Mathematika 64 (2018) 519–541 © 2018 University College London
doi:10.1112/S0025579318000153

DENOMINATORS OF BERNOULLI POLYNOMIALS

OLIVIER BORDELLÈS, FLORIAN LUCA, PIETER MOREE AND IGOR E. SHPARLINSKI

Abstract. For a positive integer n let

$$\mathfrak{P}_n = \prod_{\substack{p \\ s_p(n) \geqslant p}} p,$$

where p runs over primes and $s_p(n)$ is the sum of the base p digits of n. For all n we prove that $\mathfrak{P}_n$ is divisible by all "small" primes with at most one exception. We also show that $\mathfrak{P}_n$ is large and has many prime factors exceeding $\sqrt{n}$, with the largest one exceeding $n^{20/37}$. We establish Kellner's conjecture that the number of prime factors exceeding $\sqrt{n}$ grows asymptotically as $\kappa\sqrt{n}/\log n$ for some constant κ with $\kappa = 2$. Further, we compare the sizes of $\mathfrak{P}_n$ and $\mathfrak{P}_{n+1}$, leading to the somewhat surprising conclusion that although $\mathfrak{P}_n$ tends to infinity with n, the inequality $\mathfrak{P}_n > \mathfrak{P}_{n+1}$ is more frequent than its reverse.

§1. *Introduction.*

1.1. *Motivation.* For positive integers n and $b \geqslant 2$, let $s_b(n)$ be the sum of the base b-digits of n. The product

$$\mathfrak{P}_n = \prod_{p \text{ prime: } s_p(n) \geqslant p} p$$

has been introduced by Kellner and Sondow [**13**]. Although *a priori* this could be an infinite product, it is actually a finite product which terminates for $p > (n+1)/2$, see [**12**, **13**].

The relevance of this quantity is due to its link with denominators of *Bernoulli polynomials*

$$B_n(X) = \sum_{k=0}^{n} \binom{n}{k} B_k X^{n-k},$$

where B_k is the kth Bernoulli number. We also define the polynomials

$$\widetilde{B}_n(X) = B_n(X) - B_n,$$

which are of interest due to their connection to power sums, namely we have

$$\sum_{j=1}^{N-1} j^{n-1} = \frac{\widetilde{B}_n(N)}{N},$$

Received 7 July 2017, published online 16 April 2018.
MSC (2010): 11N37 (primary), 11B68 (secondary).

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

see [12–14]. It is shown in [12, 13] that the denominator of the polynomial $\widetilde{B}_n(X)$ is $\mathfrak{P}_n$, and thus,

$$\mathfrak{P}_n \widetilde{B}_n(X) \in \mathbb{Z}[X].$$

The celebrated von Staudt–Clausen theorem, see [9, Theorem 118], fully describes the denominator of B_n for an even n as the product of prime p with $p-1$ dividing n:

$$\mathfrak{Q}_n = \prod_{p:\ p-1|n} p,$$

(recall that $B_1 = -1/2$ and $B_n = 0$, and thus, $\mathfrak{Q}_n = 1$ for odd $n > 1$).

One thus sees that the denominator of $B_n(X)$ is $\operatorname{lcm}[\mathfrak{P}_n, \mathfrak{Q}_n]$.

In this paper, we prove some results about small and large prime factors of $\mathfrak{P}_n$. Kellner [12] has also introduced and studied the decomposition

$$\mathfrak{P}_n = \mathfrak{P}_n^- \cdot \mathfrak{P}_n^+,$$

where

$$\mathfrak{P}_n^- = \prod_{\substack{p<\sqrt{n} \\ s_p(n)\geqslant p}} p \quad \text{and} \quad \mathfrak{P}_n^+ = \prod_{\substack{p>\sqrt{n} \\ s_p(n)\geqslant p}} p.$$

Note that the definitions of $\mathfrak{P}_n^{\pm}$, with strict inequalities on p in both, are correct since $s_p(p^2) = 1$ if $n = p^2$ with p a prime. Hence, $p \nmid \mathfrak{P}_n$, even if $p = \sqrt{n}$ holds for a prime p. Kellner [12], motivated by the link with Bernoulli polynomials, has initiated the study of the arithmetic structure of $\mathfrak{P}_n^-$ and $\mathfrak{P}_n^+$.

Let, as usual, $\omega(m)$ and $P(m)$ be the number of distinct prime factors of m and the largest prime factor of m, respectively.

In particular, Kellner [12, Conjecture 1], makes the following conjecture.

CONJECTURE 1.1. *For $n > 192$ we have $P(\mathfrak{P}_n) > \sqrt{n}$.*

Conjecture 1.1 is equivalent with the conjecture that for $n > 192$ we have $\mathfrak{P}_n^+ > 1$ and is established, up to the numerical value of the threshold 192, in a much stronger form in Theorem 1.4.

Kellner [12, Conjecture 2] also makes the following conjecture.

CONJECTURE 1.2. *There is some absolute constant $\kappa > 0$ such that*

$$\omega(\mathfrak{P}_n^+) = (\kappa + o(1))\frac{\sqrt{n}}{\log n}$$

as $n \to \infty$.

In this paper we show that this conjecture is true with $\kappa = 2$. Moreover, we provide a much sharper asymptotic estimate (Theorem 1.5) for $\omega(\mathfrak{P}_n^+)$ than conjectured by Kellner.

Aside from our results on these two conjectures, we improve some of the results on $\mathfrak{P}_n$ and $\mathfrak{P}_n^{\pm}$ that have been given by Kellner [12] and also obtain several new results.

After introducing some required notation in §1.2, we state our results in §1.3.

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

1.2. *Notation.* Throughout the paper, the letters p and q always denote a prime number.

For a positive integer k and a positive real number x we write $\log_k x$ for the iteratively defined function given by $\log_1 x = \max\{1, \ln x\}$, where $\ln x$ is a natural logarithm of x and for $k \geqslant 2$ we put $\log_k x = \max\{1, \log_{k-1} x\}$.

The exponential integral is given by

$$\mathrm{E}_1(x) = \int_x^\infty \frac{\mathrm{e}^{-t}}{t}\, dt \quad (x > 0), \tag{1.1}$$

see [**1**, equation (5.1.1)].

We write $\pi(x)$ for the counting function of primes $p \leqslant x$. Define

$$\delta_c(x) = \mathrm{e}^{-c(\log x)^{3/5}(\log_2 x)^{-1/5}} \quad (x > e, c > 0). \tag{1.2}$$

Note that $\delta_c(x) \log x \ll \delta_{c_0}(x)$ for any $0 < c_0 < c$.

We define δ to be the *Erdős–Ford–Tenenbaum constant*

$$\delta = 1 - (1 + \ln\ln 2)/\ln 2 = 0.08607\ldots. \tag{1.3}$$

As usual $A = O(B)$, $A \ll B$, $B \gg A$ are equivalent to $|A| \leqslant c|B|$ for some *absolute* constant $c > 0$, whereas $A = o(B)$ means that $A/B \to 0$.

1.3. *Results.* First we show that all "small" primes p, with at most one exception, divide $\mathfrak{P}_n$, where "small" depends on n in a way which is made precise in the following statement.

THEOREM 1.3. *For any fixed $\varepsilon > 0$ there exists n_ε such that for all $n \geqslant n_\varepsilon$, all primes $p \leqslant (1/2 - \varepsilon)\log_2 n/\log_3 n$, with at most one exception, divide $\mathfrak{P}_n$.*

Remark 1. As $2 \nmid \mathfrak{P}_{2^n}$, we see that the exceptional prime sometimes exists.

We now obtain reasonably tight upper and lower bounds on the number of prime divisors and the largest prime divisor of $\mathfrak{P}_n$.

THEOREM 1.4. *We have,*

$$P(\mathfrak{P}_n) \geqslant P(\mathfrak{P}_n^+) \gg n^{20/37}.$$

This result implies that there exists n_0 such that Conjecture 1.1 is true with 192 replaced by n_0.

It is useful to recall that we always have $P(\mathfrak{P}_n) < n/2 + 1$ (see [**12**, **13**]), and it is easy to see that for any prime p we have $P(\mathfrak{P}_{2p-1}) = p$.

We now establish Conjecture 1.2.

THEOREM 1.5. *There exists $c > 0$ such that, for any positive integer n sufficiently large*

$$\omega(\mathfrak{P}_n^+) = n\mathrm{E}_1(\log\sqrt{n}) + O(\sqrt{n}\,\delta_c(\sqrt{n})),$$

where the exponential integral is defined in (1.1) and the function δ_c is given in (1.2).

