

Chains of large gaps between primes

Kevin Ford, James Maynard and Terence Tao

Abstract Let p_n denote the n -th prime, and for any $k \geq 1$ and sufficiently large X , define the quantity

$$G_k(X) := \max_{p_{n+k} \leq X} \min(p_{n+1} - p_n, \dots, p_{n+k} - p_{n+k-1}),$$

which measures the occurrence of chains of k consecutive large gaps of primes. Recently, with Green and Konyagin, the authors showed that

$$G_1(X) \gg \frac{\log X \log \log X \log \log \log \log X}{\log \log \log X}$$

for sufficiently large X . In this note, we combine the arguments in that paper with the Maier matrix method to show that

$$G_k(X) \gg \frac{1}{k^2} \frac{\log X \log \log X \log \log \log \log X}{\log \log \log X}$$

for any fixed k and sufficiently large X . The implied constant is effective and independent of k .

K. Ford

Department of Mathematics, 1409 West Green Street, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

e-mail: ford@math.uiuc.edu

J. Maynard

Mathematical Institute, Radcliffe Observatory Quarter, Woodstock Road, Oxford OX2 6GG, England

e-mail: james.alexander.maynard@gmail.com

T. Tao

Department of Mathematics, UCLA, 405 Hilgard Ave, Los Angeles CA 90095, USA

e-mail: tao@math.ucla.edu

1 Introduction

Let p_n denote the n^{th} prime, and for any $k \geq 1$ and sufficiently large X , let

$$G_k(X) := \max_{p_{n+k} \leq X} \min(p_{n+1} - p_n, \dots, p_{n+k} - p_{n+k-1}),$$

denote the maximum gap between k consecutive primes less than X . The quantity $G_1(X)$ has been extensively studied. The prime number theorem implies that

$$G_1(X) \geq (1 + o(1)) \log X,$$

with the bound being successively improved in many papers [1], [4], [25], [9], [22], [24], [23], [15], [20], [18], [10], [11]. The best lower bound currently is¹

$$G_1(X) \gg \frac{\log X \log_2 X \log_4 X}{\log_3 X},$$

for sufficiently large X and an effective implied constant, due to [11]. This result may be compared against the conjecture $G_1(X) \asymp \log^2 X$ of Cramér [7] (see also [13]), or the upper bound $G_1(X) \ll X^{0.525}$ of Baker-Harman-Pintz [3], which can be improved to $G_1(X) \ll X^{1/2} \log X$ on the Riemann hypothesis [6].

Now we turn to $G_k(X)$ in the regime where $k \geq 1$ is fixed, and X assumed sufficiently large depending on k . Clearly $G_k(X) \leq G_1(X)$, and a naive extension of the probabilistic heuristics of Cramér [7] suggest that $G_k(X) \asymp \frac{1}{k} \log^2 X$ as $X \rightarrow \infty$. The first non-trivial bound on $G_k(X)$ for $k \geq 2$ was by Erdős [9], who showed that

$$G_2(X) / \log X \rightarrow \infty$$

as $X \rightarrow \infty$. Using what is now known as the Maier matrix method, together with the arguments of Rankin [22] on $G_1(X)$, Maier [14] showed that

$$G_k(X) \gg_k \frac{\log X \log_2 X \log_4 X}{(\log_3 X)^2}$$

for any fixed $k \geq 1$ and a sequence of X going to infinity. Recently, by modifying Maier's arguments and using the more recent work on $G_1(X)$ in [10], [18], this was improved by Pintz [19] to show that

$$G_k(X) / \left(\frac{\log X \log_2 X \log_4 X}{(\log_3 X)^2} \right) \rightarrow \infty$$

for a sequence of X going to infinity.

Our main result here is as follows.

Theorem 1 *Let $k \geq 1$ be fixed. Then for sufficiently large X , we have*

¹ As usual in the subject, $\log_2 x := \log \log x$, $\log_3 x := \log \log \log x$, and so on. The conventions for asymptotic notation such as \ll and $o(\cdot)$ will be defined in Section 1.2.

$$G_k(X) \gg \frac{1}{k^2} \frac{\log X \log_2 X \log_4 X}{\log_3 X}.$$

The implied constant is absolute and effective.

Maier’s original argument required one to avoid Siegel zeroes, which restricted his results to a sequence of X going to infinity, rather than all sufficiently large X . However, it is possible to modify his argument to remove the effect of any exceptional zeroes, which allows us to extend the result to all sufficiently large X and also to make the implied constant effective. The intuitive reason for the $\frac{1}{k^2}$ factor is that our method produces, roughly speaking, k primes distributed “randomly” inside an interval of length about $\frac{\log X \log_2 X \log_4 X}{\log_3 X}$, and the narrowest gap between k independently chosen numbers in an interval of length L is typically of length about $\frac{1}{k^2} L$.

Our argument is based heavily on our previous paper [11], in particular using the hypergraph covering lemma from [11, Corollary 3] and the construction of sieve weights from [11, Theorem 5]. The main difference is in refining the probabilistic analysis in [11] to obtain good upper and lower bounds for certain sifted sets arising in the arguments in [11], whereas in the former paper only upper bounds were obtained.

We remark that in the recent paper [2], the methods from [11] were modified to obtain some information about the limit points of tuples of k consecutive prime gaps normalized by factors slightly slower than $\frac{\log X \log_2 X \log_4 X}{\log_3 X}$; see Theorem 6.4 of that paper for a precise statement.

1.1 Acknowledgments

KF thanks the hospitality of the Institute of Mathematics and Informatics of the Bulgarian Academy of Sciences. The research of JM was conducted partly while he was a CRM-ISM postdoctoral fellow at the Université de Montréal, and partly while he was a Fellow by Examination at Magdalen College, Oxford.

KF was supported by NSF grants DMS-1201442 and DMS-1501982. TT was supported by a Simons Investigator grant, the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund Endowment, and by NSF grant DMS-1266164.

The authors thank Tristan Freiberg for some corrections.

1.2 Notational conventions

In most of the paper, x will denote an asymptotic parameter going to infinity, with many quantities allowed to depend on x . The symbol $o(1)$ will stand for a quantity bounded in magnitude by $c(x)$, where $c(x)$ is a quantity that tends to zero as

$x \rightarrow \infty$. The same convention applies to the asymptotic notation $X \sim Y$, which means $X = (1 + o(1))Y$, and $X \lesssim Y$, which means $X \leq (1 + o(1))Y$. We use $X = O(Y)$, $X \ll Y$, and $Y \gg X$ to denote the claim that there is a constant $C > 0$ such that $|X| \leq CY$ throughout the domain of the quantity X . We adopt the convention that C is independent of any parameter unless such dependence is indicated, e.g. by subscript such as \ll_k . In all of our estimates here, the constant C will be effective (we will not rely on ineffective results such as Siegel's theorem). If we can take the implied constant C to equal 1, we write $f = O_{\leq}(g)$ instead. Thus for instance

$$X = (1 + O_{\leq}(\varepsilon))Y$$

is synonymous with

$$(1 - \varepsilon)Y \leq X \leq (1 + \varepsilon)Y.$$

Finally, we use $X \asymp Y$ synonymously with $X \ll Y \ll X$.

When summing or taking products over the symbol p , it is understood that p is restricted to be prime.

