Nineteen papers on normal numbers by Jean-Marie De Koninck and Imre Kátai

Édition du 16 octobre 2020

This document is a follow up of the talk *New approaches in the construction of normal numbers* given by the first author in Vienna in November 2016. Here, we produce a survey of nineteen papers on normal numbers written by Jean-Marie De Koninck and Imre Kátai since 2011. We only state the results with their motivation and at times the approach or sketch of their proofs[1].

Here is the plan:

[1]Each of the nineteen papers is available on the first author's home page at the web site address `www.jeanmariedekoninck.mat.ulaval.ca`

Notation and the concept of normal number

Throughout this survey, we let $\wp$ stand for the set of all prime numbers. The letter p, with or without subscript, stands for a prime number. The letter c, with or without subscript, stands for an absolute positive constant, but not necessarily the same at each occurrence. At times, we will be writing x_1 for $\max(1, \log x)$, x_2 for $\max(1, \log\log x)$, and so on.

We shall be using the arithmetical functions

$$\begin{aligned}
\omega(n) &= \text{the number of distinct prime factors of } n,\\
\Omega(n) &= \text{the number of prime factors of } n \text{ counting their multiplicity},\\
\phi(n) &= \#\{m \le n : \gcd(m,n) = 1\}, \text{ the Euler totient function},\\
p(n) &= \text{the smallest prime factor of } n, \text{ with } p(1) = 1,\\
P(n) &= \text{the largest prime factor of } n, \text{ with } P(1) = 1
\end{aligned}$$

as well as the functions

$$\begin{aligned}
\pi(x) &= \text{the number of prime numbers } p \le x,\\
\pi(x; k, \ell) &= \text{the number of prime numbers } p \le x \text{ such that } p \equiv \ell \pmod{k},\\
\mathrm{li}(x) &= \int_2^x \frac{dt}{\log t}, \text{ the logarithmic integral of } x,\\
\pi(B) &= \text{the number of primes belonging to the set } B.
\end{aligned}$$

Also, given a set of primes $\mathcal{P}$, we will write $\mathcal{N}(\mathcal{P})$ for the semi-group generated by $\mathcal{P}$.

A sequence $(x_n)_{n\in\mathbb{N}}$ of real numbers is said to be *uniformly distributed modulo 1* (or *mod 1*) if for every interval $[a, b) \subseteq [0, 1)$,

$$\lim_{N\to\infty} \frac{1}{N} \#\{n \le N : \{x_n\} \in [a, b)\} = b - a.$$

(Here, $\{y\}$ stands for the fractional part of y.) In other words, a sequence of real numbers is said to be uniformly distributed mod 1 if every subinterval of the unit interval gets its fair share of the fractional parts of the elements of this sequence.

Also, given a set of N real numbers $x_1, \dots, x_N$, we define the *discrepancy* of this set as the quantity

$$D(x_1, \dots, x_N) := \sup_{[a,b) \subseteq [0,1)} \left| \frac{1}{N} \sum_{\substack{n \le N \\ \{x_n\} \in [a,b)}} 1 - (b-a) \right|. \tag{0.1}$$

It is easily established that a sequence of real numbers $(x_n)_{n \in \mathbb{N}}$ is uniformly distributed mod 1 if and only if $D(x_1, \dots, x_N) \to 0$ as $N \to \infty$ (see Theorem 1.1 in Chapter 2 in the book of Kuipers and Niederreiter [50]).

The concept of a normal number goes back to 1909: it was first introduced by Émile Borel [6]. Given an integer $q \ge 2$, a *q-normal number*, or for short a *normal number*, is a real number whose q-ary expansion is such that any preassigned sequence, of length $k \ge 1$, of base q digits from this expansion, occurs at the expected frequency, namely $1/q^k$. Equivalently, given a positive real number

$$\eta = \lfloor \eta \rfloor + \sum_{j=1}^{\infty} \frac{a_j}{q^j},$$

where each $a_j \in \{0, 1, \dots, q-1\}$, we say that η is a *q-normal number* if for every integer $k \ge 1$ and $b_1 b_2 \dots b_k \in \{0, 1, \dots, q-1\}^k$, we have

$$\lim_{N \to \infty} \frac{1}{N} \#\{j \le N : a_j a_{j+1} \dots a_{j+k-1} = b_1 \dots b_k\} = \frac{1}{q^k}.$$

Also, given an integer $q \ge 2$, it can be shown (see Theorem 8.1 of Chapter 1 in the book of Kuipers and Niederreiter [50]) that a real number η is normal in base q if and only if the sequence $(\{q^n \eta\})_{n \in \mathbb{N}}$ is uniformly distributed mod 1.

Let $q \ge 2$ be a fixed integer and set $\mathcal{A}_q := \{0, 1, 2, \dots, q-1\}$. Given an integer $t \ge 1$, an expression of the form $i_1 i_2 \dots i_t$, where each $i_j \in \mathcal{A}_q$, 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* and write $\lambda(\Lambda) = 0$.

We will write $\mathcal{A}_q^k$ for the set of words of length k, while $\mathcal{A}_q^*$ will stand for the set of finite words over $\mathcal{A}_q$, including the empty word Λ. The operation on $\mathcal{A}_q^*$ is the concatenation $\alpha\beta$ for $\alpha, \beta \in \mathcal{A}_q^*$. It is clear that $\lambda(\alpha\beta) = \lambda(\alpha) + \lambda(\beta)$. Also, we will say that α is a prefix of a word γ if for some δ, we have $\gamma = \alpha\delta$.

Given $n \in \mathbb{N}$, we shall write its q-ary expansion as

$$n = \varepsilon_0(n) + \varepsilon_1(n) q + \varepsilon_2(n) q^2 + \cdots + \varepsilon_t(n) q^t, \tag{0.2}$$

where $\varepsilon_i(n) \in \mathcal{A}_q$ for $0 \le i \le t$ and $\varepsilon_t(n) \ne 0$. To this representation, we associate the word $\overline{n} = \varepsilon_0(n) \varepsilon_1(n) \dots \varepsilon_t(n) \in \mathcal{A}_q^{t+1}$. For such a word $\overline{n}$, given a word $\beta = b_1 b_2 \dots b_k \in \mathcal{A}_q^k$, we let $\nu_\beta(\overline{n})$ stand for the number of occurrences of β in the q-ary expansion of the positive integer n, that is, the number of times that $\varepsilon_j(n) \dots \varepsilon_{j+k-1}(n) = \beta$ as j varies from 0 to $t - (k-1)$.

Let $\eta_\infty = \varepsilon_1\varepsilon_2\varepsilon_3\ldots$, where each ε_i is an element of $\mathcal{A}_q$ and, for each positive integer N, let $\eta_N = \varepsilon_1\varepsilon_2\ldots\varepsilon_N$. Moreover, for each $\beta = \delta_1\delta_2\ldots\delta_k \in \mathcal{A}_q^k$ and integer $N \geq 2$, let $M(N,\beta)$ stand for the number of occurrences of β as a subsequence of the consecutive digits of η_N, that is,

$$M(N,\beta) = \#\{(\alpha,\gamma) : \eta_N = \alpha\beta\gamma,\ \alpha,\gamma \in \mathcal{A}_q^*\}.$$

We will say that η_∞ is a *normal sequence* if

$$\lim_{N\to\infty} \frac{M(N,\beta)}{N} = \frac{1}{q^{\lambda(\beta)}} \quad \text{for all } \beta \in \mathcal{A}_q^*. \tag{0.3}$$

Let $\xi < 1$ be a positive real number whose q-ary expansion is

$$\xi = 0.\varepsilon_1\varepsilon_2\varepsilon_3\ldots$$

and, for each integer $N \geq 1$, set

$$\xi_N = 0.\varepsilon_1\varepsilon_2\ldots\varepsilon_N.$$

With β and $M(N,\beta)$ as above, we will say that ξ is normal if (0.3) holds.

Given an integer $q \geq 2$ and a positive integer n, we let

$$L(n) = L_q(n) = \left\lfloor \frac{\log n}{\log q} \right\rfloor + 1, \tag{0.4}$$

that is, the number of digits of n in base q.

Preliminary results

In 1995 (see [12]), we introduced the notion of a *disjoint classification of primes*, that is a collection of $q+1$ disjoint sets of primes $\mathcal{R}, \wp_0, \wp_1, \ldots, \wp_{q-1}$, whose union is $\wp$, where $\mathcal{R}$ is a finite set (perhaps empty) and where the other q sets are of positive densities $\delta_0, \delta_1, \ldots, \delta_{q-1}$ (with clearly $\sum_{i=0}^{q-1} \delta_i = 1$). For instance, the sets $\wp_0 = \{p : p \equiv 1 \pmod 4\}$, $\wp_1 = \{p : p \equiv 3 \pmod 4\}$ and $\mathcal{R} = \{2\}$ provide a disjoint classification of primes.

We then introduced the function $H : \mathbb{N} \to \mathcal{A}_q^*$ defined[2] by $H(n) = H(p_1^{a_1}\cdots p_r^{a_r}) = \ell_1\ldots\ell_r$, where each ℓ_j is such that $p_j \in \wp_{\ell_j}$, and investigated the size of the set of positive integers $n \leq x$ for which $H(n) = \alpha$ for a given word $\alpha \in \mathcal{A}_q^k$. More precisely, we proved the following result.

Theorem A. *Let $\mathcal{R}, \wp_0, \wp_1, \ldots, \wp_{q-1}$ be a disjoint classification of primes such that, for some $c_1 \geq 5$ and each $i = 0, 1, \ldots, q-1$,*

$$\pi([u, u+v] \cap \wp_i) = \delta_i \pi([u, u+v]) + O\left(\frac{u}{\log^{c_1} u}\right) \tag{0.5}$$

holds uniformly for $2 \leq v \leq u$, where $\delta_0, \delta_1, \ldots, \delta_{q-1}$ are positive constants such that $\sum_{i=0}^{q-1} \delta_i = 1$. Let $\lim_{x\to\infty} w_x = +\infty$ with $w_x = O(x_3)$. Let A be a positive integer such

[2] Note the distinction between the use of the central dots $(\cdots)$ and that of the lower dots $(\ldots)$, the former being used for the multiplication of real numbers and the later for that of the concatenation of digits.

that $A \le x_2$ and $P(A) \le w_x$. Then, for $\sqrt{x} \le Y \le x$ and $1 \le k \le c_2 x_2$, where c_2 is an arbitrary constant, as $x \to \infty$,

$$\begin{aligned}\#\{n = An_1 \le Y : p(n_1) > w_x,\ \omega(n_1) = k,\ H(n_1) = i_1 \dots i_k\} \\ = (1+o(1))\delta_{i_1}\cdots\delta_{i_k}\frac{Y}{A\log Y} t_k(Y)\varphi_{w_x}\left(\frac{k-1}{x_2}\right)F\left(\frac{k-1}{x_2}\right),\end{aligned}$$

where $t_k(x) = \dfrac{x_2^{k-1}}{(k-1)!}$,

$$\varphi_{w_x}(z) := \prod_{p\le w_x}\left(1+\frac{z}{p}\right)^{-1} \quad \text{and} \quad F(z) := \frac{1}{\Gamma(z+1)}\prod_{p}\left(1+\frac{z}{p}\right)\left(1-\frac{1}{p}\right)^{z}.$$

Here are more results which will be used in some of our seventeen papers.

Lemma 0.1. *(*Brun-Titchmarsh Inequality*) If $1 \le k < x$ and $(k,\ell) = 1$, then*

$$\pi(x;k,\ell) < 3\frac{x}{\phi(k)\log(x/k)}.$$

Proof. This is essentially Theorem 3.8 in the book of Halberstam and Richert [43]. □

Lemma 0.2. *(*Bombieri-Vinogradov Theorem*) Given any fixed number $A > 0$, there exists a number $B = B(A) > 0$ such that*

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

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

Proof. For a proof, see Theorem 17.1 in the book of Iwaniec and Kowalski [48]. □

Lemma 0.3. *(*Siegel-Walfisz Theorem*) Let $A > 0$ be an arbitrary number. Then, there exists a positive constant $c = c(A)$ such that*

$$\pi(x;k,\ell) = \frac{\mathrm{li}(x)}{\phi(k)} + O\left(\frac{x}{e^{c\sqrt{\log x}}}\right)$$

whenever the integers k and ℓ are coprime and $k < \log^A x$.

Proof. This is Theorem 8.17 in the book of Tenenbaum [61]. □

Lemma 0.4. *Given a fixed integer $q \ge 2$, let L be defined as in (0.4). Let also $F \in \mathbb{Z}[x]$ be a polynomial of positive degree r which takes only positive integral values at positive integral arguments. Moreover, assume that κ_u is a function of u such that $\kappa_u > 1$ for all $u > e^e$. Then, given a word $\beta \in \mathcal{A}_q^k$, there exists a positive constant c such that*

$$\#\left\{p \in [u,2u] : \left|\nu_\beta(\overline{F(p)}) - \frac{L(u^r)}{q^k}\right| > \kappa_u\sqrt{L(u^r)}\right\} \le \frac{cu}{(\log u)\kappa_u^2}.$$

The above result is a particular case of Theorem 1 in the 1996 paper of Bassily and Kátai [2]. The following result is an immediate consequence of Lemma 0.4.

Lemma 0.5. *Let q, L, F, κ_u be as in Lemma 0.4. Given $\beta_1, \beta_2 \in \mathcal{A}_q^k$ with $\beta_1 \neq \beta_2$, there exists a positive constant c such that*

$$\#\left\{p \in [u, 2u] : \left|\nu_{\beta_1}(\overline{F(p)}) - \nu_{\beta_2}(\overline{F(p)})\right| > \kappa_u \sqrt{L(u^r)}\right\} \leq \frac{cu}{(\log u)\kappa_u^2}.$$

We now introduce the counting function of the y-smooth (or y friable) numbers, namely those positive integers n such that $P(n) \leq y$.

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

Lemma 0.6. *There exists an absolute constant $c > 0$ such that, uniformly for $2 \leq y \leq x$,*

$$\Psi(x, y) \leq c\, x \, \exp\left\{-\frac{1}{2}\frac{\log x}{\log y}\right\}.$$

Proof. For a proof, see the book of Tenenbaum [61]. □

Lemma 0.7. *Uniformly for $2 \leq y \leq x$, with $u = \log x / \log y$, we have*

$$\Psi(x, y) = \rho(u)\, x + O\left(\frac{x}{\log y}\right),$$

where ρ stands for the Dickman function.

Proof. See for instance Theorem 9.14 in the book of De Koninck and Luca [34]. □

Lemma 0.8. *There exists an absolute constant $c > 0$ such that, given any $\delta \in (0, 1/2)$, we have, for all $x \geq 2$,*

$$\#\{n \in [x, 2x] : P(n) < x^\delta \text{ or } P(n) > x^{1-\delta}\} < c\,\delta\, x.$$

Proof. This result is an easy consequence of Lemma 0.6. □

Lemma 0.9. *There exists an absolute constant $c > 0$ such that, given any $\delta \in (0, 1/2)$, we have, for all $x \geq 2$,*

$$\#\{p \in [x, 2x] : P(p+1) < x^\delta \text{ or } P(p+1) > x^{1-\delta}\} < c\,\delta\,\pi(x).$$

Proof. This is an immediate application of Theorem 4.2 in the book of Halberstam and Richert [43]. □

The following result will be used repetitively when trying to show that a number is normal using the known frequency of a given pattern of digits in the q-ary expansion of that number.

Lemma 0.10. *Fix an integer $q \geq 2$. Let $\gamma = \epsilon_1\epsilon_2\epsilon_3 \ldots \in \mathcal{A}_q^{\mathbb{N}}$. For each positive integer T, write γ_T for the T-digit word $\epsilon_1\epsilon_2 \ldots \epsilon_T$. Assume that, for every positive integer k and arbitrary distinct words $\beta_1, \beta_2 \in \mathcal{A}_q^k$, there exists an infinite sequence of positive integers $T_1 < T_2 < \cdots$ such that* *(i)* $\lim_{n\to\infty} \frac{T_{n+1}}{T_n} = 1$,

(ii) $\lim_{n\to\infty} \frac{1}{T_n} |\nu_{\beta_1}(\gamma_{T_n}) - \nu_{\beta_2}(\gamma_{T_n})| = 0$.

Then, the real number $0.\epsilon_1\epsilon_2\epsilon_3\ldots$ *is* q*-normal.*

Proof. It is easily seen that conditions (i) and (ii) imply that

$$\frac{1}{T} |\nu_{\beta_1}(\gamma_T) - \nu_{\beta_2}(\gamma_T)| \to 0 \qquad \text{as } T \to \infty$$

and consequently that

$$\frac{1}{T} \left| q^k \nu_{\beta_1}(\gamma_T) - \sum_{\beta_2 \in \mathcal{A}_q^k} \nu_{\beta_2}(\gamma_T) \right| \to 0 \qquad \text{as } T \to \infty. \tag{0.6}$$

But since

$$\sum_{\beta_2 \in \mathcal{A}_q^k} \nu_{\beta_2}(\gamma_T) = T + O(1),$$

it follows from (0.6) that

$$\frac{\nu_{\beta_1}(\gamma_T)}{T} = (1 + o(1)) \frac{1}{q^k} \qquad \text{as } T \to \infty,$$

thereby establishing that γ is a q-normal number and thus completing the proof of the lemma. □

Lemma 0.11. (Elliott) *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*

$$\sum_{\substack{n=1 \\ a_n \equiv 0 \pmod d}}^{N} f(n) = \kappa(d) X + R(N, d), \tag{0.7}$$

where X *and* R *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, $0 \le \kappa(p) < 1$. *Then, setting*

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

and letting z *be any real number satisfying* $\log z \ge 8 \max(\log r, S)$, *the estimate*

$$I(N, Q) := \sum_{\substack{n=1 \\ (a_n, Q)=1}}^{N} f(n) = \{1 + 2\theta_1 H\} X \prod_{p|Q} (1 - \kappa(p)) + 2\theta_2 \sum_{\substack{d|Q \\ d \le z^3}} 3^{\omega(d)} |R(N, d)| \tag{0.8}$$

holds uniformly for $r \geq 2$, where $|\theta_1| \leq 1$, $|\theta_2| \leq 1$ and

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

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

Proof. This result is Lemma 2.1 in the book of Elliott [37]. □

Lemma 0.12. *Given relatively prime polynomials $F_1, F_2 \in \mathbb{Z}[x]$, the congruences*

$$F_1(m) \equiv 0 \pmod a \quad \textit{and} \quad F_2(m) \equiv 0 \pmod a$$

have common roots for at most finitely many a's.

Proof. A proof of this result can be found in Tanaka [60]. □

Lemma 0.13. *Given any $r \in \mathbb{N}$ and setting $\pi_r(x) := \#\{n \leq x : \omega(n) = r\}$, there exist positive absolute constants c_1, c_2 such that*

$$\pi_r(x) \leq c_1 \frac{x}{\log x} \frac{(\log\log x + c_2)^{r-1}}{(r-1)!} \qquad (x \geq 3).$$

Proof. For a proof, see Hardy and Ramanujan [44]. □

Lemma 0.14 (Borel-Cantelli Lemma)**.** *Let $E_1, E_2, E_3, \ldots$ be an infinite sequence of events in some probability space. Assuming that the sum of the probabilities of the E_n's is finite, that is, $\sum_{n=1}^{\infty} P(E_n) < +\infty$, then the probability that infinitely many of them occur is 0.*

Proof. For a proof of this result, see the book of Janos Galambos [39]. □

Given a probability space $(\Omega, \mathcal{F}, P)$, we say that $A_1, A_2, \ldots$ is a list of *completely independent elements* of $\mathcal{F}$ if, given any finite increasing sequence of integers, say $i_1 < i_2 < \cdots < i_k$, we have $P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$.

The second Borel-Cantelli lemma can be considered as the converse of the classical Borel-Cantelli lemma. It can be stated as follows.

Lemma 0.15. *Let $(\Omega, \mathcal{F}, P)$ be a probability space and let $A_1, A_2, \ldots$ be a list of completely independent elements of $\mathcal{F}$. Letting E be as in Lemma 0.14 and assuming that*

$$\sum_{j=1}^{\infty} P(A_j) = \infty,$$

then $P(E) = 1$.

I. Construction of normal numbers by classified prime divisors of integers [14]

(Functiones et Approximatio, 2011)

Fix an integer $q \geq 2$ and let

$$\wp = \mathcal{R} \cup \wp_0 \cup \wp_1 \cup \cdots \cup \wp_{q-1}, \tag{1.1}$$

be a disjoint classification of primes.

Consider the function $H : \wp \to \mathcal{A}_q$ defined by

$$H(p) = \begin{cases} j & \text{if } p \in \wp_j \quad (j \in \mathcal{A}_q), \\ \Lambda & \text{if } p \in \mathcal{R}, \end{cases}$$

and further extend the domain of the function H to all prime powers p^α by simply setting $H(p^\alpha) = H(p)$.

We introduce the function $R : \mathbb{N} \to \mathcal{A}_q^*$ defined as follows. If $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, where $p_1 < \cdots < p_r$ are primes and each $\alpha_i \in \mathbb{N}$, we set

$$R(n) = H(p_1) \ldots H(p_r), \tag{1.2}$$

where on the right hand side of (1.2), we omit $H(p_i) = \Lambda$ if $p_i \in \mathcal{R}$. For convenience, we set $R(1) = \Lambda$.

For instance, choosing $\wp_0 = \{p : p \equiv 1 \pmod 4\}$, $\wp_1 = \{p : p \equiv 3 \pmod 4\}$ and $\mathcal{R} = \{2\}$, we get that

$$\{R(1), R(2), \ldots, R(15)\} = \{\Lambda, \Lambda, 1, \Lambda, 0, 1, 1, \Lambda, 1, 0, 1, 1, 0, 1, 10\}.$$

Now, consider the situation where $\wp = \mathcal{R} \cup \wp_0 \cup \ldots \cup \wp_{q-1}$ is a disjoint classification of primes, and let R be defined as in (1.2). Consider the number

$$\xi = 0.R(1)R(2)R(3)\ldots,$$

which represents an infinite sequence over $\mathcal{A}_q$ and which in turn, by concatenating the finite words $R(1)$, $R(2)$, $R(3)$, ..., can be considered as the q-ary expansion of a real number, namely the real number ξ. In what follows, we examine what other conditions are required in order to claim that the above number ξ is indeed a q-normal number.

Main results

Theorem 1.1. *Let $q \geq 2$ be a fixed integer and let $\wp = \mathcal{R} \cup \wp_0 \cup \cdots \cup \wp_{q-1}$ be a disjoint classification of primes. Assume that, for a certain constant $c \geq 5$, for each $j = 0, 1, \ldots, q-1$,*

$$\pi([u, u+v] \cap \wp_j) = \frac{1}{q}\pi([u, u+v]) + O\left(\frac{u}{\log^c u}\right)$$

uniformly for $2 \leq v \leq u$ as $u \to \infty$. Moreover, let R be as in (1.2) and set

$$\xi = 0.R(1)R(2)R(3)\ldots, \tag{1.3}$$

where the right hand side of (1.3) stands the q-ary expansion of a real number. Then ξ is a q-normal number.

Using the reduced residue class modulo a given integer $D \geq 3$, we may also create normal numbers.

Theorem 1.2. *Fix an integer $D \geq 3$ and let $h_0, h_1, \dots, h_{\phi(D)-1}$ be those positive integers $< D$ which are relatively prime with D. Then, define the function H on prime powers by*

$$H(p^a) = H(p) = \begin{cases} j & \text{if } p \equiv h_j \pmod{D}, \\ \Lambda & \text{if } p|D \end{cases}$$

and consider the corresponding arithmetic function T defined by

$$T(n) = T(p_1^{a_1} \cdots p_r^{a_r}) = H(p_1) \dots H(p_r).$$

Then, given a positive integer a with $(a, D) = 1$, the real number ξ whose $\phi(D)$-ary expansion is given by

$$\xi = 0.T(2+a)T(3+a)T(5+a)\dots T(p+a)\dots$$

is $\phi(D)$-normal.

Given a positive real number Y, then for each integer $n \geq 2$, let

$$A(n|Y) := \prod_{\substack{p^\alpha \| n \\ p \leq Y}} p^\alpha.$$

Theorem 1.3. *Let a be a positive integer. Let ε_x be a function which tends to 0 as $x \to \infty$ in such a way that $1/\varepsilon_x = o(\log\log x)$. Let $\mathcal{K}_x := \{K \in \mathbb{N} : P(K) \leq x^{\varepsilon_x}\}$. For each $K \in \mathcal{K}_x$, define*

$$\Delta_K(x) := \#\{p \leq x : A(p+a|x^{\varepsilon_x}) = K\}$$

and, for $gcd(a, K) = 1$,

$$\kappa(K) := \prod_{\substack{p < x^{\varepsilon_x} \\ gcd(p,Ka)=1}} \left(1 - \frac{1}{p-1}\right) \cdot \prod_{p|K} \left(1 - \frac{1}{p}\right) = \prod_{\substack{p < x^{\varepsilon_x} \\ gcd(p,Ka)=1}} \left(1 - \frac{1}{p-1}\right) \cdot \frac{\phi(K)}{K}. \tag{1.4}$$

Let also δ_x be a function satisfying $\lim_{x\to\infty} \delta_x = 0$ and $\lim_{x\to\infty} \delta_x/\varepsilon_x = +\infty$. Then, given any fixed $C > 0$,

$$\sum_{\substack{K \in \mathcal{K}_x,\ K < x^{\delta_x} \\ gcd(K,a)=1}} \left| \Delta_K(x) - \frac{\kappa(K)}{\phi(K)} \mathrm{li}(x) \right| \ll \exp\left\{ -\frac{1}{2} \frac{\delta_x}{\varepsilon_x} \log \frac{\delta_x}{\varepsilon_x} \right\} \cdot \pi(x) + O\left(\frac{x}{\log^C x} \right) + O(\varepsilon_x \pi(x)). \tag{1.5}$$

Moreover,

$$\lim_{x\to\infty} \frac{1}{\pi(x)} \sum_{\substack{K \in \mathcal{K}_x \\ gcd(K,a)=1}} \left| \Delta_K(x) - \frac{\kappa(K)}{\phi(K)} \mathrm{li}(x) \right| = 0. \tag{1.6}$$

We may also use the prime factors of the product of k consecutive integers to create normal numbers. the result goes as follows.

Let $k \geq 1$ be a fixed integer and set $E(n) := n(n+1)\cdots(n+k-1)$. Moreover, for each positive integer n, consider the function

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

We shall now define the sequence h_n on the prime powers q^β dividing $E(n)$ as follows:

$$h_n(q^\beta) = h_n(q) = \begin{cases} \Lambda & \text{if } q|e(n), \\ \ell & \text{if } q|n+\ell,\ 0 \leq \ell \leq k-1,\ \gcd(q, e(n)) = 1. \end{cases}$$

If $E(n) = q_1^{\beta_1} q_2^{\beta_2} \cdots q_r^{\beta_r}$ where $q_1 < q_2 < \cdots < q_r$ are primes an each $\beta_i \in \mathbb{N}$, then we set

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

Theorem 1.4. *Let k, E and S be as above. Let ξ be the real number whose k-ary expansion is given by*

$$\xi = 0.S(E(1))S(E(2))\ldots S(E(n))\ldots \tag{1.7}$$

Then, ξ is a k-normal number.

There is an analogous result for shifted primes.

Theorem 1.5. *Let $p_1 < p_2 < \cdots$ be the sequence of all primes, and let k, E and S be as above. Let ξ be the real number whose k-ary expansion is given by*

$$\xi = 0.S(E(p_1+1))S(E(p_2+1))\ldots$$

Then ξ is a k-normal number.

Here, we will only prove Theorem 1.1. To do so, we need two additional lemmas.

Lemma 1.1. *Fix an integer $q \geq 2$. Let w_x be a nondecreasing function which tends to $+\infty$ as $x \to \infty$. Moreover, let $\alpha = i_1 \ldots i_r \in \mathcal{A}_q^r$ be an arbitrary word and let R be as in (1.2), and define*

$$\begin{aligned} N_r(Y|w_x) &:= \#\{p_1^{a_1}\cdots p_r^{a_r} \leq Y : w_x < p_1 < \cdots < p_r\}, \\ N_r(Y|w_x;\alpha) &:= \#\{p_1^{a_1}\cdots p_r^{a_r} \leq Y : w_x < p_1 < \cdots < p_r,\ R(p_1^{a_1}\cdots p_r^{a_r}) = \alpha\}. \end{aligned}$$

Assume that, uniformly for $2 \leq v \leq u$, $j = 0, \ldots, q-1$,

$$\pi([u, u+v]|\wp_j) = \frac{1}{q}\pi([u, u+v]) + O\left(\frac{u}{\log^c u}\right) \qquad (u \to \infty)$$

holds for some constant $c \geq 5$. Further assume that $w_x \ll x_3$. Then, for $\sqrt{x} \leq Y \leq x$ and $1 \leq r \leq c_2 x_2$ (for some fixed positive constant c_2), as $x \to \infty$,

$$N_r(Y|w_x;\alpha) = (1+o(1))\frac{1}{q^r}N_r(Y|w_x).$$

Proof. This is a special case of Theorem 1 of De Koninck and Kátai [12]. □

For each $n \in \mathbb{N}$, define

$$e(n) := \prod_{\substack{p^a \| n \\ p \le w_x}} p^a \qquad \text{and} \qquad M(n) := \prod_{\substack{p^a \| n \\ p > w_x}} p^a.$$

Lemma 1.2. *Assume that the conditions of Lemma 1.1 are met and set*

$$\begin{aligned} S_r(Y|w_x) &:= \#\{n = e(n)M(n) \le Y : \omega(M(n)) = r\}, \\ S_r(Y|w_x;\alpha) &:= \#\{n = e(n)M(n) \le Y : \omega(M(n)) = r,\ R(M(n)) = \alpha\}. \end{aligned}$$

Then, as $x \to \infty$,

$$S_r(Y|w_x;\alpha) = (1 + o(1))\frac{1}{q^r} S_r(Y|w_x).$$

Proof. To prove Lemma 1.2, it is sufficient to observe that

$$\begin{aligned} S_r(Y|w_x;\alpha) &= \sum_{\substack{\nu \le x \\ p(\nu) \le w_x}} N_r(\frac{Y}{\nu}|w_x;\alpha), \\ S_r(Y|w_x) &= \sum_{\substack{\nu \le x \\ p(\nu) \le w_x}} N_r(\frac{Y}{\nu}|w_x), \end{aligned}$$

and thereafter to apply Lemma 1.1 and sum over all $\nu \le e^{w_x}$, say, and then show that the sum over those $\nu > e^{w_x}$ is negligible. □

Proof of Theorem 1.1.

Let $\lambda(\alpha)$ stand for the length of the word α over $\mathcal{A}_q$. Let $\beta = b_1 \dots b_k \in \mathcal{A}_q^k$ and $\omega^*(n) := \sum_{\substack{p|n \\ p \notin \mathcal{R}}} 1$, so that $\omega^*(n) = \lambda(R(n))$.

Since $\mathcal{R}$ is a finite set, it is clear that

$$T_N := \sum_{n \le N} \omega^*(n) = N \log\log N + O(N) \qquad (N \to \infty). \tag{1.8}$$

Now, for each positive integer j, let $Y_j = 2^j$ and $\eta_j := R(2^j) \dots R(2^{j+1} - 1)$, so that $\xi = 0.\eta_1\eta_2\eta_3 \dots$ Recall that $\nu_\beta(\alpha)$ stands for the number of occurrences of β as a subword in α.

It is clear that given $\beta \in \mathcal{A}_q^k$, for each positive integer j such that $Y_j < N$, we have

$$\sum_{n=Y_j}^{Y_{j+1}-1} \nu_\beta(R(n)) \le \nu_\beta(\eta_j) \le \sum_{n=Y_j}^{Y_{j+1}-1} \nu_\beta(R(n)) + (k+1)Y_j \tag{1.9}$$

and

$$\sum_{n=Y_j}^{N} \nu_\beta(R(n)) \le \nu_\beta(R(Y_j)\dots R(N)) \le \sum_{n=Y_j}^{N} \nu_\beta(R(n)) + (k+1)(N-Y_j+1). \tag{1.10}$$

Assume that $w_x \ll x_5$, let j be fixed and set $x = Y_j$. Then, for any integer $n \in [Y_j, Y_{j+1}]$, we clearly have

$$\nu_\beta(R(M(n))) \le \nu_\beta(R(n)) \le \omega(e(n)) + k + \nu_\beta(R(M(n))).$$

Observe that

$$\sum_{n=Y_j}^{N} (\omega(e(n)) + k) \le (N - Y_j)(k + \pi(w_x)).$$

We shall now provide asymptotic estimates for

$$K_j := \sum_{n=Y_j}^{Y_{j+1}-1} \nu_\beta(R(M(n))) \qquad \text{and} \qquad K_{N,Y_j} := \sum_{n=Y_j}^{N} \nu_\beta(R(M(n))). \tag{1.11}$$

To do so, we shall first find an upper bound for the number of integers $n \in [Y_j, Y_{j+1}-1]$ for which $\omega(M(n)) \ge 2x_2$. In fact, we will prove that

$$\Sigma_0 := \sum_{\substack{Y_j \le n < Y_{j+1} \\ \omega(M(n)) \ge 2x_2}} \omega(M(n)) = O(Y_j). \tag{1.12}$$

Indeed, it follows from Lemma 0.13 that

$$\pi_r(Y_j) \le \frac{c_3 Y_j}{\log Y_j} \frac{(\log\log Y_j + c_4)^{r-1}}{(r-1)!},$$

so that

$$\Sigma_0 = \sum_{r=\lfloor 2x_2 \rfloor}^{\infty} r\pi_r(Y_j) \le c_3 \sum_{r \ge 2x_2} \frac{rY_j}{\log Y_j} \frac{(\log\log Y_j + c_4)^{r-1}}{(r-1)!} \ll Y_j,$$

thereby establishing our claim (1.12).

With this result in mind, we now only need to consider those integers n for which $r = \omega(M(n)) \le 2x_2$.

So let $\alpha = e_1 \dots e_r \in \mathcal{A}_q^r$, with $r \le 2x_2$.

From Lemma 1.2, we have, as $x \to \infty$,

$$\begin{aligned} S_r(Y|w_x;\alpha) &= \#\{n = e(n)M(n) \le Y : \omega(M(n)) = r,\ R(M(n)) = \alpha\} \\ &= (1+o(1))\frac{1}{q^r} S_r(Y|w_x), \end{aligned}$$

so that

$$S_r(Y_{j+1}-1|w_x;\alpha) \quad - \quad S_r(Y_j - 1|w_x;\alpha)$$

$$= \quad (1+o(1))\frac{1}{q^r}\left(S_r(Y_{j+1}-1|w_x) - S_r(Y_j-1|w_x)\right).$$

Similarly,

$$S_r(N|w_x;\alpha) - S_r(Y_j-1|w_x,1) = (1+o(1))\frac{1}{q^r}\left(S_r(N|w_x) - S_r(Y_j|w_x)\right).$$

From these observations and in light of (1.12), it follows that, as $x\to\infty$,

$$K_j = (1+o(1))\sum_{r\le 2x_2}\frac{1}{q^r}\left(\sum_{\alpha\in\mathcal{A}_q^r}\nu_\beta(\alpha)\right)(S_r(2Y_j|w_x) - S_r(Y_j|w_x)) + O(Y_j). \tag{1.13}$$

On the other hand, clearly, for any $\beta\in\mathcal{A}_q^k$,

$$\sum_{\alpha\in\mathcal{A}_q^r}\nu_\beta(\alpha) = \begin{cases} 0 & \text{if } r<k,\\ (r-k+1)q^{r-k} & \text{if } r\ge k.\end{cases}$$

Substituting this in (19.7), it follows that, as $x\to\infty$,

$$K_j = (1+o(1))\sum_{r=k}^{\lfloor 2x_2\rfloor}\frac{r-k+1}{q^k}(S_r(2Y_j|w_x) - S_r(Y_j|w_x)) + O(Y_j). \tag{1.14}$$

Since the contribution to K_j of those integers r for which $|r-x_2|\ge x_2^{3/4}$ is clearly $o(x_2Y_j)$, estimate (1.14) becomes

$$K_j = (1+o(1))\frac{x_2}{q^k}\sum_{|r-x_2|<x_2^{3/4}}(S_r(2Y_j|w_x) - S_r(Y_j|w_x)) + o(x_2Y_j) \qquad (x\to\infty). \tag{1.15}$$

On the other hand, since the normal order of $\omega(n)$ is $\log\log n$, it is clear that

$$\sum_{|r-x_2|<x_2^{3/4}}(S_r(2Y_j|w_x) - S_r(Y_j|w_x)) = (1+o(1))(2Y_j-Y_j) = (1+o(1))Y_j \quad (x\to\infty). \tag{1.16}$$

Substituting (1.16) in (1.15), we obtain

$$K_j = (1+o(1))\frac{x_2}{q^k}Y_j \qquad (x\to\infty). \tag{1.17}$$

It remains to estimate K_{N,Y_j} (defined in (1.11)) in the case $Y_j<N<Y_{j+1}$.

Let $\varepsilon_1,\varepsilon_2,\ldots$ be a sequence of positive numbers which tends to 0 very slowly.

If $N-Y_j\ge\varepsilon_jY_j$, then, in light of Lemma 1.1 and proceeding as above, one can prove that

$$K_{N,Y_j} = (1+o(1))\frac{x_2}{q^k}(N-Y_j) \qquad (x\to\infty),$$

whereas if $N-Y_j<\varepsilon_jY_j$, we have

$$K_{N,Y_j} = O(\varepsilon_jY_j\log\log N) \qquad (Y_j\to\infty).$$

Hence, in light of these observations and of (1.17), it follows from inequalities (1.9) and (1.10) that

$$\nu_\beta(\eta_j) = (1+o(1))(Y_{j+1}-Y_j)\frac{\log\log Y_j}{q^k} \qquad (Y_j \to \infty) \tag{1.18}$$

and that

$$\nu_\beta(R(Y_j)\dots R(N)) = (1+o(1))(N-Y_j)\frac{\log\log Y_j}{q^k} + O(\varepsilon_j Y_j \log\log Y_j) \qquad (Y_j \to \infty). \tag{1.19}$$

Now, consider the q-ary expansion of the number ξ, that is $\xi = 0.R(1)R(2)\dots$. For each positive integer M, let $\xi^{(M)} = R(1)R(2)\dots R(M)$. We would like to approximate $\nu_\beta(\xi^{(M)})$. Given a fixed positive integer M, let N be defined implicitly by

$$\lambda(R(1)\dots R(N)) \le M < \lambda(R(1)\dots R(N+1)).$$

Hence, in light of (1.8), M and N are tied by the relation

$$M = T_N + O(N) = N\log\log N + O(N) \qquad (N\to\infty).$$

We therefore have that, for $Y_j \le N < Y_{j+1}$,

$$\nu_\beta(\xi^{(M)}) = \nu_\beta(R(1)\dots R(Y_j-1)) + \nu_\beta(R(Y_j)\dots R(N)) + O(\varepsilon_j N\log\log N),$$

so that

$$\frac{\nu_\beta(\xi^{(M)})}{M} = \frac{\nu_\beta(R(1)\dots R(Y_j-1))}{M} + \frac{\nu_\beta(R(Y_j)\dots R(N))}{M} + O\left(\frac{\varepsilon_j N\log\log N}{M}\right). \tag{1.20}$$

Taking into account estimates (1.18) and (1.19), it follows from (1.20) that

$$\frac{\nu_\beta(\xi^{(M)})}{M} = (1+o(1))\frac{1}{q^k}\frac{T_{Y_j}}{M} + (1+o(1))\frac{T_N - T_{Y_j}}{q^k M} + O\left(\frac{\varepsilon_j N\log\log N}{M}\right) \qquad (N\to\infty),$$

which implies, since $\varepsilon_j \to 0$ as $j\to\infty$, that

$$\lim_{M\to\infty}\frac{\nu_\beta(\xi^{(M)})}{M} = \frac{1}{q^k},$$

thus completing the proof of Theorem 1.1.

II. On a problem on normal numbers raised by Igor Shparlinski [13]

(Bulletin of the Australian Mathematical Society, 2011)

Theorem 2.1. *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 number*

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

is a normal number.

Theorem 2.2. *Let F be as in Theorem 2.1. Then the number*

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

is a normal number.

We only give here a sketch of the proof of Theorem 2.2.

Fix an integer $q \geq 2$. As usual, we let $L(n) := L_q(n) = \left\lfloor \frac{\log n}{\log q} \right\rfloor + 1$, that is, the number of digits of n in base q. Recall also that given a word $\theta = i_1 i_2 \ldots i_t \in \mathcal{A}_q^t$, we write $\lambda(\theta) = t$ and that we let $\nu_\beta(\theta)$ stand for the number of times that the subword β occurs in the word θ.

A key element of the proof of Theorem 2.2 is Lemma 0.5.

Now, given a large number x, let $I_x = [x, 2x]$ and set

$$\theta = \overline{F(P(n_0))}\,\overline{F(P(n_1))}\ldots\overline{F(P(n_T))},$$

where n_0 is the smallest integer in I_x, and n_T the largest.

It is clear that the proof of Theorem 2.2 will be complete if we can show that, given an arbitrary word $\beta \in \mathcal{A}_q^k$, we have

$$\frac{\nu_\beta(\theta)}{\lambda(\theta)} \sim \frac{1}{q^k} \qquad (x \to \infty).$$

Since the number of digits of each integer $n \in I_x$ is of order $\log x / \log q$, one can easily see, using the definition of θ, that

$$\lambda(\theta) = rx\frac{\log x}{\log q} + O(x) \approx x \log x, \tag{2.1}$$

thus revealing the true size of $\lambda(\theta)$.

Letting δ be a small positive number, it follows from Lemma 0.6 that the number of integers $n \in I_x$ for which either $P(n) < x^\delta$ or $P(n) > x^{1-\delta}$ is $\leq c\delta x$, implying that we may write

$$\nu_\beta(\theta) = \sum_{\substack{n \in I_x \\ x^\delta \leq P(n) \leq x^{1-\delta}}} \nu_\beta(\overline{F(P(n))}) + O(T) + O(\delta x \log x). \tag{2.2}$$

Let us now introduce the finite sequence $u_0, u_1, \ldots, u_H$ defined by $u_0 = x^\delta$ and $u_j = 2u_{j-1}$ for each $1 \le j \le H$, where H is the smallest positive integer for which $2^H u_0 > x^{1-\delta}$, so that $H = \left\lfloor \dfrac{(1-2\delta)\log x}{\log 2} \right\rfloor + 1$.

Now, for each prime p, let $R(p) := \#\{n \in I_x : P(n) = p\}$. We have, in light of (2.2) and the fact that $T = O(x)$,

$$\nu_\beta(\theta) = \sum_{x^\delta \le p \le x^{1-\delta}} \nu_\beta(\overline{F(p)}) R(p) + O(\delta x \log x). \tag{2.3}$$

Let $\beta_1, \beta_2 \in \mathcal{A}_q^k$ with $\beta_1 \ne \beta_2$. Then, using (2.3), we have

$$\begin{aligned} |\nu_{\beta_1}(\theta) - \nu_{\beta_2}(\theta)| &\le \sum_{x^\delta \le p \le x^{1-\delta}} \left| \nu_{\beta_1}(\overline{F(p)}) - \nu_{\beta_2}(\overline{F(p)}) \right| R(p) + O(\delta x \log x) \\ &= \sum_{j=0}^{H-1} \sum_{u_j \le p < u_{j+1}} \left| \nu_{\beta_1}(\overline{F(p)}) - \nu_{\beta_2}(\overline{F(p)}) \right| R(p) + O(\delta x \log x) \\ &= \sum_{j=0}^{H-1} S_j(x) + O(\delta x \log x), \qquad (2.4) \end{aligned}$$

say.

Using Lemma 0.7, we have, as $x \to \infty$,

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

from which it follows that

$$S_j(x) \le \frac{2x}{u_j} \sum_{u_j \le p < u_{j+1}} \left| \nu_{\beta_1}(\overline{F(p)}) - \nu_{\beta_2}(\overline{F(p)}) \right|. \tag{2.5}$$

Set $\kappa_u := \log\log u$. We will say that $p \in [u_j, u_{j+1})$ is a *good prime* if

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

and a *bad prime* otherwise.

Splitting the sum $S_j(x)$ into two sums, one running on the good primes and the other one running on the bad primes, it follows from (2.5) and Lemma 0.5 that

$$S_j(x) \le \frac{2x}{u_j}\kappa_{u_j}\sqrt{L(u_j^r)}\frac{u_j}{\log u_j} + \frac{2x}{u_j}\frac{u_j \log u_{j+1}}{(\log u_j)\kappa_{u_j}^2}$$

$$= \quad 2x \cdot \left\{ \frac{\kappa_{u_j} \sqrt{L(u_j^r)}}{\log u_j} + \frac{\log u_{j+1}}{(\log u_j)\kappa_{u_j}^2} \right\}$$

$$\leq \quad 4x \left\{ \frac{r \log\log u_j}{\sqrt{\log u_j}} + \frac{1}{(\log\log u_j)^2} \right\}.$$

Summing the above inequalities for $j = 0, 1, \ldots, H-1$, and taking into account that $H \ll \log x$, we obtain that $\sum_{j=0}^{H-1} S_j(x) = o(x \log x)$ as $x \to \infty$ and thus that, in light of (2.4), for some constant $c > 0$,

$$(2.6) \qquad |\nu_{\beta_1}(\theta) - \nu_{\beta_2}(\theta)| \leq c\delta x \log x + o(x \log x).$$

Now let ξ_N be the first N digits of the infinite word

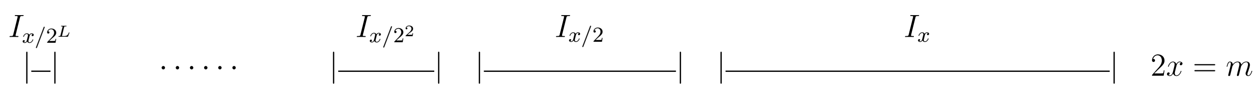
$$\overline{F(P(2))}\ \overline{F(P(3))}\ \overline{F(P(4))} \ldots$$

and let m be the unique integer such that

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

where $\lambda(\widetilde{\xi_N}) \leq N < \lambda(\widetilde{\xi_N}\overline{F(P(m+1))})$, so that $\lambda(\overline{F(P(m+1))}) \ll \log m \ll \log N$, implying in particular that ξ_N and $\widetilde{\xi_N}$ have the same digits except for at most the last $\lfloor \log N \rfloor$ ones.

Let $2x = m$ and consider the intervals $I_x, I_{x/2}, I_{x/(2^2)}, \ldots, I_{x/(2^L)}$, where $L = 2\lfloor \log\log x \rfloor$, that is,



and write

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

where a and b are the smallest and largest integers 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 $I_{x/(2^L)}$.

It is clear that

$$(2.7) \qquad \left| \nu_{\beta_1}(\widetilde{\xi_N}) - \nu_{\beta_2}(\widetilde{\xi_N}) \right| \leq |\nu_{\beta_1}(\mu) - \nu_{\beta_2}(\mu)| + \sum_{j=0}^{L} |\nu_{\beta_1}(\tau_j) - \nu_{\beta_2}(\tau_j)|$$

and that

$$(2.8) \qquad \nu_\beta(\mu) \leq \lambda(\mu) \leq \frac{x}{2^L} \cdot r \log x = o(x).$$

Applying estimate (2.6) $L+1$ times (with $\theta = \widetilde{\xi_N}$) by replacing successively $2x$ by x, $x/2$, $x/2^2$, $\ldots$, $x/2^L$, we obtain from (2.7) and in light of (2.8), that

$$(2.9) \qquad \left| \nu_{\beta_1}(\widetilde{\xi_N}) - \nu_{\beta_2}(\widetilde{\xi_N}) \right| \leq c\delta N + o(N) \qquad (N \to \infty).$$

Now, one can easily see that

$$\sum_{\gamma\in\mathcal{A}_q^k}\nu_\gamma(\theta)=\lambda(\theta)-k+1,$$

from which it follows that

$$q^k\nu_\beta(\theta)-\lambda(\theta)=\sum_{\gamma\in\mathcal{A}_q^k}(\nu_\beta(\theta)-\nu_\gamma(\theta))+O(1),$$

implying that, setting $\theta=\xi_N$ and using (2.9),

$$\begin{aligned}\left|q^k\nu_\beta(\xi_N)-\lambda(\xi_N)\right| &\le \sum_{\gamma\in\mathcal{A}_q^k}|\nu_\beta(\xi_N)-\nu_\gamma(\xi_N)|+O(1)\\ &\le (c\delta N+o(N))q^k,\end{aligned}$$

from which it follows that, observing that $\lambda(\xi_N)=N$,

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

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

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

thus establishing that ξ is q-normal.

III. Normal numbers created from primes and polynomials [16]

(Uniform Distribution Theory, 2012)

In 1995 (see [12]), we observed that one can map the set of positive integers n into the set of q-ary integers by using the multiplicative structure of the positive integers n. Indeed, we proved that if we subdivide the set of primes $\wp$ into q distinct subsets $\wp_j$, $j=0,1,\dots,q-1$, of essentially the same size, and if $p_1<\dots<p_r$ are the prime divisors of n with $p_j\in\wp_{\ell_j}$ for certain $\ell_j\in\{0,1,\dots,q-1\}$, then, for almost all n, the corresponding number $\ell_1\dots\ell_r$ appears essentially at the expected frequency, namely $1/q^r$. Using this result, we recently constructed (see [14]) large families of normal numbers.

In this paper, we further expand on this approach but this time using the prime factorization of the values taken by primitive irreducible polynomials defined on the set of positive integers.

Let $Q_1,Q_2,\dots,Q_h\in\mathbb{Z}[x]$ be distinct irreducible primitive monic polynomials each of degree no larger than 3. Recall that a polynomial with integer coefficients is said to be *primitive* if the greatest common divisor of its coefficients is 1. For each $\nu = 0, 1, 2, \ldots, D-1$, let $c_1^{(\nu)}, c_2^{(\nu)}, \ldots, c_h^{(\nu)}$ be distinct integers, $F_\nu(x) = \prod_{j=1}^h Q_j(x + c_j^{(\nu)})$, with $F_\nu(0) \neq 0$ for each ν. Moreover, assume that the integers $c_i^{(\nu)}$ are chosen in such a way that $F_\nu(x)$ are squarefree polynomials and $\gcd(F_\nu(x), F_\mu(x)) = 1$ when $\nu \neq \mu$.

Let $\wp_0$ be the set of prime numbers p for which there exist $\mu \neq \nu$ and $m \in \mathbb{N}$ such that $p|\gcd(F_\nu(m), F_\mu(m))$. It follows from Lemma 0.12 that $\wp_0$ is a finite set. Now let

$$U(n) = F_0(n)F_1(n)\cdots F_{D-1}(n) = \vartheta\, q_1^{a_1} q_2^{a_2} \cdots q_r^{a_r},$$

where $\vartheta \in \mathcal{N}(\wp_0)$ and $q_1 < q_2 < \cdots < q_r$ are primes not belonging to $\mathcal{N}(\wp_0)$ with positive integers a_i. Then, let h_n be defined on the prime divisors q^a of $U(n)$ by

$$h_n(q^a) = h_n(q) = \begin{cases} \Lambda & \text{if } q|\vartheta, \\ \ell & \text{if } q|F_\ell(n),\ q \notin \wp_0 \end{cases}$$

and further define α_n as

$$\alpha_n = h_n(q_1^{a_1}) h_n(q_2^{a_2}) \ldots h_n(q_r^{a_r}),$$

where on the right hand side we omit Λ when $h_n(q_i^{a_i}) = \Lambda$ for some i. Finally, we let η be the real number whose D-ary expansion is

$$\eta = 0.\alpha_1\alpha_2\alpha_3\ldots \tag{3.1}$$

As a simple example, take $h = 1$, $Q_1(x) = x$, $F_\nu(x) = x + \nu$ for $\nu = 0, 1, \ldots, D-1$, in which case we have $\wp_0 = \{p : p \leq D-1\}$. Then,

$$U(n) = n(n+1)\cdots(n+D-1) = e(n) q_1^{a_1} \cdots q_r^{a_r},$$

where $e(n) := \prod_{\substack{q^\alpha \| U(n) \\ q \leq D-1}} q^\alpha$, so that

$$h_n(q^a) = h_n(q) = \begin{cases} \Lambda & \text{if } q|e(n), \\ \ell & \text{if } q|n+\ell,\ q \notin \wp_0 \end{cases}$$

and

$$\alpha_n = h_n(q_1^{a_1}) h_n(q_2^{a_2}) \ldots h_n(q_r^{a_r}),$$

thus giving rise to the number

$$\eta = 0.\alpha_1\alpha_2\alpha_3\ldots$$

In the particular case $D = 5$, we get $U(n) = n(n+1)(n+2)(n+3)(n+4)$ so that $\wp_0 = \{2, 3\}$ and

$$h_n(q^a) = h_n(q) = \begin{cases} \Lambda & \text{if } q \in \{2, 3\}, \\ \ell & \text{if } q|n+\ell,\ q \geq 5 \text{ where } \ell \in \{0, 1, 2, 3, 4\}. \end{cases}$$

In this case, one can check that

$$\eta = 0.\alpha_1\alpha_2\alpha_3\alpha_4\alpha_5\ldots = 0.43241302\ldots$$

Theorem 3.1. *The number η defined by (3.1) is a normal number.*

Theorem 3.2. *With the notations above and assuming that deg$(Q_j) \leq 2$ for $j = 1, 2, \ldots, h$, then the number*

$$\xi = 0.\alpha_2\alpha_3\alpha_5 \ldots \alpha_p \ldots$$

(where the above subscripts run over primes p) is a normal number.

We will only prove Theorem 3.1. However, in order to do so, we need to prove a few extra lemmas.

We start with the well known result.

Lemma 3.1. *Let $F(m)$ be an arbitrary primitive polynomial with integer coefficients and of degree ν. Let D be the discriminant of F and assume that $D \neq 0$. Let $\rho(m)$ be the number of solutions n of $F(n) \equiv 0 \pmod{m}$. Then ρ is a multiplicative function whose values on the prime powers p^α satisfy*

$$\rho(p^\alpha) \qquad \begin{cases} = \rho(p) & \text{if } p \nmid D, \\ \leq 2D^2 & \text{if } p | D. \end{cases}$$

Moreover, there exists a positive constant $c = c(f)$ such that $\rho(p^\alpha) \leq c$ for all prime powers p^α.

Lemma 3.2. *If $g \in \mathbb{Q}[x]$ is an irreducible polynomial and $\rho(m)$ stands for the number of residue classes mod m for which $g(n) \equiv 0 \pmod{m}$, then*

(i) $\displaystyle\sum_{p \leq x} \rho(p) = \frac{x}{\log x} + O\left(\frac{x}{\log^2 x}\right);$

(ii) $\displaystyle\sum_{p \leq x} \frac{\rho(p)}{p} = \log\log x + C + O\left(\frac{1}{\log x}\right).$

Proof. This result is due to Landau [51]. □

Lemma 3.3. *Let F be a squarefree polynomial with integer coefficients and of positive degree such that the degree of each of its irreducible factors is of degree no larger than 3. Let $Y(x)$ be a function which tends to $+\infty$ as $x \to +\infty$. Then*

$$\lim_{x\to\infty} \frac{1}{x} \#\{n \leq x : p^2 | F(n) \text{ for some } p > Y(x)\} = 0.$$

Proof. For a proof, see the book of Hooley [45] (pp. 62-69). □

Lemma 3.4. *There exists a positive constant $c = c(h, D)$ such that*

$$\frac{1}{x} \sum_{n \leq x} |\omega(U(n)) - hD \log\log x|^2 \leq c \log\log x, \tag{3.2}$$

$$(3.3)\qquad \frac{1}{x}\sum_{\substack{n\le x\\ \omega(U(n))>hDx_2+cx_2^{3/4}}}\omega(U(n))\ll\sqrt{\log\log x}$$

and

$$(3.4)\qquad \frac{1}{x}\sum_{\substack{n\le x\\ \omega(U(n))<hDx_2-cx_2^{3/4}}}\omega(U(n))\ll\sqrt{\log\log x}.$$

Proof. First observe that

$$(3.5)\qquad \omega(U(n))=\sum_{j=0}^{D-1}\omega(F_j(n))+O(1),$$

where the term $O(1)$ accounts for the possible common prime divisors of $F_\nu(n)$ and $F_\mu(n)$, which as we saw are in finite number.

