21. Walking on real numbers (2013)

Paper 21: Francisco J. Aragon Artacho, David H. Bailey, Jonathan M. Borwein and Peter B. Borwein, "Walking on real numbers," *Mathematical Intelligencer*, vol. 35 (2013), p. 42–60. With permission of Springer.

Synopsis:

As mentioned earlier, an age-old question (one that has been the motivation for both ancient and computer-age computations of π and other constants) is that of whether (and why) these digits are "random," which is usually taken to be the property of normality: a number α is said to be b-normal if every m-long string in the base-b expansion of α appears, in the limit, with frequency precisely $1/b^m$. It is straightforward to show, using probability and/or measure theory, that almost all real numbers must be normal, but it has been very difficult to prove normality for any of the classical constants of mathematics.

In this paper, the authors apply some new computer-based tools to address these questions. One of these is to cast, say, the base-4 digits of a constant, as a random walk, where at each step one moves a unit right, up, left or down depending on whether the digit at the given position is 0, 1, 2 or 3 (with similar extensions to other number bases). These analyses produce some surprising results, showing, for instance, that although the Stoneham numbers are provably normal, they exhibit self-similarity properties atypical of a truly "random" sequence. Numerous results along this line are presented.

Keywords: Curiosities, Graphical Representation, Normality, Random Walks

D.H. Bailey, J.M. Borwein, *Pi: The Next Generation*,
DOI 10.1007/978-3-319-32377-0_21

Walking on Real Numbers

Francisco J. Aragón Artacho, David H. Bailey, Jonathan M. Borwein, and Peter B. Borwein

The digit expansions of $\pi, e, \sqrt{2}$, and other mathematical constants have fascinated mathematicians from the dawn of history. Indeed, one prime motivation for computing and analyzing digits of π is to explore the age-old question of whether and why these digits appear "random." The first computation on ENIAC in 1949 of π to 2037 decimal places was proposed by John von Neumann so as to shed some light on the distribution of π (and of e) [15, pg. 277–281].

One key question of some significance is whether (and why) numbers such as π and e are "normal." A real constant α is b-normal if, given the positive integer $b \geq 2$, every m-long string of base-b digits appears in the base-b expansion of α with precisely the expected limiting frequency $1/b^m$. It is a well-established, albeit counterintuitive, fact that given an integer $b \geq 2$, almost all real numbers, in the measure theory sense, are b-normal. What's more, almost all real numbers are b-normal simultaneously for all positive integer bases (a property known as "absolutely normal").

Nonetheless, it has been surprisingly difficult to prove normality for well-known mathematical constants for any given base b, much less all bases simultaneously. The first constant to be proven 10-normal is the Champernowne number, namely the constant 0.12345678910111213141516..., produced by concatenating the decimal representation of all positive integers in order. Some additional results of this sort were established in the 1940s by Copeland and Erdős [26].

At present, normality proofs are not available for any well-known constant such as $\pi, e, \log 2, \sqrt{2}$. We do not even know, say, that a 1 appears one-half of the time, in the limit, in the binary expansion of $\sqrt{2}$ (although it certainly appears to), nor do we know for certain that a 1 appears infinitely often in the decimal expansion of $\sqrt{2}$. For that matter, it is widely believed that *every* irrational algebraic number (i.e., every irrational root of an algebraic polynomial with integer coefficients) is b-normal to all positive integer bases b, but there is no proof, not for any specific algebraic number to any specific base.

In 2002, one of the present authors (Bailey) and Richard Crandall showed that given a real number r in [0,1), with r_k denoting the k-th binary digit of r, the real number

$$\alpha_{2,3}(r) := \sum_{k=1}^{\infty} \frac{1}{3^k 2^{3^k + r_k}} \tag{1}$$

is 2-normal. It can be seen that if $r \neq s$, then $\alpha_{2,3}(r) \neq \alpha_{2,3}(s)$, so that these constants are all distinct. Since r can range over the unit interval, this class of constants is uncountable. So, for example, the constant $\alpha_{2,3} = \alpha_{2,3}(0) = \sum_{k \geq 1} 1/(3^k 2^{3^k}) = 0.0418836808315030\ldots$ is provably 2-normal (this special case was proven by Stoneham in 1973 [43]). A similar result applies if 2 and 3 in formula (1) are replaced by any pair of coprime integers (b, c) with $b \geq 2$ and $c \geq 2$ [10]. More recently, Bailey and Michal Misiurewicz were able to establish 2-normality of $\alpha_{2,3}$ by a simpler argument, by utilizing a "hot spot" lemma proven using ergodic theory methods [11].

In 2004, two of the present authors (Bailey and Jonathan Borwein), together with Richard Crandall and Carl Pomerance, proved the following: If a positive real y has algebraic degree $D > 1$, then the number $\#(y, N)$ of 1-bits in the binary expansion of y through bit position N satisfies $\#(y, N) > CN^{1/D}$, for a positive number C (depending on y) and all sufficiently large N [5]. A related result has been obtained by Hajime Kaneko of Kyoto University in Japan [37]. However, these results fall far short of establishing b-normality for any irrational algebraic in any base b, even in the single-digit sense.

Supported in part by the Director, Office of Computational and Technology Research, Division of Mathematical, Information, and Computational Sciences of the U.S. Department of Energy, under contract number DE-AC02-05CH11231.

DOI 10.1007/s00283-012-9340-x

1 Twenty-First Century Approaches to the Normality Problem

In spite of such developments, there is a sense in the field that more powerful techniques must be brought to bear on this problem before additional substantial progress can be achieved. One idea along this line is to study directly the decimal expansions (or expansions in other number bases) of various mathematical constants by applying some techniques of scientific visualization and large-scale data analysis.

In a recent paper [4], by accessing the results of several extremely large recent computations [46, 47], the authors tested the first roughly four trillion hexadecimal digits of π by means of a Poisson process model: in this model, it is extraordinarily unlikely that π is not normal base 16, given its initial segment. During that work, the authors of [4], like many others, investigated visual methods of representing their large mathematical data sets. Their chosen tool was to represent these data as walks in the plane.

In this work, based in part on sources such as [22, 23, 21, 19, 14], we make a more rigorous and quantitative study of these walks on numbers. We pay particular attention to π, for which we have copious data and which—despite the fact that its digits can be generated by simple algorithms—behaves remarkably "randomly."

The organization of the article is as follows. In Section 2 we describe and exhibit uniform walks on various numbers, both rational and irrational, artificial and natural. In the next two sections, we look at quantifying two of the best-known features of random walks: the expected distance traveled after N steps (Section 3) and the number of sites visited (Section 4) In Section 5 we describe two classes for which normality and nonnormality results are known, and one for which we have only surmise. In Section 6 we show various examples and leave some open questions. Finally, in Appendix 7 we collect the numbers we have examined, with concise definitions and a few digits in various bases.

2 Walking on Numbers

2.1 Random and Deterministic Walks

One of our tasks is to compare deterministic walks (such as those generated by the digit expansion of a constant) with pseudorandom walks of the same length. For example, in Figure 1 we draw a uniform pseudorandom walk with one million base-4 steps, where at each step the path moves one unit east, north, west, or south, depending on the whether the pseudorandom iterate at that position is 0, 1, 2, or 3. The color indicates the path followed by the walk—it is shifted up the spectrum (red-orange-yellow-green-cyan-blue-purple-red) following an HSV scheme with S and V equal to one. The HSV (hue, saturation, and value) model is a cylindrical-coordinate representation that yields a rainbow-like range of colors.

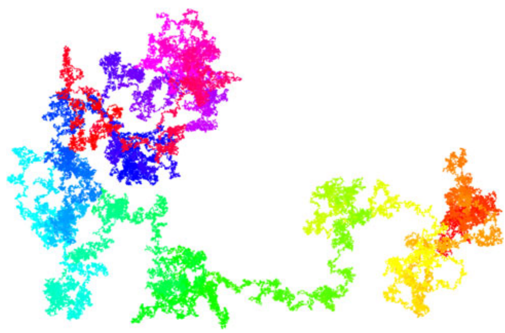
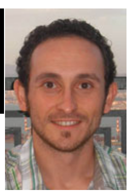


Figure 1. A uniform pseudorandom walk.

AUTHORS



FRANCISCO J. ARAGÓN ARTACHO earned his Ph.D. in optimization in 2007 at the University of Murcia, Spain. After working for a business in Madrid for a year, he took a postdoctoral position at the University of Alicante, supported by the program "Juan de la Cierva." In 2011, he became a Research Associate at the Priority Research Centre for Computer-Assisted Research Mathematics and its Applications (CARMA), University of Newcastle, Australia, under the direction of Jonathan Borwein, with whom he's currently collaborating on several projects.

Centre for Computer Assisted Research
Mathematics and its Applications
(CARMA)
University of Newcastle
Callaghan, NSW 2308
Australia
e-mail: francisco.aragon@ua.es

DAVID H. BAILEY is a Senior Scientist at Lawrence Berkeley National Laboratory. Before coming to the Berkeley Lab in 1998, he was at NASA's Ames Research Center for 14 years. Bailey has received the Chauvenet Prize from the Mathematical Association of America, the Sidney Fernbach Award from the IEEE Computer Society, and the Gordon Bell Prize from the Association for Computing Machinery. He is the author of five books, including *Mathematics by Experiment: Plausible Reasoning in the 21st Century*, coauthored with Jonathan Borwein.

Lawrence Berkeley National Laboratory
Berkeley, CA 94720
USA
e-mail: dhbailey@lbl.gov

Let us now compare this graph with that of some rational numbers. For instance, consider these two rational numbers $Q1$ and $Q2$:

```
Q1=
1049012271677499437486619280565448601617567358491560876166848380843
1443584472528755516292470277595555704537156793130587832477297720217
7081818796590637365767487981422801328592027861019258140957135748704
7122902674651513128059541953997504202061380373822338959713391954
/
1612226962694290912940490066273549214229880755725468512353395718465
1913530173488143140175045399694454793530120643833272670970079330526
2920303509209736004509554561365966493250783914647728401623856513742
95294530896122681527488756156580761624107880751845994219387748 35
```

```
Q2=
7278984857066874130428336124347736557760097920257997246066053320967
1510416153622193809833373062647935595557849662263315110631091226096 6
7568778977976821682512653537303069288477901523227013159658247897670
304354024902954939421310910639340148496028139 52
/
1118707184315428172047608747409173378543817936412916114431306628996
5259377090978187244251666337745459152093558288671765654061273733231
7877736113382974861639142628415265543797274479692427652260844707187
532155254872952853725026318685997495262134665215
```

At first glance, these numbers look completely dissimilar. However, if we examine their digit expansions, we find that they are very close as real numbers: the first 240 decimal digits are the same, as are the first 400 base-4 digits.

But even more information is exhibited when we view a plot of the base-4 digits of $Q1$ and $Q2$ as deterministic walks, as shown in Figure 2. Here, as above, at each step the path moves one unit east, north, west, or south, depending on the whether the digit in the corresponding position is 0, 1, 2, or 3, and with color coded to indicate the overall position in the walk.

