Champernowne's Number, Strong Normality, and the X Chromosome

by Adrian Belshaw and Peter Borwein

ABSTRACT.
Champernowne's number is the best-known example of a normal number, but its digits are far from random. The sequence of nucleotides in the human X-chromosome appears non-random in a similar way. We give a new test of pseudorandomness, strong normality, based on the law of the iterated logarithm. Almost all numbers are strongly normal, and we show that a strongly normal number must necessarily be normal. However, Champernowne's number fails to be strongly normal.

This paper is dedicated to Jon Borwein in celebration of his 60th birthday.

1. NORMALITY

We can write a number α in any positive integer base r as a sum of powers of the base:

$$\alpha = \sum_{j=-d}^{\infty} a_j r^{-j}.$$

The standard "decimal" notation is

$$\alpha = a_{-d}\ a_{-(d-1)}\ \ldots a_0\ .\ a_1\ a_2 \ldots$$

In either case, the sequence of digits $\{a_j\}$ the representation of α in the base r, and this representation is unique unless α is rational, in which case α may have two representations. (For example, in the base 10, $0.1 = 0.0999\ldots$.)

We call a sequence $\{a_j\}$ of digits a string. The string may be finite or infinite; we will call a finite string of t digits a t-string. An infinite string beginning in a specified position we will call a tail.

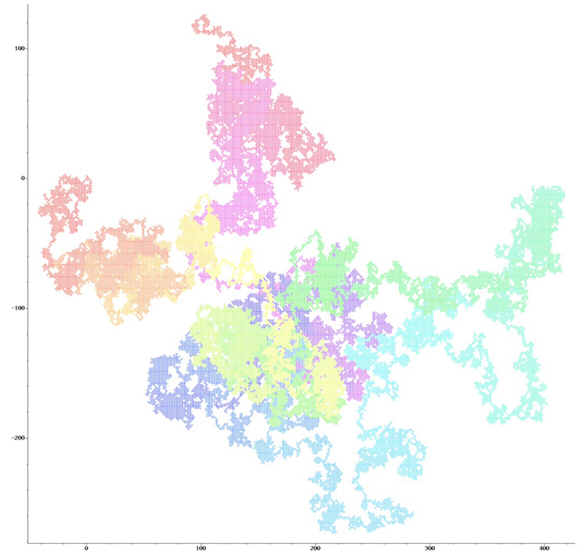

FIGURE 1
A walk on 10^6 digits of π

A number α is *simply normal* in the base r if every 1-string in its expansion in the base r occurs with an asymptotic frequency approaching $1/r$. That is, given the expansion $\{a_j\}$ of α in the base r, and letting $m_k(n)$ be the number of times that $a_j = k$ for $j \leq n$, we have

$$\lim_{n \to \infty} \frac{m_k(n)}{n} = \frac{1}{r}$$

for each $k \in \{0, 1, \ldots, r-1\}$. This is Borel's original definition [5].

A number is *normal* in the base r if every t-string in its base r expansion occurs with a frequency approaching r^{-t}. This is equivalent to Borel's definition [15]. Equivalently again, a number is normal in the base r if it is simply normal in the base r^t for every positive integer t.

A number is *absolutely normal* if it is normal in every base. Borel [5] showed that almost every real number is absolutely normal.

In 1933, Champernowne [8] produced the first concrete construction of a normal number: the Champernowne number is

$$\gamma_{10} = .1\ 2\ 3\ 4\ 5\ 6\ 7\ 8\ 9\ 10\ 11\ 12\ 13\ 14\ 15\ldots.$$

The number is written in the base 10, and its digits are obtained by concatenating the natural numbers written in the base 10. This number is probably the best-known example of a normal number.

