INTERMEDIATE PRIME FACTORS IN SPECIFIED SUBSETS

NATHAN MCNEW, PAUL POLLACK, AND AKASH SINGHA ROY

Abstract. Let $\mathcal{P}$ be a fixed set of primes possessing a limiting frequency ν, as detected by the weight $1/p$. We show that for any fixed $\alpha \in (0,1)$, the $\lceil \alpha\Omega(n) \rceil$-th smallest prime factor of n, denoted $P^{(\alpha)}(n)$, belongs to $\mathcal{P}$ on a set of n with natural density ν. We prove a similar result for the largest prime factor $P_{\le y}(n)$ of n not exceeding y, whenever $y \to \infty$. As corollaries, $P^{(\alpha)}(n)$ and $P_{\le y}(n)$ conform to Benford's leading digit law. Finally, we establish the equidistribution of $P^{(\alpha)}(n)$ in coprime residue classes, in an essentially optimal range of uniformity in the modulus.

1. Introduction

Given a positive integer n, we let $p_m(n)$ denote the mth smallest prime factor of n, where primes are counted with multiplicity. Equivalently, $p_m(n)$ is defined by the prime factorization $n = \prod_{1 \le k \le \Omega(n)} p_k(n)$, with the condition $p_1(n) \le p_2(n) \le \cdots \le p_{\Omega(n)}(n)$. For each $\alpha \in (0,1)$, we call $p_{\lceil \alpha\Omega(n) \rceil}(n)$ the α-positioned prime factor of n, and we denote it as $P^{(\alpha)}(n)$; for example, $P^{(1/2)}(n)$ is a natural candidate for the "middle" prime factor of n.

Over the past decade, there have been several investigations into the value distribution of the $P^{(\alpha)}(n)$ (see [5, 3, 7, 2, 4]). As a sample of what is known, De Koninck, Doyon, and Ouellet show[1] in [2] that (for fixed α) the quantity $\log\log P^{(\alpha)}(n)$ is asymptotically normally distributed with mean $\alpha \log\log n$ and variance $\alpha(1-\alpha)\log\log n$.

Our first theorem considers the frequency with which $P^{(\alpha)}(n)$ lands inside a specified set of primes $\mathcal{P}$. Recall that the natural density (or asymptotic density) of a set $\mathcal{S}$ of positive integers is the limit of $\frac{1}{x}\#\{n \le x : n \in \mathcal{S}\}$, as $x \to \infty$, provided this limit exists. If $\mathcal{P}$ makes up a proportion ν of all primes, it is natural to guess that the condition $P^{(\alpha)}(n) \in \mathcal{P}$ should cut out a set of n with natural density ν. We prove this is so as long as the proportion ν is detectable by the weights $1/p$ (in the sense of (1.1) below).

Theorem 1.1. *Let $\mathcal{P}$ be any set of primes such that whenever X, Y, and $Y/X \to \infty$,*

$$\sum_{\substack{p \in (X,Y] \\ p \in \mathcal{P}}} \frac{1}{p} \sim \nu \sum_{p \in (X,Y]} \frac{1}{p} \tag{1.1}$$

for some constant $\nu > 0$. Then, for any fixed $\alpha \in (0,1)$, the set of n with $P^{(\alpha)}(n) \in \mathcal{P}$ has asymptotic density ν.

2020 *Mathematics Subject Classification.* Primary 11A63; Secondary 11K16, 11N37.

Key words and phrases. Benford's law, smooth numbers, intermediate prime factors, anatomy of integers.

P.P. is supported by the NSF under award DMS-2001581.

[1] De Koninck, Doyon, and Ouellet state a slightly different definition of α-positioned than the one that we use. They take the prime factor in position $\max\{1, \lfloor \alpha(\Omega(n)+1) \rfloor\}$ rather than position $\lceil \alpha\Omega(n) \rceil$ as we do here. These two definitions often give the same position and never differ by more than 1, however we find our definition to be both easier to work with and to give more satisfactory answers in certain cases. The results about the α-positioned prime factor stated in the cited papers hold for both definitions.

As an application of Theorem 1.1, we show that Benford's leading digit law applies to the sequence of $P^{(\alpha)}(n)$. We now make this precise (see for instance [14] for further background on Benford's law).

Fix a base $g \geq 2$. Let D be a positive integer. We say that the real number $\alpha > 0$ begins with D if the digits of D appear as the most significant digits in the base g expansion of α. For example, in base 10 the number 0.02304 begins with 23 (and 2, and 230, but not 24). Benford's law is the prediction that the frequency of elements in a dataset beginning with a positive integer D should be $\log(1 + D^{-1})/\log g$.

Our interest is in datasets that are sequences of positive real numbers and we take asymptotic density as our measure of frequency. This version of Benford's law applies widely in number theory. For example, the factorials obey Benford (in every base $g \geq 2$): For each D, the asymptotic density of n for which $n!$ begins with D is $\log(1+D^{-1})/\log g$ [6]. The same statement holds if $n!$ is replaced by the nth primorial, meaning the product of the first n primes [13], or by $p(n)$, the number of partitions of n [1]. However, the sequence of natural numbers is not Benford; in fact, for any given D, the collection of natural numbers beginning with D has no asymptotic density. The sequence of prime numbers (in the usual order) also fails to obey Benford's law, for the same reason.

Recently, the second two authors showed that $P^{+}(n)$, the largest prime factor of n, obeys Benford's law [17]. More generally, let $P_1(n) = P^{+}(n)$ and define, inductively, $P_{k+1}(n) = P^{+}(n/P_1(n)\cdots P_k(n))$. Thus, $P_k(n)$ is the kth largest prime factor of n, counted with multiplicity. It is proved in [17] that $P_k(n)$ obeys Benford's law for each fixed k.

That $P^{(\alpha)}(n)$ obeys Benford's law for each fixed $\alpha \in (0, 1)$ is a consequence of Theorem 1.1, applied to the set of primes $\mathcal{P}$ beginning with a prescribed string of digits D. It is important in this application that the condition (1.1) incorporates the weight $1/p$; the simpler-to-prove version of Theorem 1.1 using the 'natural' weight 1 in place of $1/p$ (which could be established by the methods of [3]) would not suffice.

Corollary 1.2. *Fix an integer $g \geq 2$, and fix a positive integer D. Fix $\alpha \in (0, 1)$. The set of n for which $P^{(\alpha)}(n)$ begins with D in base g has asymptotic density* $\log(1 + D^{-1})/\log g$.

In our work above, the "intermediate" prime factors of n are those located a certain percentage of the way through the full list of primes dividing n. It is perhaps just as natural to consider intermediate prime factors from the point of view of their absolute rather than relative size. To address this, we obtain an analogue of Theorem 1.1 for the largest prime factor of n not exceeding y, which we denote by $P_{\leq y}(n)$. (Here we set $P_{\leq y}(n) = 1$ if n is not divisible by any prime in $[2, y]$.) In this result we require that the density of $\mathcal{P}$ be detectable by the weight $\frac{\log p}{p}$. (As shown at the start of the proof of Lemma 2.1, this implies that the density of $\mathcal{P}$ is also detectable by the weight $1/p$.)

Theorem 1.3. *Let $\mathcal{P}$ be any set of primes such that whenever X, Y, and $Y/X \to \infty$,*

$$\sum_{\substack{p \in (X,Y] \\ p \in \mathcal{P}}} \frac{\log p}{p} \sim \nu \sum_{p \in (X,Y]} \frac{\log p}{p} \tag{1.2}$$

for some constant $\nu > 0$. Then, for any function $y = y(x)$ tending to infinity as $x \to \infty$, the n with $P_{\leq y}(n) \in \mathcal{P}$ make up a set of asymptotic density ν.

