Strong Normality of Numbers

Adrian Belshaw Peter Borwein

"...the problem of knowing whether or not the digits of a number like $\sqrt{2}$ satisfy **all** the laws one could state for randomly chosen digits, still seems ... to be one of the most outstanding questions facing mathematicians."

Émile Borel [Borel 1950]

Abstract

Champernowne's number is the best-known example of a normal number, but its digits are highly patterned. We present graphic evidence of the patterning and review some relevant results in normality. We propose a strong normality criterion based on the variance of the normal approximation to a binomial distribution. Allmost all numbers pass the new test but Champernowne's number fails to be strongly normal.

1 Introduction

How can one decide whether the digits of a number behave in a random manner?

In 1909 Émile Borel introduced an analytic test for randomness. His notion of normality of numbers has posed impossibly difficult questions for mathematicians ever since. (It has not yet been shown that the decimal expansion of π, $\sqrt{2}$ or any other "natural" irrational number has infinitely many zeros, though it is almost certainly true.)

In this article we review the definition of normality and give some of the outstanding historical results. We then note that some numbers pass Borel's normality test even though they show clearly non-random behaviour in the digits. Champernowne's number is of this variety, and while it is provably normal it neither "looks" nor "behaves" like a random number. Similarly, human chromosomes, thought of as large base four numbers, behave more like Champernowne's number than like random numbers.

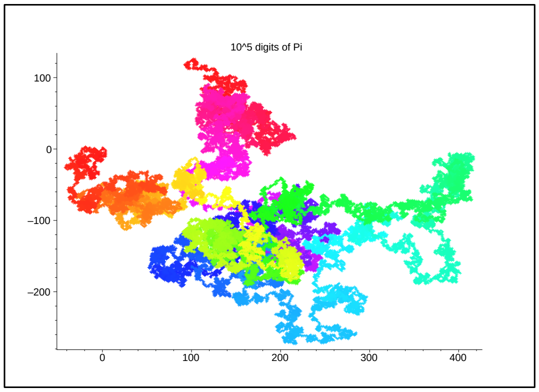

Figure 1: The first 10^5 binary digits of π as a "random walk."

In consequence we propose a stronger test for normality. Almost all numbers pass the new test, and every number that passes the test is normal in Borel's sense. However, not all normal numbers pass the new test; in particular, we show that Champernowne's number fails to meet the stronger criterion.

First, though, we present the graphic images which motivated this study.

2 Walks on the Digits of Numbers and on Chromosomes

If the binary digits of a number are "like random," then we expect the walk generated on the digits to look like a random walk generated by the toss of a coin.

We show walks generated by the digits of π, e, $\sqrt{2}$, and the base 2 Champernowne number. For comparison, we give a random walk generated by Maple.

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 walks on the digits of π, e and $\sqrt{2}$ look like random walks, although none of these numbers has been proven normal. On the other hand, the base 2 Champernowne number is known to be normal in the base 2, but its digits look far from random. Will this phenomenon disappear if we look at more digits? The answer is no – there are always more ones than zeros in the expansion, and this preponderance is unbounded even though the

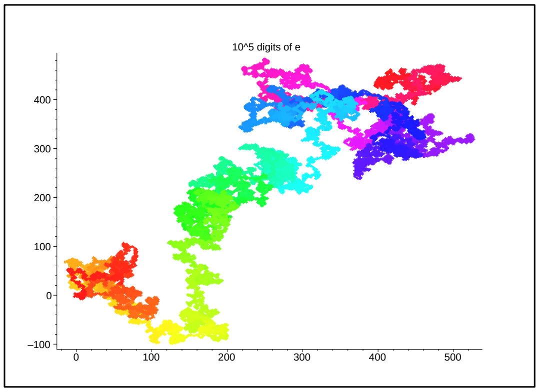

Figure 2: The first 10^5 binary digits of e as a "random walk."

