

# Walking on real numbers

Francisco J. Aragón Artacho\*    David H. Bailey†    Jonathan M. Borwein‡  
Peter B. Borwein §

July 23, 2012

## Abstract

Motivated by the desire to visualize large mathematical data sets, especially in number theory, we offer various tools for representing floating point numbers as planar (or three dimensional) walks and for quantitatively measuring their “randomness.”

## 1 Introduction

The digit expansions of  $\pi$ ,  $e$ ,  $\sqrt{2}$  and other mathematical constants have fascinated mathematicians from the dawn of history. Indeed, one prime motivation in computing and analyzing digits of  $\pi$  is to explore the age-old question of whether and why these digits appear “random.” The first computation on ENIAC in 1949 of  $\pi$  to 2037 decimal places was proposed by John von Neumann so as to shed some light on the distribution of  $\pi$  (and of  $e$ ) [14, pg. 277–281].

One key question of some significance is whether (and why) numbers such as  $\pi$  and  $e$  are “normal.” A real constant  $\alpha$  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  $\alpha$  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”).

---

\*Centre for Computer Assisted Research Mathematics and its Applications (CARMA), University of Newcastle, Callaghan, NSW 2308, Australia. [francisco.aragon@ua.es](mailto:francisco.aragon@ua.es)

†Lawrence Berkeley National Laboratory, Berkeley, CA 94720. [dhbailey@lbl.gov](mailto:dhbailey@lbl.gov). 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.

‡Centre for Computer Assisted Research Mathematics and its Applications (CARMA), University of Newcastle, Callaghan, NSW 2308, Australia. [jonathan.borwein@newcastle.edu.au](mailto:jonathan.borwein@newcastle.edu.au). Distinguished Professor, King Abdul-Aziz University, Jeddah.

§Director, IRMACS, Simon Fraser University.

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  $0.12345678910111213141516\dots$ , 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 [25].

At the present time, 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  $\alpha_{2,3} = \alpha_{2,3}(0) = \sum_{k \geq 1} 1/(3^k 2^{3^k}) = 0.0418836808315030\dots$  is provably 2-normal (this special case was proven by Stoneham in 1973 [42]). A similar result applies if 2 and 3 in formula (13) are replaced by any pair of coprime integers  $(b, c)$  with  $b \geq 2$  and  $c \geq 2$  [10]. More recently, Bailey and Michal Misieurwicz 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 [36]. However, these results fall far short of establishing  $b$ -normality for any irrational algebraic in any base  $b$ , even in the single-digit sense.

## 2 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 directly study 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 [45, 46], the authors tested the first roughly four trillion hexadecimal digits of  $\pi$  by means of a Poisson process model: in this model, it is extraordinarily unlikely that  $\pi$  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 this data as walks in the plane.

In this work, based in part on sources such as [21, 22, 20, 18, 13], we make a more rigorous and quantitative study of these walks on numbers. We pay particular attention to  $\pi$  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 paper is as follows. In Section 3 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 travelled after  $N$  steps (Section 4) and the number of sites visited (Section 5). In Section 6 we discuss measuring the fractal (actually box) dimension of our walks. In Section 7 we describe two classes for which normality and nonnormality results are known, and one for which we have only surmise. In Section 8 we show some various examples and leave some open questions. Finally, in Appendix 9 we collect the numbers we have examined, with concise definitions and a few digits in various bases.

## 3 Walking on numbers

### 3.1 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.

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

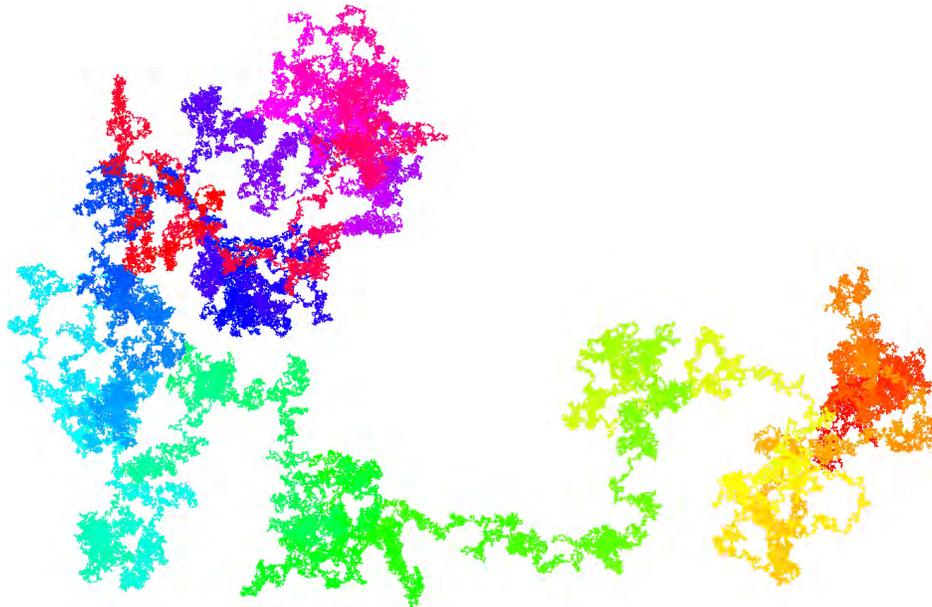


Figure 1: A uniform pseudorandom walk.

```

/
1612226962694290912940490066273549214229880755725468512353395718465
1913530173488143140175045399694454793530120643833272670970079330526
2920303509209736004509554561365966493250783914647728401623856513742
9529453089612268152748875615658076162410788075184599421938774835
  
```

```

Q2=
7278984857066874130428336124347736557760097920257997246066053320967
1510416153622193809833373062647935595578496622633151106310912260966
7568778977976821682512653537303069288477901523227013159658247897670
30435402490295493942131091063934014849602813952
  
