ACTA ARITHMETICA
LVIII.4 (1991)

Carmichael's lambda function

by

PAUL ERDŐS (Budapest), CARL POMERANCE* (Athens, Ga.) and
ERIC SCHMUTZ* (Philadelphia, Penn.)

1. Introduction. Let $\lambda(n)$ be the universal exponent for the group of residues mod n that are coprime to n. A more explicit definition of λ is:

$$\begin{array}{ll} \lambda(p^e) = \phi(p^e) = p^{e-1}(p-1) & \text{if } p \text{ is an odd prime,} \\ \lambda(2^e) = \phi(2^e) & \text{if } e = 0, 1, \text{ or } 2, \\ \lambda(2^e) = \tfrac{1}{2}\phi(2^e) & \text{if } e \geqslant 3 \end{array}$$

and finally,

$$\lambda(n) = \text{l.c.m.}\big(\lambda(p_1^{e_1}), \ldots, \lambda(p_\nu^{e_\nu})\big) \quad \text{if } n = p_1^{e_1} \ldots p_\nu^{e_\nu} \quad (p_i\text{'s distinct primes}).$$

This is Carmichael's function [3]. Not only is it an intrinsically interesting number theoretic function, $\lambda(n)$ has a connection with some primality testing algorithms [1, 11]. In this paper we investigate the average order, normal order, and minimal order of λ.

Estimates for the minimal order are already implicit in the analysis of the primality testing algorithms in [1]. But they are not immediately obvious, so it is worthwhile to make them explicit here:

THEOREM 1. *For any increasing sequence $\langle n_i \rangle_i$ of positive integers, and any positive constant $c_0 < 1/\log 2$, one has*

$$\lambda(n_i) > (\log n_i)^{c_0 \log\log\log n_i}$$

for i sufficiently large. On the other hand, there exists a sequence $n_1 < n_2 < \ldots$, and a constant c_1 with $\lambda(n_i) < (\log n_i)^{c_1 \log\log\log n_i}$ for all i.

The normal order of $\log(\lambda(n)/n)$ was stated without proof by the first author in [5]. Here we prove more:

THEOREM 2. *There is a set S of positive integers of asymptotic density* 1 *such that, for $n \in S$,*

$$\lambda(n) = n/(\log n)^{\log\log\log n + A + O((\log\log\log n)^{-1+\varepsilon})}$$

* Research supported by the National Science Foundation.

where (with q running over primes)

$$A := -1+\sum_q \frac{\log q}{(q-1)^2} = .2269688\ldots,$$

and $\varepsilon > 0$ is fixed but arbitrarily small.

Another result that was stated without proof in [5] is the following estimate for the average order: for all $\varepsilon > 0$, $k > 0$ and for $x > x_0(\varepsilon, k)$,

$$\frac{x}{\log x}(\log\log x)^k \leqslant \frac{1}{x}\sum_{n\leqslant x} \lambda(n) \leqslant \frac{x}{(\log x)^{1-\varepsilon}}.$$

We prove a sharper result here:

THEOREM 3. *For all $x \geqslant 16$, we have*

$$\frac{1}{x}\sum_{n\leqslant x} \lambda(n) = \frac{x}{\log x}\exp\left[\frac{B\log\log x}{\log\log\log x}(1+o(1))\right]$$

where (with q running over primes)

$$B = e^{-\gamma}\prod_q\left(1-\frac{1}{(q-1)^2(q+1)}\right) = .34537\ldots$$

Before proving these theorems, let us fix some global notations that will be used consistently throughout the paper. First, c, c', and c'' will be generic positive constants, not necessarily the same at different places. Second, p and q will denote primes. (Usually p will be a prime factor of n, and q a prime factor of $\lambda(n)$.) Third, let $v_q(m)$ denote the integer $v \geqslant 0$ for which $q^v \mid m$ and $q^{v+1} \nmid m$. Fourth, we let $y = \log\log x$. Finally, if S is a set, let $\omega(n, S)$ denote the number of distinct prime divisors of n that are in S; if S contains all the primes, let $\omega(n) := \omega(n, S)$.

We are grateful to Andrew Granville for calling our attention to a small error in the proof of Theorem 1 in an earlier draft of this paper.

2. Minimal order. In [1], using ideas from [14], it is shown that there is a computable constant $c_2 > 0$ with the property that, for any $x > 10$, there is a square-free number $m_x < x^2$ for which

$$\sum_{p-1\mid m_x} 1 > e^{c_2 \log x/\log\log x}.$$

Let $x_i := (\log i)^{(2/c_2)\log\log\log i}$, and let $n_i = \prod_{p-1\mid m_{x_i}} p$. Note that, for i sufficiently large, we have

$$n_i > \prod_{p-1\mid m_{x_i}} 2 > \exp\left[(\log 2)\exp\left[\frac{c_2\log x_i}{\log\log x_i}\right]\right] > i.$$

But then, for i sufficiently large,

$$\lambda(n_i) \leqslant m_{x_i} < x_i^2 = (\log i)^{(4c_2)\log\log\log i} < (\log n_i)^{c_1 \log\log\log n_i}.$$

By taking a subsequence $\langle n_{i_j}\rangle_j$, we can obtain a sequence that is increasing and satisfies the inequality for all j.

For optimality, first note that it is obvious that $\lambda(n)\to\infty$ as $n\to\infty$. Suppose that $\lambda(n)=k$, so that

$$k=\text{l.c.m.}\{\lambda(p^\alpha):\ p^\alpha\,|\,n\}.$$

Then, since we always have $p^\alpha\leqslant 4\lambda(p^\alpha)$, and since λ is at most 3-to-1 when restricted to primes and prime powers,

$$n\leqslant\prod_{\lambda(p^\alpha)|k}p^\alpha\leqslant\prod_{d|k}(4k)^3\leqslant(4k)^{3d(k)}, \tag{1}$$

where $d(m)$ denotes the number of divisors that m has. It is known [8, 17] that $d(m)\leqslant 2^{(1+o(1))\log m/\log\log m}$. Putting this in (1) gives

$$n\leqslant\exp[(3\log 4k)2^{(1+o(1))\log k/\log\log k}],$$

so that

$$\lambda(n)=k\geqslant(\log n)^{(1/\log 2+o(1))\log\log\log n}$$

as $n\to\infty$. This concludes the proof of the theorem. ∎

It has been conjectured in [1] (see Remark 6.2) that $1/\log 2$ is the "right constant".

3. Normal order. First observe that

$$\log(n/\lambda(n))=\log\phi(n)-\log\lambda(n)+\log(n/\phi(n)).$$

It is well known [9, p. 353] that $n/\log\log n\ll\phi(n)\leqslant n$. Hence, to prove the theorem, it is sufficient to show that, but for $o(x)$ choices of $n\leqslant x$, we have

$$\log\phi(n)-\log\lambda(n)=y\log y+Ay+O\left(\frac{y}{(\log y)^{1-\varepsilon}}\right). \tag{2}$$

(Recall that $y=y(x)=\log\log x$.) For all n we have

$$\log\phi(n)=\sum_q v_q(\phi(n))\log q,\quad \log\lambda(n)=\sum_q v_q(\lambda(n))\log q. \tag{3}$$

To prove (2), we break the sums in (3) into several ranges for the prime q:

$$I_1:\ q\leqslant y/\log y,\qquad I_2:\ y/\log y<q\leqslant y\log y,$$
$$I_3:\ y\log y<q\leqslant y^2,\qquad I_4:\ q>y^2.$$

(These intervals are also listed in order of declining importance for (2).)

