LE MATEMATICHE
Vol. LXXX (2025) – Issue II, pp. 521–544
doi: 10.4418/2025.80.2.2

AN EXPLICIT LOWER BOUND FOR LARGE GAPS BETWEEN SOME CONSECUTIVE PRIMES

K. SONO

Let p_n denote the nth prime and for any fixed positive integer k and $X \geq 2$, put

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

Ford, Maynard and Tao [6] proved that there exists an effective absolute constant $c_{LG} > 0$ such that

$$G_k(X) \geq \frac{c_{LG}}{k^2} \frac{\log X \log\log X \log\log\log\log X}{\log\log\log X}$$

holds for any sufficiently large X. The main purpose of this paper is to clarify the numerical value of the constant c_{LG} such that the above inequality holds. We see that c_{LG} is determined by several factors related to analytic number theory, for example, the ratio of integrals of functions in the multidimensional sieve of Maynard [14], the distribution of primes in arithmetic progressions to large moduli, and the coefficient of upper bound sieve of Selberg. We prove that the above inequality is valid at least for some $c_{LG} \geq 2.0 \times 10^{-17}$.

Received on August 12, 2024

AMS 2010 Subject Classification: 11N36
Keywords: Large gaps between primes, explicit lower bound, probabilistic method

1. Introduction

For a positive integer k and a real number $X \geq 3$, put

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

and in particular, put $G(X) := G_1(X)$, which denotes the largest prime gap below X. The prime number theorem yields the number of primes below x is asymptotically $x/\log x$, so the average gap of consecutive primes below x is $(1+o(1))\log x$. Hence we get the trivial lower bound

$$G(X) \geq (1+o(1))\log X.$$

Harald Cramér [1] constructed a simple probabilistic model of the set of prime numbers, and following his model, he conjectured that $G(X) \sim (\log X)^2$. (See also refinements of Cramér's model by Granville [8], Firoozbakht [20], p.185 and Wolf [24].) In 1931, Westzynthius [23] made the first quantitative improvement on the trivial bound and proved

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

where $\log_n$ denotes the n times composition of the logarithmic function, i.e., $\log_n X := (\underbrace{\log \circ \cdots \circ \log}_{n})X$. In particular, it follows that $G(X)/\log X \to \infty$ as $X \to \infty$, so $G(X)$ can be arbitrarily large compared with the average gap. In 1934, Ricci [21] slightly improved this and showed $G(X) \gg \log X \log_3 X$. Erdős [2] improved on Westzynthius' result and obtained

$$G(X) \gg \frac{\log X \log_2 X}{(\log_3 X)^2}.$$

Rankin [18] made a further improvement and showed that

$$G(X) \geq (c+o(1))\frac{\log X \log_2 X \log_4 X}{(\log_3 X)^2} \tag{1.1}$$

holds with $c = \frac{1}{3}$. Several mathematicians improved on the value of the coefficient c above (see Schönhage [22] with $c = \frac{1}{2}e^{\gamma}$, Rankin [19] with $c = e^{\gamma}$, Maier and Pomerance [13] with $c = 1.31256e^{\gamma}$, and Pintz [16] with $c = 2e^{\gamma}$). Erdős conjectured that (1.1) holds with arbitrarily large c. This problem had been unsolved for a long time, but Maynard [14] and a team of Ford, Green, Konyagin, Tao [5] independently solved at almost the same time in August 2014. Several months later, in a joint work [4], they obtained

$$G(X) \gg \frac{\log X \log_2 X \log_4 X}{\log_3 X}. \tag{1.2}$$

This was the first quantitative improvement of Rankin's bound (1.1) in almost 80 years.

Let k be a fixed positive integer. The $G_k(X)$ above has also been studied for a long time. Following the argument of Cramér, it is conjectured that $G_k(X) \asymp \frac{1}{k}\log^2 X$. Erdős [2] considered the case $k = 2$ and proved

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

as $X \to \infty$. Maier [12] showed

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

by combining his famous Maier matrix method and Pintz's ideas in [16], where (X_n) is some monotonically increasing sequence such that $X_n \to \infty$ as $n \to \infty$. Pintz [17] improved on Maier's result and proved

$$G_k(X_n)/\left(\frac{\log X_n \log_2 X_n \log_4 X_n}{(\log_3 X_n)^2}\right) \to \infty.$$

An issue of Maier's argument is that one has to avoid (possible) Siegel zeros, and this is the reason why the results of Maier and Pintz above are restricted to a special sequence (X_n), rather than all sufficiently large X.

In [6], Ford, Maynard and Tao succeeded in handling this difficulty and gave a lower bound for $G_k(X)$ for any sufficiently large X. Concretely, they proved that

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

holds as $X \to \infty$, and the implied constant above is absolute and effective. In other words, there exists an absolute computable constant $c_{LG} > 0$ such that

$$G_k(X) \geq \frac{c_{LG}}{k^2}\frac{\log X \log_2 X \log_4 X}{\log_3 X} \tag{1.3}$$

holds for every sufficiently large X. In addition to the technique to handle the possibility of the existence of exceptional zeros of L-functions, their result relies on hypergraph covering theorem, the construction of multidimensional sieve weight, and very clever and highly technical probabilistic arguments in [4].

Though the lower bound in their theorem is effective, the explicit value of c_{LG} (resp. the implied constant of (1.2)) is not mentioned in [6] (resp. [4]). (In the blog on the paper [4], Tao says "we manage to avoid the use of the (ineffective) Siegel-Walfisz theorem by deleting an exceptional prime from the multidimensional Selberg sieve, leading to an effective (but quite small) value of c.") However, clarifying the value of c_{LG} seems to be significant because it has the merit of making clear which parameter affects the coefficient c_{LG} and how. With this reason, the main purpose of this paper is to clarify what value is appropriate as the coefficient c_{LG} in (1.3) and how several factors in analytic number theory are related. It is fully expected that this attempt might be a clue to establish a new quantitative improvement on the lower bound of $G_k(X)$. We prove that one can take

$$c_{LG} = \frac{C_{PAP}^2 \theta c_{I,J} e^{4\gamma}}{737280000 \log 5\, C_{UB} M (1 + D_{PAP}^{-1})^4 (25 C_{UB} + 20 e^{\gamma} M)}$$

(see Theorem 3.6), where θ is a parameter which describes how primes are equidistributed in arithmetic progressions (see [14], Hypothesis 1), $c_{I,J}$ is the ratio of integrals of functions (see (2.3)) in the multidimensional sieve of Maynard [14], C_{PAP} and D_{PAP} are constants in the statement on distribution of primes in arithmetic progressions to large moduli (see Assumption 3.2), and C_{UB}, D_{UB} are the constants of upper bound sieve of Selberg (see Assumption 3.4) and $M := \max\{D_{PAP}, D_{UB}\}$. In Sections 2-3, we will see that unconditionally one can take $\theta = \frac{1}{3}$, $c_{I,J} = \frac{1}{4}$, $C_{PAP} = 1 - e^{-2}$, $D_{PAP} = 160$, $C_{UB} = 8e^{2\gamma}$ and $D_{UB} > 0$ arbitrarily. Numerically, one can take

$$c_{LG} \geq 2.0 \times 10^{-17}.$$

Obviously, this value is not the best that can be obtained with current techniques and could be improved to some extent with some effort, for example, by applying recent results on zero density estimates and explicit zero free regions of Dirichlet L-functions. In any case, however, as long as we rely on current methods, the coefficient is likely to have to be fairly small.

