ACTA ARITHMETICA
122.1 (2006)

On integers of the form $p + 2^k$

by

LAURENT HABSIEGER and XAVIER-FRANÇOIS ROBLOT (Lyon)

1. Introduction. Thoughout this paper, the symbol p will denote a prime and k will be a nonnegative integer. Romanov [5] proved that the integers of the form $p+2^k$ have positive density. He also raised the following question: does there exist an arithmetic progression consisting only of odd numbers, no term of which is of the form $p + 2^k$? Erdős [1] found such an arithmetic progression by considering integers which are congruent to 172677 modulo $5592405 = (2^{24} - 1)/3$. Thus the density of numbers of the form $p+2^k$ is less than $1/2$, the trivial bound obtained from the odd integers. For convenience we introduce

$$\underline{d} = \liminf_{x\to\infty} \frac{\#\{p + 2^k \le x\}}{x/2} \quad \text{and} \quad \bar{d} = \limsup_{x\to\infty} \frac{\#\{p + 2^k \le x\}}{x/2}.$$

The aim of this paper is to give an explicit version of the estimates $0 < \underline{d} \le \bar{d} < 1$.

THEOREM 1. *We have*

$$0.1866 < \underline{d} \le \bar{d} < 0.9819.$$

This range is pretty large and Bombieri conjectured the more precise upper bound 0.868 (see [4]).

In Section 2, we obtain the lower bound $0.1866 < \underline{d}$, by slightly refining a straightforward application of a recent result of Pintz and Ruzsa [3], in their study of Linnik's approximation of the Goldbach problem (see also [2]). In Section 3, we get the upper bound, using computations on residue classes.

2. The lower bound. Let N be a large integer and put $L = \lfloor \log N/\log 2 \rfloor$. Define the functions

$$r(n) = \#\{(p, k) : n = p + 2^k,\, p \le N,\, 1 \le k \le L\}$$

2000 *Mathematics Subject Classification*: Primary 11P32; Secondary 11Y35, 11Y60.
Partially supported by the European Community IHRP Program, within the Research Training Network "Algebraic Combinatorics in Europe", grant HPRN-CT-2001-00272.

and

$$s(N)=\#\{(p_1,p_2,k_1,k_2):p_1-p_2=2^{k_2}-2^{k_1},\ p_j\le N,\ 1\le k_j\le L,\ j=1,2\}$$

so that

$$s(N)=\sum_{n=1}^{N} r^2(n).$$

Pintz and Ruzsa [3] proved the following lemma.

LEMMA 1. *For N large enough, we have*

$$s(N)\le\frac{2}{\log^2 2}\,CN,$$

where $C<5.3636$.

Let $d(N)$ denote the number of positive integers $n\le N$ which may be written in the form $n=p+2^k$. The Cauchy–Schwarz inequality implies easily that

$$(\pi(N)L)^2\le d(N)s(N),$$

where $\pi(N)$ denotes the number of primes $p\le N$. We deduce from Lemma 1 and from the prime number theorem that $2Cd(N)\ge(1+o(1))N$, and the lower bound $\underline{d}\ge 1/C>0.1864$ follows from the definitions.

To get the bound from the theorem, we need further notations. Put

$$\varepsilon_N=\frac{\sum_{1\le n\le N,\,r(n)>0} r(n)}{\sum_{1\le n\le N,\,r(n)>0} 1}\quad\text{and}\quad\varepsilon=\frac{2}{\underline{d}\log 2}.$$

By the definitions, there exists a subsequence of $(\varepsilon_N)_{N\in\mathbb{N}}$ which converges to ε. Let us now refine the Cauchy–Schwarz inequality by studying

$$\Delta_N=\sum_{1\le n\le N,\,r(n)>0}(r(n)-\varepsilon_N)^2,$$

so that

$$\begin{aligned}\Delta_N&=\sum_{1\le n\le N} r^2(n)-\frac{(\sum_{1\le n\le N} r(n))^2}{\sum_{1\le n\le N,\,r(n)>0} 1}=s(N)-\frac{(\pi(N)L)^2}{d(N)}\\&\le\left(5.3636-\frac{1}{\underline{d}}+o(1)\right)\frac{2N}{\log^2 2}\end{aligned}$$

for infinitely many N. Without loss of generality we may assume that $\varepsilon\in\,]15,15.5[$: otherwise we would get either $\underline{d}\ge 0.19$, which would be better, or

$\underline{d} \le 0.1862$, which is false. For infinitely many N we thus have

$$\varDelta_N \ge \sum_{1\le n\le N,\, r(n)>0} (15-\varepsilon_N)^2 \ge \Big(\sum_{1\le n\le N,\, r(n)>0} (15-\varepsilon)^2 + o(1)\Big)N$$
$$= \bigg(\frac{\underline{d}}{2}\bigg(15 - \frac{2}{\underline{d}\log 2}\bigg)^2 + o(1)\bigg)N.$$

We deduce from these estimates the inequality

$$\frac{\underline{d}}{2}\bigg(15 - \frac{2}{\underline{d}\log 2}\bigg)^2 \le \frac{2}{\log^2 2}\bigg(5.3636 - \frac{1}{\underline{d}}\bigg),$$

which may be written as

$$56.25\log^2 2\underline{d}^2 - (15\log 2 + 5.3636)\underline{d} + 1 \le 0.$$

The lower bound $\underline{d} \ge 0.1866$ then follows.

