On the largest prime factors of $n$ and $n+1$

Paul Erdős and Carl Pomerance

§1. Introduction

If $n \geq 2$ is an integer, let $P(n)$ denote the largest prime factor of $n$. For every $x > 0$ and every $t$, $0 \leq t \leq 1$, let $A(x, t)$ denote the number of $n \leq x$ with $P(n) \geq x^t$. A well-known result due to Dickman [4] and others is

**Theorem A.** The function

$$a(t) = \lim_{x \to \infty} x^{-1} A(x, t)$$

is defined and continuous on $[0, 1]$.

In fact it is even shown that $a(t)$ is strictly decreasing and differentiable. Note that $a(0) = 1$ and $a(1) = 0$.

If $0 \leq t$, $s \leq 1$, denote by $B(x, t, s)$ the number of $n \leq x$ with $P(n) \geq x^t$ and $P(n+1) \geq x^s$. One might guess that

$$b(t, s) = \lim_{x \to \infty} x^{-1} B(x, t, s)$$

exists and is continuous on $[0, 1]^2$. In fact, one could guess that

$$b(t, s) = a(t)a(s);$$

that is, the largest prime factors of $n$ and $n+1$ are "independent events." We do not know how to prove the above guesses. In fact, we cannot even prove the almost certain truth that the density of integers $n$ with $P(n) > P(n+1)$ is $\frac{1}{2}$.

However we can prove:

**Theorem 1.** For each $\epsilon > 0$, there is a $\delta > 0$ such that for sufficiently large $x$,
the number of \( n \leq x \) with

\[
x^{-8} < P(n)/P(n+1) < x^8
\]

is less than \( \varepsilon x \).

That is, \( P(n) \) and \( P(n+1) \) are usually not close. We use Brun’s method in the proof. One corollary is that the lower density of integers \( n \) for which \( P(n) > P(n+1) \) is positive (see §6).

If the canonical prime factorization of \( n > 1 \) is \( \prod p_i^{\alpha_i} \), let \( f(n) = \sum a_i p_i \); and let \( f(1) = 0 \). Several authors have considered this function or the closely related \( g(n) = \sum p_i \) or \( h(n) = \sum p_i^{\alpha_i} \), among them Alladi and Erdös [1], Chawla [2], Dane [3], Hall [7], Lal [10], LeVan [12], and Nicolas [14]. In Nelson, Penney, and Pomerance [13] the following problem is raised: does the set of \( n \) for which \( f(n) = f(n+1) \) have density 0? If \( f(n) = f(n+1) \), we call \( n \) an Aaron number (see [13]). We prove here the Aaron numbers do indeed have density 0. The result follows as a corollary to Theorem 1 and

**THEOREM 2.** For every \( \varepsilon > 0 \), there is a \( \delta > 0 \) such that for sufficiently large \( x \) there are at least \( (1-\varepsilon)x \) choices for \( n \leq x \) such that

\[
P(n) < f(n) < (1 + x^{-\delta})P(n).
\]

Theorem 2 implies that usually \( f(n) \approx P(n) \) and \( f(n+1) \approx P(n+1) \). But Theorem 1 implies \( P(n) \) and \( P(n+1) \) are usually not close. Hence \( f(n) \) and \( f(n+1) \) are usually not close, and in particular, we usually have \( f(n) \neq f(n+1) \). This then establishes that the Aaron numbers have density 0. However we can prove a sharper result:

**THEOREM 3.** For every \( \varepsilon > 0 \), the number of \( n \leq x \) for which \( f(n) = f(n+1) \) is \( O(x/(\log x)^{1-\varepsilon}) \).

Actually we can prove the sharper estimate \( O(x/\log x) \), but the proof is more difficult than the proof of Theorem 3 and we do not present it here. We suspect that the estimate \( O(x/(\log x)^k) \) is true for every \( k \), but we cannot prove this for any \( k > 1 \). In fact, we cannot even get \( o(x/\log x) \). On the other hand, we cannot prove that there are infinitely many Aaron numbers (this would follow if Schinzel’s Conjecture \( H \) is true – see [13]). But by a consideration of those \( n \) for which \( P(n) \) and \( P(n+1) \) are both relatively small, we believe the number of Aaron numbers up to \( x \) is \( \Omega(x^{1-\varepsilon}) \) for every \( \varepsilon > 0 \).
There are integers $n$ for which $f(n) = f(n+1) = f(n+2)$. The least example, kindly found for us by David E. Penney in a computer search, is $n = 417162$. We cannot prove that the number of such $n \leq x$ is $o(x/\log x)$. We conjecture that for every $k$ there are integers $n$ with $f(n) = f(n+1) = \cdots = f(n+k)$.