JONATHAN M. BORWEIN is currently Laureate Professor in the School of Mathematical and Physical Sciences and is Director of the Priority Research Centre in Computer Assisted Research Mathematics and its Applications at the University of Newcastle. An ISI highly cited scientist and former Chauvenet prize winner, he has published widely in various fields of mathematics. Two of his recent books are *Convex Functions*, with Jon Vanderwerff, and *Modern Mathematical Computation with Mathematica*, with Matt Skerritt.

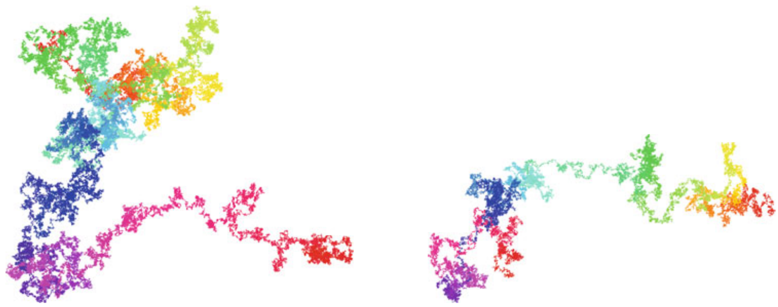
Centre for Computer Assisted Research
Mathematics and its Applications
(CARMA)
University of Newcastle
Callaghan, NSW 2308
Australia
e-mail: jonathan.borwein@newcastle.edu.au

PETER B. BORWEIN is the founder and Executive Director of the IRMACS Research Center at Simon Fraser University. He holds a Burnaby Mountain Chair and is an award-winning mathematician (Chauvenet, Ford, and Hasse Prizes, CUFA BC Academic of the Year). His primary research interests are in analysis and number theory but always with an overarching interest in the computational and experimental aspects. He has authored several hundred research papers and more than a dozen books. For more than 30 years he has collaborated extensively with his brother Jon.

IRMACS
Simon Fraser University
Burnaby, BC V5A 1S6
Canada

(a) A 440-step walk on $Q1$ base 4. **(b)** An 8240-step walk on $Q2$ base 4.

Figure 2. Walks on the rational numbers $Q1$ and $Q2$.



(a) $Q3 = \frac{3624360069}{7000000001}$ **(b)** $Q4 = \frac{123456789012}{1000000000061}$

Figure 3. Walks on the first million base-10 digits of the rationals $Q3$ and $Q4$ from [39].

The rational numbers $Q1$ and $Q2$ represent the two possibilities when one computes a walk on a rational number: either the walk is bounded as in Figure 2(a) (for any walk with more than 440 steps one obtains the same plot), or it is unbounded but repeating some pattern after a finite number of digits as in Figure 2(b).

Of course, not all rational numbers are that easily identified by plotting their walks. It is possible to create a rational number whose period is of any desired length. For example, the following rational numbers from [39],

$$Q3 = \frac{3624360069}{7000000001} \quad \text{and} \quad Q4 = \frac{123456789012}{1000000000061},$$

have base-10 periodic parts with length 1,750,000,000 and 1,000,000,000,060, respectively. A walk on the first million digits of both numbers is plotted in Figure 3. These huge periods derive from the fact that the numerators and denominators of $Q3$ and $Q4$ are relatively prime, and the denominators are not congruent to 2 or 5. In such cases, the period P is simply the discrete logarithm of the denominator D modulo 10; or, in other words, P is the smallest n such that $10^n \bmod D = 1$.

Graphical walks can be generated in a similar way for other constants in various bases—see Figures 2 through 7. Where the base $b \geq 3$, the base-b digits can be used to a select, as a direction, the corresponding base-b complex root of unity—a multiple of 120° for base three, a multiple of 90° for base four, a multiple of 72° for base 5, etc. We generally treat the case $b = 2$ as a base-4 walk, by grouping together pairs of base-2 digits (we could render a base-2 walk on a line, but the resulting images would be much less interesting). In Figure 4 the origin has been marked, but since this information is not that important for our purposes, and can be approximately deduced by the color in most cases, it is not indicated in the others. The color scheme for Figures 2 through 7 is the same as the above, except that Figure 6 is colored to indicate the number of returns to each point.

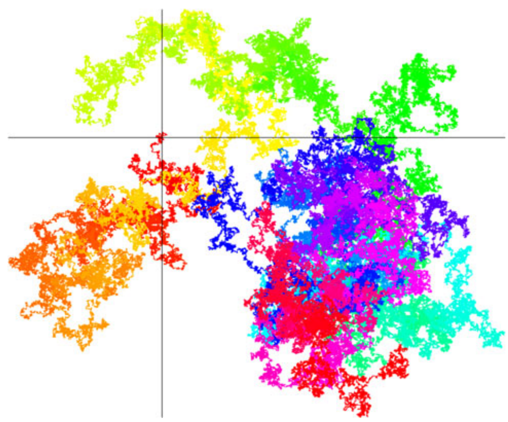


Figure 4. A million-step base-4 walk on e.

2.2 Normal Numbers as Walks

As noted previously, proving normality for specific constants of interest in mathematics has proven remarkably difficult. The tenor of current knowledge in this arena is illustrated by [45, 14, 34, 38, 40, 39, 44]. It is useful to know that, while small in measure, the "absolutely abnormal" or "absolutely non-normal" real numbers (namely those that are not b-normal for any integer b) are residual in the sense of topological category [1]. Moreover, the Hausdorff–Besicovitch dimension of the set of real numbers having no asymptotic frequencies is equal to 1. Likewise the set of Liouville numbers has measure zero but is of the second category [18, p. 352].

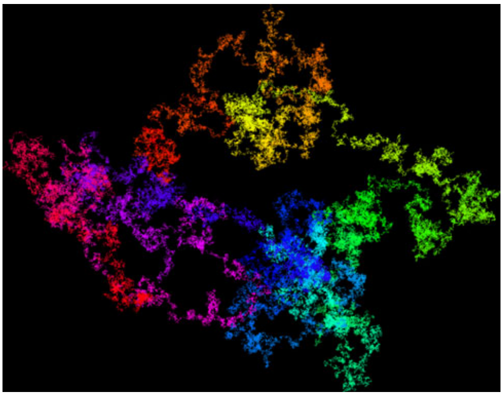


Figure 5. A walk on the first 100 billion base-4 digits of π (normal?).

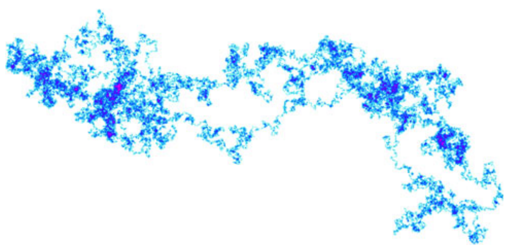


Figure 6. A walk on the first 100 million base-4 digits of π, colored by number of returns (normal?). Color follows an HSV model (green-cyan-blue-purple-red) depending on the number of returns to each point (where the maxima show a tinge of pink/red).

One question that has possessed mathematicians for centuries is whether π is normal. Indeed, part of the original motivation of the present study was to develop new tools for investigating this age-old problem.

In Figure 5 we show a walk on the first 100 billion base-4 digits of π. This may be viewed dynamically in more detail online at http://gigapan.org/gigapans/106803, where the full-sized image has a resolution of 372,224 × 290,218 pixels (108.03 gigapixels in total). This must be one of the largest mathematical images ever produced. The computations for creating this image took roughly a month, where several parts of the algorithm were run in parallel with 20 threads on CARMA's MacPro cluster.

By contrast, Figure 6 exhibits a 100 million base-4 walk on π, where the color is coded by the number of returns to the point. In [4], the authors empirically tested the normality of its first roughly four trillion hexadecimal (base-16) digits using a Poisson process model, and they concluded that, according to this test, it is "extraordinarily unlikely" that π is not 16-normal (of course, this result does not pretend to be a proof).

In what follows, we propose various methods of analyzing real numbers and visualizing them as walks. Other methods widely used to visualize numbers include the matrix representations shown in Figure 8, where each pixel is colored depending on the value of the digit to the right of the decimal point, following a left-to-right up-to-down direction (in base 4 the colors used for 0, 1, 2, and 3 are red, green, cyan, and purple, respectively). This method has been mainly used to visually test "randomness." In some cases, it clearly shows the features of some numbers; as for small periodic rationals, see Figure 8(c). This scheme also shows the nonnormality of the number $\alpha_{2,3}$—see Figure 8(d) (where the horizontal red bands correspond to the strings of zeroes)—and it captures some of the special peculiarities of the Champernowne's number C_4 (normal) in Figure 8(e). Nevertheless, it does not reveal the apparently nonrandom behavior of numbers such as the Erdős–Borwein constant; compare Figure 8(f) with Figure 7(e). See also Figure 21.

As we will see in what follows, the study of normal numbers and suspected normal numbers as walks will permit us to compare them with true random (or pseudorandom) walks, obtaining in this manner a new way to empirically test "randomness" in their digits.

3 Expected Distance to the Origin

Let $b \in \{3, 4, \ldots\}$ be a fixed base, and let $X_1, X_2, X_3, \ldots$ be a sequence of independent bivariate discrete random variables whose common probability distribution is given by

$$P\left(X = \begin{pmatrix} \cos\left(\frac{2\pi}{b}k\right) \\ \sin\left(\frac{2\pi}{b}k\right) \end{pmatrix}\right) = \frac{1}{b} \quad \text{for } k = 1, \ldots, b. \qquad (2)$$

Then the random variable $S^N := \sum_{m=1}^{N} X_m$ represents a base-b random walk in the plane of N steps.

The following result on the asymptotic expectation of the distance to the origin of a base-b random walk is probably known, but being unable to find any reference in the literature, we provide a proof.

Theorem 3.1 *The expected distance to the origin of a base-b random walk of N steps is asymptotically equal to* $\sqrt{\pi N}/2$.

Proof. By the multivariate central limit theorem, the random variable $1/\sqrt{N}\sum_{m=1}^{N}(X_m - \mu)$ is asymptotically bivariate normal with mean $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and covariance matrix M, where μ is the two-dimensional mean vector of X and M is its 2×2 covariance matrix. By applying Lagrange's trigonometric identities, one obtains

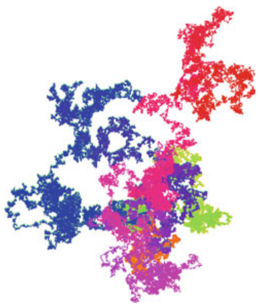
$$\mu = \begin{pmatrix} \frac{1}{b}\sum_{k=1}^{b} \cos\left(\frac{2\pi}{b}k\right) \\ \frac{1}{b}\sum_{k=1}^{b} \sin\left(\frac{2\pi}{b}k\right) \end{pmatrix} = \frac{1}{b}\begin{pmatrix} -\frac{1}{2} + \frac{\sin\left((b+1/2)\frac{2\pi}{b}\right)}{2\sin(\pi/b)} \\ \frac{1}{2}\cot(\pi/b) - \frac{\cos\left((b+1/2)\frac{2\pi}{b}\right)}{2\sin(\pi/b)} \end{pmatrix}$$
$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \qquad (3)$$

Thus,

$$M = \frac{1}{b}\begin{bmatrix} \sum_{k=1}^{b} \cos^2\left(\frac{2\pi}{b}k\right) & \sum_{k=1}^{b} \cos\left(\frac{2\pi}{b}k\right)\sin\left(\frac{2\pi}{b}k\right) \\ \sum_{k=1}^{b} \cos\left(\frac{2\pi}{b}k\right)\sin\left(\frac{2\pi}{b}k\right) & \sum_{k=1}^{b} \sin^2\left(\frac{2\pi}{b}k\right) \end{bmatrix}. \qquad (4)$$

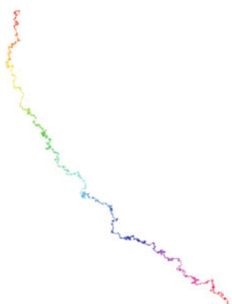
(a) A million-step walk on $\alpha_{2,3}$ base 3 (normal?).

