COUNTING COMPOSITES WITH TWO STRONG LIARS

ERIC BACH AND ANDREW SHALLUE

ABSTRACT. The strong probable primality test is an important practical tool for discovering prime numbers. Its effectiveness derives from the following fact: for any odd composite number n, if a base a is chosen at random, the algorithm is unlikely to claim that n is prime. If this does happen we call a a liar. In 1986, Erdős and Pomerance computed the normal and average number of liars, over all $n \leq x$. We continue this theme and use a variety of techniques to count $n \leq x$ with exactly two strong liars, those being the n for which the strong test is maximally effective. We evaluate this count asymptotically and give an improved algorithm to determine it exactly. We also provide asymptotic counts for the restricted case in which n has two prime factors, and for the n with exactly two Euler liars.

1. INTRODUCTION

The strong probable primality test (studied by Selfridge, Miller, Rabin, and others) is an important tool for discovering prime numbers in practice. Its success relies on the scarcity of strong liars.

Definition 1.1. Let n be an odd composite integer. Write $n - 1$ as $2^k \cdot n'$ where n' is the odd part of $n - 1$ and $k = \mathrm{ord}_2(n)$. Then a is a strong liar with respect to n if either

   (1) $a^{2^i n'} \equiv -1 \bmod n$ for some $0 \leq i < k$ or
   (2) $a^{n'} \equiv 1 \bmod n$.

Throughout we will use this convention of writing $n - 1$ as $2^k \cdot n'$ where n' is odd. We use log for the natural logarithm, and $\varphi(n)$ for the count of $1 \leq a \leq n$ with $\gcd(a, n) = 1$. When using asymptotic notation, implied constants with subscripts depend on that variable. We use the Euler constant γ defined by

$$\gamma = \int_1^\infty \left(\frac{1}{\lfloor x \rfloor} - \frac{1}{x} \right) \, \mathrm{d}x \approx 0.5772 \ .$$

If n is prime then the condition in Definition 1.1 holds for all a not divisible by n. If n is an odd composite greater than 9, then n has between 2 and $\varphi(n)/4$ strong liars (see Section 2). For n odd define

$$S(n) = \{a \mod n \ : \ a^{n'} \equiv 1 \bmod n \text{ or } a^{2^i n'} \equiv -1 \bmod n \text{ for some } 0 \leq i < k\} \ .$$

As a shorthand we will refer to elements of $S(n)$ as strong liars, even though if n prime then $a \in S(n)$ is truthfully giving us evidence that n is prime. It is useful to define two other types of liars.

Definition 1.2. Let n be an odd composite integer and $(a \mid n)$ the Jacobi symbol. Then a is a Fermat liar with respect to n if $a^{n-1} \equiv 1 \bmod n$ and a is an Euler liar with respect to n if $\gcd(a, n) = 1$ and $a^{(n-1)/2} \equiv (a \mid n) \bmod n$.

We similarly define

$$F(n) = \{a \mod n \; : \; a^{n-1} \equiv 1 \mod n\},$$

$$E(n) = \{a \mod n \; : \; \gcd(a, n) = 1 \text{ and } a^{(n-1)/2} \equiv (a \mid n) \mod n\} \; ,$$

and an important fact is that $S(n) \subseteq E(n) \subseteq F(n)$. While $E(n)$ and $F(n)$ are always subgroups of the group of units modulo n, $S(n)$ may not be.

Our primary interest will be in counting n where $|S(n)|$ is an extremal value. On the practical side, it is useful to know how often we might expect the strong primality test to be as effective as possible or as ineffective as possible. In Section 2 we discuss what is known about the worst case, but our new contribution involves counting best case composites. This occurs when $1, -1$ are the only two strong liars, and the happy consequence is that one trial of the strong primality test is sufficient to prove compositeness. We show that the proportion of such composites less than x goes to 0 as x goes to infinity, and yet more than a quarter of the numbers up to 10^9 are composites with exactly two strong liars.

For theoretical motivation, we will see that a key quantity is

$$\prod_{p \mid n} \gcd(p', n')$$

which is interesting in its own right. Finally, our work is complementary to that of Erdős and Pomerance in [5], who provide upper and lower bounds on the arithmetic and geometric mean of the size of all three sets $S(n)$, $E(n)$ and $F(n)$. They also provide some discussion of counts of n with extremal values of $|F(n)|$. We extend one of those results to $|S(n)|$.

In addressing these questions, we prove results using both analytic and algorithmic techniques. Our main result is the following.

Theorem. *The number of odd $n \leq x$ with exactly two strong liars is given by*

$$(1 + o(1)) \frac{x e^{-\gamma}}{\log \log \log x}$$

where γ is Euler's constant.

We also prove that the number of odd $n \leq x$ with exactly two Euler liars is half that amount. In [5, Section 6] the authors prove that the number of $n \leq x$ with $|F(n)| = 1$ is $(1 + o(1)) x e^{-\delta} / \log \log \log x$. All of these results utilize an argument from [3], where Erdős proves the number of $n \leq x$ with $\gcd(n, \varphi(n)) = 1$ is also $(1 + o(1)) x e^{-\gamma} / \log \log \log x$.

It would be interesting to know how many n with two strong liars have r prime factors. A start on that project is the following theorem.

Theorem. *The number of odd $n \leq x$ with $n = pq$, p, q both prime and $\gcd(p', q') = 1$ is*

$$= (1 + o(1)) \frac{C x \log \log x}{\log x}$$

$$\text{where } C := \prod_{p > 2} \left(1 - \frac{1}{(p-1)^2} \right) = 0.66016\ldots$$

is the Hardy-Littlewood twin prime constant. If C is replaced by $3C/4$, we get the count of odd $n \leq x$ with two prime factors and $|S(n)| = 2$.

We have also proven an asymptotic formula for the number of $n = pq$ with $p, q \leq x$ and $\gcd(p', q') = 1$, but will not discuss that result here.

For intuition on these two theorems, note that Mertens' theorem [9, Theorem 429] gives us

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

and that the number of positive integers $n \leq x$ that are the product of two primes is asymptotic to $x(\log \log x)/\log x$ [11]. So the count in the first theorem is driven by sieving by primes less than $\log \log x$, while the number of $n \leq x$ with two prime factors drives the second theorem.

Finally, we have designed a new algorithm that exactly counts the number of odd $n \leq x$ with two strong liars, and does so more quickly than simply applying a known formula to each n.

Theorem. *There is an algorithm that, given x, computes the number of positive integers $n \leq x$ with two strong liars. This algorithm requires $O(x(\log x)(\log \log x))$ bit operations and $O(x \log x)$ space.*

2. FACTS ABOUT STRONG LIARS

We collect in this section a number of facts related to strong liars, some of which will be useful for results in later sections. Recall that n' is the odd part of $n - 1$, so that for example $10' = 9$ and $9' = 1$. We use k to denote $\mathrm{ord}_2(n - 1)$, so that $n - 1 = 2^k \cdot n'$.

First, note that if n is odd then ± 1 are always strong liars, since $1^{n'} \equiv 1 \bmod n$ for all n and $(-1)^{n'} \equiv -1 \bmod n$ for odd n. In fact, if n is odd then $a \in S(n)$ implies $-a \in S(n)$. For if $a^{n'} \equiv \pm 1 \bmod n$ then $(-a)^{n'} \equiv \mp 1 \bmod n$ which makes $-a$ a strong liar, and if $a^{2^i n'} \equiv -1 \bmod n$ for some $1 \leq i < k$, then $(-a)^{2^i n'} = (-1)^{2^i n'} \cdot a^{2^i n'} \equiv -1 \bmod n$, which again makes $-a$ a strong liar. Altogether, we see that if n is odd, then $|S(n)|$ is even and at least 2. This means we can restrict the strong test to choosing $1 < a < (n - 1)/2$ with no loss.

More generally, we would like an explicit formula for the size of $S(n)$. This was accomplished by Monier.

