

# NOTES ON UPPER AND LOWER BOUNDING $t(N)$

TERENCE TAO, ...

## 1. INTRODUCTION

Given a natural number  $M$ , define a *factorization* of  $M$  to be a finite multiset  $\mathcal{B}$  such that the product

$$\prod \mathcal{B} := \prod_{a \in \mathcal{B}} a$$

(where the product is counted with multiplicity) is equal to  $M$ ; more generally, define a *subfactorization* of  $M$  to be a finite multiset  $\mathcal{B}$  such that  $\prod \mathcal{B}$  divides  $M$ . Given a threshold  $t$ , we say that a multiset  $\mathcal{B}$  is  *$t$ -admissible* if  $a \geq t$  for all  $a \in \mathcal{B}$ . For a given natural number  $N$ , we then define  $t(N)$  to be the largest  $t$  for which there exists a  $t$ -admissible factorization  $\mathcal{B}$  of  $N!$  of cardinality  $|\mathcal{B}| = N$ .

**Example 1.1.** The multiset

$$\{3, 3, 3, 3, 4, 4, 5, 7, 8\}$$

is a 3-admissible factorization of  $9!$  of cardinality 9, hence  $t(9) \geq 3$ . One can check that no 4-admissible factorization of  $9!$  of this cardinality exists, hence  $t(9) = 3$ .

The first few elements of this sequence are

$$1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, \dots$$

and the values of  $t(N)$  for  $N \leq 79$  were computed in [10] and also found at OEIS A034258. It is easy to see that  $t(N)$  is non-decreasing in  $N$  (basically because any cardinality  $N$  factorization of  $N!$  can be extended to a cardinality  $N + 1$  factorization of  $(N + 1)!$  by adding  $N + 1$  to the multiset). The values for  $N \leq 200$  can also be recovered from the entries of the inverse sequence of  $t$  at OEIS A034259.

When the factorial  $N!$  is replaced with an arbitrary number, this problem is essentially the bin covering problem, which is known to be NP-hard; see e.g., [2]. However, as we shall see in this paper, the special structure of the factorial (and in particular, the profusion of small prime factors such as 2 and 3) make it more tractable than the general case.

**Remark 1.2.** One can equivalently define  $t(N)$  as the greatest  $t$  for which there exists a  $t$ -admissible *subfactorization* of  $N!$  of cardinality *at least*  $N$ . This is because every such subfactorization can be converted into a  $t$ -admissible factorization of cardinality exactly  $N$  by first deleting elements from the subfactorization to make the cardinality  $N$ , and then multiplying one of the elements of the subfactorization by a natural number to upgrade the subfactorization to a factorization. This “relaxed” formulation of the problem turns out to be more convenient for both theoretical analysis of  $t(N)$  and numerical computations.

By combining the obvious lower bound

$$\prod B \geq t^{|B|} \quad (1.1)$$

for any  $t$ -admissible tuple with Stirling's formula (2.2), we obtain the trivial upper bound

$$\frac{t(N)}{N} \leq \frac{(N!)^{1/N}}{N} = \frac{1}{e} + O\left(\frac{\log N}{N}\right) \quad (1.2)$$

for  $N \geq 2$ ; see Figure 1. In [9, p.75] it was reported that an unpublished work of Erdős, Selfridge, and Straus established the asymptotic

$$\frac{t(N)}{N} = \frac{1}{e} + o(1) \quad (1.3)$$

(first conjectured in [7]) and asked if one could show the bound

$$\frac{t(N)}{N} \leq \frac{1}{e} - \frac{c}{\log N} \quad (1.4)$$

for some constant  $c > 0$  (problem #391 in <https://www.erdosproblems.com>; see also [10, Section B22, p. 122–123]); it was also noted that similar results were obtained in [1] if one restricted the  $a_i$  to be prime powers. However, as later reported in [8], Erdős “believed that Straus had written up our proof [of (1.3)]. Unfortunately Straus suddenly died and no trace was ever found of his notes. Furthermore, we never could reconstruct our proof, so our assertion now can be called only a conjecture”. In [10] the lower bound  $\frac{t(N)}{N} \geq \frac{1}{4}$  was established for sufficiently large  $N$ , by rearranging powers of 2 and 3 in the obvious factorization  $1 \times 2 \times \cdots \times N$  of  $N!$ . A variant lower bound of the asymptotic shape  $\frac{t(N)}{N} \geq \frac{3}{16} - o(1)$  obtained by rearranging only powers of 2, and which is superior for medium values of  $N$ , can also be found in [10]. The following conjectures in [10] were also made:

- (1) One has  $t(N) \leq N/e$  for  $N \neq 1, 2, 4$ .
- (2) One has  $t(N) \geq \lfloor 2N/7 \rfloor$  for  $N \neq 56$ .
- (3) One has  $t(N) \geq N/3$  for  $N \geq 300000$ . (It was also asked if the threshold 300000 could be lowered.)

In this paper we answer all of these questions.

**Theorem 1.3** (Main theorem). *Let  $N$  be a natural number.*

- (i) *If  $N \neq 1, 2, 4$ , then  $t(N) \leq N/e$ .*
- (ii) *If  $N \neq 56$ , then  $t(N) \geq \lfloor 2N/7 \rfloor$ .*
- (iii) *If  $N \geq 43632$ , then  $t(N) \geq N/3$ . The threshold 43632 is best possible.*
- (iv) *For large  $N$ , one has*

$$\frac{t(N)}{N} = \frac{1}{e} - \frac{c_0}{\log N} + O\left(\frac{1}{\log^{1+c} N}\right) \quad (1.5)$$

*for some constant  $c > 0$ , where  $c_0$  is the explicit quantity*

$$\begin{aligned} c_0 &:= \frac{1}{e} \int_0^1 \left\lfloor \frac{1}{x} \right\rfloor \log \left( ex \left\lceil \frac{1}{ex} \right\rceil \right) dx \\ &= 0.3044190 \dots \end{aligned} \quad (1.6)$$

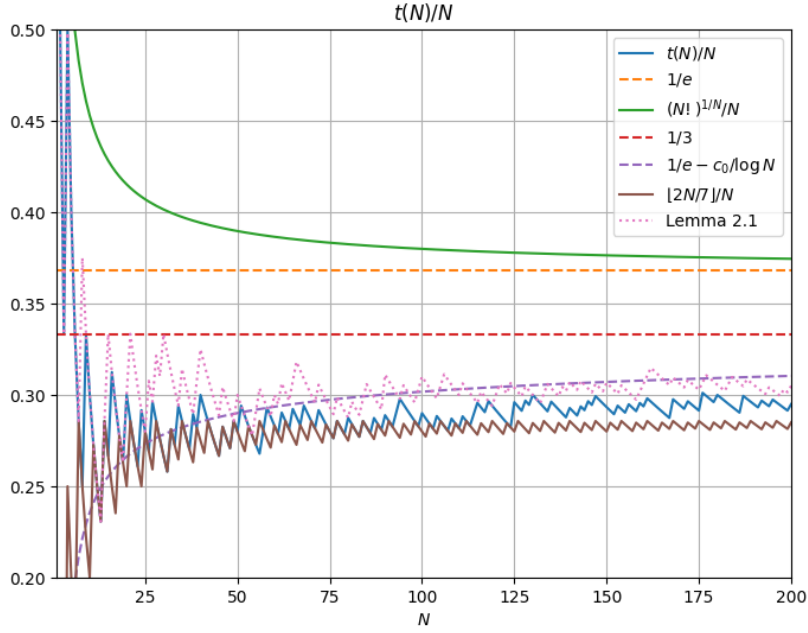


FIGURE 1. The function  $t(N)/N$  (blue) for  $N \leq 200$ , using the data from OEIS A034258, as well as the trivial upper bound  $(N!)^{1/N}/N$  (green), the improved upper bound from Lemma 3.1 (pink), which is asymptotic to  $(1.5)$  (purple), and the function  $[2N/7]/N$  (brown), which is a lower bound for  $N \neq 56$  [10]. Theorem 1.3 implies that  $t(N)/N$  is asymptotic to  $(1.5)$  (purple), which in turn converges to  $1/e$  (orange). The threshold  $1/3$  (red) is permanently crossed at  $N = 43632$ .

(see Figure 3). In particular, (1.3) and (1.4) hold.

In Appendix D we give some details on the numerical computation of the constant  $c_0$ .

**Remark 1.4.** In a previous version [14] of this manuscript, the weaker bounds

$$\frac{1}{e} - \frac{O(1)}{\log N} \leq \frac{t(N)}{N} \leq \frac{1}{e} - \frac{c_0 + o(1)}{\log N}$$

were established, which were enough to recover (1.3), (1.4), and Theorem 1.3(i).

As one might expect, the proof of Theorem 1.3 proceeds by a combination of both theoretical analysis and numerical calculations. Our main tools to obtain upper and lower bounds on  $t(N)$  can be summarized as follows:

- An *accounting identity* linking the “ $t$ -excess” of a subfactorization with its “ $p$ -surpluses” at various primes, which provides an reasonable upper bound on  $t(N)$  for all  $N$ ;
- A *linear programming* (or *integer programming*) method that provides quite accurate upper and lower bounds on  $t(N)$  for small and medium values of  $N$ ;

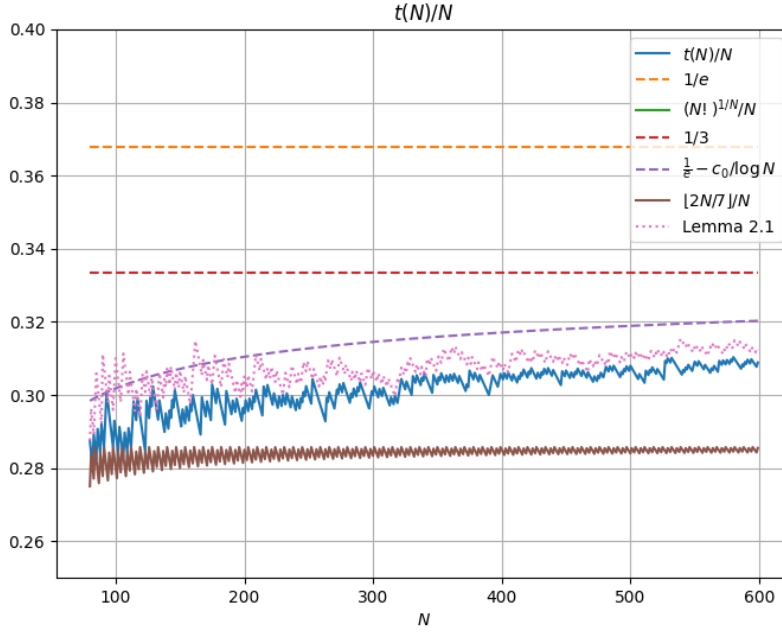


FIGURE 2. A continuation of Figure 1 to the region  $80 \leq N \leq 599$ .

- *Greedy algorithms* to construct subfactorizations, that provide quickly computable, though suboptimal, lower bounds on  $t(N)$  for small and medium values;
- A *modified approximate factorization* strategy, which provides lower bounds on  $t(N)$ , that become asymptotically quite efficient.

We now briefly discuss each of these methods in more detail.

**1.1. The accounting identity.** We use  $v_p(a/b) = v_p(a) - v_p(b)$  to denote the  $p$ -adic valuation of a positive natural number  $a/b$ , that is to say the number of times  $p$  divides the numerator  $a$ , minus the number of times  $p$  divides the denominator  $b$ . For instance,  $v_2(32/27) = 5$  and  $v_3(32/27) = -3$ . If one applies a logarithm to the fundamental theorem of arithmetic, one obtains the identity<sup>1</sup>

$$\sum_p v_p(r) \log p = \log r \quad (1.7)$$

for any positive rational  $r$ . Given a  $t$ -admissible subfactorization  $\mathcal{B}$  of  $N!$ , we can apply the above identity to  $r := N! / \prod \mathcal{B}$  and rearrange to obtain the *accounting identity*

$$\mathcal{E}_t(\mathcal{B}) + \sum_p v_p \left( \frac{N!}{\prod \mathcal{B}} \right) \log p = \log N! - |\mathcal{B}| \log t \quad (1.8)$$

where we define the  $t$ -excess  $\mathcal{E}_t(\mathcal{B})$  of the multiset  $\mathcal{B}$  by the formula

$$\mathcal{E}_t(\mathcal{B}) := \sum_{a \in \mathcal{B}} \log \frac{a}{t}. \quad (1.9)$$

<sup>1</sup>Throughout this paper, the symbol  $p$  (or  $p'$ ,  $p_1$ ,  $p_2$ , etc.) is always understood to be restricted to be prime.

