Large gaps between primes

Petru Constantinescu

Contents

1 Introduction

Let p_n denote the n-th prime, and we define

$$G(X) := \max_{p_{n+1} \le X} (p_{n+1} - p_n)$$

Our goal is to find a lower bound for $G(X)$ as good as possible. First, we note that from the prime number theorem we obtain

$$G(X) \ge (1 + o(1)) \log X$$

since the average gap between primes less than X is $\sim \log X$. The first significant improvement was achieved by Westzynthius [16] in 1931, who showed that the largest gap between consecutve can be an arbitrarily large constant of the average gap, i.e.

$$\lim_{X\to\infty} \frac{G(X)}{\log X} = \infty \ .$$

In 1936, Erdős and Rankin [15] showed that

$$G(X) \ge (c + o(1)) \frac{\log X \log_2 X \log_4 X}{(\log_3 X)^2}$$

with the constant $c = 1/3$. As it is standard in the subject, we denote $\log_2 x = \log\log x$, $\log_3 x = \log\log\log x$ and so on. During the following years, there were many improvements for the constant c. A significant breakthrough occurred in 2014, when Ford, Green, Konyagin and Tao [5] and independently Maynard [14] proved that c could be taken arbitrarily large. This asked a long standing conjecture of Erdős. After that, the 5 authors joined forces and obtained the following improvement [6]:

Theorem 1.1 *For X large enough,*

$$G(X) \gg \frac{\log X \log_2 X \log_4 X}{\log_3 X} \tag{1.1}$$

It is believed that the result above is the best obtainable with current methods.

We have the following conjectures regarding $G(X)$. Using a basic probabilistic model of primes, Cramer conjectured that

$$\limsup_{X\to\infty} \frac{G(X)}{(\log X)^2} = 1 \ .$$

Using a refinement of Cramer's model, Granville [8] conjectured that in fact

$$\limsup_{X\to\infty} \frac{G(X)}{(\log X)^2} \ge 2e^{-\gamma} \ .$$

The best unconditional upper bound for $G(X)$ is $G(X) \ll X^{0.525}$ by Baker, Harman and Pintz [1]. Assuming the Riemann Hypothesis, the best known upper bound is $G(X) \ll X^{1/2} \log X$. We remark that there is a lot of room for improvement in both the upper bounds and the lower bounds.

The goal of this project is to provide a clear exposition of the proof of Theorem 1.1. There are two key ingredients in the proof. Firstly, we need an efficient hypergraph covering theorem, on which we will concentrate in Section 4. Secondly, we need a uniform version of the multidimensional sieve approach, which was developed in [13]. This process will be described in Section 5. We will describe our plan of attack in Section 2, whilst everything will carefully be put together in Section 6.

2 Heuristics and outline of the proof

All works on lower bounds of $G(X)$ have the same starting point. We want to rephrase our problem such that it becomes a problem about sieving an interval with residue classes modulo small primes.

Definition 2.1 *Let x be a positive integer. We define $y(x)$ to be the largest integer y such that one may select one residue class $(a_p \mod p)$, for each $p \leq x$, which "sieve out" the interval $[y] = \{1, 2, \ldots, y\}$.*

Lemma 2.2 *Let $P(x) = \prod_{p \leq x} p$. Then*

$$G(P(x) + x + y(x)) \geq y(x) \ .$$

Proof Select residue classes $(a_p \mod p)$ for $p \leq x$, which cover $[y]$. Using the Chinese Remainder theorem, we find m with $x < m \leq x + P(x)$ such that $m \equiv -a_p \pmod p$, for all $p \leq x$. Then, for $1 \leq t \leq y$, there exists some p such that $t \equiv a_p \pmod p$, and hence $m + t \equiv -a_p + a_p \equiv 0 \pmod p$. Therefore $p | m + t$ and $m + t > x \geq p$, hence $m + t$ composite, for $1 \leq t \leq y$, which gives us y consecutive composite numbers. ∎

Remark Our function $y(x)$ is very similar to the Jacobsthal function $j(x)$. If x is a positive integer, then $j(x)$ is the maximal gap between numbers coprime with x. Similar as in the proof of Lemma 2.2, using the Chinese remainder theorem we see that

$$y(x) = j(P(x)) \ .$$

From the prime number theorem, we know that $P(x) = \exp((1 + o(1))x)$. Since $y(x) = e^{o(1)x}$, an immediate consequence of Lemma 2.2 is that

$$G(x) \geq y((1 + o(1)) \log x)$$

Hence, our goal is that, given an integer x, to find y as large as possible such that we cover $[y]$ with residue classes modulo primes less than x. Hence, in order to prove Theorem 1.1, we can take

$$y := cx \frac{\log x \log_3 x}{\log_2 x} \ . \tag{2.1}$$

where c is a small fixed positive constant.

Our goal is to cover $[y]$ by residue classes $(a_p \mod p)$, for $p \leq x$. We fix a number z we select later. We select our residue classes $(a_p \mod p)$ in 4 steps:

1. $a_p = 0$ for $p \in [2, \log^{10} x] \cup (z, x/4]$
2. Random uniform choice for a_p, $p \in (\log^{10} x, z]$
3. Strategic choice conditional on step 2 for $p \in (x/4, x/2]$
4. Use a_p for each $x/2 < p \leq x$ to cover each one remaining element.

Let's first discuss the final step, which is the simplest. Suppose that after the first three steps the number of survivors left in $[y]$ is less than the number of primes considered in step 4. Then one can finish off by using each prime in step 4 to remove one of the surviving elements by appropriate choice of $(a_p \mod p)$, for $x/2 < p \leq x$. By the prime number theorem, it will be enough if we can show that steps 1-3 leave at most $x/(3 \log x)$ numbers unsieved.

Now we look at the first step. The elements left uncovered are a subset of the z-smooth numbers (which are few for appropriate z) and primes greater than $x/4$, since $(x/4) \log^{10} x > y$. We expect to be left with $\approx y/\log y \approx y/\log x$ numbers uncovered. This is much better than the typical choice,

$$y \prod_{2 \leq p \leq \log^{10} x} \left(1 - \frac{1}{p}\right) \prod_{z < p \leq x/4} \left(1 - \frac{1}{p}\right) \approx y \frac{\log z}{\log_2 x \ \log x}$$

if z is large enough. We want to choose z as large as possible such that $\Psi(y, z) = o(x/\log x)$ (so that they can be covered in the final step). $\Psi(y, z)$ denotes the number of z-smooth numbers less than y. For this purpose, we use the following lemma to find an upper bound for z-smooth numbers:

Lemma 2.3 *Let $u := \frac{\log y}{\log z}$ and assume $3 \log u < \log z$. Then we can bound the number of z-smooth numbers in $[y]$ by*

$$\Psi(y, z) \ll e^{-u \log u + u} y \log z \ .$$

Proof Let $0 < \sigma < 1$ be a quantity to be optimised later. We see that $(y/n)^\sigma \geq 1$ for all $n \leq y$ and $(y/n)^\sigma > 0$, for all $n > y$. Let $P(n)$ denote the largest prime factor of n. Then

$$\Psi(y,z) = \sum_{\substack{1\leq n\leq y\\ P(n)\leq z}} 1 \leq \sum_{\substack{n\geq 1\\ P(n)\leq z}} \left(\frac{y}{n}\right)^\sigma = y^\sigma \sum_{\substack{n\geq 1\\ P(n)\leq z}} \frac{1}{n^\sigma} = y^\sigma \prod_{p\leq z}\left(1-\frac{1}{p^\sigma}\right)^{-1}$$

Denote

$$\zeta(\sigma,z) = \prod_{p\leq z}\left(1-\frac{1}{p^\sigma}\right)^{-1}$$

We choose

$$\sigma = 1 - \frac{u\log u}{\log y} = 1 - \frac{\log u}{\log z} > \frac{1}{2} \tag{2.2}$$

Hence

$$\log\zeta(\sigma,s) = \sum_{p\leq z}\frac{1}{p^\sigma} + O(1)$$

Now

$$\sum_{p\leq z}\frac{1}{p^\sigma} = \sum_{p\leq z}\left(\frac{1}{p^\sigma}-\frac{1}{p}\right) + \log\log z + O(1)$$

using Mertens' theorem. So we are left with evaluating the sum on the right hand side.

$$\sum_{p\leq z}\left(\frac{1}{p^\sigma}-\frac{1}{p}\right) = \sum_{p\leq z}\left(\frac{p^{1-\sigma}-1}{p}\right) = \sum_{p\leq z}\left(\frac{\exp\left(\frac{\log p\log u}{\log z}\right)-1}{p}\right)$$

using (2.2). Next, we use the convexity inequality

$$\exp(ct) - 1 \leq (\exp(c)-1)t$$

which holds for $c > 0$ and $0 \leq t \leq 1$. We apply with $c = \log u$ and $t = \log p/\log z$ to obtain that

$$\sum_{p\leq z}\left(\frac{1}{p^\sigma}-\frac{1}{p}\right) \leq \frac{u}{\log z}\sum_{p\leq z}\frac{\log p}{p} = \frac{u}{\log z}(\log z + O(1)) = u + O(1)$$

where we have used Mertens' theorem again. Putting everything together we obtain

$$\log\zeta(\sigma,z) \leq \log\log z + u + O(1)$$

Hence

$$\Psi(y,z) \ll e^{-u\log u + u} y\log z \quad \blacksquare$$

Remark In fact, it is known that in the range $\log^3 y \leq z \leq y$, we have that

$$\Psi(y,z) \sim e^{-u\log u + O(u\log\log(3u))} y$$

which is a theorem of de Bruijn [3]. However, this stronger asymptotic behaviour doesn't improve our bounds and the lemma above will suffice.

Since $cx \leq y \leq cx\log x$, a very efficient choice of z is

$$z := x^{\log_3 x/4\log_2 x} \tag{2.3}$$

so that $u \sim 4\frac{\log_2 x}{\log_3 x}$ and $u\log u \sim 4\log_2 x$. Applying Lemma 2.3 we obtain

$$\Psi(y,z) \ll e^{-u\log u+u} y\log z \ll \frac{y}{\log^{4+o(1)} x}\log x\frac{\log_3 x}{\log_2 x} = o\left(\frac{x}{\log x}\right)$$

as desired.

Hence essentially we are left with

$$\mathscr{Q} := \{p \text{ prime} : x/4 < p \le y\}$$

which we want to cover in steps 2 and 3 up to $O(x/\log x)$ survivors.

For step 2, we choose $(a_p \mod p)$ uniform at random, for each prime $\log^{10} x < p \le z$. Denote

$$\mathscr{S} := \{p \text{ prime} : \log^{10} x < p \le z\} \tag{2.4}$$

the primes used in step 2 and

$$\mathscr{P} = \{p \text{ prime } : \ x/4 < p \le x/2\} \tag{2.5}$$

the primes which we use in step 3.

We take residue classes $\vec{a} = (a_s \mod s)_{s\in\mathscr{S}}$ chosen with equal probability and we define the respective random variable $\vec{\mathbf{a}}$. Let

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

and let

$$\mathscr{Q}(\vec{a}) := \mathscr{Q} \cap S(\vec{a}) \tag{2.6}$$

the elements left unsieved after step 2. We expect that step 2 will sparsify $\mathscr{Q}$ by a factor of

$$\sigma := \prod_{\log^{10} x < p \le z} \left(1 - \frac{1}{p}\right) \sim \frac{10 \log_2 x}{\log z} = 40 \frac{\log_2^2 x}{\log x \log_3 x}$$

Indeed, we will see that, with high probability,

$$|\mathscr{Q}(\vec{a})| \sim \sigma \frac{y}{\log y} \sim 40c \frac{x \log_2 x}{\log x}$$

Note that if we had $\sigma \frac{y}{\log y} \ll \frac{x}{\log x}$, we would be finished by having each prime greater than $x/4$ to cover one surviving element (same as described in step 4). This is actually the argument of Erdős and Rankin which gives a value of y smaller by an order of $\log_2 x$. Hence we would like that the average number of survivors sieved by primes in step 3 is about $\log_2 x$.

We return to our problem. We want to choose $(r_p \mod p)$ for $p \in \mathscr{P}$ such that

- The sets $R_p := \{n \in \mathscr{Q}(\vec{a}) : n \equiv r_p \pmod{p}\}$ are large on average, for all $p \in \mathscr{P}$
- $\bigcup_{p\in\mathscr{P}} R_p$ covers most of $\mathscr{Q}(\vec{a})$ efficiently (little overlap)

Sets of the type R_p are hard to describe and to work with. Instead, we fix an admissible k-tuple $1 \le h_1 < \cdots < h_k \ll k^2$, where k will be determined later. For example, let $h_1, \ldots, h_k$ be the first k primes larger than k, i.e. $h_i = p_{\pi(k)+i}$. We want to work with the sets of the form

$$e_p(n, \vec{a}) := \{n + h_i p : 1 \le i \le k\} \cap \mathscr{Q}(\vec{a})$$

For each $\vec{a}$, we want to construct random variables $(\mathbf{n}_p(\vec{a}))_{p\in\mathscr{P}}$ and corresponding random sets

$$\mathbf{e}_p(\vec{a}) := \{\mathbf{n}_p + h_i p : 1 \le i \le k\} \cap \mathscr{Q}(\vec{a})$$

such that the sets $\mathbf{e}_p(\vec{a})$ are large on average and have little overlap. The random variables $\mathbf{n}_p(\vec{a})$ (and hence also $\mathbf{e}_p(\vec{a})$) will be defined depending on a fixed $\vec{a}$ in the range of $\vec{\mathbf{a}}$. So when we work with $\mathbf{n}_p$, we consider $\vec{a}$ fixed. In order to construct $(\mathbf{n}_p)(\vec{a})$, we will use sieve weights methods which appeared firstly in [12] and then further developed in [13], [2] and [14].

Heuristically, we will show that $|\mathbf{e}_p(\vec{a})| \gg \log_2 x$ on average and that they have very little overlap. Hence, we expect

$$\left| \bigcup_{p\in\mathscr{P}} \mathbf{e}_p(\vec{a}) \right| \gg |\mathscr{P}| \log_2 x \gg \frac{x \log_2 x}{\log x} \ .$$

Since we expect $|\mathscr{Q}(\vec{a})| \sim \sigma \frac{y}{\log x}$, we are done if

$$\sigma \frac{y}{\log x} \leq c \frac{x \log_2 x}{\log x}$$

This justifies our choice for y in (2.1).

