PRIME CHAINS AND PRATT TREES

KEVIN FORD, SERGEI V. KONYAGIN AND FLORIAN LUCA

ABSTRACT. Prime chains are sequences $p_1, \dots, p_k$ of primes for which $p_{j+1} \equiv 1 \pmod{p_j}$ for each j. We introduce three new methods for counting long prime chains. The first is used to show that $N(x; p) = O_\varepsilon(x^{1+\varepsilon})$, where $N(x; p)$ is the number of chains with $p_1 = p$ and $p_k \leqslant px$. The second method is used to show that the number of prime chains ending at p is $\asymp \log p$ for most p. The third method produces the first nontrivial upper bounds on $H(p)$, the length of the longest chain with $p_k = p$, valid for almost all p. As a consequence, we also settle a conjecture of Erdős, Granville, Pomerance and Spiro from 1990. A probabilistic model of $H(p)$, based on the theory of branching random walks, is introduced and analyzed. The model suggests that for most $p \leqslant x$, $H(p)$ stays very close to $\mathrm{e} \log\log x$.

1. INTRODUCTION

1.1. For positive integers a and b, write $a \prec b$ if $b \equiv 1 \pmod{a}$. We are interested in properties of *prime chains* $p_1 \prec p_2 \prec \cdots \prec p_k$, e.g. $3 \prec 7 \prec 29 \prec 59$. Prime chains are multiplicative analogs of the well-studied additive prime k-tuples (sequences $p_1 < \cdots < p_k$ of primes with $p_k - p_1$ small). Important quantities of study are $N(x)$, the number of prime chains with $p_k \leqslant x$ (k variable), $N(x; p)$, the number of prime chains with $p_1 = p$ and $p_k/p_1 \leqslant x$, $f(p)$, the number of prime chains with $p_k = p$, and $H(p)$, the length of the longest prime chain with $p_k = p$. Estimates for these quantities have arisen in investigations of iterates of Euler's totient function $\phi(n)$ and Carmichael's function $\lambda(n)$ (e.g. [5], [6], [7], [19], [28], [29], [30]), the value distribution of $\lambda(n)$ [21], common values of $\phi(n)$ and the sum-of-divisors function $\sigma(n)$ [22], and the complexity of primality certificates ([8], [33]).

In studying long chains, where the ratios $\log p_{j+1}/\log p_j$ are small on average, we require information about the large prime factors of shifted primes $p - 1$. That is, we require good estimates for $\pi(x; q, 1)$ when q is large, where $\pi(x; q, a) = |\{p \leqslant x : p \equiv a \pmod{q}\}|$. Progress, however,

---

Date: September 15, 2010.

2000 *Mathematics Subject Classification.* Primary 11N05, 11N36; Secondary 60J80.

The research of K. F. was supported in part by NSF grants DMS-0555367 and DMS-0901339, that of S. K. was supported in part by Grant 08-01-00208 from the Russian Foundation for Basic Research, and that of F. L. was supported in part by projects PAPIIT 100508 and SEP-CONACyT 79685. Some of this work was accomplished while the third author visited the University of Illinois at Urbana-Champaign in February 2007, supported by NSF grant DMS-0555367. The first and second authors enjoyed the hospitality of the Institute for Advanced Study, where some of this work was done in November 2007 and spring 2010, the first author being supported by The Ellentuck Fund and The Friends of the Institute for Advanced Study. The first author was also supported by the NSF sponsored Workshop in Linear Analysis and Probability at Texas A&M University, August 2008.

is hampered by our poor knowledge when $q > \sqrt{x}$. Let $\mathrm{li}(x) = \int_2^x dt/\log t$. The Bombieri-Vinogradov theorem ([16], Ch. 28) implies that

$$\sum_{m\leqslant Q} \max_{y\leqslant x} \left| \pi(y; m, 1) - \frac{\mathrm{li}(y)}{\phi(m)} \right| \ll R, \tag{1.1}$$

with $Q = x^{1/2}(\log x)^{-B}$ and $R = x(\log x)^{-A}$ (here $A > 0$ is arbitrary and B depends on A). The corresponding statement with $Q = x^\theta$ is not known for any fixed $\theta > 1/2$, however it is conjectured (the Elliott-Halberstam conjecture; abbreviated EH) that (1.1) holds with $Q = x^\theta$ and $R = x(\log x)^{-A}$ for any $\theta < 1$ and $A > 0$. The one-sided Brun-Titchmarsh inequality

$$\pi(x; q, 1) \leqslant \frac{2x}{(q-1)\log(x/q)}, \tag{1.2}$$

however, is useful in some situations.

If one asks just for the existence of many shifted primes $p - 1$ with a large prime factor, we can do a little bit better than Bombieri-Vinogradov. Let $P^+(n)$ denote the largest prime factor of n, and let θ_0 be the supremum of real numbers θ so that there are infinitely many primes $p \leqslant x$ such that $P^+(p-1) \geqslant x^\theta$. EH implies $\theta_0 = 1$, and the best unconditional result is due to Baker and Harman [4], who showed that $\theta_0 \geqslant 0.677$.

In this paper, we prove new bounds for $N(x; p)$, $N(x)$, $f(p)$ and $H(p)$. At the core of our arguments is a kind of duality principle: in a chain $p_1 \prec \cdots \prec p_k$, there are integers m_j with $p_{j+1} = m_j p_j + 1$, and there is an obvious bijection between the k-tuples $(p_1, \ldots, p_k)$ and $(p_1, m_1, \ldots, m_{k-1})$. It is often more efficient to focus on properties of the latter vector rather than the former.

1.2. We begin with the problem of bounding $N(x; p)$. By iterating (1.2), one arrives at a uniform bound (e.g. [19], Theorem 3.5)

$$N_k(x; p) \ll \frac{x(c \log_2 x)^{k-1}}{\log x} \tag{1.3}$$

for the number, $N_k(x; p)$, of prime chains of length k with $p_1 = p$ and $p_k/p_1 \leqslant x$. Here $\log_k x$ is the k-th iterate of the logarithm of x, and c is some constant. Summing (1.3) over $k \leqslant \frac{\log x}{\log 2} + 1$, one obtains the weak estimate $N(x; p) \ll x^{O(\log_3 x)}$.

Theorem 1. *For $p \geqslant 2$ and $x \geqslant 20$, we have the effective estimate*

$$N(x; p) \leqslant x \exp\left\{ \frac{\log x(\log_3 x + O(1))}{\log_2 x} \right\}.$$

In particular, for every $\varepsilon > 0$ there is an effective constant $C(\varepsilon)$ so that $N(x; p) \leqslant C(\varepsilon)x^{1+\varepsilon}$.

Theorem 1 has applications to problems which, at first glance, have nothing to do with prime chains. First, it is a crucial tool in the recent proof by Ford, Luca and Pomerance [22] that the equation $\phi(a) = \sigma(b)$ has infinitely many solutions, settling a well-known 50-year old problem of Erdős. In [21], Theorem 1 is used to show that for some effective q_0, if $\pi(p^{3a}; p^a, 1) - \pi(p^{3a}; p^{a+1}, 1) \geqslant 113 p^{\frac{7a-3}{4}} / \log(p^{a+1})$ for all prime powers $p^a \in (10^{10}, q_0]$, then for every positive integer n, there is another positive integer m with $\lambda(n) = \lambda(m)$. This nearly settles a conjecture from [5], the analog for λ of the famous Carmichael Conjecture for ϕ.

Theorem 1 is nearly best possible, since $N(x;p) \geqslant N_2(x;p) = \pi(px;p,1)$, which is expected to be $\gg x/(\log px)$ unless x is very small relative to p.

Conjecture 1. *We have $N(x;p) \ll x$.*

Conjecture 1 is easy to prove when p is bounded. Using $f(2)=1$ and the recursive formula

$$f(p) = 1 + \sum_{q|(p-1)} f(q), \tag{1.4}$$

we have

$$f(p) \leqslant \frac{2\log p}{\log 2} - 1 \qquad (\text{all } p). \tag{1.5}$$

Summing on $p \leqslant x$ using the prime number theorem gives $N(x) \ll x$ and hence $N(x;p) \ll px$. Lower bounds on $f(p)$ and $N(x)$ are more difficult, since $f(p)$ is sometimes very small, e.g. if $p = 1 + 2^a 3^b$ then $f(p) = 4$ (it is conjectured that there are infinitely many such primes).

Theorem 2. *(i) We have $f(p) \geqslant 0.378 \log p$ for almost all primes p. Hence, $N(x) \gg x$.*

(ii) For all $x \geqslant 3$ and any positive integer h, $|\{p \leqslant x : f(p) = h\}| \leqslant \left(\frac{6\log x}{h}\right)^h$.

In particular, part (ii) implies that primes with $f(p) = o(\log p)$ are exceptionally rare, the counting function being $x^{o(1)}$ as $x \to \infty$. Also, by (i), we have $N(x;2) = \frac{1}{2}N(x) \gg x$.

It is likely that $f(p)$ has normal order[1] $c\log p$ for some c. A very similar problem was considered in Section 2 of [19], namely the behavior of $I(n) = \min\{j : \phi_j(n) = 1\}$, where ϕ_j is the j-th iterate of ϕ. It turns out that $F(n) = I(n) - \left\{\begin{smallmatrix} 1 & n \text{ odd} \\ 0 & n \text{ even}\end{smallmatrix}\right.$ is completely additive, and $F(p) = F(p-1) = \sum_{q^a \| p-1} aF(q)$ for odd primes p. This is similar to (1.4), the only difference being the behavior at proper prime powers, which play an insignificant role in the arguments in [19]. Summing (1.4) over primes $p \leqslant x$ gives

$$N(x) = \pi(x) + \sum_{q \leqslant x/2} f(q)\pi(x;q,1).$$

Inserting this relation into the proof of Theorem 2.1 in [19] (combine Lemma 2.4, Corollary 2.5, (2.8) and Theorem 2.1 therein), we obtain the following.

Theorem A. *If* (1.1) *holds with $Q = x^{1-(\log_2 x)^{-1-\delta}}$ and $R = x(\log x)^{-2}$ for some fixed $\delta > 0$, then $N(x) \sim cx$ for some constant $c > 0$ and $f(p)$ has normal order $c\log p$.*

Conjecture 1 implies that for all $\varepsilon > 0$ and prime $q > (\log x)^{1+\varepsilon}$, for most $p \leqslant x$ there is no prime chain $q \prec \cdots \prec p$. This gives, conditionally, the first part of [19, Conjecture 1]. By contrast, the proof of Theorem 4.5 of [19] implies that if $q \leqslant (\log x)^c$, for some small constant $c > 0$, then for almost all primes $p \leqslant x$, there is a prime chain $q \prec \cdots \prec p$.

[1]for every $\varepsilon > 0$, $|f(p) - c\log p| \leqslant \varepsilon \log p$ for all primes but $o(\pi(x))$ exceptions up to x.

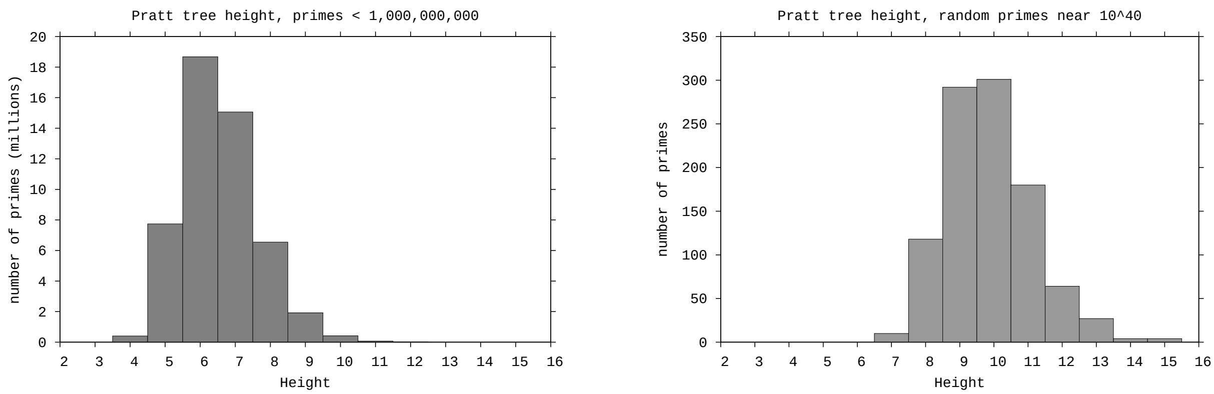

FIGURE 1. Pratt tree height

1.3. The *Pratt tree* $T(p)$ for a prime p is the tree with root node p, below p are nodes labelled with the prime factors q of $p-1$, below each q are nodes labelled with the prime factors of $q-1$, and so on. In 1975, V. Pratt [33] used it in conjunction with Lucas' primality test ([15], §4.1) to show that every prime has a short certificate (proof of primality). Pomerance [32] gave another method for producing primality certificates, but it is an open problem whether the Pratt certificate has longer complexity for most primes (see §1 of [32]). Two important statistics of the Pratt tree are the total number of nodes $f(p)$ and the height $H(p)$, the latter being the length of the longest prime chain ending at p. It is known (see [7], [18], [20]) that the number of primes at a *fixed* level n in the Pratt tree for most p is $\sim (\log_2 p)^n/n!$. The idea is that for most primes p, $p-1$ has a multiplicative structure similar to that of a typical integer of its size; namely, $p-1$ has about $\log_2 p$ prime factors, uniformly distributed on a $\log\log$-scale (see [25], Ch. 1). This, however, does not give much information about $H(p)$.