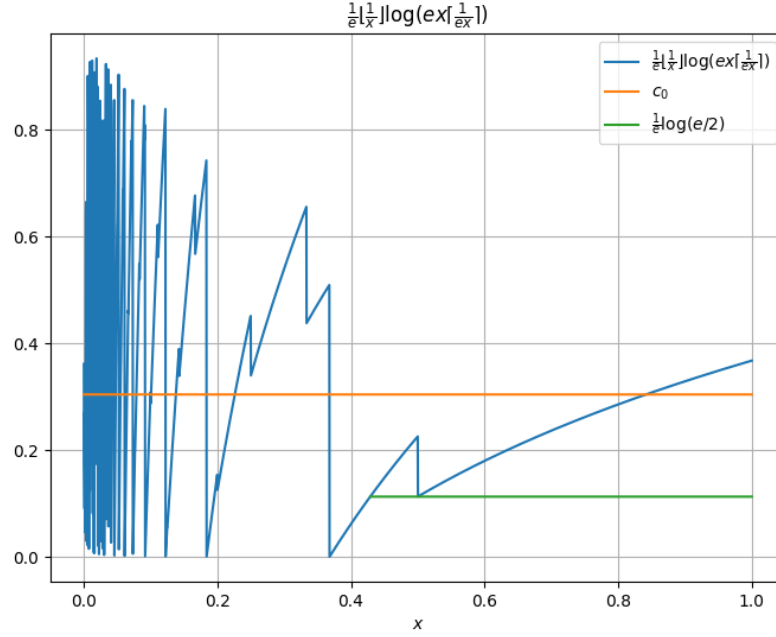


FIGURE 3. The piecewise continuous function  $x \mapsto \frac{1}{e} \left\lfloor \frac{1}{x} \right\rfloor \log\left(ex \left\lfloor \frac{1}{ex} \right\rfloor\right)$ , together with its mean value  $c_0 = 0.3044190 \dots$ . The function exhibits an oscillatory singularity at  $x = 0$  similar to  $\sin \frac{1}{x}$  (but it is always nonnegative and bounded). We also display the (crude) lower bound of  $\frac{1}{e} \log(e/2)$  for  $x \geq \frac{1}{\sqrt{2e}} = 0.4288 \dots$ . Informally, this function quantifies the difficulty that large primes in the factorization of  $N!$  have in becoming slightly larger than  $N/e$  after multiplying by a natural number.

We refer to the quantity  $v_p \left( \frac{N!}{\prod \mathcal{B}} \right)$  as the  $p$ -surplus of  $\mathcal{B}$  (as a proposed factorization of  $N!$ ); if it is negative, we refer to  $-v_p \left( \frac{N!}{\prod \mathcal{B}} \right)$  as the  $p$ -deficit, with the multiset being  $p$ -balanced if the  $p$ -surplus (or  $p$ -deficit) is zero. Note that the  $t$ -excess will be non-negative if  $\mathcal{B}$  is  $t$ -admissible, and the  $p$ -surpluses will be non-negative if  $\mathcal{B}$  is a subfactorization of  $N!$ .

**Example 1.5.** Suppose one wishes to factorize  $5! = 2^3 \times 3 \times 5$ . The attempted 3-admissible factorization  $\mathcal{B} := \{3, 4, 5, 5\}$  has a 2-surplus of  $v_2(5! / \prod \mathcal{B}) = 1$ , is in balance at 3, and has a 5-deficit of  $v_5(\prod \mathcal{B} / 5!) = 1$ , so it is not a factorization or subfactorization of  $5!$ . The 3-excess of this multiset is

$$\mathcal{E}_3(\mathcal{B}) = \log \frac{3}{3} + \log \frac{4}{3} + \log \frac{5}{3} + \log \frac{5}{3} = 1.3093 \dots$$

and the accounting identity (1.8) become

$$1.3093 \dots + \log 2 - \log 5 = 0.3930 \dots = \log 5! - 4 \log 3.$$

If one replaces one of the copies of 5 in  $\mathcal{B}$  with a 2, this erases both the 2-surplus and the 5-deficit, and creates a factorization  $\mathcal{B}' = \{2, 3, 4, 5\}$  of  $5!$ ; the 3-excess now drops to

$$\mathcal{E}_3(\mathcal{B}) = \log \frac{2}{3} + \log \frac{3}{3} + \log \frac{4}{3} + \log \frac{5}{3} = 0.3930 \dots,$$

bringing the accounting identity back into balance.

In view of Remark 1.2, one can now equivalently describe  $t(N)$  as follows:

**Lemma 1.6** (Equivalent description of  $t(N)$ ).  *$t(N)$  is the largest quantity  $t$  for which there exists a  $t$ -admissible subfactorization of  $N!$  with*

$$\mathcal{E}_t(\mathcal{B}) + \sum_p v_p \left( \frac{N!}{\prod \mathcal{B}} \right) \log p \leq \log N! - N \log t.$$

One can view  $\log N! - N \log t$  as an available “budget” that one can “spend” on some combination of  $t$ -excess and  $p$ -surpluses. For  $t$  of the form  $t = N/e^{1+\delta}$  for some  $\delta > 0$ , the budget can be computed using the Stirling approximation (2.2) to be  $\delta N + O(\log N)$ . The non-negativity of the  $t$ -excess and  $p$ -surpluses recovers the trivial upper bound (1.2); but one can improve upon this bound by observing that large prime factors of  $N!$  inevitably generate a noticeable  $t$ -excess. In ??? we develop this idea to establish the upper bound portions of Theorem 1.3, namely part (i) and the upper bound in part (iv).

## 1.2. Linear programming.

## 1.3. Greedy algorithms.

## 1.4. Modified approximate factorizations. ...

**1.5. Author contributions and data.** This project was initially conceived as a single-author manuscript by Terence Tao, but since the release of the initial preprint [14], grew to become a collaborative project organized via the Github repository [15], which also contains the supporting code and data for the project. The contributions of the individual authors, according to the CRediT categories at <https://credit.niso.org/>, are as follows:

**authors should be arranged in alphabetical order of surname.**

- ...
- Terence Tao: Conceptualization, Formal Analysis, Methodology, Project Administration, Visualization, Writing – original draft, Writing – review & editing.

## 1.6. Acknowledgments. ...

## 2. BASICS

The symbol  $p$  will always denote a prime. The primes 2, 3 will play a special role here and will be referred to as *tiny primes*.

We recall Legendre's formula

$$v_p(N!) = \sum_{j=1}^{\infty} \left\lfloor \frac{N}{p^j} \right\rfloor = \frac{N - s_p(N)}{p - 1}. \quad (2.1)$$

To bound the factorial, we have the explicit Stirling approximation [12]

$$N \log N - N + \log \sqrt{2\pi N} + \frac{1}{12N + 1} \leq \log N! \leq N \log N - N + \log \sqrt{2\pi N} + \frac{1}{12N}, \quad (2.2)$$

valid for all natural numbers  $N$ .

In addition to the usual asymptotic notation, we use  $O_{\leq}(X)$  to denote any quantity whose magnitude is bounded by at most  $X$  (note the absence of an additional constant factor).

3. CRITERIA FOR UPPER BOUNDING  $t(N)$ 

We have the trivial upper bound  $t(N) \leq (N!)^{1/N}$ . This can be improved to  $t(N) \leq N/e$  for  $N \neq 1, 2, 4$ , answering a conjecture of Guy and Selfridge [11]; see [14]. This was derived from the following slightly stronger criterion, which asymptotically gives  $\frac{t(N)}{N} \leq \frac{1}{e} - \frac{c_0 + o(1)}{\log N}$ :

**Lemma 3.1** (Upper bound criterion). [14, Lemma 2.1] *Suppose that  $1 \leq t \leq N$  are such that*

$$\sum_{\substack{t \\ p > \frac{t}{\lfloor \sqrt{t} \rfloor}}} \left\lfloor \frac{N}{p} \right\rfloor \log \left( \frac{p}{t} \left\lceil \frac{t}{p} \right\rceil \right) > \log N! - N \log t \quad (3.1)$$

*Then  $t(N) < t$ .*

A surprisingly sharp upper bound comes from linear programming.

**Lemma 3.2** (Linear programming bound). *Let  $N$  be a natural number and  $1 \leq t \leq N/2$ . Suppose for each prime  $p \leq N$ , one has a non-negative real number  $w_p$  which is weakly non-decreasing in  $p$  (thus  $w_p \leq w_{p'}$  when  $p \leq p'$ ), and such that*

$$\sum_p w_p v_p(j) \geq 1 \quad (3.2)$$

*for all  $t \leq j \leq N$ , and such that*

$$\sum_p w_p v_p(N!) < N. \quad (3.3)$$

*Then  $t(N) < t$ .*

*Proof.* We first observe that the bound (3.2) in fact holds for all  $j \geq t$ , not just for  $t \leq j \leq N$ . Indeed, if this were not the case, consider the first  $j \geq t$  where (3.2) fails. Take a prime  $p$  dividing  $j$  and replace it by a prime in the interval  $[p/2, p)$  which exists by Bertrand's postulate (or remove  $p$  entirely, if  $p = 2$ ); this creates a new  $j'$  in  $[j/2, j)$  which is still at least  $t$ . By the weakly decrasing hypothesis on  $w_p$ , we have

$$\sum_p w_p v_p(j) \geq \sum_p w_p v_p(j')$$

and hence by the minimality of  $j$  we have

$$\sum_p w_p v_p(j) > 1,$$

a contradiction.

Now suppose for contradiction that  $t(N) \geq t$ , thus we have a factorization  $N! = \prod_{j \geq t} j^{m_j}$  for some natural numbers  $m_j$  summing to  $N$ . Taking  $p$ -valuations, we conclude that

$$\sum_{j \geq t} m_j v_p(j) \leq v_p(N!)$$

for all  $p \leq N$ . Multiplying by  $w_p$  and summing, we conclude from (3.2) that

$$N = \sum_{j \geq t} m_j \leq \sum_p w_p v_p(N!),$$

contradicting (3.3). □

This bound is sharp for all  $N \leq 600$ , with the exception of  $N = 155$ , where it gives the upper bound  $t(155) \leq 46$ . A more precise integer program gives  $t(155) = 45$ .

**Remark 3.3.** A variant of the linear programming method also gives good lower bound constructions. Specifically, one can use linear programming to find non-negative real numbers  $m_j$  for  $t \leq j \leq N$  that maximize the quantity  $\sum_{t \leq j \leq N} m_j$  subject to the constraints

$$\sum_{t \leq j \leq N} m_j v_p(j) \leq v_p(N!).$$

The expression  $\prod_{t \leq j \leq N} j^{\lfloor m_j \rfloor}$  will then be a subfactorization of  $N!$  into  $\sum_{t \leq j \leq N} \lfloor m_j \rfloor$  factors  $j$ , each of which is at least  $t$ . If  $\sum_{t \leq j \leq N} \lfloor m_j \rfloor \geq N$ , this demonstrates that  $t(N) \geq t$ . Numerically, this procedure attains the exact value of  $t(N)$  for all  $N \leq 600$ ; for instance for  $N = 155$ , it shows that  $t(155) \geq 45$ .

**3.1. Asymptotic analysis of upper bound.** We refine the upper bound in [14] slightly.

**Proposition 3.4.** *For large  $N$ , one has*

$$\frac{t(N)}{N} \leq \frac{1}{e} - \frac{c_0}{\log N} + O\left(\frac{1}{\log^2 N}\right).$$



*Proof.* We apply Lemma 3.1 with

$$t := \frac{1}{e} - \frac{c_0}{\log N} + \frac{C_0}{\log^2 N}$$

with  $C_0$  a large absolute constant to be chosen later. From the Stirling approximation one sees that

$$\log N! - N \log t \geq ec_0 \frac{N}{\log N} + (C_0 - O(1)) \frac{N}{\log^2 N}$$

so it will suffice to establish the upper bound

$$\sum_{p > \frac{t}{\lfloor \sqrt{t} \rfloor}} \left\lfloor \frac{N}{p} \right\rfloor \log \left( \frac{p}{t} \left\lceil \frac{t}{p} \right\rceil \right) \leq ec_0 \frac{N}{\log N} + O \left( \frac{N}{\log^2 N} \right).$$

For  $N$  large enough, we have  $\frac{t}{\lfloor \sqrt{t} \rfloor} \leq \frac{N}{\log N}$ , so it suffices to show that

$$\sum_{\frac{N}{\log N} \leq p \leq N} \left\lfloor \frac{N}{p} \right\rfloor \log \left( \frac{p}{t} \left\lceil \frac{t}{p} \right\rceil \right) \leq ec_0 \frac{N}{\log N} + O \left( \frac{N}{\log^2 N} \right).$$

The summand is a piecewise monotone function of  $p$ , with  $O(\log N)$  pieces, and bounded in size by  $O(N)$ . A routine application of the prime number theorem (with classical error term) and summation by parts then allows one to express the left-hand side as

$$\int_{N/\log N}^N \left\lfloor \frac{N}{x} \right\rfloor \log \left( \frac{x}{t} \left\lceil \frac{t}{x} \right\rceil \right) \frac{dx}{\log x} + O \left( \frac{N}{\log^2 N} \right)$$

(in fact the error term can be made much stronger than this). We use the approximation

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

To control the error term, we observe from Taylor expansion that

$$\log \left( \frac{x}{t} \left\lceil \frac{t}{x} \right\rceil \right) \ll \frac{\left\lceil \frac{t}{x} \right\rceil - \frac{t}{x}}{t/x} \ll \frac{x}{t} \ll \frac{x}{N} \quad (3.4)$$

and the contribution of the error term is

$$\ll \int_{N/\log N}^N \frac{N}{x} \frac{x}{N} \frac{\log(N/x)}{\log^2 N} dx \ll \frac{N}{\log^2 N}$$

which is acceptable. As for the main term, we can rescale it to

$$\frac{et}{\log N} \int_{N/et \log N}^{N/et} \left\lfloor \frac{N/et}{x} \right\rfloor \log \left( ex \left\lceil \frac{1}{ex} \right\rceil \right) dx.$$

Since  $N/et = 1 + O(1/\log N)$ , we see that the integrand here is within  $O(1/\log N)$  of  $\left\lfloor \frac{1}{x} \right\rfloor \log \left( ex \left\lceil \frac{1}{ex} \right\rceil \right)$  unless  $\frac{1}{x}$  is within  $O(1/\log N)$  of an integer, which one can calculate to occur on a set of measure zero. A variant of (3.4) shows that both integrands are bounded by  $O(1)$  for all  $x \in [0, N/et]$ , so by the triangle inequality the above expression can be rewritten as

$$\frac{N}{\log N} \int_0^1 \left\lfloor \frac{1}{x} \right\rfloor \log \left( ex \left\lceil \frac{1}{ex} \right\rceil \right) dx + O \left( \frac{N}{\log^2 N} \right),$$

and the claim follows from (??). □

#### 4. A GENERAL FACTORIZATION ALGORITHM

In this section we present and then analyze an algorithm that, when given parameters  $1 \leq t \leq N$ , will attempt to construct a factorization  $N! = \prod \mathcal{B}$  of  $N!$  by a finite multiset  $\mathcal{B}$  of  $N$  elements that are all at least  $t$ . The algorithm will not always succeed, but when it does, it will certify that  $t(N) \geq t$ .