Indeed, we will prove the following theorem, which tells us that the sets $\mathbf{e}_p(\vec{a})$ are large enough on average and "well-distributed":

Theorem 2.4 *For $k \leq (\log x)^{1/5}$ and $(h_1, \ldots, h_k)$ admissible set with $h_i \ll k^2$ for all i, we have that with probability $1 - o(1)$ in $\vec{\mathbf{a}}$, we can construct random variables $(\mathbf{n}_p(\vec{a}))_{p \in \mathscr{P}}$ such that*

- *(Sets are large enough) For all $p \in \mathscr{P}$, we have that*

$$\mathbb{E}(|\mathbf{e}_p(\vec{a})|) \gg \log k$$

- *(Sparsity) For all $p \in \mathscr{P}$ and $q \in \mathscr{Q}$, we have*

$$\mathbb{P}(q \in \mathbf{e}_p(\vec{a})) \ll x^{-3/5}$$

- *(Small overlap) For distinct $q_1, q_2 \in \mathscr{Q}$,*

$$\sum_{p \in \mathscr{P}} \mathbb{P}(q_1, q_2 \in \mathbf{e}_p(\vec{a})) \ll x^{-3/5}$$

- *(Elements covered more than once in expectation) There exists a universal constant $C > 1$ such that for almost all $q \in \mathscr{Q}(\vec{a})$,*

$$\sum_{p \in \mathscr{P}} \mathbb{P}(q \in \mathbf{e}_p(\vec{a})) \sim C$$

Theorem 2.4 informally says that for almost all $\vec{a}$, we can construct random variables $\mathbf{n}_p(\vec{a})$ such that the probability of $q \in \mathscr{Q}$ to belong to $\mathbf{e}_p(\vec{a})$ is small; intersections between $\mathbf{e}_p(\vec{a})$ are very small; but that for almost all $q \in \mathscr{Q}$, the sum of probabilities is around the same.

We can think of $\mathscr{Q}(\vec{a})$ as the vertices of an hypergraph and $\mathbf{e}_p(\vec{a})$ as random edges in the hypergraph, i.e. random subsets of $\mathscr{Q}(\vec{a})$. Heuristically, we know that all edges are small ($|\mathbf{e}_p(\vec{a})| \leq k$), the degree of each vertex is small on average, there is very little overlap and that each vertex is covered in average at least once. We would like to deduce that we find an efficient covering of the vertices. This will be the subject of Section 4.

3 Probability notational conventions and useful lemmas

We will use boldface symbol (such as $\mathbf{X}$ or $\mathbf{a}$) to denote random variables. All the random variables we consider will be discrete. The range or the support of $\mathbf{X}$ are all values X such that $\mathbb{P}(\mathbf{X} = X) > 0$. We will use non-boldface symbols such as X or a to denote elements in the support of random variables.

Let E is an event of non-zero probability. For any event F, we denote

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

and for any real-valued random variable $\mathbf{X}$,

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

We recall the classical inequalities of Markov in Chebyshev. Let $\mathbf{X}$ be a positive real valued random variable and $\mu = \mathbb{E}\mathbf{X}$. Then for $\lambda > 0$, we have that

$$\mathbb{P}(\mathbf{X} \geq \lambda) \leq \frac{\mu}{\lambda}$$

and

$$\mathbb{P}(|\mathbf{X}-\mu| \geq \lambda\sqrt{\mathbb{E}|\mathbf{X}-\mu|^2}) \leq \frac{1}{\lambda^2}.$$

We will use the following lemma several times:

Lemma 3.1 *Let $A > 0$ and $0 \leq \epsilon \leq 1$ and $\mathbf{X}$ a random variable such that $\mu = \mathbb{E}\mathbf{X} = A(1+O_{\leq}(\epsilon))$ and $\mathbb{E}\mathbf{X}^2 = A^2(1+O_{\leq}(\epsilon))$. Then, for any $\delta > \epsilon$, we have that*

$$\mathbb{P}(|\mathbf{X}-A| \geq \delta A) \leq \frac{4\epsilon}{(\delta-\epsilon)^2} \ .$$

Proof We first see that we can easily bound the variance

$$\mathrm{Var}(\mathbf{X}) = \mathbb{E}|\mathbf{X}-\mu|^2 = \mathbb{E}\mathbf{X}^2 - \mu^2 = A^2(1+O_{\leq}(\epsilon)) - A^2(1+O_{\leq}\left(2\epsilon+\epsilon^2\right)) \leq 4\epsilon A^2$$

Next, using Chebyshev's inequality, we obtain

$$\mathbb{P}(|\mathbf{X}-A| \geq \delta A) \leq \mathbb{P}(|\mathbf{X}-\mu| \geq (\delta-\epsilon)A) \leq \frac{\mathrm{Var}(\mathbf{X})}{(\delta-\epsilon)^2A^2} \leq \frac{4\epsilon}{(\delta-\epsilon)^2} \ . \qquad \blacksquare$$

This lemma is very useful for showing that a random variable is very concentrated once we know estimates for the first and the second moment. Most of the time, we will have estimates of the form $\mathbb{E}\mathbf{X} = A(1+O(e))$ and $\mathbb{E}\mathbf{X}^2 = A(1+O(e))$, for some very small ϵ. Then by taking $\delta = \epsilon^{1/3}$, the lemma provides us with

$$\mathbb{P}\left(|\mathbf{X}-A| \geq \epsilon^{1/3}A\right) = O\left(\epsilon^{1/3}\right) \ .$$

We will also need Hoeffding's inequality:

Lemma 3.2 *Let $\mathbf{X}_1, \ldots, \mathbf{X}_m$ be independent random variables such that for all i, $\mathbb{E}\mathbf{X}_i = 0$ and $|\mathbf{X}_i| \leq B_i$ with probability 1. Then, for any $t > 0$,*

$$\mathbb{P}\left(|\mathbf{X}_1 + \cdots + \mathbf{X}_m| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2\sum_{i=1}^m B_i^2}\right)$$

4 Hypergraph covering lemma

4.1 Heuristic discussion

Consider the following general setting. Let $(V, E_i)_{i\in I}$ be a collection of hypergraphs, for some fixed finite set V and I an index set (so for each $i \in I$, E_i is a collection of subsets of V). We want to select a single edge $e_i \in E_i$ such that $\bigcup_{i\in I} e_i$ covers as much of V as possible. We think of $V \setminus \bigcup_{i\in I} e_i$ as a sifted version of V, where each e_i represents on step in the sieve.

Let's first look at the naive method of choosing random edge $\mathbf{e}_i$ uniformly at random from E_i, independently in i. In this case, the probability that a vertex $v \in V$ survives the sieve is

$$\mathbb{P}\left(v \notin \bigcup_{i\in I}\mathbf{e}_i\right) = \prod_{i\in I}(1-\mathbb{P}(v\in\mathbf{e}_i))$$

In practice, we assume that the probabilities $\mathbb{P}(v \in \mathbf{e}_i)$ are small, so $1-\mathbb{P}(v\in\mathbf{e}_i) \approx \exp(-\mathbb{P}(v\in\mathbf{e}_i))$. Hence

$$\mathbb{P}\left(v \notin \bigcup_{i\in I}\mathbf{e}_i\right) \approx \exp(-d_I(v))$$

where $d_I(v) = \sum_{i\in I}\mathbb{P}(v\in\mathbf{e}_i)$. If we have the uniformity assumption $d_I(v) \approx d$, for all $v \in V$, then

$$\mathbb{E}\left|V \setminus \bigcup_{i\in I}\mathbf{e}_i\right| \approx |V|\exp(-d)$$

so we expect that the sifted set $V \setminus \bigcup_{i \in I} \mathbf{e}_i$ to have density approximately $\exp(-d)$.

It turns out we can do better than this. Choosing $\mathbf{e}_i$ independently is inefficient because it allows for many overlaps between random edges. From now on we denote the random variable $\mathbf{e}_i$ to be uniformly at random from E_i, independently in i. We want an optimised choice of a random variable $\mathbf{e}'_i$ such that we can guarantee that $\mathbf{e}'_i$ are almost always disjoint. We impose the following uniformity conditions:

- *(Edges not too large)* For all $i \in I$, $|\mathbf{e}_i| \le r$ with probability 1;
- *(Sparisity)* For all $v \in V$, $\sum_{i \in I} \mathbb{P}(v \in \mathbf{e}_i) \le d$;
- *(Small codegrees)* For distinct $v_1, v_2 \in V$, $\sum_{i \in I} \mathbb{P}(v_1, v_2 \in \mathbf{e}_i) \le \delta d$, for some small δ.

We use the Rödl nibble method. We partition $I = I_1 \cup I_2 \cup \cdots \cup I_m$ and proceed in the following way:

- *(Nibble 1)* Let $I_1 \subset I$ small. For $i \in I_1$, we choose $\mathbf{e}'_i$ independently uniformly at random (so $\mathbf{e}'_i = \mathbf{e}_i$). Let
$$\mathbf{W}_1 := V \setminus \bigcup_{i \in I_1} \mathbf{e}'_i$$
Then, for each $v \in V$,
$$\mathbb{P}(v \in \mathbf{W}_1) = \prod_{i \in I_1} (1 - \mathbb{P}(v \in \mathbf{e}_i)) \approx \exp\left(-\sum_{i \in I_1} \mathbb{P}(v \in \mathbf{e}_i)\right) = \exp(-d_{I_1}(v)) =: P_1(v)$$
where $d_{I_1}(v) := \sum_{i \in I_1} \mathbb{P}(v \in \mathbf{e}_i)$. Then $\mathbb{E}|\mathbf{W}_1| = \sum_{v \in V} P_1(v)$ and for an small edge $e \subset V$,
$$\mathbb{P}(e \subset \mathbf{W}_1) \approx \prod_{v \in e} P_1(v) =: P_1(e)$$
since we assume we have small codegrees.
- *(Nibble 2)* Let $I_2 \subset I \setminus I_1$ small. For $i \in I_2$, we want to force $\mathbf{e}'_i \subset \mathbf{W}_1$. If $e \in E_i$ such that $e \subset \mathbf{W}_1$, we choose $\mathbf{e}'_i = e$ with probability proportional to $1/P_1(e)$. So take
$$\mathbb{P}(\mathbf{e}'_i = e | \mathbf{W}_1 = W_1) = \begin{cases} \frac{c_i}{P_1(e)} & \text{if } e \subset W_1 \\ 0 & \text{otherwise} \end{cases}$$
for some normalising constant c_i, for all $i \in I_2$. So $\mathbf{e}'_i$ is $\mathbf{e}_i$ conditioned to to the event $\mathbf{e}'_i \subset W_1$ and then reweighted by $P_1(\mathbf{e}_i)$ to compensate for the bias based by the conditioning. Also, they are jointly independent for $i \in I_2$. Note that we have a problem if $P_1(e) = 0$, but this is not expected behavior and we'll treat everything carefully later. Let
$$\mathbf{W}_2 := \mathbf{W}_1 \setminus \bigcup_{i \in I_2} \mathbf{e}'_i \ .$$
Fix $v \in V$. Then we will see that
$$\mathbb{P}(v \in \mathbf{W}_2) \approx \mathbb{P}(v \in \mathbf{W}_1) \prod_{i \in I_2} (1 - \mathbb{P}(v \in \mathbf{e}'_i | v \in \mathbf{W}_1)) \approx P_1(v) \exp\left(-\frac{\sum_{i \in I_2} \mathbb{P}(v \in \mathbf{e}_i)}{P_1(v)}\right) =: P_2(v)$$
For a small edge $e \subset V$, we obtain that
$$\mathbb{P}(e \subset \mathbf{W}_2) \approx \prod_{v \in w} P_2(v) =: P_2(e) \ .$$
- *(Nibble m)* For $i \in I_m$ and $e \in E_i$ and $e \subset \mathbf{W}_{m-1}$, we choose $\mathbf{e}'_i = e$ with probability proportional to $1/P_{m-1}(e)$. Let $\mathbf{W}_m = \mathbf{W}_{m-1} \setminus \bigcup_{i \in I_m} \mathbf{e}'_i$ and $d_{I_m}(v) = \sum_{i \in I_m} \mathbb{P}(v \in \mathbf{e}_i)$. Then for all $v \in V$,
$$\mathbb{P}(v \in \mathbf{W}_m) \approx P_{m-1}(v) \exp\left(-\frac{d_{I_m}(v)}{P_{m-1}(v)}\right) =: P_m(v) \ ,$$
where and for small $e \subset V$, we have
$$\mathbb{P}(e \subset \mathbf{W}_m) \approx \prod_{v \in e} P_m(v) =: P_m(e) \ .$$

In view of the algorithm described above, it is useful to have the following definitions:

Definition 4.1 *Let $(V, E_i)_{i\in I}$ a collection of hypergraphs and for each $i \in I$, let $\mathbf{e}_i$ be a random set of V supported on E_i. Let $I = I_1 \cup \cdots \cup I_m$. For all $j \in [m]$ and $v \in V$ we define the normalised degrees*

$$d_j(v) := \sum_{i\in I_j} \mathbb{P}(v \in \mathbf{e}_i) \tag{4.1}$$

and then we recursively define $P_j(v)$ by setting $P_0(v) = 1$ and

$$P_{j+1}(v) = P_j(v) \exp\left(-\frac{d_{j+1}(v)}{P_j(v)}\right) \ . \tag{4.2}$$

The key idea of the process described above is that for $i, j \in I$, $\mathbf{e}'_i$ and $\mathbf{e}'_j$ will always be disjoint, unless i and j belong to the same nibble. We formalise the process in the following theorem:

Theorem 4.2 *Let $D, r, A \geq 1$, $O < \delta, \kappa \leq 1/2$ and $m \geq 0$ an integer. Let V be a finite set. Let $I_1, \dots I_m$ be finite index sets, and for each $j \in [m]$ and $i \in I_j$, let $\mathbf{e}_i$ be a random subset of V such that:*

- *(Edges not too large) With probability 1, for all $j \in [m]$ and $i \in I_j$*

$$\#\mathbf{e}_i \leq r \tag{4.3}$$

- *(Sparsity) For all $j \in [m]$, $i \in I_j$ and $v \in V$*

$$\mathbb{P}(v \in \mathbf{e}_i) \leq \frac{\delta}{|I_j|^{1/2}} \tag{4.4}$$

- *(Small codegrees) For all $j \in [m]$ and distinct $v_1, v_2 \in V$*

$$\sum_{i\in I_j} \mathbb{P}(v_1, v_2 \in \mathbf{e}_i) \leq \delta \tag{4.5}$$

- *(Degree bounds) For $j \in [m]$ and $v \in V$, let $d_j(v)$ and $P_j(v)$ be defined as in Definition 4.1. Then, for $j \in [m]$ and $v \in V$, we have*

$$d_j(v) \leq DP_{j-1}(v) \tag{4.6}$$

  and

$$P_j(v) \geq \kappa \ . \tag{4.7}$$

- *(Smallness bound for δ) There exists a universal constant $C_0 \geq 1$ such that*

$$\delta \leq \left(\frac{\kappa^A}{C_0 \exp(AD)}\right)^{10^{m+2}} \tag{4.8}$$

Then, for all $i \in \cup_{j=1}^m I_j$, we can find random subsets $\mathbf{e}'_i$ such that the support of $\mathbf{e}'_i$ is contained in the support of $\mathbf{e}_i$ union with the empty set and for all $0 \leq J \leq m$ and $e \subset V$ with $|e| \leq A - 2rJ$, we have

$$\mathbb{P}\left(e \subset V \setminus \bigcup_{j=1}^{J} \bigcup_{i\in I_j} \mathbf{e}'_i\right) = \left(1 + O_{\leq}\left(\delta^{1/10^{J+1}}\right)\right) P_J(e) \tag{4.9}$$

where $P_J(e) := \prod_{v\in e} P_J(v)$.

