Eventually Periodic Orbits and the Vitali Set

Linas Vepstas

28 November 2023

Abstract

The Vitali set is a canonical example of a non-measurable set. It's description requires the use of the Axiom of Choice. The classical approach is to demonstrate a quotient of the reals by the rationals. A sophisticated approach is to employ the language of Borel sigma-algebras, and demonstrate a quotient, a Borel equivalence relation, that does not admit a measure. A middle ground is to work with the Bernoulli shift on the Cantor space. This provides a concrete, direct and easy-to-describe system to work with, while retaining the essential properties of the abstract approach. This approach indicates exactly why the Axiom of Choice is unavoidable: the cosets of the quotient space are not well-founded and thus do not have any unique, distinct element that can be described. Lacking the ability to isolate a specific element, one is instead forced to declare that surely, there must be some element and that it can be chosen. (Heh. It is tempting to joke "almost surely", but therein lies the rub: there is no "almost" without a measure.)

1 Introduction

Dynamical systems generically have two classes of orbits: the periodic orbits, and the chaotic ones. In between lie the eventually-periodic orbits: the ones that eventually settle down to periodic behavior, but not before a preliminary bout of irregular motion.

It came as a surprise to the author that, in an attempt to write down the class of eventually-periodic orbits, these were revealed to form the Vitali set. Perhaps this is well-known, but it is new to me. In retrospect, it is both clear and painfully obvious, and should be well-known. This text belabors the painfully obvious so as to make it perfectly clear. As to well-known, this is left to the vicissitudes of internet readership.

The Vitali set is famous for being the first example of an unmeasurable set, a set that cannot be assigned a size according to conventional measure theory. It has a more famous cousin, the Banach–Tarski paradox. The classical Vitali set is demonstrated as a quotient space $\mathbb{R}/\mathbb{Q}$ of the reals $\mathbb{R}$ by an equivalence relation $\sim$ that assigns all of the rationals $\mathbb{Q}$ to the same equivalence class. Besides being unmeasurable, it requires the use of the Axiom of Choice in order to talk about the cosets in the quotient. These properties, of non-measurability and the need for the axiom of choice, are shared by it's cousin the Banach–Tarski paradox.

Two variants of the Vitali set construction are presented below. The first is a sketch in the language of Borel sigma algebras; it is quite general, and is sufficient to present the general idea. But it is only a sketch, and lacks mathematical rigor. The second is a detailed construction making use of the Bernoulli shift on the Cantor space. This has multiple desirable properties: each step can be concretely and precisely constructed, right up to the point where the axiom of choice must be invoked. The precise and exact nature of the axiom of choice can be exposed and articulated, laying bare as to why it is unavoidable in the demonstration. The Bernoulli shift is also appealing, because it is a simple prototype for an ergodic measure-preserving dynamical system, and so is able to make contact with the broader theory of dynamical systems.

In this construction, the Cantor space 2^ω is taken as the space of all infinitely-long strings of binary digits. The Bernoulli shift acts by lopping off one digit from the front, with each iteration. Thus, specific points $p \in 2^\omega$ orbit about the space. Some orbits are periodic, some are eventually periodic, and some are chaotic, or, more precisely speaking, ergodic. The Cantor space has an invariant measure; it is uniform on the space. The uniform measure maps to the reals with the canonical map $\sum_n b_n 2^{-n}$ for a string of binary digits b_n. It also has a nonuniform measure, the Minkowski question mark measure, that maps it as a run-length encoding into Baire space, and as a map that takes the eventually-periodic orbits to the rationals. Neither of these measures are directly relevant to this text; rather, they illustrate how the Cantor space makes contact with other Polish spaces.

The quotient construction demonstrated in this text is one that assigns all eventually-periodic orbits to the same equivalence class. This set is countable, and can be identified with the rationals, after making use of the Minkowski question mark function. Thus, one can argue that the construction presented here is absolutely identical to classical Vitali set construction. The difference, and I find this to be an edifying difference, is that, by working explicitly with the Cantor space, once avoids assorted confusions and pratfalls regarding the reals. This includes such red-herrings as the half-open topology, or the need to ponder how the rationals are dense in the reals. One does not need to dive into the Hausdorff or T_1 separation axioms, nor any of the other tropes of classical general topology. One does not even need to define a Polish space. The Cantor space is discrete and totally disconnected; one can mostly avoid arcane topological issues, or, at least, push them off to a distance where they don't hurt.

