IMPROVED ERROR BOUNDS FOR THE FERMAT PRIMALITY TEST ON RANDOM INPUTS

JARED D. LICHTMAN AND CARL POMERANCE

Abstract. 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.

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 $[1, n-1]$. 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 [5, Theorem 2.2] proved that

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

as $x \to \infty$. In particular, $\lim P(x) = 0$. Kim and Pomerance [7] replaced the asymptotic inequality of (1.2) with the weaker, but explicit, inequality

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

and gave numerical bounds on $P(x)$ for $10^{60} \le x < 10^{10^5}$. In this paper we simplify the argument in [7] and obtain better upper bounds on $P(x)$ for $10^{60} \le x \le 10^{90}$,

2000 *Mathematics Subject Classification.* Primary 11Y11; Secondary 11A51, 11N25.
Key words and phrases. Fermat test, Miller–Rabin test, probable prime.

as seen in Figure 1. In particular, at the start of this range, our bound is over 700 times smaller.

FIGURE 1. New bounds on $P(x)$.

x	Bound on $P(x)$ in [7]	New bound on $P(x)$
10^{60}	$7.16E{-}2$	$1.012E{-}4$
10^{70}	$2.87E{-}3$	$1.549E{-}5$
10^{80}	$8.46E{-}5$	$2.518E{-}6$
10^{90}	$1.70E{-}6$	$4.326E{-}7$
10^{100}	$2.77E{-}8$	$7.836E{-}8$

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

With these methods, we also obtain new nontrivial bounds for $2^{40} \le x < 10^{60}$, values of x smaller than the methods in [7] could handle. These results are included in Figure 2.

FIGURE 2. Upper bound on $P(2^k)$.

k	$P(2^k) \le$	k	$P(2^k) \le$	k	$P(2^k) \le$
40	$4.475E{-}1$	140	$3.321E{-}3$	240	$1.025E{-}5$
50	$3.045E{-}1$	150	$1.827E{-}3$	250	$5.915E{-}6$
60	$1.936E{-}1$	160	$1.007E{-}3$	260	$3.434E{-}6$
70	$1.184E{-}1$	170	$5.571E{-}4$	270	$2.004E{-}6$
80	$7.066E{-}2$	180	$3.097E{-}4$	280	$1.175E{-}6$
90	$4.207E{-}2$	190	$1.731E{-}4$	290	$6.925E{-}7$
100	$2.501E{-}2$	200	$9.722E{-}5$	300	$4.100E{-}7$
110	$1.495E{-}2$	210	$5.494E{-}5$	310	$2.439E{-}7$
120	$8.973E{-}3$	220	$3.121E{-}5$	320	$1.438E{-}7$
130	$5.442E{-}3$	230	$1.783E{-}5$	330	$8.753E{-}8$

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

A number n is called y-smooth if all of its prime factors are bounded above by y. The method of [7] first computes the contribution to $P(x)$ from numbers that are not y-smooth (for an appropriate choice for y), and then enters a complicated argument based on the asymptotic method of [5] for the contribution of the y-smooth numbers. In addition to small improvements made in the non-y-smooth case, our principal new idea is to use merely that there are few y-smooth numbers. For this we use the upper bound method pioneered by Rankin for this problem, obtaining numerically explicit upper bounds on the distribution of y-smooth numbers. 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, $1 \le b \le n - 1$. 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 [8] and Rabin [10] 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 $P_1(x) \le \frac{1}{4}$, but as noted in [3], this reasoning is flawed. However, in [4] and [6], something similar to $P_1(x) \le \frac{1}{4}$ is shown. Namely, if $P_1'(2^k)$ is the analogous probability for odd k-bit integers, it is shown in [4], [6] that $P_1'(2^k) \le \frac{1}{4}$ for all $k \ge 3$. 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 [6] 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 $\lambda(n)$ denote the Carmichael universal exponent function, $\mathrm{Li}(x) = \int_2^x \frac{\mathrm{d}t}{\log t}$, and $\vartheta(x) = \sum_{p \le x} \log p$. In many instances, we take a sum over certain subsets of odd composite integers, in which cases we use $\sum_n'$ to denote $\sum_{\substack{n \text{ odd,} \\ \text{composite}}}$.

2. 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 $\pm 1 \in \mathbf{F}(n)$. 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 $x \ge 5$,

$$P(x) = \frac{\sum_{n \le x}'(F(n) - 2)}{\sum_{1 < n \le x,\, n \text{ odd}}(F(n) - 2)} = \left(1 + \frac{\sum_{2 < p \le x}(p - 3)}{\sum_{n \le x}'(F(n) - 2)}\right)^{-1}. \tag{2.1}$$

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

Theorem 2.1. *For $x \ge 3300$, we have*

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

Theorem 2.2. *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*

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

where

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

Before proving Theorems 2.1 and 2.2, we state the main result of the section, which follows from these theorems.

Theorem 2.3. *Suppose c, L_1, and L are arbitrary positive real numbers satisfying $0 < c < 1$ and $1 < L_1 < L$. Then for any $x > L^2, 3300$, we have $P(x) \le 1/(1+z^{-1})$ where*

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

and B is defined as in Theorem 2.2.

Now we give the proof of Theorem 2.1.

