Canad. Math. Bull. 2026, pp. 1–17
http://dx.doi.org/10.4153/S0008439526101982
© The Author(s), 2026. Published by Cambridge University Press on behalf of Canadian Mathematical Society. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Large prime factors of well-distributed sequences

Abhishek Bharadwaj and Brad Rodgers

Abstract. We study the distribution of large prime factors of a random element u of arithmetic sequences satisfying simple regularity and equidistribution properties. We show that if such an arithmetic sequence has level of distribution 1, the large prime factors of u tend to a Poisson–Dirichlet process, while if the sequence has any positive level of distribution the correlation functions of large prime factors tend to a Poisson–Dirichlet process against test functions of restricted support. For sequences with positive level of distribution, we also estimate the probability that the largest prime factor of u is greater than $u^{1-\varepsilon}$, showing that this probability is $O(\varepsilon)$. Examples of sequences described include shifted primes and values of single-variable irreducible polynomials. The proofs involve (i) a characterization of the Poisson–Dirichlet process due to Arratia–Kochman–Miller and (ii) an upper bound sieve.

1 Introduction

1.1 Background

The purpose of this note is to study the distribution of large prime factors of elements in sequences which satisfy only a few minimal conditions.

Let us recall the classical theory and notation: let $P^+(u)$ be the largest prime factor of a positive integer u. It is a consequence of a result of Dickman [13] that for any fixed $c \in (0,\infty)$,

$$\frac{1}{x}\left|\{u \le x : P^+(u) \le u^c\}\right| \sim \rho(1/c),$$

as $x \to \infty$, where $\rho : (0,\infty) \to (0,1]$ is a continuous function called the Dickman function. (See, e.g., [33, Chapter 7.1] for a modern account.)

This result was generalized and given a probabilistic interpretation by Billingsley [4] (and independently Knuth and Trabb Pardo [29] and Vershik [41]). Let u be chosen randomly and uniformly from the integers from 1 to x. Then the above result says that $\log P^+(u)/\log u$ tends in distribution to a nonnegative random variable L_1 with cumulative distribution function $\mathbb{P}(L_1 \le c) = \rho(1/c)$. Moreover, let $p_1 \ge p_2 \ge \cdots$ be the prime factors of u listed with multiplicity, with the convention that $p_j = 1$ if u has

Received by the editors June 15, 2025; revised January 27, 2026; accepted April 2, 2026.
Published online on Cambridge Core April 17, 2026.
B.R. is supported by an NSERC grant. A.B. is supported by a Coleman Postdoctoral Fellowship.
AMS subject classification: **11K36**, **11N64**, **60G55**.
Keywords: large prime factors, Poisson-Dirichlet, correlation functions, shifted primes, level of distribution.

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

fewer than j prime factors (so that $p_1 = P^+(u)$ and $u = p_1p_2\dots$). Billingsley showed that there is a sequence of (dependent) random variables $L_1, L_2, \dots$ such that for any $k \geq 1$, and any fixed constants $c_1, c_2, \dots, c_k \in [0, \infty)$,

$$\mathbb{P}\Big(\frac{\log p_1}{\log u} \leq c_1, \dots, \frac{\log p_k}{\log u} \leq c_k\Big) \to \mathbb{P}(L_1 \leq c_1, \dots, L_k \leq c_k),$$

as $x \to \infty$. That is, the process $\frac{\log p_1}{\log u}, \frac{\log p_2}{\log u}, \dots$ tends in distribution to $L_1, L_2, \dots$. (See, e.g., [5] for a modern probabilistic account. Here and in what follows, we discard with the case $u = 1$ by adopting the formal convention $\log 1/\log 1 = 1$.)

For each k, an explicit formula for $\mathbb{P}(L_1 \leq c_1, \dots, L_k \leq c_k)$ can be written down (see [5, Theorem 4.4] for a density formula), but the formula is somewhat complicated. The following characterization is simpler: let $U_1, U_2, \dots$ be independent and identically distributed uniform random variables on the interval $(0,1)$, and define random variables $G_1 = 1 - U_1$, $G_2 = U_1(1 - U_2)$, $G_3 = U_1U_2(1 - U_3)$, $\dots$. Then $L_1 \geq L_2 \geq \dots$ may be defined as the outcome of sorting $G_1, G_2, \dots$ into nonincreasing order.

The sequence of random variables $L_1, L_2, \dots$ is known as the *Poisson–Dirichlet process with parameter* $\theta = 1$. As the name suggests, there are Poisson–Dirichlet processes with $\theta \neq 1$, but in this article, we will only deal with $\theta = 1$, and so if there is no risk of confusion, we refer to $L_1, L_2, \dots$ as simply the Poisson–Dirichlet process. (See [5, 28] for an account of general θ, along with a more detailed introduction to the case $\theta = 1$.)

We note that $\frac{\log p_1}{\log u} + \frac{\log p_2}{\log u} + \dots = 1$, and likewise $L_1 + L_2 + \dots = 1$ almost surely.

1.2 Some well-distributed arithmetic sequences

It is natural to wonder whether this statistical pattern governing the distribution of large prime factors of random integers extends to other arithmetic sequences. Some cases which have been studied extensively include shifted primes [2, 3, 14, 17, 18, 21, 25, 43] – that is, the sequence $\{p - a\}$ for a constant a, where p ranges over the primes – and the values of irreducible polynomials [7–12, 23, 24, 27, 32, 40] – that is, the sequence $\{F(n)\}$, where F is an irreducible polynomial with integer coefficients and a positive leading coefficient and n ranges over the integers.

In this article, we study a quite general class of arithmetic sequences. Let (a_n) be a sequence of nonnegative numbers. Define the quantities

$$N(x) = \sum_{n \leq x} a_n, \qquad N_d(x) = \sum_{\substack{n \leq x \\ n \equiv 0 \pmod d}} a_n. \tag{1}$$

We say that the sequence (a_n):

(A) Has *index* α if for some $\alpha > 0$,

$$N(x) = x^{\alpha + o(1)},$$

(B) Has a *level of distribution* ϑ if for any $0 < c < \vartheta$ and any $A > 0$,

$$\sum_{d \leq x^c} |N_d(x) - g(d)N(x)| \ll_{c,A} \frac{1}{(\log x)^A} N(x), \tag{2}$$

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

where $g(d)$ is a multiplicative function with $g(d) \in [0,1]$ for all $d \geq 1$, and

$$\sum_{p \leq x} g(p) \log p = \log x + O(1), \qquad g(d) = O(C^{\Omega(d)}/d), \tag{3}$$

for some constant $C \geq 1$.

(C) Is *congruence uniform* if there are constants $B \geq 0$, $C \geq 1$ such that

$$N_d(x) \ll \Big(\frac{C^{\Omega(d)}}{d} N(x) + C^{\Omega(d)}\Big)(\log x)^B, \quad \text{for } d \leq x.$$

Additionally, we say the sequence (a_n):

- Is σ *well-distributed* if (a_n) satisfies each of (A)–(C) and σ is any positive number less than or equal to α and ϑ.

