DECOMPOSING A FACTORIAL INTO LARGE FACTORS

... AND TERENCE TAO

ABSTRACT. Let $t(N)$ denote the largest number such that $N!$ can be expressed as the product of N numbers greater than or equal to $t(N)$. The bound $t(N)/N = 1/e - o(1)$ was apparently established in unpublished work of Erdős, Selfridge, and Straus; but the proof is lost. Here we obtain the more precise asymptotic

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

for an explicit constant $c_0 = 0.3044190\dots$ and some absolute constant $c > 0$, answering a question of Erdős and Graham. With numerical assistance, we also establish several conjectures of Guy and Selfridge concerning effective estimates of this quantity, for instance establishing $t(N) \geq N/3$ for $N \geq 43632$, with the threshold shown to be best possible.

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 *sub-factorization* 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

$$\prod\{3, 3, 3, 3, 4, 4, 5, 7, 8\} = 3^4 \times 4^2 \times 5 \times 7 \times 8 = 9!$$

of cardinality

$$|\{3, 3, 3, 3, 4, 4, 5, 7, 8\}| = 9,$$

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

2020 *Mathematics Subject Classification.* 11A51.

It is easy to see that $t(N)$ is non-decreasing in N, (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 first few elements of the sequence $t(N)$ are

$$1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, \ldots$$

(OEIS A034258). The values of $t(N)$ for $N \leq 79$ were computed in [9], and the values for $N \leq 200$ can be extracted from OEIS A034259, which describes the inverse sequence to t.

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 factors at the "tiny primes" 2, 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

$$\prod \mathcal{B} \geq t^{|\mathcal{B}|} \tag{1.1}$$

for any t-admissible multiset $\mathcal{B}$ with Stirling's formula (2.6), we obtain the trivial upper bound

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

for $N \geq 2$; see Figure 1. In [8, p.75] it was reported that an unpublished work of Erdős, Selfridge, and Straus established the asymptotic

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

(first conjectured in [6]) and asked if one could show the bound

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

for some constant $c > 0$ (problem #391 in `https://www.erdosproblems.com`; see also [9, 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 [7], 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 [9] 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 [9]. The following conjectures in [9] were also made:

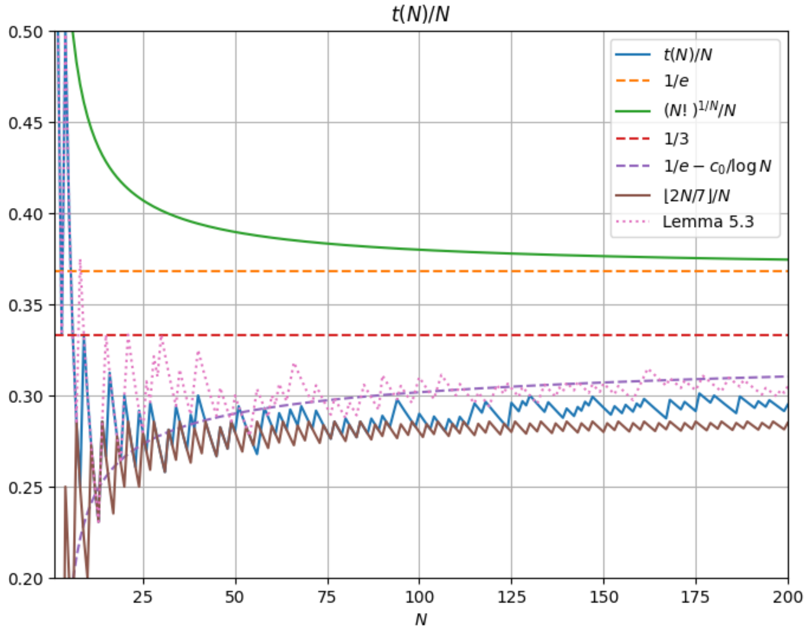

FIGURE 1. The function $t(N)/N$ (blue) for $N \leq 200$, using the data from OEIS A034258, as well as the trivial upper bound $(N!)^{1/N}/N$ (green), the improved upper bound from Lemma 5.3 (pink), which is asymptotic to (1.5) (purple), and the function $\lfloor 2N/7 \rfloor/N$ (brown), which is a lower bound for $N \neq 56$ [9]. 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 $N = 43632$. **TODO: relabel image to reflect new lemma numbering**

(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 $N \geq 3 \times 10^5$. (It was also asked if the threshold 3×10^5 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*

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

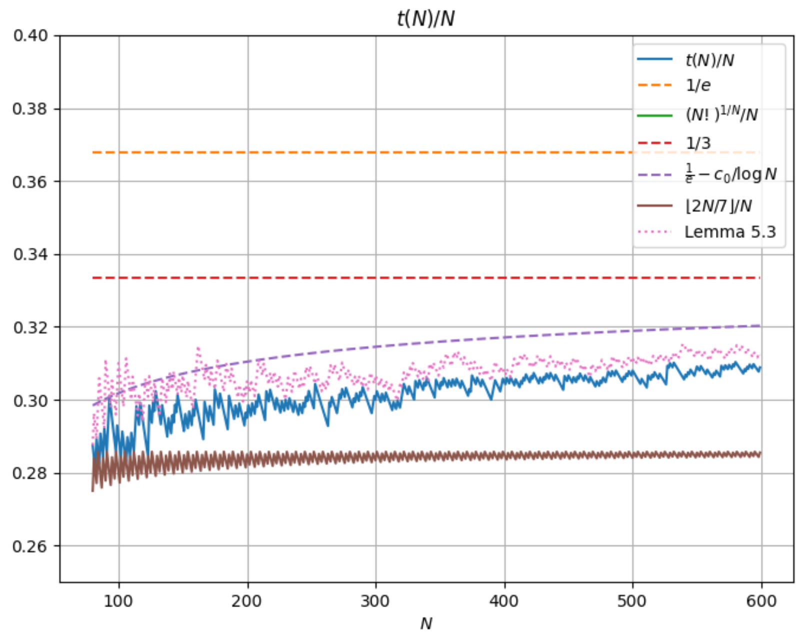

FIGURE 2. A continuation of Figure 1 to the region $80 \leq N \leq 599$.

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

$$\begin{aligned} c_0 &:= \frac{1}{e}\int_0^1 f_e(x)\,dx \\ &= 0.3044190\dots \end{aligned} \tag{1.6}$$

and for any $\alpha > 0$*,* $f_\alpha : (0,\infty) \to \mathbb{R}$ *denotes the piecewise smooth function*

$$f_\alpha(x) := \left\lfloor \frac{1}{x} \right\rfloor \log \frac{\lceil 1/\alpha x \rceil}{1/\alpha x}. \tag{1.7}$$

In particular, (1.3) *and* (1.4) *hold.*

For future reference, we observe the simple bounds

$$\begin{aligned} 0 \leq f_\alpha(x) &\leq \frac{1}{x} \log \frac{1/\alpha x + 1}{1/\alpha x} \\ &= \frac{1}{x}\log(1+\alpha x) \\ &\leq \alpha \end{aligned} \tag{1.8}$$

for all $x > 0$; in particular, f_α is a bounded function. It however has an oscillating singularity at $x = 0$; see **??**.

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

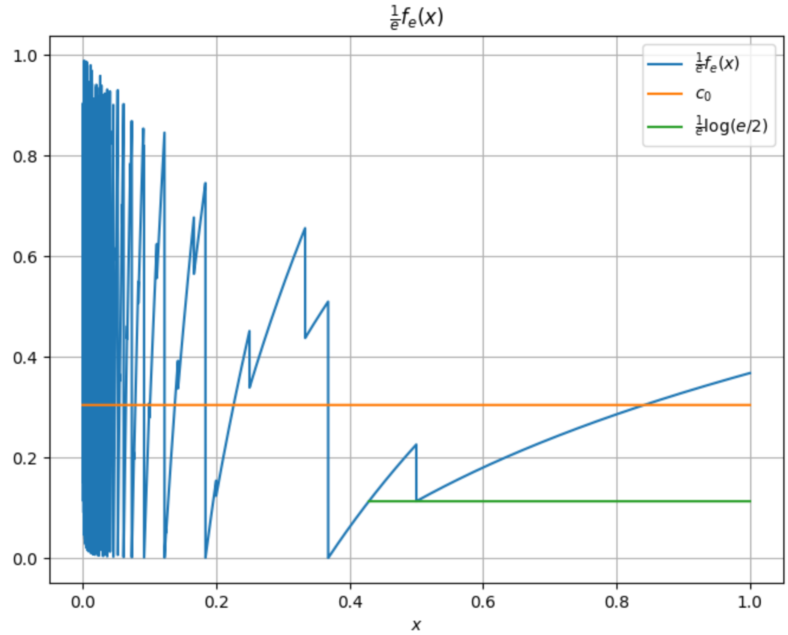

FIGURE 3. The piecewise continuous function $x \mapsto \frac{1}{e} f_e(x)$, together with its mean value $c_0 = 0.3044190\dots$ and the upper bound $\frac{\log(1+ex)}{ex}$. The function exhibits an oscillatory singularity at $x = 0$ similar to $\sin \frac{1}{x}$ (but it is always nonnegative and bounded). Informally, the function f_e 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.

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:

- In Section 4, we discuss *greedy algorithms* to construct subfactorizations, that provide quickly computable, though suboptimal, lower bounds on $t(N)$ for small and medium values;
- In Section 3, we present 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;
- In Section 5, we introduce 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, and is discussed in more detail in Section 5;

- In Section 5.1, we give *modified approximate factorization* strategy, which provides lower bounds on $t(N)$, that become asymptotically quite efficient.