Proof of Theorem 2.1. By partial summation,

$$\begin{aligned}\sum_{2<p\le x}(p-3) &= 1 - 3\pi(x) + \sum_{p\le x} p \\ &= 1 - 3\pi(x) + \vartheta(x)\frac{x}{\log x} - \int_2^x \vartheta(t)\frac{\log t - 1}{\log^2 t}\,\mathrm{d}t.\end{aligned} \tag{2.2}$$

Equations (3.6) in [11] and (5.6) in [12] state that

$$\pi(x) < 1.26\frac{x}{\log x} \quad \text{if } x > 1, \tag{2.3}$$

and

$$|\vartheta(x) - x| < \frac{\epsilon x}{\log x} \quad \text{if } x \ge \kappa \tag{2.4}$$

for $\epsilon = 0.024\,2334$, $\kappa = 758\,699$. Note that (2.3) implies that $\pi(x) < \frac{1}{3}x$ for $x \ge 50$. (We realize that this latter inequality is almost trivial, but we shall need the strength of (2.3) later.) Substituting into (2.2) gives

$$\sum_{2<p\le x}(p-3) > 1 - x + \frac{x^2}{\log x} - \epsilon\frac{x^2}{\log^2 x} - \int_2^x \vartheta(t)\frac{\log t - 1}{\log^2 t}\,\mathrm{d}t \tag{2.5}$$

for $x \ge \kappa$. Then by (2.4), the integral in (2.5) is bounded by

$$\begin{aligned}\int_2^x \vartheta(t)\frac{\log t - 1}{\log^2 t}\,\mathrm{d}t &< \int_2^\kappa \vartheta(t)\frac{\log t - 1}{\log^2 t}\,\mathrm{d}t \\ &\quad + \int_\kappa^x \frac{t(\log t - 1)}{\log^2 t}\,\mathrm{d}t + \epsilon\int_\kappa^x \frac{t(\log t - 1)}{\log^3 t}\,\mathrm{d}t.\end{aligned} \tag{2.6}$$

By [9, Theorem 1], we have $\vartheta(x) < x$ for all $0 < x \le 1.39 \cdot 10^{17}$. Thus, (2.6) implies that

(2.7)
$$\int_2^x \vartheta(t)\frac{\log t - 1}{\log^2 t}\,\mathrm{d}t < \int_2^x t\frac{\log t - 1}{\log^2 t}\,\mathrm{d}t + \epsilon\int_\kappa^x \frac{t(\log t - 1)}{\log^3 t}\,\mathrm{d}t$$
$$= \frac{x^2}{\log x} - \mathrm{Li}(x^2) - \frac{4}{\log 2} + \mathrm{Li}(4) + \epsilon\left(\frac{x^2}{2\log^2 x} - \frac{\kappa^2}{2\log^2 \kappa}\right).$$

Using this in (2.5) gives

$$\sum_{2<p\le x}(p-3) > \mathrm{Li}(x^2) - x - \frac{3\epsilon}{2}\frac{x^2}{\log^2 x} + k_0 \tag{2.8}$$

for $x \ge \kappa$, where

$$k_0 = 1 + \frac{4}{\log 2} - \mathrm{Li}(4) + \frac{\epsilon\kappa^2}{2\log^2\kappa} \approx 3.8 \cdot 10^7.$$

Call the function on the right side of (2.8) $f(x)$, and let $g(x) = x^2/(2\log x - 1/2)$. One can calculate that

$$f'(x) = \frac{x}{\log x} - 1 + 3\epsilon x\left(\frac{1}{\log^3 x} - \frac{1}{\log^2 x}\right),$$
$$g'(x) = 2x\left(\frac{1}{2\log x - 1/2} - \frac{1}{(2\log x - 1/2)^2}\right).$$

It is not hard to check that $f'(x) > g'(x)$ for $x \ge 80$ and that $f(80) > g(80)$. Thus, $f(x) > g(x)$ for $x \ge 80$ and substituting into (2.8) proves the theorem for $x \ge \kappa$. Moreover, we have checked numerically that

$$\sum_{i=1}^{k-1}(p_i - 3) > \frac{p_k^2}{2\log p_k - \frac{1}{2}}$$

for each prime p_k with $3\,299 < p_k \le \kappa$ (i.e. $463 < k \le 60\,875$). □

Proof of Theorem 2.2

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

$$\begin{aligned}\sum_{n\le x}{}'(F(n)-2) &= \sum_{\substack{n\le x\\ P^+(n)\le L}}{}'(F(n)-2) + \sum_{\substack{n\le x\\ P^+(n)>L}}{}'(F(n)-2)\\ &\le \sum_{\substack{n\le x\\ P^+(n)\le L\\ n \text{ odd}}} n + \sum_{\substack{n\le x\\ P^+(n)>L}}{}' F(n).\end{aligned} \tag{2.9}$$

For the first term in (2.9), we have for any $0 < c < 1$,

$$\sum_{\substack{n\le x\\ P^+(n)\le L\\ 2\nmid n}} n \le x^{1+c}\sum_{\substack{P^+(n)\le L\\ 2\nmid n}}\frac{1}{n^c} = x^{1+c}\prod_{2<p\le L}\left(1-p^{-c}\right)^{-1}. \tag{2.10}$$

Remark 2.4. By approximating the logarithm of the Euler product in (2.10) (with 2 included) using Lemma 3.1 in the next section and the method of [7], 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 3.1 and

