The Typical Number Is a Lexicon*

Cristian S. Calude† **Tudor Zamfirescu**‡

C.S. Calude, T. Zamfirescu. The Typical Number Is a Lexicon, *New Zealand Journal of Mathematics*, 27 (1998), 7-13.

Abstract

The parallelism between category and measure fails to hold true for random reals (i.e. reals having the sequence of digits in some base Chaitin-Martin-Löf random, [7, 4]): constructively, they are measure-one sets of first category [12]. In this note we strengthen some results in [6] by *constructively* proving that a class of pseudo-random reals has a σ-porous complement, so it is simultaneously residual and of measure-one. As a consequence, we constructively show that the typical number is a lexicon, i.e. even constructively most numbers do not obey any probability laws. To achieve our goal we prove a constructive version of (a weak form of) Lebesgue's Density Theorem, a result which might be interesting in itself.

1 Introduction

Denote by $\mathbf{N}$, $\mathbf{Q}$ the sets of natural and rational numbers, respectively. For every natural $b \geq 2$ put $B_b = \{0, 1, \ldots, b-1\}$. If X is a set, then X^+ denotes the free semigroup generated by X; the elements of X^+ are called words (over X). The length of a word $u = u_1u_2 \cdots u_n$ is $|u| = n$. The concatenation of the word u with itself i times is denoted by u^i. A word u is a *prefix* of a word v in case $v = uw$, for some word w; in this case we write $u \subset v$. A word u is *contained* in a word v in case $v = xuy$, for some words x, y.

For $u, v \in B_b^+$,

$$N_v(u) = \mathrm{card}\{1 \leq j \leq |u| \mid j \equiv 1 (\mathrm{mod}\ |v|), u_j u_{j+1} \cdots u_{j+|v|-1} = v\},$$

counts the occurrences of the word v in u. To compute $p_v(u)$, the relative frequency of the word $v \in B_b^+$ in $u \in B_b^+$, we group the elements of u in blocks of length $|v|$ (we ignore the last block in

*This paper has been completed during the second author's visit at the University of Auckland in 1995. The first author has been supported by AURC A18/XXXXX/62090/F3414030, 1994.

†Computer Science Department, The University of Auckland, Private Bag 92109, Auckland, New Zealand, e-mail: cristian@cs.auckland.ac.nz.

‡Institute of Mathematics, University of Dortmund, 44221 Dortmund, Federal Republic of Germany, e-mail: zamfi@steinitz.mathematik.uni-dortmund.de.

case it has length less than $|v|$) and we divide the number of occurrences of v in the sequence of blocks by the number of total blocks. Formally we get

$$p_v(u) = \frac{N_v(u)}{\frac{|u|}{|v|}} = \frac{|v|N_v(u)}{|u|}.$$

By B_b^ω we denote the set of all sequences $\mathbf{x} = x_1x_2\cdots x_n\cdots$, of elements in B_b. The prefix of length n of $\mathbf{x}$ is the word $\mathbf{x}(n) = x_1x_2\cdots x_n$. A sequence $\mathbf{x}$ contains a word u in case some prefix of $\mathbf{x}$ contains u.

Fix now a base $b \geq 2$. The function $\mathrm{val}_b : B_b^+ \cup B_b^\omega \to (0,1)$ is defined by: $\mathrm{val}_b(x_1x_2\cdots x_n) = \sum_{i=1}^n x_i b^{-i}$, $\mathrm{val}_b(\mathbf{x}) = \sum_{i=1}^\infty x_i b^{-i}$. The b-adic expansion $\mathrm{seq}_b(r)$ of a real number r in the interval $[0,1)$ is the unique sequence $\mathbf{x} = x_1x_2\cdots x_n\cdots \in B_b^\omega$ containing infinitely many digits different from $b-1$ such that $r = \mathrm{val}_b(\mathbf{x})$.

To each word $w \in B_b^+$ we associate the open interval $I_{b,w} = (\mathrm{val}_b(w), \mathrm{val}_b(w) + b^{-|w|}) \subset [0,1)$. The family $\{I_{b,w}\}_{w\in B_b^+}$ is a base for the natural topology on $[0,1]$. For a real $r \in [0,1)$ and a word $v \in B_b^+$ we define

$$p_{b,v}^+(r) = \limsup_{n\to\infty} p_v(\mathrm{seq}_b(r)(n)) \quad \text{and} \quad p_{b,v}^-(r) = \liminf_{n\to\infty} p_v(\mathrm{seq}_b(r)(n)).$$

We shall assume familiarity with, or access to [13] (for classical measure and category), [3] (for recursion theory), and [4] (for constructive measure and category).

