Reprinted from JOURNAL OF NUMBER THEORY
All Rights Reserved by Academic Press, New York and London

Vol. 21, No. 3, December 1985
Printed in Belgium

On Locally Repeated Values of Certain Arithmetic Functions, I

P. ERDÖS AND A. SÁRKÖZY

Mathematical Institute of the Hungarian Academy of Sciences, Budapest, Hungary

AND

C. POMERANCE*

Department of Mathematics, University of Georgia, Athens, Georgia 30602

Communicated by P. T. Bateman

Received September 12, 1983

Let $\nu(n)$ denote the number of distinct prime factors of n. We show that the equation $n+\nu(n)=m+\nu(m)$ has many solutions with $n \neq m$. We also show that if ν is replaced by an arbitrary, integer-valued function f with certain properties assumed about its average order, then the equation $n+f(n)=m+f(m)$ has infinitely many solutions with $n \neq m$. © 1985 Academic Press, Inc.

1. INTRODUCTION

If $f(n)$ is an arithmetic function, one can ask for the distribution of the integers n for which $f(n)=f(n+1)$. Depending on the function f, this question is usually either trivial or intractable. However, even when the conjectured "truth" is unobtainable, partial results are sometimes possible. Also easier questions can be asked, such as: find the distribution of the n for which $|f(n)-f(n+1)|$ is small or find the distribution of the pairs n, m for which $f(n)=f(m)$ and $|n-m|$ is small.

The aim of this paper is to study the equation

$$n+\nu(n)=m+\nu(m) \qquad (n \neq m) \tag{1}$$

and some related questions, where $\nu(n)$ is the number of distinct prime factors of n. Note that if n, m is a solution of (1), then certainly $|n-m|$ is

* Research supported in part by an NSF grant.

0022-314X/85 $3.00

Copyright © 1985 by Academic Press, Inc.
All rights of reproduction in any form reserved.

small. We show that not only does (1) have many solutions, but a generalization of (1), where $\nu(n)$ is replaced by an arbitrary, integer-valued function $f(n)$ with certain properties assumed about its average order, always has infinitely many solutions.

Let $\Omega(n)$ denote the number of prime factors of n counted according to multiplicity and let $\tau(n)$ denote the number of divisors of n. Recently, D. R. Heath-Brown ("The divisor function at consecutive integers," to appear) showed that $\tau(n)=\tau(n+1)$ has infinitely many solutions. He announced that his method also gives infinitely many solutions of $\Omega(n)=\Omega(n+1)$. In a later paper in this series we shall show that the number of $n \leqslant x$ for which $\tau(n)=\tau(n+1)$ is $O(x/\sqrt{\log\log x})$ and the same for $\Omega(n)$ and $\nu(n)$. In another paper we shall show that $|\nu(n)-\nu(n+1)|$ is bounded on at least $cx/\sqrt{\log\log x}$ values of $n \leqslant x$ and the same for $\Omega(n)$. We shall also obtain an upper bound for the number of solutions of (1) and of the equation $\phi(n)=\phi(n+1)$, where ϕ is Euler's function.

Section 2 below will be devoted to the following theorem.

THEOREM 1. *Let $f(n)$ be a positive integer-valued arithmetic function for which there is a differentiable function $F(x)$ and an x_0, such that for $x > x_0$,*

$$\text{(i)} \qquad \left|\sum_{n \leqslant x} f(n) - xF(x)\right| < \min\left\{\frac{x}{80}, \frac{1}{240F'(x)}\right\},$$

$$\text{(ii)} \qquad (\max_{n \leqslant x} f(n))^2 < \min\left\{\frac{x}{80}, \frac{1}{240F'(x)}\right\},$$

$$\text{(iii)} \qquad 0 < F'(x) < \frac{1}{60}, \qquad F'(x/2) < 3F'(x),$$

$$\text{(iv)} \qquad F'(x) \text{ is decreasing},$$

$$\text{(v)} \qquad \lim_{x \to +\infty} F(x) = +\infty.$$

Then the equation

$$n + f(n) = m + f(m) \qquad (n \neq m) \tag{2}$$

has infinitely many solutions.