The final approach is significantly more complicated than the other three, but is the only one which gives efficient lower bounds in the asymptotic limit $N \to \infty$. The key idea is to start with an approximate factorization

$$N! \approx \left(\prod_{j \in I} j\right)^A$$

for some small natural number A (e.g., $A = \lfloor \log^2 N \rfloor$) and a suitable set I of natural numbers greater than or equal to t; there is some freedom to select parameters here, and we will take I to be the natural numbers in $(t, t(1+\sigma)]$ that are coprime to 6, where t is the target lower bound for $t(N)$ we wish to establish, and $\sigma := \frac{3N}{tA}$. With a suitable choice of I, this product contains approximately the right number of copies of p for medium-sized primes p; but it has the "wrong" number of copies of large primes, and is also constructed to avoid the "tiny" primes $p = 2, 3$. One then performs a number of alterations to this approximate factorization to correct for the "surpluses" or "deficits" at various primes $p > 3$, using the supply of available tiny primes $p = 2, 3$ as a sort of "liquidity pool" to efficiently reallocate primes in the factorization. A key point will be that the incommensurability of $\log 2$ and $\log 3$ (i.e., the irrationality of $\log 3 / \log 2$) means that the 3-smooth numbers (numbers of the form $2^n 3^m$) are asymptotically dense (in logarithmic scale), allowing for other factors to be exchanged for 3-smooth factors with little loss[1].

1.1. **Author contributions and data.** This project was initially concieved 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.

1.2. **Acknowledgments.** TT is supported by NSF grant DMS-2347850. We thank Thomas Bloom for the web site `https://www.erdosproblems.com`, where the author learned of this problem, as well as Bryna Kra and Ivan Pan for corrections.

list here all contributors to the project who did not wish to be listed as co-authors.

[1]The weaker results alluded to in Remark 1.4 only used the prime 2 as a supply of "liquidity", and thus encountered inefficiencies due to the inability to "make change" when approximating another factor by a power of two.

2. Notation and basic estimates

If S is a statement, we use 1_S to denote its indicator, thus $1_S = 1$ when S is true and $1_S = 0$ when S is false. If x is a real number, we use $\lfloor x \rfloor$ to denote the greatest integer less than or equal to x, and $\lceil x \rceil$ to be the least integer greater than or equal to x.

Throughout this paper, the symbol p (or p', p_1, p_2, etc.) is always understood to be restricted to be prime. The primes $2, 3$ will play a special role in this paper and will be referred to as *tiny primes*. Call a natural number 3-*smooth* if it is the product of tiny primes, i.e., it is of the form $2^n 3^m$ for some natural numbers n, m. Given a positive real number x, we use $\lceil x \rceil^{\langle 2,3 \rangle}$ to denote the smallest 3-smooth number greater than or equal to x. For instance, $\lceil 5 \rceil^{\langle 2,3 \rangle} = 6$ and $\lceil 10 \rceil^{\langle 2,3 \rangle} = 12$. For any $L \geq 1$, let κ_L be the least quantity such that

$$x \leq \lceil x \rceil^{\langle 2,3 \rangle} \leq \exp(\kappa_L) x \tag{2.1}$$

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

$$\kappa_L \leq \log 2. \tag{2.2}$$

In fact κ_L decays to zero as L goes to infinity, due to the incommensurability of $\log 2$ and $\log 3$; we quantify this decay in Appendix A.

In practice, $\lceil x \rceil^{\langle 2,3 \rangle}$ will only be slightly larger than x; we quantify this in Appendix A.

We use (a, b) to denote the greatest common divisor of a and b, $a|b$ to denote the assertion that a divides b, and $\pi(x) = \sum_{p \leq x} 1$ to denote the usual prime counting function. The effective and asymptotic estimates over primes that we will use are summarized in Appendix C.

We use $\nu_p(a/b) = \nu_p(a) - \nu_p(b)$ 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, $\nu_2(32/27) = 5$ and $\nu_3(32/27) = -3$. If one applies a logarithm to the fundamental theorem of arithmetic, one obtains the identity

$$\sum_p \nu_p(r) \log p = \log r \tag{2.3}$$

for any positive rational r.

For a natural number n, we can write

$$\nu_p(n) = \sum_{j=1}^{\infty} 1_{p^j | n}. \tag{2.4}$$

Upon taking partial sums, we recover Legendre's formula

$$\nu_p(N!) = \sum_{j=1}^{\infty} \left\lfloor \frac{N}{p^j} \right\rfloor = \frac{N - s_p(N)}{p - 1} \tag{2.5}$$

where $s_p(N)$ is the sum of the digits of N in the base p expansion.

Given a putative factorization $\mathcal{B}$ of $N!$, we refer to the quantity $\nu_p\left(\frac{N!}{\prod \mathcal{B}}\right)$ as the *p-surplus* of $\mathcal{B}$ with respect to the target $N!$; if it is negative, we refer to $-\nu_p\left(\frac{N!}{\prod \mathcal{B}}\right)$ as the *p-deficit*, with the multiset being *p-balanced* if the p-surplus (or p-deficit) is zero. Thus, a factorization of $N!$ is achieved if and only if one is balanced at every prime p, whereas a subfactorization is achieved if one is either in balance or surplus at every prime p.

We use the usual asymptotic notation $X = O(Y)$, $X \ll Y$, or $Y \gg X$ to denote an inequality of the form $|X| \leq CY$ for some absolute constant C. We also write $X \asymp Y$ for $X \ll Y \ll X$. For effective estimates, we will use the more precise notation $O_{\leq}(Y)$ to denote any quantity whose magnitude is bounded by exactly at most Y. We also use $O_{\leq}(Y)^+$ to denote a quantity of size $O_{\leq}(Y)$ that is in addition non-negative, that is to say it lies in the interval $[0, Y]$.

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

$$N \log N - N + \log \sqrt{2\pi N} + \frac{1}{12N+1} \leq \log N! \leq N \log N - N + \log \sqrt{2\pi N} + \frac{1}{12N}, \quad (2.6)$$

valid for all natural numbers N.

3. Linear programming

A surprisingly sharp upper bound on $t(N)$ comes from linear programming.

Lemma 3.1 (Linear programming bound). *Let N be an 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*

$$\sum_p w_p \nu_p(j) \geq 1 \quad (3.1)$$

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

$$\sum_p w_p \nu_p(N!) < N. \quad (3.2)$$

Then $t(N) < t$.

Proof. We first observe that the bound (3.1) 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 (3.1) 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 decerasing hypothesis on w_p, we have

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

and hence by the minimality of j we have

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

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

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

for all $p \leq N$. Multiplying by w_p and summing, we conclude from (3.1) that

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

contradicting (3.2). □

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 (discussed below) gives $t(155) = 45$.

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

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

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 $t(155) \geq 45$.

discuss integer programming, need to restrict j to a finite set of "useful" integers

These methods also give quite precise upper and lower bounds for larger values of N, but with quite slow runtime. For instance, with $N = 3 \times 10^5$ and $t = N/3 = 10^5$, the upper bound method can be used to show that any t-admissible factorization has cardinality at most $N + 455$, while the lower bound method produces a t-admissible factorization of exactly this cardinality.

more discussion here

By using the greedy method, Theorem 1.3(ii) can be verified for $N \leq 3 \times 10^5$, and Theorem 1.3(iii) can be verified for $8 \times 10^4 \leq N \leq ???$. The linear programming method can also estabalish Theorem 1.3(iii) in the range $43632 \leq N \leq 8 \times 10^4$. Thus, to resolve these claims, it remains to only establish Theorem 1.3(iii) in the regime $N > ???$.

4. Greedy algorithms

The following simple greedy algorithm gives reasonably good performance to obtain large t-admissible subfactorizations $\mathcal{B}$ of $N!$ for a given choice of t and N:

(0) Initialize $\mathcal{B}$ to be the empty multiset.

(1) If $\mathcal{B}$ is not a factorization, locate the largest prime p which is currently in surplus: $\nu_p(N!/\prod \mathcal{B}) > 0$.
(2) If $N!/\prod \mathcal{B}$ contains a multiple of p that is greater than or equal to t, locate the smallest such multiple, add it to $\mathcal{B}$, and return to Step 1. Otherwise, `HALT` the algorithm.

This procedure clearly halts in finite time to produce a t-admissible subfactoriation of $N!$. For instance, applying this procedure with $N = 9$, $t = 3$ produces the 3-admissible subfactorization

$$\{7 \times 1, 5 \times 1, 3 \times 1, 3 \times 1, 3 \times 1, 3 \times 1, 2 \times 2, 2 \times 2, 2 \times 2\}$$

which recovers the bound $t(9) \geq 3$ from Example 1.1 (though with a slightly different subfactorization, in which the 8 is replaced by 4).

This procedure is efficient for small N, for instance attaining the exact value of $t(N)$ for all $N \leq 79$, though it begins to degrade for larger N; see Figure 4. The performance is also respectable (though not optimal) for medium N; for instance, when $N = 3\times 10^5$ and $t = N/3$, it locates a t-admissible subfactorization of $N!$ of cardinality $N + 372$, which is close to the linear programming limit of $N + 455$.

discuss modifications to the algorithm to make it perform both faster and more accurately

5. The accounting identity

Given a t-admissible multiset $\mathcal{B}$ (which we view as an approximate factorization of $N!$), we can apply (2.3) to the $r := N!/\prod \mathcal{B}$ and rearrange to obtain the *accounting identity*

$$\mathcal{E}_t(\mathcal{B}) + \sum_p \nu_p\left(\frac{N!}{\prod \mathcal{B}}\right) \log p = \log N! - |\mathcal{B}| \log t \tag{5.1}$$

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

$$\mathcal{E}_t(\mathcal{B}) := \sum_{a\in\mathcal{B}} \log \frac{a}{t}. \tag{5.2}$$

Example 5.1. 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 $\nu_2(5!/\prod \mathcal{B}) = 1$, is in balance at 3, and has a 5-deficit of $\nu_2(\prod \mathcal{B}/5!) = 1$, 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 (5.1) become

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

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

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

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 (2.6) 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, as follows.