Figure 1 shows histograms of $H(p)$ for all primes $p \leqslant 10^9$ and for 1000 randomly chosen primes near 10^{40}. Very little is known about the distribution of $H(p)$, the extremal behavior being a case in point. First, $H(p)=2$ if and only if p is a Fermat prime, that is, $p=2^{2^m}+1$ for some m. It seems plausible that $H(p)=3$ for infinitely many p, but this is hopeless to prove at this time. At the other extreme, we have the trivial upper bound $H(p) \leqslant \frac{\log p}{\log 2}+1$. Large values of $H(p)$ may be obtained using the special chain $2=q_1 \prec q_2 \prec \cdots$, where, for each j, q_{j+1} is the smallest prime $\equiv 1 \pmod{q_j}$. By Linnik's theorem, $q_{j+1} \leqslant q_j^L$ for some constant L, hence $H(q_j) \geqslant \frac{\log\log q_j}{\log L}$. It is conjectured that $q_{j+1} \leqslant q_j(\log q_j)^C$ for some fixed C and this implies a far stronger bound $H(q_j) \gg \frac{\log q_j}{\log\log q_j}$. Even showing $H(p)/\log_2 p \to \infty$ for an infinite sequence of p is extremely hard, as it implies $\theta_0=1$: a prime chain $p_1 \prec \cdots \prec p_k = p$ with $k=H(p)$ satisfies

$$\frac{\log p_k}{\log p_1} = \prod_{j=1}^{k-1} \frac{\log p_{j+1}}{\log p_j} \geqslant \left(\min_{1\leqslant j\leqslant k-1} \frac{\log p_{j+1}}{\log p_j} \right)^{k-1},$$

and hence

$$\Lambda := \limsup_{p\to\infty} \frac{H(p)}{\log_2 p} \leqslant \frac{1}{-\log\theta_0}. \tag{1.6}$$

On the other hand, Kátai [27] proved that for some constant $c > 0$, $H(p) \geqslant c\log_2 p$ for almost all primes p (for all primes $p \leqslant x$ with $o(x/\log x)$ exceptions). We prove a version with the constant made explicit in terms of the level of distribution of primes in progressions.

Theorem 3. *(a) If* (1.1) *holds with* $Q = x^\theta$ *and* $R = o(x/\log x)$*, then for any* $c < \frac{1}{\mathrm{e}^{-1} - \log\theta}$*,* $H(p) > c\log_2 p$ *for almost all primes* p*;*

(b) If (1.1) *holds with* $Q = x^\theta$ *and* $R = x(\log x)^{-A}$ *for every* $A > 1$*, then for every* $c < \frac{1}{-\log\theta}$*, there is a* K *so that* $H(p) > c\log_2 p$ *for* $\gg x/(\log x)^K$ *primes* $p \leqslant x$*. Consequently,* $\Lambda \geqslant \frac{1}{-\log\theta}$*.*

Corollary 1. *EH implies that for every* $c < \mathrm{e}$*,* $H(p) > c\log_2 p$ *for almost all* p*.*

Remark 1. By the Bombieri-Vinogradov theorem and (1.2), for every $\varepsilon > 0$, (1.1) holds with $Q = x^{1-\varepsilon}$ and $R = O_\varepsilon(x/\log x)$.

If $\Lambda \geqslant -1/\log\theta$ for some $\theta < 1$, then there are many chains $2 = p_1 \prec \cdots \prec p_k$ with $\frac{\log p_{j+1}}{\log p_j}$ at most $1/\theta$ on average. Thus, although (1.6) is likely an equality, to prove this would require strong information about the set of primes with $P^+(p-1)$ near p^{θ_0}. The same difficulty arises when trying to prove (a) with $\theta = \theta_0$.

The bound on $H(p)$ in (a) is weaker than in (b), since it is unusual in a chain $p_1 \prec \cdots \prec p_k$ for most of the ratios $\frac{\log p_{j+1}}{\log p_j}$ to be close to $1/\theta$. The constant e^{-1} appearing in (a) is likely best possible; see Conjecture 2 below.

Turning to upper bounds, before our work it was unknown if there is an infinite sequence of primes with $H(p) = o(\log p)$. A natural approach is to find p such that $P^+(p-1)$ is small and use $H(p) = 1 + \max_{q|(p-1)} H(q) \ll \max_{q|(p-1)} \log q$. However, our knowledge of smooth shifted primes is very weak (the world record is $P^+(p-1) < p^{0.2961}$ infinitely often [4]). Using a new and very different approach, we give a much stronger upper bound.

Theorem 4. *We have* $H(p) \leqslant (\log p)^{0.9503}$ *for almost all* p*.*

The proof of Theorem 4 involves showing that for most primes p, all the primes at some bounded level of the tree are small. In particular, this settles [19, Conjecture 2].

Theorem 5. *For every* $\varepsilon > 0$ *and* $\delta > 0$*, there is an integer* k *so that for large* x *and at least* $(1-\delta)x$ *integers* $n \leqslant x$*,* $P^+(\phi_k(n)) \leqslant x^\varepsilon$*.*

There is a folklore conjecture that $H(p) = O(\log_2 p)$ for most p.

Conjecture 2. $H(p)$ *has normal order* $\mathrm{e}\log_2 p$*.*

Remark 2. The lower bound in Conjecture 2 follows from EH (Corollary 1). Conversely, by (1.6), the lower bound in Conjecture 2 implies that $\theta_0 \geqslant \mathrm{e}^{-1/\mathrm{e}} = 0.6922\ldots$ (with a bit more work using (1.2), one can deduce $\theta_0 \geqslant 0.73$).

The upper bound in Conjecture 2 appears to be even more difficult. We cannot see a way to deduce it from standard conjectures in prime number theory, e.g. EH plus a uniform prime k-tuples conjecture, although Theorem 4 can be significantly improved under such hypotheses.

Understanding $H(p)$ requires detailed knowledge of the distribution of the large prime factors of shifted primes $p-1$. Making a reasonable assumption for this distribution (a consequence of EH), in Section 6 we model the Pratt tree by a branching random walk. The model provides a much more precise version of Conjecture 2.

Conjecture 3. *$H(p) = \mathrm{e}\log_2 p - \frac{3}{2}\log_3 p + E(p)$, where for some fixed $c, c' > 0$ and any $z \geqslant 0$, the number of $p \leqslant x$ for which $E(p) \geqslant z$ is $\gg \mathrm{e}^{-c'z}\pi(x)$ and $\ll \mathrm{e}^{-cz}\pi(x)$, and $E(p) \leqslant -z$ for $O(\exp\{-\mathrm{e}^{cz}\}\pi(x))$ primes $\leqslant x$.*

Notable features of Conjecture 3 are (i) the *tightness* of $E(p)$: the distribution of $H(p)$ over $p \leqslant x$ does not widen as $x \to \infty$, and (ii) the pronounced asymmetry of the distribution of $E(p)$. The analogs of these features for our probabilistic model are proved rigorously.

Assuming Median$\{H(p) : p \leqslant x\}$ grows slowly, we show that $H(p)$ is tight to the left of its median.

Theorem 6. *Suppose g and h are increasing, $0 \leqslant g(x) \leqslant h(x)$, $h(x^2) - h(x) \leqslant K$ and $g(x^2) - g(x) \leqslant K$ for $x \geqslant 1$. Suppose, for large x, that $H(p) \geqslant h(p)$ for at least $c\pi(x)$ primes $\leqslant x$. Then $H(p) \geqslant h(p) - g(p)$ for all primes $p \leqslant x$ with at most $O(\pi(x)\exp\{-\frac{c\log 2}{K}g(x)\})$ exceptions.*

We conclude this section with a conjecture about prime chains, which follows from the prime k-tuples conjecture (with $m = 2$ below) but should be "easier". It is a multiplicative analog of the statement that the primes contain arbitrarily long arithmetic progressions, recently proved by Green and Tao [23]. Even the case $k = 3$ is not known.

Conjecture 4. *For each $k \geqslant 3$, there are infinitely many prime k-tuples $(p_1, \dots, p_k)$ where, for some m, $p_{j+1} = mp_j + 1$ for $1 \leqslant j \leqslant k-1$.*

Notation. The letters p and q, with or without subscripts, always denote primes. Constants implied by the O, $\ll$ and $\asymp$ symbols do not depend on any parameter unless indicated. In Section 6, we use $\mathbf{P}$ for probability and $\mathbf{E}$ for probabilistic expectation.

2. SIFTED CHAINS: PROOF OF THEOREM 1

The underlying idea is a sieve; relax the condition that the numbers in the chain are prime, and only require that they do not have small prime factors. Let $y \geqslant 2$ and let r be the product of the primes $\leqslant y$. For $(a, r) = 1$, let $G_a(x; y)$ be the number of chains $n_1 \prec \cdots \prec n_k$ with $n_1 = a$, with $n_k/n_1 \leqslant x$ and consisting of numbers coprime to r. If $p > y$, then $N(x; p) \leqslant G_p(x, y)$. There are integers ("links") $m_1, \dots, m_{k-1}$ with $n_{j+1} = m_j n_j + 1$ for $1 \leqslant j \leqslant k-1$. For positive integers a, b and real $s > 1$,

$$S(a, b) = S(a, b; r, s) = \sum_{\substack{m \geqslant 1 \\ am+1 \equiv b \pmod r}} m^{-s}$$

encodes the possible links m from a number $n_i \equiv a \pmod r$ to a number $n_{i+1} \equiv b \pmod r$.

Fix r, s and let $U_r = (\mathbb{Z}/r\mathbb{Z})^*$. Let $A_k(a_1, a_k)$ be the sum of $(m_1 \cdots m_{k-1})^{-s}$ over all tuples $(m_1, \dots, m_{k-1})$ which could serve as links in a chain starting from a number $n_1 \equiv a_1 \pmod r$, ending with a number $n_k \equiv a_k \pmod r$ and with all numbers in the chain coprime to r. Then $A_2(a_1, a_2) = S(a_1, a_2)$ and for $k \geqslant 3$,

$$A_k(a_1, a_k) = \sum_{a_2, \dots, a_{k-1} \in U_r} S(a_1, a_2) S(a_2, a_3) \cdots S(a_{k-1}, a_k).$$

Let $V_k(a_1)$ be the column vector $(A_k(a_1, a_k) : a_k \in U_r)$. For consistency, let $V_1(a_1)$ be a vector with all zero entries except for an entry of 1 in the a_1 position. Since

$$A_{k+1}(a_1, a_{k+1}) = \sum_{a_k \in U_r} A_k(a_1, a_k) S(a_k, a_{k+1}),$$

we obtain $V_{k+1}(a_1) = MV_k(a_1)$, where $M = M(r,s) = \big(S(a,b)\big)_{b,a\in U_r}$. The rows of M are indexed by b and the columns are indexed by a. Finally, let $F_k(a_1) = \sum_{a_k} A_k(a_1, a_k)$, so that

$$F_k(a_1) = (1, \dots, 1)V_k(a_1) = (1, \dots, 1)M^{k-1}V_1(a_1);$$

i.e., $F_k(a_1)$ is the sum of the entries of column a_1 in M^{k-1}. Since $m_1 \cdots m_{k-1} \leqslant n_k/n_1 \leqslant x$,

$$G_a(x; y) \leqslant \inf_{s>1} \Big(x^s \sum_{1 \leqslant k \leqslant \frac{\log x}{\log 2}+1} F_k(a) \Big). \tag{2.1}$$

Observe that the sum on k in (2.1), if extended to $k = \infty$, is convergent if and only if M is a contracting matrix, i.e., all the eigenvalues of M have modulus < 1. Since M has positive real entries, the Perron-Frobenius Theorem implies that the eigenvalue with largest modulus is positive, real and simple. Call this eigenvalue $\lambda(s; y)$.

We show below that if y is large and $s \geqslant 1 + \frac{\log_2 y}{\log y}$, then $\lambda(s; y) < 1$. Accurate estimation of $\lambda(s; y)$ is difficult for large y, but the largest row sum of M serves as an upper bound. For a generic matrix A, let $R_b(A)$ be the sum of the entries in the row indexed by b, and let $R(A)$ be the maximum row sum of A. For row b of M, write $d = (b-1, r)$ and $b' = \frac{b-1}{d}$. Then

$$R_b(M) = \sum_{a\in U_r} \sum_{am\equiv b-1 \,(\mathrm{mod}\, r)} m^{-s} = d^{-s} \sum_{(k,r/d)=1} k^{-s} \#\{a \in U_r : ak \equiv b' \,(\mathrm{mod}\, r/d)\}.$$

The congruence $ak \equiv b' \pmod{r/d}$ has a unique solution modulo r/d, and hence has $\phi(d)$ solutions $a \in U_r$. Thus,

$$R_b(M) = \frac{\phi(d)}{d^s} \sum_{(k,r/d)=1} k^{-s} = \frac{\phi(d)}{d^s} \prod_{p\nmid(r/d)} (1 - p^{-s})^{-1} = \prod_{p>y} (1 - p^{-s})^{-1} \prod_{p|d} \frac{p-1}{p^s - 1}.$$

Therefore, since d is always even,

$$R(M) = \frac{1}{2^s - 1} \prod_{p>y} (1 - p^{-s})^{-1}. \tag{2.2}$$

Since $R(AB) \leqslant R(A)R(B)$, $R(M^{k-1}) \leqslant R(M)^{k-1}$. To bound $G_a(x; y)$, we need to bound the largest *column* sum of M^{k-1}. Lacking a better approach, we use the crude bound $\phi(r)R(M)^{k-1}$. Thus, $F_k(a) \leqslant \phi(q)R(M^{k-1}) \leqslant \phi(r)R(M)^{k-1}$. By (2.1),

$$G_a(x; y) \leqslant \phi(r) \inf_{s:R(M)<1} \frac{x^s}{1 - R(M)}.$$

By standard prime number estimates, if $1 < s \leqslant 2$, then

$$-\sum_{p>y} \log(1 - p^{-s}) = O(1/y^{2s-1}) + \sum_{p>y} p^{-s} \ll \frac{\mathrm{e}^{-(s-1)\log y}}{(s-1)\log y}.$$

Take $y = \frac{\log x}{\log_2 x}$ and $s = 1 + \frac{\log_2 y}{\log y}$. Since $2^s - 1 = 1 + (2\log 2)(s-1) + O((s-1)^2)$, (2.2) implies $1 - R(M) \sim (2\log 2)(s-1)$ as $x \to \infty$. Since $\phi(r) \leqslant r = e^{(1+o(1))y}$ as $x \to \infty$, this proves Theorem 1.