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

Note that since $P(\mathfrak{P}_n^-) < \sqrt{n}$, it follows that $\omega(\mathfrak{P}_n^-) \leqslant \pi(\sqrt{n}) \ll \sqrt{n}/\log n$. Hence, Theorem 1.5 implies the bounds

$$\frac{\sqrt{n}}{\log n} \ll \omega(\mathfrak{P}_n^+) \leqslant \omega(\mathfrak{P}_n) \leqslant \omega(\mathfrak{P}_n^+) + O\left(\frac{\sqrt{n}}{\log n}\right)$$
$$= n\mathrm{E}_1(\log\sqrt{n}) + O\left(\frac{\sqrt{n}}{\log n}\right) \ll \frac{\sqrt{n}}{\log n}.$$

Successive integrations by parts yield that for any positive integer $N \geqslant 1$

$$\mathrm{E}_1(x) = \frac{\mathrm{e}^{-x}}{x}\sum_{m=0}^{N-1}\frac{(-1)^m m!}{x^m} + (-1)^N N!\int_x^\infty \frac{\mathrm{e}^{-t}}{t^{N+1}}\,dt,$$

from which we immediately deduce the following estimate.

COROLLARY 1.6. *For all positive integers n, N, with n sufficiently large*

$$\omega(\mathfrak{P}_n^+) = \sum_{j=1}^{N}\frac{(-1)^{j-1}2^j(j-1)!\sqrt{n}}{(\log n)^j} + O\left(\frac{2^{N+1}N!\sqrt{n}}{(\log n)^{N+1}}\right).$$

The following asymptotic formula is derived in a similar way.

THEOREM 1.7. *There exists a constant $c > 0$ such that asymptotically*

$$\log\mathfrak{P}_n^+ = \sqrt{n} + O(\sqrt{n}\,\delta_c(\sqrt{n})),$$

where the function δ_c is defined in (1.2).

Finally, we study how $\mathfrak{P}_n$ changes as we go from n to $n+1$. Since $\mathfrak{P}_n$ tends to infinity with n by Theorems 1.3 and 1.4, it follows that the inequality $\mathfrak{P}_{n+1} > \mathfrak{P}_n$ holds infinitely often. Surprisingly though, the reverse inequality holds with frequency about $\ln 2 = 0.6931\ldots$, and is thus much more frequent. However, we show that also the equality $\mathfrak{P}_n = \mathfrak{P}_{n+1}$ holds for infinitely many n.

THEOREM 1.8. *For any $x \geqslant 3$ we have the following.*

(i) *The divisibility $\mathfrak{P}_{n+1} \mid \mathfrak{P}_n$ holds for all except maybe at most $O(x(\log_2 x)^{-\delta}(\log_3 x)^{-1/2})$ positive integers $n \leqslant x$.*

(ii) *The divisibility $\mathfrak{P}_{n+1} \mid \mathfrak{P}_n$ and the inequality $\mathfrak{P}_n > \mathfrak{P}_{n+1}$ hold simultaneously for at least $(\ln 2 + o(1))x$ positive integers $n \leqslant x$ as $x \to \infty$.*

(iii) *The equality $\mathfrak{P}_q = \mathfrak{P}_{q+1}$ holds for all except maybe at most $O(\pi(x)(\log_2 x)^{-c})$ primes $q \leqslant x$, where $c > 0$ is an absolute constant.*

We remark that Kellner and Sondow [**14**, Theorem 4] have shown that for odd $n \geqslant 1$ the quotient $\mathfrak{P}_n/\mathfrak{P}_{n+1}$ is an odd integer, except that $\mathfrak{P}_n/\mathfrak{P}_{n+1} = 2$ if $n = 2^k - 1$ for some $k \geqslant 2$. One can also find in [**14**] several more results about the possible values of the ratios $\mathfrak{P}_n/\mathfrak{P}_{n+1}$ for values of n which have a special structure.

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

1.4. *Underlying techniques.* It is probably interesting to note that in our approach we use a combination of various elementary, Diophantine and analytic techniques. Furthermore, the equivalence

$$s_p(n) \geqslant p \quad \Longleftrightarrow \quad \left\lfloor \frac{n-1}{p-1} \right\rfloor - \left\lfloor \frac{n}{p} \right\rfloor = 1 \tag{1.4}$$

given by Kellner [12, equation (11)], allows us to use powerful analytic tools.

In particular, for the proof of Theorem 1.3 we employ lower bounds of linear forms in logarithms due to Matveev [16].

For the proof of Theorem 1.4 we use a result about the distribution of fractional parts of reciprocals of primes as well as bounds of exponential sums with reciprocals due to Baker and Harman [2, 3].

Finally, we use a recent improvement due to McNew *et al* [17] of a result of Erdős and Wagstaff [6] on the count of positive integers n divisible by shifted primes (see also [7]), as well as a result of Luca *et al* [15] about shifted primes divisible by another shifted prime.

§2. *Proof of Theorem 1.3.*

2.1. *Sums of digits of integers in different bases.* Let $a, b \geqslant 2$ be fixed multiplicatively independent integers. Senge and Straus [18] have shown that there are only finitely many positive integers n such that the sum of digits of n in both bases a and b is at most K, with K any fixed number. Stewart [19] makes this effective and for all $n > 25$ gives a lower bound

$$s_a(n) + s_b(n) > \frac{\log_2 n}{\log_3 n + C(a,b)} - 1, \tag{2.1}$$

where $C(a, b)$ is some constant depending on a and b.

The constant $C(a, b)$ is not made explicit in [19]. Here we do so, as this is important for our purposes, and may also be of independent interest. As in [19], our approach is based on lower bounds for linear forms in logarithms, where we use the bound of Matveev [16]. We only need it for logarithms of rational numbers rather than in its full generality for logarithms of algebraic numbers. We note that for us only the asymptotic dependence of $C(a, b)$ on a and b is important, but we also use this as an opportunity to derive a completely explicit expression for $C(a, b)$.

Let $\rho = r/s$ be a rational number in a reduced form, that is, with $\gcd(r, s) = 1$ and $s \geqslant 1$. Then its height is defined as

$$h(\rho) = h(r/s) = \max\{\log |r|, \log s\}.$$

Let $\alpha_1, \ldots, \alpha_l$ be rational numbers not zero or ± 1. We put $A_i = h(\alpha_i)$ for $i = 1, \ldots, l$. We let $d_1, \ldots, d_l$ be non-zero integers and denote $\max\{|d_1|, \ldots, |d_l|\}$ by D. Let

$$\Lambda = \prod_{i=1}^{l} \alpha_i^{d_i} - 1. \tag{2.2}$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

The result below follows from [16, Corollary 2.3], and the details have been worked out as [5, Theorem 9.4].

LEMMA 2.1. *If* $\Lambda \neq 0$, *then*

$$\log|\Lambda| > -1.4 \cdot 30^{l+3} l^{4.5}(1+\log D)\prod_{i=1}^{l} A_i.$$

We are now ready to present a more explicit version of inequality (2.1).

LEMMA 2.2. *Assume that* a *and* b *are coprime integers* $\geqslant 2$. *Let* $B \geqslant \max\{a, b\}$. *Then for all* $n > \exp(10^{15}(\log B)^4)$ *the inequality*

$$s_a(n) + s_b(n) > \frac{\log_2 n}{\log_3 n + C(a,b)}$$

holds with $C(a,b) = \log(2 \cdot 10^{12}(\log B)^2)$.