Here, as usual, $\Omega(d)$ is the number of prime factors of d counted with multiplicity. Note that (3) implies that $g(p^k) < 1$ except for finitely many primes p.

Likewise note that we trivially have $N_d(x) \leq N(x)$, so that the condition for congruence uniformity is only meaningful for those integers d with prime factors $p > C$.

It will typically be the case that a_n is the indicator function of n belonging to some subset of the integers, and in that case, we will also describe that subset by the above terminology as long as there is no chance for confusion.

Remark 1 Let us explain some intuition for these conditions. (A) is self-explanatory; in (B), one may keep in mind the examples $g(d) = 1/d$ or $g(d) = 1/\phi(d)$; and (C) may be thought of as a more technical condition – the reader may check that it is trivially satisfied if the sequence (a_n) is bounded and $N(x) \gg x/(\log x)^B$ for some B.

Examples of sequences satisfying these conditions are described by the following propositions. Throughout the article, for a proposition $\mathfrak{A}$, we use the notation $\mathbf{1}[\mathfrak{A}]$ to be 1 if $\mathfrak{A}$ is true and 0 otherwise.

Proposition 2 *Shifted primes are* $1/2$ *well-distributed. That is, for a fixed integer* a*, consider the set* $\mathcal{B} = \{p - a : p \text{ is prime}\}$ *and let* $a_n = \mathbf{1}[n \in \mathcal{B}]$*. Then* (a_n) *has index* 1*, has level of distribution* $\vartheta = 1/2$*, and is congruence uniform.*

Proof That the sequence has index $\alpha = 1$ follows from the prime number theorem (or indeed Chebyshev's bounds). By Remark 1, congruence uniformity also follows from Chebyshev's lower bound $\pi(x) \gg x/\log x$. The level of distribution $1/2$ follows from the Bombieri–Vinogradov theorem [6, Theorem 9.2.1], with $g(d) = \mathbf{1}_{(a,d)=1}/\phi(d)$, where $\mathbf{1}_{(a,d)=1}$ is 1 when d is coprime to a and 0 otherwise. ∎

Proposition 3 *The values of irreducible polynomials of degree* D *are* $1/D$ *well-distributed. That is, for a polynomial* $F(X) \in \mathbb{Z}[X]$ *of degree* $D \geq 1$*, which is irreducible with positive leading coefficient, consider the set* $\mathcal{C} = \{F(n) : n \in \mathbb{N}_{\geq 1}\} \cap \mathbb{N}_{\geq 1}$ *and let* $a_n = \mathbf{1}[n \in \mathcal{C}]$*. Then* (a_n) *has index* $1/D$*, has level of distribution* $\vartheta = 1/D$*, and is congruence uniform.*

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

Proof That the index is $1/D$ follows from the fact that $N(x) \asymp x^{1/D}$.

We see that the level of distribution is $1/D$ in the following way. For a natural number d, let $h(d)$ be the number of distinct roots of F modulo d. By the Chinese Remainder Theorem (see [36, Theorem 46]), we note that h is multiplicative. Set $g(d) = h(d)/d$ and note g is multiplicative also.

Obviously $g(d) \in [0,1]$. We will show that (2) and (3) are satisfied for this function g. Let us show that (3) holds first. It is known that $\sum_{p \le x} g(p) \log p = \log x + O(1)$; this is just a weak claim that F will on average have one root modulo a prime (see, e.g., [35, Equation (4), p. 352] or [30, Corollary 4] for a more modern statement).

The second claim in (3) is just the claim that $h(d) \ll C^{\Omega(d)}$ for some constant $C > 1$. This follows from the multiplicativity of h and [36, Theorem 54] by taking $C = D \cdot \mathrm{disc}(F)^2$.

Now, let us now prove (2). We first establish that

$$|N_d(x) - g(d)N(x)| \ll h(d), \tag{4}$$

where the implicit constant depends only on F.

To see this, note that since $F(n+d) \equiv F(n) \bmod d$, in every interval of length d, we will have $F(n) \equiv 0 \pmod d$ for $h(d)$ values of n. In particular, since F is eventually increasing, there is some sufficiently large x_0 (depending on F) such that whenever $N(x)$ increases by d on an interval to the right of x_0, $N_d(x)$ will increase by $h(d)$ on the same interval. Therefore,

$$|N_d(x) - g(d)N(x)| \le h(d) + |N_d(x_0) - g(d)N(x_0)| = h(d) + O(1)$$

with the $O(1)$ term present to account for behavior of this quantity for $x \le x_0$.

If $h(d) \ge 1$, this verifies (4). On the other hand, if $h(d) = 0$, we have $N_d(x) = g(d)N(x) = 0$ for all x, so (4) is verified in this case also.

Hence, from (4),

$$\sum_{d \le x^c} |N_d(x) - g(d)N(x)| \ll \sum_{d \le x^c} h(d) \ll_c x^c (\log x)^A, \tag{5}$$

where we use Lemma 4, proved below, in the last estimate. As $N(x) \asymp x^{1/D}$, we see (2) is satisfied as long as $c < 1/D$.

To prove congruence uniformity, we use the bound $N_d(x) = g(d)N(x) + O(h(d))$. Consequently, for a constant $C > 1$ as above, we see that $N_d(x) \ll \frac{C^{\Omega(d)}}{d} N(x) + C^{\Omega(d)}$, for all $d \le x$. ∎

We have used one of the following results above and we will need them later as well.

Lemma 4 *Suppose $h(d)$ is a multiplicative function and $g(d) = h(d)/d$ satisfies $g(d) \in [0,1]$ for all d and $g(d) = O(C^{\Omega(d)}/d)$ for some constant $C \ge 1$. Then for some constant $A > 0$,*

$$\sum_{n \le x} g(n) \ll (\log x)^A$$

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

and

$$\sum_{n\le x} h(n) \ll x(\log x)^A.$$

Proof Let $p_1, \dots, p_\ell$ be the finite set of primes less than $2C$, and set $P = p_1 \dots p_\ell$. Then

$$\sum_{\substack{n\le x\\(n,P)=1}} g(n) \ll \sum_{\substack{n\le x\\(n,P)=1}} C^{\Omega(n)}/n \le \prod_{\substack{p\le x\\(p,P)=1}} \Big(1 + \frac{C}{p} + \frac{C^2}{p^2} + \cdots\Big)$$
$$= \exp\Big(\sum_{p\le x} \frac{C}{p} + O(1)\Big) \ll (\log x)^C,$$

where we have used the fact that C/p in the product above is no more than $1/2$ in order to sum the series. Using multiplicativity and $g(p_i^{e_i}) \in [0,1]$ for $1 \le i \le \ell$, we have from a crude bound

$$\sum_{n\le x} g(n) \le \sum_{p_1^{e_1}\dots p_\ell^{e_\ell}\le x} \sum_{\substack{m\le x/p_1^{e_1}\dots p_\ell^{e_\ell}\\(m,P)=1}} g(m)$$
$$\ll (\log x)^C \Big(\sum_{e_1:\, p_1^{e_1}\le x} 1\Big)\dots\Big(\sum_{e_\ell:\, p_\ell^{e_\ell}\le x} 1\Big) \ll (\log x)^{C+\ell}.$$

This proves the first estimate. For the second, we have

$$\sum_{n\le x} h(n) \le x \sum_{n\le x} g(n),$$

which implies the claim. ∎

Regarding the level of distribution of shifted primes, one expects more can be said.

