Illinois Journal of Mathematics
Volume 58, Number 1, Spring 2014, Pages 125–147
S 0019-2082

SOME ARITHMETIC PROPERTIES OF THE SUM OF PROPER DIVISORS AND THE SUM OF PRIME DIVISORS

PAUL POLLACK

ABSTRACT. For each positive integer n, let $s(n)$ denote the sum of the proper divisors of n. If $s(n) > 0$, put $s_2(n) = s(s(n))$, and define the higher iterates $s_k(n)$ similarly. In 1976, Erdős proved the following theorem: *For each $\delta > 0$ and each integer $K \geq 2$, we have*

$$-\delta < \frac{s_{k+1}(n)}{s_k(n)} - \frac{s(n)}{n}$$

for all $1 \leq k < K$, except for a set of n of asymptotic density zero. He also conjectured that

$$\frac{s_{k+1}(n)}{s_k(n)} - \frac{s(n)}{n} < \delta$$

for all $1 \leq k < K$ and all n outside of a set of density zero. This conjecture has proved rather recalcitrant and is known only when $K = 2$, a 1990 result of Erdős, Granville, Pomerance, and Spiro. We reprove their theorem in quantitative form, by what seems to be a simpler and more transparent argument.

Similar techniques are used to investigate the arithmetic properties of the sum of the distinct prime divisors of n, which we denote by $\beta(n)$. We show that for a randomly chosen n, the integer $\beta(n)$ is squarefree with the same probability as n itself. We also prove the same result with "squarefree" replaced by "abundant".

Finally, we prove that for either of the functions $f(n) = s(n)$ or $f(n) = \beta(n)$, the number of $n \leq x$ for which $f(n)$ is prime is $O(x/\log x)$.

Received November 23, 2013; received in final form April 4, 2014.
2010 *Mathematics Subject Classification.* Primary 11A25. Secondary 11N37, 11N60.

1. Introduction

Let $s(n)=\sum_{d|n,d<n} d$ denote the sum of the proper divisors of n. The function $s(n)$ has been of perennial interest in number theory, beginning with the investigations of the ancients into perfect numbers and amicable pairs. If $s(n)>0$, put $s_2(n)=s(s(n))$, and define the higher iterates $s_k(n)$ similarly. For each natural number n, the *aliquot sequence* at n is $n, s(n), s_2(n), s_3(n), \dots$, which terminates if the sequence ever hits 0. For example, the aliquot sequence at $n=12$ is 12, 16, 15, 9, 4, 3, 1, 0, while the sequence at $n=220$ is 220, 284, 220, 284, ... The following conjecture is a revision due to Dickson [Dic13] of a statement proposed by Catalan [Cat88].

Catalan–Dickson conjecture. *For every positive integer n, the aliquot sequence at n is bounded. In other words, every aliquot sequence either terminates at 0 or is eventually periodic.*

One way to disprove the Catalan–Dickson conjecture would be to produce an n for which the aliquot sequence at n is strictly increasing. It is not clear whether we should expect there to be any such n. However, it was shown by Lenstra (see [Erd76], [Len75]) that for every K, there is an n for which the sequence is increasing for the first K steps, that is,

$$n < s(n) < s\big(s(n)\big) < \cdots < s_K(n). \tag{1.1}$$

If the sequence is increasing in its first step, that is, if $s(n)>n$, then n is said to be *abundant*. It is easy to see that a positive proportion of all numbers are abundant; for instance, all proper multiples of 6 have this property. In fact, the set of abundant numbers possesses an asymptotic density, and this density is slightly less than $\frac{1}{4}$ (see [Kob10], [Kob14] for recent work on the problem of computing this density). The following remarkable sharpening of Lenstra's result was proved by Erdős [Erd76].

Proposition 1.1. *Fix a positive integer K. Then (1.1) holds for all abundant numbers n, except for a set of density zero.*

Erdős deduced Proposition 1.1 from another result, of interest in its own right:

Proposition 1.2. *Fix a positive integer K, and fix a real number $\delta>0$. Then for all n outside of a set of asymptotic density zero, we have*

$$\frac{s_{k+1}(n)}{s_k(n)} - \frac{s(n)}{n} \geq -\delta \quad \textit{for all } 1\leq k<K.$$

In the same paper, Erdős proposed the following conjecture, which is a natural dual of Proposition 1.2. (His initial claims to possess a proof were retracted in [EGPS90].)

CONJECTURE 1.3. *Fix an integer $K \geq 2$. Let $\delta > 0$. Then for all n outside of a set of asymptotic density zero, we have*

$$\frac{s_{k+1}(n)}{s_k(n)} - \frac{s(n)}{n} \leq \delta \quad \textit{for all } 1 \leq k < K.$$

This conjecture seems rather difficult and is known to hold only in the single case $K = 2$, a result of Erdős, Granville, Pomerance, and Spiro [EGPS90, Theorem 5.1]. The primary objective of this article is to present a new proof of their result in a sharper, quantitative form. For $x > 0$, let $\log_1 x = \max\{1, \log x\}$, and let $\log_k$ denote the kth iterate of $\log_1$.

THEOREM 1.4. *Let $x \geq 1$. For all but $O(x(\log_3 x)^2/(\log_2 x)^{1/4})$ positive integers $n \leq x$, we have*

$$\frac{s(s(n))}{s(n)} - \frac{s(n)}{n} \leq (\log_2 x)^{-1/4}.$$

An entirely analogous result, for the $K = 2$ case of Proposition 1.2 rather than Conjecture 1.3, was established by Kobayashi, Pollack, and Pomerance in the proof of [KPP09, Theorem 7].

In addition to yielding a quantitative result, our proof of Theorem 1.4 is simpler than the original argument of Erdős et al. [EGPS90]. For instance, we avoid using any facts about primitive α-abundant numbers. For the most part, we use only results belonging to the standard tool chest of elementary analytic number theory. The key ingredient—which seems of independent interest—is a new upper bound on the number of $n \leq x$ for which $s(n)$ possesses a prescribed divisor. We show (Lemma 2.8) that once one throws away a certain set of density zero, the number of remaining $n \leq x$ where q divides $s(n)$ is $O_\varepsilon(x/q^{1-\varepsilon})$, uniformly in a wide range of q.

In the same way that Proposition 1.1 follows from Proposition 1.2, the theorem of Erdős–Granville–Pomerance–Spiro implies that if $s(n) < n$, then almost always $s(s(n)) < s(n)$. Using Theorem 1.4 and a result of Toulmonde, we obtain a quantitative version of this result.

COROLLARY 1.5. *The number of $n \leq x$ for which $s(n) < n$ but $s(s(n)) \geq s(n)$ is at most*

$$x/\exp\left(\left(\frac{1}{10} + o(1)\right)\sqrt{\log_3 x \log_4 x}\right),$$

as $x \to \infty$.

Corollary 1.5 is an analogue of a theorem of Pomerance, who gave an upper bound of the same shape for the number of $n \leq x$ with $s(n) \geq n$ but $s(s(n)) < s(n)$ (see the proof of [Pom77, Theorem 4]).

Similar techniques can be applied to study the sum of the distinct prime factors of n, denoted $\beta(n)$. This arithmetic function, and its close relative $B(n) := \sum_{p^k \| n} kp$, have been investigated by several authors; see, for example, [Hal70], [Hal71], [Hal72], [AE77], [EP78], [Pom02], [DKL05], [BL07], [DKL08], [Jak12].

To set the stage for our first result about $\beta(n)$, we recall the 1933 theorem of Davenport [Dav33] that $s(n)/n$ has a continuous distribution function.

PROPOSITION 1.6. *For each real u, the set $\mathscr{D}(u) := \{n \in \mathbb{N} : \frac{s(n)}{n} \le u\}$ possesses an asymptotic density $D(u)$. The function $D(u)$ is continuous everywhere and satisfies $D(0) = 0$ and $\lim_{u\to\infty} D(u) = 1$.*

Call $D(u)$ the *Davenport distribution function.* We show that $s(\beta(n))/\beta(n)$ also follows the Davenport distribution.

THEOREM 1.7. *For every real number u,*

$$\lim_{x\to\infty} \frac{1}{x}\#\left\{1 < n \le x : \frac{s(\beta(n))}{\beta(n)} \le u\right\} = D(u).$$

For example, the probability that $\beta(n)$ is abundant for a randomly chosen n is the same as the probability that n itself is abundant (i.e., just under 25 percent). Note that if we replace $\beta(n)$ with $s(n)$ in Theorem 1.7, then the corresponding result follows from the $K = 2$ cases of Proposition 1.2 and Conjecture 1.3.

We also show that the set of n with $\beta(n)$ squarefree has precisely the same density as the set of squarefree numbers n.

