Arch. Math. 82 (2004) 33–39
0003–889X/04/010033–07
DOI 10.1007/s00013-003-4780-3

Archiv der Mathematik

Upper bounds for prime k-tuples of size log N and oscillations

By

CHRISTIAN ELSHOLTZ

Abstract. We prove the estimate

$$E_k(N) \ll \frac{N}{\exp\left((\frac{1}{4}+o(1))\frac{\log N\ \log\log\log N}{\log\log N}\right)},$$

for the number $E_k(N)$ of k-tuples $(n+a_1,\ldots,n+a_k)$ of primes not exceeding N, for k of size $c_1 \log N$ and N sufficiently large.

A bound of this strength was previously known in the special case $n-2^i$ $(1 \leqq i < \frac{\log n}{\log 2})$ only, (Vaughan, 1973). For general a_i this is an improvement upon the work of Hofmann and Wolke (1996).

The number of prime tuples of this size has considerable oscillations, when varying the prime pattern.

1. A discussion of the prime k-tuple conjecture. The prime k-tuple conjecture states that for any admissible set $\mathcal{A} = \{a_1,\ldots,a_k\}$ of positive integers there are infinitely many integers n such that all $n+a_i$, $(i = 1,\ldots,k)$, are prime, simultaneously. The set $\mathcal{A} = \{a_1,\ldots,a_k\}$ is admissible if there is no trivial congruence obstruction, i.e. if the set $\mathcal{A}$ does not contain a complete set of residues modulo any prime $p \leqq k$.

A quantitative version of these conjectures has been suggested by Hardy and Littlewood, see [8], and later by Bateman and Horn, see [1] and [2]. Let

$$E_{\mathcal{A}}(N) = |\{n \leqq N : n+a_i \text{ prime for } i = 1,\ldots,k\}|,$$

then they conjecture that

$$E_{\mathcal{A}}(N) \sim \prod_p \frac{(1-\frac{\omega(p)}{p})}{(1-\frac{1}{p})^k} \int_2^N \frac{du}{(\log u)^k},$$

Mathematics Subject Classification (2000): 11N05, 11N36.

where $\omega(p)$ is the number of solutions to the congruence

$$\prod_{i=1}^{k}(n+a_i) \equiv 0 \bmod p.$$

Halberstam and Richert [7], Theorem 5.1, proved upper bounds for the number of such tuples.

$$E_{\mathcal{A}}(N) \leqq 2^k \Gamma(k+1) \prod_p \frac{(1-\frac{\omega(p)}{p})}{(1-\frac{1}{p})^k} \frac{N}{(\log N)^k}\left(1+O_k\left(\frac{\log\log 3N + L}{\log N}\right)\right).$$

While in the above statements k is assumed to be constant, we investigate the prime k-tuple conjecture for non-constant k.

Note that the above result by Halberstam and Richert cannot be extended to $k \approx \log N$, for example. There are only very few results in the literature concerned with this question.

The range $k \approx c_1 \log N$ (where c_1 is a positive constant) is of particular interest since previous work emphasized the study of this problem with the set $\mathcal{A} = \{2, 4, \ldots, 2^k\}$. Let $\lfloor x \rfloor$ denote the largest integer a with $a \leqq x$, and let

$$E(N) = \left|\left\{n \leqq N, \text{all } n-2^i \text{ are prime for } i = 1, \ldots, \left\lfloor \frac{\log(n-1)}{\log 2} \right\rfloor\right\}\right|.$$

Using the large sieve method, Vaughan [17] proved that there exists a positive constant c such that

$$(1) \qquad E(N) \ll \frac{N}{\exp(c\,\frac{\log N \log_3 N}{\log_2 N})}.$$

Here $\log_i N$ denotes the i-fold iterated logarithm.

For an *arbitrary* prime pattern of this size, Hofmann and Wolke [9] implicitly obtained the following bound:

$$E_{\mathcal{A}}(N) \ll \frac{N}{\exp(C(c_1)(\log N)^{\frac{1}{2}})}.$$

In the following we improve this bound and obtain a bound comparable to the bound that Vaughan obtained in the special case where the a_i are powers of 2.

