The mysterious world of normal numbers

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

Édition du 22 juin 2015

Abstract

Given an integer $q \geq 2$, a *q-normal number* (or a *normal number*) is a real number whose q-ary expansion is such that any preassigned sequence of length $k \geq 1$, of base q digits from this expansion, occurs at the expected frequency, namely $1/q^k$. Even though there are no standard methods to establish if a given number is normal or not, it is known since 1909 that almost all real numbers are normal in every base q. This is one of the many reasons why the study of normal numbers has fascinated mathematicians for the past century. We present here a brief survey of some of the important results concerning normal numbers.

1 Introduction

Flip a coin. If you obtain heads, write 0; if you obtain tails, write 1. Keep flipping the coin, writing 0's and 1's depending on the outcome. After 100 times, count the number of 0's and 1's: you will most likely count approximately 50 of each. Then, count how many times you obtained two consecutive 0's: it will most likely be approximately 25 times, since the possible outcomes of two consecutive flips are 00, 01, 10 and 11, and the probability that any such particular outcome occurs is 1/4. Similarly, if you keep flipping the coin many times, the probability that a given sequence of length k occurs will be around $1/2^k$; that's what you expect will happen: it would be perfectly normal! This is why we say that the sequence of 0's and 1's obtained by flipping a coin creates a *random sequence*, that is, a binary *normal sequence*. This is why if $a_1, a_2, a_3, \ldots$ is the infinite sequence of 0's and 1's obtained by flipping a coin (for ever!), we say that the expression $0.a_1a_2a_3\ldots$ represents a *normal number*.

Humans have always been interested in creating random numbers. In fact, random number generators have applications in gambling, lotteries, computer simulation, cryptography, completely randomized design, and many other areas where producing an unpredictable result needs to be achieved. Normal numbers have their practical use in that they provide an infinite source of pseudorandom numbers. However, the real interest for the study of normal numbers lies in the fact tha they are extremely difficult to identify and that they are very mysterious in many other aspects.

2 Basic definitions

Given an integer $q \geq 2$, a *q-normal number* (or a *normal number*) is a real number whose q-ary expansion is such that any preassigned sequence of length $k \geq 1$, of base q digits from this expansion, occurs at the expected frequency, namely $1/q^k$. Clearly, rational numbers cannot be normal since only a particular sequence of digits is repeated infinitely often.

Equivalently, given a positive irrational number η whose expansion is

$$\eta = \lfloor \eta \rfloor + 0.a_1 a_2 a_3 \ldots = \lfloor \eta \rfloor + \sum_{j=1}^{\infty} \frac{a_j}{q^j}, \text{ with each } a_j \in \{0, 1, \ldots, q-1\},$$

where $\lfloor \eta \rfloor$ stands for the integer part of η, we say that η is a *q-normal number* if the sequence $\{q^m \eta\}$, $m = 1, 2, \ldots$ (here $\{y\}$ stands for the fractional part of y), is uniformly distributed in the interval $[0, 1)$.

Both definitions are equivalent, because the sequence $\{q^m \eta\}$, $m = 1, 2, \ldots$, is uniformly distributed in $[0, 1)$ if and only if for every integer $k \geq 1$ and $b_1 \ldots b_k \in \{0, 1, \ldots, q-1\}^k$, we have

$$\lim_{N \to \infty} \frac{1}{N} \#\{j \leq N : a_{j+1} \ldots a_{j+k} = b_1 \ldots b_k\} = \frac{1}{q^k}.$$

A real number is said to be *simply normal* in base q if each digit $d \in \{0, 1, \ldots, q-1\}$ occurs with frequency $1/q$. Of course, a number can be simply normal without being a normal number (such is the case of the binary number $0.1010101010101\ldots$).

A real number is said to be *absolutely normal* if it is normal in each base $q \geq 2$.

Normal numbers are mysterious for many reasons. For instance, the constant

$$\pi = 3.14159265358979323846264338327950288419716939937 51\ldots$$

has not yet been proved to be a normal number, although it is widely believed that it is. Similarly, the frequently used

$$\begin{aligned}
\text{Euler constant } e &= 2.71828182845904523536028747135266249775724709370 00\ldots \\
\sqrt{2} &= 1.41421356237309504880168872420969807856967187537 69\ldots \\
\log 2 &= 0.693147180559945309417232121458176568075500134360 26\ldots \\
\text{Apery number } \sum_{n=1}^{\infty} \frac{1}{n^3} &= 1.20205690315959428539973816151144999076498629234 05\ldots \\
\text{Golden number } \frac{1+\sqrt{5}}{2} &= 1.61803398874989484820458683436563811772030917980 58\ldots
\end{aligned}$$

have not yet been proven to be normal numbers, although numerical evidence seems to indicate that they are. What is even more disturbing is the fact that none of the above numbers has been shown to be simply normal. For instance, it is possible that venturing along the decimals of π, from some point on, one could not any longer find the digit 0. Even though no one believes that could be the case, we can't disprove it.