THEOREM 1.8. *The set of natural numbers n for which $\beta(n)$ is squarefree has asymptotic density $\frac{6}{\pi^2}$.*

Probably Theorem 1.8 has the following $s(n)$-analogue: *n is squarefree if and only if $s(n)$ is squarefree, up to a density zero set of exceptions.* We do not see how to show this. The reason we succeed in proving Theorem 1.8 is that we have a strong upper bound for the number of $n \le x$ with $\beta(n)$ divisible by q (Lemma 2.15), valid without *any* restriction on the size of q. It would be very interesting to know if such a *completely* uniform result could be established for $s(n)$.

We conclude the paper with some results about prime values of $s(n)$ or $\beta(n)$. Let $\pi_\beta(x)$ denote the number $n \le x$ for which $\beta(n)$ is prime. Of course, $\beta(p) = p$ for every prime p, so a more natural object of study is the difference $\pi_\beta(x) - \pi(x)$ that counts only composite n. For these n, the prime number theorem suggests that $\beta(n)$ is prime with 'probability' $1/\log\beta(n)$. Most of the time, $\beta(n) \approx P(n)$, where $P(n)$ denotes the largest prime factor of n. So we are led to predict that

$$\pi_\beta(x) - \pi(x) \approx \sum_{\text{composite } n\le x} \frac{1}{\log P(n)},$$

and this latter sum is known to be $\sim e^\gamma x/\log x$ as $x \to \infty$; see, for example, [Whe90, Theorem 9]. (That paper estimates the sum without the restriction to composite n; this is unimportant, since prime n contribute only $O(x/(\log x)^2)$.) We state the resulting conjecture formally.

CONJECTURE 1.9. *As* $x \to \infty$,

$$\pi_\beta(x) - \pi(x) \sim e^\gamma \frac{x}{\log x}.$$

Here γ *is the usual Euler–Mascheroni constant.*

One might worry that the reasoning offered for Conjecture 1.9 does not take into account the local distribution of $\beta(n)$. But in fact, $\beta(n)$ is uniformly distributed to small moduli (see Corollary 2.10 below), and this suggests that Conjecture 1.9 does not require any local correction factors. Although Conjecture 1.9 seems difficult, we are able to show an upper bound of the 'correct' order.

THEOREM 1.10. *For all* $x \ge 2$, *we have* $\pi_\beta(x) - \pi(x) \ll x/\log x$.

We can also ask about prime values of $s(n)$. Since $s(n)$ is almost always of the same order of magnitude as n, our heuristic argument now suggests that $s(n)$ is prime for asymptotically $x/\log x$ values of $n \le x$. We prove an upper bound of this order, by arguments similar to but more intricate than those offered for Theorem 1.10.

THEOREM 1.11. *For all* $x \ge 2$, *the number of* $n \le x$ *for which* $s(n)$ *is prime is* $O(x/\log x)$.

Organization. We clear the ground for the proofs of later results in Section 2. Theorems 1.4 and Corollary 1.5 are proved in Section 3. Proofs of Theorems 1.7 and 1.8 appear in Section 4. Finally, the proofs of Theorems 1.10 and 1.11 are given in Section 5.

Notation. We reserve the letter p for a prime variable. Of the two competing notations for the number of divisors of n, we pick $\tau(n)$. We let $P(n)$ denote the largest prime factor of n, with the understanding that $P(1) = 1$. We use O and o-notation, as well as the symbols $\ll$, $\gg$, and $\asymp$, with their usual meanings. Any dependence of implied constants is indicated with subscripts. Other notation will be introduced as necessary.

2. Preliminaries

2.1. General results. We begin our preparation for the proof of Theorem 1.4 with a lemma due to Pomerance.

Lemma 2.1. *Let $x \geq 3$. Let q be a positive integer. The number of $n \leq x$ for which $q \nmid \sigma(n)$ is*

$$\ll x/(\log x)^{1/\varphi(q)},$$

uniformly in q.

Proof. If $q \nmid \sigma(n)$, then there is no prime $p \equiv -1 \pmod{q}$ which appears in the prime factorization of n to only the first power. By [Pom77, Theorem 2], the number of such $n \leq x$ is $O(x/(\log x)^{1/\varphi(q)})$. □

Lemma 2.1 is complemented by the following crude upper bound on the count of n for which $\sigma(n)$ possesses a given *prime* divisor.

Lemma 2.2. *For each prime p, the number of $n \leq x$ for which $p \mid \sigma(n)$ is*

$$\ll \frac{x \log_2 x}{p^{1/2}}.$$

Proof. If $p \mid \sigma(n)$, then $p \mid \sigma(q^e)$ for some prime power q^e exactly dividing n. If $e = 1$, then $q \equiv -1 \pmod{p}$. By Brun–Titchmarsh and partial summation, the number of $n \leq x$ divisible by such a prime q is at most

$$x \sum_{\substack{q \leq x \\ q \equiv -1 \pmod{p}}} \frac{1}{q} \ll \frac{x \log_2 x}{p},$$

which fits within the asserted upper bound. If $e \geq 2$, we notice that $2q^e > \sigma(q^e) \geq p$, and so n has a squarefull divisor exceeding $\frac{p}{2}$. The number of such $n \leq x$ is $O(x/p^{1/2})$, which is again acceptable for us. □

We also need an upper bound on the count of numbers whose prime factors all belong to a specified interval. For each $x \geq 1$ and each closed interval $[z, y]$, set

$$\Psi\big(x, [z, y]\big) := \#\big\{n \leq x : p \mid n \Rightarrow p \in [z, y]\big\}.$$

Lemma 2.3. *For all $x \geq y \geq z \geq 2$, we have*

$$\Psi\big(x, [z, y]\big) \ll \frac{x}{\log z} e^{-u/2}, \quad \textit{where } u := \frac{\log x}{\log y}.$$

The proof of Lemma 2.3 requires the following result of Halberstam and Richert [HR79]. We state the result in a form close to the statement in [Ten95, Corollary 5.1, p. 309].

Lemma 2.4. *Let f be a nonnegative-valued multiplicative function. Suppose that there are positive constants λ_1 and λ_2, with $\lambda_2 < 2$, so that*

$$f\big(p^k\big) \leq \lambda_1 \lambda_2^{k-1}$$

for all prime powers p^k. Then for all $x \geq 1$,

$$\sum_{n \leq x} f(n) \ll_{\lambda_1, \lambda_2} x \exp\bigg(\sum_{p \leq x} \frac{f(p) - 1}{p}\bigg).$$

Proof of Lemma 2.3. We follow closely the proof of [Ten95, Theorem 1, p. 359]. We can suppose that $y \geq 11$, otherwise $\Psi(x,[z,y]) = O((\log x)^4)$, whereas the asserted upper bound is $\gg x^{1/4}$. Let $\chi(n)$ denote the characteristic function of those n all of whose prime factors belong to $[z,y]$. Then for any $\alpha > 0$,

$$\begin{aligned}\Psi\big(x,[z,y]\big) &= \sum_{n\leq x}\chi(n) \leq x^{3/4} + \sum_{x^{3/4}<n\leq x}\chi(n) \\ &\leq x^{3/4} + x^{-3\alpha/4}\sum_{n\leq x}\chi(n)n^{\alpha}.\end{aligned} \tag{2.1}$$

Now let $\alpha = 2/(3\log y)$. We estimate the sum on the right of (2.1) by applying Lemma 2.4 with $f(n) = \chi(n)n^{\alpha}$. Then $f(p^k) = 0$ if $p \notin [z,y]$, while if $p \in [z,y]$, then $f(p^k) = p^{k\alpha} \leq \exp(2k/3)$. This shows that the hypotheses of Lemma 2.4 are satisfied with $\lambda_1 = \lambda_2 = \exp(2/3) = 1.947\ldots$. Hence,

$$\begin{aligned}\sum_{n\leq x}\chi(n)n^{\alpha} &\ll x\exp\bigg(-\sum_{p\leq z}\frac{1}{p}\bigg)\exp\bigg(\sum_{z\leq p\leq y}\frac{p^{\alpha}-1}{p}\bigg) \\ &\ll \frac{x}{\log z}\exp\bigg(\sum_{z\leq p\leq y}\frac{p^{\alpha}-1}{p}\bigg).\end{aligned}$$

For our choice of α, we have $\alpha\log p \ll 1$ for $p \leq y$, and so $p^{\alpha} - 1 \ll \alpha\log p$. Thus, $\sum_{z\leq p\leq y}\frac{p^{\alpha}-1}{p} \ll \alpha\sum_{p\leq y}\frac{\log p}{p} \ll \alpha\log y \ll 1$. We conclude that $\sum_{n\leq x}\chi(n)n^{\alpha} \ll x/\log z$. Inserting this back into (2.1) and remembering the definition of α, we find that

$$\Psi\big(x,[y,z]\big) \ll x^{3/4} + \frac{x}{\log z}e^{-u/2}. \tag{2.2}$$

Since $y \geq 11$ and $x \geq z$, we see that $\frac{x}{\log z}e^{-u/2} \geq \frac{x}{\log x}x^{-1/(2\log 11)} \gg x^{0.79}$, and so the second term in (2.2) dominates. This completes the proof of the lemma. ☐

The case $z = 2$ is of special importance. The integers counted by $\Psi(x,[2,y])$ are called *y-smooth* (or *y-friable*), and $\Psi(x,[2,y])$ is usually abbreviated to $\Psi(x,y)$. While Lemma 2.3 would suffice in our applications, some of the arguments run more smoothly if we allow ourselves the following upper bound estimate of de Bruijn (see [dB66, Theorem 2]).