We first compute the contribution to $\log\phi(n)$ from primes in I_1 and I_2. Let $h(n):=\sum_{q\leqslant y\log y}v_q(\phi(n))\log q$, so that $h(n)$ is an additive function. The strategy is to apply the Turán–Kubilius inequality [4] to $h(n)$. First we must estimate

$$\sum_{p^k\le x}\frac{h(p^k)}{p^k}\left(1-\frac{1}{p}\right).$$

We use the inequality $h(p^k)\le\log\phi(p^k)\le\log(p^k)$, getting

$$\sum_{p^k\le x}\frac{h(p^k)}{p^k}\left(1-\frac{1}{p}\right)=\sum_{p\le x}\frac{h(p)}{p}+O(1)=\sum_{q\le y\log y}\log q\sum_{p\le x}\frac{v_q(p-1)}{p}+O(1)$$

$$=\sum_{q\le y\log y}\log q\sum_{i\ge 1}\sum_{\substack{p\le x\\ p\equiv 1\,(q^i)}}\frac{1}{p}+O(1)$$

$$=\sum_{q\le y\log y}\log q\sum_{i=1}^{\infty}\left(\frac{y}{\phi(q^i)}+O\left(\frac{\log(q^i)}{q^i}\right)\right)$$

by the estimates in [12]. This in turn is equal to

$$y\sum_{q\le y\log y}\frac{\log q}{q-1}\sum_{i=1}^{\infty}\frac{1}{q^{i-1}}+O\left(\sum_{q\le y\log y}\sum_{i=1}^{\infty}\frac{i\log^2 y}{q^i}\right)=y\sum_{q\le y\log y}\frac{q\log q}{(q-1)^2}+O(\log^3 y)$$

$$=y\sum_{q\le y\log y}\frac{\log q}{q}+y\sum_{q}\frac{(2q-1)\log q}{q(q-1)^2}-y\sum_{q>y\log y}\frac{(2q-1)\log q}{q(q-1)^2}+O(\log^3 y).$$

If we let

$$c_3:=\lim_{x\to\infty}\left(\sum_{q\le x}\frac{\log q}{q}-\log x\right)\quad\text{and}\quad c_4:=\sum_{q}\frac{(2q-1)\log q}{q(q-1)^2},$$

then this is equal to (by the prime number theorem with error term)

(4) $$y\log(y\log y)+c_3y+O(ye^{-\sqrt{\log y}})+c_4y+O(\log^3 y)$$

$$=y\log y+y\log\log y+(c_3+c_4)y+O(ye^{-\sqrt{\log y}}).$$

In order to apply the Turán–Kubilius inequality, we must also estimate the quantity

$$\sum_{p^k\le x}\frac{h(p^k)^2}{p^k}=\sum_{p\le x}\frac{h(p)^2}{p}+O(1).$$

We have

$$\sum_{p\le x}\frac{h(p)^2}{p}=\sum_{p\le x}\frac{1}{p}\Big(\sum_{q\le y\log y}v_q(p-1)\log q\Big)^2$$

$$=\sum_{p\le x}\frac{1}{p}\sum_{q_1,q_2\le y\log y}v_{q_1}(p-1)v_{q_2}(p-1)\log q_1\log q_2$$

$$=\sum_{q_1,q_2\le y\log y}\log q_1\log q_2\sum_{i,j=1}^{\infty}\sum_{\substack{p\le x,\,p\equiv 1(q_1^i)\\ p\equiv 1(q_2^j)}}1/p:=H_1+H_2$$

say, where in H_1 we have $q_1=q_2$, and in H_2 we have $q_1\ne q_2$.

For H_1 we have

$$H_1 \leqslant 2 \sum_{q \leqslant y \log y} \log^2 q \sum_{i \geqslant j \geqslant 1} \sum_{\substack{p \leqslant x \\ p \equiv 1 (q^i)}} \frac{1}{p}$$

$$= 2 \sum_{q \leqslant y \log y} \log^2 q \sum_{i \geqslant j \geqslant 1} \left(\frac{y}{\phi(q^i)} + O\left(\frac{\log(q^i)}{q^i} \right) \right)$$

$$\ll y \sum_{q \leqslant y \log y} \sum_{i=1}^{\infty} \frac{i \log^2 q}{\phi(q^i)} + \sum_{q \leqslant y \log y} \sum_{i=1}^{\infty} \frac{i^2 \log^3 q}{q^i}$$

$$\ll y \sum_{q \leqslant y \log y} \frac{\log^2 q}{q} + \sum_{q \leqslant y \log y} \frac{\log^3 q}{q} \ll y \log^2 y.$$

Also

$$H_2 = 2 \sum_{q_1 < q_2 \leqslant y \log y} \log q_1 \log q_2 \sum_{i,j=1}^{\infty} \sum_{\substack{p \leqslant x \\ p \equiv 1 (q_1^i q_2^j)}} \frac{1}{p}$$

$$= 2 \sum_{q_1 < q_2 \leqslant y \log y} \log q_1 \log q_2 \sum_{i,j=1}^{\infty} \left(\frac{y}{\phi(q_1^i q_2^j)} + O\left(\frac{\log(q_1^i q_2^j)}{q_1^i q_2^j} \right) \right)$$

$$\leqslant y \left(\sum_{q \leqslant y \log y} \sum_{i=1}^{\infty} \frac{\log q}{\phi(q^i)} \right)^2 + O\left(\left(\sum_{q \leqslant y \log y} \sum_{i=1}^{\infty} \frac{i \log q}{q^i} \right)^2 \right)$$

$$\ll y \left(\sum_{q \leqslant y \log y} \frac{\log q}{q} \right)^2 + \left(\sum_{q \leqslant y \log y} \frac{\log q}{q} \right)^2 \ll y \log^2 y.$$

Now we can apply the Turán–Kubilius inequality, and conclude that

$$\sum_{n \leqslant x} \left(h(n) - \sum_{p^k \leqslant x} \frac{h(p^k)}{p^k} \left(1 - \frac{1}{p} \right) \right)^2 \ll xy \log^2 y,$$

where

$$\sum_{p^k \leqslant x} \frac{h(p^k)}{p^k} \left(1 - \frac{1}{p} \right)$$

is given by (4). Therefore, the number of $n \leqslant x$ for which

(5) $$|h(n) - y \log y - y \log\log y - (c_3 + c_4) y| < y / \log y$$

fails is $o(x)$. We may therefore assume that (5) holds.

We must estimate the contribution to $\log\lambda(n)$ from primes q in I_1 and I_2. First we show that for all but $o(x)$ choices of $n\leqslant x$ we have

$$\sum_{\substack{q^\alpha>y^2/\log^2 y\\ \alpha>1,\, q^\alpha\|\lambda(n)}}\log q^\alpha<\log^2 y. \tag{6}$$

The average value of this quantity is found by summing:

$$\frac{1}{x}\sum_{n\leqslant x}\sum_{\substack{q^\alpha>y^2/\log^2 y\\ \alpha>1,\, q^\alpha\|\lambda(n)}}\log q^\alpha\leqslant\frac{1}{x}\sum_{\substack{\alpha>1\\ q^\alpha>y^2/\log^2 y}}(\log q^\alpha)\left(\frac{x}{q^{\alpha+1}}+\sum_{\substack{p\leqslant x\\ p\equiv 1(q^\alpha)}}\frac{x}{p}\right)$$

$$\leqslant\sum_{\substack{q^\alpha>y^2/\log^2 y\\ \alpha>1}}(\log q^\alpha)\left(\frac{y}{\phi(q^\alpha)}+O\left(\frac{\log q^\alpha}{q^\alpha}\right)\right)\ll\log y,$$

so the number of $n\leqslant x$ for which (6) fails is $O(x/\log y)=o(x)$.

Then by (6), the contribution to $\log\lambda(n)$ from the primes in I_1 is

$$\sum_{q\leqslant y/\log y}v_q(\lambda(n))\log q\ll\sum_{q\leqslant y/\log y}\log y+\log^2 y\ll y/\log y. \tag{7}$$

We now turn to the most subtle part of the argument, namely the estimation of the contribution to $\log\lambda(n)$ from primes in I_2. Let $P(q)$ denote the set of primes $p\leqslant x$ with $p\equiv 1(q)$. Also define

$P_1(q):=\{p\in P(q)\colon\ p\leqslant x^{1/y}$ and for all $q'\in I_2,\ p\not\equiv 1(qq')\}$,

$P_2(q):=\{p\in P(q)\colon\ p\equiv 1(qq')$ for some $q'\in I_2\}$,

$P_3(q):=\{p\in P(q)\colon\ x^{1/y}<p\leqslant x$ and $p\not\equiv 1(qq')$ for all $q'\in I_2\}$.

Then $P(q)$ is the union of these disjoint sets: $P(q)=P_1(q)\cup P_2(q)\cup P_3(q)$.

