On a problem on normal numbers raised by Igor Shparlinski

JEAN-MARIE DE KONINCK and IMRE KÁTAI

Édition du 7 juillet 2012

Bull. Austr. Math. Soc. **84** *(2011), 337-349*

Abstract

Given an integer $d \geq 2$, a *d-normal number*, or simply a *normal number*, is an irrational number whose d-ary expansion is such that any preassigned sequence, of length $k \geq 1$, taken within this expansion, occurs at the expected limiting frequency, namely $1/d^k$. Answering questions raised by Igor Shparlinski, we show that the numbers $0, P(2)P(3)P(4)\ldots P(n)\ldots$ and $0, P(2{+}1)P(3{+}1)P(5{+}1)\ldots P(p{+}1)\ldots$, where $P(n)$ stands for the largest prime factor of n, are both normal numbers.

AMS Subject Classification numbers: 11K16, 11A41, 11N37
Key words: normal numbers, primes, shifted primes, largest prime factor

1 Introduction

In 1909, Émile Borel [2] introduced the concept of a normal number. Given an integer $d \geq 2$, we say that an irrational number η is a *d-normal number*, or simply a *normal number*, if the d-ary expansion of η is such that any preassigned sequence, of length $k \geq 1$, taken within this expansion, occurs with the expected limiting frequency, namely $1/d^k$. Equivalently, given a positive real number $\eta < 1$ whose expansion is $\eta = 0, a_1 a_2 \ldots$, where each $a_j \in \{0, 1, \ldots, d-1\}$, that is, $\eta = \sum_{j=1}^{\infty} \frac{a_j}{d^j}$, we say that η is a d-normal number if the sequence $\{d^m \eta\}$, $m = 1, 2, \ldots$ (here $\{y\}$ stands for the fractional part of y), is uniformly distributed in the interval $[0, 1[$. Clearly, both definitions are equivalent.

The problem of determining if a given number is normal is unresolved. For instance, fundamental constants such as π, e, $\sqrt{2}$, $\log 2$ as well as the famous Apéry constant $\zeta(3)$, have not yet been proven to be normal numbers, although numerical evidence tends to indicate that they are. Interestingly, Borel [2] has shown that almost all numbers are normal.

Even constructing specific normal numbers is a no small challenge.

Several authors studied the problem of constructing normal numbers. One of the first was Champernowne [3] who, in 1933, showed that the number made up of the concatenation of the natural numbers, namely the number

$$0,123456789101112131415161718192021\ldots,$$

is normal in base 10. In 1946, Copeland and Erdős [4] proved that the same is true if one replaces the sequence of natural numbers by the sequence of primes, namely for the number

$$0,23571113171923293137\ldots$$

In the same paper, they conjectured that if $f(x)$ is any nonconstant polynomial whose values at $x = 1, 2, 3, \ldots$ are positive integers, then the number $0, f(1)f(2)f(3)\ldots$, where $f(n)$ is written in base 10, is a normal number. Six years later, Davenport and Erdős [5] proved this conjecture. In 1997, Nakai and Shiokawa [10] showed that if $f(x)$ is any nonconstant polynomial taking only positive integral values for positive integral arguments, then the number $0, f(2)f(3)f(5)f(7)\ldots f(p)\ldots$, where p runs through the prime numbers, is normal. In 2008, Madritsch, Thuswaldner and Tichy [9] extended the results of Nakai and Shiokawa by showing that, if f is an entire function of logarithmic order, then the numbers

$$0,[f(1)]_q[f(2)]_q[f(3)]_q\ldots \quad \text{and} \quad 0,[f(2)]_q[f(3)]_q[f(5)]_q[f(7)]_q\ldots,$$

where $[f(n)]_q$ stands for the base q expansion of the integer part of $f(n)$, are normal.

Recently, using our results [7] on the distribution of subsets of primes in the prime factorization of integers, we [6] constructed large families of normal numbers using classified prime divisors of integers. This motivated Igor Shparlinski to raise the following questions:

1. Letting $P(n)$ stand for the largest prime factor of the integer $n \geq 2$, is it possible to show that the number formed by the concatenation of the largest prime factors of the sequence of natural numbers $n \geq 2$, namely

   $$0, P(2)P(3)P(4)\ldots P(n)\ldots,$$

   is a normal number ?

