SUMS OF PROPER DIVISORS FOLLOW THE ERDŐS–KAC LAW

PAUL POLLACK AND LEE TROUPE

Abstract. Let \( s(n) = \sum_{d|n, \, d<n} d \) denote the sum of the proper divisors of \( n \). The second-named author proved that \( \omega(s(n)) \) has normal order \( \log \log n \), the analogue for \( s \)-values of a classical result of Hardy and Ramanujan. We establish the corresponding Erdős–Kac theorem: \( \omega(s(n)) \) is asymptotically normally distributed with mean and variance \( \log \log n \). The same method applies with \( s(n) \) replaced by any of several other unconventional arithmetic functions, such as \( \beta(n) := \sum_{p|n} p, n - \varphi(n), \) and \( n + \tau(n) \) (\( \tau \) being the divisor function).

1. Introduction

Let \( s(n) = \sum_{d|n, \, d<n} d \) denote the sum of the proper divisors of the positive integer \( n \), so that \( s(n) = \sigma(n) - n \). Interest in the value distribution of \( s(n) \) traces back to the ancient Greeks, but the modern study of \( s(n) \) could be considered to begin with Davenport [Dav33], who showed that \( s(n)/n \) has a continuous distribution function \( D(u) \). Precisely: For each real number \( u \), the set of \( n \) with \( s(n) \leq un \) has an asymptotic density \( D(u) \) which varies continuously with \( u \). Moreover, \( D(0) = 0 \) and \( \lim_{u \to \infty} D(u) = 1 \).

While the values of \( \sigma(n) = \prod_{p|n} \frac{p^{e+1}-1}{p-1} \) are multiplicatively special, we expect shifting by \(-n\) to rub out the peculiarities. That is, we expect the multiplicative statistics of \( s(n) \) to resemble those of numbers of comparable size. By Davenport’s theorem, it is usually safe to interpret “of comparable size” to mean “of the same order of magnitude as \( n \) itself”.

Various results in the literature validate this expectation about \( s(n) \). For example, the first author has shown [Pol14] that \( s(n) \) is prime for \( O(x/\log x) \) values of \( n \leq x \) (and he conjectures that the true count is \( \sim x/\log x \), in analogy with the prime number theorem). The second author [Tro20] has proved, in analogy with a classical result of Landau and Ramanujan, that there are \( \asymp x/\sqrt{\log x} \) values of \( n \leq x \) for which \( s(n) \) is a sum of two squares. Writing \( \omega(n) \) for the number of distinct prime factors of \( n \), he also showed [Tro15] that \( \omega(s(n)) \) has normal order \( \log \log n \). This is in harmony with the classical theorem of Hardy and Ramanujan [HR00] that \( \omega(n) \) itself has normal order \( \log \log n \).

In this note, we pick back up the study of \( \omega(s(n)) \). Strengthening the result of [Tro15], we prove that \( \omega(s(n)) \) satisfies the conclusion of the Erdős–Kac theorem [EK40].

Theorem 1. Fix a real number \( u \). As \( x \to \infty \),

\[
\frac{1}{x} \# \{ 1 < n \leq x : \omega(s(n)) - \log \log x \leq u \sqrt{\log \log x} \} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-\frac{1}{2}t^2} \, dt.
\]

2020 Mathematics Subject Classification. Primary 11N37; Secondary 11N56, 11N64.

P.P. is supported by NSF award DMS-2001581.
To prove Theorem 1, we adapt a simple and elegant proof of the Erdős–Kac theorem due to Billingsley ([Bil69], or [Bil95, pp. 395–397]). Making this go requires estimating, for squarefree $d$, the number of $n \leq x$ for which $d \mid s(n)$. A natural attack on this latter problem is to break off the largest prime factor $P$ of $n$, say $n = mP$. Most of the time, $P$ does not divide $m$, so that $\sigma(n) = \sigma(m)(P + 1)$. Then asking for $d$ to divide $s(n)$ amounts to imposing the congruence $s(m)P \equiv -\sigma(m) \pmod{d}$. For a given $m$, the corresponding $P$ are precisely those that lie in a certain interval $I_m$ and a certain (possibly empty) set of arithmetic progressions. At this point we adopt (and adapt) a strategy of Banks, Harman, and Shparlinski [BHS05]. Rather than analytically estimate the number of these $P$, we relate the count of such $P$ back to the total number of primes in the interval $I_m$, which we leave unestimated! This allows one to avoid certain losses of precision in comparison with [Tro15]. A similar strategy was used recently in [LLPSR] to show that $s(n)$, for composite $n \leq x$, is asymptotically uniformly distributed modulo primes $p \leq (\log x)^A$ (with $A > 0$ arbitrary but fixed).

Our proof of Theorem 1 is fairly robust. In the final section, we describe the modifications necessary to prove the corresponding result with $s(n)$ replaced by $\beta(n) := \sum_{p|n} p, n - \varphi(n)$, or $n + \tau(n)$, where $\tau(n)$ is the usual divisor-counting function.

For other recent work on the value distribution of $s(n)$, see [LP15, PP16, Pom18, PPT18].

