Solved and Unsolved Problems in Normal Numbers

Joseph Vandehey

The University of Texas at Tyler

jvandehey@uttyler.edu

May 18, 2021

May 18, 2021

1/55

Introduction

- 2 The first question
- 3 Constructing normal numbers
- 4 Normality as measure of complexity
- 5 Normality and descriptive set theory

Normality for base-*b* expansions we may think of as one of two equivalent concepts, depending on whether or not we think of a number or think of the digits in that numbers base-*b* expansion.

Symbolic interpretation: An infinite sequence of base-*b* digits is normal if every finite subsequence appears with the same limiting frequency as every other subsequence.

Dynamic interpretation: A number $x \in [0,1)$ is normal if the orbit $(T^i x)_{i \in \mathbb{N}}$ under the action

$$T: [0,1) \rightarrow [0,1)$$
 $Tx = bx \pmod{1}$

is equidistributed with respect to the Lebesgue measure.

Unless stated otherwise, normality means base-b normality for a pre-defined b.

Image: Image:

- Absolutely normal normal to all bases simultaneously.
- Normal of order k Every string of length k appears to the correct limiting frequency.
- Simply normal Normal of order 1.
- Richness/Denseness Every string appears at least once.
- Strongly normal The limiting *variance* of the strings matches that of a randomly chosen number.
- Sharply normal Instead of measuring the frequency of strings in the first N digits, measure the frequency from the Nth to roughly the $N + \sqrt{N}$ th.

All of these characteristics occur on full measure sets. Moreover sharply normal implies regular normality.

We can define normality on a variety of systems, either by specifying the exact frequency we want each string to appear with, or by specifying the transformation and measure we want the orbit to be equidistributed on. (Cigler, 1964, or Postnikov–Pyatetskii-Shapiro?)

- Generalized Lüroth series similar to base-*b* expansions, but the expected weight of a given digit could be different
- β -expansions Not every string is admissible.
- Continued fraction (CF) expansion Measure is not a product measure on the digits.
- Many, many continued fraction variants Measure is not a product measure and may not necessarily even be finite (may need to swap to ratio normality)

From an ergodic perspective, studying normality as equidistibution of the orbit $(T^i x)_{i \in \mathbb{N}}$, we often consider a transformation T that is a homogeneous \mathbb{Z} -action (or \mathbb{N} -action).

What if we lose homogeneity, so that the transformation isn't the same at every stage?

The Q-Cantor series expansion is like a base-b expansion where the base may change at every digit. (See Mance.) Most facts that are true about base-b normality do not apply here, including the equivalence of the symbolic and dynamic meanings of normality.

For things other than \mathbb{Z} -actions, we typically move beyond looking at sequences of digits $(a_i)_{i \in \mathbb{Z}}$ to arrays of digits, such as $(a_{(i,j)})_{(i,j) \in \mathbb{Z}^2}$.

We lose the natural choice of [1, n] as our segments to measure frequency over and must replace them with some choice of Folner sequence. So in these more general circumstances, we have to define normality with respect to a given Folner sequence. However, once we have, a good number of results do nicely generalize. (See Bergelson-Downarowicz-Misiurewicz and Bergelson-Downarowicz-V.) Let's take a look at the first major question in normal numbers.

The concept of normality was formalized by Borel (1909) (using a completely different definition of normality).

He declared that the problem of understanding the digits of $\sqrt{2}$ was the most important problem facing modern mathematics.

Modern mathematics has largely disagreed, but a few of us are still working on Borel's project.

Borel's questions and generalizations

Problem

Is $\sqrt{2}$ normal to any base b?

Unknown, conjectured to be normal for all bases.

Problem

Are any of π , e, ln 2, etc. normal to any base b?

Also unknown, also conjectured to be normal for all bases.

Problem

Are there any naturally occurring constants which are known to be normal to any base?

Yes. Chaitin's constant Ω , the halting probability of the universal Turing machine, is absolutely normal.