On the other hand, it is widely believed that every irrational algebraic number is normal. Nevertheless, no algebraic irrational number has yet been proved to be normal (in any base).

Despite our inability to prove that any member of this large family of numbers is normal, Émile Borel [6] showed in 1909 that almost all real numbers (with respect to the Lebesgue measure) are absolutely normal.

3 A story line

Here is a story line of some of the key results obtained concerning normal numbers.

- 1909: Borel [6] introduces the concept of normal number and proves that almost all real numbers are absolutely normal.

- 1917: Sierpiński [23] provides an alternative proof that almost all real numbers are normal. It is an existence theorem, that is Sierpiński does not point out to any particular normal numbers. Here is the general idea of Sierpiński's proof. For each number $\varepsilon \in (0,1]$, he first constructs a set $\Delta(\varepsilon)$ which is the union of countably many open intervals with rational endpoints, namely

$$\Delta(\varepsilon) := \bigcup_{q=2}^{\infty} \bigcup_{m=1}^{\infty} \bigcup_{n=n_{m,q}(\varepsilon)}^{\infty} \bigcup_{p=0}^{q-1} \Delta_{q,m,n,p} \ ,$$

  where $\Delta_{q,m,n,p}$ is the set of all open intervals of the form

$$\left(\frac{b_1}{q} + \frac{b_2}{q^2} + \cdots + \frac{b_n}{q^n} - \frac{1}{q^n}, \frac{b_1}{q} + \frac{b_2}{q^2} + \cdots + \frac{b_n}{q^n} + \frac{2}{q^n} \right)$$

  such that

$$\left| \frac{c_p(b_1, b_2, \ldots, b_n)}{n} - \frac{1}{q} \right| \geq \frac{1}{m},$$

  where each $b_i \in \{0, 1, \ldots, q-1\}$ and where $c_p(b_1, b_2, \ldots, b_n)$ represents the number of times that the digit p appears amongst the digits $b_1, b_2, \ldots, b_n$. The idea is that $\Delta_{q,m,n,p}$ contains all the numbers that are not normal in base q. He then proves that every positive real number < 1 which is external to $\Delta(\varepsilon)$ is absolutely normal. Finally, he shows that $\mu(\Delta(\varepsilon)) < \varepsilon$ for every $\varepsilon \in (0,1]$, that is that the Lebesgue measure of the set $\Delta(\varepsilon)$ tends to 0 with ε, thereby establishing that almost all numbers are normal.

- 1933: Champernowne [9], an undergraduate student, proves that the number

$$C_{10} = 0.12345678910111213141516171819 2021\ldots,$$

made up from the concatenation of the positive integers, is normal in base 10. Observe that, by concatenating the sequence of integers written in any base $q \geq 2$, one can show that it provides a q-normal number.

- 1946: Copeland and Erdős [10] prove that the number $0.23571113171923293137\ldots$, obtained by the concatenation of the prime numbers, is normal in base 10. Observe that the same result holds by concatenating the sequence of prime numbers written in any base $q \geq 2$. More generally, they prove that if $a_1, a_2, a_3, \ldots$ is an increasing sequence of positive integers (expressed in base q) such that, for each positive $\theta < 1$, $\#\{a_i \leq x\} > x^\theta$ provided $x \geq x_0(\theta)$, then $0.a_1a_2a_3\ldots$ is a q-normal number.

  Since $\pi(x) > \dfrac{x}{\log x}$ for all $x \geq 11$ (here $\pi(x)$ stands for the number of primes not exceeding x), then as a particular case we get that $0.235711131719\ldots$ is indeed normal in base 10.

  As another application of the general Copeland and Erdős result, we have that since each prime $p \equiv 1 \pmod 4$ can be written as $p = r^2 + s^2$ with $r, s \in \mathbb{N}$, and since $\#\{p \leq x : p \equiv 1 \pmod 4\} > c\,x/\log x$ for all $c < \frac{1}{2}$ provided x is large enough, it follows that $\#\{n_i \leq x : n_i = r^2 + s^2\} > cx/\log x$ for large x, thus implying that the number $0.n_1n_2n_3\ldots = 0.5131729\ldots$ is normal.

