The Mysterious World of Normal Numbers

University of Alberta

Jean-Marie De Koninck

May 3rd, 2012

Introduction

Given an integer $q \geq 2$, a *q-normal number* is an irrational 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 $1/q^k$.

Equivalently, given a positive irrational number $\eta < 1$ whose expansion is

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

Introduction-2

we say that η is a 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

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

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

Introduction-2

Interestingly :

- π, e, $\sqrt{2}$, $\log 2$ and $\zeta(3)$ have not yet been proven to be normal numbers.
- It is widely believed that every irrational algebraic number is normal.
- In fact, no algebraic irrational number has yet been proved to be normal (in any base).
- Émile Borel (1909) showed that almost all real numbers (with respect to the Lebesgue measure) are absolutely normal.

The story line

- 1909 : Borel introduces the concept of a normal number and proves that almost all real numbers are absolutely normal.
- 1917 : Sierpinski provides an example of an absolutely normal number. It is an existence theorem : for each number $\varepsilon \in (0,1]$, Sierpinski constructs a set $\Delta(\varepsilon)$ which is the union of countably many open intervals with rational endpoints, namely

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

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

The story line-2

- 1933 : Champernowne proves that the number

  $$C_{10} = 0.1234567891011121314151617181920 21\ldots,$$

  is normal in base 10.
  Mahler (1961) proved that C_{10} is transcendental. Marli and Ozcelik (2010) give a simple proof.
  (Similarly, by concatenating the sequence of integers written in any base $q \geq 2$, one can show that it provides a q-normal number.)
- 1935 : Besicovitch proves that

  $$0.14916253649\ldots,$$

The story line-3

- 1946 : Copeland and Erdős prove that the number

$$0.2357111317192329313 7\ldots$$

is normal in base 10.
(The same result holds by concatenating the sequence of prime numbers written in any base $q \geq 2$.)
More generally, they prove :

> *If $a_1, a_2, a_3, \ldots$ is an increasing sequence of positive integers 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 (where*

Applications

- Since $\pi(x) > c\dfrac{x}{\log x}$, then $0.235711131719\ldots$ is normal in base 10.

Applications

- Since $\pi(x) > c\dfrac{x}{\log x}$, then $0.235711131719\ldots$ is normal in base 10.
- Since each prime $p \equiv 1 \mod 4$ can be written as $p = r^2 + s^2$ and since

  $$\#\{p \leq x : p \equiv 1 \mod 4\} > c\frac{x}{\log x} \quad \forall c < \frac{1}{2} \text{ provided } x \text{ is large en}$$

  it follows that

  $$\#\{n_i \leq x : n_i = r^2 + s^2\} > c\frac{x}{\log x} \quad \text{for large } x,$$

  thus implying that $0.n_1 n_2 n_3 \ldots$ is a normal number.

Applications-2

- 1974 : Shiokawa proves that the Copeland–Erdős result is best possible.

Applications-2

- 1974 : Shiokawa proves that the Copeland–Erdős result is best possible.
- Copeland and Erdős (1946) 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.

Applications-2

- 1974 : Shiokawa proves that the Copeland–Erdős result is best possible.
- Copeland and Erdős (1946) 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 prove this conjecture.

Applications-3

- 1956 : Cassels proves :
  Let $f : [0,1] \to \mathbb{R}$ *be 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, 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.*

Applications-4

- 1992 : Nakai and Shiokawa 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 $\lfloor f(n) \rfloor$ is 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 $x > 0$.

Applications-5

- 1997 : Nakai and Shiokawa prove that if $f \in \mathbf{Z[X]}$ 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.

Other approaches

- 1971 : Richard Stoneham : being unable to show that

$$\log 2 = \sum_{n=1}^{\infty} \frac{1}{n2^n}$$

is a normal number, Stoneham shows that

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

is a normal number in base 2.