§2. Preliminaries

In this section we record several lemmas which will be useful in our discussion. The letter $p$ denotes a prime.

**Lemma 1.** There is an absolute constant $C$, such that if $3 < u < v$, then

$$\sum_{u \leq p \leq v} \frac{1}{p} < \frac{C + \log(v/u)}{\log u}.$$  

This lemma is used when $u$ is large compared with $v/u$. The proof follows easily from the classical result (see Hardy and Wright [8], Theorem 427 and its proof): there are absolute constants $B$, $D$ such that if $x \geq 3$, then

$$\left| \sum_{p \leq x} \frac{1}{p} - \log \log x - B \right| < \frac{D}{\log x}.$$  

Lemma 1 easily follows with $C = 2D$.

**Lemma 2.** \(\sum_{p \geq t} \frac{1}{p \log p} \sim \frac{1}{\log t}\).

**Proof.** If $p_k$ denotes the $k$-th prime, then $p_k \sim k \log k$ and

$$\sum_{p \geq t} \frac{1}{p \log p} \sim \sum_{k \pi(t)} \frac{1}{k \log^2 k} \sim \frac{1}{\log \pi(t)} \sim \frac{1}{\log t}.$$  

**Lemma 3.** If $P(n) \geq 5$, then $f(n) \leq P(n) \log n/\log P(n)$.

**Proof.** We use the fact that $t/\log t$ is increasing for $t \geq e$ and $2/\log 2 < 5/\log 5$. Write $n = \prod p_i^{a_i}$, where $p_1 = P(n)$. Then

$$f(n) = \sum a_i p_i \leq \sum a_i p_1 \log p_i/\log p_1 = P(n) \log n/\log P(n).$$
§3. Proof of Theorem 1

Let $\epsilon > 0$. From Theorem A it follows there is a $\delta_0 = \delta_0(\epsilon)$ such that $\frac{1}{4} > \delta_0 > 0$ and for large $x$ the number of $n \leq x$ with

$$P(n) < x^{\delta_0} \quad \text{or} \quad x^{1/2 - \delta_0} \leq P(n) < x^{1/2 + \delta_0}$$

is less than $\epsilon x/3$. We now consider the remaining $n \leq x$. There are 2 cases:

(i) $x^{\delta_0} \leq P(n) < x^{1/2 - \delta_0}$,

(ii) $x^{1/2 + \delta_0} \leq P(n)$.

For each pair of primes $p$, $q$, the number of $n \leq x$ for which $P(n) = p$, $P(n + 1) = q$ is at most $1 + [x/pq]$. Then for large $x$, the number of $n \leq x$ in case (i) for which (1) holds is at most (assume $0 < \delta < \delta_0/4$)

$$\sum_{\substack{x^{\delta_0} \leq p < x^{1/2 - \delta_0} \\ px^{-\delta} < q < px^\delta}} 1 + [x/pq] < x^{1-2\delta_0 + \delta} + x \sum_p \frac{1}{p} \sum_q \frac{1}{q}$$

$$< x^{1-2\delta_0 + \delta} + x \sum_p \frac{1}{p} \cdot \frac{C + \log (x^{2\delta})}{\log (px^{-\delta})} \quad \text{(Lemma 1)}$$

$$< x^{1-2\delta_0 + \delta} + 3\delta x \log x \sum_p \frac{1}{p \log p}$$

$$< x^{1-2\delta_0 + \delta} + 4\delta x / \delta_0 \quad \text{(Lemma 2)} \quad (3)$$

Hence if we choose $\delta$ so that

$$0 < \delta < \delta_0 \epsilon / 13, \quad (4)$$

then (3) implies there are fewer than $\epsilon x/3$ choices of such $n$.

Suppose now $n \leq x$ is in case (ii) and (1) holds. Let $a = n/P(n)$, $b = (n + 1)/P(n + 1)$. Then $a \leq x^{1/2 - \delta_0}$, $b < x^{1/2 - \delta_0 + \delta}$, and $x^{-\delta}/2 < a/b < 2x^\delta$. On the other hand, given integers $a$, $b$, the number of $n \leq x$ for which $n = aP(n)$ and $n + 1 = bP(n + 1)$ is at most the number of primes $p \leq x/a$ such that $(ap + 1)/b$ is prime. (Note that there is at most one such prime $p$ unless $(a, b) = 1$ and $2 \mid ab$.) All such primes $p$ are in a fixed residue class mod $b$, say $p = kb + c$ for some $k \geq 0$. Let $d = (ac + 1)/b$. Then we are counting integers $k$ with $0 \leq k < x/ab$ such that $kb + c$ and $ka + d$ are simultaneously prime. By Brun's method (see Halberstam
and Richert [6], Theorem 2.3, p. 70), we have the number of such \( k \) is at most