2. Similarly, is the number formed by the concatenation of the largest prime factor of the shifted primes, that is,

   $$0, P(2+1)P(3+1)P(5+1)P(7+1)P(11+1)\ldots P(p+1)\ldots,$$

   a normal number ?

Here, we answer in the affirmative to both these questions and actually prove more.

2 Notation

Let $\wp$ stand for the set of all the prime numbers. The letter p with or without subscript will always denote a prime number.

Given a real number $x \geq 2$ and coprime integers k and ℓ, we let $\pi(x;k,\ell)$ stand for the number of prime numbers $p \leq x$ such that $p \equiv \ell \pmod{k}$. For each real number $x \geq 2$, we set $\mathrm{li}(x) := \displaystyle\int_2^x \frac{dt}{\log t}$, a function often called the logarithmic integral. We will also be using the well known function

$$\Psi(x,y) := \#\{n \leq x : P(n) \leq y\} \qquad (2 \leq y \leq x).$$

Given an interval of real numbers I, we write $\pi(I)$ for the number of prime numbers located in the interval I, while we write $\pi(I;k,\ell)$ for the number of primes $p \in I$ such that $p \equiv \ell \pmod{k}$.

Given an integer $t \geq 1$, an expression of the form $i_1 i_2 \ldots i_t$, where each i_j is one of the numbers $0, 1, \ldots, d-1$, is called a *word* of length t. Given a word α, we shall write $\lambda(\alpha) = t$ to indicate that α is a *word* of length t. We shall also use the symbol Λ to denote the *empty word*. Finally, we will say that α is a prefix of a word γ if for some δ, we have $\gamma = \alpha\delta$.

Let $d \geq 2$ be a fixed integer and let $E = E_d = \{0, 1, 2, \ldots, d-1\}$. Then, E^t will stand for the set of words of length t over E, while E^* will stand for the set of all words over E regardless of their length, including the empty word Λ. Observe that the concatenation of two words $\alpha, \beta \in E^*$, written $\alpha\beta$, also belongs to E^*.

Given a positive integer n, we write its d-ary expansion as

$$n = \varepsilon_0(n) + \varepsilon_1(n)d + \ldots + \varepsilon_t(n)d^t,$$

where $\varepsilon_i(n) \in E$ for $0 \leq i \leq t$ and $\varepsilon_t(n) \neq 0$. To this representation, we associate the words

$$\overline{n} = \varepsilon_0(n)\varepsilon_1(n)\ldots\varepsilon_t(n) \in E^{t+1}$$

and

$$\overline{\overline{n}} = \varepsilon_t(n)\varepsilon_{t-1}(n)\ldots\varepsilon_0(n) \in E^{t+1}.$$

Let k be a fixed positive integer. For each word $\beta = b_1 \ldots b_k \in E^k$, we let $\nu_\beta(\overline{n})$ stand for the number of occurrences of β in the d-ary expansion of the positive integer n, that is, the number of times that $\varepsilon_j(n)\ldots\varepsilon_{j+k-1}(n) = \beta$ as j varies from 0 to $t-(k-1)$.

For convenience, we also introduce the function $L(n) = L_d(n) = \left[\dfrac{\log n}{\log d}\right]$, which represents roughly the number of digits in the d-ary expansion of the positive integer n.

Finally, the letter c, with or without subscript, always denotes a positive constant, but not necessarily the same at each occurrence.

3 Main results

Theorem 1. *Let $F \in \mathbb{Z}[x]$ be a polynomial of positive degree r which takes only positive integral values at positive integral arguments. Then the numbers*

$$\eta = 0, \overline{F(P(2+1))}\,\overline{F(P(3+1))}\,\overline{F(P(5+1))}\ldots\overline{F(P(p+1))}\ldots$$

and

$$\widetilde{\eta} = 0, \overline{\overline{F(P(2+1))}}\,\overline{\overline{F(P(3+1))}}\,\overline{\overline{F(P(5+1))}}\ldots\overline{\overline{F(P(p+1))}}\ldots$$

are normal numbers.

Theorem 2. *Let F be as in Theorem 1. Then the numbers*

$$\xi = 0, \overline{F(P(2))}\,\overline{F(P(3))}\,\overline{F(P(4))}\ldots\overline{F(P(n))}\ldots$$

and

$$\xi^* = 0, \overline{\overline{F(P(2))}}\,\overline{\overline{F(P(3))}}\,\overline{\overline{F(P(4))}}\ldots\overline{\overline{F(P(n))}}\ldots$$

are normal numbers.