Other approaches-2

- 2002 : Bailey and Crandall :
  Observe that
  $$\log \frac{b}{b-1} = \sum_{n=1}^{\infty} \frac{1}{nb^n}.$$
  Given positive integers b and c, let
  $$\alpha_{b,c} = \sum_{n=c^k>1} \frac{1}{nb^n} = \sum_{k=1}^{\infty} \frac{1}{c^k b^{c^k}}.$$
  If $b, c \geq 2$ are coprime integers, then the constant $\alpha_{b,c} := \sum_{n=1}^{\infty} \frac{1}{c^n b^{c^n}}$ is a normal number in base b.
  They also show that each constant $\alpha_{b,c}$ where $b > 1$ and $c > 2$ are integers, is transcendental. Indeed, a theorem of Roth states that if $|P/Q - \alpha| < 1/Q^{2+\eta}$ admits infinitely many rational solutions P/Q for some $\eta > 0$, then α is transcendental. Hence, for fixed k, writing
  $$\alpha_{b,c} = \frac{P}{Q} + \sum_{n>k} \frac{1}{c^n b^{c^n}} \qquad \text{(where } (P,Q)=1 \text{ and } Q = c^k b^{c^k}\text{)}$$
  one can show that
  $$\left|\alpha_{b,c} - \frac{P}{Q}\right| < \frac{1}{Q^{c-\delta}}$$

Other approaches-3

- 2002 : Bailey and Crandall : If $r = 0.r_1 r_2 \ldots \in [0, 1)$, 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.

Other approaches-3

- 2002 : Bailey and Crandall : 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.

- 2006 : Bailey and J.M. Borwein : $\alpha_{2,3}$ is not a 6-normal number. The idea is that since the expression

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

  is so small for large m, it implies that the number $\alpha_{2,3}$, in base 6, has long stretches of 0's

  beginning at position $3^m + 1$.

Definition

A sequence $x = (x_0, x_1, x_2, \ldots)$ in $[0, 1)$ is said to be *equidistributed* if, for any $0 \leq c < d < 1$,

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

Definition-2

- 2001 : Bailey and Crandall : Consider the sequence $x_0, x_1, \ldots$ defined by $x_0 = 1$ and, for each $n \geq 1$, by

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

If it can be proved that this sequence is equidistributed in $[0,1]$, then $\log 2$ is a binary normal number.
Similarly, consider the sequence $x_0, x_1, \ldots$ defined by $x_0 = 1$ and, for each $n \geq 1$, by

$$x_n = \left(16x_{n-1} + \frac{120n^2 - 80n + 16}{512n^4 - 1024n^3 + 712n^2 - 206n + 21}\right) \mod 1.$$

If it can be proved that this sequence is equidistributed in $[0,1]$, then π is a 16-normal number (and hence a 2-normal number).
Consider the constant

$$E = \sum_{n=1}^{\infty} \frac{1}{2^n - 1} = \sum_{a=1}^{\infty}\sum_{b=1}^{\infty} \frac{1}{2^{ab}} = \sum_{n=1}^{\infty} \frac{d(n)}{2^n} = 1.606695\ldots.$$

Definition-2

- 2001 : Bailey and Crandall : Consider the sequence $x_0, x_1, \ldots$ defined by $x_0 = 1$ and, for each $n \geq 1$, by

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

If it can be proved that this sequence is equidistributed in $[0,1]$, then $\log 2$ is a binary normal number.
Similarly, consider the sequence $x_0, x_1, \ldots$ defined by $x_0 = 1$ and, for each $n \geq 1$, by

$$x_n = \left(16x_{n-1} + \frac{120n^2 - 80n + 16}{512n^4 - 1024n^3 + 712n^2 - 206n + 21}\right) \mod 1.$$

If it can be proved that this sequence is equidistributed in $[0,1]$, then π is a 16-normal number (and hence a 2-normal number).
Consider the constant

$$E = \sum_{n=1}^{\infty} \frac{1}{2^n - 1} = \sum_{a=1}^{\infty}\sum_{b=1}^{\infty} \frac{1}{2^{ab}} = \sum_{n=1}^{\infty} \frac{d(n)}{2^n} = 1.606695\ldots.$$

- 1948 : Erdős proved that E is irrational.

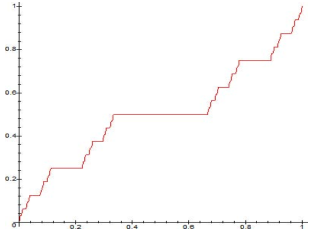
If an appropriate sequence x_n, $n = 1, 2, \ldots$, is equidistributed, then E is 2-normal.