2. Notation

Let $c > 0$ be a fixed constant (to be determined later) and x a sufficiently large real number. Put

$$y := c \frac{x \log x \log_3 x}{\log_2 x}, \quad z := x^{\frac{\log_3 x}{4 \log_2 x}}. \tag{2.1}$$

Let B_0 be either 1 or a prime number satisfying

$$\log x \ll B_0 \leq x. \tag{2.2}$$

Define three disjoint sets of primes $\mathcal{S}$, $\mathcal{P}$ and $\mathcal{Q}$ by

$$\mathcal{S} := \{s : \text{prime} \mid \log^{20} x < s \leq z,\ s \neq B_0\},$$

$$\mathcal{P} := \{p : \text{prime} \mid \frac{x}{2} < p \leq x,\ p \neq B_0\},$$

$$\mathcal{Q} := \{q : \text{prime} \mid x < q \leq y,\ q \neq B_0\}.$$

For vectors of residue classes $\vec{a} = (a_s \bmod s)_{s\in\mathcal{S}}$, $\vec{n} = (n_p \bmod p)_{p\in\mathcal{P}}$, put

$$S(\vec{a}) := \{n \in \mathbb{Z} \mid n \not\equiv a_s \pmod s,\ \forall s \in \mathcal{S}\},$$

$$S(\vec{n}) := \{n \in \mathbb{Z} \mid n \not\equiv n_p \pmod p,\ \forall p \in \mathcal{P}\}.$$

Let $0 < \theta < 1$ be a parameter of Hypothesis 1 of [14]. In our situation, unconditionally one can take

$$\theta = \frac{1}{3}$$

(see [4], Section 8).

Next, for $r \in \mathbb{Z}_{\geq 2}$, we denote by $\mathcal{F}_r$ the set of square-integrable symmetric functions $F : \mathbb{R}^r \to \mathbb{R}$ supported in

$$\mathcal{R}_r := \{(x_1,\ldots,x_r) \in \mathbb{R}^r \mid x_1,\ \ldots,x_r \geq 0, x_1+\ldots+x_r \leq 1\}.$$

For $F \in \mathcal{F}_r$, put

$$I_r(F) := \int_0^\infty \cdots \int_0^\infty F(t_1,\ldots,t_r)^2 dt_1 \cdots dt_r,$$

$$J_r(F) := \int_0^\infty \cdots \int_0^\infty \left(\int_0^\infty F(t_1,\ldots,t_r)dt_r\right)^2 dt_1 \cdots dt_{r-1}.$$

We suppose that for some positive constant $c_{I,J}$, the inequality

$$\sup_{F\in\mathcal{F}_r} \frac{J_r(F)}{I_r(F)} \geq c_{I,J} \frac{\log r}{r}\left(1 + O\left(\frac{1}{\log r}\right)\right) \tag{2.3}$$

holds as $r \to \infty$. Maynard showed that one can take $c_{I,J} = \frac{1}{4}$ (see (8.27) of [14]).

3. The sieve of intervals and the explicit lower bound for $G_k(X)$

For $X > 1$, we denote by $[X]$ the set of integers in the interval $[1,X]$.

Proposition 3.1. *Let $A \geq 1$ be an arbitrarily fixed constant and x sufficiently large real number. Let y be a parameter defined by (2.1) with*

$$c = \frac{\theta c_{I,J}}{12800 \log 5}. \tag{3.1}$$

Let B_0 be either 1 or a prime number with $\log x \ll B_0 \le x$. Then, for any prime number $p \le x$ with $p \ne B_0$, there exists a residue class a_p (mod p) for which the set

$$\mathcal{T} := \{n \in [y]\backslash[x] \mid n \not\equiv a_p \ (\mathrm{mod}\ p),\ \forall p \le x,\ p \ne B_0\}$$

satisfies the following three conditions.

· *(Upper bound)*

$$\#\mathcal{T} \le 5A(1+o(1))\frac{x}{\log x}. \tag{3.2}$$

· *(Lower bound)*

$$\#\mathcal{T} \ge A(1+o(1))\frac{x}{\log x}. \tag{3.3}$$

· *(Upper bound in short intervals) For any fixed $0 \le \alpha < \beta \le 1$, we have*

$$\#(\mathcal{T} \cap (\alpha y, \beta y]) \le 5A(2|\beta-\alpha|+\varepsilon)(1+o(1))\frac{x}{\log x}. \tag{3.4}$$

We will prove this proposition in subsequent sections. We now give a lower bound for $G_k(X)$ as a consequence of this proposition. Let $Q \ge 100$. Then it is known that there exists a positive integer B_Q which is either 1 or a prime number with

$$B_Q \gg \log_2 Q \tag{3.5}$$

such that if a Dirichlet character χ with conductor less than Q and coprime to B_Q satisfies

$$L(\sigma+it,\chi) = 0 \quad (\sigma, t \in \mathbb{R}),$$

then

$$1-\sigma \ge \frac{c_{ZFR}}{\log(Q(1+|t|))} \tag{3.6}$$

holds for some absolute constant $c_{ZFR} > 0$. We introduce the following two assumptions on distribution of primes in arithmetic progressions.

Assumption 3.2 (Assumption PAP)**.** Suppose that all L-functions associated to Dirichlet characters χ modulo q do not have any zero in the region (3.6) (with Q replaced by q). Then, there exist absolute constants $0 < C_{PAP} \le 1$ and $D_{PAP} \ge 1$ for which for any positive integers a, q with $(a,q) = 1$ and $x \ge q^{D_{PAP}}$, one has

$$\#\{p : \text{prime} \mid p \le x,\ p \equiv a \ (\mathrm{mod}\ q)\} \ge C_{PAP}(1+o(1))\frac{x}{\varphi(q)\log x}. \tag{3.7}$$

Remark 3.3. Later we will show that the above assumption is valid unconditionally for $D_{PAP} = 160$, $C_{PAP} = 1-e^{-2}$.

Assumption 3.4 (Assumption UB)**.** There exist absolute constants $C_{UB} \geq 1$, $D_{UB} > 0$ for which the following holds. For any sufficiently large x and a positive integer B_0 which is either 1 or a prime number with

$$\log x \ll B_0 \leq x, \tag{3.8}$$

put $P := P(x)/B_0$, where $P(x) := \prod_{p \leq x} p$ denotes the product of all primes equal to or less than x. Then, for any $Z \geq P^{D_{UB}}$ and $a, b \in [P]$ with $a \neq b$, one has

$$\#\{z \in [Z] \mid Pz + a,\ Pz + b : \text{prime}\} \leq C_{UB}(1 + o(1)) \left(\frac{\log x}{\log Z}\right)^2 Z. \tag{3.9}$$

Here, the $o(1)$ term is independent of a and b.

Remark 3.5. Later we will show that one can unconditionally take $C_{UB} = 8e^{2\gamma}$ and $D_{UB} > 0$ arbitrarily.

With these notations, we have the following theorem.

Theorem 3.6 (An explicit version of the theorem of Ford, Maynard and Tao [6])**.** *Let k be a fixed positive integer. Let θ be a parameter of Maynard's sieve in Hypothesis 1 of [14], $c_{I,J}$ a constant satisfying (2.3), C_{PAP} and D_{PAP} constants in Assumption PAP and C_{UB}, D_{UB} constants in Assumption UB and put $M :=$ $\max\{D_{PAP}, D_{UB}\}$. Then we have*

$$G_k(X) \geq \frac{C_{PAP}^2 \theta c_{I,J} e^{4\gamma}}{737280000 \log 5\, C_{UB} M (1 + D_{PAP}^{-1})^4 (25 C_{UB} + 20 e^{\gamma} M)} \frac{1}{k^2} \frac{\log X \log_2 X \log_4 X}{\log_3 X} \tag{3.10}$$

for any sufficiently large X.

