ACADÉMIE
DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus
Mathématique

Ofir Gorodetsky, Jared Duker Lichtman and Mo Dick Wong

On Erdős sums of almost primes

Volume 362 (2024), p. 1571-1596

Online since: 25 November 2024

https://doi.org/10.5802/crmath.650

CENTRE MERSENNE

The Comptes Rendus. Mathématique *are a member of the*
Mersenne Center for open scientific publishing
www.centre-mersenne.org — e-ISSN : 1778-3569

ACADÉMIE DES SCIENCES
INSTITUT DE FRANCE

Comptes Rendus. Mathématique
2024, Vol. 362, p. 1571-1596
https://doi.org/10.5802/crmath.650

Research article / *Article de recherche*
Number theory / *Théorie des nombres*

On Erdős sums of almost primes

Sur les sommes d'Erdős de presque premiers

Ofir Gorodetsky [a, b], **Jared Duker Lichtman** [*, c] **and Mo Dick Wong** [d]

[a] Mathematical Institute, University of Oxford, Oxford, OX2 6GG, UK

[b] Department of Mathematics, Technion – Israel Institute of Technology, Haifa 3200003, Israel

[c] Department of Mathematics, Stanford University, Stanford, CA, USA

[d] Department of Mathematical Sciences, Durham University, Stockton Road, Durham DH1 3LE, UK

E-mails: ofir.gor@technion.ac.il, jared.d.lichtman@gmail.com, mo-dick.wong@durham.ac.uk

Abstract. In 1935, Erdős proved that the sums $f_k = \sum_n 1/(n\log n)$, over integers n with exactly k prime factors, are bounded by an absolute constant, and in 1993 Zhang proved that f_k is maximized by the prime sum $f_1 = \sum_p 1/(p\log p)$. According to a 2013 conjecture of Banks and Martin, the sums f_k are predicted to decrease monotonically in k. In this article, we show that the sums restricted to odd integers are indeed monotonically decreasing in k, sufficiently large. By contrast, contrary to the conjecture we prove that the sums f_k increase monotonically in k, sufficiently large.

Our main result gives an asymptotic for f_k which identifies the (negative) secondary term, namely $f_k = 1-(a+o(1))k^2/2^k$ for an explicit constant $a = 0.0656\cdots$. This is proven by a refined method combining real and complex analysis, whereas the classical results of Sathe and Selberg on products of k primes imply the weaker estimate $f_k = 1+O_\varepsilon(k^{\varepsilon-1/2})$. We also give an alternate, probability-theoretic argument related to the Dickman distribution. Here the proof reduces to showing a sequence of integrals converges exponentially quickly $\mathrm{e}^{-\gamma}$, which may be of independent interest.

Résumé. En 1935, Erdős a prouvé que les sommes $f_k = \sum_n 1/(n\log n)$, portant sur les entiers n ayant exactement k facteurs premiers, sont majorées par une constante absolue, et en 1993, Zhang a prouvé que f_k est maximisé par la somme sur les nombres premiers $f_1 = \sum_p 1/(p\log p)$. Selon une conjecture de Banks et Martin de 2013, les sommes f_k devraient être décroissantes en fonction de k. Dans cet article, nous démontrons que les sommes restreintes aux entiers impairs sont bien décroissantes pour k suffisamment grand. En revanche, contrairement à la conjecture, nous prouvons que les sommes f_k sont croissantes en fonction de k, suffisamment grand. Notre résultat principal donne une formule asymptotique pour f_k qui identifie le terme secondaire (négatif), à savoir $f_k = 1-(a+o(1))k^2/2^k$ pour une constante explicite $a = 0{,}0656\cdots$. Ceci est prouvé par une méthode raffinée combinant analyse réelle et complexe, alors que les résultats classiques de Sathe et Selberg sur les produits de k nombres premiers impliquent l'estimation plus faible $f_k = 1+O_\varepsilon(k^{\varepsilon-1/2})$. De plus, nous donnons un argument probabiliste alternatif, lié à la distribution de Dickman. Ici, la preuve se réduit à démontrer qu'une suite d'intégrales converge exponentiellement rapidement vers $\mathrm{e}^{-\gamma}$, ce qui peut présenter un intérêt indépendant.

Keywords. Almost primes, primitive set, Dickman distribution, recursive distributional equation.

Mots-clés. Nombres presque premiers, ensemble primitif, loi de Dickman, équation en loi récursive.

2020 Mathematics Subject Classification. 11N25, 11Y60, 11A05, 60G18, 60H25.

Funding. The first author is supported by funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 851318). The second author is supported by a Clarendon Scholarship at the University of Oxford. The second and third author acknowledge support by travel grants from the Heilbronn Institute for Mathematical Research (HIMR).

Manuscript received 14 March 2023, revised 27 April 2024, accepted 12 May 2024.

* Corresponding author

1. Introduction

Let $\Omega(n)$ denote the number of prime factors of n, counted with repetition. If $\Omega(n) = k$, n is called a k-almost prime. For each $k \geq 1$ denote the series

$$f_k = \sum_{\Omega(n)=k} \frac{1}{n \log n}.$$

Here f_k is the *Erdős sum* of k-almost primes, also called the k^{th} *fingerprint number*. In 1935 Erdős [6] showed that $f_k = O(1)$ is bounded[1] and in 1993 Zhang [16] proved the primes have maximal Erdős sum, that is, $f_k \leq f_1$ holds for all k. This had given initial evidence towards the Erdős primitive set conjecture, now recently proven by the second author [13].

In 2013, Banks and Martin [1] posed a vast generalization of the Erdős primitive set conjecture (see [1, 13] for details and further discussion). In particular, they conjectured that the sums f_k decrease monotonically in k. Denote the sum $f_{k,y}$ restricting f_k to integers without prime factors $\leq y$, that is,

$$f_{k,y} = \sum_{\substack{\Omega(n)=k \\ p|n \to p>y}} \frac{1}{n \log n}.$$

Banks and Martin further conjectured that for any fixed $y \geq 1$, $f_{k,y}$ decrease monotonically.

We prove that their conjecture holds for $y \geq 2$ and k sufficiently large. By contrast, we prove that $f_k = f_{k,1}$ increases monotonically in k, sufficiently large.

Theorem 1. *Let $y \geq 2$. For k sufficiently large, we have $f_{k-1} < f_k$ and $f_{k-1,y} > f_{k,y}$.*

When $y \geq 2$, we believe $f_{k-1,y} > f_{k,y}$ should hold for all $k > 1$, in accordance with Banks–Martin [1]. When $y = 1$, we believe $f_{k-1} < f_k$ holds for all $k > 6$. These inequalities have been verified numerically up to $k \leq 20$ [10].

The classical Sathe–Selberg theorem gives asymptotics for the counting function of k-almost primes, and implies f_k converges to 1 with square-root error. That is $f_k = 1 + O_\varepsilon(k^{\varepsilon-1/2})$, see [10, Theorem 4.1]. We give an exponential refinement of this result, which identifies the (negative) secondary term to be $-a_1 k^2/2^k$ for an explicit constant $a_1 = 0.0656\cdots$.

Theorem 2. *For all $k \geq 1$ we have*

$$f_k = 1 - 2^{-k}\big(a_1 k^2 + O\big(k \log(k+1)\big)\big),$$

where $a_1 = (d \log 2)/4$ and

$$d := \frac{1}{4} \prod_{p>2} \left(1 - \frac{2}{p}\right)^{-1} \left(1 - \frac{1}{p}\right)^2 = 0.37869\cdots. \tag{1}$$

To motivate the proof of Theorem 2, we first handle the sifted Erdős sums $f_{k,y}$, whose (positive) secondary term is of order $O_y(2^{-k})$ when $y \geq 2$, and so converge more rapidly.

Theorem 3. *Let $y \geq 2$. We have*

$$f_{k,y} = \prod_{p \leq y} \left(1 - \frac{1}{p}\right) + a_y/2^k + O_y(k^3/3^k)$$

[1] Indeed, his result bounded Erdős sums $f(A) = \sum_{n \in A} 1/(n \log n)$ uniformly over any primitive set A.

uniformly for $k \geq 1$, *where* $a_y = c_y d_y$ *for*

$$c_y = \gamma + \sum_{p\leq y} \frac{\log p}{p-1} - \sum_{p>y} \frac{\log p}{(p-1)(p-2)}, \tag{2}$$

$$d_y = d \prod_{2<p\leq y} \left(1 - \frac{2}{p}\right), \tag{3}$$

for d *as in* (1), *and* $\gamma = 0.5772\cdots$ *is the Euler–Mascheroni constant.*

As usual $O_y(\cdots)$ means the implicit constant may depend on $y \geq 2$, but not on k.

In particular, from Theorems 2 and 3 we infer the even and odd terms in f_k contribute $\frac{1}{2} - (a_1 + o(1))k^2/2^k$ and $\frac{1}{2} + (a_2 + o(1))2^{-k}$, respectively. As will be shown in the proofs, this discrepancy in the size (and sign) of the secondary terms ultimately come from the different behavior of

$$\lim_{s\to 1^+} (s-1)^{-z} \sum_{p|n\to p>y} \frac{z^{\Omega(n)}}{n^s}$$

when $y = 1$ and $y \geq 2$. Namely, the singularity closest to 0 when $y = 1$ is at $z = 2$ while for $y \geq 2$ it is more distant. This provides a clean answer to the hitherto unexplained numerical observations up to $k \leq 20$.

Remark 4. Our methods can also handle $\sum_{\omega(n)=k,\, p|n\to p>y} 1/(n\log n)$, where $\omega(n) = \sum_{p|n} 1$. For this problem, the analysis does not have a discrepancy between $y = 1$ and $y \geq 2$.

Our methods should also handle the Dirichlet series $\sum_{\Omega(n)=k} n^{-t}$ for $t > 1$. For these, striking work of Banks–Martin [1] shows that $k = 1$ is maximal if and only if $t > \tau$, where $\tau = 1.14\cdots$ is the unique solution to a certain functional equation involving the Riemann zeta function. When $t < \tau$ it is not understood which $k = k_t$ is maximal, also see [4, 5]. Finally, one may also consider translated sums $\sum_{\Omega(n)=k} 1/(n(\log n + h))$ for $h \in \mathbb{R}$, where the author [12] proved that $k = 1$ is not maximal for $h > 1.04\cdots$. In fact, $k = 1$ is minimal if and only if $h > 0.803\cdots$.

Remark 5. Hankel contours over the complex plane are used in the Selberg–Delange method, concerning asymptotics for the sums $\sum_{n\leq x} z^{\Omega(n)}$, z complex. Saddle point analysis is used in the Sathe–Selberg theorem, which extracts information on the counting function of k-almost primes from these sums. These devices are powerful but lead to poor savings. By contrast, our proofs contain neither saddle point analysis nor Hankel contours.

Moreover, the Selberg–Delange method assumes a zero-free region for $\zeta(s)$, as well as a bound on $\log\zeta(s)$ to the left of $\mathrm{Re}(s) = 1$. By contrast, our proofs make no complex-analytic assumptions about $\zeta(s)$ whatsoever. We only use the (very basic) Taylor expansion $\zeta(t)(t-1) = 1 + \gamma(t-1) + O((t-1)^2)$ for real $t \in (1,2)$.

1.1. *A sequence of integrals*

We also provide an alternate, probability-theoretic argument, which gives (weaker) exponential error $f_k = 1 + O(k/2^{k/4})$, but which has potentially much wider applicability to other primitive sets beyond k-almost primes. We leave further development of this perspective to future work.

The rough strategy is to first show an asymptotic relation

$$f_k \sim \mathrm{e}^{\gamma} I_{\lfloor k/4\rfloor} \tag{4}$$

for a certain sequence of iterated integrals I_k. Here $\gamma = 0.5772\cdots$ is the Euler–Mascheroni constant. Specifically, let $I_0 = 1$ and for $k \geq 1$ let

$$I_k = \int_{[0,1]^k} \frac{\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_k}{1 + x_1(1 + x_2(\cdots(1 + x_k)\cdots))}. \tag{5}$$

Let $P_1(n)$ be the largest prime factor of n. To connect f_k to I_k, we use Mertens' theorem to show that f_k is close to e^{γ} times a *weighted average* of $\log P_1(n)/\log n$ over integers n with $\Omega(n) = k$, see (67). Then for each $n = p_1 \cdots p_k$ $(p_1 \ge p_2 \ge \cdots)$, the reciprocal of this ratio may be expressed as

$$\frac{\log n}{\log P_1(n)} = 1 + \frac{\log p_2}{\log p_1}\left(1 + \cdots \left(1 + \frac{\log p_{j+1}}{\log p_j}\left(1 + \frac{\log\big(p_{j+2}\cdots p_k\big)}{\log p_{j+1}}\right)\right)\cdots\right).$$

Further, we show the consecutive ratios $\frac{\log p_{i+1}}{\log p_i} \in [0,1]$ are independent and uniformly distributed in $[0,1]$ (with respect to a certain probability measure) which ultimately leads to (4). See (67) and (78)–(79) for the precise quantitative formulation of (4).