Theorem 1. *Let c_1 be a fixed positive constant and let N be sufficiently large. Let $\mathcal{A} = \{a_1, \ldots, a_k\} \subset [1, N]$ be a set of integers with $k \geqq c_1 \log N$. Then the number of those $n \leqq N$ for which all $n + a_i$ are prime is bounded by*

$$E_{\mathcal{A}}(N) \leqq \frac{2N}{\exp((\frac{1}{4}+o(1))\frac{\log N \log_3 N}{\log_2 N})}.$$

The same estimate also holds for primes of type $n - a_i$.

Note that the size of c_1 does not occur in the upper bound. The dependence on c_1 is hidden in the $o(1)$ notation. One would be able to work out more precise bounds with a more careful analysis of $\frac{1}{4} + o(1)$.

On the other hand, it is conjectured that there are only finitely many integers n, such that all $n - 2^i$, with $1 \leqq i < \frac{\log n}{\log 2}$ are prime. Investigations of this problem were started by Erdős [5]. He found that $n = 4, 7, 15, 21, 45, 75$ and 105 have this property and conjectured that 105 is the largest such integer. He verified this up to 203775. These calculations were extended up to 187934724677955 by Mientka and Weitzenkamp [12], and up to 2^{77} by Uchiyama and Yorinaga [16], [18]. No further examples were found. This problem is related to the order of 2 modulo primes. In fact, such an n would be a multiple of all primes p, with 2 as a primitive root and with $2^{p-1} < n$. Thus, on Artin's conjecture on primitive roots, an n of this kind must be a multiple of many small primes. Using the quantitative version of Artin's conjecture, it can be shown that

$$E(N) \ll N^{\alpha+\varepsilon}.$$

Hooley [11] (chapter 7) proved this with $\alpha = 1 - \prod_p (1 - \frac{1}{p(p-1)})$ and Narkiewicz [14] with $\alpha = 1 - \frac{1}{\log 2} \prod_p (1 - \frac{1}{p(p-1)}) \approx 0.46$.

Here we shall show that one can give a much better upper bound, if only one allows to take a few more restrictions: If $n + 2^i$ is required to be prime for $1 \leqq i \leqq c \log N$ and c sufficiently large, we prove, assuming a quantitative version of Artin's conjecture, the best possible $E_{\mathcal{A}}(N) = 0$. Unfortunately, in this variant of the original problem the $a_i = 2^i$ become large, i.e. $a_i \in [1, N^{1.9}]$.

Theorem 2. *Suppose that the Extended Riemann Hypothesis for Dedekind zeta functions holds. Let N be sufficiently large. Then the sequence* $\mathcal{A} = \{2^i : i \in \mathbb{N}\}$ *has the property that there is no* $n \in [1, N]$ *such that all* $n + 2^i$ *are simultaneously prime, for* $1 \leqq i \leqq 2.7 \log N$.

This means that for prime k-tuples of size $c \log N$ we expect considerable oscillations, depending on the set $\mathcal{A}$. Generally, it may be difficult to analyse $\prod_p (1 - \frac{\omega(p)}{p})$. In this direction, the author has proved in [4] that for many primes $p \ll (\log N)^2$ a set $\mathcal{A} \subset [1, N]$ with $|\mathcal{A}| \gg \log N$ lies in $\nu_{\mathcal{A}}(p) \gg p^{1/2}$ distinct residue classes modulo p. Similarly, it follows from the method described below that for many primes $p \ll (\log N)^2 \log_2 N$ we have $\omega(p) \gg \log N$. Therefore, the expression $\prod_p (1 - \frac{\omega(p)}{p})$ in the heuristic formula of Bateman and Horn may be of a dominating influence, for large k.