Generally, the base r Champernowne number is formed by concatenating the integers 1, 2, 3, ... in the base r. For example, the base 2 Champernowne number is written in the base 2 as

$$\gamma_2 = .1\ 10\ 11\ 100\ 101\ \ldots$$

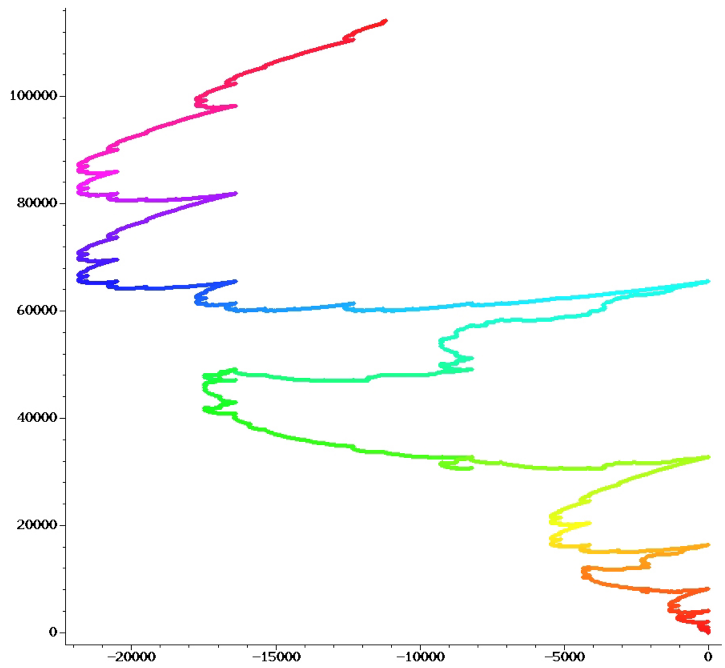

FIGURE 2
A walk on 10^6 binary digits of the base 2 Champernowne number

For any r, the base r Champernowne number is normal in the base r. However, the question of its normality in any other base (not a power of r) is open. For example, it is not known whether the base 10 Champernowne number is normal in the base 2.

It is even more disturbing that no number has been proven absolutely normal, that is, normal in every base. Every normality proof so far is only valid in one base (and its powers), and depends on a more or less artificial construction.

Most fundamental irrational constants, like $\sqrt{2}$, $\log 2$, π, and e appear to be without pattern in the digits, and statistical tests done to date are consistent with the hypothesis that they are normal. (See, for example, Kanada on π [10] and Beyer, Metropolis and Neergard on irrational square roots [4].) However, there is no proof of the normality of any of these constants.

There is an extensive literature on normality in the sense of Borel. Introductions to the literature may be found in [3] and [7].

2. Walks on the Digits of Numbers and on Chromosomes

If we compare a walk on the digits of Champernowne's number (Figure 2) with a walk on the digits of a number believed to be pseudorandom, like π (Figure 1), it will be seen that Champernowne's number is highly patterned. It is interesting that a walk on the nucleotides of a chromosome, such as the the X chromosome (Figure 3) produces a similarly patterned image.

A random walk on a million digits is expected to stay within roughly a thousand units of the origin, and this will be seen to hold for the walks on the digits of π and on the Liouville function values. On the other hand, the walks on the digits of Champernowne's number and on the X chromosome are highly patterned, and travel much farther than would be expected of a random walk.

The walk on the Liouville λ function has some properties similar to those of a random walk, and some patterned properties. The walk does moves away from the origin like $\sqrt{n}$, but it does not seem to move randomly near the origin. In fact, the positive values of λ first outweigh the negative values when $n = 906180359$ [11], which is not at all typical of a random walk.

The walks are generated on a binary sequence by converting each 0 in the sequence to -1, and then using digit pairs $(\pm 1, \pm 1)$ to walk $(\pm 1, \pm 1)$ in the plane. The shading indicates the distance travelled along the walk. The values of the Liouville λ function are already ± 1.

