ON THE NORMALITY OF NORMAL NUMBERS

DAVAR KHOSHNEVISAN

1. Introduction

Let x be a real number between zero and one. We can write it, in binary form, as $x = 0.x_1x_2\cdots$, where each x_j takes the values zero and one. Of primary interest to us, here, are numbers x such that half of their binary digits are zeros and the remaining half are ones. More precisely, we wish to know about numbers x that satisfy

$$\lim_{n\to\infty} \frac{\#\{1 \le j \le n : x_j = 1\}}{n} = \frac{1}{2}, \tag{1}$$

where # denotes cardinality.

Equation (1) characterizes some, but not all, numbers between zero and one. For example, $x = 0$ and $x = 1$ do not satisfy (1), and the following do: $0.10101010\cdots$, $0.01010101\cdots$, $0.001001001\cdots$, etc.

We can observe that the preceding three examples are all "periodic." Thus, one can ask if there are numbers that satisfy (1) whose digits are *not* periodic. Borel's normal number theorem gives an affirmative answer to this question. In fact, Borel's theorem implies, among other things, that the collection of non-normal numbers has zero length. Surprisingly, this fact is intimately connected to diverse areas in mathematics [probability, ergodic theory, b-adic analysis, analytic number theory, and logic] and theoretical computer science [source coding, random number generation, and complexity theory].

In this article, we describe briefly a general form of Borel's normal-number theorem and some of its consequences in other areas of mathematics and computer science. Our discussion complements some related papers by Berkes, Philipp, and Tichy [3], Harman [16], and Queffélec [22].

Date: April 20, 2007.
The author's work is supported in part by a grant from the National Science Foundation.

2. **Borel's theorem**

Given an integer $b \geq 2$ and a number x between zero and one, we can always write $x = \sum_{j=1}^{\infty} x_j b^{-j}$, where the x_j's take values in $\{0, \ldots, b-1\}$. This representation is unique for all but b-adic rationals; for those we opt for the representation in which $x_j = 0$ for all but a finite number of j's.

We may think of $\{0, \ldots, b-1\}$ as our "alphabet," in which case a "word" of length m is nothing but the sequence $\sigma_1 \ldots \sigma_m$, where each σ_j can take any of the values $0, \ldots, b-1$.

Let w be a fixed word of finite length m, and choose and fix integers $n \geq m$, as well as a real number $x \in [0, 1]$. We can then define $N_n^b(x\,; w)$ to be the number of times the word w appears continguously among $(x_1, \ldots, x_n)$. The reader is invited to verify that $N_n^{10}(0.5, \{5\}) = N_n^2(0.5, \{1\}) = 1$ for all $n \geq 1$.

A number x is said to be *simply normal in base* b if

$$\lim_{n\to\infty} \frac{N_n^b(x\,;\{j\})}{n} = \frac{1}{b} \qquad \text{for all letters } j \in \{0, \ldots, b-1\}. \tag{2}$$

That is, x is simply normal in base b when, and only when, all possible letters in the alphabet $\{0, \ldots, b-1\}$ are distributed equally in the b-ary representation of x. Numbers that satisfy (1) are simply normal in base 2.

More generally, a number x is said to be *normal in base* b if given any finite word w with letters from the alphabet $\{0, \ldots, b-1\}$,

$$\lim_{n\to\infty} \frac{N_n^b(x\,; w)}{n} = \frac{1}{b^{|w|}}, \tag{3}$$

where $|w|$ denotes the length of the word m. The number $a = 0.101010\cdots$ is simply normal, but not normal, in base 2. This can be seen, for example, by inspecting the two-letter word "11."

Still more generally, we say that $x \in [0, 1]$ is *simply normal* if it is simply normal in all bases $b \geq 2$, and [absolutely] *normal* if it is normal in all bases $b \geq 2$. These definitions are all due to Borel [4].

So far we have seen some "periodic" examples of numbers that are normal in base 2, say. It is also possible to construct non-periodic examples. The first such number was constructed by Champernowne [9]. He proved in 1933 that the decimal number $0.1234567891011121314\ldots$, obtained by concatenating all numerals in their natural order, is simply normal in base 10. Also, he conjectured that $0.13571113171923\ldots$, obtained by concatenating all primes, is simply normal in base 10. Champernowne's conjecture was verified in 1946 by Copeland and Erdős [10].

Also, it is possible to construct numbers that are simply normal in one base, but not in another. For example, the simply normal binary number $a = 0.101010\cdots$ is not normal in base 10, since $a = 2/3 = 0.\bar{6}$ in decimal notation.