Given a modulus q and an integer n , we use $n \bmod q$ to denote the congruence class of n in $\mathbb{Z}/q\mathbb{Z}$.

Given a set A , we use 1_A to denote its indicator function, thus $1_A(x)$ is equal to 1 when $x \in A$ and zero otherwise. Similarly, if E is an event or statement, we use 1_E to denote the indicator, equal to 1 when E is true and 0 otherwise. Thus for instance $1_A(x)$ is synonymous with $1_{x \in A}$.

We use $\#A$ to denote the cardinality of A , and for any positive real z , we let $[z] := \{n \in \mathbf{N} : 1 \leq n \leq z\}$ denote the set of natural numbers up to z .

Our arguments will rely heavily on the probabilistic method. Our random variables will mostly be discrete (in the sense that they take at most countably many values), although we will occasionally use some continuous random variables (e.g. independent real numbers sampled uniformly from the unit interval $[0, 1]$). As such, the usual measure-theoretic caveats such as “absolutely integrable”, “measurable”, or “almost surely” can be largely ignored by the reader in the discussion below. We will use boldface symbols such as \mathbf{X} or \mathbf{a} to denote random variables (and non-boldface symbols such as X or a to denote deterministic counterparts of these variables). Vector-valued random variables will be denoted in arrowed boldface, e.g. $\mathbf{a} = (\mathbf{a}_p)_{p \in \mathcal{P}}$ might denote a random tuple of random variables \mathbf{a}_p indexed by some index set \mathcal{P} .

We write \mathbb{P} for probability, and \mathbb{E} for expectation. If \mathbf{X} takes at most countably many values, we define the *essential range* of \mathbf{X} to be the set of all X such that $\mathbb{P}(\mathbf{X} = X)$ is non-zero, thus \mathbf{X} almost surely takes values in its essential range. We also employ the following conditional expectation notation. If E is an event of non-zero probability, we write

$$\mathbb{P}(F|E) := \frac{\mathbb{P}(F \wedge E)}{\mathbb{P}(E)}$$

for any event F , and

$$\mathbb{E}(\mathbf{X}|E) := \frac{\mathbb{E}(\mathbf{X}1_E)}{\mathbb{P}(E)}$$

for any (absolutely integrable) real-valued random variable \mathbf{X} . If \mathbf{Y} is another random variable taking at most countably many values, we define the conditional probability $\mathbb{P}(F|\mathbf{Y})$ to be the random variable that equals $\mathbb{P}(F|\mathbf{Y} = Y)$ on the event $\mathbf{Y} = Y$ for each Y in the essential range of \mathbf{Y} , and similarly define the conditional expectation $\mathbb{E}(\mathbf{X}|\mathbf{Y})$ to be the random variable that equals $\mathbb{E}(\mathbf{X}|\mathbf{Y} = Y)$ on the event $\mathbf{Y} = Y$. We observe the idempotency property

$$\mathbb{E}(\mathbb{E}(\mathbf{X}|\mathbf{Y})) = \mathbb{E}\mathbf{X} \quad (1)$$

whenever \mathbf{X} is absolutely integrable and \mathbf{Y} takes at most countably many values.

We will rely frequently on the following simple concentration of measure result.

Lemma 1.1 (Chebyshev inequality) *Let \mathbf{X}, \mathbf{Y} be independent random variables taking at most countably many values. Let \mathbf{Y}' be a conditionally independent copy of \mathbf{Y} over \mathbf{X} ; in other words, for every X in the essential range of \mathbf{X} , the random variables \mathbf{Y}, \mathbf{Y}' are independent and identically distributed after conditioning to the event $\mathbf{X} = X$. Let $F(\mathbf{X}, \mathbf{Y})$ be a (absolutely integrable) random variable depending on \mathbf{X} and \mathbf{Y} . Suppose that one has the bounds*

$$\mathbb{E}F(\mathbf{X}, \mathbf{Y}) = \alpha + O(\varepsilon\alpha) \quad (2)$$

and

$$\mathbb{E}F(\mathbf{X}, \mathbf{Y})F(\mathbf{X}, \mathbf{Y}') = \alpha^2 + O(\varepsilon\alpha^2) \quad (3)$$

for some $\alpha, \varepsilon > 0$ with $\varepsilon = O(1)$. Then for any $\theta > 0$, one has

$$\mathbb{E}(F(\mathbf{X}, \mathbf{Y})|\mathbf{X}) = \alpha + O_{\leq}(\theta) \quad (4)$$

with probability $1 - O(\frac{\varepsilon\alpha^2}{\theta^2})$.

Proof. See [11, Lemma 1.2].

2 Siegel zeroes

As is common in analytic number theory, we will have to address the possibility of an exceptional Siegel zero. As we want to keep all our estimates effective, we will not rely on Siegel's theorem or its consequences (such as the Bombieri-Vinogradov theorem). Instead, we will rely on the Landau-Page theorem, which we now recall. Throughout, χ denotes a Dirichlet character.

Lemma 2.1 (Landau-Page theorem) *Let $Q \geq 100$. Suppose that $L(s, \chi) = 0$ for some primitive character χ of modulus at most Q , and some $s = \sigma + it$. Then either*

$$1 - \sigma \gg \frac{1}{\log(Q(1 + |t|))},$$

or else $t = 0$ and χ is a quadratic character χ_Q , which is unique for any given Q . Furthermore, if χ_Q exists, then its conductor q_Q is square-free apart from a factor of at most 4, and obeys the lower bound

$$q_Q \gg \frac{\log^2 Q}{\log_2^2 Q}.$$

Proof. See e.g. [8, Chapter 14]. The final estimate follows from the classical bound $1 - \beta \gg q^{-1/2} \log^{-2} q$ for a real zero β of $L(s, \chi)$ with χ of modulus q .

We can then eliminate the exceptional character by deleting at most one prime factor of Q .

Corollary 1 *Let $Q \geq 100$. Then there exists a quantity B_Q which is either equal to 1 or is a prime of size*

$$B_Q \gg \log_2 Q$$

with the property that

$$1 - \sigma \gg \frac{1}{\log(Q(1 + |t|))}$$

whenever $L(\sigma + it, \chi) = 0$ and χ is a character of modulus at most Q and coprime to B_Q .

Proof. If the exceptional character χ_Q from Lemma 2.1 does not exist, then take $B_Q := 1$; otherwise we take B_Q to be the largest prime factor of q_Q . As q_Q is square-free apart from a factor of at most 4, we have $\log q_Q \ll B_Q$ by the prime number theorem, and the claim follows.

Next, we recall Gallagher's prime number theorem:

Lemma 2.2 (Gallagher's prime number theorem) *Let q be a natural number, and suppose that $L(s, \chi) \neq 0$ for all characters χ of modulus q and s with $1 - \sigma \leq \frac{\delta}{\log(Q(1 + |t|))}$, and some constant $\delta > 0$. Then there is a constant $D \geq 1$ depending only on δ such that*

$$\#\{p \text{ prime} : p \leq x; p \equiv a \pmod{q}\} \gg \frac{x}{\phi(q) \log x}$$

for all $(a, q) = 1$ and $x \geq q^D$.

Proof. See [14, Lemma 2].

This will combine well with Corollary 1 once we remove the moduli divisible by the (possible) exceptional prime B_Q .