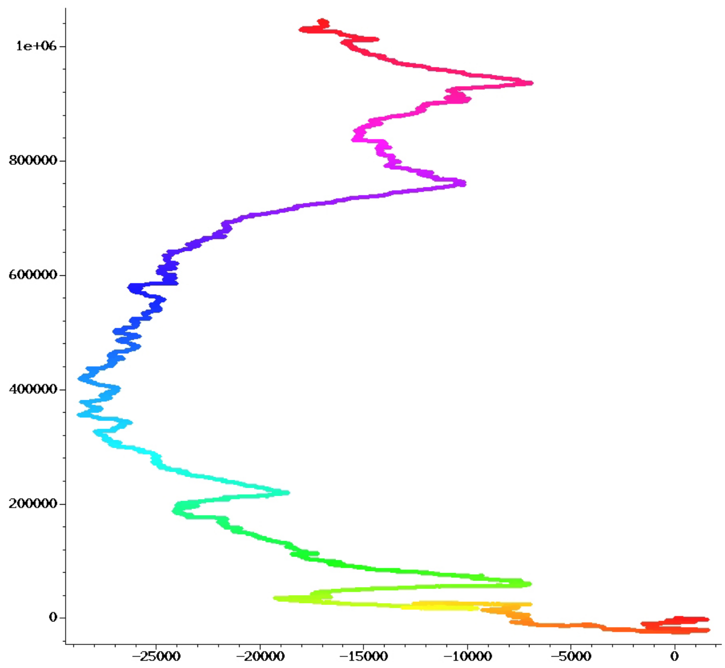

FIGURE 3
A walk on the nucleotides of the human X chromosome

There are four nucleotides in the X chromosome sequence , and each of the four is assigned one of the values $(\pm 1, \pm 1)$ to create a walk on the nucleotides. The nucleotide sequence is available on the UCSC Genome Browser [13]

3. Strong Normality

Mauduit and Sárközy [14] have shown that the digits of the base 2 Champernowne's number γ_2 fail two tests of randomness. Dodge and Melfi [9] compared values of an autocorrelation function for Champernowne's number

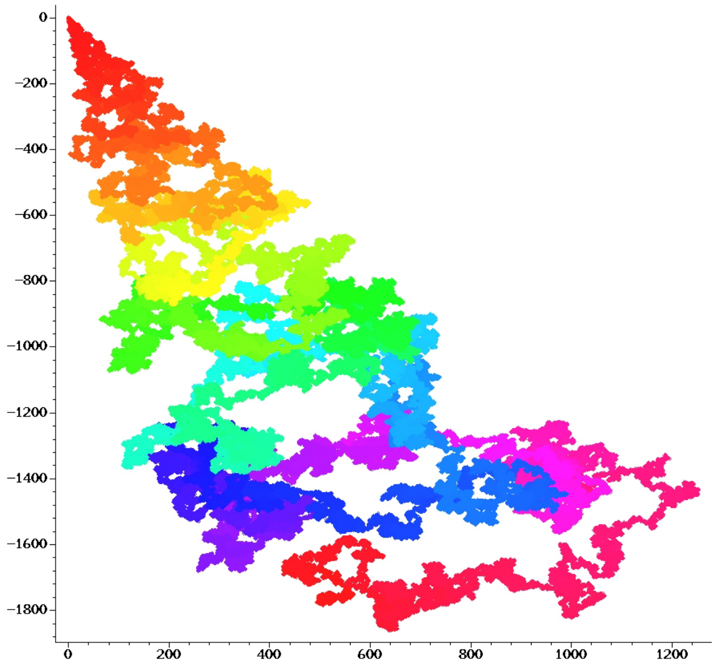

FIGURE 4
A walk on 10^6 values of the Liouville λ function

and π, and found that π had the expected pseudorandom properties but that Champernowne's number did not.

Here we provide another test of pseudorandomness, and show that it must be passed by almost all numbers. Our test is is a simple one, in the spirit of Borel's test of normality, and Champernowne's number will be seen to fail the test for almost trivial reasons.

