ON SMOOTH GAPS BETWEEN
CONSECUTIVE PRIME NUMBERS

A. Balog[1], J. Brüdern and T. D. Wooley[2]

Dedicated to the memory of Paul Erdős

1. Introduction

Investigations concerning the gaps between consecutive prime numbers have long occupied an important position on the interface between additive and multiplicative number theory. Perhaps the most famous problem concerning these gaps, the Twin Prime Conjecture, asserts that the aforementioned gaps are infinitely often as small as 2. Although a proof of this conjecture seems presently far beyond our reach (but see [5] and [10] for related results), weak evidence in its favour comes from studying unusually short gaps between prime numbers. Thus, while it follows from the Prime Number Theorem that the average gap between consecutive primes of size about x is around $\log x$, it is now known that such gaps can be infinitely often smaller than $0.249 \log x$ (this is a celebrated result of Maier [12], building on earlier work of a number of authors; see in particular [7], [13], [3] and [11]). A conjecture weaker than the Twin Prime Conjecture asserts that there are infinitely many gaps between prime numbers which are powers of 2, but unfortunately this conjecture also seems well beyond our grasp. Extending this line of thought, Kent D. Boklan has posed the problem of establishing that the gaps between prime numbers infinitely often have only small prime divisors, and here the latter divisors should be small relative to the size of the small gaps established by Maier [12]. In this paper we show that the gaps between *consecutive* prime numbers infinitely often have only small prime divisors, thereby solving Boklan's problem. It transpires that the methods which we develop to treat Boklan's problem are capable also of detecting multiplicative properties of more general type in the differences between consecutive primes, and this theme we also explore herein.

In order to describe our conclusions precisely we require some notation. We take $(p_m)_{m=1}^{\infty} = (2, 3, 5, \dots)$ to be the sequence of prime numbers, and throughout use

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TeX

the letters p, q and π to denote prime numbers. When n is a natural number, we denote by $P(n)$ the largest prime factor of n. It is convenient to describe an integer n as being y-smooth when $P(n) \leq y$. Finally, when k is a positive integer and N is a positive real number, define

$$Z(N; 2k) = \sum_{\substack{1 < p, q \leq N \\ p-q=2k}} (\log p)(\log q). \tag{1.1}$$

In the current state of knowledge, of course, we are unable to establish an asymptotic formula for $Z(N; 2k)$. However, one may employ sieve methods to obtain an upper bound of the correct order of magnitude. In this context, we note that it is expected that

$$Z(N; 2k) = \mathcal{E}(k)N + o(N),$$

where

$$\mathcal{E}(k) = \mathfrak{S} \prod_{\substack{p|k \\ p>2}} \frac{p-1}{p-2}, \tag{1.2}$$

and

$$\mathfrak{S} = 2 \prod_{p>2} \left(1 - (p-1)^{-2}\right).$$

Our conclusions depend naturally on a constant C arising in the aforementioned sieving problem, and since it is conceivable that this constant may be susceptible to improvement, we formulate our results in terms of the following hypothesis.

Hypothesis $\mathcal{H}(C)$. *Let ε and A be positive numbers, and let N be sufficiently large in terms of ε and A. Then whenever k is a natural number with $k < (\log N)^A$, one has*

$$Z(N; 2k) < (C + \varepsilon)N\mathcal{E}(k).$$

We note that the proof of Theorem 2 of Bombieri and Davenport [3] shows that the hypothesis $\mathcal{H}(C)$ holds with $C = 4$, and later work of Bombieri, Friedlander and Iwaniec [4], and Fouvry and Grupp [8], may be used to establish that $\mathcal{H}(C)$ holds with $C = 3.5$, and $C = 3.454$, respectively. Moreover, the truth of the conjectured asymptotic formula (1.2) would imply that $\mathcal{H}(C)$ holds with $C = 1$.

We now announce our solution of Boklan's problem, which we establish in §§2 and 3.

Theorem 1. *Suppose that C is a positive number exceeding 1 for which the hypothesis $\mathcal{H}(C)$ holds. Then one has the following conclusions.*

(a) *Whenever α is a real number with $\alpha > \exp(-1/C)$, there are infinitely many primes p and q with $q < p$ and*

$$P(p-q) < (\log p)^\alpha.$$

(b) *Whenever β is a real number with $\beta > \exp(-1/(2C))$, there are infinitely many consecutive primes, p_n and p_{n+1}, with*

$$P(p_{n+1} - p_n) < (\log p_{n+1})^{\beta}.$$

In view of the remarks following the statement of the hypothesis $\mathcal{H}(C)$, we have the unconditional conclusion that infinitely many gaps between primes of size about x are $(\log x)^{3/4}$-smooth, and moreover that infinitely many gaps between consecutive primes of size about x are $(\log x)^{7/8}$- smooth.

We are also able to establish the existence of chains of consecutive prime numbers with the property that the difference between each pair of successive elements of the chain is non-trivially smooth. Thus, in §4 we establish the following conclusion.

Theorem 2. *Suppose that C is a positive number exceeding 1 for which the hypothesis $\mathcal{H}(C)$ holds. Let r be a natural number with $r \geq 2$. Then one has the following conclusions.*

(a) *Whenever γ_r is a real number with $\gamma_r > \exp(-1/(2Cr^2))$, there are infinitely many chains $(p_n, p_{n+1}, \ldots, p_{n+r})$ of consecutive primes with*

$$P\Big(\prod_{i=1}^{r}(p_{n+i} - p_{n+i-1})\Big) < (\log p_{n+r})^{\gamma_r}.$$

(b) *Write*

$$A_r = \tfrac{1}{2}\left(2C - 1 + \frac{2}{r} + \sqrt{\left(2C - 1 + \frac{2}{r}\right)^2 - \frac{4}{r^2}}\right).$$

Then whenever δ_r is a real number with $\delta_r > \exp\left(-1/(A_r r^2)\right)$, there are infinitely many chains $(p_n, p_{n+1}, \ldots, p_{n+r})$ of consecutive primes with

$$P\Big(\prod_{i=1}^{r}(p_{n+i} - p_{n+i-1})\Big) < (\log p_{n+r})^{\delta_r}. \tag{1.3}$$

In particular, when r is large, the conclusion (1.3) holds with

$$\delta_r = \exp\left(-\frac{1 + O(1/r)}{(2C - 1)r^2}\right).$$

Our methods are by no means limited to detecting smooth gaps between consecutive primes, and in principle one is at liberty to impose almost any mild multiplicative constraint on the gaps. By way of illustration, in §5 we establish that there are infinitely many 5-free gaps between consecutive primes.

Theorem 3. *There are infinitely many consecutive primes, p_n and p_{n+1}, with the property that for no prime π does one have $\pi^5|(p_{n+1} - p_n)$.*

As a final illustration of the scope of the methods of this paper, in §6 we discuss erratic behaviour amongst consecutive gaps.

