

k -Parent Aliquot Numbers: Counting Pre-Images Under the Sum-of-Proper-Divisors Function

Gavin Guinn

Supervisor: Michael J. Jacobson

December 2021

1 Introduction

If we examine numbers in the set $\{6, 28, 496, 8128, \dots\}$ we notice that every member n is equal to the sum of its divisors with the exclusion of itself,

$$1 + 2 + 3 = 6.$$

This of course is the definition of a *perfect number*, a property that has fascinated number theorists since antiquity. The *sum-of-divisors* function is defined for $n > 1$ by

$$\sigma(n) = \sum_{d|n} d.$$

The *proper-divisors* of n refers to the subset of the divisors which excludes itself, the *sum-of-proper-divisors* is defined for $n > 1$ by

$$s(n) = \sigma(n) - n,$$

thus n is a perfect number if and only if $s(n) = n$. The sum-of-proper-divisors is a natural way to consider perfect numbers but is certainly not limited to that domain. For instance consider $s(9) = 4$ which has the property $n < s(n)$, such numbers are *deficient* while numbers that satisfy $n > s(n)$ are *abundant*.

Aliquot sequences are produced by iterating $s(n)$, for example consider the sequence starting at 12:

$$s(12) = 1 + 2 + 3 + 4 + 6$$

$$s(16) = 1 + 2 + 4 + 8$$

$$s(15) = 1 + 3 + 5$$

$$s(9) = 1 + 3$$

$$s(4) = 1 + 2$$

$$s(3) = 1$$

In this case the sequence has terminated as $s(n)$ is only defined for $n > 1$, this sequence is also bounded as no value is greater than 16. A sequence does not have to terminate to be bounded, since perfect numbers are defined by $s(n) = n$ an aliquot sequence that reaches a perfect number will cycle. The members of a cycle of length 2 are referred to as an *amicable pair*, for instance

$$s(1184) \rightarrow s(1210) \rightarrow s(1184) \rightarrow \dots$$

It seems that cycles can form with arbitrary length [3] note a cycle of length 24, numbers in such cycles are *sociable*. Aliquot sequences are sometimes said to terminate in a cycle, we will only refer to aliquot sequences that reach 1 to have terminated while referring to the latter case as a bounded sequence. Aliquot sequences are formalized by [1] in the following way:

Definition 1.1. The aliquot sequence starting at $a_0 > 1$ is defined by

$$a_0, a_1 = s(a_0), a_2 = s^2(a_0), \dots$$

Proposition 1.1. An aliquot sequence is bounded if and only if

$$\exists a_n, \forall a_m \text{ s.t. } a_n \geq a_m$$

The aliquot sequence of 276,

$$s(276) \rightarrow s(396) \rightarrow s(696) \rightarrow s(1104) \rightarrow s(1872) \rightarrow \dots$$

is the smallest example that has not been shown to be bounded, [11] has followed the it's growth over 2090 iterations to values with over 200 digits! The dichotomy between the seemingly unbounded growth of the 276 sequence and the boundedness of the 12 sequence gets to the most important open question relating to aliquot sequences:

Conjecture 1.1 (Catalan-Dickson). All aliquot sequences are bounded and thus must terminate or enter a cycle. [2]

Conjecture 1.2 (Guy-Selfridge). The Catalan-Dickson conjecture is false, perhaps most aliquot sequences are unbounded. [6]

The computational exploration of these sequences that has taken place since Catalan originally proposed his conjecture in 1888 has certainly given credence to the Guy-Selfridge point-of-view. However a trend in the data is far from a proof, under our current understanding it is entirely possible that the 276 sequence could trip over a prime or enter a cycle on some future iteration.

