equals

$$\begin{aligned} &\left(1 + O\left(\frac{k \log n}{\log q}\right)\right) \int_{x_k > \left(1 + \frac{C}{\log q}\right)A} \frac{e^{-x_k^2/2}}{\sqrt{2\pi}} \Phi(x_k)^{n-k} \cdot \frac{1}{(k-1)!} \\ &\quad \times \left(\int_{x > x_k} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx\right)^{k-1} dx_k + O(n^{-2k}) \\ &= \left(1 + O\left(\frac{k \log n}{\log q}\right)\right) \int_{x_1 > \dots > x_k > \left(1 + \frac{C}{\log q}\right)A} \frac{1}{(2\pi)^{k/2}} e^{-\|x\|^2/2} \Phi(x_k)^{n-k} dx_1 \dots dx_k \\ &\quad + O(n^{-2k}) \\ &= \left(1 + O\left(\frac{k \log n}{\log q}\right)\right) \cdot \\ &\quad \mathbb{P}\left(\tilde{Z}_1 > \tilde{Z}_2 > \dots > \tilde{Z}_k > \max_{k+1 \leq i \leq n} \tilde{Z}_i, \text{ and } \tilde{Z}_k > \left(1 + \frac{C}{\log q}\right)A\right) + O(n^{-2k}) \end{aligned}$$

where the \tilde{Z}_i are independent standard normal random variables. Furthermore, the same argument leading to (7.6) shows that

$$\mathbb{P}\left(\max_{k+1 \leq i \leq n} \tilde{Z}_i < \left(1 + \frac{C}{\log q}\right)A\right) \ll n^{-2k}.$$

Finally, the asymptotic (7.8) follows from combining the above estimates, and using that the probability $\mathbb{P}(\tilde{Z}_1 > \tilde{Z}_2 > \dots > \tilde{Z}_k > \max_{k+1 \leq i \leq n} \tilde{Z}_i)$ equals $(n-k)!/n!$ by symmetry of the random variables $\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_n$. \square

8 Irregularities in the densities: Proof of Theorem 1.7

As in the previous section, given distinct reduced residues $a_1, \dots, a_n \pmod q$ we let $Z = (Z_j)_{1 \leq j \leq n}$ be a multivariate normal random vector whose components have mean zero, variance one, and correlations $r_{i,j} = \mathbb{E}Z_i Z_j := \frac{B_q(a_i, a_j)}{\text{Var}(q)}$. Also, we let $2 \leq k \leq n$ be a fixed positive integer and let $f(x_1, \dots, x_k)$ denote the density function of Z_1, \dots, Z_k . We denote by $C = (r_{i,j})_{1 \leq i, j \leq k}$ the covariance matrix of Z_1, \dots, Z_k (which is certainly invertible provided q is large enough in terms of k), and let $\tilde{r}_{i,j}$ denote the entries of the inverse matrix C^{-1} .

The key ingredient in the proof of Theorem 1.7 is the following result.

Lemma 8.1 *Let $k \geq 2$ be fixed, and let q be large enough in terms of k . There exist distinct reduced residues $a_1, \dots, a_k \pmod q$ such that*

$$\begin{aligned} f(x_1, x_2, \dots, x_k) &= \left(1 + O_k\left(\frac{1}{(\log q)^2}\right)\right) \frac{1}{(2\pi)^{k/2}} \\ &\quad \times \exp\left(-\frac{\|x\|^2}{2} - x_1 x_2 \frac{\log 2 + o(1)}{\log q} + O_k\left(\frac{\|x\|^2}{(\log q)^2}\right)\right), \end{aligned}$$

where the $o(1)$ term tends to zero as $q \rightarrow \infty$.

