



ANALYSIS OF ALGORITHMS FOR MOSER'S PROBLEMS ON SUMS OF CONSECUTIVE PRIMES

Jonathan P. Sorenson

*Department of Computer Science and Software Engineering, Butler University,
Indianapolis, Indiana
jsorenso@butler.edu*

Eleanor Waiss

*West Bend Mutual Insurance, West Bend, Wisconsin
waiss.eleanor@gmail.com*

Received: 8/29/25, Accepted: 7/16/26, Published: 9/28/26

Abstract

In his 1963 paper on the sum of consecutive primes, Moser posed four open questions related to $f(n)$, the number of ways an integer n can be written as a sum of consecutive primes. In this paper, we present and analyze two algorithms that, when given a bound x , construct a histogram of values of $f(n)$ for all $n \leq x$. These two algorithms were described, but not analyzed, by Jean Charles Meyrignac and Michael S. Branicky. We show that the first algorithm takes $O(x \log x)$ time using $x^{2/3}$ space, and the second has two versions, one of which takes $O(x \log x)$ time but only $x^{3/5}$ space, and the other which takes $O(x(\log x)^2)$ time but only $O(\sqrt{x \log x})$ space. However, Meyrignac's algorithm is easier to parallelize. We then present data generated by these algorithms that address all four open questions.

1. Introduction

We say a positive integer n is *gleeful* if it can be written as the sum of consecutive primes. So, for example, $n = 10$ is gleeful since $10 = 2 + 3 + 5$. Some integers have multiple gleeful *representations*, or distinct ways of expressing them as such a sum; for instance, 17 has two representations: $2 + 3 + 5 + 7$ and 17 (a single prime is always gleeful), and for a somewhat larger example, we have

$$2357 = 773 + 787 + 797 = 461 + 463 + 467 + 487. \quad (1)$$

Let $f(n)$ count the number of gleeful representations for the integer n . Thus, $f(6) = 0$, $f(10) = 1$, and $f(17) = f(2357) = 2$.

Moser [9] proved that the average value of $f(n)$ is asymptotically equal to $\log 2$; that is,

$$\frac{1}{x} \sum_{n=1}^x f(n) = \log 2 + o(1)$$

for large x . He then posed the following four questions:

1. Is $f(n) = 1$ infinitely often?
2. Is $f(n) = k$ solvable for every integer $k \geq 0$?
3. Does the set of numbers n such that $f(n) = k$ have positive density for every integer $k \geq 0$?
4. Is $\limsup_{n \rightarrow \infty} f(n) = \infty$?

See also problem C2 from [4], and the following Online Encyclopedia of Integer Sequences (OEIS) entries [10]:

A054859, A054845, A067381, A054996, A054997, A054998, A054999,
A055000, A055001 .

As a generalization of gleeful numbers, an integer is *k-gleeful* if it can be written as the sum of k th powers of consecutive primes. For previous work on k -gleeful numbers, with $k > 1$, see [16, 11, 8, 5, 6]. In particular, Florian Luca addresses Moser’s fourth question for $k = 2$ in [6].

In this paper, we address Moser’s four questions, by describing and analyzing variations of two known algorithms that generate data on the function f . That data suggests the answers are all “yes”. In particular, given a bound x , our algorithms compute $f(n)$ for all $n \leq x$ and then construct a histogram $h(k, x)$ (oftentimes shorthand to $h(k)$ where x is implicit) where

$$h(k) = \#\{n \leq x : f(n) = k\}.$$

Our first attempt is Algorithm 1. It takes time linear in x , which is optimal for this problem, but it requires space linear in x to hold the array f (plus the prefix sums array; see [8, 11]). This is problematic on currently available hardware if we want to compute the histogram h for, say, $x = 10^{14}$. Thus, the algorithm variants we present here focus on reducing memory use while keeping the running time as close to $O(x)$ as possible.

Throughout this paper, we count arithmetic operations on integers with $O(\log x)$ bits, including branching and array indexing, as taking constant time, and we count space used in bits. In practice, all of our variables are 64-bit integers or smaller.

We now state our main results and outline the rest of this paper.

In Section 2, we look at an algorithm due to Jean Charles Meyrignac (2000) that appeared in a puzzle solution to Problem #46, posted on the primepuzzle.net

Algorithm 1: A first approach

Input: x

Output: $h[]$

- 1 Allocate an array for the values of f and initialize all values to zero.
 - 2 Construct all sums of consecutive primes $\leq x$; we can easily adapt the algorithm from [11] to find these representations in time linear in x .
 - 3 For each representation found in the previous step, if n is the sum of the primes in that representation, increment $f(n)$.
 - 4 A scan of the array holding f suffices to create the histogram h .
-

website curated by Carlos Rivera [13]. Henceforth we call this the *puzzle algorithm*. We analyze a variant of the puzzle algorithm, mostly by filling in the details of Meyrinac’s description. The algorithm parallelizes nicely, has an interesting time/space tradeoff, and generates values of $f(n)$ for all n from an interval.

In Section 3, we look at an algorithm due to Michael S. Branicky (2022), which was written in Python and appears on the OEIS [10, A054845]. Again, we analyze two variants of this priority-queue-based algorithm, depending on how primes are generated. These algorithms have the property that $f(n)$ values are generated one at a time in increasing order based on n . Although these algorithms present a better time/space tradeoff, they resist parallelization. Nevertheless, we were able to run one of our priority-queue variants up to $x = 10^{12}$ using a single core, and up to $x = 10^{13}$ using 156 cores.

In Section 4, we present histograms for x up to 10^{14} . We also discuss some of our implementation details. The curious reader may find our code and data in [17]. Finally, in Section 4.1 we give conjectured answers to Moser’s questions based on our data.

2. The Puzzle Algorithm

The *puzzle algorithm*, as we call it, is based on the following description, written by Jean Charles Meyrignac [13].

- 1) First, pre-compute a table of primes
- 2) Lower bound = 0
- 3) Higher Bound=Lower Bound + size of sliding window
- 4) Compute all consecutive sums of n elements for every element of the pre-computed table and counts every occurrence.
- 5) Display the best occurrences
- 6) Change the lower bound and go to 3)

Our version of this algorithm varies in several ways, but the main idea is the same.