```

```

/
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.

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).

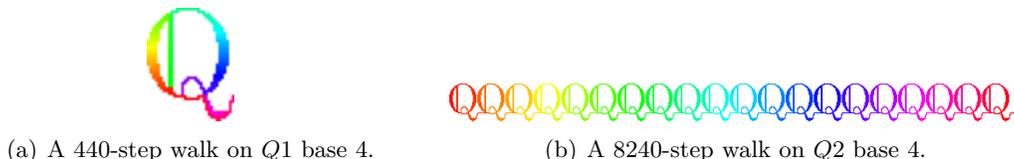


Figure 2: Walks on the rational numbers  $Q1$  and  $Q2$ .

Of course, not all rational numbers are that easily identified by plotting their walk. It is possible to create a rational number whose period is of any desired length. For example, the following rational numbers from [38],

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

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 4. 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 8. Where the base  $b \geq 3$ , the base- $b$  digits can be used to select, as a direction, the corresponding base- $b$  complex root of unity—a multiple of  $120^\circ$  for base three, a multiple of  $90^\circ$  for base four, a multiple of  $72^\circ$  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 3 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 8 is the same as the above, except that Figure 6 is colored to indicate the number of returns to each point.

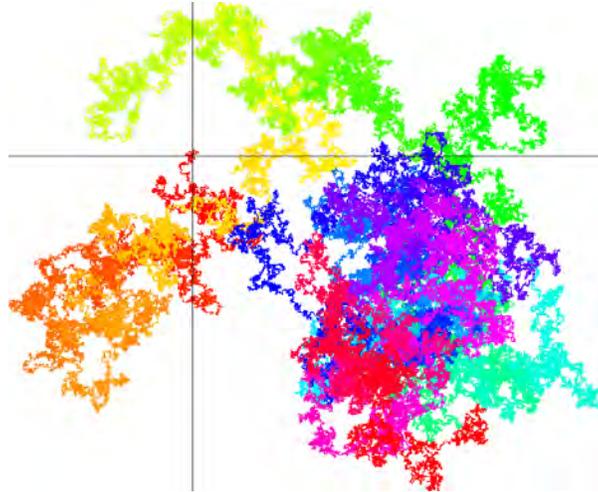


Figure 3: A million step base-4 walk on  $e$ .

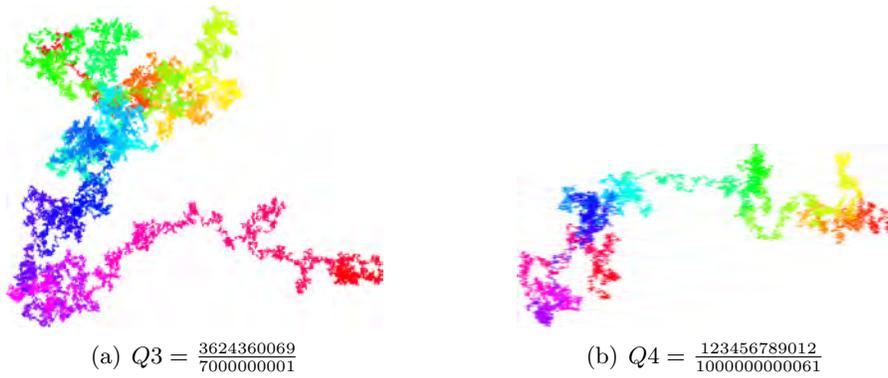


Figure 4: Walks on the first million base 10-digits of the rationals  $Q3$  and  $Q4$  from [38].

## 3.2 Normal numbers as walks

As noted above, proving normality for specific constants of interest in mathematics has proven remarkably difficult. The tenor of current knowledge in this arena is illustrated by [44, 13, 33, 37, 39, 38, 43]. It is useful to know that while small in measure, the “absolutely abnormal” or “absolutely nonnormal” 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 [17, p. 352].

One question that has possessed mathematicians for centuries is whether  $\pi$  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  $\pi$ . This may be viewed dynamically in more detail online at <http://gigapan.org/gigapans/106803>, where the full-sized image has a resolution of  $372,224 \times 290,218$  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  $\pi$ , 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 concluded that, according to this test, it is “extraordinarily unlikely” that  $\pi$  is not 16-normal (of course, this result does not pretend to be a proof).

In what follows, we propose various methods to analyze real numbers and visualize them as walks. Other methods widely used to visualize numbers include the matrix representations shown in Figure 9, 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 9(c). This scheme also shows the nonnormality of the number  $\alpha_{2,3}$ ; see Figure 9(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 9(e). Nevertheless, it does not reveal the apparently nonrandom behavior of numbers like the Erdős–Borwein constant; compare Figure 9(f) with Figure 8(e). See also Figure 23.

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.

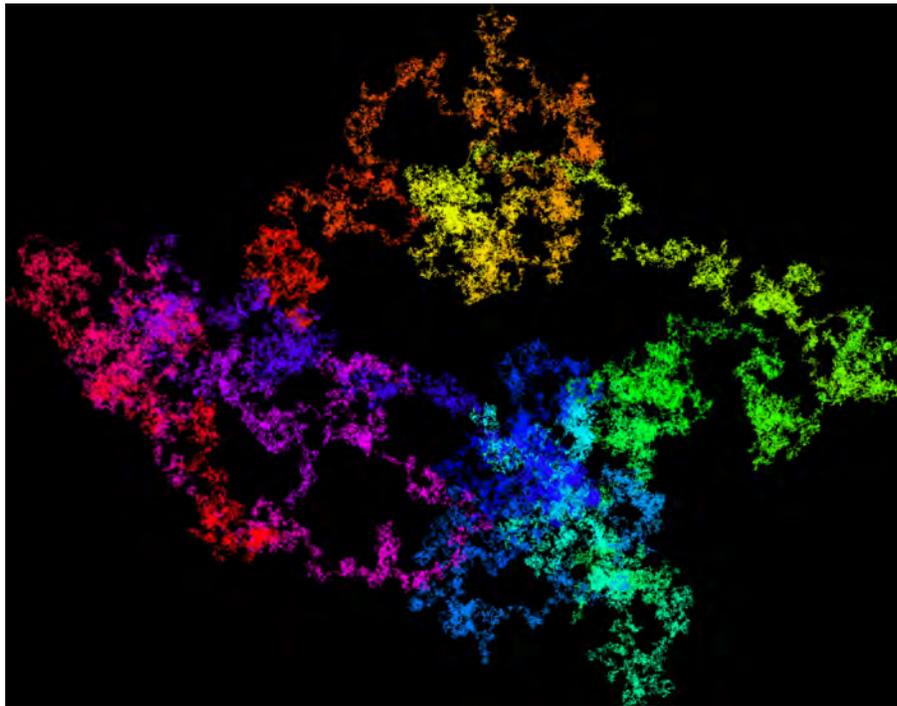


Figure 5: A walk on the first 100 billion base-4 digits of  $\pi$  (normal?).

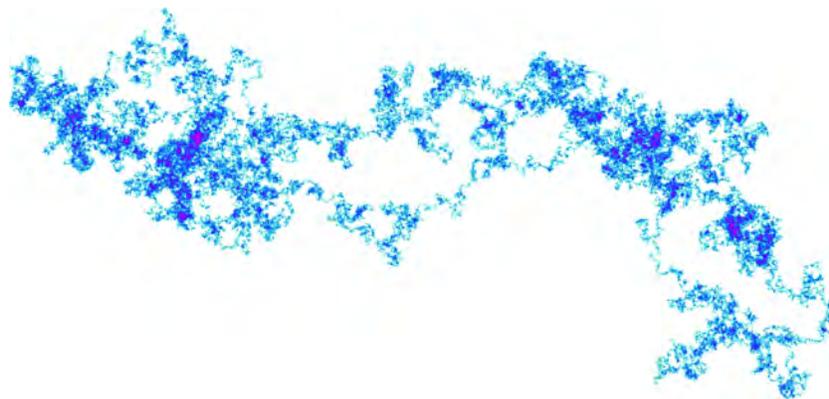


Figure 6: A walk on the 100 million base-4 digits of  $\pi$ , colored by number of returns (normal?). Color follows an HSV model (green-cyan-blue-purple-red) depending on the number of returns to each point (where the maximum is colored in pink/red).



Figure 7: A walk on the first 100,000 bits of the primes ( $CE(2)$ ) base two (normal).

## 4 Expected distance to the origin

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

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

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 4.1.** *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  $\mu$  is the two-dimensional mean vector of  $X$  and  $M$  is its  $2 \times 2$  covariance matrix. By applying Lagrange's trigonometric identities, one gets

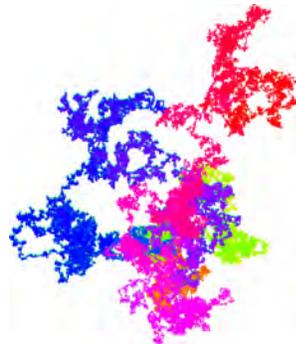
$$\mu = \begin{pmatrix} \frac{1}{b} \sum_{k=1}^b \cos\left(\frac{2\pi}{b}k\right) \\ \frac{1}{b} \sum_{k=1}^b \sin\left(\frac{2\pi}{b}k\right) \end{pmatrix} = \frac{1}{b} \begin{pmatrix} -\frac{1}{2} + \frac{\sin\left(\frac{(b+1/2)2\pi}{b}\right)}{2 \sin(\pi/b)} \\ \frac{1}{2} \cot(\pi/b) - \frac{\cos\left(\frac{(b+1/2)2\pi}{b}\right)}{2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



(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).

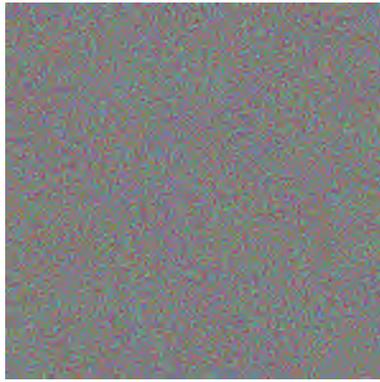


(e) A million step walk on  $EB(2)$  base 4 (normal?).

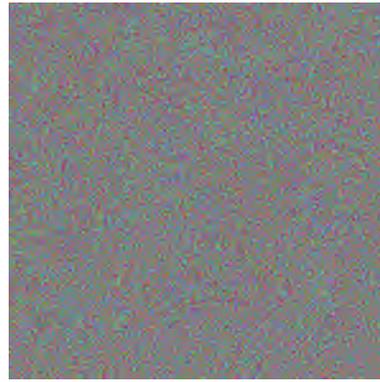


(f) A million step walk on  $CE(10)$  base 4 (normal?).

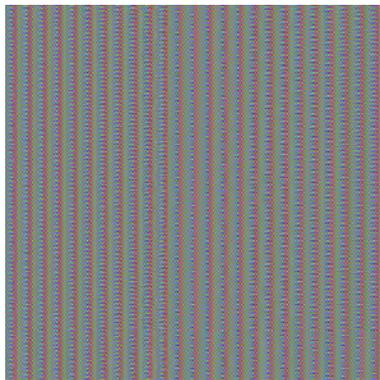
Figure 8: Walks on various numbers in different bases.



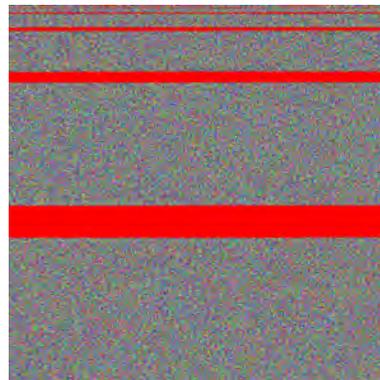
(a)  $\pi$  base 4



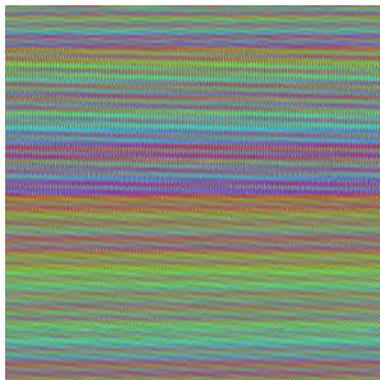
(b) (Pseudo)random number base 4



(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 9: Horizontal color representation of a million digits of various numbers.

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 $\pi$	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
$\pi$	4	1,000,000	0.84366	?
$\pi$	6	1,000,000	0.96458	?
