*Bull. Aust. Math. Soc.* (First published online 2025), page 1 of 14*
doi:10.1017/S0004972725100543
*Provisional—final page numbers to be inserted when paper edition is published

# ON THE LARGEST PRIME DIVISOR OF $n! + 1$

LI LAI

(Received 13 August 2025; accepted 28 August 2025)

**Abstract**

For an integer $m > 1$, we denote by $P(m)$ the largest prime divisor of $m$. We improve a result of Stewart ['On the greatest and least prime factors of $n! + 1$, II', *Publ. Math. Debrecen* **65**(3–4) (2004), 461–480] by showing that $\limsup_{n\to\infty} P(n! + 1)/n \geqslant 1 + 9 \log 2$. More generally, for any nonzero polynomial $f(X)$ with integer coefficients, we show that $\limsup_{n\to\infty} P(n! + f(n))/n \geqslant 1 + 9 \log 2$. This improves a result of Luca and Shparlinski ['Prime divisors of shifted factorials', *Bull. Lond. Math. Soc.* **37**(6) (2005), 809–817]. These improvements stem from an additional combinatorial idea that builds upon the works mentioned above.

2020 *Mathematics subject classification*: primary 11A41; secondary 11D75, 11J25.

*Keywords and phrases*: Diophantine inequalities, shifted factorials.

## 1. Introduction

Let $n$ be a positive integer. We are interested in the prime divisors of $n! + 1$. A folklore conjecture predicts that there exist infinitely many positive integers $n$ such that $n! + 1$ is a prime number. This conjecture remains wildly open, so we turn our attention to finding lower bounds for the largest prime divisor of $n! + 1$.

We denote by $P(m)$ the largest prime divisor of an integer $m > 1$. Clearly, $P(n! + 1) > n$. If $p$ is a prime divisor of $n! + 1$ for an odd integer $n \geqslant 3$, then Wilson's theorem implies that $p$ also divides $(p - 1 - n)! + 1$. From this, one can easily show that $\limsup_{n\to\infty} P(n! + 1)/n \geqslant 2$.

As far as the author knows, the first nontrivial result was obtained by Erdős and Stewart [1] in 1976. They showed that $\limsup_{n\to\infty} P(n! + 1)/n > 2$. This problem seemed to be forgotten until this century. In 2003, Luca and Shparlinski [3] proved that $\limsup_{n\to\infty} P(n! + f(n))/n \geqslant 5/2$ for any polynomial $f(X) \in \mathbb{Z}[X] \setminus \{0\}$. Later, Stewart [9] showed that $\limsup_{n\to\infty} P(n! + 1)/n \geqslant 11/2$. Moreover, Stewart showed that the set of positive integers $n$ for which $P(n! + 1) > (11/2 - \varepsilon)n$ has positive lower asymptotic density for any $\varepsilon > 0$. Recall that for a set $A$ of positive integers, the lower asymptotic density of $A$ is defined by $\liminf_{m\to\infty} \#(A \cap \{1, 2, \ldots, m\})/m$, where we use $\#S$ to denote the cardinality of a set $S$.

However, in 2002, Murty and Wong [6] showed that if the *abc* conjecture is true, then $P(n! + 1) > (1 + o(1))n \log n \quad$ as $n \to \infty$. Unconditionally, Luca and Shparlinski [3] showed that $P(n! + 1) > n + (1/4 + o(1)) \log n$ as $n \to \infty$, which sharpens a result from [1].

In this paper, we simultaneously improve the constants 5/2 and 11/2 from [3, 9]. Our result is as follows.

THEOREM 1.1. *Let $f(X) \in \mathbb{Z}[X] \setminus \{0\}$. Then,*

$$\limsup_{n\to\infty} \frac{P(n! + f(n))}{n} \geqslant 1 + 9 \log 2 \approx 7.238.$$

*Moreover, for any $\varepsilon > 0$, the set of positive integers $n$ for which $n! + f(n) > 1$ and $P(n! + f(n)) > (1 + 9 \log 2 - \varepsilon)n$ has positive lower asymptotic density.*

We follow most of the arguments from Luca and Shparlinski [3] and Stewart [9]. The new ingredient in this paper is Lemma 2.7. Roughly speaking, for a prime number $p$ and distinct positive integers $n_1, n_2, \ldots, n_t < p$, we can bound

$$\log p \sum_{ct\leqslant t'\leqslant t} \min_{1\leqslant k\leqslant t'} \{\mathrm{ord}_p(n_k! + f(n_k))\}$$

from above, where $c \approx 1/2$ is a constant and $\mathrm{ord}_p(m)$ is the $p$-adic order of an integer $m$. This upper bound gains a multiplicative constant compared with bounding

$$(\log p) \cdot \min_{1\leqslant k\leqslant t'} \{\mathrm{ord}_p(n_k! + f(n_k))\}$$

individually for each $t'$, as done in [3, 9].

Our new ingredient can be applied to some other related problems. For example, in [9], Stewart proved that

$$\liminf_{\substack{n\to\infty \\ n \text{ odd}}} \frac{p(n! + 1)}{n} \leqslant \frac{\sqrt{145} - 1}{8},$$

where $p(m)$ denotes the least prime divisor of an integer $m > 1$. By incorporating our new ingredient, the constant $(\sqrt{145} - 1)/8 \approx 1.380$ can be improved to

$$\frac{\sqrt{81(\log 2)^2 + 16} - 9 \log 2 + 4}{4} \approx 1.293.$$

(Note that, unlike Theorem 1.1, our methods cannot prove that the lower asymptotic density of the set $\{n : p(n! + 1) < (c - \varepsilon)n\}$ is positive.) In [4], Luca and Shparlinski showed that

$$\limsup_{n\to\infty} \frac{P(n! + 2^n - 1)}{n} \geqslant \frac{2\pi^2 + 3}{18}.$$

The constant $(2\pi^2+3)/18 \approx 1.263$ can be improved to

$$1 + \frac{2\pi^2 - 15}{6} \log \frac{3}{2} \approx 1.320.$$

This paper is organised as follows. In Section 2, we prove several lemmas. Only Lemma 2.7 is new, while the other lemmas are slight modifications of those in [3, 9]. In Section 3, we prove Theorem 1.1.