Values of $f(n)$ are computed for every integer n in a segment or window, and then these are scanned to add to a running histogram. We spend some time discussing the details of how step (4) above can be done efficiently. But before we dive into the details, we need to establish some notation and prove some useful lemmas.

2.1. Notation and Lemmas

Let us introduce some notation.

Let x be the limit for our computation. The interval $[1, x]$ is divided into equal-length segments (or windows) of size Δ each (the *sliding window*). The last segment may be shorter if Δ does not divide x . There are $\lceil x/\Delta \rceil$ segments total. For a particular segment, let x_1 be the left endpoint and x_2 be the right endpoint, but just outside the window. We have $\Delta = x_2 - x_1$, and we assume that Δ will be a fractional power of x , such as $\sqrt{x}$ or $x^{2/3}$. The basic idea is that the segment represents the values of $f(n)$ for $x_1 \leq n < x_2$, which takes Δ small integers to represent (an 8-bit integer suffices in practice).

We use the term *chain* to indicate a string of consecutive primes, and we use m to denote the number of primes, or length, of the chain. Then n is the sum of those primes. If we write c as a variable representing a chain, then $m(c)$ is its length and $n(c)$ is its sum. In practice, we store the start and stop prime indices, the length, and the sum in a chain variable. For example, if $c := [2, 3, 5, 7]$, then $m(c) = 4$ and $n(c) = 17$.

We will often wish to *slide* a chain, which means dropping its smallest prime and adding the next larger prime. A slide operation preserves length but increases the sum. Continuing our example, $\text{slide}(c) = [3, 5, 7, 11]$. If a list of primes has been precomputed, we can implement a **slide** operation so that it takes constant time.

Let $M := M(x)$ be the length of the longest chain of consecutive primes whose sum is $\leq x$. Without loss of generality, we may assume 2 is in the longest chain so that

$$2 + 3 + 5 + \cdots + p_M \leq x < 2 + 3 + 5 + \cdots + p_M + p_{M+1}.$$

For any chain of length m with sum $n \leq x$, we have $m \leq M$.

Let $p(m) = p(m, x)$ be the largest prime that participates in any chain with sum $n \leq x$ of length m .

Lemma 1. *We have*

$$\frac{x}{m+2} < p(m, x) \leq \frac{2x}{m}(1 + o(1)).$$

Proof. There exists an integer $\ell > 0$ such that

$$p_{\ell+1} + p_{\ell+2} + \cdots + p_{\ell+m} \leq x < p_{\ell+1} + p_{\ell+2} + \cdots + p_{\ell+m} + p_{\ell+m+1},$$

so that $p(m, x) = p_{\ell+m}$ here.

For the lower bound, we have

$$\begin{aligned} x &< p_{\ell+1} + p_{\ell+2} + \cdots + p_{\ell+m} + p_{\ell+m+1} \\ &\leq p_{\ell+1} + p_{\ell+2} + \cdots + p_{\ell+m} + 2p_{\ell+m} \\ &\leq (m+2)p_{\ell+m} = (m+2)p(m) \end{aligned}$$

by Bertrand's postulate.

For the upper bound, we note that

$$\sum_{i=1}^u p_i \sim \sum_{i=1}^u i \log i \sim \frac{u^2 \log u}{2} \quad (2)$$

(see Bach and Shallit [2], Chapter 2). We then have

$$\begin{aligned} x &\geq p_{\ell+1} + p_{\ell+2} + \cdots + p_{\ell+m} \\ &\sim \frac{1}{2}((\ell+m)^2 \log(\ell+m) - \ell^2 \log \ell) \\ &\sim \frac{m}{2}p_{\ell+m} + \frac{\ell}{2}(p_{\ell+m} - p_{\ell}) \\ &\geq \frac{m}{2}p(m), \end{aligned}$$

by the prime number theorem. $\square$

We have the following estimate for $M = M(x)$.

Lemma 2. *We have*

$$M(x) \sim \frac{2\sqrt{x}}{\sqrt{\log x}}.$$

Proof. Consider a chain of length M that consists of the first M primes. From (2), we have

$$\frac{M^2 \log M}{2} \leq x(1 + o(1)) < \frac{(M+1)^2 \log(M+1)}{2}.$$

Taking logarithms, we deduce that $\log x \sim 2 \log M$. The result then follows. $\square$

And we have the following bound on sums of slided chains.

Lemma 3. *If c is a chain of length m and sum n , then $n(\mathbf{slide}(c)) - n(c)$ has average value $\Theta(m \log(n/m))$.*

Proof. A single **slide** operation changes the sum of the chain by $p - q$, where p is the newest prime added to the chain and q was the smallest prime just removed from the chain by sliding. By Lemma 1, we know $p = \Theta(n/m)$, and so by the

prime number theorem, we know consecutive primes in the chain are, on average, $\log(n/m)(1 + o(1))$ apart. Thus $p \approx q + m \log(n/m)$, or

$$n(\text{slide}(c)) - n(c) = p - q = \Theta(m \log(n/m))$$

on average. $\square$

Corollary 1. *The number of chain sums n from chains of length m that fall in a segment of length Δ is, on average, proportional to*

$$\Delta / (m \log(x/m)),$$

where x is the midpoint of the interval.

2.2. Algorithm Description

We are now ready to describe our version of the puzzle algorithm. The input is a bound x , and the output is an array representing the histogram h , where $h(k) = \#\{n \leq x : f(n) = k\}$.

In step (4) of Meyrignac's description, the basic idea is to work through the chains from maximum length M down to length 1, generating all relevant chains of the current length m before moving on. If the length m is *large*, we build the first chain of length m from the previous first chain of length $m + 1$ by dropping the largest prime and then sliding as needed. (If we were to build the chain of length m from scratch, this takes $\sum m \approx M^2$ time per segment, which is way too slow.) The original chain of length M is constructed by finding primes until the sum is maximal but less than x_2 .

When m gets small enough, the first chain of length m has no overlap with the previous one of length $m + 1$, and may even be far enough away that making the first m -length chain from scratch is more efficient. This is how we define *small* values of m . We define m_{cutoff} as the largest *small* value of m .