Lemma 3.7. Assume Assumptions PAP and UB. Let B_0 be a positive integer which is either 1 or a prime number satisfying (3.8). Put $P = P(x)/B_0$. Then, for any $Z \geq P^{D_{PAP}}$ and $a \in [P]$ with $(a, P) = 1$, we have

$$\#\{z \in [Z] \mid Pz + a : \text{prime}\} \geq \frac{e^{\gamma}}{1 + D_{PAP}^{-1}} (1 + o(1)) \frac{\log x}{\log Z} Z. \tag{3.11}$$

Here, the $o(1)$ term is independent of a.

Proof. We apply the Assumption PAP with a zero free region (3.6) with $Q = P(x) = \prod_{p \leq x} p$. In the Assumption UB, we set $B_0 = 1$ if $B_{P(x)} > x$ and otherwise put $B_0 = B_{P(x)}$. Note that in the latter case, due to the condition (3.5) and a consequence of the prime number theorem ($P(x) \sim e^{(1+o(1))x}$), it follows that

$$B_0 = B_{P(x)} \gg \log_2 P(x) = \log_2 \prod_{p \leq x} p \sim \log_2 e^x = \log x,$$

so (3.8) is satisfied. Therefore, by the condition (3.7), for $Z \geq P^{D_{PAP}}$, we have

$$\begin{aligned}
&\#\{z \in [Z] \mid Pz + a : \text{prime}\} \\
&= \#\{P + a \leq p \leq PZ + a \mid p \equiv a \ (\mathrm{mod}\ P),\ p : \text{prime}\} \\
&\geq C_{PAP}(1 + o(1)) \frac{PZ + a}{\varphi(P) \log(PZ + a)} - \#\{p < P + a \mid p \equiv a \ (\mathrm{mod}\ P),\ p : \text{prime}\}.
\end{aligned} \tag{3.12}$$

By Mertens' formula, we have

$$\frac{P}{\varphi(P)} = \frac{P}{P \prod_{p \leq x, p \neq B_0} \left(1 - \frac{1}{p}\right)} = \frac{1}{\prod_{p \leq x, p \neq B_0} \left(1 - \frac{1}{p}\right)} \sim e^{\gamma} \log x. \tag{3.13}$$

Hence the first term of the last line of (3.12) is asymptotically

$$\frac{PZ}{\varphi(P) \log(PZ)} \geq e^{\gamma} \log x \frac{Z}{\log\left(Z^{1 + \frac{1}{D_{PAP}}}\right)} = \frac{e^{\gamma}}{1 + D_{PAP}^{-1}} \frac{\log x}{\log Z} Z.$$

On the other hand, the second term of the last line of (3.12) is at most

$$\#\{p < P \mid p : \text{prime}\} \leq \frac{2P}{\log P} \leq 2D_{PAP} \frac{Z^{\frac{1}{D_{PAP}}}}{\log Z} = o\left(\frac{\log x}{\log Z} Z\right).$$

Combining these estimates, we obtain (3.11). □

Let us prove (3.10). We write $\mathbb{P}$ for probability and $\mathbb{E}$ for expectation. Put $Z = P^M$, where $M := \max\{D_{UB}, D_{PAP}\}$. Let $\mathbf{z}$ be a random variable in $[Z]$ which is chosen uniformly and y, $\mathcal{T}$ and $a_p \ (\mathrm{mod}\ p)$ are those in Proposition 3.1. By Chinese remainder theorem, there exists $m \in [Z]$ such that $m \equiv -a_p \ (\mathrm{mod}\ p)$ holds for any $p \leq x, p \neq B_0$. Therefore, the interval $\mathbf{z}P + m + \mathcal{T}$ exactly consists of elements of $\mathbf{z}P + m + [y] \setminus [x]$ which is coprime to P. In particular, all prime numbers in $\mathbf{z}P + m + [y] \setminus [x]$ are contained in $\mathbf{z}P + m + \mathcal{T}$. By (3.11) and $Z = P^M = e^{(1+o(1))Mx}$, we have

$$\begin{aligned}
\mathbb{P}(\mathbf{z}P + m + a : \text{prime}) &\geq \frac{C_{PAP} e^{\gamma}}{1 + D_{PAP}^{-1}} (1 + o(1)) \frac{\log x}{\log Z} \\
&= \frac{C_{PAP} e^{\gamma}}{(1 + D_{PAP}^{-1}) M} (1 + o(1)) \frac{\log x}{x}
\end{aligned} \tag{3.14}$$

for any $a \in \mathcal{T}$. On the other hand, by Brun-Titchmarsh type estimate

$$\pi(x; q, a) \leq \frac{2x}{\varphi(q) \log \frac{x}{q}} \quad (\forall x > q)$$

and Mertens type estimate (3.13), we have

$$
\begin{aligned}
\mathbb{P}(\mathbf{z}P+m+a:\text{prime}) &= \frac{1}{Z}\#\{z\in[Z] \mid Pz+m+a:\text{prime}\} \\
&\leq \frac{1}{Z}\#\{p\leq PZ+m+a \mid p\equiv m+a \pmod{P}\} \\
&\leq \frac{1}{Z}\#\{p\leq 2PZ \mid p\equiv m+a \pmod{P}\} \\
&\leq \frac{1}{Z}\frac{2\cdot 2PZ}{\varphi(P)\log\left(\frac{2PZ}{P}\right)} \\
&\leq \frac{4e^{\gamma}\log x}{\log(2Z)} \\
&\leq \frac{4e^{\gamma}\log x}{\log Z} \sim \frac{4e^{\gamma}}{M}\frac{\log x}{x}.
\end{aligned}
\tag{3.15}
$$

Furthermore, by Assumption UB,

$$
\begin{aligned}
\mathbb{P}(\mathbf{z}P+m+a,\ \mathbf{z}P+m+b:\text{prime}) &\leq C_{UB}(1+o(1))\left(\frac{\log x}{\log Z}\right)^2 \\
&= \frac{C_{UB}}{M^2}(1+o(1))\left(\frac{\log x}{x}\right)^2
\end{aligned}
\tag{3.16}
$$

for any $a\neq b\in\mathcal{T}$. Let $\mathbf{N}$ denote the number of primes in $\mathbf{z}P+m+\mathcal{T}$ (hence in $\mathbf{z}P+m+[y]\setminus[x]$). By (3.14) and (3.3), we have

$$
\mathbb{E}\mathbf{N} = \sum_{a\in\mathcal{T}}\mathbb{P}(\mathbf{z}P+m+a:\text{prime}) \geq (1+o(1))\frac{C_{PAP}e^{\gamma}}{(1+D_{PAP}^{-1})M}A. \tag{3.17}
$$

On the other hand, by (3.15), (3.16) and (3.2),

$$
\begin{aligned}
\mathbb{E}\mathbf{N}^2 &= \sum_{a,b\in\mathcal{T}}\mathbb{P}(\mathbf{z}P+m+a,\mathbf{z}P+m+b:\text{prime}) \\
&\leq \frac{C_{UB}}{M^2}(1+o(1))\left(\frac{\log x}{x}\right)^2(\#\mathcal{T})^2 + \frac{4e^{\gamma}}{M}(1+o(1))\frac{\log x}{x}\#\mathcal{T} \\
&\leq \left(\frac{25A^2C_{UB}}{M^2}+\frac{20Ae^{\gamma}}{M}\right)(1+o(1)) \\
&\leq \frac{25C_{UB}+20e^{\gamma}M}{M^2}(1+o(1))A^2.
\end{aligned}
\tag{3.18}
$$

(The first term of the second line is an upper bound for the contribution of the terms with $a\neq b$, and the second term is that of the terms with $a=b$. In the last line, we used $A\geq 1$.) Suppose that $0\leq\alpha<\beta\leq 1$ satisfy $\beta-\alpha\leq 2\varepsilon$ for $0<\varepsilon<$

