On the number of prime factors of the k-fold iterate of the Euler function at consecutive arguments

Jean-Marie De Koninck[1] and Imre Kátai

Édition du 30 décembre 2016

Abstract

We study the distribution of the number of distinct prime factors of the k-fold iterate of the Euler totient function at consecutive arguments. We also examine the analogous problem for shifted primes.

AMS Subject Classification numbers: 11N37, 11N64, 11K65, 11N36
Key words: number of distinct prime factors, Euler function, shifted primes

1 Introduction and notation

For each integer $k \geq 1$, let $\varphi_k = \varphi \circ \varphi_{k-1}$, with $\varphi_0(n) = n$ for all $n \in \mathbb{N}$, stand for the k-fold iterate of the Euler φ function. Let also $\omega(n)$ stand for the number of distinct prime divisors of the integer $n \geq 2$, setting $\omega(1) = 0$.

Writing $\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt$ $(z \in \mathbb{R})$ for the normal distribution function, and setting

$$a_k = \frac{1}{(k+1)!}, \qquad b_k = \frac{1}{k!\sqrt{2k+1}} \qquad (k = 1, 2, \ldots),$$

Bassily, Kátai and Wijsmuller [1] proved that

$$\lim_{x\to\infty} \frac{1}{x} \# \left\{ n \leq x : \frac{\omega(\varphi_k(n)) - a_k (\log\log x)^{k+1}}{b_k\, (\log\log x)^{k+1/2}} < z \right\} = \Phi(z), \tag{1.1}$$

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \# \left\{ p \leq x : \frac{\omega(\varphi_k(p-1)) - (\log\log x)^{k+1}}{b_k\, (\log\log x)^{k+1/2}} < z \right\} = \Phi(z), \tag{1.2}$$

where $\pi(x)$ stands for the number of primes $p \leq x$.

As usual, we let $\mathrm{li}(x) := \int_2^x \frac{dt}{\log t}$ and let $\pi(x; k, \ell)$ be the number of primes $p \leq x$ such that $p \equiv \ell \pmod{k}$. We let $p(n)$ stand for the smallest prime factor of $n \geq 2$ and $P(n)$ for the largest prime factor of $n \geq 2$, with $p(1) = P(1) = 1$. For convenience, we shall write x_1 for $\max(1, \log x)$, x_2 for $\max(1, \log\log x)$, and so on. From here on,

[1]Research supported in part by a grant from NSERC.

the letter c, with or without subscript, stands for an absolute positive constant, but not necessarily the same at each occurrence, while the letters p, q, π, with or without subscript, will always denote primes.

2 Main results

Fix $k \in \mathbb{N}$. For each positive integer $n \leq x$, let

$$\ell_n := \frac{\omega(\varphi_k(n)) - a_k x_2^{k+1}}{b_k x_2^{k+\frac{1}{2}}}. \tag{2.1}$$

We can prove that, given distinct non zero integers $e_1, e_2, \ldots, e_r$ and arbitrary real numbers $z_1, z_2, \ldots, z_r$,

$$\lim_{x\to\infty} \frac{1}{x} \#\{n \leq x : \ell_{n+e_j} < z_j,\ j = 1, 2, \ldots, r\} = \Phi(z_1)\Phi(z_2)\cdots\Phi(z_r)$$

and

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \#\{p \leq x : \ell_{p+e_j} < z_j,\ j = 1, 2, \ldots, r\} = \Phi(z_1)\Phi(z_2)\cdots\Phi(z_r).$$

For the sake of simplicity, we will only prove the following two results.

Theorem 1. *Given arbitrary real numbers z_1 and z_2,*

$$\lim_{x\to\infty} \frac{1}{x} \#\{n \leq x : \ell_n < z_1,\ \ell_{n+1} < z_2\} = \Phi(z_1)\Phi(z_2). \tag{2.2}$$

Moreover, given any real number z,

$$\lim_{x\to\infty} \frac{1}{x} \#\{n \leq x : \ell_{n+1} - \ell_n < \sqrt{2}\, z\} = \Phi(z). \tag{2.3}$$

Theorem 2. *Given arbitrary real numbers z_1 and z_2,*

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \#\{p \leq x : \ell_{p-1} < z_1,\ \ell_{p+1} < z_2\} = \Phi(z_1)\Phi(z_2). \tag{2.4}$$

Moreover, given any real number z,

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \#\{p \leq x : \ell_{p+1} - \ell_{p-1} < \sqrt{2}\, z\} = \Phi(z). \tag{2.5}$$

3 Preliminary results

In preparation for the proof of Theorem 1, we introduce some preliminary results.

Let $f(t) := e^{-t^2/2}$ be the characteristic function of the Gaussian normal law. Using basic concepts from probability theory, it is easily seen that statement (1.1) is equivalent to the statement

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n\le x} e^{it\ell_n} = f(t), \tag{3.1}$$

where the convergence is uniform in $|t| \le R$ for any given real number $R > 0$, and that (2.2) is equivalent to the statement

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n\le x} e^{i(t_1\ell_n + t_2\ell_{n+1})} = f(t_1)f(t_2) \tag{3.2}$$

uniformly for $|t_1| \le R$, $|t_2| \le R$.