If the digits of a real number α are chosen at random in the base r, the asymptotic frequency $m_k(n)/n$ of each digit approaches $1/r$ with probability 1. However, the **discrepancy** $m_k(n) - n/r$ does not approach any limit, but fluctuates with an expected value equal to the standard deviation $\sqrt{(r-1)n}/r$.

Kolmogorov's law of the iterated logarithm allows us to make a precise statement about the discrepancy of a random number. We use this to define our criterion.

DEFINITION 3.1. For real α, and $m_k(n)$ as above, α is **simply strongly normal** in the base r if for each $k \in \{0, \ldots, r-1\}$

$$\limsup_{n\to\infty} \frac{m_k(n) - \dfrac{n}{r}}{\dfrac{\sqrt{r-1}}{r}\sqrt{2n\log\log n}} = 1$$

and

$$\liminf_{n\to\infty} \frac{m_k(n) - \dfrac{n}{r}}{\dfrac{\sqrt{r-1}}{r}\sqrt{2n\log\log n}} = -1 \ .$$

We make two further definitions analogous to the definitions of normality and absolute normality.

DEFINITION 3.2.
A number is **strongly normal** in the base r if it is simply strongly normal in each of the bases r^j, $j = 1, 2, 3, \ldots$.

DEFINITION 3.3. A number is **absolutely strongly normal** if it is strongly normal in every base.

These definitions of strong normality are sharper than those given by one of the authors in [2].

4. ALMOST ALL NUMBERS ARE STRONGLY NORMAL

In the following, we take the bases r to be integers no less than 2.

THEOREM 4.1.
Almost all numbers are simply strongly normal in any base r.

Proof.
Without loss of generality, we consider numbers in the interval $[0, 1)$ and fix the base $r \geq 2$. We take Lebesgue measure to be our probability measure. For any k, $0 \leq k \leq r-1$, the ith digit of a randomly chosen number is

k with probability r^{-1}. For $i \neq j$, the ith and jth digits are both k with probability r^{-2}, so the digits are pairwise independent.

We define the sequence of random variables X_j by

$$X_j = \sqrt{r-1}$$

if the jth digit is k, with probability $\frac{1}{r}$, and

$$X_j = -\frac{1}{\sqrt{r-1}}$$

otherwise, with probability $\frac{r-1}{r}$.

Then the X_j form a sequence of independent identically distributed random variables with mean 0 and variance 1. Put

$$S_n = \sum_{j=1}^{n} X_j \ .$$

By the law of the iterated logarithm (see, for example, [12]), with probability 1,

$$\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = 1 \ ,$$

and

$$\liminf_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = -1 \ .$$

Now we note that, if $m_k(n)$ is the number of occurrences of the digit k in the first n digits of our random number, then

$$S_n = m_k(n)\sqrt{r-1} - \frac{n - m_k(n)}{\sqrt{r-1}} \ .$$

Substituting this expression for S_n in the limits immediately above shows that the random number satisfies the definition of strong normality with probability 1.

This is easily extended.

COROLLARY 4.2. *Almost all numbers are strongly normal in any base r.*

Proof. By the theorem, the set of numbers in $[0, 1)$ which fail to be simply strongly normal in the base r^j is of measure zero, for each j. The countable union of these sets of measure zero is also of measure zero. Therefore the set of numbers simply strongly normal in every base r^j is of measure 1.

The following corollary is proved in the same way as the last.

COROLLARY 4.3. *Almost all numbers are absolutely strongly normal.*

The results for $[0, 1)$ are extended to $\mathbf{R}$ in the same way.