This text has two primary divisions. First, an extremely rapid and condensed review of sigma algebras, measure theory and dynamical systems is presented. This review serves only to establish just enough vocabulary so that the idea of a Borel equivalence can be made concrete, followed by a short sketch as to why some might not admit a measure. The sketch is only a sketch; it lacks rigor. The second part repeats the general ideas, this time, setting them within the context of the Bernoulli shift on the Cantor space. The result is a detailed concrete demonstration of the Vitali set. Much of the effort is expended on exposing exactly how and why the axiom of choice must appear in the demonstration. A brief commentary follows about scaling limits.

2 The general setting

This section provides a lightning review of the vocabulary of measure theory and dynamical systems. The vocabulary is just barely sufficient to present the general idea of a non-measurable Borel equivalence relation. This is followed by a critique of just how this isn't rigorous enough to be anything more than an illustration.

2.1 Definitions and sketch

A topological space X is a collection $\mathscr{T}(X)$ of open sets in X that is closed under the operation of countable unions. A sigma algebra $\mathscr{B}(X)$ is a collection of sets in X that is closed under the operation of countable unions and complementation. The smallest (coarsest) sigma algebra compatible with the topology is called the Borel sigma algebra. Elements $A \in \mathscr{B}(X)$ are termed Borel sets.

A Borel measure is a function $\nu : \mathscr{T}(X) \to \mathbb{R}$ that assigns a non-negative real number to each open set in the topology; as a sigma-additive measure, it also distributes additively over disjoint sets: $\nu(A \cup B) = \nu(A) + \nu(B)$ whenever $A \cap B = \varnothing$. The Lebesgue measure extends this to closed sets, that is, to the whole of the sigma algebra, and not just the topology.

Note, however, the number of Lebesgue measurable sets is strictly greater than the number of Borel sets; the number of Borel sets has the cardinality of the continuum $\aleph_1 = 2^{\aleph_0}$, while the number of Lebesgue measurable sets is $2^{\aleph_1}$.

The action of a monoid G acting on the sigma algebra $\mathscr{B}(X)$ is a collection of functions $T_g : \mathscr{B}(X) \to \mathscr{B}(X)$ for each $g \in G$ that commute with the monoid product, so that $T_g \circ T_h = T_{gh}$. The action is free if $T_g A = A$ implies that $g = e$ the identity element in the monoid. The action is transitive if, for every $A, B \in \mathscr{B}(X)$, there exists a $g \in G$ such that $T_g A = B$.

An invariant measure is a measure μ that does not change under the action of such a monoid. That is, for each $A \in \mathscr{B}(X)$, one has that $\mu(T_g A) = \mu(A)$. The orbits of the monoid are confined to a single slice of the measure; equivalently, the measure defines an equivalence relation $\sim$ on the sigma algebra, such that $A \sim B$ iff $\mu(A) = \mu(B)$. In physics, this collection of level sets are called "the canonical ensemble". If the monoid is one-dimensional, then it's action can be interpreted as "the passage of time", and the result is called a "measure-preserving dynamical system". If the monoid is a group, then the invariant measure is called the Haar measure.

The product topology on $X \times Y$ is the coarsest topology that allows the projection functions $\pi_1 : X \times Y \to X$ and $\pi_2 : X \times Y \to Y$ to be continuous.

Given some topological space X, a Borel equivalence relation E is a Borel set in $X \times X$, in the product topology.

An equivalence relation E on a space X allows the definition of a quotient space X/E consisting of cosets of equivalent elements. For a given point $x \in X$, the coset $x_E \in X/E$ is defined as $x_E = \{y \in X : xEy\}$. The quotient map is an exact sequence $0 \to X \to X/E \to E \to 0$. Cosets are pairwise disjoint, so that for $x_E \neq y_E$ one has $x_E \cap y_E = \varnothing$.