2 Main Results

Following Jürgensen and Thierrin [9], a real number $r \in [0,1]$ is called *disjunctive in base* b in case $\mathrm{seq}_b(r)$ contains all possible words over B_b. Let us denote by $\mathcal{L}$ the set of numbers disjunctive in any base; call such a number *absolutely disjunctive* or a *lexicon*.[1] A lexicon contains all writings, which have been or will be ever written, in any possible language.[2] Disjunctivity is a "qualitative" analogue of normality [1], a (weaker) form of pseudo-randomness.

Let F be the recursive set $\{(b,\alpha,n,v) \mid b \geq 2, \alpha \in (0,1)\cap\mathbf{Q}, n \geq 1, v \in B_b^+\}$. For $(b,\alpha,n,v) \in F$ define

$$\mathcal{R}^+_{(b,\alpha,n,v)} = \{0 \leq r \leq 1 \mid \exists\, m \geq n \ \text{ s. t. } \ p_v(\mathrm{seq}_b(r)(m)) \geq \alpha\},$$
$$\mathcal{R}^-_{(b,\alpha,n,v)} = \{0 \leq r \leq 1 \mid \exists\, m \geq n \ \text{ s. t. } \ p_v(\mathrm{seq}_b(r)(m)) \leq \alpha\}.$$

It is readily seen that

$$\begin{aligned}
\bigcap_{b,\alpha,n,v} \mathcal{R}^-_{(b,\alpha,n,v)} &= \bigcap_{b,\alpha,v} \{0 \leq r \leq 1 \mid p_{b,v}^-(r) \leq \alpha\} \\
&= \bigcap_{b,v} \{0 \leq r \leq 1 \mid p_{b,v}^-(r) = 0\} \\
&= \{0 \leq r \leq 1 \mid \forall\, b \geq 2,\ \forall\, v \in B_b^+,\ p_{b,v}^-(r) = 0\},
\end{aligned}$$

[1]Disjunctivity is not invariant under the change of base [8].

[2]For a musical analogue see Karlheinz Essl 1992 interactive, real-time composition for computer-controlled piano titled "Lexikon-Sonate" at url `http://www.ping.at/users/essl/Lexikon-Sonate.html`.

and

$$\bigcap_{b,\alpha,n,v} \mathcal{R}^{+}_{(b,\alpha,n,v)} = \{0 \le r \le 1 \mid \forall\, b \ge 2,\ \forall\, v \in B_b^{+},\ p^{+}_{b,v}(r) = 1\}.$$

A set $R \subset [0,1]$ is residual if it contains the intersection of a countable family of open dense sets. To get a constructive version of this definition we require that the family of open dense sets is enumerated by a recursively enumerable (r.e.) set, and we have a constructive "witness" to guarantee that each basic open set $I_{b,u}$ intersects the family of open dense sets.

We are led to the following definition: A set $R \subset [0,1)$ is *constructively residual* if there exists an r.e. set $E \subset \{(b,u,m) \in \mathbf{N} \times \mathbf{N}^{+} \times \mathbf{N} \mid b \ge 2, u \in B_b^{+}, m \ge 1\}$ and a recursive function $f : \mathbf{N}^{+} \times \mathbf{N} \to \mathbf{N}^{+}$ such that the following three conditions hold true:

1. For all $b \ge 2, m \ge 1, u \in B_b^{+},\ f(u,m) \in B_b^{+}$.
2. $\bigcap_{m=1}^{\infty} \left(\bigcup_{(b,w,m)\in E} I_{b,w} \right) \subset R$.
3. For all $b \ge 2,\ m \ge 1,\ u \in B_b^{+}$ we have $u \subset f(u,m)$ and $(b, f(u,m), m) \in E$.

The complement of a constructively residual set is a constructive first Baire category set; as a consequence, a constructively residual set is residual, but the converse is false ([10, 4]). The statement "constructively, the typical number has, or most numbers have, property P" means that the set of all numbers with property P is constructively residual.

