

EXTREME BIASES IN PRIME NUMBER RACES WITH MANY CONTESTANTS

KEVIN FORD, ADAM J HARPER, AND YOUNESS LAMZOURI

ABSTRACT. We continue to investigate the race between prime numbers in many residue classes modulo q , assuming the standard conjectures GRH and LI.

We show that provided $n/\log q \rightarrow \infty$ as $q \rightarrow \infty$, we can find n competitor classes modulo q so that the corresponding n -way prime number race is extremely biased. This improves on the previous range $n \geq \varphi(q)^\epsilon$, and (together with an existing result of Harper and Lamzouri) establishes that the transition from all n -way races being asymptotically unbiased, to biased races existing, occurs when $n = \log^{1+o(1)} q$.

The proofs involve finding biases in certain auxiliary races that are easier to analyse than a full n -way race. An important ingredient is a quantitative, moderate deviation, multi-dimensional Gaussian approximation theorem, which we prove using a Lindeberg type method.

1. INTRODUCTION

Let $q \geq 3$ and $2 \leq n \leq \varphi(q)$ be integers, (where the Euler function $\varphi(q)$ denotes the number of residue classes mod q that are coprime to q), and let $\mathcal{A}_n(q)$ be the set of ordered n -tuples (a_1, a_2, \dots, a_n) of distinct residue classes that are coprime to q . In this paper we are interested in the “Shanks–Rényi prime number race”, which is the following problem: if we let $\pi(x; q, a)$ denote the number of primes $p \leq x$ with $p \equiv a \pmod{q}$, is it true that for any $(a_1, a_2, \dots, a_n) \in \mathcal{A}_n(q)$, we will have the ordering

$$(1.1) \quad \pi(x; q, a_1) > \pi(x; q, a_2) > \dots > \pi(x; q, a_n)$$

for infinitely many integers x ? There is now an extensive body of work investigating different aspects of this question, and the reader may consult the expository papers of Granville and Martin [6], Ford and Konyagin [5], and Martin and Scarfy [10] for fuller discussions.

Date: November 24, 2017.

Kevin Ford was supported by National Science Foundation grant DMS-1501982. Adam Harper was supported, for part of this research, by a research fellowship at Jesus College, Cambridge. Youness Lamzouri is partially supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada. In addition, part of this work was carried out at MSRI, Berkeley during the Spring semester of 2017, supported in part by NSF grant DMS 1440140.

Assuming the Generalized Riemann hypothesis GRH and the Linear Independence hypothesis LI (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 [12] proved that the answer to this question is always Yes. More strongly, they proved that for any $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$, the set of real numbers $x \geq 2$ such that (1.1) holds has a positive logarithmic density, which we shall denote by $\delta(q; a_1, \dots, a_n)$. Recall here that the logarithmic density of a subset $S \subseteq \mathbb{R}$ is defined as

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

provided the 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. As we shall discuss, a probabilistic perspective turns out to be very helpful in this problem.

Next it is natural to ask whether all orderings of the $\pi(x; q, a_i)$ occur with approximately the same logarithmic density, in other words whether $\delta(q; a_1, \dots, a_n) \approx 1/n!$. For small q , the widely known phenomenon of *Chebyshev’s bias* implies that $\delta(q; a_1, a_2)$ can be significantly different from $1/2$, if one of the a_i is a quadratic residue and the other a non-residue mod q . For example, Theorem 1.11 of Fiorilli and Martin [4] implies that $\delta(24; 5, 1) \approx 0.99999$, assuming GRH and LI. On the other hand, Rubinstein and Sarnak [12] showed (assuming GRH and LI) that for any fixed n ,

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

in other words any biases dissolve when $q \rightarrow \infty$. Different behaviour is possible if one races “teams” of many residue classes combined against one another (e.g. all the quadratic residues mod q against all the non-residues mod q), as explored in Fiorilli’s paper [3]. Feuerverger and Martin [2] raised the question of having a uniform version of Rubinstein and Sarnak’s statement, in which $n \rightarrow \infty$ as $q \rightarrow \infty$. And they 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. Ford and Lamzouri (unpublished) formulated the following conjecture.

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

- (1) (*Uniformity for small n*) *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) (*Biases for large n*) *If $(\log q)^{1+\varepsilon} \leq n \leq \varphi(q)$, then there exist n -tuples $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$ and $(b_1, \dots, b_n) \in \mathcal{A}_n(q)$ 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 $q \rightarrow \infty$.*

The first part of this conjecture is now known to hold (assuming GRH and LI) in a slightly stronger form, as Harper and Lamzouri [7] proved that, uniformly for all $2 \leq n \leq \log q / (\log \log q)^4$ and all n -tuples $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$, we have

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

This improved on an earlier result of Lamzouri [8], where the asymptotic $\delta(q; a_1, \dots, a_n) \sim 1/n!$ was established in the range $n = o(\sqrt{\log q})$, assuming GRH and LI.

Regarding the second part of the conjecture, Harper and Lamzouri [7] proved (assuming GRH and LI) that for any $\varepsilon > 0$ and every $\varphi(q)^\varepsilon \leq n \leq \varphi(q)$, there exists an n -tuple $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$ such that

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

where $c_\varepsilon > 0$ depends only on ε . This was the first result on n -way prime number races where the biases do not dissolve when $q \rightarrow \infty$, but it is clearly far from the full statement in part (2) of Ford and Lamzouri's conjecture. In particular, we note that the bias $1 - c_\varepsilon$ is always less than 1 (whereas the conjecture asserts that biases towards both small and large values should be possible), and always close to 1 (whereas multipliers that tend to zero or to infinity with q should be possible). Our goal in this paper is to revisit this issue.

We shall prove the following result.

Theorem 1.2. *Assume GRH and LI. There exists a large absolute constant C such that the following is true. Provided n is sufficiently large and $n \leq \varphi(q)$, there exist distinct reduced residues a_1, \dots, a_n modulo q such that*

$$\delta(q; a_1, \dots, a_n) \leq \exp \left(- \frac{\min\{n, \varphi(q)^{1/50}\}}{C \log q} \right) \frac{1}{n!},$$

and there exist distinct reduced residues b_1, \dots, b_n modulo q such that

$$\delta(q; b_1, \dots, b_n) \geq \exp \left(\frac{\min\{n, \varphi(q)^{1/50}\}}{C \log q} \right) \frac{1}{n!}.$$

Notice this fully establishes part (2) of Ford and Lamzouri's conjecture, as soon as $n/\log q \rightarrow \infty$. Furthermore, as n becomes larger the relative biases become quantitatively very extreme, and for $n \leq \log q$ the bias $\exp \left(\pm \frac{n}{C \log q} \right) = 1 \pm \Theta \left(\frac{n}{\log q} \right)$ roughly matches the factor $\left(1 + O \left(\frac{n(\log n)^4}{\log q} \right) \right)$ in Harper and Lamzouri's [7] uniformity result. (For large fixed n , one can also think of this as clarifying the dependence on n in Theorem A of Lamzouri [9].)

We shall actually obtain Theorem 1.2 as a straightforward corollary of another ordering result. For any integers $1 \leq k \leq n \leq \varphi(q)$, let us 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

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

If everything were uniform, we would expect that $\delta_k(q; a_1, \dots, a_n) \approx (n-k)!/n!$. Let us also define $\delta_{2k}^\#(q; a_1, \dots, a_n)$ to be the logarithmic density of the set of real numbers $x \geq 2$ such that

$$\begin{aligned} \pi(x; q, a_1) > \pi(x; q, a_3) > \cdots > \pi(x; q, a_{2k-1}) > \max_{2k+1 \leq j \leq n} \pi(x; q, a_j) > \min_{2k+1 \leq j \leq n} \pi(x; q, a_j) > \\ > \pi(x; q, a_{2k}) > \cdots > \pi(x; q, a_4) > \pi(x; q, a_2). \end{aligned}$$

Again, if everything were uniform we would expect that $\delta_{2k}^\#(q; a_1, \dots, a_n) \approx (n-2k)!/n!$.

Harper and Lamzouri [7] proved the uniformity result that, assuming GRH and LI,

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

whenever $k(\log k)^{10} \leq (\log q)/\log n$. They also proved a non-uniformity result: for any fixed¹ $k \geq 2$, fixed $\varepsilon > 0$, and any $\varphi(q)^\varepsilon \leq n < \varphi(q)^{1/41}$, there exists an n -tuple $(a_1, \dots, a_n) \in \mathcal{A}_n(q)$ such that

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

Here we establish a significantly improved non-uniformity result.

Theorem 1.3. *Assume GRH and LI. There exists a large absolute constant A such that the following is true. Suppose q is large, and let $1 \leq k \leq n/A \leq \varphi(q)^{1/50}$. Then there exist distinct reduced residues a_1, \dots, a_n modulo q such that*

$$(1.3) \quad \delta_{2k}(q; a_1, \dots, a_n) \leq \exp \left(-\frac{k \log(n/k)}{2 \log q} \right) \frac{(n-2k)!}{n!},$$

and

$$(1.4) \quad \delta_{2k}^\#(q; a_1, \dots, a_n) \geq \exp \left(\frac{k \log(n/k)}{2 \log q} \right) \frac{(n-2k)!}{n!}.$$

¹We remark that the case $k=1$ is not excluded from the non-uniformity result for technical reasons, but because it is genuinely different. Harper and Lamzouri [7] showed that $\delta_1(q; a_1, \dots, a_n) \sim 1/n$ provided $n = o(\varphi(q)^{1/32})$, when $q \rightarrow \infty$. Notice that Theorem 1.3, below, deals with $\delta_{2k}(q; a_1, \dots, a_n)$ and $\delta_{2k}^\#(q; a_1, \dots, a_n)$, so the case of $\delta_1(q; a_1, \dots, a_n)$ is again excluded.

Proof of Theorem 1.2, assuming Theorem 1.3. Suppose first that $n \leq \varphi(q)^{1/50}$. Take $k = \lceil n/2A \rceil$, (which indeed satisfies $1 \leq k \leq n/A$ since we assume in Theorem 1.2 that n is large), and note that by Theorem 1.3 we have

$$\delta_{2k}(q; a_1, \dots, a_n) \leq \exp\left(-\frac{n \log(2A)}{4A \log q}\right) \frac{(n-2k)!}{n!}.$$

On the other hand, 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 equation (2.1) below), we get

$$\delta_{2k}(q; a_1, \dots, a_n) = \sum_{\sigma \in S_{n-2k}} \delta(q; a_1, a_2, \dots, a_{2k}, a_{\sigma(2k+1)}, \dots, a_{\sigma(n)}),$$

where we think of the symmetric group S_{n-2k} as the group of bijections of the set $\{2k+1, 2k+2, \dots, n\}$ to itself. Thus, by averaging, there exists $\sigma \in S_{n-2k}$ for which

$$\delta(q; a_1, a_2, \dots, a_{2k}, a_{\sigma(2k+1)}, \dots, a_{\sigma(n)}) \leq \exp\left(-\frac{n \log(2A)}{4A \log q}\right) \frac{1}{n!},$$

which gives the first part of Theorem 1.2 on taking $C = 4A/\log(2A)$.

Similarly for the second part of the theorem, taking $k = \lceil n/2A \rceil$ we have

$$\begin{aligned} & \sum_{\sigma \in S_{n-2k}} \delta(q; a_1, a_3, \dots, a_{2k-1}, a_{\sigma(2k+1)}, \dots, a_{\sigma(n)}, a_{2k}, \dots, a_4, a_2) = \delta_{2k}^{\#}(q; a_1, \dots, a_n) \\ & \geq \exp\left(\frac{n \log(2A)}{4A \log q}\right) \frac{(n-2k)!}{n!}, \end{aligned}$$

so by averaging there exists σ with $\delta(q; a_1, a_3, \dots, a_{2k-1}, a_{\sigma(2k+1)}, \dots, a_{\sigma(n)}, a_{2k}, \dots, a_4, a_2) \geq \exp\left(\frac{n \log(2A)}{4A \log q}\right) \frac{1}{n!}$.

