Matching the natural numbers up to n with distinct multiples in another interval

by Paul Erdös[1] and Carl Pomerance[2]

[1] *Mathematical Institute of the Hungarian Academy of Sciences*
[2] *University of Illinois at Urbana-Champaign and University of Georgia, U.S.A.*

Communicated at the meeting of June 16, 1979

§ 1. INTRODUCTION

Let $f(n)$ denote the least integer so that in the interval $(n, f(n)]$ there are distinct integers $a_1, \ldots, a_n$ with $i|a_i$ for $i=1, \ldots, n$. Thus, for example, $f(10)=24$ as can be easily seen by letting

$$a_1=11,\ a_2=22,\ a_3=21,\ a_4=16,\ a_5=15,\ a_6=12,\ a_7=14,$$
$$a_8=24,\ a_9=18,\ a_{10}=20.$$

(The fact that $f(10)\geqslant 24$ follows from the observation that there are only 9 composites in the interval $[11, 24]$.)

More generally, if m is any positive integer, let $f(n, m)$ denote the least integer so that in $(m, m+f(n, m)]$ there are distinct integers $a_1, \ldots, a_n$ with $i|a_i$ for $i=1, \ldots, n$. Thus $f(n)=n+f(n, n)$. Let $L(n)$ denote the least common multiple of $1, \ldots, n$. Then it is clear that $f(n, m)$ depends only on the residue class of m modulo $L(n)$.

We shall be concerned with the following problems:

1. Find estimates or an asymptotic formula for $f(n)$.
2. For each n, estimate the maximal value of $f(n, m)$.
3. For each n, estimate the average value of $f(n, m)$.

On Problem 1 we show that, perhaps unexpectedly, as $n \to \infty$,

$f(n)/n \to \infty$ (Theorem 1). We also show $f(n) \ll n(\log n)^{1/2}$ (Theorem 3) and that this result is nearly best possible (Theorem 2).

On Problem 2 we show that $\max_m f(n, m) \ll n^{3/2}$ (Theorem 4). We cannot show $\max_m f(n, m) > f(n, n)$ so Theorem 2 gives our best lower bound for $\max_m f(n, m)$.

On Problem 3 we show that there is a positive constant α such that

$$n(\log n)^{\alpha} \leqslant \frac{1}{L(n)} \sum_{m=1}^{L(n)} f(n, m) \leqslant n^{1+o(1)}$$

for large n (Theorems 5 and 6).

The methods we use for the lower bound theorems on $f(n)$ involve results on the function $\psi(x, y)$, the number of integers not exceeding x composed only of the primes not exceeding y. In particular we shall be concerned with estimates for $\psi(x, y)$ for "very small" y, that is, $y < \log x$.

All of our upper bound results for $f(n)$ and $f(n, m)$ rely on a theorem of König [7] and Hall [5]. We proceed now to introduce the terms needed to state the theorem. If G is a bipartite graph between the disjoint sets I, J (that is, the vertex set of G is $I \cup J$ and the edge set is contained in $I \times J$) and if $U \subset I$, then the *span* of U is the set of points of J connected to some point of U by an edge. One can similarly define the span of a set $V \subset J$. If $a \in I \cup J$, the valence of a is $|\text{span } \{a\}|$. To say that G contains a *matching of I into J* means that the edge set of G (which is a relation from I to J) contains a $1-1$ correspondence of I with a subset of J.

THEOREM (KÖNIG, HALL). Let G be a bipartite graph on the disjoint finite sets I and J. Suppose G does not contain a matching from I into J. Then both

(i) there is a $u \in I$ and a $v \in J$ with valence $u <$ valence v;

(ii) there is a $U \subset I$ with $|U| > |\text{span } U|$.

The König-Hall theorem is sometimes referred to as the "marriage theorem".

Our lower bound result for the average value of $f(n, m)$ relies on the recent work of Tenenbaum [12] for the density of the integers which contain a divisor between $n/2$ and n.

We take this opportunity to thank Harold Diamond for several interesting discussions concerning the contents of this paper.

§ 2. LOWER BOUNDS FOR $f(n)$

For each y, let $\psi(y)$ denote the set of positive integers not divisible by any prime exceeding y. Let $\psi(x, y)$ denote the number of members of $\psi(y)$ which do not exceed x.

LEMMA 1. Let n be a natural number and let k, y be positive quantities such that $1<k<y$ and

(1) $\quad \psi(n, y)-\psi(nk/y, y)>\psi(nk, y)-\psi(n, y).$

Then $f(n)>nk$.

PROOF. Assume $f(n)\leqslant nk$. Let $I=(nk/y, n]\cap\psi(y)$, $J=(n, nk]\cap\psi(y)$. Let $i=|I|$, $j=|J|$, $I=\{a_1, \ldots, a_i\}$. Then (1) implies $i>j$. The assumption $f(n)\leqslant nk$ implies there are distinct integers $b_1, \ldots, b_i\in(n, nk]$ with $a_l|b_l$ for $1\leqslant l\leqslant i$. Note that $b_l/a_l\leqslant nk/a_l<y$. Thus since $a_l\in\psi(y)$, we have $b_l\in\psi(y)$. That is, $b_1, \ldots, b_i$ are all in J. Hence $i\leqslant j$, a contradiction. □

THEOREM 1. $\lim_{n\to\infty} f(n)/n=\infty$.