1.1 Motivation

What statistics can be compiled on the behaviour of $s(n)$ to argue for or against the Calatan-Dickson Conjecture? The authors of [3] consider numerical and analytical approaches to determine the average of $s(n)/n$. The average order of an arithmetic function is a simple function that has the same behaviour on average as the less tractable function. Define $\mathcal{S}(n)$ to be the average order of $s(n)$, when computing the average of $s(n)/n$ we are numerically estimating $\mathcal{S}(n)$. When considering this metric [3] restrict the domain to evens to examine $\mathcal{S}(2n)$. I believe the impetus for this decision can be traced to [6] who heuristically argue that unbounded sequences exist:

- **if** $\mathcal{S}(2n) > 2n$ and,
- **if** other effects on the growth of a sequence are relatively minor compared to the tendency for aliquot sequences to preserve parity,
- **then** we would expect most even aliquot sequences that reach large values to be unbounded.

This argument seems to match the numerical evidence as seemingly unbounded sequences such as $s(276)$ [11] appear to maintain constant even parity.

The method used to either analytically determine or numerically compute $\mathcal{S}(2n)$ impacts the parameters value. While not explicitly stated by Guy and Selfridge, [1] indicate that arithmetic mean of $s(2n)/2n$ was used to determine $\mathcal{S}(2n)$ in constructing their heuristic argument for unboundedness.

Theorem 1.1. Arithmetic Mean of $s(2n)/2n$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \frac{s(2n)}{2n} = \left(\frac{5\pi^2}{24} - 1 \right) = 1.0562...$$

With this metric we have $\mathcal{S}(2n) = (2n)(1.0563...)$ and thus $\mathcal{S}(2n) > 2n$. However as pointed out by [1] the sequence $\{5, \frac{1}{5}, 5, \frac{1}{5}, \dots\}$ is bounded but its arithmetic mean tends to 2.6, a better metric for evaluating the boundedness of a sequence is to consider the geometric mean.

Theorem 1.2. Geometric Mean of $s(2n)/2n$

$$\mu = \lim_{N \rightarrow \infty} \exp \left(\frac{1}{N} \sum_{n=1}^N \log s(2n)/2n \right) = 0.969...$$

With this more accurate metric we have $\mathcal{S}(2n) < 2n$ which suggests that even aliquot sequences will in fact decrease on average! This result certainly throws cold water on the heuristic argument of Guy and Selfridge. However the monumental growth of sequences like $s(276)$ suggests that this metric does not completely describe sequence behaviour.

To improve this model of sequence behaviour the authors of [3] consider the image of $s(2n)$ when computing $\mathcal{S}(2n)$. Consider the solutions to $s^{-1}(316)$,

$$s(192) = s(304) = s(344) = s(412) = 316,$$

since multiple sequences are tributary at 316 it seems reasonable to conclude that the value of $s(316)/316$ impacts average sequence growth more than numbers with fewer pre-images. [3] explores this by constructing an average of $s(2n)/2n$ weighted by $\#s^{-1}(2n)$. Counting the number of pre-images under $s(2n)$ was certainly on the mind of Dr. Guy when he posed the question that motivates this project:

Think of a number!! Say 36%, which is nice and divisible. It appears that about 36% of the even numbers are "orphans".

Divide by 1. For about 36% of the (even) values of n there is just one positive integer m such that $s(m) = n$. These values of n have just one "parent".

Divide by 2. About 18% of the even values of n have exactly two parents.

Divide by 3. About 6% of the even values of n have three parents.

Divide by 4. About 1.5% of the even values of n have just 4 parents.

This suggests that $1/(e \cdot p!)$ of the even numbers have p parents.

Experiments suggests that these values are a bit large for small values of p and a bit small for larger values of p . Can anything be proved?

Definition 1.2. n is a k -parent aliquot number if $s^{-1}(n)$ has k solutions.

These percentages refer to the density of even k -parent numbers or the probability of encountering a k -parent number when choosing randomly from all even natural numbers. There are a suite of colourful names to refer to numbers outside of the range of $s(\cdot)$ including *non-aliquots* and *untouchable numbers*, for the sake of consistency I will be referring to all such numbers as *non-aliquot*.