Notation. Throughout, the letters $p$ and $P$ are reserved for primes. We write $(a, b)$ for the greatest common divisor of $a, b$. We let $P^+(n)$ denote the largest prime factor of $n$, with the convention that $P^+(1) = 1$. We write $\log_k$ for the $k$th iterate of the natural logarithm. We use $E$ for expectation and $V$ for variance.

2. Outline

We let $x$ be a large real number and we work on the probability space
\[ \Omega := \{ n \leq x : n \text{ composite, } P^+(n) > x^{1/\log_4 x}, \text{ and } P^+(n)^2 \nmid n \}, \]
equipped with the uniform measure. Standard arguments (compare with the proof of Lemma 2.2 in [Tro15]) show that as $x \to \infty$,
\[ \#\Omega = (1 + o(1))x. \]

We let $y = (\log x)^2$ and $z = x^{1/\log_3 x}$, and we define
\[ P = \{ \text{primes } p \text{ with } y < p \leq z \}. \]

For each prime $p \leq x^2$, we introduce the random variable $X_p$ on $\Omega$ defined by
\[ X_p(n) = \begin{cases} 1 & \text{if } p \mid s(n), \\ 0 & \text{otherwise.} \end{cases} \]

We let $Y_p$ be Bernoulli random variables which take the value 1 with probability $1/p$. We define
\[ X = \sum_{p \in P} X_p \quad \text{and} \quad Y = \sum_{p \in P} Y_p; \]
we think of $Y$ as an idealized model of $X$. 
Observe that
\[
\mu := \mathbb{E}[Y] = \sum_{p \in \mathcal{P}} \frac{1}{p} = \log \log z - \log \log y + o(1)
\]
\[
= \log \log x + o(\sqrt{\log \log x})
\]
(1)
while
\[
\sigma^2 := \mathbb{V}[Y] = \sum_{p \in \mathcal{P}} \frac{1}{p} \left(1 - \frac{1}{p}\right) \sim \log \log x.
\]
(2)
We renormalize \(Y\) to have mean 0 and variance 1 by defining
\[
\tilde{Y} = Y - \mu \sigma.
\]

**Lemma 2.** \(\tilde{Y}\) converges in distribution to the standard normal \(N\), as \(x \to \infty\). Moreover, \(\mathbb{E}[\tilde{Y}^k] \to \mathbb{E}[N^k]\) for each fixed positive integer \(k\).

**Proof (sketch).** Both claims follow from the proof in [Bil95, pp. 391–392] of the central limit theorem through the method of moments. One needs only that the recentered variables \(Y'_p := Y_p - \frac{1}{p}\), for \(p \in \mathcal{P}\), are independent mean 0 variables of finite variance, bounded by 1 in absolute value, with \(\sum_{p \in \mathcal{P}} \mathbb{V}[Y'_p] \to \infty\) as \(x \to \infty\). (Note that \(\sum_{p \in \mathcal{P}} \mathbb{V}[Y'_p] = \sum_{p \in \mathcal{P}} \mathbb{V}[Y_p] = \mathbb{V}[Y] = \sigma^2\) in our above notation.) \(\square\)

Let \(X = \sum_{p \in \mathcal{P}} X_p\) and \(\tilde{X} = \frac{X - \mu}{\sigma}\). The next section is devoted to the proof of the following proposition.

**Proposition 3.** For each fixed positive integer \(k\),
\[
\mathbb{E}[\tilde{X}^k] - \mathbb{E}[\tilde{Y}^k] \to 0.
\]

Lemma 2 and Proposition 3 imply that \(\mathbb{E}[\tilde{X}^k] \to \mathbb{E}[N^k]\), for each \(k\). So by the method of moments [Bil95, Theorem 30.2, p. 390], \(\tilde{X} = \frac{X - \mu}{\sigma}\) converges in distribution to the standard normal.

This is most of the way towards Theorem 1. Since \(#\Omega = (1 + o(1))x\) and \(\mu, \sigma\) satisfy the estimates (1), (2), Theorem 1 will follow if we show that \(\frac{\omega(s(\cdot)) - \mu}{\sigma}\) (viewed as a random variable on \(\Omega\)) converges in distribution to the standard normal. Observe that \(s(n) \leq \sum_{d \leq n} d < n^2 \leq x^2\) for every \(n \leq x\). So defining \(X^{(s)} = \sum_{p \leq y} X_p\) and \(X^{(l)} = \sum_{x < p \leq x^2} X_p\), we have \(\omega(s(\cdot)) = X^{(s)} + X + X^{(l)}\) on \(\Omega\) and
\[
\frac{\omega(s(\cdot)) - \mu}{\sigma} = \tilde{X} + \frac{X^{(s)}}{\sigma} + \frac{X^{(l)}}{\sigma}.
\]
Since \(\tilde{X}\) converges to the standard normal, to complete the proof of Theorem 1 it suffices to show that \(\frac{X^{(s)}}{\sigma}\) and \(\frac{X^{(l)}}{\sigma}\) converge to 0 in probability (see [Bil95, Theorem 25.4, p. 332]).
Convergence to 0 in probability is obvious for $X^{(t)}/\sigma$: A positive integer not exceeding $x^2$ has at most $\log(x^2)/\log x = 2\log x$ prime exceeding $x$, so that

$$|X^{(t)}/\sigma| \leq 2\log x/\sigma = o(1)$$
on the entire space $\Omega$. Since $\sigma \sim \sqrt{\log \log x}$, that $X^{(s)}/\sigma$ tends to 0 in probability follows from the next lemma together with Markov’s inequality.