3. Proof of Theorem 2

Let $Q(p)$ be the multiset of prime labels appearing in $T(p)$, and let $T'(p)$ be the subtree of $T(p)$ consisting of nodes with odd prime labels. There is a natural bijection between $T(p)$ and $T'(p)$, obtained by adding to every node in $T'(p)$ a child node with label 2. Let $l(n) = \prod_{p^a \| n} p^{a-1}$. The quantity $\frac{q}{q-1} l(q-1)$ measures the "loss of mass" when descending from a node labelled q to its child nodes: in fact, it is easy to see that

$$\prod_{q \in Q(p)} \left(\frac{q}{q-1} l(q-1) \right) = p.$$

If $p > 2$, exactly half of the nodes in $T(p)$ are labelled with 2, thus

$$\prod_{q \in Q(p)} l(q-1) \leqslant p 2^{-\frac{1}{2} f(p)}. \tag{3.1}$$

Consider the set of $p \leqslant x$ with $f(p) = h$, where $h = 2n$ is positive and even. Let $\mathcal{T}$ be the set of rooted trees on n nodes. For each $T' \in \mathcal{T}$, we consider separately the primes p with $T'(p)$ tree-isomorphic to T'. Form the tree T on h nodes, by adding to each node of T' an additional child node. We count in how many ways we can label with primes the nodes of T, with the leafs having label 2 and the root having label $p \leqslant x$. For a given prime p, there may be more than one way to assign primes in the Pratt tree to the nodes of T'; this occurs when some node has two or more child trees that are isomorphic (as rooted trees). Thus, for each T', we count ways to assign primes to the nodes, and divide by the number of ways in which we can permute the nodes; that is, the number $I(T')$ of isomorphisms of T'. Assign to each node an ordinal number $1, 2, \ldots, h$ so that the children of node j are assigned lower ordinals (e.g., node 1 will be a lowest leaf, and node h will be the root). To node number j, we let q_j be its prime label and $l_j = l(q_j - 1)$. Let B_j be the set of ordinal numbers of the children of node j, and observe that $B_1, \ldots, B_h$ depend only on T'. With this notation,

$$q_j - 1 = l_j \prod_{k \in B_j} q_k. \tag{3.2}$$

With T fixed, (3.2) implies a natural bijection between $(q_1, \ldots, q_h)$ and $(l_1, \ldots, l_h)$ (recall that leafs have $q_j = l_j = 2$). By (3.1), we have for any $\beta > 0$

$$|\{p \leqslant x : f(p) = h\}| \leqslant \sum_{T' \in \mathcal{T}} \frac{1}{I(T')} \sum_{l_1, \ldots, l_h} \left(\frac{x 2^{-h/2}}{l_1 \cdots l_h} \right)^{\beta}.$$

Suppose that $j \geqslant 2$ and that $l_1, \ldots, l_{j-1}$ have been chosen. If node j is a leaf of T, then $l_j = 1$. Otherwise, using (3.2), the primes q_k for $k \in B_j$ are determined by $l_1, \ldots, l_{j-1}$. Moreover, a prime

$r|l_j$ must equal one of these primes q_k. Hence

$$\sum_{l_j} l_j^{-\beta} \leqslant \prod_{k\in B_j}\left(1-q_k^{-\beta}\right)^{-1} \leqslant \left(1-2^{-\beta}\right)^{-1}\left(1-3^{-\beta}\right)^{1-|B_j|}.$$

By Borchardt's formula[2] [13] for counting labelled trees,

$$\sum_{T'\in\mathcal{T}} \frac{1}{I(T')} = \frac{n^{n-2}}{(n-1)!} = \frac{n^{n-1}}{n!} \leqslant \mathrm{e}^n.$$

Since $\sum_{|B_j|\geqslant 1}(|B_j|-1) = h/2-1$, we conclude that

$$|\{p\leqslant x : f(p)=h\}| \leqslant \mathrm{e}^{h/2}\left(x2^{-h/2}\right)^{\beta}\left(1-2^{-\beta}\right)^{-h/2}\left(1-3^{-\beta}\right)^{-h/2} \tag{3.3}$$

Taking $\beta = 0.37$, the right side of (3.3) is $\leqslant x^{0.37}(27.8371)^{h/2} \leqslant x^{0.999}$ for $h \leqslant 0.378\log x$. For the second part of Theorem 2, assume $h \leqslant (2/5)\log x$ (if $h \geqslant 3\log x$, there are no such p and for $(2/5)\log x < h \leqslant 3\log x$, $(\frac{6\log x}{h})^h > x$). Take $\beta = h/\log x$. Since $0<\beta\leqslant 2/5$, $2^{\beta}-1 \geqslant \beta\log 2$ and $1-3^{-\beta} \geqslant 0.889\beta$. Hence, the right side of (3.3) is $\leqslant x^{\beta}(4.412/\beta^2)^{h/2}$.

We remark that numerical improvements are possible by refining the above analysis; e.g. using the fact that leafs of T' must be labelled with a Fermat prime.

4. Lower bounds for $H(p)$: proof of Theorem 3

We show part (b) first, as the proof is much easier.

Proof of Theorem 3 (b). Let c' and θ' satisfy $\theta' > 1/3$ and $c < c' < \frac{1}{-\log\theta'} < \frac{1}{-\log\theta}$, and define K by $8\theta^K = \theta-\theta'$. Let x_0 be large, depending on $K, c, c', \theta, \theta'$ and put $c'' = c'\log_2(x_0^3)$. Let $\mathcal{P} = \{p : H(p) \geqslant c'\log_2 p - c''\}$. In particular, $\mathcal{P}$ contains all primes $\leqslant x_0^3$. We shall prove

$$Q(x) := |\mathcal{P}\cap(x/2,x]| \geqslant \frac{x}{(\log x)^K} \tag{4.1}$$

for $x \geqslant x_0$, which implies the desired conclusion (since $c' > c$). By the prime number theorem and the fact that $K>1$, if x_0 is large enough then (4.1) holds for $x_0 \leqslant x \leqslant x_0^3$. Suppose $y \geqslant x_0^3$ and (4.1) holds for $x_0 \leqslant x \leqslant y$. Assume $y < x \leqslant 2y$ and put $I = \mathcal{P}\cap(x^{\theta'}, x^{\theta}]$. Suppose that $x/2 < p \leqslant x$, and that $q|p-1$, where $q\in I$. Then

$$\begin{aligned} H(p) &\geqslant 1 + H(q) \geqslant 1 + c'\log_2 q - c'' \\ &\geqslant 1 + c'\log_2 x + c'\log\theta' - c'' > c'\log_2 p - c'', \end{aligned}$$

so that $p\in\mathcal{P}$. For $x/2 < p \leqslant x$, $p-1$ is divisible by at most two primes from I, hence

$$Q(x) \geqslant \frac12\sum_{q\in I}\Big(\pi(x;q,1)-\pi(x/2;q,1)\Big) \geqslant \frac{x}{4\log x}\sum_{q\in I}\frac1q + O\left(\frac{x}{(\log x)^{K+1}}\right).$$

Since (4.1) holds for $x^{\theta'} < y \leqslant x^{\theta}$, the sum on $q\in I$ is

$$\geqslant \sum_{2^j\leqslant x^{\theta-\theta'}} \frac{Q(2^j x^{\theta'})}{2^j x^{\theta'}} \geqslant \left(\frac{(\theta-\theta')\log x}{\log 2} - 1\right)\frac{1}{(\log x^{\theta})^K} \geqslant \frac{\theta-\theta'}{\theta^K(\log x)^{K-1}} = \frac{8}{(\log x)^{K-1}}.$$

[2] commonly known as Cayley's formula.

Therefore, (4.1) holds. By induction on dyadic intervals, (4.1) holds for all $x \geqslant x_0$. $\square$

Remark. The same proof gives, assuming that (1.1) holds with $Q = x^{1-\varepsilon(x)}$ and $R(x) = x(\log x)^{-g(x)}$ where $\varepsilon(x) \to 0$ and $g(x) \to \infty$ as $x \to \infty$, that $H(p) \geqslant h(p)\log_2 p$ for infinitely many p, where $h(p) \to \infty$ as $p \to \infty$ (the function h depending on the functions ε, g).

Proof of Theorem 3 (a). We proceed by induction as in part (b), but instead we iterate by many levels in the chain at once rather than one level at a time. Suppose that $c < h < c' < 1/(\mathrm{e}^{-1} - \log(\theta))$. For some constant c'', described below, let $\mathcal{P} = \{p : H(p) \geqslant c' \log_2 p - c''\}$. We will show, for some $\delta > 0$, that

$$\mathcal{P}(x) := |\{p \leqslant x : p \in \mathcal{P}\}| \geqslant \delta \frac{x}{\log x}. \tag{4.2}$$

Consequently, a positive proportion of primes p satisfy $H(p) > h \log_2 p$, and Theorem 3 (a) follows from Theorem 6.

By Stirling's formula, there is an integer $k \geqslant 2$ such that $1/c' > \frac{1}{k}(k!)^{1/k} - \log(\theta)$. Let α, β satisfy $\mathrm{e}^{-k/c'} < \beta < \theta^k \exp\left(-(k!)^{1/k}\right)$ and $\beta \exp\left((k!)^{1/k}\right) < \alpha < \theta^k$. Suppose that δ is sufficiently small, depending only on the choice of $c', \theta, k, \alpha, \beta$. Let x_0 be sufficiently large, depending on $c', \theta, k, \alpha, \beta, \delta$, and put $c'' = c' \log_2(x_0)$. Observe that (4.2) holds trivially for $2 \leqslant x \leqslant x_0$, provided δ is small enough. Throughout this proof, constants implied by the $O-$ and $\ll -$symbols may depend on $c', \theta, k, \alpha, \beta$, but not on δ.

Next, suppose that $Y \geqslant x_0$ and that inequality (4.2) holds for $2 \leqslant x \leqslant Y$. Let S be a subset of the primes in $\mathcal{P} \cap [Y^\beta, Y^{\theta^k}]$. Let $M(S)$ be the number of primes $p_0 \in (Y, 2Y]$ so that there is a prime chain $p_k \prec p_{k-1} \prec \cdots \prec p_0$ with $p_k \in S$. For such p_0, we have

$$\begin{aligned} H(p_0) &\geqslant k + H(p_k) \geqslant k + c' \log_2 p_k - c'' \\ &= c' \log_2(2Y) - c'' + c' \log \beta + k + O\left(\frac{1}{\log Y}\right) \geqslant c' \log_2 p_0 - c'' \end{aligned}$$

if x_0 is large enough. We will show, for appropriate S, that

$$M(S) \geqslant \delta \frac{Y}{\log Y}, \tag{4.3}$$

which implies $P(2Y) \geqslant P(Y) + M(S) \geqslant 2\delta Y/\log(2Y)$. By induction over dyadic intervals, (4.3) implies (4.2), and hence Theorem 3 (a).

To prove (4.3), we will consider chains satisfying not only $p_k \in S$, but also

$$p_{j+1} \leqslant p_j^\theta \quad (0 \leqslant j \leqslant k-1), \quad p_1 \leqslant Y^\theta. \tag{4.4}$$

With (4.4), we can use (1.1) to accurately count such chains. We have $M(S) \geqslant M_1(S) - M_2(S)$, where $M_1(S)$ is the number of chains satisfying $p_k \in S$ and (4.4), and $M_2(S)$ is the number of pairs of distinct chains satisfying these conditions with the same p_0. We begin with

$$M_1(S) = \sum_{p_k \in S} \sum_{p_{k-1}} \cdots \sum_{p_1} \Big(\pi(2Y, p_1, 1) - \pi(Y, p_1, 1) \Big),$$

where $p_k \prec \cdots \prec p_0$ and (4.4) in the summations. By (1.1) and induction on $1 \leqslant j \leqslant k$,

$$M_1(S) = \frac{Y}{\log Y} \sum_{p_k} \sum_{p_{k-1}} \cdots \sum_{p_j} \frac{\big(\log_2 Y^{\theta^j} - \log_2 p_j\big)^{j-1}}{p_j (j-1)!} + o\left(\frac{Y}{\log Y}\right). \tag{4.5}$$

Here, we used that for each p_j, there are $O(1)$ chains $p_k \prec \cdots \prec p_j$ with $p_k \geqslant Y^\beta$. By (4.5) with $j = k$,

$$\begin{aligned} M_1(\mathcal{P} \cap [Y^\beta, Y^\alpha]) &\geqslant \frac{\delta Y}{\log Y} \int_{Y^\beta}^{Y^\alpha} \frac{(\log_2 Y^\alpha - \log_2 t)^{k-1}}{(k-1)! t \log t}\, dt + o\left(\frac{Y}{\log Y}\right) \\ &= \frac{\delta Y}{\log Y} \frac{(\log(\alpha/\beta))^k}{k!} + o\left(\frac{Y}{\log Y}\right). \end{aligned}$$

By hypothesis, $\log(\alpha/\beta) > (k!)^{1/k}$. The summands in (4.5) (with $j = k$) are $\asymp 1/p_k$ for $Y^\beta \leqslant p_k \leqslant Y^\alpha$. Hence, if δ is small enough, there is a set $S \subseteq \mathcal{P} \cap [Y^\beta, Y^\alpha]$ such that

$$\sum_{p_k \in S} \frac{1}{p_k} \ll \delta, \qquad M_1(S) \geqslant \big(\delta + \delta^{3/2}\big) \frac{Y}{\log Y}. \tag{4.6}$$

We have $M_2 = M_{2,0} + \cdots + M_{2,k-1}$ where $M_{2,j}$ counts pairs of coupled chains

$$\left.\begin{matrix} p_k \prec \cdots \prec p_{j+1} \\ \\ p'_k \prec \cdots \prec p'_{j+1} \end{matrix}\right\rangle p_j \prec \cdots \prec p_0$$

with each of the two chains satisfying (4.4), $p_{j+1} \neq p'_{j+1}$ and $p_k, p'_k \in S$. We further write $M_{2,j} = M'_{2,j} + M''_{2,j}$, where $M'_{2,j}$ counts pairs of such chains with $p_j \leqslant p_{j+1} p'_{j+1} Y^{\delta^2}$. As before, for each pair (p_{j+1}, p'_{j+1}), there are $O(1)$ choices for $p_k, p'_k, \ldots, p_{j+2}, p'_{j+2}$. For $M'_{2,0}$, $p_1 p'_1 \geqslant Y^{1-\delta^2}$, and for each p_0, there are $O(1)$ choices for p_1, p'_1. By sieve methods (e.g. Theorem 2.2 of [24]),

$$\begin{aligned} M'_{2,0} &\ll \sum_{1 \leqslant k \leqslant Y^{\delta^2}} \left|\{n \leqslant Y : n \equiv 1 \pmod{k}, P^-(n(\tfrac{n-1}{k})) > Y^\beta\}\right| \\ &\ll \sum_{1 \leqslant k \leqslant Y^{\delta^2}} \frac{Y}{\phi(k) \log^2 Y} \ll \frac{\delta^2 Y}{\log Y}. \end{aligned}$$