In [1], the authors introduced the following arithmetic functions. First, let θ be the completely multiplicative function defined on primes p by $\theta(p) = p - 1$. Then, define the k-fold iterate of θ by $\theta_0(n) = n$ and thereafter, for each integer $k \ge 1$, by $\theta_k(n) = \theta_{k-1}(\theta(n))$. Moreover, for each integer $k \ge 0$, consider the strongly additive function τ_k defined on primes p by

$$\tau_0(p) = 1, \qquad \tau_k(p) = \sum_{q|p-1} \tau_{k-1}(q),$$

so that in particular $\tau_0(n) = \omega(n)$. In [1], the authors proved (see Lemmas 5.1 and 5.2) that, given any arbitrarily small $\varepsilon > 0$,

$$\lim_{x\to\infty} \frac{1}{x} \left\{ n \le x : \left| \frac{\omega(\theta_k(n)) - \tau_k(n)}{x_2^{k+\frac{1}{2}}} \right| > \varepsilon \right\} = 0 \tag{3.3}$$

and

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \left\{ p \le x : \left| \frac{\omega(\theta_k(p+a)) - \tau_k(p+a)}{x_2^{k+\frac{1}{2}}} \right| > \varepsilon \right\} = 0, \tag{3.4}$$

where a is any given non zero integer.

Now, for each positive integer $n \le x$, let

$$h_n := \frac{\tau_k(n) - a_k x_2^{k+1}}{b_k\, x_2^{k+\frac{1}{2}}}.$$

Therefore, in light of (3.3), in order to prove Theorem 1, it is sufficient to prove (3.1) and (3.2) with h_n in place of ℓ_n.

Finally, before we proceed with the proof of Theorem 1, let us introduce the notion of a *k-chain* which was also introduced in [1]. We say that a $(k+1)$-tuple of primes $(q_0, q_1, \ldots, q_k)$ is a *k-chain* if $q_{i-1} \mid q_i - 1$ for $i = 1, 2, \ldots, k$, in which case we write

$$q_0 \to q_1 \to \cdots \to q_k.$$

We will need the following, which is a particular case of Lemma 2.5 in Bassily, Kátai and Wisjmuller [1].

Lemma 1. *Letting* $\delta(x,k) := \sum_{\substack{p \le x \\ p \equiv 1 \pmod{k}}} \frac{1}{p}$, *there exists an absolute constant* $c > 0$ *such that*

$$\delta(x,k) \le \frac{c x_2}{\varphi(k)},$$

provided $k \le x$ *and* $x \ge 3$.

We will also be using the following standard results from analytic number theory.

Lemma 2. (Brun-Titchmarsh Inequality) *There exists a positive constant* c *such that*

$$\pi(x;k,\ell) < c\frac{x}{\varphi(k)\log(x/k)} \qquad \text{for all } k < x.$$

Proof. For a proof, see the book of Halberstam and Richert [3]. □

Lemma 3. (Bombieri-Vinogradov Theorem) *Given any fixed number* $A > 0$, *there exists a number* $B = B(A) > 0$ *such that*

$$\sum_{k \le \sqrt{x}/(\log^B x)} \max_{(k,\ell)=1} \max_{y \le x} \left| \pi(x;k,\ell) - \frac{li(x)}{\varphi(k)} \right| = O\left(\frac{x}{\log^A x} \right).$$

Moreover, an appropriate choice for $B(A)$ *is* $2A + 6$.

Proof. For a proof, see the book of Iwaniec and Kowalski [4]. □

Before concluding this section, we state the following result of Elliott.

Lemma 4. *Let* $f(n)$ *be a real valued non negative arithmetic function. Let* a_n, $n = 1, \ldots, N$, *be a sequence of integers. Let* r *be a positive real number, and let* $p_1 < p_2 < \cdots < p_s \le r$ *be prime numbers. Set* $Q = p_1 \cdots p_s$. *If* $d|Q$, *then let*

$$\sum_{\substack{n=1 \\ a_n \equiv 0 \pmod{d}}}^{N} f(n) = \kappa(d) X + T(N,d), \tag{3.5}$$

where X *and* $T(N,d)$ *are real numbers,* $X \ge 0$, *and* $\kappa(d_1 d_2) = \kappa(d_1)\kappa(d_2)$ *whenever* d_1 *and* d_2 *are co-prime divisors of* Q.

Assume that for each prime p, $0 \le \kappa(p) < 1$. Setting

$$I(N,Q) := \sum_{\substack{n=1 \\ (a_n,Q)=1}}^{N} f(n),$$

then the estimate

$$I(N,Q) = \{1 + 2\theta_1 H\} X \prod_{p|Q} (1-\kappa(p)) + 2\theta_2 \sum_{\substack{d|Q \\ d\le z^3}} 3^{\omega(d)} |T(N,d)|$$

holds uniformly for $r \ge 2$, $\max(\log r, S) \le \frac{1}{8}\log z$, where $|\theta_1| \le 1$, $|\theta_2| \le 1$, and

$$H = \exp\left(-\frac{\log z}{\log r}\left\{\log\left(\frac{\log z}{S}\right) - \log\log\left(\frac{\log z}{S}\right) - \frac{2S}{\log z}\right\}\right) \tag{3.6}$$

and

$$S = \sum_{p|Q} \frac{\kappa(p)}{1-\kappa(p)} \log p.$$

When these conditions are satisfied, there exists an absolute positive constant c such that $2H \le c < 1$.

Proof. This result is Lemma 2.1 in the book of Elliott [2]. □

4 The proof of Theorem 1

Let x be a large number. Define Y_x implicitly by $\log Y_x = \exp\{-x_2^{1/3}\} \cdot x_1$ and consider the three sets

$$\begin{aligned} \mathcal{B} &:= \{n \le x : P(n) < x_2^{3k}\}, \\ \mathcal{E} &:= \{n \le x : x_2^{3k} \le p(n) \le P(n) \le Y_x\}, \\ \mathcal{D} &:= \{n \le x : p(n) > Y_x\}. \end{aligned}$$

We can then write each positive integer $n \le x$ as

$$n = B(n)\, C(n)\, D(n),$$

where $B(n) \in \mathcal{B}$, $C(n) \in \mathcal{E}$ and $D(n) \in \mathcal{D}$.