2 Modeling the Density of k-Parent Numbers

2.1 Pollack-Pomerance Model of Non-Aliquot Density

Consider the following set,

$$N_a = \{2, 5, 52, 88, 96, 120, 124, 146, 162, 188, 206, 210, 216, 238, 246, 248, \dots\}$$

which contains non-aliquot numbers [10], if $n \in N_a$ then $\nexists m$ s.t. $s(m) = n$, in symbols $\#s^{-1}(n) = 0$. The name non-aliquot implies that these numbers do

not appear in aliquot sequences, however every sequence begins with a member of N_a ; if $\#s^{-1}(n) > 0$ then there must be some preceding number in the same sequence.

We are interested in the natural density of N_a ,

$$d(N_a) = \lim_{n \rightarrow \infty} \frac{|N_a \cap [1, n]|}{n},$$

a method of determining the proportion of numbers that are non-aliquot. Using probabilistic methods [8] precisely model the density of non-aliquot numbers.

Conjecture 2.1. *The set of non-aliquot numbers has asymptotic density Δ ,*

$$\Delta = \lim_{y \rightarrow \infty} \frac{1}{\log y} \sum_{\substack{a \leq y \\ 2|a}} \frac{1}{a} e^{-a/s(a)}.$$

This model is an application of the "balls into bins" statistics problem, estimating the result after throwing m balls randomly into n bins;

$$\#[\text{empty bins}] = n(1 - \frac{1}{n})^m.$$

To apply this technique the authors begin by defining a series of disjoint subsets,

$$A_y = \text{lcm}[1, 2, \dots, y]$$

and for each divisor $a|A_y$ uniquely define a set T_a ,

$$T_a = \{n : \gcd(n, A_y) = a\}.$$

For example consider the case $A_3 = 6$, we have $a \in \{1, 2, 3, 6\}$ and the following sets T_a :

a	T_a
1	$\{1, 5, 7, 11, 13, 17, 19, \dots\}$
2	$\{2, 4, 8, 14, 16, 20, 22, \dots\}$
3	$\{3, 9, 15, 21, 27, 33, 39, \dots\}$
6	$\{6, 12, 18, 24, 30, 36, 42, \dots\}$

Define bounds on the set,

$$T_a(x) = T_a \cap [1, x].$$

Utilizing the asymptotic properties of T_a the authors set up assumptions needed to apply *balls in bins*:

- Assume that $s(\cdot)$ maps T_a to T_a for each $a|A_y$, asymptotically true if $y \rightarrow \infty$ and $T_a = \{n > e^{e^y} \mid \gcd(n, A_y) = a\}$.

- Assume $\forall n \in T_a$ that $\sigma(n)/n \approx \sigma(a)/a$, asymptotically true up to sets of vanishing density as $y \rightarrow \infty$.
- Assume that $s(\cdot)$ is a random map.

To apply this model we need values for the number of *bins* n and *balls* m ; naturally we use the bounded sets T_a as a analog for the number of *bins*,

$$n = |T_a(x)|.$$

For an *balls* m we need to find a range of T_a such that $s(\cdot)$ maps these numbers into $T_a(x)$. Utilizing our assumptions that that $s(\cdot)$ maps T_a to T_a and $\sigma(n)/n \approx \sigma(a)/a$ we can choose a value for m such that

$$\forall h \in T_a \leq m, s(h) \in T_a(x).$$

We are choosing m to be the largest value in the set T_a such that $s(\cdot)$ maps all values less or equal to m into $T_a(x)$. It is useful to observe that every member of a T_a by assumption shares the ratio $a/s(a)$,

$$\frac{a}{s(a)} = \frac{1}{(\sigma(a)/a) - 1}.$$

Utilizing this assumption we set our number of *balls* m to

$$m = |T_a(x \cdot (a/s(a)))|.$$

Plugging in values we have the following expression,

$$\left(1 - \frac{1}{n}\right)^m = \left(1 - \frac{1}{|T_a(x)|}\right)^{|T_a(x \cdot a/s(a))|}$$

which models the probability that some member of $T_a(x)$ is non-aliquot.