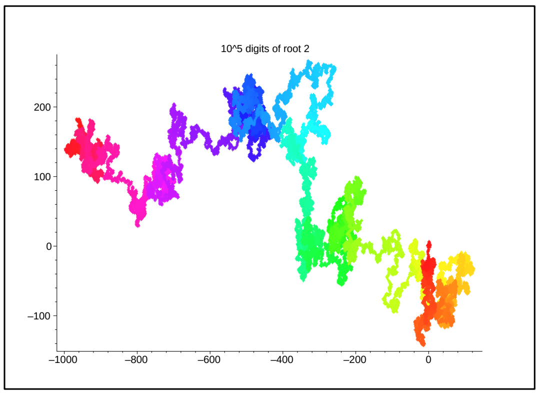

Figure 3: The first 10^5 binary digits of $\sqrt{2}$ as a "random walk."

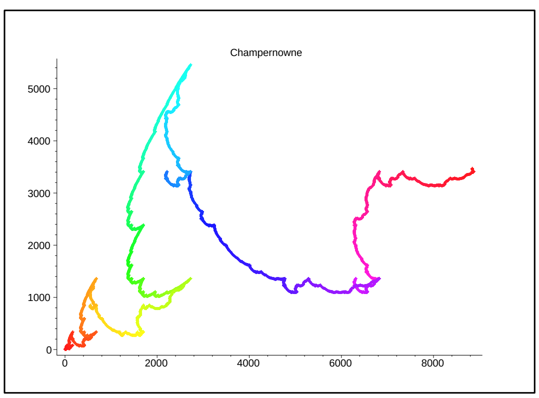

Figure 4: The first 10^5 binary digits of Champernowne's number in binary as a "random walk."

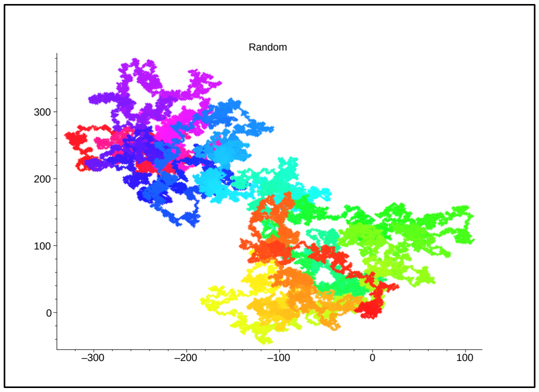

Figure 5: The first 10^5 binary digits of a random "random walk."

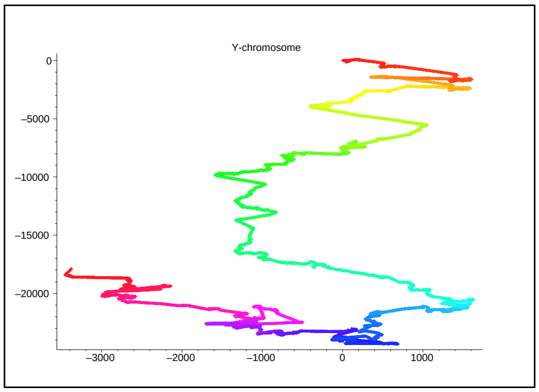

Figure 6: The first 10^5 base 4 digits of the Y-chromosome as a "random walk."

frequencies of the two digits approach equality.

The chromosomes of the human genome are sequences of nucleotides roughly 1 billion long. There are four nucleotides, so it seems natural to convert such a sequence to a sequence of base 4 digits and then make a walk on the digits.

We show walks generated by the human X and Y chromosomes.

While the chromosome sequences pass Borel-like tests of randomness on frequencies of short strings, the walks they generate are strikingly non-random, and in fact strikingly similar to the walk generated by Champernowne's number.