Here, $P^-(m)$ is the smallest prime factor of m. For $j \geqslant 1$, an argument similar to that leading to (4.5), followed by the same sieve bound, gives

$$M'_{2,j} \ll \frac{Y}{\log Y} \sum_{p_{j+1}, p'_{j+1}} \sum_{p_j} \frac{1}{p_j} \ll \sum_{k \leqslant Y^{\delta^2}} \sum_{\substack{n \leqslant Y, n \equiv 1 \pmod{k} \\ P^-(n(\frac{n-1}{k})) > Y^\beta}} \frac{1}{n} \ll \delta^2 \frac{Y}{\log Y}.$$

For chains counted by $M''_{2,j}$, the Brun-Titchmarsh inequality suffices for the estimations. When $j \geqslant 1$ and p_j is given, as before we have

$$\sum_{p_{j-1}} \cdots \sum_{p_1} \pi(2Y, p_1, 1) \ll \frac{Y}{p_j \log Y}.$$

By partial summation and (1.2), given p_{j+1} and p'_{j+1},

$$\sum_{p_j} \frac{1}{p_j} \ll \frac{1}{p_{j+1}p'_{j+1}\delta^2 \log Y} + \frac{\log(1/\delta)}{p_{j+1}p'_{j+1}} \ll \frac{\log(1/\delta)}{p_{j+1}p'_{j+1}}.$$

For $j+1 \leqslant r \leqslant k-1$,

$$\sum_{p_r} \frac{1}{p_r} \ll \frac{1}{p_{r+1}}, \qquad \sum_{p'_r} \frac{1}{p'_r} \ll \frac{1}{p'_{r+1}}. \tag{4.7}$$

Finally, by (4.6), we arrive at

$$M''_{2,j} \ll \delta^2 \log(1/\delta) \frac{Y}{\log Y}. \tag{4.8}$$

In a similar way, when $j = 0$, we have by (1.2) and partial summation,

$$\sum_{p_1 p'_1 \leqslant 2Y^{1-\delta^2}} \pi(2Y, p_1 p'_1, 1) \ll \frac{\log(1/\delta)}{p_2 p'_2}.$$

A second application of (4.7) then gives (4.8) in this case.

Finally, combining our estimates for $M'_{2,j}$ and $M''_{2,j}$, we obtain $M_2(S) \ll \delta^2 \log(1/\delta) Y/\log Y$. Together with (4.6), if δ is small enough then (4.3) holds, and this completes the proof. □

5. Proof of Theorems 4 and 5

The proofs of Theorems 4 and 5 rely on the fact that the largest prime factor of $p-1$ cannot be too large too often. At the core is a sieve upper bound for k-tuples of primes which is uniform in k, and careful averages of the associated singular series. There is a potentially troublesome factor $2^k k!$ in the sieve estimate, which is partly overcome by observing that if $H(p)$ is large, then there must be a prime chain in the Pratt tree for p which is very condensed in a multiplicative sense.

Lemma 5.1. *There is a positive constant δ so that the following holds. Let $a_1, \dots, a_k$ be positive integers, let $b_1, \dots, b_k$ be integers with $(a_j, b_j) = 1$ for all j, and let $\xi(p)$ be the number of solutions of $\prod_{i=1}^k (a_i n + b_i) \equiv 0 \pmod p$. If $x \geqslant 10$, $1 \leqslant k \leqslant \delta \frac{\log x}{\log_2 x}$ and*

$$B := \sum_p \frac{k - \xi(p)}{p} \log p \leqslant \delta \log x,$$

then the number of integers $n \leqslant x$ for which $a_1 n + b_1, \dots, a_k n + b_k$ are all prime and $> k$ is

$$\ll \frac{2^k k!}{(\log x)^k} x \mathfrak{S} \cdot \exp\left(O\left(\frac{kB + k^2 \log_2 x}{\log x} \right) \right), \qquad \mathfrak{S} = \prod_p \left(1 - \frac{\xi(p)}{p}\right)\left(1 - \frac{1}{p}\right)^{-k}.$$

Proof. Since $\xi(p) = k$ for large p, $\mathfrak{S} > 0$ if and only if $\xi(p) < p$ for all p. Also, $\xi(p) \leqslant k$ for all p. Hence, if $\mathfrak{S} = 0$, the number of n is zero. If $\mathfrak{S} > 0$, Montgomery's large sieve estimate [12, Théorème 6] implies that the number of n in question is $\ll x/G(\sqrt{x})$, where

$$G(z) = \sum_{n \leqslant z} g(n), \qquad g(n) = \mu^2(n) \prod_{p \mid n} \frac{\xi(p)}{p - \xi(p)},$$

and μ is the Möbius function. For fixed k, the argument in [24, §5.3] gives $G(z) \sim (\log z)^k/(k!\mathfrak{S})$. We sketch how to make explicit the dependence on k. By the argument in [24, p. 147–148],

$$\sum_{d\leqslant z} g(d)\log d = \sum_{d\leqslant z} g(d) \sum_{p\leqslant z/d} \frac{\xi(p)\log p}{p} + \sum_{h\leqslant z} g(h) \sum_{\substack{p\mid h\\ p>z/h}} \frac{\xi(p)\log p}{p}$$
$$= k\sum_{d\leqslant z} g(d)\log\frac{z}{d} + O\left(G(z)(B+k\log_2 z)\right).$$

Adding the sum on the right side to both sides yields

$$G(z)\log z = (k+1)\int_1^z \frac{G(t)}{t}\,dt + r(z)G(z)\log z,$$

where $r(z) \ll \frac{B+k\log_2 z}{\log z}$. If δ is small enough and $z \geqslant \sqrt{x}$, then $|r(z)| \leqslant \frac12$. By the argument in [24, p. 150], for some constant D and for $z \geqslant \sqrt{x}$,

$$(1-r(z))\frac{G(z)}{\log^k z} = D\exp\left\{O\left(\frac{kB+k^2\log_2 z}{\log z}\right)\right\}.$$

By the argument on [24, p. 151–152], $D^{-1} = k!\mathfrak{S}$. Taking $z = \sqrt{x}$ completes the proof. $\square$

For given integers $m_1,\dots,m_{k-1} \geqslant 2$, we will apply Lemma 5.1 with the forms $f_1(n) = n$, $f_{j+1}(n) = m_j f_j(n) + 1$ $(1 \leqslant j \leqslant k-1)$. We have $f_j(n) = a_j n + b_j$, where

$$a_j = m_1\cdots m_{j-1} \quad (j\geqslant 1), \qquad b_1 = 0, \quad b_j = 1 + \sum_{i=2}^{j-1} m_i\cdots m_{j-1} \ \ (j\geqslant 2). \tag{5.1}$$

Clearly, $(a_j, b_j) = 1$. Let $\mathfrak{S}(\mathbf{m}) = \mathfrak{S}$ be the associated singular series and let $\xi(p,\mathbf{m}) = \xi(p)$.

Lemma 5.2. *There is a positive constant c_1 so that $\mathfrak{S}(\mathbf{m}) \ll (c_1\log_2(4m_1\cdots m_{k-1}))^{k-1}$. Also,*

$$\sum_p \frac{k-\xi(p,\mathbf{m})}{p}\log p \leqslant k\left(\log_2(4m_1\cdots m_{k-1}) + O(1)\right).$$

Proof. We have $\xi(p,\mathbf{m}) = k$ if $p\nmid N$, where $N = m_1\cdots m_{k-1}\prod_{i<j}|a_ib_j - a_jb_i|$. Also, $\xi(p,\mathbf{m}) \geqslant 1$ for all p. Let $x = m_1\cdots m_{k-1} \geqslant 2^{k-1}$. By (5.1), $a_j \leqslant x$ and $b_j \leqslant 1 + \sum_{j=1}^{k-2} x/2^j \leqslant x$ for each j. Thus, $N \leqslant x^{k(k-1)+1} \leqslant \exp\{O(\log^3 x)\}$. Since $1 - k/p \leqslant (1-1/p)^k$ for $p > k$, if $\mathfrak{S}(\mathbf{m}) > 0$ then $\mathfrak{S}(\mathbf{m}) \leqslant \prod_{p\mid N}(1-1/p)^{1-k} \leqslant (c_1\log_2 N)^{k-1}$ for a constant c_1. The second bound follows from $\sum_{p\mid N}(\log p)/p \leqslant \log_2 N + O(1)$. $\square$

Lemma 5.3. *Let $k\geqslant 1$ and $\mathcal{I} \subseteq \{1,\dots,k-1\}$. If the variables m_i are fixed $(1\leqslant i\leqslant k-1, i\notin\mathcal{I})$, then for any prime p,*

$$\sum_{0\leqslant m_i<p\ (i\in\mathcal{I})} \xi(p,(m_1,\dots,m_{k-1})) \geqslant p^{|\mathcal{I}|+1} - (p-1)^{|\mathcal{I}|+1}.$$

Proof. Fix p and let $N(k,\mathcal{I})$ be the sum on the left side. We use induction on k, the case $k=1$ being trivial. Suppose $k\geqslant 2$ and the lemma holds with k replaced by $k-1$. If $k-1 \notin \mathcal{I}$, then $\xi(p,(m_1,\dots,m_{k-1})) \geqslant \xi(p,(m_1,\dots,m_{k-2}))$ implies $N(k,\mathcal{I}) \geqslant N(k-1,\mathcal{I})$.

If $k-1 \in \mathcal{I}$, then $N(k,\mathcal{I})$ counts the number of $(|\mathcal{I}|+1)-$tuples $(m_i(i \in \mathcal{I}), n)$ modulo p with $p|f_1(n)\cdots f_{k-1}(n)(m_{k-1}f_{k-1}(n)+1)$. The number of tuples with $p|f_1(n)\cdots f_{k-1}(n)$ is $pN(k-1,\mathcal{I}-\{k-1\})$ and the number of remaining tuples is $p^{|\mathcal{I}|}-N(k-1,\mathcal{I}-\{k-1\})$. By the inductive hypothesis, $N(k,I)=p^{|\mathcal{I}|}+(p-1)N(k-1,\mathcal{I}-\{k-1\}) \geqslant p^{|\mathcal{I}|+1}-(p-1)^{|\mathcal{I}|+1}$. $\square$

Lemma 5.4. *Let $k \geqslant 4$, and suppose that M_i, N_i are integers satisfying $M_i \geqslant 2$ and $2 \leqslant N_i \leqslant 2kM_i$ for $1 \leqslant i \leqslant k-1$. For some positive constant c_2, we have*

$$\sum_{\substack{N_i<m_i\leqslant N_i+M_i \\ (1\leqslant i\leqslant k-1)}} \mathfrak{S}(\mathbf{m}) \ll M_1\cdots M_{k-1}\,(c_2\log k)^b \exp\left\{O\left(\frac{k\log_2 k}{\log k}\right)\right\},$$

where b is the number of variables M_i which are $\leqslant 2^{k^2\log^3 k}$.

Proof. Let $L=\lfloor \log k\rfloor+1$ and $r=k^2L$. We will perform a precise averaging of the factors in $\mathfrak{S}(\mathbf{m})$ for primes $p\leqslant r$, and use crude estimates for larger p. If $p\nmid m_1\cdots m_{k-1}$, each congruence $f_j(n)\equiv 0 \pmod p$ has exactly one solution. For $h>j$, $f_j(n)\equiv 0 \pmod p$ and $f_h(n)\equiv 0 \pmod p$ have a common solution if and only if $p|(a_jb_h-a_hb_j)$. Write

$$a_jb_h-a_hb_j = m_1\cdots m_{j-1}g_{j,h}(\mathbf{m}), \qquad g_{j,h}(\mathbf{m}) := 1+\sum_{i=j+1}^{h-1} m_i\cdots m_{h-1}. \tag{5.2}$$

Define

$$\psi_r(n)=\prod_{\substack{p|n\\ p>r}}\frac{p}{p-1}, \qquad G_{j,h}(\mathbf{m})=\prod_{\substack{p>r, p|g_{j,h}(\mathbf{m})\\ p\nmid g_{i,h}(\mathbf{m})\ (j+1\leqslant i\leqslant h-2)\\ p\nmid m_1\cdots m_{k-1}}} p.$$

We then have

$$\prod_{\substack{p\nmid m_1\cdots m_{k-1}\\ p>r}}\left(1-\frac{\xi(p,\mathbf{m})}{p}\right)\left(1-\frac{1}{p}\right)^{-k} \leqslant \prod_{\substack{1\leqslant j<h\leqslant k-1\\ h\geqslant j+2}} \psi_r(G_{j,h}(\mathbf{m})).$$

Let

$$\mathcal{J}=\{(j,h): 1\leqslant j<h\leqslant k-1, h\geqslant j+2, \max_{j+1\leqslant i\leqslant h-1} M_i > 2^{k^2\log^3 k}\},$$

and put $J=|\mathcal{J}|\leqslant \frac{(k-3)(k-2)}{2}$. Also let $\mathcal{I}=\{i : M_i > 2^{k^2\log^3 k}\}$. Write $\mathcal{I}=\bigcup_{a=1}^{A}([i_a,i_a']\cap\mathbb{N})$, where $i_{a+1}\geqslant i_a'+2$ for each a. For each a and $2+i_a'\leqslant h\leqslant i_{a+1}$ (so $h\notin\mathcal{I}$),

$$\prod_{i_a'\leqslant j\leqslant h-2}\psi_r(G_{j,h}(\mathbf{m}))=\psi_r(G_h), \quad G_h=\prod_{i_a'\leqslant j\leqslant h-2} G_{j,h}(\mathbf{m}).$$