The authors note a useful asymptotic equivalence,

$$|T_a(x)| \sim \frac{\phi(A_y) \cdot x}{A_y \cdot a}.$$

Since we are interested in the asymptotic behavior we take the limit of the bound x and substitute in the equivalence,

$$\begin{aligned} \lim_{x \rightarrow \infty} \left(1 - \frac{1}{\left(\frac{\phi(A_y) \cdot x}{A_y \cdot a}\right)}\right)^{(\phi(A_y) \cdot x)/(A_y \cdot s(a))} \\ = e^{-a/s(a)}. \end{aligned}$$

This expression models the probability that a member of the unbounded set T_a is non-aliquot. An intuition for this simplification can be gained by considering a definition of Euler's number,

$$\lim_{x \rightarrow \infty} \left(1 - \frac{1}{x}\right)^x = e.$$

At this point we have a model for non-aliquot density within any set T_a , we would like to extend this to model non-aliquot density over all naturals. To accomplish this we again utilize the equivalence,

$$|T_a(x)| \sim \frac{\phi(A_y) \cdot x}{A_y \cdot a},$$

noticing that the ratio $\phi(A_y)/(a \cdot A_y)$ expresses the probability that some natural is in T_a , see [4] for a similar treatment. Supposing $n \in T_a$ and $s^{-1}(n)$ are independent events model the joint probability,

$$\mathbb{P}[T_a \text{ and non-aliquot}] = \frac{\phi(A_y)}{A_y \cdot a} \cdot e^{-a/s(a)}.$$

Since each set T_a is disjoint we can apply the additive property to find the natural density of non-aliquot numbers:

$$\begin{aligned} \Delta &= \lim_{y \rightarrow \infty} \sum_{\substack{a|A_y \\ 2|a}} \frac{\phi(A_y)}{a \cdot A_y} \cdot e^{-a/s(a)} \\ &= \lim_{y \rightarrow \infty} \frac{\phi(A_y)}{A_y} \sum_{\substack{a|A_y \\ 2|a}} \frac{1}{a} e^{-a/s(a)} \end{aligned}$$

We restrict a to even values as a 's parity determines the parity of the entire set T_a and we know that the density of odd aliquot orphans quickly vanishes

2.2 k-Parent Generalization

Note: Unedited copy from final report for NSERC workterm.

Luckily people interested in statistics have done the work for us, say we have n 'bins', m 'balls', and want to know the probability of a specific bin has exactly k balls.

$$\mathbb{P}[k \text{ balls in bin}] = \binom{m}{k} \frac{1}{n^k} \left(1 - \frac{1}{n}\right)^{m-k}$$

This gives a fairly simple strategy to generalize the Pollack/Pomerance heuristic model, simply swap out the probability of getting 0-balls in a bin for the general

equation. Plugging in the same values of:

$$n = \frac{\phi(A_y) \cdot x}{A_y \cdot a}$$

And:

$$m = \frac{\phi(A_y) \cdot x}{A_y \cdot s(a)}$$

We get the following monster:

$$\mathbb{P}[k \text{ parents}] = \lim_{x \rightarrow \infty} \binom{\frac{\phi(A_y) \cdot x}{A_y \cdot s(a)}}{k} \frac{1}{\left(\frac{\phi(A_y) \cdot x}{A_y \cdot a}\right)^k} \left(1 - \frac{1}{\left(\frac{\phi(A_y) \cdot x}{A_y \cdot a}\right)}\right)^{\frac{\phi(A_y) \cdot x}{A_y \cdot s(a)} - k}$$

This expression cleans up quite nicely (see section Probability Simplification), the probability of some $n \in T_a$ having k parents:

$$\mathbb{P}[k \text{ parents}] = \frac{a^k}{k! \cdot s(a)^k} \cdot e^{-a/s(a)}$$