PROPOSITION 2.5. *Let $x \geq y \geq 2$ satisfy $(\log x)^2 \leq y \leq x$. Whenever $u := \frac{\log x}{\log y} \to \infty$, we have*

$$\Psi(x,y) \leq x/u^{u+o(u)}.$$

2.2. Sums of divisors. For the rest of this paper, x always denotes a real number satisfying $x \geq 3$, and $\mathscr{E}(x)$ denotes the set

$$\mathscr{E}(x) := \left\{n \leq x : P(n) \leq x^{1/\log_3 x} \text{ or } P(n)^2 \mid n\right\}. \tag{2.3}$$

Lemma 2.6. *We have* $\#\mathscr{E}(x) \ll x/(\log_2 x)^4$.

Proof. If $n \in \mathscr{E}(x)$, then either $P(n) \leq x^{1/\log_3 x}$ or $P(n) > x^{1/\log_3 x}$ and $P(n)^2 \mid n$. The number of $n \leq x$ for which the former possibility holds is $O(x/(\log_2 x)^4)$ by Proposition 2.5. The number of $n \leq x$ for which the latter holds is $\ll x \sum_{p > x^{1/\log_3 x}} p^{-2} \ll x \exp(-\log x/\log_3 x)$, and this is also $O(x/(\log_2 x)^4)$. □

Lemma 2.7. *Let q be a natural number with $q \leq x^{\frac{1}{2\log_3 x}}$. The number of $n \leq x$ not belonging to $\mathscr{E}(x)$ for which q divides $s(n)$ is*

$$\ll \frac{\tau(q)}{\varphi(q)} \cdot x \log_3 x.$$

Proof. Since $n \notin \mathscr{E}$, we can write $n = Pm$, where $P := P(n) > x^{1/\log_3 x}$ and $P \nmid m$. This factorization of n induces a factorization $q = q_1 q_2$ of q, where $q_1 := \gcd(q, s(m))$ and $q_2 := q/q_1$. Our strategy is to count the number of $n \leq x$ for which q divides $s(n)$ and where n corresponds to a *fixed* factorization $q = q_1 q_2$. At the end of the proof, we sum over all possible factorizations $q_1 q_2$ of q.

Since q divides $s(Pm) = Ps(m) + \sigma(m)$, we see that $q_1 \mid \sigma(m)$, and so also

$$q_1 \mid \sigma(m) - s(m) = m.$$

Also, $q_2 = q/\gcd(q, s(m))$ is coprime to $s(m)/\gcd(q, s(m)) = s(m)/q_1$, and so the congruence

$$P\frac{s(m)}{q_1} \equiv -\frac{\sigma(m)}{q_1} \pmod{q_2}$$

places P in a uniquely determined residue class modulo q_2. Moreover,

$$P \leq x/m, \quad \text{so that in particular, } x/m > x^{1/\log_3 x}.$$

By Brun–Titchmarsh [HR74, Theorem 3.8, p. 110], the number of choices of P given m is

$$\ll \frac{x/m}{\varphi(q_2)\log \frac{x}{mq_2}} \ll \frac{x \log_3 x}{m\varphi(q_2)\log x},$$

where we have used that $\frac{x}{mq_2} \geq \frac{x}{mq} \geq x^{1/\log_3 x} q^{-1} \geq x^{\frac{1}{2\log_3 x}}$. Summing on $m \leq x^{1-\frac{1}{\log_3 x}}$ that are multiples of q_1 shows that the number of possible values of $n = mP$ is

$$\ll \frac{x \log_3 x}{q_1 \varphi(q_2)}.$$

Finally, we sum over possible factorizations $q = q_1 q_2$. We have that

$$\sum_{q_1 q_2 = q} \frac{1}{q_1 \varphi(q_2)} = \frac{1}{q} \sum_{q_2 \mid q} \frac{q_2}{\varphi(q_2)} \leq \frac{1}{q}\left(\tau(q) \frac{q}{\varphi(q)}\right) = \frac{\tau(q)}{\varphi(q)},$$

using in the second step that $q_2/\varphi(q_2) \leq q/\varphi(q)$ for every q_2 dividing q. Collecting the above estimates completes the proof. □

Lemma 2.8. *Let q be a natural number with $q \leq x^{\frac{1}{2\log_3 x}}$. Let $\varepsilon > 0$. The number of $n \leq x$ not belonging to $\mathscr{E}(x)$ for which q divides $s(n)$ is*

$$\ll_\varepsilon x/q^{1-\varepsilon}.$$

Proof. If $q > \sqrt{\log_2 x}$, this follows from Lemma 2.7 and the known results on the minimal order of φ [HW08, Theorem 327, p. 352] and the maximal order of τ [HW08, Theorem 315, p. 343]. Suppose that $q \leq \sqrt{\log_2 x}$. We can assume that $q \mid \sigma(n)$, since Lemma 2.1 implies that the number of exceptional $n \leq x$ is $\ll x/\exp(\sqrt{\log_2 x}) \ll x/q$. But if $q \mid \sigma(n)$ and $q \mid s(n)$, then $q \mid n$, and the number of these $n \leq x$ is $O(x/q)$. □

2.3. Sums of prime divisors. We require results on the distribution of $\beta(n)$ in arithmetic progressions. Many of these were noted by Hall, who made an extensive study of the frequency with which n and $\beta(n)$ are relatively prime [Hal70], [Hal71], [Hal72].

The following proposition is a weak consequence of a theorem of Halász [Hal68, Satz 2]; for several related results, see [Ell79, Chapter 9].

Proposition 2.9. *Let f be a multiplicative function. Suppose that for a certain positive integer k, one has $f(n)^k = 1$ for all n. If $\sum_{p: f(p) \neq 1} \frac{1}{p}$ diverges, then f has mean value zero.*

Corollary 2.10. *Fix a positive integer q. The values $\beta(n)$ are equidistributed modulo q. More precisely, for each a, the density of n with $\beta(n) \equiv a \pmod{q}$ is $\frac{1}{q}$.*

Proof. By the orthogonality relations for additive characters mod q, the number of $n \leq x$ with $\beta(n) \equiv a \pmod{q}$ is given by

$$\frac{\lfloor x \rfloor}{q} + \frac{1}{q} \sum_{r=1}^{q-1} e^{-2\pi i \frac{ar}{q}} \sum_{n \leq x} e^{2\pi i \frac{r\beta(n)}{q}}.$$

Thus, it suffices to show that whenever $\zeta \neq 1$ is a qth root of unity, the function $f(n) := \zeta^{\beta(n)}$ has mean value zero. Since β is an additive function, f is multiplicative. Moreover, $f(p) \neq 1$ as long as p does not divide q. Since $\sum_p 1/p$ diverges, the corollary follows from Proposition 2.9. □

The next lemma is a special case of [Hal72, Lemma 1].

LEMMA 2.11. *For all $q \le \sqrt{x}$ and all integers a,*

$$\sum_{\substack{n \le x \\ n \text{ squarefree} \\ \beta(n) \equiv a \pmod q}} 1 \ll \frac{x}{\varphi(q)} + x\frac{\log q}{\log x}.$$

Sacrificing some uniformity in q, we can remove the restriction in Lemma 2.11 to squarefree n.