4 Preliminary lemmas

The following preliminary results are fundamental for the proof of our theorems.

Lemma 1. *Let F be as in the statement of Theorem 1 (with $deg(F) = r \geq 1$). Assume that κ_u is a function of u such that $\kappa_u > 1$ for all u. Given a word $\beta \in E^k$ and setting*

$$V_\beta(u) := \#\left\{Q \in \wp \cap [u, 2u] : \left|\nu_\beta(\overline{F(Q)}) - \frac{L(u^r)}{d^k}\right| > \kappa_u\sqrt{L(u^r)}\right\},$$

then, there exists a positive constant c such that

$$V_\beta(u) \leq \frac{cu}{(\log u)\kappa_u^2}.$$

Proof. This result can be obtained as a particular case of Theorem 1 of Bassily and Kátai [1] in the particular case when

$$F_k(\gamma) := \begin{cases} 1 & \text{if } \gamma = \beta, \\ 0 & \text{otherwise.} \end{cases}$$

□

Lemma 2. *Let F be as in Lemma 1. Given $\beta_1, \beta_2 \in E^k$ with $\beta_1 \neq \beta_2$, set*

$$\Delta_{\beta_1,\beta_2}(u) := \#\left\{Q \in \wp \cap [u, 2u] : \left|\nu_{\beta_1}(\overline{F(Q)}) - \nu_{\beta_2}(\overline{F(Q)})\right| > \kappa_u\sqrt{L(u^r)}\right\}.$$

Then, for some positive constant c,

$$\Delta_{\beta_1,\beta_2}(u) \leq \frac{cu}{(\log u)\kappa_u^2}.$$

Proof. This result is an immediate consequence of Lemma 1. □

From here on, we let I_x stand for the interval $[x, 2x]$.

Lemma 3. *For all $x \geq 2$,*

$$\frac{1}{\pi(x)} \#\{p \in I_x : P(p+1) \notin [x^{\delta}, x^{1-\delta}]\} < \frac{1}{2},$$

provided $\delta > 0$ is sufficiently small.

Proof. Let $x \geq 2$. We will first prove that

$$A := \#\{p \in I_x : P(p+1) > x^{1-\delta}\} < \frac{1}{4}\pi(x) \tag{4.1}$$

provided δ is sufficiently small.

If $P(p+1) > x^{1-\delta}$ for some $p \in I_x$, then there exists a prime number $q > x^{1-\delta}$ and a positive integer $a < 2x^{\delta}$ such that $p+1 = aq$. Using Corollary 2.4.1 from the book of Halberstam and Richert [8], we have that for each fixed a, the number of pairs p, q with $p \in I_x$ and $p+1 = aq$ is less than $\dfrac{cx}{\varphi(a)\log^2 x}$, where φ stands for the Euler function. Hence, summing over all positive integers $a < 2x^{\delta}$, we obtain that

$$A < \frac{cx}{\log^2 x} \sum_{a < 2x^{\delta}} \frac{1}{\varphi(a)} \leq c_1 \frac{x}{\log^2 x} \delta \log x < \frac{1}{4}\pi(x), \tag{4.2}$$

provided δ is small enough, thus proving (4.1).

We will now prove that

$$B := \#\{p \in I_x : P(p+1) < x^{\delta}\} < \frac{1}{4}\pi(x) \tag{4.3}$$

provided δ is sufficiently small.

To do so, given $k \geq 1$, we first introduce the strongly additive function defined on primes q by

$$f(q) = \begin{cases} 1 & \text{if } x^{\delta} \leq q < x^{k\delta}, \\ 0 & \text{otherwise.} \end{cases}$$

Assume that $k\delta \leq \frac{1}{3}$. Now, by using the Bombieri-Vinogradov Inequality, one can deduce a Turán-Kubilius type of inequality, namely

$$\sum_{p \in I_x} \left| f(p+1) - \sum_{x^{\delta} < q < x^{k\delta}} \frac{1}{q} \right|^2 \leq c\pi(x) \sum_{x^{\delta} < q < x^{k\delta}} \frac{1}{q}.$$

Hence, setting $S = \displaystyle\sum_{x^{\delta} < q < x^{k\delta}} \frac{1}{q}$ and observing that

$$S = \log\left(\frac{k\delta \log x}{\delta \log x}\right) + o(1) = \log k + o(1)$$

and that $P(p+1) < x^\delta$ implies that $f(p+1) = 0$, it follows that $BS^2 \le c\pi(x)S$, so that

$$B \le \frac{c\pi(x)}{\log k + o(1)} < \frac{1}{4}\pi(x),$$

if $k = 1/3\delta$ and δ is small enough, thus proving (4.3).