- 1946: Copeland and Erdős [10] also conjecture that if $f(x)$ is any non constant polynomial whose values at $x = 1, 2, 3, \ldots$ are positive integers, then $0.f(1)f(2)f(3)\ldots$ is a normal number in base 10.

- 1952: Davenport and Erdős [11] prove this conjecture.

- 1956: Cassels [8] comes up with a large family of simply normal numbers by considering the function $f : [0,1] \to \mathbb{R}$ defined by

$$f(x) = \sum_{j=1}^{\infty} \frac{x_j}{3^j},$$

where $x_1, x_2, \ldots$ denote the binary digits of x. Then, one can easily establish that for almost all $x \in [0,1]$, $f(x)$ is simply normal with respect to every base $q \geq 2$ which is not a power of 3.

- 1992: Nakai and Shiokawa [21] prove that if $f \in \mathbb{R}[X]$ is such that $f(x) > 0$ for $x > 0$, then the real number $0.\lfloor f(1)\rfloor\lfloor f(2)\rfloor\lfloor f(3)\rfloor\ldots$, where $\lfloor f(n)\rfloor$ stands for the integer part of $f(n)$ expressed in base $q \geq 2$, is normal in base q. They also show that the same result holds if

$$f(x) = \alpha_0 x^{\beta_0} + \alpha_1 x^{\beta_1} + \cdots + \alpha_d x^{\beta_d},$$

where the α_i's and β_i's are real numbers with $\beta_0 > \beta_1 > \cdots > \beta_d \geq 0$ and $f(x) > 0$ for $x > 0$.

- 1997: Nakai and Shiokawa [22] prove that if $f \in \mathbb{Z}[X]$ is any nonconstant polynomial such that $f(x) > 0$ for $x > 0$, then the number $0.f(2)f(3)f(5)f(7)\ldots f(p)\ldots$ is normal in base 10.

- 2008: Madritsch, Thuswaldner and Tichy [19] extend the results of Nakai and Shiokawa by showing that, if f is an entire function of logarithmic order, then the numbers

  $$0.\lfloor f(1)\rfloor_q\lfloor f(2)\rfloor_q\lfloor f(3)\rfloor_q\lfloor f(4)\rfloor_q\ldots \quad \text{and} \quad 0.\lfloor f(2)\rfloor_q\lfloor f(3)\rfloor_q\lfloor f(5)\rfloor_q\lfloor f(7)\rfloor_q\ldots\ldots,$$

  where $\lfloor f(n)\rfloor_q$ stands for the base q expansion of the integer part of $f(n)$, are normal.

4 Series representing normal numbers

In 1971, Stoneham considered constants represented by convergent series as possible candidates for normality.

As we mentioned in Section 1, no one has been able to show that $\log 2 = \sum_{n=1}^{\infty} \frac{1}{n2^n}$ is a normal number. Nevertheless, Stoneham [24] was able to show that the number

$$\alpha_{2,3} := \sum_{n=3^k>1}^{\infty} \frac{1}{n2^n} = \sum_{k=1}^{\infty} \frac{1}{3^k 2^{3^k}}$$

is normal in base 2. More generally, observe that $\log \frac{b}{b-1} = \sum_{n=1}^{\infty} \frac{1}{nb^n}$. In 2002, Bailey and Crandall [4] proved that, if $b, c \geq 2$ are coprime integers, then the number $\alpha_{b,c} := \sum_{n=c^k>1} \frac{1}{nb^n} = \sum_{k=1}^{\infty} \frac{1}{c^k b^{c^k}}$ is normal in base b. They even showed that if $r = 0.r_1r_2\ldots \in [0,1)$, then $\alpha_{2,3}(r) := \sum_{n=1}^{\infty} \frac{1}{3^n 2^{3^n + r_n}}$ is a normal number in base 2, thereby providing an uncountable class of normal numbers in base 2.

Is $\alpha_{2,3}$ normal in bases other than 2? Not always! In fact, in 2006, Bailey and Borwein [1] proved that $\alpha_{2,3}$ is not a 6-normal number. Their idea was based on the fact that since the expression

$$6^{3^m}\alpha_{2,3} \mod 1 \approx \frac{(3/4)^{3^m}}{3^{m+1}}$$

(Here $x = \theta \mod 1$ means that $\theta = x - \lfloor x\rfloor$) is very small for large m, this causes the number $\alpha_{2,3}$, in base 6, to have long stretches of 0's beginning at position $3^m + 1$, and as we know this is not acceptable for a normal number!