PROOF. Let $k>1$ be arbitrary, but fixed. Let $y=k^3$. It is known (Specht [11]) that

$$\psi(x, y)=c_1(\log x)^{\pi(y)}+c_2(\log x)^{\pi(y)-1}+o((\log x)^{\pi(y)-1})$$

where

$$c_1=(\pi(y)!\cdot\prod_{p\leqslant y}\log p)^{-1},\ c_2=(c_1\pi(y)/2)\sum_{p\leqslant y}\log p,$$

and p runs through primes. Thus

$$\begin{aligned}\psi(n, y)-\psi(n/k^2, y)&=c_1(\log n)^{\pi(y)}+c_2(\log n)^{\pi(y)-1}-c_1(\log(n/k^2))^{\pi(y)}\\&\quad -c_2(\log(n/k^2))^{\pi(y)-1}+o(\log n)^{\pi(y)-1}\\&=2c_1\pi(y)\log k(\log n)^{\pi(y)-1}+o(\log n)^{\pi(y)-1},\end{aligned}$$

and similarly

$$\psi(nk, y)-\psi(n, y)=c_1\pi(y)\log k(\log n)^{\pi(y)-1}+o(\log n)^{\pi(y)-1}.$$

We thus have for all large n that (1) holds. Hence Lemma 1 implies $f(n)>nk$ for all large n. Since k is arbitrary we have our theorem. □

The above argument depends on a sharp error term for $\psi(x, y)$ available for bounded y. The existence of such a sharp formula for $\psi(x, y)$ (in the case $y=3$) was first discovered by Ramanujan (cf. Pillai [8] and Hardy [6]). To improve Theorem 1 to the actual exhibition of an explicit function which tends to infinity with n and which is a lower bound for $f(n)/n$, the above method would need a sharp formula for $\psi(x, y)$ for $y\to\infty$ slowly but explicitly. Note that the asymptotic formulas given by Ennola [2] do not have a sharp enough error term for this purpose. Possibly sharp enough formulas for $\psi(x, y)$ could be obtained, but we do not make this effort here. Instead, we find a different method to attack the problem of lower bounds for $f(n)$. In our next theorem, we use an asymptotic formula for $\log\psi(x, y)$ given by de Bruijn [1] to obtain a substantial improvement on Theorem 1. The reason we can make do with a non-sharp approximation of $\psi(x, y)$ in the proof of Theorem 2, while in Theorem 1 we need a sharp error term is the observation that knowledge about $f(m)$ gives one knowledge of $f(n)$ for all $n>m$. For example, the fact that $f(10)=24$, as seen in our opening example, can be shown to imply that $f(100)\geqslant 240$. We thus are able to get a good lower bound for all $f(n)$ by first finding a good lower bound for some $f(n)$. The method exploits the geometry of the graph of $\log \psi(x, y)$ for fixed y (cf. Pomerance [9]).

THEOREM 2. For $n\geqslant 3$, $f(n)\geqslant (2/\sqrt{e}+o(1))n\sqrt{\log n/\operatorname{loglog} n}$.

LEMMA 2. For every $\varepsilon>0$ there is an $x_0(\varepsilon)$ such that for all $x>x_0(\varepsilon)$, there is an integer $m\in[x, x^{1+\varepsilon}]$ with

$$f(m)>(1-\varepsilon)(2/\sqrt{e})m\sqrt{\log m/\operatorname{loglog} m}. \tag{2}$$

Before we prove the lemma, we show how Theorem 2 follows from it. Let $\varepsilon>0$ be arbitrary and let n be a positive integer. Let $x=(\varepsilon n)^{1/(1+\varepsilon)}$. Thus if n is sufficiently large, the lemma implies there is an integer $m\in[x, x^{1+\varepsilon}]$ for which (2) holds. Let $k=[n/m]$. In the interval $(n, f(n)]$ there are distinct integers $b_1, \ldots, b_m$ such that $ik|b_i$ for $i=1, \ldots, m$. Let $a_i=b_i/k$. Then $a_1, \ldots, a_m$ are distinct integers larger than m for which $i|a_i$ for $i=1, \ldots, m$. Thus $\max a_i\geqslant f(m)$, so that

$$\begin{aligned} f(n)\geqslant \max_{1\leqslant i\leqslant m} b_i\geqslant kf(m) &> (1-\varepsilon)(2/\sqrt{e})\, km\sqrt{\log m/\operatorname{loglog} m} \\ &> (1-\varepsilon)^2(2/\sqrt{e})(n-m)\sqrt{\log n/\operatorname{loglog} n} \\ &> (1-\varepsilon)^3(2/\sqrt{e})\, n\sqrt{\log n/\operatorname{loglog} n} \end{aligned}$$

for all sufficiently large n. We thus have Theorem 2.

PROOF OF LEMMA 2. For each $y>e^{10}$, let

$$g_y(w)=g(w)=\log \psi(e^w, y) \text{ for } w\in[(0.1)y\log y, y\log y].$$

By de Bruijn [1],

$$\begin{aligned} g(w) &= \left\{\log\left(1+\frac{y}{w}\right)\cdot\frac{w}{\log y}+\log\left(1+\frac{w}{y}\right)\cdot\frac{y}{\log y}\right\}\cdot\left\{1+o\left(\frac{1}{\log y}\right)\right\} \tag{3} \\ &= \left\{\left(\frac{y}{w}+o\left(\frac{y^2}{w^2}\right)\right)\cdot\frac{w}{\log y}+\left(\log\left(\frac{w}{y}\right)+o\left(\frac{y}{w}\right)\right)\cdot\frac{y}{\log y}\right\}\cdot\left\{1+o\left(\frac{1}{\log y}\right)\right\} \\ &= \frac{y}{\log y}\cdot\left\{\log\left(\frac{w}{y}\right)+1+o\left(\frac{y}{w}\right)\right\}\cdot\left\{1+o\left(\frac{1}{\log y}\right)\right\} \\ &= \frac{y}{\log y}\left(\log\left(\frac{w}{y}\right)+1+o\left(\frac{\operatorname{loglog} y}{\log y}\right)\right) \end{aligned}$$

uniformly for all $y>e^{10}$, $w\in[(0.1)\,y\log y, y\log y]$.

Now let $h_y(w)=h(w)$ be the function whose graph is the upper boundary of the convex hull of the graph of $g(w)$. Then

$$h(w)=\frac{y}{\log y}\left(\log\left(\frac{w}{y}\right)+1+o\left(\frac{\operatorname{loglog} y}{\log y}\right)\right) \tag{4}$$

uniformly for $y>e^{10}$, $w\in[(0.1)\,y\log y,\, y\log y]$. Indeed,

$$g(w)\leqslant h(w)\leqslant\frac{y}{\log y}\left(\log\left(\frac{w}{y}\right)+1+c\,\frac{\operatorname{loglog} y}{\log y}\right)$$

where c is the absolute constant implicit in (3). Since $g(w)$ is a step function, we have $h(w)$ piecewise linear. Thus $h'(w)$ exists everywhere but for a finite set of points which we shall call *vertex points*. A vertex point w satisfies $g(w)=h(w)$. Also e^w is an integer if the vertex point w is not an endpoint of the interval.