Proposition 2.1 ([13]). *For n odd and composite, let n' be the odd part of $n - 1$ and let r be the number of distinct prime divisors of n. Let $v = \min_{p|n}\mathrm{ord}_2(p - 1)$. Then*

$$|S(n)| = \left(1 + \frac{2^{rv} - 1}{2^r - 1}\right) \prod_{p|n} \gcd(n', p') \ .$$

It quickly follows that $|S(n)| \leq \varphi(n)/4$ when $n > 9$ is odd and composite. Thus by performing $\log_2 \left(\frac{1}{\sqrt{\epsilon}}\right)$ independent trials we can lower the probability that n is a composite falsely reported as prime to below ϵ. Note that if n is prime, Proposition 2.1 correctly gives $|S(n)| = (1 + 2^k - 1)\gcd(n', n') = n - 1$.

We briefly address the worst case, i.e. composite $n > 9$ for which $|S(n)|$ reaches the maximum of $\varphi(n)/4$. Such n are fairly easy to characterize, if not quite so easy to count. Consider the following theorem.

Theorem 2.2 ([2]). *Let C_3 be the set of odd, composite integers n with $|S(n)| > \varphi(n)/8$. Then C_3 is composed of the following:*

(1) $(m+1)(2m+1)$, *where $m+1$, $2m+1$ are odd primes,*

(2) $(m+1)(3m+1)$, *where $m+1$, $3m+1$ are primes congruent to 3 mod 4,*

(3) *Carmichael numbers n with three prime factors where there exists integer s with 2^s exactly dividing $p-1$ for all $p \mid n$,*

(4) $9, 25, 49$.

By Proposition 2.1, it follows that the n with $|S(n)| = \varphi(n)/4$ are exactly n in case (1) with $2\|m$ and Carmichael numbers in case (3) whose three prime factors are all congruent to 3 modulo 4 (this also appears in the proof to Theorem 2.2).

Unfortunately, an asymptotic formula for either case remains elusive, nor has it been proven that there are infinitely many integers in either case. On the other hand, infinitely many n of the form $(m+1)(2m+1)$ would follow from the strong prime tuples conjecture [1], and there is a precise conjecture on the number of Carmichael numbers with three prime factors.

Conjecture 2.3 ([7]). *The number of Carmichael numbers with three prime factors is asymptotic to*

$$C\frac{x^{1/3}}{\log^3 x} \ ,$$

where C is an absolute constant that can be given precisely.

We now shift to counting odd n with exactly two strong liars. The following characterization will be useful.

Proposition 2.4. *Suppose n is odd and composite. Then $|S(n)| = 2$ if and only if 1) n is divisible by $p \equiv 3$ mod 4 and 2) $\gcd(p', (n/p)') = 1$ for all primes p dividing n.*

Proof. First notice that

$$(p-1)(n/p - 1) = n - 1 - (n/p - 1) - (p-1) \ ,$$

so that $\gcd(p', n') = \gcd(p', (n/p)')$.

Thus if $\gcd(p', (n/p)') = 1$ for all $p \mid n$ the product term in Monier's formula is 1. If n is odd and divisible by $p \equiv 3$ mod 4 then $v = 1$ and we conclude that $|S(n)| = 2$.

If instead we assume $|S(n)| = 2$ then $1 + \frac{2^{rv}-1}{2^r-1} \leq 2$. If n is odd then $1 + \frac{2^{rv}-1}{2^r-1} \geq 2$, with equality only if $v = 1$. Thus n is divisible by a prime congruent to 3 modulo 4 and $\prod_{1 \leq i \leq r} \gcd(n', p_i') = \prod_{1 \leq i \leq r} \gcd(n', (n/p_i)') = 1$.   □

Monier also proved a formula for Euler liars.

Proposition 2.5 ([13]). *Let n be odd. Define $e(n) = \prod_{p|n} \gcd(\frac{n-1}{2}, p-1)$ and*

$$\delta(n) = \begin{cases} 2 & \text{if } v = \operatorname{ord}_2(n-1), \\ 1/2 & \text{if there is } p \mid n \text{ with } \operatorname{ord}_2(p-1) < \operatorname{ord}_2(n-1) \text{ and } \operatorname{ord}_p(n) \text{ odd}, \\ 1 & \text{if } \operatorname{ord}_p(n) \text{ even for all } p \mid n \text{ with } \operatorname{ord}_2(p-1) < \operatorname{ord}_2(n-1). \end{cases}$$

Then $|E(n)| = \delta(n) \cdot e(n)$.

If n is odd then the minimum number of Euler liars is 2 since ± 1 are always Euler liars.

Proposition 2.6. *Suppose n is odd and composite. Then $|E(n)| = 2$ if and only if*

(1) $n \equiv 3 \bmod 4$ *and* $\prod_{p|n} \gcd(p', n') = 1$, *or*

(2) $n \equiv 1 \bmod 4$ *with* $n = pq$, $p, q \equiv 3 \bmod 4$, *and* $\gcd(p', n') \cdot \gcd(q', n') = 1$.

Proof. First suppose that $\prod_{p|n} \gcd(p', n') = 1$. If n is equivalent to 3 modulo 4 then $\gcd(\frac{n-1}{2}, p-1) = 1$ for all $p \mid n$. Additionally, $\delta(n) = 2$ since $\mathrm{ord}_2(p-1)$ cannot be any smaller. If instead $n = pq$ with $p, q \equiv 3 \bmod 4$, then $\mathrm{ord}_2(n-1) = 2$ and so $\delta(n) = 1/2$, while $\prod_{p|n} \gcd(\frac{n-1}{2}, p-1) = 4$.

Now suppose that $|E(n)| = 2$. It is impossible to have $e(n) = 2$. For if $n \equiv 3 \bmod 4$ then the product will be odd, while if $n \equiv 1 \bmod 4$, n odd means the product will be divisible by at least one factor of 2 for each prime factor of n. Thus the only two possibilities are 1) $\delta(n) = 2$ and $e(n) = 1$; 2) $\delta(n) = 1/2$ and $e(n) = 4$.

In case 1), $e(n) = 1$ implies $n \equiv 3 \bmod 4$ and $\prod_{p|n} \gcd(n', p') = 1$, since otherwise $e(n)$ would be larger. With $n \equiv 3 \bmod 4$, it must be divisible by a prime congruent to 3 modulo 4, and so it follows that $\delta(n) = 2$.

In case 2), $\delta(n) = 1/2$ implies $n \equiv 1 \bmod 4$ and divisible by a prime congruent to 3 modulo 4. Then $e(n) = 4$ implies $\prod_{p|n} \gcd(n', p') = 1$ and n is the product of two distinct prime factors, for otherwise the power of 2 dividing $e(n)$ would be greater. $\square$

3. Preliminaries

The proofs of our asymptotic formulas will utilize a number of results from analytic number theory. Our goal is to craft an account that is readable and self-contained, and hence will not necessarily include best-possible results.

One tool will be counts of primes in arithmetic progressions. The classic result is the prime number theorem for arithmetic progressions.

Lemma 3.1. *If $\gcd(d, a) = 1$, let $\pi(x, d, a)$ denote the number of primes $\leq x$ that are congruent to $a \bmod d$. Then*

$$\pi(x, d, a) = (1 + o_d(1)) \frac{x}{\varphi(d) \log x}.$$

Next we have a version of the Brun-Titchmarsh inequality from Montgomery and Vaughan [14]. Note that the constant is absolute for arbitrary d smaller than x.

Lemma 3.2. *For $x > d \geq 1$, we have*

$$\pi(x, d, a) < \frac{2x}{\varphi(d) \log(x/d)}.$$

The Siegel-Walfisz theorem [17] gives an absolute lower bound, but the range of possible d is much smaller.

Lemma 3.3. *Assume $d \leq \log x$. Then*

$$\pi(x, d, a) = (1 + o(1)) \frac{x}{\varphi(d) \log x}.$$

We will rely on a number of prime reciprocal sums. The most basic is a result of Landau in [12, v. 1, p. 197].

Lemma 3.4. *We have*

$$\sum_{p \leq x} \frac{1}{p} = \log \log x + A + O((\log x)^{-1})$$

where A is an absolute constant.

Bounds on $\pi(x, d, a)$ lead to asymptotic formulas for prime reciprocal sums over arithmetic progressions. It is doubtful the following lemma is new, but a good reference is elusive.