$\pi$	10	1,000,000	0.82167	?
$\pi$	10	10,000,000	0.56856	?
$\pi$	10	100,000,000	0.94725	?
$\pi$	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 $\gamma$	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

Table 1: Average of the normalized distance to the origin (i.e. multiplied by  $2/\sqrt{\pi N}$ , where  $N$  is the number of steps) of the walk of various constants in different bases.

Thus,

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

Since

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

formula (4) reduces to

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

Hence,  $1/\sqrt{N}S^N$  is asymptotically bivariate normal with mean  $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$  and covariance matrix  $M$ . Since its components  $(1/\sqrt{N}S_1^N, 1/\sqrt{N}S_2^N)^T$  are uncorrelated, then they are independent random variables, whose distribution is (univariate) normal with mean 0 and variance 1/2. Therefore, the random variable

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

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

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

where  $E(\cdot)$  stands for the expectation of a random variable. Therefore, the expected distance to the origin of the random walk is asymptotically equal to  $\sqrt{\pi N}/2$ .  $\square$

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 (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<sup>1</sup>, 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, while 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  $\pi$ . The resulting arithmetic mean is 1.0008, while 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 this data, such as the fact that by this measure, Champernowne’s number  $C_{10}$  is far from what is expected of a truly “random” number.

## 5 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, e.g., [35, pg. 338] or [12, pg. 27]. This result was first proven by Dvoretzky and Erdős [32, Thm. 1]. The main practical problem with this asymptotic result is that its convergence is rather slow; specifically, it has order of  $O(N \log \log N/(\log N)^2)$ . In [30, 31], Downham and Fotopoulos show the following bounds on the expectation of the number of distinct points,

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

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), while  $\pi N/\log(N) = 227396$ , 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  $\pi$  fit quite well the expectation. On the other hand, numbers that are known to be normal like  $\alpha_{2,3}$  or

---

<sup>1</sup>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 the number of protons in the universe, which is approximately  $10^{80}$ .

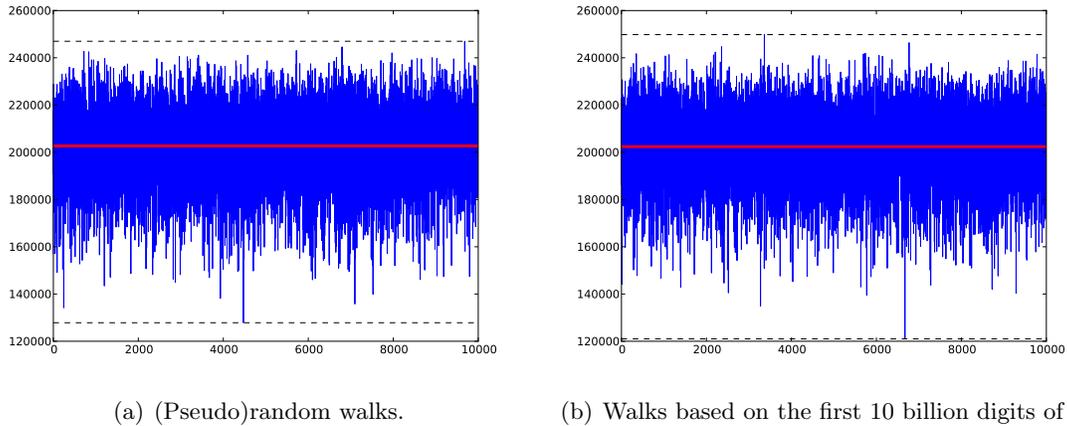


Figure 10: Number of points visited by  $10^4$  base-4 million steps walks.

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 8(d).

At first look the walk on  $\alpha_{2,3}$  might seem random, see Figure 8(c). A closer look, shown in Figure 11, reveals 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 given in Section 7.2, where the precise structure of Figure 11 is conjectured.

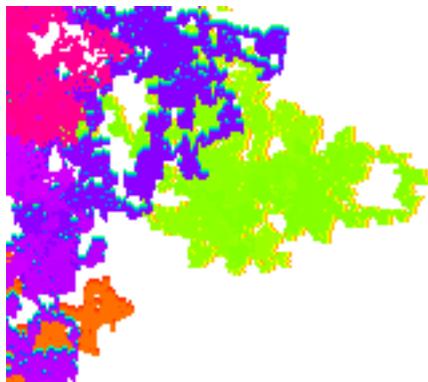
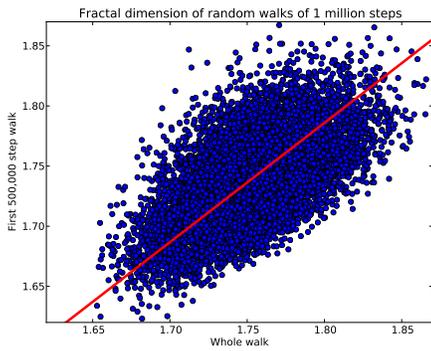


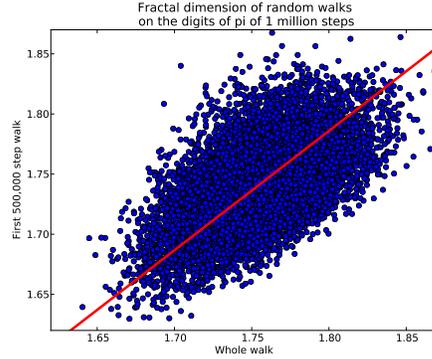
Figure 11: Zooming in on the base-4 walk on  $\alpha_{2,3}$  of Figure 8(c) and Conjecture 7.6.

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 $\pi$	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
$\pi$	1,000,000	204,148	199,256	203,060
$\pi$	10,000,000	1,933,903	1,738,645	1,767,533
$\pi$	100,000,000	16,109,429	15,421,296	15,648,132
$\pi$	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 $\gamma$	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

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 [30, Theorem 2] and [31].



(a) Whole walk/half walk, random.



(b) Whole walk/half walk, digits of  $\pi$ .

Figure 12: Comparison of approximate box dimension of 10,000 walks. On the left are results for a sample of (pseudo)random walks, while on the right we used the first ten billion digits of  $\pi$  in blocks of one million digits.

## 6 Fractal and box-dimension

Another approach that can be taken is to estimate the fractal dimensions of walks. (One can observe in each of the pictures in Figures 1 through 8 that the walks on numbers exhibit a fractal-like structure.) The fractal dimension is an appropriate tool 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–Bouligand dimension*, permits us to estimate the fractal dimension of a given set and often coincides with the fractal dimension. If we denote by  $\#\text{box}_\epsilon(A)$  the number of boxes of side length  $\epsilon > 0$  required to cover a compact set  $A \subset \mathbb{R}^n$ , the box-counting dimension is defined as

$$d_{\text{box}}(A) := \lim_{\epsilon \rightarrow 0} \frac{\log(\#\text{box}_\epsilon(A))}{\log(1/\epsilon)}. \quad (10)$$