Proof Let $p_1 < p_2$ be the smallest prime numbers such that $(p_1 p_2, q) = 1$. Then one has $p_1 < p_2 \leq 2 \log q$ in view of the fact that $\prod_{p \leq z} p = e^{z(1+o(1))}$, which follows from the prime number theorem. Mimicking a construction used in Theorem 2 of Lamzouri [23], we let $a_1 = 1, a_2 = -1$ and $a_j = (p_1 p_2)^j$ for $3 \leq j \leq k$. Then by Lemma 2.2 and Eq. (2.6), it follows that

$$r_{1,2} = r_{2,1} = -\frac{(\log 2) + o(1)}{\log q}, \quad (8.1)$$

and

$$r_{i,j} \ll \frac{(2 \log q)^{2k+1}}{\varphi(q)}, \quad (8.2)$$

for all $i \neq j$ such that $\{i, j\} \neq \{1, 2\}$. Now, by (5.4) we have

$$f(x_1, \dots, x_k) = \left(1 + O_k\left(\frac{1}{(\log q)^2}\right)\right) \frac{1}{(2\pi)^{k/2}} \exp\left(-\frac{1}{2} \mathbf{x}^T C^{-1} \mathbf{x}\right).$$

Recall that $\tilde{r}_{i,i} = 1 + O_k(1/(\log q)^2)$ by (5.3). Furthermore, by (5.3) and (8.2) we have

$$\tilde{r}_{i,j} \ll_k \frac{1}{(\log q)^2},$$

for all $i \neq j$ such that $\{i, j\} \neq \{1, 2\}$. This implies

$$\begin{aligned} & f(x_1, \dots, x_k) \\ &= \left(1 + O_k\left(\frac{1}{(\log q)^2}\right)\right) \frac{1}{(2\pi)^{k/2}} \\ &\quad \times \exp\left(-\frac{\|x\|^2}{2} - x_1x_2\frac{\tilde{r}_{2,1} + \tilde{r}_{1,2}}{2} + O_k\left(\frac{\|x\|^2}{(\log q)^2}\right)\right). \end{aligned} \tag{8.3}$$

Now, let \mathcal{A} be the matrix obtained from \mathcal{C} by removing the first row and the second column. Then arguing as in the proof of Lemma 3.3, we obtain

$$\begin{aligned} \tilde{r}_{2,1} &= -\frac{\det(\mathcal{A})}{\det(\mathcal{C})} = -\left(1 + O_k\left(\frac{1}{(\log q)^2}\right)\right) \det(\mathcal{A}) \\ &= -\left(1 + O_k\left(\frac{1}{(\log q)^2}\right)\right) \left(r_{2,1} + O_k\left(\frac{(2 \log q)^{2k+1}}{\varphi(q)}\right)\right). \end{aligned}$$

Thus by (8.1) we obtain $\tilde{r}_{2,1} = \frac{\log 2 + o(1)}{\log q}$, and a similar estimate holds for $\tilde{r}_{1,2}$. Inserting these estimates in (8.3) completes the proof. \square

Proof of Theorem 1.7 By the same argument that was used in the Introduction to deduce Theorem 1.3 from Theorem 1.7, it will suffice to prove Theorem 1.7 in the case $k = 2$, since this implies the result for larger k . (Using Lemma 8.1 one could in fact prove the result directly for any fixed k , but this would be more complicated and require an unwanted additional assumption of the shape $n < \varphi(q)^{c/k}$ at the end.)

Let $a_1 = 1$ and $a_2 = -1$ be as in Lemma 8.1, and let a_3, \dots, a_n be distinct reduced residues modulo q that are different from a_1, a_2 . As in the proof of Theorem 1.6, Normal Approximation Result 1 implies that

$$\left| \delta_2(q; a_1, \dots, a_n) - \mathbb{P}\left(Z_1 > Z_2 > \max_{3 \leq i \leq n} Z_i\right) \right| \ll \frac{n^3}{\varphi(q)^{1/8}}.$$