3 Sieving an interval

We now give the key sieving result that will be used to prove Theorem 1.

Theorem 2 (Sieving an interval) *There is an absolute constants $c > 0$ such that the following holds. Fix $A \geq 1$ and $\varepsilon > 0$, and let x be sufficiently large depending on A and ε . Suppose y satisfies*

$$y = c \frac{x \log x \log_3 x}{\log_2 x}, \quad (5)$$

and suppose that $B_0 = 1$ or that B_0 is a prime satisfying

$$\log x \ll B_0 \leq x.$$

Then one can find a congruence class $a_p \pmod p$ for each prime $p \leq x$, $p \neq B_0$ such that the sieved set

$$\mathcal{T} := \{n \in [y] \setminus [x] : n \not\equiv a_p \pmod p \text{ for all } p \leq x, p \neq B_0\}$$

obeys the following size estimates:

- (Upper Bound) *One has*

$$\#\mathcal{T} \ll A \frac{x}{\log x}. \quad (6)$$

- (Lower Bound) *One has*

$$\#\mathcal{T} \gg A \frac{x}{\log x}. \quad (7)$$

- (Upper bound in short intervals) *For any $0 \leq \alpha \leq \beta \leq 1$, one has*

$$\#(\mathcal{T} \cap [\alpha y, \beta y]) \ll A(|\beta - \alpha| + \varepsilon) \frac{x}{\log x}. \quad (8)$$

We remark that if one lowers y to be of order $\frac{x \log x \log_3 x}{(\log_2 x)^2}$ rather than $\frac{x \log x \log_3 x}{\log_2 x}$, then this theorem is essentially [14, Lemma 6]. It is convenient to sieve $[y] \setminus [x]$ instead of $[y]$ for minor technical reasons (we will use the fact that the residue class $0 \pmod p$ avoids all the primes in $[y] \setminus [x]$ whenever $p \leq x$). The arguments in [11] already can give much of this theorem, with the exception of the lower bound (7), which is the main additional technical result of this paper that is needed to extend the results of that paper to longer chains.

We will prove Theorem 2 in later sections. In this section, we show how this theorem implies Theorem 1. Here we shall use the Maier matrix method, following the arguments in [14] closely (although we will use probabilistic notation rather than matrix notation). Let $k \geq 1$ be a fixed integer, let $c_0 > 0$ be a small constant, and let $A \geq 1$ and $0 < \varepsilon < 1/2$ be large and small quantities depending on k to be chosen later.

We now recall (a slight variant of) some lemmas from [14].

Lemma 3.1 *There exists an absolute constant $D \geq 1$ such that, for all sufficiently large x , there exists a natural number B_0 which is either equal to 1 or a prime, with*

$$\log x \ll B_0 \leq x, \quad (9)$$

and is such that the following holds. If one sets $P := P(x)/B_0$ (where we recall that $P(x)$ is the product of the primes up to x), then one has

$$\#\{z \in [Z] : Pz + a \text{ prime}\} \gg \frac{\log x}{\log Z} Z \quad (10)$$

for all $Z \geq P^D$ and $a \in P$ coprime to P , and

$$\#\{z \in [Z] : Pz + a, Pz + b \text{ both prime}\} \ll \left(\frac{\log x}{\log Z}\right)^2 Z \quad (11)$$

for all $Z \geq P^D$ and all distinct $a, b \in [P]$ coprime to P .

Proof. We first prove (10). We apply Corollary 1 with $Q := P(x)$ to obtain a quantity $B_{P(x)}$ with the stated properties. We set $B_0 = 1$ if $B_{P(x)} > x$, and $B_0 := B_{P(x)}$ otherwise. Then from Mertens' theorem we have (9) if $B_0 \neq 1$. From Corollary 1 and Lemma 2.2, we then have

$$\#\{z \in [Z] : Pz + a \text{ prime}\} \gg \frac{PZ}{\phi(P) \log(PZ)}$$

for any $Z \geq P^D$ and a suitable absolute constant $D \geq 1$. Note that $\log(PZ) \ll \log Z$. From Mertens' theorem (and (9)) we also have

$$\frac{P}{\phi(P)} \asymp \log x, \quad (12)$$

and (10) follows.

Finally, the estimate (11) follows from standard upper bound sieves (cf. [14, Lemma 3]).

Now set $Z := P^D$ with x and D as in Lemma 3.1, and let \mathbf{z} be chosen uniformly at random from $[Z]$. Let y , \mathcal{T} and $a_p \bmod p$ be as in Theorem 2. By the Chinese remainder theorem, we may find $m \in [P]$ such that $m \equiv -a_p \pmod{p}$ for all $p \leq x$ with $p \neq B_0$. Thus, $\mathbf{z}P + m + \mathcal{T}$ consists precisely of those elements of $\mathbf{z}P + m + [y] \setminus [x]$ that are coprime to P . In particular, any primes that lie in the interval $\mathbf{z}P + m + [y] \setminus [x]$ lie in $\mathbf{z}P + m + \mathcal{T}$.

From (10) and Mertens' theorem we have

$$\mathbb{P}(\mathbf{z}P + m + a \text{ prime}) \gg \frac{\log x}{x}$$

for all $a \in \mathcal{T}$ (we allow implied constants to depend on D). Similarly, from (11) and Mertens' theorem we have

$$\mathbb{P}(\mathbf{z}P + m + a, \mathbf{z}P(x) + m + b \text{ both prime}) \ll \left(\frac{\log x}{x}\right)^2 \quad (13)$$

for any distinct $a, b \in \mathcal{T}$. If we let \mathbf{N} denote the number of primes in $\mathbf{z}P + m + \mathcal{T}$ (or equivalently, in $\mathbf{z}P + m + [y] \setminus [x]$), we thus have from (6) and (7) that

$$\mathbb{E}\mathbf{N} \gg A$$

and

$$\mathbb{E}\mathbf{N}^2 \ll A^2.$$

From this we see that with probability $\gg 1$, we have

$$A \ll \mathbf{N} \ll A, \tag{14}$$

where all implied constants are independent of ε and A . (This is because the contribution to $\mathbb{E}\mathbf{N}$ when \mathbf{N} is much larger than A is much smaller than A .)

Next, if $0 \leq \alpha \leq \beta \leq 1$ and $\beta - \alpha \leq 2\varepsilon$, then from (13), (8) and the union bound we see that the probability that there are at least two primes in $\mathbf{z}P + m + [\alpha y, \beta y]$ is at most

$$O\left(\left(A\varepsilon \frac{x}{\log x}\right)^2 \left(\frac{\log x}{x}\right)^2\right) = O(A^2 \varepsilon^2).$$

Note that one can cover $[0, 1]$ with $O(1/\varepsilon)$ intervals of length at most 2ε , with the property that any two elements a, b of $[0, 1]$ with $|a - b| \leq \varepsilon$ may be covered by at least one of these intervals. From this and the union bound, we see that the probability that $\mathbf{z}P + m + [y] \setminus [x]$ contains two primes separated by at most εy is bounded by $O(\frac{1}{\varepsilon} A^2 \varepsilon^2) = O(A^2 \varepsilon)$. In particular, if we choose ε to be a sufficiently small multiple of $\frac{1}{A^2}$, we may find $z \in [Z]$ such that the interval $\mathbf{z}P + m + [y] \setminus [x]$ contains $\gg A$ primes and has no prime gap less than εy . If we choose A to be a sufficiently large multiple of k , we conclude that

$$G_k(\mathbf{z}P + m + y) \geq \varepsilon y \gg \frac{1}{k^2} y.$$

By Mertens' theorem, we have $\mathbf{z}P + m + y \ll \exp(O(x))$, and Theorem 1 then follows from (5).