5 Equidistribution

A sequence of positive numbers $x_0, x_1, x_2, \ldots$, each smaller than 1, is said to be *equidistributed* if, for any $0 \le c < d < 1$,

$$\lim_{N\to\infty} \frac{1}{N}\#\{0 \le j < N : x_j \in [c,d)\} = d - c.$$

In 2001, Bailey and Crandall [3] considered the sequence $x_0, x_1, \ldots$ defined by $x_0 = 1$ and, for each $n \ge 1$, by

$$x_n = \left(2x_{n-1} + \frac{1}{n}\right) \mod 1.$$

They showed that if one could prove that this sequence is equidistributed in $[0,1]$, then it would imply that $\log 2$ is a binary normal number.

Similarly, consider the sequence $y_0, y_1, \ldots$ defined by $y_0 = 1$ and, for each $n \ge 1$, by

$$y_n = \left(16y_{n-1} + \frac{120n^2 - 89n + 16}{512n^4 - 1024n^3 + 712n^2 - 206n + 21}\right) \mod 1.$$

They showed that if one could prove that this sequence is equidistributed in $[0,1]$, then it would imply that π is a 16-normal number (and hence a 2-normal number as well).

These results raise a natural question: Is it easier to prove the equidistribution of the sequence $(x_n)_{n\ge1}$ or the normality of $\log 2$? What about the sequence $(y_n)_{n\ge1}$ and its corresponding number π? Nobody knows!

6 Abnormal numbers

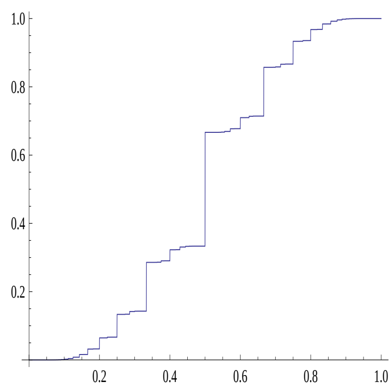
Surely, if we have so much difficulty finding normal numbers, it should be easy to find many numbers which are not normal. It turns out that, except for the rational numbers, this task is not so easy!

A number is said to be *abnormal* in base q if it is not normal in base q. For instance, the binary number

$$\sum_{n=1}^{\infty} \frac{n}{2^{n^2}} = 0.1010000110000100000000101000000000110000000000001110\ldots$$

is clearly abnormal since one can easily show that almost all of its digits in base 2 are zeros.

A less obvious example of a binary abnormal number is the amazing *Devil's staircase* number, namely the number $f(x) := \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{2^n}$ with $x \in [0,1]$. Here is the graph of $f(x)$:



This function has amazing properties. Bailey and Crandall [4] studied this function and proved that, for $x \in (0,1)$,

- f is monotone increasing,
- f is continuous at every irrational x, but discontinuous at every rational x,
- $f(x) \in \mathbb{R} \setminus \mathbb{Q}$ if and only if $x \in \mathbb{R} \setminus Q$,
- if x is irrational, then $f(x)$ is transcendental,
- the range of $f([0,1])$ is a set of measure zero,
- if $x = a/b$ with $a, b \in \mathbb{N}$ and $a/b < 1$, then $f(x) = \dfrac{1}{2^b - 1} + \displaystyle\sum_{m=1}^{\infty} \frac{1}{2^{\lfloor m/x \rfloor}}$, while if x is irrational, then $f(x) = \displaystyle\sum_{m=1}^{\infty} \frac{1}{2^{\lfloor m/x \rfloor}}$.

But then the most interesting property of $f(x)$ shown by Bailey and Crandall is that it is never 2-normal.

7 Absolutely abnormal numbers

A number is said to be *absolutely abnormal* if it is not normal in every base $q \geq 2$. In May 2000, during a survey talk by Glynn Harman, Andrew Granville asked about a specific absolutely abnormal number. In response, Carl Pomerance suggested considering the Liouville number $\ell := \sum_{n=1}^{\infty} (n!)^{-n!}$. Recall that a number β is said to be a *Liouville number* if, given any large integer m, there exists a rational p/q such that

$$0 < \left| \beta - \frac{p}{q} \right| < \frac{1}{q^m}.$$

Observe that it is known that every Liouville number is transcendental. As of now, no one has proved that ℓ is absolutely abnormal. Intrigued by Granville's question, Martin [20] considered the very fast growing sequence

$$d_2 = 2^2,\ d_3 = 3^2,\ d_4 = 4^3,\ d_5 = 5^{16},\ d_6 = 6^{30\,517\,578\,125}, \ldots$$

with the recursive rule

$$d_j = j^{d_{j-1}/(j-1)} \qquad (j \geq 3).$$

Then he proved that the number

$$\prod_{j=2}^{\infty} \left(1 - \frac{1}{d_j}\right) = 0.6562499999956991\ \underbrace{999\ldots999}_{23{,}747{,}291{,}559\ \text{9's}}\ 85284042016\ldots$$

is a Liouville number and in fact an absolutely abnormal normal.