To set a more precise value for m_{cutoff} , we set the work threshold to the time needed to process the segment for that value of m after that first chain is found. By Corollary 1, this is $O(\Delta / (m_{\text{cutoff}} \log x))$ operations. A key in the analysis is estimating this cutoff point, which we do now.

Lemma 4. *On average, we have*

$$m_{\text{cutoff}} = \Theta(x_1 / \Delta).$$

Proof. Let c_a be the first chain of length m we use (after), and c_b be the first chain of length $m + 1$ (before). Both chains have sums near x_1 , as in

$$\begin{aligned} n(c_b) &\geq x_1 & \text{and} & & n(\text{slide}^{-1}(c_b)) < x_1, \\ n(c_a) &\geq x_1 & \text{and} & & n(\text{slide}^{-1}(c_a)) < x_1 \end{aligned}$$

are both true. By Lemma 3, we can deduce that

$$|n(c_a) - n(c_b)| = O(m \log(x/m)) \quad (3)$$

on average.

By Lemma 1, the largest prime in c_b is $p(m+1, x_1) = \Theta(x_1/m)$. Let s denote the number of **slide** operations to obtain c_a from c_b after $p(m+1, x_1)$ is removed from c_b . Then we have

$$n(c_a) = n(\text{slide}^s(c_b)) - p(m+1, x_1).$$

Note that removing $p(m+1, x_1)$ before or after sliding gives the same sum.

By Lemma 3, sliding a chain of length m with sum near x changes its sum by $\Theta(m \log(x/m))$. This gives us

$$n(c_a) = n(c_b) + \Theta(sm \log(x_1/m)) - p(m+1, x_1)$$

or

$$p(m+1, x_1) = n(c_b) - n(c_a) + \Theta(sm \log(x_1/m))$$

on average. Using Equation (3) then gives

$$p(m+1, x_1) = \Theta(sm \log(x_1/m))$$

or, with Lemma 1,

$$\frac{x_1}{\log x_1} \approx s \cdot m_{cutoff}^2$$

on average for when $m = m_{cutoff}$. But, we have

$$s = \Theta\left(\frac{\Delta}{m_{cutoff} \log x}\right),$$

as the number of slide operations matches the work for the segment for that value of m by Corollary 1. A bit of simplification completes the proof. $\square$

With an estimate for m_{cutoff} in hand, we are now ready to describe the Puzzle Algorithm in detail (Algorithm 2).

2.3. Time and Space Analysis

In our analysis, we assume that Δ is a fractional power of x and, in fact, at least $M(x)$. We assume that $x_2 \approx x$. The number of loop iterations when x_2 is substantially smaller than x is limited, and we can assume, therefore, that this case contributes negligibly.

The space used by our algorithm will be at least Δ bits. If possible, we want to keep space usage as low as possible while keeping our running time as close to linear in x as we can.

Algorithm 2: The Puzzle Algorithm

Input: x, Δ
Output: $h[]$

- 1 Initialize $h := []$, of length 20, to zeroes;
- 2 Construct primes up to $O(\Delta)$;
- 3 Initialize $f := []$, of length Δ , to zeroes;
- 4 $x_1 := 2, x_2 := \Delta + x_1$;
- 5 **while** $x_1 \leq x$ **do**
 - 6 **if** $x_2 > x + 1$ **then** $x_2 := x + 1$;
 - 7 Clear f to zeroes;
 - 8 Build chain c such that $c.\text{length} \leq M(x_2)$, $c.\text{sum} \leq x_2$, $c.\text{pmin} := 2$, and $2 + \text{slide}(c).\text{sum} > x_2$;
 - 9 $m_{\text{cutoff}} := 0$;
 - 10 **while** $c.\text{length} > m_{\text{cutoff}}$ **do**
 - 11 $s := \Delta / (c.\text{length} \cdot \log(x_1 / c.\text{length}))$;
 - 12 $i := 0$;
 - 13 **while** $c.\text{sum} < x_1$ **do**
 - 14 Increment i ;
 - 15 $c := \text{slide}(c)$;
 - 16 **if** $i \geq s$ **then** $m_{\text{cutoff}} := m$;
 - 17 $c_0 := c$;
 - 18 **while** $c.\text{sum} < x_2$ **do**
 - 19 Increment $f[c.\text{sum} - x_1]$;
 - 20 $c := \text{slide}(c)$;
 - 21 $c := c_0$;
 - 22 Drop the largest prime from c ;
 - 23 $m := m_{\text{cutoff}} - 1$;
 - 24 **while** $m > 0$ **do**
 - 25 Find all primes in $[p(x_1, m) - m \log(x_2), p(x_2, m)]$;
 - 26 Build chain c such that $c.\text{length} = m$, $x_1 \leq c.\text{sum} \leq x_2$, and $\text{slide}^{-1}(c).\text{sum} < x_1$;
 - 27 **while** $c.\text{sum} < x_2$ **do**
 - 28 Increment $f[c.\text{sum} - x_1]$;
 - 29 $c := \text{slide}(c)$;
 - 30 Decrement m ;
 - 31 **for each** n , $x_1 \leq n < x_2$ **do** Increment $h[f[n - x_1]]$;
 - 32 $x_1 := x_2$;
 - 33 $x_2 := x_2 + \Delta$;

Since the average value of $f(n)$ is constant, most values of $f(n)$ can be represented in a constant number of bits, with larger values stored in a secondary data structure to keep the overall space use linear. In practice, we simply used an array of 8-bit integers. We have not yet found an integer n with $f(n) > 14$.

Throughout, we mention the use of average-case analysis. For example, by the prime number theorem we know that consecutive primes near x differ by $\log x$ on average. For a particular step on a particular segment, this will vary, but over the course of the entire algorithm, significant variations from the average must cancel out, as the algorithm examines all primes and all chains up to x .

2.3.1. Preprocessing

It only takes $O(\Delta)$ time to do steps 1 to 4. In particular, finding the primes up to $O(\Delta)$ can be done in linear time using any number of prime sieves, including the sieve of Eratosthenes with a wheel or the Atkin-Bernstein sieve [1]. In practice, we use an incremental sieve [15] to find more primes as needed, since we do not know the implied constant in $O(\Delta)$.