(b) A 100,000-step walk on $\alpha_{2,3}$ base 6 (nonnormal).

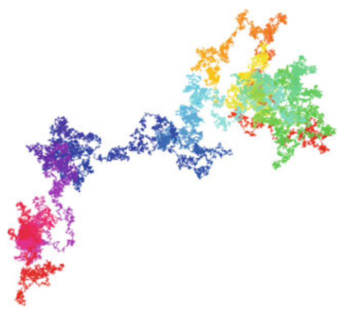


(c) A million-step walk on $\alpha_{2,3}$ base 2 (normal).

(d) A 100,000-step walk on Champernowne's number C_4 base 4 (normal).



(e) A million-step walk on $EB(2)$ base 4 (normal?).



(f) A million-step walk on $CE(10)$ base 4 (normal?).

Figure 7. Walks on various numbers in different bases.

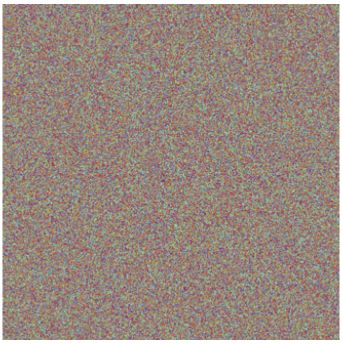
Since

$$\sum_{k=1}^{b}\cos^2\left(\frac{2\pi}{b}k\right) = \sum_{k=1}^{b}\frac{1+\cos\left(\frac{4\pi}{b}k\right)}{2} = \frac{b}{2},$$
$$\sum_{k=1}^{b}\sin^2\left(\frac{2\pi}{b}k\right) = \sum_{k=1}^{b}\frac{1-\cos\left(\frac{4\pi}{b}k\right)}{2} = \frac{b}{2},$$
$$\sum_{k=1}^{b}\cos\left(\frac{2\pi}{b}k\right)\sin\left(\frac{2\pi}{b}k\right) = \sum_{k=1}^{b}\frac{\sin\left(\frac{4\pi}{b}k\right)}{2} = 0, \tag{5}$$

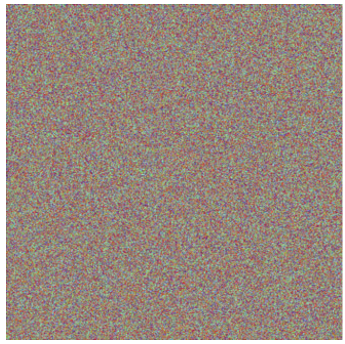
formula (4) reduces to

$$M = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{bmatrix}. \tag{6}$$

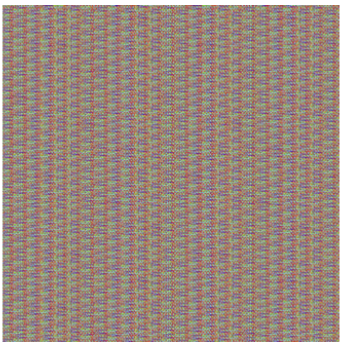
Hence, $S^N/\sqrt{N}$ is asymptotically bivariate normal with mean $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and covariance matrix M. Because its components $(S_1^N/\sqrt{N},\ S_2^N/\sqrt{N})^T$ are uncorrelated, then they are independent random variables, whose distribution is (univariate)



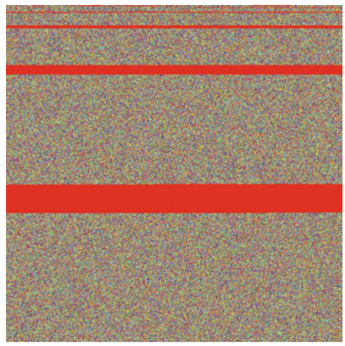
(a) π base 4



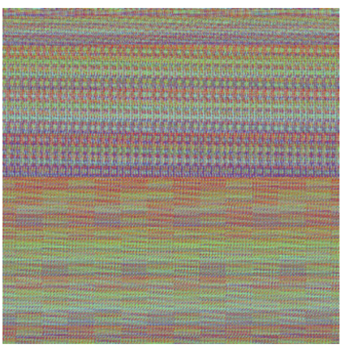
(b) (Pseudo)random number base 4



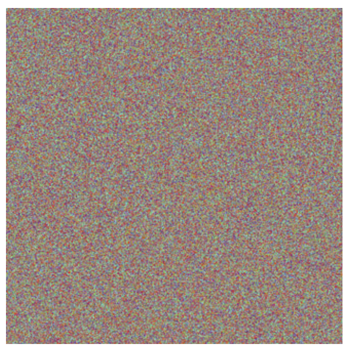
(c) The rational number $Q1$ base 4



(d) $\alpha_{2,3}$ base 6



(e) Champernowne's number C_4 base 4



(f) $EB(2)$ base 4

Figure 8. Horizontal color representation of a million digits of various numbers.

normal with mean 0 and variance 1/2. Therefore, the random variable

$$\sqrt{\left(\frac{\sqrt{2}}{\sqrt{N}}S_1^N\right)^2+\left(\frac{\sqrt{2}}{\sqrt{N}}S_2^N\right)^2} \tag{7}$$

converges in distribution to a χ random variable with two degrees of freedom. Then, for N sufficiently large,

$$E\left(\sqrt{(S_1^N)^2+(S_2^N)^2}\right)=\frac{\sqrt{N}}{\sqrt{2}}E\left(\sqrt{\left(\frac{\sqrt{2}}{\sqrt{N}}S_1^N\right)^2+\left(\frac{\sqrt{2}}{\sqrt{N}}S_2^N\right)^2}\right)$$
$$\approx\frac{\sqrt{N}}{\sqrt{2}}\frac{\Gamma(3/2)}{\Gamma(1)}=\frac{\sqrt{\pi N}}{2}, \tag{8}$$

where $E(\cdot)$ stands for the expectation of a random variable. Therefore, the expected distance to the

Table 1. Average of the normalized distance to the origin of the walk of various constants in different bases

Number	Base	Steps	Average normalized distance to the origin	Normal
Mean of 10,000 random walks	4	1,000,000	1.00315	Yes
Mean of 10,000 walks on the digits of π	4	1,000,000	1.00083	?
$\alpha_{2,3}$	3	1,000,000	0.89275	?
$\alpha_{2,3}$	4	1,000,000	0.25901	Yes
$\alpha_{2,3}$	5	1,000,000	0.88104	?
$\alpha_{2,3}$	6	1,000,000	108.02218	No
$\alpha_{4,3}$	3	1,000,000	1.07223	?
$\alpha_{4,3}$	4	1,000,000	0.24268	Yes
$\alpha_{4,3}$	6	1,000,000	94.54563	No
$\alpha_{4,3}$	12	1,000,000	371.24694	No
$\alpha_{3,5}$	3	1,000,000	0.32511	Yes
$\alpha_{3,5}$	5	1,000,000	0.85258	?
$\alpha_{3,5}$	15	1,000,000	370.93128	No
π	4	1,000,000	0.84366	?
π	6	1,000,000	0.96458	?
π	10	1,000,000	0.82167	?
π	10	10,000,000	0.56856	?
π	10	100,000,000	0.94725	?
π	10	1,000,000,000	0.59824	?
e	4	1,000,000	0.59583	?
$\sqrt{2}$	4	1,000,000	0.72260	?
log 2	4	1,000,000	1.21113	?
Champernowne C_{10}	10	1,000,000	59.91143	Yes
$EB(2)$	4	1,000,000	6.95831	?
$CE(10)$	4	1,000,000	0.94964	?
Rational number Q_1	4	1,000,000	0.04105	No
Rational number Q_2	4	1,000,000	58.40222	No
Euler constant γ	10	1,000,000	1.17216	?
Fibonacci $\mathcal{F}$	10	1,000,000	1.24820	?
$\zeta(2) = \frac{\pi^2}{6}$	4	1,000,000	1.57571	?
$\zeta(3)$	4	1,000,000	1.04085	?
Catalan's constant G	4	1,000,000	0.53489	?
Thue–Morse $\mathcal{TM}_2$	4	1,000,000	531.92344	No
Paper-folding $\mathcal{P}$	4	1,000,000	0.01336	No

origin of the random walk is asymptotically equal to $\sqrt{\pi N}/2$.

As a consequence of this result, for any random walk of N steps in any given base, the expectation of the distance to the origin multiplied by $2/\sqrt{\pi N}$ (which we will call *normalized distance* to the origin) must approach 1 as N goes to infinity. Therefore, for a "sufficiently" big random walk, one would expect that the arithmetic mean of the normalized distances

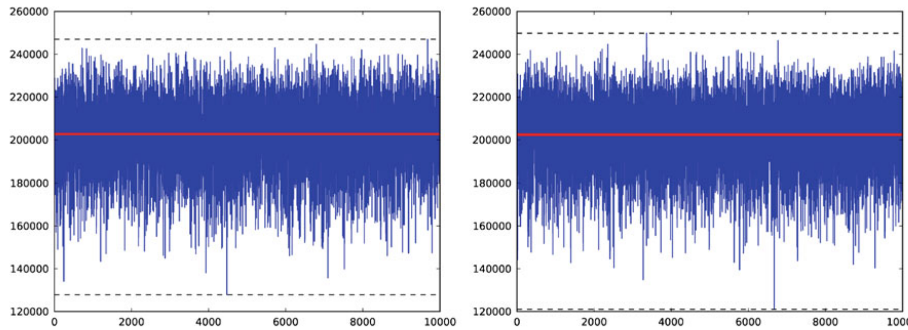
Table 2. Number of points visited in various *N*-step base-4 walks. The values of the two last columns are upper and lower bounds on the expectation of the number of distinct sites visited during an *N*-step random walk; see [31, Theorem 2] and [32]

Number	Steps	Sites visited	Bounds on the expectation of sites visited by a random walk	
			Lower bound	Upper bound
Mean of 10,000 random walks	1,000,000	202,684	199,256	203,060
Mean of 10,000 walks on the digits of π	1,000,000	202,385	199,256	203,060
$\alpha_{2,3}$	1,000,000	95,817	199,256	203,060
$\alpha_{4,3}$	1,000,000	68,613	199,256	203,060
$\alpha_{3,2}$	1,000,000	195,585	199,256	203,060
π	1,000,000	204,148	199,256	203,060
π	10,000,000	1,933,903	1,738,645	1,767,533
π	100,000,000	16,109,429	15,421,296	15,648,132
π	1,000,000,000	138,107,050	138,552,612	140,380,926
e	1,000,000	176,350	199,256	203,060
$\sqrt{2}$	1,000,000	200,733	199,256	203,060
log 2	1,000,000	214,508	199,256	203,060
Champernowne C_4	1,000,000	548,746	199,256	203,060
$EB(2)$	1,000,000	279,585	199,256	203,060
CE(10)	1,000,000	190,239	199,256	203,060
Rational number Q_1	1,000,000	378	199,256	203,060
Rational number Q_2	1,000,000	939,322	199,256	203,060
Euler constant γ	1,000,000	208,957	199,256	203,060
$\zeta(2)$	1,000,000	188,808	199,256	203,060
$\zeta(3)$	1,000,000	221,598	199,256	203,060
Catalan's constant G	1,000,000	195,853	199,256	203,060
$\mathcal{TM}_2$	1,000,000	1,000,000	199,256	203,060
Paper-folding $\mathcal{P}$	1,000,000	21	199,256	203,060

(which will be called *average normalized distance* to the origin) should be close to 1.