It remains to prove Theorem 2. This is the objective of the remaining sections of the paper.

4 Sieving a set of primes

Theorem 2 concerns the problem of deterministically sieving an interval $[y] \setminus [x]$ of size (5) so that the sifted set \mathcal{T} has certain size properties. We use a variant of the Erdős-Rankin method to reduce this problem to a problem of *probabilistically* sieving a set \mathcal{Q} of *primes* in $[y] \setminus [x]$, rather than integers in $[y] \setminus [x]$.

Given a real number $x \geq 1$, and a natural number B_0 , define

$$z := x^{\log_3 x / (4 \log_2 x)}, \quad (15)$$

and introduce the three disjoint sets of primes

$$\mathcal{S} := \{s \text{ prime} : \log^{20} x < s \leq z; s \neq B_0\}, \quad (16)$$

$$\mathcal{P} := \{p \text{ prime} : x/2 < p \leq x; p \neq B_0\}, \quad (17)$$

$$\mathcal{Q} := \{q \text{ prime} : x < q \leq y; q \neq B_0\}. \quad (18)$$

For residue classes $\mathbf{a} = (a_s \bmod s)_{s \in \mathcal{S}}$ and $\mathbf{n} = (n_p \bmod p)_{p \in \mathcal{P}}$, define the sifted sets

$$S(\mathbf{a}) := \{n \in \mathbb{Z} : n \not\equiv a_s \pmod{s} \text{ for all } s \in \mathcal{S}\}$$

and likewise

$$S(\mathbf{n}) := \{n \in \mathbb{Z} : n \not\equiv n_p \pmod{p} \text{ for all } p \in \mathcal{P}\}.$$

We reduce Theorem 2 to

Theorem 3 (Sieving primes) *Let $A \geq 1$ be a real number, let x be sufficiently large depending on A , and suppose that y obeys (5). Let B_0 be a natural number. Then there is a quantity*

$$A' \asymp A, \quad (19)$$

and some way to choose the vectors $\mathbf{a} = (a_s \bmod s)_{s \in \mathcal{S}}$ and $\mathbf{n} = (n_p \bmod p)_{p \in \mathcal{P}}$ at random (not necessarily independent of each other), such that for any fixed $0 \leq \alpha < \beta \leq 1$ (independent of x), one has with probability $1 - o(1)$ that

$$\#(\mathcal{Q} \cap S(\mathbf{a}) \cap S(\mathbf{n}) \cap (\alpha y, \beta y]) \sim A' |\beta - \alpha| \frac{x}{\log x}. \quad (20)$$

The $o(1)$ decay rates in the probability error and implied in the \sim notation are allowed to depend on A, α, β .

In [11, Theorem 2], a weaker version of this theorem was established in which B_0 was not present, and only the upper bound in (20) was proven. Thus, the main new contribution of this paper is the lower bound in (20).

We prove Theorem 3 in subsequent sections. In this section, we show how this theorem implies Theorem 2 (and hence Theorem 1). The arguments here are almost identical to those in [11, §2].

Fix $A \geq 1, 0 < \varepsilon \leq 1$. We partition $(0, 1]$ into $O(1/\varepsilon)$ intervals $[\alpha_i, \beta_i]$ of length between $\varepsilon/2$ and ε . Applying Theorem 3 with the pairs $(\alpha, \beta) = (\alpha_i, \beta_i)$ and the pair $(\alpha, \beta) = (0, 1)$, and invoking a union bound (and the fact that ε is independent of x), we see that if x is sufficiently large (depending on A, ε), there are A', y obeying (19), (5) and tuples of residue classes $\mathbf{a} = (a_s \bmod s)_{s \in \mathcal{S}}$ and $\mathbf{n} = (n_p \bmod p)_{p \in \mathcal{P}}$ such that

$$\#(\mathcal{Q} \cap S(\mathbf{a}) \cap S(\mathbf{n})) \sim A' \frac{x}{\log x}$$

and

$$\#(\mathcal{Q} \cap S(\mathbf{a}) \cap S(\mathbf{n})) \cap (\alpha_i y, \beta_i y] \ll A \varepsilon \frac{x}{\log x}$$

for all i . A covering argument then gives

$$\#(\mathcal{Q} \cap S(\mathbf{a}) \cap S(\mathbf{n}) \cap [\alpha y, \beta y]) \ll A(|\beta - \alpha| + \varepsilon) \frac{x}{\log x}$$

for any $0 \leq \alpha < \beta \leq 1$. Now we extend the tuple \mathbf{a} to a tuple $(a_p)_{p \leq x}$ of congruence classes $a_p \bmod p$ for all primes $p \leq x$ by setting $a_p := n_p$ for $p \in \mathcal{P}$ and $a_p := 0$ for $p \notin \mathcal{S} \cup \mathcal{P}$, and consider the sifted set

$$\mathcal{T} := \{n \in [y] \setminus [x] : n \not\equiv a_p \pmod{p} \text{ for all } p \leq x\}.$$

The elements of \mathcal{T} , by construction, are not divisible by any prime in $(0, \log^{20} x]$ or in $(z, x/2]$, except possibly for B_0 . Thus, each element must either be a z -smooth number (i.e. a number with all prime factors at most z) times a power of B_0 , or must consist of a prime greater than $x/2$, possibly multiplied by some additional primes that are all either at least $\log^{20} x$ or equal to B_0 . However, from (5) we know that $y = o(x \log x)$, and by hypothesis we know that $B_0 \gg \log x$. Thus, we see that an element of \mathcal{T} is either a z -smooth number times a power of B_0 or a prime in \mathcal{Q} . In the second case, the element lies in $\mathcal{Q} \cap S(\mathbf{a}) \cap S(\mathbf{n})$. Conversely, every element of $\mathcal{Q} \cap S(\mathbf{a}) \cap S(\mathbf{n})$ lies in \mathcal{T} . Thus, \mathcal{T} only differs from $\mathcal{Q} \cap S(\mathbf{a}) \cap S(\mathbf{n})$ by a set \mathcal{R} consisting of z -smooth numbers in $[y]$ multiplied by powers of B_0 .

To estimate $\#\mathcal{R}$, let

$$u := \frac{\log y}{\log z},$$

so from (5), (15) one has $u \sim 4 \frac{\log_2 x}{\log_3 x}$. The number of powers of B_0 in $[y]$ is $O(\log x)$. By standard counts for smooth numbers (e.g. de Bruijn's theorem [5]) and (5), we thus have

$$\begin{aligned} \#\mathcal{R} &\ll \log x \times y e^{-u \log u + O(u \log \log(u+2))} \\ &= \log x \times \frac{y}{\log^{4+o(1)} x} = o\left(\frac{x}{\log x}\right). \end{aligned}$$

Thus the contribution of \mathcal{R} to \mathcal{T} is negligible for the purposes of establishing the bounds (6), (7), (8), and Theorem 2 follows from (20).