1. Then by (3.16) and (3.4), the probability that the interval $\mathbf{z}P+m+[\alpha y,\beta y]$ contains at least two primes is at most

$$\left(6A(4\varepsilon+\varepsilon)\frac{x}{\log x}\right)^2\frac{C_{UB}}{M^2}\left(\frac{\log x}{x}\right)^2\leq\frac{900\varepsilon^2A^2C_{UB}}{M^2}.$$

We cover the interval $[0,y]$ by intervals

$$I_1=[0,2\varepsilon y],\, I_2=[\varepsilon y,3\varepsilon y],\,\ldots,\, I_{n_\varepsilon}=[(n_\varepsilon-1)\varepsilon y,y],$$

where $n_\varepsilon=\lceil\frac{1}{\varepsilon}\rceil$. Then two elements $a,b\in[0,y]$ with $|a-b|\leq\varepsilon y$ are contained in the same interval I_j for some $1\leq j\leq n_\varepsilon$. Therefore, the probability that "The interval $\mathbf{z}P+m+[y]\backslash[x]$ contains at least two primes with gap at most εy" is equal to or less than the probability that "some of the n_ε intervals

$$\mathbf{z}P+m+[0,2\varepsilon y],\,\mathbf{z}P+m+[\varepsilon y,3\varepsilon y],\,\ldots,\,\mathbf{z}P+m+[(n_\varepsilon-1)\varepsilon y,y]$$

contains at least two primes", and by the above consideration, this probability is at most

$$n_\varepsilon\frac{900\varepsilon^2A^2C_{UB}}{M^2}<\frac{1800\varepsilon A^2C_{UB}}{M^2},$$

since $n_\varepsilon=\lceil\frac{1}{\varepsilon}\rceil<\frac{2}{\varepsilon}$. We take

$$\varepsilon=\frac{M^2}{1800A^2C_{UB}}\frac{e^{2\gamma}}{4(1+D_{PAP}^{-1})^2(25C_{UB}+20e^{\gamma}M)}. \tag{3.19}$$

(Later we will see that $\varepsilon<1$ with an appropriate choice of A.) Then the above probability is less than

$$\frac{e^{2\gamma}}{4(1+D_{PAP}^{-1})^2(25C_{UB}+20e^{\gamma}M)}.$$

In other words, all primes in $\mathbf{z}P+m+[y]\backslash[x]$ are separated more than εy to each other with probability greater than

$$1-\frac{e^{2\gamma}}{4(1+D_{PAP}^{-1})^2(25C_{UB}+20e^{\gamma}M)}. \tag{3.20}$$

On the other hand, for any fixed $c_1>0$, by Cauchy-Schwarz inequality, we have

$$\mathbb{E}\mathbf{N}\leq c_1A\mathbb{P}(\mathbf{N}\leq c_1A)+\mathbb{E}\mathbf{N}1_{\mathbf{N}>c_1A}\leq c_1A+(\mathbb{E}\mathbf{N}^2)^{\frac{1}{2}}\mathbb{P}(\mathbf{N}>c_1A)^{\frac{1}{2}}.$$

Therefore, by (3.17) and (3.18), we have

$$\mathbb{P}(\mathbf{N} > c_1 A) \geq \frac{(\mathbb{E}\mathbf{N} - c_1 A)^2}{\mathbb{E}\mathbf{N}^2} \geq \frac{\left(\frac{C_{PAP}e^{\gamma}A}{(1+D_{PAP}^{-1})M} - c_1 A\right)^2}{\frac{(25C_{UB}+20e^{\gamma}M)A^2}{M^2}} = \frac{\left(\frac{C_{PAP}e^{\gamma}}{1+D_{PAP}^{-1}} - c_1 M\right)^2}{25C_{UB}+20e^{\gamma}M}, \tag{3.21}$$

provided that $c_1 < \frac{C_{PAP}e^{\gamma}}{(1+D_{PAP}^{-1})M}$. We take

$$c_1 = \frac{C_{PAP}e^{\gamma}}{2(1+D_{PAP}^{-1})M}.$$

Then,

$$\mathbb{P}(\mathbf{N} > c_1 A) \geq \frac{e^{2\gamma}}{4(1+D_{PAP}^{-1})^2(25C_{UB}+20e^{\gamma}M)}. \tag{3.22}$$

Since the sum of (3.20) and the right hand side of (3.22) is equal to 1, it follows that there exists some integer $z \in [Z]$ such that the interval $zP+m+[y]\backslash[x]$ contains at least

$$c_1 A = \frac{e^{\gamma}A}{2M(1+D_{PAP}^{-1})}$$

primes and all of them are separated at least

$$\varepsilon y = \frac{yM^2}{1800A^2C_{UB}} \frac{e^{2\gamma}}{4(1+D_{PAP}^{-1})^2(25C_{UB}+20e^{\gamma}M)}$$

to each other, where

$$y = c\frac{x\log x \log_3 x}{\log_2 x}, \quad c = \frac{\theta c_{I,J}}{12800\log 5}.$$

We take

$$A = 2e^{-\gamma}C_{PAP}^{-1}M(1+D_{PAP}^{-1})k. \tag{3.23}$$

Then, the interval $zP+m+[y]\backslash[x]$ contains k consecutive primes which are separated at least εy to each other. Therefore,

$$\begin{aligned} G_k(ZP+m+y) &\geq \varepsilon y \\ &= \frac{\theta c_{I,J}}{12800\log 5}\frac{M^2}{1800A^2C_{UB}}\frac{e^{2\gamma}}{4(1+D_{PAP}^{-1})^2(25C_{UB}+20e^{\gamma}M)}\frac{x\log x\log_3 x}{\log_2 x}. \end{aligned}$$

Since

$$ZP+m+y \leq 2ZP = 2P^{M+1} = 2e^{(1+o(1))(M+1)x} \leq e^{1.5Mx},$$

by putting $X = e^{1.5Mx}$, we have

$$
\begin{aligned}
G_k(X) &\geq \frac{\theta c_{I,J}}{12800\log 5}\frac{M^2}{1800A^2C_{UB}}\frac{e^{2\gamma}(1+o(1))}{4(1+D_{PAP}^{-1})^2(25C_{UB}+20e^{\gamma}M)}\frac{2}{3M}\frac{\log X\log_2 X\log_4 X}{\log_3 X}\\
&\geq \frac{\theta c_{I,J}}{12800\log 5}\frac{M^2}{1800A^2C_{UB}}\frac{e^{2\gamma}}{4(1+D_{PAP}^{-1})^2(25C_{UB}+20e^{\gamma}M)}\frac{1}{2M}\frac{\log X\log_2 X\log_4 X}{\log_3 X}.
\end{aligned}
$$

Finally, by substituting (3.23), we obtain (3.10). □

4. Possible values of C_{UB} and D_{UB} in Assumption UB

To establish the upper bound (3.9) in Assumption UB, we use the following theorem.

Theorem 4.1 ([9], Theorem 5.7). *For $g \in \mathbb{N}$ and integers a_i, b_i $(i = 1,\ldots,g)$, suppose*

$$
E := \prod_{i=1}^{g} a_i \prod_{1\leq r<s\leq g}(a_r b_s - a_s b_r) \neq 0.
$$

For any prime number p, let $\rho(p)$ denote the number of solutions n (mod p) of the equation

$$
\prod_{i=1}^{g}(a_i n + b_i) \equiv 0 \pmod{p},
$$

and suppose that $\rho(p) < p$ holds for any primes p. Then, for any $1 < y \leq x$, we have

$$
\begin{aligned}
&\#\{n \mid x-y < n \leq x, a_i n + b_i : \text{prime for } i = 1,\ldots,g\}\\
&\quad \leq 2^g g! \prod_p \left(1 - \frac{\rho(p)-1}{p-1}\right)\left(1-\frac{1}{p}\right)^{-g+1}\frac{y}{\log^g y}\left(1 + O((\log y)^{-1}(\log_2 3y + \log_2 3|E|))\right).
\end{aligned}
$$

We apply this theorem with $x = y = Z$, $g = 2$, $a_1 = a_2 = P$, and $b_1 = a$, $b_2 = b$. Then, $E = P^3(a-b) \neq 0$ and

$$
|E| \ll P^3 Z \ll e^{3(1+o(1))x} Z.
$$

Note that since $a, b \in \mathcal{T} \subset [y]\backslash[x]$, we have $|a-b| \leq y = o(x\log x)$. We will use this later (see (4.5) below).