Lemma 5.3 (Upper bound criterion). *Suppose that $1 \le t \le N$ are such that*

$$\sum_{\frac{t}{\lfloor\sqrt{t}\rfloor} < p \le N} f_{N/t}(p/N) > \log N! - N\log t, \tag{5.3}$$

where $f_{N/t}$ was defined in (1.7). *Then $t(N) < t$.*

Proof. Suppose for contradiction that $t(N) \ge t$, then we can find a t-admissible factorization $\mathcal{B}$ of $N!$. The accounting identity then gives

$$\sum_{a\in\mathcal{B}} \log\frac{a}{t} = \mathcal{E}_t(\mathcal{B}) = \log N! - N\log t. \tag{5.4}$$

We write $f_{N/t}(p/N) = \lfloor\frac{N}{p}\rfloor g_t(p)$, where $g_t(p) := \log(\frac{p}{t}\lceil\frac{t}{p}\rceil)$. We claim that

$$\log\frac{a}{t} \ge g_t(p_{a,1}) + \cdots + g_t(p_{a,k_a}) \tag{5.5}$$

for all $a \in \mathcal{B}$, where $p_{a,1}, \dots, p_{a,k_a}$ are the primes greater than $\frac{t}{\lfloor\sqrt{t}\rfloor}$ that divide a (counting multiplicity). For $k_a = 0$ this is clear since $a \ge t$. For $k_a = 1$, we can write $a = d_a p_{a,1}$ where $p_{a,1} > \frac{t}{\sqrt{t}+1}$ and $d_a \ge \lceil\frac{t}{p_{a,1}}\rceil$, so that

$$\log\frac{a}{t} = \log\left(\frac{p_{a,1}}{t}d_a\right) \ge g_t(p_{a,1}),$$

again giving (5.5). For $k_a \geq 2$, we have $a \geq p_{a,1} \dots p_{a,k}$, hence

$$\begin{aligned}\log \frac{a}{t} - \sum_{j=1}^{k_a} g_t(p_{a,j}) &\geq \sum_{j=1}^{k_a} (\log p_{a,j} - g_t(p_{a,j})) - \log t \\ &= \sum_{j=1}^{k_a} \left(\log t - \log \left\lceil \frac{t}{p_{a,j}} \right\rceil \right) - \log t \\ &\geq \sum_{j=1}^{k_a} \left(\log t - \log \sqrt{t} \right) - \log t \\ &\geq 0\end{aligned}$$

which again gives (5.5). Summing (5.5) over all $a \in \mathcal{B}$ and inserting into (5.4), we conclude that

$$\sum_{p > \frac{t}{\lfloor \sqrt{t} \rfloor}} \nu_p(N!) g_t(p) \leq \log N! - N \log t.$$

By (2.5), we can bound $\nu_p(N!)g_t(p)$ by $\lfloor N/p \rfloor g_t(p) = f_{N/t}(p/N)$. This contradicts (5.3), giving the claim. $\square$

In practice, Lemma 5.3 gives reasonable upper bounds on N, especially when N is large, although for medium N the linear programming method is superior: see Figure 1, Figure 2, Figure 4

We can now prove the upper bound portion of Theorem 1.3(iv):

Proposition 5.4. *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 5.3 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 Taylor expansion and 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

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

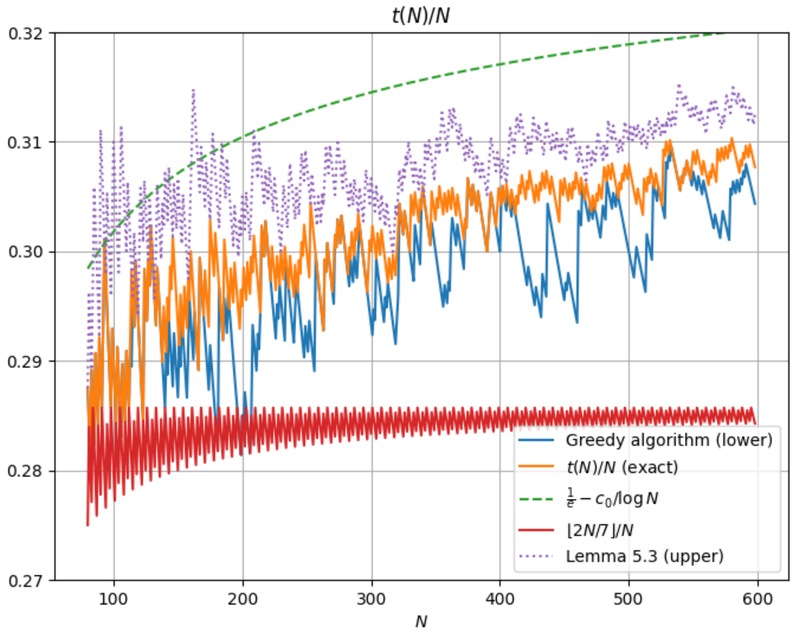

FIGURE 4. An enlarged version of Figure 2, displaying the lower bound from the greedy algorithm and the upper bound from Lemma 5.3. The linear programming upper and lower bounds are exact in this region, except for $N = 155$ in which the upper bound is off by one.

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

$$\sum_{\frac{N}{\log N} \leq p \leq N} f_{N/t}(p/N) \leq ec_0 \frac{N}{\log N} + O\left(\frac{N}{\log^2 N}\right).$$

On the interval $[1/\log N, 1]$, the piecewise smooth function $f_{N/t}$ is bounded by $O(1)$ thanks to (1.8), and has a total variation of $O(\log N)$; the same is then true for the rescaled function $x \mapsto f_{N/t}(x/N)$ on $[N/\log N, 1]$. By Lemma C.2, (C.7), the left-hand side is then

$$\int_{N/\log N}^{N} \left(1 - \frac{2}{\sqrt{x}}\right) f_{N/t}(x/N) \frac{dx}{\log x} + O\left(N \exp(-c\sqrt{\log N})\right)$$

for some $c > 0$. Discarding the $\frac{2}{\sqrt{x}}$ term, performing a change of variable, and using (1.6), we reduce to showing that

$$\int_{1/\log N}^{1} f_{N/t}(x) \frac{\log N}{\log(Nx)}\, dx \leq \int_0^1 f_e(x)\, dx + O\left(\frac{1}{\log N}\right).$$

We have the Taylor approximation

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

Applying (1.8) and the integrability of $\log(1/x)$, we see that the contribution of the error term is acceptable. Applying a further rescaling by $N/et = 1+O(1/\log N)$, we reduce to showing that

$$\int_{N/et\log N}^{N/et} f_{N/t}(Nx/et)\,dx = \int_0^1 f_e(x)\,dx + O\left(\frac{1}{\log N}\right).$$

But observe that $f_{N/t}(Nx/et) = f_e(x)$ unless $\frac{1}{x}$ is within $O(1/\log N)$ of an integer, which one can calculate to occur on a set of measure $O(1/\log N)$ for $x \in [0, N/et]$. By (1.8), both integrands are bounded by $O(1)$ for all $x \in [0, N/et]$, and the claim follows from the triangle inequality. $\square$

We can now establish Theorem 1.3(i):

Proposition 5.5. *One has $t(N)/N < 1/e$ for $N \neq 1, 2, 4$.*

Proof. From existing data on $t(N)$ (or the linear programming method) one can verify this claim for $N < 80$ (see Figure 1), so we assume that $N \geq 80$.

Applying Lemma 5.3, (2.6), it suffices to show that

$$\sum_{p \geq \frac{N/e}{\lfloor\sqrt{N/e}\rfloor}} f_e(p/N) > \frac{1}{2}\log(2\pi N) + \frac{1}{12N}. \tag{5.6}$$

This may be easily verified numerically in the range $80 \leq N \leq 5000$ (see Figure 5). We will discard the $\lfloor\sqrt{N/e}\rfloor$ denominator, and reduce to showing

$$\sum_{N/e<p\leq N} f_e(p/N) > \frac{1}{2}\log(2\pi N) + \frac{1}{12N} \tag{5.7}$$

for $N > 5000$. On $[1/e, 1]$, one can compute

$$\|f_e\|_{\mathrm{TV}^*(1/e,1]} = 4 - 2\log 2$$

so by Lemma C.2, (C.8) (noting that $5000 > 1423e$) we have

$$\sum_{N/e<p\leq N} f_e(p/N)\log p \geq N\int_{1/e}^1 \left(1 - \frac{2}{\sqrt{Nx}}\right) f_e(x)\,dx - (4-2\log 2)\tilde{E}(N).$$

By upper bounding $\log p$ by $\log N$ and lower bounding $\left(1 - \frac{2}{\sqrt{Nx}}\right)$ by $1 - \frac{2}{\sqrt{N/e}}$, it suffices to show that

$$\left(1 - \frac{2}{\sqrt{N/e}}\right)\int_{1/e}^1 f_e(x)\,dx \geq (4-2\log 2)\frac{\tilde{E}(N)}{N} + \frac{\log N\log(2\pi N)}{2N} + \frac{\log N}{12N^2},$$

which is easily verified for $N \geq 5000$ (one has $\int_{1/e}^1 f_e(x)\,dx = \frac{2}{e} - \frac{\log 2}{2} = 0.3891\ldots$ and $4 - 2\log 2 = 2.613\ldots$, while $\tilde{E}(N)/N \leq 0.015$, and the other two terms on the right-hand side are negligible). $\square$

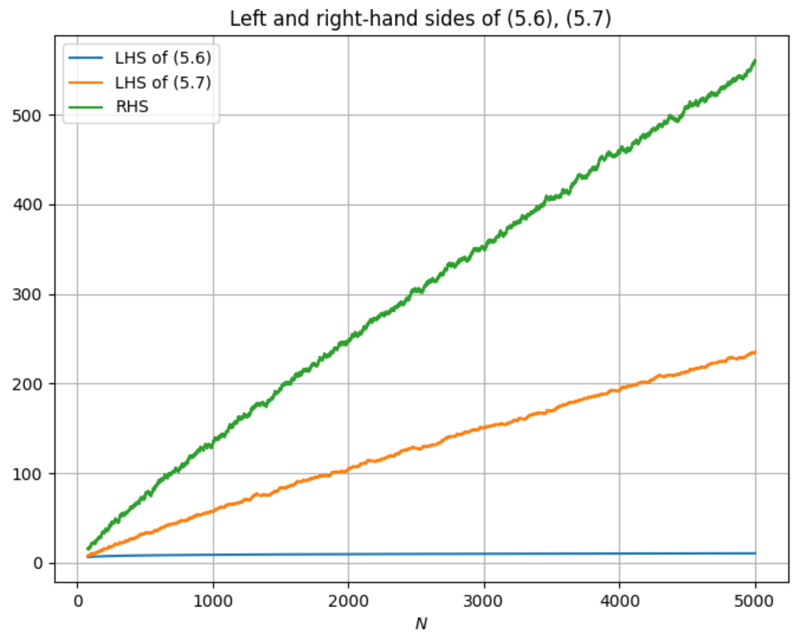

FIGURE 5. A plot of the left and right-hand sides of (??), (5.7) for $80 \leq N < 5000$.

