sciendo

uniform distribution theory

DOI: 10.2478/UDT-2023-0010
Unif. Distrib. Theory **18** (2023), no.2, 1–18

CREATING NORMAL NUMBERS USING THE PRIME DIVISORS OF CONSECUTIVE INTEGERS

Jean-Marie De Koninck[1] — Imre Kátai[2]

[1]Université Laval, Quebec, CANADA

[2]Eötvös Loránd University, Budapest, HUNGARY

ABSTRACT. For each integer $n \geq 2$, let $p_1 \leq p_2 \leq \cdots \leq p_k$ be the complete list of the prime factors of $a(n) := n(n+1)$. Consider the function $s_n : \{p_1, \ldots, p_k\} \to \{0, 1\}$ defined by $s_n(p_j) = 0$ if $p_j \mid n$ and 1 if $p_j \mid n+1$. Then consider the binary number $h(n) := s_n(p_1) \ldots s_n(p_k)$. In an earlier paper, we proved that the number $0.h(2)\, h(3)\, h(4) \ldots$ is a binary normal number and in fact we proved the more general statement when, for a fixed integer $t \geq 2$, we set $a(n) := n(n+1)\cdots(n+t-1)$, thus allowing for the construction of a normal number in base t. Here, we give a much shorter and simpler proof of this result and then we consider a more general result when $a(n)$ is the product of linear functions.

Communicated by Georges Grekos

1. Introduction

Given an integer $t \geq 2$, a *t-normal number*, or for short a *normal number*, is a real number whose t-ary expansion is such that any preassigned sequence, of length $k \geq 1$, of base t digits in this expansion, occurs at the expected frequency, namely $1/t^k$. Given an integer $r \geq 1$, we say that an expression of the form $i_1 i_2 \ldots i_r$, where each $i_j \in \{0, 1, \ldots, t-1\}$, is a *word* of length r. The symbol Λ will denote the *empty word*.

2010 Mathematics Subject Classification: 11K16.
Keywords: Normal numbers.
The work of the first author is supported by a grant from the Natural Sciences and Engineering Research Council of Canada.

In a series of papers (see [2] through [16]), we produced several methods for creating large families of normal numbers using the multiplicative structure of integers. In one of these [3], we proved the following result.

Theorem A. *Let $t \geq 2$ be a fixed integer and set $E(n) := n(n+1)\cdots(n+t-1)$. Moreover, for each positive integer n, set*

$$e(n) := \prod_{\substack{q^\beta \| E(n) \\ q \leq t-1}} q^\beta.$$

Define the sequence h_n on the prime powers q^a of $E(n)$ as follows:

$$h_n(q^a) = h_n(q) = \begin{cases} \Lambda & \text{if } q | e(n) \\ \ell & \text{if } q | n+\ell,\ \mathrm{GCD}\big(q, e(n)\big) = 1. \end{cases}$$

If $E(n) = q_1^{a_1} q_2^{a_2} \cdots q_r^{a_r}$, where $q_1 < q_2 < \cdots < q_r$ are primes and each $a_i \in \mathbb{N}$, then set

$$S\big(E(n)\big) = h_n(q_1) h_n(q_2) \cdots h_n(q_r).$$

Then, the number $0.S\big(E(1)\big)\, S\big(E(2)\big) \ldots S\big(E(n)\big) \ldots$ is a t-normal number.

The proof of Theorem A that we gave in [3] is lengthly and somewhat complicated. Here, we provide a shorter and simpler proof of Theorem A and thereafter, we show the more general result when $E(n)$ is the product of linear functions.

2. Main results

We start by a simple reformulation of Theorem A for the binary case.

Theorem 2.1. *For each integer $n \geq 2$, let $p_1 \leq p_2 \leq \cdots \leq p_k$ be the complete list of the prime factors of $n(n+1)$. Consider the function*

$$s_n : \{p_1, \ldots, p_k\} \to \{0, 1\}$$

defined by

$$s_n(p_j) = \begin{cases} 0 & \text{if } p_j \mid n, \\ 1 & \text{if } p_j \mid n+1 \end{cases}$$

and the corresponding binary number

$$h(n) := s_n(p_1) \ldots s_n(p_k).$$

Then, the number $\xi = 0.h(2)\, h(3)\, h(4) \ldots$ is a binary normal number.

Then comes a (simpler) formulation of Theorem A for the case of an arbitrary base $t \geq 2$.

Theorem 2.2. *Let $t \geq 2$ be a fixed integer, set $a(n) := n(n+1)\cdots(n+t-1)$ and write its factorisation as $a(n) = p_1 p_2 \cdots p_k$, where $p_1 \leq p_2 \leq \cdots \leq p_k$ are all the primes dividing $a(n)$. Consider the function $s_n : \{p_1, \ldots, p_k\} \to \{0, 1, \ldots, t-1\}$ defined by*

$$s_n(p) = \begin{cases} j & \text{if} \quad p \geq t \quad \text{and} \quad p \mid n+j, \\ \Lambda & \text{if} \quad p < t, \end{cases}$$

and set $h(n) := s_n(p_1)s_n(p_2)\ldots s_n(p_k)$. Then, the number $0.h(2)\,h(3)\,h(4)\ldots$ is a t-normal number.

Finally, we have the more general result when $a(n)$ is the product of linear functions.

