Numbers free of large prime factors

Jared Duker Lichtman

Dartmouth College
Advisor: Carl Pomerance

A thesis submitted for the degree of
Artium Baccalaurei

November 2017

Dedicated to Jonina Duker

תייבל בת רחל לאה

ושלמה יהושע

Acknowledgements

I must first thank Carl Pomerance for bringing me into the world of mathematics research. The sense of agency I gained from creating new knowledge was made possible by his guidance. Carl has been so generous with his time, teaching me the tools of the trade, and I am deeply thankful. It speaks volumes that he was willing to mentor me even after retirement. I have benefited from his patience and wisdom, and collaborating with him these past two years has been a joy. I am also grateful to Steven J. Miller for his guidance and advice from SMALL and beyond.

I am moved by the level of support I have received from the mathematics department these four years. I thank Vladimir Chernov, Tom Shemanske, and Peter Winkler, who encouraged me from early on at Dartmouth. I thank Pierre Clare, Erik van Erp, Carolyn Gordon, Mits Kobayashi, John Voight, and Dana Williams for their encouragement as I advanced in the major. I would also like to express my gratitude to the Byrne family for their generous contributions to Dartmouth mathematics. I thank Peter Doyle, Sergi Elizalde, Rosa Orellana, and Scott Pauls for their involvement in the Byrne Scholars Program.

I also thank James Schafer, Will Rose, and Katey Koblenz for being exceptional teachers, who in no small part influenced me towards a career in mathematics.

Most importantly, I thank my parents who, at great expense, have acted and continue to act in my best interest.

Abstract

This thesis consists of three chapters, each of which tackles a separate number-theoretic problem and may stand alone as an individual research paper. Nevertheless, these problems are united by a common thread of numbers free of large prime factors, so-called *smooth* numbers. From the vantage point of smooth numbers, these three problems (and their methods of proof) build upon one another in a natural progression of ideas. Chapters 1 and 2 appear as [30, 31] in Mathematics of Computation and the Journal of Number Theory, respectively, joint work with Carl Pomerance.

In Chapter 1, we investigate the probability that a random odd composite number passes a random Fermat primality test, improving on earlier estimates in moderate ranges. For example, with random numbers to 2^{200}, our results improve on prior estimates by close to 3 orders of magnitude.

In Chapter 2, we investigate the distribution of smooth numbers. There is a large literature on the asymptotic distribution of smooth numbers. But there is very little known about this distribution that is numerically explicit. We follow the general plan for the saddle point argument of Hildebrand and Tenenbaum, giving explicit and fairly tight intervals in which the true count lies. We give two numerical examples of our method, and with the larger one, our interval is so tight we can exclude the famous Dickman–de Bruijn asymptotic estimate as too small and the Hildebrand–Tenenbaum main term as too large.

In Chapter 3, we investigate the reciprocal sum of so-called *primitive nondeficient numbers*, or pnds. Erdős showed that the reciprocal sum of pnds converges, which he used to prove that nondeficient numbers have a natural density. However no one has investigated the value of this series! We provide the first known bound by showing the reciprocal sum of pnds is at most 18.6.

Contents

List of Figures

Chapter 1

Improved error bounds for the Fermat primality test on random inputs

1.1 Introduction

Part of the basic landscape in elementary number theory is the Fermat congruence: If n is a prime and $1 \le b \le n-1$, then

$$b^{n-1} \equiv 1 \pmod{n}. \tag{1.1}$$

It is attractive in its simplicity and ease of verification: using fast arithmetic subroutines, (1.1) may be checked in $(\log n)^{2+o(1)}$ bit operations. Further, its converse (apparently) seldom lies. In practice, if one has a large random number n that satisfies (1.1) for a random choice for b, then almost certainly n is prime. To be sure, there are infinitely many composites (the Carmichael numbers) that satisfy (1.1) for all b coprime to n, see [1]. And in [2] it is shown that there are infinitely many Carmichael numbers n such that (1.1) holds for $(1-o(1))n$ choices for b in $[1, n-1]$. (Specifically, for each fixed k there are infinitely many Carmichael numbers n such that the probability a random b in $[1, n-1]$ has $(b, n) > 1$ is less than $1/\log^k n$.) However, Carmichael numbers are rare, and if a number n is chosen at random, it is unlikely to be one.

We say n is a *probable prime to the base* b if (1.1) holds. A probable prime is either prime or composite, but the terminology certainly suggests that it is probably prime! Specifically, let $P(x)$ denote the probability that an integer n is composite given that

(i) n is chosen at random with $1 < n \le x$, n odd,

(ii) b is chosen at random with $1 < b < n-1$, and

(iii) n is a probable prime to the base b.

It is known that if x is sufficiently large, then $P(x)$ is small. Indeed, Erdős and Pomerance [18, Theorem 2.2] proved that

$$P(x) \leq \exp(-(1+o(1))\log x \log\log\log x / \log\log x) \tag{1.2}$$

as $x \to \infty$. In particular, $\lim P(x) = 0$. Kim and Pomerance [29] replaced the asymptotic inequality of (1.2) with the weaker, but explicit, inequality

$$P(x) \leq (\log x)^{-197} \quad \text{for } x \geq 10^{10^5}$$

and gave numerical bounds on $P(x)$ for $10^{60} \leq x < 10^{10^5}$. In this paper we simplify the argument in [29] and obtain better upper bounds on $P(x)$ for $10^{60} \leq x \leq 10^{90}$, as seen in Figure 1.1. In particular, at the start of this range, our bound is over 700 times smaller.

Figure 1.1: New bounds on $P(x)$.

x	Bound on $P(x)$ in [29]	New bound on $P(x)$
10^{60}	$7.16E{-2}$	$1.002E{-4}$
10^{70}	$2.87E{-3}$	$1.538E{-5}$
10^{80}	$8.46E{-5}$	$2.503E{-6}$
10^{90}	$1.70E{-6}$	$4.304E{-7}$
10^{100}	$2.77E{-8}$	$7.798E{-8}$

The notation aEm means $a \times 10^m$.

With these methods, we also obtain new nontrivial bounds for $2^{40} \leq x < 10^{60}$, values of x smaller than the methods in [29] could handle. These results are included in Figure 1.2.

Figure 1.2: Upper bound on $P(2^k)$.

k	$P(2^k) \leq$	k	$P(2^k) \leq$	k	$P(2^k) \leq$
40	$4.306E{-}1$	140	$3.265E{-}3$	240	$1.017E{-}5$
50	$2.904E{-}1$	150	$1.799E{-}3$	250	$5.876E{-}6$
60	$1.848E{-}1$	160	$9.932E{-}4$	260	$3.412E{-}6$
70	$1.127E{-}1$	170	$5.505E{-}4$	270	$1.992E{-}6$
80	$6.728E{-}2$	180	$3.064E{-}4$	280	$1.169E{-}6$
90	$4.017E{-}2$	190	$1.714E{-}4$	290	$6.888E{-}7$
100	$2.388E{-}2$	200	$9.634E{-}5$	300	$4.080E{-}7$
110	$1.435E{-}2$	210	$5.447E{-}5$	310	$2.428E{-}7$
120	$8.612E{-}3$	220	$3.097E{-}5$	320	$1.451E{-}7$
130	$5.229E{-}3$	230	$1.770E{-}5$	330	$8.713E{-}8$

We compute the exact values of $P(x)$ for $x = 2^k$ with $3 \leq k \leq 36$. Additionally, we estimate $P(x)$ for $x = 2^k$ with $30 \leq k \leq 50$, using random sampling. Calibrating these estimates against the true values for $30 \leq k \leq 36$ suggest that the estimates are fairly close to the true values for $37 \leq k \leq 50$, and almost certainly within an order of magnitude from the truth.

A number n is called L-smooth if all of its prime factors are bounded above by L. The method of [29] first computes the contribution to $P(x)$ from numbers that are not L-smooth (for an appropriate choice for L), and then enters a complicated argument based on the asymptotic method of [18] for the contribution of the L-smooth numbers. In addition to small improvements made in the non-L-smooth case, our principal new idea is to use merely that there are few L-smooth numbers. For this we use the upper bound method pioneered by Rankin in [41] for this problem, obtaining numerically explicit upper bounds on sums over L-smooth numbers, c.f. equation (1.11) and Remark 1.3.4. These upper bounds should prove useful in other contexts.

One possible way to gain an improvement is to replace the Fermat test with the strong probable prime test of Selfridge. Also known as the Miller–Rabin test, it is just as simple to perform and it returns fewer false positives. To describe this test, let $n > 1$ be an odd number. First one computes s, t with $n - 1 = 2^s t$ and t odd. Next, one chooses a number b, $1 \leq b \leq n - 1$. The number n passes the test (and is called a *strong probable prime to the base* b) if either

$$b^t \equiv 1 \pmod{n} \quad \text{or} \quad b^{2^i t} \equiv -1 \pmod{n} \quad \text{for some } i < s. \tag{1.3}$$

Every odd prime must pass this test. Moreover, Monier [33] and Rabin [40] have shown that if n is an odd composite, then the probability that it is a strong probable prime to a random base b in $[1, n-1]$ is less than $\frac{1}{4}$.

Let $P_1(x)$ denote the same probability as $P(x)$, except that (iii) is replaced by

(iii)′ n is a strong probable prime to the base b.

Based on the Monier-Rabin theorem, one might assume that $P_1(x) \leq \frac{1}{4}$, but as noted in [6], this reasoning is flawed. However, in [10] and [14], something similar to $P_1(x) \leq \frac{1}{4}$ is shown. Namely, if $P_1'(2^k)$ is the analogous probability for odd k-bit integers, it is shown in [10], [14] that $P_1'(2^k) \leq \frac{1}{4}$ for all $k \geq 3$. We show below how our estimates can be used to numerically bound $P_1(x)$. In particular, the results here improve on the estimates of [14] up to 2^{300}.

Notation

We have (a, b), $[a, b]$ as the greatest common divisor, least common multiple of the positive integers a, b, respectively. We use p and q to denote prime numbers, and p_i to denote the ith prime. For $n > 1$, we let $P^+(n)$ denote the largest prime factor of n. Let φ denote Euler's function, λ the Carmichael universal exponent function, ζ the Riemann zeta-function, $\mathrm{Li}(x) = \int_2^x \frac{\mathrm{d}t}{\log t}$, and $\vartheta(x) = \sum_{p \leq x} \log p$. In many instances, we take a sum over certain subsets of odd composite integers, in which cases we use $\sum_n'$ to denote $\sum_{\substack{n \text{ odd,} \\ \text{composite}}}$.

1.2 Preliminary lemmas

In this section, we prove some preliminary lemmas which are needed for the rest of the paper, and which may be of interest in their own right.

Lemma 1.2.1. *Given real numbers a, b and a nonnegative, decreasing function f on the interval $[a, b]$, we have that*

$$\int_{\lceil a \rceil}^{b} f(t)\, \mathrm{d}t \leq \sum_{a \leq n \leq b} f(n) \leq f(a) + \int_a^b f(t)\, \mathrm{d}t.$$

The proof is clear. Note that since $\sum_{a < n \leq b} f(n) \leq \sum_{a \leq n \leq b} f(n)$, we may apply the upper bound for the sum on the half open interval.

Lemma 1.2.2. *For $x \geq 2$, we have that*

$$\frac{x}{\zeta(2)} - \log x \leq \sum_{n \leq x} \frac{\varphi(n)}{n} \leq \frac{x}{\zeta(2)} + \log x.$$

Proof. The result holds for $2 \leq x < 18$, so assume $x \geq 18$. We have that

$$\begin{aligned} \sum_{n \leq x} \frac{\varphi(n)}{n} &= \sum_{n \leq x} \sum_{d|n} \frac{\mu(d)}{d} = \sum_{d \leq x} \frac{\mu(d)}{d} \sum_{n \leq x/d} 1 = \sum_{d \leq x} \frac{\mu(d)}{d} \left\lfloor \frac{x}{d} \right\rfloor \\ &= x \sum_{d \leq x} \frac{\mu(d)}{d^2} - \sum_{d \leq x} \frac{\mu(d)}{d} \left\{ \frac{x}{d} \right\} \\ &= \frac{x}{\zeta(2)} - x \sum_{d > x} \frac{\mu(d)}{d^2} - \sum_{d \leq x} \frac{\mu(d)}{d} \left\{ \frac{x}{d} \right\}, \end{aligned} \tag{1.4}$$

where $\{ \}$ denotes the frational part. By Lemma 1.2.1,

$$\begin{aligned} \sum_{d > x} \frac{\mu(d)}{d^2} &\leq \sum_{d > x} \frac{1}{d^2} \leq \frac{1}{x^2} + \int_x^\infty \frac{dt}{t^2} = \frac{1}{x^2} + \frac{1}{x}, \\ \sum_{d > x} \frac{\mu(d)}{d^2} &\geq - \sum_{d > x} \frac{1}{d^2} \geq - \frac{1}{x^2} - \frac{1}{x}. \end{aligned} \tag{1.5}$$

Since $\sum_{1 < d \leq 18,\, \mu(d) \neq -1} \frac{1}{d} = \frac{367}{336} > 1.09$, we have

$$- \sum_{d \leq x} \frac{\mu(d)}{d} \left\{ \frac{x}{d} \right\} \leq \sum_{\substack{d \leq x \\ \mu(d) = -1}} \frac{1}{d} \leq \sum_{1 < d \leq x} \frac{1}{d} - \sum_{\substack{1 < d \leq 18 \\ \mu(d) \neq -1}} \frac{1}{d} < \log x - 1.09.$$

Substituting this and (1.5) back into (1.4) gives

$$\sum_{n \leq x} \frac{\varphi(n)}{n} \leq \frac{x}{\zeta(2)} + \frac{1}{x} + 1 + \log x - 1.09 < \frac{x}{\zeta(2)} + \log x.$$

Similarly, direct computation shows that

$$\sum_{d \leq x} \frac{\mu(d)}{d} \left\{ \frac{x}{d} \right\} \leq \sum_{\substack{d \leq x \\ \mu(d) = 1}} \frac{1}{d} \leq \sum_{1 < d \leq x} \frac{1}{d} - \sum_{\substack{1 < d \leq 4 \\ \mu(d) \neq 1}} \frac{1}{d} < \log x - \frac{13}{12}.$$

and thus

$$\sum_{n \leq x} \frac{\varphi(n)}{n} \geq \frac{x}{\zeta(2)} - \frac{1}{x} - 1 - \log x + \frac{13}{12} > \frac{x}{\zeta(2)} - \log x.$$

□

Lemma 1.2.3. *For $x \geq 1$, we have that*

$$\frac{\log x}{\zeta(2)} + 1 - \frac{\log 2}{\zeta(2)} < \sum_{n \leq x} \frac{\varphi(n)}{n^2} \leq \frac{\log x}{\zeta(2)} + 1.$$

Proof. The inequalities are easily verified for $x < 40$, so assume $x \geq 40$. Partial summation gives

$$\sum_{n\leq x}\frac{\varphi(n)}{n^2}=\sum_{n\leq 39}\frac{\varphi(n)}{n^2}+\frac{1}{x}\sum_{n\leq x}\frac{\varphi(n)}{n}-\frac{1}{40}\sum_{n\leq 39}\frac{\varphi(n)}{n}+\int_{40}^{x}\frac{1}{t^2}\sum_{n\leq t}\frac{\varphi(n)}{n}\,\mathrm{d}t.$$

Evaluating the two sums to 39 and using the upper and lower bounds in Lemma 1.2.2 for the sums to x and t, we obtain the stronger result,

$$\frac{\log x}{\zeta(2)}+0.58<\sum_{n\leq x}\frac{\varphi(n)}{n^2}<\frac{\log x}{\zeta(2)}+0.82.$$

Note that the upper bound in the lemma is tight at $x = 1$ and the lower bound cannot be improved as $x \to 2^-$. □

Lemma 1.2.4. *If* $2 \leq y < x$ *and* $0 < c < 1$, *then*

$$\sum_{y<p\leq x} p^{-c} < f(x,y),$$

where

$$f(x,y) := (1+2.3\cdot 10^{-8})\left(\mathrm{Li}(x^{1-c})-\mathrm{Li}(y^{1-c})+\frac{y^{1-c}}{\log y}\right)-\vartheta(y)\frac{y^{-c}}{\log y}.$$

Proof. We use the inequalities

$$\vartheta(x) < x \quad (0 < x \leq 10^{19}), \qquad |x-\vartheta(x)| < \epsilon x \quad (x > 10^{19}), \tag{1.6}$$

where $\epsilon = 2.3 \times 10^{-8}$, see [11], [12], improving on recent work in [38] (also see Proposition 2.2.1 in Chapter 2). Let $f(t) = 1/(t^c \log t)$. By partial summation,

$$\sum_{y<p\leq x} p^{-c}=\sum_{y<p\leq x} f(p)\log p=\vartheta(x)f(x)-\vartheta(y)f(y)-\int_{y}^{x}\vartheta(t)f'(t)\,\mathrm{d}t.$$

Note that (1.6) implies that $\vartheta(t) < (1+\epsilon)t$ for all $t > 0$. Since $f'(t) < 0$ for $t \geq 2$, we have

$$\begin{aligned}\sum_{y<p\leq x} p^{-c} &< (1+\epsilon)xf(x)-(1+\epsilon)\int_{y}^{x} tf'(t)\,\mathrm{d}t-\vartheta(y)f(y)\\ &=(1+\epsilon)\left(\mathrm{Li}(x^{1-c})-\mathrm{Li}(y^{1-c})+yf(y)\right)-\vartheta(y)f(y),\end{aligned}$$

where we have integrated by parts and used that $\int f(t)\,dt = \mathrm{Li}(t^{1-c})$. This completes the proof. □

Lemma 1.2.5. *We have*

$$(i)\ \sum_{n>y}\frac{1}{n^2}<\frac{5}{3y} \qquad \textit{for } y>0,$$

$$(ii)\ \sum_{n\geq y}\frac{1}{n^3}\leq\frac{4(\zeta(3)-1)}{y^2} \qquad \textit{for } y>1.$$

Proof. The first claim is stated and proved in (4.7) in [29]. We proceed similarly for the second claim. When $1<y\leq 2$, we have

$$\sum_{n\geq y}\frac{1}{n^3}=\sum_{n\geq 2}\frac{1}{n^3}=\zeta(3)-1=\frac{4(\zeta(3)-1)}{4}\leq\frac{4(\zeta(3)-1)}{y^2}.$$

When $2<y\leq 3$, direct computation shows that

$$\sum_{n\geq y}\frac{1}{n^3}=\sum_{n\geq 3}\frac{1}{n^3}=\zeta(3)-1-\frac{1}{8}<\frac{4(\zeta(3)-1)}{y^2}.$$

When $3<y\leq 4$, direct computation shows that

$$\sum_{n\geq y}\frac{1}{n^3}=\sum_{n\geq 4}\frac{1}{n^3}=\zeta(3)-1-\frac{1}{8}-\frac{1}{27}<\frac{4(\zeta(3)-1)}{y^2}.$$

When $y>4$, by Lemma 1.2.1, direct computation shows that

$$\sum_{n\geq y}\frac{1}{n^3}\leq\frac{1}{y^3}+\int_y^\infty\frac{\mathrm{d}t}{t^3}=\frac{1}{y^3}+\frac{1}{2y^2}<\frac{4(\zeta(3)-1)}{y^2}.$$

□

1.3 The basic method

Let

$$\mathbf{F}(n)=\{b\in(\mathbb{Z}/n\mathbb{Z})^\times : b^{n-1}=1\}$$

and let $F(n)=\#\mathbf{F}(n)$. If $n>1$ is odd, then $\pm1\in\mathbf{F}(n)$. Thus, for these n, $F(n)-2$ counts the number of integers b, $1<b<n-1$, with $b^{n-1}\equiv 1 \pmod n$. Also note that by Fermat's little theorem, $F(p)=p-1$ for primes p. We thus have for $x\geq 5$,

$$P(x)=\frac{\sum'_{n\leq x}(F(n)-2)}{\sum_{1<n\leq x,\, n\text{ odd}}(F(n)-2)}=\left(1+\frac{\sum_{2<p\leq x}(p-3)}{\sum'_{n\leq x}(F(n)-2)}\right)^{-1}. \tag{1.7}$$

Hence to obtain an upper bound for $P(x)$, we shall be interested in obtaining a lower bound for $\sum_{2<p\leq x}(p-3)$ and an upper bound for $\sum'_{n\leq x}(F(n)-2)$. To this end, we shall prove two theorems.

Theorem 1.3.1. *For $x \geq 2657$, we have*

$$\sum_{2<p\leq x} (p-3) > \frac{x^2}{2\log x - \frac{1}{2}}.$$

Theorem 1.3.2. *Suppose c, L_1, and L are arbitrary real numbers with $0 < c < 1$, $1 < L_1 < L$. Then for any $x > L^2$, we have*

$$\sum_{n\leq x}{}' (F(n) - 2) < x^{c+1} \prod_{2<p\leq L} \left(1 - p^{-c}\right)^{-1} + x^2 B,$$

where

$$B = \frac{1}{4L_1} + \frac{\log L_1}{\zeta(2)}\Big(\frac{1}{2(L-1)} + \frac{1}{x^{1/2}}\Big) + \frac{.5}{L-1} + \frac{.8}{x^{1/2}} + \frac{L_1}{(x^{1/2}-1)^2} + \frac{(1+\log L_1)}{2(x^{1/2}-1)} + \frac{1}{(L-1)^2}\Big(\frac{L_1}{\zeta(2)} + \log L_1\Big).$$

Before proving Theorems 1.3.1 and 1.3.2, we state the main result of the section, which follows from these theorems.