5.1. **Modified approximate factorizations.** In this section we present and then analyze an algorithm that, when given parameters $1 \leq t \leq N$, will attempt to construct a t-admissible subfactorization $\prod \mathcal{B}$ of $N!$ that obeys the criterion in Lemma 5.2. The algorithm will not always succeed, but when it does, it will certify that $t(N) \geq t$.

5.2. **Description of algorithm.** In addition to the given parameters $1 \leq t \leq N$, we require additional natural number parameters A, K obeying the hypotheses

$$K^2(1+\sigma) < t; \quad K\sqrt{N} < t; \quad \sigma < 1; \quad K \geq 3; \tag{5.8}$$

where

$$\sigma := \frac{3N/t}{A}. \tag{5.9}$$

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

With such parameters in hand, we can consider the following algorithm.

(1) Let I denote the elements of the interval[2] $(t, t(1+\sigma)]$ 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

[2] 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 $\pi(x)$.

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 $\nu_p(N!)$ 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 $\nu_p(N!/\prod \mathcal{B}) > 0$, replace $\nu_p(N!/\prod \mathcal{B})$ copies of p in the prime factorizations of elements of $\mathcal{B}^{(3)}$ with $\lceil p \rceil^{\langle 2,3 \rangle}$ 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 $\nu_p(\prod \mathcal{B}/N!) > 0$, 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 $x_i\lceil t/x_i \rceil^{\langle 2,3 \rangle}$ or $x_*\lceil t/x_* \rceil^{\langle 2,3 \rangle}$ 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 criterion in Lemma 5.2 is obeyed by $\mathcal{B}^{(6)}$, then we have successfully established[3] that $t(N) \geq t$. Otherwise, `HALT` the algorithm with an error.

To analyze this algorithm, it will be convenient to divide the set of primes into the following ranges:

- *Tiny primes* $p = 2, 3$.
- *Small primes* $3 < p \leq K$.
- *Borderline small primes* $K < p \leq K(1 + \sigma)$.
- *Medium primes* $K(1 + \sigma) < p \leq t/K$.
- *Large primes* $p > t/K$.

[3] If desired, one could implement the proof of Lemma 5.2 as a final component of this algorithm, that is to say one removes elements from $\mathcal{B}^{(6)}$ to make the cardinality exactly N, and then distributes any surplus primes arbitrarily to create a t-admissible factorization of $N!$ of cardinality exactly N.

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 $\mathcal{B}^{(i)}$, $i = 1, \dots, 6$; 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. The final step $\mathcal{B}^{(7)}$ is an optional one, if one wishes to convert the subfactorization $\mathcal{B}^{(6)}$ to an exact factorization.

5.3. **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

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

with the usual convention $+\infty \times 0 = 0$. 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 can then rewrite the termination condition of Lemma 5.2 (using the fact that $\mathcal{B}^{(6)}$ is a subfactorization of $N!$) as

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

If we assume that $t = N/e^{1+\delta}$ for some $\delta > 0$, we can use the Stirling approximation (2.6) to reduce to the sufficient condition

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

which we choose to normalize as

$$\frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(6)}) + \frac{1}{N} \sum_p \left| v_p\left(\frac{N!}{\prod \mathcal{B}^{(6)}}\right)\right|_{\log p, \infty} \leq \delta + \frac{\log \sqrt{2\pi N}}{N}. \tag{5.10}$$

5.4. **Analysis of Step 6.** Now we analyze Step 6, using the quantity κ_L introduced in (2.1). The main tool we need is the following efficient subfactorization of 3-smooth numbers.

TODO: draw picture of n_0, m_0, etc.

Lemma 5.6 (Efficient subfactorization of 3-smooth numbers)**.** *Let $L \geq 1$. Let $t > 3L$ and let $2^n 3^m$ be a 3-smooth number with n, m positive and obeying the condition*

$$\frac{\log(3L) + \kappa_L}{\log t - \log(3L)} \leq \frac{n \log 2}{m \log 3} \leq \frac{\log t - \log(2L)}{\log(2L) + \kappa_L}. \tag{5.11}$$

Then one can find a t-admissible subfactorization $\mathcal{B}$ of $2^n 3^m$ such that

$$\mathcal{E}_t(\mathcal{B}) \leq \kappa_L \frac{n \log 2 + m \log 3}{\log t} \tag{5.12}$$

and

$$\sum_{p_0 = 2,3} |\nu_{p_0}(2^n 3^m / \mathcal{B})|_{\log p_0, \infty} \leq 2(\log t + \kappa_L). \tag{5.13}$$

In practice, $\log t$ will be significantly larger than $\log(2L)$ or $\log(3L)$, and so the hypothesis (5.11) will be relatively 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

$$L \leq \frac{t}{2^{n_0}} < 2L \tag{5.14}$$

and

$$L \leq \frac{t}{3^{m_0}} < 3L. \tag{5.15}$$

From (2.1), the 3-smooth numbers $2^{n_1} 3^{m_1} := \lceil t/2^{n_0} \rceil^{\langle 2,3 \rangle}$, $2^{n_2} 3^{m_2} := \lceil t/3^{m_0} \rceil^{\langle 2,3 \rangle}$ obey the estimates

$$\frac{t}{2^{n_0}} \leq 2^{n_1} 3^{m_1} \leq e^{\kappa_L} \frac{t}{2^{n_0}} \tag{5.16}$$

and

$$\frac{t}{3^{m_0}} \leq 2^{n_2} 3^{m_2} \leq e^{\kappa_L} \frac{t}{3^{m_0}}, \tag{5.17}$$

or equivalently

$$t \leq 2^{n_0 + n_1} 3^{m_1}, 2^{n_2} 3^{m_0 + m_2} \leq e^{\kappa_L} t. \tag{5.18}$$

We can use (5.14), (5.16) to bound

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

(with the convention that this bound is vacuously true for $m_1 = 0$). Similarly, from (5.15), (5.17) we have

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

and hence by (5.11)

$$\frac{n_2}{m_0+m_2} \leq \frac{n}{m} \leq \frac{n_0+n_1}{m_1}. \tag{5.19}$$

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 (5.18), each term $2^{n'}3^{m'}$ here is admissible and contributes a t-excess of at most κ_L, which is in turn bounded by $\kappa_L \frac{n'\log 2 + m'\log 3}{\log t}$. Adding these bounds together, we obtain (5.12).

As a subfactorization of $2^n 3^m$, the multiset $\mathcal{B}$ has a 2-surplus of at most $n_0+n_1+n_2$ and a 3-surplus of at most $m_0+m_2+m_1$, hence

$$\sum_{p_0=2,3} \nu_{p_0}\left(\frac{2^n3^m}{\prod \mathcal{B}}\right)\log p_0 \leq \log 2^{n_0+n_1}3^{m_1} + \log 2^{n_2}3^{m_0+m_2},$$

and the bound (5.13) follows from (5.18). □

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

Proposition 5.7. *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 5.2 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)*

$$\begin{aligned}\frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(5)}) + \frac{1}{N}\sum_{p>3}\left|\nu_p\left(\frac{N!}{\prod \mathcal{B}^{(5)}}\right)\right|_{\log p,\infty} \\ + \frac{\kappa_L \log\sqrt{12}}{\log t} + \frac{3}{2}\frac{\log N}{N} \leq \delta + \frac{\log\sqrt{2\pi}}{N}.\end{aligned} \tag{5.20}$$

(ii) *(Large surpluses at tiny primes) One has*

$$\frac{1}{N}\sum_{p=2,3}\nu_p\left(\prod \mathcal{B}^{(5)}\right)\log p < \min(Q_{N,t,L}, Q'_{N,t,L}) \tag{5.21}$$

where

$$Q_{N,T,L} := \log 2 - \frac{\log(3L)+\kappa_L}{\log t - \log(3L)}\frac{\log 3}{2} - \frac{\log(2N)}{N} \tag{5.22}$$

and

$$Q'_{N,T,L} := \frac{\log 3}{2} - \frac{\log(2L)+\kappa_L}{\log t - \log(2L)}\log 2 - \frac{\log(3N)}{N}. \tag{5.23}$$

Then $t(N) \geq t$.

Since $\log 2 = 0.6931\dots$ is larger than $\frac{\log 3}{2} = 0.5493\dots$, the $Q'_{N,T,L}$ term will dominate in practice.