Conjecture 5 *The shifted primes have level of distribution $\vartheta = 1$.*

Indeed this is a slightly weaker version of the Elliott–Halberstam conjecture [6, Chapter 9.2].

On the other hand, it does not seem that the values of irreducible polynomials of degree 2 or greater have level of distribution 1.

Some interesting arithmetic sequences are known to have level of distribution 1, however, for instance, those positive integers $\mathcal{S}$ which indicate a `0` in the Thue–Morse sequence [38]. $\mathcal{S}$ may be characterized in the following way: it is the collection of positive integers n, where n has an even number of `1`s in its binary expansion.

Proposition 6 *The values of the Thue–Morse sequence are 1 well-distributed. That is, let $a_n = \mathbf{1}[n \in \mathcal{S}]$. Then (a_n) has index 1, has level of distribution $\vartheta = 1$, and is congruence uniform.*

Proof The claim that the sequence has index 1 follows from classical results of Gelfond [20] which implies $N(x) \sim x/2$ (see Theorem A in [38]). The claim that the

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

sequence has level of distribution $\vartheta = 1$ follows from Theorem 1.1 of [38]. And again by Remark 1, congruence uniformity is trivial. ∎

1.3 Main results

Our main results depend on the following setup. As above, we let (a_n) be a sequence of nonnegative numbers, and for a parameter x, we let u be a random integer such that

$$\mathbb{P}(u = m) = \frac{a_m}{N(x)}\mathbf{1}[1 \le m \le x].$$

In the case that a_n is the indicator function of a subset of natural numbers, u is uniformly distributed on elements of the set no more than x. As before, we let $p_1 \ge p_2 \ge \cdots$ be the prime factors of u listed with multiplicity, with the convention that $p_j = 1$ if n has fewer than j prime factors.

We will prove results comparing the distribution of $p_1, p_2, \dots$ to a Poisson–Dirichlet process (Theorem 7 and Lemma 8) and also an upper bound for the likelihood that $p_1 = P^+(u)$ is exceptionally large (Theorem 11). These results are related in that they use almost the same information, but because they may be of independent interest, we have written this note so that their proofs may be read independently.

Theorem 7 *If (a_n) is 1 well-distributed, then*

$$\mathbb{P}\Big(\frac{\log p_1}{\log u} \le c_1, \dots, \frac{\log p_k}{\log u} \le c_k\Big) \to \mathbb{P}(L_1 \le c_1, \dots, L_k \le c_k),$$

as $x \to \infty$. That is, the process $\frac{\log p_1}{\log u}, \frac{\log p_2}{\log u}, \dots$ tends in distribution to the Poisson–Dirichlet process $L_1, L_2, \dots$.

This generalizes to multiple prime factors a result noted by Granville [22] (with details of the proof provided by Wang [42]) that for shifted primes the Elliott–Halberstam conjecture implies the distribution of the largest prime divisor is governed by the Dickman function (a phenomenon first conjectured by Pomerance [37]). And this result proves unconditionally that large prime factors of the Thue–Morse tend to a Poisson–Dirichlet distribution.

On the other hand, while the values of irreducible polynomials do not appear to have level of distribution 1, it is reasonable to believe that their prime factors tend to a Poisson–Dirichlet distribution. That the distribution of the largest prime factor is governed by the Dickman function was given a conditional proof by Martin [31] on the assumption of a prime number theorem for polynomial sequences. It may be possible to formulate a relaxed version of level of distribution 1 which applies to the values of irreducible polynomials and which also implies a Poisson–Dirichlet distribution for large prime factors, but we do not pursue this further here. (Indeed the condition (D_σ) in the very recent preprint [34] might be a correct starting point.)

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

Even for a sequence with level of distribution less than 1, one may still compare the *correlation functions* of its large prime factors to those of the Poisson–Dirichlet process, at least against test functions with restricted support.

Lemma 8 *If (a_n) is σ well-distributed, then for any $k \geq 1$ and any continuous $\eta : \mathbb{R}^k \to \mathbb{C}$ with* $\operatorname{supp} \eta \subset \{y \in \mathbb{R}_+^k : y_1 + \cdots + y_k < \sigma\}$ *(so, in particular, η is compactly supported),*

$$\mathbb{E} \sum_{\substack{j_1,\dots,j_k \\ \text{distinct}}} \eta\Big(\frac{\log p_{j_1}}{\log u}, \dots, \frac{\log p_{j_k}}{\log u}\Big) \to \mathbb{E} \sum_{\substack{j_1,\dots,j_k \\ \text{distinct}}} \eta(L_{j_1}, \dots, L_{j_k}), \tag{6}$$

as $x \to \infty$.

Here and throughout the article, we adopt the convention that $\mathbb{R}_+ = (0, \infty)$. So η being supported in $\mathbb{R}_+^k$ means that $\eta(y_1, \dots, y_k)$ will vanish when any y_i is sufficiently close to 0. (Recall that the support of a function is the *closure* of the set on which it does not vanish.)

Remark 9 In Theorem 7 and Lemma 8 and in other places in this article, it should be possible to adopt a weaker definition of level of distribution, in which the bound (2) need only hold for sums over d in which $\Omega(d) \leq k$ and all prime factors of d are larger than x^ε, with implicit constants depending on k and ε, for all $k \geq 1$ and $\varepsilon > 0$, but we do not pursue this generalization here.

In fact, the right-hand side in (6) has a simple evaluation in general: for any continuous η with compact support in $\mathbb{R}_+^k$,

$$\mathbb{E} \sum_{\substack{j_1,\dots,j_k \\ \text{distinct}}} \eta(L_{j_1}, \dots, L_{j_k}) = \int_{\mathbb{R}_+^k} \frac{\mathbf{1}[t_1 + \cdots + t_k \leq 1]}{t_1 \dots t_k} \eta(t)\, d^k t. \tag{7}$$

This is [41, Equation (4)] (see also the closely related [1, Equation (14)]).

Theorem 7 will be seen to follow from Lemma 8 and a characterization of the Poisson–Dirichlet process due to Arratia–Kochman–Miller [1].