It was conjectured by Borel [5] in 1950 that all irrational algebraic numbers are normal; see also Mahler's 1976 lectures [20] wherein he proved, among other things, that Champernowne's number is transendental. Unfortunately, not much further progress has been made in this direction. For example, it is not known whether household numbers such as e, π, $\ln 2$, or $\sqrt{2}$ are simply normal in any given base.[1] We do not even know if $\sqrt{2}$ has infinitely-many 5's [say] in its decimal expansion!

I hasten to add that there are compelling arguments that support the conjecture that e, π, $\sqrt{2}$, and a host of other nice algebraic irrationals, are indeed normal; see Bailey and Crandall [1].

The preceding examples, and others, were introduced in order to better understand the remarkable *normal number theorem* of Borel [4] from 1909:

Theorem 2.1 (Borel)**.** *Almost every number in* $[0\,,1]$ *is normal.*

The veracity of this result is now beyond question. However, to paraphrase Doob [12, p. 591], Borel's original derivation contains an "unmendably faulty" error. Borel himself was aware of the gap in his proof, and asked for a complete argument. His plea was answered a year later by Faber [15, p. 400], and also later by Hausdorff [17].

Theorem 2.1 suggests that it should be easy to find normal numbers. But I am not aware of any easy-to-describe numbers that are even simply normal. Recently, Becker and Figueira [2] have built on a constructive proof of Theorem 2.1, due to Sierpiński [26], to prove the existence of computable normal numbers. Their arguments suggest possible ways for successively listing out the digits of some normal numbers. But a direct implementation of this program appears to be at best arduous.

Borel's theorem is generally considered to be one of the first contributions to the modern theory of mathematical probability; a fact which Borel himself was aware of [4]. In order to describe this connection to probability, let us select a number X uniformly at random from the interval $[0\,,1]$. The key feature of this random selection process is that for all Borel sets $A \subseteq [0\,,1]$,

$$\mathrm{P}\{X \in A\} = \text{Lebesgue measure of } A, \tag{4}$$

[1] $x > b$ is said to be [simply] normal in base b when x/b is [simply] normal in base b.

where P denotes probability.

We can write X in b-ary form as $\sum_{j=1}^{\infty} X_j b^{-j}$. Borel's central observation was that $\{X_j\}_{j=1}^{\infty}$ is a collection of *independent* random variables, each taking the values 0, 1, ..., $b-1$ with equal probability. Then he proceeded to [somewhat erroneously] prove his *strong law of large numbers*, which was the first of its kind. Borel's law of large numbers states that for all letters $j \in \{0, \ldots, b-1\}$,

$$\mathrm{P}\left\{\lim_{n\to\infty} \frac{\mathbf{1}_{\{X_1=j\}} + \cdots + \mathbf{1}_{\{X_n=j\}}}{n} = \frac{1}{b}\right\} = 1, \tag{5}$$

where $\mathbf{1}_A$ denotes the characteristic function of A. It follows readily from (5) that with probability one X is simply normal in base b. Because there are only a countable number of integers $b \geq 2$, this proves that X is simply normal. Normality of X is proved similarly, but one analyses blocks of digits in place of single digits at a time.

Let $\mathscr{N}_b$ denote the collection of all numbers normal in base b. The preceding argument implies that $\mathrm{P}\{X \in \bigcap_{b=2}^{\infty} \mathscr{N}_b\} = 1$. This and (4) together imply Theorem 2.1.

We conclude this section by making a few more comments:

- In 1916 Weyl [28] described a tantalizing generalization of Theorem 2.1 that is nowadays called Weyl's equidistribution theorem. In this connection, we mention also the thesis of Wall [27].
- Riesz [23] devised a slightly more direct proof of Theorem 2.1. His derivation appeals to Birkhoff's ergodic theorem in place of Borel's (or more generally, Kolmogorov's) strong law of large numbers. But the general idea is not dissimilar to the proof outlined above.
- The probabilistic interpretation of Theorem 2.1 has the following striking implication:

  Finite-state, finite-time random number generators do not exist. (6)

  Of course, this does not preclude the possibility of generating a random number one digit at a time. But it justifies our present day use of *psuedo* random-number generators; see Knuth [18] for more on this topic. Remarkably, a complexity theory analogue to (6) completely characterizes all normal numbers; see Schnorr and Stimm [25] and Bourke, Hitchcock, and Vinochandran [6]. In this general direction, see also the interesting works of Chaitin [8] and Lutz [19].