Finally, if $\varphi(q)^{1/50} < n \leq \varphi(q)$ then set $m := \lfloor \varphi(q)^{1/50} \rfloor$ and assume that q is so large that $m \geq 2$, hence $m \geq \frac{1}{2}\phi(q)^{1/50}$. As in the previous discussion there exists an m -tuple $(a_1, \dots, a_m) \in \mathcal{A}_m(q)$ for which

$$\delta(q; a_1, \dots, a_m) \leq \exp\left(-\frac{\varphi(q)^{1/50}}{C \log q}\right) \frac{1}{m!},$$

and there exists an m -tuple $(b_1, \dots, b_m) \in \mathcal{A}_m(q)$ for which

$$\delta(q; b_1, \dots, b_m) \geq \exp\left(\frac{\varphi(q)^{1/50}}{C \log q}\right) \frac{1}{m!},$$

with $C = 8A/\log(2A)$. Then if we choose any other coprime residues $a_{m+1}, \dots, a_n \pmod q$, we have

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

There are $n!/m!$ terms in the sum, so it follows that for at least one permutation σ we must have $\delta(q; a_{\sigma(1)}, a_{\sigma(2)}, \dots, a_{\sigma(n)}) \leq \exp\left(-\frac{\varphi(q)^{1/50}}{C \log q}\right) \frac{1}{n!}$, as desired. The analogous lower bound with the b_i is proved exactly similarly. \square

We conclude this introduction by discussing some of the ideas from our proofs. As we shall recall in section 2, under GRH and LI the logarithmic density $\delta(q; a_1, \dots, a_n)$ is the same as an ordering probability for certain random variables $X(q, a_1), \dots, X(q, a_n)$. The ultimate source of the biases in our theorems is the fact that some of these random variables are not independent of one another, but have correlations of size $\xi \sim -\frac{\log 2}{\log q}$. Notice this is a completely different source of bias than the influence of being a quadratic residue or non-residue, which leads to Chebyshev's bias for small q because the *mean values* of the corresponding $X(q, a_i)$ are slightly unequal.

As we mentioned, Harper and Lamzouri [7] proved a non-uniformity result for the auxiliary quantities $\delta_k(q; a_1, \dots, a_n)$, and then deduced their non-uniformity result for $\delta(q; a_1, \dots, a_n)$ by averaging. The advantage of looking at $\delta_k(q; a_1, \dots, a_n)$ is that in the corresponding ordering probability, namely

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

the maximum of the $X(q, a_j)$ (when they are normalised by their standard deviations) is close to $\sqrt{2 \log n}$ with very high probability. This means that $X(q, a_1), \dots, X(q, a_k)$ are all larger than $\sqrt{2 \log n}$ with very high probability, so if we can arrange to have negative correlations $\sim -\frac{\log 2}{\log q}$ between $\asymp k$ of those random variables, we obtain a bias $\asymp -\frac{1}{\log q} k (\sqrt{2 \log n})^2 \asymp -\frac{k \log n}{\log q}$ in the exponent in the probability distribution function on that event.

Here we follow the same broad strategy, with two main changes. Firstly, in order to produce biases towards both large and small probabilities (unlike in [7] which only handled the small case) we work with the auxiliary density $\delta_{2k}^\#(q; a_1, \dots, a_n)$, in addition to $\delta_{2k}(q; a_1, \dots, a_n)$. In the ordering probability corresponding to $\delta_{2k}^\#(q; a_1, \dots, a_n)$, we have $X(q, a_1), X(q, a_3), \dots, X(q, a_{2k-1}) > \sqrt{2 \log n}$ and $X(q, a_2), X(q, a_4), \dots, X(q, a_{2k}) < -\sqrt{2 \log n}$ with very high probability (roughly speaking, when everything is normalised). So if we can arrange to have negative correlations $\sim -\frac{\log 2}{\log q}$ between pairs of the odd and even indexed $X(q, a_i)$, we obtain a positive bias $\asymp -\frac{1}{\log q} k \sqrt{2 \log n} (-\sqrt{2 \log n}) \asymp \frac{k \log n}{\log q}$ in the probability distribution function on that event. Secondly, whereas Harper and Lamzouri [7] worked with fixed k , here we allow k to be as large as a small constant times n . Thus, whereas Harper and Lamzouri obtain biases in the probability distribution of size at most $\asymp \frac{\log n}{\log q}$, we can obtain large biases of size $\asymp \frac{n}{\log q}$.

To implement the above strategy with k large, there are two technical issues that must be overcome. Firstly, if we have arranged some correlations of size ξ to produce a bias, we need to ensure that any other correlations do not interfere with that bias. Harper and Lamzouri [7] did this using a “large sieve” kind of average estimate for correlations, but here we take a different approach by working with specially chosen residues a_1, \dots, a_n where the only correlations, except those of size ξ , are extremely small. This is perfectly acceptable if one only wants to exhibit some residues producing large biases, and makes the argument simpler (we don’t need any sophisticated tools from Gaussian process theory) and possible to execute for very large k .

Secondly, in the above discussion and in our proofs we want to treat the $X(q, a_i)$ as jointly Gaussian random variables. As we shall discuss in section 2, in fact the $X(q, a_i)$ are a sum of independent random variables, and have variance $\asymp \varphi(q) \log q$, so we expect standard Berry–Esseen type ideas to produce a multivariate Gaussian approximation with an error term saving a polynomial in q , and with a polynomial dependence on the dimension n . Now comparing with Theorem 1.2, we are dealing with probabilities of size $1/n! = e^{-n \log n + O(n)}$ with n as large as a power of q , which is much smaller than the possible Berry–Esseen saving. To get around this, we need a multivariate Gaussian approximation with a *relative* error term rather than an absolute one, which can therefore be useful even for very improbable events. This kind of multivariate “moderate deviation” result does exist in the probabilistic literature, but we were unable to locate a result that applied to our situation without quite a lot of reworking. (See Theorem 1 of Bentkus [1] for a sample of what is available— Bentkus obtains a moderate deviation Gaussian approximation for the norm of a sum of independent, identically distributed random elements in a Hilbert space.) Thus we prove our own bespoke result (see Proposition 3.4 and Proposition 3.5, below), using a special construction of smooth test functions and an inductive “replacement” argument.

1.1. Notation. We adopt familiar order of magnitude notation of Bachmann–Landau, Vinogradov and Knuth. The notations $f = O(g)$, $f \ll g$ and $g \gg f$ mean that there exists a positive constant C such that $|f| \leq Cg$ throughout the range of f (either implicitly understood or explicitly given). The notations $f \asymp g$, $f = \Theta(g)$ mean that both $f \ll g$ and $f \gg g$ hold. We have $f(x) \sim g(x)$ as $x \rightarrow a$ if $\lim_{x \rightarrow a} f(x)/g(x) = 1$, where a may be finite, ∞ or $-\infty$.

We use \mathbb{P} and \mathbb{E} to denote probability and probabilistic expectation, respectively. The underlying probability spaces will be described explicitly or understood from context.

2. PRELIMINARIES ON PRIME NUMBER RACES, AND CORRESPONDING RANDOM VARIABLES

In this section, we review the connection between quantities like $\delta(q; a_1, \dots, a_n)$ and ordering probabilities for suitable random variables. See section 2 of Harper and Lamzouri [7] for a similar but more detailed review of this material. We also show the existence of special sets \mathcal{A} of residues mod q , for which the correlation structure of the corresponding random variables has nice properties that we can exploit.

Given distinct reduced residues a_1, \dots, a_n modulo q , we define

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

where

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

and $\pi(x)$ denotes the total number of primes less than x . It turns out that the normalization is such that, if we assume GRH, $E_{q;a_1,\dots,a_n}(x)$ varies roughly boundedly as x varies. Notice also that

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

Let us introduce the temporary notation

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

The work of Rubinstein and Sarnak [12] shows, under GRH and LI, that for any Lebesgue measurable set $S \subset \mathbb{R}^n$ whose boundary has measure zero, the logarithmic density $\delta_{q;a_1,\dots,a_n}(S)$ exists.

Furthermore, Rubinstein and Sarnak provide a probabilistic characterisation of $\delta_{q;a_1,\dots,a_n}(S)$. For a non-principal Dirichlet character χ modulo q , denote by $\{\gamma_\chi\}$ the sequence of imaginary parts of the nontrivial zeros of $L(s, \chi)$. If we assume LI then all of the non-negative values of γ_χ are linearly independent over \mathbb{Q} , and in particular are distinct. Let χ_0 denote the principal character modulo q and define $\Gamma = \bigcup_{\chi \neq \chi_0 \pmod q} \{\gamma_\chi\}$. Furthermore, let $\{U(\gamma_\chi)\}_{\gamma_\chi \in \Gamma, \gamma_\chi > 0}$ be a sequence of independent random variables uniformly distributed on the unit circle. Then we have

$$(2.1) \quad \delta_{q;a_1,\dots,a_n}(S) = \int_S d\mu_{q;a_1,\dots,a_n},$$

where $\mu_{q;a_1,\dots,a_n}$ is a certain probability measure on \mathbb{R}^n (which is absolutely continuous when $n < \varphi(q)$). Specifically, $\mu_{q;a_1,\dots,a_n}$ is the probability measure corresponding to the \mathbb{R}^n -valued

random vector $(X(q, a_1), \dots, X(q, a_n))$, where

$$(2.2) \quad X(q, a) := -C_q(a) + \sum_{\substack{\chi \neq \chi_0 \\ \chi \pmod{q}}} \operatorname{Re} \left(2\chi(a) \sum_{\gamma_\chi > 0} \frac{U(\gamma_\chi)}{\sqrt{\frac{1}{4} + \gamma_\chi^2}} \right),$$

with

$$(2.3) \quad C_q(a) := -1 + |\{b \pmod{q} : b^2 \equiv a \pmod{q}\}|.$$

It will turn out that the deterministic shifts $C_q(a)$ are sufficiently small that they can essentially be ignored when $q \rightarrow \infty$.

Let $\operatorname{Cov}_{q; a_1, \dots, a_n}$ be the covariance matrix of $(X(q, a_1), \dots, X(q, a_n))$. We need to understand this in order to understand the probabilities and dependencies of events involving the $X(q, a_i)$. Assuming GRH, one may compute (see section 2 of Harper and Lamzouri [7] for fuller details and references) 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

$$(2.4) \quad \operatorname{Var}(q) := 2 \sum_{\substack{\chi \neq \chi_0 \\ \chi \pmod{q}}} \sum_{\gamma_\chi > 0} \frac{1}{\frac{1}{4} + \gamma_\chi^2} \sim \varphi(q) \log q \quad \text{as } q \rightarrow \infty,$$

and

$$(2.5) \quad B_q(a, b) := \sum_{\substack{\chi \neq \chi_0 \\ \chi \pmod{q}}} \sum_{\gamma_\chi > 0} \frac{\chi\left(\frac{b}{a}\right) + \chi\left(\frac{a}{b}\right)}{\frac{1}{4} + \gamma_\chi^2} \ll \varphi(q), \quad a \neq b.$$

In view of equations (2.4) and (2.5), we have the normalised correlation bound $\frac{B_q(a, b)}{\operatorname{Var}(q)} \ll \frac{1}{\log q}$ for all distinct residues $a, b \pmod{q}$. This bound can be attained by certain pairs a, b (though only when $B_q(a, b)$ is negative), but otherwise it can be improved, depending on the arithmetic properties of a, b and q . Understanding this in general is somewhat complicated, (see section 3 of Harper and Lamzouri [7], for example), but here we can afford to restrict to special sets of residues where the behaviour is nice. The following lemma will provide us with such sets.

Lemma 2.1. *Assume GRH. Let q be large and let $1 \leq k \leq n < q/(20 \log^2 q)$. Then there is a set $\mathcal{A} = \{b_1, -b_1, b_2, -b_2, \dots, b_k, -b_k, b_{k+1}, \dots, b_n\}$ of reduced residues modulo q , with*

b_1, \dots, b_n distinct integers in $\{1, \dots, \lfloor q/2 \rfloor\}$, and satisfying the correlation bounds

$$B_q(a, a') = -\varphi(q)(\log 2)\ell_q(a, a') + O(n \log^4 q)$$

for every pair of distinct elements $a, a' \in \mathcal{A}$, where $\ell_q(a, a') = 1$ if $a + a' \equiv 0 \pmod q$, and $\ell_q(a, a') = 0$ otherwise.

Proof of Lemma 2.1. Let b_1, \dots, b_n be any distinct primes in the interval $(5n \log^2 q, 10n \log^2 q]$ that do not divide q . By the prime number theorem, this interval contains $\sim \frac{5n \log^2 q}{\log(10n \log^2 q)} \geq 4n \log q$ primes, provided q is large enough. And since q has at most $\log q$ prime factors, if we remove all those we are still left with at least $3n \log q \geq n$ primes (say), so we can indeed choose b_1, \dots, b_n in this way.

The claimed estimate for $B_q(a, a')$ now follows from Proposition 6.1 of Lamzouri [9], noting that we have $1/2 < \frac{|a|}{|a'|} < 2$ for all $a, a' \in \mathcal{A}$, so the term $\Lambda_0(\cdot)$ in that proposition always vanishes. \square

3. GAUSSIAN APPROXIMATION

In this section, we prove results that will allow us to replace the actual random variables $X(q, a_i)$ arising from prime number races by Gaussian random variables with the same covariance structure. This will be important because jointly Gaussian random variables have an explicit probability density, depending only on the means and covariances, which we can work with to analyse the probabilities of events.

3.1. Smooth test functions. We begin by constructing smooth weight functions that closely approximate indicators of the “ordering” events that we are interested in. Smooth weights will allow us to use Taylor expansion as a tool when we come to our Gaussian approximation.