2.3.2. Main Loop

We move on to the the main loop, Line 5, which repeats $O(x/\Delta)$ times.

Inside the main loop, all the steps outside of processing large and small values of m can be easily seen to take at most $O(\Delta)$ time, especially when we recall that $M \ll \Delta$.

2.3.3. Large m

Processing chains with large m begins on Line 10. Everything, aside from the loops starting on Lines 13 and 18, takes constant time. Summing over all large m gives us $O(M) = O(\Delta)$.

By Corollary 1, the loop at Line 18 takes at most $O(\Delta/(m \log(x/m)))$ time. We can simplify this to $O(\Delta/(m \log x))$ because $1 \leq m \leq M = \Theta(\sqrt{x}/\log x)$ by Lemma 2, so that $\log x = \Theta(\log(x/m))$.

The value of s in Line 13 is chosen specifically to make sure that the cost of the loop at Line 13 is not more than the cost of the one at Line 18.

For large m , then, we have a cost proportional to

$$\frac{x}{\Delta} \left(\Delta + \sum_{m=m_{\text{cutoff}}}^M \frac{\Delta}{m \log x} \right).$$

By Lemma 2, this simplifies to $O(x)$.

2.3.4. Small m

Processing chains with small m begins on Line 24. For small m , the largest prime of any chain of length m will range from $p(m, x_1)$ to $p(m, x_2)$. This explains the need for Line 25. By Lemma 1 this is from roughly x_1/m to x_2/m , or an interval of size $O(\Delta/m)$. The smaller primes of any chain either fall in that interval, or just below it by roughly $m \log x$ by the prime number theorem.

Now, the Atkin-Bernstein sieve will find primes of size y on an interval of size $y^{1/3}$ or larger in time linear in the interval length, using Galway's improvements [3]. Here, $y = p(m, x_2) = O(x/m)$, so we need

$$\begin{aligned} \frac{\Delta}{m} + m \log x &\gg (x/m)^{1/3} \quad \text{or} \\ m &\ll \Delta^{3/2}/x^{1/2}. \end{aligned}$$

Using Lemma 4 and a bit of work, this gives us an upper bound of $x^{2/5}$ for m_{cutoff} , as $m \leq m_{cutoff}$ for "small" m . As we will see later, $m_{cutoff} \approx x^{1/3}$, and so the Atkin-Bernstein sieve with Galway's improvements will work fine here. (Note that in practice we used something akin to the pseudosquares prime sieve [14], but with Miller-Rabin prime tests using particular bases to achieve deterministic primality testing.)

Summing this over all segments and over all $m \leq m_{cutoff}$ gives a time bound of

$$\frac{x}{\Delta} \sum_{m=1}^{m_{cutoff}} \left(\frac{\Delta}{m} + m \log x \right) \sim x \log m_{cutoff} + \frac{x \log x}{2\Delta} \cdot m_{cutoff}^2$$

To keep this close to linear, we want

$$\Delta \gg m_{cutoff}^2.$$

Line 26 can be done in $O(m)$ time now that the necessary primes are available. Line 27 takes at most $O(\Delta/(m \log x))$ time by Corollary 1, since the primes are available so that **slide** operations take constant time.

Theorem 1. *On input $x > 0$, the Puzzle Algorithm computes a histogram h as described above in $O(x \log x)$ arithmetic operations and $O(x^{2/3})$ space.*

Proof. From our discussion above, everything outside processing small values of m takes $O(x)$ time.

The equation to describe the asymptotic running time, within a constact factor,

for small values of m , is

$$\begin{aligned} & \frac{x}{\Delta} \sum_{m=1}^{m_{\text{cutoff}}} \left(\frac{\Delta}{m} + m \log x + m + \frac{\Delta}{m \cdot \log x} \right) \\ & \sim x \log m_{\text{cutoff}} + \frac{m_{\text{cutoff}}^2 \cdot x \log x}{2\Delta} + \frac{m_{\text{cutoff}}^2 \cdot x}{2\Delta} + \frac{x \log m_{\text{cutoff}}}{\log x} \\ & \sim x \log m_{\text{cutoff}} + \frac{m_{\text{cutoff}}^2 \cdot x \log x}{2\Delta}. \end{aligned}$$

This argues for making m_{cutoff} as small as possible, but we have Lemma 4. Next, we multiply the numerator and denominator of the second term by m_{cutoff} and write $\Delta \cdot m_{\text{cutoff}} = \Theta(x)$ to give us

$$x \log m_{\text{cutoff}} + m_{\text{cutoff}}^3 \cdot \log x.$$

We obtain that $m_{\text{cutoff}} = x^{1/3}$ and $\Delta = x^{2/3}$ and the proof is complete. $\square$

On current hardware, $x^{2/3}$ space is problematic if we want to compute the histogram h up to, say, $x = 10^{15}$, as this means using 10^{10} space, which is too much. What happens if we allow our runtime to creep up – does this allow for less space use? The answer is yes, as the runtime is inversely proportional to Δ . If we set $\Delta = x^{2/3-u}$, then by Lemma 4 we have $m_{\text{cutoff}} = O(x/\Delta) = x^{1/3+u}$, then the runtime becomes

$$O\left(x \log x + \frac{(x^{1/3+u})^2 x \log x}{x^{2/3-u}}\right) = O(x^{1+3u} \log x).$$

So for example, setting $\Delta = \sqrt{x}$, or $u = 1/6$, gives a running time of $O(x^{3/2} \log x)$.