Theorem 4. *There is a positive number δ with the property that, for infinitely many natural numbers n, one has either*

$$\frac{p_{n+2} - p_{n+1}}{p_{n+1} - p_n} > 1 + \delta \quad or \quad \frac{p_{n+2} - p_{n+1}}{p_{n+1} - p_n} < 1 - \delta.$$

Moreover, there is a positive number ω such that, for infinitely many natural numbers n, one has

$$|p_{n+2} - 2p_{n+1} + p_n| > \omega \log p_n.$$

Since, on average, one has $p_{n+2} - p_{n+1} = (1+o(1))(p_{n+1} - p_n)$, the first conclusion of Theorem 4 shows that, infinitely often, a gap between consecutive primes is either significantly larger, or significantly smaller, than its predecessor. We note that one may take $\delta = 0.0045$ in the first conclusion of Theorem 4. Presumably, by making use of the most modern sieve estimates, it would be possible to increase the permissible choices for δ, but we have not pursued such matters in this paper. The second conclusion of Theorem 4 shows that the second order differences between consecutive prime numbers are infinitely often "large". A priori, there is no reason to suppose that the second order gaps are even as large as $(\log x)^\varepsilon$ for primes of size about x. As should be evident from the discussion of §6, the methods we develop are capable of showing that for any fixed natural number r, the rth order differences between consecutive prime numbers of size about x are infinitely often $\gg_r \log x$. In the interests of concision, we do not discuss the latter application in detail, confining ourselves herein to merely illustrating such ideas in Theorem 4 above.

The basic strategy implicit in the proof of Theorems 1, 2 and 3 is simple to describe. We first obtain an ample supply of gaps between primes of size not much larger than the average gap. Such a supply of gaps is guaranteed by the work of Bombieri and Davenport [3], and indeed the latter already suffices for our proof of Theorem 1(a). In order to establish conclusions concerning *consecutive* primes, we are forced to make some elementary observations concerning the distribution function of the consecutive gaps. We then employ the upper bound embodied in the hypothesis $\mathcal{H}(C)$ to estimate the number of such gaps divisible by a given modulus $2k$. Thus, by summing the contributions from a number of such moduli, we are able to detect weak multiplicative properties by engineering an elementary sieving procedure.

Throughout this paper ε denotes a sufficiently small positive number, and implicit constants in the notations of Landau and Vinogradov depend at most on ε, the parameter r, and the hypothetical constant C, unless stated otherwise. As is usual, we denote the number of primes at most x in size by $\pi(x)$.

2. Smooth gaps between prime numbers

We begin our investigations concerning the gaps between prime numbers by establishing Theorem 1(a). In so doing we take advantage of the opportunity to record several estimates useful in later sections. The following result of Bombieri and Davenport [3] yields a supply of typical prime gaps sufficient for our deliberations in this section.

Lemma 2.1. *Suppose that ε and A are fixed positive numbers, and that N is a positive real number sufficiently large in terms of ε and A. Then whenever k is a positive number with $1 \le k \le (\log N)^A$, one has*

$$\sum_{1 \le n \le k} \left(1 - \frac{n}{k}\right) Z(N; 2n) > N \sum_{1 \le n \le k} \left(1 - \frac{n}{k}\right) \mathcal{E}(n) - \left(\frac{1}{4} + \varepsilon\right) N \log N. \quad (2.1)$$

Proof. This is immediate from [3, Theorem 1], on taking $t(n) = 1 - |n|/k$.

Fortunately, the sum on the right hand side of the inequality (2.1) is easily estimated by exploiting the smoothing factor, and making use of another result of Bombieri and Davenport [3].

Lemma 2.2. *For each positive number k, one has*

$$\sum_{1 \le n \le k} \mathcal{E}(n) = 2k + O((\log(2k))^2)$$

and

$$\sum_{1 \le n \le k} (1 - n/k)\mathcal{E}(n) = k + O((\log(2k))^2).$$

Proof. Bearing in mind our modest adjustments to the notation of [3], the first assertion of the lemma follows from [3, equation (33)]. By partial summation, moreover,

$$\sum_{1 \le n \le k} (1 - n/k)\mathcal{E}(n) = \frac{1}{k} \sum_{1 \le \nu \le k-1} \sum_{1 \le n \le \nu} \mathcal{E}(n) = \frac{1}{k} \sum_{1 \le \nu \le k-1} \left(2\nu + O((\log(2\nu))^2)\right),$$

and thus the second assertion of the lemma follows immediately from the first.

Next we estimate the contribution to the left hand side of (2.1) arising from those n possessing a prime divisor exceeding some parameter D.

Lemma 2.3. *Suppose that C is a positive number exceeding 1 for which the hypothesis $\mathcal{H}(C)$ holds. Let ε and A be fixed positive numbers. Suppose that N is a positive number sufficiently large in terms of ε and A, and that k is a real number*

with $1 \leq k \leq (\log N)^A$. *Then whenever D is a real number with $2 \leq D \leq k$, one has*

$$\sum_{p>D} \sum_{\substack{1 \leq n \leq k \\ p|n}} \left(1 - \frac{n}{k}\right) Z(N; 2n) < (C+\varepsilon)Nk \left(\log\left(\frac{\log k}{\log D}\right) + O\left(\frac{1}{\sqrt{\log D}}\right)\right).$$

Proof. For each prime number p with $p > 2$, it follows from the hypothesis $\mathcal{H}(C)$ that

$$\sum_{\substack{1 \leq n \leq k \\ p|n}} \left(1 - \frac{n}{k}\right) Z(N; 2n) = \sum_{1 \leq m \leq k/p} \left(1 - \frac{pm}{k}\right) Z(N; 2pm)$$

$$< (C+\varepsilon)N \sum_{1 \leq m \leq k/p} \left(1 - \frac{pm}{k}\right) \mathcal{E}(pm). \tag{2.2}$$

But in view of (1.2), one has

$$\sum_{1 \leq m \leq k/p} \left(1 - \frac{pm}{k}\right) \mathcal{E}(pm) \leq \mathfrak{S} \sum_{1 \leq m \leq k/p} \left(1 - \frac{pm}{k}\right) \frac{p-1}{p-2} \prod_{\substack{\pi|m \\ \pi>2}} \frac{\pi-1}{\pi-2}$$

$$= \frac{p-1}{p-2} \sum_{1 \leq m \leq k/p} \left(1 - \frac{pm}{k}\right) \mathcal{E}(m),$$

whence by Lemma 2.2,

$$\sum_{1 \leq m \leq k/p} \left(1 - \frac{pm}{k}\right) \mathcal{E}(pm) \leq \frac{p-1}{p-2} \left(\frac{k}{p} + O\left((\log(2k/p))^2\right)\right). \tag{2.3}$$

Next one observes that a well-known version of the Prime Number Theorem with error term (see, for example, [6]) yields the estimates

$$\sum_{D<p\leq k} \frac{p-1}{p(p-2)} = \log\left(\frac{\log k}{\log D}\right) + O\left(\frac{1}{\log D}\right), \tag{2.4}$$

and

$$\sum_{D<p\leq k} \frac{p-1}{p-2} (\log(2k/p))^2 = \sum_{D<p\leq k(\log k)^{-2}} \frac{p-1}{p-2} (\log(2k/p))^2$$

$$+ \sum_{k(\log k)^{-2}<p\leq k} \frac{p-1}{p-2} (\log(2k/p))^2$$

$$\ll \frac{k(\log(2k))^2}{(\log k)^3} + \frac{k(\log\log(3k))^2}{\log k}. \tag{2.5}$$

On combining (2.2)-(2.5), therefore, we deduce that

$$\sum_{\substack{p>D}}\sum_{\substack{1\leq n\leq k \\ p|n}}\left(1-\frac{n}{k}\right)Z(N;2n)$$

$$< (C+\varepsilon)Nk\left(\log\left(\frac{\log k}{\log D}\right)+O\left(\frac{1}{\log D}+\frac{1}{\sqrt{\log k}}\right)\right),$$

and the desired conclusion follows immediately.