Since $G_h \leqslant (k2^{k^2\log^3 k})^k$, G_h has $O(k^3\log^3 k)$ prime factors, and thus for some constant $C > 1$, $\psi_r(G_h) \leqslant \exp\{\sum_{p|G_h, p>r} \frac{1}{p-1}\} \leqslant C$. There are b such numbers h. Hence, by Hölder's inequality,

$$\begin{aligned}\sum_{\mathbf{m}} \mathfrak{S}(\mathbf{m}) &\leqslant C^b \sum_{\mathbf{m}} \prod_{p\leqslant r}\left(1-\frac{\xi(p,\mathbf{m})}{p}\right)\left(1-\frac{1}{p}\right)^{-k} \prod_{i=1}^{k-1}\psi_r(m_i)^{k-1} \prod_{(j,h)\in\mathcal{J}} \psi_r(G_{j,h}(\mathbf{m})) \\ &\leqslant C^b \biggl(\sum_{\mathbf{m}}\biggl[\prod_{p\leqslant r}\left(1-\frac{\xi(p,\mathbf{m})}{p}\right)\left(1-\frac{1}{p}\right)^{-k}\biggr]^{\frac{L}{L-1}}\biggr)^{1-\frac{1}{L}} \\ &\quad\times \prod_{i=1}^{k-1}\biggl(\sum_{\mathbf{m}}\psi_r(m_i)^{2L(k-1)^2}\biggr)^{\frac{1}{2L(k-1)}} \prod_{(j,h)\in\mathcal{J}}\biggl(\sum_{\mathbf{m}}\psi_r(G_{j,h}(\mathbf{m}))^{2JL}\biggr)^{\frac{1}{2JL}}.\end{aligned} \tag{5.3}$$

If we write $\psi_r^s = 1 * \beta_s$, then β_s is multiplicative and supported on square-free integers composed of primes $> r$. Furthermore, if $p > r \geqslant s+1$, then

$$\beta_s(p) = \left(\frac{p}{p-1}\right)^s - 1 \leqslant \mathrm{e}^{s/(p-1)} - 1 \leqslant \frac{4s}{p}. \tag{5.4}$$

Thus, for each i,

$$\begin{aligned}\sum_{N_i<m_i\leqslant N_i+M_i} \psi_r(m_i)^{2L(k-1)^2} &\leqslant \sum_{d\leqslant N_i+M_i} \beta_{2L(k-1)^2}(d)\frac{N_i+M_i}{d} \\ &\leqslant (2k+1)M_i\prod_{p>r}\left(1+\frac{\beta_{2L(k-1)^2}(p)}{p}\right) \ll kM_i.\end{aligned} \tag{5.5}$$

For fixed $(j,h)\in\mathcal{J}$, let $M_l = \max(M_{j+1},\ldots,M_{h-1}) > 2^{k^2\log^3 k}$ and write

$$g_{j,h}(\mathbf{m}) = m_l(m_{l+1}\cdots m_{h-1})g_{j,l}(\mathbf{m}) + g_{l,h}(\mathbf{m}). \tag{5.6}$$

We'll use

$$G_{j,h}(\mathbf{m})|G'_{j,h}(\mathbf{m}) := \prod_{\substack{p>r,\, p|g_{j,h}(\mathbf{m}) \\ p\nmid m_1\cdots m_{k-1}g_{l,h}(\mathbf{m})}} p,$$

and note that $G'_{j,h}(\mathbf{m}) \leqslant g_{j,h}(\mathbf{m}) \leqslant k(2k+1)^k M_{j+1}\ldots M_{h-1} \leqslant (6kM_l)^k$ by (5.2). Fix all of $m_{j+1},\ldots,m_{h-1}$ except for m_l. By (5.4) and (5.6),

(5.7)

$$\begin{aligned}\sum_{m_l}\psi_r(G'_{j,h}(\mathbf{m}))^{2JL} &= \sum_{\substack{d\leqslant(6kM_l)^k \\ (d,g_{l,h}(\mathbf{m})\prod_{i\neq l}m_i)=1}} \beta_{2JL}(d) \sum_{\substack{N_l<m_l\leqslant N_l+M_l \\ d|G'_{j,h}(\mathbf{m})}} 1 \leqslant \sum_{d\leqslant(6kM_l)^k}\left(\frac{M_l}{d}+1\right)\beta_{2JL}(d) \\ &\leqslant M_l\prod_{p>r}\left(1+\frac{8JL}{p^2}\right) + \prod_{r<p\leqslant(6kM_l)^k}\left(1+\frac{8JL}{p}\right) \ll M_l.\end{aligned}$$

Also,

$$\left[\prod_{p\leqslant r}\left(1-\frac{\xi(p,\mathbf{m})}{p}\right)\left(1-\frac{1}{p}\right)^{-k}\right]^{\frac{L}{L-1}-1} \leqslant \prod_{p\leqslant r}\left(1-\frac{1}{p}\right)^{-\frac{k}{L-1}} = \exp\left\{O\left(\frac{k\log_2 k}{\log k}\right)\right\}.$$

Therefore, by (5.3), (5.5) and (5.7),

$$\sum_{\mathbf{m}} \mathfrak{S}(\mathbf{m}) \ll C^b\,(M_1\cdots M_{k-1})^{\frac{1}{L}}\, S^{1-\frac{1}{L}} \exp\left\{O\left(\frac{k\log_2 k}{\log k}\right)\right\},$$

where

$$S = \sum_{\mathbf{m}}\prod_{p\leqslant r}\left(1-\frac{\xi(p,\mathbf{m})}{p}\right)\left(1-\frac{1}{p}\right)^{-k}.$$

Let $M' = \prod_{p\leqslant r} p$. Since $M' \leqslant \mathrm{e}^{2r}$, for each $i\in\mathcal{I}$, $M_i \gg kM'$. Hence, the number of $m_i \in (N_i, N_i+M_i]$ lying in a given residue class modulo M' is $\leqslant M_i/M'+1 \leqslant (1+O(1/k))M_i/M'$. Thus, by Lemma 5.3 and the Chinese Remainder Theorem,

$$\begin{aligned} S &\leqslant \sum_{\substack{N_i<m_i\leqslant N_i+M_i\\(i\notin\mathcal{I})}}\prod_{i\in\mathcal{I}}\frac{M_i}{M'}\left(1+O\left(\frac{1}{k}\right)\right)\sum_{\substack{m_i \bmod M'\\(i\in\mathcal{I})}}\prod_{p\leqslant r}\left(1-\frac{\xi(p,\mathbf{m})}{p}\right)\left(1-\frac{1}{p}\right)^{-k}\\ &\ll \prod_{i\in\mathcal{I}} M_i \sum_{\substack{N_i<m_i\leqslant N_i+M_i\\(i\notin\mathcal{I})}}\prod_{p\leqslant r}\frac{1}{p^{|\mathcal{I}|}}\left(\frac{p}{p-1}\right)^k\left[p^{|\mathcal{I}|}-\frac{1}{p}\sum_{\substack{0\leqslant m_i<p\\(i\in\mathcal{I})}}\xi(p,\mathbf{m})\right]\\ &\ll M_1\cdots M_{k-1}\prod_{p\leqslant r}\left(\frac{p}{p-1}\right)^{k-1-|\mathcal{I}|}. \end{aligned}$$

Noting that $b=k-1-|\mathcal{I}|$, the lemma follows from Mertens' estimate. □

Theorem 7. *Suppose that $\eta>0$, $r\geqslant 1$, and l and x are sufficiently large as a function of η. There are*

$$\ll \frac{x}{\log x}(2\eta\mathrm{e}^{1+\eta})^{l/2} + \sum_{j=1}^{r} x(\log_2 x)^{O(jl)}\left(\frac{(jl)^{3+\eta}}{\log x}\right)^{\lfloor jl/(\log_2 jl)^2\rfloor}$$

primes $p\leqslant x$, such that there is a prime chain $p_{rl}\prec p_{rl-1}\prec\cdots\prec p_0=p$ with $p_{rl}>x^{(r+1)^{-\eta}}$.

Proof. Suppose $2\eta\mathrm{e}^{1+\eta}<1$ and $rl\leqslant(\log x)^{1/(3+\eta)}$, else the theorem is trivial. Put $k_j=jl$ and $x_j=x^{(j+1)^{-\eta}}$ for $0\leqslant j\leqslant r$. Suppose $p\leqslant x$ and there are even integers $h_1,\ldots,h_{k_r}$ so that

$$p=p_0=h_1p_1+1, p_1=h_2p_2+1,\ldots,p_{k_r-1}=h_{k_r}p_{k_r}+1, \tag{5.8}$$

with $p_0,\ldots,p_{k_r}$ prime and $p_{k_r}\geqslant x_r$. The vector $(h_1,\ldots,h_{k_r})$ may not be unique, but we associate to each such p a single such vector. Each p lies in $Q_1\cup\cdots\cup Q_r$, where Q_j is the set of primes p so that $p_{k_i}<x_i$ $(i<j)$ and $p_{k_j}\geqslant x_j$. By assumption, $k_r\leqslant(\log x_r)^{1/3}$.

Fix j and even integers $h_1,\ldots,h_{k_j}$ satisfying $h_1\cdots h_{k_j}\leqslant x/x_j$. By Lemma 5.2,

$$\sum_p \frac{k_j-\xi(p;(h_{k_j},\ldots,h_1))}{p}\log p \ll k_j\log_2 x \ll (\log x_r)^{1/3}\log_2 x.$$

By Lemma 5.1 and Stirling's formula, the number of $p = p_0 \leqslant x$ satisfying (5.8) is

$$\ll \frac{x}{h_1 \cdots h_{k_j}} \frac{(2k_j/\mathrm{e})^{k_j+3/2}\mathfrak{S}(h_{k_j}, \dots, h_1)}{(\log x_j)^{k_j+1}}. \tag{5.9}$$

Let $1 \leqslant b_j \leqslant k_j$ be a parameter to be chosen later, and put $A_j = 2^{2k_j^2 \log^3 k_j}$. Let $Q_{j,1}$ be the set of $p \in Q_j$ for which at least b_j of the variables $h_1, \dots, h_{k_j}$ are $\leqslant A_j$, and $Q_{j,2} = Q_j \backslash Q_{j,1}$.

To estimate $|Q_{j,1}|$, fix a set $\mathcal{B} \subseteq \{1, \dots, k_j\}$ of size b_j so that $h_i \leqslant A_j$ for each $i \in \mathcal{B}$. Let $\mathcal{I} = \{1 \leqslant i \leqslant k_j : i \notin \mathcal{B}\}$ and, for $0 \leqslant i \leqslant j-1$, put $a_i = |\mathcal{B} \cap \{k_i+1, \dots, k_{i+1}\}|$, $\mathcal{I}_i = \mathcal{I} \cap \{k_i+1, \dots, k_{i+1}\}$. By the definition of Q_j,

$$\prod_{g \in \mathcal{I}_i \cup \cdots \cup \mathcal{I}_{j-1}} h_g \leqslant h_{k_i+1} \cdots h_{k_j} \leqslant \frac{x_i}{x_j} \quad (0 \leqslant i \leqslant j-1). \tag{5.10}$$

Since $h_i \geqslant 2$ for all i,

$$\prod_{g \in \mathcal{B}} \sum_{2 \leqslant h_g \leqslant A_j} \frac{1}{h_g} \leqslant (2k_j^2 \log^3 k_j)^{b_j}. \tag{5.11}$$

Let $\alpha = l/\log x_j$. By (5.10) and the elementary estimate $\sum_{2 \leqslant h \leqslant y} h^{-1-s} \leqslant 1/s$,

$$\begin{aligned}
\sum_{h_g\ (g \in \mathcal{I})} \frac{1}{\prod_{g \in \mathcal{I}} h_g} &\leqslant \sum_{h_g \geqslant 2\ (g \in \mathcal{I})} \frac{1}{\prod_{g \in \mathcal{I}} h_g} \prod_{i=0}^{j-1} \left(\frac{x_i}{x_j}\right)^{\alpha} \frac{1}{\prod_{g \in \mathcal{I}_i \cup \cdots \cup \mathcal{I}_{j-1}} h_g^{\alpha}} \\
&= \prod_{i=0}^{j-1} \left(\frac{x_i}{x_j}\right)^{\alpha} \sum_{h_g \geqslant 2\ (g \in \mathcal{I}_i)} \frac{1}{\prod_{g \in \mathcal{I}_i} h_g^{1+(i+1)\alpha}} \\
&\leqslant \prod_{i=0}^{j-1} \left(\frac{x_i}{x_j}\right)^{\alpha} \left(\frac{1}{(i+1)\alpha}\right)^{k_{i+1}-k_i-a_i} \\
&= \left(\frac{1}{\alpha}\right)^{k_j-b_j} \frac{1^{a_0} 2^{a_1} \cdots j^{a_{j-1}}}{(j!)^l} \exp\left\{ l \sum_{i=0}^{j} \left[\left(\frac{j+1}{i+1}\right)^{\eta} - 1 \right] \right\}.
\end{aligned}$$

The last sum is $\leqslant \frac{\eta}{1-\eta}(j+1) \leqslant 2j$. Also, $1^{a_0} 2^{a_1} \cdots j^{a_{j-1}} \leqslant j^{b_j}$ and $j! \geqslant (j/\mathrm{e})^j$. Hence,

$$\sum_{h_g\ (g \in \mathcal{I})} \frac{1}{\prod_{g \in \mathcal{I}} h_g} \leqslant \mathrm{e}^{3k_j} \left(\frac{\log x_j}{k_j}\right)^{k_j-b_j}. \tag{5.12}$$

The number of choices for $\mathcal{B}$ is $\binom{k_j}{b_j} \leqslant (\mathrm{e}k_j/b_j)^{b_j}$. By (5.9), (5.11), (5.12), and Lemma 5.2,

$$\begin{aligned} |Q_{j,1}| &\ll \frac{x\,(c_1 k_j \log_2 x)^{k_j+3/2}}{(\log x_j)^{k_j+1}} \left(\frac{2\mathrm{e}k_j^3 \log^3 k_j}{b_j}\right)^{b_j} \mathrm{e}^{3k_j} \left(\frac{\log x_j}{k_j}\right)^{k_j-b_j} \\ &= \frac{x}{\log x_j} k_j^{3/2} \left(c_1 \mathrm{e}^3 \log_2 x\right)^{k_j+3/2} \left(\frac{2\mathrm{e}k_j^3 (k_j/b_j) \log^3 k_j (j+1)^\eta}{\log x}\right)^{b_j} \\ &\ll x \exp\left\{ O(k_j \log_3 x) + b_j \left[(3+\eta)\log k_j + \log\left(\frac{k_j}{b_j}\right) - \log_2 x \right] \right\}. \end{aligned} \tag{5.13}$$