For $n\leqslant x$, we see from (6) that $\sum_{q\in I_2}v_q(\lambda(n))\log q$, the contribution to $\log\lambda(n)$ from all $q\in I_2$, is given by

$$\sum_{\substack{q\in I_2\\ \omega(n,P_1(q))>0}}\log q+O\Big(\sum_{q\in I_2}\sum_{\substack{p|n\\ p\in P_2(q)}}\log q\Big)+O\Big(\sum_{q\in I_2}\sum_{\substack{p|n\\ p\in P_3(q)}}\log q\Big)+O(\log^2 y). \tag{8}$$

We show that normally the contributions from $p\in P_2(q)$ and from $p\in P_3(q)$ are negligible by averaging. The average contribution from $p\in P_2(q)$ is

$$\frac{1}{x}\sum_{n\leqslant x}\sum_{q\in I_2}\sum_{p|n,\,p\in P_2(q)}\log q\leqslant\sum_{q\in I_2}\log q\sum_{q'\in I_2}\sum_{p\leqslant x,\,p\equiv 1(qq')}\frac{1}{p}$$

$$=\sum_{q\in I_2}\log q\sum_{q'\in I_2}\left(\frac{y}{\phi(qq')}+O\left(\frac{\log(qq')}{qq'}\right)\right)$$

$$\ll y\log y\Big(\sum_{q\in I_2}\frac{1}{q}\Big)^2+\log^2 y\Big(\sum_{q\in I_2}\frac{1}{q}\Big)^2\ll\frac{y(\log\log y)^2}{\log y}.$$

Thus the number of $n \leqslant x$ for which

$$\sum_{q\in I_2} \sum_{\substack{p|n \\ p\in P_2(q)}} \log q < y(\log\log y)^3/\log y \tag{9}$$

fails is $O(x/\log\log y) = o(x)$. We may therefore assume that (9) holds.

We now consider the contribution to $\log\lambda(n)$ from $q\in I_2$ and $p\in P_3(q)$. Since the normal number of prime factors of $n \leqslant x$ that are larger than $x^{1/y}$ is $\log y$, we may assume that the numbers n that we are looking at have fewer than $2\log y$ prime factors larger than $x^{1/y}$. For these n,

$$\sum_{q\in I_2} \sum_{\substack{p|n \\ p\in P_3(q)}} \log q \ll \log^2 y. \tag{10}$$

Finally, we consider the contribution to $\log\lambda(n)$ from $q\in I_2$ and $p\in P_1(q)$. We are concerned with the expected number of $q\in I_2$ for which n is divisible by a prime $p\in P_1(q)$. Towards this end, we estimate the number that do *not* have this property. Let

$$g(n) := \sum_{\substack{q\in I_2 \\ \omega(n,P_1(q))=0}} 1.$$

We would like to apply the Turán–Kubilius inequality to $g(n)$. But it is not an additive function, nor does it resemble an additive function. Nevertheless, we can still establish a normal order for the function $g(n)$. To do this, we shall establish asymptotic formulas for the average value of $g(n)$ and $g(n)^2$. We have

$$\sum_{n\leqslant x} g(n) = \sum_{q\in I_2} \sum_{\substack{n\leqslant x \\ \omega(n,P_1(q))=0}} 1 = \sum_{q\in I_2} \left\{ x \prod_{p\in P_1(q)} \left(1-\frac{1}{p}\right) + O\left(\frac{x}{\log^2 x}\right)\right\} \tag{11}$$

by the fundamental lemma of Brun's sieve [7, Theorem 2.5]. To estimate the product in (11) we need to estimate

$$\begin{aligned}\sum_{p\in P_1(q)} \frac{1}{p} &= \sum_{\substack{p\leqslant x^{1/y} \\ p\equiv 1(q)}} \frac{1}{p} - \sum_{\substack{p\leqslant x^{1/y} \\ p\in P_2(q)}} \frac{1}{p} \\ &= \frac{y-\log y}{q-1} + O\left(\frac{\log q}{q}\right) + O\left(\sum_{q'\in I_2} \sum_{\substack{p\leqslant x \\ p\equiv 1(qq')}} \frac{1}{p}\right) \\ &= \frac{y}{q} + O\left(\frac{\log y}{q}\right) + O\left(\sum_{q'\in I_2} \frac{y}{qq'}\right) = \frac{y}{q} + O\left(\frac{y\log\log y}{q\log y}\right).\end{aligned}$$

Therefore, from (11) we have

$$\sum_{n\leqslant x} g(n) = x \sum_{q\in I_2} \exp\left\{\frac{-y}{q} + O\left(\frac{y\log\log y}{q\log y}\right)\right\} + O\left(\frac{x}{\log x}\right). \tag{12}$$

For $y/\log y < q \leqslant y/(2\log\log y)$ and all large x we have

$$\exp\left\{\frac{-y}{q}+O\left(\frac{y\log\log y}{q\log y}\right)\right\} \ll \frac{1}{\log^2 y}, \tag{13}$$

so that the contribution to (12) from the values of $q \leqslant y/(2\log\log y)$ i $O(xy/\log^2 y)$.

For $q > y/(2\log\log y)$,

$$\exp\left\{O\left(\frac{y\log\log y}{q\log y}\right)\right\} = 1+O\left(\frac{y\log\log y}{q\log y}\right).$$

Together with (12) and (13), this implies that

$$\sum_{n\leqslant x} g(n) = x\sum_{q\in I_2} \exp\left\{\frac{-y}{q}\right\}\left(1+O\left(\frac{y\log\log y}{q\log y}\right)\right)+O\left(\frac{xy}{\log^2 y}\right).$$

Thus, using $0<\exp\{-y/q\}<1$, we have

$$\sum_{n\leqslant x} g(n) = x\sum_{q\in I_2} \exp\left\{\frac{-y}{q}\right\}+O\left(\frac{xy(\log\log y)^2}{\log^2 y}\right). \tag{14}$$

We shall save the estimation of the last sum until later.

First we estimate

$$\sum_{n\leqslant x} g(n)^2 = \sum_{n\leqslant x} \sum_{\substack{q_1,q_2\in I_2\\ \omega(n,P_1(q_i))=0,\, i=1,2}} 1$$

$$= \sum_{n\leqslant x} g(n)+2\sum_{\substack{q_1,q_2\in I_2\\ q_1\neq q_2}} \sum_{\substack{n\leqslant x\\ \omega(n,P_1(q_i))=0,\, i=1,2}} 1.$$

By the fundamental lemma of Brun's sieve, this is

$$= \sum_{n\leqslant x} g(n)+2\sum_{\substack{q_1,q_2\in I_2\\ q_1\neq q_2}} x \prod_{p\in P_1(q_1)\cup P_1(q_2)} \left(1-\frac{1}{p}\right)+O\left(\frac{x}{\log x}\right).$$

Since $P_1(q_1)$ and $P_1(q_2)$ are disjoint for $q_1\neq q_2$, this is equal to

$$\sum_{n\leqslant x} g(n)+x\left(\sum_{q\in I_2}\prod_{p\in P_1(q)}\left(1-\frac{1}{p}\right)\right)^2-x\sum_{q\in I_2}\prod_{p\in P_1(q)}\left(1-\frac{1}{p}\right)^2+O\left(\frac{x}{\log x}\right) \tag{15}$$

$$= (1/x)\Big(\sum_{n\leqslant x} g(n)\Big)^2+O(xy),$$

using (11) and the observation that $g(n)\ll y$ for all n.

It remains to estimate the sum in (14). We have

$$(16)\qquad \sum_{q\in I_2} \exp\left\{\frac{-y}{q}\right\}$$

$$= e^{-1/\log y}(\pi(y\log y)-\pi(y/\log y)) - \int_{y/\log y}^{y\log y} e^{-y/t}\frac{y}{t^2}\left(\pi(t)-\pi\left(\frac{y}{\log y}\right)\right)dt.$$

But note that

$$e^{-1/\log y}(\pi(y\log y)-\pi(y/\log y)) = y - \frac{y\log\log y}{\log y} + O\left(\frac{y(\log\log y)^2}{\log^2 y}\right).$$

In addition,

$$\int_{y/\log y}^{y\log y} e^{-y/t}\frac{y}{t^2}\left(\pi(t)-\pi\left(\frac{y}{\log y}\right)\right)dt$$

$$= \int_{y/\log y}^{y\log y} e^{-y/t}\frac{y}{t^2}\left(\frac{t}{\log t}+O\left(\frac{t}{\log^2 t}\right)\right)dt - \pi\left(\frac{y}{\log y}\right)(e^{-1/\log y}-e^{-\log y})$$

$$= \int_{y/\log y}^{y\log y} e^{-y/t}\frac{y}{t}\left(\frac{1}{\log y}+O\left(\frac{\log\log y}{\log^2 y}\right)\right)dt + O\left(\frac{y}{\log^2 y}\right)$$

$$= \int_{1/\log y}^{\log y} e^{-1/u}\frac{y}{u\log y}du + O\left(\frac{y(\log\log y)^2}{\log^2 y}\right)$$

$$= \frac{y}{\log y}(e^{-1/\log y}\log\log y + e^{-\log y}\log\log y)$$

$$- \int_{1/\log y}^{\log y} e^{-1/u}\frac{y\log u}{u^2\log y}du + O\left(\frac{y(\log\log y)^2}{\log^2 y}\right)$$

$$= \frac{y\log\log y}{\log y} - \frac{y}{\log y}\int_0^\infty e^{-1/u}\frac{\log u}{u^2}du + O\left(\frac{y(\log\log y)^2}{\log^2 y}\right).$$