3. The Priority Queue Algorithm

The main idea behind the *priority queue* algorithm is to generate all the chains with sums between 1 and x , our upper bound, and insert them into a priority queue (a min-heap) using their sums as their priority. Thus, the front of the queue will have the chains of smallest sums. Then we can find all chains that sum to 1 from the front of the queue, counting them as we remove them from the priority queue, and incrementing the histogram as appropriate. Rinse and repeat with 2, 3, 4, $\dots$, x . This works, but there are too many chains to store all at once – $O(x)$ according to Moser. However, observe that a particular value of n has at most one representation of any given length. This means we need only store one chain of each length in the priority queue. By Lemma 2 we know that this is $O(\sqrt{x}/\log x)$, much better than $O(x)$. When a chain is removed from the priority queue, we simply perform a **slide**

and then insert it back into the queue, prioritized by its new, larger sum. This is precisely the idea in Michael S. Branicky’s algorithm, described as Python code from [10, A054845] in 2022.

```
(Python) # alternate version for going to large n
import heapq
from sympy import prime
from itertools import islice
def agen(): # generator of terms
    p = v = prime(1); h = [(p, 1, 1)]; nextcount = 2; oldv = ways = 0
    while True:
        (v, s, l) = heapq.heappop(h)
        if v == oldv: ways += 1
        else:
            yield ways
            for n in range(oldv+1, v): yield 0
            ways = 1
        if v >= p:
            p += prime(nextcount)
            heapq.heappush(h, (p, 1, nextcount))
            nextcount += 1
        oldv = v
        v -= prime(s); s += 1; l += 1; v += prime(l)
        heapq.heappush(h, (v, s, l))
    print(list(islice(agen(), 102))) # Michael S. Branicky, Feb 17 2022
```

The reader need not examine this Python program too closely, as we will restate the algorithm, using the notation in this paper, elaborating on details, and analyzing the running time as we go.

3.1. Algorithm Description

Let the interval $[x_1, x_2)$ be the input to the algorithm. For ease of notation, let $\Delta := x_2 - x_1$. As in the last section, the output is a histogram h where $h[k] = \#\{n : x_1 \leq n < x_2 \text{ and } f(n) = k\}$.

It is fine to keep in mind $[1, x + 1)$ as the input for now, but we are setting the stage for parallelization later.

3.1.1. Setup

A chain object is stored as 3 integers: the smallest prime of the chain (**pmin**), the largest prime of the chain (**pmax**), and its sum (**sum**). We store chain objects in an array, and the array position indicates its length, so the length of the chain is stored implicitly, and array position zero is left empty.

To create this array of chain objects, we start by computing the longest chain c that includes 2, with maximal sum at most x_2 . Let M be the length of c . Allocate

the array `happy[]` of length $M + 1$, and set `happy[M] := c`. We then fill the array from position $M - 1$ down to 1. For long chains of length m , we copy the previous chain of length $m + 1$, drop the largest prime, and then **slide** it as necessary until its sum is minimal, but greater than or equal to x_1 . For short chains of length m , we estimate the largest prime of the chain to be x_1/m (see Lemma 1), find consecutive primes below that to get m of them in total, and then **slide** as necessary to get a sum that is minimal but at least x_1 . We prove below that if we split long from short chains at $x_2^{1/3}$, the number of **slide** operations it takes to generate all the chains is $O(x_2^{2/3})$. This also tells us that we want $x_2 - x_1 \gg x_2^{2/3}$ for the setup time to be negligible compared to our main loop. Unlike the puzzle algorithm, we do not need any space for the interval.

Lemma 5. *Let $x_1 < x_2$ be positive integers. A list of chains, one of each length from 1 to $M(x_2)$, where the chain of length m has minimal sum $\geq x_1$, can be constructed in a total number of **slide** operations that does not exceed $O(\sqrt{x_2/\log x_2} + x_1^{2/3})$.*

Proof. Let $m_c := x_1^{1/3}$ be the cutoff length between long and short chains as described above. The first chain of length $M(x_2)$ can be made with $M(x_2) = O(\sqrt{x_2/\log x_2})$ **slide** operations or the equivalent by Lemma 2. Each subsequent long chain of length m is created from the previous chain of length $m + 1$ as described above, and this takes

$$1 + \pi(p(m, x_1)) - \pi(p(m + 1, x_1))$$

slide operations, where $\pi(x)$ is the prime counting function. This telescopes when summed, giving

$$\begin{aligned} & \sum_{m=m_c}^{M(x_2)} (1 + \pi(p(m, x_1)) - \pi(p(m + 1, x_1))) \\ & \leq O(M) + \pi(p(m_c, x_1)) - \pi(p(M, x_1)) \\ & = O(M + x_1/m_c) \end{aligned}$$

by Lemma 1 and the prime number theorem. A short chain of length m needs m primes near x_1/m , the equivalent of $O(m)$ **slide** operations. Summing over all short lengths m gives $O(m_c^2)$ total. We then optimize the cost $O(M + x_1/m_c + m_c^2)$ giving $m_c = x_1^{1/3}$ and our proof is complete. $\square$

We conclude the setup by inserting the integers 1 to M into the priority queue, where the priority of the integer i is `happy[i].sum`.

Later, we discuss two possible implementations of the priority queue: as a binary heap, or using a hash table. In either case, the initial queue can be constructed in $O(M)$ time: We can use a *build* operation for the heap, or $O(M)$ hash table insertions of chain objects, depending on our choice for data structure.

3.1.2. Main Loop

We describe the main loop of our algorithm below (Algorithm 3).

Algorithm 3: Main Loop for the Priority Queue Algorithm

```

1 for  $n := x_1$  to  $x_2 - 1$  do
2    $f := 0$ ;
3   while there are objects with priority  $n$  in the priority queue: do
4     Let  $c$  be the chain object at the front of the queue with  $c.\text{sum} = n$ ;
5     Remove  $c$  from the priority queue;
6     Increment  $f$ ;
7      $c := \text{slide}(c)$ ;
8     if  $c.\text{sum} < x_2$  then Insert  $c$  back into the priority queue;;
9   Increment  $h[f]$ ;
```

The inner while loop, which starts at Line 3, iterates once per chain, which is once per gleeful representation, which is linear, or $O(\Delta)$ on average, over the course of the algorithm, due to Moser’s theorem. We have one delete-Min or dequeue operation, at most one insertion, and one **slide**.

If we use a hash table to store the priority queue (where the key is the chain sum), insertions and deletions are constant time on average. If we use a standard binary heap, these operations take $O(\log M)$ time each. As $M < \sqrt{x_2}$, this is $O(\log x_2)$, for a total of $O(\Delta \log x_2)$ time for the interval.