We next estimate $|Q_{j,2}|$. Place each variable h_i into an interval J_i. If $h_i \leqslant A_j$, then take $J_i = (2^{l_i-1}, 2^{l_i}]$ for an integer $l_i \geqslant 1$, and if $h_i > A_j$, then take

$$J_i = \left(\lfloor A_j(1+1/k_j)^{l_i-1}\rfloor, \lfloor A_j(1+1/k_j)^{l_i}\rfloor\right]$$

for some integer $l_i \geqslant 1$. For brevity, write $J_i = (H_i, K_i]$ for each i. Since $K_i - H_i \geqslant H_i/(2k_j)$, there are at most b_j values of i with $K_i - H_i \leqslant A_j$. Lemma 5.4 then gives

$$\begin{aligned} \sum_{h_1 \in J_1} \cdots \sum_{h_{k_j} \in J_{k_j}} \frac{\mathfrak{S}(h_{k_j}, \ldots, k_1)}{h_1 \cdots h_{k_j}} &\leqslant \frac{1}{H_1 \cdots H_{k_j}} \sum_{h_1 \in J_1} \cdots \sum_{h_{k_j} \in J_{k_j}} \mathfrak{S}(h_{k_j}, \ldots, h_1) \\ &\ll (c_2 \log k_j)^{b_j} \exp\left\{ O\left(\frac{k_j \log_2 k_j}{\log k_j}\right)\right\} \prod_{i=1}^{k_j} \frac{K_i - H_i}{H_i}. \end{aligned}$$

By our definition of the intervals J_i,

$$\begin{aligned} \prod_{i=1}^{k_j} \frac{K_i - H_i}{H_i} &\leqslant \prod_{\substack{1 \leqslant i \leqslant k_j \\ K_i - H_i < A_j}} 2 \sum_{H_i < h_i \leqslant K_i} \frac{1}{h_i} \prod_{\substack{1 \leqslant i \leqslant k_j \\ K_i - H_i \geqslant A_j}} \left(1 + O\left(\frac{1}{k_j}\right)\right) \sum_{H_i < h_i \leqslant K_i} \frac{1}{h_i} \\ &\ll 2^{b_j} \sum_{h_1 \in J_1} \cdots \sum_{h_{k_j} \in J_{k_j}} \frac{1}{h_1 \cdots h_{k_j}}. \end{aligned}$$

Thus, after summing over all possibilities for $J_1, \ldots, J_{k_j}$, we obtain by (5.9)

$$|Q_{j,2}| \ll \frac{x(2k_j/\mathrm{e})^{k_j+3/2}}{(\log x_j)^{k_j+1}} \exp\left\{ O\left(\frac{k_j \log_2 k_j}{\log k_j} + b_j \log_2 k_j\right)\right\} \sum_{h'_1, \ldots h'_{k_j}} \frac{1}{h'_1 \cdots h'_{k_j}},$$

where $h'_{k_i+1} \cdots h'_{k_j} \leqslant 2^{k_j - k_i} h_{k_i+1} \cdots h_{k_j} \leqslant 2^{k_j} x_i/x_j$ for $0 \leqslant i \leqslant j-1$. For positive $\alpha_0, \ldots, \alpha_{j-1}$,

$$\begin{aligned} \sum_{h'_1, \ldots, h'_{k_j}} \frac{1}{h'_1 \cdots h'_{k_j}} &\leqslant \prod_{i=0}^{j-1} \left[\left(2^{k_j} \frac{x_i}{x_j}\right)^{\alpha_i} \sum_{h'_{k_i+1}, \ldots, h'_{k_{i+1}} = 2}^{\infty} \frac{1}{(h'_{k_i+1} \cdots h'_{k_{i+1}})^{1+\alpha_0+\cdots+\alpha_i}} \right] \\ &\leqslant \prod_{i=0}^{j-1} \left(2^{k_j} \frac{x_i}{x_j}\right)^{\alpha_i} \left(\frac{1}{\alpha_0 + \cdots + \alpha_i}\right)^{l}. \end{aligned}$$

If we ignore the factors $2^{k_j\alpha_i}$, the optimal choice of parameters is

$$\alpha_i = \frac{l}{(j+1)^\eta \log x_j}\left[\frac{(i+2)^\eta (i+1)^\eta}{(i+2)^\eta - (i+1)^\eta} - \frac{(i+1)^\eta i^\eta}{(i+1)^\eta - i^\eta}\right], \qquad i = 0, \dots, j-1.$$

Since $(i+2)^\eta - (i+1)^\eta \geqslant \eta(i+2)^{\eta-1}$,

$$\alpha_0 + \cdots + \alpha_i \in \Big[\frac{l(i+1)(i+2)^\eta}{\eta(j+1)^\eta \log x_j}, \frac{l(i+2)(i+1)^\eta}{\eta(j+1)^\eta \log x_j}\Big].$$

Recalling $k_j = jl$, the sum on $h'_1, \dots, h'_{k_j}$ is at most

$$2^{k_j(\alpha_0+\cdots+\alpha_{j-1})} \exp\left\{\sum_{i=0}^{j-1} \alpha_i \left[\left(\frac{j+1}{i+1}\right)^\eta - 1\right] \log x_j\right\} \left(\frac{\eta(j+1)^\eta \log x_j}{l}\right)^{k_j} \frac{1}{(j!)^l((j+1)!)^{\eta l}}.$$

The exponential factor is $\mathrm{e}^{lj} = \mathrm{e}^{k_j}$ and $(j+1)! \geqslant e^{-j-1}(j+1)^{j+1}$, so

$$\sum_{h'_1,\dots,h'_{k_j}} \frac{1}{h'_1 \cdots h'_{k_j}} \leqslant 2^{2k_j^2/(\eta \log x_j)} \left(\frac{\eta \mathrm{e}^{2+\eta} \log x_j}{k_j}\right)^{k_j}.$$

Therefore,

$$|Q_{j,2}| \ll \frac{x}{\log x}\left(2\eta \mathrm{e}^{1+\eta}\right)^{k_j} \exp\left\{O\left(\frac{k_j \log_2 k_j}{\log k_j} + b_j \log_2 k_j\right)\right\}. \tag{5.14}$$

Finally, put $b_j = \lfloor k_j/(\log_2 k_j)^2 \rfloor$, and sum the inequalities (5.13) and (5.14) for $1 \leqslant j \leqslant r$. □

Proof of Theorem 4. Let $\eta = 0.15718$, $l = \lfloor (\log x)^\varepsilon \rfloor$ and $r = \lfloor (\log x)^\beta \rfloor$, where ε and β are fixed and satisfy $0 < \varepsilon + \beta < \frac{1}{3+\eta}$. Then $\log x_r \asymp (\log x)^{1-\eta\beta}$. For the primes p not counted in Theorem 7, the primes at level rl of the Pratt tree are all $< x_r$, so $H(p) \leqslant \frac{\log x_r}{\log 2} + 1 + rl \ll (\log x)^{0.95022}$ if we take β sufficiently close to $\frac{1}{3+\eta}$. By Theorem 7, the number of exceptional primes $p \leqslant x$ is $O(x \exp\{-(\log x)^\delta\})$ for some $\delta > 0$. □

Proof of Theorem 5. Let x be large, $x/\log x < n \leqslant x$ and suppose there is a prime $p > x^{\varepsilon/2}$ such that $p|\phi_k(n)$. Then either (i) there is a prime $q > x^{\varepsilon/2}$ and $0 \leqslant j \leqslant k$ such that $q^2|\phi_j(n)$, or (ii) there is a prime chain $p = p_k \prec p_{k-1} \prec \cdots \prec p_1 \prec p_0$ with $p_0|n$. In case (i), let j be the smallest such index. Using the uniform estimate

$$\sum_{\substack{p \leqslant x \\ p \equiv 1 \pmod m}} \frac{1}{p} \ll \frac{\log_2 x}{\phi(m)},$$

coming from the Brun-Titchmarsh inequality, the number of integers in category (i) is

$$\leqslant \sum_{q > x^{\varepsilon/2}} \frac{x}{q^2} + \sum_{j=1}^{k} \sum_{x^{\varepsilon/2} < q \leqslant x} \sum_{\substack{p_{j-1} \equiv 1 \pmod{q^2} \\ p_{j-1} \leqslant x}} \sum_{\substack{p_{j-2} \equiv 1 \pmod{p_{j-1}} \\ p_{j-2} \leqslant x}} \cdots \sum_{\substack{p_0 \equiv 1 \pmod{p_1} \\ p_0 \leqslant x}} \frac{x}{p_0}$$

$$\ll_{\varepsilon,k} \frac{x^{1-\varepsilon/2}}{\log x} + \sum_{j=1}^{k} \frac{x^{1-\varepsilon/2}(\log_2 x)^j}{\log x} \ll_{\varepsilon,k} x^{1-\varepsilon/2}.$$

Consider n in category (ii). Take $\eta = \frac{1}{7}$, let r be the smallest integer with $(r+1)^{-\eta} < \varepsilon/2$, let l be sufficiently large, $l \leqslant \log_2 x$ and $k = rl$. By Theorem 7, for $x^{\varepsilon/2} < y \leqslant x$, the number of $p_0 \leqslant y$ is $O(y/\log^2 y + y(2\eta e^{1+\eta})^{-l/2}/\log y)$. By partial summation, the number of n is $\ll_\varepsilon x(2\eta e^{1+\eta})^{-l/2}$. Taking l large enough, depending on ε and δ, completes the proof. □

6. Stochastic model of Pratt trees

In this section, we develop a model of the Pratt trees which explains Conjectures 2 and 3. Factor n as $n = \prod_{j=1}^{\Omega(n)} p_j(n)$, with $p_1(n) \geqslant p_2(n) \geqslant \cdots$. Put $p_j(n) = 1$ for $j > \Omega(n)$ and let

$$S(n) = \left(\frac{\log p_1(n)}{\log n}, \frac{\log p_2(n)}{\log n}, \ldots \right).$$

The distribution of the first component of $S(n)$ has been greatly studied, the results having wide application in the theory of numbers (see e.g. the comprehensive survey article [26]). We have[3] $\mathbf{P}(\log p_1(n) \leqslant \frac{1}{u} \log n) = \rho(u)$, where ρ is the *Dickman function*, the unique continuous solution of the differential-delay equations $\rho(u) = 1$ $(0 \leqslant u \leqslant 1)$, $u\rho'(u) = -\rho(u-1)$ $(u > 1)$. The complete distribution of $S(n)$, found by Billingsly in 1972 [11], corresponds to the Poisson-Dirichlet distribution with parameter 1, $PD(1)$ for short (more precisely, for each j, the first j components of $S(n)$ are distributed as the first j components of the $PD(1)$ distribution). The joint distribution of the components in the $PD(1)$ distribution can easily be expressed in terms of ρ. There is a simpler characterization of the distribution, found by Donnelly and Grimmett [17]. Let $U_1, U_2, \ldots$ be independent random variables with uniform distribution on $[0, 1]$. Let $\mathbf{x} = (x_1, x_2, \ldots)$ be the infinite dimensional vector formed from the decreasing rearrangement of the numbers

$$y_1 = U_1, \quad y_2 = (1 - U_1)U_2, \quad y_3 = (1 - U_1)(1 - U_2)U_3, \ldots. \tag{6.1}$$

Then $\mathbf{x}$ has the $PD(1)$ distribution. The paper [17] gives a simple, transparent proof that $(x_1, \ldots, x_k)$ and the first k components of $S(n)$ have the same distribution.

Since $\sum x_i = 1$ with probability 1, we can interpret the $PD(1)$ distribution as a random partition of the unit interval $[0, 1]$ into an infinite number of parts achieved by cutting $[0, 1]$ at a random place (with uniform distribution), then cutting the right sub-interval at a random place, and so on.

Conjecture 5. *As p runs over the set of primes, $S(p-1)$ has $PD(1)$ distribution.*

Conjecture 5 is widely believed, and is a simple consequence of EH. Unconditionally, we know little about primes in progressions to very large moduli. Assume that $S(p-1)$ has $PD(1)$ distribution, $S(q-1)$ has $PD(1)$ distribution for each prime $q|(p-1)$, the vectors $S(q-1)$ for $q|(p-1)$ are independent, and so forth. The primes o4n the first level of the tree, on a logarithmic scale, correspond to a random partition of $[0, 1]$. The primes on the second level correspond to randomly partitioning each of the parts of the original partition, etc. The entire procedure corresponds to what is known as a discrete-time *random fragmentation process*. Random fragmentation processes have been used to model a variety of physical phenomena (e.g., genetic mutations, planet formation) and the growth of certain data structures in computer science. Discrete time fragmentation processes may be recast in the language of *branching random walks*, which we now describe.

[3] If $\mathcal{B} \subseteq \mathcal{A} \subseteq \mathbb{N}$, we say $\mathbf{P}(n \in \mathcal{B} | n \in \mathcal{A}) = \alpha$ if $\displaystyle\lim_{x\to\infty} \frac{|\{n \in \mathcal{B} : n \leqslant x\}|}{|\{n \in \mathcal{A} : n \leqslant x\}|} = \alpha$.

Let M_n be the size of the largest object at time n. Then M_n is a model of $Q_n := \frac{\log q_n}{\log p}$, where q_n is the largest prime at level n of the tree. The event $\{M_n < \frac{\log 2}{\log p}\}$ is a model of the statement "all the primes at level n of the Pratt tree for p are < 2"; that is, $H(p) < n$. Thus, $H(p)$ is modeled by the random variable $T(\frac{\log 2}{\log p})$, where $T(\varepsilon) = \min\{n : M_n \leqslant \varepsilon\}$.