COROLLARY. *Each of the equations*

$$n + \nu(n) = m + \nu(m) \qquad (n \neq m)$$

$$n + \Omega(n) = m + \Omega(m) \qquad (n \neq m)$$

$$n + \tau(n) = m + \tau(m) \qquad (n \neq m)$$

has infinitely many solutions.

In fact, the Corollary can be derived easily from Theorem 1 by using the well-known formulas

$$\sum_{n \leqslant x} \nu(n) = x \log\log x + c_1 x + O\left(\frac{x}{\log x}\right),$$

$$\sum_{n \leqslant x} \Omega(n) = x \log\log x + c_2 x + O\left(\frac{x}{\log x}\right),$$

$$\sum_{n \leqslant x} \tau(n) = x \log x + c_3 x + O(\sqrt{x}).$$

It is not hard to construct examples to show that condition (i) on the average-order function $F(x)$ stated in Theorem 1 is nearly best possible. For example, let $f(n) = [\log\log(n+2)]$, so that it is clear that $n + f(n) = m + f(m)$ has no solutions $n \neq m$. We have

$$\sum_{n \leqslant x} f(n) = x \log\log x + O(x).$$

If we let $F(x) = \log\log x$, then every condition of the theorem is satisfied except that it is not true that

$$\left| \sum_{n \leqslant x} f(n) - x \log\log x \right| < \frac{x}{80} \qquad \text{for } x > x_0.$$

Rather

$$\left| \sum_{n \leqslant x} f(n) - x \log\log x \right| \leqslant (1 + o(1))\, x,$$

so that apart from the precise value of the constant our condition is best possible.

Another example is $f(n) = [n^{1/4}]$. It is again immediate that $n + f(n) = m + f(m)$ has no solutions $n \neq m$. We have

$$\sum_{n \leqslant x} f(n) = \tfrac{4}{5} x^{5/4} - \tfrac{1}{2} x - (\tfrac{1}{3} + 2\{x^{1/4}\}^2 - 2\{x^{1/4}\})\, x^{3/4} + O(x^{1/2}).$$

If we let $F(x) = \frac{4}{5} x^{1/4} - \frac{1}{2} - \frac{1}{12} x^{-1/4}$, then every condition of Theorem 1 is satisfied except that

$$\left| \sum_{n \leqslant x} f(n) - xF(x) \right| < \frac{1}{240 F'(x)} \qquad \text{for } x > x_0$$

fails. Rather

$$\left| \sum_{n \leqslant x} f(n) - xF(x) \right| \leqslant \frac{1}{(20 + o(1))\, F(x)},$$

so that again, apart from the precise value "1/240," the condition of the theorem is best possible.

It is interesting to note that Theorem 1 can be used to give an "Ω-theorem" for the mean value of some non-decreasing, integer-valued functions $f(n)$. Indeed, the conclusion of the theorem is not satisfied, so at least one hypothesis for the function $F(x)$ must also fail. For example, if $f(n)$ is integer-valued, non-decreasing, and $f(n) \sim \log\log n$ as $n \to \infty$, then for every value of the constant c we have

$$\limsup_{x \to \infty} \frac{1}{x} \left| \sum_{n \leqslant x} f(n) - x \log\log x - cx \right| \geqslant \frac{1}{80}.$$

In Sections 3 and 4 we shall prove (by a different method) that Eq. (1) has "many" solutions:

THEOREM 2. *There is an x_1 such that for $x > x_1$, the equation*

$$n + v(n) = m + v(m), \qquad n \leqslant x, m \leqslant x, n \neq m$$

has more than $x \cdot \exp\{-4000 \log\log x \log\log\log x\}$ solutions.

