*Bull. Aust. Math. Soc.* **100** (2019), 378–387
doi:10.1017/S0004972719000467

# POSITIVE PROPORTION OF SHORT INTERVALS CONTAINING A PRESCRIBED NUMBER OF PRIMES

DANIELE MASTROSTEFANO

(Received 30 January 2019; accepted 6 March 2019; first published online 17 May 2019)

**Abstract**

We prove that for every $m \geq 0$ there exists an $\varepsilon = \varepsilon(m) > 0$ such that if $0 < \lambda < \varepsilon$ and $x$ is sufficiently large in terms of $m$ and $\lambda$, then

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

The value of $\varepsilon(m)$ and the dependence of the implicit constant on $\lambda$ and $m$ may be made explicit. This is an improvement of the author's previous result. Moreover, we will show that a careful investigation of the proof, apart from some slight changes, can lead to analogous estimates when allowing the parameters $m$ and $\lambda$ to vary as functions of $x$ or replacing the set $\mathbb{P}$ of all primes by primes belonging to certain specific subsets.

2010 *Mathematics subject classification*: primary 11N05; secondary 11N36.

*Keywords and phrases*: primes in short intervals, Maynard's sieve theory.

## 1. Introduction

Let $\mathbb{P}$ denote the set of prime numbers and fix $\lambda > 0$ a real number and $m$ a nonnegative integer. The author [2] has recently proved that the proportion of short intervals of the form $[n, n + \lambda \log n]$, for $n \leq x$, containing exactly $m$ primes can be bounded below by $1/\log x$ if we choose $\lambda$ sufficiently small. More precisely, it was shown that

$$d_{\lambda,m}(x) := \frac{|\{n \leq x : |[n, n + \lambda \log n] \cap \mathbb{P}| = m\}|}{x} \gg_{m,\lambda} \frac{1}{\log x}, \tag{1.1}$$

whenever $0 < \lambda < \varepsilon$ for a certain $\varepsilon = \varepsilon(m) > 0$ and $x$ large enough in terms of $m$ and $\lambda$. (Note that the dependences on $\lambda$ of the implied constant and of $x$ were not stated explicitly in [2].) Under these circumstances, it is a considerable improvement of a result of Freiberg [1], who gave the lower bound $d_{\lambda,m}(x) \gg x^{-\varepsilon'(x)}$, with $\varepsilon'(x) = (\log\log\log\log x)^2/(\log\log\log x)$, true for any choice of the parameters $\lambda$ and $m$.

The author is funded by a Departmental Award and by an EPSRC Doctoral Training Partnership Award.

The idea behind both those results is that we can make use of Maynard's sieve method [3] to find clusters of consecutive primes inside particular sets and then construct short intervals of a specific form around them. Indeed, Maynard's sieve method allows us to show that any subset of the primes, which is well distributed in arithmetic progressions, contains many elements that are close together. The work of Freiberg showed that the subset of primes which belongs to the image of certain admissible sets of linear functions is well distributed in arithmetic progressions and is suitable for the application of Maynard's results. A combinatorial process is then used to detect a fixed number among them that are contained in our selected set of intervals.

The major difference between the work of Freiberg and that of the author is in the way the admissible set of linear forms is generated. In the former case, an Erdős–Rankin-type construction [1, Lemma 3.3] was considered, providing a lower bound for the density related to each choice of $\lambda$ and $m$. However, this freedom inevitably forces us to lose precision and obtain weak estimates. In the latter case, the set of linear forms was chosen implicitly by means of Maynard's sieve, producing better information on the density, but only for very small values of $\lambda$.

The aim of the present note is to improve the author's previous work, showing that a better exploration of the approach generates a positive proportion of short intervals containing a prescribed number of primes. The key idea is that at the start of the process we need to select clusters of primes in which the elements are also well spaced.

From now on, $m$ denotes a nonnegative integer and $k = C\exp(49m/C')$, for certain suitable constants $C, C' > 0$, and $\lambda$ denotes a positive real number smaller than $\varepsilon = \varepsilon(k) := k^{-4}(\log k)^{-2}$. The result is the following theorem.

THEOREM 1.1. *We have*

$$d_{\lambda,m}(x) \gg \lambda^{k+1} e^{-Dk^4 \log k}, \tag{1.2}$$

*for a certain absolute constant $D > 0$, if $x$ is sufficiently large in terms of $m$ and $\lambda$.*

It is interesting to note that, from a heuristic point of view, we expect a positive proportion result for all short intervals of the form $[n, n + \lambda \log n]$ (and for all nonnegative integers $m$). More precisely, we conjecture that

$$d_{\lambda,m}(x) \sim \frac{\lambda^m e^{-\lambda}}{m!} \quad \text{as } x \to \infty$$

for every $\lambda$ and $m$. (See the expository article [5] by Soundararajan for further discussions.)