\[
\frac{Ax}{ab \log^2(x/ab) \prod_{p \mid ab} \left(1 - \frac{1}{p}\right)^{-1}} = \frac{Ax}{\varphi(a)\varphi(b) \log^2(x/ab)}
\]

where \( A \) is an absolute constant (independent of the choice of \( a, b \)) and \( \varphi \) is Euler's function. Hence for sufficiently large \( x \), the number of \( n \leq x \) in case (ii) for which (1) holds is at most

\[
Ax \sum_{\frac{1}{2} \leq \delta \leq \delta_0, \quad a \leq x, \quad ax^{-\delta} < b < 2ax^\delta} \frac{1}{\varphi(a)\varphi(b)} \log^2(x/ab) \tag{5}
\]

\[
\leq \frac{2Ax}{(2\delta_0 - \delta)^2 \log^2 x} \sum \frac{1}{\varphi(a)} \sum \frac{1}{\varphi(b)}. \tag{5}
\]

We now use the result of Landau [11], that if \( E = \zeta(2)\zeta(3)/\zeta(6) \), then

\[
\sum_{n \leq x} 1/\varphi(n) = E \log x + o(1).
\]

Hence for large \( x \) the quantity in (5) is less than

\[
\frac{3EAx}{(2\delta_0 - \delta)^2 \log^2 x} \sum \frac{\log(x^{2\delta})}{\varphi(a)} \tag{6}
\]

\[
= \frac{6\delta EAx}{(2\delta_0 - \delta)^2 \log x} \sum \frac{1}{\varphi(a)} \tag{6}
\]

\[
< \frac{7\delta E^2 Ax}{(2\delta_0 - \delta)^2 \log x} \log(x^{1/2 - \delta_0}) \tag{6}
\]

\[
< \frac{4\delta E^2 Ax}{(2\delta_0 - \delta)^2}. \tag{6}
\]

If we now choose \( \delta \) so that

\[
0 < \delta < \delta_0^2 \epsilon/4E^2A \quad \text{and} \quad \delta < \delta_0/4, \quad \tag{7}
\]

then (6) implies there are fewer than \( \epsilon x/3 \) choices for such \( n \). Hence if we choose \( \delta \) so that (4) and (7) hold, it follows that the number of \( n \leq x \) for which (1) holds is
less than $\varepsilon x$ for every sufficiently large value of $x$ (depending, of course, on $\varepsilon$). This completes our proof.

Note that using a known explicit estimate for the upper bound sieve result we may take $A = 8 + o_x(1)$.

§4. The proof of Theorem 2

Since any integer $n \leq x$ is divisible by at most $\log x / \log 2$ primes, we have for large $x$ and composite $n \leq x$

$$f(n) = P(n) + f(n/P(n)) \leq P(n) + P(n/p(n)) \log x / \log 2$$

$$< P(n) + P(n/P(n)) x^{\delta}.$$  \hspace{1cm} (8)

If (2) fails, then, but for $o(x)$ choices of $n \leq x$, we have

$$f(n) \geq (1 + x^{-\delta}) P(n),$$

so that from (8) and (9) we have

$$P(n/P(n)) > x^{-2\delta} P(n).$$  \hspace{1cm} (10)

Let $\varepsilon > 0$. From Theorem A there is a $\delta_0 = \delta_0(\varepsilon) > 0$ such that for large $x$, the number of $n \leq x$ with $P(n) < x^{\delta_0}$ is at most $\varepsilon x / 3$. For each pair of primes $p$, $q$ the number of $n \leq x$ with $P(n) = p$ and $P(n/P(n)) = q$ is at most $[x/pq]$. Hence from (10), for large $x$ the number of $n \leq x$ for which (2) fails is at most (assume $0 < \delta < \delta_0 / 7$)

$$o(x) + \varepsilon x / 3 + \sum_{x^{\delta_0} \leq p \leq x^{\frac{1}{2}} \leq q \leq x^{\delta_0}} [x/pq] < \varepsilon x / 2 + x \sum_{p \leq x^{\delta_0}} \frac{1}{p} \sum_{q \leq p} \frac{1}{q}$$

$$< \varepsilon x / 2 + x \sum_{p \leq x^{\delta_0}} \frac{1}{p} \cdot \frac{C + \log (x^{2\delta})}{\log (x^{2\delta} p)}$$

$$< \varepsilon x / 2 + 3\delta x \log x \sum_{p \leq x^{\delta_0}} \frac{1}{p \log p}$$

$$< \varepsilon x / 2 + 4\delta x / \delta_0$$

$$\leq \varepsilon x,$$

(Lemma 1)