Lemma 3.5. *Let $P(x, d)$ be the prime reciprocal sum over primes congruent to 1 modulo d. That is,*

$$P(x, d) = \sum_{\substack{p \leq x \\ p \equiv 1(d)}} \frac{1}{p}$$

where the sum is over primes. Then

(1) *for $1 < d \leq x$ we have $P(x, d) = (1 + o_d(1))(\log \log x)/\varphi(d)$,*
(2) *for $1 < d \leq \log x$ we have $P(x, d) = (1 + o(1))(\log \log x)/\varphi(d)$,*
(3) *for $1 < d \leq \sqrt{x}$ we have $P(x, d) < 2\varphi(d)^{-1}(\log \log x + O(1))$.*

Proof. Since $d \geq 2$, the smallest prime $p = 1 \bmod d$ is 3. Replacing the sum by a Stieltjes integral and integrating by parts, we get

$$(3.1) \qquad \sum_{\substack{p \leq x \\ p \equiv 1(d)}} \frac{1}{p} = \left.\frac{\pi(t, d, 1)}{t}\right|_{3-}^{x} + \int_3^x \frac{\pi(t, d, 1)dt}{t^2} \ .$$

The first two cases are easier. For general $d \leq x$ we apply Lemma 3.1 to get

$$\frac{\pi(x, d, 1)}{x} + \int_3^x (1 + o_d(1))\frac{1}{\varphi(d)t \log t} \ dt = \frac{1 + o_d(1)}{\varphi(d) \log x} + O(1) + \frac{1 + o_d(1)}{\varphi(d)} \log \log x$$

which is equivalent to $(1 + o_d(1))(\log \log x)/\varphi(d)$. If $d \leq \log x$ we apply Lemma 3.3 to get the same result, except that the constant in the $o(1)$ does not depend on d.

For part 3), the first term of (3.1) is

$$\frac{\pi(x, d, 1)}{x} \leq \frac{2x}{x\varphi(d) \log(x/d)} = O\left(\frac{1}{\varphi(d)}\right),$$

by Lemma 3.2 and the assumption that $d \leq \sqrt{x}$. We wish to push the lower bound of the integral to $2d$, which at worst costs us one term of the sum, and only if $d + 1$ is prime. Using Lemma 3.2 again, the revised integral is bounded by

$$\int_{2d}^{xd} \frac{\pi(t, d, 1)dt}{t^2} \leq \int_{2d}^{xd} \frac{2dt}{\varphi(d)t \log(t/d)} = 2 \int_2^x \frac{du}{\varphi(d)u \log u} \ .$$

This is $2\varphi(d)^{-1}(\log \log x + O(1))$. The lost term of the sum makes no difference, since $1/p = 1/(d + 1) \leq \varphi(d)^{-1}$. $\qquad \square$

Next we give a brief introduction to sieve theory; interested readers are encouraged to peruse [8] or [10]. Sieve theory is a collection of results for estimating the number of "survivors" that remain after we start with an interval (or other large set) and remove elements that satisfy congruence conditions. Typically, the exact formula for the number of survivors is of exponential complexity, and so one seeks approximations that are easier to evaluate but still reasonably accurate.

We use $S(x, \mathcal{P})$ to denote the count of integers up to x coprime to the elements of $\mathcal{P}$, where $\mathcal{P}$ is a set of primes. When $\mathcal{P}$ is the set of primes up to z we instead use $S(x, z)$, and we replace x with $\mathcal{X}$ when our base set is a subset of the integers up to x. Our first sieve is the Legendre sieve, an exercise in keeping track of the errors from the Sieve of Eratosthenes.

Theorem 3.6 (Legendre sieve). *Let $\mathcal{P}$ be a set of primes. Then*

$$S(x, \mathcal{P}) = x \prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p}\right) + E$$

where the error term E satisfies $|E| \leq 2^{|\mathcal{P}|}$.

Proof. Let P be the product of all primes in $\mathcal{P}$, and let d be an arbitrary divisor. Using inclusion-exclusion we obtain

$$S(x, \mathcal{P}) = \sum_{d|P} \mu(d) \left\lfloor \frac{x}{d} \right\rfloor \leq x \sum_{d|P} \left(\frac{\mu(d)}{d} + 1\right) = x \prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p}\right) + 2^{|\mathcal{P}|} .$$

For the lower bound we instead use $\mu(d)\lfloor x/d \rfloor \geq \mu(d)(x/d) - 1$. $\qquad\square$

Corollary 3.7. *Let $\mathcal{P}$ be the set of primes up to z, where $z \leq \log x$ and $z \to \infty$. Then*

$$S(x, \mathcal{P}) = S(x, z) = (1 + o(1))\frac{e^{-\gamma}x}{\log z}$$

where $\gamma \approx 0.5772$ is Euler's constant.

Proof. The product term is given by Mertens' theorem [9, Theorem 429]. For the error term note that for z large enough

$$2^{\pi(z)} \leq 2^{\frac{2z}{\log z}} \leq 2^{\frac{2\log x}{\log z}} = x^{\frac{2}{(\log_2 e)(\log z)}} = o\left(\frac{x}{\log z}\right) . \qquad\square$$

The asymptotic notation in this statement should be interpreted as follows. The variables x, z lie in a region defined by $f(x) \leq z \leq \log x$, where $f(x) \to \infty$ as $x \to \infty$. In this way the implied constant in the $o(1)$ term depends only on x.

Despite the logarithmic bound on z, Corollary 3.7 will be strong enough to give the main term in Theorem 4.3. If we are willing to settle for an upper bound, we can generalize the set of sieving primes.

Corollary 3.8. *Let $\mathcal{P}$ be an arbitrary set of primes smaller than $\log x$. Then*

$$S(x, \mathcal{P}) \leq (1 + o(1)) \cdot x \cdot \exp\left(-\sum_{p \in \mathcal{P}} \frac{1}{p}\right) .$$

Proof. By convexity we have $\log(1 - 1/p) \leq -1/p$. The error term is dealt with in similar fashion to Corollary 3.7. $\qquad\square$

The following application of the Legendre sieve will be used in the next section.

Corollary 3.9 ([3]). *Let $p \to \infty$, $x \to \infty$ with $p + \log p \leq \log x$. Denote by $C_p(x)$ the number of integers $n \leq x$ for which the least prime factor of n is p. Then*

$$C_p(x) = (1 + o(1))\frac{xe^{-\gamma}}{p \log p} .$$

Proof. Note the least prime factor of n is p if and only if n/p is coprime to all primes smaller than p. Since $p + \log p \leq \log x$ implies $p \leq \log(x/p)$, we apply Corollary 3.7 to obtain

$$C_p(x) = S(x/p, p) = (1 + o(1))\frac{e^{-\gamma}x}{p \log p} \ . \qquad \square$$

The Legendre sieve can also be extended to other initial sets. For example, Let $\mathcal{X}$ be the set of integers $n \leq x$ that are congruent to 3 modulo 4. Then by the Chinese Remainder Theorem, the size of the subset of $\mathcal{X}$ divisible by d odd is $x/(4d) + O(1)$. The subset is empty if d is even.

Theorem 3.10. *Assume that $z \leq \log x$ with $z \to \infty$, and let $\mathcal{P}$ be the set of odd primes up to z. Then*

$$S(\mathcal{X}, \mathcal{P}) = (1 + o(1))\frac{e^{-\gamma}x}{2 \log z} \ .$$

Proof. Let P be the product of all odd primes up to z. Applying the Legendre sieve, we have

$$S(\mathcal{X}, \mathcal{P}) = \sum_{d|P} \left(\mu(d)\frac{x}{4d} + O(1) \right) = \frac{x}{4} \prod_{p|P} \left(1 - \frac{1}{p} \right) + O(2^{\pi(z)})$$

$$= \frac{x}{2} \prod_{p \leq z} \left(1 - \frac{1}{p} \right) + O(2^{\pi(z)}) \ .$$

With $z \leq \log x$, Corollary 3.7 gives the result. $\qquad \square$

For some results we will need a stronger sieve, i.e. one where z can grow larger than $\log x$. The following special case of the Brun sieve adapted from [8, Section 3.2.3] will suffice. For sifting density we use the simpler characterization found in [8, Section 1.3.5].