We note that the error term occurring in the conclusion of Lemma 2.3 could be sharpened with somewhat greater effort. However, such would be surplus to our requirements.

Equipped now with the apparatus necessary for performing the sieving alluded to in the introduction, the proof of Theorem 1(a) may be swiftly disposed of. We suppose that C is a positive number exceeding 1 for which the hypothesis $\mathcal{H}(C)$ holds, and that ε is a positive number sufficiently small in terms of C. Write $\alpha = \exp(-1/(C+2\varepsilon))$, take N to be a real number sufficiently large in terms of ε, and take also $k = \varepsilon^{-2}\log N$ and $D = (\log N)^{\alpha}$. By Lemma 2.1 we have

$$\sum_{\substack{1\leq n\leq k \\ p|n\Rightarrow p\leq D}}\left(1-\frac{n}{k}\right)Z(N;2n) > N\sum_{1\leq n\leq k}\left(1-\frac{n}{k}\right)\mathcal{E}(n)-\left(\frac{1}{4}+\varepsilon\right)Nk\varepsilon^2$$

$$-\sum_{\substack{p>D}}\sum_{\substack{1\leq n\leq k \\ p|n}}\left(1-\frac{n}{k}\right)Z(N;2n). \tag{2.6}$$

Thus, on applying Lemma 2.2 to estimate the first sum on the right hand side of (2.6), and estimating the second such sum by recourse to Lemma 2.3, we obtain

$$\sum_{\substack{1\leq n\leq k \\ p|n\Rightarrow p\leq D}}\left(1-\frac{n}{k}\right)Z(N;2n) > Nk\left(1-\varepsilon^2-(C+\varepsilon)\log\left(\frac{\log k}{\log D}\right)+O\left(\frac{1}{\sqrt{\log D}}\right)\right)$$

$$\geq Nk\left(1-2\varepsilon^2+(C+\varepsilon)\log\alpha\right) > \frac{\varepsilon Nk}{2C}. \tag{2.7}$$

In order to complete the proof of Theorem 1(a), it remains only to show that the D-smooth gaps supplied by (2.7) do not correspond exclusively to excessively small primes. However, by applying Lemma 2.2 in combination with the hypothesis $\mathcal{H}(C)$, it follows that for each large number M one has

$$\sum_{1\leq n\leq k}\left(1-\frac{n}{k}\right)Z(M;2n) < (C+\varepsilon)Mk+O\left(M(\log(2k))^2\right).$$

Consequently, on taking $M = \varepsilon N/(8C^2)$, we deduce from (2.7) that

$$\sum_{\substack{1\leq n\leq k \\ p|n\Rightarrow p\leq D}}\left(1-\frac{n}{k}\right)(Z(N;2n)-Z(M;2n)) > \frac{\varepsilon Nk}{2C}-2CMk = \frac{\varepsilon Nk}{4C}.$$

We therefore conclude that there exist D-smooth values of n for which there exist primes p and q with $q < p \le N$ and $p > \varepsilon N/(8C^2)$, and satisfying the equation $p - q = 2n$. Since $D \le (\log p)^{\alpha + \varepsilon}$, the conclusion of Theorem 1(a) follows immediately.

3. Smooth gaps between consecutive prime numbers

In the remainder of this paper we restrict attention to gaps between *consecutive* prime numbers. In order to adapt the ideas of the previous section to successfully analyse such gaps, we require information concerning the distribution function for these gaps. When n is a natural number, we denote by d_n the nth prime gap $d_n = p_{n+1} - p_n$. We fix x to be some large real number, and when Λ is a nonnegative real number, we define the distribution function $f_x(\Lambda)$ by

$$f_x(\Lambda) = \frac{\log x}{x} \operatorname{card}\{n \in \mathbb{N} : d_n > \Lambda \log x \text{ and } p_{n+1} \le x\}.$$

Lemma 3.1. *The function $f_x(\Lambda)$ has the following properties:*
  (i) *for each fixed x one has that $f_x(\Lambda)$ is a piecewise continuous function of Λ which is monotonic decreasing on $[0, \infty)$,*
  (ii) *for each $\Lambda \in [0, \infty)$, one has $0 \le f_x(\Lambda) \le 1 + O(1/\log x)$,*
  (iii) *one has $\int_0^\infty f_x(\Lambda) d\Lambda \le 1$,*
  (iv) *whenever r is a natural number and ε is a positive number with $\varepsilon < 1$, there exists a real number Λ_ε, with $\varepsilon \le \Lambda_\varepsilon \le 2r + 2\varepsilon$, which satisfies the property that*
$$f_x(\Lambda_\varepsilon) \le \frac{1}{r}\left(1 - \frac{\Lambda_\varepsilon}{2r + 2\varepsilon}\right).$$

Proof. That $f_x(\Lambda)$ is piecewise continuous and monotonic decreasing on $[0, \infty)$, and that $f_x(\Lambda) \ge 0$ for each Λ, is immediate from the definition of $f_x(\Lambda)$. Thus the property (i) claimed in the lemma holds true. In order to establish property (ii) we have merely to observe that, as a consequence of a familiar version of the Prime Number Theorem with error term (see [6]), one has

$$f_x(\Lambda) \le \frac{\log x}{x} \sum_{p_{n+1} \le x} 1 = 1 + O(1/\log x).$$

Also, it follows from the definition of $f_x(\Lambda)$ that

$$\int_0^\infty f_x(\Lambda) d\Lambda = \frac{\log x}{x} \int_0^\infty \sum_{\substack{d_n > \Lambda \log x \\ p_{n+1} \le x}} 1 \, d\Lambda$$

$$= \frac{\log x}{x} \sum_{p_{n+1} \le x} \int_0^{d_n/\log x} d\Lambda = \frac{1}{x} \sum_{p_{n+1} \le x} d_n \le 1,$$

and so property (iii) holds.