Theorem 1.3.3. *Suppose c, L_1, and L are arbitrary positive real numbers satisfying $0 < c < 1$ and $1 < L_1 < L$. Then for any $x > \max\{L^2, 2657\}$, we have $P(x) \leq 1/(1+z^{-1})$ where*

$$z = \left(B + x^{c-1} \prod_{2<p\leq L} \left(1 - p^{-c}\right)^{-1}\right)\left(2\log x - \tfrac{1}{2}\right),$$

and B is defined as in Theorem 1.3.2.

In principle the prime sum is much larger than the composite sum, so the probability $P(x)$ may be approximately viewed as their quotient. We remark that the prime sum in Theorem 1.3.1 is asymptotically equal to $x^2/(2\log x)$, so the result is close to best possible. Additionally, in the application of Theorem 1.3.2, L and c are used as parameters for smoothness and Rankin's upper bound, respectively.

We now prove Theorem 1.3.1 using (1.6) and the additional inequalities from [11], [12] that

$$\vartheta(x) > x - 2\sqrt{x} \quad (1423 \leq x \leq 10^{19}), \qquad \pi(x) < (1+\epsilon)\mathrm{li}(x) \quad (x \geq 2), \qquad (1.8)$$

where $\mathrm{li}(x) = \int_0^x \mathrm{d}t/\log t$ and $\epsilon = 2.3 \times 10^{-8}$.

Proof of Theorem 1.3.1. Let $A = 1500$. By partial summation,

$$\begin{aligned}\sum_{2<p\le x}(p-3) &= 1 - 3\pi(x) + \sum_{p\le x} p \\ &= 1 - 3\pi(x) + \sum_{2<p\le A}(p-3) + \frac{x\vartheta(x)}{\log x} - \frac{A\vartheta(A)}{\log A} - \int_A^x \vartheta(t)\frac{\log t - 1}{\log^2 t}\,\mathrm{d}t.\end{aligned} \tag{1.9}$$

By (1.8), we have $-3\pi(x) > -3(1+\epsilon)\mathrm{li}(x)$. Suppose that $A \le x \le 10^{19}$. By (1.6), (1.8), we have

$$\begin{aligned}\frac{x\vartheta(x)}{\log x} - \int_A^x \vartheta(t)\frac{\log t - 1}{\log^2 t}\,\mathrm{d}t &> \frac{x^2 - 2x^{3/2}}{\log x} - \int_A^x \frac{t}{\log t} - \frac{t}{\log^2 t}\,\mathrm{d}t \\ &= \mathrm{li}(x^2) - \frac{2x^{3/2}}{\log x} - \mathrm{li}(A^2) + \frac{A^2}{\log A}.\end{aligned}$$

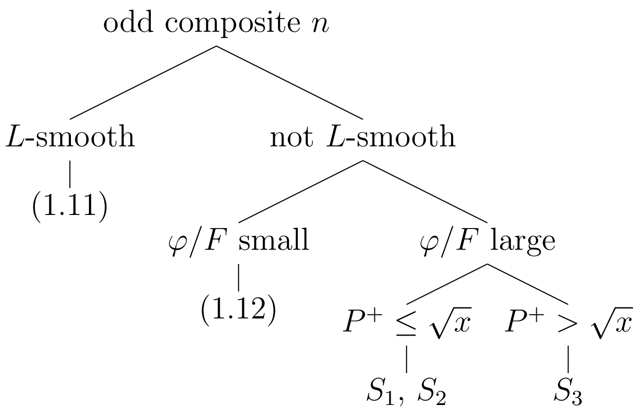
Using these estimates in (1.9), we have

$$\sum_{2<p\le x}(p-3) > \mathrm{li}(x^2) - \frac{2x^{3/2}}{\log x} - 3(1+\epsilon)\mathrm{li}(x) + 5\,875.$$

It is now routine to verify the theorem for $17\,000 \le x \le 10^{19}$. Similar calculations with (1.6), (1.8) establish the theorem for $x > 10^{19}$. A simple check then verifies the theorem in the stated range. □

Proof of Theorem 1.3.2

The bulk of the work is devoted to the proof of Theorem 1.3.2. The basic method is to divide the eligible n into five parts, depending on the largest prime factor $P^+(n)$ as well as the quotient $\varphi(n)/F(n)$, indicating how close n is to being a Carmichael number. We summarize this in the diagram below, which may help guide the reader through the proof.



For any $x > L^2$ with $L > L_1 > 1$, we have

$$\begin{aligned}\sum_{n\le x}{}' (F(n)-2) &= \sum_{\substack{n\le x\\ P^+(n)\le L}}{}' (F(n)-2) + \sum_{\substack{n\le x\\ P^+(n)>L}}{}' (F(n)-2)\\ &\le \sum_{\substack{n\le x\\ P^+(n)\le L\\ n \text{ odd}}} n + \sum_{\substack{n\le x\\ P^+(n)>L}}{}' F(n).\end{aligned} \tag{1.10}$$

For the first term in (1.10), we have for any $0 < c < 1$,

$$\sum_{\substack{n\le x\\ P^+(n)\le L\\ 2\nmid n}} n \le x^{1+c} \sum_{\substack{P^+(n)\le L\\ 2\nmid n}} \frac{1}{n^c} = x^{1+c} \prod_{2<p\le L} \left(1-p^{-c}\right)^{-1}. \tag{1.11}$$

Remark 1.3.4. By approximating the logarithm of the Euler product in (1.11) (with 2 included) using Lemma 1.2.4 and the method of [29], we can write a closed, numerically explicit upper bound on the distribution of L-smooth numbers: If $\frac{1}{2} < c < 1$ and $37 \le L < x$, then

$$\sum_{\substack{n\le x\\ P^+(n)\le L}} 1 \le x^c f_0 \exp(A + f(L,36)),$$

where the notation $f(a,b)$ is defined in Lemma 1.2.4 and

$$f_0 := \prod_{p<37} \left(1-p^{-c}\right)^{-1}, \quad A := \frac{1}{2c-1}\left(\frac{1}{2} + \frac{1}{3(37^c-1)}\right)\left(36^{1-2c} - \frac{1}{2}\cdot 37^{1-2c}\right).$$

There has been a very recent improvement of this Rankin-type upper bound due to Granville and Soundarajan, see Appendix 2.5, that is suitable for numerical estimates. It would be interesting to adapt that method to this paper.

Now we bound the second term in (1.10). Since $\mathbf{F}(n)$ is a subgroup of $(\mathbb{Z}/n\mathbb{Z})^\times$, by Lagrange's Theorem we have $F(n) \mid \varphi(n)$, where φ is Euler's function. Then for each k, it makes sense to define $\mathbf{C}_k(x)$ as the set of odd, composite $n \le x$ such that $F(n) = \varphi(n)/k$. Let $\mathbf{C}'_k(x)$ be the set of $n \in \mathbf{C}_k(x)$ for which $P^+(n) > L$, and let $C'_k(x) = \#\mathbf{C}'_k(x)$. Thus, we have

$$\begin{aligned}\sum_{\substack{n\le x\\ P^+(n)>L}}{}' F(n) &= \sum_{k=1}^{\infty} \sum_{n\in \mathbf{C}'_k(x)} F(n) = \sum_{k=1}^{\infty} \sum_{n\in \mathbf{C}'_k(x)} \frac{\varphi(n)}{k}\\ &= \sum_{k\le L_1} \frac{1}{k} \sum_{n\in \mathbf{C}'_k(x)} \varphi(n) + \sum_{k>L_1} \frac{1}{k} \sum_{n\in \mathbf{C}'_k(x)} \varphi(n)\\ &\le x \sum_{k\le L_1} \frac{C'_k(x)}{k} + \frac{1}{L_1} \sum_{\substack{1<n\le x\\ n \text{ odd}}} (n-2) \le x \sum_{k\le L_1} \frac{C'_k(x)}{k} + \frac{x^2}{4L_1}.\end{aligned} \tag{1.12}$$

It will thus be desirable to obtain an upper bound for $\sum_{k\le L_1}\frac{C'_k(x)}{k}$. We remark that in the case $k>L_1$ we do not use $P^+(n)>L$; this observation will be useful in the next section.

Given a prime $p>L$, $d\mid p-1$, let

$$\mathbf{S}_{p,d}(x)=\{n : n\le x \text{ odd, composite, } n\equiv p \pmod{p(p-1)/d}\}.$$

Let $S_{p,d}(x)=\#\mathbf{S}_{p,d}(x)$. Note that $S_{p,d}\le\frac{xd}{p(p-1)}$ by the Chinese Remainder Theorem. We prove that

$$\bigcup_{k\le L_1}\mathbf{C}'_k(x)\subset\bigcup_{\substack{d\le L_1\\ d\mid p-1\\ L<p\le x}}\mathbf{S}_{p,d}(x).$$

Take n in the left set. Then $p=P^+(n)>L$ and $k=\varphi(n)/F(n)\le L_1$. By Lemma 2.4 in [29], we have $n\equiv 1 \pmod{\frac{p-1}{(k,p-1)}}$. Letting $d=(k,p-1)$, we have that $n\in\mathbf{S}_{p,d}$ (via the Chinese remainder theorem) and $d\le k\le L_1$, so n is in the right set.

Additionally, for a given p,d pair, $S_{p,d}$ counts integers $n=mp$ for which $m\equiv 1 \pmod{\frac{p-1}{d}}$. Write $m=1+u(\frac{p-1}{d})$ for some u. Letting $g=(u,d)$, we have that $m=1+(\frac{u}{g})(\frac{p-1}{d/g})$, so $n\in\mathbf{S}_{p,d/g}$, meaning that n will be counted multiple times if $g>1$. Thus we require $(u,d)=1$. In particular, if d is even, then u is odd. Since $m=1+u(\frac{p-1}{d})$ is odd, we have $u(\frac{p-1}{d})$ even. That is, if d is even then u is odd and $\frac{p-1}{d}$ is even, so $2d\mid p-1$. On the other hand, if d is odd, we of course have $2d\mid p-1$. Thus $2d\mid p-1$ always, and so

$$\begin{aligned}\sum_{k\le L_1}\frac{C'_k(x)}{k}&\le\sum_{d\le L_1}\frac{1}{d}\sum_{\substack{L<p\le x\\ 2d\mid p-1}}\sum_{\substack{u\le\frac{xd}{p(p-1)}\\ (u,d)=1}}1\\ &=\sum_{d\le L_1}\frac{1}{d}\sum_{\substack{L<p\le x^{1/2}\\ 2d\mid p-1}}\sum_{\substack{u\le\frac{xd}{p(p-1)}\\ (u,d)=1}}1+\sum_{d\le L_1}\frac{1}{d}\sum_{\substack{x^{1/2}<p\le x\\ 2d\mid p-1}}\sum_{\substack{u\le\frac{xd}{p(p-1)}\\ (u,d)=1}}1\\ &<\sum_{d\le L_1}\frac{\varphi(d)}{d}\sum_{\substack{L<p\le x^{1/2}\\ 2d\mid p-1}}\Big(\frac{x}{p(p-1)}+1\Big)+\sum_{d\le L_1}\sum_{\substack{x^{1/2}<n\le x\\ 2d\mid n-1}}\frac{x}{n(n-1)}\\ &<S_1+S_2+S_3,\end{aligned}\tag{1.13}$$

where

$$\begin{aligned}S_1&=x\sum_{d\le L_1}\frac{\varphi(d)}{d}\sum_{\substack{L<n\le x^{1/2}\\ 2d\mid n-1}}\frac{1}{(n-1)^2},\quad S_2=\sum_{d\le L_1}\frac{\varphi(d)}{d}\sum_{\substack{1<n\le x^{1/2}\\ 2d\mid n-1}}1,\\ S_3&=\sum_{d\le L_1}\sum_{\substack{x^{1/2}<n\le x\\ 2d\mid n-1}}\frac{x}{(n-1)^2}.\end{aligned}\tag{1.14}$$

It is worth noting that in S_1, S_2, S_3, we have dropped the condition that n be prime. An alternative bound using the condition of primality may be handled as an application of the Brun-Titchmarsh inequality. However, such a method is less effective for the small values of x considered here.

Consider S_1 in (1.14). For a given $d \le L_1$, by Lemma 1.2.1 we have that

$$\begin{aligned} \sum_{\substack{n>L \\ 2d|n-1}} \frac{1}{(n-1)^2} = \sum_{2du+1>L} \frac{1}{4d^2u^2} &\le \frac{1}{(L-1)^2} + \frac{1}{4d^2}\int_{(L-1)/2d}^{\infty} \frac{\mathrm{d}t}{t^2} \\ &= \frac{1}{(L-1)^2} + \frac{1}{2d(L-1)}. \end{aligned} \tag{1.15}$$

Thus, by Lemma 1.2.2 and Lemma 1.2.3,

$$\begin{aligned} S_1 &< x\sum_{d\le L_1} \frac{\varphi(d)}{d}\Big(\frac{1}{(L-1)^2} + \frac{1}{2d(L-1)}\Big) \\ &\le \frac{x}{(L-1)^2}\big(\frac{L_1}{\zeta(2)} + \log L_1\big) + \frac{x}{2(L-1)}\Big(\frac{\log L_1}{\zeta(2)} + 1\Big). \end{aligned} \tag{1.16}$$

By Lemma 1.2.3, S_2 in (1.14) is bounded by

$$S_2 \le \sum_{d\le L_1} \frac{\varphi(d)}{d}\frac{x^{1/2}}{2d} \le x^{1/2}\Big(\frac{\log L_1}{2\zeta(2)} + .8\Big). \tag{1.17}$$

We now consider S_3 in (1.14). For a fixed $d \le L_1$, we have, as in (1.15),

$$\sum_{\substack{x^{1/2}<n\le x \\ 2d|n-1}} \frac{1}{(n-1)^2} \le \frac{1}{(x^{1/2}-1)^2} + \frac{1}{2d(x^{1/2}-1)}.$$

So,

$$S_3 \le x\sum_{d\le L_1} \frac{1}{(x^{1/2}-1)^2} + \frac{1}{2d(x^{1/2}-1)} \le \frac{xL_1}{(x^{1/2}-1)^2} + \frac{x(1+\log L_1)}{2(x^{1/2}-1)}. \tag{1.18}$$

By (1.16), (1.17), and (1.18), we obtain from (1.13) that

$$\sum_{k\le L_1} \frac{C_k'(x)}{k} < x\Big(B - \frac{1}{4L_1}\Big) \tag{1.19}$$

for B as in Theorem 1.3.2. Thus, using (1.19) in (1.12) gives the following result.

Theorem 1.3.5. *Suppose L_1 and L are arbitrary real numbers satisfying $1 < L_1 < L$. Then for any $x > L^2$, we have*

$$\sum_{\substack{n\le x \\ P^+(n)>L}}{}' (F(n)-2) < x^2B.$$

where B is as in Theorem 1.3.2.

Thus, (1.10), (1.11), and Theorem 1.3.5 give us Theorem 1.3.2.

1.4 A refinement of the basic method

We refine the basic method as done analogously in [29], by considering the *two* largest prime factors of n. This refinement provides a modest improvement over Theorem 1.3.3 for x starting around 2^{140}.

Theorem 1.4.1. *Suppose c, L_1, L, and M are arbitrary real numbers satisfying $0 < c < 1$, $10 < L_1 < L$, $2L < M < L^2$. Then for any $x > L^2$, we have*

$$\sum_{n\le x}{}'(F(n)-2) < x^{c+1}\big(1+f(L,M^{1/2})\big)\prod_{2<p\le M^{1/2}}\big(1-p^{-c}\big)^{-1} + x^2(B+C),$$

where f is as in Lemma 1.2.4, B is as in Theorem 1.3.2, and

$$\begin{aligned} C =& \frac{L^2}{2x}(1+\log L_1) + \frac{2(1+\log L_1)^2}{M\log M} \\ &+ \frac{1}{12(M-2L)}(1+\log L)(4+\log L_1)^4\Big(\frac{5}{12}+(\zeta(3)-1)(1+\log L)\Big). \end{aligned}$$

Proof. For each odd, composite $n \le x$, letting P, Q be the two largest prime factors of n (i.e. $P = P^+(n), Q = P^+(n/P)$), we have three possible cases,

(i) $P > L$ or $F(n) < \varphi(n)/L_1$,

(ii) $P \le L$ and $PQ \le M$,

(iii) $P \le L$, $PQ > M$, and $F(n) \ge \varphi(n)/L_1$.

It is worth noting that cases (i) and (ii) are not in general mutually exclusive. We retain Theorem 1.3.5 and the remark following (1.12) to handle case (i). For case (ii), let $0 < c < 1$. When $P \le M^{1/2}$, we have

$$\sum_{\substack{n\le x, 2\nmid n\\ P\le M^{1/2}}} 1 \le x^c \sum_{\substack{2\nmid n\\ P^+(n)\le M^{1/2}}} n^{-c}.$$

Similarly, when $P > M^{1/2}$ we have $Q \le \frac{M}{P} < M^{1/2}$, so

$$\begin{aligned} \sum_{\substack{n\le x, 2\nmid n\\ M^{1/2}<P\le L\\ Q\le M^{1/2}}} 1 &\le \sum_{M^{1/2}<p\le L}\ \sum_{\substack{n\le x/p\\ P^+(n)\le M^{1/2}\\ 2\nmid n}} 1 \le \sum_{M^{1/2}<p\le L}\ \sum_{\substack{P^+(n)\le M^{1/2}\\ 2\nmid n}} \Big(\frac{x}{np}\Big)^c \\ &= x^c \sum_{M^{1/2}<p\le L} p^{-c} \sum_{\substack{2\nmid n\\ P^+(n)\le M^{1/2}}} n^{-c}. \end{aligned}$$

Using Lemma 1.2.4,

$$\begin{aligned}
\sum_{\substack{n\le x, 2\nmid n\\ P\le L\\ PQ\le M}} 1 &\le \sum_{\substack{n\le x, 2\nmid n\\ P\le M^{1/2}}} 1 + \sum_{\substack{n\le x, 2\nmid n\\ M^{1/2}<P\le L\\ Q\le M^{1/2}}} 1 \\
&\le x^c \sum_{\substack{2\nmid n\\ P^+(n)\le M^{1/2}}} n^{-c} + x^c \sum_{M^{1/2}<p\le L} p^{-c} \sum_{\substack{2\nmid n\\ P^+(n)\le M^{1/2}}} n^{-c} \\
&= x^c\Bigg(1+\sum_{M^{1/2}<p\le L} p^{-c}\Bigg) \sum_{\substack{2\nmid n\\ P^+(n)\le M^{1/2}}} n^{-c} \le x^c\big(1+f(L,M^{1/2})\big) \sum_{\substack{2\nmid n\\ P^+(n)\le M^{1/2}}} n^{-c} \\
&= x^c\big(1+f(L,M^{1/2})\big) \prod_{2<p\le M^{1/2}} \big(1-p^{-c}\big)^{-1}.
\end{aligned} \tag{1.20}$$

We now have the following result.

Theorem 1.4.2. *If* $0 < c < 1$, $1 < L < x$, *and* $L < M < L^2$, *then*

$$\sum_{\substack{n\le x, n \text{ odd}\\ P\le L\\ PQ\le M}} n \le x^{c+1}\big(1+f(L,M^{1/2})\big) \prod_{2<p\le M^{1/2}} \big(1-p^{-c}\big)^{-1},$$

where f *is as in Lemma 1.2.4.*

Consider n belonging to case (iii). For each k, let $\mathbf{B}_k(x)$ denote the set of such n with $\varphi(n)/F(n) = k$ and let $B_k(x) = \#\mathbf{B}_k(x)$. Thus,

$$\sideset{}{'}\sum_{n \text{ in case (iii)}} F(n) \le x \sum_{k\le L_1} \frac{B_k(x)}{k}. \tag{1.21}$$

By (2.11) in [18], we have $\lambda(n) \mid k(n-1)$ for all $n \in \mathbf{B}_k(x)$. Since $PQ \mid n$, we have $\lambda(PQ) \mid \lambda(n)$, so n satisfies the set of congruences

$$n \equiv 0 \pmod{PQ}, \quad k(n-1) \equiv 0 \pmod{\lambda(PQ)}. \tag{1.22}$$

Suppose first that $P = Q$. Then $\lambda(PQ) = P(P-1)$, so that (1.22) implies that $P \mid k$. For such a prime P, the number of $n \le x$ with $P^2 \mid n$ is at most $x/P^2 < x/M$. Thus, the contribution for n in this case is at most

$$\frac{x}{M}\sum_{k\le L_1}\frac{x}{k}\sum_{\substack{P\mid k\\ P>M^{1/2}}} 1 < \frac{x^2}{M}\Big(\sum_{k\le L_1}\frac{1}{k}\Big)\frac{\log L_1}{\log M^{1/2}} < \frac{2x^2}{M\log M}(1+\log L_1)^2. \tag{1.23}$$

Now consider the case $P > Q$. The latter congruence in (1.22) is equivalent to

$$n \equiv 1 \pmod{\frac{\lambda(PQ)}{(k, \lambda(PQ))}},$$

from which we also note

$$\left(PQ\,,\ \frac{\lambda(PQ)}{(k, \lambda(PQ))}\right) = 1.$$

Thus for arbitrary fixed primes $p > q$, the Chinese remainder theorem gives that the number of integers $n \le x$ satisfying the system $n \equiv 0 \pmod{pq}$, $k(n-1) \equiv 0 \pmod{\lambda(pq)}$ as in (1.22) is at most

$$1 + \frac{x(k, \lambda(pq))}{pq\lambda(pq)}.$$

Summing over choices for p, q, we have the number of n in this case is at most

$$\sum_{\substack{q<p\le L\\ pq>M}} \left(1 + \frac{x(k, \lambda(pq))}{pq\lambda(pq)}\right) \le \frac{1}{2}L^2 + \frac{1}{2}x \sum_{\substack{p,q\le L\\ pq>M\\ p\neq q}} \frac{(k, [p-1, q-1])}{pq[p-1, q-1]}. \tag{1.24}$$

This is (4.4) in [29] where "L_2" there is our "L". Following the argument in [29] from there, and letting $M' = M - 2L$ and with $u_1, u_2, u_3, u_4, \mu, \nu, \delta$ positive integer variables, we have that

$$\sum_{\substack{q,p\le L\\ pq>M\\ p\neq q}} \frac{(k, [p-1, q-1])}{pq[p-1, q-1]} \le \sum_{\substack{u_1u_2u_3u_4=k\\ (u_1,u_2)=1}} \sum_{\substack{\mu\le L/u_1\\ \nu\le L/u_2}} \sum_{u_1u_2u_3^2\mu\nu\delta^2>M'} \frac{1}{\mu^2\nu^2\delta^3u_1u_2u_3^2}. \tag{1.25}$$

which is the initial inequality of (4.6) in [29] and with a typo corrected (the variable "δ" under the second summation there should be "μ").