More generally, given any sequence of positive integers $n_2, n_3, \ldots$, set $d_2 = 2^{n_2}$ and

$$d_j = j^{n_j d_{j-1}/(j-1)} \qquad (j \geq 3)$$

and consider the number

$$\alpha := \prod_{j=2}^{\infty} \left(1 - \frac{1}{d_j}\right).$$

Martin proved that α is an absolutely abnormal number, thus providing an uncountable family of absolutely abnormal numbers.

8 Using the prime factorization to construct normal numbers

As of 2011, all known normal numbers were essentially of one of the types described in Sections 3 and 4. In 2011, a totally different approach was initiated. It is based on the idea that the prime factorization of integers is locally chaotic but globally very regular. Here is how it goes.

Let $q \geq 2$ be a fixed integer and let $\wp$ stand for the set of all primes. Let $\wp_0, \wp_1, \ldots, \wp_{q-1}$ be disjoint sets of primes such that

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

where $\mathcal{R}$ is a given finite (perhaps empty) set of primes. We call $\mathcal{R}, \wp_0, \wp_1, \ldots, \wp_{q-1}$ a *disjoint classification of primes*.

A simple example of a disjoint classification of primes is obtained by letting $q = 2$ and setting $\mathcal{R} = \{2\}$, $\wp_0 = \{p \in \wp : p \equiv 1 \pmod 4\}$ and $\wp_1 = \{p \in \wp : p \equiv 3 \pmod 4\}$.

Now, for each integer $q \geq 2$, let $A_q := \{0, 1, \ldots, q-1\}$. Given an integer $t \geq 1$, we say that an expression of the form $i_1 i_2 \ldots i_t$, where each $i_j \in A_q$, is a *word* of length

t. The symbol Λ will denote the *empty word*. Now, given a disjoint classification of primes $\mathcal{R}, \wp_0, \wp_1, \ldots, \wp_{q-1}$, let the function $H : \wp \to A_q$ be defined by

$$H(p) = \begin{cases} j & \text{if } p \in \wp_j \text{ for some } j \in A_q, \\ \Lambda & \text{if } p \in \mathcal{R}. \end{cases}$$

Let A_q^* be the set of finite words over A_q and consider the function $T : \mathbb{N} \to A_q^*$ defined by

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

where we omit $H(p_i) = \Lambda$ if $p_i \in \mathcal{R}$. For convenience, we set $T(1) = \Lambda$. Finally, given a set of integers S, let $\pi(S) := \#\{p \in \wp \cap S\}$. In 2011, De Koninck and Kátai [13] proved the following result.

Theorem 1. *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 $c \geq 5$,*

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

uniformly for $2 \leq v \leq u$, $j = 0, 1, \ldots, q-1$, as $u \to \infty$. Moreover, let T be defined on $\mathbb{N}$ by

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

where

$$H(p) = \begin{cases} j & \text{if } p \in \wp_j \text{ for some } j \in A_q, \\ \Lambda & \text{if } p \in \mathcal{R}. \end{cases}$$

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

Example: Let $q = 2$, $\mathcal{R} = \{2\}$, $\wp_0 = \{p : p \equiv 1 \pmod 4\}$ and $\wp_1 = \{p : p \equiv 3 \pmod 4\}$. In particular, $\{T(1), T(2), \ldots, T(15)\} = \{\Lambda, \Lambda, 1, \Lambda, 0, 1, 1, \Lambda, 1, 0, 1, 1, 0, 1, 10\}$. Then, it follows from Theorem 1 that $\xi = 0.T(1)T(2)T(3)T(4)\ldots = 0.101110110110\ldots$ is a binary normal number.

Although we will not give here a proof of Theorem 1, let us at least mention that a key element of its proof is a 1995 result of De Koninck and Kátai [12] which we state here as Theorem A.