2. Proof of Theorem 1

Assume that the hypotheses of Theorem 1 hold. For $k = 1, 2, \ldots$, let $\mathscr{S}(k)$ denote the set of integers n such that

$$0 < n \leqslant k \qquad \text{and} \qquad n + f(n) \geqslant k + 1.$$

Denote by $g(k)$ the number of elements of the set $\mathscr{S}(k)$.

Assume first that for some positive integer k,

$$g(k) > g(k+1). \tag{3}$$

Clearly $k + 1 \in \mathscr{S}(k+1)$ and $k + 1 \notin \mathscr{S}(k)$, so that (3) implies that $\mathscr{S}(k) - \mathscr{S}(k+1)$ has at least two elements. That is, (3) implies there exist integers m, n with $m \neq n$, such that both are in $\mathscr{S}(k)$ and neither is in $\mathscr{S}(k+1)$. Then clearly, we have

$$m + f(m) = n + f(n) = k + 1,$$

giving a solution of (2). Furthermore, distinct values of k determine distinct solutions of (2). Thus if (3) holds for infinitely many k, then also (2) has infinitely many solutions.

So let us assume the opposite, that

$$g(k) \leqslant g(k+1) \qquad \text{for } k \geqslant k_0. \tag{4}$$

We now show that the functions g and f have about the same average order. We have for $x > x_0$

$$\begin{aligned}\left|\sum_{k\leqslant x} g(k) - \sum_{n\leqslant x} f(n)\right| &= \left|\sum_{k\leqslant x}\ \sum_{\substack{n\leqslant k\\ n+f(n)\geqslant k+1}} 1 - \sum_{n\leqslant x} f(n)\right| \\ &= \left|\sum_{n\leqslant x}\left(-f(n) + \sum_{\substack{n\leqslant k<n+f(n)\\ k\leqslant x}} 1\right)\right| \\ &= \left|\sum_{\substack{n\leqslant x\\ n+f(n)>x+1}} (-f(n)+[x]+1-n)\right| \\ &\leqslant \sum_{\substack{n\leqslant x\\ n>x-f(n)+1}} f(n) < (\max_{n\leqslant x} f(n))^2 \\ &< \min\left\{\frac{x}{80}, \frac{1}{240F'(x)}\right\},\end{aligned}$$

by condition (ii) of Theorem 1. Thus by condition (i), for $x > x_0$

$$\begin{aligned}\left|\sum_{k\leqslant x} g(k) - xF(x)\right| &\leqslant \left|\sum_{k\leqslant x} g(k) - \sum_{n\leqslant x} f(n)\right| + \left|\sum_{n\leqslant x} f(n) - xF(x)\right| \\ &< 2\cdot\min\left\{\frac{x}{80}, \frac{1}{240F'(x)}\right\} = \min\left\{\frac{x}{40}, \frac{1}{120F'(x)}\right\}.\end{aligned} \tag{5}$$

By (5) and conditions (iii) and (v) of Theorem 1 it follows that $g(k)$ is not bounded, so that for all x_1 there is an integer k_1 with

$$k_1 > x_1,\ g(k_1) < g(k_1+1). \tag{6}$$

Let x_1 be any number larger than $2\cdot\max\{x_0, k_0, 10\}$ and let k_1 be such that (6) holds. Put

$$x = \frac{4}{3}k_1, \qquad t = \left[\min\left\{\frac{x}{4}, \frac{1}{12F'(x)}\right\}\right].$$

Then by (4) and (6) we have

$$\sum_{k=k_1+1}^{k_1+t} g(k) - \sum_{k=k_1-t+1}^{k_1} g(k) \geqslant t. \tag{7}$$