- The proof of Borel's theorem is more interesting than the theorem itself, because it identifies the digits of a uniform random variable as independent and identically distributed. Such sequences have interesting properties that are not described by Theorem 2.1. Next we mention one of the many possible examples that support our claim.

  Let $R_n(x)$ denote the length of the largest run of ones in the first n binary digits of x. [A run of ones is a continguous sequences of ones.] Then, according to a theorem of Erdős and Rényi [14] from 1970,

$$\lim_{n\to\infty} \frac{R_n(x)}{\log_2(n)} = 1 \qquad \text{for almost every } x \in [0\,,1]. \tag{7}$$

  Because this involves words of arbitrarily large length, it is not a statement about normal number per se. There are variants of (7) that are valid in all bases, as well.

3. Unbiased sampling

As was implied earlier, one of the perplexing features of normal numbers is that they are abundant (Theorem 2.1), and yet we do not know of a single concrete number that is normal. This has puzzled many researchers, but appears to be a fact that goes beyond normal numbers, or even the usual structure of the real line.

Next we present two examples that examine analogous problems in similar settings. These examples suggest the following general principle: *Quite often, schemes that involve taking "unbiased samples from large sets" lead to notions of normality that are hard to pinpoint concretely.* I believe that this principle explains our inability in deciding whether or not a given number is normal. But I have no proof [nor disproof].

3.1. Cantor's set.

For our first example, let us consider the ternary Cantor set C, which we can think of as all numbers $x \in [0\,,1]$ whose ternary expansion $\sum_{j=1}^\infty x_j 3^{-j}$ consists only of digits $x_j \in \{0\,,2\}$.

In order to take an "unbiased sample" from C, it is necessary and sufficient to find a probability measure on C that is as "flat" as possible. [We are deliberately being vague here.] There are many senses in which the most flat probability measure on C can be identified with the restriction m_C of the usual $\log_3(2)$-dimensional Hausdorff measure to C. That is, m_C is the Cantor–Lebesgue measure. Now it is not difficult to show that m_C can be defined directly as follows:

$$m_C(A) := \mathrm{P}\left\{\sum_{j=1}^{\infty}\frac{X_j}{3^j}\in A\right\} \qquad \text{for all Borel sets } A\subseteq[0\,,1], \tag{8}$$

where $X_1, X_2, \ldots$ are independent random variables, taking the values zero and two with probability 1/2 each. A ready application of the strong law of large numbers then reveals that the following holds for m_C-almost every $x \in C$:

$$\lim_{n\to\infty}\frac{N_n^3(x\,;w)}{n} = \frac{1}{2^{|w|}} \qquad \text{for all words } w\in\bigcup_{k=1}^{\infty}\{0\,,2\}^k. \tag{9}$$

We say that a number $x \in C$ is normal in the Cantor set C if it satisfies (9). Although m_C-almost every number in C is normal in C, I am not aware of any concrete examples. On the other hand, I point out that we do not know very many concrete numbers in C at all—be they normal or otherwise. By analogy, this suggests the sightly uncomfortable fact that we do not know very many numbers—normal as well as non-normal—in $[0\,,1]$.

3.2. **Wiener's measure.** Our second example is a little more sophisticated, and assumes that the reader has a good background in probability and infinite-dimensional analysis. This example can be skipped on first reading.

Let $C^0[0\,,1]$ denote the colletion of all continuous functions $f : [0\,,1] \to \mathbf{R}$ such that $f(0) = 0$. As usual, we endow $C^0[0\,,1]$ with the [compact-open] topology of uniform convergence. We seek to find a flat measure on $C^0[0\,,1]$. The latter space is an abelian group under pointwise addition. But it fails to support a Haar measure, primarily because it is not locally compact. Nonetheless, there *are* measures on $C^0[0\,,1]$ that are relatively flat. We describe one next.

Define $C^0_n[0\,,1]$ to be the collection of all piecewise continuous functions $f : [0\,,1] \to \mathbf{R}$ that have the following properties:

(1) $f(0) = 0$;
(2) $|f((k+1)/n) - f(k/n)| = 1/\sqrt{n}$ for all $k = 0, \ldots, n-1$;
(3) for all x between k/n and $(k+1)/n)$, we define $f(x)$ by interpolating linearly the values $f(k/n)$ and $f((k+1)/n)$.