We then simply slap this probability expression into the Pollack and Pomerance result giving the density of even k parent aliquot numbers over all naturals:

$$\Delta_k = \lim_{y \rightarrow \infty} \frac{\phi(A_y)}{A_y} \sum_{\substack{a|A_y \\ 2|a}} \frac{a^{k-1}}{k! \cdot s(a)^k} \cdot e^{-a/s(a)}$$

The authors continue to establish an easier to compute expression for Δ :

$$\Delta = \lim_{y \rightarrow \infty} \frac{\phi(A_y)}{A_y} \sum_{\substack{a|A_y \\ 2|a}} \frac{1}{a} e^{-a/s(a)} \quad (3.1)$$

$$= \lim_{y \rightarrow \infty} \frac{1}{\log y} \sum_{\substack{a \leq y \\ 2|a}} \frac{1}{a} e^{-a/s(a)} \quad (3.4)$$

This suggests that the density of k parent numbers can be simplified to:

$$\Delta_k = \lim_{y \rightarrow \infty} \frac{1}{\log y} \sum_{\substack{a \leq y \\ 2|a}} \frac{a^{k-1}}{k! \cdot s(a)^k} \cdot e^{-a/s(a)}$$

2.3 Open Work

1. Prove:

$$\begin{aligned}\Delta_k &= \lim_{y \rightarrow \infty} \frac{\phi(A_y)}{A_y} \sum_{\substack{a|A_y \\ 2|a}} \frac{a^{k-1}}{k! \cdot s(a)^k} \cdot e^{-a/s(a)} \\ &= \lim_{y \rightarrow \infty} \frac{1}{\log y} \sum_{\substack{a \leq y \\ 2|a}} \frac{a^{k-1}}{k! \cdot s(a)^k} \cdot e^{-a/s(a)}\end{aligned}$$

2. Given that:

$$\Delta_k = \lim_{y \rightarrow \infty} \frac{1}{\log y} \sum_{\substack{a \leq y \\ 2|a}} \frac{a^{k-1}}{k! \cdot s(a)^k} \cdot e^{-a/s(a)}$$

Estimates the density of **even** k parent numbers over all naturals, does:

$$\Delta_k = \lim_{y \rightarrow \infty} \frac{1}{\log y} \sum_{a \leq y} \frac{a^{k-1}}{k! \cdot s(a)^k} \cdot e^{-a/s(a)}$$

Estimate the density of **all** k -parent numbers over all naturals? (note that the condition $2|a$ is removed from the sum)

3 Tabulation of k-Parent Numbers

3.1 Brute-Forcing Preimages

Consider the equation $s(n) = m$, where the integer m is the *image* of n under the sum-of-proper-divisors function. Computing $s(n)$ is equivalent to the prime-factorization of n , using the method of [5]. A harder problem is to compute the *preimages* of m , all integers n such that $s(n) = m$ expressed as $s^{-1}(m)$. Restricting to even values of m this problem can be brute-forced by computing $s(n)$ for ranges of n , a conjecture similar to [7] is useful to select the values of n .

Proposition 3.1. $s(n)$ is even if and only if

$$n = (2m + 1)^2, \text{ or}$$

$$2|n \text{ and } (n \neq m^2 \text{ or } n \neq 2m^2)$$

Conjecture 3.1. *If $2|m$ and $s(n) = m$ then $n \in \{x \mid \text{even } x, 2 \leq x \leq 2m\} \cup \{x^2 \mid \text{odd } x, 1 \leq x \leq m\}$*

- If $2|n$ then $s(n) = 1 + 2 + \frac{n}{2} + x$ so $s(n) > \frac{n}{2}$;
thus $s(n) > m$ if $n > 2m$.
- If $2 \nmid n$ then $s(n) = m$ if and only if n is an odd square,
thus $s(n) = 1 + \sqrt{n} + x$ and $s(n) > \sqrt{n}$. So $s(n) > m$ if $n > m^2$.