Abnormal numbers

- A number is said to be *abnormal* in base q if it is not normal in base q. Intriguing example : $f(x) = \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{2^n}$. Then (Bailey and Crandall), for $x \in (0,1)$,

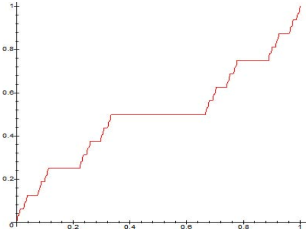
Abnormal numbers-2

Intriguing example : $f(x) = \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{2^n}$.



Abnormal numbers-2

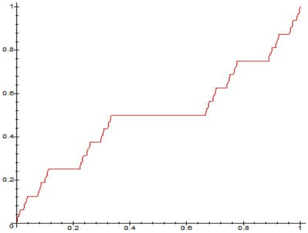
Intriguing example : $f(x) = \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{2^n}$.



1. f is monotone increasing.

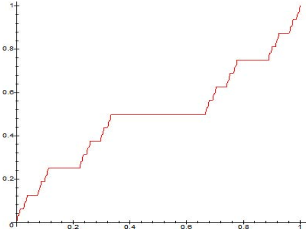
Abnormal numbers-3

Intriguing example : $f(x) = \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{2^n}$.



Abnormal numbers-3

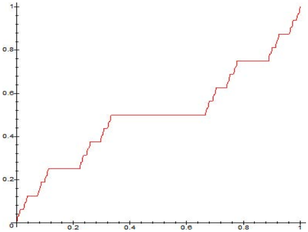
Intriguing example : $f(x) = \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{2^n}$.



1. f is continuous at every irrational x, but discontinuous at every rational x.

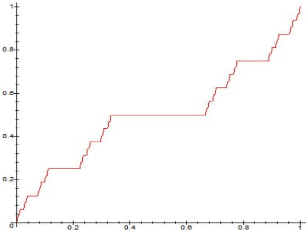
Abnormal numbers-4

Intriguing example : $f(x) = \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{2^n}$.



Abnormal numbers-4

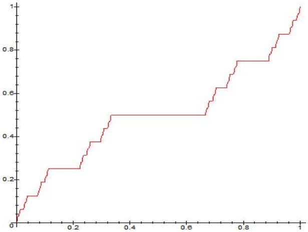
Intriguing example : $f(x) = \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{2^n}$.



1. $f(x) \in \mathbb{R} \setminus \mathbb{Q} \iff x \in \mathbb{R} \setminus Q$.

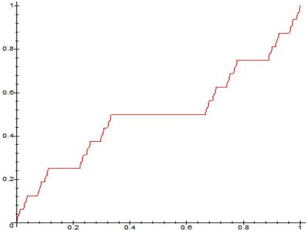
Abnormal numbers-5

Intriguing example : $f(x) = \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{2^n}.$



Abnormal numbers-5

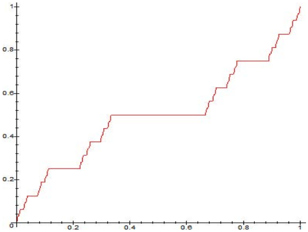
Intriguing example : $f(x) = \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{2^n}$.



1. If x is irrational, then $f(x)$ is transcendental.

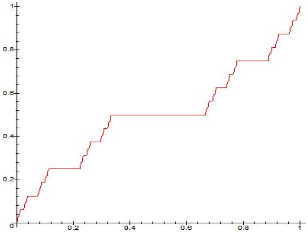
Abnormal numbers-6

Intriguing example : $f(x) = \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{2^n}$.



Abnormal numbers-6

Intriguing example : $f(x) = \sum_{n=1}^{\infty} \frac{\lfloor nx \rfloor}{2^n}$.



1. $f(x)$ is never 2-normal.

Abnormal numbers-8

- 2001 : Greg Martin considers the 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 proves that the number

$$\alpha = \prod_{j=2}^{\infty} \left(1 - \frac{1}{d_j}\right) = 0.656249999956991 \quad \underbrace{999\ldots999}_{23,747,291,559 \text{ 9's}} \quad 852840420$$

is a Liouville number and an absolutely abnormal normal.