Theorem A. *Let $\mathcal{R}, \wp_0, \wp_1, \ldots, \wp_{q-1}$ be a disjoint classification of primes such that*

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

holds uniformly for $2 \leq v \leq u$, $i = 0, 1, \ldots, q-1$, where $c_1 \geq 5$ is a constant, $\delta_0, \delta_1, \ldots, \delta_{q-1}$ are positive constants such that $\sum_{i=0}^{q-1} \delta_i = 1$. Assume that $\lim_{x\to\infty} w_x = +\infty$, $w_x = O(\log\log\log x)$, $\sqrt{x} \leq Y \leq x$ and $1 \leq k \leq c_2 \log\log x$, where c_2 is an *arbitrary constant. Let $A \leq \log\log x$ with $P(A) \leq w_x$. Then, as $x \to \infty$, letting $\omega(n)$ stand for the number of distinct prime factors of n,*

$$\#\{n = An_1 \leq Y : p(n_1) > w_x,\ \omega(n_1) = k,\ H(n_1) = i_1 \dots i_k\}$$
$$= (1 + o(1))\delta_{i_1} \cdots \delta_{i_k} \frac{Y}{A \log Y} \frac{(\log\log x)^{k-1}}{(k-1)!} \varphi_{w_x}\left(\frac{k-1}{\log\log x}\right) F\left(\frac{k-1}{\log\log x}\right),$$

where

$$\varphi_w(z) := \prod_{p \leq w} \left(1 + \frac{z}{p}\right)^{-1} \quad \text{and} \quad F(z) := \frac{1}{\Gamma(z+1)} \prod_p \left(1 + \frac{z}{p}\right)\left(1 - \frac{1}{p}\right)^z.$$

De Koninck and Kátai [15] also proved the following.

Theorem 2. *Let $q \geq 2$ be a fixed integer. Given a positive integer*

$$n = p_1^{e_1} \cdots p_{k+1}^{e_{k+1}}$$

(here, k can be zero), let

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

Define the arithmetic function H by

$$H(n) = H(p_1^{e_1} \cdots p_{k+1}^{e_{k+1}}) = \begin{cases} c_1(n) \dots c_k(n) & \text{if } k \geq 1, \\ \Lambda & \text{if } k \leq 0. \end{cases}$$

Then the number $\xi = 0.H(1)H(2)H(3)\dots$ is a q-normal number.

9 A question raised by Shparlinski

Let $P(n)$ stand for the largest prime factor of the integer $n \geq 2$. In 2010, Igor Shparlinski asked if the number

$$0.P(2)P(3)P(4)P(5)P(6)\dots$$

is normal in base 10.

In 2011, De Koninck and Kátai [14] answered Shparlinski's question in the affirmative and actually proved more, as stated in Theorem 3 below.

But first some notation. Given a positive integer n, write its q-ary expansion as

$$n = \varepsilon_0(n) + \varepsilon_1(n)q + \cdots + \varepsilon_t(n)q^t,$$

where each $\varepsilon_i(n) \in A_q$ and $\varepsilon_t(n) \neq 0$. Then write

$$\overline{n} = \varepsilon_0(n)\varepsilon_1(n)\dots\varepsilon_t(n).$$

Theorem 3. *Let $F \in \mathbb{Z}[x]$ be a polynomial with positive leading coefficient and positive degree, and such that $F(x) > 0$ if $x > 0$. Then the number*

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

is normal.

We only give here a sketch of the proof of Theorem 3.

Let $L(n) := L_q(n) = \left\lfloor \dfrac{\log n}{\log q} \right\rfloor + 1$, that is, the number of digits of n in base q. Given a word $\theta = i_1 i_2 \ldots i_t \in A_q^t$, we write $\lambda(\theta) = t$. Also, 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 3 is the following 1996 result of Bassily and Kátai [5].

> *Let $F \in \mathbb{Z}[x]$ be a polynomial with positive leading coefficient and of positive degree r. Let $\beta \in A_q^k$. Assume that κ_u is a function of u such that $\kappa_u > 1$ for all u. Setting*
>
> $$V_\beta(u) := \#\left\{p \in \wp \cap [u, 2u] : \left|\nu_\beta(\overline{F(p)}) - \frac{L(u^r)}{q^k}\right| > \kappa_u \sqrt{L(u^r)}\right\},$$
>
> *then, there exists a positive constant c such that*
>
> $$V_\beta(u) \le \frac{cu}{(\log u)\kappa_u^2}.$$

One can easily see that from this result it follows that given $\beta_1, \beta_2 \in A_q^k$ with $\beta_1 \neq \beta_2$ and setting

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

then, for some positive constant c,

$$\Delta_{\beta_1,\beta_2}(u) \le \frac{cu}{(\log u)\kappa_u^2}. \tag{9.1}$$

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 3 will be complete if we can show that, given an arbitrary word $\beta \in A_q^k$, we have

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

Since the number of digits of each prime $p \in I_x$ is of order $\log x$, it follows by the definition of θ that

$$\lambda(\theta) \approx r\, x\, \log x,$$

which reveals the true size of $\lambda(\theta)$.