3. The upper bound

A. *Basic ideas.* Let us introduce further notations. Let M be a positive odd integer and let ω denote the order of 2 in $(\mathbb{Z}/M\mathbb{Z})^*$. For $\overline{m}$ a residue class modulo M, put

$$f_M(\overline{m}) = \{\bar{k} \in \mathbb{Z}/\omega\mathbb{Z} : \overline{m} - 2^{\bar{k}} \in (\mathbb{Z}/M\mathbb{Z})^*\}$$

and

$$\delta_M(\nu) = |\{\overline{m} \in \mathbb{Z}/M\mathbb{Z} : |f_M(\overline{m})| = \nu\}|.$$

The basic tool to get an upper bound for $\bar{d}$ is the following lemma.

Lemma 2. *With the previous notations, we have*

$$\bar{d} \le \sum_{\nu=0}^{\omega} \delta_M(\nu)\min\bigg(\frac{1}{M}, \frac{2\nu}{\omega\varphi(M)\log 2}\bigg),$$

where φ denotes Euler's function.

Proof. Let $\overline{m}$ be a congruence class modulo M, with $|f_M(\overline{m})| = \nu$. Let us study the proportion of odd integers congruent to $\overline{m}$ that may be written in the form $p + 2^k$. This proportion is clearly at most $1/M$, and we only need to prove the alternative upper bound.

Since all the primes but a finite number are invertible modulo M, there exist ν congruence equations $\overline{m} = \overline{p}_i + 2^{\bar{k}_i}$, $i \in \{1, \ldots, \nu\}$, such that all but finitely many representations $p + 2^k$ come from one of these congruence equations. The number of primes up to N which are congruent to p_i modulo M is asymptotic to $N/(\varphi(M)\log N)$, while the number of powers of 2 which are congruent to 2^{k_i} modulo M is asymptotic to $\log N/(\omega\log 2)$. Thus the number of integers congruent to $\overline{m}$ that may be written in the form $p + 2^k$ is at most $(\nu/(\varphi(M)\omega\log 2) + o(1))N$. This implies that the proportion of odd integers enjoying these properties is at most $2\nu/(\varphi(M)\omega\log 2)$ and the lemma follows. ∎

This lemma provides a nontrivial upper bound for $\bar{d}$ as soon as there exist residue classes $\overline{m}$ modulo M such that

$$f_M(\overline{m}) < \frac{\omega\varphi(M)\log 2}{2M}, \tag{1}$$

a condition that occurs for a small number of classes. The main problem is to compute the distribution of the $f_M(\overline{m})$'s in an efficient way. The direct computation of all the $f_M(\overline{m})$'s is quickly limited by memory problems. However one can obtain significant results this way.

Take $M = 23205 = (2^{24}-1)/723$, so that $\omega = 24$ and $\varphi(M) = 9216$. Condition (1) is equivalent to $f_M(\overline{m}) \le 3$. We find

$$(\delta_M(0), \delta_M(1), \delta_M(2), \delta_M(3)) = (0, 48, 720, 320),$$

and we get this way $\bar{d} \le 0.985049$.

B. *Refined algorithms and results.* It appears that the function f_M takes very few possible values, when compared to the set of subsets of $\mathbb{Z}/\omega\mathbb{Z}$. So let us introduce

$$g_M(I) = \{\overline{m} \in \mathbb{Z}/M\mathbb{Z} : f_M(\overline{m}) = I\} \quad \text{and} \quad G_M(I) = |g_M(I)|$$

for $I \subset \mathbb{Z}/\omega\mathbb{Z}$. Note that

$$\delta_M(\nu) = \sum_{|I|=\nu} G_M(I).$$

So it is sufficient to know the distribution of the $G_M(I)$'s to compute an upper bound for $\bar{d}$.

The main advantage of the function g_M is that it is easily computable by induction on the number of prime factors of M. The initial case is given by $g_0(\{0\}) = \{0\}$.