It remains to establish Theorem 3. This is the objective of the remaining sections of the paper.

5 Using a hypergraph covering theorem

In the previous section we reduced matters to obtaining random residue classes \mathbf{a} , \mathbf{n} such that the sifted set $\mathcal{Q} \cap S(\mathbf{a}) \cap S(\mathbf{n})$ is small. In this section we use a hypergraph

covering theorem from [11] to reduce the task to that of finding random residue classes \mathbf{n} that have large intersection with $\mathcal{Q} \cap S(\mathbf{a})$. More precisely, we will use the following result:

Theorem 4 *Let $x \rightarrow \infty$. Let \mathcal{P}' , \mathcal{Q}' be sets of primes in $(x/2, x]$ and $(x, x \log x]$, respectively, with $\#\mathcal{Q}' > (\log_2 x)^3$. For each $p \in \mathcal{P}'$, let \mathbf{e}_p be a random subset of \mathcal{Q}' satisfying the size bound*

$$\#\mathbf{e}_p \leq r = O\left(\frac{\log x \log_3 x}{\log_2^2 x}\right) \quad (p \in \mathcal{P}'). \quad (21)$$

Assume the following:

- (Sparsity) For all $p \in \mathcal{P}'$ and $q \in \mathcal{Q}'$,

$$\mathbb{P}(q \in \mathbf{e}_p) \leq x^{-1/2-1/10}. \quad (22)$$

- (Uniform covering) For all but at most $\frac{1}{(\log_2 x)^2} \#\mathcal{Q}'$ elements $q \in \mathcal{Q}'$, we have

$$\sum_{p \in \mathcal{P}'} \mathbb{P}(q \in \mathbf{e}_p) = C + O\left(\frac{1}{(\log_2 x)^2}\right) \quad (23)$$

for some quantity C , independent of q , satisfying

$$\frac{5}{4} \log 5 \leq C \ll 1. \quad (24)$$

- (Small codegrees) For any distinct $q_1, q_2 \in \mathcal{Q}'$,

$$\sum_{p \in \mathcal{P}'} \mathbb{P}(q_1, q_2 \in \mathbf{e}_p) \leq x^{-1/20}. \quad (25)$$

Then for any positive integer m with

$$m \leq \frac{\log_3 x}{\log 5}, \quad (26)$$

we can find random sets $\mathbf{e}'_p \subseteq \mathcal{Q}'$ for each $p \in \mathcal{P}'$ such that

$$\#\{q \in \mathcal{Q}' : q \notin \mathbf{e}'_p \text{ for all } p \in \mathcal{P}'\} \sim 5^{-m} \#\mathcal{Q}'$$

with probability $1 - o(1)$. More generally, for any $\mathcal{Q}'' \subset \mathcal{Q}'$ with cardinality at least $(\#\mathcal{Q}')/\sqrt{\log_2 x}$, one has

$$\#\{q \in \mathcal{Q}'' : q \notin \mathbf{e}'_p \text{ for all } p \in \mathcal{P}'\} \sim 5^{-m} \#\mathcal{Q}''$$

with probability $1 - o(1)$. The decay rates in the $o(1)$ and \sim notation are uniform in \mathcal{P}' , \mathcal{Q}' , \mathcal{Q}'' .

Proof. See [11, Corollary 3].

In view of the above result, we may now reduce Theorem 3 to the following claim.

Theorem 5 (Random construction) *Let x be a sufficiently large real number, let B_0 be a natural number and suppose y satisfies (5). Then there is a quantity C with*

$$C \asymp \frac{1}{c} \quad (27)$$

with the implied constants independent of c , and some way to choose random vectors $\mathbf{a} = (\mathbf{a}_s \bmod s)_{s \in \mathcal{S}}$ and $\mathbf{n} = (\mathbf{n}_p)_{p \in \mathcal{P}}$ of congruence classes $\mathbf{a}_s \bmod s$ and integers \mathbf{n}_p , obeying the following axioms:

- For every \mathbf{a} in the essential range of \mathbf{a} , one has

$$\mathbb{P}(q \equiv \mathbf{n}_p \pmod{p} | \mathbf{a} = \mathbf{a}) \leq x^{-1/2-1/10}$$

uniformly for all $p \in \mathcal{P}$.

- For fixed $0 \leq \alpha < \beta \leq 1$, we have with probability $1 - o(1)$ that

$$\#(\mathcal{Q} \cap S(\mathbf{a}) \cap [\alpha y, \beta y]) \sim 80c|\beta - \alpha| \frac{x}{\log x} \log_2 x. \quad (28)$$

- Call an element \mathbf{a} in the essential range of \mathbf{a} good if, for all but at most $\frac{x}{\log x \log_2 x}$ elements $q \in \mathcal{Q} \cap S(\mathbf{a})$, one has

$$\sum_{p \in \mathcal{P}} \mathbb{P}(q \equiv \mathbf{n}_p \pmod{p} | \mathbf{a} = \mathbf{a}) = C + O_{\leq} \left(\frac{1}{(\log_2 x)^2} \right). \quad (29)$$

Then \mathbf{a} is good with probability $1 - o(1)$.

We now show why Theorem 5 implies Theorem 3. By (27), we may choose $0 < c < 1/2$ small enough so that (24) holds. Let $A \geq 1$ be a fixed quantity. Then we can find an integer m obeying (26) such that the quantity

$$A' := 5^{-m} \times 80c \log_2 x$$

is such that $A' \asymp A$ with implied constants independent of A .

Suppose that we are in the probability $1 - o(1)$ event that \mathbf{a} takes a value \mathbf{a} which is good and such that (28) holds. On each sub-event $\mathbf{a} = \mathbf{a}$ of this probability $1 - o(1)$ event, we may apply Theorem 4 (for the random variables \mathbf{n}_p conditioned to this event) define the random variables \mathbf{n}'_p on this event with the stated properties. For the remaining events $\mathbf{a} = \mathbf{a}$, we set \mathbf{n}'_p arbitrarily (e.g. we could set $\mathbf{n}'_p = 0$). The claim (20) then follows from Theorem 4 and (28), thus establishing Theorem 3.

It remains to establish Theorem 5. This will be achieved in the next section.

6 Using a sieve weight

If r is a natural number, an *admissible r -tuple* is a tuple (h_1, \dots, h_r) of distinct integers h_1, \dots, h_r that do not cover all residue classes modulo p , for any prime p . For instance, the tuple $(p_{\pi(r)+1}, \dots, p_{\pi(r)+r})$ consisting of the first r primes larger than r is an admissible r -tuple.

We will establish Theorem 5 by a probabilistic argument involving a certain weight function. More precisely, we will deduce this result from the following construction from [11].