5. Champernowne's Number is Not Strongly Normal

We begin by examining the digits of Champernowne's number in the base 2,

$$\gamma_2 = .1\ 10\ 11\ 100\ 101\ldots\ .$$

When we concatenate the integers written in base 2, we see that there are 2^{n-1} integers of n digits. As we count from 2^n to $2^{n+1}-1$, we note that every integer begins with the digit 1, but that every possible selection of zeros and ones occurs exactly once in the other digits, so that apart from the excess of initial ones there are equally many zeros and ones in the non-initial digits.

Thus, $m_1(N) > N/2$, and γ_2 fails the $\liminf$ criterion for strong normality. It can be shown that γ_2 fails the $\limsup$ criterion as well.

We thus have:

THEOREM 5.1. *The base 2 Champernowne number is not strongly normal in the base 2.*

The theorem can be generalized to every Champernowne number, since there is a shortage of zeros in the base r representation of the base r Champernowne number. Each base r Champernowne number fails to be strongly normal in the base r.

6. Strongly Normal Numbers are Normal

Our definition of strong normality is strictly more stringent than Borel's definition of normality:

THEOREM 6.1.
If a number α is simply strongly normal in the base r, then α is simply normal in the base r.

Proof. It will suffice to show that if a number is not simply normal, then it cannot be simply strongly normal.

Let $m_k(n)$ be the number of occurrences of the 1-string k in the first n digits of the expansion of α in the base r, and suppose that α is not simply normal in the base r. This implies that for some k

$$\lim_{n\to\infty} \frac{rm_k(n)}{n} \neq 1.$$

Then there is some $Q > 1$ and infinitely many n_i such that either

$$rm_k(n_i) > Qn_i$$

or

$$rm_k(n_i) < \frac{n_i}{Q}.$$

If infinitely many n_i satisfy the former condition, then for these n_i,

$$m_k(n_i) - \frac{n_i}{r} > Q\frac{n_i}{r} - \frac{n_i}{r} = n_i P$$

where P is a positive constant.

Then for any $R > 0$,

$$\limsup_{n\to\infty} R\frac{m_k(n) - \frac{n}{r}}{\sqrt{2n\log\log n}} \geq \limsup_{n\to\infty} R\frac{nP}{\sqrt{2n\log\log n}} = \infty,$$

so α is not simply strongly normal.

On the other hand, if infinitely many n_i satisfy the latter condition, then for these n_i,

$$\frac{n_i}{r} - m_k(n_i) > \frac{n_i}{r} - \frac{n_i}{Qr} = n_i P,$$

and once again the constant P is positive. Now

$$\liminf_{n\to\infty} \frac{m_k(n) - \frac{n}{r}}{\sqrt{2n\log\log n}} = -\limsup_{n\to\infty} \frac{\frac{n}{r} - m_k(n)}{\sqrt{2n\log\log n}}$$

and so, in this case also, α fails to be simply strongly normal.

The general result is an immediate corollary.

COROLLARY 6.2.
If α is strongly normal in the base r, then α is normal in the base r.

7. No Rational Number is Simply Strongly Normal

A rational number cannot be normal, but it is simply normal to the base r if each 1-string occurs the same number of times in the repeating string in the tail. However, such a number is not simply strongly normal.

If α is rational and simply normal in the base r, then if we restrict ourselves to the first n digits in the repeating tail of the expansion, the frequency of any 1-string k is exactly n/r whenever n is a multiple of the length of the repeating string. The excess of occurences of k can never exceed the constant number of times k occurs in the repeating string. Therefore, with $m_k(n)$ defined as before,

$$\limsup_{n\to\infty}\left(m_k(n)-\frac{n}{r}\right)=Q,$$

with Q a constant due in part to the initial non-repeating block, and in part to the maximum excess in the tail.

But

$$\limsup_{n\to\infty}\frac{Q}{\sqrt{2n\log\log n}}=0\ ,$$

so α does not satisfy the first part of the definition of strong normality.

8. Further Questions

We have not produced an example of a strongly normal number. Can such a number be constructed explicitly ?