Combining (4.1) and (4.3) completes the proof of Lemma 3.

□

5 Proof of Theorem 1

Given a fixed real number x, write $p_1 < p_2 < \cdots < p_T$ for the whole list of primes belonging to I_x, while for each $Q \in \wp$, let

$$M(Q) := \#\{p \in I_x : P(p+1) = Q\}$$

and observe that by the Brun-Titchmarsh Theorem,

$$M(Q) \le \pi(I_x; Q, -1) \le \frac{cx}{Q\log(x/Q)}. \tag{5.1}$$

Let δ be a small positive number. Then, as we did in order to establish (4.2), it is easy to see that, for some absolute positive constant $c > 0$,

$$\#\{p \in I_x : P(p+1) > x^{1-\delta}\} \le \frac{c\delta x}{\log x}. \tag{5.2}$$

With the primes $p_1 < p_2 < \cdots < p_T$ defined above, consider the number θ defined by

$$\theta = \overline{F(P(p_1+1))}\;\overline{F(P(p_2+1))}\ldots\overline{F(P(p_T+1))}.$$

Since, for each $j \in \{1, \ldots, T\}$,

$$\lambda(\overline{F(P(p_j+1))}) = L(F(P(p_j+1))) + O(1) = rL(P(p_j+1)) + O(1),$$

it follows that

$$\lambda(\theta) = r\sum_{j=1}^{T} L(P(p_j+1)) + O(T). \tag{5.3}$$

Now, since $L(P(p_j+1)) \le L(p_j+1) \le L(2x)$, it follows, combining (5.2) and (5.3), that

$$\sum_{P(p_j+1)>x^{1-\delta}} L(P(p_j+1)) \le L(2x) \sum_{P(p_j+1)>x^{1-\delta}} 1 \le \frac{c\delta x}{\log x}L(2x). \tag{5.4}$$

On the other hand,

$$\sum_{P(p_j+1)<x^\delta} L(P(p_j+1)) \le L(2x) \sum_{P(p_j+1)<x^\delta} 1 \le \frac{c\delta x}{\log x} L(2x). \tag{5.5}$$

Using (5.4) and (5.5) in (5.3), we may conclude that there exist two positive numbers $d_1 < d_2$ such that

$$d_1 < \frac{\lambda(\theta)}{rL(2x)\pi(I_x)} < d_2,$$

where we used Lemma 3, thereby implying that the true order of $\lambda(\theta)$ is x.

We will now subdivide the interval $[x^\delta, x^{1-\delta}]$ into subintervals $[u_j, u_{j+1}]$, where $u_j = x^\delta 2^j$ with $j = 0, 1, \ldots, Z$, where Z is the unique positive integer satisfying $u_Z \le x^{1-\delta} < u_{Z+1}$.

Our intention is to show that $\nu_\beta(\theta) \sim \frac{1}{d^k}\lambda(\theta)$ as $\lambda(\theta) \to \infty$. We will do so by establishing that $\nu_{\beta_1}(\theta) - \nu_{\beta_2}(\theta)$ is small if $\beta_1 \ne \beta_2$.

Let us now choose $\kappa_u = \log\log u$.

We shall say that the prime $Q \in [u_j, u_{j+1}]$ is a *good prime* (with respect to β_1 and β_2) if

$$\max_{i=1,2} \left| \nu_{\beta_i}(\overline{F(Q)}) - \frac{rL(u_j)}{d^k} \right| < \kappa_{u_j} \sqrt{L(u_j)}, \tag{5.6}$$

while we say that it is a *bad prime* if (5.6) does not hold.

