Tools for visualizing real numbers.

Part I: Planar number walks

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

May 13, 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

In the recent paper [4], by accessing the results of several extremely large recent computa-
tions [44, 45], the authors tested positively the normality of a prefix of roughly four trillion
hexadecimal digits of π. This result was used by a Poisson process model of normality of
π: in this model, it is extraordinarily unlikely that π is not asymptotically normal base 16,
given the normality of its initial segment. During that work, the authors of [4] like many
others looked for visual methods of representing their large mathematical data sets. Their
chosen tool was representation as walks in the plane. In this work, based in part on sources
such as [21, 22, 20, 18, 14] we make a more rigorous and quantitative study of such walks
on numbers.

The organization of the paper is as follows. In Section 2 we describe and exhibit
uniform walks on various numbers, both rational and irrational, artificial and ‘natural’. In

∗Centre for Computer Assisted Research Mathematics and its Applications (CARMA), University of
Newcastle, Callaghan, NSW 2308, Australia. francisco.aragon@ua.es
†Lawrence Berkeley National Laboratory, Berkeley, CA 94720. 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. Distinguished Pro-
fessor King Abdul-Aziz Univ, Jeddah.
§Director IRMACS, Simon Fraser University.
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 3) and the number of sites
visited (Section 4). In Section 5 we discuss measuring the fractal (actually box) dimension
of our walks. In Section 6 we describe two classes for which normality and non-normality
results are known and one for which we have only surmise. In Section 7 we show some
further examples and leave some open questions. Finally, in Appendix 8 we collect the
numbers we have examined.

2 Walking on numbers

One of our tasks is to compare pseudo-random walks to deterministic walks of the same
length. For example, in Figure 1 we draw a uniform pseudo random-walk with a million
base-4 steps.

\[
\begin{align*}
Q_1 & = \frac{1}{3} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \cdots \\
Q_2 & = \frac{1}{3} - \frac{1}{4} + \frac{1}{8} - \frac{1}{16} + \cdots
\end{align*}
\]

Below one will see very similar and widely dissimilar walks on the numbers we study.

2.1 A first few steps

At first glance the following two rational numbers \( Q_1 \) and \( Q_2 \) look quite different:
Their decimal expansion gives us some additional information: they both agree in the first 240 digits (in base 4, their first 400 digits are the same). In Figure 2 a plot of the walk on the digits of $Q_1$ and $Q_2$ in base 4 is shown. In order to create it, we have transformed the rational numbers $Q_1$ and $Q_2$ into their base 4 decimal expansion, obtaining two (periodic) sequences whose elements are either 0, 1, 2 or 3. At each step a new point is drawn adjacent to the previous one: 0, 1, 2 and 3 draw the new point to the east, north, west and south, respectively.

2.2 Random and deterministic walks

The idea is the same as when a random walk is plotted, with the difference being that the direction followed at each step is encoded in the digits of the number instead of being chosen randomly. The color is shifted up the spectrum (red-orange-yellow-green-cyan-blue-purple-red) following an HSV\(^1\) model with \(S\) and \(V\) equal to one, which permits to visualize the path followed by the walk. Following this same method, in each of the pictures in Figures 2 through 8, a digit string for a given number in a particular base is used to determine the angle of unit steps (multiples of 120 degrees for base three, 90 degrees for base four, etc). In Figure 3 the origin has been marked. Since this information is not that important for our purposes and it can be approximately deduced by the color in most of the cases (and

\(^1\)HSV (hue, saturation, and value) is a cylindrical-coordinate representation that permits to get an easy rainbow-like range of colors.
the pictures look less esthetic, let us face it), we have not marked it in the rest. Figure 6 is colored in a different way: by the number of returns to each point.

For the reader’s convenience, we have included an appendix with the definitions of the mathematical constants addressed in this paper, see Section 8. We also exhibit there a few digits in various bases for each number.

The rational numbers $Q_1$ and $Q_2$ represent the two possibilities when one computes a walk on a rational number: either the walk is bounded like 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 like in Figure 2(b).

![Figure 2: Walks on the rational numbers $Q_1$ and $Q_2$.](image)

(a) A 440 step walk on $Q_1$ base 4. (b) A 8,240 step walk on $Q_2$ base 4.

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 [37],

$$Q_3 = \frac{3624360069}{7000000001} \quad \text{and} \quad Q_4 = \frac{123456789012}{1000000000061},$$

have a periodic part in base 10 with length\(^2\) 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.

2.3 Numbers as walks

In the following, given some positive integer base $b$, we will say that a real number $x$ is $b$-normal if every $m$-long string of base-$b$ digits appears in the base-$b$ expansion of $x$ with precisely the expected limiting frequency $1/b^m$. It follows, from basic measure theory, as shown by Borel, that almost all real numbers are $b$-normal for any specific base $b$ and even for all bases simultaneously [15].

But proving normality for specific constants of interest in mathematics has proven remarkably difficult. It is useful to know that while small in measure, the absolutely abnormal numbers 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 Liouville numbers are measure zero but second category [17, p. 352].

