Chapter "Normal numbers and Computer Science" In "Sequences, Groups, and Number Theory",
Valérie Berthé and Michel Rigó editors.
Trends in Mathematics Series, Birkhäuser/Springer, 2018

Chapter 7
Normal Numbers and Computer Science

Verónica Becher and Olivier Carton

Abstract Émile Borel defined normality more than 100 years ago to formalize the most basic form of randomness for real numbers. A number is normal to a given integer base if its expansion in that base is such that all blocks of digits of the same length occur in it with the same limiting frequency. This chapter is an introduction to the theory of normal numbers. We present five different equivalent formulations of normality, and we prove their equivalence in full detail. Four of the definitions are combinatorial, and one is, in terms of finite automata, analogous to the characterization of Martin-Löf randomness in terms of Turing machines. All known examples of normal numbers have been obtained by constructions. We show three constructions of numbers that are normal to a given base and two constructions of numbers that are normal to all integer bases. We also prove Agafonov's theorem that establishes that a number is normal to a given base exactly when its expansion in that base is such that every subsequence selected by a finite automaton is also normal.

7.1 Introduction

Flip a coin a large number of times, and roughly half of the flips will come up heads and half will come up tails. *Normality* makes analogous assertions about the digits in the expansion of a real number. Precisely, let b be an integer greater than or equal to 2. A real number is *normal* to base b if each of the digits $0, \ldots, b-1$ occurs in its expansion with the same asymptotic frequency $1/b$, each of the blocks of two

V. Becher (✉)
Departamento de Computación, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Pabellón I, Ciudad Universitaria, C1428EGA Buenos Aires, Argentina
e-mail: vbecher@dc.uba.ar

O. Carton
Institut de Recherche en Informatique Fondamentale, Université Paris Diderot, F-75205 Paris Cedex 13, France
e-mail: Olivier.Carton@irif.fr

V. Berthé, M. Rigo (eds.), *Sequences, Groups, and Number Theory*,
Trends in Mathematics, https://doi.org/10.1007/978-3-319-69152-7_7

digits occurs with frequency $1/b^2$, each of the blocks of three digits occurs with frequency $1/b^3$, and so on, for every block length. A number is *absolutely normal* if it is normal to every base. Émile Borel [99] defined normality more than 100 years ago to formalize the most basic form of randomness for real numbers. Many of his questions are still open, such as whether any of π, e, or $\sqrt{2}$ is normal in some base, as well as his conjecture that the irrational algebraic numbers are absolutely normal [100].

In this chapter, we give an introduction to the theory of normal numbers. We start by considering five different equivalent formulations of normality, and we prove their equivalence in full detail. These proofs have not appeared all together in the literature before. Four of the definitions are combinatorial, and one is, in terms of finite automata, analogous to the characterization of Martin-Löf randomness [198] in terms of Turing machines. This characterization of normality holds for various enrichments of finite automata [57, 131], but the relation with deterministic pushdown automata remains unsolved. We also briefly mention another well-known equivalent definition of normality, in terms of uniform distribution modulo 1, that will be further considered in Chapter 8.

All known examples of normal numbers have been obtained by constructions. We first focus in three selected constructions of numbers that are normal to a given base. We then present two constructions of absolutely normal numbers, one is a slightly simplified version of the pioneer work done by Alan Turing and the other is a simplified version of the polynomial time algorithm in [53].

Finally we consider the problem of preserving normality by selection by finite automata of a subsequence of a give sequence. We give the proof of Agafonov's theorem [6] showing that a number is normal to a given base exactly when its expansion in that base is such that every subsequence selected by a finite automata is also normal.

Notation Let A be finite set of symbols that we refer as the alphabet. We write A^ω for the set of all infinite words in alphabet A, A^* for the set of all finite words, $A^{\leq k}$ for the set of all words of length up to k, and A^k for the set of words of length exactly k. The length of a finite word w is denoted by $|w|$. The positions of finite and infinite words are numbered starting at 1. To denote the symbol at position i of a word w, we write $w[i]$, and to denote the substring of w from position i to j, we write $w[i \dots j]$. The empty word is denoted by λ.

For two words w and u, the number $|w|_u$ of *occurrences* of u in w and the number $\|w\|_u$ of *aligned occurrences* of u in w are, respectively, given by

$$|w|_u = |\{i : w[i \dots i + |u| - 1] = u\}|,$$

$$\|w\|_u = |\{i : w[i \dots i + |u| - 1] = u \text{ and } i \equiv 1 \mod |u|\}|.$$

For example, $|aaaaa|_{aa} = 4$ and $\|aaaaa\|_{aa} = 2$. Notice that the definition of aligned occurrences has the condition $i \equiv 1 \mod |u|$ instead of $i \equiv 0 \mod |u|$, because the positions are numbered starting at 1. When a word u is just a symbol, $|w|_u$ and

$\|w\|_u$ coincide. Counting aligned occurrences of a word of length r over alphabet A is the same as counting occurrences of the corresponding symbol over alphabet A^r. Precisely, consider alphabet A, a length r, and an alphabet B with $|A|^r$ symbols. The set of words of length r over alphabet A and the set B are isomorphic, as witnessed by the isomorphism $\pi : A^r \to B$ induced by the lexicographic order in the respective sets. Thus, for any $w \in A^*$ such that $|w|$ is a multiple of r, $\pi(w)$ has length $|w|/r$ and $\pi(u)$ has length 1, as it is just a symbol in B. Then, for any $u \in A^r$, $\|w\|_u = |\pi(w)|_{\pi(u)}$.

7.2 Borel's Definition of Normality

A *base* is an integer greater than or equal to 2. For a real number x in the unit interval, the *expansion* of x in base b is a sequence $a_1a_2a_3\ldots$ of integers from $\{0, 1, \ldots, b-1\}$ such that

$$x = \sum_{k\geq 1} a_k b^{-k} = 0.a_1a_2a_3\ldots$$

To have a unique representation of all rational numbers, we require that expansions do not end with a tail of $b-1$. We will abuse notation, and whenever the base b is understood, we will denote the first n digits in the expansion of x with $x[1\ldots n]$.

Definition 7.2.1 (Strong Aligned Normality, Borel [99]). A real number x is simply normal to base b if, in the expansion of x in base b, each digit d occurs with limiting frequency equal to $1/b$,

$$\lim_{n\to\infty} \frac{|x[1\ldots n]|_d}{n} = \frac{1}{b}$$

A real number x is normal to base b if each of the reals $x, bx, b^2x, \ldots$ are simply normal to bases $b^1, b^2, b^3, \ldots$. A real x is absolutely normal if x is normal to every integer base greater than or equal to 2.

Theorem 7.2.2 (Borel [99]). *Almost all real numbers (with respect to Lebesgue measure) are absolutely normal.*

Are the usual mathematical constants, such as π, e, or $\sqrt{2}$, absolutely normal? Or at least simply normal to *some* base? The question remains open.

Conjecture 7.2.3 (Borel [100]). Irrational algebraic numbers are absolutely normal.

The most famous example of a normal number is due to Champernowne [141]. He proved that the number

$$0.1234567891011121314151617181920212223242526 27\ldots$$

is normal to base 10. The construction can be done in any base, obtaining a number normal to that base. It is unknown whether Champernowne numbers are normal to the bases that are multiplicatively independent to the base used in the construction. Champernowne's construction has been generalized in many interesting ways. There are also some other methods to obtain examples of numbers that are normal to a given base. In Section 7.7, we comment on the different methods, and we present three selected constructions.

All known examples of absolutely normal numbers are given by constructions. The oldest were not even computable. The first computable construction is due to A. Turing [52, 570]. In Section 7.8, we give references of known constructions, and we present two of them.

7.3 Equivalences Between Combinatorial Definitions of Normality

Borel's original definition of normality turned out to be redundant. Pillai in 1940, see [118, Theorem 4.2], proved the equivalence between Definition 7.2.1 and the following.

Definition 7.3.1 (Aligned Normality). A real number x is normal to base b if x is simply normal to bases $b^1, b^2, b^3, \ldots$.

Niven and Zuckerman in 1951, see [118, Theorem 4.5], proved yet another equivalent formulation of normality by counting occurrences of blocks but not aligned. This formulation was stated earlier by Borel himself, without proof.

Definition 7.3.2 (Non-aligned Normality). A real number x is normal to base b if for every block u,

$$\lim_{n\to\infty} \frac{|x[1\ldots n]|_u}{n} = \frac{1}{b^{|u|}}.$$

We will prove that Definitions 7.2.1, 7.3.1 and 7.3.2 are equivalent. The following lemma gives a central limit theorem that bounds the number of words in alphabet A of length k having too few or too many occurrences of some block w.

Definition 7.3.3. Let A be an alphabet of b symbols. We define the set of words of length k such that a given word w has a number of occurrences that differs from the expected value in plus or minus εk,

$$Bad(A,k,w,\varepsilon) = \left\{ v \in A^k : \left| \frac{|v|_w}{k} - b^{-|w|} \right| \geq \varepsilon \right\}.$$

Lemma 7.3.4 (Adapted from Hardy and Wright [283, proof of Theorem 148]). *Let b be an integer greater than or equal to 2, and let k be a positive integer. If $6/k \leq \varepsilon \leq 1/b$, then for every $d \in A$,*

$$|Bad(A,k,d,\varepsilon)| < 4b^k e^{-b\varepsilon^2 k/6}.$$

Proof. Observe that for any $d \in A$,

$$Bad(A,k,d,\varepsilon) = \sum_{n \leq k/b - \varepsilon k} \binom{k}{n}(b-1)^{k-n} + \sum_{n \geq k/b + \varepsilon k} \binom{k}{n}(b-1)^{k-n}$$

Fix b and k and write $N(n)$ for

$$\binom{k}{n}(b-1)^{k-n}.$$

For all $n < k/b$, we have that $N(n) < N(n+1)$ and the quotients

$$\frac{N(n)}{N(n+1)} = \frac{(n+1)(b-1)}{k-n}$$

decrease as n increases. Similarly, for all $n > k/b$, we have that $N(n) < N(n-1)$ and the quotients

$$\frac{N(n)}{N(n-1)} = \frac{k-n+1}{n(b-1)}$$

increase as n decreases. The strategy will be to "shift" each of the sums m positions.

We bound the first sum as follows. For any n we can write

$$N(n) = \frac{N(n)}{N(n+1)} \cdot \frac{N(n+1)}{N(n+2)} \cdot \ldots \cdot \frac{N(n+m-1)}{N(n+m)} \cdot N(n+m)$$

Let

$$m = \lfloor \varepsilon k/2 \rfloor \text{ and } p = \lfloor k/b - \varepsilon k \rfloor$$

For each n such that $n \leq p + m - 1$, we have that $n + m < k/b$, so

$$\begin{aligned}
\frac{N(n)}{N(n+1)} &\leq \frac{N(p+m-1)}{N(p+m)} \\
&= \frac{(p+m)(b-1)}{k-p-m+1} \\
&< \frac{(k/b - \varepsilon k/2)(b-1)}{k - k/b + \varepsilon k/2}
\end{aligned}$$

$$
\begin{aligned}
&= 1 - \frac{\varepsilon b/2}{1 - 1/b + \varepsilon/2} \\
&< 1 - \varepsilon b/2 \qquad \text{(using the hypothesis } \varepsilon \leq 1/b\text{).} \\
&< e^{-b\varepsilon/2}.
\end{aligned}
$$

Then,

$$
\begin{aligned}
N(n) &< \left(e^{-b\varepsilon/2}\right)^m \; N(n+m) \\
&\leq e^{-b\varepsilon(\varepsilon k/2-1)/2} \, N(n+m) \\
&\leq 2e^{-b\varepsilon^2 k/4} \, N(n+m), \qquad \text{(the hypothesis } \varepsilon \leq 1/b \text{ implies } e^{b\varepsilon/2} < 2\text{)}
\end{aligned}
$$

We obtain,

$$
\sum_{n \leq u} N(n) < 2e^{-b\varepsilon^2 k/2} \sum_{n \leq u} N(n+m) \leq 2\, b^k e^{-b\varepsilon^2 k/4}.
$$

