The Number Density Paradox
(Thin or Fat it Matters Not)

Ken Seton
kenseton@gmail.com

Abstract

In the world of transfinite cardinality, any talk of number density, dart-board hits, proportions or probability is just a pouring from the empty into the void. Transfinite cardinality is unrelated to any normal concept of proportion or density over an interval and no understanding of it can obtained from probabilistic analogies. This is demonstrated by comparing extremely sparse and extremely dense collections of reals which are normally understood as uncountable and countable respectively. Should this incline one to think that Cantor's equivalence definition is inappropriate for identifying the "size" of an infinite collection ?

This paper is expository and surely contains no original perspective or point that has not been made and over again during the last century. Nevertheless, the author feels that it needs communicating in this form because a common-sense description and appreciation of the situation that does not obfuscate with opaque jargon seems to be missing from textbooks. And we do our students no service by obscuring or avoiding the obvious.

Keywords

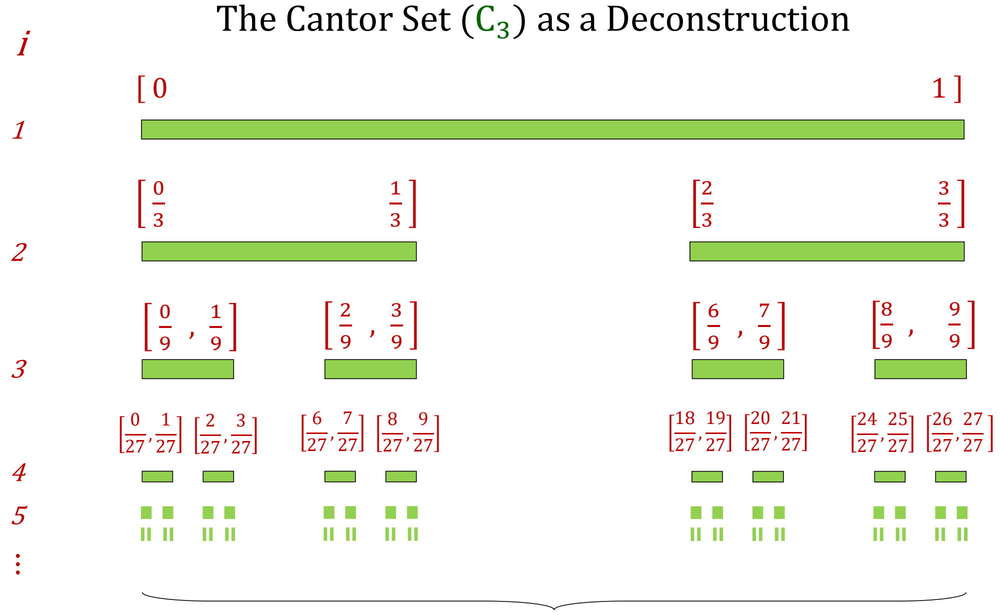
Set Theory; Thin and Fat Sets; Number Density.

Contents

1. The Cantor Set

The Cantor Set is a subset of the continuum on the real interval $[0,1]$. One way to obtain this set is to start with the fully populated continuum and remove the middle-third open sub-interval, then for each remaining interval, again remove the middle-third open sub-interval *and so on* forever over the infinite totality of all such iterations. The set of real points that remain are the points of the Cantor Set, here designated C_3.

The Cantor Set (C_3) as a Deconstruction



C_3 contains the infinitely thin froth and bubble
that remains after continuing the process *forever*.

The Cantor Set contains all the rational end-points of any and every segment because the removal of any *open* middle-third always leaves such points in place. The surprise is that C_3 also contains an infinite number of other reals, including *irrationals*.
Cantor's own description defines the elements of C_3 as infinite ternary bit-strings:

$$c_j = \frac{c_{j1}}{3^1} + \frac{c_{j2}}{3^2} + \cdots + \frac{c_{jk}}{3^k} + \ \ldots \quad \text{where } c_{jk} = 0 \text{ or } 2 \qquad \forall\, k = 1,2,3, \ldots$$

whereby a real is contained in the Cantor Set if and only if there exists a ternary representation consisting entirely of *zeros* and *twos*. But this also means that the elements of the set of reals and the set C_3 can be placed into a one-to-one correspondence. For example, the binary real $^1/_{2\pi} = 001010001011 \ldots$ will correspond to the ternary real $002020002022 \ldots$. By definition, this makes the two sets equivalent. This is the surprise of the Cantor Set C_3. It has the same cardinality as the continuum.