**4.1. Notational preliminaries.** We begin with some key definitions.

By the fundamental theorem of arithmetic (or (1.7)), we see that a finite multiset  $\mathcal{B}$  is a factorization of  $M$  if and only if

$$v_p(M / \prod \mathcal{B}) = 0$$

for all primes  $p$ , and a subfactorization if and only if

$$v_p(M / \prod \mathcal{B}) \geq 0$$

for all primes  $p$ . We refer to  $v_p(M / \prod \mathcal{B})$  as the  $p$ -surplus of  $\mathcal{B}$  (as an attempted factorization of  $M$ ) at prime  $p$ , and  $-v_p(M / \prod \mathcal{B}) = v_p(\prod \mathcal{B} / M)$  as the  $p$ -deficit, and say that the factorization is  $p$ -balanced if  $v_p(M / \prod \mathcal{B}) = 0$ . Thus a subfactorization (resp. factorization) occurs when one has non-negative surpluses (resp. balance) at all primes  $p$ .

**Example 4.1.** Suppose one wishes to factorize  $5! = 2^3 \times 3 \times 5$ . The attempted factorization  $\mathcal{B} := \{3, 4, 5, 5\}$  has a 2-surplus of  $v_2(5! / \prod \mathcal{B}) = 1$ , is in balance at 3, and has a 5-deficit of  $v_5(\prod \mathcal{B} / 5!) = 1$ , so it is not a factorization or subfactorization of  $5!$ . However, if one replaces one of the copies of 5 in  $\mathcal{B}$  with a 2, this erases both the 2-surplus and the 5-deficit, and creates a factorization  $\{2, 3, 4, 5\}$  of  $5!$ .

A finite multiset  $\mathcal{B}$  is said to be  $t$ -admissible for some  $t > 0$  if  $b \geq t$  for all  $b \in \mathcal{B}$ . Then  $t(N)$  is largest quantity such that there exists a  $t(N)$ -admissible factorization of  $N!$  of cardinality  $N$ .

Call a natural number 3-smooth if it is of the form  $2^n 3^m$  for some natural numbers  $n, m$ . Given a positive real number  $x$ , we use  $\lceil x \rceil^{(2,3)}$  to denote the smallest 3-smooth number greater than or equal to  $x$ . For instance,  $\lceil 5 \rceil^{(2,3)} = 6$  and  $\lceil 10 \rceil^{(2,3)} = 12$ .

**4.2. Description of algorithm.** We now describe an algorithm that, for given  $1 \leq t \leq N$ , either successfully demonstrates that  $t(N) \geq t$ , or halts with an error.

(0) Select natural numbers  $A, K$  such that  $K^2(1 + \sigma) < t$ , where

$$\sigma := \frac{3N/t}{A}. \quad (4.1)$$

There is some freedom to select parameters here, but roughly speaking one would like to have  $1 \lll A \lll K \lll \sqrt{t}$ .

- (1) Let  $I$  denote the elements of the interval<sup>2</sup>  $(t, t(1 + \sigma)]$  that are coprime to 6. Let  $\mathcal{B}^{(1)}$  be the elements of  $I$ , each occurring with multiplicity  $A$ . This multiset is  $t$ -admissible, and  $\prod \mathcal{B}^{(1)}$  is not divisible by tiny primes 2, 3. (It will have approximately the right number of primes for  $3 < p \leq t/K$ , though it may have quite different prime factorization at primes  $p > t/K$ .)
- (2) Remove any element from  $\mathcal{B}^{(1)}$  that contains a prime factor  $p$  with  $p > t/K$ , and call this new multiset  $\mathcal{B}^{(2)}$ . It remains  $t$ -admissible with no tiny prime factors, though it tends to acquire a  $p$ -surplus in the range  $3 < p \leq K$ .
- (3) For each  $p > t/K$ , add in  $v_p(N!)$  copies of the number  $p \lceil t/p \rceil$  to  $\mathcal{B}^{(2)}$ , and call this new multiset  $\mathcal{B}^{(3)}$ . Now  $\mathcal{B}^{(3)}$  is  $t$ -admissible and in balance at all primes  $p > t/K$ , but will typically be in a slight deficit at primes  $3 < p \leq t/K$ , particularly in the range  $3 < p \leq K$ . (It will now also contain a few tiny prime factors, but will generally still have a large surplus at those primes.)
- (4) For each prime  $3 < p \leq t/K$  at which there is a surplus  $v_p(N! / \prod \mathcal{B}) > 0$ , replace  $v_p(N! / \prod \mathcal{B})$  copies of  $p$  in the prime factorizations of elements of  $\mathcal{B}^{(3)}$  with  $\lceil p \rceil^{(2,3)}$  instead, and call this new multiset  $\mathcal{B}^{(4)}$ . Thus  $\mathcal{B}^{(4)}$  has no surplus at primes  $3 < p \leq t/K$  (and is still  $t$ -admissible and in balance for  $p > t/K$ ).
- (5) For the primes  $3 < p \leq t/K$  at which there is a deficit  $v_p(\prod \mathcal{B} / N!) > 0$ , multiply all these primes together, and use the greedy algorithm to group them into factors  $x_1, \dots, x_M$  in the range  $(\sqrt{t/K}, t/K]$ , together with possibly one exceptional factor  $x_*$  in the range  $(1, t/K]$ . For each of these factors  $x_i$  or  $x_*$ , add the quantity  $x_i \lceil t/x_i \rceil^{(2,3)}$  or  $x_* \lceil t/x_* \rceil^{(2,3)}$  to  $\mathcal{B}^{(4)}$ , and call this new multiset  $\mathcal{B}^{(5)}$ .
- (6) By construction,  $\mathcal{B}^{(5)}$  is  $t$ -admissible and will be in balance at all primes  $p > 3$ , and is thus  $N! / \prod \mathcal{B}^{(5)}$  is of the form  $2^n 3^m$  for some integers  $n, m$ . If at least one of  $n, m$  is negative, then HALT the algorithm with an error. Otherwise, select a 3-smooth number  $2^{n_1} 3^{m_1}$  greater than equal to  $t$  with  $n_1/m_1 \leq n/m$  (which one can interpret as  $n_1 m \leq n m_1$  in case some of the denominators here vanish), and similarly select a 3-smooth number  $2^{n_2} 3^{m_2}$  greater than or equal to  $t$  with  $n_2/m_2 \geq n/m$ . (It is reasonable to select the smallest such 3-smooth numbers in both cases, although this is not absolutely necessary for the algorithm to be successful.) By construction, we can express  $(n, m)$  as a positive linear combination  $\alpha_1(n_1, m_1) + \alpha_2(n_2, m_2)$  of  $(n_1, m_1)$  and  $(n_2, m_2)$ . Add  $\lfloor \alpha_1 \rfloor$  copies of  $2^{n_1} 3^{m_1}$  and  $\lfloor \alpha_2 \rfloor$  copies of  $2^{n_2} 3^{m_2}$  to  $\mathcal{B}^{(5)}$ , and call this tuple  $\mathcal{B}^{(6)}$ . (This will largely eliminate the surplus at 2 and 3.)
- (7) If the multiset  $\mathcal{B}^{(6)}$  has cardinality less than  $N$ , HALT the algorithm with an error. Otherwise, delete elements from  $\mathcal{B}^{(6)}$  to bring the cardinality to  $N$ , and arbitrarily distribute any surplus primes to one of the remaining elements, and call the resulting multiset  $\mathcal{B}^{(7)}$ . By construction,  $\mathcal{B}^{(7)}$  is a  $t$ -admissible factorization of  $N!$  into  $N$  numbers, demonstrating that  $t(N) \geq t$ .

It will be convenient to divide the set of primes into the following ranges:

- *Tiny primes*  $p = 2, 3$ .
- *Small primes*  $3 < p \leq K$ .

<sup>2</sup>Numerically, it would be slightly better to use the closed interval  $[t, t(1 + \sigma)]$  instead of the half-open interval  $(t, t(1 + \sigma)]$ , but we will consistently aim to use half-open intervals here to be compatible with standard notation for the prime counting function  $\pi(x)$ .

- *Borderline small primes*  $K < p \leq K(1 + \sigma)$ .
- *Medium primes*  $K(1 + \sigma) < p \leq t/K$ .
- *Large primes*  $p > t/K$ .

The expected  $p$ -surpluses or  $p$ -deficits at various stages of this process are summarized in Table 1.

	Tiny $p$	Small $p$	Borderline $p$	Medium $p$	Large $p$
$\mathcal{B}^{(1)}$	Max. surplus	Near balance	Near balance	Near balance	???
$\mathcal{B}^{(2)}$	Max. surplus	Med. surplus	Med. surplus?	Near balance	Max. surplus
$\mathcal{B}^{(3)}$	Lg. surplus	Sm. surplus?	Med. surplus?	Near balance	Balance
$\mathcal{B}^{(4)}$	Lg. surplus	Balance?	Balance?	Balance/sm. deficit	Balance
$\mathcal{B}^{(5)}$	Lg. surplus	Balance	Balance	Balance	Balance
$\mathcal{B}^{(6)}$	Sm. surplus	Balance	Balance	Balance	Balance
$\mathcal{B}^{(7)}$	Balance	Balance	Balance	Balance	Balance

TABLE 1. Evolution of the surpluses and deficits of the multisets  $\mathcal{B}^{(i)}$ ,  $i = 1, \dots, 7$ ; we describe the size of these surpluses and deficits informally as “small”, “medium”, “large”, or “maximal”. For entries with a question mark, we allow the possibility of a tiny deficit. For the entry marked ???, all behavior from large surpluses to large deficits are possible.

**4.3. Analysis of Step 7.** We now analyze the above algorithm, starting from the final Step 7 and working backwards to Step 1, to establish sufficient conditions for the algorithm to successfully demonstrate that  $t(N) \geq t$ .

It will be convenient to introduce the following notation. For  $a_+, a_- \in [0, +\infty]$ , we define the asymmetric norm  $|x|_{a_+, a_-}$  of a real number  $x$  by the formula

$$|x|_{a_+, a_-} := \begin{cases} a_+ |x| & x \geq 0 \\ a_- |x| & x \leq 0. \end{cases}$$

If  $a_+, a_-$  are finite, this function is Lipschitz with constant  $\max(a_+, a_-)$ . One can think of  $a_+$  as the “cost” of making  $x$  positive, and  $a_-$  as the “cost” of making  $x$  negative.