Using the concept of k-chain introduced in Section 3, it follows that

$$\tau_k(n) = \#\{q_0 \mid n : q_0 \to q_1 \to \cdots \to q_k\}.$$

Thus, using Lemma 1, we obtain that

$$\sum_{n\le x} \tau_k(B(n)) \le \sum_{\substack{q_0 \le x_2^{3k} \\ q_0 \to \cdots \to q_k}} \left\lfloor \frac{x}{q_k} \right\rfloor \le \sum_{\substack{q_0 \le x_2^{3k} \\ q_0 \to \cdots \to q_k}} \frac{x}{q_k} \le c x x_2^k x_4. \tag{4.1}$$

Similarly, we have

$$\sum_{n\le x}\tau_k(D(n)) \le \sum_{\substack{Y_x<q_k\le x\\ q_0\to q_1\cdots\to q_k}} \left\lfloor \frac{x}{q_k}\right\rfloor \le x \sum_{\substack{Y_x<q_k\le x\\ q_0\to q_1\cdots\to q_k}} \frac{1}{q_k}. \tag{4.2}$$

To estimate the right hand side of (4.2) for fixed $q_0, q_1, \dots, q_{k-1}$, observe that

$$\sum_{Y_x<q_k<x} \frac{1}{q_k} < \begin{cases} \frac{c}{q_{k-1}} \log\left(\frac{\log x}{\log Y_x}\right) \le c\frac{x_2^{1/3}}{q_{k-1}} & \text{if } q_{k-1}\le Y_x^{1/2},\\ \frac{cx_2}{q_{k-1}} & \text{if } q_{k-1} > Y_x^{1/2}. \end{cases} \tag{4.3}$$

We can then use (4.3) to estimate the inner sum in (4.2) and get

$$\sum_{\substack{Y_x<q_k\le x,\ q_{k-1}\le Y_x^{1/2}\\ q_0\to q_1\cdots\to q_k}} \frac{1}{q_k} < cx_2^{1/3} \sum_{q_0\to\cdots\to q_{k-1}\le x} \frac{1}{q_{k-1}} \tag{4.4}$$

and

$$\sum_{\substack{Y_x<q_k\le x,\ q_{k-1}> Y_x^{1/2}\\ q_0\to q_1\cdots\to q_k}} \frac{1}{q_k} < cx_2 \sum_{\substack{q_0\to\cdots\to q_{k-1}\le x\\ q_{k-1}>Y_x^{1/2}}} \frac{1}{q_{k-1}}. \tag{4.5}$$

Iterating the above procedure provided by (4.4) and (4.5), it follows that (4.2) yields

$$\sum_{n\le x}\tau_k(D(n)) \le cx x_2^{k+\frac{1}{3}}. \tag{4.6}$$

Combining estimates (4.6) and (4.1), we obtain that

$$\frac{1}{x}\#\{n\le x : |h_n - h_{C(n)}| > \varepsilon\} \to 0 \qquad (x\to\infty)$$

for every $\varepsilon > 0$, which allows us to claim that

$$\frac{1}{x}\sum_{n\le x} e^{ith_n} = \frac{1}{x}\sum_{n\le x} e^{ith_{C(n)}} + o(1) \qquad (x\to\infty) \tag{4.7}$$

uniformly for $|t|\le R$ for an arbitrary $R>0$.

Now, since one can easily see that, given a function r_x which tends to ∞ arbitrarily slowly as $x\to\infty$,

$$\frac{1}{x}\#\{n\le x : C(n) > Y_r^{r_x}\} \to 0 \qquad (x\to\infty), \tag{4.8}$$

we can now estimate

$$y_C := \frac{1}{x}\#\{n\le x : C(n) = C\}$$

for those $C \in \mathcal{E}$ such that $C \le Y_x^{r_x}$.

Using the Eratosthenian sieve we have that

$$y_C = (1+o(1)) \frac{x}{C} \prod_{x_2^{3k} \le \pi < Y_x} \left(1 - \frac{1}{\pi}\right) \qquad (x \to \infty). \tag{4.9}$$

Setting

$$\mu(Y) := \prod_{\pi < Y} \left(1 - \frac{1}{\pi}\right),$$

estimate (4.9) can be written as

$$y_C = (1+o(1)) \frac{x}{C} \frac{\mu(Y_x)}{\mu(x_2^{3k})} \qquad (x \to \infty). \tag{4.10}$$

Now, in light of (4.7), of the expression for $f(t)$ given in (3.1) and of our comments given in Section 3, we have that

$$f(t) = (1+o(1)) \frac{1}{x} \sum_{\substack{C \le Y_x^{r_x} \\ C \in \mathcal{E}}} y_C \, e^{ith_C} \qquad (x \to \infty). \tag{4.11}$$

Therefore, it follows from (4.10) and (4.11) that

$$f(t) = (1+o(1)) \frac{\mu(Y_x)}{\mu(x_2^{3k})} \sum_{\substack{C \le Y_x^{r_x} \\ C \in \mathcal{E}}} \frac{e^{ith_C}}{C} \qquad (x \to \infty). \tag{4.12}$$

Let us now consider the counting function

$$z(B_1, C_1; B_2, C_2) := \#\{n \le x : B(n) = B_1, C(n) = C_1, B(n+1) = B_2, C(n+1) = C_2\}.$$