LEMMA 2.12. *For all $q \le x^{1/4}$ and all integers a,*

$$\sum_{\substack{n \le x \\ \beta(n) \equiv a \pmod q}} 1 \ll \frac{x}{\varphi(q)} + x\frac{\log q}{\log x}. \tag{2.4}$$

Proof. Write $n = n_1 n_2$, where n_1 is squarefree, n_2 is squarefull, and n_1 and n_2 are coprime. We can assume that $n_2 \le x^{1/2}$, since the count of $n \le x$ possessing a squarefull divisor exceeding $x^{1/2}$ is $O(x^{3/4})$, which is dominated by the right-hand side of (2.4). Now $\beta(n) = \beta(n_1) + \beta(n_2)$. Moreover, $q \le x^{1/4} \le \sqrt{x/n_2}$, and so the number of $n \le x$ with $\beta(n) \equiv a \pmod q$ corresponding to a given value of n_2 is at most

$$\sum_{\substack{n_1 \le x/n_2 \\ n_1 \text{ squarefree} \\ \beta(n_1) \equiv a - \beta(n_2) \pmod q}} 1 \ll \frac{x/n_2}{\varphi(q)} + \frac{x}{n_2}\frac{\log q}{\log(x/n_2)} \ll \frac{x}{n_2\varphi(q)} + \frac{x}{n_2}\frac{\log q}{\log x}.$$

Summing over squarefull $n_2 \le x^{1/4}$ completes the proof. ☐

The next result is an easy variant of [Hal72, Lemma 1]. We include the short proof. Recall the definition (2.3) of $\mathscr{E} = \mathscr{E}(x)$.

LEMMA 2.13. *Let q be a positive integer. The number of $n \in (1, x]$ not belonging to $\mathscr{E}$ for which q divides $\beta(n)$ is*

$$\ll \frac{x(\log x)^2}{q}.$$

Proof. Suppose that $n \in (1, x]$ and that $n \notin \mathscr{E}(x)$. Write $n = mP$, where $P = P(n)$. Suppose that q divides $\beta(n)$. Then

$$q \le \beta(n) \le P\omega(n) \le 2P\log x \le 2\frac{x}{m}\log x,$$

and so $mq \le 2x\log x$. Since n does not belong to $\mathscr{E}$, we have that $P \nmid m$, and $\beta(n) = P + \beta(m)$. Thus, $P \equiv -\beta(m) \pmod q$. So given m, the number of possibilities for $P \le x/m$ is trivially at most

$$\frac{x}{mq} + 1 = \frac{x}{mq}\left(1 + \frac{mq}{x}\right) \ll \frac{x\log x}{mq}.$$

Summing over $m \le x$ gives the estimate of the lemma. ☐

The following result is an analogue of Lemma 2.7 from the preceding section.

LEMMA 2.14. *Let q be a positive integer with $q \le x^{\frac{1}{2\log_3 x}}$. The number of $n \in (1, x]$ not belonging to $\mathscr{E}$ for which $q \mid \beta(n)$ is*

$$\ll \frac{x \log_3 x}{\varphi(q)}.$$

Proof (sketch). Write $n = mP$, where $P = P(n)$. Since $n \notin \mathscr{E}$, we see that $x/m \ge P > x^{1/\log_3 x}$ and that $P \nmid m$. So if $q \mid \beta(n)$, then $P \equiv -\beta(m) \pmod q$. To complete the proof, we apply Brun–Titchmarsh in the same way as in the proof of Lemma 2.7. □

Assembling the preceding estimates yields the following 'master result'. Note that unlike Lemma 2.8, the following result holds without any restriction on the size of q.

LEMMA 2.15. *Let q be a positive integer. For each $\varepsilon > 0$, the number of $n \in (1, x]$ not belonging to $\mathscr{E}$ for which $q \mid \beta(n)$ is*

$$\ll_\varepsilon x/q^{1-\varepsilon}.$$

Proof. For $q \le \log x$, this follows from Lemma 2.12. For $\log x < q \le \exp(\sqrt{\log x})$, we apply Lemma 2.14. Finally, if $q > \exp(\sqrt{\log x})$, this is a consequence of Lemma 2.13. □

3. Abundancy of $s(n)$ versus $s(s(n))$

3.1. Proof of Theorem 1.4. We start by noting that $\frac{\sigma(m)}{m} = \sum_{d \mid m} \frac{1}{d}$ for every positive integer m. Let $y = x^{1/\log_3 x}$. As long as $n > 1$ (so that $s(n) > 0$), we have

$$\begin{aligned}\frac{s(s(n))}{s(n)} - \frac{s(n)}{n} &= \frac{\sigma(s(n))}{s(n)} - \frac{\sigma(n)}{n} \\ &= \sum_{d \mid s(n)} \frac{1}{d} - \sum_{d \mid n} \frac{1}{d} \le \Sigma_1(n; x) + \Sigma_2(n; x),\end{aligned}$$

where we define

$$\Sigma_1(n; x) := \sum_{\substack{d \mid s(n),\, d \nmid n \\ d \le y^{1/2}}} \frac{1}{d}, \quad \text{and} \quad \Sigma_2(n; x) := \sum_{\substack{d \mid s(n),\, d \nmid n \\ d > y^{1/2}}} \frac{1}{d}.$$

It turns out that $\Sigma_2(n; x)$ is always quite small. To see this, observe that $s(n) = \sum_{d \mid n, d < n} d < n\tau(n) \le n^2 \le x^2$ and that any number not exceeding x^2 has at most

$$\exp\left((\log 2 + o(1)) \frac{\log(x^2)}{\log_2(x^2)} \right)$$

divisors, as $x \to \infty$ (see [HW08, Theorem 317, p. 345]). Thus,

$$\Sigma_2(n;x) \le y^{-1/2}\tau\big(s(n)\big) < y^{-1/3} < \frac{1}{2}(\log_2 x)^{-1/4},$$

once x is sufficiently large.

To complete the proof of Theorem 1.4, it is enough to show that

$$\Sigma_1(n;x) \le \frac{1}{2}(\log_2 x)^{-1/4} \tag{3.1}$$

for all but $O(x(\log_3 x)^2/(\log_2 x)^{1/4})$ values of $n \le x$. In counting exceptions to (3.1), we may assume that $n \notin \mathscr{E}(x)$, since $\#\mathscr{E}(x) = O(x/(\log_2 x)^4)$. We may also assume that $n \notin \mathscr{E}'(x)$, where

$$\mathscr{E}'(x) := \big\{n \le x : \text{there is a } d \le \sqrt{\log_2 x} \text{ with } d \nmid \sigma(n)\big\}.$$

Indeed, from Lemma 2.1,

$$\begin{aligned}\#\mathscr{E}'(x) &\ll \sum_{d \le \sqrt{\log_2 x}} \frac{x}{(\log x)^{1/\varphi(d)}} \le \frac{x}{(\log x)^{1/\sqrt{\log_2 x}}} \sum_{d \le \sqrt{\log_2 x}} 1 \\ &\ll x/\exp\left(\frac{1}{2}(\log_2 x)^{1/2}\right),\end{aligned}$$

which is negligible. Now the number of $n \le x$ not belonging to $\mathscr{E}$ or $\mathscr{E}'$ for which (3.1) fails is at most

$$2(\log_2 x)^{1/4} \sum_{\substack{n \le x \\ n \notin \mathscr{E} \cup \mathscr{E}'}} \Sigma_1(n;x). \tag{3.2}$$

To estimate the sum in (3.2), we notice that

$$\sum_{\substack{n \le x \\ n \notin \mathscr{E} \cup \mathscr{E}'}} \Sigma_1(n;x) = \sum_{d \le y^{1/2}} \frac{1}{d} \sum_{\substack{n \le x \\ n \notin \mathscr{E} \cup \mathscr{E}' \\ d \mid s(n),\, d \nmid n}} 1 \le \sum_{(\log_2 x)^{1/2} < d \le y^{1/2}} \frac{1}{d} \sum_{\substack{n \le x \\ n \notin \mathscr{E} \\ d \mid s(n)}} 1.$$

In the second step, we have used that $n \notin \mathscr{E}'$, so that each $d \le \sqrt{\log_2 x}$ either divides both of $s(n)$ and n or neither. By Lemma 2.7,

$$\sum_{(\log_2 x)^{1/2} < d \le y^{1/2}} \frac{1}{d} \sum_{\substack{n \le x \\ n \notin \mathscr{E} \\ d \mid s(n)}} 1 \ll x \log_3 x \sum_{d > (\log_2 x)^{1/2}} \frac{\tau(d)}{d\varphi(d)}. \tag{3.3}$$

We will estimate the remaining sum by partial summation. For every $t \ge 1$, put

$$S(t) = \sum_{d \le t} \tau(d) \frac{d}{\varphi(d)}.$$

Define an auxiliary arithmetic function g so that $\tau(d)\frac{d}{\varphi(d)} = \sum_{r|d} g(r)$ for every d. We easily compute that on primes p, we have $g(p) = 2\frac{p}{\varphi(p)} - 1$ and $g(p^k) = \frac{p}{\varphi(p)}$ for $k \geq 2$. In particular, g is nonnegative, and so

$$(3.4)\quad S(t) = \sum_{d\leq t}\sum_{r|d} g(r) \leq t\sum_{r\leq t}\frac{g(r)}{r} \leq t\prod_{p\leq t}\left(1+\frac{g(p)}{p}+\frac{g(p^2)}{p^2}+\cdots\right)$$
$$\leq t\exp\left(\sum_{p\leq t}\sum_{k=1}^{\infty}\frac{g(p^k)}{p^k}\right).$$