In order to establish property (iv), we consider the function

$$g_r(\Lambda) = \begin{cases} \frac{1}{r}\left(1 - \frac{\Lambda}{2r+2\varepsilon}\right), & \text{when } 0 \leq \Lambda \leq 2r + 2\varepsilon, \\ 0, & \text{otherwise,} \end{cases}$$

and observe that, in view of the conclusion of part (iii) of the lemma, one has

$$\int_\varepsilon^{2r+2\varepsilon} (f_x(\Lambda) - g_r(\Lambda))\, d\Lambda \leq \int_0^\infty f_x(\Lambda)d\Lambda - \int_\varepsilon^{2r+2\varepsilon} g_r(\Lambda)d\Lambda$$

$$\leq 1 - \frac{1}{2r}(2r + \varepsilon)\left(1 - \frac{\varepsilon}{2r + 2\varepsilon}\right)$$

$$= -\frac{\varepsilon^2}{4r(r + \varepsilon)} < 0.$$

Consequently, for some real number Λ_ε with $\varepsilon \leq \Lambda_\varepsilon \leq 2r + 2\varepsilon$, one has $f_x(\Lambda_\varepsilon) < g_r(\Lambda_\varepsilon)$, and the desired conclusion follows immediately.

Before advancing to prove Theorem 1(b), we pause to convert the hypothesis $\mathcal{H}(C)$ into an unweighted version more convenient for the application at hand. When k is a positive integer, define

$$Z^*(N; 2k) = \sum_{\substack{1 < p,q \leq N \\ p-q=2k}} 1. \tag{3.1}$$

Lemma 3.2. *Suppose that C is a positive number exceeding 1 for which the hypothesis $\mathcal{H}(C)$ holds. Let ε and A be positive numbers, and let N be sufficiently large in terms of C, ε and A. Then whenever k is a natural number with $k < (\log N)^A$, one has*

$$Z^*(N; 2k) < (C + \varepsilon)\mathcal{E}(k)N(\log N)^{-2}.$$

Proof. With the hypotheses of the statement of the lemma, it follows from (1.1) that whenever $k < (\log N)^A$, one has

$$\sum_{\substack{N(\log N)^{-3} < p,q \leq N \\ p-q=2k}} 1 \leq \left(\log(N(\log N)^{-3})\right)^{-2} \sum_{\substack{1 < p,q \leq N \\ p-q=2k}} (\log p)(\log q)$$

$$< (C + \tfrac{1}{2}\varepsilon)\mathcal{E}(k)N(\log N)^{-2}\left(1 + O\left(\frac{\log\log N}{\log N}\right)\right). \tag{3.2}$$

On the other hand, again from (1.1),

$$\sum_{\substack{1 < p,q \leq 2N(\log N)^{-3} \\ p-q=2k}} 1 \leq \sum_{\substack{1 < p,q \leq 2N(\log N)^{-3} \\ p-q=2k}} (\log p)(\log q) < 2(C + \varepsilon)\mathcal{E}(k)N(\log N)^{-3}.$$

$$\tag{3.3}$$

On combining (3.2) and (3.3), therefore, we find that the conclusion of the lemma follows from (3.1).

We are now equipped to prove Theorem 1(b). Suppose that C is a positive number exceeding 1 for which the hypothesis $\mathcal{H}(C)$ holds. Let ε be a positive number sufficiently small in terms of C, and let x be a real number sufficiently large in terms of ε and C. Then according to Lemma 3.1, there is a real number Λ_ε, with $\varepsilon \leq \Lambda_\varepsilon \leq 2 + 2\varepsilon$, which satisfies the property that

$$f_x(\Lambda_\varepsilon) \leq 1 - \frac{\Lambda_\varepsilon}{2 + 2\varepsilon}.$$

It therefore follows from the definition of the function $f_x(\Lambda)$, together with a version of the Prime Number Theorem with error term, that

$$\sum_{\substack{d_n \leq \Lambda_\varepsilon \log x \\ p_{n+1} \leq x}} 1 = \pi(x) - 1 - f_x(\Lambda_\varepsilon) \frac{x}{\log x}$$

$$\geq \frac{\Lambda_\varepsilon}{2 + 2\varepsilon} \frac{x}{\log x} + O\left(\frac{x}{(\log x)^2}\right). \tag{3.4}$$

We next estimate the number of the gaps of size at most $\Lambda_\varepsilon \log x$ which are divisible by large primes. Let D be a real parameter with $2 \leq D \leq \Lambda_\varepsilon \log x$. Then by Lemma 3.2,

$$\sum_{D < p \leq \Lambda_\varepsilon \log x} \sum_{\substack{d_n \leq \Lambda_\varepsilon \log x \\ p_{n+1} \leq x \\ p | d_n}} 1 \leq \sum_{D < p \leq \Lambda_\varepsilon \log x} \sum_{1 \leq k \leq \frac{\Lambda_\varepsilon \log x}{2p}} Z^*(x; 2kp)$$

$$< (C + \varepsilon) \frac{x}{(\log x)^2} \sum_{D < p \leq \Lambda_\varepsilon \log x} \sum_{1 \leq k \leq \frac{\Lambda_\varepsilon \log x}{2p}} \mathcal{E}(kp). \tag{3.5}$$

But by (1.2) and Lemma 2.2, one has

$$\sum_{1 \leq k \leq \frac{\Lambda_\varepsilon \log x}{2p}} \mathcal{E}(kp) \leq \mathfrak{S} \sum_{1 \leq k \leq \frac{\Lambda_\varepsilon \log x}{2p}} \frac{p-1}{p-2} \prod_{\substack{\pi | k \\ \pi > 2}} \frac{\pi - 1}{\pi - 2}$$

$$= \frac{p-1}{p-2} \sum_{1 \leq k \leq \frac{\Lambda_\varepsilon \log x}{2p}} \mathcal{E}(k)$$

$$\leq \frac{p-1}{p-2} \left(\frac{\Lambda_\varepsilon \log x}{p} + O\left((\log((\Lambda_\varepsilon \log x)/p))^2\right)\right). \tag{3.6}$$

Thus, on substituting (3.6) into (3.5), and making use of the Prime Number Theorem as in the argument leading to (2.4) and (2.5), we obtain

$$\sum_{\substack{D<p\leq\Lambda_\varepsilon\log x}} \sum_{\substack{d_n\leq\Lambda_\varepsilon\log x\\ p_{n+1}\leq x\\ p|d_n}} 1$$

$$< (C+\varepsilon)\Lambda_\varepsilon \frac{x}{\log x} \sum_{D<p\leq\Lambda_\varepsilon\log x} \left(\frac{p-1}{p(p-2)} + O\left(\frac{(\log((\Lambda_\varepsilon\log x)/p))^2}{\log x} \right) \right)$$

$$= (C+\varepsilon)\Lambda_\varepsilon \frac{x}{\log x} \left(\log\left(\frac{\log(\Lambda_\varepsilon\log x)}{\log D} \right) + O\left(\frac{1}{\sqrt{\log D}} \right) \right). \tag{3.7}$$

Finally, on taking

$$\alpha = \exp\left(-\frac{1}{2(1+\varepsilon)(C+2\varepsilon)} \right)$$

and $D = (\log x)^\alpha$, and collecting together (3.4) and (3.7), we obtain

$$\sum_{\substack{d_n\leq\Lambda_\varepsilon\log x\\ p_{n+1}\leq x\\ p|d_n\Rightarrow p\leq D}} 1 \geq \frac{\Lambda_\varepsilon}{2+2\varepsilon} \frac{x}{\log x} \left(1 + 2(1+\varepsilon)(C+\varepsilon)\log\alpha + O\left((\log\log x)^{-1/2} \right) \right)$$

$$= \frac{\Lambda_\varepsilon}{2+2\varepsilon} \frac{x}{\log x} \frac{\varepsilon}{C+2\varepsilon} \left(1 + O\left((\log\log x)^{-1/2} \right) \right).$$