Theorem 2.3. *Fix an integer $t \geq 2$. Consider the t linear functions $L_j(x) = a_j x + b_j$, $j = 0, 1, \ldots, t-1$, where*

$$\begin{aligned} a_i b_j - a_j b_i \neq 0 \qquad & \text{for all} \qquad 0 \leq i < j \leq t-1, \\ a_j, b_j > 0 \qquad & \text{for} \qquad j = 0, 1, \ldots, t-1. \end{aligned}$$

Set $a(n) := \prod_{j=0}^{t-1} L_j(n) = p_1 p_2 \cdots p_k$, where $p_1 \leq p_2 \leq \cdots \leq p_k$ are all the primes dividing $a(n)$. Moreover, let $N_0 > 0$ be a constant which is such that if $p > N_0$, then if $1 \leq i < j \leq t-1$, $\mathrm{GCD}\big(L_i(n), L_j(n)\big)$ is not a multiple of p. Then, let s_n be defined on the prime divisors of $a(n)$ by

$$s_n(p) = \begin{cases} j & \text{if } p \geq N_0 \quad \text{and} \quad p \mid L_j(n), \\ \Lambda & \text{if } p < N_0. \end{cases}$$

Moreover, let

$$h(n) := s_n(p_1)s_n(p_2)\ldots s_n(p_k).$$

Then the number $0.h(2)\,h(3)\,h(4)\ldots$ is a base t normal number.

3. Basic techniques on normal numbers

In the first section, we introduced the notion of "word". In the present section, we focus on words in base 2, namely binary words. A *binary word* is a finite sequence of 0's and 1's, that is a sequence of digits in base 2. Let E be the set of binary words. Given $\beta = b_1 b_2 \ldots b_k \in E$, we denote by $\lambda(\beta)$ the *length* of the word β, that is k, the number of binary digits that it contains. In particular, $\lambda(\Lambda) = 0$. Given $\kappa, \beta \in E$, we say that κ is a *subword* of β if there exist $u, v \in E$ such that $\beta = u\kappa v$. Given $\beta, \kappa \in E$, we let $\rho(\beta \mid \kappa)$ stand for the number of ways

of writing β as $\beta = u\kappa v$ for some $u, v \in E$. In particular, it is clear that if $\beta = \beta_1\beta_2$, then

$$\rho(\beta_1 \mid \kappa) + \rho(\beta_2 \mid \kappa) \le \rho(\beta_1\beta_2 \mid \kappa) \le \rho(\beta_1 \mid \kappa) + \rho(\beta_2 \mid \kappa) + \lambda(\kappa) - 1.$$

With this set up, we introduce the following proposition.

Proposition 3.1. *Let $\varepsilon_1, \varepsilon_2, \dots$ be an infinite binary sequence and consider the (finite) words*

$$A_M := \varepsilon_1\varepsilon_2 \dots \varepsilon_M \qquad (M = 1, 2, \dots)$$

and further consider the real number $\alpha := 0.\varepsilon_1\varepsilon_2 \dots$. If for every $\kappa \in E$, there exists an increasing sequence of integers $M_1, M_2, \dots$ such that

$$\lim_{r\to\infty} \frac{M_{r+1} - M_r}{M_r} = 0 \quad \text{and} \quad \lim_{r\to\infty} \frac{\rho(A_{M_r} \mid \kappa)}{M_r} = \frac{1}{2^{\lambda(\kappa)}}, \tag{1}$$

then,

$$\lim_{N\to\infty} \frac{\rho(A_N \mid \kappa)}{N} = \frac{1}{2^{\lambda(\kappa)}}, \tag{2}$$

implying that α is a binary normal number.

Proof. Fix an arbitrary $\kappa \in E$ and let $(M_r)_{r\in\mathbb{N}}$ be the corresponding sequence of integers satisfying (1). Let $N > 0$ be a large integer and let r be such that $M_r < N \le M_{r+1}$. Then, define D_N implicitly by $A_N = A_{M_r}D_N$. It follows that

$$\lambda(D_N) \le N - M_r \le M_{r+1} - M_r$$

and therefore,

$$\rho(A_{M_r} \mid \kappa) \le \rho(A_N \mid \kappa) \le \rho(A_{M_r} \mid \kappa) + \lambda(D_N),$$

which in turn implies that

$$\frac{\rho(A_{M_r} \mid \kappa)}{N} \le \frac{\rho(A_N \mid \kappa)}{N} \le \frac{\rho(A_{M_r} \mid \kappa)}{N} + \frac{\lambda(D_N)}{N}. \tag{3}$$

Since, as N becomes large and therefore r as well, we have

$$\frac{\rho(A_{M_r} \mid \kappa)}{M_{r+1}} \le \frac{\rho(A_{M_r} \mid \kappa)}{N} \quad \text{and} \quad \frac{\rho(A_{M_r} \mid \kappa)}{N} + \frac{\lambda(D_N)}{N} \le \frac{\rho(A_{M_r} \mid \kappa)}{M_r} + o(1),$$

it follows that (3) can be replaced by

$$\frac{\rho(A_{M_r} \mid \kappa)}{M_r} \cdot \frac{M_r}{M_{r+1}} \le \frac{\rho(A_N \mid \kappa)}{N} \le \frac{\rho(A_{M_r} \mid \kappa)}{M_r} + o(1). \tag{4}$$

Since $\dfrac{M_r}{M_{r+1}} = 1 + o(1)$ as $r \to \infty$ (due to the first relation in (1)) and because of the second relation in (1), we may replace (4) by

$$\frac{1}{2^{\lambda(\kappa)}}\big(1 + o(1)\big) \le \frac{\rho(A_N \mid \kappa)}{N} \le \frac{1}{2^{\lambda(\kappa)}}\big(1 + o(1)\big) \qquad (r \to \infty),$$

which clearly implies (2). □

4. Additional notation and preliminary results

From here on, the letters p and q with or without subscripts will always denote prime numbers.

We now state some classical results from prime number theory. We start with Mertens' theorem, which in fact can be formulated in three equivalent forms.

