LARGE PRIME GAPS AND PROBABILISTIC MODELS

WILLIAM BANKS, KEVIN FORD, AND TERENCE TAO

Abstract. We introduce a new probabilistic model of the primes consisting of integers that survive the sieving process when a random residue class is selected for every prime modulus below a specific bound. From a rigorous analysis of this model, we obtain heuristic upper and lower bounds for the size of the largest prime gap in the interval $[1, x]$. Our results are stated in terms of the extremal bounds in the interval sieve problem. The same methods also allow us to rigorously relate the validity of the Hardy-Littlewood conjectures for an arbitrary set (such as the actual primes) to lower bounds for the largest gaps within that set.

1. Introduction

In this paper, we introduce a new probabilistic model $\mathcal{R} \subset \mathbb{N}$ for the primes $\mathcal{P} := \{2, 3, 5, \ldots\}$ which can be analyzed rigorously to make a variety of heuristic predictions. In contrast to the well known prime model $\mathcal{C}$ of Cramér [6] and the subsequent refinement $\mathcal{G}$ of Granville [16], in which random sets are formed by including positive integers with specific probabilities, the model $\mathcal{R}$ proposed here is comprised of integers that survive the sieve when a random residue class is selected for every prime modulus below a specific bound. We determine the asymptotic behavior of the largest gap function, $G_{\mathcal{R}}(x)$, for the set $\mathcal{R}$, where for any subset $\mathcal{A} \subset \mathbb{N}$ we denote

$$G_{\mathcal{A}}(x) := \max\{b - a : [a, b] \subset [1, x] \text{ and } [a, b] \cap \mathcal{A} = \varnothing\}.$$

We conjecture that the primes $\mathcal{P}$ have similar behavior. Our bounds, given in Theorem 1.1 below, are stated in terms of the extremal bounds in the interval sieve problem.

Date: May 3, 2023.

MSC Primary: 11N05; 11B83.

Keywords: primes, prime gaps, Hardy-Littlewood conjectures, probabilistic models, sieves.

Acknowledgements: The first author was supported by a grant from the University of Missouri Research Board. The second author was supported by National Science Foundation grants DMS-1501982 and DMS-1802139. The third author 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-1764034. Part of the work was accomplished during a visit of the the second author to the University of Missouri (August 2017), UCLA (November 2018), the University of Cambridge (March 2019), the University of Oxford (March-June, 2019) and the Institute of Mathematics of the Bulgarian Academy of Sciences (June-July, 2019). He thanks them all for their support.

At present, the strongest unconditional lower bound on $G_{\mathcal{P}}(x)$ is due to Ford, Green, Konyagin, Maynard, and Tao [11], who have shown that[1]

$$G_{\mathcal{P}}(x) \gg \frac{\log x \, \log_2 x \, \log_4 x}{\log_3 x},$$

for sufficiently large x, with $\log_k x$ the k-fold iterated natural logarithm of x, whereas the strongest unconditional upper bound is

$$G_{\mathcal{P}}(x) \ll x^{0.525},$$

a result due to Baker, Harman, and Pintz [2]. Assuming the Riemann Hypothesis, Cramér [5] showed that

$$G_{\mathcal{P}}(x) \ll x^{1/2} \log x.$$

1.1. **Cramér's random model.** In 1936, Cramér [6] introduced a probabilistic model $\mathcal{C}$ of primes, where each natural number $n \geqslant 3$ is selected for inclusion in $\mathcal{C}$ with probability $1/\log n$, the events $n \in \mathcal{C}$ being jointly independent in n. By Hoeffding's inequality (or Lemma 3.3 below), for any fixed $\varepsilon > 0$ one has

$$\pi_{\mathcal{C}}(x) := |\{n \in \mathcal{C} : n \leqslant x\}| = \int_2^x \frac{dt}{\log t} + O(x^{1/2+\varepsilon}) \tag{1.1}$$

with probability one.[2] The analogous statement for primes is equivalent to the Riemann Hypothesis. In 1936, Cramér [6] proved that $\limsup_{x\to\infty} \frac{G_{\mathcal{C}}(x)}{\log^2 x} = 1$ almost surely, and remarked: "Obviously we may take this as a suggestion that, for the particular sequence of ordinary prime numbers p_n, some similar relation may hold." Later, Shanks [40] conjectured the stronger bound $G_{\mathcal{P}}(x) \sim \log^2 x$, also based on the analysis of a random model very similar to Cramér's model. This is a natural conjecture in light of the fact that

$$G_{\mathcal{C}}(x) \sim \log^2 x \tag{1.2}$$

holds with probability one (although (1.2) doesn't appear to have been observed before). In the literature, the statements $G_{\mathcal{P}}(x) = O(\log^2 x)$ and $G_{\mathcal{P}}(x) \asymp \log^2 x$ are sometimes referred to as "Cramér's conjecture." Several people have made refined conjectures, e.g., Cadwell [4] has suggested that $G_{\mathcal{P}}(x)$ is well-approximated by $(\log x)(\log x - \log_2 x)$, a conjecture which is strongly supported by numerical calculations of gaps. We refer the reader to Granville [16] or Soundararajan [41] for additional information about the Cramer model and subsequent developments.

Tables of prime gaps have been computed up to 10^{18} and beyond (see [35]), thus

$$\sup_{x \leqslant 10^{18}} \frac{G_{\mathcal{P}}(x)}{\log^2 x} \approx 0.9206,$$

a consequence of the gap of size 1132 following the prime 1693182318746371. See also Figure 1 for a plot of $G(x)$ versus various approximations.

[1] See Section 3.1 for the asymptotic notation used in this paper.

[2] See also [6, eq. (5)] for a more precise version of (1.1).

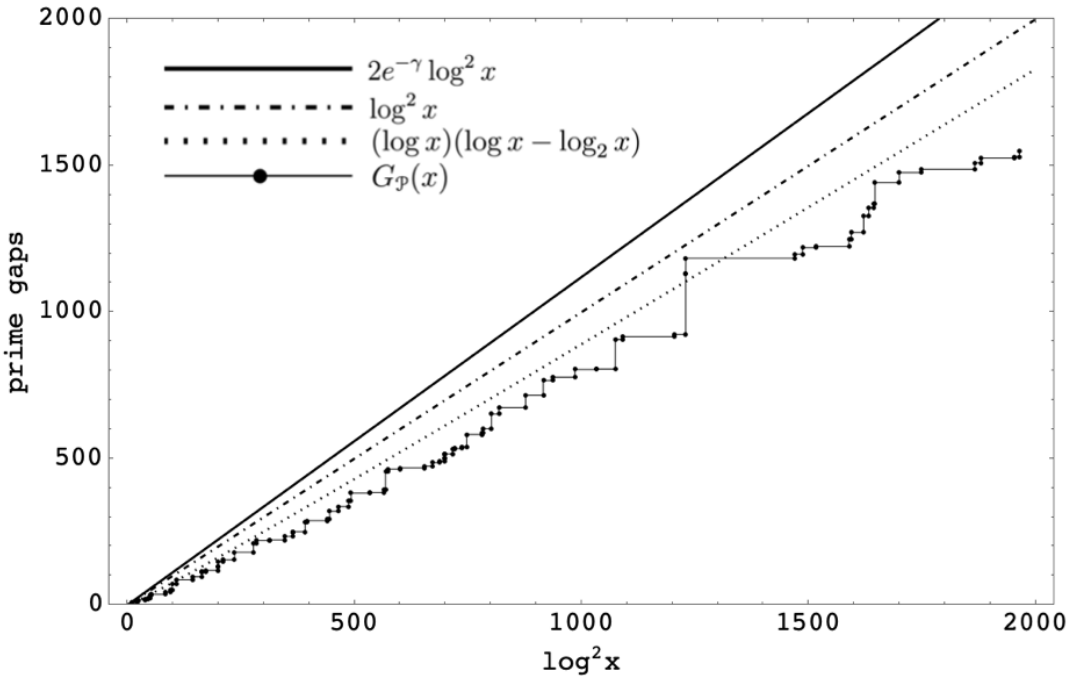

FIGURE 1. $G_{\mathcal{P}}(x)$ vs. various approximations

Despite its utility, the Cramér model has several well-documented weaknesses, the most dramatic one being that the model does not predict the expected asymptotics for prime k-tuples. Indeed, for *any* finite set $\mathcal{H} \subset \mathbb{Z}$, Cramér's model gives

$$|\{n \leqslant x : n + h \in \mathcal{C} \text{ for all } h \in \mathcal{H}\}| \sim \frac{x}{\log^{|\mathcal{H}|} x} \qquad (x \to \infty)$$

with probability one, whereas the analogous assertion for prime numbers is false in general (for example, there is no integer n such that $n + h$ is prime for all $h \in \{0, 1, 2\}$). The reason for the disparity is simple: for any prime p, every prime other than p must lie in one of the residue classes $\{1, \dots, p-1\}$ modulo p (we refer to this as the *bias* of the primes modulo p), whereas $\mathcal{C}$ is equidistributed over *all* residue classes modulo p.

See Pintz [36] and Section 2.5 below, for further discussion of flaws in the Cramér model.

1.2. **Granville's random model.** To correct this flaw in the Cramér model $\mathcal{C}$, Granville [16] altered the model, constructing a random set $\mathcal{G}$ as follows. For each interval $(x, 2x]$ (with x being a power of two, say), let A be a parameter such that $A = \log^{1-o(1)} x$ as $x \to \infty$, and put $Q := \prod_{p \leqslant A} p$. Discard those n for which $(n, Q) > 1$, and select for inclusion in $\mathcal{G}$ each of the remaining integers $n \in (x, 2x]$ with probability $\frac{Q/\phi(Q)}{\log n}$, where ϕ is the Euler totient function, the events $n \in \mathcal{G}$ being jointly independent in n. Since $\phi(Q)/Q$ is the density in $\mathbb{Z}$ of the set of integers coprime to Q, this model captures the correct global distribution of primes; that is, an analog of (1.1) holds with $\mathcal{C}$ replaced by $\mathcal{G}$. Unlike Cramér's model, however, Granville's model also captures the bias of

primes in residue classes modulo the primes $p \leqslant A$. In particular, for any finite set $\mathcal{H}$ of integers, Granville's set satisfies the appropriate analog of the Hardy-Littlewood conjectures for counts of prime k-tuples (see (1.4) below).

In contrast with the Cramér model, Granville's random set $\mathcal{G}$ satisfies

$$G_{\mathcal{G}}(x) \gtrsim \xi \log^2 x, \qquad \xi := 2e^{-\gamma} = 1.1229\cdots, \tag{1.3}$$

with probability one. Granville establishes (1.3) by choosing starting points a with $Q \mid a$. If $y \asymp \log^2 x$, then there are about $y/\log y$ numbers $n \in [a, a+y]$ that are coprime to every $p \leqslant A$; this is a factor ξ smaller than the corresponding quantity for a random starting point a, and it accounts for the difference between (1.2) and (1.3). We elaborate on this idea in our analysis of $G_{\mathcal{R}}(x)$.

1.3. **A new probabilistic model for primes.** Hardy and Littlewood [19] conjectured that the asymptotic relation

$$|\{n \leqslant x : n + h \in \mathcal{P} \text{ for all } h \in \mathcal{H}\}| = \big(\mathfrak{S}(\mathcal{H}) + o(1)\big) \int_2^x \frac{dt}{\log^{|\mathcal{H}|} t} \tag{1.4}$$

holds for any finite set $\mathcal{H} \subset \mathbb{Z}$, where $\mathfrak{S}(\mathcal{H})$ is the singular series given by

$$\mathfrak{S}(\mathcal{H}) := \prod_p \left(1 - \frac{|\mathcal{H} \bmod p|}{p}\right)\left(1 - \frac{1}{p}\right)^{-|\mathcal{H}|}. \tag{1.5}$$

Note that the left side of (1.4) is bounded if $|\mathcal{H} \bmod p| = p$ for some prime p, since then for every integer n, one has $p \mid n + h$ for some $h \in \mathcal{H}$. In this case, $\mathfrak{S}(\mathcal{H}) = 0$. We say that $\mathcal{H}$ is *admissible* if $|\mathcal{H} \bmod p| < p$ for every prime p.

To motivate our model set $\mathcal{R}$, we first reinterpret (1.4) probabilistically. The rapid convergence of the product (1.5) implies that $\mathfrak{S}(\mathcal{H})$ is well approximated by the truncation

$$\mathfrak{S}_z(\mathcal{H}) := \prod_{p \leqslant z} \left(1 - \frac{|\mathcal{H} \bmod p|}{p}\right)\left(1 - \frac{1}{p}\right)^{-|\mathcal{H}|} = V_{\mathcal{H}}(z)\Theta_z^{-|\mathcal{H}|},$$

where

$$V_{\mathcal{H}}(z) := \prod_{p \leqslant z} \left(1 - \frac{|\mathcal{H} \bmod p|}{p}\right) \qquad \text{and} \qquad \Theta_z := \prod_{p \leqslant z} \left(1 - \frac{1}{p}\right). \tag{1.6}$$

We interpret $V_{\mathcal{H}}(z)$ as a product of local densities, and Θ_z as a kind of global density. In order to match the global density of primes as closely as possible, we take $z = z(t)$ be the largest prime number for which $1/\Theta_{z(t)} \leqslant \log t$; this is well-defined for $t \geqslant e^2$, and by the prime number theorem we have

$$z(t) \sim t^{1/e^\gamma} \qquad \text{and} \qquad \Theta_{z(t)}^{-1} = \log t + O(t^{-1/e^\gamma}). \tag{1.7}$$

It follows that the right side of (1.4) is

$$\sim \int_{e^2}^x V_{\mathcal{H}}(z(t))\, dt.$$

On the other hand, the quantity $V_{\mathcal{H}}(z)$ can be written probabilistically as

$$V_{\mathcal{H}}(z) = \mathbb{P}(\mathcal{H} \subset \mathcal{S}_z), \tag{1.8}$$

where $\mathbb{P}$ denotes probability over a uniform choice of residue classes $a_p \bmod p$, for every prime p, with the random variables $a_p \bmod p$ being jointly independent in p, and $\mathcal{S}_z$ is the random set

$$\mathcal{S}_z := \mathbb{Z} \setminus \bigcup_{p \leqslant z} (a_p \bmod p). \tag{1.9}$$

Thus, $\mathcal{H} \subset \mathcal{S}_z$ is the event that $\mathcal{H}$ survives sieving by random residue classes modulo primes $p \leqslant z$. Consequently, for admissible $\mathcal{H}$, (1.4) takes the form

$$|\{n \leqslant x : n + h \in \mathcal{P} \text{ for every } h \in \mathcal{H}\}| \sim \int_{e^2}^{x} \mathbb{P}(\mathcal{H} \subset \mathcal{S}_{z(t)})\, dt.$$

Thus, (1.4) asserts that the probability that a random shift of $\mathcal{H}$ lies in $\mathcal{P}$ is asymptotically the same as the probability that $\mathcal{H}$ lies in a randomly sifted set.

Motivated by this probabilistic interpretation of (1.4), we now define

$$\mathcal{R} := \{n \geqslant e^2 : n \in \mathcal{S}_{z(n)}\} \tag{1.10}$$

as our random set of integers. Note that the number of primes being sieved out increases as n increases in order to mimic the slowly decreasing density of the primes. This can be compared with the description of $\mathcal{P}$ using the sieve of Eratosthenes, in which $z(n)$ is replaced by $n^{1/2}$ and the a_p are replaced by 0.

We believe that the random set $\mathcal{R}$ is a useful model for primes, especially for studying local statistics such as gaps. On the other hand, the analysis of $\mathcal{R}$ presents more difficulties than the analysis of $\mathcal{C}$ or $\mathcal{G}$, owing to the more complicated coupling between events such as $n_1 \in \mathcal{R}$ and $n_2 \in \mathcal{R}$ for $n_1 \neq n_2$.

1.4. **Large gaps from the model.** The behavior of $G_{\mathcal{R}}(x)$ is intimately tied to extremal properties of the *interval sieve*. To describe this connection, for any $y \geqslant 2$ let W_y denote the (deterministic) quantity

$$W_y := \min \left| [0, y] \cap \mathcal{S}_{(y/\log y)^{1/2}} \right|, \tag{1.11}$$

where $\mathcal{S}_z$ is defined in (1.9) and the minimum in (1.11) is taken over all choices of the residue classes $\{a_p \bmod p : p \leqslant (y/\log y)^{1/2}\}$. At present, the sharpest known bounds on W_y are

$$(4 + o(1)) \frac{y \log_2 y}{\log^2 y} \leqslant W_y \leqslant \frac{y}{\log y} + O\left(\frac{y \log_2 y}{\log^2 y}\right), \tag{1.12}$$

the lower bound being a consequence of Iwaniec's theory (see [12, Theorem 12.14] or [21]) of the linear sieve, and the upper bound resulting from the particular choice $a_p := 0 \bmod p$ for all primes $p \leqslant (y/\log y)^{1/2}$. There is a folklore conjecture that the upper bound in (1.12) is closer to the truth. The problem of bounding W_y belongs to a circle of problems centered on the question about the maximum number of primes in some interval of length x; see e.g., [20] and [9].

THEOREM 1.1 (Asymptotic for largest gap in the random model). *Put*

$$g(u) := \max\{y : W_y \log y \leqslant u\} \tag{1.13}$$

and define $\xi := 2e^{-\gamma} = 1.1229\ldots$. *For any* $\varepsilon > 0$, *with probability one, we have*

$$g((\xi - \varepsilon) \log^2 x) \leqslant G_{\mathcal{R}}(x) \leqslant g((\xi + \varepsilon) \log^2 x)$$

for all large x.

The function $g(u)$ is evidently increasing, and by (1.12) we see that

$$(1+o(1))u \leqslant g(u) \leqslant (1+o(1))\frac{u\log u}{4\log_2 u} \qquad (u\to\infty) \tag{1.14}$$

and so Theorem 1.1 implies that for every $\varepsilon > 0$, almost surely we have

$$(\xi-\varepsilon)\log^2 x \leqslant G_{\mathcal{R}}(x) \leqslant (\xi+\varepsilon)\frac{\log^2 x\log_2 x}{2\log_3 x} \tag{1.15}$$

for all large x.

It seems likely that $g((\xi\pm\varepsilon(x))\log^2 x) \sim g(\xi\log^2 x)$ whenever $\varepsilon(x)$ goes to zero as $x\to\infty$, although we cannot prove this. Assuming this, Theorem 1.1 leads us to the following prediction for gaps between primes:

Conjecture 1.2 (Asymptotic for largest gap in the primes). *We have*

$$G_{\mathcal{P}}(x) \sim g(\xi\log^2 x) \qquad (x\to\infty).$$

Assuming the previously mentioned folklore conjecture that the lower bound in (1.14) is asymptotically tight in the sense that $g(u)\sim u$ as $u\to\infty$, we are then led to the prediction that

$$G_{\mathcal{P}}(x) \sim \xi\log^2 x \qquad (x\to\infty).$$

This matches the lower bound (1.3) for the gap in the Granville model $\mathcal{G}$.