Let M_1, M_2 be two positive odd squarefree integers, with $M_2 = pM_1$ for some prime p not dividing M_1. Let ω_1, ω_2 and ω_p denote the order of 2 in $(\mathbb{Z}/M_1\mathbb{Z})^*$, $(\mathbb{Z}/M_2\mathbb{Z})^*$ and $(\mathbb{Z}/p\mathbb{Z})^*$, respectively. The image of f_p is easy to compute. There is the subset

$$I_{p,0} = \{\overline{2}^{\bar{k}} \in (\mathbb{Z}/p\mathbb{Z})^* : \bar{k} \in \mathbb{Z}/\omega_p\mathbb{Z}\}$$

with $G_p(I_{p,0}) = p - \omega_p$, and for each $\bar{j} \in \mathbb{Z}/\omega_p\mathbb{Z}$ the subset

$$I_{p,\bar{j}} = \{\overline{2}^{\bar{k}} \in (\mathbb{Z}/p\mathbb{Z})^* : \bar{k} \in \mathbb{Z}/\omega_p\mathbb{Z},\ \bar{k} \ne \bar{j}\}$$

with $G_p(I_{p,\bar{j}}) = 1$. Now, let I_2 and I_p be in the image of f_{M_2} and f_p respectively. Denote by $\widetilde{I_2}$ and $\widetilde{I_p}$ the subsets of $\mathbb{Z}/M_1\mathbb{Z}$ which are inverse images of I_2 and I_p under the map on subsets induced by the natural surjections

$\mathbb{Z}/M_1\mathbb{Z} \to \mathbb{Z}/M_2\mathbb{Z}$ and $\mathbb{Z}/M_1\mathbb{Z} \to \mathbb{Z}/p\mathbb{Z}$ respectively. Then it is easy to see that $\widetilde{I}_2 \cap \widetilde{I}_p$ is in the image of f_{M_1} with

$$G_{M_1}(\widetilde{I}_2 \cap \widetilde{I}_p) = G_{M_2}(I_2)G_p(I_p),$$

and that all subsets in the image of f_{M_1} are obtained in this way.

This construction allows us to build recursively the image of f_M. It also enables us to find how many classes have the same image. Therefore, one can compute $G_M(I)$ without knowing $g_M(I)$.

Let us give an example. For

$$M = 5592405 = 3 \cdot 5 \cdot 7 \cdot 13 \cdot 17 \cdot 241 = (2^{24} - 1)/3,$$

we have $\omega = 24$. There are 16401 subsets in the image of f_M, which is much fewer than 2^{24}. Each of these subsets is obtained in r ways, with $1 \leq r \leq 250068$. Only subsets of cardinality at most 3 lead to an improved upper bound. The empty set appears 48 times. Each of the singletons from $\mathbb{Z}/24\mathbb{Z}$ appears 540 times. For 2-subsets, the situation is slightly more complicated to describe. The subsets of the form $\{a, a \pm 8\}$ appear 3625 times (there are 24 of them) while those of the form $\{a, a + 12\}$ appear 7170 times (there are 12 of them). There are 224 interesting 3-subsets, appearing 3, 6, 225 or 9520 times.

This method requires much less memory than the algorithm from the previous subsection. It is still possible to save a bit more memory. Indeed, the representation problem (by an invertible plus a power of 2) is invariant when multiplied by a power of 2. So we can use a representative of a collection of subsets, each of them being obtained by translation from the representative, instead of subsets of $\mathbb{Z}/\omega\mathbb{Z}$.

The best result found so far is given by

$$M = 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19 \cdot 31 \cdot 41 \cdot 73 \cdot 241 \cdot 257.$$

It leads to the improvement

$$\bar{d} < 0.9818818607968211912960156368,$$

and the upper bound from Theorem 1 follows. This computation took 35 minutes on an Intel Xeon 2.4 GHz with a memory stack of 2.1 GB. Indeed, the real limitation is the memory. Note that during the computations, subsets for which $G_M(I)$ was quite large and thus unlikely to contribute to the density were dropped (still there were a total of 4469837 different subsets at the end). Hence the density obtained may be a little greater than the actual density for this value of M.

Addendum. The referee informed the authors that, while the paper was being refereed, János Pintz improved on the lower bound. In a paper to appear in Acta Math. Hungar., he showed $\underline{d} \geq 0.18734$ by a more elaborate method.

References

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

[2] D. R. Heath-Brown and J.-C. Puchta, *Integers represented as a sum of primes and powers of two*, Asian J. Math. 6 (2002), 535–565.

[3] J. Pintz and I. Z. Ruzsa, *On Linnik's approximation to Goldbach's problem, I*, Acta Arith. 109 (2003), 169–194.

[4] F. Romani, *Computations concerning primes and powers of two*, Calcolo 20 (1983), 319–336.

[5] N. P. Romanov, *Über einige Sätze der additiven Zahlentheorie*, Math. Ann. 109 (1934), 668–678.

Institut Camille Jordan
CNRS UMR 5208 Mathématiques
Université Claude Bernard Lyon 1
43 boulevard du 11 novembre 1918
69622 Villeurbanne Cedex, France
E-mail: Laurent.Habsieger@math.univ-lyon1.fr
roblot@math.univ-lyon1.fr

Received on 8.2.2005
and in revised form on 20.12.2005 (4934)