Remark 10 It may be worthwhile for number theorists unfamiliar with correlation sums to write out an explicit example of the sort of sum which appears on the left-hand side of (6). For instance, if $u = p_1 p_2 p_3$ with $p_1 \geq p_2 \geq p_3$ all primes, then we have for the 2-point correlation sum,

$$\begin{aligned}\sum_{\substack{j_1,j_2 \\ \text{distinct}}} \eta\Big(\frac{\log p_{j_1}}{\log u}, \frac{\log p_{j_2}}{\log u}\Big) &= \eta\Big(\frac{\log p_1}{\log u}, \frac{\log p_2}{\log u}\Big) + \eta\Big(\frac{\log p_1}{\log u}, \frac{\log p_3}{\log u}\Big) \\ &\quad + \eta\Big(\frac{\log p_2}{\log u}, \frac{\log p_1}{\log u}\Big) + \eta\Big(\frac{\log p_2}{\log u}, \frac{\log p_3}{\log u}\Big) \\ &\quad + \eta\Big(\frac{\log p_3}{\log u}, \frac{\log p_1}{\log u}\Big) + \eta\Big(\frac{\log p_3}{\log u}, \frac{\log p_2}{\log u}\Big).\end{aligned}$$

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

We have adopted the convention for such a u that $p_j = 1$ for $j \geq 4$, but because of the support of η such terms do not appear in this sum. Note that the sum is symmetric in p_1, p_2, and p_3.

The left-hand side of (6) will then be an average over u of sums of this type.

Lemma 8 has a surface-level resemblance to results that can be proven about zeros of L-functions. The reader unfamiliar with correlation sums as occur in the lemma may consult [26, Chapter 1] for a general introduction and further information.

Lemma 8 gives information about prime divisors of intermediate size, but because of restrictions on the support of η, it does not entail an asymptotic formula for the distribution of the largest prime factor of u. Our last result shows that even this partial information about the level of distribution entails an upper bound for how often the largest prime factor can be especially large.

Theorem 11 *If (a_n) is σ well-distributed for some $\sigma > 0$, then for any $\varepsilon > 0$,*

$$\limsup_{x\to\infty} \mathbb{P}(P^+(u) \geq u^{1-\varepsilon}) \ll \varepsilon,$$

where the implicit constant depends on the sequence (a_n).

We note that this is an essentially optimal result, since for sequences with level of distribution 1 by Theorem 7, we have $\mathbb{P}(P^+(u) \geq u^{1-\varepsilon}) \sim 1 - \rho(1/(1-\varepsilon))$, and for ε small enough that $1 \leq 1/(1-\varepsilon) \leq 2$, we have $1 - \rho(1/(1-\varepsilon)) = \log(1/(1-\varepsilon)) \approx \varepsilon$ (see [33, Equation (7.10)] for the evaluation of ρ in this range).

In the case of sampling shifted primes $p - 1 \leq x$, Theorem 11 recovers the following corollary.

Corollary 12 (Erdős) *For any $\varepsilon > 0$,*

$$\limsup_{x\to\infty} \frac{1}{\pi(x)} |\{p \leq x : P^+(p-1) \geq p^{1-\varepsilon}\}| \ll \varepsilon.$$

This result appears implicitly, though somewhat obscurely, in a paper of Erdős (see the line beginning with "the sum in a is less than" in [16, p. 213]). A recent paper of Ding [15] gives a proof with explicit constants and explains some of the history around this estimate.

As in these other proofs of Corollary 12, the proof of Theorem 11 relies on an upper bound sieve.

2 Resemblance to Poisson–Dirichlet: A proof of Theorem 7 and Lemma 8

Proof of Lemma 8 Let us rewrite the left-hand side of (6) as

$$\frac{1}{N(x)} \sum_{n\leq x} a_n \sum_{\substack{j_1,\dots,j_k \\ \text{distinct}}} \eta\Big(\frac{\log p_{j_1}}{\log n}, \dots, \frac{\log p_{j_k}}{\log n}\Big), \tag{8}$$

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

where in the inner sum, $p_1, p_2, \dots$ are the prime factors of n listed according to multiplicity. We have given an expression for the right-hand side of (6) in the evaluation (7). We will show convergence to this expression in the following steps: first, we show that $\log n$ can be replaced by $\log x$ in the denominators above; second, we show that those n with repeated large prime factors contribute only a negligible amount to the sum; and third, we show that the resulting expression can be expressed in more conventional language of analytic number theory and in that way easily evaluated.

Let us note from the start there are constants $a > 0$ and $c < \sigma$ such that $\eta(y_1, \dots, y_k)$ vanishes whenever $y_i \le a$ or $y_1 + \cdots + y_k \ge c$; this is because of the restricted support of η.

Thus, in our first step, we show that (8) is

$$= \frac{1}{N(x)} \sum_{n \le x} a_n \sum_{\substack{j_1, \dots, j_k \\ \text{distinct}}} \eta\Big(\frac{\log p_{j_1}}{\log x}, \dots, \frac{\log p_{j_k}}{\log x}\Big) + o_{x\to\infty}(1). \tag{9}$$

To see this, note that in (8) for each tuple $j_1, \dots, j_k$ for which the summand is non-vanishing, $p_{j_1}, \dots, p_{j_k} \ge n^a$. But n can have no more than $\lfloor 1/a \rfloor$ prime factors $p_j \ge n^a$. Thus, we have the following crude bound: for any n,

$$\Big| \sum_{\substack{j_1, \dots, j_k \\ \text{distinct}}} \eta\Big(\frac{\log p_{j_1}}{\log n}, \dots, \frac{\log p_{j_k}}{\log n}\Big) \Big| \le \max(|\eta|) \cdot k! \binom{\lfloor 1/a \rfloor}{k} = O(1), \tag{10}$$

where the implicit constant depends on k and η but does not depend on n.

Hence, for arbitrary $\varepsilon > 0$, the left-hand side of (8) is

$$= \frac{1}{N(x)} \sum_{x^{1-\varepsilon} < n \le x} a_n \sum_{\substack{j_1, \dots, j_k \\ \text{distinct}}} \eta\Big(\frac{\log p_{j_1}}{\log n}, \dots, \frac{\log p_{j_k}}{\log n}\Big) + O\Big(\frac{N(x^{1-\varepsilon})}{N(x)}\Big). \tag{11}$$

Due to continuity and compact support, η is uniformly continuous. Hence, for any $p_{j_1}, \dots, p_{j_k}$ in the sum above, for $x^{1-\varepsilon} < n \le x$,

$$\begin{aligned} \eta\Big(\frac{\log p_{j_1}}{\log n}, \dots, \frac{\log p_{j_k}}{\log n}\Big) &= \eta\Big(\frac{\log p_{j_1}}{\log x} + O(\varepsilon), \dots, \frac{\log p_{j_k}}{\log x} + O(\varepsilon)\Big) \\ &= \eta\Big(\frac{\log p_{j_1}}{\log x}, \dots, \frac{\log p_{j_k}}{\log x}\Big) + O(\delta(\varepsilon)), \end{aligned}$$

for a quantity $\delta(\varepsilon) \to 0$ as $\varepsilon \to 0$. Thus, (11) is

$$\begin{aligned} &= \frac{1}{N(x)} \sum_{x^{1-\varepsilon} < n \le x} a_n \sum_{\substack{j_1, \dots, j_k \\ \text{distinct}}} \eta\Big(\frac{\log p_{j_1}}{\log x}, \dots, \frac{\log p_{j_k}}{\log x}\Big) \\ &\qquad + O\Big(\delta(\varepsilon) \cdot \frac{N(x) - N(x^{1-\varepsilon})}{N(x)}\Big) + O\Big(\frac{N(x^{1-\varepsilon})}{N(x)}\Big). \end{aligned} \tag{12}$$

This is because in both (11) and (12), for each n, only a bounded number of tuples $j_1, \dots, j_k$ will give rise to a nonzero summand (there will be at most $k!\binom{\lfloor 1/a \rfloor}{k}$ such

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

tuples, as in (10)), and in the difference between such summands in (11) and (12) will be $O(\delta(\varepsilon))$ always.