The following corollary of Theorem 1.3 shows that Benford behavior also emerges for $P_{\leq y}(n)$.

Corollary 1.4. *Fix an integer $g \geq 2$ and a positive integer D. For any function $y = y(x)$ tending to infinity with x, the set of n for which $P_{\leq y}(n)$ begins with D in base g has asymptotic density $\log(1 + D^{-1})/\log g$.*

Besides the Benford behavior highlighted in Corollaries 1.2 and 1.4, many other properties of primes occur with a frequency detectable by the weights $\frac{1}{p}$ and $\frac{\log p}{p}$. For instance, it follows immediately from Theorems 1.1 and 1.3 that $P^{(\alpha)}(n)$ and $P_{\leq y}(n)$ are equidistributed in coprime residue classes mod q, for each fixed modulus q.[2] Our final theorem establishes a uniform result of this kind for $P^{(\alpha)}(n)$, with a best possible range for q.

Theorem 1.5. *Fix $\alpha \in (0, 1)$ and $\epsilon > 0$, and let $c(\alpha) := 1 - 2^{-\alpha/(1-\alpha)}$. The number of $n \leq x$ for which $P^{(\alpha)}(n) \equiv a \pmod q$ is $(1 + o(1))x/\varphi(q)$ as $x \to \infty$, uniformly in moduli $q \leq (\log x)^{c(\alpha)-\epsilon}$ and in coprime residue classes a mod q.*

In particular ($\alpha = 1/2$), the middle prime factor is equidistributed among the coprime residue classes to moduli $q \leq (\log x)^{1/2-\epsilon}$. In a remark following the proof of Theorem 1.5, we show that the range of uniformity in q is essentially optimal in the power of $\log x$, in that uniformity fails for $q > (\log x)^{c(\alpha)+\epsilon}$ for any fixed $\epsilon > 0$.

Notation and conventions. Most of our notation is standard. A possible exception is our use of $\Omega_E(n)$ for the number of primes of E dividing n, counted with multiplicity; explicitly, $\Omega_E(n) := \sum_{p^k \| n,\ p \in E} k$. We set $\Omega_{>z}(n) := \Omega_{(z,\infty)}(n)$. We continue to use $P^+(n)$ for the largest prime factor of n (with $P^+(1) = 1$) and we use $P^-(n)$ for the smallest prime factor of n (taking $P^-(1) = \infty$). We say n is y-smooth if $P^+(n) \leq y$, and we call n y-rough when $P^-(n) > y$. The y-smooth and y-rough parts of n are the largest y-smooth and y-rough divisors of n, respectively. We write $\log_k x$ for the k-fold iterate of the natural logarithm.

2. Benford behavior of intermediate prime factors: Proofs of Corollaries 1.2 and 1.4

The following lemma allows us to deduce Corollaries 1.2 and 1.4 from Theorems 1.1 and 1.3, respectively. For related results see [21], [19, p. 76], [8, Exercises 7.15, 7.16; p. 243–244], and [18].

Lemma 2.1. *Fix an integer $g \geq 2$, and fix a positive integer D. Let $\mathcal{P}$ be the set of primes beginning with D in base g. Whenever X, Y, and Y/X tend to infinity,* (1.1) *and* (1.2) *hold with $\nu = \frac{\log(1+D^{-1})}{\log g}$.*

Proof. To prove Lemma 2.1 it is enough to treat (1.2). More generally, for any given set of primes $\mathcal{P}$ and any $\nu > 0$, if (1.2) holds whenever X, Y, and $Y/X \to \infty$, then so does (1.1). To see this, partition $(X, Y]$ into disjoint intervals $(X', Y']$ where each multiplicative width $Y'/X' \to \infty$ and where $\log(X') \sim \log(Y')$. (For example, we might write $Y/X = X^{\theta/\log\log X}$ and split $(X, Y]$ into $\lceil\theta\rceil$ pieces of equal multiplicative width.) The relation (1.1) follows after applying (1.2) to each $(X', Y']$.

To handle (1.2) we use a strong form of a theorem of Mertens, which follows from the prime number theorem with the usual de la Vallée Poussin error term: There is a real number C and a positive constant K such that for all $x \geq 3$,

$$\sum_{p \leq x} \frac{\log p}{p} = \log x + C_2 + O(\exp(-K\sqrt{\log x})). \tag{2.1}$$

[2] In fact, since primes are equidistributed in coprime progressions mod q already when counted with the 'natural' weight 1 (the prime number theorem for arithmetic progressions), this result on $P^{(\alpha)}(n)$ would already be accessible by the method of [3].

Now suppose X, Y, and $Y/X \to \infty$. For each integer j, put $I_j = [D \cdot g^j, D \cdot g^{j+1})$, and let $\mathcal{J}$ be the collection of j with $I_j \subset (X, Y]$. Set $j_0 = \min \mathcal{J}$ and $j_1 = \max \mathcal{J}$, so that $j_0 = \frac{\log X}{\log g} + O(1)$ and $j_1 = \frac{\log Y}{\log g} + O(1)$. The primes $p \in (X, Y] \setminus \cup_{j \in \mathcal{J}} I_j$ all lie within a bounded factor of X or Y, so that by (2.1) the sum of $\frac{\log p}{p}$ taken over $p \in (X, Y] \setminus \cup_{j \in \mathcal{J}} I_j$ is $O(1)$. The estimate (2.1) also implies that

$$\sum_{\substack{p \in I_j \\ p \text{ begins with } D}} \frac{\log p}{p} = \sum_{p \in [Dg^j, (D+1)g^j)} \frac{1}{p} = \log(1 + D^{-1}) + O(1/j^2).$$

Summing this estimate on j, we find that the left-hand side of (1.2) is $\sim \frac{\log(1+D^{-1})}{\log g} \log(Y/X)$. Since $\sum_{p \in (X,Y]} \frac{\log p}{p} \sim \log(Y/X)$, the asymptotic (1.2) follows. □

3. Behavior of α-positioned prime factors: Proof of Theorem 1.1

We begin with a lemma belonging to the study of 'anatomy of integers'. It is a precise version of the claim that $\Omega_E(n)$, for $n \le x$, is typically of size $\sum_{p \le x,\ p \in E} 1/p$, uniformly across all sets of primes E (cf. Theorem 08 on p. 5 of [11]). Put

$$Q(y) = y \log y - y + 1.$$

Note that $Q(y) = \int_1^y \log t \, \mathrm{d}t$, so that $Q(y) \ge 0$ for all $y > 0$, with equality only when $y = 1$.

Lemma 3.1. *Let $x \ge 3$ and let E be a nonempty set of primes with smallest element p_0. Define $E(x) = \sum_{p \le x,\ p \in E} 1/p$. For $1 \le y \le \min\{100, 0.9p_0\}$,*

$$\frac{1}{x} \sum_{\substack{n \le x \\ \Omega_E(n) \ge yE(x)}} 1 \ll \exp(-E(x) \cdot Q(y)), \qquad \frac{1}{\log x} \sum_{\substack{n \le x \\ \Omega_E(n) \ge yE(x)}} \frac{1}{n} \ll \exp(-E(x) \cdot Q(y)).$$

When $0 < y \le 1$, the same inequalities hold with the $\Omega_E(n)$ condition replaced by $\Omega_E(n) \le yE(x)$. Throughout this statement, implied constants are absolute.