Next, we have

$$\begin{aligned} & \mathbb{P}\left(Z_1 > Z_2 > \max_{3 \leq i \leq n} Z_i\right) \\ &= \mathbb{P}\left(Z_1 > Z_2 > \max_{3 \leq i \leq n} Z_i, \text{ and } Z_2 > \sqrt{1.99 \log n}\right) \\ &\quad + O\left(\mathbb{P}\left(\max_{3 \leq i \leq n} Z_i \leq \sqrt{1.99 \log n}\right)\right), \end{aligned}$$

and using Normal Comparison Result 2 we obtain

$$\mathbb{P}\left(\max_{3 \leq i \leq n} Z_i \leq \sqrt{1.99 \log n}\right) \ll e^{-n^{c_1}} + n^{-c_2 \log q} \ll n^{-4}, \tag{8.4}$$

for some positive constants c_1, c_2 . Moreover, similarly to (7.2) we derive

$$\begin{aligned} & \mathbb{P}(Z_1 > Z_2 > \max_{3 \leq i \leq n} Z_i, \text{ and } Z_2 > \sqrt{1.99 \log n}) \\ &= \int_{\substack{x_1 > x_2 > \sqrt{1.99 \log n} \\ \|x\| < 3\sqrt{2 \log n}}} f(x_1, x_2) \cdot \mathbb{P}\left(\max_{3 \leq i \leq n} Z_i < x_2 \mid Z_1 = x_1, Z_2 = x_2\right) dx_1 dx_2 \\ &+ O\left(n^{-4}\right). \end{aligned} \quad (8.5)$$

Next, it follows from Lemma 7.2 that for all $x = (x_1, x_2)$ such that $x_1 > x_2 > \sqrt{1.99 \log n}$ and $\|x\| < 3\sqrt{2 \log n}$ we have

$$\begin{aligned} & \mathbb{P}\left(\max_{3 \leq i \leq n} Z_i \leq x_2 \mid Z_1 = x_1, Z_2 = x_2\right) = \prod_{i=3}^n \Phi(w_i) \\ &+ O\left(\frac{1}{n^{9/10} \log q} \left(\prod_{i=3}^n \Phi\left((1 + O(1/\log q))w_i\right) + n^{-4}\right)\right), \end{aligned} \quad (8.6)$$

where $w_i = (1 + O(1/\log^2 q))x_2 + O(\|x\| \sum_{s=1}^2 |r_{i,s}|)$. Then, similarly to (7.4) one gets

$$\begin{aligned} \prod_{i=3}^n \Phi(w_i) &= \prod_{i=3}^n \Phi\left((1 + O(1/\log^2 q))x_2\right) \\ &\times \left(1 + O\left(\|x\| \cdot e^{-(1/2 + O(\sqrt{1/\log q}))x_2^2} \sum_{i=3}^n \sum_{s=1}^2 |r_{i,s}|\right)\right) \\ &= \Phi\left((1 + O(1/\log^2 q))x_2\right)^{n-2} \left(1 + O\left(\frac{1}{n^{2/5} \log q}\right)\right), \end{aligned}$$

by Correlation Estimate 1 and the fact that $x_2 > \sqrt{1.99 \log n}$. Moreover, in the same range for the x_i we deduce from Lemma 8.1 that

$$\begin{aligned} f(x_1, x_2) &= \left(1 + O\left(\frac{\log n}{(\log q)^2}\right)\right) \frac{1}{(2\pi)^{2/2}} \exp\left(-\frac{\|x\|^2}{2} - x_1 x_2 \frac{\log 2 + o(1)}{\log q}\right) \\ &\leq \left(1 - \frac{c \log n}{\log q}\right) \frac{1}{2\pi} e^{-\|x\|^2/2} \end{aligned}$$

for some positive constant $c > 0$, provided q is large enough.

Combining the above estimates, and using our assumption that $n \geq \varphi(q)^\epsilon$, we deduce that the main term in (8.5) equals

$$\int_{\substack{x_1 > x_2 > \sqrt{1.99 \log n} \\ \|x\| < 3\sqrt{2 \log n}}} f(x_1, x_2) \cdot \prod_{i=3}^n \Phi(w_i) dx_1 dx_2$$

$$\leq (1 - c_\epsilon) \int_{\substack{x_1 > x_2 > \sqrt{1.99 \log n} \\ \|x\| < 3\sqrt{2 \log n}}} \frac{1}{2\pi} e^{-\|x\|^2/2} \Phi\left((1 + O(1/\log^2 q))x_2\right)^{n-2} dx_1 dx_2,$$

for some positive constant c_ϵ . Furthermore, using an exactly similar argument as in the proof of (7.8), together with (8.4), we derive

$$\int_{\substack{x_1 > x_2 > \sqrt{1.99 \log n} \\ \|x\| < 3\sqrt{2 \log n}}} \frac{1}{2\pi} e^{-\|x\|^2/2} \Phi\left((1 + O(1/\log^2 q))x_2\right)^{n-2} dx_1 dx_2$$

$$= \left(1 + O\left(\frac{\log n}{\log^2 q}\right)\right) \frac{(n-2)!}{n!} = \left(1 + O\left(\frac{1}{\log q}\right)\right) \frac{(n-2)!}{n!}.$$