Lemma 3.1. *There exists an increasing function $\theta : \mathbb{R} \rightarrow (0, 1)$ that is three times continuously differentiable, and satisfies*

$$\theta(-x) \ll e^{-x}, \quad 1 - \theta(x) \ll e^{-x} \quad \forall x \geq 0,$$

as well as

$$\left| \frac{d^m}{dx^m} \theta(x) \right| \ll \theta(x) \quad \forall 1 \leq m \leq 3, \quad \forall x \in \mathbb{R}.$$

Proof of Lemma 3.1. We take $\theta(x)$ to be the probability distribution function corresponding to a density proportional to $e^{-|x|}$, since repeatedly differentiating the exponential continues to yield an exponential. Let $f(x) = e^{-|x|}$ for $|x| \geq 1$, and when $|x| < 1$, let $f(x)$ be an

appropriate quadratic polynomial which ensures that $f \in C^2(\mathbb{R})$; the polynomial $\frac{7-4x^2+x^4}{4e}$ satisfies these requirements. If $m_0 := \int_{-\infty}^{\infty} f(t)dt$, then clearly

$$\theta(x) := \frac{1}{m_0} \int_{-\infty}^x f(t)dt$$

satisfies all the required properties of the lemma. In particular, for $x \leq -1$, we have $|\theta^{(j)}(x)| = \theta(x) = \frac{1}{m_0}e^x$ for $j = 1, 2, 3$ and for $x \geq -1$, $\theta(x) \geq 1/(em_0) \gg |\theta^{(j)}(x)|$ for $j = 1, 2, 3$. \square

By taking a product of (shifted and dilated) copies of the function $\theta(x)$, we can obtain multi-dimensional smooth test functions, which is what we shall actually need. We shall use the following notation: let $S \subseteq \{1, 2, \dots, n\} \times \{1, 2, \dots, n\}$ and let

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

We denote by $\mathbf{1}_{R(S)}(x_1, \dots, x_n)$ the indicator function of the set $R(S)$. We say that the set S is *admissible* if $R(S)$ is nonempty. In particular, S does not contain any diagonal pairs (i, i) . For an admissible set S and $\delta > 0$ we also define the two functions $h_{S,\delta}^+$ and $h_{S,\delta}^-$ by

$$(3.1) \quad h_{S,\delta}^{\pm}(x_1, \dots, x_n) := \prod_{(i,j) \in S} g(x_i - x_j \pm \sqrt{\delta}), \quad \text{where } g(x) := \theta(x/\delta).$$

We next show that these functions are good approximations of $\mathbf{1}_{R(S)}(x_1, \dots, x_n)$.

Proposition 3.2. *Let S be an admissible set. Each of $h_{S,\delta}^-$ and $h_{S,\delta}^+$ is a three times continuously differentiable function from \mathbb{R}^n to $[0, \infty)$, satisfying*

$$h_{S,\delta}^{\pm}(\mathbf{x}) \leq 1, \quad \sup_{1 \leq i \leq n} \left| \frac{\partial}{\partial x_i} h_{S,\delta}^{\pm}(\mathbf{x}) \right| \ll \frac{n}{\delta} h_{S,\delta}^{\pm}(\mathbf{x}), \quad \sup_{1 \leq i,j,k \leq n} \left| \frac{\partial^3}{\partial x_i \partial x_j \partial x_k} h_{S,\delta}^{\pm}(\mathbf{x}) \right| \ll \left(\frac{n}{\delta}\right)^3 h_{S,\delta}^{\pm}(\mathbf{x}).$$

Furthermore, provided that $e^{-1/\sqrt{\delta}} \leq 1/n^2$, we have

$$h_{S,\delta}^-(x_1, \dots, x_n) - O(e^{-1/\sqrt{\delta}}) \leq \mathbf{1}_{R(S)}(x_1, \dots, x_n) \leq h_{S,\delta}^+(x_1, \dots, x_n) + O(n^2 e^{-1/\sqrt{\delta}}).$$

An important feature of this result is the fact that the derivatives of $h_{S,\delta}^{\pm}$ are always controlled by $h_{S,\delta}^{\pm}$ itself, even at points where $h_{S,\delta}^{\pm}$ is very small. We will exploit this later in our Gaussian approximation.

Proof of Proposition 3.2. The partial derivative bounds in Proposition 3.2 follow exactly as in Lemma 4.3 of Harper and Lamzouri [7], for example, by repeated application of the product rule together with the fact that $\frac{d^m}{dx^m} g(x) = (1/\delta)^m \frac{d^m}{dx^m} \theta(x/\delta) \ll (1/\delta)^m \theta(x/\delta) = (1/\delta)^m g(x)$ for $1 \leq m \leq 3$.

For the lower bound on $\mathbf{1}_{R(S)}$, it will suffice to show that whenever $(x_1, \dots, x_n) \notin R(S)$ we have a uniform upper bound

$$h_{S,\delta}^-(x_1, \dots, x_n) \ll e^{-1/\sqrt{\delta}}.$$

Indeed, if $(x_1, \dots, x_n) \notin R(S)$ then for at least one pair $(i, j) \in S$, we have $x_i - x_j - \sqrt{\delta} < -\sqrt{\delta}$, and therefore

$$h_{S,\delta}^-(x_1, \dots, x_n) \leq g(-\sqrt{\delta}) = \theta(-1/\sqrt{\delta}) \ll e^{-1/\sqrt{\delta}}.$$

Similarly, for the upper bound it will suffice to show that whenever $(x_1, \dots, x_n) \in R(S)$ we have

$$h_{S,\delta}^+(x_1, \dots, x_n) = 1 - O(n^2 e^{-1/\sqrt{\delta}}).$$

Indeed, if $(x_1, \dots, x_n) \in R(S)$ then we have $x_i - x_j + \sqrt{\delta} \geq \sqrt{\delta}$ for each pair $(i, j) \in S$, and so

$$h_{S,\delta}^+(x_1, \dots, x_n) \geq g(\sqrt{\delta})^{\#S} \geq \theta(1/\sqrt{\delta})^{n^2},$$

which implies the result since we have $\theta(1/\sqrt{\delta}) = 1 - O(e^{-1/\sqrt{\delta}})$.

□

3.2. A Lindeberg type argument. In this subsection, we shall establish a “moderate deviation” type of Gaussian approximation result relative to smooth test functions h . By “moderate deviation”, we mean that we want the theorem to remain useful even in the tails of the Gaussian, where an approximation with a simple absolute error term would not be useful. Instead, we are seeking a result with a relative error (together with an absolute error term that is extremely small).

We shall obtain our Gaussian approximation with a version of the Lindeberg replacement strategy, which was originally used to prove the classical central limit theorem for sums of independent random variables. The idea is to Taylor expand the test function h to third order, and replace the independent summands $\mathbf{X}^{(j)}$ by corresponding Gaussian random variables $\mathbf{Z}^{(j)}$ one at a time. The slightly non-standard assumptions we shall make on h , that its partial derivatives are controlled by h itself, will allow us to make the biggest error term a relative rather than absolute one. (So far as we are aware, the use of such non-standard h is novel in this context. The use of a Lindeberg type method is certainly not novel, for example this is how Bentkus [1] proceeds.)

Let $C_1, C_3 \geq 1$, and let $h : \mathbb{R}^n \rightarrow [0, \infty)$ be a three times continuously differentiable function, such that for all $\mathbf{x} \in \mathbb{R}^n$ we have

$$h(\mathbf{x}) \leq 1, \quad \sup_{1 \leq i \leq n} \left| \frac{\partial}{\partial x_i} h(\mathbf{x}) \right| \leq C_1 h(\mathbf{x}), \quad \sup_{1 \leq i, j, k \leq n} \left| \frac{\partial^3}{\partial x_i \partial x_j \partial x_k} h(\mathbf{x}) \right| \leq C_3 h(\mathbf{x}).$$

Lemma 3.3. *Let $0 < \epsilon \leq \min\{1/(2C_1), 1/(2C_3^{1/3})\}$. Then uniformly for any fixed $\mathbf{Y} \in \mathbb{R}^n$, any real random vector $\mathbf{X} = (X_1, \dots, X_n)$ whose components have mean zero, we have the following. If $\mathbf{Z} = (Z_1, \dots, Z_n)$ is a multivariate normal random vector whose components have mean zero and the same covariances $r_{i,j} := \mathbb{E}X_iX_j$ as \mathbf{X} , then*

$$\begin{aligned} \mathbb{E}h(\mathbf{Y} + \mathbf{X}) &= (1 + \gamma(\epsilon))\mathbb{E}h(\mathbf{Y} + \mathbf{Z}) + \\ &+ O\left(C_3 \left(\sum_{i=1}^n \mathbb{E}\left(\mathbf{1}_{|X_i| > \epsilon/n} (\epsilon^3 + n^2 |X_i|^3)\right) + \sum_{i=1}^n (\epsilon^3 + n^2 r_{i,i}^{3/2}) e^{-\epsilon^2/(2n^2 r_{i,i})} \right)\right), \end{aligned}$$

where $\gamma(\epsilon) = \gamma(\epsilon, h, \mathbf{Y}, \mathbf{X})$ satisfies $|\gamma(\epsilon)| \leq 6C_3\epsilon^3$.

Proof of Lemma 3.3. By the multivariate form of Taylor's theorem, we have

$$h(\mathbf{Y} + \mathbf{X}) = h(\mathbf{Y}) + \sum_{i=1}^n X_i \frac{\partial}{\partial x_i} h(\mathbf{Y}) + \sum_{i,j=1}^n * X_i X_j \frac{\partial^2}{\partial x_i \partial x_j} h(\mathbf{Y}) + R(h, \mathbf{Y}, \mathbf{X}),$$

where the * on the sum indicates that the terms $i = j$ should be counted with weight 1/2, and where the error term satisfies

$$\begin{aligned} |R(h, \mathbf{Y}, \mathbf{X})| &\leq \sup_{\theta \in [0,1]} \sup_{1 \leq i,j,k \leq n} \left| \frac{\partial^3}{\partial x_i \partial x_j \partial x_k} h(\mathbf{Y} + \theta \mathbf{X}) \right| \left(\sum_{i=1}^n |X_i| \right)^3 \\ &\leq C_3 \sup_{\theta \in [0,1]} h(\mathbf{Y} + \theta \mathbf{X}) \left(\sum_{i=1}^n |X_i| \right)^3. \end{aligned}$$

One has the same expansion for $h(\mathbf{Y} + \mathbf{Z})$. Taking expectations and then taking the difference $\mathbb{E}h(\mathbf{Y} + \mathbf{X}) - \mathbb{E}h(\mathbf{Y} + \mathbf{Z})$, all of the main terms cancel because we assume \mathbf{Z} and \mathbf{X} have the same means and covariances, so we get

$$\begin{aligned} |\mathbb{E}h(\mathbf{Y} + \mathbf{X}) - \mathbb{E}h(\mathbf{Y} + \mathbf{Z})| &\leq \mathbb{E}|R(h, \mathbf{Y}, \mathbf{X})| + \mathbb{E}|R(h, \mathbf{Y}, \mathbf{Z})| \\ &\leq C_3 \left(\mathbb{E} \left\{ \sup_{\theta \in [0,1]} h(\mathbf{Y} + \theta \mathbf{X}) \left(\sum_{i=1}^n |X_i| \right)^3 \right\} + \mathbb{E} \left\{ \sup_{\theta \in [0,1]} h(\mathbf{Y} + \theta \mathbf{Z}) \left(\sum_{i=1}^n |Z_i| \right)^3 \right\} \right). \end{aligned}$$

Now using Taylor's theorem again, for any $\theta \in [0, 1]$ we have

$$\begin{aligned} |h(\mathbf{Y} + \theta \mathbf{X}) - h(\mathbf{Y} + \mathbf{X})| &\leq \sup_{\phi \in [0,1]} \sup_{1 \leq i \leq n} \left| \frac{\partial}{\partial x_i} h(\mathbf{Y} + \phi \mathbf{X}) \right| \sum_{i=1}^n |X_i| \\ &\leq C_1 \sup_{\phi \in [0,1]} h(\mathbf{Y} + \phi \mathbf{X}) \sum_{i=1}^n |X_i|. \end{aligned}$$

In particular, if we happen to have $\sum_{i=1}^n |X_i| \leq \epsilon \leq 1/(2C_1)$ then it follows that

$$\sup_{\theta \in [0,1]} h(\mathbf{Y} + \theta \mathbf{X}) \leq 2h(\mathbf{Y} + \mathbf{X}).$$

So we always have the bound

$$\mathbb{E} \sup_{\theta \in [0,1]} h(\mathbf{Y} + \theta \mathbf{X}) \left(\sum_{i=1}^n |X_i| \right)^3 \leq 2\epsilon^3 \mathbb{E} h(\mathbf{Y} + \mathbf{X}) + \mathbb{E} \mathbf{1}_{\sum_{i=1}^n |X_i| > \epsilon} \left(\sum_{i=1}^n |X_i| \right)^3.$$

The same argument applies to $\sup_{\theta \in [0,1]} h(\mathbf{Y} + \theta \mathbf{Z})$.