Proof. Write $n := \nu_2(N!/\prod \mathcal{B}^{(5)})$ and $m := \nu_3(N!/\prod \mathcal{B}^{(5)})$. From (2.5) we have $n \leq N$ and $m \leq N/2$, hence

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

From (5.21), (2.5) we have

$$\begin{aligned} n &= \nu_2(N!) - \nu_2\left(\prod \mathcal{B}^{(5)}\right) \\ &> N - (1+\log N/\log 2) - \frac{NQ_{N,T,L}}{\log 2} \\ &\geq \frac{\log(3L)+\kappa_L}{\log t - \log(3L)}\frac{N\log 3}{2\log 2} \\ &\geq \frac{\log(3L)+\kappa_L}{\log t - \log(3L)}\frac{\nu_3(N!)\log 3}{\log 2} \end{aligned}$$

and similarly

$$m > \frac{\log(2L)+\kappa_L}{\log t - \log(2L)}\frac{\nu_2(N!)\log 2}{\log 3}.$$

In particular, n, m are positive. Since we also have $n \leq \nu_2(N!)$ and $m \leq \nu_3(N!)$, the condition (5.11) holds. Applying Lemma 5.6, we can find a subfactorization $\mathcal{B}'$ of 2^n3^m with an excess of at most $(\kappa_L \log\sqrt{12})\frac{N}{\log t}$, and with

$$\sum_{p_0=2,3}\left|\nu_{p_0}\left(\frac{2^n3^m}{\prod \mathcal{B}'}\right)\right|_{\log p_0,\infty} \leq 2(\log t + \kappa_L) \leq 2\log N$$

where we have used (2.2) and the fact that $\log t \leq \log N - 1$. Then $\mathcal{B}^{(6)} = \mathcal{B}^{(5)} \cup \mathcal{B}'$ is another t-admissible multiset, and from (5.20), we obtain the previous sufficient condition (5.10). ☐

5.5. **Analysis of Step 5.** We can use the surplus tiny primes to efficiently deal with larger surplus primes, as follows.

Proposition 5.8. *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 $\delta > 0$, and (5.8) holds. Suppose that the algorithm in Section 5.2 with the indicated parameters reaches the end of Step 4 to produce a multiset $\mathcal{B}^{(4)}$ such that the* *quantities*

$$Y_1 := \frac{1}{N} \sum_{3<p\leq K} \left| \nu_p \left(\frac{N!}{\prod \mathcal{B}^{(3)}} \right) \right|_{\frac{\log p}{\log(t/K^2)}, \infty} \tag{5.24}$$

$$Y_2 := \frac{1}{N} \sum_{K<p\leq t/K} \left| \nu_p \left(\frac{N!}{\prod \mathcal{B}^{(3)}} \right) \right|_{1,\infty} \tag{5.25}$$

obeying the bounds

$$\begin{aligned} &\frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(4)}) + \kappa_K Y_1 + \kappa_K Y_2 \\ &\quad + \frac{\kappa_L \log \sqrt{12}}{\log t} + \frac{3}{2}\frac{\log N}{N} \leq \delta. \end{aligned} \tag{5.26}$$

and

$$\begin{aligned} &\frac{1}{N} \sum_{p_0=2,3} \nu_{p_0}\left(\prod \mathcal{B}^{(4)}\right) \log p_0 + (\log(K^2) + \kappa_K) Y_1 + (\log(t/K) + \kappa_K)(Y_2 + \frac{1}{N}) \\ &\leq \min(Q_{N,t,L}, Q'_{N,t,L}). \end{aligned}$$

Then $t(N) \geq t$.

Proof. By (5.26), $\mathcal{B}^{(4)}$ is a subfactorization of $N!$, and by construction it is in balance at all large primes $p > t/K$. Consider all the p-surplus small primes $3 < p \leq K$, thus each such prime is considered with multiplicity $\nu_p(N!/\prod \mathcal{B}^{(4)})$. Using the greedy algorithm, one can factor the product of all these primes into M factors $c_1, \dots, c_M$ in the interval $(t/K^2, t/K]$ for some M, times at most one exceptional factor c_* in $(1, t/K^2]$. We have the bound

$$\left(\frac{t}{K^2}\right)^{M'} \leq \prod_{3<p\leq K} \nu_p \left(\frac{N!}{\prod \mathcal{B}^{(4)}} \right)$$

and hence on taking logarithms and using (5.24) we have

$$M \leq NY_1.$$

Also, the number of borderline or medium primes $K < p \leq t/K$ appearing in $N!/\prod \mathcal{B}^{(4)}$, counting multiplicity, is NY_2.

For each of the M factors c_i, we introduce the 3-smooth number $\lceil t/c_i \rceil^{\langle 2,3 \rangle} = 2^{n_i}3^{m_i}$, which by (2.1) 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 $\lceil t/c_* \rceil^{\langle 2,3 \rangle} = 2^{n_*}3^{m_*}$ in the interval $[t/c_*, e^{\kappa_K} t/c_*]$. Finally, for the Y_2 borderline or medium primes $K < p \leq t/K$ discussed above, we introduce the 3-smooth number $\lceil t/p \rceil^{\langle 2,3 \rangle} = 2^{n_p}3^{m_p}$, which lies in the interval $[t/p, e^{\kappa_K} t/p]$. We adjoin the 3-smooth numbers $\lceil t/c_i \rceil^{\langle 2,3 \rangle} c_i = 2^{n_i}3^{m_i} c_i$ for $i = 1, \dots, M$ as well $\lceil t/c_* \rceil^{\langle 2,3 \rangle} c_* = 2^{n_*}3^{m_*} c_*$ and $\lceil t/p \rceil^{\langle 2,3 \rangle} p = 2^{n_p}3^{m_p} c_p$ to the t-admissible multiset $\mathcal{B}^{(4)}$ to create a new t-admissible multiset $\mathcal{B}^{(5)}$. The quantities $\log \lceil t/c_i \rceil^{\langle 2,3 \rangle} = n_i \log 2 + m_i \log 3$ are bounded by $\log K^2 + \kappa_K$, and the quantities $\log \lceil t/c_* \rceil^{\langle 2,3 \rangle} = n_* \log 2 + m_* \log 3$ and $\log \lceil t/p \rceil^{\langle 2,3 \rangle} = n_p \log 2 + m_p \log 3$ are similarly bounded by $\log(t/K) + \kappa_K$, hence

$$\begin{aligned}
&\sum_{p_0=2,3} \nu_{p_0}\left(\prod \mathcal{B}^{(5)}\right) \log p_0 \\
&\quad \leq \sum_{p_0=2,3} \nu_{p_0}\left(\prod \mathcal{B}^{(4)}\right) \log p_0 \\
&\quad + NY_1(\log K^2 + \kappa_K) + (NY_2 + 1)(\log(t/K) + \kappa_K).
\end{aligned}$$

Each of the new factors in $\mathcal{B}^{(5)}$ contributes an excess of at most κ_K, so that

$$\mathcal{E}_t(\mathcal{B}^{(5)}) \leq \mathcal{E}_t(\mathcal{B}^{(4)}) + \kappa_K(NY_1 + NY_2 + 1).$$

We conclude that $\mathcal{B}^{(5)}$ obeys the hypotheses of Proposition 5.7 (using (2.2) to bound κ_K by $\log\sqrt{2\pi}$ to clean up some lower order terms), and the claim follows. □

5.6. **Analysis of Step 4.** The surplus tiny primes can also be used to deal with any larger primes that are in deficit, as follows.

Proposition 5.9. *Let $L \geq 1$. Let $9L < t = N/e^{1+\delta}$ for some $\delta > 0$ be such that* (5.8) *holds, and suppose that the algorithm reaches the end of Step 3 to produce a multiset $\mathcal{B}^{(3)}$ such that the quantities*

$$Y_1^+ := \frac{1}{N} \sum_{3<p\leq K} \left|\nu_p\left(\frac{N!}{\prod \mathcal{B}^{(3)}}\right)\right|_{\frac{\log p}{\log(t/K^2)},0} \tag{5.27}$$

$$Y_1^- := \frac{1}{N} \sum_{3<p\leq K} \left|\nu_p\left(\frac{N!}{\prod \mathcal{B}^{(3)}}\right)\right|_{0,1}. \tag{5.28}$$

$$Y_2^\pm := \frac{1}{N} \sum_{K<p\leq t/K} \left|\nu_p\left(\frac{N!}{\prod \mathcal{B}^{(3)}}\right)\right| \tag{5.29}$$

obey the hypotheses

$$\begin{aligned}
&\frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(3)}) + \kappa_K Y_1^+ + \kappa_5 Y_1^- + \kappa_K Y_\pm \\
&\quad + \frac{\kappa_L \log\sqrt{12}}{\log t} + \frac{3}{2}\frac{\log N}{N} \leq \delta.
\end{aligned} \tag{5.30}$$

and

$$\begin{aligned}
&\frac{1}{N}\sum_{p_0=2,3} \nu_{p_0}\left(\prod \mathcal{B}^{(3)}\right) \log p_0 \\
&\quad + (\log K^2 + \kappa_K)Y_1^+ + (\log K + \kappa_5)Y_1^- + (\log(t/K) + \kappa_K)(Y_2^\pm + \frac{1}{N}) \\
&\quad \leq \min(Q_{N,t,L}, Q'_{N,t,L})
\end{aligned} \tag{5.31}$$

Then $t(N) \geq t$.