Proof. We follow [19, pp. 66–69] with the appropriate modifications. Consider the following a-ary and b-ary expansions of $n > B^4$:

$$n = a_1 a^{m_1} + a_2 a^{m_2} + \cdots + a_r a^{m_r}, \quad a_i \in \{1, \ldots, a-1\},\ 1 \leqslant i \leqslant r,$$
$$n = b_1 b^{\ell_1} + b_2 b^{\ell_2} + \cdots + b_t b^{\ell_t}, \quad b_i \in \{1, \ldots, b-1\},\ 1 \leqslant i \leqslant t,$$

where

$$m_1 > \cdots > m_r \geqslant 0 \quad \text{and} \quad \ell_1 > \ell_2 > \cdots > \ell_t \geqslant 0.$$

We put

$$\vartheta = c_0 \log_2 n,$$

where c_0 is an explicit constant depending on B that is to be chosen later. We now consider the intervals

$$\Theta_1 = [1, \vartheta), \qquad \Theta_2 = [\vartheta, \vartheta^2), \ldots, \Theta_k = [\vartheta^{k-1}, \vartheta^k),$$

where k satisfies the inequalities

$$\vartheta^k \leqslant \frac{\log n}{4 \log B} < \vartheta^{k+1}. \tag{2.3}$$

Since $n \geqslant B^4$, we have that $k \geqslant 1$ in (2.3). We now show that for an appropriate c_0 and sufficiently large n each interval Θ_s contains either a term of the form $m_1 - m_i$ for $i = 2, \ldots, r$ or a term of the form $\ell_1 - \ell_j$ for $j = 2, \ldots, t$. Let us suppose that it is not so. Then for some $1 \leqslant s \leqslant k$, the interval Θ_s does not contain any $m_1 - m_i$ and any $\ell_1 - \ell_j$. So, let u, v be given by

$$\begin{aligned} m_1 - m_u &< \vartheta^{s-1}, & m_1 - m_{u+1} &\geqslant \vartheta^s, \\ \ell_1 - \ell_v &< \vartheta^{s-1}, & \ell_1 - \ell_{v+1} &\geqslant \vartheta^s. \end{aligned} \tag{2.4}$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

In the case $m_1 - m_r < \vartheta^{s-1}$, we choose $u = r$. Similarly, we choose $v = t$ in the case $\ell_1 - \ell_t < \vartheta^{s-1}$. Note that

$$m_1 = \left\lfloor \frac{\log n}{\log a} \right\rfloor \geqslant \frac{\log n}{\log a} - 1 > \frac{\log n}{4 \log B} > \vartheta^k \geqslant \vartheta^s$$

(because $n > B^4$) and a similar inequality holds for ℓ_1. Write

$$n = a^{m_u} \alpha_a + \zeta_a \quad \text{and} \quad n = b^{\ell_v} \beta_b + \zeta_b,$$

with

$$\alpha_a = a_1 a^{m_1 - m_u} + \cdots + \alpha_a \quad \text{and} \quad \beta_b = b_1 b^{\ell_1 - \ell_v} + \cdots + \beta_b.$$

Clearly, $\zeta_a \in [0, a^{m_{u+1}+1})$ and $\zeta_b \in [0, b^{\ell_{v+1}+1})$. If $u = r$, then $\zeta_a = 0$. Similarly, if $v = t$, then $\zeta_b = 0$. We then have

$$|\alpha_a a^{m_u} - \beta_b b^{\ell_v}| \leqslant \max\{\zeta_a, \zeta_b\}. \tag{2.5}$$

We show that the left-hand side of (2.5) is non-zero. In order to do so, recall that

$$m_1 - m_u \leqslant \vartheta^{s-1} < \vartheta^k < \frac{\log n}{4 \log B},$$

so

$$\begin{aligned} m_u = m_1 - (m_1 - m_u) &> \left\lfloor \frac{\log n}{\log a} \right\rfloor - \frac{\log n}{4 \log B} \\ &\geqslant \frac{\log n}{\log a} - 1 - \frac{\log n}{4 \log B} \geqslant \frac{3 \log n}{4 \log a} - 1 > \frac{\log n}{2 \log a}, \end{aligned}$$

provided $n > B^4$. In particular, $a^{m_u} > \sqrt{n}$. A similar argument shows that $b^{\ell_v} > \sqrt{n}$. Since a and b are coprime, it follows that if the left-hand side of (2.5) is zero, then $a^{m_u} \alpha_a = b^{\ell_v} \beta_b$, so $a^{m_u} \mid \beta_b$ and $b^{\ell_v} \mid \alpha_a$. In particular, both α_a and a^{m_u} exceed $\sqrt{n}$, so

$$n = a^{m_u} \alpha_a + \zeta_a \geqslant a^{m_u} \alpha_a > \sqrt{n} \cdot \sqrt{n} = n,$$

a contradiction. Since the left-hand side of (2.5) is non-zero, the right-hand side of it is non-zero also. In particular, at least one of the inequalities $u < r$ or $v < t$ holds and

$$|\alpha_a a^{m_u} - \beta_b b^{\ell_v}| \leqslant \max\{\zeta_a, \zeta_b\} < \max\{a^{m_{u+1}+1}, b^{\ell_{v+1}+1}\} \tag{2.6}$$

(with the convention that in the maximum the term which is not defined is simply omitted). Assuming that the maximum on the right-hand side of (2.6) is $b^{\ell_{v+1}+1}$, and putting $(U, V) = (\alpha_a, \beta_b)$, we divide both sides of (2.6) by $V b^{\ell_v}$, getting

$$|(U/V) a^{m_u} b^{-\ell_v} - 1| < \frac{b^{\ell_{v+1}+1}}{b^{\ell_v} \beta_b} \leqslant \frac{1}{b^{\ell_v - \ell_{v+1} - 1}}. \tag{2.7}$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

A similar inequality holds when the maximum on the right-hand side of (2.6) is $a^{m_{u+1}+1}$, namely

$$|(V/U)a^{-m_u}b^{\ell_v} - 1| < \frac{1}{a^{m_1-m_{u+1}-1}}. \tag{2.8}$$

So, we are all set to apply Lemma 2.1 to find a lower bound on the left-hand side of (2.7) or (2.8).

We take

$$l = 3, \qquad \alpha_1 = U/V, \qquad \alpha_2 = a, \qquad \alpha_3 = b$$

and

$$(d_1, d_2, d_3) = \varepsilon(1, m_u, -l_v),$$

where $\varepsilon \in \{\pm 1\}$. More precisely, we set:

- $\varepsilon = 1$ if in (2.6) we have $\max\{a^{m_{u+1}+1}, b^{\ell_{v+1}+1}\} = b^{\ell_{v+1}+1}$, and thus (2.7) holds;
- $\varepsilon = -1$ if in (2.6) we have $\max\{a^{m_{u+1}+1}, b^{\ell_{v+1}+1}\} = a^{m_{u+1}+1}$, and thus (2.8) holds.

As a and b are multiplicatively independent, these events are clearly mutually exclusive.

Obviously, we have

$$\max\{m_u, \ell_v\} = \max\left\{\left\lfloor \frac{\log n}{\log a}\right\rfloor, \left\lfloor \frac{\log n}{\log b}\right\rfloor\right\} \leqslant \frac{\log n}{\log 2} < 2\log n,$$

so we can take $D = 2\log n$. Furthermore, $A_2 = \log a$ and $A_3 = \log b$. As for A_1, we have

$$\begin{aligned} h(\alpha_1) &\leqslant \max\{\log \alpha_a, \log \beta_b\} \leqslant \max\{\log(a^{m_1-m_u+1}), \log(b^{\ell_1-\ell_v+1})\} \\ &\leqslant \max\{m_1 - m_u + 1, \ell_1 - \ell_v + 1\}\log B \leqslant (\vartheta^{s-1} + 1)\log B \\ &\leqslant 2\vartheta^{s-1}\log B, \end{aligned}$$

so we take $A_1 = 2\vartheta^{s-1}\log B$. Now Lemma 2.1 combined with inequalities (2.7) or (2.8) tells us that

$$\begin{aligned} &-1.4 \times 30^6 \times 3^{4.5}(1 + \log(2\log n))2\vartheta^{s-1}\log a \log b \log B \\ &\quad < -\min\{\log(a^{m_1-m_{u+1}-1}), \log(b^{\ell_1-\ell_{v+1}-1})\}. \end{aligned}$$

Assuming that $n > 230$ (so that $2\log_2 n > 1 + \log(2\log n)$), we derive

$$\begin{aligned} &\min\{(m_1 - m_{u+1} - 1)\log a, (\ell_1 - \ell_v - 1)\log b\} \\ &\quad < 6 \cdot 10^{11}\vartheta^{s-1}(\log a)(\log b)(\log B)(\log_2 n). \end{aligned}$$