The strength of Maynard's sieve is its flexibility, making it applicable to counting primes in sparser subsets as well. In fact, the same proof that leads to Theorem 1.1 can be adapted to study a variety of different situations, in which for instance we restrict the primes to lie on an arithmetic progression or allow for uniformity of the parameters $\lambda$ and $m$. This gives the following results.

THEOREM 1.2. *Let $x$ be sufficiently large in terms of $m$ and $\lambda$. Suppose that $q \le f(x)$ is a positive integer with $(\log x)/f(x) \to \infty$ as $x \to \infty$. Take $0 \le a < q$ with $(a, q) = 1$. Then*

$$d_{\lambda,m}^{a,q}(x) \gg \frac{\lambda^{k+1} e^{-Dk^4 \log k}}{q^{k+1}}, \tag{1.3}$$

*for a certain $D > 0$, where $d_{\lambda,m}^{a,q}(x)$ is defined as in (1.1) but with $\mathbb{P}$ replaced by $\mathbb{P}_{a,q}$, the intersection of $\mathbb{P}$ with $a \pmod q$.*

THEOREM 1.3. *Fix a small parameter $\epsilon_1 > 0$ and $0 < \epsilon_2 < 1$. Let $x \ge x_0(\epsilon_1, \epsilon_2)$, $m \le \epsilon_1 \log\log x$ and $\lambda \ge (\log x)^{\epsilon_2 - 1}$, satisfying $k^4(\log k)^2\lambda \le 1$ and $\lambda > k \log k(\log x)^{-1}$. Then the estimate (1.2) continues to hold.*

THEOREM 1.4. *Let $\mathbb{K}/\mathbb{Q}$ be a Galois extension of $\mathbb{Q}$ with discriminant $\Delta_{\mathbb{K}}$. There exist constants $C_{\mathbb{K}}, C'_{\mathbb{K}} > 0$ depending only on $\mathbb{K}$ such that the following holds. Let $C \subset \mathrm{Gal}(\mathbb{K}/\mathbb{Q})$ be a conjugacy class in the Galois group of $\mathbb{K}/\mathbb{Q}$ and let*

$$\mathcal{P} = \Big\{p \text{ prime} : p \nmid \Delta_{\mathbb{K}}, \Big[\frac{\mathbb{K}/\mathbb{Q}}{p}\Big] = C\Big\},$$

*where $[\frac{\mathbb{K}/\mathbb{Q}}{\cdot}]$ denotes the Artin symbol. Let $m \in \mathbb{N}$, $k = C'_{\mathbb{K}} \exp(C_{\mathbb{K}} m)$ and $\lambda < \varepsilon$. Then*

$$d_{\lambda,m}^{\mathbb{K}}(x) \gg \lambda^{k+1} e^{-Dk^4 \log k},$$

*provided $x \ge x_0(\mathbb{K}, \lambda, m)$, where $d_{\lambda,m}^{\mathbb{K}}(x)$ is defined as in (1.1) except that $\mathbb{P}$ is replaced by $\mathcal{P}$.*

If we consider values of $\lambda$ slightly bigger than $k^{-4}(\log k)^{-2}$, a small variation of the sieve method used to prove Theorem 1.1 leads to the following improvement on the Freiberg bound in [1].

THEOREM 1.5. *For every nonnegative integer $m$ and positive real number $\lambda$ smaller than $k^{-1}(\log k)^{-1}$, with $k = C \exp(49m/C')$ for suitable constants $C, C' > 0$,*

$$d_{\lambda,m}(x) \gg \frac{\lambda e^{-Dk^4 \log k}}{(\log x)^k},$$

*for a certain $D > 0$, if $x$ is sufficiently large in terms of $m$ and $\lambda$.*

## 2. Notation

Throughout, $\mathbb{P}$ denotes the set of all primes, $\mathbf{1}_S : \mathbb{N} \to \{0, 1\}$ the indicator function of a set $S \subset \mathbb{N}$ and $p$ a prime. As usual, $\varphi$ will denote the Euler totient function and $(m, n)$ the greatest common divisor of integers $n$ and $m$. We will always denote by $x$ a sufficiently large real number. By $o(1)$ we mean a quantity that tends to 0 as $x$ tends to infinity. The expressions $A = O(B), A \ll B$ and $B \gg A$ mean $|A| \le c|B|$, where $c$ is some positive (absolute, unless stated otherwise) constant.