\(^2\)The numerators and denominators of $Q_3$ and $Q_4$ are relatively prime, and the denominators are not congruent to 2 or 5, thus the period is simply the discrete logarithm of the denominator.
Figure 3: A million step base-4 walk on $e$.

Figure 4: Walks on the first million base 10-digits of the rationals $Q_3$ and $Q_4$ from [37].

(a) $Q_3 = \frac{3624360009}{10000000001}$

(b) $Q_4 = \frac{123456789012}{1000000000061}$
For example, a reasonable conjecture is that every irrational algebraic number (i.e., the irrational roots of polynomial equations with integer coefficients) is $b$-normal for every integer base $b$. Yet there is no proof of this conjecture, not for any specific algebraic number or for any specific integer base. The tenor of current knowledge is illustrated by [43, 14, 33, 36, 38, 37, 42].

In Figure 5 we show a walk on the first 10 billion base 4 digits of $\pi$. This may be viewed in more detail online at http://gigapan.org/gigapans/99214/. In Figure 6 a 100 million base 4 walk on $\pi$ is shown where the color is determined by the number of returns to the point. It is unknown whether $\pi$ is normal or not. In [4] the authors have empirically tested the normality of roughly four trillion of its hexadecimal digits by using a Poisson process model, concluding that it is “extraordinarily unlikely” that $\pi$ is not 16-normal.

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 rep-

\footnote{The full-size picture has a resolution of 149,818 $\times$ 136,312 pixels, that is, 20.42 gigapixels.}
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).
(a) A million step walk on $\alpha_{2,3}$ base 3 (normal?).

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

(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 prefixes of various numbers in different bases.
resentations shown in Figure 9, where each pixel is colored depending on the value of the digit, 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). It also shows the nonnormality of the number \( \alpha_{2,3} \), see Figure 9(e) (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(d). Nevertheless, it does not reveal the apparently non-random behavior of numbers like the Erdős-Borwein constant: compare Figure 9(f) with Figure 8(e).

As we will see in what follows, the study of real numbers as walks will permit us to compare them with random 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\(^4\), 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

\[
P \left( X = \left( \cos \left( \frac{2\pi}{b} k \right), \sin \left( \frac{2\pi}{b} k \right) \right) \right) = \frac{1}{b} \quad \text{for } k = 1, \ldots, b.
\]

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 \( 0_2 \) and covariance matrix \( M \), where \( \mu \) is the bi-dimensional mean vector of \( X \) and \( M \) is its \( 2 \times 2 \) covariance matrix. By applying Lagrange’s trigonometric identities, one gets

\[
\mu = \left( \frac{1}{b} \sum_{k=1}^{b} \cos \left( \frac{2\pi}{b} k \right), \frac{1}{b} \sum_{k=1}^{b} \sin \left( \frac{2\pi}{b} k \right) \right) = \frac{1}{b} \left( -\frac{1}{2} + \frac{\sin \left( (b+1/2) \frac{2\pi}{b} \right)}{2 \sin \left( \frac{\pi}{b} \right)} \cos \left( \frac{2\pi}{b} \right) \right) = 0_2.
\]