The proof will be described in detail in the next section. We now focus on the main application of it:

Lemma 4.3 *Let $f(x)$ such that $\lim_{x\to\infty} f(x) = \infty$. Let I and V be sets with $|I| \leq x$ and $f(x)^2 \leq |V| \leq x^{50}$. Let $\delta \leq x^{-1/50}$. For each $i \in I$, let $\mathbf{e}_i$ be a random subset of V such that*

- *(Edges not too large) With probability 1, for all $i \in I$*

$$\#\mathbf{e}_i \leq r(x) \tag{4.10}$$

- *(Conditions on growth of $f(x)$ and $r(x)$) We have that*

$$r(x)f(x)^{3/2} \leq \log x \tag{4.11}$$

- *(Sparsity) For all $i \in I$ and $v \in V$*

$$\mathbb{P}(v \in \mathbf{e}_i) \leq \delta x^{-1/2} \tag{4.12}$$

- *(Small codegrees) For any two distinct $v_1, v_2 \in V$,*

$$\sum_{i \in I} \mathbb{P}(v_1, v_2 \in \mathbf{e}_i) \leq \delta \tag{4.13}$$

- *(Elements covered more than once in expectation) For all but at most $o\left(\frac{|V|}{f(x)}\right)$ elements $v \in V$, we have*

$$\sum_{i \in I} \mathbb{P}(v \in \mathbf{e}_i) = C + O_{\leq}\left(\frac{1}{f(x)^2}\right) \tag{4.14}$$

  for some global constant C satisfying $\frac{5}{4}\log 5 \leq C \ll 1$.

Then, for all $i \in I$ we can find random subsets $\mathbf{e}'_i \subseteq V$ which is either empty or a subset of V which $\mathbf{e}_i$ attains with positive probability such that

$$\#\{v \in V \ : \ v \notin \mathbf{e}'_i, \ \textit{for all } i \in I\} \sim \frac{|V|}{f(x)} \tag{4.15}$$

with probability $1 - o(1)$.

Before we proceed with the proof, we would like to discuss why this lemma is very useful. Very informally, it states that if we have a collection on hypergraphs such that all edges are small, the degrees are on average small and the codegrees very small, and that if every vertex is covered on average at least once, we can find a very effective almost covering of V with one edge from each hypergraph. If we compare it with the naive approach of choosing uniformly independently at random as discussed in the beginning of the section, we expect that the density of the elements left uncovered is around $\exp(-C)$. This method provides us a density of at most $1/f(x)$.

We chose to expose this lemma in slightly more generality than in [6] because we have liberty in choosing our parameters $f(x)$ and $r(x)$ that might be useful in further applications. Sometimes it might be useful to have larger bound for the size of edges, which means $r(x)$ larger, and sometimes we might want to guarantee better covering, which means $f(x)$ larger.

We now proceed with the proof of Lemma 4.3 assuming Theorem 4.2.

Proof The number of elements in V that fail (4.14) is $o\left(\frac{|V|}{f(x)}\right)$, so we may discard these elements from V and assume (4.14) holds for all $v \in V$ (it is easy to see that this does not influence our assumptions or the conclusion we want to prove).

Let $m = \lfloor \frac{\log f(x)}{\log 5} \rfloor$ (so m is the largest integer such that $5^m \leq f(x)$). We want to find $I_1, I_2, \dots I_m$ disjoint subsets of I such that uniformly for all $v \in V$ and $j \in [m]$:

$$\sum_{i \in I_j} \mathbb{P}(v \in \mathbf{e}_i) = 5^{1-j}\log 5 + O\left(\frac{1}{f(x)^2}\right) \tag{4.16}$$

For this purpose, let $\mathbf{t}_i$ be uniform in $[0,1]$ and independent for each $i \in I$. Let $\vec{\mathbf{t}} = (\mathbf{t}_i)_{i \in I}$. Since $C \geq \frac{5}{4}\log 5$, we find disjoint intervals $\mathscr{I}_1, \dots \mathscr{I}_m$ in $[0,1]$ with $|\mathscr{I}_j| = \frac{5^{1-j}\log 5}{C}$. We define the random sets

$$I_j(\vec{\mathbf{t}}) := \{i \in I : \mathbf{t}_i \in \mathscr{I}_j\} \ .$$

For $v \in V$, $j \in [m]$ and $i \in I$ we consider the independent random variables

$$\mathbf{X}_{v,i,j}(\vec{\mathbf{t}}) = \begin{cases} \mathbb{P}(v \in \mathbf{e}_i) & \text{if } i \in I_j(\vec{\mathbf{t}}) \\ 0 & \text{otherwise.} \end{cases}$$

Then we clearly have that

$$\sum_{i\in I}\mathbb{E}\mathbf{X}_{v,i,j}=\sum_{i\in I}\mathbb{P}(v\in\mathbf{e}_i)\mathbb{P}(i\in I_j(\vec{\mathbf{t}})=|\mathscr{I}_j|\sum_{i\in I}\mathbb{P}(v\in\mathbf{e}_i)=5^{1-j}\log 5+O_{\le}\left(\frac{4/5}{f(x)^2}\right) \tag{4.17}$$

using (4.14). Since $|\mathbf{X}_{v,i,j}|\le\delta x^{-1/2}$ by (4.12), we apply Lemma 3.2 to obtain

$$\mathbb{P}\left(|\sum_{i\in I}(\mathbf{X}_{v,i,j}(\vec{\mathbf{t}})-\mathbb{E}\mathbf{X}_{v,i,j}(\vec{\mathbf{t}})|\ge\frac{1}{f(x)^2}\right)\le 2\exp\left(-\frac{f(x)^{-4}}{2\delta^2x^{-1}|I|}\right)\le 2\exp\left(-\frac{1}{2\delta^2f(x)^4}\right)\ll x^{-100}$$

since $\delta\le x^{-1/50}$ and $f(x)\le\log x$. Since $|V|\le x^{50}$ and $m\le x$, there exists a deterministic choice $\vec{t}$ of $\vec{\mathbf{t}}$ such that for all $v\in V$ and $j\in[m]$, we have

$$|\sum_{i\in I}(\mathbf{X}_{v,i,j}(\vec{t})-\mathbb{E}\mathbf{X}_{v,i,j}(\vec{t})|\le\frac{1}{f(x)^2} \tag{4.18}$$

Our choice of $\vec{t}$ will give us a deterministic choice for I_j. Putting together (4.17) and (4.18), we obtain

$$\sum_{i\in I}\mathbf{X}_{v,i,j}(\vec{t})=\sum_{i\in I_j}\mathbb{P}(v\in\mathbf{e}_i)=5^{1-j}\log 5+O_{\le}\left(\frac{2}{f(x)^2}\right) \tag{4.19}$$

for all $v\in V$, $j\in J$, which is exactly what we wanted to prove in (4.16).

We are now ready to look at the normalised degrees. Since $5^m\le f(x)$, the last equation implies that

$$d_{I_j}(v)=5^{1-j}\log 5\left(1+O_{\le}\left(\frac{2}{f(x)}\right)\right)$$

for all $v\in V$ and $j\in[m]$. We prove by induction that

$$P_j(v)=5^{-j}(1+O_{\le}(4^j/f(x))) \tag{4.20}$$

Indeed, it is easy to check for $j=1$ and

$$\begin{aligned}P_{j+1}(v)&=P_j(v)\exp\bigl(-d_{I_{j+1}}(v)/P_j(v)\bigr)=5^{-j}(1+O_{\le}(4^j/f(x)))\exp\left(-\frac{5^{-j}\log 5(1+O_{\le}(2/f(x)))}{5^{-j}(1+O_{\le}(4^j/f(x)))}\right)\\&=5^{-(j+1)}(1+O_{\le}(4^j/f(x)))\exp\bigl(O_{\le}(2\cdot 4^j/f(x))\bigr)=5^{-(j+1)}(1+O_{\le}(4^{j+1}/f(x)))\end{aligned}$$

as desired. Hence, since $j\le m\le\log f(x)/\log 5$, if we take $\nu=1-\frac{\log 4}{\log 5}$, we obtain that

$$P_j(v)=5^{-j}(1+O_{\le}(f(x)^{-\nu})) \tag{4.21}$$

for all $v\in V$ and $j\in[m]$.

We check that we satisfy all the conditions of Theorem 4.2. Let $A=2rm+2$. We see that (4.11) is chosen such that (4.8) is satisfied and all other conditions follow easily. Indeed, let $\kappa=\frac{1}{2f(x)}$ and $D=1$ to see that (4.6) and (4.7) are satisfied. Also

$$f(x)^{\frac{\log 10}{\log 5}}r(x)\log f(x)\ll\log x$$

since $\log 10/\log 5<3/2$ and then (4.8) follows.

Hence, we obtain random variables $\mathbf{e}'_i$, for $i\in\cup_{j=1}^m I_j$ such that

$$\mathbb{P}\left(v\notin\bigcup_{j=1}^m\bigcup_{i\in I_j}\mathbf{e}'_i\right)=\left(1+O_{\le}\left(\delta^{1/10^{m+1}}\right)\right)P_m(v) \tag{4.22}$$

for all $v\in V$, and

$$\mathbb{P}\left(v_1,v_2\notin\bigcup_{j=1}^m\bigcup_{i\in I_j}\mathbf{e}'_i\right)=\left(1+O_{\le}\left(\delta^{1/10^{m+1}}\right)\right)P_m(v_1)P_m(v_2) \tag{4.23}$$

for distinct $v_1, v_2 \in V$. We set $\mathbf{e}'_i = \emptyset$ for $i \in I \setminus \bigcup_{j=1}^m I_j$.

We want to use Lemma 3.1 to approximate $\#\{v \in V : v \notin \mathbf{e}'_i, \text{ for all } i \in I\}$. For this purpose, we define the random variable

$$\mathbf{X} := \#\{v \in V \;:\; v \notin \mathbf{e}'_i, \text{ for all } i \in I\} = \sum_{v \in V} 1_{v \notin \cup_{j=1}^m \cup_{i \in I_j} \mathbf{e}'_i}$$

and we need to approximate $\mathbb{E}\mathbf{X}$ and $\mathbb{E}\mathbf{X}^2$. From (4.22) and from the upper bound on δ we see that

$$\mathbb{E}\mathbf{X} = \frac{1}{f(x)} \left(1 + O\left(f(x)^{-\nu}\right)\right) |V|$$

and from (4.23) we obtain that

$$\begin{aligned}
\mathbb{E}\mathbf{X}^2 &= \frac{1}{f(x)^2} \left(1 + O\left(f(x)^{-\nu}\right)\right) |V|(|V|-1) + \frac{1}{f(x)^2} \left(1 + O\left(f(x)^{-\nu}\right)\right) |V| \\
&= \frac{1}{f(x)^2} \left(1 + O\left(f(x)^{-\nu}\right)\right) |V|^2
\end{aligned}$$

Using that $|V| \geq f(x)^2$ we obtain that

$$\mathbb{P}\left(\left|\mathbf{X} - \frac{|V|}{f(x)}\right| \geq |V| f(x)^{-1-\nu/3}\right) \ll f(x)^{-\nu/3}$$

which implies the conclusion. ∎

4.2 Proof of Theorem 4.2

We proceed by induction on m. There is nothing to prove for $m = 0$, so suppose $m \geq 1$ and that the conclusion is proved for $m-1$. Hence, for $i \in \cup_{j=1}^{m-1} I_j$ we find random variables $\mathbf{e}'_i$ such that

$$\mathbb{P}\left(e \subset V \setminus \bigcup_{j=1}^{m-1} \bigcup_{i \in I_j} \mathbf{e}'_i\right) = \left(1 + O_{\leq}\left(\delta^{1/10^m}\right)\right) P_{m-1}(e) \tag{4.24}$$

for all $e \subset V$ of cardinality at most $A - 2r(m-1)$. Denote the random variable

$$\mathbf{W} = V \setminus \bigcup_{j=1}^{m-1} \bigcup_{i \in I_j} \mathbf{e}'_i \,.$$

Our goal is to find random variables $\mathbf{e}'_i$, for $i \in I_m$, whose support is contained in that of $\mathbf{e}_i$ together with the empty set such that

$$\mathbb{P}\left(e \subset \mathbf{W} \setminus \bigcup_{i \in I_m} \mathbf{e}'_i\right) = \left(1 + O_{\leq}\left(\delta^{1/10^{m+1}}\right)\right) P_m(e) \tag{4.25}$$

for all $e \subset V$ with $|e| \leq A - 2rm$. Hence we may assume $A \geq 2rm$ if we want our theorem to have any meaning.

As discussed earlier, we want to choose $\mathbf{e}'_i$ with probability proportional to $1/P_{m-1}(e)$ conditional on $\mathbf{e}_i \subset \mathbf{W}$. For this purpose, for each W in the essential range of $\mathbf{W}$ and $i \in I_m$, we define the normalisation factor

$$X_i(W) = \sum_{e_i \subset W} \frac{\mathbb{P}(\mathbf{e}_i = e_i)}{P_{m-1}(e_i)} \tag{4.26}$$

We want to avoid the cases where $X_i(W) = 0$. In fact, we will show that $X_i(\mathbf{W})$ is very close to 1 with high probability. We define

$$F_i(W) = \begin{cases} 1 & \text{if } |X_i(W) - 1| \leq \delta^{\frac{1}{3 \times 10^m}} \\ 0 & \text{otherwise.} \end{cases} \tag{4.27}$$

Now we are ready to state the definition of $\mathbf{e}_i'$. If $F_i(W) = 0$, we let $\mathbf{e}_i' = \emptyset$, or in other words, $\mathbb{P}(\mathbf{e}_i' = \emptyset | \mathbf{W} = W) = 1$. Otherwise, if $F_i(W) = 1$, for each $e_i \in E_i$ we define

$$\mathbb{P}(\mathbf{e}_i' = e_i | \mathbf{W} = W) = \begin{cases} \frac{\mathbb{P}(\mathbf{e}_i = e_i)}{X_i(W) P_{m-1}(e_i)} & \text{if } e_i \subset W \\ 0 & \text{otherwise.} \end{cases} \tag{4.28}$$

We check this is a well defined probability distribution:

$$\sum_{e_i \in E_i} \mathbb{P}(\mathbf{e}_i = e_i | \mathbf{W} = W) = \frac{1}{X_i(W)} \sum_{e_i \in W} \frac{\mathbb{P}(\mathbf{e}_i = e_i)}{P_{m-1}(e_i)} = 1$$

In order to make the proof easier to follow, we split it into several lemmas. First we check that $F_i(W) = 1$ with high probability:

Lemma 4.4

$$\mathbb{P}(F_i(\mathbf{W}) = 1) = \mathbb{E}(F_i(\mathbf{W})) = 1 - O\left(\delta^{\frac{1}{3\times 10^m}}\right) .$$