In the following, we will always consider admissible $k$-tuples of linear forms $\{gn + h_1, \ldots, gn + h_k\}$, where $0 \le h_1 < h_2 < \cdots < h_k < \lambda \log x$, $k$ a sufficiently large integer and $g$ a positive integer which is coprime with $B$, squarefree and such that $\log x < g \le 2\log x$. Here, $B = 1$ or $B$ is a prime with $\log\log x^{\eta} \ll B \ll x^{2\eta}$, where we put $\eta := c/500k^2$ with $0 < c < 1$. A finite set $\mathcal{L} := \{L_1, \dots, L_k\}$ of linear functions is admissible if the set of solutions modulo $p$ to $L_1(n)\cdots L_k(n) \equiv 0 \pmod p$ does not form a complete residue system modulo $p$ for any prime $p$. In our case, in which $L_i(n) = gn + h_i$ for every $i = 1, \dots, k$, the set $\{L_1, \dots, L_k\}$ is admissible if and only if the set $\mathcal{H} := \{h_1, \dots, h_k\}$ is admissible, in the sense that the elements $h_1, \dots, h_k$ do not cover all the residue classes modulo $p$ for any prime $p$.

The proof of Theorem 1.1 (and of its variations, Theorems 1.2–1.5) follows by mimicking [2], taking into account a new crucial assumption on the set of linear forms. We will briefly rewrite the main estimates and passages already contained in the proof in [2], highlighting the main differences and the new computations. In particular, several types of notation will not be introduced here because they are not essential for the general understanding of the argument or already present in [2].

## 3. Application of Maynard's sieve

As at the start of [2, Section 3], and following the notation there introduced, we define the double sum

$$S = \sum_{\mathcal{H}}^{*} \sum_{x<n\le 2x} S(\mathcal{H}, n), \tag{3.1}$$

where

$$S(\mathcal{H}, n) = \Big(\sum_{i=1}^{k} \mathbf{1}_{\mathbb{P}}(gn + h_i) - m - k\sum_{i=1}^{k} \sum_{\substack{p\mid gn+h_i\\ p\le x^{\rho},\, p\nmid B}} 1 - k \sum_{\substack{h\le 5\lambda\log x\\ (h,g)=1\\ h\notin\mathcal{H}}} \mathbf{1}_{S(\rho,B)}(gn+h)\Big) w_n(\mathcal{H}). \tag{3.2}$$

In (3.1), $\Sigma^{*}_{\mathcal{H}}$ means that the sum is taken over all the admissible sets $\mathcal{H}$ such that $0 \le h_1 < h_2 < \cdots < h_k < \lambda\log x$ and $|h_i - h_j| > C_0^{-1}\lambda\log x$, for $1 \le i \ne j \le k$, where $C_0$ will be chosen later. Note that, unlike the corresponding definition in [2], in the innermost sum in (3.1) we now have $m$ instead of $m - 1$.

Following closely the discussion at the beginning of [2, Section 3], we deduce that

$$S \ll k(\log x)^{2k} \exp(O(k/\rho))|I(x)|, \tag{3.3}$$

where now the set $I(x)$ contains intervals of the form $[gn, gn + 5\lambda\log x]$, for $x < n \le 2x$, with the property that $|[gn, gn + 5\lambda\log x] \cap \mathbb{P}| = |\{gn + h_1, \dots, gn + h_k\} \cap \mathbb{P}| \ge m + 1$ for a unique admissible set $\mathcal{H}$ such that $0 \le h_1 < \cdots < h_k < \lambda\log x$ and also $|h_i - h_j| > C_0^{-1}\lambda\log x$ for $1 \le i \ne j \le k$. We recall that the intervals in $I(x)$ are pairwise disjoint if for instance $\lambda < 1/5$.

https://doi.org/10.1017/S0004972719000467 Published online by Cambridge University Press

We also need a lower bound for $S$. Using [2, Propositions 2.1–2.5],

$$S \geq \sum_{\mathcal{H}}^{*}\Big[(1+o(1))\frac{B^{k-1}}{\varphi(B)^{k-1}}\mathfrak{S}_B(\mathcal{H})(\log R)^{k+1}J_k\sum_{i=1}^{k}\frac{\varphi(g)}{g}\sum_{x<n\leq 2x}\mathbf{1}_{\mathbb{P}}(gn+h_i)$$
$$-m(1+o(1))\frac{B^k}{\varphi(B)^k}\mathfrak{S}_B(\mathcal{H})x(\log R)^k I_k$$
$$+O\Big(\rho^2k^6(\log k)^2\frac{B^k}{\varphi(B)^k}\mathfrak{S}_B(\mathcal{H})x(\log R)^k I_k\Big)$$
$$+O\Big(k\frac{B^k}{\varphi(B)^k}\mathfrak{S}_B(\mathcal{H})x(\log R)^{k-1}I_k\Big)$$
$$+O\Big(\frac{k}{\rho}\frac{B^{k+1}}{\varphi(B)^{k+1}}\mathfrak{S}_B(\mathcal{H})x(\log R)^{k-1}I_k\sum_{\substack{h\leq 5\lambda\log x\\ (h,g)=1\\ h\notin\mathcal{H}}}\frac{\Delta_{\mathcal{L}}}{\varphi(\Delta_{\mathcal{L}})}\Big)\Big]. \tag{3.4}$$