We now show that if w is not a vertex point, then

$$h'(w)=\frac{y}{w\log y}\left(1+o\left(\sqrt{\frac{\operatorname{loglog} y}{\log y}}\right)\right) \tag{5}$$

uniformly. Indeed for each $\delta>0$ and w such that w is not a vertex point and

$$(0.1)\,y\log y\leqslant(1-\delta)w,\ (1+\delta)w\leqslant y\log y,$$

we have (since h is concave down)

$$(h((1+\delta)w)-h(w))/\delta w\leqslant h'(w)\leqslant(h(w)-h((1-\delta)w))/\delta w.$$

Hence by (4),

$$\frac{y}{\delta w\log y}\left(\log((1+\delta)w)-\log\, w+o\left(\frac{\operatorname{loglog} y}{\log y}\right)\right)\leqslant h'(w)$$
$$\leqslant\frac{y}{\delta w\log y}\left(\log w-\log\,((1-\delta)w)+o\left(\frac{\operatorname{loglog} y}{\log y}\right)\right),$$

$$\frac{y}{w\log y}\left(1+o(\delta)+o\left(\frac{\operatorname{loglog} y}{\delta\log y}\right)\right)\leqslant h'(w)\leqslant\frac{y}{w\log y}\left(1+o(\delta)+o\left(\frac{\operatorname{loglog} y}{\delta\log y}\right)\right).$$

Thus choosing $\delta=\sqrt{\operatorname{loglog} y/\log y}$, we have (5).

Let $0<\varepsilon<\frac{1}{2}$ be arbitrary. Let b be a constant to be chosen later with $0.1<b<\frac{1}{2}$. Then for all sufficiently large y, depending on the choice of ε, we have by (5) that

$$h'(b\,y\log y)>h'((1+\varepsilon)b\,y\log y)$$

if neither argument is a vertex point. Thus there is a vertex point $W_y=W$ satisfying

$$b\,y\log y\leqslant W\leqslant(1+\varepsilon)b\,y\log y.$$

Let $m_y = m = e^W$, an integer. Let a be a positive constant to be chosen later, and let

$$\begin{aligned} A &= (\tfrac{1}{2})\log y + \log a, & B &= (\tfrac{1}{2})\log y - \log a, \\ \Delta_1 &= g(W) - g(W-A), & \Delta_2 &= g(W+B) - g(W), \\ \delta_1 &= \psi(m, y) - \psi(m/e^A, y), & \delta_2 &= \psi(m\cdot e^B, y) - \psi(m, y). \end{aligned}$$

Note that if y is sufficiently large, then (0.1) $y \log y \leqslant W - A$ and $W + B \leqslant$ $\leqslant y \log y$.

We shall show that for sufficiently large y and for suitable fixed choices of b, a, we have $\delta_1 > \delta_2$.

Note that

$$\delta_1 = \psi(m, y)(1 - e^{-\Delta_1}), \quad \delta_2 = \psi(m, y)(e^{\Delta_2} - 1). \tag{6}$$

If $h'_+(W)$ denotes the right hand derivative at W, we have by (5)

$$\begin{aligned} \Delta_2 &\leqslant B \cdot h'_+(W) = \frac{B\,y}{W \log y}\left(1 + o\left(\sqrt{\frac{\operatorname{loglog} y}{\log y}}\right)\right) \\ &\leqslant \frac{B}{b \log^2 y}\left(1 + o\left(\sqrt{\frac{\operatorname{loglog} y}{\log y}}\right)\right) = \frac{1}{2b \log y}\left(1 + o\left(\sqrt{\frac{\operatorname{loglog} y}{\log y}}\right)\right). \end{aligned} \tag{7}$$

Also for large y, $m \cdot e^B \geqslant 2m$, so that there is a power of 2 in the interval $(m, m\cdot e^B]$. Thus $\delta_2 > 0$, so that (6) implies $\Delta_2 > 0$.

Assume $\Delta_1 \geqslant 1$. Since (6), (7) imply $\delta_2 = \psi(m, y)\cdot o(1/\log y)$, we would thus have by (6) that $\delta_1 > \delta_2$. Thus we may assume $\Delta_1 < 1$. With this assumption we have

$$\delta_1 > \psi(m, y)(\Delta_1 - \tfrac{1}{2}\Delta_1^2), \;\; \delta_2 \leqslant \psi(m, y)(\Delta_2 + \tfrac{1}{2}\Delta_2^2 + o(\Delta_2^3)).$$

We thus have

$$\begin{aligned} \delta_1 - \delta_2 &\geqslant \psi(m, y)(\Delta_1 - \tfrac{1}{2}\Delta_1^2 - \Delta_2 - \tfrac{1}{2}\Delta_2^2 + o(\Delta_2^3)) \\ &= \psi(m, y)((\Delta_1 - \Delta_2) - \Delta_2^2 - \Delta_2(\Delta_1 - \Delta_2) - \tfrac{1}{2}(\Delta_1 - \Delta_2)^2 + 0(\Delta_2^3)). \end{aligned} \tag{8}$$

Since W is a vertex point, we have $\Delta_1/A > \Delta_2/B$, so that

$$\Delta_1 - \Delta_2 > \left(\frac{A}{B} - 1\right)\Delta_2 = \frac{2 \log a}{B}\Delta_2. \tag{9}$$

The assumption $\Delta_1 < 1$ implies $\Delta_1 - \Delta_2 < 1 - \Delta_2$, so that (7), (8), (9) give

$$\delta_1 - \delta_2 \geqslant \psi(m, y)\Delta_2 \cdot \left\{\frac{2 \log a}{B} - \Delta_2\left(1 + \frac{2 \log a}{B} + \frac{2 \log^2 a}{B^2} + o(\Delta_2)\right)\right\}$$

$$= \psi(m, y)\Delta_2 \cdot$$

$$\left\{\frac{4 \log a}{\log y} + o\left(\frac{1}{\log^2 y}\right) - \frac{1}{2b \log y}\left(1 + o\left(\sqrt{\frac{\operatorname{loglog} y}{\log y}}\right)\right)\left(1 + o\left(\frac{1}{\log y}\right)\right)\right\}$$

$$= \psi(m, y)\frac{\Delta_2}{\log y}\left(4 \log a - \frac{1}{2b} + o\left(\sqrt{\frac{\operatorname{loglog} y}{\log y}}\right)\right).$$

We now choose $a=e^{1/8b}(1+\varepsilon/2)$, so that for all sufficiently large y we have $\delta_1>\delta_2$.