We now bound the second sum, shifting it m positions. For any n we can write

$$
N(n) = \frac{N(n)}{N(n-1)} \cdot \frac{N(n-1)}{N(n-2)} \cdot \ldots \cdot \frac{N(n-m+1)}{N(n-m)} \cdot N(n-m)
$$

Let

$$
m = \lfloor \varepsilon k/2 \rfloor \text{ and } q = \lceil k/b + \varepsilon k \rceil .
$$

For each n such that $n \geq q - m + 1$, we have $n - m > k/b$, so

$$
\begin{aligned}
\frac{N(n)}{N(n-1)} &\leq \frac{N(q-m+1)}{N(q-m)} \\
&= \frac{k-q+m}{(q-m+1)(b-1)} \\
&= \frac{k - \lceil k/b + \varepsilon k \rceil + \lfloor \varepsilon k/2 \rfloor}{(\lceil k/b + \varepsilon k \rceil - \lfloor \varepsilon k/2 \rfloor + 1)(b-1)} \\
&\leq \frac{k - k/b - \varepsilon k/2}{(k/b + \varepsilon k/2 + 1)(b-1)} \\
&< \frac{1 - 1/b - \varepsilon/2}{(1/b + \varepsilon/2)(b-1)}
\end{aligned}
$$

Now

$$\begin{array}{rrl}
 & \frac{1-1/b-\varepsilon/2}{(1/b+\varepsilon/2)(b-1)} \leq & 1-b\varepsilon/3 \\
\Leftrightarrow & 1-1/b-\varepsilon/2 \leq & (1-b\varepsilon/3)(1/b+\varepsilon/2)(b-1) \\
\Leftrightarrow & (b-1)/b-\varepsilon/2 \leq & (1/b+\varepsilon/2)(b-1)-(b\varepsilon/3)(1/b+\varepsilon/2)(b-1) \\
\Leftrightarrow & (b\varepsilon/3)(1/b+\varepsilon/2)(b-1) \leq & b\varepsilon/2 \\
\Leftrightarrow & (1/b+\varepsilon/2)(b-1) \leq & 3/2.
\end{array}$$

Since $\varepsilon \leq 1/b$, we obtain the required inequality,

$$(1/b+\varepsilon/2)(b-1) \leq (1/b+1/(2b))(b-1) = 3/(2b)(b-1) \leq 3/2$$

We conclude,

$$\frac{N(n)}{N(n-1)} \leq 1-b\varepsilon/3 \leq e^{-b\varepsilon/3}.$$

Then,

$$\begin{aligned}
N(n) &< \left(e^{-b\varepsilon/3}\right)^m N(n-m) \\
&\leq e^{-b\varepsilon\lfloor \varepsilon k/2\rfloor/3} N(n-m) \\
&\leq e^{-b\varepsilon(\varepsilon k/2-1)/3} N(n-m) \\
&\leq 2\, e^{-b\varepsilon^2 k/6} N(n-m), \qquad \text{(the hypothesis } \varepsilon \leq 1/b \text{ implies } e^{b\varepsilon/3} < 2).
\end{aligned}$$

Thus,

$$\sum_{n\geq q} N(n) < 2\, b^k e^{-b\varepsilon^2 k/6}.$$

This completes the proof.

The next lemma bounds the number of words of k symbols in alphabet A that contain too many or too few occurrences of some block of length ℓ, with respect to a toleration specified by ε.

Lemma 7.3.5. *Let A be an alphabet of b symbols. Let k, ℓ be positive integers and ε a real such that $6/\lfloor k/\ell\rfloor \leq \varepsilon \leq 1/b^\ell$. Then,*

$$\left|\bigcup_{w\in A^\ell} Bad(A,k,w,\varepsilon)\right| < 2\ell\, b^{k+2\ell}\, e^{-b^\ell \varepsilon^2 k/(6\ell)}.$$

Proof. Split the set $\{1, 2, \ldots, k\}$ into the congruence classes modulo ℓ. Each of these classes contains either $\lfloor k/\ell \rfloor$ or $\lceil k/\ell \rceil$ elements. Let M_0 denote the class of all indices which leave remainder zero when being reduced modulo ℓ. Let $n_0 = |M_0|$.

For each x in A^k, consider the word in $(A^\ell)^{n_0}$

$$x[i_1 \ldots (i_1 + \ell - 1)]x[i_2 \ldots (i_2 + \ell - 1)] \ldots x[i_{n_0} \ldots (i_{n_0} + \ell - 1)]$$

for $i_1, \ldots i_{n_0} \in M_0$. By Lemma 7.3.4, we have

$$\left|Bad(A^\ell, n_0, w, \varepsilon)\right| < 4\,(b^\ell)^{n_0} e^{-b^\ell \varepsilon^2 n_0/6}.$$

Clearly, similar estimates hold for the indices in the other residue classes. Let $n_1, \ldots, n_{\ell-1}$ denote the cardinalities of these other residue classes. By assumption $n_0 + \cdots + n_{\ell-1} = k$. Then,

$$\begin{aligned}
|Bad(A, k, w, \varepsilon)| &\leq \sum_{j=0}^{\ell-1} \left|Bad(A^\ell, n_j, w, \varepsilon)\right| \\
&\leq \sum_{j=0}^{\ell-1} 4(b^\ell)^{n_j} e^{-b^\ell \varepsilon^2 n_j/6} \\
&\leq \sum_{j=0}^{\ell-1} 4(b^\ell)^{k/\ell+1} e^{-b^\ell \varepsilon^2 k/(6\ell)} \\
&= 4\,\ell\, b^{k+\ell}\, e^{-b^\ell \varepsilon^2 k/(6\ell)}.
\end{aligned}$$

The last inequality holds because

$$(b^\ell)^{\lceil k/\ell \rceil} e^{-b^\ell \varepsilon^2 \lceil k/\ell \rceil/6} < (b^\ell)^{k/\ell+1} e^{-b^\ell \varepsilon^2 k/(6\ell)}$$

and $\varepsilon \leq 1/b^\ell$ ensures

$$(b^\ell)^{\lfloor k/\ell \rfloor} e^{-b^\ell \varepsilon^2 \lfloor k/\ell \rfloor/6} \leq b^k e^{-b^\ell \varepsilon^2 k/(6\ell)} e^{1/(6b^\ell)} \leq b^k e^{-b^\ell \varepsilon^2 k/(6\ell)} b^\ell.$$

Now, summing up over all $w \in A^\ell$, we obtain

$$\left|\bigcup_{w \in A^\ell} Bad(A, k, w, \varepsilon)\right| < 2\ell\, b^{k+2\ell} e^{-b^\ell \varepsilon^2 k/(6\ell)}\ .$$

Instead of the factor 4, we can put the factor 2 because if a word $w \in A^\ell$ occurs fewer times than expected in a given word $x \in A^k$, then there is another word $v \in A^\ell$ that occurs in x more times than expected.

Lemma 7.3.6. *Let* $(x_{1,n})_{n\geq 0}, (x_{2,n})_{n\geq 0}, \ldots, (x_{k,n})_{n\geq 0}$ *be sequences of real numbers such that* $\sum_{i=1}^{k} x_{i,n} = 1$*, and let* $c_1, c_2, \ldots, c_k$ *be real numbers such that* $\sum_{i=1}^{k} c_i = 1$*. Then,*

1. *If for each i,* $\liminf_{n\to\infty} x_{i,n} \geq c_i$ *then for each i,* $\lim_{n\to\infty} x_{i,n} = c_i$*.*
2. *If for each i,* $\limsup_{n\to\infty} x_{i,n} \leq c_i$ *then for each i,* $\lim_{n\to\infty} x_{i,n} = c_i$*.*

Proof. For any i in $\{1, \ldots, k\}$,

$$\begin{aligned}
\limsup_{n\to\infty} x_{i,n} &= \limsup_{n\to\infty} (1 - \sum_{j\neq i} x_{j,n}) \\
&= 1 + \limsup_{n\to\infty} (-\sum_{j\neq i} x_{j,n}) \\
&= 1 - \liminf_{n\to\infty} (\sum_{j\neq i} x_{j,n}) \\
&\leq 1 - \sum_{j\neq i} \liminf_{n\to\infty} x_{j,n} \\
&\leq 1 - \sum_{j\neq i} c_j \\
&= c_i.
\end{aligned}$$

Since

$$\liminf \leq \limsup \text{ and } \limsup_{n\to\infty} x_{i,n} \leq c_i \leq \liminf_{n\to\infty} x_{i,n},$$

necessarily,

$$\liminf_{n\to\infty} x_{i,n} = \limsup_{n\to\infty} = c_i \text{ and } \lim_{n\to\infty} x_{i,n} = c_i.$$

Theorem 7.3.7. *Definitions 7.2.1, 7.3.1 and 7.3.2 are equivalent.*

Proof. Let x be a real number. We use the fact that for every block $w \in A^*$,

$$\lim_{n\to\infty} \frac{|x[1\ldots n]|_w}{n} = b^{-|w|}$$

if and only if there is a positive integer r such that

$$\lim_{n\to\infty} \frac{|x[1\ldots nr]|_w}{nr} = b^{-|w|}.$$

A similar fact is true for the limit of $\|x[1\dots n\ell]\|_w/n$.

1. We show that *strong aligned normality* implies *non-aligned normality*.
   Observe that for any $w \in A^\ell$,

$$|x[1\dots n]|_w = \sum_{i=0}^{\ell-1} \|(b^i x)[1\dots n-i]\|_w$$

Then,

$$\lim_{n\to\infty} \frac{|x[1\dots n]|_w}{n} = \sum_{i=0}^{\ell-1} \lim_{n\to\infty} \frac{\|(b^i x)[1\dots n-i]\|_w}{n} = \sum_{i=0}^{\ell-1} b^{-\ell}/\ell = b^{-\ell}.$$

2. We prove that *non-aligned normality* implies *aligned normality*. Define

$$\|v\|_{w,r} = |\{i : v[i..i+|w|-1] = w \text{ and } i = r \bmod |w|\}|.$$

$$\|v\|_{w,*} = \max_{1\le r\le |w|} \|v\|_{w,r}$$

$$V(w,k,\varepsilon) = \{v \in A^{k|w|-1} : \|v\|_{w,*} > (k-1)(b^{-|w|}+\varepsilon)\}$$

Given $w \in A^*$, let d be corresponding digit in $A^{|w|}$, and observe that for each $v \in V(w,k,\varepsilon)$, there is $\tilde{v} \in Bad(A^{|w|}, k-1, d, \varepsilon)$ and there are words $s,t \in A^*$ such that $|s|+|t| = |w|-1$ and $v = s\tilde{v}t$. Thus,

$$|V(w,k,\varepsilon)| \le |w| b^{|w|-1} |Bad(A^{|w|}, k-1, d, \varepsilon)|.$$

So by Lemma 7.3.5, for every positive real δ, there is k_0 such that for every $k > 0$,

$$|V(w,k,\varepsilon)|\, b^{-(k|w|-1)} < \delta.$$

Fix ℓ and assume $w \in A^\ell$. Then, for any $k \ge \max(2, k_0)$,

$$\begin{aligned}
\limsup_{n\to\infty} \frac{\|x[1\dots n\ell]\|_w}{n} &\le \limsup_{n\to\infty} \frac{1}{n(k-1)\ell} \sum_{t=1}^{n\ell-\ell+1} \|x[t\dots t+(k-1)\ell+\ell-2]\|_{w,2-t} \\
&\le \limsup_{n\to\infty} \frac{1}{n(k-1)\ell} \sum_{t=1}^{n\ell-\ell+1} \|x[t\dots t+(k-1)\ell+\ell-2]\|_{w,*} \\
&= \limsup_{n\to\infty} \sum_{v\in A^{k\ell-1}} \frac{|x[1\dots(n+k-1)\ell-1]|_v}{n\ell} \frac{\|v\|_{w,*}}{k-1}
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{v\in A^{k\ell-1}} \left(\limsup_{n\to\infty} \frac{|x[1\ldots(n+k-1)\ell-1]|_v}{n\ell}\right) \frac{\|v\|_{w,*}}{k-1}\\
&= \sum_{v\in A^{k\ell-1}} \left(\limsup_{n\to\infty} \frac{|x[1\ldots n\ell]|_v}{n\ell}\right) \frac{\|v\|_{w,*}}{k-1}\\
&= \sum_{v\in A^{k\ell-1}} b^{-(k\ell-1)} \frac{\|v\|_{w,*}}{k-1}\\
&= \sum_{v\in A^{k\ell-1}\setminus V(w,k,\varepsilon)} b^{-(k\ell-1)} \frac{\|v\|_{w,*}}{k-1} + \sum_{v\in V(w,k,\varepsilon)} b^{-(k\ell-1)} \frac{\|v\|_{w,*}}{k-1}\\
&\leq (b^{-\ell}+\varepsilon) \sum_{v\in A^{k\ell-1}\setminus V(w,k,\varepsilon)} b^{-(k\ell-1)} + \sum_{v\in A^{k\ell-1}\setminus V(w,k,\varepsilon)} b^{-(k\ell-1)}\\
&\leq b^{-\ell}+\varepsilon+\delta.
\end{aligned}
$$

To obtain the inequality in the second line, observe that each aligned occurrence of w in a position $j\ell+1$, where $k-1 \leq j < n$, is counted $(k-1)\ell$ times by $\|x[t\ldots t+k\ell-2]\|_{w,2-t}$ for $(j+1-k)\ell+1 \leq t \leq j\ell+1$.