One might speculate that the walks look alike because both Champernowne's number and the chromosomes contain coded information – but that

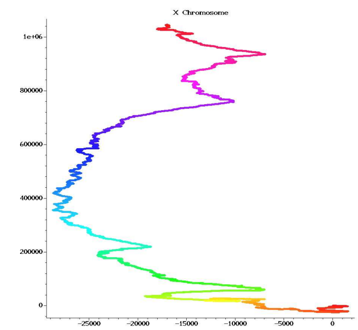

Figure 7: The whole X-chromosome as a "random walk."

goes far beyond the scope of this article.

3 Borel's Normality Criterion

Recall that 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, we call 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 will use the term *string* to denote a sequence $\{a_j\}$ of digits. The string may be finite or infinite; we will call a finite string of t digits a t-string.

A finite string of digits beginning in some specified position we will refer to as a *block*. An infinite string beginning in a specified position we will call a *tail*.

Since we are interested in the asymptotic frequencies of digit strings, we will work in $\mathbb{R}/\mathbb{Z}$ by discarding all digits to the left of the decimal point in the representation of α. If the real numbers α and β have the same fractional part, we write

$$\alpha \equiv \beta \pmod 1.$$

A number α is *simply normal* to the base r if every 1-string in its expansion to the base r occurs with a frequency approaching $1/r$. That is, given the expansion $\{a_j\}$ of α to 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 [Borel 1909].

The number $1/3$ has the binary representation $0.010101\ldots$. Since the digits 0 and 1 occur equally frequently, this number is simply normal in the base 2. One wants to exclude such repeating patterns.

A number is *normal* to the base r if every t-string in its base r expansion occurs with a frequency approaching r^{-t}. In other words, it is simply normal to the base r^t for every positive integer t. This differs from Borel's original definition; it was used as though equivalent to Borel's for some time, but the equivalence was first proved by Wall in 1949 [Wall 1949].

A number is *absolutely normal* if it is normal to every base.

It's easy to show that no rational number is normal in any base. In his original paper, Borel [Borel 1909] proved that almost every real number is normal in every base. That is, the set of all numbers which are not absolutely normal has Lebesgue measure zero. We will use Borel's method of proof below, when we show that almost every number meets our stronger normality criterion.

Borel gave no example of a normal number. It wasn't until 1917, eight years after Borel's paper, that Sierpiński produced the first example [Sierpiński 1917]. (Lebesgue apparently constructed a normal number in 1909, but didn't publish his work until 1917 [Lebesgue 1917]; the papers by Sierpiński and Lebesgue appeared side by side in the same journal.) But neither construction produced a tangible string of digits.

Finally, in 1933, Champernowne [Champernowne 1933] produced an easy and concrete construction of a normal number: the Champernowne number is

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

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.

For full proofs of these basic results, we refer the reader to Niven [Niven 1956]. Niven also gives an easy proof of Wall's beautiful result that a number α is normal to the base r if and only if the sequence $\{r^j\alpha\}$ is uniformly distributed modulo 1 [Wall 1949].

A concatenated number is formed in the base r by taking a sequence of integers a_1, a_2, a_3, ... and writing the integers in the base r to the right of the decimal point:

$$\alpha = .a_1\ a_2\ a_3\ \ldots$$

We have seen that Champernowne's number is formed by concatenating the positive integers in their natural sequence in the base 10. More generally, a 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

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

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.

Champernowne made the conjecture that the number obtained by concatenating the primes, $\alpha = .2\ 3\ 5\ 7\ 11\ 13\ldots$, was normal in the base 10. Copeland and Erdös [Copeland and Erdös 1946] proved this in 1946 , as a corollary of a more general result : if an increasing sequence of positive integers $\{a_j\}$ has the property that, for large enough N, the number of a_j less than N is greater than N^θ for any $\theta < 1$, then concatenating the a_j written in any base gives a normal number in that base.

The number formed by concatenating the primes is commonly called the Copeland-Erdös number.

Copeland and Erdös conjectured that, if $p(x)$ is a polynomial in x taking positive integer values whenever x is a positive integer, then the number

$$\alpha = .p(1\ )p(2)\ p(3)\ \ldots$$

formed by concatenating the base 10 values of the polynomial at positive integer values of x is normal in the base 10. This result was proved by Davenport and Erdös in 1952 [Davenport and Erdös 1952].