Putting everything together, we get

$$\begin{aligned} |\mathbb{E} h(\mathbf{Y} + \mathbf{X}) - \mathbb{E} h(\mathbf{Y} + \mathbf{Z})| &\leq 2C_3 \epsilon^3 (\mathbb{E} h(\mathbf{Y} + \mathbf{X}) + \mathbb{E} h(\mathbf{Y} + \mathbf{Z})) + \\ &+ C_3 \left(\mathbb{E} \left(\mathbf{1}_{\sum_{i=1}^n |X_i| > \epsilon} \left(\sum_{i=1}^n |X_i| \right)^3 \right) + \mathbb{E} \left(\mathbf{1}_{\sum_{i=1}^n |Z_i| > \epsilon} \left(\sum_{i=1}^n |Z_i| \right)^3 \right) \right). \end{aligned}$$

Furthermore, our assumptions on ϵ imply that $2C_3 \epsilon^3 \leq 1/4$, and so

$$\frac{1 + 2C_3 \epsilon^3}{1 - 2C_3 \epsilon^3} = 1 + \frac{4C_3 \epsilon^3}{1 - 2C_3 \epsilon^3} \leq 1 + 6C_3 \epsilon^3,$$

and similarly $\frac{1 - 2C_3 \epsilon^3}{1 + 2C_3 \epsilon^3} \geq 1 - 4C_3 \epsilon^3$. So rearranging our above bound, we find that

$$\begin{aligned} \mathbb{E} h(\mathbf{Y} + \mathbf{X}) &= (1 + \gamma(\epsilon)) \mathbb{E} h(\mathbf{Y} + \mathbf{Z}) + \\ &O \left(C_3 \left(\mathbb{E} \left(\mathbf{1}_{\sum_{i=1}^n |X_i| > \epsilon} \left(\sum_{i=1}^n |X_i| \right)^3 \right) + \mathbb{E} \left(\mathbf{1}_{\sum_{i=1}^n |Z_i| > \epsilon} \left(\sum_{i=1}^n |Z_i| \right)^3 \right) \right) \right), \end{aligned}$$

where $|\gamma(\epsilon)| \leq 6C_3 \epsilon^3$.

Next we want to work with the ‘‘big Oh’’ term a little, to replace terms depending on $\sum_{i=1}^n |X_i|$ and $\sum_{i=1}^n |Z_i|$ by terms that only require information about individual components X_i and Z_i , as in the statement of the lemma. First, Hölder’s inequality implies that $(\sum_{i=1}^n |X_i|)^3 \leq n^2 \sum_{i=1}^n |X_i|^3$. Furthermore, by splitting into cases according as each $|X_i| > \epsilon/n$ or not, we get

$$\begin{aligned} n^2 \mathbb{E} \left(\mathbf{1}_{\sum_{i=1}^n |X_i| > \epsilon} \sum_{i=1}^n |X_i|^3 \right) &\leq n^2 \mathbb{E} \left(\mathbf{1}_{\sum_{i=1}^n |X_i| > \epsilon} \sum_{i=1}^n ((\epsilon/n)^3 + \mathbf{1}_{|X_i| > \epsilon/n} |X_i|^3) \right) \\ &\leq n^2 \sum_{i=1}^n \mathbb{E} \left(\mathbf{1}_{|X_i| > \epsilon/n} |X_i|^3 \right) + \epsilon^3 \mathbb{E} \left(\mathbf{1}_{\sum_{i=1}^n |X_i| > \epsilon} \right). \end{aligned}$$

And we can bound this further using the inequality

$$\mathbb{E} \mathbf{1}_{\sum_{i=1}^n |X_i| > \epsilon} \leq \mathbb{E} \mathbf{1}_{\max_i |X_i| > \epsilon/n} \leq \sum_{i=1}^n \mathbb{E} \mathbf{1}_{|X_i| > \epsilon/n},$$

so overall we have

$$\mathbb{E} \mathbf{1}_{\sum_{i=1}^n |X_i| > \epsilon} \left(\sum_{i=1}^n |X_i| \right)^3 \leq n^2 \mathbb{E} \mathbf{1}_{\sum_{i=1}^n |X_i| > \epsilon} \sum_{i=1}^n |X_i|^3 \leq \sum_{i=1}^n \mathbb{E} \mathbf{1}_{|X_i| > \epsilon/n} (\epsilon^3 + n^2 |X_i|^3).$$

We have the analogous bound for the term involving the Z_i .

Finally, in the case of the Z_i , since we know that $Z_i \sim N(0, r_{i,i})$ we can further say that $\mathbb{E}(\mathbf{1}_{|Z_i| > \epsilon/n}) = \mathbb{P}(|Z_i| > \epsilon/n) \ll e^{-\epsilon^2/(2n^2 r_{i,i})}$, and that

$$\mathbb{E}(\mathbf{1}_{|Z_i| > \epsilon/n} |Z_i|^3) \ll r_{i,i}^{3/2} (1 + (\epsilon/n \sqrt{r_{i,i}})^3) e^{-\epsilon^2/(2n^2 r_{i,i})} \ll \left(r_{i,i}^{3/2} + \frac{\epsilon^3}{n^3} \right) e^{-\epsilon^2/(2n^2 r_{i,i})}.$$

Inserting these estimates in the “big Oh” term completes the proof. \square

Applying Lemma 3.3 inductively, we shall prove our Gaussian approximation result for sums of m independent random vectors.

Proposition 3.4. *For $1 \leq j \leq m$, let $\mathbf{X}^{(j)} = (X_1^{(j)}, \dots, X_n^{(j)})$ be independent \mathbb{R}^n -valued random vectors whose components have mean zero, and let $\mathbf{Z}^{(j)}$, $1 \leq j \leq m$, be independent multivariate normal random vectors whose components have mean zero and the same covariances $r^{(j)}_{i,k} = \mathbb{E}X_i^{(j)}X_k^{(j)}$ as $\mathbf{X}^{(j)}$.*

Let $0 < \epsilon \leq \min\{1/(2C_1), 1/(3C_3^{1/3}m^{1/3})\}$ be a small parameter.

Then we have

$$\begin{aligned} \mathbb{E}h\left(\sum_{j=1}^m \mathbf{X}^{(j)}\right) &= e^{\Delta(\epsilon)} \mathbb{E}h\left(\sum_{j=1}^m \mathbf{Z}^{(j)}\right) + O\left(C_3 \sum_{j=1}^m \sum_{i=1}^n \mathbb{E}\left(\mathbf{1}_{|X_i^{(j)}| > \epsilon/n} (\epsilon^3 + n^2 |X_i^{(j)}|^3)\right)\right) \\ &\quad + O\left(C_3 \sum_{j=1}^m \sum_{i=1}^n (\epsilon^3 + n^2 r^{(j)}_{i,i})^{3/2} e^{-\epsilon^2/(2n^2 r^{(j)}_{i,i})}\right), \end{aligned}$$

where $\Delta(\epsilon) = \Delta_m(\epsilon, h, \{\mathbf{X}^{(j)}\})$ satisfies $|\Delta(\epsilon)| \leq 12mC_3\epsilon^3$.

To get an idea of the potential usefulness of this result, the reader might consider the case where all components $X_i^{(j)}$ have variance $\asymp 1/m$ (and so all components of the sum $\sum_{j=1}^m \mathbf{X}^{(j)}$ have variance $\asymp 1$), and C_1, C_3, n are fairly small compared with m . Because $\Delta(\epsilon)$ decays cubically with ϵ , any choice of ϵ that is rather smaller than $1/(mC_3)^{1/3}$ will make the relative error term $e^{\Delta(\epsilon)}$ close to 1. Meanwhile, if the random components $X_i^{(j)}$ are reasonably concentrated on the order of their standard deviations, we expect any choice of ϵ rather larger than n/\sqrt{m} to yield substantial savings in the “big Oh” terms. So we have room to make a choice of ϵ that simultaneously controls all these terms.

Proof of Proposition 3.4. We proceed by induction on m . Our inductive hypothesis will be the estimate stated in the proposition, with an additional multiplier $e^{6C_3\epsilon^3(m-1)}$ in the “big Oh” terms. If we can prove this inductively we will be done, since our conditions on m imply this factor is $\leq e^{6/27} \ll 1$.

When $m = 1$, Proposition 3.4 is a direct consequence of Lemma 3.3, with \mathbf{Y} taken as the 0 vector.

For the inductive step, if we first take $\mathbf{Y} = \sum_{j=1}^{m-1} \mathbf{X}^{(j)}$ and condition on the value of \mathbf{Y} (which is independent of $\mathbf{X}^{(m)}$ and $\mathbf{Z}^{(m)}$), then Lemma 3.3 implies that

$$\begin{aligned} \mathbb{E}h(\mathbf{Y} + \mathbf{X}^{(m)}) &= (1 + \gamma(\epsilon))\mathbb{E}h(\mathbf{Y} + \mathbf{Z}^{(m)}) + O\left(C_3 \sum_{i=1}^n \mathbb{E}\left(\mathbf{1}_{|X_i^{(m)}| > \epsilon/n} (\epsilon^3 + n^2 |X_i^{(m)}|^3)\right)\right) \\ &\quad + O\left(C_3 \sum_{i=1}^n (\epsilon^3 + n^2 r(m)_{i,i}^{3/2}) e^{-\epsilon^2/(2n^2 r(m)_{i,i})}\right). \end{aligned}$$

Now if we condition on the value of $\mathbf{Z}^{(m)}$, and apply the inductive hypothesis with $h(\cdot)$ replaced by $h(\cdot + \mathbf{Z}^{(m)})$ (which obeys all the same partial derivative bounds), we get that $\mathbb{E}h(\sum_{j=1}^{m-1} \mathbf{X}^{(j)} + \mathbf{Z}^{(m)})$ is

$$\begin{aligned} e^{\Delta_{m-1}(\epsilon)} \mathbb{E}h\left(\sum_{j=1}^{m-1} \mathbf{Z}^{(j)} + \mathbf{Z}^{(m)}\right) &+ O\left(e^{6C_3\epsilon^3(m-1)} C_3 \sum_{j=1}^{m-1} \sum_{i=1}^n \mathbb{E}\mathbf{1}_{|X_i^{(j)}| > \epsilon/n} (\epsilon^3 + n^2 |X_i^{(j)}|^3)\right) \\ &+ O\left(e^{6C_3\epsilon^3(m-1)} C_3 \sum_{j=1}^{m-1} \sum_{i=1}^n (\epsilon^3 + n^2 r(j)_{i,i}^{3/2}) e^{-\epsilon^2/(2n^2 r(j)_{i,i})}\right). \end{aligned}$$

The above is then multiplied by $(1 + \gamma(\epsilon))$ in the expression for $\mathbb{E}h(\mathbf{Y} + \mathbf{X}^{(m)})$. Using the fact that

$$e^{-12C_3\epsilon^3} \leq e^{-2|\gamma(\epsilon)|} \leq (1 + \gamma(\epsilon)) \leq e^{|\gamma(\epsilon)|} \leq e^{6C_3\epsilon^3},$$

we complete the induction. \square

3.3. Application to prime number races. In this subsection, we specialise the discussion in the preceding propositions to the case of prime number races. Recall the random variables

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

from section 2, and recall the sets $R(S) = \{(x_1, \dots, x_n) \in \mathbb{R}^n : x_i \geq x_j \forall (i, j) \in S\} \subseteq \mathbb{R}^n$ and the corresponding smooth test functions $h_{S,\delta}^-, h_{S,\delta}^+$ from section 3.1.

Proposition 3.5. *Let q be large, and suppose that $2 \leq n \leq \varphi(q)^{1/12}$.*

Let $\mathbf{X} := \left(\frac{X(q, a_i)}{\sqrt{\text{Var}(q)}} \right)_{1 \leq i \leq n}$, where a_1, \dots, a_n are coprime residue classes mod q , and let $\mathbf{Z} = (Z_i)_{1 \leq i \leq n}$ 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 small parameter $(\frac{n^5 \log q}{\sqrt{\varphi(q)}})^{1/3} \leq \delta \leq \frac{1}{\log^2 n}$ and any admissible set S , we have

$$\begin{aligned} & \left(1 + O\left(\frac{1}{\log q}\right) \right) \mathbb{E}h_{S, \delta}^-(\mathbf{Z}) - O\left(e^{-2/\sqrt{\delta}} + ne^{-\Theta\left(\frac{\delta^2 \varphi(q)^{1/3}}{n^4 \log^{2/3} q}\right)}\right) \\ & \leq \mathbb{P}(\mathbf{X} \in R(S)) \leq \left(1 + O\left(\frac{1}{\log q}\right) \right) \mathbb{E}h_{S, \delta}^+(\mathbf{Z}) + O\left(n^2 e^{-2/\sqrt{\delta}} + ne^{-\Theta\left(\frac{\delta^2 \varphi(q)^{1/3}}{n^4 \log^{2/3} q}\right)}\right). \end{aligned}$$

We remark that the condition $n \leq \varphi(q)^{1/12}$ is stronger than necessary, but is convenient at one point in the proof, and without it the proposition is trivial because the ‘‘big Oh’’ term will not be less than 1.