Since the last inequality is true for any $\delta, \varepsilon > 0$, we conclude that

$$\limsup_{n\to\infty} \frac{\|x[1\ldots n\ell]\|_w}{n} \leq b^{-\ell}.$$

Applying Lemma 7.3.6, we conclude,

$$\lim_{n\to\infty} \frac{\|x[1\ldots n\ell]\|_w}{n} = b^{-|w|}.$$

3. We prove that *aligned normality* implies *strong aligned normality*. It is sufficient to prove that if x has aligned normality, then bx also has aligned normality. Define

$$U(k,w,i) = \{u \in A^k \ : \ u[i\ldots i+|w|-1] = w\}.$$

Fix a positive integer ℓ. For any $w \in A^\ell$ and for any positive integer r,

$$
\begin{aligned}
\liminf_{n\to\infty} \frac{\|(bx)[1\ldots nr\ell]\|_w}{nr} &\geq \liminf_{n\to\infty} \frac{1}{r} \sum_{k=0}^{r-2} \sum_{u\in U(r\ell,w,2+\ell k)} \frac{\|x[1\ldots nr\ell]\|_u}{n}\\
&= \frac{1}{r} \sum_{k=0}^{r-2} \sum_{u\in U(\ell r,w,2+\ell k)} b^{-r\ell}\\
&= \frac{r-1}{r} b^{-\ell}.
\end{aligned}
$$

For every r, the following equality holds:

$$\liminf_{n\to\infty} \frac{\|(bx)[1\dots n\ell]\|_w}{n} = \liminf_{n\to\infty} \frac{\|(bx)[1\dots nr\ell]\|_w}{nr}.$$

Then, using the inequality obtained above, we have

$$\liminf_{n\to\infty} \frac{\|(bx)[1\dots n\ell]\|_w}{n} \geq \frac{r-1}{r} b^{-\ell}.$$

Since this last inequality holds for every r, we obtain,

$$\liminf_{n\to\infty} \frac{\|(bx)[1\dots n\ell]\|_w}{n} \geq b^{-\ell}.$$

Finally, this last inequality is true for every $w \in A^\ell$; hence, by Lemma 7.3.6,

$$\lim_{n\to\infty} \frac{\|(bx)[1\dots n\ell]\|_w}{n} = b^{-\ell}.$$

7.4 Normality as a Seemingly Weaker Condition

The following result is due to Piatetski-Shapiro in 1957 [481] and was rediscovered later by Borwein and Bailey [101] who called it the hot spot lemma. In Theorem 7.4.1, we present two versions of this result, one with non-aligned occurrences and one with aligned occurrences. The theorem has been extended relaxing the constant C to a sublinear function; see [118] for the references.

Theorem 7.4.1. *Let x be a real and let b be an integer greater than or equal to 2. Let $A = \{0, \dots, b-1\}$. The following conditions are equivalent,*

1. *The real x is normal to base b.*
2. *There is a constant C such that for infinitely many lengths ℓ and for every w in A^ℓ*

$$\limsup_{n\to\infty} \frac{\|x[1\dots n|w|]\|_w}{n} < C \cdot b^{-|w|}.$$

3. *There is a constant C such that for infinitely many lengths ℓ and for every w in A^ℓ*

$$\limsup_{n\to\infty} \frac{|x[1\dots n]|_w}{n} < C \cdot b^{-|w|}.$$

Proof. The implications 1 ⇒ 2 and 1 ⇒ 3 follow from Theorem 7.3.7.

We now prove 2 ⇒ 1. Define,

$$\widetilde{Bad}(A^{|w|}, k, w, \varepsilon) = \left\{ v \in A^{k|w|} : \left| \frac{\|v\|_w}{k} - b^{-|w|} \right| > \varepsilon \right\}$$

Lemma 7.3.5 implies that the size of $\widetilde{Bad}(A^{|w|}, k, w, \varepsilon)$ shrinks exponentially as k increases. Suppose there is C such that for infinitely many lengths ℓ and for every $w \in A^{\ell}$,

$$\limsup_{n\to\infty} \frac{\|x[1\dots n\ell]\|_w}{n} < C \cdot b^{-\ell}.$$

Fix ℓ and $w \in A^{\ell}$. Fix $\varepsilon > 0$ and take k large enough.

$$\begin{aligned}
\liminf_{n\to\infty} \frac{\|x[1\dots nk\ell\|_w}{nk} &= \liminf_{n\to\infty} \sum_{v\in A^{k\ell}} \frac{\|x[1\dots nk\ell\|_v}{n} \frac{\|v\|_w}{k} \\
&\geq \liminf_{n\to\infty} \sum_{v\in A^{k\ell}\setminus\widetilde{Bad}(A^{\ell},k,w,\varepsilon)} \frac{\|x[1\dots nk\ell\|_v}{n} \frac{\|v\|_w}{k} \\
&\geq (1-\varepsilon)b^{-\ell} \liminf_{n\to\infty} \sum_{v\in A^{k\ell}\setminus\widetilde{Bad}(A^{\ell},k,w,\varepsilon)} \frac{\|x[1\dots nk\ell\|_v}{n} \\
&= (1-\varepsilon)b^{-\ell} \liminf_{n\to\infty} \left(1 - \sum_{v\in\widetilde{Bad}(A^{\ell},k,w,\varepsilon)} \frac{\|x[1\dots nk\ell\|_v}{n} \right) \\
&= (1-\varepsilon)b^{-\ell} \left(1 - \limsup_{n\to\infty} \sum_{v\in\widetilde{Bad}(A^{\ell},k,w,\varepsilon)} \frac{\|x[1\dots nk\ell]\|_v}{n} \right) \\
&\geq (1-\varepsilon)b^{-\ell} \left(1 - \sum_{v\in\widetilde{Bad}(A^{\ell},k,w,\varepsilon)} \limsup_{n\to\infty} \frac{\|x[1\dots nk\ell]\|_v}{n} \right) \\
&\geq (1-\varepsilon)b^{-\ell} \left(1 - \sum_{v\in\widetilde{Bad}(A^{\ell},k,w,\varepsilon)} C \cdot b^{-k\ell} \right) \\
&\geq (1-\varepsilon)b^{-\ell}(1 - C\varepsilon).
\end{aligned}$$

Since this is true for all $\varepsilon > 0$,

$$\liminf_{n\to\infty} \frac{\|x[1\dots nk\ell\|_w}{nk} \geq b^{-\ell}.$$

Finally, this last inequality is true for every $w \in A^\ell$; hence, by Lemma 7.3.6

$$\lim_{n\to\infty} \frac{\|x[1\dots n\ell\|_w}{n} = b^{-\ell}.$$

The proof of implication 3 ⇒ 1 is similar to 2 ⇒ 1. Consider the set $Bad(A, w, k, \varepsilon)$ from Definition 7.3.3, the bound in Lemma 7.3.5, and the following fact. Fix w of length ℓ. Then for any n and k,

$$|x[1\dots n]|_w \leq \frac{1}{k} \sum_{v\in A^k} \sum_{r=0}^{k-1} \|x[1\dots n]\|_{v,r}\, (|v|_w + \ell - 1)$$

$$|x[1\dots n]|_w \geq \frac{1}{k-\ell+1} \sum_{v\in A^k} \sum_{r=0}^{k-1} \|x[1\dots n]\|_{v,r} |v|_w$$

Then,

$$\begin{aligned}\lim_{n\to\infty} \frac{|x[1\dots n]|_w}{n} &\leq \lim_{n\to\infty} \frac{1}{k} \sum_{v\in A^k} \sum_{r=0}^{k-1} \frac{\|x[1\dots n]\|_{v,r}}{n} (|v|_w + \ell - 1)\\ &= \lim_{n\to\infty} \frac{1}{k} \sum_{v\in A^k} \sum_{r=0}^{k-1} \frac{\|x[1\dots n]\|_{v,r}}{n} |v|_w.\end{aligned}$$

And

$$\begin{aligned}\lim_{n\to\infty} \frac{|x[1\dots n]|_w}{n} &\geq \lim_{n\to\infty} \frac{1}{k-\ell+1} \sum_{v\in A^k} \sum_{r=0}^{k-1} \frac{\|x[1\dots n]\|_{v,r}}{n} |v|_w\\ &\geq \lim_{n\to\infty} \frac{1}{k} \sum_{v\in A^k} \sum_{r=0}^{k-1} \frac{\|x[1\dots n]\|_{v,r}}{n} |v|_w\end{aligned}$$

Hence,

$$\lim_{n\to\infty} \frac{1}{k} \frac{|x[1\dots n]|_w}{n} = \lim_{n\to\infty} \sum_{v\in A^k} \sum_{r=0}^{k-1} \frac{\|x[1\dots n]\|_{v,r}}{n} |v|_w = \lim_{n\to\infty} \frac{1}{k} \sum_{v\in A^k} \frac{|x[1\dots n]|_v}{n} |v|_w.$$

7.5 Normality as Incompressibility by Finite Automata

The definition of normality can be expressed as a notion of incompressibility by finite automata with output also known as transducers. We consider *nondeterministic transducers*. We focus on transducers that operate in real time, that is, they process exactly one input alphabet symbol per transition. We start with the definition of a transducer (see Section 1.5.4 for the definition of automata without output).

Definition 7.5.1. A *nondeterministic transducer* is a tuple $\mathscr{T} = \langle Q, A, B, \delta, I, F \rangle$, where

- Q is a finite set of *states*,
- A and B are the input and output alphabets, respectively,
- $\delta \subset Q \times A \times B^* \times Q$ is a finite *transition* relation,
- $I \subseteq Q$ and $F \subseteq Q$ are the sets of *initial* and *final* states, respectively.

A transition of such a transducer is a tuple $\langle p, a, v, q \rangle$ which is written $p \xrightarrow{a|v} q$. A finite (respectively infinite) *run* is a finite (respectively infinite) sequence of consecutive transitions,

$$q_0 \xrightarrow{a_1|v_1} q_1 \xrightarrow{a_2|v_2} q_2 \cdots q_{n-1} \xrightarrow{a_n|v_n} q_n$$

A finite path is written $q_0 \xrightarrow{a_1 \cdots a_n | v_1 \cdots v_n} q_n$. An infinite path is *final* if the state q_n is final for infinitely many integers n. In that case, the infinite run is written $q_0 \xrightarrow{a_1 a_2 a_3 \cdots | v_1 v_2 v_3 \cdots} \infty$. An infinite run is accepting if it is final and furthermore its first state q_0 is initial. This is the classical Büchi acceptance condition. For two infinite words $x \in A^\omega$ and $y \in B^\omega$, we write $\mathscr{T}(x, y)$ whenever there is an accepting run $q_0 \xrightarrow{x|y} \infty$ in $\mathscr{T}$.