Next, we discuss the cost of **slide** operations. If a list of primes is available, we can store the index of the prime in our chain objects, and then finding the next prime (for both **pmax** and **pmin**) takes constant time. However, we need primes as large as x_2 , and finding all primes up to x_2 takes too much time, and storing them takes too much space. But we can pick a bound B , and compute and store all primes up to B . For primes needed beyond B , we can generate them either using an incremental prime sieve (see [15]) or using prime tests, with some support to make this as efficient as possible (see [14]).

Let us now derive optimal values for B , based on whether we use an incremental sieve or prime tests. We begin with an incremental sieve. Galway [3] showed how to modify the Atkin-Bernstein sieve [1] to find primes in an interval $[a, b]$ using at most $O(b - a + b^{1/3})$ time and $O(b^{1/3})$ space. This sieve can, theoretically, be constructed to generate the primes as needed [15]. We divide the chains stored in our priority queue into those that use primes up to B , and those that use this incremental sieve to generate needed primes. Thus, each chain object that uses the sieve, would have attached to it its own copy of an Atkin-Bernstein-Galway incremental sieve to generate primes as needed for sliding. We now prove that the optimal value for B is $x_2^{3/5}$ in this case.

Theorem 2. *After setup, the priority queue algorithm takes $O(\Delta \log x_2 + x_2^{3/5})$ time to find and count all representations of integers in the interval $[x_1, x_2]$ where $\Delta = x_2 - x_1$. The space used is $O(x_2^{3/5})$.*

In practice, we want $\Delta \gg x_1^{2/3}$ so that the setup cost is negligible. Thus, for the interval $(1, x]$ we get a running time of $O(x \log x)$ using $O(x^{3/5})$ space. We can parallelize this, keeping an optimal amount of work, using up to $O(x^{1/3})$ processors, where each processor gets its own interval of length $x^{2/3}$.

Proof. Set $B := x_2^{3/5}$. The algorithm starts by finding all primes up to $O(B)$. Any linear-time prime sieve is fine for this. The primes can be stored in $O(B)$ bits.

By Lemma 5 the time for initializing the priority queue is negligible.

For **slide** operations on chains whose primes fall below B , **slide** operations take constant time. No extra space beyond the $O(B)$ to store the prime list is required. Let m be the length of any such chain. Then $p(m, x_2) \leq B$. By Lemma 1, we have $p(m, x_2) = \Theta(x_2/m)$, giving $m \geq \Theta(x_2/B) = \Theta(x_2^{2/5})$.

Chains of length m with $m \leq \Theta(x_2^{2/5})$, then, must have an incremental sieve attached. Each one uses space $O(p(m, x_2)^{1/3})$. Again using Lemma 1, the total space used is proportional to

$$\sum_{m=1}^{x_2^{2/5}} \left(\frac{x_2}{m}\right)^{1/3} = O(x_2^{1/3} \cdot (x_2^{2/5})^{2/3}) = O(x_2^{3/5}).$$

For these shorter chains, **slide** operations take, on average, $O(\log x_2)$ time each, as the Atkin-Bernstein-Galway sieve takes linear time to find primes up to a given bound, and primes are, on average, $O(\log x_2)$ apart. $\square$

To emphasize a point, as we have defined them, short chains and small chains overlap, but are not the same concept. *Short* chains must generate their primes, so that **slide** operations require $\Theta(\log x_2)$ time on average, whereas *long* chains can find their primes from a precomputed list bounded above by B , and thus **slide** operations take constant time. These are only relevant to the priority queue algorithms. *Large* chains are created from truncating a larger chain and sliding, but *small* chains are created from scratch. These are relevant to the puzzle algorithm and the setup phase of the priority queue algorithms.

If an overwhelming majority of chains are long, this might pull the average cost per **slide** below $\log x_2$, but this is not the case. The number of short chains (for the input $[1, x]$) is (see [9])

$$\begin{aligned} \sum_{m=1}^{x^{2/5}} \pi(x/m) &\sim \sum_{m=1}^{x^{2/5}} \frac{x}{m \log(x/m)} \sim \int_1^{x^{2/5}} \frac{x}{y \log(x/y)} dy \\ &\sim x(\log \log x - \log \log(x/x^{2/5})) \approx 0.511 \cdot x. \end{aligned} \quad (4)$$

Overall there are $x \log 2 \approx 0.693 \cdot x$ chains, so roughly 74% are short. Given the running time, both a binary heap and a hash table data structure are reasonable for this approach.

Next, we consider generating primes as needed using a prime test. In particular, given a starting point n , we find the smallest prime larger than n as follows:

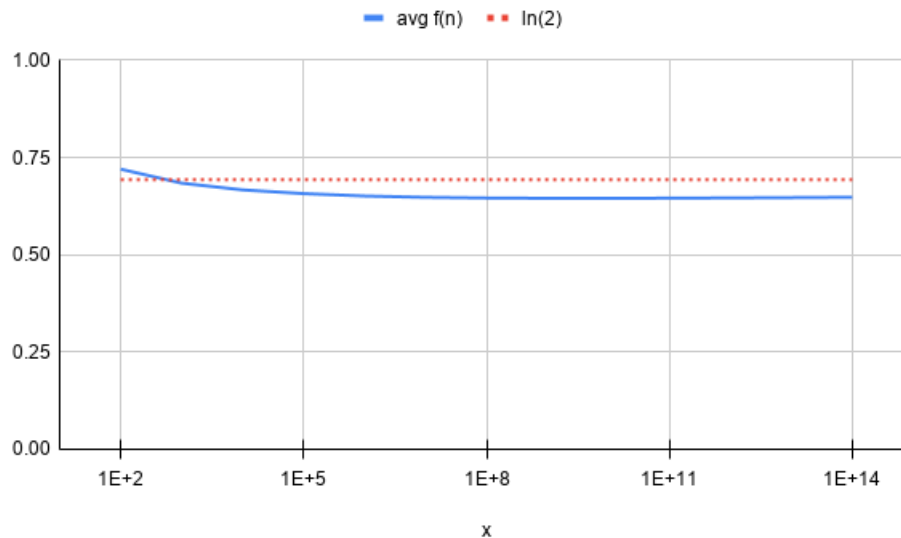
- If $n < B$ we use the precomputed list of primes to find the next prime in constant time. This implies storing prime indices instead of prime values. We distinguish indices from prime values by storing indices as negative numbers (so the prime $p_3 = 5$ is stored as -3).
- For larger n we follow the ideas in [14]:
 1. Sieve an interval of size $2 \log n$ using primes up to $\log n$. This takes $O(\log n)$ time leaving $O(\log n / \log \log n)$ possible primes in the interval $(n, n + 2 \log n]$.
 2. Scan this interval for a possible prime, and then perform a base-2 strong pseudoprime test.
 3. If the test is passed, the number is likely prime, so do a prime test. In theory, we use a pseudosquares prime test which takes $O((\log n)^2)$ time in theory. In practice, we could use a Baille-PSW prime test, which works on 64-bit integers, or a Miller-Rabin style test on a fixed, short list of appropriate bases.