**Lemma 4.** $\mathbb{E}[X^{(s)}] \ll \log x \log_4 x$ for all large $x$.

**Proof.** Put $L = x^{1/\log_4 x}$, and for each $m \leq x$, let $L_m = \max\{x^{1/\log_4 x}, P^+(m)\}$. The $n$ belonging to $\Omega$ are precisely the positive integers $n$ that admit a decomposition $n = mP$, where $m > 1$ and $L_m < P \leq x/m$. Note that this decomposition of $n$ is unique whenever it exists, since one can recover $P$ from $n$ as $P^+(n)$.

Let $n \in \Omega$ and write $n = mP$ as above. Then $s(mP) = \sigma(m)(P + 1) - mP = Ps(m) + \sigma(m)$. Hence, for each $p \leq y$,

$$\sum_{n \in \Omega} X_p(n) = \sum_{n \in \Omega} 1 = \sum_{1 < m < x/L} \sum_{L_m < P \leq x/m} \sum_{Ps(m) \equiv -\sigma(m) \pmod{p}} 1.$$

If $p \nmid s(m)$, then the congruence $Ps(m) \equiv -\sigma(m) \pmod{p}$ puts $P$ in a determined congruence class mod $p$ (possibly 0 mod $p$). By Brun–Titchmarsh, the number of such $P \leq x/m$ is

$$\ll \frac{x}{mp \log(x/mp)} \ll \frac{x \log_4 x}{mp \log x}.$$

(We use here that $x/mp > L/p > L^{1/2}$ and $\log(L^{1/2}) \gg \log x/\log_4 x$.) On the other hand, if $p \mid s(m)$, then the congruence $Ps(m) \equiv -\sigma(m) \pmod{p}$ has integer solutions $P$ only when $p \mid \sigma(m)$, in which case $p \mid \sigma(m) - s(m) = m$. In that scenario, every prime $P$ satisfies $Ps(m) \equiv -\sigma(m) \pmod{p}$. Since there are $\ll \frac{x \log_4 x}{m \log x}$ primes $P \leq x/m$, we conclude that

$$\sum_{1 < m < x/L} \sum_{L_m < P \leq x/m} \sum_{Ps(m) \equiv -\sigma(m) \pmod{p}} 1 \ll \sum_{m \leq x} \frac{x \log_4 x}{m \log x} + \sum_{m \leq x} \frac{x \log_4 x}{mp \log x} \ll \frac{x \log_4 x}{p}.$$

Keeping in mind that $|\Omega| \sim x$,

$$\mathbb{E}[X^{(s)}] \ll \frac{1}{x} \sum_{p \leq y} \frac{x \log_4 x}{p} \ll \log_4 x \log_2 y \ll \log_4 x \log_3 x. \quad \square$$

3. Completion of the proof of Theorem 1: Proof of Proposition 3

Throughout this section, $k$ is a fixed positive integer. All estimates are to be understood as holding for $x$ large enough, allowed to depend on $k$, and implied constants in Big-oh relations and $\ll$ symbols may depend on $k$. 
Recalling the definitions of \( \tilde{X}, \tilde{Y} \) and expanding,
\[
\mathbb{E}[\tilde{X}^k] - \mathbb{E}[\tilde{Y}^k] = \frac{1}{\sigma^k} \sum_{j=1}^{k} \binom{k}{j} (-\mu)^{k-j} (\mathbb{E}[X^j] - \mathbb{E}[Y^j])
\]
\[
\ll (\log_2 x)^O(1) \sum_{j=1}^{k} |\mathbb{E}[X^j] - \mathbb{E}[Y^j]|.
\]
For each \( j = 1, 2, \ldots, k \),
\[
\mathbb{E}[X^j] - \mathbb{E}[Y^j] = \sum_{p_1, \ldots, p_j \in \mathcal{P}} (\mathbb{E}[X_{p_1} \cdots X_{p_j}] - \mathbb{E}[Y_{p_1} \cdots Y_{p_j}]).
\]
Writing \( d \) for the product of the distinct primes from the list \( p_1, \ldots, p_j \), we have \( X_{p_1} \cdots X_{p_j} = \prod_{p \mid d} X_p, Y_{p_1} \cdots Y_{p_j} = \prod_{p \mid d} Y_p \), and
\[
\mathbb{E}[X_{p_1} \cdots X_{p_j}] - \mathbb{E}[Y_{p_1} \cdots Y_{p_j}] = \frac{1}{|\Omega|} \sum_{n \in \Omega} \frac{1}{d} - \frac{1}{d}.
\]
Observe that given \( d \) and \( j \), there are only \( O(1) \) possibilities for the original list \( p_1, \ldots, p_j \). Since there are \( O(1) \) possibilities for \( j \), we conclude that
\[
\mathbb{E}[\tilde{X}^k] - \mathbb{E}[\tilde{Y}^k] \ll (\log_2 x)^O(1) \sum_{\substack{\text{squarefree} \\ \omega(d) \leq k}} \frac{1}{|\Omega|} \sum_{n \in \Omega} \frac{1}{d} - \frac{1}{d}.
\]
We will show that
\[
(3) \quad \sum_{\substack{\text{squarefree} \\ \omega(d) \leq k}} \frac{1}{|\Omega|} \sum_{n \in \Omega} \frac{1}{d} - \frac{1}{d} \ll (\log_2 x)^O(1) \frac{\log x}{\log x}.
\]
Hence, \( \mathbb{E}[\tilde{X}^k] - \mathbb{E}[\tilde{Y}^k] \to 0 \) as claimed.