We then have

$$\begin{aligned} |\nu_{\beta_1}(\theta) - \nu_{\beta_2}(\theta)| \;\le\; & c \sum_{j=0}^{Z} \kappa_{u_j} \sqrt{L(u_j)} \sum_{Q \in [u_j, u_{j+1}]} M(Q) \\ & + O\left(r \sum_{j=0}^{Z} L(u_j) \sum_{\substack{Q \in [u_j, u_{j+1}] \\ Q \text{ bad prime}}} M(Q) \right) + O(Z) \\ & + O\left(\frac{c\delta x}{\log x} L(2x) \right), \end{aligned} \tag{5.7}$$

where the first term on the right hand side of this inequality was concerned with the good primes Q, while the fourth (and last) term is to account for the primes p for which $p < x^\delta$ or $p > x^{1-\delta}$.

Using inequality (5.1), we obtain

$$\sum_{\substack{Q \in [u_j, u_{j+1}] \\ Q \text{ bad prime}}} M(Q) \le \frac{cx}{u_j \log(x/u_j)} \sum_{\substack{Q \in [u_j, u_{j+1}] \\ Q \text{ bad prime}}} \frac{1}{Q}. \tag{5.8}$$

On the other hand, it follows from Lemma 2 that

$$\sum_{\substack{Q\in[u_j,u_{j+1}] \\ Q \text{ bad prime}}} \frac{1}{Q} \le \frac{cu_j}{(\log u_j)\kappa_{u_j}^2}. \tag{5.9}$$

Hence, using (5.9) in (5.8), we obtain

$$\sum_{\substack{Q\in[u_j,u_{j+1}] \\ Q \text{ bad prime}}} M(Q) \le \frac{cx}{u_j \log(x/u_j)} \cdot \frac{cu_j}{(\log u_j)\kappa_{u_j}^2}. \tag{5.10}$$

Using (5.10) in (5.7), we may write

$$|\nu_{\beta_1}(\theta) - \nu_{\beta_2}(\theta)| \le \Sigma_1 + \Sigma_2 + cZ + c\delta x, \tag{5.11}$$

where

$$\begin{aligned} \Sigma_1 &= c\sum_{j=0}^{Z} \frac{\kappa_{u_j}}{\sqrt{\log u_j}} \cdot \frac{x}{u_j \log(x/u_j)}, \\ \Sigma_2 &= c\sum_{j=0}^{Z} \frac{x}{\kappa_{u_j}^2 \log(x/u_j)}. \end{aligned}$$

First, it is clear that

$$\Sigma_1 = o(x). \tag{5.12}$$

On the other hand, since

$$\Sigma_2 \le cx \cdot \frac{1}{\kappa_{u_0}^2} \sum_{j=0}^{Z} \frac{1}{\log(x/u_j)},$$

it follows that

$$\Sigma_1 = o(x) \tag{5.13}$$

as well.

Now, by the way we chose Z, it is clear that $Z \le cx/\log x$. Hence, gathering (5.12) and (5.13) in (5.11), we get that

$$|\nu_{\beta_1}(\theta) - \nu_{\beta_2}(\theta)| \le c\delta x + o(x). \tag{5.14}$$

Since $\sum_{\gamma\in E^k} \nu_\gamma(\theta) = \lambda(\theta) - k + 1$, it follows that

$$d^k \nu_\beta(\theta) - \lambda(\theta) = \sum_{\gamma\in E^k} (\nu_\beta(\theta) - \nu_\gamma(\theta)) + O(1).$$

Using this last estimate in (5.14), we have

$$\left| \nu_\beta(\theta) - \frac{\lambda(\theta)}{d^k} \right| \le c\delta x + o(x). \tag{5.15}$$

Now, let η_N be the prefix of length N of the infinite sequence

$$\overline{F(P(2+1))}\ \overline{F(P(3+1))}\ \overline{F(P(5+1))} \dots$$

and let p^* be the largest prime for which

$$\lambda\left(\overline{F(P(2+1))}\,\overline{F(P(3+1))} \dots \overline{F(P(p^*+1))}\right) < N.$$

Moreover, set

$$\eta_N^* = \overline{F(P(2+1))}\ \overline{F(P(3+1))} \dots \overline{F(P(p^*+1))}.$$

We then have

$$0 \le N - \lambda(\eta_N^*) \le cr \log p^* \le crN/\log N. \tag{5.16}$$

We now define the sequence $Y_0, Y_{-1}, Y_{-2}, \dots, Y_{-H}$ as follows:

$$Y_0 = p^*, \quad Y_{-1} = \frac{1}{2} Y_0, \quad \dots \quad , Y_{-(j+1)} = \frac{1}{2} Y_{-j}, \quad \dots \quad , Y_{-H},$$

where H is the smallest integer for which $2^H > \log p^*$, implying that $H \log 2 \sim \log\log p^*$ as p^* grows.