The box-counting dimension of a given image is easily estimated by dividing the image into a non-overlapping regular grid and counting the number of nonempty boxes for different grid sizes. We estimated the box-counting dimension  $c$  by the slope of a linear regression model on  $\log(1/\epsilon)$  and the logarithm of the number of nonempty boxes for different values of box-size  $\epsilon$ . This method seems to be both efficient and stable for analyzing “randomness.”

A random walk on a two-dimensional lattice, being space-filling (see e.g. [35, pp. 124–125]), has fractal dimension two—if we were able to use infinitely many steps. It also returns to any point with probability one. Note in Figure 5 that after 100 billion steps on  $\pi$  we are back close to the origin.

For the aforementioned sample of 10,000 pseudorandom walks of one million steps, the average of their box-counting dimension is 1.752, with a variance of 0.0011. The average of the box-counting dimension of these same walks with 500,000 steps is somewhat lower at 1.738, with a variance of 0.0013. We also computed the box-counting dimension of 10,000 walks based on the first ten billion digits of  $\pi$ . The average dimension of the one million steps walks is 1.753, with a variance of 0.0011. The average dimension of these same walks with the smaller length of 500,000 steps is necessarily somewhat lower at 1.739, with a variance of 0.0013. Note how random  $\pi$  seems as shown in Figure 12.

## 7 Copeland–Erdős, Stoneham, and Erdős–Borwein constants

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

### 7.1 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, i.e.,  $C_{10} = 0.12345678910111213141516\dots$ . Champernowne proved that  $C_{10}$  is 10-normal in 1933 [23]. This was later extended to base- $b$  normality (for base- $b$  versions of the Champernowne constant) as in Theorem 7.1. In 1946, Copeland and Erdős established that the corresponding concatenation of primes  $0.23571113171923\dots$  and the concatenation of composites  $0.46891012141516\dots$ , among others, are also 10-normal [25]. In general they proved that concatenation leads to normality if the sequence grows slowly enough. We call such numbers *concatenation numbers*:

**Theorem 7.1** ([25]). *If  $a_1, a_2, \dots$  is an increasing sequence of integers such that for every  $\theta < 1$  the number of  $a_i$ 's up to  $N$  exceeds  $N^\theta$  provided  $N$  is sufficiently large, then the infinite decimal*

$$0.a_1a_2a_3\dots$$

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

This result clearly applies to the Champernowne numbers (Figure 8(d)), to the primes of the form  $ak+c$  with  $a$  and  $c$  relatively prime, in any given base, and to the integers which are the sum of two squares (since every prime of the form  $4k+1$  is included). In further illustration, using the primes in binary leads to normality in base two of the number

$$CE(2) = 0.1011101111101111011000110011101111110111111100101101001101011\dots_2,$$

as shown as a planar walk in Figure 7.

### 7.1.1 Strong normality

In [13] it is shown that  $C_{10}$  fails the following stronger test of normality which we now discuss. The test is a simple one, in the spirit of Borel's test of normality, as opposed to other more statistical tests discussed in [13]. If the digits of a real number  $\alpha$  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  $m_k(n) := \#\{i: a_i = k, i \leq n\}$  when  $\alpha$  has fractional part  $0.a_0 a_1 a_2 \dots$  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 [13] use this to define their criterion and then show that almost every number is absolutely strongly normal.

**Definition 7.2** (Strong normality [13]). *For real  $\alpha$ , and  $m_k(n)$  as above,  $\alpha$  is simply strongly normal in the base  $b$  if for each  $0 \leq k \leq b-1$  one has*

$$\limsup_{n \rightarrow \infty} \frac{m_k(n) - n/b}{\sqrt{2n \log \log n}} = \frac{\sqrt{b-1}}{b} \quad \text{and} \quad \liminf_{n \rightarrow \infty} \frac{m_k(n) - n/b}{\sqrt{2n \log \log n}} = -\frac{\sqrt{b-1}}{b}. \quad (11)$$

*A number is strongly normal in base  $b$  if it is simply strongly normal in each base  $b^j$ ,  $j = 1, 2, 3, \dots$ , 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  $b^{\nu-1}$  digits of length  $\nu$  and they all start with a digit between 1 and  $b-1$  while the following  $\nu-1$  digits take values between 0 and  $b-1$  equally. In consequence, there is a dearth of zeroes. This is easiest to analyze 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  $\nu = 3$  there are 4 zeroes and 8 ones, for  $\nu = 4$  there are 12 zeroes and 20 ones, and for  $\nu = 5$  there are 32 zeroes and 48 ones.

Since the details were not given in [13] we give them here.

**Theorem 7.3.** (Belshaw and P. Borwein) *Champernowne's base-2 number is not 2-strongly normal.*

*Proof.* In general, let  $n_k := 1 + (k-1)2^k$  for  $k \geq 1$ . 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  $n \leq n_k$ , and proceed by induction on  $k$ . Let us arrange the digits of the integers  $2^k, 2^k + 1, \dots, 2^k + 2^{k-1} - 1$  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 first  $n_{k-1}$  digits of Champernowne's base-2 number, where by our induction hypothesis  $m_1(n) > m_0(n)$  for  $n \leq n_{k-1}$ .