Finally, the total contribution of the various error terms [notably the one from Normal Approximation Result 1 and the one from (8.6)] is

$$\ll \frac{n^3}{\varphi(q)^{1/8}} + n^{-4} + \frac{1}{n^{0.9} \log q} \frac{(n-2)!}{n!}.$$

Recalling our assumption that $\varphi(q)^\epsilon \leq n < \varphi(q)^{1/41}$, we see the error is negligible compared with $\frac{(n-2)!}{n!}$, which completes the proof. □

Acknowledgements We would like to thank Daniel Fiorilli for a helpful clarification of the discussion of one of our results, and the referee for their comments and corrections.

Appendix A: Sketch proof of Lemma 4.1

In this appendix we very briefly indicate how to deduce Lemma 4.1 from Theorem 2.1 of Reinert and Röllin [30].

Indeed, the exchangeable pair construction and calculations used to deduce Central Limit Theorem 1, in an appendix of the preprint [10], transfer directly to this situation and imply that

$$|\mathbb{E}h(W) - \mathbb{E}h(Z)| \ll |h|_2 \sum_{a,b \in \mathcal{A}} \sqrt{\sum_{i=1}^m |c_i(a)|^2 |c_i(b)|^2 \mathbb{E}|V_i|^4}$$

$$+ |h|_3 \sum_{i=1}^m \mathbb{E}|V_i|^3 \left(\sum_{a \in \mathcal{A}} |c_i(a)|\right)^3.$$

In Lemma 4.1 we assume the uniform fourth moment bound $\mathbb{E}|V_i|^4 \leq K^4/m^2$, and so the first term is

$$\ll \frac{|h|_2 K^2}{m} \sum_{a,b \in \mathcal{A}} \sqrt{\sum_{i=1}^m |c_i(a)|^2 |c_i(b)|^2},$$

which is acceptable for Lemma 4.1. We also note that

$$\mathbb{E}|V_i|^3 \leq \frac{K^3}{m^{3/2}} + \mathbb{E}|V_i|^3 \mathbf{1}_{|V_i| > K/\sqrt{m}} \leq \frac{K^3}{m^{3/2}} + \frac{\sqrt{m}}{K} \mathbb{E}|V_i|^4 \leq 2 \frac{K^3}{m^{3/2}},$$

and therefore the second term above is

$$\ll \frac{|h|_3 K^3}{m^{3/2}} \sum_{i=1}^m \left(\sum_{a \in \mathcal{A}} |c_i(a)| \right)^3,$$