Lemma 2.1. *Constructively, most numbers are in*

$$\mathcal{R}^{+} = \bigcap_{(b,\alpha,n,v)\in F} \mathcal{R}^{+}_{(b,\alpha,n,v)}.$$

Proof. Fix a recursive bijection $\psi : \mathbf{N} \to F$ and define the auxiliary recursive functions $t : \mathbf{N} \times \mathbf{N} \times ([0,1) \cap \mathbf{Q}) \to \mathbf{N}$ and $\theta : F \times \mathbf{N}^{+} \to \mathbf{N}^{+}$ by

$$t(q,m,\alpha) = \left\lfloor \frac{\alpha q}{m(1-\alpha)} \right\rfloor + 1,$$

and

$$\theta((b,\alpha,n,v),u) = u0^{\max(n-|u|,0)} v^{t(\max(|u|,n),|v|,\alpha)}.$$

Fix $(b,\alpha,n,v) \in F$ and $u \in B_b^{+}$. We notice that

$$\begin{aligned} p_v(\theta((b,\alpha,n,v),u)) &\ge \frac{|v|}{|\theta((b,\alpha,n,v),u)|}\, t(\max(|u|,n),|v|,\alpha) \\ &= \frac{|v|\; t(\max(|u|,n),|v|,\alpha)}{\max(|u|,n) + |v|\; t(\max(|u|,n),|v|,\alpha)} \ge \alpha, \end{aligned}$$

and

$$m = |\theta((b,\alpha,n,v),u)| \ge n,$$

so

$$I_{b,\theta((b,\alpha,n,v),u)} \subset \mathcal{R}^+_{(b,\alpha,n,v)}.$$

For every word $u \in B_b^+$,

$$I_{b,u} \cap I_{b,\theta((b,\alpha,n,v),u)} \neq \emptyset,$$

so the open set

$$\bigcup_{u \in B_b^+} I_{b,\theta((b,\alpha,n,v),u)}$$

is dense in $[0,1]$.

In conclusion, the set of real numbers the lemma speaks about is constructively residual using the r.e. set

$$E = \{(b, \theta((b,\alpha,n,v),u), m) \mid b \geq 2, u \in B_b^+, m \geq 1, \psi(m) = ((b,\alpha,n,v),u)\},$$

and the recursive function $f : \mathbf{N}^+ \times \mathbf{N} \to \mathbf{N}^+$ defined by $f(u,m) = \theta(\psi(m), u)$. □

In view of the fact that for every rational $\alpha \in (0,1)$, and all words $u, v \in B_b^+$ there exists a word $w \in B_b^+$ such that $N_v(uw) \leq \alpha$, we can modify the definition of θ in the above proof appropriately to guarantee the inequality $p_v(\theta((b,\alpha,n,v),u)) \leq \alpha$. So, the set

$$\mathcal{R}^- = \bigcap_{(b,\alpha,n,v) \in F} \mathcal{R}^-_{(b,\alpha,n,v)}$$

is constructively residual. Finally, the set $\mathcal{R}^- \cap \mathcal{R}^+$ is constructively residual too. We have proven:

Theorem 2.2. *Constructively, for most numbers $r \in [0,1]$, using any base b and choosing any word $v \in B_b^+$*

$$p^-_{b,v}(r) = 0 \text{ and } p^+_{b,v}(r) = 1.$$

As an immediate consequence we derive a constructive version of a result due to Oxtoby and Ulam [12], p. 877:

Corollary 2.3. *Constructively, a typical number does not obey the law of large numbers.*

Proof. Indeed, the set of all reals $r \in [0,1]$ such that in their dyadic expansion the digits 0 and 1 appear with probability one-half lies in the complement of the constructively residual set from Theorem 2.2. □

Random numbers are transcendental, as they are non-computable [4]. Liouville numbers, i.e. numbers in which arbitrarily sparse ones occur, are transcendental numbers which appear to be "typically" non-random. Jürgensen and Thierrin [9] have proved the existence, for one arbitrary base, of uncountably many Liouville disjunctive numbers. In fact, a stronger result can be proven:

Corollary 2.4. *Constructively, the typical Liouville number is a lexicon.*

Proof. Since the constructively residual set in Theorem 2.2 is a subset of $\mathcal{L}$, constructively most numbers from $[0,1]$ are in $\mathcal{L}$. But most reals are constructively Liouville numbers, as the proof from [13] p. 8 can be readily constructivized. □