We have created a sample of 10,000 (pseudo)random walks base-4 of one million points each in Python[1], and we have computed their average normalized distance to the origin. The arithmetic mean of these numbers for the mentioned sample of pseudorandom walks is 1.0031, whereas its standard deviation is 0.3676, so the asymptotic result fits quite well. We have also computed the normalized distance to the origin of 10,000 walks of one million steps each generated by the first ten billion digits of π. The resulting arithmetic mean is 1.0008, whereas the standard deviation is 0.3682. In Table 1 we show the average normalized distance to the origin of various numbers. There are several surprises in there data, such as the

[1]Python uses the Mersenne Twister as the core generator and produces 53-bit precision floats, with a period of $2^{19937}-1 \approx 10^{6002}$. Compare the length of this period to the comoving distance from Earth to the edge of the observable universe in any direction, which is approximately $4.6 \cdot 10^{37}$ nanometers, or to the number of protons in the universe, which is approximately 10^{80}.



(a) (Pseudo)random walks.

(b) Walks based on the first 10 billion digits of π.

Figure 9. Number of points visited by 10^4 base-4 million-step walks.

fact that by this measure, Champernowne's number C_{10} is far from what is expected of a truly "random" number.

4 Number of Points Visited during an *N*-Step base-4 Walk

The number of distinct points visited during a walk of a given constant (on a lattice) can be also used as an indicator of how "random" the digits of that constant are. It is well known that the expectation of the number of distinct points visited by an N-step random walk on a two-dimensional lattice is asymptotically equal to $\pi N/\log(N)$; see, for example, [36, pg. 338] or [13, pg. 27]. This result was first proven by Dvoretzky and Erdős [33, Thm. 1]. The main practical problem with this asymptotic result is that its convergence is rather slow; specifically, it has order of $O\left(N \log\log N/(\log N)^2\right)$. In [31, 32], Downham and Fotopoulos show the following bounds on the expectation of the number of distinct points,

$$\left(\frac{\pi(N+0.84)}{1.16\pi - 1 - \log 2 + \log(N+2)}, \frac{\pi(N+1)}{1.066\pi - 1 - \log 2 + \log(N+1)}\right), \tag{9}$$

which provide a tighter estimate on the expectation than the asymptotic limit $\pi N/\log(N)$. For example, for $N = 10^6$, these bounds are (199256.1, 203059.5), whereas $\pi N/\log(N) = 227396$, which overestimates the expectation.

In Table 2 we have calculated the number of distinct points visited by the base-4 walks on several constants. One can see that the values for different step walks on π fit quite well the expectation. On the other hand, numbers that are known to be normal such as $\alpha_{2,3}$ or the base-4 Champernowne number substantially differ from the expectation of a random walk. These constants, despite being normal, do not have a "random" appearance when one draws the associated walk, see Figure 7(d).

At first look, the walk on $\alpha_{2,3}$ might seem random, see Figure 7(c). A closer look, shown in Figure 12, shows a more complex structure: the walk appears to be somehow self-repeating. This helps explain why the number of sites visited by the base-4 walk on $\alpha_{2,3}$ or $\alpha_{4,3}$ is smaller than the expectation for a random walk. A detailed discussion of the Stoneham constants and their walks is provided in Section 5.2, where the precise structure of Figure 12 is conjectured.

5 Copeland–Erdős, Stoneham, and Erdős–Borwein Constants

As well as the classical numbers—such as e, π, γ—listed in the Appendix, we also considered various other constructions, which we describe in the next three subsections.

5.1 Champernowne Number and Its Concatenated Relatives

The first mathematical constant proven to be 10-normal is the *Champernowne number*, which is defined as the concatenation of the decimal values of the positive integers, that is, $C_{10} = 0.12345678910111213141516\ldots$ Champernowne proved that C_{10} is 10-normal in 1933 [24]. This was later extended to base-b normality (for base-b versions of the Champernowne constant) as in Theorem 5.1. In 1946, Copeland and Erdős established that the corresponding concatenation of primes $0.23571113171923\ldots$ and the concatenation of composites $0.46891012141516\ldots$, among others, are also 10-normal [26]. In general they proved that concatenation leads to normality if the sequence grows slowly enough. We call such numbers *concatenation numbers*:

Theorem 5.1 ([26]). *If $a_1, a_2, \ldots$ is an increasing sequence of integers such that for every $\theta < 1$ the number of a_i's up to N exceeds N^θ provided N is sufficiently large, then the infinite decimal*

$$0.a_1a_2a_3\cdots$$

is normal with respect to the base b in which these integers are expressed.

This result clearly applies to the Champernowne numbers (Fig. 7(d)), to the primes of the form $ak + c$ with a and c relatively prime, in any given base, and to the integers that are the sum of two squares (since every prime of the form $4k + 1$ is

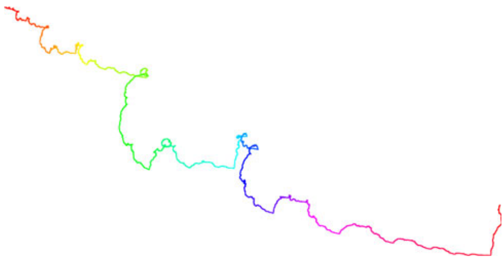


Figure 10. A walk on the first 100,000 bits of the primes ($CE(2)$) base two (normal).

included). In further illustration, using the primes in binary leads to normality in base 2 of the number

$$CE(2) = 0.1011101111110111110110001100111011111111011\,1111100101101001101011\ldots_2,$$

shown as a planar walk in Figure 10.

5.1.1 Strong Normality

In [14] it is shown that C_{10} fails the following stronger test of normality, which we now discuss. The test is is a simple one, in the spirit of Borel's test of normality, as opposed to other more statistical tests discussed in [14]. If the digits of a real number α are chosen at random in the base b, the *asymptotic frequency* $m_k(n)/n$ of each 1-string approaches $1/b$ with probability 1. However, the *discrepancy* $m_k(n) - n/b$ does not approach any limit, but fluctuates with an expected value equal to the standard deviation $\sqrt{(b-1)n}/b$. (Precisely $m_k(n) := \#\{i: a_i = k, i \le n\}$ when α has fractional part $0.a_0 a_1 a_2 \cdots$ in base b.)

Kolmogorov's law of the iterated logarithm allows one to make a precise statement about the discrepancy of a random number. Belshaw and P. Borwein [14] use this to define their criterion and then to show that almost every number is absolutely strongly normal.

Definition 5.2 (Strong normality [14]). For real α, and $m_k(n)$ as above, α is *simply strongly normal* in the base b if for each $0 \le k \le b-1$ one has

$$\limsup_{n\to\infty} \frac{m_k(n) - n/b}{\sqrt{2n \log\log n}} = \frac{\sqrt{b-1}}{b} \quad \text{and} \quad \liminf_{n\to\infty} \frac{m_k(n) - n/b}{\sqrt{2n \log\log n}} = -\frac{\sqrt{b-1}}{b}. \tag{10}$$

A number is *strongly normal* in base b if it is simply strongly normal in each base $b^j, j = 1, 2, 3, \ldots$, and is *absolutely strongly normal* if it is strongly normal in every base.

In paraphrase (absolutely) strongly normal numbers are those that distributionally oscillate as much as is possible.

Belshaw and Borwein show that strongly normal numbers are indeed normal. They also make the important observation that Champernowne's base-b number is not strongly normal in base b. Indeed, there are b^{v-1} digits of length v and they all start with a digit between 1 and $b-1$ whereas the following $v-1$ digits take values between 0 and $b-1$ equally. In consequence, there is a dearth of zeroes. This is easiest to analyze in base 2. As illustrated below, the concatenated numbers start

$$1, 10, 11, 100, 101, 110, 111, 1000, 1001, 1010, 1011, 1100, 1101, 1110, 1111$$

For $v = 3$ there are 4 zeroes and 8 ones, for $v = 4$ there are 12 zeroes and 20 ones, and for $v = 5$ there are 32 zeroes and 48 ones.

Because the details were not provided in [14], we present them here.

Theorem 5.3 (Belshaw and P. Borwein) *Champernowne's base-2 number is is not 2-strongly normal.*

Proof. In general, let $n_k := 1 + (k-1)2^k$ for $k \ge 1$. One has $m_0(n_k) = 1 + (k-1)2^k$ and so

$$m_1(n_k) - m_0(n_k) = n_k - 2m_0(n_k) = 2^k - 1.$$

In fact $m_1(n) > m_0(n)$ for all n. To see this, suppose it true for $n \le n_k$, and proceed by induction on k. Let us arrange the digits of the integers $2^k, 2^k + 1, \ldots, 2^k + 2^{k-1} - 1$ in a 2^{k-1} by $k+1$ matrix, where the i-th row contains the digits of the integer $2^k + i - 1$. Each row begins 10, and if we delete the first two columns we obtain a matrix in which the i-th row is given by the digits of $i - 1$, possibly preceded by some zeroes. Neglecting the first row and the initial zeroes in each subsequent row, we see the first n_{k-1} digits of Champernowne's base-2 number, where by our induction hypothesis $m_1(n) > m_0(n)$ for $n \le n_{k-1}$.