If a particular integer $n \le x$ is counted by $z(B_1, C_1; B_2, C_2)$, then we must have $(B_1, B_2) = (C_1, C_2) = 1$. By the Eratosthenian sieve, we have

$$z(B_1, C_1; B_2, C_2) = (1+o(1)) \frac{x}{B_1 B_2 C_1 C_2} \prod_{\pi < Y_x} \left(1 - \frac{\rho(\pi)}{\pi}\right) \qquad (x \to \infty), \tag{4.13}$$

where $\rho(\pi)$ is the function defined for each prime $\pi < Y_x$ by

$$\rho(\pi) = \begin{cases} 1 & \text{if } \pi \mid B_1 C_1 B_2 C_2, \\ 2 & \text{if } \pi \nmid B_1 C_1 B_2 C_2. \end{cases} \tag{4.14}$$

Indeed, since n and $n+1$ can be written as

$$n = B_1 C_1 w_1, \quad n+1 = B_2 C_2 w_2, \qquad \text{for some } w_1, w_2 \in \mathcal{D}, \tag{4.15}$$

if we let $w_1^{(0)}$, $w_2^{(0)}$ correspond to the smallest solution of (4.15), then all the other solutions of (4.15) are those

$$w_1 = w_1^{(0)} + tB_2C_2, \qquad w_2 = w_2^{(0)} + tB_1C_1$$

for which $g(t) := (w_2^{(0)} + tB_1C_1)(w_1^{(0)} + tB_2C_2)$ belongs to $\mathcal{D}$, thus establishing (4.14).

It follows from this observation and (4.13) that, as $x \to \infty$,

$$\frac{1}{x}\sum_{B_1,B_2} z(B_1,C_1;B_2,C_2) = (1+o(1))\frac{1}{C_1C_2}\prod_{x_2^{3k}<\pi<Y_x}\left(1-\frac{\rho(\pi)}{\pi}\right)\cdot S_0(x), \tag{4.16}$$

where

$$S_0(x) = \sum_{(B_1,B_2)=1}\frac{1}{B_1B_2}\prod_{\pi\mid B_1B_2}\left(1-\frac{1}{\pi}\right)\prod_{\substack{\pi\nmid B_1B_2\\ \pi<x_2^{3k}}}\left(1-\frac{2}{\pi}\right). \tag{4.17}$$

Now, let us show that

$$S_0(x) = 1 + o(1) \qquad (x\to\infty). \tag{4.18}$$

Indeed, observe first that since $2 \mid B_1B_2$, it follows that the sum in $S_0(x)$ is symmetric in B_1, B_2. Therefore, we may assume that B_1 is even and B_2 is odd and then double the sum in the end. So, let $B_1 = 2^\alpha B_1'$, where $\alpha \geq 1$ and B_1' is odd, in which case we have

$$\begin{aligned}
S_0(x) &= 2\left(\frac{1}{2}+\frac{1}{2^2}+\cdots\right)\frac{1}{2}\sum_{(B_1',B_2)=1}\frac{1}{B_1'B_2}\prod_{\pi\mid B_1'B_2}\frac{1-1/\pi}{1-2/\pi}\cdot\prod_{2<\pi<x_2^{3k}}\left(1-\frac{2}{\pi}\right)+o(1)\\
&= \sum_{(B_1',B_2)=1}\frac{1}{B_1'B_2}\prod_{\pi\mid B_1'B_2}\frac{\pi-1}{\pi-2}\cdot\prod_{2<\pi<x_2^{3k}}\left(1-\frac{2}{\pi}\right)+o(1)\\
&= \prod_{2<\pi<x_2^{3k}}\left(1+2\left(\frac{1}{\pi}+\frac{1}{\pi^2}+\cdots\right)\frac{\pi-1}{\pi-2}\right)\cdot\prod_{2<\pi<x_2^{3k}}\left(\frac{\pi-2}{\pi}\right)+o(1)\\
&= \prod_{2<\pi<x_2^{3k}}\left(1+\frac{2}{\pi-2}\right)\cdot\prod_{2<\pi<x_2^{3k}}\left(\frac{\pi-2}{\pi}\right)+o(1) = 1+o(1),
\end{aligned}$$

where the term $o(1)$ comes from the fact that on the first and third of the above five lines of equations, we assumed that the sum of the reciprocals of the powers of 2 and the sum of the reciprocals of the powers of π were infinite series, while in reality they are finite sums. We have thus established (4.18). Using this, we can replace estimate (4.16) by

$$\frac{1}{x}\sum_{B_1,B_2} z(B_1,C_1;B_2,C_2) = (1+o(1))\frac{1}{C_1C_2}\prod_{x_2^{3k}<\pi<Y_x}\left(1-\frac{\rho(\pi)}{\pi}\right). \tag{4.19}$$

Writing the last product appearing in (4.19) as $L(C_1,C_2)$, we have

$$L(C_1,C_2) = \left(\frac{\mu(Y_x)}{\mu(x_2^{3k})}\right)^2 \cdot \prod_{\substack{x_2^{3k}<\pi<Y_x \\ \pi\nmid C_1C_2}} \frac{1-2/\pi}{(1-1/\pi)^2} \cdot \prod_{\substack{x_2^{3k}<\pi<Y_x \\ \pi\mid C_1C_2}} \frac{1-1/\pi}{(1-1/\pi)^2}$$

$$(4.20) \qquad = \left(\frac{\mu(Y_x)}{\mu(x_2^{3k})}\right)^2 \cdot U(C_1C_2)\cdot V(C_1C_2),$$

say. Now, on the one hand, it is clear that

$$\log U^{-1}(C_1C_2) \le c \sum_{\pi > x_2^{3k}} \frac{1}{\pi^2} \le \frac{c}{x_2^{3k}},$$

so that

$$(4.21) \qquad 1 \ge U(C_1C_2) \ge 1 - \frac{c}{x_2^{3k}}.$$

On the other hand, using the fact that $\omega(C_1) \le \tau_k(n) \ll x_2^{k+1}$ and $\omega(C_2) \le \tau_k(n+1) \ll x_2^{k+1}$, we have that the inequalities

$$(4.22) \qquad 0 \le \log V(C_1C_2) = \sum_{\substack{\pi > x_2^{3k} \\ \pi\mid C_1C_2}} \log\frac{1}{1-1/\pi} \le \sum_{\substack{\pi > x_2^{3k} \\ \pi\mid C_1C_2}} \frac{1}{\pi} \ll \frac{1}{x_2^2}$$

hold for almost all positive integers $n \le x$.