Various other classes of "artificial" numbers have been shown to be normal; for an overview, we refer the reader to Berggren, Borwein and Borwein [Berggren, Borwein and Borwein 2004] or to Borwein and Bailey [Borwein and Bailey 2004].

And other classes of irrational numbers have been shown not to be normal. It's easy to prove the non-normality of Liouville numbers , for example, and to construct similar irrationals like this one in the base 2:

$$\alpha = .10\ 100\ 1000\ \ldots$$

Martin gave a construction of a number not normal to any base [Martin 2001].

But the central mystery remains intact. Most familiar 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 [Kanada 1988] and Beyer, Metropolis and Neergard on irrational square roots [Beyer, Metropolis and Neergaard 1970].)

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 the number being constructed and written in that base.

4 Binomial Normality

Borel's original definition of normality [Borel 1909] had the advantage of great simplicity. None of the current profusion of concatenated monsters had been studied at the time, so there was no need for a stronger definition.

However, one would like a test or a set of tests to eliminate exactly those numbers that do not behave in the limit in every way as a binomially random number defined informally as follows: let each of the numbers $a_1, a_2, a_3, \ldots$ be chosen with equal probability from the set of integers $\{0, 1, \ldots, r-1\}$, and let $\alpha = .a_1\ a_2\ a_3 \ldots$ be the number represented in the base r by the concatenation of the digits a_j. Then α is a *binomially random* number in the base r.

This leads to another informal definition: A number is *binomially normal* to the base r if it passes every reasonable asymptotic test on the frequencies of the digits that would be passed with probability 1 by a binomially random number.

We leave the definition in this intuitive form, to guide us in the search for normality criteria. It's easy to devise an uncountable set of "unreasonable" asymptotic tests, each of them passed with probability 1 by a random number, in such a way that no number can pass all the tests. Our goal here is simply to begin to take up Borel's challenge in deciding "whether or not the digits of a number like $\sqrt{2}$ satisfy **all** the laws one could state for randomly chosen digits."

Borel's test of normality is passed with probability 1 by a binomially random number [Borel 1909], so it would certainly be passed by a binomially normal number as well. However, in this article we give an asymptotic test that is failed by some normal numbers, but passed with probability 1 by a binomially random number.

5 Strong Normality

In this section, we define strong normality, and in the following sections we prove that almost all numbers are strongly normal and that Champernowne's number is not strongly normal.

The definition is motivated as follows: let a number α be represented in the base r, and let $m_k(n)$ represent the number of occurrences of the kth 1-string in the first n digits. Then α is simply normal to the base r if

$$\frac{rm_k(n)}{n} \to 1$$

as $n \to \infty$, for each $k \in \{0, 1, \ldots, r-1\}$. But if a number is binomially random, then the discrepancy $m_k(n) - n/r$ should fluctuate, with its expected value equal to the standard deviation $\sqrt{(r-1)n}/r$.

The following definition makes this idea precise:

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

$$\limsup_{n\to\infty} \frac{(m_k(n) - n/r)^2}{n^{1+\varepsilon}} = 0$$

and

$$\limsup_{n\to\infty} \frac{(m_k(n) - n/r)^2}{n^{1-\varepsilon}} = \infty$$

for any $\varepsilon > 0$.

It follows from the definition that a number will be simply strongly normal if the maximum discrepancy grows like $\sqrt{n}$. This definition can certainly be tightened, but it is easy to apply and it will suffice for our present purpose.

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

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

Definition 3. *A number is* **absolutely strongly normal** *if it is strongly normal to every base.*

6 Almost All Numbers are Strongly Normal

The proof that almost all numbers are strongly normal is based on Borel's original proof [Borel 1909] that almost all numbers are normal.