Joseph Vandehey (UTT)

The closest we have come to showing the normality of a commonly used irrational number is in the binary expansion of $\sqrt{2}$. Let A(N) denote the number of 1's in the first N binary digits of $\sqrt{2}$. Here are the current records, and remember, we want $A(N) \sim N/2$:

- $A(N) \ge C\sqrt{N}$ (Bailey, Borwein, Crandall, Pomerance, 2004)
- $A(N) \ge \sqrt{N}(1 + o(1))$ (Rivoal?, 2008)

•
$$A(N) \ge \sqrt{2N}(1+o(1))$$
 (V., 2018)

- $\limsup_{N \to \infty} A(N) / \sqrt{N} \ge 2 / \sqrt{2\sqrt{2} 1}$ (V., 2018)
- $\limsup_{N \to \infty} A(N) / \sqrt{N} \ge \sqrt{8/\pi}$ (Dubickas, 2019)
- $A(N) \ge \sqrt{2.119N}(1 + o(1))$? (V., unpublished)

The problem of $\sqrt{2}$

Here's the idea of these proofs. If

$$\sqrt{2}=\sum\frac{1}{2^{a_i}},$$

where the a_i 's are the indices where 1's occur in the binary expansion of $\sqrt{2}$, then

$$2 = \left(\sum \frac{1}{2^{a_i}}\right)^2 = \sum \frac{1}{2^{a_i+a_j}}$$

If the a_i 's are sparse, we expect the carries not to amount to much. So this question is closely related to the size of $A \subset \mathbb{N}$ for which $A + A = \mathbb{N}$.

Problem

Sets $A \subset \mathbb{N}$ such that $A + A = \mathbb{N}$ tend to have lots of redundancy, so there should be lots of carries. Why can't we take this into account?

May 18, 2021

12 / 55

Joseph Vandehey (UTT)

Problem

Can the BBCP result be extended to bound how many non-zero digits algebraic irrationals must have in β -expansions?

(日)

Three main types of constructions:

- Combinatorial method
- Analytic method
- Rational-analytic method

(This is not meant to be all inclusive, just a marking of several major techniques.)

Key idea: Most finite length strings should be close to normal, so if we concatenate enough strings that become arbitrarily close to normal, we should get something that is normal.

Here "close to normal" means some variant of (ϵ, k) -normal, which states that for all strings of length k, the string appears to within ϵ of the desired frequency.

Theorem

The number of base-b strings of length ℓ that are NOT (ϵ, k) -normal is at most $b^{\ell(1-\delta)}$ for some $\delta = \delta(\epsilon, k) > 0$.

Problem

What is the precise asymptotic for the number of non- (ϵ, k) -normal numbers?

The number $0.(a_1)(a_2)(a_3)...$ formed by concatenating the a_n 's is normal when...

- $a_n = n$ (Champernowne, 1933)
- $a_n = n^2$ (Besicovitch, 1935)
- *a_n* is the *n*th prime number (Copeland-Erdős, 1946)
- (Many constructions involving the distribution of prime numbers in De Koninck and Kátai.)
- A variety of constructions including $a_n = \lfloor \sqrt{n} \rfloor$ in Szüsz and Volkmann (1994)
- a_n is the Euler totient function $\phi(n)$ (Pollack-V., 2015)
- a_n is the sum of divisors function $\sigma(n)$ (Pollack-V., 2015)

Problems in the combinatorial method

One thing one might notice of the many examples is that all the a_n 's are increasing in the long run and growing at most linearly (up to some log factors).

With one major exception... $a_n = n^2$.

The combinatorial technique relies on non- (ϵ, k) -normal numbers being sparse, so any sufficiently dense set of natural numbers will consist mostly of (ϵ, k) -normal numbers. But the squares are already sparse! Besicovitch's solution was to break $a_n = n^2$ into two pieces, looking at the

first half and second half separately.

Problem (Pollack-V.)

Can you prove the normality of

```
0.(1^3)(2^3)(3^3)(4^3)\dots
```

using the combinatorial technique?

Also, many of these sequences grow fairly steadily. A function like $\phi(n)$ is not a non-decreasing function, but it is close to one, in the sense that $\phi(n) \approx n$ up to some log factors.

What about functions that are much more chaotic in their overall size?

Problem (Katie Anders)

If you concatenate the Stern sequence, do you get a normal number?

 $0.(a_1)(a_2)(a_3)...$

where $a_0 = 0$, $a_1 = 1$, $a_{2n} = a_n$ and $a_{2n+1} = a_n + a_{n+1}$.

0.01121323143525341...

