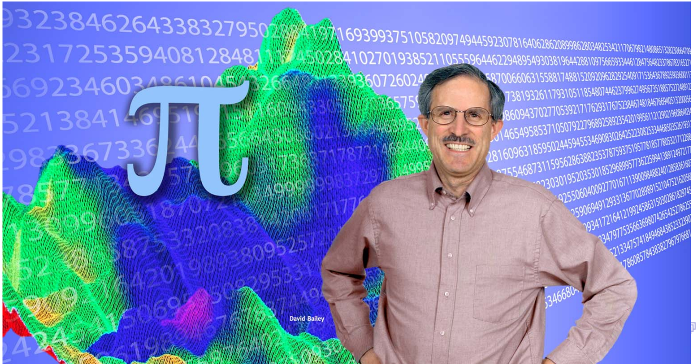
Are the digits of Pi random?

David H. Bailey
https://www.davidhbailey.com
Lawrence Berkeley National Laboratory (retired)
Updated 1 March 2024

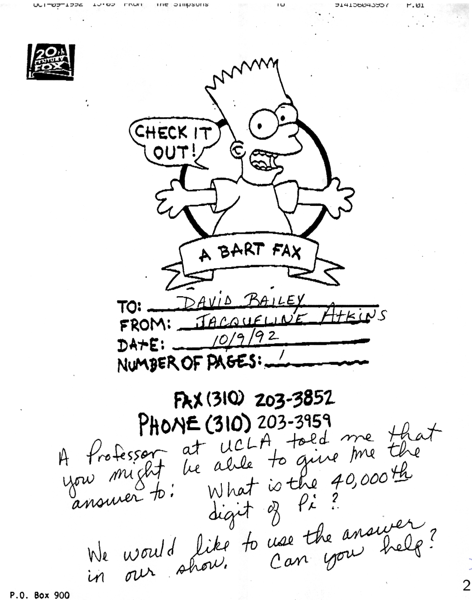


DHB receives a fax from the Simpsons

In October 1992, while working at NASA, DHB received this fax from the Simpsons TV crew, requesting the 40,000th digit of π. He computed and faxed back 40,000 digits, noting that the 40,000th digit is a 1.

In the 6 May 1993 episode, Apu (manager of a convenience store) was challenged by Marge's attorney in a courtroom. He replied that he has a perfect memory; he can recite 40,000 digits of π, and the last digit is a 1.

No other mathematical constant has this level of public interest in its digits!



Normal numbers: An embarrassing ignorance

Given integer $b \geq 2$, a real number x is b-normal (or "normal base b") if every m-long string of digits appears in the base-b expansion of x with limiting frequency $1/b^m$.

Using measure theory, it can be shown that almost all real numbers are b-normal for any given integer base b. It is widely believed that all of these are b-normal, for all integer bases $b \geq 2$:

1. $\pi = 3.14159265358979323846\ldots$
2. $e = 2.71828182845904523540\ldots$
3. $\sqrt{2} = 1.41421356237309504880\ldots$
4. $\log(2) = 0.69314718055994530942\ldots$
5. Every irrational algebraic number (Borel's conjecture).

But there are no proofs for any of these constants in any number base. We do not even know for sure whether the digit 7 appears 1/10 of the time in the decimal expansion of π, or whether the digit 1 appears 1/2 of the time in the binary expansion.

Until recently, normality proofs were known only for a few contrived examples such as Champernowne's constant $= 0.12345678910111213141516\ldots$ (which is 10-normal).

One very weak result for algebraic numbers

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

Simple special case: The first n binary digits of $\sqrt{2}$ must have at least $\sqrt{n}$ one bits.