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 we showed above,  $m_1(n_k) - m_0(n_k) = 2^k - 1$ , so  $m_1(n) > m_0(n) + 2^{k-1}$  for every  $n \leq n_k + (k + 1)2^{k-1}$ . 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  $n \leq n_{k+1}$ . Therefore,  $2m_1(n) > m_0(n) + m_1(n) = n$  for all  $n$ , and so

$$\liminf_{n \rightarrow \infty} \frac{m_1(n) - n/2}{\sqrt{2n \log \log n}} \geq 0 \neq -\frac{1}{2},$$

and as asserted Champernowne's base-2 number is not 2-strongly normal.  $\square$

It seems likely that by appropriately shuffling the integers, one should be able to display a strongly normal variant. Along this line, Martin [39] has shown how to construct an explicit absolutely nonnormal number.

Finally, while the log log limiting behavior required by (11) appears hard 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 7.1, fail to be strongly normal for similar reasons.

Heuristically, we would expect the number  $CE(2)$  above to fail to be strongly normal, since each prime of length  $k$  both starts and ends with a one, while intermediate bits should show no skewing. Indeed, for  $CE(2)$  we have checked that  $2m_1(n) > n$  for all  $n \leq 10^9$ , see also Figure 13(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 13(b) and leads us to conjecture that it is 2-strongly normal.

## 7.2 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

$$\alpha_{b,c} := \sum_{m \geq 1} \frac{1}{c^m b^{c^m}}, \tag{12}$$

for relatively prime integers  $b, c$  [10].

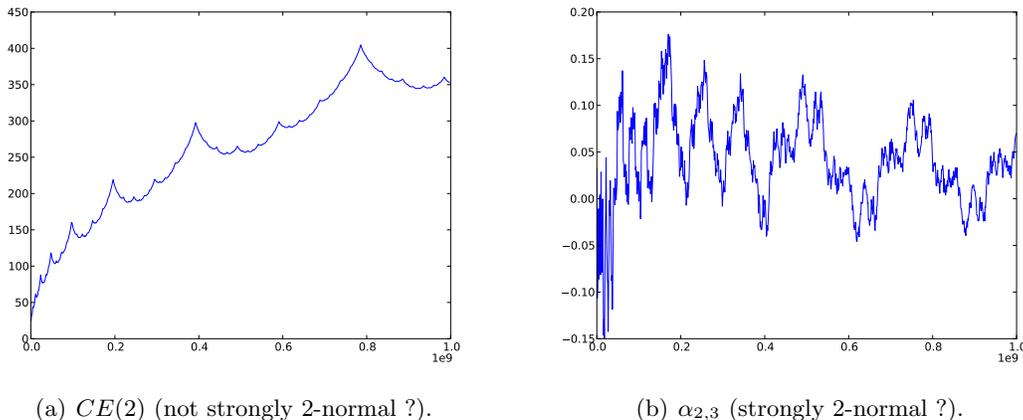


Figure 13: Plot of the first  $10^9$  values of  $\frac{m_1(n) - n/2}{\sqrt{2n \log \log n}}$ .

**Theorem 7.4** (Normality of Stoneham constants [3]). *For every coprime pair of integers  $(b, c)$  with  $b \geq 2$  and  $c \geq 2$ , the constant  $\alpha_{b,c} = \sum_{m \geq 1} 1/(c^m b^{c^m})$  is  $b$ -normal.*

So, for example, the constant  $\alpha_{2,3} = \sum_{k \geq 1} 1/(3^k 2^{3^k}) = 0.0418836808315030\dots$  is provably 2-normal. This special case was proven by Stoneham in 1973 [42]. More recently, Bailey and Misiurewicz were able to establish this normality result by a much simpler argument, based on techniques of ergodic theory [11] [15, pg. 141–173].

Equally interesting is the following result:

**Theorem 7.5** (Nonnormality of Stoneham constants [3]). *Given coprime integers  $b \geq 2$  and  $c \geq 2$ , and integers  $p, q, r \geq 1$ , 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  $\alpha_{b,c} = \sum_{k \geq 0} 1/(c^k b^{c^k})$  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 8(b) to Figure 8(c) and Figure 8(a). Now compare the values given in Table 1 and Table 2. Clearly, from this sort of visual and numeric data, the discovery of other cases of Theorem 7.5 is very easy.

As illustrated also in the “zoom” of Figure 11, 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):



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 one [10] [15, pg. 141–173]. We have not yet studied this generalized class by graphical methods.

### 7.3 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* [17]:

$$EB(b) := \sum_{n \geq 1} \frac{1}{b^n - 1} = \sum_{n \geq 1} \frac{\tau(n)}{b^n}, \quad (14)$$

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 non-zero rational and  $b = 2, 3, \dots$  such that  $r \neq b^n$  for all  $n$  [19]. Whence, as provably irrational numbers other than the standard examples are few and far between, it is interesting to consider their normality.

Crandall [26] has observed that the structure of (14) is analogous to the “BBP” formula for  $\pi$  (see [7, 15]), as well as some nontrivial knowledge of the arithmetic properties of  $\tau$ , 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 [26] 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 **4359105565638** zeroes and **4436987456570** 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 8(e) but contrast it to Figure 9(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 [26].

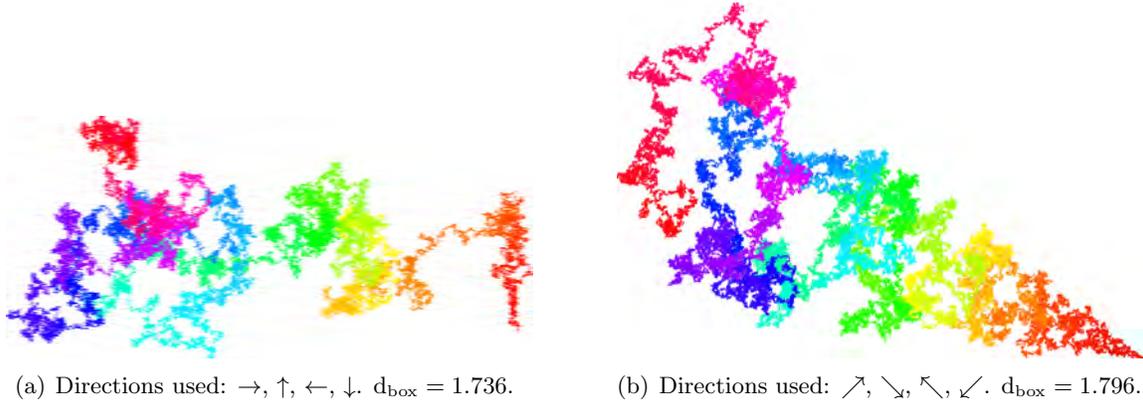


Figure 15: Two different rules for plotting a base-2 walk on the first two million values of  $\lambda(n)$  (the Liouville number  $\lambda_2$ ).

## 8 Other avenues and concluding remarks

Let us recall two further examples utilized in [13], that of the *Liouville function* which counts the parity of the number of prime factors of  $n$ , namely  $\Omega(n)$  (see Figure 15) and of the human genome taken from the *UCSC Genome Browser* at <http://hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/>, see Figure 18. 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 7.1 and 7.2 and indicated in Appendix 9. While 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  $\lambda_2$  (see Figure 15) exhibits a much more structured drift than  $\pi$  or  $e$ , but looks more like them than like Figure 18(a).

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