as required for Lemma 4.1. □

References

1. Bourgain, J.: On $\Lambda(p)$ -subsets of squares. *Israel J. Math.* **67**(3), 291–311 (1989)
2. Davenport, H.: *Multiplicative Number Theory*. Revised and with a Preface by Hugh L. Montgomery. Graduate Texts in Mathematics, vol. 74, xiv+177 pp. Springer, New York (2000)
3. Feuerverger, A., Martin, G.: Biases in the Shanks–Rényi prime number race. *Experiment. Math.* **9**(4), 535–570 (2000)
4. Fiorilli, D.: Highly biased prime number races. *Algebra Number Theory* **8**(7), 1733–1767 (2014)
5. Fiorilli, D., Martin, G.: Inequities in the Shanks–Rényi prime number race: an asymptotic formula for the densities. *J. Reine Angew. Math.* **676**, 121–212 (2013)
6. Ford, K., Konyagin, S.: The prime number race and zeros of L -functions off the critical line. *Duke Math. J.* **113**(2), 313–330 (2002)
7. Ford, K., Konyagin, S.: Chebyshev’s Conjecture and the Prime Number Race. *Modern Problems of Number Theory and its Applications, Topical Problems Part II* (Tula, Russia 2001)
8. Ford, K., Konyagin, S., Lamzouri, Y.: The prime number race and zeros of Dirichlet L -functions off the critical line: part III. *Q. J. Math.* **64**(4), 1091–1098 (2013)
9. Granville, A., Martin, G.: Prime number races. *Am. Math. Mon.* **113**(1), 1–33 (2006)
10. Harper, A.J.: A note on the maximum of the Riemann zeta function, and log-correlated random variables. [arXiv:1304.0677](https://arxiv.org/abs/1304.0677)
11. Kaczorowski, J.: A contribution to the Shanks–Rényi race problem. *Quart. J. Math. Oxford Ser. (2)* **44**(176), 451–458 (1993)
12. Kaczorowski, J.: Results on the distribution of primes. *J. Reine Angew. Math.* **446**, 89–113 (1994)
13. Kaczorowski, J.: On the Shanks–Rényi race problem. *Acta Arith.* **74**(1), 31–46 (1996)
14. Knapowski, S., Turán, P.: Comparative prime-number theory. I. *Acta Math. Acad. Sci. Hung.* **13**, 299–314 (1962a)
15. Knapowski, S., Turán, P.: Comparative prime-number theory. II. *Acta Math. Acad. Sci. Hung.* **13**, 315–342 (1962b)
16. Knapowski, S., Turán, P.: Comparative prime-number theory. III. *Acta Math. Acad. Sci. Hung.* **13**, 343–364 (1962c)
17. Knapowski, S., Turán, P.: Comparative prime-number theory. IV. *Acta Math. Acad. Sci. Hung.* **14**, 31–42 (1963a)
18. Knapowski, S., Turán, P.: Comparative prime-number theory. V. *Acta Math. Acad. Sci. Hung.* **14**, 43–63 (1963b)
19. Knapowski, S., Turán, P.: Comparative prime-number theory. VI. *Acta Math. Acad. Sci. Hung.* **14**, 65–78 (1963c)

20. Knapowski, S., Turán, P.: Comparative prime-number theory. VII. *Acta Math. Acad. Sci. Hung.* **14**, 241–250 (1963d)
21. Knapowski, S., Turán, P.: Comparative prime-number theory. VIII. *Acta Math. Acad. Sci. Hung.* **14**, 251–268 (1963e)
22. Lamzouri, Y.: The Shanks–Rényi prime number race with many contestants. *Math. Res. Lett.* **19**(03), 649–666 (2012)
23. Lamzouri, Y.: Prime number races with three or more competitors. *Math. Ann.* **356**(3), 1117–1162 (2013)
24. Li, W., Shao, Q.-M.: A normal comparison inequality and its applications. *Probab. Theory Relat. Fields* **122**, 494–508 (2002)
25. Littlewood, J.E.: Distribution des nombres premiers. *C. R. Acad. Sci. Paris* **158**, 1869–1872 (1914)
26. Martin, G., Scarfy, J.: Comparative prime number theory: a survey. 37 pages. [arXiv:1202.3408](https://arxiv.org/abs/1202.3408)
27. Montgomery, H.L.: *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*. Published for the Conference Board of the Mathematical Sciences by the American Mathematical Society (1994)
28. Montgomery, H.L., Vaughan, R.C.: *Multiplicative Number Theory I: Classical Theory*, 1st edn. Cambridge University Press, Cambridge (2007)
29. Piterbarg, V.: *Asymptotic Methods in the Theory of Gaussian Processes and Fields*. Volume 148 of the American Mathematical Society Translations of Mathematical Monographs (1996)
30. Reinert, G., Röllin, A.: Multivariate normal approximation with Stein’s method of exchangeable pairs under a general linearity condition. *Ann. Probab.* **37**(6), 2150–2173 (2009)
31. Rubinstein, M., Sarnak, P.: Chebyshev’s bias. *Experiment. Math.* **3**(3), 173–197 (1994)