Theorem 3.11 (Brun sieve). *Let $\mathcal{P}$ be a set of primes all less than z and let d be a divisor of $\prod_{p \in \mathcal{P}} p$. Assume $\mathcal{P}$ has sifting density $\kappa > 0$, i.e. there is a constant $A > 1$ such that*

$$\sum_{w \leq p < z} \frac{\log p}{p - 1} \leq \kappa \log\left(\frac{z}{w}\right) + A \text{ when } 2 \leq w < z \text{ and } p \in \mathcal{P} \ .$$

Then

$$S(x, \mathcal{P}) \sim x \prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p} \right)$$

as $x \to \infty$, uniformly in $z \leq x^{1/(c\kappa \log \log x)}$, where c is an absolute constant.

As an application we give an upper bound on the count of $n \leq x$ divisible by only primes $\equiv 1 \bmod 4$. Despite being far from best-possible, it is adequate for our needs in a later proof.

Corollary 3.12. *The count of $n \leq x$ divisible by only primes $\equiv 1 \bmod 4$ is $o(x/\log \log x)$.*

Proof. The count desired can be obtained by sieving all primes $p \equiv 3 \bmod 4$, and if we restrict the set of sieving primes the count only gets larger. So let $\mathcal{P}$ be the set of primes $p \leq x^{1/(\log \log x)^2}$ with $p \equiv 3 \bmod 4$. With x large enough we have $x^{1/(\log \log x)^2} \leq x^{1/(c\kappa \log \log x)}$ and the Brun sieve applies, giving a count after sieving

of $(1 + o(1))x \prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p}\right)$. Then Corollary 3.8 and part (2) of Lemma 3.5 yields an upper bound of

$$(1 + o(1))x \cdot \exp\left(-\sum_{p \in \mathcal{P}} \frac{1}{p}\right) \leq (1 + o(1))x \cdot \exp\left(-\frac{(1 + o(1))}{2} \log \log x\right) \ . \qquad \square$$

Finally, in Section 6 we will frequently use various measures for the average number of prime factors of a number.

Lemma 3.13. *Let $\omega(n)$ be the number of distinct prime factors of n and $\Omega(n)$ the total number of prime factors of n. Let p be a prime. Then*

$$\sum_{n \leq x} \omega(n) = O(x \log \log x) \ ,$$

$$\sum_{n \leq x} \Omega(n) = O(x \log \log x) \ ,$$

$$\sum_{p \leq x} \Omega(p - 1) = O\left(\frac{x \log \log x}{\log x}\right) \ .$$

Proof. For the first two see [9, Theorem 430]. The third is a special case of [4, Lemma 2.1]. $\qquad \square$

4. Two strong liars

Denote by $A(x)$ the number of odd $n \leq x$ with $\prod_{p|n} \gcd(n', p') = 1$. Let $A_r(x)$ be the count of such n whose least prime dividing n' is r. Then $A(x) = \lfloor \log_2 x \rfloor + \sum_r A_r(x)$, where $\lfloor \log_2 x \rfloor$ counts n for which $n - 1$ is a power of 2.

We will break $\sum_r A_r(x)$ into three sums depending on whether $r < (\log \log x)^{1-\epsilon}$, $(\log \log x)^{1-\epsilon} \leq r \leq (\log \log x)^{1+\epsilon}$, or $r > (\log \log x)^{1+\epsilon}$. Call these, respectively, $\sum_1, \sum_2, \sum_3$. For ease of notation we use z_1 for $(\log \log x)^{1-\epsilon}$ and z_2 for $(\log \log x)^{1+\epsilon}$, while z will denote a generic bound on r. This strategy mirrors closely an argument from [3] (thanks to Carl Pomerance for help with a particularly perplexing point). In fact, upper bounds on all three of $\sum_1, \sum_2, \sum_3$ are identical to those used by Erdős. However, the new definition of $A_r(x)$ required for the current work does necessitate a different approach for the lower bound to $\sum_3$. A new writeup is useful for other reasons: we have streamlined the discussion of prime reciprocal sums and clarified the derivation of the upper bound to $\sum_2$.

Here and in the next section, we wish to prove that $\lim_{x \to \infty} f(x) = a$ where $f(x)$ is the quotient of our target function and a simpler approximation. To prove a sequence a_n has the limit a, it is sufficient to show that for every $\epsilon > 0$,

$$a - \epsilon \leq \liminf a_n \leq \limsup a_n \leq a + \epsilon \ .$$

Lemma 4.1. *Let $0 < \epsilon < 1$ and $z_1 = (\log \log x)^{1-\epsilon}$. We have*

$$\sum_{r < z_1} A_r(x) = o_\epsilon \left(\frac{x}{\log \log x}\right) \ .$$

Proof. Suppose that n is counted by $A_r(x)$ with r an odd prime less than z_1. Then $n \equiv 1 \bmod r$, but must not be divisible by any $p \equiv 1 \bmod r$. So an upper bound on $A_r(x)$ is given by the count of n not divisible by any $p \equiv 1 \bmod r$, and the

count is further enlarged if we restrict our sieving set $\mathcal{P}$ to primes $p \equiv 1 \bmod r$ with $p < x^{1/(\log \log x)^2}$.

Now the Brun sieve applies. We use the upper bound from Corollary 3.8 and the unconditional lower bound from Lemma 3.5 (note z_1 is small enough so $r \le \log x$). For every ϵ, we can take x large enough so that

$$A_r(x) \sim x \prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p}\right) \le x \cdot \exp\left(-\sum_{p \in \mathcal{P}} \frac{1}{p}\right) \le x \cdot \exp\left(-\frac{(1 + o(1)) \log \log x}{2\varphi(r)}\right) \quad .$$

Since $r \le (\log \log x)^{1-\epsilon}$, that last term is $o(x/(\log \log x)^2)$. Then

$$\sum_{r < z_1} A_r(x) \le (\log \log x)^{1-\epsilon} \cdot o\left(\frac{x}{(\log \log x)^2}\right) = o_\epsilon\left(\frac{x}{\log \log x}\right) \quad . \qquad \square$$

Lemma 4.2. *Let $0 < \epsilon < 4/5$, $z_1 = (\log \log x)^{1-\epsilon}$, and $z_2 = (\log \log x)^{1+\epsilon}$. Then there exists a $c > 0$ such that*

$$\sum_{z_1 \le r \le z_2} A_r(x) \le c \frac{\epsilon x}{\log \log \log x}$$

for x sufficiently large.

Proof. For $z_1 \le r \le z_2$ we use a different upper bound on $A_r(x)$, namely the count of $n \le x$ with r as the smallest prime factor of $n - 1$. This is at most one away from the count of $n \le x$ whose least prime factor is r. By Corollary 3.9, for large enough x this count is upper bounded by

$$\frac{2xe^{-\gamma}}{r \log r} \quad .$$

Then for $x > x_0$ (independent of ϵ),

$$\sum_{r=z_1}^{z_2} \frac{2xe^{-\gamma}}{r \log r} \le \frac{2xe^{-\gamma}}{\log((\log \log x)^{1-\epsilon})} \sum_{r=z_1}^{z_2} \frac{1}{r} \le \frac{10xe^{-\gamma}}{\log \log \log x} \left(\log\left(\frac{1+\epsilon}{1-\epsilon}\right) + o(1)\right)$$

where the sum is resolved via Lemma 3.4. Note $\epsilon < 4/5$ implies $\log(\frac{1+\epsilon}{1-\epsilon}) \le 3\epsilon$, and that the $o(1)$ term will be smaller than ϵ for $x > x_1(\epsilon)$. $\qquad \square$

The final term is the one that will have the largest magnitude. If n is counted by $\sum_{r > z_2} A_r(x)$ then $n - 1$ has no odd prime factor smaller than z_2. We apply the Legendre sieve.