\(^4\)We treat the case \( b = 2 \) as a base-4 walk. We could consider it as a base-2 walk on a line, but the pictures would be much less interesting.
Figure 9: Horizontal color representation of a million digits of various numbers.
<table>
<thead>
<tr>
<th>Number</th>
<th>Base</th>
<th>Steps</th>
<th>Average normalized distance to the origin</th>
<th>Normal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of 10,000 random walks</td>
<td>4</td>
<td>1,000,000</td>
<td>1.00315</td>
<td>Yes</td>
</tr>
<tr>
<td>$\alpha_{2,3}$</td>
<td>3</td>
<td>1,000,000</td>
<td>0.89275</td>
<td>?</td>
</tr>
<tr>
<td>$\alpha_{2,3}$</td>
<td>4</td>
<td>1,000,000</td>
<td>0.25901</td>
<td>Yes</td>
</tr>
<tr>
<td>$\alpha_{2,3}$</td>
<td>5</td>
<td>1,000,000</td>
<td>0.88104</td>
<td>?</td>
</tr>
<tr>
<td>$\alpha_{2,3}$</td>
<td>6</td>
<td>1,000,000</td>
<td>108.02218</td>
<td>No</td>
</tr>
<tr>
<td>$\alpha_{4,3}$</td>
<td>3</td>
<td>1,000,000</td>
<td>1.07223</td>
<td>?</td>
</tr>
<tr>
<td>$\alpha_{4,3}$</td>
<td>4</td>
<td>1,000,000</td>
<td>0.24268</td>
<td>Yes</td>
</tr>
<tr>
<td>$\alpha_{4,3}$</td>
<td>6</td>
<td>1,000,000</td>
<td>94.54563</td>
<td>No</td>
</tr>
<tr>
<td>$\alpha_{4,3}$</td>
<td>12</td>
<td>1,000,000</td>
<td>371.24994</td>
<td>No</td>
</tr>
<tr>
<td>$\alpha_{3,5}$</td>
<td>3</td>
<td>1,000,000</td>
<td>0.32511</td>
<td>Yes</td>
</tr>
<tr>
<td>$\alpha_{3,5}$</td>
<td>5</td>
<td>1,000,000</td>
<td>0.85258</td>
<td>?</td>
</tr>
<tr>
<td>$\alpha_{3,5}$</td>
<td>15</td>
<td>1,000,000</td>
<td>370.93128</td>
<td>No</td>
</tr>
<tr>
<td>$\pi$</td>
<td>4</td>
<td>1,000,000</td>
<td>0.84366</td>
<td>?</td>
</tr>
<tr>
<td>$\pi$</td>
<td>6</td>
<td>1,000,000</td>
<td>0.96458</td>
<td>?</td>
</tr>
<tr>
<td>$\pi$</td>
<td>10</td>
<td>1,000,000</td>
<td>0.82167</td>
<td>?</td>
</tr>
<tr>
<td>$\pi$</td>
<td>10</td>
<td>10,000,000</td>
<td>0.56856</td>
<td>?</td>
</tr>
<tr>
<td>$\pi$</td>
<td>10</td>
<td>100,000,000</td>
<td>0.94725</td>
<td>?</td>
</tr>
<tr>
<td>$\pi$</td>
<td>10</td>
<td>1,000,000,000</td>
<td>0.59824</td>
<td>?</td>
</tr>
<tr>
<td>$e$</td>
<td>4</td>
<td>1,000,000</td>
<td>0.59583</td>
<td>?</td>
</tr>
<tr>
<td>$\sqrt{2}$</td>
<td>4</td>
<td>1,000,000</td>
<td>0.72260</td>
<td>?</td>
</tr>
<tr>
<td>$\log 2$</td>
<td>4</td>
<td>1,000,000</td>
<td>1.21113</td>
<td>?</td>
</tr>
<tr>
<td>Champernownne $C_{10}$</td>
<td>10</td>
<td>1,000,000</td>
<td>59.91143</td>
<td>Yes</td>
</tr>
<tr>
<td>$EB(2)$</td>
<td>4</td>
<td>1,000,000</td>
<td>6.95831</td>
<td>?</td>
</tr>
<tr>
<td>$CE(10)$</td>
<td>4</td>
<td>1,000,000</td>
<td>0.94964</td>
<td>?</td>
</tr>
<tr>
<td>Rational number $Q_1$</td>
<td>4</td>
<td>1,000,000</td>
<td>0.04105</td>
<td>No</td>
</tr>
<tr>
<td>Rational number $Q_2$</td>
<td>4</td>
<td>1,000,000</td>
<td>58.40222</td>
<td>No</td>
</tr>
<tr>
<td>Euler constant $\gamma$</td>
<td>10</td>
<td>1,000,000</td>
<td>1.17216</td>
<td>?</td>
</tr>
<tr>
<td>Fibonacci $F$</td>
<td>10</td>
<td>1,000,000</td>
<td>1.24820</td>
<td>?</td>
</tr>
<tr>
<td>$\zeta(2) = \frac{\pi^2}{6}$</td>
<td>4</td>
<td>1,000,000</td>
<td>1.57571</td>
<td>?</td>
</tr>
<tr>
<td>$\zeta(3)$</td>
<td>4</td>
<td>1,000,000</td>
<td>1.04085</td>
<td>?</td>
</tr>
<tr>
<td>Catalan’s constant $G$</td>
<td>4</td>
<td>1,000,000</td>
<td>0.53489</td>
<td>?</td>
</tr>
<tr>
<td>Thue-Morse $TM_2$</td>
<td>4</td>
<td>1,000,000</td>
<td>0.92344</td>
<td>?</td>
</tr>
<tr>
<td>Paper-folding $P$</td>
<td>4</td>
<td>1,000,000</td>
<td>0.01336</td>
<td>?</td>
</tr>
</tbody>
</table>

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, 

\[ M = \frac{1}{b} \left[ \sum_{k=1}^{b} \cos \left( \frac{2\pi}{b} k \right) \sin \left( \frac{2\pi}{b} k \right) \sum_{k=1}^{b} \cos \left( \frac{4\pi}{b} k \right) \sum_{k=1}^{b} \sin \left( \frac{4\pi}{b} k \right) \sum_{k=1}^{b} \sin \left( \frac{4\pi}{b} k \right) \right] \]

Since

\[
\begin{align*}
\sum_{k=1}^{b} \cos^2 \left( \frac{2\pi}{b} k \right) &= \sum_{k=1}^{b} \frac{1 + \cos \left( \frac{4\pi}{b} k \right)}{2} = \frac{b}{2}, \\
\sum_{k=1}^{b} \sin^2 \left( \frac{2\pi}{b} k \right) &= \sum_{k=1}^{b} \frac{1 - \cos \left( \frac{4\pi}{b} k \right)}{2} = \frac{b}{2}, \\
\sum_{k=1}^{b} \cos \left( \frac{2\pi}{b} k \right) \sin \left( \frac{2\pi}{b} k \right) &= \sum_{k=1}^{b} \frac{\sin \left( \frac{4\pi}{b} k \right)}{2} = 0,
\end{align*}
\]

one has

\[ M = \left[ \frac{1}{2} \quad 0 \quad \frac{1}{2} \right]. \]