Substituting (4.21) and (4.22) in (4.20), we obtain that, as $x\to\infty$,

$$L(C_1,C_2) = (1+o(1))\left(\frac{\mu(Y_x)}{\mu(x_2^{3k})}\right)^2$$

for every $n \le x$ with at most $o(x)$ exceptions.

Using this last estimate, we have thus established that, as $x\to\infty$,

$$\frac{1}{x}\sum_{n\le x} e^{i(t_1h_n+t_2h_{n+1})} = \sum_{\substack{C_1,C_2\in\mathcal{E} \\ (C_1,C_2)=1}} \frac{e^{i(t_1h_{C_1}+t_2h_{C_2})}}{C_1C_2}\cdot(1+o(1))\left(\frac{\mu(Y_x)}{\mu(x_2^{3k})}\right)^2 + o(1)$$

$$= (1+o(1))\left\{\sum_{C_1\in\mathcal{E}} \frac{e^{it_1h_{C_1}}}{C_1}\frac{\mu(Y_x)}{\mu(x_2^{3k})}\right\}\left\{\sum_{C_2\in\mathcal{E}} \frac{e^{it_1h_{C_2}}}{C_2}\frac{\mu(Y_x)}{\mu(x_2^{3k})}\right\}$$

$$(4.23) \qquad + \quad o(1) + E(x),$$

where $E(x)$ stands for the error term generated from the product of those terms for which $C_1,C_2\in\mathcal{E}$ with $(C_1,C_2)>1$. In fact, the size of this error term can be estimated as follows.

$$(4.24) \qquad E(x) \le \left(\frac{\mu(Y_x)}{\mu(x_2^{3k})}\right)^2 \left\{\sum_{C_1',C_2'\in\mathcal{E}} \frac{1}{C_1'C_2'}\right\}\left\{\sum_{\substack{d>1 \\ p(d)>x_2^{3k}}} \frac{1}{d^2}\right\}.$$

It is clear that

$$\sum_{C_1',C_2'\in\mathcal{E}} \frac{1}{C_1'C_2'} \le \prod_{x_2^{3k}<\pi<Y_x} \left(1+\frac{2}{\pi}+\frac{2}{\pi^2}+\cdots\right),$$

so that

$$\left(\frac{\mu(Y_x)}{\mu(x_2^{3k})}\right)^2 \left\{\sum_{C_1',C_2'\in\mathcal{E}} \frac{1}{C_1'C_2'}\right\} = O(1), \tag{4.25}$$

while

$$\sum_{\substack{d>1\\ p(d)>x_2^{3k}}} \frac{1}{d^2} \ll \frac{1}{x_3}. \tag{4.26}$$

Substituting (4.25) and (4.26) in (4.24), it follows that $E(x)=o(1)$ as $x\to\infty$. Using this in (4.23), we have thus established that

$$\lim_{x\to\infty}\frac{1}{x}\sum_{n\le x} e^{i(t_1h_n+t_2h_{n+1})} = f(t_1)f(t_2) = e^{-\frac{t_1^2}{2}-\frac{t_2^2}{2}}, \tag{4.27}$$

thereby proving that (2.2) holds. To complete the proof of Theorem 1, it remains to prove (2.3). But this is a direct consequence of (2.2). Indeed, by choosing $t_2=-t_1$ in (4.27), it follows that

$$\lim_{x\to\infty}\frac{1}{x}\sum_{n\le x} e^{it(h_n-h_{n+1})} = e^{-t^2},$$

which is precisely the characteristic function of the Gaussian law with mean value 0 and variance $\sqrt{2}$. This establishes (2.3) and therefore completes the proof of Theorem 1.

5 The proof of Theorem 2

The proof of Theorem 2 goes along the same lines as that of Theorem 1. Therefore, we shall not provide all the details. The proof is essentially based on the Brun-Titchmarsh inequality, the Bombieri-Vinogradov theorem and a result of Elliott on arithmetic functions, which are all stated in Section 3.

First of all, let $\mathcal{B}$, $\mathcal{E}$, and $\mathcal{D}$ be the subsets of integers introduced in Section 4. We start by writing the shifted primes as

$$\begin{aligned} p-1 &= B(p-1)\,C(p-1)\,D(p-1), \\ p+1 &= B(p+1)\,C(p+1)\,D(p+1), \end{aligned}$$

where $B(p\pm 1)\in\mathcal{B}$, $C(p\pm 1)\in\mathcal{E}$, and $D(p\pm 1)\in\mathcal{D}$.