1.5. **Hardy-Littlewood from the model.** It has been conjectured that a much more precise version of (1.4) holds (see, e.g., Montgomery and Soundararajan [28]), namely:

$$|\{n\leqslant x : n+h\in\mathcal{P} \text{ for all } h\in\mathcal{H}\}| = \mathfrak{S}(\mathcal{H})\int_2^x \frac{dt}{\log^{|\mathcal{H}|}t} + O(x^{1/2+\varepsilon}). \tag{1.16}$$

There is some computational evidence for this strong estimate for certain small sets $\mathcal{H}$; see Section 2.1. Granville's model set $\mathcal{G}$, by contrast, satisfies the analogous relation with an error term that cannot be made smaller than $O(x/\log^{|\mathcal{H}|+1}x)$. This occurs because $\mathcal{G}$ is only capturing the bias of $\mathcal{P}$ modulo primes $p\leqslant A$; that is, the set $\mathcal{G}$ satisfies the analog of (1.16) with $\mathfrak{S}(\mathcal{H})$ replaced by $\mathfrak{S}_A(\mathcal{H})$.

The model set $\mathcal{R}$ given by (1.10) has been designed with the Hardy-Littlewood conjectures in mind. We establish a uniform analog of (1.16) that holds in a wide range of $\mathcal{H}$.

Theorem 1.3 (Hardy-Littlewood conjecture for the random model). *Fix $c\in[1/2,1)$ and $\varepsilon>0$. Almost surely, we have*

$$|\{n\leqslant x : n+h\in\mathcal{R} \text{ for all } h\in\mathcal{H}\}| = \mathfrak{S}(\mathcal{H})\int_2^x \frac{dt}{\log^{|\mathcal{H}|}t} + O\big(x^{1-\frac{1-c}{8c^2-2c}+\varepsilon}\big)$$

uniformly for all admissible tuples $\mathcal{H}$ satisfying $|\mathcal{H}|\leqslant\log^c x$ and in the range $\mathcal{H}\subset[0,\exp(\frac{\log^{1-c}x}{\log_2 x})]$.

In particular, when $c=\frac{1}{2}$ the error term is $O(x^{1/2+o(1)})$, which matches (1.16) provided that $\mathcal{H}\subseteq[0,\exp\{\frac{\log^{1/2}x}{\log_2 x}\}]$ and $|\mathcal{H}|\leqslant\log^{1/2}x$. As we will invoke the Borel-Cantelli lemma in the proof, the constant implied by the $O-$symbol exists

almost surely, but we cannot give any uniform bound on it. This remark applies to the next result as well.

For the special case $\mathcal{H} = \{0\}$ we have the following more precise statement.

THEOREM 1.4 (Riemann hypothesis for the random model). *Fix $c > 3/2$. Almost surely, we have*

$$|\{n \in \mathcal{R} : n \leqslant x\}| = \int_2^x \frac{dt}{\log t} + O(x^{1/2} \log^c x).$$

Similar results can be obtained for any *fixed* tuple $\mathcal{H}$; we leave this to the interested reader.

1.6. **Large gaps from Hardy-Littlewood.** The results stated above have a partial deterministic converse. We show that *any* set of integers that satisfies a uniform analogue of the Hardy-Littlewood conjecture (1.16) has large gaps. The maximal length of the gaps depends on the range of uniformity of (1.16), and comes close to order $\log^2 x$ with a strong uniformity assumption. Our result extends a theorem of Gallagher [14], who showed that if, for every fixed $k \in \mathbb{N}$ and real $c > 1$, the primes obey the Hardy-Littlewood conjectures uniformly for every admissible k-tuple $\mathcal{H} \subset [0, c \log x]$, then the gaps normalized by $\frac{1}{\log x}$ enjoy an exponential distribution asymptotically. His approach applies to any set $\mathcal{A}$ in place of the primes $\mathcal{P}$.

THEOREM 1.5 (Hardy-Littlewood implies large gaps). *Assume $\frac{2 \log_2 x}{\log x} \leqslant \kappa \leqslant 1/2$ and that $\mathcal{A} \subset \mathbb{N}$ satisfies the Hardy-Littlewood type conjecture*

$$|\{n \leqslant x : n + h \in \mathcal{A} \text{ for all } h \in \mathcal{H}\}| = \mathfrak{S}(\mathcal{H}) \int_2^x \frac{dt}{\log^{|\mathcal{H}|} t} + O(x^{1-\kappa}) \tag{1.17}$$

uniformly over all tuples $\mathcal{H} \subset [0, \log^2 x]$ with $|\mathcal{H}| \leqslant \frac{\kappa \log x}{2 \log_2 x}$. Then

$$G_{\mathcal{A}}(x) \gg \frac{\kappa \log^2 x}{\log_2 x}$$

for all large x, where the implied constant is absolute.

We also have the following variant of Theorem 1.5, which has a stronger conclusion but requires a uniform Hardy-Littlewood conjecture for larger tuples (of cardinality as large as $\log x \log_2 x$); on the other hand, this conjecture is only needed in a certain averaged sense.

THEOREM 1.6 (Averaged Hardy-Littlewood implies large gaps). *Fix $0 < c < 1$. Suppose that $\mathcal{A} \subset \mathbb{N}$ satisfies the averaged Hardy-Littlewood type conjecture*

$$\sum_{\substack{\mathcal{H} \subset [0,y] \\ |\mathcal{H}| = k}} |\{n \leqslant x : n + h \in \mathcal{A} \text{ for all } h \in \mathcal{H}\}| = \sum_{\substack{\mathcal{H} \subset [0,y] \\ |\mathcal{H}| = k}} \int_2^x \frac{\mathfrak{S}_{z(t)}(\mathcal{H})}{\log^k t}\, dt + O(x^{1-c}) \tag{1.18}$$

uniformly for $k \leqslant \frac{Cy}{\log x}$ *and* $\log x \leqslant y \leqslant (\log^2 x)\log_2 x$, *where* C *is a sufficiently large absolute constant. Then*

$$G_{\mathcal{A}}(x) \geqslant g((c\,\xi - o(1))\log^2 x) \qquad (x \to \infty),$$

where g *is defined in* (1.13).

One could combine Theorem 1.3 with Theorem 1.5 (taking $\kappa := (\log x)^{c-1+\varepsilon}$ with fixed $c < 1$, say) to obtain results similar to Theorem 1.1. However, the conclusion is considerably weaker than that of Theorem 1.1, and it does not appear that this approach is going to come close to recovering the bounds we obtain using a direct argument.

Below we summarize, in rough form, the various results and conjectures for the primes $\mathcal{P}$, the various random models $\mathcal{C}, \mathcal{G}, \mathcal{R}$ for the primes, and for arbitrary sets $\mathcal{A}$ obeying a Hardy-Littlewood type conjecture:

Set	Hardy-Littlewood conjecture?	Asymptotic largest gap up to x
$\mathcal{C}$	No (singular series is missing)	$\sim \log^2 x$
$\mathcal{G}$	Yes (with weak error term)	$g((\xi \pm o(1))\log^2 x)$
$\mathcal{R}$	Yes (with error $O(x^{1-c})$)	$g((\xi \pm o(1))\log^2 x)$
$\mathcal{P}$	Yes (conjecturally)	$\sim \xi \log^2 x$ (conjecturally)
$\mathcal{A}$	Assumed (error $O(x^{1-c})$)	$\gg c\frac{\log^2 x}{\log_2 x}$
$\mathcal{A}$	Assumed on average (error $O(x^{1-c})$) for tuples of size up to $(\log x)\log_2 x$	$\gtrsim g((c\,\xi - o(1))\log^2 x)$

Of course, one can combine this table's conclusions with the unconditional bounds in (1.14), or the conjecture $g(u) \sim u$, to obtain further rigorous or predicted upper and lower bounds for the largest gap.

1.7. **Open Problems.**

(1) Improve upon the bounds (1.12); alternatively, give some heuristic reason for why the upper bound in (1.12) should be closer to the truth.
(2) Show that $g(a) \sim g(b)$ whenever $a \sim b$. This will clean up the statement of Theorem 1.1.
(3) Analyze the distribution of large gaps between special elements of $\mathcal{R}$. For example, what is the largest gap between elements of $\{n : n \in \mathcal{R}, n + 2 \in \mathcal{R}\}$ below x? This should be a good predictor for the maximal gap between pairs of twin primes and likely will involve a different extremal sieve problem.

1.8. **Plan of the paper.** Following further remarks and background inequalities in Sections 2 and 3, we prove Theorems 1.3 and 1.4 in Section 4 using first and second moment bounds. Section 5 and 6 contain probability estimates on $|[0, y] \cap \mathcal{S}_w|$ for various ranges of w. These are then used to prove Theorem 1.1 in Section 7 and Theorems 1.5 and 1.6 in Section 8. In Section 2.4, we connect the interval sieve problem to the problem of "exceptional zeros," made explicit in Theorem 2.2; this is proved in Section 9.

2. Background and Further Remarks

The discussion here is not needed for the proofs of the main theorems and may be omitted on the first reading.

2.1. **Remarks on the Hardy-Littlewood conjectures.** For any $\mathcal{H} \subseteq [0, y]$, we have $\mathfrak{S}(\mathcal{H}) \leqslant e^{O(|\mathcal{H}| \log_2 y)}$ (see Lemma 3.4 below), and thus when $y \leqslant (\log x)^{O(1)}$, the main terms in (1.16) and (1.17) are smaller than one for $c_1 \frac{\log x}{\log_2 x} \leqslant |\mathcal{H}| \leqslant \exp\{(\log x)^{c_2}\}$, where $c_1, c_2 > 0$ are appropriate constants. Therefore, we cannot have a genuine asymptotic when $|\mathcal{H}| > c_1 \frac{\log x}{\log_2 x}$.

In the case of primes, it may be the case that (1.16) fails when $|\mathcal{H}| > \frac{\log x}{\log_2 x}$ owing to potentially large fluctuations in both the size of $\mathfrak{S}(\mathcal{H})$ and in the prime counts themselves. We note that Elsholtz [8] has shown that for any $c > 0$, the left side of (1.16) is bounded by

$$O\left(x \exp\left(-(\tfrac{1}{4} + o(1)) \frac{\log x \log_3 x}{\log_2 x}\right)\right)$$

when $|\mathcal{H}| \geqslant c \log x$, where the implied function $o(1)$ depends on c. On the other hand, there are admissible tuples with $|\mathcal{H}| \ll \log x$ for which the left side of (1.16) is zero (see [8] for a construction of such $\mathcal{H}$).

Our assumption in Theorem 1.6 is more speculative, in light of the above remarks, since we need to deal with tuples $\mathcal{H}$ satisfying $k = |\mathcal{H}| > \log x$. Also, simply considering subsets $\mathcal{H}$ of the primes in $(y/2, y]$ (which are automatically admissible), we see that there are at least $(\frac{y}{k \log y})^k > (\log x)^{k/2}$ tuples $\mathcal{H}$ in the summation, and this means that when $k > \log x$, (1.18) implies a great deal of cancellation in the error terms of (1.17) over tuples $\mathcal{H}$.

In a few special cases, e.g., $\mathcal{H} = \{0, 2\}$, $\mathcal{H} = \{0, 2, 6\}$, and $\mathcal{H} = \{0, 4, 6\}$, there is extensive numerical evidence (cf. [19, pp. 43–44, 62–64], [32], [24], [33], [34]) in support of the conjecture (1.16) with such a strong error term[3]. Note that the special case of (1.16) with $\mathcal{H} = \{0\}$ is equivalent to the Riemann Hypothesis. Theorem 1.3 makes plausible the notion that (1.16) may hold uniformly for $\mathcal{H} \subset [0, Y]$ with $|\mathcal{H}| \leqslant K$, where Y, K are appropriate functions of x.

2.2. **The cutoff $z(t)$.** In [37], Pólya suggests using a truncation x^{1/e^γ} to justify the Hardy-Littlewood conjectures. The observation that the cutoff $\sqrt{x}$ leads to erroneous prime counts was made by Hardy and Littlewood [19, Section 4.3] and is occasionally referred to as "the Mertens Paradox" (see [31]). In discussing the probabilistic heuristic for counting the number of primes below x, Hardy and Littlewood write (here ϖ denotes a prime) "One might well replace $\varpi < \sqrt{n}$ by $\varpi < n$, in which case we should obtain a probability half as large. This remark is in itself enough to show the unsatisfactory character of the argument" and later "*Probability* is not a notion of pure mathematics, but of philosophy or physics."

[3]Most of this work appears only on web pages, rather than in books or journals.

2.3. **Connection to Jacobsthal's function.** Any improvement of the lower bound in (1.12) leads to a corresponding improvement of the known upper bound on Jacobsthal's function $J(w)$, which we define to be the largest gap which occurs in the set of integers that have no prime factor $\leqslant w$. Equivalently, $J(w)$ is the largest gap in $\mathcal{S}_w$. Iwaniec [21] proved that $J(w) \ll w^2$ using his linear sieve bounds. Using Montgomery and Vaughan's explicit version of the Brun-Titchmarsh inequality [29], the cardinality of the set $\mathcal{S}_w(y) := [0,y] \cap \mathcal{S}_w$ for $w > (y/\log y)^{1/2}$ can be bounded from below by

$$\begin{aligned} |\mathcal{S}_w(y)| &\geqslant |\mathcal{S}_{(y/\log y)^{1/2}}(y)| - \sum_{(y/\log y)^{1/2} < p \leqslant w} |\mathcal{S}_{(y/\log y)^{1/2}}(y) \cap (a_p \bmod p)| \\ &\geqslant W_y - \sum_{(y/\log y)^{1/2} < p \leqslant w} \frac{2y/p}{\log(2y/p)}. \end{aligned}$$

If the right side is positive, it follows that $J(w) < y$. Suppose, for example, that $W_y \geqslant \alpha y/\log y$ for large y, where $0 < \alpha \leqslant 1$ is fixed. Mertens' estimates then imply that

$$J(w) \ll w^{1+e^{-\alpha/2}+o(1)} \qquad (w \to \infty),$$

which improves Iwaniec's upper bound.

We remark that all of the unconditional lower bounds on $G_{\mathcal{P}}(x)$, including the current record [11], have utilized the simple inequality $G_{\mathcal{P}}(x) \geqslant J(y)$, where $y \sim \log x$.

2.4. **The interval sieve problem and exceptional zeros.** The problem of determining W_y asymptotically is connected with the famous problem about exceptional zeros of Dirichlet L-functions (also known as Siegel zeros or Landau-Siegel zeros); see, e.g., [7, Sections 14, 20, 21, 22] for background on these and [22] for further discussion.

Definition 2.1. We say that *exceptional zeros exist* if there is an infinite set $\mathcal{E} \subset \mathbb{N}$, such that for every $q \in \mathcal{E}$ there is a real Dirichlet character χ_q and a zero $1-\delta_q$ with $L(1-\delta_q, \chi_q) = 0$ and $\delta_q = o(1/\log q)$ as $q \to \infty$. □

Theorem 2.2. *Suppose that exceptional zeros exist. Then*

$$\liminf_{y\to\infty} \frac{W_y}{y/\log y} = 0 \qquad \textit{and} \qquad \limsup_{u\to\infty} \frac{g(u)}{u} = \infty.$$

Hence, we almost surely have

$$\limsup_{x\to\infty} \frac{G_{\mathcal{R}}(x)}{\log^2 x} = \infty$$

and Conjecture 1.2 implies that

$$\limsup_{x\to\infty} \frac{G_{\mathcal{P}}(x)}{\log^2 x} = \infty.$$

Our proof of Theorem 2.2, given in Section 9, is quantitative, exhibiting an upper bound for W_y in terms of the decay of δ_q. Siegel's theorem [7, Sec. 21] implies that $\frac{\log 1/\delta_q}{\log q} \to 0$, but we cannot say anything about the rate at which this

occurs (i.e., the bound is *ineffective*). If the rate of decay to zero is extremely slow, then our proof shows that, infinitely often, $W_y = f(y)\frac{y\log_2 y}{\log y}$, with $f(y) \to \infty$ extremely slowly. Consequently, $G_{\mathcal{R}}(x)$ is infinitely often close to the upper bound in (1.15).

The related quantity

$$\widetilde{W}_y := \max |S_{\sqrt{y}} \cap [0, y]|,$$

is known by the theory of upper bound sieves to satisfy $\widetilde{W}_y \leqslant \frac{2y}{\log y}$ (see, e.g., [30]), and it is well known that an improvement of the constant two would imply that exceptional zeros do not exist; see, e.g., Selberg's paper [39]. Theorem 2.2 (in the contrapositive) similarly asserts that an improvement of the constant zero in the trivial lower bound $W_y \geqslant 0 \cdot \frac{y}{\log y}$ implies that exceptional zeroes do not exist. Extending our ideas and those of Selberg, Granville [17] has recently shown that if exceptional zeros exist, then for any real $r > 1$,

$$\liminf_{y\to\infty} \frac{\min_{(a_p)} |[0, y] \cap \mathcal{S}_{y^{1/r}}|}{e^{-\gamma} y/\log y^{1/r}} = f(r),$$

$$\limsup_{y\to\infty} \frac{\max_{(a_p)} |[0, y] \cap \mathcal{S}_{y^{1/r}}|}{e^{-\gamma} y/\log y^{1/r}} = F(r),$$

where f, F are the lower and upper linear sieve functions. In particular, $f(r) = 0$ for $r \leqslant 2$ and $f(r) > 0$ for $r > 2$.

It is widely believed that exceptional zeros do not exist, and this is a famous unsolved problem. Theorem 2.2 indicates that to fully understand W_y, it is necessary to solve this problem. Iwaniec's lectures [22] give a nice overview of the problem of exceptional zeros, attempts to prove that they do not exist, and various consequences of their existence. In the paper [10], the second author shows that if there is a sequence of moduli q with $\delta_q \ll (\log q)^{-2}$, then one can deduce larger lower bounds for $J(w)$ and $G_{\mathcal{P}}(x)$ than are currently known unconditionally.