On recalling that $\Lambda_\varepsilon \geq \varepsilon$, therefore, we conclude that

$$\sum_{\substack{d_n\leq\Lambda_\varepsilon\log x\\ p_{n+1}\leq x\\ p|d_n\Rightarrow p\leq D}} 1 > \frac{\varepsilon^2 x}{3C\log x}. \tag{3.8}$$

The proof of Theorem 1(b) will be completed on accounting for the contribution of the small primes. But the Prime Number Theorem shows that

$$\sum_{\substack{d_n\leq\Lambda_\varepsilon\log x\\ p_{n+1}\leq\varepsilon^3 x\\ p|d_n\Rightarrow p\leq D}} 1 \leq \pi(\varepsilon^3 x) < \frac{2\varepsilon^3 x}{\log x}, \tag{3.9}$$

whence by (3.8),

$$\sum_{\substack{d_n\leq\Lambda_\varepsilon\log x\\ \varepsilon^3 x<p_{n+1}\leq x\\ p|d_n\Rightarrow p\leq D}} 1 > \frac{\varepsilon^2 x}{4C\log x}.$$

Consequently, there exist D-smooth values of d_n for which $p_{n+1} > \varepsilon^3 x$. Since $D \leq (\log p_{n+1})^{\alpha+\varepsilon}$, the conclusion of Theorem 1(b) is immediate.

4. Chains of smooth gaps between consecutive primes

It is evident that the approach described in the previous section can be modified so as to demonstrate the existence of chains of smooth gaps between consecutive primes, provided that one has sufficient knowledge concerning upper bounds for the number of solutions to simultaneous gap problems. Such simultaneous problems, unfortunately, are not well understood, and so we are forced to sidestep such difficulties with some elementary observations concerning $f_x(\Lambda)$.

Lemma 4.1. *Suppose that C is a positive number exceeding 1 for which the hypothesis $\mathcal{H}(C)$ holds. Then one has the following.*

(i) *For each positive number ε with $\varepsilon \le 1/(2C)$, whenever x is sufficiently large in terms of C and ε, one has for $\varepsilon \le \Lambda \le 1/(C+\varepsilon)$,*

$$f_x(\Lambda) > 1 - (C+\varepsilon)\Lambda.$$

(ii) *Suppose that r is a natural number with $r \ge 2$, and ε is a positive number with $\varepsilon \le 1/(2C)$. Write $C_\varepsilon = C(1+3\varepsilon)$, and define B_r by*

$$B_r = \tfrac{1}{2}(C+\varepsilon)\left(2C_\varepsilon - 1 + \frac{2}{r} - \sqrt{\left(2C_\varepsilon - 1 + \frac{2}{r}\right)^2 - \frac{4}{r^2}}\right). \qquad (4.1)$$

Then there exists a real number Λ_ε, with

$$\frac{1}{C+\varepsilon} \le \Lambda_\varepsilon \le \frac{1}{rB_r},$$

which satisfies the property that

$$f_x(\Lambda_\varepsilon) \le \frac{1}{r} - B_r\Lambda_\varepsilon.$$

Proof. Suppose that C is a real number exceeding 1 for which the hypothesis $\mathcal{H}(C)$ holds. We start by applying Lemma 3.2 to obtain, for each positive number Λ,

$$\sum_{\substack{d_n > \Lambda \log x \\ p_{n+1} \le x}} 1 = \pi(x) - 1 - \sum_{\substack{d_n \le \Lambda \log x \\ p_{n+1} \le x}} 1$$

$$\ge \pi(x) - 1 - \sum_{1 \le d \le \frac{1}{2}\Lambda \log x} Z^*(x; 2d)$$

$$> \pi(x) - 1 - (C + \tfrac{1}{2}\varepsilon)\frac{x}{(\log x)^2} \sum_{1 \le d \le \frac{1}{2}\Lambda \log x} \mathcal{E}(d).$$

Then on applying the Prime Number Theorem with a suitable error term, we deduce from Lemma 2.2 together with the definition of $f_x(\Lambda)$ that for $\varepsilon \leq \Lambda \leq 1/(C+\varepsilon)$,

$$f_x(\Lambda) = \frac{\log x}{x} \sum_{\substack{d_n > \Lambda \log x \\ p_{n+1} \leq x}} 1 > 1 - (C + \tfrac{1}{2}\varepsilon)\Lambda + O((\log x)^{-1/2}),$$

whence part (i) of the lemma follows immediately.

In order to establish part (ii) of the lemma we consider the function $g_r(\Lambda)$ defined by

$$g_r(\Lambda) = \begin{cases} 1 - (C + \varepsilon)\Lambda, & \text{when } 0 \leq \Lambda \leq 1/(C+\varepsilon), \\ \frac{1}{r} - B_r\Lambda, & \text{when } 1/(C+\varepsilon) < \Lambda \leq 1/(rB_r), \\ 0, & \text{otherwise.} \end{cases}$$

¿From Lemma 3.1(iii) together with part (i) of this lemma, we have

$$\int_{1/(C+\varepsilon)}^{\infty} f_x(\Lambda)d\Lambda = \int_0^{\infty} f_x(\Lambda)d\Lambda - \int_0^{1/(C+\varepsilon)} f_x(\Lambda)d\Lambda$$

$$< 1 - \int_{\varepsilon}^{1/(C+\varepsilon)} \left(1 - (C+\varepsilon)\Lambda\right)d\Lambda$$

$$< 1 + \varepsilon - \frac{1}{2(C+\varepsilon)}.$$

But then, on recalling (4.1), a modest computation reveals that

$$\int_{1/(C+\varepsilon)}^{1/(rB_r)} f_x(\Lambda) - g_r(\Lambda)d\Lambda \leq \int_{1/(C+\varepsilon)}^{\infty} f_x(\Lambda)d\Lambda - \int_{1/(C+\varepsilon)}^{1/(rB_r)} g_r(\Lambda)d\Lambda$$

$$< 1 + \varepsilon - \frac{1}{2(C+\varepsilon)} - \frac{1}{2B_r}\left(\frac{1}{r} - \frac{B_r}{C+\varepsilon}\right)^2$$

$$= 1 + \varepsilon - \frac{1}{2(C+\varepsilon)} - \frac{2C_\varepsilon - 1}{2(C+\varepsilon)},$$

whence

$$\int_{1/(C+\varepsilon)}^{1/(rB_r)} f_x(\Lambda) - g_r(\Lambda)d\Lambda < 0.$$

We may therefore conclude that for some Λ_ε with $1/(C+\varepsilon) \leq \Lambda_\varepsilon \leq 1/(rB_r)$, one has $f_x(\Lambda_\varepsilon) < g_r(\Lambda_\varepsilon)$, and so part (ii) of the lemma follows.