This approach is expensive if computing $s^{-1}(m)$ for a single value of m , most values will be disregarded when searching for $s(n) = m$. It is more palatable to compute preimages for a range of m , as determining $s^{-1}(m)$ requires $\{s^{-1}(x) \mid \text{even } x, 1 \leq x \leq m\}$ anyway. This computation is useful to count the occurrence and density of k -parent numbers, these counts can be compared to the predictions of the k -parent heuristic model to provide evidence for correctness.

3.2 Pomerance-Yang Algorithm

We can certainly do better than this brute-force approach for enumerating preimages, a technique of [9] is a useful improvement. Algorithm 1 presents the Pomerance-Yang algorithm modified to determine $\#s^{-1}(n)$ for all integers less than or equal to the bound x .

Algorithm 1 Pomerance-Yang

```
1: procedure ENUMERATE_KPARENT( $x$ )
2:    $f[i] \leftarrow 1, \forall i \in [1, x]$ 
3:   for all odd  $m \in [1, x]$  do
4:     if  $2 \mid \sigma(m)$  then
5:        $t \leftarrow 3\sigma(m) - 2m$ 
6:       while  $t \leq x$  do
7:          $f[t] \leftarrow f[t] + 1$ 
8:          $t \leftarrow 2t + \sigma(m)$ 
9:       end while
10:    end if
11:    if  $\sigma(m) = m + 1$  then  $\triangleright$  if  $m$  is prime
12:       $f[m + 1] \leftarrow f[m + 1] + 1$ 
13:    end if
14:  end for
15:  for all odd composite  $m \in [1, x^{2/3})$  do
16:    if  $s(m^2) \leq x$  then
17:       $f[s(m^2)] \leftarrow f[s(m^2)] + 1$ 
18:    end if
19:  end for
20:  return  $f$ 
21: end procedure
```

Algorithm 1 was computed on a large scale by [3] who used the result to compute the geometric mean of $s(n)/n$ weighted by $\#s^{-1}(n)$. Extending this work to compute the density of k -parent numbers I have re-implemented the algorithm and can report a substantial performance improvement.

References

- [1] W. Bosma and B. Kane, *The aliquot constant*, The Quarterly Journal of Mathematics **63** (2010), no. 2, 309–323.
- [2] E. Catalan, *Propositions et questions diverses*, Bulletin de la Société Mathématique de France **2** (1888), 128–129.
- [3] K. Chum, R. K. Guy, Jr. M. J. Jacobson, and A. S. Mosunov, *Numerical and statistical analysis of aliquot sequences*, Experimental Mathematics **29** (2018), no. 4, 414–425.
- [4] Matthew Conroy, *Examples of integer sequence density calculations*.
- [5] Dennis DeTurck, *Finding the sum of the divisors of n* , 2005, <https://www2.math.upenn.edu/deturck/m170/wk3/lecture/sumdiv.html>.

- [6] Richard K. Guy and J. L. Selfridge, *What drives an aliquot sequence?*, Mathematics of Computation **29** (1975), no. 129, 101–101.
- [7] Florian Luca and Carl Pomerance, *The range of the sum-of-proper-divisors function*, Acta Arithmetica **168** (2015), 187–199.
- [8] Paul Pollack and Carl Pomerance, *Some problems of erdős on the sum-of-divisors function*, Transactions of the American Mathematical Society, Series B **3** (2016), 1–26.
- [9] Carl Pomerance and Hee-Sung Yang, *Variant of a theorem of erdős on the sum-of-proper-divisors function.*, Mathematics of Computation **83** (2014), no. 288, 1903–1913.
- [10] Neil Sloane, *A005114*, 2022.
- [11] Paul Zimmermann, *Aliquot sequences 276, 552, 564, 660, 996, 1074 and 1134*,
http://christophe.clavier.free.fr/Aliquot/site/zimmermann_table.html,
 Aug 2016.