We now diverge from the argument in [29], and split up the sum on the right side of (1.25) into two cases, $\delta = 1$ and $\delta \ge 2$. When $\delta = 1$, by Lemma 1.2.5(i) we have

$$\begin{aligned} \sum_{\substack{u_1u_2u_3u_4=k\\ (u_1,u_2)=1}} \sum_{\substack{\mu\le L/u_1\\ \nu\le L/u_2}} \sum_{\mu\nu u_1u_2u_3^2>M'} \frac{1}{\mu^2\nu^2u_1u_2u_3^2} &< \frac{5}{3M'} \sum_{u_1u_2u_3u_4=k} \sum_{\nu\le L/u_2} \frac{1}{\nu} \\ &\le \frac{5}{3M'}(1+\log L) \sum_{u_1u_2u_3u_4=k} 1, \end{aligned} \tag{1.26}$$

When $\delta \ge 2$, let $D := \sqrt{M'/u_1u_2u_3^2\mu\nu}$. By Lemma 1.2.5(ii) we have

$$\begin{aligned} \sum_{\substack{u_1u_2u_3u_4=k\\ (u_1,u_2)=1}} \sum_{\substack{\mu\le L/u_1\\ \nu\le L/u_2}} \sum_{\delta\ge\max\{2,D\}} \frac{1}{\mu^2\nu^2\delta^3u_1u_2u_3^2} &\le \frac{4(\zeta(3)-1)}{M'} \sum_{u_1u_2u_3u_4=k} \sum_{\substack{\mu\le L/u_1\\ \nu\le L/u_2}} \frac{1}{\mu\nu} \\ &\le \frac{4(\zeta(3)-1)}{M'}(1+\log L)^2 \sum_{u_1u_2u_3u_4=k} 1. \end{aligned} \tag{1.27}$$

Substituting (1.26) and (1.27) back into (1.25) and then (1.24), we have

$$\begin{aligned}\sum_{k\le L_1}\frac{1}{k}\sum_{\substack{q<p\le L\\ pq>M}}\left(1+\frac{x(k,\lambda(pq))}{pq\lambda(pq)}\right) &< \frac{1}{2}L^2(1+\log L_1)\\ &+ x(1+\log L)\Big(\frac{5}{6M'}+\frac{2(\zeta(3)-1)}{M'}(1+\log L)\Big)\sum_{k\le L_1}\frac{\tau_{(4)}(k)}{k},\end{aligned} \tag{1.28}$$

where $\tau_{(i)}(k)$ is the number of ordered factorizations of k into i positive factors. In [29] (see (4.9)), an easy induction argument shows that

$$\sum_{k\le y}\frac{\tau_{(i)}(k)}{k}\le\frac{1}{i!}(i+\log y)^i$$

for any natural number i and any $y\ge 1$. Using this in (1.28) and then combining with (1.23) gives

$$x\sum_{k\le L_1}\frac{B_k(x)}{k}\le x^2C,$$

where C is as in Theorem 1.4.1. Thus, from (1.21) we have the following result.

Theorem 1.4.3. *If* $10<L_1<L<M/2$ *and* $x>L^2>M$, *then*

$$\sum_{n\text{ in case (iii)}}{}' F(n)\le x^2C,$$

where C *is as in Theorem 1.4.1.*

Combining Theorems 1.3.5, 1.4.2 and 1.4.3 yield Theorem 1.4.1. □

Finally, Theorems 1.3.1 and 1.4.1 give the following result.

Theorem 1.4.4. *If* $0<c<1$, $10<L_1<L$, $2L<M<L^2<x$, *and* $x\ge 2657$, *then* $P(x)\le 1/(1+z^{-1})$ *where*

$$z=\left(x^{c-1}\big(1+f(L,M^{1/2})\big)\prod_{2<p\le M^{1/2}}\big(1-p^{-c}\big)^{-1}+B+C\right)\big(2\log x-\tfrac{1}{2}\big),$$

f *is as in Lemma 1.2.4,* B *is as in Theorem 1.3.2, and* C *is as in Theorem 1.4.1.*

1.5 The strong probable prime test

The next theorem extends the applicability of Theorems 1.3.3 and 1.4.4 to the probability, $P_1(x)$, that an odd composite $n \le x$ passes the strong probable prime test to a random base. For an odd number n, let $S(n)$ denote the number of integers $1 \le b \le n-1$ such that n is a strong probable prime to the base b, cf. (1.3). Thus,

$$P_1(x) = \frac{\sum'_{n\le x}\big(S(n)-2\big)}{\sum'_{n\le x}\big(S(n)-2\big) + \sum_{2<p\le x}(p-3)}.$$

The following theorem together with Theorems 1.3.1, 1.3.2, and 1.4.1 allows for a numerical estimation of $P_1(x)$ for various values of x.

Theorem 1.5.1. *For $x \ge 1$, we have that*

$$\sum_{n\le x}{}' \big(S(n)-2\big) \le \frac{1}{2}\sum_{n\le x}{}' \big(F(n)-2\big).$$

Proof. By (2.1) in [14], we have that $S(n) \le 2^{1-\omega(n)}F(n)$, where $\omega(n)$ denotes the number of distinct prime factors of n. So, if n is odd and divisible by at least 2 different primes, we have $S(n) \le \frac{1}{2}F(n)$. Further, if $n = p^a$ is an odd prime power then $S(p^a) = F(p^a) = p-1$. Therefore we have

$$\begin{aligned}\sum_{n\le x}{}' \big(S(n)-2\big) &\le \sum_{n\le x}{}' \Big(\frac{1}{2}F(n)-2\Big) + \frac{1}{2}\sum_{\substack{2<p^a\le x\\ a\ge 2}}(p-1)\\ &= \frac{1}{2}\sum_{n\le x}{}' \big(F(n)-2\big) - \sum_{n\le x}{}' 1 + \frac{1}{2}\sum_{\substack{2<p\le x^{1/a}\\ a\ge 2}}(p-1),\end{aligned}$$

so to prove the theorem it is enough to show that

$$\sum_{n\le x}{}' 1 \ge \frac{1}{2}\sum_{\substack{2<p\le x^{1/a}\\ a\ge 2}}(p-1). \tag{1.29}$$

Since 3 times an odd integer > 1 is an odd composite number, we have

$$\sum_{n\le x}{}' 1 \ge \sum_{\substack{1<m\le x/3\\ m\,\mathrm{odd}}} 1 = \left\lfloor \frac{x}{6} - \frac{1}{2}\right\rfloor > \frac{1}{6}x - \frac{3}{2}.$$

Also, since the primes larger than 2 are odd, for a given value of a we have

$$\frac{1}{2}\sum_{2<p\le x^{1/a}}(p-1) \le \sum_{j\le\frac{1}{2}(x^{1/a}-1)} j \le \frac{1}{2}\Big(\frac{1}{2}x^{1/a}-1\Big)\Big(\frac{1}{2}x^{1/a}+1\Big) < \frac{1}{8}x^{2/a}.$$

Adding these inequalities for $a = 2, 3, \ldots, \lfloor \log x/\log 3 \rfloor$, we see that (1.29) will follow if we show that

$$\frac{1}{6}x - \frac{3}{2} > \frac{1}{8}x + \frac{1}{8}x^{2/3} + \frac{1}{8}x^{1/2}(\log x/\log 3 - 3).$$

This inequality holds for $x \geq 254$. For $9 \leq x < 254$, (1.29) can be verified directly. Indeed, the prime sum in (1.29) increases only at the 8 powers of odd primes to 254 and it is enough to compute the two sums at those points. For $x < 9$,

$$\sum_{n \leq x}{}' \big(F(n) - 2\big) = \sum_{n \leq x}{}' \big(S(n) - 2\big) = 0,$$

so the theorem holds here as well. This completes the proof. □

We remark that the same result holds for the Euler probable prime test (also known as the Solovay–Strassen test). This involves verifying that the odd number n satisfies $a^{(n-1)/2} \equiv \left(\frac{a}{n}\right) \pmod{n}$, where $\left(\frac{a}{n}\right)$ is the Jacobi symbol. Indeed, from Monier's formula, see [18, (5.4)], we have that the number of bases $a \pmod{n}$ for which the Euler congruence holds is also $\leq 2^{1-\omega(n)}F(n)$. Like the strong test (as discussed in the introduction), an advantage with the Euler probable prime test is that more liars may be weeded out by repeating the test.

1.6 Numerical results

We apply Theorems 1.3.3 and 1.4.4 to obtain numerical bounds on $P(x)$ for various values of x. In Figure 1.3, bounds on $P(2^k)$ are computed via Theorem 1.3.3 for $40 \leq k \leq 130$ and Theorem 1.4.4 for $140 \leq k \leq 330$, at which point the methods of this paper lose their edge over those in [29]. To select values for parameters L, L_1, M, c, we started with an initial guess based on [29], and then optimized each parameter in turn (holding the others fixed). The reported values were determined by repeated this process five times.

Note that the upper bounds in Theorems 1.3.3, 1.4.4 are decreasing functions in x, so one can use the Figure 1.3 data to compute upper bounds for values of x between consecutive entries.

We also compute the exact values of $P(x)$ for $x = 2^k$ when $k \leq 36$. By definition,

$$P(x) = \frac{S_c(x)}{S_c(x) + S_p(x)}$$

for

$$S_p(x) = \sum_{2 < p \leq x} (p - 3), \quad S_c(x) = \sum_{n \leq x}{}' \big(F(n) - 2\big).$$

For ease, we have split up the computation into dyadic intervals $(2^{k-1}, 2^k)$. Letting

$$S_p(x/2, x) = \sum_{x/2<p\leq x} (p-3), \quad S_c(x/2, x) = {\sum_{x/2<n\leq x}}' \big(F(n) - 2\big),$$

we have that

$$P(2^k) = \frac{\sum_{j=3}^{k} S_c(2^{j-1}, 2^j)}{\sum_{j=3}^{k} S_p(2^{j-1}, 2^j) + S_c(2^{j-1}, 2^j)}. \tag{1.30}$$

Figure 1.3: Upper bound on $P(2^k)$.

k	L	L_1	$M^{1/2}$	c	$P(2^k) \leq$
40	307^-	135		0.5440	$4.306E{-1}$
50	727^-	318		0.5831	$2.904E{-1}$
60	$1.860E{+3}$	831		0.6235	$1.848E{-1}$
70	$4.000E{+3}$	$1.75E{+3}$		0.6491	$1.127E{-1}$
80	$8.500E{+3}$	$3.72E{+3}$		0.6704	$6.728E{-2}$
90	$1.804E{+4}$	$7.55E{+3}$		0.6906	$4.017E{-2}$
100	$3.505E{+4}$	$1.54E{+4}$		0.7052	$2.388E{-2}$
110	$7.351E{+4}$	$3.27E{+4}$		0.7217	$1.435E{-2}$
120	$1.354E{+5}$	$5.95E{+4}$		0.7321	$8.612E{-3}$
130	$2.507E{+5}$	$1.10E{+5}$		0.7423	$5.229E{-3}$
140	$9.90E{+5}$	$1.57E{+5}$	$2.379E{+5}$	0.7444	$3.265E{-3}$
150	$2.20E{+6}$	$3.19E{+5}$	$3.739E{+5}$	0.7504	$1.799E{-3}$
160	$4.88E{+6}$	$6.21E{+5}$	$5.689E{+5}$	0.7554	$9.932E{-4}$
170	$1.05E{+7}$	$1.21E{+6}$	$8.669E{+5}$	0.7602	$5.505E{-4}$
180	$2.21E{+7}$	$2.30E{+6}$	$1.315E{+6}$	0.7648	$3.064E{-4}$
190	$4.55E{+7}$	$4.55E{+6}$	$1.990E{+6}$	0.7692	$1.714E{-4}$
200	$9.23E{+7}$	$8.69E{+6}$	$2.990E{+6}$	0.7734	$9.634E{-5}$
210	$1.84E{+8}$	$1.66E{+7}$	$4.455E{+6}$	0.7773	$5.447E{-5}$
220	$3.62E{+8}$	$3.16E{+7}$	$6.627E{+6}$	0.7811	$3.097E{-5}$
230	$7.19E{+8}$	$5.74E{+7}$	$9.644E{+6}$	0.7845	$1.770E{-5}$
240	$1.38E{+9}$	$1.09E{+8}$	$1.410E{+7}$	0.7878	$1.017E{-5}$
250	$2.62E{+9}$	$2.01E{+8}$	$2.049E{+7}$	0.7911	$5.876E{-6}$
260	$4.96E{+9}$	$3.66E{+8}$	$2.946E{+7}$	0.7941	$3.412E{-6}$
270	$9.29E{+9}$	$6.64E{+8}$	$4.204E{+7}$	0.7969	$1.992E{-6}$
280	$1.73E{+10}$	$1.19E{+9}$	$5.998E{+7}$	0.7996	$1.169E{-6}$
290	$3.16E{+10}$	$2.18E{+9}$	$8.558E{+7}$	0.8023	$6.888E{-7}$
300	$5.83E{+10}$	$3.97E{+9}$	$1.197E{+8}$	0.8048	$4.080E{-7}$
310	$1.06E{+11}$	$6.87E{+9}$	$1.678E{+8}$	0.8072	$2.428E{-7}$
320	$1.90E{+11}$	$1.20E{+10}$	$2.346E{+8}$	0.8094	$1.451E{-7}$
330	$3.38E{+11}$	$2.10E{+10}$	$3.297E{+8}$	0.8117	$8.713E{-8}$

Note that the probability that an odd composite in the interval $(2^{k-1}, 2^k)$ passes the Fermat test is given by

$$P(2^{k-1}, 2^k) := \frac{S_c(2^{k-1}, 2^k)}{S_p(2,^{k-1}, 2^k) + S_c(2,^{k-1}, 2^k)}.$$

We have directly computed $S_p(2^{k-1}, 2^k)$ and $S_c(2^{k-1}, 2^k)$ for $k \leq 36$, with the latter computation aided by the formula $F(n) = \prod_{p|n}(p-1, n-1)$. Specifically, S_p is computed directly from the available list of primes up to 2^{36}. To compute S_c we use a sieve-like procedure. We initialize an array representing the odd numbers from 2^{k-1} and 2^k with all 1's. For each prime p to $2^k/3$, we let m run over the odd numbers between $2^{k-1}/p$ and $2^k/p$. For each m, we locate mp in the array, multiplying the entry there by $\gcd(m-1, p-1)$. At the end of the run the non-1 entries in our array correspond to the numbers $F(n)$ for n odd and composite. Note this avoids factoring integers n in $(2^{k-1}, 2^k)$, though a brute force method to the modest level of 2^{36} would have worked too.

In Figure 1.4, we provide the values of $S_p(2^k)$ and $S_c(2^k)$, as well as $P(2^k)$ and $P(2^{k-1}, 2^k)$.

Figure 1.4: Exact values of data.

k	$S_p(2^{k-1}, 2^k)$	$S_c(2^{k-1}, 2^k)$	$P(2^{k-1}, 2^k)$	$P(2^k)$
3	6	0	0	0
4	18	2	$1.000E{-1}$	$7.692E{-2}$
5	104	4	$3.704E{-2}$	$4.478E{-2}$
6	320	24	$6.977E{-2}$	$6.276E{-2}$
7	1180	114	$8.810E{-2}$	$8.126E{-2}$
8	4292	316	$6.858E{-2}$	$7.210E{-2}$
9	16338	1114	$6.384E{-2}$	$6.605E{-2}$
10	57416	3056	$5.054E{-2}$	$5.492E{-2}$
11	208576	10890	$4.962E{-2}$	$5.109E{-2}$
12	780150	28094	$3.476E{-2}$	$3.922E{-2}$
13	2837158	74528	$2.600E{-2}$	$2.936E{-2}$
14	10673384	231514	$2.123E{-2}$	$2.342E{-2}$
15	39467286	582318	$1.454E{-2}$	$1.695E{-2}$
16	148222234	1636968	$1.092E{-2}$	$1.254E{-2}$
17	559288478	4521166	$8.019E{-3}$	$9.224E{-3}$
18	2106190104	11682336	$5.516E{-3}$	$6.503E{-3}$
19	7995006772	33290330	$4.147E{-3}$	$4.770E{-3}$
20	30299256236	88781082	$2.922E{-3}$	$3.410E{-3}$
21	115430158810	230250774	$1.991E{-3}$	$2.364E{-3}$
22	440353630422	628735800	$1.426E{-3}$	$1.672E{-3}$
23	1683364186642	1680806136	$9.975E{-4}$	$1.174E{-3}$
24	6448755473484	4408788648	$6.832E{-4}$	$8.115E{-4}$
25	24754014371036	11552686982	$4.665E{-4}$	$5.565E{-4}$
26	95132822935752	30756273488	$3.232E{-4}$	$3.840E{-4}$
27	366232744269106	82133627362	$2.242E{-4}$	$2.657E{-4}$
28	1411967930053822	215629423796	$1.527E{-4}$	$1.820E{-4}$
29	5450257882815404	565834872742	$1.038E{-4}$	$1.241E{-4}$
30	21065843780715212	1504267288346	$7.140E{-5}$	$8.504E{-5}$
31	81507897575948416	3999812059436	$4.907E{-5}$	$5.837E{-5}$
32	315718919767278610	10350692466866	$3.278E{-5}$	$3.940E{-5}$
33	1224166825030041460	27472503360964	$2.244E{-5}$	$2.682E{-5}$
34	4750936696054816476	72288538641772	$1.522E{-5}$	$1.821E{-5}$
35	18454541611019193346	190806759987694	$1.034E{-5}$	$1.237E{-5}$
36	71745407298862105164	498526567616818	$6.949E{-6}$	$8.342E{-6}$

Additionally, we have estimated $P(2^k)$ in the range $30 \le k \le 50$ using random sampling. More precisely, we randomly sample $\lfloor 2^{k/2} \rfloor$ odd composite numbers in the interval $(2^{k-1}, 2^k)$, estimating $S_p(2^{k-1}, 2^k)$ by

$$\widehat{S}_p(2^{k-1}, 2^k) = \int_{2^{k-1}}^{2^k} \frac{t-3}{\log t}\,\mathrm{d}t = \mathrm{Li}(2^{2k}) - \mathrm{Li}(2^{2(k-1)}) - 3\big(\mathrm{Li}(2^k) - \mathrm{Li}(2^{k-1})\big),$$

in order to smooth out some noise from the experiment. To estimate $S_c(2^{k-1}, 2^k)$, we add up $F(n) - 2$ for each odd composite n sampled, and scale this sum by

$$\frac{2^{k-2} - \mathrm{Li}(2^k) + \mathrm{Li}(2^{k-1})}{2^{k/2}},$$

representing the ratio between the number of composites in the interval and the number of samples taken. We repeat this procedure ten times, and compute the mean, $\widehat{S}_{\text{mean}}(2^{k-1}, 2^k)$, and median, $\widehat{S}_{\text{median}}(2^{k-1}, 2^k)$, of the data. Using these statistics, we estimate $P(2^{k-1}, 2^k)$ by

$$\widehat{P}_{\text{mean}}(2^{k-1}, 2^k) = \frac{\widehat{S}_{\text{mean}}(2^{k-1}, 2^k)}{\widehat{S}_p(2^{k-1}, 2^k) + \widehat{S}_{\text{mean}}(2^{k-1}, 2^k)},$$
$$\widehat{P}_{\text{median}}(2^{k-1}, 2^k) = \frac{\widehat{S}_{\text{median}}(2^{k-1}, 2^k)}{\widehat{S}_p(2^{k-1}, 2^k) + \widehat{S}_{\text{median}}(2^{k-1}, 2^k)}.$$

For $30 \le k \le 36$, $P(2^{k-1}, 2^k)$ is known, in which case we compute the relative errors, $\widehat{P}_{\text{mean}}/P - 1$ and $\widehat{P}_{\text{median}}/P - 1$, to get a sense of the accuracy of the experiment. Then we estimate $P(2^k)$ by

$$\widehat{P}_{\text{mean}}(2^k) = \frac{\widehat{S}_{\text{mean}}(2^k)}{\widehat{S}_p(2^k) + \widehat{S}_{\text{mean}}(2^k)}$$

where

$$\widehat{S}_{\text{mean}}(2^k) = \begin{cases} S_c(2^{k-1}) + \widehat{S}_{\text{mean}}(2^{k-1}, 2^k) & \text{for } 30 \le k \le 36, \\ S_c(2^{36}) + \sum_{j=37}^{k} \widehat{S}_{\text{mean}}(2^{j-1}, 2^j) & \text{for } 37 \le k \le 50, \end{cases}$$

and

$$\widehat{S}_p(2^k) = \begin{cases} S_p(2^{k-1}) + \widehat{S}_p(2^{k-1}, 2^k) & \text{for } 30 \le k \le 36, \\ S_p(2^{36}) + \sum_{j=37}^{k} \widehat{S}_p(2^{j-1}, 2^j) & \text{for } 37 \le k \le 50. \end{cases}$$

Results of the random sampling experiment are summarized in Figures 1.5 and 1.6.