We need to compute $\rho(p)$. First, for $p \leq x$ with $p \neq B_0$, if either $p|a$ or $p|b$ holds, then

$$
Pn + a \equiv 0 \pmod{p} \quad \text{or} \quad Pn + b \equiv 0 \pmod{p}
$$

holds for any $n \in \mathbb{N}$, so

$$\#\{z \in [Z] \mid Pz+a, Pz+b : \text{prime}\} = 0.$$

Therefore, it suffices to assume that $p \nmid ab$ $(\forall p \leq x)$ or $p = B_0$ (in case of B_0: prime).
I) If $p \leq x$, $p \neq B_0$, then the condition

$$(Pn+a)(Pn+b) \equiv 0 \pmod p$$

is equivalent to

$$ab \equiv 0 \pmod p$$

which does not hold due to our assumption. Hence $\rho(p) = 0$.
II) If B_0 is a prime number and $p = B_0$, then the equation

$$(Pn+a)(Pn+b) \equiv 0 \pmod p$$

has two solutions. Hence $\rho(B_0) = 2$.
III) If $p > x$, then the number of the solutions $n \pmod p$ of the equation

$$(Pn+a)(Pn+b) \equiv 0 \pmod p$$

is 1 if $a \equiv b \pmod p$, and 2 if $a \not\equiv b \pmod p$.
Consequently,

$$\begin{aligned}
&\#\{z \in [Z] \mid Pz+a, Pz+b : \text{prime}\} \\
&\quad \leq 2^2 \cdot 2! \prod_{\substack{p \leq x \\ p \neq B_0}} \left(1+\frac{1}{p-1}\right)\left(1-\frac{1}{p}\right)^{-1} \left[\left(1-\frac{1}{B_0-1}\right)\left(1-\frac{1}{B_0}\right)^{-1}\right] \\
&\qquad \times \prod_{\substack{p > x \\ p \nmid (a-b)}} \left(1-\frac{1}{p-1}\right)\left(1-\frac{1}{p}\right)^{-1} \prod_{\substack{p > x \\ p \mid (a-b)}} \left(1-\frac{1}{p}\right)^{-1} \\
&\qquad \times \frac{Z}{\log^2 Z}\left(1+O\left(\frac{\log_2 Z + \log x}{\log Z}\right)\right).
\end{aligned} \tag{4.1}$$

Here, the factor $[\cdots]$ in the second line appears only when B_0 is prime. Now,

$$\prod_{\substack{p \leq x \\ p \neq B_0}} \left(1+\frac{1}{p-1}\right)\left(1-\frac{1}{p}\right)^{-1} \leq \prod_{p \leq x} \left(1-\frac{1}{p}\right)^{-2} \sim e^{2\gamma} \log^2 x, \tag{4.2}$$

$$\left(1-\frac{1}{B_0-1}\right)\left(1-\frac{1}{B_0}\right)^{-1} \sim 1, \tag{4.3}$$

$$\prod_{\substack{p>x \\ p\nmid(a-b)}} \left(1-\frac{1}{p-1}\right)\left(1-\frac{1}{p}\right)^{-1} \leq 1, \tag{4.4}$$

and

$$\prod_{\substack{p>x \\ p\mid(a-b)}} \left(1-\frac{1}{p}\right)^{-1} = \exp\left(-\sum_{\substack{p>x \\ p\mid(a-b)}} \log\left(1-\frac{1}{p}\right)\right) \leq \exp\left(\sum_{\substack{p>x \\ p\mid(a-b)}} \frac{2}{p}\right) \leq \exp\left(\frac{2\omega(|a-b|)}{x}\right) \sim 1, \tag{4.5}$$

where $\omega(n)$ denotes the number of distinct prime factors of n. Substituting (4.2)-(4.5) into (4.1), we have

$$\begin{aligned} \#\{z \in [Z] \mid Pz+a, Pz+b : \text{prime}\} &\leq 8e^{2\gamma}\log^2 x \frac{Z}{\log^2 Z}(1+o(1)) \\ &= 8e^{2\gamma}(1+o(1))\left(\frac{\log x}{\log Z}\right)^2 Z, \end{aligned} \tag{4.6}$$

provided that $\log x = o(\log Z)$. Since $Z \geq P^{D_{UB}} = e^{(1+o(1))D_{UB}x}$, this condition is valid for any $D_{UB} > 0$. Consequently, the Assumption UB is valid for

$$C_{UB} = 8e^{2\gamma} \tag{4.7}$$

and arbitrary $D_{UB} > 0$.

5. Possible values of C_{PAP} and D_{PAP} in Assumption PAP

To establish an inequality in Assumption PAP, we need some results on zero-free regions of Dirichlet L-functions. We use the following results by McCurley [11]. For a positive integer q, let $\mathcal{L}_q(s)$ be the product of $\varphi(q)$ L-functions associated to Dirichlet characters modulo q. Write $s = \sigma + it$, where $\sigma, t \in \mathbb{R}$.

Proposition 5.1 ([11], Theorem 1)**.** *Let $M := \max\{q, q|t|, 10\}$, $R = 9.6459\ldots$. Then $\mathcal{L}_q(s)$ has at most one zero in*

$$\left\{ s = \sigma + it \mid \sigma \geq 1 - \frac{1}{R\log M} \right\}.$$

The only possible zero is a real zero of the L-function associated to a non-principal real character modulo q.

Proposition 5.2 ([11], Theorem 2)**.** *For $i = 1,2$, let χ_i be a real character modulo q_i such that $\chi_1 \neq \chi_2$ and β_i be a real zero of $L(s,\chi_i)$. Put*

$$M_1 = \max\left\{\frac{q_1 q_2}{17}, 13\right\}, \quad R_1 = \frac{5-\sqrt{5}}{15-10\sqrt{2}}.$$

Then we have

$$\min\{\beta_1,\beta_2\} < 1 - \frac{1}{R_1 \log M_1}.$$

For $Q \geq 100$ and $q_1, q_2 \leq Q$, let χ_1, χ_2 be distinct primitive characters modulo q_1, q_2 respectively. Suppose that for $j = 1,2$, $L(s,\chi_j)$ has a zero $\rho_j = \beta_j + i\gamma_j$ in the domain

$$\mathcal{R} : \left\{ s = \sigma + it \mid \sigma \geq 1 - \frac{1}{4R_1 \log(Q(1+|t|))} \right\}.$$

Then for $M_{(j)} := \max\{q_j, q_j|t|, 10\}$ $(j = 1,2)$,

$$1 - \frac{1}{R \log M_{(j)}} \leq 1 - \frac{1}{4R_1 \log(Q(1+|t|))}$$

holds. Hence by Proposition 5.1, ρ_j is a real zero of $L(s,\chi_j)$ for $j = 1,2$. By Proposition 5.2, we have

$$\min\{\beta_1,\beta_2\} \leq 1 - \frac{1}{R_1 \log M_1}, \tag{5.1}$$

where $M_1 = \max\{\frac{q_1 q_2}{17}, 13\}$. However, since

$$1 - \frac{1}{R_1 \log M_1} \leq 1 - \frac{1}{4R_1 \log Q},$$

the inequality (5.1) contradicts our assumption that both $\rho_1 = \beta_1$ and $\rho_2 = \beta_2$ lie in $\mathcal{R}$. Combining this and an argument in the proof of Corollary 1 of [6], we have the following conclusion.