Definition 7.5.2. A transducer T is *bounded-to-one* if the function $y \mapsto |\{x : \mathscr{T}(x, y)\}|$ is bounded.

Definition 7.5.3. An infinite word $x = a_1 a_2 a_3 \cdots$ is *compressible* by a nondeterministic transducer if it has an accepting run $q_0 \xrightarrow{a_1|v_1} q_1 \xrightarrow{a_2|v_2} q_2 \xrightarrow{a_3|v_3} q_3 \cdots$ satisfying

$$\liminf_{n \to \infty} \frac{|v_1 v_2 \cdots v_n|}{n} \frac{\log |B|}{\log |A|} < 1.$$

It follows from the results in [175, 531] that the words which are not compressible by one-to-one deterministic transducers are exactly the normal words. A direct proof of this result appears in [56]. Extensions of this characterization for nondeterminisms and extra memory appear in [57, 131].

Theorem 7.5.4. *An infinite word is normal if and only if it not compressible by a bounded-to-one nondeterministic transducer.*

We first show that a non-normal word is compressible. We show a slightly stronger result since the transducer can be chosen deterministic and one-to-one.

Lemma 7.5.5. *A non-normal infinite word is compressible by a deterministic one-to-one transducer.*

Proof. Assume $x \in A^\omega$ is not normal. Let us show that x is compressible regardless of the choice of an output alphabet B. Since x is not normal, there is some word u_0 of length k such that

$$\lim_{n\to\infty} \frac{\|x[1\ldots n]\|_{u_0}}{n/k} \neq \frac{1}{|A|^k}$$

meaning that the limit on the left side either does not exist or it does exist but it is different from $1/|A|^k$. There exists then an increasing sequence $(n_i)_{i\geq 0}$ of integers such that the limit $f_u = \lim_{i\to\infty} \|x[1\ldots n_i]\|_u/(n_i/k)$ does exists for each word u of length k and furthermore $f_{u_0} \neq 1/|A|^k$. Note that $\sum_{u\in A^k} f_u = 1$. Let m be an integer to be fixed later. For each word $w \in A^{km}$, let f_w be defined by $f_w = \prod_{i=1}^m f_{u_i}$ where w is factorized $w = u_1 \cdots u_m$ with $|u_i| = k$ for each $1 \leq i \leq m$. Since $\sum_{w\in A^{km}} f_w = 1$, a word $v_w \in B^*$ can be associated with each word $w \in A^{km}$ such that $v_w \neq v_{w'}$ for $w \neq w'$, the set $\{v_w : w \in A^{mk}\}$ is prefix-free, and for each $w \in A^{km}$,

$$|v_w| \leq \lceil -\log f_w/\log|B|\rceil.$$

We claim that the words $(v_w)_{w\in A^{km}}$ can be used to construct a deterministic transducer $\mathscr{T}_m$ which compresses x for m large enough. The state set Q_m of $\mathscr{T}_m$ is the set $A^{<km}$ of words of length less than km. Its initial state is the empty word λ, and all states are final. Its set E_m of transitions is given by

$$E_m = \{w \xrightarrow{a|\lambda} wa : |wa| < km\} \cup \{w \xrightarrow{a|v_{wa}} \lambda : |wa| = km\}.$$

Let us denote by $\mathscr{T}_m(z)$ the output of the transducer $\mathscr{T}_m$ on some finite input word z. Suppose that the word z is factorized $z = w_1 \cdots w_n w'$ where $|w_i| = km$ for each $1 \leq i \leq n$ and $|w'| < km$. Note that $n = \lfloor |z|/km \rfloor$. Note also that the transducer $\mathscr{T}_m$ always comes back to its initial state λ after reading km symbols.

$$\begin{aligned}
|\mathscr{T}_m(z)| &= \sum_{i=1}^n |v_{w_i}| \\
&\leq \sum_{i=1}^n \lceil -\log f_{w_i}/\log|B|\rceil \\
&\leq \frac{|z|}{km} + \sum_{i=1}^n -\log f_{w_i}/\log|B| \\
&\leq \frac{|z|}{km} + \sum_{w\in A^{km}} \|z\|_w \frac{-\log f_w}{\log|B|} \\
&\leq \frac{|z|}{km} + \sum_{u\in A^k} \|z\|_u \frac{-\log f_u}{\log|B|}.
\end{aligned}$$

Applying this computation to the prefix $z = x[1..n]$ of x gives

$$\begin{aligned}\liminf_{n\to\infty} \frac{|\mathscr{T}_m(x[1..n])\log|B|}{n\log|A|} &\le \lim_{i\to\infty} \frac{|\mathscr{T}_m(x[1..n_i])\log|B|}{n_i\log|A|} \\ &\le \frac{\log|B|}{km\log|A|} + \frac{1}{k\log|A|}\sum_{u\in A^k} f_u(-\log f_u).\end{aligned}$$

Since at least one number f_u is not equal to $1/|A|^k$, the sum $\sum_{u\in A^k} f_u(-\log f_u)$ is strictly less than $k\log|A|$. For m chosen large enough, we obtain that $\mathscr{T}_m$ compresses x.

The following lemma is the key lemma to prove the converse.

Lemma 7.5.6. *Let ℓ be a positive integer, and let $u_1, u_2, u_3, \ldots$ be words of length ℓ over the alphabet A such that $u_1u_2u_3\cdots$ is simply normal to word length ℓ. Let*

$$C_0 \xrightarrow{u_1|v_1} C_1 \xrightarrow{u_2|v_2} C_2 \xrightarrow{u_3|v_3} C_3\cdots$$

be a run where each C_i is a configuration of some kind of transducer. Assume there is a real $\varepsilon > 0$ and a set $U \subseteq A^\ell$ of at least $(1-\varepsilon)|A|^\ell$ words such that $u_i \in U$ implies $|v_i| \ge \ell(1-\varepsilon)$. Then,

$$\liminf_{n\to\infty} \frac{|v_1v_2\cdots v_n|}{n\ell} \ge (1-\varepsilon)^3.$$

Proof. Assume words u_i as in the hypothesis. By definition of normality to word length ℓ, let n_0 be such that for every $u \in A^\ell$ and for every $n \ge n_0$,

$$|\{i : 1 \le i \le n, u_i = u\}| \ge n|A|^{-\ell}(1-\varepsilon).$$

Then, for every $n \ge n_0$,

$$\begin{aligned}|v_1v_2\cdots v_n| &= \sum_{i=1}^{n} |v_i| \\ &\ge \sum_{1\le i\le n, u_i\in U} |v_i| \\ &\ge \sum_{1\le i\le n, u_i\in U} \ell(1-\varepsilon) \\ &\ge n|A|^{-\ell}(1-\varepsilon)\sum_{u\in U}\ell(1-\varepsilon) \\ &\ge n|A|^{-\ell}(1-\varepsilon)(1-\varepsilon)|A|^\ell\ell(1-\varepsilon) \\ &\ge (1-\varepsilon)^3 n\ell.\end{aligned}$$

We now come back to the proof that normal words are not compressible by bounded-to-one transducers.

Proof. Fix a normal infinite word $x = a_1a_2a_3\cdots$, a real $\varepsilon > 0$, a bounded-to-one nondeterministic transducer $T = \langle Q, A, B, \delta, q_0, F\rangle$, and an accepting run $q_0 \xrightarrow{a_1|v_1} q_1 \xrightarrow{a_2|v_2} q_2 \xrightarrow{a_3|v_3} q_3 \cdots$. It suffices to show that there is ℓ and U such that Lemma 7.5.6 applies to this arbitrary choice of ε, T, and accepting run. For each word $u \in A^*$, let

$$h_u = \min\{|v| : \exists i, j, 0 \le i \le j, q_i \xrightarrow{u|v} q_j\}$$

be the minimum number of symbols that the processing of u can contribute to the output in the run we fixed. Let

$$U_\ell = \{u \in A^\ell : h_u \ge (1-\varepsilon)\ell\}$$

be the set of words of length ℓ with relatively large contribution to the output. Let t be such that T is t-to-one. For each length ℓ, pair of states p, q that appear in the run, and for each word v, consider the set

$$U' = \{u \in A^\ell : p \xrightarrow{u|v} q\}.$$

Since p and q appear in the run, let $q_0 \xrightarrow{u_0|v_0} p$ be a prefix of the run and $q \xrightarrow{x_0|y_0} \infty$ be a suffix of the run. This implies $q \xrightarrow{x_0|y_0} \infty$ goes infinitely often through an accepting state. Thus, for different $u_1, u_2 \in U'$, there are accepting runs $q_0 \xrightarrow{u_0u_1x_0|v_0vy_0} \infty$ and $q_0 \xrightarrow{u_0u_2x_0|v_0vy_0} \infty$, from which it follows that $\mathscr{T}(u_0u_1x_0, v_0vy_0)$ and $\mathscr{T}(u_0u_2x_0, v_0vy_0)$. Therefore, by definition of t, $|U'| \le t$.

$$|\{u \in A^\ell : p \xrightarrow{u|v} q\}| \le t.$$

Thus,

$$|U_\ell| \ge |A|^\ell - |Q|^2 t |B|^{(1-\varepsilon)\ell+1}.$$

Fix ℓ such that $|U_\ell| > |A|^\ell(1-\varepsilon)$ and apply Lemma 7.5.6 with $U = U_\ell$ to the considered run. This completes the proof.

7.6 Normality as Uniform Distribution Modulo 1

Let $(x_j)_{j\ge1}$ be a sequence of real numbers in the unit interval. The discrepancy of the N first elements is

$$D_N((x_j)_{j\ge1}) = \sup_{0\le u<v\le1} \left| \frac{|\{j : 1 \le j \le N \text{ and } u \le x_j \le v\}|}{N} - (v-u) \right|.$$

The sequence $(x_j)_{j\geq 1}$ is uniformly distributed in the unit interval if

$$\lim_{N\to\infty} D_N((x_j)_{j\geq 1}) = 0.$$

Schmidt [530] proved that for every sequence $(x_j)_{j\geq 1}$ of reals in the unit interval, there are infinitely many Ns such that

$$D_N((x_j)_{j\geq 1}) \geq \frac{\log N}{100\,N}.$$

There are sequences that achieve this lower bound, see [199].

Normality can be expressed in terms of uniform distribution modulo 1.

Theorem 7.6.1 (Wall 1949 [578]). *A real number x is normal to base b if and only if the sequence $(b^j x)_{j\geq 0}$ is uniformly distributed modulo* 1.

The discrepancy modulo 1 of the sequence $(b^j x)_{j\geq 0}$ gives the speed of convergence to normality to base b. Gál and Gál [236] and Philipp [480] proved that for almost all real numbers x, the discrepancy modulo 1 of the sequence $(b^j x)_{j\geq 0}$ is essentially the same and it obeys the law of iterated logarithm up to a constant factor that depends on b. Fukuyama [233] obtained the precise constant factor.

For a real number x, we write $\{x\} = x - \lfloor x \rfloor$ to denote the fractional part of x.

Theorem 7.6.2 (Fukuyama 2008 [233]). *For every real $\theta > 1$, there is a constant C_θ such that for almost all real numbers x (with respect to Lebesgue measure),*

$$\limsup_{N\to\infty} \frac{D_N(\{\theta^j x\}_{j\geq 0})\sqrt{N}}{\sqrt{\log\log N}} = C_\theta.$$

For instance, in case θ is an integer greater than or equal to 2*,*

$$C_\theta = \begin{cases} \sqrt{84}/9, & \text{if } \theta = 2 \\ \sqrt{2(\theta+1)/(\theta-1)}/2, & \text{if } \theta \text{ is odd} \\ \sqrt{2(\theta+1)\theta(\theta-2)/(\theta-1)^3}/2, & \text{if } \theta \geq 4 \text{ is even.} \end{cases}$$

It remains an open problem to establish the minimal discrepancy that can be achieved by a sequence $(\{b^j x\})_{j\geq 0}$ for some x.