On the other hand, by (5), we have

$$\begin{aligned}
&\sum_{k=k_1+1}^{k_1+t} g(k) - \sum_{k=k_1-t+1}^{k_1} g(k) \\
&\quad = \sum_{k \leqslant k_1+t} g(k) - 2\sum_{k\leqslant k_1} g(k) + \sum_{k \leqslant k_1 - t} g(k) \\
&\quad \leqslant (k_1+t)\,F(k_1+t) - 2k_1 F(k_1) + (k_1 - t)\,F(k_1 - t) \\
&\qquad + 4\cdot \min\left\{\frac{x}{40}, \frac{1}{120F'(x)}\right\} \\
&\quad \leqslant k_1(F(k_1+t) - 2F(k_1) + F(k_1-t)) + t(F(k_1+t) - F(k_1-t)) + \frac{t}{2}
\end{aligned}$$

Now condition (iv) of Theorem 1 implies F is concave downward, so that

$$\begin{aligned}
\sum_{k=k_1+1}^{k_1+t} g(k) - \sum_{k=k_1-t+1}^{k_1} g(k) &< t(F(k_1+t) - F(k_1-t)) + \frac{t}{2} \\
&< 2t^2F'(k_1 - t) + \frac{t}{2} \leqslant 2t^2 F'\left(\frac{3}{4}x - \frac{1}{4}x\right) + \frac{t}{2} \\
&< 6t^2F'(x) + \frac{t}{2},
\end{aligned} \tag{8}$$

by condition (iii) of Theorem 1. Thus from (7) and (8),

$$t < 6t^2F'(x) + \frac{t}{2},$$

so that

$$t > \frac{1}{12F'(x)}. \tag{9}$$

But (9) contradicts the definition of t, which shows that there is no k_0 for which (4) holds. Thus (3) holds for infinitely many k, which, as we have seen, is sufficient for (2) to have infinitely many solutions.

3. Preparation for Theorem 2

We shall use the following lemmas in the proof of Theorem 2.

Lemma 1. *Let $\mathscr{P}$ be an arbitrary set of primes. Let $x \geq 1$ and let a, b be integers with $0 \leq a < b$ such that every prime factor of b is in $\mathscr{P}$. Let*

$$g(n) = \sum_{\substack{p \leq x^{1/3}, p \notin \mathscr{P} \\ p \mid n}} 1, \qquad E = \sum_{p \leq x^{1/3}, p \notin \mathscr{P}} \frac{1}{p}.$$

Then

$$\left| \sum_{n \leq x} (g(a+bn) - E)^2 - Ex \right| < c_4 x \tag{10}$$

where the constant c_4 is absolute (independent of each $\mathscr{P}$, a, b, x).

Proof. The Lemma can be proved easily by Turán's method; see [1, 2]. For the sake of completeness we give the proof.

We may assume that x is an integer. Then

$$\sum_{n \leq x} (g(a+bn) - E)^2 = \sum_{n \leq x} g(a+bn)^2 - 2E \sum_{n \leq x} g(a+bn) + xE^2. \tag{11}$$

The first sum is

$$\begin{aligned}
&\sum_{n \leq x} g(a+bn)^2 \\
&= \sum_{n \leq x} \Bigg(\sum_{\substack{p \leq x^{1/3}, p \notin \mathscr{P} \\ p \mid a+bn}} 1 \Bigg)^2 \\
&= \sum_{n \leq x} \sum_{\substack{p \leq x^{1/3}, p \notin \mathscr{P} \\ p \mid a+bn}} 1 + \sum_{n \leq x} \sum_{\substack{p,q \leq x^{1/3}, p,q \notin \mathscr{P} \\ p \neq q, pq \mid a+bn}} 1 \\
&= \sum_{p \leq x^{1/3}, p \notin \mathscr{P}} \sum_{\substack{n \leq x \\ p \mid a+bn}} 1 + \sum_{\substack{p,q \leq x^{1/3}, p,q \notin \mathscr{P} \\ p \neq q}} \sum_{\substack{n \leq x \\ pq \mid a+bn}} 1 \\
&= xE + O(x^{1/3}) + x \sum_{\substack{p,q \leq x^{1/3}, p,q \notin \mathscr{P} \\ p \mid a+bn}} \frac{1}{pq} + O(x^{2/3}) = xE + xE^2 + O(x),
\end{aligned} \tag{12}$$

where the implied constant is absolute.