The combinatorial method is currently the only method which extends nicely to questions about other systems besides base-*b*. The same basic concept (that most finite strings should be close to normal) is quite general and has been rediscovered in multiple contexts. (Postnikov–Pyatetskii-Shapiro for CF, Betrand-Mathis–Volkmann for β -expansions, and Madritsch-Mance for the most general construction.)

For continued fractions, we concatenate the finite expansions of the rationals in order

$$\frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{1}{4}, \frac{2}{4}, \frac{3}{4}, \dots$$

(Adler-Keane-Smorodinsky)

Several simple variants of this, like only considering those rationals whose numerator is prime, are also possible. (V.)

< □ > < □ > < □ > < □ > < □ > < □ >

It is known that Copeland-Erdős's result is functionally sharp. Namely, for any $\epsilon > 0$, there exist subsets of the positive integers which contain at least $N^{1-\epsilon}$ of the first N integers that do NOT result in a normal number when concatenated (Shiokawa, 1974).

Problem

What is the equivalent optimal bound for continued fractions (ordered as in Adler-Keane-Smorodinsky)?

The order of the rationals is important. If they are ordered as in the Kepler or Farey tree, then the resulting concatenation will be "Minkowski" normal (Dajani-de Lepper-Robinson).

Efficient constructions

Both Champernowne's construction and the Adler-Keane-Smorodinsky construction are what we might consider efficient constructions:

- Every (or functionally every) finite string is used in the concatenation.
- **2** A given string is only used once in the concatenation.

However, for most systems, the general construction (see Madritch-Mance) is very inefficient. A given string of length k might get used k^{2k} times.

Problem

Does an efficient combinatorial construction exist for most reasonable systems? And if so, what is a good way to order the strings?

The answer is yes (V.) for certain finite digit Lüroth series (the order is by the size of the corresponding cylinder), and yes (Bertrand-Mathis–Volkmann) for certain β -expansions, but unknown in more general situations.

Analytic and rational-analytic methods

The analytic method: prove a concatenation of function values

 $0.(f(1))(f(2))(f(3))(f(4))\dots$

is base-b normal by studying

$$\sum_{n\leq N} e^{2\pi i \frac{a}{b^k} f(n)}$$

The rational-analytic method: prove a number of the form

$$\sum_{m} \frac{p_{m}}{q_{m} \cdot b^{a_{m}}} \qquad \text{or} \qquad \prod_{m} \left(1 + \frac{p_{m}}{q_{m} \cdot b^{a_{m}}} \right)$$

is base-b normal by studying

$$\sum_{n\leq N} e^{2\pi i b^n \frac{p_m}{q_m}}$$

The analytic method

The number

$0.(f(1))(f(2))(f(3))\dots$

is normal when f is...

- an integer valued polynomial (Davenport-Erdős, 1952)
- an integer-valued polynomial evaluated at primes. (Nakai-Shiokawa, 1997)
- a polynomial evaluated at shifted primes (De Koninck-Kátai, 2011)
- the floor of a pseudo-polynomial (Nakai, Shiokawa, 1990)
- the floor of a pseudo-polynomial, evaluated at primes (Madritsch, 2014)
- the floor of an entire function of small order (Madritsch-Thuswaldner-Tichy, 2008)
- the *latter half* of the digits of ω , the distinct prime counting function. (V., 2013)

Problem (Pillai's problem)

Is 0.(2)(4)(8)(16)..., formed by concatenating the powers of 2, normal?

This seems very far from what is currently possible.

Problem

Is there an additive function (f(nm) = f(n) + f(m) when n, m relatively prime) that can be concatenated into a normal number? This would result in an exponential sum that can be interpreted as the sum of a multiplicative function.

The rational method

Numbers of the type

$$\sum_{m} \frac{p_m}{q_m \cdot b^{a_m}}$$

are generally called Stoneham-Korobov type numbers. Numbers of the type

$$\prod_{m} \left(1 + \frac{p_m}{q_m \cdot b^{a_m}} \right)$$

are Wagner-type numbers.

In both cases, most constructions let

- p_m be small, usually ± 1 .
- q_m be growing in m, typically $q_m = c^m$ for some c.
- a_m has to be usually close to the size of q_m .