We can conjecture that such naturally occurring constants as the real irrational numbers π, e, $\sqrt{2}$, and $\log 2$ are strongly normal, since they appear on the evidence to be as though random.

On the other hand, we speculate that the binary Liouville number, created in the obvious way from the λ function values, is normal but not strongly normal.

Bailey and Crandall [1] proved normality for a large class of numbers with Borwein-Bailey-Plouffe formulas

$$\alpha=\sum_{j=0}^{\infty}\frac{p(j)}{r^j q(j)}\ ,$$

where p and q are polynomials. This class of numbers may be a good place to look for a first proof of strong normality.

Finally, one can foresee non-random numbers passing our test. Consider, for example, a number with discrepancy cycling in a predictable way between the $\limsup$ and the $\liminf$ of our definition. We want, in the end, a deeper test passed only by those numbers whose digits generate typical Brownian motions.

Acknowledgments: Many thanks are due to Stephen Choi for his comments on the earlier ideas in [2] . We also give many thanks to Richard Lockhart for his help with ideas in probability.

REFERENCES

[1] D. H. Bailey and R. E. Crandall, "Random Generators and Normal Numbers," *Experiment. Math. 11*, no. 4, 527-546, 2003.

[2] A. Belshaw, *On the Normality of Numbers*, M.Sc. thesis, Simon Fraser University, Burnaby, BC, 2005.

[3] L. Berggren, J. Borwein and P. Borwein, *Pi : a Source Book*, 3rd ed., Springer-Verlag, New York, 2004.

[4] W. A. Beyer, N. Metropolis and J. R. Neergaard, “Statistical study of digits of some square roots of integers in various bases,” *Math. Comp. 24*, 455 - 473, 1970.

[5] E. Borel, “Les probabilités dénombrables et leurs applications arithmétiques,” *Supplemento ai rendiconti del Circolo Matematico di Palermo 27*, 247 - 271, 1909.

[6] E. Borel, “Sur les chiffres décimaux de $\sqrt{2}$ et divers problèmes de probabilités en chaîne,” *C. R. Acad. Sci. Paris 230*, 591- 593, 1950.

[7] J. Borwein and D. Bailey, *Mathematics by Experiment*, A K Peters Ltd., Natick, MA, 2004.

[8] D. G. Champernowne, “The construction of decimals normal in the scale of ten,” *Journal of the London Mathematical Society 3*, 254- 260, 1933.

[9] Y. Dodge and G. Melfi, "On the Reliability of Random Number Generators," http : //pictor.math.uqam.ca/ plouffe/articles/reliability.pdf

[10] Y. Kanada, “Vectorization of Multiple-Precision Arithmetic Program and 201,326,395 Decimal Digits of π Calculation,” *Supercomputing 88 (II, Science and Applications)*, 1988.

[11] R. S. Lehman, "On Liouville’s Function," *Mathematical Computation 14*, 311-320, 1960.

[12] R. G. Laha and V. K. Rohatgi, *Probability Theory*, John Wiley and Sons, Inc.,, New York, NY , 1979.

[13] UCSC Genome Browser, http : //hgdownload.cse.ucsc.edu/goldenPath/hg19/chromosomes/.

[14] Christian Mauduit and András Sárközy, “On Finite Pseudorandom Binary Sequences II. The Champernowne, Rudin-Shapiro, and Thue-Morse Sequences, A Further Construction," *J. Num. Th. 73*, 256-276, 1998.

[15] D. D. Wall, *Normal Numbers*, Ph.D. thesis, University of California, Berkeley, CA, 1949.

Adrian Belshaw

Department of Mathematics and Statistics
Capilano University
PO Box 1609
Sechelt, BC, V0N 3A0
Canada
e-mail: abelshaw@capilanou.ca

Peter Borwein

Department of Mathematics
Simon Fraser University
8888 University Drive
Burnaby, BC, V5A 1S6
Canada
e-mail: pborwein@cecm.sfu.ca