By the inequality [2, (2.7)],

$$\frac{\varphi(B)}{B}\frac{\varphi(g)}{g}\sum_{i=1}^{k}\sum_{x<n\leq 2x}\mathbf{1}_{\mathbb{P}}(gn+h_i) > \frac{kx}{2\log x}.$$

Note that the hypotheses of [2, Theorem 2.2] are satisfied. Using this estimate together with [2, Lemma 3.1], and choosing $\rho := k^{-3}(\log k)^{-1}$, we find that

$$S \geq \sum_{\mathcal{H}}^{*}\frac{B^k}{\varphi(B)^k}\mathfrak{S}_B(\mathcal{H})x(\log R)^k\Big[(1+o(1))kJ_k\frac{\log R}{2\log x} - mI_k(1+o(1)) + O(I_k)$$
$$+O(kI_k(\log R)^{-1}) + O(k^4(\log k)I_k(\log R)^{-1}\lambda\log x\log k)\Big]. \tag{3.5}$$

We remark here that the aforementioned Theorem 2.2 and Lemma 3.1 need $x$ to be large enough with respect to $k$ and $\lambda$. Now, by [2, Propositions 2.1 and 2.6], we know that $J_k \geq C'k^{-1}\log kI_k$ for a certain $C' > 0$. We should consider $k$ sufficiently large in terms of $m$. For example, we may take $k := C\exp(49m/C')$ with $C > 0$. Choosing $\lambda \leq \varepsilon$, with $\varepsilon = \varepsilon(k) := k^{-4}(\log k)^{-2}$, and taking $x$ and $C$ suitably large, we may conclude that

$$S \gg \sum_{\mathcal{H}}^{*}\frac{B^k}{\varphi(B)^k}\mathfrak{S}_B(\mathcal{H})x(\log R)^k I_k. \tag{3.6}$$

By the estimates [2, Propositions 2.1 and 2.6], we know that $I_k \gg (2k\log k)^{-k}$ and $\mathfrak{S}_B(\mathcal{H}) \gg \exp(-C_1k)$ for a certain $C_1 > 0$. Remember also that $R = x^{1/24}$. Finally, we may certainly use $B^k/\varphi(B)^k \geq 1$. Inserting all of these in (3.6),

$$S \gg x(\log x)^k e^{-C_2k^2}\sum_{\mathcal{H}}^{*}1 \tag{3.7}$$

for a suitable constant $C_2 > 0$. Thus, we are left with obtaining a lower bound for the sum in (3.7). We greedily sieve the interval $[0, \lambda\log x]$ by removing for each prime $p \leq k$ in turn any elements from the residue class modulo $p$ which contains the fewest

https://doi.org/10.1017/S0004972719000467 Published online by Cambridge University Press

elements. The resulting set $\mathcal{A}$, say, has size

$$|\mathcal{A}| \geq \lambda \log x \prod_{p \leq k}\Big(1 - \frac{1}{p}\Big) \geq c' \frac{\lambda \log x}{\log k},$$

by Mertens's theorem, with $c' > 0$.

Any choice of $k$ distinct $h_i$ from $\mathcal{A}$ will be an admissible set $\mathcal{H} = \{h_1, \ldots, h_k\}$ such that $0 \leq h_1 < h_2 < \cdots < h_k < \lambda \log x$. Now, we count how many of them have the property that $|h_i - h_j| > C_0^{-1}\lambda \log x$ for $1 \leq i \neq j \leq k$. Certainly, we can choose $h_1$ in $|\mathcal{A}|$ ways. Let us write $\mathcal{A}_1 := \mathcal{A}$ and define

$$\mathcal{A}_2 := \mathcal{A}_1 \setminus \mathcal{A}_1 \cap [h_1 - \lfloor \lambda \log x/C_0 \rfloor, h_1 + \lfloor \lambda \log x/C_0 \rfloor].$$

We will pick $h_2 \in \mathcal{A}_2$, having then $|\mathcal{A}_2| \geq |\mathcal{A}_1| - 2\lfloor \lambda \log x/C_0 \rfloor$ possibilities. Iterating this process, we can count the number of admissible choices for any $h_i$ up to $h_k$, which will be an element in

$$\mathcal{A}_k := \mathcal{A}_{k-1} \setminus \mathcal{A}_{k-1} \cap [h_{k-1} - \lfloor \lambda \log x/C_0 \rfloor, h_{k-1} + \lfloor \lambda \log x/C_0 \rfloor]$$

and which will have cardinality $|\mathcal{A}_k| \geq |\mathcal{A}_1| - 2(k-1)\lfloor \lambda \log x/C_0 \rfloor$.