The formulation of normality in terms of uniform distribution modulo 1 has been used in constructions of numbers that are normal to one base and not normal to another, where analytic tools come into play by way of Weyl's criterion of equidistribution [118, 364]. We give some references in Section 7.8.

7.7 Constructions of Numbers That Are Normal to a Given Base

Copeland and Erdős [166] generalized Champernowne's construction [141]. They show that for any increasing sequence of integers which does not grow too fast, the concatenation of its terms yields the expansion of a normal number. In particular, one can take the sequence of prime numbers. There are many other generalizations, such as [180, 435].

Other examples of normal numbers are defined by arithmetic constructions, the first ones are due to Stoneham [553] and Korobov [359]. For b, c be relatively prime integers greater than 1, the real numbers

$$\alpha_{b,c} = \sum_{n=1}^{\infty} \frac{1}{c^n b^{c^n}}$$

are normal to base b. Bailey and Borwein [31] showed that $\alpha_{2,3}$ is normal to base 2 but not to base 6. Noticeably, for any given integer base b, Levin [376] gives an arithmetic construction of a real number x, subtler than the series for $\alpha_{b,c}$, such that $D_N(\{b^n x\}_{n\geq 0})$is in $O((\log N)^2/N)$. This is the lowest discrepancy obtained so far, and it is close to the lower bound of $O(\log(N)/N)$ proved by Schmidt for arbitrary sequences (see Section 7.6 above). It is an open question whether there exists a real x for which $D_N(\{b^n x\}_{n\geq 0})$ reaches Schmidt's general lower bound.

Yet there is a very different kind of construction of expansions of normal numbers, based on combinatorics on words, specifically on de Bruijn words. This is due to Ugalde in [571].

In all the cases, the constructions have the form of an algorithm or can be turned into an algorithm. Recall that a real number x is computable if there is an algorithm that produces the expansion of x in some base, one digit after the other. The algorithm computes in linear time or has linear time complexity if it produces the first n digits in the expansion of x after performing a number of operations that is linear in n. Similarly, we consider polynomial, exponential, or hyper-exponential complexity. Algorithms with exponential complexity cannot run in human time, but algorithms with sub-exponential complexity can. In this monograph we analyze the computational complexity by counting the number of mathematical operations required to output the first k digits of the expansion of the computed number in a designated base. Thus, we do not count how many elementary operations are implied by each of the mathematical operations, which means that we neglect the computational cost of performing arithmetical operations with arbitrary precision.

In this section we present three constructions of real numbers that are insured to be normal to a given base. Since we care about the normality to just one base, we will just construct infinite words in a given alphabet. We first present the simplest possible construction à la Champernowne. Then we present Ugalde's construction, and we give a much simpler proof than the one in [571]. Finally we present a subtle construction of a normal word which has a self-similarity condition: the whole infinite word is identical to its subsequence at the even positions. This result is due to Becher, Carton, and Heiber (see [50, Theorem 4.2]).

7.7.1 À la Champernowne

Theorem 7.7.1. *Let A be an alphabet. Let w_j be the concatenation of all words over A of length j, in lexicographic order. The infinite word $w = w_1w_2w_3\ldots$ is normal to alphabet A.*

Proof. Let $w = w_1w_2w_3\ldots = a_1a_2\ldots$ where each a_i is a symbol in A. Fix N and let n be such that

$$\sum_{j=1}^{n} j|A|^j \leq N < \sum_{j=1}^{n+1} j|A|^j$$

Let u be a block of symbols in alphabet A. The occurrences of u in the prefix of $w[1..N]$ are divided into two classes: those that are fully contained in a single block of length i in some w_i and those that overlap several blocks.

$$\begin{aligned}\frac{|a_1a_2\ldots a_N|_u}{N} &\leq \frac{|a_1a_2\ldots a_{x_{n+1}}|_u}{n|A|^n}\\ &\leq \frac{1}{n|A|^n}\left(\sum_{j=|u|}^{n+1}(j-|u|+1)|A|^{j-|u|} + \sum_{j=1}^{n+1}(|u|-1)|A|^j\right)\\ &\leq \frac{(n+1)|A|^{-|u|}}{n|A|^n}\sum_{j=1}^{n+1}|A|^j + \frac{|u|}{n|A|^n}\sum_{j=1}^{n+1}|A|^j\\ &\leq \frac{(n+1)}{n(|A|-1)}|A|^{-|u|} + \frac{|u||A|^2}{n(|A|-1)}.\end{aligned}$$

The first term accounts for occurrences fully contained in a block and the second of for those that overlap several blocks. It follows that

$$\limsup_{N\to\infty}\frac{|a_1a_2\ldots a_N|_u}{N} \leq \frac{2}{|A|-1}|A|^{-|u|}.$$

By Lemma 7.4.1, w is normal to alphabet A.

The infinite word w can be computed very efficiently: the first N symbols can be produced in at most $O(N)$ elementary operations. It is also possible to produce just the N-th symbol of w in $O(\log N)$ many elementary operations.

7.7.2 Infinite de Bruijn Words

See [76] for a fine presentation and history of de Bruijn words.

Definition 7.7.2 ([182, 517]). A (noncyclic) *de Bruijn word* of order n over alphabet A is a word of length $|A|^n + n - 1$ such that every word of length n occurs in it exactly once.

Every de Bruijn word of order n over A with $|A| \geq 3$ can be extended to a de Bruijn word of order $n + 1$. Every de Bruijn word of order n over A with $|A| = 2$ can *not* be extended to order $n + 1$, but it can be extended to order $n + 2$. See [55] for a complete proof of this fact. This allows us to define infinite de Bruijn words, as follows.

Definition 7.7.3. An infinite de Bruijn word $w = a_1a_2\ldots$ in an alphabet of at least three symbols is an infinite word such that, for every n, $a_1 \ldots a_{|A|^n+n-1}$ is a de Bruijn word of order n. In case the alphabet has two symbols, an infinite de Bruijn word $w = a_1a_2\ldots$ is such that, for every odd n, $a_1 \ldots a_{|A|^n+n-1}$ is a de Bruijn word of order n.

Ugalde [571] was the first to prove that infinite de Bruijn words are normal.

Theorem 7.7.4. *Infinite de Bruijn words are normal.*

Proof. In case the alphabet A has two symbols, consider instead the words in the alphabet A' of four symbols obtained by the morphism mapping blocks two symbols in A to one symbol in A', and prove normality for alphabet A'.

Suppose that the alphabet A has at least 3 symbols. Let $x = a_1a_2\ldots$ be an infinite de Bruijn word over A. Fix a word u of length ℓ and $n > |A|^\ell + \ell - 1$. Then u occurs in a de Bruijn word of order $n \geq \ell$ between $|A|^{n-\ell}$ and $|A|^{n-\ell} + n - \ell$ times. To see this, observe if u occurs at a position i, for some i such that $1 \leq i \leq |A|^n$, then position i is the beginning of an occurrence of a word of length n. There are exactly $|A|^{n-\ell}$ words of length n whose first ℓ symbols are u. In addition, there are exactly $n - \ell$ other positions in a de Bruijn word of order n at which a subword of length ℓ may start. Since x is infinite de Bruijn, by definition, for each n, $a_1 \ldots a_{|A|^n+n-1}$ is a de Bruijn word or order n. Fix a position N, and let n be such that

$$|A|^n + n - 1 \leq N < |A|^{n+1} + n.$$

Then,

$$\frac{|a_1 \ldots a_N|_u}{N} \leq \frac{|a_1 \ldots a_{|A|^{n+1}+n}|_u}{|A|^n + n - 1} \leq \frac{|A|^{n+1-\ell} + n - \ell}{|A|^n + n - 1} \leq 2\,|A|^{-\ell+1}.$$

Thus,

$$\limsup_{N\to\infty} \frac{|a_1 \dots a_N|_u}{N} < 2\,|A|^{-\ell+1}.$$

By Lemma 7.4.1, using $C = 2\,|A|$, x is normal.

There is an obvious algorithm to compute an infinite de Bruijn word which, for each $n \geq 1$, extends a Hamiltonian cycle in a de Bruijn graph of order n to an Eulerian cycle in the same graph. This is done in time exponential in n. No efficient algorithm is known to compute the N-th symbol of an infinite de Bruijn word without computing the first N symbols.

7.7.3 A Normal and Self-Similar Word

For a given finite or infinite word $x = a_1a_2a_3\dots$ where each a_i is a symbol in alphabet A, define $\text{even}(x) = a_2a_4a_6\cdots$ and $odd(x) = a_1a_3a_5\cdots$. Thus, $x = \text{even}(x)$ means that $a_n = a_{2n}$ for all n.

Theorem 7.7.5 ([50, Theorem 4.2]). *There is a normal word x such that $x = \text{even}(x)$.*

We construct a normal word $x = a_1a_2a_3\cdots$ over the alphabet $\{0, 1\}$ such that $a_{2n} = a_n$ for every n. The construction can be extended to an alphabet of size k to obtain a word $a_1a_2a_3\cdots$ such that $a_{kn} = a_n$ for each integer $n \geq 1$.

A finite word w is called *ℓ-perfect* for an integer $\ell \geq 1$, if $|w|$ is a multiple of ℓ and all words of length ℓ have the same number $|w|/(\ell 2^\ell)$ of aligned occurrences in w.

Lemma 7.7.6. *Let w be an ℓ-perfect word such that $|w|$ is a multiple of $\ell 2^{2\ell}$. Then, there exists a 2ℓ-perfect word z of length $2|w|$ such that $\text{even}(z) = w$.*

Proof. Since $|w|$ is a multiple of $\ell 2^{2\ell}$ and w is ℓ-perfect, for each word u of length ℓ, $\|w\|_u$ is a multiple of 2^ℓ. Consider a factorization of $w = w_1w_2\cdots w_r$ such that for each i, $|w_i| = \ell$. Thus, $r = |w|/\ell$. Since w is ℓ-perfect, for any word u of length ℓ, the set $\{i : w_i = u\}$ has cardinality $r/2^\ell$. Define z of length $2|w|$ as $z = z_1z_2\cdots z_r$ such that for each i, $|z_i| = 2\ell$, $even(z_i) = w_i$ and for all words u and u' of length ℓ, the set $\{i : z_i = u' \vee u\}$ has cardinality $r/2^{2\ell}$. This latter condition is achievable because, for each word u of length ℓ, the set $\{i : \text{even}(z_i) = u\}$ has cardinality $r/2^\ell$ which is a multiple of 2^ℓ, the number of possible words u'.

Corollary 7.7.7. *Let w be an ℓ-perfect word for some even integer ℓ. Then there exists an ℓ-perfect word z of length $2|w|$ such that $\text{even}(z) = w$.*

Proof. Since w is ℓ-perfect, it is also $\ell/2$-perfect. Furthermore, if u and v are words of length $\ell/2$ and ℓ, respectively, then $\|w\|_u = 2^{\ell/2+1}\|w\|_v$. Thus, the hypothesis of Lemma 7.7.6 is fulfilled with $\ell/2$.

Corollary 7.7.8. *There exist a sequence $(w_n)_{n\geq 1}$ of words and a sequence of positive integers $(\ell_n)_{n\geq 1}$ such that $|w_n| = 2^n$, $\mathrm{even}(w_{n+1}) = w_n$, w_n is ℓ_n-perfect and $(\ell_n)_{n\geq 1}$ is nondecreasing and unbounded. Furthermore, it can be assumed that $w_1 = 01$.*

Proof. We start with $w_1 = 01$, $\ell_1 = 1$, $w_2 = 1001$, and $\ell_2 = 1$. For each $n \geq 2$, if $\ell_n 2^{2\ell_n}$ divides $|w_n|$, then $\ell_{n+1} = 2\ell_n$ and w_{n+1} is obtained by Lemma 7.7.6. Otherwise, $\ell_{n+1} = \ell_n$ and w_{n+1} is obtained by Corollary 7.7.7. Note that the former case happens infinitely often, so $(\ell_n)_{n\geq 1}$ is unbounded. Also note that each ℓ_n is a power of 2.