## 2. Preliminary lemmas and notation

We need the following two lemmas from Luca and Shparlinski [3].

LEMMA 2.1. *Let $f(X) \in \mathbb{Z}[X] \setminus \{0\}$. Then, there exists a positive integer $n_0 \geqslant 2$, which depends only on f, such that the equation*

$$f(n)\prod_{i=1}^{k}(n+i) - f(n+k) = 0$$

*does not have any integer solution $(n,k)$ with $n \geqslant n_0$ and $k \geqslant 1$. Moreover, for any integer $n \geqslant n_0$, we have $n! + f(n) > 1$ and $f(n) \neq 0$.*

PROOF. If we delete the last sentence, this corresponds to [3, Lemma 3]. However, adding the last sentence is trivial. □

LEMMA 2.2 [3, Lemma 4] *Let $f(X) \in \mathbb{Z}[X] \setminus \{0\}$. Then, there exists a positive constant $C_0$, which depends only on f, such that for any prime number p and any interval $J \subset [1,p)$ with length $|J| \geqslant 1$,*

$$\#\{n \in \mathbb{Z}_{>0} \cap J \mid p \text{ divides } n! + f(n)\} \leqslant C_0|J|^{2/3}.$$

From now on, we fix a polynomial $f(X) \in \mathbb{Z}[X] \setminus \{0\}$ and let $n_0$ be the constant as stated in Lemma 2.1. We also fix an $\varepsilon_0 \in (0, 1/100)$.

We use the Vinogradov symbol $\ll$ as well as the Landau symbols $O$ and $o$. Throughout this paper, all the constants implied by $\ll$ or $O$ depend at most on $f$ and $\varepsilon_0$ (in particular, these implicit constants do not depend on $x$, a parameter we introduce later). For convenience, we specify some constants; one of them is $C_1$. We fix a nonnegative integer $C_1$ such that

$$|f(n)| \leqslant n^{C_1} \text{ for any integer } n \geqslant 2. \tag{2.1}$$

Let $\lambda := 1 + 9\log 2 - 100\varepsilon_0$. To prove Theorem 1.1, we will show that the set

$$B(\lambda) := \{n \in \mathbb{Z}_{\geqslant n_0} \mid P(n! + f(n)) > \lambda n\}$$

has lower asymptotic density $\geqslant \varepsilon_0$. In fact, we will show that for any sufficiently large integer $x$,

$$|B(\lambda) \cap \{1, 2, \ldots, x\}| \geqslant \varepsilon_0 x.$$

We argue by contradiction. In the following, we fix a sufficiently large integer $x$ and suppose that $|B(\lambda) \cap \{1, 2, \ldots, x\}| < \varepsilon_0 x$. Define

$$Z := \prod_{\substack{n=\lceil \varepsilon_0 x\rceil \\ n\notin B(\lambda)}}^{x} (n! + f(n)). \tag{2.2}$$

(We assume $x$ is large enough so that $\varepsilon_0 x > n_0$. For $n \geqslant \varepsilon_0 x$, we have $n! + f(n) > 1$ and $f(n) \neq 0$.) By Stirling's formula, $\log(n! + f(n)) = n \log n + O(n)$, so that

$$\log Z > \biggl(\sum_{n=\lceil \varepsilon_0 x\rceil}^{x} (n \log n + O(n))\biggr) - \varepsilon_0 x(x \log x + O(x)).$$

Therefore, under the assumption that $x$ is sufficiently large,

$$\log Z > (\tfrac{1}{2} - 2\varepsilon_0)x^2 \log x. \tag{2.3}$$

We denote by $|I|$ the length of an interval $I$. The following notation will be frequently used in this paper.

DEFINITION 2.3. Let $n_1, n_2, \ldots, n_t$ be distinct positive integers ($t \geqslant 2$). We define $\mathcal{I}(n_1, n_2, \ldots, n_t)$ to be the following set of intervals. If $\sigma$ is the unique permutation on the index set $\{1, 2, \ldots, t\}$ such that $n_{\sigma(1)} < n_{\sigma(2)} < \cdots < n_{\sigma(t)}$, then

$$\mathcal{I}(n_1, n_2, \ldots, n_t) := \{I \mid I = (n_{\sigma(i)}, n_{\sigma(i+1)}] \text{ for some } i \in \{1, 2, \ldots, t-1\}\}.$$

We also define

$$\begin{aligned}\mathcal{I}_{\text{good}}(n_1, n_2, \ldots, n_t) := \{I \in \mathcal{I}(n_1, n_2, \ldots, n_t) \mid |I| \leqslant x^{0.99}/t \\ \text{or } I \text{ contains at least } 2C_1 + 1 \text{ prime numbers}\}.\end{aligned}$$

Note that $\mathcal{I}(n_1, n_2, \ldots, n_t)$ is determined solely by the set $\{n_1, n_2, \ldots, n_t\}$. However, $\mathcal{I}_{\text{good}}(n_1, n_2, \ldots, n_t)$ depends also on $t$, $x$ and the constant $C_1$ from (2.1). Clearly, $\#\mathcal{I}(n_1, n_2, \ldots, n_t) = t - 1$ and the intervals in $\mathcal{I}(n_1, n_2, \ldots, n_t)$ are disjoint. It is straightforward to check from the definition that

$$\#(\mathcal{I}(n_1, n_2, \ldots, n_t) \setminus \mathcal{I}(n_1, n_2, \ldots, n_{t-1})) \leqslant 2$$

for any integer $t \geqslant 3$ and any $t$ distinct positive integers $n_1, \ldots, n_t$. This fact will be used later in Lemma 2.7.

The following lemma is a variation of an observation in [9].