Evidently, $\cup_{n=1}^{\infty} C^0_n[0\,,1]$ is dense in $C^0[0\,,1]$. Because each $C^0_n[0\,,1]$ is a finite collection, the uniform measure m_n on $C^0_n[0\,,1]$ is decidedly the flattest—or least biased—measure on $C^0_n[0\,,1]$. In 1952, Donsker [11] proved that m_n converges (weak-$*$) to Wiener's measure [29]. In this sense, Wiener's measure W is a relatively flat measure on $C^0[0\,,1]$. Moreover, it can be shown that given any complete orthonormal basis $\{\phi_k\}_{k=1}^\infty$ of $L^2[0\,,1]$, there exist independent, standard normal variables $\{X_k\}_{k=1}^\infty$—defined on $C^0[0\,,1]$—such that

$$f \simeq \sum_{k=1}^{\infty} \phi_k X_k(f) \qquad \text{for } W\text{-almost all } f \in C^0[0\,,1]. \tag{10}$$

Choose and fix a function $G \in L^1(e^{-x^2/2}dx)$. Then, by the strong law of large numbers,

$$\lim_{n\to\infty} \frac{\sum_{k=1}^n G(X_k(f))}{n} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(x) e^{-x^2/2}\, dx, \tag{11}$$

for W-almost all $f \in C^0[0\,,1]$. A density argument shows that [29] holds for all $G \in L^1(e^{-x^2/2}dx)$ outside a single W-null set of functions f. Thus, we call a function $f \in C^0[0\,,1]$ *normal* if it satisfies (11) for all $G \in L^1(e^{-x^2/2}dx)$.

Choose and fixed a function $f \in C^0[0\,,1]$. We now ask the following question:

$$\text{Is } f \text{ is normal?} \tag{12}$$

The $X_k(f)$'s are defined abstractly via $X_k := \lambda(\phi_k)$, where $\lambda : L^2[0\,,1] \to L^2(W)$ denote Wiener's isometry. Because each X_k is defined via a Hilbertian isometry, it is not clear how one understands $X_k(f)$ for a fixed $f \in C^0[0\,,1]$. Thus, there is little hope in answering (12), even when f is nice.

4. **Non-normal numbers**

At first glance, one might imagine that because normal numbers are so complicated, non-normal numbers are not. Unfortunately, this is not the case. We conclude this article by mentioning two striking results that showcase some of the complex beauty of non-normal numbers.

4.1. **Eggleston's theorem.**

Let us choose and fix a base $b \geq 2$ and a probability vector $\boldsymbol{p} := (p_0\,, \ldots, p_{b-1})$; that is, $0 \leq p_j \leq 1$ and $p_0 + \cdots + p_{b-1} = 1$. Consider the set

$$\mathscr{E}(\boldsymbol{p}) := \left\{ x \in [0\,,1] : \lim_{n\to\infty} \frac{N_n^b(x\,; \{j\})}{n} = p_j \text{ for all } j = 0\,, \ldots, b-1 \right\}. \tag{13}$$

Note that if any one of the p_j's is different from $1/b$, then all elements of $\mathscr{E}(\boldsymbol{p})$ are non-normal. In 1949, Eggleston [13] confirmed a conjecture of I. J. Good by deriving the following result.

Theorem 4.1 (Eggleston)**.** *The Hausdorff dimension of* $\mathscr{E}(\boldsymbol{p})$ *is the thermodynamic entropy* $\sum_{j=0}^{b-1} p_j \log_b(1/p_j)$*, where* $0 \times \log_b(1/0) := 0$*.*

This theorem is true even if $p_0 = \cdots = p_{b-1} = 1/b$, but yields a weaker result than Borel's theorem in that case. Ziv and Lempel [30] developed related ideas in the context of source coding.

4.2. **Cassels's theorem.** For the second, and final, example of this article we turn to a striking theorem of Cassels [7] from 1959:

Theorem 4.2 (Cassels)**.** *Define the function* $f : [0\,,1] \to \mathbf{R}$ *by*

$$f(x) := \sum_{j=1}^{\infty} \frac{x_j}{3^j}, \tag{14}$$

where $x_1, x_2, \ldots$ *denote the binary digits of* x*. Then, for almost every* $x \in [0\,,1]$*,* $f(x)$ *is simply normal with respect to every base* b *that is not a power of* 3*.*

It is manifestly true that Cassels's $f(x)$ is not normal in bases 3, 9, etc. Hence, non-normal numbers too have complicated structure. We end our discussion by making two further remarks:

- Cassels's theorem answered a question of Hugo Steinhaus, and was later extended by Schmidt [24]. Pollington [21] derived a further refinement.
- Because $2f$ is a bijection between $[0\,,1]$ and the Cantor set C, Cassels's theorem constructs an uncountable number of points in $\frac{1}{2}C$ that are simply normal with respect to every base b that is not a power of 3. Not surprisingly, we do not have any concrete examples of such numbers.