Abnormal numbers-9

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

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

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

Abnormal numbers-10

- 2010 : Igor Shparlinski asks if the number

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

is normal in base 10. Here $P(n)$ stands for the largest prime factor of n.

A different approach for constructing normal numbers

Let $q \geq 2$ be a fixed integer.
$\wp =$ 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.

The general idea

Given a positive integer n, we associate to each prime factor appearing in its prime factorization the index i of the set of primes $\wp_i$ to which this prime factor belongs, that is :

$$n = p_1^{a_1} \cdots p_r^{a_r} \quad \longmapsto \quad \ell_1 \ldots \ell_r,$$

where each ℓ_j is such that $p_j \in \wp_{\ell_j}$

The general idea-2

We call $\mathcal{R}, \wp_0, \wp_1, \ldots, \wp_{q-1}$ a *disjoint classification of primes*.

Example of a disjoint classification of primes :

$$R = \{2\}, \quad \wp_0 = \{p : p \equiv 1 \pmod 4\}, \\ \wp_1 = \{p : p \equiv 3 \pmod 4\}$$

The general idea-3

For each integer $q \geq 2$, let $A_q := \{0, 1, \ldots, q-1\}$.
Given an integer $t \geq 1$, an expression of the form

$$i_1 i_2 \ldots i_t, \qquad \text{where each } i_j \in A_q$$

is called a *word* of length t.

The general idea-4

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.
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 $T(1) = \Lambda$.

The general idea-5

Given a set of integers S, we let

$$\pi(S) = \#\{p \in \wp \cap S\}.$$

Theorem

Let $q \geq 2$ be an integer and let $\wp = \mathcal{R} \cup \wp_0 \cup \ldots \cup \wp_{q-1}$ 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 $u \to \infty$. Further, 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,

$$\xi = 0.T(1)T(2)T(3)T(4)\ldots$$

Example

Let $q = 2$.

$$\mathcal{R} = \{2\}, \quad \wp_0 = \{p : p \equiv 1 \pmod 4\}, \quad \wp_1 = \{p : p \equiv 3 \pmod 4\}.$$

Then,

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

and

$$\xi = 0.T(1)T(2)T(3)T(4)\ldots = 0.101110110110\ldots$$

is a normal number

Main tool

Theorem A. (JMDK & IK, Acta Arith., 1995) *Let* $\mathcal{R}, \wp_0, \wp_1, \dots, \wp_{q-1}$ *be a disjoint classification of primes such that*

$$\pi([u, u+v] \cap \wp_i) = \delta_i \pi([u, u+v]) + O\left(\frac{u}{(\log u)^{c_1}}\right) \tag{1.1}$$

holds uniformly for $2 \le v \le u$, $i = 0, 1, \dots, q-1$, *where* $c_1 \ge 5$ *is a constant,* $\delta_0, \delta_1, \dots, \delta_{q-1}$ *are positive constants such that* $\sum_{i=0}^{q-1} \delta_i = 1$. *Let* $\lim_{x\to\infty} w_x = +\infty$, $w_x = O(x_3)$, $\sqrt{x} \le Y \le x$ *and* $1 \le k \le c_2 x_2$, *where* c_2 *is an arbitrary constant. Let* $A \le x_2$ *with* $P(A) \le w_x$. *Then,*

$$\#\{n = An_1 \le Y : p(n_1) > w_x,\ \omega(n_1) = k,\ H(n_1) = i_1 \dots i_k\}$$

$$= (1 + o(1))\delta_{i_1} \cdots \delta_{i_k} \frac{Y}{A \log Y} \frac{x_2^{k-1}}{(k-1)!} \varphi_{w_x}\left(\frac{k-1}{x_2}\right) F\left(\frac{k-1}{x_2}\right),$$

where

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

Theorem A-2

Theorem

Given two co-prime positive integers a and D, let $\wp_h := \{p : p \equiv h \pmod{D}\}$ *for* $\gcd(h, D) = 1$*. Let* $h_0, h_1, \ldots, h_{\varphi(D)-1}$ *be those positive integers* $< D$ *which are relatively prime with D. Further let* $\mathcal{R} = \{p : p|D\}$ *and set*

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

and

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

Let ξ *be the real number whose* $\varphi(D)$*-ary expansion is given by*

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

where $p+a$ *is the sequence of shifted primes. Then* ξ *is a* $\varphi(D)$*-normal number.*

Theorem A-3

Theorem

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

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

We shall now define the sequence ρ_n on the prime powers q^a of $E(n)$ as follows :

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

Theorem A-4

Theorem

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.