Since $g(p)/p = 1/p + O(1/p^2)$ while $g(p^k)/p^k \ll p^{-k}$ for $k \geq 2$, we find that $\sum_{k=1}^{\infty}\frac{g(p^k)}{p^k} \leq 1/p + O(1/p^2)$, and so

$$\sum_{p\leq t}\sum_{k=1}^{\infty}\frac{g(p^k)}{p^k} \leq \log_2 t + O(1).$$

Inserting this into (3.4) shows that $S(t) \ll t\log t$ for $t \geq 2$, and so

$$\sum_{d>(\log_2 x)^{1/2}}\frac{\tau(d)}{d\varphi(d)} = \int_{(\log_2 x)^{1/2}}^{\infty}\frac{1}{t^2}\,\mathrm{d}S(t) \ll \int_{(\log_2 x)^{1/2}}^{\infty}\frac{\log t}{t^2}\,\mathrm{d}t \ll \frac{\log_3 x}{(\log_2 x)^{1/2}}.$$

Hence, the right-hand side of (3.3) is $O(x(\log_3 x)^2/(\log_2 x)^{1/2})$, and the expression (3.2) is $O(x(\log_3 x)^2/(\log_2 x)^{1/4})$, as desired.

Remark. In his 1976 article [Erd76], Erdős briefly sketches his own proof of the $K = 2$ case of Conjecture 1.3. He remarks that this is a simple consequence of the following result:

Erdős's Lemma 4. *To every $\varepsilon > 0$, there is an ℓ so that for all x,*

$$\sum_{1<n\leq x}\sum_{\substack{p|s(n)\\ p>\ell}}\frac{1}{p} < \varepsilon x.$$[1]

Concerning this lemma, he writes:

> Unfortunately, I have at present only a very messy proof of the lemma and this is the reason that I suppress it. I am fairly sure that an elegant and simple proof exists.

Erdős's Lemma 4 follows from quickly from our methods: One discards $n \in \mathscr{E}(x)$ and restricts the inner sum to $p \leq y^{1/2}$. For large x, this changes the double sum by at most $\varepsilon x/2$. That the remaining sum is $<\varepsilon x/2$, if ℓ was initially chosen large, follows from Lemma 2.8.

[1] Erdős allows $n = 1$ in the double sum, but this is a typo.

3.2. Proof of Corollary 1.5. We need a result of Toulmonde that is contained in the proof of [Tou06, Théorème 1]; see especially pp. 382–383 and his remarks in §10.

PROPOSITION 3.1. *Fix a number ρ in the range of $\sigma(n)/n$. Then for all t sufficiently large (where 'large' may depend on ρ) and all $x \geq 1$, the number of $n \leq x$ for which $\sigma(n)/n \in [\rho - 1/t, \rho)$ is*

$$\ll x/\exp\left(\frac{1}{5}\sqrt{\log t \log_2 t}\right).$$

Proof of Corollary 1.5. Suppose $n \leq x$, that $s(n) < n$, and that $s(s(n)) \geq s(n)$. If $s(n)/n < 1 - (\log_2 x)^{-1/4}$, then n is in the exceptional set counted by Theorem 1.4. The number of these n is $O(x(\log_3 x)^2/(\log_2 x)^{1/4})$, which fits within the upper bound claimed in Corollary 1.5. In the opposite case, the asserted upper bound follows from Proposition 3.1 with $\rho = 2$ and $t = (\log_2 x)^{1/4}$. (Certainly $\rho = 2$ is in the range of $\sigma(n)/n$, since $\sigma(6) = 2 \cdot 6$.) $\square$

4. Abundant and squarefree values of $\beta(n)$

4.1. The abundancy of $\beta(n)$. Our task is to show that $\frac{s(n)}{n}$ and $\frac{s(\beta(n))}{\beta(n)}$ have the same distribution. We reformulate this problem in terms of $h(n)$ and $h(\beta(n))$, where $h(n) := \frac{n}{\sigma(n)}$. This ends up being simpler, since h is multiplicative and universally bounded between 0 and 1. Define the distribution function

$$\tilde{D}(u) := \lim_{x\to\infty} \frac{1}{x}\#\{n \leq x : h(n) \leq u\},$$

and note that Davenport's function $D(u)$ satisfies the relation $D(u) = 1 - \tilde{D}((u+1)^{-1})$. Let

$$F(u) := \lim_{x\to\infty} \frac{1}{x}\#\left\{n \leq x : \frac{s(\beta(n))}{\beta(n)} \leq u\right\} \quad \text{and}$$
$$\tilde{F}(u) := \lim_{x\to\infty} \frac{1}{x}\#\{n \leq x : h\big(\beta(n)\big) \leq u\}.$$

We will show that $\tilde{F}(u) = \tilde{D}(u)$ for every u. Theorem 1.7 follows in view of the relation $F(u) = 1 - \tilde{F}((u+1)^{-1})$.

We use the method of moments as described, for example, in the textbook of Billingsley (see especially [Bil95, Theorems 30.1 and 30.2, pp. 388–390]). Since both $h(n)$ and $h(\beta(n))$ are bounded between 0 and 1, to show that $\tilde{F} = \tilde{D}$, it is enough to show that their sequences of moments agree. Equivalently, it suffices to show that for every positive integer k,

$$\lim_{x\to\infty} \frac{1}{x}\sum_{1<n\leq x} h(n)^k = \lim_{x\to\infty} \frac{1}{x}\sum_{1<n\leq x} h\big(\beta(n)\big)^k.$$

For each k, the left-hand limit is straightforward to determine by a direct calculation, already carried out in [Pol14]: Define an arithmetic function g so that

$$h(n) = \sum_{d \mid n} g(d) \tag{4.1}$$

for every natural number n. Then

$$\lim_{x\to\infty} \frac{1}{x} \sum_{1<n\le x} h(n)^k = \mu_k,$$

where μ_k is given by the convergent sum

$$\mu_k := \sum_{d_1,\dots,d_k} \frac{g(d_1)\cdots g(d_k)}{\mathrm{lcm}[d_1,\dots,d_k]}, \tag{4.2}$$

and the sum is taken over all k-tuples of positive integers $d_1,\dots,d_k$ [Pol14, Lemma 2.2]. So the proof of Theorem 1.7 has been reduced to establishing the following result.

Lemma 4.1. *For each positive integer k,*

$$\lim_{x\to\infty} \frac{1}{x} \sum_{1<n\le x} h\big(\beta(n)\big)^k = \mu_k,$$

where μ_k is defined by (4.2).

Proof. In estimating $\sum_{1<n\le x} h(\beta(n))^k$, we may ignore the contribution from $n \in \mathscr{E}(x)$. Indeed, since $0 \le h \le 1$ and $\#\mathscr{E}(x) = o(x)$, removing the terms $n \in \mathscr{E}(x)$ changes the sum by only $o(x)$. Using a $'$ to denote a sum restricted to $n \notin \mathscr{E}(x)$, we find that

$$\begin{aligned} \frac{1}{x} \sideset{}{'}\sum_{1<n\le x} h\big(\beta(n)\big)^k &= \frac{1}{x} \sideset{}{'}\sum_{1<n\le x} \Bigg(\sum_{d\mid\beta(n)} g(d)\Bigg)^k \\ &= \frac{1}{x} \sum_{d_1,\dots,d_k} g(d_1)\cdots g(d_k) \sideset{}{'}\sum_{\substack{1<n\le x\\ \mathrm{lcm}[d_1,\dots,d_k]\mid\beta(n)}} 1. \end{aligned} \tag{4.3}$$

Let w be a large, fixed positive integer. Consider the contribution to (4.3) from tuples $(d_1,\dots,d_k)$ with some $d_i > w$. This contribution is bounded in absolute value by

$$\frac{1}{x} \sum_{\substack{d_1,\dots,d_k\\ \text{some } d_i>w}} \big|g(d_1)\big|\cdots\big|g(d_k)\big| \sideset{}{'}\sum_{\substack{1<n\le x\\ \mathrm{lcm}[d_1,\dots,d_k]\mid\beta(n)}} 1. \tag{4.4}$$

Recalling the definition (4.1) of g, we calculate that for each prime p and each integer $e \geq 1$,

$$g(p^e) = \frac{p^e}{\sigma(p^e)} - \frac{p^{e-1}}{\sigma(p^{e-1})} = -\frac{p^{e-1}}{\sigma(p^e)\sigma(p^{e-1})},$$

so that $|g(p^e)| \leq p^{-e}$. Hence, each $|g(d_i)| \leq d_i^{-1}$. Using Lemma 2.15 with $\varepsilon = \frac{1}{4}$, we find that (4.4) is

$$\ll \sum_{\substack{d_1,\dots,d_k \\ \text{some } d_i > w}} \frac{1}{d_1 \cdots d_k \cdot \operatorname{lcm}[d_1,\dots,d_k]^{3/4}}.$$

Now $\operatorname{lcm}[d_1,\dots,d_k] \geq \max\{d_1,\dots,d_k\} \geq (d_1 \cdots d_k)^{1/k}$, so that

$$\begin{aligned} \sum_{\substack{d_1,\dots,d_k \\ \text{some } d_i > w}} \frac{1}{d_1 \cdots d_k \cdot \operatorname{lcm}[d_1,\dots,d_k]^{\frac{3}{4}}} &\leq \sum_{\substack{d_1,\dots,d_k \\ \text{some } d_i > w}} \frac{1}{(d_1 \cdots d_k)^{1+\frac{3}{4k}}} \\ &\ll_k \Bigg(\sum_{d_1 > w} \frac{1}{d_1^{1+\frac{3}{4k}}}\Bigg)\Bigg(\sum_{d \geq 1} \frac{1}{d^{1+\frac{3}{4k}}}\Bigg)^{k-1} \\ &\ll_k w^{-\frac{3}{4k}}, \end{aligned}$$

which tends to 0 with w.