Theorem 1. *Almost all numbers are simply strongly normal to any integer base $r > 1$.*

Proof. Let α be a binomially random number in the base r, so that the nth digit of the representation of α in the base r is, with equal probability, randomly chosen from the numbers 0, 1, 2, …, $r - 1$. Let $m_k(n)$ be the number of occurrences of the 1-string k in the first n digits of α.

Then $m_k(n)$ is a random variable of binomial distribution with mean n/r and variance $\frac{n(r-1)}{r^2}$. As $n \to \infty$, the random variable approaches a normal distribution with the same mean and variance.

The probability that

$$\left(m_k(n) - \frac{n}{r}\right)^2 > \frac{r-1}{r^2} n^{1+\varepsilon/2}$$

is the probability that

$$\left|m_k(n) - \frac{n}{r}\right| > B\sqrt{n} n^{\varepsilon/4},$$

where $B = \sqrt{r-1}/r$, and this probability rapidly approaches zero as $n \to \infty$.

With probability 1, only finitely many $m_k(n)$ satisfy the inequality, and so with probability 1

$$\limsup_{n\to\infty} \frac{(m_k(n) - n/r)^2}{\frac{r-1}{r^2} n^{1+\varepsilon/2}} < 1.$$

We have

$$\limsup_{n\to\infty} \frac{(m_k(n) - n/r)^2}{\frac{r-1}{r^2} n^{1+\varepsilon}} = \limsup_{n\to\infty} \left(\frac{(m_k(n) - n/r)^2}{\frac{r-1}{r^2} n^{1+\varepsilon/2}}\right) \left(\frac{1}{n^{\varepsilon/2}}\right).$$

The first factor in the right hand limit is less than 1 (with probability 1), and the second factor is zero in the limit.

With probability one, this supremum limit is zero, and the first condition of strong normality is satisfied.

The same argument, word for word, but replacing $1 + \varepsilon/2$ with $1 - \varepsilon/2$ and reversing the inequalities, establishes the second condition. $\square$

As with the corresponding result for normality, this is easily extended.

Corollary 2. *Almost all numbers are strongly normal to any base r.*

Proof. By the theorem, the set of numbers in $[0, 1)$ which fail to be simply strongly normal to 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 to every base r^j is of measure 1. $\square$

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

Corollary 3. *Almost all numbers are absolutely strongly normal.*

7 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.

As we concatenate the integers from 1 to $2^k - 1$, we write the first

$$\sum_{n=1}^{k} n2^{n-1} = (k-1)2^k + 1$$

digits of γ_2. The excess of ones in the digits is

$$2^k - 1.$$

The locally greatest excess of ones occurs at the first digit of 2^k, since each power of 2 is written as a 1 followed by zeros. At this point the number of digits is $(k-1)2^k + 2$ and the excess of ones is 2^k. That is, the actual number of ones in the first $N = (k-1)2^k + 2$ digits is

$$m_1(N) = (k-2)2^{k-1} + 1 + 2^k.$$

This gives

$$m_1(N) - \frac{N}{2} = 2^{k-1}$$

and

$$\left((m_1(N) - \frac{N}{2}\right)^2 = 2^{2(k-1)}.$$

Thus, we have

$$\frac{\left((m_1(N) - \frac{N}{2}\right)^2}{\frac{1}{4}N^{1+\varepsilon}} \geq \frac{2^{2(k-1)}}{\frac{1}{4}\left((k-1)2^k\right)^{1+\varepsilon}}.$$

The limit of the right hand expression as $k \to \infty$ is infinity for any sufficiently small positive ε. Since the left hand limit is a constant multiple of the first supremum limit in our definition of strong normality, we have proved the following theorem:

Theorem 4. *The base 2 Champernowne number is not strongly normal to 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.

8 Strongly Normal Numbers are Normal

If any strongly normal number failed to be normal, then the definition of strong normality would be inappropriate. Fortunately, this does not happen.