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

Now we bound the second term in (2.9). 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

$$\begin{aligned}\sum_{\substack{n\le x\\ P^+(n)>L}}{}' F(n) &= \sum_{k=1}^{\infty} \sum_{n\in\mathbf{C}'_k(x)} F(n) = \sum_{k=1}^{\infty} \sum_{n\in\mathbf{C}'_k(x)} \frac{\varphi(n)}{k}\\ &= \sum_{k\le L_1} \frac{1}{k} \sum_{n\in\mathbf{C}'_k(x)} \varphi(n) + \sum_{k>L_1} \frac{1}{k} \sum_{n\in\mathbf{C}'_k(x)} \varphi(n) \qquad (2.11)\\ &\le x \sum_{k\le L_1} \frac{C'_k(x)}{k} + \frac{1}{L_1} \sum_{\substack{1<n\le x\\ n \text{ odd}}} (n-2) \le x \sum_{k\le L_1} \frac{C'_k(x)}{k} + \frac{x^2}{4L_1}.\end{aligned}$$

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 $S_{p,d} \le \frac{xd}{p(p-1)}$. 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 [7], we have $n \equiv 0 \pmod{\frac{p-1}{(k,p-1)}}$. Letting $d = (k, p-1)$, we have that $n \in \mathbf{S}_{p,d}$ 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 $m \equiv 1 \pmod{\frac{p-1}{d}}$. Then $m = 1 + u(\frac{p-1}{d})$ 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. If d is even, we then have u odd and

$\frac{p-1}{d}$ even, so $2d \mid p-1$. On the other hand, if d is odd, we of course have $2d \mid p-1$. Thus, we always have $2d \mid p-1$, and so

(2.12)
$$\begin{aligned}\sum_{k\le L_1}\frac{C_k'(x)}{k} &\le \sum_{d\le L_1}\frac{1}{d}\sum_{\substack{L<p\le x\\ 2d\mid p-1}}\sum_{\substack{u\le \frac{xd}{p(p-1)}\\ (u,d)=1}} 1\\ &= \sum_{d\le L_1}\frac{1}{d}\sum_{\substack{L<p\le x^{1/2}\\ 2d\mid p-1}}\sum_{\substack{u\le \frac{xd}{p(p-1)}\\ (u,d)=1}} 1 + \sum_{d\le L_1}\frac{1}{d}\sum_{\substack{x^{1/2}<p\le x\\ 2d\mid p-1}}\sum_{\substack{u\le \frac{xd}{p(p-1)}\\ (u,d)=1}} 1\\ &< \sum_{d\le L_1}\frac{\varphi(d)}{d}\sum_{\substack{L<p\le x^{1/2}\\ 2d\mid p-1}}\Big(\frac{x}{p(p-1)}+1\Big) + \sum_{d\le L_1}\sum_{\substack{x^{1/2}<n\le x\\ 2d\mid n-1}}\frac{x}{n(n-1)}\\ &< S_1+S_2+S_3,\end{aligned}$$

where

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

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.

Before obtaining our bounds, we first need some lemmas.

Lemma 2.5. *Given real numbers a, b and a nonnegative, decreasing function f on the interval $[a, b]$, we have that*

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

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

Lemma 2.6. *For $x \ge 2$, we have that*

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

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

(2.13)
$$\begin{aligned}\sum_{n\le x}\frac{\varphi(n)}{n} &= \sum_{n\le x}\sum_{d\mid n}\frac{\mu(d)}{d} = \sum_{d\le x}\frac{\mu(d)}{d}\sum_{n\le x/d} 1 = \sum_{d\le x}\frac{\mu(d)}{d}\left\lfloor\frac{x}{d}\right\rfloor\\ &= x\sum_{d\le x}\frac{\mu(d)}{d^2} - \sum_{d\le x}\frac{\mu(d)}{d}\left\{\frac{x}{d}\right\}.\end{aligned}$$

By Lemma 2.5,

$$\begin{aligned}\sum_{d\le x}\frac{\mu(d)}{d^2} &= \sum_{d=1}^{\infty}\frac{\mu(d)}{d^2} - \sum_{d>x}\frac{\mu(d)}{d^2} \le \frac{1}{\zeta(2)} + \sum_{d>x}\frac{1}{d^2}\\ &\le \frac{1}{\zeta(2)} + \frac{1}{x^2} + \int_x^{\infty}\frac{\mathrm{d}t}{t^2} = \frac{1}{\zeta(2)} + \frac{1}{x^2} + \frac{1}{x}.\end{aligned} \tag{2.14}$$

Direct computation shows that

$$\begin{aligned}-\sum_{d\le x}\frac{\mu(d)}{d}\Big\{\frac{x}{d}\Big\} &\le \sum_{\substack{d\le x\\ \mu(d)=-1}}\frac{1}{d} \le \sum_{1<d\le x}\frac{1}{d} - \sum_{\substack{1<d\le 18\\ \mu(d)\ne -1}}\frac{1}{d}\\ &\le \log x - 1 - \frac{1}{18}.\end{aligned}$$

Substituting this and (2.14) back into (2.13) gives

$$\sum_{n\le x}\frac{\varphi(n)}{n} \le \frac{x}{\zeta(2)} + \frac{1}{x} + 1 + \log x - 1 - \frac{1}{18} \le \frac{x}{\zeta(2)} + \log x.$$