We now begin the analysis of Step 9. This procedure will terminate successfully as long as the length  $|\mathcal{B}^{(6)}|$  of the tuple is at least  $N$ . To ensure this, we introduce the  $t$ -excess of a multiset  $\mathcal{B}$  by the formula

$$\mathcal{E}_t(\mathcal{B}) := \prod_{b \in \mathcal{B}} \log \frac{b}{t} = \log \prod \mathcal{B} - |\mathcal{B}| \log t.$$

Thus, to ensure the success of this step, it suffices to establish the inequality

$$\mathcal{E}_t(\mathcal{B}^{(6)}) \leq \log \prod \mathcal{B}^{(6)} - N \log t.$$

From (1.7) we have

$$\log \prod \mathcal{B}^{(6)} = \log N! - \sum_p v_p \left( \frac{N!}{\prod \mathcal{B}^{(6)}} \right) \log p,$$

so we can rewrite the previous condition (using the fact that  $B^{(6)}$  is a subfactorization of  $N!$ ) as

$$\mathcal{E}_t(B^{(6)}) + \sum_p \left| v_p \left( \frac{N!}{\prod B^{(6)}} \right) \right|_{\log p, \infty} \leq \log N! - N \log t.$$

If we assume that  $t = N/e^{1+\delta}$  for some  $\delta > 0$ , we can use the Stirling approximation (2.2) to reduce to the sufficient condition

$$\mathcal{E}_t(B^{(6)}) + \sum_p \left| v_p \left( \frac{N!}{\prod B^{(6)}} \right) \right|_{\log p, \infty} \leq \delta N + \log \sqrt{2\pi N}. \quad (4.2)$$

**4.4. Analysis of Step 6.** Now we analyze Step 6. For any  $L \geq 1$ , let  $\kappa_L$  be the least quantity such that

$$x \leq \lceil x \rceil^{(2,3)} \leq \exp(\kappa_L)x \quad (4.3)$$

holds for all  $x \geq L$ . Just from considering the powers of two, we have the trivial upper bound

$$\kappa_L \leq \log 2. \quad (4.4)$$

We shall obtain better estimates on this quantity in Section A. For now we use this quantity to help achieve efficient subfactorizations of 3-smooth numbers, as follows.

**Lemma 4.2.** *Let  $L \geq 1$ . Let  $t > 3L$  and let  $2^n 3^m$  be a 3-smooth number with  $n, m > 0$  obeying the conditions*

$$\frac{\log(3L) + \kappa_L}{\log t - \log(3L)} \leq \frac{n \log 2}{m \log 3} \leq \frac{\log t - \log(2L)}{\log(2L) + \kappa_L}. \quad (4.5)$$

*Then one can find a  $t$ -admissible subfactorization  $B$  of  $2^n 3^m$  such that*

$$\mathcal{E}_t(B) \leq \kappa_L \frac{n \log 2 + m \log 3}{\log t} \quad (4.6)$$

and

$$|v_2(2^n 3^m / B)|_{\log 2, \infty} + |v_3(2^n 3^m / B)|_{\log 3, \infty} \leq 2(\log t + \kappa_L). \quad (4.7)$$

In practice,  $\log t$  will be significantly larger than  $\log(2L)$  or  $\log(3L)$ , and so the hypothesis (4.5) will be quite mild, as long as  $n$  and  $m$  are both reasonably large.

*Proof.* Let  $2^{n_0}, 3^{m_0}$  be the largest powers of 2 and 3 less than or equal to  $t/L$  respectively, thus

$$L \leq \frac{t}{2^{n_0}} \leq 2L \quad (4.8)$$

and

$$L \leq \frac{t}{3^{m_0}} \leq 3L. \quad (4.9)$$

From (4.3), the 3-smooth numbers  $\lceil t/2^{n_0} \rceil^{(2,3)} = 2^{n_1} 3^{m_1}$ ,  $\lceil t/3^{m_0} \rceil^{(2,3)} = 2^{n_2} 3^{m_2}$  obey the estimates

$$\frac{t}{2^{n_0}} \leq 2^{n_1} 3^{m_1} \leq e^{\kappa_L} \frac{t}{2^{n_0}} \quad (4.10)$$

and

$$\frac{t}{3^{m_0}} \leq 2^{n_2} 3^{m_2} \leq e^{\kappa_L} \frac{t}{3^{m_0}}, \quad (4.11)$$

or equivalently

$$t \leq 2^{n_0+n_1} 3^{m_1}, 2^{n_2} 3^{m_0+m_2} \leq e^{\kappa_L} t. \quad (4.12)$$

We can use (4.8), (4.10) to bound

$$\begin{aligned} \frac{n_0 + n_1}{m_1} &\geq \frac{n_0}{\log(e^{\kappa_L} \frac{t}{2^{m_0}}) / \log 3} \\ &\geq \frac{(\log t - \log(2L)) / \log 2}{(\log(2L) + \kappa_L) / \log 3} \end{aligned}$$

(with the convention that this bound is vacuously true for  $m_1 = 0$ ). Similarly, from (4.9), (4.11) we have

$$\begin{aligned} \frac{n_2}{m_0 + m_2} &\leq \frac{\log(e^{\kappa} \frac{t}{3^{m_0}}) / \log 2}{m_0} \\ &\leq \frac{(\log(3L) + \kappa) / \log 2}{(\log t - \log(3L)) / \log 3} \end{aligned}$$

and hence by (4.5)

$$\frac{n_2}{m_0 + m_2} \leq \frac{n}{m} \leq \frac{n_0 + n_1}{m_1}. \quad (4.13)$$

Thus we can write  $(n, m)$  as a non-negative linear combination

$$(n, m) = \alpha_1(n_0 + n_1, m_1) + \alpha_2(n_2, m_0 + m_2)$$

for some real  $\alpha_1, \alpha_2 \geq 0$ . We now take our subfactorization  $\mathcal{B}$  to consist of  $\lfloor \alpha_1 \rfloor$  copies of the 3-smooth number  $2^{n_0+n_1}3^{m_1}$  and  $\lfloor \alpha_2 \rfloor$  copies of the 3-smooth number  $2^{n_2}3^{m_0+m_2}$ . By (4.12), each term  $2^{n'}3^{m'}$  here is admissible and contributes a  $t$ -excess of at most  $\kappa_L$ , which is in turn bounded by  $\kappa_L \frac{n' \log 2 + m' \log 3}{\log t}$ . Adding these bounds together, we obtain (4.6).

The expression  $2^{n_0+n_1}3^{m_1} / \prod \mathcal{B}$  contains at most  $n_0 + n_1 + n_2$  factors of 2 and at most  $m_0 + m_2 + m_1$  factors of 3, hence

$$v_2(2^{n_0+n_1}3^{m_1} / \prod \mathcal{B}) \log 2 + v_3(2^{n_2}3^{m_0+m_2} / \prod \mathcal{B}) \log 3 \leq \log 2^{n_0+n_1}3^{m_1} + \log 2^{n_2}3^{m_0+m_2},$$

and the bound (4.7) follows from (4.12).  $\square$

We now use this lemma to analyze Step 6 as follows.

**Proposition 4.3.** *Let  $L \geq 1$ . Let  $3L < t = N/e^{1+\delta}$  for some  $\delta > 0$ , and let  $1 \leq K \leq t$  and  $A \geq 1$ . Suppose that the algorithm in Section 4.2 with the indicated parameters reaches the end of Step 5 with a multiset  $\mathcal{B}^{(5)}$  obeying the following hypotheses:*

(i) *(Small excess and surplus at non-tiny primes)*

$$\mathcal{E}_t(\mathcal{B}^{(5)}) + \sum_{p>3} \left| v_p \left( \frac{N!}{\prod \mathcal{B}^{(5)}} \right) \right|_{\log p, \infty} \leq \delta N + \log \sqrt{2\pi} - \frac{3}{2} \log N - (\kappa_L \log \sqrt{12}) \frac{N}{\log t}. \quad (4.14)$$

(ii) *(Large surpluses at tiny primes) The surpluses  $v_2(N! / \prod \mathcal{B}^{(5)})$ ,  $v_3(N! / \prod \mathcal{B}^{(5)})$  lie in the sector  $\Gamma_{t,L} \subset \mathbb{R}^2$ , defined to be the set of pairs  $(n, m)$  with  $n, m > 0$  and*

$$\frac{\log(3L) + \kappa_L}{\log t - \log(3L)} \leq \frac{n \log 2}{m \log 3} \leq \frac{\log t - \log(2L)}{\log(2L) + \kappa_L}.$$

Then  $t(N) \geq t$ .

*Proof.* Write  $n := v_2(N! / \prod B^{(5)})$  and  $m := v_3(N! / \prod B^{(5)})$ . From (2.1) we have  $n \leq N$  and  $m \leq N/2$ , hence

$$n \log 2 + m \log 3 \leq N \log \sqrt{12}.$$

Applying Lemma 4.2, we can find a subfactorization  $B'$  of  $2^n 3^m$  with an excess of at most  $(\kappa_L \log \sqrt{12}) \frac{N}{\log t}$ , and with

$$\left| v_2 \left( \frac{2^n 3^m}{\prod B'} \right) \right|_{\log 2, \infty} + \left| v_3 \left( \frac{2^n 3^m}{\prod B'} \right) \right|_{\log 3, \infty} \leq 2(\log t + \kappa_L) \leq 2 \log N$$

where we have used (4.4) and the fact that  $\log t \leq \log N - 1$ . Then  $B^{(6)} = B^{(5)} \cup B'$  is another  $t$ -admissible multiset, and from (4.14), we obtain the previous sufficient condition (4.2).  $\square$

#### 4.5. Analysis of Step 5.

**Proposition 4.4.** *Let  $1 \leq K \leq t \leq N$ ,  $A \geq 1$ , and  $L \geq 1$  be parameters such that  $9L < t = N/e^{1+\delta}$  for some  $\delta > 0$ . Suppose that the algorithm in Section 4.2 with the indicated parameters reaches the end of Step 4 to produce a multiset  $B^{(4)}$  obeying the following hypotheses.*

(i) *(Small excess and surplus at small/medium primes)*

$$\begin{aligned} \mathcal{E}_t(B^{(4)}) + \sum_{3 < p \leq t/K} \left| v_p \left( \frac{N!}{\prod B^{(4)}} \right) \right|_{\kappa_K \min(\frac{\log p}{\log \sqrt{t/K}}, 1), \infty} \\ \leq \delta N - \frac{3}{2} \log N - \kappa_L (\log \sqrt{12}) \frac{N}{\log t}. \end{aligned} \quad (4.15)$$

(ii) *(Large surpluses at tiny primes) Whenever  $n_{**}, m_{**}$  are natural numbers obeying the bounds*

$$n_{**} \log 2 + m_{**} \log 3 \leq \sum_{3 < p \leq t/K} \left| v_p \left( \frac{N!}{\prod B^{(4)}} \right) \right|_{\frac{\log \sqrt{tK} + \kappa_K}{\log \sqrt{t/K}} \log p, \infty} + \log t + \kappa_L,$$

*then one has*

$$\left( v_2 \left( \frac{N!}{\prod B^{(4)}} \right) - n_{**}, v_3 \left( \frac{N!}{\prod B^{(4)}} \right) - m_{**} \right) \in \Gamma_{t,L}.$$

*Then  $t(N) \geq t$ .*

*Proof.* By (4.15),  $B^{(4)}$  is a subfactorization of  $N!$ , and by construction it is in balance at all primes  $p > t/K$ . Consider all the  $p$ -surplus primes in the small, borderline small, and medium range  $3 < p \leq t/K$ , thus each such prime is considered with multiplicity  $v_p(N! / \prod B^{(4)})$ . Using the greedy algorithm, one can factor the product of all these primes into  $M$  factors  $c_1, \dots, c_M$  in the interval  $(\sqrt{t/K}, t/K]$ , times at most one exceptional factor  $c_*$  in  $(1, \sqrt{t/K}]$ ,

for some  $M$ . If we let  $M'$  denote the number of factors in  $c_1, \dots, c_M$  that are not divisible by a prime larger than  $\sqrt{t/K}$ , we have the bound

$$\left(\sqrt{t/K}\right)^{M'} \leq \prod_{3 < p \leq \sqrt{t/K}} v_p \left( \frac{N!}{\prod B^{(4)}} \right)$$

and hence on taking logarithms

$$M' \leq \sum_{3 < p \leq \sqrt{t/K}} \left| v_p \left( \frac{N!}{\prod B^{(4)}} \right) \right|_{\frac{\log p}{\log \sqrt{t/K}}, \infty}.$$

Restoring the factors divisible by primes  $p > \sqrt{t/K}$ , we conclude that

$$M \leq \sum_{3 < p \leq t/K} \left| v_p \left( \frac{N!}{\prod B^{(4)}} \right) \right|_{\min(\frac{\log p}{\log \sqrt{t/K}}, 1), \infty}. \quad (4.16)$$

For each of the  $M$  factors  $c_i$ , we introduce the 3-smooth number  $[t/c_i]^{(2,3)} = 2^{n_i} 3^{m_i}$ , which by (4.3) lies in the interval  $[t/c_i, e^{\kappa_K} t/c_i]$ ; similarly, for the exceptional factor  $c_*$  we introduce a 3-smooth number  $[t/c_*]^{(2,3)} = 2^{n_*} 3^{m_*}$  in the interval  $[t/c_*, e^{\kappa_K} t/c_*]$ . If we then adjoin the 3-smooth numbers  $[t/c_i]^{(2,3)} c_i = 2^{n_i} 3^{m_i} c_i$  for  $i = 1, \dots, M$  as well as  $[t/c_*]^{(2,3)} c_* = 2^{n_*} 3^{m_*} c_*$  to the  $t$ -admissible multiset  $B^{(4)}$  to create a new  $t$ -admissible multiset  $B^{(5)}$ . The quantity  $\log [t/c_*]^{(2,3)} = n_* \log 2 + m_* \log 3$  is bounded by  $\log \sqrt{tK} + \kappa_K$ , and the quantity  $\log [t/c_i]^{(2,3)} = n_i \log 2 + m_i \log 3$  is similarly bounded by  $\log t + \kappa_K$ , hence if we denote  $n_{**} := n_1 + \dots + n_M + n_*$  and  $m_{**} := m_1 + \dots + m_M + m_*$ , we have

$$n_{**} \log 2 + m_{**} \log 3 \leq (\log \sqrt{tK} + \kappa_K) \sum_{3 < p \leq t/K} \left| v_p \left( \frac{N!}{\prod B^{(4)}} \right) \right|_{\min(\frac{\log p}{\log \sqrt{t/K}}, 1), \infty} + \log t + \kappa_K.$$