Problem (Bailey-Crandall)

Show that the numbers

$$\log 2 = \sum_{n} \frac{1}{n2^{n}}, \quad \pi = \sum_{n} \frac{1}{16^{n}} \left(\frac{4}{8n+1} - \frac{2}{8n+4} - \frac{1}{8n+5} - \frac{1}{8n+6} \right)$$
are say base-2 normal

However, the numbers we can show are base-2 normal look more like

$$\sum \frac{1}{3^n 2^{2^n}}$$
 or $\prod \left(1 + \frac{1}{3^n 2^{2^n}}\right)$.

So this problem does not seem reasonable to solve in the near future.

May 18, 2021

27 / 55

Joseph Vandehey (UTT)

Some more practical problems...

Problem What is the slowest growing sequence a_n such that $\sum \frac{1}{n2^{a_n}}$ or $\prod \left(1 + \frac{1}{n2^{a_n}}\right)$ or $\sum \frac{1}{n! \cdot 2^{a_n}} \qquad \text{or} \qquad \prod \left(1 + \frac{1}{n! \cdot 2^{a_n}}\right)$ is base-2 normal?

These problems are still non-trivial because the estimates on $\sum e^{2\pi i b^n \frac{p_m}{q_m}}$ assume that q_m has only a fixed set of prime factors.

▲圖 ▶ ▲ 国 ▶ ▲ 国 ▶ …

Some curious construction problems

Can one exhibit a closed form of a number x such that x is normal and...

Problem (Mendés France)	
1/x is normal?	

Problem

 $\dots x^2$ is normal?

Problem

...f(x) is normal, for a given polynomial $f(x) \in \mathbb{Z}[x]$?

Problem

...f(x) is normal, for all polynomials $f(x) \in \mathbb{Z}[x]$? Is this even possible?

Some curious construction problems

One possible idea for solving these problems: the Wagner-type construction.

$$x = \prod_{m} \left(1 + \frac{p_m}{q_m b^{a_m}} \right)$$

then

lf

$$x^2=\prod_m\left(1+rac{2p_m}{q_mb^{a_m}}+rac{p_m^2}{q_m^2b^{2a_m}}
ight)$$

and

$$\frac{1}{x} = \prod_m \left(1 - \frac{p_m}{q_m b^{a_m}} + \frac{p_m^2}{q_m^2 b^{2a_m}} - \dots \right)$$

all of which have relatively similar forms...

The reason why these doubly-exponential terms keep showing up in these rational-analytic constructions is because we need to give the early terms time to dominate the behavior of the number before tweaking it:

$$\frac{1}{3^{10}} + \frac{1}{3^{1000}2^{3^{1000}}} + \dots$$

But it feels to me like we shouldn't need to wait.

Problem

Is there some meaningful way to develop a notion of two numbers being "disjointly normal" so that when added together, we get a number that reflects the properties of both?

Problem

Give a closed-form or algorithmic construction of a number that is strongly normal.

Problem (Rivoal)

Given an example of a number x that is normal but that 1/x is not normal.

Given two dynamical systems on the same space, say (X, T, μ) and (X, S, ν) , when does *T*-normality imply *S*-normality?

The simple case: base-*b* normality.

It is know that base-*b* normality is equivalent to base-*c* normality if and only if $b^r = c^s$ for some integers *r*, *s*. And if this condition is not satisfied, then there are uncountably many base-*b*-normal base-*c*-non-normal numbers. In fact, let A, B be a partition of the integers that are at least two, such that if b belongs to a set, then so does every integer power of b, then there exist numbers that are b-normal for every $b \in A$ and b-non-normal for every $b \in B$.

Schweiger's conjecture: In general, T-normality is equivalent to S-normality if and only if $T^r = S^s$ for some positive integers r, s

This conjecture is false. At this point we now have several counterexamples.

- Regular continued fraction normality is equivalent to continued fraction from below normality and nearest integer continued fraction normality (Kraaikamp-Nakada, 2000)
- Regular continued fraction normality and odd continued fraction normality are equivalent (V. 2014)

Outside of the classical, base-b case, we have few examples of systems where normality is NOT equivalent, despite it ostensibly being the thing we expect to occur easiest.

The traditional way of showing, say, base-2 normal is not equivalent to base-3 normal is an intricate argument using a Riesz product measure.

Theorem

(Jackson-Mance-V., TBA) There exist uncountably many CF-normal, absolutely abnormal (not normal to any base b) numbers.