Let \( d \) be a product of at most \( k \) distinct primes from \( \mathcal{P} \). It will be useful in subsequent arguments to keep in mind that \( d = x^{O(1/\log_4 x)} \), and so is of size \( L^{o(1)} \). Decomposing each \( n \in \Omega \) in the form \( mP \), as in the proof of Lemma 4, we see that
\[
(4) \quad \sum_{n \in \Omega} \frac{1}{d} = \sum_{1 < m \leq x/L} \sum_{\substack{1 \leq s \leq x/m \\ P(s(m)) \equiv \sigma(m) \equiv 0 \pmod{d}}} 1,
\]
where as before \( L = x^{1/\log_4 x} \) and \( L_m = \max\{x^{1/\log_4 x}, P^+(m)\} \). To analyze the right-hand double sum, we consider various cases for \( m \).

Say that \( m \) is \textit{d-compatible} if for every prime \( p \) dividing \( d \), either \( p \) divides both \( s(m) \) and \( \sigma(m) \) or \( p \) divides neither. Then \( m \) is \textit{d-compatible} precisely when the congruence \( u s(m) + \sigma(m) \equiv 0 \pmod{d} \) has a solution \( u \) coprime to \( d \); in this case, the primes \( P \) with \( P s(m) + \sigma(m) \equiv 0 \pmod{d} \) are precisely those belonging to a certain coprime residue class modulo \( d/(d, s(m)) \). We call \( m \) \textit{d-ideal} if \( \gcd(d, s(m) \sigma(m)) = 1 \); equivalently, \( m \) is \textit{d-ideal} if \( m \) is \textit{d-compatible} and \( \gcd(d, s(m)) = 1 \). Note that only \textit{d-compatible} values of \( m \) contribute to the right side of \( (4) \).

When \( m \) is \( d \)-ideal,
\[
\sum_{L_m < P \leq x/m \atop Ps(m)+\sigma(m) \equiv 0 \pmod{d}} 1 = \frac{1}{\varphi(d)} \sum_{L_m < P \leq x/m \atop \text{not } d\text{-ideal}} 1 + O(E(x/m; d)),
\]
where
\[
E(T; q) := \max_{2 \leq t \leq T} \max_{\gcd(a, q) = 1} \left| \pi(t; q, a) - \frac{\pi(t)}{\varphi(q)} \right|.
\]
So the contribution to the right-hand side of (4) from \( d \)-ideal \( m \) is
\[
\left( \frac{1}{\varphi(d)} - \frac{1}{\varphi(d)} \right) \sum_{1 < m < x/L \atop \text{not } d\text{-ideal}} 1 + O \left( \sum_{m < x/L} E(x/m; d) \right).
\]
Since \( d \) is a product of \( O(1) \) primes all of which exceed \( y \), the first main term here admits the estimate
\[
\left( \frac{|\Omega|}{\varphi(d)} \right) = \frac{|\Omega|}{d} \left( 1 + O(1/y) \right) = \frac{|\Omega|}{d} + O(x/ dy).
\]
We bound the second main term, involving the double sum on \( m, P \), from above. The inner sum is no more than \( \pi(x/m) \ll \frac{x}{m \log(x/m)} \ll \frac{x \log x}{m \log x} \), so that
\[
\frac{1}{\varphi(d)} \sum_{1 < m < x/L \atop \text{not } d\text{-ideal}} 1 \ll \frac{x \log x}{\log x} \sum_{1 < m < x/L \atop \text{not } d\text{-ideal}} \frac{1}{md}.
\]
Next, we investigate the contribution to the right-hand side of (4) from \( m \) that are \( d \)-compatible but not \( d \)-ideal. For these \( m \), the corresponding primes \( P \) are restricted to a progression mod \( d/(d, s(m)) \), and so by the Brun–Titchmarsh inequality these \( m \) contribute
\[
\ll \sum_{1 < m < x/L \atop \text{d-Compat} \atop \text{not } d\text{-ideal}} \frac{x}{m \cdot \varphi(d/(d, s(m))) \log(x(d, s(m))/md)} \ll \frac{x \log x}{\log x} \sum_{1 < m < x/L \atop \text{d-Compat} \atop \text{not } d\text{-ideal}} \frac{(d, s(m))}{md}.
\]
We derive from (5), (6), (7), and (8) that
\[
\left| \frac{1}{|\Omega|} \sum_{n \in \Omega \atop d \mid s(n)} 1 - \frac{1}{d} \right| \ll \frac{1}{x} \sum_{m < x/L} |E(x/m; d)| + \frac{1}{dy} + \frac{\log x}{\log x} \sum_{1 < m < x/L \atop \text{not } d\text{-ideal}} \frac{1}{md} + \log x \sum_{1 < m < x/L \atop \text{d-Compat} \atop \text{not } d\text{-ideal}} \frac{(d, s(m))}{md}.
\]
Now we sum on \( d \).
First off, the Bombieri–Vinogradov theorem implies that