### 8.1 Three dimensions

We have also explored three-dimensional graphics—using base-6 for directions—both in perspective as in Figure 16, and in a large passive (glasses-free) three-dimensional viewer outside the CARMA laboratory; but have not yet quantified these excursions.

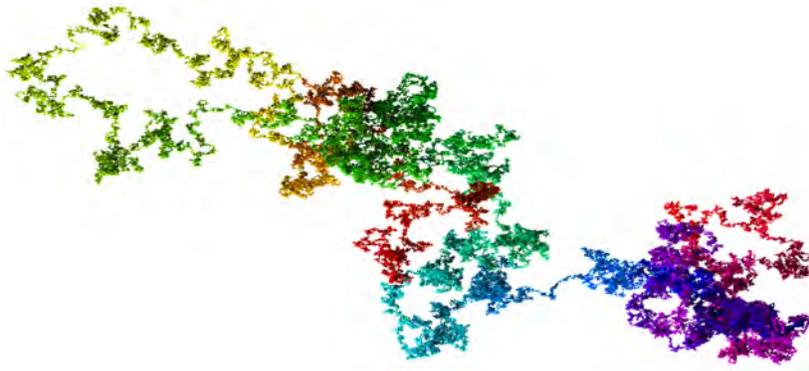


Figure 16: A 3D walk on the first million base 6 digits of  $\pi$ .

## 8.2 Genome comparison

Genomes are made up of so called purine and pyrimidines nucleotides. In DNA, purine nucleotide bases are adenine and guanine (A and G), while 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-four uniform walk on the symbols (A,G,T,C) illustrated in Figure 18 (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.<sup>2</sup> As an illustration we show twelve games in Figure 17: four on a triangle, four on a square, and four on a hexagon. At each step we go from the current point halfway towards 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 7.2 (the non-normality of  $\alpha_{2,3}$  and  $\alpha_{3,2}$  in base 6 yields a paler color, while the repeating structure of  $\alpha_{2,3}$  and  $\alpha_{3,5}$  is the origin of the purple tone, see Conjectures 7.6 and 7.7).

## 8.3 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 9.

---

<sup>2</sup>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 [34]. For a recent summary see [16, pp. 194–205].

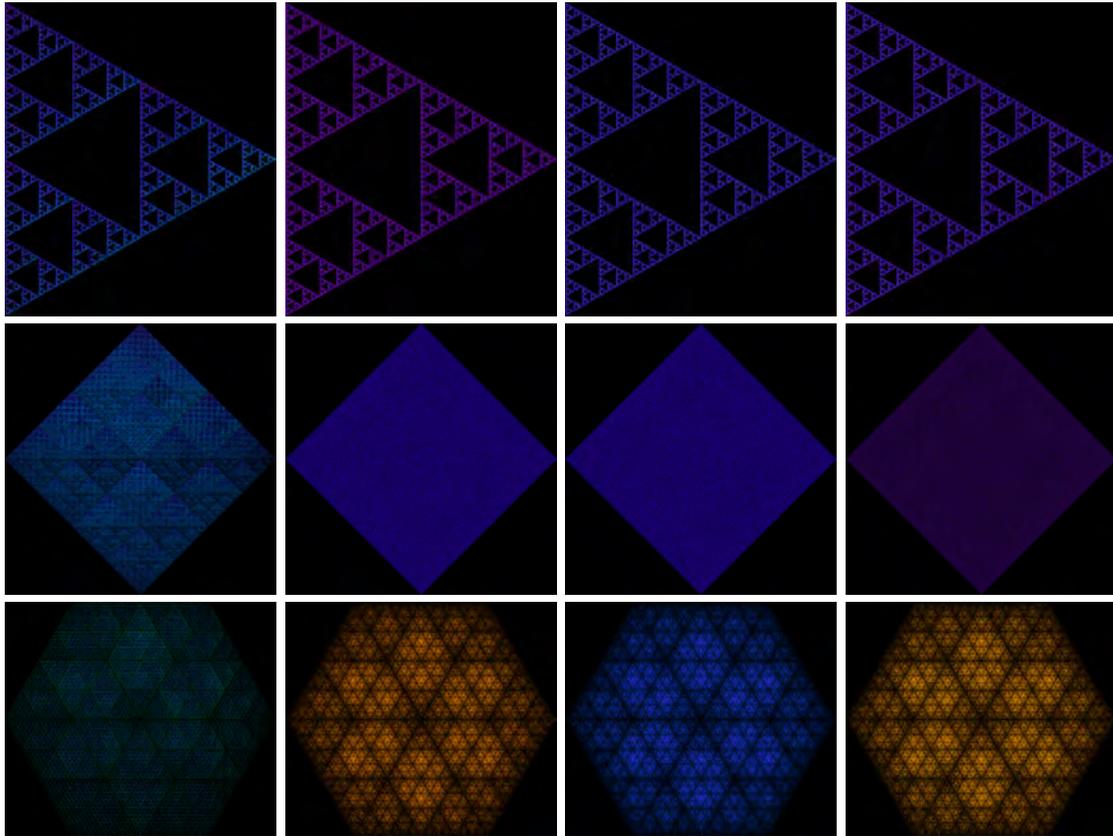


Figure 17: Chaos games on various numbers, colored by frequency. Row 1:  $C_3$ ,  $\alpha_{3,5}$ , a (pseudo)random number, and  $\alpha_{2,3}$ . Row 2:  $C_4$ ,  $\pi$ , a (pseudo)random number, and  $\alpha_{2,3}$ . Row 3:  $C_6$ ,  $\alpha_{3,2}$ , a (pseudo)random number, and  $\alpha_{2,3}$ .

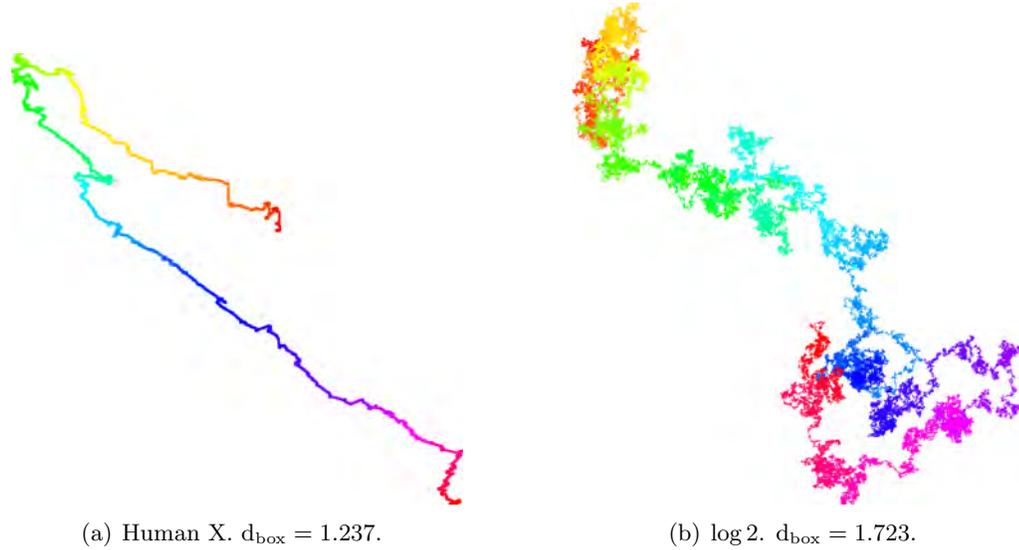


Figure 18: Base four walks on  $10^6$  bases of the X-chromosome and  $10^6$  digits of  $\log 2$ .

Automatic numbers are never normal and are typically transcendental; by comparison the Liouville number  $\lambda_2$  is not  $p$ -automatic for any prime  $p$  [24].

The walks on  $\mathcal{P}$  and  $\mathcal{TM}_2$  have a similar shape, see Figure 19, but while the Thue–Morse sequence generates very large pictures, the paper-folding sequence generates very small ones since 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 one or “rotate the Logo turtle” a fixed angle, exhibits some of their striking features (see Figure 20). For instance, drawn with a rotating angle of  $\pi/3$ ,  $\mathcal{TM}_2$  converges to a Koch snowflake [40], see Figure 20(c). We show a corresponding turtle graphic of  $\pi$  in Figure 20(d). Analogous features occur for the paper-folding sequence as described in [27, 28, 29], and two variants are shown in Figures 20(a) and 20(b).

## 8.4 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 number is rational or a quadratic irrational, respectively; see [15, 7]. By contrast, the simple continued fractions for  $\pi$  and  $e$  start as follows in the standard compact form:

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

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

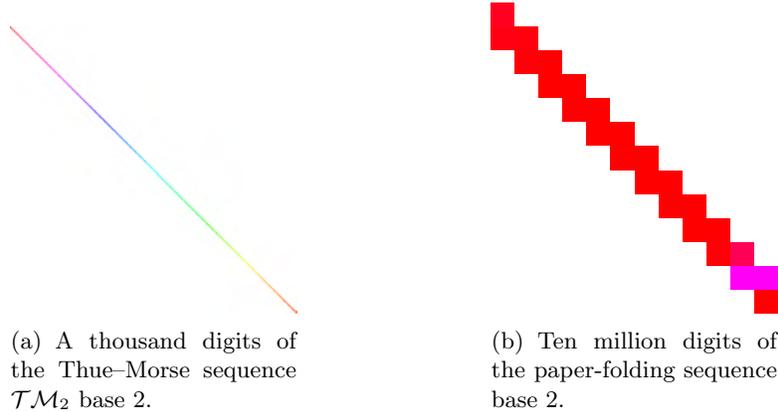


Figure 19: Walks on two automatic and nonnormal numbers.

from which the surprising regularity of  $e$  and apparent irregularity of  $\pi$  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  $\alpha = [a_1, a_2, \dots, a_n, \dots]$ , for which the *Gauss–Kuzmin distribution* holds [15]: for each  $k = 1, 2, 3, \dots$

$$\text{Prob}\{a_n = k\} = -\log_2 \left( 1 - \frac{1}{(k+1)^2} \right), \quad (15)$$

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

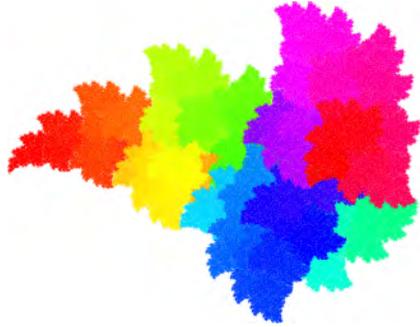
In Figure 21 we show a histogram of the first 100 million terms, computed by Neil Bickford and accessible at <http://neilbickford.com/picf.htm>, of the continued fraction of  $\pi$ . We have not yet found a satisfactory way to embed this in a walk on a continued fraction but in Figure 22 we show base-4 walks on  $\pi$  and  $e$  where we use the remainder modulo four to build the walk (with 0 being right, 1 being up 2 being left and 3 being down). We also show turtle plots on  $\pi, e$ .

Andrew Mattingly has observed that:

**Proposition 8.1.** *With probability one, a mod four 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:*

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

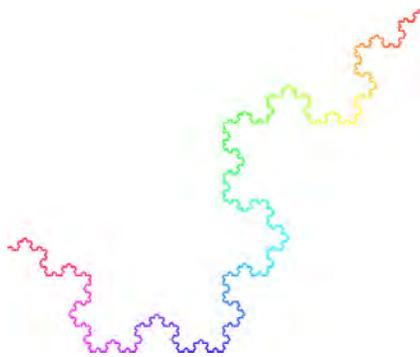
*Proof.* The result comes by summing the expected Gauss–Kuzmin probabilities of each step being taken as given by (15). □



(a) Ten million digits of the paper-folding sequence with rotating angle  $\pi/3$ .  $d_{\text{box}} = 1.921$ .



(b) Dragon curve from one million digits of the paper-folding sequence with rotating angle  $2\pi/3$ .  $d_{\text{box}} = 1.783$ .

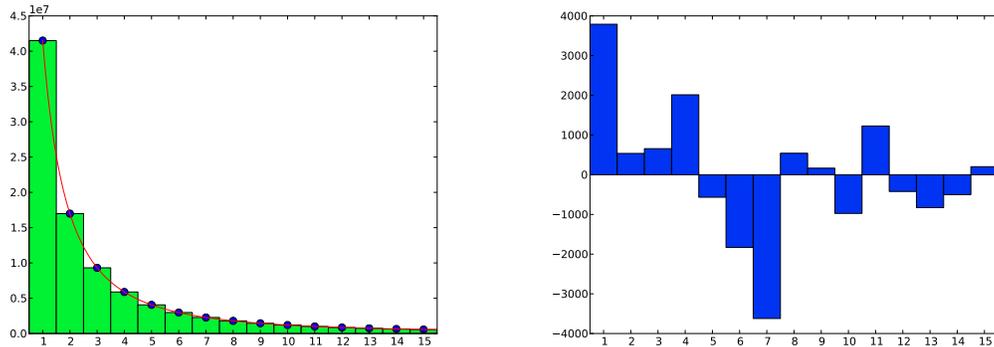


(c) Koch snowflake from 100,000 digits of the Thue–Morse sequence  $\mathcal{TM}_2$  with rotating angle  $\pi/3$ .  $d_{\text{box}} = 1.353$ .



(d) One million digits of  $\pi$  with rotating angle  $\pi/3$ .  $d_{\text{box}} = 1.760$ .

Figure 20: Turtle plots on various constants with different rotating angles in base 2—where ‘0’ gives forward motion and ‘1’ rotation by a fixed angle.