Hence, \( \frac{1}{\sqrt{N}} S^N \) is asymptotically bivariate normal with mean \( \mathbf{0}_2 \) and covariance matrix \( M \). Since its components \((1/\sqrt{N} S^N_1, 1/\sqrt{N} S^N_2)^T \) are uncorrelated, then they are independent random variables, whose distribution is (univariate) normal with mean 0 and variance \( 1/2 \). Therefore, the random variable

\[
\sqrt{\left( \frac{\sqrt{2}}{\sqrt{N}} S^N_1 \right)^2 + \left( \frac{\sqrt{2}}{\sqrt{N}} S^N_2 \right)^2}
\]

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

\[
E \left( \sqrt{(S^N_1)^2 + (S^N_2)^2} \right) = \frac{\sqrt{N}}{\sqrt{2}} E \left( \sqrt{\left( \frac{\sqrt{2}}{\sqrt{N}} S^N_1 \right)^2 + \left( \frac{\sqrt{2}}{\sqrt{N}} S^N_2 \right)^2} \right) \\
\approx \frac{\sqrt{N}}{\sqrt{2}} \frac{\Gamma(3/2)}{\Gamma(1)} = \frac{\sqrt{\pi N}}{2},
\]

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 asymptotically to 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 denominated as the 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\(^5\), 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 variance is 0.1351, so the asymptotic result fits quite well. In Table 1 we show the average normalized distance to the origin of various numbers.

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 an \( N \)-step random walk on a two-dimensional lattice is asymptotically equal to \( \pi N / \log(N) \), see e.g. [34, page 338] or [13, page 27]. This result was first proven by Dvoretzky and Erdős in [32, Theorem 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 [30, 31], Downham and Fotopoulos show the following bounds on the expectation of the number of distinct points,

\[
\left[ \frac{\pi(N + 0.84)}{1.16\pi - 1 - \log 2 + \log(N + 2)} , \frac{\pi(N + 1)}{1.066\pi - 1 - \log 2 + \log(N + 1)} \right],
\]

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 \([199,256.1, 203,059.5]\) while \( \pi N / \log(N) = 227,396 \), 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 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 visit, the walk on the constant \( \alpha_{2,3} \) 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 6.2 where the precise structure of Figure 11 is conjectured.

\(^5\)Python uses the Mersenne Twister as the core generator and produces 53-bit precision floats, with a period of \( 2^{19937} - 1 \approx 10^{6006} \). Compare the length of this period to the comoving distance from Earth to the edge of the observable universe in any direction, which has a magnitude of \( 4.6 \cdot 10^{29} \) millimeters.
Figure 10: Number of points visited by $10^4$ base-4 (pseudo)random million steps walks.

Figure 11: Zooming in on the base-4 walk on $\alpha_{2,3}$ of Figure 8(c) and Conjecture 6.5.
<table>
<thead>
<tr>
<th>Number</th>
<th>Steps</th>
<th>Sites visited</th>
<th>Bounds on the expectation of sites visited by a random walk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean of 10,000 random walks</td>
<td>1,000,000</td>
<td>202,684</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>( \alpha_{2,3} )</td>
<td>1,000,000</td>
<td>95,817</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>( \alpha_{4,3} )</td>
<td>1,000,000</td>
<td>68,613</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>( \alpha_{3,2} )</td>
<td>1,000,000</td>
<td>195,585</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>( \pi )</td>
<td>1,000,000</td>
<td>204,148</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>( \pi )</td>
<td>10,000,000</td>
<td>1,933,903</td>
<td>1,738,645 1,767,533</td>
</tr>
<tr>
<td>( \pi )</td>
<td>100,000,000</td>
<td>16,109,429</td>
<td>15,421,296 15,648,132</td>
</tr>
<tr>
<td>( \pi )</td>
<td>1,000,000,000</td>
<td>138,107,050</td>
<td>138,552,612 140,380,926</td>
</tr>
<tr>
<td>( e )</td>
<td>1,000,000</td>
<td>176,350</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>( \sqrt{2} )</td>
<td>1,000,000</td>
<td>200,733</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>( \log 2 )</td>
<td>1,000,000</td>
<td>214,508</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>Champernowne ( C_4 )</td>
<td>1,000,000</td>
<td>548,746</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>( \mathit{EB}(2) )</td>
<td>1,000,000</td>
<td>279,585</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>( \mathit{CE}(10) )</td>
<td>1,000,000</td>
<td>190,239</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>Rational number ( Q_1 )</td>
<td>1,000,000</td>
<td>378</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>Rational number ( Q_2 )</td>
<td>1,000,000</td>
<td>939,322</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>Euler constant ( \gamma )</td>
<td>1,000,000</td>
<td>208,957</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>( \zeta(2) )</td>
<td>1,000,000</td>
<td>188,808</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>( \zeta(3) )</td>
<td>1,000,000</td>
<td>221,598</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>Catalan’s constant ( G )</td>
<td>1,000,000</td>
<td>195,853</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>( \mathit{TM}_2 )</td>
<td>1,000,000</td>
<td>1,000,000</td>
<td>199,256 203,060</td>
</tr>
<tr>
<td>Paper-folding ( P )</td>
<td>1,000,000</td>
<td>21</td>
<td>199,256 203,060</td>
</tr>
</tbody>
</table>