Theorem B (Mertens). *For large x, we have*

(i) $\displaystyle\sum_{p\le x}\frac{\log p}{p}=\log x+O(1)$,

(ii) $\displaystyle\sum_{p\le x}\frac{1}{p}=\log\log x+B+O\left(\frac{1}{\log x}\right)$ *for some constant B,*

(iii) $\displaystyle\prod_{p\le x}\left(1-\frac{1}{p}\right)=\frac{e^D}{\log x}\left(1+O\left(\frac{1}{\log x}\right)\right)$ *for some constant D.*

P r o o f. The above is Theorem 10.1 in the book of De Koninck and Doyon [1], where a detailed proof is given. □

Lemma 4.1. *Given real numbers $y>x\ge e$, we have*

(a) $\displaystyle\sum_{x<p\le y}\frac{\log p}{p}=\log y-\log x+O\left(\frac{1}{\log x}\right)$,

(b) $\displaystyle\sum_{x<p\le y}\frac{1}{p}=\log\left(\frac{\log y}{\log x}\right)+O\left(\frac{1}{\log x}\right)$.

P r o o f. To prove part (a), one will not succeed by simply using part (i) of Theorem B. A stronger estimate is required. In 1962, Rosser and Schoenfeld [19] proved that there exist constants $E<0$ and $a>0$ such that

$$\sum_{p\le x}\frac{\log p}{p}=\log x+E+O\left(\frac{1}{e^{a\sqrt{\log x}}}\right). \tag{5}$$

It is then clear that part (a) is an easy consequence of (5).

Part (b) is an immediate consequence of Theorem B (ii). □

Lemma 4.2. *Given $x\ge e^{e^e}$, let $\varepsilon(x)$ be a function which tends to 0 as $x\to\infty$ but at the same time satisfies $\varepsilon(x)\ge 1/\log\log\log x$. Further set*

$$z_1=z_1(x)=\exp\left\{(\log x)^{\varepsilon(x)}\right\}\qquad\text{and}\qquad z_2=z_2(x)=\exp\left\{(\log x)^{1-\varepsilon(x)}\right\}.$$

Then, for $j = 1, 2$, there exists an absolute constant $c > 0$ for which, for all $x \ge e^{e^e}$,

$$W_j(x; z_1, z_2) := \sum_{n \le x} \left(\sum_{\substack{p \mid n \\ p < z_1}} 1 \right)^j + \sum_{n \le x} \left(\sum_{\substack{p \mid n \\ z_2 < p \le x}} 1 \right)^j \le c\, \varepsilon(x)\, x\, (\log\log x)^j. \quad (6)$$

P r o o f. First consider the case $j = 1$. Using Theorem B(ii), we have

$$\begin{aligned} W_1(x; z_1, z_2) &= \sum_{p < z_1} \left\lfloor \frac{x}{p} \right\rfloor + \sum_{z_2 < p \le x} \left\lfloor \frac{x}{p} \right\rfloor \le x \sum_{p < z_1} \frac{1}{p} + x \sum_{z_2 < p \le x} \frac{1}{p} \\ &= x \left(\log \left((\log x)^{\varepsilon(x)} \right) \right) + O(x) \\ &\qquad + x \left(\log\log x - \log \left((\log x)^{1-\varepsilon(x)} \right) \right) + O(x) \\ &= \varepsilon(x)\, x\, \log\log x + O(x), \end{aligned}$$

thus proving (6) in the case $j = 1$.

For the case $j = 2$, we follow the method of proof of Theorem 7.2 in the book of De Koninck and Luca [17] by first writing that

$$\sum_{n \le x} \left(\sum_{\substack{p \mid n \\ p < z_1}} 1 \right)^2 = \sum_{n \le x} \left(\sum_{\substack{p \mid n \\ p < z_1}} 1 + 2 \sum_{\substack{pq \mid n \\ p < q < z_1}} 1 \right) = S_1(x) + 2S_2(x),$$

say, and similarly

$$\sum_{n \le x} \left(\sum_{\substack{p \mid n \\ z_2 < p \le x}} 1 \right)^2 = \sum_{n \le x} \left(\sum_{\substack{p \mid n \\ z_2 < p \le x}} 1 + 2 \sum_{\substack{pq \mid n \\ z_2 < p < q \le x}} 1 \right) = T_1(x) + 2T_2(x),$$

say.

Clearly,

$$S_1(x) \le \sum_{n \le x} \sum_{p \mid n} 1 = \sum_{n \le x} \omega(n) = O(x \log\log x), \quad (7)$$

where $\omega(n)$ stands for the number of distinct prime factors of n.

On the other hand,

$$\begin{aligned} S_2(x) &= \sum_{\substack{p<q<z_1\\ pq\le x}} \left\lfloor \frac{x}{pq} \right\rfloor \le x \sum_{p<q<z_1} \frac{1}{pq} \\ &= x \sum_{q<z_1} \frac{1}{q} \sum_{p<q} \frac{1}{p} \ll x \sum_{q<z_1} \frac{\log\log q}{q} \\ &= x \int_2^{z_1} \frac{\log\log t}{t}\, d\pi(t) \ll x \int_2^{z_1} \frac{\log\log t}{t\log t}\, dt \\ &= \frac{x}{2} (\log\log t)^2 \Big|_2^{z_1} \ll x(\log\log z_1)^2 \\ &= x(\varepsilon(x)\log\log x)^2 < x\varepsilon(x)(\log\log x)^2. \end{aligned} \tag{8}$$

Similarly as we did to obtain (7), one easily obtains that

$$T_1(x) = O(x\log\log x). \tag{9}$$

On the other hand,

$$\begin{aligned} T_2(x) &\le x \sum_{z_2<p\le x} \frac{1}{p} \sum_{p<q\le x} \frac{1}{q} \ll x \sum_{z_2<p\le x} \frac{1}{p} \log\log x \\ &\ll x\log\log x\,(\log\log x - \log\log z_2) \\ &= x\log\log x\,\big(\log\log x - \big(1-\varepsilon(x)\big)\log\log x\big) \\ &= x\log\log x\big(\varepsilon(x)\log\log x\big) \\ &= x\varepsilon(x)(\log\log x)^2. \end{aligned} \tag{10}$$

Gathering estimates (7), (8), (9) and (10) completes the proof of (6) in the case $j = 2$ and thereby the proof of Lemma 4.2. □