Remark 1. Pomerance, Sárközy and Stewart [15] proved that for $k < \log N$ there exist sets $\mathcal{A}, \mathcal{B} \subset [1, N]$ with $\mathcal{A} + \mathcal{B} \subset \mathcal{P}$ if $|\mathcal{B}| = k$ and $|\mathcal{A}| < \frac{N}{k(\log N)^k}$. In particular, this implies that there exists such sets of size $|\mathcal{A}| \geqq c_1 \log N$ and $|\mathcal{B}| = k \gg \frac{\log N}{\log \log N}$. The proof is a combinatorial existence proof and the patterns described by the sets $\mathcal{A}$ and $\mathcal{B}$ may depend on N.

Remark 2. While remark 1 describes that there exists patterns of size $c_1 \log N$ that occur at least $\gg \frac{\log N}{\log\log N}$ times in $[1, N]$ we know by theorem 2 an explicit pattern of size $c_1 \log N$ which (on the stated assumption) does not occur at all. Another explicit pattern (which depends on N) is a long arithmetic progression of primes. Suppose that for some $d < N$ all numbers $n + d, n + 2d, \dots, n + kd$ are prime. Then the following argument shows that $k \leqq (1 + o(1)) \log N$ must hold. The common difference d is divisible by all primes $p \leqq k$. Let $P = \prod_{p \leqq k} p$, then $P \mid d$. Suppose that $k \geqq (1 + \varepsilon) \log N$. This would imply that $P = \exp(\sum_{p \leqq k} \log p) > \exp((1 + \frac{\varepsilon}{2}) \log N) > N$ but $d < N$, which is a contradiction. Therefore the size of the largest arithmetic progression of primes in $[1, N]$ is less than $(1 + o(1)) \log N$.

The author would like to thank anonymous referees for comments.

2. The details. A crucial observation is that $\mathcal{A}$ must necessarily lie in many residue classes modulo many small primes. We will make use of Gallagher's larger sieve and of Montgomery's large sieve.

Let us state Montgomery's sieve:

Lemma 1 (Montgomery [13]). *Let $\mathcal{P}$ denote the set of primes. Let $\mathcal{A} \subset [1, N]$ denote a set of integers which lies outside $\nu_{\mathcal{A}}(p)$ residue classes modulo the prime p. Here $\nu_{\mathcal{A}} : \mathcal{P} \to \mathbb{N}$ with $0 \leqq \nu_{\mathcal{A}}(p) \leqq p - 1$. Then the number $A(N)$ of elements satisfies*

$$A(N) \leqq \frac{2N}{L}, \text{ where } L = \sum_{q \leqq N^{1/2}} \mu^2(q) \prod_{p \mid q} \frac{\nu_{\mathcal{A}}(p)}{p - \nu_{\mathcal{A}}(p)}.$$

Vaughan [17] gives a suitable evaluation of L if $\sum_{p \leqq y} \frac{\nu_{\mathcal{A}}(p)}{p}$ is known.

Lemma 2 (Vaughan [17]). *The following lower bound holds:*

$$L \geqq \sum_m \exp\left(m \log\left(\frac{1}{m} \sum_{p \leqq N^{1/(2m)}} \frac{\nu_{\mathcal{A}}(p)}{p}\right)\right).$$

The size of this sum can be approximated by choosing a value of m which maximizes the summand. Since $p \geqq 2$ we can assume that $1 \leqq m \leqq \frac{\log(N^{1/2})}{\log 2}$.

We recall Gallagher's larger sieve.

Lemma 3 (Gallagher [6]). *Let S denote a set of primes such that $\mathcal{A} \subseteqq [1, N]$ lies in at most $\nu_{\mathcal{A}}(p)$ residue classes modulo p (for $p \in S$). Then the following inequality holds, provided the denominator is positive:*

$$A(N) \leqq \frac{-\log N + \sum_{p \in S} \log p}{-\log N + \sum_{p \in S} \frac{\log p}{\nu_{\mathcal{A}}(p)}}.$$