If we now count all the zeroes as we read the matrix in the natural order, any excess of zeroes must come from the initial zeroes, and there are exactly $2^{k-1} - 1$ of these. As shown above, $m_1(n_k) - m_0(n_k) = 2^k - 1$, so $m_1(n) > m_0(n) + 2^{k-1}$ for every $n \le n_k + (k+1)\, 2^{k-1}$. A similar argument for the integers from $2^k + 2^{k-1}$ to $2^{k+1} - 1$ shows that $m_1(n) > m_0(n)$ for every $n \le n_{k+1}$. Therefore, $2m_1(n) > m_0(n) + m_1(n) = n$ for all n, and so

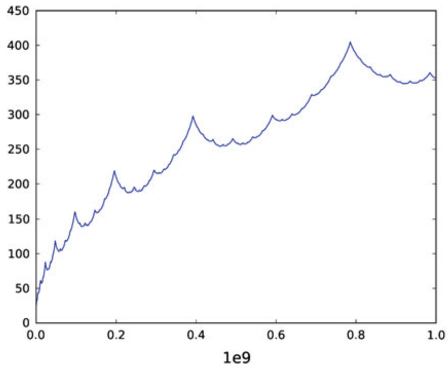
$$\liminf_{n\to\infty} \frac{m_1(n) - n/2}{\sqrt{2n \log\log n}} \ge 0 \ne -\frac{1}{2},$$

and, as asserted, Champernowne's base-2 number is not 2-strongly normal.

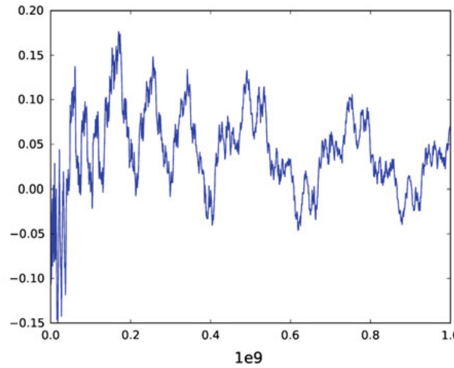
It seems likely that by appropriately shuffling the integers, one should be able to display a strongly normal variant. Along this line, Martin [40] has shown how to construct an explicit absolutely nonnormal number.

Finally, although the log log limiting behavior required by (10) appears difficult to test numerically to any significant level, it appears reasonably easy computationally to check whether other sequences, such as many of the concatenation sequences of Theorem 5.1, fail to be strongly normal for similar reasons.

Heuristically, we would expect the number $CE(2)$ above to fail to be strongly normal, because each prime of length k both starts and ends with a one, whereas intermediate bits should



(a) $CE(2)$ (not strongly 2-normal ?).



(b) $\alpha_{2,3}$ (strongly 2-normal ?).

Figure 11. Plot of the first 10^9 values of $\frac{m_1(n)-n/2}{\sqrt{2n\log\log n}}$.

show no skewing. Indeed, for $CE(2)$ we have checked that $2m_1(n) > n$ for all $n \le 10^9$, see also Figure 11(a). Thus motivated, we are currently developing tests for strong normality of numbers such as $CE(2)$ and $\alpha_{2,3}$ below in binary.

For $\alpha_{2,3}$, the corresponding computation of the first 10^9 values of $\frac{m_1(n)-n/2}{\sqrt{2n\log\log n}}$ leads to the plot in Figure 11(b) and leads us to conjecture that it is 2-strongly normal.

5.2 Stoneham Numbers: A Class Containing Provably Normal and Nonnormal Constants

Giving further motivation for these studies is the recent provision of rigorous proofs of normality for the *Stoneham numbers*, which are defined by

$$\alpha_{b,c} := \sum_{m\ge 1} \frac{1}{c^m b^{c^m}}, \tag{11}$$

for relatively prime integers b, c [10].

Theorem 5.4 (Normality of Stoneham constants [3]). *For every coprime pair of integers (b, c) with $b \ge 2$ and $c \ge 2$, the constant $\alpha_{b,c} = \sum_{m\ge 1} 1/(c^m\, b^{c^m})$ is b-normal.*

So, for example, the constant $\alpha_{2,3} = \sum_{k\ge 1} 1/(3^k 2^{3^k}) = 0.0418836808315030\ldots$ is provably 2-normal. This special case was proven by Stoneham in 1973 [43]. More recently, Bailey and Misiurewicz were able to establish this normality result by a much simpler argument, based on techniques of ergodic theory [11] [16, pg. 141–173].

Equally interesting is the following result:

Theorem 5.5 (Nonnormality of Stoneham constants [3]). *Given coprime integers $b \ge 2$ and $c \ge 2$, and integers p, q, $r \ge 1$, with neither b nor c dividing r, let $B = b^p c^q r$. Assume that the condition $D = c^{q/p} r^{1/p}/b^{c-1} < 1$ is satisfied. Then the constant $\alpha_{b,c} = \sum_{k\ge 0} 1/(c^k b^{c^k})$ is B-nonnormal.*

In various of the Figures and Tables, we explore the striking differences of behavior—proven and unproven—for $\alpha_{b,c}$ as we vary the base. For instance, the nonnormality of $\alpha_{2,3}$ in base-6 digits was proved just before we started to draw walks. Contrast Figure 7(b) to Figure 7(c) and Figure 7(a). Now compare the values presented in Table 1 and Table 2. Clearly, from this sort of visual and numeric data, the discovery of other cases of Theorem 5.5 is very easy.

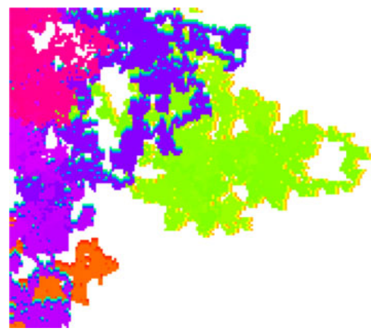


Figure 12. Zooming in on the base-4 walk on $\alpha_{2,3}$ of Figure 7(c) and Conjecture 5.6.

As illustrated also in the "zoom" of Figure 12, we can use these images to discover more subtle structure. We conjecture the following relations on the digits of $\alpha_{2,3}$ in base 4 (which explain the values in Tables 1 and 2):

Conjecture 5.6 (Base-4 structure of $\alpha_{2,3}$). *Denote by a_k the k^{th} digit of $\alpha_{2,3}$ in its base-4 expansion; that is, $\alpha_{2,3} = \sum_{k=1}^{\infty} a_k/4^k$, with $a_k \in \{0,1,2,3\}$ for all k. Then, for all $n = 0, 1, 2, \ldots$ one has:*

(i) $\displaystyle\sum_{k=\frac{3}{2}(3^n+1)}^{\frac{3}{2}(3^n+1)+3^n} e^{a_k \pi i/2} = \frac{(-1)^{n+1}-1}{2} + \frac{(-1)^n - 1}{2} i = -\begin{cases} i, & n \text{ odd} \\ 1, & n \text{ even}\end{cases};$

(ii) $a_k = a_{k+3^n} = a_{k+2\cdot 3^n}$ *for all* $k = \frac{3}{2}(3^n+1), \frac{3}{2}(3^n+1)+1, \ldots, \frac{3}{2}(3^n+1)+3^n-1$.

In Figure 13, we show the position of the walk after $\frac{3}{2}(3^n+1), \frac{3}{2}(3^n+1)+3^n$ and $\frac{3}{2}(3^n+1)+2\cdot 3^n$ steps for

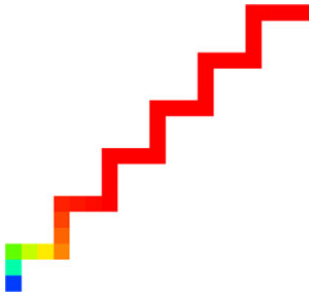


Figure 13. A pattern in the digits of $\alpha_{2,3}$ base 4. We show only positions of the walk after $\frac{3}{2}(3^n+1), \frac{3}{2}(3^n+1)+3^n$ and $\frac{3}{2}(3^n+1)+2\cdot 3^n$ steps for $n=0,1,\ldots,11$.

$n=0,1,\ldots,11$, which, together with Figures 7(c) and 12, graphically explain Conjecture 5.6. Similar results seem to hold for other Stoneham constants in other bases. For instance, for $\alpha_{3,5}$ base 3 we conjecture the following.

Conjecture 5.7 (Base-3 structure of $\alpha_{3,5}$). *Denote by a_k the k^{th} digit of $\alpha_{3,5}$ in its base-3 expansion; that is, $\alpha_{3,5} = \sum_{k=1}^{\infty} a_k/3^k$, with $a_k \in \{0,1,2\}$ for all k. Then, for all $n=0,1,2,\ldots$ one has:*

(i) $\displaystyle\sum_{k=2+5^{n+1}}^{2+5^{n+1}+4\cdot 5^n} e^{a_k\pi i/2} = (-1)^n\left(\frac{-1+\sqrt{3}i}{2}\right) = e^{(3n+2)\pi i/3}$;

(ii) $a_k = a_{k+4\cdot 5^n} = a_{k+8\cdot 5^n} = a_{k+12\cdot 5^n} = a_{k+16\cdot 5^n}$ *for* $k = 5^{n+1}+j, j=2,\ldots,2+4\cdot 5^n$.

Along this line, Bailey and Crandall showed that, given a real number r in [0,1), and r_k denoting the k-th binary digit of r, the real number

$$\alpha_{2,3}(r) := \sum_{k=0}^{\infty} \frac{1}{3^k 2^{3^k+r_k}} \tag{12}$$

is 2-normal. It can be seen that if $r \neq s$, then $\alpha_{2,3}(r) \neq \alpha_{2,3}(s)$, so that these constants are all distinct. Thus, this generalized class of Stoneham constants is uncountably infinite. A similar result applies if 2 and 3 in this formula are replaced by any pair of co-prime integers (b, c) greater than 1, [10] [16, pg. 141–173]. We have not yet studied this generalized class by graphical methods.

5.3 The Erdős–Borwein Constants

The constructions of the previous two subsections exhaust most of what is known for concrete irrational numbers. By contrast, we finish this section with a truly tantalizing case:

In a base $b \geq 2$, we define the *Erdős–(Peter) Borwein constant* $EB(b)$ by the *Lambert series* [18]:

$$EB(b) := \sum_{n\geq 1} \frac{1}{b^n-1} = \sum_{n\geq 1} \frac{\tau(n)}{b^n}, \tag{13}$$

where $\tau(n)$ is the number of divisors of n. It is known that the numbers $\sum_{n\geq 1} 1/(b^n - r)$ are irrational for r a nonzero rational and $b=2,3,\ldots$ such that $r \neq b^n$ for all n [20]. Whence, as provably irrational numbers other than the standard examples are few and far between, it is interesting to consider their normality.

Crandall [27] has observed that the structure of (13) is analogous to the "BBP" formula for π (see [7, 16]) and used this, as well as some nontrivial knowledge of the arithmetic properties of τ, to establish results such as that the googol-th bit (namely, the bit in position 10^{100} to the right of the "decimal" point) of $EB(2)$ is a 1.

In [27] Crandall also computed the first 2^{43} bits (one Tbyte) of $EB(2)$, which required roughly 24 hours of computation, and found that there are **435**9105565638 zeroes and **443**6987456570 ones. There is a corresponding variation in the second and third place in the single-digit hex (base-16) distributions. This certainly leaves some doubt as to its normality. Likewise, Crandall finds that in the first 1,000 decimal positions after the quintillionth digit (10^{18}), the respective digit counts for digits 0, 1, 2, 3, 4, 5, 6, 7, 8, 9 are 104, 82, 87, 100, 73, 126, 87, 123, 114, 104. Our own more modest computations of $EB(10)$ base-10 again leave it far from clear that $EB(10)$ is 10-normal. See also Figure 7(e) but contrast it to Figure 8(f).