Since

$$\min\{m_1 - m_{u+1} - 1, \ell_1 - \ell_{v+1} - 1\} \geqslant \vartheta^s - 1 > \frac{\vartheta^s}{2}$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

for $s \geqslant 1$ (by inequality (2.4)), we see that

$$\frac{\vartheta^s}{2} < 6 \cdot 10^{11} \vartheta^{s-1} (\log B)^2 \log_2 n,$$

giving

$$\vartheta < 12 \cdot 10^{11} (\log B)^2 \log_2 n,$$

which is false if we choose $c_0 = 2 \cdot 10^{12} (\log B)^2$. The starting value for n, such that $k \geqslant 1$, remains to be proved. That is,

$$\vartheta < \frac{\log n}{4 \log B}.$$

This is equivalent to

$$\frac{\log n}{\log_2 n} > 4c_0 \log B. \tag{2.9}$$

Note that $4c_0 \log B = 8 \cdot 10^{12} (\log B)^3$. The inequality

$$\frac{x}{\log x} > A$$

is fulfilled when $x > 2A \log A$ provided $A > e$. We now take $A = 10^{13} (\log B)^3$ and we see that (2.9) holds, provided that

$$\log n > 2 \cdot 10^{13} (\log B)^3 (13 \log 10 + 3 \log_2 B).$$

Since $\log_2 B < \log B$, $B \geqslant 3$ and $13 \log 10 + 3 < 33$, it follows that the desired inequality holds for $\log n > 10^{15} (\log B)^4$. Thus, indeed each interval Θ_s for $s = 1, \ldots, k$ contains one of $m_1 - m_i$, or $\ell_1 - \ell_j$ for $i = 1, \ldots, r$, or $j = 1, \ldots, t$. Hence, $r - 1 + s - 1 \geqslant k$. Therefore,

$$\begin{aligned} s_a(n) + s_b(n) &\geqslant r + s \geqslant k + 2 = (k+1) + 1 \\ &> \frac{\log((\log n)/(4 \log B))}{\log \vartheta} + 1 \\ &= \frac{\log_2 n + (\log \vartheta - \log(4 \log B))}{\log_3 n + \log c_0}. \end{aligned}$$

Clearly, $\vartheta > 4 \log B$. Thus, the inequality

$$s_a(n) + s_b(n) > \frac{\log_2 n}{\log_3 n + C}$$

holds with $C = \log(2 \cdot 10^{12} (\log B)^2)$, which concludes the proof. □

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

2.2. *The proof of Theorem 1.3.* Let $\varepsilon > 0$ be arbitrary. We consider the primes $p \leqslant B$ with $B = \log_2 n$. Then, as $n \to \infty$, we have

$$C = \log(2 \cdot 10^{12} (\log B)^2) = 2 \log_4 n + O(1).$$

Further, we need to check that $n > \exp(10^{15} (\log B)^4)$. This is equivalent to $\log n > 10^{15} (\log_3 n)^4$, which holds for all $n > \exp(10^{18})$. Thus, assuming n is this large, for any two distinct primes $p, q \leqslant B$, it follows by inequality (2.1) that, as $n \to \infty$,

$$s_p(n) + s_q(n) > (1 + o(1)) \frac{\log_2 n}{\log_3 n}$$

holds uniformly for primes p, q in the interval $[2, B]$. If, in particular, we take $P = (1/2 - \varepsilon) \log_2 n / \log_3 n$, then it follows that for every $n > n_\varepsilon$, there is at most one prime $q \leqslant \log_2 n$ such that

$$s_q(n) < P.$$

For all other primes $p \leqslant \log_2 n$, we then have $s_p(n) \geqslant P$. So, if, in addition, we also have $p < P$, then $s_p(n) \geqslant P > p$, so $p \mid \mathfrak{P}_n$. Thus, indeed, for all $\varepsilon > 0$, there exists n_ε such that if $n > n_\varepsilon$, then all primes $p < (1/2 - \varepsilon) \log_2 n / \log_3 n$ divide $\mathfrak{P}_n$ with at most one exception.

§3. *The proof of Theorem 1.4.*

3.1. *Fractional parts of reciprocals of primes.* In what follows, for a real number x we write $\lfloor x \rfloor$ and $\{x\}$ for its integer and fractional parts, respectively. We now use the following result from [**2**, Proposition 2].

LEMMA 3.1. *For all v and w that satisfy*

$$v^{37/20} \leqslant w \leqslant v^2,$$

and are sufficiently large, we have

$$\#\left\{ p : 2v < p < 3v, \left\{ \frac{w}{p} \right\} \geqslant 1 - \frac{w}{16v^2} \right\} \gg \frac{w}{v \log w}.$$

3.2. *Concluding the proof.* We now derive one of our main technical results.

LEMMA 3.2. *For all n and v that satisfy*

$$v^{37/20} \leqslant n \leqslant v^2,$$

and are sufficiently large, we have

$$\#\{p : 2v < p < 3v, p \mid \mathfrak{P}_n^+\} \gg \frac{n}{v \log n}.$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

Proof. Let p be a prime counted in Lemma 3.1 taken with $w = n$, that is,

$$1 > \left\{\frac{n}{p}\right\} \geqslant 1 - \frac{n}{16v^2}.$$

Clearly, $p^2 > 4v^2 > n$. Thus, writing n in base p, we have $n = ap + b$, where

$$a = \lfloor n/p \rfloor \geqslant n/p - 1,$$

and

$$b = p\left\{\frac{n}{p}\right\} \geqslant p - \frac{np}{16v^2}.$$

Thus,

$$\begin{aligned} s_p(n) = a + b &\geqslant p + \frac{n}{p} - \frac{np}{16v^2} - 1 \\ &= p + \frac{n}{p}\left(1 - \frac{p^2}{16v^2}\right) - 1 > p + \frac{n}{p}\left(1 - \frac{9}{16}\right) - 1 \\ &= p + \frac{7n}{16p} - 1 > p, \end{aligned}$$

assuming p (and hence, n) is sufficiently large. □

We now use Lemma 3.2 with $v = n^{20/37}$ and obtain the lower bound claimed for $P(\mathfrak{P}_n^+)$.

§4. *Proof of Theorems 1.5 and 1.7.*

4.1. *Preliminary comments.* The proofs of both Theorems 1.5 and 1.7 may be unified in only one proof by making use of the number $\kappa \in \{0, 1\}$. We also use the functions $\mathbf{e}(x) = \exp(2\pi i x)$ and $\psi(x) = x - \lfloor x \rfloor - \frac{1}{2}$.

4.2. *Tools.* We start with the following simple result.

LEMMA 4.1. *Let $\kappa \in \{0, 1\}$. For any real numbers $\varepsilon > 0$, $x \geqslant 2$ and $1 \leqslant x^\varepsilon \leqslant y \leqslant z < x$*

$$\sum_{z < p \leqslant x} \left(\left\lfloor \frac{x+y}{p} \right\rfloor - \left\lfloor \frac{x}{p} \right\rfloor\right)(\log p)^\kappa < 2\varepsilon^{-1}(y+1)(\log x)^\kappa.$$

Proof. As $\log p \leqslant \log x$ for $p \leqslant x$ it suffices to show that the inequality holds if $\kappa = 0$. We write

$$\sum_{z < p \leqslant x} \left(\left\lfloor \frac{x+y}{p} \right\rfloor - \left\lfloor \frac{x}{p} \right\rfloor\right) = \sum_{z < p \leqslant x} \sum_{x/p < m \leqslant (x+y)/p} 1.$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

Collecting together products $k = mp \in (x, x+y]$ and changing the order of summation we obtain

$$\sum_{z<p\leqslant x}\left(\left\lfloor\frac{x+y}{p}\right\rfloor-\left\lfloor\frac{x}{p}\right\rfloor\right)=\sum_{x<k\leqslant x+y}\sum_{\substack{p\mid k\\ z<p\leqslant x}}1\leqslant\sum_{x<k\leqslant x+y}\frac{\log k}{\log z}.$$

Since for $k \in (x, x+y]$ we have

$$\frac{\log k}{\log z}\leqslant\frac{\log(x+y)}{\log z}\leqslant\frac{\log(2x)}{\log y}\leqslant 2\varepsilon^{-1},$$

the result now follows. □

COROLLARY 4.2. *Let $\kappa \in \{0,1\}$ and $\alpha \in (n^{-7/16}, 1)$. Then*

$$\sum_{\alpha^{-1}\sqrt{n}<p\leqslant n}\left(\left\lfloor\frac{n}{p}+\frac{n-p}{p(p-1)}\right\rfloor-\left\lfloor\frac{n}{p}\right\rfloor\right)(\log p)^{\kappa}\ll\alpha\sqrt{n}(\log n)^{\kappa}.$$