Proof (of Theorem 7.7.5). Let $(w_n)_{n\geq 1}$ be a sequence given by Corollary 7.7.8. Let $x = 11w_1w_2w_3\cdots$ We first prove that x satisfies $x = \mathrm{even}(x)$. Note that $x[2^k + 1..2^{k+1}] = w_k$ for each $k \geq 1$ and $x[1..2^{k+1}] = 11w_1\cdots w_k$. The fact that $w_n = \mathrm{even}(w_{n+1})$ implies $x[2n] = x[n]$, for every $n \geq 3$. The cases for $n = 1$ and $n = 2$ hold because $x[1..4] = 1101$.

We prove that x is normal. Consider an arbitrary index n_0. By construction, w_{n_0} is ℓ_{n_0}-perfect, and for each $n \geq n_0$, w_n is also ℓ_{n_0}-perfect. For every word u of length ℓ_{n_0} and for every $n \geq n_0$,

$$\|x[1..2^{n+1}]\|_u \leq \|x[1..2^{n_0}]\|_u + \|w_{n_0}\ldots w_n\|_u.$$

Then, for every N such that $2^n \leq N < 2^{n+1}$ and $n \geq n_0$,

$$\begin{aligned}
\frac{\|x[1..N]\|_u}{N/\ell_{n_0}} &\leq \frac{\|x[1..2^{n+1}]\|_u}{N/\ell_{n_0}} \\
&\leq \frac{\|x[1..2^{n_0}]\|_u + \|w_{n_0}\ldots w_n\|_u}{N/\ell_{n_0}} \\
&\leq \frac{\|x[1..2^{n_0}]\|_u}{2^n/\ell_{n_0}} + \frac{\|w_{n_0}\ldots w_n\|_u}{2^n/\ell_{n_0}} \\
&= \frac{\|x[1..2^{n_0}]\|_u}{2^n/\ell_{n_0}} + \frac{(2^{n_0} + \ldots + 2^n)/(\ell_{n_0}2^{\ell_{n_0}})}{2^n/\ell_{n_0}} \\
&< \frac{\|x[1..2^{n_0}]\|_u}{2^n/\ell_{n_0}} + \frac{2}{2^{\ell_{n_0}}}.
\end{aligned}$$

For large values of N and n such that $2^n \leq N < 2^{n+1}$, the expression $\|x[1..2^{n_0}]\|_u/(2^n/\ell_{n_0})$ becomes arbitrarily small. We obtain for every word u of length ℓ_{n_0},

$$\limsup_{N\to\infty} \frac{\|x[1..N]\|_u}{N/\ell_{n_0}} \leq 3\; 2^{-\ell_{n_0}}.$$

Since the choice of ℓ_{n_0} was arbitrary, the above inequality holds for each ℓ_n. Since $(\ell_n)_{n\geq 1}$ is unbounded, the hypothesis of Lemma 7.4.1 is fulfilled, with $C = 3$, so we conclude that x is normal.

It is possible to compute a normal word x such that $x = even(x)$ in linear time.

7.8 Constructions of Absolutely Normal Numbers

The first constructions of absolutely normal numbers were given, independently, by Lebesgue [371] and Sierpiński [547], when the theory of computing was undeveloped. The numbers defined by these two constructions cannot be computed because they are just determined as the infimum of a set defined by infinite unions and intersections. The first example of a computable absolutely normal number was given by Turing [52, 570], and, unfortunately, it has doubly exponential time complexity. The computable reformulation of Sierpiński's construction [51] has also doubly exponential time complexity.

There are exponential algorithms that use analytic tools, such as Levin's construction [19, 375] of an absolutely normal number with fast convergence to normality and Schmidt's construction [529] of a number that is normal to all the bases in a prescribed set but not normal to the bases in the complement, see Theorem 7.9.3.

Some years ago, several efficient algorithms were published. Figueira and Nies gave in [222] an algorithm based on martingales with polynomial time complexity. Becher, Heiber, and Slaman [53] reworked Turing's strategy and obtained an algorithm with just above quadratic time complexity. Madritsch, Scheerer, and Tichy [397] adapted it and obtained an efficient algorithm to compute a number that is normal to all Pisot bases. Recently Lutz and Mayordomo [395] obtained an algorithm based on martingales with poly-logarithmic linear time complexity.

Another aspect in constructions of absolutely normal numbers is the speed of convergence to normality. Aistleitner et al. [8] constructed an absolutely normal real number x, so that for every integer b greater than or equal to 2 the discrepancy modulo 1 of the sequence $(b^n x)_{n\geq 0}$ is strictly below that realized by almost all real numbers (see Section 7.6) The construction yields an exponential algorithm that achieves a discrepancy estimate lower than that in Levin's work [375]. According to Scheerer's analysis [525], currently there are no other known constructions achieving a smaller discrepancy. The problem of the existence of an absolutely normal number computable with polynomial complexity having fast rate of convergence to normality remains open.

We will present two algorithms, and we will analyze their computational complexity. We first need some notation.

If v is a block of digits in base b, I_v denotes b-ary interval

$$(.v, .v + b^{-|v|})$$

Definition 7.8.1. Let x be a real in the unit interval, and let x_b be its expansion in base b. We define

$$\Delta_N(x_b) = \max_{d\in\{1,\dots,b\}} \left| \frac{|x_b[1\dots N]|_d}{N} - \frac{1}{b} \right| .$$

If w is a finite block of digits in base b, we just write $\Delta(w)$ instead of $\Delta_{|w|}(w)$.

7.8.1 Turing's Construction of Absolutely Normal Numbers

Theorem 7.8.2 (Turing 1937? [52, 570]). *There is an algorithm that computes the expansion in base* 2 *of an absolutely normal number y in the unit interval.*

The construction is done by steps. We will use n as the step number, and we will define the following functions of n: N_n is the number of digits looked at step n, b_n is the largest base considered at step n, and ε_n is the maximum difference between the expected frequency of digits and the tolerated frequency of digits at step n. It is required that b_n be nondecreasing and unbounded and ε_n be nonincreasing and goes to zero. Many instantiations of these functions can work.

Definition 7.8.3. Define the following functions of n,

$$\begin{aligned} N_n &= 2^{n_0+2n}, \text{where} n_0 = 11, \\ b_n &= \lfloor \log N_n \rfloor \\ \varepsilon_n &= 1/b_n. \end{aligned}$$

Define the following sets of real numbers,

$$\begin{aligned} E_0 &= (0,1), \text{ and for each } n \\ E_n &= \bigcap_{b\in\{2,\dots,b_n\}} \{x \in (0,1) : \Delta_{N_n}(x_b) < \varepsilon_n\}. \end{aligned}$$

The value n_0 has been selected so that the forthcoming calculations are simple. Observe that for every n, $b_n \geq 2$. Thus, for each n the set E_n consists of all the real numbers whose expansion in the bases 2,3,…, b_n exhibit good frequencies of digits in the first N_n digits. We write μ for Lebesgue measure.

Proposition 7.8.4. *For each n, E_n is a finite union of open intervals with rational endpoints, $E_{n+1} \subset E_n$, and $\mu E_n > 1 - N_n^2$.*

Proof. The values of N_n and ε_n satisfy the hypotheses of Lemma 7.3.5 with digits in base b (i.e., let k be N_n, let ℓ be 1, and let ε be ε_n),

$$\mu\{x \in (0,1) : \Delta_{N_n}(x_b) \geq \varepsilon_n\} < 2b^2 e^{-\varepsilon_n^2 b N_n/6.}$$

Then, for $b_n \leq \log N_n$, $\varepsilon \geq 1/\log N_n$ and $N_n > e^{10}$ can be checked that

$$\sum_{b=2}^{b_n} 2b^2 e^{-\varepsilon^2 b N_n/6} < 1/N_n^2.$$

Hence,

$$\mu E_n \geq 1 - \sum_{b=2}^{b_n} 2b^2 e^{-\varepsilon^2 b N_n/6} \geq 1 - 1/N_n^2.$$

Proposition 7.8.5. *The set $\bigcap_{n\geq 0} E_n$ has positive measure and consists just of absolutely normal numbers.*

Proof. From Proposition 7.8.4 follows that $\bigcap_{n\geq 0} E_n$ has positive measure. Suppose $x \in \bigcap_{n\geq 0} E_n$. Then, for every n, $x \in E_n$, so for each $b = 2, 3, \ldots, b_n$,

$$\Delta_{N_n}(x_b) \leq \varepsilon_n.$$

Let b be an arbitrary base, and let M be an arbitrary position. Let n be such that

$$N_n \leq M < N_{n+1}.$$

For each b smaller than b_n we have that for each digit d in $\{0, \ldots, b-1\}$,

$$\begin{aligned}
\frac{|x_b[1\ldots M]|_d}{M} &< \frac{|x_b[1\ldots N_{n+1}]|_d}{N_n} < \frac{N_{n+1}}{N_n}\left(\frac{1}{b} + \varepsilon_{n+1}\right) = 4\left(\frac{1}{b} + \varepsilon_{n+1}\right)\\
\frac{|x_b[1\ldots M]|_d}{M} &> \frac{|x_b[1\ldots N_n]|_d}{N_{n+1}} > \frac{N_n}{N_{n+1}}\left(\frac{1}{b} - \varepsilon_n\right) = \frac{1}{4}\left(\frac{1}{b} - \varepsilon_n\right).
\end{aligned}$$

Since ε_n is decreasing in n and goes to 0, we conclude that for each base $b = 2, 3 \ldots$,

$$\limsup_{N\to\infty} \frac{|x_b[1\ldots N]|_d}{N} < 4\frac{1}{b}.$$

Using the morphism that maps digits in base b^ℓ to words in base b, this is equivalent to say that for each base b, for every length ℓ, and for every word u of length ℓ,

$$\limsup_{N\to\infty} \frac{\|x_b[1\ldots \ell N]\|_u}{N} < 4\frac{1}{b^\ell}.$$

By Theorem 7.4.1, x is normal to every base b, hence absolutely normal.

Turing's construction selects nested binary intervals $I_1, I_2, \ldots$ such that, for each n, $\mu I_n = 1/2^n$. Each interval I_{n+1} is either the left half or the right half of I_n. The base-2 expansion of the computed number y is denoted with the sequences $y_1, y_2, \ldots$ which is the trace of the left/right selection at each step. Recall Definition 7.8.3 where the sets E_n are defined, for every $n \geq 0$.

```
*Initial step, $n = 0$.* $I_0 = (0,1)$, $E_0 = (0,1)$.

*Recursive step, $n > 1$.* Assume that in the previous step we have computed $I_{n-1}$.
    Let $I_n^0$ be left half of $I_{n-1}$ and $I_n^1$ be right half of $I_{n-1}$.
    If $\mu\left(I_n^0 \cap \bigcap_{j=0}^n E_j\right) > 1/N_n$ then let $I_n = I_n^0$ and $y_n = 0$.
    Else let $I_n = I_n^1$ and $y_n = 1$.
```

Proof (of Theorem 7.8.2). From Algorithm 7.8.1 follows that the intervals $I_1, I_2, \ldots$ are nested, and for each n, $\mu I_n = 1/2^n$. To prove the correctness of the algorithm, we need to prove that the following condition is invariant along every step n of the algorithm:

$$\mu\left(I_n \cap \bigcap_{j=1}^{n} E_j\right) > 0.$$

We prove it by induction on n. Recall $N_n = 2^{n_0+2n}$.

Base case $n = 0$.

$$\mu(I_0 \cap E_0) = \mu((0,1)) > \frac{1}{N_0^2} = \frac{1}{2^{2n_0}}.$$

Inductive case, $n > 0$. Assume as inductive hypothesis that

$$\mu\left(I_n \cap \bigcap_{j=0}^{n} E_j)\right) > \frac{1}{N_n}.$$

We now show it holds for $n+1$. Recall $\mu E_n > 1 - 1/N_n^2$. Then,

$$\mu\left(I_n \cap \bigcap_{j=0}^{n+1} E_j\right) = \mu\left(I_n \cap \bigcap_{j=0}^{n} E_j \cap E_{n+1}\right) > \frac{1}{N_n} - \frac{1}{N_{n+1}^2} > \frac{2}{N_{n+1}}.$$

Since the algorithm chooses I_{n+1} among I_n^0 and I_n^1 ensuring $\mu(I_{n+1} \cap \bigcap_{j=0}^{n+1} E_j) > 1/N_{n+1}$, we conclude $\mu(I_{n+1} \cap \bigcap_{j=0}^{n+1} E_j) > 1/N_{n+1}$ as required.