Finally, we prove that I_k converges exponentially quickly to $\mathrm{e}^{-\gamma}$, which may be of independent interest. The qualitative convergence $I_k \to \mathrm{e}^{-\gamma}$ may be deduced from work of Chamayou [3] (cf. [14, Proposition 2.1]).

Theorem 6. *We have* $I_k = \mathrm{e}^{-\gamma} + O(2^{-k})$.

Our approach to Theorem 6 is self-contained and relies on a probabilistic reformulation of the problem, which turns out to be related to the Dickman–Goncharov distribution. See the extensive survey of Molchanov and Panov [14] for background on this distribution, which appears across probability and theoretical computer science [8].

1.2. *Proof of monotonicity results*

Here we quickly deduce Theorem 1 assuming Theorems 2 and 3: Indeed, for large k we have

$$\begin{aligned} f_k - f_{k-1} &= \frac{\log 2}{4} d\big((k-1)^2/2^{k-1} - k^2/2^k\big) \; + \; o(k^2/2^k) \\ &= \frac{\log 2}{4} d k^2/2^k + o(k^2/2^k) \; > \; 0 \end{aligned}$$

by Theorem 2. Similarly, by Theorem 3 we have

$$\begin{aligned} f_{k-1,y} - f_{k,y} &= c_y d_y (1/2^{k-1} - 1/2^k) \; + \; o(1/2^k) \\ &= c_y d_y/2^k + o(1/2^k) \; > \; 0. \end{aligned}$$

Here we used $c_y > 0$ for $y \ge 2$. Indeed, c_y from (2) is clearly increasing in $y \ge 2$, so $c_y \ge c_2 > 0$ since

$$\sum_{p>2} \frac{\log p}{(p-1)(p-2)} < .9 < 1.2 < \log 2 + \gamma.$$

This completes the proof of Theorem 1.

1.3. *Proof method: permutations*

As a short illustration of the proof method, we consider a permutation analogue of the sums f_k. This result also may be of independent interest. Namely, define $f_{k,\pi}$ by

$$f_{k,\pi} := \sum_{m \ge 1} \frac{a_{m,k}}{m}$$

where $a_{m,k}$ is the probability that a permutation, chosen uniformly at random from S_m, has exactly k disjoint cycles.

Remark 7. To see the analogy with the original sum, given a permutation $\pi \in S_m$ let $C(\pi)$ be the number of disjoint cycles in π, $d(\pi)$ be m and $|\pi|$ be $|S_m| = m!$. The sum $f_{k,\pi}$ may be expressed as

$$f_{k,\pi} = \sum_{\substack{C(\pi)=k \\ \pi \in \cup_{m\geq 1} S_m}} \frac{1}{|\pi| d(\pi)}.$$

The weight $1/(|\pi|d(\pi))$ is analogous to $1/(n \log n)$.

Recall the Riemann zeta function $\zeta(s) = \sum_{n\geq 1} n^{-s}$.

Proposition 8. *For any $k \geq 1$ we have*

$$f_{k,\pi} = \zeta(k+1) = 1 + 2^{-k-1} + O(3^{-k}).$$

Proof. Consider the exponential generating function of $\sum_{\pi \in S_n} z^{C(\pi)}$:

$$F(u,z) := \sum_{m\geq 0} \frac{1}{m!} \left(\sum_{\pi \in S_m} z^{C(\pi)} \right) u^m = 1 + \sum_{k\geq 1} z^k \sum_{m\geq 1} a_{m,k} u^m.$$

The exponential formula for permutations shows

$$F(u,z) = \exp\left(\sum_{n\geq 1} z \frac{u^n}{n} \right) = \exp\big(-z \log(1-u)\big). \tag{6}$$

The Taylor series for $\exp(-z\log(1-u))$ in z is

$$\exp\big(-z\log(1-u)\big) = \sum_{k\geq 0} \frac{z^k}{k!} \big(-\log(1-u)\big)^k,$$

so extracting the coefficient of z^k in $F(u,z)$ yields, by (6),

$$\sum_{m\geq 1} a_{m,k} u^m = \frac{\big(-\log(1-u)\big)^k}{k!}. \tag{7}$$

Since $1/m = \int_0^1 u^{m-1}\,\mathrm{d}u$, we can integrate (7) over $u \in [0,1]$ to obtain

$$f_{k,\pi} = \sum_{m\geq 1} \frac{a_{m,k}}{m} = \int_0^1 \frac{\big(-\log(1-u)\big)^k}{k!} \frac{\mathrm{d}u}{u} = \int_0^\infty \frac{x^k}{\Gamma(k+1)} \frac{\mathrm{d}x}{\mathrm{e}^x - 1} = \zeta(k+1).$$

This completes the proof. Here in the third equality we performed the change of variables $u = 1 - \mathrm{e}^{-x}$, and in the fourth we applied Riemann's famous integral representation for $\zeta(s)$, with $s = k+1$,

$$\Gamma(s)\zeta(s) = \int_0^\infty \frac{x^{s-1}}{\mathrm{e}^x - 1}\,\mathrm{d}x. \qquad \square$$

Remark 9. In the same way that we shall relate f_k to the sequence of integrals I_k in Theorem 6, one can relate $f_{k,\pi}$ to I_k as well.

2. Sifted sums

2.1. *Preparation*

We now collect a few analytic properties of the generating functions that will play central roles in the proof of Theorem 3. Let us start with the function

$$F(s,z) := \sum_{n\geq 1} \frac{z^{\Omega(n)}}{n^s} = \prod_p \left(1 - \frac{z}{p^s}\right)^{-1}, \qquad s \in \mathbb{R}. \tag{8}$$

Lemma 10. *The function $F(s,z)$ converges absolutely for $s>1$ and $|z|<2$. In this domain F has a power series representation in z and it defines a smooth function in s.*[2]

Proof. Observe that for $s \geq 1+\varepsilon$ and $|z| \leq 2^{-\varepsilon}$ the pth factor in (8) is

$$\left(1-\frac{z}{p^s}\right)^{-1} = \exp\left(\sum_{i\geq 1}\frac{z^i}{i p^{si}}\right) = \exp\left(O_\varepsilon(|z|p^{-s})\right)$$

and so the product in (8) is a uniform limit of power series in z which are smooth in s. □

For $s>1$ and $|z|<2$ we define

$$G(s,z) := F(s,z)(s-1)^z.$$

The following lemma extends the range of definition of G. For any $y \geq 1$ denote by y_1 the smallest prime greater than y.

Lemma 11. *For every prime q, $\prod_{p\leq q}(1-z/p^s)G(s,z)$, as well as its derivatives in s, are smooth in $s\geq 1$ and have a power series expansion in z with radius q_1. Moreover, for $s \geq 1$ and $k \geq 0$, the kth derivative of $G(s,z)$ with respect to s has a meromorphic continuation to $z \in \mathbb{C}$ with poles of order k at $z=p^s$ for every prime p. This continuation satisfies*

$$G(1,z) = \prod_p \left(1-\frac{z}{p}\right)^{-1}\left(1-\frac{1}{p}\right)^z.$$

Proof. We write G as

$$G(s,z) = F(s,z)\zeta(s)^{-z}(\zeta(s)(s-1))^z. \tag{9}$$

It is well known that $\lim_{s\to 1^+}\zeta(s)(s-1)=1$, and that extending $\zeta(s)(s-1)$ to $s=1$, by setting it to equal 1 there, it is a smooth function in $s\geq 1$.[3] Hence, $(\zeta(s)(s-1))^z = \exp(z\log(\zeta(s)(s-1))$ is smooth in $s \geq 1$ and has a power series representation in z with infinite radius of convergence:

$$(\zeta(s)(s-1))^z = \sum_{i\geq 0}\frac{z^i}{i!}(\log(\zeta(s)(s-1)))^i. \tag{10}$$

At $s=1$, (10) is equal to 1. It remains to consider $F(s,z)\zeta(s)^{-z}$. It has the following Euler product:

$$F(s,z)\zeta(s)^{-z} = \prod_p\left(1-\frac{z}{p^s}\right)^{-1}\left(1-\frac{1}{p^s}\right)^z. \tag{11}$$

For every prime q, we can use the product in (11) to define $F(s,z)\zeta(s)^{-z}$ as a product of the rational function $\prod_{p\leq q}(1-z/p^s)^{-1}$ (which has simple poles in z at $z=p^s$ for every prime $p\leq q$, and is defined for $s\geq 1$) with a function that has a power series representation in z with radius of convergence q_1 and is defined for $s\geq 1$. This is because if $|z|<q_1^{1-\varepsilon}$ and $p>q$ then the pth term in the right-hand side of (11) equals

$$\exp\left(\sum_{i=2}^{\infty}\frac{z^i - z}{i p^{si}}\right) = \exp\left(O_\varepsilon\left((|z|^2+|z|)p^{-2s}\right)\right),$$

and the product over $p>q$ converges absolutely and uniformly. □

[2] In this paper we will only need first and second derivatives with respect to s, which is always seen as a real-valued variable.

[3] Throughout we identify $\zeta(s)(s-1)$ at $s=1$ with 1. A function f is smooth on $[1,\infty)$ if it belongs to $\bigcap_{k=0}^{\infty} C^k([1,\infty))$ where for $k\geq 1$, $C^k([1,\infty)) := \{f\colon [1,\infty)\to\mathbb{R}, f \text{ is differentiable}, f' \in C^{k-1}([1,\infty))\}$ and differentiability at 1 is defined via the right derivative. For $k=0$, $C^0([1,\infty))$ consists of continuous functions on $[1,\infty)$, with continuity at 1 defined by right-continuity.

We also introduce

$$F_y(s,z) := \sum_{\substack{n\geq 1\\ p|n\to p>y}} \frac{z^{\Omega(n)}}{n^s} = \prod_{p>y}\left(1-\frac{z}{p^s}\right)^{-1} = \prod_{p\leq y}\left(1-\frac{z}{p^s}\right)F(s,z), \tag{12}$$

and

$$\begin{aligned} G_y(s,z) &:= \prod_{p\leq y}\left(1-\frac{z}{p^s}\right)G(s,z) \\ &= \prod_{p\leq y}\left(1-\frac{1}{p^s}\right)^z \prod_{p>y}\left(1-\frac{z}{p^s}\right)^{-1}\left(1-\frac{1}{p^s}\right)^z (\zeta(s)(s-1))^z. \end{aligned} \tag{13}$$

For any smooth function $H(s,z)$, we shall denote by $H^{(a,b)}$ the mixed partial derivative

$$H^{(a,b)} = \frac{\partial^{a+b}}{\partial s^a \partial z^b} H.$$

For every $j\geq 0$ and fixed $s\geq 1$, $G_y^{(j,0)}(s,z)$ has a meromorphic continuation to $\mathbb{C}$ with poles at $z=p^s$ for every prime $p>y$ by Lemma 10. Taking the logarithmic derivative of (13) with respect to s gives

$$\begin{aligned} \frac{G_y^{(1,0)}}{G_y}(s,z) &= \frac{\partial}{\partial s}\left[\log G_y(s,z)\right] \\ &= \frac{\partial}{\partial s}\left[\sum_{p\leq y} z\log\left(1-\frac{1}{p^s}\right) + \sum_{p\geq y_1}\left(z\log\left(1-\frac{1}{p^s}\right)-\log\left(1-\frac{z}{p^s}\right)\right)+z\log\big(\zeta(s)(s-1)\big)\right] \end{aligned}$$

so that

Lemma 12. *We have*

$$G_y^{(1,0)}(s,z) = zG_y(s,z)\left(\sum_{p\leq y}\frac{\log p}{p^s-1} - \sum_{p\geq y_1}\frac{(z-1)\log p}{(p^s-1)(p^s-z)} + (\log(\zeta(s)(s-1)))'\right). \tag{14}$$

Remark 13. These generating functions and their values have natural connections to related questions about k-almost primes [2, 11]. In particular $G_2(1,2) = \frac{1}{4}\prod_{p>2}(1-\frac{2}{p})^{-1}(1-\frac{1}{p})^2 = d$ as in (1) equals $2\beta_2$, from the main term of [2, Theorem 1.2].