The second sum on the right of (11) is

$$-2E \sum_{n \leq x} g(a+bn) = -2E \sum_{n \leq x} \sum_{\substack{p \leq x^{1/3}, p \notin \mathscr{P} \\ p \mid a+bn}} 1 = -2xE^2 + O(Ex^{1/3}) \tag{13}$$

(as in the calculation in (12)) where again the implied constant is absolute.

Finally, (11), (12), and (13) yield that

$$\sum_{n \leqslant x} (g(a+bn)-E)^2 = xE + xE^2 - 2xE^2 + xE^2 + O(x) = xE + O(x),$$

which gives (10).

LEMMA 2. *Let $\mathscr{P}$ be a set of primes with*

$$\sum_{p \in \mathscr{P}} \frac{1}{p} \leqslant 1. \tag{14}$$

Let a, b be integers with $0 \leqslant a < b$ and such that every prime factor of b is in $\mathscr{P}$. Let

$$f(n) = \sum_{p \notin \mathscr{P}, p \mid n} 1.$$

Then for

$$x > b^2, \tag{15}$$

and $x > x_2$ (where x_2 is an absolute constant, independent of $\mathscr{P}$, a, b) and for all $t > 0$, the number of integers n with $a + bn \leqslant x$ and

$$f(a+bn) < \log\log x - t\sqrt{\log\log x} \tag{16}$$

is less than $5x/t^2 b$.

Proof. The lemma is clearly true if $0 < t \leqslant 1$, so assume now that $t > 1$. Let $\mathscr{T}$ denote the set of integers n with $a + bn \leqslant x$ and satisfying (16). Define

$$g(n) = \sum_{\substack{p \leqslant ((x-a)/b)^{1/3}, p \notin \mathscr{P} \\ p \mid n}} 1,$$

so that $g(n) \leqslant f(n)$ for all n. Thus

$$g(a+bn) \leqslant f(a+bn) < \log\log x - t\sqrt{\log\log x} \qquad (\text{for } n \in \mathscr{T}). \tag{17}$$

In view of (14) and (15) we have

$$E = \sum_{p \leqslant ((x-a)/b)^{1/3}, p \notin \mathscr{P}} \frac{1}{p} = \log\log x + O(1), \tag{18}$$

where the error term is uniformly bounded. Thus from Lemma 1 we have

$$\sum_{n \leqslant (x-a)/b} (g(a+bn)-E)^2 = E\frac{x-a}{b} + O\left(\frac{x-a}{b}\right)$$
$$= \frac{x}{b}(\log\log x + O(1)). \tag{19}$$

On the other hand, by (17) and (18) we have for x large that

$$\sum_{n \leqslant (x-a)/b} (g(a+bn)-E)^2 \geqslant \sum_{n \in \mathscr{T}} (g(a+bn)-E)^2 > \sum_{n \in \mathscr{T}} \left(\frac{t}{2}\sqrt{\log\log x}\right)^2$$
$$= |\mathscr{T}|\frac{t^2}{4}\log\log x.$$

Combining this estimate with (19) gives the lemma.

LEMMA 3. *For $x > x_3$ and all $t > 0$, the number of integers $n \leqslant x$ with $v(n) \geqslant \log\log x + t\sqrt{\log\log x}$ is less than $5x/t^2$.*

This result is well known [1, 2] and, in fact, is a consequence of Lemma 1.

4. THE PROOF OF THEOREM 2

The idea of the proof is to show there are many disjoint intervals $[u, v] \subset [1, x]$ such that the function $n + v(n)$ maps most of $[u, v]$ into an interval $[u', v']$, where $v' - u'$ is a bit smaller than $v - u$. In fact u, v, u', v' will be found so that more integers in $[u, v]$ are mapped into $[u', v']$ than there are integers in $[u', v']$. Thus in $[u, v]$ there are at least two numbers n, m with $n + v(n) = m + v(m)$. The interval $[u, v]$ is found so that just above u, the function $v(n)$ is for most n, unusually large (so that u' can be taken large), while just below v, the function $v(n)$ behaves normally.