LEMMA 2.4. *Suppose that $x$ is larger than some constant depending at most on $f$ and $\varepsilon_0$. Let $t$ and $t'$ be integers such that $2 \leqslant t' \leqslant t$. Let $p$ be a prime number and $n_1, n_2, \ldots, n_t$ be distinct positive integers in the interval* $[\varepsilon_0 x, \min\{p-1, x\}]$.

https://doi.org/10.1017/S0004972725100543 Published online by Cambridge University Press

*Then, for any two distinct intervals* $I_1, I_2 \in \mathcal{I}(n_1, n_2, \ldots, n_{t'}) \cap \mathcal{I}_{\text{good}}(n_1, n_2, \ldots, n_t)$ *with* $|I_2| \geqslant |I_1|$,

$$(\log p) \cdot \min_{1 \leqslant k \leqslant t'} \{\mathrm{ord}_p(n_k! + f(n_k))\} \leqslant |I_1| \log\left(\frac{ex}{|I_1|}\right) + \left(|I_2| - |I_1| + \frac{x^{0.99}}{t} + O(1)\right) \log x. \tag{2.4}$$

PROOF. There exist indices $i_1, j_1, i_2, j_2 \leqslant t'$ such that $I_1 = (n_{i_1}, n_{j_1}]$ and $I_2 = (n_{i_2}, n_{j_2}]$. We denote

$$D = p^{\min_{1 \leqslant k \leqslant t'} \{\mathrm{ord}_p(n_k! + f(n_k))\}}.$$

From $D \mid (n_{i_1}! + f(n_{i_1}))$ and $D \mid (n_{j_1}! + f(n_{j_1}))$, we obtain

$$D \,\Bigg|\, \left(f(n_{i_1}) \prod_{k=1}^{n_{j_1} - n_{i_1}} (n_{i_1} + k) - f(n_{j_1})\right). \tag{2.5}$$

Since $n_{j_1} > n_{i_1} \geqslant \varepsilon_0 x > n_0$, by Lemma 2.1,

$$f(n_{i_1}) \prod_{k=1}^{n_{j_1} - n_{i_1}} (n_{i_1} + k) - f(n_{j_1}) \neq 0.$$

Hence, we deduce from (2.5) and (2.1) that

$$D \leqslant \left| f(n_{i_1}) \prod_{k=1}^{n_{j_1} - n_{i_1}} (n_{i_1} + k) - f(n_{j_1}) \right| \leqslant 2x^{C_1 + |I_1|}.$$

If $|I_1| \leqslant x^{0.99}/t$, then $\log D \leqslant (x^{0.99}/t + O(1)) \log x$ and (2.4) is true. Similarly,

$$D \,\Bigg|\, \left(f(n_{i_2}) \prod_{k=1}^{n_{j_2} - n_{i_2}} (n_{i_2} + k) - f(n_{j_2})\right) \neq 0. \tag{2.6}$$

If $|I_2| \leqslant x^{0.99}/t$, then (2.4) is true.

It remains to consider the case that both $|I_1|$ and $|I_2|$ are greater than $x^{0.99}/t$. Then, by the definition of $\mathcal{I}_{\text{good}}(n_1, n_2, \ldots, n_t)$, there exist distinct primes $q_1, q_2, \ldots, q_{2C_1+1}$ in the interval $I_1$, and there exist distinct primes $q'_1, q'_2, \ldots, q'_{2C_1+1}$ in the interval $I_2$. By (2.5) and (2.6),

$$D \,\Bigg|\, \left(f(n_{i_1}) f(n_{j_2}) \prod_{k=1}^{n_{j_1} - n_{i_1}} (n_{i_1} + k) - f(n_{i_2}) f(n_{j_1}) \prod_{k'=1}^{n_{j_2} - n_{i_2}} (n_{i_2} + k')\right). \tag{2.7}$$

We claim that

$$f(n_{i_1}) f(n_{j_2}) \prod_{k=1}^{n_{j_1} - n_{i_1}} (n_{i_1} + k) - f(n_{i_2}) f(n_{j_1}) \prod_{k'=1}^{n_{j_2} - n_{i_2}} (n_{i_2} + k') \neq 0. \tag{2.8}$$

In fact, since $I_1 = (n_{i_1}, n_{j_1}]$ and $I_2 = (n_{i_2}, n_{j_2}]$ are disjoint, there are only two cases. The first case is that $n_{i_1} < n_{j_1} \leqslant n_{i_2} < n_{j_2}$. If (2.8) does not hold, then we deduce that

https://doi.org/10.1017/S0004972725100543 Published online by Cambridge University Press

$q_1'q_2'\cdots q_{2C_1+1}' \mid f(n_{i_1})f(n_{j_2})$. Since $n_{i_1}, n_{j_1}, n_{i_2}, n_{j_2} \geqslant \varepsilon_0 x > n_0$, we have $f(n_{i_1})f(n_{j_2}) \neq 0$. Thus, $q_1'q_2'\cdots q_{2C_1+1}' \leqslant |f(n_{i_1})f(n_{j_2})| \leqslant x^{2C_1}$. However, $q_1'q_2'\cdots q_{2C_1+1}' \geqslant (\varepsilon_0 x)^{2C_1+1}$, which leads to a contradiction because $x$ is large. The second case is that $n_{i_2} < n_{j_2} \leqslant n_{i_1} < n_{j_1}$. If (2.8) does not hold in this case, we will arrive at a similar contradiction from $q_1 q_2 \cdots q_{2C_1+1} \mid f(n_{i_2})f(n_{j_1})$. Hence, the claim (2.8) holds.