Since $e^{A+B}=y$, the inequality $\delta_1>\delta_2$ implies by Lemma 1 that

$$f(m)>m\cdot e^B=m\sqrt{y}/a.$$

Now $m=e^W$, so that

$$\text{(10)}\qquad b\,y\log y\leqslant\log m\leqslant(1+\varepsilon)\,b\,y\log y.$$

Thus for large y, $\log y<\operatorname{loglog} m$, so that

$$y>\frac{1}{b(1+\varepsilon)}\cdot\frac{\log m}{\operatorname{loglog} m}.$$

Hence

$$\begin{aligned} f(m) &> \frac{m}{a\sqrt{b}\sqrt{1+\varepsilon}}\cdot\sqrt{\frac{\log m}{\operatorname{loglog} m}} \\ &= \frac{m}{e^{1/8b}(1+\varepsilon/2)\sqrt{b}\sqrt{1+\varepsilon}}\sqrt{\frac{\log m}{\operatorname{loglog} m}} \\ &> \frac{(1-\varepsilon)\,m}{e^{1/8b}\sqrt{b}}\sqrt{\frac{\log m}{\operatorname{loglog} m}}. \end{aligned}$$

Thus choosing $b=\frac{1}{4}$, we have (2).

For each x, let y be such that $\log x=\frac{1}{4}y\log y$. We have seen that for all sufficiently large x there is an integer m for which both (2) and (10) hold. But (10) implies $x\leqslant m\leqslant x^{1+\varepsilon}$. □

§ 3. UPPER BOUNDS FOR $f(n)$

THEOREM 3. For $n\geqslant 2$, $f(n)\leqslant(2+o(1))n\sqrt{\log n}$.

PROOF. Let $\varepsilon>0$ be arbitrary, but fixed. For $i\in(n/\sqrt{\log n}, n]$, let $a_i=i([\sqrt{\log n}]+1)$. Then the a_i are distinct and $a_i\in(n, n(\sqrt{\log n}+1)]$. Let

$$I=[1, n/\sqrt{\log n}]\cap\mathbf{Z},\quad J=(n(\sqrt{\log n}+1),\ (2+\varepsilon)n\sqrt{\log n}]\cap\mathbf{Z}.$$

Let G be the bipartite graph from I to J where $(i,j)\in I\times J$ is an edge if and only if j/i is prime.

If $i\in I$, then the valence of i is

$$\pi((2+\varepsilon)n\sqrt{\log n}\,/i)-\pi(n(\sqrt{\log n}\,+1)/i)$$
$$>\left(1+\frac{\varepsilon}{2}\right)\frac{n\sqrt{\log n}}{i\log(n\sqrt{\log n}/i)}\geqslant\left(1+\frac{\varepsilon}{2}\right)\frac{\log n}{\operatorname{loglog} n}$$

for all sufficiently large n by the prime number theorem. If $j\in J$, then the valence of j is at most $\omega(j)$, the number of distinct prime factors of j.

But again from the prime number theorem, we have for all large n

$$\omega(j) < (1+\varepsilon/2) \log n/\log\log n.$$

Thus by the König-Hall theorem, it follows that G contains a matching of I into J. Hence for all large n, $f(n) \leqslant (2+\varepsilon)n\sqrt{\log n}$. □

We can improve the theorem slightly. Let r be the solution of the equation $e^{-r}=r$ and let $c=\sqrt{r}/(1-r)=1.7398\ \ldots$. Then

$$(11) \qquad f(n) < (c+o(1))n\sqrt{\log n}.$$

We now sketch a proof of (11). Let $\gamma \in (0, 1)$ and let k be a natural number. Let

$$I_j = (\gamma^j n/\sqrt{\log n},\ \gamma^{j-1}\, n/\sqrt{\log n}] \cap \mathbf{Z} \text{ for } j=-k+1,\ -k+2,\ \ldots,\ k,$$
$$I_{-k} = (\gamma^{-k}\, n/\sqrt{\log n},\ n] \cap \mathbf{Z},\ \ I_{k+1} = [1,\ \gamma^k n/\sqrt{\log n}] \cap \mathbf{Z}.$$

Thus the I_j are disjoint and $\bigcup_{j=-k}^{k+1} I_j = [1, n] \cap \mathbf{Z}$. Now let b be a positive number and let

$$J_1 = (n,\ \gamma^k n(\sqrt{\log n}+1)] \cap \mathbf{Z}, \quad J_2 = (\gamma^k n(\sqrt{\log n}+1),\ (\gamma^k+b)n\sqrt{\log n}].$$

If $i \in I_{-k}$, let $a_i = i([\gamma^k\sqrt{\log n}]+1) \in J_1$. Let

$$I = \bigcup_{j=-k+1}^{k+1} I_j = ([1, n] \cap \mathbf{Z}) - I_{-k}.$$

Let G be the bipartite graph on I, J_2 with $(i, j) \in I \times J_2$ an edge if and only if j/i is prime. We shall show that for a suitable choice of γ, k, b, G contains a matching of I into J_2. It will thus follow that $f(n) \leqslant (\gamma^k+b)n\sqrt{\log n}$.