An index $\alpha \geq \sigma > 0$ implies that the error terms in (12) are $O(\delta(\varepsilon)) + o_{x\to\infty}(1)$, and because $\delta(\varepsilon)$ can be taken arbitrarily small, this implies (9).

In our second step, we show that in (9), the sum over n can be replaced by a sum only over those n which have no repeated large prime factors. Let us consider the complementary set of n which do have a repeated large prime factor; define $S(x)$ to be the set of positive integers $n \leq x$ such that in the prime factorization $n = p_1 p_2 \dots$ some $p_i \in [x^a, x^c]$ occurs with multiplicity at least 2. We have

$$\begin{aligned}\sum_{n\in S(x)} a_n &\leq \sum_{n\leq x} a_n \sum_{x^a\leq p\leq x^c} \mathbf{1}[\, p^2|n\,] = \sum_{x^a\leq p\leq x^c} N_{p^2}(x)\\ &\ll N(x)\sum_{x^a\leq p\leq x^c}\frac{1}{p^2}(\log x)^B + \log(x)^B \sum_{x^a\leq p\leq x^c} 1\\ &= N(x)\log(x)^B \sum_{x^a\leq p\leq x^c}\frac{1}{p^2} + O_B\Big(x^c(\log x)^B\Big) = o(N(x)),\end{aligned} \tag{13}$$

where the second to last equation holds for some $B > 0$ and follows from the congruence uniformity property (C), and the last line follows from the prime number theorem and the assumption (A) that the sequence has an index satisfying $\alpha \geq \sigma > c$.

Furthermore, observe that the crude bound (10) remains true if $\log n$ is replaced by $\log x$ in the denominators, for all $n \leq x$; the argument remains the same as in (10). Hence, using this estimate and (13), we see that (9) is

$$= \frac{1}{N(x)} \sum_{\substack{n\leq x\\ n\notin S(x)}} a_n \sum_{\substack{j_1,\dots,j_k\\ \text{distinct}}} \eta\Big(\frac{\log p_{j_1}}{\log x},\dots,\frac{\log p_{j_k}}{\log x}\Big) + o_{x\to\infty}(1). \tag{14}$$

Now coming to the third and final step, we observe that (14) can be rewritten:

$$= \frac{1}{N(x)} \sum_{\substack{n\leq x\\ n\notin S(x)}} a_n \sum_{\substack{p_{j_1},\dots,p_{j_k}\\ \text{distinct}}} \eta\Big(\frac{\log p_{j_1}}{\log x},\dots,\frac{\log p_{j_k}}{\log x}\Big) + o_{x\to\infty}(1).$$

The distinction between this sum and (14) is that now the prime factors $p_{j_1},\dots,p_{j_k}$ must be distinct, whereas before we needed only that the indices $j_1,\dots,j_k$ be distinct. But because $n \notin S(x)$ in the sum and because of the support of η, these coincide whenever the inner summand is non-vanishing.

But there is a one-to-one correspondence between tuples $(p_{j_1},\dots,p_{j_k})$ of prime factors of n, all distinct, and tuples $(q_1,\dots,q_k)$ of distinct primes in which $q_1\dots q_k|n$. So we can further rewrite the above as

$$= \frac{1}{N(x)} \sum_{\substack{n\leq x\\ n\notin S(x)}} a_n \sum_{\substack{q_1,\dots,q_k\\ \text{prime, distinct}}} \mathbf{1}[\, q_1\dots q_k|n\,]\eta\Big(\frac{\log q_1}{\log x},\dots,\frac{\log q_k}{\log x}\Big) + o_{x\to\infty}(1). \tag{15}$$

By the same argument as in (10),

$$\sum_{\substack{q_1,\dots,q_k\\ \text{prime, distinct}}} \mathbf{1}[\, q_1\dots q_k|n\,]\eta\Big(\frac{\log q_1}{\log x},\dots,\frac{\log q_k}{\log x}\Big) = O(1)$$

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

uniformly in n. So we may use the bound (13) for contributions from $n \in S(x)$ to see that (15) is

$$= \frac{1}{N(x)} \sum_{n \le x} a_n \sum_{\substack{q_1,\dots,q_k \\ \text{prime, distinct}}} \mathbf{1}\big[\, q_1 \dots q_k | n \,\big] \eta\Big(\frac{\log q_1}{\log x}, \dots, \frac{\log q_k}{\log x}\Big) + o_{x\to\infty}(1).$$

But this is

$$= \frac{1}{N(x)} \sum_{\substack{q_1,\dots,q_k \\ \text{prime, distinct}}} N_{q_1 \dots q_k}(x) \eta\Big(\frac{\log q_1}{\log x}, \dots, \frac{\log q_k}{\log x}\Big) + o_{x\to\infty}(1).$$

We now simplify the above sum using that the level of distribution is $\vartheta \ge \sigma > c$. Note that the above sum can be rewritten as

$$= \frac{k!}{N(x)} \sum_{\substack{q_1 < q_2 < \cdots < q_k \\ \text{prime, distinct}}} N_{q_1 \dots q_k}(x) \eta\Big(\frac{\log q_1}{\log x}, \dots, \frac{\log q_k}{\log x}\Big) + o_{x\to\infty}(1).$$

Since $q_1 \dots q_k \le x^c$, one may think of the above sum as occurring over a subset of integers less than or equal to x^c, and using (2), this simplifies to

$$= \sum_{\substack{q_1,\dots,q_k \\ \text{prime, distinct}}} g(q_1 \dots q_k) \eta\Big(\frac{\log q_1}{\log x}, \dots, \frac{\log q_k}{\log x}\Big) + o_{x\to\infty}(1). \tag{16}$$

Observe that the upper bound in (3) implies that for any $j \ge 2$,

$$\sum_{\substack{q \ge x^a \\ \text{prime}}} g(q)^j = o_{x\to\infty}(1).$$

Utilizing the multiplicativity of g and then this bound, we see that (16) is

$$= \sum_{\substack{q_1,\dots,q_k \\ \text{prime, distinct}}} g(q_1) \dots g(q_k) \eta\Big(\frac{\log q_1}{\log x}, \dots, \frac{\log q_k}{\log x}\Big) + o_{x\to\infty}(1)$$

$$= \sum_{\substack{q_1,\dots,q_k \\ \text{prime}}} g(q_1) \dots g(q_k) \eta\Big(\frac{\log q_1}{\log x}, \dots, \frac{\log q_k}{\log x}\Big) + o_{x\to\infty}(1). \tag{17}$$