Theorem 6 (Existence of good sieve weight) *Let x be a sufficiently large real number, let B_0 be an integer, and let y be any quantity obeying (5). Let \mathcal{P}, \mathcal{Q} be defined by (17), (18). Let r be a positive integer with*

$$r_0 \leq r \leq \log^{c_0} x \quad (30)$$

for some sufficiently small absolute constant c_0 and sufficiently large absolute constant r_0 , and let (h_1, \dots, h_r) be an admissible r -tuple contained in $[2r^2]$. Then one can find a positive quantity

$$\tau \geq x^{-o(1)} \quad (31)$$

and a positive quantity $u = u(r)$ depending only on r with

$$u \asymp \log r \quad (32)$$

and a non-negative function $w : \mathcal{P} \times \mathbb{Z} \rightarrow \mathbb{R}^+$ supported on $\mathcal{P} \times (\mathbb{Z} \cap [-y, y])$ with the following properties:

- *Uniformly for every $p \in \mathcal{P}$, one has*

$$\sum_{n \in \mathbb{Z}} w(p, n) = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \tau \frac{y}{\log^r x}. \quad (33)$$

- *Uniformly for every $q \in \mathcal{Q}$ and $i = 1, \dots, r$, one has*

$$\sum_{p \in \mathcal{P}} w(p, q - h_i p) = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \tau \frac{u}{r} \frac{x}{2 \log^r x}. \quad (34)$$

- *Uniformly for every $h = O(y/x)$ that is not equal to any of the h_i , one has*

$$\sum_{q \in \mathcal{Q}} \sum_{p \in \mathcal{P}} w(p, q - hp) = O\left(\frac{1}{\log_2^{10} x} \tau \frac{x}{\log^r x} \frac{y}{\log x}\right). \quad (35)$$

- *Uniformly for all $p \in \mathcal{P}$ and $n \in \mathbb{Z}$,*

$$w(p, n) = O(x^{1/3+o(1)}). \quad (36)$$

Proof. See² [11, Theorem 5]. We remark that the construction of the weights and the verification of the required estimates relies heavily on the previous work of the second author in [17].

It remains to show how Theorem 6 implies Theorem 5. The analysis will be based on that in [11, §5], which used a weight with slightly weaker hypotheses than in Theorem 6 to obtain somewhat weaker conclusions than Theorem 5 (in which the condition $q \equiv \mathbf{n}_p \pmod{p}$ was replaced by the stronger condition that $q = \mathbf{n}_p + h_i p$ for some $i = 1, \dots, r$).

Let $x, B_0, c, y, z, \mathcal{S}, \mathcal{P}, \mathcal{Q}$ be as in Theorem 5. Let c_0 be a sufficiently small absolute constant. We set r to be the maximum value permitted by Theorem 6, namely

$$r := \lfloor \log^{c_0} x \rfloor \quad (37)$$

and let (h_1, \dots, h_r) be the admissible r -tuple consisting of the first r primes larger than r , thus $h_i = p_{\pi(r)+i}$ for $i = 1, \dots, r$. From the prime number theorem we have $h_i = O(r \log r)$ for $i = 1, \dots, r$, and so we have $h_i \in [2r^2]$ for $i = 1, \dots, r$ if x is large enough (there are many other choices possible, e.g. $(h_1, \dots, h_r) = (1^2, 3^2, \dots, (2r-1)^2)$). We now invoke Theorem 6 to obtain quantities τ, u and a weight $w : \mathcal{P} \times \mathbb{Z} \rightarrow \mathbb{R}^+$ with the stated properties.

For each $p \in \mathcal{P}$, let $\tilde{\mathbf{n}}_p$ denote the random integer with probability density

$$\mathbb{P}(\tilde{\mathbf{n}}_p = n) := \frac{w(p, n)}{\sum_{n' \in \mathbb{Z}} w(p, n')}$$

for all $n \in \mathbb{Z}$ (we will not need to impose any independence conditions on the $\tilde{\mathbf{n}}_p$). From (33), (34) we have

$$\sum_{p \in \mathcal{P}} \mathbb{P}(q = \tilde{\mathbf{n}}_p + h_i p) = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \frac{u}{r} \frac{x}{2y} \quad (38)$$

for every $q \in \mathcal{Q}$ and $i = 1, \dots, r$, and similarly from (33), (35) we have

$$\sum_{q \in \mathcal{Q}} \sum_{p \in \mathcal{P}} \mathbb{P}(q = \tilde{\mathbf{n}}_p + h p) \ll \frac{1}{\log_2^{10} x} \frac{x}{\log x} \quad (39)$$

for every $h = O(y/x)$ not equal to any of the h_i . Finally, from (33), (36), (31) one has

$$\mathbb{P}(\tilde{\mathbf{n}}_p = n) \ll x^{-1/2-1/6+o(1)} \quad (40)$$

for all $p \in \mathcal{P}$ and $n \in \mathbb{Z}$.

We choose the random vector $\mathbf{a} := (\mathbf{a}_s \bmod s)_{s \in \mathcal{S}}$ by selecting each $\mathbf{a}_s \bmod s$ uniformly at random from $\mathbb{Z}/s\mathbb{Z}$, independently in s and independently of the $\tilde{\mathbf{n}}_p$. The resulting sifted set $S(\mathbf{a})$ is a random periodic subset of \mathbb{Z} with density

² The integer B_0 was not deleted from the sets \mathcal{P} or \mathcal{Q} in that theorem, however it is easy to see (using (36)) that deleting at most one prime from either \mathcal{P} or \mathcal{Q} will not significantly worsen any of the estimates claimed by the theorem.

$$\sigma := \prod_{s \in \mathcal{S}} \left(1 - \frac{1}{s}\right).$$

From the prime number theorem (with sufficiently strong error term), (15) and (16),

$$\sigma = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \frac{\log(\log^{20} x)}{\log z} = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) \frac{80 \log_2 x}{\log x \log_3 x / \log_2 x},$$

so in particular we see from (5) that

$$\sigma y = \left(1 + O\left(\frac{1}{\log_2^{10} x}\right)\right) 80cx \log_2 x. \quad (41)$$

We also see from (37) that

$$\sigma^r = x^{o(1)}. \quad (42)$$

We have a useful correlation bound:

Lemma 6.1 *Let $t \leq \log x$ be a natural number, and let n_1, \dots, n_t be distinct integers of magnitude $O(x^{o(1)})$. Then one has*

$$\mathbb{P}(n_1, \dots, n_t \in S(\mathbf{a})) = \left(1 + O\left(\frac{1}{\log^{16} x}\right)\right) \sigma^t.$$

Proof. See [11, Lemma 5.1].

Among other things, this gives the claim (28):

Corollary 2 *For any fixed $0 \leq \alpha < \beta \leq 1$, we have with probability $1 - o(1)$ that*

$$\#(\mathcal{Q} \cap [\alpha y, \beta y] \cap S(\mathbf{a})) \sim \sigma |\beta - \alpha| \frac{y}{\log x} \sim 80c |\beta - \alpha| \frac{x}{\log x} \log_2 x. \quad (43)$$

Proof. See [11, Corollary 4], replacing \mathcal{Q} with $\mathcal{Q} \cap [\alpha y, \beta y]$.