Proof We see that

$$\mathbb{P}(F_i(\mathbf{W}) = 0) = \mathbb{P}(|X_i(\mathbf{W}) - 1| \geq \delta^{\frac{1}{3\times 10^m}})$$

and this suggests using Lemma 3.1. So we want to provide good approximations for $\mathbb{E}X_i(\mathbf{W})$ and $\mathbb{E}(X_i(\mathbf{W})^2)$. We begin with the first one:

$$\begin{aligned} \mathbb{E}X_i(\mathbf{W}) &= \sum_W \sum_{e_i \subset W} \frac{\mathbb{P}(\mathbf{e}_i = e_i)}{P_{m-1}(e_i)} \mathbb{P}(\mathbf{W} = W) = \sum_{e_i} \frac{\mathbb{P}(\mathbf{e}_i = e_i)}{P_{m-1}(e_i)} \mathbb{P}(e_i \subset \mathbf{W}) \\ &= \sum_{e_i} \frac{\mathbb{P}(\mathbf{e}_i = e_i)}{P_{m-1}(e_i)} \left(1 + O_{\leq}\left(\delta^{1/10^m}\right)\right) P_{m-1}(e_i) = \left(1 + O_{\leq}\left(\delta^{1/10^m}\right)\right) \end{aligned}$$

using (4.24). Now we compute the second moment:

$$\begin{aligned} \mathbb{E}(X_i(\mathbf{W})^2) &= \sum_W \sum_{e_i, f_i \subset W} \frac{\mathbb{P}(\mathbf{e}_i = e_i)}{P_{m-1}(e_i)} \frac{\mathbb{P}(\mathbf{e}_i = f_i)}{P_{m-1}(f_i)} \mathbb{P}(\mathbf{W} = W) = \sum_{e_i, f_i} \frac{\mathbb{P}(\mathbf{e}_i = e_i)}{P_{m-1}(e_i)} \frac{\mathbb{P}(\mathbf{e}_i = f_i)}{P_{m-1}(f_i)} \mathbb{P}(e_i \cup f_i \subset \mathbf{W}) \\ &= \left(1 + O_{\leq}\left(\delta^{1/10^m}\right)\right) \sum_{e_i, f_i} \mathbb{P}(\mathbf{e}_i = e_i) \mathbb{P}(\mathbf{e}_i = f_i) \frac{1}{P_{m-1}(e_i \cap f_i)} \end{aligned}$$

using again (4.24) and that $P_{m-1}(e_i)P_{m-1}(f_i) = P_{m-1}(e_i \cup f_i)P_{m-1}(e_i \cap f_i)$. We note that $P_{m-1}(e_i \cap f_i)$ is 1 if $e_i \cap f_i = \emptyset$ and it is at least κ^r otherwise, using (4.7). Hence

$$\begin{aligned} \sum_{e_i, f_i} \mathbb{P}(\mathbf{e}_i = e_i) \mathbb{P}(\mathbf{e}_i = f_i) \frac{1}{P_{m-1}(e_i \cap f_i)} &= 1 + \sum_{e_i \cap f_i \neq \emptyset} \mathbb{P}(\mathbf{e}_i = e_i) \mathbb{P}(\mathbf{e}_i = f_i) \left(\frac{1}{P_{m-1}(e_i \cap f_i)} - 1\right) \\ &= 1 + O\left(k^{-r} \sum_{e_i} \mathbb{P}(\mathbf{e}_i = e_i) \sum_{v \in e_i} \mathbb{P}(v \in \mathbf{e}_i)\right) = 1 + O\left(r\delta k^{-r}\right) \end{aligned}$$

where we used (4.4) for the last equality. But it is easy to see from (4.8) that

$$rk^{-r} \leq Ak^{-A} \leq \delta^{-\frac{1}{10^{m+2}}}$$

which clearly implies $r\delta\kappa^{-r} \leq \delta^{\frac{1}{10^m}}$. Hence

$$\mathbb{E}(X_i(\mathbf{W})^2) = \left(1 + O\left(\delta^{1/10^m}\right)\right) .$$

Now just applying Lemma 3.1, we obtain $\mathbb{P}(|X_i(\mathbf{W}) - 1| \geq \delta^{\frac{1}{3\times 10^m}}) \ll \delta^{\frac{1}{3\times 10^m}}$, as desired. ∎

Fix $e \subset V$ with $|e| \leq A - 2rm$. Recall that our aim is to prove that (4.25). We note that

$$\begin{aligned} \mathbb{P}\left(e \subset \mathbf{W} \setminus \bigcup_{i \in I_m} \mathbf{e}_i'\right) &= \sum_W \mathbb{P}\left(e \subset W \setminus \bigcup_{i \in I_m} \mathbf{e}_i' | \mathbf{W} = W\right) \mathbb{P}(\mathbf{W} = W) \\ &= \sum_{e \subset W} \mathbb{P}\left(e \subset W \setminus \bigcup_{i \in I_m} \mathbf{e}_i' | \mathbf{W} = W\right) \mathbb{P}(\mathbf{W} = W) \end{aligned}$$

since $\mathbb{P}\left(e \subset W \setminus \bigcup_{i\in I_m} \mathbf{e}'_i | \mathbf{W} = W\right) = 0$ if $e \cap W = \emptyset$. By the induction hypothesis (4.24), we know that

$$\mathbb{P}(e \subset \mathbf{W}) = \sum_{e\subset W} \mathbb{P}(\mathbf{W} = W) = P_{m-1}(e)\left(1 + O_{\leq}\left(\delta^{1/10^m}\right)\right).$$

Hence it is enough to prove that

$$\frac{1}{\mathbb{P}(e \subset \mathbf{W})} \sum_{e\subset W} \mathbb{P}\left(e \subset W \setminus \bigcup_{i\in I_m} \mathbf{e}'_i | \mathbf{W} = W\right) \mathbb{P}(\mathbf{W} = W) = \left(1 + O\left(\delta^{\frac{1}{9\times 10^m}}\right)\right) \exp\left(-\sum_{v\in e} \frac{d_{I_m}(v)}{P_{m-1}(v)}\right) \tag{4.29}$$

Let $Y(W) = \mathbb{P}\left(e \subset W \setminus \bigcup_{i\in I_m} \mathbf{e}'_i | \mathbf{W} = W\right)$. Then (4.29) is equivalent to

$$\mathbb{E}(Y(\mathbf{W}) | e \subset \mathbf{W}) = \left(1 + O\left(\delta^{\frac{1}{9\times 10^m}}\right)\right) \exp\left(-\sum_{v\in e} \frac{d_{I_m}(v)}{P_{m-1}(v)}\right) \tag{4.30}$$

Lemma 4.5 *Let W in the essential range of $\mathbf{W}$ such that $e \subset W$. Then*

$$Y(W) = \mathbb{P}\left(e \subset W \setminus \bigcup_{i\in I_m} \mathbf{e}'_i | \mathbf{W} = W\right) = \left(1 + O\left(\delta^{1/10^m}\right)\right) \exp\left(-\sum_{i\in I_m}\sum_{v\in e} \mathbb{P}(v \in \mathbf{e}'_i | \mathbf{W} = W)\right)$$

Proof We observe that, since $\mathbf{e}'_i$ are jointly independent for each $i \in I_m$ conditional on the event $\mathbf{W} = W$:

$$\mathbb{P}\left(e \subset W \setminus \bigcup_{i\in I_m} \mathbf{e}'_i | \mathbf{W} = W\right) = \prod_{i\in I_m} (1 - \mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W))$$

Of course, we want to approximate the product on the right hand side by an exponential, so we need to find upper bounds for the probabilities in order to estimate the error terms. Clearly $\mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W) = 0$ is 0 if $F_i(W) = 0$, so now we consider the case $F_i(W) = 1$.

$$\begin{aligned}\mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W) &\leq \sum_{v\in e} \mathbb{P}(v \in \mathbf{e}'_i | \mathbf{W} = W) \leq \sum_{v\in e}\sum_{e_i : v\in e_i} \mathbb{P}(\mathbf{e}'_i = e_i | \mathbf{W} = W) \\ &\ll A\kappa^{-r}\mathbb{P}(v \in \mathbf{e}_i) \ll A\kappa^{-r}\delta |I_m|^{-1/2}\end{aligned}$$

where we have used the definition of $\mathbf{e}'_i$ (4.28), $|e| \leq A$ and the sparsity assumption (4.4). Hence

$$1 - \mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W) = \exp\left(-\mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W)\right) + O\left(\frac{(A\kappa^{-r}\delta)^2}{|I_m|}\right)$$

Using (4.8), we notice again that $A^2\kappa^{-2r} \leq A^2\kappa^{-A} \leq \delta^{-\frac{1}{10^{m+2}}}$, and so $(A\kappa^{-2r}\delta)^2 \leq \delta^{1/10^m}$. Hence

$$\prod_{i\in I_m} (1 - \mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W)) = \left(1 + O\left(\delta^{1/10^m}\right)\right) \exp\left(-\sum_{i\in I_m} \mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W)\right) \tag{4.31}$$

Next we see that

$$\mathbb{P}(e \cap \mathbf{e}'_i \neq \emptyset | \mathbf{W} = W) = \sum_{v\in e} \mathbb{P}(v \in \mathbf{e}'_i | \mathbf{W} = W) + O\left(\sum_{\substack{v,w,\in e\\ v\neq w}} \mathbb{P}(v, w \in \mathbf{e}'_i | \mathbf{W} = W)\right) \tag{4.32}$$

We estimate the error term:

$$\begin{aligned}\sum_{i\in I_m}\sum_{\substack{v,w,\in e\\ v\neq w}} \mathbb{P}(v, w \in \mathbf{e}'_i | \mathbf{W} = W) &= \sum_{i\in I_m}\sum_{\substack{v,w,\in e\\ v\neq w}}\sum_{e_i : v,w\in e_i} \mathbb{P}(\mathbf{e}'_i = e_i | \mathbf{W} = W) \\ &\ll \sum_{i\in I_m}\sum_{\substack{v,w,\in e\\ v\neq w}} \kappa^{-r}\ \mathbb{P}(v, w \in \mathbf{e}_i) \ll A^2\kappa^{-r}\delta\end{aligned}$$

using the condition (4.5). Similarly as before, $A^2\kappa^{-r}\delta \leq \delta^{1/10^m}$, so putting together (4.31) and (4.32) we obtain

$$\prod_{i\in I_m}(1-\mathbb{P}(e\cap \mathbf{e}_i' \neq \emptyset|\mathbf{W}=W)) = \left(1+O\left(\delta^{1/10^m}\right)\right)\exp\left(-\sum_{i\in I_m}\sum_{v\in e}\mathbb{P}(v\in \mathbf{e}_i'|\mathbf{W}=W)+O\left(\delta^{1/10^m}\right)\right)$$
$$= \left(1+O\left(\delta^{1/10^m}\right)\right)\exp\left(-\sum_{i\in I_m}\sum_{v\in e}\mathbb{P}(v\in \mathbf{e}_i'|\mathbf{W}=W)\right) \quad \blacksquare$$

Lemma 4.6 *Conditionally on $e\subset \mathbf{W}$, we have that*

$$\sum_{i\in I_m}\mathbb{P}(v\in \mathbf{e}_i'|\mathbf{W}) = \sum_{i\in I_m}\sum_{\substack{e_i:v\in e_i\\ e_i\subset W}}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)} + O\left(\delta^{\frac{1}{8\times 10^m}}\right)$$

with probability $1-O\left(\delta^{\frac{1}{8\times 10^m}}\right)$.

Proof We note that $\mathbb{P}(v\in \mathbf{e}_i'|\mathbf{W}=W) = F_i(W)\mathbb{P}(v\in \mathbf{e}_i'|\mathbf{W}=W)$. Hence

$$\mathbb{P}(v\in \mathbf{e}_i'|\mathbf{W}=W) = \frac{F_i(W)}{X_i(W)}\sum_{\substack{e_i:v\in e_i\\ e_i\subset W}}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)} = \left(1+O\left(1-F_i(W)+\delta^{\frac{1}{3\times 10^m}}\right)\right)\sum_{\substack{e_i:v\in e_i\\ e_i\subset W}}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)}$$

We use this approximation because, from Lemma 4.4, we expect $F_i(W)=1$ most of the time. Hence

$$\sum_{i\in I_m}\mathbb{P}(v\in \mathbf{e}_i'|\mathbf{W}=W) = \sum_{i\in I_m}\sum_{\substack{e_i:v\in e_i\\ e_i\subset W}}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)} + \sum_{i\in I_m}O\left(1-F_i(W)+\delta^{\frac{1}{3\times 10^m}}\right)\sum_{\substack{e_i:v\in e_i\\ e_i\subset W}}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)}$$

We aim to find an bound for the error term. From Lemma 4.4, we know that $\mathbb{E}(F_i(\mathbf{W}))=1-O\left(\delta^{\frac{1}{3\times 10^m}}\right)$, which implies that

$$\mathbb{E}(F_i(\mathbf{W})|e\subset \mathbf{W}) = \frac{1}{\mathbb{P}(e\subset \mathbf{W})}\sum_{e\subset W}F_i(W)\mathbb{P}(\mathbf{W}=W) = \frac{\mathbb{P}(e\subset \mathbf{W})-O\left(\delta^{\frac{1}{3\times 10^m}}\right)}{\mathbb{P}(e\subset \mathbf{W})} = 1-\frac{O\left(\delta^{\frac{1}{3\times 10^m}}\right)}{P_{m-1}(e)}$$

Let

$$Z(W) := \sum_{i\in I_m}\left(1-F_i(W)+\delta^{\frac{1}{3\times 10^m}}\right)\sum_{\substack{e_i:v\in e_i\\ e_i\subset W}}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)}$$

Then clearly

$$Z(W) \ll \kappa^{-r}\sum_{i\in I_m}\left(1-F_i(W)+\delta^{\frac{1}{3\times 10^m}}\right)\mathbb{P}(v\in \mathbf{e}_i)$$

hence

$$\mathbb{E}(Z(\mathbf{W})|e\subset \mathbf{W}) \ll \frac{\kappa^{-r}}{P_{m-1}(e)}\delta^{\frac{1}{3\times 10^m}}d_{I_m}(v) \ll \kappa^{-2r}\delta^{\frac{1}{3\times 10^m}}d_{I_m}(v) \ll \kappa^{-A}\delta^{\frac{1}{3\times 10^m}} \ll \delta^{\frac{1}{4\times 10^m}}$$

We now use Markov's inequality to obtain that

$$\mathbb{P}(Z(\mathbf{W})\geq \delta^{\frac{1}{8\times 10^m}}) \ll \delta^{\left(\frac{1}{4\times 10^m}-\frac{1}{8\times 10^m}\right)} = \delta^{\frac{1}{8\times 10^m}}\ . \quad \blacksquare$$

Lemma 4.7 *Conditionally on $e\subset \mathbf{W}$, we have*

$$\sum_{i\in I_m}\sum_{e_i:v\in e_i}1_{e_i\subset \mathbf{W}}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)} = \frac{d_{I_m}(v)}{P_{m-1}(v)} + O\left(\delta^{\frac{1}{3\times 10^m}}\right)$$

with probability $1-O\left(\delta^{\frac{1}{3\times 10^m}}\right)$.