Say G does not contain a matching of I into J_2. Then by the König-Hall theorem there is a set $U \subset I$ with $x=|U|>|\text{span } U|=y$. Let $V=\text{span } U$ and let $x_j=|U \cap I_j|$ for $j=-k+1, \ldots, k$.

If $u \in I_j$, $-k+1 \leqslant j \leqslant k+1$, the valence of u is at least

$$(b\gamma^{-j+1}+o(1))\log n/\log\log n.$$

If $v \in J_2$, the number of $u \in I_j$ which are connected to v by an edge is at most the lesser of

$$(b+\gamma^k+o(1))(\gamma^{-j}-\gamma^{-j+1})\log n/\log\log n \text{ and } (1+o(1))\log n/\log\log n.$$

Now the number of edges incident to $U \cap I_j$ is at least the number of edges incident to V with an endpoint in I_j. Thus for $-k+1 \leqslant j \leqslant k$,

$$(b\gamma^{-j+1}+o(1))x_j \leqslant (b+\gamma^k+o(1))(\gamma^{-j}-\gamma^{-j+1})y.$$

Hence for $-k+1 \leqslant j \leqslant k$ and using $y < x$,

$$(12) \qquad bx_j \leqslant (b+\gamma^k+o(1))(\gamma^{-1}-1)x.$$

The number of edges incident to V is at least as big as the number of edges incident to U. But the number of edges incident to V is at most

$(1+o(1))\, y \log n/\log\log n$. Thus

$$(1+o(1))x \geq (1+o(1))y \geq \sum_{j=-k+1}^{k} (b\gamma^{-j+1}+o(1))x_j + (b\gamma^{-k}+o(1))(x - \sum_{j=-k+1}^{k} x_j)$$

$$= \sum_{j=-k+1}^{k} (b\gamma^{-j+1} - b\gamma^{-k}+o(1))x_j \;+\; (b\gamma^{-k}+o(1))x$$

$$\geq \sum_{j=-k+1}^{k} (b+\gamma^{k}+o(1))(\gamma^{-1}-1)(\gamma^{-j+1}-\gamma^{-k})x \;+\; (b\gamma^{-k}+o(1))x$$

using the negative of (12). Thus dividing by x and multiplying by γ^{k+1}, we have

$$\gamma^{k+1}(1+o(1)) \geq (b+\gamma^{k})(1-\gamma) \sum_{j=-k+1}^{k} (\gamma^{k-j+1}-1) \;+\; b\gamma$$

$$= -(b+\gamma^{k})(\gamma^{2k+1}-\gamma-2k\gamma+2k)+b\gamma.$$

Let $\beta = b+\gamma^{k}$. Then, by letting $n \to \infty$, we have

$$2\gamma^{k+1} \geq \beta(-\gamma^{2k+1}+(2k+2)\gamma-2k).$$

We thus conclude that if γ, k, b are chosen so that $-\gamma^{2k+1}+(2k+2)\gamma-2k>0$ and

$$\beta > 2\gamma^{k+1}/(-\gamma^{2k+1}+(2k+2)\gamma-2k) \tag{13}$$

then for all sufficiently large n, $f(n) \leq \beta n\sqrt{\log n}$.

Let r be the solution of the equation $e^{-r}=r$ and let $\gamma = 1-r/2k$. Then the right side of (13) is

$$\frac{2(1-r/2k)^{k+1}}{-(1-r/2k)^{2k+1}+(2k+2)(1-r/2k)-2k}$$

$$= \frac{2e^{-r/2}+o(1/k)}{-e^{-r}+2-r+o(1/k)} = \frac{\sqrt{r}}{1-r} + o\left(\frac{1}{k}\right) = c + o\left(\frac{1}{k}\right).$$

Thus letting $k \to \infty$, we have (11).

§ 4. EXTREME VALUES OF $f(n, m)$

THEOREM 4. For all positive integers m, n, we have

$$f(n, m) \leq 4n([\sqrt{n}]+1).$$

PROOF. Let m, n be arbitrary positive integers, let $I_1 = [1, n] \cap \mathbb{Z}$, $J_1 = (m, m+4n[\sqrt{n}]] \cap \mathbb{Z}$. Using the intervals $(m+(k-1)n, m+kn]$ for $k = 1, 2, \ldots, 4[\sqrt{n}]$, we have a partition of J_1 into $4[\sqrt{n}]$ consecutive intervals of length n. If $j \in J_1$, let $\langle j \rangle$ denote the interval to which j belongs. Let G_1 be the bipartite graph from I_1 to J_1 where (i, j) is an edge if and only if $i | j$.

Say the valence of every $j \in J_1$ is at most $[\sqrt{n}]$. Since the valence of every $i \in I_1$ is at least $4[\sqrt{n}]$, it follows from the König-Hall theorem that G_1 contains a matching of I_1 into J_1. Thus we would have our result. Thus assume some $j_1 \in J_1$ has valence larger than $[\sqrt{n}]$. Say span $\{j_1\} = K_1 \subset I_1$ where $|K_1| \geq [\sqrt{n}]+1$. Send each $i \in K_1$ to $j_1 + i = a_i$. Then $i | a_i$ and these a_i are all distinct. Moreover these a_i lie in $\langle j_1 \rangle \cup \langle j_1 + n \rangle$. If $K_1 = I_1$, we are all done. So assume $K_1 \neq I_1$. Let $I_2 = I_1 - K_1$, $J_2 = J_1 - (\langle j_1 - n \rangle \cup \langle j_1 \rangle \cup \langle j_1 + n \rangle)$. Let G_2 be the subgraph of G_1 determined by I_2, J_2.