From the Turán-Kubilius inequality,

$$(3.6)\qquad \frac{1}{x}\sum_{n\le x}\left(\omega(F_\nu(n))-\sum_{p\le x}\frac{\rho_{F_\nu}(p)}{p}\right)^2<c\left(1+\sum_{p\le x}\frac{\rho_{F_\nu}(p)}{p}\right).$$

On the other hand, it follows from Lemma 3.2 (ii) that

$$(3.7)\qquad \sum_{p\le x}\frac{\rho_{F_\nu}(p)}{p}=\sum_{j=0}^{h-1}\sum_{p\le x}\frac{\rho_{Q_j(x+c_j^{(\nu)})}(p)}{p}=h\log\log x+O(1).$$

Combining (3.5), (3.6) and (3.7), inequality (3.2) follows.

Setting

$$\Sigma_A:=\sum_{\substack{n\le x\\ \omega(U(n))>hDx_2+cx_2^{3/4}}}1$$

and using the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
&\sum_{\substack{n\le x\\ \omega(U(n))>hDx_2+cx_2^{3/4}}}\omega(U(n))\\
&\quad=\sum_{\substack{n\le x\\ \omega(U(n))>hDx_2+cx_2^{3/4}}}(\omega(U(n))-hDx_2)\quad+\quad hDx_2\sum_{\substack{n\le x\\ \omega(U(n))>hDx_2+cx_2^{3/4}}}1\\
(3.8)\qquad&\quad\le\Sigma_A^{1/2}\times\left(\sum_{\substack{n\le x\\ \omega(U(n))>hDx_2+cx_2^{3/4}}}|\omega(U(n))-hDx_2|^2\right)^{1/2}+hDx_2\,\Sigma_A.
\end{aligned}$$

Now, it follows from (3.2) that

$$(3.9)\qquad \Sigma_A\le\frac{x}{\sqrt{x_2}}.$$

Hence, in light of (3.2) and (3.9), estimate (3.8) yields

$$\sum_{\substack{n\le x\\ \omega(U(n))>hDx_2+cx_2^{3/4}}} \omega(U(n)) \ll \Sigma_A^{1/2}\sqrt{x}\cdot x_2^{1/2} + x_2\Sigma_A \ll x\,x_2^{1/4} + x\,x_2^{1/2} \ll x\,x_2^{1/2},$$

thereby completing the proof of inequality (3.3). Clearly, (3.4) can be obtained in a similar way. □

Let $\varepsilon_x = 1/\sqrt{x_2}$, $Y_x = \exp\{x_1^{\varepsilon_x}\}$ and $Z_x = \exp\{x_1^{1-\varepsilon_x}\}$. Also, let

$$\wp_1 = \{p : p \le Y_x, p \notin \wp_0\}, \quad \wp_2 = \{p : Y_x < p < Z_x\}, \quad \wp_3 = \{p : p \ge Z_x\}.$$

Finally, for each $j = 0, 1, 2, 3$, set $\omega_j(n) = \sum_{\substack{p\mid n\\ p\in\wp_j}} 1$.

Lemma 3.5. *With the above notation, we have*

$$\sum_{n\le x}\omega_1(U(n)) \ll x\sum_{p\le Y_x}\frac{1}{p} \ll x\varepsilon_x x_2 = x\sqrt{x_2}, \tag{3.10}$$

$$\sum_{n\le x}\omega_3(U(n)) \ll x\sum_{Z_x\le p<x^{1/4}}\frac{1}{p} + O(x) \ll x\sqrt{x_2}. \tag{3.11}$$

Proof. These two estimates are straightforward. □

Let us write each positive integer n as $n = A(n)B(n)C(n)$, where $A(n) \in \mathcal{N}(\wp_0 \cup \wp_1)$, $B(n) \in \mathcal{N}(\wp_2)$ and $C(n) \in \mathcal{N}(\wp_3)$.

Lemma 3.6. *Let $m_0, m_1, \ldots, m_{D-1}$ be squarefree numbers belonging to $\mathcal{N}(\wp_2)$, with $M = m_0m_1\cdots m_{D-1} \le \sqrt{x}$. Let $T(x|m_0, m_1, \ldots, m_{D-1})$ be the number of those integers $n \le x$ for which $B(F_j(n)) = m_j$ for $j = 0, 1, \ldots, D-1$. Then,*

$$\left|T(x|m_0,m_1,\ldots,m_{D-1}) - \frac{x\rho(M)\phi(M)}{M^2}\prod_{p\in\wp_2}\left(1-\frac{D\rho_F(p)}{p}\right)K(M)\right| \ll \frac{x\rho(M)}{M}\exp\{-x_1^{\varepsilon_x}\}, \tag{3.12}$$

where

$$K(M) = \prod_{p\mid M}\left(1-\frac{D\rho_F(p)}{p}\right)^{-1}.$$

Remark 3.1. *Observe that $K(M) = 1 + o(1)$ as $M \to \infty$.*

Proof. First observe that $\rho_{F_\nu}(n) = \rho_{F_\mu}(n)$ for every ν and μ, while $\gcd(m_\nu, m_\mu) = 1$ whenever $\nu \ne \mu$. Thus, M is squarefree as well. For convenience, let $\rho = \rho_{F_\nu}$. Using these facts, it is clear that the congruences

$$B(F_j(m)) \equiv 0 \pmod{m_j} \qquad (j = 0, 1, \ldots, D-1)$$

hold for $n \equiv \ell_i \pmod M$, $i = 1, 2, \ldots, \rho(M)$.

Let us now consider $\ell = \ell_i$ for a fixed $i \in [1, \rho(M)]$ and define

$$\begin{aligned} \varphi_j(k) &= \frac{F_j(\ell + kM)}{m_j} \qquad (j = 0, 1, \ldots, D-1), \\ \Phi(k) &= \varphi_0(k)\varphi_1(k)\cdots\varphi_{D-1}(k). \end{aligned} \tag{3.13}$$

Finally, let $Q = \prod_{p \in \wp_2} p$.

We now apply Lemma 0.11 with $f(k) = 1$, $a_k = \Phi(k)$ and $X = x/M$, and obtain an estimate for each corresponding $I_i(X, Q)$ (to the function $I(X, Q)$ defined in relation (0.8)) for the particular choice $\ell = \ell_i$. With this set up, we have

$$T(x|m_0, m_1, \ldots, m_{D-1}) = \sum_{i=1}^{\rho(M)} I_i(X, Q). \tag{3.14}$$

Observe that $\eta(p^\alpha) = \eta(p) = 0$ if $p \in \wp_1$. On the other hand, for $p \in \wp_2 \cap \wp_3$, we have $\rho_{\varphi_j}(p^\alpha) = \rho_{\varphi_j}(p)$ and also that if $p|m_j$, then $\rho_{\varphi_j}(p) = 1$ and $\rho_{\varphi_\ell}(p) = 0$ for $\ell \neq j$, while on the other hand if $(p, M) = 1$, then $\rho_{\varphi_j}(p) = \rho(p)$ for $j = 0, 1, \ldots, D-1$.

Now we denote by $\eta(M)$ the number of those k mod M such that $\Phi(k) \equiv 0 \pmod M$. Then one can easily show that

$$\eta(p^\alpha) = \eta(p) = \begin{cases} 0 & \text{if } p \in \wp_1, \\ \rho_{\varphi_j}(p) = 1 & \text{if } p|m_j, \\ \rho(p) & \text{if } p \in \wp_2 \cap \wp_3,\ (p, M) = 1. \end{cases} \tag{3.15}$$

It is also clear that the error term in (0.7) satisfies

$$|R(X, d)| \leq D\rho(d). \tag{3.16}$$

It follows from Lemma 0.11 that

$$I_i(X, Q) = (1 + O(H))\frac{x}{M}\prod_{p|Q}\left(1 - \frac{\eta(p)}{p}\right) + O\left(\sum_{\substack{d|Q \\ d \leq z^3}} 3^{\omega(d)}|R(X, d)|\right). \tag{3.17}$$

Using the notation of Lemma 0.11, we have

$$S = \sum_{p|Q} \frac{\eta(p)}{p - \eta(p)} \log p,$$

and one can show that there exist two positive constants $c_1 < c_2$ such that

$$c_1 < \frac{S}{(\log x)^{\varepsilon_x}} < c_2. \tag{3.18}$$

Moreover, we have that $\log r = (\log x)^{\varepsilon_x}$. So, we choose $\log z = (\log x)^{\delta_x}$, with $0 < \varepsilon_x < \delta_x$, where δ_x is a function which tends to 0 as $x \to \infty$ and which will be determined later.

We can prove that for $z \geq 2$,

$$\sum_{\substack{d|Q \\ d \leq z^3}} 3^{\omega(d)} \eta(d) \leq c z^3 (\log z)^K, \tag{3.19}$$

for a suitable large constant K. Indeed,

$$\begin{aligned} \sum_{d \leq Y} 3^{\omega(d)} \eta(d) |\mu(d)| \log d &\leq \sum_{pu \leq Y} 3^{\omega(pu)} (\log p) \eta(p) \eta(u) |\mu(u)| \\ &\leq 3 \sum_{u \leq Y} 3^{\omega(u)} \eta(u) |\mu(u)| \sum_{p \leq Y/u} \eta(p) \log p. \end{aligned} \tag{3.20}$$

Since $\sum_{p \leq Y/u} \eta(p) \log p \leq c \frac{Y}{u}$, (3.20) becomes

$$\begin{aligned} \sum_{d \leq Y} 3^{\omega(d)} \eta(d) |\mu(d)| \log d &\leq cY \sum_{u \leq Y} \frac{3^{\omega(u)} \eta(u)}{u} |\mu(u)| \\ &\leq cY \prod_{p \leq Y} \left(1 + \frac{3\eta(p)}{p}\right) \leq cY \exp\left\{3 \sum_{p \leq Y} \frac{\eta(p)}{p}\right\} \\ &\leq cY \exp(3h \log\log Y) = cY (\log Y)^{3h}. \end{aligned} \tag{3.21}$$

Let us write

$$\sum_{d \leq Y} 3^{\omega(d)} \eta(d) |\mu(d)| = \sum_{d \leq \sqrt{Y}} + \sum_{\sqrt{Y} < d \leq Y} = S_1 + S_2, \tag{3.22}$$

say. Clearly we have

$$S_1 \ll \sqrt{Y} \cdot Y^{\varepsilon}, \tag{3.23}$$

where $\varepsilon > 0$ can be taken arbitrarily small. On the other hand, in light of (3.21), we have

$$S_2 \leq \frac{2}{\log Y} \cdot cY (\log Y)^{3h} \ll Y (\log Y)^{3h-1}. \tag{3.24}$$

Setting $Y = z^3$ and using (3.23) and (3.24) in (3.22) proves (3.19).

Coming back to our choice of z and to the size of S given by (3.18), we have

$$\frac{\log z}{(\log x)^{\varepsilon_x}} = x_1^{\delta_x - \varepsilon_x}, \qquad \frac{\log z}{S} \approx x_1^{\delta_x - \varepsilon_x}.$$

Therefore, by choosing $\delta_x = 2\varepsilon_x$, we obtain

$$H \leq C \exp\left\{-\frac{1}{2}(\delta_x - \varepsilon_x) x_2 \cdot x_1^{\delta_x - \varepsilon_x})\right\} = C \exp\{-\frac{1}{2}\varepsilon_x \cdot x_2 \cdot x_1^{\varepsilon_x}\}. \tag{3.25}$$

Moreover,

$$\begin{aligned}
\prod_{p|Q}\left(1-\frac{\eta(p)}{p}\right) &= \prod_{\substack{p\in\wp_2\\(p,M)=1}}\left(1-\frac{D\rho_F(p)}{p}\right)\prod_{p|M}\left(1-\frac{1}{p}\right)\\
&= \frac{\phi(M)}{M}K(M)\prod_{p\in\wp_2}\left(1-\frac{D\rho_F(p)}{p}\right). \qquad (3.26)
\end{aligned}$$

Using (3.19), (3.25) and (3.26) in (3.17), and then using this in (3.14), we obtain that inequality (3.12) follows immediately, thus completing the proof of Lemma 3.6.

□

Proof of Theorem 3.1

Recall that given a word $\beta = b_1b_2\dots b_k \in \mathcal{A}_D^k$, $\nu_\beta(\delta)$ stands for the number of occurrences of β in δ, that is the number of solutions $\tau_1, \tau_2 \in \mathcal{A}_D^*$ such that $\delta = \tau_1\beta\tau_2$. Note that it is clear that

$$\nu_\beta(\gamma_1) + \nu_\beta(\gamma_2) \le \nu_\beta(\gamma_1\gamma_2) \le \nu_\beta(\gamma_1) + \nu_\beta(\gamma_2) + k.$$

Let N be a large integer and let θ_N be the prefix of length N of the infinite sequence $\alpha_1\alpha_2\dots$. Moreover, let x be the largest integer for which

$$\lambda(\alpha_1\dots\alpha_x) \le N < \lambda(\alpha_1\dots\alpha_x\alpha_{x+1}).$$

Since $\lambda(\alpha_{x+1}) \le \omega(U(x+1)) \le c\log x$, we have

$$N + O(\log x) = \sum_{n\le x}\lambda(\alpha_n) = \sum_{n\le x}(\omega(U(n)) + O(1)) = O(x) + hDx\log\log x.$$

We may therefore write

$$x = \frac{N}{hD\log\log N} + O\left(\frac{N}{(\log\log N)^2}\right). \qquad (3.27)$$

Let $\theta_N = \alpha_1\dots\alpha_x$. For each $n\in[1,x]$, let $\alpha_n = \gamma_n\kappa_n\delta_n$, where γ_n is the word composed from $h_n(q)$ where q runs over those prime divisors of $U(n)$ which belong to the set $\wp_1$ and similarly δ_n is composed from those $h_n(q)$ where q runs over the prime divisors of $U(n)$ which belong to $\wp_3$.

We have $\lambda(\gamma_n) \le \omega_1(U(n))$ and $\lambda(\delta_n) \le \omega_3(U(n))$, so that by (3.10) and (3.11), we obtain that

$$\sum_{n\le x}\lambda(\gamma_n) \ll x\sqrt{x_2} \qquad \text{and} \qquad \sum_{n\le x}\lambda(\delta_n) \ll x\sqrt{x_2},$$

thereby implying that

$$\nu_\beta(\theta_N) = \sum_{n=1}^{x}\nu_\beta(\kappa_n) + O(x\sqrt{x_2}). \qquad (3.28)$$

Using estimates (3.3) and (3.4) of Lemma 3.4, it follows from (3.28) that

$$\nu_\beta(\theta_N) = \sum_{\substack{n=1 \\ n\in\mathcal{J}}}^{x} \nu_\beta(\kappa_n) + O(x\sqrt{x_2}), \tag{3.29}$$

where

$$\mathcal{J} := \{n : |\omega(U(n)) - hDx_2| \le cx_2^{3/4}\}.$$

Now, let

$$\mathcal{J}' := \{n \in \mathcal{J} : q^2|U(n) \text{ for } q \in \wp_2\}.$$

We claim that we can drop from the sum in (3.29) those $n \in \mathcal{J}'$, since one can show by Lemma 3.3 that

$$\sum_{\substack{n\le x \\ n\in\mathcal{J}'}} \nu_\beta(\kappa_n) = o(x\log\log x) \qquad (x\to\infty).$$

For the remaining integers $n \le x$, $n \in \mathcal{J}\setminus\mathcal{J}'$, we have

$$B(F_\nu(n)) = m_\nu \qquad (\nu = 0,1,\dots,D-1),$$

with $M = m_0m_1\cdots m_{D-1}$, M squarefree, $|\omega(M) - hD\log\log x| \le Cx_2^{3/4}$. We then have

$$M \le Z_x^{2hDx_2} \le x^{\varepsilon_x},$$

say.

Now, let $M \in \mathcal{N}(\wp_2)$, squarefree, $M \le x^{\varepsilon_x}$, $M = q_1\cdots q_S$ for primes $q_1 < \cdots < q_S$, $|S - hDx_2| \le cx_2^{3/4}$.

With $M = m_0m_1\cdots m_{D-1}$ being any representation, we have by Lemma 3.6,

$$\begin{aligned} T(x|m_0,m_1,\dots,m_{D-1}) &= x\frac{\rho(M)\phi(M)}{M}\prod_{p\in\wp_2}\left(1-\frac{D\rho_F(p)}{p}\right)\cdot K(M) \\ &\quad + O\left(x\frac{\rho(M)}{M}\exp\{-x^{\varepsilon_x}\}\right). \end{aligned}$$

For a fixed M, consider all those $m_0,m_1,\dots,m_{D-1}$ for which $M = m_0m_1\cdots m_{D-1}$. Let $\tau_D(M)$ be the number of solutions of $M = m_0m_1\cdots m_{D-1}$. It is clear that τ_D is a multiplicative function and that $\tau_D(p) = D$. If $m_0,m_1,\dots,m_{D-1}$ run over all the possible choices, then the corresponding β_n's run over all the possible words of length S in $\mathcal{A}_D^k$. Indeed, let $\varepsilon_1\dots\varepsilon_S \in \mathcal{A}_D^S$ and let $m_j = \prod_{\varepsilon_\ell=j} q_\ell$ $(j = 0,1,\dots,S-1)$. We then have

$$\begin{aligned} \nu_\beta(\theta_N) &= x\sum_{\substack{M\le x^{\varepsilon_x} \\ M \text{ squarefree } \in\mathcal{N}(\wp_2) \\ |\omega(M)-hDx_2|\le cx_2^{3/4}}} \frac{\rho(M)\phi(M)}{M^2}K(M)\prod_{p\in\wp_2}\left(1-\frac{D\rho_F(p)}{p}\right)\sum_{\rho\in\mathcal{A}_D^S}\nu_\beta(\rho) \\ &\quad + O\left(\sum_{M\le x^{\varepsilon_x}} x\frac{\rho(M)\omega(M)\tau_D(M)}{M}\exp\{-x_1^{\varepsilon_x}\}\right) + O\left(x\cdot x_2^{3/4}\right). \end{aligned} \tag{3.30}$$

Letting Σ_0 be the first error term above, we have that

$$\Sigma_0 \ll x \exp\{-x_1^{\varepsilon_x}\} x_2 \prod_{p \in \wp_2} \left(1 + \frac{\rho(p)\tau_D(p)}{p}\right) \ll x \exp\{-x_1^{\varepsilon_x}\} x_2 \cdot (\log x)^\kappa \ll x.$$

From this and observing that $\sum_{\rho \in \mathcal{A}_D^S} \nu_\beta(\rho) = (s-k+1)D^{s-k}$, it follows that, given arbitrary distinct words β_1, β_2 belonging to $\mathcal{A}_D^k$,

$$|\nu_{\beta_1}(\theta_N) - \nu_{\beta_2}(\theta_N)| \ll x \cdot x_2^{3/4}.$$

Since

$$\sum_{\beta \in \mathcal{A}_D^k} \nu_\beta(\theta_N) = N + O(\log N)$$

and since by (3.27) we have $x \approx N/(\log\log N)$, it follows that

$$\left| \nu_\beta(\theta_N) - \frac{N}{D^k} \right| \le \frac{1}{D^k} \sum_{\beta_1 \in E_k} |\nu_\beta(\theta_N) - \nu_{\beta_1}(\theta_N)| + O\left(\frac{N}{(\log\log N)^{1/4}} \right),$$

thus establishing that

$$\limsup_{N \to \infty} \frac{\nu_\beta(\theta_N)}{N} = \frac{1}{D^k}$$

and thereby completing the proof of Theorem 3.1.

IV. Some new methods for constructing normal numbers [17]

(Annales des Sciences Mathématiques du Québec, 2012)

First method

Fix an integer $q \ge 2$. Let $\mathcal{B}$ be an infinite set of positive integers and let $B(x) = \#\{b \le x : b \in \mathcal{B}\}$. Further, let $F : \mathcal{B} \to \mathbb{N}$ be a function for which, for some positive integer r and constants $0 < c_1 < c_2 < +\infty$,

$$c_1 \le \frac{F(b)}{b^r} \le c_2 \qquad \text{for all } b \in \mathcal{B}.$$

Let x be a large number and set $N = \left\lfloor \frac{\log x}{\log q} \right\rfloor + 1$.

Let $0 \le \ell_1 < \cdots < \ell_h$ $(\le rN)$ be integers and let $a_1, \ldots, a_h \in \mathcal{A}_q$. Using the notation given in (0.2), we further let

$$\mathcal{B}_F \left(x \,\middle|\, \begin{matrix} \ell_1, \ldots, \ell_h \\ a_1, \ldots, a_h \end{matrix} \right) = \{ b \le x : b \in \mathcal{B},\ \varepsilon_{\ell_j}(F(b)) = a_j, j = 1, \ldots, h \}$$

and

$$B_F\left(x\,\middle|\,\begin{array}{l}\ell_1,\dots,\ell_h\\ a_1,\dots,a_h\end{array}\right) = \#\mathcal{B}_F\left(x\,\middle|\,\begin{array}{l}\ell_1,\dots,\ell_h\\ a_1,\dots,a_h\end{array}\right).$$

We say that $F(\mathcal{B})$ is a *q-ary smooth sequence* if there exists a positive constant $\alpha < 1$ and a function $\varepsilon(x)$, which tends to 0 as x tends to infinity, such that for every fixed integer $h \geq 1$,

$$\sup_{N^\alpha \leq \ell_1 < \cdots < \ell_h \leq rN - N^\alpha} \left| \frac{q^h B_F\left(x\,\middle|\,\begin{array}{l}\ell_1,\dots,\ell_h\\ a_1,\dots,a_h\end{array}\right)}{B(x)} - 1 \right| \leq c(h)\varepsilon(x) \tag{4.1}$$

(where $c(h)$ is a positive constant depending only on h) and also such that $B(x) \gg \dfrac{x}{\log x}$.

Theorem 4.1. *Let $F(\mathcal{B})$ be a q-ary smooth sequence. Let $b_1 < b_2 < b_3 < \cdots$ stand for the list of all elements of $\mathcal{B}$. Let also*

$$\xi_n = \overline{F(b_n)} = \varepsilon_0(F(b_n))\dots\varepsilon_t(F(b_n))$$

and set

$$\eta = 0.\xi_1\xi_2\xi_3\dots.$$

Consider η as the real number whose q-ary expansion is the concatenation of the numbers $\xi_1, \xi_2, \xi_3, \dots$. Then η is a q-normal number.

Theorem 4.2. *Let $n_1 < n_2 < n_3 < \cdots$ be a sequence of integers such that $\#\{j \in \mathbb{N} : n_j \leq x\} > \rho x$ provided $x > x_0$, for some positive constant ρ. Then, using the notation of Theorem 4.1, let*

$$\mu = 0.\xi_{n_1}\xi_{n_2}\xi_{n_3}\dots$$

Then μ is a q-normal number.

Second method

Theorem 4.3. *Let $q \geq 2$ be a fixed integer. Given a positive integer $n = p_1^{e_1}\cdots p_{k+1}^{e_{k+1}}$ with primes $p_1 < \cdots < p_{k+1}$ and positive exponents $e_1, \dots, e_{k+1}$, we introduce the numbers $c_1(n), \dots, c_k(n)$ defined by*

$$c_j(n) := \left\lfloor \frac{q \log p_j}{\log p_{j+1}} \right\rfloor \in \mathcal{A}_q \qquad (j = 1, \dots, k)$$

and consider the arithmetic function $H : \mathbb{N} \to \mathcal{A}_q^*$ defined by

$$H(n) = \begin{cases} c_1(n)\dots c_k(n) & \text{if } \omega(n) \geq 2,\\ \Lambda & \text{if } \omega(n) \leq 1.\end{cases}$$

Then, the number

$$\xi = 0.H(1)H(2)H(3)\dots$$

is a q-normal number.

Proof. As we will see, this theorem is an easy consequence of a variant of the Turán-Kubilius inequality.

Let $b_1, \ldots, b_k$ be fixed digits in $\mathcal{A}_q$. Then, for each sequence of $k+1$ primes $p_1 < \cdots < p_{k+1}$, define the function

$$f(p_1, \ldots, p_{k+1}) = \begin{cases} 1 & \text{if } \left\lfloor \frac{q \log p_j}{\log p_{j+1}} \right\rfloor = b_j \text{ for each } j \in \{1, \ldots, k\}, \\ 0 & \text{otherwise.} \end{cases}$$

From this, we define the arithmetic function F as follows. If $n = q_1^{\alpha_1} \cdots q_\mu^{\alpha_\mu}$, where $q_1 < \cdots < q_\mu$ are prime numbers and $\alpha_1, \ldots, \alpha_\mu \in \mathbb{N}$, let

$$F(n) = F(n|b_1, \ldots, b_k) = \sum_{j=0}^{\mu-k-1} f(q_{j+1}, \ldots, q_{j+k+1}).$$

We will now show that $F(n)$ is close to $\dfrac{1}{q^k}\omega(n)$ for almost all positive integers n.

Let $Y_x = \exp\exp\{\sqrt{\log\log x}\}$ and $Z_x = x/Y_x$ and further set

$$\begin{aligned} F_0(n) &= \sum_{q_{j+1} \le Y_x} f(q_{j+1}, \ldots, q_{j+k+1}), \\ F_1(n) &= \sum_{Y_x < q_{j+1} \le Z_x} f(q_{j+1}, \ldots, q_{j+k+1}), \\ F_2(n) &= \sum_{q_{j+1} > Z_x} f(q_{j+1}, \ldots, q_{j+k+1}), \end{aligned}$$

so that

$$F(n) = F_0(n) + F_1(n) + F_2(n). \tag{4.2}$$

It is clear that

$$F_0(n) \le \omega_{Y_x}(n) := \sum_{\substack{p|n \\ p \le Y_x}} 1$$

and that

$$F_2(n) \le \sum_{\substack{p|n \\ p > Z_x}} 1.$$

Therefore,

$$\sum_{n \le x} F_0(n) \le x \sum_{p \le Y_x} \frac{1}{p} \le cx\sqrt{\log\log x} \tag{4.3}$$

and

$$\sum_{n \le x} F_2(n) \le x \sum_{Z_z < p \le x} \frac{1}{p} \le cx \log\left(\frac{\log x}{\log Z_x}\right) \ll cx \frac{e^{\sqrt{\log\log x}}}{\log x}. \tag{4.4}$$

We now move on to estimate $\sum_{n\le x}(F_1(n)-A(x))^2$ for a suitable expression $A(x)$, which shall later be given explicitly.

We first write this sum as follows:

$$\begin{aligned}\sum_{n\le x}(F_1(n)-A(x))^2 &= \sum_{n\le x}F_1(n)^2-2A(x)\sum_{n\le x}F_1(n)+A(x)^2\lfloor x\rfloor \\ &= S_1(x)-2A(x)S_2(x)+A(x)^2\lfloor x\rfloor,\end{aligned} \tag{4.5}$$

say.

Let $Y_x<p_1<\cdots<p_{k+1}$. We say that $p_1,\ldots,p_{k+1}$ is a *chain of prime divisors* of n, which we note as $p_1\mapsto p_2\mapsto\cdots\mapsto p_{k+1}|n$, if $\gcd\Big(\frac{n}{p_1\cdots p_{k+1}},p\Big)=1$ for all primes p in the interval $[p_1,p_{k+1}]$ with the possible exception of the primes p belonging to the set $\{p_1,\ldots,p_{k+1}\}$.

Observe that the contribution to the sums $S_1(x)$ and $S_2(x)$ of those positive integers $n\le x$ for which $p^2|n$ for some prime p is small, since the contribution of those particular integers $n\le x$ is less than

$$cxk\sum_{p>Y_x}\frac{1}{p^2}\le\frac{cxk}{Y_x}=o(x).$$

Hence we can assume that the sums $S_1(x)$ and $S_2(x)$ run only over squarefree integers n.

We now introduce the function

$$\Gamma(u,v):=\prod_{\substack{p\in\wp\\ u\le p<v}}\left(1-\frac{1}{p}\right)$$

and observe that it follows from Theorem 5.3 of Prachar [56] that

$$\Gamma(u,v)=\frac{\log u}{\log v}\left(1+O\left(\exp\{-\sqrt{\log u}\}\right)\right). \tag{4.6}$$

Now, using Lemma 0.11, one can establish that

$$\begin{aligned}\#\left\{\nu\le\frac{x}{p_1\cdots p_{k+1}}:\gcd\left(\nu,\prod_{p_1\le p\le p_{k+1}}p\right)=1\right\} \\ =\frac{x}{p_1\cdots p_{k+1}}\Gamma(p_1,p_{k+1})\left(1+O\left(\log^{-C}p_1\right)\right),\end{aligned} \tag{4.7}$$

where C is an arbitrary but fixed positive constant.

It follows from (4.7) using (4.6) that

$$\begin{aligned}S_2(x) &= x\sum_{\substack{p_1<\cdots<p_{k+1}\le x\\ Y_x<p_1\le Z_x}}\frac{f(p_1,\ldots,p_{k+1})}{p_1\cdots p_{k+1}}\Gamma(p_1,p_{k+1}) \\ &\quad +O\left(\sum_{\substack{p_1<\cdots<p_{k+1}\le x\\ Y_x<p_1\le Z_x}}\frac{f(p_1,\ldots,p_{k+1})}{p_1\cdots p_{k+1}}\frac{\Gamma(p_1,p_{k+1})}{\log^C p_1}\right)\end{aligned}$$

$$
(4.8)\qquad = x \sum_{\substack{p_1 < \cdots < p_{k+1} \le x \\ Y_x < p_1 < Z_x}} \frac{f(p_1,\ldots,p_{k+1})}{p_1 \cdots p_{k+1}} \frac{\log p_1}{\log p_{k+1}} + O\left(\sum_{\substack{p_1 < \cdots < p_{k+1} \le x \\ Y_x < p_1 < Z_x}} \frac{f(p_1,\ldots,p_{k+1})}{p_1 \cdots p_{k+1}} \frac{1}{\log^C p_1} \right).
$$

In order to estimate the main term on the right hand side of (4.8), we let

$$
L(x) = \sum_{\substack{p_1 < \cdots < p_{k+1} \le x \\ Y_x < p_1 \le Z_x}} \frac{f(p_1,\ldots,p_{k+1})}{p_1 \cdots p_{k+1}} \frac{\log p_1}{\log p_{k+1}}
$$

and we also consider the sum $L_0(x)$, that is essentially the same sum as the sum $L(x)$ but where we drop the condition $Y_x < p_1 \le Z_x$ in the summation.

Note that, in light of (4.3), the error $L_0(x) - L(x)$ satisfies

$$
(4.9)\qquad 0 \le L_0(x) - L(x) \le c \frac{1}{x} \sum_{n \le x} \omega_{Y_x}(n) \ll \sqrt{\log\log x}.
$$

Now, since, for each $j \in \{1, 2, \ldots, k\}$, we have

$$
\sum_{\left\lfloor \frac{q \log p_j}{\log p_{j+1}} \right\rfloor = b_j} \frac{\log p_j}{p_j} = \frac{1}{q} \log p_{j+1} + O(1),
$$

it follows that, after iteration, we have

$$
(4.10)\qquad L_0(x) = \frac{1}{q^k} \sum_{p_{k+1} \le x} \frac{1}{p_{k+1}} + O(1) = \frac{1}{q^k} \log\log x + O(1).
$$

Now, because of (4.9), we have that $L(x) - L_0(x) = o(\log\log x)$, so that it follows from (4.10) that

$$
(4.11)\qquad L(x) = \frac{1}{q^k} \log\log x + O(1).
$$

Substituting (4.11) in (4.8), we get

$$
(4.12)\qquad S_2(x) = \frac{1}{q^k} x \log\log x + O(x).
$$

In order to estimate S_1, we proceed as follows. We have

$$
\begin{aligned}
S_1 &= \sum_{n \le x} F_1(n)^2 \\
(4.13)\qquad &= 2 \sum_{n \le x} \sum_{\substack{p_1 \mapsto \cdots \mapsto p_{k+1} | n \\ q_1 \mapsto \cdots \mapsto q_{k+1} | n \\ p_{k+1} < q_1}} f(p_1,\ldots,p_{k+1}) f(q_1,\ldots,q_{k+1}) + E(x),
\end{aligned}
$$

where the error term $E(x)$ arises from those $k+1$ tuples $\{p_1,\ldots,p_{k+1}\}$ and $\{q_1,\ldots,q_{k+1}\}$ which have common elements. One can see that the sum of $f(p_1,\ldots,p_{k+1})f(q_1,\ldots,q_{k+1})$ on such $k+1$ tuples is less than $k\omega(n)$, implying that

$$E(x) \ll x\log\log x. \tag{4.14}$$

Using the fact that

$$\#\left\{\nu \le \frac{x}{p_1\cdots p_{k+1}q_1\cdots q_{k+1}} : \left(\nu, \prod_{p_1<p<p_{k+1}} p \times \prod_{q_1<p<q_{k+1}} p\right) = 1\right\}$$
$$= x\frac{\Gamma(p_1,p_{k+1})\Gamma(q_1,q_{k+1})}{p_1\cdots p_{k+1}q_1\cdots q_{k+1}}\left(1+O\left(\frac{1}{\log^C p_1}\right)\right),$$

for some positive constant C. It follows from (4.13) and (4.14), while arguing as we did for the estimation of S_2, that

$$S_1(x) = x\left(\sum_{\substack{p_1<\cdots<p_{k+1}\le x\\ Y_x<p_1<Z_x}} \frac{f(p_1,\ldots,p_{k+1})\Gamma(p_1,p_{k+1})}{p_1\cdots p_{k+1}}\right)^2 + O(x\log\log x). \tag{4.15}$$

Hence, in light of (4.12) and (4.15), we get that

$$S_1(x) = x\left(\frac{\log\log x}{q^k}+O(1)\right)^2 = x\left(\frac{\log\log x}{q^k}\right)^2 + O(x\log\log x).$$

Hence, choosing $A(x) = \frac{1}{q^k}\log\log x$, it follows that the left hand side of (4.5) satisfies

$$\sum_{n\le x}\left(F_1(n) - \frac{1}{q^k}\log\log x\right)^2 \ll \frac{1}{q^k}x\log\log x. \tag{4.16}$$

Recall that $F_1(n)$, as well as $F(n)$, depends on $b_1,\ldots,b_k$, while $A(x)$ does not. Hence, setting

$$G(n) = \sum_{\{b_1,\ldots,b_k\}\in\mathcal{A}_q^k} F(n|b_1,\ldots,b_k),$$

a sum containing q^k terms, we get that

$$\sum_{n\le x}\left(F(n|b_1,\ldots,b_k) - \frac{G(n)}{q^k}\right)^2 \ll x\log\log x,$$

so that $\dfrac{G(n)}{q^k}$ does not depend on the choice of $(b_1,\ldots,b_k)\in\mathcal{A}_q^k$.

Now, by using the Cauchy-Schwarz inequality along with (4.3) and (4.4), we obtain, in light of (4.2), that

$$\sum_{n\le x}\left|F(n) - \frac{1}{q^k}x_2\right| \;\le\; \sum_{n\le x}\left|F_1(n)-\frac{1}{q^k}x_2\right| + \sum_{n\le x}|F_0(n)| + \sum_{n\le x}|F_2(n)|$$

$$(4.17) \qquad \leq \sqrt{x}\left(\sum_{n\leq x}\left|F_1(n)-\frac{1}{q^k}x_2\right|^2\right)^{1/2}+O(x\sqrt{\log\log x}).$$

Hence, it follows from (4.16) and (4.17) that

$$(4.18) \qquad \sum_{n\leq x}\left|F(n)-\frac{1}{q^k}x_2\right|\leq Cx\sqrt{\log\log x}.$$

Hence, given any two k-tuples $(b_1,\dots,b_k)$ and $(b'_1,\dots,b'_k)$ both belonging to $\mathcal{A}_q^k$, it follows from (4.18) that

$$\sum_{n\leq x}|F(n|b_1,\dots,b_k)-F(n|b'_1,\dots,b'_k)|\leq 2Cx\sqrt{\log\log x},$$

thus implying that the probability of the occurrence of $b_1,\dots,b_k$ in the chain of prime divisors $p_1\mapsto\cdots\mapsto p_{k+1}|n$ is almost the same (that is, essentially of the same order) as that of the occurrence of $b'_1,\dots,b'_k$ for any $(b'_1,\dots,b'_k)\in\mathcal{A}_q^k$. This final observation proves that ξ is a normal number and thus completes the proof of Theorem 4.3.

□

Final remarks

This last method can easily be applied to prove the following more general theorem.

Theorem 4.4. *Let $R[x]\in\mathbb{Z}[x]$, the leading coefficient of which is positive. Let m_0 be a positive integer such that $R(m)\geq 0$ for all $m\geq m_0$. Moreover, let $H(n)$ be defined as in Theorem 4.3 and set*

$$\xi=0.H(R(m_0))H(R(m_0+1))H(R(m_0+2))\dots$$

Also, let $m_0\leq p_1<p_2<\cdots$ be the sequence of all primes no smaller than m_0 and set

$$\eta=0.H(R(p_1))H(R(p_2))H(R(p_3))\dots$$

Then ξ and η are q-normal numbers.

Even more is true, namely the following.

Theorem 4.5. *Let $(m_0<)n_1<n_2<\cdots$ be a sequence of integers for which $\#\{n_j\leq x\}>\rho x$ provided $x>x_0$, for some positive constant ρ. Then, using the notations of Theorem 4.4, let*

$$\tau=0.H(R(n_1))H(R(n_2))\dots$$

Then τ is a q-normal number.

Moreover, let $(m_0<)\pi_1<\pi_2<\cdots$ be a sequence of primes for which $\#\{\pi_j\leq x\}>\delta\pi(x)$ provided $x>x_0$, for some positive constant δ. Let

$$\kappa=0.H(R(\pi_1))H(R(\pi_2))\dots$$

Then κ is a q-normal number.

V. Construction of normal numbers by classified prime divisors of integers II [18]

(Funct. Approx. Comment. Math., 2013)

In 2011 (see paper I above), we used Theorem A to construct large families of normal numbers, namely by establishing the following result.

Theorem B. *Let $q \geq 2$ be an integer and let $\mathcal{R}, \wp_0, \wp_1, \ldots, \wp_{q-1}$ be a disjoint classification of primes. Assume that, for a certain constant $c_1 \geq 5$,*

$$\pi([u, u+v] \cap \wp_i) = \frac{1}{q}\pi([u, u+v]) + O\left(\frac{u}{\log^{c_1} u}\right) \tag{5.1}$$

uniformly for $2 \leq v \leq u$, $i = 0, 1, \ldots, q-1$, as $u \to \infty$. Furthermore, let $H : \wp \to \mathcal{A}_q^*$ be defined by

$$H(p) = \begin{cases} \Lambda & \text{if } p \in \mathcal{R}, \\ \ell & \text{if } p \in \wp_\ell \text{ for some } \ell \in \mathcal{A}_q \end{cases} \tag{5.2}$$

and further let $T : \mathbb{N} \to \mathcal{A}_q^*$ be defined by $T(1) = \Lambda$ and for $n \geq 2$ by

$$T(n) = T(p_1^{a_1} \cdots p_r^{a_r}) = H(p_1) \ldots H(p_r). \tag{5.3}$$

Then, the number $0.T(1)T(2)T(3)T(4)\ldots$ is a q-normal number.

As one will notice, Theorem B does not use the full power of Theorem A. Indeed, it is clear that condition (5.1) is much more restrictive than condition (0.5) since it does not allow for subsets of primes $\wp_j$ of distinct densities. In this paper, we first weaken condition (5.1) to allow for the construction of even larger families of normal numbers. Then, we extend our method in order to construct normal numbers using the sequence of shifted primes, and thereafter using the sequence n^2+1, $n = 1, 2, \ldots$

Finally, let us mention that throughout this text, unless specified otherwise, the letters p, $p_1, p_2, \ldots$, $q_1, q_2, \ldots$, $\pi_0, \pi_1, \pi_2, \ldots$ will always denote primes.

Main results

Theorem 5.1. *Assume that $\mathcal{R}, \wp_0, \ldots, \wp_{q-1}$ are disjoint sets of primes, whose union is $\wp$, and assume that there exists a positive number $\delta < 1$ and a real number $c_1 \geq 5$ such that*

$$\pi([u, u+v] \cap \wp_i) = \delta\pi([u, u+v]) + O\left(\frac{u}{\log^{c_1} u}\right) \tag{5.4}$$

holds uniformly for $2 \leq v \leq u$, $i = 0, 1, \ldots, q-1$, and similarly

$$\pi([u, u+v] \cap \mathcal{R}) = (1 - q\delta)\pi([u, u+v]) + O\left(\frac{u}{\log^{c_1} u}\right).$$

Let H and T be defined as in (18.24) and (5.3). Then,

$$\xi = 0.T(1)T(2)T(3)\ldots$$

is a q-normal number.

Examples

1. Let $\wp_0 = \{p : p \equiv 1 \pmod 8\}$, $\wp_1 = \{p : p \equiv 7 \pmod 8\}$ and $\mathcal{R} = \{2\} \cup \{p : p \equiv 3, 5 \pmod 8\}$. With H, T and ξ as in the statement of Theorem 5.1, we may conclude that the corresponding number ξ is a binary normal number.

2. Let $P(x) = e_k x^k + \cdots + e_1 x \in \mathbb{R}[x]$ be a polynomial with at least one irrational coefficient. Let I_0 and I_1 be two disjoint intervals in $[0, 1)$ of equal length. Consider the set of primes $\wp_0 = \{p : \{P(p)\} \in I_0\}$, $\wp_1 = \{p : \{P(p)\} \in I_1\}$ and $\mathcal{R} = \wp \setminus (\wp_0 \cup \wp_1)$. (Here, $\{P(p)\}$ stands for the fractional part of $P(p)$.) With H, T and ξ as in Theorem 5.1, we may conclude that ξ is a binary normal number.

3. It is well known that, given a prime $p \equiv 1 \pmod 4$, there exists a prime $\rho \in \mathbb{Z}[i]$ (the set of Gaussian integers) such that $\dfrac{\arg \rho}{\pi/2} \in [0, 1)$ and $p = \rho \cdot \overline{\rho}$. So, let the subsets of primes $\wp_0, \ldots, \wp_{q-1}$ be defined in such a way that $p \in \wp_j$ if the corresponding Gaussian prime ρ satisfies
$$\frac{\arg \rho}{\pi/2} \in \left[\frac{j}{q}, \frac{j+1}{q}\right) \qquad (j = 0, 1, \ldots, q-1)$$
and let $\mathcal{R} = \{2\} \cup \{p : p \equiv 3 \pmod 4\}$. Then, letting H, T and ξ be defined as in Theorem 5.1, we may claim that ξ is a normal number in base q.

Theorem 5.2. *Let $\mathcal{R}, \wp_0, \ldots, \wp_{q-1}$, H and T be as in the statement of Theorem 5.1. Then the number*
$$\eta = 0.T(1)T(2)T(4)T(6)T(10)\ldots T(p-1)\ldots,$$
where p runs through the sequence of primes, is a q-normal number.

Theorem 5.3. *Let $f : \mathbb{N} \to \mathbb{N}$ be defined by $f(n) = n^2 + 1$. Consider the subset of primes $\widetilde{\wp} := \{p \in \wp : p \equiv 1 \pmod 4\}$. Assume that the sets $\wp_0, \wp_1, \ldots, \wp_{q-1} \subseteq \widetilde{\wp}$ satisfy (5.4) and let*
$$\mathcal{R} = \wp \setminus \left(\bigcup_{j=0}^{q-1} \wp_j \right).$$
Let also H and T be defined as in (18.24) and (5.3). Then
$$\tau = 0.T(f(1))T(f(2))T(f(3))T(f(4))\ldots$$
is a q-normal number.

We will only prove Theorem 5.1. To do so, we will need three additional lemmas. But first, we introduce important functions. Let Z_x be a function tending to infinity but with the condition $\frac{\log Z_x}{\log x} \to 0$ as $x \to \infty$. Furthermore, let $K_x \to \infty$ as $x \to \infty$, but also satisfying $\dfrac{K_x \log Z_x}{\log x} \to 0$ as $x \to \infty$.

Let $Q = \displaystyle\prod_{p \le Z_x} p$. Given an integer $m \ge 2$ such that $P(m) \le Z_x$, we set
$$\mathcal{D}(x|m) = \#\{p \le x : p \equiv 1 \pmod m,\ \gcd\left(\frac{p-1}{m}, Q\right) = 1\}.$$

Further set $\nu(Q) = \prod_{\substack{p|Q \\ p>2}} \left(1 - \frac{1}{p-1}\right)$.

We now introduce the strongly multiplicative function $\kappa(n)$ defined on primes p by

$$\kappa(p) = \begin{cases} 1 & \text{if } p = 2, \\ \frac{p-1}{p-2} & \text{if } p > 2. \end{cases} \tag{5.5}$$

Lemma 5.1. *Let* Z_x *and* K_x *be defined by* $\log Z_x = (\log x)/x_2^2$ *and* $K_x = Bx_2$*, where* B *is a large constant. Then, given any arbitrarily large constant* C*,*

$$\sum_{\substack{m \le Z_x^{K_x} \\ P(m) \le Z_x}} \left| \mathcal{D}(x|m) - \frac{\nu(Q)\kappa(m)}{m} li(x) \right| \ll \frac{x}{\log^C x}.$$

Proof. For now, we fix an integer $m \le Z_x^{K_x}$ such that $P(m) \le Z_x$. We plan to use Lemma 0.11. For this, we set $r = \pi(Z_x)$ and we let $q_1 < \cdots < q_T$ be the sequence of those primes $q_j \le x$ satisfying $q_j - 1 \equiv 0 \pmod{m}$ for $j = 1, \ldots, T$ (so that $T = \pi(x; m, 1)$); and also we let $a_n = (q_n - 1)/m$ for $n = 1, 2, \ldots, T$ and set $f(n) = 1$. Now, define $R(m, d)$ implicitly by

$$\pi(x; dm, 1) = \sum_{\substack{p \le x \\ \frac{p-1}{m} \equiv 0 \pmod{d}}} 1 = \eta(d)\pi(x; m, 1) + R(m, d), \tag{5.6}$$

where $\eta(d)$ is the strongly multiplicative function defined on primes p by

$$\eta(p) = \begin{cases} \frac{1}{p} & \text{if } p|m, \\ \frac{1}{p-1} & \text{if } (p, m) = 1. \end{cases}$$

Hence, as a consequence of Lemma 0.11, we obtain

$$\mathcal{D}(x|m) = \{1 + 2\theta_1 H\}\pi(x; m, 1) \prod_{p|Q} (1 - \eta(p)) + 2\theta_2 \sum_{\substack{d|Q \\ d \le z^3}} 3^{\omega(d)} |R(m, d)|. \tag{5.7}$$

Now, since

$$S = \sum_{\substack{p|Q \\ p>2}} \frac{\log p}{p-2} = (1 + o(1)) \log Z_x \qquad (x \to \infty)$$

and

$$r = \pi(Z_x) \qquad \text{and} \qquad \log r = \log Z_x + O(\log\log x),$$

and since

$$\log z = K_x \log Z_x, \qquad \frac{\log z}{\log r} \sim K_x, \qquad \log\left(\frac{\log z}{S}\right) = \log K_x \qquad (x \to \infty),$$

we have, for x large,

$$H = \exp\{-K_x(\log K_x - \log\log K_x - z/K_x)\} \le \exp\left\{-\frac{K_x}{2}\log K_x\right\}.$$

Hence, it follows from (5.7) that

$$\left|\mathcal{D}(x|m) - \pi(x;m,1)\frac{\phi(m)}{m}\kappa(m)\nu(Q)\right| \tag{5.8}$$
$$\leq 2H\pi(x;m,1)\nu(Q)\kappa(m) + 2\sum_{\substack{d|Q \\ d\leq z^3}} 3^{\omega(d)}|R(m,d)|,$$

where $R(m,d)$ satisfies, in light of (5.6),

$$|R(m,d)| \leq E(dm) + \frac{E(m)}{\phi(d)}, \tag{5.9}$$

where

$$E(r) := \left|\pi(x;r,1) - \frac{\mathrm{li}(x)}{\phi(r)}\right|.$$

Using (5.9), we have that

$$\begin{aligned}
\sum_{\substack{d|Q \\ d\leq z^3}} 3^{\omega(d)}|R(m,d)| &\leq \sum_{\substack{d|Q \\ d\leq z^3}} 3^{\omega(d)}\left(E(dm) + \frac{E(m)}{\phi(d)}\right) \\
&= \sum_{\substack{d|Q \\ d\leq z^3}} 3^{\omega(d)}E(dm) + \sum_{\substack{d|Q \\ d\leq z^3}} 3^{\omega(d)}\frac{E(m)}{\phi(d)} \\
&= \Sigma_1 + \Sigma_2, \qquad (5.10)
\end{aligned}$$

say. Now, on the one hand,

$$\Sigma_1 = \sum_{k\leq z^4} E(k)\prod_{p|k}(1+3) = \sum_{k\leq z^4} E(k)2^{2\omega(k)}. \tag{5.11}$$

On the other hand, we have

$$\Sigma_2 \leq E(m)\sum_{d|Q}\frac{3^{\omega(d)}}{\phi(d)} \leq E(m)\prod_{p|Q}\left(1+\frac{3}{p-1}\right) \leq cE(m)(\log Z_x)^3. \tag{5.12}$$

Thus, using (5.11) and (5.12) in (5.10), we obtain that

$$\begin{aligned}
\sum_{\substack{d|Q \\ d\leq z^3}} 3^{\omega(d)}|R(m,d)| &\leq c(\log Z_x)^3E(m) + \sum_{k\leq z^4} E(k)2^{2\omega(k)} \\
&= T_1 + T_2, \qquad (5.13)
\end{aligned}$$

say. Now, because of Lemma 0.2, we have that, given any fixed constant C,

$$T_1 \ll \frac{x}{\log^C x}. \tag{5.14}$$

On the other hand, observe that since $a \le b + \frac{1}{b}a^2$ for all $a, b \in \mathbb{R}^+$, we have

$$T_2 \le 2^{2Bx_2} \sum_{k \le z^4} E(k) + 2^{-2Bx_2} \sum_{k \le z^4} E(k) 2^{4\omega(k)} = U_1 + U_2, \tag{5.15}$$

say. Using Lemmas 0.1 and 0.2 in order to estimate U_1 and U_2, respectively, it follows that (5.15) can be replaced by

$$T_2 \le \frac{x}{(\log x)^{(A')(2B\log 2)}} + \frac{x}{\log x}(\log x)^{-2B\log 2} \sum_{k \le z^4} \frac{2^{4\omega(k)}}{\phi(k)}, \tag{5.16}$$

where B and A' are arbitrary positive constants. Hence, by an appropriate choice of B and A', it follows from (5.16) that

$$T_2 \ll \frac{x}{\log^C x}. \tag{5.17}$$

Then, using (5.14) and (5.17) in (5.13), placing the result in (5.8) and then summing the first term on the right hand side of (5.8) over m, we obtain from Lemma 0.1 that it is $\ll x/(\log^C x)$, thus completing the proof of Lemma 5.1. □

Lemma 5.2. *Given positive integers k and A, set*

$$B_k(x, A) = \sum_{\substack{m_1 \le Z_x^{K_x} \\ \omega(m_1) = k \\ p(m_1) > w_x,\ P(m_1) \le Z_x}} \mathcal{D}(x|Am_1).$$

Let $\wp_0, \ldots, \wp_{q-1}$ be a disjoint classification of primes with corresponding densities $\delta_0, \ldots, \delta_{q-1}$. Then, given an arbitrary constant $C > 0$,

$$\sum_{\substack{A \le w_x^{w_x} \\ P(A) \le w_x}} \sum_{k \le Bx_2} \sum_{i_1 \ldots i_k \in \mathcal{A}_q^k} \left| \sum_{\substack{m_1 \le Z_x^{K_x} \\ H(m_1) = i_1 \ldots i_k \\ p(m_1) > w_x,\ P(m_1) \le Z_x}} \mathcal{D}(x|Am_1) - \delta_{i_1} \cdots \delta_{i_k} B_k(x, A) \right| \ll \frac{x}{\log^C x}.$$

Moreover,

$$\sum_{\substack{A \le w_x^{w_x} \\ P(A) \le w_x}} \sum_{k \le Bx_2} \left| B_k(x, A) \quad - \quad \nu(Q)\, li(x) \frac{\kappa(A)}{A} \sum_{\substack{m_1 \le Z_x^{K_x} \\ \omega(m_1) = k \\ p(m_1) > w_x,\ P(m_1) \le Z_x}} \frac{\kappa(m_1)}{m_1} \right| \ll \frac{x}{\log^C x}.$$

Proof. The result is a direct consequence of Theorem A and Lemma 5.1. □

Recall that $\nu_\beta(\alpha)$ stands for the number of occurrences of β as a subword of the word α. In other words,

$$\nu_\beta(\alpha) = \#\{(\gamma_1, \gamma_2) : \alpha = \gamma_1 \beta \gamma_2, \text{ where } \gamma_1, \gamma_2 \in \mathcal{A}_q^*\}.$$

We then have the following.

Lemma 5.3. *Given positive integers $h \geq 2k$,*

$$\sum_{\alpha \in \mathcal{A}_q^h} \left(\nu_\beta(\alpha) - \frac{h}{q^k} \right)^2 \leq c \frac{hkq^h}{q^k}, \tag{5.18}$$

where c is some absolute constant.

Proof. On the one hand, we have

$$\Sigma_1 := \sum_{\alpha \in \mathcal{A}_q^h} \nu_\beta(\alpha) = \sum_{\ell=0}^{h-k} q^\ell q^{h-\ell-k} = q^{h-k}(h-k+1), \tag{5.19}$$

while on the other hand

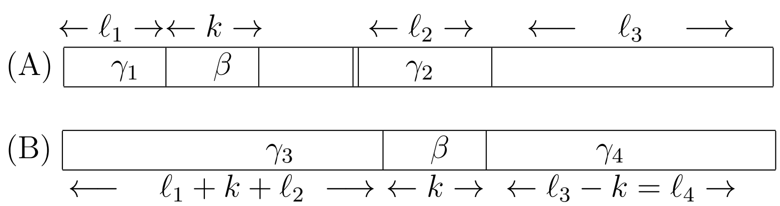
$$\Sigma_2 := \sum_{\alpha \in \mathcal{A}_q^h} \nu_\beta^2(\alpha) = \#\{(\gamma_1, \gamma_2, \gamma_3, \gamma_4) : \alpha = \gamma_1 \beta \gamma_2 = \gamma_3 \beta \gamma_4\}. \tag{5.20}$$

Now, write

$$\Sigma_2 = \Sigma_{2,0} + \Sigma_{2,1} + \Sigma_{2,2},$$

where in $\Sigma_{2,0}$, we impose the condition $\lambda(\gamma_1) = \lambda(\gamma_3)$, in $\Sigma_{2,1}$, we impose the condition $\lambda(\gamma_1) > \lambda(\gamma_3)$, and finally in $\Sigma_{2,2}$, we are restricted to $\lambda(\gamma_1) < \lambda(\gamma_3)$. In $\Sigma_{2,0}$, we have $\gamma_1 = \gamma_3$, so that $\Sigma_{2,0} = \Sigma_1$.