Lemma 4.3. *Let H be a large number. Then,*

$$\sum_{\substack{p<q\\ \frac{\log q}{\log p} < e^H}} \frac{\log^2 p}{p} = \sum_{q^{e^{-H}} < p < q} \frac{\log^2 p}{p} = \big(1 - e^{-2H}\big)\, \frac{\log^2 q}{2} + O(\log q). \tag{11}$$

P r o o f. Writing the sum Σ_q in (11) as a Stieltjes integral and then using the prime number theorem in the form

$$\pi(x) := \sum_{p\le x} 1 = \frac{x}{\log x} + O\left(\frac{x}{\log^2 x}\right),$$

we obtain that

$$\begin{aligned}\Sigma_q &= \int_{q^{e^{-H}}}^{q} \frac{\log^2 t}{t}\mathrm{d}\pi(t) = \pi(t)\frac{\log^2 t}{t}\Big|_{q^{e^{-H}}}^{q} \\ &\quad + \int_{q^{e^{-H}}}^{q} \pi(t)\frac{\log^2 t}{t^2}\left(1+O\left(\frac{1}{\log t}\right)\right)\mathrm{d}t \\ &= O(\log q) + \int_{q^{e^{-H}}}^{q} \frac{\log t}{t}\left(1+O\left(\frac{1}{\log t}\right)\right)\mathrm{d}t \\ &= \frac{1}{2}\left(1-e^{-2H}\right)\log^2 q + O(\log q),\end{aligned}$$

as claimed. □

Lemma 4.4. *Let $H > 0$ be a fixed positive number. Let $\tau_H(n)$ be the number of prime divisors p of $n(n+1)$ for which there exist no prime divisors of $n(n+1)$ located in the interval $\left(p, p^{e^H}\right)$. Then, for $j = 1, 2$, there exists an absolute constant $c > 0$ such that*

$$\sum_{n\le x} \tau_H(n)^j \le \frac{c}{H}\, x\,(\log\log x)^j. \tag{12}$$

P r o o f. In the case of $j = 1$, using standard sieve techniques as well as Theorem B (iii), we have that

$$\sum_{n\le x} \tau_H(n) \ll \sum_{p\le x} \frac{x}{p} \prod_{p<q<p^{e^H}} \left(1-\frac{1}{q}\right) \ll x \sum_{p\le x} \frac{1}{p}\cdot\frac{1}{e^H} \ll \frac{x}{e^H}\log\log x,$$

which clearly implies (12) in the case $j = 1$. The case $j = 2$ can be handled in a similar manner. □

5. The sketch of the proof of Theorem 2.1

Given an integer $n \ge 2$ with $n(n+1) = p_1 p_2 \cdots p_k$, it is clear that

$$k = \Omega(n) + \Omega(n+1),$$

where $\Omega(n)$ stands for the number of prime factors of n counting their multiplicities.

Now, given a word $\delta_1 \dots \delta_\ell \in \{0,1\}^\ell$ with $\ell \le k$, we let $K(n \mid \delta_1 \dots \delta_\ell)$ stand for the number of occurrences of the word $\delta_1 \dots \delta_\ell$ in $h(n)$, that is,

$$K(n \mid \delta_1 \dots \delta_\ell) := \#\big\{j \in \{1, \dots, k-r\} : s_n(p_{j+r}) = \delta_r \quad \text{with} \quad r = 1, \dots, \ell\big\}.$$

In particular, it is easy to see that

$$\sum_{\delta_1\dots\delta_\ell\in\{0,1\}^\ell} K(n \mid \delta_1\dots\delta_\ell) = \Omega(n)+\Omega(n+1)-(\ell-1)$$
$$= k-(\ell-1).$$

It is known (see for instance Elliott [18]) that, given any fixed numbers $\delta>0$ and $\varepsilon>0$,

$$\lim_{x\to\infty}\frac{1}{x}\#\left\{n\le x : \frac{|\Omega(n)-\log\log n|}{(\log\log n)^{\frac{1}{2}+\delta}}>\varepsilon\right\}=0. \tag{13}$$

Our goal is to show that, for every word $\delta_1\dots\delta_\ell\in\{0,1\}^\ell$, we have

$$K(n \mid \delta_1\dots\delta_\ell) = \big(1+o(1)\big)\frac{\Omega(n)+\Omega(n+1)}{2^\ell}$$

for almost all $n\le x$ as $x\to\infty$, which amounts, in light of (13), to show that

$$K(n \mid \delta_1\dots\delta_\ell) = \big(1+o(1)\big)\frac{\log\log x}{2^{\ell-1}} \tag{14}$$

for almost all $n\le x$ as $x\to\infty$.

For each word $\delta_1\dots\delta_\ell\in\{0,1\}^\ell$, we set

$$T(x \mid \delta_1\dots\delta_\ell) := \sum_{n\le x} K(n \mid \delta_1\dots\delta_\ell).$$

It then follows from (14) that

$$T(x \mid \delta_1\dots\delta_\ell) = \big(1+o(1)\big)x\frac{\log\log x}{2^{\ell-1}} \qquad (x\to\infty) \tag{15}$$

and actually more is true as we will prove that

$$\sum_{n\le x}\left(K(n \mid \delta_1\dots\delta_\ell)-\frac{\log\log x}{2^{\ell-1}}\right)^2 = o\big(x(\log\log x)^2\big). \tag{16}$$

Clearly, as a consequence of (16), we have that

$$\sum_{n\le x}\left|K(n \mid \delta_1\dots\delta_\ell)-\frac{\log\log x}{2^{\ell-1}}\right| = o(x\log\log x) \qquad (x\to\infty), \tag{17}$$

which means that for any fixed number $\varepsilon>0$, we have

$$\lim_{x\to\infty}\frac{1}{x}\#\left\{n\le x : \max_{\delta_1\dots\delta_\ell\in\{0,1\}^\ell}\left|\frac{K(n \mid \delta_1\dots\delta_\ell)}{2\log\log x}-\frac{1}{2^\ell}\right|>\varepsilon\right\}=0,$$

from which, as we will now see, it will follow that the number ξ of Theorem 2.1 is indeed a binary normal number.