In conclusion, for our particular choice of admissible sets we have at least a number of possibilities equal to

$$\frac{1}{k!}\prod_{i=1}^{k}|\mathcal{A}_i| \geq \frac{1}{k^k}\prod_{i=1}^{k}(|\mathcal{A}_1| - 2(i-1)\lfloor \lambda \log x/C_0 \rfloor) \geq \frac{1}{k^k}\Big(c'\frac{\lambda \log x}{\log k} - 2k\frac{\lambda \log x}{C_0}\Big)^k. \quad (3.8)$$

Take $C_0 = C_0(k) := 4k(\log k)/c'$. We immediately see that the right-hand side of (3.8) is $\gg \lambda^k e^{-C_3 k^2}(\log x)^k$, which leads to $S \gg \lambda^k e^{-C_4 k^2} x(\log x)^{2k}$ for certain constants $C_3, C_4 > 0$. Finally, by combining (3.3) with the above information on $S$,

$$|I(x)| \gg \lambda^k e^{-C_5 k^4 \log k} x \quad (3.9)$$

with an absolute constant $C_5 > 0$.

## 4. Modification of the combinatorial process

Consider an interval $I \in I(x)$. There exist an integer $n$ with $x < n \leq 2x$ and an admissible set $\mathcal{H}$, with $0 \leq h_1 < h_2 < \cdots < h_k < \lambda \log x$ and $|h_i - h_j| > \lambda \log x/C_0$ for $1 \leq i \neq j \leq k$, such that $I = [gn, gn + 5\lambda \log x]$ and

$$|[gn, gn + 5\lambda \log x] \cap \mathbb{P}| = |\{gn + h_1, \ldots, gn + h_k\} \cap \mathbb{P}| \geq m + 1.$$

In order to avoid having a trivial gap between the elements of $\mathcal{H}$, we ask for $x$ to be sufficiently large with respect to $\lambda$ and $k$. Let us define

$$I_j = [N_j, N_j + \lambda \log N_j], \qquad N_j = gn + j,$$

for $j = 0, \ldots, \lfloor \lambda \log N_0 \rfloor$. We recall here the following properties of the intervals $I_j$ that are stated and proved in detail in [2]:

(1) $I_j \subseteq I$ for any such $j$;
(2) $I_j \cap \{gn + h_1, \ldots, gn + h_k\} = \{gn + h_1, \ldots, gn + h_k\}$ for the choice $j = h_1$;
(3) $I_j \cap \{gn + h_1, \ldots, gn + h_k\} = \emptyset$ for the value $j = \lfloor \lambda \log N_0 \rfloor$;
(4) if $|I_j \cap \mathbb{P}| < |I_{j+1} \cap \mathbb{P}|$ for a certain $j$, then $|I_{j+1} \cap \mathbb{P}| = |I_j \cap \mathbb{P}| + 1$.

https://doi.org/10.1017/S0004972719000467 Published online by Cambridge University Press

Now, let us define

$$\tilde{j} := \max\{0 \le j \le \lfloor \lambda \log N_0 \rfloor : |I_j \cap \mathbb{P}| \ge m+1\}.$$

Note that necessarily $N_{\tilde{j}} = gn + \tilde{j}$ is prime. Consequently, $|I_{\tilde{j}+1} \cap \mathbb{P}| = m$. But from our assumption on $\mathcal{H}$,

$$|I_{\tilde{j}+l} \cap \mathbb{P}| = m \quad \text{for any } l \text{ with } 1 \le l \le \lfloor \lambda \log x / C_0 \rfloor.$$

This means that we have found $\lfloor \lambda \log x / C_0 \rfloor$ different intervals $[N, N + \lambda \log N]$ containing exactly $m$ primes, with $N < 5x \log x$, if $x$ is sufficiently large. Together with the lower bound (3.9), we have shown that for every $m \ge 0$ and for each $\lambda \le \varepsilon$,

$$|\{N \le 5x \log x : |[N, N + \lambda \log N] \cap \mathbb{P}| = m\}| \gg \frac{\lambda^{k+1}}{k \log k} e^{-C_5 k^4 \log k} x \log x,$$

which is equivalent to

$$|\{N \le X : |[N, N + \lambda \log N] \cap \mathbb{P}| = m\}| \gg \lambda^{k+1} e^{-C_6 k^4 \log k} X, \tag{4.1}$$

when $X$ is large enough in terms of $\lambda$ and $k$, for a certain constant $C_6 > 0$, which proves Theorem 1.1.