Since $|I_2| = n_{j_2} - n_{i_2} \geqslant |I_1| = n_{j_1} - n_{i_1}$, the factorial $(n_{j_1} - n_{i_1})!$ is a divisor of the right-hand side of (2.7). Since $D$ is a power of the prime $p$ and $n_{i_1}, n_{j_1}, n_{i_2}, n_{j_2} < p$, the factorial $(n_{j_1} - n_{i_1})!$ is coprime to $D$. Thus,

$$D \,\Bigg|\, \frac{1}{(n_{j_1} - n_{i_1})!}\bigg(f(n_{i_1})f(n_{j_2}) \prod_{k=1}^{n_{j_1}-n_{i_1}} (n_{i_1} + k) - f(n_{i_2})f(n_{j_1}) \prod_{k'=1}^{n_{j_2}-n_{i_2}} (n_{i_2} + k')\bigg). \tag{2.9}$$

Noting that $m! \geqslant (m/e)^m$ for any positive integer $m$ and

$$\bigg|f(n_{i_1})f(n_{j_2}) \prod_{k=1}^{n_{j_1}-n_{i_1}} (n_{i_1} + k) - f(n_{i_2})f(n_{j_1}) \prod_{k'=1}^{n_{j_2}-n_{i_2}} (n_{i_2} + k')\bigg| \leqslant 2x^{2C_1+|I_2|},$$

we deduce from (2.9) and (2.8) that

$$D \leqslant 2x^{|I_2|-|I_1|+2C_1} \cdot \Big(\frac{ex}{|I_1|}\Big)^{|I_1|},$$

which completes the proof of (2.4). □

We require the intervals in $\mathcal{I}_{\text{good}}(n_1, n_2, \ldots, n_t)$ to contain at least $2C_1 + 1$ prime numbers solely to ensure that the left-hand side of (2.8) is nonzero. The following result of Heath-Brown will help us demonstrate that most intervals in $\mathcal{I}(n_1, n_2, \ldots, n_t)$ are 'good' (when $t \leqslant C_0 x^{2/3}$). This is the only nonelementary result that we need in this paper.

Lemma 2.5 (Heath-Brown, [2, Theorem 1]) *Let $2 = p_1 < p_2 < \cdots$ denote the sequence of prime numbers. For any $\varepsilon > 0$, there exists a constant $c_\varepsilon$ depending only on $\varepsilon$, such that for any $y \geqslant 2$,*

$$\sum_{p_k \leqslant y} (p_{k+1} - p_k)^2 \leqslant c_\varepsilon y^{23/18+\varepsilon}.$$

The exponent $23/18$ in Lemma 2.5 is less than $4/3$, which is sufficient for us to demonstrate that most intervals in $\mathcal{I}(n_1, n_2, \ldots, n_t)$ are 'good' when $t \leqslant C_0 x^{2/3}$, as stated below in Corollary 2.6. The exponent $23/18$ in Lemma 2.5 has been improved to $5/4$ by Peck [7] (see also Maynard [5]) and to $123/100$ by Stadlmann [8]. However, these improvements to Lemma 2.5 do not yield any enhancement to Theorem 1.1.

Corollary 2.6. *Suppose that $x$ is larger than some constant depending at most on $f$ and $\varepsilon_0$. Let $t$ be an integer such that $2 \leqslant t \leqslant C_0 x^{2/3}$, where $C_0$ is the constant in Lemma 2.2. Let $n_1, n_2, \ldots, n_t$ be any distinct integers in the interval $[\varepsilon_0 x, x]$. Then,*

$$\#\mathcal{I}_{\text{good}}(n_1, n_2, \ldots, n_t) \geqslant t - 2t^{0.99}.$$

PROOF. Since $x$ is large, we have $t \leqslant C_0 x^{2/3} < x^{2/3+0.01} < x$. By Lemma 2.5, there exists an absolute constant $C$ such that

$$\sum_{p_k \leqslant y} (p_{k+1} - p_k)^2 \leqslant C y^{23/18+0.001} \quad \text{for any } y \geqslant 2. \tag{2.10}$$

(Recall that we denote by $2 = p_1 < p_2 < \cdots$ the sequence of prime numbers.)

For any interval $I \in \mathcal{I}(n_1, n_2, \ldots, n_t) \setminus \mathcal{I}_{\text{good}}(n_1, n_2, \ldots, n_t)$, we have $|I| > x^{0.99}/t$ and $I$ contains at most $2C_1$ prime numbers. There are indices $i, j$ such that $I = (n_i, n_j]$. Let $p_{k+1}$ be the least prime number such that $p_{k+1} > n_j$. Then, we have $p_k \leqslant n_j \leqslant x$ and $p_{k+1} > n_j \geqslant \varepsilon_0 x$. Note that $k > 2C_1$ since $x$ is large ($x > p_{2C_1+1}/\varepsilon_0$). Given that $I$ contains at most $2C_1$ prime numbers, we must have $p_{k-2C_1} \leqslant n_i$, so $I \subset (p_{k-2C_1}, p_{k+1})$.

Let

$$m = \#(\mathcal{I}(n_1, n_2, \ldots, n_t) \setminus \mathcal{I}_{\text{good}}(n_1, n_2, \ldots, n_t)).$$

For any index $k$ such that $k > 2C_1$ and $p_k \leqslant x$, we denote by $m_k$ the number of intervals in $\mathcal{I}(n_1, n_2, \ldots, n_t) \setminus \mathcal{I}_{\text{good}}(n_1, n_2, \ldots, n_t)$ that are contained in $(p_{k-2C_1}, p_{k+1})$. Then,

$$\sum_{\substack{p_k \leqslant x \\ k > 2C_1}} m_k \geqslant m$$

and

$$\sum_{\substack{p_k \leqslant x \\ k > 2C_1}} (p_{k+1} - p_{k-2C_1})^2 \geqslant \sum_{\substack{p_k \leqslant x \\ k > 2C_1}} m_k^2 \Big(\frac{x^{0.99}}{t}\Big)^2 \geqslant \frac{x^{1.98}}{t^2} \sum_{\substack{p_k \leqslant x \\ k > 2C_1}} m_k \geqslant m \cdot \frac{x^{1.98}}{t^2}. \tag{2.11}$$