As earlier, we find that

$$\sum_{p\le x}\tau_k(B(p\pm 1)) = \sum_{\substack{q_k<x_2^{3k}\\ q_0\to\cdots\to q_k}}\pi(x;q_k,\pm 1)\ll \mathrm{li}(x)\cdot x_2^k\cdot x_3. \tag{5.1}$$

Observe that we can drop those primes $p\le x$ for which

$$P(p-1)>x^{1-1/x_3}\qquad\text{or}\qquad P(p+1)>x^{1-1/x_3}, \tag{5.2}$$

since the number of those $p\le x$ satisfying (5.2) is $o(\mathrm{li}(x))$, as $x\to\infty$. We can therefore assume that $q=P(p\pm 1)\le x^{1-1/x_3}$, in which case we have, by Lemma 2,

$$\pi(x;q,\pm 1)\le c\,\mathrm{li}(x)\cdot x_3.$$

Using this inequality and proceeding as we did to estimate the sum in (4.2) and obtain the upper bound given in (4.6), we find that

$$\begin{aligned}\sum_{P(p\pm1)\le x^{1-1/x_3}}\tau_k(D(p\pm 1)) &\le \sum_{\substack{Y_x<q_k\le x^{1-1/x_3}\\ q_0\to\cdots\to q_k}}\pi(x;q_k,\pm 1)\\ &\le c\,\mathrm{li}(x)x_3\sum_{\substack{Y_x<q_k\le x^{1-1/x_3}\\ q_0\to\cdots\to q_k}}\frac{1}{q_k}\\ &\le c\,\mathrm{li}(x)x_3\,x_2^{k+1/3}.\end{aligned}$$

As before, we can conclude from these inequalities that

$$\frac{1}{\pi(x)}\sum_{p\le x}e^{ith_{p\pm1}} = \frac{1}{\pi(x)}\sum_{p\le x}e^{ith_{C(p\pm1)}}+o(1)\qquad(x\to\infty). \tag{5.3}$$

We therefore need to estimate the two quantities

$$y_R^{(+,-)} := \#\{p\le x : C(p\pm1)=R\} = \#\left\{p\le x : R\mid p\pm 1,\ \left(\frac{p\pm 1}{R},\prod_{x_2^{3k}<\pi\le Y_x}\pi\right)=1\right\}.$$

To do so, we apply Lemma 4. To simplify the notation, we will only consider the case of the shifted primes $p+1$.

Using the notation of Lemma 4, we have

$$Q=\prod_{x_2^{3k}<\pi<Y_x}\pi,\qquad \sum_{\substack{p\le x\\ p+1\equiv 0 \pmod{dR}}}1=\pi(x;dR,-1),$$

implying that equation (3.5) will be written as

$$\pi(x;dR,-1)=\kappa(d)\pi(x;R,-1)+T(d,R), \tag{5.4}$$

where κ is the multiplicative function defined on prime powers p^α by

$$\kappa(p^\alpha) = \begin{cases} 1/\varphi(p^\alpha) & \text{if } (p,R)=1, \\ 1/p^\alpha & \text{if } p \mid R. \end{cases}$$

Setting

$$\Delta(x,k) := \max_{\substack{\ell \pmod k \\ (\ell,k)=1}} \left| \pi(x;k,\ell) - \frac{\mathrm{li}(x)}{\varphi(k)} \right|,$$

it follows from (5.4) that

$$T(d,R) \le \Delta(x,dR) + \kappa(d)\Delta(x,R). \tag{5.5}$$

Therefore, it follows from Lemma 4 that

$$\begin{aligned} \#\{p \le x : C(p+1) = R\} &= (1+2\theta_1 H)\pi(x;R,-1) \prod_{p \mid R} \left(1 - \frac{1}{p}\right) \prod_{\substack{p \nmid R \\ p \mid Q}} \left(1 - \frac{1}{p-1}\right) \\ &\quad + 2\theta_2 \sum_{\substack{d \le z^3 \\ d \mid Q}} 3^{\omega(d)} T(d,R). \end{aligned} \tag{5.6}$$

Then, with H as in (3.6), we have the following representation for S:

$$S = \sum_{p \mid Q} \frac{\kappa(p)}{1-\kappa(p)} \log p = \sum_{x_2^{3k} < p < Y_x} \frac{\log p}{p-2} + \sum_{p \mid R} \left(\frac{1}{p-1} - \frac{1}{p-2} \right) \log p. \tag{5.7}$$

Since the first of these last two sums is clearly $\log Y_x + O(x_3)$, while the second one is $O(1)$, it follows from (5.7) that

$$S = \log Y_x + O(x_3). \tag{5.8}$$

On the other hand, by definition,

$$r = \pi(Y_x) - \pi(x_2^{3k}),$$

so that

$$\log r = \log Y_x + O(x_2). \tag{5.9}$$

Choosing z so that $\log z = \dfrac{S}{\delta_x}$, where δ_x is a function tending to 0 very slowly, it follows from (5.8) and (5.9) that

$$\frac{\log z}{\log r} = \frac{1}{\delta_x} \frac{S}{\log r} = \frac{1}{\delta_x} \frac{\log Y_x + O(x_3)}{\log Y_x + O(x_2)} = \frac{1}{\delta_x} \left(1 + O\left(\frac{x_2}{x_1} \right) \right). \tag{5.10}$$