We establish Theorem 2 following the trail laid down in §3. Suppose that C is a real number exceeding 1 for which the hypothesis $\mathcal{H}(C)$ holds. Let r be a natural number with $r \geq 2$, let ε be a positive number sufficiently small in terms of r and C, and let x be a real number sufficiently large in terms of r, ε and C. Then according

to Lemma 4.1(ii), there is a real number Λ_ε with $1/(C+\varepsilon) \leq \Lambda_\varepsilon \leq 1/(rB_r)$ (where B_r is defined as in (4.1)), which satisfies the property that

$$f_x(\Lambda_\varepsilon) \leq \frac{1}{r} - B_r \Lambda_\varepsilon.$$

It therefore follows from the definition of the function $f_x(\Lambda)$, together with a version of the Prime Number Theorem with error term, that one has

$$\sum_{\substack{d_{n+i-1} \leq \Lambda_\varepsilon \log x \ (1 \leq i \leq r) \\ p_{n+r} \leq x}} 1 \geq \pi(x) - r - r f_x(\Lambda_\varepsilon) \frac{x}{\log x}$$

$$\geq r B_r \Lambda_\varepsilon \frac{x}{\log x} + O\left(\frac{x}{(\log x)^2}\right). \tag{4.2}$$

Meanwhile, the number of consecutive r-tuples $(d_n, d_{n+1}, \ldots, d_{n+r-1})$ of gaps, in which at least one of the d_{n+i-1} $(1 \leq i \leq r)$ is divisible by a large prime, may be estimated as follows. We take D to be a real parameter with $2 \leq D \leq \Lambda_\varepsilon \log x$. Then again making use of the Prime Number Theorem, it follows from (3.7) that

$$\sum_{D < p \leq \Lambda_\varepsilon \log x} \sum_{\substack{d_{n+i-1} \leq \Lambda_\varepsilon \log x \ (1 \leq i \leq r) \\ p_{n+r} \leq x \\ p | d_{n+i-1} \text{ some } i}} 1$$

$$\leq r \sum_{D < p \leq \Lambda_\varepsilon \log x} \sum_{\substack{d_n \leq \Lambda_\varepsilon \log x \\ p_{n+1} \leq x \\ p | d_n}} 1$$

$$< r(C+\varepsilon) \Lambda_\varepsilon \frac{x}{\log x} \left(\log\left(\frac{\log(\Lambda_\varepsilon \log x)}{\log D} \right) + O\left(\frac{1}{\sqrt{\log D}} \right) \right). \tag{4.3}$$

Finally, on taking

$$\delta_r = \exp\left(-\frac{B_r}{C+2\varepsilon} \right)$$

and $D = (\log x)^{\delta_r}$, and combining (4.2) and (4.3), we deduce that

$$\sum_{\substack{d_{n+i-1} \leq \Lambda_\varepsilon \log x \ (1 \leq i \leq r) \\ p_{n+r} \leq x \\ p | d_{n+i-1} \Rightarrow p \leq D \ (1 \leq i \leq r)}} 1 \geq r B_r \Lambda_\varepsilon \frac{x}{\log x} \left(1 + \frac{C+\varepsilon}{B_r} \log \delta_r + O\left(\frac{1}{\sqrt{\log \log x}} \right) \right)$$

$$= \frac{\varepsilon r B_r \Lambda_\varepsilon}{C+2\varepsilon} \frac{x}{\log x} \left(1 + O(1/\sqrt{\log \log x}) \right).$$

On recalling that $\Lambda_\varepsilon \geq \varepsilon$, therefore, we arrive at the conclusion

$$\sum_{\substack{d_{n+i-1} \leq \Lambda_\varepsilon \log x \ (1 \leq i \leq r) \\ p_{n+r} \leq x \\ p | d_{n+i-1} \Rightarrow p \leq D \ (1 \leq i \leq r)}} 1 > \frac{\varepsilon^2 r B_r x}{2C \log x}.$$

Moreover the estimate (3.9) may again be employed to discard the small primes occurring in the latter sum, so that there exist chains $(d_n, d_{n+1}, \ldots, d_{n+r-1})$, with $p_{n+r} > \varepsilon^3 x$, and satisfying the property that whenever $\pi | d_{n+i-1}$ then $\pi \leq D$ $(1 \leq i \leq r)$. Since $D = (\log p_{n+r})^{\delta_r + \varepsilon}$, and

$$B_r = \frac{C + \varepsilon}{\frac{1}{2}\left(2C_\varepsilon - 1 + \frac{2}{r} + \sqrt{\left(2C_\varepsilon - 1 + \frac{2}{r}\right)^2 - \frac{4}{r^2}}\right) r^2},$$

the main conclusion of part (b) of Theorem 2 follows immediately. Part (a) of Theorem 2, meanwhile, follows from part (b) on noting that whenever $r \geq 2$, one has

$$A_r \leq 2C - 1 + 2/r \leq 2C.$$

Moreover, a simple expansion yields

$$A_r = 2C - 1 + \frac{2}{r} + O\left(\frac{1}{r^2}\right),$$

thereby establishing the final assertion of Theorem 2(b).

5. The existence of k-free gaps between consecutive prime numbers

Since our methods fall short of showing that there are infinitely many 4-free gaps between consecutive primes, in the interest of concision we will be crude in our proof of Theorem 3. Suppose that C is a positive number exceeding 1 for which the hypothesis $\mathcal{H}(C)$ holds. Let ε be a positive number sufficiently small in terms of C, and let x be a real number sufficiently large in terms of ε and C. We begin by noting that the lower bound (3.4) holds for some real number Λ_ε with $\varepsilon \leq \Lambda_\varepsilon \leq 2 + 2\varepsilon$. Moreover, when k is a natural number with $k \geq 2$, the number of k-free prime gaps of size at most $\Lambda_\varepsilon \log x$ is given by

$$\sum_{\substack{d_n \leq \Lambda_\varepsilon \log x \\ p_{n+1} \leq x \\ m^k | d_n \Rightarrow m=1}} 1 \geq \sum_{\substack{d_n \leq \Lambda_\varepsilon \log x \\ p_{n+1} \leq x}} 1 - \sum_p \sum_{\substack{d_n \leq \Lambda_\varepsilon \log x \\ p_{n+1} \leq x \\ p^k | d_n}} 1.$$

Thus we deduce from (3.4) and Lemma 3.2 that

$$\sum_{\substack{d_n \leq \Lambda_\varepsilon \log x \\ p_{n+1} \leq x \\ m^k | d_n \Rightarrow m=1}} 1 \geq \frac{\Lambda_\varepsilon}{2 + 2\varepsilon} \frac{x}{\log x} - \sum_p \sum_{\substack{1 \leq d \leq \frac{1}{2}\Lambda_\varepsilon \log x \\ p^k | 2d}} Z^*(x; 2d) + O\left(\frac{x}{(\log x)^2}\right)$$

$$> \frac{\Lambda_\varepsilon}{2 + 2\varepsilon} \frac{x}{\log x} - (C + \varepsilon)\frac{x}{(\log x)^2} \mathcal{M}_k(x) + O\left(\frac{x}{(\log x)^2}\right),$$