But finally the asymptotic formula in (3) and partial summation implies that if ν is the indicator function of an interval $[A, B] \subset (0, \infty)$,

$$\sum_{q \text{ prime}} g(q)\, \nu\Big(\frac{\log q}{\log x}\Big) = \int_{\mathbb{R}_+} \frac{\nu(t)}{t}\, dt + o_{x\to\infty}(1).$$

This implies that if η is the indicator function of a rectangle $[A_1, B_1] \times \cdots \times [A_k, B_k] \subset (0, \infty)^k$, the right-hand side of (17) is

$$= \Big(\sum_{q_1 \text{ prime}} g(q_1) \mathbf{1}_{[A_1,B_1]}\Big(\frac{\log q_1}{\log x}\Big) \Big) \dots \Big(\sum_{q_k \text{ prime}} g(q_k) \mathbf{1}_{[A_k,B_k]}\Big(\frac{\log q_k}{\log x}\Big) \Big) + o_{x\to\infty}(1)$$

$$= \int_{\mathbb{R}_+^k} \frac{1}{t_1 \dots t_k} \eta(t)\, d^k t + o_{x\to\infty}(1). \tag{18}$$

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

But because linear combinations of such functions are dense in the space of continuous functions with compact support in $\mathbb{R}_+^k$, a standard approximation argument implies that (18) is true for this class of functions as well.

This gives an asymptotic evaluation of the left-hand side of (8). Due to the restricted support of η, the indicator function in the expression (7) for Poisson–Dirichlet plays no role here, and we have therefore that the left-hand side tends to the right-hand side of (8). ∎

We will verify Theorem 7 using the above lemma and the following criterion, from [1, Lemma 2].

Lemma 13 (Arratia–Kochman–Miller) *If for each x, $(L_1(x), L_2(x), \dots)$ is a random process with $L_1(x) \geq L_2(x) \geq \cdots \geq 0$ satisfying $\sum L_i(x) = 1$, and if for any collection of disjoint intervals $I_i = [a_i, b_i] \subset (0,1]$ with $b_1 + \cdots + b_k < 1$, we have*

$$\liminf_{x\to\infty} \mathbb{E} \prod_{i=1}^{k} |\{j : L_j(x) \in I_i\}| \geq \prod_{i=1}^{k} \log(b_i/a_i) = \int_{I_1\times\cdots\times I_k} \frac{d^k t}{t_1 \dots t_k}, \tag{19}$$

then the process $L_1(x), L_2(x), \dots$ tends in distribution to the Poisson–Dirichlet process $L_1, L_2, \dots$ as $x \to \infty$.

Proof of Theorem 7 Let $L_i(x) = \log p_i / \log u$. If $\eta^* = \mathbf{1}_{I_1\times\cdots\times I_k}$ were continuous this would be implied by Lemma 8, as

$$\prod_{i=1}^{k} |\{j : L_j(x) \in I_i\}| = \sum_{\substack{j_1,\dots,j_k \\ \text{distinct}}} \eta^*\Big(\frac{\log p_{j_1}}{\log u}, \dots, \frac{\log p_{j_k}}{\log u}\Big).$$

But for any $\varepsilon > 0$, we can find a continuous function η^- with support in $\{y \in \mathbb{R}_+^k : y_1 + \cdots + y_k < 1\}$ such that

$$\eta^* \geq \eta^-$$

and

$$\int_{\mathbb{R}_+^k} (\eta^* - \eta^-) \frac{d^k t}{t_1 \dots t_k} \leq \varepsilon.$$

(Indeed, if a box $I_1^- \times \cdots \times I_k^-$ is inside $I_1 \times \cdots \times I_k$ and only slightly smaller, a continuous function η^- sandwiched between indicator functions will satisfy these conditions.) Thus, lower-bounding η^* by η^-, applying Lemma 8 for η^-, and then using the correlation function formula (7), we have

$$\liminf_{x\to\infty} \mathbb{E} \sum_{\substack{j_1,\dots,j_k \\ \text{distinct}}} \eta^*\Big(\frac{\log p_{j_1}}{\log u}, \dots, \frac{\log p_{j_k}}{\log u}\Big) \geq \int_{\mathbb{R}_+^k} \eta^- \frac{d^k t}{t_1 \dots t_k}.$$

But the right-hand side is within ε of

$$\int \eta^* \frac{d^k t}{t_1 \dots t_k} = \int_{I_1\times\dots I_k} \frac{d^k t}{t_1 \dots t_k},$$

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

and because ε is arbitrary, this verifies that (19) is true in this case, and the result follows. ∎

Remark 14 In effect, what Lemma 13 of Arratia–Kochman–Miller shows is that if the result of Lemma 8 holds for $\sigma = 1$ for some ordered process, then that process tends in distribution to the Poisson–Dirichlet process. That is, the convergence of correlation functions implies convergence in distribution in this context.

3 Upper bounds on largest primes: A proof of Theorem 11

In proving Theorem 11, we will use the Selberg upper bound sieve. We recall the setup from [19].

Let $\mathcal{P}$ be a finite set of primes, and define $P = \prod_{p\in\mathcal{P}} p$. Let $g(d) \in [0,1)$ be defined for $d|P$ and be a multiplicative function for this set of d. In the notation (1) as before, define

$$r_d = N_d(x) - g(d)N(x),$$

and suppose there are constants $\kappa, K > 0$ such that

$$\prod_{\substack{w\le p<z\\ p\in\mathcal{P}}} \frac{1}{1-g(p)} \le K\Big(\frac{\log z}{\log w}\Big)^{\kappa}, \tag{20}$$

for all $z > w \ge 2$.

Theorem 15 (An explicit upper bound sieve) *For $\mathcal{P}$, P, $g(d)$, and r_d as just described, with g satisfying (20), define $k = \kappa + \log K$. If a parameter D is chosen such that all $p \in \mathcal{P}$ satisfy $p < D^{1/(4k)}$, then*

$$\sum_{\substack{n\le x\\ (n,P)=1}} a_n \le C\cdot V\cdot N(x) + \sum_{\substack{d|P\\ d<D}} \tau_3(d)|r_d|,$$

where $C > 0$ is a constant which depends only on k, we have defined

$$V = \prod_{p\in\mathcal{P}} (1-g(p)),$$

and $\tau_3(d) = \sum_{d_1d_2d_3=d} 1$ is the threefold divisor function.