One sees a negative bias in these data with the results of random sampling undershooting the true figures. The referee has pointed out to us that this may be due to Jensen's inequality applied to the convex function $x/(a + x)$, so that $\mathrm{E}[X/(a + X)] \ge \mathrm{E}[X]/(a + \mathrm{E}[X])$. The undershoot may also be due to the fact that on average $F(n)$ is much larger than it is typically. In fact, it is shown in [18] that on a set of asymptotic density 1, we have $F(n) = n^{o(1)}$, yet the average behavior is $> n^{15/23}$. The exponent $15/23$, after more recent work of Baker and Harman [4], can be replaced with 0.7039. It follows from an old conjecture of Erdős on the distribution of Carmichael numbers that on average $F(n)$ behaves like $n^{1-o(1)}$.

Figure 1.5: Random sampling estimates in range where $P(2^k)$ is known.

k	$\widehat{P}_{\text{mean}}(2^{k-1}, 2^k)$	rel. err.	$\widehat{P}_{\text{median}}(2^{k-1}, 2^k)$	rel. err.	$\widehat{P}_{\text{mean}}(2^k)$	rel. err.
30	$5.541E{-5}$	-0.224	$5.045E{-5}$	-0.293	$7.319E{-5}$	-0.139
31	$4.800E{-5}$	-0.022	$3.616E{-5}$	-0.263	$5.758E{-5}$	-0.014
32	$2.706E{-5}$	-0.175	$1.899E{-5}$	-0.421	$3.515E{-5}$	-0.108
33	$2.223E{-5}$	-0.009	$1.248E{-5}$	-0.444	$2.666E{-5}$	-0.006
34	$1.387E{-5}$	-0.088	$1.013E{-5}$	-0.334	$1.721E{-5}$	-0.055
35	$7.603E{-6}$	-0.265	$6.506E{-6}$	-0.371	$1.033E{-5}$	-0.165
36	$4.433E{-6}$	-0.362	$4.123E{-6}$	-0.407	$6.474E{-6}$	-0.224

Figure 1.6: Random sampling estimates in range where $P(2^k)$ is unknown.

k	$\widehat{P}_{\text{mean}}(2^{k-1}, 2^k)$	$\widehat{P}_{\text{median}}(2^{k-1}, 2^k)$	$\widehat{P}_{\text{mean}}(2^k)$
37	$4.113E{-6}$	$3.675E{-6}$	$5.200E{-6}$
38	$4.807E{-6}$	$2.677E{-6}$	$4.908E{-6}$
39	$3.008E{-6}$	$1.463E{-6}$	$3.496E{-6}$
40	$1.519E{-6}$	$1.097E{-6}$	$2.026E{-6}$
41	$9.078E{-7}$	$5.697E{-7}$	$1.194E{-6}$
42	$7.747E{-7}$	$3.772E{-7}$	$8.822E{-7}$
43	$3.472E{-7}$	$2.334E{-7}$	$4.842E{-7}$
44	$1.968E{-7}$	$1.677E{-7}$	$2.704E{-7}$
45	$1.639E{-7}$	$1.687E{-7}$	$1.911E{-7}$
46	$1.186E{-7}$	$1.198E{-7}$	$1.372E{-7}$
47	$1.051E{-7}$	$6.597E{-8}$	$1.133E{-7}$
48	$4.076E{-8}$	$3.947E{-8}$	$5.928E{-8}$
49	$3.791E{-8}$	$3.213E{-8}$	$4.337E{-8}$
50	$2.361E{-8}$	$1.318E{-8}$	$2.865E{-8}$

Chapter 2

Explicit estimates for the distribution of numbers free of large prime factors

2.1 Introduction

For a positive integer $n > 1$, denote by $P(n)$ the largest prime factor of n, and let $P(1) = 1$. Let $\Psi(x, y)$ denote the number of $n \leq x$ with $P(n) \leq y$. Such integers n are known as y-smooth, or y-friable. Asymptotic estimates for $\Psi(x, y)$ are quite useful in many applications, not least of which is in the analysis of factorization and discrete logarithm algorithms.

One of the earliest results is due to Dickman [15] in 1930, who gave an asympotic formula for $\Psi(x, y)$ in the case that x is a fixed power of y. Dickman showed that

$$\Psi(x, y) \sim x\rho(u) \qquad (y \to \infty,\ x = y^u) \tag{2.1}$$

for every fixed $u \geq 1$, where $\rho(u)$ is the "Dickman–de Bruijn" function, defined to be the continuous solution of the delay differential equation

$$\begin{aligned} u\rho'(u) + \rho(u-1) &= 0 \qquad (u > 1), \\ \rho(u) &= 1 \qquad (0 \leq u \leq 1). \end{aligned}$$

There remain the questions of the error in the approximation (2.1), and also the case when $u = \log x / \log y$ is allowed to grow with x and y. In 1951, de Bruijn [9] proved that

$$\Psi(x, y) = x\rho(u)\Big(1 + O_\varepsilon\Big(\frac{\log(1+u)}{\log y}\Big)\Big)$$

holds uniformly for $x \geq 2$, $\exp\{(\log x)^{5/8+\varepsilon}\} < y \leq x$, for any fixed $\varepsilon > 0$. After improvements in the range of this result by Maier and Hensley, Hildebrand [23] showed that the de Bruijn estimate holds when $\exp(\{(\log\log x)^{5/3+\varepsilon}\}) \leq y \leq x$.

In 1986, Hildebrand and Tenenbaum [24] provided a uniform estimate for $\Psi(x, y)$ for all $x \geq y \geq 2$, yielding an asymptotic formula when y and u tend to infinity. The starting point for their method is an elementary argument of Rankin [41] from 1938, commonly known now as Rankin's "trick". For complex s, define

$$\zeta(s, y) = \sum_{\substack{n \geq 1 \\ P(n) \leq y}} n^{-s} = \prod_{p \leq y} (1 - p^{-s})^{-1}$$

(where p runs over primes) as the partial Euler product of the Riemann zeta function $\zeta(s)$. In the case that $s = \sigma$ is real and $0 < \sigma < 1$, we have

$$\Psi(x, y) = \sum_{\substack{n \leq x \\ P(n) \leq y}} 1 \leq \sum_{P(n) \leq y} (x/n)^{\sigma} = x^{\sigma} \zeta(\sigma, y). \tag{2.2}$$

Then σ can be chosen optimally to minimize $x^{\sigma}\zeta(\sigma, y)$.

Let

$$\phi_j(s, y) = \frac{\partial^j}{\partial s^j} \log \zeta(s, y).$$

The function

$$\phi_1(s, y) = -\sum_{p \leq y} \frac{\log p}{p^s - 1}$$

is especially useful since the solution $\alpha = \alpha(x, y)$ to $\phi_1(\alpha, y) + \log x = 0$ gives the optimal σ in (2.2). We also denote $\sigma_j(x, y) = |\phi_j(\alpha(x, y), y)|$.

In this language, Hildebrand and Tenenbaum [24] proved that the estimate

$$\Psi(x, y) = \frac{x^{\alpha}\zeta(\alpha, y)}{\alpha\sqrt{2\pi\sigma_2(x, y)}} \Big(1 + O\Big(\frac{1}{u} + \frac{\log y}{y}\Big)\Big)$$

holds uniformly for $x \geq y \geq 2$. As suggested by this formula, quantities $\alpha(x, y)$ and $\sigma_2(x, y)$ are of interest, and were given uniform estimates which imply the formulae

$$\alpha(x, y) \sim \frac{\log(1 + y/\log x)}{\log y}$$

and

$$\sigma_2(x, y) \sim \Big(1 + \frac{\log x}{y}\Big) \log x \log y,$$

together which imply

$$\Psi(x,y) \sim \frac{x^\alpha \zeta(\alpha,y)}{\sqrt{2\pi u}\log(y/\log x)} \qquad (\text{if } y/\log x \to \infty),$$
$$\Psi(x,y) \sim \frac{x^\alpha \zeta(\alpha,y)}{\sqrt{2\pi y/\log y}} \qquad (\text{if } y/\log x \to 0).$$

These formulae indicate that $\Psi(x,y)$ undergoes a "phase change" when y is of order $\log x$, see [8]. This paper concentrates on the range where y is considerably larger, say $y > (\log x)^4$.

The primary aim of this paper is to make the Hildebrand–Tenenbaum method explicit and so effectively construct an algorithm for obtaining good bounds for $\Psi(x,y)$.

2.1.1 Explicit Results

Beyond the Rankin upper bound $\Psi(x,y) \le x^\alpha \zeta(\alpha,y)$, we have the explicit lower bound

$$\Psi(x,y) \ge x^{1-\log\log x/\log y} = \frac{x}{(\log x)^u}$$

due to Konyagin and Pomerance [29]. Recently Granville and Soundararajan [22] found an elementary improvement of Rankin's upper bound, which they have graciously permitted us to include, see Appendix 2.5. In particular, they show that

$$\Psi(x,y) \le 1.39 y^{1-\sigma} x^\sigma \zeta(\sigma,y)/\log x$$

for every value of $\sigma \in [1/\log y, 1]$, see Theorem 2.5.1.

In another direction, by relinquishing the goal of a compact formula, several authors have devised algorithms to compute bounds on $\Psi(x,y)$ for given x,y as inputs. For example, using an accuracy parameter c, Bernstein [7] created an algorithm to generate bounds $B^-(x,y) \le \Psi(x,y) \le B^+(x,y)$ with

$$\frac{B^-}{\Psi} \ge 1 - \frac{\log x}{c\log 3/\log 2} \quad \text{and} \quad \frac{B^+}{\Psi} \le 1 + \frac{2\log x}{c\log 3/\log 2},$$

running in

$$O\Big(\frac{y}{\log_2 y} + \frac{y\log x}{\log^2 y} + c\log x\log c\Big)$$

time. Parsell and Sorenson [36] refined this algorithm to run in

$$O\Big(c\frac{y^{2/3}}{\log y} + c\log x\log c\Big)$$

time, as well as obtaining faster and tighter bounds assuming the Riemann Hypothesis. The largest example computed by this method was an approximation of $\Psi(2^{255}, 2^{28})$.

As seen in Figure 2.1, the lower bound presented in this paper does better than the Konyagin–Pomerance lower bound by 10 orders of magnitude in the smaller example and 26 orders of magnitude in the larger example. The upper bound presented is about 2 to 3 orders of magnitude better than the Rankin estimate and about 1.5 orders of magnitude better than the new Granville–Soundararajan estimate.

As a point of reference we also give the main-term estimates $x^\alpha\zeta(\alpha, y)/\alpha\sqrt{2\pi\sigma_2}$ from [24] and $\rho(u)x$ from [15]. It is interesting that our lower and upper estimates in the second example create an interval for the true count that is tight enough to exclude both the Dickman–de Bruijn and Hildebrand–Tenenbaum main terms. The second-named author has asked if $\Psi(x, y) \geq x\rho(u)$ holds in general for $x \geq 2y \geq 2$, see [21, (1.25)]. This inequality is known for u bounded and x sufficiently large, see the discussion in [33, Section 9].

Figure 2.1: Examples.

x	10^{100}	10^{500}
y	10^{15}	10^{35}
KP	$1.786 \cdot 10^{84}$	$1.857 \cdot 10^{456}$
R	$4.599 \cdot 10^{96}$	$9.639 \cdot 10^{484}$
GS	$5.350 \cdot 10^{95}$	$6.596 \cdot 10^{483}$
DD	$2.523 \cdot 10^{94}$	$1.472 \cdot 10^{482}$
HT	$2.652 \cdot 10^{94}$	$1.5127 \cdot 10^{482}$
Ψ^-	$2.330 \cdot 10^{94}$	$1.4989 \cdot 10^{482}$
Ψ^+	$2.923 \cdot 10^{94}$	$1.5118 \cdot 10^{482}$

Here,

KP is the Konyagin–Pomerance lower bound $x/(\log x)^u$,

R is the Rankin upper bound $x^\alpha\zeta(\alpha, y)$,

GS is the Granville–Soundararajan upper bound $1.39y^{1-\alpha}x^\alpha\zeta(\alpha, y)/\log x$,

DD is the Dickman–de Bruijn main term $\rho(u)x$,

HT is the Hildebrand–Tenenbaum main term $x^\alpha\zeta(\alpha, y)/(\alpha\sqrt{2\pi\sigma_2})$, and

Ψ^-, Ψ^+ are the lower and upper bounds obtained in this paper.

Our principal result, which benefits from some notation developed over the course of the paper, is Theorem 2.3.11. It is via this theorem that we were able to estimate $\Psi(10^{100}, 10^{15})$ and $\Psi(10^{500}, 10^{35})$ as in the table above.

2.2 Plan for the paper

The basic strategy of the saddle-point method relies on Perron's formula, which implies the identity[1]

$$\Psi(x,y) = \frac{1}{2\pi i}\int_{\sigma-i\infty}^{\sigma+i\infty} \zeta(s,y)\frac{x^s}{s}\,ds,$$

for any $\sigma > 0$. A convenient value of σ to use is the saddle point $\alpha = \alpha(x,y)$ discussed in the Introduction: For any $\sigma > 0$, the integrand is maximized on the vertical line with real part σ at $s = \sigma$, and it is minimized for $\sigma > 0$ at α.

We are interested in abridging the integral at a certain height T and then approximating the contribution given by the tail. To this end, we have

$$\Psi(x,y) = \frac{1}{2\pi i}\int_{\alpha-iT}^{\alpha+iT} \zeta(s,y)\frac{x^s}{s}\,ds + \text{Error}. \tag{2.3}$$

There is a change in behavior occurring in $\zeta(s,y)$ when $t = \mathfrak{Im}(s)$ is on the order $1/\log y$. In [24] it is shown that

$$\begin{aligned}\Big|\frac{\zeta(s,y)}{\zeta(\alpha,y)}\Big| &= \prod_{p\le y}\Big|\frac{1-p^{-\alpha}}{1-p^{-s}}\Big| = \prod_{p\le y}\Big(1+\frac{2(1-\cos(t\log p))}{p^\alpha(1-p^{-\alpha})^2}\Big)^{-1/2}\\ &\le \exp\Big\{-\sum_{p\le y}\frac{1-\cos(t\log p)}{p^\alpha}\Big\}.\end{aligned} \tag{2.4}$$

Thus when t is small (compared to $1/\log y$) the oscillatory terms are in resonance, and when t is large the oscillatory terms should exhibit cancellation. This behavior suggests we should divide our range of integration into $|t| \le T_0$ and $T_0 < |t| < T$, where $T_0 \approx 1/\log y$ is a parameter to be optimized.

The contribution for $|t| \le T_0$ will constitute a "main term", and so we will try to estimate this part very carefully. In this range we forgo (2.4) and attack the integrand $\zeta(s,y)x^s/s$ directly. The basic idea is to expand $\phi(s,y) = \log\zeta(s,y)$ as a Taylor series in t. This approach, when carefully done, gives us fairly close upper and lower bounds for the integral. In our smaller example, the upper bound is less than

[1] The right side should be increased by $\frac{1}{2}$ in the case that x itself is a y-smooth integer.

1% higher than the lower bound, and in the larger example, this is better by a factor of 20. Considerably more noise is encountered beyond T_0 and in the Error in (2.3).

For the second range $T_0 < |t| < T$, we focus on obtaining a satisfactory lower bound on the sum over primes,

$$\sum_{p \le y} \frac{1 - \cos(t \log p)}{p^\alpha}.$$

Our strategy is to sum the first L terms directly, and then obtain an analytic formula $W(y, w)$ to lower bound the remaining terms starting at some $w \ge L$, where essentially

$$W(y, w) = \frac{y^{1-\alpha} - w^{1-\alpha}}{1 - \alpha} + \text{error}.$$

With an explicit version of Perron's formula, the Error in (2.3) may be handled by

$$\begin{aligned} \big|\text{Error}\big| &\le x^\alpha \sum_{\substack{P(n) \le y \\ T|\log(x/n)| > T^d}} \frac{1}{n^\alpha} \frac{1}{\pi T |\log(x/n)|} + \sum_{\substack{P(n) \le y \\ T|\log(x/n)| \le T^d}} \left(\frac{x}{n}\right)^\alpha \\ &\le \frac{x^\alpha \zeta(\alpha, y)}{\pi T^d} + e^{\alpha T^{d-1}} \Big[\Psi(xe^{T^{d-1}}, y) - \Psi(xe^{-T^{d-1}}, y)\Big]. \end{aligned}$$

Here $d \approx \frac{1}{2}$ is a parameter of our choosing, which we set to balance the two terms above. Thus the problem of bounding |Error| is reduced to estimating the number of y-smooth integers in the "short" interval $\big(xe^{-T^{d-1}}, xe^{T^{d-1}}\big]$.

This latter portion is better handled when T is large, but the earlier portion in the range $[T_0, T]$ is better handled when T is small. Thus, T is numerically set to balance these two forces.

In our proofs we take full advantage of some recent calculations involving the prime-counting function $\pi(x)$ and the Chebyshev functions

$$\psi(x) = \sum_{p^m \le x} \log p, \quad \vartheta(x) = \sum_{p \le x} \log p,$$

with p running over primes and m running over positive integers. As a corollary of the papers [11], [12] of Büthe we have the following excellent result.

Proposition 2.2.1. *For* $1427 \le x \le 10^{19}$ *we have*

$$.05\sqrt{x} \le x - \vartheta(x) \le 1.95\sqrt{x}.$$

We have

$$\frac{|\vartheta(x) - x|}{x} < \begin{cases} 2.3 \cdot 10^{-8}, & \text{when } x > 10^{19}, \\ 1.2 \cdot 10^{-8}, & \text{when } x > e^{45}, \\ 1.2 \cdot 10^{-9}, & \text{when } x > e^{50}, \\ 2.9 \cdot 10^{-10}, & \text{when } x > e^{55}. \end{cases}$$

Proof. The first assertion is one of the main results in Büthe [12]. Let H be a number such that all zeros of the Riemann zeta-function with imaginary parts in $[0, H]$ lie on the 1/2-line. Inequality (7.4) in Büthe [11] asserts that if $x/\log x \le H^2/4.92^2$ and $x \ge 5000$, then

$$\frac{|\vartheta(x) - x|}{x} < \frac{(\log x - 2)\log x}{8\pi\sqrt{x}}.$$

We can take $H = 3 \cdot 10^{10}$, see Platt [37]. Thus, we have the result in the range $10^{19} \le x \le e^{45}$. For $x \ge e^{45}$ we have from Büthe [11] that $|\psi(x) - x|/x \le 1.118 \cdot 10^{-8}$. Further, we have (see [42, (3.39)]) for $x > 0$,

$$\psi(x) \ge \vartheta(x) > \psi(x) - 1.02x^{1/2} - 3x^{1/3}.$$

(This result can be improved, but it is not important to us.) Thus, for $x \ge e^{45}$ we have $|\vartheta(x) - x|/x \le 1.151 \cdot 10^{-8}$, establishing our result in this range. For the latter two ranges we argue similarly, using $|\psi(x) - x| \le 1.165 \cdot 10^{-9}$ when $x \ge e^{50}$ and $|\psi(x) - x| \le 2.885 \cdot 10^{-10}$ for $x \ge e^{55}$, both of these inequalities coming from [11]. □

We remark that there are improved inequalities at higher values of x, found in [11] and [19], which one would want to use if estimating $\Psi(x, y)$ for larger values of y than we have done here.