Now consider the contribution to (4.3) from tuples $(d_1,\dots,d_k)$ with each component $d_i \leq w$. By Corollary 2.10, for any such tuple, the inner sum in (4.3) is asymptotic to $x/\operatorname{lcm}[d_1,\dots,d_k]$, as $x \to \infty$. Note that the restriction to $n \notin \mathscr{E}(x)$ does not affect the asymptotics since $\#\mathscr{E}(x) = o(x)$. Hence, as $x \to \infty$, the contribution to (4.3) from these $d_1,\dots,d_k$ tends to

$$\sum_{d_1,\dots,d_k \leq w} \frac{g(d_1)\cdots g(d_k)}{\operatorname{lcm}[d_1,\dots,d_k]}.$$

This expression can be made arbitrarily close to μ_k by initially choosing the fixed parameter w large enough.

It follows that both the $\liminf_{x\to\infty}$ and $\limsup_{x\to\infty}$ of $\frac{1}{x}\sum_{1<n\leq x} h(\beta(n))^k$ are within $O_k(w^{-3/4k})$ of μ_k. Letting $w \to \infty$ completes the proof. □

4.2. Squarefree values of $\beta(n)$; proof of Theorem 1.8. Let w be a large, fixed positive integer, and let $W = \prod_{p \leq w} p$. The condition that $\beta(n)$ not be divisible by p^2 for any prime $p \leq w$ places $\beta(n)$ into one of $\prod_{p\leq w}(p^2-1)$ residue classes modulo W^2. By Corollary 2.10 with $q = W^2$, the density of the corresponding set of n is

$$\prod_{p \leq w}\left(1 - \frac{1}{p^2}\right).$$

By choosing the fixed parameter w large enough, this can be made arbitrarily close to $\frac{6}{\pi^2}$. Thus, it suffices to prove that the upper density of n where $\beta(n)$ is divisible by p^2 for some $p > w$ tends to 0, as $w \to \infty$. Now the number of $n \le x$ where $\beta(n)$ has a divisor p^2 with $p > w$ is at most

$$\#\mathscr{E}(x) + \sum_{p>w} \sum_{\substack{n\le x\\ n\notin\mathscr{E}\\ p^2\mid\beta(n)}} 1 \ll \#\mathscr{E}(x) + x\sum_{p>w} p^{-3/2},$$

using Lemma 2.15 with $q = p^2$ and $\varepsilon = \frac{1}{4}$ in the last step. Dividing by x and letting $x \to \infty$, we get that the upper density in question is $\ll \sum_{p>w} p^{-3/2} \ll w^{-1/2}$, which tends to 0 as $w \to \infty$.

5. Prime values of $\beta(n)$ and $s(n)$

5.1. Proof of Theorem 1.10. For each composite integer $n \le x$, write $n = mP$, where $P := P(n)$ and $m > 1$. We begin by discarding those $n \le x$ with $P(n) \le x^{1/\log_2 x}$ or $P(n)^2 \mid n$; this eliminates only $O(x/(\log x)^2)$ integers (see Proposition 2.5). We partition the remaining n as follows. Let $L := \lceil \log x \rceil$, and let $\mathscr{L} := \{0, 1/L, 2/L, \dots, (L-1)/L\}$. We can find $\lambda = \lambda(n) \in \mathscr{L}$ with $\lambda < \frac{\log m}{\log x} \le \lambda + \frac{1}{L}$; then

$$x^\lambda < m \le ex^\lambda.$$

We now fix $\lambda \in \mathscr{L}$ and count the corresponding values of n; at the end of the argument, we sum over λ. If λ corresponds to at least one value of n under consideration, then

$$x^{1-\lambda} > x/m \ge P > x^{1/\log_2 x},$$

and so we may assume this lower bound on $x^{1-\lambda}$ in what follows. Since both P and $\beta(n) = P + \beta(m)$ are prime, the upper bound sieve (see, e.g., [HR74, Corollary 2.4.1, p. 80]) shows that for a given m, the number of possible $P \le x/m$ is

$$\ll \frac{x/m}{(\log \frac{x}{m})^2} \prod_{p\mid\beta(m)} \left(1 - \frac{1}{p}\right)^{-1} \ll \frac{x^{1-\lambda}}{(1-\lambda)^2(\log x)^2} \sum_{d\mid\beta(m)} \frac{1}{d}. \tag{5.1}$$

Since $P(m) \le P \le x^{1-\lambda}$, we obtain an upper bound for the number of n corresponding to a given λ by summing (5.1) over $m \in (1, ex^\lambda]$ with $P(m) \le x^{1-\lambda}$. By Cauchy–Schwarz,

$$\sum_{\substack{1<m\le ex^\lambda\\ P(m)\le x^{1-\lambda}}} \sum_{d\mid\beta(m)} \frac{1}{d} \le \Bigg(\sum_{\substack{1<m\le ex^\lambda\\ P(m)\le x^{1-\lambda}}} 1 \Bigg)^{1/2} \Bigg(\sum_{1<m\le ex^\lambda} \Bigg(\sum_{d\mid\beta(m)} \frac{1}{d} \Bigg)^2 \Bigg)^{1/2}. \tag{5.2}$$

The first sum on the right-hand side of (5.2) is at most $\Psi(ex^\lambda, x^{1-\lambda})$. Take first the case when $\lambda > \frac{1}{2}$. Applying Lemma 2.3 (with $z=2$), we get that $\Psi(ex^\lambda, x^{1-\lambda}) \ll x^\lambda e^{-u/2}$, where

$$u = \frac{1+\lambda\log x}{(1-\lambda)\log x} > \frac{\lambda}{1-\lambda} > \frac{1}{2(1-\lambda)}.$$

Since $e^{-u/2} \ll u^{-4}$ for $u \geq 1$, we conclude that

$$\sum_{\substack{1<m\leq ex^\lambda \\ P(m)\leq x^{1-\lambda}}} 1 \ll x^\lambda (1-\lambda)^4.$$

All of this was obtained under the assumption that $\lambda > \frac{1}{2}$, but the final estimate is trivially also valid for $\lambda \leq \frac{1}{2}$.

We now turn to the second parenthesized expression on the right of (5.2). We split this into two pieces:

$$\sum_{1<m\leq ex^\lambda} \Bigg(\sum_{d\mid\beta(m)} \frac{1}{d}\Bigg)^2 \leq \sum_{\substack{1<m\leq ex^\lambda \\ m\in\mathscr{E}(ex^\lambda)}} \Bigg(\sum_{d\mid\beta(m)} \frac{1}{d}\Bigg)^2 + \sum_{\substack{1<m\leq ex^\lambda \\ m\notin\mathscr{E}(ex^\lambda)}} \Bigg(\sum_{d\mid\beta(m)} \frac{1}{d}\Bigg)^2. \tag{5.3}$$

It is straightforward to prove that $\beta(m) \leq m$ for every m (see, for instance, [Jak12, Theorem 2.3]); so from the maximal order of the sum-of-divisors function (see [HW08, Theorem 323, p. 350]), we see that $\sum_{d\mid\beta(m)} 1/d = \sigma(\beta(m))/\beta(m) \ll \log_2(ex^\lambda)$ uniformly for all $1 < m \leq ex^\lambda$. So the first piece on the right of (5.3) is

$$\ll \#\mathscr{E}\big(ex^\lambda\big) \cdot \big(\log_2\big(ex^\lambda\big)\big)^2 \ll x^\lambda.$$

To handle the second piece, notice that

$$\sum_{\substack{1<m\leq ex^\lambda \\ m\notin\mathscr{E}(ex^\lambda)}} \Bigg(\sum_{d\mid\beta(m)} \frac{1}{d}\Bigg)^2 = \sum_{d_1,d_2} \frac{1}{d_1 d_2} \sum_{\substack{1<m\leq ex^\lambda \\ m\notin\mathscr{E}(ex^\lambda) \\ [d_1,d_2]\mid\beta(m)}} 1 \ll x^\lambda \sum_{d_1,d_2} \frac{1}{d_1 d_2 [d_1,d_2]^{1/2}};$$

in the last step, we used Lemma 2.15 with $\varepsilon = \frac{1}{2}$. Since $\operatorname{lcm}[d_1,d_2] \geq (d_1 d_2)^{1/2}$, the final double sum is bounded by $(\sum_d 1/d^{5/4})^2$. We conclude that the right-hand side of (5.3) is $O(x^\lambda)$.