For each $p \in \mathcal{P}$, we consider the quantity

$$X_p(\mathbf{a}) := \mathbb{P}(\tilde{\mathbf{n}}_p + h_i p \in S(\mathbf{a}) \text{ for all } i = 1, \dots, r), \quad (44)$$

and let $\mathcal{P}(\mathbf{a})$ denote the set of all the primes $p \in \mathcal{P}$ such that

$$X_p(\mathbf{a}) = \left(1 + O_{\leq} \left(\frac{1}{\log^3 x}\right)\right) \sigma^r. \quad (45)$$

In light of Lemma 6.1, we expect most primes in \mathcal{P} to lie in $\mathcal{P}(\mathbf{a})$, and this will be confirmed below (Lemma 6.2). We now define the random variables \mathbf{n}_p as follows. Suppose we are in the event $\mathbf{a} = \mathbf{a}$ for some \mathbf{a} in the range of \mathbf{a} . If $p \in \mathcal{P} \setminus \mathcal{P}(\mathbf{a})$, we set $\mathbf{n}_p = 0$. Otherwise, if $p \in \mathcal{P}(\mathbf{a})$, we define \mathbf{n}_p to be the random integer with conditional probability distribution

$$\mathbb{P}(\mathbf{n}_p = n | \mathbf{a} = \mathbf{a}) := \frac{Z_p(\mathbf{a}; n)}{X_p(\mathbf{a})}, \quad Z_p(\mathbf{a}; n) = 1_{n+h_j p \in S(\mathbf{a}) \text{ for } j=1, \dots, r} \mathbb{P}(\tilde{\mathbf{n}}_p = n). \quad (46)$$

with the \mathbf{n}_p jointly conditionally independent on the event $\mathbf{a} = \mathbf{a}$. From (45) we see that these random variables are well defined.

Substituting definition (46) into the left hand side of (29), and observing that $\mathbf{n}_p \equiv q \pmod{p}$ is only possible if $p \in \mathcal{P}(\mathbf{a})$, we see that to prove (29), it suffices to show that with probability $1 - o(1)$ in \mathbf{a} , for all but at most $\frac{x}{\log x \log_2 x}$ primes in $\mathcal{Q} \cap S(\mathbf{a})$, we have

$$\sigma^{-r} \sum_{p \in \mathcal{P}(\mathbf{a})} \sum_h Z_p(\mathbf{a}; q - hp) = C + O\left(\frac{1}{\log_2^3 x}\right). \quad (47)$$

We now confirm that $\mathcal{P} \setminus \mathcal{P}(\mathbf{a})$ is small with high probability.

Lemma 6.2 *With probability $1 - O(1/\log^3 x)$, $\mathcal{P}(\mathbf{a})$ contains all but $O(\frac{1}{\log^3 x} \frac{x}{\log x})$ of the primes $p \in \mathcal{P}$. In particular, $\mathbb{E} \# \mathcal{P}(\mathbf{a}) = \# \mathcal{P}(1 + O(1/\log^3 x))$.*

Proof. See [11, Lemma 5.3].

The left side of relation (47) breaks naturally into two pieces, a ‘main term’ consisting of summands where $h = h_i$ for some i , and an ‘error terms’ consisting of the remaining summands. We first take care of the error terms.

Lemma 6.3 *With probability $1 - o(1)$ we have*

$$\sigma^{-r} \sum_{p \in \mathcal{P}(\mathbf{a})} \sum_{\substack{h \ll y/x \\ h \notin \{h_1, \dots, h_r\}}} Z_p(\mathbf{a}; q - hp) \ll \frac{1}{\log_2^3 x} \quad (48)$$

for all but at most $\frac{x}{2 \log x \log_2 x}$ primes $q \in \mathcal{Q} \cap S(\mathbf{a})$.

Proof. We first extend the sum over all $p \in \mathcal{P}$. By Markov’s inequality, it suffices to show that

$$\mathbb{E} \sum_{q \in \mathcal{Q} \cap S(\mathbf{a})} \sigma^{-r} \sum_{p \in \mathcal{P}} \sum_{\substack{h \ll y/x \\ h \notin \{h_1, \dots, h_k\}}} Z_p(\mathbf{a}; q - hp) = o\left(\frac{x}{\log x \log_2^4 x}\right). \quad (49)$$

The left-hand side of (49) equals

$$\sigma^{-r} \sum_{q \in \mathcal{Q}} \sum_{\substack{h \ll y/x \\ h \notin \{h_1, \dots, h_k\}}} \sum_{p \in \mathcal{P}} \mathbb{P}(q \in S(\mathbf{a}), q + h_j p - hp \in S(\mathbf{a}) \text{ for } j=1, \dots, r) \mathbb{P}(q = \tilde{\mathbf{n}}_p + hp).$$

We note that for any h in the above sum, the $r+1$ integers $q, q + h_1 p - hp, \dots, q + h_r p - hp$ are distinct. Applying Lemma 6.1, followed by (39), we may thus bound this expression by

$$\ll \sum_{\substack{h \ll y/x \\ h \notin \{h_1, \dots, h_k\}}} \sigma \frac{x/\log x}{\log_2^{10} x} \ll \sigma \frac{1}{\log_2^{10} x} \frac{y}{\log x}.$$

The claim now follows from (41).

Next, we deal with the main term of (47), by showing an analogue of (38).

Lemma 6.4 *With probability $1 - o(1)$, we have*

$$\sigma^{-r} \sum_{i=1}^r \sum_{p \in \mathcal{P}(\mathbf{a})} Z_p(\mathbf{a}; q - h_i p) = \left(1 + O\left(\frac{1}{\log_2^3 x}\right)\right) \frac{u}{\sigma} \frac{x}{2y} \quad (50)$$

for all but at most $\frac{x}{2\log x \log_2 x}$ of the primes $q \in \mathcal{Q} \cap S(\mathbf{a})$.

Proof. We first show that replacing $\mathcal{P}(\mathbf{a})$ with \mathcal{P} has negligible effect on the sum, with probability $1 - o(1)$. Fix i and substitute $n = q - h_i p$. By Markov's inequality, it suffices to show that

$$\mathbb{E} \sum_n \sigma^{-r} \sum_{p \in \mathcal{P} \setminus \mathcal{P}(\mathbf{a})} Z_p(\mathbf{a}; n) = o\left(\frac{u}{\sigma} \frac{x}{2y} \frac{1}{r} \frac{1}{\log_2^3 x} \frac{x}{\log x \log_2 x}\right). \quad (51)$$

By Lemma 6.1, we have

$$\begin{aligned} \mathbb{E} \sum_n \sigma^{-r} \sum_{p \in \mathcal{P}} Z_p(\mathbf{a}; n) &= \sigma^{-r} \sum_{p \in \mathcal{P}} \sum_n \mathbb{P}(\mathbf{n}_p = n) \mathbb{P}(n + h_j p \in S(\mathbf{a}) \text{ for } j = 1, \dots, r) \\ &= \left(1 + O\left(\frac{1}{\log^{16} x}\right)\right) \#\mathcal{P}. \end{aligned}$$

Next, by (45) and Lemma 6.2 we have

$$\begin{aligned} \mathbb{E} \sum_n \sigma^{-r} \sum_{p \in \mathcal{P}(\mathbf{a})} Z_p(\mathbf{a}; n) &= \sigma^{-r} \sum_{\mathbf{a}} \mathbb{P}(\mathbf{a} = \mathbf{a}) \sum_{p \in \mathcal{P}(\mathbf{a})} X_p(\mathbf{a}) \\ &= \left(1 + O\left(\frac{1}{\log^3 x}\right)\right) \mathbb{E} \#\mathcal{P}(\mathbf{a}) = \left(1 + O\left(\frac{1}{\log^3 x}\right)\right) \#\mathcal{P}; \end{aligned}$$

subtracting, we conclude that the left-hand side of (51) is $O(\#\mathcal{P}/\log^3 x) = O(x/\log^4 x)$. The claim then follows from (5) and (30).