Does absolute normality imply CF-normality?

Why is this unsolved while the reverse direction is known?

Because we have examples of numbers which are close to CF-normal but far from base-2 normal, notably most dyadic rationals. However, we do not have good examples of numbers which are close to base-2 normal but far from CF-normal. (This changes if we know $\sqrt{2}$ is base-2 normal.)

There is very much we still don't know here.

Problem

Since Schweiger's conjecture was wrong, what is the right condition for when T-normality implies S-normality?

Problem

Since the previous problem is likely very difficult, is there a better general technique for showing two systems are NOT normal equivalent?

Let's briefly consider why base-100 and base-1000 are equivalent. If the base-100 expansion of a number starts with

0.(12)(34)(56)...

then the base-1000 expansion starts with

0.(123)(456)...

So iterating the base-100 expansion forward 3 times will result in the same number as if we iterated the base-1000 expansion forward 2 times.

All the examples of normal-equivalent system have the same phenomenon occurring although the number of iterations that match up may not be the same at every point. So maybe that's the important consideration.

Exhibit a closed-form construction of a number that is both base-2 and base-3 normal.

Problem

Exhibit a closed-form construction of a number that is base-2 normal and base-3 non-normal. (Examples of base-p normal base-pq non-normal are easy.)

Problem

How does base-b normality relate to Lüroth series normality? To β -expansion normality?

From many different perspectives, normality is seen as a property of being highly complex. For example, from one information theoretical perspective, normality equates to being incompressible.

Therefore if we do something of minimal complexity to a normal number, either we should get something trivial or we should still have something normal.

Our notion of "minimal complexity" will be "deterministic."

As there are many definitions of normality, there are also many definitions of determinism.

Symbolic interpretation: A sequence is deterministic if it has low subword complexity (if one excludes a negligible set of subwords)—that is, the number of strings of length k grows subexponentially in k.

Dynamic interpretation: A sequence is deterministic if it has Kamae entropy 0.

??? interpretation: A sequence is deterministic if it has zero noise.

So let's briefly talk about noise.

Noise

The noise function was defined by Rauzy. Given a sequence (c_n) of digits, we define

$$\beta_{\ell}((c_n), N) = \inf_{F} \frac{1}{N} \sum_{n \leq N} \min\{1, |c_n - F(c_{n+1}, c_{n+2}, \dots, c_{n+\ell})|\}$$

where *F* runs over all functions from ℓ -tuples of digits into one of the digits. Essentially, this measures how predictable (c_n) is.

We can then define the noise of a sequence to be $\beta((c_n)) = \lim_{\ell \to \infty} \lim_{N \to \infty} \beta_{\ell((c_n),N)}$, if these limits exist.

If this feels a lot like a weird entropy calculation, it should!

It turns out that base-b normal numbers will always have noise (b-1)/b. And deterministic numbers will always have noise 0.

Theorem (Weiss, Kamae)

Suppose $0.a_1a_2a_3...$ is base-b normal and $n_1 < n_2 < n_3 < ...$ is of positive density in \mathbb{N} . Then $0.a_{n_1}a_{n_2}a_{n_3}...$ is base-b normal if and only if the n_i 's are indices of non-zero terms in a deterministic sequence. (Selection along a deterministic sequence)

Note that arithmetic progressions are the indices of non-zero terms in a deterministic sequence.

Theorem (Rauzy)

Addition by a number x preserves normality if and only if the digits of x are deterministic.

The digits of any rational number are deterministic.

イロト イヨト イヨト

Normality versus determinism

Theorem (Wall?)

Multiplication by a non-zero rational number preserves base-b normality.

Problem (Mance)

What is the set of numbers that preserve base-b normality when multiplied by?

My suspicion is that the answer is no.

Problem

Show that multiplication by

$$\sum_{n} \frac{1}{10^{n!!}}$$

does not preserve base-10 normality.

Joseph Vandehey (UTT)

However, all of the prior discussion applies to base-b expansions. What about others?