\[
\sum_{\substack{d \text{ squarefree} \\ p | d \Rightarrow p \in \mathcal{P} \\ \omega(d) \leq k}} \left( \frac{1}{x} \sum_{m < x/L} \left| E(x/m; d) \right| \right) \leq \frac{1}{x} \sum_{m < x/L} \sum_{d \leq (x/m)^{1/3}} \left| E(x/m; d) \right|
\]

\[
\ll \frac{1}{x} \sum_{m < x/L} \frac{x/m}{(\log (x/m))^2} \ll \frac{(\log_4 x)^2}{(\log x)^2} \sum_{m < x/L} \frac{1}{m} \ll \frac{(\log_4 x)^2}{\log x}.
\]

Next,

\[
\sum_{\substack{d \text{ squarefree} \\ p | d \Rightarrow p \in \mathcal{P} \\ \omega(d) \leq k}} \frac{1}{dy} \leq \frac{1}{y} \sum_{j=0}^k \frac{1}{j!} \left( \sum_{p \in \mathcal{P}} \frac{1}{p} \right)^j \ll \frac{(\log_2 x)^k}{(\log x)^2}.
\]

Continuing, note that if \( m \) is not \( d \)-ideal, then there is a prime \( p | d \) with \( p | s(m) \sigma(m) \). Hence,

\[
\sum_{\substack{d \text{ squarefree} \\ p | d \Rightarrow p \in \mathcal{P} \\ \omega(d) \leq k}} \left( \frac{\log_4 x}{\log x} \sum_{1 < m < x/L} \frac{1}{md} \right) \leq \frac{\log_4 x}{\log x} \sum_{1 < m < x/L} \frac{1}{m} \sum_{d \text{ squarefree} \atop p | d} \sum_{\omega(d) \leq k} \frac{1}{d} \ll \frac{(\log_2 x)^{O(1)}}{\log x} \sum_{1 < m < x/L} \frac{1}{m} \sum_{p \in \mathcal{P}} \frac{1}{p}.
\]

Since each \( p \in \mathcal{P} \) exceeds \( y \), the final sum on \( p \) is \( \ll \omega(s(m) \sigma(m))/y \ll \log x/y = 1/\log x \), and so the last displayed expression is

\[
\ll \frac{(\log_2 x)^{O(1)}}{(\log x)^2} \sum_{1 < m < x/L} \frac{1}{m} \ll \frac{(\log_2 x)^{O(1)}}{\log x}.
\]

Finally, suppose \( m \) is \( d \)-compatible but not \( d \)-ideal. Then \( (d, s(m)) > 1 \), \( (d, s(m)) | \sigma(m) \), and \( (d, s(m)) | \sigma(m) - s(m) = m \). Hence, thinking of \( d' \) as \( (d, s(m)) \),

\[
\sum_{\substack{d \text{ squarefree} \\ p | d \Rightarrow p \in \mathcal{P} \\ \omega(d) \leq k}} \left( \frac{\log_4 x}{\log x} \sum_{1 < m < x/L} \frac{(d, s(m))}{md} \right) \leq \frac{\log_4 x}{\log x} \sum_{1 < m < x/L} \frac{1}{m} \sum_{d \text{ squarefree} \atop d' | d} \sum_{d' \text{d'COMPAT} \atop d' > 1} \sum_{\omega(d) \leq k} \frac{1}{d'} \frac{1}{d' \omega(m)}.
\]

Let us estimate the inner sum on \( m \). Write \( m = d'm' \). The contribution to that sum from \( m \) with \( (d', m') > 1 \) is at most

\[
\sum_{p | d'} \frac{1}{d'} \sum_{m' < x \atop p | m'} \frac{1}{m} \ll \frac{\log x}{d'} \sum_{p | d'} \frac{1}{p} \ll \frac{\log x}{d'} \log x \omega(d') \ll \frac{1}{d' \log x}.
\]

Suppose now that \( \gcd(d', m') = 1 \). If \( d' | s(m) \), then \( P^+(d') | \sigma(d') s(m') \), while \( P^+(d') > P^+(\sigma(d')) \) (since \( d' \) is a squarefree product of odd primes and \( d' > 1 \)). Thus, \( P^+(d') | \sigma(m') \). Choose a prime power \( q^e \parallel m' \) with \( P^+(d') | \sigma(q^e) \). If \( e \geq 2 \), then \( y < P^+(d') \leq \sigma(q^e) < 2q^e \),
and so \( m' \) has squarefull part exceeding \( y/2 \). If \( e = 1 \), then \( q \parallel m' \) with \( q \equiv -1 \pmod{P^+(d')} \). Hence, these \( m \) make a contribution to the inner sum bounded by