The set of all numbers each of which is a lexicon is large not only in the sense of constructive category, but also in the sense of constructive measure theory: this set contains all random numbers [5], so it has constructive measure-one by a result of Martin-Löf [10, 4]. This suggests that constructively $\mathcal{L}$ may contain nearly all elements of $[0,1]$. But what does "nearly all" mean? Classically, a set contains nearly all numbers if its complement is σ-porous [15]. The complement of a σ-porous set is simultaneously residual and of measure-one (but the complement of a null set of first category may well not contain nearly all elements, [14]). The fact that a porous set has measure-one is a consequence of Lebesgue's Density Theorem [13] which, to the best of our knowledge [2], has not (yet) been proven constructively.

Call a set $M \subset [0,1]$ *constructively megaporous* if there exist a base $b \geq 2$, a rational number $r \in (0,1)$ and a recursive function $f : B_b^+ \to B_b^+$ such that each interval $I_{b,u}$ of length less than r contains a subinterval $I_{b,f(u)}$ disjoint from M and having length greater than $rb^{-|u|}$. An r.e. union of constructively megaporous sets is called *constructively σ-megaporous.* More precisely, M is constructively σ-megaporous if $M = \cup_{n=1}^{\infty} M_n$, and there exist two recursive functions $T : \mathbf{N} \times \mathbf{N}^+ \to \mathbf{N}^+$, $R : \mathbf{N} \to \mathbf{Q}$ such that M_n is constructively megaporous under $T(n,.)$ and $R(n)$. We say that "constructively, nearly every point of $[0,1]$ enjoys property P" if the set of points not enjoying P is constructively σ-megaporous.

Theorem 2.5. *Constructively, nearly every number is a lexicon.*

Proof. Let $\gamma : \{(b,w) \mid b \geq 2,\ w \in B_b^+\} \to \mathbf{N}$ be a recursive bijection, and define the recursive functions $T(n,u) = uw,\ R(n) = b^{-|w|} - 1$, whenever $n = \gamma(b,w)$. Again, if $n = \gamma(b,w)$, we put $L_n = \{0 \leq x \leq 1 \mid w \text{ is not contained in } \mathrm{seq}_b(x)\}$. It is seen that $[0,1] \setminus \mathcal{L} = \cup_{i=1}^{\infty} L_i$, and each L_n is constructively megaporous with respect to the base b, the recursive function $T(n,.)$ and the rational $R(n)$. □

Recall that, following [10, 4], a set $S \subset [0,1]$ is constructively null (with respect to the Lebesgue measure μ) if there exists a base $b \geq 2$ and an r.e. set $G \subset B_b^+ \times \mathbf{N}$ such that

$$S \subset \bigcap_{n=1}^{\infty} \left(\bigcup_{(x,n)\in G} I_{b,x} \right),$$

and

$$\lim_{n\to\infty} \mu \left(\bigcup_{(x,n)\in G} I_{b,x} \right) = 0, \text{ constructively.}$$

The following result is a constructive version of (a weak form of) Lebesgue's Density Theorem.

Theorem 2.6. *Every constructively σ-megaporous set is constructively null.*

Proof. Due to a theorem of Martin-Löf [10, 4], the union of all constructive null sets is a (maximal) constructive null set. Consequently, it is enough to prove the theorem for constructive megaporous sets. Let M be constructively megaporous with respect to the base b, the rational r and the recursive function f. To estimate the size of M we will generate, in a recursive way, smaller and smaller coverings of M. We start with an integer n such that $b^{-n} < r$. For a word $w \in B_b^+$ put $E(w) = \{y \in B_b^+ \mid w \subset y, |y| = |f(w)|, \text{ and } y \neq f(w)\}$. The first covering is

$$M \subset \bigcup_{|u|=n} I_{b,u}.$$

The second iteration is

$$M \subset \bigcup_{|u|=n} \bigcup_{v_1 \in E(u)} I_{b,v_1} = \bigcup_{|u|=n} I_{b,u} \setminus I_{b,f(u)}.$$