## 5. Concluding remarks

**5.1. Explicit constants.** Since we let $k = C \exp(49m/C')$, with $C, C'$ as above, note that we can rewrite the final estimate using only the relationship between $\lambda$ and $m$. Remembering the choice of $\varepsilon(m)$, we see immediately that the connection between $\lambda$ and $m$ is given by

$$\lambda(49m + c_1)^2 \exp(196 c_2 m) \ll 1 \tag{5.1}$$

for certain constants $c_1, c_2 > 0$.

REMARK 5.1. Notice that, if we could take $k = m+1$ and if we were able to improve the constants in $k$ in the sieve method, then we would end up with an explicit constant in (4.1) that almost matches the expected one for values of $\lambda$ close to 0.

**5.2. The case of primes in arithmetic progressions.** Suppose that $q$ is a squarefree positive integer coprime with $B$ and $q \le f(x)$ with $(\log x)/f(x) \to \infty$ as $x \to \infty$. Take $0 \le a < q$ with $(a, q) = 1$. In order to extend the result of Theorem 1.1 to this situation, we go over its proof again. In particular, in (3.1) we now average over admissible sets $\mathcal{H} = \{h_1, \ldots, h_k\}$ such that $0 \le h_1 := a + qb_1 < h_2 := a + qb_2 < \cdots < h_k := a + qb_k < \lambda \log x$ and $|b_i - b_j| > C_0^{-1} q^{-1} \lambda \log x$ for $1 \le i \ne j \le k$. Moreover, we need to take $g$ as a squarefree multiple of $q$ coprime with $B$ and such that $\log x < g \le 2 \log x$. Such a set of linear functions satisfies the hypotheses of [2, Proposition 2.1 and Theorem 2.2] and the images of all its elements lie on the arithmetic progression $a \pmod q$. In particular, as in Section 3,

$$|I(x)| \gg \frac{\lambda^k e^{-C_8 k^4 \log k}}{q^k} x,$$

if $x$ is sufficiently large in terms of $\lambda$ and $k$, for a suitable constant $C_8 > 0$.

https://doi.org/10.1017/S0004972719000467 Published online by Cambridge University Press

Using the notation $\mathbb{P}_{a,q}$ to indicate the primes in the arithmetic progression $a \pmod q$, the set $I(x)$ contains intervals of the form $[gn, gn + 5\lambda \log x]$, for $x < n \le 2x$, with $g \equiv 0 \pmod q$ and $|[gn, gn + 5\lambda \log x] \cap \mathbb{P}| = |\{gn + h_1, \dots, gn + h_k\} \cap \mathbb{P}_{a,q}| \ge m + 1$ for a unique admissible set $\mathcal{H} = \{h_1, \dots, h_k\}$ such that $0 \le h_1 := a + qb_1 < h_2 := a + qb_2 < \cdots < h_k := a + qb_k < \lambda \log x$ and $|b_i - b_j| > C_0^{-1} q^{-1} \lambda \log x$ for $1 \le i \ne j \le k$. Following the computations in Section 4,

$$|\{N \le X : |[N, N + \lambda \log N] \cap \mathbb{P}_{a,q}| = m\}| \gg \frac{\lambda^{k+1} e^{-C_9 k^4 \log k}}{q^{k+1}} X,$$

when $X$ is sufficiently large in terms of $\lambda$ and $k$, for a suitable absolute constant $C_9 > 0$, which proves Theorem 1.2. The restriction on $q$ to be squarefree and coprime with $B$ can be removed at the cost of slightly modifying the proof of [2, Theorem 2.2]. In particular, at the start of its proof we need to replace $B$ with the largest prime factor of $\tilde{l}$ coprime with $g$, with $\tilde{l}$ being the modulus of a possible exceptional character among all the primitive Dirichlet characters $\chi$ mod $l$ with moduli $l \le x^{2\eta}$.