2.5. **Primes in longer intervals.** With probability one, the Cramér model $\mathcal{C}$ also satisfies

$$\pi_{\mathcal{C}}(x + y) - \pi_{\mathcal{C}}(x) \sim \frac{y}{\log x} \tag{2.1}$$

as long as $x \to \infty$, $y \leqslant x$, and $y/\log^2 x \to \infty$. However, Maier [25] has shown that the analogous statement for primes is false, namely that for any fixed $A > 1$ one has

$$\liminf_{x\to\infty} \frac{\pi(x + (\log x)^A) - \pi(x)}{(\log x)^{A-1}} < 1 \quad \text{and} \quad \limsup_{x\to\infty} \frac{\pi(x + (\log x)^A) - \pi(x)}{(\log x)^{A-1}} > 1. \tag{2.2}$$

The disparity between (2.1) and (2.2) again stems from the uniform distribution of $\mathcal{C}$ in residue classes modulo primes. Both models $\mathcal{G}$ and $\mathcal{R}$ satisfy the analogs of (2.2); we omit the proofs. Moreover, the ideas behind Theorem 1.1 can be used to sharpen (2.2), by replacing the right sides of the inequalities by quantities defined in terms of the extremal behavior of $|[0, y] \cap S_{y^{1/u}}|$ for fixed $u > 1$; we refer the reader to [23, Exercise 30.1] for details. The authors thank Dimitris Koukoulopoulos for this observation.

By contrast, on the Riemann Hypothesis, Selberg [38] showed that

$$\pi(x+y) - \pi(x) \sim \frac{y}{\log x}$$

holds for *almost all* x provided that $y = y(x) \leqslant x$ satisfies $y/\log^2 x \to \infty$ as $x \to \infty$.

On a related note, Granville and Lumley [18] have developed heuristics and conjectures concerning the *maximum* number of primes $\leqslant x$ lying in intervals of length L, where L varies between $\log x$ and $\log^2 x$.

2.6. **Remarks on the singular series and prime gaps.** If y is small compared to x, the difference $\pi_{\mathcal{C}}(x+y) - \pi_{\mathcal{C}}(x)$ is a random variable with (essentially) a binomial distribution. Letting $y \to \infty$ with $y/\log x$ fixed, the result is a Poisson distribution: for any real $\lambda > 0$ and any integer $k \geqslant 0$, we have

$$\big|\big\{m \leqslant x : \pi_{\mathcal{C}}(m + \lambda \log m) - \pi_{\mathcal{C}}(m) = k\big\}\big| \sim e^{-\lambda}\frac{\lambda^k}{k!}x \qquad (x \to \infty) \tag{2.3}$$

with probability one. In particular, using $\mathcal{C}$ as a model for the primes $\mathcal{P}$, this leads to the conjecture that

$$\lim_{x\to\infty} \pi(x)^{-1}\big|\{p_n \leqslant x : p_{n+1} - p_n \geqslant \lambda \log p_n\}\big| = e^{-\lambda} \qquad (\lambda > 0). \tag{2.4}$$

Gallagher [14] showed that if the Hardy-Littlewood conjectures (1.4) are true uniformly for $\mathcal{H} \subset [0, \log^2 x]$ with fixed cardinality $|\mathcal{H}|$, then (2.4) follows. His analysis relies on the relation

$$\sum_{\substack{\mathcal{H}\subset[0,y]\\ |\mathcal{H}|=k}} \mathfrak{S}(\mathcal{H}) \sim \binom{y}{k} \qquad (y \to \infty), \tag{2.5}$$

which asserts that the singular series has an average value of one. Sharper versions of (2.5) exist (see, e.g., Montgomery and Soundararajan [28]); such results, however, are uniform only in a range $|\mathcal{H}| \ll \log_2 y$ or so, far too restrictive for our use. Reinterpreting the sum on the left side of (2.5) probabilistically, as we have done above, allows us to adequately deal with a much larger range of sizes $|\mathcal{H}|$. In particular, it is possible to deduce from a uniform version of (1.16) a uniform version of (2.4), although we have not done so in this paper.

We take this occasion to mention a recent unconditional theorem of Mastrostefano [26, Theorem 1.1], which is related to (2.5), and which states that for any integer $m \geqslant 0$ there is an $\varepsilon = \varepsilon(m) > 0$ so that whenever $0 < \lambda < \varepsilon$, we have

$$|\{n \leqslant x : |[n, n + \lambda \log n] \cap \mathcal{P}| = m\}| \gg_{\lambda,\varepsilon} x.$$

Establishing the Poisson distribution (2.3) unconditionally, even for some fixed λ, seems very difficult.

2.7. **The maximal gap in Granville's model.** The claimed bounds in Theorem 1.1 are also satisfied by Granville's random set $\mathcal{G}$, i.e., one has

$$g((\xi - o(1)) \log^2 x) \leqslant G_{\mathcal{G}}(x) \leqslant g((\xi + o(1)) \log^2 x).$$

The proof is very short, and we sketch it here as a prelude to the proof of Theorem 1.1. Consider the elements of $\mathcal{G}$ in $(x, 2x]$ for x a power of two. In accordance with (1.14), let y satisfy $\log^2 x \leqslant y = o(\log^2 x \log_2 x)$ and put $A :=$

$(y/\log y)^{1/2}$, so that $A = o(\log x)$. Let $\theta := \prod_{p\leqslant A}(1-1/p)^{-1} \sim (e^\gamma/2)\log y$ and $Q := \prod_{p\leqslant A} p$. For simplicity, we suppose that each $n \in (x, 2x]$ with $(n, Q) = 1$ is chosen for inclusion in $\mathcal{G}$ with probability $\theta/\log x$; this modification has a negligible effect on the size of the largest gap. Fix $\varepsilon > 0$ arbitrarily small. Let X_m denote the event $(m, m+y] \cap \mathcal{G} = \varnothing$.

Let D_m denote the number of integers in $(m, m+y]$, all of whose prime factors are larger than A. If we take $y := g((\xi+\varepsilon)\log^2 x)$, then

$$\begin{aligned}\mathbb{E}\big|\{x < m \leqslant 2x : X_m\}\big| &= \sum_{x<m\leqslant 2x} (1-\theta/\log x)^{D_m}\\ &\leqslant x(1-\theta/\log x)^{W_y} \leqslant xe^{-\theta W_y/\log x}\\ &\ll x^{-\varepsilon/2}\end{aligned}$$

by our assumption that $W_y \log y \sim (\xi+\varepsilon)\log^2 x$. Summing on x and applying Borel-Cantelli, we see that almost surely, only finitely many X_m occur.

For the lower bound, we take $y := g((\xi-\varepsilon)\log^2 x)$ and restrict to special values of m, namely $m \equiv b \bmod Q$, where b is chosen so that

$$D_b = W_y.$$

Let $\mathcal{M} := \{x < m \leqslant 2x : m \equiv b \bmod Q\}$ and let N be the number of $m \in \mathcal{M}$ for which X_m occurs. By the above argument, we see that

$$\mathbb{E}N = |\mathcal{M}|(1-\theta/\log x)^{W_y}.$$

By assumption, $|\mathcal{M}| = x^{1-o(1)}$ and hence the right side is $> x^{\varepsilon/2}$ for large x. Similarly,

$$\begin{aligned}\mathbb{E}N^2 &= |\mathcal{M}|(1-\theta/\log x)^{W_y} + (|\mathcal{M}|^2 - |\mathcal{M}|)(1-\theta/\log x)^{2W_y}\\ &= (\mathbb{E}N)^2 + O(\mathbb{E}N).\end{aligned}$$

By Chebyshev's inequality, $\mathbb{P}(N < \frac{1}{2}\mathbb{E}N) \ll 1/\mathbb{E}N \ll x^{-\varepsilon/2}$. Considering all x and using Borel-Cantelli, we conclude that almost surely every sufficiently large dyadic $(x, 2x]$ contains an m for which X_m occurs.

We remark that our lower bound argument above works as well for the Cramér model, showing (1.2). We take $A = Q = \theta = b = 1$, and the details are simpler.

Acknowledgements. The authors thank Andrew Granville, Ben Green, D. R. Heath-Brown, Henryk Iwaniec, Dimitris Koukoulopoulos, James Maynard, Carl Pomerance and Joni Teräväinen for useful discussions, especially concerning the interval sieve problem.

3. Preliminaries

3.1. **Notation.** The indicator function of any set $\mathcal{T}$ is denoted $\mathbf{1}_{\mathcal{T}}(n)$. We select residue classes $a_p \bmod p$ uniformly and independently at random for each prime p, and then for any set of primes $\mathcal{Q}$ we denote by $\mathcal{A}_{\mathcal{Q}}$ the ordered tuple $(a_p : p \in \mathcal{Q})$; often we condition our probabilities on $\mathcal{A}_{\mathcal{Q}}$ for a fixed choice of $\mathcal{Q}$.

Probability, expectation, and variance are denoted by $\mathbb{P}$, $\mathbb{E}$, and $\mathbb{V}$ respectively. We use $\mathbb{P}_{\mathcal{Q}}$ and $\mathbb{E}_{\mathcal{Q}}$ to denote the probability and expectation, respectively, with respect to random $\mathcal{A}_{\mathcal{Q}}$. When $\mathcal{Q}$ is the set of primes in $(c, d]$, we write $\mathcal{A}_{c,d}$, $\mathbb{P}_{c,d}$

and $\mathbb{E}_{c,d}$; if $\mathcal{Q}$ is the set of primes $\leqslant c$, we write $\mathcal{A}_c$, $\mathbb{P}_c$ and $\mathbb{E}_c$. In particular, $\mathbb{P}_{c,d}$ refers to the probability over random $\mathcal{A}_{c,d}$, often with conditioning on $\mathcal{A}_c$.

Throughout the paper, any implied constants in symbols O, $\ll$ and $\gg$ are *absolute* (independent of any parameter) unless otherwise indicated. The notations $F \ll G$, $G \gg F$ and $F = O(G)$ are all equivalent to the statement that the inequality $|F| \leqslant c|G|$ holds with some constant $c > 0$. We write $F \asymp G$ to indicate that $F \ll G$ and $G \ll F$ both hold. The notation $o(1)$ is used to indicate a function that tends to zero as $x \to \infty$; in expressions like $1 - o(1)$, the function is assumed to be positive. We write $F \sim G$ when $F = (1 + o(1))G$ as $x \to \infty$.

For a set $\mathcal{H}$ of integers, we denote $\mathcal{H} - \mathcal{H} := \{h - h' : h, h' \in \mathcal{H}\}$, and for any integer m, $\mathcal{H} + m := \{h + m : h \in \mathcal{H}\}$.

3.2. **Various inequalities.** We collect here some standard inequalities from sieve theory and probability that are used in the rest of the paper.

LEMMA 3.1 (Upper bound sieve, [30, Theorem 3.8]). *For $1 \leqslant w \leqslant p \leqslant y$, p prime, $b \in \mathbb{Z}/p\mathbb{Z}$, and an arbitrary interval $\mathcal{I}$ of length y, we have uniformly*

$$\big|\{n \in \mathcal{I} : n \equiv b \bmod p, \big(n, \prod_{q \leqslant w} q\big) = 1\}\big| \ll \frac{y/p}{1 + \min\{\log w, \log(y/p)\}}.$$

LEMMA 3.2 (Azuma's inequality [1]). *Suppose that $X_0, \dots, X_n$ is a martingale with $|X_{j+1} - X_j| \leqslant c_j$ for each j. Then*

$$\mathbb{P}\left(|X_n - X_0| \geqslant t\right) \leqslant 2\exp\left\{-\frac{t^2}{2(c_0^2 + \cdots + c_{n-1}^2)}\right\} \qquad (t > 0).$$

LEMMA 3.3 (Bennett's inequality [3]). *Suppose that $X_1, \dots, X_n$ are independent random variables such that for each j, $\mathbb{E}X_j = 0$, and $|X_j| \leqslant M$ holds with probability one. Then*

$$\mathbb{P}\bigg(\Big|\sum_{1 \leqslant j \leqslant n} X_j\Big| \geqslant t\bigg) \leqslant 2\exp\bigg\{-\frac{\sigma^2}{M^2}\,\mathscr{L}\Big(\frac{Mt}{\sigma^2}\Big)\bigg\} \qquad (t > 0),$$

where $\sigma^2 := \sum_j \mathbb{V}X_j$, and

$$\mathscr{L}(u) := \int_1^{1+u} \log t\, dt = (1 + u)\log(1 + u) - u.$$

LEMMA 3.4. *For any $\mathcal{H} \subset [0, y]$ with $|\mathcal{H}| = k$, we have*

$$\mathfrak{S}_z(\mathcal{H}) = \mathfrak{S}(\mathcal{H})\bigg(1 + O\Big(\frac{k^2}{z}\Big)\bigg) \qquad (z > \max(y, k^2)) \tag{3.1}$$

and

$$\mathfrak{S}(\mathcal{H}) \leqslant e^{O(k \log_2(y))}. \tag{3.2}$$

Proof. Estimate (3.1) follows from the definition of $\mathfrak{S}(\mathcal{H})$ and the fact that for $p > y$, $|\mathcal{H} \bmod p| = k$. Estimate (3.2) is trivial if $\mathcal{H}$ is inadmissible, since then $\mathfrak{S}(\mathcal{H}) = 0$, and otherwise (3.2) is a special case of [15, (6.16)]. □

LEMMA 3.5. *If $\mathcal{H} \subseteq [0,y]$ is an admissible k-tuple and $t \geqslant 2$ satisfies $z(t) > y$ and $k \leqslant t^{1/100}$, then*

$$V_{\mathcal{H}}(z(t)) = \frac{\mathfrak{S}(\mathcal{H})}{(\log t)^k}\big(1 + O(1/t^{0.55})\big).$$

Proof. Let $z := z(t)$. By (1.7), $z \gg t^{1/e^\gamma} \gg t^{0.561}$. Using Lemma 3.4 and (1.7), we have

$$\begin{aligned} V_{\mathcal{H}}(z(t)) &= \mathfrak{S}_z(\mathcal{H})\Theta_z^k \\ &= \mathfrak{S}(\mathcal{H})\left(1 + O\left(\frac{k^2}{z}\right)\right)\left(\frac{1}{\log t} + O(t^{-1/e^\gamma})\right)^k. \end{aligned}$$

The lemma now follows since $k \leqslant t^{1/100}$. □

4. Uniform Hardy-Littlewood from the model

In this section, we prove Theorems 1.3 and 1.4 using the first and second moment bounds provided by the following proposition.

PROPOSITION 4.1 (First and second moment bounds). *Suppose that x and y are integers with $x \geqslant 3$ and $\sqrt{x} \leqslant y \leqslant x$, and suppose that $0 \leqslant D \leqslant \sqrt{x}$. Let $\mathcal{H} \subset [0,D]$ be an admissible tuple with $k := |\mathcal{H}| \leqslant \frac{\log x}{(\log_2 x)^2}$, and put*

$$X_n := \prod_{h\in\mathcal{H}} \mathbf{1}_{\mathcal{R}}(n+h) \qquad (n \in \mathbb{N}).$$

Then

$$\mathbb{E}\Big(\sum_{x<n\leqslant x+y} X_n\Big) = \mathfrak{S}(\mathcal{H})\int_x^{x+y} \frac{dt}{(\log t)^k} + O\Big(\frac{yD}{x} + \frac{y}{x^{0.54}}\Big). \tag{4.1}$$

Furthermore,

$$\mathbb{V}\Big(\sum_{x<n\leqslant x+y} X_n\Big) \ll y\left(\frac{D}{x} + \frac{yD^2}{x^2} + V_{\mathcal{H}}(z(x))(k^2 + yD/x) + V_{\mathcal{H}}(z(x))^2 F\right), \tag{4.2}$$

where

$$F := \begin{cases} (\log x)^{k^2} & \text{if } k \leqslant \frac{(\log x)^{1/2}}{\log_2 x}, \\ y^{\frac{4\varrho^2-1}{4\varrho^2-\varrho}} \exp\Big\{O\Big(\frac{\log x \log_3 x}{\log_2 x}\Big)\Big\} & \text{if } \frac{(\log x)^{1/2}}{\log_2 x} \leqslant k = (\log x)^\varrho \leqslant \frac{\log x}{(\log_2 x)^2}. \end{cases}$$

Before turning to the proof of the proposition, we first indicate how it is used to prove the two theorems, starting with Theorem 1.4.

Proof of Theorem 1.4. Fix $c > 3/2$. For any integers $u \geqslant 2$ and $v \geqslant 0$, we let

$$\Delta(u, u+v) := \sum_{u<n\leqslant u+v} \mathbf{1}_{\mathcal{R}}(n) - \int_u^{u+v} \frac{dt}{\log t}.$$

We apply Proposition 4.1 in the case that $\mathcal{H} = \{0\}$, $k = 1$ and $D = 0$. By (4.1), if $v \geqslant \sqrt{u}$ then

$$\mathbb{E}\Delta(u, u+v) \ll \frac{v}{u^{0.54}} \ll u^{0.46}. \tag{4.3}$$

Inequality (4.2) implies that

$$\mathbb{V}\big(\Delta(u, u+v)\big) \ll v\big(V_{\mathcal{H}}(z(u)) + V_{\mathcal{H}}(z(u))^2 \log u\big) \ll \frac{v}{\log u}.$$

Let x be a large integer. For integers h, m with $2\sqrt{x} \leqslant 2^m \leqslant x$ and $0 \leqslant h \leqslant x/2^m - 1$, let $G_{m,h}$ be the event that

$$\big|\Delta(x + h \cdot 2^m, x + (h+1)2^m)\big| \leqslant x^{1/2}(\log x)^{c-1}.$$

For large x, (4.3) implies that

$$\big|\mathbb{E}\Delta(x + h \cdot 2^m, x + (h+1)2^m)\big| \leqslant \frac{x^{1/2}(\log x)^{c-1}}{2}.$$

Hence, Chebyshev's inequality yields the bound

$$\mathbb{P}\big(\text{not } G_{h,m}\big) \ll \frac{2^m}{x(\log x)^{2c-1}}.$$

Let F_x denote the event that $G_{h,m}$ holds for all such h, m. By a union bound, we see that $\mathbb{P}F_x = 1 - O((\log x)^{2-2c})$. On this event F_x, for any integer y with $1 \leqslant y \leqslant x$, we have

$$\begin{aligned}
|\Delta(x, x+y)| &= \bigg| \sum_{2\sqrt{x} \leqslant 2^m \leqslant y} \Delta\Big(x + \lfloor y/2^{m+1} \rfloor 2^{m+1}, x + \lfloor y/2^m \rfloor 2^m\Big) \bigg| + O(\sqrt{x}) \\
&\leqslant \sum_{2\sqrt{x} \leqslant 2^m \leqslant y} x^{1/2}(\log x)^{c-1} + O(\sqrt{x}) \\
&\ll x^{1/2}(\log x)^c.
\end{aligned}$$

Since $2c - 2 > 1$, the Borel-Cantelli lemma implies that with probability one, F_{2^s} is true for all large integers s. On this event, $\Delta(2, x) \ll x^{1/2}(\log x)^c$ for all real $x \geqslant 2$, proving the theorem. □

Proof of Theorem 1.3. Fix $c \in [1/2, 1)$ and $\varepsilon > 0$. For integers $a \geqslant 2$, $b \geqslant 0$ and a tuple $\mathcal{H}$, define

$$\Delta(a, a+b; \mathcal{H}) := \sum_{a < n \leqslant a+b} \prod_{h \in \mathcal{H}} \mathbf{1}_{\mathcal{R}}(n+h) - \mathfrak{S}(\mathcal{H}) \int_a^{a+b} \frac{dt}{(\log t)^{|\mathcal{H}|}}.$$

Let

$$\lambda := 1 - \frac{1-c}{8c^2 - 2c}.$$

Let u be a large integer in terms of c and ε, and let F_u denote the event that

$$|\Delta(a, a+b; \mathcal{H})| \leqslant u^{\lambda + \varepsilon}$$

for all integers a, b satisfying $u \leqslant a \leqslant a + b \leqslant 2u$ and all admissible tuples $\mathcal{H}$ satisfying

$$|\mathcal{H}| = k \leqslant 10(\log u)^c, \quad \mathcal{H} \subset \Big[0, \exp\Big\{10(\log u)^{1-c}/\log_2 u\Big\}\Big]. \tag{4.4}$$

The number of such $\mathcal{H}$ does not exceed $u^{100/\log_2 u} = u^{o(1)}$ as $u \to \infty$.