However, by (2.10) and the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\sum_{\substack{p_k \leqslant x \\ k > 2C_1}} (p_{k+1} - p_{k-2C_1})^2 \\
&\quad \leqslant (2C_1 + 1) \sum_{\substack{p_k \leqslant x \\ k > 2C_1}} ((p_{k+1} - p_k)^2 + (p_k - p_{k-1})^2 + \cdots + (p_{k-2C_1+1} - p_{k-2C_1})^2) \\
&\quad \leqslant (2C_1 + 1)^2 \sum_{p_k \leqslant x} (p_{k+1} - p_k)^2 \\
&\quad \leqslant (2C_1 + 1)^2 \cdot C \cdot x^{23/18+0.001} < x^{23/18+0.01} \quad \text{(because $x$ is large).}
\end{aligned} \tag{2.12}$$

Combining (2.11) and (2.12), we obtain

$$m \leqslant \frac{t^2}{x^{13/18-0.03}} < \frac{t}{x^{1/18-0.04}} < t^{0.99}.$$

(We have used $t < x^{2/3+0.01}$ and $t < x$.) Therefore,

$$\#\mathcal{I}_{\text{good}}(n_1, n_2, \ldots, n_t) = t - 1 - m \geqslant t - 2t^{0.99}.$$

The proof of Corollary 2.6 is complete. ◻

The following lemma is the key step of this paper.

https://doi.org/10.1017/S0004972725100543 Published online by Cambridge University Press

LEMMA 2.7. *Suppose that $x$ is larger than some constant depending at most on $f$ and $\varepsilon_0$. Let $p$ be a prime number. Let $t$ be an integer such that*

$$\left(\frac{10}{\varepsilon_0}\right)^{100} \leqslant t \leqslant C_0 x^{2/3},$$

*where $C_0$ is the constant in Lemma 2.2. Let $J$ be an interval such that*

$$J \subset [\varepsilon_0 x, \min\{p-1, x\}].$$

*If $n_1, n_2, \ldots, n_t$ are distinct integers in the interval $J$, then*

$$\log p \sum_{t'=\lceil (1+\varepsilon_0)t/2 \rceil}^{t} \min_{1 \leqslant j \leqslant t'} \{\mathrm{ord}_p(n_j! + f(n_j))\} \leqslant \frac{|J|}{2} \log t + \frac{x \log x}{t^{0.98}} + O(x). \tag{2.13}$$

PROOF. Let $t_1 = \#\mathcal{I}_{\mathrm{good}}(n_1, n_2, \ldots, n_t)$. Obviously, $t_1 \leqslant t-1$. By Corollary 2.6,

$$t_1 \geqslant t - 2t^{0.99}. \tag{2.14}$$

We list the lengths of all intervals in $\mathcal{I}_{\mathrm{good}}(n_1, n_2, \ldots, n_t)$ in ascending order:

$$\gamma_1 \leqslant \gamma_2 \leqslant \cdots \leqslant \gamma_{t_1}. \tag{2.15}$$

Since $n_1, n_2, \ldots, n_t \in J$ and the intervals in $\mathcal{I}_{\mathrm{good}}(n_1, n_2, \ldots, n_t)$ are disjoint,

$$\sum_{k=1}^{t_1} \gamma_k \leqslant |J|.$$

Let

$$t_2 = \lfloor t - 3t^{0.99} \rfloor. \tag{2.16}$$

From $(t_1 - t_2 + 1)\gamma_{t_2} \leqslant \sum_{t'=t_2}^{t_1} \gamma_{t'} \leqslant |J|$, and using (2.14) and (2.16), we deduce that

$$\gamma_{t_2} \leqslant \frac{|J|}{t^{0.99}}. \tag{2.17}$$

For any $t' \leqslant t$, from the definition of $\mathcal{I}(n_1, n_2, \ldots, n_{t'})$ (see Definition 2.3), we immediately see that

$$\#(\mathcal{I}(n_1, n_2, \ldots, n_{t'}) \setminus \mathcal{I}(n_1, n_2, \ldots, n_{t'-1})) \leqslant 2.$$

Let $k$ be an integer in the range

$$0 \leqslant k \leqslant t_3 \quad \text{where } t_3 = \lfloor \tfrac{1}{2}(t - 5t^{0.99}) \rfloor - 1. \tag{2.18}$$

Then,

$$\#(\mathcal{I}(n_1, n_2, \ldots, n_{t-k}) \cap \mathcal{I}_{\mathrm{good}}(n_1, n_2, \ldots, n_t)) \geqslant t_1 - 2k. \tag{2.19}$$

In particular, for any $k$ in the range (2.18), we have (recall (2.14))

$$\#(\mathcal{I}(n_1, n_2, \ldots, n_{t-k}) \cap \mathcal{I}_{\mathrm{good}}(n_1, n_2, \ldots, n_t)) \geqslant 3t^{0.99}.$$

https://doi.org/10.1017/S0004972725100543 Published online by Cambridge University Press

Let $I_1^{(k)}, I_2^{(k)}, \ldots, I_{\lceil t^{0.99}\rceil+1}^{(k)}$ be the shortest $\lceil t^{0.99}\rceil + 1$ intervals in $\mathcal{I}(n_1, n_2, \ldots, n_{t-k}) \cap \mathcal{I}_{\text{good}}(n_1, n_2, \ldots, n_t)$ such that

$$|I_1^{(k)}| \leqslant |I_2^{(k)}| \leqslant \cdots \leqslant |I_{\lceil t^{0.99}\rceil+1}^{(k)}|.$$

Then, by (2.19) and (2.15),

$$|I_i^{(k)}| \leqslant \gamma_{2k+i} \quad \text{for } i = 1, 2, \ldots, \lceil t^{0.99}\rceil + 1. \tag{2.20}$$

In particular, taking (2.18) and (2.17) into consideration,

$$|I_{\lceil t^{0.99}\rceil+1}^{(k)}| \leqslant \gamma_{t_2} \leqslant \frac{|J|}{t^{0.99}}. \tag{2.21}$$