Theorem A-5

Theorem

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

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

let

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

Define the arithmetic function H by

$$H(n) = H(p_1^{e_1} \cdots p_{k+1}^{e_{k+1}}) = \begin{cases} c_1(n) \ldots c_k(n) & \text{if } \omega(n) \geq 2, \\ \Lambda & \text{if } \omega(n) \leq 1. \end{cases}$$

Then the number

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

is a q-normal number.

Main steps of the proof

Let $b_1, \ldots, b_k \in A_q$ fixed.
For each sequence of $k+1$ primes $p_1 < \cdots < p_{k+1}$, let

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

Define $F : \mathbb{N} \to \mathbb{N}$ as follows : if $n = q_1^{\alpha_1} \cdots q_\mu^{\alpha_\mu}$, then

$$F(n) = F(q_1^{\alpha_1} \cdots q_\mu^{\alpha_\mu} | b_1, \ldots, b_k) = \sum_{j=0}^{\mu-k-1} f(q_{j+1}, \ldots, q_{j+k+1}).$$

Main steps of the proof-2

$F(n)$ counts how many times the sequence of digits $b_1, \ldots, b_k$ occurs in $H(n)$.

We will 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

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

so that

$$F(n) = F_0(n) + F_1(n) + F_2(n).$$

Main steps of the proof-3

We easily show that

$$\sum_{n\le x} F_0(n) \ll x\sqrt{\log\log x},$$

$$\sum_{n\le x} F_2(n) \ll x\frac{e^{\sqrt{\log\log x}}}{\log x}.$$

We can assume that we only need to consider those $p_i | n$ such that

$$Y_x < p_1 < \cdots < p_{k+1},$$

$$p_i \| n,$$

$$\text{with} \quad \gcd\left(\frac{n}{p_1\cdots p_{k+1}}, p\right) = 1 \ \forall p \in [p_1, p_{k+1}], \ p \ne p_i$$

Main steps of the proof-4

Consider

$$\begin{aligned}\sum_{n\le x}(F_1(n)-A)^2 &= \sum_{n\le x}F_1(n)^2 - 2A\sum_{n\le x}F_1(n) + A^2\lfloor x\rfloor\\ &= S_1 - 2AS_2 + A^2\lfloor x\rfloor.\end{aligned}$$

Then, using a sieve method,

$$\begin{aligned}S_2=\sum_{n\le x}F_1(n) &= \sum_{\substack{n=\nu p_1\cdots p_{k+1}\le x\\ \left(\nu,\prod_{p_1\le p\le p_{k+1}}p\right)=1}}\ \sum_{Y_x<p_1\le Z_x} f(p_1,\ldots,p_{k+1})\\ &= (1+o(1))x\sum_{\substack{p_1<\cdots<p_{k+1}\le x\\ Y_x<p_1<Z_x}}\frac{f(p_1,\ldots,p_{k+1})}{p_1\cdots p_{k+1}}\prod_{p_1\le p\le p_{k+1}}\left(1-\frac{1}{p}\right)\\ &= (1+o(1))x\sum_{\substack{p_1<\cdots<p_{k+1}\le x\\ Y_x<p_1<Z_x}}\frac{f(p_1,\ldots,p_{k+1})}{p_1\cdots p_{k+1}}\frac{\log p_1}{\log p_{k+1}}\\ &= (1+o(1))x\left(\frac{1}{q^k}\log\log x+O(1)\right).\end{aligned}$$

Main steps of the proof-5

Also,

$$S_1 = \sum_{n \le x} F_1(n)^2 = x \left(\frac{\log\log x}{q^k} \right)^2 + O(x \log\log x).$$

Choosing $A = \frac{1}{q^k} \log\log x$, we get

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

By the Cauchy-Schwarz inequality, we obtain that

$$\sum_{n \le x} \left| F(n) - \frac{1}{q^k} \log\log x \right| \le \sum_{n \le x} \left| F_1(n) - \frac{1}{q^k} \log\log x \right|$$

Main steps of the proof-6

Theorem

Let $R \in \mathbb{Z}[x]$, 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 5 and set

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

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

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

Then ξ and η are q-normal numbers.