Proof (sketch). We first show the estimates for the sums on $1/n$. The usual Mertens estimate for $\sum_{p \le x} 1/p$ implies that for all y with $0 < y \le \min\{100, 0.9p_0\}$,

$$\sum_{n \le x} \frac{y^{\Omega_E(n)}}{n} \le \sum_{n:\ p \mid n \Rightarrow p \le x} \frac{y^{\Omega_E(n)}}{n} = \prod_{\substack{p \le x \\ p \in E}} \left(1 - \frac{y}{p}\right)^{-1} \prod_{\substack{p \le x \\ p \notin E}} \left(1 - \frac{1}{p}\right)^{-1} \ll (\log x) \exp((y-1)E(x)). \tag{3.1}$$

When $y \ge 1$, the initial sum in (3.1) is of size at least $y^{yE(x)} \sum_{n \le x,\ \Omega_E(n) \ge yE(x)} 1/n$. Hence,

$$\sum_{\substack{n \le x \\ \Omega_E(n) \ge yE(x)}} \frac{1}{n} \ll (\log x) \exp((y-1)E(x)) y^{-yE(x)} = (\log x) \exp(-E(x)Q(y)),$$

as desired. The claimed bound on $\sum_{n \le x,\ \Omega_E(n) \le yE(x)} 1/n$, when $0 < y \le 1$, is shown in the same way. To handle the sums with 1 in place of $1/n$, one argues analogously, replacing (3.1) with

$$\sum_{n \le x} y^{\Omega_E(n)} \ll x \exp((y-1)E(x)). \tag{3.2}$$

The estimate (3.2) can be deduced from (3.1) by a (weak form of a) mean value theorem of Halberstam and Richert, appearing as Theorem 01 in [11]; for a somewhat different treatment of (3.2), see the discussion on pp. 690–691 of [16]. □

We now turn to the 'proof proper' of Theorem 1.1. For the rest of section, $\alpha \in (0,1)$ is fixed. Since we may change X to $X - \frac{1}{2}$ when $X \in \mathbb{Z}$, we see that the hypotheses (1.1) and (1.2) also hold with $(X, Y]$ replaced everywhere by $[X, Y]$, and similar arguments would allow one to take (X, Y) or $[X, Y]$. In what follows we shall make free use of such variants of (1.1) and (1.2).

We now define the α-prefixed prime factor of n to be $p_J(n)$ where $J = \lceil \alpha(\Omega(n) + 1) \rceil$. (If $J > \Omega(n)$ then n does not have an α-prefixed prime factor.) The term 'prefixed' is based on the following simple observation, which will be crucial later: If p is the α-prefixed prime factor of n, then p is the α-positioned prime factor of $n\ell$ for any prime $\ell \geq p$.

Proposition 3.2. *Let $\mathcal{P}$ be any set of primes such that, for some constant $\nu > 0$, the relation (1.1) holds whenever X, Y, and $Y/X \to \infty$. The set of n whose α-prefixed prime factor lies in $\mathcal{P}$ has logarithmic density ν.*

Proof. We follow a similar strategy to [3] but work with weights of $1/p$ instead of 1.

We let x be a large real number (which will be taken to infinity eventually) and we let $n \leq x$ be a positive integer. We restrict attention to n with $\Omega(n) \in W := (\frac{1}{2}\log_2 x, \frac{3}{2}\log_2 x)$. This restriction is harmless, since Lemma 3.1 (with E the full set of primes in $[2, x]$) guarantees that the exceptional $n \leq x$ have reciprocal sum $o(\log x)$. We assume that x is taken sufficiently large (depending on α) so that every integer n with $\Omega(n) \in W$ has an α-prefixed prime factor.

Now fix an integer $w \in W$ and define $J = \lceil \alpha(w+1) \rceil$. Every $n \leq x$ with $\Omega(n) = w$ can be uniquely written as $n = ApB$, where $p = p_J(n)$ is the α-prefixed prime factor of n, $A = \prod_{j<J} p_j(n)$, and $B = \prod_{J<j\leq\Omega(n)} p_j(n)$. Then $\Omega(A) = J-1$, $\Omega(B) = w - J$, and $\min\{x/AB, P^-(B)\} \geq p \geq P^+(A)$. We impose the additional constraint on n that $(A, B) \in \mathcal{L}_w$, where

$$\mathcal{L}_w := \left\{(A,B) : \Omega(A) = J-1,\ \Omega(B) = w - J,\ P^+(A) > \log x,\ \frac{\min\{x/AB, P^-(B)\}}{P^+(A)} \geq \log_3 x\right\}.$$

The n which fail this last constraint for some w have reciprocal sum $o(\log x)$. Indeed, writing $n = ApB$ as above, any offending n has either

(i) $P^+(A) \leq \log x$,

(ii) $P^+(A) > \log x$, but $\frac{\min\{x/AB, P^-(B)\}}{P^+(A)} < \log_3 x$.

In case (i), we apply Lemma 3.1 with E the set of primes not exceeding $\log x$. Each n arising in case (i) has $\Omega_E(n) \geq J - 1 > 2\log_3 x = (2 + o(1))E(x)$, and Lemma 3.1 shows that these n have reciprocal sum $o(\log x)$. In case (ii), n is divisible by a product p_1p_2 of two primes exceeding $\log x$ with $1 \leq p_2/p_1 < \log_3 x$, viz. $p_1 = P^+(A)$ and $p_2 = p$. But the reciprocal sum of all such $n \leq x$ is bounded above by

$$\sum_{p_1 > \log x} \sum_{p_2 \in [p_1, p_1 \log_3 x)} \sum_{m \leq x} \frac{1}{p_1 p_2 m} \ll \log x \sum_{p_1 > \log x} \frac{1}{p_1} \cdot \frac{\log_4 x}{\log p_1} \ll \frac{\log x \log_4 x}{\log_2 x},$$

which is $o(\log x)$.

The n not discarded so far are precisely those that can be uniquely written as ApB, where $(A,B) \in \mathcal{L}_w$ for some $w \in W$ and p is a prime belonging to $I_{A,B} := [P^+(A), \min\{P^-(B), x/AB\}]$. Therefore,

$$\begin{aligned}\sum_{w\in W}\sum_{(A,B)\in\mathcal{L}_w}\frac{1}{AB}\sum_{p\in I_{A,B}}\frac{1}{p} &= \sum_{n\le x}\frac{1}{n} + o(\log x)\\ &= (1+o(1))\log x. \end{aligned}\tag{3.3}$$

By the hypothesis (1.1) we have uniformly for $w \in W$ and $(A,B) \in \mathcal{L}_w$ that

$$\sum_{\substack{p\in I_{A,B}\\ p\in\mathcal{P}}}\frac{1}{p} \sim \nu \sum_{p\in I_{A,B}}\frac{1}{p}.$$

Summing on $w \in W$ and $(A,B) \in \mathcal{L}_w$, and comparing with (3.3),

$$\sum_{w\in W}\sum_{(A,B)\in\mathcal{L}_w}\frac{1}{AB}\sum_{\substack{p\in I_{A,B}\\ p\in\mathcal{P}}}\frac{1}{p} \sim \nu\log x.$$

But the left-hand triple sum is, up to an error of $o(\log x)$, the reciprocal sum of $n \le x$ whose α-prefixed prime factor is in $\mathcal{P}$. Proposition 3.2 follows. □

To deduce Theorem 1.1 from Proposition 3.2, we adapt a method of Erdős (compare with the proof of [9, Theorem I]).