Say the valence of every $j \in J_2$ is at most $[\sqrt{n}]$. Since the valence of every $i \in I_2$ is at least $4[\sqrt{n}] - 3 \geq [\sqrt{n}]$, it follows from the König-Hall theorem that G_2 contains a matching of I_2 into J_2. Thus we would have our result. So assume some $j_2 \in J_2$ has valence larger than $[\sqrt{n}]$. Say span $\{j_2\} = K_2 \subset I_2$ where $|K_2| \geq [\sqrt{n}]+1$. Send each $i \in K_2$ to $j_2 + i = a_i$. Then $i | a_i$, the a_i are distinct, and the a_i all lie in $\langle j_2 \rangle \cup \langle j_2 + n \rangle$. These two intervals are disjoint from $\langle j_1 \rangle \cup \langle j_1 + n \rangle$. If $K_2 = I_2$, we are done. So assume $K_2 \neq I_2$. Let $I_3 = I_2 - K_2$, $J_3 = J_2 - (\langle j_2 - n \rangle \cup \langle j_2 \rangle \cup \langle j_2 + n \rangle)$. Note that we might have $\langle j_2 + n \rangle = \langle j_1 - n \rangle$ or $\langle j_2 - n \rangle = \langle j_1 + n \rangle$. Let G_3 be the subgraph of G_2 determined by I_3, J_3.

Say we continue this procedure until we reach the bipartite graph G_{t+1} from I_{t+1} to J_{t+1}. We have that

$$(14) \qquad |I_{t+1}| \leq n - t([\sqrt{n}]+1)$$

and that J_{t+1} consists of at least $4[\sqrt{n}] - 3t$ disjoint intervals of length n. From (14) we may assume $t \leq [\sqrt{n}]$, so that J_{t+1} consists of at least $[\sqrt{n}]$ disjoint intervals of length n. Thus every $i \in I_{t+1}$ has valence at least $[\sqrt{n}]$.

We thus conclude that our procedure must terminate at some t and when it does, one of two events must occur. Either $I_t = \emptyset$ or G_t contains a matching from I_t to J_t. In either case, we are done. □

We can lower the constant "4" in Theorem 4 somewhat, but we do not know how to prove $f(n, m) = o(n^{3/2})$. We conjecture that $f(n, m) \leq n^{1+o(1)}$.

§ 5. THE AVERAGE VALUE OF $f(n, m)$

THEOREM 5. Let $\alpha = 1 - \log(e \log 2)/\log 2 = .08607 \ldots$. Then for all sufficiently large n,

$$L(n)^{-1} \sum_{m=1}^{L(n)} f(n, m) > n(\log n)^{\alpha}.$$

PROOF. From Tenenbaum [12] we have that the density d_n of the integers which have a divisor between $[n/2]$ and n is $o((\log n)^{-\alpha})$. In the interval $(m, m+f(n, m)]$ there are at least $n/2$ distinct integers with a divisor between $[n/2]$ and n. Let $S(n, x)$ denote the number of $j \leq x$ which have a divisor d with $[n/2] < d \leqslant n$. Then

$$\sum_{m=1}^{L(n)} (S(n, m+2n(\log n)^\alpha) - S(n, m)) \leqslant 2n(\log n)^\alpha \cdot S(n, L(n))$$
$$= 2n(\log n)^\alpha d_n L(n)$$
$$< 2n(\log n)^\alpha \cdot \tfrac{1}{8}(\log n)^{-\alpha} \cdot L(n) = \tfrac{1}{4} n L(n)$$

for all sufficiently large n. Therefore, the number Z of m, $1 \leqslant m \leqslant L(n)$, such that $S(n, m+2n(\log n)^\alpha) - S(n, m) \geqslant n/2$ satisfies $(n/2)\ Z \leqslant (n/4)L(n)$. That is, for all large n, $Z \leqslant L(n)/2$. Thus for at least $L(n)/2$ choices of m, $1 \leqslant m \leqslant L(n)$, we have $f(n, m) > 2n(\log n)^\alpha$. Thus

$$\sum_{m=1}^{L(n)} f(n, m) > \tfrac{1}{2} L(n) \cdot 2n(\log n)^\alpha = L(n) \cdot n(\log n)^\alpha,$$

for all large n. □

THEOREM 6. Let $\beta = \log(5^{5/2}/2 \cdot 3^{3/2}) = 1.6825 \ldots$. Then

$$L(n)^{-1} \sum_{m=1}^{L(n)} f(n, m) \leqslant n \cdot \exp\left((\beta + o(1))\log n/\text{loglog } n\right).$$

PROOF. Let $\varepsilon > 0$ be arbitrary and let $c = 3/2 + 4\varepsilon$. Let

$$b = \log(1+c) + c\log(1+c^{-1}),$$

so that as $\varepsilon \to 0$, we have $b \to \beta$.

For any integer m let $d_n(m)$ denote the number of divisors d of m with $d \leqslant n$ and let $\omega_n(m)$ denote the number of prime divisors p of m with $p \leqslant n$. Let T_n denote the number of $m \in [1, L(n)]$ with

$$d_n(m) \geqslant \exp(b \log n/\text{loglog } n) \overset{\text{def}}{=} e(n).$$

We now show

$$(15) \qquad T_n \leqslant L(n)/n^{3/2+\varepsilon}.$$

First we note that for any m, if $\omega_n(m) = s$, then $d_n(m) \leqslant \psi(n, p_s)$ where p_s denotes the s-th prime. Indeed, if $q_1, q_2, \ldots, q_s$ are the prime factors of m not exceeding n, then every divisor d of m with $d \leqslant n$ is composed of just the q's. We thus observe that an upper bound for $d_n(m)$ is the number of integers not exceeding n composed of just $p_1, p_2, \ldots, p_s$; that is, $\psi(n, p_s)$.

Now if $p_s < (3/2 + 3\varepsilon)\log n$, it follows from de Bruijn [1], that for all large n, $\psi(n, p_s) < e(n)$. Thus, for large n, $d_n(m) \geqslant e(n)$ implies

$$p_s \geqslant (3/2 + 3\varepsilon)\log n$$

where $s = \omega_n(m)$. This in turn implies that

$$\omega_n(m) \geqslant (3/2 + 2\varepsilon)\log n/\text{loglog } n \overset{\text{def}}{=} r_n.$$

Thus T_n is at most the number of $m \leqslant L(n)$ with $\omega_n(m) \geqslant r_n$. Hence, for large n,

$$T_n \leqslant L(n)\left(\sum_{p \leqslant n} 1/p\right)^{r_n} / r_n! \leqslant L(n) \cdot (2 \operatorname{loglog} n)^{r_n} / (r_n/e)^{r_n}$$

$$= L(n) \cdot \exp\,(r_n \operatorname{logloglog} n + r_n(1 + \log 2) - r_n \log r_n)$$

$$\leqslant L(n) \cdot \exp\,(-(3/2 + \varepsilon) \log n),$$

which gives (15).

