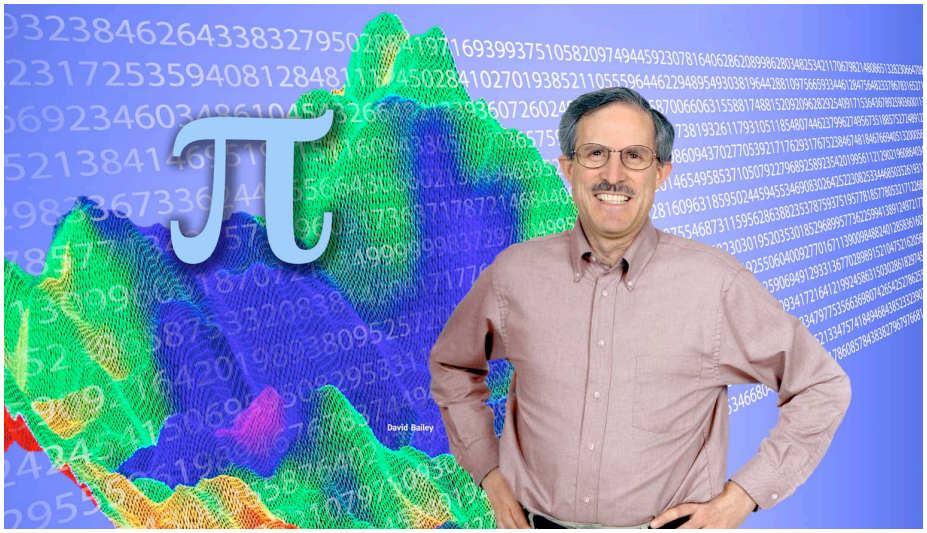
Normal Numbers

David H Bailey
Lawrence Berkeley Laboratory
http://crd.lbl.gov/~dhbailey



Normal numbers

Given an integer $b > 1$, a real number x is ***b*-normal** (or "normal base b") if every m-long string of digits in the base-b expansion of x appears with limiting frequency b^{-m}.

Using measure theory, it is easy to show that almost all reals are b-normal. In fact, almost all reals are b-normal for all integer bases $b > 1$.

These are widely believed to be b-normal, for all integer bases $b > 1$:

π = 3.1415926535…

e = 2.7182818284…

sqrt(2) = 1.4142135623…

log(2) = 0.6931471805…

Every irrational algebraic number.

But there are no normality proofs for any of these constants, not for any base b, nor are there any non-normality results.

Until recently, normality proofs were known only for contrived examples such as Champernowne's constant = 0.12345678910111213141 5… and equivalents in other bases.

A recent result for algebraic numbers

If x is algebraic of degree $d > 1$, then its binary expansion through position n must have at least $C\, n^{1/d}$ 1-bits, for all sufficiently large n and some C that depends on x.

Example: The first n binary digits of sqrt(2) must have at least sqrt(n) 1-bits. In this case, the proof is easy – it follows by noting that the 1-bit count of the product of two integers is less than or equal to the product of the 1-bit counts of the two integers.

A number of other related results are established in the paper below. These results are still a far cry from full normality.

DHB, J. M. Borwein, R. E. Crandall and C. Pomerance, "On the Binary Expansions of Algebraic Numbers," *Journal of Number Theory Bordeaux*, vol. 16 (2004), pg. 487-518.

The Borwein-Plouffe observation

In 1996, Peter Borwein and Simon Plouffe of SFU in Canada observed that the following well-known formula for log 2

$$\log 2 = \sum_{n=1}^{\infty} \frac{1}{n 2^n} = 0.693147180559945309417232121458176568075 5\ldots_{10}$$
$$= 0.10110001011100100001011111101111101000111001111011\ldots_2$$

leads to a simple scheme for computing binary digits of log 2 at an arbitrary starting position (here {} denotes fractional part):

$$\{2^d \log 2\} = \left\{ \sum_{n=1}^{d} \frac{2^{d-n}}{n} \right\} + \sum_{n=d+1}^{\infty} \frac{2^{d-n}}{n}$$
$$= \left\{ \sum_{n=1}^{d} \frac{2^{d-n} \mod n}{n} \right\} + \sum_{n=d+1}^{\infty} \frac{2^{d-n}}{n}$$

The numerator in the first portion of the RHS can be computed very rapidly using the binary algorithm for exponentiation mod *n*.

Fast exponentiation mod *n*

Problem:

What is 3^{17} mod 10?

Algorithm A:

$3^{17} = 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 \times 3 = 129140163$, so answer = 3.

Algorithm B (faster):

$3^{17} = (((3^2)^2)^2)^2 \times 3 = 129140163$, so answer = 3.