□

Lemma 2.7. *For $x \ge 1$, we have that*

$$\sum_{n\le x}\frac{\varphi(n)}{n^2} \le \frac{\log x}{\zeta(2)} + 1.6.$$

Proof. We may assume that $x \ge 2$. By Lemma 2.6 and partial summation,

$$\begin{aligned}\sum_{n\le x}\frac{\varphi(n)}{n^2} &= 1 + \sum_{2\le n\le x}\frac{\varphi(n)}{n^2} = 1 + \frac{1}{x}\sum_{2\le n\le x}\frac{\varphi(n)}{n} + \int_2^x\frac{1}{t^2}\sum_{2\le n\le t}\frac{\varphi(n)}{n}\mathrm{d}t\\ &\le 1 + \frac{1}{x}\left(\frac{x}{\zeta(2)} + \log x - 1\right) + \int_2^x\frac{1}{t^2}\left(\frac{t}{\zeta(2)} + \log t - 1\right)\mathrm{d}t\\ &= \frac{\log x}{\zeta(2)} + \frac{1-\log 2}{\zeta(2)} + \frac{2+\log 2}{2} - \frac{2}{x},\end{aligned}$$

so the result follows. □

Consider S_1 in (2.12). For a given $d \le L_1$, by Lemma 2.5 we have that

$$\begin{aligned}\sum_{\substack{n>L\\ 2d\mid n-1}}\frac{1}{(n-1)^2} &= \sum_{2du+1>L}\frac{1}{4d^2u^2} \le \frac{1}{(L-1)^2} + \frac{1}{4d^2}\int_{(L-1)/2d}^{\infty}\frac{\mathrm{d}t}{t^2}\\ &= \frac{1}{(L-1)^2} + \frac{1}{2d(L-1)}.\end{aligned} \tag{2.15}$$

Thus, by Lemma 2.6 and Lemma 2.7,

$$\begin{aligned}S_1 &< x\sum_{d\le L_1}\frac{\varphi(d)}{d}\Big(\frac{1}{(L-1)^2} + \frac{1}{2d(L-1)}\Big)\\ &\le \frac{x}{(L-1)^2}\Big(\frac{L_1}{\zeta(2)} + \log L_1\Big) + \frac{x}{2(L-1)}\Big(\frac{\log L_1}{\zeta(2)} + 1.6\Big).\end{aligned} \tag{2.16}$$

By Lemma 2.7, S_2 in (2.12) is bounded by

$$S_2 \le \sum_{d\le L_1} \frac{\varphi(d)}{d}\frac{x^{1/2}}{2d} \le x^{1/2}\Big(\frac{\log L_1}{2\zeta(2)} + .8\Big). \tag{2.17}$$

We now consider S_3 in (2.12). For a fixed $d \le L_1$, we have, as in (2.15),

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

So,

$$S_3 \le x\sum_{d\le L_1} \frac{1}{(x^{1/2}-1)^2} + \frac{1}{2d(x^{1/2}-1)} \le \frac{xL_1}{(x^{1/2}-1)^2} + \frac{x(1+\log L_1)}{2(x^{1/2}-1)}. \tag{2.18}$$

By (2.16), (2.17), and (2.18), we obtain from (2.12) that

$$\sum_{k\le L_1} \frac{C_k'(x)}{k} < xB \tag{2.19}$$

for B as in Theorem 2.2. Thus, using (2.19) in (2.11) gives the following result.

Theorem 2.8. *Suppose c, L_1, and L are arbitrary real numbers satisfying $0 < c < 1$ and $1 < L_1 < L$. Then for any $x > L^2$, we have*

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

where B is as in Theorem 2.2.

Thus, (2.9), (2.10), and Theorem 2.8 give us Theorem 2.2.

3. A refinement of the basic method

We refine the basic method as done analogously in [7]. This refinement provides a modest improvement over Theorem 2.3 for x starting around 2^{140}. Before stating our principal result we require a lemma.

Lemma 3.1. *If $1 < y < x$ and $0 < c < 1$, then*

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

where

$$f(x,y) := (1+7.5\cdot 10^{-7})\left(\mathrm{Li}(x^{1-c}) - \mathrm{Li}(y^{1-c}) + \frac{y^{1-c}}{\log y}\right) - \vartheta(y)\frac{y^{-c}}{\log y}.$$

Remark 3.2. In applying this result numerically it can be convenient to use (see [12, (5.2)]):

$$\vartheta(y) \ge 0.998\,684y \quad \text{if } y \ge 1\,319\,000.$$

However, we have not used this approximation in our results.

Proof. By partial summation,

$$\sum_{y<p\le x} p^{-c} = \sum_{y<p\le x} \log p\Big(\frac{p^{-c}}{\log p}\Big)$$
$$\le \vartheta(x)\frac{x^{-c}}{\log x} - \vartheta(y)\frac{y^{-c}}{\log y} + \int_y^x \vartheta(t)\Big(\frac{ct^{-1-c}\log t + t^{-1-c}}{\log^2 t}\Big)\,\mathrm{d}t.$$

By [9, Cor. 1], we have $\vartheta(x) < (1 + 7.5\cdot 10^{-7})x$ for all $x > 0$, so

$$\sum_{y<p\le x} p^{-c} \le (1+7.5\cdot 10^{-7})\bigg(\frac{x^{1-c}}{\log x} + \int_y^x \frac{c}{t^c\log t} + \frac{1}{t^c\log^2 t}\,\mathrm{d}t\bigg) - \vartheta(y)\frac{y^{-c}}{\log y}$$
$$= (1+7.5\cdot 10^{-7})\bigg(\mathrm{Li}(x^{1-c}) - \mathrm{Li}(y^{1-c}) + \frac{y^{1-c}}{\log y}\bigg) - \vartheta(y)\frac{y^{-c}}{\log y}.$$

□

Theorem 3.3. *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 3.1, B is as in Theorem 2.2, and

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

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 $F(n) \ge \varphi(n)/L_1$.