We therefore have

$$(17)\qquad \sum_{q\in I_2} \exp\left\{\frac{-y}{q}\right\} = y - \frac{2y\log\log y}{\log y} + \frac{c_5 y}{\log y} + O\left(\frac{y(\log\log y)^2}{\log^2 y}\right)$$

where

$$c_5 = \int_0^\infty e^{-1/u}\frac{\log u}{u^2}du = -\int_0^\infty e^{-v}\log v\,dv = \gamma, \text{ Euler's constant.}$$

From (15) we get

$$\sum_{n\leqslant x}\left(g(n)-\frac{1}{x}\sum_{m\leqslant x}g(m)\right)^2=O(xy),$$

so that from (14) and (17), the number of $n\leqslant x$ for which

$$\left|g(n)-\left(y-\frac{2y\log\log y}{\log y}+\frac{c_5 y}{\log y}\right)\right|<\frac{y(\log\log y)^3}{\log^2 y} \tag{18}$$

fails is

$$O\left(\frac{x\log^4 y}{y(\log\log y)^6}\right)=o(x).$$

Thus we may assume that (18) holds.

Note that

$$\pi(y\log y)-\pi(y/\log y)=y-\frac{y\log\log y}{\log y}+\frac{y}{\log y}+O\left(\frac{y(\log\log y)^2}{\log^2 y}\right).$$

Note also that, for $q\in I_2$, we have

$$\log q=\log y+O(\log\log y).$$

Hence, by (8), (9), (10), and (18), we have for all but $o(x)$ choices of $n\leqslant x$

$$\sum_{q\in I_2}v_q(\lambda(n))\log q=\sum_{\substack{q\in I_2\\ \omega(n,P_1(q))>0}}\log q+O\left(\frac{y(\log\log y)^3}{\log y}\right) \tag{19}$$

$$=(\log y+O(\log\log y))\sum_{\substack{q\in I_2\\ \omega(n,P_1(q))>0}}1+O\left(\frac{y(\log\log y)^3}{\log y}\right)$$

$$=(\log y+O(\log\log y))\Big(\pi(y\log y)-\pi(y/\log y)-\sum_{\substack{q\in I_2\\ \omega(n,P_1(q))=0}}1\Big)$$

$$+O\left(\frac{y(\log\log y)^3}{\log y}\right)$$

$$=(\log y+O(\log\log y))\left(\frac{y\log\log y}{\log y}\right.$$

$$\left.+(1-c_5)\frac{y}{\log y}+O\left(\frac{y(\log\log y)^3}{\log^2 y}\right)\right)$$

$$=y\log\log y+(1-c_5)y+O\left(\frac{y(\log\log y)^3}{\log y}\right).$$

We now turn our attention to the range I_3. Since we may assume that $q^2 \nmid n$ for $q \in I_3$, we have by (6)

$$(20)\quad -\log^2 y + \sum_{q\in I_3} \big(v_q(\phi(n)) - v_q(\lambda(n))\big)\log q \leqslant \sum_{\substack{q\in I_3\\ v_q(\lambda(n))=1}} \big(v_q(\phi(n))-1\big)\log q$$

$$\leqslant \sum_{\substack{q\in I_3\\ \omega(n,P(q))>1}} \omega(n, P(q))\log q \overset{\text{def}}{=} G(n).$$

We now compute the average value of $G(n)$. We have

$$\sum_{n\leqslant x} G(n) = \sum_{q\in I_3} \log q \sum_{i\geqslant 2} i \sum_{\substack{n\leqslant x\\ \omega(n,P(q))=i}} 1$$

$$\leqslant \sum_{q\in I_3} \log q \sum_{i\geqslant 2} i \sum_{p_1<\ldots<p_i\in P(q)} \frac{x}{p_1\cdots p_i} \leqslant \sum_{q\in I_3} \log q \sum_{i\geqslant 2} \frac{x}{(i-1)!}\Big(\sum_{p\in P(q)} \frac{1}{p}\Big)^i$$

$$\leqslant \sum_{q\in I_3} \log q \sum_{i\geqslant 2} \frac{x}{(i-1)!}\Big(\frac{y}{q-1} + O\Big(\frac{\log q}{q}\Big)\Big)^i \ll \sum_{q\in I_3} \frac{xy^2 \log q}{q^2} \ll \frac{xy}{\log y}.$$

Therefore the number of $n \leqslant x$ for which

$$(21)\qquad G(n) < y \log\log y/\log y$$

fails is $O(x/\log\log y) = o(x)$. We thus may assume that (21) holds.

Finally, we turn our attention to the range I_4. It is easy to see that, for all but $o(x)$ values of $n \leqslant x$, we have

$$(22)\qquad \sum_{q>y^2} \big(v_q(\phi(n)) - v_q(\lambda(n))\big)\log q = 0.$$

Indeed, the number of $n \leqslant x$ divisible by some q^2, or by two primes in $P(q)$, with $q > y^2$ is

$$\ll \sum_{q>y^2} \frac{x}{q^2} + x \sum_{q>y^2} \Big(\frac{y}{q-1} + O\Big(\frac{\log q}{q}\Big)\Big)^2 \ll \frac{x}{\log y} = o(x).$$

We now assemble all of our results. From (5), (7), (19), (20), (21), and (22), we have

$$\log \phi(n) - \log \lambda(n)$$

$$= y\log y + y\log\log y + (c_3+c_4)y - y\log\log y + (c_5-1)y + O\Big(\frac{y(\log\log y)^3}{\log y}\Big)$$

$$= y\log y + (c_3+c_4+c_5-1)y + O\Big(\frac{y(\log\log y)^3}{\log y}\Big)$$

for all but $o(x)$ choices of $n \leqslant x$.

Finally, we evaluate the constant $A \overset{\text{def}}{=} c_3+c_4+c_5-1$. From [16] we have

$$c_3 = -\gamma - \sum_p \sum_{n\geqslant 2} \frac{\log p}{p^n} = -\gamma - \sum_p \frac{\log p}{p(p-1)}.$$

Hence

$$A = -1-\sum_p \frac{\log p}{p(p-1)}+\sum_p \frac{(2p-1)\log p}{p(p-1)^2}$$
$$= -1+\sum_p \frac{\log p}{(p-1)^2} = -1+\sum_{k=1}^{\infty} k\sum_p \frac{\log p}{p^{k+1}}.$$

Then, with the help of the numerical approximations in [16], it is straightforward to compute that $A = .2269688\ldots$ ∎

It is worth mentioning that, as an immediate consequence of (22), we have the following:

COROLLARY. *The largest prime factor of* $\phi(n)/\lambda(n)$ *is less than* $(\log\log n)^2$ *for all* n *in a set of asymptotic density* 1.

4. Average order. In this section, we estimate the average order

$$F(x):=\frac{1}{x}\sum_{n\leqslant x}\lambda(n).$$

It turns out that most of the contribution to $F(x)$ comes from integers which are atypical in the sense that they have only $\Theta(y/\log y)$ prime divisors. Even if we restrict our attention to integers with $\Theta(y/\log y)$ prime factors, most of the contribution is from a small exceptional set on which λ is large.

Before embarking on the proof, let us first fix some notation. Let $\pi'(x)$ denote the number of primes and powers of primes up to x. Let $S_1, S_2, \ldots, S_D$ be disjoint sets whose union is the set of odd primes less than or equal to x. Define

$$E_i:=\sum_{\substack{p^\alpha\leqslant x\\ p\in S_i}} 1/p^\alpha.$$

For us, $\boldsymbol{j}$ is a vector $(j_1, j_2, \ldots, j_D)$ with each j_i a non-negative integer, and $\|\boldsymbol{j}\|:=j_1+j_2+\ldots+j_D$. Finally let $C(x,\boldsymbol{j})$ be the set of integers $\leqslant x$ with exactly j_i distinct prime divisors in S_i. The following proposition is of independent interest:

PROPOSITION. *There is an absolute constant* $c>0$ *such that, for any* z *with* $1<z<x$, *and all vectors* $\boldsymbol{j}\neq\boldsymbol{0}$ *as defined above, we have*

$$\#C(x,\boldsymbol{j})\leqslant \Psi(x, z)+\frac{cx}{\log z}\left(\prod_{i=1}^{D}\frac{E_i^{j_i}}{j_i!}\right)\left(\sum_{i=1}^{D}\frac{j_i}{E_i}\right)$$

where $\Psi(x, z)$ *is the number of integers* $\leqslant x$ *whose prime factors are all* $\leqslant z$. *(If* S_i *is empty, then* $0/E:=0$ *and* $0^0:=1$.)