Suppose now m is such that in the interval $J = (m, m + n \cdot e(n)] \cap \mathbb{Z}$ there is no integer j with $d_n(j) \geqslant e(n)$. Then $f(n, m) \leqslant n \cdot e(n)$. Indeed, if we consider the bipartite graph from $I = [1, n] \cap \mathbb{Z}$ to J where $i \in I$ is connected to $j \in J$ if $i | j$, then the minimum valence of an $i \in I$ is at least $e(n)$, while the maximum valence of a $j \in J$ is less than $e(n)$. Thus the König-Hall theorem applies.

Now the number of $m \leqslant L(n)$ for which there is an integer $j \in (m, m + n \cdot e(n)]$ with $d_n(j) \geqslant e(n)$ is at most

$$T_n \cdot n \cdot e(n) \leqslant L(n) \cdot n^{-1/2-\varepsilon} \cdot e(n)$$

by (15) for large n. For these m we have $f(n, m) \ll n^{3/2}$ by Theorem 4. We have seen that for the remaining m we have $f(n, m) < n \cdot e(n)$. Thus

$$\sum_{m=1}^{L(n)} f(n, m) \leqslant n^{3/2+\varepsilon} \cdot L(n) \cdot n^{-1/2-\varepsilon} \cdot e(n) + n \cdot e(n) \cdot L(n)$$

$$= 2L(n) \cdot n \cdot e(n)$$

$$< L(n) \cdot n \cdot \exp\,((b - \varepsilon) \log n / \operatorname{loglog} n)$$

for all large n. Thus letting $\varepsilon \to 0$, we have already seen that $b \to \beta$, and so our theorem follows. □

Improvements on the size of β in Theorem 6 are attainable. The limit of the method gives $\beta = \log 4$. However, we believe much more is true. We conjecture that

$$L(n)^{-1} \sum_{m=1}^{L(n)} f(n, m) \ll n (\log n)^{\gamma}$$

for some $\gamma > 0$.

§ 6. OTHER PROBLEMS

If $1 \leqslant k \leqslant n$, let $g(n, k)$ denote the smallest number so that for each choice of integers $1 \leqslant a_1 < \ldots < a_k \leqslant n$, there are distinct integers $b_1, \ldots, b_k$ in $(n, g(n, k)]$ with $a_i | b_i$ for $i = 1, \ldots, k$. Also let $h(n, k)$ denote the least number so that in any interval of length $h(n, k)$ we can find a set of distinct multiples for each k-element subset of $\{1, \ldots, n\}$. Thus $g(n, k)$

$< h(n, k)+n$. In our previous notation we have $f(n)=g(n, n)$,

$$\max_m f(n, m)=h(n, n).$$

By a similar argument as the one which gives Theorem 2 from Lemma 2, we have

$$g(n, k) > [n/k]f(k)$$

so that

(16) $$\liminf_{n\to\infty} g(n, k)/n \geqslant f(k)/k.$$

Mimicking the proof of Theorem 4, we have

(17) $$h(n, k) \ll n\sqrt{k}$$

uniformly for all k, n (with $k<n$). Thus

(18) $$\limsup_{n\to\infty} g(n, k)/n \ll \sqrt{k}.$$

We do not know how to narrow the gap between (16) and (18), but we feel (16) is closer to the truth.

Now we look at particular subsets of $\{1, \ldots, n\}$ that are of interest. Let $f_{\mathscr{P}}(n)$ denote the smallest number so that in $(n, f_{\mathscr{P}}(n)]$ we can find distinct $b_1, \ldots, b_{\pi(n)}$ where $p_i|b_i$ for each i (p_i denotes the i-th prime). It is not too hard to show that for each $n>1$, $f_{\mathscr{P}}(n)=2p_{\pi(n)}$ except that $f_{\mathscr{P}}(4)=8$ and $f_{\mathscr{P}}(10)=16$. More interesting is the function $f_{\mathscr{P}}(n, m)$, the least number so that in $(m, m+f_{\mathscr{P}}(n, m)]$ there are distinct numbers $b_1, \ldots, b_{\pi(n)}$ such that $p_i|b_i$ for each i. The question is, what is the average value of $f_{\mathscr{P}}(n, m)$, that is, what is

$$g_{\mathscr{P}}(n)=M(n)^{-1}\sum_{m=1}^{M(n)} f_{\mathscr{P}}(n, m)$$

where $M(n)$ is the product of the primes not exceeding n? By Theorem 6, we have $g_{\mathscr{P}}(n) \leqslant n^{1+o(1)}$. Perhaps it is possible to show that $g_{\mathscr{P}}(n)/n$ is bounded above by a power of log n. We cannot show (nor are we sure we believe) that $g_{\mathscr{P}}(n)/n$ is unbounded.

Now let

$$h_{\mathscr{P}}(n)=\max_m f_{\mathscr{P}}(n, m).$$

We know very little about $h_{\mathscr{P}}(n)$. Erdös and Selfridge can show, using Brun's method, that

(19) $$\limsup_{n\to\infty} h_{\mathscr{P}}(n)/n \geqslant 3.$$

Using (17) in the case $k=\pi(n)$, we have

(20) $$h_{\mathscr{P}}(n)/n \ll \sqrt{n/\log n}.$$

We do not know how to narrow the immense gulf between (19) and (20).

Related to these questions, we ask if there is a large constant c so that in any interval of length cn there are $\pi(n)-\pi(n/2)$ distinct multiples of the primes in $(n/2, n]$ (there need not be a matching). If yes, what is the smallest value of c? The same question can be asked if $p_1<\ldots<p_k$ is any set of primes, but now "cn" should be replaced by "cp_k". Erdös and Selfridge have shown that for every k there is a set of primes $p_1<\ldots<p_{k^2}$ with only $2k$ multiples in some interval of length $(3-o(1))p_{k^2}$. This is how (19) is established.