\[
\frac{1}{d'} \left( \sum_{r > y/2 \text{ squarefree}} \frac{1}{m'} + \sum_{p|d'} \sum_{q < x \text{ prime}} \sum_{m'|m} \frac{1}{m'} \right) \ll \frac{\log x}{d'} \left( \sum_{r > y/2 \text{ squarefull}} \frac{1}{r} + \sum_{p|d'} \sum_{q < x \text{ prime}} \frac{1}{q} \right)
\]

\[
\ll \frac{\log x}{d'} \left( \frac{1}{\log x} + \sum_{p|d'} \frac{\log x}{p} \right) = \frac{1}{d'} + \frac{(\log x)^2}{d'} \sum_{p|d'} \frac{1}{p} \ll \frac{1}{d'} + \frac{(\log x)^2}{d'} \frac{1}{y} \ll \frac{1}{d'}.
\]

Inserting these estimates back above, the right-hand side of (9) is seen to be

\[
\ll \frac{\log_4 x}{\log x} \sum_{d \text{ squarefree}} \frac{1}{d} \sum_{d'|d, d'>1} 1 \ll \frac{\log_4 x}{\log x} \sum_{d \text{ squarefree}} \frac{1}{d} \ll \frac{(\log_3 x)^{O(1)}}{\log x}.
\]

Assembling the last several estimates yields (3), which completes the proof of Theorem 1.

Remark. As with most variants of Erdős–Kac, Theorem 1 remains valid if we count prime factors with multiplicity. Define \( \omega'(n) = \sum_{p^k \mid n} k \). (We avoid the more familiar notation \( \Omega(n) \), since \( \Omega \) denotes our sample space.) It is shown in [Tro15] that, for a certain subset \( \Omega' \) of \( (1, x] \) containing \( (1 + o(1))x \) elements,

\[
\frac{1}{x} \sum_{n \in \Omega'(x)} (\omega'(s(n)) - \omega(s(n))) \ll (\log x)^2.
\]

(See p. 133 of [Tro15].) It follows that away from a set of \( o(x) \) elements of \( (1, x] \), we have \( \omega'(s(n)) - \omega(s(n)) < (\log \log x)^{0.49} \) (say). Hence, the Erdős–Kac theorem for \( \omega'(s(n)) \) is a consequence of the corresponding theorem for \( \omega(s(n)) \).

4. Other arithmetic functions

The astute reader will observe that many of the calculations above do not depend on properties specific to \( s(n) \). In this section, we discuss how to adapt the previous argument for other arithmetic functions.

Let \( f \) be an integer-valued arithmetic function with \( f(n) \) nonzero for \( n > 1 \) and \( |f(n)| \leq x^{O(1)} \) for all \( n \leq x \). Assume that for all positive integers \( m \) and all primes \( P \) not dividing \( m \), there are integers \( a(m) \) and \( b(m) \) such that \( f(mP) = Pa(m) + b(m) \), with \( a(m), b(m) \) nonzero for \( m > 1 \). Finally, assume that \( |a(m)|, |b(m)| \leq x^{O(1)} \) for all \( 1 < m \leq x \). (For \( f(n) = s(n) \), we have \( 0 < s(n) \leq x^2 \) when \( 1 < n \leq x \), and \( s(mP) = Ps(m) + \sigma(m) \) for any positive integer \( m \) and any prime \( P \nmid m \).) All symbols are defined as in Section 2, except that the random variable \( X_p \) is now equal to 1 if \( p \mid f(n) \) and is 0 otherwise.

To obtain an Erdős–Kac-type result for \( \omega(f(n)) \), we follow the same general strategy as in the case \( f(n) = s(n) \). By the method of moments, Lemma 2 and the analogue of Proposition 3 (once shown) will establish that

\[
\bar{X} = \frac{\bar{S} - \mu}{\sigma}
\]

converges in distribution to the standard normal.
Recall that $y = (\log x)^2$ and $z = x^{1/\log_3 x}$; then

$$\frac{\omega(f(\cdot)) - \mu}{\sigma} = \bar{X} + \frac{X(s)}{\sigma} + \frac{X(l)}{\sigma},$$

where $X(s) = \sum_{p \leq y} X_p$ and $X(l) = \sum_{z < p \leq x^c} X_p$, where $c > 0$ is a constant such that $|f(n)| \leq x^c$ for all $n \leq x$.

As before, our task is to show that $\frac{X(s)}{\sigma}$ and $\frac{X(l)}{\sigma}$ converge to 0 in probability. The argument for $X(l)$ is the same, with the exponent 2 replaced by $c$. For $X(s)$, we again hope to use Markov’s inequality coupled with an upper bound for $\mathbb{E}[X(s)]$ of size $o(\sqrt{\log_2 x})$, analogous to Lemma 4. The argument there yields, in this case,

$$\mathbb{E}[X(s)] \ll \log_3 x \log_4 x + \frac{\log_4 x}{\log x} \sum_{p \leq y} \sum_{m \leq x} \frac{1}{m}.$$

Thus, the aim is to show that

$$\sum_{p \leq y} \sum_{p \mid a(m) \text{ and } p \mid b(m)} \frac{1}{m} = o\left(\frac{\sqrt{\log_2 x}}{\log_4 x} \log x\right).$$