Each of the new factors in  $B^{(5)}$  contributes an excess of at most  $\kappa_K$ , so the total excess of  $B^{(5)}$  is at most

$$\mathcal{E}_t(B^{(4)}) + \kappa_K M + \kappa_K$$

which by (4.16) is bounded by

$$\mathcal{E}_t(B^{(4)}) + \sum_{3 < p \leq t/K} \left| v_p \left( \frac{N!}{\prod B^{(4)}} \right) \right|_{\kappa_K \min(\frac{\log p}{\log \sqrt{t/K}}, 1), \infty} + \kappa_K.$$

We conclude that  $B^{(5)}$  obeys the hypotheses of Proposition 4.3 (using (4.4) to bound  $\kappa_K$  by  $\log \sqrt{2\pi}$ ), and the claim follows.  $\square$

#### 4.6. Analysis of Step 4.

**Proposition 4.5.** *Let  $L \geq 1$ . Let  $9L < t = N/e^{1+\delta}$  for some  $\delta > 0$ , and suppose that the algorithm reaches the end of Step 3 to produce a multiset  $B^{(3)}$  obeying the following hypotheses:*



(i) (Small excess and surplus at small/medium primes) One has

$$\begin{aligned} \mathcal{E}_t(\mathcal{B}^{(3)}) + \sum_{3 < p \leq t/K} \left| v_p \left( \frac{N!}{\prod \mathcal{B}^{(3)}} \right) \right|_{\kappa_K \min(\frac{\log p}{\log \sqrt{t/K}}, 1), \kappa_p} \\ \leq \delta N - \frac{3}{2} \log N - \kappa_L (\log \sqrt{12}) \frac{N}{\log t}. \end{aligned} \quad (4.17)$$

(ii) (Large surpluses at tiny primes) Whenever  $n_{**}, m_{**}$  are natural numbers obeying the bounds

$$\begin{aligned} n_{**} \log 2 + m_{**} \log 3 \leq \sum_{3 < p \leq t/K} \left| v_p \left( \frac{N!}{\prod \mathcal{B}^{(3)}} \right) \right|_{(\log \sqrt{tK} + \kappa_K) \min(\frac{\log p}{\log \sqrt{t/K}}, 1), \log p + \kappa_p} \\ + \log t + \kappa_K, \end{aligned} \quad (4.18)$$

then one has

$$\left( v_2 \left( \frac{N!}{\prod \mathcal{B}^{(3)}} \right) - n_{**}, v_3 \left( \frac{N!}{\prod \mathcal{B}^{(3)}} \right) - m_{**} \right) \in \Gamma_{t,L}. \quad (4.19)$$

Then  $t(N) \geq t$ .

*Proof.* Suppose there is a non-tiny prime  $p > 3$  with a positive  $p$ -deficit  $|v_p(N! / \prod \mathcal{B}^{(3)})|_{0,1} > 0$ . Since  $\mathcal{B}^{(3)}$  is in balance at all large primes, we have  $3 < p \leq t/K$ . We locate an element of  $\mathcal{B}^{(3)}$  that contains  $p$  as a factor, and replaces it with  $[p]^{(2,3)} = 2^{n_p} 3^{m_p}$ , which increases that factor by at most  $\exp(\kappa_p)$  thanks to (4.3). This procedure reduces the  $p$ -deficit by one, adds at most  $\kappa_p$  to the  $t$ -excess, and decrements  $v_2(N! / \prod \mathcal{B}^{(3)})$ ,  $v_3(N! / \prod \mathcal{B}^{(3)})$  by  $n_p, m_p$  respectively. Since  $n_p \log 2 + m_p \log 3 \leq \log p + \kappa_p$ , if we apply this procedure to clear all deficits at non-tiny primes, the resulting multiset  $\mathcal{B}^{(4)}$  has a  $t$ -excess of

$$\mathcal{E}_t(\mathcal{B}^{(4)}) \leq \mathcal{E}_t(\mathcal{B}^{(3)}) + \sum_{p > 3} |v_p(N! / \prod \mathcal{B})|_{0, \kappa_p}$$

and we have

$$v_2(N! / \prod \mathcal{B}^{(4)}) = v_2(N! / \prod \mathcal{B}^{(3)}) - n', \quad v_3(N! / \prod \mathcal{B}^{(4)}) = v_3(N! / \prod \mathcal{B}^{(3)}) - m'$$

with

$$n' \log 2 + m' \log 3 \leq \sum_{p > 3} \left| v_p \left( \frac{N!}{\prod \mathcal{B}^{(3)}} \right) \right|_{0, \log p + \kappa_p}.$$

The hypotheses of Proposition 4.4 are now satisfied, and we are done.  $\square$

To simplify the criteria here, we introduce the quantities

$$X_1 := \sum_{3 < p \leq K} \left| v_p \left( \frac{N!}{\prod \mathcal{B}^{(3)}} \right) \right|_{\frac{\log p}{\log \sqrt{t/K}}, 0} + \sum_{K < p \leq t/K} \left| v_p \left( \frac{N!}{\prod \mathcal{B}^{(3)}} \right) \right| + \quad (4.20)$$

$$X_2 := \sum_{3 < p \leq K} \left| v_p \left( \frac{N!}{\prod \mathcal{B}^{(3)}} \right) \right|_{0,1}. \quad (4.21)$$

Since  $\kappa_K \min(\frac{\log p}{\log \sqrt{t/K}}, 1)$ ,  $\kappa_p$  are both bounded by  $\kappa_K$  for  $p \geq K$ , and bounded by  $\kappa_K \frac{\log p}{\log \sqrt{t/K}}$ ,  $\kappa_5$  respectively for  $3 < p \leq K$ , we can replace (4.17) with

$$\mathcal{E}_t(\mathcal{B}^{(3)}) + \kappa_K X_1 + \kappa_5 X_2 \leq \delta N - \frac{3}{2} \log N - \kappa_L (\log \sqrt{12}) \frac{N}{\log t}. \quad (4.22)$$

Similarly, since  $\log p + \kappa_p$  is bounded by  $\log \sqrt{tK} + \kappa_K$  for  $K < p \leq t/K$ , and  $\log p + \kappa_p$  is bounded by  $\log K + \kappa_5$  for  $3 < p \leq K$ , we can replace (4.18) with

$$n_{**} \log 2 + m_{**} \log 3 \leq (\log \sqrt{tK} + \kappa_K)(X_1 + 2) + (\log K + \kappa_5)X_2. \quad (4.23)$$

**4.7. Analysis of Steps 1,2,3.** To apply Proposition 4.5, we now compute the various statistics of  $\mathcal{B}^{(3)}$  produced by Steps 1-3.

We begin with the analysis of  $\mathcal{B}^{(1)}$ , constructed in Step 1 of the algorithm. To count elements coprime to 6, we use the following lemma:

**Lemma 4.6.** *For any interval  $(a, b]$  with  $0 \leq a \leq b$ , the number of natural numbers in the interval that are coprime to 6 is  $\frac{b-a}{3} + O_{\leq}(4/3)$ .*

*Proof.* By the triangle inequality, it suffices to show that the number of natural numbers coprime to 6 in  $[0, a]$ , minus  $a/3$ , is  $O_{\leq}(2/3)$ . The claim is easily verified for  $0 \leq a \leq 6$ , and the quantity in question is 6-periodic in  $a$ , giving the claim.  $\square$

The excess of  $\mathcal{B}^{(1)}$  can be computed as

$$\mathcal{E}_t(\mathcal{B}^{(1)}) = A \sum_{n \in I} \log \frac{n}{t}.$$

By the fundamental theorem of calculus, this is

$$A \int_0^{3t/A} |I \cap (t, t+h]| \frac{dh}{t+h}.$$

Bounding  $\frac{1}{t+h}$  by  $\frac{1}{t}$  and applying Lemma 4.6, we conclude that

$$\mathcal{E}_t(\mathcal{B}^{(1)}) \leq A \int_0^{3N/A} \left( \frac{h}{3} + \frac{4}{3} \right) \frac{dh}{t} = \frac{3N^2}{2tA} + 4. \quad (4.24)$$

Next, we compute  $p$ -valuations  $v_p(\mathcal{B}^{(1)})$ . By construction, this quantity vanishes at tiny primes  $p = 2, 3$ . For  $p > 3$ , we can use Lemma 4.6 again to conclude

$$\begin{aligned} v_p(\mathcal{B}^{(1)}) &= A \sum_{1 \leq j \leq \frac{\log N}{\log p}} |I \cap p^j \mathbb{Z}| \\ &= A \sum_{1 \leq j \leq \frac{\log N}{\log p}} \left( \frac{N}{p^j A} + O_{\leq}(4/3) \right) \\ &= \frac{N}{p-1} + O_{\leq} \left( \frac{1}{p-1} \right) + O_{\leq} \left( \frac{4A}{3} \left\lceil \frac{\log N}{\log p} \right\rceil \right) \\ &= \frac{N}{p-1} + O_{\leq} \left( \frac{4A + 0.75}{3} \left\lceil \frac{\log N}{\log p} \right\rceil \right) \end{aligned}$$

since  $\frac{1}{p-1} \leq \frac{0.75}{3}$ . Meanwhile, from (2.1) one has

$$v_p(N!) = \frac{N}{p-1} + O_{\leq} \left( \left\lceil \frac{\log N}{\log p} \right\rceil \right)$$

and thus

$$v_p(N!/\mathcal{B}^{(1)}) = O_{\leq} \left( \frac{4A + 3.75}{3} \left\lceil \frac{\log N}{\log p} \right\rceil \right). \quad (4.25)$$

Now we pass to  $\mathcal{B}^{(2)}$  by performing Step 2 of the algorithm. Removing elements from a  $t$ -admissible multiset cannot increase the  $t$ -excess, so from (4.24) we have

$$\mathcal{E}_t(\mathcal{B}^{(2)}) \leq \frac{3N^2}{2tA} + 4. \quad (4.26)$$

The elements removed are of the form  $pm$  with  $m \leq K(1+\nu)$  coprime to 6, and  $p$  in the interval  $(\frac{t}{\min(m, K)}, \frac{t(1+\sigma)}{m}]$ . We conclude that

$$v_p(N!/\mathcal{B}^{(2)}) = v_p(N!/\mathcal{B}^{(1)}) \quad (4.27)$$

for medium primes  $K(1+\sigma) < p \leq t/K$ . For small and borderline small primes  $3 < p \leq K(1+\sigma)$  one has

$$v_p(N!/\mathcal{B}^{(2)}) = v_p(N!/\mathcal{B}^{(1)}) + A \sum_{\substack{m \leq K(1+\sigma) \\ (m, 6)=1}} v_p(m) \left( \pi \left( \frac{t(1+\sigma)}{m} \right) - \pi \left( \frac{t}{\min(m, K)} \right) \right). \quad (4.28)$$

Finally, for tiny primes  $p = 2, 3$  we have the maximal surplus:

$$v_p(N!/\mathcal{B}^{(2)}) = v_p(N!).$$

We now pass to  $\mathcal{B}^{(3)}$  by performing Step 3 of the algorithm. In other words, we add  $v_p(N!)$  copies of  $p \lceil t/p \rceil$  for each largeprime  $p > t/K$ . The  $t$ -excess is now given by

$$\mathcal{E}_t(\mathcal{B}^{(3)}) = \mathcal{E}_t(\mathcal{B}^{(2)}) + \sum_{p > t/K} v_p(N!) \log \frac{\lceil t/p \rceil}{t/p}$$

and hence by (4.26)

$$\mathcal{E}_t(\mathcal{B}^{(3)}) = \frac{3N^2}{2tA} + 4 + \sum_{p > t/K} v_p(N!) \log \frac{\lceil t/p \rceil}{t/p}. \quad (4.29)$$

By construction one has balance (4.30) at large primes  $p > t/K$ ,

$$v_p(N!/B^{(3)}) = 0 \quad (4.30)$$

and no modification at borderline small or medium primes  $K < p \leq t/K$ ,

$$v_p(N!/B^{(3)}) = v_p(N!/B^{(2)}) \quad (4.31)$$

but now the  $p$ -surplus or  $p$ -deficit at small primes  $3 < p \leq K$  is modified:

$$v_p(N!/B^{(3)}) = v_p(N!/B^{(2)}) - \sum_{p' > t/K} v_{p'}(N!) v_p(\lceil t/p' \rceil). \quad (4.32)$$