(Lemma 2)

if we take $\delta = \delta_0 / 8$. This completes the proof.
§5. Aaron numbers

In this section we prove Theorem 3. Let \( x \) be large, \( n \leq x \), and \( f(n) = f(n + 1) \). We distinguish two cases:

(i) \( P(n) > x^{1/2} \),
(ii) \( P(n) \leq x^{1/2} \).

Let \( n \) be in case (i). We first show that

\[
P(n + 1) > P(n)/3.
\] (11)

Indeed we have

\[
x^{1/2} < P(n) \leq f(n) = f(n + 1) \leq P(n + 1) \log (x + 1)/\log 2
\]
so that \( P(n + 1) > x^{1/2} \log 2/\log (x + 1) \). Hence Lemma 3 implies

\[
P(n) < P(n + 1) \log (x + 1)/\log (x^{1/2} \log 2/\log (x + 1)) < 3P(n + 1)
\]
for large \( x \), which proves (11). We next show that

\[
|P(n) - P(n + 1)| < 4x/P(n).
\] (12)

Indeed, \( f(n) = f(n + 1) \) implies

\[
P(n + 1) - P(n) = f(n/P(n)) - f((n + 1)/P(n + 1)) \leq n/P(n),
P(n) - P(n + 1) \leq (n + 1)/P(n + 1),
\]
so that using (11) we have (12). We next show that

\[
P(n) < 3x^{2/3}.
\] (13)

We use the congruence

\[
(P(n + 1) - P(n)) \frac{n + 1}{P(n + 1)} = 1 \pmod{P(n)}.
\] (14)

From (11) we have \( P(n) \) and \( P(n + 1) \) both odd primes so the left side of (14) is
not 1. Then (11), (12), and (14) imply

$$P(n) \leq \left| P(n) - P(n+1) \right| \frac{n+1}{P(n+1)} + 1 < \frac{4x}{P(n)} \cdot \frac{x+1}{P(n+1)} + 1$$

$$< \frac{12x(x+1)}{P(n)^2} + 1 < \frac{14x^2}{P(n)^2}$$

for large $x$, so that (13) follows.

If $p, q$ are primes with $x^{1/2} < p, q > p/3$, then there are at most 3 integers $n \leq x$ with $P(n) = p$ and $P(n+1) = q$. Hence from (11), (12), (13) we have for large $x$ that the number of $n \leq x$ in case (i) for which $f(n) = f(n+1)$ is at most

$$3 \sum_{x^{1/2} < p < 3x^{2/3}} \sum_{|p-q| < 4x/p} \frac{x/p}{\log (x/p)}$$

$$\ll \sum \frac{x}{p \log x} \ll \frac{x}{\log x},$$

where we use the well-known result of Hardy and Littlewood (see [9], p. 66) for the number of primes in an interval and Lemma 1.

We now turn our attention to case (ii). We have (see Erdős [5], proof of Lemma 1 or Rankin [15], Lemma II) the number of $n \leq x$ for which we do not have

$$P(n) > x^{1/3 \log \log x}$$

is $O(x/\log x)$. So we may assume (16) holds. Then using Lemma 3 and the argument which establishes (11), we have from the equation $f(n) = f(n+1)$ that

$$P(n)/4 \log \log x < P(n+1) < 3P(n) \log \log x.$$  \hspace{1cm} (17)

For each pair of primes $p, q$, there are at most $1 + [x/pq]$ integers $n \leq x$ with $P(n) = p$ and $P(n+1) = q$. Hence from (16) and (17), for large $x$ the number of $n \leq x$ in case (ii) for which $f(n) = f(n+1)$ is at most

$$\sum_{x^{1/2} \log \log x < p \leq x^{1/2}} \pi(x^{1/2}) \pi(3x^{1/2} \log \log x) + x \sum_p \frac{1}{P} \sum_q \frac{1}{q}$$

$$\ll \frac{x}{\log x} + x \sum_p \frac{1}{p} \cdot \frac{\log \log \log x}{\log p}$$

(Lemma 1)
\[
\frac{x \log \log x \log \log \log x}{\log x}.
\]

(Lemma 2)

This completes the proof of Theorem 3.

§6. The probability that \( P(n) > P(n+1) \).