Proof. Since $(n-p)/(p-1) = (n-1)/(p-1) - 1$ is a decreasing function of $p \geqslant 2$, if $p > \alpha^{-1}\sqrt{n}$, then

$$\alpha\sqrt{n}-\frac{n-p}{p-1}>\alpha\sqrt{n}-\frac{n-\alpha^{-1}\sqrt{n}}{\alpha^{-1}\sqrt{n}-1}=\frac{\sqrt{n}(1+\alpha)(1-\alpha)}{\sqrt{n}-\alpha}>0.$$

Consequently,

$$\begin{aligned}&\sum_{\alpha^{-1}\sqrt{n}<p\leqslant n}\left(\left\lfloor\frac{n}{p}+\frac{n-p}{p(p-1)}\right\rfloor-\left\lfloor\frac{n}{p}\right\rfloor\right)(\log p)^{\kappa}\\&\quad\leqslant\sum_{\alpha^{-1}\sqrt{n}<p\leqslant n}\left(\left\lfloor\frac{n+\alpha\sqrt{n}}{p}\right\rfloor-\left\lfloor\frac{n}{p}\right\rfloor\right)(\log p)^{\kappa},\end{aligned}$$

and the proof is achieved on invoking Lemma 4.1 with $x = n$, $y = \alpha\sqrt{n}$, $z = \alpha^{-1}\sqrt{n}$ and $\varepsilon = 1/16$. □

LEMMA 4.3. *Let $\kappa \in \{0,1\}$ and let M be a positive integer. For any real valued function $g : \mathbb{Z} \to \mathbb{R}$ and any positive integer H we have*

$$\begin{aligned}&\left|\sum_{M<p\leqslant 2M}(\log p)^{\kappa}\psi\left(\frac{n}{p}+g(p)\right)\right|\\&\quad\ll\frac{M}{H(\log M)^{1-\kappa}}+\sum_{h\leqslant H}\frac{1}{h}\left|\sum_{M<p\leqslant 2M}(\log p)^{\kappa}\mathbf{e}\left(\frac{nh}{p}\right)\mathbf{e}(hg(p))\right|.\end{aligned}$$

Proof. For any $0 < |t| < 1$ we put $\Phi(t) = \pi t(1-|t|)\cot(\pi t) + |t|$. Note that $0 < \Phi(t) < 1$ for $0 < |t| < 1$.

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

We follow the proof of [4, Corollary 6.2]. It follows from the result of Vaaler [20], which we use in the form given by [4, Theorem 6.1], that for any real number $x \geqslant 1$ and any positive integer H,

$$\psi(x) = -\sum_{0<|h|\leqslant H} \Phi\left(\frac{h}{H+1}\right)\frac{\mathbf{e}(hx)}{2\pi i h} + \mathcal{R}_H(x), \tag{4.1}$$

where the error term $\mathcal{R}_H(x)$ satisfies

$$|\mathcal{R}_H(x)| \leqslant \frac{1}{2H+2}\sum_{|h|\leqslant H}\left(1-\frac{|h|}{H+1}\right)\mathbf{e}(hx). \tag{4.2}$$

The right-hand side of (4.2) does not appear to be, but is in fact a non-zero real number since it can be shown, cf. [4, Exercise 3, p. 350], that

$$\sum_{|h|\leqslant H}\left(1-\frac{|h|}{H+1}\right)\mathbf{e}(hx) = \frac{1}{H+1}\left|\sum_{h=0}^{H}\mathbf{e}(hx)\right|^2.$$

Using (4.1) with $x = n/p + g(p)$, multiplying by $(\log p)^\kappa$ and summing over all the primes p in $(M, 2M]$ yields the estimate

$$\sum_{M<p\leqslant 2M}(\log p)^\kappa \psi\left(\frac{n}{p}+g(p)\right) = \Sigma_1 + \Sigma_2,$$

where

$$\Sigma_1 = -\sum_{0<|h|\leqslant H}\Phi\left(\frac{h}{H+1}\right)\frac{1}{2\pi i h}\sum_{M<p\leqslant 2M}(\log p)^\kappa \mathbf{e}\left(\frac{nh}{p}+hg(p)\right),$$

$$\Sigma_2 = \sum_{M<p\leqslant 2M}(\log p)^\kappa \mathcal{R}_H\left(\frac{n}{p}+g(p)\right).$$

Now

$$\begin{aligned}|\Sigma_1| &\leqslant \sum_{0<|h|\leqslant H}\frac{1}{2\pi|h|}\left|\sum_{M<p\leqslant 2M}(\log p)^\kappa e\left(\frac{nh}{p}+hg(p)\right)\right| \\ &= \frac{1}{\pi}\sum_{h\leqslant H}\frac{1}{h}\left|\sum_{M<p\leqslant 2M}(\log p)^\kappa \mathbf{e}\left(\frac{nh}{p}\right)e(hg(p))\right|,\end{aligned}$$

and by (4.2) we have

$$\begin{aligned}|\Sigma_2| &\leqslant \sum_{M<p\leqslant 2M}(\log p)^\kappa\left|\mathcal{R}_H\left(\frac{n}{p}+g(p)\right)\right| \\ &\leqslant \frac{1}{2H+2}\sum_{|h|\leqslant H}\left(1-\frac{|h|}{H+1}\right)\sum_{M<p\leqslant 2M}(\log p)^\kappa e\left(\frac{nh}{p}+hg(p)\right)\end{aligned}$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

$$= \frac{1}{2H+2} \sum_{M<p\leqslant 2M} (\log p)^{\kappa} + \frac{1}{H+1} \sum_{h\leqslant H} \left(1 - \frac{h}{H+1}\right)$$
$$\times \operatorname{Re}\left(\sum_{M<p\leqslant 2M} (\log p)^{\kappa} \mathbf{e}\left(\frac{nh}{p} + hg(p)\right)\right)$$
$$\leqslant \frac{1}{2H+2} \sum_{M<p\leqslant 2M} (\log p)^{\kappa}$$
$$+ \sum_{h\leqslant H} \frac{1}{h} \left| \sum_{M<p\leqslant 2M} (\log p)^{\kappa} \mathbf{e}\left(\frac{nh}{p}\right) \mathbf{e}(hg(p))\right|,$$

where in the third line the cases $h = 0$ and $|h| > 0$ are separated, concluding the proof. □

Recall that the von Mangoldt function is defined by

$$\Lambda(m) = \begin{cases} \log p & \text{if } m \text{ is a power of a prime } p, \\ 0 & \text{otherwise.} \end{cases}$$

LEMMA 4.4. *Let $\kappa \in \{0, 1\}$ and let $n \geqslant 5$ and $2 \leqslant M < n/2$ be positive integers. Let $g : \mathbb{Z} \to \mathbb{R}$ be a real valued function such that either*

$$g(p) = 0$$

for all primes p or

$$g(p) = \frac{n-p}{p(p-1)}$$

for all primes p. Then for any positive integer H we have

$$\left| \sum_{M<p\leqslant 2M} (\log p)^{\kappa} \psi\left(\frac{n}{p} + g(p)\right)\right|$$
$$\ll \frac{1}{(\log M)^{1-\kappa}} \sum_{h\leqslant H} \frac{1}{h} \left(1 + \frac{nh}{M^2}\right) \max_{M<N\leqslant 2M} \left| \sum_{M<m\leqslant N} \Lambda(m) \mathbf{e}\left(\frac{hn}{m}\right)\right|$$
$$+ \frac{M}{H(\log M)^{1-\kappa}} + \frac{nH}{M^{3/2}(\log M)^{1-\kappa}} + \frac{M^{1/2}\log H}{(\log M)^{1-\kappa}}.$$

Proof. From Lemma 4.3,

$$\left| \sum_{M<p\leqslant 2M} (\log p)^{\kappa} \psi\left(\frac{n}{p} + g(p)\right)\right|$$
$$\ll \frac{M}{H(\log M)^{1-\kappa}} + \sum_{h\leqslant H} \frac{1}{h} \left| \sum_{M<p\leqslant 2M} (\log p)^{\kappa} \mathbf{e}\left(\frac{nh}{p}\right) \mathbf{e}(hg(p))\right|.$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