Theorem 4.3. *The number of odd $n \le x$ with exactly two strong liars is given by*

$$(1 + o(1)) \frac{xe^{-\gamma}}{\log \log \log x} \quad .$$

Proof. The main work is in counting odd $n \le x$ with $\prod_{p|n} \gcd(n', p') = 1$, and the main term is $\sum_{r > z_2} A_r(x)$. This is smaller than the count of n where $n - 1$ has no prime divisor smaller than z_2. With $z_2 < \log x$, Corollay 3.7 gives us

$$\sum_{r > z_2} A_r(x) \le S(x, z_2) = (1 + o(1)) \frac{xe^{-\gamma}}{\log z_2} = \frac{(1 + o(1))xe^{-\gamma}}{(1 + \epsilon) \log \log \log x} \quad .$$

For a lower bound we exclude $n \equiv 1 \bmod r$ that are divisible by a prime $p \equiv 1 \bmod r$, and do this for all prime $r > z_2$. For a given r the number of n excluded is

$$\sum_{p \equiv 1 \bmod r} \frac{x}{pr} + O(1)$$

since the condition $n \equiv 0 \bmod p$ and $n \equiv 1 \bmod r$ repeats every pr integers by the Chinese Remainder Theorem. Note that $p \equiv 1 \bmod r$ and $pr \leq x$ implies that $r \leq \sqrt{x}$. So applying Lemma 3.5 (uniform upper bound) and Lemma 3.2 gives

$$\sum_{z_2 < r < \sqrt{x}} \sum_{p \equiv 1 \bmod r} \frac{x}{pr} + O(1) \leq \sum_{z_2 < r < \sqrt{x}} \frac{x}{r} \frac{2}{\varphi(r)} (\log\log x + O(1)) + O\left(\frac{x}{\varphi(r) \log x}\right)$$

$$\leq O\left(\sum_{z_2 < r < \sqrt{x}} \frac{x \log\log x}{r^2}\right) + O\left(\sum_{z_2 < r < \sqrt{x}} \frac{x}{r^2}\right) + O\left(\frac{x \log\log x}{\log x}\right) \; .$$

Taking the sum over integers rather than over primes, we have

$$\sum_{z_2 < r < \sqrt{x}} \frac{1}{r^2} < \int_{z_2}^{\sqrt{x}} \frac{1}{r^2} \, \mathrm{d}r = -\frac{1}{\sqrt{x}} + \frac{1}{z_2} < \frac{1}{(\log\log x)^{1+\epsilon}} \; .$$

So the amount we are subtracting is upper bounded by

$$O\left(\frac{x}{(\log\log x)^\epsilon}\right) + O\left(\frac{x}{(\log\log x)^{1+\epsilon}}\right) + O\left(\frac{x \log\log x}{\log x}\right) = o_\epsilon\left(\frac{x}{\log\log\log x}\right) \; .$$

Let $0 < \epsilon < 4/5$ be arbitrary. By Lemma 4.1, for large enough x we have $\frac{\Sigma_1}{x/\log\log\log x} < \epsilon$. Then by Lemma 4.2 and the work above, we see that

$$\frac{e^{-\gamma}}{1+\epsilon} - \epsilon \leq \liminf \frac{A(x)}{x/\log\log\log x} \leq \limsup \frac{A(x)}{x/\log\log\log x} \leq \frac{e^{-\gamma}}{1+\epsilon} + \epsilon + c\epsilon \; .$$

Since $0 < \epsilon < 4/5$ was arbitrary, the limit exists and the proper constant is indeed $e^{-\gamma}$. As far as being divisible by at least one prime $\equiv 3 \bmod 4$, by Corollary 3.12 the number of $n \leq x$ only divisible by primes $\equiv 1 \bmod 4$ is $o(x/\log\log x)$. The characterization in Proposition 2.4 now finishes the proof. $\square$

The same proof technique can be extended to counting n with exactly two Euler liars.

Theorem 4.4. *The number of $n \leq x$ with exactly two Euler liars is given by*

$$(1 + o(1)) \frac{x e^{-\gamma}}{2 \log\log\log x} \; .$$

Proof. The characterization is given by Proposition 2.6; we start with the first case. We use the same proof technique as that for Theorem 4.3. For all terms except the main term, we can drop the condition that $n \equiv 3 \bmod 4$ at no loss. It does affect the main term however: by Theorem 3.10 the count of $n \leq x$ with $n \equiv 3 \bmod 4$ and not divisible by any factor less than $\log\log x$ is given by

$$(1 + o(1)) \frac{x}{2 e^\gamma \log\log\log x} \; .$$

The second case is asymptotically smaller, since the number of $n \leq x$ with two prime factors is $O((x \log\log x)/(\log x))$. $\square$

5. TWO STRONG LIARS AND TWO PRIME FACTORS

Our goal in this section is to prove the second of the three main theorems given in the introduction, thus providing an asymptotic formula for the count of odd $n \leq x$ with two strong liars and two prime factors. Before discussing this in detail, we note that there are $(1 + o(1))(x \log \log x)/(\log x)$ numbers $n \leq x$ that are a product of two primes (this result is due to Landau [11], see also Wright [18]). The constant

$$C = \prod_{p > 2} \left(1 - \frac{1}{(p-1)^2} \right)$$

is what we would expect from the following heuristic assumption: the two prime factors of n are chosen independently, and fall into congruence classes in the "correct" proportion. The task, therefore, is to make this rigorous. The main idea of the proof will be to approximate a count using a fixed number of terms of the inclusion-exclusion formula, and then use a union bound to show that the approximate count is good enough. Hooley [10] has called this strategy the "simple asymptotic sieve".

In this section, p and q denote odd primes with $p \leq q$, and d denotes a positive integer. We now introduce several sets:

$$T = \{pq \leq x : \gcd(p', q') = 1\},$$

$$T' = \{pq \leq x : \gcd(p', q') > 1\},$$

$$S = \{pq \leq x : p, q \text{ odd }\},$$

$$S_d = \{pq \leq x : p \equiv q \equiv 1(d)\},$$

$$S^{(B)} = \{n \in T' : \gcd(p', q') > 1 \text{ and has no primes} < B \}.$$

Note that if d is odd, $S_d = \{pq \leq x : d \text{ divides } \gcd(p', q')\}$.

Our first two tasks are to show that $S^{(B)}$ is not too large, then to approximately count S_d. A good tool for the first job is the Brun-Titchmarsh theorem (Lemma 3.2), but the factor $\log(x/d)$ in its denominator can give trouble when d is close to x. Our way around this is inspired by the chess player's gambit: give up a piece now to win later. More precisely, we will increase x, thereby bringing the log factor under control at the price of a slightly worse upper bound that is still good enough. For the second job, since we will only be concerned with a fixed number of d's (depending on B), we can rely on a non-uniform version of the prime number theorem for arithmetic progressions (Lemma 3.1).

Lemma 5.1. *We have*

$$|S^{(B)}| = O \left(\frac{x \log \log x}{\sqrt{B} \log x} \right) ,$$

where the implied constant is absolute.

Proof. Let b be an odd prime. We first find an upper bound for $|S_b|$, namely

$$\sum_{\substack{p \leq \sqrt{x} \\ p \equiv 1(b)}} \#\{q : p \leq q \leq x/p \text{ and } q \equiv 1(b)\} \leq \sum_{\substack{p \leq \sqrt{x} \\ p \equiv 1(b)}} \#\{q : q \leq b^{1/2}x/p \text{ and } q \equiv 1(b)\} .$$

We may assume that $b \leq \sqrt{x}$, since the sum vanishes otherwise (note that $b^2 \leq pq = x$). Then we are guaranteed that $b^{1/2}x/p > b$ since $p\sqrt{b} < pb \leq pq \leq x$. This

allows us to estimate the summand using Lemma 3.2, and thereby get

$$|S_b| \leq \sum_{\substack{p \leq \sqrt{x} \\ p \equiv 1(b)}} \frac{2b^{1/2}x}{p\varphi(b)\log(x/(pb^{1/2}))} \ .$$

We know that $p \leq \sqrt{x}$ and $\sqrt{b} \leq x^{1/4}$, making $x/(p\sqrt{b}) \geq x^{1/4}$. Therefore,

$$\frac{1}{\log(x/(p\sqrt{b}))} \leq \frac{4}{\log x}$$

and thus

$$|S_b| \leq \frac{8\sqrt{b}x}{\varphi(b)\log x} \sum_{\substack{p \leq \sqrt{x} \\ p \equiv 1(b)}} \frac{1}{p} \ .$$

By Lemma 3.5, the inner sum has an upper bound of $2\varphi(b)^{-1}(\log\log x + O(1))$. Summing over all primes $b \geq B$, we get the result. $\qquad \square$