Using some computer estimates of the function \( a(t) \) made with the generous assistance of Don R. Wilhelmsen, it can be shown that the number of integers \( n \leq x \) such that

\[
x^{0.31} \leq P(n) < x^{0.46}
\]

is more than \( 0.2002x \) for sufficiently large \( x \). By an elementary argument similar to the proof of case (i) in Theorem 1 (see §3) one can show the number of \( n \leq x \) for which (18) holds and for which

\[
P(n) < P(n+1) < P(n)x^{0.08}
\]

is less than \( 0.0763x \) for sufficiently large \( x \). Hence the number of \( n \leq x \) for which (19) fails is more than

\[
0.2002x - 0.0763x = 0.1239x
\]

for sufficiently large \( x \). Now for every \( k \) choices of \( n \leq x \) for which \( P(n+1) \geq P(n)x^{0.08} \), there must be at least \( \lfloor 0.08k \rfloor \) integers \( n \) in the same interval for which \( P(n) > P(n+1) \). Hence the lower density of integers \( n \) for which \( P(n) > P(n+1) \) is at least

\[
(0.08) \cdot (0.1239) > 0.0099.
\]

Note that the same is true for integers \( n \) for which \( P(n) < P(n+1) \). Undoubtedly improvements in this type of result are possible.

§7. Comments on three or more consecutive numbers.

It is easy to show that the patterns

\[
P(n) < P(n+1), \ P(n+1) > P(n+2);
P(n) > P(n+1), \ P(n+1) < P(n+2),
\]
both occur infinitely often. However we cannot prove either of these two patterns occurs for a positive density of \( n \), although this certainly must be the case. Suppose now \( p \) is an odd prime and

\[
k_0 = \inf \{ k : P(p^{2^k} + 1) > p \}
\]

(note that \( P(p^{2^k} + 1) \equiv 1 \pmod{2^{k_0+1}} \), so \( k_0 < \infty \)). Then

\[
P(p^{2^{k_0}} - 1) < P(p^{2^{k_0}}) < P(p^{2^{k_0}} + 1).
\]

On the other hand, we cannot find infinitely many \( n \) for which

\[
P(n) > P(n + 1) > P(n + 2),
\]

but perhaps we overlook a simple proof.

Suppose now

\[
\epsilon_n = \begin{cases} 1, & \text{if } P(n) > P(n + 1), \\ 0, & \text{if } P(n) < P(n + 1). \end{cases}
\]

Then \( \sum_{n=2}^{\infty} \epsilon_n 2^n \) is irrational. Indeed, suppose not, so that \( \{\epsilon_n\} \) is eventually periodic with period length \( K \). Let \( p > K \) be a fixed prime. An old and well-known result of Pólya implies that there are only finitely many pairs of consecutive integers in the set \( M = \{ n : P(n) \leq p \} \). (In fact, from the work of Baker, the largest consecutive pair in \( M \) is effectively computable.) Note that \( p^i, 2p^i, \ldots, Kp^i \) are all in \( M \) for every \( i \). Hence for large \( i \), none of \( p^i + 1, 2p^i + 1, \ldots, Kp^i + 1 \) is in \( M \), so that \( \epsilon_n = 0 \) for \( m = p^i, 2p^i, \ldots, Kp^i \). But these numbers form a complete residue system mod \( K \). Hence \( \epsilon_n = 0 \) for every large \( n \), an absurdity.

For each \( k \), let \( h(k) \) denote the number of different patterns of \( k \) consecutive terms of \( \{\epsilon_n\} \) which occur infinitely often. Surely we must have \( h(k) = 2^k \). This is easy for \( k = 1 \), but already for \( k = 2 \), all we can prove is \( h(2) \geq 3 \). (If there are infinitely many \( n \) for which (20) holds, then \( h(2) = 4 \).) It follows from the non-periodicity of \( \{\epsilon_n\} \) that for every \( k \),

\[
h(k) \geq k + 1.
\]

To see this, it is sufficient to show \( h(k) \) is strictly increasing (since \( h(1) = 2 \)). But if \( h(k) = h(k + 1) \) (clearly \( h(k) > h(k + 1) \) is impossible), then sufficiently far out in the sequence \( \{\epsilon_n\} \) we have each term determined by the previous \( k \) terms. Then as soon as a \( k \)-tuple repeats, the sequence repeats and hence is periodic.
We remark that \( h(k) = 2^k \) can be seen to follow from the prime \( k \)-tuples conjecture.

REFERENCES


Department of Mathematics, University of Georgia, Athens, Georgia 30602 U.S.A.