(5.1)

where

$$\mathcal{M}_k(x) = \sum_{1 \le d \le \frac{\Lambda_\varepsilon \log x}{2^k}} \mathcal{E}(2^{k-1}d) + \sum_{p>2} \sum_{1 \le d \le \frac{\Lambda_\varepsilon \log x}{2p^k}} \mathcal{E}(p^k d). \tag{5.2}$$

Next we estimate $\mathcal{M}_k(x)$. By (1.2) and Lemma 2.2, it follows from (5.2) that

$$\mathcal{M}_k(x) \le \sum_{1 \le d \le \frac{\Lambda_\varepsilon \log x}{2^k}} \mathcal{E}(d) + \sum_{\substack{p>2 \\ p^k \le \frac{1}{2}\Lambda_\varepsilon \log x}} \sum_{1 \le d \le \frac{\Lambda_\varepsilon \log x}{2p^k}} \frac{p-1}{p-2} \mathcal{E}(d)$$

$$= 2^{1-k}\Lambda_\varepsilon \log x + \sum_{\substack{p>2 \\ p^k \le \frac{1}{2}\Lambda_\varepsilon \log x}} \left(\frac{p-1}{p-2} \frac{\Lambda_\varepsilon \log x}{p^k} + O\left(\left(\log\left(p^{-k}\Lambda_\varepsilon \log x\right) \right)^2 \right) \right)$$

$$\le \omega(k)\Lambda_\varepsilon \log x + O((\log x)^{1/2+\varepsilon}), \tag{5.3}$$

where

$$\omega(k) = 2^{1-k} + \sum_{p>2} \frac{p-1}{p^k(p-2)}.$$

We note, in particular, that when $k \ge 5$, a modest computation reveals that $\omega(k) < 1/12$. On substituting (5.3) into (5.1), therefore, we conclude that whenever $k \ge 2$, one has

$$\sum_{\substack{d_n \le \Lambda_\varepsilon \log x \\ p_{n+1} \le x \\ m^k | d_n \Rightarrow m=1}} 1 > \frac{\Lambda_\varepsilon}{2+2\varepsilon} \frac{x}{\log x} \left(1 - (2+2\varepsilon)(C+\varepsilon)\omega(k) + O((\log x)^{\varepsilon-1/2}) \right).$$

Consequently, whenever $C < 6$ we deduce from the lower bound $\Lambda_\varepsilon \ge \varepsilon$ that

$$\sum_{\substack{d_n \le \Lambda_\varepsilon \log x \\ p_{n+1} \le x \\ m^k | d_n \Rightarrow m=1}} 1 > \frac{\varepsilon^2 x}{3 \log x}.$$

Since an estimate similar to (3.9) again shows that the small primes in the latter sum are inconsequential, we may conclude that there are infinitely many 5-free gaps between consecutive primes. This completes the proof of Theorem 3.

6. Inequitable consecutive prime gaps

We establish Theorem 4 through the use of another sieve estimate, which again we formulate as a hypothesis. When h and k are positive integers, and N is a large real number, we define

$$Z(N; 2k, 2h) = \sum_{\substack{1 < p,q,r \le N \\ q = p+2h \\ r = p+2h+2k}} 1,$$

in which the summation is over prime numbers p, q and r. In view of the prime k-tuples conjecture of Bateman and Horn [2] (see also [1] for a discussion of this conjecture), it is expected that

$$Z(N; 2k, 2h) = (1 + o(1))\mathcal{E}(k, h)N(\log N)^{-3},$$

where

$$\mathcal{E}(k, h) = \prod_{\pi} \left(1 - \frac{1}{\pi}\right)^{-3} \left(1 - \frac{\rho(\pi)}{\pi}\right),$$

and $\rho(\pi)$ denotes the number of solutions of the congruence

$$n(n + 2h)(n + 2h + 2k) \equiv 0 \pmod{\pi}.$$

Hypothesis $\mathcal{K}(C)$. *Let ε and A be positive numbers, and let N be sufficiently large in terms of ε and A. Then whenever h and k are natural numbers with $\max\{h, k\} < (\log N)^A$, one has*

$$Z(N; 2k, 2h) < (C + \varepsilon)\mathcal{E}(k, h)N(\log N)^{-3}.$$

By applying standard sieve methods it is possible to show that the hypothesis $\mathcal{K}(C)$ holds for some fixed positive number C (indeed, [9, Theorem 5.7] shows that $C = 48$ is permissible). Moreover, as may be verified with a little effort, there is a positive number B with the property that for any fixed positive numbers Δ_1 and Δ_2 with $\Delta_2 > \Delta_1$, one has for each large number H the upper bound

$$\sum_{1 \leq h \leq H} \sum_{\Delta_1 h \leq k \leq \Delta_2 h} \mathcal{E}(k, h) \leq (\Delta_2 - \Delta_1)(B + o(1))H^2. \tag{6.1}$$

We note that while the expenditure of sufficient effort would establish that $B = 4$ is permissible, the mere existence of such a number suffices for the proof of Theorem 4.

Before proceeding further, we require some additional information concerning the distribution function $f_x(\Lambda)$.

Lemma 6.1. *Whenever ε is a positive number with $\varepsilon < 1$, there exists a real number Λ_ε, with $\varepsilon \leq \Lambda_\varepsilon \leq \frac{3}{2}(1 + \varepsilon)$, which satisfies the property that*

$$f_x(\Lambda_\varepsilon) \leq 1 - \frac{4\Lambda_\varepsilon^2}{9(1 + \varepsilon)^2}.$$

Proof. Consider the function $b(\Lambda)$, defined by

$$b(\Lambda) = \begin{cases} 1 - \frac{4\Lambda^2}{9(1+\varepsilon)^2}, & \text{when } 0 \leq \Lambda \leq \frac{3}{2}(1 + \varepsilon), \\ 0, & \text{otherwise,} \end{cases}$$

and observe that, in view of Lemma 3.1(iii), one has

$$\int_{\varepsilon}^{3(1+\varepsilon)/2} f_x(\Lambda) - b(\Lambda) d\Lambda \leq \int_0^{\infty} f_x(\Lambda) d\Lambda - \int_{\varepsilon}^{3(1+\varepsilon)/2} b(\Lambda) d\Lambda$$
$$< 1 - \left(\tfrac{3}{2}(1+\varepsilon) - \tfrac{1}{2}(1+\varepsilon)\right) + \varepsilon = 0.$$

Thus, for some real number Λ_ε with $\varepsilon \leq \Lambda_\varepsilon \leq \frac{3}{2}(1+\varepsilon)$, one has $f_x(\Lambda_\varepsilon) < b(\Lambda_\varepsilon)$, whence the lemma follows immediately.