Algorithm C (fastest):

$((((3^2 \bmod 10)^2 \bmod 10)^2 \bmod 10)^2 \bmod 10) \times 3 \bmod 10 = 3$.

Note that in Algorithm C, we never have to deal with integers larger than 81 = $(n - 1)^2$. Thus it can be implemented using ordinary 64-bit integer arithmetic, even for very large n.

The BBP formula for π

In 1996, at the suggestion of Peter Borwein, Simon Plouffe used DHB's PSLQ integer relation program to discover this new formula for π:

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

This formula permits one to compute binary (or hexadecimal) digits of π beginning at an arbitrary starting position, using a very simple scheme that can run on any system with standard 64-bit or 128-bit arithmetic.

Recently it was proven that no base-*b* formulas of this type exist for π, except for when *b* is a power of two.

1. DHB, P. B. Borwein and S. Plouffe, "On the Rapid Computation of Various Polylogarithmic Constants," *Mathematics of Computation*, vol. 66, no. 218 (Apr 1997), pg. 903-913.

2. J. M. Borwein, W. F. Galway and D. Borwein, "Finding and Excluding *b*-ary Machin-Type BBP Formulae," *Canadian Journal of Mathematics*, vol. 56 (2004), pg 1339-1342.

Some other BBP-type formulas

$$\log\frac{9}{10} = -\sum_{k=1}^{\infty}\frac{1}{k10^k}$$

$$\pi^2 = \frac{9}{8}\sum_{k=0}^{\infty}\frac{1}{64^k}\left(\frac{16}{(6k+1)^2}-\frac{24}{(6k+2)^2}-\frac{8}{(6k+3)^2}-\frac{6}{(6k+4)^2}+\frac{1}{(6k+5)^2}\right)$$

$$\begin{aligned}\pi^2 = \frac{2}{27}\sum_{k=0}^{\infty}\frac{1}{729^k}\Bigg(&\frac{243}{(12k+1)^2}-\frac{405}{(12k+2)^2}-\frac{81}{(12k+4)^2}-\frac{27}{(12k+5)^2}\\&-\frac{72}{(12k+6)^2}-\frac{9}{(12k+7)^2}-\frac{9}{(12k+8)^2}-\frac{5}{(12k+10)^2}+\frac{1}{(12k+11)^2}\Bigg)\end{aligned}$$

$$\begin{aligned}\zeta(3) = \frac{1}{1792}\sum_{k=0}^{\infty}\frac{1}{2^{12k}}\Bigg(&\frac{6144}{(24k+1)^3}-\frac{43008}{(24k+2)^3}+\frac{24576}{(24k+3)^3}+\frac{30720}{(24k+4)^3}\\&-\frac{1536}{(24k+5)^3}+\frac{3072}{(24k+6)^3}+\frac{768}{(24k+7)^3}-\frac{3072}{(24k+9)^3}-\frac{2688}{(24k+10)^3}\\&-\frac{192}{(24k+11)^3}-\frac{1536}{(24k+12)^3}-\frac{96}{(24k+13)^3}-\frac{672}{(24k+14)^3}-\frac{384}{(24k+15)^3}\\&+\frac{24}{(24k+17)^3}+\frac{48}{(24k+18)^3}-\frac{12}{(24k+19)^3}+\frac{120}{(24k+20)^3}+\frac{48}{(24k+21)^3}\\&-\frac{42}{(24k+22)^3}+\frac{3}{(24k+23)^3}\Bigg)\end{aligned}$$

BBP formulas and normality

Consider the general BBP-type constant

$$\alpha = \sum_{n=0}^{\infty} \frac{p(n)}{b^n q(n)}$$

where *p* and *q* are integer polynomials, deg *p* < deg *q*, and *q* has no zeroes for nonnegative arguments. Let {} denote fractional part.

In 2001, DHB and Richard Crandall proved that α is *b*-normal iff the sequence $x_0 = 0$, and

$$x_n = \left\{ b x_{n-1} + \frac{p(n)}{q(n)} \right\}$$

is equidistributed in the unit interval. Here "equidistributed" means that the sequence visits each subinterval [*c*, *d*) with limiting frequency *d* - *c*.

DHB and R. E. Crandall, "On the Random Character of Fundamental Constant Expansions," *Experimental Mathematics*, vol. 10, no. 2 (Jun 2001), pg. 175-190.

Two specific examples

Let {} denote fractional part, and consider the sequence $x_0 = 0$, and

$$x_n = \left\{ 2x_{n-1} + \frac{1}{n} \right\}$$

Then log 2 is 2-normal iff this sequence is equidistributed in the unit interval.