We retain Theorem 2.8 and the remark following (2.11) 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

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

Using Lemma 3.1,

(3.1)

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

We now have the following result.

Theorem 3.4. *If $0<c<1$, $1<L<x$, and $L<M<L^2$, then*

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

where f is as in Lemma 3.1.

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,

$$\sum_{n \text{ in case (iii)}}{}' F(n) \le x \sum_{k\le L_1} \frac{B_k(x)}{k}. \tag{3.2}$$

By (2.11) in [5], 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

$$n\equiv 0 \pmod{PQ}, \quad k(n-1)\equiv 0 \pmod{\lambda(PQ)}. \tag{3.3}$$

Suppose first that $P=Q$. Then $\lambda(PQ)=P(P-1)$, so that (3.3) 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

$$\frac{x}{M}\sum_{k\le L_1}\frac{x}{k}\sum_{\substack{P\mid k\\ P>M^{1/2}}} 1 < \frac{x^2}{M}\log L_1 \sum_{k\le L_1}\frac{1}{k} < \frac{x^2}{M}(1+\log L_1)^2. \tag{3.4}$$

Now consider the case $P>Q$. The latter congruence in (3.3) is equivalent to

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

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 (3.3) 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

$$\sum_{\substack{q<p\le L\\ pq>M}}\left(1+\frac{x(k,\lambda(pq))}{pq\lambda(pq)}\right)\le\frac{1}{2}L^2+\frac{1}{2}x\sum_{\substack{p,q\le L\\ pq>M\\ p\ne q}}\frac{(k,[p-1,q-1])}{pq[p-1,q-1]}. \tag{3.5}$$

This is (4.4) in [7]. Following the argument in [7] from there, and letting $M'=M-2L$, we have that

$$\sum_{\substack{q,p\le L\\ pq>M\\ p\ne q}}\frac{(k,[p-1,q-1])}{pq[p-1,q-1]}\le\sum_{\substack{u_1u_2u_3u_4=k\\ (u_1,u_2)=1}}\sum_{\substack{\mu\le L/u_1\\ \nu\le L/u_2}}\sum_{u_1u_2u_3^2\mu\nu\delta^2>M'}\frac{1}{\mu^2\nu^2\delta^3u_1u_2u_3^2}. \tag{3.6}$$

We now split up the sum on the right side of (3.6) into two cases, $\delta=1$ and $\delta\ge 2$. When $\delta=1$, we have

$$\begin{aligned}\sum_{\substack{u_1u_2u_3u_4=k\\ (u_1,u_2)=1}}\sum_{\substack{\mu\le L/u_1\\ \nu\le L/u_2}}\sum_{\mu\nu u_1u_2u_3^2>M'}\frac{1}{\mu^2\nu^2u_1u_2u_3^2}&<\frac{5}{3M'}\sum_{u_1u_2u_3u_4=k}\sum_{\nu\le L/u_2}\frac{1}{\nu}\\ &\le\frac{5}{3M'}(1+\log L)\sum_{u_1u_2u_3u_4=k}1,\end{aligned} \tag{3.7}$$

where we used (4.7) in [7], which states

$$\sum_{\mu>y}\frac{1}{\mu^2}<\frac{5}{3y}\quad\text{for } y>0.$$

When $\delta\ge 2$, let $D:=\sqrt{M'/u_1u_2u_3^2\mu\nu}$, and we have

$$\begin{aligned}\sum_{\substack{u_1u_2u_3u_4=k\\ (u_1,u_2)=1}}\sum_{\substack{\mu\le L/u_1\\ \nu\le L/u_2}}\sum_{\delta\ge\max\{2,D\}}\frac{1}{\mu^2\nu^2\delta^3u_1u_2u_3^2}&\le\frac{4(\zeta(3)-1)}{M'}\sum_{u_1u_2u_3u_4=k}\sum_{\substack{\mu\le L/u_1\\ \nu\le L/u_2}}\frac{1}{\mu\nu}\\ &\le\frac{4(\zeta(3)-1)}{M'}(1+\log L)^2\sum_{u_1u_2u_3u_4=k}1,\end{aligned} \tag{3.8}$$

where we used the following lemma.

Lemma 3.5. *For $y>1$, we have*

$$\sum_{n\ge y}\frac{1}{n^3}\le\frac{4(\zeta(3)-1)}{y^2}.$$

Proof. When $1<y\le 2$, we have

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

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

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

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

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

When $4 < y$, by Lemma 2.5, direct computation shows that

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

□

Substituting (3.7) and (3.8) back into (3.6) and then (3.5), we have

$$\begin{aligned}\sum_{k\le L_1}\frac{1}{k}\sum_{\substack{q<p\le L\\ pq>M}}\Big(1+\frac{x(k,\lambda(pq))}{pq\lambda(pq)}\Big) &< \frac{1}{2}L^2(1+\log L_1)\\ &\quad + x(1+\log L)\Big(\frac{5}{6M'} + \frac{2(\zeta(3)-1)}{M'}(1+\log L)\Big)\sum_{k\le L_1}\frac{\tau_{(4)}(k)}{k},\end{aligned} \tag{3.9}$$

where $\tau_{(i)}(k)$ is the number of ordered factorizations of k into i positive factors. By (4.9) in [7], we have

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

for any natural number i and any $y \ge 1$. Using this in (3.9) and then combining with (3.4) gives

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

where C is as in Theorem 3.3. Thus, from (3.2) we have the following result.