Below, we use $v = v(x)$ for a function of x that tends to infinity with x but at a sufficiently slow rate (to be specified precisely below). To start with, we assume that

$$v \le \log_2 x.$$

We set

$$z = x^{1/v}.$$

Let $\mathcal{D}$ be the set of $n \le x$ whose α-positioned prime factor is in $\mathcal{P}$. Fixing $\epsilon \in (0,1)$, we let $\mathcal{D}'$ be the subset of $\mathcal{D}$ containing those $n \in \mathcal{D}$ with

$$(1-\epsilon)\log v < \Omega_{>z}(n) < (1+\epsilon)\log v. \tag{3.4}$$

By Lemma 3.1, only $o(x)$ values of $n \le x$ violate (3.4), and so

$$\frac{1}{x}\#\mathcal{D} = \frac{1}{x}\#\mathcal{D}' + o(1). \tag{3.5}$$

Each $n \in \mathcal{D}'$ has between $(1-\epsilon)\log v$ and $(1+\epsilon)\log v$ representations in the form $n = mp$, where $p > z$, and so if we define

$$N = \#\{(m,p) : m \in \mathbb{N}, mp \in \mathcal{D}', p > z\},$$

then

$$(1-\epsilon)\log v \cdot \#\mathcal{D}' \le N \le (1+\epsilon)\log v \cdot \#\mathcal{D}'. \tag{3.6}$$

We will show that, as long as v is chosen to be suitably slow growing, $N \sim \nu x \log v$. The desired asymptotic relation $\#\mathcal{D} \sim \nu x$ then follows from (3.5) and (3.6), since ϵ can be taken arbitrarily close to 0.

To estimate N, we fix m and count corresponding values of p. Clearly we need only consider $m \le x/z$ for which

(i) $(1-\epsilon)\log v < \Omega_{>z}(m) + 1 < (1+\epsilon)\log v$.

We will impose two further assumptions on m which in total exclude only $o(x \log v)$ pairs (m, p). For the estimates below, it will be helpful below to note that the number of p corresponding to a given m is always bounded above by $\pi(x/m) \ll \frac{x}{m \log(x/m)} \ll \frac{xv}{m \log x}$.

We assume that

(ii) $\Omega(m) > \frac{1}{2} \log\log x$.

By Lemma 3.1, the number of pairs (m, p) excluded in this way is

$$\ll \frac{xv}{\log x} \sum_{\substack{m \le x \\ \Omega(m) \le \frac{1}{2}\log\log x}} \frac{1}{m} \ll \frac{xv}{(\log x)^{Q(1/2)}};$$

this is $o(x)$ and so in particular $o(x \log v)$. We also assume that

(iii) the α-prefixed prime factor of m is at most z.

Suppose (iii) fails for m. Once x is large, (ii) guarantees there are no less than $\frac{1-\alpha}{4} \log_2 x$ (say) prime factors of m that are at least as large as the α-prefixed prime factor of m. All of these exceed z and so also exceed $x^{1/\log_2 x}$. We apply Lemma 3.1 with E the set of primes exceeding $x^{1/\log_2 x}$. Then $E(x) = \sum_{\ell \le x,\ \ell \in E} 1/\ell = \log_3 x + O(1)$. Using a $'$ to denote a sum on m violating (iii), and noting that $\frac{1-\alpha}{4} \log_2 x > 10E(x)$, Lemma 3.1 shows that the number of exceptional pairs (m, p) of N introduced by (iii) is

$$\ll \frac{xv}{\log x} {\sum}' \frac{1}{m} \ll xv \exp(-(\log_3 x) \cdot Q(10)) \ll x.$$

(In the last step, we use that $Q(10) > 1$ and that $v \le \log_2 x$.) This is of course $o(x \log v)$.

Since (iii) holds for m, if there is any $p > z$ for which $mp \in \mathcal{D}'$, then

(iv) the α-prefixed prime factor of m is in $\mathcal{P}$.

(We use here our earlier observation relating the α-prefixed prime factor of n and the α-positioned prime factor of $n\ell$.) On the other hand, if all of (i)–(iv) hold for a positive integer $m \le x/z$, then $mp \in \mathcal{D}'$ for all $p \in (z, x/m]$. Thus, up to an error of $o(x \log v)$, the target quantity N is given by

$$\sum_{\substack{m \le x/z \\ \text{(i)–(iv) hold}}} (\pi(x/m) - \pi(z)).$$

We can go a bit further. Since $\sum_{m \le x/z} \pi(z) < (x/z) \cdot z = x$, we in fact have

$$N = \sum_{\substack{m \le x/z \\ \text{(i)–(iv) hold}}} \pi(x/m) + o(x \log v). \tag{3.7}$$

The final sum on m will be handled via the prime number theorem conjoined with the estimate

$$S(y) := \sum_{\substack{m \le y \\ \text{(i)–(iv) hold}}} \frac{1}{m} \sim \nu \log y, \tag{3.8}$$

which we assert holds uniformly for $x^{1/2} \le y \le x/z$, as $x \to \infty$. If only condition (iv) were imposed, the claimed estimate (3.8) would follow immediately from Proposition 3.2. But imposing (i)–(iii) changes the sum by only $o(\log y)$, by a further application of Lemma 3.1. (It is straightforward to check this for (i) and (ii). For (iii), we argue as above: exceptional m have unusually many prime factors exceeding $x^{1/\log_2 x}$.)

Each integer $m \le x/z$ belongs to an interval $I_j := (x/z^{j+1}, x/z^j]$ for some positive integer j. Let J be the largest positive integer with $z^{j+1} \le x^{1/3}$, so that $J = \frac{1}{3}\frac{\log x}{\log z} + O(1) = \frac{1}{3}v + O(1)$. We first consider the contribution to (3.7) from m that belong to I_j for some $j \le J$.

Uniformly for $j \le J$ and $m \in I_j$, we have $\pi(x/m) \sim \frac{x}{m\log(x/m)}$, and $\frac{x}{(j+1)m\log z} \le \frac{x}{m\log(x/m)} \le \frac{x}{jm\log z}$. Hence,

$$(1+o(1))\frac{x}{(j+1)\log z}S_j \le \sum_{\substack{m\in I_j\\ \text{(i)–(iv) hold}}} \pi(x/m) \le (1+o(1))\frac{x}{j\log z}S_j, \tag{3.9}$$

where

$$\begin{aligned} S_j &:= \sum_{\substack{m\in I_j\\ \text{(i)–(iv) hold}}} \frac{1}{m} \\ &= S(x/z^j) - S(x/z^{j+1}). \end{aligned}$$

Rephrasing (3.8), there is a nonnegative-valued function $\epsilon(x)$ such that $\epsilon(x) \to 0$ as $x \to \infty$ such that

$$|S(y) - \nu\log y| \le \epsilon(x)\log y$$

whenever $x^{1/2} \le y \le x$. Then for all $j \le J$,

$$|S_j - \nu\log z| \le 2\epsilon(x)\log x.$$

Keeping in mind that $z = x^{1/v}$, we enforce the condition on $v = v(x)$ that $\epsilon(x)$ decay to zero faster than $1/v$. For example, we can choose $v = \min\{\log_2 x, \epsilon(x)^{-1/2}\}$, where the first term in the minimum is taken if $\epsilon(x) = 0$. With this choice of v, the above guarantees that

$$S_j \sim \nu\log z,$$

uniformly for positive integers $j \le J$. Plugging this back into (3.9) and summing on j, we deduce that

$$\sum_{\substack{m\in\cup_{j\le J}I_j\\ \text{(i)–(iv) hold}}} \pi(x/m) \sim \nu x\log v.$$

Finally, if $m \le x/z$ but $m \notin \cup_{j\le J}I_j$, then $m \le z\cdot x^{2/3} < x^{3/4}$ say. But $\sum_{m\le x^{3/4}}\pi(x/m) \ll \frac{x}{\log x}\sum_{m\le x^{3/4}} 1/m \ll x$, and so these m contribute $o(x\log v)$. Hence, $N = \sum_{m\le x/z,\ \text{(i)–(iv) hold}}\pi(x/m) + o(x\log v) = (\nu+o(1))x\log v$, as desired.