To see this, we proceed as follows. Recall that

$$\xi = 0.h(2)\,h(3)\ldots = 0.\epsilon_1\epsilon_2\epsilon_3\ldots,$$

where each $\epsilon_i \in \{0,1\}$, and also set

$$A_N := \epsilon_1\epsilon_2\ldots\epsilon_N.$$

Further set

$$C_r := h(2)\,h(3)\ldots h(r) \qquad (r = 2,3,\ldots).$$

We know that

$$\begin{aligned}\lambda(C_r) &= \sum_{j=2}^{r} \lambda\big(h(j)\big)\\ &= 2\sum_{i=2}^{r}\Omega(i) + O(\log r)\\ &= 2r\log\log r + O(r). \end{aligned} \tag{18}$$

Consequently, if for the sequence $(M_r)_{r\in\mathbb{N}}$ appearing in Proposition 3.1, we choose

$$M_r := \lambda(C_r),$$

it will follow from (18) that

$$\lim_{r\to\infty}\frac{M_{r+1}}{M_r} = 1.$$

On the other hand, it is clear that for an arbitrary word κ, we have

$$\sum_{j=2}^{r}\rho\big(h(j)\mid\kappa\big) \le \rho(C_r\mid\kappa) \le \sum_{j=2}^{r}\rho(h(j)\mid\kappa) + r\lambda(\kappa). \tag{19}$$

Using inequalities (17) and (19), we have that

$$\frac{\rho(C_r\mid\kappa)}{M_r} \to \frac{1}{2^{\lambda(\kappa)}} \qquad (r\to\infty). \tag{20}$$

Applying Proposition 3.1, we may conclude that

$$\lim_{N\to\infty}\frac{\rho(A_N\mid\kappa)}{N} = \lim_{r\to\infty}\frac{\rho(C_r\mid\kappa)}{r} = \frac{1}{2^{\lambda(\kappa)}},$$

thus establishing that ξ is indeed a binary normal number.

6. The proofs of the various steps

Given positive integers $m < n$, we set

$$Q(m,n) := \prod_{m<p<n} p.$$

In order to prove (15), we start with the case of $K(n \mid 01)$. First observe that setting $T(x \mid 01) := \sum_{n\le x} K(n \mid 01)$, we have

$$\begin{aligned} T(x \mid 01) =& \sum_{p<q\le x} \#\big\{n \le x : p \,|\, n, q \,|\, n+1 \text{ and } \operatorname{GCD}\big(n(n+1), Q(p,q)\big) = 1\big\} \\ =:& \sum_{p<q\le x} S(x;p,q), \end{aligned}$$

say. We separate the above sum in two parts, as follows

$$T_1(x \mid 01) := \sum_{\substack{z_1<p<q<z_2 \\ \frac{\log q}{\log p} < e^H}} S(x;p,q) \text{ and } T_2(x \mid 01) := T(x \mid 01) - T_1(x \mid 01), \tag{21}$$

where z_1 and z_2 were defined in Lemma 4.2. It follows from Lemmas 4.2 and 4.4 that there exist absolute constants $c_2 > 0$ and $c_3 > 0$ for which we have

$$T_2(x \mid 01) \le \Big(c_2\varepsilon(x) + \frac{c_3}{H}\Big)\, x \,\log\log x. \tag{22}$$

Using sieve theory techniques and Lemma 4.1, one can prove that, as $x \to \infty$,

$$\begin{aligned} S(x;p,q) &= \big(1+o(1)\big)\frac{x}{pq}\prod_{p<\pi<q}\left(1-\frac{2}{\pi}\right) \\ &= \big(1+o(1)\big)\frac{x}{pq}\exp\left\{-2\sum_{p<\pi<q}\frac{1}{\pi} + O\left(\sum_{p<\pi<q}\frac{1}{\pi^2}\right)\right\} \\ &= \big(1+o(1)\big)\frac{x}{pq}\exp\left\{-2\log\left(\frac{\log q}{\log p}\right) + O\left(\frac{1}{\log p}\right)\right\} \\ &= \big(1+o(1)\big)\frac{x}{pq}\frac{\log^2 p}{\log^2 q}. \end{aligned} \tag{23}$$

Using (23) in the representation of $T_1(x \mid 01)$ given in (21), we get that

$$T_1(x \mid 01) = \big(1+o(1)\big)x\sum_{q<z_2}\frac{1}{q\log^2 q}\sum_{z_1<p<q<p^{e^H}}\frac{\log^2 p}{p} \qquad (x\to\infty). \tag{24}$$

Using Lemma 4.3 in (24), we obtain that, as $x \to \infty$,

$$T_1(x \mid 01) = \big(1+o(1)\big)\big(1-e^{-2H}\big)\,\frac{x}{2}\sum_{q<z_2}\frac{1}{q} = \big(1+o(1)\big)\big(1-e^{-2H}\big)\,\frac{x}{2}\log\log x,$$

which, combined with estimate (22) and since H can be taken arbitrarily large, proves that

$$T_1(x \mid 01) = \big(1+o(1)\big)\frac{x}{2}\log\log x \qquad (x\to\infty),$$

thus establishing (15) in the case of $\ell = 2$.