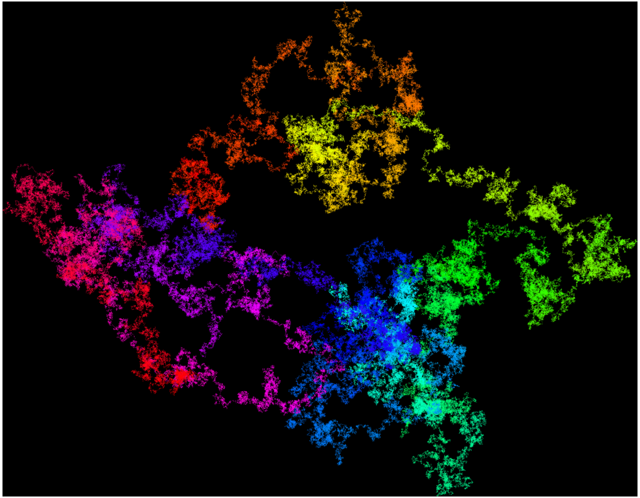
This result follows by noting that the one-bit count of the product of two integers is less than or equal to the product of the one-bit counts of the two integers. The more general result above requires a more sophisticated approach.

$$
\begin{array}{rrrrrrrrrr}
 & & & & & 1 & 0 & 1 & 1 & 1 \\
 & & & & \times & 1 & 1 & 0 & 1 & 1 \\
\hline
1 & 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 1
\end{array}
$$

However, note that these results are still a far cry from even single-digit normality.

1. D. H. Bailey, J. M. Borwein, R. E. Crandall and C. Pomerance, "On the binary expansions of algebraic numbers," *J. Number Theory Bordeaux*, v. 16 (2004), 487–518; preprint at DHB's website.

A random walk generated from the base-4 digits of π



See this user-searchable random walk on the first 100 billion base-4 digits of π (courtesy F. Aragon-Artacho), available at http://gigapan.com/gigapans/106803.

Experimental mathematics: Data mining meets mathematical research

Methodology:

1. Compute various mathematical entities (integrals, infinite series sums, limits, etc.) to very high numeric precision, typically 100–100,000 digits.
2. Use algorithms such as PSLQ to recognize these numerical values as results of formulas involving known mathematical constants.
3. Seek formal mathematical proofs of the discovered relations.

Many results have recently been found using this methodology, in both pure and applied mathematics.

> *If mathematics describes an objective world just like physics, there is no reason why inductive methods should not be applied in mathematics just the same as in physics.*
> — *Kurt Godel*

The PSLQ integer relation algorithm

Let (x_n) be a given vector of real numbers. An integer relation algorithm either finds integers (a_n) such that

$$a_1 x_1 + a_2 x_2 + \cdots + a_n x_n = 0$$

(to within the "epsilon" of the arithmetic being used), or else finds bounds within which no relation can exist.

The "PSLQ" algorithm and its variants, due to mathematician-sculptor Helaman Ferguson, is the most widely used integer relation algorithm.

Integer relation detection by any algorithm requires very high precision (at least $n \times d$ digits, where d is the size in digits of the largest a_k), both in the input data and in the operation of the algorithm.

1. H. R. P. Ferguson, D. H. Bailey and S. Arno, "Analysis of PSLQ, an integer relation finding algorithm," *Math. of Computation*, v. 68 (Jan 1999), 351–369; preprint at DHB's website.
2. D. H. Bailey and D. J. Broadhurst, "Parallel integer relation detection: Techniques and applications," *Math. of Computation*, v. 70 (Oct 2000), 1719–1736; preprint at DHB's website.

The BBP formula for π

In 1996, a PSLQ-like computer program discovered this new formula for π:

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

This formula permits one to compute base-2 (binary) or base-16 (hexadecimal) digits of π beginning at an arbitrary starting position, using a very simple scheme that requires only standard 64-bit or 128-bit floating-point arithmetic.

In 2004, Borwein, Galway and Borwein proved that the only BBP-type formulas for π have base $n = 2^m$ for some m. But there is both a base-2 and a base-3 formula for π^2.

1. D. H. Bailey, P. B. Borwein and S. Plouffe, "On the rapid computation of various polylogarithmic constants," *Math. of Computation*, v. 66 (Apr 1997), 903–913; preprint at DHB's website.
2. J. M. Borwein, W. F. Galway and D. Borwein, "Finding and excluding b-ary Machin-type BBP formulae," *Canadian J. of Mathematics*, v. 56 (2004), 1339–1342.