This is repeated as needed until a prime is found.

The average time to find a prime is dominated by the cost of the final prime test, at $O((\log n)^2)$ time. In practice, we used a fixed wheel of size 30, performed some fixed trial divisions, and then used the Miller-Rabin test as described above, taking something closer to $O((\log n)^2 / \log \log n)$ heuristic time. We also need a previous prime function, but the cost is the same. We leave the details to the reader.

Theorem 3. *After setup, the priority queue algorithm takes $O(\Delta(\log x_2)^2)$ time to find and count all representations of integers in the interval $[x_1, x_2)$ where $\Delta = x_2 - x_1$. The space used is $O(\sqrt{x_2 \log x_2})$.*

Proof. After setup, the number of **slide** operations is $O(\Delta)$. From the discussion above, using the pseudosquares prime test we can perform each **slide** in $O((\log x_2)^2)$ time. The only significant use of space is to store the $O(M)$ chain objects, each of which uses $O(\log x_2)$ bits. This completes the proof. $\square$


 Figure 1: Average value of $f(n)$ for all $n \leq x$

4. Computational Results

Several different practical implementations of the algorithms presented in Section 2 and Section 3 were created and run over several upper bounds, and we present both their results and timing information below. Values of the histogram function $h(k, x)$, where x ranges over powers of ten are given in Table 1. We also give the average value of f on these intervals, in Figure 1.

The puzzle algorithm from Section 2 was implemented using a version of the Miller-Rabin primality test with a fixed list of bases, and ran in parallel on 560 AMD EPYC9734 cores at 2.20GHz. We used $\Delta = 5 \cdot 10^8$, or half of a GB per core. The interval $[10^{13}, 10^{14}]$ took 3 weeks, and the program needed 4 cumulative days to compute the interval $[1, 10^{13}]$. We estimate that it would take on the order of one year to run this implementation on the interval $[10^{14}, 10^{15}]$, where we can heuristically expect to find the first integer with 15 gleeful representations.

The priority queue algorithm from Section 3 was implemented using the C++ standard library `unordered_map` as a hash table, and generated primes using an incremental sieve based on the sieve of Eratosthenes. This program ran on a single thread of an Intel Xeon E5-2650 v4 at 2.20GHz, running with an upper bound of $x = 10^{12}$ and taking just over 26 days to terminate. Lower-order results include taking 3 days for $x = 10^{11}$ and 6 hours cumulative for all smaller powers of ten.

We also created a version of the priority queue algorithm using a custom min-

k	$x = 10^2$	10^3	10^4	10^5	10^6	10^7
0	46	520	5191	51462	518001	5205110
1	38	310	3290	34538	344100	3427038
2	14	140	1213	11236	111132	1099545
3	2	28	275	2396	22916	228659
4	0	0	29	323	3409	35009
5	0	2	2	44	403	4197
6	0	0	0	1	37	412
7	0	0	0	0	2	28
8	0	0	0	0	0	2
k	$x = 10^8$	10^9	10^{10}	10^{11}		
0	52209546	522955756	5232998716	52334124696		
1	34146573	340693986	3402573440	34000179027		
2	10950371	109272550	1092269461	10929758863		
3	2292360	23003362	230955228	2319354726		
4	353614	3584873	36207780	365678388		
5	42946	440748	4493266	45727635		
6	4205	44623	458971	4726851		
7	356	3793	39944	415858		
8	28	299	2980	31743		
9	1	10	204	2108		
10	0	0	10	101		
11	0	0	0	4		
k	$x = 10^{12}$	10^{13}	10^{14}			
0	523187120742	5229093083950	52255406421916			
1	339871838389	3398259361550	33983734540429			
2	109433799948	1096096081344	10980796354963			
3	23298396623	234059213343	2351331657312			
4	3691321145	37239962166	375496312242			
5	464500544	4710770197	47717060499			
6	48377037	493740792	5027735200			
7	4291580	44107969	451927961			
8	330053	3428047	35376934			
9	22539	235471	2452073			
10	1332	14382	151480			
11	66	757	8546			
12	2	31	430			
13	0	1	14			
14	0	0	1			

Table 1: Histogram data, $h(k, x)$, for x up to 10^{14}

heap implementation with primality testing, storing chain objects indexed by length in an array. We considered using two heaps – one for short chains and one for long ones – in the hope that the short chains would all fit in cache. From Equation (4) we know we would work from cache well over half the time. However, for $x = 10^{15}$, there are 10^6 chains, and each chain requires three 64-bit integers, or 24 bytes, about 8 times the L1 cache available. Caching a smaller number of chains works, but we would miss cache more often. In the end, we decided that implementing separate heaps for short and long chains was not worth the effort.

4.1. Moser’s Questions

Let us recall the questions Moser posed in his 1963 paper.

1. Is $f(n) = 1$ infinitely often?
2. Is $f(n) = k$ solvable for every k ?
3. Do the numbers for which $f(n) = k$ have a density for every k ?
4. Is $\limsup_{n \rightarrow \infty} f(n) = \infty$?