Proof This is Theorem 7.4 in [19], where in their notation, we have taken $s = 4k$ and $X = N(x)$. (Note that the hypothesis of their theorem requires $g(d) \in (0,1)$ for $d|P$, but one may check the proof works with no modification if $g(d) \in [0,1)$ for $d|P$.) ∎

We now apply this result to get an upper bound on the frequency with which a number n has a prime factor larger than $n^{1-\varepsilon}$. It is only when ε is small that Theorem 11 is nontrivial so we may suppose with no loss of generality that $\varepsilon < 1/2$. The idea behind the proof is easy to state: if n has a prime factor larger than $n^{1-\varepsilon}$, it will have no prime factors in between n^{ε} and $n^{1-\varepsilon}$. (Since $\varepsilon < 1/2$, we have $n^{\varepsilon} < n^{1-\varepsilon}$.) The upper bound is obtained by sieving by (a subset of) such primes.

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

Proof of Theorem 11 Note that by (3), there is a constant z_0 such that $g(p) \le 1/2$ for all $p \ge z_0$. Further, by (3), we have

$$\prod_{z_0 \le p < z} (1 - g(p)) = \exp[-\log\log z + O(1)],$$

so that under the hypothesis of Theorem 11, we have that (20) is satisfied for any subset $\mathcal{P}$ of primes larger than z_0, for $\kappa = 1$ and some constant K. As in Theorem 15, define $k = \kappa + \log K$.

Let $\alpha > 0$ be the index. Using $N(x^{1/2}) = x^{\alpha/2+o(1)} = o(N(x))$, we have

$$\sum_{n \le x} a_n \mathbf{1}[P^+(n) \ge n^{1-\varepsilon}] = \sum_{x^{1/2} < n \le x} a_n \mathbf{1}[P^+(n) \ge n^{1-\varepsilon}] + o_{x\to\infty}(N(x)). \tag{21}$$

But if $P^+(n) \ge n^{1-\varepsilon}$, then n is not divisible by any primes strictly in between n^ε and $n^{1-\varepsilon}$. If $x^{1/2} < n \le x$, then this implies such n is not divisible by any primes strictly in between x^ε and $x^{(1-\varepsilon)/2}$.

We will sieve out by a sparser set of primes even than these. Let $\delta > 0$ be some number smaller than σ. (So (a_n) has level of distribution and index larger than δ.)

Let $D = x^\delta$ and then set $\delta_0 = \min[(1-\varepsilon)/2, \delta/(4k)]$. Now define $\mathcal{P}$ to be the primes larger than z_0 and strictly in between x^ε and x^{δ_0}. (Let $\mathcal{P}$ be empty if there are no such primes.) We have $\mathcal{P}$ is a subset of the primes in between x^ε and $x^{(1-\varepsilon)/2}$, and also all $p \in \mathcal{P}$ satisfy $p < D^{1/4k}$.

Thus, if we set $P = \prod_{p\in\mathcal{P}} p$, the right-hand side of (21) is

$$\begin{aligned} &\le \sum_{\substack{n \le x \\ (n,P)=1}} a_n + o_{x\to\infty}(N(x)) \\ &\le C' \cdot V \cdot N(x) + \sum_{\substack{d<D \\ d|P}} \tau_3(d)|r_d| + o_{x\to\infty}(N(x)), \end{aligned}$$

where C' is a constant which depends only on the sequence (a_n).

Now note that for sufficiently small ε, once x is sufficiently large, the set $\mathcal{P}$ will not be empty. Without loss of generality, we may assume that ε is this small and x is at least this large in the remainder of the proof.

We have

$$V = \prod_{x^\varepsilon < p < x^{\delta_0}} (1 - g(p)) = \exp[-\log\log(x^{\delta_0}) + \log\log(x^\varepsilon) + O(1)] \ll \varepsilon.$$

Moreover, there are constants B and C such that

$$|r_d| \ll \frac{(\log x)^B C^{\Omega(d)}}{d} N(x) + (\log x)^B C^{\Omega(d)} \ll \frac{(\log x)^B C^{\Omega(d)}}{d} N(x) \tag{22}$$

for $d \le D$. (The first inequality follows from congruence uniformity, and the second from the index relation $N(x)/d \gg x^{\alpha+o(1)}/x^\delta \gg 1$.) Taking such a constant C, we note

$$\sum_{\substack{d<D \\ d|P}} \tau_3(d)|r_d| \le \Big(\sum_{\substack{d \le x^\delta \\ d|P}} \frac{C^{\Omega(d)}\tau_3(d)^2}{d} \Big)^{1/2} \Big(\sum_{\substack{d \le x^\delta \\ d|P}} \frac{d}{C^{\Omega(d)}} |r_d|^2 \Big)^{1/2}.$$

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

Note $\tau_3(n) \le 3^{\Omega(n)}$ and for sufficiently large x, we have $(9C)^{\Omega(d)}/d \le 1$ for all $d|P$. So using Lemma 4 to estimate the first parentheses and (22) to estimate the second, for sufficiently large x, the above is

$$\ll (\log x)^A \Big((\log x)^B N(x) \sum_{d \le x^\delta} |r_d| \Big)^{1/2}$$

for some constant $A > 0$.

Using that (a_n) has level of distribution greater than δ, the above is

$$\ll N(x)/(\log x)^{A'} = o_{x\to\infty}(N(x)),$$

for any constant $A' > 0$.

Putting matters together, we have

$$\sum_{n \le x} a_n \mathbf{1}[P^+(n) \ge n^{1-\varepsilon}] \ll \varepsilon N(x) + o_{x\to\infty}(N(x)),$$

where the implicit constant depends only on the sequence (a_n), which implies the theorem. ∎

Remark 16 Theorem 11 says that the likelihood that $\log P^+(u)/\log u \ge 1 - \varepsilon$ is $O(\varepsilon)$. Although in its proof, we have imported Theorem 15 directly from sieve theory, it is likely possible and would be interesting to abstract the combinatorial content of this sieve bound to prove a version of Theorem 11 for general point processes on the simplex with correlation functions known to agree with those of a Poisson–Dirichlet process against test functions with restricted support, in the sense of Lemma 8. We do not pursue this here however.

Acknowledgments We thank Yuchen Ding, Ofir Gorodetsky, Ram Murty, and anonymous referees for comments, suggestions, and corrections to earlier versions of this manuscript. We have used ChatGPT 5.2 to proofread a version of the manuscript and to help format references but no parts of the article were computer generated.

References

[1] R. Arratia, F. Kochman, and V. S. Miller, *Extensions of Billingsley's theorem via multi-intensities*. Preprint, *2014*. arXiv:1401.1555.
[2] R. Baker and G. Harman, *The Brun-Titchmarsh theorem on average*. In: B. C. Berndt, H. G. Diamond and A. J. Hildebrand (eds.), *Analytic number theory. Vol. 1* (Allerton Park, IL, 1995), Progress in Mathematics, 138, Birkhäuser, Boston, MA, 1996, pp. 39–103.
[3] 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.
[4] P. Billingsley, *On the distribution of large prime divisors*. Period. Math. Hung. 2(1972), 283–289.
[5] P. Billingsley, *Convergence of probability measures*. Second ed., Wiley Series in Probability and Statistics: Probability and Statistics, A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York, NY, 1999, pp. x+277.
[6] A. C. Cojocaru and M. R. Murty, *An introduction to sieve methods and their applications*, London Mathematical Society Student Texts, 66, Cambridge University Press, Cambridge, 2006, pp. xii+224.
[7] C. Dartyge, *Le problème de Tchébychev pour le douzième polynôme cyclotomique*. Proc. Lond. Math. Soc. **111**(2015), no. 1, 1–62.