A key feature of C_3 is that it is *sparse*, and we might therefore be surprised that it has the same cardinality as the set of *all* reals. A comparison will help us appreciate why we should be *astonished*.

2. One Thing (The Vanishing Cantor Set)

This is a general differentiation model for a real quantity v_j on the unit interval:

$$v_j = \frac{v_{j1}}{d_1} + \frac{v_{j2}}{d_1\, d_2} + \frac{v_{j3}}{d_1\, d_2\, d_3} + \cdots + \frac{v_{jk}}{d_1\, d_2\, d_3 \ldots d_k} + \cdots$$

$$\text{where } v_{jk} \in \{0, 1, 2, \ldots (d_k - 1)\} \quad \text{and } 1 < d_{k-1} \le d_k$$

In the case of the Cantor Set d_k is always 3 and we have:

$$v_j = \frac{v_{j1}}{3} + \frac{v_{j2}}{3.3} + \frac{v_{j3}}{3.3.3} + \cdots + \frac{v_{jk}}{3.3 \ldots 3} + \cdots \qquad \text{where } v_{jk} = 0 \text{ or } 2$$

The set of reals of this form (the Cantor Set C_3) may feel pretty thin, but there are subsets of the reals that are *much* thinner. Consider the set C_V of all elements:

$$v_j = \frac{v_{j1}}{d_1} + \frac{v_{j2}}{d_1\, d_2} + \cdots + \frac{v_{jk}}{d_1\, d_2\, d_3 \ldots d_k} + \cdots \qquad \text{where } v_{jk} = 0 \text{ or } 2$$

$$\text{where } \quad d_1 = 3 \text{ and } d_k = \left({d_{k-1}}^{d_{k-1}} \right)! \quad \forall k = 2,3, \ldots$$

The number of differentiated pieces of the unit interval becomes *very* numerous *very* fast. Understanding C_V as a *deconstruction* of the unit interval (similar to the way that we can visualise C_3) all but the first and third segments of each new differentiation of any previously remaining pieces are removed. The first few values of d_k are:

k	d_k	
1	3	$= 3$ (*init*)
2	$(3^3)! = 27!$	$\sim 10^{28}$
3	$(27!^{27!})!$	$\sim 10^{(10^{788})}$
$\vdots$	$\vdots$	$\vdots$

The proportion of the unit interval which can contain elements of C_V becomes *vanishingly* small almost immediately. It happens *so fast* it is not possible for us to appreciate just how quickly the number of removed segments accelerates and how *thin* the set C_V becomes.

But as with C_3, the elements of the set C_V have a one-to-one correspondence with the set of *all reals* expressed in binary. Therefore the set C_V has the same cardinality as the set of all reals. If we are not *amazed* ... we should at least be *in awe*.

3. And Another (Extending the Countable Reals)

We will now consider classes of reals that we might say push the boundaries in the other direction. The Algebraic Reals A were shown by Cantor (1874) to have the same cardinality as the natural numbers. Cantor used a zig-zag traversal of the existential possibilities provided by the Fundamental Theorem of Algebra.

Algebraic numbers include all numbers which can be obtained from the integers using a finite number of integer additions, subtractions, multiplications, divisions and taking n'th roots *plus* an infinite number of additional irrationals that cannot be obtained in this way including all their rational multiples. The Algebraic Reals are *very dense*.

By definition all reals not algebraic are *transcendental*. However, we can extend Cantor's approach to enumerate beyond the algebraic reals and include a very large class of transcendentals. By the Gelfond-Schneider Theorem (1934) :

> If a and b are algebraic numbers and b is irrational
> (and $a \neq 0$ or 1) then a^b is transcendental.

By enumerating all the forms a^b using a zig-zag traversal of all the algebraic possibilities for a and all the algebraic possibilities for b we enumerate a new countable class of number E_{GS} which includes the resulting Gelfond-Schneider transcendentals. This is the same approach as Cantor's enumeration of the rationals using a zig-zag traversal of all natural number pairs, laid out as an infinite table.