Theorem 5. *If a number α is simply strongly normal to the base r, then α is simply normal to 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 to the base r, and suppose that α is not simply normal to 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 $R > 0$ and small ε,

$$\limsup_{n\to\infty} R\frac{(m_k(n) - \frac{n}{r})^2}{n^{1+\varepsilon}} \geq \limsup_{n\to\infty} R\frac{n^2P^2}{n^{1+\varepsilon}} = \infty,$$

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 and the rest of the argument follows. □

The general result is an immediate corollary.

Corollary 6. *If α is strongly normal to the base r, then α is normal to the base r.*

9 No Rational Number is Simply Strongly Normal

A rational number cannot be normal, but it will be 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 to 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,

$$\limsup_{n\to\infty} \left(m_k(n) - \frac{n}{r}\right)^2 = Q,$$

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

But

$$\limsup_{n\to\infty} \frac{Q}{n^{1-\varepsilon}} = 0$$

if ε is small, so α does not satisfy the second criterion of strong normality.

Simple strong normality is not enough to imply normality. As an illustration of this, consider the number

$$\alpha = .01\ 0011\ 000111\ldots,$$

a concatenation of binary strings of length $2l$ in each of which l zeros are followed by l ones. After the first $l-1$ such strings, the zeros and ones are equal in number. and the number of digits is

$$\sum_{k=1}^{l-1} 2k = 2l(l-1).$$

After the next l digits there is a locally maximal excess of $l/2$ zeros and the total number of digits is $2l^2 - l$. Thus, the greatest excess of zeros grows like the square root of the number of digits, and so does the greatest shortage of ones. It is not hard to verify that α satisfies the definition of simple strong normality to the base 2. However, α is not normal to the base 2.

10 Further Questions

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

It is natural to conjecture that such naturally occurring constants as the real algebraic irrational numbers, π, e, and $\log 2$ are strongly normal, since they appear on the evidence to be binomially normal.

It is easy to construct normal concatenated numbers which, like Champernowne's number, are not strongly normal. Do all the numbers

$$\alpha = .\ p(1)\ p(2)\ p(3)\ldots,$$

where p is a polynomial taking positive integer values at each positive integer, fail to be strongly normal? Does the Copeland-Erdös concatenation of the primes fail to be strongly normal? We conjecture that the answer is "yes" to both these questions.

NOTE Much of the material here appeared in Belshaw's M. Sc. thesis [Belshaw 2005]. Many thanks are due to Stephen Choi for his editorial comments on that manuscript.

References

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

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

[Beyer, Metropolis and Neergaard 1970] 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.

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

[Borel 1950] 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.

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

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

[Copeland and Erdös 1946] A. H.Copeland and P. Erdös, "Note on Normal Numbers," *Bulletin of the American Mathematical Society 52(1)*, 857-860, 1946.

[Davenport and Erdös 1952] H. Davenport, and P. Erdös, "Note on normal decimals," *Canadian J. Math. 4*, 58- 63, 1952.

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

[Lebesgue 1917] H. Lebesgue, "Sur Certaines Démonstrations d'Existence," *Bulletin de la Société Mathématique de France 45*, 132-144, 1917.

[Martin 2001] G. Martin, “Absolutely abnormal numbers,” *Amer. Math. Monthly 108 (8)*, 746- 754, 2001.

[Niven 1956] I. Niven, *Irrational Numbers*, The Mathematical Association of America, John Wiley and Sons, Inc., New York, N.Y., 1956.

[Sierpiński 1917] W.Sierpiński, “Démonstration Élémentaire du Théorème de M. Borel sur les Nombres Absolument Normaux et Déteermination Effective d'un Tel Nombre,” *Bulletin de la Société Mathématique de France 45*, 125-132, 1917.

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

Adrian Belshaw , Simon Fraser University, Burnaby, BC, V5A 1S6. (abelshaw@capcollege.bc.ca)

Peter Borwein, Simon Fraser University, Burnaby, BC, V5A 1S6. (pborwein@cecm.sfu.ca)