We again invoke the moment bounds in Proposition 4.1. Assume $\mathcal{H}$ satisfies (4.4) and that $u \leqslant a \leqslant 2u$ and $\sqrt{a} \leqslant b \leqslant a$. It follows from (4.1) that

$$\mathbb{E}\Delta(a, a+b; \mathcal{H}) \ll \frac{bu^{o(1)}}{a} + \frac{b}{a^{0.54}} \ll u^{0.46},$$

and inequality (4.2) implies

$$\mathbb{V}\Delta(a, a+b; \mathcal{H}) \ll b^{1+\frac{4c^2-1}{4c^2-c}+o(1)} a^{o(1)} \ll bu^{2\lambda-1+o(1)},$$

where the implied function $o(1)$ is uniform over all such $\mathcal{H}$, a and b. For integers h, m with $2\sqrt{u} \leqslant 2^m \leqslant u$ and $0 \leqslant h \leqslant u/2^m - 1$, let $G_{h,m}$ be the event that for all $\mathcal{H}$ satisfying (4.4),

$$|\Delta(u + h \cdot 2^m, u + (h+1) \cdot 2^m; \mathcal{H})| \leqslant u^{\lambda+\varepsilon/2}.$$

Again, if u is large enough, the expectation of the left side is at most $\frac{1}{2}u^{\lambda+\varepsilon/2}$, uniformly over all $h, m, \mathcal{H}$. By a union bound and Chebyshev's inequality,

$$\begin{aligned}\mathbb{P}\big(\cup_{h,m} (\text{not } G_{h,m})\big) &\leqslant \sum_{h,m} \sum_{\mathcal{H}} \mathbb{P}\big(\big|\Delta(u + h \cdot 2^m, u + (h+1) \cdot 2^m; \mathcal{H})\big| \geqslant \tfrac{1}{2}u^{\lambda+\varepsilon/2}\big) \\ &\ll \sum_{h,m} \sum_{\mathcal{H}} \frac{2^m}{u^{1+\varepsilon+o(1)}} \ll \frac{1}{u^{\varepsilon/2}}.\end{aligned}$$

Furthermore, as in the proof of Theorem 1.4, we see that if u is large enough (in terms of c, ε) and if $G_{h,m}$ holds for all h, m, then F_u holds. Therefore,

$$\mathbb{P}F_u = 1 - O\big(1/u^{\varepsilon/2}\big).$$

By Borel-Cantelli, almost surely F_{2^s} is true for all sufficiently large integers s.

Now assume that we are in the event that F_{2^s} holds for all $s \geqslant s_0$. Let x be sufficiently large such that $x \geqslant 2^{3s_0+1}$ and $2^{t-1} < x \leqslant t^s$, and let $\mathcal{H}$ be an admissible tuple with

$$k := |\mathcal{H}| \leqslant (\log x)^c, \qquad \mathcal{H} \subseteq \Big[0, \exp\Big\{\frac{(\log x)^{1-c}}{\log_2 x}\Big\}\Big].$$

Note that whenever $x^{1/3} \leqslant u = 2^s \leqslant x$ we have (4.4). Thus, using (3.2),

$$\begin{aligned}\Big|\sum_{n\leqslant x} \prod_{h\in\mathcal{H}} \mathbf{1}_{\mathcal{R}}(n+h) - \mathfrak{S}(\mathcal{H}) \int_2^x \frac{dt}{\log^{|\mathcal{H}|} t}\Big| &\leqslant O(x^{1/3+o(1)}) + \\ &+ \sum_{x^{1/3} < 2^s \leqslant x/2} |\Delta(2^s, 2^{s+1}; \mathcal{H})| + |\Delta(2^{s+1}, x; \mathcal{H})| \\ &\ll x^{\lambda+\varepsilon/2},\end{aligned}$$

as required for Theorem 1.3. □

The following lemma is needed in the proof of Proposition 4.1. When an admissible tuple $\mathcal{H}$ is fixed, define

$$\psi_t := V_{\mathcal{H}}(z(t)).$$

LEMMA 4.2. *Let $2 \leqslant u \leqslant v \leqslant 3u$, and suppose $\mathcal{H}$ is an admissible tuple with $k := |\mathcal{H}| \geqslant 1$. Then*

$$\psi_u - \psi_v \ll k\psi_u \left(\frac{1}{u^{1/e^\gamma}} + \frac{v-u}{u \log u} \right).$$

Proof. We begin with the simple bound

$$\begin{aligned} \psi_u - \psi_v &= \psi_u \Bigg(1 - \prod_{z(u) < p \leqslant z(v)} (1 - \nu_p/p) \Bigg) \\ &\leqslant \psi_u \sum_{z(u) < p \leqslant z(v)} \frac{\nu(p)}{p} \\ &\leqslant k\psi_u \sum_{z(u) < p \leqslant z(v)} \frac{1}{p}. \end{aligned} \tag{4.5}$$

By multiple applications of (1.7),

$$\begin{aligned} \sum_{z(u) < p \leqslant z(v)} \frac{1}{p} &\leqslant \sum_{z(u) < p \leqslant z(v)} -\log(1 - 1/p) = \log \left(\Theta_{z(u)} / \Theta_{z(v)} \right) \\ &= \log \left(\frac{\log v}{\log u} \Big(1 + O(1/z(u)) \Big) \right) \\ &\ll \frac{1}{z(u)} + \log \left(1 + \frac{\log(v/u)}{\log u} \right) \\ &\ll \frac{1}{z(u)} + \frac{\log(v/u)}{\log u} \\ &\ll \frac{1}{u^{1/e^\gamma}} + \frac{v-u}{u \log u}. \end{aligned}$$

This completes the proof. □

Proof of Proposition 4.1. Suppose that $\mathcal{H} \subset [0, D]$ with $k := |\mathcal{H}| \leqslant \frac{\log x}{(\log_2 x)^2}$. We may assume that D is an integer. Write $\nu_p := |\mathcal{H} \bmod p|$ for every prime p. Since $z(t)$ is increasing and ψ_u is decreasing in u,

$$\psi_{n+D} \leqslant \mathbb{E} X_n \leqslant \psi_n.$$

Hence,

$$\mathbb{E} \sum_{x < n \leqslant x+y} X_n = \sum_{x < n \leqslant x+y} \psi_n + O\Bigg(\sum_{j=1}^{D} \big(\psi_{x+j} - \psi_{x+y+j} \big) \Bigg).$$

By Lemma 4.2 and the bound $\psi_u \ll 1/\log u$, the big-O term is

$$\ll \frac{kD}{\log x} \left(\frac{1}{x^{1/e^\gamma}} + \frac{y}{x \log x} \right) \ll \frac{kDy}{x \log^2 x},$$

since $y \geqslant \sqrt{x}$ and $1/e^\gamma > 1/2$. This proves that

$$\mathbb{E} \sum_{x < n \leqslant x+y} X_n = \sum_{x < n \leqslant x+y} \psi_n + O\left(\frac{kDy}{x \log^2 x} \right). \tag{4.6}$$

Lemma 3.5 implies that for each integer $n \in (x, x+y]$ we have

$$\psi_n = \frac{\mathfrak{S}(\mathcal{H})}{(\log n)^k}\left(1 + O(1/x^{0.55})\right) = \mathfrak{S}(\mathcal{H})\int_{n-1}^{n} \frac{dt}{(\log t)^k} + O\left(\frac{\mathfrak{S}(\mathcal{H})}{x^{0.55}}\right).$$

Estimate (3.2) implies that $\mathfrak{S}(\mathcal{H}) \leqslant x^{o(1)}$ and this proved the estimate (4.1) of the proposition.

For the second moment bound, let v be a parameter in $[4k, \log x]$ and set $Q := \prod_{p \leqslant v} p$. Given integers n_1 and n_2 with $x < n_1 < n_2 \leqslant x+y$, define m and b by

$$m := n_2 - n_1, \qquad b \equiv m \bmod Q \quad \text{with} \quad b \in [0, Q).$$

We consider separately the primes $\leqslant v$ and those $> v$, setting

$$\psi_n' := \prod_{v < p \leqslant z(n)} \left(1 - \frac{\nu_p}{p}\right), \qquad \xi_b := \prod_{p \leqslant v} \left(1 - \frac{|(\mathcal{H} \cup (\mathcal{H}+b)) \bmod p|}{p}\right).$$

Then

$$\begin{aligned} \mathbb{E}X_{n_1}X_{n_2} &\leqslant \prod_{p \leqslant z(n_1)} \left(1 - \frac{|(\mathcal{H} \cup (\mathcal{H}+m)) \bmod p|}{p}\right) \prod_{z(n_1) < p \leqslant z(n_2)} \left(1 - \frac{\nu_p}{p}\right) \\ &= \frac{\psi_{n_2}'}{\psi_{n_1}'}\, \xi_b \prod_{v < p \leqslant z(n_1)} \left(1 - \frac{|(\mathcal{H} \cup (\mathcal{H}+m)) \bmod p|}{p}\right). \end{aligned} \tag{4.7}$$

For technical reasons, we use the trivial bound $\mathbb{E}X_{n_1}X_{n_2} \leqslant \psi_{n_1} \leqslant \psi_x$ when $m \in \mathcal{H} - \mathcal{H}$; the total contribution from such terms is $\leqslant \psi_x k^2 y$, which is an acceptable error term for (4.2).

Now suppose that $m \notin \mathcal{H} - \mathcal{H}$. For any prime $p > v$ and integer $a \in (-p/2, p/2)$, let

$$\lambda_a(p) := |(\mathcal{H} \cap (\mathcal{H}+a)) \bmod p|.$$

Then, given $v < p \leqslant z(x+y)$ and m we have

$$|(\mathcal{H} \cup (\mathcal{H}+m)) \bmod p| = 2\nu_p - \lambda_a(p),$$

where a is the unique integer such that

$$a \equiv m \bmod p \quad \text{and} \quad |a| < p/2.$$

Clearly, $\lambda_a(p) \leqslant \nu_p \leqslant k$, and $\lambda_a(p) = 0$ unless $a \in (\mathcal{H} - \mathcal{H}) \cap (-p/2, p/2)$. In addition,

$$\sum_a \lambda_a(p) = \nu_p^2. \tag{4.8}$$

Consequently, for any $p > v$ we have

$$1 - \frac{|(\mathcal{H} \cup (\mathcal{H}+m)) \bmod p|}{p} = \left(1 - \frac{2\nu_p}{p}\right)(1 + f_a(p))$$

with

$$f_a(p) := \frac{\lambda_a(p)}{p - 2\nu_p}$$

We remark that $f_a(p) \in (0, 1]$ since $p > v \geqslant 4k \geqslant 4\nu_p \geqslant 4\lambda_a(p)$. For a fixed choice of $a \in \mathcal{H} - \mathcal{H}$ and fixed n_1, extend f_a to a multiplicative function supported on squarefree integers whose prime factors all lie in $I(n_1, a) := (\max\{v, 2|a|\}, z(n_1)]$.

If an integer r has a prime factor outside the interval $I(n_1, a)$ or r is not squarefree, we set $f_a(r) := 0$. Then

$$\begin{aligned}\prod_{v<p\leqslant z(n_1)} \Big(1 - \frac{|(\mathcal{H} \cup (\mathcal{H}+m)) \bmod p|}{p}\Big) &= \prod_{v<p\leqslant z(n_1)} \Big(1-\frac{2\nu_p}{p}\Big) \prod_{a\in\mathcal{H}-\mathcal{H}} \prod_{\substack{v<p\leqslant z(n_1)\\ p\,|\,m-a}} (1+f_a(p)) \\ &= \prod_{v<p\leqslant z(n_1)} \Big(1-\frac{2\nu_p}{p}\Big) \prod_{a\in\mathcal{H}-\mathcal{H}} \sum_{d_a\,|\,(m-a)} f_a(d_a)\end{aligned}$$

(since $m \notin \mathcal{H}-\mathcal{H}$, we always have $m-a \neq 0$). Recalling (4.7) we obtain that

$$\mathbb{E} X_{n_1} X_{n_2} \leqslant \psi'_{n_1}\psi'_{n_2}\xi_b \prod_{v<p\leqslant z(n_1)} \Big(\frac{p^2-2p\nu_p}{(p-\nu_p)^2}\Big) S(n_1,n_2), \tag{4.9}$$

where

$$S(n_1,n_2) := \prod_{a\in\mathcal{H}-\mathcal{H}} \sum_{d_a\,|\,(m-a)} f_a(d_a).$$

We now fix n_1 and sum over n_2. Let

$$\begin{aligned}\mathcal{D}(n_1) := \big\{\mathbf{d} = (d_a)_{a\in\mathcal{H}-\mathcal{H}} : \exists\, m\in[1,y]\setminus(\mathcal{H}-\mathcal{H}) \text{ such that } \forall\, a,\ d_a \mid (m-a),\\ \text{each } d_a \text{ is squarefree with all of its prime factors in } I(n_1,a)\big\},\end{aligned}$$

i.e., $\mathcal{D}(n_1)$ is the set of all possible vectors of the numbers d_a. We compute

$$\sum_{\substack{n_1<n_2\leqslant x+y\\ n_2-n_1\notin\mathcal{H}-\mathcal{H}}} \psi'_{n_2}\xi_b\, S(n_1,n_2) \leqslant \sum_{\mathbf{d}\in\mathcal{D}(n_1)} \Big(\prod_a f_a(d_a)\Big) \sum_{b \bmod Q} \xi_b \sum_{\substack{n_1<n_2\leqslant x+y\\ n_2\equiv n_1+b \bmod Q\\ \forall a,\ n_2\equiv n_1+a \bmod d_a}} \psi'_{n_2},$$

where we have dropped the condition $n_2-n_1\notin\mathcal{H}-\mathcal{H}$ on the right side. A crucial observation is that for every $\mathbf{d}\in\mathcal{D}(n_1)$, the components d_a are pairwise coprime. Indeed, if a, a' are two distinct elements of $\mathcal{H}-\mathcal{H}$ and a prime $p > \max\{v, 2|a|, 2|a'|\}$ divides both d_a and $d_{a'}$, then there is some $m\in[1,y]\setminus(\mathcal{H}-\mathcal{H})$ so that $p \mid d_a \mid (m-a)$ and $p \mid d_{a'} \mid (m-a')$. This implies $a\equiv a' \pmod p$, a contradiction. Hence, the innermost sum is a sum over a single residue class modulo $d := Q\prod_a d_a$. For any $e\in\mathbb{Z}$ we have by (4.5) that

$$\begin{aligned}\sum_{\substack{n_1<n\leqslant x+y\\ n\equiv e \bmod d}} \psi'_n &= \sum_{\substack{n_1<n\leqslant x+y\\ n\equiv e \bmod d}} \Big[\frac{1}{d}(\psi'_n+\cdots+\psi'_{n+d-1}) + O\Big(k\psi'_x \sum_{z(n)<p\leqslant z(n+d)} \frac{1}{p}\Big)\Big] \\ &= O(\psi'_x) + \frac{1}{d}\sum_{n_1<n\leqslant x+y} \psi'_n,\end{aligned}$$

where we used that $k\leqslant \log x$ and

$$\sum_{z(x)<p\leqslant z(x+y+d)} \frac{1}{p} \ll \frac{1}{\log x}.$$

Therefore,

$$\sum_{\substack{n_1<n_2\leqslant x+y\\ n_2-n_1\notin\mathcal{H}-\mathcal{H}}}\psi'_{n_2}\xi_b\,S(n_1,n_2)\leqslant\frac{1}{Q}\sum_{b\bmod Q}\xi_b\sum_{n_1<n_2\leqslant x+y}\psi'_{n_2}\sum_{\mathbf{d}\in\mathcal{D}(n_1)}\prod_a\frac{f_a(d_a)}{d_a}\\ +O\bigg(\psi'_x\sum_{b\bmod Q}\xi_b\sum_{\mathbf{d}\in\mathcal{D}(n_1)}\prod_a f_a(d_a)\bigg). \tag{4.10}$$

Now (4.8) implies that

$$\begin{aligned}\sum_{b\bmod Q}\xi_b&=\prod_{p\leqslant v}\sum_{c=0}^{p-1}\bigg(1-\frac{|(\mathcal{H}\cup(\mathcal{H}+c))\bmod p|}{p}\bigg)\\&=\prod_{p\leqslant v}\bigg(p-2\nu_p+\frac{1}{p}\sum_a\lambda_a(p)\bigg)\\&=Q\prod_{p\leqslant v}\bigg(1-\frac{\nu_p}{p}\bigg)^2.\end{aligned}$$

Hence, combining (4.9) and (4.10), and reinserting terms with $n_2-n_1\in\mathcal{H}-\mathcal{H}$, for each n_1 we obtain that

$$\mathbb{E}\sum_{n_1<n_2\leqslant x+y}X_{n_1}X_{n_2}\leqslant\psi_{n_1}\sum_{n_1<n_2\leqslant x+y}\psi_{n_2}\prod_{v<p\leqslant z(n_1)}\bigg(\frac{p^2-2p\nu_p}{(p-\nu_p)^2}\bigg)\sum_{\mathbf{d}\in\mathcal{D}(n_1)}\prod_a\frac{f_a(d_a)}{d_a}\\ +O\bigg(\psi_x^2Q\sum_{\mathbf{d}\in\mathcal{D}(n_1)}\prod_a f_a(d_a)+\psi_xk^2\bigg).$$