Proof. Suppose there is a non-tiny prime $p > 3$ with a positive p-deficit $|\nu_p(N!/\prod \mathcal{B}^{(3)})|_{0,1} > 0$. 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 $\lceil p\rceil^{\langle 2,3\rangle} = 2^{n_p}3^{m_p}$, which increases that factor by at most $\exp(\kappa_p)$ thanks to (2.1). This procedure reduces the p-deficit by one, adds at most κ_p to the t-excess, and increments $\sum_{p_0=2,3} \nu_{p_0}(N!/\prod \mathcal{B}^{(3)}) \log p_0$ by $n_p \log 2 + m_p \log 3$. 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

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

and hence

$$\frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(4)}) + \kappa_K Y_1 + \kappa_K Y_2$$

$$\frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(4)}) + \kappa_K Y_1^+ + \kappa_5 Y_1^- + \kappa_K Y_2^\pm ..$$

Similarly we have

$$\sum_{p'=2,3} \nu_p \left(\prod \mathcal{B}^{(4)} \right) \log p \leq \sum_{p'=2,3} \nu_p \left(\prod \mathcal{B}^{(3)} \right) \log p + \sum_{p>3} \left| \nu_p \left(\frac{N!}{\prod \mathcal{B}^{(3)}} \right) \right|_{0,\log p+\kappa_p}$$

and hence

$$\frac{1}{N} \sum_{p_0=2,3} \nu_{p_0} \left(\prod \mathcal{B}^{(4)} \right) \log p_0 + (\log(K^2) + \kappa_K) Y_1 + (\log(t/K) + \kappa_K)(Y_2 + \frac{1}{N})$$
$$\leq \frac{1}{N} \sum_{p_0=2,3} \nu_{p_0} \left(\prod \mathcal{B}^{(3)} \right) \log p_0 + (\log(K^2) + \kappa_K) Y_1^+ + (\log K + \kappa_K) Y_2^+ + (\log(t/K) + \kappa_K)(Y_2^\pm + \frac{1}{N}).$$

The hypotheses of Proposition 5.8 are now satisfied, and we are done. □

5.7. **Analysis of Step 3.** From direct inspection of Step 3, we see that borderline or medium primes $K < p \leq t/K$ are unaffected by this step:

$$\nu_p \left(\frac{N!}{\prod \mathcal{B}^{(3)}} \right) = \nu_p \left(\frac{N!}{\prod \mathcal{B}^{(2)}} \right). \tag{5.32}$$

For tiny or small primes $p_1 \leq K$, we instead have

$$\nu_{p_1} \left(\frac{N!}{\prod \mathcal{B}^{(3)}} \right) = \nu_{p_1} \left(\frac{N!}{\prod \mathcal{B}^{(2)}} \right) - \sum_{p>t/K} \nu_p(N!) \nu_{p_1}(\lceil t/p \rceil).$$

From (5.8) we see that all large primes $p > t/K$ are larger than $\sqrt{N}$, hence by (2.5) $\nu_p(N!) = \lfloor N/p \rfloor$. Meanwhile, $\lceil t/p \rceil$ is at most K. Thus, making the change of variables $m := \lceil t/p \rceil$, we can also write

$$\nu_{p_1} \left(\frac{N!}{\prod \mathcal{B}^{(3)}} \right) = \nu_{p_1} \left(\frac{N!}{\prod \mathcal{B}^{(2)}} \right) - \sum_{m \leq K} \nu_{p_1}(m) \sum_{\frac{t}{m} \leq p < \frac{t}{m-1}} \lfloor \frac{N}{p} \rfloor. \tag{5.33}$$

Finally, the t-excess after Step 3 can be computed as

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

and hence by (2.5) and (1.7)

$$\mathcal{E}_t(\mathcal{B}^{(3)}) = \mathcal{E}_t(\mathcal{B}^{(2)}) + \sum_{t/K<p\le N} f_{N/t}(p/N).$$

If we bound $f_{N/t}(p/N)$ by $\frac{1}{\log(t/K)} f_{N/t}(p/N)\log p$ and apply Lemma C.2, we conclude

$$\frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(3)}) \le \frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(2)}) + \frac{1}{\log(t/K)}\int_{t/NK}^{1} f_{N/t}(x)\,dx + \|f_{N/t}\|_{\mathrm{TV}^*((t/NK,1])}\frac{E(N)}{N\log(t/K)}. \tag{5.34}$$

5.8. **Analysis of Step 2.** Step 2 removes A copies of every factor of the form mp when $p > t/K$ and $mp \in I$, which forces m coprime to 6 and $m \le K(1+\sigma)$. Thus this procedure does not affect medium primes $K(1+\sigma) < p \le t/K$:

$$\nu_p\left(\frac{N!}{\prod \mathcal{B}^{(2)}}\right) = \nu_p\left(\frac{N!}{\prod \mathcal{B}^{(1)}}\right). \tag{5.35}$$

It also does not generate any tiny primes $p_0 = 2, 3$, thus:

$$\sum_{p_0=2,3} \nu_{p_0}\left(\frac{N!}{\prod \mathcal{B}^{(2)}}\right)\log p_0 = 0. \tag{5.36}$$

For borderline primes $K < p_1 \le K(1+\sigma)$, this step removes A copies of $p_1 p$ whenever $p_1 p \in I$ and $p \ge t/K$, adding to the p_1-surplus p_1, but otherwise does not affect this surplus; thus

$$\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(2)}}\right) = \nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(1)}}\right) + A\left(\pi\left(\frac{t(1+\sigma)}{p_1}\right) + \pi\left(\frac{t}{K}\right)\right). \tag{5.37}$$

For small primes $3 < p_1 \le K$, we instead have

$$\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(2)}}\right) = \nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(1)}}\right) + A\sideset{}{^*}\sum_{m\le K(1+\sigma)} \nu_{p_1}(m)\left(\pi\left(\frac{t(1+\sigma)}{m}\right) - \pi\left(\frac{t}{\min(K,m)}\right)\right),$$

where the notation $\sum^*$ indicates that the summation variable m is required to be coprime to 6. We can split this as

$$\nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(2)}}\right) = \nu_{p_1}\left(\frac{N!}{\prod \mathcal{B}^{(1)}}\right) + A\sideset{}{^*}\sum_{m\le K} \nu_{p_1}(m)\left(\pi\left(\frac{t(1+\sigma)}{m}\right) - \pi\left(\frac{t}{m}\right)\right) + NZ_{p_1} \tag{5.38}$$

where

$$Z_{p_1} := \frac{A}{N}\sideset{}{^*}\sum_{K<m\le K(1+\sigma)} \nu_{p_1}(m)\left(\pi\left(\frac{t(1+\sigma)}{m}\right) - \pi\left(\frac{t}{K}\right)\right). \tag{5.39}$$

As for the excess, we simply use the trivial bound

$$\mathcal{E}_t(\mathcal{B}^{(2)}) \le \mathcal{E}_t(\mathcal{B}^{(1)}). \tag{5.40}$$

5.9. **Analysis of Step 1.** To count elements coprime to 6, we use the following lemma:

Lemma 5.10. *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* $\frac{b-a}{3} + O_{\leq}(4/3)$*.*

TODO: display the sawtooth function used in the proof

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. ☐

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, and noting that $t\sigma = 3N/A$, this is

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

Bounding $\frac{1}{t+h}$ by $\frac{1}{t}$ and applying Lemma 5.10, we conclude that

$$\mathcal{E}_t(\mathcal{B}^{(1)}) \leq A \int_0^{3N/A} \left(\frac{h}{3} + \frac{4}{3}\right) \frac{dh}{t} = \frac{3N^2}{2tA} + 4. \tag{5.41}$$

Next, we compute p-valuations $v_p(\mathcal{B}^{(1)})$. For small, borderline, or large primes $3 < p \leq t/K$, one has

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

Meanwhile, from (2.5) one has

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

and thus

$$v_p(N!/\mathcal{B}^{(1)}) = O_{\leq}\left(\frac{4A + 3}{3} \left\lceil \frac{\log N}{\log p} \right\rceil\right). \tag{5.42}$$

Combining these bounds with (5.32), (5.33), (5.34), (5.35), (5.37), (5.38), (5.40) and (5.27), (5.28), (5.29), we see that

$$\frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(3)}) \leq \frac{3N}{2tA} + \frac{4}{N} + \frac{1}{\log(t/K)}\int_{t/NK}^{1} f_{N/t}(x)\,dx + \frac{1}{\log(t/K)}\|f_{N/t}\|_{\mathrm{TV}^*((t/NK,1])}\frac{E(N)}{N\log(t/K)} \tag{5.43}$$

$$Y_1^+ \leq \frac{4A+3}{3N}\sum_{3<p_1\leq K}\left\lceil\frac{\log N}{\log p_1}\right\rceil\frac{\log p_1}{\log(t/K^2)} + \sum_{3<p_1\leq K} Z_{p_1}\frac{\log p}{\log(t/K^2)} + \sum_{3<p\leq K}|W_{p_1}|_{\frac{\log p}{\log(t/K^2)},0} \tag{5.44}$$

$$Y_1^- \leq \frac{4A+3}{3N}\sum_{3<p_1\leq K}\left\lceil\frac{\log N}{\log p_1}\right\rceil + \sum_{3<p\leq K}|W_{p_1}|_{0,1} \tag{5.45}$$

$$Y_2^\pm \leq \frac{4A+3}{3N}\sum_{K<p\leq t/K}\left\lceil\frac{\log N}{\log p}\right\rceil + \frac{A}{N}\sum_{K<p\leq K(1+\sigma)}\left(\pi\left(\frac{t(1+\sigma)}{p_1}\right)-\pi\left(\frac{t}{K}\right)\right) \tag{5.46}$$

$$\sum_{p_0=2,3}\nu_{p_0}\left(\prod\mathcal{B}^{(3)}\right)\log p_0 \leq \sum_{p_0=2,3}\sum_{m\leq K}(\nu_{p_0}(m)\log p_0)\sum_{\frac{t}{m}\leq p<\frac{t}{m-1}}\lfloor\frac{N}{p}\rfloor \tag{5.47}$$

where for any small prime $3 < p_1 \leq K$, we define W_{p_1} to be the quantity

$$W_{p_1} := \frac{A}{N}\sum_{m\leq K}^{*}\nu_{p_1}(m)\left(\pi\left(\frac{t(1+\sigma)}{m}\right)-\pi\left(\frac{t}{m}\right)\right) - \frac{1}{N}\sum_{m\leq K}\nu_{p_1}(m)\sum_{\frac{t}{m}\leq p<\frac{t}{m-1}}\lfloor\frac{N}{p}\rfloor. \tag{5.48}$$

A key point will be that in practice, the W_{p_1} will be positive (or only barely negative), so that their contribution to (5.45) vanishes (or is small).

6. The asymptotic regime

With the above estimates, we can now establish the lower bound in Theorem 1.3(iv). Thus we aim to show that $t(N) \geq t$ for sufficiently large N, where

$$t := \frac{N}{e} - \frac{c_0 N}{\log N} + \frac{N}{\log^{1+c_1} N} \asymp N \tag{6.1}$$

and $0 < c_1 < 1$ is a small absolute constant. 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

$$A := \lfloor \log^2 N \rfloor \tag{6.2}$$

$$K := \lfloor \log^3 N \rfloor \tag{6.3}$$

$$L := N^{0.1}, \tag{6.4}$$

so from (5.9) one has

$$\sigma = \frac{3N}{tA} \asymp \frac{1}{A} \asymp \frac{1}{\log^2 N}.$$

The conditions (5.8) and $t > 9L$ are easily verified for N large enough.