Let us now estimate the expression

$$E(x \mid 01) \;:=\; \sum_{n\le x}\left(\sum_{\substack{z_1<p<q<z_2\\ p\mid n,\ q\mid n+1\\ (n(n+1),Q(p,q))=1\\ \frac{\log q}{\log p}<e^H}} 1\right)^2. \tag{25}$$

Recalling the definition of $T_1(x \mid 01)$ given above, we may write that

$$\begin{aligned} E(x \mid 01) &= T_1(x \mid 01) + 2\sum_{n\le x}\sum_{\substack{z_1<p_1<q_1<p_2<q_2<z_2\\ p_1p_2\mid n,\ q_1q_2\mid n+1\\ (n(n+1),Q(p,q))=1\\ \frac{\log q_j}{\log p_j}<e^H,\ j=1,2}} 1 \\ &= T_1(x \mid 01) + 2\sum_{\substack{z_1<p_1<q_1<p_2<q_2<z_2\\ \frac{\log q_j}{\log p_j}<e^H,\ j=1,2}} R(x;p_1,p_2,q_1,q_2), \end{aligned} \tag{26}$$

say. Recalling the definition of $S(x;p,q)$ given above, we obtain that

$$\frac{R(x;p_1,p_2,q_1,q_2)}{x} = \big(1+o(1)\big)\frac{S(x;p_1,q_1)}{x}\cdot\frac{S(x;p_2,q_2)}{x} \qquad (x\to\infty),$$

which substituted in (26) implies that

$$\frac{E(x \mid 01)}{x} = \big(1+o(1)\big)\left(\sum_{\substack{z_1<p<q<z_2\\ \frac{\log q}{\log p}<e^H}}\frac{S(x;p,q)}{x}\right)^2 + o\big((\log\log x)^2\big) \qquad (x\to\infty).$$

Using the above and applying Lemmas 4.2 and 4.4, we then obtain that

$$\frac{1}{x}\sum_{n\le x}\left(\sum_{\substack{p\mid n,\ q\mid n+1\\ p<q\\ (n(n+1),Q(p,q))=1}} 1-\frac{1}{2}\log\log x\right)^2 = O\left(\frac{1}{H}(\log\log x)^2\right). \tag{27}$$

Since H can be chosen arbitrarily large, it follows from (27) that

$$\frac{1}{x}\sum_{n\le x}\left(\sum_{\substack{p\mid n,\ q\mid n+1\\ p<q\\ (n(n+1),Q(p,q))=1}} 1-\frac{1}{2}\log\log x\right)^2 = o\left((\log\log x)^2\right) \qquad (x\to\infty), \tag{28}$$

which proves (16) in the case $\delta_1\delta_2 = 01$ and therefore (17) as well.

So far, we have only considered the particular case $\delta_1\delta_2 = 01$. Clearly the same holds when $\delta_1\delta_2 = 00$ or 10 or 11.

For the general case, we proceed as follows. Given $\delta_1\dots\delta_\ell \in \{0,1\}^\ell$, we set

$$\begin{aligned}
&\nu(n \mid \delta_1\dots\delta_\ell)\\
&:= \sum_{p_1<\dots<p_\ell} \#\left\{p_j \mid n+\delta_j, j=1,\dots,\ell, \text{ with } \left(\frac{n(n+1)}{p_1\cdots p_\ell}, Q(p_1,p_\ell)\right)=1\right\}\\
&\nu_1(n \mid \delta_1\dots\delta_\ell)\\
&:= \sum_{\substack{z_1<p_1<\dots<p_\ell<z_2\\ \frac{\log p_{j+1}}{\log p_j}<e^H,\ j=1,\dots,\ell-1}} \#\left\{p_j \mid n+\delta_j, \text{ with } \left(\frac{n(n+1)}{p_1\cdots p_\ell}, Q(p_1,p_\ell)\right)=1\right\}\\
&\nu_2(n \mid \delta_1\dots\delta_\ell)\\
&:= \nu(n \mid \delta_1\dots\delta_\ell) - \nu_1(n \mid \delta_1\dots\delta_\ell)
\end{aligned}$$

and we also set

$$\begin{aligned}
T(x \mid \delta_1\dots\delta_\ell) &:= \sum_{n\le x}\nu(n \mid \delta_1\dots\delta_\ell)\\
T_j(x \mid \delta_1\dots\delta_\ell) &:= \sum_{n\le x}\nu_j(n \mid \delta_1\dots\delta_\ell) \qquad (j=1,2).
\end{aligned}$$

It follows from Lemmas 4.2 and 4.4 that, for some absolute constant $c>0$,

$$\frac{1}{x}\sum_{n\le x}\nu_2(n \mid \delta_1\dots\delta_\ell)^j \le \frac{c}{H}(\log\log x)^j \qquad (j=1,2). \tag{29}$$

Similarly, proceeding as we did for estimating $T(x \mid 01)$, we first write that

$$T(x \mid \delta_1 \dots \delta_\ell) = T_1(x \mid \delta_1 \dots \delta_\ell) + T_2(x \mid \delta_1 \dots \delta_\ell) = T_1(x) + T_2(x),$$

say, where, as $x \to \infty$,

$$\begin{aligned} \frac{T_1(x)}{x} &= \big(1+o(1)\big) \sum_{\substack{z_1<p_1<\dots<p_\ell<z_2 \\ \frac{\log p_{j+1}}{\log p_j}<e^H,\ j=1,\dots,\ell-1}} \frac{1}{p_1\cdots p_\ell} \prod_{p_1<\pi<p_\ell}\left(1-\frac{2}{\pi}\right) \\ &= \big(1+o(1)\big) \sum_{\substack{z_1<p_1<\dots<p_\ell<z_2 \\ \frac{\log p_{j+1}}{\log p_j}<e^H,\ j=1,\dots,\ell-1}} \frac{1}{p_1\cdots p_\ell}\frac{\log^2 p_1}{\log^2 p_\ell}. \end{aligned} \tag{30}$$