4. Behavior of the largest prime divisor not exceeding y: Proof of Theorem 1.3

The cases of Theorem 1.3 where $y = x^{o(1)}$ are simplest and we treat them first. We use the following special case of the Fundamental Lemma of the Sieve (see, for instance, Theorem 2.5 on p. 82 of [10]).

Lemma 4.1. *Let $x \to \infty$, and let $y = x^{o(1)}$. Uniformly across sets of primes $E \subset [2, y]$, we have*

$$\#\left\{n \le x : \gcd(n, \prod_{p\in E} p) = 1\right\} \sim x\prod_{p\in E}\left(1 - \frac{1}{p}\right).$$

Proof of Theorem 1.3 when $y = x^{o(1)}$. We begin by discarding those $n \le x$ with $P_{\le y}(n) \le y^{1/\log_2 y}$. These are precisely the n with no prime factor in $(y^{1/\log_2 y}, y]$, and by Lemma 4.1 the number of these is $\sim x\prod_{y^{1/\log_2 y}<\ell\le y}(1-1/\ell)$, which is $o(x)$. We now count $n \le x$ with $P_{\le y}(m) = p$ for a given

$p \in (y^{1/\log_2 y}, y]$. Writing $n = mp$ where $m \le x/p$, and using Lemma 4.1 to count ms, we deduce that the number of such n is

$$\sim \frac{x}{p} \prod_{p<\ell\le y} \left(1 - \frac{1}{\ell}\right) \sim \frac{x}{\log y} \cdot \frac{\log p}{p}.$$

Hence, by the hypothesis (1.2) and (2.1), the number of $n \le x$ for which $P_{\le y}(n) \in \mathcal{P}$ is

$$\frac{x}{\log y} \sum_{\substack{p\in(y^{1/\log_2 y}, y] \\ p\in\mathcal{P}}} \frac{\log p}{p} + o(x) = (\nu + o(1))\, x,$$

as desired. □

To complete the proof of Theorem 1.3, it suffices to treat those cases when $y > x^\epsilon$ for a fixed $\epsilon > 0$.

It will be helpful to recall some of the fundamental 'anatomical theory' of large prime factors. Recall that $P_k(n)$ denotes the kth largest prime factor of n. We set

$$\Psi_k(X, Y) = \#\{n \le X : P_k(n) \le Y\}.$$

When $k = 1$, it is customary to omit the subscripts.

An asymptotic formula for $\Psi_k(X, Y)$, when Y is at least a small power of X, was established by Knuth and Trabb-Pardo in [12].[3]

Define functions $\rho_k(U)$, for integers $k \ge 0$ and real U, as follows:

$$\rho_k(U) = 0 \quad \text{if} \quad U \le 0 \text{ or } k = 0,$$

$$\rho_k(U) = 1 \quad \text{for} \quad 0 < U \le 1 \text{ and } k \ge 1,$$

$$\rho_k(U) = 1 - \int_1^U (\rho_k(t-1) - \rho_{k-1}(t-1)) \frac{\mathrm{d}t}{t}, \quad \text{for} \quad U > 1 \text{ and } k \ge 1. \tag{4.1}$$

Lemma 4.2 (see eq. (4.7) of [12]). *Fix $k \in \mathbb{N}$, and fix a real number $U_{\max} \ge 1$. Whenever $X, Y \to \infty$ with $\frac{\log X}{\log Y} \le U_{max}$, we have*

$$\Psi_k(X, Y) \sim X\rho_k(U), \quad \text{where } U := \frac{\log X}{\log Y}.$$

One can show (see again [12]) that the ρ_k ($k = 1, 2, 3, \dots$) are positive-valued and weakly decreasing on $(0, \infty)$, and that $\lim_{U\to\infty} \rho_k(U) = 0$ for each fixed k.

Proof of Theorem 1.3 when $y > x^\epsilon$, for $\epsilon > 0$ is fixed. Each $n \le x$ can be decomposed uniquely as $n = ApB$, where B is the y-rough part of n and $p = P_{\le y}(n)$. We fix a small $\eta > 0$ and restrict attention to n satisfying

(i) $p > x^\eta$,

(ii) $Bp \le x^{1-\eta}$,

(iii) $B \le x^{1-3\eta}$.

[3]The case $k = 1$ was studied prior to [12]; see [15] for the history.

Call $n \leq x$ satisfactory if (i)–(iii) hold.

Since $y > x^\epsilon$, each $n \leq x$ has $\Omega(B) < 1/\epsilon$. Letting $k = \lfloor 1/\epsilon \rfloor + 3$, we see that if n is not satisfactory, then $P_k(n) \leq x^\eta$. Hence, the number of nonsatisfactory $n \leq x$ is at most $(\rho_k(1/\eta) + o(1))x$, as $x \to \infty$. Since the coefficient $\rho_k(1/\eta)$ tends to 0 as $\eta \downarrow 0$, it will suffice to prove that (for each small, fixed $\eta > 0$)

$$\#\{\text{satisfactory } n \leq x : P_{\leq y}(n) \in \mathcal{P}\} \sim \nu \#\{\text{satisfactory } n \leq x\}, \quad \text{as } x \to \infty. \tag{4.2}$$

Say that n is B-satisfactory if n is satisfactory and has B as its y-rough part. We will prove (4.2) by showing the analogous asymptotic formula for B-satisfactory n and then summing on B.

Fix a y-rough number $B \leq x^{1-3\eta}$. In order for a prime p to be $P_{\leq y}(n)$ for some B-satisfactory number n, it must be that

$$x^\eta < p \leq \min\{x^{1-\eta}/B, y\}. \tag{4.3}$$

Conversely, if (4.3) holds, then for every p-smooth value of $A \leq x/pB$, the number ApB is a B-satisfactory number in $[1, x]$ having $p = P_{\leq y}(n)$.

We now split the range (4.3) into more manageable pieces. Define δ_0 by $\min\{x^{1-\eta}/B, y\}/x^\eta = x^{\delta_0}$. Since we may assume $\eta < \epsilon/2$, we have $1 > \delta_0 \geq \eta$. We let

$$N = \lfloor \log_2 x \rfloor,$$

we let $\delta = \delta_0/N$. Putting $I_k = (x^{\eta+(k-1)\delta}, x^{\eta+k\delta}]$, we then have

$$\bigcup_{1 \leq k \leq N} I_k = (x^\eta, \min\{x^{1-\eta}/B, y\}].$$

With an imminent application of (1.1) in mind, we note that the "multiplicative width" of each I_k is $x^{\delta_0/N} \geq x^{\eta/N}$, and that $x^{\eta/N}$ tends to infinity with x.