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].
5 Fractal and box-dimension

We now turn to our estimates of fractal dimension of number walks. One can appreciate in each of the pictures in Figures 1 through 8 that the walks on numbers exhibit a fractal-like structure. The fractal dimension has been used as an appropriate tool to measure the geometrical complexity of a set, characterizing its space-filling capacity (see e.g. [5] for a nice introduction about fractals). The box-counting dimension, also known as the Minkowski-Bouligand dimension, permits to estimate the fractal dimension of a given set. 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

\[
\text{d}_{\text{box}}(A) := \lim_{\epsilon \to 0} \frac{\log (\#\text{box}_\epsilon(A))}{\log(1/\epsilon)}.
\]

The box-counting dimension of a given image can be easily estimated by dividing the image into a non-overlapping regular grid and counting the number of filled boxes for different grid sizes. Then the box-counting dimension can be approximated 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 \( \epsilon \), see Figure 12.

A random walk, being space-filling, has fractal dimension 2. For the mentioned sample of 10,000 pseudo-random 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 1.738, with a variance of 0.0013. This method seems to be both efficient and stable for analyzing “randomness”, see also Figure 13. The box-counting dimension of various constants is collected in Table 3.

6 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. Most notably as described in the next three subsections.

6.1 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., \( C_{10} = 0.12345678910111213141516 \ldots \). 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 6.1. In 1946, Copeland and Erdős established that the corresponding concatenation of primes \( 0.23571113171923 \ldots \) and also the concatenation of composites \( 0.46891012141516 \ldots \), 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:
Figure 12: Approximate box-counting dimensions of prefixes of various walks.
Figure 13: Comparison of the approximate box dimension of 10,000 random walks.
Theorem 6.1 ([25]). 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 \( N^\theta \) provided \( N \) is sufficiently large, then the infinite decimal
\[
0.a_1a_2a_3\ldots
\]
is normal with respect to the base \( b \) in which these integers are expressed.

This result clearly applies 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 lead to normality in base two of the number
\[
0.1011101111011010011001111011110111101011001101101111010100110111101\ldots,
\]
as shown as a planar walk in Figure 7.

6.1.1 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 \( \alpha \) 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 \( m_k(n) := m_k(n) = \#\{i : a_i = k, i \leq n\} \) when \( \alpha \) has fractional part \( 0.a_0a_1a_2\ldots \) in base \( b \).)

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

Definition 6.2 (Strong normality [14]). 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
\[
\limsup_{n \to \infty} \frac{m_k(n) - n/b}{\sqrt{2n \log \log n}} = -\liminf_{n \to \infty} \frac{m_k(n) - n/b}{\sqrt{2n \log \log n}} = \sqrt{1 - \frac{1}{b}}.
\]

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 \( b^{k-1} \) digits of length \( k \) and they all start with a digit between 1 and \( b - 1 \) while the following \( k - 1 \) digits take values between 0 and \( b - 1 \) equally.
In consequence, there is a dearth of zeroes. This is easiest to analyse base 2. As illustrated below, the concatenated numbers start

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

where for \( k = 3 \) there are 4 zeroes and 8, for \( k = 4 \) there are 12 zeroes and 20 ones, and for \( k = 5 \) there are 32 zeroes and 48 ones.

In general, let \( n_k := 1 + (k - 2)2^k \). Then one has \( m_0(n_k) = 1 + (k - 2)2^{k-1} \) and so

\[ m_1(n_k) = n_k - m_0(n_k) = k2^{k-1} \]

and so \( 2m_1(n_k) \geq n_k \) for \( k \geq 1 \). Thence

\[ \lim \inf \frac{m_1(n) - n/2}{\sqrt{2n \log \log n}} \geq 0 \neq -\sqrt{1 - \frac{1}{2}}. \]

It seems likely that by appropriately shuffling the integers, one should be able to display a strongly normal variant. Relatedly Martin [38] has shown how to construct an explicit absolutely abnormal number.

Finally, while the log log limiting-behaviour required by (3) 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 6.1 fail to be strongly normal for similar reasons. Heuristically, we would expect the number \( CE(2) \) 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.