Similarly, at tiny primes  $p = 2, 3$  we have

$$v_p(N!/B^{(3)}) = v_p(N!) - \sum_{p' > t/K} v_{p'}(N!) v_p(\lceil t/p' \rceil). \quad (4.33)$$

At medium primes,  $K(1 + \sigma) < p \leq t/K$ , we see from (4.31), (4.25) that

$$\left| v_p \left( \frac{N!}{\prod B^{(3)}} \right) \right| \leq \frac{4A + 3.75}{3} \left\lceil \frac{\log N}{\log p} \right\rceil.$$

For borderline primes  $K \leq p < K(1 + \sigma)$ , we have from (4.31), (4.28) that

$$\left| v_p \left( \frac{N!}{\prod B^{(3)}} \right) \right| \leq \frac{4A + 3.75}{3} \left\lceil \frac{\log N}{\log p} \right\rceil + A \sum_{\substack{m \leq K(1+\sigma) \\ (m,6)=1}} v_p(m) \left( \pi \left( \frac{t(1+\sigma)}{m} \right) - \pi \left( \frac{t}{\min(m, K)} \right) \right).$$

The only  $m$  which contributes here is  $m = p$ , thus we may simplify to

$$\begin{aligned} \left| v_p \left( \frac{N!}{\prod B^{(3)}} \right) \right| &\leq \frac{4A + 3.75}{3} \left\lceil \frac{\log N}{\log p} \right\rceil + A \left( \pi \left( \frac{t}{p} + \frac{\sigma t}{p} \right) - \pi \left( \frac{t}{K} \right) \right) \\ &\leq \frac{4A + 3.75}{3} \left\lceil \frac{\log N}{\log p} \right\rceil + A \left( \pi \left( \frac{t(1+\sigma)}{K} \right) - \pi \left( \frac{t}{K} \right) \right). \end{aligned}$$

For the small primes  $3 < p \leq K$ , we see from (4.25), (4.28), (4.32) that we have the upper bound

$$v_p \left( \frac{N!}{\prod B^{(3)}} \right) \leq \frac{4A + 3.75}{3} \left\lceil \frac{\log N}{\log p} \right\rceil + Y_p + Z_p$$

and the lower bound

$$v_p \left( \frac{N!}{\prod B^{(3)}} \right) \leq -\frac{4A + 3.75}{3} \left\lceil \frac{\log N}{\log p} \right\rceil + Y_p$$

where  $Y_p$  is the quantity

$$\begin{aligned} Y_p &:= A \sum_{\substack{m \leq K \\ (m,6)=1}} v_p(m) \left( \pi \left( \frac{t(1+\sigma)}{m} \right) - \pi \left( \frac{t}{\min(m, K)} \right) \right) \\ &\quad - \sum_{p' > t/K} \left\lceil \frac{N}{p'} \right\rceil v_p(\lceil t/p' \rceil) \end{aligned}$$

and  $Z_p$  is the (non-negative) error term

$$Z_p := A \sum_{\substack{K < m \leq K(1+\sigma) \\ (m,6)=1}} \nu_p(m) \left( \pi \left( \frac{t(1+\sigma)}{m} \right) - \pi \left( \frac{t}{\min(m, K)} \right) \right).$$

Here we have used (2.1) to write  $\nu_{p'}(N!)$  as  $\lceil N/p' \rceil$ . An important phenomenon for us will be that  $Y_p$  is usually positive (so that  $\mathcal{B}^{(3)}$  typically enjoys a (modest) surplus at small primes rather than a deficit). From the triangle inequality we now have

$$\begin{aligned} X_1 &\leq \frac{4A + 3.75}{3} \sum_{3 < p \leq t/K} \left\lceil \frac{\log N}{\log p} \right\rceil \\ &\quad + A(\pi(K(1+\sigma)) - \pi(K)) \left( \pi \left( \frac{t(1+\sigma)}{K} \right) - \pi \left( \frac{t}{K} \right) \right) \\ &\quad + \sum_{3 < p \leq K} \frac{\log p}{\log \sqrt{t/K}} (|Y_p|_{1,0} + Z_p) \end{aligned} \quad (4.34)$$

and

$$X_2 \leq \frac{4A + 3.75}{3} \sum_{3 < p \leq K} \left\lceil \frac{\log N}{\log p} \right\rceil + \sum_{3 < p \leq K} |Y_p|_{0,1}. \quad (4.35)$$

## 5. ASYMPTOTIC EVALUATION OF $t(N)$

In this section we establish the lower bound

$$\frac{t}{N} \geq \frac{1}{e} - \frac{c_0}{\log N} + \frac{1}{\log^{1+c_1} N} \quad (5.1)$$

for some small absolute constant  $0 < c_1 < 1$ , if  $N$  is sufficiently large. With this choice of parameters, one has

$$\delta = \frac{ec_0}{\log N} + \frac{1}{\log^{1+c_1} N} + O\left(\frac{1}{\log^2 N}\right).$$

Let  $N$  be sufficiently large. We introduce parameters

$$A := \lfloor \log^2 N \rfloor$$

and

$$K := \lfloor \log^3 N \rfloor$$

and

$$L := N^{0.1}.$$

We apply the algorithm from Section 4.2, using the first option for Step 3, and invoke Proposition 4.5. With these parameters, we see from Lemma A.3 that the right-hand side of (4.17) is at least

$$ec_0 \frac{N}{\log N} + \frac{N}{2 \log^{1+c_1} N}$$

if  $c_1$  is small enough and  $N$  is large enough. Thus, in view of (4.22), (4.23), it suffices to establish the bound

$$\mathcal{E}_t(\mathcal{B}^{(3)}) + \kappa_K X_1 + \kappa_5 X_2 \leq ec_0 \frac{N}{\log N} + O\left(\frac{N(\log \log N)^3}{\log^2 N}\right) \quad (5.2)$$

as well as the condition (4.19) whenever (4.23) holds.

By repeating the proof of Proposition 3.4, we see that

$$\sum_{p > t/K} v_p(N!) \log \frac{[t/p]}{t/p} = ec_0 \frac{N}{\log N} + O\left(\frac{N}{\log^2 N}\right)$$

and hence by (4.29)

$$\mathcal{E}_t(\mathcal{B}^{(2)}) = ec_0 \frac{N}{\log N} + O\left(\frac{N}{\log^2 N}\right).$$

Next, from (4.34), (4.35) and the prime number theorem, we can calculate that

$$X_1 \ll \sum_{3 < p \leq K} \frac{\log p}{\log N} (|Y_p|_{1,0} + Z_p) + \frac{N}{\log^2 N}$$

and

$$X_2 \ll \sum_{3 < p \leq K} |Y_p|_{0,1} + \log^6 N$$

so to verify (5.2) will suffice by Mertens' theorem to show that

$$|Y_p|_{1,0}, Z_p \ll \frac{N(\log \log N)^2}{p \log N} \quad (5.3)$$

and

$$|Y_p|_{0,1} \ll \frac{N(\log \log N)^2}{p \log^2 N}, \quad (5.4)$$

as this will imply that

$$X_1, X_2 \ll \frac{N(\log \log N)^3}{\log^2 N}. \quad (5.5)$$

Note the need to obtain stronger control on  $|Y_p|_{0,1}$  than on  $|Y_p|_{1,0}$ .

The error term  $Z_p$  is easily disposed of:

$$\begin{aligned} Z_p &\ll A \sum_{K < m \leq K(1+\sigma)} v_p(m) \left( \pi\left(\frac{t(1+\sigma)}{K}\right) - \pi\left(\frac{t}{K}\right) \right) \\ &\ll \frac{A\sigma t}{K \log N} \sum_{K < m \leq K(1+\sigma)} \sum_{j \ll \log K} 1_{p^j | m} \\ &\ll \frac{N}{K \log N} \sum_{j \ll \log K} \left( \frac{K\sigma}{p^j} + 1 \right) \\ &\ll \frac{N}{p \log^2 N}. \end{aligned}$$

As for  $Y_1$ , we see from Corollary C.2 that

$$\begin{aligned}
& A \sum_{m \leq K(1+\sigma); (m,6)=1} v_p(m) \left( \pi \left( \frac{t(1+\sigma)}{m} \right) - \pi \left( \frac{t}{m} \right) \right) \\
&= (1 + O(\sigma)) \sum_{m \leq K(1+\sigma); (m,6)=1} v_p(m) \frac{t\sigma}{\log(t/m)} \\
&= \left( 1 + O \left( \frac{\log \log N}{\log N} \right) \right) \frac{N}{\log N} \sum_{m \leq K(1+\sigma); (m,6)=1} v_p(m) \frac{3}{m};
\end{aligned}$$

since

$$\begin{aligned}
\sum_{m \leq K(1+\sigma)} \frac{v_p(m)}{m} &= \sum_{j=1}^{\infty} \sum_{m \leq K(1+\sigma); p^j | m} \frac{1}{m} \\
&\ll \sum_{j=1}^{\infty} \frac{\log K}{p^j} \\
&\ll \frac{\log \log N}{p}
\end{aligned} \tag{5.6}$$

we conclude the desired upper bound

$$|Y_1|_{1,0} \ll \frac{N \log \log N}{p \log N}.$$

To control the negative part, we apply Corollary C.2 again to calculate

$$\begin{aligned}
\sum_{p' > t/K} v_{p'}(N!) v_p(\lceil t/p' \rceil) &\leq \sum_{p' > t/K} \frac{N}{p'} v_p(\lceil t/p' \rceil) \\
&= \sum_{m \leq K} v_p(m) \sum_{t/m \leq p' < t/(m-1)} \frac{N}{p'} \\
&= \left( 1 + O \left( \frac{\log \log N}{\log N} \right) \right) \sum_{m \leq K} v_p(m) \int_{t/m}^{t/(m-1)} \frac{N}{x} \frac{dx}{\log N} \\
&= \left( 1 + O \left( \frac{\log \log N}{\log N} \right) \right) \frac{N}{\log N} \sum_{m \leq K} v_p(m) \log \frac{m}{m-1} \\
&= \frac{N}{\log N} \sum_{m \leq K} v_p(m) \log \frac{m}{m-1} + O \left( \frac{N(\log \log N)^2}{p \log^2 N} \right)
\end{aligned}$$

where to justify the last line we used (5.6). Putting this together, we find that

$$|Y_1|_{0,1} \ll \frac{N}{\log N} \left| \sum_{m \leq K(1+\sigma); (m,6)=1} v_p(m) \frac{3}{m} - \sum_{m \leq K} v_p(m) \log \frac{m}{m-1} \right|_{0,1} + \frac{N(\log \log N)^2}{p \log^2 N}. \tag{5.7}$$

We now have a crucial inequality:

**Lemma 5.1** (Key inequality). *For  $p \geq 5$  and  $K > 0$ , we have*

$$0 \leq \sum_{m \leq K; (m,6)=1} v_p(m) \frac{3}{m} - \sum_{m \leq K} v_p(m) \log \frac{m}{m-1} \leq \frac{2}{p-1}.$$

We defer the proof of this lemma to Appendix B. Because of this lemma, the first term on the right-hand side of (5.7) vanishes, and we recover the desired bounds (5.3), (5.4), and hence (5.5), (5.2), (4.22), and (4.17).

Now we turn to the verification of (4.19) assuming (4.23). In view of (5.5), the latter condition implies that

$$n_{**}, m_{**} \ll \frac{N(\log \log N)^3}{\log N}.$$

Meanwhile, for a tiny prime  $p = 2, 3$ , we see from (4.33), (2.1), the prime number theorem, and (5.6) that

$$\begin{aligned} v_p(N!/B^{(3)}) &= \frac{N}{p-1} + O(\log N) - O\left(\sum_{p' > t/K} \frac{N}{p'} v_p(\lceil t/p' \rceil)\right) \\ &= \frac{N}{p-1} + O(\log N) - O\left(\sum_{m \leq K} v_p(m) \sum_{t/m \leq p' < t/(m-1)} \frac{N}{p'}\right) \\ &= \frac{N}{p-1} + O(\log N) - O\left(\frac{N}{\log N} \sum_{m \leq K} \frac{v_p(m)}{m}\right) \\ &= \frac{N}{p-1} + O\left(\frac{N \log \log N}{\log N}\right). \end{aligned}$$

We conclude that

$$\begin{aligned} v_2(N!/B^{(3)}) - n_{**} &= N + O\left(\frac{N(\log \log N)^2}{\log N}\right) \\ v_3(N!/B^{(3)}) - m_{**} &= \frac{N}{2} + O\left(\frac{N(\log \log N)^2}{\log N}\right). \end{aligned}$$

By choice of  $L$ , this implies (4.19) for  $N$  large enough. The proof of (5.1) is now complete.