**5.3. The case of uniform parameters.** In Section 3, we applied [2, Proposition 2.1], which is a specific case of [3, Proposition 6.1], in which a uniformity in $k \le (\log x)^{1/5}$, say, is allowed. A careful examination of [2, Theorem 2.2] and of the computations in Sections 3 and 4 in the present paper and in [2] shows that the estimate (1.2) continues to hold when $m \le \epsilon_1 \log\log x$ and $\lambda \ge (\log x)^{\epsilon_2 - 1}$ satisfy (5.1) together with $\lambda > k \log k (\log x)^{-1}$. Here, $\epsilon_1$ is a fixed sufficiently small constant (for example, smaller than $C'/294$) and $0 < \epsilon_2 < 1$.

Remark 5.2. Notice that, in the case in which $\lambda$ goes to 0 together with $x$ and $m$ and $\lambda$ varies in the range defined above, the Cramér model used in [5] still gives us an expected asymptotic value for $d_{\lambda,m}$, which now takes the form

$$d_{\lambda,m}(x) \sim \frac{\lambda^m}{m!} \quad \text{as } x \to \infty.$$

Obviously, since the constant in $m$ in the lower bound (1.2) is not optimal, the value of $d_{\lambda,m}(x)$ now will be far away from what the model suggests.

**5.4. The case of primes in Chebotarev sets.** As already mentioned, so far we have only used a very special case of [3, Proposition 6.1]. In particular, we can replace the set of all the primes with a smaller one, as long as it satisfies a suitable variant of [2, Theorem 2.2]. More specifically, we would like to concentrate on the so-called primes in Chebotarev sets.

Let $\mathbb{K}/\mathbb{Q}$ be a Galois extension of $\mathbb{Q}$ with discriminant $\Delta_{\mathbb{K}}$. Let $C \subset \mathrm{Gal}(\mathbb{K}/\mathbb{Q})$ be a conjugacy class in the Galois group of $\mathbb{K}/\mathbb{Q}$ and let

$$\mathcal{P} = \left\{ p \text{ prime} : p \nmid \Delta_{\mathbb{K}}, \left[\frac{\mathbb{K}/\mathbb{Q}}{p}\right] = C \right\},$$

where $[\frac{\mathbb{K}/\mathbb{Q}}{\cdot}]$ denotes the Artin symbol. Fix $m \in \mathbb{N}$, $k = C'_{\mathbb{K}} \exp(C_{\mathbb{K}} m)$, for suitable $C_{\mathbb{K}}, C'_{\mathbb{K}} > 0$, and $\lambda < \varepsilon$. Finally, let $\log x < g \le 2 \log x$ be a squarefree number with

https://doi.org/10.1017/S0004972719000467 Published online by Cambridge University Press

$(g, \Delta_{\mathbb{K}}) = 1$, bearing in mind that now $B = \Delta_{\mathbb{K}}$, and consider admissible sets $\mathcal{H}$ of the usual form.

Murty and Murty proved in their main theorem in [4] that the primes in $\mathcal{P}$ are well distributed among arithmetic progressions of moduli $q \le x^{\theta}$, with $\theta < \min(1/2, 2/|G|)$, and such that $\mathbb{K} \cap \mathbb{Q}(\zeta_q) = \mathbb{Q}$. An adaptation of the argument in the proof of [2, Theorem 2.2] leads to the second estimate stated there, where $\mathbb{P}$ is replaced by $\mathcal{P}$ and the sum over $q$ is over all the moduli $q \le x^{\theta/4}$ satisfying the algebraic condition described above. Regarding the first estimate in [2, Theorem 2.2],

$$\frac{1}{k}\frac{B}{\varphi(B)}\frac{\varphi(g)}{g}\sum_{i=1}^{k}\sum_{x<n\le 2x}\mathbf{1}_{\mathcal{P}}(gn+h_i) \ge (1+o(1))\frac{\Delta_{\mathbb{K}}}{\varphi(\Delta_{\mathbb{K}})}\frac{|C|}{|G|}\frac{x}{\log x},$$

which essentially follows from the Chebotarev density theorem. Working as in Section 3, we find that

$$|I(x)| \gg \lambda^k e^{-C_{10}k^4\log k}x,$$

if $x \ge x_0(\mathbb{K}, \lambda, m)$, for a suitable constant $C_{10} > 0$. Here, the set $I(x)$ contains intervals of the form $[gn, gn + 5\lambda\log x]$, for $x < n \le 2x$ and $\log x < g \le 2\log x$, having the property that $|[gn, gn + 5\lambda\log x] \cap \mathbb{P}| = |\{gn + h_1, \ldots, gn + h_k\} \cap \mathcal{P}| \ge m + 1$ for a unique admissible set $\mathcal{H}$ such that $0 \le h_1 < h_2 < \cdots < h_k < \lambda\log x$ and also $|h_i - h_j| > C_0^{-1}\lambda\log x$ for $1 \le i \ne j \le k$. Following the computations in Section 4,

$$|\{N \le X : |[N, N + \lambda\log N] \cap \mathcal{P}| = m\}| \gg \lambda^{k+1}e^{-C_{11}k^4\log k}X,$$

when $X \ge X_0(\mathbb{K}, \lambda, m)$, for a suitable absolute constant $C_{11} > 0$, which proves Theorem 1.4.