Proof Let

$$X(\mathbf{W}) = \sum_{i\in I_m}\sum_{e_i:v\in e_i} 1_{e_i\subset \mathbf{W}}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)}\ .$$

We would like to use once again Lemma 3.1, so we want to approximate $\mathbb{E}(X(\mathbf{W})|e \subset \mathbf{W})$ and $\mathbb{E}(X(\mathbf{W})^2|e \subset \mathbf{W})$. We begin with the first one:

$$\begin{aligned}\mathbb{E}(X(\mathbf{W})|e\subset\mathbf{W}) &= \frac{1}{\mathbb{P}(e\subset\mathbf{W})}\sum_{W:e\subset W}\sum_{i\in I_m}\sum_{e_i:v\in e_i} 1_{e_i\subset W}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)}\mathbb{P}(\mathbf{W}=W)\\ &= \sum_{i\in I_m}\sum_{e_i:v\in e_i}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)}\frac{\mathbb{P}(e\cup e_i\subset\mathbf{W})}{\mathbb{P}(e\subset\mathbf{W})}\\ &= \left(1+O\left(\delta^{1/10^m}\right)\right)\sum_{i\in I_m}\sum_{e_i:v\in e_i}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)}\frac{P_{m-1}(e\cup e_i)}{P_{m-1}(e)}\end{aligned}$$

using the induction hypothesis (4.24). We proceed similarly to the proof of Lemma 4.4. Since $v \in e\cap e_i$, we have that

$$\frac{P_{m-1}(e\cup e_i)}{P_{m-1}(e_i)P_{m-1}(e)} = \frac{1}{P_{m-1}(v)P_{m-1}(e_i\cap e\setminus\{v\})}$$

Now, $P_{m-1}(e_i \cap e \setminus \{v\})$ is 1 if e_i and $e \setminus \{v\}$ are disjoint and at least κ^r otherwise. Hence

$$\begin{aligned}\sum_{i\in I_m}\sum_{e_i:v\in e_i}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)}\frac{P_{m-1}(e\cup e_i)}{P_{m-1}(e)} &= \sum_{i\in I_m}\frac{\mathbb{P}(v\in\mathbf{e}_i)}{P_{m-1}(v)} + O\left(\kappa^{-r}\sum_{i\in I_m}\sum_{w\in e\setminus\{v\}}\frac{\mathbb{P}(v,w\in\mathbf{e}_i)}{P_{m-1}(v)}\right)\\ &= \frac{d_{I_m}(v)}{P_{m-1}(v)} + O\left(\kappa^{-(r+1)}\delta A\right) = \frac{d_{I_m}(v)}{P_{m-1}(v)} + O\left(\delta^{1/10^m}\right)\end{aligned}$$

Hence, we have that

$$\mathbb{E}(X(\mathbf{W})|e\subset\mathbf{W}) = \left(1+O\left(\delta^{1/10^m}\right)\right)\frac{d_{I_m}(v)}{P_{m-1}(v)} \tag{4.33}$$

We now study $\mathbb{E}(X(\mathbf{W})^2|e \subset \mathbf{W})$:

$$\begin{aligned}\mathbb{E}(X(\mathbf{W})^2|e\subset\mathbf{W}) &= \frac{1}{\mathbb{P}(e\subset\mathbf{W})}\sum_{W:e\subset W}\sum_{i,j\in I_m}\sum_{e_i:v\in e_i}\sum_{f_j:v\in f_j} 1_{e_i\subset W}1_{f_j\subset W}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)}\frac{\mathbb{P}(\mathbf{e}_j=f_j)}{P_{m-1}(f_j)}\mathbb{P}(\mathbf{W}=W)\\ &= \left(1+O\left(\delta^{1/10^m}\right)\right)\sum_{i,j\in I_m}\sum_{e_i:v\in e_i}\sum_{f_j:v\in f_j}\frac{\mathbb{P}(\mathbf{e}_i=e_i)}{P_{m-1}(e_i)}\frac{\mathbb{P}(\mathbf{e}_j=f_j)}{P_{m-1}(f_j)}\frac{P_{m-1}(e\cup e_i\cup f_j)}{P_{m-1}(e)}\end{aligned}$$

We note that

$$Y(e_i,f_j) := \frac{P_{m-1}(v)^2P_{m-1}(e\cup e_i\cup f_j)}{P_{m-1}(e)P_{m-1}(e_i)P_{m-1}(f_j)}$$

is 1 when $e \setminus \{v\}$, $e_i \setminus \{v\}$ and $f_j \setminus \{v\}$ are pairwise disjoint and otherwise is at most κ^{-2r}. Hence

$$Y(e_i,f_j) = 1 + O\left(\kappa^{-2r}\left(\sum_{w\in e\setminus\{v\}}(1_{w\in e_i}+1_{w\in f_j}) + \sum_{w\in e_i\setminus\{v\}}1_{w\in f_j}\right)\right)$$

We note that

$$\sum_{i,j\in I_m}\sum_{w\in e\setminus\{v\}}\mathbb{P}(v,w\in\mathbf{e}_i, v\in\mathbf{e}_j) \le d_{I_m}(v)\sum_{w\in e\setminus\{v\}}\sum_{i\in I_m}\mathbb{P}(v,w\in\mathbf{e}_i)\le DA\delta$$

using (4.6) and (4.5), since $d_{I_m}(v) \le DP_{m-1}(v) \le D$. Similarly,

$$\sum_{i,j\in I_m}\sum_{w\in e\setminus\{v\}}\mathbb{P}(v\in\mathbf{e}_i, v,w\in\mathbf{e}_j)\le DA\delta$$

and

$$\sum_{i,j\in I_m}\sum_{e_i:v\in\mathbf{e}_i}\sum_{f_j:v\in f_j}\mathbb{P}(\mathbf{e}_i=e_i)\sum_{w\in e_i\setminus\{v\}}\mathbb{P}(v,w\in\mathbf{e}_j)\le rd_{I_m}(v)\delta\le Dr\delta$$

Using again the smallness condition on δ (4.8), we know that $k^{-2r}DA\delta \le \delta^{1/10^m}$, hence

$$\mathbb{E}(X(\mathbf{W})^2|e \subset \mathbf{W}) = \left(\frac{d_{I_m}(v)}{P_{m-1}(v)}\right)^2 + O\left(\delta^{1/10^m}\right) .$$

and the conclusion follows as an easy application of Lemma 3.1. ∎

Now we are ready to put everything together and complete the proof of Theorem 4.2. Recall from (4.30), that it is enough to prove

$$\mathbb{E}(Y(\mathbf{W})|e \subset \mathbf{W}) = \left(1 + O\left(\delta^{\frac{1}{9\times 10^m}}\right)\right) \exp\left(-\sum_{v\in e} \frac{d_{I_m}(v)}{P_{m-1}(v)}\right)$$

where $Y(W) = \mathbb{P}\left(e \subset W \setminus \bigcup_{i\in I_m} \mathbf{e}'_i|\mathbf{W} = W\right)$. From Lemma 4.5, we know that

$$Y(W) = \left(1 + O\left(\delta^{1/10^m}\right)\right) \exp\left(-\sum_{i\in I_m}\sum_{v\in e} \mathbb{P}(v \in \mathbf{e}'_i|\mathbf{W} = W)\right)$$

Next we use Lemma 4.6 and that $0 \le Y(W) \le 1$, for all W to obtain

$$\begin{aligned}
\mathbb{E}(Y(\mathbf{W})|e \subset \mathbf{W}) &= \left(1 + O\left(\delta^{1/10^m}\right)\right) \exp\left(-\sum_{v\in e}\left(\sum_{\substack{e_i: v\in e_i\\ e_i\subset W}} \frac{\mathbb{P}(\mathbf{e}_i = e_i)}{P_{m-1}(e_i)} + O\left(\delta^{\frac{1}{8\times 10^m}}\right)\right)\right) + O\left(\delta^{\frac{1}{8\times 10^m}}\right)\\
&= \left(1 + O\left(\delta^{1/10^m}\right)\right) \exp\left(-\sum_{v\in e}\left(\sum_{\substack{e_i: v\in e_i\\ e_i\subset W}} \frac{\mathbb{P}(\mathbf{e}_i = e_i)}{P_{m-1}(e_i)}\right) + O\left(\delta^{\frac{1}{9\times 10^m}}\right)\right) + O\left(\delta^{\frac{1}{8\times 10^m}}\right)\\
&= \left(1 + O\left(\delta^{\frac{1}{9\times 10^m}}\right)\right) \exp\left(-\sum_{v\in e}\sum_{\substack{e_i: v\in e_i\\ e_i\subset W}} \frac{\mathbb{P}(\mathbf{e}_i = e_i)}{P_{m-1}(e_i)}\right) + O\left(\delta^{\frac{1}{9\times 10^m}}\right)
\end{aligned}$$

where we used that $|e| \le A \le \delta^{-1/10^{m+2}}$. Finally, using Lemma 4.7, we obtain

$$\begin{aligned}
\mathbb{E}(Y(\mathbf{W})|e \subset \mathbf{W}) &= \left(1 + O\left(\delta^{\frac{1}{9\times 10^m}}\right)\right) \exp\left(-\sum_{v\in e}\left(\frac{d_{I_m}(v)}{P_{m-1}(v)} + O\left(\delta^{\frac{1}{3\times 10^m}}\right)\right)\right) + O\left(\delta^{\frac{1}{9\times 10^m}}\right)\\
&= \left(1 + O\left(\delta^{\frac{1}{9\times 10^m}}\right)\right) \exp\left(-\sum_{v\in e} \frac{d_{I_m}(v)}{P_{m-1}(v)}\right) + O\left(\delta^{\frac{1}{9\times 10^m}}\right) .
\end{aligned}$$

The conclusion follows from the fact that

$$\exp\left(-\sum_{v\in e} \frac{d_{I_m}(v)}{P_{m-1}(v)}\right) \ge \exp(-AD) \ge \delta^{\frac{1}{10^{m+2}}}$$

using once again (4.8) and (4.6). This concludes (4.30) and hence the proof of Theorem 4.2. ∎

5 Sieve estimates

5.1 Introduction

Consider the following problem. We fix an admissible k-tuple $(h_1, \dots, h_k)$ and we would like to show there exists infinitely many n such that many of $n + h_i$ are prime. We recall the key idea from [12]. We want to construct sieve weights $w(n)$ which are large when many of $n+h_i$ are prime and small otherwise. Then we look at sums of the form

$$S_1(x) = \sum_{x\le n\le 2x} w(n)$$

$$S_2(x) = \sum_{x\le n\le 2x}\sum_{i=1}^{k} 1(n + h_i \text{ prime})w(n)$$

If we can show that $S_2(x) \geq uS_1(x)$, then there exist some values of $n \in [x, 2x]$ such that at least u of $n + h_i$ are prime.

Theorem 5.1 (Maynard) *For $k \leq (\log x)^{1/5}$ and admissible k-tuple $(h_1, \dots, h_k)$ with $h_i \ll k^2$ for all i, there exist weights $w(n)$ such that $S_2(n) \gg \log k\ S_1(n)$.*

Recall from the discussion from the first section that we want to construct sets of the form $\{n + h_i p : 1 \leq i \leq k\}$ that contain many primes on average, for $p \in \mathscr{P}$. Hence in our case, we want to construct sieve weights $w(p, n)$ such that $w(p, n)$ is large when many of $n + h_i p$ are primes, for $p \in \mathscr{P}$. Indeed, we will deduce the following theorem:

Theorem 5.2 (Ford, Green, Konyagin, Maynard, Tao) *Let $k \leq (\log x)^{1/5}$ and $(h_1, \dots, h_k)$ an admissible k-tuple contained in $[2k^2]$. Then we can find weights $w : \mathscr{P} \times (\mathbb{Z} \cap [-y, y]) \to \mathbb{R}^+$ such that there exists a quantity t and $u \approx \log k$ with the following properties:*

(a) Uniformly for every $p \in \mathscr{P}$,

$$\sum_{|n| \leq y} w(p, n) \sim ty \tag{5.1}$$

(b) Uniformly for every $q \in \mathscr{Q}$,

$$\sum_{p \in \mathscr{P}} \sum_{i=1}^{k} w(p, q - h_i p) \sim t|\mathscr{P}|u \tag{5.2}$$

(c) Uniformly for every $p \in \mathscr{P}$ and $|n| \leq y$,

$$w(p, n) = O\left(x^{1/3}\right) \tag{5.3}$$

We note that (5.1) is similar to a sum of type S_1 and that (5.2) is similar to a sum of type S_2. We will see in Section 6 how this sieve weights $w(p, n)$ help us to construct sets of the form $\{n + h_i p : 1 \leq i \leq k\}$ with the required properties. The focus of this section will be showing how to prove Theorem 5.2.

We begin by recalling the classical Bombieri-Vinogradov theorem, which tells us that the average error term in the prime number theorem for arithmetic progressions is small.

Theorem 5.3 (Bombieri-Vinogradov) *Let $A > 0$ and $x^{1/2}(\log x)^{-A} \leq Q \leq x^{1/2}$. Then*

$$\sum_{q \leq Q} \max_{z \leq x} \max_{a : (a,q) = 1} \left| \pi(z; q, a) - \frac{\pi(z)}{\phi(q)} \right| \ll_A x^{1/2} Q (\log x)^5$$

We want to obtain a better upper bound. We need to take into consideration possible Siegel zeros. From the Landau-Page page theorem [4, Chapter 14], we know that there exists a constant c such that for $Q \geq 100$, there exists at most one primitive character χ of modulus at most Q with a zero in the region

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

Taking this into account, we have the following modified version of Bombieri-Vinogradov [4, Chapter 28]:

Theorem 5.4 *There exists a universal constant $c > 0$ and $B = B(x) \leq x$ which is either 1 or a prime such that*

$$\sum_{\substack{q \leq x^{1/3} \\ (q,B) = 1}} \max_{z \leq x} \max_{a : (a,q) = 1} \left| \pi(z; q, a) - \frac{\pi(z)}{\phi(q)} \right| \ll x \exp\left(-c\sqrt{\log x}\right)$$

5.2 Construction of sieve weights

Definition 5.5 *A linear form is a function $L : \mathbb{Z} \to \mathbb{Z}$ of the form $L(n) = an + b$, with integer coefficients a, b and $a \neq 0$.*
A finite set $\mathscr{L} = \{L_1, \dots L_k\}$ of linear form is admissible if $\prod_{i=1}^{k} L_i(n)$ has no fixed prime divisor, i.e. for every prime, there exists an integer n_p such that $\prod_{i=1}^{k} L_i(n_p)$ not divisible by p.

Fix $\mathscr{L}$ an admissible set of linear forms. We define the multiplicative functions $\omega = \omega_{\mathscr{L}}$ and $\varphi = \varphi_{\mathscr{L}}$ and the series $\mathfrak{S}_D(\mathscr{L})$, for an integer D by

$$\omega(p) = \begin{cases} \#\{1 \leq n \leq p : \prod_{i=1}^k L_i(n) \equiv 0 \mod p\}, & p \nmid B \\ 0, & p | B \end{cases} \tag{5.4}$$

$$\varphi(d) = \prod_{p|d}(p - \omega(p)) \tag{5.5}$$

$$\mathfrak{S}_D(\mathscr{L}) = \prod_{p \nmid D} \left(1 - \frac{\omega(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k} \tag{5.6}$$

Since $\mathscr{L}$ is admissible, we have $\omega(p) < p$, for all primes p, hence $\varphi(n) > 0$ and $\mathfrak{S}_D(\mathscr{L}) > 0$ for all integers D. Also, it is not hard to see that $\omega_p = k$ for all $p \nmid \prod_{i=1}^k a_i \prod_{i \neq j}(a_i b_j - b_i a_j)$. This implies that indeed the product in the definition of $\mathfrak{S}_D(\mathscr{L})$ is convergent. We denote $\mathfrak{S}_1(\mathscr{L}) =: \mathfrak{S}(\mathscr{L})$.