Using (5.10) in (3.6), we obtain

$$H = \exp\left\{-(1+o(1))\frac{1}{\delta_x}\left(\log(1/\delta_x) - \log\log(1/\delta_x) - 2\delta_x\right)\right\} \qquad (x\to\infty),$$

so that by choosing $\delta_x = 1/x_3$, we have that

$$H = \exp\{-(1+o(1))x_3x_4\} = o(1) \qquad \text{as } x\to\infty.$$

Using this last estimate along with (5.5), we obtain that, letting t_x be a function which tends to ∞ very slowly,

$$\begin{aligned}
\frac{1}{\pi(x)}\sum_{p\le x} e^{ith_{C(p+1)}} &= \frac{1}{\pi(x)}\sum_{R\in\mathcal{E}} e^{ith_R}\#\{p\le x : C(p+1)=R\}\\
&= (1+o(1))\sum_{\substack{R\in\mathcal{E},R\le Y_x^{t_x}\\ R\,\text{squarefree}}} e^{ith_R}\frac{\pi(x;R,-1)}{\pi(x)}\frac{\varphi(R)}{R}\prod_{\substack{p\nmid R\\ p\mid Q}}\left(1-\frac{1}{p-1}\right)\\
&\quad +\frac{2\theta_2}{\pi(x)}\sum_{\substack{R\in\mathcal{E},R\le Y_x^{t_x}\\ R\,\text{squarefree}}}\sum_{\substack{d\le z^3\\ d\mid Q}} z^{\omega(d)}T(d,R)\\
&= (1+o(1))S_1(x)+S_2(x), \qquad (5.11)
\end{aligned}$$

say. As we will see, the main contribution will come from $S_1(x)$.

We start by estimating $S_2(x)$. We have, using (5.5),

$$\begin{aligned}
S_2(x) &\le \frac{1}{\pi(x)}\left\{\sum_{\substack{R\in\mathcal{E},R\le Y_x^{t_x}\\ R\,\text{squarefree}}}\sum_{\substack{d\le z^3\\ d\mid Q}} z^{\omega(d)}\Delta(x,dR)+\kappa(d)\Delta(x,R)\right\}\\
&\le \frac{1}{\pi(x)}\left\{\sum_{\substack{R\in\mathcal{E},R\le Y_x^{t_x}\\ R\,\text{squarefree}}}\Delta(x,R)\sum_{\substack{d\le z^3\\ d\mid Q}}\kappa(d)+\sum_{M}\Delta(x,M)\sum_{\substack{d\mid M\\ d\mid Q}} z^{\omega(d)}\right\}. \qquad (5.12)
\end{aligned}$$

Since

$$\sum_{\substack{d\le z^3\\ d\mid Q}}\kappa(d)\le\sum_{d\le z^3}\frac{1}{\varphi(d)}\ll\log z$$

and

$$\sum_{\substack{d\mid M\\ d\mid Q}} z^{\omega(d)}\le 4^{\omega(M)},$$

it follows from (5.12) that

$$S_2(x)\ll\frac{1}{\pi(x)}\left\{\log z\sum_{\substack{R\in\mathcal{E},R\le Y_x^{t_x}\\ R\,\text{squarefree}}}\Delta(x,R)+\sum_{M\le Y_x^{t_x}z^3}4^{\omega(M)}\Delta(x,M)\right\}. \qquad (5.13)$$

From Lemma 3, we have that

$$\sum_{\substack{R\in\mathcal{E},R\le Y_x^{t_x}\\ R\,\text{squarefree}}} \Delta(x,R) \ll \frac{x}{\log^A x} \tag{5.14}$$

and that

$$\begin{aligned}\sum_{M\le Y_x^{t_x}z^3} 4^{\omega(M)}\Delta(x,M) &= \sum_{\substack{M\le Y_x^{t_x}z^3\\ \omega(M)\le 10x_2}} 4^{\omega(M)}\Delta(x,M) + \sum_{\substack{M\le Y_x^{t_x}z^3\\ \omega(M)>10x_2}} 4^{\omega(M)}\Delta(x,M)\\ &= S_3(x)+S_4(x). \end{aligned}\tag{5.15}$$

say. By Lemma 3,

$$S_3(x) \ll 4^{10x_2}\frac{x}{\log^A x}Y_x^{t_x}z^3 = o\left(\frac{x}{\log x}\right), \tag{5.16}$$

while using Lemma 2,

$$\begin{aligned} S_4(x) &\le c\frac{x}{\log x}\sum_{\substack{M\le Y_x^{t_x}z^3\\ \omega(M)>10x_2}} \frac{4^{\omega(M)}}{\varphi(M)} \ll \frac{x}{\log x}2^{-20x_2}\sum_{M\le Y_x^{t_x}z^3}\frac{4^{\omega(M)}}{\varphi(M)}\\ &\ll \frac{x}{\log x}2^{-20x_2}\prod_{p|Q}\left(1+\frac{8}{p-1}\right) \ll \frac{x}{\log x}2^{-20x_2}(\log(Y_x^{t_x}z^3))^8\\ &\ll \frac{x}{\log x}\frac{1}{(\log x)^{20\log 2}}t_x^8\cdot\log^8 x\cdot x_2^8 = o\left(\frac{x}{\log x}\right). \end{aligned}\tag{5.17}$$

Gathering estimates (5.16) and (5.17) in (5.15), which combined with (5.14) in (5.13) yields

$$S_2(x) = o(1) \qquad (x\to\infty). \tag{5.18}$$

On the other hand,

$$\begin{aligned} S_1(x) &= (1+o(1))\left\{\sum_{\substack{R\in\mathcal{E},R\le Y_x^{t_x}\\ R\,\text{squarefree}}} \frac{e^{ith_R}}{R}\prod_{p|R}\frac{p-1}{p-2}\right\}\cdot\prod_{p|Q}\left(1-\frac{1}{p-1}\right)\\ &\quad +O\left(\frac{1}{\pi(x)}\sum_{\substack{R\in\mathcal{E},R\le Y_x^{t_x}\\ R\,\text{squarefree}}}\Delta(x,R)\right). \end{aligned}\tag{5.19}$$

It is easily shown that

$$\prod_{p|Q}\left(1-\frac{1}{p-1}\right)\sum_{R|Q}\frac{1}{R}\left(\prod_{p|R}\frac{p-1}{p-2}-1\right)\to 0 \qquad (x\to\infty).$$

Therefore, from (5.19) and in light of (5.11) and (5.18), it follows that

$$\frac{1}{\pi(x)}\sum_{p\le x} e^{ith_{C(p+1)}} = \prod_{p\mid Q}\left(1-\frac{1}{p-1}\right)\sum_{\substack{R\mid Q\\ R\le Y_x^{t_x}}}\frac{e^{ith_R}}{R}+o(1)\qquad (x\to\infty). \tag{5.20}$$

Observe that (5.20) remains valid if we drop the two conditions $R\le Y_x^{t_x}$ and R squarefree, and if we replace $\prod_{p\mid Q}\left(1-\frac{1}{p-1}\right)$ by $\prod_{p\mid Q}\left(1-\frac{1}{p}\right)$.