Extending the first sum over $\mathbf{d}$ to all pairwise coprime tuples $\mathbf{d}$ composed of prime factors in $(v,z(n_1)]$, and applying (4.8) again, we find that

$$\begin{aligned}\sum_{\mathbf{d}\in\mathcal{D}(n_1)}\prod_a\frac{f_a(d_a)}{d_a}&\leqslant\prod_{v<p\leqslant z(n_1)}\bigg(1+\sum_a\frac{f_a(p)}{p}\bigg)\\&=\prod_{v<p\leqslant z(n_1)}\bigg(1+\frac{\nu_p^2}{p(p-2\nu_p)}\bigg).\end{aligned}$$

Finally, summing over n_1 we conclude that

$$\mathbb{E}\sum_{x<n_1<n_2\leqslant x+y}X_{n_1}X_{n_2}\leqslant\sum_{x<n_1<n_2\leqslant x+y}\psi_{n_1}\psi_{n_2}+O(\psi_xk^2y+\psi_x^2QTy),$$

where

$$T:=\max_{n_1}\sum_{\mathbf{d}\in\mathcal{D}(n_1)}\prod_a f_a(d_a).$$

Since $X_n^2=X_n$ we arrive at

$$\mathbb{E}\bigg(\sum_{x<n\leqslant x+y}X_n\bigg)^2\leqslant\mathbb{E}\sum_{x<n\leqslant x+y}X_n+\sum_{\substack{x<n_1,n_2\leqslant x+y\\ n_1\neq n_2}}\psi_{n_1}\psi_{n_2}+O(\psi_xk^2y+\psi_x^2QTy),$$

Comparing this with (4.6), it follows that the variance in question satisfies

$$\begin{aligned}\mathbb{V}\sum_{x<n\leqslant x+y} X_n &\leqslant \sum_{x<n\leqslant x+y}\big(\psi_n-\psi_n^2\big)+O\big(\psi_x k^2 y+\psi_x^2 QTy\big)+\\ &\qquad +O\bigg(\frac{yD}{x}\sum_{x<n\leqslant x+y}\psi_n+\frac{y^2D^2}{x^2}+\frac{yD}{x}\bigg)\\ &\ll y\psi_x+k^2y\psi_x+\psi_x^2QTy+\frac{y^2D}{x}\psi_x+\frac{y^2D^2}{x^2}+\frac{yD}{x}\\ &\ll k^2y\psi_x+\psi_x^2QTy+\frac{yD}{x}\Big[1+y(\psi_x+D/x)\Big].\end{aligned} \tag{4.11}$$

To bound T, we consider two cases. First, suppose that $k\leqslant(\log x)^{1/2}/\log_2 x$, and let $v:=4k$. In this case, we argue crudely, using (4.8) and $\nu_p\leqslant k$ for all p, obtaining

$$\begin{aligned}T&\leqslant\prod_{v<p\leqslant z(2x)}\Big(1+\sum_{|a|<p/2}f_a(p)\Big)\\ &=\prod_{4k<p\leqslant z(2x)}\bigg(1+\frac{k^2}{p-2k}\bigg)\\ &\leqslant\exp\big(k^2(\log_2 x-\log_2 k+O(1))\big)\ll e^{-k^2}(\log x)^{k^2}.\end{aligned}$$

The prime number theorem implies that $\log Q\ll v$ and thus $QT\ll(\log x)^{k^2}$. Therefore, (4.11) implies (4.2).

Next, suppose that

$$\frac{(\log x)^{1/2}}{\log_2 x}\leqslant k\leqslant\frac{\log x}{(\log_2 x)^2},\qquad\text{with}\quad k=(\log x)^{\varrho}, \tag{4.12}$$

and put

$$v:=\frac{4\log x}{\log_2 x}, \tag{4.13}$$

so that $v\geqslant 4k$ and $Q=x^{o(1)}$. For a parameter $U\leqslant x^5$, to be chosen later, let

$$\begin{aligned}\mathcal{D}_U^-&:=\big\{\mathbf{d}\in\mathcal{D}(n_1):\textstyle\prod d_a\leqslant U\big\},\\ \mathcal{D}_U^+&:=\big\{\mathbf{d}\in\mathcal{D}(n_1):\textstyle\prod d_a> U\big\}.\end{aligned}$$

We begin with $\mathcal{D}_U^-$. For any parameter $\alpha>0$ we have, by (4.8),

$$\begin{aligned}\sum_{\mathbf{d}\in\mathcal{D}_U^-}\prod_a f_a(d_a)&\leqslant U^\alpha\sum_{\mathbf{d}\in\mathcal{D}_U^-}\prod_a\frac{f_a(d_a)}{d_a^\alpha}\\ &\leqslant U^\alpha\prod_{v<p\leqslant z(2x)}\bigg(1+\frac{k^2}{p^\alpha(p-2k)}\bigg)\\ &\leqslant U^\alpha\exp\bigg\{2k^2\sum_{v<p\leqslant z(2x)}\frac{1}{p^{1+\alpha}}\bigg\}\\ &\leqslant U^\alpha\exp\bigg\{O\bigg(\frac{k^2}{\alpha v^\alpha\log v}\bigg)\bigg\}.\end{aligned}$$

Let

$$\alpha := 2\varrho - 1 + \frac{3\log_3 x}{\log_2 x},$$

so that $\frac{\log_3 x}{\log_2 x} \leqslant \alpha \leqslant 1$ by (4.12). Recalling (4.13), we see that

$$\alpha v^\alpha \log v \gg \alpha (\log_2 x)^{1-\alpha} (\log x)^\alpha \gg (\log x)^\alpha = k^2 (\log_2 x)^3 / \log x,$$

hence it follows that

$$\sum_{\mathbf{d}\in\mathcal{D}_U^-} \prod_a f_a(d_a) \leqslant U^{2\varrho-1} \exp\left\{ O\Big(\frac{\log x \log_3 x}{\log_2 x} \Big) \right\}. \tag{4.14}$$

Next, we turn to $\mathcal{D}_U^+$, and make use of the special structure of $\mathcal{D}(n_1)$. For any parameter $\beta \in [0,1)$ we have

$$\begin{aligned}
\sum_{\mathbf{d}\in\mathcal{D}_U^+} \prod_a f_a(d_a) &\leqslant U^{-\beta} \sum_{\mathbf{d}\in\mathcal{D}(n_1)} \prod_a (f_a(d_a) d_a^\beta) \\
&\leqslant U^{-\beta} \sum_{\substack{1\leqslant m\leqslant y \\ m\notin \mathcal{H}-\mathcal{H}}} \sum_{\substack{\mathbf{d}\in\mathcal{D}(n_1) \\ \forall a,\, d_a \mid (m-a)}} \prod_a (f_a(d_a) d_a^\beta) \\
&\leqslant U^{-\beta} \sum_{\substack{1\leqslant m\leqslant y \\ m\notin \mathcal{H}-\mathcal{H}}} \prod_{a\in\mathcal{H}-\mathcal{H}} \prod_{\substack{p\mid m-a \\ \max\{v,2|a|\}<p\leqslant z(2x)}} \left(1 + \frac{\lambda_a(p) p^\beta}{p - 2\nu_p}\right).
\end{aligned}$$

Note that each prime p can appear at most once in the double product, since $p \mid (m-a)$ and $p \mid (m-a')$ implies $p \mid (a-a')$, which forces $a = a'$. We split the last product into two pieces according to whether $p \leqslant w$ or $p > w$, where w is a parameter to be chosen later. For any $m \notin \mathcal{H} - \mathcal{H}$ we have

$$\begin{aligned}
\prod_{a\in\mathcal{H}-\mathcal{H}} \prod_{\substack{p\mid m-a \\ \max\{v,2|a|\}<p\leqslant w}} \left(1 + \frac{\lambda_a(p) p^\beta}{p - 2\nu_p}\right) &\leqslant \prod_{v<p\leqslant w} \big(1 + 2k p^{\beta-1}\big) \\
&\leqslant \exp\big\{2k w^\beta \log_2 x\big\}
\end{aligned}$$

for large x. We bound the contribution of larger primes trivially using the fact that any integer $m-a$ is divisible by $\ll \frac{\log x}{\log_2 x}$ such primes (here, it is crucial that $m \neq a$). Thus, for any $m \notin \mathcal{H} - \mathcal{H}$ we have

$$\prod_{a\in\mathcal{H}-\mathcal{H}} \prod_{\substack{p\mid m-a \\ \max\{w,2|a|\}<p\leqslant z(2x)}} \left(1 + \frac{\lambda_a(p) p^\beta}{p - 2\nu_p}\right) \leqslant \exp\left\{ O\Big(k^3 w^{\beta-1} \frac{\log x}{\log_2 x} \Big) \right\}.$$

We now put

$$w := k^2 \log x \qquad \text{and} \qquad \beta := \frac{1 - \varrho - 2\frac{\log_3 x}{\log_2 x}}{2\varrho + 1}.$$

By (4.12) we have $\beta \geqslant 0$, and clearly $\beta < 1$. It follows that

$$\sum_{\mathbf{d}\in\mathcal{D}_U^+} \prod_a f_a(d_a) \leqslant y U^{-\frac{1-\varrho}{2\varrho+1}} \exp\left\{ O\Big(\frac{\log x\ \log_3 x}{\log_2 x} \Big) \right\}. \tag{4.15}$$

Comparing (4.14) with (4.15), we choose U so that $U^{2\varrho-1} = yU^{-\frac{1-\varrho}{2\varrho+1}}$, that is,

$$U := y^{\frac{2\varrho+1}{4\varrho^2-\varrho}}.$$

Since $1/2 + o(1) \leqslant \rho \leqslant 1 + o(1)$, the exponent of y is $\leqslant 4 + o(1) \leqslant 5$ for large x. This gives

$$T \leqslant y^{\frac{4\varrho^2-1}{4\varrho^2-\varrho}} \exp\Big\{O\Big(\frac{\log x \log_3 x}{\log_2 x}\Big)\Big\}.$$

Inserting this into (4.11) yields the inequality (4.2), and completes the proof of Proposition 4.1. □

5. Random sieving by small primes

Throughout the sequel, we employ the notation

$$\Theta_z := \prod_{p \leqslant z} \Big(1 - \frac{1}{p}\Big) \qquad \text{and} \qquad \Theta_{z_1,z_2} := \prod_{z_1 < p \leqslant z_2} \Big(1 - \frac{1}{p}\Big) = \frac{\Theta_{z_2}}{\Theta_{z_1}}. \tag{5.1}$$

Throughout this section, we assume that x and y are large real numbers that satisfy

$$W_y \log y \in [\alpha(\log x)^2, \beta(\log x)^2], \tag{5.2}$$

where W_y is given by (1.11), and α, β are fixed with $0 < \alpha < \beta$. Note that (1.12) and (5.2) yield the estimates

$$(\log x)^2 \ll y \ll \frac{\log_2 x}{\log_3 x}(\log x)^2. \tag{5.3}$$

We adopt the convention that any constants implied by O and $\ll$ may depend on α, β but are independent of other parameters.

We define

$$\mathcal{S}_w(y) := [0, y] \cap \mathcal{S}_w$$

and when the value of y is clear from context we put

$$S_w := |\mathcal{S}_w(y)|.$$

Using a variety of tools, we give sharp probability bounds for S_w at five different "checkpoint" values $w_1 < w_2 < w_3 < w_4 < w_5$ (defined below), with each $S_{w_{i+1}}$ controlled in terms of S_{w_i} for $i = 1, 2, 3, 4$. Our arguments are summarised as follows, where the range is a range of primes:

Range	Estimation technique
$[2, w_1]$	Lower bound by W_y (5.4)
$(w_1, w_2]$	Buchstab identity, sieve upper bound (Lemma 5.1)
$(w_2, w_3]$	Buchstab identity, large sieve, Bennett inequality (Lemma 5.2)
$(w_3, w_4]$	Martingale interpretation, Azuma inequality (Lemma 5.3)
$(w_4, w_5]$	Graph interpretation, combinatorial expansion (Lemma 6.1)
$(w_5, z]$	Combinatorial expansion (Lemmas 6.3, 6.5, Corollary 6.4)

The most delicate part of the argument is dealing with primes p near $\log x$, that is, $w_1 \leqslant p \leqslant w_3$ (see Lemmas 5.1 and 5.2). To initialize the argument, we observe from definition (1.11) of W_y that we have the lower bound

$$S_{w_1} \geqslant W_y. \tag{5.4}$$

Now we successively increase the sieving range from S_{w_1} to S_{w_2}, and so on, up to S_{w_5}.

LEMMA 5.1 (Sieving for $w_1 < p \leqslant w_2$). *Let $w_1 := (y/\log y)^{1/2}$ and $w_2 := \log x \log_3 x$. With probability one, we have*

$$S_{w_2} = \left(1 + O\left(\frac{\log_4 x}{\log_3 x}\right)\right) S_{w_1}.$$

Proof. In this section and the next one, we adopt the notation R_p for the residue class $a_p \bmod p$. From the Buchstab identity

$$S_{w_2} = S_{w_1} - \sum_{w_1 < p \leqslant w_2} |\mathcal{S}_{p-1}(y) \cap R_p|$$

we have

$$S_{w_1} \geqslant S_{w_2} \geqslant S_{w_1} - \sum_{w_1 < p \leqslant w_2} |\mathcal{S}_{w_1}(y) \cap R_p|. \tag{5.5}$$

The sieve upper bound (Lemma 3.1) and Mertens' theorem together imply that

$$\sum_{w_1 < p \leqslant w_2} |\mathcal{S}_{w_1}(y) \cap R_p| \ll \frac{y}{\log y} \log\left(\frac{\log w_2}{\log w_1}\right) = S_{w_1} C_y \log\left(\frac{\log w_2}{\log w_1}\right), \tag{5.6}$$

where

$$C_y := \frac{y}{S_{w_1} \log y}.$$

By (5.2) and (5.3) we have

$$C_y \leqslant \frac{y}{W_y \log y} \ll \frac{\log_2 x}{\log_3 x}. \tag{5.7}$$

Using (5.2) and the lower bound $w_1^2 = S_{w_1} C_y \geqslant W_y C_y$ we see that

$$\log w_1 \geqslant \log_2 x - \tfrac{1}{2}(\log_2 y - \log C_y) + O(1),$$

hence

$$\begin{aligned} \log\left(\frac{\log w_2}{\log w_1}\right) &\leqslant \log\left(\frac{\log_2 x + \log_4 x}{\log_2 x - \frac{1}{2}(\log_2 y - \log C_y) + O(1)}\right) \\ &\ll \frac{\log_2 y - \log C_y}{\log_2 x} \ll \frac{\log_3 x - \log C_y}{\log_2 x}. \end{aligned}$$

Inserting this bound into (5.6) we find that

$$\sum_{w_1 < p \leqslant w_2} |\mathcal{S}_{w_1}(y) \cap R_p| \ll S_{w_1} \frac{C_y(\log_3 x - \log C_y)}{\log_2 x}.$$

The function $z(\log_3 x - \log z)$ is increasing for $z \leqslant e^{-1} \log_2 x$, hence by (5.7) we have

$$\sum_{w_1 < p \leqslant w_2} |\mathcal{S}_{w_1}(y) \cap R_p| \ll S_{w_1} \frac{\log_4 x}{\log_3 x}$$

and the stated result follows from (5.5). □

LEMMA 5.2 (Sieving for $w_2 < p \leqslant w_3$). *Let $w_2 := \log x \log_3 x$ and $w_3 := \log x\,(\log_2 x)^2$. Conditional on $\mathcal{A}_{w_2}$ satisfying $S_{w_2} \geqslant \frac{1}{2}W_y$, we have*

$$\mathbb{P}_{w_2,w_3}\biggl(S_{w_3} \leqslant \biggl(1 - \frac{1}{\log_3 x}\biggr)S_{w_2}\biggr) \ll x^{-100}.$$

Proof. As in the previous lemma, we start with

$$S_{w_3} \geqslant S_{w_2} - \sum_{w_2<p\leqslant w_3} |\mathcal{S}_{w_2}(y) \cap R_p|. \tag{5.8}$$

Let $X_p := |\mathcal{S}_{w_2}(y) \cap R_p| - p^{-1}S_{w_2}$ for each prime $p \in (w_2, w_3]$. The variables X_p are independent and have a mean value of zero, and by the sieve upper bound (Lemma 3.1) it follows that

$$|X_p| \ll \frac{y}{p\log y} \ll \frac{y}{w_2 \log_2 x},$$

hence

$$|X_p| \leqslant M := \frac{c\,y}{\log x\,\log_2 x\,\log_3 x} \qquad (w_2 < p \leqslant w_3) \tag{5.9}$$

for some absolute constant $c > 0$. Using Montgomery's Large Sieve inequality (see [12, Equation (9.18)] or [27]),

$$\sum_{w_2<p\leqslant w_3} p^2\,\mathbb{V}X_p = \sum_{w_2<p\leqslant w_3} p \sum_{a\in\mathbb{Z}/p\mathbb{Z}} \Bigl(\bigl|\mathcal{S}_{w_2}(y) \cap (a \bmod p)\bigr| - p^{-1}S_{w_2}\Bigr)^2 \leqslant 2w_3^2\,S_{w_2},$$

which implies that

$$\sigma^2 := \sum_{w_2<p\leqslant w_3} \mathbb{V}X_p \leqslant 2w_2^{-2}w_3^2\,S_{w_2} \ll \frac{(\log_2 x)^4}{(\log_3 x)^2}S_{w_2}. \tag{5.10}$$

We apply Bennett's inequality (Lemma 3.3) with $t := S_{w_2}/(2\log_3 x)$. By (5.9), (5.10) and (5.3), we have

$$\frac{Mt}{\sigma^2} \gg \frac{y}{\log x\,(\log_2 x)^5} \gg \frac{\log x}{(\log_2 x)^5},$$

and therefore

$$\frac{\sigma^2}{M^2}\mathscr{L}\Bigl(\frac{Mt}{\sigma^2}\Bigr) \gg \frac{t}{M}\log\Bigl(\frac{Mt}{\sigma^2}\Bigr) \gg \frac{S_{w_2}\log x\,(\log_2 x)^2}{y} \gg \log x\,\log_3 x,$$

where the last bound follows from (1.12) and our assumption that $S_{w_2} \geqslant \frac{1}{2}W_y$. Lemma 3.3 now shows that for some constant $c' > 0$,

$$\mathbb{P}\biggl(\Bigl|\sum_{w_2<p\leqslant w_3} X_p\Bigr| \geqslant \frac{S_{w_2}}{2\log_3 x}\biggr) \leqslant 2\exp\bigl\{-c'\log x\,\log_3 x\bigr\} \ll x^{-100}.$$

Thus, with probability at least $1 - O(x^{-100})$ we have

$$\sum_{w_2<p\leqslant w_3} \bigl|\mathcal{S}_{w_2}(y) \cap R_p\bigr| \leqslant S_{w_2}\biggl(\frac{1}{2\log_3 x} + \sum_{w_2<p\leqslant w_3}\frac{1}{p}\biggr) \leqslant \frac{S_{w_2}}{\log_3 x}$$

for sufficiently large x. Recalling (5.8), the proof is complete. □

LEMMA 5.3 (Sieving for $w_3 < p \leqslant w_4$). *Let $w_3 := \log x\,(\log_2 x)^2$ and $w_4 := y^{4/3}$. Conditional on $\mathcal{A}_{w_3}$ satisfying $S_{w_3} \geqslant \frac{1}{4}W_y$, we have*

$$\mathbb{P}_{w_3,w_4}\left(\left|S_{w_4} - \tfrac{3}{8}S_{w_3}\right| \geqslant \frac{S_{w_3}}{(\log_2 x)^{1/2}}\right) \ll x^{-100}.$$

Proof. Let $p_0 := w_3$ and let $p_1 < \ldots < p_m$ be the primes in $(w_3, w_4]$. Using the notation (5.1), we define random variables by

$$X_j := \Theta_{w_3,p_j}^{-1} S_{p_j} \qquad (j = 0, 1, \ldots, m).$$

The sequence $X_0, X_1, \ldots, X_m$ is a martingale since

$$\mathbb{E}(X_{j+1}|X_j) = \Theta_{w_3,p_{j+1}}^{-1}\mathbb{E}(S_{p_{j+1}}|\mathcal{A}_{p_j}) = \Theta_{w_3,p_{j+1}}^{-1}\left(1 - p_{j+1}^{-1}\right) S_{p_j} = X_j.$$

Note that

$$X_0 = S_{w_3} \geqslant \tfrac{1}{4}W_y \gg \frac{y \log_3 x}{(\log_2 x)^2}, \tag{5.11}$$

where we have used (1.12) in the last step.