$\mathbf{E}_{\mathbf{GS}}$	$\boldsymbol{a^b}$	b	b_1	b_2	b_3	b_4	…
	a						
	a_1		e_1	e_2	e_4	e_7	…
	a_2		e_3	e_5	e_8		…
	a_3		e_6	e_9			…

And we could even go further using other parameterized forms and repeated applications of the same method. We could for example, enumerate finite linear combinations of reals from E_{GS} or re-apply Gelfond-Schneider all over again.
Such *Extension* classes E are *extremely* dense but nevertheless *countable*.

4. Independent Ideas

A century after Cantor there is hardly a text on general mathematics that does not enthusiastically single out his most famous result as it applies to the existence and cardinality of the real numbers, variously stated as:

The number of real numbers (quantities)
on the unit interval is *uncountable*.

In a well-defined sense real quantities form a
distinctly larger infinity than the infinity of rationals.

Real quantities are *all irrational* except for a
vanishingly small proportion of rational exceptions.

Real quantities are *all transcendental* except for
a *vanishingly small* proportion of algebraic exceptions.

But such statements can be problematic. Many of them imply, intentionally or otherwise, that there is a relationship between uncountability and *proportion*. This is clearly misleading. And we have always known this. After all, the primes form a vanishingly small proportion of the natural numbers even though they can be placed into correspondence and therefore have the same transfinite cardinality $\aleph_0$.

If we arrange a simple hierarchy of some of the number classes according to their density on an interval and note their countability status, we might write something like:

R	Reals	(*maximally* dense)	*Uncountable*
E	Extension Reals	(*extremely* dense)	*countable*
A	Algebraic Reals	(*very* dense)	*countable*
Q	Rationals	(dense)	*countable*
C_3	Cantor Set	(sparse)	*Uncountable*
C_V	Vanishing Cantor Set	(*very* sparse)	*Uncountable*

The set C_V is *nowhere dense* and the reals E is *everywhere extremely dense*.
The uncountability of C_V in juxtaposition to the countability of E is astonishing.

Our initial intuition might be that something is seriously wrong. But there it is.
The cardinality of a subset of real points is independent of its density on the interval.

5. An Intrinsic Property or Just a Categorization ?

Countable or uncountable, thin or fat, it matters not. The cardinality distinction between the two transfinite cardinals $\aleph_0$ and $\aleph = 2^{\aleph_0}$ has nothing whatsoever to do with what we normally regard as proportion or density on the number line.

Imagine a dart board of all points whose radial position is one of the elements of the classes C_V and E from the unit interval. Do we have a greater chance of hitting an element of C_V with a dart throw ? Maybe this is unfair because it relies on spatial distribution. It also might be argued that there is no chance of hitting an element of *either* class. So alternatively and better (and to give set theory its due) imagine a bag containing all the reals on the unit interval from the classes C_V and E . Convention says that the elements of C_V form an uncountable population that is effectively *all* of the reals in the bag and the elements of E represent a countable and vanishingly small proportion of the reals in the bag. But would you bet real money that a single selection from the bag would return an element of C_V ? Maybe you would.

How many of us can say, hand-on-heart, that we intuitively *understand* why some *extremely* sparse sets (the vanishing Cantor set C_V) are uncountable while others that are *really dense* (extensions E of the algebraic reals) are uncountable ?

Of course the Density paradox is not actually a paradox. It is the natural outcome of definition. Nevertheless, it is a *puzzle* for anyone who cannot supress the feeling that the first demonstrated transfinite cardinality distinction should *somehow* be related to the idea of number density, however indirect or opaque that connection might be.

So although - as seems almost obligatory to be asserted in popular texts - we might accept that the reals are *infinitely more numerous* than the rationals … one is nevertheless left with the overriding sense that countability versus uncountability is all about whether or not there exist finitely parameterized definitions for classes of number *rather* than whether they are more or less *numerous* in some even remotely normal sense of that word. How certain can we be that the cardinality of infinite classes of number is *an actual property* of the classes themselves - some measure of their intrinsic numerosity – and not just a *categorization* of number based on whether or not there exist finitely parameterized algorithms for their generation ? The distinction is only unimportant if one already assumes equivalence is a good measure of "size" or if one lacks philosophical heart. For the realist, it should be a question full of meaning.

In the world of transfinite cardinality, ramblings on number density, dart-board hits, proportions and probability is just a pouring from the empty into the void. Transfinite cardinality is unrelated to any normal concept of proportion or density over an interval and no understanding of it can obtained from probabilistic analogies.