Proof. Suppose $n\in C(x,\boldsymbol{j})$ and n has a prime factor $p>z$. Say $p\in S_{i_0}$. Then $n=mp^\alpha$ for some m, $\alpha\geqslant 1$ with $p\nmid m$ and $m\in C(x/z, \boldsymbol{j}-\boldsymbol{e}_{i_0})$. For each

$m \in C(x/z, \boldsymbol{j}-\boldsymbol{e}_{i_0})$, the number of $p^\alpha \leqslant x/m$ with $p \in S_{i_0}$ is at most (for some absolute positive constant c)

$$\pi'\left(\frac{x}{m}\right) < \frac{cx}{m\log(x/m)} \leqslant \frac{cx}{m\log z}.$$

But clearly

$$\sum_{m\in C(x/z,\boldsymbol{j}-\boldsymbol{e}_{i_0})} \frac{1}{m} < \left(\prod_{k=1}^{D} \frac{E_k^{j_k}}{j_k!}\right)\left(\frac{j_{i_0}}{E_{i_0}}\right).$$

Putting these two bounds together and summing over all choices of i_0 gives the result. ∎

COROLLARY. *There is an absolute positive constant $c > 0$ such that for all $x > e^e$ and all vectors $\boldsymbol{j}$ as defined above, we have*

$$\# C(x, \boldsymbol{j}) \leqslant \frac{cx}{(\log x)^{\log y}} + \frac{cxy}{\log x}\left(\prod_{i=1}^{D} \frac{E_i^{j_i}}{j_i!}\right)\left(\sum_{i=1}^{D} \frac{j_i}{E_i}\right).$$

Proof. Note that $C(x, \boldsymbol{0})$ is the set of powers of 2 up to x, so the corollary is true for $\boldsymbol{j} = \boldsymbol{0}$. For $\boldsymbol{j} \neq \boldsymbol{0}$, take $z = x^{1/y}$, and apply well-known estimates of de Bruijn [2] for $\Psi(x, z)$. (Recall that $y = \log\log x$.) ∎

Now we shall specialize; that is, we make a particular choice for the "partition" $S_1, S_2, \ldots, S_D$. Let $m = [y/\log^3 y]$, and let $D = m!$. From now on, we define $S_k := \{p \leqslant x\colon \text{g.c.d.}\ (p-1, D) = 2k\}$. With this particular choice of a partition, we can estimate the E_i's that appear in the proposition.

LEMMA 1. *For $k \leqslant \log^2 y$ we have the uniform asymptotic estimate*

$$E_k = \frac{y}{\log y}\cdot P_k \cdot (1+o(1)),$$

where

$$P_k = \frac{e^{-\gamma}}{k} \prod_{q>2}\left(1-\frac{1}{(q-1)^2}\right) \prod_{q|2k,\, q>2} \frac{q-1}{q-2}.$$

There is also a constant $c_6 > 0$ such that, for all $2k|D$, $E_k > 1/D^{c_6}$.

Proof. Let $k \leqslant \log^2 y$ and let $s_k(t) = \#\{p \leqslant t\colon \text{g.c.d.}(p-1, D) = 2k\}$. First we shall use the fundamental lemma of Brun's sieve to estimate $s_k(t)$. Let $\xi := (\log x)^{7\log y}$, and for $t > \xi$, let

$$A = A(t) := \{(p-1)/2k\colon\ p \leqslant t\ \&\ p \equiv 1(2k)\}.$$

Let

$$\mathfrak{p} = \{q\colon\ q \text{ divides } D/2k\}.$$

Finally, let

$$\omega(q) = \begin{cases} q/(q-1) & \text{if } v_q(2k) = 0 < v_q(D), \\ 1 & \text{if } 0 < v_q(2k) < v_q(D), \\ 0 & \text{else.} \end{cases}$$

The restriction $t > \xi$ is more than enough to ensure that the conditions of Theorem 2.5′ of [7] are satisfied. Hence

$$s_k(t) = S(A, \mathfrak{p}, y) = \left(\frac{\mathrm{li}(t)}{\phi(2k)} \prod_{q|(D/2k)} \left(1 - \frac{\omega(q)}{q}\right)\right)(1 + o(1)),$$

where the function implicit in the $o(1)$ can be chosen uniformly with respect to k. But

$$\begin{aligned}\frac{\mathrm{li}(t)}{\phi(2k)} \prod_{q|(D/2k)} \left(1 - \frac{\omega(q)}{q}\right) &= \frac{\mathrm{li}(t)}{\prod_{q|2k} q^{v_q(2k)-1}(q-1)} \prod_{\substack{q|D \\ q\nmid 2k}} \left(1 - \frac{1}{q-1}\right) \prod_{\substack{q|(D/2k) \\ q|2k}} \left(1 - \frac{1}{q}\right) \\ &= \frac{\mathrm{li}(t)}{2k} \prod_{q|2k} \frac{q}{q-1} \prod_{\substack{q|D \\ q\nmid 2k}} \left(1 - \frac{1}{q-1}\right) \prod_{\substack{q|(D/2k) \\ q|2k}} \left(1 - \frac{1}{q}\right) \\ &= \frac{\mathrm{li}(t)}{2k} \prod_{\substack{q|D \\ q\nmid 2k}} \left(1 - \frac{1}{q-1}\right) \prod_{\substack{q|2k \\ q\nmid (D/2k)}} \frac{q}{q-1} \\ &= \frac{\mathrm{li}(t)}{2k} \prod_{\substack{q|D \\ q>2}} \left(1 - \frac{1}{q}\right) \prod_{\substack{q|D \\ q>2}} \left(1 - \frac{1}{(q-1)^2}\right) \prod_{\substack{q|2k \\ q>2}} \frac{q-1}{q-2} \prod_{\substack{q|2k \\ q\nmid (D/2k)}} \frac{q}{q-1} \\ &= \frac{\mathrm{li}(t)}{\log y} P_k (1 + o(1)).\end{aligned}$$

In the last step, we have used Mertens' theorem that

$$\prod_{q \le T} \left(1 - \frac{1}{q}\right) \sim \frac{e^{-\gamma}}{\log T}$$

and the fact that

$$\prod_{\substack{q|2k \\ q\nmid (D/2k)}} \frac{q}{q-1} = 1$$

for y large, i.e. the product is empty.

With this estimate for $s_k(t)$, it is easy to estimate E_k: we have

$$E_k = \sum_{p \in S_k, p > \xi} \frac{1}{p} + \sum_{p \in S_k, p \le \xi} \frac{1}{p} + \sum_{p \in S_k, j > 1} \frac{1}{p^j}.$$

The first sum is

$$\frac{s_k(x)}{x} - \frac{s_k(\xi)}{\xi} + \int_\xi^x \frac{s_k(t)}{t^2} dt = o(1) + (1 + o(1)) \frac{P_k}{\log y} \int_\xi^x \frac{\mathrm{li}(t)}{t^2} dt = (1 + o(1)) \frac{y P_k}{\log y}.$$

The second and third sums are at most

$$\sum_{p \le \xi} \frac{1}{p} + \sum_{p, j \ge 1} \frac{1}{p^j} \ll \log y$$

and thus are negligible. This completes the proof of the first part of the lemma.