Lemma 5.2. *Let $d \geq 1$. Then*

$$|S_d| = (1 + o_d(1))\frac{x\log\log x}{\varphi(d)^2\log x} \ .$$

Proof. We have

$$|S_d| = \sum_{\substack{p \leq \sqrt{x} \\ p \equiv 1(d)}} \#\{q : p \leq q \leq x/p \text{ and } q \equiv 1(d)\} \ .$$

If we drop the lower bound on q, we incur an error that is no more than

$$\sum_{p \leq \sqrt{x}} \#\{q : q \leq p\} \leq \sum_{p \leq \sqrt{x}} \#\{q : q \leq \sqrt{x}\} = O\left(\frac{x}{\log^2 x}\right) \ .$$

Accordingly, we can work with the simpler sum

$$(5.1) \qquad \sum_{\substack{p \leq \sqrt{x} \\ p \equiv 1(d)}} \#\{q : q \leq x/p \text{ and } q \equiv 1(d)\} \ .$$

Fix ϵ with $0 < \epsilon < 1/2$. We will split the sum, using the break point $p = x^\epsilon$. The contribution to (5.1) from the $p \leq x^\epsilon$ is

$$(5.2) \qquad \sum_{\substack{p \leq x^\epsilon \\ p \equiv 1(d)}} \#\{q : q \leq x/p \text{ and } q \equiv 1(d)\} = \sum_{\substack{p \leq x^\epsilon \\ p \equiv 1(d)}} (1 + o_d(1))\frac{x}{\varphi(d)p\log(x/p)}$$

by Lemma 3.1. The assumption $p \leq x^\epsilon$ implies $x/p \geq x^{1-\epsilon}$. Thus $(1-\epsilon)\log x \leq \log(x/p) \leq \log x$, which gives

$$\frac{(1 + o_d(1))x}{\varphi(d)\log x} \sum_{\substack{p \leq x^\epsilon \\ p \equiv 1(d)}} \frac{1}{p} \leq (5.2) \leq \frac{(1 + o_d(1))x}{(1-\epsilon)\varphi(d)\log x} \sum_{\substack{p \leq x^\epsilon \\ p \equiv 1(d)}} \frac{1}{p} \ .$$

For the prime reciprocal sum we apply Lemma 3.5, which yields

$$\sum_{\substack{p \leq x^\epsilon \\ p \equiv 1(d)}} \frac{1}{p} = (1 + o_d(1))\frac{\log\log x^\epsilon}{\varphi(d)} = (1 + o_d(1))\frac{\log\log x + \log \epsilon}{\varphi(d)} \ .$$

This gives

$$(1 + o_d(1))\frac{x(\log\log x + \log\epsilon)}{\varphi(d)^2 \log x} \le (5.2) \le (1 + o_d(1))\frac{x(\log\log x + \log\epsilon)}{(1-\epsilon)\varphi(d)^2 \log x}$$

and since ϵ was arbitrary and $x \to \infty$, we conclude

$$(5.2) = (1 + o_d(1))\frac{x\log\log x}{\varphi(d)^2 \log x}.$$

To finish off the proof, we will show that the sum over primes larger than x^ϵ does not grow this quickly. This sum is

$$(5.3) \qquad \sum_{\substack{x^\epsilon < p \le \sqrt{x} \\ p \equiv 1(d)}} \#\{q : q \le x/p \text{ and } q \equiv 1(d)\}$$

and it has an upper bound of

$$\sum_{x^\epsilon < p \le \sqrt{x}} \#\{q : q \le x/p\} = \sum_{x^\epsilon < p \le \sqrt{x}} (1 + o(1))\frac{x}{p\log(x/p)} \le \frac{(1 + o(1))x}{(1/2)\log x} \sum_{x^\epsilon < p \le \sqrt{x}} \frac{1}{p}.$$

By Lemma 3.4 the inner sum is $\log(1/2) - \log\epsilon + O((\log x)^{-1})$. So

$$(5.3) = O_\epsilon\left(\frac{x}{\log x}\right)$$

and we are done. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

We are now ready for the main event. Our strategy will be to first estimate how many odd pq's satisfy only condition (2) in the characterization of strong liars in Proposition 2.4 (this is of interest by itself). Then, we make a similar estimate under the additional requirement that at least one of p and q be 3 modulo 4.

In the next two theorems,

$$C := \prod_{p>2}\left(1 - \frac{1}{(p-1)^2}\right) = 0.66016... .$$

Theorem 5.3. *The number of odd $n = pq \le x$ with $\gcd(p', q') = 1$ is*

$$(5.4) \qquad\qquad (1 + o(1))\frac{Cx\log\log x}{\log x}.$$

Proof. We have

$$\bigcup_{\substack{2 < r < B \\ r \text{ prime}}} S_r \subseteq T' \subseteq S^{(B)} \cup \bigcup_{\substack{2 < r < B \\ r \text{ prime}}} S_r.$$

The set of odd primes up to B is finite, so just as with the Legendre sieve we can use inclusion-exclusion and get

$$\sum_r |S_r| - \sum_{r,s} |S_{rs}| + \sum_{r,s,t} |S_{rst}| + \cdots \le |T'| \le |S^{(B)}| + \sum_r |S_r| - \sum_{r,s} |S_{rs}| + \sum_{r,s,t} |S_{rst}| + \cdots$$

where $r, s, t, \ldots$ denote distinct odd primes $< B$. Letting P be the product of the odd primes less than B and d a divisor of P, this can be rewritten as

$$(5.5) \qquad\qquad \sum_{\substack{d|P \\ d>1}} -\mu(d)|S_d| \le |T'| \le |S^{(B)}| + \sum_{\substack{d|P \\ d>1}} -\mu(d)|S_d|.$$

Since S is the disjoint union of T and T', and $S = S_1$, we have

$$|T| = |S| - |T'| = |S_1| - |T'|.$$

Combining this with (5.5) we get

$$-|S^{(B)}| + \sum_{d|P} \mu(d)|S_d| \leq |T| \leq \sum_{d|P} \mu(d)|S_d|.$$

Since B is fixed, we can use Lemma 5.2 to express the sum over d as

$$(C_B + o_B(1))\frac{x \log \log x}{\log x}.$$

where

$$C_B = \prod_{3 \leq p < B} \left(1 - \frac{1}{(p-1)^2}\right) = \sum_{d|P} \frac{\mu(d)}{\varphi(d)^2}.$$

Combining these results with Lemma 5.1, we then get

$$\frac{|T|}{x \log \log x / \log x} = C_B + o_B(1) + O(B^{-1/2}).$$

For any $\delta > 0$, we can choose a B for which both $C_B - C$ and the $B^{-1/2}$ term are bounded by $\delta/4$ in absolute value. With any such choice of B, the $o_B(1)$ term will be no more than $\delta/2$ for sufficiently large x, so

$$C - \delta \leq \liminf \frac{|T|}{x \log \log x / \log x} \leq \limsup \frac{|T|}{x \log \log x / \log x} \leq C + \delta.$$

Since δ is arbitrary, we conclude that the limit as $x \to \infty$ exists and equals C. $\square$

Theorem 5.4. *The number of odd $n = pq \leq x$ with $|S(n)| = 2$ is*

$$(5.6) \qquad\qquad (1 + o(1))\frac{3C \cdot x \log \log x}{4 \log x}.$$

Proof. By Proposition 2.4 we need to count odd $n \leq x$ with $n = pq$, $\gcd(p', q') = 1$, and at least one of p, q congruent to 3 modulo 4. Since we have the count of $n \leq x$ with two prime factors and $\gcd(p', q') = 1$, it suffices to subtract those where $p \equiv q \equiv 1 \bmod 4$.