2.2. *Proof of Theorem 3*

Let $y\geq 1$, and let y_1 be the smallest prime greater than y. We have the integral representation

$$f_{k,y} = \sum_{\substack{\Omega(n)=k\\ p|n\to p>y}} \frac{1}{n\log n} = \int_1^\infty \sum_{\substack{\Omega(n)=k\\ p|n\to p>y}} n^{-s}\,\mathrm{d}s. \tag{15}$$

The smallest number n with $\Omega(n)=k$ and all prime factors greater than y is y_1^k, so the contribution of $s\geq 2$ to (15) is

$$\int_2^\infty \sum_{\substack{\Omega(n)=k\\ p|n\to p>y}} n^{-s}\,\mathrm{d}s = \sum_{\substack{\Omega(n)=k\\ p|n\to p>y}} \frac{1}{n^2\log n} \leq \sum_{n\geq y_1^k}\frac{1}{n^2\log n} \ll \frac{1}{k}\int_{y_1^k-1}^\infty \frac{\mathrm{d}t}{t^2} \ll \frac{1}{ky_1^k}.$$

Thus we have

$$f_{k,y} = I_{k,y} + O\left(\frac{1}{ky_1^k}\right), \tag{16}$$

where $I_{k,y}$ is the corresponding integral over $s\in[1,2]$, namely,

$$I_{k,y} := \int_1^2 \sum_{\substack{\Omega(n)=k\\ p|n\to p>y}} n^{-s}\mathrm{d}s = \int_1^2 \frac{1}{k!}F_y^{(0,k)}(s,0)\mathrm{d}s. \tag{17}$$

Here we term-wise differentiated $F_y(s,z)$ in (12) with respect to z, for $s>1$.

Next from the Taylor series

$$(s-1)^{-z} = \exp\big(-z\log(s-1)\big) = \sum_{i\ge 0} \frac{\big(-\log(s-1)\big)^i}{i!} z^i, \tag{18}$$

we apply the product rule to $F_y(s,z) = (s-1)^{-z} G_y(s,z)$, giving

$$\frac{1}{k!} F_y^{(0,k)}(s,0) = \sum_{i=0}^{k} \frac{\big(-\log(s-1)\big)^{k-i}}{(k-i)!} \frac{1}{i!} G_y^{(0,i)}(s,0).$$

Thus (17) becomes

$$I_{k,y} = \sum_{i=0}^{k} \int_1^2 \frac{\big(-\log(s-1)\big)^{k-i}}{(k-i)!} \frac{1}{i!} G_y^{(0,i)}(s,0)\,\mathrm{d}s. \tag{19}$$

Now we introduce a similar integral $I'_{k,y}$, given by evaluating $G^{(0,i)}(s,0)$ in the integrand at $s=1$, namely,

$$I'_{k,y} := \sum_{i=0}^{k} \frac{1}{i!} G_y^{(0,i)}(1,0) \int_1^2 \frac{\big(-\log(s-1)\big)^{k-i}}{(k-i)!}\,\mathrm{d}s. \tag{20}$$

To handle $I'_{k,y}$, we substitute $s = 1+\mathrm{e}^{-t}$ and obtain, for any $j\ge 0$,

$$\int_1^2 \frac{\big(-\log(s-1)\big)^j}{j!}\,\mathrm{d}s = \int_0^\infty \frac{\mathrm{e}^{-t}\,t^j}{j!}\,\mathrm{d}t = \frac{\Gamma(j+1)}{j!} = 1, \tag{21}$$

as we are evaluating the Gamma function at $j+1$. Hence (20) simplifies as

$$I'_{k,y} = \sum_{i=0}^{k} \frac{1}{i!} G_y^{(0,i)}(1,0). \tag{22}$$

In the upcoming subsections, we shall estimate $I_{k,y}$ by means of the following lemmas for $I'_{k,y}$ and $I_{k,y} - I'_{k,y}$.

Lemma 14. *Let $y\ge 1$. We have $I'_{k,y} = G_y(1,1) + O_y(y_1^{-k})$.*

Lemma 15. *Let $y\ge 2$. We have $I_{k,y} = I'_{k,y} + G_y^{(1,0)}(1,2)/2^{k+1} + O_y(k^3/3^k)$.*

Proof of Theorem 3 assuming Lemmas 14 and 15. Recalling (16), we have

$$\begin{aligned} f_{k,y} &= I_{k,y} + O(1/(k y_1^k)) \\ &= I'_{k,y} + G_y^{(1,0)}(1,2)/2^{k+1} + O_y(k^3/3^k) \\ &= G_y(1,1) + G_y^{(1,0)}(1,2)/2^{k+1} + O_y(k^3/3^k) \end{aligned} \tag{23}$$

for $y\ge 2$. To compute the constants above, we first note $G_y(1,1) = \prod_{p\le y}(1-1/p)$. Next, by (13) and (14) with $(s,z) = (1,2)$ we have

$$\begin{aligned} G_y(1,2) &= \prod_{p\le y}\left(1-\frac{1}{p}\right)^2 \prod_{p\ge y_1}\left(1-\frac{2}{p}\right)^{-1}\left(1-\frac{1}{p}\right)^2 = d_y, \\ G_y^{(1,0)}(1,2) &= 2G_y(1,2)\left(\sum_{p\le y}\frac{\log p}{p-1} - \sum_{p\ge y_1}\frac{\log p}{(p-1)(p-2)} + \gamma\right) = 2d_y c_y \end{aligned} \tag{24}$$

for $y\ge 2$. Here we used $(\log(\zeta(s)(s-1)))'|_{s=1} = \gamma$. Hence plugging (24) and $G_y(1,1) = \prod_{p\le y}(1-1/p)$ back into (23), we conclude

$$f_{k,y} = \prod_{p\le y}\left(1-\frac{1}{p}\right) + c_y d_y/2^k + O_y(k^3/3^k). \tag{25}$$

□

2.3. *Proof of Lemma 14*

We use the notation $[z^n]A(z) = \frac{1}{n!}\big(\frac{\mathrm{d}^n}{\mathrm{d}z^n}A\big)(0)$ to denote the coefficient of z^n in A, where A is a function with Taylor series representation at $z=0$. From the representation of $I'_{k,y}$ in (22), we have for all $k \ge 1$,

$$I'_{k,y} = [z^k]\frac{G_y(1,z)}{1-z} = \frac{1}{2\pi i}\int_{|z|=1/2}\frac{G_y(1,z)}{1-z}\frac{\mathrm{d}z}{z^{k+1}},$$

using Cauchy's integral formula. Here the integral ranges over a circle centered around $z=0$, oriented counterclockwise, with radius $1/2$.

The function $G_y(1,z)$ has simple poles at $z=p$ for every prime $p>y$; these are its only poles. The rational function $1/(1-z)$ has a simple pole at $z=1$. So recalling the smallest prime $y_1 > y$, the only poles of $G_y(1,z)/(1-z)$ in the range $1/2 < |z| < y_1 + 1/2 =: R$ occur at $z=1$ and $z=y_1$. Thus by Cauchy's residue theorem,

$$I'_{k,y} = \frac{1}{2\pi i}\int_{|z|=R}\frac{G_y(1,z)}{1-z}\frac{\mathrm{d}z}{z^{k+1}} + G_y(1,1) - \frac{\lim_{z\to y_1}(z-y_1)\,G_y(1,z)}{(1-y_1)y_1^{k+1}}. \tag{26}$$

Note that $\lim_{z\to y_1} G_y(1,z)(z-y_1) \ll_y 1$. Then we claim $|G_y(1,z)| \ll_y 1$ in the integrand of (26), from which we conclude

$$I'_{k,y} = \int_{|z|=R}\frac{O_y(1)}{R-1}\frac{\mathrm{d}z}{R^{k+1}} + G_y(1,1) - \frac{O_y(1)}{(1-y_1)y_1^{k+1}} \;=\; G_y(1,1) + O_y(y_1^{-k}). \tag{27}$$

To show this claim, note that if $|z|=R$ and $p>2R$, then

$$\left|\left(1-\frac{1}{p}\right)^z\left(1-\frac{z}{p}\right)^{-1}\right| = \left|\exp\left(\sum_{i\ge 2}\frac{z^i - z}{ip^i}\right)\right| \le \exp\left(\sum_{i\ge 2}\frac{|z|^i}{p^i}\right) \le \exp\left(\frac{2R^2}{p^2}\right).$$

Hence, recalling (13) with $s=1$ we obtain

$$\begin{aligned}\max_{|z|=R}\left|G_y(1,z)\right| &= \max_{|z|=R}\left|\prod_{p\le y}\left(1-\frac{1}{p}\right)^z\prod_{p\ge y_1}\left(1-\frac{1}{p}\right)^z\left(1-\frac{z}{p}\right)^{-1}\right|\\ &\le \prod_{p\le y}\left(1-\frac{1}{p}\right)^{-R}\prod_{p>2R}\exp\left(\frac{2R^2}{p^2}\right)\cdot\max_{|z|=R}\prod_{y_1\le p\le 2R}\left|\left(1-\frac{1}{p}\right)^z\left(1-\frac{z}{p}\right)^{-1}\right| \ll_y 1.\end{aligned}$$

This proves the claim, and hence Lemma 14 follows.

2.4. *Proof of Lemma 15*

By Taylor expansion at $s=1$, we have uniformly for $s\in[1,2]$

$$G_y^{(0,i)}(s,0) = G_y^{(0,i)}(1,0) + (s-1)b_i + O\big((s-1)^2 c_i\big),$$

for coefficients

$$b_i := G_y^{(1,i)}(1,0) \qquad \text{and} \qquad c_i := \max_{s'\in[1,2]}\left|G_y^{(2,i)}(s',0)\right|. \tag{28}$$

Thus subtracting (19) from (20), we have

$$\begin{aligned}I_{k,y} - I'_{k,y} &= \sum_{i=0}^{k}\int_1^2\frac{\big(-\log(s-1)\big)^{k-i}}{(k-i)!}\frac{1}{i!}\Big(G_y^{(0,i)}(s,0) - G_y^{(0,i)}(1,0)\Big)\,\mathrm{d}s\\ &= \sum_{i=0}^{k}\int_0^1\frac{\big(-\log s\big)^{k-i}}{(k-i)!\,i!}\big(sb_i + O(s^2c_i)\big)\,\mathrm{d}s.\end{aligned} \tag{29}$$

Substituting $s = \mathrm{e}^{-t}$ shows that

$$\begin{aligned}\sum_{i=0}^{k} \frac{b_i}{i!(k-i)!}\int_0^1 (-\log s)^{k-i} s\,\mathrm{d}s &= \sum_{i=0}^{k} \frac{b_i}{i!(k-i)!}\int_0^\infty t^{k-i}\mathrm{e}^{-2t}\,\mathrm{d}t \\ &= \sum_{i=0}^{k} \frac{b_i 2^{i-k-1}}{i!(k-i)!}\int_0^\infty v^{k-i}\mathrm{e}^{-v}\,\mathrm{d}v = \sum_{i=0}^{k}\frac{b_i}{i!}2^{i-k-1}\end{aligned} \tag{30}$$

and similarly

$$\sum_{i=0}^{k} \frac{c_i}{i!(k-i)!}\int_0^1 (-\log s)^{k-i} s^2\,\mathrm{d}s = \sum_{i=0}^{k}\frac{c_i}{i!}3^{i-k-1}. \tag{31}$$

Plugging (30) and (31) back into (29) gives

$$I_{k,y} - I'_{k,y} = \sum_{i=0}^{k}\left(\frac{b_i}{i!}2^{i-k-1} + O\Big(\frac{c_i}{i!}3^{i-k-1}\Big)\right). \tag{32}$$

So proceeding as in the proof of Lemma 14 (as in (26)), by Cauchy's integral formula and residue theorem,

$$\begin{aligned}\sum_{i=0}^{k}\frac{b_i}{i!}2^{i-k-1} &= 2^{-1}[z^k]\frac{G_y^{(1,0)}(1,z)}{1-\frac{z}{2}} = \frac{1}{4\pi i}\int_{|z|=1/2}\frac{G_y^{(1,0)}(1,z)}{1-\frac{z}{2}}\frac{\mathrm{d}z}{z^{k+1}} \\ &= \frac{1}{2\pi i}\int_{|z|=R}\frac{G_y^{(1,0)}(1,z)}{2-z}\frac{\mathrm{d}z}{z^{k+1}} + G_y^{(1,0)}(1,2)2^{-k-1} - \frac{\lim_{z\to y_1}(z-y_1)\,G_y^{(1,0)}(1,z)}{(2-y_1)y_1^{k+1}} \\ &= G_y^{(1,0)}(1,2)2^{-k-1} + O_y(y_1^{-k})\end{aligned} \tag{33}$$

holds where $R := y_1 + 1/2$.