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

[8] C. Dartyge, G. Martin, and G. Tenenbaum, *Polynomial values free of large prime factors*. Period. Math. Hungar. **43**(2001), nos. 1–2, 111–119.
[9] C. Dartyge and J. Maynard, *On the largest prime factor of quartic polynomial values: The cyclic and dihedral cases*. J. Eur. Math. Soc. published online first 2025.
[10] R. de la Bretèche, *Plus grand facteur premier de valeurs de polynômes aux entiers*. Acta Arith. **169**(2015), no. 3, 221–250.
[11] R. de la Bretèche and S. Drappeau, *Niveau de répartition des polynômes quadratiques et crible majorant pour les entiers friables*. J. Eur. Math. Soc. **22**(2020), no. 5, 1577–1624.
[12] J.-M. Deshouillers and H. Iwaniec, *On the greatest prime factor of* $n^2 + 1$. Ann. Inst. Fourier (Grenoble) **32**(1982/1983), no. 4, 1–11.
[13] K. Dickman, *On the frequency of numbers containing prime factors of a certain relative magnitude*. Ark. Mat., Astron. Fysik. **22A**(1930), no. 10, 1–14.
[14] Y. Ding, *On a conjecture on shifted primes with large prime factors*. Arch. Math. **120**(2023), no. 3, 245–252.
[15] Y. Ding, *On a conjecture on shifted primes with large prime factors, II*. Bull. Aust. Math. Soc. **111**(2025), no. 1, 48–55.
[16] P. Erdős, *On the normal number of prime factors of* $p - 1$ *and some related problems concerning Euler's* ϕ*-function*. Q. J. Math. **os-6**(1935) no. 1, 205–213.
[17] B. Feng and J. Wu, *On the density of shifted primes with large prime factors*. Sci China Math. **61**(2018), no. 1, 83–94.
[18] É. Fouvry, *Théorème de Brun-Titchmarsh: Application au théorème de Fermat*. Invent. Math. **79**(1985), no. 2, 383–407.
[19] J. Friedlander and H. Iwaniec, *Opera de cribro*, American Mathematical Society Colloquium Publications, 57, American Mathematical Society, Providence, RI, 2010, pp. xx+527.
[20] A. O. Gelfond. *Sur les nombres qui ont des propriétés additives et multiplicatives données. (French)*. Acta Arith. **13**(1967/1968), 259–265.
[21] M. Goldfeld, *On the number of primes p for which* $p + a$ *has a large prime factor*. Mathematika **16**(1969), 23–27.
[22] A. Granville, *Smooth numbers: Computational number theory and beyond*. In: J. P. Buhler and P. Stevenhagen (eds.), *Algorithmic number theory: Lattices, number fields, curves and cryptography*, Mathematical Sciences Research Institute Publications, 44, Cambridge University Press, Cambridge, 2008, pp. 267–323.
[23] D. R. Heath-Brown, *The largest prime factor of* $x^3 + 2$. Proc. Lond. Math. Soc. **82**(2001), no. 3, 554–596.
[24] C. Hooley, *On the greatest prime factor of a quadratic polynomial*. Acta Math. **117**(1967), 281–299.
[25] C. Hooley, *On the largest prime factor of* $p + a$. *Mathematika* **20**(1973), 135–143.
[26] J. B. Hough, M. Krishnapur, Y. Peres, and B. Virág, *Zeros of Gaussian analytic functions and determinantal point processes*, University Lecture Series, 51, American Mathematical Society, Providence, RI, 2009, pp. x+154.
[27] A. J. Irving, *The largest prime factor of* $x^3 + 2$. Acta Arith. **171**(2015), no. 1, 67–80.
[28] J. F. C. Kingman, *Random discrete distributions*. J. Roy. Statist. Soc. Ser. B **37**(1975), 1–22.
[29] D. E. Knuth and L. Trabb Pardo, *Analysis of a simple factorization algorithm*. Theor. Comput. Sci. **3**(1976/1977), no. 3, 321–348.
[30] E. S. Lee, *Explicit Mertens' theorems for number fields*. Bull. Aust. Math. Soc. **108**(2023), no. 1, 169–172.
[31] G. Martin, *An asymptotic formula for the number of smooth values of a polynomial*. J. Number Theory **93**(2002), no. 2, 108–182.
[32] J. Merikoski, *On the largest prime factor of* $n^2 + 1$. J. Eur. Math. Soc. **25**(2023), no. 4, 1253–1284.
[33] H. L. Montgomery and R. C. Vaughan, *Multiplicative number theory. I. Classical theory*, Cambridge Studies in Advanced Mathematics, 97, Cambridge University Press, Cambridge, 2007, pp. xviii+552.
[34] A. Mounier, *Un crible minorant effectif pour les entiers friables*. Preprint, 2024. arXiv:2402.13198
[35] T. Nagel, *Généralisation d'un théorème de Tchebycheff*. J. Math. **8**(1921), no. 4, 343–356.
[36] T. Nagell, *Introduction to number theory*, John Wiley & Sons, Inc., Almqvist & Wiksell, New York, NY, 1951, pp. 309.
[37] C. Pomerance, *Popular values of Euler's function*. *Mathematika* **27**(1980), no. 1, 84–89.
[38] L. Spiegelhofer, *The level of distribution of the Thue-Morse sequence*. Compos. Math. **156**(2020), no. 12, 2560–2587.
[39] T. Tao, *The Poisson-Dirichlet process, and large prime factors of a random number*, (2013). https://terrytao.wordpress.com/2013/09/21/the-Poisson-Dirichlet-process-and-large-prime-factors-of-a-random-number/

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press

[40] G. Tenenbaum, *Sur Une question d'Erdős et Schinzel II*. Invent. Math. **99**(1990), 215–224.
[41] A. M. Vershik, *Asymptotic distribution of decompositions of natural numbers into prime divisors. (Russian)*. Dokl. Akad. Nauk SSSR **289**(1986), no. 2, 269–272.
[42] Z. Wang, *Autour des plus grands facteurs premiers d'entiers consécutifs voisins d'un entier criblé*. Q. J. Math. **69**(2018), no. 3, 995–1013.
[43] J. Wu, *On shifted primes with large prime factors and their products*. Arch. Math. **112**(2019), no. 4, 387–393.

Chennai Mathematical Institute, India
e-mail: bharadwaj.work@outlook.com

Mathematics and Statistics, Queen's University, Kingston, Canada
e-mail: brad.w.rodgers@gmail.com

https://doi.org/10.4153/S0008439526101982 Published online by Cambridge University Press