Proof of Proposition 3.5. In the first place, the condition $\delta \leq \frac{1}{\log^2 n}$ means that Proposition 3.2 is applicable with δ replaced by $\delta/4$, and gives

$$\mathbb{E}h_{S, \delta/4}^-(\mathbf{X}) - O(e^{-2/\sqrt{\delta}}) \leq \mathbb{P}(\mathbf{X} \in R(S)) \leq \mathbb{E}h_{S, \delta/4}^+(\mathbf{X}) + O(n^2 e^{-2/\sqrt{\delta}}).$$

Now the shifts $D_i := \frac{C_q(a_i)}{\sqrt{\text{Var}(q)}}$ in the components $X_i := \frac{X(q, a_i)}{\sqrt{\text{Var}(q)}}$ (see (2.2)) are a little awkward, since they cause the components to have non-zero mean. However, recalling our expressions for $C_q(a)$ (2.3) and $\text{Var}(q)$ (2.4), together with the lower bound $\delta \geq (\frac{n^5 \log q}{\sqrt{\varphi(q)}})^{1/3} \geq \frac{1}{\varphi(q)^{1/6}}$, we find that

$$|D_i| \leq \frac{\sum_{m|q} 1}{\sqrt{\text{Var}(q)}} \ll \frac{1}{\varphi(q)^{0.49}} \leq \frac{\sqrt{\delta}}{4}.$$

It follows that for any pair (i, j) we have

$$X_i - X_j + \sqrt{\delta/4} \leq \tilde{X}_i - \tilde{X}_j + \sqrt{\delta}$$

and

$$X_i - X_j - \sqrt{\delta/4} \geq \tilde{X}_i - \tilde{X}_j - \sqrt{\delta},$$

where

$$\tilde{X}_i = X_i + D_i \quad (1 \leq i \leq n).$$

So recalling the definitions (3.1) of $h_{S, \delta/4}^+$, $h_{S, \delta/4}^-$, we may remove the shifts D_i and still have the same upper and lower bounds for $\mathbb{P}(\mathbf{X} \in R(S))$, at the cost of replacing $h_{S, \delta/4}^\pm$ by $h_{S, \delta}^\pm$.

That is, we have

$$\mathbb{E}h_{S,\delta}^-(\tilde{\mathbf{X}}) - O(e^{-2/\sqrt{\delta}}) \leq \mathbb{P}(\mathbf{X} \in R(S)) \leq \mathbb{E}h_{S,\delta}^+(\tilde{\mathbf{X}}) + O(n^2 e^{-2/\sqrt{\delta}}).$$

Now we want to show that $\mathbb{E}h_{S,\delta}^+(\tilde{\mathbf{X}})$ may be replaced by $\mathbb{E}h_{S,\delta}^+(\mathbf{Z})$, and $\mathbb{E}h_{S,\delta}^-(\tilde{\mathbf{X}})$ by $\mathbb{E}h_{S,\delta}^-(\mathbf{Z})$, up to acceptable error terms. We apply Proposition 3.4 with the sum over $1 \leq j \leq m$ replaced by a sum over characters $\chi \neq \chi_0 \pmod{q}$ (so $m = \varphi(q) - 1$), and with \tilde{X}_j replaced by

$$\tilde{X}^{(x)} := \left(\frac{1}{\sqrt{\text{Var}(q)}} \text{Re} \left(2\chi(a_i) \sum_{\gamma_\chi > 0} \frac{U(\gamma_\chi)}{\sqrt{\frac{1}{4} + \gamma_\chi^2}} \right) \right)_{1 \leq i \leq n}.$$

Notice that these are indeed independent \mathbb{R}^n -valued random vectors whose components have mean zero, since the underlying random variables $U(\gamma_\chi)$ are independent and have mean zero. For the test functions $h_{S,\delta}^\pm$, we may take $C_1 \asymp n/\delta$ and $C_3 \asymp (n/\delta)^3$, so in Proposition 3.4 we are permitted to make any choice of $0 < \epsilon \leq \delta/(n\varphi(q))^{1/3}$.

Turning to the error terms in Proposition 3.4, for each $1 \leq i \leq n$ and each $\chi \neq \chi_0$ we have

$$r(\chi)_{i,i} = \frac{1}{\text{Var}(q)} \sum_{\gamma_\chi > 0} \frac{\mathbb{E}(\text{Re} 2\chi(a_i) U(\gamma_\chi))^2}{\frac{1}{4} + \gamma_\chi^2} = \frac{2}{\text{Var}(q)} \sum_{\gamma_\chi > 0} \frac{1}{\frac{1}{4} + \gamma_\chi^2} \ll \frac{1}{\varphi(q)},$$

where the final inequality uses the standard estimate $\sum_{\gamma_\chi > 0} \frac{1}{\frac{1}{4} + \gamma_\chi^2} \ll \log q$ (see Corollary 10.18 of Montgomery and Vaughan [11], for example). Furthermore, an exponential moment calculation (as in the proof of Lemma 2.3 of Lamzouri [8]) implies that

$$\mathbb{P}(|\tilde{X}_i^{(x)}| > r) \ll e^{-\Theta(r^2 \varphi(q))}$$

for each $1 \leq i \leq n$, each $\chi \neq \chi_0$ and any $r \geq 0$. This simply says that, as we might expect, the components $\tilde{X}_i^{(x)}$ have Gaussian-type tails. A consequence of this bound is

$$\mathbb{E} \mathbf{1}_{|\tilde{X}_i^{(x)}| > r} |\tilde{X}_i^{(x)}|^3 \ll \left(r^3 + \frac{1}{\varphi(q)^{3/2}} \right) e^{-\Theta(r^2 \varphi(q))}.$$

Inserting all this information in Proposition 3.4, we get for any $0 < \epsilon \leq \delta/(n\varphi(q))^{1/3}$ that

$$\mathbb{E}h_{S,\delta}^\pm(\tilde{\mathbf{X}}) = e^{\Delta(\epsilon)} \mathbb{E}h_{S,\delta}^\pm(\mathbf{Z}) + O\left(\left(\frac{n}{\delta} \right)^3 \varphi(q) n \left(\epsilon^3 + \frac{n^2}{\varphi(q)^{3/2}} \right) e^{-\Theta(\epsilon^2 \varphi(q)/n^2)} \right),$$

where $|\Delta(\epsilon)| \ll \varphi(q)(n\epsilon/\delta)^3$.

Finally, if we take $\epsilon = \frac{\delta}{n\varphi(q)^{1/3} \log^{1/3} q}$ then, in view of the condition $\delta \geq \left(\frac{n^5 \log q}{\sqrt{\varphi(q)}} \right)^{1/3}$, we have $\epsilon^3 \geq \frac{n^2}{\varphi(q)^{3/2}}$. Then a quick calculation verifies that our ‘‘big Oh’’ term is of the form claimed in the proposition. \square

the last $n - 2k$ components of \mathbf{x} , and by $\|\mathbf{x}\| = (x_1^2 + \cdots + x_n^2)^{1/2}$ the Euclidean norm of the full vector \mathbf{x} . We prove the following result.

Proposition 4.1. *Let q be sufficiently large and assume (4.1). Let $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$. Then we have*

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \exp \left(\xi \sum_{j=1}^k x_{2j-1} x_{2j} - \frac{\|\mathbf{x}_{2k}\|^2}{2} \left(1 + O \left(\frac{1}{(\log q)^2} \right) \right) \right. \\ \left. - \frac{\|\mathbf{x}^{n-2k}\|^2}{2} (1 + O(\varepsilon n)) + O \left(\frac{k}{(\log q)^2} \right) \right)$$

where ε satisfies (4.3). We also have the cruder estimate

$$(4.5) \quad f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2}} \exp \left(-\frac{\|\mathbf{x}_{2k}\|^2}{2} \left(1 + O \left(\frac{1}{\log q} \right) \right) - \frac{\|\mathbf{x}^{n-2k}\|^2}{2} (1 + O(\varepsilon n)) + O \left(\frac{k}{(\log q)^2} \right) \right).$$

To establish this result we first prove the following lemma.

Lemma 4.2. *Suppose q is sufficiently large that $|\xi| \leq 1/2$ and $\varepsilon \leq \frac{1}{4n}$. Then*

- $\det C = (\det A)(1 + O(\varepsilon n))$;
- C is invertible, and $C^{-1} = A^{-1} + F$, where the entries of F are bounded in absolute value by 8ε .

Proof. Write $C = A + E = A(I - E')$, where $E' = -A^{-1}E$. As $|\xi| \leq 1/2$, and the inverse of $\begin{pmatrix} 1 & \xi \\ \xi & 1 \end{pmatrix}$ is $(1 - \xi^2)^{-1} \begin{pmatrix} 1 & -\xi \\ -\xi & 1 \end{pmatrix}$, we easily see that the entries of E' are bounded in absolute value by 2ε . Then $\det C = (\det A) \det(I - E')$ and, writing the determinant as a sum over permutations $\sigma \in S_n$, with t the number of fixed points of σ ,

$$\det(I - E') = (1 + O(\varepsilon))^n + O \left(\sum_{\text{id} \neq \sigma \in S_n} (2\varepsilon)^{n-t} (1 + 2\varepsilon)^t \right) \\ = 1 + O(\varepsilon n) + O \left(\sum_{t=0}^{n-2} n^{n-t} (2\varepsilon)^{n-t} \right) = 1 + O(\varepsilon n).$$

Here we made several uses of our assumption that $|\varepsilon n| \leq 1/4$.

Also

$$C^{-1} = (I + E' + (E')^2 + (E')^3 + \cdots) A^{-1},$$

where the infinite series converges because the entries of $(E')^j$ are (by an easy induction) bounded in absolute value by $n^{j-1} (2\varepsilon)^j \leq \frac{2\varepsilon}{2^j - 1}$. Hence, $C^{-1} = (I + E'') A^{-1}$, where the entries of E'' are bounded in absolute value by 4ε . The second part now follows. \square

Proof of Proposition 4.1. First, note that $\det A = (1 - \xi^2)^k$ and recall that the inverse of $\begin{pmatrix} 1 & \xi \\ \xi & 1 \end{pmatrix}$ is $(1 - \xi^2)^{-1} \begin{pmatrix} 1 & -\xi \\ -\xi & 1 \end{pmatrix}$. Therefore, it follows from (4.4) and Lemma 4.2 that

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} (1 - \xi^2)^{-\frac{k}{2}} (1 + O(\varepsilon n)) \times \\ \times \exp \left\{ -\frac{1}{2(1 - \xi^2)} \sum_{j=1}^k (x_{2j-1}^2 + x_{2j}^2 - 2\xi x_{2j-1}x_{2j}) - \frac{1}{2} \sum_{j=2k+1}^n x_j^2 + \sum_{h,j=1}^n \varepsilon_{h,j} x_h x_j \right\},$$

where $|\varepsilon_{h,j}| \leq 8\varepsilon$ for every $1 \leq h, j \leq n$. By the Cauchy–Schwarz inequality we obtain

$$\left| \sum_{h,j=1}^n \varepsilon_{h,j} x_h x_j \right| \leq 8\varepsilon \left(\sum_{h=1}^n |x_h| \right)^2 \leq 8\varepsilon n \|\mathbf{x}\|^2 = 8\varepsilon n (\|\mathbf{x}_{2k}\|^2 + \|\mathbf{x}^{n-2k}\|^2),$$

and thus we deduce that

$$(4.6) \quad f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} (1 - \xi^2)^{-\frac{k}{2}} (1 + O(\varepsilon n)) \times \\ \times \exp \left\{ \frac{\xi}{1 - \xi^2} \sum_{j=1}^k x_{2j-1}x_{2j} - \|\mathbf{x}_{2k}\|^2 \left(\frac{1}{2(1 - \xi^2)} + O(\varepsilon n) \right) - \|\mathbf{x}^{n-2k}\|^2 \left(\frac{1}{2} + O(\varepsilon n) \right) \right\}.$$

Invoking (4.2), we have that $\frac{1}{1 - \xi^2} = 1 + O(1/\log^2 q)$. Also, by our assumption (4.1) and the estimate (4.3), it follows that

$$\varepsilon n \ll \frac{1}{\log^2 q}.$$

Another application of Cauchy–Schwarz yields

$$(4.7) \quad \sum_{j=1}^k x_{2j-1}x_{2j} \leq \|\mathbf{x}_{2k}\|^2,$$

and therefore

$$\frac{\xi}{1 - \xi^2} \sum_{j=1}^k x_{2j-1}x_{2j} = \xi \sum_{j=1}^k x_{2j-1}x_{2j} + O(\xi^3 \|\mathbf{x}_{2k}\|^2) \\ = \xi \sum_{j=1}^k x_{2j-1}x_{2j} + O((\log q)^{-3} \|\mathbf{x}_{2k}\|^2).$$

Combining these estimates with (4.6) gives the first estimate in Proposition 4.1.