A sample of the many BBP-type formulas found using PSLQ

$$\pi^2 = \frac{1}{8}\sum_{k=0}^{\infty}\frac{1}{64^k}\left(\frac{144}{(6k+1)^2} - \frac{216}{(6k+2)^2} - \frac{72}{(6k+3)^2} - \frac{54}{(6k+4)^2} + \frac{9}{(6k+5)^2}\right)$$

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

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

1. D. H. Bailey, "A compendium of BBP-type formulas," preprint at DHB's website.

How to compute binary digits of log 2 at an arbitrary position

The BBP algorithm for computing binary digits at an arbitrary position can be illustrated as follows, based Euler's formula for log 2:

$$\log 2 = \sum_{n=1}^{\infty} \frac{1}{n2^n} \;=\; 0.10110001011100100001011111110111110100011100111011\ldots_2$$

Note that the binary digits of log 2 beginning at position $d+1$ can be written as $\mathrm{frac}(2^d \log 2)$, where $\mathrm{frac}(\cdot)$ denotes fractional part. Thus we can write:

$$\begin{aligned}\mathrm{frac}(2^d \log 2) &= \mathrm{frac}\left(\sum_{n=1}^{d} \frac{2^{d-n}}{n}\right) + \mathrm{frac}\left(\sum_{n=d+1}^{\infty} \frac{2^{d-n}}{n}\right) \\ &= \mathrm{frac}\left(\sum_{n=1}^{d} \frac{2^{d-n} \bmod n}{n}\right) + \mathrm{frac}\left(\sum_{n=d+1}^{\infty} \frac{2^{d-n}}{n}\right),\end{aligned}$$

where we have inserted $\bmod n$ since were are only interested in the fractional part when divided by n. Now note that the numerator $2^{d-n} \bmod n$ can be calculated very rapidly by the binary algorithm for exponentiation.

The binary algorithm for exponentiation

Problem: What is 3^{17} mod 10? In other words, what is the last decimal digit of 3^{17}?

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

Algorithm B (much faster): $3^{17} = ((((3^2)^2)^2)^2) \times 3 = 129140163$, so answer $= 3$.

Algorithm C (much faster still):
$3^{17} \bmod 10 = ((((3^2 \bmod 10)^2 \bmod 10)^2 \bmod 10)^2 \bmod 10) \times 3 \bmod 10 = 3$.
Note that we never have to deal with integers larger than $9 \times 9 = 81$.

The BBP scheme for π, combined with the binary algorithm for exponentiation, has been used to calculate a string of binary digits of π beginning at position 2 quadrillion.

All recent large computations of π (the current record is 100 trillion decimal digits) employ the BBP scheme as a check.

BBP formulas and normality

Consider a general BBP-type constant (i.e., a formula that permits the BBP scheme):

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

where p and q are integer polynomials, $\deg p < \deg q$, and q has no zeroes for nonnegative arguments.

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

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

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

1. D. H. Bailey and R. E. Crandall, "On the random character of fundamental constant expansions," *Experimental Mathematics*, v. 10 (Jun 2001), 175–190; preprint at DHB's website.

Two specific examples

Consider the sequence $x_0 = 0$ and

$$x_n = \mathrm{frac}\left(2x_{n-1} + \frac{1}{n}\right).$$

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

Similarly, consider the sequence $x_0 = 0$ and

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

Then π is 16-normal (and hence 2-normal) iff this sequence is equidistributed in the unit interval. By the way, this iteration gives the sequence of base-16 (hexadecimal) digits of π without any errors, as far as we can determine.

These results are promising first steps, but the question of whether $\log 2$ or π is 2-normal (or 10-normal) remains unanswered. But perhaps some smart young person...

A class of provably normal constants

DHB and Crandall proved that an infinite class of constants are 2-normal, for example:

$$\alpha_{2,3} = \sum_{n=1}^{\infty} \frac{1}{3^n 2^{3^n}}$$
$$= 0.0418836808315029850712528986245716824260967\ldots_{10}$$
$$= 0.0ab8e38f684bda12f684bf35ba781948b0fcd6e9e0\ldots_{16}$$

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 ≥ 2. We also extended this result to an uncountable class: For any real r in $[0,1)$, the constant

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

is 2-normal, where r_n is the n-th bit in the binary expansion of r in $[0,1)$. These constants are all distinct, so the class is uncountable.