**5.5. The case of slightly bigger values of $\lambda$.** Let us fix an admissible $k$-tuple of linear functions $\mathcal{L} = \{gn + h_1, \ldots, gn + h_k\}$ with the usual form. We replace the last sum in parenthesis in (3.2) with

$$\sum_{\substack{h\le 5\lambda\log x\\ (h,g)=1\\ h\notin\mathcal{H}}}\mathbf{1}_{S(1/80,1)}(gn+h)$$

and we remove the average over $\mathcal{H}$ in (3.1). With these variations in mind, we see immediately that (3.3) still continues to hold, but now we can only say that for every interval $I \in I(x)$ there exists an integer $x < n \le 2x$ such that $I = [gn, gn + 5\lambda\log x]$ and

$$|[gn, gn + 5\lambda\log x] \cap \mathbb{P}| = |\{gn + h_1, \ldots, gn + h_k\} \cap \mathbb{P}| \ge m + 1.$$

Arguing as in Section 3 with the appropriate variations, but essentially carrying over all the computations, we deduce that

$$S \gg x(\log x)^k e^{-C_{12}k^2}, \quad |I(x)| \gg e^{-C_{13}k^4\log k}\frac{x}{(\log x)^k} \tag{5.2}$$

https://doi.org/10.1017/S0004972719000467 Published online by Cambridge University Press

for suitable $C_{12}, C_{13} > 0$. The only key difference in proving (5.2) is that the last big-$O$ in (3.4) now assumes the shape

$$O\Big(80k\frac{B^k}{\varphi(B)^k}\mathfrak{S}_B(\mathcal{H})x(\log R)^{k-1}I_k \sum_{\substack{h\leq 5\lambda\log x\\ (h,g)=1\\ h\notin\mathcal{H}}}\frac{\Delta_{\mathcal{L}}}{\varphi(\Delta_{\mathcal{L}})}\Big).$$

Consequently, this also modifies the last big-$O$ in (3.5), which will be

$$O(k(\log k)I_k(\log R)^{-1}\lambda\log x) = O(I_k) \quad \text{if } \lambda < \frac{1}{k\log k}.$$

The rest of the argument goes through as before and we conclude that

$$|\{N \leq X : |[N, N+\lambda\log N]\cap\mathbb{P}| = m\}| \gg \lambda e^{-C_{14}k^4\log k}\frac{X}{(\log X)^k},$$

when $X$ is large enough in terms of $\lambda$ and $k$, for a certain $C_{14} > 0$, which proves Theorem 1.5.

Remark 5.3. We would like to observe that many of the variables and parameters have not been chosen in the best possible way, since finding their precise range of definition is not in the spirit of the paper and does not significantly improve the final results. We refer to [6] for several arithmetic consequences of finding primes of a given splitting type and note that they may be translated into our context. Finally, we would like to point out that we are able to combine the results presented in this section, paying attention to the possible relations between the different parameters.

## References

[1] T. Freiberg, 'Short intervals with a given number of primes', *J. Number Theory* **163** (2016), 159–171.

[2] D. Mastrostefano, 'Short intervals containing a prescribed number of primes', *J. Number Theory* **197** (2019), 49–61.

[3] J. Maynard, 'Dense clusters of primes in subsets', *Compos. Math.* **152**(7) (2016), 1517–1554.

[4] M. R. Murty and V. K. Murty, 'A variant of the Bombieri–Vinogradov theorem', in: *Number Theory (Montreal, Quebec, 1985)*, CMS Conference Proceedings, 7 (American Mathematical Society, Providence, RI, 1987), 243–272.

[5] K. Soundararajan, 'The distribution of prime numbers', in: *Equidistribution in Number Theory, An Introduction*, NATO Science Series II: Mathematics, Physics and Chemistry, 237 (eds. A. Granville and Z. Rudnick) (Springer, Dordrecht, 2007), 59–83.

[6] J. Thorner, 'Bounded gaps between primes in Chebotarev sets', *Res. Math. Sci.* **1**(4) (2014), 16 pages.

DANIELE MASTROSTEFANO, University of Warwick,
Mathematics Institute, Zeeman Building,
Coventry, CV4 7AL, UK
e-mail: Daniele.Mastrostefano@warwick.ac.uk

https://doi.org/10.1017/S0004972719000467 Published online by Cambridge University Press