We should note that for computational purposes, we employed the identity

$$\sum_{n\geq 1} \frac{1}{b^n-1} = \sum_{n\geq 1} \frac{b^n+1}{b^n-1}\frac{1}{b^{n^2}},$$

for $|b|>1$, due to Clausen, as did Crandall [27].

6 Other Avenues and Concluding Remarks

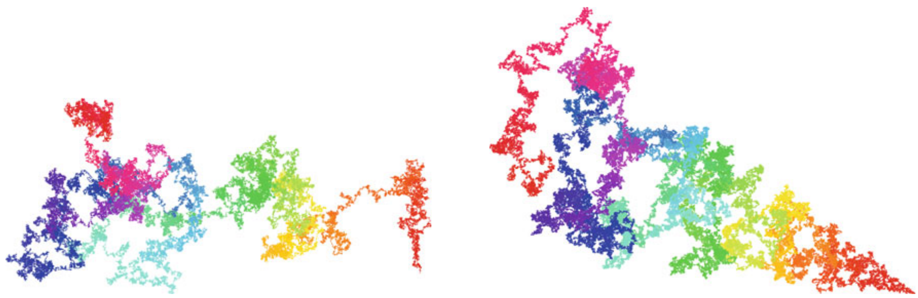
Let us recall two further examples used in [14], that of $\Omega(n)$, the *Liouville function*, which counts the parity of the number of prime factors of n (see Figure 14), and the human genome taken from the *UCSC Genome Browser* at http://hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/ (see Fig. 15). Note the similarity of the genome walk to the those of concatenation sequences. We have explored a wide variety of walks on genomes, but we will reserve the results for a future study.

We should emphasize that, to the best of our knowledge, the normality and transcendence status of the numbers explored is unresolved other than in the cases indicated in sections 5.1 and 5.2 and indicated in Appendix 7. Although one of the clearly nonrandom numbers (say Stoneham or Copeland–Erdős) may pass muster on one or other measure of the walk, it is generally the case that it fails another. Thus, the Liouville number λ_2 (see Fig. 14) exhibits a much more structured drift than π or e, but looks more like them than like Figure 15(a).

This situation provides hope for more precise future analyses. We conclude by remarking on some unresolved issues and plans for future research.

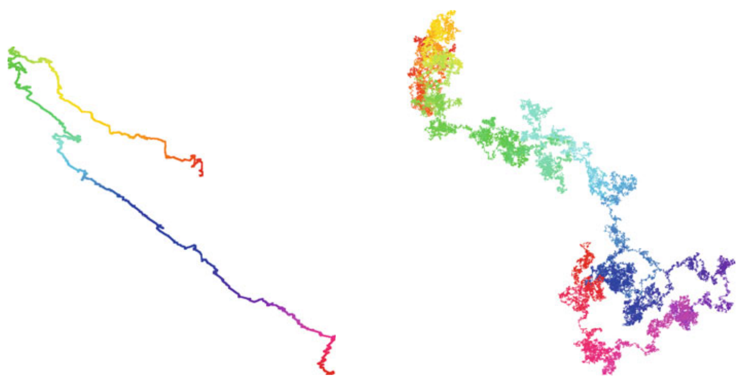
6.1 Fractal and Box-Dimension

Another approach is to estimate the fractal dimensions of walks, which is an appropriate tool with which to measure the geometrical complexity of a set, characterizing its space-filling capacity (see, e.g., [6] for a nice introduction about fractals). The *box-counting dimension*, also known as the *Minkowski–*



(a) Directions used: →, ↑, ←, ↓. **(b)** Directions used: ↗, ↘, ↖, ↙.

Figure 14. Two different rules for plotting a base-2 walk on the first two million values of $\lambda(n)$ (the Liouville number λ_2).



(a) Human X. **(b)** log 2.

Figure 15. Base-4 walks on 10^6 bases of the X-chromosome and 10^6 digits of log 2.

Bouligand dimension, permits us to estimate the fractal dimension of a given set and often coincides with the fractal dimension. The box-dimension of the walk of numbers such as π turns out to be close to 2, whereas for nonrandom numbers as $\alpha_{2,3}$ in base 6 or Champernowne's number, the box-dimension is nearly 1.

6.2 Three Dimensions

We have also explored three-dimensional graphics—using base-6 for directions—both in perspective and in a large passive (glasses-free) three-dimensional viewer outside the CARMA laboratory; but we have not yet quantified these excursions.

6.3 Genome Comparison

Genomes are made up of so-called purine and pyrimidine nucleotides. In DNA, purine nucleotide bases are adenine and guanine (A and G), whereas the pyrimidine bases are thymine and cytosine (T and C). Thymine is replaced by uracyl in RNA. The haploid human genome (i.e., 23 chromosomes) is estimated to hold about 3.2 billion base pairs and so to contain 20,000-25,000 distinct genes. Hence there are many ways of representing a stretch of a chromosome as a walk, say as a base-4 uniform walk on the symbols (A, G, T, C) illustrated in Figure 15 (where A, G, T, and C draw the new point to the south, north, west, and east, respectively, and we have not plotted undecoded or unused portions), or as a three-dimensional logarithmic walk inside a tetrahedron.

We have also compared random *chaos games* in a square with genomes and numbers plotted by the same rules.[2] As an illustration, we show twelve games in Figure 16: four on a triangle, four on a square, and four on a hexagon. At each step we go from the current point halfway toward one of the vertices, chosen depending on the value of the digit. The color indicates the number of hits, in a similar manner as in Figure 6. The nonrandom behavior of the Champernowne numbers is apparent in the coloring patterns, as are the special features of the Stoneham numbers described in Section 5.2 (the nonnormality of $\alpha_{2,3}$ and $\alpha_{3,2}$ in base 6 yields a paler color, whereas the repeating structure of $\alpha_{2,3}$ and $\alpha_{3,5}$ is the origin of the purple tone, see Conjectures 5.6 and 5.7).

[2]The idea of a chaos game was described by Barnsley in his 1988 book *Fractals Everywhere* [6]. Games on amino acids seem to originate with [35]. For a recent summary see [17, pp. 194–205].

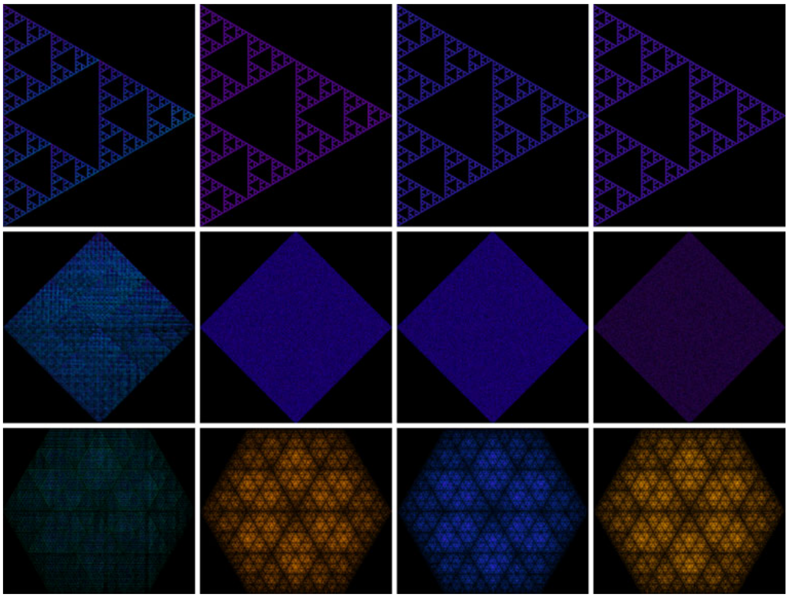
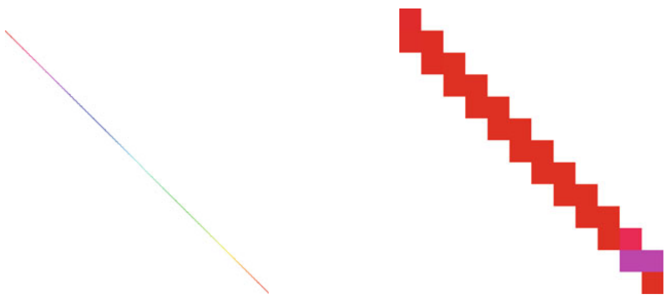


Figure 16. Chaos games on various numbers, colored by frequency. Row 1: C_3, $\alpha_{3,5}$, a (pseudo)random number, and $\alpha_{2,3}$. Row 2: C_4, π, a (pseudo)random number, and $\alpha_{2,3}$. Row 3: C_6, $\alpha_{3,2}$, a (pseudo)random number, and $\alpha_{2,3}$.



(a) A thousand digits of the Thue–Morse sequence $\mathcal{TM}_2$ base 2.

(b) Ten million digits of the paper-folding sequence base 2.

Figure 17. Walks on two automatic and nonnormal numbers.

6.4 Automatic Numbers

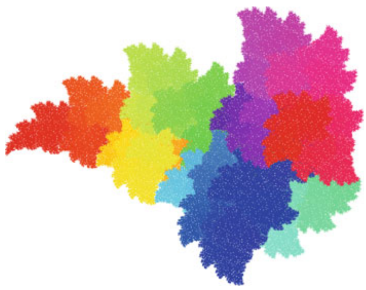
We have also explored numbers originating with finite state automata, such as those of the *paper-folding* and the *Thue–Morse* sequences, $\mathcal{P}$ and $\mathcal{TM}_2$, see [2] and Section 7. Automatic numbers are never normal and are typically transcendental; by comparison, the Liouville number λ_2 is not p-automatic for any prime p [25].

The walks on $\mathcal{P}$ and $\mathcal{TM}_2$ have a similar shape, see Figure 17, but while the Thue–Morse sequence generates very large pictures, the paper-folding sequence generates very small ones, because it is highly self-replicating; see also the values in Tables 1 and 2.

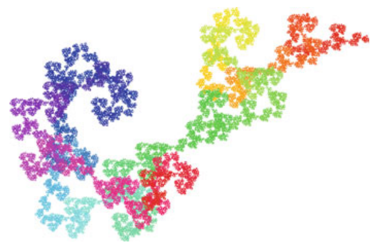
A *turtle plot* on these constants, where each binary digit corresponds to either "forward motion" of length 1 or "rotate the Logo turtle" in a fixed angle, exhibits some of their striking features (see Fig. 18). For instance, drawn with a rotating angle of $\pi/3, \mathcal{TM}_2$ converges to a Koch snowflake [41]; see Figure 18(c). We show a corresponding turtle graphic of π in Figure 18(d). Analogous features occur for the paper-folding sequence as described in [28, 29, 30], and two variants are shown in Figures 18(a) and 18(b).