Proof of Theorem 1. We first apply Lemma 3 to the set $\mathcal{A}$ that we use in Theorem 1. Put $y_0 = \lfloor 2\log N \rfloor$, $m = \lfloor \frac{\log N}{4\log_2 N + 2\log_3 N} \rfloor$ and $y = N^{\frac{1}{2m}}$ so that $(\log N)^2 \log_2 N \ll y \ll (\log N)^2 \log_2 N$. The set of primes used in Lemma 3 is defined to be

$$S = \{p \in [y_0, y] : \nu_{\mathcal{A}}(p) \leqq c_2 \log N\}$$

and similarly we define

$$T = \{p \in [y_0, y] : \nu_{\mathcal{A}}(p) > c_2 \log N\}.$$

Suppose that $\sum\limits_{p \in S} \log p \geqq c_3 y$ for some positive constant $c_3 > \frac{c_2}{c_1} + \varepsilon > 0$. Note that for fixed c_1 we may choose $\varepsilon > 0$ and c_2 such that c_3 can be arbitrarily small. We then have, for $N \geqq N(\varepsilon)$,

$$A(N) \leqq \frac{-\log N + y}{-\log N + \frac{c_3 y}{c_2 \log N}} \leqq \frac{c_2}{c_3 - \varepsilon} \log N < c_1 \log N,$$

which contradicts $A(N) = k \geqq c_1 \log N$.

So the set S contains only an arbitrarily small proportion of the primes of the interval $[y_0, y]$. Therefore, we find that

$$\sum_{p \in T} \frac{\nu_{\mathcal{A}}(p)}{p} \geqq \frac{1}{2} c_2 \log N (\log\log y - \log\log y_0).$$

We also have

$$\begin{aligned} \log\log y - \log\log y_0 &\sim \log(2\log_2 N + \log_3 N + O(1)) - \log\log(2\log N) \\ &= \log 2 + \log_3 N + o(1) - (\log_3 N + o(1)) \geqq c_4. \end{aligned}$$

In a second step we apply Lemma 2 to those integers n to be counted in Theorem 1. We obtain

$$\begin{aligned} L &\geqq \exp\left((m + o(1)) \log \left(\frac{1}{m} \sum_{p \leqq y} \frac{\nu_{\mathcal{A}}(p)}{p} \right) \right) \\ &\geqq \exp\left((m + o(1)) \log \left(\frac{2\log y}{\log N} \frac{c_2}{2} c_4 \log N \right) \right) \\ &\geqq \exp\left(\frac{\log N}{(4 + o(1)) \log_2 N} \log \left(c_5 \log_2 N \right) \right) \\ &\geqq \exp\left(\left(\frac{1}{4} + o(1) \right) \frac{\log N \log_3 N}{\log_2 N} \right), \end{aligned}$$

for sufficiently large N. An application of Lemma 1 shows that $E_{\mathcal{A}}(N) \leqq \frac{2N}{L}$, which establishes the theorem.

In the case of $n - a_i$ we proceed as before for intervals $[2^i, 2^{i+1}]$. Summing up over the $\lfloor \frac{\log N}{\log 2} \rfloor$ intervals gives the same upper bound.

Using partial summation for the estimation of $\sum\limits_{p \leqq y} (\frac{\log p}{p})^{1/2}$ and an application of the Cauchy-Schwarz inequality would lead to

$$\left(\sum_{p \leqq y} \frac{\log p}{\nu_{\mathcal{A}}(p)} \right) \left(\sum_{p \leqq y} \frac{\nu_{\mathcal{A}}(p)}{p} \right) \geqq \left(\sum_{p \leqq y} \left(\frac{\log p}{p} \right)^{1/2} \right)^2 \gg \frac{y}{\log y}$$

only and thus to a bound weaker by the $\log_3 N$ factor. $\square$

P r o o f o f T h e o r e m 2. It was proved by Hooley [11] that the Extended Riemann Hypothesis for certain Dedekind zeta functions implies a strong form of Artin's conjecture on primitive roots. Let $F_2(N)$ denote the number of primes $p \leqq N$ such that 2 is a primitive root of p. Hooley proved under this assumption that $F_2(N) = \prod\limits_p (1 - \frac{1}{p(p-1)}) \frac{N}{\log N} + O(\frac{N \log \log N}{(\log N)^2})$. By partial summation we deduce that $F(N) := \sum\limits_{p \leqq N, \mathrm{ord}_p(2) = p-1} \log p \sim \prod\limits_p (1 - \frac{1}{p(p-1)}) N \geqq 0.373N$, for sufficiently large N. Let $\mathcal{A} = \{2^i : i \in \mathbb{N}\}$. Since $n + 2^i$ shall be prime we have for all primes $p \leqq 2.7 \log N$ with $\mathrm{ord}_p(2) = p - 1$ that n lies in only one residue class (namely the zero class) mod p. Because of