Theorem 3.6. *If $10 < L_1 < L < M/2$ and $x > L^2 > M$, then*

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

where C is as in Theorem 3.3.

Combining Theorems 2.8, 3.4 and 3.6 yield Theorem 3.3. □

Finally, Theorems 2.1 and 3.3 give the following result.

Theorem 3.7. *If $0 < c < 1$, $10 < L_1 < L$, $2L < M < L^2 < x$, and $x \ge 3300$, then $P(x) \le 1/(1+z^{-1})$ where*

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

f is as in Lemma 3.1, B is as in Theorem 2.2, and C is as in Theorem 3.3.

4. The strong probable prime test

The next theorem extends the applicability of Theorems 2.3 and 3.7 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,

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

The following theorem together with Theorems 2.1, 2.2, and 3.3 allows for a numerical estimation of $P_1(x)$ for various values of x.

Theorem 4.1. *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 [6], 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

$$\sum_{n\le x}{}' 1 \ge \frac{1}{2}\sum_{\substack{2<p\le x^{1/a}\\ a\ge 2}}(p-1). \tag{4.1}$$

Since the primes larger than 2 are odd, we have $\pi(x)-1 < \frac{1}{2}x$ when $x > 0$. We also have the bound (2.3). Thus,

$$\begin{aligned}\frac{1}{2}\sum_{\substack{2<p^a\le x\\ a\ge 2}}(p-1) &\le \frac{1}{2}(\pi(x^{1/2})-1)(x^{1/2}-1) + \frac{1}{2}\sum_{3\le a\le \log x}(\pi(x^{1/a})-1)(x^{1/a}-1)\\ &< \frac{1.26(x-x^{1/2})}{\log x} + \frac{1}{4}(x^{2/3}-x^{1/3})(\log x - 2).\end{aligned}$$

Also, by (2.3), we have,

$$\sum_{n\le x}{}' 1 \ge \frac{1}{2}x - 1.26\frac{x}{\log x}.$$

We thus have for $x \geq 610$,

$$\sum_{n\leq x}{}' 1 - \frac{1}{2}\sum_{2<p\leq x}(p-1)$$
$$\geq \frac{1}{2}x - 1.26\frac{x}{\log x} - 1.26\frac{x-x^{1/2}}{\log x} - \frac{1}{4}(x^{2/3}-x^{1/3})(\log x - 2) > 0,$$

and (4.1) holds. For $9 \leq x \leq 610$, the inequality can be verified directly. Indeed, the prime sum in (4.1) increases only at the 13 powers of odd primes to 610 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. □

5. Numerical results

We apply Theorems 2.3 and 3.7 to obtain numerical bounds on $P(x)$ for various values of x. In Figure 3, bounds on $P(2^k)$ are computed using Theorem 2.3 for $40 \leq k \leq 130$ and Theorem 3.7 for $140 \leq k \leq 330$, at which point the methods of this paper lose their edge over those in [7]. Note that the upper bounds in Theorems 2.3, 3.7 are decreasing functions in x, so one can use the Figure 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 $k \leq 36$. We have that

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

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

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

we have that

$$P(2^k) = \frac{\sum_{j=3}^{k} S'_c(2^j)}{\sum_{j=3}^{k}\big(S'_p(2^j)+S'_c(2^j)\big)}. \tag{5.1}$$

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

$$P'(2^k) = \frac{S'_c(2^k)}{S'_p(2^k)+S'_c(2^k)}.$$

We have directly computed $S'_p(2^k)$ and $S'_c(2^k)$ for $k \leq 36$, with the latter computation aided by the formula $F(n) = \prod_{p|n}(p-1, n-1)$. In Figure 4, we provide the values of $S_p(2^k)$ and $S_c(2^k)$, as well as $P(2^k)$ and $P'(2^k)$ up to 7 significant digits.

FIGURE 3. Upper bound on $P(2^k)$.