2.3 The main argument

As in the Introduction, for complex s, define

$$\zeta(s, y) = \sum_{\substack{n \ge 1 \\ P(n) \le y}} n^{-s} = \prod_{p \le y} (1 - p^{-s})^{-1},$$

which is the Riemann zeta function restricted to y-smooth numbers, and for $j \ge 0$, let

$$\phi_j(s, y) = \frac{\partial^j}{\partial s^j} \log \zeta(s, y).$$

We have the explicit formulae,

$$\begin{aligned}
\phi_1(s,y) &= -\sum_{p\le y} \frac{\log p}{p^s-1},\\
\phi_2(s,y) &= \sum_{p\le y} \frac{p^s\log^2 p}{(p^s-1)^2},\\
\phi_3(s,y) &= -\sum_{p\le y} \frac{(p^{2s}+p^s)\log^3 p}{(p^s-1)^3},\\
\phi_4(s,y) &= \sum_{p\le y} \frac{(p^{3s}+4p^{2s}+p^s)\log^4 p}{(p^s-1)^4},\\
\phi_5(s,y) &= -\sum_{p\le y} \frac{(p^{4s}+11p^{3s}+11p^{2s}+p^s)\log^5 p}{(p^s-1)^5}.
\end{aligned}$$

Note that for $y \ge 2$, $\sigma > 0$, $\phi_1(\sigma, y)$ is strictly increasing from 0, so there is a unique solution $\alpha = \alpha(x, y) > 0$ to the equation

$$\log x + \phi_1(\alpha, y) = 0.$$

Since we cannot exactly solve this equation, we shall assume any choice of α that we use is a reasonable approximation to the exact solution, and we must take into account an upper bound for the difference between our value and the exact value. We denote

$$\phi_j = \phi_j(\alpha, y), \quad \sigma_j = |\phi_j| = (-1)^j\phi_j, \quad B_j = B_j(t) = \sigma_j t^j/j!$$

so that the Taylor series of $\phi(s, y) = \log \zeta(s, y)$ about $s = \alpha$ is

$$\phi(\alpha+it,y) = \sum_{j\ge 0} \frac{\sigma_j}{j!}(-it)^j = \sum_{j\ge 0}(-i)^j B_j.$$

Our first result, which is analogous to Lemma 10 in [24], sets the stage for our estimates.

Lemma 2.3.1. *Let* $0 < d < 1$ *and* $T > 1$. *We have that*

$$\begin{aligned}
&\left|\Psi(x,y) - \frac{1}{2\pi i}\int_{\alpha-iT}^{\alpha+iT} \zeta(s,y)\frac{x^s}{s}\,ds\right|\\
&\qquad \le \frac{x^\alpha\zeta(\alpha,y)}{\pi T^d} + e^{\alpha T^{d-1}}\left[\Psi(xe^{T^{d-1}},y) - \Psi(xe^{-T^{d-1}},y)\right].
\end{aligned}$$

Proof. We have

$$\begin{aligned}
\frac{1}{2\pi i}\int_{\alpha-iT}^{\alpha+iT} \zeta(s,y)\frac{x^s}{s}\,ds &= \frac{1}{2\pi i}\int_{\alpha-iT}^{\alpha+iT} \sum_{P(n)\le y}\left(\frac{x}{n}\right)^s \frac{ds}{s}\\
&= \sum_{P(n)\le y} \frac{1}{2\pi i}\int_{\alpha-iT}^{\alpha+iT} \left(\frac{x}{n}\right)^s \frac{ds}{s},
\end{aligned}$$

where the interchange of sum and integral is justified since $\zeta(s,y)$ is a finite product, hence uniformly convergent as a sum.

By Perron's formula (see [25, Theorem G] and its proof), we have

$$\left|\frac{1}{2\pi i}\int_{\alpha-iT}^{\alpha+iT}\left(\frac{x}{n}\right)^s\frac{ds}{s}\right| \le \frac{(x/n)^\alpha}{\max\Big(1,\pi T|\log(x/n)|\Big)} \qquad \text{if } n>x,$$

$$\left|1-\frac{1}{2\pi i}\int_{\alpha-iT}^{\alpha+iT}\left(\frac{x}{n}\right)^s\frac{ds}{s}\right| \le \frac{(x/n)^\alpha}{\max\Big(1,\pi T|\log(x/n)|\Big)} \qquad \text{if } n\le x.$$

Together these imply

$$\begin{aligned}\left|\Psi(x,y)-\frac{1}{2\pi i}\int_{\alpha-iT}^{\alpha+iT}\zeta(s,y)\frac{x^s}{s}\,ds\right| &\le x^\alpha\sum_{P(n)\le y}\frac{n^{-\alpha}}{\max\Big(1,\pi T|\log(x/n)|\Big)}\\ &\le x^\alpha\sum_{\substack{P(n)\le y\\ |\log(x/n)|>T^{d-1}}}\frac{1}{n^\alpha}\frac{1}{\pi T|\log(x/n)|}+x^\alpha\sum_{\substack{P(n)\le y\\ |\log(x/n)|\le T^{d-1}}}\frac{1}{n^\alpha}\\ &\le \frac{x^\alpha\zeta(\alpha,y)}{\pi T^d}+e^{\alpha T^{d-1}}\Big[\Psi(xe^{T^{d-1}},y)-\Psi(xe^{-T^{d-1}},y)\Big].\end{aligned}$$

This completes the proof. □

In using this result we have the problems of performing the integration from $\alpha-iT$ to $\alpha+iT$ and estimating the number of y-smooth integers in the interval $\big(xe^{-T^{d-1}},xe^{T^{d-1}}\big]$. We turn first to the integral evaluation.

Recall that $B_j=B_j(t)=\sigma_j(x,y)t^j/j!$ and let $B_1^*=B_1^*(t)=t\log x-B_1(t)$. Note that $B_1^*=0$ if α is chosen perfectly.

Lemma 2.3.2. *For $s=\alpha+it$, we have*

$$\mathfrak{Re}\Big\{\zeta(s,y)\frac{x^s}{s}\Big\}=$$
$$\frac{x^\alpha\zeta(\alpha,y)}{\alpha^2+t^2}\big(\alpha\cos(B_3+B_1^*+b_5)+t\sin(B_3+B_1^*+b_5)\big)\exp\big\{-B_2+B_4+a_5\big\},$$

where a_5,b_5 are real numbers, depending on the choice of t, with $|a_5+ib_5|\le B_5(t)$.

Proof. We expand $\phi(\alpha+it,y)=\log\zeta(\alpha+it,y)$ in a Taylor series around $t=0$. There exists some real ξ between 0 and t such that

$$\begin{aligned}\phi(\alpha+it,y)&=\phi(\alpha,y)+it\phi_1-\frac{t^2}{2}\phi_2-\frac{it^3}{3!}\phi_3+\frac{t^4}{4!}\phi_4-i\frac{t^5}{5!}(\alpha+i\xi,y)\\ &=B_0-iB_1-B_2+iB_3+B_4-i\frac{t^5}{5!}\phi_5(\alpha+i\xi,y).\end{aligned}$$

Since $\zeta(s,y)=\exp(\phi(s,y))$, we obtain

$$\begin{aligned}\zeta(s,y)\frac{x^s}{s} &= \frac{\zeta(\alpha,y)x^\alpha}{\alpha+it}\exp\Big\{it\log x - iB_1 - B_2 + iB_3 + B_4 + i\frac{t^5}{5!}\phi_5(\alpha+i\xi,y)\Big\}\\ &= \frac{x^\alpha\zeta(\alpha,y)}{\alpha+it}\exp\Big\{-B_2+B_4+i(B_1^*+B_3)+i\frac{t^5}{5!}\phi_5(\alpha+i\xi,y)\Big\}.\end{aligned}$$

Letting $i\phi_5(\alpha+i\xi)t^5/5! = a_5+b_5 i$, we have

$$\zeta(s,y)\frac{x^s}{s} =$$
$$\frac{x^\alpha\zeta(\alpha,y)}{\alpha^2+t^2}(\alpha-it)\big(\cos(B_1^*+B_3+b_5)+i\sin(B_1^*+B_3+b_5)\big)\exp\big\{-B_2+B_4+a_5\big\},$$

and taking the real part gives the result. □

The main contribution to the integral in Lemma 2.3.1 turns out to come from the interval $[-T_0,T_0]$, where T_0 is fairly small. We have

$$\frac{1}{2\pi i}\int_{\alpha-iT_0}^{\alpha+iT_0}\zeta(s,y)\frac{x^s}{s}\,ds = \frac{1}{2\pi}\int_{-T_0}^{T_0}\zeta(\alpha+it,y)\frac{x^{\alpha+it}}{\alpha+it}\,dt.$$

Note that the integrand, written as a Taylor series around $s=\alpha$, has real coefficients, so the real part is an even function of t and the imaginary part is an odd function. Thus, the integral is real, and its value is double the value of the integral on $[0,T_0]$.

Consider the cosine, sine combination in Lemma 2.3.2:

$$f(t,v) := \alpha\cos(B_3(t)+v)+t\sin(B_3(t)+v),$$

and let

$$v_0(t) = |B_1^*(t)| + B_5(t).$$

We have, for each value of t, the constraint that $|v|\le v_0(t)$. The partial derivative of $f(t,v)$ with respect to v is zero when $\arctan(t/\alpha)-B_3(t)\equiv 0 \pmod{\pi}$. Let

$$u(t) = \arctan(t/\alpha) - B_3(t).$$

If $u(t)\notin[-v_0(t),v_0(t)]$, then $f(t,v)$ is monotone in v on that interval; otherwise it has a min or max at $u(t)$. Let T_3,T_2,T_1,T_0 be defined, respectively, as the least positive solutions of the equations

$$u(t)=v_0(t),\quad u(t)=-v_0(t),\quad u(t)+\pi=v_0(t),\quad u(t)+\pi=-v_0(t).$$

Then $0<T_3<T_2<T_1<T_0$. We have the following properties for $f(t,v)$:

1. For t in the interval $[0, T_3]$ we have $f(t, v)$ increasing for $v \in [-v_0(t), v_0(t)]$, so that
$$f(t, -v_0(t)) \leq f(t, v) \leq f(t, v_0(t)).$$
2. For t in the interval $[T_3, T_2]$, we have $f(t, v)$ increasing for $-v_0(t) \leq v \leq u(t)$ and then decreasing for $u(t) \leq v \leq v_0(t)$. Thus,
$$\min\{f(t, -v_0(t)), f(t, v_0(t))\} \leq f(t, v) \leq f(t, u(t)).$$
3. For $t \in [T_2, T_1]$, $f(t, v)$ is decreasing for $v \in [-v_0(t), v_0(t)]$, so that
$$f(t, v_0(t)) \leq f(t, v) \leq f(t, -v_0(t)).$$
4. For $t \in [T_1, T_0]$, we have $f(t, v)$ decreasing for $v \in [-v_0(t), u(t) + \pi]$ and increasing for $v \in [u(t) + \pi, v_0(t)]$; that is,
$$f(t, u(t) + \pi) \leq f(t, v) \leq \max\{f(t, -v_0(t)), f(t, v_0(t))\}.$$

Note too that $f(t, v)$ has a sign change from positive to negative in the interval $[T_2, T_1]$. Let Z^-, Z^+ be, respectively, the least positive roots of $f(t, v(t)) = 0$, $f(t, -v(t)) = 0$.

Let I_0^+ be an upper bound for the function appearing in Lemma 2.3.2 on $[0, T_0]$ using $|a_5|, |b_5| \leq B_5$ and the above facts about $f(t, v)$, and let I_0^- be the corresponding lower bound. We choose $a_5 = B_5$ in I_0^+ when the cos, sin combination is positive, and $a_5 = -B_5$ when it is negative. For I_0^-, we choose a_5 in the reverse way.

Let
$$J_0^+ = \int_0^{T_0} I_0^+(t)\, dt, \quad J_0^- = \int_0^{T_0} I_0^-(t)\, dt. \tag{2.5}$$
We thus have the following result, which is our analogue of Lemma 11 in [24].

Lemma 2.3.3. *We have*
$$\frac{x^\alpha \zeta(\alpha, y)}{\pi} J_0^- \leq \frac{1}{2\pi i} \int_{\alpha - iT_0}^{\alpha + iT_0} \zeta(s, y) \frac{x^s}{s}\, ds \leq \frac{x^\alpha \zeta(\alpha, y)}{\pi} J_0^+.$$

In order to estimate the integral in Lemma 2.3.1 when $|t| > T_0$ we must know something about prime sums to y.

Lemma 2.3.4. *We have*
$$\Big| \int_{\alpha + iT_0}^{\alpha + iT} \zeta(s, y) \frac{x^s}{s}\, ds \Big| \leq x^\alpha \zeta(\alpha, y) J_1,$$

where

$$J_1 := \int_{T_0}^{T} \exp\big(-W(y,1,t)\big)\,\frac{dt}{\sqrt{\alpha^2+t^2}}$$

and

$$W(v,w,t) := \sum_{w<p\le v} \frac{1-\cos(t\log p)}{p^\alpha}. \tag{2.6}$$

Proof. For $0 \le v \le 1 < t$, equation (3.14) in [24] states that

$$(1 + 4vt/(t-1)^2)^{-1} \le \exp\{-4v/t\}.$$

Applied to (3.17) in [24] with $v = (1-\cos(t\log p))/2$, we have that

$$\begin{aligned} \Big|\frac{\zeta(s,y)}{\zeta(\alpha,y)}\Big| &= \prod_{p\le y}\Big|\frac{1-p^{-\alpha}}{1-p^{-s}}\Big| = \prod_{p\le y}\Big(1+\frac{2(1-\cos(t\log p))}{p^\alpha(1-p^{-\alpha})^2}\Big)^{-1/2} \\ &\le \exp\Big\{-\sum_{p\le y}\frac{1-\cos(t\log p)}{p^\alpha}\Big\}. \end{aligned} \tag{2.7}$$

This completes the proof. □

Our goal now is to find a way to estimate $W(v,w,t)$. The following result is analogous to Lemma 6 in [24].

Lemma 2.3.5. *Let s be a complex number, let $1 < w < v$, and define*

$$F_s(v,w) := \sum_{w<p\le v} \frac{\log p}{p^s} - \frac{v^{1-s} - w^{1-s}}{1-s}.$$

(i) If $v \le 10^{19}$ we have

$$|F_s(v,w)| \le 2(v^{1/2-\alpha} + w^{1/2-\alpha}) + 2|s|\frac{w^{1/2-\alpha} - v^{1/2-\alpha}}{\alpha - 1/2}.$$

(ii) If $10^{19} \le w \le v$ we have

$$|F_s(v,w)| \le \varepsilon_w\Big(v^\beta + w^\beta + |s|\frac{v^\beta - w^\beta}{\beta}\Big),$$

where $\beta = 1-\alpha$ and

$$\varepsilon_w = \begin{cases} 2.3\cdot 10^{-8}, & w \in (10^{19}, e^{45}], \\ 1.2\cdot 10^{-8}, & w \in (e^{50}, e^{55}], \\ 1.2\cdot 10^{-9}, & w \in (e^{50}, e^{55}], \\ 2.9\cdot 10^{-10}, & w > e^{55}. \end{cases}$$

Proof. (i) By partial summation,

$$\sum_{w<p\le v}\frac{\log p}{p^s}=\frac{\vartheta(v)}{v^s}-\frac{\vartheta(w)}{w^s}+\int_w^v s\frac{\vartheta(t)}{t^{s+1}}\,dt$$
$$=\frac{v^{1-s}-w^{1-s}}{1-s}-\frac{E(v)}{v^s}+\frac{E(w)}{w^s}-\int_w^v s\frac{E(t)}{t^{s+1}}\,dt,$$

so by the first part of Proposition 2.2.1,

$$|F_s(v,w)|\le\frac{|E(v)|}{v^\alpha}+\frac{|E(w)|}{w^\alpha}+|s|\int_w^v\frac{E(t)}{t^{1+\alpha}}\,dt$$
$$\le 2v^{1/2-\alpha}+2w^{1/2-\alpha}+2|s|\frac{v^{1/2-\alpha}-w^{1/2-\alpha}}{1/2-\alpha}.$$

(ii) Similarly, by the second part of Proposition 2.2.1,

$$|F_s(v,w)|\le\frac{|E(v)|}{v^\alpha}+\frac{|E(w)|}{w^\alpha}+|s|\int_w^v\frac{E(t)}{t^{1+\alpha}}\,dt\le\varepsilon_w\Big(v^{1-\alpha}+w^{1-\alpha}+|s|\int_w^v\frac{dt}{t^\alpha}\Big)$$
$$=\varepsilon_w\Big(v^{1-\alpha}+w^{1-\alpha}+|s|\frac{v^{1-\alpha}-w^{1-\alpha}}{1-\alpha}\Big).$$

□

The following result plays the role of Corollary 6.1 in [24].

Lemma 2.3.6. *For $t\in\mathbb{R}$, $z>1$, and $\beta=1-\alpha$, let*

$$\delta_z:=t\log z-\arctan(t/\beta).$$

(i) For $1427\le w<v\le 10^{19}$ we have that $W(v,w,t)\ge W_0(v,w,t)$, where

$$W_0(v,w,t)\log v=\frac{v^\beta-w^\beta}{\beta}-\frac{v^\beta\cos\delta_v-w^\beta\cos\delta_w}{\sqrt{\beta^2+t^2}}$$
$$-4(v^{1/2-\alpha}+w^{1/2-\alpha})-2(\alpha+|s|)\frac{w^{1/2-\alpha}-v^{1/2-\alpha}}{\alpha-1/2}.$$

(ii) For $10^{19}\le w<v$ we have that $W(v,w,t)\ge W_0(v,w,t)$, where

$$W_0(v,w,t)\log v=\frac{v^\beta-w^\beta}{\beta}-\frac{v^\beta\cos\delta_v-w^\beta\cos\delta_w}{\sqrt{\beta^2+t^2}}$$
$$-2\varepsilon_w\big(v^\beta+w^\beta\big)-\varepsilon_w(\alpha+|s|)\Big(\frac{v^\beta-w^\beta}{\beta}\Big).$$

Proof. We apply Lemma 2.3.5 with $s = 1-\beta$ and $s = 1-\beta+it$, and take the real part of the difference. Letting the difference of the sums be S, we have that

$$\begin{aligned} S :&= \sum_{w<p\le v} \Big(\frac{\log p}{p^{1-\beta}} - \frac{\log p}{p^{1-\beta+it}}\Big) = \sum_{w<p\le v} \frac{\log p}{p^{1-\beta}}(1-p^{-it}), \text{ so} \\ \mathfrak{Re}(S) &= \sum_{w<p\le v} \frac{\log p}{p^{1-\beta}}(1-\cos(t\log p)), \end{aligned}$$

which is the sum we wish to bound.

For a positive real number z, let $S_z := \frac{z^\beta}{\beta} - \frac{z^{\beta-it}}{\beta-it}$. We have that

$$\begin{aligned} S_z &= \frac{z^\beta}{\beta}\Big(1-\frac{\beta}{\beta-it}z^{-it}\Big) = \frac{z^\beta}{\beta}\Big(1-\beta\frac{\beta+it}{\beta^2+t^2}e^{-it\log z}\Big) \\ &= \frac{z^\beta}{\beta}\Big(1-\beta\frac{\beta+it}{\beta^2+t^2}[\cos(t\log z)-i\sin(t\log z)]\Big), \end{aligned}$$

so by Lemma 2.3.6,

$$\begin{aligned} \mathfrak{Re}(S_z) &= \frac{z^\beta}{\beta}\Big(1-\frac{\beta}{\beta^2+t^2}[\beta\cos(t\log z)+t\sin(t\log z)]\Big) \\ &= \frac{z^\beta}{\beta}\Big(1-\frac{\beta}{\sqrt{\beta^2+t^2}}\Big[\frac{\beta\cos(t\log z)}{\sqrt{\beta^2+t^2}}+\frac{t\sin(t\log z)}{\sqrt{\beta^2+t^2}}\Big]\Big) \\ &= \frac{z^\beta}{\beta}\Big(1-\frac{\beta}{\sqrt{\beta^2+t^2}}\cos(t\log z+\arctan(\beta/t))\Big) = \frac{z^\beta}{\beta}\Big(1-\frac{\beta\cos\delta_z}{\sqrt{\beta^2+t^2}}\Big). \end{aligned}$$

Thus,

$$\mathfrak{Re}(S_v - S_w) = \frac{v^\beta - w^\beta}{\beta} - \frac{v^\beta\cos\delta_v - w^\beta\cos\delta_w}{\sqrt{\beta^2+t^2}}. \tag{2.8}$$

Recalling the definition of $F_s(v,w)$, we have

$$\begin{aligned} \mathfrak{Re}(S) &= \mathfrak{Re}(S_v - S_w + F_\alpha(v,w) - F_s(v,w)) \\ &\ge \mathfrak{Re}(S_v - S_w) - |F_\alpha(v,w)| - |F_s(v,w)| \end{aligned}$$

which gives the desired result by (2.8) and Lemma 2.3.5. □

From Lemma 2.3.4, we see that a goal is to bound $W(y,1,t)$ from below, and pieces of this sum are bounded by Lemma 2.3.6. Ideally, if y were sufficiently small W could be computed directly and the problem settled. In practice W might only be computed up to some convenient number L, suitable for numerical integration, after which the analytic bound $W_0(y,w,t)$ may be used. Still, there are further refinements to be made. Just as $x/\log x$ loses out to $\mathrm{li}(x)$, W_0 on a long interval is smaller than W_0 summed on a partition of the interval into shorter parts. This plan is reflected in the following lemma.