Assuming EH, Lamzouri [28] showed that Q_n has the same distribution as M_n for each *fixed* n (he studies the distribution of $P^+(\phi_n(m))$ for all integers m; the same proofs give the distribution of Q_n). Further, on EH, Lamzouri shows that $\mathbf{P}\{Q_n \leqslant \frac{1}{u}\} = \mathbf{P}\{M_n \leqslant \frac{1}{u}\} = \rho_n(u)$, where, for each fixed n,

$$\rho_n(u) = \left(\frac{1+o(1)}{\log_{n-1}(u)\log_n(u)}\right)^u \qquad (u \to \infty), \tag{6.2}$$

with $\log_0(u) = u$. Our goal is to understand the distribution of M_n as $n \to \infty$.

Create a tree structure from the random fragmentation process as follows: label the root node with zero, beneath the root node put an infinite number of child nodes, each corresponding to one of the fragments of the initial segment $[0, 1]$. Each of these nodes has an infinite number of child nodes, corresponding to the fragments in the second step of the process, and so on. Each node is labeled with the number $-\log x$, where x is the fragment size. This randomly labeled tree corresponds to a *branching random walk* (BRW). More generally, an initial ancestor is at the origin, and who forms the zeroth generation. This parent then produces children, the first generation, which are randomly displaced from the parent according to some law. Each of these children behaves like an independent copy of the parent, their children randomly displaced from their parent according to the same law, and forming the second generation, and so on. In our case, each parent produces an infinite number of offspring, the displacements from their parent given by $V = \{-\log y : y \in Z\}$, where Z is a point set with $PD(1)$ distribution. We'll say that V has LPD (logarithmic Poisson-Dirichlet) distribution from now on.

Let B_n be the minimum label of an individual at time n, so that $B_n = -\log M_n$. The first order behavior of the analog of B_n (law of large numbers) for a general BRW was determined in the 1970s by Biggins, Hammersley and Kingman (see [9]). In our case, Biggins' theorem [9] implies $B_n \sim \frac{n}{\mathrm{e}}$ as $n \to \infty$ almost surely. Thus, $T\left(\frac{\log 2}{\log p}\right) \sim \mathrm{e}\log_2 p$ as $p \to \infty$ almost surely, which justifies Conjecture 2.

Let $b_n = \mathrm{median}(B_n)$. The study of B_n naturally breaks into two parts: (i) global behavior: asymptotics for b_n, and (ii) local behavior: the distribution of $B_n - b_n$. A result of McDiarmid [31] can be used to prove $b_n = \frac{n}{\mathrm{e}} + O(\log n)$, and this was sharpened by Addario-Berry and Ford [1] to

Theorem 8. *We have* $b_n = \frac{n}{\mathrm{e}} + \frac{3}{2\mathrm{e}}\log n + O(1)$.

Corollary 2. *We have median*$\big(T(\varepsilon)\big) = \mathrm{e}\log(1/\varepsilon) - \frac{3}{2}\log_2(1/\varepsilon) + O(1)$.

This justifies part of Conjecture 3. One ingredient in the proof is the following expectation identity. Let $Z_n(t)$ be the number of generation n individuals with position $\leqslant t$, and let $\mathbf{z}^{(n)}$ be the set of positions of generation n individuals. If $\mathbf{v} = (v_1, v_2, \ldots)$ has $PD(1)$ distribution, (6.1) gives

$$\mathbf{E}\sum_{j=1}^{\infty} v_j^s = \mathbf{E}\sum_{k=1}^{\infty}\big((1-U_1)\cdots(1-U_{k-1})U_k\big)^s = \sum_{k=1}^{\infty}\frac{1}{(1+s)^k} = \frac{1}{s},$$

since $\mathbf{E}U_i^s = \mathbf{E}(1-U_i)^s = 1/(1+s)$. By the branching property,

$$\mathbf{E}\sum_{z_n\in\mathbf{z}^{(n)}} \mathrm{e}^{-sz_n} = \mathbf{E}\sum_{z_{n-1}\in\mathbf{z}^{(n-1)}} \mathrm{e}^{-sz_{n-1}}\mathbf{E}\sum_{z_1\in\mathbf{z}^{(1)}} \mathrm{e}^{-sz_1} = \frac{1}{s}\mathbf{E}\sum_{z_{n-1}\in\mathbf{z}^{(n-1)}} \mathrm{e}^{-sz_{n-1}}.$$

By induction, the left side is $1/s^n$, so $\int_0^\infty \mathrm{e}^{-st} d\mathbf{E}Z_n(t) = 1/s^n$. Therefore, $\mathbf{E}Z_n(t) = t^n/n!$. Because $t^n/n! \approx 1$ when $t = \frac{n}{\mathrm{e}} + \frac{1}{2\mathrm{e}}\log n + O(1)$, a naive guess would be $b_n = \frac{n}{\mathrm{e}} + \frac{1}{2\mathrm{e}}\log n + O(1)$. However, for reasons clearly explained in [2] and executed in [1], the leftmost point in the n-th generation of a branching random walk has an atypical ancestry with high probability. Denote the locations of points in the ancestral line of this leftmost point by $0, z_1, z_2, \ldots, z_n = B_n$ with B_n close to b_n. Then usually $z_j \geqslant \frac{j}{n}z_n - O(1)\ \ (1 \leqslant j \leqslant n/2)$. A randomly chosen point $z_j \in \mathbf{z}^{(n)}$ has this property with probability of order $1/n$, so the expected number of such z_j is in fact of order $t^n/(n\cdot n!)$, which is $\geqslant 1$ when $t \geqslant \frac{n}{\mathrm{e}} + \frac{3}{2\mathrm{e}}\log n + O(1)$.

We next discuss the local behavior of B_n. Under very general conditions on the BRW, it is known that $B_n - b_n$ is a tight sequence.[4] The basic idea is that a single individual will, with high probability, produce many offspring a few generations later which are close by. In our situation, tightness on the left for $H(p)$ is relatively easy to prove unconditionally:

Proof of Theorem 6. The conclusion is trivial if $g(x^{1/2}) \leqslant 3K$, so we will assume that $g(x^{1/2}) > 3K$. Let $m = \lfloor g(x^{1/2})/K \rfloor$ so that $m \geqslant 3$. Put $Q = x^{2^{-m}}$ and let T be the set of primes $x^{1/2} < p \leqslant x$ such that there is a prime $q|(p-1)$ with $Q < q \leqslant x^{1/4}$ and $H(q) \geqslant h(q)$. For $p \in T$,

$$H(p) \geqslant 1 + h(q) \geqslant h(Q) \geqslant h(x) - mK \geqslant h(p) - g(p),$$

while by sieve methods (Theorem 4.2 of [24]), for large x

$$|\{x^{1/2} < p \leqslant x : p \notin T\}| \ll \frac{x}{\log x} \prod_{\substack{Q<q\leqslant x^{1/4}\\ H(q)\geqslant h(q)}} \left(1 - \frac{1}{q}\right) \ll \frac{x}{\log x}2^{-mc}. \qquad \square$$

It is also known that under certain conditions on the displacement law of the BRW (e.g. [3]), the analog of $B_n - b_n$ converges in probability to a random variable as $n \to \infty$. This is not known in our case.

Conjecture 6. *$B_n - b_n \to X$ as $n \to \infty$ for a random variable X with continuous distribution.*

If X exists, and the medians satisfy $b_{n+1} - b_n \to \mathrm{e}^{-1}$ as $n \to \infty$ (plausible in light of Theorem (8)), it is easy to see that $X \overset{d}{=} -1/\mathrm{e} + \min_i(z_i + X_i)$, where $(z_1, z_2, \ldots)$ has LPD distribution, $X_1, X_2, \ldots$ are independent copies of X, and $\overset{d}{=}$ means "has the same distribution as". This follows by conditioning on the positions of the first generation individuals (the points z_i); that is, using $B_n \overset{d}{=} \min_i(z_i + B_{n-1}^{(i)})$, where $B_{n-1}^{(i)}$ are independent copies of B_{n-1}. The solutions X of this recursive distributional equation are not known, however.

Unconditionally (whether X exists or not), we prove that $B_n - b_n$ has an exponentially decreasing left tail and doubly-exponentially decreasing right tail. Consequently, if Conjecture 6 holds, then all moments of X exist.

[4]A sequence $X_1, X_2, \ldots$ of random variables is tight if for every $\varepsilon > 0$ there is a number M so that for all j, $\mathbf{P}(|X_j| > M) \leqslant \varepsilon$. In other words, the distribution of X_j does not spread out as $j \to \infty$.

Theorem 9. *(a) For any $c_1 < \mathrm{e}$, we have*

$$\mathbf{P}\{B_n - b_n \leqslant -x\} \ll_{c_1} \mathrm{e}^{-c_1 x} \qquad (n \geqslant 1, x \geqslant 0),$$

and for any $c_2 > 2\mathrm{e}\log(2\mathrm{e})$ and $\eta > 0$,

$$\mathbf{P}\{B_n - b_n \leqslant -x\} \gg_{c_2,\eta} \mathrm{e}^{-c_2 x} \qquad (n \geqslant 1, 0 \leqslant x \leqslant (1/2\mathrm{e} - \eta)n);$$

(b) for any $c_3 < 1$ there is a constant c_4, depending on c_3, so that

$$\mathbf{P}\{B_n \geqslant b_n + x\} \leqslant \exp\left(-\mathrm{e}^{c_3(x-c_4)}\right) \qquad (n \geqslant 1, x \geqslant 0).$$

Remark 3. By (6.2), part (b) is nearly best possible; that is, the conclusion is false if $c_3 > 1$.

The next two lemmas hold for very general branching random walks. A notable feature is that they are *local* results, and tightness of $B_n - b_n$ can be proved without knowing anything about the growth of b_n. We will use Theorem 8 to prove the stronger tail estimates.

Lemma 6.1. *For positive integers m, n and positive real numbers M, N,*

$$\mathbf{P}\{B_{m+n} \geqslant M + N\} \leqslant \mathbf{E}[(\mathbf{P}\{B_n \geqslant N\})^{Z_m(M)}].$$

Proof. Suppose $B_{m+n} \geqslant M + N$ and $Z_m(M) = k$. For each of these k individuals, all of their descendants in generation $m + n$ are offset from their generation m ancestor by at least N. $\square$

Lemma 6.2. *Let m, n be positive integers and let $M > 0, \varepsilon > 0$ be real. If $\mathbf{E}\{(1-\varepsilon)^{Z_m(M)}\} \leqslant \frac{1}{2}$, then $\mathbf{P}\{B_n \leqslant b_{n+m} - M\} \leqslant \varepsilon$. In particular, the conclusion holds if $\mathbf{P}\{Z_m(M) < 1/\varepsilon\} \leqslant \frac{1}{5}$.*

Proof. Let q be the ε-quantile of B_n, that is, $\mathbf{P}\{B_n \leqslant q\} = \varepsilon$. By Lemma 6.1,

$$\mathbf{P}\{B_{m+n} \geqslant M + q\} \leqslant \mathbf{E}\left[(\mathbf{P}\{B_n \geqslant q\})^{Z_m(M)}\right] \leqslant \frac{1}{2}.$$

Therefore, $M + q \geqslant b_{m+n}$, and thus $\mathbf{P}\{B_n \leqslant b_{m+n} - M\} \leqslant \mathbf{P}\{B_n \leqslant q\} = \varepsilon$. To prove the second part, assume that $\mathbf{P}\{Z_m(M) < 1/\varepsilon\} \leqslant \frac{1}{5}$. Then

$$\mathbf{E}\left\{(1-\varepsilon)^{Z_m(M)}\right\} \leqslant \mathbf{P}\{Z_m(m) < \tfrac{1}{\varepsilon}\} + (1 - \mathbf{P}\{Z_m(M) < \tfrac{1}{\varepsilon}\})(1-\varepsilon)^{1/\varepsilon} \leqslant \frac{1}{5} + \frac{4}{5\mathrm{e}} < \frac{1}{2}. \quad \square$$

Lemma 6.3. *For real $t \geqslant 1$ and integer $k \geqslant 1$, we have $\mathbf{P}\{Z_1(t) \geqslant k\} \leqslant (\mathrm{e}t/k)^{k-1}$.*

Proof. The conclusion is trivial if $k \leqslant \mathrm{e}t$, so we suppose $k > \mathrm{e}t$. Take $s = \frac{k}{t} - 1$. By (6.1),

$$\begin{aligned}\mathbf{P}\{Z_1(t) \geqslant k\} &\leqslant \mathbf{P}\left\{(1-U_1)\cdots(1-U_{k-1}) \geqslant \mathrm{e}^{-t}\right\} \\ &\leqslant \mathrm{e}^{st}\int_0^1 \cdots \int_0^1 [(1-u_1)\cdots(1-u_{k-1})]^s du_1 \cdots du_{k-1} = \frac{\mathrm{e}^{st}}{(1+s)^{k-1}}. \quad \square\end{aligned}$$

Lemma 6.4. *For all $r \geqslant 1$, $\theta > 1$ and $\varepsilon > 0$, if x is large then $\mathbf{P}\{Z_r(x) \geqslant \theta^x\} \leqslant \exp\{-(\theta-\varepsilon)^x\}$.*

Proof. When $r = 1$, this follows from Lemma 6.3. Assume it to be true for some $r \geqslant 1$, let θ and ε be given, and assume without loss of generality that $\theta - \varepsilon > 1$. The probability that $Z_r(x) \geqslant (\theta - \varepsilon/3)^x$ is $\leqslant \exp\{-(\theta-\varepsilon/2)^x\}$ for large x. Now suppose $Z_r(x) = j < (\theta-\varepsilon/3)^x$ and $Z_{r+1}(x) \geqslant \theta^x$. Let m_i be the number of children of the i-th largest point in $\mathbf{z}^{(r)}$ which are offset at most x from their parent. Let $\mathcal{I}$ be the set of indices with $m_i \geqslant 100x$. Note that $m_1+\cdots+m_j \geqslant \theta^x$. With $j, m_1, \dots, m_j$ fixed, by Lemma 6.3 the probability that $Z_{r+1}(x) \geqslant \theta^x$ is at most

$$\prod_{i=1}^{j} \mathbf{P}\{Z_1(x) \geqslant m_i\} \leqslant \prod_{i\in\mathcal{I}} \mathrm{e}^{-2m_i} \leqslant \mathrm{e}^{-2(\theta^x-100xj)} \leqslant \exp\{-\theta^x\}.$$

As $m_i \leqslant \mathrm{e}^x$, the number of choices for $j, m_1, \dots, m_j$ is at most $\exp\{(\theta-\varepsilon/4)^x\}$. For large x,

$$\mathbf{P}\{Z_r(x) \geqslant \theta^x\} \leqslant \exp\{-(\theta-\varepsilon/2)^x\} + \exp\{(\theta-\varepsilon/4)^x - \theta^x\} \leqslant \exp\{-(\theta-\varepsilon)^x\}. \qquad \square$$

Proof of Theorem 9. Let $a > 1/\mathrm{e}$ and $0 < \eta < a\mathrm{e}/2$. By [10, Theorem 2], for large r we have $\mathbf{P}\{Z_r(ar) \leqslant (a\mathrm{e}-\eta)^r\} \leqslant \frac{1}{5}$. Let r be so large that, in addition, $b_{n+r} \geqslant b_n + (1/\mathrm{e}-\eta)r$ for all n (r exists by Theorem 8). Apply Lemma 6.2 with $M = ar$, $m = r$, $\varepsilon = (a\mathrm{e}-\eta)^{-r}$. For large integers r,

$$\mathbf{P}\{B_n \leqslant b_n - (a - 1/\mathrm{e} + \eta)r\} \leqslant \mathbf{P}\{B_n \leqslant b_{n+r} - ar\} \leqslant (a\mathrm{e}-\eta)^{-r}.$$

The first estimate follows with $c_1 = \frac{\log(a\mathrm{e}-\eta)}{(a-1/\mathrm{e}+\eta)}$. Fix a, let $\eta \to 0$, then let $a \to 1/\mathrm{e}$, so that $c_1 \to \mathrm{e}$.