Making repetitive use of Lemma 4.3 in evaluating the sum in (30), we obtain that

$$\begin{aligned} \frac{T_1(x)}{x} &= \big(1+o(1)\big)\frac{1-e^{-2H}}{2} \cdot \sum_{\substack{z_1<p_2<\dots<p_\ell<z_2 \\ \frac{\log p_{j+1}}{\log p_j}<e^H,\ j=2,\dots,\ell-1}} \frac{1}{p_2\cdots p_\ell}\frac{\log^2 p_2}{\log^2 p_\ell} \\ &\vdots \\ &= \big(1+o(1)\big)\left(\frac{1-e^{-2H}}{2}\right)^{\ell-1} \sum_{z_1<q<z_2}\frac{1}{q} \\ &= \big(1+o(1)\big)\left(\frac{1-e^{-2H}}{2}\right)^{\ell-1} \log\log x. \end{aligned} \tag{31}$$

On the other hand, as in the case of $\ell = 2$, in light of (29), it is clear that

$$T_2(x) = O\left(\frac{\log\log x}{H}\right). \tag{32}$$

Combining (31) and (32), and observing that H can be chosen arbitrarily large, we may conclude that

$$\frac{T(x)}{x} = \big(1+o(1)\big)\frac{\log\log x}{2^{\ell-1}}. \tag{33}$$

It remains to prove that

$$\frac{1}{x}\sum_{n\le x}\left(\nu(n \mid \delta_1 \dots \delta_\ell) - \frac{\log\log x}{2^{\ell-1}}\right)^2 = o\big((\log\log x)^2\big) \quad (x\to\infty). \tag{34}$$

First observe that, in light of Lemmas 4.2 and 4.4, it will be sufficient to prove that

$$\frac{1}{x}\sum_{n\le x}\left(\nu_1(n\mid \delta_1\dots\delta_\ell)-\frac{\log\log x}{2^{\ell-1}}\right)^2 = O\left(\frac{(\log\log x)^2}{H}\right). \tag{35}$$

To do so, we start by observing that the left hand side of (35) is equal to

$$\begin{aligned}&\frac{1}{x}\sum_{n\le x}\nu_1(n\mid \delta_1\dots\delta_\ell)^2\\ &\quad -2\frac{\log\log x}{2^{\ell-1}}\frac{1}{x}\sum_{n\le x}\nu_1(n\mid \delta_1\dots\delta_\ell)+\frac{\lfloor x\rfloor}{x}\left(\frac{\log\log x}{2^{\ell-1}}\right)^2\\ &=\frac{1}{x}\Sigma_1-2\frac{\log\log x}{2^{\ell-1}}\cdot\frac{1}{x}T_1(x)+\frac{\lfloor x\rfloor}{x}\left(\frac{\log\log x}{2^{\ell-1}}\right)^2,\end{aligned}$$

say. We already know from (31) that

$$\frac{T_1(x)}{x}=\big(1+o(1)\big)\left(1+O\left(\frac{1}{H}\right)\right)\frac{\log\log x}{2^{\ell-1}} \qquad (x\to\infty). \tag{36}$$

Therefore, in order to prove (35), it will be enough to show that

$$\frac{1}{x}\Sigma_1=\big(1+o(1)\big)\cdot\left(\frac{\log\log x}{2^{\ell-1}}\right)^2\cdot\left(1+O\left(\frac{1}{H}\right)\right) \quad (x\to\infty). \tag{37}$$

By definition, we have that

$$\nu_1(n\mid \delta_1\dots\delta_\ell)^2 \;=\; \sum_{\substack{p_j\mid n+\delta_j,\ q_j\mid n+\delta_j\ ,j=1,\dots,\ell\\ z_1<p_1<\dots<p_\ell<z_2,\ z_1<q_1<\dots<q_\ell<z_2\\ \frac{\log p_{j+1}}{\log p_j}<e^H,\ \frac{\log q_{j+1}}{\log q_j}<e^H,\ j=1,\dots,\ell-1}} 1.$$

Notice that the above sum runs over two ℓ-tuples of increasing primes, namely

$$p_1<\dots<p_\ell \quad\text{and}\quad q_1<\dots<q_\ell.$$

For each such pair of ℓ-tuples, let us consider their intersection

$$I:=\{p_1,\dots,p_\ell\}\cap\{q_1,\dots,q_\ell\}.$$

The contribution to Σ_1 of those pairs for which $I\neq\emptyset$ can be neglected. To see this, let us assume, for simplicity, that $\delta_1=0$, in which case we have that $p_1\mid n$. This means that q_1 must be equal to one of the primes $p_2,\dots,p_\ell$, allowing for only $\ell-1$ possibilities. Hence, the number of those pairs of ℓ-tuples for which

the corresponding set I is non empty does not exceed $(\ell-1)\Omega\big(n(n+1)\big)$, and therefore their contribution to Σ_1 does not exceed

$$\frac{1}{x}\sum_{n\le x} c\big(\Omega(n)+\Omega(n+1)\big) \ll \log\log x$$

for some constant c which may depend on ℓ.

We can therefore assume that $I=\emptyset$ and without loss of generality that $p_\ell < q_1$, in which case we have

$$\frac{1}{x}\Sigma_1 = \frac{1}{x} \sum_{\substack{z_1<p_1<\cdots<p_\ell<q_1<\cdots<q_\ell<z_2 \\ \frac{\log p_{j+1}}{\log p_j}<e^H,\ \frac{\log q_{j+1}}{\log q_j}<e^H,\ j=1,\dots,\ell-1}} \sum_{\substack{n\le x \\ p_j\mid n+\delta_j,\ q_j\mid n+\delta_j \\ j=1,\dots,\ell}} 1 + O(\log\log x).$$

Then, following the same pattern that lead to estimate (33), we finally obtain that

$$\frac{1}{x}\Sigma_1 = \left(1+O\left(\frac{1}{H}\right)\right)\frac{(\log\log x)^2}{2^{2(\ell-1)}},$$

which allows us to conclude that (37) holds, and therefore (35) and (34) as well, thus proving our claim.