We apply Azuma's inequality (Lemma 3.2). If $p_{j+1} > y$, then $|X_{j+1} - X_j| \ll 1$ since $\Theta_{w_3,p_j}^{-1} \ll 1$. In the case that $p_{j+1} \leqslant y$, Lemma 3.1 shows that for any value of $R_{p_{j+1}}$ we have

$$\begin{aligned}|X_{j+1} - X_j| &= \Theta_{w_3,p_j}^{-1}\left|\left(1 - p_{j+1}^{-1}\right)^{-1} S_{p_{j+1}} - S_{p_j}\right| \ll \frac{S_{p_{j+1}}}{p_{j+1}} + S_{p_j} - S_{p_{j+1}} \\ &= \frac{S_{p_{j+1}}}{p_{j+1}} + \left|\mathcal{S}_{p_j}(y) \cap R_{p_{j+1}}\right| \ll \frac{y/p_{j+1}}{1 + \log(y/p_{j+1})}.\end{aligned}$$

Consequently,

$$\sum_{j=0}^{m-1} |X_{j+1} - X_j|^2 \ll \frac{y^2}{w_3 \log w_3 \log^2 y} + y^{4/3} \ll \frac{y^2}{\log x\, \log_2^5 x}.$$

Thus, if $c > 0$ is sufficiently small, then Lemma 3.2 shows that

$$\mathbb{P}_{w_3,w_4}\left(|X_m - X_0| \geqslant \frac{X_0}{(\log_2 x)^{1/2}}\right) \ll \exp\left\{-\frac{c\, X_0^2 \log x\, (\log_2 x)^4}{y^2}\right\} \ll x^{-100} \tag{5.12}$$

since by (5.11) we have

$$\frac{X_0^2 \log x\, (\log_2 x)^4}{y^2} \gg \log x\, (\log_3 x)^2.$$

Using (5.1) and (5.3) we write

$$\lambda := \Theta_{w_3,w_4}^{-1} = \tfrac{8}{3}(1 + r_x) \qquad \text{with} \quad r_x \ll \frac{\log_3 x}{\log_2 x};$$

then noting that

$$\left|S_{w_4} - \tfrac{3}{8}S_{w_3}\right| = \left|\lambda^{-1}X_m - \tfrac{3}{8}X_0\right| = \lambda^{-1}|X_m - (1 + r_x)X_0|,$$

for any $Z > 0$ we have

$$\mathbb{P}_{w_3,w_4}\big(\left|S_{w_4} - \tfrac{3}{8}S_{w_3}\right| \geqslant Z\big) \leqslant \mathbb{P}_{w_3,w_4}\big(\left|X_m - X_0\right| \geqslant \lambda Z - r_x X_0\big).$$

In view of (5.12) this implies that

$$\mathbb{P}_{w_3,w_4}\big(\big|S_{w_4} - \tfrac{3}{8}S_{w_3}\big| \geqslant Z\big) \ll x^{-100}$$

holds provided that

$$\lambda Z - r_x X_0 \geqslant \frac{X_0}{(\log_2 x)^{1/2}}.$$

The result follows by taking $Z := \frac{X_0}{(\log_2 x)^{1/2}} = \frac{S_{w_3}}{(\log_2 x)^{1/2}}$ and noting that $\lambda \geqslant 2$. $\square$

6. Random sieving by large primes

In this section, we adopt the notation

$$S_w := |\mathcal{S}_w(y)| = |[0,y] \cap \mathcal{S}_w|$$

from the previous section; however, we *do not* assume inequalities (5.2) and (5.3), except in Corollary 6.2 below. We do assume that y is sufficiently large. Sieving by large primes ($p > y^4$, say) is easier because there is a relatively low probability that $\mathcal{S} \cap R_p \neq \varnothing$ and we are able to deploy combinatorial methods.

Lemma 6.1 (Sieving for $w_4 < p \leqslant w_5$). *Let v be a real number greater than $w_4 := y^{4/3}$, and let $\vartheta \in [y^{-1/4}, 1)$. Conditional on $\mathcal{A}_{w_4}$, we have*

$$\mathbb{P}_{w_4,v}\Big(\big|S_v - \Theta_{w_4,v} S_{w_4}\big| \geqslant \vartheta S_{w_4}\Big) \leqslant \exp\{-0.1\vartheta^2 S_{w_4}\}.$$

Proof. Put $\mathcal{S} := \mathcal{S}_{w_4}(y)$, $\ell := |\mathcal{S}| = S_{w_4}$, and let $\mathcal{P}$ be the set of primes in $(w_4, v]$. The random residue classes $\{R_p : p \in \mathcal{P}\}$ give rise to a bipartite graph $\mathcal{G}$ that has vertex sets $\mathcal{S}$ and $\mathcal{P}$, with edges connecting the vertices $s \in \mathcal{S}$ and $p \in \mathcal{P}$ if and only if $s \in R_p$ (i.e., $s \equiv a_p \bmod p$). Since $0 \leqslant s \leqslant y < w_4$, for every p there is at most one vertex s joined to it. For any $s \in \mathcal{S}$, let $d(s)$ be its degree,

$$d(s) := \big|\{p \in \mathcal{P} : s \in R_p\}\big|,$$

and let $\mathcal{S}^+$ be the set of vertices in $\mathcal{S}$ of positive degree:

$$\mathcal{S}^+ := \{s \in \mathcal{S} : d(s) > 0\} = \bigcup_{p \in \mathcal{P}} (\mathcal{S} \cap R_p).$$

Finally, we denote by $\mathbf{d}$ the vector $\big\langle d(s) : s \in \mathcal{S}^+ \big\rangle$. In this manner, the random residue classes $\{R_p : p \in \mathcal{P}\}$ determine a subset $\mathcal{S}^+ \subset \mathcal{S}$ and a vector $\mathbf{d}$.

For any subset $\mathcal{T} = \{t_1, \dots, t_m\}$ in $\mathcal{S}$ and a vector $\mathbf{r} = \langle r_1, \dots, r_m \rangle$ whose entries are positive integers, let $E(\mathcal{T}, \mathbf{r})$ be the event that the random graph $\mathcal{G}$ described above has $\mathcal{S}^+ = \mathcal{T}$ and $\mathbf{d} = \mathbf{r}$. Since $\mathcal{S} \subset [0,y]$ and $w_4 > y$, we have $|\mathcal{S} \cap R_p| \leqslant 1$ for all $p \in \mathcal{P}$, and thus

$$h := r_1 + \cdots + r_m = \sum_{s \in \mathcal{S}^+} d(s) = \big|\{p \in \mathcal{P} : \mathcal{S} \cap R_p \neq \varnothing\}\big|.$$

Fixing the primes $p_1, \dots, p_h \in \mathcal{P}$ with $R_p \cap \mathcal{S} \neq \varnothing$, there are $\binom{h}{r_1 \cdots r_m}$ ways to choose the graph's edges connecting the p_i to $\mathcal{T}$. Consequently,

$$\begin{aligned}\mathbb{P}_{w_4,v}(E(\mathcal{T},\mathbf{r})) &= \sum_{\substack{p_1,\dots,p_h\in\mathcal{P}\\ p_1<\dots<p_h}} \frac{1}{p_1\cdots p_h}\binom{h}{r_1\ r_2\ \cdots\ r_m} \prod_{p\in\mathcal{P}\setminus\{p_1,\dots,p_h\}}\Big(1-\frac{\ell}{p}\Big)\\ &= \binom{h}{r_1\ r_2\ \cdots\ r_m}\prod_{p\in\mathcal{P}}\Big(1-\frac{\ell}{p}\Big)\sum_{\substack{p_1,\dots,p_h\in\mathcal{P}\\ p_1<\dots<p_h}}\prod_{j=1}^{h}\frac{1}{p_j-\ell}. \end{aligned} \tag{6.1}$$

Relaxing the conditions on the last sum in (6.1), we find that

$$\mathbb{P}_{w_4,v}(E(\mathcal{T},\mathbf{r})) \leqslant \frac{VU^h}{r_1!\cdots r_m!} \qquad \text{with} \quad V := \prod_{p\in\mathcal{P}}\Big(1-\frac{\ell}{p}\Big) \quad \text{and} \quad U := \sum_{p\in\mathcal{P}}\frac{1}{p-\ell}.$$

For fixed m, there are $\binom{\ell}{m}$ choices for $\mathcal{T}$; thus, summing over all $r_1, \dots, r_m$ we conclude that

$$\mathbb{P}_{w_4,v}(S_{w_4}-S_v=m) \leqslant V\binom{\ell}{m}(e^U-1)^m. \tag{6.2}$$

The complete sum over m of the right side of (6.2) is equal to $Ve^{U\ell}$, and the peak occurs when $m = (1-e^{-U})\ell + O(1)$. We also have

$$1-e^{-U} = 1-\Theta_{w_4,v}\bigg(1+O\bigg(\frac{\ell}{w_4\log w_4}\bigg)\bigg), \tag{6.3}$$

Standard large-deviation results for the binomial distribution (such as Lemma 3.2) imply that for any $\delta > 0$,

$$e^{-U\ell}\sum_{|m-(1-e^{-U})\ell|\geqslant\delta\ell}\binom{\ell}{m}(e^U-1)^m \leqslant 2e^{-\delta^2\ell/2}.$$

Recalling that $\ell := S_{w_4}$, we see that the inequality

$$\big|S_v - \Theta_{w_4,v}\ell\big| \geqslant \vartheta\ell$$

implies via (6.3) that

$$|m-(1-e^{-U})\ell| \geqslant \vartheta\ell - |e^{-U}-\Theta_{w_4,v}|\ell \geqslant \vartheta\ell - O(y^{-1/3}\ell) \geqslant \vartheta\ell/2$$

for all large x since $w_4 := y^{4/3}$ and $\ell \leqslant y$. Combining our results above, we conclude that

$$\begin{aligned}\mathbb{P}_{w_4,v}\left(\big|S_v-\Theta_{w_4,v}\ell\big|\geqslant\vartheta\ell\right) &\ll Ve^{U\ell}e^{-\vartheta^2\ell/8}\\ &\ll e^{-\vartheta^2\ell/8+O(\ell^2/w_4)}\\ &\leqslant e^{-\vartheta^2\ell/10}\end{aligned}$$

for all large x, and the proof is complete. □

Combining Lemmas 5.1, 5.2, 5.3 and 6.1 (with $v := y^8$ and $\vartheta := y^{-1/10}$) we obtain the following result.

COROLLARY 6.2 (Sieving for $w_1 < p \leqslant w_5$). *Assume* (5.2), *let* $w_1 := (y/\log y)^{1/2}$ *and* $w_5 := y^8$. *Conditional on* $\mathcal{A}_{w_1}$, *we have with probability* $1 - O(x^{-100})$ *that*

$$\left| S_{w_5} - \frac{S_{w_1}}{16} \right| \ll_{\alpha,\beta} \frac{\log_4 x}{\log_3 x} S_{w_1}.$$

Our next result is a very general tool for handling primes larger than y^4.

LEMMA 6.3 (Sieving for $w_5 < p \leqslant z$, I). *Let* $w \geqslant y^4$ *and* $\mathcal{P}$ *be a set of primes larger than* w *such that* $\sum_{p\in\mathcal{P}} 1/p \geqslant 1/10$. *Let* $\mathcal{S} \subseteq \mathcal{S}_w$ *with* $|\mathcal{S}| \leqslant 10y$, *and such that for all* $p \in \mathcal{P}$, $\mathcal{S}$ *is distinct modulo* p. *Conditional on* $\mathcal{A}_w$, *we have for all* $0 \leqslant g \leqslant |\mathcal{S}|$*:*

$$\mathbb{P}_{\mathcal{P}}\left(\left| \mathcal{S} \setminus \bigcup_{p\in\mathcal{P}} R_p \right| = g \right) = (1-\Theta)^{|\mathcal{S}|-g} \Theta^g \binom{|\mathcal{S}|}{g} (1 + O(y^2/w)),$$

where

$$\Theta := \prod_{p\in\mathcal{P}} (1 - 1/p).$$

Proof. Put $\ell := |\mathcal{S}|$, and assume that $\ell \geqslant 1$ (the case $\ell := 0$ being trivial). Take $m := \ell - g$, and let $\mathcal{T}$, $\mathbf{r}$, $E(\mathcal{T}, \mathbf{r})$ and h be defined as in Lemma 6.1 with $|\mathcal{T}| = m = \ell - g$. As before (see (6.1)) we have

$$\mathbb{P}_{\mathcal{P}}(E(\mathcal{T}, \mathbf{r})) = \binom{h}{r_1\ r_2\ \cdots\ r_m} \prod_{p\in\mathcal{P}} \left(1 - \frac{\ell}{p}\right) \sum_{\substack{p_1,\dots,p_h\in\mathcal{P} \\ p_1<\cdots<p_h}} \prod_{j=1}^{h} \frac{1}{p_j - \ell}. \tag{6.4}$$

For any prime $p \in \mathcal{P}$ the elements of $\mathcal{S}$ lie in distinct residue classes modulo p; this implies that $\mathbb{P}_{\mathcal{P}}(E(\mathcal{T}, \mathbf{r}))$ can only be nonzero when $m = h$ in (6.4) (that is, every $r_j := 1$). Let T_h be the sum over $p_1, \dots, p_h$ in (6.4). Then

$$\begin{aligned} T_h &= \frac{1}{h!}\left(\sum_{p\in\mathcal{P}} \frac{1}{p-\ell} + O\left(\frac{h}{w}\right)\right)^h \\ &= \frac{1}{h!}\left(\sum_{p\in\mathcal{P}} \frac{1}{p} + O\left(\frac{\ell}{w}\right)\right)^h \\ &= \frac{(-\log\Theta + O(y/w))^h}{h!}. \end{aligned}$$

Also, note that

$$V := \prod_{p\in\mathcal{P}} \left(1 - \frac{\ell}{p}\right) = \Theta^\ell (1 + O(y^2/w)).$$

Hence, summing over all vectors $\mathbf{r}$, we find that

$$\begin{aligned}\mathbb{P}_{\mathcal{P}}\big(\big|\mathcal{S}\setminus\cup_{p\in\mathcal{P}}R_p\big|=\ell-m\big) &= \sum_{\substack{\mathcal{T}\subset\mathcal{S}\\|\mathcal{T}|=m}}\sum_{h}\sum_{r_1+\cdots+r_m=h}\binom{h}{r_1\ \cdots\ r_m}VT_h\\ &= (1+O(y^2/w))\Theta^\ell\sum_{\substack{\mathcal{T}\subset\mathcal{S}\\|\mathcal{T}|=m}}\sum_{r_1,\dots,r_m\geqslant 1}\frac{(-\log\Theta+O(y/w))^{r_1+\cdots+r_m}}{r_1!\cdots r_m!}\\ &= (1+O(y^2/w))\binom{\ell}{m}\Theta^\ell\big(e^{-\log\Theta+O(y/w)}-1\big)^m\\ &= (1+O(y^2/w))\binom{\ell}{m}\Theta^\ell\big(\Theta^{-1}-1\big)^{\ell-g}.\end{aligned}$$

This completes the proof. □

Corollary 6.4 (Sieving for $w_5<p\leqslant z$, II). *Uniformly for $z^{1/2}\geqslant w\geqslant y^4$, we have*

$$\mathbb{E}_{w,z}\binom{S_z}{k}=\Theta_{w,z}^k\binom{S_w}{k}(1+O(y^2/w)).$$

Proof. Let $\Theta:=\Theta_{w,z}$. By Lemma 6.3 with $\mathcal{S}:=\mathcal{S}_w\cap[0,y]$ and $\mathcal{P}$ the set of primes in $(w,z]$, we have

$$\begin{aligned}\mathbb{E}_{w,z}\binom{S_z}{k} &= (1+O(y^2/w))\sum_{g=k}^{S_w}(1-\Theta)^{S_w-g}\Theta^g\binom{S_w}{g}\binom{g}{k}\\ &= (1+O(y^2/w))\Theta^k\binom{S_w}{k}\sum_{j=0}^{S_w-k}(1-\Theta)^{S_w-k-j}\Theta^j\binom{S_w-k}{S_w-k-j}\\ &= (1+O(y^2/w))\Theta^k\binom{S_w}{k}.\end{aligned}$$

□

The next lemma has a weaker conclusion than Lemma 6.3 but is more general and is needed for a second moment argument below in which we derive a lower bound for the largest prime gap in $[0,x]$.