Proposition 5.3. *Put*

$$c_{ZFR} := \frac{1}{24} \left(< \frac{1}{4R_1} = \frac{15 - 10\sqrt{2}}{4(5-\sqrt{5})} \right).$$

Let $Q \geq 100$ be a fixed positive integer. Suppose that $L(s,\chi) = 0$ holds for some primitive Dirichlet character modulo at most Q. Then either

$$1 - \sigma \geq \frac{c_{ZFR}}{\log(Q(1+|t|))} \quad (s = \sigma + it, \sigma, t \in \mathbb{R})$$

or $t = 0$ and $\chi = \chi_Q$ is a real character which is determined uniquely by Q.

Gallagher [7] proved that there exist some absolute constants $a,b>0$ for which

$$\sum_{q\le Q}\sum_{\chi\,(\mathrm{mod}\,q)}^{*}\left|\sum_{p=x}^{x+h}\chi(p)\log p\right|\ll h\exp\left(-a\frac{\log x}{\log Q}\right) \tag{5.2}$$

holds uniformly for $x/Q\le h\le x$, $\exp(x^{\frac{1}{2}}\log x)\le Q\le x^{b}$. We need to make clear the value of a. Due to the above consideration, the L-functions associated to primitive characters modulo at most T do not have zero in

$$\sigma>1-\frac{3c_{ZFR}}{\log T},\quad |t|\le T$$

with at most one exception, and if the exceptional zero exists, then it is a simple real zero. Hence one can take $c_1=3c_{ZFR}$ in [7], (26). The estimate of [7], (30) yields

$$\sum_{q\le Q}\sum_{\chi\,(\mathrm{mod}\,q)}^{*}\left|\sum_{p=x}^{x+h}\chi(p)\log p\right|\ll h\left(\sum_{q\le Q}\sum_{\chi\,(\mathrm{mod}\,q)}^{*}\sum_{\beta}x^{\beta-1}+\frac{Q^4}{T}\right), \tag{5.3}$$

where β runs over zeros of non-trivial zeros of $L(s,\chi)$ in $\sigma\le 1-3c_{ZFR}/\log T$, $|t|\le T$. Due to the argument of [7], we have

$$\sum_{q\le Q}\sum_{\chi\,(\mathrm{mod}\,q)}^{*}\sum_{\beta}x^{\beta-1}\ll x^{-\frac{1}{2}\frac{3c_{ZFR}}{\log T}}.$$

Therefore, by putting $T=Q^5$, the right hand side of (5.3) is at most

$$\exp\left(-\frac{3}{10}c_{ZFR}\frac{\log x}{\log Q}\right).$$

Therefore, one can take

$$a=\frac{3}{10}c_{ZFR}$$

in (5.2).

Next, by Theorem 6 of [7], for some constant $c_{ZD}>0$,

$$\sum_{q\le T}\sum_{\chi\,(\mathrm{mod}\,q)}^{*}N(\alpha,T,\chi)\ll T^{c_{ZD}(1-\alpha)} \tag{5.4}$$

holds uniformly for $0\le\alpha\le 1$. Here, $N(\alpha,T,\chi)$ denotes the number of zeros of $L(s,\chi)$ in the domain $\alpha\le\Re(s)\le 1$, $|\Im(s)|\le T$. With this c_{ZD}, Theorem 7 of [7] is valid for $T^{c_{ZD}}\le x^{\frac{1}{2}}$, $T=Q^5$. Therefore, for $\exp(\log^{\frac{1}{2}}x)\le Q\le x^{\frac{1}{10c_{ZD}}}$, (5.2) holds for $x/Q \le h \le x$. Therefore, the proof of Lemma 2 of Maier [12] is valid for $x \ge Q^{10c_{ZD}}$.

In [12], the contribution of the terms with $\chi = \chi_0$ to $\psi(x;q,a)$ is $(1+o(1))x/\varphi(q)$, whereas that of the terms with $\chi \ne \chi_0$ is $O_{\le}(xe^{-aD}/\varphi(q))$. Here, $f(x) = O_{\le}(g(x))$ means that $|f(x)| \le g(x)$ holds for any sufficiently large x. Therefore, we have

$$\psi(x;q,a) \ge (1-e^{-aD})(1+o(1))\frac{x}{\varphi(q)}$$

provided that $D > 0$. Hence one can take $D_{PAP} = 10c_{ZD}$, $C_{PAP} = 1 - e^{-aD_{PAP}}$.

Next, we evaluate c_{ZD} above. Put

$$N^*(\alpha,T,Q) := \sum_{q \le Q} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} N(\alpha,T,\chi).$$

By Theorem 1 of Jutila [10], for any fixed $\varepsilon > 0$,

$$N^*(\alpha,T,Q) \ll_\varepsilon (Q^2T)^{(2+\varepsilon)(1-\alpha)}$$

holds uniformly for $\frac{4}{5} \le \alpha \le 1$, $T \ge 1$. Put $Q = T$. Then it follows that

$$\sum_{q \le T} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} N(\alpha,T,\chi) \ll_\varepsilon T^{(6+\varepsilon)(1-\alpha)}. \tag{5.5}$$

On the other hand, if $0 \le \alpha \le \frac{4}{5}$, by using the trivial estimate

$$N(\alpha,T,\chi) \ll T\log T,$$

which holds uniformly for $q \le T$, we have

$$\sum_{q \le T} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} N(\alpha,T,\chi) \ll \sum_{q \le T} qT\log T \ll T^3\log T \ll T^{(15+\varepsilon)(1-\alpha)}. \tag{5.6}$$

By (5.5) and (5.6),

$$\sum_{q \le T} \sideset{}{^*}\sum_{\chi \,(\mathrm{mod}\, q)} N(\alpha,T,\chi) \ll_\varepsilon T^{(15+\varepsilon)(1-\alpha)}$$

holds uniformly for $0 \le \alpha \le 1$. Therefore, one can take $c_{ZD} = 16$. Consequently, one can take

$$C_{PAP} = 1 - e^{-2}, \quad D_{PAP} = 160 \tag{5.7}$$

as the values of C_{PAP} and D_{PAP}.

6. Proof of Proposition 3.1

Proposition 3.1 can be derived from the following proposition.

Proposition 6.1 (An explicit version of Theorem 3 of [6])**.** *Let $A \geq 1$ be an arbitrary real number and x be a sufficiently large real number and B_0 be a positive integer. Put*

$$y = c\frac{x\log x\log_3 x}{\log_2 x}, \quad c = \frac{\theta c_{I,J}}{12800\log 5}$$

and let $S(\vec{a})$, $S(\vec{n})$ be the subsets of $\mathbb{Z}$ defined in Section 2. Then, there exist $A' = A'(x)$ and $c' = c'(x)$ satisfying

$$A' \sim c'A, \quad 1 \leq c' \leq 5$$

and choices of vectors $\vec{\mathbf{a}} = (\mathbf{a}_s \bmod s)_{s\in\mathcal{S}}$, $\vec{\mathbf{n}} = (\mathbf{n}_p \bmod p)_{p\in\mathcal{P}}$ such that for arbitrarily fixed $0 \leq \alpha < \beta \leq 1$,

$$\#(\mathcal{Q}\cap S(\vec{\mathbf{a}})\cap S(\vec{\mathbf{n}})\cap(\alpha y, \beta y]) \sim A'|\beta-\alpha|\frac{x}{\log x} \tag{6.1}$$

holds with probability $1-o(1)$. The implied constant of $o(1)$ in the probability may depend on A, α, β, but that of (6.1) doesn't.