To Finalize DAVID to test strong normality for \( CE(2) \) and \( \alpha_{2,3} \)

### 6.2 Stoneham numbers: a class containing provably normal and abnormal constants

Giving further motivation for these studies is the recent provision of rigorous proofs of normality for an uncountably infinite class of explicit real numbers, the Stoneham numbers defined by

\[ \alpha_{b,c} := \sum_{m \geq 1} \frac{1}{cm^b c^m} \]  

for relatively prime integers \( b, c \).

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

\[ \alpha_{2,3}(r) := \sum_{k=0}^{\infty} \frac{1}{3^k 2^{3^k} + r_k} \]
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 class of 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 \([10] \ [15, \text{pg. 141–173}]\). More recently, Bailey and Misieurwicz were able to establish this normality result by a much simpler argument, based on techniques of ergodic theory \([11] \ [15, \text{pg. 141–173}]\).

**Theorem 6.3** (Normality of Stoneham constants \([3]\)). For every coprime pair of integers \((b,c)\) with \( b \geq 2 \) and \( c \geq 2 \), the constant \( \alpha_{b,c} = \sum_{m \geq 1} 1/(c^m b^m) \) is \( b \)-normal.

So, for example, the constant \( \alpha_{2,3} = \sum_{k \geq 0} 1/(3^k 2^k) = 0.0418836808315030\ldots \) is provably 2-normal. This special case was proven by Stoneham in 1973 \([41]\). A similar result applies for all \( \alpha_{b,c}(r) \) as above.

Equally interesting is the following result:

**Theorem 6.4** (Abnormality 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 \( D = c^q/p \cdot r^1/b - 1 < 1 \) is satisfied. Then the constant \( \alpha_{b,c} = \sum_{k \geq 0} 1/(c^k b^k) \) is \( B \)-nonnormal.

In various of the Figures and Tables we explore the striking differences of behaviour—proven and unproven—for \( \alpha_{b,c} \) as we vary the base. For instance, the abnormality of \( \alpha_{2,3} \) base six 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 6.4 is very easy.

As illustrated also in the ‘zoom’ of Figure 11, we can use the pictures 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 6.5** (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:

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

\[(ii) \quad a_k = a_{k+3^n} = a_{k+2.3^n} \text{ for all } k = \frac{3}{2}(3^n + 1), \frac{3}{2}(3^n + 1) + 1, \ldots, \frac{3}{2}(3^n + 1) + 3^n - 1.
\]

In Figure 14 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 \( n = 0, 1, \ldots, 11 \), which, together with Figures 8(c) and 11, graphically explains Conjecture 6.5. Similar results seem to hold for other Stoneham constants in other bases. For instance, for \( \alpha_{3,5} \) base 3 we conjecture the following.
Figure 14: 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 $n = 0, 1, \ldots, 11$.

**Conjecture 6.6** (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:

\[(i) \quad \sum_{k=2+5^n+1}^{2+5^n+4 \cdot 5^n} e^{a_k \pi i/2} = \frac{(-1)^n (-1 + \sqrt{3}i)}{2} = e^{(3n+2)\pi i/3},\]

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

6.3 The Erdős-Borwein constants

The constructions of the previous two subsections exhaust most of what is known of concrete irrational examples. 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]:

$$EB(b) := \sum_{n \geq 1} \frac{1}{b^n - 1} = \sum_{n \geq 1} \frac{\sigma(n)}{b^n}$$

(6)

where $\sigma_k$ the sum of the k-th power of the divisors and $\sigma = \sigma_1$. It is known that the numbers $\sum 1/(q^n - r)$ are irrational for $r$ rational and $q = 1/b, b = 2, 3, \ldots$ [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 used the BBP-like structure [7, 15] made obvious in (6), and some non-trivial knowledge of the arithmetic properties of $\sigma$ to establish results such as that the googol-th bit—that is, the bit in position $10^{100}$ to the right of the floating point—is a 1.

In [26] Crandall also computed the full first $2^{43}$ bits of $EB(2)$ (a Terabyte in about a day), and finds that there are $435910556538$ zeroes and $4436987456570$ ones. There is
corresponding variation in the second and third place in the single digit hex distributions. This certainly leaves some doubt as to its normality. See also Figure 8(e) but contrast it to Figure 9(f).

Our own more modest computations of $EB(10)$ base-ten again leave it far from clear that $EB(10)$ is 10-normal. Likewise, Crandall finds that in the first 10,000 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.

We should note that for computational purposes we used the identity

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

for $|q| < 1$, due to Clausen, as did Crandall [26].

---

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

7 Other avenues and concluding remarks

Let us recall two further examples utilized in [14], that of the Liouville function which counts the parity of the number of primes 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 17. Note the similarity of the genome walk to the those of concatenation sequences. We have explored a wide variety of walks on genomes but reserve the results for future paper.

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
Table 3: Box-counting dimensions of various walks (Fig. 12) and turtle plots (Fig. 19).