Is it true that for a large enough c, every interval of length cn contains a number divisible by precisely one prime in $(n/2, n]$? What if we replace the primes in $(n/2, n]$ with the primes in $[1, n]$?

Let $f_0(n)$ denote the least number so that in $(n, f_0(n)]$ we can find distinct numbers $b_1, \ldots, b_t$ where $a_i|b_i$ for each i and $\{a_1, \ldots, a_t\}$ is the set of numbers not exceeding n divisible by no prime exceeding $\log n$. Theorems 1 and 3 immediately give inequalities for $f_0(n)$. However, Theorem 2 does not seem to carry over for $f_0(n)$, although Lemma 2 does. Is it true that $f_0(n)=f(n)$ for all sufficiently large n, or for almost all n?

Let $f_{\mathscr{D}}(n)$ denote the least number so that in $(n, f_{\mathscr{D}}(n)]$ we can find distinct numbers b_d for each $d|n$ such that $d|b_d$. We at first thought that $f_{\mathscr{D}}(n)$ could be as large as $f_0(n)$ by considering highly composite choices for n. But a very simple proof shows $f_{\mathscr{D}}(n)=2n$ for every choice of n. Indeed, let $b_d=n+d$.

Given a particular set of integers $0<a_1<\ldots<a_k$, what is the length of the shortest interval which contains distinct numbers $b_1, \ldots, b_k$ with $a_i|b_i$ for each i? Say, for example, p, q, r are distinct odd primes and $a_1=pq$, $a_2=pr$, $a_3=qr$. Let d_1, d_2, d_3 be the minimal integers with

$$d_1p=e_1q+f_1r,\ \ d_2q=e_2p+f_2r,\ \ d_3r=e_3p+f_3q$$

such that the e_i, f_i are positive integers. Then it is easy to show that the shortest interval which contains distinct numbers b_1, b_2, b_3 with $a_i|b_i$ for each i has length $l=\min\{d_1p, d_2q, d_3r\}$. If $p<q<r$,

$$\tfrac{1}{2}(r-p)q=\tfrac{1}{2}(r-q)p+\tfrac{1}{2}(q-p)r,$$

so that $d_2\leqslant\frac{1}{2}(r-p)$. Thus $l\leqslant\frac{1}{2}(r-p)q$ which is half the length of $[pq, qr]$. Does equality hold infinitely often?

In the introduction we remarked that we cannot show $\max_m f(n, m) > f(n, n)$. Nevertheless, we believe this to be the case for all $n\geqslant 5$. In fact we conjecture

$$\max_m f(n, m)-f(n, n)\to\infty.$$

All we can prove is that there are infinitely many n with

$$\max_m f(n, m)-f(n, n)\geqslant 1. \tag{21}$$

In fact (21) holds if n is a sufficiently large prime p. In this case $f(p, p-1) - f(p, p) = 1$. To see this, suppose not, so that there exist

$$a_1, \ldots, a_p \in [p, p+f(p, p)-1]$$

distinct with $i|a_i$. Thus $f(p-1, p-1) \leqslant f(p, p)$. But this inequality is untrue for all large p. What is true is that $f(p-1, p-1) = 1 + f(p, p)$, since given the mapping of $\{1, \ldots, p-1\}$ into $\{p, \ldots, p-1+f(p-1, p-1)\}$ we note that p need not be used as an image. Thus we can map $\{1, \ldots, p\}$ into $\{p+1, \ldots, p-1+f(p-1, p-1)\}$ by sending p to $2p$. Note that $2p$ need not be used as an image for 1 or 2 – we may use p' for 1 and $2p'$ for 2 where $p' > p$ is prime. Probably it is possible to show the left side of (21) is unbounded, but we are not sure of the details.

Another problem that is perhaps of some interest is to estimate $\delta(n, c)$, the asymptotic density of the set of n with $f(n, m) \leqslant cn$. It is clear the density exists since $f(n, m)$ is periodic in m. Moreover if $c \geqslant 1$, then $\delta(n, c) > 0$. Even the case $c = 1$ provides some interesting considerations.

REFERENCES

1. Bruijn, N. G. de – On the number of positive integers $\leqslant x$ and free of prime factors $>y$. II. Nederl. Akad. Wetensch. Proc. Ser. **A 69** = Indag. Math. **28,** 239–247 (1966).
2. Ennola, V. – On numbers with small prime divisors, Ann. Acad. Sci. Fenn. Ser. A I No. **440,** 16 pp. (1969).
3. Erdös, P. and J. L. Selfridge – Some problems on the prime factors of consecutive integers. II. Proc. Washington State U. Conf. on Number Th. at Pullman, 13–21 (1971).
4. Grimm, C. A. – A conjecture on consecutive composite numbers, Amer. Math. Monthly **76,** 1126–1128 (1969).
5. Hall, P. – On representatives of subsets, J. London Math. Soc. **10,** 26–30 (1935).
6. Hardy, G. H. – *Ramanujan*, Cambridge, Ch. V (1940).
7. König, D. – Über Graphen und ihre Anwendung auf Determinantentheorie und Mengenlehre, Math. Ann. **77,** 453–465 (1915–16).
8. Pillai, S. S. – On numbers of the form $2^a . 3^b$. I. Proc. Nat. Acad. Sci. India Sect. **A 15,** 128–132 (1942).
9. Pomerance, C. – The prime number graph, Math. Comp. **33,** 399–408 (1979).
10. Ramachandra, K. T., N. Shorey and R. Tijdeman – On Grimm's problem relating to factorization of a block of consecutive integers, J. reine angew. Math. **273,** 109–124 (1975).
11. Specht, W. – Zahlenfolgen mit endlich vielen Primteilern, Bayer. Akad. Wiss. Math.-Natur. Abt. S.-B. **1948,** 149–169 (1949).
12. Tenenbaum, G. – Sur la répartition des diviseurs, Sem. Delange-Pisot-Poitou **17,** No. G14, 5 pp. (1975/76).