To obtain the crude estimate (4.5), we combine (4.2) and (4.7) to get

$$\xi \sum_{j=1}^k x_{2j-1}x_{2j} \ll \frac{\|\mathbf{x}_{2k}\|^2}{\log q}.$$

Inserting this bound into the first estimate of Proposition 4.1 completes the proof. \square

We conclude this section with an estimate for the tails of our multidimensional Gaussian $\mathbf{Z} = (Z_1, \dots, Z_n)$.

Lemma 4.3. *Let q be sufficiently large, and set $\mathcal{Q} := \log q$. Also assume (4.1). Then*

$$\begin{aligned} \mathbb{P}\left(\sum_{i=2k+1}^n Z_i^2 > 10n \log(n\mathcal{Q})\right) &\leq e^{-3n \log(n\mathcal{Q}) + O(n)}, \\ \mathbb{P}\left(\sum_{i=1}^{2k} Z_i^2 > 10k \log(n\mathcal{Q})\right) &\leq e^{-3k \log(n\mathcal{Q}) + O(k)}. \end{aligned}$$

Notice that both of these probabilities are negligible compared with the target probability $\exp\left(-\frac{0.6k \log(n/k)}{\log q}\right) \frac{(n-2k)!}{n!}$ that we shall be aiming for later (see (5.1)).

Proof of Lemma 4.3. Arguing as before, the density function of $(Z_i)_{2k+1 \leq i \leq n}$ takes the form

$$\frac{(1 + O(\varepsilon n))}{(2\pi)^{(n-2k)/2}} \exp\left\{-\frac{1}{2}\|\mathbf{x}^{n-2k}\|^2(1 + O(\varepsilon n))\right\},$$

and since $\varepsilon n \ll \frac{1}{\log^2 q}$ we get

$$\begin{aligned} \mathbb{P}\left(\sum_{i=2k+1}^n Z_i^2 > 10n \log(n\mathcal{Q})\right) &\ll \int_{\|\mathbf{x}^{n-2k}\|^2 > 10n \log(n\mathcal{Q})} \frac{e^{-\frac{1}{2}\|\mathbf{x}^{n-2k}\|^2(1+O(\frac{1}{\log^2 q}))}}{(2\pi)^{(n-2k)/2}} dx_{2k+1} \dots dx_n \\ &\leq e^{-3n \log(n\mathcal{Q})} \left(\int_{-\infty}^{\infty} \frac{e^{-\frac{x^2}{6}}}{\sqrt{2\pi}} dx\right)^{n-2k} = e^{-3n \log(n\mathcal{Q}) + O(n)}. \end{aligned}$$

The proof of the second part of the lemma is exactly similar, using that (similarly as in Proposition 4.1) the density function of $(Z_i)_{1 \leq i \leq 2k}$ takes the form

$$\frac{1}{(2\pi)^k} \exp\left\{-\frac{1}{2}\|\mathbf{x}\|^2 \left(1 + O\left(\frac{1}{\log q}\right)\right) + O\left(\frac{k}{(\log q)^2}\right)\right\}.$$

□

5. PROOF OF THEOREM 1.3: SOME REDUCTIONS

In this section, we carry out some preliminary reductions and calculations for the proof of Theorem 1.3. Again, we let $(a_1, \dots, a_n) = (b_1, -b_1, \dots, b_k, -b_k, b_{k+1}, \dots, b_{n-k})$ be a tuple of residues as in section 4 (i.e. with the b_i satisfying the conclusion of Lemma 2.1).

We begin by discussing the first bound (1.3) claimed in Theorem 1.3. We let

$$S := \{(i, i+1) : 1 \leq i \leq 2k-1\} \cup \{(2k, j) : 2k+1 \leq j \leq n\}.$$

Then in view of (2.1) and the surrounding discussion in section 2, and applying Proposition 3.5 with the choice $\delta = 1/(n \log q)^5$, we have

$$\begin{aligned} \delta_{2k}(q; a_1, \dots, a_n) &= \mathbb{P} \left(\frac{X(q, a_1)}{\sqrt{\text{Var}(q)}} > \frac{X(q, a_2)}{\sqrt{\text{Var}(q)}} > \dots > \frac{X(q, a_{2k})}{\sqrt{\text{Var}(q)}} > \max_{2k+1 \leq j \leq n} \frac{X(q, a_j)}{\sqrt{\text{Var}(q)}} \right) \\ &\leq \left(1 + O \left(\frac{1}{\log q} \right) \right) \mathbb{E} h_{S, \delta}^+(\mathbf{Z}) + O \left(n^2 e^{-2(n \log q)^{5/2}} + n e^{-\Theta \left(\frac{\varphi(q)^{1/3}}{n^{14} (\log q)^{10+2/3}} \right)} \right), \end{aligned}$$

where $\mathbf{Z} = (Z_i)_{1 \leq i \leq n}$ are the jointly Gaussian random variables from Proposition 3.5 that correspond to the residues a_i . If we now restrict to the range $1 \leq k \leq n/A \leq \varphi(q)^{1/50}$ in Theorem 1.3, then we have $(\frac{n^5 \log q}{\sqrt{\varphi(q)}})^{1/3} \leq \delta \leq \frac{1}{\log^2 n}$ in Proposition 3.5, and furthermore the “big Oh” term above is $\ll e^{-(n \log q)^2}$. This is negligible for (1.3).

Thus, to prove (1.3), it will suffice to show that

$$(5.1) \quad \mathbb{E} h_{S, \delta}^+(\mathbf{Z}) \leq \exp \left(-\frac{0.6k \log(n/k)}{\log q} \right) \frac{(n-2k)!}{n!},$$

for q sufficiently large.

We expect (since this is how we constructed things) that the smooth function $h_{S, \delta}^+$ should behave a lot like the indicator function $\mathbf{1}_{R(S)}$. Thus, if we apply $h_{S, \delta}^+$ to a vector of *independent* standard normal random variables, and take expectations, we expect the answer to be close to $\frac{(n-2k)!}{n!}$, the probability that such a vector would satisfy the ordering dictated by S . Our next auxiliary lemma is an upper bound that reflects this expectation.

Lemma 5.1. *Let $1 \leq k \leq n/2$, and let $0 < \delta \leq \frac{1}{5k^5 \log^2 n}$. If $\mathbf{W} = (W_1, \dots, W_n)$ is a vector of independent standard normal random variables, and S is as above, we have*

$$\mathbb{E} h_{S, \delta}^+(\mathbf{W}) \leq (1 + O(\sqrt{\delta \log(1/\delta) k^4 \log n})) \frac{(n-2k)!}{n!}.$$

We remark that our previous choice $\delta = 1/(n \log q)^5$ satisfies the hypotheses of this Lemma.

Proof of Lemma 5.1. By definition, we have $h_{S, \delta}^+(\mathbf{W}) \leq g(-\sqrt{\delta}) \ll e^{-1/\sqrt{\delta}}$ unless $W_i - W_j \geq -2\sqrt{\delta}$ for all pairs $(i, j) \in S$. Furthermore, in that case we still have $h_{S, \delta}^+(\mathbf{W}) \leq 1$. So recalling our construction of S , if we define $\widetilde{W}_i := W_i + (2k+1-i)2\sqrt{\delta}$ for $1 \leq i \leq 2k$, and define $\widetilde{W}_i := W_i$ for $2k+1 \leq i \leq n$, then

$$\mathbb{E} h_{S, \delta}^+(\mathbf{W}) \leq \mathbb{P} \left(\widetilde{W}_1 > \widetilde{W}_2 > \dots > \widetilde{W}_{2k} > \max_{2k < i \leq n} \widetilde{W}_i \right) + O(e^{-1/\sqrt{\delta}}).$$

Here our upper bound condition on δ implies that

$$e^{-1/\sqrt{\delta}} \ll \sqrt{\delta \log(1/\delta) k^4 \log n} \cdot n^{-2k} \leq \sqrt{\delta \log(1/\delta) k^4 \log n} \frac{(n-2k)!}{n!},$$

and so the “big Oh” term is acceptable for Lemma 5.1.

Next, by independence the probability here is simply

$$\begin{aligned} &= \int_{x_1 > x_2 > \dots > x_{2k}} \frac{e^{-\sum_{i=1}^{2k} (x_i - (2k+1-i)2\sqrt{\delta})^2/2}}{(2\pi)^k} \left(\int_{-\infty}^{x_{2k}} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \right)^{n-2k} dx_1 \dots dx_{2k} \\ &= \int_{x_1 > x_2 > \dots > x_{2k}} \frac{e^{-\frac{1}{2}\|\mathbf{x}_{2k}\|^2 + O(k^{3/2}\sqrt{\delta}\|\mathbf{x}_{2k}\| + k^3\delta)}}{(2\pi)^k} \left(\int_{-\infty}^{x_{2k}} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \right)^{n-2k} dx_1 \dots dx_{2k}. \end{aligned}$$

The part of the integral where $\|\mathbf{x}_{2k}\|^2 > 10(k \log n + \log(1/\delta))$ is

$$\ll \int_{\|\mathbf{x}_{2k}\|^2 > 10(k \log n + \log(1/\delta))} \frac{e^{-\frac{1}{2}\|\mathbf{x}_{2k}\|^2 + O(k^{3/2}\sqrt{\delta}\|\mathbf{x}_{2k}\|)}}{(2\pi)^k} dx_1 \dots dx_{2k} \leq e^{-3(k \log n + \log(1/\delta)) + O(k)},$$

which again is acceptable for Lemma 5.1. And the complementary part of the integral, where $\|\mathbf{x}_{2k}\|^2 \leq 10(k \log n + \log(1/\delta))$, is

$$\leq e^{O(\sqrt{\delta k^3(k \log n + \log(1/\delta))})} \int_{x_1 > x_2 > \dots > x_{2k}} \frac{e^{-\frac{1}{2}\|\mathbf{x}_{2k}\|^2}}{(2\pi)^k} \left(\int_{-\infty}^{x_{2k}} \frac{e^{-t^2/2}}{\sqrt{2\pi}} dt \right)^{n-2k} dx_1 \dots dx_{2k}.$$

Finally, the integral here is simply $\mathbb{P}(W_1 > W_2 > \dots > W_{2k} > \max_{2k < i \leq n} W_i)$, which by symmetry is equal to $\frac{(n-2k)!}{n!}$. \square

Now we turn to the second bound (1.4) claimed in Theorem 1.3. Here we let

$$\begin{aligned} S^\# &:= \{(2i-1, 2i+1) : 1 \leq i \leq k-1\} \cup \{(2i+2, 2i) : 1 \leq i \leq k-1\} \\ &\quad \cup \{(2k-1, j) : 2k \leq j \leq n\} \cup \{(i, 2k) : 2k+1 \leq i \leq n\}. \end{aligned}$$

With the choice $\delta = 1/(n \log q)^5$ as before, we see that to prove (1.4) it will suffice to show that

$$(5.2) \quad \mathbb{E}h_{S^\#, \delta}^-(\mathbf{Z}) \geq \exp\left(\frac{0.6k \log(n/k)}{\log q}\right) \frac{(n-2k)!}{n!}.$$

Lemma 5.2. *Let $1 \leq k \leq n/2$, and let $0 < \delta \leq \frac{1}{5k^5 \log^2 n}$. If $\mathbf{W} = (W_1, \dots, W_n)$ is a vector of independent standard normal random variables, we have*

$$(1 - O(n^2\sqrt{\delta})) \frac{(n-2k)!}{n!} \leq \mathbb{E}h_{S^\#, \delta}^-(\mathbf{W}) \leq \mathbb{E}h_{S^\#, \delta}^+(\mathbf{W}) \leq (1 + O(\sqrt{\delta \log(1/\delta) k^4 \log n})) \frac{(n-2k)!}{n!}.$$

Proof of Lemma 5.2. For the upper bound, take

$$\begin{aligned} \widetilde{W}_{2i-1} &= W_{2i-1} + (k+1-i)2\sqrt{\delta} \quad (1 \leq i \leq k), \\ \widetilde{W}_{2i} &= W_{2i} - (k+1-i)2\sqrt{\delta} \quad (1 \leq i \leq k), \\ \widetilde{W}_i &= W_i \quad (2k+1 \leq i \leq n). \end{aligned}$$

Then, if $W_i - W_j \geq -2\sqrt{\delta}$ for all $(i, j) \in S^\#$, we have

$$\widetilde{W}_1 > \widetilde{W}_3 > \cdots > \widetilde{W}_{2k-1} > \max_{2k+1 \leq j \leq n} \widetilde{W}_j, \quad \widetilde{W}_2 < \widetilde{W}_4 < \cdots < \widetilde{W}_{2k} < \min_{2k+1 \leq j \leq n} \widetilde{W}_j.$$

The upper bound then follows similarly as in Lemma 5.1.