For each $p \in I_k$, the number of B-satisfactory n with $P_{\leq y}(n) = p$ is

$$\Psi\left(\frac{x}{Bp}, p\right) \sim \frac{x}{Bp} \rho\left(\frac{\log(x/Bp)}{\log p}\right),$$

uniformly in B and p. Here every argument of ρ is bounded: $\frac{\log(x/Bp)}{\log p} \leq 1/\eta$ for p belonging to any I_k. We now exploit that the ρ-term is essentially constant within I_k. Let π_k be the left endpoint of I_k. Our choice of N guarantees that $\frac{\log(x/Bp)}{\log p}$ is within $O(1/\log_2 x)$ of $\frac{\log(x/B\pi_k)}{\log \pi_k}$ for every $p \in I_k$. Since $\rho(u) \gg 1$ for u from the interval $(0, 1/\eta]$ and $\rho(u)$ is uniformly continuous on that same interval[4], we conclude that

$$\rho\left(\frac{\log(x/Bp)}{\log p}\right) \sim \rho\left(\frac{\log(x/B\pi_k)}{\log \pi_k}\right),$$

[4]for example, since ρ is continuous on $[0, 1/\eta]$ if we redefine $\rho(0) = 1$

uniformly for $1 \le k \le N$ and $p \in I_k$. Setting $\theta_k = \frac{x}{B}\rho\left(\frac{\log(x/B\pi_k)}{\log \pi_k}\right)$, it follows that the number of B-satisfactory $n \le x$ for which $p = P_{\le y}(n)$ is $\sim \theta_k/p$, uniformly for $p \in I_k$. But then by (1.1),[5]

$$\begin{aligned}\#\{B\text{-satisfactory } n \le x : P_{\le y}(n) \in \mathcal{P}\} &\sim \sum_{k=1}^{N} \theta_k \sum_{\substack{p \in I_k\\ p \in \mathcal{P}}} \frac{1}{p}\\ &\sim \nu \sum_{k=1}^{N} \theta_k \sum_{p \in I_k} \frac{1}{p}\\ &\sim \nu \#\{B\text{-satisfactory } n \le x\}.\end{aligned}$$

Since this relation holds uniformly in B, summing on B gives (4.2). $\square$

5. The α-positioned prime divisor in arithmetic progressions: Proof of Theorem 1.5

Again the proof requires some preparation. We start by providing a uniform upper bound on the sum of reciprocals of smooth numbers with a given number of prime divisors.

Lemma 5.1. *Uniformly in real numbers $z \ge 3$ and integers $J \ge 1$,*

$$\sum_{\substack{A:\ P^+(A) \le z\\ \Omega(A) = J}} \frac{1}{A} \ll \frac{J}{2^J} (\log z)^2. \tag{5.1}$$

Proof. For any $0 < y < 2$ we have

$$\begin{aligned}\sum_{\substack{A:\ P^+(A) \le z\\ \Omega(A) = J}} \frac{1}{A} < y^{-J} \sum_{A:\ P^+(A) \le z} \frac{y^{\Omega(A)}}{A} &= y^{-J} \left(1 - \frac{y}{2}\right)^{-1} \prod_{3 \le p \le z} \left(1 - \frac{y}{p}\right)^{-1}\\ &\ll \frac{y^{-J}}{2-y} \exp\left(y \sum_{3 \le p \le z} \frac{1}{p}\right) \ll \frac{y^{-J}}{2-y} (\log z)^y.\end{aligned}$$

Letting $y = 2 - 1/J$ and noting that $(1 - 1/2J)^{-J} \asymp 1$, we obtain the desired estimate. $\square$

We also need the following consequence of the Sathe–Selberg theorem concerning the distribution of numbers with a given number of prime factors (see, e.g., [20, Theorem 6.5, p.304]).

Proposition 5.2. *Fix $\delta \in (0, 1)$. For all sufficiently large values of x,*

$$\sum_{\substack{n \le x\\ \Omega(n) = k}} 1 \asymp \frac{x}{\log x} \frac{(\log_2 x)^{k-1}}{(k-1)!}$$

uniformly for positive integers $1 \le k \le (2 - \delta) \log_2 x$.

It is also helpful to have at hand a simple upper bound on the count $\Psi(X, Y)$ of Y-smooth numbers not exceeding X; the following appears as Theorem 5.1 on p. 512 of [20].

[5] Recall that (1.1) follows from (1.2), as discussed in §2.

Proposition 5.3. *For all $X \ge Y \ge 2$,*

$$\Psi(X,Y) \ll X\exp(-U/2),$$

where $U \coloneqq \log X/\log Y$.

Proof of Theorem 1.5. We claim that with $z = \exp((\log x)^{\epsilon/3})$, the $n \le x$ with $P^{(\alpha)}(n) \le z$ give a negligible contribution to our count, meaning a count of size $o(x/\varphi(q))$. More precisely,

$$\sum_{\substack{n\le x\\ P^{(\alpha)}(n)\le z}} 1 \ll \frac{x}{(\log x)^{c(\alpha)-3\epsilon/4}} \tag{5.2}$$

where $c(\alpha) = 1-2^{-\alpha/(1-\alpha)}$. (Recall that $q \le (\log x)^{c(\alpha)-\epsilon}$.)

To show (5.2), we first discard all $n \le x$ for which $\Omega(n) > k(\alpha)\log_2 x$ where $k(\alpha) \coloneqq 1.9/(1-\alpha)$.

An application of (3.2) with $y = 1.7$ (and E the full set of primes in $[2,x]$) yields

$$\sum_{\substack{n\le x\\ \Omega(n)>k(\alpha)\log_2 x}} 1 \ll \frac{x(\log x)^{0.7}}{(1.7)^{k(\alpha)\log_2 x}} = \frac{x}{(\log x)^{k(\alpha)\log(1.7)-0.7}} \ll \frac{x}{(\log x)^{c(\alpha)}},$$

where in the last step we use that $k(\alpha)\log(1.7)-0.7-c(\alpha) > 0$ for all $\alpha\in(0,1)$. (To check that inequality, it helps to first show that $k(\alpha)\log(1.7)-0.7-c(\alpha)$ is an increasing function of α on $(0,1)$.)

So for the sake of proving (5.2), we can restrict to n with $\Omega(n) \le k(\alpha)\log_2 x$. In what follows, let $J \coloneqq \lceil \alpha\Omega(n)\rceil$ so that $P^{(\alpha)}(n) = p_J(n)$. If $J = \Omega(n)$, then n is z-smooth and by Proposition 5.3, the count of such $n \le x$ is $O(x\exp(-\frac{1}{2}(\log x)^{1-\epsilon/3}))$, which is negligible in comparison to the upper bound claimed in (5.2).