Let p_i denote the ith prime. Let x be a large integer, put

$$y = [42\sqrt{\log\log x}]$$

and let

$$\mathscr{P} = \{p_i : y^2 < i \leqslant 2y^2\},$$

$$b = \prod_{p \in \mathscr{P}} p.$$

A simple computation shows that for large enough x,

$$\sum_{p \in \mathscr{P}} \frac{1}{p} < 1, \tag{20}$$

and

$$\begin{aligned} b < \prod_{i \leqslant 2y^2} p_i &= \exp((1+o(1))\, p_{2y^2}) \\ &= \exp((1+o(1))\, 2y^2 \log(2y^2)) \\ &< \exp(4000 \log\log \times \log\log\log x) < \sqrt{x}. \end{aligned} \tag{21}$$

Let $h = h_0$ denote the least positive solution of the linear congruence system

$$h + j \equiv 0 \left(\bmod \prod_{i = y^2 + (j-1)y + 1}^{y^2 + jy} p_i \right), \qquad j = 1, 2, \ldots, y. \tag{22}$$

Thus $0 < h_0 < b$ and h satisfies (22) if and only if

$$h \equiv h_0 (\bmod b). \tag{23}$$

Let $\mathscr{H}$ denote the set of integers h with $0 < h \leqslant x - b$ and such that (23) holds. Put

$$f(n) = \sum_{p \notin \mathscr{P}, p \mid n} 1.$$

For each $h \in \mathscr{H}$, let $\mathscr{J}_h$ denote the set of positive integers j such that

$$h + j + v(h+j) < h + \log\log x + \frac{y}{2}. \tag{24}$$

(Note that if $j \in \mathscr{J}_h$, then $j < \log\log x + y/2$.) Finally, let $\mathscr{H}_1$ denote the set of $h \in \mathscr{H}$ with

$$|\mathscr{J}_h| < \frac{y}{16}. \tag{25}$$

We now show that at least half of the elements of $\mathscr{H}$ are in $\mathscr{H}_1$. By the construction of the sequence $\mathscr{H}$, for $1 \leqslant j \leqslant y$ we have

$$v(h+j) = \sum_{p \mid h+j} 1 = \sum_{\substack{p \mid h+j \\ p \in \mathscr{P}}} 1 + \sum_{\substack{p \mid h+j \\ p \notin \mathscr{P}}} 1 = y + f(h+j)$$

so that (24) implies

$$f(h+j)=\nu(h+j)-y<\log\log x-j-y/2<\log\log x-y/2 \quad (\text{for } 1\leqslant j\leqslant y, j\in \mathscr{J}_h). \tag{26}$$

On the other hand, for $j>y$ we obtain from (24) that

$$f(h+j)\leqslant \nu(h+j)<\log\log x-j+y/2 \quad (\text{for } y<j\leqslant \log\log x+y/2, j\in \mathscr{J}_h). \tag{27}$$

Obviously, we have

$$\sum_{h\in\mathscr{H}}|\mathscr{J}_h|\geqslant \sum_{h\in\mathscr{H}-\mathscr{H}_1}\frac{y}{16}=\frac{1}{16}y\,|\mathscr{H}-\mathscr{H}_1|. \tag{28}$$

We now obtain an upper bound for the sum on the left of (28). We have

$$\sum_{h\in\mathscr{H}}|\mathscr{J}_h|=\sum_{h\in\mathscr{H}}\sum_{j\in\mathscr{J}_h}1=\sum_{j\geqslant 1}\sum_{\substack{h\in\mathscr{H}\\ j\in\mathscr{J}_h}}1$$

$$=\sum_{1\leqslant j\leqslant y}\sum_{\substack{h\in\mathscr{H}\\ j\in\mathscr{J}_h}}1+\sum_{j>y}\sum_{\substack{h\in\mathscr{H}\\ j\in\mathscr{J}_h}}1.$$