<table>
<thead>
<tr>
<th>Number</th>
<th>Steps</th>
<th>Box dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random average base 4</td>
<td>500,000</td>
<td>1.738</td>
</tr>
<tr>
<td>Random average base 4</td>
<td>1,000,000</td>
<td>1.752</td>
</tr>
<tr>
<td>$e$ base 4</td>
<td>1,000,000</td>
<td>1.804</td>
</tr>
<tr>
<td>$C_4$ base 4</td>
<td>1,000,000</td>
<td>1.090</td>
</tr>
<tr>
<td>$\pi$ base 4</td>
<td>1,000,000</td>
<td>1.754</td>
</tr>
<tr>
<td>$\pi$ base 4</td>
<td>1,000,000,000</td>
<td>1.842</td>
</tr>
<tr>
<td>$\alpha_{2,3}$ base 6</td>
<td>1,000,000</td>
<td>1.057</td>
</tr>
<tr>
<td>$\alpha_{2,3}$ base 3</td>
<td>1,000,000</td>
<td>1.754</td>
</tr>
<tr>
<td>$P$ angle $\pi/3$</td>
<td>10,000,000</td>
<td>1.921</td>
</tr>
<tr>
<td>$P$ angle $2\pi/3$</td>
<td>1,000,000</td>
<td>1.783</td>
</tr>
<tr>
<td>$T,M_2$ angle $\pi/3$</td>
<td>100,000</td>
<td>1.353</td>
</tr>
<tr>
<td>$\pi$ angle $\pi/3$</td>
<td>1,000,000</td>
<td>1.783</td>
</tr>
</tbody>
</table>

sections 6.1 and 6.2 and indicated in Appendix 8. While one of the clearly non-random 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, shows a much more structured drift than $\pi$ or $e$ but looks more like them than like Figure 17(a).

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

### 7.1 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.

### 7.2 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. Thence, 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 17 (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
7.3 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 \( \mathcal{T}_2 \), see [2] and Section 8. 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 \( \mathcal{T}_2 \) have a similar shape, see Figure 18, 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*\(^6\) on these constants exhibits some of their striking features, see Figure 19. For instance, drawn with a rotating angle of \( \pi/3 \), \( \mathcal{T}_2 \) converges to a Koch snowflake [39], see Figure 19(c). We show a corresponding turtle graphic of \( \pi \) in Figure 19(d). Corresponding features occur for the paper folding sequence as described in [27, 28, 29] and two variants are shown in Figures 19(a) and 19(b). While the walk specific metrics make little sense to measure, we do supply their fractal dimensions in Table 3 below. As might be expected turtle plots of the same length appear to exhibit higher box-dimensions. The corresponding dimensions for our two variants of Liouville walks are recorded below their pictures in Figure 15.

\(^6\)In base 2, each digit correspond to one of the following orders: either “forward motion” of length one or “rotate the Logo turtle” a fixed angle.
Figure 17: Base four walks on $10^6$ bases of the X-chromosome and $10^6$ digits of log 2.

Figure 18: Walks on two automatic and abnormal numbers.
7.4 Continued fractions

Simple continued fractions often encode more information 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 irrationality [15, 7]. By contrast, the simple continued fractions for $\pi$ and $e$ start as follows in the standard compact form:

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

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

from which the surprising regularity of $e$ and apparent irregularity of $\pi$ as continued fraction 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 [15]: for each $k = 1, 2, 3, \ldots$

\[
\text{Prob}\{a_n = k\} = -\log_2 \left(1 - \frac{1}{(k+1)^2}\right)
\]

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

In Figure 20 we show an histogram of the first 100 million terms\(^7\) 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 the Figure 21 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 7.1.** With probability one, such a mod four random walk on the simple continued fraction coefficients of a real number is asymptotic to a line making a positive angle with the $x$-axis of:

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

**Proof.** The result comes by summing the expected Gauss-Kuzmin probabilities of each step being taken as given by (7).

This is illustrated in Figure 21(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 and continued fractions of numbers as horizontal matrix plots of the form already used in Figure 9. In Figure 22 we show six pairs of million terms digits strings and their corresponding fractions. In some cases both look random, in others one or the other does.

---

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

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

(c) Koch snowflake from 100,000 digits of the Thue-Morse sequence $T_M_2$ with rotating angle $\pi/3$. $d_{box} = 1.353$.

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

Figure 19: 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 20: Expected values of the Gauss-Kuzmin distribution of (7) 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 π/3.  
(d) A 100 step turtle walk on the continued fraction of e modulo 2 with rotating angle π/3.

Figure 21: Simple continued fraction based uniform walks on π and e.
Figure 22: Million step comparisons of base-4 digits and fractions. Row 1: $\alpha_{2,3}$ (base 6) and $C_4$. Row 2: $e$ and $\pi$. Row 3: $Q_1$ and pseudorandoms; as listed from top left to bottom right.
In conclusion, we have only tapped the surface of what is becoming possible in a period in which data—five hundred million terms of the continued fraction or five trillion bits of $\pi$, full genomes and much more—can be downloaded from the internet, then rendered—and visually mined—with fair rapidity.