In the spirit of Cramér’s model on the distribution of primes, the histograms in Table 1 give rise to a conjecture on a probabilistic model for Moser’s function f . Namely, although not a true random probability distribution, $f(n)$ is roughly Poisson with mean $\lambda = \log 2$. Compare the histogram data for $x = 10^{14}$ (restated for convenience) with the expected counts of a Poisson variable (Table 2). In this spirit, we can heuristically answer in the affirmative to Moser’s four questions.

In particular interest to Moser’s second and fourth questions, OEIS [10, A054859] lists the smallest integer having exactly k representations. Prior to our work, this sequence was known to $k = 13$; we present the first known integer with 14 representations as a sum of consecutive primes. Table 3 gives the complete updated OEIS table, and Table 4 gives the 14 representations for our new entry. It is interesting, but not surprising, to note that 11 of the 14 records in Table 3 are themselves prime. We found five more integers with 14 gleeful representations each when we pushed our computations a bit past 10^{14} :

150419978330323, 157780612689449, 159544009953396, 218546221676323,
269239358885729.

With regard to Moser’s first and third questions, we can conjecture that the density of numbers n for which $f(n) = k$ follows the Poisson distribution’s density: asymptotically $\frac{(\log 2)^k}{2 \cdot k!}$. Indeed, we expect half of the integers will have no gleeful representation $\log 2/2$ (roughly 1/3) of the integers will have exactly one representation, and so forth.

k	$h(k, 10^{14})$	$10^{14} \cdot \Pr[X = k]$
0	52255406421916	50000000080183
1	33983734540429	34657359083576
2	10980796354963	12011325367217
3	2351331657312	2775205437692
4	375496312242	480906456153
5	47717060499	66667790839
6	5027735200	7701765209
7	451927961	762636691
8	35376934	66077434
9	2452073	5089043
10	151480	352746
11	8546	22228
12	430	1284
13	14	68
14	1	3

Table 2: $h(k)$ is heuristically Poisson with mean $\lambda = \log 2$

k	n
0	1
1	2
2	5
3	41
4	1 151
5	311
6	34 421
7	218 918
8	3 634 531
9	48 205 429
10	1 798 467 197
11	12 941 709 050
12	166 400 805 323
13	6 123 584 726 269
14	84 941 668 414 584

Table 3: Smallest n with $f(n) = k$ representations

ℓ	$p_{\min}$	$p_{\max}$
2 117 074	21 797 833	58 785 359
361 092	231 753 581	238 710 779
288 268	291 853 531	297 473 801
199 390	424 030 259	427 989 799
112 544	753 590 641	755 886 067
73 026	1 162 407 049	1 163 930 791
68 854	1 232 927 929	1 234 369 457
296	286 965 092 209	286 965 099 727
294	288 917 235 553	288 917 243 497
206	412 338 193 609	412 338 198 731
146	581 792 247 697	581 792 251 207
86	987 693 817 667	987 693 819 859
26	3 266 987 246 389	3 266 987 247 019
2	42 470 834 207 273	42 470 834 207 311

Table 4: The 14 representations for $g = 84\,941\,668\,414\,584$.

Acknowledgements. This work was supported in part by a grant from the Holcomb Awards Committee. Thanks to Frank Levinson for his generous support of Butler’s research computing infrastructure. Thanks to Ankur Gupta and Jonathan Webster for several helpful discussions. This work began when the second author was an undergraduate student at Butler University.

References

- [1] A. O. L. Atkin and D. J. Bernstein, Prime sieves using binary quadratic forms, *Math. Comp.* **73** (246) (2004), 1023–1030.
- [2] E. Bach and J. Shallit, *Algorithmic Number Theory. Vol. 1*, MIT Press, Cambridge, MA, 1996.
- [3] W. F. Galway, Dissecting a sieve to cut its need for space, in W. Bosma, editor, *Proceedings of the 4th International Algorithmic Number Theory Symposium (ANTS-IV)*, Springer, Berlin, Germany, 2000, pp. 297–312, LNCS 1838, ISBN 3-540-67695-3.
- [4] R. K. Guy, *Unsolved Problems in Number Theory*, Springer-Verlag, New York, 2004.
- [5] F. Luca, Multiply gleeful numbers, *Fibonacci Quart.*, to appear.
- [6] F. Luca, Multiply gleeful numbers, II, preprint.
- [7] G. L. Miller, Riemann’s hypothesis and tests for primality, *J. Comput. System Sci.* **13** (3) (1976), 300–317.
- [8] S. Moore and J. P. Sorenson, Explicit bounds and parallel algorithms for counting multiply gleeful numbers, *Integers* **26** (2026). #A84, 18 pp.

- [9] L. Moser, Notes on number theory. III. On the sum of consecutive primes, *Canad. Math. Bull.* **6** (1963), 159–161.
- [10] OEIS Foundation Inc., The On-Line Encyclopedia of Integer Sequences, <https://oeis.org>.
- [11] C. O’Sullivan, J. P. Sorenson, and A. Stahl, Algorithms and bounds on the sums of powers of consecutive primes, *Integers* **24** (2024), #A4, 14 pp.
- [12] M. O. Rabin, Probabilistic algorithm for testing primality, *J. Number Theory* **12** (1) (1980), 128–138.
- [13] C. Rivera, Puzzle 46.- primes expressible as sum of consecutive primes in k ways, https://primepuzzles.net/puzzles/puzz_046.htm.
- [14] J. P. Sorenson, The pseudosquares prime sieve, in F. Hess, S. Pauli, and M. Pohst, editors, *Proceedings of the 7th International Symposium on Algorithmic Number Theory (ANTS-VII)*, Springer, Berlin, Germany, 2006, pp. 193–207, LNCS 4076, ISBN 3-540-36075-1.
- [15] J. P. Sorenson, Two compact incremental prime sieves, *LMS Journal of Computation and Mathematics* **18** (1) (2015), 675–683.
- [16] J. Tongsomporn, S. Wananiyakul, and J. Steuding, Sums of consecutive prime squares, *Integers* **22** (2022), #A9, 5 pp.
- [17] E. Waiss, *GitHub*, <https://eleanorwaiss.github.io/gleeful>.