By (26), the first inner sum is at most the number of terms of the arithmetic progression h_0+j+bn in $(0, x]$ with

$$f(h_0+j+bn)<\log\log x-y/2.$$

By (27), the second inner sum is at most the number of h_0+j+bn in $(0, x]$ with

$$f(h_0+j+bn)<\log\log x-j+y/2.$$

By (20) and (21), for large x Lemma 2 can be applied to estimate each of the inner sums, so that

$$\begin{aligned}\sum_{h\in\mathscr{H}}|\mathscr{J}_h| &\leqslant \sum_{1\leqslant j\leqslant y}\frac{5x}{(y/2\sqrt{\log\log x})^2\,b}+\sum_{j>y}\frac{5x}{((j-y/2)/\sqrt{\log\log x})^2\,b}\\ &\leqslant \frac{20x\log\log x}{yb}+\frac{5x\log\log x}{b}\sum_{j>y}\frac{1}{(j-y/2)^2}\\ &<\frac{20x\log\log x}{yb}+\frac{5x\log\log x}{b}\cdot\frac{1}{y/2-1}\\ &<\frac{31x\log\log x}{yb}.\end{aligned} \tag{29}$$

Thus (28) and (29) imply that

$$|\mathscr{H}-\mathscr{H}_1|<\frac{496x\log\log x}{y^2b}<\frac{x}{3b}.$$

With this and (21) we have for large x that

$$|\mathscr{H}_1|=|\mathscr{H}|-|\mathscr{H}-\mathscr{H}_1|>\frac{x}{b}-2-\frac{x}{3b}>\frac{x}{2b}, \tag{30}$$

thus showing that at least half of the elements of $\mathscr{H}$ are in $\mathscr{H}_1$. The members of $\mathscr{H}_1$ are our candidates for the numbers "u" described in the beginning of this section, while the numbers $h+\log\log x+y/2$ for $h\in\mathscr{H}_1$ are the candidates for the numbers "u'." However we have to do some more thinning out to allow for "v" and "v'."

Let $\mathscr{H}_2$ denote the set of $h\in\mathscr{H}_1$ for which there is an integer l_h with

$$h+y<l_h\leqslant h+b \tag{31}$$

and (letting $z=\log\log x+y/4$)

$$\sum_{\substack{h<n\leqslant l_h\\ n+v(n)\geqslant l_h+z}}1\leqslant\frac{y}{16}. \tag{32}$$

In order to give a lower bound for $|\mathscr{H}_2|$, we need an upper bound for $|\mathscr{H}_1-\mathscr{H}_2|$, i.e., for the number of $h\in\mathscr{H}_1$ such that

$$\sum_{\substack{h<n\leqslant l\\ n+v(n)\geqslant l+z}}1>\frac{y}{16}$$

for all l satisfying

$$h+y<l\leqslant h+b.$$

We have

$$\begin{aligned}\sum_{h\in\mathscr{H}_1-\mathscr{H}_2}\sum_{l=h+y+1}^{h+b}\sum_{\substack{h<n\leqslant l\\ n+v(n)\geqslant l+z}}1&>\sum_{h\in\mathscr{H}_1-\mathscr{H}_2}\sum_{l=h+y+1}^{h+b}\frac{y}{16}=\sum_{h\in\mathscr{H}_1-\mathscr{H}_2}\frac{y}{16}(b-y)\\&>|\mathscr{H}_1-\mathscr{H}_2|\frac{yb}{17}\end{aligned} \tag{33}$$

for large enough x.