The total variation of the sequence $(hg(k))_k$ in $(M, 2M]$ is

$$\sum_{M<k\leqslant 2M} |hg(k+1) - hg(k)| = h \sum_{M<k\leqslant 2M} \frac{2n-(k+1)}{k(k^2-1)}$$
$$\leqslant 2nh \sum_{M<k\leqslant 2M} \frac{1}{k(k^2-1)} \ll \frac{nh}{M^2},$$

so that by Abel summation and the inequality $|\mathbf{e}(u) - \mathbf{e}(v)| \ll |u - v|$ for any $u, v \in \mathbb{R}$, we obtain

$$\left| \sum_{M<p\leqslant 2M} (\log p)^\kappa \mathbf{e}\left(\frac{nh}{p}\right) \mathbf{e}(hg(p)) \right|$$
$$\ll \left(1 + \frac{nh}{M^2}\right) \max_{M<N\leqslant 2M} \left| \sum_{M<p\leqslant N} (\log p)^\kappa \mathbf{e}\left(\frac{nh}{p}\right) \right|,$$

and by Abel summation again

$$\left| \sum_{M<p\leqslant N} (\log p)^\kappa \mathbf{e}\left(\frac{nh}{p}\right) \right|$$
$$\leqslant \frac{2}{(\log M)^{1-\kappa}} \max_{M<L\leqslant N} \left| \sum_{M\leqslant p\leqslant L} (\log p) \mathbf{e}\left(\frac{nh}{p}\right) \right|$$
$$\ll \frac{1}{(\log M)^{1-\kappa}} \left(\max_{M<L\leqslant N} \left| \sum_{M\leqslant m\leqslant L} \Lambda(m) \mathbf{e}\left(\frac{nh}{m}\right) \right| + N^{1/2} \right).$$

Putting everything together gives the asserted estimate. □

We also recall that using [8, Theorem 9] with $k = 2$ we obtain the following.

LEMMA 4.5. *If $M \leqslant \frac{1}{5} x^{3/5}$, then, for any $N \in (M, 2M]$*

$$\left| \sum_{M<m\leqslant N} \Lambda(m) \mathbf{e}\left(\frac{x}{m}\right) \right| < 17(x^2 M^{19})^{1/24} (\log 16M)^{11/4}.$$

LEMMA 4.6. *Let $\kappa \in \{0, 1\}$. There exists some absolute constant $c > 0$ such that, for any large real number $t > 1$, we have*

$$\sum_{p>t} \frac{(\log p)^\kappa}{p(p-1)} = F_\kappa(t) + O(t^{-1} \delta_c(t)),$$

where $F_0(t) = \mathrm{E}_1(\log t)$ and $F_1(t) = t^{-1}$.

Proof. The Prime Number Theorem in the version of Korobov and Vinogradov (see [10, Theorem 12.2 and equation (12.27)] or [11, Corollary 8.30], for instance) states that

$$\pi(u) = \mathrm{li}(u) + O(u \delta_{c_0}(u)),$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

where $\mathrm{li}(u)$ is the logarithmic integral

$$\mathrm{li}(u) = \int_2^u \frac{dt}{\log t} \quad (u > 0).$$

Using it we then derive by partial summation that

$$\begin{aligned}
\sum_{p>t} \frac{(\log p)^\kappa}{p(p-1)} &= \sum_{p>t} \frac{(\log p)^\kappa}{p^2} + \sum_{p>t} \frac{(\log p)^\kappa}{p^2(p-1)} \\
&= -\frac{\pi(t)(\log t)^\kappa}{t^2} + \int_t^\infty \frac{\pi(u)(2(\log u)^\kappa - \kappa(\log u)^{\kappa-1})}{u^3}\, du \\
&\quad + O\left(\frac{1}{t^2(\log t)^{1-\kappa}}\right) \\
&= -\frac{(\log t)^\kappa}{t^2}(\mathrm{li}(t) + O(t\delta_{c_0}(t))) \\
&\quad + \int_t^\infty \frac{2(\log u)^\kappa - \kappa(\log u)^{\kappa-1}}{u^3}(\mathrm{li}(u) + O(u\delta_{c_0}(u)))\, du \\
&\quad + O\left(\frac{1}{t^2(\log t)^{1-\kappa}}\right) \\
&= -\frac{\mathrm{li}(t)(\log t)^\kappa}{t^2} + \int_t^\infty \frac{\mathrm{li}(u)(2(\log u)^\kappa - \kappa(\log u)^{\kappa-1})}{u^3}\, du \\
&\quad + O(t^{-1}(\log t)^\kappa \delta_{c_0}(t)).
\end{aligned}$$

Now, integrating by parts we derive

$$\begin{aligned}
&-\frac{\mathrm{li}(t)(\log t)^\kappa}{t^2} + \int_t^\infty \frac{\mathrm{li}(u)(2(\log u)^\kappa - \kappa(\log u)^{\kappa-1})}{u^3}\, du \\
&= \int_t^\infty \frac{du}{u^2(\log u)^{1-\kappa}} = F_\kappa(t).
\end{aligned}$$

The result follows with any constant $c \in (0, c_0)$. □

4.3. *First step.* Let $\kappa \in \{0, 1\}$ and define

$$\Sigma_\kappa = \sum_{\sqrt{n}<p\leqslant n} \left(\left\lfloor \frac{n-1}{p-1} \right\rfloor - \left\lfloor \frac{n}{p} \right\rfloor\right)(\log p)^\kappa.$$

Notice that, if $\sqrt{n} < p \leqslant n$, then

$$0 \leqslant \frac{n-p}{p-1} < \sqrt{n},$$

and hence

$$\frac{n-1}{p-1} - \frac{n}{p} = \frac{n-p}{p(p-1)} < \frac{\sqrt{n}}{p} < 1,$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

so that

$$\omega(\mathfrak{p}_n^+) = \Sigma_0 \quad \text{and} \quad \log \mathfrak{p}_n^+ = \Sigma_1. \tag{4.3}$$

We now recall (1.4) and split the sum Σ_κ into two subsums as

$$\Sigma_\kappa = S_1 + S_2,$$

where

$$S_1 = \sum_{\sqrt{n} < p \leqslant \sqrt{n}\,(\delta_c(\sqrt{n}))^{-1}} \left(\left\lfloor \frac{n-1}{p-1} \right\rfloor - \left\lfloor \frac{n}{p} \right\rfloor \right) (\log p)^\kappa,$$

$$S_2 = \sum_{\sqrt{n}\,(\delta_c(\sqrt{n}))^{-1} < p \leqslant n} \left(\left\lfloor \frac{n-1}{p-1} \right\rfloor - \left\lfloor \frac{n}{p} \right\rfloor \right) (\log p)^\kappa,$$

and $c > 0$ is the constant given in Lemma 4.6.

4.4. *The sum S_2.* We use Corollary 4.2 with $\alpha = \delta_c(\sqrt{n})$ giving immediately

$$S_2 \ll \sqrt{n}\,\delta_c(\sqrt{n})(\log n)^\kappa \ll \sqrt{n}\,\delta_{c_1}(\sqrt{n}) \tag{4.4}$$

for any $c_1 \in (0, c]$.

4.5. *The sum S_1.* Using

$$\frac{n-1}{p-1} = \frac{n}{p} + \frac{n-p}{p(p-1)},$$

and recalling the definition of the function $\psi(x)$, we write

$$S_1 = S_{11} - S_{12},$$

where

$$S_{11} = \sum_{\sqrt{n} < p \leqslant \sqrt{n}\,(\delta_c(\sqrt{n}))^{-1}} \frac{(\log p)^\kappa (n-p)}{p(p-1)},$$

$$S_{12} = \sum_{\sqrt{n} < p \leqslant \sqrt{n}\,(\delta_c(\sqrt{n}))^{-1}} (\log p)^\kappa \left(\psi\left(\frac{n}{p} + \frac{n-p}{p(p-1)} \right) - \psi\left(\frac{n}{p} \right) \right).$$

The main term. For any large integer $n \geqslant 1$, we derive from Lemma 4.6 that

$$\sum_{\sqrt{n} < p \leqslant \sqrt{n}(\delta_c(\sqrt{n}))^{-1}} \frac{(\log p)^\kappa}{p(p-1)} = F_\kappa(\sqrt{n}) - F_\kappa\left(\frac{\sqrt{n}}{\delta_c(\sqrt{n})} \right) + O(n^{-1/2}\delta_c(\sqrt{n})).$$

From [1, equation (5.1.19)] we have the inequalities

$$\frac{\mathrm{e}^{-x}}{x+1} < \mathrm{E}_1(x) < \frac{\mathrm{e}^{-x}}{x} \quad (x > 0),$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

which imply that

$$F_\kappa\left(\frac{\sqrt{n}}{\delta_c(\sqrt{n})}\right) \ll n^{-1/2}\delta_c(\sqrt{n}),$$

so that

$$\sum_{\sqrt{n}<p\leqslant\sqrt{n}(\delta_c(\sqrt{n}))^{-1}} \frac{(\log p)^\kappa}{p(p-1)} = F_\kappa(\sqrt{n}) + O(n^{-1/2}\delta_c(\sqrt{n})).$$

Therefore,

$$\begin{aligned} S_{11} &= n \sum_{\sqrt{n}<p\leqslant\sqrt{n}(\delta_c(\sqrt{n}))^{-1}} \frac{(\log p)^\kappa}{p(p-1)} - \sum_{\sqrt{n}<p\leqslant\sqrt{n}(\delta_c(\sqrt{n}))^{-1}} \frac{(\log p)^\kappa}{p-1} \\ &= n \sum_{\sqrt{n}<p\leqslant\sqrt{n}(\delta_c(\sqrt{n}))^{-1}} \frac{(\log p)^\kappa}{p(p-1)} + O(\log n) \\ &= nF_\kappa(\sqrt{n}) + O(n^{1/2}\delta_c(\sqrt{n})). \end{aligned} \tag{4.5}$$

The error term. It remains to prove that, for n sufficiently large

$$|S_{12}| \ll n^{1/2}\delta_c(\sqrt{n}). \tag{4.6}$$

This estimate follows from the next result which actually gives a power saving.