References

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

[2] Becker, Verónica, and Santiago Figueira (2002). An example of a computable absolutely normal number, *Theoret. Comput. Sci.* **270**, 947–958

[3] Berkes, István, Walter Philipp, and Robert F. Tichy (2007), Empirical processes in probabilistic number theory: the LIL for the discrepancy of $(n_k\omega) \bmod 1$, *Illinois J. Math.* **50**(*1–4*), 107–145

[4] Borel, Émile (1909). Les probabilités dénombrables et leurs applications arithmétiques, *Supplemento di rend. circ. Mat. Palermo* **27**, 247–271

[5] Borel, Émile (1950). Sur les chiffres décimaux de $\sqrt{2}$ et divers probléme de probabilités en chaîne, *C. R. Acad. Sci. Paris* **230**, 591–593

[6] Bourke, C., J. M. Hitchcock, and N. V. Vinochandran (2005). Entropy rates and finite-state dimension, *Theoret. Comput. Sci.* **349**(*3*), 392–406.

[7] Cassels, J. W. S. (1959). On a problem of Steinhaus about normal numbers, *Colloq. Math.* **7**, 95–101

[8] Chaitin, G. J. (1975). A theory of program size formally identical to information theory, *ACM* **22**, 329–340

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

[10] Copeland, A. H. and P. Erdős (1946). Note on normal numbers, *Bull. Amer. Math. Soc.* **52**, 857–860

[11] Donsker, Monroe D. (1952). Justification and extension of Doob's heuristic approach to the Kolmogorov–Smirnov statistic, *Ann. Math. Statistics* **23**, 277–281

[12] Doob, Joseph L. (1996). The development of rigor in mathematical probability, *The Amer. Math. Monthly* **103**(4), 586–595

[13] Eggleston, H. G. (1949). The fractional dimension of a set defined by decimal properties, *Quart. J. Math.*, Oxford Ser. 20, 31–36

[14] Erdős, P. and A. Rényi (1970). On a new law of large numbers, *J. Analyse Math.* **23**, 103–111.

[15] Faber, Georg (1910). Über stetige Funktionen, II, *Math. Ann.* **69**, 372–443

[16] Harman, Glyn (2002). One hundred years of normal numbers, *Number Theory for the Millennium, II* (Urbana, IL, 2000), 149–166, A. K. Peters, Natick, MA

[17] Hausdorff, Felix (1914). *Grundzüge der Mengerlehre*, Verlag von Veit & Comp., Leipzig

[18] Knuth, Donald E. (1981). *The Art of Computer Programming*, Vol. 2, Addison-Wesley Publishing Co., Reading, Mass. second Edition

[19] Lutz, Jack H. (2003). The dimensions of individual strings and sequences *Inform. and Comput.* **187**(*1*), 49–79 (English summary)

[20] Mahler, Kurt (1976). *Lectures on Transcendental Numbers*, Lecture Notes in Math. **546**, Springer Verlag, Berlin, New York

[21] Pollington, A. D. (1981). The Hausdorff dimension of a set of normal numbers, *Pacific J. Math.* **95**(*1*), 193–204

[22] Queffélec, Martine (2006). Old and new results on normality, *IMS Lecture Notes–Monograph Series* [Dynamics and Stochastics], Vol. 48, 225–236, Institute of Mathematical Statistics, Hayward, California

[23] Riesz, Frédéric (1945). Sur la théorie ergodique, *Comment. Math. Helv.* **17**, 221–239

[24] Schmidt, Wolfgang M. (1960). On normal numbers, *Pacific J. Math.* **10**, 661–672

[25] Schnorr, C. P. and H. Stimm (1972). Endliche Automaten und Zufallsfolgen, *Acta Inform.* **1**, 345–359

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

[27] Wall, Donald Dines (1949). *Normal numbers*, Ph.D. Thesis, University of California, Berkeley, Berkeley, California.

[28] Weyl, Herman (1916). Über die Gleichverteilung von Zahlen mod. Eins, *Math. Ann.* **77**, 313–352

[29] Wiener, Norbert (1923). Differential space, *J. Math. Physics* **2**, 131–174

[30] Ziv, Jacob, and Abraham Lempel (1978). Compression of individual sequences via variable-rate coding, *IEEE Trans. Inform. Theory* **24**(*5*), 530–536

Davar Khoshnevisan: Department of Mathematics, The University of Utah, 155 S. 1400 E. Salt Lake City, UT 84112–0090

E-mail address: davar@math.utah.edu

URL: http://www.math.utah.edu/~davar