Let $k$ be in the range (2.18). By Lemma 2.4, for any $i \in \{1, 2, \ldots, \lceil t^{0.99}\rceil\}$,

$$\begin{aligned}(\log p) \cdot &\min_{1\leqslant j\leqslant t-k}\{\operatorname{ord}_p(n_j!+f(n_j))\} \\ &\leqslant |I_i^{(k)}| \log\left(\frac{ex}{|I_i^{(k)}|}\right) + \left(|I_{i+1}^{(k)}| - |I_i^{(k)}| + \frac{x^{0.99}}{t} + O(1)\right)\log x.\end{aligned} \tag{2.22}$$

Taking the average of (2.22) over $i \in \{1, 2, \ldots, \lceil t^{0.99}\rceil\}$ and noticing that the function $y \mapsto y\log(ex/y)$ is increasing on $y \in (0, x)$, we deduce that

$$\begin{aligned}(\log p) \cdot &\min_{1\leqslant j\leqslant t-k}\{\operatorname{ord}_p(n_j!+f(n_j))\} \\ &\leqslant |I_{\lceil t^{0.99}\rceil}^{(k)}| \log\left(\frac{ex}{|I_{\lceil t^{0.99}\rceil}^{(k)}|}\right) + \left(\frac{|I_{\lceil t^{0.99}\rceil+1}^{(k)}|}{t^{0.99}} + \frac{x^{0.99}}{t} + O(1)\right)\log x.\end{aligned}$$

Then, bounding $|I_{\lceil t^{0.99}\rceil}^{(k)}|$ by (2.20) and bounding $|I_{\lceil t^{0.99}\rceil+1}^{(k)}|$ by (2.21), we obtain

$$\begin{aligned}(\log p) \cdot &\min_{1\leqslant j\leqslant t-k}\{\operatorname{ord}_p(n_j!+f(n_j))\} \\ &\leqslant \gamma_{2k+\lceil t^{0.99}\rceil} \log\left(\frac{ex}{\gamma_{2k+\lceil t^{0.99}\rceil}}\right) + \left(\frac{|J|}{t^{1.98}} + \frac{x^{0.99}}{t} + O(1)\right)\log x.\end{aligned} \tag{2.23}$$

Summing (2.23) over $k$ in the range (2.18),

$$\begin{aligned}\log p &\sum_{k=0}^{t_3} \min_{1\leqslant j\leqslant t-k}\{\operatorname{ord}_p(n_j!+f(n_j))\} \\ &\leqslant \left(\sum_{k=0}^{t_3} \gamma_{2k+\lceil t^{0.99}\rceil} \log\left(\frac{ex}{\gamma_{2k+\lceil t^{0.99}\rceil}}\right)\right) + \left(\frac{|J|}{t^{0.98}} + x^{0.99} + O(t)\right)\log x \\ &\leqslant \left(\sum_{k=0}^{t_3} \gamma_{2k+\lceil t^{0.99}\rceil} \log\left(\frac{ex}{\gamma_{2k+\lceil t^{0.99}\rceil}}\right)\right) + \frac{x\log x}{t^{0.98}} + O(x),\end{aligned} \tag{2.24}$$

where in the last inequality, we have used $|J| < x$ and $t \leqslant C_0 x^{2/3}$.

https://doi.org/10.1017/S0004972725100543 Published online by Cambridge University Press

Finally, since the function $y \mapsto y\log(ex/y)$ is concave and increasing on $y \in (0, x)$, and since

$$\sum_{k=0}^{t_3} \gamma_{2k+\lceil t^{0.99}\rceil} \leqslant \frac{1}{2}\sum_{k=0}^{t_3}(\gamma_{2k+\lceil t^{0.99}\rceil} + \gamma_{2k+\lceil t^{0.99}\rceil+1}) < \frac{1}{2}\sum_{k=1}^{t_1}\gamma_k \leqslant \frac{|J|}{2},$$

we have (by Jensen's inequality and monotonicity)

$$\begin{aligned}\sum_{k=0}^{t_3} \gamma_{2k+\lceil t^{0.99}\rceil}\log\Big(\frac{ex}{\gamma_{2k+\lceil t^{0.99}\rceil}}\Big) &< (t_3+1)\cdot\frac{|J|/2}{t_3+1}\log\Big(\frac{ex}{(|J|/2)/(t_3+1)}\Big)\\ &< \frac{|J|}{2}\log\Big(\frac{ext}{|J|/2}\Big) = \frac{|J|}{2}\log t + \frac{|J|}{2}\log\Big(\frac{ex}{|J|/2}\Big)\\ &< \frac{|J|}{2}\log t + x, \end{aligned} \tag{2.25}$$

where the last inequality holds because $|J|/2 < x$ and the function $y \mapsto y\log(ex/y)$ is increasing on $y \in (0, x)$.

By (2.24), (2.25) and the fact that $t - t_3 < \lceil((1+\varepsilon_0)t/2)\rceil$ (because $t > (10/\varepsilon_0)^{100}$ and $t_3 = \lfloor(t - 5t^{0.99})/2\rfloor - 1$), we complete the proof of (2.13). □

## 3. Proof of Theorem 1.1

By the definitions of $B(\lambda)$ and $Z$ (see (2.2)), we have $P(Z) \leqslant \lambda x$. For any prime number $p$, we define the set

$$N_p := \{n \in \mathbb{Z} \cap [\varepsilon_0 x, x] \mid n \notin B(\lambda) \text{ and } p \mid (n! + f(n))\}.$$

Then,

$$(\log p)\,\mathrm{ord}_p(Z) = \log p \sum_{n\in N_p} \mathrm{ord}_p(n! + f(n)). \tag{3.1}$$

If $n \geqslant p$ and $p \mid (n! + f(n))$, then $f(n) \equiv 0 \pmod p$. Let $a \neq 0$ be the leading coefficient of $f(X)$. If $p > |a|$, then $f(X) \pmod p$ has degree $\deg f$. Hence, for any two real numbers $y_2 > y_1 \geqslant p$, we have $\#(N_p \cap [y_1, y_2]) \leqslant ((y_2 - y_1)/p + 1)\deg f$. That is,

$$\#(N_p \cap [y_1, y_2]) \ll \frac{y_2 - y_1}{p} + 1 \quad \text{for any } y_2 > y_1 \geqslant p. \tag{3.2}$$

If $p \leqslant |a|$, then $\#(N_p \cap [y_1, y_2]) \leqslant y_2 - y_1 + 1 \leqslant |a|(y_2 - y_1)/p + 1$, so (3.2) still holds.