This is very similar to the proof of Theorem 5.3, so we only note the differences. First, the "universe" S is no longer S_1 but S_4. Second, we define $T_{4,1}$ and $T'_{4,1}$ similarly to T and T', but with the additional requirement that $p \equiv q \equiv 1 \bmod 4$. We still use P for the product of all odd primes up to B. Then, as before, a combinatorial argument gives

$$-|S^{(B)}| + \sum_{d|P} \mu(d)|S_{4d}| \leq |T_{4,1}| \leq \sum_{d|P} \mu(d)|S_{4d}| \ ,$$

where we reuse $S^{(B)}$ since keeping primes that are 3 mod 4 creates a larger set and hence a smaller sum. By Lemma 5.2 we have

$$\sum_{d|P} \mu(d)|S_{4d}| = \left(\frac{1}{4} \prod_{3 \leq p < B} \left(1 - \frac{1}{(p-1)^2}\right) + o_B(1)\right)\frac{x \log \log x}{\log x}.$$

The rest of the proof proceeds just as for Theorem 5.3. $\square$

In the table below, count 1 is the number of $n \leq x$ with $n = pq$ and $\gcd(p', q') = 1$. Count 2 adds the condition that $p \equiv q \equiv 1 \bmod 4$. As can be seen from the first two columns, Theorem 5.3 is reasonably accurate, despite the slowly growing $\log \log x$ factor. In the last two columns the asymptotic expression can be seen to be a bit of an overestimate. We believe this reflects "Chebyshev's bias", whereby the residue class 1 mod 4 gets, among small primes, noticeably less than its fair share.

x	count 1	prediction	count 2	prediction
10^3	166	184.70	28	46.17
5×10^3	795	830.16	149	207.54
10^4	1544	1591.44	298	397.86
5×10^4	7246	7264.91	1473	1816.23
10^5	14027	14011.09	2872	3502.77
5×10^5	65442	64754.58	13681	16188.65
10^6	127207	125471.12	26792	31367.78
5×10^6	595382	585478.01	126898	146369.50
10^7	1159409	1138603.46	248242	284650.87
5×10^7	5459378	5353378.05	1178844	1338344.51
10^8	10653388	10441331.16	2307619	2610332.79
5×10^8	50424160	49392155.46	10991685	12348038.86
10^9	98596968	96563937.17	21542038	24140984.29

6. TABULATION ALGORITHM

In this section we design and analyze an algorithm that tabulates all $n \leq x$ with exactly two strong liars, thus giving an exact count. This appears to be unstudied, so we start with naive ideas and improve upon them.

First, we address the costs of basic operations. We can add two numbers with k bits using $O(k)$ bit operations, and we use $M(k)$ to denote the cost of multiplying two k bit numbers. A classic fast multiplication algorithm is that of Schönhage and Strassen with $M(k) = O(k \log k \log \log k)$, but it has been recently superseded by [6]. A good discussion along with a table comparing different multiplication algorithms may be found in [16, Section 8.3]. Finally, for integers of k bits the best gcd algorithm takes $O(M(k) \log k)$ bit operations [15].

Turning to tabulation algorithms, a truly naive method would be to consider each n in turn by factoring and then applying Monier's formula. Since factoring is expensive for an individual n but has a cheap amortized cost when factoring a range of n, we instead generate all factorizations first before applying Proposition 2.1.

To factor all positive integers $n \leq x$, we will generate an array where the largest prime factor of n is stored at index n. To do so, initialize the array with all zeros. Starting with $p = 2$, let p be the next largest index whose value is 0. Then take all indices that are a multiple of p and overwrite the value with p. Do this for all $p \leq x$. Each operation is an addition, and the total number of operations is

$$\sum_{p \leq x} \frac{x}{p} = O(x \log \log x) \ ,$$

making the total complexity $O(x \log x)$ space and $O(x(\log x)(\log \log x))$ bit operations. One can retrieve the factorization of n at an amortized cost of $M(\log n) \log \log n$

by dividing n by p and then recursively looking up the largest prime factor of n/p in the table. The average of $\log\log n$ for the number of prime factors of n comes from Lemma 3.13.

This then gives Algorithm 0: factor all integers $n \leq x$, then apply Monier's formula to each n. Generating the array with the largest prime factor of each $n \leq x$ costs $O(x(\log x)(\log\log x))$ bit operations and uses $O(x \log x)$ space. We then have a gcd check for each distinct prime divisor of n, which by Lemma 3.13 is a total of $O(x \log\log x)$ gcd's at a total cost of $O(xM(\log x)(\log\log x)^2)$ bit operations. The total cost of the factorization retrievals is $O(xM(\log x)(\log\log x))$ by the same theorem, and doesn't affect the asymptotic running time.

Our first improvement will be to reduce the number of gcd checks. For all $p \mid n$, form $\ell = \prod_{p \mid n} p'$. Then checking Monier's formula only requires a single gcd application. This improvement is implemented in Algorithm 1.

Algorithm 1: Naive tabulation

---

1 Factor all $n \leq x$ using a sieve ;
  /* For each n build p' product and gcd with n' */
2 for $n \leq x$ do
3 | $\ell = 1$;
4 | for $p \mid n$ do
5 | | $\ell \leftarrow \ell \cdot p'$;
6 | if $\gcd(\ell, n') \neq 1$ then
7 | | set 0 ;
8 | if n odd and all $p \equiv 1 \bmod 4$ then
9 | | set 0 ;
10 | set 1 ;

Note that the tabulation includes even n with $\prod_{p \mid n} \gcd(p', n') = 1$, but it is trivial to isolate the odd survivors.

Theorem 6.1. *Algorithm* 1 *stores at most* $O(x \log\log x)$ *integers* $\leq x$ *and runs using* $O(x \cdot M(\log x) \log\log x)$ *bit operations.*

Proof. As discussed, the factoring step costs $O(x(\log x)(\log\log x))$ time and $O(x \log x)$ space to generate the array and $O(xM(\log x)(\log\log x))$ bit operations to generate all the factorizations over the course of the algorithm.

Algorithm 1 then does a multiplication for every distinct prime divisor of every $n \leq x$, a total of $O(x \log\log x)$ multiplications by Lemma 3.13. The algorithm also does x gcd computations at a cost of $O(xM(\log x)(\log\log x))$ bit operations, and the multiplications have the same total cost. $\qquad\square$

In developing a better tabulation algorithm we seek to use more of a sieve strategy. Note that for a given prime p with $p-1$ divisible by an odd prime r, we can cross off any $n = p \cdot d$ where $d \equiv 1 \bmod r$. For in this case $\gcd(p', (n/p)') \neq 1$ and thus $|S(n)| > 2$ by Proposition 2.4. Such n are exactly those in the arithmetic progression

$$\{n = p + kpr \ : \ k \in \mathbb{N}\}.$$

In Algorithm 2, checking all prime factors for one that is congruent to 3 modulo 4 would be too expensive. Thus we add another sieving step, and introduce three states for each integer. Integers start out labeled "2". If they fail to have $\prod_{p|n} \gcd(p', n') = 1$ they get labeled "0". Finally, those divisible by a prime congruent to 3 modulo 4 are labeled "1" and counted.

Algorithm 2: Sieving tabulation

---

1 Generate an array with the largest prime factor of n for all $n \leq x$;

2 Initialize new array with 2 in each odd entry, 0 in each even ;

3 for *primes $p \leq x$* **do**

4 $\quad$ generate factorization of $p - 1$;

5 $\quad$ **for** *odd prime $r \mid (p - 1)$* **do**

6 $\quad\quad$ **for** $n \equiv p \bmod pr$ **do**

7 $\quad\quad\quad$ set 0 ;

8 for *primes $p \leq x$ with $p \equiv 3 \bmod 4$* **do**

9 $\quad$ **for** *multiples of p with value 2* **do**

10 $\quad\quad$ set 1 ;

11 count n with value 1 ;

Theorem 6.2. *Algorithm 2 stores at most $O(x)$ integers $\leq x$ and runs using $O(x(\log x)(\log \log x))$ bit operations.*

Proof. Generating the array of largest prime factors takes $O(x(\log x)(\log \log x))$ bit operations and $O(x \log x)$ space, as does the final sieving at line (8). The main difficulty of the algorithm is the loop at line (3).

With the array of largest prime factors in hand, identifying primes is easy. Then the main loop has two components. The first is generating the factorization of $p - 1$ for all primes up to x. Since $\sum_{p \leq x} \Omega(p - 1) = O(x \log \log x / \log x)$ by Lemma 3.13, the total cost in bit operations is

$$O\left(\frac{x M(\log x) \log \log x}{\log x} \right).$$

Even using a naive multiplication algorithm with $M(\log x) = O(\log x)^2$, this is no worse than $O(x(\log x)(\log \log x))$.