Therefore,

$$\frac{1}{\pi(x)}\sum_{p\le x} e^{ith_{C(p+1)}} = \prod_{p\mid Q}\left(1-\frac{1}{p}\right)\sum_{R\in\mathcal{E}}\frac{e^{ith_R}}{R}+o(1)\quad (x\to\infty). \tag{5.21}$$

It was proved in [1] (see (1.2)) that, for each $t\in\mathbb{R}$,

$$\frac{1}{\pi(x)}\sum_{p\le x} e^{ith_{p-1}} \to e^{-t^2/2}\qquad (x\to\infty). \tag{5.22}$$

It is clear that this estimate is still true if we replace $p-1$ by $p+1$. Therefore,

$$\frac{1}{\pi(x)}\sum_{p\le x} e^{ith_{p+1}} \to e^{-t^2/2}\qquad (x\to\infty). \tag{5.23}$$

Hence, it follows from (5.21), (5.22), and (5.23) that, for each $\varepsilon>0$,

$$\frac{1}{\pi(x)}\#\{p\le x : |h_{p\pm1}-h_{C(p\pm1)}|>\varepsilon\}\to 0\qquad (x\to\infty),$$

which implies that, as $x\to\infty$,

$$\begin{aligned} e^{-t^2/2}+o(1) &= \frac{1}{\pi(x)}\sum_{p\le x} e^{ith_{C(p+1)}}\\ &= \prod_{\pi\mid Q}\left(1-\frac{1}{\pi}\right)\sum_{R\in\mathcal{E}}\frac{e^{ith_R}}{R}+o(1). \end{aligned}\tag{5.24}$$

Proceeding in the same manner, we obtain that, as $x\to\infty$,

$$\begin{aligned} e^{-t^2/2}+o(1) &= \frac{1}{\pi(x)}\sum_{p\le x} e^{ith_{C(p-1)}}\\ &= \prod_{\pi\mid Q}\left(1-\frac{1}{\pi}\right)\sum_{R\in\mathcal{E}}\frac{e^{ith_R}}{R}+o(1), \end{aligned}\tag{5.25}$$

Combining (5.24) and (5.25) completes the proof of (2.4).

It remains to prove (2.5). First write

$$\frac{1}{\pi(x)}\sum_{p\le x} e^{i(t_1h_{p+1}+t_2h_{p-1})} = \sum_{R_1,R_2\in\mathcal{E}} e^{i(t_1h_{R_1}+t_2h_{R_2})}\cdot W_{R_1,R_2}, \tag{5.26}$$

where

$$W_{R_1,R_2} = \frac{1}{\pi(x)}\#\left\{p\le x : R_1 \mid p+1, R_2 \mid p-1, \left(\frac{p+1}{R_1},Q\right)=1, \left(\frac{p-1}{R_2},Q\right)=1\right\}.$$

Using Lemma 4 and proceeding as in the proof of (2.4), we obtain that, assuming $(R_1,R_2)=1$ and $R_1,R_2\le Y_x^{t_x}$, we obtain that

$$W_{R_1,R_2} = \prod_{\pi\mid Q}\left(1-\frac{2}{\pi}\right)\frac{1}{R_1R_2} + E(R_1,R_2),$$

where the error term $E(R_1,R_2)$ is $o(1)$ as $x\to\infty$. Moreover, accounting for the fact that

$$\sum_{\max(R_1,R_2)>Y_x^{t_x} \text{ or } (R_1,R_2)>1} W_{R_1,R_2} = o(1) \qquad (x\to\infty),$$

we then deduce, as we did in the proof of Theorem 1, that the left hand side of (5.26) is

$$(1+o(1))e^{-t_1^2/2}\cdot e^{-t_2^2/2} \qquad (x\to\infty),$$

thus establishing (2.5) and therefore completing the proof of Theorem 2.

References

[1] N. L. Bassily, I. Kátai, M. Wijsmuller, *Number of prime divisors of $\varphi_k(n)$, where φ_k is the k-fold iterate of φ*, J. Number Theory **65** (1997), No. 2, 226–239.

[2] P.D.T.A. Elliott, *Probabilistic Number Theory I, Mean Value Theorems*, Springer-Verlag, Berlin, 1979.

[3] H.H. Halberstam and H.E. Richert, *Sieve Methods*, Academic Press, London, 1974.

[4] H. Iwaniec and E. Kowalski, *Analytic Number Theory*, AMS Colloquium Publications, Vol. 53, 2004.

Jean-Marie De Koninck
Dép. de mathématiques et de statistique
Université Laval
Québec
Québec G1V 0A6
Canada
jmdk@mat.ulaval.ca

Imre Kátai
Computer Algebra Department
Eötvös Loránd University
1117 Budapest
Pázmány Péter Sétány I/C
Hungary
katai@compalg.inf.elte.hu

JMDK, le 30 décembre 2016; fichier: iterates-euler-consecutive-arguments-dec-2016.tex