k	L	L_1	$M^{1/2}$	c	$P(2^k) \leq$
40	307^-	135		0.5440	$4.475E{-1}$
50	727^-	318		0.5850	$3.045E{-1}$
60	$1.860E{+3}$	831		0.6235	$1.936E{-1}$
70	$4.000E{+3}$	$1.75E{+3}$		0.6491	$1.184E{-1}$
80	$8.500E{+3}$	$3.72E{+3}$		0.6704	$7.066E{-2}$
90	$1.804E{+4}$	$7.55E{+3}$		0.6906	$4.207E{-2}$
100	$3.505E{+4}$	$1.54E{+4}$		0.7052	$2.501E{-2}$
110	$7.351E{+4}$	$3.27E{+4}$		0.7217	$1.495E{-2}$
120	$1.354E{+5}$	$5.95E{+4}$		0.7321	$8.973E{-3}$
130	$2.506E{+5}$	$1.10E{+5}$		0.7423	$5.442E{-3}$
140	$1.05E{+6}$	$1.57E{+5}$	$2.379E{+5}$	0.7445	$3.321E{-3}$
150	$2.33E{+6}$	$3.19E{+5}$	$3.739E{+5}$	0.7506	$1.827E{-3}$
160	$5.20E{+6}$	$6.02E{+5}$	$5.689E{+5}$	0.7555	$1.007E{-3}$
170	$1.10E{+7}$	$1.21E{+6}$	$8.669E{+5}$	0.7603	$5.571E{-4}$
180	$2.31E{+7}$	$2.30E{+6}$	$1.315E{+6}$	0.7648	$3.097E{-4}$
190	$4.73E{+7}$	$4.55E{+6}$	$1.990E{+6}$	0.7692	$1.731E{-4}$
200	$9.65E{+7}$	$8.69E{+6}$	$2.990E{+6}$	0.7734	$9.722E{-5}$
210	$1.93E{+8}$	$1.66E{+7}$	$4.455E{+6}$	0.7773	$5.494E{-5}$
220	$3.77E{+8}$	$3.16E{+7}$	$6.627E{+6}$	0.7811	$3.121E{-5}$
230	$7.51E{+8}$	$5.74E{+7}$	$9.644E{+6}$	0.7845	$1.783E{-5}$
240	$1.44E{+9}$	$1.09E{+8}$	$1.409E{+7}$	0.7878	$1.025E{-5}$
250	$2.73E{+9}$	$2.01E{+8}$	$2.049E{+7}$	0.7911	$5.915E{-6}$
260	$5.11E{+9}$	$3.66E{+8}$	$2.946E{+7}$	0.7941	$3.434E{-6}$
270	$9.59E{+9}$	$6.64E{+8}$	$4.204E{+7}$	0.7969	$2.004E{-6}$
280	$1.79E{+10}$	$1.19E{+9}$	$5.998E{+7}$	0.7996	$1.175E{-6}$
290	$3.28E{+10}$	$2.18E{+9}$	$8.558E{+7}$	0.8023	$6.925E{-7}$
300	$6.03E{+10}$	$3.97E{+9}$	$1.197E{+8}$	0.8048	$4.100E{-7}$
310	$1.09E{+11}$	$6.87E{+9}$	$1.678E{+8}$	0.8072	$2.439E{-7}$
320	$2.01E{+11}$	$1.22E{+10}$	$2.347E{+8}$	0.8095	$1.438E{-7}$
330	$3.47E{+11}$	$2.10E{+10}$	$3.297E{+8}$	0.8117	$8.753E{-8}$

Additionally, we have estimated $P(2^k)$ in the range $30 \leq k \leq 50$ 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 $S'_p(2^k)$ by

$$R'_p(2^k) = \int_{2^{k-1}}^{2^k} \frac{t-3}{\log t}\, \mathrm{d}t = \mathrm{Li}(2^{2k}) - \mathrm{Li}(2^{2(k-1)}) - 3\big(\mathrm{Li}(2^k) - \mathrm{Li}(2^{k-1})\big),$$

to smooth out some noise from the experiment. To estimate $S'_c(2^k)$, 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, $\bar{R}'_c$, and median, $\tilde{R}'_c$, of the data. Using these statistics, we estimate $P'(2^k)$ by

$$\bar{Q}'(2^k) = \frac{\bar{R}'_c(2^k)}{R'_p(2^k) + \bar{R}'_c(2^k)}, \quad \tilde{Q}'(2^k) = \frac{\tilde{R}'_c(2^k)}{R'_p(2^k) + \tilde{R}'_c(2^k)}.$$

FIGURE 4. Exact values of $S'_p(2^k)$, $S'_c(2^k)$, $P'(2^k)$, and $P(2^k)$.

k	$S'_p(2^k)$	$S'_c(2^k)$	$P'(2^k)$	$P(2^k)$
3	6	0	0	0
4	18	2	$1.000000E{-}1$	$7.692308E{-}2$
5	104	4	$3.703704E{-}2$	$4.477612E{-}2$
6	320	24	$6.976744E{-}2$	$6.276151E{-}2$
7	1180	114	$8.809892E{-}2$	$8.126411E{-}2$
8	4292	316	$6.857639E{-}2$	$7.210031E{-}2$
9	16338	1114	$6.383223E{-}2$	$6.604565E{-}2$
10	57416	3056	$5.053579E{-}2$	$5.492029E{-}2$
11	208576	10890	$4.962044E{-}2$	$5.109129E{-}2$
12	780150	28094	$3.475931E{-}2$	$3.922073E{-}2$
13	2837158	74528	$2.559617E{-}2$	$2.936153E{-}2$
14	10673384	231514	$2.123028E{-}2$	$2.342189E{-}2$
15	39467286	582318	$1.453992E{-}2$	$1.695170E{-}2$
16	148222234	1636968	$1.092337E{-}2$	$1.254137E{-}2$
17	559288478	4521166	$8.018958E{-}3$	$9.224140E{-}3$
18	2106190104	11682336	$5.516072E{-}3$	$6.503488E{-}3$
19	7995006772	33290330	$4.146624E{-}3$	$4.769917E{-}3$
20	30299256236	88781082	$2.921580E{-}3$	$3.410027E{-}3$
21	115430158810	230250774	$1.990748E{-}3$	$2.364213E{-}3$
22	440353630422	628735800	$1.425762E{-}3$	$1.672109E{-}3$
23	1683364186642	1680806136	$9.974844E{-}4$	$1.174178E{-}3$
24	6448755473484	4408788648	$6.831980E{-}4$	$8.115041E{-}4$
25	24754014371036	11552686982	$4.664818E{-}4$	$5.564524E{-}4$
26	95132822935752	30756273488	$3.231937E{-}4$	$3.839294E{-}4$
27	366232744269106	82133627362	$2.242159E{-}4$	$2.657319E{-}4$
28	1411967930053822	215629423796	$1.526922E{-}4$	$1.820302E{-}4$
29	5450257882815404	565834872742	$1.038072E{-}4$	$1.240823E{-}4$
30	21065843780715212	1504267288346	$7.140278E{-}5$	$8.503905E{-}5$
31	81507897575948416	3999812059436	$4.907029E{-}5$	$5.837036E{-}5$
32	315718919767278610	10350692466866	$3.278344E{-}5$	$3.939180E{-}5$
33	1224166825030041460	27472503360964	$2.244129E{-}5$	$2.681455E{-}5$
34	4750936696054816476	72288538641772	$1.521541E{-}5$	$1.820515E{-}5$
35	18454541611019193346	190806759987694	$1.033918E{-}5$	$1.236484E{-}5$
36	71745407298862105164	498526567616818	$6.948502E{-}6$	$8.342128E{-}6$