6.5 Continued Fractions

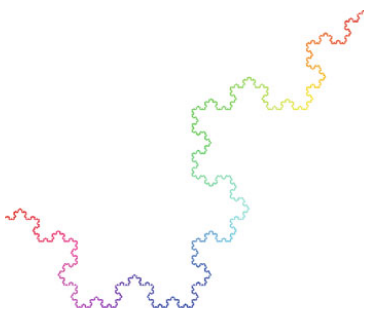
Simple continued fractions often encode more information than base expansions about a real number. Basic facts are that a continued fraction terminates or repeats if and only if the



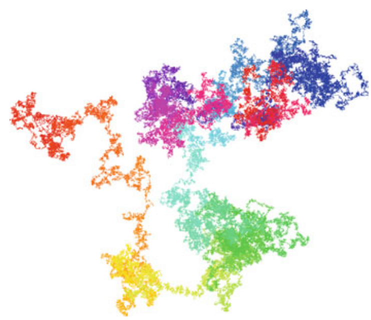
(a) Ten million digits of the paper-folding sequence with rotating angle $\pi/3$.



(b) Dragon curve from one million digits of the paper-folding sequence with rotating angle $2\pi/3$.



(c) Koch snowflake from 100,000 digits of the Thue–Morse sequence $\mathcal{TM}_2$ with rotating angle $\pi/3$.



(d) One million digits of π with rotating angle $\pi/3$.

Figure 18. Turtle plots on various constants with different rotating angles in base 2—where "0" gives forward motion and "1" rotation by a fixed angle.

number is rational or a quadratic irrational, respectively; see [16, 7]. By contrast, the simple continued fractions for π and e start as follows in the standard compact form:

$$\pi = [3, 7, 15, 1, 292, 1, 1, 1, 2, 1, 3, 1, 14, 2, 1, 1, 2, 2, 2, 2, 1, 84, 2, 1, 1, 15, 3, 13, 1, 4, \ldots]$$

$$e = [2, 1, 2, 1, 1, 4, 1, 1, 6, 1, 1, 8, 1, 1, 10, 1, 1, 12, 1, 1, 14, 1, 1, 16, 1, 1, 18, 1, 1, 20, 1, \ldots],$$

from which the surprising regularity of e and apparent irregularity of π as continued fractions is apparent. The counterpart to Borel's theorem—that almost all numbers are normal—is that almost all numbers have "normal" continued fractions $\alpha = [a_1, a_2, \ldots, a_n, \ldots]$, for which the *Gauss–Kuzmin distribution* holds [16]: for each $k = 1, 2, 3, \ldots$

$$\text{Prob}\{a_n = k\} = -\log_2\left(1 - \frac{1}{(k+1)^2}\right), \qquad (14)$$

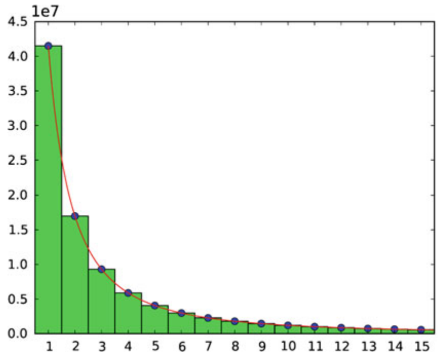
so that roughly 41.5% of the terms are 1, 16.99% are 2, 9.31% are 3, etc.

In Figure 19, we show a histogram of the first 100 million terms, computed by Neil Bickford and accessible at http://neilbickford.com/picf.htm, of the continued fraction of π. We have not yet found a satisfactory way to embed this in a walk on a continued fraction, but in Figure 20 we show base-4 walks on π and e where we use the remainder modulo 4 to build the walk (with 0 being right, 1 being up, 2 being left, and 3 being down). We also show turtle plots on π, e.

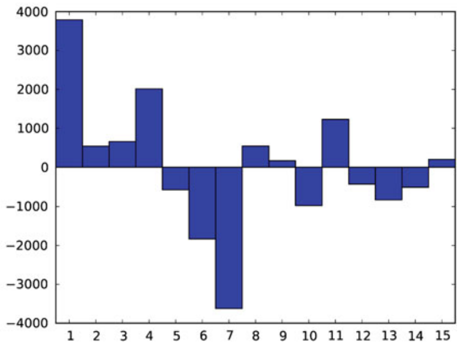
Andrew Mattingly has observed that:

Proposition 6.1 *With probability* 1, *a mod-*4 *random walk (with* 0 *being right,* 1 *being up,* 2 *being left, and* 3 *being down) on the simple continued fraction coefficients of a real number is asymptotic to a line making a positive angle with the x-axis of:*

$$\arctan\left(\frac{1}{2}\frac{\log_2(\pi/2) - 1}{\log_2(\pi/2) - 2\log_2(\Gamma(3/4))}\right) \approx 110.44°.$$

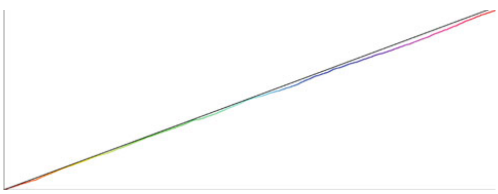


(a) Histogram of the terms in green, Gauss–Kuzmin function in red.



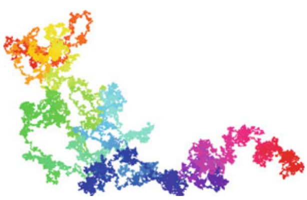
(b) Difference between the expected and computed values of the Gauss–Kuzmin function.

Figure 19. Expected values of the Gauss–Kuzmin distribution of (14) and the values of 100 million terms of the continued fraction of π.



(a) A 100,000-step walk on the continued fraction of π modulo 4.

(b) A 100-step walk on the continued fraction of e modulo 4.



(c) A one million-step turtle walk on the continued fraction of π modulo 2 with rotating angle $\pi/3$.

(d) A 100-step turtle walk on the continued fraction of e modulo 2 with rotating angle $\pi/3$.

Figure 20. Uniform walks on π and e based on continued fractions.

Proof. The result comes by summing the expected Gauss–Kuzmin probabilities of each step being taken as given by (14).

This is illustrated in Figure 20(a) with a 90° anticlockwise rotation; thus making the case that one must have some *a priori* knowledge before designing tools.

It is also instructive to compare digits and continued fractions of numbers as horizontal matrix plots of the form already used in Figure 8. In Figure 21, we show six pairs of million-term digit-strings and the corresponding continued fraction. In some cases both look random, in others one or the other does.

In conclusion, we have only scratched the surface of what is becoming possible in a period in which data—for example, five-hundred million terms of the continued fraction or five-trillion binary digits of π, full genomes, and much more—can be downloaded from the Internet, then rendered and visually mined, with fair rapidity.

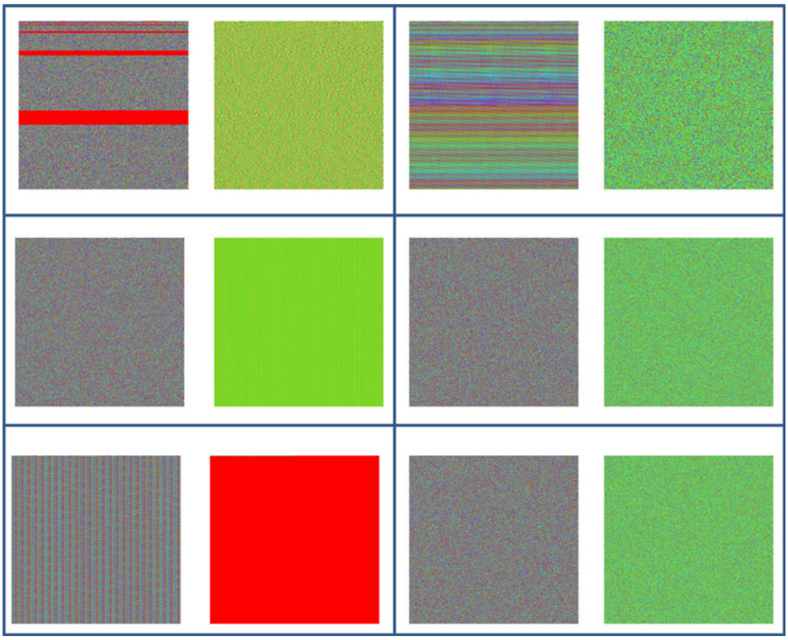


Figure 21. Million-step comparisons of base-4 digits and continued fractions. Row 1: $\alpha_{2,3}$ (*base* 6) and C_4. Row 2: e and π. Row 3: Q_1 and pseudorandom iterates; as listed from top left to bottom right.

7 Appendix Selected Numerical Constants

In what follows, $x := 0.a_1a_2a_3a_4\ldots_b$ denotes the base-b expansion of the number x, so that $x = \sum_{k=1}^{\infty} a_k b^{-k}$. Base-10 expansions are denoted without a subscript.

Catalan's constant (irrational?; normal?):

$$G := \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^2} = 0.9159655941\ldots \quad (15)$$

Champernowne numbers (irrational; normal to corresponding base):

$$C_b := \sum_{k=1}^{\infty} \frac{\sum_{m=b^{k-1}}^{b^k-1} m b^{-k\left[m-(b^{k-1}-1)\right]}}{b\sum_{m=0}^{k-1} m(b-1)b^{m-1}} \quad (16)$$

$$C_{10} = 0.123456789101112\ldots$$
$$C_4 = 0.1231011121320212223\ldots_4$$

Copeland–Erdős constants (irrational; normal to corresponding base):

$$CE(b) := \sum_{k=1}^{\infty} p_k b^{-\left(k+\sum_{m=1}^{k}\lfloor \log_b p_m \rfloor\right)}, \quad (17)$$

where p_k is the k^{th}prime number

$$CE(10) = 0.2357111317\ldots$$
$$CE(2) = 0.1011101111\ldots_2$$

Exponential constant (transcendental; normal?):

$$e := \sum_{k=0}^{\infty} \frac{1}{k!} = 2.7182818284\ldots \quad (18)$$

Erdős–Borwein constants (irrational; normal?):

$$EB(b) := \sum_{k=1}^{\infty} \frac{1}{b^k - 1} \quad (19)$$

$$EB(2) = 1.6066951524\ldots = 1.212311001\ldots_4$$

Euler–Mascheroni constant (irrational?; normal?):

$$\gamma := \lim_{m\to\infty}\left(\sum_{k=1}^{m}\frac{1}{k} - \log m\right) = 0.5772156649\ldots \quad (20)$$