Let us write η_N^* as

$$\eta_N^* = \rho\theta_{-H}\theta_{-(H-1)} \dots \theta_0 \qquad (\approx \eta_N),$$

where ρ is the word $\overline{F(P(2+1))}\,\overline{F(P(3+1))} \dots \overline{F(P(q_0+1))}$, where q_0 is the largest prime number which is smaller than Y_{-H} and where

$$\theta_{-j} = \overline{F(P(p_1+1))}\,\overline{F(P(p_2+1))} \dots \overline{F(P(p_r+1))},$$

where $p_1 < p_2 < \dots < p_r$ are all the primes contained in the interval $[Y_{-(j+1)}, Y_{-j}]$. With this set up, it is clear that

$$\nu_\beta(\eta_N^*) = \nu_\beta(\rho) + \nu_\beta(\theta_{-H}) + \dots + \nu_\beta(\theta_0) + O((H+1)k). \tag{5.17}$$

But since

$$\nu_\beta(\rho) \le crq_0 \le crY_{-H} < c\frac{p*}{\log p*},$$

it follows from (5.17) that

$$\nu_{\beta_1}(\eta_N^*) - \nu_{\beta_2}(\eta_N^*) \quad = \quad \sum_{j=-H}^{0} (\nu_{\beta_1}(\theta_j) - \nu_{\beta_2}(\theta_j))$$

$$+O(\log\log p^*) + O\left(\frac{p^*}{\log p^*}\right). \tag{5.18}$$

In light of (5.14), we obtain from (5.18) that

$$\begin{aligned} |\nu_{\beta_1}(\eta_N^*) - \nu_{\beta_2}(\eta_N^*)| &\leq c\delta \sum_{j=0}^{H} (Y_{-j} - Y_{-(j+1)}) + O\left(\frac{p^*}{\log p^*}\right) \\ &\leq c\delta Y_0 = c\delta p^*, \end{aligned}$$

so that, proceeding as we did to obtain (5.15), we obtain

$$\left| \nu_\beta(\eta_N) - \frac{\lambda(\eta_N)}{d^k} \right| \leq c\delta p^*. \tag{5.19}$$

Since

$$\lambda(\eta_N) = \sum_{p \leq p^*} L(P(p+1)) = \frac{rp^*}{\log d} + O(\pi(p^*)),$$

it follows from (5.19) that

$$\lim_{N\to\infty} \left| \frac{\nu_\beta(\eta_N)}{\lambda(\eta_N)} - \frac{1}{d^k} \right| \leq c\delta. \tag{5.20}$$

Since $\delta > 0$ is arbitrary, we may conclude that the left hand side of (5.20) is 0.

This completes the proof that the number η is normal. The proof that $\widetilde{\eta}$ is normal can be obtained along the same lines.

6 Proof of Theorem 2

The proof is much easier than that of Theorem 1. Indeed, as previously, let $I_x = [x, 2x]$ and set

$$\theta = \overline{F(P(n_0))} \ldots \overline{F(P(n_T))},$$

where n_0 is the smallest integer in I_x, while n_T is the largest. We then have

$$\lambda(\theta) = rx\frac{\log x}{\log d} + O(x).$$

Let δ be a small positive number. One can easily show that the number of integers $n \in I_x$ for which either $P(n) < x^\delta$ or $P(n) > x^{1-\delta}$ is $\leq c\delta x$. In light of this, we have

$$\nu_\beta(\theta) = \sum_{\substack{n\in I_x \\ x^\delta \leq P(n) \leq x^{1-\delta}}} \nu_\beta(\overline{F(P(n))}) + O(T) + O(\delta x \log x). \tag{6.1}$$

Let us choose

$$u_0 = x^\delta \text{ and thereafter } u_j = 2u_{j-1} \text{ for each } 1 \leq j \leq H, \tag{6.2}$$

where H is the smallest positive integer for which $2^H u_0 > x^{1-\delta}$, so that

$$H = \left[\frac{(1-2\delta)\log x}{\log 2}\right] + O(1). \tag{6.3}$$

Letting $z = \log x/\log y$, it is known that

$$\Psi(x,y) = \alpha(z)\,x + O\left(\frac{x}{\log y}\right) \qquad \text{uniformly for } 2 \le y \le x, \tag{6.4}$$

where α stands for the Dickman function (see for instance Tenenbaum [11]).