- Selection along arithmetic progressions preserves normality for generalized Lüroth series (Auslander-Dowker, 1979)
- Addition or multiplication by non-zero rationals preserves CF-normality (V.)
- Selection along a non-trivial arithmetic progression does NOT preserve CF-normality (Heersink-V.)
- Selection along a non-trivial arithmetic progression does not preserve normality for β -expansions.
- Addition by periodic points preserves normality for certain β-expansions (V., in progress)
- When working with more general actions, selection along subsets (appropriately defined) preserves normality if and only if the subset is "deterministic" (Bergelson-Downarowicz-V., 2021)

What is the right notion of "determinism" for non-base-b systems? Is there one?

Note: Furstenberg's notion of disjoint systems is not necessarily helpful here.

Problem

Are the only sequences for which selection preserves CF-normality those which consist of almost everything?

Problem

Does addition or multiplication by $\sqrt{2}$ preserve CF-normality?

Problem

Can one construct a base-b number with any given possible noise?

Joseph Vandehey (UTT)

While there is not much time to discuss it here, automata allow a natural generalization of "low complexity" actions where the specific action performed may depend on the number it is being performed to.

For example, Carton recently used this to give a selection method that preserves β -normality.

Time for some descriptive set theory...

- Π_0^0 denotes the collection of all closed sets of \mathbb{R} .
- Σ_0^0 denotes the collection of all open sets of \mathbb{R} .
- Π^0_1 denotes the collection of all countable intersections of open sets of $\mathbb R$
- Σ^0_1 denotes the collection of all countable unions of closed sets of $\mathbb R$
- Π^0_{α} denotes the collection of all countable intersections of $\Sigma^0_{\alpha-1}$ sets
- Σ^0_{α} denotes the collection of all countable unions of $\Pi^0_{\alpha-1}$ sets

The Borel hierarchy

For a given interesting set, we are most interested in where it first shows up in the hierarchy. If it first shows up in Γ , we say it is Γ -complete. (If it shows up no sooner than Γ , then we say it is Γ -hard.)

The set of normal numbers in a given base b is Π_3^0 -complete (Ki-Linton, 1994)

Consider the following description of the set of normal numbers in base-*b*: The set of base-*b* normal numbers consists of all numbers *x* such that for all finite strings *s* of base-*b* digits and all $\epsilon > 0$, there exists $n_0 \in \mathbb{N}$ such that for all $n \ge N$, we have

$$\frac{N_s(x,n)}{n}-\frac{1}{b^{|s|}}\bigg|<\epsilon.$$

This has a universal, existential, and then universal quantifier, or, in terms of sets, an intersection, union, then intersection. The fact that this set is Π_3^0 -complete means there is no simpler way to write this set out. $\exists v \in \mathcal{I}_3$ and $\exists v \in \mathcal{I}_3$

Other things we know.

- The set of CF-normal numbers is Π₃⁰-complete (Airey-Jackson-Kwietniak-Mance).
- The set of absolutely normal numbers is Π_3^0 -complete (Becher-Heiber-Slaman).
- The set of numbers which are normal to at least one base is $\Sigma_4^0\mbox{-}complete$ (Becher-Slaman)

For a given level Γ in the Borel, hierarchy, we define $D_2(\Gamma)$ as the collection of sets of the form $A \setminus B$ for $A, B \in \Gamma$.

If $A \setminus B$ belongs to $D_2(\Gamma)$, then this generally means that A and B are logically distinct in a very powerful way. Not only does being in A not imply being in B, being in A and satisfying a property simpler than belonging to Γ cannot imply being in B.

It is known that the set of numbers that are normal of order 1 but not normal of order 2 is $D_2(\Pi_3^0)$ -complete. (Beros)

Theorem (Jackson-Mance-V., in progress)

The set of numbers that are continued fraction normal but not base-2 normal are $D_2(\Pi_3^0)$ -complete.

So not only does CF-normality not imply base-2 normality, this result also implies that CF-normality and something like base-2 richness do not imply base-2 normality.

Theorem (Jackson-Mance-V., in progress)

The set of numbers that are base-2 normal but not base-3 normal is $D_2(\Pi_3^0)$ -complete.

Surprisingly, the above result is obtained by working with the continued fraction expansion of the numbers.

Where does the set of strongly normal numbers (or sharply normal numbers) exist on the Borel hierarchy?

Problem

Is the set of numbers x that are normal but which 1/x is not a $D_2(\Pi_3^0)$ -complete set?

The end!

Thank you for coming to my talk.



Image: A matrix