Recall that $|f(n)| \leqslant n^{C_1}$ for all $n \geqslant 2$. Since $f(n) \neq 0$ for $n \geqslant n_0$, we have $\mathrm{ord}_p(f(n)) \leqslant C_1 \log n/\log p$. We can take a large constant $C_2$ depending only on $C_1$ and $n_0$ (so $C_2$ depends only on $f$) such that

$$\begin{cases} C_2 - C_1\log C_2/\log 2 > C_1 + 1,\\ C_2 > C_1/\log 2,\\ C_2 > n_0. \end{cases}$$

https://doi.org/10.1017/S0004972725100543 Published online by Cambridge University Press

(For example, $C_2 = 100(C_1+1)^2 + n_0$ is sufficient.) Then, for any prime $p$ and any integer $n \geqslant C_2 p$, we have $n/p - 1 > C_1 \log n/\log p$. Since $\mathrm{ord}_p(n!) > n/p - 1$, we deduce that $\mathrm{ord}_p(n! + f(n)) = \mathrm{ord}_p(f(n)) = O(\log n/\log p)$ for any prime $p$ and any $n \geqslant C_2 p$. Since $\{n \in N_p \mid n \geqslant C_2 p\} \subset [C_2 p, x]$, it follows that $\#\{n \in N_p \mid n \geqslant C_2 p\} \ll (x/p + 1)$ by (3.2). Therefore, for any prime $p$,

$$\log p \sum_{\substack{n\in N_p\\ n\geqslant C_2 p}} \mathrm{ord}_p(n! + f(n)) \ll \left(\frac{x}{p} + 1\right) \cdot \log x \ll x \log x. \tag{3.3}$$

By (3.2), $|N_p \cap [p, C_2 p)| = O(1)$. Since $\log(n! + f(n)) = O(x \log x)$ for $n_0 < n \leqslant x$,

$$\log p \sum_{\substack{n\in N_p\\ p\leqslant n< C_2 p}} \mathrm{ord}_p(n! + f(n)) \ll 1 \cdot x \log x \ll x \log x. \tag{3.4}$$

For any $n \in N_p \cap [1, p)$, by the definition of $B(\lambda)$, we have $p \leqslant \lambda n$. Consider the interval

$$J_p := [\max\{p/\lambda, \varepsilon_0 x\}, \min\{p-1, x\}].$$

Then,

$$N_p \cap [1, p) \subset J_p.$$

We claim that

$$\log p \sum_{\substack{n\in N_p\\ n<p}} \mathrm{ord}_p(n! + f(n)) \leqslant \frac{1}{9\log(2/(1+\varepsilon_0))} |J_p| \log^2 x + O(x \log x). \tag{3.5}$$

If $N_p \cap [1, p) = \emptyset$, then there is nothing to prove. In the following, we assume that $T := \#(N_p \cap [1, p)) \geqslant 1$ and $N_p \cap [1, p) = \{n_1, n_2, \ldots, n_T\}$. Relabelling if necessary, we may assume that

$$\mathrm{ord}_p(n_1! + f(n_1)) \geqslant \mathrm{ord}_p(n_2! + f(n_2)) \geqslant \cdots \geqslant \mathrm{ord}_p(n_T! + f(n_T)). \tag{3.6}$$

If $|J_p| \geqslant 1$, then Lemma 2.2 implies that $T \leqslant C_0 |J_p|^{2/3} < C_0 x^{2/3}$. If $|J_p| < 1$, then $T \leqslant 1 < C_0 x^{2/3}$. Therefore, we always have

$$T < C_0 x^{2/3}. \tag{3.7}$$

Let $k^*$ be the least nonnegative integer such that

$$\left\lfloor \left(\frac{1+\varepsilon_0}{2}\right)^{k^*} T \right\rfloor < \left(\frac{10}{\varepsilon_0}\right)^{100}.$$

We denote

$$T_k := \left(\frac{1+\varepsilon_0}{2}\right)^{k} T \quad \text{for } k = 0, 1, \ldots, k^*.$$

https://doi.org/10.1017/S0004972725100543 Published online by Cambridge University Press

Note that

$$k^* = \frac{\log T}{\log(2/(1+\varepsilon_0))} + O(1). \tag{3.8}$$

For $0 \leqslant k < k^*$, by Lemma 2.7 (with $t = \lfloor T_k \rfloor$ and $J = J_p$) and (3.6),

$$\log p \sum_{i=\lfloor T_{k+1} \rfloor + 1}^{\lfloor T_k \rfloor} \mathrm{ord}_p(n_i! + f(n_i)) \leqslant \frac{|J_p|}{2} \log T_k + \frac{x \log x}{\lfloor T_k \rfloor^{0.98}} + O(x). \tag{3.9}$$

Since $\log(n! + f(n)) = O(x \log x)$ for $n_0 < n \leqslant x$ and $T_{k^*} = O(1)$,

$$\log p \sum_{i=1}^{\lfloor T_{k^*} \rfloor} \mathrm{ord}_p(n_i! + f(n_i)) = O(x \log x). \tag{3.10}$$

Summing (3.9) over $k = 0, 1, \ldots, k^* - 1$ and adding (3.10),

$$\begin{aligned} &\log p \sum_{i=1}^{T} \mathrm{ord}_p(n_i! + f(n_i)) \\ &\quad \leqslant \frac{|J_p|}{2} \sum_{k=0}^{k^*-1} \log T_k + x \log x \sum_{k=0}^{k^*-1} \frac{1}{\lfloor T_k \rfloor^{0.98}} + O(x \log x) + O(k^* x). \end{aligned} \tag{3.11}$$