On the other hand, by Lemma 3 we have

$$\sum_{h \in \mathscr{H}_1 - \mathscr{H}_2} \sum_{l=h+y+1}^{h+b} \sum_{\substack{h < n \leqslant l \\ n + v(n) \geqslant l + z}} 1$$

$$\leqslant \sum_{h \in \mathscr{H}_1 - \mathscr{H}_2} \sum_{h \leqslant n \leqslant h+b} \sum_{n \leqslant l \leqslant n + v(n) - z} 1$$

$$\leqslant \sum_{n \leqslant x} \sum_{z \leqslant k \leqslant v(n)} 1 = \sum_{k \geqslant z} \sum_{\substack{n \leqslant x \\ v(n) \geqslant k}} 1 \tag{34}$$

$$< \sum_{k \geqslant z} \frac{5x}{((k - \log\log x)/\sqrt{\log\log x})^2}$$

$$= 5x \log\log x \sum_{k \geqslant z} \frac{1}{(k - \log\log x)^2}$$

$$< 5x \log\log x \cdot \frac{1}{y/4 - 1} < \frac{25x \log\log x}{y}$$

for large x.

Thus (33) and (34) imply that

$$|\mathscr{H}_1 - \mathscr{H}_2| < \frac{17}{by} \cdot \frac{25x \log\log x}{y} = \frac{425x \log\log x}{by^2} < \frac{x}{4b},$$

so that by (30),

$$|\mathscr{H}_2| = |\mathscr{H}_1| - |\mathscr{H}_1 - \mathscr{H}_2| > \frac{x}{2b} - \frac{x}{4b} = \frac{x}{4b}. \tag{35}$$

Now for $h \in \mathscr{H}_2$ consider the interval $[h+1, l_h]$, where l_h satisfies (31) and (32). By (24) and (25), but for at most $y/16$ exceptions, every $n \in [h+1, l_h]$ has

$$n + v(n) \geqslant h + \log\log x + \frac{y}{2}. \tag{36}$$

By (32), but for at most $y/16$ exceptions, every $n \in [h+1, l_h]$ has

$$n + v(n) < l_h + \log\log x + \frac{y}{4}. \tag{37}$$

Thus, but for at most $y/8$ exceptions, every $n \in [h+1, l_h]$ has both (36) and (37) holding. So we have at least $l_h - h - y/8$ numbers n mapped by $n + v(n)$ to the interval

$$\left[h + \log\log x + \frac{y}{2},\, l_h + \log\log x + \frac{y}{4}\right),$$

which has at most $l_h - h - [y/4]$ integers. There are therefore at least $[y/8]$ pairs $n, m \in [h+1, l_h]$ with $n \neq m$ and $n + \nu(n) = m + \nu(m)$.

As h runs over $\mathscr{H}_2$, the intervals $[h+1, l_h]$ are disjoint and contained in $[1, x]$. Thus, below x, there are at least (using (21), (35) and assuming x is large)

$$[y/8] |\mathscr{H}_2| > 4 |\mathscr{H}_2| > \frac{x}{b} > x \cdot \exp(-4000 \log\log x \log\log\log x)$$

pairs $n, m \leqslant x$ with $n \neq m$ and $n + \nu(n) = m + \nu(m)$.

Remarks. Probably Theorem 2 is far from the truth. We conjecture that there are positive constants c_5, c_6 such that

$$|\{n \leqslant x : \exists m \neq n, n + \nu(n) = m + \nu(m)\}| \sim c_5 x,$$
$$|\{(n, m) : n \leqslant x, m \leqslant x, n + \nu(n) = m + \nu(m)\}| \sim c_6 x.$$

Almost the same proof as for Theorem 2 can show the analogous result with $\Omega(n)$ replacing $\nu(n)$. With a little more difficulty, the same can be proved with $\tau(n)$ replacing $\nu(n)$.

ACKNOWLEDGMENT

Thanks are due to Paul T. Bateman, who suggested several improvements and corrections.

REFERENCES

1. P. D. T. A. ELLIOTT, "Probabilistic Number Theory," Vol. 1, Springer-Verlag, New York, 1980.
2. P. TURÁN, On a theorem of Hardy and Ramanujan, *J. London Math. Soc.* **9** (1934), 274–276.

Printed by the St. Catherine Press Ltd., Tempelhof 41, Bruges, Belgium