Fibonacci constant (irrational [12, Theorem 2]; normal?):

$$\mathcal{F} := \sum_{k=1}^{\infty} F_k 10^{-\left(1+k+\sum_{m=1}^{k}\lfloor \log_{10} F_m \rfloor\right)}, \text{ where}$$

$$F_k = \frac{\left(\frac{1+\sqrt{5}}{2}\right)^k - \left(\frac{1-\sqrt{5}}{2}\right)^k}{\sqrt{5}}$$

$$= 0.0112358132134 55\ldots$$

Liouville number (irrational; not p-automatic):

$$\lambda_2 := \sum_{k=1}^{\infty}\left(\frac{\lambda(k)+1}{2}\right)2^{-k} \quad (22)$$

where $\lambda(k) := (-1)^{\Omega(k)}$ and $\Omega(k)$ counts prime factors of k

$$= 0.5811623188\ldots = 0.10010100110\ldots_2$$

Logarithmic constant (transcendental; normal?):

$$\log 2 := \sum_{k=1}^{\infty} \frac{1}{k2^k} \tag{23}$$

$$= 0.6931471806\ldots = 0.10110001011100100001\ldots_2$$

Pi (transcendental; normal?):

$$\pi := 2\int_{-1}^{1} \sqrt{1-x^2}\mathrm{d}x = 4\sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \tag{24}$$

$$= 3.1415926535\ldots = 11.00100100001111110110\ldots_2$$

Riemann zeta function at integer arguments (transcendental for n even; irrational for $n = 3$; unknown for $n \geq 5$ odd; normal?):

$$\zeta(s) := \sum_{k=1}^{\infty} \frac{1}{k^s} \tag{25}$$

In particular:

$$\zeta(2) = \frac{\pi^2}{6} = 1.6449340668\ldots$$

$$\zeta(2n) = (-1)^{n+1}\frac{(2\pi)^{2n}}{2(2n)!}B_{2n}$$

(where B_{2n} are Bernoulli numbers)

$$\zeta(3) = \text{Apery's constant} = \frac{5}{2}\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^3\binom{2k}{k}}$$

$$= 1.2020569031\ldots$$

Stoneham constants (irrational; normal in some bases; nonnormal in different bases; normality still is unknown other bases):

$$\alpha_{b,c} := \sum_{k=1}^{\infty} \frac{1}{b^{c^k} c^k} \tag{26}$$

$$\alpha_{2,3} = 0.0418836808\ldots = 0.0022232032\ldots_4$$
$$= 0.0130140430003334\ldots_6$$
$$\alpha_{4,3} = 0.0052087571\ldots = 0.0001111111301\ldots_4$$
$$= 0.0010430041343502130000\ldots_6$$
$$\alpha_{3,2} = 0.0586610287\ldots = 0.0011202021212121\ldots_3$$
$$= 0.0204005200030544000002\ldots_6$$
$$\alpha_{3,5} = 0.0008230452\ldots = 0.00000012101210121\ldots_3$$
$$= 0.002ba00000061d2\ldots_{15}$$

Thue–Morse constant (transcendental; 2-automatic, hence nonnormal):

$$\mathcal{TM}_2 := \sum_{k=1}^{\infty} \frac{1}{2^{t(n)}} \text{ where } t(0) = 0, \text{ while } t(2n) = t(n)$$
$$\text{and } t(2n+1) = 1 - t(n) \tag{27}$$

$$= 0.4124540336\ldots$$
$$= 0.01101001100101101001011001101001\ldots_2$$

Paper-folding constant (transcendental; 2-automatic, hence nonnormal):

$$\mathcal{P} := \sum_{k=0}^{\infty} \frac{8^{2^k}}{2^{2^{k+2}} - 1} = 0.8507361882\ldots$$
$$= 0.1101100111001001\ldots_2 \tag{28}$$

ACKNOWLEDGMENTS

The authors thank the referee for a thoughtful and thorough report, David Allingham for his kind help in running the algorithms for plotting the "big walks" on π, Adrian Belshaw for his assistance with strong normality, Matt Skerritt for his 3D image of π, and Jake Fountain who produced a fine Python interface for us as a 2011-2012 Vacation Scholar at CARMA. We are also most grateful for several discussions with Andrew Mattingly (IBM) and Michael Coons (CARMA), who helped especially with continued fractions and automatic numbers, respectively.

REFERENCES

[1] S. Albeverioa, M. Pratsiovytyie, and G. Torbine G, ''Topological and fractal properties of real numbers which are not normal''. *Bulletin des Sciences Mathématiques*, **129** (2005), 615–630.

[2] J.-P. Allouche and J. Shallit, *Automatic Sequences: Theory, Applications, Generalizations*. Cambridge University Press, Cambridge, 2003.

[3] D. H. Bailey and J. M. Borwein, ''Normal numbers and pseudorandom generators,'' *Proceedings of the Workshop on Computational and Analytical Mathematics in Honour of Jonathan Borwein's 60th Birthday*, Springer, 2012, in press.

[4] D. H. Bailey, J. M. Borwein, C. S. Calude, M. J. Dinneen, M. Dumitrescu, and A. Yee, ''An empirical approach to the normality of pi''. *Experimental Mathematics*, 2012; in press.

[5] D. H. Bailey, J. M. Borwein, R. E. Crandall, and C. Pomerance. ''On the binary expansions of algebraic numbers''. *Journal of Number Theory Bordeaux*, **16** (2004), 487–518.

[6] M. Barnsley, *Fractals Everywhere*, Academic Press, Inc., Boston, MA, 1988.

[7] D. H. Bailey, P. B. Borwein, and S. Plouffe, ''On the rapid computation of various polylogarithmic constants''. *Mathematics of Computation*, **66**, no. 218 (1997), 903–913.

[8] D. H. Bailey and D. J. Broadhurst, ''Parallel integer relation detection: Techniques and applications''. *Mathematics of Computation*, **70**, no. 236 (2000), 1719–1736.

[9] D. H. Bailey and R. E. Crandall, ''On the random character of fundamental constant expansions''. *Experimental Mathematics*, **10**, no. 2 (2001), 175–190.

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

[11] D. H. Bailey and M. Misiurewicz, ''A strong hot spot theorem,'' *Proceedings of the American Mathematical Society*, **134** (2006), no. 9, 2495–2501.

[12] G. Barat, R. F. Tichy, and R. Tijdeman, Digital blocks in linear numeration systems. *Number theory in progress*, **2** (Zakopane-Kościelisko, 1997), de Gruyter, Berlin (1999), 607–631.

[13] M. N. Barber and B. W. Ninham, *Random and Restricted Walks: Theory and Applications*, Gordon and Breach, New York, 1970.

[14] A. Belshaw and P. B. Borwein, ''Champernowne's number, strong normality, and the X chromosome,'' *Proceedings of the Workshop on Computational and Analytical Mathematics in Honour of Jonathan Borwein's 60th Birthday*, Springer, 2012, in press.

[15] L. Berggren, J. M. Borwein, and P. B. Borwein, *Pi: a Source Book*, Springer-Verlag, Third Edition, 2004.

[16] J. M. Borwein and D. H. Bailey, *Mathematics by Experiment: Plausible Reasoning in the 21st Century*, 2nd ed., A. K. Peters, Natick, MA, 2008.

[17] J. Borwein, D. Bailey, N. Calkin, R. Girgensohn, R. Luke, V. Moll, *Experimental Mathematics in Action*. A. K. Peters, Natick, MA, 2007.

[18] J. M. Borwein and P. B. Borwein, *Pi and the AGM: A Study in Analytic Number Theory and Computational Complexity*, John Wiley, New York, 1987, paperback 1998.

[19] J. M. Borwein, P. B. Borwein, R. M. Corless, L. Jörgenson, and N. Sinclair, ''What is organic mathematics?'' *Organic mathematics* (Burnaby, BC, 1995), CMS Conf. Proc., **20**, Amer. Math. Soc., Providence, RI, 1997, 1–18.

[20] P. B. Borwein, ''On the irrationality of certain series.'' *Math. Proc. Cambridge Philos. Soc.* **112** (1992) 141–146.

[21] P. B. Borwein and L. Jörgenson, '' Visible structures in number theory,'' *Amer. Math. Monthly* **108** (2001), no. 10, 897–910.

[22] C. S. Calude, ''Borel normality and algorithmic randomness,'' in G. Rozenberg, A. Salomaa (eds.), *Developments in Language Theory*, World Scientific, Singapore, 1994, 113–129.

[23] C.S. Calude, *Information and Randomness: An Algorithmic Perspective*, 2nd ed., Revised and Extended, Springer-Verlag, Berlin, 2002.

[24] D. G. Champernowne, ''The construction of decimals normal in the scale of ten.'' *Journal of the London Mathematical Society*, **8** (1933) 254–260.

[25] M. Coons, ''(Non)automaticity of number theoretic functions,'' *J. Théor. Nombres Bordeaux*, **22** (2010), no. (2), 339–352.

[26] A. H. Copeland and P. Erdős, ''Note on normal numbers,'' *Bulletin of the American Mathematical Society*, **52** (1946), 857–860.

[27] R. E. Crandall, ''The googol-th bit of the Erdős–Borwein constant,'' *Integers*, A23, 2012.

[28] M. Dekking, M. Mendès France, and A. van der Poorten, ''Folds,'' *Math. Intelligencer* **4** (1982), no. 3, 130–138.

[29] M. Dekking, M. Mendès France, and A. van der Poorten, ''Folds II,'' *Math. Intelligencer* **4** (1982), no. 4, 173–181.

[30] M. Dekking, M. Mendès France, and A. van der Poorten, ''Folds III,'' *Math. Intelligencer* **4**(1982), no. (4), 190–195.

[31] D. Y. Downham and S. B. Fotopoulos, ''The transient behaviour of the simple random walk in the plane,'' *J. Appl. Probab.* **25** (1988), no. 1, 58–69.

[32] D. Y. Downham and S. B. Fotopoulos, ''A note on the simple random walk in the plane,'' *Statist. Probab. Lett.*, 17 (1993), no. 3, 221–224.

[33] A. Dvoretzky and P. Erdős, ''Some problems on random walk in space,'' *Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability*, (1951), 353–367.

[34] P. Hertling, ''Simply normal numbers to different bases,'' *Journal of Universal Computer Science*, **8**, no. 2 (2002), 235–242.

[35] H. J. Jeffrey, Chaos game representation of gene structure, *Nucl. Acids Res.* **18** no 2, (1990) 2163–2170.

[36] B. D. Hughes, *Random Walks and Random Environments*, *Vol. 1. Random Walks*, Oxford Science Publications, New York, (1995).

[37] H. Kaneko, ''On normal numbers and powers of algebraic numbers,'' *Integers*, **10** (2010), 31–64.

[38] D. Khoshnevisan, ''Normal numbers are normal,'' *Clay Mathematics Institute Annual Report* (2006), 15 & 27–31.

[39] G. Marsaglia, ''On the randomness of pi and other decimal expansions,'' preprint, 2010.

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

[41] J. Mah and J. Holdener, ''When Thue–Morse meets Koch,'' *Fractals*, **13** (2005), no. 3, 191–206.

[42] S. M. Ross, *Stochastic Processes*. John Wiley & Sons, New York, 1983.

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

[44] M. Queffelec, ''Old and new results on normality,'' *Lecture Notes – Monograph Series*, **48**, *Dynamics and Stochastics*, 2006, Institute of Mathematical Statistics, 225–236.

[45] W. Schmidt, ''On normal numbers,'' *Pacific Journal of Mathematics*, **10** (1960), 661–672.

[46] A. J. Yee, ''y-cruncher-multi-threaded pi program,'' http://www.numberworld.org/y-cruncher, 2010.

[47] A. J. Yee and S. Kondo, ''10 trillion digits of pi: A case study of summing hypergeometric series to high precision on multicore systems,'' preprint, 2011, available at http://hdl.handle.net/2142/28348.