Finally, since $(I_n)_{n\geq 0}$ is a nested sequence of intervals and $\mu(I_n \cap \bigcap_{j=0}^{n} E_j) > 0$, for every n, we obtain that

$$\bigcap_{n\geq 0} I_n = \bigcap_{n\geq 0}\left(I_n \cap \bigcap_{j=0}^{n} E_j\right).$$

contains a unique real number y. By Lemma 7.8.5, all the elements in $\bigcap_{j\geq 0} E_j$ are absolutely normal. This concludes the proof of Theorem 7.8.2.

We now bound the number of mathematical operations computed by the algorithm to output the first n digits of the expansion of the computed number in a designated base. We do not count how many elementary operations are implied by each of the mathematical operations, which means that we neglect the computational cost of performing arithmetical operations with arbitrary precision.

Proposition 7.8.6. *Turing's algorithm has double exponential time complexity.*

Proof. At step n the algorithm computes the set $I_{n-1} \cap E_n$ by computing first the set

$$I_{n-1} \cap E_n = \bigcap_{b\in\{2,..,b_n\}} \{x \in I_{n-1} \cap E_{n-1} : \Delta_{N_n}(x_b) < \varepsilon_n\}$$

and choosing one of its halves. Then, the number of words to be examined to compute $I_n \cap E_n$ is

$$(b_n)^{N_n - N_{n-1} - (n-1)}.$$

Since $N_n = 2^{n_0+2n}$ and $b_n = \lfloor \log N_n \rfloor$, this number of words is in the order of

$$O\big((2n)^{2^{2n}}\big).$$

The examination of all these words requires $O\big((2n)^{2^{2n}}\big)$ mathematical operations. We conclude by noticing that using the set $I_n \cap E_n$ at step n the algorithm determines the $n - th$ binary digit of the computed number.

7.8.2 A Fast Construction of Absolutely Normal Numbers

We give a simplified version of the algorithm given by Becher, Heiber, and Slaman in [53].

Theorem 7.8.7. *There is an algorithm that computes an absolutely normal number x in nearly quadratic time completely: the first n digits in the expansion of x in base 2 are obtained by performing $O\big(n^2 \sqrt[4]{\log n}\big)$ mathematical operations.*

The following two lemmas are not hard to prove.

Lemma 7.8.8 ([53, Lemma 3.1]). *Let u and v be blocks and let ε be a positive real number.*

1. *If $\Delta(u) < \varepsilon$ and $\Delta(v) < \varepsilon$ then $\Delta(uv) < \varepsilon$.*
2. *If $\Delta(u) < \varepsilon$, $v = a_1 \dots a_{|v|}$ and $|v|/|u| < \varepsilon$ then $\Delta(vu) < 2\varepsilon$, and for every ℓ such that $1 \le \ell \le |v|$, $\Delta(ua_1a_2 \dots a_\ell) < 2\varepsilon$.*

Lemma 7.8.9 (Lemma 3.4 [53]). *For any interval I and any base b, there is a b-ary subinterval J such that $\mu J \ge \mu I/(2b)$.*

The next two definitions are the core of the construction.

Definition 7.8.10. A t-sequence $\overrightarrow{\sigma}$ is a sequence of intervals $(\sigma_2, \dots, \sigma_t)$ such that for each base $b = 2, \dots, t$, σ_b is b-ary, for each base $b = 3, \dots, t$, $\sigma_b \subset \sigma_{b-1}$ and $\mu\sigma_b \ge \mu\sigma_{b-1}/(2b)$.

Observe that the definition implies $\mu\sigma_t \ge (\mu\sigma_2)/(2^t t!)$.

Definition 7.8.11. A t-sequence $\overrightarrow{\tau} = (\tau_2, \dots, \tau_t)$ *refines* a t'-sequence $\overrightarrow{\sigma} = (\sigma_2, \dots, \sigma_{t'})$ if $t' \le t$ and $\tau_b \subset \sigma_b$ for each $b = 2, \dots, t'$. A refinement has *discrepancy less than* ε if for each $b = 2, ..t'$, there are words u, v such that $\sigma_b = I_u$, $\tau_b = I_{uv}$, and $\Delta(v) < \varepsilon$.

We say that an interval is b-ary of *order n* if it is of the form

$$\left(\frac{a}{b^n}, \frac{a+1}{b^n}\right)$$

for some integer a such that $0 \le a < b^n$. If σ_b and τ_b are b-ary intervals, and $\tau_b \subseteq \sigma_b$, we say that the *relative order* of τ_b with respect to σ_b is the *order* of τ_b minus the *order* of σ_b.

Lemma 7.8.12. *Let t be an integer greater than or equal to 2, let t' be equal to t or to $t+1$, and let ε be a positive real less than $1/t$. Then, any t-sequence $\overrightarrow{\sigma} = (\sigma_2, \dots, \sigma_t)$ admits a refinement $\overrightarrow{\tau} = (\tau_2, \dots, \tau_{t'})$ with discrepancy less than ε. The relative order of τ_2 can be any integer greater than or equal to* $\max(6/\varepsilon, 24(\log_2 t)(\log(t!))/\varepsilon^2)$.

Proof. First assume $t' = t$. We must pick a t-sequence $(\tau_2, \dots, \tau_t)$ that refines $(\sigma_2, \dots, \sigma_t)$ in a zone of low discrepancy. This is possible because the measure of the zones of large discrepancy decreases at an exponential rate in the order of the interval. To prove the lemma, we need to determine the relative order N of τ_2 such that the measure of the union of the bad zones inside σ_2 for the bases $b = 2, \dots t$ is strictly less than the measure of the set all the possible t-ary subintervals τ_t of σ_2.

Let L be the largest binary subinterval in σ_t. Consider the partition of L in 2^N binary intervals τ_2 of equal length. For each τ_2, apply iteratively Lemma 7.8.9 to define $\tau_3, \ldots, \tau_{t_n}$. In this form, we have defined 2^N many t_n-sequences $(\tau_2, \ldots \tau_t)$. Let S be the union of the set of all possible intervals τ_t over these 2^N many t_n-sequences. Hence, by the definition of t-sequence,

$$\mu S \geq \mu L/(2^t t!).$$

By Lemma 7.8.9,

$$\mu L \geq \mu \sigma_t/4.$$

And by the definition of t-sequence again,

$$\mu \sigma_t \geq \mu \sigma_2/(2^t t!).$$

Combining inequalities we obtain,

$$\mu S \geq \mu \sigma_2/(2^t t!\ 4\ 2^t t!)$$

Now consider the bad zones inside σ_2. For each $b = 2, \ldots t$, for a length N and a real value ε, consider the following set of intervals of relative order $\lceil N/\log_2 b \rceil$ with respect to σ_2,

$$B_{b, \lceil N/\log_2 b \rceil, \varepsilon} = \bigcup_{\substack{u \in \{0,\ldots,b-1\}^{\lceil N/\log_2 b \rceil} \\ \Delta(u) \geq \varepsilon}} I_u.$$

Thus, the actual measure of the bad zones is

$$\mu \sigma_2\ \mu\Big(\bigcup_{b=2,..,t} \mu B_{b, \lceil N/\log_2 b \rceil, \varepsilon} \Big)$$

Then, N must be such that

$$\mu \sigma_2\ \mu\Big(\bigcup_{b=2,..,t} B_{b, \lceil N/\log_2 b \rceil, \varepsilon} \Big) < \mu S.$$

Using Lemma 7.3.5 on the left and the inequality above for μS on the right it suffices that N be greater than $6/\varepsilon$ and also N be such that

$$2t^2 \cdot e^{-\varepsilon^2 (N/3 \log_2 t)} < \frac{1}{2^t t!} \frac{1}{4} \frac{1}{2^t t!}.$$

We can take N greater than or equal to $\max(6/\varepsilon, 24(\log_2 t)(\log(t!))/\varepsilon^2)$.

The case $t' = t + 1$ follows easily by taking first a t-sequence $\overrightarrow{\tau}$ refining $\overrightarrow{\sigma}$ with discrepancy less than ε. Definition 7.8.11 does not require any discrepancy considerations for τ_{t+1}. Take τ_{t+1} the largest $(t+1)$-ary subinterval of τ_t. By Lemma 7.8.9, $\mu\tau_{t+1} \geq (\mu\tau_t)/(2(t+1))$. This completes the proof of the lemma.

The algorithm considers three functions of the step number n: t_n is the maximum base to be considered at step n, ε_n is the maximum discrepancy tolerated at step n, and N_n is the number of digits in base 2 added at step n. It is required that t_n be increasing and ε_n be decreasing. Many instantiations of this functions can work.

The algorithm constructs $\overrightarrow{\sigma}_0, \overrightarrow{\sigma}_1, \overrightarrow{\sigma}_2, \ldots$ such that $\overrightarrow{\sigma}_0 = (0,1)$, and for each $n \geq 1$, $\overrightarrow{\sigma}_n$ is t_n-sequence that refines $\overrightarrow{\sigma}_{n-1}$ with discrepancy ε_n and such that the order of $\sigma_{n,2}$ is N_n plus the order of $\sigma_{n-1,2}$.

Definition 7.8.13. Define the following functions of n,

$$t_n = \max(2, \lfloor \sqrt[4]{\log n} \rfloor),$$
$$\varepsilon_n = 1/t_n,$$
$$N_n = \lfloor \log n \rfloor + n_{start},$$

where n_{start} is the minimum integer such that it validates the condition in Lemma 7.8.12. Thus, we require that for every positive n,

$$\lfloor \log n \rfloor + n_{start} \geq 6/\varepsilon_n \qquad \text{and}$$
$$\lfloor \log n \rfloor + n_{start} \geq 24(\log_2 t_n)(\log(t_n!))/\varepsilon_n^2.$$

Initial step, $n = 1$. $\overrightarrow{\sigma}_1 = (\sigma_2)$, with $\sigma_2 = (0,1)$.

Recursive step, $n > 1$. Assume $\overrightarrow{\sigma}_{n-1} = (\sigma_2, \ldots, \sigma_{t_{n-1}})$. Take $\overrightarrow{\sigma}_n = (\tau_2, \ldots, \tau_{t_n})$ the leftmost t_n-sequence such that it is refinement of $\overrightarrow{\sigma}_{n-1}$ with discrepancy less than ε_n such that the relative order of τ_2 is N_n.

Proof (of Theorem 7.8.7). Consider Algorithm 7.8.2. The existence of the sequence $\overrightarrow{\sigma}_1, \overrightarrow{\sigma}_2, \ldots$ is guaranteed by Lemma 7.8.12. We have to prove that the real number x defined by the intersection of all the intervals in the sequence is absolutely normal. We pick a base b and show that x is simply normal to base b. Let $\tilde{\varepsilon} > 0$. Choose n_0 so that $t_{n_0} \geq b$ and $\varepsilon_{n_0} \leq \tilde{\varepsilon}/4$. At each step n after n_0 the expansion of x in base b was constructed by appending blocks u_n such that $\Delta(u_n) < \varepsilon_{n_0}$. Thus, by Lemma 7.8.8 (item 1) for any $n > n_0$,

$$\Delta(u_{n_0} \ldots u_n) < \varepsilon_{n_0}.$$

Applying Lemma 7.8.8 (item 2a), we obtain n_1 such that for any $n > n_1$

$$\Delta(u_1 \ldots u_n) < 2\varepsilon_{n_0}.$$

Let $N_n^{(b)}$ be the relative order of τ_b with respect to σ_b. By Lemma 7.8.9,

$$\frac{N_n}{\log_2 b} \leq N_n^{(b)} \leq \frac{N_n + 1}{\log_2 b} + 1.$$

Since $N_n = \lfloor \log n \rfloor + n_{start}$, N_n grows logarithmically and so does $N_n^{(b)}$ for each base b. Then, for n sufficiently large,

$$N_n^{(b)} \leq \frac{N_n + 1}{\log_2 b} + 1 \leq 2\varepsilon_{n_0} \sum_{j=1}^{n-1} \frac{N_j}{\log_2 b} \leq 2\varepsilon_{n_0} \sum_{j=1}^{n-1} N_j^{(b)}.$$

By Lemma 7.8.8 (item 2b), we conclude that for n sufficiently large, if $u_n = a_1 \dots a_{|u_n|}$, then for every ℓ such that $1 \leq \ell \leq |u_n|$,

$$\Delta_\ell(u_1 \dots u_{n-1} a_1 \dots a_\ell) < \; 4\varepsilon_{n_0} < \tilde{\varepsilon}.$$

So, x is simply normal to base b for every $b \geq 2$.