We now suppose that $J < \Omega(n)$, or equivalently that $\Omega(n) \ge 1/(1-\alpha)$. For each positive integer w satisfying $1/(1-\alpha) \le w \le k(\alpha)\log_2 x$, we factor those $n \le x$ having $\Omega(n) = w$ and $P^{(\alpha)}(n) \le z$ as $n = AB$, where $A = \prod_{j\le J} p_j(n)$ and $B = \prod_{J<j\le\Omega(n)} p_j(n)$. We then count values of B given A. Observe that $\Omega(A) = J$ while $P^+(A) = p_J(n) = P^{(\alpha)}(n) \le z$, so that $A \le z^J < \exp((\log x)^{\epsilon})$. Moreover, $\Omega(B) = w-J \le (1-\alpha)w \le 1.9\log_2 x$. By Proposition 5.2, the number of B corresponding to a given A is

$$\ll \frac{x/A}{\log(x/A)}\cdot\frac{(\log_2(x/A))^{w-J-1}}{(w-J-1)!} \ll \frac{x}{A\log x}\cdot\frac{(\log_2 x)^{w-J-1}}{(w-J-1)!}. \tag{5.3}$$

Summing (5.3) over z-smooth A with $\Omega(A) = J$, we deduce from Lemma 5.1 that for each $w \in [1/(1-\alpha), k(\alpha)\log_2 x]$,

$$\sum_{\substack{n\le x\\ \Omega(n)=w\\ P^{(\alpha)}(n)\le z}} 1 \ll \frac{x}{(\log x)^{1-2\epsilon/3}}\cdot\frac{J}{2^J}\cdot\frac{(\log_2 x)^{w-J-1}}{(w-J-1)!} \ll \frac{x}{(\log x)^{1-3\epsilon/4}}\cdot\frac{(\log_2 x)^{w-J-1}}{2^J(w-J-1)!}$$

where (recall) $J = \lceil\alpha w\rceil$.

We now sum on w. Write $v = w-J-1 = w-\lceil\alpha w\rceil-1$. Then v is a nonnegative integer. Furthermore, as $v = (1-\alpha)w+O(1)$, each v arises from only $O(1)$ choices of w and $2^J = 2^{\lceil\alpha w\rceil} \asymp 2^{\frac{\alpha}{1-\alpha}(1-\alpha)w} \asymp 2^{\frac{\alpha}{1-\alpha}v}$. Hence,

$$\sum_{\frac{1}{1-\alpha}\le w\le k(\alpha)\log_2 x} \frac{x}{(\log x)^{1-3\epsilon/4}}\cdot\frac{(\log_2 x)^{w-J-1}}{2^J(w-J-1)!} \ll \frac{x}{(\log x)^{1-3\epsilon/4}}\sum_{v\ge0}\frac{1}{v!}\left(\frac{\log_2 x}{2^{\alpha/(1-\alpha)}}\right)^v = \frac{x}{(\log x)^{c(\alpha)-3\epsilon/4}},$$

finishing the proof of (5.2).

To complete the proof of Theorem 1.5, it will therefore suffice to show that

$$\sum_{\substack{n\le x\\ P^{(\alpha)}(n)>z\\ P^{(\alpha)}(n)\equiv a \ (\mathrm{mod}\ q)}} 1 \sim \frac{x}{\varphi(q)} \tag{5.4}$$

as $x\to\infty$, uniformly for $q\le(\log x)^{c(\alpha)-\epsilon}$. Here we adapt the anatomical splitting considered in the proof of Proposition 3.2: We decompose each n counted on the left of (5.4) as ApB, where $A=\prod_{j<J}p_j(n)$, $p=p_J(n)$, and $B=\prod_{J<j\le\Omega(n)}p_j(n)$, with $J=\lceil\alpha\Omega(n)\rceil$. Then $P^+(A)\le p\le P^-(B)$ and $z<p\le x/AB$, so that given A and B the count of p is

$$\sum_{\substack{m<p\le M\\ p\equiv a \ (\mathrm{mod}\ q)}} 1+O(1)$$

where $m:=\max\{z,P^+(A)\}$ and $M:=\min\{P^-(B),x/AB\}$. Recalling that $z=\exp((\log x)^{\epsilon/3})$ (so that $q\le(\log z)^{O(1)}$), we derive from the Siegel-Walfisz Theorem that the sum in the last display has size

$$\frac{1}{\varphi(q)}\sum_{m<p\le M}1+O\left(\frac{x}{AB}\exp\left(-K(\log x)^{\epsilon/6}\right)\right).$$

(Here K is a positive absolute constant.) Finally we sum this last expression over all possible A,B. The O-term gives us a negligible error. In the main term, we pick up $\frac{1}{\varphi(q)}$ multiplied by the count of $n\le x$ with $P^{(\alpha)}(n)>z$, and this is $\sim x/\varphi(q)$ (using again (5.2)). □

Remark (optimality of the range of uniformity). A modification of the argument used to establish (5.2) shows that

$$\sum_{\substack{n\le x\\ P^{(\alpha)}(n)=2}} 1 \gg \frac{x}{(\log x)^{c(\alpha)}} \tag{5.5}$$

which immediately implies that uniformity in the modulus q in Theorem 1.5 fails as soon as $q\ge(\log x)^{c(\alpha)+\epsilon}$ for any fixed $\epsilon>0$.

In order to show (5.5), we note that any positive integer $n\le x$ with $\Omega(n)=w$ and $P^{(\alpha)}(n)=2$ can be uniquely written as $2^{\lceil\alpha w\rceil}m$ for some positive integer $m\le x/2^{\lceil\alpha w\rceil}$ having $\Omega(m)=w-\lceil\alpha w\rceil$. Hence, by Proposition 5.2, the sum in (5.5) is

$$\gg \sum_{\frac{1}{1-\alpha}\le w\le k(\alpha)\log_2 x}\ \sum_{\substack{m\le x/2^{\lceil\alpha w\rceil}\\ \Omega(m)=w-\lceil\alpha w\rceil}} 1 \gg \frac{x}{\log x}\sum_{\frac{1}{1-\alpha}<w\le k(\alpha)\log_2 x}\frac{1}{2^{\lceil\alpha w\rceil}}\frac{(\log_2 x)^{w-1-\lceil\alpha w\rceil}}{(w-1-\lceil\alpha w\rceil)!}$$

where $k(\alpha)=1.9/(1-\alpha)$ as above. For each positive integer v satisfying $0\le v\le 1.9\log_2 x-2$, there is at least one positive integer w with $w-\lceil\alpha w\rceil-1=v$. (An easy way to see this uses that $v\to\infty$ as $w\to\infty$, while increasing w by 1 either leaves v constant or increases v by 1.) Furthermore, any such w lies in the interval $[1/(1-\alpha),k(\alpha)\log_2 x]$. Keeping in mind our previous estimate $2^{\lceil\alpha w\rceil}\asymp 2^{\frac{\alpha}{1-\alpha}v}$, the last expression in the previous display is seen to be

$$\gg \frac{x}{\log x}\sum_{0\le v\le 1.9\log_2 x-2}\frac{1}{v!}\left(\frac{\log_2 x}{2^{\alpha/(1-\alpha)}}\right)^v \ge \frac{x}{\log x}\sum_{0\le v\le\frac{\log_2 x}{2^{\alpha/(1-\alpha)}}}\frac{1}{v!}\left(\frac{\log_2 x}{2^{\alpha/(1-\alpha)}}\right)^v.$$

Finally, since $\sum_{0\le u\le X} X^u/u! \gg e^X$ as $X\to\infty$, the last sum here is $\gg x/(\log x)^{c(\alpha)}$, yielding (5.5).

Remark (restriction to squarefree inputs). Since $P^{(\alpha)}(n)$ is defined via the count of prime factors *with multiplicity*, it is reasonable to wonder how important nonsquarefree numbers are to Theorem 1.1 (and Corollary 1.2) and Theorem 1.5. For Theorem 1.1 (and Corollary 1.2), nothing essential in the statement or the proof changes upon restricting to squarefree numbers. (Of course the statements should now refer to "asymptotic density relative to the set of squarefrees.")