Let $\Sigma_{2,1,1}$ be the number of those γ_1, γ_3 for which $\lambda(\gamma_3) \leq \lambda(\gamma_1) + k$, and $\Sigma_{2,1,2}$ be the number of those γ_1, γ_3 for which $\lambda(\gamma_3) > \lambda(\gamma_1) + k$. Since γ_3 is a prefix of $\gamma_1 \beta$, it follows that it has no more than k distinct values for a fixed γ_1, and therefore that $\Sigma_{2,1,1} \leq k\Sigma_1$. Assume now that $\lambda(\gamma_3) > \lambda(\gamma_1) + k$. Thus we have the following scheme:



Let us fix the position of β in (A) and in (B), that is the lengths ℓ_1 and ℓ_2. Then $\ell_1+\ell_2+\ell_4$ digits can be distributed freely, which yields $q^{\ell_1+\ell_2+\ell_4} = q^{h-2k}$ integers. Hence the number of those nonnegative integers ℓ_1, ℓ_2, ℓ_4 for which $\ell_1 + \ell_2 + \ell_4 = h - 2k$ is equal to

$$\sum_{\ell_4=0}^{h-2k} (h - 2k - \ell_4 + 1) = \sum_{\nu=1}^{h-2k} \nu = \frac{(h-2k)(h-2k+1)}{2}.$$

Thus

$$\Sigma_{2,1,2} = \frac{(h-2k)(h-2k+1)}{2q^{2k}} q^h = \frac{h^2 q^h}{2q^{2k}} + O\left(\frac{khq^h}{q^{2k}} \right),$$

so that (5.20) can be written as

$$\Sigma_2 = \frac{h^2 q^h}{q^{2k}} + O\left(\frac{khq^h}{q^{2k}} \right), \tag{5.21}$$

Therefore, combining (5.19) and (5.21), inequality (5.18) follows, thus completing the proof of Lemma 5.3. □

Proof of Theorem 5.1

Let $\wp^* = \bigcup_{j=0}^{q-1} \wp_j$ and define

$$\omega_{\wp^*}(n) := \sum_{\substack{p|n \\ p\in\wp^*}} 1.$$

For each real number $u \geq 2$, let us set

$$\rho_u := T([u]+1)\dots T([2u]).$$

It is clear that

$$\lambda(\rho_u) = u \sum_{\substack{p\leq 2u \\ p\in\wp^*}} \frac{1}{p} + O(u) = q\delta u \log\log u + O(u). \tag{5.22}$$

Now let k be a fixed positive integer and consider the word $\beta = i_1 \dots i_k \in \mathcal{A}_q^k$. We shall prove that

$$\max_{\beta\in\mathcal{A}_q^k} \left| \nu_\beta(\rho_u) - \frac{\lambda(\rho_u)}{q^k} \right| \leq \varepsilon(u)\lambda(\rho_u), \tag{5.23}$$

where $\varepsilon(u)$ tends to 0 monotonically as $u \to \infty$.

Once we will have proven (5.23), Theorem 5.1 will follow. Indeed, let ξ_N stand for the q-ary expansion of ξ up to the N-th digit. Now, given N, let u be a real number which satisfies the inequalities

$$N_1 := \sum_{j\leq 2u} \omega_{\wp^*}(j) \leq N < \sum_{j\leq 2u+1} \omega_{\wp^*}(j).$$

Let us further set $\xi_{N_1} := T(1)T(2)\dots T([2u])$. With this definition, we have that

$$0 \leq \lambda(\xi_N) - \lambda(\xi_{N_1}) = O(\log N). \tag{5.24}$$

Now, given an arbitrary positive integer ℓ satisfying $2^\ell < u$, let us write

$$\xi_{N_1} = \chi^{(\ell)} \ \rho_{u/2^\ell} \ \rho_{u/2^{\ell-1}} \ \dots \ \rho_u,$$

where

$$\rho_v := T([v]+1)\dots T([2v]).$$

It follows that

$$\nu_\beta(\xi_{N_1}) = \nu_\beta(\chi^{(\ell)}) + \nu_\beta(\rho_{u/2^\ell}) + \dots + \nu_\beta(\rho_u) + O(\ell+1).$$

Hence, using (5.23) and (5.24), we obtain that

$$\nu_\beta(\xi_N) = \nu_\beta(\xi_{N_1}) + O(\log N) = \frac{\lambda(\xi_N)}{q^k} + O\left(\varepsilon(u/2^\ell)N + \lambda(\chi^{(\ell)})\right). \tag{5.25}$$

Now, choosing ℓ to be the unique integer satisfying $2^\ell \le \sqrt{u} < 2^{\ell+1}$ and using the fact that $\lambda(\chi^{(\ell)})/N \to 0$ as $N \to \infty$, we then obtain from (5.25) that

$$\frac{\nu_\beta(\xi_N)}{N} \to \frac{1}{q^k} \qquad \text{as } N \to \infty, \tag{5.26}$$

thus proving that ξ is a q-normal number.

Thus, it remains to prove (5.23). To do that, we will make repetitive use of (5.22). First we set $w_u = \log\log\log u$ and $Z_u = \exp\{(\log u)^{1-\varepsilon_u}\}$, where $\varepsilon_u \to 0$ as $u \to \infty$, and write each integer $n \ge 2$ as

$$n = \prod_{\substack{p^a \| n \\ p \le w_u}} p^a \cdot \prod_{\substack{p^a \| n \\ w_u < p \le Z_u}} p^a \cdot \prod_{\substack{p^a \| n \\ p > Z_u}} p^a = A(n) \cdot B(n) \cdot C(n),$$

say. Since

$$\sum_{u \le n \le 2u} \omega(A(n)) + \sum_{u \le n \le 2u} \omega(C(n)) = o(u \log\log u) \qquad (u \to \infty),$$

it follows that

$$\nu_\beta(\rho_u) = \sum_{u \le n \le 2u} \nu_\beta(T(B(n))) + o(u \log\log u) \qquad (u \to \infty). \tag{5.27}$$

Let $\mathcal{M}_u$ be the set of those positive integers m for which there exists at least one integer $n \in [u, 2u]$ such that $B(n) = m$, in which case we let

$$D(m) = \#\{n \in [u, 2u] : B(n) = m\}.$$

Then, from (5.27), we have

$$\nu_\beta(\rho_u) = \sum_{m \in \mathcal{M}_u} \nu_\beta(T(m)) D(m) + o(u \log\log u) \qquad (u \to \infty). \tag{5.28}$$

Further define $\mathcal{M}_u^{(1)}$ as the set of those $m \in \mathcal{M}_u$ for which at least one of the following conditions holds:

(1) m is not squarefree,

(2) $m \ge Z_u^{K_u}$, $K_u = (\log u)^{\varepsilon_u/2}$,

(3) there exist $p_1 | m$ and $p_2 | m$ such that $p_1 < p_2 < 2p_1$,

(4) $|\omega(m) - \log\log u| > (\log\log u)^{3/4}$.

Let $\mathcal{M}_u^{(0)} = \mathcal{M}_u \setminus \mathcal{M}_u^{(1)}$. Observing that $\nu_\beta(T(m)) \le \omega(m)$, we easily obtain that

$$\sum_{m \in \mathcal{M}_u^{(1)}} \nu_\beta(T(m)) D(m) = o(u \log\log u) \qquad (u \to \infty). \tag{5.29}$$

By a standard sieve argument, we easily get that, as $u \to \infty$,

$$D(m) = (1+o(1))\frac{u}{m}\prod_{w_u \le p \le Z_u}\left(1-\frac{1}{p}\right) = (1+o(1))\frac{u}{m}\frac{\log w_u}{\log Z_u} \qquad (m \in \mathcal{M}_u^{(0)}). \tag{5.30}$$

Thus, using (5.29) and (5.30) in (5.28), we obtain

$$\nu_\beta(\rho_u) = (1+o(1))u\frac{\log w_u}{\log Z_u}\sum_{m\in\mathcal{M}_u^{(0)}}\frac{\nu_\beta(T(m))}{m} + o(u\log\log u) \qquad (u\to\infty).$$

Hence, it remains to prove that, given arbitrary distinct words β_1 and β_2 belonging to $\mathcal{A}_q^k$,

$$\sum_{m\in\mathcal{M}_u^{(0)}}\frac{\nu_{\beta_1}(T(m))}{m} = (1+o(1))\sum_{m\in\mathcal{M}_u^{(0)}}\frac{\nu_{\beta_2}(T(m))}{m} \qquad (u\to\infty). \tag{5.31}$$

We shall now use a technique we have already used to prove Theorem 1 of our 1995 paper [12]. We define the sequence $\ell_0 < \ell_1 < \cdots$ as follows:

$$\ell_0 = w_u, \quad \ell_{j+1} = \ell_j + \frac{\ell_j}{(\log \ell_j)^5} \quad \text{for } j = 0, 1, \ldots.$$

Let r be defined implicitly by $\ell_r \le Z_u < \ell_{r+1}$ and set $I_j = [\ell_j, \ell_{j+1})$ for each integer $j \ge 0$.

Let h be fixed, $|h - \log\log u| \le (\log\log u)^{3/4}$, $0 \le j_1 < j_2 < \cdots < j_h \le r-1$ with $j_{\ell+1} \ge 2j_\ell$. Further define $\mathcal{M}_u^{(0)}(j_1,\ldots,j_h)$ as the set of those $m = \pi_1\pi_2\cdots\pi_h$ for which $\pi_j \in I_{\ell_j}$ for $j = 1,\ldots,h$.

Observe that any $m \in \mathcal{M}_u^{(0)}(j_1,\ldots,j_h)$ satisfies

$$\ell_{j_1+1}\cdot\ell_{j_2+1}\cdots\ell_{j_h+1} \ge m \ge \ell_{j_1}\cdot\ell_{j_2}\cdots\ell_{j_h}$$

and that

$$\begin{aligned} 1 &\le \frac{\ell_{j_1+1}\cdot\ell_{j_2+1}\cdots\ell_{j_h+1}}{\ell_{j_1}\cdot\ell_{j_2}\cdots\ell_{j_h}} \le \prod_{j=1}^{h}\left(1+\frac{1}{(\log \ell_j)^5}\right) \\ &\le \exp\left\{\sum_{j=1}^{h}\frac{1}{(\log\ell_j)^5}\right\} \le \exp\left\{\sum_{j=0}^{h-1}\frac{1}{(\log w_u + j\log 2)^5}\right\} \\ &= 1+o(1) \qquad (u\to\infty). \end{aligned}$$

This means that instead of proving (5.31), we only need to prove

$$\sum_{m\in\mathcal{M}_u^{(0)}(\ell_{j_1},\ldots,\ell_{j_h})}\frac{\nu_{\beta_1}(T(m))}{m} = (1+o(1))\sum_{m\in\mathcal{M}_u^{(0)}(\ell_{j_1},\ldots,\ell_{j_h})}\frac{\nu_{\beta_2}(T(m))}{m} \qquad (u\to\infty). \tag{5.32}$$

Now let $\mathcal{M}_u^{(0)}(\ell_{j_1},\ldots,\ell_{j_h}|\wp_{\nu_1},\ldots,\wp_{\nu_h})$ be the set of those $m = \pi_1\pi_2\cdots\pi_h \in \mathcal{M}_u^{(0)}(\ell_{j_1},\ldots,\ell_{j_h})$ for which $\pi_\ell \in \wp_{\nu_\ell}$.

Then, repeating the computation done in [12], we obtain that

$$\text{(5.33)} \qquad \frac{\#\mathcal{M}_u^{(0)}(\ell_{j_1},\dots,\ell_{j_h}|\wp_{\nu_1},\dots,\wp_{\nu_h})}{\#\mathcal{M}_u^{(0)}(\ell_{j_1},\dots,\ell_{j_h})} = (1+o(1))\tau(\nu_1)\cdots\tau(\nu_h) \qquad (u\to\infty),$$

where $\tau(\nu)=\delta$ if $\nu\in\{0,1,\dots,q-1\}$ and $\tau(q)=1-q\delta$. Assume that among $\nu_1,\dots,\nu_h$, the value q occurs t_1 times. Then, on the right hand side of (5.33), we have $\tau(\nu_1)\cdots\tau(\nu_h)=(1-q\delta)^{t_1}\cdot\delta^{h-t_1}$, which depends only on t_1. It is clear that $\nu_\beta(T(m))$ is constant in every set $\mathcal{M}_u^{(0)}(\ell_{j_1},\dots,\ell_{j_h}|\wp_{\nu_1},\dots,\wp_{\nu_h})$. So, let $e_1<\cdots<e_{t_1}\le h$ be arbitrary integers and consider those $\wp_{\nu_1},\dots,\wp_{\nu_h}$ for which $\nu_{e_j}=q$ for $j=1,\dots,t_1$ and $\nu_\ell\ne q$ if $\ell\ne e_j$. Further let $v_0<v_1<\cdots<v_{h-t_1-1}$ be the sequence of integers defined by

$$\{v_0,\dots,v_{h-t_1-1}\}=\{1,\dots,h\}\setminus\{e_1,\dots,e_{t_1}\}.$$

Moreover, for $j=0,1,\dots,h-t_1-1$, let $\nu_{v_j}\in\{0,1,\dots,q-1\}$ be arbitrary digits. If $m\in\mathcal{M}_u^{(0)}(\ell_{j_1},\dots,\ell_{j_h}|\wp_{\nu_1},\dots,\wp_{\nu_h})$, then

$$\text{(5.34)} \qquad \nu_\beta(T(m))=\nu_\beta(\nu_{v_0}\nu_{v_1}\dots\nu_{v_{h-t_1-1}}).$$

Now, one can easily show that the number of those $n\in[u,2u]$ for which $h-t_1\le k^2$ is $o(u)$. Hence, we may assume that $h-t_1>k^2$. Then, in light of (5.33), (5.34) and Lemma 5.3, we easily obtain (5.32) and thereby (5.23) and (5.26), thus completing the proof of Theorem 5.1.

VI. Construction of normal numbers using the distribution of the k-th largest prime factor [20]

(Bull. Australian Mathematical Society, 2013)

In [13], we showed that if $F\in\mathbb{Z}[x]$ is a polynomial of positive degree with $F(x)>0$ for $x>0$, then the real numbers

$$0.\overline{F(P(2))}\,\overline{F(P(3))}\dots\overline{F(P(n))}\dots$$

and

$$0.\overline{F(P(2+1))}\,\overline{F(P(3+1))}\dots\overline{F(P(p+1))}\dots,$$

where p runs through the sequence of primes, are q-normal numbers.

Here, we prove that the same result holds if $P(n)$ is replaced by $P_k(n)$, the k-largest prime factor of n. The case of $P_k(n)$ relies on the same basic tool we used to study the case of $P(n)$, namely the 1996 result of Bassily and Kátai [2], stated in Lemma 0.5 above. However, the $P_k(n)$ case raises new technical challenges and the proof is not straightforward. Interestingly, the family of normal numbers thus created is much larger. To conclude, we raise an open question.

Main results

Given an integer $k \geq 1$, for each integer $n \geq 2$, we let $P_k(n)$ stand for the k-largest prime factor of n if $\omega(n) \geq k$, while we set $P_k(n) = 1$ if $\omega(n) \leq k-1$. Thus, if $n = p_1^{\alpha_1} p_2^{\alpha_2} \cdots p_s^{\alpha_s}$ stands for the prime factorization of n, where $p_1 < p_2 < \cdots < p_s$, then

$$P_1(n) = P(n) = p_s, \qquad P_2(n) = p_{s-1}, \qquad P_3(n) = p_{s-2}, \ldots$$

Let $F \in \mathbb{Z}[x]$ be a polynomial of positive degree satisfying $F(x) > 0$ for $x > 0$. Also, let $T \in \mathbb{Z}[x]$ be such that $T(x) \to \infty$ as $x \to \infty$ and assume that $\ell_0 = \deg T$. Fix an integer $k \geq \ell_0$. We then have the following results.

Theorem 6.1. *The number*

$$\theta = 0.\overline{F(P_k(T(2)))}\,\overline{F(P_k(T(3)))} \ldots \overline{F(P_k(T(n)))} \ldots$$

is a q-normal number.

Theorem 6.2. *Assuming that $k \geq \ell_0 + 1$, the number*

$$\rho = 0.\overline{F(P_k(T(2+1)))}\,\overline{F(P_k(T(3+1)))} \ldots \overline{F(P_k(T(p+1)))} \ldots$$

is a q-normal number.

The following lemma will come handy in the proofs of our theorems.

Lemma 6.1. *Let $\varepsilon > 0$ be a small number. Given any integer $k \geq \ell_0 + 1$, there exists $x_0 = x_0(\varepsilon)$ such that, for all $x \geq x_0$,*

$$\#\{p \in I_x : P_k(T(p+1)) < x^\varepsilon\} \leq c\varepsilon \frac{x}{\log x}. \tag{6.1}$$

Moreover, for each integer $k \geq \ell_0$, there exists $x_0 = x_0(\varepsilon)$ such that, for all $x \geq x_0$,

$$\#\{n \in I_x : P_k(T(n)) < x^\varepsilon\} \leq c\varepsilon x. \tag{6.2}$$

Proof. For a proof of (6.1) in the case $k = 1$ and $T(n) = n$, see the proof of Theorem 1 in our paper [13]. The more general case $k \geq 2$ and $T \in \mathbb{Z}[x]$ can be handled along the same lines. The estimate (6.2) also follows easily. □

The proof of Theorem 6.1

Let x be a fixed large number. Let $I_x = [x, 2x]$, $N_0 = \lceil x \rceil$, $N_1 = \lfloor 2x \rfloor$ and set

$$\theta^{(x)} := \overline{F(P_k(T(N_0)))}\,\overline{F(P_k(T(N_0+1)))} \ \ldots \ \overline{F(P_k(T(N_1)))}.$$

Given any prime p, we know that

$$\#\{n \in I_x : T(n) \equiv 0 \pmod{p}\} = \frac{\rho(p)}{p} x + O(1), \tag{6.3}$$

where $\rho(p)$ stands for the number of solutions n of the congruence $T(n) \equiv 0 \pmod{p}$.

On the other hand, since we have assumed that $k \geq \ell_0$, there exists a constant $c > 1$ such that $P_k(T(n)) < cx$ for all $n \in I_x$. We then have

$$\#\{n \in I_x : P_k(T(n)) \geq x\} \ll \pi([x, cx]) + x \sum_{x<p<cx} \frac{\rho(p)}{p} = O\left(\frac{x}{\log x}\right) = o(x). \tag{6.4}$$

Finally, given a fixed small positive number $\delta = \delta(k)$, setting

$$\omega_\delta(T(n)) := \sum_{\substack{p \mid T(n) \\ x^\delta < p < x^{1/2}}} 1,$$

one can show, using a type of Turán-Kubilius inequality, that a positive proportion of the integers $n \in I_x$ satisfy the inequality $\omega_\delta(T(n)) \geq k$. It follows from this observation and from (6.4) that

$$\nu_\beta(\theta^{(x)}) = \sum_{n \in I_x} \nu_\beta(\overline{F(P_k(T(n)))}) + O(x) \approx x \log x, \tag{6.5}$$

where the constant implied by the $\approx$ symbol may depend on k as well as on the degrees of T and F.

In order to complete the proof of the theorem it will be sufficient, in light of (6.5), to prove that given any two distinct words $\beta_1, \beta_2 \in \mathcal{A}_q^\ell$, we have

$$\left|\nu_{\beta_1}(\theta^{(x)}) - \nu_{\beta_2}(\theta^{(x)})\right| = o(x \log x) \qquad \text{as } x \to \infty. \tag{6.6}$$

Indeed, since $\mathcal{A}_q^\ell$ contains exactly q^ℓ distinct words and since their respective occurrences are very close in the sense of (6.6), it will follow that

$$\frac{\nu_\beta(\theta^{(x)})}{x \log x} \to \frac{1}{q^\ell} \qquad \text{as } x \to \infty, \tag{6.7}$$

thus establishing that θ is a q-normal number.

In the spirit of Lemma 0.4, we will say that the prime $Q \in I_u$ is a *bad prime* if

$$\max_{\beta \in \mathcal{A}_q^\ell} \left|\nu_\beta(\overline{F(Q)}) - \frac{L(u^r)}{q^\ell}\right| > \kappa_u \sqrt{L(u^r)} \tag{6.8}$$

and a *good prime* if

$$\left|\nu_\beta(\overline{F(Q)}) - \frac{L(u^r)}{q^\ell}\right| \leq \kappa_u \sqrt{L(u^r)}. \tag{6.9}$$

First observe that

$$\left|\nu_{\beta_1}(\theta^{(x)}) - \nu_{\beta_2}(\theta^{(x)})\right| \leq \Sigma_1 + \Sigma_2 + \Sigma_3 + O(x), \tag{6.10}$$

where

- in Σ_1, we sum the expression $m_n := \left|\nu_{\beta_1}(\overline{F(P_k(T(n)))}) - \nu_{\beta_2}(\overline{F(P_k(T(n)))})\right|$ over those integers $n \in I_x$ for which $P_k(T(n)) < x^\varepsilon$;
- in Σ_2, we sum the expression m_n over those integers $n \in I_x$ for which $p = P_k(T(n)) \geq x^\varepsilon$ with p being a good prime;
- in Σ_3, we sum the expression m_n over those integers $n \in I_x$ for which $p = P_k(T(n)) \geq x^\varepsilon$ with p being a bad prime.

It is clear that, in light of estimate (6.2) of Lemma 6.1,

$$\Sigma_1 \leq c\varepsilon x \log x. \tag{6.11}$$

On the other hand, choosing $\kappa_u = \log\log u$ in the range $x^\varepsilon < u < x$,

$$\Sigma_2 \leq cx\sqrt{\log x}\log\log x. \tag{6.12}$$

Finally,

$$\Sigma_3 = \sum_{\substack{n\in I_x \\ p=P_k(T(n))\geq x^\varepsilon \\ p \text{ bad prime}}} m_n \leq c\log x \sum_{\substack{n\in I_x \\ p=P_k(T(n))\geq x^\varepsilon \\ p \text{ bad prime}}} 1 = c\log x\,\Sigma_4, \tag{6.13}$$

say.

Subdivide the interval $[x^\varepsilon, \sqrt{x}]$ into disjoint intervals $[u, 2u)$ as follows. Let j_0 be the smallest positive integer such that $2^{j_0+1}x^\varepsilon \geq \sqrt{x}$, so that

$$[x^\varepsilon, \sqrt{x}] \subset \bigcup_{j=0}^{j_0} J_j,$$

where

$$J_j = [u_j, u_{j+1}) := [2^j x^\varepsilon, 2^{j+1}x^\varepsilon), \qquad j = 0, 1, \ldots, j_0.$$

Using (6.3), we get

$$\begin{aligned}
\Sigma_4 &\leq \sum_{j=0}^{j_0} \sum_{\substack{p\in[u_j,2u_j) \\ p \text{ bad prime}}} \#\{n \in I_x : T(n) \equiv 0 \pmod p\} \\
&\leq cx\sum_{j=0}^{j_0} \sum_{\substack{p\in[u_j,2u_j) \\ p \text{ bad prime}}} \frac{\rho(p)}{p} \\
&\leq cx\sum_{j=0}^{j_0} \frac{1}{(\log\log u_j)^2\log u_j} \\
&\ll \frac{1}{\varepsilon}\frac{x}{(\log\log x)^2}. \qquad (6.14)
\end{aligned}$$

Substituting (6.14) in (6.13), we obtain that

$$\Sigma_3 = O\left(\frac{x\log x}{(\log\log x)^2}\right). \tag{6.15}$$

Thus, gathering estimates (6.11), (6.12) and (6.15) in (6.10), estimate (6.6) follows immediately and therefore (6.7) as well, thereby completing the proof of Theorem 6.1.

The proof of Theorem 6.2

First observe that the additional condition $k \geq \ell_0 + 1$ guarantees that, for $p \leq x$, we have $Q = P_k(T(p+1)) < x^{\ell_0/k}$, with $\ell_0/k < 1$. Hence, it follows from the Brun-Titchmarsh Inequality (Lemma 0.1) that

$$\sum_{\substack{p\in[x,2x]\\ T(p+1)\equiv 0 \pmod Q}} 1 \ll \frac{\rho(Q)x}{\phi(Q)\log(x/Q)} \ll \frac{\rho(Q)}{Q}\frac{x}{\log x}. \tag{6.16}$$

From here on, the proof is somewhat similar to that of Theorem 6.1 but with various adjustments. It goes as follows.

Let

$$\rho^{(x)} := \overline{F(P_k(T(\rho_1+1)))}\dots\overline{F(P_k(T(\rho_S+1)))},$$

where $\rho_1 < \cdots < \rho_S$ is the sequence of primes appearing in the interval I_x.

Observe that, since $S = \pi([x,2x]) \approx \dfrac{x}{\log x}$, we may write

$$\nu_\beta(\rho^{(x)}) = \sum_{i=1}^{S} \nu_\beta(\overline{F(P_k(T(\rho_i+1)))}) + O\left(\frac{x}{\log x}\right) \approx x. \tag{6.17}$$

As in the proof of Theorem 6.1, in order to complete the proof of Theorem 6.2, it will be sufficient, in light of (6.17), to prove that given any two arbitrary distinct words $\beta_1,\beta_2 \in \mathcal{A}_q^\ell$, we have

$$\left|\nu_{\beta_1}(\rho^{(x)}) - \nu_{\beta_2}(\rho^{(x)})\right| = o(x) \qquad \text{as } x\to\infty. \tag{6.18}$$

Indeed, since $\mathcal{A}_q^\ell$ contains exactly q^ℓ distinct words and since their respective occurrences will be proved to be very close in the sense of (6.18), it will follow that

$$\frac{\nu_\beta(\rho^{(x)})}{x} \to \frac{1}{q^\ell} \qquad \text{as } x\to\infty, \tag{6.19}$$

thus establishing that ρ is a q-normal number.

Hence, our main task will be to prove (6.18). To do so, we once more use the concepts of bad prime and good prime defined in (19.17) and (6.9), respectively. We first write

$$\begin{aligned}\left|\nu_{\beta_1}(\rho^{(x)}) - \nu_{\beta_2}(\rho^{(x)})\right| &\leq \sum_{i=1}^{S}\left|\nu_{\beta_1}(\overline{F(P_k(T(\rho_i+1)))}) - \nu_{\beta_2}(\overline{F(P_k(T(\rho_i+1)))})\right| + O(S)\\ &= \Sigma_1+\Sigma_2+\Sigma_3 + O\left(\frac{x}{\log x}\right), \qquad (6.20)\end{aligned}$$

where, letting $m_j := |\nu_{\beta_1}(F(P_k(T(\rho_j+1)))) - \nu_{\beta_2}(F(P_k(T(\rho_j+1))))|$,

- in Σ_1, we sum m_j over those j for which $p = P_k(T(\rho_j + 1)) < x^\varepsilon$,
- in Σ_2, we sum m_j over those j for which $p = P_k(T(\rho_j + 1)) \geq x^\varepsilon$, when p is a good prime,
- in Σ_3, we sum m_j over those j for which $p = P_k(T(\rho_j + 1)) \geq x^\varepsilon$, when p is a bad prime.

Now observe that

$$\nu_\beta(\overline{F(Q)}) \leq cL(u^r) \leq c_1 \log u \qquad \text{for all primes } Q \in I_u. \tag{6.21}$$

Thus, using Lemma 6.1, we have, in light of (6.21), that

$$\Sigma_1 \ll \log x \cdot \frac{\varepsilon x}{\log x} = \varepsilon x. \tag{6.22}$$

Using Lemma 6.1 and estimate (6.21), we also have that

$$\Sigma_2 \leq c\frac{u}{\log u} \cdot \frac{1}{(\log\log u)^2} \cdot \log u = o\left(\frac{x}{\log x} \cdot \log x\right) = o(x). \tag{6.23}$$

Finally, it is clear, using (6.21), that

$$\Sigma_3 = \sum_{\substack{p = P_k(T(\rho_j+1)) \geq x^\varepsilon \\ p \text{ bad prime}}} m_j \leq c \log x \sum_{\substack{p = P_k(T(\rho_j+1)) \geq x^\varepsilon \\ p \text{ bad prime}}} 1 = c \log x \Sigma_4, \tag{6.24}$$

say. Since

$$\Sigma_4 \leq \sum_{j=0}^{j_0} \sum_{\substack{p \in [u_j, 2u_j) \\ p \text{ bad prime}}} \#\{j : T(\rho_j + 1) \equiv 0 \pmod p\},$$

it follows, by (18.23) and by adopting essentially the same approach used to establish (6.14), that

$$\begin{aligned} \Sigma_4 &\leq c \sum_{j=0}^{j_0} \frac{u_j}{\log u_j} \sum_{\substack{p \in [u_j, 2u_j) \\ p \text{ bad prime}}} \frac{\rho(p)}{p} \\ &\leq c \frac{x}{\log x} \sum_{j=0}^{j_0} \frac{1}{(\log\log u_j)^2 \log u_j} \\ &\ll \frac{x}{\log x (\log\log x)^2}. \end{aligned} \tag{6.25}$$

Substituting (6.25) in (6.24), we obtain

$$\Sigma_3 = O\left(\frac{x}{(\log\log x)^2}\right). \tag{6.26}$$

Substituting (6.22), (6.23) and (6.26) in (6.20), we get that, given any two distinct words $\beta_1, \beta_2 \in \mathcal{A}_q^\ell$,

$$\left|\nu_{\beta_1}(\rho^{(x)}) - \nu_{\beta_2}(\rho^{(x)})\right| < \varepsilon x,$$

which proves (6.18) and in consequence (6.19), thus completing the proof of Theorem 6.2.

A related open problem

Let q be a fixed prime number. Let n be a positive integer such that $(n, q) = 1$ and consider its sequence of divisors $1 = d_1 < d_2 < \cdots < d_{\tau(n)} = n$, where $\tau(n)$ stands for the number of divisors of n. Given any positive integer m, we associate to it its congruence class modulo q, thus introducing the function $f_q(m) = \ell$, that is, $m \equiv \ell \pmod q$. Let us now consider the arithmetical function ξ defined by

$$\xi(n) = f_q(d_1)\ldots f_q(d_{\tau(n)}) \in \mathcal{A}_q^{\tau(n)}.$$

Given $\beta \in \mathcal{A}_q^k$ and $\alpha \in \mathcal{A}_q^*$, let $M(\alpha|\beta)$ stand for the number of occurrences of the word β in the word α.

Is it true that the quantity

$$Q_k(n) := \max_{\beta \in \mathcal{A}_q^k} \left| \frac{M(\xi(n)|\beta)(q-1)^k}{\tau(n)} - 1 \right|$$

tends to 0 for almost all positive integers n for which $(n, q) = 1$ and $\tau(n) \to \infty$?

This seems to be a difficult problem. Even proving the particular case $Q_2(n) \to 0$ appears to be quite a challenge. But observe that the case $k = 1$ is easy to establish. Indeed, let χ stand for a Dirichlet character and let

$$S_\chi(n) = \sum_{d|n} \chi(d) = \prod_{p^\alpha \| n} \left(1 + \chi(p) + \cdots + \chi(p^\alpha)\right).$$

Then, letting ϕ stand for the Euler function, we have, letting χ_0 stand for the principal character,

$$\begin{aligned}
\#\{d|n : d \equiv \ell \pmod q\} &= \frac{1}{\phi(q)} \sum_\chi \overline{\chi}(\ell) S_\chi(n) \\
&= \frac{1}{\phi(q)} \overline{\chi_0}(\ell) S_{\chi_0}(n) + \frac{1}{\phi(q)} \sum_{\chi \neq \chi_0} \overline{\chi}(\ell) S_\chi(n) \\
&= \frac{1}{q-1}\tau(n) + \frac{1}{q-1} \sum_{\chi \neq \chi_0} \overline{\chi}(\ell) S_\chi(n) \\
&= \frac{1}{q-1}\tau(n) + o(\tau(n)),
\end{aligned}$$

for almost all n such that $\tau(n) \to \infty$, thus establishing the case $Q_1(n) \to 0$.

VII. Using large prime divisors to construct normal numbers [19]

(Annales Univ. Sci. Budapest, Sect. Comput., 2013)

Let $\eta(x)$ be a slowly increasing function, that is an increasing function satisfying $\lim_{x\to\infty} \dfrac{\eta(cx)}{\eta(x)} = 1$ for any fixed constant $c > 0$. Being slowly increasing, it satisfies in particular the condition $\dfrac{\log \eta(x)}{\log x} \to 0$ as $x \to \infty$.

We then let $Q(n)$ be the smallest prime divisor of n which is larger than $\eta(n)$, while setting $Q(n) = 1$ if $P(n) < \eta(n)$. Then, we show that the real number $0.\overline{Q(1)}\,\overline{Q(2)}\,\overline{Q(3)}\ldots$ is a q-normal number. With various similar constructions, we create large families of normal numbers in any given base $q \geq 2$.

Finally, we consider exponential sums involving the $Q(n)$ function.

Main results

Theorem 7.1. *Given an arbitrary base $q \geq 2$, the number*

$$\xi_1 = 0.\overline{Q(1)}\,\overline{Q(2)}\,\overline{Q(3)}\ \ldots$$

is a q-normal number.

Given an integer $q \geq 2$, let $\mathcal{R}, \wp_0, \wp_1, \ldots, \wp_{q-1}$ be a disjoint set of primes such that, uniformly for $2 \leq v \leq u$ as $u \to \infty$,

$$\pi([u, u+v] \cap \wp_j) = \frac{1}{q}\pi([u, u+v]) + O\left(\frac{u}{\log^5 u}\right) \qquad (j = 0, 1, \ldots, q-1),$$

so that, in particular,

$$\pi([u, u+v] \cap \mathcal{R}) = O\left(\frac{u}{\log^5 u}\right).$$

Then, consider the function κ defined on $\wp$ as follows:

$$\kappa(p) = \begin{cases} \ell & \text{if } p \in \wp_\ell, \\ \Lambda & \text{if } p \in \mathcal{R}. \end{cases}$$

With this notation, we have

Theorem 7.2. *The number*

$$\xi_2 = 0.\kappa(Q(1))\kappa(Q(2))\kappa(Q(3))\ldots$$

is a q-normal number.

Remark 7.1. *In an earlier paper [14], we used such classification of prime numbers to create normal numbers, but by simply concatenating the numbers $\kappa(1)$, $\kappa(2)$, $\kappa(3)$, …*

Let a be a fixed positive integer. Then we have the following result.

Theorem 7.3. *The number*

$$\xi_3 = 0.\kappa(Q(2+a))\kappa(Q(3+a))\kappa(Q(5+a))\ldots\kappa(Q(p+a))\ldots,$$

where p runs through the set of primes, is a q-normal number.

Define $\wp^*$ as the set of all the prime numbers $p \equiv 1 \pmod 4$. Then, let $\mathcal{R}^*, \wp_0^*, \wp_1^*, \ldots, \wp_{q-1}^*$ be disjoint sets of prime numbers such that

$$\wp^* = \mathcal{R}^* \cup \wp_0^* \cup \wp_1^* \cup \cdots \cup \wp_{q-1}^*,$$

and such that, uniformly for $2 \le v \le u$ as $u \to \infty$,

$$\pi([u, u+v] \cap \wp_j^*) = \frac{1}{q}\pi([u, u+v] \cap \wp^*) + O\left(\frac{u}{\log^5 u}\right) \qquad (j = 0, 1, \ldots, q-1),$$

so that, in particular,

$$\pi([u, u+v] \cap \mathcal{R}^*) = O\left(\frac{u}{\log^5 u}\right).$$

Then, consider the function ν defined on primes p as follows

$$\nu(p) = \begin{cases} \ell & \text{if } p \in \wp_\ell^*, \\ \Lambda & \text{if } p \notin \bigcup_{\ell=0}^{q-1} \wp_\ell^*. \end{cases}$$

With this notation, we have the following result.

Theorem 7.4. *The number*

$$\xi_4 = 0.\nu(Q(1))\nu((Q(2))\nu(Q(3))\ldots$$

is a q-normal number.

Now, consider the arithmetic function $f(n) = n^2 + 1$. We then have the following result.

Theorem 7.5. *The two numbers*

$$\begin{aligned} \xi_5 &= 0.\kappa(Q(f(1)))\kappa(Q(f(2)))\kappa(Q(f(3)))\ldots, \\ \xi_6 &= 0.\kappa(Q(f(2)))\kappa(Q(f(3)))\kappa(Q(f(5)))\ldots\kappa(Q(f(p)))\ldots, \end{aligned}$$

where p runs through the set of primes, are q-normal numbers.

Remark 7.2. *One can show that this last result remains true if $f(n)$ is replaced by another non constant irreducible polynomial.*

We now introduce the product function $F(n) = n(n+1)\cdots(n+q-1)$. Observe that if for some positive integer n, we have $p = Q(F(n)) > q$, then $p | n+\ell$ only for one $\ell \in \{0, 1, \ldots, q-1\}$, implying that ℓ is uniquely determined for all positive integers n such that $Q(F(n)) > q$. This allows us to properly define the function

$$\tau(n) = \begin{cases} \ell & \text{if } p = Q(F(n)) > q \text{ and } p | n+\ell, \\ \Lambda & \text{otherwise.} \end{cases}$$

Using this notation, we have the following result.

Theorem 7.6. *The number*

$$\xi_7 = 0.\tau(q+1)\tau(q+2)\tau(q+3)\ldots$$

is a q-normal number.

We now introduce the product function $G(n) = (n+1)(n+2)\cdots(n+q)$ and further define the function

$$\rho(n) = \begin{cases} \ell & \text{if } p = Q(G(n)) > q+1 \text{ and } p|n+\ell+1, \\ \Lambda & \text{otherwise.} \end{cases}$$

Moreover, let $(p_j)_{j\geq 1}$ be the sequence of all primes larger than q, that is, $q < p_1 < p_2 < \cdots$ With this notation, we have the following result.

Theorem 7.7. *The number*

$$\xi_8 = 0.\rho(p_1)\rho(p_2)\rho(p_3)\ldots$$

is a q-normal number.

Let α be an arbitrary irrational number. We will be using the standard notation $e(y) = \exp\{2\pi i y\}$. We then have the following.

Theorem 7.8. *Let*

$$A(x) := \sum_{n\leq x} f(n) e(\alpha Q(n)),$$

where f is any given multiplicative function satisfying $|f(n)| = 1$ for all positive integers n. Then,

$$\lim_{x\to\infty} \frac{A(x)}{x} = 0. \tag{7.1}$$

We will only prove Theorems 7.1 and 7.2. However, we will first prove Theorem 7.2 since its content will be useful for the proof of Theorem 7.1.

Proof of Theorem 7.2

Let $I_x = [x, 2x]$ and first observe that, given any fixed small $\varepsilon > 0$, we may assume that $Q(n) \leq \eta(x)^{1/\varepsilon}$. Indeed,

$$\#\{n \in I_x : Q(n) > \eta(x)^{1/\varepsilon}\} \ll x \prod_{\eta(x) < p \leq \eta(x)^{1/\varepsilon}} \left(1 - \frac{1}{p}\right) \ll \varepsilon x. \tag{7.2}$$

Now let $p_0, p_1, \ldots, p_{k-1}$ be any distinct primes belonging to the interval $(\eta(x), \eta(x)^{1/\varepsilon})$, and let $p_0^* < p_1^* < \cdots < p_{k-1}^*$ be the unique permutation of the primes $p_0, p_1, \ldots, p_{k-1}$, namely the one such that has all its members appear in increasing order, so that we have

$$\eta(x) < p_0^* < p_1^* < \cdots < p_{k-1}^* < \eta(x)^{1/\varepsilon}.$$

Our first goal will be to estimate the size of

$$N(x|p_0, p_1, \ldots, p_{k-1}) := \#\{n \leq x : Q(n+j) = p_j,\ j = 0, 1, \ldots, k-1\}.$$

We must therefore estimate the number of those integers $n \in I_x$ for which $p_j | n+j$ ($j = 0,1,\ldots,k-1$), while at the same time $(\pi_j, n+j) = 1$ if $\eta(x) < \pi_j < p_j$ ($j = 0,1,\ldots,k-1$). Before moving on, let us set

$$Q_k = p_0 p_1 \cdots p_{k-1} \quad \text{and} \quad T_j = \prod_{\eta(x)<\pi<p_j} \pi \qquad (j = 0,1,\ldots,k-1),$$

where this last product runs over primes π. It is then easy to see that, say by using the Eratosthenian sieve (see for instance Chapter 12 in the book of De Koninck and Luca [34]), we have

$$N(x|p_0,p_1,\ldots,p_{k-1}) = (1+o(1))\frac{x}{Q_k}\Sigma_0 \qquad (x \to \infty), \tag{7.3}$$

where

$$\Sigma_0 = \sum_{\substack{\delta_0,\ldots,\delta_{k-1} \\ \delta_j | T_j \ (j=0,1,\ldots,k-1) \\ (\delta_i,\delta_j)=1 \ \text{if} \ i \neq j}} \frac{\mu(\delta_0)\cdots\mu(\delta_{k-1})}{\delta_0 \cdots \delta_{k-1}}$$

(here μ stands for the Möbius function.) One can see that, as $x \to \infty$,

$$\begin{aligned} \Sigma_0 &= \prod_{\eta(x)<\pi<p_0^*}\left(1-\frac{k}{\pi}\right)\cdot\prod_{p_0^*<\pi<p_1^*}\left(1-\frac{k-1}{\pi}\right)\cdots\prod_{p_{k-2}^*<\pi<p_{k-1}^*}\left(1-\frac{1}{\pi}\right) \\ &= (1+o(1))\left(\frac{\log p_0^*}{\log \eta(x)}\right)^{-k}\left(\frac{\log p_1^*}{\log p_0^*}\right)^{-k+1}\cdots\left(\frac{\log p_{k-1}^*}{\log p_{k-2}^*}\right)^{-1}. \end{aligned} \tag{7.4}$$

Hence, if we set $\sigma(p) := \dfrac{\log \eta(x)}{\log p}$, it follows from (7.4) that

$$\Sigma_0 = (1+o(1))\sigma(p_0)\cdots\sigma(p_{k-1}) \qquad (x \to \infty). \tag{7.5}$$

Substituting (7.5) in (7.3), we obtain

$$N(x|p_0,p_1,\ldots,p_{k-1}) = (1+o(1))x\prod_{j=0}^{k-1}\frac{\sigma(p_j)}{p_j} \qquad (x \to \infty), \tag{7.6}$$

an estimate which holds uniformly for $\eta(x) \le p_j \le \eta(x)^{1/\varepsilon}$ ($j = 0,1,\ldots,k-1$).

We will now use a technique which we first used in [12] to study the distribution of subsets of primes in the prime factorization of integers. We first introduce the sequence

$$u_0 = \eta(x), \qquad u_{j+1} = u_j + \frac{u_j}{\log^2 u_j} \quad \text{for each } j = 0,1,2,\ldots$$

and then let T be the unique positive integer satisfying $u_{T-1} < \eta(x)^{1/\varepsilon} \le u_T$. Then, consider the intervals

$$J_0 := [u_0,u_1), \quad J_1 := [u_1,u_2),\ldots, \quad J_{T-1} := [u_{T-1},u_T).$$

Choose k arbitrary integers $j_0, \dots, j_{k-1} \in \{0, 1, \dots, T-1\}$, as well as k arbitrary integers $i_0, \dots, i_{k-1}$ from the set $\{0, 1, \dots, q-1\}$, and consider the quantity

$$M\left(x \,\middle|\, \begin{array}{l} j_0, j_1, \dots, j_{k-1} \\ i_0, i_1, \dots, i_{k-1} \end{array}\right) := \sum_{p_\ell \in J_\ell \cap \wp_{i_\ell}} N(x|p_0, \dots, p_{k-1}). \tag{7.7}$$

Observe that $\dfrac{\sigma(p_h)}{p_h} = (1+o(1))\dfrac{\sigma(u_h)}{u_h}$ as $x \to \infty$ if $p \in J_h$. It follows from this observation and using (7.6) and (7.7) that, as $x \to \infty$,

$$M\left(x \,\middle|\, \begin{array}{l} j_0, j_1, \dots, j_{k-1} \\ i_0, i_1, \dots, i_{k-1} \end{array}\right) = (1+o(1))x \sum_{p_\ell \in J_\ell \cap \wp_{i_\ell}} \prod_{j=0}^{k-1} \frac{\sigma(u_j)}{u_j}. \tag{7.8}$$

Using Theorem 1 of our 1995 paper [12] in combination with (7.8), we obtain that

$$M\left(x \,\middle|\, \begin{array}{l} j_0, j_1, \dots, j_{k-1} \\ i_0, i_1, \dots, i_{k-1} \end{array}\right) = (1+o(1)) M\left(x \,\middle|\, \begin{array}{l} j_0, j_1, \dots, j_{k-1} \\ i'_0, i'_1, \dots, i'_{k-1} \end{array}\right) \qquad (x \to \infty), \tag{7.9}$$

where $(i'_0, i'_1, \dots, i'_k)$ is any arbitrary sequence of length k composed of integers from the set $\{0, 1, \dots, q-1\}$.

Finally, consider the expression

$$A_x := \kappa(Q(\lfloor x \rfloor)) \dots \kappa(Q(\lfloor 2x \rfloor - 1)).$$

It follows from (7.9) that, for any given word $\beta \in \mathcal{A}_q^k$, the number of occurrences of β as a subword in the word A_x is equal to $(1+o(1))\dfrac{x}{q^k}$ as $x \to \infty$, thus completing the proof of Theorem 7.2.

Proof of Theorem 7.1

Let

$$B_x = \overline{Q(\lfloor x \rfloor)} \dots \overline{Q(\lfloor 2x \rfloor - 1)}.$$

Also, let $Q^*(n) = \min\limits_{\substack{p|n \\ p > \eta(x)}} p$ and observe that $Q^*(n) \le Q(n)$, whereas if $Q^*(n) \ne Q(n)$, we have $p|n$ if $\eta(x) < p < \eta(2x)$.

Moreover, let

$$B_x^* = \overline{Q^*(\lfloor x \rfloor)} \dots \overline{Q^*(\lfloor 2x \rfloor - 1)}.$$

Clearly, since $\eta(x)$ was chosen to be a slowly oscillating function, we have

$$0 \le \lambda(B_x) - \lambda(B_x^*) \le cx \sum_{\eta(x) < p < \eta(2x)} \frac{\log p}{\log q} \le c_1 x \log \frac{\eta(2x)}{\eta(x)} = o(x) \qquad (x \to \infty). \tag{7.10}$$

It follows from (7.10) that we now only need to estimate $\lambda(B_x^*)$. To do so, we first let δ_x be a function tending to 0 very slowly as $x \to \infty$, in a manner specified below. If $p < x^{\delta_x}$, we have

$$R_p(x) := \#\{n \in I_x : Q^*(n) = p\} \;=\; (1+o(1))\frac{x}{p} \prod_{\eta(x) < \pi < p} \left(1 - \frac{1}{\pi}\right)$$

$$(7.11) \qquad = (1+o(1))\frac{x}{p}\frac{\log\eta(x)}{\log p} \qquad (x\to\infty),$$

whereas if $x^{\delta_x} \le p \le 2x$, we have

$$(7.12) \qquad R_p(x) < c\frac{x}{p}\frac{\log\eta(x)}{\log p}.$$

Now, observe that, as $x\to\infty$,

$$\begin{aligned}
\lambda(B_x^*) &= \sum_{\eta(x)<p\le 2x} R_p(x)\lambda(p) = \sum_{\eta(x)<p\le 2x} R_p(x)\left\lfloor\frac{\log p}{\log q}\right\rfloor \\
&= (1+o(1))\frac{x}{\log q}\sum_{\eta(x)<p\le 2x}\frac{\log\eta(x)}{p} + O\left(x\log\eta(x)\sum_{x^{\delta_x}<p\le x}\frac{1}{p}\right) \\
(7.13) \qquad &= (1+o(1))x\frac{\log\eta(x)}{\log q}\log\frac{\log x}{\log\eta(x)} + O\left(x\log\eta(x)\log\frac{1}{\delta_x}\right).
\end{aligned}$$

Choosing the function δ_x in such a way that

$$\log\frac{1}{\delta_x} = o\left(\log\frac{\log x}{\log\eta(x)}\right)$$

allows us to replace (7.13) with

$$(7.14) \qquad \lambda(B_x^*) = (1+o(1))x\frac{\log\eta(x)}{\log q}\log\frac{\log x}{\log\eta(x)} \qquad (x\to\infty).$$

Now, pick any two distinct words $\beta_1,\beta_2\in\mathcal{A}_q^k$. First write

$$[\eta(x), x^{\delta_x}] = \bigcup_{j=0}^{T} I_{u_j},$$

where

$$I_{u_j} = [u_j, u_{j+1}), \qquad \text{with } u_0=\eta(x), \qquad u_j = 2^j\eta(x) \qquad \text{for } j=1,2,\dots,T+1,$$

where T is defined as the unique positive integer satisfying $u_T < x^{\delta_x} \le u_{T+1}$.

In the spirit of Lemma 0.4, we will say that the prime $p\in I_u$ is a *bad prime* if

$$\max_{\beta\in\mathcal{A}_q^\ell}\left|\nu_\beta(\overline{p}) - \frac{L(u)}{q^\ell}\right| > \kappa_u\sqrt{L(u)}$$

and a *good prime* if

$$\left|\nu_\beta(\overline{p}) - \frac{L(u)}{q^\ell}\right| \le \kappa_u\sqrt{L(u)}.$$

We will now separate the sum $\sum R_p(x)\lambda(p)$ running over the primes p located in the intervals $[u_j,u_{j+1})$ into two categories, namely the bad primes and the good primes.

First, using (7.11) and (7.12), we have

$$\sum_{\substack{p\in[u_j,u_{j+1})\\ p\,\text{bad}}} R_p(x)\lambda(p) \le c\kappa(u_j) \sum_{p\in[u_j,u_{j+1})} \frac{x\log\eta(x)}{p\log p} \ll x\frac{\log\eta(x)}{\log\eta(x)+j\log 2}. \tag{7.15}$$

On the other hand, if p is a good prime, one can easily establish that the number of occurrences of the words β_1 and β_2 in the word B_x^* are close to each other, in the sense that

$$\nu_{\beta_1}(B_x^*) - \nu_{\beta_2}(B_x^*) = o(\lambda(B_x^*)). \tag{7.16}$$

Hence, proceeding as in [13] (see paper II above – page 16), it follows, considering the true size of $\lambda(B_x^*)$ given by (7.14) and in light of (7.10), (7.15) and (7.16), that the number of words $\beta \in \mathcal{A}_q^k$ appearing in B_x is equal to $(1+o(1))\dfrac{\lambda(B_x)}{q^k}$ as $x \to \infty$.

We then proceed in the same manner in order to obtain similar estimates successively for the intervals $I_{x/2}, I_{x/2^2}, \ldots$ Thus, repeating the argument used in [13], Theorem 7.1 follows immediately.

VIII. Prime-like sequences leading to the construction of normal numbers [21]

(Funct. Approx. Comment. Math., 2013)

Given an integer $q \ge 3$, we consider the sequence of primes reduced modulo q and examine various possibilities for constructing normal numbers using this sequence. We create a sequence of independent random variables that mimics the sequence of primes and then show that for almost all outcomes we obtain a normal number.

Given a fixed integer $q \ge 3$, let

$$f_q(n) = \begin{cases} \Lambda & \text{if } (n,q) \ne 1, \\ \ell & \text{if } n \equiv \ell \pmod q, \quad (\ell,q)=1. \end{cases}$$

Further, letting ϕ stand for the Euler function, let

$$B_{\phi(q)} = \{\ell_1, \ldots, \ell_{\phi(q)}\}$$

be the set of reduced residues modulo q.

Let $\wp$ stand for the set of all primes, writing $p_1 < p_2 < \cdots$ for the sequence of consecutive primes, and consider the infinite word

$$\xi_q = f_q(p_1)f_q(p_2)f_q(p_3)\ldots$$

We first state the following conjecture.

Conjecture 8.1. *The word ξ_q is a normal sequence over $B_{\phi(q)}$ in the sense that given any integer $k \geq 1$ and any word $\beta = r_1 \dots r_k \in B^k_{\phi(q)}$, then, setting*

$$\xi_q^{(N)} = f_q(p_1)f_q(p_2)\dots f_q(p_N) \quad \text{for each} \quad N \in \mathbb{N}$$

and

$$M_N(\xi_q|\beta) := \#\{(\gamma_1, \gamma_2)|\xi_q^{(N)} = \gamma_1\beta\gamma_2\},$$

we have

$$\lim_{N\to\infty} \frac{M_N(\xi_q|\beta)}{N} = \frac{1}{\phi(q)^k}.$$

`Recently added comment:` See comments on Page 71 regarding progress on this conjecture.

Now, with the above notation, consider the following weaker conjecture.

Conjecture 8.2. *For every finite word β, there exists a positive integer N such that $M_N(\xi_q|\beta) > 0$.*

Remark 8.1. *Observe that, in 2000, Shiu [58] provided some hope in the direction of a proof of this last conjecture by proving that given any positive integer k, there exists a string of congruent primes of length k, that is a set of consecutive primes $p_{n+1} < p_{n+2} < \cdots < p_{n+k}$ (where p_i stands for the i-th prime) such that*

$$p_{n+1} \equiv p_{n+2} \equiv \cdots \equiv p_{n+k} \equiv a \pmod q,$$

for some positive integer n, for any given modulus q and positive integer a relatively prime with q.

Let ε_n be a real function which tends monotonically to 0 as $n \to \infty$ but in such a way that $(\log\log n)\varepsilon_n \to \infty$ as $n \to \infty$. Letting $p(n)$ stand for the smallest prime factor of n, consider the set

$$\mathcal{N}^{(\varepsilon_n)} := \{n \in \mathbb{N} : p(n) > n^{\varepsilon_n}\} = \{n_1, n_2, \dots\}. \tag{8.1}$$

We then have the following conjecture.

Conjecture 8.3. *Let $n_1 < n_2 < \cdots$ be the sequence defined in (8.1). Then the infinite word*

$$\xi_q := f_q(n_1)f_q(n_2)\dots$$

is a normal sequence over the set $\{\ell \mod q : (\ell, q) = 1\}$.

Although the problem of generating normal numbers using the sequence of primes does seem inaccessible, we will nevertheless manage to create large families of normal numbers, in the direction of Conjectures 8.1, 8.2 and 8.3, but this time using prime-like sequences.

Main results

Theorem 8.1. *Let $n_1 < n_2 < \cdots$ be the sequence defined in (8.1). Then the infinite word*

$$\eta_q := res_q(n_1) res_q(n_2) \ldots,$$

where $res_q(n) = \ell$ if $n \equiv \ell \pmod q$, contains every finite word whose digits belong to $B_{\phi(q)}$ infinitely often.

Remark 8.2. *It is now convenient to recall a famous conjecture concerning the distribution of primes.*

Let $F_1, \ldots, F_g$ be distinct irreducible polynomials in $\mathbb{Z}[x]$ (with positive leading coefficients) and assume that the product $F := F_1 \cdots F_g$ has no fixed prime divisor. Then the famous Hypothesis H of Schinzel and Sierpinski [57] states that there exist infinitely many integers n such that each $F_i(n)$ ($i = 1, \ldots, g$) is a prime number. The following quantitative form of Hypothesis H was later given by Bateman and Horn ([3],[4]):

> *(*Bateman-Horn Hypothesis*) If $Q(F_1, \ldots, F_g; x)$ stands for the number of positive integers $n \le x$ such that each $F_i(n)$ ($i = 1, \ldots, g$) is a prime number, then*
>
> $$Q(F_1, \ldots, F_g; x) = (1 + o(1)) \frac{C(F_1, \ldots, F_g)}{h_1 \cdots h_g} \frac{x}{\log^g x} \qquad (x \to \infty),$$
>
> *where $h_i = deg\, F_i$ and*
>
> $$C(F_1, \ldots, F_g) = \prod_p \left(\left(1 - \frac{1}{p}\right)^{-g} \left(1 - \frac{\rho(p)}{p}\right) \right),$$
>
> *with $\rho(p)$ denoting the number of solutions of $F_1(n) \cdots F_g(n) \equiv 0 \pmod p$.*

Theorem 8.2. *Let β be an arbitrary word belonging to $B^k_{\phi(q)}$ and let ξ_q be defined as in Conjecture 3. If the Bateman-Horn Hypothesis holds, then*

$$M_N(\xi_q|\beta) \to \infty \qquad \text{as } N \to \infty.$$

Let

$$\lambda_m = \begin{cases} 0 & \text{if } m = 1, 2, \ldots, 10, \\ 1/\log m & \text{if } m \ge 11. \end{cases}$$

Let ξ_m be a sequence of independent random variables defined by $P(\xi_m = 1) = \lambda_m$ and $P(\xi_m = 0) = 1 - \lambda_m$. Let Ω be the set of all possible events ω in this probability space.