Since $\omega(p) \leq \min(k, p-1)$, we observe that

$$\begin{aligned} \mathfrak{S}_D(\mathscr{L}) &= \prod_{p \nmid D} \left(1 - \frac{\omega(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k} \geq \prod_{p \leq k, p \nmid D} \frac{1}{p} \prod_{p > k, p \nmid D} \left(1 - \frac{k}{p}\right)\left(1 - \frac{1}{p}\right)^{-k} \\ &\geq \prod_{p \leq k} \frac{1}{p} \prod_{p > k} \left(1 + O\left(\frac{k^2}{p^2}\right)\right) \geq \exp(-Ck) \end{aligned} \tag{5.7}$$

for some universal constant C. In the products above, we dropped the restriction $p \nmid D$ since all the terms are less than 1.

Let B as in Theorem 5.4 and $k = (\log x)^{1/5}$. Let $W := \prod_{p \leq 2k^2, p \nmid B} p$. Given a prime $p \nmid WB$, let $1 \leq r_{p,1}(\mathscr{L}) < \cdots < r_{p,\omega(p)}(\mathscr{L})$ be the $\omega(p)$ residue classes for each $\prod_{i=1}^k L_i(n)$ vanishes modulo p. For such a prime p, for each $1 \leq a \leq \omega(p)$, let $j_{p,a} = j_{p,a}(\mathscr{L})$ the smallest element of $[k]$ such that

$$L_{j_{p,a}}(r_{p,a}) \equiv 0 \pmod p .$$

For any $L \in \mathscr{L}$, there is at most one residue class for which L vanishes mod p. Thus, $j_{p,1}, \dots, j_{p,\omega(p)}$ must be distinct.

We construct sieve weights $w(n) = w_{\mathscr{L}}(n)$ such that $w(n) = 0$ if $(\prod_{i=1}^k L_i(n), W) \neq 1$. This condition guarantees that if $w(n) \neq 0$, then all $L_i(n)$ do not have small prime factors. Otherwise, we let

$$w(n) = \left(\sum_{d_i | L_i(n)} \lambda_{d_1,\dots,d_k}\right)^2 = \left(\sum_{d_i | L_i(n)} \lambda_{\mathbf{d}}\right)^2 \tag{5.8}$$

for some real coefficients $\lambda_{\mathbf{d}}$ depending on $\mathbf{d} = (d_1, \dots, d_k)$.

We want to restrict the support of $\lambda_{\mathbf{d}}$ to $\mathscr{D}_k(\mathscr{L})$, where

$$\begin{aligned} \mathscr{D}_k(\mathscr{L}) := \{\mathbf{d} \in \mathbb{N}^k \ : \ &\mu^2(d_1 \dots d_k) = 1; (d_1 \dots d_k, WB) = 1; \\ &(d_j, p) = 1 \text{ if } j \notin \{j_{p,1}, \dots, j_{p,\omega(p)}\}\} \end{aligned} \tag{5.9}$$

The reason we choose there restrictions is that we want different components $\mathbf{d}$ and $\mathbf{e}$ appearing in the sum (5.8) to be relatively coprime. Indeed, let $\mathbf{d}$ and $\mathbf{e}$ both appearing an the sum and $p|(d_i, e_j)$, for some $i \neq j$. Then $p|L_i(n)$, hence i must be the chosen index for the residue class $n \pmod p$. But similarly j must be the chosen index for the same residue class, which is a contradiction. Hence a prime will divide only elements with the same index appearing in the sum.

We define

$$\lambda_{d_1,\dots,d_k} := \mu(d_1 \dots d_k) d_1 \dots d_k \mathfrak{S}_{WB}(\mathscr{L}) \sum_{\substack{\mathbf{r} \in \mathscr{D}_k(\mathscr{L}) \\ d_i | r_i, \text{ for all } i}} \frac{1}{\varphi(r_1 \dots r_k)} F\left(\frac{\log r_1}{\log R}, \dots, \frac{\log r_k}{\log R}\right) \tag{5.10}$$

Here, $F:\mathbb{R}^k \to \mathbb{R}$ will be a smooth function supported on $\{(t_1,\dots,t_k)\in\mathbb{R}_+^k \ : \ t_1+\dots+t_k\le 1\}$. This condition makes sure that the terms in the sum in (5.10) are non-zero only if $r_1\dots r_k \le R$. There is some liberty in the choice of R, but for our purposes we take $R = x^{1/9}$.

We use the following key theorem from [13] about the constructions of weights $w(n)$:

Theorem 5.6 *There exists a universal constant C such that the following holds. Let $\mathscr{L}=\{L_1,\dots,L_k\}$ be an admissible set of linear forms such that $C\le k\le \log^{1/5} x$. Assume that the coefficients a_i, b_i of the linear forms $L_i(n)=a_in+b_i$ satisfy the bounds $|a_i|,|b_i|\le x^2$. for all $i\in[k]$. Then there exists a smooth function $F:\mathbb{R}^k\to\mathbb{R}$ and quantities I_k, J_k depending only on k satisfying*

$$I_k \gg (2k\log k)^{-k} \quad \text{and} \quad J_k \asymp \frac{\log k}{k} I_k \tag{5.11}$$

such that for $w(n)$ described above, the following statements hold uniformly for $x^{1/30}\le R\le x^{1/9}$:

- *For all $n\in\mathbb{Z}$, we have*
$$w(n) \ll x^{1/3} \tag{5.12}$$
-
$$\sum_{x\le n\le 2x} w(n) = \left(1+O\left(\frac{1}{\log^{1/10} x}\right)\right) x\frac{B^k}{\phi(B)^k}\mathfrak{S}(\mathscr{L})(\log R)^k I_k \tag{5.13}$$
- *Let $L(n)=n$. If $L\in\mathscr{L}$, we have*
$$\sum_{\substack{x\le n\le 2x\\ n \text{ prime}}} w(n) = \left(1+O\left(\frac{1}{\log^{1/10} x}\right)\right)\mathfrak{S}(\mathscr{L})\frac{x}{\log x}(\log R)^{k+1}J_k\frac{B^{k-1}}{\phi(B)^{k-1}}$$
$$+O\left(\frac{B^k}{\phi(B)^k}\mathfrak{S}(\mathscr{L})x(\log R)^{k-1}I_k\right) \tag{5.14}$$

Let $h_1,\dots,h_k$ be the first k primes larger than k, i.e. $h_i = p_{\pi(k)+i}$. In particular, we have that $h_i\le 2k^2$, for k large enough. For each $i\in[k]$, let $L_i(n)=n+h_ip$. For all $p\in\mathscr{P}$, we define $\mathscr{L}_p=\{L_1,\dots,L_k\}$. It is easy to see that $\mathscr{L}_p$ is an admissible set of linear forms.

We then have that
$$\omega_p(q) := \omega_{\mathscr{L}_p}(q) = \#\{h_i \pmod q\} \ \text{ if } q\ne p$$
so ω_p are all very close to being equal. It makes to define $\omega(q)=\#\{h_i \pmod q\}$, for all primes q. Then

$$\mathfrak{S}_D := \prod_{q\nmid D}\left(1-\frac{\omega(q)}{q}\right)\left(1-\frac{1}{q}\right)^{-k} = \left(1+O\left(\frac{k}{x}\right)\right)\mathfrak{S}_D(\mathscr{L}_p) \tag{5.15}$$

for all integers D, since $\frac{x}{2}\le p\le\frac{x}{4}$ and $\omega(q),\omega_p(q)\le k$, for all primes q and $p\in\mathscr{P}$.

We observe that for $q>2k^2$, $q\nmid B$, all h_i are distinct mod q (since $h_i<2k^2$, for all i). So if $q\ne p$, we have that $\omega_p(q)=k$ and $\{j_{q,1}(\mathscr{L}_p),\dots,j_{q,\omega_p(q)}(\mathscr{L}_p)\}=\{1,2,\dots,k\}$.

Also, when $q\le R$, we have that $q\ne p$, for all $p\in\mathscr{P}$ since $R\le x/4$. Hence,

$$\mathscr{D}_k(\mathscr{L}_p)\cap\{(d_1,\dots,d_k):\prod_{i=1}^k d_i\le R\} = \{\mathbf{d}\in\mathbb{N}^k : \mu^2(d)=1,(d,WB)=1,d\le R\}$$

is independent of p. Hence, using (5.10), we get

$$\lambda_{\mathbf{d}}(\mathscr{L}_p) = \frac{\mathfrak{S}_D(\mathscr{L}_p)}{\mathfrak{S}_D}\lambda_{\mathbf{d}} = \left(1+O\left(\frac{k}{x}\right)\right)\lambda_{\mathbf{d}}$$

for some $\lambda_{\mathbf{d}}$ independent of p, for all integers D.

We define
$$w(p,n) := 1_{[-y,y]}(n)w_{\mathscr{L}_p}(n) \tag{5.16}$$

We are now ready to apply Theorem 5.6. Recall that we want to estimate $\sum_{n\in\mathbb{Z}} w(p,n)$. For this purpose, we define the set $\mathscr{L}'_p$ composed of linear forms $L_i(n) = n + h_i p - 3\lfloor y \rfloor$, for $i \in [k]$. Clearly $\mathscr{L}'_p$ is admissible, $\omega_{\mathscr{L}'_p} = \omega_{\mathscr{L}_p}$ and $\mathfrak{S}_D(\mathscr{L}'_p) = \mathfrak{S}_D(\mathscr{L}_p)$. Also,

$$\sum_{n\in\mathbb{Z}} w(p,n) = \sum_{-y\le n\le y} w_{\mathscr{L}_p}(n) = \sum_{y\le n\le 2y} w_{\mathscr{L}'_p}(n)$$

From (5.13) and (5.15), we see that

$$\sum_{n\in\mathbb{Z}} w(p,n) = \left(1 + O\left(\frac{1}{\log^{1/10} x}\right)\right) y\, \mathfrak{S}\frac{B^k}{\phi(B)^k}(\log R)^k I_k \tag{5.17}$$

Let

$$t := \mathfrak{S}\frac{B^k}{\phi(B)^k}(\log R)^k I_k \tag{5.18}$$

Using (5.7), (5.11), we see that $t \ge 1$. Hence

$$\sum_{n\in\mathbb{Z}} w(p,n) = \left(1 + O\left(\frac{1}{\log^{1/10} x}\right)\right) ty \tag{5.19}$$

which gives us (5.1).

Fix $i \in [k]$ and $q \in \mathscr{Q}$. We want to approximate $\sum_{p\in\mathscr{P}} w(p, q - h_i p)$. For this purpose, we define the set of linear forms $\mathscr{L}_q^{(i)} = \{L_1, L_2, \ldots, L_k\}$, where $L_i(n) = n$ and $L_j(n) = q + (h_j - h_i)n$, for $i \ne j$.

We check that $\mathscr{L}_q^{(i)}$ is admissible. Indeed, for any prime $s \ne q$, the number of solutions mod s to

$$n \prod_{i\ne j}(n + (h_j - h_i)q) \equiv 0 \pmod{s}$$

is $\#\{h_j \pmod{s} \;:\; j \in [k]\} < s$ since $(h_1, \ldots, h_k)$ admissible. Hence, similar as before we obtain that

$$\mathfrak{S}_D(\mathscr{L}_q^{(i)}) = \left(1 + O\left(\frac{k}{x}\right)\right)\mathfrak{S}_D \tag{5.20}$$

for all integers D and that $\mathscr{D}_k(\mathscr{L}_q^{(i)}) \cap \{(d_1, \ldots, d_k) : d_1 \ldots d_k \le R\}$ is independent of q and i, so once again

$$\lambda_{\mathbf{d}}(\mathscr{L}_q^{(i)}) = \left(1 + O\left(\frac{k}{x}\right)\right)\lambda_{\mathbf{d}}\ .$$

Now, since $q - h_i p \in [-y, y]$, for all $q \in \mathscr{Q}$ and $i \in [k]$, then

$$w_{\mathscr{L}_q^{(i)}}(p) = \left(1 + O\left(\frac{k}{x}\right)\right) w_{\mathscr{L}_p}(q - h_i p) = \left(1 + O\left(\frac{k}{x}\right)\right) w(p, q - h_i p)\ .$$

Hence

$$\sum_{p\in\mathscr{P}} w(p, q - h_i p) = \left(1 + O\left(\frac{k}{x}\right)\right)\sum_{p\in\mathscr{P}} w_{\mathscr{L}_q^{(i)}}(p)$$

We apply (5.14) to get

$$\sum_{p\in\mathscr{P}} w(p, q - h_i p) = \left(1 + O\left(\frac{1}{\log^{1/10} x}\right)\right)\mathfrak{S}\frac{x}{4\log x}(\log R)^{k+1} J_k \frac{B^{k-1}}{\phi(B)^{k-1}} + O\left(x\frac{B^k}{\phi(B)^k}\mathfrak{S}(\log R)^{k-1} I_k\right)$$

where we used (5.20). From our choice of R and (5.11), we see that the second error term can be absorbed into the first one.

Let

$$u := \frac{\phi(B)}{B}\frac{\log R}{\log x}\frac{kJ_k}{I_k}$$

Then we have that

$$\sum_{p\in\mathscr{P}} w(p, q - h_i p) = \left(1 + O\left(\frac{1}{\log^{1/10} x}\right)\right)\frac{tu}{k}\frac{x}{4} \tag{5.21}$$

6 Proof of large gaps

6.1 Random uniform choice

Recall that $\mathscr{S} = \{p \text{ prime } : \log^{10} x < p \leq z\}$, where $z = x^{\log_3 x / 4 \log_2 x}$.