## 6. GUY–SELFIDGE CONJECTURE

We now establish the Guy–Selfridge conjecture  $t(N) \geq \delta$  in the range

$$N \geq ???.$$

We will apply Proposition 4.5 with the choice of parameters

$$t := N/3$$

$$A := ???$$

$$K := 342$$

$$L := 342.$$

Clearly  $\delta = \log \frac{3}{e} = 0.09861 \dots$ , and

$$\sigma = \frac{9}{A}.$$



From Lemma A.1, we have

$$\kappa_K = \kappa_L \leq \log \frac{9}{8} = 0.11778 \dots \quad (6.1)$$

Thus the right-hand side of (4.22) is at least

$$N \log \frac{3}{e} - \frac{3}{2} \log N - (\log \frac{9}{8})(\log \sqrt{12}) \frac{N}{\log(N/3)}.$$

Now we compute the  $t$ -excess. From (4.29) and (2.1) (noting that  $N/3K \geq \sqrt{N}$ ) we have

$$\mathcal{E}_t(\mathcal{B}^{(3)}) \leq \frac{9N}{2A} + 4 + \sum_{p > N/3K} \left\lfloor \frac{N}{p} \right\rfloor \log \frac{[N/3p]}{N/3p}$$

and thus

$$\mathcal{E}_t(\mathcal{B}^{(3)}) \leq \frac{9N}{2A} + 4 + \frac{1}{\log(N/3K)} \sum_{N/3K < p \leq N} f(p/N) \log p$$

where  $f : (1/3K, 1] \rightarrow \mathbb{R}$  is the piecewise smooth function

$$f(x) := \left\lfloor \frac{1}{x} \right\rfloor \log \frac{[1/3x]}{1/3x}.$$

Applying Lemma C.3 and a change of variables, we thus have

$$\mathcal{E}_t(\mathcal{B}^{(3)}) \leq \frac{9N}{2A} + 4 + \frac{N}{\log(N/3K)} \int_{1/3K}^1 \left(1 - \frac{2}{\sqrt{Nx}}\right) f(x) dx + \frac{E(N)}{\log(N/3K)} (f(1/3K+) + f(1) + \|f\|_{\text{TV}}).$$

We discard the  $\frac{2}{\sqrt{Nx}}$  term as it gives a negative contribution. Direct numerical calculation (cf. Figure 4) reveals that

$$\int_{1/3K}^1 f(x) dx \leq 0.9201$$

$$f(1/3K+) + f(1) + \|f\|_{\text{TV}} \leq 2044$$

and thus

$$\mathcal{E}_t(\mathcal{B}^{(3)}) \leq \frac{N}{20} + 4 + \frac{N}{\log(N/3K)} (0.9201 + 2044 \frac{E(N)}{N}).$$

Now we consider the expression

$$\sum_{3 < p \leq t/K} \left\lfloor \frac{\log N}{\log p} \right\rfloor$$

appearing in (4.34). The quantity  $\left\lfloor \frac{\log N}{\log p} \right\rfloor$  equals 1 for  $p > \sqrt{t}$  and is at most  $\frac{\log N}{\log 5} + 1$  for  $3 < p \leq \sqrt{t}$ , so we have

$$\sum_{3 < p \leq t/K} \left\lfloor \frac{\log N}{\log p} \right\rfloor \leq \pi(t/K) + \frac{\log N}{\log 5} \pi(\sqrt{t}).$$

One can bound this using for instance the bound

$$\pi(x) \leq \frac{x}{\log x} + \frac{1.2762x}{\log^2 x}$$

for  $x > 1$  (see [6, §3]).

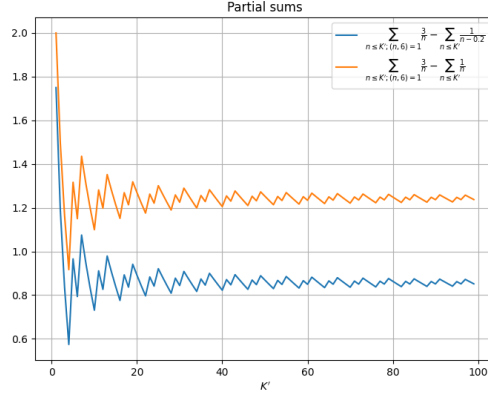


FIGURE 4. A plot of  $f(x)$ . The integral  $c_1 = \int_{1/K}^1 f(x) dx \approx 0.9200$  is slightly larger than  $ec_0 \approx 0.8244$ .

As for the quantity  $\sum_{3 < p \leq K} \left\lceil \frac{\log N}{\log p} \right\rceil$  appearing in (4.35), we crudely bound it by

$$\sum_{3 < p \leq K} \left\lceil \frac{\log N}{\log p} \right\rceil \leq \left( \frac{\log N}{\log 5} + 1 \right) (\pi(K) - \pi(3)) = 66 \left( \frac{\log N}{\log 5} + 1 \right).$$

#### APPENDIX A. POWERS OF 2 AND 3

We now obtain good bounds on the quantity  $\kappa_L$ . Clearly  $\kappa_L$  is a non-increasing function of  $L$  with  $\kappa_1 = \log 2$ . The following lemma gives improved control on  $\kappa_L$  for large  $L$ :

**Lemma A.1.** *If  $n_1, n_2, m_1, m_2$  are natural numbers such that  $n_1 + n_2, m_1 + m_2 \geq 1$  and*

$$1 \leq \frac{3^{m_1}}{2^{n_1}}, \frac{2^{n_2}}{3^{m_2}}$$

*then*

$$\kappa_{\min(2^{n_1+n_2}, 3^{m_1+m_2})/6} \leq \log \max \left( \frac{3^{m_1}}{2^{n_1}}, \frac{2^{n_2}}{3^{m_2}} \right).$$

*Proof.* If  $\min(2^{n_1+n_2}, 3^{m_1+m_2})/6 \leq t \leq 2^{n_2-1}3^{m_1-1}$ , then we have

$$t \leq 2^{n_2-1}3^{m_1-1} \leq \max \left( \frac{3^{m_1}}{2^{n_1}}, \frac{2^{n_2}}{3^{m_2}} \right) t, \quad (\text{A.1})$$

so we are done in this case. Now suppose that  $t > 2^{n_2-1}3^{m_1-1}$ . If we write  $\lceil t \rceil^{(2,3)} = 2^n 3^m$  be the smallest 3-smooth number that is at least  $t$ , then we must have  $n \geq n_2$  or  $m \geq m_1$  (or both). Thus at least one of  $\frac{2^{n_1}}{3^{m_1}} 2^n 3^m$  and  $\frac{3^{m_2}}{2^{n_2}} 2^n 3^m$  is an integer, and is thus at most  $t$  by construction. This gives (A.1), and the claim follows.  $\square$

Some efficient choices of parameters for this lemma are given in Table 2. For instance,  $\kappa_{4.5} \leq 0.28768 \dots$  and  $\kappa_{40.5} \leq 0.16989 \dots$ .

$n_1$	$m_1$	$n_2$	$m_2$	$\min(2^{n_1+n_2}, 3^{m_1+m_2})/6$	$\log \max(3^{m_1}/2^{n_1}, 2^{n_2}/3^{m_2})$
1	1	<b>1</b>	<b>0</b>	$1/2 = 0.5$	$\log 2 = 0.69314 \dots$
<b>1</b>	<b>1</b>	2	1	$2^2/3 = 1.33 \dots$	$\log(3/2) = 0.40546 \dots$
3	2	<b>2</b>	<b>1</b>	$3^2/2 = 4.5$	$\log(2^2/3) = 0.28768 \dots$
3	2	<b>5</b>	<b>3</b>	$3^4/2 = 40.5$	$\log(2^5/3^3) = 0.16989 \dots$
<b>3</b>	<b>2</b>	8	5	$2^{10}/3 = 341.33 \dots$	$\log(3^2/2^3) = 0.11778 \dots$
<b>11</b>	<b>7</b>	8	5	$2^{18}/3 = 87381.33 \dots$	$\log(3^7/2^{11}) = 0.06566 \dots$
19	12	<b>8</b>	<b>5</b>	$3^{17}/2 \approx 6.4 \times 10^7$	$\log(2^8/3^5) = 0.05211 \dots$
19	12	<b>27</b>	<b>17</b>	$3^{29}/2 \approx 3.4 \times 10^{13}$	$\log(2^{27}/3^{17}) = 0.03856 \dots$
19	12	<b>46</b>	<b>29</b>	$3^{41}/2 \approx 1.8 \times 10^{19}$	$\log(2^{46}/3^{29}) = 0.02501 \dots$

TABLE 2. Efficient parameter choices for Lemma A.1. The parameters used to attain the minimum or maximum are indicated in **boldface**. Note how the number of rows in each group matches the terms 1, 1, 2, 2, 3, ... in the continued fraction expansion.

**Remark A.2.** It should be unsurprising that the continued fraction convergents  $1/1, 2/1, 3/2, 8/5, 19/12, \dots$  to

$$\frac{\log 3}{\log 2} = 1.5849\dots = [1; 1, 1, 2, 2, 3, 1, \dots]$$

are often excellent choices for  $n_1/m_1$  or  $n_2/m_2$ , although other approximants such as  $5/3$  or  $11/7$  are also usable.

Asymptotically, we have logarithmic-type decay:

**Lemma A.3** (Baker bound). *We have*

$$\kappa_L \ll \log^{-c} L$$

for all  $L \geq 2$  and some absolute constant  $c > 0$ .

*Proof.* From the classical theory of continued fractions, we can find rational approximants

$$\frac{p_{2j}}{q_{2j}} \leq \frac{\log 3}{\log 2} \leq \frac{p_{2j+1}}{q_{2j+1}} \quad (\text{A.2})$$

to the irrational number  $\log 3 / \log 2$ , where the convergents  $p_j/q_j$  obey the recursions

$$p_j = b_j p_{j-1} + p_{j-2}, \quad q_j = b_j q_{j-1} + q_{j-2}$$

with  $p_{-1} = 1, q_{-1} = 0, p_0 = b_0, q_0 = 1$ , and

$$[b_0; b_1, b_2, \dots] = [1; 1, 1, 2, 2, 3, 1, \dots]$$

is the continued fraction expansion of  $\frac{\log 3}{\log 2}$ . Furthermore,  $p_{2j+1}q_{2j} - p_{2j}q_{2j+1} = 1$ , and hence

$$\frac{\log 3}{\log 2} - \frac{p_{2j}}{q_{2j}} = \frac{1}{q_{2j}q_{2j+1}}. \quad (\text{A.3})$$

By Baker's theorem,  $\frac{\log 3}{\log 2}$  is a Diophantine number, giving a bound of the form

$$q_{2j+1} \ll q_{2j}^{O(1)} \quad (\text{A.4})$$

and a similar argument (using  $p_{2j+2}q_{2j+1} - p_{2j+1}q_{2j+2} = -1$ ) gives

$$q_{2j+2} \ll q_{2j+1}^{O(1)}. \quad (\text{A.5})$$

We can rewrite (A.2) as

$$1 \leq \frac{3^{q_{2j}}}{2^{p_{2j}}}, \frac{2^{p_{2j+1}}}{3^{q_{2j+1}}}$$

and routine Taylor expansion using (A.3) gives the upper bounds

$$\frac{3^{q_{2j}}}{2^{p_{2j}}}, \frac{2^{p_{2j+1}}}{3^{q_{2j+1}}} \leq \exp\left(O\left(\frac{1}{q_{2j}}\right)\right).$$

From Lemma A.1 we obtain

$$\kappa_{\min(2^{p_{2j}+p_{2j+1}}, 3^{q_{2j}+q_{2j+1}})/6} \ll \frac{1}{q_{2j}}.$$

The claim then follows from (A.4), (A.5) after optimizing in  $j$ .

□

It seems reasonable to conjecture that  $c$  can be taken to be arbitrarily close to 1, but this is essentially equivalent to the open problem of determining that irrationality measure of  $\log 3 / \log 2$  is equal to 2.