(*Proof of Proposition 6.1 ⇒ Proposition 3.1*) Let $0 < \varepsilon < 1$. We decompose the interval $(0,1]$ into $O(\varepsilon^{-1})$ disjoint intervals $(\alpha_i, \beta_i]_{i\in I}$ of lengths between $\varepsilon/2$ and ε. We apply Proposition 6.1 with $(\alpha,\beta) = (\alpha_i,\beta_i)$ ($i \in I$) and $(\alpha,\beta) = (0,1)$. Then there exist A', c' with

$$A' \sim c'A, \quad 1 \leq c' \leq 5$$

and some vectors of residue classes $\vec{a} = (a_s \bmod s)_{s\in\mathcal{S}}$, $\vec{n} = (n_p \bmod p)_{p\in\mathcal{P}}$ for which

$$\#(\mathcal{Q}\cap S(\vec{a})\cap S(\vec{n})) \sim A'\frac{x}{\log x} \tag{6.2}$$

and

$$\#(\mathcal{Q}\cap S(\vec{a})\cap S(\vec{n})\cap(\alpha_i y, \beta_i y]) \leq A'\varepsilon(1+o(1))\frac{x}{\log x} \tag{6.3}$$

hold. For any real numbers $0 \leq \alpha < \beta \leq 1$, the interval $(\alpha y, \beta y]$ can be covered by at most $\lfloor\frac{2}{\varepsilon}\rfloor|\beta-\alpha|+1$ intervals $(\alpha_i y, \beta_i y]$. Hence by (6.3) we have

$$\begin{aligned}\#(\mathcal{Q}\cap S(\vec{a})\cap S(\vec{n})\cap(\alpha y, \beta y]) &\leq A'\varepsilon\left(\left\lfloor\frac{2}{\varepsilon}\right\rfloor|\beta-\alpha|+1\right)(1+o(1))\frac{x}{\log x} \\ &\leq A'(2|\beta-\alpha|+\varepsilon)(1+o(1))\frac{x}{\log x}.\end{aligned} \tag{6.4}$$

We extend $\vec{a}$ to $(a_p)_{p\le x}$ by

$$a_p := \begin{cases} n_p & (p \in \mathcal{P}) \\ 0 & (p \notin \mathcal{S} \cup \mathcal{P}), \end{cases}$$

and set

$$\mathcal{T} := \{n \in [y] \backslash [x] \mid n \not\equiv a_p \ (\mathrm{mod}\ p), \forall p \le x, p \ne B_0\}.$$

Following the argument of Section 4 of [6], we see that the set $\mathcal{T}$ only differs from $\mathcal{Q} \cap S(\vec{a}) \cap S(\vec{n})$ by a set $\mathcal{R}$ consisting of z-smooth numbers in $[y]$ multiplied by powers of B_0, and that $\mathcal{R} = o(x/\log x)$. Consequently we obtain Proposition 3.1. □

Hence to prove the main theorem, it suffices to prove Proposition 6.1. This proposition can be derived from the following two propositions.

Proposition 6.2 (An explicit version of Theorem 5 of [6]). *Let x be a sufficiently large real number and put*

$$y = c\frac{x\log x \log_3 x}{\log_2 x}, \quad c = \frac{\theta c_{I,J}}{12800\log 5}.$$

Then, there exist a quantity $C = C(x)$ satisfying

$$\frac{\theta c_{I,J}}{9600c}(1+o(1)) \le C \le \frac{\theta c_{I,J}}{4800c}(1+o(1)) \tag{6.5}$$

and choices of random vectors $\vec{\mathbf{a}} = (\mathbf{a}_s \bmod s)_{s\in\mathcal{S}}$, $\vec{\mathbf{n}} = (\mathbf{n}_p \bmod p)_{p\in\mathcal{P}}$, for which the following three conditions hold.
· *For any vector $\vec{a}$ in the essential range of $\vec{\mathbf{a}}$,*

$$\mathbb{P}(q \equiv \mathbf{n}_p \ (\mathrm{mod}\ p) \mid \vec{\mathbf{a}} = \vec{a}) \le x^{-\frac{1}{2}-\frac{1}{10}} \tag{6.6}$$

holds uniformly for all $p \in \mathcal{P}$.
· *For any fixed $0 \le \alpha < \beta \le 1$,*

$$\#(\mathcal{Q} \cap S(\vec{\mathbf{a}}) \cap (\alpha y, \beta y]) \sim 80c|\beta-\alpha|\frac{x}{\log x}\log_2 x \tag{6.7}$$

holds with probability $1-o(1)$.
· *We say that a vector $\vec{a}$ in the essential range of $\vec{\mathbf{a}}$ is "good" if*

$$\sum_{p\in\mathcal{P}} \mathbb{P}(q \equiv \mathbf{n}_p \ (\mathrm{mod}\ p) \mid \vec{\mathbf{a}} = \vec{a}) = C + O_{\le}\left(\frac{1}{(\log_2 x)^2}\right) \tag{6.8}$$

holds for any $q \in \mathcal{Q} \cap S(\vec{\mathbf{a}})$ with at most $x/\log x \log_2 x$ exceptionals. Then $\vec{\mathbf{a}}$ is "good" with probability $1-o(1)$.

Proposition 6.3 ([6], Theorem 4)**.** *Let x be a sufficiently large real number, and $\mathcal{P}', \mathcal{Q}'$ sets of primes in $(\frac{x}{2}, x]$, $(x, \log x]$ with $\#\mathcal{Q}' > (\log_2 x)^3$, respectively. For each $p \in \mathcal{P}'$, let $\mathbf{e}_p$ be a random subset of $\mathcal{Q}'$ satisfying*

$$\#\mathbf{e}_p \le r = O\left(\frac{\log x \log_3 x}{\log_2^2 x}\right) \quad (p \in \mathcal{P}'). \tag{6.9}$$

Assume the following two conditions.
· For any $p \in \mathcal{P}'$, $q \in \mathcal{Q}'$,

$$\mathbb{P}(q \in \mathbf{e}_p) \le x^{-\frac{1}{2}-\frac{1}{10}}. \tag{6.10}$$

· For any $q \in \mathcal{Q}'$ with at most $\frac{1}{(\log_2 x)^2}\#\mathcal{Q}'$ exceptions, there exists some quantity C (which is independent of q) with

$$\frac{5}{4}\log 5 \le C \ll 1 \tag{6.11}$$

such that

$$\sum_{p \in \mathcal{P}'} \mathbb{P}(q \in \mathbf{e}_p) = C + O_{\le}\left(\frac{1}{(\log_2 x)^2}\right) \tag{6.12}$$

holds.
Then, for any positive integer m with

$$m \le \frac{\log_3 x}{\log 5}, \tag{6.13}$$

there exist random subsets $\mathbf{e}'_p \subset \mathcal{Q}'$ for each $p \in \mathcal{P}$ such that

$$\#\{q \in \mathcal{Q}' \mid q \notin \mathbf{e}'_p, \forall p \in \mathcal{P}'\} \sim 5^{-m}\#\mathcal{Q}'$$

holds with probability $1 - o(1)$. More generally, for any subset $\mathcal{Q}'' \subset \mathcal{Q}'$ with $\#\mathcal{Q}'' \ge (\#\mathcal{Q}')/\sqrt{\log_2 x}$,

$$\#\{q \in \mathcal{Q}'' \mid q \notin \mathbf{e}'_p, \forall p \in \mathcal{P}'\} \sim 5^{-m}\#\mathcal{Q}''$$

holds with probability $1 - o(1)$.