Finally, we claim $c_i \ll_y i!(i+1)^2/y_1^i$, in which case

$$\sum_{i=0}^{k}\frac{c_i}{i!}3^{i-k-1} \ll \sum_{i=0}^{k} y_1^{-i}(i+1)^2 3^{i-k-1} \ll \sum_{i=0}^{k}(i+1)^2 3^{-k} \ll_y k^3/3^k$$

since $y_1 \geq 3$ (as $y \geq 2$). Thus combined with (33), we conclude

$$I_{k,y} - I'_{k,y} = G_y^{(1,0)}(1,2)/2^{k+1} + O_y(k^3/3^k). \tag{34}$$

Hence to complete the proof of Lemma 15, it suffices to show $c_i \ll_y i!(i+1)^2/y_1^i$, which by definition means that uniformly for $s \in [1,2]$,

$$[z^i]G_y^{(2,0)}(s,z) \ll_y (i+1)^2/y_1^i. \tag{35}$$

To this end, recall $G_y^{(1,0)}(s,z) = zG_y(s,z)c(s)$ by (14), where

$$c(s) = c_y(s,z) := \sum_{p\leq y}\frac{\log p}{p^s-1} - \sum_{p\geq y_1}\frac{(z-1)\log p}{(p^s-1)(p^s-z)} + (\log(\zeta(s)(s-1)))'.$$

So differentiating again with respect to s we obtain

$$\begin{aligned}G_y^{(2,0)}(s,z) &= zG_y^{(1,0)}(s,z)c(s) + zG_y(s,z)c'(s) \\ &= G_y(s,z)\big(z^2c(s)^2 + zc'(s)\big).\end{aligned} \tag{36}$$

The derivative of c with respect to s is

$$c'(s) = -\sum_{p\leq y}\frac{p^s(\log p)^2}{(p^s-1)^2} + \sum_{p\geq y_1}\frac{(z-1)(\log p)^2(2p^{2s}-(z+1)p^s)}{(p^s-1)^2(p^s-z)^2} + \big(\log(\zeta(s)(s-1))\big)''.$$

For fixed $s \in [1,2]$, note c and c' are meromorphic functions on $\mathbb{C}$, with poles located only at $z = p^s$ for each $p \ge y_1$. Thus

$$G_y(s,z) = \left(1 - \frac{z}{y_1^s}\right)^{-1} H_{y,1}(s,z),$$

$$z^2 c(s)^2 + zc'(s) = \left(1 - \frac{z}{y_1^s}\right)^{-2} H_{y,2}(s,z),$$

for functions $H_{y,1}, H_{y,2}$, whose smallest pole is at $z = y_2^s$ where y_2 is the smallest prime larger than y_1.

Letting $H_y := H_{y,1}H_{y,2}$, we see (36) becomes

$$G_y^{(2,0)}(s,z) = \left(1 - \frac{z}{y_1^s}\right)^{-3} H_y(s,z). \tag{37}$$

Note H_y has no poles inside $|z| \le y_2^{s-\varepsilon}$ so $\max_{s\in[1,2]} |H_y(s,z)| \ll_{y,\varepsilon} 1$ uniformly for $|z| \le y_2^{s-\varepsilon}$, as we take the maximum of the continuous function $|H_y(s,z)|$ over the compact set $\{(s,z) : 1 \le s \le 2, |z| \le y_2^{s-\varepsilon}\}$.

Thus by Cauchy's integral formula,

$$[z^i]H_y(s,z) = \frac{1}{2\pi i}\int_{|z|=y_2^{s-\varepsilon}} \frac{H_y(s,z)}{z^{i+1}}\mathrm{d}z \ll y_2^{-i(s-\varepsilon)} \max_{|z|=y_2^{s-\varepsilon}} |H_y(s,z)| \ll_{y,\varepsilon} y_2^{-i(1-\varepsilon)}$$

uniformly for $s \in [1,2]$. By the binomial theorem,

$$[z^i]\left(1 - \frac{z}{y_1^s}\right)^{-3} = y_1^{-is}\binom{i+2}{2} \ll (i+1)^2/y_1^i$$

uniformly for $s \in [1,2]$ and $i \ge 0$. Hence by the product rule, from (37) we conclude

$$[z^i]G_y^{(2,0)}(s,z) = \sum_{i_1+i_2=i} [z^{i_1}]\left(1 - \frac{z}{y_1^s}\right)^{-3} [z^{i_2}]H_y(s,z) \ll_y (i+1)^2/y_1^i.$$

This gives (35) as desired, which completes the proof.

Remark 16. By a similar proof as of (35) above, for any $y \ge 2$, $m \ge 0$,

$$[z^i]G_y^{(m,0)}(s,z) \ll_{m,y} (i+1)^m/y_1^i \tag{38}$$

holds uniformly for $s \in [1,2]$ and $i \ge 0$.

3. Proof of Theorem 2

Recall $f_k = f_{k,1}$, $F = F_1$ and $G = G_1$. By (16) with $y = 1$, we have

$$f_k = I_k + O\left(\frac{1}{k2^k}\right)$$

where

$$I_k = \int_1^2 \frac{1}{k!}F^{(0,k)}(s,0)\mathrm{d}s.$$

We apply the product rule to

$$F(s,z) = (s-1)^{-z}G(s,z) = (s-1)^{-z}\left(1 - \frac{z}{2^s}\right)^{-1} G_2(s,z),$$

giving

$$\frac{1}{k!}F^{(0,k)}(s,0) = \sum_{i+j+l=k} \frac{\left(-\log(s-1)\right)^l}{l!} 2^{-js}\frac{1}{i!}G_2^{(0,i)}(s,0),$$

using the power series in (18), as well as $\left(1-\frac{z}{2^s}\right)^{-1}=\sum_{j\geq 0}2^{-js}z^j$. Thus we have

$$I_k=\sum_{i+j+l=k}\int_1^2\frac{\left(-\log(s-1)\right)^l}{l!}2^{-js}\frac{1}{i!}G_2^{(0,i)}(s,0)\,\mathrm{d}s. \tag{39}$$

Now we introduce a similar integral I'_k, given by evaluating $G_2^{(0,i)}(s,0)$ in the integrand at $s=1$, namely,

$$I'_k=\sum_{i+j+l=k}\int_1^2\frac{\left(-\log(s-1)\right)^l}{l!}2^{-js}\frac{1}{i!}G_2^{(0,i)}(1,0)\,\mathrm{d}s. \tag{40}$$

Hence to establish Theorem 2, it suffices to prove the following lemmas for I'_k and $I_k-I'_k$.

Lemma 17. *We have* $I'_k=1-\frac{\log 2}{4}2^{-k}\left(dk^2+O(k\log(k+1))\right)$.

Lemma 18. *We have* $I_k=I'_k+O(k/2^k)$.

These are the $y=1$ analogues of Lemmas 14 and 15.

3.1. *Proof of Lemma 18*

By the mean value theorem we have, uniformly for $s\in[1,2]$,

$$\left|G_2^{(0,i)}(s,0)-G_2^{(0,i)}(1,0)\right|\leq(s-1)b_i$$

for coefficients

$$b_i:=\max_{s'\in[1,2]}\left|G_2^{(1,i)}(s',0)\right|.$$

Thus subtracting (39) from (40), we find

$$\begin{aligned}\left|I_k-I'_k\right|&=\left|\sum_{i+j+l=k}\int_1^2\frac{\left(-\log(s-1)\right)^l}{l!}2^{-js}\frac{1}{i!}\left(G_2^{(0,i)}(s,0)-G_2^{(0,i)}(1,0)\right)\mathrm{d}s\right|\\&\leq\sum_{i+j+l=k}\int_1^2\frac{\left(-\log(s-1)\right)^l}{l!}2^{-js}(s-1)\frac{b_i}{i!}\,\mathrm{d}s.\end{aligned} \tag{41}$$

By (38), we have uniformly for $t\in[1,2]$,

$$\frac{1}{i!}G_2^{(1,i)}(t,0)=[z^i]G_2^{(1,0)}(t,z)\ll(i+1)3^{-i}. \tag{42}$$

Hence $\frac{b_i}{i!}\ll(i+1)3^{-i}$, so that (41) implies

$$\begin{aligned}I_k-I'_k&\ll\sum_{i+j+l=k}\int_0^\infty\frac{t^l}{l!}2^{-j(1+\mathrm{e}^{-t})}\,\mathrm{e}^{-2t}(i+1)3^{-i}\,\mathrm{d}t\\&\leq\sum_{i+j+l=k}(i+1)3^{-i}2^{-j}\int_0^\infty\frac{t^l}{l!}\,\mathrm{e}^{-2t}\,\mathrm{d}t\\&=\sum_{i+j+l=k}(i+1)3^{-i}2^{-j-l-1}\int_0^\infty\frac{u^l}{l!}\,\mathrm{e}^{-u}\,\mathrm{d}u\;=\sum_{i+j+l=k}(i+1)3^{-i}2^{-j-l-1}\end{aligned}$$

where the last equalities follow from substituting $u/2$ for t and recalling the integral form (21) of the Gamma function. Hence we conclude

$$I_k-I'_k\ll\sum_{i\leq k}(i+1)3^{-i}2^{i-k}\sum_{j\leq k-i}1\leq k2^{-k}\sum_{i\leq k}(i+1)(2/3)^i\ll k2^{-k}.$$

3.2. *Proof of Lemma 17*

Recall

$$I'_k = \sum_{i=0}^{k} \frac{1}{i!} G_2^{(0,i)}(1,0) \sum_{j+l=k-i} \int_1^2 \frac{\big(-\log(s-1)\big)^l}{l!} 2^{-js}\,\mathrm{d}s. \tag{43}$$

We will prove in the next subsection that

Lemma 19. *For $k \ge 1$, we have*

$$\sum_{j+l=k} \int_1^2 \frac{\big(-\log(s-1)\big)^l}{l!} 2^{-sj}\,\mathrm{d}s = 2 - \frac{\log 2}{4} 2^{-k}\Big(k^2 + O\big((k+1)\log(k+2)\big)\Big).$$

Remark 20. The relative saving is $k/\log k$ and it appears sharp. We find it to be an unusual saving.

Using Lemma 19 we simplify the inner sum in (43) and find

$$\begin{aligned} I'_k &= \sum_{i=0}^{k} \frac{1}{i!} G_2^{(0,i)}(1,0)\left(2 - \frac{\log 2}{4} 2^{i-k}\Big((k-i)^2 + O\big((k+1-i)\log(k+2-i)\big)\Big)\right) \\ &= 2I'_{k,2} - \frac{\log 2}{4} 2^{-k} B \end{aligned} \tag{44}$$

where

$$I'_{k,2} := \sum_{i=0}^{k} \frac{1}{i!} G_2^{(0,i)}(1,0), \tag{45}$$

$$B := \sum_{i=0}^{k} \frac{2^i}{i!}\Big((k-i)^2 + O\big((k-i+1)\log(k-i+2)\big)\Big) G_2^{(0,i)}(1,0). \tag{46}$$

Lemma 14 with $y = 2$ yields

$$I'_{k,2} = G_2(1,1) + O(3^{-k}) = \frac{1}{2} + O(3^{-k}).$$

Similarly as in the proof of Lemma 14, Cauchy's integral formula and residue theorem imply

$$\begin{aligned} B' &:= \sum_{i=0}^{k} \frac{2^i}{i!} G_2^{(0,i)}(1,0) = [z^k]\frac{G_2(1,2z)}{1-z} \\ &= \frac{1}{2\pi i}\int_{|z|=1/4} \frac{G_2(1,2z)}{1-z}\frac{\mathrm{d}z}{z^{k+1}} \\ &= \frac{1}{2\pi i}\int_{|z|=2} \frac{G_2(1,2z)}{1-z}\frac{\mathrm{d}z}{z^{k+1}} + G_2(1,2) - \frac{\lim_{z\to 3/2}(z-3/2)\,G_y(1,2z)}{(1-3/2)(3/2)^{k+1}} \\ &= d + O((3/2)^{-k}). \end{aligned} \tag{47}$$

Here $G_2(1,2) = \frac{1}{4}\prod_{p>2}(1-\frac{2}{p})^{-1}(1-\frac{1}{p})^2 = d$. We also note $G_2(1,z)$ is meromorphic in $|z| < 4$ with simple pole at $z = 3$ (so $z = 3/2$ is the smallest pole of $G_2(1,2z)$). In particular,

$$\frac{1}{i!} G_2^{(0,i)}(1,0) = [z^i]G_2(1,z) \ll 3^{-i}. \tag{48}$$

Thus combining (46), (47) and (48), we obtain

$$\begin{aligned} B - k^2 B' &= \sum_{i=0}^{k} \frac{2^i}{i!} G_2^{(0,i)}(1,0)((k-i)^2 + O((k-i+1)\log(k-i+2)) - k^2) \\ &\ll \sum_{i=0}^{k} (2/3)^i (ik + O(k\log(k+1))) \ll k\log(k+1). \end{aligned}$$

Hence $B = dk^2 + O(k\log(k+1))$, so plugging back into (44) we conclude

$$I_k' = 1 - \frac{\log 2}{4} 2^{-k}\big(dk^2 + O(k\log(k+1))\big). \tag{49}$$

3.3. *Proof of Lemma 19*

We may suppose $k \geq 2$. Multiplying through by 2^k, we aim to prove

$$A_k = 2^{k+1} - \frac{\log 2}{4}k^2 + O(k\log k), \tag{50}$$

for

$$A_k := 2^k \sum_{i+j=k} \int_1^2 \frac{\big(-\log(s-1)\big)^i}{i!} 2^{-js}\mathrm{d}s = \sum_{i=0}^{k} \frac{2^i}{i!} J(i)$$

where, substituting $s = 1 + \mathrm{e}^{-u}$,

$$J(i) := \int_1^2 \big(-\log(s-1)\big)^i 2^{-(k-i)(s-1)}\mathrm{d}s = \int_0^\infty u^i\,\mathrm{e}^{-u}\, 2^{-(k-i)\,\mathrm{e}^{-u}}\mathrm{d}u.$$

Note the trivial bound $J(i) \leq i!$, using $2^{-(k-i)\,\mathrm{e}^{-u}} \leq 1$.