Hence, if for each prime q, we let $R(q) := \#\{n \in I_x : P(n) = q\}$, it follows from (6.4) that

$$\begin{aligned} R(q) &= \Psi\left(\frac{2x}{q}, q\right) - \Psi\left(\frac{x}{q}, q\right) \\ &= \alpha\left(\frac{\log(2x/q)}{\log q}\right)\frac{2x}{q} - \alpha\left(\frac{\log(x/q)}{\log q}\right)\frac{x}{q} + O\left(\frac{x}{q\log q}\right) \\ &= (1+o(1))\alpha\left(\frac{\log x}{\log q} - 1\right)\frac{x}{q}, \end{aligned} \tag{6.5}$$

where we used the fact that $q \in [x^\delta, x^{1-\delta}]$.

Now, it follows from (6.1) that

$$\nu_\beta(\theta) = \sum_{x^\delta \le q \le x^{1-\delta}} \nu_\beta(\overline{F(q)})R(q) + O(T) + O(\delta x \log x). \tag{6.6}$$

Let $\beta_1, \beta_2 \in E_k$ with $\beta_1 \ne \beta_2$. Then, it follows from (6.6) that

$$|\nu_{\beta_1}(\theta) - \nu_{\beta_2}(\theta)| \le \sum_{x^\delta \le q \le x^{1-\delta}} \left|\nu_{\beta_1}(\overline{F(q)}) - \nu_{\beta_2}(\overline{F(q)})\right| R(q) + O(x) + O(\delta x \log x). \tag{6.7}$$

Taking $u_0, u_1, \ldots, u_H$ as in (6.2) with H as in (6.3), the sum on the right hand side of (6.7), which we shall denote by $S^*(x)$, can be rewritten and handled as follows.

$$S^*(x) = \sum_{j=0}^{H-1} \sum_{u_j \le q < u_{j+1}} \left|\nu_{\beta_1}(\overline{F(q)}) - \nu_{\beta_2}(\overline{F(q)})\right| R(q) = \sum_{j=0}^{H-1} S_j, \tag{6.8}$$

say. Now, clearly, in light of (6.5),

$$S_j \le \frac{2\alpha(u_j)}{u_j} x \sum_{u_j \le q < u_{j+1}} \left|\nu_{\beta_1}(\overline{F(q)}) - \nu_{\beta_2}(\overline{F(q)})\right|. \tag{6.9}$$

We now define $\kappa_u := \log\log u$ and classify the primes $q \in [u_j, u_{j+1})$ as good or bad primes. We say that $q \in [u_j, u_{j+1})$ is a *good prime* if

$$\left|\nu_{\beta_1}(\overline{F(q)}) - \nu_{\beta_2}(\overline{F(q)})\right| \le \kappa_u \sqrt{L(u^r)},$$

while we say that it is a *bad prime* otherwise.

Splitting the sum on the right hand side of (6.9) into two sums, one running on the good primes and one running on the bad primes, it follows from Lemma 2 that

$$\begin{aligned} S_j &\le \frac{2\alpha(u_j)}{u_j} x\kappa_{u_j}\sqrt{L(u_j^r)}\frac{u_j}{\log u_j} + \frac{2\alpha(u_j)}{u_j} x\frac{u_j}{(\log u_j)\kappa_{u_j}^2} \\ &= 2\alpha(u_j)x \cdot \left\{ \frac{\kappa_{u_j}\sqrt{L(u_j^r)}}{\log u_j} + \frac{1}{(\log u_j)\kappa_{u_j}^2} \right\} \\ &\le 4r\alpha(u_j)x\frac{\log\log u_j}{\sqrt{\log u_j}}. \end{aligned} \tag{6.10}$$

Summing the inequalities in (6.10) for $j = 0, 1, \ldots, H-1$, we obtain from (6.8) that $S^*(x) = o(\mathrm{li}(x))$ as $x \to \infty$. Using this estimate in (6.7), we obtain that

$$|\nu_{\beta_1}(\theta) - \nu_{\beta_2}(\theta)| \le c\delta x \log x + o(x \log x). \tag{6.11}$$

Now let ξ_N be the prefix of length N of

$$\overline{F(P(2))}\ \overline{F(P(3))} \ldots$$

and let

$$\widetilde{\xi_N} = \overline{F(P(2))}\ \overline{F(P(3))} \ldots \overline{F(P(m))},$$

where $\lambda(\widetilde{\xi_N}) \le N < \lambda(\widetilde{\xi_N}\overline{F(P(m+1))})$.