We have the random vector $\vec{\mathbf{a}} := (\mathbf{a}_s \mod s)_{s \in \mathscr{S}}$ where each residue class $\mathbf{a}_s$ is selected uniformly at random independently in s. We define the random set

$$S(\vec{\mathbf{a}}) := \{n \in \mathbb{Z} : n \not\equiv \mathbf{a}_s \pmod{s}, \text{ for all } s \in \mathscr{S}\} \tag{6.1}$$

which is a subset of $\mathbb{Z}$ with density

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

Using Mertens' Theorem, we deduce that

$$\sigma := \prod_{\log^{10} x < p \leq z} \left(1 - \frac{1}{p}\right) \sim \frac{10 \log_2 x}{\log z} = 40 \frac{\log_2^2 x}{\log x \log_3 x} \tag{6.2}$$

Lemma 6.1 *Let $t \leq \log x$ and $n_1, n_2, \ldots, n_t$ distinct integers such that $|n_i| \leq x^2$, for all i. Then*

$$\mathbb{P}(n_1, n_2, \ldots, n_t \in S(\vec{\mathbf{a}}) = \left(1 + o\left(\frac{1}{\log^6 x}\right)\right) \sigma^t$$

Proof We note that if $n_1, n_2, \ldots, n_t$ are not distinct modulo some prime $s \in \mathscr{S}$, then s divides one of $n_i - n_j$, for some $1 \leq i < j \leq t$. But for each $1 \leq i < j \leq t$, we have that $|n_i - n_j| \leq 2x^2$, so it can be divisible by at most $O\left(\frac{\log x}{\log_2 x}\right)$ primes which are at least $\log^{10} x$. Hence there are $o(t^2 \log x)$ possible values of s such that $n_1, \ldots, n_t$ are not distinct residue classes modulo s. Hence the probability that $\mathbf{a}_s$ avoids all $n_1, \ldots, n_t$ is $1 - \frac{t}{s}$ except for $o(\log^3 x)$ values of s, where it is

$$\left(1 - \frac{t}{s}\right)\left(1 + O\left(\frac{t}{s}\right)\right) = \left(1 - \frac{t}{s}\right)\left(1 + O\left(\frac{1}{\log^9 x}\right)\right)$$

Hence

$$\begin{aligned}
\mathbb{P}(n_1, n_2, \ldots, n_t \in S(\vec{\mathbf{a}})) &= \prod_{s \in \mathscr{S}} \left(1 - \frac{t}{s}\right) \left(1 + O\left(\frac{1}{\log^9 x}\right)\right)^{o(\log^3 x)} \\
&= \left(1 + o\left(\frac{1}{\log^6 x}\right)\right) \sigma^t \prod_{s \in \mathscr{S}} \left(1 + O\left(\frac{t^2}{s^2}\right)\right) \\
&= \sigma^t \left(1 + o\left(\frac{1}{\log^6 x}\right)\right)
\end{aligned}$$

Lemma 6.2 *With probability $1 - o(1)$, we have that*

$$|\mathscr{Q} \cap S(\vec{\mathbf{a}})| \sim \sigma |\mathscr{Q}| \sim 40c \frac{x \log_2 x}{\log x}$$

More precisely, with probability $1 - o(1)$,

$$|\mathscr{Q} \cap S(\vec{\mathbf{a}})| = \left(1 + O\left(\frac{1}{\log_2 x}\right)\right) 40c \frac{x \log_2 x}{\log x}$$

Proof Let

$$X(\vec{\mathbf{a}}) = \#(\mathscr{Q} \cap S(\vec{\mathbf{a}})) = \sum_{q \in \mathscr{Q}} 1_{q \in S(\vec{\mathbf{a}})} .$$

Then, using Lemma 6.1

$$\mathbb{E} X(\vec{\mathbf{a}}) = \left(1 + O\left(\frac{1}{\log^6 x}\right)\right) \sigma |\mathscr{Q}|$$

and

$$\mathbb{E}X(\vec{\mathbf{a}})^2 = \left(1 + O\left(\frac{1}{\log^6 x}\right)\right)(\sigma|\mathscr{Q}| + \sigma^2|\mathscr{Q}|(|\mathscr{Q}|-1)) = \left(1 + O\left(\frac{1}{\log^6 x}\right)\right)\sigma^2|\mathscr{Q}|^2$$

since $|\mathscr{Q}| \sim \frac{y}{\log x}$ by the prime number theorem and using our choice of y (2.1). Then the conclusion follows from Lemma 3.1.

6.2 Weighted choice

Our goal is to construct random variables $\mathbf{n}_p$ such that $\{\mathbf{n}_p + h_i p : i \in [k]\} \cap \mathscr{Q}(\vec{\mathbf{a}})$ large on average. First, we will apply Theorem 5.2 to guarantee that sets of the form $\{n + h_i p : i \in [k]\}$ contain many primes.

For each $p \in \mathscr{P}$,we select a random number $\mathbf{m}_p \in [0, y]$ with probability proportional to $w(p, n)$:

$$\mathbb{P}(\mathbf{m}_p = n) := \frac{w(p,n)}{\sum_{z\in\mathbb{Z}} w(p,z)} \tag{6.3}$$

Informally, the probability is big when many of $n + h_i p$ are prime.

First, we notice from (5.12) and (5.19) that for all $p \in \mathscr{P}$ and $n \in \mathbb{Z}$

$$\mathbb{P}(\mathbf{m}_p = n) \ll x^{-2/3} \tag{6.4}$$

Using (5.19) and (5.21), we note that for all $q \in \mathscr{Q}$ and $i \in [k]$:

$$\sum_{p\in\mathscr{P}} \mathbb{P}(\mathbf{m}_p = q - h_i p) = \sum_{p\in\mathscr{P}} w(p, q - h_i p)\frac{1}{\sum_{n\in\mathbb{Z}} w(p,n)} = \left(1 + O\left(\frac{1}{\log^{1/10} x}\right)\right)\frac{u}{k}\frac{x}{4y} \tag{6.5}$$

where $u \approx \log_2 x$.

For each $p \in \mathscr{P}$, we define $X_p(\vec{a})$ by

$$X_p(\vec{a}) = \mathbb{P}(\mathbf{m}_p + h_i p \in S(\vec{a}),\ \text{for all } i \in [k])$$

In light of Lemma 6.1, we expect that $X_p(\vec{a}) \sim \sigma^k$. Indeed, denote $\mathscr{P}(\vec{a})$ the set of primes $p \in \mathscr{P}$ such that

$$X_p(\vec{a}) = \left(1 + O_{\le}\left(\frac{1}{\log^2 x}\right)\right)\sigma^k$$

We will see in Lemma 6.3 that $|\mathscr{P}(\vec{a})| \sim |\mathscr{P}|$ with probability $1 - o(1)$.

We define

$$Z_p(\vec{a}, n) = \begin{cases} 1 & \text{if } n + h_j p \in S(\vec{a}),\ \text{for all } j \in [k] \\ 0 & \text{otherwise.} \end{cases}$$

We want to define random variable $\mathbf{n}_p$ proportional to $\mathbf{m}_p$ and conditional on $\vec{\mathbf{a}}$ such that we can assure $n + h_1 p, \ldots, n + h_k p \in \mathscr{S}(\vec{a})$ if $\mathbb{P}(\mathbf{n}_p = n | \vec{\mathbf{a}} = \vec{a}) \neq 0$. If $p \in \mathscr{P}(\vec{a})$, we let

$$\mathbb{P}(\mathbf{n}_p = n | \vec{\mathbf{a}} = \vec{a}) = \frac{Z_p(\vec{a}, n)}{X_p(\vec{a})}\mathbb{P}(\mathbf{m}_p = n)$$

Otherwise, if $p \in \mathscr{P} \setminus \mathscr{P}(\vec{a})$, we let $\mathbf{n}_p = 0$. From our choice of $X_p(\vec{a})$, we see that this is a well defined random variable. This is a very similar construction to the choices of $\mathbf{e}'_i$ in the proof of the Theorem 4.2.

First, we prove indeed that $|\mathscr{P}(\vec{\mathbf{a}})| \sim |\mathscr{P}|$ with high probability:

Lemma 6.3 $|\mathscr{P}(\vec{\mathbf{a}})| \ge \left(1 - \frac{1}{\log x}\right)|\mathscr{P}|$ *with probability* $1 - O\left(\frac{1}{\log x}\right)$.

Proof First, we observe that

$$\begin{aligned}\mathbb{E}X_p(\vec{\mathbf{a}}) &= \sum_{\vec{a}}\sum_n Z_p(\vec{a},n)\mathbb{P}(\mathbf{m}_p=n)\mathbb{P}(\vec{\mathbf{a}}=\vec{a})\\ &= \sum_n \mathbb{P}(\mathbf{m}_p=n)\mathbb{P}(n+h_1p,\dots,n+h_kp\in S(\vec{\mathbf{a}}))\\ &= \left(1+O\left(\frac{1}{\log^6 x}\right)\right)\sigma^k\end{aligned}$$

using Lemma 6.1. Secondly, we see that

$$\begin{aligned}\mathbb{E}X_p(\vec{\mathbf{a}})^2 &= \sum_{\vec{a}}\left(\sum_{n_1,n_2} Z_p(\vec{a},n_1)Z_p(\vec{a},n_2)\mathbb{P}(\mathbf{m}_p=n_1)\mathbb{P}(\mathbf{m}_p=n_2)\right)\mathbb{P}(\vec{\mathbf{a}}=\vec{a})\\ &= \sum_{n_1,n_2}\mathbb{P}(\mathbf{m}_p=n_1)\mathbb{P}(\mathbf{m}_p=n_2)\mathbb{P}(n_1+h_1p,\dots,n_1+h_kp,n_2+h_1p,\dots,n_2+h_kp\in S(\vec{\mathbf{a}}))\\ &= \left(1+O\left(\frac{1}{\log^6 x}\right)\right)\mathbb{E}\sigma^{\#\{\mathbf{m}_p^{(l)}+h_ip\ :i\in[k],l=1,2\}}\end{aligned}$$

From (6.4) and from $k\le\log x$, we know that $\#\{\mathbf{m}_p^{(l)}+h_ip\ \ : i\in[k],l=1,2\}$ is $2k$ with probability $1-O\left(x^{-2/3+o(1)}\right)$. Hence

$$\mathbb{E}X_p(\vec{\mathbf{a}})^2 = \left(1+O\left(\frac{1}{\log^6 x}\right)\right)\sigma^{2r}\left(1+O\left(x^{-1/2}\right)\right) = \left(1+O\left(\frac{1}{\log^6 x}\right)\right)\sigma^{2r}$$

Now we are ready to apply Lemma 3.1. We get that

$$\mathbb{P}\left(|X_p(\vec{\mathbf{a}})-\sigma^k|\ge\frac{\sigma^k}{\log^2 x}\right) = O\left(\frac{1}{\log^2 x}\right)$$

This means that for each prime $p\in\mathscr{P}$, we have that $\mathbb{P}(p\in\mathscr{P}(\vec{\mathbf{a}}))=1-O\left(\frac{1}{\log^2 x}\right)$.

Let

$$Y_p(\vec{\mathbf{a}}) := \begin{cases}1 & \text{if } p\in\mathscr{P}(\vec{\mathbf{a}});\\ 0 & \text{otherwise.}\end{cases}$$

Then clearly $\#\mathscr{P}(\vec{\mathbf{a}})=\sum_{p\in\mathscr{P}}Y_p(\vec{\mathbf{a}})$. Using Markov's inequality and linearity of expectation, we get

$$\mathbb{P}\left(|\mathscr{P}\setminus\mathscr{P}(\vec{\mathbf{a}})|\ge\frac{|\mathscr{P}|}{\log x}\right)\le\frac{\sum_{p\in\mathscr{P}}\mathbb{P}(p\notin\mathscr{P}(\vec{\mathbf{a}}))}{|\mathscr{P}|/\log x} = O\left(\frac{1}{\log x}\right)$$

which is what we wanted to prove. ∎

Let

$$C := \frac{ux}{4\sigma y} \tag{6.6}$$

From (2.1), (6.2) and that $u\approx\log_2 x$, we deduce that $C\approx\frac{1}{c}$, where c is the small constant in the definition of y.

Recall that for fixed $\vec{a}$, we let

$$\mathbf{e}_p(\vec{a})=\{\mathbf{n}_p+h_ip : i=1,\dots,k\}$$

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

$$\sum_{p\in\mathscr{P}}\mathbb{P}(q\in\mathbf{e}_p(\vec{\mathbf{a}})) = C+O_{\le}\left(\frac{1}{\log_2^2 x}\right)$$

for all but at most $\frac{x}{\log x\log_2 x}$ *primes in* $\mathscr{Q}\cap\mathscr{S}(\vec{\mathbf{a}})$.

Proof We notice that

$$\sum_{p\in\mathscr{P}} \mathbb{P}(q\in \mathbf{e}_p(\vec{a})|\vec{\mathbf{a}}=\vec{a}) = \sum_{i=1}^{k}\sum_{p\in\mathscr{P}(\vec{a})} \mathbb{P}(\mathbf{n}_p = q-h_ip|\vec{\mathbf{a}}=\vec{a})$$
$$= \left(1+O\left(\frac{1}{\log^2 x}\right)\right)\sigma^{-k}\sum_{i=1}^{k}\sum_{p\in\mathscr{P}(\vec{a})} Z_p(\vec{a},q-h_ip)\mathbb{P}(\mathbf{m}_p=q-h_ip)$$

since $\mathbb{P}(q\in\mathbf{e}_p(\vec{a}))=0$ if $p\notin\mathscr{P}(\vec{a})$. Hence it will suffice to show that

$$\sigma^{-k}\sum_{i=1}^{k}\sum_{p\in\mathscr{P}(\vec{\mathbf{a}})} Z_p(\vec{\mathbf{a}},q-h_ip)\mathbb{P}(\mathbf{m}_p=q-h_ip) = \left(1+O\left(\frac{1}{\log_2^3 x}\right)\right)C \tag{6.7}$$

with probability $1-o(1)$ for all but $\frac{x}{\log x\log_2 x}$ primes $q\in\mathscr{Q}$.

We would like to be able to replace the inner sum in (6.7) to $p\in\mathscr{P}$. Clearly,

$$\sum_{q\in\mathscr{Q}}\sum_{i=1}^{k}\sum_{p\in\mathscr{P}\setminus\mathscr{P}(\vec{a})} Z_p(\vec{a},q-h_ip)\mathbb{P}(\mathbf{m}_p=q-h_ip) \le k\sum_{n\in\mathbb{Z}}\sum_{p\in\mathscr{P}\setminus\mathscr{P}(\vec{a})} Z_p(\vec{a},n)\mathbb{P}(\mathbf{m}_p=n)$$

We note that

$$\mathbb{E}\left(\sum_{n}\sum_{p\in\mathscr{P}} Z_p(\vec{\mathbf{a}},n)\mathbb{P}(\mathbf{m}_p=n)\right) = \sum_{p\in\mathscr{P}}\sum_{n}\mathbb{P}(\mathbf{m}_p=n)\mathbb{P}(n+h_jp\in S(\vec{\mathbf{a}}),\ \text{for all } j\in[k])$$
$$= \left(1+O\left(\frac{1}{\log^6 x}\right)\right)\sigma^k|\mathscr{P}| \tag{6.8}$$

and that

$$\mathbb{E}\left(\sum_{n}\sum_{p\in\mathscr{P}(\vec{\mathbf{a}})} Z_p(\vec{\mathbf{a}},n)\mathbb{P}(\mathbf{m}_p=n)\right) = \sum_{\vec{a}}\mathbb{P}(\vec{\mathbf{a}}=\vec{a})\sum_{p\in\mathscr{P}(\vec{a})} X_p(\vec{a})$$
$$= \left(1+O\left(\frac{1}{\log x}\right)\right)\sigma^k|\mathscr{P}| \tag{6.9}$$

where we have used Lemma 6.3. Subtracting (6.9) from (6.8) we obtain

$$\sum_{q\in\mathscr{Q}}\mathbb{E}\left(\sigma^{-k}\sum_{i=1}^{k}\sum_{p\in\mathscr{P}\setminus\mathscr{P}(\vec{\mathbf{a}})} Z_p(\vec{a},q-h_ip)\mathbb{P}(\mathbf{m}_p=q-h_ip)\right) = O\left(\frac{|\mathscr{P}|}{k\log x}\right)$$

From Markov's inequality, it follows that

$$\sum_{q\in\mathscr{Q}}\mathbb{P}\left(\sigma^{-k}\sum_{i=1}^{k}\sum_{p\in\mathscr{P}\setminus\mathscr{P}(\vec{\mathbf{a}})} Z_p(\vec{a},q-h_ip)\mathbb{P}(\mathbf{m}_p=q-h_ip)\ge\frac{1}{\log_2^3 x}\right) = O\left(\frac{|\mathscr{P}|\log_2^3}{k\log x}\right) = O\left(\frac{x}{\log x\log_2^3 x}\right)$$