The second component of the main loop involves checking each element of the sequence $n = p + kpr$, where p runs over primes up to x and r runs over the distinct prime divisors of $p - 1$. Generating such a sequence requires $x/(pr)$ additions, making the total number of additions

$$\sum_{p \leq x} \sum_{r | p-1} \frac{x}{pr}.$$

To evaluate this sum, we reverse the order of summation. This same sum appeared in the proof of Theorem 4.3; note that once again $r \mid p - 1$ and $pr \leq x$ implies

$r \leq \sqrt{x}$. We have

$$x \sum_{r \leq \sqrt{x}} \frac{1}{r} \sum_{\substack{p \leq x \\ p \equiv 1 \bmod r}} \frac{1}{p} = O\left(x \sum_{r \leq \sqrt{x}} \frac{1}{r} \frac{\log \log x}{\varphi(r)} \right) = O\left(x \log \log x \sum_{r \leq \sqrt{x}} \frac{1}{r^2} \right).$$

The sum over primes has a constant upper bound. Thus there are $O(x \log \log x)$ additions at a cost of $O(x(\log x)(\log \log x))$ bit operations. $\quad\square$

Algorithm 2 was implemented, giving the following values of $C(x)$, where $C(x)$ is the count of odd composite $n \leq x$ with exactly two strong liars.

x	$C(x)$	$C(x) \cdot \frac{\log \log \log x}{x}$
10^3	242	0.1595
10^4	2552	0.2036
10^5	25954	0.2319
10^6	261279	0.2522
10^7	2616236	0.2675
10^8	26140022	0.2795
10^9	260899380	0.2893

With $e^{-\gamma} \approx 0.5615$, we see that convergence to the asymptotic formula is quite slow. As Daniel Shanks once wrote, $\log \log \log x$ does go to infinity, but "with great dignity".

7. Conclusions and future work

It is interesting that counts of $n \leq x$ with $\gcd(n, \varphi(n)) = 1$, with $|F(n)| = 1$, and with $|S(n)| = 2$ all have the same asymptotic formula, and it suggests that there might be some general class of arithmetic sets whose size can be approximated by the set of n with no prime factor smaller than $\log \log x$. The set of n satisfying the best case for the Lucas pseudoprime test would be well worth studying next. It is worth remarking that the three sets $\{n \leq x \ : \ \gcd(n, \varphi(n)) = 1\}, \{n \leq x \ : \ |F(n)| = 1\}, \{n \leq x \ : \ |S(n)| = 2\}$ are not the same. For 9 has two strong liars, but $\gcd(9, \varphi(9)) \neq 1$ and 9 has two Fermat liars rather than one. Also, 15 satisfies $\gcd(15, \varphi(15)) = 1$, but 15 has more than one Fermat liar.

The authors of [5] give a number of other results regarding the size of $|F(n)|$, and it would be worth extending those results to $|S(n)|$ and $|E(n)|$. Our Theorem 5.4 is in a different vein, and it would be nice to extend it to counts of n with two strong liars and k prime factors for $k > 2$.

The slow rate of convergence of exact counts of n with two strong liars to the asymptotic formula cries out for a more precise formula with a faster rate of convergence. It seems that a large part of the error comes from the fact that the Mertens bound is not very accurate when one only sieves by small primes. As for why the count of n with two strong liars is approximated by sieving up to $\log \log x$, consider the following heuristic argument. A typical n will have $\log \log n$ prime factors p. For a given prime r of size roughly $\log \log n$, the expected number of p with $r \mid p-1$ is one. Since we need $n-1$ to not be divisible by r, we exclude all the $n \leq x$ with $n-1$ divisible by a prime less than $\log \log x$.

Though asymptotically the number of n with $|S(n)| = 2$ is density 0, for quite some time the proportion is more than a quarter of all integers. It would be worth knowing at what point the proportion is less than an arbitrary constant $0 < c < 1$, as well as the proportion of n with $|S(n)| = \ell$ for values of ℓ greater than two.

Our algorithm counts odd $n \leq x$ with two strong liars by tabulating them. If Theorem 4.3 could be improved by finding an explicit error bound, one could find an approximate count much faster through the use of that formula.

Acknowledgment

References

[1] P. T. Bateman and R. A. Horn, *A heuristic asymptotic formula concerning the distribution of prime numbers*, Math. Comp. **16** (1962), 363–367. MR0148632 (26 #6139)

[2] I. Damgård, P. Landrock, and C. Pomerance, *Average case error estimates for the strong probable prime test*, Math. Comp. **61** (1993), no. 203, 177–194, DOI 10.2307/2152945. MR1189518 (94b:11124)

[3] P. Erdös, *Some asymptotic formulas in number theory*, J. Indian Math. Soc. (N.S.) **12** (1948), 75–78. MR0029406 (10,594d)

[4] P. Erdős and C. Pomerance, *On the normal number of prime factors of $\phi(n)$*, Rocky Mountain J. Math. **15** (1985), no. 2, 343–352, DOI 10.1216/RMJ-1985-15-2-343. Number theory (Winnipeg, Man., 1983). MR823246 (87e:11112)

[5] P. Erdős and C. Pomerance, *On the number of false witnesses for a composite number*, Math. Comp. **46** (1986), no. 173, 259–279, DOI 10.2307/2008231. MR815848 (87i:11183)

[6] M. Fürer, *Faster integer multiplication*, SIAM J. Comput. **39** (2009), no. 3, 979–1005, DOI 10.1137/070711761. MR2538847 (2011b:68296)

[7] A. Granville and C. Pomerance, *Two contradictory conjectures concerning Carmichael numbers*, Math. Comp. **71** (2002), no. 238, 883–908, DOI 10.1090/S0025-5718-01-01355-2. MR1885636 (2003d:11148)

[8] G. Greaves, *Sieves in number theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 43, Springer-Verlag, Berlin, 2001. MR1836967 (2002i:11092)

[9] G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 5th ed., The Clarendon Press, Oxford University Press, New York, 1979. MR568909 (81i:10002)

[10] C. Hooley, *Applications of Sieve Methods to the Theory of Numbers*, Cambridge University Press, Cambridge, New York, Melbourne, 1976. Cambridge Tracts in Mathematics, No. 70. MR0404173 (53 #7976)

[11] E. Landau, *Sur quelques problèmes relatifs à la distribution des nombres premiers* (French), Bull. Soc. Math. France **28** (1900), 25–38. MR1504359

[12] E. Landau, *Handbuch der Lehre von der Verteilung der Primzahlen. 2 Bände* (German), Chelsea Publishing Co., New York, 1953. 2d ed; With an appendix by Paul T. Bateman. MR0068565 (16,904d)

[13] L. Monier, *Evaluation and comparison of two efficient probabilistic primality testing algorithms*, Theoret. Comput. Sci. **12** (1980), no. 1, 97–108, DOI 10.1016/0304-3975(80)90007-9. MR582244 (82a:68078)

[14] H. L. Montgomery and R. C. Vaughan, *The large sieve*, Mathematika **20** (1973), 119–134. MR0374060 (51 #10260)

[15] D. Stehlé and P. Zimmermann, *A binary recursive gcd algorithm*, Algorithmic number theory, Lecture Notes in Comput. Sci., vol. 3076, Springer, Berlin, 2004, pp. 411–425, DOI 10.1007/978-3-540-24847-7_31. MR2138011 (2006e:11194)

[16] J. von zur Gathen and J. Gerhard, *Modern Computer Algebra*, 2nd ed., Cambridge University Press, Cambridge, 2003. MR2001757 (2004g:68202)

[17] A. Walfisz, *Zur additiven Zahlentheorie. II* (German), Math. Z. **40** (1936), no. 1, 592–607, DOI 10.1007/BF01218882. MR1545584

[18] E. M. Wright, *A simple proof of a theorem of Landau*, Proc. Edinburgh Math. Soc. (2) **9** (1954), 87–90. MR0065579 (16,448e)

University of Wisconsin–Madison, 1210 W. Dayton St., Madison, Wisconsin 53706
E-mail address: `bach@cs.wisc.edu`

Illinois Wesleyan University, 1312 Park St., Bloomington, Illinois 61701
E-mail address: `ashallue@iwu.edu`