LEMMA 4.7. *For $n \geqslant 2$, we have $S_{12} \ll n^{49/100+o(1)}$.*

Proof. Split the interval $(\sqrt{n}, \sqrt{n}\,(\delta_c(\sqrt{n}))^{-1}]$ into $O(\log n)$ dyadic subintervals of the form $(M, 2M]$, so that

$$\begin{aligned} &\left|\sum_{\sqrt{n}<p\leqslant\sqrt{n}\,(\delta_c(\sqrt{n}))^{-1}} (\log p)^\kappa \psi\left(\frac{n}{p}+g(p)\right)\right| \\ &\quad\ll \max_{\sqrt{n}<M\leqslant\sqrt{n}\,(\delta_c(\sqrt{n}))^{-1}} \left|\sum_{M<p\leqslant 2M} (\log p)^\kappa \psi\left(\frac{n}{p}+g(p)\right)\right| \log n, \end{aligned}$$

where, as in Lemma 4.4,

$$g(p) = 0 \quad \text{or} \quad g(p) = \frac{n-p}{p(p-1)}.$$

From Lemmas 4.4 and 4.5

$$\begin{aligned} &\left|\sum_{M<p\leqslant 2M} (\log p)^\kappa \psi\left(\frac{n}{p}+g(p)\right)\right| \\ &\quad\ll (\log M)^{7/4+\kappa} \sum_{h\leqslant H} \frac{1}{h}\left(1+\frac{nh}{M^2}\right)(h^2n^2M^{19})^{1/24} \end{aligned}$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

$$+\frac{M}{H(\log M)^{1-\kappa}}+\frac{nH}{M^{3/2}(\log M)^{1-\kappa}}+\frac{M^{1/2}\log H}{(\log M)^{1-\kappa}}$$
$$\ll((n^{26}H^{26}M^{-29})^{1/24}+(n^2H^2M^{19})^{1/24})(\log M)^{11/4}$$
$$+\frac{M}{H(\log M)^{1-\kappa}}+\frac{nH}{M^{3/2}}+M^{1/2}\log H.$$

Choose

$$H=\lfloor M^{53/50}n^{-13/25}\rfloor$$

to balance the first and the third terms in the above bound up to logarithmic factors, which we all replace by $n^{o(1)}$. We also note that since $M\geqslant n^{1/2}$ we have $H\geqslant 1$ provided that n is large enough. Hence,

$$\left|\sum_{M<p\leqslant 2M}(\log p)^{\kappa}\psi\left(\frac{n}{p}+g(p)\right)\right|$$
$$\ll((n^{26}M^{-3})^{1/50}+(nM^{22})^{1/25}+(n^{12}M^{-11})^{1/25}+M^{1/2})n^{o(1)}$$

and so, using that $M=n^{1/2+o(1)}$, we obtain

$$|S_{12}|\ll(n^{49/100}+n^{13/50}+n^{1/4})n^{o(1)}\ll n^{49/100+o(1)},$$

whenever n is sufficiently large, concluding the proof. □

4.6. *Completion of the proof of Theorems 1.5 and 1.7.* Follows at once from (4.3), (4.4), (4.5) and (4.6).

§5. *The proof of Theorem 1.8.*

5.1. *Preliminary considerations.* Let us fix some prime p and write n in base p as

$$n=\sum_{i=0}^{k}a_i p^{k-i},$$

with p-ary digits

$$a_i\in\{0,\dots,p-1\},\quad i=0,\dots,k,\ \text{and}\ a_0\neq 0.$$

In the above, $k=\lfloor\log n/\log p\rfloor+1$.

We distinguish the following two cases.

Case 1: $a_k\neq p-1$. In this case,

$$n+1=a_0p^k+\cdots+(a_k+1)$$

is the base p-representation of $n+1$. Hence, $s_p(n+1)=s_p(n)+1$. In particular, the following events

$$p\mid\mathfrak{P}_{n+1}\quad\text{and}\quad p\nmid\mathfrak{P}_n,\tag{5.1}$$

and thus,

$$s_p(n+1)\geqslant p>s_p(n),$$

are simultaneously possible only when

$$s_p(n)=p-1.\tag{5.2}$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

Case 2: $a_k = p-1$. Let $i \in [0,k]$ be such that $a_k = a_{k-1} = \cdots = a_{k-i} = p-1$, but $a_{k-i-1} < p-1$. Then

$$n+1 = a_0 p^k + \cdots + (a_{k-i-1}+1)p^i,$$

and we obtain

$$s_p(n+1) = a_0 + \cdots + a_{k-i-1} + 1 \leqslant a_0 + \cdots + a_{k-i-1} + a_k \leqslant s_p(n).$$

Hence, for each p with $a_k = p-1$, it follows that if $p \mid \mathfrak{P}_{n+1}$, then also $p \mid \mathfrak{P}_n$. However, the opposite to (5.1), that is

$$p \nmid \mathfrak{P}_{n+1} \quad \text{and} \quad p \mid \mathfrak{P}_n, \tag{5.3}$$

is also possible in this case. Actually, this case plays an important role in our argument in §5.2 below.

5.2. *Proof of (i): divisibility.* As we have seen, primes p which belong to Case 2 of §5.1 do not appear in the denominator of the ratio $\mathfrak{P}_n/\mathfrak{P}_{n+1}$. We now show that for almost all $n \leqslant x$ (with an explicit bound on the size of the exceptional set) the primes which belong to Case 1 of §5.1 do not appear in the denominator of this ratio either, and thus, we have the desired divisibility.

More precisely, using the characterization of (5.1) given by (5.2), we conclude that $\mathfrak{P}_{n+1} \mid \mathfrak{P}_n$ holds for all positive integers n that have the property that there is no prime p with $s_p(n) = p-1$. It remains to prove that the complementary set has asymptotic density zero.

Thus, we define

$$\mathcal{A}(x) = \{n \leqslant x \ : \ s_p(n) = p-1 \text{ for some prime } p\}.$$

First of all note that since $s_p(n) = p-1$, it follows that $(p-1) \mid n$. McNew *et al* [**17**], improving on the previous result of Erdős and Wagstaff [**6**], have shown that uniformly in $3 \leqslant y \leqslant x$, the estimate

$$\#\{n \leqslant x : (p-1) \mid n \text{ for some } p \geqslant y\} \ll \frac{x}{(\log y)^{\delta}(\log_2 y)^{1/2}}, \tag{5.4}$$

holds, where δ is the Erdős–Ford–Tenenbaum constant defined at (1.3). We also recall that Ford [**7**] has recently established more precise results, which, however, cannot be used to improve our bounds.