In order to conclude (50), it suffices to prove the following two estimates and apply them with $T = 15\log k$:

$$\sum_{0\leq i\leq T} \frac{2^i}{i!} J(i) = \sum_{0\leq i\leq T} 2^i + O(kT) \qquad \text{for} \quad k \geq T \geq \log k, \tag{51}$$

$$\sum_{T<i\leq k} \frac{2^i}{i!} J(i) = 2^{k+1} - \frac{\log 2}{4}k^2 - \sum_{0\leq i\leq T} 2^i + O(kT) \qquad \text{for} \quad k \geq T \geq 15\log k. \tag{52}$$

We first prove (51). For $T \geq \log k$, the contribution of $\mathrm{e}^u \geq k$ to $J(i)$, i.e. $u \geq \log k$, is handled by the Taylor expansion $2^{-(k-i)\,\mathrm{e}^{-u}} = 1 - O(k\,\mathrm{e}^{-u})$. Thus

$$\begin{aligned}
J(i) &= \int_0^{\log k} u^i\,\mathrm{e}^{-u}\, 2^{-(k-i)\,\mathrm{e}^{-u}}\,\mathrm{d}u + \int_{\log k}^\infty u^i\,\mathrm{e}^{-u}(1 + O(k\,\mathrm{e}^{-u}))\mathrm{d}u \\
&= \int_0^{\log k} u^i\,\mathrm{e}^{-u}\, 2^{-(k-i)\,\mathrm{e}^{-u}}\,\mathrm{d}u + \int_{\log k}^\infty u^i\,\mathrm{e}^{-u}\,\mathrm{d}u + O(k2^{-i})\int_0^\infty v^i\,\mathrm{e}^{-v}\,\mathrm{d}v \\
&= \int_0^{\log k} u^i\,\mathrm{e}^{-u}\, 2^{-(k-i)\,\mathrm{e}^{-u}}\,\mathrm{d}u + \int_0^\infty u^i\,\mathrm{e}^{-u}\,\mathrm{d}u - \int_0^{\log k} u^i\,\mathrm{e}^{-u}\,\mathrm{d}u + O(k2^{-i}\,i!) \\
&= i! + \int_0^{\log k} u^i\,\mathrm{e}^{-u}(2^{-(k-i)\,\mathrm{e}^{-u}} - 1)\,\mathrm{d}u + O(k2^{-i}\,i!).
\end{aligned}$$

Summing over $i \leq T$, we obtain

$$\sum_{0\leq i\leq T} \frac{2^i}{i!} J(i) = \sum_{0\leq i\leq T} 2^i - \sum_{0\leq i\leq T} \frac{2^i}{i!}\int_0^{\log k} u^i\,\mathrm{e}^{-u}(1 - 2^{-(k-i)\,\mathrm{e}^{-u}})\,\mathrm{d}u + O(kT). \tag{53}$$

Thus to conclude (51) from (53), it remains show

$$\sum_{0\leq i\leq T} \frac{2^i}{i!}\int_0^{\log k} u^i\,\mathrm{e}^{-u}(1 - 2^{-(k-i)\,\mathrm{e}^{-u}})\,\mathrm{d}u = O(k\log k). \tag{54}$$

The contribution of $i \leq \log k$ is

$$\sum_{0\leq i\leq \log k} \frac{2^i}{i!}\int_0^\infty u^i\,\mathrm{e}^{-u}\,\mathrm{d}u \leq \sum_{0\leq i\leq \log k} 2^i = O(2^{\log k}) = O(k) \tag{55}$$

since $2 < e$. Next recall the function $u \mapsto u^i \mathrm{e}^{-u}$ is increasing for $u \le i$, implying

$$\sum_{\log k < i \le T} \frac{2^i}{i!} \int_0^{\log k} u^i \mathrm{e}^{-u}\,\mathrm{d}u \le \sum_{\log k < i \le T} \frac{2^i}{i!} \log k (\log k)^i \mathrm{e}^{-\log k} \le \frac{\log k}{k} \sum_{i=0}^{\infty} \frac{(2\log k)^i}{i!} = k\log k. \tag{56}$$

Combining (55) and (56) gives (54) as desired. This completes the proof of (51).

Now to prove (52), we begin by handling the contribution of small u to the integral $J(i)$. Since the function $u \mapsto u^i \mathrm{e}^{-u}$ increases for $u \le i$, for fixed $a \in (0,1)$, the contribution of $u \le ai$ to $J(i)$ is at most

$$\int_0^{ai} u^i \mathrm{e}^{-u} 2^{-(k-i)\mathrm{e}^{-u}} \mathrm{d}u \le \int_0^{ai} u^i \mathrm{e}^{-u}\,\mathrm{d}u \le (ai)(ai)^i \mathrm{e}^{-ai} \ll_a i!\sqrt{i}(ea/\mathrm{e}^a)^i$$

by Stirling's approximation. Thus when a is small enough to satisfy $\mathrm{e}^a > 2ea$, we have

$$\int_0^{ai} u^i \mathrm{e}^{-u} 2^{-(k-i)\mathrm{e}^{-u}} \mathrm{d}u \ll_a i! 2^{-i}. \tag{57}$$

For concreteness, we fix $a = 0.21$. Now let $c := a/3$. For $T < i \le k$, if $u \ge ci$ we have

$$(k-i)\mathrm{e}^{-u} \le k\mathrm{e}^{-ci} < k\mathrm{e}^{-cT} \le k^{1-15c} = k^{-.05}$$

since $T \ge 15\log k$. In particular, we may use the 2nd order Taylor expansion

$$2^{-(k-i)\mathrm{e}^{-u}} = 1 - \log 2(k-i)\mathrm{e}^{-u} + O(k^2 \mathrm{e}^{-2u}).$$

Thus by (57) we have for $T < i \le k$,

$$\begin{aligned} J(i) &= \int_{ci}^{\infty} u^i \mathrm{e}^{-u} 2^{-(k-i)\mathrm{e}^{-u}} \mathrm{d}u + O(i!2^{-i}) \\ &= \int_{ci}^{\infty} u^i \mathrm{e}^{-u} \left(1 - \log 2\,(k-i)\mathrm{e}^{-u} + O\left(k^2 \mathrm{e}^{-2u}\right)\right) \mathrm{d}u + O(i!2^{-i}) \\ &= A_{i,1} - \log 2(k-i) A_{i,2} + O(k^2 A_{i,3}) + O(i!2^{-i}) \end{aligned} \tag{58}$$

where, for $j = 1,2,3$,

$$A_{i,j} := \int_{ci}^{\infty} u^i \mathrm{e}^{-ju}\,\mathrm{d}u = j^{-i-1} \int_{cij}^{\infty} v^i \mathrm{e}^{-v}\,\mathrm{d}v = j^{-i-1}\, i!\,(1 + O(2^{-i})). \tag{59}$$

In the last equality in (59) we used (57). Plugging (59) back into (58) and dividing through by $i!$, we find that

$$\begin{aligned} \frac{J(i)}{i!} &= (1 + O(2^{-i})) - \log 2(k-i) 2^{-i-1}(1 + O(2^{-i})) + O(k^2 3^{-i-1} + 2^{-i}) \\ &= 1 - \log 2(k-i)2^{-i-1} + O(k^2 3^{-i} + 2^{-i}). \end{aligned} \tag{60}$$

Summing over $i \in (T,k]$, we conclude that

$$\begin{aligned} \sum_{T<i\le k} \frac{2^i}{i!} J(i) &= \sum_{T<i\le k} \left(2^i - \log 2(k-i)2^{-1} + O(k^2(2/3)^i + 1)\right) \\ &= 2^{k+1} - \sum_{i\le T} 2^i - \frac{\log 2}{2} \sum_{T<i\le k} (k-i) \; + \; O(k^2(2/3)^T + k) \\ &= 2^{k+1} - \sum_{i\le T} 2^i - \frac{\log 2}{2}\left(\frac{k^2}{2} + O(Tk)\right) \end{aligned}$$

holds. Here $k^2(2/3)^T \le k^{2+15\log(2/3)} < k^{-4}$ since $T \ge 15\log k$. This gives (52) as desired, and hence completes the proof of Lemma 19.

4. Probability-theoretic argument

In this section, we give an alternative probabilistic interpretation of Erdős sums, showing

Proposition 21. *We have $f_k = 1 + O(k/2^{k/4})$.*

In view of Theorem 2 we haven't tried to optimize the exponent $2^{k/4}$.

For an integer $a \geq 1$, let $P^+(a)$ and $P^-(a)$ denote the largest and smallest prime factors of a, respectively (here $P^+(1) := 1$ and $P^-(1) := 1$). Also let $P_j(n)$ denote the jth largest prime of n, with multiplicity, so that $n = P_1(n) \cdots P_k(n)$. In particular $P_1(n) = P^+(n)$.

Define the set of L-multiples L_a,[4]

$$\mathrm{L}_a := \{ba \in \mathbb{N} : P^-(b) \geq P^+(a)\}.$$

We define the (natural) density of a set $A \subseteq \mathbb{N}$ to be $\mathrm{d}(A) := \lim_{x\to\infty} |A \cap [1,x]|/x$ as long as this limit exists. Note $\mathrm{d}(\mathrm{L}_a) = \frac{1}{a}\prod_{p<P^+(a)}(1-1/p)$.

4.1. *Preliminary lemmas*

We begin with some preliminaries.

Lemma 22. *For any $a \in \mathbb{N}$, we have*

$$\sum_{p \geq P^+(a)} \mathrm{d}(\mathrm{L}_{ap}) = \mathrm{d}(\mathrm{L}_a).$$

Proof. Let $y > 1$. Consider the set of positive integers without prime factors smaller than y, and partition it according to the smallest prime factor $q \geq y$. This gives the disjoint union,

$$\{b \in \mathbb{N} : P^+(b) \geq y\} \ = \ \bigcup_{q \geq y} \{bq \in \mathbb{N} : P^-(b) \geq q\}.$$

Taking the density of both sides, we find that

$$\prod_{p<y}\left(1-\frac{1}{p}\right) = \sum_{q\geq y}\frac{1}{q}\prod_{p<q}\left(1-\frac{1}{p}\right). \tag{61}$$

Now choosing $y = P^+(a)$, we divide (61) by a to conclude $\mathrm{d}(\mathrm{L}_a) = \sum_{q\geq P^+(a)} \mathrm{d}(\mathrm{L}_{aq})$. □

From Lemma 22, a simple induction argument on $j \geq 1$ gives

$$\sum_{\substack{\Omega(b)=j \\ p(b)\geq P^+(a)}} \mathrm{d}(\mathrm{L}_{ab}) = \mathrm{d}(\mathrm{L}_a). \tag{62}$$

In particular when $a = 1$, for any $j \geq 1$ we have $\sum_{\Omega(b)=j} \mathrm{d}(\mathrm{L}_b) = 1$. We shall refine this result in the lemma below.