1. D. H. Bailey and R. E. Crandall, "Random generators and normal numbers," *Experimental Mathematics*, v. 11 (2002), 527–546; preprint at DHB's website.

A "hot spot" lemma to prove normality

A known result from measure theory: Let $\{\text{frac}(b^j \alpha)\}$ be shifts of the base-b expansion of the constant α. If there exists some B such that for every subinterval (c, d) of $(0, 1)$,

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

then α is b-normal.

In plain English: If α is not b-normal, then there exists some interval (c, d) that is visited 10 times too often relative to its size 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.

Conversely, if one can show that no interval (c, d) is visited, say, more than 10 times too often by the sequence of shifts $\{\text{frac}(b^j \alpha)\}$, then this proves α is b-normal.

This result can be employed to dramatically simplify the proofs of normality for constants such as $\alpha_{2,3}$ and $\alpha_{p,q}$. Perhaps also for constants such as π and $\log 2$?

A strong "hot spot" lemma

In 2006, DHB and Michal Misiurewicz proved a stronger version of this result, using methods of ergodic theory: Let $X_n = (0.x_1 x_2 x_3 \ldots x_n)$ be the base-b expansion of x out to position n. If for every $x \in (0,1)$, it is true that

$$\liminf_{n \geq 1} \limsup_{m \geq 1} \frac{\#_{0 \leq j < m} \operatorname{frac}(b^j \alpha) \in (X_n, X_n + b^{-n})}{m b^{-n}} < \infty,$$

then α is b-normal.

In plain English: 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 neighborhoods of x too often, by an arbitrarily large factor. Note that this avoids the possibility, mentioned in the previous viewgraph, that the intervals might be different.

1. DHB and M. Misiurewicz, "A strong hot spot theorem," *Proc. of the American Mathematical Society*, v. 134 (2006), 2495–2501; preprint at DHB's website.

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

First 900 base-6 digits of $\alpha_{2,3}$:

3130140430003334251130502130000001243555045432233011500243525320551352
3435410104300000000000000000051411300540405554553031442504334351012413 45
2351125142125134505503545015053522052044340452151505102411552500425130
0511244540010441311500324203032130000000000000000000000000000000000000
0000014212034311121452013525445342113412240220525301054204423552411055
4150155204350414555400310145303033532002534340401301240104453254343502
1420204324150255551010040433000455441145010313314511510144514123443342
3412400551313335045423530553151153501533452435450250055521453054234342
1530350125024205404135451231323245353031534552304115020154242121145201
5422225343403404505301233255344404431033324453321414150142334545424124
3203125340050134150245514404300000000000000000000000000000000000000000
0000000000000000000000000000000000000000000000000000000000000000000000
0000000000313350542444431111055534141052014540213412313001424333133115

Note the long stretches of zeroes after power-of-three positions, e.g., after 3, 9, 27, 81, 243, 727, etc. This observation can be fashioned into a rigorous proof of non-normality.

Conclusion

1. Despite literally millennia of wonder and effort by mathematicians throughout history, the question of whether the digits of π are "random" remains unanswered.
2. We do not even know for sure whether the digit 7 appears $1/10$ of the time in the decimal expansion of π, or whether the digit 1 appears $1/2$ of the time in the binary expansion.
3. The same conclusion applies to numerous other well-known constants of mathematics, including $\log 2$, e, $\sqrt{2}$, ϕ (golden ratio), and more.
4. Large statistical analyses on literally trillions of digits, binary and decimal, have failed to find any significant abnormalities. But there are no proofs.
5. On the other hand, several new results and techniques, computational as well as theoretical, have been obtained in the past few years, suggesting that perhaps some very bright young person...

This talk is available at:

http://www.davidhbailey.com/dhbtalks/dhb-west-point-2024.pdf