(*Proof of Proposition 6.2 $\wedge$ Proposition 6.3 $\Rightarrow$ Proposition 6.1*) Let m be a positive integer with

$$m \le \frac{\log_3 x}{\log 5}, \tag{6.14}$$

and put $A' = 5^{-m} 80c \log_2 x$. If

$$m = \left\lfloor \frac{\log\left(\frac{80c\log_2 x}{A}\right)}{\log 5} \right\rfloor,$$

then $A \leq A' \leq 5A$ holds. For this m satisfy (6.14), we need

$$1 \ll \frac{80c}{A} \leq 1.$$

Write

$$c = \frac{A}{80k}\varepsilon_0, \quad \varepsilon_0 > 0.$$

Then by (6.5), it follows that

$$\frac{k\theta c_{I,J}}{120A\varepsilon_0}(1+o(1)) \leq C \leq \frac{k\theta c_{I,J}}{60A\varepsilon_0}(1+o(1)).$$

This C must satisfy the condition (6.11) in Proposition 6.3. Hence the ε_0 must satisfy

$$1 \ll \varepsilon_0 \leq \frac{k\theta c_{I,J}}{150A\log 5}(1+o(1)).$$

We take $\varepsilon_0 = \frac{k\theta c_{I,J}}{160A\log 5}$. Then

$$c = \frac{k\theta c_{I,J}}{12800\log 5}.$$

Following the argument in Section 5 of [6], we see that the statement of Proposition 6.1 is valid with this c.

With the above discussion, all that remains is to prove Proposition 6.2. In practice, however, we only need to check how the remaining coefficients are determined following the arguments in [4] and [6].

(*Proof of Proposition 6.2*) Proposition 6.2 can be obtained by applying the sieve of Maynard (Theorem 6 of [6]. For the proof, see Section 6 of [4]). In particular, the constant c_0 in (6.1) of [6] can be chosen as $c_0 = \frac{1}{5}$, following the argument of Maynard [14]. The quantity C in Theorem 5 of [6] is computed in the following way. Put

$$\sigma = \prod_{s\in\mathcal{S}}\left(1-\frac{1}{s}\right), \quad y = c\frac{x\log x\log_3 x}{\log_2 x}.$$

By Mertens formula, we have

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

Following (8.3) of [4], we take

$$u = \frac{\varphi(B)}{B} \frac{\log R}{\log x} \frac{rJ_r}{2I_r},$$

where B is either 1 or prime, so

$$\frac{1}{2} \leq \frac{\varphi(B)}{B} \leq 1.$$

Following Theorem 6 of [4], we take $R = x^{\frac{\theta}{3}}$, where $\theta > 0$ is a constant which satisfies the condition in Definition 2 of [4]. (As mentioned above, one can take $\theta = \frac{1}{3}$.) Suppose

$$J_r = c_{I,J}(1+o(1))\frac{\log r}{r} I_r$$

holds as $r \to \infty$. Then

$$\frac{\theta}{12} c_{I,J}(1+o(1))\log r \leq u \leq \frac{\theta}{6} c_{I,J}(1+o(1))\log r.$$

The parameter r above can be chosen as

$$r = \lfloor \log^{c_0} x \rfloor = \left\lfloor \log^{\frac{1}{5}} x \right\rfloor.$$

Hence

$$\frac{\theta}{60} c_{I,J}(1+o(1))\log_2 x \leq u \leq \frac{\theta}{30} c_{I,J}(1+o(1))\log_2 x.$$

Finally, following [6], one can take

$$C = \frac{u}{\sigma} \frac{x}{2y}.$$

Hence due to the above estimate of u, the statement of Proposition 6.1 holds if the parameter satisfies the condition (6.5).

7. Acknowledgements

This work is partially supported by the JSPS, KAKENHI Grant Number 24K06697. The author would like to thank Professor Terence Tao for telling him an idea to obtain a lower bound (3.21) in his blog. He would also like to express his sincere gratitude to the reviewer for his/her careful reading of this paper and many helpful suggestions.

REFERENCES

[1] H. Cramér, *On the order of magnitude of the difference between consecutive prime numbers*, Acta Arith. **2** (1936), 396–403.

[2] P. Erdős, *On the difference of consecutive primes*, Quart. J. Math. Oxford Ser. **6** (1935), 124–128.

[3] P. Erdős, *Some of my favourite unsolved problems, in A Tribute to Paul Erdős (A. Baker, B. Bollobás, A. Hajnal, eds.)*, Cambridge Univ. Press, 1990, pp. 467–478.

[4] K. Ford, B. Green, S. Konyagin, J. Maynard, T. Tao, *Long gaps between primes*, J. Amer. Math. Soc. **31** (1): 65–105 (2018).

[5] K. Ford, B. Green, S. Konyagin, T. Tao, *Large gaps between consecutive prime numbers*, Ann. of Math. **183** (2016), Issue 3, 935-974.

[6] K. Ford, J. Maynard. T. Tao, *Chains of large gaps between primes*, Irregularities in the Distribution of Prime Numbers pp. 1–21 (2018) arXiv:1511.04468.

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

[8] A. Granville, *Harald Cramér and the distribution of prime numbers*, Scandinavian Actuarial J. **1** (1995), 12–28.

[9] H. Halberstam, H. E. Richert, *Sieve Methods*, London Mathematical Society Monographs. Vol. 4. Academic Press (1974, by Dover in 2011).

[10] M. Jutila, *On Linnik's constant*, Mathematica Scandinavica Vol. 41, No. 1 (1977), 45-62.

[11] K. S. McCurley, *Explicit zero-free regions for Dirichlet L-functions*, J. Number Theory **19**, 7-32 (1984).

[12] H. Maier, *Chains of large gaps between consecutive primes*. Adv. Math. **39**, 257-269 (1981).

[13] H. Maier, C. Pomerance, *Unusually large gaps between consecutive primes.*, Trans. Amer. Math. Soc. **322** (1990), No. 1, 201–237.

[14] J. Maynard, *Dense clusters of primes in subsets*, Compositio Math. **152** (2016), 1517-1554.

[15] J. Maynard, *Large gaps between primes*, Ann. of Math. **183** (2016) Issue 3, 915-933.

[16] J. Pintz, *Very large gaps between consecutive primes*. J. Number Theory **63** (1997), No. 2, 286–301.

[17] J. Pintz, *On the distribution of gaps between consecutive primes*, arXiv:1407.2213.

[18] R. A. Rankin, *The difference between consecutive prime numbers*, J. London Math. Soc. **13** (1938), 242–247.

[19] R. A. Rankin, *The difference between consecutive prime numbers*. V, Proc. Edinburgh Math. Soc. (2) **13** (1962/63), 331–332.

[20] P. Ribenboim, *The Little Book of Bigger Primes* Second Edition. Springer-Verlag. p. 185. (2004) ISBN 9780387201696.

[21] G. Ricci, *Ricerche arithmetiche sui polinomi. II. Intorno a una proposizione non vera di Legendre*, Rend. Circ. Mat. Palermo, **58** (1934), 190-207.

[22] A. Schönhage, *Eine Bemerkung zur Konstruktion grosser Primzahll ücken*, Arch. Math. **14** (1963), 29–30.

[23] E. Westzynthius, *Über die Verteilung der Zahlen, die zu den n ersten Primzahlen teilerfremd sind*, Commentationes Physico - Mathematicae, Societas Scientarium Fennica, Helsingfors **5**, No. 25, (1931) 1–37.

[24] M. Wolf, *Nearest-neighbor-spacing distribution of prime numbers and quantum chaos*, Phys. Rev. E, **89** (2): 022922 (2014).

K. SONO
Kanto Gakuin University,
Kanazawa, Yokohama,
Kanagawa, Japan
e-mail: `sono@kanto-gakuin.ac.jp`