We take $y = (\log x)^{1/2}$ in (5.4), getting that the number of $n \in \mathcal{A}(x)$ such that $s_p(n) = p-1$ for some $p \geqslant y$ is

$$\frac{x}{(\log y)^{\delta}(\log_2 y)^{1/2}} \ll \frac{x}{(\log_2 x)^{\delta}(\log_3 x)^{1/2}}. \tag{5.5}$$

Assume now that $p < y$. We remark that for each p and

$$k = \left\lfloor \frac{\log n}{\log p} \right\rfloor + 1,$$

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

the condition $s_p(n) = p - 1$ leads to the equation

$$a_0 + \cdots + a_k = p - 1$$

on the p-ary digits $a_0, \ldots, a_k$ of n (we possibly append some leading zeros to make all p-ary expansions of the same length). Thus, for each p there are at most

$$\binom{k+p}{p} \leqslant (k+p)^p \leqslant (p + \log x)^p$$

possible values for the string of digits $(a_0, \ldots, a_k)$, and therefore for the number of such $n \in \mathcal{A}(x)$. Hence, the total contribution from all $p \leqslant y$ to $\#\mathcal{A}(x)$ is at most

$$\sum_{p \leqslant y} (p + \log x)^p \ll (y + \log x)^y \leqslant \exp(2(\log x)^{1/2} \log_2 x), \tag{5.6}$$

which is negligible compared to (5.5), and concludes the proof.

5.3. *Proof of (ii): divisibility and strict inequality.* Since we have established that $\mathfrak{P}_{n+1} \mid \mathfrak{P}_n$ for almost all $n \leqslant x$, it is enough to produce a sequence of n of asymptotic density $\ln 2$ such that for each of them there exists a prime p with the property (5.3).

Let $n \leqslant x$ be of the form $n = ap - 1$, where $p > \sqrt{x+1}$ and $a \geqslant 2$. Then $a \leqslant (x+1)/p < \sqrt{x+1} < p$.

In particular, $n + 1 = ap$ with $a < p$, so

$$s_p(n+1) = a < p. \tag{5.7}$$

On the other hand, since $n = (a-1)p + p - 1$ and $2 \leqslant a < p$, it follows that

$$s_p(n) = a - 1 + (p-1) = (a-2) + p \geqslant p. \tag{5.8}$$

Comparing (5.7) and (5.8), we see that we have (5.3). Obviously, each integer $n \leqslant x$ of this form can be generated by only one prime $p > \sqrt{x+1}$. Fixing p, we have that $2 \leqslant a \leqslant (x+1)/p$. Hence, there are $\lfloor (x+1)/p \rfloor - 1 = x/p + O(1)$ possibilities for a. Thus, using the Mertens formula (see [**9**, Theorem 427]), we see that we generate

$$\begin{aligned} \sum_{\sqrt{x+1} < p \leqslant x+1} \left(\frac{x}{p} + O(1) \right) &= x \sum_{\sqrt{x+1} < p < x+1} \frac{1}{p} + O(\pi(x)) \\ &= x \ln 2 + o(x) \end{aligned}$$

such integers $n \leqslant x$ as $x \to \infty$ in this way, which concludes the proof.

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

5.4. *Proof of (iii): equality.* We take $n = q - 1$ for a prime $q \leqslant x$. Clearly,

$$s_q(n) = q - 1 < q \quad \text{and} \quad s_q(n+1) = 1 < q.$$

Thus,

$$q \nmid \mathfrak{P}_n \mathfrak{P}_{n+1}.$$

We remark that q is the only prime that divides $n + 1$. Thus, Case 2 of §5.1 is impossible for any prime $p \neq q$. Hence, the only primes in which $\mathfrak{P}_n$ and $\mathfrak{P}_{n+1}$ may differ are the primes p as in Case 1 of §5.1 for which $s_p(n) = p - 1$. In particular, $p-1 \mid q-1$ with $p \neq n$. It is shown in [**15**], that there exists a positive constant c such that uniformly in $2 \leqslant y \leqslant x$, we have an analogue of (5.4) for shifted primes $q - 1$ in place of n; that is,

$$\#\{q \leqslant x : (p-1) \mid (q-1) \text{ for some } p \geqslant y\} \ll \frac{x}{\log x (\log y)^c}. \tag{5.9}$$

Again we take $y = \sqrt{\log x}$ and proceeding as in the last part of §5.2, in particular, using (5.6) and thus ignoring the fact that $n = q - 1$ is a shifted prime in this part of the argument, together with (5.9), we conclude the proof.

Acknowledgements. The authors are very grateful to Bernd Kellner and Jonathan Sondow for many useful comments on the initial version of the manuscript. The authors would also like to express their gratitude to the anonymous referee for the careful reading of the manuscript and for many valuable suggestions and corrections. F.L., P.M. and I.S. would like to thank the Max Planck Institute for Mathematics in Bonn for its hospitality and support during the period when this paper was written. The second author was supported in part by NRF (South Africa) Grants CPRR160325161141 and an A-rated researcher award and by CGA (Czech Republic) Grant 17-02804S. The fourth author was supported in part by ARC (Australia) Grant DP170100786.

References

1. M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, National Bureau of Standards, Wiley-Interscience (New York, 1972), 10th printing, with corrections.
2. R. Baker and G. Harman, The sequence x/n and its subsequences. *Rocky Mountain J. Math.* **26** (1996), 795–814.
3. R. C. Baker and G. Harman, Small remainder of a vector to suitable modulus. *Math. Zeit.* **221** (1996), 59–71.
4. O. Bordellès, *Arithmetic Tales* (*Universitext*), Springer (London, 2012).
5. Y. Bugeaud, M. Mignotte and S. Siksek, Classical and modular approaches to exponential Diophantine equations I. Fibonacci and Lucas perfect powers. *Ann. of Math. (2)* **163** (2006), 969–1018.
6. P. Erdős and S. Wagstaff, The fractional parts of the Bernoulli numbers. *Illinois J. Math.* **24** (1980), 104–112.
7. K. Ford, Integers divisible by a large shifted prime. *Acta Arith.* **178** (2017), 163–180.
8. A. Granville and O. Ramaré, Explicit bounds on exponential sums and the scarcity of squarefree binomial coefficients. *Mathematika* **43** (1996), 73–107.
9. G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, Oxford University Press (Oxford, 1979).

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press

10. A. Ivić, *The Riemann Zeta-Function. Theory and Applications*, Dover (Mineola, NY, 2003).
11. H. Iwaniec and E. Kowalski, *Analytic Number Theory*, American Mathematical Society (Providence, RI, 2004).
12. B. C. Kellner, On a product of certain primes. *J. Number Theory* **179** (2017), 126–141.
13. B. C. Kellner and J. Sondow, Power-sum denominators. *Amer. Math. Monthly* **124** (2017), 695–709.
14. B. C. Kellner and J. Sondow, The denominators of power sums of arithmetic progressions. *Preprint*, 2017, arXiv:1705.05331.
15. F. Luca, A. Pizarro-Madariaga and C. Pomerance, On the counting function of irregular primes. *Indag. Math. (N.S.)* **26** (2015), 147–161.
16. E. M. Matveev, An explicit lower bound for a homogeneous rational linear form in the logarithms of algebraic numbers, II. *Izv. Ross. Akad. Nauk Ser. Mat.* **64** (2000), 125–180; Engl. transl. *Izv. Math.* **64** (2000), 1217–1269.
17. N. McNew, P. Pollack and C. Pomerance, Numbers divisible by a large shifted primes and torsion subgroups of CM ellipitic curves. *Int. Math. Res. Not. IMRN* **2017** (2017), 5525–5553.
18. H. G. Senge and E. G. Straus, PV-numbers and sets of multiplicity. *Period. Math. Hungar.* **3** (1973), 93–100.
19. C. L. Stewart, On the representation of an integer in two different bases. *J. Reine Angew. Math.* **319** (1980), 63–72.
20. J. D. Vaaler, Some extremal functions in Fourier analysis. *Bull. Amer. Math. Soc. (N.S.)* **12** (1985), 183–215.

Olivier Bordellès,
Allée de la combe,
43000 Aiguilhe,
France
E-mail: borde43@wanadoo.fr

Florian Luca,
School of Mathematics,
University of the Witwatersrand,
Private Bag X3,
Wits 2050,
South Africa
and
Max Planck Institute for Mathematics,
Vivatsgasse 7,
53111 Bonn,
Germany
and
Department of Mathematics,
Faculty of Sciences,
University of Ostrava,
30 Dubna 22,
701 03 Ostrava 1,
Czech Republic
E-mail: florian.luca@wits.ac.za

Pieter Moree,
Max Planck Institute for Mathematics,
Vivatsgasse 7,
53111 Bonn,
Germany
E-mail: moree@mpim-bonn.mpg.de

Igor E. Shparlinski,
Department of Pure Mathematics,
University of New South Wales,
2052 NSW,
Australia
E-mail: igor.shparlinski@unsw.edu.au

https://doi.org/10.1112/S0025579318000153 Published online by Cambridge University Press