It is clear that $m \sim c\dfrac{N}{\log N}$ for some constant $c > 0$, which implies that $\lambda(\overline{F(P(m+1))}) \ll r \log m$.

Let $2x = m$ and consider the intervals $I_x, I_{x/2}, I_{x/(2^2)}, \ldots, I_{x/(2^L)}$, where $L = 2[\log\log x]$, and write

$$\tau_j = \overline{F(P(a))} \ldots \overline{F(P(b))} \qquad (j = 0, 1, \ldots, L),$$

where a is the smallest and b the largest integer in $I_{x/(2^j)}$. Moreover, let

$$\mu = \overline{F(P(2))} \ldots \overline{F(P(s))},$$

where s is the largest integer which is less than the smallest integer in $I_{x/(2^{j+1})}$.

It is clear that

$$\left|\nu_{\beta_1}(\widetilde{\xi_N}) - \nu_{\beta_2}(\widetilde{\xi_N})\right| \le |\nu_{\beta_1}(\mu) - \nu_{\beta_2}(\mu)| + \sum_{j=0}^{L} |\nu_{\beta_1}(\tau_j) - \nu_{\beta_2}(\tau_j)| \tag{6.12}$$

and that

$$\nu_\beta(\mu) \le \lambda(\mu) \le \frac{x}{2^L} \cdot r \log x = o(x). \tag{6.13}$$

Applying estimate (6.11) $L+1$ times by replacing successively $2x$ by x, $x/2$, $x/2^2$, $\ldots$, $x/2^L$, we obtain from (6.12) and in light of (6.13), that

$$\left|\nu_{\beta_1}(\widetilde{\xi_N}) - \nu_{\beta_2}(\widetilde{\xi_N})\right| \le c\delta N + o(N) \qquad (N\to\infty). \tag{6.14}$$

Then, using the same argument as in the proof of Theorem 1, it follows from (6.14) that

$$\limsup_{N\to\infty}\left|\frac{\nu_\beta(\xi_N)}{N} - \frac{1}{d^k}\right| \le c\delta.$$

Since $\delta>0$ can be chosen arbitrarily small, it follows that

$$\limsup_{N\to\infty}\frac{\nu_\beta(\xi_N)}{N} = \frac{1}{d^k},$$

thus establishing that ξ is normal.

The proof that ξ^* is normal is similar.

This completes the proof of Theorem 2.

References

[1] N.L. Bassily and I. Kátai, *Distribution of consecutive digits in the q-ary expansions of some sequences of integers*, Journal of Mathematical Sciences **78**, no. 1 (1996), 11-17.

[2] E. Borel, *Les probabilités dénombrables et leurs applications arithmétiques*, Rend. Circ. Mat. Palermo **27** (1909), 247–271.

[3] D.G. Champernowne, *The construction of decimals normal in the scale of ten*, J. London Math. Soc. **8** (1933), 254-260.

[4] A.H. Copeland and P. Erdős, *Note on normal numbers*, Bull. Amer. Math. Soc. **52** (1946), 857-860.

[5] H. Davenport and P. Erdős, *Note on normal decimals*, Can. J. Math. **4** (1952), 58-63.

[6] J.M. De Koninck and I. Kátai, *Construction of normal numbers by classified prime divisors of integers*, Functiones et Approximatio **45.2** (2011), 231-253.

[7] J.M. De Koninck and I. Kátai, *On the distribution of subsets of primes in the prime factorization of integers*, Acta Arithmetica **72**, no.2 (1995), 169-200.

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

[9] M.G. Madritsch, J.M. Thuswaldner and R.F. Tichy, *Normality of numbers generated by the values of entire functions*, J. of Number Theory **128** (2008), 1127-1145.

[10] Y. Nakai and I. Shiokawa, *Normality of numbers generated by the values of polynomials at primes*, Acta Arith. **81** (1997), no. 4, 345-356.

[11] G. Tenenbaum, *Introduction à la théorie analytique des nombres*, Collection Échelles, Belin, 2008.

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 7 juillet 2012; fichier: normal-IS-2010-2011.tex