Lemma 2.3.7. *If v, w satisfy the hypotheses of Lemma 2.3.5, let*

$$W_*(v,w,t) := W_0(v/e^{\lfloor \log(y/w) \rfloor}, w, t) + \sum_{j=0}^{\lfloor \log(v/w) \rfloor - 1} W_0(v/e^j, v/e^{j+1}, t).$$

Suppose that w, L satisfy $1427, L \le w$. If $y \le 10^{19}$, then

$$J_1 \le \int_{T_0}^{T} \exp\big(- W_*(y,w,t) - W(L,1,t)\big) \, \frac{dt}{\sqrt{\alpha^2+t^2}}.$$

If $y > e^{55}$ and $1427, L \le w \le 10^{19}$, let

$$\begin{aligned} &W_1 = W_*(10^{19}, w, t), \quad W_2 = W_*(e^{45}, 10^{19}, t), \quad W_3 = W_*(e^{50}, e^{45}, t), \\ &W_4 = W_*(e^{55}, e^{50}, t), \quad W_5 = W_*(y, e^{55}, t). \end{aligned}$$

Then

$$J_1 \le \int_{T_0}^{T} \exp\big(- W_1 - W_2 - W_3 - W_4 - W_5 - W(L,1,t)\big) \, \frac{dt}{\sqrt{\alpha^2+t^2}}.$$

We remark that if $10^{19} < y \le e^{55}$, then there is an appropriate inequality for J_1 involving fewer W_j's. If y is much larger than our largest example of $y = 10^{35}$, one might wish to use better approximations to $\vartheta(y)$ than were used in Proposition 2.2.1.

Proof. If $1427 \le w < v$ and $[w, v]$ satisfy the hypotheses of Lemma 2.3.5, we have

$$\begin{aligned} W(v,w,t) &= W(v/e^{\lfloor \log(v/w) \rfloor}, w, t) + \sum_{j=0}^{\lfloor \log(v/w) \rfloor - 1} W(v/e^j, v/e^{j+1}, t) \\ &\ge W_0(v/e^{\lfloor \log(v/w) \rfloor}, w, t) + \sum_{j=0}^{\lfloor \log(v/w) \rfloor - 1} W_0(v/e^j, v/e^{j+1}, t). \end{aligned}$$

The result then follows from Lemma 2.3.4. □

Remark 2.3.8. We implement Lemma 2.3.7 by choosing L as large as possible so as not to interfere overly with numerical integration. We have found that $L = 10^6$ works well. The ratio e in the definition of W_* is convenient, but might be tweaked for slightly better results. The individual terms in the sum $W(L,1,t)$ are as in (2.6), except for the first 30 primes, where instead we forgo using the inequality in (2.7), using instead the slightly larger expression

$$\frac{1}{2} \log \Big(1 + \frac{2(1 - \cos(t \log p))}{p^{\alpha}(1 - p^{-\alpha})^2} \Big).$$

We choose w as a function $w(t)$ in such a way that the bound in Lemma 2.3.6 is minimized. For simplicity, we ignore the oscillating terms, i.e., we set

$$\frac{\partial}{\partial w}\Big[-w^\beta/\beta - 4w^{1/2-\alpha} + 2(\alpha+|s|)w^{1/2-\alpha}/(1/2-\alpha)\Big]$$
$$= -w^{\beta-1} - 4w^{-1/2-\alpha}/(1/2-\alpha) + 2(\alpha+|s|)w^{-1/2-\alpha}$$

equal to 0. Multiplying by $w^{1/2+\alpha}$ and solving for w gives

$$w(\alpha,t) := \Big(\frac{4}{\alpha-1/2} + 2\alpha + 2\sqrt{\alpha^2+t^2}\Big)^2.$$

We let

$$w(t) := \max\{L, w(\alpha,t)\}.$$

Our next result, based on [24, Lemma 9], gives a bound on the number of y-smooth integers in a short interval.

Lemma 2.3.9. *Let* $0 < d < 1$, $T > 1$ *be such that* $z := (e^{2T^{d-1}} - 1)^{-1} > 1$. *We have*

$$\Psi(xe^{T^{d-1}}, y) - \Psi(xe^{-T^{d-1}}, y) \le e^{\alpha^2/2z^2 - \alpha T^{d-1}} x^\alpha \zeta(\alpha,y)\sqrt{\frac{2e}{\pi}\frac{J_2}{z}}.$$

where, with $W(y,w,t)$ *as in Lemma 2.3.6,*

$$J_2 := \int_0^\infty \exp\Big\{-\frac{t^2}{2z^2} - W(y,1,t)\Big\}\,dt.$$

Proof. Let $\xi = xe^{-T^{d-1}}$, so that

$$\Psi(xe^{T^{d-1}}, y) - \Psi(xe^{-T^{d-1}}, y) = \Psi(\xi + \xi/z, y) - \Psi(\xi, y). \tag{2.9}$$

For $\xi < n \le \xi + \frac{\xi}{z}$, we have that

$$1 > \frac{\xi}{n} \ge \Big(1 + \frac{1}{z}\Big)^{-1},$$

so $0 > \log(\xi/n) \ge -\log(1+1/z) \ge -\frac{1}{z}$, which implies that $0 < [z\log(\xi/n)]^2 \le 1$. Thus,

$$\Psi(\xi+\xi/z, y) - \Psi(\xi,y) = \sum_{\substack{P(n)\le y\\ \xi<n\le\xi+\xi/z}} 1 \le \sqrt{e}\sum_{\substack{P(n)\le y\\ \xi<n\le\xi+\xi/z}} \exp\Big\{-\frac{1}{2}[z\log(\xi/n)]^2\Big\}.$$

For $\sigma, v \in \mathbb{R}$, we have the formula

$$e^{-v^2/2} = \frac{1}{\sqrt{2\pi}} e^{\sigma^2/2-\sigma v}\int_{-\infty}^{+\infty} \exp\Big\{-\frac{1}{2}t^2 + it(\sigma - v)\Big\}\,dt$$

Letting $\sigma = \alpha/z$, $v = -z\log(\xi/n)$, we obtain

$$\begin{aligned}\Psi(\xi+\xi/z,y)-&\Psi(\xi,y)\\ &\le\sqrt{\frac{e}{2\pi}}\sum_{\substack{P(n)\le y\\ \xi<n\le\xi+\xi/z}} e^{\sigma^2/2-\sigma v}\int_{-\infty}^{+\infty}\exp\Big\{-\frac{1}{2}t^2+it(\sigma-v)\Big\}\,dt\\ &=e^{\alpha^2/2z^2}\sqrt{\frac{e}{2\pi}}\int_{-\infty}^{+\infty}\exp\Big\{-\frac{1}{2}t^2+it\alpha/z\Big\}\sum_{\substack{P(n)\le y\\ \xi<n\le\xi+\xi/z}}\Big(\frac{\xi}{n}\Big)^{\alpha+itz}\,dt.\end{aligned}$$

Since $\alpha \le 1 \le z$, changing variables $t \mapsto t/z$ and taking the modulus gives

$$\begin{aligned}\Psi(\xi+\xi/z,y)-\Psi(\xi,y)&\\ &\le z^{-1}e^{\alpha^2/2z^2}\sqrt{\frac{e}{2\pi}}\int_{-\infty}^{+\infty}\exp\Big\{-\frac{t^2}{2z^2}+it\alpha/z^2\Big\}\sum_{\substack{P(n)\le y\\ \xi<n\le\xi+\xi/z}}\Big(\frac{\xi}{n}\Big)^{\alpha+it}\,dt\\ &\le\frac{\xi^\alpha}{z}e^{\alpha^2/2z^2}\sqrt{\frac{e}{2\pi}}\int_{-\infty}^{+\infty}e^{-t^2/2z^2}|\zeta(\alpha+it,y)|\,dt\\ &=\frac{\xi^\alpha}{z}e^{\alpha^2/2z^2}\sqrt{\frac{2e}{\pi}}\int_{0}^{\infty}e^{-t^2/2z^2}|\zeta(\alpha+it,y)|\,dt.\end{aligned}$$

This last integral may be estimated by the method of Lemma 2.3.4, giving

$$\int_0^\infty e^{-t^2/2z^2}|\zeta(\alpha+it,y)|\,dt \le \zeta(\alpha,y)\int_0^\infty \exp\Big(-\frac{t^2}{2z^2}-W(y,1,t)\Big)\,dt = \zeta(\alpha,y)J_2.$$

We have

$$\Psi(\xi+\xi/z,y)-\Psi(\xi,y)\le \xi^\alpha\zeta(\alpha,y)e^{\alpha^2/2z^2}\sqrt{\frac{2e}{\pi}}\frac{J_2}{z},$$

and the lemma now follows from (2.9) and the definition of ξ. □

Remark 2.3.10. For t large, say $t > 2z\log z$, we can ignore the term $W(y,1,t)$ in J_2, getting a suitably tiny numerical estimate for the tail of this rapidly converging integral. The part for t small may be integrated numerically with $w(t), L$ as in Remark 2.3.8.

With these lemmas, we now have our principal result.

Theorem 2.3.11. *Let d, T, z be as in Lemma 2.3.9, let $J_0^\pm$ be as in* (2.5)*, J_1 as in Lemma 2.3.4, and J_2 as in Lemma 2.3.9. We have*

$$\Psi(x,y)\ge\frac{x^\alpha\zeta(\alpha,y)}{\pi}\Big(J_0^- - J_1 - T^{-d} - e^{\alpha^2/2z^2}\sqrt{2\pi e}\frac{J_2}{z}\Big)$$

and

$$\Psi(x,y)\le\frac{x^\alpha\zeta(\alpha,y)}{\pi}\Big(J_0^+ + J_1 + T^{-d} + e^{\alpha^2/2z^2}\sqrt{2\pi e}\frac{J_2}{z}\Big).$$

2.4 Computations

In this section we give some guidance on how, for a given pair x, y, the numbers α, $\zeta(\alpha, y)$, and σ_j for $j \le 5$ may be numerically approximated. Further, we discuss how these data may be used to numerically approximate $\Psi(x, y)$ via Theorem 2.3.11.

2.4.1 Computing α

Given a number $a \in (0, 1)$ and a large number y we may obtain upper and lower bounds for the sum

$$\sigma_1(a, y) = \sum_{p \le y} \frac{\log p}{p^a - 1}.$$

First, we choose a moderate bound $w_0 \le y$ where we can compute the sum $\sigma_1(a, w_0)$ relatively easily, such as $w_0 = 179{,}424{,}673$, the ten-millionth prime. The sum

$$\sum_{w_0 < p \le y} \frac{\log p}{p^a} \tag{2.10}$$

may be approximated easily with Proposition 2.2.1 and partial summation. Let $l^-(a, w_0, y)$ be a lower bound for this sum and let $l^+(a, w_0, y)$ be an upper bound. Then

$$l^-(a, w_0, y) + \sigma_1(a, w_0) \le \sigma_1(a, y) \le \frac{w_0^a}{w_0^a - 1} l^+(a, w_0, y) + \sigma_1(a, w_0).$$

We choose α as a number a where $\log x$ lies between these two bounds. If a given trial for a is too small, this is detected by our lower bound for $\sigma_1(a, y)$ lying above $\log x$, and if a is too large, we see this if our upper bound for $\sigma_1(a, y)$ lies below $\log x$. It does not take long via linear interpolation to find a reasonable choice for α. While narrowing in, one might use a less ambitious choice for w_0.

The partial summation used to estimate (2.10) and similar sums may be summarized in the following result.

Lemma 2.4.1. *Suppose $f(t)$ is positive and $f'(t)$ is negative on $[w_0, w_1]$. Suppose too that $t - 2\sqrt{t} < \vartheta(t) \le t$ on $[w_0, w_1]$. Then*

$$\int_{w_0}^{w_1} (1 - 1/\sqrt{t}) f(t)\, dt + (w_0 - \vartheta(w_0) - 2\sqrt{w_0}) f(w_0)$$
$$\le \sum_{w_0 < p \le w_1} f(p) \log p \le \int_{w_0}^{w_1} f(t)\, dt + (w_0 - \vartheta(w_0)) f(w_0).$$

Because of Proposition 2.2.1, the condition on ϑ holds if $[w_0, w_1] \subset [1427, 10^{19}]$. For intervals beyond 10^{19}, it is easy to fashion an analogue of Lemma 2.4.1 using the other estimates of Proposition 2.2.1.

2.4.2 Computing $\sigma_0 = \log \zeta(\alpha, y)$ and the other σ_j's

Once a choice for α is computed it is straightforward to compute σ_0 and the other σ_j's.

We have

$$\sigma_0(\alpha, y) = \sum_{p \le y} -\log(1 - p^{-\alpha}).$$

We may compute this sum up to some moderate w_0 as with the α computation. For the range $w_0 < p \le y$ we may approximate the summand by $p^{-\alpha}$ and sum this over $(w_0, y]$ using partial summation (Lemma 2.4.1) and Proposition 2.2.1, say a lower bound is l_0^- and an upper bound is l_0^+. Then

$$l_0^- + \sigma_0(\alpha, w_0) \le \sigma_0(\alpha, y) \le \frac{-\log(1 - w_0^{-\alpha})}{w_0^{-\alpha}} l_0^+ + \sigma_0(\alpha, w_0).$$

The other σ_j's are computed in a similar manner.

2.4.3 Data

We record our calculations of α and the numbers σ_j for two examples. Note that we obtain bounds for ζ via $\sigma_0 = \log \zeta$.

Figure 2.2: Data.

x	10^{100}	10^{500}
y	10^{15}	10^{35}
α	.9111581	.94932677
ζ	$352{,}189 \pm 16$	$2.09222 \cdot 10^{10} \pm 5 \cdot 10^5$
σ_1^*	$4.3 \cdot 10^{-4}$	$5.6 \cdot 10^{-4}$
σ_2	$5{,}763.47 \pm 0.03$	$71{,}689.2 \pm 0.02$
σ_3	$159{,}066.8 \pm 0.5$	$4{,}779{,}948.5 \pm 0.5$
σ_4	$4{,}604{,}079 \pm 8$	$330{,}260{,}722 \pm 21$
σ_5^+	$1.3725 \cdot 10^8$	$2.3353 \cdot 10^{10}$

Note that σ_1^* is an upper bound for $|\sigma_1 - \log x|$, and σ_5^+ is an upper bound for σ_5.

The functions $\alpha(x, y)$ and $\sigma_j(x, y)$ are of interest in their own right. A simple observation from their definitions allows for more general bounds on α and σ_j using the data in Figure 2.2, as described in the following remark.

Remark 2.4.2. For pairs x, y and x', y', if $x \ge x'$ and $y \le y'$ then $\alpha(x, y) \le \alpha(x', y')$. Similarly, if $\alpha(x, y) \ge \alpha(x', y')$ and $y \le y'$ then $\sigma_j(x, y) \le \sigma_j(x', y')$.

2.4.4 A word on numerical integration

The numerical integration needed to estimate J_1, J_2 is difficult, especially when we choose a large value of L, like $L = 10^6$. We performed these integrals independently on both Mathematica and Sage platforms. It helps to segment the range of integration, but even so, the software can report an error bound in addition to the main estimate. In such cases we have always added on this error bound and then rounded up, since we seek upper bounds for these integrals. In a case where one wants to be assured of a rigorous estimate, there are several options, each carrying some costs. One can use a Simpson or midpoint quadrature with a mesh say of 0.1 together with a careful estimation of the higher derivatives needed to estimate the error. An alternative is to do a Riemann sum with mesh 0.1, where on each interval and for each separate cosine term appearing, the maximum contribution is calculated. If this is done with $T = 4 \cdot 10^5$ and $L = 10^6$, there would be magnitude 10^{11} of these calculations. The extreme value of the cosine contribution would either be at an endpoint of an interval or -1 if the argument straddles a number that is $\pi \bmod 2\pi$. We have done a mild form of this method in our estimation of the integrals $J_0^{\pm}$, see the discussion leading up to Lemma 2.3.3.

2.4.5 Example estimates

We list some example values of x, y and the corresponding estimates in the figure below.

Figure 2.3: Results.

x	10^{100}	10^{500}
y	10^{15}	10^{35}
T_3	.00642708	.00114940
T_2	.00644109	.00115038
Z^-	.0385260	.0124202
Z^+	.0403125	.0127461
T_1	.0478624	.0155272
T_0	.0514483	.0161799
T	$4 \cdot 10^5$	10^9
d	0.57	0.58
J_0^-	$1.78554 \cdot 10^{-2}$	$4.90043 \cdot 10^{-3}$
J_0^+	$1.80312 \cdot 10^{-2}$	$4.92738 \cdot 10^{-3}$
J_1	$7.236 \cdot 10^{-4}$	$1.717 \cdot 10^{-6}$
J_2	$1.758 \cdot 10^{-2}$	$4.745 \cdot 10^{-3}$
Ψ^-	$2.3302 \cdot 10^{94}$	$1.4989 \cdot 10^{482}$
Ψ^+	$2.9227 \cdot 10^{94}$	$1.5118 \cdot 10^{482}$

2.5 Appendix: Rankin revisited

We prove the following theorem.

Theorem 2.5.1 (Granville and Soundararajan)**.** *If $3 \le y \le x$ and $1/\log y \le \sigma \le 1$, then*

$$\Psi(x,y) \le 1.39 \frac{y^{1-\sigma}}{\log x} x^\sigma \zeta(\sigma, y).$$

Proof. By the identity $\log n = \sum_{d|n} \Lambda(d)$, we have

$$\sum_{\substack{n \le x \\ P(n) \le y}} \log n = \sum_{\substack{m \le x \\ P(m) \le y}} \sum_{\substack{d \le x/m \\ P(d) \le y}} \Lambda(d) = \sum_{\substack{m \le x \\ P(m) \le y}} \sum_{p \le \min\{y, x/m\}} \log p \left\lfloor \frac{\log(x/m)}{\log p} \right\rfloor$$
$$\le \sum_{\substack{m \le x \\ P(m) \le y}} \pi\big(\min\{y, x/m\}\big) \log(x/m).$$

Thus,

$$\Psi(x,y)\log x = \sum_{\substack{n \le x \\ P(n) \le y}} (\log n + \log(x/n)) \le \sum_{\substack{n \le x \\ P(n) \le y}} \big(1 + \pi(\min\{y, x/n\})\big) \log(x/n).$$

Using the estimates in [42] we see that the maximum of $(1+\pi(t))/(t/\log t)$ occurs at $t = 7$, so that

$$1 + \pi(t) < 1.39t/\log t$$

for all $t > 1$. The above estimate then gives

$$\Psi(x,y)\log x < 1.39 \sum_{\substack{x/y<n\le x\\ P(n)\le y}} x/n + 1.39 \sum_{\substack{n\le x/y\\ P(n)\le y}} y\log(x/n)/\log y.$$

We now note that if $1/\log y \le \sigma \le 1$, then

$$y^{1-\sigma}(x/n)^{\sigma} \ge \begin{cases} x/n, & \text{if } x/y < n \le x,\\ y\log(x/n)/\log y, & \text{if } n \le x/y.\end{cases}$$

Indeed, in the first case, since $t^{1-\sigma}$ is non-decreasing in t, we have $(x/n)^{1-\sigma} \le y^{1-\sigma}$. And in the second case, since $t^{-\sigma}\log t$ is decreasing in t for $t \ge y$, we have $(x/n)^{-\sigma}\log(x/n) \le y^{-\sigma}\log y$.

We thus have

$$\Psi(x,y)\log x < 1.39 \sum_{\substack{n\le x\\ P(n)\le y}} y^{1-\sigma}(x/n)^{\sigma} < 1.39y^{1-\sigma}x^{\sigma}\zeta(\sigma,y).$$

This completes the proof. □

Chapter 3

The reciprocal sum of primitive nondeficient numbers

3.1 Introduction

The ancients were enamored by numbers that were equal to the sum of their own proper divisors, and hailed such numbers as perfect. In modern notation, $\sigma(n)$ denotes the sum of divisors of n, so n is perfect if $\sigma(n) = 2n$. Then n is called abundant if $\sigma(n) > 2n$, and deficient if $\sigma(n) < 2n$. One can show that if n is nondeficient, that is, either perfect or abundant, then multiples of n are also nondeficient. Thus, we are led to define a *primitive nondeficient number* (pnd) to be a nondeficient number all of whose proper divisors are deficient. Often in the literature the alternate term primitive abundant number (pan) is used, justified by redefining the term abundant to signify numbers n with $\sigma(n) \geq 2n$.

With the goal of understanding the distribution of $\sigma(n)/n$, much effort has been put toward studying the set of pnds. For example, Erdős [16] found an elementary proof that the set of abundant has a natural density which hinges on showing that the reciprocal sum of pnds converges. The convergence was shown by determining a sufficiently small upper bound on the counting function for pnds. Denoting the number of pnds $\leq x$ by $N(x)$, the paper showed that

$$N(x) = o\left(\frac{x}{(\log x)^2}\right),$$

which is enough to prove that the sum of reciprocals of the pnds converges. A more detailed study by Erdős in [17] found that, for sufficiently large x,

$$x\exp(-c_1\sqrt{\log x \log\log x}) \leq N(x) \leq x\exp(-c_2\sqrt{\log x \log\log x}),$$

where $c_1 = 8$ and $c_2 = 1/25$. Subsequent improvements to the constants c_1, c_2 were made by Ivić [26] and Avidon [3] so that we now know that we may take $c_1 = \sqrt{2}+\varepsilon$, $c_2 = 1-\varepsilon$ for any fixed $\varepsilon > 0$ which can be made arbitrarily small by taking sufficiently large x.