8 Appendix: Selected numerical constants

In what follows, we denote $x := 0.a_1a_2a_3a_4\ldots$ the base-$b$ expansion of the number $x$, that is, $x = \sum_{k=1}^{\infty} \frac{a_k}{b^k}$. Base 10 is denoted without a subindex.

Archimedes constant (transcendental):

$$\pi := 2 \int_{-1}^{1} \sqrt{1 - x^2} \, dx$$

$$= \sum_{k=0}^{\infty} \frac{1}{16^k} \left( \frac{4}{8k + 1} - \frac{2}{8k + 4} - \frac{1}{8k + 5} - \frac{1}{8k + 6} \right)$$

$$= 3.1415926535\ldots$$

Catalan’s constant (irrational?):

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

Champernowne numbers (irrational and normal):

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

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

$$C_4 = 0.41231011121320212223\ldots$$

Copeland-Erdős constants (irrational and normal):

$$CE(b) := \sum_{k=1}^{\infty} p_k b^{-(k+\sum_{m=1}^{k} \log_2 p_m)}, \text{ where } p_k \text{ is the } k^{\text{th}} \text{ prime number.}$$

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

$$CE(2) = 0.21011101111\ldots$$
**Exponential constant** (transcendental):
\[ e := \sum_{k=0}^{\infty} \frac{1}{k!} = 2.7182818284 \ldots \] (12)

**Erdös-Borwein constants** (irrational):
\[ EB(b) := \sum_{k=1}^{\infty} \frac{1}{b^k - 1}. \] (13)
\[ EB(2) = 1.6066951524 \ldots = 0.4212311001 \ldots \]

**Euler-Mascheroni constant** (irrational?):
\[ \gamma := \lim_{m \to \infty} \left( \sum_{k=1}^{m} \frac{1}{k} - \log m \right) = 0.5772156649 \ldots \] (14)

**Fibonacci constant** (irrational?):
\[ \mathcal{F} := \sum_{k=1}^{\infty} F_k 10^{-(1+k+\sum_{m=1}^{k} \lfloor \log_{10} F_m \rfloor)} \text{, where } F_k = \frac{(1+\sqrt{5})^k - (1-\sqrt{5})^k}{\sqrt{5}}, \] (15)
\[ = 0.011235813213455 \ldots \]

**Liouville number** (irrational, not \( p \)-automatic):
\[ \lambda_2 := \sum_{k=1}^{\infty} \frac{1}{2^{(\lambda(k)+1)/2}} \text{ where } \lambda(k) := (-1)^{\Omega(k)} \] (16)
\[ \text{where } \Omega(k) \text{ counts prime factors of } k. \]

**Logarithmic constant** (transcendental):
\[ \log 2 := \sum_{k=1}^{\infty} \frac{1}{k2^k} = 0.6931471806 \ldots \] (17)

**Riemann zeta function** (transcendental for \( n \) even, irrational for \( n = 3 \)):
\[ \zeta(s) := \sum_{k=1}^{\infty} \frac{1}{k^s}, \] (18)
whence

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

\[ \zeta(2n) = (-1)^{n+1} \frac{(2\pi)^{2n}}{2(2n)!} B_{2n} \quad \text{(in terms of Bernoulli numbers)} \]

\[ \zeta(3) = \text{Apéry’s constant} = \frac{5}{2} \sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k^3 \binom{2k}{k}} = 1.2020569031 \ldots \]

**Stoneham constants** (normal or abnormal irrationals):

\[ \alpha_{b,c} := \sum_{k=1}^{\infty} \frac{1}{b^c c^k} \quad (19) \]

- \( \alpha_{2,3} = 0.0418836808 \ldots = 0.40022232032 \ldots = 0.0130140430003334 \ldots \)
- \( \alpha_{4,3} = 0.0052087571 \ldots = 0.400011111111301 \ldots = 0.001040304134350213000 \ldots \)
- \( \alpha_{3,2} = 0.0586610287 \ldots = 0.30011202021212121 \ldots = 0.0204005200030544000002 \ldots \)
- \( \alpha_{3,5} = 0.008230452 \ldots = 0.30000012101210121 \ldots = 0.15002b00000061d2 \ldots \)

**Thue-Morse constant** (transcendental, 2-automatic hence abnormal):

\[ \mathcal{T} \mathcal{M}_2 := \sum_{k=1}^{\infty} \frac{1}{2^{t(n)}} \quad \text{where } t(0) = 0, \text{ while } t(2n) = t(n) \text{ and } t(2n+1) = 1 - t(n). \quad (20) \]

\[ = 0.4124540336 \ldots = 0.20110100110010101100110101001 \ldots \]

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

\[ \mathcal{P} := \sum_{k=0}^{\infty} \frac{8^{2k}}{2^{2k+2} - 1} = 0.8507361882 \ldots = 0.21101100111001001 \ldots \quad (21) \]

**Acknowledgements**

The authors would like to thank Dr David Allingham for his kind help in running the algorithms for plotting the ‘big walks’ on \( \pi \), Michael Rose for his assistance with strong normality computations, 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