Let ω be a particular outcome, say $m_1, m_2, \ldots$, that is one for which $\xi_{m_j} = 1$ for $j = 1, 2, \ldots$ and $\xi_\ell = 0$ if $\ell \notin \{m_1, m_2, \ldots\}$. Now, for a fixed integer $q \ge 3$, set $\text{res}_q(m) = \ell$ if $m \equiv \ell \pmod q$, with $\ell \in \mathcal{A}_q$. Then, let $\eta_q(\omega)$ be the real number whose q-ary expansion is given by

$$\eta_q(\omega) = 0.\text{res}_q(m_1)\text{res}_q(m_2) \ldots$$

We then have the following result.

Theorem 8.3. *The number $\eta_q(\omega)$ is a q-normal number for almost all outcomes ω.*

We only prove Theorems 8.1 and 8.2. Before doing so, we prove three important lemmas.

Lemma 8.1. *Let $q \geq 2$, $k \geq 1$ and $M \geq 1$ be fixed integers. Given any nonnegative integer $n < q^M$, write its q-ary expansion as*

$$n = \sum_{j=0}^{M-1} \varepsilon_j(n) q^j, \qquad \varepsilon_j(n) \in \mathcal{A}_q$$

and, given any word $\alpha = b_1 \dots b_k \in \mathcal{A}_q^k$, set

$$E_\alpha(n) := \#\{j \in \{0, 1, \dots, M-k\} : \varepsilon_j(n) \dots \varepsilon_{j+k-1}(n) = \alpha\}.$$

Then, there exists a constant $c = c(k, q)$ such that

$$\sum_{0 \leq n < q^M} \left(E_\alpha(n) - \frac{M}{q^k} \right)^2 \leq c\, q^M M.$$

Proof. Let

$$f(c_1, \dots, c_k) = \begin{cases} 1 & \text{if } (c_1, \dots, c_k) = (b_1, \dots, b_k), \\ 0 & \text{otherwise.} \end{cases}$$

Then,

$$\Sigma_1 := \sum_{0 \leq n < q^M} E_\alpha(n) = \sum_{0 \leq n < q^M} \sum_{j=0}^{M-k-1} f(\varepsilon_j(n), \dots, \varepsilon_{j+k-1}(n)) = q^{M-k}(M-k).$$

Similarly,

$$\begin{aligned} \Sigma_2 &:= \sum_{0 \leq n < q^M} E_\alpha(n)^2 \\ &= \sum_{0 \leq n < q^M} \sum_{j_1=0}^{M-k-1} \sum_{j_2=0}^{M-k-1} f(\varepsilon_{j_1}(n), \dots, \varepsilon_{j_1+k-1}(n)) \cdot f(\varepsilon_{j_2}(n), \dots, \varepsilon_{j_2+k-1}(n)) \\ &= \sum_{0 \leq n < q^M} \sum_{|j_1 - j_2| \leq k} f(\varepsilon_{j_1}(n), \dots, \varepsilon_{j_1+k-1}(n)) \cdot f(\varepsilon_{j_2}(n), \dots, \varepsilon_{j_2+k-1}(n)) \\ &\quad + \sum_{0 \leq n < q^M} \sum_{|j_1 - j_2| > k} f(\varepsilon_{j_1}(n), \dots, \varepsilon_{j_1+k-1}(n)) \cdot f(\varepsilon_{j_2}(n), \dots, \varepsilon_{j_2+k-1}(n)) \\ &= \Sigma_{2,1} + \Sigma_{2,2}, \end{aligned}$$

say.

On the one hand, it is clear that

$$0 \leq \Sigma_{2,1} \leq (2k+1) q^{M-k}(M-k) \leq c q^M M. \tag{8.2}$$

On the other hand, to estimate $\Sigma_{2,2}$, first observe that for fixed j_1, j_2 with $|j_1 - j_2| > k$, we have to sum 1 over those $n \in [0, q^M - 1[$ for which

$$\varepsilon_{j_1}(n) \dots \varepsilon_{j_1+k-1}(n) = \alpha = \varepsilon_{j_2}(n) \dots \varepsilon_{j_2+k-1}(n).$$

But this occurs exactly for q^{M-2k} many n's. Thus,

$$\Sigma_{2,2} = q^{M-2k} \sum_{\substack{|j_1-j_2|>k \\ 0\le j_1,j_2\le M-k-1}} 1 = q^{M-2k}M^2 + O(q^M M). \tag{8.3}$$

In light of (8.2) and (8.3), it follows that

$$\begin{aligned}
\sum_{0\le n<q^M} \left(E_\alpha(n) - \frac{M}{q^k}\right)^2 &= \Sigma_2 - 2\frac{M}{q^k}\Sigma_1 + \frac{M^2}{q^{2k}}q^M \\
&= q^{M-2k}M^2 + O(q^M M) - 2\frac{M^2}{q^{2k}}q^M + \frac{M^2}{q^{2k}}q^M \\
&= O(q^M M),
\end{aligned}$$

thus completing the proof of the lemma. □

Lemma 8.2. *Given a fixed positive integer R, consider the word $\kappa = c_1 \dots c_R \in \mathcal{A}_q^R$. Fix another word $\alpha = b_1 \dots b_k \in \mathcal{A}_q^k$, with $k \le R$. Let K_1 stand for the number of solutions (γ_1, γ_2) of $\kappa = \gamma_1\alpha\gamma_2$, that is the number of those j's for which $c_{j+1}\dots c_{j+k} = \alpha$. Then, given fixed indices $i_1, \dots, i_H$, let K_2 be the number of solutions of $c_{j+1}\dots c_{j+k} = \alpha$ for which $\{j+1, \dots, j+k\} \cap \{i_1, \dots, i_H\} = \emptyset$ holds. Then,*

$$0 \le K_1 - K_2 \le 2kH.$$

Proof. The proof is obvious. □

Lemma 8.3. *Let $F_1, \dots, F_g$ be distinct irreducible polynomials in $\mathbb{Z}[x]$ (with positive leading coefficients) and set $F := F_1 \cdots F_g$. Let $\rho(p)$ stand for the number of solutions of $F(n) \equiv 0 \pmod p$ and assume that $\rho(p) < p$ for all primes p. Write $p(n)$ for the smallest prime factor of the integer $n \ge 2$ and assume that u and x are real numbers satisfying $u \ge 1$ and $x^{1/u} \ge 2$. Then,*

$$\begin{aligned}
&\#\{n \le x : F_i(n) = q_i \text{ for } i = 1, \dots, k\} \\
&= x \prod_{p<x^{1/u}} \left(1 - \frac{\rho(p)}{p}\right) \\
&\quad \times \left\{1 + O_F(\exp(-u(\log u - \log\log 3u - \log k - 2))) + O_F(\exp(-\sqrt{\log x}))\right\}.
\end{aligned}$$

Proof. This is Theorem 2.6 in the book of Halbertsam and Richert [43]. □

Proof of Theorem 8.1

Theorem 8.1 is essentially a consequence of Lemma 8.3. Indeed, letting $a_1 < \cdots < a_k$ be positive integers coprime to q and considering the product of linear polynomials

$$F(n) := (qn + a_1)\cdots(qn + a_k), \tag{8.4}$$

we have, by Lemma 8.3, that, as $x \to \infty$,

$$\#\{n \in [x, 2x] : p(F(n)) > (2qx + a_k)^{\varepsilon_x}\} = (1 + o(1))x \prod_{p<x^{\varepsilon_x}} \left(1 - \frac{\rho(p)}{p}\right). \tag{8.5}$$

If n is counted in the set on the left hand side of (8.5), we certainly have that $p(qn+a_j) > (qn+a_j)^{\varepsilon_{qn+a_j}}$ for $j=1,\ldots,k$. On the other hand, the desired numbers $qn+a_j$, $j=1,\ldots,k$, are consecutive integers with no small prime factors for all but a negligible number of them. Indeed, if they were not consecutive, then there would be an integer $b \in (a_1, a_k)$ such that $p(qn+b) > x^{\varepsilon_x}$. In this case, set $G_b(n) := qn+b$. Then, by (8.5), we would have

$$\#\{n \in [x,2x] : p(F(n)G_b(n)) > x^{\varepsilon_x}\} = (1+o(1))x \prod_{p<x^{\varepsilon_x}} \left(1 - \frac{\rho_b(p)}{p}\right), \tag{8.6}$$

where $\rho_b(p)$ stands for the number of solutions of $F(n)G_b(n) \equiv 0 \pmod p$. Since $\rho(p)=k$ (recall that each factor on the right hand side of (8.4) is linear) and $\rho_b(p) = k+1$ if $p \nmid q$ and $p > a_k$, it follows that we have the following two "opposite" inequalities:

$$\begin{aligned} \prod_{p<x^{\varepsilon_x}} \left(1 - \frac{\rho(p)}{p}\right) &\geq C(a_1,\ldots,a_k)\,(\varepsilon_x \log x)^{-k}, \\ \prod_{p<x^{\varepsilon_x}} \left(1 - \frac{\rho_b(p)}{p}\right) &\leq C(a_1,\ldots,a_k)\,(\varepsilon_x \log x)^{-k-1}. \end{aligned}$$

Now, for the choice of b, we clearly have $a_k - a_1 + 1 - k$ possible values. We have thus proved that for every large number x, there is at least one $n \in [x,2x]$ for which the numbers $qn+a_1,\ldots,qn+a_k$ are consecutive integers without small prime factors, that is for which $p(qn+a_j) > (qn+a_j)^{\varepsilon_{qn+a_j}}$, thus completing the proof of Theorem 8.1.

Proof of Theorem 8.2

The proof of Theorem 8.2 is almost similar to that of Theorem 8.1. Indeed assume that the Bateman-Horn Hypothesis holds (see Remark 8.2 above). Then, let a_1 be a positive integer such that $a_1 \equiv b_1 \pmod q$ and $a_1 \equiv 0 \pmod D$, where $D = \prod_{\substack{\pi \leq k \\ \pi \nmid q}} \pi$, where π are primes. Similarly, let a_2 be a positive integer such that $a_2 \equiv b_2 \pmod q$ and $a_2 \equiv 0 \pmod D$, with $a_2 > a_1$. Continuing in this manner, that is if $a_1,\ldots,a_{\ell-1}$ have been chosen, we let $a_\ell \equiv b_\ell \pmod q$ with $D | a_\ell$ and $a_\ell > a_{\ell-1}$. Then, applying the Bateman-Horn Hypothesis, we get that if $0 < a_1 < \cdots < a_k$ are k integers satisfying $(a_j,q)=1$ for $j=1,\ldots,k$, then for each positive integer n, setting

$$F(n) = (qn+a_1)\cdots(qn+a_k),$$

letting

$$\rho(m) = \#\{\nu \pmod m : F(\nu) \equiv 0 \pmod m\},$$

so that $\rho(m)=0$ if $(m,q)>1$ and $\rho(p)<p$ for each prime p, and further setting

$$\Pi_x := \prod_{\substack{p \in \wp \\ p \leq \sqrt{qx+a_k}}} p,$$

we have that, as $x \to \infty$, letting μ stand for the Moebius function,

$$\sum_{\substack{n \leq x \\ (F(n),\Pi_x)=1}} 1 = \sum_{n \leq x} \sum_{\delta | (F(n),\Pi_x)} \mu(\delta) = \sum_{\delta | \Pi_x} \mu(\delta) \sum_{\substack{n \leq x \\ F(n) \equiv 0 \pmod \delta}} 1$$

$$
\begin{aligned}
&= (1+o(1))x \sum_{\delta \mid \Pi_x} \frac{\mu(\delta)\rho(\delta)}{\delta} = (1+o(1))x \prod_{p \le \sqrt{qx+a_k}} \left(1 - \frac{\rho(p)}{p}\right) \\
(8.7) \qquad &= (1+o(1))c\frac{x}{\log^k x},
\end{aligned}
$$

where c is a positive constant which depends only on $a_1, \ldots, a_k$.

Now, we can show that almost all prime solutions $\pi_1 < \cdots < \pi_k$ represent a chain of consecutive primes. To see this, assume the contrary, that is that the primes $\pi_1 < \cdots < \pi_k$ are not consecutive, meaning that there exists a prime π satisfying $\pi_1 < \pi < \pi_k$ and $\pi \notin \{\pi_2, \ldots, \pi_{k-1}\}$. Assume that $\pi_\ell < \pi < \pi_{\ell+1}$ for some $\ell \in \{1, \ldots, k-1\}$. We then have

$$
\begin{aligned}
\pi_2 &= \pi_1 + a_2 - a_1, \\
\pi_3 &= \pi_1 + a_3 - a_1, \\
&\vdots \\
\pi_\ell &= \pi_1 + a_\ell - a_1, \\
&\vdots \\
\pi_k &= \pi_1 + a_k - a_1, \\
\pi &= \pi_1 + d, \text{ where } a_\ell - a_1 < d < a_{\ell+1} - a_1.
\end{aligned}
$$

We can now find an upper bound for the number of such $k+1$ tuples. Indeed, by using the Brun-Selberg sieve, one can obtain that the number of such solutions up to x is no larger than $c\dfrac{x}{\log^{k+1} x}$, which in light of (8.7) proves our claim, thus completing the proof of Theorem 8.2.

IX. Normal numbers and the middle prime factor of an integer [24]

(Colloquium Mathematicum, 2014)

Given an integer $n \ge 2$, consider its prime factorisation $n = q_1^{a_1} \cdots q_k^{a_k}$. We let $p_m(n)$ stand for the *middle prime factor* of n, that is,

$$
p_m(n) = \begin{cases} q_1 & \text{if } k = 1, \\ q_{\frac{k+1}{2}} & \text{if } k \text{ is odd}, \\ q_{k/2} & \text{if } k \text{ is even}. \end{cases}
$$

Recently, De Koninck and Luca [34] showed that as $x \to \infty$,

$$
\sum_{n \le x} \frac{1}{p_m(n)} = \frac{x}{\log x} \exp\left((1+o(1))\sqrt{2 \log\log x \, \log\log\log x}\right),
$$

thus answering in part a question raised by Paul Erdős.

Here, we first establish that the size of $\log p_m(n)$ is, for almost all n, close to $\sqrt{\log n}$, and then we show how one can use the middle prime factor of an integer to generate a normal number in any given base $D \geq 2$. Finally, we study the behavior of exponential sums involving the middle prime factor function.

Main results

Theorem 9.1. *Let $g(x)$ be a function which tends to infinity with x but arbitrarily slowly. Set $x_2 = \log\log x$. Then, as $x \to \infty$,*

$$\frac{1}{x}\#\left\{n \in [x, 2x] : e^{-\sqrt{x_2}\,g(x)} \leq \frac{\log p_m(n)}{\sqrt{\log x}} \leq e^{\sqrt{x_2}\,g(x)}\right\} \ \to\ 1, \tag{9.1}$$

$$\frac{1}{x}\#\left\{n \leq x : e^{-\sqrt{x_2}\,g(x)} \leq \frac{\log p_m(n)}{\sqrt{\log x}} \leq e^{\sqrt{x_2}\,g(x)}\right\} \ \to\ 1. \tag{9.2}$$

Analogously, as $x \to \infty$,

$$\frac{1}{x}\#\left\{n \leq x : \left|\log\log p_m(n) - \frac{1}{2}x_2\right| \leq \sqrt{x_2}\,g(x)\right\} \to 1. \tag{9.3}$$

Theorem 9.2. *The sequence Concat$(\overline{p_m(n)} : n \in \mathbb{N})$ is D-normal in every base $D \geq 2$.*

From here on, we will be using the standard notation $e(y) := \exp\{2\pi i y\}$. We now introduce the sum

$$T(x) := \sum_{n \leq x} \log p_m(n).$$

Theorem 9.3. *Consider the real valued polynomial $Q(x) = \alpha_k x^k + \cdots + \alpha_1 x$, where at least one of the coefficients $\alpha_k, \ldots, \alpha_1$ is irrational, and set*

$$E_Q(x) := \sum_{n \leq x} \log p_m(n) \cdot e(Q(p_m(n))).$$

Then,

$$E_Q(x) = o(T(x)) \qquad (x \to \infty).$$

Remark 9.1. *Observe that Theorem 9.3 includes the interesting case $Q(x) = \alpha x$, where α is an arbitrary irrational number.*

Proofs of the theorems

We will first prove the following lemmas.

Lemma 9.1. *Given a positive integer k, let β_1 and β_2 be two distinct words belonging to $\mathcal{A}_D^k$. Let $c_0 > 0$ be an arbitrary number and consider the intervals*

$$J_w := \left[w, w + \frac{w}{\log^{c_0} w}\right] \qquad (w > 1).$$

Further let $\pi(J_w)$ stand for the number of prime numbers belonging to the interval J_w. Then,

$$\frac{1}{\pi(J_w)} \sum_{p \in J_w} \frac{|\nu_{\beta_1}(\overline{p}) - \nu_{\beta_2}(\overline{p})|}{\log p} \to 0 \qquad \text{as } w \to \infty.$$

Proof. This is a reformulation of Lemma 0.5. □

Lemma 9.2. *Let*

$$E_x := \sum_{\substack{n\le x\\ qp_m(n)\mid n\\ \frac{p_m(n)}{3}<q<3p_m(n)}} \log p_m(n).$$

Then, there exists a positive constant c such that

$$E_x \le cx\log\log x.$$

Proof. We have that

$$E_x \le \sum_{p\le x}\log p \sum_{\substack{qpr\le x\\ p/3<q<3p}} 1 \le x\sum_{p\le x}\frac{\log p}{p}\sum_{p/3<q<3p}\frac{1}{q} \le c_1 x\sum_{p\le x}\frac{1}{p} \le c_2 x\log\log x,$$

thus completing the proof of Lemma 9.2. □

Lemma 9.3. *Let $Q(x) = \alpha_k x^k + \cdots + \alpha_1 x$ be a real-valued polynomial such that at least one of its coefficients $\alpha_k, \ldots, \alpha_1$ is irrational. If $p_1 < p_2 < \cdots$ stands for the sequence of primes, then*

$$\sum_{n\le x} e(Q(p_n)) = o(x) \qquad \text{as } x\to\infty.$$

Proof. For a proof of this result, see Chapters 7 and 8 in the book of I.M. Vinogradov [63]. □

Proof of Theorem 9.1

Let

$$y = \exp\{\sqrt{\log x}\}, \quad \text{so that } \log\log y = \frac{1}{2}x_2. \tag{9.4}$$

Then set

$$\omega_y(n) = \sum_{\substack{p\mid n\\ p<y}} 1, \qquad R_y(n) = \sum_{\substack{p\mid n\\ p>y}} 1, \qquad \Delta_y(n) = \omega_y(n) - R_y(n).$$

It is well known that, if $\varepsilon_x \to 0$ arbitrarily slowly as $x\to\infty$, then

$$\frac{1}{x}\#\{n\le x : |\omega(n) - x_2| > \frac{1}{\varepsilon_x}\sqrt{x_2}\} \to 0 \quad \text{as } x\to\infty.$$

On the other hand, from the Turán-Kubilius inequality and in light of our choice of y given by (9.4), we have

$$\sum_{n\le x}\left(\omega_y(n) - \frac{1}{2}x_2\right)^2 = \sum_{n\le x}|\omega_y(n) - \log\log y|^2 = O(xx_2).$$

Secondly,

$$\left|R_y(n) - \frac{1}{2}x_2\right|^2 \le \left(|\omega(n) - x_2| + \left|\omega_y(n) - \frac{1}{2}x_2\right|\right)^2$$

$$(9.5) \qquad \leq \ 2\left((\omega(n)-x_2)^2 + \left(\omega_y(n) - \frac{1}{2}x_2\right)^2 \right),$$

where we used the basic inequality $(a+b)^2 \leq 2(a^2+b^2)$ valid for all real numbers a and b. Then, summing both sides of (9.5) for $n \leq x$, we obtain that for some positive constant C,

$$(9.6) \qquad \sum_{n\leq x} |\Delta_y(n)|^2 \leq \sum_{n\leq x} 2|\omega_y(n) - \frac{1}{2}x_2|^2 + \sum_{n\leq x} 2|R_y(n) - \frac{1}{2}x_2|^2 \leq C\,x\,x_2.$$

It follows from (9.6) that

$$(9.7) \qquad |\Delta_y(n)| \leq \frac{1}{\varepsilon_x}\sqrt{x_2} \qquad \text{for all but at most } o(x) \text{ integers } n \leq x.$$

Let us now choose z and w so that

$$\log z = (\log y) e^{-\sqrt{x_2}\, g(x)}, \qquad \log w = (\log y) e^{\sqrt{x_2}\, g(x)}.$$

Since

$$\sum_{z<p<y} \frac{1}{p} = \log\left(\frac{\log y}{\log z}\right) + o(1) = \sqrt{x_2}\, g(x) + o(1) = A(x) + o(1),$$

say, and similarly,

$$\sum_{y<p<w} \frac{1}{p} = \log\left(\frac{\log w}{\log y}\right) + o(1) = \sqrt{x_2}\, g(x) + o(1) = A(x) + o(1),$$

then setting

$$\omega_{[a,b]}(n) := \sum_{\substack{p|n \\ p\in[a,b]}} 1,$$

we have, again using the Turán-Kubilius inequality, that

$$\sum_{n\leq x} \left(\omega_{[z,y]}(n) - A(x)\right)^2 \leq C\,x\,A(x) \text{ and } \sum_{n\leq x} \left(\omega_{[y,w]}(n) - A(x)\right)^2 \leq C\,x\,A(x).$$

from which it follows that

$$(9.8) \qquad \left|\omega_{[z,y]}(n) - A(x)\right| \leq \frac{1}{\varepsilon_x}\sqrt{A(x)},$$

$$(9.9) \qquad \left|\omega_{[y,w]}(n) - A(x)\right| \leq \frac{1}{\varepsilon_x}\sqrt{A(x)}.$$

Now, recall that from (9.7), we only need to consider those $n \leq x$ for which

$$|\omega_y(n) - R_y(n)| \leq \frac{1}{\varepsilon_x}\sqrt{x_2}$$

and for which (9.8) and (9.9) hold. So, let us choose $\varepsilon_x = 2/g(x)$, in which case we have $A(x) = \sqrt{x_2}\cdot g(x) = (2/\varepsilon_x)\sqrt{x_2}$. Thus, assuming first that $0 \leq R_y(n) - \omega_y(n) < \frac{1}{\varepsilon_x}\sqrt{x_2}$, we have $p_m(n) > y$ and by (9.9), $p_m(n) < w$ provided x is large enough. On the other hand, if $-\frac{1}{\varepsilon_x}\sqrt{x_2} \leq R_y(n) - \omega_y(n) \leq 0$, then we have $p_m(n) \leq y$ and by (9.8), $p_m(n) > z$ provided x is large enough. Hence, in any case, we get

$$z \leq p_m(n) \leq w,$$

which proves (9.2), from which (9.1) and (9.3) follow as well, thus completing the proof of Theorem 9.1.

Proof of Theorem 9.2

Let x be a fixed large number. Let $L_x := \{n \in \mathbb{N} : \lfloor x \rfloor \leq n \leq \lfloor 2x \rfloor - 1\}$ and set

$$\rho_x := \text{Concat}(\overline{p_m(n)} : n \in L_x).$$

It is clear that

$$\lambda(\rho_x) = \sum_{n \in L_x} \lambda(\overline{p_m(n)}), \tag{9.10}$$

$$\nu_\beta(\rho_x) = \sum_{n \in L_x} \nu_\beta(\overline{p_m(n)}) + O(x), \tag{9.11}$$

$$\lambda(\overline{p}) = \frac{\log p}{\log D} + O(1). \tag{9.12}$$

It follows from (9.10), (9.12) and Theorem 9.1 that there exists $c_1 > 0$ such that

$$\lambda(\rho_x) \geq c_1 x \sqrt{\log x} \exp\{-\sqrt{x_2} g(x)\}. \tag{9.13}$$

Given arbitrary distinct words $\beta_1, \beta_2 \in \mathcal{A}_D^k$, we set

$$\Delta(\alpha) := \nu_{\beta_1}(\alpha) - \nu_{\beta_2}(\alpha) \qquad (\alpha \in \mathcal{A}_D^*).$$

Our main task will be to prove that

$$\lim_{x \to \infty} \frac{\Delta(\rho_x)}{\lambda(\rho_x)} = 0. \tag{9.14}$$

This will prove that, for any word $\beta \in \mathcal{A}_D^k$,

$$\frac{\nu_\beta(\rho_x)}{\lambda(\rho_x)} - \frac{1}{D^k} = o(1) \qquad \text{as } x \to \infty \tag{9.15}$$

and therefore that the sequence $\text{Concat}(\overline{p_m(n)} : n \in \mathbb{N})$ is D-normal, thus completing the proof of Theorem 9.2.

To see how (9.15) follows from (9.14), observe that, in light of the fact that, for $k \in \mathbb{N}$ fixed,

$$\sum_{\gamma \in \mathcal{A}_D^k} \nu_\gamma(\rho_x) = \lambda(\rho_x) - k + 1 = \lambda(\rho_x) + O(1), \tag{9.16}$$

we have, as $x \to \infty$,

$$\begin{aligned}
\nu_\beta(\rho_x) - \frac{\lambda(\rho_x)}{D^k} &= \frac{\nu_\beta(\rho_x)D^k - \lambda(\rho_x)}{D^k} \\
&= \frac{\nu_\beta(\rho_x)D^k - \sum_{\gamma\in\mathcal{A}_D^k}\nu_\gamma(\rho_x) + O(1)}{D^k} \\
&= \frac{1}{D^k}\sum_{\gamma\in\mathcal{A}_D^k}(\nu_\beta(\rho_x) - \nu_\gamma(\rho_x)) + O(1) \\
&= \frac{1}{D^k}D^k \cdot o(\lambda(\rho_x)) \\
&= o(\lambda(\rho_x)),
\end{aligned}$$

thus proving (9.15).

Hence, we only need to prove (9.14).

Now, from (9.11), it follows that

$$\Delta(\rho_x) = \sum_{n\in L_x}\Delta(\overline{p_m(n)}) + O(x). \tag{9.17}$$

Let us further introduce the sets

$$\begin{aligned}
L_x^{(0)} &= \left\{n \in L_x : q\,p_m(n) \mid n \ \text{ for some prime } q \in \left(\frac{p_m(n)}{3}, 3p_m(n)\right)\right\}, \\
L_x^{(1)} &= \left\{n \in L_x : \log p_m(n) \le \sqrt{\log x}\exp\{-2\sqrt{x_2}\,g(x)\}\right\}.
\end{aligned}$$

With this notation, we then have, in light of Lemma 9.2 and of (9.13), that

$$\begin{aligned}
\sum_{n\in L_x^{(0)}\cup L_x^{(1)}}\log p_m(n) &\le cx\log\log x + x\sqrt{\log x}\ \exp\{-2\sqrt{x_2}\,g(x)\} \\
&= o\left(x\sqrt{\log x}\exp\{-\sqrt{x_2}\,g(x)\}\right) \\
&= o(\lambda(\rho_x)). \qquad (9.18)
\end{aligned}$$

Hence, setting $L_x^{(2)} = L_x \setminus \left(L_x^{(0)} \cup L_x^{(1)}\right)$, it follows from (9.17) and (9.18) that

$$\Delta(\rho_x) = \sum_{n\in L_x^{(2)}}\Delta(\overline{p_m(n)}) + o(\lambda(\rho_x)). \tag{9.19}$$

Let us now write each integer $n \in L_x^{(2)}$ as $n = a\,p_m(n)\,b$, where

$$P(a) \le p_m(n) \le p(b).$$

Thus setting $M = ab$ and given an arbitrarily small $\varepsilon > 0$, we have from Theorem 9.1 that

$$M \le \frac{2x}{e^{(\log x)^{\frac{1}{2}-\varepsilon}}}. \tag{9.20}$$

Now, let us fix $M = ab$. It is clear that we may ignore those integers $n \le x$ for which $p_m(n)^2 \mid n$ since they are at most $o(x)$ of them anyway. Once this is done, it is clear that in the factorization $n = a p_m(n) b$, we have $P(a) < p(b)$, so that M determines a and b uniquely. Then, in light of (9.20), we may consider the set

$$\mathcal{E}_M := \{n \in L_x^{(2)} : n = a\, p_m(n)\, b = M\, p_m(n)\}.$$

Let $n_1 < n_2 < \cdots < n_H$ be the list of all elements of $\mathcal{E}_M$ and further set $\pi_j = p_m(n_j)$ for $j = 1, 2, \ldots, H$. By construction, it is clear that $\pi_1 < \pi_2 < \cdots < \pi_H$, all consecutive primes, and that, since x/M is large by (9.20), it follows that $\pi_H > (3/2)\pi_1$.

Then, let $\mathcal{K}$ be the set of those M's such that the corresponding set $\mathcal{E}_M$ contains at least one $n \in L_x^{(2)}$, since the others need not be accounted for. Hence, for those $ab = M$, we have that $\mathcal{E}_M$ contains at least $\dfrac{\pi_1}{2\log \pi_1}$ elements, thus implying that $H \ge \dfrac{\pi_1}{2\log \pi_1}$, provided x is chosen to be large enough.

Using Lemma 9.1, it follows that, when $M \in \mathcal{K}$, we have

$$\frac{1}{H}\sum_{j=1}^{H} \frac{\left|\Delta(\overline{p_m(n_j)})\right|}{\log p_m(n_j)} \to 0 \quad \text{as } x \to \infty.$$

From this, it follows that, for $M \in \mathcal{K}$, there exists a function $\varepsilon_x \to 0$ as $x \to \infty$ such that

$$\sum_{M\in\mathcal{K}} \sum_{n\in\mathcal{E}_M} \left|\Delta(\overline{p_m(n)})\right| < \varepsilon_x \sum_{M\in\mathcal{K}} \sum_{n\in\mathcal{E}_M} \lambda(\overline{p_m(n)}). \tag{9.21}$$

Using (9.21), estimate (9.14) follows, thus completing the proof of Theorem 9.2.

Proof of Theorem 9.3

We first write

$$E_Q(2x) - E_Q(x) = \sum_{x\le n\le 2x} \log p_m(n) \cdot e(Q(p_m(n))). \tag{9.22}$$

Using the notation introduced in the proof of Theorem 9.2, we can, in the above sum, drop all those $n \in L_x^{(0)} \cup L_x^{(1)}$. It follows that we only need to consider those $M \in \mathcal{K}$. Now for a fixed $M \in \mathcal{K}$, we only need to examine the sum

$$\sum_{j=1}^{H} \log \pi_j \cdot e(Q(\pi_j)),$$

where $\pi_1, \ldots, \pi_H$ are consecutive primes and $\pi_H > (3/2)\pi_1$. Using Lemma 9.3, we then obtain that

$$\left|\sum_{j=1}^{H} \log \pi_j \cdot e(Q(\pi_j))\right| \le \varepsilon_x \left|\sum_{j=1}^{H} \log \pi_j\right|.$$

Using this in (9.22), it follows that, as $x \to \infty$,

$$|E_Q(2x) - E_Q(x)| \;=\; \left|\sum_{\substack{x\le n\le 2x\\ n\in L_x^{(2)}}} \log p_m(n) \cdot e(Q(p_m(n)))\right| + o(T(x))$$

$$
\begin{aligned}
&\leq \varepsilon_x T(x) + o(T(x)) \\
&= o(T(x)),
\end{aligned}
$$

as requested.

Final remarks

Instead of considering the middle prime factor of an integer, that is the prime factor whose rank amongst the $\omega(n)$ distinct prime factors of an integer n is the $\lfloor \frac{1}{2}\omega(n) \rfloor$-th one, we could have also studied the one whose rank is the $\lfloor \alpha\omega(n) \rfloor$-th one, for any given real number $\alpha \in (0,1)$. In this more general case, say with $p^{(\alpha)}(n)$ in place of $p_m(n)$, the same type of results as above would also hold, meaning in particular that $\log p^{(\alpha)}(n)$ would be close to $\log^\alpha n$ instead of $\sqrt{\log n}$.

X. Constructing normal numbers using residues of selective prime factors of integers [23]

(Annales Univ. Sci. Budapest. Sect. Comp., 2014)

Given an integer $N \geq 1$, for each integer $n \in J_N := [e^N, e^{N+1})$, let $q_N(n)$ be the smallest prime factor of n which is larger than N; if no such prime factor exists, set $q_N(n) = 1$. Fix an integer $Q \geq 3$ and consider the function $f(n) = f_Q(n)$ defined by $f(n) = \ell$ if $n \equiv \ell \pmod{Q}$ with $(\ell, Q) = 1$ and by $f(n) = \Lambda$ otherwise, where Λ stands for the empty word. Then consider the sequence $(\kappa(n))_{n\geq 1} = (\kappa_Q(n))_{n\geq 1}$ defined by $\kappa(n) = f(q_N(n))$ if $n \in J_N$ with $q_N(n) > 1$ and by $\kappa(n) = \Lambda$ if $n \in J_N$ with $q_N(n) = 1$. Then, for each integer $N \geq 1$, consider the concanetation of the numbers $\kappa(1), \kappa(2), \ldots$, that is define $\theta_N := \text{Concat}(\kappa(n) : n \in J_N)$. Then, set $\alpha_Q := \text{Concat}(\theta_N : N = 1, 2, 3, \ldots)$. Finally, let $B_Q = \{\ell_1, \ell_2, \ldots, \ell_{\phi(Q)}\}$ be the set of reduced residues modulo Q, where ϕ stands for the Euler function. We show that α_Q is a normal sequence over B_Q.

In previous papers ([13], [20], [22]), we showed how one could construct normal numbers by concatenating the digits of the numbers $P(2)$, $P(3)$, $P(4), \ldots$, where $P(n)$ stands for the largest prime factor of n, then similarly by using the k-th largest prime factor instead of the largest prime factor and finally by doing the same replacing $P(n)$ by $p(n)$, the smallest prime factor of n.

Here, we consider a different approach which uses the residue modulo an integer $Q \geq 3$ of the smallest element of a particular set of prime factors of an integer n.

Given a fixed integer $Q \geq 3$, let

$$
f_Q(n) := \begin{cases} \Lambda & \text{if} \quad (n, Q) \neq 1, \\ \ell & \text{if} \quad n \equiv \ell \pmod{Q}, \quad (\ell, Q) = 1. \end{cases} \tag{10.1}
$$

Write $p_1 < p_2 < \cdots$ for the sequence of consecutive primes, and consider the infinite word

$$
\xi_Q = f_Q(p_1) f_Q(p_2) f_Q(p_3) \ldots
$$

Let

$$B_Q = \{\ell_1, \ell_2, \dots, \ell_{\phi(Q)}\}$$

be the set of reduced residues modulo Q, where ϕ stands for the Euler totient function.

In an earlier paper [21] (see Conjecture 8.1 on Page 58), we conjectured that the word ξ_Q is a normal sequence over B_Q in the sense that given any integer $k \geq 1$ and any word $\beta = r_1 \dots r_k \in B_Q^k$, and further setting

$$\xi_Q^{(N)} = f_Q(p_1) f_Q(p_2) \dots f_Q(p_N) \quad \text{for each} \quad N \in \mathbb{N}$$

and

$$M_N(\xi_Q|\beta) := \#\{(\gamma_1, \gamma_2) | \xi_Q^{(N)} = \gamma_1 \beta \gamma_2\},$$

we have

$$\lim_{N\to\infty} \frac{M_N(\xi_Q|\beta)}{N} = \frac{1}{\phi(Q)^k}.$$

In this paper, we consider a somewhat similar but more simple problem, namely by using the residue of the smallest prime factor of n (modulo Q) which is larger than a certain quantity, and this time we obtain an effective result.

Our main result

Given an integer $N \geq 1$, for each integer $n \in J_N := [x_N, x_{N+1}) := [e^N, e^{N+1})$, let $q_N(n)$ be the smallest prime factor of n which is larger than N; if no such prime factor exists, set $q_N(n) = 1$. Fix an integer $Q \geq 3$ and consider the function $f(n) = f_Q(n)$ defined by (10.1). Then consider the sequence $(\kappa(n))_{n\geq 1} = (\kappa_Q(n))_{n\geq 1}$ defined by $\kappa(n) = f(q_N(n))$ if $n \in J_N$ with $q_N(n) > 1$ and by $\kappa(n) = \Lambda$ if $n \in J_N$ with $q_N(n) = 1$. Then, for each integer $N \geq 1$, consider the concatenation of $\kappa(1), \kappa(2), \kappa(3), \dots$, that is define

$$\theta_N := \text{Concat}(\kappa(n) : n \in J_N).$$

Then, setting

$$\alpha_Q := \text{Concat}(\theta_N : N = 1, 2, 3, \dots),$$

we will prove the following result.

Theorem 10.1. *The sequence α_Q is a normal sequence over B_Q.*

Proof of the main result

We first introduce the notation $\lambda_N = \log\log N$. Moreover, from here one, the letters p and π, with or without subscript, always stand for primes. Finally, let $\wp$ stand for the set of all primes.

Fix an arbitrary large integer N and consider the interval $J := [x, x+y] \subseteq J_N$. Let $p_1, p_2, \dots, p_k$ be k distinct primes belonging to the interval $(N, N^{\lambda_N}]$. Then, set

$$S_J(p_1, p_2, \dots, p_k) := \#\{n \in J : q_N(n+j) = p_j \text{ for } j = 1, 2, \dots, k\}.$$

We know by the Chinese Remainder Theorem that the system of congruences

$$n + j \equiv 0 \pmod{p_j}, \qquad j = 1, 2, \ldots, k \tag{10.2}$$

has a unique solution $n_0 < p_1 p_2 \cdots p_k$ and that any solution $n \in J$ of (10.2) is of the form

$$n = n_0 + m p_1 p_2 \cdots p_k \quad \text{for some non negative integer } m.$$

Let us now reorder the primes $p_1, p_2, \ldots, p_k$ as

$$p_{i_1} < p_{i_2} < \cdots < p_{i_k}.$$

If $\pi \in \wp$ and $N < \pi < p_{i_1}$, it is clear that we will have $(n+j, \pi) = 1$ for each $j \in \{1, 2, \ldots, k\}$. Similarly, if $\pi \in \wp$ and $p_{i_1} < \pi < p_{i_2}$, then $(n+j, \pi) = 1$ for each $j \in \{1, 2, \ldots, k\} \setminus \{i_1\}$, and so on. Let us now introduce the function $\rho : \wp \cap (N, p_{i_k}] \to \{0, 1, 2, \ldots, k\}$ defined by

$$\rho(\pi) = \begin{cases} k & \text{if } \; N < \pi < p_{i_1}, \\ k-1 & \text{if } \; p_{i_1} < \pi < p_{i_2}, \\ \vdots & \qquad \vdots \\ 1 & \text{if } \; p_{i_{k-1}} < \pi < p_{i_k}, \\ 0 & \text{if } \; \pi \in \{p_1, p_2, \ldots, p_k\}. \end{cases}$$

By using the Eratosthenian sieve, we easily obtain that, as $y \to \infty$,

$$S_J(p_1, \ldots, p_k) = (1 + o(1)) \frac{y}{p_1 \cdots p_k} \prod_{N < \pi < p_{i_k}} \left(1 - \frac{\rho(\pi)}{\pi}\right). \tag{10.3}$$

Setting $U := \prod_{N < \pi < p_{i_k}} \left(1 - \frac{\rho(\pi)}{\pi}\right)$, one can see that, as $N \to \infty$,

$$\begin{aligned} \log U &= k \log\log N - k \log\log p_{i_1} - (k-1) \log\log p_{i_2} + (k-1) \log\log p_{i_1} \\ &\qquad - \cdots - \log\log p_{i_k} + \log\log p_{i_{k-1}} + o(1) \\ &= k \log\log N - \log\log p_{i_1} - \cdots - \log\log p_{i_k} + o(1), \end{aligned}$$

implying that

$$U = (1 + o(1)) \prod_{j=1}^{k} \frac{\log N}{\log p_j} \qquad (N \to \infty). \tag{10.4}$$

Hence, in light of (10.4), relation (10.3) can be replaced by

$$S_J(p_1, \ldots, p_k) = (1 + o(1)) \frac{y}{p_1 \cdots p_k} \prod_{j=1}^{k} \frac{\log N}{\log p_j} \qquad (y \to \infty). \tag{10.5}$$

Now let $r_1, \dots, r_k$ be an arbitrary collection of reduced residues modulo Q and let us define

$$B_y(r_1,\dots,r_k) := \sum_{\substack{p_j \equiv r_j \pmod{Q} \\ N<p_j\le N^{\lambda_N} \\ j=1,\dots,k}} S_J(p_1,\dots,p_k).$$

From the Prime Number Theorem in arithmetic progressions, we have that

$$\sum_{\substack{u\le p\le u+u/(\log u)^{10} \\ p\equiv \ell \pmod{Q}}} \frac{1}{p\log p} = (1+o(1))\frac{1}{\phi(Q)} \sum_{u\le p\le u+u/(\log u)^{10}} \frac{1}{p\log p} \quad (u\to\infty). \tag{10.6}$$

On the other hand, it is clear that, from the Prime Number Theorem,

$$\sum_{N<p\le N^{\lambda_N}} \frac{1}{p\log p} = (1+o(1))\int_N^{N^{\lambda_N}} \frac{du}{u\log^2 u} = \frac{1+o(1)}{\log N} \qquad (N\to\infty). \tag{10.7}$$

Combining (10.5), (10.7), and (10.6), it follows that, as $y\to\infty$,

$$\begin{aligned} B_y(r_1,\dots,r_k) &= (1+o(1))y \sum_{\substack{p_j\equiv r_j \pmod{Q} \\ N<p_j<N^{\lambda_N} \\ j=1,\dots,k}} \prod_{j=1}^{k} \frac{\log N}{p_j\log p_j} \\ &= (1+o(1))\frac{y}{\phi(Q)^k}. \end{aligned} \tag{10.8}$$

Observe also that

$$\frac{1}{x_N}\#\{n\in J_N : q_N(n) > N^{\lambda_N}\} \to 0 \qquad \text{as } x_N\to\infty. \tag{10.9}$$

Indeed, it is clear that if $q_N(n) > N^{\lambda_N}$, then $\left(n, \prod_{N<\pi<N^{\lambda_N}} \pi\right) = 1$. Therefore, for some absolute constants $C_1 > 0$ and $C_2 > 0$, we have

$$\#\{n\in J_N : q_N(n) > N^{\lambda_N}\} \le Cx_N \prod_{N<\pi\le N^{\lambda_N}} \left(1-\frac{1}{\pi}\right) \le C\frac{x_N}{\lambda_N}, \tag{10.10}$$

which proves (10.9).

We now examine the first M digits of α_Q, say $\alpha_Q^{(M)}$. Let N be such that $x_N \le M < x_{N+1}$ and set $x := x_N$, $y := M - x_N$ and $J_0 = [x, x+y]$.

It follows from (10.8) and (10.10) that, as $y\to\infty$,

$$\#\{n\in J_0 : q_N(n+j)\equiv r_j \pmod{Q} \text{ for } j=1,\dots,k\} = (1+o(1))\frac{y}{\phi(Q)^k} + O\left(\frac{x_N}{\lambda_N}\right), \tag{10.11}$$

where the above error term accounts (as measured by (10.10)) for those integers $n\in J_N$ for which $q_N(n) > N^{\lambda_N}$. Running the same procedure for each positive integer $H < N$, each time choosing $J_H = [x_H, x_{H+1})$, we then obtain a formula similar to the one in (10.11).

Gathering the resulting relations allows us to obtain that, for $X = x + y$,

$$\begin{aligned}
&\lim_{X\to\infty} \frac{1}{X} \# \{n \le X : q_N(n+j) \equiv r_j \pmod Q \text{ for } j = 1, 2, \dots, k\} \\
&= \lim_{X\to\infty} \frac{1}{X} \Big(\sum_{H=1}^{N-1} \# \{n \in J_H : q_N(n+j) \equiv r_j \pmod Q \text{ for } j = 1, 2, \dots, k\} \\
&\qquad + \#\{n \in J_0 : q_N(n+j) \equiv r_j \pmod Q \text{ for } j = 1, \dots, k\} \Big) \\
&\qquad = \frac{1}{\phi(Q)^k},
\end{aligned}$$

thus completing the proof of Theorem 10.1.

Final remarks

Let $\Omega(n) := \sum_{p^\alpha \| n} \alpha$ stand for the number of prime factors of n counting their multiplicity. Fix an integer $Q \ge 3$ and consider the function $u_Q(m) = \ell$, where ℓ is the unique non negative number $\le Q-1$ such that $m \equiv \ell \pmod Q$. Now consider the infinite sequence

$$\xi_Q = \text{Concat}\,(u_Q(\Omega(n)) : n \in \mathbb{N}).$$

We conjecture that ξ_Q is a normal sequence over $\{0, 1, \dots, Q-1\}$.

Moreover, let $\widetilde{\wp} \subset \wp$ be a subset of primes such that $\sum_{p\in\widetilde{\wp}} 1/p = +\infty$ and consider the function $\Omega_{\widetilde{\wp}}(n) := \sum_{\substack{p^\alpha \| n \\ p \in \widetilde{\wp}}} \alpha$. We conjecture that

$$\xi_Q(\widetilde{\wp}) := \text{Concat}\,(u_Q(\Omega_{\widetilde{\wp}}(n)) : n \in \mathbb{N})$$

is also a normal sequence over $\{0, 1, \dots, Q-1\}$.

Finally, observe that we can also construct normal numbers by first choosing a monotonically growing sequence $(w_N)_{N\ge1}$ such that $w_N > N$ for each positive integer N and such that $(\log w_N)/N \to 0$ as $N \to \infty$, and then defining $q_N(n)$ as the smallest prime factor of n larger than w_N if $n \in J_N$, setting $q_N(n) = 1$ otherwise. The proof follows along the same lines as the one of our main result.

XI. The number of prime factors function on shifted primes and normal numbers [25]

(Topics in Mathematical Analysis and Applications, Springer, Volume 94, 2014)

Let $\omega(n)$ stand for the number of distinct prime factors of the positive integer n. One can easily show that the concatenation of the successive values of $\omega(n)$, say by considering the real number $\xi := 0.\overline{\omega(2)}\,\overline{\omega(3)}\,\overline{\omega(4)}\,\overline{\omega(5)}\dots$, where each $\overline{m}$ stands for the q-ary expansion of the integer m, will not yield a normal number. Indeed, since the interval $I := [e^{e^{r-1}}, e^{e^r}]$, where $r := \lfloor \log\log x \rfloor$, covers most of the interval $[1, x]$ and since $\left| \frac{\omega(n)}{r} - 1 \right| < \frac{1}{r^{1/4}}$, say, with the exception of a small number of integers $n \in I$, it follows that ξ cannot be normal in base q.

Recently, Vandehey [62] used another approach to yet create normal numbers using certain small additive functions. He considered irrational numbers formed by concatenating some of the base q digits from additive functions $f(n)$ that closely resemble the prime counting function $\Omega(n) := \sum_{p^\alpha \| n} \alpha$. More precisely, he used the concatenation of the last $\lceil y \frac{\log\log\log n}{\log q} \rceil$ digits of each $f(n)$ in succession and proved that the number thus created turns out to be normal in base q if and only if $0 < y \le 1/2$.

In this paper, we show that the concatenation of the successive values of $|\omega(n) - \lfloor \log\log n \rfloor|$, as n runs through the integers $n \ge 3$, yields a normal number in any given base $q \ge 2$. We show that the same result holds if we consider the concatenation of the successive values of $|\omega(p+1) - \lfloor \log\log(p+1) \rfloor|$, as p runs through the prime numbers.

So, let us first introduce the arithmetic function $\delta(n) := |\omega(n) - \lfloor \log\log n \rfloor|$.

Main results

Theorem 11.1. *Let $R \in \mathbb{Z}[x]$ be a polynomial of positive degree such that $R(y) \ge 0$ for all $y \ge 0$. Let*

$$\eta = Concat(\overline{R(\delta(n))} : n = 3, 4, 5, \ldots).$$

Then, η is a normal sequence in any given base $q \ge 2$.

Theorem 11.2. *Let*

$$\xi = Concat(\overline{\delta(p+1)} : p \in \wp).$$

Then, ξ is a normal sequence in any given base $q \ge 2$.

Remark 11.1. *We shall only provide the proof of Theorem 11.2, the reason being that it is somewhat harder than that of Theorem 11.1. Indeed, for the proof of Theorem 11.1, one can use the fact that*

$$\pi_k(x) := \#\{n \le x : \omega(n) = k\} = (1 + o(1)) \frac{x}{x_1} \frac{x_2^{k-1}}{(k-1)!}$$

uniformly for $|k - x_2| \le \sqrt{x_2}\, x_3$, say, and also the Hardy-Ramanujan inequality

$$\pi_k(x) < c_1 \frac{x}{x_1} \frac{(x_2 + c_2)^{k-1}}{(k-1)!}$$

which is valid uniformly for $1 \le k \le 10x_2$ and $x \ge x_0$ (see for instance the book of De Koninck and Luca [34], p. 157). Hence, using these estimates, one can easily prove Theorem 11.1 essentially as we did to prove that Concat$(P(m) : m \in \mathbb{N})$ is a normal sequence in any given base $q \ge 2$ (see [13]). Now, since there are no known estimate for the asymptotic behavior of $\#\{p \le x : \omega(p+1) = k\}$, we need to find another approach for the proof of Theorem 11.2.

Remark 11.2. *It will be clear from our approach that if $\omega(n)$ is replaced by $\Omega(n)$ or if we consider the function $\delta_2(n) := |\lfloor \log \tau(n) \rfloor - \lfloor \log\log n \rfloor|$ (where $\tau(n)$ stands for the number of positive divisors of n), the same results hold.*

Preliminary lemmas

For each real number $u > 0$, let $\Phi(u) := \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{u} e^{-t^2/2}\, dt$.

Lemma 11.1. *(a) As $x \to \infty$,*

$$\frac{1}{\pi(x)} \# \left\{ p \le x : \frac{\delta(p+1)}{\sqrt{\log\log x}} < u \right\} = (1 + o(1)) \left(\Phi(u) - \Phi(-u) \right).$$

(b) Letting ε_x a function which tends to 0 as $x \to \infty$. Then, as $x \to \infty$,

$$\frac{1}{\pi(x)} \# \left\{ p \le x : \delta(p+1) \le \varepsilon_x \sqrt{\log\log x} \right\} \to 0.$$

Proof. For a proof of part (a), see the book of Elliott [38], page 30. Part (b) is an immediate consequence of part (a). □

Let x be a fixed large number. For each integer $n \ge 2$, we now introduce the function

$$\delta^*(n) := \left| \omega(n) - \lfloor \log\log x \rfloor \right|.$$

Lemma 11.2. *For all $x \ge 2$,*

$$\sum_{p \le x} (\delta^*(p+1))^2 \le c\pi(x) \log\log x.$$

Proof. To obtain this inequality, we may argue as in the proof of the Turán-Kubilius inequality, using the fact that the contribution of those prime divisors which are larger than $x^{1/6}$, say, is small. □

Lemma 11.3. *Given an arbitrary $\kappa \in (0, 1/2)$, then, for all $x \ge 2$,*

$$\#\{p \le x : P(p+1) < x^{\kappa}\} + \#\{p \le x : P(p+1) > x^{1-\kappa}\} \le c\kappa\pi(x).$$

Proof. This is an immediate application of Theorem 4.2 in the book of Halberstam and Richert [43]. □

Lemma 11.4. *Let a and b be two non zero co-prime integers, one of which is even. Then, as $x \to \infty$, we have, uniformly in a and b,*

$$\#\{p \le x : ap + b \in \wp\} \le 8 \prod_{p > 2} \left(1 - \frac{1}{(p-1)^2} \right) \prod_{\substack{p > 2 \\ p \mid ab}} \frac{p-1}{p-2} \frac{x}{\log^2 x} \left(1 + O\left(\frac{\log\log x}{\log x} \right) \right).$$

Proof. This is Theorem 3.12 in the book of Halberstam and Richert [43] for the particular case $k = 1$. □

Lemma 11.5. *Let $M \ge 2k$, $\beta_1, \beta_2 \in \mathcal{A}_q^k$. Set $\Delta(\alpha) = |\nu_{\beta_1}(\alpha) - \nu_{\beta_2}(\alpha)|$. Then,*

$$\sum_{\alpha \in \mathcal{A}_q^{M+1}} \Delta^2(\alpha) \le cMq^M.$$

Proof. Let $\beta = b_{k-1} \dots b_0 \in \mathcal{A}_q^k$. Consider the function $f_\beta : \mathcal{A}_q^k \to \{0,1\}$ defined by

$$f_\beta(u_{k-1}, \dots, u_0) = \begin{cases} 1 & \text{if } u_{k-1} \dots u_0 = \beta, \\ 0 & \text{otherwise.} \end{cases}$$

Let $M \in \mathbb{N}$, $M \geq 2k$. Let $\alpha = \varepsilon_M \dots \varepsilon_0$ run over elements of $\mathcal{A}_q^{M+1}$. It is clear that

$$\begin{aligned} A &:= \sum_{\alpha \in \mathcal{A}_q^{M+1}} \nu_\beta(\alpha) \\ &= \sum_{\nu=0}^{M+1-k} \#\{\alpha \in \mathcal{A}_q^{M+1} : \varepsilon_{\nu+k-1} \dots \varepsilon_\nu = \beta\} \\ &= (M+1-k) q^{M+1-k}. \end{aligned} \tag{11.1}$$

On the other hand,

$$\begin{aligned} B &:= \sum_{\alpha \in \mathcal{A}_q^{M+1}} \nu_\beta^2(\alpha) \\ &= \sum_{\nu_1=0}^{M+1-k} \sum_{\nu_2=0}^{M+1-k} \sum_{\varepsilon_0, \dots, \varepsilon_M} f_\beta(\varepsilon_{\nu_1+k-1}, \dots, \varepsilon_{\nu_1}) f_\beta(\varepsilon_{\nu_2+k-1}, \dots, \varepsilon_{\nu_2}) \\ &= A + 2 \sum_{\substack{\nu_1,\nu_2=0 \\ \nu_1<\nu_2}}^{M+1-k} \sum_{\varepsilon_0, \dots, \varepsilon_M} f_\beta(\varepsilon_{\nu_1+k-1}, \dots, \varepsilon_{\nu_1}) f_\beta(\varepsilon_{\nu_2+k-1}, \dots, \varepsilon_{\nu_2}) \\ &= A + 2 \sum_{\substack{\nu_1,\nu_2=0 \\ \nu_1<\nu_2\leq\nu_1+k}}^{M+1-k} \sum_{\varepsilon_0, \dots, \varepsilon_M} f_\beta(\varepsilon_{\nu_1+k-1}, \dots, \varepsilon_{\nu_1}) f_\beta(\varepsilon_{\nu_2+k-1}, \dots, \varepsilon_{\nu_2}) \\ &\quad + 2 \sum_{\substack{\nu_1,\nu_2=0 \\ \nu_2>\nu_1+k}}^{M+1-k} \sum_{\varepsilon_0, \dots, \varepsilon_M} f_\beta(\varepsilon_{\nu_1+k-1}, \dots, \varepsilon_{\nu_1}) f_\beta(\varepsilon_{\nu_2+k-1}, \dots, \varepsilon_{\nu_2}). \end{aligned} \tag{11.2}$$

Now, on the one hand we have

$$\sum_{\substack{\nu_1,\nu_2=0 \\ \nu_1<\nu_2\leq\nu_1+k}}^{M+1-k} \sum_{\varepsilon_0, \dots, \varepsilon_M} f_\beta(\varepsilon_{\nu_1+k-1}, \dots, \varepsilon_{\nu_1}) f_\beta(\varepsilon_{\nu_2+k-1}, \dots, \varepsilon_{\nu_2}) \leq ckMq^{M+1-k}, \tag{11.3}$$

while on the other hand,

$$\begin{aligned} &\sum_{\substack{\nu_1,\nu_2=0 \\ \nu_2>\nu_1+k}}^{M+1-k} \sum_{\varepsilon_0, \dots, \varepsilon_M} f_\beta(\varepsilon_{\nu_1+k-1}, \dots, \varepsilon_{\nu_1}) f_\beta(\varepsilon_{\nu_2+k-1}, \dots, \varepsilon_{\nu_2}) \\ &= \sum_{\substack{\nu_1,\nu_2=0 \\ \nu_1+k<\nu_2}}^{M+1-k} q^{M+1-2k} = q^{M+1-2k} \left((M+1)^2 - O(kM) \right). \end{aligned} \tag{11.4}$$

Combing (11.1) and (11.2), using estimates (11.3) and (11.4), we conclude that

$$\sum_{\alpha=\varepsilon_M\ldots\varepsilon_0}\left(\nu_\beta(\alpha)-\frac{M+1}{q^k}\right)^2\le cMq^M. \tag{11.5}$$

Note that here we summed over those $\varepsilon_M=0$ as well. But (11.5) remains true if we drop those $\varepsilon_M=0$. This allows us to conclude that

$$\sum_{\alpha\in\mathcal{A}_q^{M+1}}\left(\nu_\beta(\alpha)-\frac{M+1}{q^k}\right)^2\le cMq^M,$$

thus completing the proof of Lemma 11.5.