By Proposition 5.9, it suffices to verify the criteria (5.30), (5.31). From (**??**), (5.23) we have

$$Q_{N,t,L}, Q'_{N,t,L} \asymp 1$$

so by Lemma A.3 it will suffice to establish the bounds

$$\frac{1}{N}\mathcal{E}_t(\mathcal{B}^{(3)}) \leq \frac{ec_0}{\log N} + O\left(\frac{(\log\log N)^{O(1)}}{\log^2 N}\right) \tag{6.5}$$

$$Y_1^+ \ll \frac{(\log\log N)^{O(1)}}{\log N} \tag{6.6}$$

$$Y_1^- \ll \frac{(\log\log N)^{O(1)}}{\log^2 N} \tag{6.7}$$

$$Y_2^\pm \ll \frac{(\log\log N)^{O(1)}}{\log^2 N} \tag{6.8}$$

$$\frac{1}{N}\sum_{p_0=2,3} \nu_{p_0}\left(\prod \mathcal{B}^{(3)}\right)\log p_0 \ll \frac{(\log\log N)^{O(1)}}{\log N}. \tag{6.9}$$

We begin with (6.5). The function $f_{N/t}$ is piecewise monotone on $(t/NK, 1]$ with $O(K)$ pieces, and is bounded by 1, so that

$$\|f_{N/t}\|_{\mathrm{TV}^*((t/NK,1])} \ll K.$$

Applying (5.43), (C.7), (6.1), (6.2), (6.3) we see that

$$\mathcal{E}_t(\mathcal{B}^{(3)}) \leq \frac{1}{\log(t/K)}\int_{t/NK}^1 f_{N/t}(x)\,dx + O\left(\frac{1}{\log^2 N}\right).$$

But by repeating the arguments used to prove Proposition 5.5 we have

$$\begin{aligned}\frac{1}{\log(t/K)}\int_{t/NK}^1 f_{N/t}(x)\,dx &= \left(1 + O\left(\frac{\log\log N}{\log^2 N}\right)\right)\int_{1/eK}^{N/et} f_{N/t}(etx/N)\,dx \\ &= \left(1 + O\left(\frac{\log\log N}{\log^2 N}\right)\right)\left(ec_0 + O\left(\frac{\log\log N}{\log^2 N}\right)\right),\end{aligned}$$

giving the claim.

From (5.46) and the Brun–Titchmarsh inequality one has

$$Y_2^{\pm} \ll \frac{A}{N}\frac{t/K}{\log N} + \frac{A}{N}\frac{K\sigma}{\log K}\frac{t\sigma/K}{\log K}$$

and the claim (6.8) follows from (6.2), (6.3), (6.1).

From (5.47) and the Brun–Titchmarsh inequality as well as the trivial bound $\nu_{p_0}(m) \ll \log\log N$ for $m \ll K$, one has

$$\sum_{p_0=2,3} \nu_{p_0}\left(\prod \mathcal{B}^{(3)}\right)\log p_0 \ll \sum_{m\leq K}(\log\log N)\frac{t}{m\log N}$$

and from summing the harmonic series we obtain (6.9).

Applying similar estimates to (5.39) gives

$$\begin{aligned} Z_{p_1} &\ll \frac{A}{N}\sum_{K<m\leq K(1+\sigma):\, p_1|m}(\log\log N)\frac{\sigma t}{K\log N} \\ &\ll \frac{\log\log N}{K\log N}(\frac{K\sigma}{p_1}+1) \\ &\ll \frac{\log\log N}{p_1\log^2 N} + \frac{\log\log N}{K\log N} \end{aligned}$$

and hence by (5.44) and Mertens' theorem

$$\begin{aligned} Y_1^{+} &\ll \frac{A}{N}\sum_{p_1\leq K}\log N\frac{\log\log N}{\log N} + \sum_{p_1\leq K}\frac{\log\log N}{p_1\log^2 N} + \frac{\log\log N}{K\log N} \\ &\quad + \sum_{3<p\leq K}|W_{p_1}|_{\frac{\log p}{\log(t/K^2)},0} \\ &\ll \sum_{3<p\leq K}|W_{p_1}|_{\frac{\log p}{\log(t/K^2)},0} + \frac{(\log\log N)^{O(1)}}{\log N} \end{aligned}$$

while from (5.45) we similarly have

$$\begin{aligned} Y_1^{-} &\ll \frac{A}{N}\sum_{p_1\leq K}\log N + \sum_{3<p\leq K}|W_{p_1}|_{0,1} \\ &\ll \sum_{3<p\leq K}|W_{p_1}|_{0,1} + \frac{\log^{O(1)} N}{N} \end{aligned}$$

and so to verify the remaining claims (6.6), (6.7) it will suffice from Mertens' theorem to show the bounds

$$|W_{p_1}|_{1,0} \ll \frac{(\log\log N)^{O(1)}}{p_1\log N} \tag{6.10}$$

$$|W_{p_1}|_{0,1} \ll \frac{(\log\log N)^{O(1)}}{p_1\log^2 N} \tag{6.11}$$

for all small primes $3 < p_1 \leq K$.

For the first bound (6.10), we crudely discard the negative term in (5.48) and use the Brun–Titchmarsh inequality and the crude bound

$$\nu_{p_1}(m) \ll 1_{p_1|m} \log\log N \tag{6.12}$$

for $m \ll K$ to obtain

$$\begin{aligned}
|W_{p_1}|_{1,0} &\leq \frac{A}{N} \sum_{m \leq K}^{*} \nu_{p_1}(m) \left(\pi\left(\frac{t(1+\sigma)}{m}\right) - \pi\left(\frac{t}{m}\right) \right) \\
&\ll \frac{A}{N} \sum_{m \leq K : p_1 | m} (\log\log N) \frac{t\sigma/m}{\log N} \\
&\ll \frac{(\log\log N)^2}{p_1 \log N}
\end{aligned}$$

as required. For the second bound (6.11), we need to be more careful. From Lemma C.2, (C.7) we have

$$\begin{aligned}
&\frac{A}{N} \sum_{m \leq K}^{*} \nu_{p_1}(m) \left(\pi\left(\frac{t(1+\sigma)}{m}\right) - \pi\left(\frac{t}{m}\right) \right) \\
&\quad = \frac{A}{N} \sum_{m \leq K}^{*} \nu_{p_1}(m)(1 + O(\frac{\log\log N}{\log N})) \frac{t\sigma m}{\log N} \\
&\quad = \frac{1}{\log N} \sum_{m \leq K}^{*} \frac{3\nu_{p_1}(m)}{m} + O\left(\frac{N(\log\log N)^{O(1)}}{p_1 \log^2 N} \right)
\end{aligned}$$

where we used (6.12) to control the error. We can also use Lemma C.2, (C.7) to bound

$$\begin{aligned}
\frac{1}{N} \sum_{m \leq K} \nu_{p_1}(m) \sum_{\frac{t}{m} \leq p < \frac{t}{m-1}} \lfloor \frac{N}{p} \rfloor &\leq \left(1 + O\left(\frac{\log\log N}{\log N} \right) \right) \frac{1}{N} \sum_{m \leq K} \nu_{p_1}(m) \sum_{\frac{t}{m} \leq p < \frac{t}{m-1}} \frac{N}{p} \log p \\
&\leq \left(1 + O\left(\frac{\log\log N}{\log N} \right) \right) \frac{1}{N} \sum_{m \leq K} \nu_{p_1}(m) \int_{\frac{t}{m}}^{\frac{t}{m-1}} \frac{N}{x}\, dx + O\left(\frac{1}{p_1 \log^2 N} \right) \\
&\leq \sum_{m \leq K} \nu_{p_1}(m) \log \frac{m}{m-1} + O\left(\frac{(\log\log N)^{O(1)}}{p_1 \log^2 N} \right)
\end{aligned}$$

where we again used (6.12) to control the error. Inserting these bounds into (5.48), it will now suffice to establish the following bound.

Lemma 6.1 (Key inequality)**.** *For any prime $p \geq 5$ and any real $K > 0$, we have*

$$0 \leq \sum_{m \leq K}^{*} \nu_p(m) \frac{3}{m} - \sum_{m \leq K} \nu_p(m) \log \frac{m}{m-1} \leq \frac{2}{p-1}.$$

But this can be easily verified; see Appendix B. The proof of (6.1) is now complete.

7. Guy–Selfridge conjecture

We now establish the Guy–Selfridge conjecture $t(N) \geq N/3$ in the range

$$N \geq ???.$$

We will apply Proposition 5.9 with the choice of parameters

$$\begin{aligned} t &:= N/3 \\ A &:= ??? \\ K &:= 342 \\ L &:= 342. \end{aligned}$$

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

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

From Lemma A.1, we have

$$\kappa_K = \kappa_L \leq \log \frac{9}{8} = 0.11778\dots \tag{7.1}$$

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

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

Direct numerical calculation (cf. Figure 6) reveals that

$$\begin{aligned} \int_{1/3K}^{1} f(x)\, dx &\leq 0.9201 \\ \|f\|_{\mathrm{TV}^*(1/3K,1]} &\leq 2044 \end{aligned}$$

and thus

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

The 2044 factor may seem large, but for large N the quantity $\frac{E(N)}{N}$ is so small that this term is in fact negligible.

We can directly compute $\pi(K) - \pi(3) = 66$.