Collecting all of the above estimates, (5.1) and (5.2) yield that the number of n corresponding to a given λ is

$$\ll \frac{x^{1-\lambda}}{(1-\lambda)^2(\log x)^2} \cdot \big(x^\lambda(1-\lambda)^4\big)^{1/2} \cdot \big(x^\lambda\big)^{1/2} = \frac{x}{(\log x)^2}.$$

Summing on the $O(\log x)$ possible values of λ completes the proof.

5.2. Proof of Theorem 1.11. Write $n = mP$, where $P := P(n)$. We can assume that $P > x^{1/\log_2 x}$ and that $P \nmid m$, since the number of $n \le x$ for which either condition fails is $O(x/(\log x)^2)$. Thus,

$$m < x/\exp(\log x/\log_2 x). \tag{5.4}$$

Suppose that $s(n)$ is prime. Then $m > 1$. Moreover, $\gcd(m, \sigma(m)) = 1$. Indeed,

$$\gcd\big(m,\sigma(m)\big) \mid (P+1)\sigma(m) - mP = s(n), \quad \text{and}$$
$$s(n) = Ps(m) + \sigma(m) > \gcd\big(m,\sigma(m)\big),$$

so that $s(n)$ cannot be prime unless $\gcd(m, \sigma(m)) = 1$.

We begin by crudely bounding the number of n corresponding to a given m. Since P and $Ps(m) + \sigma(m)$ are both prime, Brun's sieve (see [HR74, Theorem 2.4] for the specific formulation used here) shows that given m, the number of possibilities for the prime $P \le x/m$ is

$$\ll \frac{x/m}{(\log \frac{x}{m})^2} \prod_{p \mid s(m)} (1-1/p)^{\rho(p)-1} \prod_{p \mid \sigma(m)} (1-1/p)^{-1}, \tag{5.5}$$

where

$$\rho(p) := \#\big\{u \bmod p : us(m) + \sigma(m) \equiv 0 \pmod p\big\}.$$

Since m and $\sigma(m)$ are relatively prime, $\rho(p) = 0$ whenever $p \mid s(m)$, and the upper bound (5.5) is

$$\ll \frac{x/m}{(\log \frac{x}{m})^2} \prod_{p \mid s(m)\sigma(m)} (1+1/p). \tag{5.6}$$

From (5.4), we have $\log(x/m) > \log x/\log_2 x$. Also, $\prod_{p \mid s(m)\sigma(m)}(1+1/p) \le \frac{\sigma(s(m)\sigma(m))}{s(m)\sigma(m)}$. Since $s(m)\sigma(m) \le \sigma(m)^2 < x^2$, we get from the maximal order of the σ-function that $\frac{\sigma(s(m)\sigma(m))}{s(m)\sigma(m)} \ll \log_2 x$. Thus, the number of n corresponding to a given m is

$$\ll \frac{x(\log_2 x)^3}{m(\log x)^2}. \tag{5.7}$$

We use the crude upper bound (5.7) to show that we can impose a number of additional assumptions on m. For instance, we can suppose that

$$m \ge \exp\big(\log x/(\log_2 x)^4\big), \tag{5.8}$$

since summing (5.7) over m below this bound gives a count of corresponding n that is $o(x/\log x)$. We now argue that we can also assume all of

(i) $\prod_{p \le \sqrt{\log_2 x}} p$ divides $\sigma(m)$,

(ii) $\sum_{\substack{p \mid \sigma(m) \\ p > (\log_2 x)^{10}}} \frac{1}{p} \le 1$,

(iii) $\sum_{\substack{p \mid s(m)\\ p>(\log_2 x)^{10}}} \frac{1}{p} \le 1$.

Let $L = \lceil \log x \rceil$, and choose $\lambda \in \{0, 1/L, \dots, (L-1)/L\}$ as large as possible with $m > x^\lambda$. Then $x^\lambda < m \le ex^\lambda$. Let us estimate the contribution to (5.7) from $m \in (x^\lambda, ex^\lambda]$ failing at least one of (i), (ii), and (iii). By Lemma 2.1, the number of $m \in (x^\lambda, ex^\lambda]$ where (i) fails is

$$\begin{aligned}
&\ll \sum_{d \le \sqrt{\log_2 x}} x^\lambda / \big(\log x^\lambda\big)^{1/\varphi(d)} \ll x^\lambda \sum_{d \le \sqrt{\log_2 x}} \exp\big(-\log_2\big(x^\lambda\big)/\sqrt{\log_2 x}\big)\\
&\ll x^\lambda / \exp\left(\frac{1}{2}\sqrt{\log_2 x}\right) \ll x^\lambda/(\log_2 x)^4.
\end{aligned}$$

To go from the first line to the second, we used that $x^\lambda \asymp m$ and that m satisfies the lower bound (5.8). From Lemma 2.2, the count of $m \in (x^\lambda, ex^\lambda]$ where (ii) fails does not exceed

$$\begin{aligned}
\sum_{m \le ex^\lambda} \sum_{\substack{p \mid \sigma(m)\\ p > (\log_2 x)^{10}}} \frac{1}{p} &\le \sum_{p > (\log_2 x)^{10}} \frac{1}{p} \sum_{\substack{m \le ex^\lambda\\ p \mid \sigma(m)}} 1\\
&\ll x^\lambda \log_2 x \sum_{p > (\log_2 x)^{10}} \frac{1}{p^{3/2}} \ll \frac{x^\lambda}{(\log_2 x)^4}.
\end{aligned}$$

We proceed similarly to handle (iii). Observe that since $s(m) < \sigma(m) < x$,

$$\sum_{\substack{p \mid s(m)\\ p > \log x}} \frac{1}{p} \le \frac{1}{\log x} \sum_{\substack{p \mid s(m)\\ p > \log x}} 1 \le \frac{1}{\log x} \cdot \frac{\log s(m)}{\log_2 x} \le \frac{1}{\log_2 x}.$$

So if m fails (iii), then (assuming x is large)

$$\sum_{\substack{p \mid s(m)\\ (\log_2 x)^{10} < p \le \log x}} \frac{1}{p} \ge \frac{1}{2}. \tag{5.9}$$

Now the size of the set $\mathscr{E}(ex^\lambda)$ is $O(x^\lambda/(\log_2 x)^4)$, while the number of $m \in (x^\lambda, ex^\lambda]$ satisfying (5.9) and not belonging to $\mathscr{E}(ex^\lambda)$ is at most

$$\begin{aligned}
2 \sum_{\substack{m \le ex^\lambda\\ m \notin \mathscr{E}(ex^\lambda)}} \sum_{\substack{p \mid s(m)\\ (\log_2 x)^{10} < p \le \log x}} \frac{1}{p} &= 2 \sum_{(\log_2 x)^{10} < p \le \log x} \frac{1}{p} \sum_{\substack{m \le ex^\lambda\\ m \notin \mathscr{E}(ex^\lambda)\\ p \mid s(m)}} 1\\
&\ll x^\lambda \log_3 x \sum_{p > (\log_2 x)^{10}} \frac{1}{p^2} \ll \frac{x^\lambda}{(\log_2 x)^9}.
\end{aligned}$$

Here we applied Lemma 2.7 (with x replaced by ex^λ and q replaced by p). Collecting everything, we see that the number of $m \in (x^\lambda, ex^\lambda]$ failing one of (i), (ii), and (iii) is $O(x^\lambda/(\log_2 x)^4)$. Summing the bound (5.7) over these m, we see that the number of corresponding n is

$$\ll \frac{x(\log_2 x)^3}{(\log x)^2}\cdot\frac{1}{x^\lambda}\cdot\frac{x^\lambda}{(\log_2 x)^4} \ll \frac{x}{(\log x)^2\log_2 x}.$$

Summing over the $O(\log x)$ possible values of λ shows that there are $o(x/\log x)$ total values of $n \le x$ that arise. So we may indeed assume (i), (ii), and (iii).