7. The proof of Theorem 2.2

The proof of Theorem 2.2 runs along the same lines as that of Theorem 2.1. The only difference is that we need to treat separately the prime divisors of $a(n)=n(n+1)\cdots(n+t-1)$ which are smaller than t, and in fact ignore them altogether in listing the digits that form the number $0.h(2)\,h(3)\,h(4)\dots$

8. The proof of Theorem 2.3

In order to avoid repeating the same arguments as in the proof of Theorem 2.1, we only provide a sketch of the proof.

Given an integer $n\ge 2$ and a word $\delta_1\dots\delta_\ell\in\{0,1,\dots,t-1\}^\ell$, we set

$$\widetilde{K}(n\mid \delta_1\dots\delta_\ell) = \#\{j\in\{0,1,\dots,k-r\} : s_n(p_{j+r})=\delta_r \text{ for } r=1,\dots,\ell\}.$$

Proceeding as we did in the proof of Theorem 2.1, one can easily establish that, as $x\to\infty$,

$$\frac{1}{x}\sum_{n\le x}\left(\widetilde{K}(n \mid \delta_1 \dots \delta_\ell) - \frac{\log\log x}{t^{\ell-1}}\right)^2 = o\big((\log\log x)^2\big).$$

It follows from this that the number $0.h(2)h(3)h(4)\dots$ is a normal number in base t, thus proving Theorem 2.3.

Acknowledgement. The authors would like to thank the referee for remarks and suggestions that greatly improved the quality of the presentation.

REFERENCES

[1] DE KONINCK, J.-M.—DOYON, N.: *The Life of Primes in 37 Episodes*, American Mathematical Society, Providence, RI, 2021.

[2] DE KONINCK, J.-M.—KÁTAI, I.: *On a problem on normal numbers raised by Igor Shparlinski*, Bull. Aust. Math. Soc. **84** (2011), 337–349.

[3] DE KONINCK, J.-M.—KÁTAI, I.: *Construction of normal numbers by classified prime divisors of integers*, Funct. Approx. Comment. Math. **4**5 (2011), no. 2, 231–253.

[4] DE KONINCK, J.-M.—KÁTAI, I.: *Normal numbers created from primes and polynomials*, Unif. Distrib. Theory **7** (2012), no. 2, 1–20.

[5] DE KONINCK, J.-M.—KÁTAI, I.: *Some new methods for constructing normal numbers*, Ann. Sci. Math. Québec **3**6 (2013), no. 2, 349–359.

[6] DE KONINCK, J.-M.—KÁTAI, I.: *Construction of normal numbers by classified prime divisors of integers II*, Funct. Approx. Comment. Math. **4**9 (2013), no. 1, 7–27.

[7] DE KONINCK, J.-M.—KÁTAI, I.: *Using large prime divisors to construct normal numbers*, Ann. Univ. Sci. Budapest Sect. Comput. **3**9 (2013), 45–62.

[8] DE KONINCK, J.-M.—KÁTAI, I.: *Construction of normal numbers using the distribution of the k-th largest prime factor*, J. Australian Math. Soc. **8**8 (2013), 158–168.

[9] DE KONINCK, J.-M.—KÁTAI, I.: *Prime-like sequences leading to the construction of normal numbers*, Funct. Approx. Comment. Math. **4**9 (2013), 291—302.

[10] DE KONINCK, J.-M.—KÁTAI, I.: *Normal numbers generated using the smallest prime factor function*, Ann. Math. Qué. **3**8 (2014), no. 2, 133–144.

[11] DE KONINCK, J.-M.—KÁTAI, I.: *Constructing normal numbers using residues of selective prime factors of integers*, Ann. Univ. Sci. Budapest. Sect. Comput. **4**2 (2014), 127–133.

[12] DE KONINCK, J.-M.—KÁTAI, I.: *Normal numbers and the middle prime factor of an integer*, Colloq. Math. **1**35 (2014), no. 1, 69–77.

[13] DE KONINCK, J.-M.—KÁTAI, I.: *The number of prime factors function on shifted primes and normal numbers. Topics in Mathematical Analysis and Applications.* In: (Themistocles M. Rassias, László Tóth, eds.) *Springer Optim. Appl. Vol. 94*, Springer, Cham., 2014, pp. 315–326.

[14] DE KONINCK, J.-M.—KÁTAI, I.: *Complex roots of unity and normal numbers*, Journal of Numbers, Vol. 2014 (June), Article ID 437814, 4 pp.

[15] DE KONINCK, J.-M.—KÁTAI, I.: *The number of large prime factors of integers and normal numbers*, Algébre et théorie des nombres, Publ. Math. Besançon Algèbre Théorie Vol. 2015, pp. 5–12.

[16] DE KONINCK, J.-M.—KÁTAI, I.: *Prime factorization and normal numbers*, Researches in Mathematics and Mechanics **2**0 (2015), no. 2, 69–80.

[17] DE KONINCK, J.-M.—LUCA, F.: *Analytic Number Theory: Exploring the Anatomy of Integers*. Graduate Studies in Math. Vol. 134, American Mathematical Society, Providence, RI, 2012.

[18] ELLIOTT, P. D. T. A.: *Probabilistic Number Theory II. Central Limit Theorems. Fundamental Principles of Mathematical Sciences, Vol. 240*. Springer-Verlag, Berlin-New York, (1980).

[19] ROSSER, J.B.—SCHOENFELD, L.: *Approximate formulas for some functions of prime numbers*, Illinois J. Math. **6** (1962), 64–94.

Received November 21, 2022
Accepted June 26, 2023

Jean-Marie De Koninck
Département de mathématiques
Faculté des sciences et de génie
Université Laval
1045 Avenue de la médecine
G1V 0A6, Quebec
CANADA
E-mail: jmdk@mat.ulaval.ca

Imre Kátai
Computer Algebra Departement
Eötvös Loránd University
Pázmány Péter Sétány I/C
1117 Budapest
HUNGARY
E-mail: katai@inf.elte.hu