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

$$\nu_\beta(\theta) = \sum_{\substack{n \in I_x \\ x^\delta \leq P(n) \leq x^{1-\delta}}} \nu_\beta(\overline{F(P(n))}) + O(T) + O(\delta x \log x). \tag{9.2}$$

Let us now introduce the finite sequence $u_0, u_1, \ldots, u_H$ defined by $u_0 = x^\delta$ and thereafter by $u_j = 2u_{j-1}$ for each $1 \leq j \leq H$, 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 (9.2) and the fact that $T = O(x)$,

$$\nu_\beta(\theta) = \sum_{x^\delta \leq p \leq x^{1-\delta}} \nu_\beta(\overline{F(p)}) R(p) + O(\delta x \log x). \tag{9.3}$$

Let $\beta_1, \beta_2 \in A_q^k$ with $\beta_1 \neq \beta_2$. Then, using (9.3), we have

$$\begin{aligned}
|\nu_{\beta_1}(\theta) - \nu_{\beta_2}(\theta)| &\leq \sum_{x^\delta \leq p \leq x^{1-\delta}} \left| \nu_{\beta_1}(\overline{F(p)}) - \nu_{\beta_2}(\overline{F(p)}) \right| R(p) + O(\delta x \log x) \\
&= \sum_{j=0}^{H-1} \sum_{u_j \leq p < u_{j+1}} \left| \nu_{\beta_1}(\overline{F(p)}) - \nu_{\beta_2}(\overline{F(p)}) \right| R(p) + O(\delta x \log x) \\
&= \sum_{j=0}^{H-1} S_j(x) + O(\delta x \log x), \qquad (9.4)
\end{aligned}$$

say.

Set $\Psi(x,y) := \#\{n \leq x : P(n) \leq y\}$. Then, letting $z = \log x / \log y$, it is well known that

$$\Psi(x,y) = \rho(z)\, x + O\left(\frac{x}{\log y}\right) \qquad \text{uniformly for } 2 \leq y \leq x,$$

where ρ stands for the Dickman function (see for instance Theorem 9.14 in the book of De Koninck and Luca [18]).

We then have, as $x \to \infty$,

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

$$= \quad (1+o(1))\rho\left(\frac{\log x}{\log p}-1\right)\frac{x}{p},$$

from which it follows that

$$S_j(x) \le \frac{2x}{u_j}\sum_{u_j\le p<u_{j+1}}\left|\nu_{\beta_1}(\overline{F(p)})-\nu_{\beta_2}(\overline{F(p)})\right|. \tag{9.5}$$

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 one running on the bad primes, it follows from (9.5) and the Bassily-Kátai result (9.1) that

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

Summing the above inequalities for $j = 0, 1, \ldots, H-1$, 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 (9.4), for some constant $c > 0$,

$$|\nu_{\beta_1}(\theta)-\nu_{\beta_2}(\theta)| \le c\delta x\log x + o(x\log x). \tag{9.6}$$

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 $\lambda(\widetilde{\xi_N}) \le N < \lambda(\widetilde{\xi_N}\overline{F(P(m+1))})$, 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 $L = 2[\log\log x]$, that is,

$$\begin{array}{ccccc} I_{x/2^L} & & I_{x/2^2} & I_{x/2} & I_x & \\ |\!-\!| & \cdots\cdots & |\!-\!-\!-\!| & |\!-\!-\!-\!-\!-\!| & |\!-\!-\!-\!-\!-\!-\!-\!-\!-\!-\!| & 2x = m \end{array}$$

and write

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

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

Moreover, let

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

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

It is clear that

$$\left| \nu_{\beta_1}(\widetilde{\xi_N}) - \nu_{\beta_2}(\widetilde{\xi_N}) \right| \le |\nu_{\beta_1}(\mu) - \nu_{\beta_2}(\mu)| + \sum_{j=0}^{L} |\nu_{\beta_1}(\tau_j) - \nu_{\beta_2}(\tau_j)| \tag{9.7}$$

and that

$$\nu_\beta(\mu) \le \lambda(\mu) \le \frac{x}{2^L} \cdot r \log x = o(x). \tag{9.8}$$

Applying estimate (9.6) $L+1$ times (with $\theta = \widetilde{\xi_N}$) by replacing successively $2x$ by x, $x/2$, $x/2^2$, …, $x/2^L$, we obtain from (9.7) and in light of (9.8), that

$$\left| \nu_{\beta_1}(\widetilde{\xi_N}) - \nu_{\beta_2}(\widetilde{\xi_N}) \right| \le c\delta N + o(N) \qquad (N \to \infty). \tag{9.9}$$

Now, one can easily see that

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

from which it follows that

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

implying that, setting $\theta = \xi_N$ and using (9.9),

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

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 normal.