Once a series is found to converge, it is natural to wonder what its value may be. For example, by Brun's Theorem, it is known that the reciprocal sum of twin primes converges. This sum, called Brun's constant, is approximately 1.902160583104, which is found by extrapolating via the Hardy–Littlewood heuristics. However, the best proven upper bound is 2.347. (See [13, 28].) Similarly, Pomerance [39] proved that the reciprocal sum of numbers in amicable pairs converges, and work has also been done to determine bounds on this value, the Pomerance constant, the current bounds being 0.0119841556 and 222. (See [5, 34, 35]) In light of such results, it is somewhat surprising that there do not appear to be any serious attempts to determine the value of the reciprocal sum of pnds. The purpose of this paper is to initiate such an attempt. Given such precedent, we define the value of the reciprocal sum of pnds the *Erdős constant*. The principal result of this paper is to provide the first known upper bound for the Erdős constant as 18.6, proved in Theorem 3.3.2.

Notation

Let $\sum_n' = \sum_{\text{pnd } n}$. Let $\log_k n$ denote the k-fold logarithm $\log\log\cdots\log n$. Let $\zeta(s)$ be the Riemann zeta function. Let $P(n), \omega(n)$ denote the largest prime factor of n, and number of distinct prime factors of n, respectively. We say a positive integer n is y-smooth if $P(n) \le y$. We say n is square-full if $p^2 \mid n$ for all primes $p \mid n$.

3.2 Setting up the bound

To bound the reciprocal sum of pnds,

$$\sum_n{}' \frac{1}{n} = \sum_{n\le x_0}{}' \frac{1}{n} + \sum_{n>x_0}{}' \frac{1}{n} \tag{3.1}$$

we may first compute the sum directly up to some convienient $x_0 \in \mathbb{Z}$. For example, Mits Kobayashi computed that for $x_0 = 10^{10}$

$$\sum_{n\le 10^{10}}{}' \frac{1}{n} = 0.34816486577357275, \tag{3.2}$$

as well as that the number of pnds up to 10^{10} is 1123430. For the remaining part of the series, we shall split up by y-smoothness, for y to be determined. Before proceeding, we first study the related sum

$$M(x,y) = \sideset{}{'}\sum_{\substack{n\le x\\ P(n)>y}} 1,$$

which will prove useful for handling the non-y-smooth contribution.

3.2.1 An upper bound for $M(x,y)$

We first prove a preliminary lemma for square-full numbers.

Lemma 3.2.1. *For $\lambda = \zeta(3/2)/\zeta(3)$, we have*

$$B(x,y) := \sum_{\substack{n\le x\\ s(n)\ge y}} 1 \le \lambda x y^{-1/2} + 3xy^{-2/3}.$$

Proof. Denoting $K(y)$ as the number of square-full integers up to y, by (8) in [20] we have

$$-3\sqrt[3]{y} \le K(y) - \lambda\sqrt{y} \le 0, \qquad \lambda = \frac{\zeta(3/2)}{\zeta(3)}. \tag{3.3}$$

For each square-full number $s \in [y,x]$, there are at most x/s such pnds n up to x with $s = s(n)$. Then by partial summation we have

$$\sum_{\substack{n\le x\\ s(n)\ge y}} 1 \le \sum_{\substack{s>y\\ s\ \square\text{-full}}} \frac{x}{s} = x\Big(-\frac{K(y)}{y} + \int_y^\infty \frac{K(t)}{t^2}\,dt\Big) \le x\Big(-\frac{\lambda y^{1/2} - 3y^{1/3}}{y} + \lambda\int_y^\infty \frac{dt}{t^{3/2}}\Big)$$
$$= x\Big(-\lambda y^{-1/2} + 3xy^{-2/3} + \lambda\Big[-2t^{-1/2}\Big]_y^\infty\Big) = \lambda x y^{-1/2} + 3xy^{-2/3}.$$

which completes the proof. ∎

Now to bound $M(x,y)$, we roughly follow the developments in [17,26], and split up into two cases. Every integer n may be decomposed uniquely into $n = qs$, where $s = s(n)$ is square-full, $q = q(n)$ is square-free, and $(s,q) = 1$.

In the first case, suppose $s(n) \ge y^a$ for some parameter $a \in (0, \frac{1}{3})$ to be determined later. Then by Lemma 3.2.1, we have

$$\sideset{}{'}\sum_{\substack{n\le x\\ P(n)>y\\ s(n)\ge y^a}} 1 \le \sideset{}{'}\sum_{\substack{n\le x\\ s(n)\ge y^a}} 1 = B(x,y^a) \le \lambda x y^{-a/2} + 3xy^{-2a/3}. \tag{3.4}$$

In the second case, we have $s(n) < y^a$. We prove a lemma, adapted from [17,26].

Lemma 3.2.2. *Assume $x > y > 8$. Let $\{n_1, \ldots, n_m\}$ be the set of pnds $n_i \leq x$ with $P(n_i) > y$ and square-full part $s(n_i) < y^a$. Let $b = 3a/2$ and suppose that b satisfies*

$$b \in \left(0, \tfrac{1}{2}\right), \qquad 2 \geq \left(2 - y^{-b}\right)\left(1 + \sqrt{2}y^{-1/2}\right)^{2\log x/\log(y/2)}. \tag{3.5}$$

Then for each n_i there exists a square-free divisor $d_i \mid q(n_i)$ with $d_i \in [y^{b/3}, \frac{1}{\sqrt{2}}y^{1/2}]$. Moreover, $\{n_1/d_1, \ldots, n_m/d_m\}$ is a set of m distinct numbers at most $xy^{-b/3}$, and therefore

$$\sideset{}{'}\sum_{\substack{n\leq x\\ P(n)>y\\ s(n)<y^a}} 1 = m \leq xy^{-b/3}. \tag{3.6}$$

Proof. To prove existence, we proceed by contradiction. Take any $n = n_i$, and suppose there is no divisor $d \mid q(n)$ with $d \in [y^{b/3}, \frac{1}{\sqrt{2}}y^{1/2}]$. Since $s(n) < y^a = y^{2b/3}$, this is equivalent to having no prime divisors $p \mid n$ in the interval $[y^{b/3}, \frac{1}{\sqrt{2}}y^{1/2}]$. Then we may decompose n as $n = uv$ where $r < y^{b/3}$ and $p > \frac{1}{\sqrt{2}}y^{1/2}$ for all primes $r \mid u$ and $p \mid v$, respectively. We have that

$$\begin{aligned}\omega(v) = \sum_{p|v} 1 \leq \sum_{p|v} \frac{\log p}{\log(y/2)/2} &= \frac{2}{\log(y/2)} \log \prod_{p|v} p \\ &= \frac{2\log v}{\log(y/2)} \leq \frac{2\log x}{\log(y/2)}.\end{aligned}$$

Suppose $u \leq y^b$, where we recall that b satisfies (3.5). Also recall $y < P(n) \leq n$ so $u < n$ is a proper divisor of the pnd n. Thus u is deficient so $\sigma(u) \leq 2u - 1$, and since the function $h(n) = \sigma(n)/n$ is multiplicative,

$$\begin{aligned}2 \leq h(n) = h(u)h(v) &\leq \left(2 - \frac{1}{u}\right) \prod_{p|v} \left(1 + \frac{1}{p}\right) \\ &\leq \left(2 - y^{-b}\right)\left(1 + \sqrt{2}y^{-1/2}\right)^{\omega(v)} \\ &< \left(2 - y^{-b}\right)\left(1 + \sqrt{2}y^{-1/2}\right)^{2\log x/\log(y/2)}.\end{aligned}$$

This contradicts the assumption (3.5) for b. Hence we deduce $u > y^b$.

Write the square-free part of u as $q(u) = q_1 q_2 \cdots q_t$ in ascending order of primes. Since $s(n) < y^a < \frac{1}{\sqrt{2}}y^{1/2}$, all the primes of $s(n)$ are less than $\frac{1}{\sqrt{2}}y^{1/2}$. Therefore $s(n) \mid u$ and so $s(n) = s(u)$. Additionally, we have $u > y^b$ so that

$$q(u) = \frac{u}{s(u)} = \frac{u}{s(n)} > y^{b-a} = y^{b/3}.$$

Since each $q_i \mid u$ we have each $q_i < y^{b/3}$, so there must exist $l \in [1, t]$ such that

$$q_1 \cdots q_{l-1} \le y^{b/3} < q_1 \cdots q_l < y^{2b/3}.$$

Since $b < 1/2$ and $y > 8$, we have $y^{2b/3} < \frac{1}{\sqrt{2}} y^{1/2}$. Thus $d = q_1 \cdots q_l$ is a divisor of n in the interval $[y^{b/3}, \frac{1}{\sqrt{2}} y^{1/2}]$. However, this contradicts assumption, and thus each $q(n_i)$ has a divisor d_i in the interval.

The proof of distinctness is unchanged from [17, 26], but we provide it for completeness. For all $n = n_i$, since the square-full part $s(n)$ is less than $y^a < y < P(n)$ we have that $P(n)^2$ does not divide n so

$$\begin{aligned} 2 \le h(n) &= h(P(n)) h\Big(\frac{n}{P(n)}\Big) \\ &= \Big(1 + \frac{1}{P(n)}\Big) h\Big(\frac{n}{P(n)}\Big) \le 2 + \frac{2}{P(n)} < 2 + 2/y. \end{aligned}$$

Thus for all n_i, n_j we have

$$\frac{h(n_i)}{h(n_j)} < \frac{2 + 2/y}{2} = 1 + 1/y. \tag{3.7}$$

Suppose $n_i/d_i = n_j/d_j$ for some $i \ne j$. Since $n_i \ne n_j$ we have $d_i \ne d_j$. Then by multiplicativity,

$$\frac{h(n_i)}{h(d_i)} = h\Big(\frac{n_i}{d_i}\Big) = h\Big(\frac{n_j}{d_j}\Big) = \frac{h(n_j)}{h(d_j)}.$$

Since d_i and d_j are square-free, we have that $h(d_i) \ne h(d_j)$ so we may assume $h(d_i) > h(d_j)$. Thus

$$1 < \frac{h(d_i)}{h(d_j)} = \frac{\sigma(d_i) d_j}{\sigma(d_j) d_i}$$

so that $\sigma(d_i) d_j \ge \sigma(d_j) d_i + 1$. Since d_i divides the pnd n_i, d_i is deficient so $\sigma(d_i) < 2d_i$. And since $d_i \le \frac{1}{\sqrt{2}} y^{1/2}$ we deduce

$$\begin{aligned} \frac{h(n_i)}{h(n_j)} = \frac{h(d_i)}{h(d_j)} = \frac{\sigma(d_i) d_j}{\sigma(d_j) d_i} &\ge 1 + \frac{1}{\sigma(d_j) d_i} \\ &> 1 + \frac{1}{2 d_i d_j} > 1 + 1/y. \end{aligned}$$

contradicting (3.7). Hence each n_i/d_i must be distinct. □

Combining (3.4) and Lemma (3.2.2) gives our desired bound on $M(x, y)$.

Theorem 3.2.3. *Assume $x > y > 8$. Let $b = b(x, y)$ be defined by*

$$y^{-b} = 2 - 2\left(1 + \sqrt{2/y}\right)^{-2\log x/\log(y/2)}. \tag{3.8}$$

Then so long as $0 < b < \frac{1}{2}$, we have the upper bound

$$M(x, y) = \sideset{}{'}\sum_{\substack{n\le x\\ P(n)>y}} 1 \le (\lambda + 1)xy^{-b/3} + 3xy^{-4b/9}. \tag{3.9}$$

Proof. The definition of b is constructed to satisfy (3.5). By (3.4) and Lemma 3.2.2 we have

$$\begin{aligned} M(x, y) &= \sideset{}{'}\sum_{\substack{n\le x\\ P(n)>y\\ s(n)>y^a}} 1 + \sideset{}{'}\sum_{\substack{n\le x\\ P(n)>y\\ s(n)\le y^a}} 1 \\ &\le \lambda xy^{-a/2} + 3xy^{-2a/3} + xy^{-b/3}. \end{aligned}$$

The result then follows from $a = 2b/3$. □

The utility of Theorem 3.2.3 to us comes as the following Corollary.

Corollary 3.2.3.1. *With b as in Theorem 3.2.3, for $y > 8$ we have that*

$$\sideset{}{'}\sum_{\substack{x_1\le n\le x_2\\ P(n)>y}} \frac{1}{n} \le (1 + \log(x_2/x_1))[(\lambda + 1)y^{-b/3} + 3y^{-4b/9}].$$

Proof. Since $C = (\lambda + 1)y^{-b/3} + 3y^{-4b/9}$ is constant with respect to x, by partial summation and Theorem 3.2.3,

$$\begin{aligned} \sideset{}{'}\sum_{\substack{x_1\le n\le x_2\\ P(n)>y}} \frac{1}{n} &= \frac{M(x_2, y)}{x_2} - \frac{M(x_1, y)}{x_1} + \int_{x_1}^{x_2} M(x, y)\frac{dx}{x^2} \\ &\le C + C\int_{x_1}^{x_2} \frac{dx}{x} = (1 + \log(x_2/x_1))C. \end{aligned}$$

□

3.2.2 Bounding the tail

Recall the contribution of pnds less than 10^{10} was computed directly. On the other end, we may bound the tail of the reciprocal sum of pnds greater than e^{1600}. In this range, we bifurcate based on the relative size of $\omega(n)$ compared to $4\log_2(n)$. Note that if $\omega(n) \le 4\log_2(n)$, then there exists a prime power $q^a \mid n$ such that

$$q^a \ge n^{1/4\log_2(n)} = \exp\left(\frac{\log n}{4\log_2 n}\right) =: y(n).$$

If $a = 1$, then $P(n)$ is large, and if $a \geq 2$ then $s(n)$ is large. Thus it suffices to consider the following cases:

(i) $\omega(n) > 4\log_2(n)$,

(ii) $P(n) \geq y(n)$,

(iii) $s(n) \geq y(n)$.

In case (i), by Proposition 3.2 in [35],

$$\sum_{\substack{n>e^{1600} \\ \omega(n)>4\log_2 n}} \frac{1}{n} \leq \frac{1}{24} \sum_{k\geq 1600} \frac{(k+5)^4}{k^{4\log 4}} \leq \frac{1}{24} \int_{1599}^{\infty} \frac{(t+5)^4}{t^{4\log 4}}\, dt \leq 0.00138. \tag{3.10}$$

In case (ii), let $y_k = y(e^k) = \exp\left(\frac{k}{4\log k}\right)$ and define $b_k = b(e^{k+1}, y_k)$ from Theorem 3.2.3. A calculation shows that $b_k \in \left(0, \frac{1}{2}\right)$ for $k \geq 191$. Then by Corollary 3.2.3.1

$$\sum_{\substack{n>e^{1600} \\ P(n)>y(n)}} \frac{1}{n} \leq \sum_{k\geq 1600} \sum_{\substack{e^k<n\leq e^{k+1} \\ P(n)>y_k}} \frac{1}{n} \leq 2 \sum_{k\geq 1600} (\lambda+1) y_k^{-b_k/3} + 3y_k^{-4b_k/9}.$$

We may compute this sum directly up to, say, 10^4, which contributes at most 0.4471. We bound the remaining series by the integral,

$$\sum_{k\geq 10^4} (\lambda+1) y_k^{-b_k/3} + 3y_k^{-4b_k/9} \leq 6 \int_{10^4}^{\infty} \exp\left(\frac{-bt}{12\log t}\right) dt.$$

Note that the definition of $b = b_k$ in (3.8) using $x = e^{k+1}$ and $y = y_k$ ensures $b \geq 0.4$ for $k \geq 10^4$. Then since $b \geq 0.4$ and $k \geq 10^4$, we have $.7\log k \geq \log(12/b) + \log_2 k$ which implies

$$-t^{0.3} \geq \frac{-t}{(12/b)\log t}. \tag{3.11}$$

Hence we have

$$\begin{aligned} \sum_{\substack{n>e^{1600} \\ P(n)>y(n)}} \frac{1}{n} &\leq 2(0.4471) + 12 \int_{10^4}^{\infty} \exp(-t^{0.3})\, dt \\ &= 2(0.4471) + 12(0.00191572) \leq .918. \end{aligned} \tag{3.12}$$

In case (iii), by partial summation and Lemma 3.2.1,

$$\begin{aligned}
\sum_{\substack{n>e^{1600}\\ s(n)>y(n)}} \frac{1}{n} \le \sum_{k\ge 1600} \sum_{\substack{e^k\le n\le e^{k+1}\\ s(n)>y_k}} \frac{1}{n} &\le \sum_{k\ge 1600} \Big(\frac{B(e^{k+1},y_k)}{e^{k+1}} - \frac{B(e^k,y_k)}{e^k} + \int_{e^k}^{e^{k+1}} B(t,y_k)\frac{dt}{t^2}\Big) \\
&\le \sum_{k\ge 1600} (\lambda y_k^{-1/2} + 3y_k^{-2/3})\Big(1+\int_{e^k}^{e^{k+1}} \frac{dt}{t}\Big) \\
&\le 2\sum_{k\ge 1600} (\lambda y_k^{-1/2} + 3y_k^{-2/3}) \le 6\int_{1599}^{\infty} \exp\Big(\frac{-t}{8\log t}\Big)\, dt \\
&\le 6\int_{1599}^{\infty} \exp(-3t^{.3})\, dt \le 1.5\cdot 10^{-9} \qquad (3.13)
\end{aligned}$$

by a bound analogous to (3.11).

3.2.3 Intermediate Range

We are left to deal with the contribution of pnds lying in the intermediate range $10^{10} \le n \le e^{1600}$. We further split up the range at e^{300}, and first deal with the upper subrange $e^{300} \le n \le e^{1600}$ in which the bounds for $M(x,y)$ still have potency.

For smooths in the upper range, we appeal to Lemma 2.10 in [35], which we state below.

Lemma 3.2.4 (Nguyen-Pomerance)**.** *Let $x > y \ge 2$ and $u = \log x/\log y$. For $u \ge 3$ and $\log(u\log u)/\log y \le 1/3$, we have*

$$\sum_{\substack{n>x\\ P(n)\le y}} \frac{1}{n} \le \frac{25e^{(1+\varepsilon)u}(u\log u)^{-u}}{2^{\log(u\log u)/\log y}-1}$$

where $\varepsilon = 2.3\times 10^{-8}$.

We apply this bound to the interval $k\in[300,1600]$ in segments of length 5, where in each we find a reasonable value of $u = u_k$, which then determines $y_k = k/u_k$. Note y_k here is different from the previous section. Thus by Corollary 3.2.3.1,

$$\begin{aligned}
\sideset{}{'}\sum_{e^{300}\le n\le e^{1600}} \frac{1}{n} &\le \sum_{\substack{300\le k\le 1600\\ 5|k}} \Big(\sum_{\substack{n>e^k\\ P(n)\le y}} \frac{1}{n} + \sideset{}{'}\sum_{\substack{e^k\le n\le e^{k+5}\\ P(n)>y}} \frac{1}{n}\Big) \\
&\le \sum_{\substack{300\le k\le 1600\\ 5|k}} \Big(\frac{25e^{(1+\varepsilon)u_k}(u_k\log u_k)^{-u_k}}{2^{\log(u_k\log u_k)/\log y_k}-1} + 6[(\lambda+1)y^{-b_k/3} + 3y^{-4b_k/9}]\Big).
\end{aligned}$$

We find computationally that the optimal choice at $k = 300$ is $u = 6.3$, and $u = 15$ at $k = 1600$, between which a linear fit was used. With such choices of u, we obtain

$$\sum_{e^{300} \le n \le e^{1600}}{}' \frac{1}{n} \le 1.634184 + 0.809179 \le 2.444. \tag{3.14}$$

3.3 Abundant density estimates

The remaining range $10^{10} \le n \le e^{300}$ is too small for the pnd methods to be effective, so we use a different method related to abundant density estimates. We first illustrate the approach with a simple example from [34, 35] before proceeding to the general case.

The function $h(n) = \sigma(n)/n$ is multiplicative, so

$$h(n)^j = \sum_{d|n} f_j(d)$$

for $f_j(n)$ multiplicative with $f_j(p^a) = h(p^a)^j - h(p^{a-1})^j$ on prime powers, by Möbius inversion.