Lemma 6.5 (Sieving for $w_5<p\leqslant z$, III). *Let w and z be real numbers for which $z^{1/2}\geqslant w\geqslant y^8$. Let $\mathcal{S}\subset\mathcal{S}_w\cap[0,e^y]$ with $|\mathcal{S}|\leqslant y$ and such that for every prime $p>w$, no more than two numbers in $\mathcal{S}$ lie in any given residue class modulo p. Then*

$$\mathbb{P}_{w,z}\Big(\mathcal{S}\cap\mathcal{S}_z=\varnothing\Big)=(1-\Theta_{w,z})^{|\mathcal{S}|}(1+O(y^4/w)).$$

Proof. Put $\ell:=|\mathcal{S}|$, and let $\mathcal{P}$ be the set of primes in $(w,z]$, and put

$$\mathcal{Q}:=\big\{p\in\mathcal{P}:p\mid s-s'\text{ for some }s,s'\in\mathcal{S},s\neq s'\big\}.$$

Note that the bound

$$|\mathcal{Q}|\leqslant\frac{\ell^2y}{\log w}\leqslant y^3 \tag{6.5}$$

holds if y is large enough.

By assumption, for every $p \in \mathcal{Q}$, $|\mathcal{S} \cap R_p| \leqslant 2$. Let E_m be the event that for $\mathcal{S} \cap R_p \neq \varnothing$ holds for precisely m primes $p \in \mathcal{Q}$. Since for any prime $p \in \mathcal{P}$ the probability that $\mathcal{S} \cap R_p \neq \varnothing$ does not exceed ℓ/p, using (6.5) we have

$$\mathbb{P}_{\mathcal{Q}}(E_m) \leqslant \frac{1}{m!}\Big(\sum_{p\in\mathcal{Q}} \frac{\ell}{p}\Big)^m \leqslant \left(\frac{e\ell|\mathcal{Q}|}{mw}\right)^m \leqslant (ey^4/w)^m \qquad (m \geqslant 1). \tag{6.6}$$

Assume the event E_m occurs, and fix $\mathcal{A}_{\mathcal{Q}}$. If $\mathcal{S}$ has precisely n elements covered by $\bigcup_{p\in\mathcal{Q}} R_p$, then $0 \leqslant n \leqslant 2m$, the upper bound being a consequence of our hypothesis on $\mathcal{S}$. Put

$$\mathcal{S}' := \big\{s \in \mathcal{S} : s \notin R_p \text{ for all } p \in \mathcal{Q}\big\},$$

so that $|\mathcal{S}'| = \ell - n$. Lemma 6.3 implies that

$$\begin{aligned}\mathbb{P}_{\mathcal{P}\setminus\mathcal{Q}}\Big(\mathcal{S}' \subset \bigcup_{p\in\mathcal{P}\setminus\mathcal{Q}} R_p\Big) &= (1 + O(y^2/w))\Big(1 - \Theta_{w,z}\prod_{p\in\mathcal{Q}}\big(1-p^{-1}\big)^{-1}\Big)^{\ell-n}\\ &= (1 + O(y^4/w))\Big(1 - \Theta_{w,z}\Big)^{\ell-n}\\ &\ll \Big(1 - \Theta_{w,z}\Big)^{\ell-2m},\end{aligned}$$

since

$$\prod_{p\in\mathcal{Q}}\big(1-p^{-1}\big)^{-1} = 1 + O(|\mathcal{Q}|/w) = 1 + O(y^3/w)$$

by (6.5). Now $\mathbb{P}_{\mathcal{Q}}(E_0) = 1 - O(y^4/w)$ by (6.6), so we conclude that

$$\begin{aligned}\mathbb{P}_{w,z}\Big(\mathcal{S} \subset \bigcup_{p\in\mathcal{P}} R_p\Big) &= \sum_{m=0}^{|\mathcal{Q}|} \mathbb{P}_{\mathcal{Q}}(E_m)\cdot\mathbb{E}_{\mathcal{Q}}\Big(\mathbb{P}_{\mathcal{P}\setminus\mathcal{Q}}\Big(\mathcal{S}' \subset \bigcup_{p\in\mathcal{P}\setminus\mathcal{Q}} R_p\Big)\Big|E_m\Big)\\ &= (1 + O(y^4/w))\big(1 - \Theta_{w,z}\big)^{\ell} + O\Big(\sum_{m\geqslant 1}(ey^4/w)^m\big(1 - \Theta_{w,z}\big)^{\ell-2m}\Big)\\ &= (1 + O(y^4/w))\big(1 - \Theta_{w,z}\big)^{\ell}.\end{aligned}$$

This completes the proof. □

7. The behavior of the largest gap

In this section we use the estimates from the previous section to complete the proof of Theorem 1.1. In Theorems 7.1 and 7.2 below, we suppose that

$$\varepsilon = \varepsilon(x) := \frac{1}{(\log_3 x)^{1/3}}. \tag{7.1}$$

We also note that

$$u < W_{g(u)+1}\log(g(u)+1) \leqslant (W_{g(u)}+1)\log(g(u)+1).$$

and hence

$$W_{g(u)}\log g(u) = u + O(\log u). \tag{7.2}$$

Theorem 7.1 (Probabilistic upper bound for gap). *For large x,*

$$\mathbb{P}\big[G_{\mathcal{R}}(x) \leqslant g\big((1+\varepsilon)\xi(\log\tfrac{x}{2})^2\big)\big] \geqslant 1 - x^{-\varepsilon/2}.$$

THEOREM 7.2 (Probabilistic lower bound for gap). *If x is large then*

$$\mathbb{P}\big[G_{\mathcal{R}}(x) \geqslant g\big((1-\varepsilon)\xi(\log 2x)^2\big)\big] \geqslant 1 - O\big((\log x)^{-8}\big).$$

Proof of Theorem 7.1. Let $y := g((1+\varepsilon)\xi(\log \frac{x}{2})^2)$, so that by (7.2) we have

$$W_y \log y = (1+\varepsilon)\xi(\log x)^2 + O(\log x). \tag{7.3}$$

We also have by (1.12) the bounds

$$\log^2 x \ll y \ll (\log^2 x)\log_2 x.$$

Let $z := z(x)$. The probability that $\mathcal{R} \cap [0,x]$ has a gap of size $\geqslant y$ does not exceed the probability that $\mathcal{S}_z \cap [0,x]$ has a gap of size $\geqslant y$, which in turn is at most

$$\mathbb{E}\big|\{n \leqslant x : [n,n+y] \cap \mathcal{S}_z = \varnothing\}\big| \leqslant x \cdot \mathbb{P}(\mathcal{S}_z = 0).$$

Let $w_1 := (y/\log y)^{1/2}$ and $w_5 := y^8$ as before. Also put $\eta := \frac{\log_4 x}{\log_3 x}$. Applying Corollary 6.2 together with (7.3), it follows that with probability $1 - O(x^{-100})$ we have

$$\begin{aligned} S_{w_5} &= (1+O(\eta))\frac{S_{w_1}}{16} \geqslant (1+O(\eta))\frac{W_y}{16} \\ &\geqslant \frac{(1+\varepsilon+O(\eta))\,\xi(\log x)^2}{32\log_2 x} \\ &\geqslant \frac{(1+2\varepsilon/3)\,\xi(\log x)^2}{32\log_2 x} \end{aligned}$$

using (7.1) in the final step. Fix $\mathcal{A}_{w_5}$ so that S_{w_5} satisfies this inequality. Taking into account that

$$\Theta_{w_5,z} = \frac{32\log_2 x}{\xi \log x}\left(1 + O\left(\frac{1}{\log_2 x}\right)\right),$$

Lemma 6.3 now shows that

$$\mathbb{P}_{w_5,z}(S_z = 0) \ll (1-\Theta_{w_5,z})^{S_{w_5}} \ll x^{-1-\varepsilon/2},$$

as required. □

Proof of Theorem 7.2. Set $y := g((1-\varepsilon)\xi(\log 2x)^2)$, so that

$$W_y \log y = (1-\varepsilon)\xi \log^2 x + O(\log x). \tag{7.4}$$

Again, (1.12) implies that

$$\log^2 x \ll y \ll (\log^2 x)\frac{\log_2 x}{\log_3 x}.$$

Let $z := z(x/2)$, $w_1 := (y/\log y)^{1/2}$, $w_5 := y^8$ and $\eta := \frac{\log_4 x}{\log_3 x}$. In particular, $z \sim (x/2)^{1/e^\gamma}$ by (1.7), and

$$w_1 \ll \frac{\log x}{(\log_3 x)^{1/2}}. \tag{7.5}$$

It suffices to show that with high probability, $\mathcal{S}_z \cap (x/2, x]$ has a gap of size $\geqslant y$, for this implies that $\mathcal{R}$ has a gap of size $\geqslant y$ within $[0, x]$. For the sake of brevity we write

$$\mathcal{F}(u, v) := [u, u+y] \setminus \bigcup_{p \leqslant v} R_p, \qquad F(u, v) := |\mathcal{F}(u, v)|.$$

That is, $F(u, v)$ counts the number of elements in $[u, u+y]$ sieved by the primes $\leqslant v$. In particular, $S_w = F(0, w)$. There is some vector $(b_p)_{p \in w_1}$ so that there are exactly W_y integers in $[0, y]$ that avoid the residue classes $(b_p \bmod p)_{p \leqslant w_1}$. Setting

$$Q := \prod_{p \leqslant w_1} p,$$

for any $\mathcal{A}_{w_1}$, there is a progression $b \bmod Q$ such that

$$F(u, w_1) = W_y \qquad \text{whenever} \quad u \equiv b \bmod Q.$$

Specifically, choose b such that $b \equiv a_p - b_p \bmod p$ for all primes $p \leqslant w_1$. Let $\mathcal{U}$ be the set of integers $u \equiv b \bmod Q$ such that $[u, u+y] \subset (x/2, x]$. We show that with high probability, $F(u, z) = 0$ for at least one $u \in \mathcal{U}$.

By Corollary 6.2, with probability at least $1 - O(x^{-100})$, we have for any given $u \in \mathcal{U}$ the bound

$$F(u, w_5) = (\tfrac{1}{16} + O(\eta))F(u, w_1) = (\tfrac{1}{16} + O(\eta))W_y. \tag{7.6}$$

Let E be the event that this bound holds for *every* $u \in \mathcal{U}$. By the union bound, $\mathbb{P}_{w_1, w_5}(E) \geqslant 1 - O(x^{-99})$. Conditioning on E, we denote

$$\mathcal{U}_r := \{u \in \mathcal{U} : F(u, w_5) = r\} \qquad (r \geqslant 0).$$

The sets $\mathcal{U}_r$ depend only on $\mathcal{A}_{w_5}$, and $\mathcal{U}_r = \varnothing$ unless $r = (\frac{1}{16} + O(\eta))W_y$ by (7.6). Rather than work with all r, we focus on a popular value of r; thus, let ℓ be fixed with the property that $|\mathcal{U}_\ell| \geqslant |\mathcal{U}_r|$ for all r. By (7.5), we have

$$|\mathcal{U}_\ell| \gg \frac{|\mathcal{U}|}{\eta W_y} \gg \frac{x}{QW_y} = xe^{-O(w_1)} \gg x^{1 - O((\log_3 x)^{-1/2})}. \tag{7.7}$$

Combining (7.4) with (7.6) and (7.1), we have

$$\ell \leqslant (\tfrac{1}{16} + O(\eta))W_y \leqslant \frac{(1 - (2/3)\varepsilon)\xi(\log x)^2}{32 \log_2 x}. \tag{7.8}$$

Next, let

$$M := \big|\{u \in \mathcal{U}_\ell : F(u, z) = 0\}\big|,$$

which counts those intervals indexed by $u \in \mathcal{U}_\ell$ for which $\mathcal{F}(u, w_5)$ is covered by $\bigcup_{w_5 < p \leqslant z} R_p$. We analyze M using first and second moments. Firstly, by Lemma 6.3,

$$\mathbb{E}_{w_5, z} M = \sum_{u \in \mathcal{U}_\ell} \mathbb{P}_{w_5, z}(F(u, z) = 0) = |\mathcal{U}_\ell|(1 - \Theta)^\ell (1 + O(y^2/w_5)),$$

where

$$\Theta := \Theta_{w_5, z} = \frac{32 \log_2 x}{\xi \log x} \left(1 + O\left(\frac{1}{\log_2 x}\right)\right). \tag{7.9}$$

To bound the second moment of M, apply Lemma 6.5 with $\mathcal{S} := \mathcal{F}(u, w_5) \cup \mathcal{F}(u', w_5)$, where u and u' are distinct elements of $\mathcal{U}_\ell$. The hypotheses of Lemma 6.5

are satisfied as any prime $p > w_5 > y$ can divide at most two elements of $\mathcal{S}$. We obtain

$$\begin{aligned}\mathbb{E}_{w_5,z}M^2 &= \mathbb{E}_{w_5,z}M + \sum_{\substack{u,u'\in\mathcal{U}_\ell\\ u\neq u'}} \mathbb{P}_{w_5,z}\left(F(u,z)=F(u',z)=0\right)\\ &= |\mathcal{U}_\ell|^2(1-\Theta)^{2\ell}(1+O(y^4/w_5)) + O\big(|\mathcal{U}_\ell|(1-\Theta)^\ell\big).\end{aligned}$$

By (7.7), (7.8) and (7.9) we have

$$|\mathcal{U}_\ell|(1-\Theta)^\ell \geqslant x^{2\varepsilon/3-O((\log_3 x)^{-1/2})} \geqslant x^{\varepsilon/2}$$

for large x, and hence we bound the variance by

$$\sigma^2 := \mathbb{V}_{w_5,z}M = \mathbb{E}_{w_5,z}M^2 - (\mathbb{E}_{w_5,z}M)^2 \ll |\mathcal{U}_\ell|^2(1-\Theta)^{2\ell}y^4/w_5.$$

Thus, Chebyshev's inequality implies

$$\mathbb{P}_{w_5,z}\left(M \geqslant \tfrac{1}{2}|\mathcal{U}_\ell|(1-\Theta)^\ell\right) \geqslant 1 - O(y^4/w_5) = 1 - O(1/y^4).$$

In particular, with probability at least $1-O(y^{-4}) = 1-O((\log x)^{-8})$ there is an interval $[u, u+y]$ in $(x/2, x]$ completely sieved out by $\mathcal{A}_z$. □

Proof of Theorem 1.1. Let $x_j := 2^j$ vary over positive integers j, and let $\varepsilon > 0$ be fixed. Theorem 7.1 implies that for large j we have

$$\mathbb{P}\big[G_{\mathcal{R}}(x_j) \leqslant g((1+\varepsilon)\xi\log^2 x_{j-1})\big] \geqslant 1 - x_j^{-\varepsilon/2} \qquad (j \text{ large}).$$

The convergence of $\sum_j x_j^{-\varepsilon/2}$ implies, via the Borel-Cantelli lemma, that almost surely there is a J so that

$$G_{\mathcal{R}}(x_j) \leqslant g((1+\varepsilon)\xi\log^2 x_{j-1}) \qquad (j \geqslant J).$$

As $G_{\mathcal{R}}$ and g are both increasing functions, the above relation implies that for all $x_{j-1} < x \leqslant x_j$ and $j > J$ we have

$$G_{\mathcal{R}}(x) \leqslant G_{\mathcal{R}}(x_j) \leqslant g((1+\varepsilon)\xi\log^2 x_{j-1}) \leqslant g((1+\varepsilon)\xi\log^2 x),$$

In a similar manner, Theorem 7.2 and Borel-Cantelli imply that almost surely there is a J so that

$$G_{\mathcal{R}}(x_j) \geqslant g((1-\varepsilon)\xi\log^2 x_{j+1}) \qquad (j \geqslant J).$$

As before, this implies that

$$G_{\mathcal{R}}(x) \geqslant g\big((1-\varepsilon)\xi\log^2 x\big) \qquad (x \geqslant x_J).$$

□

8. Large gaps from Hardy-Littlewood

To prove Theorems 1.5 and 1.6, we start with a simple inclusion-exclusion result (a special case of the Bonferroni inequalities or the "Brun pure sieve").

Lemma 8.1 (Brun's sieve). *Suppose that $y \geqslant 1$, let $\mathcal{N}, \mathcal{A}$ be sets of positive integers, and put*

$$T := \sum_{n\in\mathcal{N}} \prod_{h\in[0,y]} (1 - \mathbf{1}_{\mathcal{A}}(n+h))$$

and

$$U_K := \sum_{k=0}^{K} (-1)^k \sum_{\substack{\mathcal{H}\subset[0,y]\\ |\mathcal{H}|=k}} \sum_{n\in\mathcal{N}} \prod_{h\in\mathcal{H}} \mathbf{1}_{\mathcal{A}}(n+h) \qquad (K \geqslant 0).$$

Then, for any even K we have $T \leqslant U_K$, and for any odd K we have $T \geqslant U_K$.

Proof. For any integers $K, m \geqslant 0$ let

$$\delta_K(m) := \sum_{k=0}^{K} (-1)^k \binom{m}{k} \qquad \text{and} \qquad \delta(m) := \begin{cases} 1 & \text{if } m = 0, \\ 0 & \text{if } m \geqslant 1. \end{cases}$$

Observe that

$$\delta(m) \leqslant \delta_K(m) \quad (K \text{ even}) \qquad \text{and} \qquad \delta(m) \geqslant \delta_K(m) \quad (K \text{ odd});$$

hence, taking $A(n) := \big|\{0 \leqslant h \leqslant y : n+h \in \mathcal{A}\}\big|$ we have

$$T = \sum_{n\in\mathcal{N}} \delta(A(n)) = \sum_{n\in\mathcal{N}} \delta_K(A(n)) + \theta,$$

where $\theta \geqslant 0$ if K is even and $\theta \leqslant 0$ if K is odd. Also,

$$\sum_{n\in\mathcal{N}} \delta_K(A(n)) = \sum_{k=0}^{K} (-1)^k \sum_{n\in\mathcal{N}} \binom{A(n)}{k} = U_K$$

since

$$\binom{A(n)}{k} = \sum_{\substack{\mathcal{H}\subset[0,y]\\ |\mathcal{H}|=k}} \prod_{h\in\mathcal{H}} \mathbf{1}_{\mathcal{A}}(n+h) \qquad (n \in \mathcal{N}),$$

and the lemma is proved. □

Proof of Theorem 1.5. Although Theorem 1.5 concerns the behavior of a specific set $\mathcal{A}$, our first task is to express the gap-counting function for $\mathcal{A}$ in terms of the random quantities with which we have been working in the past few sections.