Later, in De Koninck and Kátai [16], we showed how the concatenation of the successive values of the smallest prime factor $p(n)$, as n runs through the positive integers, can also yield a normal number.

10 Using the number of prime factors of an integer to create normal numbers

In the previous section, we showed that the number $0.P(2)P(3)P(4)\ldots$ is a normal number. What if we replace the function $P(n)$ by some other arithmetic function $f(n)$? Will we still get a normal number? Not always. Take for instance the function $\omega(n)$ which counts the number of distinct prime factors of n. One can easily show that the concatenation of the successive values of $\omega(n)$, say by considering the real number $\xi := 0.\overline{\omega(2)}\,\overline{\omega(3)}\,\overline{\omega(4)}\,\overline{\omega(5)}\ldots$, 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 $\left| \dfrac{\omega(n)}{r} - 1 \right| < \dfrac{1}{r^{1/4}}$, say, with the exception of a small number of integers $n \in I$, it follows that ξ cannot be normal in basis q.

Recently, Vandehey [25] 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 $\Omega(n) := \sum_{p^\rho \| n} \rho$. 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 basis q if and only if $0 < y \le 1/2$.

In De Koninck and Kátai [17], we showed 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 basis $q \ge 2$. Moreover, we showed 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.

11 Final remarks

In 2004, Bailey, Borwein, Crandall and Pomerance [2] proved that if x is an algebraic number of degree $d > 1$, then there exists a positive constant C such that the binary expansion of x through position n has at least $Cn^{1/d}$ ones, provided n is sufficiently large. For instance, choose $x = \sqrt{2}$. It is algebraic of degree 2. Hence according to this result, the first n digits of $\sqrt{2}$ must include at least $c\sqrt{n}$ ones (for some positive constant c). Of course, if we could prove that $\sqrt{2}$ is normal, then the first n digits should include approximately $n/2$ ones. This means that we are far from the truth.

Many authors have shown a great interest for the study of normal numbers. The recent book of Bugeaud [7] contains many other results concerning this fascinating topic along with many open problems on normal numbers.

References

[1] D.H. Bailey and J.M. Borwein, *Nonnormality of the Stoneham constants*, 2011, available at http://crd.lbl.gov/dhbailey/dhbpapers/nonnormality.pdf.

[2] D.H. Bailey, J. M. Borwein, R. E. Crandall and C. Pomerance, *On the Binary Expansions of Algebraic Numbers*, Journal of Number Theory Bordeaux **16** (2004), 487–518.

[3] D.H. Bailey and R.E. Crandall, *On the random character of fundamental constant expansions*, Experimental Math., **10** (2001), (2), 175–190.

[4] D.H. Bailey and R.E. Crandall, *Random generators and normal numbers*, Exp. Math. **11** (2002), no. 4, 527–546.

[5] 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.

[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 (no. 193), Cambridge University Press, 2012.

[8] J.W.S. Cassels, *On a problem of Steinhaus about normal numbers*, Colloquium Mathematicun**7** (1959), 96–101.

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

[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, *Construction of normal numbers by classified prime divisors of integers*, Functiones et Approximatio **45** (2011), no.2, 231–253.

[14] J.M. De Koninck and I. Kátai, *On a problem on normal numbers raised by Igor Shparlinski*, Bulletin of the Australian Mathematical Society **84** (2011), 337–349.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] 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.

[20] G. Martin, *Absolutely abnormal numbers*, Amer. Math. Monthly **108** (2001), no.8, 746–754.

[21] Y. Nakai and I. Shiokawa, *Discrepancy estimates for a class of normal numbers*, Acta Arith. **62** (1992), no.3, 271–284.

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

[23] W. Sierpiński, *Démonstration élémentaire du théorème de M. Borel sur les nombres absolument normaux et détermination effective d'un tel nombre*, Bull. Soc. Math. France **45** (1917), 127–132.

[24] R. Stoneham, *On absolute* $(j,)$*-normality in the rational fractions with applications to normal numbers*, Acta Arith. **22** (1973), 277–286.

[25] J. Vandehey, *The normality of digits in almost constant additive functions*, Monatsh. Math. 171 (2013), no.3-4, 481–497.