We advance now to dispose of the proof of Theorem 4. Suppose that C is a real number exceeding 1 for which the hypothesis $\mathcal{K}(C)$ holds. Let ε be a positive number sufficiently small in terms of C, and let x be a real number sufficiently large in terms of ε and C. Also, let δ be a small positive parameter to be chosen later. By Lemma 6.1, there is a real number Λ_ε, with $\varepsilon \leq \Lambda_\varepsilon \leq \frac{3}{2}(1+\varepsilon)$, which satisfies the property that

$$f_x(\Lambda_\varepsilon) \leq 1 - \frac{4\Lambda_\varepsilon^2}{9(1+\varepsilon)^2}.$$

By the definition of the function $f_x(\Lambda)$, together with a version of the Prime Number Theorem with error term, therefore, we have

$$\sum_{\substack{d_n \leq \Lambda_\varepsilon \log x \\ p_{n+1} \leq x}} 1 \geq \pi(x) - 1 - f_x(\Lambda_\varepsilon) \frac{x}{\log x} \geq \frac{4\Lambda_\varepsilon^2}{9(1+\varepsilon)^2} \frac{x}{\log x} + O\left(\frac{x}{(\log x)^2}\right). \quad (6.2)$$

Let N_0 denote the number of prime gaps d_n counted in the latter sum, for which the subsequent prime gap d_{n+1} satisfies $|d_{n+1}/d_n - 1| \leq \delta$. Then one has

$$N_0 \leq \sum_{p \leq x} \sum_{\substack{1 \leq h \leq \frac{1}{2}\Lambda_\varepsilon \log x \\ p+2h \text{ prime}}} \sum_{\substack{(1-\delta)h \leq k \leq (1+\delta)h \\ p+2h+2k \text{ prime}}} 1, \quad (6.3)$$

and so by combining (6.1) with the hypothesis $\mathcal{K}(C)$, we arrive at the estimate

$$N_0 \leq (C+\varepsilon) \frac{x}{(\log x)^3} \sum_{1 \leq h \leq \frac{1}{2}\Lambda_\varepsilon \log x} \sum_{(1-\delta)h \leq k \leq (1+\delta)h} \mathcal{E}(k,h)$$
$$\leq \tfrac{1}{2}\delta(C+\varepsilon)(B+\varepsilon)\Lambda_\varepsilon^2 \frac{x}{\log x}. \quad (6.4)$$

On recalling (6.2), we deduce from (6.3) and (6.4) that

$$\sum_{\substack{d_n \leq \Lambda_\varepsilon \log x \\ p_{n+1} \leq x \\ |d_{n+1}/d_n - 1| > \delta}} 1 \geq \frac{4\Lambda_\varepsilon^2}{9(1+\varepsilon)^2} \frac{x}{\log x} \left(1 - \tfrac{9}{8}\delta(C+\varepsilon)(B+\varepsilon)(1+\varepsilon)^2 + O(1/\log x)\right).$$

Consequently, whenever

$$\delta < \frac{8}{9(C+\varepsilon)(B+\varepsilon)(1+\varepsilon)^2},$$

one finds that

$$\sum_{\substack{d_n \leq \Lambda_\varepsilon \log x \\ p_{n+1} \leq x \\ |d_{n+1}/d_n - 1| > \delta}} 1 \gg \frac{x}{\log x}, \tag{6.5}$$

where here the implicit constant may depend on δ, B and C. Again, by an estimate similar to (3.9), one finds that the small primes provide a negligible contribution, and thus we deduce that for infinitely many natural numbers n, one has $|d_{n+1}/d_n - 1| > \delta$. This completes the proof of the first assertion of Theorem 4.

We remark that the choices $C = 48$ and $B = 4$ yield the conclusion that δ is permissible whenever $\delta < 1/216$, thereby justifying the assertion made in the introduction.

Finally, we observe that whenever μ is a positive number, it follows from Lemmata 2.2 and 3.2 that

$$\sum_{\substack{d_n \leq \mu \log x \\ p_{n+1} \leq x}} 1 < 2(C+\varepsilon)\mu \frac{x}{\log x},$$

whence, on taking μ_ε to be a positive number sufficiently small in terms of ε and δ, we deduce from (6.5) that

$$\sum_{\substack{\mu_\varepsilon \log x \leq d_n \leq \Lambda_\varepsilon \log x \\ p_{n+1} \leq x \\ |d_{n+1}/d_n - 1| > \delta}} 1 \gg \frac{x}{\log x}.$$

Thus we conclude that for infinitely many natural numbers n, one has $d_n \geq \mu_\varepsilon \log x$, $p_{n+1} \leq x$, and

$$|d_{n+1} - d_n| > \delta d_n \geq \mu_\varepsilon \delta \log x.$$

The second assertion of Theorem 4 follows immediately.

REFERENCES

1. A. Balog, *The prime k-tuplets conjecture on average*, Analytic number theory (Allerton Park, IL, 1989), Progr. Math. 85, Birkhäuser Boston, Boston, MA, 1990, pp. 47–75.
2. P. T. Bateman and R. A. Horn, *A heuristic asymptotic formula concerning the disribution of prime numbers*, Math. Comp. **16** (1962), 363–367.
3. E. Bombieri and H. Davenport, *Small differences between prime numbers*, Proc. Roy. Soc. Ser. A **293** (1966), 1–18.
4. E. Bombieri, J. B. Friedlander and H. Iwaniec, *Primes in arithmetic progressions to large moduli*, Acta Math. **156** (1986), 203–251.
5. J.-R. Chen, *On the representation of a large even integer as the sum of a prime and the product of at most two primes*, Sci. Sinica **16** (1973), 157–176.

6. H. Davenport, *Multiplicative number theory, 2nd ed. revised by H. L. Montgomery*, Graduate Texts in Mathematics, 74, Springer-Verlag, Berlin, 1980.

7. P. Erdős, *The difference of consecutive primes*, Duke Math. J. **6** (1940), 438–441.

8. E. Fouvry and F. Grupp, *On the switching principle in sieve theory*, J. Reine Angew. Math. **370** (1986), 101–126.

9. H. Halberstam and H.-E. Richert, *Sieve methods*, London Mathematical Society Monographs, No. 4, Academic Press, London-New York, 1974.

10. D. R. Heath-Brown, *Prime twins and Siegel zeros*, Proc. London Math. Soc. (3) **47** (1983), 193–224.

11. M. N. Huxley, *Small differences between consecutive primes II*, Mathematika **24** (1977), 142–152.

12. H. Maier, *Small differences between prime numbers*, Michigan Math. J. **35** (1988), 323–344.

13. R. A. Rankin, *The difference between consecutive prime numbers. III*, J. London Math. Soc. **22** (1947), 226–230.

AB: Mathematical Institute, Budapest 1364, Hungary
E-mail address: `balog@math-inst.hu`

JB: Mathematisches Institut A, Universität Stuttgart, Postfach 80 11 40, D-70511 Stuttgart, Germany
E-mail address: `bruedern@mathematik.uni-stuttgart.de`

TW: Department of Mathematics, University of Michigan, East Hall, 525 East University Avenue, Ann Arbor, MI 48109-1109, U.S.A.
Current address: Department of Mathematics, Princeton University, Fine Hall, Washington Road, Princeton, NJ 08544-1000, U.S.A.
E-mail address: `wooley@math.lsa.umich.edu, wooley@math.princeton.edu`