Now suppose that $2k$ divides D. Let

$$Q = \prod_{\substack{q|(D/2k)\\ q\nmid 2k}} q \quad \text{and} \quad T = \prod_{q|2k} q^{v_q(D)}.$$

By the Chinese remainder theorem, we can choose $\alpha < QT$ so that $\alpha \equiv 2(Q)$ and $\alpha \equiv 2k+1(T)$. By a well known theorem of Linnik [15], there is a prime $p < (QT)^{c_6} \leqslant D^{c_6}$ for which $p \equiv \alpha(QT)$. Evidently, $p \in S_k$. Thus $E_k > 1/D^{c_6}$. ■

With these results available, we can now prove the upper bound. Certainly

$$\frac{1}{x}\sum_{n\leqslant x}\lambda(n) = \frac{1}{x}\sum_{\substack{n\leqslant x\\ \omega(n)<y^2}}\lambda(n) + \frac{1}{x}\sum_{\substack{n\leqslant x\\ \omega(n)\geqslant y^2}}\lambda(n).$$

The second sum is negligible because there are only $O(x/\log^2 x)$ integers $n \leqslant x$ with more than y^2 prime divisors (set $D := 1$ in the corollary to the proposition, or apply the well known inequality of Hardy–Ramanujan). The first sum is equal to

$$S := \frac{1}{x}\sum_{\|j\|<y^2}\sum_{n\in C(x,j)}\lambda(n).$$

(This would be true for any partition $S_1, S_2, \ldots, S_D$, so it is certainly true for the one we have chosen.)

For $n \in C(x, j)$, we have

$$\lambda(n) < \frac{D\phi(n)}{\prod_{k=1}^{D}(2k)^{j_k}} < \frac{Dx}{\prod_{k=1}^{D}(2k)^{j_k}}.$$

Combining this estimate with the corollary to the proposition, we get the upper bound

$$S \leqslant \left(\frac{cxyD}{\log x}\right)\sum_{\|j\|<y^2}\left(\prod_{k=1}^{D}\frac{E_k^{j_k}}{(2k)^{j_k}j_k!}\right)\left(\sum_{i=1}^{D}\frac{j_i}{E_i}\right) + \frac{cxD}{(\log x)^{\log y}}\sum_{\|j\|<y^2}\prod_{k=1}^{D}\frac{1}{(2k)^{j_k}}.$$

To estimate the second term, note that

$$\sum_{\|j\|<y^2}\prod_{k=1}^{D}\frac{1}{(2k)^{j_k}} \leqslant \prod_{k=1}^{D}\sum_{j_k\leqslant y^2}\frac{1}{(2k)^{j_k}} < \prod_{k=1}^{D}\frac{1}{1-(1/2k)} \leqslant 2D.$$

Thus the second term is negligible. For the first, note that for $\|j\| < y^2$, we have by Lemma 1

$$\left(\sum_{i=1}^{D}\frac{j_i}{E_i}\right) < \frac{y^2 D}{D^{c_6}}.$$

Hence, we need only estimate

$$\left(\frac{xy^3D^c}{\log x}\right)\sum_{\|j\|<y^2}\left(\prod_{k=1}^{D}\frac{E_k^{j_k}}{(2k)^{j_k}j_k!}\right).$$

But this is less than

$$\left(\frac{xy^3D^c}{\log x}\right)\exp\left(\sum_{k=1}^{D}\frac{E_k}{2k}\right) = \left(\frac{x}{\log x}\right)\exp\left(\sum_{k=1}^{D}\frac{E_k}{2k} + o\left(\frac{y}{\log y}\right)\right)$$

for our choice of D as $[y/\log^3 y]!$. Now let $l := [\log y]$, and consider the sum in the exponent:

$$\sum_{k=1}^{D} \frac{E_k}{2k} = \sum_{k=1}^{l^2} \frac{E_k}{2k} + \sum_{k=l^2+1}^{D} \frac{E_k}{2k}.$$

First we show that the second sum is negligible. Using the Brun–Titchmarsh inequality, it is easy to verify that $E_k \ll y/\phi(k)$. Hence

$$\sum_{k=l^2+1}^{D} \frac{E_k}{2k} \ll \sum_{k>l^2} \frac{y}{k\phi(k)} \ll \frac{y}{l^2} = o\left(\frac{y}{\log y}\right).$$

By Lemma 1, the first sum $\sum_{k=1}^{l^2} E_k/2k$ is asymptotic to

$$\frac{y}{\log y} e^{-\gamma} \prod_{q>2} \left(1 - \frac{1}{(q-1)^2}\right) \sum_{k=1}^{l^2} \left(\frac{1}{2k^2} \prod_{q|2k,\, q>2} \frac{q-1}{q-2}\right) \sim B \frac{y}{\log y},$$

where

$$B :\stackrel{\text{def}}{=} \frac{e^{-\gamma}}{2} \prod_{q>2} \left(1 - \frac{1}{(q-1)^2}\right) \sum_{k=1}^{\infty} \frac{1}{k^2} \prod_{q|2k,\, q>2} \frac{q-1}{q-2}.$$

Observe that

$$\frac{1}{k^2} \prod_{q|2k,\, q>2} \frac{q-1}{q-2}$$

is multiplicative. Hence our expression for the constant B can be simplified:

$$B = \frac{e^{-\gamma}}{2}\left(1 + \frac{1}{4} + \frac{1}{16} + \dots\right) \prod_{q>2}\left(1 - \frac{1}{(q-1)^2}\right)$$

$$\times \prod_{q>2}\left(1 + \frac{1}{q^2}\frac{q-1}{q-2} + \frac{1}{q^4}\frac{q-1}{q-2} + \dots\right)$$

$$= \frac{2e^{-\gamma}}{3} \prod_{q>2}\left(1 - \frac{1}{(q-1)^2}\right)\left(1 + \frac{1}{(q+1)(q-2)}\right)$$

$$= e^{-\gamma} \prod_{q}\left(1 - \frac{1}{(q-1)^2(q+1)}\right) = .34557\dots$$

We have proved the upper bound in Theorem 3. Before proving the lower bound, we need some notation. Define

$\Omega_1(x; j)$:= the set of integers that can be formed by picking $v = \|j\|$ distinct primes $p_1, p_2, \dots, p_v$ in such a way that

(a) $\forall i,\ p_i < x^{1/y^3}$, and

(b) the first j_1 primes are in S_1, the next j_2 are in S_2, etc.

$\Omega_2(x;\mathbf{j})$ consists of those integers $m=(p_1p_2\dots p_v)\in\Omega_1(x;\mathbf{j})$ with the additional property that g.c.d.$(p_i-1,\ p_j-1)$ divides $D=[y/\log^3 y]!$, $\forall i\neq j$.

$\Omega_3(x;\mathbf{j})$ consists of all integers n of the form $n=mp$ where $m\in\Omega_2(x;\mathbf{j})$ and $p\in S_1$ satisfies $\max(x/2m,\ x^{1/y})<p\leqslant x/m$.

$\Omega_4(x;\mathbf{j})$ consists of all integers $n=(p_1p_2\dots p_v)p$ in $\Omega_3(x;\mathbf{j})$ with the additional property that g.c.d.$(p-1,\ p_i-1)=2$ for all i.

Now we can proceed with the proof of the lower bound. To help make the overall argument clear, we postpone several lemmas until afterwards. Let $l:=[\log y]$, and let J denote the set of $\mathbf{j}$'s with $0\leqslant j_k\leqslant E_k/k$ for $k\leqslant l$, and $j_k=0$ for $k>l$. Evidently,

$$\frac{1}{x}\sum_{n\leqslant x}\lambda(n)\geqslant\frac{1}{x}\sum_{\mathbf{j}\in J}\sum_{n\in\Omega_4(x;\mathbf{j})}\lambda(n).$$

Lemma 2 yields the lower bound (using $j_k=0$ for $k>l$)

$$\frac{1}{x}\sum_{n\leqslant x}\lambda(n)\geqslant\left(\frac{c}{y}\right)\sum_{\mathbf{j}\in J}\prod_{k=1}^{l}(2k)^{-j_k}\sum_{n\in\Omega_4(x;\mathbf{j})}1.$$

To estimate the innermost sum, note that

$$\sum_{n\in\Omega_4(x;\mathbf{j})}1=\sum_{m\in\Omega_2(x;\mathbf{j})}\ \sum_{\{p:\,(mp)\in\Omega_4(x;\mathbf{j})\}}1.$$

By Lemma 3, this is greater than

$$\sum_{m\in\Omega_2(x;\mathbf{j})}\frac{cx}{my\log x}.$$

Of course one must check that the hypothesis $\|\mathbf{j}\|\leqslant y^2$ of Lemma 3 is satisfied. But for $\mathbf{j}\in J$, we have by Lemma 1

$$\|\mathbf{j}\|\leqslant\sum_{k\leqslant l}\frac{E_k}{k}\ll\frac{y}{\log y}. \tag{23}$$

Thus

$$\frac{1}{x}\sum_{n\leqslant x}\lambda(n)>\left(\frac{cx}{y^2\log x}\right)\sum_{\mathbf{j}\in J}\prod_{k=1}^{l}(2k)^{-j_k}\sum_{m\in\Omega_2(x;\mathbf{j})}\frac{1}{m}.$$