## APPENDIX B. KEY INEQUALITY

We now prove Lemma 5.1. Writing  $\nu_p(m) = \sum_{j \geq 1} 1_{p^j | m}$ , it suffices to show that

$$0 \leq \sum_{m \leq K; (m,6)=1, p^j | m} \frac{3}{m} - \sum_{m \leq K, p^j | m} \log \frac{m}{m-1} \leq \frac{2}{p^j}$$

for all  $j$ . Making the change of variables  $m = p^j n$ , it suffices to show that

$$0 \leq \sum_{n \leq K'} \frac{3}{n} 1_{(n,6)=1} - p^j \log \frac{p^j n}{p^j n - 1} \leq 2$$

for any  $K' > 0$ . Using the bound

$$\log \frac{p^j n}{p^j n - 1} = \int_{p^j n - 1}^{p^j n} \frac{dx}{x} \in \left[\frac{1}{p^j n}, \frac{1}{p^j n - 1}\right]$$

and  $p^j \geq 5$ , we have

$$\frac{1}{n} \leq p^j \log \frac{p^j n}{p^j n - 1} \leq \frac{1}{n - 0.2}$$

and so it suffices to show that

$$0 \leq \sum_{n \leq K'} \frac{3}{n} 1_{(n,6)=1} - \frac{1}{n - 0.2} \leq \sum_{n \leq K'} \frac{3}{n} 1_{(n,6)=1} - \frac{1}{n} \geq 2. \quad (\text{B.1})$$

Since

$$\sum_{n=1}^{\infty} \frac{1}{n - 0.2} - \frac{1}{n} = \psi(0.8) - \psi(1) = 0.353473,$$

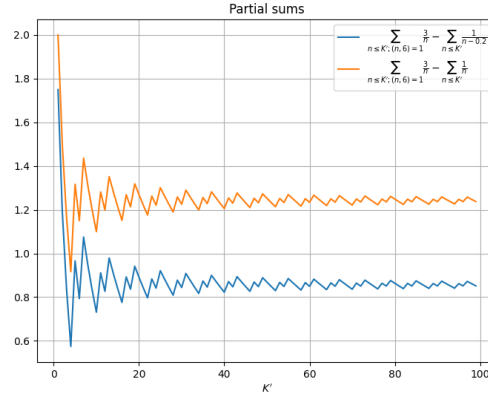


FIGURE 5. A plot of (B.1).

where  $\psi$  here denotes the digamma function rather than the von Mangoldt summatory function, it will suffice to show that

$$0.4 \leq \sum_{n \leq K'} \frac{3}{n} 1_{(n,6)=1} - \frac{1}{n} \geq 2. \quad (\text{B.2})$$

This can be numerically verified for  $K' \leq 100$ , with substantial room to spare for  $K'$  large; see Figure 5. On a block  $6a - 1 \leq n \leq 6a + 4$ , the sum is positive:

$$\begin{aligned} \sum_{6a-1 \leq n \leq 6a+4} \frac{3}{n} 1_{(n,6)=1} - \frac{1}{n} &= \left( \frac{1}{6a-1} - \frac{1}{6a} \right) + \left( \frac{1}{6a-1} - \frac{1}{6a+2} \right) \\ &\quad + \left( \frac{1}{6a+1} - \frac{1}{6a+3} \right) + \left( \frac{1}{6a+1} - \frac{1}{6a+4} \right) \\ &> 0. \end{aligned}$$

Similarly, on a block  $6a - 4 \leq n \leq 6a + 1$ , the sum is negative:

$$\begin{aligned} \sum_{6a-4 \leq n \leq 6a+1} \frac{3}{n} 1_{(n,6)=1} - \frac{1}{n} &= \left( \frac{1}{6a+1} - \frac{1}{6a} \right) + \left( \frac{1}{6a+1} - \frac{1}{6a-2} \right) \\ &\quad + \left( \frac{1}{6a-1} - \frac{1}{6a-3} \right) + \left( \frac{1}{6a-1} - \frac{1}{6a-4} \right) \\ &< 0. \end{aligned}$$

Thus the sum in (B.2) is increasing for  $K' = 4$  (6) and decreasing for  $K' = 1$  (6), and the inequality for  $K' > 100$  is then easily verified from the  $K' \leq 100$  data and the triangle inequality

From this and the triangle inequality one can easily establish (B.1) in the remaining ranges  $K' \geq 98$ .

## APPENDIX C. ESTIMATING SUMS OVER PRIMES

**Lemma C.1** (Integration by parts). *Let  $(y, x]$  be a half-open interval in  $(0, +\infty)$ . Suppose that one has a function  $a : \mathbb{N} \rightarrow \mathbb{R}$  and a continuous function  $f : (y, x] \rightarrow \mathbb{R}$  such that*

$$\sum_{y < n \leq z} a_n = \int_z^y f(t) dt + C + O_{\leq}(A)$$

*for all  $y \leq z \leq x$ , and some  $C \in \mathbb{R}$ ,  $A > 0$ . Then, for any function  $b : (y, x] \rightarrow \mathbb{R}$  of bounded total variation, one has*

$$\sum_{y < n \leq x} b(n)a_n = \int_x^y b(t)f(t) dt + O_{\leq}(A(|b(y^+)| + |b(x)| + \|b\|_{TV})), \quad (\text{C.1})$$

where  $b(y^+) := \lim_{t \rightarrow y^+} b(t)$  denotes the right limit of  $b$  at  $y$ .

*Proof.* If, for every natural number  $y < n \leq x$ , one modifies  $b$  to be equal to the constant  $b(n)$  in a small neighborhood of  $n$ , then one does not affect the left-hand side of (C.1) or increase the total variation of  $b$ , while only modifying the integral in (C.1) by an arbitrarily small amount. Hence, by the usual limiting argument, we may assume without loss of generality that  $b$  is locally constant at each such  $n$ . If we define the function  $g : (y, x] \rightarrow \mathbb{R}$  by

$$g(z) := \sum_{y < n \leq z} a_n - \int_z^y f(u) du - C$$

then  $g$  has jump discontinuities at the natural numbers, but is otherwise continuously differentiable, and is also bounded uniformly in magnitude by  $A$ . We can then compute the Riemann–Stieltjes integral

$$\int_{(y,x]} b dg = \sum_{y < n \leq x} b(n)a_n - \int_y^x f(t)b(t) dt.$$

Since the discontinuities of  $g$  and  $b$  do not coincide, we may integrate by parts to obtain

$$\int_{(y,x]} b dg = b(x)g(x) - b(y^+)g(y^+) - \int_{(y,x]} g db.$$

The left-hand side is  $O_{\leq}(A(|b(y^+)| + |b(x)| + \|b\|_{TV}))$ , and the claim follows.  $\square$

By combining the above lemma with the prime number theorem with classical error term, we obtain

**Corollary C.2.** *Under the above hypotheses with  $1 \leq y \leq x$ , we have*

$$\sum_{y < p \leq x} b(p) = \int_y^x b(t) \frac{dt}{\log t} + O\left((|b(y^+)| + |b(x)| + \|b\|_{TV})x \exp(-c\sqrt{\log x})\right)$$

for some absolute constant  $c > 0$ .

We have the following explicit version of the above estimate, where it is convenient to apply the Chebyshev weighting of assigning each prime  $p$  a weight of  $\log p$ .

**Lemma C.3** (Buthe effective prime number theorem). *Under the above hypotheses with  $1423 \leq y \leq x$ , one has*

$$\sum_{y < p \leq x} b(p) \log p = \int_y^x b(t) \left(1 - \frac{2}{\sqrt{t}}\right) dt + O\left((|b(y^+)| + |b(x)| + \|b\|_{TV})E(x)\right)$$

where

$$E(x) := 0.95\sqrt{x} + \frac{\sqrt{x}}{8\pi} \log x (\log x - 3) 1_{x \geq 10^{19}} + 1.11742 \times 10^{-8} x 1_{x \geq e^{45}}.$$

*Proof.* Observe that  $E$  is monotone non-decreasing. Thus by Lemma C.1, it will suffice to show that

$$\sum_{p \leq x} \log p = x - \sqrt{x} + O_{\leq}(E(x)) = \int_0^x \left(1 - \frac{2}{\sqrt{t}}\right) dt + O_{\leq}(E(x))$$

where

$$E(x) := 0.95\sqrt{x} + \frac{\sqrt{x}}{8\pi} \log x (\log x - 3) 1_{x \geq 10^{19}} + 1.11742 \times 10^{-8} x 1_{x \geq e^{45}}.$$

For  $1423 \leq x \leq 10^{19}$ , this follows from [4, Theorem 2]. In the range For  $10^{19} \leq x \leq 10^{21} \approx e^{48.35}$ , we use the bound

$$\psi(x) = x + O_{\leq}\left(\frac{\sqrt{x}}{8\pi} \log x (\log x - 3)\right)$$

which was established for  $5000 \leq x \leq 10^{21}$  in [3, (7.3)], where  $\psi(x) := \sum_{n \leq x} \Lambda(n)$  is the usual von Mangoldt summatory function. To use this, we apply [3, (6.10), (6.11)] to conclude that

$$\sum_{p \leq x} \log p = \psi(x) - \psi(\sqrt{x}) + O_{\leq}(1.03883(x^{1/3} + x^{1/5} + 2(\log x)x^{1/13})).$$

From [13, Theorems 10,12] we have

$$\psi(\sqrt{x}) = \sqrt{x} + O_{\leq}(0.18\sqrt{x}).$$

Since

$$0.18\sqrt{x} + 1.03883(x^{1/3} + x^{1/5} + 2(\log x)x^{1/13}) \leq 0.95\sqrt{x}$$

for  $x \geq 10^{19}$ , the claim follows.

Finally, in the range  $x \geq 10^{21}$ , we see from [3, Theorem 1, Table 1] that one has the bound

$$\psi(x) = x + O_{\leq}(1.11742 \times 10^{-8} x)$$

for  $x \geq e^{45} \approx 10^{19.54}$ , and the claim follows by repeating the previous arguments.  $\square$

We remark that assuming the Riemann hypothesis, the final term in the definition of  $E(x)$  may be deleted, since [3, (7.3)] then holds for all  $x \geq 5000$ .

APPENDIX D. COMPUTATION OF  $c_0$ 

In this appendix we give some details regarding the rigorous numerical estimation of the constant  $c_0$  defined in (??). As one might imagine from an inspection of Figure 3, direct application of numerical quadrature converges quite slowly due to the oscillatory singularity. To resolve the singularity, we can perform a change of variables  $x = 1/y$  to express  $c_0$  as an improper integral:

$$c_0 = \frac{1}{e} \int_1^\infty \lfloor y \rfloor \log \frac{\lceil y/e \rceil}{y/e} \frac{dy}{y^2}. \quad (\text{D.1})$$

The integrand is piecewise smooth and the integral can be computed explicitly on any interval  $[a, b]$  of the form

$$[a, b] \subset [n, n+1] \cap [(m-1)e, me]$$

for some non-negative integers  $n, m$  as

$$\int_a^b \lfloor y \rfloor \log \frac{\lceil y/e \rceil}{y/e} \frac{dy}{y^2} = n \left( \frac{\log(b/m)}{b} - \frac{\log(a/m)a}{b} \right).$$

This formula permits one to evaluate  $\int_1^b \lfloor y \rfloor \log \frac{\lceil y/e \rceil}{y/e} \frac{dy}{y^2}$  exactly for any finite  $b$ . To control the tail, we see from the crude bounds  $0 \leq \lfloor y \rfloor \leq y$  and

$$0 \leq \log \frac{\lceil y/e \rceil}{y/e} \leq \log \left( 1 + \frac{e}{y} \right) \leq \frac{e}{y}$$

that

$$0 \leq \int_b^\infty \lfloor y \rfloor \log \frac{\lceil y/e \rceil}{y/e} \frac{dy}{y^2} \leq \frac{e}{b} \quad (\text{D.2})$$

which allows for rigorous upper and lower bounds on the improper integral. For instance, this procedure gives

$$0.304419004 \leq c_0 \leq 0.304419017.$$

Heuristically, the tail integral (??) should be approximately  $e/2b$  due to the equidistribution properties of the fractional part of  $y/e$ . Using this heuristic approximation, one obtains the prediction

$$c_0 \approx 0.30441901087$$

## REFERENCES

- [1] K. Alladi, C. Grinstead, *On the decomposition of  $n!$  into prime powers*, J. Number Theory **9** (1977) 452–458.
- [2] S. F. Assmann, D. S. Johnson, D. J. Kleitman, J. Y.-T. Leung, *On a dual version of the one-dimensional bin packing problem*, J. Algorithms **5** (1984) 502–525.
- [3] J. Büthe, *Estimating  $\pi(x)$  and related functions under partial RH assumptions*, Math. Comp., **85**(301), 2483–2498, Jan. 2016.
- [4] J. Büthe, *An analytic method for bounding  $\psi(x)$* . Math. Comp., **87** (312), 1991–2009.
- [5] P. Dusart, *Inégalités explicites pour  $\psi(x)$ ,  $\theta(x)$ ,  $\pi(x)$  et les nombres premiers*, C. R. Math. Rep. Acad. Sci. Canada Vol. **21** (2) 1999, pp. 53–59.
- [6] P. Dusart, *Explicit estimates of some functions over primes*, Ramanujan J. **45** (2018) 227–251.
- [7] P. Erdős, *Some problems in number theory*, in Computers in Number Theory, Academic Press, London New York, 1971, pp. 405–414.



- [8] P. Erdős, *Some problems I presented or planned to present in my short talk*, Analytic number theory, Vol. 1 (Allerton Park, IL, 1995) (1996), 333–335.
- [9] P. Erdős, R. Graham, *Old and new problems and results in combinatorial number theory*, Monographies de L'Enseignement Mathématique 1980.
- [10] R. K. Guy, *Unsolved Problems in Number Theory*, 3rd Edition, Springer, 2004.
- [11] R. K. Guy, J. L. Selfridge, *Factoring factorial  $n$* , Amer. Math. Monthly **105** (1998) 766–767.
- [12] H. Robbins, *A Remark on Stirling's Formula*, Amer. Math. Monthly **62** (1955) 26–29.
- [13] J. Rosser, L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. 6 (1962), 64–94.
- [14] T. Tao, *Decomposing factorials into bounded factors*, preprint, 2025. <https://arxiv.org/abs/2503.20170v2>
- [15] T. Tao, *Verifying the Guy–Selfridge conjecture*, Github repository, 2025. <https://github.com/teorth/erdos-guy-selfridge>.