For $30 \leq k \leq 36$, $P'(2^k)$ is known, in which case we compute the relative errors, $\bar{Q}'/P' - 1$ and $\tilde{Q}'/P' - 1$, to get a sense of the accuracy of the experiment. Then

we estimate $P(2^k)$ by

$$Q(2^k) = \frac{R_c(2^k)}{R_p(2^k) + R_c(2^k)}$$

where

$$R_c(2^k) = \begin{cases} S_c(2^{k-1}) + \bar{R}_c'(2^k) & \text{for } 30 \le k \le 36, \\ S_c(2^{36}) + \sum_{j=37}^{k} \bar{R}_c'(2^j) & \text{for } 37 \le k \le 50, \end{cases}$$

and

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

The results of a random sampling experiment are summarized in Figures 5 and 6. It is difficult to give rigorous probabilities since $P(x)$ is inordinately influenced by a small fraction of composite numbers n where $F(n)$ is large. In fact, it is shown in [5] that the normal order of $F(n)$ for n composite is far different than the average order.

FIGURE 5. Random sampling estimates in range where $P(2^k)$ is known.

k	$\bar{Q}'(2^k)$	rel. err.	$\tilde{Q}'(2^k)$	rel. err.	$Q(2^k)$
30	$5.541E{-5}$	-0.224	$5.045E{-5}$	-0.293	$7.319E{-5}$
31	$4.800E{-5}$	-0.022	$3.616E{-5}$	-0.263	$5.758E{-5}$
32	$2.706E{-5}$	-0.175	$1.899E{-5}$	-0.421	$3.515E{-5}$
33	$2.223E{-5}$	-0.009	$1.248E{-5}$	-0.444	$2.666E{-5}$
34	$1.387E{-5}$	-0.088	$1.013E{-5}$	-0.334	$1.721E{-5}$
35	$7.603E{-6}$	-0.265	$6.506E{-6}$	-0.371	$1.033E{-5}$
36	$4.433E{-6}$	-0.362	$4.123E{-6}$	-0.407	$6.474E{-6}$

FIGURE 6. Random sampling estimates in range where $P(2^k)$ is unknown.

k	$\bar{Q}'(2^k)$	$\tilde{Q}'(2^k)$	$Q(2^k)$
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	$2.361E{-8}$	$1.318E{-8}$	$2.865E{-8}$

References

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. 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.
4. R. J. Burthe, Jr., *Further investigations with the strong probable prime test.* Math. Comp. **65** (1996), 373–381.
5. P. Erdős, C. Pomerance, *On the number of false witnesses for a composite number.* Math. Comp. **46** (1986), 259–279.
6. I. Damgård, P. Landrock, C. Pomerance, *Average case error estimates for the strong probable prime test.* Math. Comp. **61** (1993), 177–194.
7. S. H. Kim, C. Pomerance, *The probability that a random probable prime is composite.* Math. Comp. **53** (1989), 721–741.
8. L. Monier, *Evaluation and comparison of two efficient probabilistic primality testing algorithms.* Theoret. Comput. Sci. **12** (1980), 97–108.
9. D. J. Platt, T. S. Trudgian, *On the first sign change of $\vartheta(x) - x$.* Math. Comp. **85** (2016), 1539–1547.
10. M. O. Rabin, *Probabilistic algorithms for testing primality.* J. Number Theory, **12** (1980), 128–138.
11. J. Rosser, L. Schoenfeld, *Approximate formulas for some functions of prime numbers.* Illinois J. Math. **6** (1962), 64–94.
12. J. Rosser, L. Schoenfeld, *Sharper bounds for the Chebyshev functions $\vartheta(x)$ and $\psi(x)$.* Math. Comp. **29** (1975), 243–265.

Department of Mathematics, Dartmouth College, Hanover, NH 03755
E-mail address: lichtman.18@dartmouth.edu

Department of Mathematics, Dartmouth College, Hanover, NH 03755
E-mail address: carl.pomerance@dartmouth.edu