The exact sequence states that all cosets are isomorphic to one-another, and are isomorphic to E. If one were able to assign a "size" to each coset, that is, a real-valued invariant measure, then each coset would have the same "size". The idea is that such a measure would need to respect the isomorphism.

The inability to assign such a measure arises when X/E is countable; that is, when there are a countable number of cosets. The cosets can then be labeled with natural numbers n so that there is one $x_n \in X/E$ for each $n \in \mathbb{N}$. The problem arises in attempting to define an invariant $\mu(x_n)$, since the idea of invariance implies that $\mu(x_n) = \mu(x_m)$ for all m, n. By abuse of notation, $\mu(x_n) = \mu(E)$. Measures must be bounded, and yet, clearly the sum $\sum_n \mu(x_n) = \aleph_0 \mu(E)$ demonstrates that $\mu(E) = 0$. To be explicit: this is in direct contradiction to the idea that the cosets make up all of the space, so that $X = \bigcup_n x_n$ with $\mu(X) = 1$. Since the cosets are disjoint, with $x_n \cap x_m = \varnothing$ when $n \neq m$, the sigma-additive property of the measure implies a contradiction: $1 = \mu(X) = \mu(\bigcup_n x_n) = \sum_n \mu(x_n) = \aleph_0 \mu(E) = 0$.

Perhaps, in nonstandard analysis, it is possible to assign an infinitesimal size to E, but then, such a measure is no longer real-valued; it would be nonstandard-real-valued.

2.2 Critique

The above provides a fast-and-furious presentation of the issue with measurability. Unfortunately, it lacks rigor, and glosses over important details. For example, many of the definitions above are coherent only when the space X is a Polish space. To some extent, this can be glossed over, since most of the usual spaces encountered in analysis are Polish spaces; this includes the reals, including $\mathbb{R}^n$; separable Banach spaces, including Hilbert space; and, of course, the poster children of Cantor space and Baire space. Precise proofs need to take this into account.

The jump from the definition of a Borel equivalence relation to the construction of quotient spaces was also a bit hand-wavey; the disjointness of cosets, and their measurability properties should not be taken for granted, but firmly rooted with appropriate theorems.

This cannot and will not be done here. The next steps in a more precise formulation are provided by some search-engine keyword hits: the Glimm–Effros dichotomy, the Luzin–Novikov theorem and the Feldman–Moore theorem.

2.3 A more concrete approach

The critique above can be avoided by focusing on a specific, concrete example. The example used here will be the Bernoulli shift on the Cantor space 2^ω. As already noted, the Cantor space is a Polish space. The Bernoulli shift is concrete and well-understood: it is a prototypical measure-preserving dynamical system; it is conservative, and it is ergodic, demonstrating both periodic and chaotic orbits. If it has a fault, it is that it is not mixing; but mixing does not seem to be relevant to the present context.

Working with the Bernoulli shift allows all of the above concepts to be given in a concrete, precisely-defined form, through which, ideally, no smuggler's truck can be driven through. Alas, this ideal is not entirely met, either. The problem lies with the surreptitious appearance of the Axiom of Choice.

The construction of the Vitali set necessarily requires the use of the axiom of choice, so as to select a representative element from each coset. A fair amount of effort is devoted below in illustrating exactly why this is hard, and how this is a result of a fundamental ambiguity in labeling the elements of the coset. More precisely, the elements of a coset have no natural label; they cannot be placed into a well-founded order relation that selects a single, unique element in the coset. Lacking the ability to clearly, completely and unambiguously identify a specific element in the coset, one is forced to use the axiom of choice. I almost wrote "forced to select one at random" but therein lies the rub: there's no conception of randomness that can be applied, as there is no measure. The axiom of choice rears it's (ugly?) head whenever one has a set that one can show is nonempty, but one does not have any function capable of reliably, repeatedly selecting a specific element from that set. Absent such a function, one is forced to fall back to the axiom of choice, to select "any old member of the set, cause we know there must be one."