(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 21: Expected values of the Gauss–Kuzmin distribution of (15) and the values of 100 million terms of the continued fraction of  $\pi$ .

This is illustrated in Figure 22(a) with a  $90^\circ$  anticlockwise rotation; thus making the case that one must have some a priori knowledge before designing tools.

It is also instructive to compare at digits and continued fractions of numbers as horizontal matrix plots of the form already used in Figure 9. In Figure 23 we show six pairs of million-term digit-strings and their corresponding fraction. In some cases both look random, in others one or the other does.

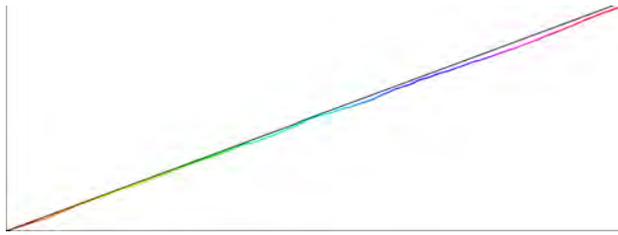
In conclusion, we have only tapped the surface of what is becoming possible in a period in which data—e.g., five hundred million terms of the continued fraction or five trillion binary digits of  $\pi$ , full genomes and much more—can be downloaded from the internet, then rendered and visually mined, with fair rapidity.

## 9 Appendix: Selected numerical constants

In what follows,  $x := 0.a_1a_2a_3a_4 \dots_b$  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?):

$$G := \sum_{k=0}^{\infty} \frac{(-1)^k}{(2k+1)^2} = 0.9159655941 \dots \quad (16)$$



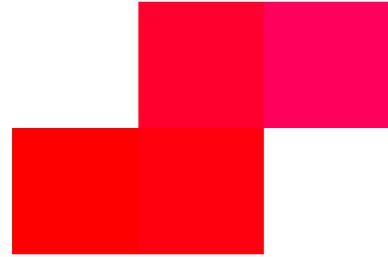
(a) A 100,000 step walk on the continued fraction of  $\pi$  modulo 4.



(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  $\pi$  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 22: Uniform walks on  $\pi$  and  $e$  based on continued fractions.

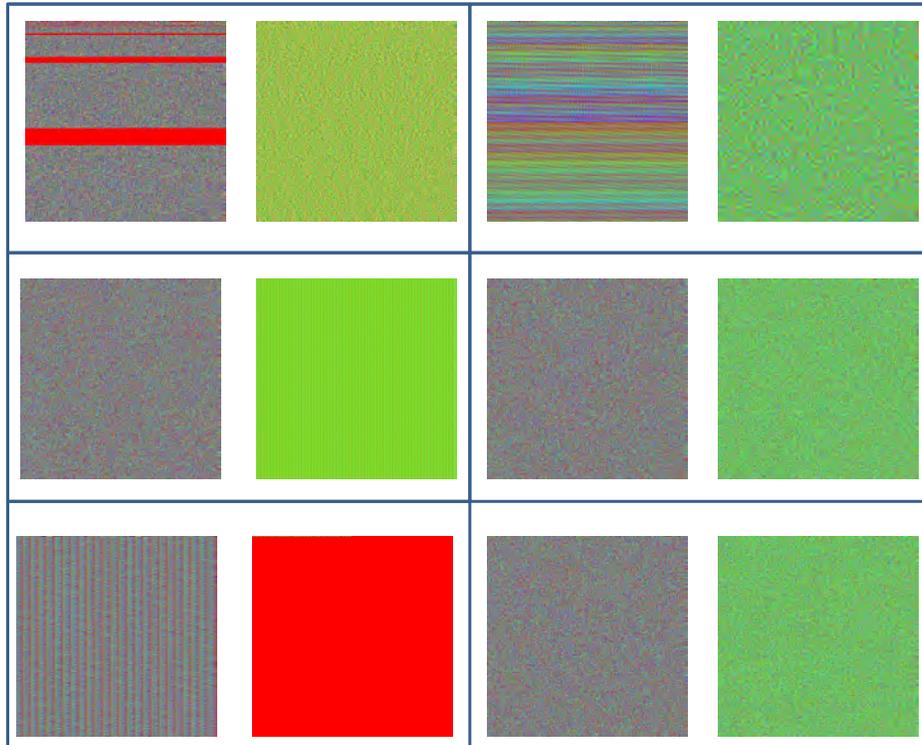


Figure 23: Million step comparisons of base-4 digits and continued fractions. Row 1:  $\alpha_{2,3}$ (base 6) and  $C_4$ . Row 2:  $e$  and  $\pi$ . Row 3:  $Q_1$  and pseudorandom iterates; as listed from top left to bottom right.

**Champernowne numbers** (irrational; normal to corresponding base):

$$C_b := \sum_{k=1}^{\infty} \frac{\sum_{m=b^{k-1}}^{b^k-1} mb^{-k[m-(b^{k-1}-1)]}}{b \sum_{m=0}^{k-1} m(b-1)b^{m-1}} \quad (17)$$

$$C_{10} = 0.123456789101112\dots$$

$$C_4 = 0.1231011121320212223\dots_4$$

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

$$CE(b) := \sum_{k=1}^{\infty} p_k b^{-(k+\sum_{m=1}^k \lfloor \log_b p_m \rfloor)}, \text{ where } p_k \text{ is the } k^{\text{th}} \text{ prime number} \quad (18)$$

$$CE(10) = 0.2357111317\dots$$

$$CE(2) = 0.1011101111\dots_2$$

**Exponential constant** (transcendental; normal?):

$$e := \sum_{k=0}^{\infty} \frac{1}{k!} = 2.7182818284\dots \quad (19)$$

**Erdős–Borwein constants** (irrational; normal?):

$$EB(b) := \sum_{k=1}^{\infty} \frac{1}{b^k - 1} \quad (20)$$

$$EB(2) = 1.6066951524\dots = 1.212311001\dots_4$$

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

$$\gamma := \lim_{m \rightarrow \infty} \left( \sum_{k=1}^m \frac{1}{k} - \log m \right) = 0.5772156649\dots \quad (21)$$

**Fibonacci constant** (irrational?; normal?):

$$\mathcal{F} := \sum_{k=1}^{\infty} F_k 10^{-(1+k+\sum_{m=1}^k \lfloor \log_{10} F_m \rfloor)}, \text{ where } F_k = \frac{\left(\frac{1+\sqrt{5}}{2}\right)^k - \left(\frac{1-\sqrt{5}}{2}\right)^k}{\sqrt{5}} \quad (22)$$

$$= 0.011235813213455\dots$$

**Liouville number** (irrational; not  $p$ -automatic):

$$\lambda_2 := \sum_{k=1}^{\infty} \left( \frac{\lambda(k) + 1}{2} \right) 2^{-k} \quad (23)$$

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

$$= 0.5811623188\dots = 0.10010100110\dots_2$$

**Logarithmic constant** (transcendental; normal?):

$$\begin{aligned} \log 2 &:= \sum_{k=1}^{\infty} \frac{1}{k2^k} \\ &= 0.6931471806\dots = 0.10110001011100100001\dots_2 \end{aligned} \quad (24)$$

**Pi** (transcendental; normal?):

$$\begin{aligned} \pi &:= 2 \int_{-1}^1 \sqrt{1-x^2} dx = 4 \sum_{k=0}^{\infty} \frac{(-1)^k}{2k+1} \\ &= 3.1415926535\dots = 11.00100100001111110110\dots_2 \end{aligned} \quad (25)$$

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

$$\zeta(s) := \sum_{k=1}^{\infty} \frac{1}{k^s} \quad (26)$$

In particular:

$$\begin{aligned} \zeta(2) &= \frac{\pi^2}{6} = 1.6449340668\dots \\ \zeta(2n) &= (-1)^{n+1} \frac{(2\pi)^{2n}}{2(2n)!} B_{2n} \quad (\text{where } B_{2n} \text{ are Bernoulli numbers}) \\ \zeta(3) &= \text{Apéry's constant} = \frac{5}{2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^3 \binom{2k}{k}} = 1.2020569031\dots \end{aligned}$$

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

$$\alpha_{b,c} := \sum_{k=1}^{\infty} \frac{1}{b^{c^k} c^k} \quad (27)$$

$$\alpha_{2,3} = 0.0418836808\dots = 0.0022232032\dots_4 = 0.0130140430003334\dots_6$$

$$\alpha_{4,3} = 0.0052087571\dots = 0.000111111301\dots_4 = 0.0010430041343502130000\dots_6$$

$$\alpha_{3,2} = 0.0586610287\dots = 0.0011202021212121\dots_3 = 0.0204005200030544000002\dots_6$$

$$\alpha_{3,5} = 0.0008230452\dots = 0.00000012101210121\dots_3 = 0.002ba00000061d2\dots_{15}$$

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

$$\begin{aligned} \mathcal{TM}_2 &:= \sum_{k=1}^{\infty} \frac{1}{2^{t(n)}} \text{ where } t(0) = 0, \text{ while } t(2n) = t(n) \text{ and } t(2n+1) = 1 - t(n) \\ &= 0.4124540336\dots = 0.01101001100101101001011001101001\dots_2 \end{aligned} \quad (28)$$