The measure of this covering is

$$\begin{aligned} \mu\left(\bigcup_{|u|=n} I_{b,u} \setminus I_{b,f(u)}\right) &= \sum_{|u|=n} \mu(I_{b,u} \setminus I_{b,f(u)}) \\ &= \sum_{|u|=n} (b^{-|u|} - b^{-|f(u)|}) \\ &\leq \sum_{|u|=n} b^{-|u|}(1-r) = 1-r. \end{aligned}$$

In general, a proof by induction shows that

$$\begin{aligned} M &\subset \bigcup_{|u|=n} \bigcup_{v_1 \in E(u)} \cdots \bigcup_{v_k \in E(v_{k-1})} \bigcup_{v_{k+1} \in E(v_k)} I_{b,v_{k+1}} \\ &= \bigcup_{|u|=n} \bigcup_{v_1 \in E(u)} \cdots \bigcup_{v_k \in E(v_{k-1})} I_{b,v_k} \setminus I_{b,f(v_k)} \end{aligned}$$

and

$$\mu\left(\bigcup_{|u|=n} \bigcup_{v_1 \in E(u)} \cdots \bigcup_{v_k \in E(v_{k-1})} \bigcup_{v_{k+1} \in E(v_k)} I_{b,v_{k+1}}\right) < (1-r)^{k+1}.$$

We conclude that M is constructively null with respect to the r.e. family $G = \{(w,n) \in B_b^+ \times \mathbf{N} \mid w \in F_n, n = 1, 2, \ldots\}$, where $F_0 = \{u \in B_b^+ \mid |u| = n\}$ and $F_{k+1} = \{u \in B_b^+ \mid u \in E(w), \text{ for some } w \in F_k\}$. $\square$

The above result is stronger than the classical one as, for instance, constructive null sets are even smaller than classical null sets: the union of all null sets coincides with the whole space, while, the union of all constructive null sets is a constructive null set [10, 4].

By constructively proving that a typical real number is a lexicon we have shown that most numbers do not obey any probability laws. In particular, a typical number does not obey the law of large numbers. The collapse of the law of large numbers has been noticed in non equilibrium processes, processes which seem to hesitate among various possible directions of evolution (cf. Nicolis and Prigogine [11]). Finally, our result is "global" and gives no information about specific numbers (is π a lexicon?).

Acknowledgment

The authors are indebted to Peter Hertling for correcting a mistake in the proof of Theorem 2.6 and the anonymous referees for their valuable critical remarks.

References

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

[2] D. S. Bridges. Private communication to C. Calude, 28 April 1995.

[3] C. Calude. *Theories of Computational Complexity*, North-Holland, Amsterdam, 1988.

[4] C. Calude. *Information and Randomness: An Algorithmic Perspective*, Springer-Verlag, Berlin, 1994.

[5] C. Calude, H. Jürgensen. Randomness as an invariant for number representations, in H. Maurer, J. Karhumäki, G. Rozenberg (eds.). *Results and Trends in Theoretical Computer Science*, Springer-Verlag, Berlin, 1994, 44–66.

[6] C. Calude, T. Zamfirescu. Most numbers obey no probability laws, *Technical Report No 112, 1995*, Department of Computer Science, The University of Auckland, New Zealand, 4 pp.

[7] G. J. Chaitin. *Information, Randomness and Incompleteness, Papers on Algorithmic Information Theory*, World Scientific, Singapore, 2nd ed., 1990.

[8] P. Hertling. Disjunctive ω-words and real numbers, *J. UCS* 2 (1996), 549-568.

[9] H. Jürgensen, G. Thierrin. Some structural properties of ω-languages, *13th Nat. School with Internat. Participation "Applications of Mathematics in Technology"*, Sofia, 1988, 56–63.

[10] P. Martin-Löf. *Notes on Constructive Mathematics*, Almqvist & Wiksell, Stockholm, 1970.

[11] G. Nicolis, I. Prigogine. *Exploring Complexity*, W. H. Freeman, New York, 1989.

[12] J. C. Oxtoby, S. M. Ulam. Measure-preserving homeomorphisms and metrical transitivity, *Ann. Math.* 42(1941), 874–925.

[13] J. C. Oxtoby. *Measure and Category*, Springer-Verlag, Berlin, 1971.

[14] L. Zajíček. Porosity and σ-porosity, *Real Analysis Exchange* 13 (1987/88), 314–350.

[15] T. Zamfirescu. Porosity in convexity, *Real Analysis Exchange* 15 (1989/90), 424–436.