□

Proof of Theorem 11.2

Let

$$\xi_x=\mathrm{Concat}(\overline{\delta(p+1)}:p\le x).$$

Our first goal is to prove that there exist two positive constants c_1 and c_2 such that

$$c_1\le\frac{\lambda(\xi_x)}{\pi(x)\,x_3}\le c_2, \tag{11.6}$$

provided x is sufficiently large, from which it will follow that the order of $\lambda(\xi_x)$ is $\pi(x)x_3$.

We have

$$\lambda(\xi_x)=\sum_{\substack{p\le x\\ \delta(p+1)\ne0}}\left\lfloor\frac{\log\delta(p+1)}{\log q}\right\rfloor+O(\pi(x))=\Sigma_1+\Sigma_2+O(\pi(x)), \tag{11.7}$$

say, where the sum in Σ_1 runs over the primes $p\le x/x_2$, while that of Σ_2 runs over the primes located in the interval $J_x:=(x/x_2,x]$.

It follows from Lemma 11.1 that, for each $u>0$ there exists $c(u)>0$ such that

$$\#\left\{p\le x:\frac{\delta(p+1)}{\sqrt{x_2}}>u\right\}>c(u)\pi(x),$$

from which it follows that

$$\Sigma_2\ge c\pi(x)\,x_3 \tag{11.8}$$

and therefore, from (11.7), that, if $x>x_0$, the inequality $\dfrac{\lambda(\xi_x)}{\pi(x)\,x_3}>c$ holds for some positive constant c, thereby establishing the first inequality in (11.6).

To obtain the upper bound in (11.6), first observe that

$$\Sigma_1\le2\pi(x/x_2)x_2=O(\pi(x)). \tag{11.9}$$

On the other hand, from the definitions of the functions δ and δ^*, it is clear that

$$|\delta^*(p+1)-\delta(p+1)| \le 1 \qquad \text{for all } p \in J_x.$$

Hence,

$$\begin{aligned}
\Sigma_2 &\le c \sum_{x/x_2<p\le x} \log \delta^*(p+1) = c \sum_{\substack{x/x_2<p\le x\\ \delta^*(p+1)\le 4\sqrt{x_2}}} \log \delta^*(p+1) + c \sum_{\substack{x/x_2<p\le x\\ \delta^*(p+1)>4\sqrt{x_2}}} \log \delta^*(p+1)\\
&\le \left(2\log 2 + \frac{x_3}{2}\right)\pi(x) + c \sum_{\substack{x/x_2<p\le x\\ \delta^*(p+1)>4\sqrt{x_2}}} \log \delta^*(p+1)\\
&\le c_3\pi(x)x_3 + \Sigma_3, \qquad (11.10)
\end{aligned}$$

say.

From Lemma 11.2, we obtain that for every $A \ge 1$,

$$\#\left\{p \in J_x : \frac{\delta^*(p+1)}{\sqrt{x_2}} > A\right\} \le \frac{c\pi(x)}{A^2}. \tag{11.11}$$

We now apply (11.11) successively with $A = 2^j$, $j = 2, 3, \ldots$, thus obtaining

$$\begin{aligned}
\Sigma_3 &\le c\pi(x) \sum_{j\ge 2} \frac{\log\left(2^{j+1}\sqrt{x_2}\right)}{2^{2j}}\\
&\le c\pi(x) \sum_{j\ge 2} \left(\frac{(j+1)\log 2}{4^j} + \frac{x_3}{2\cdot 4^j}\right)\\
&\le c_4\pi(x)\, x_3,
\end{aligned}$$

from which we may conclude, in light of (11.7), (11.9) and (11.10), that the right hand side of (11.6) holds as well.

We will now prove that, given any fixed integer $k \ge 1$ and distinct words $\beta_1, \beta_2 \in \mathcal{A}_q^k$, and setting $\Delta(\alpha) := \nu_{\beta_1}(\alpha) - \nu_{\beta_2}(\alpha)$ for each word $\alpha \in \mathcal{A}_q^*$,

$$\lim_{x\to\infty} \frac{|\Delta(\xi_x)|}{\lambda(\xi_x)} = 0. \tag{11.12}$$

In order to achieve this, now that we know (from (11.6)) that the true order of $\lambda(\xi_x)$ is $\pi(x)\,x_3$, we essentially need to prove that $\Delta(\xi_x)$ is of smaller order than $\pi(x)\,x_3$.

Let θ_x be an arbitrary function which tends monotonically to 0 very slowly. Then consider the sets

$$\begin{aligned}
D_1 &= \{p \in \wp : p \le x/x_2\},\\
D_2 &= \{p \in \wp : p \le x \text{ and } \delta(p+1) \le \theta_x\sqrt{x_2}\},\\
D_3 &= \{p \in \wp : p \le x \text{ and } \delta(p+1) > \frac{1}{\theta_x}\sqrt{x_2}\},
\end{aligned}$$

and let $D = D_1 \cup D_2 \cup D_3$.

Because $\Delta(\delta(p+1)) \le cx_3$ if $p \in D_1$ and $p \le cx_2$, and since (11.11) holds for $p \in D_3$, it follows from Lemma 11.1 and (11.9) that

$$\begin{aligned}\sum_{p\in D} |\Delta(\delta(p+1))| \;&\le\; cx_3\pi(x)\left(\Phi(\theta_x) - \Phi(-\theta_x)\right) + c\pi(x/x_2)x_2 \\ &\quad + \sum_{j=0}^{\infty} \#\left\{p \in J_x : \frac{\delta^*(p+1)}{\sqrt{x_2}} \in \left[\frac{2^j}{\theta_x}, \frac{2^{j+1}}{\theta_x}\right]\right\} \cdot \log\left(\sqrt{x_2}\cdot\frac{2^{j+1}}{\theta_x}\right). \end{aligned} \tag{11.13}$$

Since this last sum is less than

$$\pi(x)\sum_{j\ge 0}(x_3 + j + \log(1/\theta_x))\cdot\frac{\theta_x^2}{2^{2j}} \le c\left(\log(1/\theta_x) + x_3\right)\theta_x^2\pi(x),$$

it follows from (11.13) that

$$\sum_{p\in D} |\Delta(\delta(p+1))| = o(\pi(x)\, x_3) \qquad (x\to\infty). \tag{11.14}$$

Using (11.14), we then have

$$\Delta(\xi_x) = \sum_{p\notin D}\Delta(\delta(p+1)) + o(\pi(x)\,x_3) = \Sigma_A + o(\pi(x)\,x_3), \tag{11.15}$$

say.

Let $\kappa \in (0,1/2)$. From Lemma 11.3, we obtain, using the fact that $p \notin D_3$ (since $p \notin D$), that

$$\sum_{\substack{p\notin D\\ P(p+1)\notin[x^\kappa, x^{1-\kappa}]}} |\Delta(\delta(p+1))| \le c\kappa\pi(x)\log\left(\frac{1}{\theta_x}\sqrt{x_2}\right) \le c_1\kappa\pi(x)x_3, \tag{11.16}$$

provided that θ_x is chosen so that $1/\theta_x < x_2$, say.

Now let $K = \lfloor x_2 \rfloor$ and then, for ℓ satisfying $\varepsilon_x\sqrt{K} \le |\ell| \le \frac{1}{\varepsilon_x}\sqrt{K}$, where ε_x is a function which tends to infinity very slowly as $x\to\infty$ and which will be chosen appropriately later on.

Further set

$$R_\kappa(\ell) := \#\{p \in J_x : P(p+1) \in (x^\kappa, x^{1-\kappa}) \text{ and } \omega(p+1) = K+\ell\}.$$

Using Lemma 11.4, we obtain that

$$\begin{aligned} R_\kappa(\ell) \;&\le\; \#\{p\in J_x : p+1 = aq,\ a < x^{1-\kappa},\ q > x^\kappa/x_2,\ \omega(a) = K+\ell-1\} \\ &\le\; \frac{1}{\kappa^2}\frac{x}{\log^2 x}\sum_{\omega(n)=K+\ell-1}\frac{1}{a}\prod_{\substack{p>2\\ p\mid a}}\frac{p-1}{p-2} + O\left(x^{1-\kappa}\right), \end{aligned} \tag{11.17}$$

where the $O(\dots)$ term accounts for the contribution of those q such that $q^2 \mid p+1$.

It then follows from (11.17) that

$$
\begin{aligned}
R_\kappa(\ell) &\le \frac{c_2 x}{\kappa^2 \log^2 x}\left(\sum_{p\le x}\frac{1}{p}+c\right)^{K+\ell-1}\frac{1}{(K+\ell-1)!}+O\left(x^{1-\kappa}\right)\\
&\le \frac{c_3 x}{\kappa^2\log^2 x}\frac{(K+c)^{K+\ell-1}}{(K+\ell-1)!}. \qquad (11.18)
\end{aligned}
$$

Now, observe that, if $\omega(p+1)=K+\ell$, $p\in J_x$, then $\delta(p+1)\in\{|\ell|-1,|\ell|,|\ell|+1\}$. Therefore, recalling (11.16),

$$
\begin{aligned}
|\Sigma_A| &\le \sum_{\varepsilon_x\sqrt{K}\le|\ell|\le\frac{1}{\varepsilon_x}\sqrt{K}}\left(\Delta(|\bar{\ell}|)+\Delta(|\bar{\ell}|-1)+\Delta(|\bar{\ell}|+1)\right)\cdot\left(R_\kappa(-\ell)+R_\kappa(\ell)\right)\\
&\qquad +c_1\kappa\pi(x)x_3\\
&= \Sigma_B+c_1\kappa\pi(x)x_3, \qquad (11.19)
\end{aligned}
$$

say.

Using (11.18), we obtain that

$$
\Sigma_B \le \frac{c_4 x}{\kappa^2\log^2 x}\sum_{\varepsilon_x\le\frac{\ell}{\sqrt{K}}\le\frac{1}{\varepsilon_x}}\left(\Delta(|\bar{\ell}|)+\Delta(|\bar{\ell}|-1)+\Delta(|\bar{\ell}|+1)\right)\cdot\left(\frac{(K+c)^{K+\ell-1}}{(K+\ell-1)!}+\frac{(K+c)^{K-\ell-1}}{(K-\ell-1)!}\right). \qquad (11.20)
$$

Since we can easily establish that

$$
\max_{0\le\ell\le\frac{\sqrt{K}}{\varepsilon_x}}\left(\frac{(K+c)^{K+\ell-1}}{(K+\ell-1)!}+\frac{(K+c)^{K-\ell-1}}{(K-\ell-1)!}\right)<\frac{(K+c)^{K-1}}{(K-1)!}\exp\left\{c_5\left(\frac{1}{\varepsilon_x}\right)^2\right\},
$$

it follows from (11.20) that

$$
\Sigma_B\le\frac{c_4 x}{\kappa^2\log^2 x}\exp\left\{c_5\left(\frac{1}{\varepsilon_x}\right)^2\right\}\frac{(K+c)^{K-1}}{(K-1)!}\Sigma_C, \qquad (11.21)
$$

where

$$
\begin{aligned}
\Sigma_C &= \sum_{\varepsilon_x\le\frac{\ell}{\sqrt{K}}\le\frac{1}{\varepsilon_x}}\left(\Delta(\bar{\ell})+\Delta(\bar{\ell}-1)+\Delta(\bar{\ell}+1)\right)\\
&\le 3\sum_{\varepsilon_x\le\frac{\ell}{\sqrt{K}}\le\frac{1}{\varepsilon_x}}\Delta(\bar{\ell})+O(x_3)=3\Sigma_D+O(x_3), \qquad (11.22)
\end{aligned}
$$

say.

To estimate Σ_D, we will use Lemma 11.5. Indeed, let M_0 be the largest integer for which $q^{M_0}\le\varepsilon_x\sqrt{K}$ and let M_1 be the smallest integer for which $q^{M_1}>\frac{1}{\varepsilon_x}\sqrt{K}$. Set $\mathcal{K}_M=[q^M,q^{M+1}-1]$. With this set up, we clearly have that

$$
\Sigma_D\le\sum_{M_0\le M\le M_1}T_M, \qquad (11.23)
$$

where $T_M = \sum_{\ell \in \mathcal{K}_M} \Delta(\ell)$. Now, it follows from Lemma 11.5 that

$$T_M \le c(q^{M+1})^{1/2}(Mq^{M-k})^{1/2} \le c\sqrt{M}q^M. \tag{11.24}$$

Using (11.24) in (11.23), we obtain that

$$\Sigma_D \le c\sqrt{M_1}q^{M_1}\left(1+\frac{1}{q}+\frac{1}{q^2}+\cdots\right) < \frac{c_6}{\varepsilon_x}\sqrt{K}\,\sqrt{\log K} < \frac{c_6\sqrt{x_2}\,\sqrt{x_3}}{\varepsilon_x}. \tag{11.25}$$

Gathering (11.21), (11.22) and (11.25), we have that

$$\Sigma_B \le \frac{c_7 x}{\kappa^2 \log^2 x}\exp\left\{c_5\left(\frac{1}{\varepsilon_x}\right)^2\right\}\frac{(K+c)^{K-1}}{(K-1)!}\cdot\frac{\sqrt{x_2}\sqrt{x_3}}{\varepsilon_x}. \tag{11.26}$$

Setting $\ell_K = \dfrac{(K+c)^{K-1}}{(K-1)!}$ and using Stirling's formula $n! = n^n e^{-n}\sqrt{2\pi n}(1+O(1/n))$, we have that

$$\begin{aligned}
\log \ell_K &= (K-1)\log(K+c) - (K-1)\log\left(\frac{K-1}{e}\right) - \frac{1}{2}\log K + O(1)\\
&= (K-1)\log\frac{K+c}{K-1} - \frac{1}{2}\log K + O(1) + K - 1,
\end{aligned}$$

from which it follows that

$$\ell_K \le c_8 \frac{x_1}{\sqrt{x_2}}.$$

Using this last estimate in (11.26), we obtain that

$$\Sigma_B \ll \frac{\exp\{c_5/\varepsilon_x^2\}}{\kappa^2\varepsilon_x}\pi(x)\,\sqrt{x_3}. \tag{11.27}$$

Choosing $\varepsilon_x = x_5$, say, while using (11.27) and (11.6), we conclude that

$$\limsup_{x\to\infty}\frac{\Sigma_B}{\lambda(\xi_x)} = 0. \tag{11.28}$$

Combining (11.28), (11.19) and (11.15), we obtain that

$$\limsup_{x\to\infty}\frac{\Delta(\xi_x)}{\lambda(\xi_x)} \le c\kappa. \tag{11.29}$$

Since κ can be taken arbitrarily small, we may finally conclude that (11.12) holds, thus completing the proof of Theorem 11.2.

XII. Normal numbers generated using the smallest prime factor function [22]

(Annales mathématiques du Québec, 2014)

In a series of recent papers, we constructed large families of normal numbers using the distribution of the values of the largest prime factor function (see for instance [13], [17] and [20]). What if we consider instead the function $p(n)$, which stands for the smallest prime factor of an integer $n \geq 2$? At first, one might think that the (base 10) real number $\eta_1 := 0.p(2)p(3)p(4)p(5)\ldots$ is not a normal number because $p(n) = 2$ for every even number. But, on the contrary, as we will show here, η_1 is indeed a normal number. In fact, it turns out that the smallest prime factor of an (odd) integer is often very large with a decimal expansion which "most of the times" contains all ten digits at essentially the same frequency.

Here,we examine various constructions of real numbers involving the smallest prime factor function $p(n)$, including ones where the integers n run through the set of shifted primes.

Main results

Theorem 12.1. *The expression $n_1 = Concat(\overline{p(n)} : n \in \mathbb{N})$ is a normal sequence.*

Theorem 12.2. *Let $R \in \mathbb{Z}[x]$ be a polynomial such that $R(x) > 0$ for all $x > 0$ and satisfying $\lim_{x\to\infty} R(x) = \infty$. The expression $n_2 = Concat(\overline{R(p(n))} : n \in \mathbb{N})$ is a normal sequence.*

Theorem 12.3. *Let $a \in \mathbb{N} \cup \{0\}$ be an even integer. The expression $n_3 = Concat(\overline{p(\pi + a)} : \pi \in \wp)$ is a normal sequence.*

Remark 12.1. *Observe that the particular case $a = 0$ has been proved by Davenport and Erdős [11].*

Theorem 12.4. *Let $a \in \mathbb{N} \cup \{0\}$ be an even integer and let R be as in Theorem 12.2. The expression $n_4 = Concat(\overline{R(p(\pi + a))} : \pi \in \wp)$ is a normal sequence.*

We will only provide the proofs of Theorems 12.1 and 12.3, since those of Theorems 12.2 and 12.4 can be obtained along the same lines.

Proof of Theorem 12.1

Let x be a large number, but fixed. Consider the interval

$$I_x := \left[\left\lfloor \frac{x}{2} \right\rfloor + 1, \lfloor x \rfloor\right)$$

and the following two subwords of n_1:

$$\eta_x := \text{Concat}(\overline{p(n)} : n \leq x), \qquad \rho_x := \text{Concat}(\overline{p(n)} : n \in I_x).$$

Let β be an arbitrary word in $\mathcal{A}_q^k$.

Letting ℓ_0 be the largest integer such that $2^{\ell_0} < x$, it is clear that

$$\nu_\beta(\eta_x) \;=\; \sum_{\ell=0}^{\ell_0} \nu_\beta(\rho_{x/2^\ell}) + O(\log x), \tag{12.1}$$

$$\nu_\beta(\rho_{x/2^\ell}) = \sum_{n\in I_{x/2^\ell}} \nu_\beta(\overline{p(n)}) + O\left(\frac{x}{2^\ell}\right), \tag{12.2}$$

where the error term on the right hand side of (12.1) accounts for the cases where the word β overlaps two consecutive intervals $I_{x/2^{\ell+1}}$ and $I_{x/2^\ell}$. Note that here and throughout this section, the constants implied by the Landau notation $O(\cdots)$ may depend on the particular base q and on the particular word β.

Hence, in light of (12.1) and (12.2), in order to prove that n_1 is a normal sequence, it will be sufficient to show that, given any two distinct words $\beta_1, \beta_2 \in \mathcal{A}_q^k$, we have

$$\frac{|\nu_{\beta_1}(\rho_x) - \nu_{\beta_2}(\rho_x)|}{\lambda(\rho_x)} \to 0 \qquad \text{as } x \to \infty. \tag{12.3}$$

We first start by establishing the exact order of $\lambda(\rho_x)$.

For each $Q \in \wp$, we let

$$\mathfrak{F}_x(Q) = \#\{n \in I_x : p(n) = Q\}.$$

Let ε_x be a function such that $\lim_{x\to\infty} \varepsilon_x = 0$. Let also $Y_x < Z_x$ be two positive functions tending to infinity with x, that we will specify later. It is clear, using Mertens' formula, that, as $x \to \infty$,

$$\mathfrak{F}_x(Q) = (1 + o(1))\frac{x}{2Q}\prod_{\substack{\pi<Q\\ \pi\in\wp}}\left(1 - \frac{1}{\pi}\right) = (1 + o(1))\frac{e^{-\gamma}}{2}\frac{x}{Q\log Q} \tag{12.4}$$

uniformly for $Y_x < Q \le x^{\varepsilon_x}$ (here γ stands for the Euler-Mascheroni constant). By a sieve approach, we may say that for some absolute constant $c_1 > 0$, we have

$$\mathfrak{F}_x(Q) \begin{cases} \le c_1 \frac{x}{Q\log Q} & \text{for all } Q \le \sqrt{x}, \\ \le \frac{x}{Q} & \text{for } \sqrt{x} < Q \le x. \end{cases} \tag{12.5}$$

We may then write

$$\begin{aligned} \lambda(\rho_x) &= \sum_{Q<Y_x} \mathfrak{F}_x(Q)\lambda(\overline{Q}) + \sum_{Y_x\le Q<Z_x} \mathfrak{F}_x(Q)\lambda(\overline{Q}) + \sum_{Z_x\le Q\le x} \mathfrak{F}_x(Q)\lambda(\overline{Q}) + O(x) \\ &= \Sigma_1 + \Sigma_2 + \Sigma_3 + O(x), \end{aligned} \tag{12.6}$$

say. As we will see, the main contribution will come from the term Σ_2.

Using (12.4) and (12.5), we easily obtain

$$\Sigma_1 \le c_2 x \sum_{Q<Y_x} \frac{1}{Q\log Q} \cdot \log Q \le c_3 x \log\log Y_x, \tag{12.7}$$

$$\Sigma_3 \le c_4 x \sum_{Z_x\le Q\le x} \frac{1}{Q} \le c_5 x \log\left(\frac{\log x}{\log Z_x}\right). \tag{12.8}$$

Choosing Y_x so that $\log Y_x = (\log x)^{\varepsilon_x}$ and Z_x so that $\dfrac{\log x}{\log Z_x} = (\log x)^{\varepsilon_x}$, it follows from (12.7) and (12.8) that, as $x \to \infty$,

$$\Sigma_1 = o(x \log\log x), \tag{12.9}$$
$$\Sigma_3 = o(x \log\log x). \tag{12.10}$$

Now, in light of (12.4), we have, as $x \to \infty$,

$$\begin{aligned}
\Sigma_2 &= \sum_{Y_x \le Q < Z_x} \mathfrak{F}_x(Q)\lambda(\overline{Q}) \\
&= (1+o(1))c_6 x \sum_{Y_x \le Q < Z_x} \frac{\lambda(\overline{Q})}{Q \log Q} = (1+o(1)) \frac{c_7 x}{\log q} \sum_{Y_x \le Q < Z_x} \frac{1}{Q} \\
&= (1+o(1))c_7 x \log\left(\frac{\log Z_x}{\log Y_x}\right) = (1+o(1))c_7 x \log\log x + O(x\varepsilon_x \log\log x), \qquad (12.11)
\end{aligned}$$

for some positive constants c_6 and c_7.

Hence, gathering estimates (12.9), (12.10) and (12.11) and substituting them into (12.6), we obtain that

$$\lambda(\rho_x) = c_7 x \log\log x + o(x \log\log x),$$

thus establishing that the true order of $\lambda(\rho_x)$ is $x \log\log x$. Therefore, in light of our ultimate goal (12.3), we now only need to show that

$$|\nu_{\beta_1}(\rho_x) - \nu_{\beta_2}(\rho_x)| = o(x \log\log x) \qquad (x \to \infty). \tag{12.12}$$

To accomplish this, using the same approach as above, we easily get that

$$|\nu_{\beta_1}(\rho_x) - \nu_{\beta_2}(\rho_x)| \le \sum_{Y_x < Q < Z_x} \left|\nu_{\beta_1}(\overline{Q}) - \nu_{\beta_2}(\overline{Q})\right| \mathfrak{F}_x(Q) + o(x \log\log x). \tag{12.13}$$

We further set ℓ_1 as the largest integer such that $2^{\ell_1+1} \le Y_x$ and ℓ_2 as the smallest integer such that $2^{\ell_2+1} \ge Z_x$. We then write the interval $[Y_x, Z_x]$ as a subset of the union of a finite number of intervals, namely as follows:

$$[Y_x, Z_x] \subseteq \bigcup_{\ell=\ell_1}^{\ell_2} \left[\frac{x}{2^{\ell+1}}, \frac{x}{2^\ell}\right], \tag{12.14}$$

that is the union of a finite number of intervals of the form $[u, 2u]$.

For each of these intervals $[u, 2u]$, we have

$$T(u) := \sum_{u \le Q \le 2u} \left|\nu_{\beta_1}(\overline{Q}) - \nu_{\beta_2}(\overline{Q})\right| \mathfrak{F}_x(Q) = S_1(u) + S_2(u), \tag{12.15}$$

where $S_1(u)$ is the same as $T(u)$ but with the restriction that the sum runs only over those primes $Q \in [u, 2u]$ for which

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

while $S_2(u)$ accounts for the other primes $Q \in [u, 2u]$, namely those for which

$$\left|\nu_{\beta_1}(\overline{Q}) - \nu_{\beta_2}(\overline{Q})\right| > \kappa_u \sqrt{L(u)}.$$

Using Lemma 0.5 and (12.5), we thus have that, for some positive constants c_8 and c_9,

$$\begin{aligned} S_1(u) &\leq c_8 \sum_{u \leq Q \leq 2u} \kappa_u \sqrt{\log u}\, \mathfrak{F}_x(Q) \leq c_8 \kappa_u \sqrt{\log u}\; x \sum_{u \leq Q \leq 2u} \frac{1}{Q \log Q} \\ &\leq c_9 \frac{x \kappa_u}{(\log u)^{3/2}}. \end{aligned} \tag{12.16}$$

On the other hand, using the trivial estimate $\nu_{\beta_i}(\overline{Q}) \leq \lambda(\overline{Q}) \ll \log u$, we easily get, again using Lemma 0.5 and (12.5), that, for some positive constant c_{10},

$$S_2(u) \leq \frac{c_{10} x}{u} \frac{u}{(\log u)\kappa_u^2} = \frac{c_{10} x}{(\log u)\kappa_u^2}. \tag{12.17}$$

Substituting (12.16) and (12.17) in (12.15), we obtain that

$$T(u) \leq cx \left(\frac{\kappa_u}{(\log u)^{3/2}} + \frac{1}{(\log u) \cdot \kappa_u^2} \right). \tag{12.18}$$

We now choose $\kappa_u = \log\log\log x$. Then, in light of (12.14) and using (12.18), we may conclude that

$$\sum_{Y_x < Q < Z_x} |\nu_{\beta_1}(\rho_x) - \nu_{\beta_2}(\rho_x)| \leq \sum_{\ell=\ell_1}^{\ell_2} T\left(\frac{x}{2^\ell}\right) \leq o(x \log\log x),$$

which in light of (12.13) proves (12.12), thereby completing the proof of Theorem 12.1.

Proof of Theorem 12.3

We let x be a large number and turn our attention to the truncated word

$$\sigma_x = \text{Concat}(\overline{p(\pi + a)} : \pi \in I_x),$$

of which we first plan to estimate the size of $\lambda(\sigma_x)$.

For each prime number U, let

$$M_x(U) = \#\{\pi \in I_x : p(\pi + a) = U\}.$$

This allows us to write

$$\lambda(\sigma_x) = \sum_{U \in \wp} M_x(U) \lambda(\overline{U}) = \sum_{\substack{U < x^{\varepsilon_x} \\ U \in \wp}} + \sum_{\substack{U \geq x^{\varepsilon_x} \\ U \in \wp}} = \Sigma_1 + \Sigma_2, \tag{12.19}$$

say. Using Theorem 4.2 of Halberstam and Richert [43], we get that

$$\Sigma_2 \;\leq\; (\log x) \cdot \#\{\pi < x : p(\pi + a) \geq x^{\varepsilon_x}\}$$

$$\leq \; c\frac{x\log x}{\log x}\prod_{p<x^{\varepsilon_x}}\left(1-\frac{1}{p}\right)\leq c_1\frac{x}{\varepsilon_x\log x}, \tag{12.20}$$

by Mertens' estimate.

Let us choose ε_x so that $1/\varepsilon_x$ tends monotonically to infinity, but very slowly. We will now use Lemma 0.11 and the Bombieri-Vinogradov theorem to estimate $M_x(U)$ for $U < x^{\varepsilon_x}$ for almost all U. Choose $\kappa_U = 1/\sqrt{\varepsilon_U}$.

Following the notation of Lemma 0.11, we have

$$T_U = \prod_{p<U} p, \quad p_1 < \cdots < p_s(\leq U), \quad \Delta = \pi(U) - 1,$$

$$\left(\frac{x}{2}\leq\right)\pi_1 < \cdots < \pi_N(\leq x), \quad \pi_j + a \equiv 0 \pmod{U},$$

$$a_n = \pi_n + a \text{ for } n = 1,2,\ldots,N, \quad f(n) = 1 \text{ for all } n \in \mathbb{N}.$$

Moreover, for each $d|T_U$,

$$\pi(I_x; dU, -a) = \sum_{a_n\equiv 0 \pmod d} f(n) = \frac{1}{\phi(d)(U-1)}\left(\operatorname{li}(x) - \operatorname{li}(x/2)\right) + R(N, dU, -a),$$

say. We have

$$|R(N,dU,-a)| \leq \left|\pi(x;dU,-a) - \frac{\operatorname{li}(x)}{\phi(dQ)}\right| + \left|\pi(\frac{x}{2};dU,-a) - \frac{\operatorname{li}(x/2)}{\phi(dQ)}\right|.$$

Let η be the multiplicative function defined on the squarefree integers by

$$\eta(p) = \begin{cases} 1/(p-1) & \text{if } p \nmid a, \\ 0 & \text{if } p \mid a. \end{cases}$$

We then have

$$S = \sum_{\substack{p|T_U \\ p\nmid a}} \frac{\log p}{p-2} = \log U + O(1).$$

Then, the condition

$$\frac{1}{8}\log z \geq \max(\log \pi(U), \log U)$$

clearly holds for every large U. Further set

$$H = H_U = \exp\left\{-\kappa_U\left(\log\kappa_U - \log\log\kappa_U - \frac{2}{\kappa_U}\right)\right\}.$$

We then have

$$M_x(U) = \{1 + 2\theta_1 H\}\frac{\operatorname{li}(x)-\operatorname{li}(x/2)}{U-1}\prod_{\substack{2<p<U \\ p\nmid a}}\left(1-\frac{1}{p-1}\right) + B(U), \tag{12.21}$$

where

$$B(U) = 2\theta_2 \sum_{\substack{d|T_U \\ d \le U^{\kappa_U}}} 3^{\omega(d)} |R(N,d)|,$$

and where $|\theta_1| \le 1$, $|\theta_2| \le 1$.

On the one hand, there exists a constant $A_1 = A_1(a) > 0$ such that

$$\prod_{\substack{2<p<U \\ p\nmid a}} \left(1 - \frac{1}{p-1}\right) = (1+o(1))\frac{A_1}{\log U} \qquad (U \to \infty). \tag{12.22}$$

On the other hand,

$$\begin{aligned} \sum_{U \le x^{\varepsilon_x}} B(U) \quad &\le \quad \sum_{U \le x^{\varepsilon_x}} \sum_{\substack{d|T_U \\ d \le U^{\kappa_U}}} 3^{\omega(d)} \left| \pi(x; dU, -a) - \frac{\mathrm{li}(x)}{\phi(dU)} \right| \\ &\quad + \sum_{U \le x^{\varepsilon_x}} \sum_{\substack{d|T_U \\ d \le U^{\kappa_U}}} 3^{\omega(d)} \left| \pi(x/2; dU, -a) - \frac{\mathrm{li}(x/2)}{\phi(dU)} \right| \\ &= \quad S_1(x) + S_2(x), \end{aligned} \tag{12.23}$$

say. We have $dU \le U^{\kappa_U+1}$. Set $m = dU$. Since $U = P(m)$, it follows that m determines d and U uniquely.

We shall now provide an estimate for $S_1(x)$ by using the Brun-Titchmarsh inequality (Lemma 0.1) and the Bombieri-Vinogradov theorem (Lemma 0.2). So, let $B > 0$ and $E > 0$ be arbitrary numbers. We then have

$$\begin{aligned} S_1(x) \quad &\ll \quad \sum_{\substack{m \le x^{\sqrt{\varepsilon_x}+\varepsilon_x} \\ \omega(m) \le Bx_2}} 3^{Bx_2} \left| \pi(x; m, -a) - \frac{\mathrm{li}(x)}{\phi(m)} \right| + \sum_{\substack{m \le x^{\sqrt{\varepsilon_x}+\varepsilon_x} \\ \omega(m) > Bx_2}} 3^{\omega(m)} \frac{\mathrm{li}(x)}{\phi(m)} \\ &\ll \quad \frac{x \cdot 3^{Bx_2}}{x_1^E} + \frac{\mathrm{li}(x)}{3^{Bx_2}} \sum_{\substack{m \le x^{1/4} \\ P(m) < x^{\varepsilon_x}}} \frac{3^{2\omega(m)}}{\phi(m)} \\ &\ll \quad \frac{x \cdot 3^{Bx_2}}{x_1^E} + \frac{\mathrm{li}(x)}{3^{Bx_2}} \prod_{p < x^{\varepsilon_x}} \left(1 + \frac{9p}{(p-1)^2}\right). \end{aligned} \tag{12.24}$$

It follows from (12.24) that, given any fixed number $A > 0$, an appropriate choice of B and E will lead to

$$S_1(x) \ll \frac{\mathrm{li}(x)}{\log^A x}. \tag{12.25}$$

Proceeding in a similar manner, we easily obtain that

$$S_2(x) \ll \frac{\mathrm{li}(x)}{\log^A x}. \tag{12.26}$$

Using (12.25) and (12.26) in (12.23), and combining this with (12.22) and (12.21) in our estimate (12.20), and recalling (12.19), we obtain

$$\lambda(\sigma_x) = \sum_{U \in \wp} M_x(U)\lambda(\overline{U}) = \Sigma_1 + \Sigma_2 \ll \Sigma_1 + \frac{x}{(\log x)\varepsilon_x}. \tag{12.27}$$

Let us now write

$$\Sigma_1 = \sum_{U<\log x} + \sum_{\log x \leq U < x^{\varepsilon_x}} = T_1 + T_2, \tag{12.28}$$

say.

First observe that, using (12.21), as $x \to \infty$,

$$\begin{aligned} T_1 &= \sum_{U<\log x} (1+o(1)) \frac{A_1(\mathrm{li}(x) - \mathrm{li}(x/2))}{(U-1)\log U} \lambda(\overline{U}) + O\left(\frac{\mathrm{li}(x)}{\log^A x}\right) \\ &\ll \mathrm{li}(x) \sum_{U<\log x} \frac{1}{U} + O\left(\frac{\mathrm{li}(x)}{\log^A x}\right) \\ &\ll \mathrm{li}(x) \cdot x_3, \end{aligned} \tag{12.29}$$

while

$$\begin{aligned} T_2 &\ll (\mathrm{li}(x) - \mathrm{li}(x/2)) \sum_{\log x \leq U < x^{\varepsilon_x}} \frac{1}{(U-1)\log U} \left\lceil \frac{\log U}{\log q} \right\rceil \\ &\ll \frac{1}{\log q} (\mathrm{li}(x) - \mathrm{li}(x/2)) \sum_{\log x \leq U < x^{\varepsilon_x}} \frac{1}{U} \\ &= (1+o(1)) \frac{1}{\log q} (\mathrm{li}(x) - \mathrm{li}(x/2)) \log\log x. \end{aligned} \tag{12.30}$$

Gathering (12.29), (12.30) and (12.28) in (12.27), we get

$$\lambda(\sigma_x) \ll \frac{1}{2\log q \cdot \log x} (\mathrm{li}(x) - \mathrm{li}(x/2))\, x_2 \ll \frac{x x_2}{\log x}. \tag{12.31}$$

Let $\beta_1, \beta_2 \in \mathcal{A}_q^k$ and set $\Delta(\alpha) = \nu_{\beta_1}(\alpha) - \nu_{\beta_2}(\alpha)$. We will prove that

$$\lim_{x\to\infty} \frac{|\Delta(\sigma_x)|}{\lambda(\sigma_x)} = 0. \tag{12.32}$$

First, observe that it is clear that

$$|\Delta(\sigma_x)| \leq \sum_{U\in\wp} M_x(U)|\Delta(\overline{U})| + O(1) \sum_{U\in\wp} M_x(U).$$

By using (12.20), we obtain that

$$\sum_{U > x^{\varepsilon_x}} M_x(U) \leq c \frac{x}{\varepsilon_x \log^2 x}.$$

By using (12.25) and (12.26), we obtain that

$$\sum_{\substack{U\in\wp\\ U\le x^{\varepsilon_x}}} B(U)|\Delta(\overline{U})| \le \log x\cdot \sum_{\substack{U\in\wp\\ U\le x^{\varepsilon_x}}} B(U) \le \frac{x}{\log^2 x},$$

provided $x > x_0$.

Thus, by using (12.21) and (12.29), we obtain that

$$|\Delta(\sigma_x)| \le \sum_{\substack{U\in\wp\\ \log x\le U\le x^{\varepsilon_x}}} c\frac{x}{\log x}\cdot\frac{\Delta(\overline{U})}{U\log U} + O\left(\frac{x\cdot x_3}{\log x}\right).$$

By using Lemma 0.5, it follows that

$$\sum_{\substack{U\in\wp\\ V\le U\le 2V}} |\Delta(\overline{U})| \le \frac{cV\log V}{\log V\cdot \kappa_V^2} + \frac{cV}{\log V}\cdot\kappa_V\cdot\log V = \frac{cV}{\kappa_V^2} + cV\kappa_V.$$

Thus,

$$\sum_{\substack{U\in\wp\\ V\le U\le 2V}} \frac{|\Delta(\overline{U})|}{U\log U} \le \frac{c}{\log V\cdot\kappa_V^2} + \frac{c\kappa_V}{\log^{3/2} V}. \tag{12.33}$$

Let us apply this with $V = V_j$ for $j = 0, 1, \ldots, j_0$, where $V_0 = \log x$, $V_j = 2^j V_0$, with $V_{j_0} \le x^{\varepsilon_x} < V_{j_0+1}$.

Thus, it follows from (12.33) that

$$\begin{aligned}\sum_{\substack{U\in\wp\\ \log x\le U\le x^{\varepsilon_x}}} \frac{|\Delta(\overline{U})|}{U\log U} &\le \frac{c}{\kappa_{V_0}^2}\sum_{j=0}^{j_0}\frac{1}{\log(V_0\cdot 2^j)} + c\kappa_{V_{j_0+1}}\sum_{j=0}^{j_0}\frac{1}{\log^{3/2} V_j}\\ &= W_1 + W_2,\end{aligned} \tag{12.34}$$

say. Since

$$W_1 \le \frac{c_1}{\kappa_{V_0}^2}\log j_0 \le \frac{c_1 x_2}{\kappa_{V_0}^2}$$

and noting that $\kappa_{V_0} \to \infty$ as $x\to\infty$, and since

$$W_2 \le c\kappa_x \sum_{j\ge 0}\frac{1}{(\log V_0 + j)^{3/2}} \le \frac{c_2\kappa_x}{x_2^{1/2}},$$

it follows from (12.34), that if we choose $\kappa_x \le \sqrt{x_2}$ say, then

$$\sum_{\substack{U\in\wp\\ \log x\le U\le x^{\varepsilon_x}}} \frac{|\Delta(\overline{U})|}{U\log U} = o(x_2),$$

which, in light of (12.31), proves (12.32) and thus completes the proof of Theorem 12.3.

Further remarks

Using the same approach, one can also prove the following two theorems.

Theorem 12.5. *Let $G(n) = n^2 + 1$ and set*

$$\begin{aligned}\xi_1 &= Concat(\overline{p(G(n))} : n \in \mathbb{N}),\\ \xi_2 &= Concat(\overline{p(G(\pi))} : \pi \in \wp).\end{aligned}$$

Then ξ_1 and ξ_2 are q-normal sequences.

We further let $p_k(n)$ stand for the k-th smallest prime factor of n, that is, if $n = q_1^{\alpha_1} \cdots q_r^{\alpha_r}$, where $q_1 < \cdots < q_r$ are primes and each α_i an integer, then

$$p_k(n) = \begin{cases} q_k & \text{if } k \leq r,\\ 1 & \text{if } k > r.\end{cases}$$

Theorem 12.6. *Let $G(x) = a_r x^r + a_{r-1}x^{r-1} + \cdots + a_0 \in \mathbb{Z}[x]$ be irreducible and satisfying $(a_r, a_{r-1}, \ldots, a_0) = 1$, $a_r > 0$ and $G(x) > 0$ for $x > x_0$. Then*

$$\eta_k = Concat\left(\overline{p_k(G(n))} : x_0 < n \in \mathbb{N}\right)$$

is a q-normal sequence.

Observe that the proof of Theorem 12.6 is very similar to that of Theorem 12.1. Indeed, we first define

$$\kappa_x := \text{Concat}\left(\overline{p_k(G(n))} : n \in I_x\right),$$

where $I_x = [\lfloor x/2 \rfloor + 1, \lfloor x \rfloor)$. Then, for each prime Q, we set

$$T(Q) := \#\{n \in I_x : p_k(G(n)) = Q\},$$

so that

$$\lambda(\kappa_x) = \sum_{n \in I_x} \lambda(\overline{p_k(G(n))}) = \sum_{Q \leq x} \lambda(\overline{Q})T(Q).$$

As can be shown using sieve methods, the main contribution to the above sum comes from those primes $Q \leq x^{1/2k}$, while that coming from the primes $Q > x^{1/2k}$ can be neglected. This allows us to establish that the order of $\lambda(\kappa_x)$ is $x(\log\log x)^k$.

Then, it is enough to prove that, given an arbitrary $t \in \mathbb{N}$ and any two words $\beta_1, \beta_2 \in \mathcal{A}_q^t$,

$$\frac{|\nu_{\beta_1}(\kappa_x) - \nu_{\beta_2}(\kappa_x)|}{\lambda(\kappa_x)} \to 0 \qquad \text{as } x \to \infty$$

and this is done by showing that

$$|\nu_{\beta_1}(\kappa_x) - \nu_{\beta_2}(\kappa_x)| = o(x(\log\log x)^k) \qquad \text{as } x \to \infty.$$

XIII. Complex roots of unity and normal numbers [26]

(Journal of Numbers, 2014)

Given an arbitrary prime number q, set $\xi = e^{2\pi i/q}$. We use a clever selection of the values of ξ^α, $\alpha = 1, 2, \ldots$, in order to create normal numbers. We also use a famous result of André Weil concerning Dirichlet characters to construct a family of normal numbers.

Let $\lambda(n)$ be the Liouville function (defined by $\lambda(n) := (-1)^{\Omega(n)}$ where $\Omega(n) := \sum_{p^\alpha \| n} \alpha$). It is well known that the statement "$\sum_{n\le x} \lambda(n) = o(x)$ as $x \to \infty$" is equivalent to the Prime Number Theorem. It is conjectured that if $b_1 < b_2 < \cdots < b_k$ are arbitrary positive integers, then $\sum_{n\le x} \lambda(n)\lambda(n+b_1)\cdots\lambda(n+b_k) = o(x)$ as $x \to \infty$. This conjecture seems presently out of reach since we cannot even prove that $\sum_{n\le x}\lambda(n)\lambda(n+1) = o(x)$ as $x \to \infty$.

The Liouville function belongs to a particular class of multiplicative functions, namely the class $\mathcal{M}^*$ of completely multiplicative functions. Recently, Indlekofer, Kátai and Klesov [47] considered a very special function $f \in \mathcal{M}^*$ constructed in the following manner. Let $\wp$ stand for the set of all primes. For each $q \in \wp$, let $C_q = \{\xi \in \mathbb{C} : \xi^q = 1\}$ be the group of complex roots of unity of order q. As p runs through the primes, let ξ_p be independent random variables distributed uniformly on C_q. Then, let $f \in \mathcal{M}^*$ be defined on $\wp$ by $f(p) = \xi_p$, so that $f(n)$ yields a random variable. In their 2011 paper, Indlekofer, Kátai and Klesov proved that, if $(\Omega, \mathcal{A}, \wp)$ stands for a probability space where ξ_p $(p \in \wp)$ are the independent random variables, then for almost all $\omega \in \Omega$, the sequence $\alpha = f(1)f(2)f(3)\ldots$ is a normal sequence over C_q (see Definition 13.1 below).

Let us now consider a somewhat different set up. Let $q \ge 2$ be a fixed prime number and set $\mathcal{A}_q := \{0, 1, \ldots, q-1\}$. Given an integer $t \ge 1$, an expression of the form $i_1 i_2 \ldots i_t$, where each $i_j \in \mathcal{A}_q$, is called a *word* of length t. We use the symbol Λ to denote the *empty word*. Then, $\mathcal{A}_q^t$ will stand for the set of words of length t over $\mathcal{A}_q$, while $\mathcal{A}_q^*$ will stand for the set of all words over $\mathcal{A}_q$ regardless of their length, including the empty word Λ. Similarly, we define C_q^* to be the set of words over C_q regardless of their length.

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

$$n = \varepsilon_0(n) + \varepsilon_1(n)q + \cdots + \varepsilon_t(n)q^t,$$

where $\varepsilon_i(n) \in \mathcal{A}_q$ for $0 \le i \le t$ and $\varepsilon_t(n) \ne 0$. To this representation, we associate the word

$$\overline{n} = \varepsilon_0(n)\varepsilon_1(n)\ldots\varepsilon_t(n) \in \mathcal{A}_q^{t+1}.$$

Definition 13.1. *Given a sequence of integers $a(1), a(2), a(3), \ldots$, we will say that the concatenation of their q-ary digit expansions $\overline{a(1)}\,\overline{a(2)}\,\overline{a(3)}\ldots$, denoted by Concat$(\overline{a(n)} : n \in \mathbb{N})$, is a normal sequence if the number $0.\overline{a(1)}\,\overline{a(2)}\,\overline{a(3)}\ldots$ is a q-normal number.*

It can be proved using a theorem of Halász (see [43]) that if $f \in \mathcal{M}^*$ is defined on the primes p by $f(p) = \xi_a$ $(a \ne 0)$, then $\sum_{n\le x} f(n) = o(x)$ as $x \to \infty$.

Now, given $u_0, u_1, \ldots, u_{\ell-1} \in \mathcal{A}_q$, let $Q(n) := \prod_{j=0}^{\ell-1}(n+j)^{u_j}$. We believe that if $\max_{j\in\{0,1,\ldots,\ell-1\}} u_j > 0$, then

$$\sum_{n\le x} f(Q(n)) = o(x) \qquad \text{as } x \to \infty. \tag{13.1}$$

If this were true, it would follow that

$$\mathrm{Concat}(f(n) : n \in \mathbb{N}) \quad \text{is a normal sequence over } C_q.$$

We cannot prove (13.1), but we can prove the following. Let $q \in \wp$ and set $\xi := e^{2\pi i/q}$. Further set $x_k = 2^k$ and $y_k = x_k^{1/\sqrt{k}}$ for $k = 1, 2, \ldots$ Then, consider the sequence of completely multiplicative functions f_k, $k = 1, 2, \ldots$, defined on the primes p by

$$f_k(p) = \begin{cases} \xi & \text{if } k \le p \le y_k, \\ 1 & \text{if } p < k \text{ or } p > y_k. \end{cases} \tag{13.2}$$

Then, set

$$\eta_k := f_k(x_k) f_k(x_k + 1) f_k(x_k + 2) \ldots f_k(x_{k+1} - 1) \qquad (k \in \mathbb{N})$$

and

$$\theta := \mathrm{Concat}(\eta_k : k \in \mathbb{N}).$$

Theorem 13.1. *The sequence θ is a normal sequence over C_q.*

We now use a famous result of André Weil to construct a large family of normal numbers.

Let q be a fixed prime and set $\xi := e^{2\pi i/q}$ and $\xi_a := e^{2\pi i a/q} = \xi^a$. Recall that C_q stands for the group of complex roots of unity of order q, that is,

$$C_q = \{\varsigma \in \mathbb{C} : \varsigma^q = 1\} = \{\xi^a : a = 0, 1, \ldots, q-1\}.$$

Let $p \in \wp$ be such that $q | p - 1$. Moreover, let χ_p be a Dirichlet character modulo p of order q, meaning that the smallest positive integer t for which $\chi_p^t = \chi_0$ is q. (Here χ_0 stands for the principal character.)

Let $u_0, u_1, \ldots, u_{k-1} \in \mathcal{A}_q$ and consider the polynomial

$$F(z) = F_{u_0,\ldots,u_{k-1}}(z) = \prod_{j=0}^{k-1} (z + j)^{u_j} \tag{13.3}$$

and assume that its degree is at least 1, that is, that there exists one $j \in \{0, \ldots, k-1\}$ for which $u_j \neq 0$. Further set

$$S_{u_0,\ldots,u_{k-1}}(\chi_p) = \sum_{n \pmod p} \chi_p\left(F_{u_0,\ldots,u_{k-1}}(n)\right).$$

According to a 1948 result of André Weil [64],

$$\left|S_{u_0,\ldots,u_{k-1}}(\chi_p)\right| \le (k-1)\sqrt{p}. \tag{13.4}$$

For a proof, see Proposition 12.11 (page 331) in the book of Iwaniec and Kowalski [48].

We can prove the following.

Theorem 13.2. *Let $p_1 < p_2 < \cdots$ be an infinite set of primes such that $q \mid p_j - 1$ for all $j \in \mathbb{N}$. For each $j \in \mathbb{N}$, let χ_{p_j} be a character modulo p_j of order q. Further set*

$$\Gamma_p = \chi_p(1)\chi_p(2)\ldots\chi_p(p-1) \qquad (p = p_1, p_2, \ldots)$$

and

$$\eta := \Gamma_{p_1}\Gamma_{p_2}\ldots \tag{13.5}$$

Then η is a normal sequence over C_q.

As an immediate consequence of this theorem, we have the following corollary.

Corollary 13.1. *Let $\varphi : C_q \to \mathcal{A}_q$ be defined by $\varphi(\xi_a) = a$. Extend the function φ to $\varphi : C_q^* \to \mathcal{A}_q^*$ by $\varphi(\alpha\beta) = \varphi(\alpha)\varphi(\beta)$ and let*

$$\varphi(\eta) = \varphi(\Gamma_{p_1})\varphi(\Gamma_{p_2})\ldots$$

and consider the q-ary expansion of the real number

$$\kappa = 0.\varphi(\Gamma_{p_1})\varphi(\Gamma_{p_2})\ldots \tag{13.6}$$

Then κ is a normal number in base q.

Example 13.1. *Choosing $q = 3$ and $\{p_1, p_2, p_3, \ldots\} = \{7, 13, 19, \ldots\}$ as the set of primes $p_j \equiv 1 \pmod 3$, then, the sequence η defined by (13.5) is normal sequence over $\{0, e^{2\pi i/3}, e^{4\pi i/3}\}$, while κ defined by (13.6) is a ternary normal number.*

XIV. The number of large prime factors of integers and normal numbers [27]

(Publications mathématiques de Besançon, 2015)

Letting $\omega(n)$ stand for the number of distinct prime factors of the positive integer n, we have shown in [25] (see paper XI above) that the concatenation of the successive values of $|\omega(n) - \lfloor \log\log n \rfloor|$ in a fixed base $q \geq 2$, as n runs through the integers $n \geq 3$, yields a normal number.

Given an integer $N \geq 1$, for each integer $n \in J_N := (e^N, e^{N+1})$, let $q_N(n)$ be the smallest prime factor of n which is larger than N; if no such prime factor exists, set $q_N(n) = 1$. Fix an integer $Q \geq 3$ and consider the function $f(n) = f_Q(n)$ defined by $f(n) = \ell$ if $n \equiv \ell \pmod Q$ with $(\ell, Q) = 1$ and by $f(n) = \Lambda$ otherwise, where Λ stands for the empty word. Then consider the sequence $(\kappa(n))_{n\geq 3} = (\kappa_Q(n))_{n\geq 3}$ defined by $\kappa(n) = f(q_N(n))$ if $n \in J_N$ with $q_N(n) > 1$ and by $\kappa(n) = \Lambda$ if $n \in J_N$ with $q_N(n) = 1$. Then, given an integer $N \geq 1$ and writing $J_N = \{j_1, j_2, j_3, \ldots\}$, consider the concatenation of the numbers $\kappa(j_1), \kappa(j_2), \kappa(j_3), \ldots$, that is define

$$\theta_N := \text{Concat}(\kappa(n) : n \in J_N) = 0.\kappa(j_1)\kappa(j_2)\kappa(j_3)\ldots.$$

Then, set $\alpha_Q := \text{Concat}(\theta_N : N = 1, 2, 3, \ldots)$ and let $B_Q = \{\ell_1, \ell_2, \ldots, \ell_{\phi(Q)}\}$ be the set of reduced residues modulo Q, where ϕ stands for the Euler function. In [23], we showed that α_Q is a normal sequence over B_Q, that is, the real number $0.\alpha_Q$ is a normal number over B_Q.

Here we prove the following. Let $q \geq 2$ be a fixed integer. Given an integer $n \geq n_0 = \max(q, 3)$, let N be the unique positive integer satisfying $q^N \leq n < q^{N+1}$ and let $h(n, q)$ stand for the residue modulo q of the number of distinct prime factors of n located in the interval $[\log N, N]$. Setting $x_N := e^N$, we then create a normal number in base q using the concatenation of the numbers $h(n, q)$, as n runs through the integers $\geq x_{n_0}$.