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

$$\mathcal{P} := \sum_{k=0}^{\infty} \frac{8^{2^k}}{2^{2^{k+2}} - 1} = 0.8507361882\dots = 0.1101100111001001\dots_2 \quad (29)$$

## Acknowledgements

The authors would like to thank Dr David Allingham, for his kind help in running the algorithms for plotting the “big walks” on  $\pi$ ; Adrian Belshaw, for his assistance with strong normality; Matt Skerritt for his 3D image of  $\pi$ ; and Jake Fountain, who produced a fine Python interface for us as a 2011/12 Vacation Scholar at CARMA. We are also most grateful for several discussions with Andrew Mattingly (IBM) and Michael Coons (CARMA), who helped especially with our discussions of continued fractions and automatic numbers, respectively.

## References

- [1] S. Albeverio, M. Pratsiovytyie and, G. Torbine, “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, in press 2012.
- [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*, in press 2012.
- [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), pg. 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] M.N. Barber and B.W. Ninham, *Random and Restricted Walks: Theory and Applications*, Gordon and Breach, New York, 1970.
- [13] 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, in press, 2012.
- [14] L. Berggren, J.M. Borwein and P.B. Borwein, *Pi: a Source Book*, Springer-Verlag, Third Edition, 2004.
- [15] J.M. Borwein and D.H. Bailey, *Mathematics by Experiment: Plausible Reasoning in the 21st Century*, A K Peters, Natick, MA, second edition, 2008.
- [16] J. Borwein, D. Bailey, N. Calkin, R. Girgensohn, R. Luke, and V. Moll, *Experimental Mathematics in Action*, A.K. Peters, 2007.
- [17] 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.
- [18] 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.
- [19] P.B. Borwein, “On the irrationality of certain series.” *Math. Proc. Cambridge Philos. Soc.* **112** (1992), 141–146.
- [20] P.B. Borwein and L. Jörgenson, “Visible structures in number theory,” *Amer. Math. Monthly* **108** (2001), no. 10, 897–910.
- [21] C.S. Calude, “Borel normality and algorithmic randomness,” in G. Rozenberg, A. Salomaa (eds.), *Developments in Language Theory*, World Scientific, Singapore, 1994, 113–129.
- [22] C.S. Calude, *Information and Randomness: An Algorithmic Perspective*, 2nd Edition, Revised and Extended, Springer-Verlag, Berlin, 2002.
- [23] D.G. Champernowne, “The construction of decimals normal in the scale of ten,” *Journal of the London Mathematical Society*, **8** (1933), 254–260.
- [24] M. Coons, “(Non)automaticity of number theoretic functions,” *J. Théor. Nombres Bordeaux*, **22** (2010), no. 2, 339–352.
- [25] A.H. Copeland and P. Erdős, “Note on normal numbers,” *Bulletin of the American Mathematical Society*, **52** (1946), 857–860.
- [26] R.E. Crandall, “The googol-th bit of the Erdős–Borwein constant,” *Integers*, A23, 2012.
- [27] M. Dekking, M. Mendès France and A. van der Poorten, “Folds,” *Math. Intelligencer* **4** (1982), no. 3, 130–138.
- [28] M. Dekking, M. Mendès France and A. van der Poorten, “Folds II,” *Math. Intelligencer* **4** (1982), no. 4, 173–181.
- [29] M. Dekking, M. Mendès France and A. van der Poorten, “Folds III,” *Math. Intelligencer* **4** (1982), no. 4, 190–195.

- [30] 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.
- [31] 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.
- [32] 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.
- [33] P. Hertling, “Simply normal numbers to different bases,” *Journal of Universal Computer Science*, **8**, no. 2 (2002), 235–242.
- [34] H.Joel Jeffrey, Chaos game representation of gene structure, *Nucl. Acids Res.* **18** no 2, (1990) 2163–2170. doi:10.1093/nar/18.8.2163.
- [35] B.D. Hughes, *Random Walks and Random Environments, Vol. 1. Random Walks*, Oxford Science Publications, New York, 1995.
- [36] H. Kaneko, “On normal numbers and powers of algebraic numbers,” *Integers*, **10** (2010), 31–64.
- [37] D. Khoshnevisan, “Normal numbers are normal,” *Clay Mathematics Institute Annual Report* (2006), 15 & 27–31.
- [38] G. Marsaglia, “On the randomness of pi and other decimal expansions,” preprint, 2010.
- [39] G. Martin, “Absolutely abnormal numbers,” *Amer. Math. Monthly*, **108** (2001), no. 8, 746–754.
- [40] J. Mah and J. Holdener, “When Thue–Morse meets Koch,” *Fractals*, **13** (2005), no. 3, 191–206.
- [41] S.M. Ross, *Stochastic Processes*. John Wiley & Sons, New York, 1983.
- [42] R. Stoneham, “On absolute  $(j, \varepsilon)$ -normality in the rational fractions with applications to normal numbers,” *Acta Arithmetica*, **22** (1973), 277–286.
- [43] M. Queffelec, “Old and new results on normality.” *Lecture Notes – Monograph Series*, **48**, *Dynamics and Stochastics*, 2006, Institute of Mathematical Statistics, 225–236.
- [44] W. Schmidt, “On normal numbers,” *Pacific Journal of Mathematics*, **10** (1960), 661–672.
- [45] A.J. Yee, “y-cruncher-multi-threaded pi program,” <http://www.numberworld.org/y-cruncher>, 2010.
- [46] 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>.