(If $k^* = 0$, the empty sum $\sum_{k=0}^{k^*-1}$ is zero.) We estimate each term on the right-hand side of (3.11) as follows. Since

$$\begin{aligned} \sum_{k=0}^{k^*-1} \log T_k &= \sum_{k=0}^{k^*-1} \left( \log T - k \log\left(\frac{2}{1+\varepsilon_0}\right)\right) = k^* \log T - \frac{(k^*-1)k^*}{2} \log\left(\frac{2}{1+\varepsilon_0}\right) \\ &= \frac{1}{2\log(2/(1+\varepsilon_0))} \log^2 T + O(\log T) \quad \text{(by (3.8))} \\ &\leqslant \frac{2}{9\log(2/(1+\varepsilon_0))} \log^2 x + O(\log x) \quad \text{(by (3.7))}, \end{aligned}$$

we have

$$\frac{|J_p|}{2} \sum_{k=0}^{k^*-1} \log T_k \leqslant \frac{1}{9\log(2/(1+\varepsilon_0))} |J_p| \log^2 x + O(x \log x).$$

Since $T_k$ decreases exponentially with respect to $k$,

$$\sum_{k=0}^{k^*-1} \frac{1}{\lfloor T_k \rfloor^{0.98}} \ll \frac{1}{\lfloor T_{k^*-1} \rfloor^{0.98}} \ll 1.$$

https://doi.org/10.1017/S0004972725100543 Published online by Cambridge University Press

By (3.8) and (3.7),

$$k^*x = O(x\log x).$$

By substituting these estimates into (3.11), we deduce that

$$\log p\sum_{i=1}^{T}\operatorname{ord}_p(n_i!+f(n_i)) \leqslant \frac{1}{9\log(2/(1+\varepsilon_0))}|J_p|\log^2 x + O(x\log x),$$

which is exactly the statement in (3.5) that we claimed.

Combining (3.1), (3.3), (3.4) and (3.5),

$$(\log p)\operatorname{ord}_p(Z) \leqslant \frac{1}{9\log(2/(1+\varepsilon_0))}|J_p|\log^2 x + O(x\log x), \tag{3.12}$$

which holds for any prime number $p$. Since $P(Z) \leqslant \lambda x$ and $|J_p| \leqslant \min\{p, x\} - p/\lambda$, summing (3.12) over primes $p \leqslant \lambda x$, we obtain

$$\log Z \leqslant \frac{1}{9\log(2/(1+\varepsilon_0))}\bigg(\sum_{p\leqslant x}(1-1/\lambda)p + \sum_{x<p\leqslant\lambda x}(x - p/\lambda)\bigg)\log^2 x + O(x^2),$$

where for the error term, we have used $\pi(\lambda x) \ll x/\log x$. Now, we note that because $\pi(y) = (1+o(1))y/\log y$ and

$$\sum_{p\leqslant y} p = \Big(\frac{1}{2} + o(1)\Big)\frac{y^2}{\log y} \quad \text{as } y \to +\infty,$$

we can deduce that

$$\log Z \leqslant \Big(\frac{1}{9\log(2/(1+\varepsilon_0))}\frac{\lambda-1}{2} + o(1)\Big)x^2\log x \quad \text{as } x \to +\infty. \tag{3.13}$$

By comparing (2.3) with (3.13),

$$\frac{1}{9\log(2/(1+\varepsilon_0))}\frac{\lambda-1}{2} \geqslant \frac{1}{2} - 2\varepsilon_0,$$

which implies that

$$\lambda \geqslant 1 + 9(1-4\varepsilon_0)\log\Big(\frac{2}{1+\varepsilon_0}\Big).$$

However, this contradicts the assumptions $\lambda = 1 + 9\log 2 - 100\varepsilon_0$ and $\varepsilon_0 \in (0, 1/100)$.

Therefore, the lower asymptotic density of $B(\lambda)$ must be greater than or equal to $\varepsilon_0$. The proof of Theorem 1.1 is complete.

## Acknowledgement

The author is very grateful to the referee for helpful comments and suggestions. In particular, the author thanks the referee (and also Runbo Li) for kindly reminding him of the known improvements to Lemma 2.5.

## References


[1] P. Erdős and C. L. Stewart, 'On the greatest and least prime factors of $n! + 1$', *J. Lond. Math. Soc. (2)* **13**(3) (1976), 513–519.

[2] D. R. Heath-Brown, 'The difference between consecutive primes, III', *J. Lond. Math. Soc. (2)* **20**(2) (1979), 177–178.

[3] F. Luca and I. E. Shparlinski, 'Prime divisors of shifted factorials', *Bull. Lond. Math. Soc.* **37**(6) (2005), 809–817.

[4] F. Luca and I. E. Shparlinski, 'On the largest prime factor of $n! + 2^n - 1$', *J. Théor. Nombres Bordeaux* **17**(3) (2005), 859–870.

[5] J. Maynard, 'On the difference between consecutive primes', Preprint, 2012, arXiv:1201.1787 [math.NT].

[6] M. R. Murty and S. Wong, 'The *ABC* conjecture and prime divisors of the Lucas and Lehmer sequences', in: *Number Theory for the Millennium, III (Urbana, IL, 2000)* (eds. M. A. Bennett, B. Berndt, N. Boston, A. J. Hildebrand, H. G. Diamond and W. Philipp) (A. K. Peters, Natick, MA, 2002), 43–54.

[7] A. Peck, *On the Differences Between Consecutive Primes*, PhD Thesis, University of Oxford, 1996.

[8] J. Stadlmann, 'On the mean square gap between primes', Preprint, 2022, arXiv:2212.10867 [math.NT].

[9] C. L. Stewart, 'On the greatest and least prime factors of $n! + 1$, II', *Publ. Math. Debrecen* **65**(3–4) (2004), 461–480.



LI LAI, Xiamen University,
Fujian, China
e-mail: lilaimath@gmail.com


https://doi.org/10.1017/S0004972725100543 Published online by Cambridge University Press