Lemma 4 implies that, for some constant $c'>0$, this is greater than

$$\frac{x}{\log x}\exp\left[\frac{-c'y}{\log y(\log\log y)^2}\right]\sum_{\mathbf{j}\in J}\prod_{k=1}^{l}\frac{E_k^{j_k}}{(2k)^{j_k}j_k!} \tag{24}$$

$$=\frac{x}{\log x}\exp\left[\frac{-c'y}{\log y(\log\log y)^2}\right]\prod_{k=1}^{l}\sum_{j_k=0}^{[E_k/k]}\frac{(E_k/2k)^{j_k}}{j_k!}.$$

Note that

$$\sum_{j=0}^{[2w]} \frac{w^j}{j!} > \frac{e^w}{2} \quad \text{for} \quad w \geqslant 1.$$

Thus the quantity in (24) is greater than

$$\frac{x}{\log x} \exp\left[\frac{-c'y}{\log y (\log\log y)^2}\right] 2^{-l} \exp\left[\sum_{k=1}^{l} \frac{E_k}{2k}\right] = \frac{x}{\log x} \exp\left[\frac{By}{\log y} + o\left(\frac{y}{\log y}\right)\right]. \ \blacksquare$$

Finally, we prove the lemmas that were just used in the lower bound argument.

LEMMA 2. *If $n \in \Omega_4(x; j)$ then*

$$\lambda(n) > \frac{cx}{y} \prod_{k=1}^{D} (2k)^{-j_k}$$

where c is an absolute, positive constant.

Proof. Suppose $n = (p_1 p_2 \dots p_v) p \in \Omega_4(x; j)$. Let d_i = g.c.d. $(p_i - 1, D)$, and let $m_i := (p_i - 1)/d_i$. Then

$$\lambda(n) = \text{l.c.m.}\ (p_1 - 1, p_2 - 1, \dots, p_v - 1, p - 1)$$
$$\geqslant (m_1 m_2 \dots m_v) \frac{p-1}{2} = \frac{\phi(n)}{2 \prod_{i=1}^{v} d_i} = \frac{\phi(n)}{2 \prod_{k=1}^{D} (2k)^{j_k}}$$
$$\gg \frac{n}{y \prod_{k=1}^{D} (2k)^{j_k}} \gg \frac{x}{y \prod_{k=1}^{D} (2k)^{j_k}}. \ \blacksquare$$

LEMMA 3. *If $m \in \Omega_2(x; j)$, and $\|j\| < y^2$, then*

$$\#\{p\colon (mp) \in \Omega_4(x; j)\} > cx/(my \log x)$$

where c is an absolute, positive constant.

Proof. In the proof of this lemma, let

$$\{q_1, q_2, \dots, q_s\} = \{q\colon 2 < q \leqslant y\} \cup \{q\colon q > 2,\ q \mid \phi(m)\}.$$

Then

$$\#\{p\colon (mp) \in \Omega_4(x; j)\} \geqslant \#\left\{p \in \left(\frac{x}{2m}, \frac{x}{m}\right]\colon p \equiv 3(4) \text{ and for } i \leqslant s,\ q_i \nmid \frac{(p-1)}{2}\right\}.$$

To estimate this quantity, we use Brun's sieve. Let $\mathfrak{p} := \{q_1, \dots, q_s\} \cup \{2\}$, and let

$$A := \left\{\frac{p-1}{2}\colon p \in \left(\frac{x}{2m}, \frac{x}{m}\right]\right\}.$$

Observe that m is relatively small: $m < (x^{1/y^3})^{y^2} = x^{1/y}$. Then by Theorem 2.5′ of [7], we have

$$S(A, \mathfrak{p}, \max(m, y)) > \frac{cx}{m\log(x/m)} \prod_{i=1}^{s}\left(1-\frac{1}{q_i-1}\right) > \frac{c'x}{m\log(x/m)} \prod_{i=1}^{s}\left(1-\frac{1}{q_i}\right).$$

Note that $s \ll \log x$. Hence the last expression is greater than $c''x/(my\log x)$. ∎

LEMMA 4. *If $j \in J$, then for all sufficiently large x,*

$$\sum_{m\in\Omega_2(x;j)} \frac{1}{m} > \exp\left[\frac{-cy\log\log y}{\log^2 y}\right] \prod_{k=1}^{l} \frac{E_k^{j_k}}{j_k!},$$

where c is a positive, absolute constant.

Proof. Since $j \in J$, we have $j_k = 0$ for $k > l$. Thus

$$\sum_{m\in\Omega_2(x;j)} \frac{1}{m} > \frac{1}{j_1!j_2!\dots j_l!} \sum_{\langle p_i\rangle} \frac{1}{p_1p_2\dots p_v}, \tag{25}$$

where the sum in (25) is over all sequences $\langle p_i\rangle_{i=1}^{v}$ of primes for which $v = \|j\| = j_1+j_2+\dots+j_l$, and

(A) The first j_1 primes $p_1, p_2, \dots, p_{j_1}$ are in S_1, the next j_2 in S_2, etc.,
(B) $\forall i$, p_i-1 has no prime factors in $[y/\log^3 y, y\log^3 y]$,
(C) $\forall i$, $p_i \leqslant x^{1/y^3}$,
(D) $\forall i$, $\omega(p_i-1) < y\log\log y$ and $\omega(p_i-1, [y, y^3]) < \log\log y$,
(E) $\forall i \neq j$, $p_i \neq p_j$,
(F) $\forall i \neq j$, g.c.d.(p_i-1, p_j-1) divides $D = [y/\log^3 y]!$.

Let us examine the rth sum in the v-fold summation on the right side of (25):

$$\sum 1/p_r. \tag{26}$$

Suppose that $p_1, p_2, \dots, p_{r-1}$ have already been specified. In order to satisfy condition (F), p_r-1 must avoid certain primes that may appear in the various p_i-1 for $i < r$. For this lemma, let

$$\{q_1, q_2, \dots, q_t\} = \{q \in [y, y^3]\colon\ q|p_i-1 \text{ for some } i < r\},$$

$$\{q_{t+1}, q_{t+2}, \dots, q_s\} = \{q > y^3\colon\ q|p_i-1 \text{ for some } i < r\}.$$

There is some $k \leqslant l$ such that $p_r \in S_k$; in fact k is such that

$$j_1+j_2+\dots+j_{k-1} < r \leqslant j_1+j_2+\dots+j_k.$$

Let $E'_k = \sum' 1/p$, where the sum is over those $p \in S_k$ for which condition (B) holds. Since

$$\sum_{q\in[y/\log^3 y,\, y\log^3 y]} \frac{1}{q} \sim \frac{6\log\log y}{\log y},$$

it follows from the proof of Lemma 1 (i.e., from the fundamental lemma of the sieve) that

$$E'_k = E_k\left(1+O\left(\frac{\log\log y}{\log y}\right)\right). \tag{27}$$

The sum in (26) is at least

$$E'_k - T_C - T_D - T_E - T_F,$$

where

$$T_C := \sum_{x^{1/y^3} < p \leqslant x} \frac{1}{p},$$

$$T_D := \sum_{\substack{p \leqslant x \\ \omega(p-1) \geqslant y\log\log y}} \frac{1}{p} + \sum_{\substack{p \leqslant x \\ \omega(p-1,[y,y^3]) \geqslant \log\log y}} \frac{1}{p},$$

$$T_E := \sum_{i=1}^{r-1} \frac{1}{p_i}, \quad T_F := \sum_{i=1}^{s} \sum_{\substack{p \leqslant x \\ p \equiv 1(q_i)}} \frac{1}{p}.$$

Indeed, T_C, T_D, T_E, T_F respectively take care of those p for which (C), (D), (E), and (F) fail.