Lemma 23. *Uniformly for $0 < v < 1$ and $a \in \mathbb{Z}_{>1}$, we have*

$$\sum_{q \geq P_1(a)^{\frac{1}{v}}} \mathrm{d}(\mathrm{L}_{aq}) = v\,\mathrm{d}(\mathrm{L}_a)\left(1 + O\left(\frac{1}{\log P_1(a)}\right)\right). \tag{63}$$

Proof. Take $0 < v < 1$. We first recall Mertens' product theorem states that

$$\prod_{p<x}\left(1-\frac{1}{p}\right) = \frac{\mathrm{e}^{-\gamma}}{\log x}\left(1+O\left(\frac{1}{\log x}\right)\right)$$

[4] L for lexicographic

holds for $x \ge 2$. In particular, for $x = P_1(a) \ge 2$,

$$\prod_{P_1(a)\le p<P_1(a)^{\frac{1}{v}}}\left(1-\frac{1}{p}\right) = \prod_{p<P_1(a)}\left(1-\frac{1}{p}\right)^{-1}\prod_{p<P_1(a)^{\frac{1}{v}}}\left(1-\frac{1}{p}\right)$$
$$= \frac{\log P_1(a)}{\log P_1(a)^{\frac{1}{v}}}\left(1+O\left(\frac{1}{\log P_1(a)}\right)\right) \;=\; v\left(1+O(1/\log P_1(a))\right). \tag{64}$$

So by (64) and (61) with $y = P_1(n)^{\frac{1}{v}}$,

$$\sum_{q\ge P_1(a)^{\frac{1}{v}}}\frac{1}{q}\prod_{p<q}\left(1-\frac{1}{p}\right) = \prod_{p<P_1(a)^{\frac{1}{v}}}\left(1-\frac{1}{p}\right) = v\left(1+O(1/\log P_1(a))\right)\prod_{p<P_1(a)}\left(1-\frac{1}{p}\right).$$

Dividing through by a completes the proof. □

Lemma 24. *For $k \ge 1$, let $c_1 \ge \cdots \ge c_k \ge 0$. If $d_1, D_1, E_1, \ldots, d_k, D_k, E_k \ge 0$ satisfy $E_i \le \sum_{j=1}^{i} d_j \le D_i$ for all $1 \le i \le k$ (and let $d_0 = E_0 = D_0 = 0$), then we have*

$$\sum_{i=1}^{k} c_i(E_i - E_{i-1}) \;\le\; \sum_{i=1}^{k} c_i d_i \;\le\; \sum_{i=1}^{k} c_i(D_i - D_{i-1}).$$

Proof. We have

$$\sum_{i=1}^{k} c_i d_i = \sum_{i=1}^{k} c_i\left(\sum_{j=1}^{i} d_j - \sum_{j=0}^{i-1} d_j\right) = \sum_{i=1}^{k-1}(c_i - c_{i+1})\sum_{j=1}^{i} d_j \;+\; c_k\sum_{i=1}^{k} d_i \tag{65}$$

by summation by parts. Since $c_i - c_{i+1} \ge 0$ and $\sum_{j\le i} d_j \le D_i$, from (65) we obtain that

$$\sum_{i=1}^{k} c_i d_i \le \sum_{i=1}^{k-1}(c_i - c_{i+1})D_i \;+\; c_k D_k = \sum_{i=1}^{k} c_i(D_i - D_{i-1})$$

holds. Similarly, since $\sum_{j\le i} d_j \ge E_i \ge 0$, from (65) we obtain that

$$\sum_{i=1}^{k} c_i d_i \ge \sum_{i=1}^{k-1}(c_i - c_{i+1})E_i \;+\; c_k E_k = \sum_{i=1}^{k} c_i(E_i - E_{i-1})$$

holds. □

To handle the contribution of smooth numbers, we use a simple bound of Erdős and Sárközy [7, Lemma 2], whose proof we provide for completeness.

Lemma 25 (Erdős–Sárközy). *For any $k \ge 1$, $y > 1$, we have*

$$\sum_{\substack{\Omega(n)=k\\ P_1(n)<e^y}}\frac{1}{n} \;\ll\; y^2 k/2^k.$$

Proof. Observe that 2^k times our given sum is bounded by the following Euler product,

$$\sum_{\substack{\Omega(n)=k\\ P_1(n)<e^y}}\frac{2^k}{n} \le \prod_{p<e^y}\left(1+\frac{2}{p}+\cdots+\frac{2^k}{p^k}\right)$$
$$\le (k+1)\prod_{2<p<e^y}\left(1-\frac{2}{p}\right)^{-1} \ll ky^2$$

by Mertens' product theorem. Dividing by 2^k completes the proof. □

Corollary 26. *For any $1 \le j \le k$ and $y > 1$, we have*

$$\sum_{\substack{\Omega(n)=k\\ P_{j+1}(n)<e^y}}\frac{1}{n\log n} \ll \sum_{\substack{\Omega(n)=k\\ P_{j+1}(n)<e^y}} \mathrm{d}(\mathrm{L}_n) \ll y^2 k 2^{j-k}.$$

Proof. First, for each n with $\Omega(n)=k$, one can factor n uniquely as ab with $\Omega(b)=j$ and $p(b)\geq P_1(a)$ (namely take $b=\prod_{i=1}^{j}P_i(n)$ and $a=n/b$). Thus by (62) we have

$$\sum_{\substack{\Omega(n)=k\\ P_{j+1}(n)<\mathrm{e}^y}} \mathrm{d}(\mathrm{L}_n) = \sum_{\substack{\Omega(a)=k-j\\ P_1(a)<\mathrm{e}^y}} \sum_{\substack{\Omega(b)=j\\ p(b)\geq P_1(a)}} \mathrm{d}(\mathrm{L}_{ab}) = \sum_{\substack{\Omega(a)=k-j\\ P_1(a)<\mathrm{e}^y}} \mathrm{d}(\mathrm{L}_a).$$

On the right-hand side of the above identity we apply the simple bound $\mathrm{d}(\mathrm{L}_a)\ll 1/a$, and on the left-hand side we apply $\mathrm{d}(\mathrm{L}_n)\gg 1/(n\log n)$. This gives

$$\sum_{\substack{\Omega(n)=k\\ P_{j+1}(n)<\mathrm{e}^y}} \frac{1}{n\log n} \ll \sum_{\substack{\Omega(n)=k\\ P_{j+1}(n)<\mathrm{e}^y}} \mathrm{d}(\mathrm{L}_n) \ll \sum_{\substack{\Omega(a)=k-j\\ P_1(a)<\mathrm{e}^y}} \frac{1}{a} \ll y^2 k 2^{j-k}$$

by Lemma 25 with k replaced by $k-j$. □

4.2. *Proof of Proposition 21*

Let $k\geq 1$ be sufficiently large. We shall choose $y=2^j$ for $j=\lfloor k/4\rfloor$, and $N=4^k$. Let f_k' denote the sum f_k restricted by $P_{j+1}(n)\geq \mathrm{e}^y$. Thus by Corollary 26,

$$f_k = \sum_{\Omega(n)=k} \frac{1}{n\log n} = f_k' + O(y^2 k 2^{j-k}) = f_k' + O(k/2^{k/4}) \tag{66}$$

where, by Mertens' product theorem,

$$f_k' := \sum_{\substack{\Omega(n)=k\\ P_{j+1}(n)\geq \mathrm{e}^y}} \frac{1}{n\log n} = \sum_{\substack{\Omega(n)=k\\ P_{j+1}(n)\geq \mathrm{e}^y}} \left(\mathrm{e}^\gamma + \frac{O(1)}{\log P_1(n)}\right) \frac{\log P_1(n)}{\log n} \mathrm{d}(\mathrm{L}_n). \tag{67}$$

Next, we rewrite the identity $n=P_1(n)\cdots P_k(n)$ as

$$\frac{\log n}{\log P_1(n)} = 1 + \frac{\log P_2(n)}{\log P_1(n)}\left(1+\cdots\left(1+\frac{\log P_{j+1}(n)}{\log P_j(n)}\left(1+\frac{\log\big(P_{j+2}(n)\cdots P_k(n)\big)}{\log P_{j+1}(n)}\right)\right)\cdots\right).$$

Taking the reciprocal of identity above gives

$$\frac{\log P_1(n)}{\log n} = u_{j+1}\left(\frac{\log P_2(n)}{\log P_1(n)},\ldots,\frac{\log P_{j+1}(n)}{\log P_j(n)},\frac{\log(P_{j+2}(n)\cdots P_k(n))}{\log P_{j+1}(n)}\right),$$

for the functions $u_j:\mathbb{R}^j\to\mathbb{R}$ given by

$$u_j(x_1,\ldots,x_j) := \frac{1}{1+x_1(1+x_2(\cdots(1+x_j)\cdots))}. \tag{68}$$

In particular, from $P_1(n)\geq\cdots\geq P_k(n)$ we infer the inequalities

$$\frac{\log P_1(n)}{\log n} \leq u_j\left(\frac{\log P_2(n)}{\log P_1(n)},\ldots,\frac{\log P_{j+1}(n)}{\log P_j(n)}\right), \tag{69}$$

$$\frac{\log P_1(n)}{\log n} \geq u_j\left(\frac{\log P_2(n)}{\log P_1(n)},\ldots,\frac{\log P_{j+1}(n)}{\log P_j(n)}(k-j)\right). \tag{70}$$

By (69), we see (67) implies that

$$f_k' \leq \sum_{\substack{\Omega(a)=k-j\\ P_1(a)\geq \mathrm{e}^y}} \left(\mathrm{e}^\gamma + \frac{O(1)}{\log P_1(a)}\right) \sum_{P_1(a)\leq p_j\leq\cdots\leq p_1} u_j\left(\frac{\log p_2}{\log p_1},\ldots,\frac{\log P_1(a)}{\log p_j}\right) \mathrm{d}(\mathrm{L}_{ap_j\cdots p_1}). \tag{71}$$

Lemma 27. *There is an absolute constant $C>1$ such that for any $a\in\mathbb{Z}_{>1}$,*

$$\sum_{P_1(a)\le p_j\le\cdots\le p_1} u_j\left(\frac{\log p_2}{\log p_1},\dots,\frac{\log P_1(a)}{\log p_j}\right)\mathrm{d}(\mathrm{L}_{ap_j\cdots p_1}) \le \frac{\left(1+\frac{C}{\log P_1(a)}\right)^j}{N^j}\sum_{i_1,\dots,i_j=1}^{N} u_j\left(\frac{i_1-1}{N},\dots,\frac{i_j-1}{N}\right)\mathrm{d}(\mathrm{L}_a).$$

Proof. For each $1\le r\le j$, it suffices to show that

$$N^{1-r}\sum_{i_1,\dots i_{r-1}=1}^{N}\sum_{P_1(a)\le p_j\le\cdots\le p_{r+1}\le p_r} u_j\left(\frac{i_1-1}{N},\dots,\frac{i_{r-1}-1}{N},\frac{\log p_{r+1}}{\log p_r},\dots,\frac{\log P_1(a)}{\log p_j}\right)\mathrm{d}(\mathrm{L}_{ap_j\cdots p_r})$$
$$\le N^{-r}\sum_{i_1,\dots,i_r=1}^{N}\sum_{P_1(a)\le p_j\le\cdots\le p_{r+1}} u_j\left(\frac{i_1-1}{N},\dots,\frac{i_r-1}{N},\frac{\log p_{r+2}}{\log p_{r+1}},\dots,\frac{\log P_1(a)}{\log p_j}\right)\mathrm{d}(\mathrm{L}_{ap_j\cdots p_{r+1}})$$
$$\times\left(1+\frac{C}{\log P_1(a)}\right) \quad (72)$$

holds. Indeed, iterating (72) (with each $r=1,2,\dots,j$ in turn) completes the proof of the lemma.

To show that (72) holds, fix indices $i_1,\dots i_{r-1}\le N$ and primes $p_j\le\cdots\le p_{r+1}$ ($p_j\ge P_1(a)$). Define c_{i_r} and d_{i_r} by

$$c_{i_r} := u_j\left(\frac{i_1-1}{N},\dots,\frac{i_r-1}{N},\frac{\log p_{r+2}}{\log p_{r+1}},\dots,\frac{\log P_1(a)}{\log p_j}\right)$$
$$d_{i_r} := \sum_{p_r\in[p_{r+1}^{N/i_r},\,p_{r+1}^{N/(i_r-1)})}\mathrm{d}(\mathrm{L}_{ap_j\cdots p_r}).$$

(For $i_r=1$, the range of p_r in the definition of d_{i_r} is to be interpreted as $[p_{r+1}^N,\infty)$.) Note for any $u\le N$, by Lemma 23 we have

$$\sum_{i_r=1}^{u} d_{i_r} = \sum_{p_r\ge p_{r+1}^{N/u}}\mathrm{d}(\mathrm{L}_{ap_j\cdots p_r}) \le \frac{u}{N}\mathrm{d}(\mathrm{L}_{ap_j\cdots p_{r+1}})\left(1+\frac{C}{\log P_1(a)}\right) =: D_u.$$

In particular $D_u-D_{u-1}=\frac{1}{N}\mathrm{d}(\mathrm{L}_{ap_j\cdots p_{r+1}})(1+C/\log P_1(a))$. Splitting up the sum over $p_r\ge p_{r+1}$ below according to the i_r for which $p_r\in\left[p_{r+1}^{N/i_r},p_{r+1}^{N/(i_r-1)}\right)$ holds, and then applying Lemma 24, we find that

$$\sum_{p_r\ge p_{r+1}} u_j\left(\frac{i_1-1}{N},\dots,\frac{i_{r-1}-1}{N},\frac{\log p_{r+1}}{\log p_r},\dots,\frac{\log P_1(a)}{\log p_j}\right)\mathrm{d}(\mathrm{L}_{ap_j\cdots p_r})$$
$$\le\sum_{i_r=1}^{N} c_{i_r}\sum_{p_r\in[p_{r+1}^{N/i_r},\,p_{r+1}^{N/(i_r-1)})}\mathrm{d}(\mathrm{L}_{ap_j\cdots p_r}) = \sum_{i_r=1}^{N} c_{i_r}d_{i_r}$$
$$\le\sum_{i_r=1}^{N} c_{i_r}(D_{i_r}-D_{i_r-1}) = \left(1+\frac{C}{\log P_1(a)}\right)\frac{1}{N}\sum_{i_r=1}^{N} c_{i_r}\mathrm{d}(\mathrm{L}_{ap_j\cdots p_{r+1}}) \quad (73)$$
$$=\left(1+\frac{C}{\log P_1(a)}\right)\frac{1}{N}\sum_{i_r=1}^{N} u_j\left(\frac{i_1-1}{N},\dots,\frac{i_r-1}{N},\frac{\log p_{r+2}}{\log p_{r+1}},\dots,\frac{\log P_1(a)}{\log p_j}\right)\mathrm{d}(\mathrm{L}_{ap_j\cdots p_{r+1}}).$$