We now analyze the computational complexity of the algorithm. Lemma 7.8.12 ensures the existence of the wanted t-sequence at each step n. To effectively find it, we proceed as follows. Divide the interval σ_2 into

$$2^{N_n}$$

equal binary intervals. In the worst case, for each of them, we need to check if it allocates a t_n-sequence $(\tau_2, \dots, \tau_{t_n})$ that refines $(\sigma_2 \dots, \sigma_{t_{n-1}})$ with discrepancy less than ε_n. Since we are just counting the number of mathematical operations ignoring the precision, at step n the algorithm performs

$$O\big(2^{N_n} t_n\big)$$

many mathematical operations. Since N_n is logarithmic in n and t_n is a rational power of $\log(n)$, we conclude that at step n the algorithm performs

$$O(n \sqrt[4]{\log n})$$

mathematical operations. Finally, in the first k steps, the algorithm will output at least k many digits of the binary expansion of the computed number having performed

$$O(k^2 \sqrt[4]{\log k})$$

many mathematical operations. This completes the proof of Theorem 7.8.7.

7.9 Normality, Non-normality, and Other Mathematical Properties

Recall that two positive integers are *multiplicatively dependent* if one is a rational power of the other. Then, 2 and 8 are dependent, but 2 and 6 are independent.

Theorem 7.9.1 (Maxfield 1953 [118]). *Let b and b' multiplicatively dependent. For any real number x, x is normal to base b if and only if x is normal to base b'.*

Theorem 7.9.2 (Cassels 1959 [135]). *Almost all real numbers in the middle third Cantor set (with respect to the uniform measure) are normal to every base which is not a power of* 3.

Theorem 7.9.3 (Schmidt 1961 [529]). *For any given set S of bases closed under multiplicative dependence, there are real numbers normal to every base in S and not normal to any base in its complement. Furthermore, there is a real x computable from S.*

Theorem 7.9.3 was improved in [58] to obtain lack of simple normality for the bases outside S instead of just lack of normality. Then Becher, Bugeaud, and Slaman [49] obtained the necessary and sufficient conditions on a set S for the existence of real numbers simply normal to every base in S and not simply normal to any base in its complement.

Theorem 7.9.4 (Becher, Bugeaud, and Slaman [49]). *Let S be a set of bases. There is a real x that is simply normal to exactly the elements in S if and only if*

1. *for each b, if b^k in S then b in S,*
2. *if infinitely many powers of b belong to S, then all powers of b belong to S.*

Moreover, the real x is computable from the set S. Furthermore, the set of real numbers that satisfy this condition has full Hausdorff dimension.

We end the section with references on the relation of normality and Diophantine approximations. The irrationality exponent m of a real number x reflects how well x can be approximated by rational numbers. Precisely, it is the supremum of the set of real numbers z for which the inequality

$$0 < \left| x - \frac{p}{q} \right| < \frac{1}{q^z}$$

is satisfied by an infinite number of integer pairs (p, q) with $q > 0$. Rational numbers have irrationality exponent equal to 1. Liouville numbers are those with infinite irrationality exponent. It follows from the fundamental work by [347] that almost all irrational numbers (with respect to Lebesgue measure) have irrationality exponent equal to 2. On the other hand, it follows from the theory of continued fractions that for every m greater than 2 or equal to infinity, there is a real number x with irrationality exponent equal to m.

Absolute normality places no restriction on irrationality exponents of irrational numbers. For every real number z greater than or equal to 2, there is an absolutely normal number with irrationality exponent equal to z. This existential result follows from Kaufman [338]. Bugeaud [117] showed there is an absolutely normal Liouville. In both cases, existence of such real numbers follows from the existence of a measure whose Fourier transform vanishes sufficiently quickly at infinity and which is supported by a subset of the real numbers with the appropriate irrationality exponent. Bugeaud's argument employs an adaptation of Kaufman's methods to the set of Liouville numbers due to Bluhm [92]. Becher, Heiber, and Slaman [54] exhibit a computable construction of an absolutely number Liouville number.

7.10 Selection

We consider the selection of symbols from an infinite word and define a word with the selected symbols. The general problem is which forms of selection preserve normality, that is, which families of functions f performing selection guarantee that $f(x)$ is normal when x is normal. Notice that if a selection procedure is allowed to read the symbol being decided, it would be possible to "select only zeroes" or yield similar schemes that do not preserve normality.

We consider three forms of selection. *Prefix selection* looks at just the prefix of length $i-1$ to decide whether the symbol at position i is selected. *Suffix selection* looks at just the suffix starting at position $i+1$ to decide whether symbol at position i is selected. *Two-sided selection* looks at the prefix of length $i-1$ and the suffix starting at position $i+1$ to decide the selection of the symbol at position i. Prefix selection is the selection defined by Agafonov [6].

Let $x = a_0a_1a_2\cdots$ be an infinite word over alphabet A. Let $L \subseteq A^*$ be a set of finite words over A and $X \subseteq A^\omega$ a set of infinite words over A.

The word obtained by *prefix selection* of x by L is $x \upharpoonright L = a_{i_0}a_{i_1}a_{i_2}a_{i_3}\cdots$ where $i_0, i_1, i_2, \cdots$ is the enumeration in increasing order of all the integers i such that $a_0a_2\cdots a_{i-1} \in L$.

The word obtained by *suffix selection* of x by X is $x \upharpoonleft X = a_{i_0}a_{i_1}a_{i_2}a_{i_3}\cdots$ where $i_0, i_1, i_2, \cdots$ is the enumeration in increasing order of all the integers i such that $a_{i+1}a_{i+2}a_{i+3}\cdots \in X$.

Theorem 7.10.1 (Agafonov [6]). *If $x \in A^\omega$ is normal and $L \subset A^*$ is rational then $x \upharpoonright L$ is also normal.*

Before giving the proof of Theorem 7.10.1, we discuss some other results. Agafanov's theorem can be extended to suffix selection by replacing the rational set of finite words L by a rational set of infinite words X. The proof of this theorem is quite technical, so we do not give it here.

Theorem 7.10.2 ([57]). *If $x \in A^\omega$ is normal and $X \subset A^\omega$ is rational, then $x \upharpoonleft X$ is also normal.*

The prefix and suffix selections cannot be combined to preserve normality: in general, two-sided selection does not preserve normality. For instance, selecting all symbols surrounded by two symbols 1 in a normal word over $\{0, 1\}$ always destroys normality: the factor 11 occurs more frequently than the factor 00 in the resulting word.

We now give three lemmas to be used in the proof of Theorem 7.10.1.

Lemma 7.10.3. *For any set of finite words L, the function $x \mapsto \langle x \upharpoonright L, x \upharpoonright A^* \setminus L \rangle$ is one-to-one.*

Proof. Let $y_1 = x \upharpoonright L$ and $y_2 = x \upharpoonright A^* \setminus L$. By definition, y_1 contains some symbols of x, in the same relative order, and y_2 contains the complement, also in the same relative order. It is possible to reconstruct x by interleaving appropriately the symbols in y_1 and y_2. For each $i \geq 1$, the i-th symbol of x comes from y_1 if and only if the prefix of length i of x is in L. Thus, there is a unique x such that $y_1 = x \upharpoonright L$ and $y_2 = x \upharpoonright A^* \setminus L$.

A (deterministic) *two-output transducer* is like transducer, but it has two output tapes. Each of its transitions has the form $p \xrightarrow{a|v,w} q$ where a is the symbol read on the input tape and v and w are the words written to the first and the second output tape, respectively.

An infinite word $x = a_0a_1a_2\cdots$ is *compressible* by a two-output transducer if there is an accepting run $q_0 \xrightarrow{a_0|v_0,w_0} q_1 \xrightarrow{a_1|v_1,w_1} q_2 \xrightarrow{a_2|v_2,w_2} \cdots$ that satisfies

$$\liminf_{n\to\infty} \frac{(|v_0v_2\cdots v_n| + |w_0w_2\cdots w_n|)}{n+1}\frac{\log|B|}{\log|A|} < 1.$$

The following lemma states that an extra output tape does not help for compressing.

Lemma 7.10.4. *An infinite word is compressible by a bounded-to-one two-output transducer if and only if it is compressible by a bounded-to-one transducer.*

Proof. The "if" part is immediate by not using one of the output tapes.

Suppose that x is compressible by the bounded-to-one two-output transducer $\mathscr{T}_2$. We construct a transducer $\mathscr{T}_1$ with a single output tape which also compresses x. The main idea is to merge the two outputs into the single tape without losing the bounded-to-one assumption. Let m be an integer to be fixed later. The transducer $\mathscr{T}_1$ simulates $\mathscr{T}_2$ on the input and uses two buffers of size m to store the outputs made by $\mathscr{T}_2$. Whenever one of the two buffers is full and contains m symbols, its content is copied to the output tape of $\mathscr{T}_1$ with an additional symbol in front of it. This symbol is either 0 or 1 to indicate whether the m following symbols comes from the first or the second buffer. This trick preserves the bounded-to-one assumption. This additional symbol for each block of size m increases the length of the output by a factor $(m+1)/m$. For m large enough, the transducer $\mathscr{T}_1$ also compresses x.

Lemma 7.10.5. *Let $x = a_0a_1a_2\cdots$ be a normal word, and let $q_0 \xrightarrow{a_0} q_1 \xrightarrow{a_1} q_2 \xrightarrow{a_2} \cdots$ be a run in a deterministic automaton. If the state q is visited infinitely often then* $\liminf_{n\to\infty} |\{i \le n : q_i = q\}|/n > 0$.

Proof. Let $\mathscr{A}$ be a deterministic automaton. For a state p and a finite word w, the unique state q such that $p \xrightarrow{w} q$ is denoted $p \cdot w$.

Let $q = q_1, \dots, q_n$ be the states occurring infinitely often in the run. For $1 \le i, j$ *len*, let $u_{i,j}$ be a word such that $q_i \cdot u_{i,j} = q_j$. Let us define the sequence of words $(w_k)_{1\le k\le n}$ by $w_1 = \lambda$ and $w_{k+1} = w_k u_{i,1}$ where $q_{k+1} \cdot w_k = q_i$. By definition, $q_k \cdot w_k = q$, and thus the finite run $q_i \xrightarrow{w_n} q_i \cdot w_n$ visits the state q for each i since w_i is a prefix of w_n. Since the number of occurrences of w_n in x converges to $1/|A|^{|w_n|}$, the result holds.

Proof (of Theorem 7.10.1). Let x be a normal word. Let $L \subset A^*$ be a rational language. We suppose by constriction that $x \upharpoonright L$ is not normal, and we show that x can be compressed, contradicting its normality.

Let $\mathscr{A}$ be a deterministic automaton accepting L. This automaton can be turned into a two-output transducer that outputs $x \upharpoonright L$ and $x \upharpoonright A^* \setminus L$ on its first and second output tapes, respectively. Each transition that leaves a final state copies its input symbol to the first output tape, and each transition that leaves a nonfinal state copies its input symbol to the second output tape. By hypothesis, $x \upharpoonright L$ is not normal and therefore can be compressed by some deterministic transducer. Combining, these two transducers yield a two-output transducer that compresses x. This later result holds because, by Lemma 7.10.5, the states that select symbols from x are visited at least linearly often. Then, by Lemma 7.10.4, x can be compressed and is not normal.