Similarly, consider the sequence $x_0 = 0$, and

$$x_n = \left\{ 16x_{n-1} + \frac{120n^2 - 89n + 16}{512n^4 - 1024n^3 + 712n^2 - 206n + 21} \right\}$$

Then π is 16-normal (and hence 2-normal) iff this sequence is equidistributed in the unit interval.

A class of provably normal constants

DHB and Crandall have also shown that an infinite class of mathematical constants is 2-normal, including

$$\begin{aligned}\alpha_{2,3} &= \sum_{n=1}^{\infty} \frac{1}{3^n 2^{3^n}} \\ &= 0.041883680831502985071252898624571682426096\ldots_{10} \\ &= 0.0ab8e38f684bda12f684bf35ba781948b0fcd6e9e0\ldots_{16}\end{aligned}$$

This constant was proven 2-normal by Stoneham in 1971, but we have extended this to the case where (2,3) are any pair (*p*,*q*) of relatively prime integers > 1. We also extended this result to an uncountable class:

$$\alpha_{2,3}(r) = \sum_{n=1}^{\infty} \frac{1}{3^n 2^{3^n + r_n}}$$

Here r_n is the *n*-th bit in the binary expansion of *r* in (0,1). These constants are all distinct.

DHB and R. E. Crandall, "Random Generators and Normal Numbers," *Experimental Mathematics*, vol. 11, no. 4 (2002), pg. 527-546.

A “hot spot” lemma

Given the real constant α, if there exists some *B* such that for every subinterval [*c,d*) of [0,1),

$$\limsup_{m \geq 1} \frac{\#_{0 \leq j < m} \left(\{b^j \alpha\} \in [c, d) \right)}{m(d-c)} \quad \leq \quad B$$

then α is *b*-normal.

In other words, if α is not *b*-normal, then there is some interval [*c*, *d*) that is visited 10 times too often by shifts of the base-*b* expansion of α; there is some other interval [*c*’, *d*’) that is visited 100 times too often; there is some other interval [*c*’’, *d*’’) that is visited 1000 times too often, etc. However, one cannot conclude that these intervals are nested.

L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, 1974, pg. 77.

A strong "hot spot" lemma

Recently DHB and Michal Misiurewicz proved a stronger version of this result, using methods of ergodic theory:

Let $0.x_1x_2\ldots x_n$ be the base-b expansion of x to position n. If for every x in (0,1),

$$\liminf_{n\geq 1} \limsup_{m\geq 1} \frac{\#_{0\leq j<m}\left[\{b^j\alpha\} \in [0.x_1x_2\ldots x_n, 0.x_1x_2\ldots x_n + b^{-n})\right]}{mb^{-n}} \quad < \quad \infty$$

then α is b-normal.

In other words, if α is not b-normal, then there is at least one x in (0,1) such that shifts of the base-b expansion of α visit all sufficiently small digit neighborhoods of x too often, by an arbitrarily large factor.

DHB and M. Misiurewicz, "A Strong Hot Spot Theorem," *Proceedings of the American Mathematical Society*, vol. 134 (2006), no. 9, pg. 2495-2501.

Simple examples of “hot spots”

Consider the rational fraction 1/28:

0.0357142857142857142857142857142857142857142857142857142857142857142857...

Examine the sequence of shifts of its decimal expansion:

0.3571428571428571428571428571428571428571428571428571428571428571428571…

0.5714285714285714285714285714285714285714285714285714285714285714285714…

0.7142857142857142857142857142857142857142857142857142857142857142857142…

0.1428571428571428571428571428571428571428571428571428571428571428571428…

0.4285714285714285714285714285714285714285714285714285714285714285714285…

0.2857142857142857142857142857142857142857142857142857142857142857142857…

0.8571428571428571428571428571428571428571428571428571428571428571428571…

This sequence visits each of the six points 1/7, 2/7, 3/7, …, 6/7 one-sixth of the time, in the limit, which, relative to the size a sufficiently small neighborhood around each of these points, is too often by an arbitrarily large factor. Thus these six points are “hot spots” for 1/28. Thus 1/28 is not a 10-normal number.

In a similar way, the constant $1/10 + 1/10^4 + 1/10^9 + 1/10^{16} + 1/10^{25} + \ldots =$

0.10010000100000010000000010000000000100000000000010000000000000010000000…

is irrational but not 10-normal, since zero is clearly a “hot spot” for this constant.