Appendix A. Powers of 2 and 3

We now obtain good bounds on the quantity κ_L introduced in (2.1). Clearly κ_L is a non-increasing function of L with $\kappa_1 = \log 2$. The following lemma gives improved control on κ_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*

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

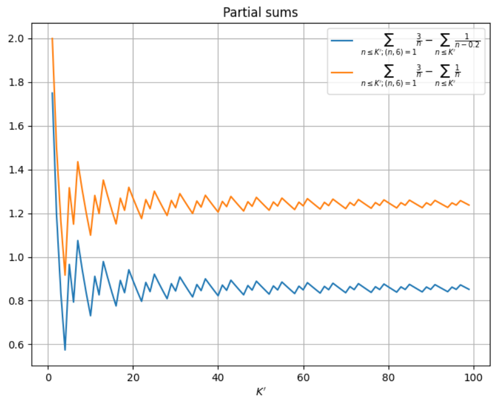

FIGURE 6. A plot of $f(x)$. The integral $c_1 = \int_{1/K}^1 f(x)\,dx \approx 0.9200$ is slightly larger than $ec_0 \approx 0.8244$.

then

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

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

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

so we are done in this case. Now suppose that $t > 2^{n_2-1}3^{m_1-1}$. If we write $\lceil t \rceil^{\langle 2,3\rangle} = 2^n 3^m$ be the smallest 3-smooth number that is at least t, then we must have $n \ge n_2$ or $m \ge m_1$ (or both). Thus at least one of $\frac{2^{n_1}}{3^{m_1}} 2^n 3^m$ and $\frac{3^{m_2}}{3^{n_2}} 2^n 3^m$ 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, $\kappa_{4.5} \le 0.28768\dots$ and $\kappa_{40.5} \le 0.16989\dots$.

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	**1**	**0**	$1/2 = 0.5$	$\log 2 = 0.69314\dots$
1	**1**	2	1	$2^2/3 = 1.33\dots$	$\log(3/2) = 0.40546\dots$
3	2	**2**	**1**	$3^2/2 = 4.5$	$\log(2^2/3) = 0.28768\dots$
3	2	**5**	**3**	$3^4/2 = 40.5$	$\log(2^5/3^3) = 0.16989\dots$
3	**2**	8	5	$2^{10}/3 = 341.33\dots$	$\log(3^2/2^3) = 0.11778\dots$
11	**7**	8	5	$2^{18}/3 = 87381.33\dots$	$\log(3^7/2^{11}) = 0.06566\dots$
19	12	**8**	**5**	$3^{17}/2 \approx 6.4 \times 10^7$	$\log(2^8/3^5) = 0.05211\dots$
19	12	**27**	**17**	$3^{29}/2 \approx 3.4 \times 10^{13}$	$\log(2^{27}/3^{17}) = 0.03856\dots$
19	12	**46**	**29**	$3^{41}/2 \approx 1.8 \times 10^{19}$	$\log(2^{46}/3^{29}) = 0.02501\dots$

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

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

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

$$\frac{p_{2j}}{q_{2j}} \leq \frac{\log 3}{\log 2} \leq \frac{p_{2j+1}}{q_{2j+1}} \tag{A.2}$$

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

$$\frac{\log 3}{\log 2} - \frac{p_{2j}}{q_{2j}} = \frac{1}{q_{2j}q_{2j+1}}. \tag{A.3}$$

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

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

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

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

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.

$\square$

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 6.1. 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

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

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

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

Since

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

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

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

This can be numerically verified for $K' \leq 100$, with substantial room to spare for K' large; see Figure 7. On a block $6a-1 \leq n \leq 6a+4$, the sum is positive:

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

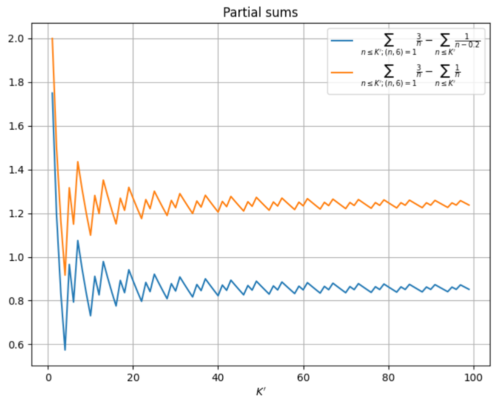

FIGURE 7. A plot of (B.1).

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

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

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

In this section we collect some estimates on sums over primes from the literature that we will use in this paper.

We recall the effective prime number theorem from [5, Corollary 5.2], which asserts that

$$\pi(x) \geq \frac{x}{\log x} + \frac{x}{\log^2 x} \tag{C.1}$$

for $x \geq 599$ and

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

for $x > 1$.

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

$$\sum_{y<n\leq x} b(n) a_n = \int_x^y b(t) f(t)\, dt + O_{\leq}(A\|b\|_{\mathrm{TV}^*(y,x]}), \tag{C.3}$$

where the augmented total variation $\|b\|_{\mathrm{TV}^*(y,x]}$ *is defined as*

$$\|b\|_{\mathrm{TV}^*(y,x]} := |b(y^+)| + |b(x)| + \|b\|_{\mathrm{TV}(y,x]},$$

$b(y^+) := \lim_{t\to y^+} b(t)$ *denotes the right limit of* b *at* y*, and the total variation* $\|b\|_{\mathrm{TV}(y,x]}$ *is defined as the supremum of the quantities* $\sum_{j=0}^{J-1} |b(x_{j+1}) - b(x_j)|$ *for* $y < x_0 \leq \cdots \leq x_J \leq x$*.*

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.3) or increase the total variation of b, while only modifying the integral in (C.3) 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 $g\colon (y, x] \to \mathbb{R}$ 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 $O_{\leq}(A\|b\|_{\mathrm{TV}^*(y,x]})$, and the claim follows. □

By combining this lemma with effective prime number estimates, we obtain

Lemma C.2 (Effective prime number theorem)**.** *Under the above hypotheses with* $1423 \leq y \leq x$*, one has*

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

where

$$E(x) := 0.95\sqrt{x} + \frac{\sqrt{x}}{8\pi} \log x(\log x - 3) 1_{x\geq 10^{19}} + \min(\varepsilon_0, \varepsilon_1(x), \varepsilon_2(x)) x 1_{x\geq e^{45}}$$

and

$$\begin{aligned}\varepsilon_0 &:= 1.11742 \times 10^{-8}\\ \varepsilon_1(x) &:= 9.39(\log^{1.515} x)\exp(-0.8274\sqrt{\log x})\\ \varepsilon_2(x) &:= 0.026(\log^{1.801} x)\exp(-0.1853(\log^{3/5} x)(\log\log x)^{-1/5})\end{aligned}$$

for some absolute constant $c > 0$.

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

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

for all $x \ge 1423$.

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

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

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

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

From [13, Theorems 10,12] we have

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

Since

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

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

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

$$\psi(x) = x + O_{\le}(\varepsilon_0)$$

for $x \ge e^{45} \approx 10^{19.54}$, while from [11, Theorems 1.1, 1.4] one has

$$\psi(x) = x + O_{\le}(\varepsilon_1(x))$$

and

$$\psi(x) = x + O_{\le}(\varepsilon_2(x))$$

for all $x \ge 2$. The claim then follows by repeating the previous arguments. □

Remark C.3. Assuming the Riemann hypothesis, the final term in the definition of $E(x)$ may be deleted, since [3, (7.3)] then holds for all $x \ge 5000$.

Applying the above lemma to $b(t) = 1/\log t$, we conclude in particular that

$$\pi(x) - \pi(y) = \int_y^x (1 - \frac{2}{\sqrt{t}}) \frac{dt}{\log t} + O_{\leq}(\frac{2E(x)}{\log y}) \tag{C.4}$$

for $1423 \leq y \leq x$. In particular, we have the slightly crude upper bound

$$\pi(x) - \pi(y) \leq \frac{x-y}{\log y} + \frac{2E(x)}{\log y} \tag{C.5}$$

in this range, as well as the lower bound

$$\pi(x) - \pi(y) \geq \left(1 - \frac{2}{\sqrt{y}}\right) \frac{x-y}{\log y} - \frac{2E(x)}{\log y} \tag{C.6}$$

The quantity $E(x)$ is somewhat complicated. In our paper we will only need simpler upper bounds on this quantity. Firstly, from the ε_1 term we obtain the classical error term

$$E(x) \ll x \exp(-c\sqrt{\log x}) \tag{C.7}$$

for some absolute constant $c > 0$. Secondly, because

$$\frac{\log x(\log x - 3)}{8\pi\sqrt{x}} \leq 2.244 \times 10^{-8}$$

for $x \geq 10^{19}$, we also have the upper bound

$$E(x) \leq \tilde{E}(x) \tag{C.8}$$

where

$$\tilde{E}(x) := 0.95\sqrt{x} + 2.25 \times 10^{-8} x. \tag{C.9}$$

One small advantage of working with $\tilde{E}$ instead of E is that in addition to being monotone increasing, $\tilde{E}(x)/x$ is monotone decreasing.

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 (1.6). 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:

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

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

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

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

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

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 (D.2) 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

$$c_0 \approx 0.30441901087.$$

It should be possible to obtain this level of precision more rigorously (using interval arithmetic to preclude any possibility of roundoff error), but we have not attempted to do so.

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, *Explicit estimates of some functions over primes*, Ramanujan J. **45** (2018) 227–251.

[6] P. Erdős, *Some problems in number theory*, in Computers in Number Theory, Academic Press, London New York, 1971, pp. 405–414.

[7] 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.

[8] P. Erdős, R. Graham, *Old and new problems and results in combinatorial number theory*, Monographies de L'Enseignement Mathematique 1980.

[9] R. K. Guy, *Unsolved Problems in Number Theory*, 3rd Edition, Springer, 2004.

[10] R. K. Guy, J. L. Selfridge, *Factoring factorial n*, Amer. Math. Monthly **105** (1998) 766–767.

[11] D. Johnston, A. Yang, *Some explicit estimates for the error term in the prime number theorem*, J. Math. Anal. Appl., **527** (2) (2023), Paper No. 127460.

[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`.

???

Email address: ???

UCLA DEPARTMENT OF MATHEMATICS, LOS ANGELES, CA 90095-1555.

Email address: tao@math.ucla.edu