The main result

Theorem 14.1. *Let $q \geq 2$ be a fixed integer. Given an integer $n \geq n_0 = \max(q, 3)$, let N be the unique positive integer satisfying $q^N \leq n < q^{N+1}$ and let $h(n, q)$ stand for the residue modulo q of the number of distinct prime factors of n located in the interval $[\log N, N]$. For each integer $N \geq 1$, set $x_N := e^N$. Then, Concat$(h(n, q) : x_{n_0} \leq n \in \mathbb{N})$ is a q-ary normal sequence.*

Proof. For each integer $N \geq 1$, let $J_N = (x_N, x_{N+1})$. Further let S_N stand for the set of primes located in the interval $[\log N, N]$ and T_N for the product of the primes in S_N. Let $n_0 = \max(q, 3)$. Given a large integer N, consider the function

$$f(n) = f_N(n) = \sum_{\substack{p \mid n \\ \log N \leq p \leq N}} 1. \tag{14.1}$$

Let us further introduce the following sequences:

$$\begin{aligned} U_N &= \text{Concat}\left(h(n, q) : n \in J_N\right), \\ V_\infty &= \text{Concat}\left(U_N : N \geq n_0\right) = \text{Concat}\left(h(n, q) : n \geq x_{n_0}\right), \\ V_x &= \text{Concat}\left(h(n, q) : x_{n_0} \leq n \leq x\right). \end{aligned}$$

Let us set $\mathcal{A}_q := \{0, 1, \ldots, q-1\}$. If we fix an arbitrary integer r, it is sufficient to prove that given any particular word $w \in \mathcal{A}_q^r$, the number of occurrences $F_w(V_x)$ of w in V_x satisfies

$$F_w(V_x) = (1 + o(1)) \frac{x}{q^r} \qquad (x \to \infty). \tag{14.2}$$

For each integer $r \geq 1$, considering the polynomial

$$Q_r(u) = u(u+1) \cdots (u + r - 1).$$

and letting

$$\rho_r(d) = \#\{u \pmod d : Q_r(u) \equiv 0 \pmod d\},$$

it is clear that, since N is large,

$$\rho_r(p) = r \qquad \text{if } p \in S_N. \tag{14.3}$$

Observe that it follows from the Turán-Kubilius inequality that for some positive constant C,

$$\sum_{n\in J_N} (f(n) - \log\log N)^2 \le Ce^N \log\log N. \tag{14.4}$$

Letting $\varepsilon_N = 1/\log\log\log N$, it follows from (14.4) that

$$\frac{1}{x_N}\#\{n \in J_N : |f(n) - \log\log N| > \frac{1}{\varepsilon_N}\sqrt{\log\log N}\} \to 0 \qquad (\varepsilon_N \to 0). \tag{14.5}$$

This means that in the estimation of $F_w(V_x)$, we may ignore those integers n appearing in the concatenation $h(2,q)h(3,q)\dots h(\lfloor x\rfloor,q)$ for which the corresponding $f(n)$ is "far" from $\log\log N$ in the sense described in (14.5).

Let X be a large number. Then there exists a large integer N such that $\frac{X}{e} < x_N \le X$. Letting $\mathscr{L} = \left]\frac{X}{e}, X\right]$, we write

$$\mathscr{L} = \left]\frac{X}{e}, x_N\right] \cup \left]x_N, X\right] = \mathscr{L}_1 \cup \mathscr{L}_2,$$

say, and $\lambda(\mathscr{L}_i)$ for the length of the interval $\mathscr{L}_i$ for $i = 1, 2$.

Given an arbitrary function δ_N which tends to 0 arbitrarily slowly, it is sufficient to consider those $\mathscr{L}_1$ and $\mathscr{L}_2$ such that

$$\lambda(\mathscr{L}_1) \ge \delta_N X \quad \text{and} \quad \lambda(\mathscr{L}_2) \ge \delta_N X. \tag{14.6}$$

The reason for this is that those $n \in \mathscr{L}_1$ (resp. $n \in \mathscr{L}_2$) for which $\lambda(\mathscr{L}_1) < \delta_N X$ (resp. $\lambda(\mathscr{L}_2) < \delta_N X$) are $o(x)$ in number and can therefore be ignored in the proof of (14.2).

Let us first consider the set $\mathscr{L}_2$. We start by observing that any subword taken in the concatenation $h(n,q)h(n+1,q)\dots h(n+r-1,q)$ is made of co-prime divisors of T_N (since no two members of the sequence $h(n,q), h(n+1,q), \dots, h(n+r-1,q)$ of r elements may have a common prime divisor $p > \log N$). So, let $d_0, d_1, \dots, d_{r-1}$ be co-prime divisors of T_N and let $B_N(\mathscr{L}_2; d_0, d_1, \dots, d_{r-1})$ stand for the number of those $n \in \mathscr{L}_2$ for which $d_j \mid n+j$ for $j = 0, 1, \dots, r-1$ and such that $\left(Q_r(n), \frac{T_N}{d_0 d_1 \cdots d_{r-1}}\right) = 1$. We can assume that each of the d_j's is squarefree, since the number of those $n + j \le X$ for which $p^2 \mid n+j$ for some $p > \log N$ is $\ll X \sum_{p>\log N} \frac{1}{p^2} = o(X)$.

In light of (14.4), we may assume that

$$\omega(d_j) \le 2\log\log N \qquad \text{for } j = 0, 1, \dots, r-1. \tag{14.7}$$

By using the Eratosthenian sieve (see for instance the book of De Koninck and Luca [34]) and recalling that condition (14.6) ensures that $X - x_N$ is large, we obtain that, as $N \to \infty$,

$$B_N(\mathscr{L}_2; d_0, d_1, \dots, d_{r-1}) = \frac{X - x_N}{d_0 d_1 \cdots d_{r-1}} \prod_{p \mid T_N/(d_0 d_1 \cdots d_{r-1})} \left(1 - \frac{r}{p}\right)$$

$$+o\left(\frac{x_N}{d_0d_1\cdots d_{r-1}}\prod_{p|T_N/(d_0d_1\cdots d_{r-1})}\left(1-\frac{r}{p}\right)\right). \tag{14.8}$$

Letting $\theta_N := \prod_{p|T_N}\left(1-\frac{r}{p}\right)$, one can easily see that

$$\theta_N = (1+o(1))\frac{(\log\log N)^r}{(\log N)^r} \qquad (N\to\infty). \tag{14.9}$$

Let us also introduce the strongly multiplicative function κ defined on primes p by $\kappa(p) = p-r$. Then, (14.8) can be written as

$$B_N(\mathscr{L}_2; d_0, d_1, \ldots, d_{r-1}) = \frac{X-x_N}{\kappa(d_0)\kappa(d_1)\cdots\kappa(d_{r-1})}\theta_N + o\left(\frac{x_N}{\kappa(d_0)\kappa(d_1)\cdots\kappa(d_{r-1})}\theta_N\right) \tag{14.10}$$

as $N\to\infty$. For each integer $N > e^e$, let

$$R_N := \left[\log\log N - \frac{\sqrt{\log\log N}}{\varepsilon_N}, \log\log N + \frac{\sqrt{\log\log N}}{\varepsilon_N}\right].$$

Let $\ell_0, \ell_1, \ldots, \ell_{r-1}$ be an arbitrary collection of non negative integers $< q$. Note that there are q^r such collections. Our goal is to count how many times, amongst the integers $n \in \mathscr{L}_2$, we have $f(n+j) \equiv \ell_j \pmod q$ for $j = 0, 1, \ldots, r-1$. In light of (14.5), we only need to consider those $n \in \mathscr{L}_2$ for which

$$f(n+j) \in R_N \qquad (j = 0, 1, \ldots, r-1).$$

Let

$$\mathscr{S}(\ell_0, \ell_1, \ldots, \ell_{r-1}) := \sideset{}{^*}\sum_{\substack{f(d_j)\equiv \ell_j \pmod q\\ d_j|T_N\\ j=0,1,\ldots,r-1}} \frac{1}{\kappa(d_0)\kappa(d_1)\cdots\kappa(d_{r-1})}, \tag{14.11}$$

where the star over the sum indicates that the summation runs only on those d_j satisfying $f(d_j) \in R_N$ for $j = 0, 1, \ldots, r-1$.

From (14.10), we therefore obtain that

$$\begin{aligned}&\#\{n\in\mathscr{L}_2 : f(n+j)\equiv \ell_j \pmod q,\ j=0,1,\ldots,r-1\}\\ &\qquad = (X-x_N)\theta_N\mathscr{S}(\ell_0,\ell_1,\ldots,\ell_{r-1}) + o\left(x_N\theta_N\mathscr{S}(\ell_0,\ell_1,\ldots,\ell_{r-1})\right)\end{aligned} \tag{14.12}$$

as $N\to\infty$. Let us now introduce the function

$$\eta = \eta_N = \sum_{p|T_N}\frac{1}{\kappa(p)}.$$

Observe that, as $N\to\infty$,

$$\eta \ = \ \sum_{\log N\le p\le N}\frac{1}{p(1-r/p)} = \sum_{\log N\le p\le N}\frac{1}{p} + O\left(\sum_{\log N\le p\le N}\frac{1}{p^2}\right)$$

$$= \log\log N - \log\log\log N + o(1) + O\left(\frac{1}{\log N}\right)$$

$$= \log\log N - \log\log\log N + o(1). \tag{14.13}$$

From the definition (14.11), one easily sees that

$$\mathscr{S}(\ell_0, \ell_1, \ldots, \ell_{r-1}) = (1+o(1)) \sum_{\substack{t_j \equiv \ell_j \pmod q \\ t_j \in R_N}} \frac{\eta^{t_0+t_1+\cdots+t_{r-1}}}{t_0! t_1! \cdots t_{r-1}!} \qquad (N \to \infty), \tag{14.14}$$

where we ignore in the denominator of the summands the factors $\kappa(p)^a$ (with $a \geq 2$) since their contribution is negligible.

Moreover, for $t \in R_N$, one can easily establish that

$$\frac{\eta^{t+1}}{(t+1)!} = (1+o(1))\frac{\eta^t}{t!} \qquad (N \to \infty)$$

and consequently that, for each $j \in \{0, 1, \ldots, r-1\}$,

$$\sum_{\substack{t_j \equiv \ell_j \pmod q \\ t_j \in R_N}} \frac{\eta^{t_j}}{t_j!} = (1+o(1))\frac{1}{q}\sum_{t \in R_N} \frac{\eta^t}{t!} = (1+o(1))\frac{e^\eta}{q} \qquad (N \to \infty). \tag{14.15}$$

Using (14.15) in (14.14), we obtain that

$$\mathscr{S}(\ell_0, \ell_1, \ldots, \ell_{r-1}) = (1+o(1))\frac{e^{\eta r}}{q^r} \qquad (N \to \infty). \tag{14.16}$$

Combining (14.12) and (14.16), we obtain that

$$\#\{n \in \mathscr{L}_2 : f(n+j) \equiv \ell_j \pmod q,\ j = 0, 1, \ldots, r-1\}$$

$$= (X - x_N)\theta_N \frac{e^{\eta r}}{q^r} + o\left(x_N \theta_N \frac{e^{\eta r}}{q^r}\right)$$

$$= \frac{X - x_N}{q^r} + o\left(x_N \frac{1}{q^r}\right) \qquad (N \to \infty), \tag{14.17}$$

where we used (14.9) and (14.13).

Since the first term on the right hand side of (14.17) does not depend on the particular collection $\ell_0, \ell_1, \ldots, \ell_{r-1}$, we may conclude that the frequency of those integers $n \in \mathscr{L}_2$ for which $f(n+j) \equiv \ell_j \pmod q$ for $j = 0, 1, \ldots, r-1$ is the same independently of the choice of $\ell_0, \ell_1, \ldots, \ell_{r-1}$.

The case of those $n \in \mathscr{L}_1$ can be handled in a similar way.

We have thus shown that the number of occurrences of any word $w \in \mathcal{A}_q^r$ in $h(n,q)h(n+1,q)\ldots h(n+r-1,q)$ as n runs over the $\lfloor X - X/e \rfloor$ elements of $\mathscr{L}$ is $(1+o(1))\dfrac{(X - X/e)}{q^r}$. Repeating this for each of the intervals

$$\left]\frac{X}{e^{j+1}}, \frac{X}{e^j}\right] \qquad (j = 0, 1, \ldots, \lfloor \log x \rfloor),$$

we obtain that the number of occurrences of w for $n \le x$ is $(1+o(1))\dfrac{x}{q^r}$, as claimed.

The proof of (14.2) is thus complete and the Theorem is proved.

□

Final remarks

First of all, let us first mention that our main result can most likely be generalized in order that the following statement will be true:

> Let $a(n)$ and $b(n)$ be two monotonically increasing sequences of n for $n = 1, 2, \ldots$ such that $n/b(n)$, $b(n)/a(n)$ and $a(n)$ all tend to infinity monotonically as $n \to \infty$. Let $f(n)$ stand for the number of prime divisors of n located in the interval $[a(n), b(n)]$ and let $h(n, q)$ be the residue of $f(n)$ modulo q; then, the sequence $h(n, q)$, $n = 1, 2, ...$, is a q-ary normal sequence.

Secondly, let us first recall that it was proven by Pillai [55] (with a more general result by Delange [36]) that the values of $\omega(n)$ are equally distributed over the residue classes modulo q for every integer $q \ge 2$, and that the same holds for the function $\Omega(n)$, where $\Omega(n) := \sum_{p^\alpha \| n} \alpha$. We believe that each of the sequences $\mathrm{Concat}(\omega(n) \pmod q : n \in \mathbb{N})$ and $\mathrm{Concat}(\Omega(n) \pmod q : n \in \mathbb{N})$ represents a normal sequence for each base $q = 2, 3, \ldots$. However, the proof of these statements could be very difficult to obtain. Indeed, in the particular case $q = 2$, such a result would imply the famous Chowla conjecture

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n \le x} \lambda(n)\lambda(n+a_1)\cdots\lambda(n+a_k) = 0,$$

where $\lambda(n) := (-1)^{\Omega(n)}$ is known as the Liouville function and where $a_1, a_2, \ldots, a_k$ are k distinct positive integers (see Chowla [9]).

Thirdly, we had previously conjectured that, given any integer $q \ge 2$ and letting $\mathrm{res}_q(n)$ stand for the residue of n modulo q, it may not be possible to create an infinite sequence of positive integers $n_1 < n_2 < \cdots$ such that

$$0.\mathrm{Concat}(\mathrm{res}_q(n_j) : j = 1, 2, \ldots)$$

is a q-normal number. However, we now have succeeded in creating such a monotonic sequence. It goes as follows. Let us define the sequence $(m_k)_{k\ge1}$ by

$$m_k = f(k) + k!,$$

where f is the function defined by

$$f(n) = f_N(n) = \sum_{\substack{p|n \\ \log N \le p \le N}} 1.$$

In this case, we obtain that

$$m_{k+1} - m_k = k! \cdot k + f(k+1) - f(k),$$

a quantity which is positive for all integers $k \geq 1$ provided

$$f(k+1) - f(k) > -k! \cdot k, \tag{14.18}$$

that is if

$$f(k) < k! \cdot k. \tag{14.19}$$

But since we trivially have

$$f(k) \leq \omega(k) \leq 2 \log k \leq k! \cdot k,$$

then (14.19) follows and therefore (14.18) as well.

Hence, in light of Theorem 14.1, if we choose $n_k = m_k$, our conjecture is disproved.

XV. Multidimensional sequences uniformly distributed modulo 1 created from normal numbers [28]

(Contemporary Mathematics, Vol. 655, AMS, 2015)

Recall that if α is an irrational number, then the sequence $(\alpha n)_{n\geq 1}$ is uniformly distributed modulo 1 (see for instance Example 2.1 in the book of Kuipers and Neiderreiter [50]). Here, given a prime number $q \geq 3$, we construct an infinite sequence of normal numbers in base $q-1$ which, for any fixed positive integer r, yields an r-dimensional sequence which is uniformly distributed on $[0,1)^r$. More precisely, our main result consists in creating an infinite sequence $\alpha_1, \alpha_2, \ldots$ of normal numbers in base $q-1$ such that, for any fixed positive integer r, the r-dimensional sequence $(\{\alpha_1(q-1)^n\}, \ldots, \{\alpha_r(q-1)^n\})$ is uniformly distributed on $[0,1)^r$, where as usual $\{y\}$ stands for the fractional part of y.

Fix a positive integer r. For each integer $j \in \{1, \ldots, r\}$, write the $(q-1)$-ary expansion of each α_j as

$$\alpha_j = 0.a_{j,1}a_{j,2}a_{j,3}\ldots$$

To prove our claim we only need to prove that for every positive integer k and arbitrary integers $b_{j,\ell} \in \mathcal{A}_{q-1} := \{0, 1, \ldots, q-2\}$ (for $1 \leq j \leq r$, $1 \leq \ell \leq k$), the proportion of those positive integers $n \leq x$ for which $a_{j,n+\ell} = b_{j,\ell}$ simultaneously for $j = 1, \ldots, r$ and $\ell = 1, \ldots, k$ is asymptotically equal to $1/(q-1)^{kr}$.

To do so, we first construct the proper set up. For each positive integer N, consider the semi-open interval $J_N := [x_N, x_{N+1})$, where $x_N = e^N$. For each integer $N > e^e$, we introduce the expression $\lambda_N = \log\log N$ and consider the corresponding interval $K_N := [N, N^{\lambda_N}]$. Given an integer $n \in J_N$, we define the function $q_N(n)$ as the smallest prime factor of n which belongs to K_N, while we let $q_N(n) = 1$ if $(n,p) = 1$ for all primes $p \in K_N$.

Further let $\pi_1 \leq \pi_2 \leq \cdots \leq \pi_{h(n)}$ be the prime factors of n which belong to K_N (written with multiplicity). With this definition, we clearly have $(n/\pi_1 \cdots \pi_{h(n)}, p) = 1$ for each prime $p \in K_N$.

For each positive integer i and each $n \in K_N$, we let

$$q_N^{(i)}(n) = \begin{cases} \pi_i & \text{if} \quad 1 \le i \le h(n), \\ 1 & \text{if} \quad i > h(n), \end{cases}$$

so that in particular $q_N^{(1)}(n) = q_N(n)$.

We further set

$$f_q(m) = \begin{cases} \ell - 1 & \text{if} \quad m \equiv \ell \pmod q \text{ and } \ell \neq 0, \\ \Lambda & \text{if} \quad q \mid m. \end{cases}$$

Let r and k be fixed positive integers. Let $Q_{i,\ell}$, for $i = 1, \dots, r$ and $\ell = 1, \dots, k$ be distinct primes belonging to K_N such that $Q_{1,\ell} < Q_{2,\ell} < \cdots < Q_{r,\ell}$. For a given interval $J = [x, x+y] \subseteq J_N$, where $y > x_N$, we let $S_J(Q_{i,\ell} \mid i = 1, \dots, r,\ \ell = 1, \dots, k)$ be the number of those integers $n \in J$ for which $q_N^{(i)}(n + \ell) = Q_{i,\ell}$.

For each integer $r \ge 1$, let $\sigma(1), \dots, \sigma(k)$ be the permutation of the set $\{1, \dots, k\}$ which allows us to write

$$Q_{r,\sigma(1)} < Q_{r,\sigma(2)} < \cdots < Q_{r,\sigma(k)}.$$

Using the Eratosthenian sieve, we obtain that, as $N \to \infty$,

$$\begin{aligned} & S_J(Q_{i,\ell} \mid i = 1, \dots, r,\ \ell = 1, \dots, k) \\ & \qquad = (1 + o(1)) \frac{y}{\prod\limits_{\substack{1 \le i \le r \\ 1 \le \ell \le k}} Q_{i,\ell}} \cdot \prod_{N \le \pi < Q_{r,\sigma(k)}} \left(1 - \frac{\rho(\pi)}{\pi}\right), \end{aligned} \tag{15.1}$$

where

$$\rho(\pi) = \begin{cases} k & \text{if} \quad N \le \pi < Q_{r,\sigma(1)}, \\ k-1 & \text{if} \quad Q_{r,\sigma(1)} < \pi < Q_{r,\sigma(2)}, \\ \vdots & \qquad \vdots \\ 1 & \text{if} \quad Q_{r,\sigma(k-1)} < \pi < Q_{r,\sigma(k)}, \\ 0 & \text{if} \quad \pi \in \{Q_{i,\ell} : i = 1, \dots, r,\ \ell = 1, \dots, k\}. \end{cases}$$

Let $t_{i,\ell}$ $(i = 1, \dots, r,\ \ell = 1, \dots, k)$ be any collection of the (non zero) reduced residues modulo q and set

$$B_J(t_{i,\ell} \mid i = 1, \dots, r,\ \ell = 1, \dots, k) := \sum_{\substack{Q_{i,\ell} \equiv t_{i,\ell} \pmod q \\ N \le Q_{i,\ell} < N^{\lambda_N}}} S_J(Q_{i,\ell} \mid i = 1, \dots, r,\ \ell = 1, \dots, k). \tag{15.2}$$

Now, letting $\pi(x; k, \ell)$ stand for the number of primes $p \le x$ such that $p \equiv \ell \pmod k$, it follows from the Prime Number Theorem in arithmetical progressions that, with $2 \le v \le u$, as $u \to \infty$,

$$\pi(u + v; q, \ell) - \pi(u; q, \ell) = (1 + o(1)) \frac{1}{q-1} \left(\pi(u+v) - \pi(u)\right) + O\left(\frac{u}{\log^{10} u}\right),$$

from which we obtain that

$$\sum_{\substack{u<p\le u+v\\ p\equiv \ell \pmod q}} \frac{1}{p\log p} = (1+o(1))\frac{1}{q-1}\frac{1}{\log u}\log\frac{\log(u+v)}{\log u} + O\left(\frac{1}{\log^{10} u}\right) \tag{15.3}$$

and

$$\sum_{\substack{u<p\le u+v\\ p\equiv \ell \pmod q}} \frac{1}{p} = (1+o(1))\frac{1}{q-1}\log\frac{\log(u+v)}{\log u} + O\left(\frac{1}{\log^{10} u}\right) \tag{15.4}$$

Substituting (15.3) and (15.4) in (15.1), we obtain

$$\begin{aligned} &S_J(Q_{i,\ell} \mid i=1,\dots,r,\ \ell=1,\dots,k)\\ &\quad = (1+o(1))\frac{y}{\prod_{1\le i\le r, 1\le \ell\le k} Q_{i,\ell}}\exp\{k\log\log N - k\log\log Q_{r,\sigma(1)}\\ &\qquad\qquad -(k-1)\log\log Q_{r,\sigma(2)} + (k-1)\log\log Q_{r,\sigma(1)} - \ldots - \log\log Q_{r,\sigma(k)}\}\\ &\quad = (1+o(1))\frac{y}{\prod_{1\le i\le r, 1\le \ell\le k} Q_{i,\ell}}\prod_{\ell=1}^{k}\frac{\log N}{\log Q_{r,\ell}} \qquad (y\to\infty). \end{aligned} \tag{15.5}$$

Using (15.5) and definition (15.2), we obtain that, as $y\to\infty$,

$$B_J(t_{i,\ell} \mid i=1,\dots,r,\ \ell=1,\dots,k) = (1+o(1))\frac{y}{(q-1)^{kr}}\sum_{\pi_{i,\ell}}\frac{1}{\prod \pi_{i,\ell}}\prod_{\ell=1}^{k}\frac{\log N}{\log \pi_{r,\ell}}, \tag{15.6}$$

where the summation runs over those subsets of primes $\pi_{i,\ell}$ for which

$$N<\pi_{1,\ell}<\pi_{2,\ell}<\cdots<\pi_{r,\ell}<N^{\lambda_N} \qquad (\ell=1,\dots,k).$$

Now, observe that, as $N\to\infty$,

$$\begin{aligned} &\sum_{N<\pi_{1,\ell}<\cdots<\pi_{r-1,\ell}<\pi_{r,\ell}<N^{\lambda_N}} \frac{1}{\pi_{1,\ell}\cdots\pi_{r-1,\ell}}\cdot\frac{1}{\pi_{r,\ell}\log\pi_{r,\ell}}\\ &\quad = (1+o(1))\sum_{\pi_{r,\ell}}\frac{1}{(r-1)!}\left(\sum_{N<\pi<\pi_{r,\ell}}\frac{1}{\pi}\right)^{r-1}\cdot\frac{1}{\pi_{r,\ell}\log\pi_{r,\ell}}\\ &\quad = (1+o(1))\sum_{\pi_{r,\ell}}\frac{1}{(r-1)!}\left(\log\left(\frac{\log\pi_{r,\ell}}{\log N}\right)\right)^{r-1}\cdot\frac{1}{\pi_{r,\ell}\log\pi_{r,\ell}}\\ &\quad = (1+o(1))\int_N^{N^{\lambda_N}}\frac{1}{(r-1)!}\left(\log\left(\frac{\log u}{\log N}\right)\right)^{r-1}\frac{du}{u\log^2 u}\\ &\quad = (1+o(1))\int_{\log N}^{\lambda_N\log N}\frac{1}{(r-1)!}\left(\log\left(\frac{v}{\log N}\right)\right)^{r-1}\frac{dv}{v^2}. \end{aligned} \tag{15.7}$$

Setting $v = y \log N$ in this last integral, we obtain that the above expression can be replaced by

$$\frac{(1+o(1))}{\log N}\int_1^{\lambda_N}\frac{1}{(r-1)!}\frac{(\log y)^{r-1}}{y^2}\,dy = \frac{(1+o(1))}{\log N}\frac{1}{(r-1)!}\int_1^{\infty}\frac{(\log y)^{r-1}}{y^2}\,dy,$$

which in turn, after setting $z = \log y$, becomes

$$\frac{(1+o(1))}{\log N}\int_0^{\infty}\frac{e^{-z}z^{r-1}}{(r-1)!}\,dz = \frac{(1+o(1))}{\log N},$$

which substituted in (15.7) yields

$$\sum_{N<\pi_{1,\ell}<\cdots<\pi_{r-1,\ell}<\pi_{r,\ell}<N^{\lambda_N}}\frac{1}{\pi_{1,\ell}\cdots\pi_{r-1,\ell}}\cdot\frac{1}{\pi_{r,\ell}\log\pi_{r,\ell}} = \frac{(1+o(1))}{\log N}\qquad (N\to\infty). \tag{15.8}$$

Using (15.8) in (15.6), we obtain that

$$B_J(t_{i,\ell}\mid i=1,\ldots,r,\ \ell=1,\ldots,k) = (1+o(1))\frac{y}{(q-1)^{kr}}\qquad (y\to\infty). \tag{15.9}$$

We now define, for each integer $N\in\mathbb{N}$,

$$\theta_N^{(i)} = \text{Concat}\{f_q(q_N^{(i)}(n)) : n\in J_N\}\qquad (i=1,2,\ldots).$$

Then consider the number

$$\theta^{(i)} = \theta_1^{(i)}\theta_2^{(i)}\ldots$$

and from these numbers, introduce the number

$$\alpha_i := 0.\theta^{(i)},$$

that is the number whose q-ary expansion is $0.\theta^{(i)}$.

Recall that, for $n\in J_N$, we defined $h(n)$ as the number of prime divisors of n located in the interval $[N, N^{\lambda_N}]$. Thus, setting

$$U_N := \sum_{N<p<N^{\lambda_N}}\frac{1}{p} = \log\lambda_N + o(1)\qquad (N\to\infty),$$

we obtain, using the Turán-Kubilius inequality, that for some absolute constant $c>0$,

$$\sum_{n\in J_N}(h(n)-U_N)^2 \le c x_N \log\lambda_N. \tag{15.10}$$

On the one hand, it follows from (15.10) that for each integer $r\ge 1$, there exists a constant $c_r>0$ such that

$$\#\{n\in J_N : h(n)\le r\} \le \frac{c_r x_N}{\log\lambda_N}. \tag{15.11}$$

On the other hand, it is easy to see that, as $y \to \infty$,

$$\#\{n \in J_N : p^2 | n \text{ for some prime } p > N\} \le c x_N \sum_{p>N} \frac{1}{p^2} = O\left(\frac{x_N}{N}\right). \tag{15.12}$$

We therefore have, in light of (15.9), keeping in mind (15.11) and (15.12), that, as $y \to \infty$ (and thus as $N \to \infty$),

(15.13)

$$\#\{n \in J : f_q(q_N^{(i)}(n+\ell)) = t_{i,\ell} - 1 : i = 1, \ldots, r,\ \ell = 1, \ldots, k\} = (1+o(1))\frac{y}{(q-1)^{kr}} + o(x_N).$$

Now, to prove the normality of α_i in base $q-1$, we need to estimate the quantity

$$H(x) := \#\{n \le x : f_q(q_N^{(i)}(n+\ell)) = t_{i,\ell} - 1 : i = 1, \ldots, r,\ \ell = 1, \ldots, k\}.$$

For this, let us set

$$K_N := \#\{n \in J_N : f_q(q_N^{(i)}(n+\ell)) = t_{i,\ell} - 1 : i = 1, \ldots, r,\ \ell = 1, \ldots, k\}.$$

Let x be a large number. Then, $x \in J_{N_0}$ for some N_0. Hence, applying (15.13), we get

$$\begin{aligned} H(x) &= O(1) + K_3 + K_4 + \cdots + K_{N_0-1} \\ &\qquad + \#\{J_{N_0-1} \le x : f_q(q_{N_0-1}^{(i)}(n+\ell)) = t_{i,\ell} - 1 : i = 1, \ldots, r,\ \ell = 1, \ldots, k\} \\ &= \frac{(1+o(1))}{(q-1)^{kr}}\left((x_2 - x_1) + (x_3 - x_2) + \cdots + (x_{N_0} - x_{N_0-1}) + (x - x_{N_0})\right) + O(1) \\ &= (1+o(1))\frac{x - x_1}{(q-1)^{kr}} = (1+o(1))\frac{x}{(q-1)^{kr}}, \end{aligned}$$

thus completing the proof of our main result.

XVI. On sharp normality [31]

(Uniform Distribution Theory, 2016)

In this paper[3], we identify a very special family of normal numbers – that we will call *sharp normal numbers* – which are connected with arithmetical functions that have a local

[3]Our original paper on sharp normality appeared in *Uniform Distribution Theory* under the title *On strong normality*. After its publication, we became aware that the term "strongly normal" had been used by other authors with a different meaning. For instance, Adrian Belshaw and Peter Borwein [5] call α strongly normal in base b if every string of digits in the base b expansion of α appears with the frequency expected for random digits and the discrepancy fluctuates as is expected by the law of the iterated logarithm. With this concept of "strong normality", they then showed that almost all numbers are strongly normal (as we do in the present document, but for different reasons). This being said, in order to avoid confusion, in this survey and in other papers in which we will further expand on properties regarding this new concept, we shall always talk about "sharp normal numbers".

normal distribution, such as the function $\omega(n)$ which counts the number of distinct prime factors of n.

Let us first recall some definitions already given on Page 2.

A sequence $(x_n)_{n\in\mathbb{N}}$ of real numbers is said to be *uniformly distributed modulo 1* (or *mod 1*) if for every interval $[a,b) \subseteq [0,1)$,

$$\lim_{N\to\infty} \frac{1}{N}\#\{n \le N : \{x_n\} \in [a,b)\} = b-a.$$

In other words, a sequence of real numbers is said to be uniformly distributed mod 1 if every subinterval of the unit interval gets its fair share of the fractional parts of the elements of this sequence.

Recall also that, given a set of N real numbers $x_1,\ldots,x_N$, the *discrepancy* of this set is defined as the quantity

$$D(x_1,\ldots,x_N) := \sup_{[a,b)\subseteq[0,1)} \left| \frac{1}{N} \sum_{\substack{n\le N \\ \{x_n\}\in[a,b)}} 1 - (b-a) \right|.$$

It is known that a sequence $(x_n)_{n\in\mathbb{N}}$ of real numbers is uniformly distributed mod 1 if and only if $D(x_1,\ldots,x_N) \to 0$ as $N\to\infty$ (see Theorem 1.1 in the book of Kuipers and Niederreiter [50]).

Also, given an integer $q \ge 2$, it can be shown (see Theorem 8.1 in the book of Kuipers and Niederreiter [50]) that a real number α is normal in base q if and only if the sequence $(\{q^n\alpha\})_{n\in\mathbb{N}}$ is uniformly distributed mod 1.

We are now ready to introduce the concept of *sharp normality*. For each positive integer N, let

$$M = M_N := \lfloor \delta_N \sqrt{N} \rfloor, \text{ where } \delta_N \to 0 \text{ and } \delta_N \log N \to \infty \text{ as } N\to\infty. \tag{16.1}$$

We shall say that an infinite sequence of real numbers $(x_n)_{n\ge1}$ is *sharply uniformly distributed* mod 1 if

$$D(x_{N+1},\ldots,x_{N+M}) \to 0 \quad \text{as} \quad N\to\infty$$

for every choice of δ_N satisfying (16.1).

Remark 16.1. *Observe that if a sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ is sharply uniformly distributed mod 1, then it must be uniformly distributed mod 1 as well. The proof goes as follows. Assume that $(x_n)_{n\in\mathbb{N}}$ is sharply uniformly distributed mod 1 and define the sequence $(\epsilon_k)_{k\in\mathbb{N}}$ by*

$$\epsilon_k = \begin{cases} 1 & \textit{if} \quad k \le e, \\ 1/\log k & \textit{if} \quad k > e. \end{cases}$$

Also, for each integer $k \ge 1$, let $U_k = \lfloor k^2\epsilon_k \rfloor$ and $V_k = U_{k+1} - U_k - 1$. Moreover, setting $N = U_k$ and $M = M_N = V_k$, one can verify that (16.1) is satisfied as $k\to\infty$. To see this, observe that

$$V_k = (k+1)^2\epsilon_{k+1} - k^2\epsilon_k + O(1) = 2k\epsilon_{k+1} + k^2(\epsilon_{k+1}-\epsilon_k) + O(1)$$

$$(16.2) \qquad = 2k\epsilon_{k+1} + O\left(\frac{k}{\log^2 k}\right) = (1+o(1))2k\epsilon_k \qquad \textit{as } k\to\infty.$$

Now, for each $k\in\mathbb{N}$, define δ_{U_k} implicitly by $V_k = \lfloor \delta_{U_k}\sqrt{U_k}\rfloor$. Using this in (16.2), it follows that

$$2k\epsilon_k(1+o(1)) = \delta_{U_k} k\sqrt{\epsilon_k}(1+o(1)) \qquad (k\to\infty),$$

from which we obtain that

$$\delta_{U_k} = (1+o(1))2\sqrt{\epsilon_k} \qquad (k\to\infty).$$

Hence, it follows that

$$\delta_N = \delta_{U_k} \to 0 \textit{ and } \delta_N \log N = (1+o(1))2\sqrt{\epsilon_k}\log U_k = (1+o(1))4\sqrt{\log k} \to \infty \quad (k\to\infty),$$

implying that condition (16.1) is satisfied and also, using the fact that $(x_n)_{n\in\mathbb{N}}$ is sharply uniformly distributed mod 1, that

$$(16.3) \qquad D(x_{U_k},\dots,x_{U_{k+1}-1}) = D(x_N,\dots,x_{N+M}) \to 0 \qquad (k\to\infty).$$

We shall now use this result to prove that

$$(16.4) \qquad D(x_1,\dots,x_N) \to 0 \qquad (N\to\infty).$$

To do so, for each $N\in\mathbb{N}$, let t_N be the unique integer k for which $U_k \le N < U_{k+1}$, from which it follows that

$$(16.5) \qquad \frac{N-U_{t_N}}{N} \le \frac{U_{t_N+1}-U_{t_N}}{N} \to 0 \qquad (N\to\infty).$$

With this set up, we have

$$(16.6) \qquad ND(x_1,\dots,x_N) \le \sum_{\ell=1}^{t_N-1}(U_{\ell+1}-U_\ell)D(x_{U_\ell},\dots,x_{U_{\ell+1}-1}) + (N-U_{t_N}).$$

Applying (16.3) successively with $k=\ell$ for $\ell=1,\dots,t_N-1$, it follows, in light of (16.5), that the right hand side of (16.6) is $o(N)$ as $N\to\infty$. From this, (16.4) follows immediately, thus proving our claim.

Remark 16.2. *It follows from the above that if α is a sharp normal number, then it must also be a normal number. Indeed, by definition, the sequence $(\{\alpha q^n\})_{n\in\mathbb{N}}$ is sharply uniformly distributed mod 1 and therefore, in light of Remark 16.1, it must then be uniformly distributed mod 1, which in turn (as we saw above) is equivalent to the statement that α is a normal number.*

Given a fixed integer $q\ge 2$, we say that an irrational number α is a *sharp normal number* in base q (or a *sharp q-normal number*) if the sequence $(x_n)_{n\in\mathbb{N}}$, defined by $x_n = \{q^n\alpha\}$, is sharply uniformly distributed mod 1. First, observe that there exist normal numbers which are not sharp normal. For instance, consider the Champernowne number

$$\theta := 0.1\,10\,11\,100\,101\,110\,111\,1000\,1001\,1010\,1011\,1100\,1101\,1110\,1111\,\dots$$

that is the number made up of the concatenation of the positive integers written in base 2. It is known since Champernowne [8] that θ is normal. However, one can show that θ is not a sharp normal number. Indeed, given a positive integer n, let $S_n = \lfloor 2^n/(\sqrt{n}\log n)\rfloor$ and consider the sequence

$$2^{2n}+1, 2^{2n}+2, 2^{2n}+3, \ldots, 2^{2n}+S_n, \tag{16.7}$$

writing each of the above S_n integers in binary. Each of the resulting binary integers contains $2n+1$ digits, implying that the total number of digits appearing in the sequence (16.7) is equal to $(2n+1)S_n$.

Now, letting $\lambda(m)$ stand for the number of digits in the integer m, the total number N of digits of the concatenated integers preceding the number $2^{2n}+1$ is, as n becomes large,

$$N = \sum_{m\le 2^{2n}} \lambda(m) = 2n+1+\sum_{m\le 2^{2n}} \left\lceil \frac{\log m}{\log 2}\right\rceil = (1+o(1))2n\cdot 2^{2n}. \tag{16.8}$$

We can write the first digits of the Champernowne number as

$$\begin{aligned} \theta &= 0.\epsilon_1\epsilon_2\ldots\epsilon_N\overline{2^{2n}+1}\,\overline{2^{2n}+2}\ \ldots\overline{2^{2n}+S_n}\ldots \\ &= 0.\epsilon_1\epsilon_2\ldots\epsilon_N\,\rho\ \ldots, \end{aligned}$$

say, where in fact, $\rho = \overline{2^{2n}+1}\,\overline{2^{2n}+2}\ \ldots\overline{2^{2n}+S_n} = \epsilon_{N+1}\ldots\epsilon_{N+\lambda(\rho)}$. (Here, $\overline{n_1}\,\overline{n_2}\ldots\overline{n_r}$ stands for the concatenation of all the digits appearing successively in the integers $n_1, n_2, \ldots, n_r$.) We will first show that the proportion of zeros in the word ρ is too large. For this we shall first count the number of 1's in ρ. Setting $\beta(m)$ as the number of 1's in the integer m, the total number of 1's in ρ is equal to

$$\sum_{m\le S_n} \beta(m) = \frac{1}{2}\frac{S_n\log S_n}{\log 2} + O(S_n),$$

from which we can deduce that the total number of zeros in ρ is

$$\sum_{m=1}^{S_n} n + \sum_{m=1}^{S_n}(n-\beta(m)) = 2nS_n - \frac{1}{2}\frac{S_n\log S_n}{\log 2} + O(S_n). \tag{16.9}$$

Since $\lambda(\rho) = (2n+1)S_n$ and recalling that $S_n = \lfloor 2^n/(\sqrt{n}\log n)\rfloor$, it follows from (16.9) that the proportion of zeros in ρ is equal to, as $n\to\infty$,

$$\begin{aligned} \frac{1}{\lambda(\rho)}\times \text{the number of zeros in } \rho &= \frac{2n}{2n+1} - \frac{1}{2}\frac{\log S_n}{(2n+1)\log 2} + o(1) \\ &= 1+o(1) - \frac{1}{2}\frac{n\log 2 - \frac{1}{2}\log n}{(2n+1)\log 2} + o(1) \\ &= 1-\frac{1}{4}+o(1) = \frac{3}{4}+o(1). \end{aligned}$$

Then, since

$$\left|\sum_{\substack{N+1\le\nu\le N+M\\ \{2^\nu\theta\}<\frac{1}{2}}} 1 - \frac{1}{2}(2n+1)S_n\right| \ge \frac{1}{4}(2n+1)S_n,$$

it follows that, setting $x_n := \{2^n\theta\}$ and choosing

$$M = M_N = (2n+1)S_n \approx \sqrt{N}/\log\log N$$

(where we used (16.8)), thereby complying with condition (16.1), the discrepancy of the sequence of numbers $x_{N+1}, \ldots, x_{N+M}$ is

$$\begin{aligned} D(x_{N+1}, \ldots, x_{N+M}) & \\ &= \sup_{[a,b)\subseteq[0,1)} \frac{1}{(2n+1)S_n} \left| \sum_{\substack{N+1\le\nu\le N+M \\ \{2^\nu\theta\}\in[a,b)}} 1 - (b-a)\left((2n+1)S_n\right) \right| \\ &\geq \frac{\frac{1}{4}(2n+1)S_n}{(2n+1)S_n} = \frac{1}{4} \end{aligned}$$

and therefore does not tend to 0, thereby implying that θ is not sharply normal.

Remark 16.3. *Observe that instead of choosing $M_N = \lfloor \delta_N \sqrt{N} \rfloor$ as we did in (16.1), we could have set $M_N = \lfloor \delta_N N^\gamma \rfloor$, where γ is fixed real number belonging to the interval $(0,1)$, and then introduce the corresponding concept of a γ-strongly uniformly distributed sequence mod 1, with corresponding γ-strong normal numbers. In this case, one could easily show that if $0 < \gamma_1 < \gamma_2 < 1$, then any γ_1-strong normal number is also be a γ_2-strong normal number.*

Remark 16.4. *A further discussion on appropriate choices of M_N in the definition of sharp normality is exposed below.*

Identifying which real numbers are normal is not an easy task. For instance, no one has been able to prove that any of the classical constants π, e, $\sqrt{2}$ and $\log 2$ is normal, even though numerical evidence indicates that all of them are. Even constructing normal numbers is not an easy task. Hence, one might believe that constructing sharp normal numbers will even be more difficult. So, here we first show how one can construct large families of sharp normal numbers. On the other hand, it has been shown by Borel [6] that almost all real numbers are normal. Although the set of sharp normal numbers is "much smaller" than the whole set of normal numbers, in this paper, we prove that almost all numbers are sharply normal. After studying the multidimensional case, we examine the relation between arithmetic functions with local normal distribution and sharp normality.

Our first two propositions provide a simple criteria for sharp uniform distribution mod 1 and for sharp normality. They are direct consequences of the definition of sharp normality.

Proposition 16.1. *Let $\mathcal{D}$ be the set of all continuous functions $f : [0,1] \to [0,1)$ such that $\int_0^1 f(x)\,dx = 0$. Then, the sequence $(x_n)_{n\geq 1}$ is sharply uniformly distributed mod 1 if and only if, for all $f \in \mathcal{D}$, letting $M = M_N$ be as in (16.1),*

$$\frac{1}{M}\sum_{j=1}^{M} f(\{x_{N+j}\}) \to 0 \qquad \text{as } N \to \infty.$$

Given a positive real number $\alpha < 1$ whose q-ary expansion is written as $\alpha = 0.\epsilon_1\epsilon_2\ldots$, where each $\epsilon_j \in \mathcal{A}_q := \{0, 1, \ldots, q-1\}$. For an arbitrary word $\beta = \delta_1 \ldots \delta_k \in \mathcal{A}_q^k$, let $R_{N,M}(\beta)$ stand for the number of times that the word β appears as a subword of the word $\epsilon_{N+1} \ldots \epsilon_{N+M}$.

Proposition 16.2. *A positive real number $\alpha < 1$ is sharply q-normal if and only if, given an arbitrary word $\beta = \delta_1 \ldots \delta_k \in \mathcal{A}_q^k$ and $M = M_N$ as in (16.1),*

$$\lim_{N\to\infty} \frac{R_{N,M}(\beta)}{M} = \frac{1}{q^k}.$$

The construction of sharp normal numbers

We first show how one can go about constructing sharp normal numbers. One way is as follows. First, we start with a normal number in base $q \geq 2$, say $\alpha = 0.\epsilon_1\epsilon_2\ldots$, and then for each positive integer T, we consider the corresponding word $\alpha_T = \epsilon_1\epsilon_2\ldots\epsilon_T$. One can show that, if the sequences of integers $T_1 < T_2 < \cdots$ and $m_1 < m_2 < \cdots$ are chosen appropriately, and if, for short, we write γ^m for the concatenation of m times the word γ, that is $\gamma^m = \underbrace{\gamma\ldots\gamma}_{m \text{ times}}$, then the number

$$\beta = 0.\alpha_{T_1}^{m_1}\alpha_{T_2}^{m_2}\ldots$$

is a sharp normal number in base q.

We first show that the choice $T_\ell = \ell$ and $m_\ell = \ell$ is an appropriate one and in fact we state this as a proposition.

Proposition 16.3. *Let α be a q-normal number. Then, using the above notation, the number*

$$\beta = 0.\alpha_1^1\alpha_2^2\alpha_3^3\ldots$$

is a sharp normal number in base q.

Remark 16.5. *Other choices of T_ℓ and m_ℓ can also lead to the construction of sharp normal numbers. For instance, let $R > 0$ be a fixed integer and, for each real number $x > 0$, define*

$$x_1 := \log_+ x = \max(1, \log x), \qquad x_{\ell+1} = \log_+ x_\ell \qquad (\ell = 1, 2, \ldots).$$

Given a real number

$$\alpha = 0.\epsilon_1\epsilon_2\ldots \in \mathcal{A}_q^{\mathbb{N}},$$

set

$$F(\alpha; \beta) = \#\{(\gamma_1, \gamma_2) : \alpha = \gamma_1\beta\gamma_2\},$$

that is the number of occurrences of the word β in the digits of the word α. One can construct a real number α such that, for every integer $k \geq 1$,

$$\max_{\beta\in\mathcal{A}_q^k} \left| \frac{1}{M_N} F(\epsilon_{N+1}\ldots\epsilon_{N+M_N}; \beta) - \frac{1}{q^k} \right| \to 0 \quad \text{as } N \to \infty. \tag{16.10}$$

Indeed, for each integer $\ell \geq 1$, let us choose $T_\ell = \ell$ and $m_\ell = 2^{2^{\cdot^{\cdot^{\cdot^{2^\ell}}}}}$, that is $\ell = \underbrace{\log_2 \log_2 \ldots \log_2}_{R+1 \text{ times}} m_\ell$. Now, starting with a q-ary normal number $\gamma = 0.\epsilon_1\epsilon_2\ldots$, and, for each positive integer T, set $\gamma_T = 0.\epsilon_1\epsilon_2\ldots\epsilon_T$. Then, one can show that the number

$$\beta = 0.\gamma_1^{m_1}\gamma_2^{m_2}\ldots$$

does indeed satisfy condition (16.10) and is therefore a sharp q-normal number.

Preliminary lemmas

A real number is *simply normal* in base q if in its base q expansion, every digit $0, 1, \ldots, q-1$ occurs with the same frequency $1/q$. The following lemma offers a simple way of establishing if a given real number is a normal number.

Lemma 16.1. *Let $q \geq 2$ be an integer. If a real number α is simply normal in base q^r for each $r \in \mathbb{N}$, then α is normal in base q.*

Proof. A proof of this result can be found in the book of Kuipers and Niederreiter [50]. □

In the spirit of Proposition 16.2, we will say that a real number $\alpha < 1$ is a *simply sharp normal number in base q* if for every digit $d \in \mathcal{A}_q$,

$$\lim_{N\to\infty} \frac{R_{N,M}(d)}{M} = \frac{1}{q}.$$

Lemma 16.2. *Let $q \geq 2$ be an integer. If a real number α is a simply sharp normal in base q^r for each $r \in \mathbb{N}$, then α is sharply normal in base q.*

Proof. This result can be proved along the same lines as one would use to prove Lemma 16.1. □

Lemma 16.3. *For each integer $k \geq 1$, let*

$$\pi_k(x) := \#\{n \leq x : \omega(n) = k\}.$$

Then, the relation

$$\pi_k(x) = (1 + o(1)) \frac{x}{\log x} \frac{(\log\log x)^{k-1}}{(k-1)!} \qquad (x \to \infty)$$

holds uniformly for

$$|k - \log\log x| \leq \frac{1}{\delta_x}\sqrt{\log\log x}, \tag{16.11}$$

where δ_x is some function of x chosen appropriately and which tends to 0 as $x \to \infty$.

Proof. This follows from Theorem 10.4 stated in the book of De Koninck and Luca [34]. □

Lemma 16.4. *Letting δ_x be as in the statement of Lemma 16.3,*

$$\max_{\substack{k \ \text{satisfying (16.11)} \\ \ell\in[0,\lceil\delta_x^{3/2}\sqrt{\log\log x}\rceil]}} \left|\frac{\pi_{k+\ell}(x)}{\pi_k(x)}-1\right| \to 0 \qquad \text{as } x\to\infty.$$

Proof. Given k satisfying (16.11), let θ_k be defined implicitly by $k=\log\log x+\theta_k$, and let $\ell\in[0,\lceil\delta_x^{3/2}\sqrt{\log\log x}\rceil]$. Then, in light of Lemma 16.3, we have, as $x\to\infty$,

$$\begin{aligned}
\frac{\pi_{k+\ell}(x)}{\pi_k(x)} &= (1+o(1))\,\frac{(\log\log x)^\ell}{k^\ell\prod_{\nu=0}^{\ell-1}\left(1+\frac{\nu}{k}\right)}\\
&= (1+o(1))\left(\frac{\log\log x}{k}\right)^\ell \exp\left\{-\frac{\ell(\ell-1)}{2k}+O\left(\frac{\ell^3}{k^2}\right)\right\}\\
&= (1+o(1))\left(\frac{1}{1+\theta_k/\log\log x}\right)^\ell (1+o(1))\\
&= (1+o(1))\exp\left\{-\frac{\ell\theta_k}{\log\log x}+O\left(\frac{\ell\theta_k^2}{(\log\log x)^2}\right)\right\}\\
&= 1+o(1),
\end{aligned}$$

thereby completing the proof of Lemma 16.4. □

For any particular set of primes $\mathcal{P}$, we introduce the expressions

(16.12) $$\Omega_{\mathcal{P}}(n):=\sum_{\substack{p^a\|n\\ p\in\mathcal{P}}} a \qquad\text{and}\qquad E(x):=\sum_{\substack{p\le x\\ p\in\mathcal{P}}}\frac{1}{p}.$$

The following two results, which we also state as lemmas, are due respectively to Halász [42] and Kátai [49].

Lemma 16.5. (Halász) *Let $0<\delta\le 1$ and let $\mathcal{P}$ be a set of primes with corresponding functions $\Omega_{\mathcal{P}}(n)$ and $E(x)$ given in (16.12). Then, the estimate*

$$\sum_{\substack{n\le x\\ \Omega_{\mathcal{P}}(n)=k}} 1 = \frac{xE(x)^k}{k!}e^{-E(x)}\left\{1+O\left(\frac{|k-E(x)|}{E(x)}\right)+O\left(\frac{1}{\sqrt{E(x)}}\right)\right\}$$

holds uniformly for all integers k and real numbers $x\ge 3$ satisfying

$$E(x)\ge\frac{8}{\delta^3} \qquad \text{and} \qquad \delta\le\frac{k}{E(x)}\le 2-\delta.$$

Lemma 16.6. (Kátai) *For $1\le h\le x$, let*

$$\begin{aligned}
A_k(x,h) &:= \sum_{\substack{x\le n\le x+h\\ \omega(n)=k}} 1,\\
\delta_k(x,h) &:= \frac{A_k(x,h)}{h}-\frac{\pi_k(x)}{x},
\end{aligned}$$

$$E(x,h) \;:=\; \sum_{k=1}^{\infty} \delta_k^2(x,h).$$

Letting $\varepsilon > 0$ be an arbitrarily small number and $x^{7/12+\varepsilon} \le h \le x$, then

$$E(x,h) \ll \frac{1}{\log^2 x \cdot \sqrt{\log\log x}}.$$

Our main results

Theorem 16.1. *The Lebesgue measure of the set of all those real numbers $\alpha \in [0,1]$ which are not sharply q-normal is equal to 0.*

Let r be a fixed positive integer and set $E := [0,1)^r$. Consider an r dimensional sequence $(\underline{x}_n)_{n\in\mathbb{N}} := (x_1^{(n)}, \ldots, x_r^{(n)})_{n\in\mathbb{N}}$ in $\mathbb{R}^r$. This sequence is said to be *uniformly distributed mod E* if, for all intervals $[a_j, b_j) \subseteq [0,1)$, $j = 1, \ldots, r$, we have

$$\lim_{N\to\infty} \frac{1}{N} \#\{n \le N : \{x_j^{(n)}\} \in [a_j, b_j) \text{ for } j = 1, \ldots, r\} = \prod_{j=1}^{r} (b_j - a_j).$$

Accordingly, the *discrepancy* of the finite sequence $\underline{x}_1, \ldots, \underline{x}_N$ in $\mathbb{R}^r$ is defined as

$$D(\underline{x}_1, \ldots, \underline{x}_N) = \sup_{\substack{[a_j,b_j)\subseteq[0,1)\\ j=1,\ldots,r}} \left| \frac{1}{N} \sum_{\substack{\{x_j^{(n)}\}\in[a_j,b_j)\\ j=1,\ldots,r}} 1 - \prod_{j=1}^{r} (b_j - a_j) \right|.$$

Then, we shall say that an infinite sequence $(\underline{x}_n)_{n\in\mathbb{N}}$ is *sharply uniformly distributed mod E* if

$$D(\underline{x}_N, \ldots, \underline{x}_{N+M}) \to 0 \qquad \text{as } N \to \infty$$

for every choice of δ_N satisfying (16.1).

In what follows, we let $q_1, \ldots, q_r$ be fixed integers ≥ 2.

Theorem 16.2. *The Lebesgue measure of the set of all those r-tuples $(\alpha_1, \ldots, \alpha_r) \in [0,1)^r$ for which the sequence $(\underline{x}_n)_{n\in\mathbb{N}}$, where $\underline{x}_n := (\{\alpha_1 q_1^n\}, \ldots, \{\alpha_r q_r^n\})$, is not sharply uniformly distributed in $[0,1)^r$ is equal to 0.*

Theorem 16.3. *Assume that for each $i = 1, 2, \ldots, r$, the number α_i is sharply q_i-normal. Let $E = [0,1)^r$ and assume that f is a continuous periodic function mod E and that it satisfies $\int_0^1 \cdots \int_0^1 f(x_1, \ldots, x_r)\, dx_1 \cdots dx_r = 0$. Further set*

$$y_n = f(\alpha_1 q_1^{\omega(n)}, \ldots, \alpha_r q_r^{\omega(n)}) \qquad (n = 1, 2, \ldots).$$

Then,

$$\frac{1}{x} \sum_{n \le x} y_n \to 0 \qquad \text{as } x \to \infty. \tag{16.13}$$

Moreover, further defining $\underline{z}_n := \left(\{\alpha_1 q_1^{\omega(n)}\}, \ldots, \{\alpha_r q_r^{\omega(n)}\}\right)$ *for* $n = 1, 2, \ldots$, *we have that* $(\underline{z}_n)_{n\in\mathbb{N}}$ *is uniformly distributed in* E.

The following result is a direct consequence of Theorem 16.3 and is related to the result stated in Lemma 16.5.

Theorem 16.4. *Let* g *be any one of the arithmetic functions*

$$\omega(n) := \sum_{p|n} 1, \qquad \Omega(n) := \sum_{p^a \| n} a, \qquad \Omega_{\mathcal{P}}(n) := \sum_{\substack{p^a \| n \\ p\in\mathcal{P}}} a$$

and let $\underline{x}_n := (\{\alpha_1 q_1^{g(n)}\}, \ldots, \{\alpha_r q_r^{g(n)}\})$. *Then, for almost all* $(\alpha_1, \ldots, \alpha_r) \in [0,1)^r$, *the sequence* $(\underline{x}_n)_{n\geq 1}$ *is uniformly distributed in* $[0,1)^r$.

The following result is a consequence of Lemma 16.6 and its proof is essentially along the same lines as that of Theorem 16.3.

Theorem 16.5. *For each integer* $i = 1, \ldots, r$, *assume that* α_i *is sharply* q_i*-normal and set*

$$\underline{x}_n := (\{\alpha_1 q_1^{\omega(n)}\}, \ldots, \{\alpha_r q_r^{\omega(n)}\}).$$

Then, with $M = M_N$ *as in (16.1),*

$$D(\underline{x}_{N+1}, \ldots, \underline{x}_{N+M}) \to 0 \qquad \text{as } N \to \infty.$$

We only provide here the proof of Theorem 16.1, which will follow essentially from the following lemma.