Main steps of the proof-7

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

$$\begin{aligned} \overline{n} &= \varepsilon_0(n)\varepsilon_1(n)\ldots\varepsilon_t(n) \\ \overline{\overline{n}} &= \varepsilon_t(n)\varepsilon_{t-1}(n)\ldots\varepsilon_0(n) \end{aligned}$$

Main steps of the proof-8

Theorem

Let $F \in \mathbb{Z}[x]$ be a polynomial with positive leading coefficient, with $F(x) > 0$ if $x > 0$, and of positive degree. Then the numbers

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

and

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

are normal.

Main steps of the proof-9

Theorem

Let F be as in Theorem 7. Then the numbers

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

and

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

are normal.

Main steps of the proof of Theorem 7

Set $L(n) = L_q(n) = \left\lfloor \frac{\log n}{\log q} \right\rfloor$.
Let $\nu_\beta(\theta)$ stand for the number of times that the subword β occurs in the word θ.
For the proof, we shall use a 1996 result of Bassily and Kátai :

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

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

then, there exists a positive constant c such that

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

Main steps of the proof of Theorem 7-2

from which it follows that

Given $\beta_1, \beta_2 \in A_q^k$ *with* $\beta_1 \neq \beta_2$, *set*

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

Then, for some positive constant c,

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

Main steps of the proof of Theorem 7-3

Let $I_x = [x, 2x]$ and set

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

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

Let δ be a small positive number.

One can easily show that the number of integers $n \in I_x$ for which either $P(n) < x^\delta$ or $P(n) > x^{1-\delta}$ is $\leq c\delta x$.

Main steps of the proof of Theorem 7-4

In light of this, we have

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

Let us choose

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

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

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

x^δ	$2x^\delta$		$x^{1-\delta}$
$\bullet$	$\bullet$	$\cdots\cdots$	$\bullet$
$\uparrow$	$\uparrow$		$\uparrow$
u_0	$u_1 = 2u_0$	$\cdots\cdots$	$u_H = 2u_{H-1} = 2^H u_0$

Main steps of the proof of Theorem 7-5

Now, setting $R(q) = \#\{n \in I_x : P(n) = q\}$, we have

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

Let $\beta_1, \beta_2 \in A_q^k$ with $\beta_1 \neq \beta_2$. Then,

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

say.

Main steps of the proof of Theorem 7-6

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

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

where α stands for the Dickman function.
We then have

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

Main steps of the proof of Theorem 7-7

But

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

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

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

while we say that it is a *bad prime* otherwise.
Splitting the sum into two sums, one running on the good primes and one running on the bad primes, it follows from the Bassily-Kátai result that

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

Main steps of the proof of Theorem 7-8

Summing the above inequalities for $j = 0, 1, \ldots, H-1$, we obtain that $\sum_{j=0}^{H} S_j(x) = o(\mathrm{li}(x))$ as $x \to \infty$ and thus that

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

Main steps of the proof of Theorem 7-9

Now let ξ_N be the prefix of length N of

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

and let

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

where $\lambda(\widetilde{\xi_N}) \le N < \lambda(\widetilde{\xi_N}\overline{F(P(m+1))})$.
It is clear that $m \sim c\dfrac{N}{\log N}$ for some constant $c > 0$, which implies that $\lambda(\overline{F(P(m+1))}) \ll r \log m$.
Let $2x = m$ and consider the intervals $I_x, I_{x/2}, I_{x/(2^2)}, \dots, I_{x/(2^L)}$, where $L = 2[\log\log x]$, and write

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

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

Main steps of the proof of Theorem 7-10