Summing (73) over $i_1,\dots,i_{r-1}\le N$ and $p_j\le\cdots\le p_{r+1}$, we obtain (72) as desired. □

Plugging Lemma 27 into (71) we obtain that

$$f_k'\le \mathrm{e}^{\gamma}\sum_{\substack{\Omega(a)=k-j\\ P_1(a)\ge \mathrm{e}^y}}\mathrm{d}(\mathrm{L}_a)\,\frac{(1+C/y)^{j+1}}{N^j}\sum_{i_1,\dots,i_j=1}^{N} u_j\left(\frac{i_1-1}{N},\dots,\frac{i_j-1}{N}\right) \quad (74)$$

holds. By Corollary 26 and (62) with $a = 1$,

$$\sum_{\substack{\Omega(a)=k-j\\ P_1(a)\geq e^y}} \mathrm{d}(\mathrm{L}_a) = \sum_{\Omega(a)=k-j} \mathrm{d}(\mathrm{L}_a) + O(y^2 k 2^{j-k}) = 1 + O(k/2^{k/4}),$$

recalling the definitions $y = 2^j$, $j = \lfloor k/4 \rfloor$. Thus

$$f_k' \leq e^{\gamma} \frac{\left(1 + O(k/2^{k/4})\right)^{j+1}}{N^j} \sum_{i_1,\ldots,i_j=1}^{N} u_j\left(\frac{i_1-1}{N},\ldots,\frac{i_j-1}{N}\right). \tag{75}$$

By an analogous argument (using the lower bound in (70), and $E_i = \frac{i}{N}\mathrm{d}(\mathrm{L}_{p_2\cdots p_j a})(1 - C/\log P_1(a))$ in Lemma 24 instead of D_i), we may obtain a similar lower bound

$$f_k' \geq e^{\gamma} \frac{\left(1 - O(k/2^{k/4}))\right)^{j+2}}{N^j} \sum_{i_1,\ldots,i_j=1}^{N} u_j\left(\frac{i_1}{N},\ldots,\frac{i_j}{N}(k-j)\right). \tag{76}$$

Now for a sequence $(c_j)_j$, define the integral

$$I_j(c_j) := \int_{[0,1]^j} u_j(x_1,\ldots,x_{j-1},c_j x_j)\,\mathrm{d}x_1\cdots\mathrm{d}x_j. \tag{77}$$

Observe that the sum in (75) is the upper Riemann sum for the integral $I_j(1)$, noting that $u_j : [0,1]^j \to [0,1]$ is decreasing in each component. And since the upper and lower Riemann sums (which squeeze $I_j(1)$) overlap in $(N-1)^j$ points, their difference is $\ll (N^j - (N-1)^j)/N^j = 1-(1-1/N)^j \ll j/N$. In particular the sum in (75) equals $I_j(1) + O(j/N)$.

Similarly (75) is the lower Riemann sum for $I_j(k-j)$. Thus we obtain

$$N^{-j} \sum_{i_1,\ldots,i_j=1}^{N} u_j\left(\frac{i_1-1}{N},\ldots,\frac{i_j-1}{N}\right) = I_j(1) + O(j/N),$$

$$N^{-j} \sum_{i_1,\ldots,i_j=1}^{N} u_j\left(\frac{i_1}{N},\ldots,\frac{i_j}{N}(k-j)\right) = I_j(k-j) + O(j/N).$$

Recalling $N = 4^k$ and $j = \lfloor k/4 \rfloor$, we see that (75) and (76) become

$$f_k' \leq (e^{\gamma} + O(k/2^{k/4}))\, I_j(1), \tag{78}$$

$$f_k' \geq (e^{\gamma} - O(k/2^{k/4}))\, I_j(k-j). \tag{79}$$

In the next section, we shall establish the following quantitative result.

Theorem 28. *Let $(c_j)_j$ be any nonnegative sequence. Then $I_j(c_j)$, as in* (77)*, satisfies*

$$I_j(c_j) = e^{-\gamma} + O\left(2^{-j}\left(1 + c_j\right)\right).$$

Remark 29. The qualitative result that $I_n(c_n) = e^{-\gamma} + o(1)$ may be established in the wider regime where $\limsup_{n\to\infty} \frac{1}{n}\log c_n < 1$ holds (see Lemma 35 for more precise statement) but the bound above is sufficient for the purpose of this article.

In particular, Theorem 28 gives

$$I_j(1) = e^{-\gamma} + O(2^{-j}),$$

$$I_j(k-j) = e^{-\gamma} + O(k/2^j).$$

Thus plugging into (78) and (79) we obtain $f_k' = 1 + O(k/2^{k/4})$. Hence by (66) we conclude that

$$f_k = f_k' + O(k/2^{k/4}) = 1 + O(k/2^{k/4}). \tag{80}$$

This completes the proof of Proposition 21.

5. A sequence of integrals

In this section, we prove Theorem 28. This implies Theorem 6 for $I_j(1)$. Recalling u_j in (68), we defined the following sequence of integrals, for a sequence $(c_j)_j$,

$$I_j = I_j(c_j) := \int_{[0,1]^j} \frac{\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_j}{1+x_1(1+x_2(\cdots(1+x_{j-1}(1+c_jx_j))\cdots))}. \tag{81}$$

This sequence of iterated integrals is closely related to the so-called Dickman–Goncharov distribution, the properties of which are well studied in the literature (see e.g. [14, Props. 2.1 and 2.4], [3] and [15]). Since we need small refinements of existing results, we will provide self-contained explanations for all the results below. Our approach will be based on techniques from random iterated functions/stochastic fixed-point equations.

5.1. *Probabilistic setup*

In the following, all random variables are assumed to live in a common reference probability space $(\Omega, \mathscr{F}, \mathbb{P})$.

Lemma 30. *Let $U, U_1, U_2, \dots$ be i.i.d.* Uniform$[0,1]$ *random variables. Define*

$$F_n(x) := 1 + U_n x \qquad \forall x \in \mathbb{R}, \quad n \in \mathbb{N},$$

and consider the sequence of iterated random functions

$$S_0(x) := x, \qquad S_n(x) := F_1 \circ F_2 \circ \cdots \circ F_n(x) \qquad \forall n \in \mathbb{N}. \tag{82}$$

Then the following statements hold.

(i) *The random variable*

$$S_\infty := \lim_{n\to\infty} S_n(1) = 1 + \sum_{j=1}^{\infty}\prod_{k=1}^{j} U_k \tag{83}$$

exists almost surely and satisfies $\mathbb{P}(1 \le S_\infty < \infty) = 1$.

(ii) *Let $\theta \in (1, e)$. If $(V_n)_n$ is a sequence of random variables such that* $\lim_{n\to\infty}\theta^{-n}|V_n| = 0$ *almost surely, then*

$$\lim_{n\to\infty} S_n(V_n) = S_\infty \qquad \textit{almost surely.} \tag{84}$$

Remark 31. The composition of maps $S_n(x) := F_1 \circ \cdots \circ F_n(x)$ in Lemma 30 may be identified with products of random matrices, i.e.

$$\begin{pmatrix} S_n(x) \\ 1 \end{pmatrix} = \begin{pmatrix} U_1 & 1 \\ 0 & 1 \end{pmatrix} \cdots \begin{pmatrix} U_n & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ 1 \end{pmatrix} \qquad \forall x \in \mathbb{R}, \quad n \in \mathbb{N}.$$

We choose the current formulation because many results in this section have natural extensions to nonlinear random functions F_n with similar assumptions on their Lipschitz constants.

Proof. By definition,

$$S_n(1) = 1 + \sum_{j=1}^{n}\prod_{k=1}^{j} U_k$$

and it is immediate that $S_{n+1}(1) \ge S_n(1) \ge 1$ for all $n \in \mathbb{N}$. Therefore, the almost sure limit in (83) exists by monotone convergence and we have $\mathbb{P}(S_\infty \ge 1) = 1$. Moreover,

$$\mathbb{E}[S_\infty] = 1 + \sum_{j=1}^{\infty}\mathbb{E}\left[\prod_{k=1}^{j} U_k\right] = 1 + \sum_{j=1}^{\infty}\mathbb{E}[U]^j = \sum_{j=0}^{\infty} 2^{-j} = 2 \tag{85}$$

which implies $\mathbb{P}(S_\infty < \infty) = 1$. Thus we have verified (i).

Now suppose $\theta \in (1, e)$, and $(V_n)_n$ is a sequence of random variables such that $\theta^{-n}|V_n|$ converges almost surely to 0 as $n \to \infty$. Since each of the functions F_n is linear with Lipschitz constant $\|F_n\|_{\mathrm{Lip}} = U_n$, we have

$$\begin{aligned}|S_n(V_n) - S_n(1)| &= |F_1 \circ F_2 \circ \cdots \circ F_n(V_n) - F_1 \circ F_2 \circ \cdots \circ F_n(1)| \\ &= \left[\prod_{j=1}^{n} \|F_j\|_{\mathrm{Lip}}\right] |V_n - 1| = \left[\prod_{j=1}^{n} U_n\right] |V_n - 1| \\ &\le \exp\left(\sum_{j=1}^{n} \log U_j\right)(1 + |V_n|).\end{aligned} \tag{86}$$

Since $\mathbb{E}[\log U_i] = \int_0^1 \log u \,\mathrm{d}u = -1$, the strong law of large numbers gives

$$\frac{1}{n}\sum_{j=1}^{n} \log U_j \xrightarrow[n\to\infty]{a.s.} -1.$$

In particular, if we choose $\varepsilon \in (0, 1 - \log\theta)$, then almost surely there exists some (random) $n_0 = n_0(\varepsilon) \in \mathbb{N}$ such that

$$\frac{1}{n}\sum_{j=1}^{n} \log U_j \le -1 + \varepsilon \qquad \text{for all } n \ge n_0.$$

Substituting this into (86), we obtain

$$|S_n(V_n) - S_n(1)| \le (\theta / \mathrm{e}^{1-\varepsilon})^n \left[\theta^{-n}(1 + |V_n|)\right] = O\big(\theta^{-n}(1 + |V_n|)\big).$$

The assumption on V_n implies that $S_n(V_n) - S_n(1)$ converges to 0 almost surely. Since $S_n(1)$ converges to S_∞ almost surely by (i), we conclude that (ii) holds, i.e. $S_n(V_n)$ also converges to S_∞ almost surely. □

Corollary 32. *For the integral I_n as in* (81)*, we have*

$$I_n = \mathbb{E}\left[\frac{1}{S_{n-1}(1 + c_n U_n)}\right].$$

In particular, for any sequence $(c_n)_n$ satisfying $0 \le c_n = o(\theta^n)$ for some $\theta \in (1, e)$,

$$\lim_{n\to\infty} I_n = \mathbb{E}[S_\infty^{-1}] =: I_\infty.$$

Proof. The probabilistic representation of the iterated integral I_n follows immediately by construction. If we let $V_{n-1} := 1 + c_n U_n$, then $\theta^{-n} V_n$ converges to 0 almost surely as $n \to \infty$, and by Lemma 30 we also obtain that $S_{n-1}(V_{n-1})$ converges almost surely to S_∞. Since

$$\frac{1}{S_{n-1}(V_{n-1})} \le \frac{1}{S_{n-1}(1)} \le 1,$$

we conclude by dominated convergence that

$$\lim_{n\to\infty} I_n = \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{S_{n-1}(V_{n-1})}\right] = \mathbb{E}[S_\infty^{-1}].$$ □

The next step is a simple but crucial characterisation of the distribution of S_∞.

Lemma 33. *Let U, X be two independent random variables such that $U \sim$ Uniform$[0,1]$ and*

$$X \overset{(d)}{=} 1 + UX. \tag{87}$$

If $\mathbb{P}(|X| < \infty) = 1$, then $X \overset{(d)}{=} S_\infty$.

Remark 34. While Lemma 33 is conveniently formulated in terms of random variables, the statement ultimately concerns the law of X only and does not require the knowledge of the underlying probability space. For instance, the distributional equality (87) can be reformulated as

$$\mathbb{E}[g(X)] = \int_0^1 \mathbb{E}[g(1+uX)]\,\mathrm{d}u$$

for all suitable test functions g (provided both sides are well defined), and the conclusion of the lemma says that we necessarily have $\mathbb{E}[g(X)] = \mathbb{E}[g(S_\infty)]$, or equivalently $\mathbb{P}(X \le x) = \mathbb{P}(S_\infty \le x)$ for any $x \in \mathbb{R}$.