3.3.1 Case $2 \nmid n$

We have

$$\begin{aligned} \sum_{\substack{10^{10} \le n < e^K \\ n \text{ odd, abundant}}} \frac{1}{n} < 2^{-j} \sum_{\substack{10^{10} \le n < e^K \\ n \text{ odd}}} \frac{h(n)^j}{n} &= 2^{-j} \sum_{\substack{10^{10} \le n < e^K \\ n \text{ odd}}} \sum_{d|n} \frac{f_j(d)}{n} \\ &= 2^{-j} \sum_{\substack{d < e^K \\ d \text{ odd}}} \frac{f_j(d)}{d} \sum_{\substack{10^{10}/d \le m < e^K \\ m \text{ odd}}} \frac{1}{m}. \end{aligned}$$

By Corollary 2.2 in [35] with $u = 2$,

$$\sum_{\substack{m < e^K \\ n \text{ odd}}} \frac{1}{m} \le \frac{K + \gamma + \log 2}{2} + \frac{2}{e^K} \le K/2 + .64$$

for $K \ge 20$. Thus

$$\sum_{\substack{10^10 \le n < e^K \\ n \text{ odd, abundant}}} \frac{1}{n} < 2^{-j-1}(K + 1.28) \sum_{\substack{d < e^K \\ d \text{ odd}}} \frac{f_j(d)}{d}.$$

Now expanding the Euler product, we have

$$\sum_{d\text{ odd}} \frac{f_j(d)}{d} = \prod_{p>2}\sum_{a\ge 0}\frac{f_j(p^a)}{p^a} = \prod_{p>2}\Big(1+\sum_{a\ge 1}\frac{h(p^a)^j - h(p^{a-1})^j}{p^a}\Big)$$
$$= \prod_{p>2}\Big(1-\frac{1}{p}\Big)\sum_{a\ge 0}\frac{h(p^a)^j}{p^a}.$$

We may compute this product for primes up to B directly, where B is some convenient number, after which

$$\prod_{p>B}\Big(1-\frac{1}{p}\Big)\sum_{a\ge 0}\frac{h(p^a)^j}{p^a} \le \prod_{p>B}\Big(1-\frac{1}{p}\Big)\Big(1+\Big(\frac{p}{p-1}\Big)^j\sum_{a\ge 1}p^{-a}\Big)$$
$$= \prod_{p>B}\Big(1-\frac{1}{p}\Big)\Big(1+\frac{1}{p-1}\Big(\frac{p}{p-1}\Big)^j\Big)$$
$$= \prod_{p>B}\Big(1-p^{-1}+p^{-1}\Big(\frac{p}{p-1}\Big)^j\Big) \le \exp\sum_{p>B}\frac{\big(\frac{p}{p-1}\big)^j - 1}{p}$$
$$\le \exp\sum_{p>B}\frac{2j}{p(p-1)} \le \exp\sum_{x>B}\frac{j}{x(x-1)}$$
$$\le e^{j/B},$$

using

$$\sigma(p^a) = 1+p+\cdots+p^a = \frac{p^{a+1}-1}{p-1},$$
$$h(p^a) = \frac{\sigma(p^a)}{p^a} = \frac{p-p^{-a}}{p-1} \le \frac{p}{p-1} \quad (\text{for } a\ge 1).$$

Thus we have

$$\sum_{d\text{ odd}}\frac{f_j(d)}{d} = \prod_{p>2}\Big(1-\frac{1}{p}\Big)\sum_{a\ge 0}\frac{h(p^a)^j}{p^a}$$
$$\le e^{j/B}\prod_{2<p\le B}\Big(1-\frac{1}{p}\Big)\Bigg(1+\sum_{a=1}^{A}p^{-a}\Big(\frac{p-p^{-a}}{p-1}\Big)^j + \sum_{a=A+1}^{\infty}p^{-a}\Big(\frac{p}{p-1}\Big)^j\Bigg)$$
$$= e^{j/B}\prod_{2<p\le B}\Big(1-\frac{1}{p}\Big)\Bigg(1+\sum_{a=1}^{A}p^{-a}\Big(1+\frac{1-p^{-a}}{p-1}\Big)^j + \frac{1}{p^A(p-1)}\Big(\frac{p}{p-1}\Big)^j\Bigg). \tag{3.15}$$

By computing with $A=500$, $B=10^6$ we find that this is optimized at $j=18$ to give the bound

$$\sideset{}{'}\sum_{\substack{10^{10}<n<e^K\\ 2^i\| n}}\frac{1}{n} < 0.0118865K + 0.015.$$

This is in the proof of Proposition 3.5 from [35].

3.3.2 Scheme A

Consider the pnds $n \in [10^{10}, e^K]$ for which $2^i \| n$. We shall require $2^{2i+1} < 10^{10}$, i.e. $i < 16$. Since $n' = 2^i p$ is a pnd for all primes $p < 2^{i+1}$ and $n' \leq n$, this forces $n' \nmid n$ so $p \nmid n$. Thus letting $u = \prod_{p \leq 2^{i+1}} p$ and $n = 2^i l$, we deduce $\gcd(l, u) = 1$.

Writing $n = 2^i l$ for l with $\gcd(l, u) = 1$, we have

$$\begin{aligned}\sum_{\substack{10^{10} < n < e^K \\ 2^i \| n}}{}' \frac{1}{n} &< 2^{-j} \sum_{\substack{n < e^K \\ 2^i \| n}}{}' \frac{h(n)^j}{n} \leq 2^{-j} \frac{h(2^i)^j}{2^i} \sum_{\substack{l < e^K/2^i \\ \gcd(l,u)=1}} \frac{h(l)^j}{l} \\ &= 2^{-i} \Big(1 - 2^{-i-1}\Big)^j \sum_{\substack{l < e^K/2^i \\ \gcd(l,u)=1}} \sum_{d|n} \frac{f_j(d)}{l} \\ &\leq 2^{-i} \Big(1 - 2^{-i-1}\Big)^j \sum_{\substack{d < e^K \\ \gcd(d,u)=1}} \frac{f_j(d)}{d} \sum_{\substack{m < e^K/2^i \\ \gcd(m,u)}} \frac{1}{m}.\end{aligned}$$

By Corollary 2.2 in [35],

$$\begin{aligned}\sum_{\substack{m < e^K/2^i \\ \gcd(m,u)=1}} \frac{1}{m} &\leq \frac{\varphi(u)}{u}(K - i \log 2 + \gamma) - \sum_{d|u} \frac{\mu(d) \log d}{d} + 2^{\omega(u)+i} e^{-K} \\ &= (K - i \log 2 + \gamma) \prod_{p \leq 2^{i+1}} \Big(1 - \frac{1}{p}\Big) - \sum_{d|u} \frac{\mu(d) \log d}{d} + 2^{\pi(2^{i+1})+i} e^{-K}.\end{aligned}$$

Then, analogously to (3.15), we have

$$\sum_{\substack{d < e^K/2^i \\ \gcd(d,u)=1}} \frac{f_j(d)}{d} < e^{j/B} \prod_{2^{i+1} < p \leq B} \Big(1 - \frac{1}{p}\Big)\Bigg(1 + \sum_{a=1}^{A} p^{-a}\Big(1 + \frac{1 - p^{-a}}{p-1}\Big)^j + \frac{1}{p^A(p-1)}\Big(\frac{p}{p-1}\Big)^j\Bigg),$$

with j to be optimized.

In summary, we have

$$\sum_{\substack{10^{10} < n < e^K \\ 2^i \| n}}{}' \frac{1}{n} < 2^{-i}\Big(1 - 2^{-i-1}\Big)^j F(i) L(K, i), \tag{3.16}$$

where

$$u = \prod_{p \le 2^{i+1}} p, \tag{3.17}$$

$$L(K,i) = (K - i\log 2 + \gamma) \prod_{p \le 2^{i+1}} \Big(1 - \frac{1}{p}\Big) - \sum_{d|u} \frac{\mu(d)\log d}{d} + 2^{\pi(2^{i+1})+i} e^{-K}, \tag{3.18}$$

$$F(i) = e^{j/B} \prod_{2^{i+1} < p \le B} \Big(1 - \frac{1}{p}\Big)\Big(1 + \sum_{a=1}^{A} p^{-a}\Big(1 + \frac{1-p^{-a}}{p-1}\Big)^j + \frac{1}{p^A(p-1)}\Big(\frac{p}{p-1}\Big)^j\Big). \tag{3.19}$$

We use scheme A for $i = 4$. That is, we have

$$\sideset{}{'}\sum_{\substack{10^{10} < n < e^K \\ 2^i \| n}} \frac{1}{n} < b + mK,$$

where

i	j	A	B	b	m
4	95	30	500	0.0021484	0.002272899

For $i \ge 5$, we use Corollary 2.2 in [35]. That is, for $i = 5$

$$\begin{aligned}\sideset{}{'}\sum_{\substack{n<e^K \\ 2^i \mid n}} \frac{1}{n} &\le 2^{-i} \sum_{\substack{l < e^K/2^i \\ \gcd(l,u/2)=1}} \frac{1}{l} \\ &\le 2^{-i}(K - i\log 2 + \gamma) \prod_{2 < p \le 2^{i+1}} \Big(1 - \frac{1}{p}\Big) - 2^{-i} \sum_{d|u/2} \frac{\mu(d)\log d}{d} + 2^{\pi(2^{i+1})-1} e^{-K} \\ &\le 0.008224209K.\end{aligned}$$

3.3.3 Scheme B with a prime q

Scheme B is based on the following result of Erdős [17], which we state below.

Proposition 3.3.1 (Erdős)**.** *For any $i, k \ge 2$, we have that $m = 2^i q_1 \cdots q_k$ is a pnd for any k distinct primes $q_1, \ldots, q_k \in \big[(k-1)2^{i+1}, k2^{i+1}\big]$.*

Consider the pnds $n \in [10^{10}, e^K]$ for which $2^i \| n$ for some $i \le 9$. By Proposition 3.3.1 with $k = 2$, we have $m = 2^i qq'$ is a pnd for any distinct primes $q, q' \in [2^{i+1}, 2^{i+2}]$. Then $m \nmid n$ since $m \le 2^{3i+4} < 10^{10} \le n$ are distinct pnds, and so $qq' \nmid n$. Thus at most one prime q from the interval $[2^{i+1}, 2^{i+2}]$ can divide n. Writing $n = 2^i l$ and $u = \prod_{p<2^{i+2}} p$, we have $\gcd(l, u/q) = 1$.

Thus

$$\sum_{\substack{10^{10}<n<e^K\\ 2^i\|n}}{}' \frac{1}{n} < 2^{-j}\sum_{\substack{n<e^K\\ 2^i\|n}}{}' \frac{h(n)^j}{n} \le 2^{-j}\frac{h(2^i)^j}{2^i}\sum_{\substack{l<e^K/2^i\\ \gcd(l,u/q)=1}}\frac{h(l)^j}{l}$$
$$= 2^{-i}\Big(1-2^{-i-1}\Big)^j \sum_{\substack{l<e^K/2^i\\ \gcd(l,u/q)=1}}\sum_{d|n}\frac{f_j(d)}{l}$$
$$\le 2^{-i}\Big(1-2^{-i-1}\Big)^j \sum_{\substack{d<e^K\\ \gcd(d,u/q)=1}}\frac{f_j(d)}{d}\sum_{\substack{m<e^K/2^i\\ \gcd(m,u/q)}}\frac{1}{m}.$$

Again, by Corollary 2.2 in [35],

$$\sum_{\substack{m<e^K/2^i\\ \gcd(m,u)=1}}\frac{1}{m} \le \frac{\varphi(u/q)}{u/q}(K-i\log 2+\gamma) - \sum_{d|u/q}\frac{\mu(d)\log d}{d} + 2^{\omega(u/q)+i}e^{-K}$$
$$= (K-i\log 2+\gamma)\prod_{\substack{p\le 2^{i+2}\\ p\ne q}}\Big(1-\frac{1}{p}\Big) - \sum_{d|u/q}\frac{\mu(d)\log d}{d} + 2^{\pi(2^{i+2})-1+i}e^{-K}.$$

Then, analogously to (3.15), we have

$$\sum_{\substack{d<e^K/2^i\\ \gcd(d,u/q)=1}}\frac{f_j(d)}{d} <$$
$$e^{j/B}\prod_{\substack{2^{i+2}<p\le B\\ \text{or } p=q}}\Big(1-\frac{1}{p}\Big)\bigg(1+\sum_{a=1}^{A}p^{-a}\Big(1+\frac{1-p^{-a}}{p-1}\Big)^j + \frac{1}{p^A(p-1)}\Big(\frac{p}{p-1}\Big)^j\bigg),$$

with j to be optimized.

In summary, have

$$\sum_{\substack{10^{10}<n<e^K\\ 2^i\|n}}{}' \frac{1}{n} < 2^{-i}\Big(1-2^{-i-1}\Big)^j\sum_{2^{i+1}<q<2^{i+2}}F(i,q)L(K,i,q), \tag{3.20}$$

where

$$u = \prod_{p<2^{i+2}} p, \tag{3.21}$$

$$L(K,i,q) = (K-i\log 2+\gamma)\prod_{\substack{p\le 2^{i+2}\\ p\ne q}}\Big(1-\frac{1}{p}\Big) - \sum_{d|u/q}\frac{\mu(d)\log d}{d} + 2^{\pi(2^{i+2})-1+i}e^{-K}, \tag{3.22}$$

$$F(i,q) = e^{j/B}\prod_{\substack{2^{i+2}<p\le B\\ \text{or } p=q}}\Big(1-\frac{1}{p}\Big)\bigg(1+\sum_{a=1}^{A}p^{-a}\Big(1+\frac{1-p^{-a}}{p-1}\Big)^j + \frac{1}{p^A(p-1)}\Big(\frac{p}{p-1}\Big)^j\bigg). \tag{3.23}$$

We apply scheme B for $1 \le i \le 3$. That is, we have

$$\sum_{\substack{10^{10} < n < e^K \\ 2^i \| n}}' \frac{1}{n} < b + mK,$$

where

i	q	j	A	B	b	m
1	5	23	30	500	0.0190246	0.0131128
1	7	35	30	500	0.0028889	0.0018895
2	11	40	30	500	0.0061241	0.0044578
2	13	46	30	500	0.0035739	0.0025531
3	17	69	30	500	0.0034296	0.0023470
3	19	80	30	500	0.0022583	0.0015313
3	23	97	30	500	0.0010876	0.0007271
3	29	115	30	500	0.0004906	0.0003232
3	31	119	30	500	0.0004039	0.0002650

Hence we obtain the following result:

$$\sum_{10^{10} < n < e^K}' \frac{1}{n} \le .0496K + 0.056. \tag{3.24}$$

Thus for $K = 300$, combining with the previous results, we obtain our desired final bound.

Theorem 3.3.2. *The Erdős constant is bounded between*

$$0.34816486577357275 \ \le \ \sum_n' \frac{1}{n} \ \le \ 18.6.$$

Proof. The lower bound was computed directly, as in (3.2). For the upper bound, collecting the above results gives

$$\begin{aligned}\sum_n' \frac{1}{n} &= \sum_{n \le 10^{10}}' \frac{1}{n} + \sum_{10^{10} \le n \le e^{300}}' \frac{1}{n} + \sum_{e^{300} \le n \le e^{1600}}' \frac{1}{n} + \sum_{n > e^{1600}}' \frac{1}{n} \\ &\le 0.34816486577357275 + (.0496 \cdot 300 + 0.056) + (0.00138 + .918) + 2.444 \\ &\le 18.6.\end{aligned}$$

□

Bibliography

[1] W. R. Alford, A. Granville, and C. Pomerance, *There are infinitely many Carmichael numbers.* Ann. of Math. (2) **139** (1994), 703–722.

[2] ———, *The difficulty of finding reliable witnesses.* Algorithmic Number Theory Proceedings (ANTS-I), L. M. Adleman and M.-D. Huang, eds., Lecture Notes in Computer Sci. **877** (1994), Springer-Verlag, Berlin, pp. 1–16.

[3] M. Avidon, *On the distribution of primitive abundant numbers*, Acta Arith. **77** (2) (1996), 195–205.

[4] R. C. Baker and G. Harman, *Shifted primes without large prime factors.* Acta Arith. **83** (1998), 331–361.

[5] J. Bayless, D. Klyve, *On the sum of reciprocals of amicable numbers*, Integers **11A** (2011), Article 5.

[6] P. Beauchemin, G. Brassard, C. Crépeau, C. Goutier, and C. Pomerance, *The generation of random numbers that are probably prime.* J. Cryptology **1** (1988), 53-64.

[7] D. J. Bernstein, *Arbitrarily tight bounds on the distribution of smooth numbers*, in M. Bennett, et al., eds., Proceedings of the Millennial Conference on Number Theory, volume 1, pages 49–66. A. K. Peters, 2002.

[8] N. G. de Bruijn, *On the number of positive integers $\leq x$ and free of prime factors $> y$*, Nederl. Akad. Wetensch. Proc. Ser. A **54** (1951), 50–60.

[9] ———, *On the number of positive integers $\leq x$ and free of prime factors $> y$. II*, Nederl. Akad. Wetensch. Proc. Ser. A **69** = Indag. Math. **28** (1966), 239–247.

[10] R. J. Burthe, Jr., *Further investigations with the strong probable prime test.* Math. Comp. **65** (1996), 373–381.

[11] J. Büthe, *Estimating $\pi(x)$ and related functions under partial RH assumptions*, Math. Comp. **85** (2016), 2483–2498.

[12] ———, *An analytic method for bounding $\psi(x)$*, Math. Comp., Published electronically: October 26, 2017. https://doi.org/10.1090/mcom/3264.

[13] R. Crandall, C. Pomerance, *Prime Numbers: A Computational Perspective* 2nd ed. Springer (2005).

[14] I. Damgård, P. Landrock, C. Pomerance, *Average case error estimates for the strong probable prime test.* Math. Comp. **61** (1993), 177–194.

[15] K. Dickman, *On the frequency of numbers containing prime factors of a certain relative magnitude*, Ark. Mat. Astr. Fys. **22** (1930), 1–14.

[16] P. Erdős, *On the density of the abundant numbers*, J. London Math. Soc., **9** (1934), 278–282.

[17] ———, *On primitive abundant numbers*, J. London Math. Soc., **37** (1935), 49–58.

[18] P. Erdős, C. Pomerance, *On the number of false witnesses for a composite number.* Math. Comp. **46** (1986), 259–279.

[19] L. Faber and H. Kadiri, *New bounds for $\psi(x)$*, Math. Comp. **84** (2015), 1339–1357.

[20] S. W. Golomb, *Powerful numbers*, Amer. Math. Monthly **77**, (1970), 848–852.

[21] A. Granville, *Smooth numbers: computational number theory and beyond*, in Algorithmic number theory: lattices, number fields, curves and cryptography, 267–323, Math. Sci. Res. Inst. Publ. 44, Cambridge Univ. Press, Cambridge, 2008.

[22] A. Granville and K. Soundararajan, *Multiplicative number theory: The pretentious approach* (tentative title), in progress.

[23] A. Hildebrand, *On the number of positive integers $\le x$ free of prime factors $> y$*, J. Number Theory **22** (1986), 289–307.

[24] A. Hildebrand and G. Tenenbaum, *On integers free of large prime factors*, Trans. Amer. Math. Soc. **296** (1986), 265–290.

[25] A. E. Ingham, *The distribution of prime numbers*, Cambridge U. Press, 1932.

[26] A. Ivić, *The distribution of primitive abundant numbers*, Studia Sci. Math. Hungar., **20** (1985), 183–187.

[27] S. H. Kim, C. Pomerance, *The probability that a random probable prime is composite.* Math. Comp. **53** (1989), 721–741.

[28] D. Klyve, *Explicit bounds on twin primes and Brun's constant*, PhD thesis, Dartmouth College (2007).

[29] S. V. Konyagin and C. Pomerance, *On primes recognizable in deterministic polynomial time*, The mathematics of Paul Erdős, R. L. Graham and J. Nesetril, eds., Springer-Verlag, Berlin, 1997, pp. 176–198.

[30] J. D. Lichtman and C. Pomerance, *Improved error estimates for the Fermat primality test on random inputs*, Math. Comp. `https://doi.org/10.1090/mcom/3314`

[31] ———, *Explicit estimates for the distribution of numbers free of large prime factors*, J. Number Theory 183 (2018), 1–23. `https://doi.org/10.1016/j.jnt.2017.08.039`

[32] L. Monier, *Evaluation and comparison of two efficient probabilistic primality testing algorithms.* Theoret. Comput. Sci. **12** (1980), 97–108.

[33] P. Moree, *Integers without large prime factors: from Ramanujan to de Bruijn*, Integers **14A** (2014), paper A5, 13 pp.

[34] H. M. Nguyen, *The reciprocal sum of the amicable numbers*, Bachelors thesis, Dartmouth College (2014).

[35] H. M. Nguyen, C. Pomerance, *The reciprocal sum of the amicable numbers*, submitted for publication.

[36] S. T. Parsell, J. P. Sorenson, (2006) *Fast Bounds on the Distribution of Smooth Numbers*, Hess F., Pauli S., Pohst M. (eds) Algorithmic Number Theory. ANTS 2006. Lecture Notes in Computer Science, vol 4076. Springer, Berlin, Heidelberg

[37] D. J. Platt, *Isolating some nontrivial zeros of Zeta*, Math. Comp. **86** (2017), 2449–2467.

[38] D. J. Platt, T. S. Trudgian, *On the first sign change of $\vartheta(x) - x$.* Math. Comp. **85** (2016), 1539–1547.

[39] C. Pomerance, *On the distribution of amicable numbers*, II, J. reine angew. Math. **325** (1981), 183–188.

[40] M. O. Rabin, *Probabilistic algorithm for testing primality.* J. Number Theory, **12** (1980), 128–138.

[41] R. A. Rankin, *The difference between consecutive prime numbers*, J. London Math. Soc. **13** (1938), 242–247.

[42] J. Rosser, L. Schoenfeld, *Approximate formulas for some functions of prime numbers*, Illinois J. Math. **6** (1962), 64–94.

[43] ______, *Sharper bounds for the Chebyshev functions $\vartheta(x)$ and $\psi(x)$.* Math. Comp. **29** (1975), 243–265.

[44] H. Salié, *Über die Dichte abundanter Zahlen*, Math. Nachr. **14** (1955), 39–46.