Hence there are at most $\frac{x}{\log x\log_2^2 x}$ primes $q\in\mathscr{Q}$ for which

$$\mathbb{P}\left(\sigma^{-k}\sum_{i=1}^{k}\sum_{p\in\mathscr{P}\setminus\mathscr{P}(\vec{\mathbf{a}})} Z_p(\vec{a},q-h_ip)\mathbb{P}(\mathbf{m}_p=q-h_ip)\ge\frac{1}{\log_2^3 x}\right)\ge\frac{1}{\log_2 x}\ .$$

Therefore, it is enough to show that, with probability $1-o(1)$,

$$\sigma^{-k}\sum_{i=1}^{k}\sum_{p\in\mathscr{P}} Z_p(\vec{\mathbf{a}},q-h_ip)\mathbb{P}(\mathbf{m}_p=q-h_ip) = \left(1+O\left(\frac{1}{\log_2^3 x}\right)\right)C \tag{6.10}$$

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

For $q \in \mathscr{Q}$, let $Y_q(\vec{a}) = \sum_{i=1}^{k} \sum_{p \in \mathscr{P}} Z_p(\vec{a}, q - h_i p) \mathbb{P}(\mathbf{m}_p = q - h_i p)$. Clearly $Y_q(\vec{a}) = 0$ if $q \notin \mathscr{Q} \cap \mathscr{S}(\vec{a})$. From (6.5), we have that

$$\begin{aligned}\mathbb{E}Y_q(\vec{\mathbf{a}}) &= \sum_{i=1}^{k} \sum_{p \in \mathscr{P}} \mathbb{P}\left(q + (h_j - h_i)p \in \mathscr{S}(\vec{\mathbf{a}}) \text{ for all } j \in [k]\right) \mathbb{P}(\mathbf{m}_p = q - h_i p) \\ &= \left(1 + O\left(\frac{1}{\log_2^{20} x}\right)\right) \sigma^k u \frac{x}{4y}\end{aligned}$$

and

$$\mathbb{E}Y_q(\vec{\mathbf{a}})^2 = \sum_{p_1, p_2 \in \mathscr{P}} \sum_{i_1, i_2} \mathbb{P}(q + (h_j - h_{i_l} p_l \in \mathscr{S}(\vec{\mathbf{a}}) \text{ for } j \in [k], l = 1, 2) \mathbb{P}(\mathbf{m}_{p_1} = q - h_{i_1} p_1) \mathbb{P}(\mathbf{m}_{p_2} = q - h_{i_2} p_2)$$

We note that $\#\{q + (h_j - h_{i_l} p_l) : j \in [k], l = 1, 2\} = 2k - 1$ unless $p_1 = p_2$, since $h_i < x/4$, for all i. We use again (6.4) to see that the terms with $p_1 = p_2$ have negligible contribution. Hence we obtain

$$\mathbb{E}Y_q(\vec{\mathbf{a}})^2 = \left(1 + O\left(\frac{1}{\log_2^{20} x}\right)\right) \sigma^{2k-1} \left(u \frac{x}{4y}\right)^2$$

We know from Lemma 6.1 that $\mathbb{P}(q \in S(\vec{\mathbf{a}}) = \sigma \left(1 + O\left(\frac{1}{\log^6 x}\right)\right)$, hence

$$\mathbb{E}(Y_q(\vec{\mathbf{a}}) | q \in S(\vec{\mathbf{a}})) = \left(1 + O\left(\frac{1}{\log_2^{20} x}\right)\right) \sigma^{k-1} u \frac{x}{4y}$$

and

$$\mathbb{E}(Y_q(\vec{\mathbf{a}})^2 | q \in S(\vec{\mathbf{a}})) = \left(1 + O\left(\frac{1}{\log_2^{20} x}\right)\right) \left(\sigma^{k-1} u \frac{x}{4y}\right)^2 .$$

Therefore, using Lemma 3.1, we get that

$$\mathbb{P}\left(\left|\sigma^{-k} Y_q(\vec{\mathbf{a}}) - \frac{u}{\sigma} \frac{x}{4y}\right| \geq \frac{1}{\log_2^6 x} \frac{u}{\sigma} \frac{x}{4y} \mid q \in S(\vec{\mathbf{a}})\right) \leq \frac{1}{\log_2^6 x}$$

Hence

$$\mathbb{E}\left(\sum_{q \in \mathscr{Q} \cap S(\vec{\mathbf{a}})} Y_q(\vec{\mathbf{a}})\right) = \mathbb{E}\left(\sum_{q \in \mathscr{Q}} Y_q(\vec{\mathbf{a}})\right) = \left(1 + O\left(\frac{1}{\log_2^{20} x}\right)\right) \sigma^k u \frac{x}{4y} |\mathscr{Q}|$$

and

$$\mathbb{E}\left(\sum_{q \in \mathscr{Q} \cap S(\vec{\mathbf{a}})} Y_q(\vec{\mathbf{a}})^2\right) = \left(1 + O\left(\frac{1}{\log_2^{20} x}\right)\right) \sigma^{2k-1} \left(u \frac{x}{4y}\right)^2 |\mathscr{Q}|$$

We obtain that

$$\mathbb{P}\left(\sum_{q \in \mathscr{Q} \cap S(\vec{\mathbf{a}})} \left|Y_q(\vec{\mathbf{a}}) - \sigma^{k-1} u \frac{x}{4y}\right| \geq \frac{1}{\log_2^6 x} |\mathscr{Q}| \sigma^k u \frac{x}{4y}\right) \leq \frac{1}{\log_2^6 x}$$

In conclusion, with probability $1 - \frac{1}{\log_2^6 x}$, the number of $q \in \mathscr{Q} \cap S(\vec{\mathbf{a}})$ such that $\sigma^{-k} Y_q(\vec{\mathbf{a}}) \geq \frac{1}{\log_2^3 x} \frac{u}{\sigma} \frac{x}{4y}$ is at most

$$O\left(\frac{\sigma |\mathscr{Q}|}{\log_2^3 x}\right) = O\left(\frac{x}{\log x \log_2^2 x}\right)$$

using (6.2) and (2.1). ∎

6.3 Applying the covering lemma

Let's summarise what we have achieved so far. For each $\vec{a}$ in the range of $\vec{\mathbf{a}}$, we have constructed random integers $\mathbf{n}_p$ and random subsets $\mathbf{e}_p(\vec{a}) = \{\mathbf{n}_p + h_i p : i \in [k]\} \cap \mathscr{Q}(\vec{a})$. From Lemma 6.4, we know that that with probability $1 - o(1)$ in $\vec{\mathbf{a}}$,

$$\sum_{p \in \mathscr{P}} \mathbb{P}(q \in \mathbf{e}_p(\vec{\mathbf{a}})) = C + O_{\leq}\left(\frac{1}{\log_2^2 x}\right) \tag{6.11}$$

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

From Lemma 6.2, we know that with probability $1 - o(1)$, we have that

$$|\mathscr{Q}(\vec{\mathbf{a}})| = |\mathscr{Q} \cap S(\vec{\mathbf{a}})| = \left(1 + O\left(\frac{1}{\log_2 x}\right)\right) 40c \frac{x \log_2 x}{\log x} \tag{6.12}$$

From now on, fix $\vec{a}$ in the range of $\vec{\mathbf{a}}$ such that (6.11) and (6.12) both hold.

For all $q \in \mathscr{Q}$, $p \in \mathscr{P}$:

$$\mathbb{P}(q \in \mathbf{e}_p(\vec{a}) | \vec{\mathbf{a}} = \vec{a}) = \sum_{i=1}^{k} \mathbb{P}(\mathbf{n}_p = q - h_i p) \ll \sigma^{-k} \sum_{i=1}^{k} \mathbb{P}(\mathbf{m}_p = q - h_i p) \leq x^{-3/5} \tag{6.13}$$

for x large enough, where we used (6.4). Also, for distinct integers $q_1, q_2 \in \mathscr{Q}$, if $q_1, q_2 \in \mathbf{e}_p(\vec{a})$, then $p | q_1 - q_2$. But $q_1 - q_2 \leq x \log x$, so the difference is divisible by at most on prime $p_0 \in \mathscr{P}$. Hence

$$\sum_{p \in \mathscr{P}} \mathbb{P}(q_1, q_2 \in \mathbf{e}_p(\vec{a})) = \mathbb{P}(q_1, q_2 \in e_{p_0}(\vec{a})) \leq x^{-3/5} \ . \tag{6.14}$$

Note that we satisfy all the conditions of Theorem 2.4 stated in the first section. Now we are ready to apply our hypergraph covering Lemma 4.3. Set $V = \mathscr{Q}(\vec{a})$, $I = \mathscr{P}$, $\mathbf{e}_i = \mathbf{e}_p(\vec{a})$, $r(x) = k = (\log x)^{1/5}$, $f(x) = \log_2 x$, $\delta = x^{-1/20}$. Then we see that all the conditions in Lemma 4.3 are satisfied. We conclude there exist random variables $\mathbf{e}'_p(\vec{a})$ whose support is contained in the support of $\mathbf{e}_p(\vec{a})$ together with $\emptyset$ such that

$$\#\left\{q \in \mathscr{Q}(\vec{a}) \ : \ q \notin \mathbf{e}'_p(\vec{a}), \text{ for all } p \in \mathscr{P}\right\} \sim \frac{1}{\log_2 x} |\mathscr{Q}(\vec{a})| \leq \frac{x}{6 \log x}$$

with probability $1 - o(1)$. But $\mathbf{e}'_p(\vec{a}) = \{\mathbf{n}'_p + h_i p : 1 \leq i \leq k\} \cap \mathscr{Q}(\vec{a})$ or $\emptyset$ for some random integers $\mathbf{n}'_p$ with the same support as $\mathbf{n}_p$. Hence

$$\{q \in \mathscr{Q}(\vec{a}) \ : \ q \not\equiv \mathbf{n}'_p \pmod p \text{ for all } p \in \mathscr{P}\} \subseteq \{q \in \mathscr{Q}(\vec{a}) \ : \ q \notin \mathbf{e}'_p(\vec{a}), \text{ for all } p \in \mathscr{P}\}$$

Now just take n_p in the range of $\mathbf{n}'_p$ such that

$$\{q \in \mathscr{Q}(\vec{a}) : q \not\equiv n_p \pmod p \text{ for all } p \in \mathscr{P}\} \leq \frac{x}{6 \log x} \ .$$

This completes the proof, since after using all primes less than $x/2$ in our sieving process, we are left with at most $\frac{x}{5 \log x}$ which can be cleared using primes in $[x/2, x]$.

References

[1] R. C. Baker, G. Harman, and J. Pintz. "The difference between consecutive primes. II". In: *Proc. London Math. Soc. (3)* 83.3 (2001), pp. 532–562. ISSN: 0024-6115. DOI: 10.1112/plms/83.3.532. URL: https://doi.org/10.1112/plms/83.3.532.

[2] William D. Banks, Tristan Freiberg, and James Maynard. "On limit points of the sequence of normalized prime gaps". In: *Proc. Lond. Math. Soc. (3)* 113.4 (2016), pp. 515–539. ISSN: 0024-6115. DOI: 10.1112/plms/pdw036. URL: https://doi.org/10.1112/plms/pdw036.

[3] N.G. Bruijn, de. "On the number of positive integers $\leq x$ and free of prime factors $> y$. II". English. In: *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen: Series A: Mathematical Sciences* 69.3 (1966), pp. 239–247. ISSN: 0023-3358.

[4] Harold Davenport. *Multiplicative number theory.* Third. Vol. 74. Graduate Texts in Mathematics. Revised and with a preface by Hugh L. Montgomery. Springer-Verlag, New York, 2000, pp. xiv+177. ISBN: 0-387-95097-4.

[5] Kevin Ford et al. "Large gaps between consecutive prime numbers". In: *Ann. of Math. (2)* 183.3 (2016), pp. 935–974. ISSN: 0003-486X. DOI: 10.4007/annals.2016.183.3.4. URL: https://doi.org/10.4007/annals.2016.183.3.4.

[6] Kevin Ford et al. "Long gaps between primes". In: *J. Amer. Math. Soc.* 31.1 (2018), pp. 65–105. ISSN: 0894-0347. DOI: 10.1090/jams/876. URL: https://doi.org/10.1090/jams/876.

[7] P. Frankl and V. Rödl. "Near perfect coverings in graphs and hypergraphs". In: *European J. Combin.* 6.4 (1985), pp. 317–326. ISSN: 0195-6698. DOI: 10.1016/S0195-6698(85)80045-7. URL: https://doi.org/10.1016/S0195-6698(85)80045-7.

[8] Andrew Granville. "Harald Cramér and the distribution of prime numbers". In: *Scand. Actuar. J.* 1 (1995). Harald Cramér Symposium (Stockholm, 1993), pp. 12–28. ISSN: 0346-1238. DOI: 10.1080/03461238.1995.10413946. URL: https://doi.org/10.1080/03461238.1995.10413946.

[9] Andrew Granville. "Primes in intervals of bounded length". In: *Bull. Amer. Math. Soc. (N.S.)* 52.2 (2015), pp. 171–222. ISSN: 0273-0979. DOI: 10.1090/S0273-0979-2015-01480-1. URL: https://doi.org/10.1090/S0273-0979-2015-01480-1.

[10] H. Halberstam and H.-E. Richert. *Sieve methods.* London Mathematical Society Monographs, No. 4. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], London-New York, 1974, xiv+364 pp. (loose errata).

[11] Henryk Iwaniec and Emmanuel Kowalski. *Analytic number theory.* Vol. 53. American Mathematical Society Colloquium Publications. American Mathematical Society, Providence, RI, 2004, pp. xii+615. ISBN: 0-8218-3633-1. DOI: 10.1090/coll/053. URL: https://doi.org/10.1090/coll/053.

[12] James Maynard. "Small gaps between primes". In: *Ann. of Math. (2)* 181.1 (2015), pp. 383–413. ISSN: 0003-486X. DOI: 10.4007/annals.2015.181.1.7. URL: https://doi.org/10.4007/annals.2015.181.1.7.

[13] James Maynard. "Dense clusters of primes in subsets". In: *Compos. Math.* 152.7 (2016), pp. 1517–1554. ISSN: 0010-437X. DOI: 10.1112/S0010437X16007296. URL: https://doi.org/10.1112/S0010437X16007296.

[14] James Maynard. "Large gaps between primes". In: *Ann. of Math. (2)* 183.3 (2016), pp. 915–933. ISSN: 0003-486X. DOI: 10.4007/annals.2016.183.3.3. URL: https://doi.org/10.4007/annals.2016.183.3.3.

[15] R. A. Rankin. "The Difference between Consecutive Prime Numbers". In: *J. London Math. Soc.* 11.4 (1936), pp. 242–245. DOI: 10.1112/jlms/s1-13.4.242. URL: https://doi.org/10.1112/jlms/s1-13.4.242.

[16] Erik Westzynthius. "Uber die Verteilung der Zahlen, die zu den n ersten Primzahlen teilerfremd sind". In: *Commentationes Physico–Mathematicae* 5 (1931), pp. 1–37.