We now turn our attention to the analogue of Proposition 3. Say that $m$ is $d$-compatible if for every $p \mid d$, either $p$ divides both $a(m)$ and $b(m)$ or $p$ divides neither; and $m$ is $d$-ideal if $\gcd(d, a(m)b(m)) = 1$. Equivalently, $m$ is $d$-ideal if $m$ is $d$-compatible and $\gcd(d, a(m)) = 1$. Tracing through the argument in Section 3, we see that few of the calculations depend on specific properties of $f(n)$; in fact, the analogue of Proposition 3 is established if

$$\sum_{d \text{ squarefree}} \sum_{\substack{1 < m < x \mid \text{d-ideal} \text{ and } \omega(d) \leq k}} \frac{(d, a(m))}{md} \ll (\log_2 x)^{O(1)}.$$ 

We summarize the above discussion in the following proposition.

**Proposition 5.** Suppose $f(n)$ is an integer-valued arithmetic function with $f(n)$ nonzero when $n > 1$ and $|f(n)| \leq x^{O(1)}$ for all $n \leq x$. Suppose also that for every positive integer $m$, there are $a(m)$ and $b(m)$ such that $f(mP) = Pa(m) + b(m)$ for all primes $P$ not dividing $m$,

and that $|a(m)|, |b(m)| \leq x^{O(1)}$ whenever $m \leq x$.

Suppose also that $a(m), b(m)$ are nonzero whenever $m > 1$. Then, if

$$\sum_{p \leq y} \sum_{p \mid a(m) \text{ and } p \mid b(m)} \frac{1}{m} = o\left(\frac{\sqrt{\log_2 x}}{\log_4 x} \log x\right)$$

(10)
Theorem 1 is true with $f$ in place of $s(n)$.

4.1. The sum of prime divisors. For each positive integer $n$, let $\beta(n):=\sum_{p|n}p$ denote the sum of the prime divisors of $n$. If $1 < n \leq x$, then $0 < \beta(n) \leq n \leq x$. If $P$ is a prime not dividing the integer $m$, then

$$\beta(mP) = P + \beta(m).$$

We apply Proposition 5, with $f(n) = \beta(n)$, $a(m) = 1$, and $b(m) = \beta(m)$. Since $a(m) = 1$, one quickly observes that the sums on the left-hand sides of (10) and (11) are empty. Thus, Theorem 1 holds with $\beta(n)$ in place of $s(n)$. The same argument applies, verbatim, with $\beta(n)$ replaced by $A(n) = \sum_{p|n} kP$, where prime factors are summed with multiplicity. For other work on the value distribution of $\beta(n)$ and $A(n)$, see [Hal70, Hal71, Hal72, AE77, Pol14, Gol17].

4.2. A shifted divisor function. Let $f(n) = n + \tau(n)$, where $\tau(n)$ denotes the number of divisors of $n$. Then if $n \leq x$, $f(n) < x^{O(1)}$ trivially. If $P$ is a prime not dividing the positive integer $m$, then

$$f(mP) = mP + \tau(mP) = Pm + 2\tau(m),$$

so $a(m) = m$ and $b(m) = 2\tau(m)$ in this case. For $m \leq x$, the largest exponent appearing in the prime factorization of $m$, and hence the largest prime divisor of $\tau(m)$, is $\ll \log x$. This means there is no value of $m$ that is $d$-compatible but not $d$-ideal, since every prime $p \mid d$ satisfies $p > (\log x)^2$. Equation (11) is therefore satisfied, since the sum is empty.

Equation (10) is handled nearly as easily. Ignoring the condition $p \mid b(m)$ in the inner sum, the left-hand side of (10) is at most

$$\sum_{p \leq y} \frac{1}{m} \ll \log x \sum_{p \leq y} \frac{1}{p} \ll \log x \log_3 x = o\left(\frac{\sqrt{\log x}}{\log_4 x} \log x\right),$$

as desired. Thus, by Proposition 5, Theorem 1 holds with $f(n) = n + \tau(n)$ in place of $s(n)$. Similar arguments apply to $n - \tau(n)$ and $n \pm \omega(n)$. The functions $n - \tau(n)$ and $n - \omega(n)$ appear in work of Luca [Luc05]; for each of these two functions, he shows that the range is missing infinitely many positive integers.

4.3. The cototient function. Let $f(n) = n - \varphi(n)$, where (as above) $\varphi(n)$ is Euler’s totient function. (See [Erd73, BS95, FL00, GM05, PY14, PP16] for studies of the range of $n - \varphi(n)$.) Note that $0 < f(n) < n$ for $n > 1$ and, if $P$ is a prime not dividing $m$,

$$f(mP) = Pm - \varphi(Pm) = Pm - (P - 1)\varphi(m) = P(m - \varphi(m)) + \varphi(m).$$

We apply Proposition 5 with $f(n) = n - \varphi(n)$, $a(m) = m - \varphi(m)$, and $b(m) = \varphi(m)$. We first observe that equation (10) can be established as in (12), noting that if $p \mid a(m)$ and $p \mid b(m)$, then $p \mid a(m) + b(m) = m$. To show (11), use the argument surrounding (9), replacing $s(m)$ by $a(m) = m - \varphi$ and $\sigma(m)$ by $b(m) = \varphi(m)$. The argument carries through with only the
Sums of Proper Divisors Follow Erdős–Kac

... of modifications. By Proposition 5, Theorem 1 holds with \( f(n) = n - \varphi(n) \) in place of \( s(n) \).