The BBP sequence corresponding to $\alpha_{2,3}$

It is fairly easy to show that the BBP sequence (z_n) corresponding to

$$\alpha_{2,3} = \sum_{n=1}^{\infty} \frac{1}{3^n 2^{3^n}}$$

is the following. Note that each section is repeated three times, and the sequence evenly fills in the unit interval with all fractions of the form $k\,3^{-p}$.

$$0,\ 0,\ 0,$$
$$\frac{1}{3},\ \frac{2}{3},\ \frac{1}{3},\ \frac{2}{3},\ \frac{1}{3},\ \frac{2}{3},$$
$$\frac{4}{9},\ \frac{8}{9},\ \frac{7}{9},\ \frac{5}{9},\ \frac{1}{9},\ \frac{2}{9},\quad \text{(repeated 3 times)},$$
$$\frac{13}{27},\ \frac{26}{27},\ \frac{25}{27},\ \frac{23}{27},\ \frac{19}{27},\ \frac{11}{27},\ \frac{22}{27},\ \frac{17}{27},\ \frac{7}{27},\ \frac{14}{27},\ \frac{1}{27},\ \frac{2}{27},\ \frac{4}{27},\ \frac{8}{27},\ \frac{16}{27},\ \frac{5}{27},$$
$$\frac{10}{27},\ \frac{20}{27},\quad \text{(repeated 3 times), etc.}$$

It is also easy to show that the sequence of shifted bits satisfies, for all $n > 0$,

$$|\{2^n \alpha_{2,3}\} - z_n| < \frac{1}{2n}$$

Proof that $\alpha_{2,3}$ is 2-normal

Given any x in (0,1), let $c = 0.x_1x_2\ldots x_n$, and $d = 0.x_1x_2\ldots(x_n+1)$. Let m be any integer $> 2^{2n}$, and let 3^p such that $3^p <= m < 3^{p+1}$. Note that for $j > 2^n$, $[c - 1/(2j), d + 1/(2j))$ is a subset of $[c - 2^{-n-1}, d + 2^{-n-1})$. Since the length of this interval is 2^{-n+1}, it contains at most $3^p\, 2^{-n+1} + 1$ instances of $k\, 3^{-p}$. Therefore

$$\frac{\#_{0\le j<m}\left(\{2^j\alpha\} \in [c,d)\right)}{m2^{-n}} \le \frac{2^n + \#_{2^n\le j<m}\left(z_j \in [c-2^{-n-1}, d+2^{-n-1})\right)}{m2^{-n}}$$

$$\le \frac{2^n + 3(3^p 2^{-n+1} + 1)}{m2^{-n}} < 8$$

Thus by the hot spot lemma, $\alpha_{2,3}$ is 2-normal.

See paper by DHB and Misiurewicz for full details.

$\alpha_{2,3}$ is not 6-normal

It is also possible to establish non-normality results for $\alpha_{2,3}$ in certain number bases, such as base 6. Note that we can write

$$\{6^n \alpha_{2,3}\} = \left\{ \sum_{m=1}^{\lfloor \log_3 n \rfloor} 3^{n-m} 2^{n-3^m} \right\} + \left\{ \sum_{m=\lfloor \log_3 n \rfloor + 1}^{\infty} 3^{n-m} 2^{n-3^m} \right\}.$$

The first portion of this expression is zero, since all of the terms in the summation are zero. When $n = 3^m$, the second portion is, very accurately,

$$\{6^{3^m} \alpha_{2,3}\} \approx \frac{\left(\frac{3}{4}\right)^{3^m}}{3^{m+1}}$$

Thus the base-6 expansion of $\alpha_{2,3}$ has long stretches of zeroes beginning at positions $3^m + 1$. This observation can be fashioned into a rigorous proof of non-normality.

DHB, "A Non-Normality Result," manuscript, Aug 2007, http://crd.lbl.gov/~dhbailey/dhbpapers/alpha-6.pdf.

A pseudorandom number generator based on the binary digits of $\alpha_{2,3}$

Given a seed *q* in the range $3^{33} + 100 < q < 2^{53}$, use the first line to compute x_0, and use the second line to compute all successive iterates:

$$x_0 = (2^{q-3^{33}} \cdot \lfloor 3^{33}/2 \rfloor) \mod 3^{33}$$
$$x_k = (2^{53} \cdot x_{k-1}) \mod 3^{33}$$

Divide the results by 3^{33} to obtain pseudorandom 64-bit floating-point iterates in (0,1).

This generator has several desirable properties:

- Iterates contain successive 53-bit sections of the binary digits of $\alpha_{2,3}$.
- It is not subject to power-of-two stride problems that plague other schemes.
- It passes all standard tests for randomness.
- It is well-suited for parallel processing – each individual processor can quickly jump to its own starting point in the sequence.
- Efficient implementations are as fast as several other widely used schemes.

DHB, "A Pseudo-Random Number Generator Based on Normal Numbers," Dec 2004, http://crd.lbl.gov/~dhbailey/dhbpapers/normal-random.pdf.