We have $T_C \sim 3\log y$. Further, it is easy to see that T_D is small. Indeed, note that

$$T_D \leqslant \sum_{\substack{m \leqslant x \\ \omega(m) \geqslant y\log\log y}} \frac{1}{m} + \sum_{\substack{q|m \Rightarrow q \in [y,y^3] \\ \omega(m) \geqslant \log\log y}} \sum_{\substack{p \leqslant x \\ p \equiv 1(m)}} \frac{1}{p}$$

$$\ll \sum_{\substack{m \leqslant x \\ \omega(m) \geqslant y\log\log y}} \frac{1}{m} + \sum_{\substack{q|m \Rightarrow q \in [y,y^3] \\ \omega(m) \geqslant \log\log y}} \frac{y}{\phi(m)}$$

$$\ll \sum_{i \geqslant y\log\log y} \frac{1}{i!}\left(\sum_{q^\alpha \leqslant x} \frac{1}{q^\alpha}\right)^i + y \sum_{i \geqslant \log\log y} \frac{1}{i!}\left(\sum_{q \in [y,y^3], \alpha \geqslant 1} \frac{1}{\phi(q^\alpha)}\right)^i$$

$$\leqslant \sum_{i \geqslant y\log\log y} \frac{1}{i!}(c+y)^i + y \sum_{i \geqslant \log\log y} \frac{1}{i!} c^i$$

$$\ll \frac{1}{[y\log\log y]!}(c+y)^{[y\log\log y]} + \frac{y}{[\log\log y]!} c^{[\log\log y]}$$

$$\ll \left(\frac{y}{\log^{10} y}\right).$$

Since $r \leqslant v = \|j\|$, we see from (23) that

$$T_E \leqslant \log\log y + O(1).$$

Since the primes $p_1, p_2, \ldots, p_{r-1}$ already chosen satisfy (D), we see from (23) that

$$t < r\log\log y \leqslant v\log\log y \ll (y\log\log y)/\log y,$$

$$s < ry\log\log y \leqslant vy\log\log y \ll (y^2\log\log y)/\log y.$$

Thus, from (B),

$$T_F \ll y\sum_{i=1}^{t}\frac{1}{q_i} + y\sum_{i=t+1}^{s}\frac{1}{q_i} \leqslant \frac{ty}{y\log^3 y} + \frac{sy}{y^3} \ll \frac{y\log\log y}{\log^4 y}.$$

Combining these estimates, we deduce from Lemma 1 that

$$T_C + T_D + T_E + T_F \ll \frac{y\log\log y}{\log^4 y} = o\left(\frac{E_k}{\log y}\right).$$

Thus the sum in (26) is

$$E_k\left(1 + O\left(\frac{\log\log y}{\log y}\right)\right)$$

and so the lemma follows immediately from (23) and (25). ∎

5. Further problems. There are many questions about Euler's ϕ function that remain interesting when put in terms of the λ function. It has been known since I. J. Schoenberg proved this in the 1920's that $n/\phi(n)$ has a continuous distribution function. That is, $D(u)$, the asymptotic density of the n for which $n/\phi(n) \leqslant u$, exists for every u and is a continuous function of u. In this sense, the correct "measuring stick" for $\phi(n)$ is the function n.

It follows from Theorem 2 that, if $\lambda(n)$ has a "measuring stick", it would be about $n/(\log n)^{\log\log\log n + A}$. However, we suspect that there is no monotone function that stays within a constant factor of $\lambda(n)$ for most n. In fact, the following is probably true: there is a function $\psi(x) \to \infty$ such that if $c > 0$ is arbitrary, if $x \geqslant x_0(c)$, and if $A \subseteq [1, x]$ is any set of integers with $|A| > cx$, then

$$\max_{a,b\in A}\frac{\lambda(a)}{\lambda(b)} \geqslant \psi(x).$$

Although we think we can prove the above statement, it may be a hard problem to find the fastest growing function $\psi(x)$ for which it holds. We suspect that it holds for $\psi(x) = \exp[\sqrt{\log\log x}]$, but it is not clear whether this is close to the best possible.

Let $N(k)$ be the number of solutions to $\lambda(n) = k$. From the proof of Theorem 1, it is possible to show that the maximal order of $N(k)$ is very large.

In fact, we have

(28) $$N(k) > \exp\left[\exp\left[(c_2 - o(1))\log k/\log\log k\right]\right]$$

for infinitely many k. On the other hand,

$$N(k) < \exp\left[\exp\left[(\log 2 + o(1))\log k/\log\log k\right]\right].$$

This contrasts sharply with what is known about $\phi(n)$. The number of solutions to $\phi(n) = k$ is always less than the much smaller bound

$$k\exp\left[(-1 + o(1))\log k \log\log\log k/\log\log k\right].$$

Perhaps this is the best possible, but all we can prove is that there is some $c > 0$ such that the number of solutions to $\phi(n) = k$ is greater than k^c for infinitely many k—see [13] for a history of the problem. It is known that

$$\#\{n\colon\ \phi(n) \leqslant x\} \sim cx,$$

where $c = \zeta(2)\zeta(3)/\zeta(6)$. In contrast, we see from (28) that no such result can hold for $\lambda(n)$. We have

$$\exp\left[\exp\left[\frac{(c_2 - o(1))\log x}{\log\log x}\right]\right] < \#\{n\colon\ \lambda(n) \leqslant x\}$$
$$< \exp\left[\exp\left[\frac{(\log 2 + o(1))\log x}{\log\log x}\right]\right].$$

Let $R_\phi(x) = \#\{m \leqslant x\colon m = \phi(n) \text{ for some } n\}$. It is known (see [10]) that

$$R_\phi(x) = \frac{x}{\log x}\exp\left[(c + o(1))(\log\log\log x)^2\right].$$

What about $R_\lambda(x)$? Since few numbers have a large divisor of the form $p-1$ (see [6]), it is clear that $R_\lambda(x) = o(x)$. In fact, the number of values of λ up to x is at most $x/(\log x)^c$ for some $c > 0$. On the other hand, $R_\lambda(x) \gg x/\log x$ trivially because this is already attained on the primes. Perhaps one can find a constant c_7 for which $R_\lambda(x) = x/(\log x)^{c_7 + o(1)}$. Probably $0 < c_7 < 1$, but we do not know what to suggest for the "correct" value of c_7.

References

[1] L. M. Adleman, C. Pomerance and R. S. Rumely, *On distiguishing prime numbers from composite numbers*, Ann. of Math. 117 (1983), 173–206.

[2] N. G. de Bruijn, *On the number of positive integers $\leqslant x$ and free of prime factors $> y$*, Nederl. Akad. Wetensch. Proc. Ser. A 54 (1951), 50–60.

[3] R. D. Carmichael, *On composite numbers p which satisfy the Fermat congruence $a^{p-1} \equiv 1 \bmod p$*, Amer. Math. Monthly 19 (1912), 22–27.

[4] P. D. T. A. Elliott, *Probabilistic Number Theory, I, II*, Springer, New York–Berlin 1979, 1980.

[5] P. Erdős, *On pseudoprimes and Carmichael numbers*, Publ. Math. Debrecen 4 (1955–6), 201–206.

[6] P. Erdős and S. S. Wagstaff, Jr., *The fractional parts of the Bernoulli numbers*, Illinois J. Math. 24 (1980), 112–120.

[7] H. Halberstam and H.-E. Richert, *Sieve Methods*, Academic Press, London–New York 1974.

[8] G. H. Hardy and S. Ramanujan, *The normal number of prime factors of a number n*, Quart. J. Math. 48 (1917), 76–92.

[9] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford Univ. Press, London 1983.

[10] H. Maier and C. Pomerance, *On the number of distinct values of Euler's ϕ-function*, Acta Arith. 49 (1988), 263–275.

[11] G. Miller, *Riemann's hypothesis and tests for primality*, J. Comput. System Sci. 13 (1976), 300–317.

[12] C. Pomerance, *On the distribution of amicable numbers*, J. Reine Angew. Math. 293/4 (1970), 217–222.

[13] —, *Two methods in elementary analytic number theory*, in *Number Theory and Applications*, R. Mollin (ed.), Kluwer Academic Publishers, Dordrecht 1989, 135–161.

[14] K. Prachar, *Über die Anzahl der Teiler einer natürlichen Zahl, welche die Form $p-1$ haben*, Monatsh. Math. 59 (1955), 91–97.

[15] —, *Primzahlverteilung*, Springer, Berlin–Göttingen–Heidelberg 1957, Chapter X.

[16] J. B. Rosser and L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. 6 (1962), 64–94.

[17] S. Wigert, *Sur l'ordre de grandeur du nombre des diviseurs d'un entier*, Arkiv för Math, Astr. Fys. 3 (18) (1907), 1–9.

MATHEMATICAL INSTITUTE
HUNGARIAN ACADEMY OF SCIENCES
Realtanoda u. 13-15
H-1364 Budapest, Pf. 127
Hungary

DEPARTMENT OF MATHEMATICS
UNVERSITY OF GEORGIA
Athens, Georgia 30602
U.S.A.

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE
DREXEL UNIVERSITY
Philadelphia, Pennsylvania 19104
U.S.A.

Received on 2.5.1990
and in revised form on 18.5.1990 (2038)