The situation is different for Theorem 1.5. Here one finds that, for any fixed $\alpha\in(0,1)$, uniform distribution holds for all moduli $q\le(\log x)^{1-\epsilon}$. The crux of the argument is an analogue of the estimate (5.2) which states that for an appropriate $\delta>0$ depending only on ϵ and α, the number of squarefree $n\le x$ with $P^{(\alpha)}(n)\le z:=\exp((\log x)^\delta)$ is $O(x/(\log x)^{1-\epsilon/2})$. This can be established by considering the same splitting $n=AB$ with $A=\prod_{j\le J}p_j(n)$ and $B=\prod_{J<j\le\Omega(n)}p_j(n)$ as in the proof of (5.2), counting values of B by the Hardy-Ramanujan Theorem (instead of Proposition 5.2) and then bounding the sum of reciprocals of A by $\frac{1}{J!}(\sum_{p\le z}1/p)^J$, both of which are validated by the squarefreeness of A and B. Moreover, the range $q\le(\log x)^{1-\epsilon}$ is best possible in that the exponent $1-\epsilon$ cannot be replaced with a number larger than 1: Indeed, given $\alpha\in(0,1)$, fix integers $a,b\ge1$ such that $a-1<\alpha b\le a$. Then, with p_j denoting the jth prime in the usual order, there are $\gg x/\log x$ many squarefree integers n of the form $p_1\cdots p_{b-1}p$ having $p_{\lceil\alpha\Omega(n)\rceil}=p_{\lceil\alpha b\rceil}=p_a$, as p varies over the primes in $(p_{b-1},x/p_1\cdots p_{b-1})$.

6. Concluding remarks

Our proof of Theorem 1.1 was somewhat roundabout. A more direct plan of attack might involve first finding an asymptotic formula for the count of $n\le x$ with $P^{(\alpha)}(n)=p$, uniformly in a wide range of p. One could then hope to recover Theorem 1.1 by summing on $p\in\mathcal{P}$. This strategy can indeed be made to work. To give some sense of what comes out of this approach, fix $\alpha=1/2$. (This is done only for simplicity of exposition; the method applies for each fixed $\alpha\in(0,1)$.) Assume $\beta:=\frac{\log_2 p}{\log_2 x}$ satisfies $0.201\le\beta\le0.999$. Then (as $x\to\infty$), the number of $n\le x$ with $P^{(1/2)}(n)=p$ is

$$\sim C_\beta\frac{x}{p(\log x)^{1-2\sqrt{\beta(1-\beta)}}\sqrt{\log_2 x}}, \tag{6.1}$$

where (with γ the usual Euler–Mascheroni constant)

$$C_\beta=\frac{\exp\left(\frac{\gamma(1-2\beta)}{\sqrt{\beta(1-\beta)}}\right)}{\Gamma\left(1+\sqrt{\frac{\beta}{1-\beta}}\right)}\frac{\sqrt{\beta}+\sqrt{1-\beta}}{2\sqrt{\pi}\beta^{1/4}(1-\beta)^{3/4}}\prod_p\left(1-\frac{1}{p}\right)^{\sqrt{\frac{1-\beta}{\beta}}}\left(1-\frac{\sqrt{\frac{1-\beta}{\beta}}}{p}\right)^{-1}.$$

Despite its complicated-seeming shape, once (6.1) is known it is not so hard to rederive the $\alpha=1/2$ case of Theorem 1.1. One first argues that it is enough to consider $p\in\mathcal{P}$ having $\beta\sim1/2$; note that for these primes, $C_\beta\sim C_{1/2}=\sqrt{2/\pi}$. One then places p into intervals on which $1/(\log x)^{1-2\sqrt{\beta(1-\beta)}}\sqrt{\log_2 x}$ is essentially constant and compares the sum over $p\in\mathcal{P}$ with the unrestricted sum on p. A similar idea was used above in the $y>x^\epsilon$ half of the proof of Theorem 1.3.

Formula (6.1), and its generalizations, are established by a very different set of tools than those employed here. We intend to present the proofs of these results, along with a discussion of applications, in a sequel paper.

References

1. Anderson, T.C., Rolen, L., Stoehr, R.: Benford's law for coefficients of modular forms and partition functions. Proc. Amer. Math. Soc. **139**, 1533–1541 (2011)
2. De Koninck, J.-M., Doyon, N., Ouellet, V.: The limit distribution of the middle prime factors of an integer. Integers **19** #A56 (2019)
3. De Koninck, J.-M., Kátai, I.: Normal numbers and the middle prime factor of an integer. Colloq. Math. **135**, 69–77 (2014)
4. ______: On key applications of the Turán-Kubilius inequality. Lith. Math. J. **61**, 312–322 (2021)
5. De Koninck, J.-M., Luca, F.: On the middle prime factor of an integer. J. Integer Seq. **16**, no. 5 (2013)
6. Diaconis, P.: The distribution of leading digits and uniform distribution mod 1. Ann. Probability **5**, 72–81 (1977)
7. Doyon, N., Ouellet, V.: On the sum of the reciprocals of the middle prime factor counting multiplicity. Ann. Univ. Sci. Budapest. Sect. Comput. **47**, 249–259 (2018)
8. Ellison, W., Ellison, F.: Prime numbers. A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York; Hermann, Paris (1992)
9. Erdős, P.: On the integers having exactly k prime factors. Ann. of Math. (2) **49**, 53–66 (1948)
10. Halberstam, H., Richert, H.-E.: Sieve methods. Academic Press, London-New York (1974)
11. Hall, R.R., Tenenbaum, G.: Divisors, Volume 90 of Cambridge Tracts in Mathematics. Cambridge University Press, Cambridge (1988)
12. Knuth, D.E., Trabb Pardo, L.: Analysis of a simple factorization algorithm. Theoret. Comput. Sci. **3**, 321–348 (1976/77)
13. Massé, B., Schneider, D.: Fast growing sequences of numbers and the first digit phenomenon. Int. J. Number Theory **11**, 705–719 (2015)
14. Miller, S.J. (ed.): Benford's Law: theory and applications. Princeton University Press, Princeton, NJ (2015)
15. Norton, K.K.: Numbers with small prime factors, and the least kth power non-residue, Volume 106 of the Memoirs of the American Mathematical Society. American Mathematical Society, Providence, RI (1971)
16. ______: On the number of restricted prime factors of an integer. I. Illinois J. Math. **20**, 681–705 (1976)
17. Pollack, P., Singha Roy, A.: Benford behavior and distribution in residue classes of large prime factors. Canad. Math. Bull. (to appear)
18. ______: Dirichlet, Sierpiński, and Benford. J. Number Theory **239**, 352–364 (2022)
19. Serre, J.-P.: A course in arithmetic, Volume 7 of Graduate Texts in Mathematics. Springer-Verlag, New York-Heidelberg (1973)
20. Tenenbaum, G.: Introduction to analytic and probabilistic number theory, third ed, Volume 163 of Graduate Studies in Mathematics. American Mathematical Society, Providence, RI (2015)
21. Whitney, R.E.: Initial digits for the sequence of primes. Amer. Math. Monthly **79**, 150–152 (1972)

Email address: `nmcnew@towson.edu`

Department of Mathematics, Towson University, Towson, MD 21252

Email address: `pollack@uga.edu`

Email address: `akash01s.roy@gmail.com`

Department of Mathematics, University of Georgia, Athens, GA 30602