The meta-study would then need to be able to articulate certain questions: "under what conditions is it impossible to have a function which selects a unique member from a set?". Equivalently, "when does a set not admit a unique labeling of it's contents?" Equivalently, "when does a category fail to have a unique morphism from the initial object to a specific object, thus uniquely selecting that object?". Whenever a unique morphism fails to exist, the axiom of choice must be invoked to select some arbitrary element from the known-to-be-nonempty category.

Based on the work below, a necessary condition seems to be the lack of a well-founded order on the set. If one had an order, and if that order had a unique minimal element, then that unique minimal element could be used as a label, from which the rest of the set can be generated via a free, transitive action. Is this a sufficient condition? What if one has a well-founded order, but not a free, transitive action? The questions here spiral out of control.

There are two possibilities to arrive at a more precise formulation. One is to rephrase the construction below, but this time explicitly using first-order logic, so that each step of the construction becomes a predicate. That predicate can then be examined to see which axioms of set theory were required to state it. It can be assigned to a specific level in the Borel hierarchy. This is a large task, and would inflate this text to twice it's current size.

Another alternative is to return to the literature. The ultrafilter lemma and the Boolean prime ideal theorem are, in a sense, equivalent to each other. Compactness properties of subsets of the Cantor space provide insight into the structure of each. Either can be used to provide a weaker form of the axiom of choice. But, again, to deploy this, the construction below would need to be formalized using first-order logic.

With all these caveats, disclaimers and apologies out of the way, let's proceed and see where we get.

3 Vitali sets inside the Cantor Space

This section reviews the construction a certain class of quotient spaces that can legitimately be called Vitali sets. The word "legitimately" is used, because the construction does not proceed through the conventional quotient of $\mathbb{R}/\mathbb{Q}$. The end result, however, is isomorphic to it.

The demonstration proceeds by creating quotient spaces, obtained by equating all eventually-periodic sequences. The prototypical example is the quotient space $E_0 = \Delta/\mathbb{D}$ of the Cantor space $\Delta = 2^\omega = \{0,1\}^\omega$ modulo the space of finite length strings $\mathbb{D} = 2^{<\omega}$. The equivalence relation generates cosests that consist of infinite-length binary strings that differ at a finite number of locations.

The set E_0 can legitimately be called "the Vitali set". That this is an appropriate name can be sketched as follows. Note that the cardinality of Δ is $\aleph_1$ while the cardinality of $\mathbb{D}$ is $\aleph_0$ and so we conclude that E_0 consists of $\aleph_1$ cosets, each of which contains $\aleph_0$ members. One then defines a set of points V_0 by choosing one representative from each coset. Finally, one maps the points from V_0 to the unit interval of the real numbers by applying the canonical binary expansion mapping from Δ to the reals. The image of V_0 is then the conventional Vitali set.

The rest of this text expands on the above. It tends to veer into the territory of unneeded detailing and over-explaining the obvious. My apologies. The details do seem to reveal the shape of some monster lurking in the depths.

3.1 Prelude

The alert reader might notice that $\mathbb{D}$ corresponds to the dyadic rationals, whereas the usual Vital set is constructed using the rationals. No matter: if this objection arises, then the inverse $?^{-1}$ of the Minkowski question mark function $?$ can be applied, to map the dyadics to the rationals. A short side-trip to explain this remark is worthwhile. The set $\mathbb{D}$, and all of Δ, can be mapped to the unit interval $[0,1]$ of the reals, by writing $x = \sum_n b_n 2^{-n}$ for a string of binary digits b_n. The finite-length strings correspond to the dyadic rationals in this mapping. There is also a different mapping, given by the run-length encoding of the string b_n. Write a sequence $a_1, a_2, \cdots$ as the count of the number of sequential zeros appearing in the string b_n, followed by the count of the sequential ones, then zeros again. This provides a map from Cantor space to Baire space $\mathbb{N}^\omega$. Elements in Baire space can again be mapped to the unit interval; this is the continued-fraction mapping

$$y = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots}}$$

When expressed as real numbers, these two maps are related as $x = ?(y)$ where $?$ is the Minkowski question mark function. It has a property of mapping the rationals to the dyadic rationals. It has a large variety of other interesting properties, such as mapping the quadratic irrationals to the rationals (thus proving the quadratic irrationals are countable). But all of this is a grand distraction from the task at hand. This mapping is mentioned only as a prelude, reminding the reader that the Cantor space (or Baire space) can be mapped to the reals, and so the Vitali set constructions below can be likewise mapped. As we vowed not to work with the reals, further mentions will be avoided.