We complete the proof in the same way as the proof of Theorem 1.10. We continue to assume that $\lambda \in \{0, 1/L, \dots, (L-1)/L\}$ is chosen so that $\lambda < \frac{\log m}{\log x} \le \lambda + \frac{1}{L}$. Then $P(m) \le P \le \frac{x}{m} < x^{1-\lambda}$. So from (5.6), the number of n corresponding to a given λ is

$$\ll \frac{x^{1-\lambda}}{(1-\lambda)^2(\log x)^2} \sideset{}{'}\sum_{\substack{m\in(x^\lambda, ex^\lambda] \\ P(m)\le x^{1-\lambda}}} \prod_{p\mid s(m)\sigma(m)}\left(1+\frac{1}{p}\right), \tag{5.10}$$

where the $'$ on the sum indicates that m satisfies all of the restrictions so far imposed. By (ii) and (iii),

$$\prod_{p\mid s(m)\sigma(m)}\left(1+\frac{1}{p}\right) \le \prod_{p\le(\log_2 x)^{10}}\left(1+\frac{1}{p}\right) \prod_{\substack{p\mid s(m)\sigma(m)\\ p>(\log_2 x)^{10}}}\left(1+\frac{1}{p}\right) \ll \log_3 x,$$

uniformly for the m appearing above. From (i) and the condition that m is relatively prime to $\sigma(m)$, we know that m has no prime factors up to $\sqrt{\log_2 x}$. It follows that

$$\sideset{}{'}\sum_{\substack{m\in(x^\lambda, ex^\lambda] \\ P(m)\le x^{1-\lambda}}} 1 \le \Psi\big(ex^\lambda, \big[\sqrt{\log_2 x}, x^{1-\lambda}\big]\big).$$

For $\lambda > \frac{1}{2}$, we may apply Lemma 2.3 to see that

$$\begin{aligned}\Psi\big(ex^\lambda, \big[\sqrt{\log_2 x}, x^{1-\lambda}\big]\big) &\ll \frac{x^\lambda}{\log_3 x}e^{-u/2}, \quad \text{where } u = \frac{\log(ex^\lambda)}{\log(x^{1-\lambda})}\\ &\ll \frac{x^\lambda}{\log_3 x}(1-\lambda)^2.\end{aligned}$$

When $\lambda \le \frac{1}{2}$, the same estimate holds by an elementary sieve (e.g., inclusion–exclusion is sufficient). Collecting these results, (5.10) is seen to be

$$\ll \frac{x^{1-\lambda}}{(1-\lambda)^2(\log x)^2}\cdot\log_3 x\cdot\frac{x^\lambda}{\log_3 x}(1-\lambda)^2 = \frac{x}{(\log x)^2}.$$

Summing on the $O(\log x)$ possible values of λ completes the proof of Theorem 1.11.

Acknowledgments. The author thanks Carl Pomerance and Enrique Treviño for stimulating conversations. He also thanks the referee for a careful reading of the manuscript as well as a number of helpful suggestions that improved the exposition.

References

[AE77] K. Alladi and P. Erdős, *On an additive arithmetic function*, Pacific J. Math. **71** (1977), 275–294. MR 0447086

[Bil95] P. Billingsley, *Probability and measure*, 3rd ed., Wiley Series in Probability and Mathematical Statistics, Wiley, New York, 1995. MR 1324786

[BL07] W. D. Banks and F. Luca, *Sums of prime divisors and Mersenne numbers*, Houston J. Math. **33** (2007), 403–413 (electronic). MR 2308986

[Cat88] E. Catalan, *Propositions et questions diverses*, Bull. Soc. Math. France **16** (1888), 128–129. MR 1504036

[Dav33] H. Davenport, *Über numeri abundantes*, S.-Ber. Preuß. Akad. Wiss., Math.-Nat. Kl. (1933), 830–837.

[dB66] N. G. de Bruijn, *On the number of positive integers $\leq x$ and free prime factors $> y$. II*, Nederl. Akad. Wetensch. Proc. Ser. A Indag. Math. **28** (1966), 239–247. MR 0205945

[Dic13] L. E. Dickson, *Theorems and tables on the sum of the divisors of a number*, Quart. J. Math. **44** (1913), 264–296.

[DKL05] J.-M. De Koninck and F. Luca, *Integers divisible by the sum of their prime factors*, Mathematika **52** (2005), 69–77. MR 2261843

[DKL08] J.-M. De Koninck and F. Luca, *Integers divisible by sums of powers of their prime factors*, J. Number Theory **128** (2008), 557–563. MR 2389855

[EGPS90] P. Erdős, A. Granville, C. Pomerance and C. Spiro, *On the normal behavior of the iterates of some arithmetic functions*, Analytic number theory (Allerton Park, IL, 1989), Progr. Math., vol. 85, Birkhäuser, Boston, MA, 1990, pp. 165–204. MR 1084181

[Ell79] P. D. T. A. Elliott, *Probabilistic number theory. I: Mean-value theorems*, Grundlehren der Mathematischen Wissenschaften, vol. 239, Springer, New York, 1979. MR 0551361

[EP78] P. Erdős and C. Pomerance, *On the largest prime factors of n and $n+1$*, Aequationes Math. **17** (1978), 311–321. MR 0480303

[Erd76] P. Erdős, *On asymptotic properties of aliquot sequences*, Math. Comp. **30** (1976), 641–645. MR 0404115

[Hal68] G. Halász, *Über die Mittelwerte multiplikativer zahlentheoretischer Funktionen*, Acta Math. Acad. Sci. Hungar. **19** (1968), 365–403. MR 0230694

[Hal70] R. R. Hall, *On the probability that n and $f(n)$ are relatively prime*, Acta Arith. **17** (1970), 169–183. Corrigendum in **19** (1971), 203–204. MR 0268148

[Hal71] R. R. Hall, *On the probability that n and $f(n)$ are relatively prime. II*, Acta Arith. **19** (1971), 175–184. MR 0289436

[Hal72] R. R. Hall, *On the probability that n and $f(n)$ are relatively prime. III*, Acta Arith. **20** (1972), 267–289. MR 0304339

[HR74] H. Halberstam and H.-E. Richert, *Sieve methods*, London Mathematical Society Monographs, vol. 4, Academic Press, London–New York, 1974. MR 0424730

[HR79] H. Halberstam and H.-E. Richert, *On a result of R. R. Hall*, J. Number Theory **11** (1979), 76–89. MR 0527762

[HW08] G. H. Hardy and E. M. Wright, *An introduction to the theory of numbers*, 6th ed., Oxford University Press, Oxford, 2008. MR 2445243

[Jak12] R. Jakimczuk, *Sum of prime factors in the prime factorization of an integer*, Int. Math. Forum **7** (2012), 2617–2621. MR 2971264

[Kob10] M. Kobayashi, *On the density of abundant numbers*, Ph.D. thesis, Dartmouth College, 2010. MR 2996025

[Kob14] M. Kobayashi, *A new series for the density of the abundant numbers*, Int. J. Number Theory **10** (2014), 73–84. MR 3189967

[KPP09] M. Kobayashi, P. Pollack and C. Pomerance, *On the distribution of sociable numbers*, J. Number Theory **129** (2009), 1990–2009. MR 2522719

[Len75] H. W. Lenstra, *Problem 6064*, Amer. Math. Monthly **82** (1975), 1016. Solution by the proposer in **84** (1977), 580.

[Pol14] P. Pollack, *Equidistribution mod q of abundant and deficient numbers*, Unif. Distrib. Theory **9** (2014), 99–114. MR 3237077

[Pom77] C. Pomerance, *On the distribution of amicable numbers*, J. Reine Angew. Math. **293/294** (1977), 217–222. MR 0447087

[Pom02] C. Pomerance, *Ruth–Aaron numbers revisited*, Paul Erdős and his mathematics, I (Budapest, 1999), Bolyai Soc. Math. Stud., vol. 11, János Bolyai Math. Soc., Budapest, 2002, pp. 567–579. MR 1954715

[Ten95] G. Tenenbaum, *Introduction to analytic and probabilistic number theory*, Cambridge Studies in Advanced Mathematics, vol. 46, Cambridge University Press, Cambridge, 1995. MR 1342300

[Tou06] V. Toulmonde, *Module de continuité de la fonction de répartition de $\varphi(n)/n$*, Acta Arith. **121** (2006), 367–402. MR 2224402

[Whe90] F. S. Wheeler, *Two differential-difference equations arising in number theory*, Trans. Amer. Math. Soc. **318** (1990), 491–523. MR 0963247

Paul Pollack, Department of Mathematics, Boyd Graduate Studies Research Center, University of Georgia, Athens, GA 30602, USA

E-mail address: pollack@uga.edu