For the lower bound, we note that if $(x_1, \dots, x_n) \in R(S^\#)$ is such that $|x_i - x_j| \geq 2\sqrt{\delta}$ for all $i \neq j$, then checking the definition (3.1) of $h_{S^\#, \delta}^-$ we obtain

$$h_{S^\#, \delta}^-(x_1, \dots, x_n) \geq \prod_{(i, j) \in S^\#} g(\sqrt{\delta}) \geq 1 - O(ne^{-1/\sqrt{\delta}}).$$

Hence, if we temporarily let G denote the event that $|W_i - W_j| \geq 2\sqrt{\delta}$ for all $i \neq j$, then we see that $\mathbb{E}h_{S^\#, \delta}^-(\mathbf{W})$ is

$$\geq (1 - O(ne^{-1/\sqrt{\delta}}))\mathbb{P}(W_1 > W_3 > \dots > W_{2k-1} > \{W_i\}_{2k < i \leq n} > W_{2k} > \dots > W_4 > W_2, \text{ and } G).$$

However, by symmetry the probability here is equal to $\frac{(n-2k)!}{n!}\mathbb{P}(G)$, and by the union bound we have

$$\mathbb{P}(G) \geq 1 - \sum_{i \neq j} \mathbb{P}(|W_i - W_j| \leq 2\sqrt{\delta}) = 1 - O(n^2\sqrt{\delta}). \quad \square$$

6. PROOF OF THEOREM 1.3: THE CONCLUSION

In this section, we assume as in Theorem 1.3 that q is large, and that $1 \leq k \leq n/A \leq \varphi(q)^{1/50}$ for a sufficiently large absolute constant A . Let us observe at the outset that these conditions imply, in section 4 and specifically in Proposition 4.1, that $\varepsilon n \ll \frac{n^2 \log^3 q}{\varphi(q)} \leq \frac{1}{n \log^{10} q}$, say. We also recall the random variables Z_1, \dots, Z_n from sections 4 and 5, which have density function $f(x_1, \dots, x_n)$. Finally, as in Lemma 4.3, we shall sometimes write $\mathcal{Q} := \log q$.

We begin by proving the statement (1.3) in Theorem 1.3, in the special case where $k \log(n/k) > 1000 \log \log q$, say. (The case where $k \log(n/k) \leq 1000 \log \log q$ is a bit more delicate because the saving factor $\exp\left(-\frac{0.6k \log(n/k)}{\log q}\right)$ we are aiming for is only slightly smaller than 1, so that will be dealt with later.) In view of the discussion in section 5, to do this it will suffice to prove the upper bound (5.1) for $\mathbb{E}h_{S^\#, \delta}^+(\mathbf{Z})$. Let us introduce notation for the following four ‘‘good’’ events:

$$\mathcal{A} := \left\{ \sum_{i=2k+1}^n Z_i^2 \leq 10n \log(n\mathcal{Q}) \right\}, \quad \mathcal{B} := \left\{ \sum_{i=1}^{2k} Z_i^2 \leq 10k \log(n\mathcal{Q}) \right\},$$

$$\mathcal{C} := \left\{ \max_{2k < i \leq n} Z_i > \sqrt{\log(n/k)} \right\}, \quad \mathcal{U} := \left\{ \min(Z_1, \dots, Z_{2k}) \geq \max_{2k < i \leq n} Z_i - 2\sqrt{\delta} \right\},$$

where $\delta = 1/(n \log q)^5$ is the same as in section 5.

Firstly, Lemma 4.3 and the definition of $h_{S,\delta}^+$ (in particular the fact that $0 \leq h_{S,\delta}^+ \leq 1$) implies that

$$\begin{aligned}
|\mathbb{E}h_{S,\delta}^+(\mathbf{Z}) - \mathbb{E}h_{S,\delta}^+(\mathbf{Z})\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{U}}| &\leq (1 - \mathbb{P}(\mathcal{A} \cap \mathcal{B})) + \mathbb{E}(h_{S,\delta}^+(\mathbf{Z})(1 - \mathbf{1}_{\mathcal{U}})) \\
&\ll e^{-3k \log(n\mathcal{Q}) + O(k)} + g(-\sqrt{\delta}) \\
&\ll e^{-3k \log(n\mathcal{Q}) + O(k)} + e^{-(n \log q)^{5/2}} \\
(6.1) \quad &\ll e^{-3k \log(n\mathcal{Q}) + O(k)} \leq \frac{(n-2k)!}{n!} e^{-k \log(n\mathcal{Q}) + O(k)}.
\end{aligned}$$

Next, introduce the ‘‘truncated’’ set

$$S' = \{(i, i+1) : 1 \leq i \leq 2k-1\},$$

considered as a subset of $\{1, \dots, 2k\}^2$, the associated set $R(S') \subset \mathbb{R}^{2k}$ and function $h_{S',\delta}^+ : \mathbb{R}^{2k} \rightarrow \mathbb{R}$. Using the crude estimate from Proposition 4.1 we have

$$\begin{aligned}
\mathbb{E}h_{S,\delta}^+(Z)\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{U} \cap \{\mathcal{C} \text{ fails}\}} &\leq e^{O(\frac{k}{\log^2 q})} \int \cdots \int_{\|\mathbf{x}_{2k}\|^2 \leq 10k \log(n\mathcal{Q})} h_{S',\delta}^+(x_1, \dots, x_{2k}) \frac{e^{-\frac{\|\mathbf{x}_{2k}\|^2}{2}(1+O(\frac{1}{\log q}))}}{(2\pi)^k} dx_1 \cdots dx_{2k} \\
(6.2) \quad &\times \int \cdots \int_{\substack{\max_{2k+1 \leq i \leq n} x_i \leq \sqrt{\log(n/k)}, \\ \|\mathbf{x}^{n-2k}\|^2 \leq 10n \log(n\mathcal{Q})}} \frac{e^{-\frac{\|\mathbf{x}^{n-2k}\|^2}{2}(1+O(\frac{1}{n \log^{10} q}))}}{(2\pi)^{(n-2k)/2}} dx_{2k+1} \cdots dx_n.
\end{aligned}$$

Now the integral on the first line here is

$$\leq e^{O(\frac{k \log(n\mathcal{Q})}{\log q})} \int h_{S',\delta}^+(x_1, \dots, x_{2k}) \frac{e^{-\frac{\|\mathbf{x}_{2k}\|^2}{2}}}{(2\pi)^k} dx_1 \cdots dx_{2k} \ll e^{O(\frac{k \log(n\mathcal{Q})}{\log q})} \frac{1}{(2k)!} = \frac{e^{O(k)}}{k^{2k}},$$

where the second inequality follows from Lemma 5.1 (with n replaced by $2k$), and the final equality follows from Stirling’s formula. Meanwhile, the integral on the second line is

$$\ll \left(\int_{-\infty}^{\sqrt{\log(n/k)}} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \right)^{n-2k} = \left(1 - \Theta \left(\frac{e^{-\log(n/k)/2}}{\sqrt{\log(n/k)}} \right) \right)^{n-2k} = \exp \left\{ -\Theta \left(\sqrt{\frac{nk}{\log(n/k)}} \right) \right\}.$$

(Notice here that $n/k \geq A$ is large, under our hypotheses.) So putting things together, we have shown that

$$\begin{aligned}
\mathbb{E} \{ h_{S,\delta}^+(\mathbf{Z})\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{U} \cap \{\mathcal{C} \text{ fails}\}} \} &\leq \frac{e^{O(k)}}{k^{2k}} e^{-\Theta(\sqrt{\frac{nk}{\log(n/k)}})} = \frac{1}{n^{2k}} e^{O(k \log(n/k)) - \Theta(k \sqrt{\frac{n/k}{\log(n/k)}})} \\
(6.3) \quad &\leq \frac{1}{n^{2k}} e^{-1000k \log(n/k)} \\
&\leq \frac{(n-2k)!}{n!} e^{-1000k \log(n/k)},
\end{aligned}$$

say.

On the other hand, the first part of Proposition 4.1 implies that on those tuples (x_1, \dots, x_n) corresponding to all four events $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{U}$, the density function $f(x_1, \dots, x_n)$ satisfies

$$\begin{aligned} f(\mathbf{x}) &\leq \frac{e^{\xi k \log(n/k) + O\left(\frac{k}{\log^2 q}\right)}}{(2\pi)^{n/2}} \exp \left\{ -\frac{\|\mathbf{x}_{2k}\|^2}{2} \left(1 + O\left(\frac{1}{\log^2 q}\right)\right) - \frac{\|\mathbf{x}^{n-2k}\|^2}{2} \left(1 + O\left(\frac{1}{n \log^{10} q}\right)\right) \right\} \\ &= \frac{e^{\xi k \log(n/k) + O\left(\frac{k \log(n\mathcal{Q})}{\log^2 q}\right)}}{(2\pi)^{n/2}} \exp \left\{ -\frac{\|\mathbf{x}\|^2}{2} \right\}, \end{aligned}$$

where $\xi \sim -\frac{\log 2}{\log q}$ (recall (4.2)). The crucial thing to notice here is the emergence of the bias term $e^{\xi k \log(n/k)}$, which came from the non-trivial correlations of size ξ amongst pairs $(Z_{2i-1}, Z_{2i})_{1 \leq i \leq k}$, together with the fact that we have arranged to have $Z_1, \dots, Z_{2k} \geq \sqrt{\log(n/k)} - 2\sqrt{\delta}$. Using this upper bound, we get

$$\begin{aligned} \mathbb{E} \{ h_{S,\delta}^+(\mathbf{Z}) \mathbf{1}_{\mathcal{A}\mathcal{B}\mathcal{C}\mathcal{U}} \} &\leq e^{\xi k \log(n/k) + O\left(\frac{k \log(n\mathcal{Q})}{\log^2 q}\right)} \int h_{S,\delta}^+(x_1, \dots, x_n) \frac{e^{-\frac{\|\mathbf{x}\|^2}{2}}}{(2\pi)^{n/2}} dx_1 \dots dx_n \\ &= e^{\xi k \log(n/k) + O\left(\frac{k \log(n\mathcal{Q})}{\log^2 q}\right)} \mathbb{E} h_{S,\delta}^+(\mathbf{W}) \leq e^{\xi k \log(n/k) + O\left(\frac{k \log(n\mathcal{Q})}{\log^2 q}\right)} \frac{(n-2k)!}{n!}, \end{aligned}$$

where $\mathbf{W} = (W_1, \dots, W_n)$ is a vector of independent standard Gaussian random variables, and where the second inequality follows from Lemma 5.1 on noting that

$$\left(1 + O\left(\sqrt{\delta \log(1/\delta) k^4 \log n}\right)\right) \leq \left(1 + O\left(\sqrt{\frac{\log^2(n\mathcal{Q})}{n \log^5 q}}\right)\right) \leq e^{O\left(\frac{k \log(n\mathcal{Q})}{\log^2 q}\right)}.$$

The error in the exponent satisfies

$$O\left(\frac{k \log(n\mathcal{Q})}{\log^2 q}\right) = O\left(k \left(\frac{\log \log q}{\log^2 q} + \frac{1}{\log q}\right)\right) = O\left(\frac{k}{\log q}\right),$$

and thus

$$(6.4) \quad \mathbb{E} \{ h_{S,\delta}^+(\mathbf{Z}) \mathbf{1}_{\mathcal{A}\mathcal{B}\mathcal{C}\mathcal{U}} \} \leq e^{\xi k \log(n/k) + O\left(\frac{k}{\log q}\right)} \frac{(n-2k)!}{n!}.$$

Again recall that $\xi \sim -\frac{\log 2}{\log q}$ as $q \rightarrow \infty$. Hence, for q sufficiently large, and A sufficiently large, the O -term in the exponent is smaller than $0.01|\xi k \log(n/k)|$ and consequently

$$\xi k \log(n/k) + O\left(\frac{k}{\log q}\right) \leq -0.65 \frac{k \log(n/k)}{\log q}.$$

Adding this together with (6.1) and (6.3), and recalling that we are in the special case where $k \log(n/k) > 1000 \log \log q$, we indeed have a bound $\leq \exp\left(-\frac{0.6k \log(n/k)}{\log q}\right) \frac{(n-2k)!}{n!}$ as we wanted for (5.1).