For the second part of (a), take $\eta > 0$ and r as before (but fixed here), and let $\delta = (1/\mathrm{e}-\eta)r$, so that $b_{n+r} \geqslant b_n + \delta$ for all n. Since $\rho(u) = 1 - \log u$ for $1 \leqslant u \leqslant 2$, $\mathbf{P}(Z_1(\varepsilon) \geqslant 1) = 1 - \rho(\mathrm{e}^\varepsilon) = \varepsilon$ when $0 \leqslant \varepsilon \leqslant \log 2$. Considering the "leftmost child of the leftmost child of the …of the initial ancestor" in the branching random walk, we have $\mathbf{P}\{B_{kr} \leqslant \delta k/2\} \geqslant \mathbf{P}\{Z_1(\delta/2r) \geqslant 1\}^{kr} = (\delta/2r)^{kr}$ for every $k \geqslant 1$. Hence,

$$\mathbf{P}\{B_n \leqslant b_{n-kr} + \delta k/2\} \geqslant \mathbf{P}\{B_{n-kr} \leqslant b_{n-kr}\}\mathbf{P}\{B_{kr} \leqslant \delta k/2\} \geqslant \frac{1}{2}\left(\frac{\delta}{2r}\right)^{kr}.$$

By assumption, $b_{n-kr} + \delta k/2 \leqslant b_n - \delta k/2$. Hence, for $0 \leqslant k \leqslant n/r$ we have

$$\mathbf{P}\{B_n \leqslant b_n - \delta k/2\} \geqslant \frac{1}{2}\left(\frac{\delta}{2r}\right)^{kr}.$$

This gives the desired bound when $0 \leqslant x \leqslant \frac{\delta n}{2r}$, with $c_1' = \frac{2r}{\delta}\log\frac{2r}{\delta}$.

To show part (b), we use induction on n to show that

$$\mathbf{P}\{B_n \geqslant b_n + x\} \leqslant 2^{-\exp\{c_3(x-c_5)\}} \tag{6.3}$$

for $n \geqslant 1$ and $x \geqslant 0$, where c_5 is sufficiently large. Theorem 9 (c) then follows with $c_4 = c_5 - \frac{\log\log 2}{c_3}$. As (6.3) is trivial for $0 \leqslant x \leqslant c_5$, we may assume $x \geqslant c_5$. Let r, δ be such that $b_{n+r} - b_n \geqslant \delta > 0$ for all n (the relationship between r and δ is unimportant). Let A be a large integer, so that if $R = Ar$ and $\Delta = A\delta$, then $2\mathrm{e}^{2-\Delta} \leqslant 1 - c_3$. Also suppose $c_3 \geqslant \frac{1}{2}$. For $1 \leqslant n \leqslant R$, (6.2) implies $\mathbf{P}\{B_n \geqslant b_n + x\} \leqslant \mathbf{P}\{B_n \geqslant x\} = \rho_n(\mathrm{e}^x) \leqslant \exp\{-\mathrm{e}^x\}$ if c_5 is large enough. Suppose now that (6.3) has been proved for $1 \leqslant n \leqslant m-1$, where $m-1 \geqslant R$. Define $\lambda_j = \Delta + \frac{\log j - 1}{c_3}$ for $j \geqslant 1$ Let j_0 be the largest index j with $\lambda_j \leqslant x + \Delta$. Let $z_1 \leqslant z_2 \leqslant \cdots$ be the points in $\mathbf{z}^{(R)}$. For $1 \leqslant j \leqslant j_0$, let P_j be the event $\{z_i > \lambda_i \ (i<j), z_j \leqslant \lambda_j\}$, and the Q be the event $\{z_i > \lambda_i \ (1 \leqslant i \leqslant j_0)\}$. If P_j, then the generation m points descending from each of the j

points $z_1, \dots, z_j$ are offset from their generation R ancestor by at least $b_m + x - \lambda_j$. So

$$\mathbf{P}\{B_m \geqslant b_m + x\} \leqslant \sum_{j=1}^{j_0} \mathbf{P}[P_j]\mathbf{P}\{B_{m-R} \geqslant b_m + x - \lambda_j\}^j + \mathbf{P}[Q].$$

Since $b_m \geqslant b_{m-R} + \Delta$, the induction hypothesis implies that the sum on j is

$$\leqslant \sum_{j=1}^{j_0} \mathbf{P}[P_j] 2^{-j\exp\{c_3(x-c_5+\Delta-\lambda_j)\}} \leqslant \sum_{j=1}^{j_0} \mathbf{P}[P_j] 2^{-\exp\{c_3(x-c_5)+1\}} \leqslant 2^{-1-\exp\{c_3(x-c_5)\}}.$$

Now suppose that Q holds. By the assumption on A,

$$\sum_{j \leqslant j_0} \mathrm{e}^{-z_j} \leqslant \sum_{j=1}^{j_0} \mathrm{e}^{-\lambda_j} \leqslant \mathrm{e}^{-\Delta+1/c_3} \sum_{j=1}^{\infty} j^{-1/c_3} \leqslant \frac{1}{2}.$$

As $\lambda_{j_0} \geqslant x + \Delta - (\lambda_{j_0+1} - \lambda_{j_0}) \geqslant x$,

$$\sum_{\substack{z \in \mathbf{z}^{(R)} \\ z \geqslant x}} \mathrm{e}^{-z} = 1 - \sum_{\substack{z \in \mathbf{z}^{(R)} \\ z < x}} \mathrm{e}^{-z} \geqslant \frac{1}{2}.$$

Let $\varepsilon = \frac{1}{3}(\mathrm{e} - \mathrm{e}^{c_3})$ and $\theta = \mathrm{e} - \varepsilon$, so that $\mathrm{e}^{c_3} < \theta - \varepsilon < \theta < \mathrm{e}$. For some integer $k \geqslant x$, there are $\geqslant \theta^k$ points of $\mathbf{z}^{(R)}$ in $[k-1, k)$, for otherwise

$$\sum_{\substack{z \in \mathbf{z}^{(R)} \\ z \geqslant x}} \mathrm{e}^{-z} \leqslant \sum_{k \geqslant x} \mathrm{e} \left(\frac{\theta}{\mathrm{e}}\right)^k < \frac{1}{2}.$$

By Lemma 6.4, $\mathbf{P}[Q] \leqslant \sum_k \mathbf{P}\{Z_R(k) \geqslant \theta^k\} \leqslant 2\mathrm{e}^{-(\theta-\varepsilon)^x}$. This completes the proof of (b). □

Acknowledgments. The authors wish to thank L. Addario-Berry, J. Bertoin, C. Elsholtz, A. Granville, A. Hildebrand, I. Kátai, C. Pomerance, Y. Sinai, and R. Song for helpful conversations.

REFERENCES

[1] L. Addario-Berry and K. Ford, *Poisson-Dirichlet branching random walks*, preprint.
[2] L. Addario-Berry and B. Reed, *Minima in branching random walks*, Ann. Prob. **37** (2009), 1044–1079.
[3] M. Bachman, *Limit theorems for the minimal position in a branching random walk with independent log-concave displacements*, Adv. Appl. Prob. **32** (2000), 159–176.
[4] R. C. Baker and G. Harman, *Shifted primes without large prime factors*, Acta Arith. **83** (1998), 331–361.
[5] W. D. Banks, J. Friedlander, F. Luca, F. Pappalardi and I. E. Shparlinski, *Coincidences in the values of the Euler and Carmichael functions*, Acta Arith. **122** (2006), 207–234.
[6] W. D. Banks and I. E. Shparlinski, *On values taken by the largest prime factor of shifted primes*, J. Aust. Math. Soc. **82** (2007), no. 1, 133–147.
[7] N. L. Bassily, I. Kátai and M. Wijsmuller, *Number of prime divisors of $\phi_k(n)$, where ϕ_k is the k-fold iterate of ϕ*, J. Number Theory **65** (1997), 226–239.
[8] J. Bayless, *The Lucas-Pratt primality tree*, Math. Comp. **77** (2008), 495-502.
[9] J. D. Biggins, *The first- and last-birth problems for a multitype age-dependent branching process*, Adv. Appl. Prob. **31** (1976), 446–459.
[10] J. D. Biggins, *Chernoff's theorem in the branching randon walk*, J. Appl. Prob. **14** (1977), 630–636.

[11] P. Billingsly, *On the distribution of large prime divisors*, Collection of articles dedicated to the memory of Alfréd Rényi, I. Period. Math. Hungar. **2** (1972), 283–289.

[12] E. Bombieri, *Le grand crible dans la théorie analytique des nombres*, 2nd ed. (French. English summary) Astérisque No. 18. Société Mathématique de France, Paris, 1987.

[13] C. W. Borchardt, *Über eine Interpolationsformel für eine Art Symmetrischer Functionen und über Deren Anwendung*, Math. Abh. der Akademie der Wissenschaften zu Berlin (1860), 1–20.

[14] M. A. Cherepnev, *Some properties of large prime divisors of numbers of the form* $p-1$, Mat. Zametki **80** (2006), 920–925. (Russian). English translation in Math. Notes **80** (2006), 863–867.

[15] R. Crandall and C. Pomerance, *Prime numbers: a computational perspective*, 2nd ed., Springer-Verlag, 2005.

[16] H. Davenport, *Multiplicative number theory*, 3rd ed., Graduate Texts in Mathematics vol. 74, Springer-Verlag, New York, 2000.

[17] P. Donnelly and G. Grimmett, *On the asymptotic distribution of large prime factors*, J. London Math. Soc. (2) **47** (1993), 395–404.

[18] P. Erdős, *On the normal number of prime factors of* $p-1$ *and some related problems concerning Euler's* ϕ*-function*, Quart. J. Math. Oxford **6** (1935), 205-213.

[19] P. Erdős, A. Granville, C. Pomerance, and C. Spiro, *On the normal behavior of the iterates of some arithmetic functions*, in Analytic Number Theory, Proceedings of a conference in honor of Paul T. Bateman, Birkhäuser, Boston, 1990, 165–204.

[20] P. Erdős and C. Pomerance, *On the normal number of prime factors of* $\phi(n)$, Number theory (Winnipeg, Man., 1983). Rocky Mountain J. Math. **15** (1985), no. 2, 343–352.

[21] K. Ford and F. Luca, *The number of solutions of* $\lambda(x) = n$, to appear in INTEGERS, special volume in honor of Melvyn Nathanson and Carl Pomerance.

[22] K. Ford, F. Luca and C. Pomerance, *Common values of the arithmetic functions* ϕ *and* σ, Bull. London Math. Soc. **42** (2010), 478–488.

[23] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. Math. **167** (2008), 481–547.

[24] H. Halberstam and H.-E. Richert, *Sieve methods*, Academic Press, London, 1974.

[25] R. R. Hall and G. Tenenbaum, *Divisors*, Cambridge Tracts in Math., 1988.

[26] A. Hildebrand and G. Tenenbaum, *Integers without large prime factors*, J. de Théorie des Nombres de Bordeaux, **5** (1993), 411–484.

[27] I. Kátai, *Some problems on the iteration of multiplicative number-theoretical functions*, Acta Math. Acad. Scient. Hung. **19** (1968), 441–450.

[28] Y. Lamzouri, *Smooth values of iterates of the Euler phi-function*, Canad. J. Math. **59** (2007), 127–147.

[29] F. Luca and C. Pomerance, *Irreducible radical extensions and Euler-function chains*, in "Combinatorial number theory", 351–361, de Gruyter, Berlin (2007).

[30] G. Martin and C. Pomerance, *The iterated Carmichael* λ*-function and the number of cycles of the power generator*, Acta Arith. **188** (2005), 305–335.

[31] C. McDiarmid, *Minimal positions in a branching random walk*, Ann. Appl. Prob. **5** (1995), no. 1, 128–139.

[32] C. Pomerance, *Very short primality proofs*, Math. Comp. **48** (1987), 315–322.

[33] V. Pratt, *Every prime has a succinct certificate*, SIAM J. Comput. **4** (1975), no. 3, 214–220.

KF: DEPARTMENT OF MATHEMATICS, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN URBANA, 1409 WEST GREEN ST., URBANA, IL 61801, USA

E-mail address: `ford@math.uiuc.edu`

SVK: DEPARTMENT OF MECHANICS AND MATHEMATICS, MOSCOW STATE UNIVERSITY, MOSCOW, 119992, RUSSIA

E-mail address: `konyagin@ok.ru`

FL : INSTITUTO DE MATEMÁTICAS, UNIVERSIDAD NACIONAL AUTONOMA DE MÉXICO, AP. POSTAL 61-3 (XANGARI), C.P. 58089, MORELIA, MICHOACÁN, MÉXICO
E-mail address: fluca@matmor.unam.mx