Lemma 16.7. *Let* $(\Omega, \mathcal{A}, P)$ *be a probability space, where* $\Omega = [0,1)$, $\mathcal{A}$ *is the ring of Borel sets and* P *is the Lebesgue measure. Let* $q \geq 2$ *be a fixed integer and set* $\mathcal{A}_q := \{0, 1, \ldots, q-1\}$. *Let* $\epsilon_n \in \mathcal{A}_q$, $n = 1, 2, \ldots$, *be independent random variables such that* $P(\epsilon_n = a) = 1/q$ *for each* $a \in \mathcal{A}_q$. *For each* $\omega \in \Omega$, *let*

$$\alpha(\omega) := 0.\epsilon_1(\omega)\epsilon_2(\omega)\ldots$$

For an arbitrary $\delta > 0$, *let*

$$E_\delta := \left\{ \omega \in \Omega : \limsup_{N\to\infty} \max_{d\in\mathcal{A}_q} \left| \frac{1}{M} \sum_{\substack{n=N+1 \\ \epsilon_n = d}}^{N+M} 1 - \frac{1}{q} \right| > \delta \right\},$$

where M *satisfies (16.1). Then,*

$$P(E_\delta) = 0 \qquad \text{for every } \delta > 0. \tag{16.14}$$

Moreover, setting

$$E^* := \left\{ \omega \in \Omega : \limsup_{N\to\infty} \max_{d\in\mathcal{A}_q} \left| \frac{1}{M} \sum_{\substack{n=N+1 \\ \epsilon_n = d}}^{N+M} 1 - \frac{1}{q} \right| \neq 0 \right\},$$

we have $P(E^*) = 0$.

Proof of Lemma 16.7. Let $U \in \mathbb{N}$ and given any $d \in \mathcal{A}_q$, let

$$\alpha_d(\epsilon_1, \ldots, \epsilon_U) = \sum_{\substack{i \in \{1,\ldots,U\} \\ \epsilon_i = d}} 1.$$

It is clear that

$$P(\alpha_d(\epsilon_1, \ldots, \epsilon_U) = j) = \frac{1}{q^U} \binom{U}{j} (q-1)^{U-j}.$$

For each $0 < \delta < 1/q$, set

$$S = S(\delta) := \left\{ \omega \in \Omega : \max_{d \in \mathcal{A}_q} \left| \alpha_d(\epsilon_1, \ldots, \epsilon_U) - \frac{U}{q} \right| > \delta U \right\}. \tag{16.15}$$

If $\omega \in S$, then clearly the inequality

$$\alpha_d(\epsilon_1, \ldots, \epsilon_U) < \frac{U}{q} - \delta \frac{U}{q}$$

holds for at least one $d \in \mathcal{A}_q$, in which case we have

$$P(S) \le \frac{q}{q^U} \sum_{0 \le j \le (1-\delta)U/q} \binom{U}{j} \cdot (q-1)^{U-j} = q \left(1 - \frac{1}{q}\right)^U \sum_{0 \le j \le V} \binom{U}{j} \frac{1}{(q-1)^j}, \tag{16.16}$$

where $V = \lfloor (1-\delta)U/q \rfloor$.

Now let

$$t_j = \binom{U}{j} \frac{1}{(q-1)^j} \qquad (j = 0, 1, \ldots, V).$$

Then, for each integer $j \ge 1$, we have

$$\frac{t_{j-1}}{t_j} = (q-1)\frac{j}{U-j+1} < \frac{(q-1)(1-\delta)U/q}{U+1-(1-\delta)U/q} < \frac{(q-1)(1-\delta)}{q-(1-\delta)} < 1-\delta,$$

so that $t_{j-1} < (1-\delta)t_j$, thus implying that

$$\sum_{0 \le j \le V} t_j \le t_V \left\{1 + (1-\delta) + (1-\delta)^2 + \ldots\right\} = \frac{t_V}{\delta}.$$

Using the Stirling formula in the form

$$\log n! = n \log(n/e) + \frac{1}{2}\log(2\pi n) + \theta_n \qquad \text{with } \theta_n \to 0$$

and setting $V = \kappa U$, where $\kappa = \dfrac{\lfloor \frac{1-\delta}{q} U \rfloor}{U} = \dfrac{1-\delta}{q} + O\left(\dfrac{1}{U}\right)$, we then have

$$\begin{aligned} \log t_V &= U \log U - \kappa U \log(\kappa U) - (1-\kappa)U \log((1-\kappa)U) - \kappa U \log(q-1) \\ &\quad + \frac{1}{2} \log \frac{1}{\kappa(1-\kappa)} - \frac{1}{2}\log(2\pi) + O(\theta_V) \end{aligned}$$

$$= \quad (-\kappa\log\kappa-(1-\kappa)\log(1-\kappa)-\kappa\log(q-1))\,U$$
$$+\frac{1}{2}\log\frac{1}{\kappa(1-\kappa)}-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log U+O(\theta_V).$$

Letting $h(\kappa)=\kappa\log\dfrac{1}{(q-1)\kappa}+(1-\kappa)\log\dfrac{1}{1-\kappa}$, it follows that

$$\log t_V = Uh(\kappa)+\frac{1}{2}\log\frac{1}{\kappa(1-\kappa)}-\frac{1}{2}\log(2\pi)-\frac{1}{2}\log U+O(\theta_V).$$

Observe that $h(1/q)=\log\dfrac{q}{q-1}$ and that

$$h(\kappa)<(1-c(\delta))\log\frac{q}{q-1},$$

where $c(\delta)>0$ provided $\delta>0$.

Using this in (16.16), we obtain that

$$(16.17)\quad P(S)\le q\exp\left\{U\log(1-1/q)+\log V+U(1-c(\delta))\log\frac{q}{q-1}\right\}<\exp\{-c_1(\delta)U\},$$

where $c_1(\delta)>0$ is some constant depending only on δ and q.

For each integer $r\ge 1$, let $N_r=q^r$ and consider the interval $\mathcal{L}_r=[N_r,N_{r+1}-1]$. Let us cover a given interval $\mathcal{L}_r$ by the union of $K_r:=1+\lfloor\frac{(q-1)q^r}{r^2}\rfloor$ consecutive intervals $\mathcal{T}_1^{(r)},\mathcal{T}_2^{(r)},\dots,\mathcal{T}_{K_r+1}^{(r)}$, each of length $U_r:=r^2$. Now, we define the sets $S_i^{(r)}$, for $i=1,\dots,K_r+1$, as we did for the set S in (16.15), but this time with the independent variables

$$\epsilon_{N_r+(i-1)U_r+\ell}\qquad(\ell=1,2,\dots,U_r).$$

For these new independent variables, if we proceed as we did to obtain (16.17), we then have

$$P(S_i^{(r)})\le q^r\exp\{-c_1(\delta)r^2\}\qquad(i=1,\dots,K_r+1),$$

so that

$$P\left(\bigcup_{i=1}^{K_r+1}S_i^{(r)}\right)\ \ll\ K_rq^r\exp\{-c_1(\delta)r^2\}\le\frac{q^{2r+1}}{r^2}\exp\{-c_1(\delta)r^2\}$$
$$=\ \exp\{-c_1(\delta)r^2+(2r+1)\log q-2\log r\}<\frac{1}{r^3},$$

provided r is sufficiently large.

Since the series $\sum 1/r^3$ converges, we may apply Lemma 0.14 and conclude that the set

$$E_\delta:=\#\{\omega:\omega\in\bigcup_{i=1}^{K_r+1}S_i^{(r)}\text{ for infinitely many }r\}$$

is such that $P(E_\delta)=0$. From this result, it then follows also that $P(E^*)=0$. □

Final remarks

When we introduced the notion of *sharply normal number in base* q, we chose for simplicity to consider intervals $[N+1, N+M]$ with $M=\lfloor \delta_N \sqrt{N} \rfloor$. However, it is interesting to observe that we could have chosen much smaller intervals, namely with $M=\lfloor \log^2 N \rfloor$, and nevertheless still preserve the property that almost all real numbers are sharply normal. Indeed, following the proof used in Lemma 16.7, as we consider an arbitrary sequence of digits $\epsilon_{N+1}\epsilon_{N+2}\ldots\epsilon_{N+M}$, with $M=\lfloor \log^2 N \rfloor$, and examine the occurrence of an arbitrary digit $d \in \mathcal{A}_q$ in this sequence, we could define r as the unique integer such that $q^r \le n < q^{r+1}$, in which case we would have

$$r^2 \le \left(\frac{\log n}{\log q}\right)^2 < (r+1)^2.$$

In the end, we would see that

$$\left| \frac{1}{\log^2 n} \sum_{\substack{\nu=n+1 \\ \epsilon_\nu = d}}^{n+\lfloor \log^2 n \rfloor} 1 - \frac{1}{q} \right| > \delta$$

holds only for finitely many n's and that this is true for each $\delta > 0$. We can conclude from this that, for almost all α,

$$\lim_{n\to\infty} \max_{d \in \mathcal{A}_q} \left| \frac{1}{\log^2 n} \sum_{\substack{\nu=n+1 \\ \epsilon_\nu = d}}^{n+\lfloor \log^2 n \rfloor} 1 - \frac{1}{q} \right| = 0,$$

thus also establishing that we could have defined the notion of *sharply normal numbers* with $M=\lfloor \log^2 N \rfloor$ instead of with $M=\lfloor \delta_N \sqrt{N} \rfloor$.

Now, could we have chosen M even smaller, say $M=\lfloor \log N \rfloor$? Not really! Indeed, assume that $(\epsilon_n)_{n\ge 1}$ are independent random variables such that $P(\epsilon_n = a) = 1/q$ for each $a \in \mathcal{A}_q$. For $N \in \mathbb{N}$, let $H = H_N = \left\lfloor \dfrac{q^{N+1}-q^N}{N} \right\rfloor$ and set

$$B_\ell^{(N)} := \{\omega : \epsilon_{q^N+\ell N+\nu} = 0,\ \nu = 0,1,\ldots,N-1\} \qquad (\ell = 0,1,\ldots,H-1).$$

The events $B_\ell^{(N)}$ $(\ell = 0,1,\ldots,H-1)$ are independent and $P\left(B_\ell^{(N)}\right) = 1/q^N$. Hence, with $D_N = \bigcup_{\ell=0}^{H-1} B_\ell^{(N)}$, we have

$$P(D_N) = \frac{H}{q^N} \ge \frac{1}{2^N}.$$

On the other hand $D_1, D_2, \ldots$ are independent and $\displaystyle\sum_{N=1}^{\infty} P(D_N) = \infty$. Hence, by the second Borel-Cantelli lemma (see Lemma 0.15), we may conclude that for almost all events ω, there exists an infinite sequence of N's, say $n_1, n_2, \ldots$ such that

$$\epsilon_{n_\nu+1} = 0, \epsilon_{n_\nu+2} = 0, \ldots, \epsilon_{n_\nu+m_\nu} = 0,$$

where $m_\nu \geq c\dfrac{\log n_\nu}{\log q}$. We have thus shown that one could encounter a normal number α with sequences of digits covering intervals of the form $[N+1, N+M]$, with $M \approx \log N$, made up only of zeros.

XVII. Prime factorization and normal numbers [29]

(Researches in Mathematics and Mechanics, 2015)

In two papers [14], [18], we used the fact that the prime factorization of integers is locally chaotic but at the same time globally very regular in order to create various families of normal numbers.

Here, we create a new family of normal numbers again using the factorization of integers but with a different approach. Write each integer $n \geq 2$ as $n = p_1 p_2 \cdots p_r$, where $p_1 \leq p_2 \leq \cdots \leq p_r$ represent all the prime factors of n. Then, setting $\ell(1) = 1$ and, for each integer $n \geq 2$, letting $\ell(n)$ represent the concatenation of the primes $p_1, p_2, \ldots, p_r$, we show that by concatenating $\ell(1), \ell(2), \ell(3), \ldots$, we can create a normal number, that is that the real number $0.\ell(1)\ell(2)\ell(3)\ldots$ is a normal number. Actually, we prove more general results.

Main results

Let $q \geq 2$ be a fixed integer. From here on, we let $S(x) \in \mathbb{Z}[x]$ be an arbitrary polynomial (of positive degree r_0) such that $S(n) > 0$ for all integers $n \geq 1$. Moreover, for each integer $n \geq 2$, we write its prime factorization as $n = p_1 p_2 \cdots p_r$, where $p_1 \leq p_2 \leq \cdots \leq p_r$ are all the prime factors of n and set

$$\ell(n) := \overline{S(p_1)}\,\overline{S(p_2)}\,\ldots\,\overline{S(p_r)},$$

where each $S(p_i)$ is expressed in base q. For convenience, we set $\ell(1) = 1$.

Theorem 17.1. *The real number*

$$\xi := 0.\ell(1)\ell(2)\ell(3)\ell(4)\ldots$$

is a q-normal number.

Theorem 17.2. *Given an arbitrary positive integer a, the real number*

$$\eta := 0.\ell(2+a)\ell(3+a)\ell(5+a)\ldots\ell(p+a)\ldots,$$

where p runs through all primes, is a q-normal number.

Let $1 = d_1 < d_2 < \cdots < d_{\tau(n)} = n$ be the sequence of divisors of n and let $t(n) = \overline{S(d_1)}\,\overline{S(d_2)}\,\ldots\,\overline{S(d_{\tau(n)})}$. Then, let

$$\begin{aligned} \theta &:= 0.\mathrm{Concat}(t(n) : n \in \mathbb{N}), \\ \kappa &:= 0.\mathrm{Concat}(t(p+a) : p \in \wp), \end{aligned}$$

where a is a fixed positive integer.

Theorem 17.3. *The above real numbers θ and κ are q-normal numbers.*

Let $S(x)$ be as above and let $Q(x) \in \mathbb{Z}[x]$ be a polynomial of positive degree such that $Q(n) > 0$ for each integer $n \geq 1$. Then, consider the expression

$$Q(n) = \prod_{p^a \| Q(n)} p^a = p_1 p_2 \cdots p_r,$$

where $p_1 \leq p_2 \leq \cdots \leq p_r$ are all the prime factors of $Q(n)$, so that

$$\ell(Q(n)) = \overline{S(p_1)}\,\overline{S(p_2)}\ \ldots\ \overline{S(p_r)}.$$

Then, let

$$\begin{aligned} \alpha &:= 0.\mathrm{Concat}(\ell(Q(n)) : n \in \mathbb{N}), \\ \beta &:= 0.\mathrm{Concat}(\ell(Q(p)) : p \in \wp). \end{aligned}$$

Theorem 17.4. *The above real number α is a q-normal number and, if $Q(0) \neq 0$, the real number β is also a q-normal number.*

Let $Q(x)$ be as above. Then, let $1 = e_1 < e_2 < \cdots < e_{\delta(n)}$ be the sequence of all the divisors of $Q(n)$ which do not exceed n, consider the expression

$$h(Q(n)) := \overline{S(e_1)}\,\overline{S(e_2)}\ \ldots\ \overline{S(e_{\delta(n)})}$$

and set

$$\psi := 0.\mathrm{Concat}(h(Q(n)) : n \in \mathbb{N})$$

Theorem 17.5. *The above real number ψ is a q-normal number.*

Here we shall only prove Theorems 17.1, 17.2 and 17.3. For this, we will need the following two lemmas.

Lemma 17.1. *Let $S \in \mathbb{Z}[x]$ be as above. Given a positive integer k, let β_1 and β_2 be any two distinct words belonging to $\mathcal{A}_q^k$. Let $c_0 > 0$ be an arbitrary number and consider the intervals*

$$J_w := \left[w, w + \frac{w}{\log^{c_0} w}\right] \qquad (w > 1).$$

Then,

$$\frac{1}{\pi(J_w) \cdot \log w} \sum_{p \in J_w} \left| \nu_{\beta_1}(\overline{S(p)} - \nu_{\beta_2}(\overline{S(p)}) \right| \to 0 \qquad \text{as } w \to \infty.$$

Proof. This result is a consequence of Lemma 0.5. □

Given an infinite sequence $\gamma = a_1 a_2 \ldots \in \mathcal{A}_q^{\mathbb{N}}$ and a positive integer T, we write γ^T for the word $a_1 a_2 \ldots a_T$.

Lemma 17.2. *The infinite sequence γ is a q-normal sequence if for every positive integer k and arbitrary words $\beta_1, \beta_2 \in \mathcal{A}_q^k$, there exists an infinite sequence of positive integers $T_1 < T_2 < \cdots$ such that* *(i)* $\lim_{n\to\infty} \frac{T_{n+1}}{T_n} = 1$,

(ii) $\lim_{n\to\infty} \frac{1}{T_n} \left|\nu_{\beta_1}(\gamma^{T_n}) - \nu_{\beta_2}(\gamma^{T_n})\right| = 0$.

Proof. It is easily seen that conditions (i) and (ii) imply that

$$\frac{1}{T}\left|\nu_{\beta_1}(\gamma^T) - \nu_{\beta_2}(\gamma^T)\right| \to 0 \quad \text{as } T \to \infty$$

and consequently that

$$\frac{1}{T}\left|q^k \nu_{\beta_1}(\gamma^T) - \sum_{\beta_2 \in \mathcal{A}_q^k} \nu_{\beta_2}(\gamma^T)\right| \to 0 \quad \text{as } T \to \infty. \tag{17.1}$$

But since

$$\sum_{\beta_2 \in \mathcal{A}_q^k} \nu_{\beta_2}(\gamma^T) = T + O(1),$$

it follows from (17.1) that

$$\frac{\nu_{\beta_1}(\gamma^T)}{T} = (1 + o(1))\frac{1}{q^k} \quad \text{as } T \to \infty,$$

thereby establishing that γ is a q-normal number and thus completing the proof of the lemma. □

Proof of Theorem 17.1

Let x be a large number and set

$$\xi^{(x)} := \ell(1)\ell(2)\ell(3)\ldots\ell(\lfloor x \rfloor).$$

Since $\log S(p) = (1 + o(1))r_0 \log p$ as $p \to \infty$, we find that

$$\begin{aligned}
\lambda(\xi^{(x)}) &= \sum_{n \le x}\left(\left\lfloor \frac{\log \ell(n)}{\log q}\right\rfloor + 1\right) \\
&= \frac{1}{\log q}\sum_{n \le x}\sum_{p^a \| n} a \log S(p) + O(x) \\
&= \frac{1}{\log q}\sum_{\substack{p^a \le x \\ a \ge 1}} a \log S(p)\left(\frac{x}{p^a} + O(1)\right) + O(x) \\
&= \frac{x}{\log q}\sum_{p \le x} \frac{\log S(p)}{p} + O(x) \\
&= (1 + o(1))r_0 \frac{x \log x}{\log q} + O(x),
\end{aligned}$$

thereby establishing that the number of digits of $\xi^{(x)}$ is of order $x\log x$, that is that

$$\lambda(\xi^{(x)}) \approx x\log x. \tag{17.2}$$

Now, we easily obtain that

$$\nu_\beta(\xi^{(x)}) = \sum_{p^a\le x} \nu_\beta(\overline{S(p)}) \left\lfloor \frac{x}{p^a} \right\rfloor + O(x) = x\sum_{p\le x} \frac{\nu_\beta(\overline{S(p)})}{p} + O(x).$$

and therefore that, given any two distinct words $\beta_1,\beta_2 \in A_q^k$, the exists a positive constant C such that, as $x\to\infty$,

$$\frac{1}{\lambda(\xi^{(x)})}\left|\nu_{\beta_1}(\xi^{(x)}) - \nu_{\beta_2}(\xi^{(x)})\right| \le \frac{C}{\log x}\sum_{p\le x}\frac{\left|\nu_{\beta_1}(\overline{S(p)}) - \nu_{\beta_2}(\overline{S(p)})\right|}{p} + o(1). \tag{17.3}$$

On the other hand, it is clear from Lemma 17.1 that

$$\frac{1}{\pi([x,2x])\log x}\sum_{x\le p<2x}\left|\nu_{\beta_1}(\overline{S(p)}) - \nu_{\beta_2}(\overline{S(p)})\right| \to 0 \qquad (x\to\infty). \tag{17.4}$$

Observe that, in light of (17.4), as $x\to\infty$,

$$\begin{aligned}\sum_{p\le x}\frac{\left|\nu_{\beta_1}(\overline{S(p)}) - \nu_{\beta_2}(\overline{S(p)})\right|}{p} &\le \sum_{\substack{2^\ell\le x\\ \ell\ge 1}}\frac{1}{2^\ell}\sum_{2^\ell\le p<2^{\ell+1}}\left|\nu_{\beta_1}(\overline{S(p)}) - \nu_{\beta_2}(\overline{S(p)})\right| \\ &= \sum_{\substack{2^\ell\le x\\ \ell\ge 1}}\frac{1}{2^\ell}\, o\left(\frac{2^\ell\log 2^\ell}{\ell}\right) = o(\log x),\end{aligned}$$

which used in (17.3) along with (17.2) yields

$$\frac{1}{\lambda(\xi^{(x)})}\left|\nu_{\beta_1}(\xi^{(x)}) - \nu_{\beta_2}(\xi^{(x)})\right| = o\left(\frac{1}{\log x}\log x\right) + o(1) = o(1),$$

which, in light of Lemma 17.2, completes the proof of Theorem 17.1.

Proof of Theorem 17.2

Let x be a large number and set

$$\eta^{(x)} := \text{Concat}(\ell(p+a) : p\le x).$$

First observe that the number of digits in the word $\eta^{(x)}$ is of order x, since

$$\lambda(\eta^{(x)}) \approx \pi(x)\log x \approx x. \tag{17.5}$$

On the other hand, letting $\delta>0$ be an arbitrary small number, it follows from Lemma 0.9 that there exists a positive constant $c>0$ such that

$$\#\{\pi\le x : P(\pi+a) > x^{1-\delta}\} \le c\delta\pi(x). \tag{17.6}$$

Arguing as in the proof of Theorem 17.1, we have that, given any two distinct words $\beta_1, \beta_2 \in \mathcal{A}_q^k$, for some positive constant C_1,

$$\begin{aligned} \left|\nu_{\beta_1}(\eta^{(x)}) - \nu_{\beta_2}(\eta^{(x)})\right| &\leq \sum_{p \leq x^{1-\delta}} \left|\nu_{\beta_1}(\overline{S(p)}) - \nu_{\beta_2}(\overline{S(p)})\right| \cdot \pi(x;p,-a) \\ &\quad + C_1 \sum_{x^{1-\delta} < p \leq x} (\log p)\pi(x;p,-a) + O(\pi(x) \log\log x). \end{aligned} \tag{17.7}$$

It follows from Lemma 0.1 that

$$\pi(x;p,-a) \ll \frac{x}{p\log(x/p)}, \tag{17.8}$$

which implies, in light of (17.6), that

$$\sum_{x^{1-\delta} < p \leq x} (\log p)\pi(x;p,-a) \ll \log x \cdot \delta\pi(x) \ll \delta x. \tag{17.9}$$

Using Lemma 0.5, it follows from (17.7), 17.8) and (17.9) that, for some positive constant C_2,

$$\lim_{x\to\infty} \frac{\left|\nu_{\beta_1}(\eta^{(x)}) - \nu_{\beta_2}(\eta^{(x)})\right|}{\lambda(\eta^{(x)})} \leq C_2\delta. \tag{17.10}$$

Since $\delta > 0$ was chosen to be arbitrarily small, it follows that the left hand side of (17.10) must be 0. Combining this with observation (17.5), the result follows.

Proof of Theorem 17.3

The proof that θ is a normal number is somewhat similar to the proof that η is normal as shown in Theorem 17.2. Hence, we will focus our attention on the proof that κ is normal.

Let x be a large number and set $\kappa^{(x)} := \text{Concat}(t(p+a) : p \leq x)$. First we observe that

$$\begin{aligned} \lambda(\kappa^{(x)}) &= \sum_{d \leq x} \lambda(\overline{S(d)})\pi(x;d,-a) + O(\text{li}(x)) \\ &= \sum_{d \leq x} \left(\left\lfloor \frac{\log S(d)}{\log q} \right\rfloor + 1 \right) \pi(x;d,-a) + O(\text{li}(x)) \\ &= r_0 \sum_{d \leq x} \frac{\log d}{\log q} \pi(x;d,-a) + O\left(\sum_{p \leq x} \tau(p+a) \right) + O(\text{li}(x)) \\ &= \frac{r_0}{\log q} \sum_{d \leq x} (\log d)\pi(x;d,-a) + O(x), \end{aligned} \tag{17.11}$$

where we used the fact that $\sum_{p \leq x} \tau(p+a) = O(x)$.

Let $\delta > 0$ be an arbitrarily small number. On the one hand, for some positive constant C_1,

$$\sum_{x^{1-\delta} < d \leq x} (\log d)\pi(x;d,-a) \leq (\log x) \sum_{\substack{x^{1-\delta} < d \leq x \\ dv = p+a,\ p \leq x}} 1$$

$$\le \quad (\log x) \sum_{v \le x^\delta} \pi(x; v, -a)$$

$$\le \quad C_1 (\log x) \sum_{v \le x^\delta} \frac{x}{\phi(v)\log(x/v)} \le \delta C_1 x \log x. \tag{17.12}$$

and, for some positive constant C_2,

$$\sum_{d \le x^{1-\delta}} (\log d)\pi(x; d, -a) \le (\log x) \sum_{d \le x^{1-\delta}} \frac{C_2 x}{\phi(d)\log(x/d)} \le C_2 x. \tag{17.13}$$

On the other hand, using Lemmas 0.1 and 0.2, for some positive constant C_3,

$$\begin{aligned}
\sum_{d \le x} (\log d)\pi(x; d, -a) \quad &\ge \quad \sum_{d \le x^{1/3}} (\log d) \frac{\mathrm{li}(x)}{\phi(d)} - \sum_{d \le x^{1/3}} (\log d) \left| \pi(x; d, -a) - \frac{\mathrm{li}(x)}{\phi(d)} \right| \\
&= \quad C_3(1 + o(1)) x \log x + O\left(\frac{x}{\log^A x} \right) \\
&\gg \quad x \log x. \qquad (17.14)
\end{aligned}$$

Hence combining relations (17.11), (17.12), (17.13) and (17.14), we find that

$$\lambda(\theta^{(x)}) \approx x \log x. \tag{17.15}$$

Now, we easily obtain that, for any distinct words $\beta_1, \beta_2 \in \mathcal{A}_q^k$,

$$\begin{aligned}
\left| \nu_{\beta_1}(\theta^{(x)}) - \nu_{\beta_2}(\theta^{(x)}) \right| \quad &\le \quad \sum_{d \le x^{1-\delta}} \left| \nu_{\beta_1}(\overline{S(d)}) - \nu_{\beta_2}(\overline{S(d)}) \right| \pi(x; d, -a) + c\delta x \log x \\
&\le \quad C_4 \sum_{d \le x^{1-\delta}} \frac{\left| \nu_{\beta_1}(\overline{S(d)}) - \nu_{\beta_2}(\overline{S(d)}) \right|}{\phi(d)\log(x/d)} + c\delta x \log x, \qquad (17.16)
\end{aligned}$$

where we used Lemma 0.1. Combining (17.16) with Lemma 17.1, we obtain that

$$\limsup_{x \to \infty} \left| \frac{\nu_{\beta_1}(\theta^{(x)}) - \nu_{\beta_2}(\theta^{(x)})}{\lambda(\theta^{(x)})} \right| \le \delta,$$

thereby implying, arguing as in the previous proofs and in light of (17.15), that

$$\limsup_{x \to \infty} \left| \frac{\nu_{\beta_1}(\theta^{(x)}) - \nu_{\beta_2}(\theta^{(x)})}{\lambda(\theta^{(x)})} \right| = 0,$$

thus completing the proof of Theorem 17.3.

XVIII. On properties of sharp normal numbers and of non-Liouville numbers [32]

(Annales mathématiques du Québec, 2018)

We show that some sequences of real numbers involving sharp normal numbers or non-Liouville numbers are uniformly distributed modulo 1. In particular, we prove that if $\tau(n)$ stands for the number of divisors of n and α is a binary sharp normal number, then the sequence $(\alpha\tau(n))_{n\geq 1}$ is uniformly distributed modulo 1 and that if $g(x)$ is a polynomial of positive degree with real coefficients and whose leading coefficient is a non-Liouville number, then the sequence $(g(\tau(\tau(n))))_{n\geq 1}$ is also uniformly distributed modulo 1.

Recall the concept of *sharp normality* introduced by De Koninck, Kátai and Phong [31] (see paper XVI above). Before we move on, observe that instead of choosing $M_N = \lfloor \delta_N \sqrt{N} \rfloor$ in (16.1), we could have chosen $M_N = \lfloor \delta_N N^\gamma \rfloor$ for some fixed number $\gamma \in (0,1)$, thereby introducing the notion of γ-sharp distribution modulo 1 and the corresponding notion of γ-sharp normal number. With such definitions, it can be shown that, given $0 < \gamma_1 < \gamma_2 < 1$, any γ_1-sharp normal number is also a γ_2-sharp normal number. One can then show that, given $\gamma \in (0,1)$, almost all real numbers are γ-sharp normal numbers. Various alternatives for the choice of $M = M_N$ in (16.1) are discussed in De Koninck, Kátai and Phong [31].

We shall also need the concept of discrepancy of a set of N t-tuples $\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_N$, where $\underline{y}_n = (x_1^{(n)}, \ldots, x_t^{(n)})$ for $n = 1, 2, \ldots, N$, with each $x_i^{(n)} \in \mathbb{R}$. The discrepancy of a set of N such vectors $\underline{y}_1, \ldots, \underline{y}_N$ is defined as the quantity

$$D(\underline{y}_1, \ldots, \underline{y}_N) := \sup_{I \subseteq [0,1)^t} \left| \frac{1}{N} \sum_{\substack{n=1 \\ \{\underline{y}_n\} \in I}}^{N} 1 - \prod_{i=1}^{t} (\beta_i - \alpha_i) \right|,$$

where $\{\underline{y}_n\}$ stands for $(\{x_1\}, \ldots, \{x_n\})$ and where the above supremum runs over all possible subsets $I = [\alpha_1, \beta_1) \times \cdots \times [\alpha_t, \beta_t)$ of the t-dimensional unit interval $[0,1)^t$.

Recall also that an irrational number β is said to be a *Liouville number* if for each integer $m \geq 1$, there exist two integers t and $s > 1$ such that

$$0 < \left| \beta - \frac{t}{s} \right| < \frac{1}{s^m}.$$

In a sense, one might say that a Liouville number is an irrational number which can be well approximated by a sequence of rational numbers.

Here, we show that some sequences of real numbers involving sharp normal numbers or non-Liouville numbers are uniformly distributed modulo 1. We also study the discrepancy of a sequence of t-tuples of real numbers involving sharp normal numbers.

Throughout this paper, $\wp$ stands for the set of all primes. Given an integer $n \geq 2$, we let $\gamma(n)$ (resp. $\omega(n)$) stand for the product (resp. number) of distinct prime factors of n, with $\gamma(1) = 1$ and $\omega(1) = 0$. Moreover, given a set $\mathcal{B} \subseteq \wp$, we let

$$\omega_{\mathcal{B}}(n) = \sum_{\substack{p \mid n \\ p \in \mathcal{B}}} 1.$$

We also let τ stand for the number of divisors function. More generally, given an integer $\ell \geq 2$, we let $\tau_\ell(n)$ stand for the number of ways of writing n as the product of ℓ positive integers. Also, we let φ stand for the Euler function and write $e(y)$ for $e^{2\pi i y}$. Finally, by $\log_2 x$ (resp. $\log_3 x$) we mean $\max(2, \log\log x)$ (resp. $\max(2, \log\log_2 x)$).

Main results

If α is an irrational number, it is well known that the sequence $(\alpha n)_{n\geq 1}$ is uniformly distributed modulo 1, while there is no guarantee that the sequence $(\alpha\tau(n))_{n\geq 1}$ will itself be uniformly distributed modulo 1. However, if α is a sharp normal number, the situation is different, as is shown in our first result.

Theorem 18.1. *Let $q \geq 2$ be a fixed integer. If α is a sharp q-normal number, then the sequence $(\alpha\tau_q(n))_{n\geq 1}$ is uniformly distributed modulo 1.*

In an earlier paper [30], we showed that if $g(x) = \alpha x^k + \alpha_{k-1}x^{k-1} + \cdots + \alpha_1 x + \alpha_0 \in \mathbb{R}[x]$ is a polynomial of positive degree, where α is a non-Liouville number, and if h belongs to a particular set of arithmetic functions, then the sequence $(g(h(n))_{n\geq 1}$ is uniformly distributed modulo 1. Our next result goes along the same lines.

Theorem 18.2. *Let $g(x) = \alpha x^k + \alpha_{k-1}x^{k-1} + \cdots + \alpha_1 x + \alpha_0 \in \mathbb{R}[x]$ be a polynomial of positive degree, where α is a non-Liouville number. Then, the sequence $(g(\tau(\tau(n))))_{n\geq 1}$ is uniformly distributed modulo 1.*

Now, consider the following (plausible) conjecture.

Conjecture 18.4. *Let ε_x be some function which tends to 0 as $x \to \infty$. Then, if $|k-\ell| \leq \varepsilon_x\sqrt{\log_2 x}$, we have, uniformly for $|k - \log_2 x| \leq \frac{1}{\varepsilon_x}\sqrt{\log_2 x}$ and $|\ell - \log_2 x| \leq \frac{1}{\varepsilon_x}\sqrt{\log_2 x}$, as $x \to \infty$,*

$$\begin{aligned}\frac{1}{x}&\#\{n \leq x : \omega(n) = k \text{ and } \omega(n+1) = \ell\}\\ &= (1+o(1))\frac{1}{x}\#\{n \leq x : \omega(n) = k\} \cdot \frac{1}{x}\#\{n \leq x : \omega(n+1) = \ell\}\end{aligned}$$

and more generally, if $|\ell_i - \ell_j| \leq \varepsilon_x\sqrt{\log_2 x}$ for all $i \neq j$, then, uniformly for $|\ell_j - \log_2 x| \leq \frac{1}{\varepsilon_x}\sqrt{\log_2 x}$, for each $j = 0, 1, \ldots, t-1$, as $x \to \infty$,

$$\begin{aligned}\frac{1}{x}&\#\{n \leq x : \omega(n+j) = \ell_j, \text{ with } j = 0, 1, \ldots, t-1\}\\ &= (1+o(1))\prod_{j=0}^{t-1}\frac{1}{x}\#\{n \leq x : \omega(n+j) = \ell_j\}.\end{aligned}$$

It is interesting to observe that, using the ideas mentioned at the beginning of Theorem 18.3, the following result would follow immediately from Conjecture 18.4.

> Let $q_0, q_1, \ldots, q_{t-1}$ be integers larger than 1 and, for each $j = 0, 1, \ldots, t-1$, let α_j be a sharp q_j-normal number. Consider the sequence of t-tuples $(\underline{x}_n)_{n\geq 1}$ defined by
>
> $$\underline{x}_n := \left(\{\alpha_0 q_0^{\omega(n)}\}, \{\alpha_1 q_1^{\omega(n+1)}\}, \ldots, \{\alpha_{t-1} q_{t-1}^{\omega(n+t-1)}\}\right) \in [0,1)^t.$$
>
> Then, the sequence $(\underline{x}_n)_{n\geq 1}$ is uniformly distributed modulo $[0,1)^t$.

This observation explains the importance of the following result.

Theorem 18.3. *Let w_x and Y_x be two increasing functions both tending to ∞ as $x \to \infty$ and satisfying the conditions*

$$\frac{\log Y_x}{\log x} \to 0, \qquad \frac{Y_x}{\log x} \to \infty, \qquad w_x \ll \log_2 x \qquad (x \to \infty).$$

Set $\mathcal{B} = \mathcal{B}_x = \{p \in \wp : w_x < p < Y_x\}$ and let $q_0, q_1 \dots, q_{t-1}$ be t integers larger than 1 and for each $i = 0, 1, \dots, t-1$, let α_i be a sharp normal number in base q_i. Consider the sequence of t-tuples $(\underline{y}_n)_{n\geq 1}$ defined by

$$\underline{y}_n := \left(\{\alpha_0 q_0^{\omega_{\mathcal{B}}(n)}\}, \{\alpha_1 q_1^{\omega_{\mathcal{B}}(n+1)}\}, \dots, \{\alpha_{t-1} q_{t-1}^{\omega_{\mathcal{B}}(n+t-1)}\} \right) \in [0,1)^t.$$

If $D_{\lfloor x \rfloor}$ stands for the discrepancy of the set $\{\underline{y}_1, \dots, \underline{y}_{\lfloor x \rfloor}\}$, then $D_{\lfloor x \rfloor} \to 0$ as $x \to \infty$.

Finally, the following result is essentially the case $t = 1$ of the previous theorem.

Corollary 18.2. *Given an integer $q \geq 2$, let α be a sharp q-normal number. Let w_x, Y_x and $\mathcal{B} = \mathcal{B}_x$ be as in Theorem 18.3 and consider the sequence $(y_n)_{n\geq 1}$ defined by $y_n = \{\alpha q^{\omega_{\mathcal{B}}(n)}\}$. Then, the discrepancy $D(y_1, y_2, \dots, y_{\lfloor x \rfloor})$ tends to 0 as $x \to \infty$.*

Preliminary results

Lemma 18.1. *If α is a sharp q-normal number and m a positive integer, then $m\alpha$ is also a sharp q-normal number.*

Proof. Let $x_n \in [0,1)$ for $n = 1, 2, \dots, N$ and consider the corresponding numbers $y_n = \{m x_n\}$ for $n = 1, 2, \dots, N$. If we can prove the inequality

$$D(y_1, y_2, \dots, y_N) \leq m D(x_1, x_2, \dots, x_N), \tag{18.1}$$

the proof of Lemma 18.1 will be complete. In order to prove (18.1), first observe that, for each integer $n \in \{1, 2, \dots, N\}$, we have that $y_n \in [a, b) \subseteq [0, 1)$ if and only if $m x_n \in \bigcup_{\ell=0}^{m-1} [\ell + a, \ell + b)$, which is equivalent to

$$x_n \in \bigcup_{\ell=0}^{m-1} \left[\frac{\ell}{m} + \frac{a}{m}, \frac{\ell}{m} + \frac{b}{m} \right) =: \bigcup_{\ell=0}^{m-1} J_\ell.$$

Since

$$\left| \frac{1}{N} \sum_{\substack{n=1 \\ x_n \in J_\ell}}^{N} 1 - \frac{b-a}{m} \right| \leq D(x_1, x_2, \dots, x_N),$$

it follows that

$$\left| \frac{1}{N} \sum_{\substack{n=1 \\ y_n \in [a,b)}}^{N} 1 - (b-a) \right| \leq \sum_{\ell=0}^{m-1} \left| \frac{1}{N} \sum_{\substack{n=1 \\ x_n \in J_\ell}}^{N} 1 - \frac{b-a}{m} \right| \leq m D(x_1, x_2, \dots, x_N).$$

Taking the supremum of the first two of the above quantities over all possible subintervals $[a, b)$ of $[0, 1)$, inequality (18.1) follows immediately. $\square$

The following result is Lemma 3 in Spiro [59].

Lemma 18.2. *Let B_1, B_2 and B_3 be three fixed positive numbers. Assume that $x \geq 3$ and that both y and ℓ are positive integers satisfying $y \leq B_1 \log_2 x$, $\ell \leq \exp\{\log^{B_2} x\}$ and $\gamma(\ell) \leq \log^{B_3} x$. Then, uniformly for y and ℓ,*

$$\begin{aligned} \pi_\ell(x,y) &:= \#\{n \leq x : \omega(n) = y,\ \mu^2(n) = 1,\ (n,\ell) = 1\} \\ &= \frac{x(\log_2 x)^{y-1}}{(y-1)!\log x}\left\{F\left(\frac{y-1}{\log_2 x}\right)F_\ell\left(\frac{y-1}{\log_2 x}\right) + O_{B_1,B_3}\left(y\frac{(\log_3(16\ell))^3}{(\log_2 x)^2}\right)\right\}, \end{aligned}$$

where

$$F(z) = \frac{1}{\Gamma(z+1)}\prod_p\left(1+\frac{z}{p}\right)\left(1-\frac{1}{p}\right)^z, \qquad F_\ell(z) = \prod_{p\mid\ell}\left(1+\frac{z}{p}\right)^{-1}.$$

Lemma 18.3. *Let w_x, Y_x and $\mathcal{B} = \mathcal{B}_x$ be as in Theorem 18.3 and let $\mathcal{N}(\mathcal{B})$ be the semigroup generated by $\mathcal{B}$. Further let r_x be a function which tends to ∞ as $x \to \infty$, while satisfying the two conditions*

$$r_x \ll \log_3 x \qquad \text{and} \qquad \lim_{x\to\infty}\frac{r_x \log Y_x}{\log x} = 0. \tag{18.2}$$

Moreover, let $D_j \in \mathcal{N}(\mathcal{B})$, $j = 0,1,\dots,t-1$, with $(D_i,D_j) = 1$ for $i \neq j$, and let

$$\mathcal{N}_{D_0,D_1,\dots,D_{t-1}}(x) := \#\left\{n \leq x : D_j \mid n+j, j = 0,1,\dots,t-1,\ \left(\frac{n+j}{D_j},\mathcal{B}\right) = 1\right\}. \tag{18.3}$$

Then, as $x \to \infty$,

$$\frac{1}{x}\#\{n \leq x : D_j \mid n+j, j = 0,1,\dots,t-1 \text{ and } \max(D_0,D_1,\dots,D_{t-1}) > Y_x^{r_x}\} \to 0 \tag{18.4}$$

and, uniformly for $D_j \leq Y_x^{r_x}$, $j = 0,1,\dots,t-1$,

$$\mathcal{N}_{D_0,D_1,\dots,D_{t-1}}(x) = (1+o(1))x\,\kappa(D_0)\kappa(D_1)\cdots\kappa(D_{t-1})L_x^t$$

as $x \to \infty$, where κ is the multiplicative function defined on primes p by

$$\kappa(p) = \frac{1}{p}\cdot\frac{p-t+1}{p-t}$$

and $L_x := \dfrac{\log w_x}{\log Y_x}$.

Proof. First observe that (18.4) is easily proved. We may therefore assume that $D_j \leq Y_x^{r_x}$ for $j = 0,1,\dots,t-1$. In order to use the same notation as in Lemma 0.11, we set

$$\mathcal{B} = \{p_1,\dots,p_s\}, \quad Q = p_1\cdots p_s, \quad E = D_0D_1\cdots D_{t-1}, \quad D_j \mid Q \text{ for } j = 0,1,\dots,t-1.$$

Observe that the condition $D_j \mid n+j$ for ($j = 0,1,\dots,t-1$) in the definition of $\mathcal{N}_{D_0,D_1,\dots,D_{t-1}}(x)$ (see (18.3)) holds for exactly one residue class $n \pmod E$. Letting this residue class be $\ell \pmod E$, we then have

$$\mathcal{N}_{D_0,D_1,\dots,D_{t-1}}(x) = \#\left\{m \leq \left\lfloor\frac{x}{E}\right\rfloor : \left(\frac{\ell+mE+j}{D_j},Q\right) = 1,\ j = 0,1,\dots,t-1\right\} + O(1).$$

Choose $N = \left\lfloor \frac{x}{E} \right\rfloor$ and $f(m) = 1$, while further setting $a_m := \prod_{j=0}^{t-1} \frac{\ell + mE + j}{D_j}$.

Using Lemma 0.11 with $X = N$, we then get that if $d \mid Q$, relation (0.7) can be written as

$$\sum_{\substack{m=1 \\ a_m \equiv 0 \pmod d}}^{N} 1 = \rho(d)N + R(N, d).$$

Here, $\rho(d)$ is multiplicative and defined by

$$\rho(p) = \begin{cases} t/p & \text{if } p \mid Q/E, \\ (t-1)/p & \text{if } p \mid E. \end{cases}$$

On the other hand, $|R(N, d)| \le \tau_t(d) = (t+1)^{\omega(d)}$ (since d is squarefree), which implies that

$$\sum_{\substack{d \mid Q \\ d \le z^3}} 3^{\omega(d)} |R(N, d)| \le \sum_{\substack{d \mid Q \\ d \le z^3}} 3^{\omega(d)} \tau_t(d) \le \sum_{d \le z^3} (3(t+1))^{\omega(d)} \le C z^3 \log^A z,$$

where A and C are suitable constants depending only on t. Again, with the notation used in Lemma 0.11, we have

$$\begin{aligned} S &= \sum_{p \mid Q} \frac{\rho(p)}{1 - \rho(p)} \log p = \sum_{p \mid Q/E} \frac{t \log p}{p(1 - t/p)} + \sum_{p \mid E} \frac{(t-1) \log p}{p(1 - (t-1)/p)} \\ &= t \sum_{p \mid Q/E} \frac{\log p}{p} + (t-1) \sum_{p \mid E} \frac{\log p}{p} + O(1) \\ &= t \sum_{p \mid Q} \frac{\log p}{p} - \sum_{p \mid E} \frac{\log p}{p} + O(1). \qquad (18.5) \end{aligned}$$

Observing that $\sum_{p \mid E} \frac{\log p}{p} \le t \frac{r_x \log Y_x}{w_x} \to 0$ as $x \to \infty$ (because of (18.2)), it follows from (18.5) that

$$S = t \log(Y_x / w_x) + O\left(\frac{r_x \log Y_x}{w_x} \right).$$

Choosing $r = p_s$ and since

$$s = \pi(Y_x) - \pi(w_x) = \pi(Y_x) \left(1 - \frac{\pi(w_x)}{\pi(Y_x)} \right),$$

it follows, since $\log r = \log s + \log \log s + O(1)$, that

$$\log r = \log Y_x + O(\log \log Y_x).$$

Finally, choose $z = Y_x^{8t\nu_x}$, where $\nu_x \to \infty$ very slowly as $x \to \infty$. One can then easily check that the conditions of Lemma 0.11 are satisfied, thus allowing us to conclude that

$$H = \exp\left(-8t\nu_x \left(\log(8\nu_x) - \log\log(8\nu_x) + O(1) \right) \right),$$

thereby implying, since $\nu_x \to \infty$ as $x \to \infty$, that

$$H = H_{x,\nu_x} = o(1) \qquad (x \to \infty). \tag{18.6}$$

Now, writing

$$\prod_{p|Q}(1-\rho(p)) = \prod_{p|Q}\left(1-\frac{t}{p}\right)\cdot\prod_{p|E}\frac{1-\frac{t-1}{p}}{1-t/p} =: \lambda(E),$$

we may conclude from (18.6) that

$$\mathcal{N}_{D_0,D_1,\dots,D_{t-1}}(x) = (1+o(1))\frac{x}{E}\lambda(E) + O(z^3\log^A z). \tag{18.7}$$

It remains to check that the above error term is not too large compared to the main term $\dfrac{x}{E}\lambda(E)$. Indeed, if ν_x tends to ∞ slowly enough, this will guarantee that $z^4 \le \sqrt{x}$, say, while on the other hand, in light of conditions (18.2), we have that, for any small $\varepsilon > 0$,

$$\frac{x}{E} \ge \frac{x}{Y_x^{tr_x}} = \frac{x}{e^{tr_x \log Y_x}} \ge \frac{x}{e^{t\varepsilon \log x}} = \frac{x}{x^{t\,\varepsilon}} > x^{3/4},$$

say. Finally, since $\lambda(E) \ge C/\log Y_x$ for some constant $C > 0$, we may conclude that indeed the error term in (18.7) is of smaller order than the main term of (18.7). Consequently, uniformly for $D_j \le Y_x^{r_x}$, $j = 0,1,\dots,t-1$, we find that

$$\begin{aligned}
\mathcal{N}_{D_0,D_1,\dots,D_{t-1}}(x) &= (1+o(1))\frac{x}{D_0D_1\cdots D_{t-1}}\prod_{\substack{p\nmid D_0D_1\cdots D_{t-1}\\ p\in\mathcal{B}}}\left(1-\frac{t}{p}\right)\cdot\prod_{p|D_0D_1\cdots D_{t-1}}\left(1-\frac{t-1}{p}\right)\\
&= (1+o(1))\frac{x}{D_0D_1\cdots D_{t-1}}\prod_{p|D_0D_1\cdots D_{t-1}}\frac{1-\frac{t-1}{p}}{1-\frac{t}{p}}\cdot\prod_{p\in\mathcal{B}}\left(1-\frac{t}{p}\right).
\end{aligned}$$

Since

$$\prod_{p\in\mathcal{B}}\left(1-\frac{t}{p}\right) = (1+o(1))L_x^t \qquad (x\to\infty),$$

the proof of Lemma 18.3 is complete. □

The following result is Lemma 1 in our paper [30].

Lemma 18.4. *Let $g(x) = \alpha x^k + \alpha_{k-1}x^{k-1} + \dots + \alpha_1 x + \alpha_0 \in \mathbb{R}[x]$ be a polynomial of positive degree, where α be a non-Liouville number. Then,*

$$\sup_{U\ge 1}\frac{1}{N}\left|\sum_{n=U+1}^{U+N} e(g(n))\right| \to 0 \qquad \textit{as } N\to\infty.$$

Lemma 18.5. *Assume that the set of natural integers $\mathbb{N}$ is written as a disjoint union of sets N_K, where K runs through the elements of a particular set $\mathcal{P}$ of positive integers, that is, $\mathbb{N} = \bigcup_{K\in\mathcal{P}} N_K$. Assume that, for each $K \in \mathcal{P}$, the counting function $N_K(x) := \#\{n \le x : n \in N_K\}$ satisfies*

$$\lim_{x\to\infty}\frac{N_K(x)}{x} = c_K,$$

where the c_K *are positive real numbers such that* $\sum_{K\in\mathcal{P}} c_K = 1$. *Moreover, let* $(x_n)_{n\ge1}$ *be a sequence of real numbers which is such that, for each* $K \in \mathcal{P}$, *the corresponding sequence* $(x_n)_{n\in N_K}$ *is uniformly distributed modulo 1, that is, for each integer* $h \ge 1$,

$$S_K^{(h)}(x) := \sum_{\substack{n\le x\\ n\in N_K}} e(hx_n) = o(N_K(x)) \qquad \text{as } x\to\infty. \tag{18.8}$$

Then, the sequence $(x_n)_{n\ge1}$ *is uniformly distributed modulo 1.*

Proof. According to an old and very important result of Weyl [65], a sequence $(x_n)_{n\ge1}$ is uniformly distributed modulo 1 if for every non negative integer h,

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} e(hx_n) = 0.$$

Therefore, in light of Weyl's criteria, we only need to prove that, for each positive integer h,

$$S^{(h)}(x) := \sum_{K\in\mathcal{P}} S_K^{(h)}(x) \to 0 \text{ as } x\to\infty. \tag{18.9}$$

Given any $z > 0$ and writing

$$S^{(h)}(x) = \sum_{\substack{K\in\mathcal{P}\\ K<z}} S_K^{(h)}(x) + \sum_{\substack{K\in\mathcal{P}\\ K\ge z}} S_K^{(h)}(x),$$

it follows that

$$\left|\frac{S^{(h)}(x)}{x}\right| \le \sum_{K<z,K\in\mathcal{P}} \frac{N_K(x)}{x}\cdot\frac{1}{N_K(x)}|S_K^{(h)}(x)| + \frac{1}{x}\#\left\{n\le x : n\in \bigcup_{K\in\mathcal{P},K\ge z} N_K\right\}. \tag{18.10}$$

Since, in light of (18.8), we have that $\frac{1}{N_K(x)}|S_K^{(h)}(x)| = o(1)$ as $x\to\infty$, it follows from (18.10) that, for some $C>0$,

$$\limsup_{x\to\infty}\left|\frac{S^{(h)}(x)}{x}\right| \le C\cdot(\sum_{K<z,K\in\mathcal{P}} c_K)\cdot o(1) + \sum_{K\ge z,K\in\mathcal{P}} c_K,$$

which is as small as we want provided z is chosen large enough, thus proving (18.9). □

Proof of Theorem 18.1

An integer n is called *squarefull* if $p \mid n$ implies that $p^2 \mid n$. Let $\mathcal{P}$ be the set of all squarefull numbers. For convenience, we let $1\in\mathcal{P}$. To each squarefull number K, we associate the set $N_K := \{n = Km : (m,K)=1,\ \mu^2(m)=1\}$, where μ stands for the Möbius function. Since each positive integer n belongs to one and only one such set N_K, we have that

$$\mathbb{N} = \bigcup_{K\in\mathcal{P}} N_K.$$

For any $n \in N_K$, we have $\tau_q(n) = \tau_q(Km) = \tau_q(K) q^{\omega(m)}$.

Now, in light of Lemma 18.5, the theorem will follow if we can prove that for each fixed $K \in \mathcal{P}$,

$$\text{the sequence } (\{\alpha \tau_q(n)\})_{n \in N_K} \text{ is uniformly distributed modulo 1 over } N_K. \tag{18.11}$$

To prove this last statement, we use Lemma 18.2. First, observe that for $\ell = K$ fixed, we have that $\gamma(\ell) = \gamma(K)$ is bounded and that we can also assume that, given any function δ_x which tends to 0 sufficiently slowly as $x \to \infty$, say with $1/\delta_x < \log_3 x$,

$$|y - \log_2 x| \le \frac{1}{\delta_x} \sqrt{\log_2 x}, \tag{18.12}$$

so that each of the two quantities $F\left(\dfrac{y-1}{\log_2 x}\right)$ and $F_\ell\left(\dfrac{y-1}{\log_2 x}\right)$ is equal to $1 + o(1)$ as $x \to \infty$ for y in the range (18.12). From there and the fact that α is a sharp normal number, it is clear that (18.11) follows.

Proof of Theorem 18.2

Given a squarefull number K, let N_K and $\mathcal{P}$ be as in the proof of Theorem 18.1. Any integer $n \in N_K$ can be written as $n = Km$, where $(K, m) = 1$ and $\mu^2(m) = 1$. Moreover, write $\tau(K) = k_1 \cdot 2^{\rho_K}$ for some odd positive integer k_1 and some non negative integer ρ_K. From this set up, it follows that $\tau(n) = \tau(Km) = k_1 \cdot 2^{\rho_K + \omega(m)}$, from which it follows that

$$\tau(\tau(n)) = \tau(k_1) \left(\omega(m) + \rho_K + 1\right). \tag{18.13}$$

Now, for $n \in N_K$ with $\omega(m) = t$, we have, using (18.13),

$$g(\tau(\tau(n))) = \alpha \tau(k_1)^k (t + \rho_K + 1)^k + \cdots = \alpha \tau(k_1)^k t^k + P_{k-1}(t), \tag{18.14}$$

where $P_{k-1}(t)$ stands for some polynomial of degree no larger than $k - 1$.

We shall now use Weyl's criteria, already stated in the proof of Lemma 18.5. So, let h be an arbitrary positive integer. For each $K \in \mathcal{P}$, set

$$S_K(x) := \sum_{\substack{n \le x \\ n \in N_K}} e(hg(\tau(\tau(n)))).$$

In light of (18.14), we have, writing t for $\omega(m)$,

$$S_K(x) = \sum_{t \ge 1} e(h\alpha \tau(k_1)^k t^k + P_{k-1}(t)) \cdot \pi_K(x, t),$$

were $\pi_k(x, t)$ was defined in Lemma 18.2. Setting $R(t) := \alpha \tau(k_1)^k t^k + P_{k-1}(t)$, we may write the above as

$$S_K(x) = \sum_{t \ge 1} e(hR(t)) \cdot \pi_K(x, t).$$

Our goal will be to establish that, given any $K \in \mathcal{P}$,

$$S_K(x) = o(x) \qquad (x \to \infty). \tag{18.15}$$

If we can accomplish this, then, in light of Lemma 18.5, the proof of Theorem 18.2 will be complete.