Next we turn to the other case where $k \log(n/k) \leq 1000 \log \log q$. Notice that in this case, in the definition of the event \mathcal{B} we have $10k \log(n\mathcal{Q}) = 10k(\log k + \log(n/k) + \log \mathcal{Q}) \ll (\log \log q)^2$, and so we get $\|\mathbf{x}_{2k}\|^2 \ll (\log \log q)^2$. Thus, if W_1, \dots, W_n are independent standard normal random variables, and if we write ‘primed’ events (e.g., \mathcal{A}') for the corresponding events with each Z_i replaced by W_i , then the above calculations in fact imply that

$$\begin{aligned} \mathbb{E}h_\delta^+(\mathbf{Z})\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{U} \cap \mathcal{C}} &\leq e^{\xi k \log(n/k) + O(\frac{k}{\log q})} \mathbb{E}h_\delta^+(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \mathcal{U}' \cap \mathcal{C}'} \\ &= (1 + \xi k \log(n/k) + O(\frac{k}{\log q})) \mathbb{E}h_\delta^+(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \mathcal{U}' \cap \mathcal{C}'}, \end{aligned}$$

and that (using the crude density estimate from Proposition 4.1, as in (6.2))

$$\begin{aligned} &|\mathbb{E}h_\delta^+(\mathbf{Z})\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{U} \cap \{\mathcal{C} \text{ fails}\}} - \mathbb{E}h_\delta^+(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \mathcal{U}' \cap \{\mathcal{C}' \text{ fails}\}}| \\ &\ll \int_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{U} \cap \{\mathcal{C} \text{ fails}\}} h_\delta^+(x_1, \dots, x_n) \frac{e^{-\frac{\|\mathbf{x}\|^2}{2}}}{(2\pi)^{n/2}} \left(\frac{k}{\log^2 q} + \frac{\|\mathbf{x}_{2k}\|^2}{\log q} \right) dx_1 \dots dx_n. \end{aligned}$$

The part of this integral where $\|\mathbf{x}_{2k}\|^2 \leq 10k \log k$ contributes $\ll \frac{k \log k}{\log q} \mathbb{E}h_\delta^+(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \mathcal{U}' \cap \{\mathcal{C}' \text{ fails}\}}$, which is $\ll \frac{1}{\log q} \frac{(n-2k)!}{n!}$ by the calculations leading to (6.3). Meanwhile, another short calculation shows the other part of the integral contributes

$$\begin{aligned} &\ll \frac{1}{\log q} \int_{\|\mathbf{x}_{2k}\|^2 > 10k \log k} \frac{e^{-\frac{\|\mathbf{x}_{2k}\|^2}{2}}}{(2\pi)^k} \|\mathbf{x}_{2k}\|^2 dx_1 \dots dx_{2k} \cdot \mathbb{P}(\max_{2k < i \leq n} W_i \leq \sqrt{\log(n/k)}) \\ &\ll \frac{1}{\log q} e^{-3k \log k + O(k)} e^{-\Theta(\sqrt{\frac{nk}{\log(n/k)}})} \ll \frac{1}{\log q} \frac{(n-2k)!}{n!}. \end{aligned}$$

The calculations leading to (6.3) further imply that $\xi k \log(n/k) \mathbb{E}h_\delta^+(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \mathcal{U}' \cap \{\mathcal{C}' \text{ fails}\}} \ll \frac{k \log(n/k)}{\log q} \mathbb{E}h_\delta^+(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \mathcal{U}' \cap \{\mathcal{C}' \text{ fails}\}} \ll \frac{1}{\log q} \frac{(n-2k)!}{n!}$. So we can collect together the above computations, along with (6.1), in the form

$$\begin{aligned} \mathbb{E}h_\delta^+(\mathbf{Z}) &= \mathbb{E}h_\delta^+(\mathbf{Z})\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{U} \cap \mathcal{C}} + \mathbb{E}h_\delta^+(\mathbf{Z})\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{U} \cap \{\mathcal{C} \text{ fails}\}} + (\mathbb{E}h_\delta^+(\mathbf{Z}) - \mathbb{E}h_\delta^+(\mathbf{Z})\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{U}}) \\ &\leq (1 + \xi k \log(n/k) + O(\frac{k}{\log q})) \mathbb{E}h_\delta^+(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \mathcal{U}'} + O(\frac{1}{\log q} \frac{(n-2k)!}{n!}). \end{aligned}$$

And Lemma 5.1 implies that $\mathbb{E}h_\delta^+(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \mathcal{U}'} \leq \mathbb{E}h_\delta^+(\mathbf{W}) \leq (1 + O(\sqrt{\frac{\log \log q}{\log^4 q}})) \frac{(n-2k)!}{n!}$, which suffices to establish (5.1) in the case where $k \log(n/k) \leq 1000 \log \log q$.

Next we shall prove the other statement (1.4) in Theorem 1.3. To do this it will suffice to prove the lower bound (5.2) for $\mathbb{E}h_{S\#, \delta}^-(\mathbf{Z})$. In addition to the three typical events $\mathcal{A}, \mathcal{B}, \mathcal{C}$

considered above, we introduce the following:

$$\mathcal{D} := \left\{ \min_{2k < i \leq n} Z_i < -\sqrt{\log(n/k)} \right\},$$

$$\mathcal{V} := \left\{ Z_1, Z_3, \dots, Z_{2k-1} \geq \max_{2k < i \leq n} Z_i - 2\sqrt{\delta}, \quad \text{and} \quad Z_2, Z_4, \dots, Z_{2k} \leq \min_{2k < i \leq n} Z_i + 2\sqrt{\delta} \right\}.$$

Similarly as above, on those tuples (x_1, \dots, x_n) corresponding to all the events $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}, \mathcal{V}$, the density function $f(x_1, \dots, x_n)$ satisfies

$$f(\mathbf{x}) \geq \frac{e^{-\xi k \log(n/k) + O\left(\frac{k \log(n\mathcal{Q})}{\log^2 q}\right)}}{(2\pi)^{n/2}} \exp\left(-\frac{\|\mathbf{x}\|^2}{2}\right).$$

Again, the key thing to notice here is the large bias factor $e^{-\xi k \log(n/k)}$, which emerged because we arranged to have

$$\begin{aligned} \min(Z_1, Z_3, \dots, Z_{2k-1}) &\geq \sqrt{\log(n/k)} - 2\sqrt{\delta}, \quad \text{and} \\ \max(Z_2, Z_4, \dots, Z_{2k}) &\leq -(\sqrt{\log(n/k)} - 2\sqrt{\delta}). \end{aligned}$$

So if we again let $\mathbf{W} = (W_1, \dots, W_n)$ denote a vector of independent standard Gaussian random variables, we have that

$$\begin{aligned} \mathbb{E}h_{S^\#, \delta}^-(\mathbf{Z}) &\geq \mathbb{E}h_{S^\#, \delta}^-(\mathbf{Z})\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{V}} = \mathbb{E}h_{S^\#, \delta}^-(\mathbf{Z})\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{C} \cap \mathcal{D} \cap \mathcal{V}} + \mathbb{E}h_{S^\#, \delta}^-(\mathbf{Z})\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{V} \cap \{\mathcal{C} \text{ or } \mathcal{D} \text{ fails}\}} \\ &\geq e^{-\xi k \log(n/k) + O\left(\frac{k \log(n\mathcal{Q})}{\log^2 q}\right)} \mathbb{E}h_{S^\#, \delta}^-(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \mathcal{C}' \cap \mathcal{D}' \cap \mathcal{V}'} + \mathbb{E}h_{S^\#, \delta}^-(\mathbf{Z})\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \mathcal{V} \cap \{\mathcal{C} \text{ or } \mathcal{D} \text{ fails}\}}, \end{aligned}$$

where the ‘primed’ events (e.g., \mathcal{A}') are the corresponding events with each Z_i replaced by W_i . As before, for A and q large enough, the O -term in the exponent is negligible and we have $e^{-\xi k \log(n/k) + O\left(\frac{k \log(n\mathcal{Q})}{\log^2 q}\right)} \geq e^{0.65k \frac{\log(n/k)}{\log q}}$, and so

$$\begin{aligned} \mathbb{E}h_{S^\#, \delta}^-(\mathbf{Z}) &\geq e^{0.65k \frac{\log(n/k)}{\log q}} \mathbb{E}h_{S^\#, \delta}^-(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \mathcal{V}'} \\ (6.5) \quad &+ \mathbb{E}h_{S^\#, \delta}^-(\mathbf{Z})\mathbf{1}_{\mathcal{A} \cap \mathcal{B} \cap \{\mathcal{C} \text{ or } \mathcal{D} \text{ fails}\} \cap \mathcal{V}} - e^{0.65k \frac{\log(n/k)}{\log q}} \mathbb{E}h_{S^\#, \delta}^-(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \{\mathcal{C}' \text{ or } \mathcal{D}' \text{ fails}\} \cap \mathcal{V}'}. \end{aligned}$$

Furthermore (the proofs of) Lemmas 4.3 and 5.2 together imply, with our choice $\delta = 1/(n \log q)^5$, that

$$\mathbb{E}h_{S^\#, \delta}^-(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \mathcal{V}'} \geq (1 - O(n^2 \sqrt{\delta})) \frac{(n-2k)!}{n!} + O(e^{-3k \log(n\mathcal{Q}) + O(k)}) = (1 - O\left(\frac{1}{\log^{5/2} q}\right)) \frac{(n-2k)!}{n!}.$$

We can also mimic the calculations leading to (6.3), this time using Lemma 5.2 in place of Lemma 5.1 to bound $\int h_{S^\#, \delta}^-(x_1, \dots, x_{2k}) \frac{e^{-\frac{1}{2}\|\mathbf{x}_{2k}\|^2}}{(2\pi)^k} dx_1 \cdots dx_{2k}$, and obtain that

$$\mathbb{E} \left\{ h_{S^\#, \delta}^-(\mathbf{W})\mathbf{1}_{\mathcal{A}' \cap \mathcal{B}' \cap \mathcal{V}' \cap \{\mathcal{C}' \text{ or } \mathcal{D}' \text{ fails}\}} \right\} \leq \frac{1}{n^{2k}} e^{-1000k \log(n/k)} \leq \frac{(n-2k)!}{n!} e^{-1000k \log(n/k)},$$

and the same for $\mathbb{E} \left\{ h_{S^\#, \delta}^-(\mathbf{Z}) \mathbf{1}_{A \cap B \cap \mathcal{V} \cap \{\mathcal{C} \text{ or } \mathcal{D} \text{ fails}\}} \right\}$. Hence when $k \log(n/k) > 1000 \log \log q$, the final two terms in (6.5) make a negligible contribution, and the desired lower bound (5.2) for $\mathbb{E} h_{S^\#, \delta}^-(\mathbf{Z})$ follows. In the remaining case where $k \log(n/k) \leq 1000 \log \log q$, we can again duplicate our previous approach and show the *difference* of the final two terms in (6.5) is $\ll \frac{1}{\log q} \frac{(n-2k)!}{n!}$, which is negligible for (5.2). \square

REFERENCES

- [1] V. Y. Bentkus, *Large deviations in Banach spaces*. Theory Probab. Appl. **31** (1986), no. 4, 627-632.
- [2] A. Feuerverger and G. Martin, *Biases in the Shanks-Rényi prime number race*. Experiment. Math. **9** (2000), no. 4, 535-570.
- [3] D. Fiorilli, *Highly biased prime number races*. Algebra Number Theory **8** (2014), no. 7, 1733-1767.
- [4] D. Fiorilli and G. Martin, *Inequities in the Shanks-Rényi Prime Number Race: An asymptotic formula for the densities*. J. Reine Angew. Math. **676** (2013), 121-212.
- [5] K. Ford and S. Konyagin, *Chebyshev's conjecture and the prime number race*. Modern Problems of Number Theory and its Applications; Topical Problems Part II (Tula, Russia, 2001).
- [6] A. Granville and G. Martin, *Prime number races*. Amer. Math. Monthly **113** (2006), no. 1, 1-33.
- [7] A. J. Harper and Y. Lamzouri, *Orderings of weakly correlated random variables, and prime number races with many contestants*. To appear in Probab. Theory Related Fields.
- [8] Y. Lamzouri, *The Shanks-Rényi prime number race with many contestants*. Math. Res. Lett. **19** (2012), no. 03, 649-666.
- [9] Y. Lamzouri, *Prime number races with three or more competitors*, Math. Annalen **356** (2013), 1117-1162.
- [10] G. Martin and J. Scarfy, *Comparative prime number theory: A survey*. 37 pages. [arXiv:1202.3408](https://arxiv.org/abs/1202.3408)
- [11] H. L. Montgomery and R. C. Vaughan, *Multiplicative Number Theory I: Classical Theory*. First edition, published by Cambridge University Press, 2007.
- [12] M. Rubinstein and P. Sarnak, *Chebyshev's bias*. Experiment. Math. **3** (1994), no. 3, 173-197.

KF: DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, 1409 WEST GREEN STREET, URBANA, IL 61801, USA

E-mail address: ford@math.uiuc.edu

AJH: MATHEMATICS INSTITUTE, ZEEMAN BUILDING, UNIVERSITY OF WARWICK, COVENTRY CV4 7AL, ENGLAND

E-mail address: A.Harper@warwick.ac.uk

YL: DEPARTMENT OF MATHEMATICS AND STATISTICS, YORK UNIVERSITY, 4700 KEELE STREET, TORONTO, ON, M3J1P3 CANADA

E-mail address: lamzouri@mathstat.yorku.ca