$$\prod_{\substack{p \leqq 2.7 \log N \\ \mathrm{ord}_p(2) = p-1}} p = \exp \sum_{\substack{p \leqq 2.7 \log N \\ \mathrm{ord}_p(2) = p-1}} \log p \geqq \exp(2.7 \times 0.373 \log N) > N$$

it follows by an application of the Chinese Remainder Theorem that there cannot be any such $n \in [1, N]$. (An application of Gallagher's larger sieve would also show that $E_{\mathcal{A}}(N) = O(1)$). Moreover $a_i \leqq 2^{2.7 \log N} < N^{1.9}$. $\square$

References

[1] P. T. BATEMAN and R. A. HORN, A heuristic asymptotic formula concerning the distribution of prime numbers. Math. Comp. **16**, 363–367 (1962).

[2] P. T. BATEMAN and R. A. HORN, Primes represented by irreducible polynomials in one variable. Proc. Sympos. Pure Math., Vol. VIII, 119–132, Amer. Math. Soc. Providence, R.I. 1965.

[3] C. ELSHOLTZ, The inverse Goldbach problem. Mathematika **48**, 151–158 (2001).

[4] C. ELSHOLTZ, The distribution of sequences in residue classes. Proc. Amer. Math. Soc. **130**, 2247–2250 (2002).

[5] P. ERDŐS, On integers of the form $2^k + p$ and some related problems. Summa Brasil. Math. **2**, 113–123 (1950).

[6] P. X. GALLAGHER, A larger sieve. Acta Arith. **18**, 77–81 (1971).

[7] H. HALBERSTAM and H. -E. RICHERT, Sieve methods. London Math. Soc. Monographs **4**, London 1974.

[8] G. H. HARDY and J. E. LITTLEWOOD, Some problems of Partitio Numerorum: III, On the expression of a number as a sum of primes. Acta Mathematica **44**, 1–70 (1922).

[9] A. HOFMANN and D. WOLKE, On additive decompositions of the set of primes. Arch. Math. **67**, 379–382 (1996).
[10] C. HOOLEY, On Artin's conjecture. J. Reine Angew. Math. **225**, 209–220 (1967).
[11] C. HOOLEY, Applications of sieve methods to the theory of numbers. Cambridge Tracts in Math. **70**, Cambridge 1976.
[12] W. MIENTKA and R. WEITZENKAMP, On f-plentiful numbers. J. Combin. Theory **7**, 374–377 (1969).
[13] H. L. MONTGOMERY, The Analytic Principle of the Large Sieve. Bull. Amer. Math. Soc. **84**, 547–567 (1978).
[14] W. NARKIEWICZ, On a conjecture of Erdős. Colloq. Math. **37**, 313–315 (1977).
[15] C. POMERANCE, A. SÁRKÖZY and C. L. STEWART, On divisors of sums of integers III. Pacific J. Math. **133**, 363–379 (1988).
[16] S. UCHIYAMA and M. YORINAGA, Notes on a conjecture of P. Erdős I. Math. J. Okayama Univ. **19**, 129–140 (1976/77).
[17] R. C. VAUGHAN, Some Applications of Montgomery's Sieve. J. Number Theory **5**, 64–79 (1973).
[18] M. YORINAGA and S. UCHIYAMA, Notes on a conjecture of P. Erdős II. Math. J. Okayama Univ. **20**, 41–49 (1978/79).

Received: 20 December 2002

Christian Elsholtz
Department of Mathematics
Royal Holloway University of London
Egham
Surrey TW20 0EX
United Kingdom
christian.elsholtz@rhul.ac.uk