To prove (18.15), we first observe that

$$\sum_{\substack{t\geq 1\\ |t-\log_2 x|>\sqrt{\log_2 x}/\varepsilon_x}} \pi_K(x,t) = o(x) \qquad (x\to\infty) \tag{18.16}$$

and furthermore that

$$\max_{\substack{t_1\\ |t_1-\log_2 x|\leq\sqrt{\log_2 x}/\varepsilon_x}} \max_{\substack{t_2\\ |t_2-t_1|\leq\varepsilon_x\sqrt{\log_2 x}}} \left|\frac{\pi_K(x,t_1)}{\pi_K(x,t_2)}-1\right| \to 0 \text{ as } x\to\infty. \tag{18.17}$$

Now, consider the sequence of real numbers $(z_n)_{n\geq 0}$ defined by

$$z_0 = \log_2 x - \frac{\sqrt{\log_2 x}}{\varepsilon_x} \text{ and for each } m\geq 1 \text{ by } z_m = z_{m-1}+\varepsilon_x\sqrt{\log_2 x},$$

and, setting $M = \left\lfloor \dfrac{(2/\varepsilon_x)\sqrt{\log_2 x}}{\varepsilon_x\sqrt{\log_2 x}}\right\rfloor = \left\lfloor \dfrac{2}{\varepsilon_x^2}\right\rfloor$, further consider the intervals

$$I_j := [\lfloor z_j\rfloor, z_{j+1}) \qquad (j=0,1,\ldots,M).$$

Now, observe that, uniformly for $j\in\{0,1,\ldots,M\}$, as $x\to\infty$,

$$\left|\sum_{t\in I_j} e(hR(t))\pi_K(x,t) - \pi_K(x,\lfloor z_j\rfloor)\sum_{t\in I_j} e(hR(t))\right| \leq o(1)\sum_{t\in I_j}\pi_K(x,t). \tag{18.18}$$

Using the fact that the above intervals I_j are all of the same length, say $\mathcal{L}=\mathcal{L}_x$, it follows from Lemma 18.4 that, uniformly for $j\in\{0,1,\ldots,M\}$,

$$\frac{1}{\mathcal{L}}\sum_{t\in I_j} e(hR(t)) \to 0 \qquad (x\to\infty). \tag{18.19}$$

Combining (18.18) and (18.19) allows us to conclude that

$$\left|\sum_{j=0}^{M}\sum_{t\in I_j} e(hR(t))\pi_K(x,t)\right| = o(x).$$

Using this last estimate and recalling estimates (18.16) and (18.17), it follows that estimate (18.15) holds, thus completing the proof of Theorem 18.2.

Proof of Theorem 18.3

Given a large number x, let $T=T_x := \displaystyle\sum_{w_x\leq p\leq Y_x}\frac{1}{p}$, and observe that

$$T = \log\left(\frac{\log Y_x}{\log w_x}\right) + o(1) = \log L_x^{-1} + o(1) \qquad (x\to\infty). \tag{18.20}$$

Further let δ_x be a function which tends to 0 as $x \to \infty$, but not too fast in the sense that $\dfrac{1}{\delta_x} = O(\log_2 T)$.

We will be using the fact that, as a consequence of Lemma 18.3, as $x \to \infty$,

$$\frac{1}{x}\#\{n \le x : \omega_{\mathcal{B}}(n+j) = k_j,\ j = 0, 1, \ldots, t-1\} = (1 + o(1)) \prod_{j=0}^{t-1} \frac{1}{x}\#\{n \le x : \omega_{\mathcal{B}}(n) = k_j\}$$

uniformly for positive integers $k_0, k_1, \ldots, k_{t-1}$ satisfying $|k_j - T| \le \dfrac{1}{\delta_x}\sqrt{T}$ and also that

$$\frac{1}{x}\#\left\{n \le x : \frac{|\omega_{\mathcal{B}}(n) - T|}{\sqrt{T}} > \frac{1}{\delta_x}\right\} \to 0 \quad \text{as} \quad x \to \infty.$$

We begin by obtaining an upper bound for the sum

$$S := \sum_{\substack{D_0, D_1, \ldots, D_{t-1} \\ D_\nu \in \mathcal{N}(\mathcal{B}),\ D_\nu \le Y_x^{r_x} \\ (D_i, D_j) > 1 \text{ for some } i \ne j}} \kappa(D_0)\kappa(D_1)\cdots\kappa(D_{t-1}) L_x^t,$$

where r_x is as in Lemma 18.3, keeping in mind that we allow the above sum to run only over those $D_\nu \le Y_x^{r_x}$, because, as was shown in (18.4), the total contribution of those terms for which at least one of the D_ν exceeds $Y_x^{r_x}$ is negligible. So, let us fix i, j and consider the sum

$$S_{i,j} := \sum_{\substack{D_i, D_j \in \mathcal{N}(\mathcal{B}) \\ (D_i, D_j) > 1 \\ D_i, D_j \le Y_x^{r_x}}} \kappa(D_i)\kappa(D_j) L_x^2.$$

Writing $D_i = UD_i'$ and $D_j = VD_j'$, where U and V have the same prime divisors, $(D_i', D_j') = (U, D_i') = (V, D_j') = 1$, we then have

$$\kappa(D_i)\kappa(D_j) = \kappa(D_i')\kappa(D_j')\kappa(U)\kappa(V).$$

Observe also that, for some positive constant c_1, we have

$$\kappa(U)\kappa(V) < c_1 \left(\prod_{p | U} p^2\right)^{-1}.$$

From these observations, it follows that, for some positive constant c_2,

$$\begin{aligned} S_{i,j} &< c_2 \sum_{\substack{m=2 \\ m \in \mathcal{N}(\mathcal{B})}}^{\infty} \frac{1}{m^2} \cdot \left(L_x \sum_{D \in \mathcal{N}(\mathcal{B})} \kappa(D)\right)^2 \\ &= c_2 \sum_{\substack{m=2 \\ m \in \mathcal{N}(\mathcal{B})}}^{\infty} \frac{1}{m^2} \cdot \prod_{p \in \mathcal{B}} (1 + \kappa(p))^2 \cdot L_x^2. \qquad (18.21) \end{aligned}$$

On the other hand, using (18.20),

$$\begin{aligned}\prod_{p\in\mathcal{B}}(1+\kappa(p)) &= \exp\left(\sum_{p\in\mathcal{B}}\log(1+\kappa(p))\right)\\ &= \exp\left(\sum_{p\in\mathcal{B}}\frac{1}{p}+O(1)\right)=\exp(T+O(1))\\ &= \exp(-\log L_x+O(1)).\end{aligned}$$

Using this last estimate and the fact that

$$\sum_{\substack{m=2\\ m\in\mathcal{N}(\mathcal{B})}}^{\infty}\frac{1}{m^2}<\sum_{m>w_x}\frac{1}{m^2}<\frac{2}{w_x},$$

say, it follows from (18.21) that, for some positive constant c_3,

$$S_{i,j}\le\frac{c_3}{w_x}\cdot\frac{1}{L_x^2}\cdot L_x^2=\frac{c_3}{w_x}.$$

Moreover, in light of the fact that

$$L_x\sum_{\substack{D_\nu\in\mathcal{N}(\mathcal{B})\\ D_\nu\le Y_x^{r_x}\\ \text{for every }\ \nu=0,1,\dots,t-1}}\kappa(D_\nu)\le c_4$$

for some absolute constant $c_4>0$, we obtain after gathering our estimates that

$$S=O\left(\frac{1}{w_x}\right). \tag{18.22}$$

Now, given arbitrary subsets $E_0,E_1,\dots,E_{t-1}$ of $\{D:D\in\mathcal{N}(\mathcal{B}),\ D\le Y_x^{r_x}\}$, we have, as $x\to\infty$, in light of (18.22),

$$\sum_{\substack{D_0\in E_0,\dots,D_{t-1}\in E_{t-1}\\ (D_i,D_j)=1\ \text{ for }\ i\ne j}}\kappa(D_0)\kappa(D_1)\cdots\kappa(D_{t-1})L_x^t=\prod_{j=0}^{t-1}\left(L_x\sum_{D\in E_j}\kappa(D)\right)+o(x). \tag{18.23}$$

Observe that to the discrepancy $D_N:=D(x_1,\dots,x_N)$ of the real numbers $x_1,\dots,x_N$ (as defined by (0.1)), one can associate the so-called *star discrepancy*

$$D_N^*=D^*(x_1,\dots,x_N):=\sup_{0\le\beta<1}\left|\frac{1}{N}\sum_{\substack{i=1\\ \{x_i\}<\beta}}^{N}1-\beta\right|$$

and establish that $D_N^*\le D_N\le 2D_N^*$. In light of this observation, defining the function $H_u:[0,1)\to\{0,1\}$ by

$$H_u(y):=\begin{cases}1 & \text{if}\quad 0\le y<u,\\ 0 & \text{if}\quad u\le y<1,\end{cases} \tag{18.24}$$

one can easily establish that

$$D_N^* = \max_{u\in[0,1)} \left(\frac{1}{N} \sum_{n=1}^{N} H_u(x_n) - u \right),$$

implying that if we can show that this last expression tends to 0 as $N \to \infty$, it will allow us to conclude that $D_N = D_{\lfloor x \rfloor} \to 0$ as $N \to \infty$.

To do so, given real numbers $u_0, u_1, \ldots, u_{t-1} \in [0,1)$, choose

$$E_j := \{D \in \mathcal{N}(\mathcal{B}) : |\omega(D) - T| \le \sqrt{T}/\delta_x,\ D \le Y_x^{r_x},\ H_{u_j}(\{\alpha_j q_j^{\omega(D)}\}) = 1\}$$

and apply estimate (18.23).

It follows from this that, if we can prove that

$$\left(\{\alpha_j q_j^{\omega_{\mathcal{B}}(n+j)}\}\right)_{n\ge 1} \qquad \text{is uniformly distributed modulo 1} \tag{18.25}$$

for each $j = 0, 1, \ldots, t-1$, it will imply that, as $x \to \infty$,

$$\sum_{\substack{D_j \in \mathcal{N}(\mathcal{B}) \\ D_j \le Y_x^{r_x}}} H_{u_j}(\{\alpha_j q_j^{\omega(D_j)}\}) \kappa(D_j) L_x \to u_j \qquad (j = 0, 1, \ldots, t-1),$$

thus allowing us to conclude that

$$\prod_{j=0}^{t-1} \left(\sum_{\substack{D_j \in \mathcal{N}(\mathcal{B}) \\ D_j \le Y_x^{r_x}}} H_{u_j}(\{\alpha_j q_j^{\omega(D_j)}\}) \kappa(D_j) L_x \right) = u_0 u_1 \cdots u_{t-1} + o(1) \qquad (x \to \infty),$$

thereby establishing that the sequence $(\underline{y}_n)_{n\ge1}$ is uniformly distributed mod $[0,1)^t$.

Thus, it remains to prove (18.25). To do so, it is enough to prove Corollary 18.2.

Proof of Corollary 18.2

Let

$$A(n) := \prod_{\substack{p^a \| n \\ p \in \mathcal{B}}} p^a \quad \text{and} \quad M_x := \prod_{p \in \mathcal{B}} \left(1 - \frac{1}{p}\right).$$

For every $D \in \mathcal{N}(\mathcal{B})$ with $D \le Y_x^{r_x}$, we have

$$\#\{n \le x : A(n) = D\} = \left(1 + O\left(\frac{1}{\log w_x}\right)\right) \frac{x}{D} M_x \qquad (x \to \infty),$$

from which it follows that, as $x \to \infty$,

$$\begin{aligned} B_k(x) &:= \frac{1}{x}\#\{n \le x : \omega_{\mathcal{B}}(n) = k\} \\ &= (1 + o(1)) M_x \sum_{\substack{D \in \mathcal{N}(\mathcal{B}) \\ \omega(D) = k}} \frac{1}{D} + O(U_k(x)), \end{aligned} \tag{18.26}$$

where

$$U_k(x) = M_x \sum_{\substack{D\in\mathcal{N}(\mathcal{B})\\ \omega(D)=k\\ D>Y_x^{r_x}}} \frac{1}{D} + \frac{1}{x}\#\{n \le x : A(n) > Y_x^{r_x},\ \omega(A(n)) = k\},$$

thereby implying that

$$\sum_{k\ge 1} U_k(x) \to 0 \qquad \text{as } x\to\infty. \tag{18.27}$$

For each positive integer k, let $z_k = \{\alpha q^k\}$. Further, let $H_u(y)$ be the function defined in the proof of Theorem 18.3 (see (18.24)).

In light of estimate (18.26), we have, as $x\to\infty$,

$$\begin{aligned} R_x &:= \frac{1}{x}\sum_{n\le x} H_u(y_n) = \sum_{k\ge1} H_u(z_k)B_k(x) \\ &= (1+o(1))\sum_{k\ge1} H_u(z_k) M_x \sum_{\substack{D\in\mathcal{N}(\mathcal{B})\\ \omega(D)=k}} \frac{1}{D} + O\left(\sum_{k\ge1} U_k(x)\right). \end{aligned} \tag{18.28}$$

Observing that

$$\sum_{a\ge1, p\in\mathcal{B}} \frac{1}{ap^a} = \sum_{p\in\mathcal{B}} \frac{1}{p} + O\left(\frac{1}{w_x}\right)$$

allows us to write that

$$M_x = \exp\left\{-\sum_{p\in\mathcal{B}}\frac{1}{p} + O\left(\frac{1}{w_x}\right)\right\} = \exp\left\{-T + O\left(\frac{1}{w_x}\right)\right\}, \tag{18.29}$$

say. Hence, it follows from (18.27), (18.28) and (18.29) that

$$R_x = (1+o(1))\sum_{k\ge1} H_u(z_k)\exp\{-T\}\cdot\frac{T^k}{k!} + o(1) \qquad (x\to\infty). \tag{18.30}$$

Now, since, for any function δ_x which tends to 0 as $x\to\infty$,

$$\sum_{\frac{|k-T|}{\sqrt{T}} > \frac{1}{\delta_x}} \exp\{-T\}\cdot\frac{T^k}{k!} \to 0 \qquad \text{as } x\to\infty,$$

we obtain that (18.30) can be replaced by

$$R_x = (1+o(1)) \sum_{\frac{|k-T|}{\sqrt{T}} \le \frac{1}{\delta_x}} H_u(z_k)K_k + o(1) \qquad (x\to\infty), \tag{18.31}$$

where $K_k := \exp\{-T\}\cdot\dfrac{T^k}{k!}$.

On the other hand, observe that for any function ε_x which tends to 0 as $x \to \infty$, we have

$$\max_{\substack{k_1 \\ \left|\frac{k_1-T}{\sqrt{T}}\right| \le \frac{1}{\delta_x}}} \max_{\substack{k_2 \\ |k_2-k_1|<\varepsilon_x\sqrt{T}}} \left| \frac{K_{k_2}}{K_{k_1}} - 1 \right| \to 0 \qquad \text{as } x \to \infty. \tag{18.32}$$

Let us now subdivide the interval $[T - \sqrt{T}/\delta_x, T + \sqrt{T}/\delta_x]$ into intervals $I_1, I_2, \ldots, I_s$, where $s = \lfloor 2/(\delta_x \varepsilon_x) \rfloor$, each of length $\varepsilon_x \sqrt{T}$. Since, in light of (18.32), we have

$$\max_{j=1,\ldots,s} \max_{k_1,k_2 \in I_j} \left| \frac{K_{k_2}}{K_{k_1}} - 1 \right| \to 0 \qquad \text{as } x \to \infty \tag{18.33}$$

and since α is a sharp q-normal number, it follows that, for each $j \in \{1, \ldots, s\}$,

$$\sum_{k \in I_j} H_u(z_k) = (1 + o(1)) \sum_{k \in I_j} 1 \qquad (x \to \infty).$$

Using this last statement in (18.31), recalling (18.33), and writing $|I_j|$ for the length of the interval I_j, we obtain that, as $x \to \infty$,

$$\begin{aligned} R_x &= (1+o(1)) \sum_{j=1}^{s} \sum_{k \in I_j} H_u(z_k) K_k \\ &= (1+o(1)) \sum_{j=1}^{s} \sum_{k \in I_j} H_u(z_k) \left(\frac{1}{|I_j|} \sum_{k_1 \in I_j} K_{k_1} \right) \\ &= (1+o(1)) \sum_{j=1}^{s} \left(\frac{1}{|I_j|} \sum_{k_1 \in I_j} K_{k_1} \right) \sum_{k \in I_j} H_u(z_k) \\ &= (1+o(1)) \sum_{j=1}^{s} \left(\frac{1}{|I_j|} \sum_{k_1 \in I_j} K_{k_1} \right) (1+o(1)) u |I_j| \\ &= (1+o(1)) u \sum_{j=1}^{s} \sum_{k \in I_j} K_k \\ &= (1+o(1)) u \sum_{\substack{k \\ |k-T| \le \sqrt{T}/\delta_x}} K_k \\ &= (1+o(1)) u. \end{aligned}$$

Since this last estimate holds for every real $u \in [0, 1)$, it follows that $R_x = o(1)$ as $x \to \infty$ and the proof of Corollary 18.2 is complete.

Final remarks

Using the same techniques as above, one could prove the following result regarding the discrepancy of a t-tuples sequence.

Let $f_1, f_2, \ldots, f_t \in \mathbb{R}[x]$ be polynomials of positive degree such that the coefficient of the leading term of each f_j is some non-Liouville number α_j. Moreover, let $a_1, a_2, \ldots, a_t$ be distinct integers and let $\mathcal{B}$ be as in Theorem 18.3. Set

$$\underline{y}_n := (f_1(\omega_{\mathcal{B}}(n+a_1)), f_2(\omega_{\mathcal{B}}(n+a_2)), \ldots, f_t(\omega_{\mathcal{B}}(n+a_t)))\,.$$

Then,

$$D(\underline{y}_1, \underline{y}_2, \ldots, \underline{y}_{\lfloor x \rfloor}) \to 0 \qquad \text{as } x \to \infty$$

and similarly, if p_i and $\pi(x)$ stand respectively for the i-th prime and the number of primes not exceeding x,

$$D(\underline{y}_2, \underline{y}_3, \underline{y}_5, \ldots, \underline{y}_{p_{\pi(x)}}) \to 0 \qquad \text{as } x \to \infty.$$

XIX. Distinguishing between sharp and non-sharp normal numbers [33]
(Mathematica Pannonica, 2018)

In 2015, De Koninck, Kátai and Phong introduced the concept of sharp normal numbers and proved that almost all real numbers are sharp normal numbers in the sense of the Lebesgue measure. They also proved that although the Champernowne number is normal in base 2, it is not sharp in that base. Here, we prove that various real numbers are sharp normal numbers, while others are not.

Given an integer $q \geq 2$, De Koninck, Kátai and Phong [31] introduced the concept of *base q strong normal number*, shortly after called *base-q sharp normal number*. In particular, they showed that, given a fixed base $q \geq 2$, the Lebesgue measure of the set of all those real numbers $\alpha \in [0,1]$ which are not sharp q-normal is equal to 0.

In a more recent paper [32], we proved that, given a fixed integer $q \geq 2$ and letting $\tau_q(n)$ stand for the number of ways of writing n as a product of q positive integers, then, if α is a sharp normal number in base q, the sequence $(\alpha\tau_q(n))_{n\geq 1}$ is uniformly distributed modulo 1. In that same paper, other properties of sharp normal numbers were established.

Given an integer $q \geq 2$ and a real number $\gamma \in (0,1)$, we will say that a real number α is a γ-sharp normal number in base q if, by setting $x_n = \{\alpha q^n\}$ for $n = 1, 2, \ldots$ and

$$M = M_N = \lfloor \delta_N N^\gamma \rfloor, \qquad \text{where } \delta_N \to 0 \text{ and } \delta_N \log N \to \infty \text{ as } N \to \infty, \tag{19.1}$$

we have that

$$D(x_{N+1}, \ldots, x_{N+M}) \to 0 \qquad \text{as} \quad N \to \infty$$

for every choice of δ_N satisfying (19.1).

Observe that in [31], it was shown that the binary Champernowne number

$$\theta := 0.1\,10\,11\,100\,101\,110\,111\,1000\,1001\,1010\,1011\,1100\,1101\,1110\,1111\,\ldots$$

is not a sharp normal number. Similarly, one can prove that θ is not a γ-sharp normal number for any $\gamma \in (0,1)$.

Here, we further explore the topic of γ-sharp normal numbers.

MAIN RESULTS

From here on, we let q stand for a fixed integer ≥ 2. Let $\wp = \{p_1, p_2, \ldots\}$ stand for the set of all primes. Given a positive integer n, we let $\overline{n}$ stand for the concatenation of the base q digits of the number n.

In 1946, Copeland and Erdős [10] showed that the now called *Copeland-Erdős number*

$$\theta := 0.\overline{p_1}\,\overline{p_2}\,\overline{p_3}\ldots$$

is q-normal. Here, we will prove the following.

Theorem 19.1. *Given any $\gamma \in (0,1)$, the number θ is not a binary γ-sharp normal number.*

In the same 1946 paper, Copeland and Erdős conjectured that if $f \in \mathbb{Z}[x]$ is a polynomial of positive degree such that $f(x) > 0$ for $x > 0$, then the number $\beta = 0.\overline{f(1)}\,\overline{f(2)}\,\overline{f(3)}\ldots$ is a normal number in base 10. This was proved to be true in 1952 by Davenport and Erdős [11]. Here we prove the following.

Theorem 19.2. *Given a positive integer r, the real number*

$$\beta = 0.\overline{1^r}\,\overline{2^r}\,\overline{3^r}\,\ldots$$

is not a binary sharp normal number.

Fix an integer $q \geq 2$. Given an integer $n \geq 2$, let $p(n)$ stand for its smallest prime factor and write $\overline{p(n)}$ for the concatenation of the digits of $p(n)$ in base q. In 2014, we showed [22] that the number $\eta = 0.\overline{p(2)}\,\overline{p(3)}\,\overline{p(4)}\ldots$ is a q-normal number. Here, we prove the following.

Theorem 19.3. *Given an arbitrary real number $\gamma \in (0,1)$, the real number*

$$\eta = 0.\overline{p(2)}\,\overline{p(3)}\,\overline{p(4)}\ldots$$

is a γ-sharp normal number in base q.

Fix an integer $q \geq 2$. Let $\wp_0, \wp_1, \ldots, \wp_{q-1}, \mathcal{R}$ be disjoint sets of primes such that

$$\wp = \wp_0 \cup \wp_1 \cup \cdots \cup \wp_{q-1} \cup \mathcal{R}$$

and such that $\#\mathcal{R} < \infty$. Assume also that

$$\max_{0 \leq i < j \leq q-1} \max_{\frac{x}{\log^5 x} \leq y \leq x} \left| \frac{\pi([x, x+y] \cap \wp_i)}{\pi([x, x+y] \cap \wp_j)} - 1 \right| \to 0 \quad \text{as } x \to \infty.$$

More over let Λ stand for the empty word and for each $p \in \wp$, let

$$H(p) := \begin{cases} \ell & \text{if} \quad p \in \wp_\ell, \\ \Lambda & \text{if} \quad p \in \mathcal{R}. \end{cases}$$

Given an integer $n \geq 2$ written as $n = q_1^{a_1} \cdots q_r^{a_r}$, where $q_1 < \cdots < q_r$ are primes and each $a_i \in \mathbb{N}$, let

$$S(n) := H(q_1) \dots H(q_r).$$

Further set $S(1) = 1$. In 2011, we showed [14] that the number $0.\text{Concat}(\overline{S(n)} : n \in \mathbb{N})$ is a q-normal number. Here, we prove the following.

Theorem 19.4. *Given an arbitrary real number $\gamma \in (0,1)$, the real number*

$$0.\overline{S(1)}\,\overline{S(2)}\,\overline{S(3)}\,\dots$$

is a γ-sharp normal number in base q.

We also have the following.

Theorem 19.5. *Fix an integer $q \geq 2$. Given any pair of prime numbers $u < v$, let $\epsilon(u,v)$ stand for the unique integer $\ell \in \{0, 1, \dots, q-1\}$ such that*

$$\frac{\ell}{q} \leq \frac{\log u}{\log v} < \frac{\ell+1}{q}.$$

For each positive integer $n = q_1^{a_1} \cdots q_r^{a_r}$, let

$$\xi(n) = \begin{cases} \epsilon(q_1, q_2)\, \epsilon(q_2, q_3) \dots \epsilon(q_{r-1}, q_r) & \text{if} \quad \omega(n) \geq 2, \\ \Lambda & \text{if} \quad \omega(n) \leq 1. \end{cases}$$

Then, given any real number $\gamma \in (0,1)$, the number

$$0.\mathit{Concat}(\xi(n) : n \in \mathbb{N})$$

is a γ-sharp normal number in base q.

Let $\mathcal{P}$ be a set of primes and set $\pi_{\mathcal{P}}(x) := \#\{p \leq x : p \in \mathcal{P}\}$. Moreover, let $\mathcal{N} = \{n_1, n_2, \dots\}$ be the semi-group generated by $\mathcal{P}$. Let $F(x) \in \mathbb{Z}[x]$ be a monic polynomial of positive degree t. Assume that there exists a positive constant τ such that

$$\lim_{x \to \infty} \frac{\pi_{\mathcal{P}}(x)}{\text{li}(x)} = \tau,$$

where $\text{li}(x) := \displaystyle\int_2^x \frac{dt}{\log t}$. Fix an integer $q \geq 2$. Given a positive integer n, let $\overline{n}$ stand for the concatenation of the digits of n in base q and consider the real number

$$\eta_0 = 0.\overline{F(n_1)}\,\overline{F(n_2)}\,\overline{F(n_3)}\,\dots$$

It was proved by German and Kátai [40] that η_0 is a q-normal number. Their proof uses essentially the same method as the one used in the paper of Bassily and Kátai [2], along with other ideas of E. Wirsing, H. Davenport and L.K. Hua. Using these ideas, one could prove the following.

Theorem 19.6. *The q-normal number η_0 is not sharp.*

Proof of Theorem 19.1

First observe that it has been proved by Montgomery [53] that, given any small $\varepsilon > 0$,

$$\pi(x+y) - \pi(x) = (1+o(1))\frac{y}{\log x} \quad \text{uniformly for} \quad x^{\frac{7}{12}+\varepsilon} \le y \le x. \tag{19.2}$$

Let $t \ge 2$ be an integer sufficiently large so that $\gamma \le 1 - \dfrac{1}{2^t}$. Moreover, for each integer $k \ge 1$, let $x_k = 2^{2^k}$ and $y_k = x_k^{1-1/2^t} = 2^{2^k - 2^{k-t}}$. Then, let $q_1 < q_2 < \cdots < q_R$ be all the primes located in the interval $(x_k, x_k + y_k]$, where clearly $R = R(k)$. For each $j \in \{1, \ldots, R\}$, let a_j be defined implicitly by $q_j = x_k + a_j$. Then, $a_j \le y_k$ and in light of (19.2), we have

$$R = \pi(x_k + y_k) - \pi(x_k) = (1+o(1))y_k/\log x_k \qquad (k \to \infty).$$

Given an integer $n \ge 1$, let $\alpha(n)$ stand for the sum of its binary digits. Adopting the argument of Erdős and Copeland used in [10], we can say that for every arbitrarily small $\delta > 0$, there exists a constant $\kappa = \kappa(\delta) > 0$ such that

$$\#\left\{ m \le y_k : \alpha(m) > (1+\delta)2^{k-1}\left(1 - \frac{t}{2}\right)\right\} < y_k^{1-\kappa},$$

provided k is sufficiently large. It follows from this observation that

$$\begin{aligned} T &:= \sum_{j=1}^{R} \alpha(q_j) = R + \sum_{j=1}^{R} \alpha(a_j) \\ &\le R + (1+\delta)2^{k-1}\left(1 - \frac{t}{2}\right) R + 2^k y_k^{1-\kappa} \le (1+2\delta)2^{k-1}\left(1 - \frac{t}{2}\right) R, \end{aligned} \tag{19.3}$$

provided k is large enough.

Letting $\lambda(n)$ stand for the number of binary digits of n and observing that $\lambda(\overline{q_j}) = 2^k + 1$ for $j = 1, \ldots, R$, it follows from (19.3) that

$$T < \left(\frac{1}{2} - \varepsilon\right) \sum_{j=1}^{R} \lambda(\overline{q_j}). \tag{19.4}$$

However, if θ were to be a binary γ-sharp normal number, we would need to have

$$\frac{T}{\sum_{j=1}^{R} \lambda(\overline{q_j})} \to \frac{1}{2} \qquad (k \to \infty),$$

which clearly contradicts (19.4). We may therefore conclude that θ is not a binary γ-sharp normal number.

Proof of Theorem 19.2

Given an integer $n \in [2^k, 2^{k+1})$, write its binary expansion as $n = \sum_{\nu=0}^{k} \epsilon_\nu(n) 2^\nu$. In [2], the following result was proved.

Lemma 19.1. *Let* $N = \left\lfloor \dfrac{\log x}{\log 2} \right\rfloor$ *and let* $F(x) \in \mathbb{Z}[x]$ *be a polynomial of positive degree* r *such that* $F(n) > 0$ *for* $n \geq 1$. *If*

$$N^{1/3} \leq \ell \leq rN - N^{1/3},$$

then,

$$\frac{1}{x}\#\{n \leq x : \epsilon_\ell(F(n)) = 1\} = \frac{1}{2} + O\left(\frac{1}{\log^A x}\right),$$

where A *is some positive constant which may depend on the particular polynomial* F.

In order to prove Theorem 19.2, we use Lemma 19.1 with $F(n) = n^r$.

Let $M = M_k := 2^k$ and let

$$f(m) = (4M^2 + m)^r = (2M)^{2r} + g(m), \tag{19.5}$$

where

$$g(m) = \sum_{j=0}^{r-1} \binom{r}{j} (2M)^{2j} m^{r-j}.$$

Recalling that $\alpha(n)$ stands for the sum of the binary digits of n, whereas $\lambda(n)$ stands for the number of binary digits of n, our goal will be to estimate $A_M := \sum_{m=1}^{M} \alpha(f(m))$ and to compare it with $L_M := \sum_{m=1}^{M} \lambda(f(m))$.

Now, let

$$I_0 = [0, 2k], \quad I_1 = [2k+1, 4k], \quad \ldots \quad , I_{r-1} = [2(r-1)k+1, 2rk].$$

Given any $I \subseteq \mathbb{N} \cup \{0\}$, we shall be using the function $\alpha_I(n) := \sum_{\nu \in I} \epsilon_\nu(n)$.

It follows from (19.5) that

$$\alpha(f(m)) = 1 + \alpha(g(m)) = 1 + \sum_{j=0}^{r-1} \alpha_{I_j}(g(m)).$$

With M fixed, consider the expression

$$K_j := \sum_{m=1}^{M} \alpha_{I_j}(g(m)) \qquad (j = 0, 1, \ldots, r-1).$$

Observing that $\alpha_{I_0}(g(m)) = \alpha_{I_0}(m^r)$ and choosing $A = 2/3$ in Lemma 19.1, we get that

$$K_0 = kM + O(k^{1/3}M).$$

Similarly, we obtain that

$$K_1 = \sum_{m=1}^{M} \alpha_{I_1}\left(m^r + \binom{r}{1}(2M)^2 m^{r-1}\right) = kM + O(k^{1/3}M) \tag{19.6}$$

and more generally that

$$K_j = KM + O(k^{1/3}M) \qquad (j = 2, \dots, r-2). \tag{19.7}$$

We also get that

$$\alpha_{I_{r-1}}(g(m)) = \alpha_{I_{r-1}}\left(\binom{r}{r-1} 2^{(k+1)2(r-1)} m\right) = \alpha_{[0,k]}(m), \tag{19.8}$$

implying that

$$K_{r-1} = \frac{k}{2}M + O(k^{1/3}M). \tag{19.9}$$

Therefore, gathering (19.6), (19.7), (19.8) and (19.9), we obtain that

$$A_M = M + (r-1)kM + \frac{k}{2}M + O(k^{1/3}M) = \left(r - \frac{1}{2}\right) kM + O(k^{1/3}M).$$

Since $\lambda(f(m)) = 2(k+1)r + 1$ for $m = 1, \dots, M$, it follows that

$$L_M = \sum_{m=1}^{M} \lambda(f(m)) = (2(k+1)r + 1)M.$$

Combining these last two relations, we find that

$$\limsup_{M\to\infty} \frac{A_M}{L_M} = \frac{1}{2} - \frac{1}{2r}. \tag{19.10}$$

However, if β were to be a binary sharp normal number, we would need to have

$$\limsup_{M\to\infty} \frac{A_M}{L_M} = \frac{1}{2},$$

which is clearly in contradiction with (19.10). We may therefore conclude that β is not a binary sharp binary normal number.

Proof of Theorem 19.3

Given large numbers x and $y = y(x)$, we set

$$\begin{aligned} \eta_x &:= \overline{p(2)}\,\overline{p(3)}\,\overline{p(4)} \dots \overline{p(\lfloor x \rfloor)}, \\ \mu = \mu_{x,y} &:= \overline{p(\lfloor x \rfloor + 1)}\,\overline{p(\lfloor x \rfloor + 2)} \dots \overline{p(\lfloor x \rfloor + \lfloor y \rfloor)}. \end{aligned}$$

In [22], we proved that there exists an absolute constant $c > 0$ such that

$$\lambda(\eta_x) = (1 + o(1))cx \log\log x \qquad (x \to \infty). \tag{19.11}$$

Pick an arbitrary positive number $\delta < 1$, let $y = y(x) = x^\delta$ and consider the interval $J_x = [x, x+y]$. Using standard sieve methods, given a fixed small number $\varepsilon > 0$, one can prove that, for any prime $Q \le x^\varepsilon$, for some absolute constants $C_1 > 0$ and $C_2 > 0$,

$$\sum_{\substack{n \in J_x \\ p(n) = Q}} 1 \le C_1 \frac{y}{Q} \prod_{\pi < Q} \left(1 - \frac{1}{\pi}\right) \le C_2 \frac{y}{Q \log Q} \tag{19.12}$$

and that, for some absolute constant $C_3 > 0$,

$$\#\{n \in J_x : p(n) > x^\varepsilon\} \le C_3 \frac{y}{\log x}. \tag{19.13}$$

In light of (19.11), it is easily seen that, for some absolute constant $c_1 > 0$,

$$\lambda(\mu_{x,y}) = (1 + o(1)) c_1 y \log\log x \qquad (x \to \infty). \tag{19.14}$$

Let $\mathcal{A}_q := \{0, 1, \dots, q-1\}$. Moreover, let K be an arbitrary positive integer and let Υ_K be the set of the q-ary words of length K. Here, by a q-ary word of length K, we mean a block of K base q digits. Choose an arbitrary $\beta \in \Upsilon_K$. Given a word ξ whose digits belong to $\mathcal{A}_q$, let $\sigma(\xi, \beta)$ be the number of times that β appears as a subword of the word ξ. It is clear that

$$\sigma(\mu, \beta) = \sum_{n=\lfloor x \rfloor + 1}^{\lfloor x \rfloor + \lfloor y \rfloor} \sigma(\overline{p(n)}, \beta) + O(y\, K)$$

and therefore that, if $\beta_1, \beta_2 \in \Upsilon_K$ with $\beta_1 \ne \beta_2$, then

$$|\sigma(\mu, \beta_1) - \sigma(\mu, \beta_2)| \le \sum_{n=\lfloor x \rfloor + 1}^{\lfloor x \rfloor + \lfloor y \rfloor} \left| \sigma(\overline{p(n)}, \beta_1) - \sigma(\overline{p(n)}, \beta_2) \right| + O(y\, K). \tag{19.15}$$

Clearly, the theorem will be proved if we can show that

$$\max_{\substack{\beta_1, \beta_2 \in \Upsilon_K \\ \beta_1 \ne \beta_2}} \frac{|\sigma(\mu, \beta_1) - \sigma(\mu, \beta_2)|}{\lambda(\mu)} \to 0 \quad (x \to \infty). \tag{19.16}$$

Indeed, if (19.16) holds, then, given any $\beta \in \Upsilon_K$,

$$\max_{\beta \in \Upsilon_K} \frac{1}{\lambda(\mu)} \left| \sigma(\mu, \beta) - \frac{\lambda(\mu)}{q^K} \right| \to 0 \quad (x \to \infty),$$

thereby implying that μ is a q-normal sequence, as requested.

Arguing as Copeland and Erdős did in their paper [10], we have that, given a fixed $\varepsilon_1 > 0$,

$$\# \left\{ Q \in \wp \cap [U, 2U] : \max_{\substack{\beta_1, \beta_2 \in \Upsilon_K \\ \beta_1 \ne \beta_2}} \frac{|\sigma(\overline{Q}, \beta_1) - \sigma(\overline{Q}, \beta_2)|}{\lambda(\overline{Q})} > \varepsilon_1 \right\} \le c_2 U^{1-\kappa}, \tag{19.17}$$

where κ and c_2 are positive constants depending on ε_1 and K.

Let us now say that Q is a *bad prime* if

$$\max_{\substack{\beta_1, \beta_2 \in \Upsilon_K \\ \beta_1 \ne \beta_2}} \frac{|\sigma(\overline{Q}, \beta_1) - \sigma(\overline{Q}, \beta_2)|}{\lambda(\overline{Q})} > \varepsilon_1.$$

Now, observe that, for each $\beta \in \Upsilon_K$,

$$\sum_{n=\lfloor x \rfloor + 1}^{\lfloor x \rfloor + \lfloor y \rfloor} \sigma(\overline{p(n)}, \beta) \quad = \quad \sum_{Q < x^\varepsilon} \sigma(\overline{Q}, \beta) \cdot \#\{n \in J_x : p(n) = Q\}$$

$$+O\left(\#\{n \in J_x : p(n) > x^\varepsilon\} \cdot \log x\right),$$

which in light of (19.13) can be written as

$$\sum_{n=\lfloor x\rfloor+1}^{\lfloor x\rfloor+\lfloor y\rfloor} \sigma(\overline{p(n)}, \beta) = \sum_{Q<x^\varepsilon} \sigma(\overline{Q}, \beta) \cdot \#\{n \in J_x : p(n) = Q\} + O(y).$$

It follows from this last estimate that

$$\begin{aligned} S &:= \sum_{n=\lfloor x\rfloor+1}^{\lfloor x\rfloor+\lfloor y\rfloor} \left|\sigma(\overline{p(n)}, \beta_1) - \sigma(\overline{p(n)}, \beta_2)\right| \\ &\le O(y) + \varepsilon_1 \sum_{Q<x^\varepsilon} \lambda(\overline{Q}) \cdot \#\{n \in J_x : p(n) = Q\} + B(x), \end{aligned} \tag{19.18}$$

where $B(x)$ stands for the contribution of the bad primes.

Now, since, in light of (19.17), the number of bad primes $Q \in [2^u, 2^{u+1}]$ is no larger than $c_2 \cdot (2^u)^{1-\kappa}$, it follows, using (19.12), that there exists a positive constant c_3 such that

$$\begin{aligned} B(x) &\le c_3 \sum_{\substack{Q<x^\varepsilon \\ Q \text{ bad primes}}} \lambda(\overline{Q}) \frac{y}{Q \log Q} \le c_3 y \sum_{\substack{Q<x^\varepsilon \\ Q \text{ bad primes}}} \frac{1}{Q} \\ &\le c_3 y \sum_{2^u \le x^\varepsilon} \frac{1}{2^u} \#\{Q \in [2^u, 2^{u+1}] : Q \text{ is a bad prime}\} \\ &\le c_3 c_2 y \sum_{2^u \le x^\varepsilon} \frac{1}{2^u} 2^{u(1-\kappa)} \\ &= c_3 c_2 y \sum_{2^u \le x^\varepsilon} \frac{1}{2^{u\kappa}} \le c_3 c_2 y \sum_{u=1}^{\infty} \frac{1}{2^{u\kappa}} < c_4 y \end{aligned} \tag{19.19}$$

for some positive constant c_4.

Substituting (19.19) in (19.18) and recalling (19.14), it follows from (19.15) that

$$\max_{\substack{\beta_1,\beta_2 \in \Upsilon_K \\ \beta_1 \ne \beta_2}} \frac{|\sigma(\mu, \beta_1) - \sigma(\mu, \beta_2)|}{\lambda(\mu)} \le \frac{O(y) + \varepsilon_1 \lambda(\mu) + O(y)}{\lambda(\mu)} \le \varepsilon_1 + o(1) \quad (x \to \infty),$$

which implies (19.16), thereby completing the proof of the theorem.

Proofs of Theorems 19.4, 19.5 and 19.6

The proofs of Theorems 19.4, 19.5 and 19.6 are similar to that of Theorem 19.3 and we will therefore omit them.

Open problems and conjectures

1. Consider the Liouville function $\lambda(n) := (-1)^{\Omega(n)}$ and define the sequence $(\epsilon_m)_{m\geq 1}$ as follows:

$$\epsilon_m = \begin{cases} 0 & \text{if} \quad \lambda(m) = -1, \\ 1 & \text{if} \quad \lambda(m) = 1 \end{cases}$$

and consider the number

$$\xi = 0.\epsilon_1\epsilon_2\epsilon_3\ldots$$

It is not known if ξ is a binary normal number.

Observe that this would be an immediate consequence of the Chowla conjecture, which can be stated as follows: *Given a positive integer k, for every choice of integers $0 < a_1 < a_2 < \cdots < a_k$, we have*

$$\lim_{x\to\infty} \frac{1}{x} \sum_{n\leq x} \lambda(n)\lambda(n+a_1)\cdots\lambda(n+a_k) = 0.$$

It is clear that if the Chowla conjecture is true, then the number ξ is a binary normal number. Observe that recently, some partial results concerning the Chowla conjecture have been obtained (see K. Matom-Aki, M. Radziwi and T. Tao, *An average form of Chowla's conjecture*, arxiv.org/pdf/1503.05121v1.pdf).

2. Given an integer $q \geq 2$, let $\mathcal{A}_q = \{0, 1, \ldots, q-1\}$ and $\Omega(n) = \sum_{p^\alpha \| n} \alpha$. Consider the following generalisation of Chowla's conjecture: *Given arbitrary positive integers $a_1 < \cdots < a_k$,*

$$\lim_{x\to\infty} \frac{1}{x} \#\{n \leq x : \Omega(n+a_j) \equiv \ell_j \pmod{q},\ j = 1, \ldots, k\} = \frac{1}{q^k}$$

for every choice of $(\ell_1, \ldots, \ell_k) \in \mathcal{A}_q^k$. Now, consider the function $R_q(m)$ defined by $R_q(m) = \ell$ where $\ell \in \mathcal{A}_q$ is the unique integer such that $m \equiv \ell \pmod{q}$ and the real number

$$\eta = 0.R_q(\Omega(1))R_q(\Omega(2))R_q(\Omega(3))\ldots$$

If the above generalisation of Chowla's conjecture is true, then the number η is a q-normal number.

3. Let $p_1 < p_2 < p_3 < \cdots$ be the sequence of all primes and consider the set $B_q = \{\ell_0, \ldots, \ell_{\phi(q)-1}\}$ of reduced residues modulo q. With the function $R_q(m)$ defined above (in **2**), consider the function

$$\overline{R_q}(m) = \begin{cases} R_q(m) & \text{if} \quad R_q(m) \in B_q, \\ \Lambda & \text{if} \quad (m,q) \neq 1 \end{cases}$$

and the corresponding real number

$$\rho = 0.\overline{R_q}(p_1)\overline{R_q}(p_2)\overline{R_q}(p_3)\ldots$$

We make the conjecture that ρ is a q-normal number, although we are not absolutely sure that it is true.

The following conjecture of Rényi is somewhat simpler: *Let* $t_1, t_2, \ldots, t_h$ *be arbitrary integers belonging to* B_q. *Then, there exist infinitely many positive integers* n *such that* $p_{n+j} \equiv t_j \pmod{q}$ *for* $j = 1, 2, \ldots, h$. Interestingly, this conjecture has been solved in the particular case $t_1 = t_2 = \cdots = t_h$ (see Shiu [58] and Remark 8.1 on Page 58).

4. Let $\mathcal{M}$ be the semi-group generated by the integers 2 and 3. Let $m_1 < m_2 < \cdots$ be the list of all the elements of $\mathcal{M}$. Is it possible to construct a real number α such that the sequence $(y_n)_{n\in\mathbb{N}}$, where $y_n = \{m_n\alpha\}$ (here $\{x\}$ stands for the fractional part of x), is uniformly distributed in the interval $[0, 1)$?

5. Is it possible to construct a real number β for which the corresponding sequence $(s_n)_{n\in\mathbb{N}}$, where $s_n = \{(\sqrt{2})^n\beta\}$, is uniformly distributed in the interval $[0, 1)$?

6. Given a fixed integer $q \geq 2$, let $1 = \ell_0 < \ell_1 < \cdots < \ell_{\phi(q)-1}$ be the list of reduced residues modulo q. Further let $p_1 < p_2 < \cdots$ be all those prime numbers which do not divide q. Denote by $\wp_q$ the set of these primes. For each $p \in \wp_q$, let $h(p) = \nu$ if $p \equiv \ell_\nu \pmod{q}$ and consider the real number α whose $\phi(q)$-ary expansion is given by $\alpha = 0.h(p_1)\,h(p_2)\,h(p_3)\ldots$ Concerning this number, we state three conjectures:

- `Conjecture 1`: α is a $\phi(q)$-ary normal number.
- `Conjecture 2` (somewhat weaker): α is a $\phi(q)$-ary normal number with weight $1/n$, in the following sense: For every positive integer k, given $e_1 \ldots e_k$, an arbitrary block of k digits in $\{0, 1, \ldots, \phi(q) - 1\}$, we have

$$\lim_{N\to\infty} \frac{1}{\log N} \sum_{\substack{n\leq N \\ h(p_{n+1})\ldots h(p_{n+k})=e_1\ldots e_k}} \frac{1}{n} = \frac{1}{\phi(q)^k}.$$

- `Conjecture 3`: The sequence $(\{\phi(q)^n\alpha\})_{n\in\mathbb{N}}$ is everywhere dense in the interval $[0, 1)$.

7. Given a fixed integer $q \geq 2$, consider the two sequences $(\varepsilon_n)_{n\in\mathbb{N}}$ and $(\delta_n)_{n\in\mathbb{N}}$ defined by $\varepsilon_n = \omega(n) \pmod{q}$ and $\delta_n = \Omega(n) \pmod{q}$, where $\omega(n) = \sum_{p|n} 1$ and $\Omega(n) = \sum_{p^a\|n} a$. Then, let $\alpha_q := 0.\varepsilon_1\varepsilon_2\ldots$ and $\beta_q := 0.\delta_1\delta_2\ldots$ Moreover, consider the sequence $(\kappa_n)_{n\in\mathbb{N}}$ defined by $\kappa_n = \Omega(p_n + 1) \pmod{q}$, where p_n stands for the n-th prime, and let $\gamma_q := 0.\kappa_1\kappa_2\ldots$ We state the following conjecture:

`Conjecture:` The numbers α_q, β_q and γ_q are all q-ary normal numbers.

Observe that the case of δ_2 is essentially Chowla's conjecture, which we already stated on Page 99.

References

[1] N.L. Bassily and I. Kátai, *Distribution of the values of q-additive functions on polynomial sequences*, Acta Math. Hungar. **68 (4)** (1995), 353-361.

[2] 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.

[3] P.T. Bateman and R.A. Horn, *A heuristic asymptotic formula concerning the distribution of prime numbers*, Math. Comp. **16** (1962), 363–367.

[4] P.T. Bateman and R.A. Horn, *Primes represented by irreducible polynomials in one variable*, Proc. Sympos. Pure Math. **8** (1965), 119–135.

[5] A. Belshaw and P. Borwein, *Champernowne's number, strong normality, and the X chromosome*, Computational and analytical mathematics, 29-–44, Springer Proc. Math. Stat., **50**, Springer, New York, 2013.

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

[7] Y. Bugeaud, *Distribution modulo one and Diophantine approximation*. Cambridge Tracts in Mathematics, 193. Cambridge University Press, Cambridge, 2012. xvi+300 pp.

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

[9] S. Chowla, *The Riemann Hypothesis and Hilbert's Tenth Problem*, Gordon and Breach, New York, 1965.

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

[11] H. Davenport and P. Erdős, *Note on normal decimals*, Canadian J. Math. **4** (1952). 58—63.

[12] 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.

[13] J.M. De Koninck and I. Kátai, *On a problem on normal numbers raised by Igor Shparlinski*, *Bulletin of the Austr. Math. Soc.* **84** (2011), 337–349.

[14] 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.

[15] J.M. De Koninck and I. Kátai, *Exponential sums involving the largest prime factor function*, Acta Arithmetica **146** (2011), 233–245.

[16] J.M. De Koninck and I. Kátai, *Normal numbers created from primes and polynomials*, Uniform Distribution Theory **7** (2012), no.2, 1–20.

[17] J.M. De Koninck and I. Kátai, *Some new methods for constructing normal numbers*, Annales des Sciences Mathématiques du Québec **36** (2012), No.2, 349–359.

[18] J.M. De Koninck and I. Kátai, *Construction of normal numbers by classified prime divisors of integers II*, Functiones et Approximatio **49** (2013), No. 1, 7–27.

[19] J.M. De Koninck and I. Kátai, *Using large prime divisors to construct normal numbers*, Ann. Univ. Sci. Budapest Sect. Comput. **39** (2013), 45–62.

[20] J.M. De Koninck and I. Kátai, *Construction of normal numbers using the distribution of the k-th largest prime factor*, Journal of the Australian Mathematical Society **88** (2013), 158–168.

[21] J.M. De Koninck and I. Kátai, *Prime-like sequences leading to the construction of normal numbers*, Funct. Approx. Comment. Math. **49** (2013), No.2 (Dec), 291—302.

[22] J.M. De Koninck and I. Kátai, *Normal numbers generated using the smallest prime factor function*, Annales mathématiques du Québec **38** (2014), No.2, 133–144.

[23] J.M. De Koninck and I. Kátai, *Constructing normal numbers using residues of selective prime factors of integers*, Annales Univ. Sci. Budapest., Sect. Comp. **42** (2014), 127–133.

[24] J.M. De Koninck and I. Kátai, *Normal numbers and the middle prime factor of an integer*, Colloquium Mathematicum **135** (2014), No.1 (Jan), 69–77.

[25] J.M. De Koninck and I. Kátai, *The number of prime factors function on shifted primes and normal numbers*, Topics in Mathematical Analysis and Applications, Series: Springer Optimization and Its Applications, Rassias, Themistocles M., Tóth, László (Eds.) Springer, Volume 94, 2014, 315-326.

[26] J.M. De Koninck and I. Kátai, *Complex roots of unity and normal numbers*, Journal of Numbers, Vol. 2014 (June), Article ID 437814, 4 pages.

[27] J.M. De Koninck and I. Kátai, *The number of large prime factors of integers and normal numbers*, Publications mathématiques de Besançon, Année 2015, 5–12.

[28] J.M. De Koninck and I. Kátai, *Multidimensional sequences uniformly distributed modulo 1 created from normal numbers*; in Contemporary Mathematics, Vol. 655, AMS, 2015, pp. 77–82.

[29] J.M. De Koninck and I. Kátai, *Prime factorization and normal numbers*, Researches in Mathematics and Mechanics **20** no. 2 (2015), 69–80.

[30] J.M. De Koninck and I. Kátai, *On a property of non Liouville numbers*, Acta Cybernetica, **22** (2015), no. 2, 335–347.

[31] J.M. De Koninck, I. Kátai and B.M. Phong, *On strong normality*, Uniform Distribution Theory **11** (2016), no. 1, 59–78.

[32] J.M. De Koninck and I. Kátai, *On properties of sharp normal numbers and of non-Liouville numbers*, Annales mathématiques du Québec **42** (2018), no. 1, 31–47.

[33] J.M. De Koninck and I. Kátai, *Distinguishing between sharp and non-sharp normal numbers*, Mathematica Pannonica **26** (2017/2018), no. 1, 3–14.

[34] J.M. De Koninck and F. Luca, *Analytic Number Theory: Exploring the Anatomy of Integers*, Graduate Studies in Mathematics, Vol. 134, American Mathematical Society, Providence, Rhode Island, 2012.

[35] J.M. De Koninck and F. Luca, *On the middle prime factor of an integer*, Journal of Integer Sequences **13** (2013), Article 13.5.5.

[36] H. Delange, *Sur la distribution des valeurs de certaines fonctions arithmétiques*, Colloque sur la Théorie des Nombres, Bruxelles, 1955, pp. 147–161. Georges Thone, Liège; Masson and Cie, Paris, 1956.

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

[38] P.D.T.A. Elliott: Probabilistic Number Theory II, Central limit theorems, Fundamental Principles of Mathematical Sciences, 240. Springer-Verlag, Berlin-New York (1980).

[39] J. Galambos, *Introductory Probability Theory*, Marcel Dekker, New York, 1984.

[40] L. Germán and I. Kátai, *Distribution of the values of q-additive functions on some multiplicative semigroups II*, Ann. Univ. Sci. Budapest. Sect. Comput. **38** (2012), 147—159.

[41] G. Halász, *Über die Mittelwerte multiplikativen zahlentheoretischer Funktionen*, Acta Math. Acad. Scient. Hungaricae **19** (1968), 365–404.

[42] G. Halász, *On the distribution of additive and the mean values of multiplicative arithmetic functions*, Studia Sci. Math. Hungar. **6** (1971), 211–233.

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

[44] G.H.Hardy and S.Ramanujan, *The normal number of prime factors of a number n*, Quart.J.Math. **48** (1917), 76-92.

[45] C. Hooley, *Applications of Sieve Methods to the Theory of Numbers*, Cambridge University Press, 1976.

[46] L.K. Hua, *Additive Theory of Prime Numbers*, AMS (Providence, Rhode Island, 1965). Translations of Mathematical Monographs, Vol. 13.

[47] K.H. Indlekofer, I. Kátai and O. Klesov, *Renewal theorems for some weighted renewal functions*, Ann. Univ. Sci. Budapest. Sect. Comput. **34** (2011), 179—194.

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

[49] I. Kátai, *A remark on a paper of K. Ramachandra* in Number theory (Ootacamund, 1984), 147–152, Lecture Notes in Math., 1122, Springer, Berlin, 1985.

[50] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, John Wiley & Sons, New York, 1974.

[51] E. Landau, *Neuer Beweis der Primzahlsatzes und Beweis der Primidealsatzes*, Math. Ann. **56** (1903), 645-670.

[52] 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.

[53] H. Montgomery, *Primes in arithmetic progressions*, Mich Math. J. **17** (1970), 33–39.

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

[55] S.S. Pillai, *Generalisation of a theorem of Mangoldt*, Proc. Indian Acad. Sci., Sect. A. **11** (1940), 13–20.

[56] K. Prachar, *Primzahlverteilung*, Springer, Berlin, 1957.

[57] A. Schinzel and W. Sierpinski, *Sur certaines hypothèses concernant les nombres premiers*, Acta Arith. **4** (1958), 185-208; Corrigendum: ibid. **5** (1959), 259.

[58] D. Shiu, *Strings of congruent primes*, J. London Math. Soc. (2) **61** (2000), no.2, 359–363.

[59] C. Spiro, *How often is the number of divisors of n a divisor of n ?* J. Number Theory **21** (1985), no. 1, 81–100.

[60] M. Tanaka, *On the number of prime factors of integers I*, Japan J. Math. **25** (1955), 1–20.

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

[62] J. Vandehey, *The normality of digits in almost constant additive functions*, Monatsh. Math. **171** (2013), no. 3-4, 481—497.

[63] I.M. Vinogradov, *The method of trigonometric sums in the theory of numbers*, translated, revised and annotated by A. Davenport and K.F.Roth, Interscience, New York, 1954.

[64] A. Weil, *On some exponential sums*, Proc. Nat. Acad. Sci.U. S. A. **34** (1948), 204–207.

[65] H. Weyl, *Ueber die Gleichverteilung von Zahlen mod. Eins*, Math. Ann. **77** (1916), no. 3, 313–352.

JMDK; fichier: nineteen-papers-on-normal-numbers.tex; le 16 octobre 2020.