Proof. Without loss of generality (by Remark 34), assume that X is defined on the same probability space in Lemma 30 such that X is independent of all the uniformly distributed random variables $U, U_1, U_2, U_3, \dots$; our goal is to verify the following two claims:

(1) The random variable $S_\infty := \lim_{n\to\infty} S_n(1)$ satisfies the distributional equality (87), i.e. $S_\infty \overset{(d)}{=} 1 + US_\infty$.

This is straightforward by a quick reordering of the underlying i.i.d. random variables/iterated maps. Indeed,

$$\begin{aligned} S_n(1) = F_1 \circ \cdots \circ F_n(1) &\overset{(d)}{=} F_n \circ F_1 \circ F_2 \circ \cdots \circ F_{n-1}(1) \\ &= 1 + U_n S_{n-1}(1) \xrightarrow[n\to\infty]{(d)} 1 + US_\infty. \end{aligned}$$

(2) If X satisfies $\mathbb{P}(|X| < \infty)$ and (87), then $X \overset{(d)}{=} S_\infty$.

To establish this claim, observe that for any $n \in \mathbb{N}$ we have

$$X \overset{(d)}{=} F_1(X) \overset{(d)}{=} \cdots \overset{(d)}{=} F_1 \circ F_2 \circ \cdots \circ F_n(X) = S_n(X).$$

Since $\mathbb{P}(|X| < \infty)$ and in particular $2^{-n}|X| \xrightarrow[n\to\infty]{a.s.} 0$, we apply Lemma 30 with $V_n = X$ and obtain $S_n(X) \xrightarrow[n\to\infty]{a.s.} S_\infty$. In other words,

$$X \overset{(d)}{=} S_\infty = 1 + \sum_{j=1}^{\infty} \prod_{k=1}^{j} U_k,$$

which concludes the proof. □

5.2. *Proof of Theorem 28*

The recursive distributional equation (87) is a very convenient tool that helps us control the rate of convergence of $S_n(\cdot)$ and extract information about the statistical behaviour of S_∞ at the same time. We first explain how to estimate the difference between $I_n(c_n)$ and its limit I_∞.

Lemma 35. *For any nonnegative sequence* $(c_n)_n$*, we have*

$$|I_n(c_n) - I_\infty| \le 2^{1-n} + \mathbb{E}\left[\min\left(1, c_n \prod_{j\le n} U_j\right)\right] \le 2^{-n}(2 + c_n). \tag{88}$$

Proof. Recall that $V_{n-1} := 1 + c_n U_n$ and

$$I_n(c_n) = \mathbb{E}\left[\frac{1}{S_{n-1}(V_{n-1})}\right].$$

On the other hand, if we introduce a new random variable $T \overset{(d)}{=} S_\infty$ that is independent of all of the U_j's, we see that

$$S_{n-1}(T) \overset{(d)}{=} S_\infty \qquad \text{and hence} \qquad I_\infty = \mathbb{E}[S_\infty^{-1}] = \mathbb{E}\left[\frac{1}{S_{n-1}(T)}\right]$$

by the distributional fixed point equation (87).

Since $S_{n-1}(\cdot)$ is linear with Lipschitz constant $\|S_{n-1}\|_{\mathrm{Lip}} = \prod_{j\le n-1} U_j$, we have

$$\begin{aligned} |I_n(c_n) - I_\infty| &= \left|\mathbb{E}\left[\frac{1}{S_{n-1}(V_{n-1})} - \frac{1}{S_{n-1}(T)}\right]\right| \\ &\le \mathbb{E}\left[\frac{\|S_{n-1}\|_{\mathrm{Lip}}|V_{n-1}-T|}{S_{n-1}(V_{n-1})S_{n-1}(T)}\right] \\ &\le \mathbb{E}\left[\frac{\|S_{n-1}\|_{\mathrm{Lip}}|T-1|}{S_{n-1}(T)}\right] + \mathbb{E}\left[\frac{\|S_{n-1}\|_{\mathrm{Lip}}|V_{n-1}-1|}{S_{n-1}(V_{n-1})}\right]. \end{aligned} \tag{89}$$

Since $\mathbb{E}[U_j] = 1/2$, $\mathbb{E}[T] = \mathbb{E}[S_\infty] = 2$ by (85) and $\mathbb{P}(T \ge 1) = \mathbb{P}(S_{n-1}(T) \ge 1) = \mathbb{P}(S_\infty \ge 1) = 1$, the first term on the right-hand side of (89) satisfies

$$\begin{aligned} \mathbb{E}\left[\frac{\|S_{n-1}\|_{\mathrm{Lip}}|T-1|}{S_{n-1}(T)}\right] &\le \mathbb{E}\left[\|S_{n-1}\|_{\mathrm{Lip}}|T-1|\right] \\ &= \mathbb{E}\left[\left(\prod_{j\le n-1} U_j\right)|T-1|\right] = \mathbb{E}[T-1]\prod_{j\le n-1}\mathbb{E}\left[U_j\right] = 2^{1-n}. \end{aligned}$$

Next, observe that $V_{n-1} - 1 = c_n U_n$ and $S_{n-1}(x) \ge 1$ for any $x \ge 0$. This means

$$\frac{\|S_{n-1}\|_{\mathrm{Lip}}|V_{n-1}-1|}{S_{n-1}(V_{n-1})} \le c_n \prod_{j\le n} U_j.$$

On the other hand, $S_{n-1}(V_{n-1}) \ge 1 + V_{n-1}\|S_{n-1}\|_{\mathrm{Lip}} \ge \|S_{n-1}\|_{\mathrm{Lip}}|V_{n-1}-1|$, which leads to a slightly improved bound

$$\frac{\|S_{n-1}\|_{\mathrm{Lip}}|V_{n-1}-1|}{S_{n-1}(V_{n-1})} \le \min\left(1, c_n \prod_{j\le n} U_j\right).$$

Taking expectation both sides and plugging this back into (89) yields the first inequality in (88), and the second inequality in (88) follows from $\mathbb{E}[c_n \prod_{j\le n} U_j] = 2^{-n} c_n$. □

It remains to show that the value of I_∞ equals $\mathrm{e}^{-\gamma}$. This will be achieved using the recursive distributional equation (87) with the help of Laplace transform $\phi(t) := \mathbb{E}[\mathrm{e}^{-tS_\infty}]$, which is intrinsically related to our problem because

$$I_\infty = \mathbb{E}[S_\infty^{-1}] = \mathbb{E}\left[\int_0^\infty \mathrm{e}^{-tS_\infty}\,\mathrm{d}t\right] = \int_0^\infty \phi(t)\,\mathrm{d}t \tag{90}$$

by Fubini's theorem. Let us first highlight that:

Lemma 36. *The Laplace transform $\phi(t) := \mathbb{E}[\mathrm{e}^{-tS_\infty}]$ satisfies*

$$t\,\mathrm{e}^t \phi(t) = \int_0^t \phi(v)\,\mathrm{d}v, \qquad t \ge 0. \tag{91}$$

In particular,

$$I_\infty = \lim_{t\to\infty} t\,\mathrm{e}^t \phi(t). \tag{92}$$

Proof. From the recursive distribution equation (87), we have

$$\begin{aligned} \phi(t) := \mathbb{E}[\mathrm{e}^{-tS_\infty}] &= \mathbb{E}[\mathrm{e}^{-t(1+US_\infty)}] \\ &= \mathrm{e}^{-t}\int_0^1 \mathbb{E}[\mathrm{e}^{-tuS_\infty}]\,\mathrm{d}u = \frac{\mathrm{e}^{-t}}{t}\int_0^t \mathbb{E}[\mathrm{e}^{-vS_\infty}]\,\mathrm{d}v. \end{aligned}$$

Hence $t\,\mathrm{e}^t\phi(t) = \int_0^t \phi(v)\mathrm{d}v$, as claimed. In particular $\lim_{t\to\infty} t\,\mathrm{e}^t\phi(t) = \int_0^\infty \phi(v)\,\mathrm{d}v = I_\infty$. □

Proof of Theorem 28. Differentiating the equality (91) yields $t\,\mathrm{e}^t\,\phi'(t)+(t\,\mathrm{e}^t)'\phi(t)=\phi(t)$, which may be rewritten as

$$\frac{\phi'(t)}{\phi(t)}=\frac{1-(t\,\mathrm{e}^t)'}{t\,\mathrm{e}^t}=\frac{\mathrm{e}^{-t}-1}{t}-1.$$

Since $\phi(0)=1$, we then obtain

$$\begin{aligned}\log\phi(x)=\log\phi(x)-\log\phi(0)&=\int_0^x\frac{\mathrm{d}}{\mathrm{d}u}\big[\log\phi(u)\big]\,\mathrm{d}u=\int_0^x\frac{\phi'}{\phi}(u)\,\mathrm{d}u\\&=\int_0^x\left(\frac{\mathrm{e}^{-u}-1}{u}-1\right)\mathrm{d}u=\int_0^x\left(\mathrm{e}^{-u}-1\right)\frac{\mathrm{d}u}{u}-x\\&=\big[(\mathrm{e}^{-u}-1)\log u\big]_0^x+\int_0^x\mathrm{e}^{-u}\log u\,\mathrm{d}u-x.\end{aligned}$$

From Euler's identity for γ [9, Equation (2.2.8)],

$$\gamma=-\int_0^\infty\mathrm{e}^{-u}\log u\,\mathrm{d}u,$$

we see as $x\to\infty$,

$$\log\phi(x)=-\log x-x-\gamma+o(1).$$

Substituting this into (92), we obtain $I_\infty=\lim_{x\to\infty}x\,\mathrm{e}^x\,\phi(x)=\mathrm{e}^{-\gamma}$. Combining this with Lemma 35, we conclude that $I_n=\mathrm{e}^{-\gamma}+O(2^{-n}(1+c_n))$. □

Acknowledgments

Collaboration for this paper stems in part from a problem session during the '50 Years of Number Theory and Random Matrix Theory Conference' at the Institute for Advanced Study (IAS) in Princeton. The authors would like to thank the conference organizers, as well as Paul Kinlaw and the anonymous referee for careful feedback.

Declaration of interests

The authors do not work for, advise, own shares in, or receive funds from any organization that could benefit from this article, and have declared no affiliations other than their research organizations.

References

[1] W. D. Banks and G. Martin, "Optimal primitive sets with restricted primes", *Integers* **13** (2013), article no. A69 (10 pages).

[2] J. Bayless, P. Kinlaw and J. D. Lichtman, "Higher Mertens constants for almost primes", *J. Number Theory* **234** (2022), pp. 448–475.

[3] J.-M.-F. Chamayou, "A probabilistic approach to a differential-difference equation arising in analytic number theory", *Math. Comput.* **27** (1973), pp. 197–203.

[4] T. H. Chan, J. D. Lichtman and C. Pomerance, "A generalization of primitive sets and a conjecture of Erdős", *Discrete Anal.* **2020** (2020), article no. 16 (13 pages).

[5] T. H. Chan, J. D. Lichtman and C. Pomerance, "On the critical exponent for k-primitive sets", *Combinatorica* **42** (2022), no. 5, pp. 729–747.

[6] P. Erdős, "Note on Sequences of Integers No One of Which is Divisible By Any Other", *J. Lond. Math. Soc.* **10** (1935), no. 2, pp. 126–128.

[7] P. Erdős and A. Sárközy, "On the number of prime factors in integers", *Acta Sci. Math.* **42** (1980), no. 3-4, pp. 237–246.

[8] H.-K. Hwang and T.-H. Tsai, "Quickselect and the Dickman function", *Comb. Probab. Comput.* **11** (2002), no. 4, pp. 353–371.

[9] J. C. Lagarias, "Euler's constant: Euler's work and modern developments", *Bull. Am. Math. Soc.* **50** (2013), no. 4, pp. 527–628.

[10] J. D. Lichtman, "Almost primes and the Banks–Martin conjecture", *J. Number Theory* **211** (2020), pp. 513–529.

[11] J. D. Lichtman, "Mertens' prime product formula, dissected", *Integers* **21A** (2021), no. Ron Graham Memorial Volume, article no. A17 (15 pages).

[12] J. D. Lichtman, "Translated sums of primitive sets", *C. R. Math. Acad. Sci. Paris* **360** (2022), pp. 409–414.

[13] J. D. Lichtman, "A proof of the Erdős primitive set conjecture", *Forum Math. Pi* **11** (2023), article no. e18 (21 pages).

[14] S. A. Molchanov and V. A. Panov, "The Dickman–Goncharov distribution", *Russ. Math. Surv.* **75** (2020), no. 6, pp. 1089–1132.

[15] M. D. Penrose and A. R. Wade, "Random minimal directed spanning trees and Dickman-type distributions", *Adv. Appl. Probab.* **36** (2004), no. 3, pp. 691–714.

[16] Z. X. Zhang, "On a problem of Erdős concerning primitive sequences", *Math. Comput.* **60** (1993), no. 202, pp. 827–834.