First, observe that (1.17) with $\mathcal{H} = \{0\}$ implies that

$$\big|\{n \leqslant x : n \in \mathcal{A}\}\big| \sim x/\log x,$$

and it follows trivially that $G_{\mathcal{A}}(x) \gg \log x$. Therefore, by adjusting the implied constant in the conclusion of the theorem, we may assume that

$$\kappa \geqslant D\frac{\log_2 x}{\log x} \tag{8.1}$$

for a sufficiently large constant D.

Let x be a large real number, put $\mathcal{N} := [x/2, x]$ and let y, K be integer parameters to be chosen later, with K odd and with $K \leqslant \frac{\kappa \log x}{2\log_2 x}$. Define T and U_K as in Lemma 8.1. Since $T \geqslant U_K$ by Lemma 8.1, our aim is to show that $U_K \geqslant 1$. Using (1.17) we see that

$$U_K = \sum_{k=0}^{K} (-1)^k \int_{x/2}^{x} \frac{1}{(\log t)^k} \sum_{\substack{\mathcal{H}\subset[0,y]\\ |\mathcal{H}|=k}} \mathfrak{S}(\mathcal{H})\, dt + O(E),$$

where

$$E := Kx^{1-\kappa}\binom{y+1}{K}.$$

By Lemma 3.5, replacing $\mathfrak{S}(\mathcal{H})/\log^k t$ with $V_{\mathcal{H}}(z(t))$ induces an additive error of size $O(E)$ since $\kappa \leqslant 1/2$. Also, (1.8) implies that

$$\sum_{\substack{\mathcal{H}\subset[0,y]\\ |\mathcal{H}|=k}} V_{\mathcal{H}}(z(t)) = \mathbb{E}_{z(t)}\binom{S_{z(t)}}{k},$$

and we get

$$U_K = \int_{x/2}^{x} \mathbb{E}_{z(t)} \sum_{k=0}^{K} (-1)^k \binom{S_{z(t)}}{k}\, dt + O(E).$$

Since K is odd, the sum on k is a lower bound for $\mathbb{P}(S_{z(t)} = 0)$; adding the term $k = K+1$ switches the inequality (cf. the proof of Lemma 8.1) and thus

$$U_K \geqslant \int_{x/2}^{x} \mathbb{P}(S_{z(t)} = 0) - \mathbb{E}_{z(t)}\binom{S_{z(t)}}{K+1}\, dt + O(E). \tag{8.2}$$

Let

$$w := y^4, \qquad z := z(x/2).$$

The upper bound sieve (Lemma 3.1) implies the crude bound $S_w \leqslant Cy/\log y$ for some absolute constant C. We now put

$$y := \frac{\kappa\,\xi \log^2 x}{400C \log_2 x} \qquad \text{and} \qquad K := 2\left\lfloor \frac{100Cy}{\log x} \right\rfloor - 1. \tag{8.3}$$

With these choices, $K \leqslant \frac{\kappa \log x}{2\log_2 x}$ and, using (8.1), we have

$$y \geqslant \frac{D}{400C} \log x. \tag{8.4}$$

It also follows that

$$E \ll x^{1-\kappa}(\log x)^K \ll x^{1-\kappa+\kappa\,\xi/2} \ll x^{1-\kappa/3}.$$

for all large x. Corollary 6.4 and the crude bound $\Theta_{w,z} \leqslant 8\frac{\log y}{\log x}$ imply that

$$\begin{aligned}
\mathbb{E}_{z(t)}\binom{S_{z(t)}}{K+1} &\leqslant \mathbb{E}_z\binom{S_z}{K+1} \\
&\ll \Theta_{w,z}^{K+1}\mathbb{E}_w\binom{S_w}{K+1} \\
&\ll \left(\Theta_{w,z}\frac{eCy}{K\log y}\right)^{K+1} \\
&\ll e^{-K} \ll e^{-200Cy/\log x},
\end{aligned}$$

where we used (8.3) in the last step. It remains to show that $\mathbb{P}_{z(t)}(S_{z(t)} = 0)$ is substantially larger. Lemma 6.3 implies immediately that

$$\begin{aligned}
\mathbb{P}_z(S_{z(t)} = 0) \geqslant \mathbb{P}_z(S_z = 0) &\gg (1-\Theta_{w,z})^{S_w} \\
&\gg e^{-\Theta_{w,z}(Cy/\log y)} \geqslant e^{-8Cy/\log x},
\end{aligned}$$

as required. Combining these estimates with (8.2) gives

$$U_K \gg xe^{-8Cy/\log x} + O(xe^{-200Cy/\log x} + x^{1-c/3}) \gg xe^{-8Cy/\log x},$$

the last inequality following from (8.4), the fact that D is sufficiently large, and that $y/\log x \ll \kappa/\log_2 x$. This completes the proof of Theorem 1.5. □

Proof of Theorem 1.6. Let x be large, let $\varepsilon > 0$, and let $y := g((1-\varepsilon)c\,\xi \log^2 x)$. By (7.2),

$$W_y \log y = (1-\varepsilon)c\,\xi \log^2 x + O_c(\log_2 x). \tag{8.5}$$

In particular, (5.2) holds, with α, β depending on c. Also, from (1.12) we have

$$(c/2)\log^2 x \leqslant y = o((\log^2 x)\log_2 x). \tag{8.6}$$

Let

$$w_1 := (y/\log y)^{1/2}, \qquad w_5 := y^8, \qquad z := z(x/2).$$

Again, let $\mathcal{N} := (x/2, x]$, and define C as in the previous proof. We apply Lemma 8.1 with

$$K := 2\left\lfloor \frac{100Cy}{\log x} \right\rfloor - 1,$$

so that $K \leqslant \frac{200Cy}{\log x}$. Similarly to (8.2) we get that

$$U_K \geqslant \int_{x/2}^{x} \mathbb{P}(S_{z(t)} = 0) - \mathbb{E}_{z(t)}\binom{S_{z(t)}}{K+1}\,dt + O(E), \tag{8.7}$$

where, because the function $\mathfrak{S}_{z(t)}(\mathcal{H})$ appears already in (1.18), as does the averaging over $\mathcal{H}$, we have

$$E \ll Kx^{1-c} \ll x^{1-c}\log^2 x. \tag{8.8}$$

By the same reasoning as in the proof of Theorem 1.5, we get that

$$\mathbb{E}_{z(t)}\binom{S_{z(t)}}{K+1} \ll e^{-K} \ll x^{-10c}, \tag{8.9}$$

where we used (8.6) in the last step.

Let $w := y^8$ and fix $\mathcal{A}_w$. By Lemma 6.3 we have

$$\mathbb{P}_{w,z}(S_z = 0) = (1-\Theta_{w,z})^{S_w}(1 + O(y^{-6})). \tag{8.10}$$

Now put $w_1 := (y/\log y)^{1/2}$, and let $\mathcal{A}_{w_1}$ be fixed such that $S_{w_1} = W_y$. This occurs with probability $\geqslant x^{-o(1)}$, since $(y/\log y)^{1/2} = o(\log x)$ by (8.6). Conditional on $\mathcal{A}_{w_1}$, Corollary 6.2 implies that with probability at least $1 - O(x^{-100})$ we have

$$S_w = (\tfrac{1}{16} + O(\eta))S_{w_1} = (\tfrac{1}{16} + O(\eta))W_y,$$

where $\eta := \frac{\log_4 x}{\log_3 x}$ as before and the implied constants may depend on c. Now fix w such that the above holds. Since

$$\Theta_{w,z} = (1 + O(\eta))\frac{16\log y}{\xi \log x},$$

(8.5) implies that

$$\Theta_{w,z} S_w = (1 + O(\eta))(1-\varepsilon)c\log x,$$

where we have used (8.5) in the last step. Inserting this last estimate into (8.10), we conclude that

$$\mathbb{P}_z(S_z = 0) \gg e^{-(1+O(\eta))(1-\varepsilon)c\log x} \gg x^{-(1-\varepsilon/2)c} \tag{8.11}$$

In particular, the right side of (8.11) has larger order than the right sides in (8.8) and (8.9). Thus, inserting (8.8), (8.9) and (8.11) into (8.7), we conclude that $U_K \geqslant 1$ if x is sufficiently large depending on ε. By a simple diagonalization argument, the same claim then holds for some $\varepsilon = \varepsilon(x) = o(1)$ going to zero sufficiently slowly as $x \to \infty$. This completes the proof of Theorem 1.6. □

9. The influence of exceptional zeros

In this section, we show that the existence of exceptional zeros implies that W_y is rather smaller than the upper bound in (1.12) infinitely often.

Theorem 9.1. *Let $q \in \mathbb{N}$, and suppose that there is a real Dirichlet character χ_q mod q such that $L(1-\delta_q, \chi_q) = 0$ and $0 < \delta_q \leqslant \frac{c}{\log q}$, where $c := 1/11^2$. For*

$$y := \exp\left\{\left(\frac{\log q}{\delta_q}\right)^{1/2}\right\} \tag{9.1}$$

we have

$$W_y \ll \delta_q y = \frac{y\log q}{\log^2 y}.$$

Proof. Denote by $\pi(x; q, a)$ the number of primes $p \leqslant x$ lying in the progression a mod q. By hypothesis, $qy \geqslant q^{1+1/\sqrt{c}} = q^{12}$, therefore we may apply [42, Corollary 1.4], obtaining

$$\begin{aligned}\pi(qy+1; q, 1) &\leqslant \sqrt{y/q} + \frac{2}{\log(qy)} \sum_{\substack{\sqrt{qy} < p \leqslant qy \\ p \equiv 1 \pmod q}} \log p \\ &\ll \sqrt{y/q} + \frac{\lambda qy}{\phi(q)\log(qy)},\end{aligned}$$

where

$$\lambda := 1 - (qy)^{-\delta_q}/(1-\delta_q) \ll \delta_q \log(qy).$$

By Siegel's Theorem [7, §21], for any $\varepsilon > 0$, $\delta_q \gg_\varepsilon q^{-\varepsilon}$. We conclude that

$$\pi(qy+1; q, 1) \ll \frac{\delta_q qy}{\phi(q)}.$$

This may also be deduced from Gallagher's prime number theorem [13, Theorem 7].

Define the residue classes a_p by $qa_p + 1 \equiv 0 \bmod p$ when $p \nmid q$. Let $\mathcal{T}$ denote the set of $n \leqslant y$ with $n \not\equiv a_p \bmod p$ for all $p \nmid q$ such that $p \leqslant \sqrt{y/\log y}$. Then for any $n \in \mathcal{T}$, $qn+1$ is either prime or the product of two primes $> \sqrt{y/\log y}$. Then we make a greedy choice of a_p for $p \mid q$, choosing successively a_p so that

$a_p \bmod p$ covers a proportion at least $1/p$ of the remaining elements of $\mathcal{T}$. This shows that

$$\begin{aligned} W_y &\leqslant \frac{\phi(q)}{q}|\mathcal{T}| \\ &\leqslant \frac{\phi(q)}{q}\left[\pi(qy+1;q,1) + \sum_{\sqrt{y/\log y}<p\leqslant\sqrt{qy+1}} \pi\left(\frac{qy+1}{p};q,\overline{p}\right)\right], \end{aligned}$$

where $\overline{p}$ is the inverse of p modulo q. Siegel's theorem implies that $\log y \leqslant q^{o(1)}$. Applying the Brun-Titchmarsh theorem to the sum over p, we see that

$$W_y \ll \frac{\phi(q)}{q}\left[\frac{qy\delta_q}{\phi(q)} + \frac{qy\log(q\log y)}{\phi(q)\log^2 y}\right] \ll y\left[\delta_q + \frac{\log q}{\log^2 y}\right] \ll \delta_q y.$$

This completes the proof. □

Proof of Theorem 2.2. Let $q \in Q$, and apply Thereom 9.1 with $y = y_q$ defined by (9.1). By assumption, $\frac{\log y_q}{\log q} \to \infty$ as $q \to \infty$, and hence that

$$\delta_q = \frac{\log q}{\log^2 y_q} = o\left(\frac{1}{\log y_q}\right).$$

This shows that $W_{y_q} = o(y_q/\log y_q)$, and the remaining parts of Theorem 2.2 follow immediately. □

References

[1] K. Azuma, *Weighted sums of certain dependent random variables*, Tôhoku Math. J., **19** (1967), 357–367.

[2] 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.

[3] G. Bennett, *Probability inequalities for the sum of independent random variables*, J. Amer. Stat. Assoc. **57** (1962), 33–45.

[4] J. H. Cadwell, *Large intervals between consecutive primes*, Math. Comp. **25** (1971), 909–913,

[5] H. Cramér, *Some theorems concerning prime numbers*, Arkiv för Mat. o. Fys. **15** (1920), no. 5, 1–32.

[6] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arith. **2** (1936), no. 1, 23–46.

[7] H. Davenport, *Multiplicative number theory* (3rd edition). Graduate Texts in Mathematics vol. 74, Springer-Verlag, New York, 2000.

[8] C. Elsholtz, *Upper bounds for prime k-tuples of size* $\log N$ *and oscillations*, Arch. Math. (Basel), **82** (2004), 33–39.

[9] P. Erdős and I. Richards, Density functions for prime and relatively prime numbers. *Monatsh. Math.* 83 (1977), 99–112.

[10] K. Ford, *Large prime gaps and progressions with few primes*, Rivista di Matematica della Universita di Parma **12**, no. 1, (2021), 41–47. Proceedings of Second Symposium on Analytic Number Theory Cetraro, Italy, July 8–12, 2019.

[11] K. Ford, B. Green, S. Konyagin, J. Maynard and T. Tao, *Long gaps between primes*, J. Amer. Math. Soc., **31** (2018), no. 1, 65–105.

[12] J. Friedlander and H. Iwaniec, *Opera de Cribro*. American Math. Soc., 2010.

[13] P. X. Gallagher, *A large sieve density estimate near* $\sigma = 1$, Invent. Math. **11** (1970), 329–339.

[14] P. X. Gallagher, *On the distribution of primes in short intervals*, Mathematika **23** (1) (1976), 4–9.

[15] D. A. Goldston, J. Pintz, and C. Y. Yıldırım, *Primes in tuples I*, Ann. Math. **170** (2009), 819–862.
[16] A. Granville, *Harald Cramér and the distribution of prime numbers*, Harald Cramér Symposium (Stockholm, 1993). *Scand. Actuar. J.* (1995), no. 1, 12–28.
[17] A. Granville, *Sieving intervals and Siegel zeros.* Acta Arith. **205** (2022), no. 1, 1–19.
[18] A. Granville and A. Lumley, *Primes in short intervals: Heuristics and calculations*, preprint. `arXiv:2009.05000`.
[19] G. H. Hardy and J. E. Littlewood, *Some problems of Partitio Numerorum (III): On the expression of a number as a sum of primes*, Acta Math. **44** (1922), no. 1, 1–70.
[20] D. Hensley and I. Richards, Primes in intervals. *Acta Arith.* 25 (1973/74), 375–391.
[21] H. Iwaniec, *On the error term in the linear sieve*, Acta Arith. **19** (1971), 1–30.
[22] H. Iwaniec, *Conversations on the exceptional character*, in the book *Analytic number theory*, lectures given at the C.I.M.E. summer school in Cetraro, Italy, (A. Perelli, C. Viola, eds.), Lecture Notes in Mathematics vol. 1891, Springer-Verlag 2002, 97–132.
[23] D. Koukoulopoulos, *The distribution of prime numbers.* To be published.
[24] P. Leikauf and J. Waldvogel, *Finding clusters of primes, I.*, preprint. `https://www.sam.math.ethz.ch/~waldvoge/Projects/clprimes03.pdf`
[25] H. Maier, *Primes in short intervals*, Michigan Math. J. **32** (1985), no. 2, 221–225.
[26] D. Mastrostefano, *Positive proportion of short intervals containing a prescribed number of primes*, Bull. Aust. Math. Soc. **100** (2019), no. 3, 378–387.
[27] H. L. Montgomery, *A note on the large sieve*, J. London Math. Soc. **43** (1968), 93–98.
[28] H. L. Montgomery and K. Soundararajan, Primes in short intervals. Comm. Math. Phys. **252** (2004), no. 1-3, 589–617.
[29] H. L. Montgomery and R. C. Vaughan, *The large sieve*, Mathematika **20** (1973), 119–135.
[30] H. L. Montgomery and R. C. Vaughan, *Multiplicative Number Theory I. Classical Theory*, Cambridge Studies in Advanced Mathematics, 97. Cambridge University Press, Cambridge, 2007.
[31] H. L. Montgomery and S. Wagon, *A heuristic for the prime number theorem*, Math. Intelligencer **28** (2006), no. 3, 6–9.
[32] T. R. Nicely, *Enumeration to* 10^{14} *of the twin primes and Brun's constant.* Virginia J. Sci. **46** (1995), no. 3, 195–204.
[33] T. R. Nicely, *New evidence for the infinitude of some prime constellations*, preprint. `http://www.trnicely.net/ipc/ipc1d.html`
[34] T. R. Nicely, *Prime constellations research project*, web pages: `http://www.trnicely.net/counts.html`
[35] T. R. Nicely, *First occurrence prime gaps*, web page: `http://www.trnicely.net/gaps/gaplist.html`
[36] J. Pintz, *Cramér vs. Cramér. On Cramér's probabilistic model for primes*, Funct. Approx. Comment. Math. **37** (2007), part 2, 361–376.
[37] G. Pólya, *Heuristic reasoning in the theory of numbers*, Amer. Math. Monthly **66** (1959), 375–384.
[38] A. Selberg, *On the normal density of primes in small intervals, and the difference between consecutive primes*, Arch. Math. Naturvid. **47** (1943), no. 6, 87–105.
[39] A. Selberg, *Remarks on sieves*, Proc. 1972 Number Theory Conference, University of Colorado at Boulder (August 14–18), 205–216.
[40] D. Shanks, *On maximal gaps between successive primes.* Math. Comp. **18** (1964), 646–651.
[41] K. Soundararajan, *The distribution of prime numbers*, in the book *Equidistribution in number theory, an introduction*, 59–83. NATO Sci. Ser. II Math. Phys. Chem. **237**, Springer, Dordrecht, 2007.
[42] J. Thorner and A. Zaman, *Refinements to the prime number theorem for arithmetic progressions*, preprint. `arXiv:2108.10878`.

Department of Mathematics, University of Missouri, Columbia MO 65211, USA.
Email address: bankswd@missouri.edu

Department of Mathematics, University of Illinois, 1409 West Green St, Urbana, IL 61801, USA.
Email address: ford126@illinois.edu

Department of Mathematics, UCLA, 405 Hilgard Ave, Los Angeles CA 90095, USA.
Email address: tao@math.ucla.edu