Several other applications of our method could be given, although some require slight changes to the framework. For example, fix an integer \( a \neq 0 \) and consider the shifted totient function \( f(n) = \varphi(n) + a \). It is not hard to prove that the hypotheses of Proposition 5 are satisfied with \( a(m) = \varphi(m) \) and \( b(m) = -\varphi(m) + a \), with one exception: If \( a > 0 \) is in the range of \( \varphi \), then \( b(m) \) will vanish at some \( m > 1 \). However, it is still true that \( b(m) \) is nonvanishing for all \( m > m_0(a) \), and one can simply run our argument with the condition \( n/P^+(n) > m_0(a) \) added to the definition of \( \Omega \).

References

[AE77] K. Alladi and Erdős, P., *On an additive arithmetic function*, Pacific J. Math. 71 (1977), 275–294.
[BHS05] W.D. Banks, G. Harman, and I.E. Shparlinski, *Distributional properties of the largest prime factor*, Michigan Math. J. 53 (2005), 665–681.
[Bil69] P. Billingsley, *On the central limit theorem for the prime divisor functions*, Amer. Math. Monthly 76 (1969), 132–139.
[Bil95] ________, *Probability and measure*, third ed., Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York, 1995.
[BS95] J. Browkin and A. Schinzel, *On integers not of the form \( n = \sigma(f(n)) \)*, Acta Arith. 71 (1995), 55–58.
[Dav33] H. Davenport, *Über numeri abundantes*, S.-Ber. Preuß. Akad. Wiss., math.-nat. Kl. (1933), 830–837.
[DK40] P. Erdős and M. Kac, *The Gaussian law of errors in the theory of additive number theoretic functions*, Amer. J. Math. 62 (1940), 738–742.
[Erd73] P. Erdős, *Über die Zahlen der Form \( \sigma(n) - n \) und \( n - \varphi(n) \)*, Elem. Math. 28 (1973), 83–86.
[FL00] A. Flammenkamp and F. Luca, *Infinite families of noncototients*, Colloq. Math. 86 (2000), 37–41.
[GM05] A. Grytczuk and B. Medryk, *On a result of Flammenkamp-Luca concerning noncototent sequence*, Tsukuba J. Math. 29 (2005), 533–538.
[Gol17] D. Goldfeld, *On an additive prime divisor function of Alladi and Erdős*, Analytic number theory, modular forms and \( q \)-hypergeometric series, Springer Proc. Math. Stat., vol. 221, Springer, Cham, 2017, pp. 297–309.
[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.
[Hal71] ________, *On the probability that \( n \) and \( f(n) \) are relatively prime. II*, Acta Arith. 19 (1971), 175–184.
[Hal72] ________, *On the probability that \( n \) and \( f(n) \) are relatively prime. III*, Acta Arith. 20 (1972), 267–289.
[HR00] G.H. Hardy and S. Ramanujan, *The normal number of prime factors of a number \( n \)* [Quart. J. Math. 48 (1917), 76–92], Collected papers of Srinivasa Ramanujan, AMS Chelsea Publ., Providence, RI, 2000, pp. 262–275.
[LLPSR] N. Lebowitz-Lockard, P. Pollack, and A. Singha Roy, *Distribution mod \( p \) of Euler’s totient and the sum of proper divisors*, submitted.
[LP15] F. Luca and C. Pomerance, *The range of the sum-of-proper-divisors function*, Acta Arith. 168 (2015), 187–199.
[Luc05] F. Luca, *On numbers not of the form \( n - \omega(n) \)*, Acta Math. Hungar. 106 (2005), 117–135.
[Pol14] P. Pollack, *Some arithmetic properties of the sum of proper divisors and the sum of prime divisors*, Illinois J. Math. 58 (2014), 125–147.
[Pom18] C. Pomerance, *The first function and its iterates*, Connections in discrete mathematics, Cambridge Univ. Press, Cambridge, 2018, pp. 125–138.
[PP16] P. Pollack and C. Pomerance, *Some problems of Erdős on the sum-of-divisors function*, Trans. Amer. Math. Soc. Ser. B 3 (2016), 1–26.
[PPT18] P. Pollack, C. Pomerance, and L. Thompson, *Divisor-sum fibers*, Mathematika 64 (2018), 330–342.
[PY14] C. Pomerance and H.-S. Yang, *Variant of a theorem of Erdős on the sum-of-proper-divisors function*, Math. Comp. 83 (2014), 1903–1913.
[Tro15] L. Troupe, On the number of prime factors of values of the sum-of-proper-divisors function, J. Number Theory 150 (2015), 120–135.
[Tro20] , Divisor sums representable as the sum of two squares, Proc. Amer. Math. Soc. 148 (2020), 4189–4202.

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

Email address: pollack@uga.edu

Department of Mathematics, Mercer University, 1501 Mercer University Drive, Macon, GA 31207

Email address: troupe_lt@mercer.edu