By (51), it suffices to show that with probability $1 - o(1)$, for all but at most $\frac{x}{2\log x \log_2 x}$ primes $q \in \mathcal{Q} \cap S(\mathbf{a})$, one has

$$\sum_{i=1}^r \sum_{p \in \mathcal{P}} Z_p(\mathbf{a}; q - h_i p) = \left(1 + O_{\leq}\left(\frac{1}{\log_2^3 x}\right)\right) \sigma^{r-1} u \frac{x}{2y}. \quad (52)$$

Call a prime $q \in \mathcal{Q}$ *bad* if $q \in \mathcal{Q} \cap S(\mathbf{a})$ but (52) fails. Using Lemma 6.1 and (38), we have

$$\begin{aligned}
& \mathbb{E} \left[\sum_{q \in \mathcal{Q} \cap S(\mathbf{a})} \sum_{i=1}^r \sum_{p \in \mathcal{P}} Z_p(\mathbf{a}; q - h_i p) \right] \\
&= \sum_{q, i, p} \mathbb{P}(q + (h_j - h_i)p \in S(\mathbf{a}) \text{ for all } j = 1, \dots, r) \mathbb{P}(\tilde{\mathbf{n}}_p = q - h_i p) \\
&= \left(1 + O\left(\frac{1}{\log_2^{10} x}\right) \right) \frac{\sigma y}{\log x} \sigma^{r-1} u \frac{x}{2y}
\end{aligned}$$

and

$$\begin{aligned}
& \mathbb{E} \left[\sum_{q \in \mathcal{Q} \cap S(\mathbf{a})} \left(\sum_{i=1}^r \sum_{p \in \mathcal{P}} Z_p(\mathbf{a}; q - h_i p) \right)^2 \right] \\
&= \sum_{\substack{p_1, p_2, q \\ i_1, i_2}} \mathbb{P}(q + (h_{j_1} - h_{i_1})p_1 \in S(\mathbf{a}) \text{ for } j_1 = 1, \dots, r; \ell = 1, 2) \\
&\quad \times \mathbb{P}(\tilde{\mathbf{n}}_{p_1}^{(1)} = q - h_{i_1} p_1) \mathbb{P}(\tilde{\mathbf{n}}_{p_2}^{(2)} = q - h_{i_2} p_2) \\
&= \left(1 + O\left(\frac{1}{\log_2^{10} x}\right) \right) \frac{\sigma y}{\log x} \left(\sigma^{r-1} u \frac{x}{2y} \right)^2,
\end{aligned}$$

where $(\tilde{\mathbf{n}}_{p_1}^{(1)})_{p_1 \in \mathcal{P}}$ and $(\tilde{\mathbf{n}}_{p_2}^{(2)})_{p_2 \in \mathcal{P}}$ are independent copies of $(\tilde{\mathbf{n}}_p)_{p \in \mathcal{P}}$ over \mathbf{a} . In the last step we used the fact that the terms with $p_1 = p_2$ contribute negligibly.

By Chebyshev's inequality (Lemma 1.1) it follows that the number of bad q is $\ll \frac{\sigma y}{\log x} \frac{1}{\log_2^3 x} \ll \frac{x}{\log x \log_2^2 x}$ with probability $1 - O(1/\log_2 x)$. This concludes the proof.

We now conclude the proof of Theorem 5. We need to prove (47); this follows immediately from Lemma 6.3 and Lemma 6.4 upon noting that by (37), (32) and (41),

$$C := \frac{u}{\sigma} \frac{x}{2y} \sim \frac{1}{c}.$$

References

1. R. J. Backlund, *Über die Differenzen zwischen den Zahlen, die zu den ersten n Primzahlen teilerfremd sind*, Commentationes in honorem E. L. Lindelöf. Annales Acad. Sci. Fenn. **32** (1929), Nr. 2, 1–9.
2. R. C. Baker, T. Freiberg, *Limit points and long gaps between primes*, preprint.
3. R. C. Baker, G. Harman and J. Pintz, *The difference between consecutive primes. II.*, Proc. London Math. Soc. (3) **83** (2001), no. 3, 532–562.
4. A. Brauer, H. Zeitz, *Über eine zahlentheoretische Behauptung von Legendre*, Sber. Berliner Math. Ges. **29** (1930), 116–125.
5. N. G. de Bruijn, *On the number of positive integers $\leq x$ and free of prime factors $> y$* , Nederl. Acad. Wetensch. Proc. Ser. A. **54** (1951) 50–60.

6. H. Cramér, *Some theorems concerning prime numbers*, Ark. Mat. Astr. Fys. **15** (1920), 1–33.
7. H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arith. **2** (1936), 396–403.
8. H. Davenport, *Multiplicative number theory*, 3rd ed., Graduate Texts in Mathematics vol. 74, Springer-Verlag, New York, 2000.
9. P. Erdős, *On the difference of consecutive primes*, Quart. J. Math. Oxford Ser. **6** (1935), 124–128.
10. K. Ford, B. Green, S. Konyagin, T. Tao, *Large gaps between consecutive prime numbers*, Ann. Math. **183** (2016), 935–974.
11. K. Ford, B. Green, S. Konyagin, J. Maynard, T. Tao, *Long gaps between primes*, preprint.
12. P. X. Gallagher, *A large sieve density estimate near $\sigma = 1$* , Invent. Math. **11** (1970), 329–339.
13. A. Granville, *Harald Cramér and the distribution of prime numbers*, Scandanavian Actuarial J. **1** (1995), 12–28.
14. H. Maier, *Chains of large gaps between consecutive primes*, Advances in Mathematics **39** (1981), 257–269.
15. H. Maier and C. Pomerance, *Unusually large gaps between consecutive primes*. Trans. Amer. Math. Soc. **322** (1990), no. 1, 201–237.
16. J. Maynard, *Small gaps between primes*, Ann. of Math. (2) **181** (2015), no. 1, 383–413.
17. J. Maynard, *Dense clusters of primes in subsets*, preprint.
18. J. Maynard, *Large gaps between primes*, Ann. Math. **183** (2016), 915–933.
19. J. Pintz, *On the distribution of gaps between consecutive primes*, preprint.
20. J. Pintz, *Very large gaps between consecutive primes*. J. Number Theory **63** (1997), no. 2, 286–301.
21. N. Pippenger, J. Spencer, *Asymptotic behavior of the chromatic index for hypergraphs*, J. Combin. Theory Ser. A **51** (1989), no. 1, 24–42.
22. R. A. Rankin, *The difference between consecutive prime numbers*, J. London Math. Soc. **13** (1938), 242–247.
23. R. A. Rankin, *The difference between consecutive prime numbers. V*, Proc. Edinburgh Math. Soc. (2) **13** (1962/63), 331–332.
24. A. Schönhage, *Eine Bemerkung zur Konstruktion grosser Primzahlücken*, Arch. Math. **14** (1963), 29–30.
25. E. Westzynthius, *Über die Verteilung der Zahlen, die zu den n ersten Primzahlen teilerfremd sind*, Commentationes Physico-Mathematicae, Societas Scientiarum Fennica, Helsingfors **5**, no. 25, (1931) 1–37.