3.2 Finite binary strings

To provide a proper anchor for the discussion, a vast ocean of thoroughly conventional and quite tedious notation will be provided.

Let $\Delta = 2^\omega = \{0,1\}^\omega$ be the Cantor space, represented as the space of all infinite-length binary strings. The goal of this section is to more closely define the space

$\mathbb{D} = 2^{<\omega}$ of all finite-length binary strings. Although the idea is seemingly obvious, the additional notation is needed to avoid later ambiguities.

Let $b = (b_k)_{k=1}^{\infty} \in \Delta$ be an infinite-length string of binary digits b_k. Define the length of such a string as $\text{len}(b) = |b| = k$, where k is the largest integer for which $b_k = 1$; write $|b| = \infty$ if there is no such largest integer. A string is finite-length if $k < \infty$. Informally, the finite-length strings are exactly the ones for which all digits are zero after a certain point.

Use this to define the set of finite-length strings as

$$\mathbb{D} = \{b \in \Delta : \text{len}(b) < \infty\}$$

This defines $\mathbb{D}$ such that the subset relation $\mathbb{D} \subset \Delta$ is made explicit. The strings in $\mathbb{D}$ still have an infinite number of bits; just that almost all of them are zero. By construction, $\mathbb{D}$ is countable; the cardinality of $\mathbb{D}$ is $\aleph_0$.

There are other conceptions of length. The "one-based length" is given by $\text{len}_1(b) = |b|_1 = k$ where k is the largest integer for which $b_k = 0$; if there is no such largest integer, then $|b|_1 = \infty$. Informally, these are the strings which have a finite-length prefix, after which all digits are one. This leads to another set $\mathbb{D}_1$ defined analogously, as $\mathbb{D}_1 = \{b \in \Delta : \text{len}_1(b) < \infty\}$. It is convenient to disambiguate all these by adding a subscript 0 to the earlier definitions, so as to write $|b| = |b|_0 = \text{len}_0(b)$ and also $\mathbb{D} = \mathbb{D}_0$.

Of course, $\mathbb{D}_0$ and $\mathbb{D}_1$ are isomorphic in the obvious, conventional sense. Let's belabor this into absurdity. The binary-not provides the conventional complementation map $\neg : \{0,1\} \to \{0,1\}$ defined as $\neg : 0 \mapsto 1$ and $\neg : 1 \mapsto 0$. This extends to an involution $\neg : \Delta \to \Delta$ on the Cantor space, by bit-wise extension to the product space. This is done in the obvious, conventional manner, as the bit-wise complement: $b \mapsto \neg b = \neg (b_k)_{k=1}^{\infty} = (\neg b_k)_{k=1}^{\infty}$. This provides the desired isomorphism $\neg : \mathbb{D}_0 \to \mathbb{D}_1$.

If one is too lazy to work with an infinite number of digits, one can just drop the trailing repeated digits, and work with the finite prefixes. For this purpose, it is useful to define the set of finite-length prefixes as[1]

$$\mathbb{P}_0 = \left\{ p | p = (p_k)_{k=1}^{N} \text{ for } N < \infty \text{ and } p_N = 1 \right\}$$

and likewise $\mathbb{P}_1 = \neg \mathbb{P}_0$. For ease of notation, the subscript can be dropped, to write $\mathbb{P} = \mathbb{P}_0$. These are the "true" finite length strings; one gets $\mathbb{D}_0$ from $\mathbb{P}_0$ by appending an infinite string of zeros. Clearly, $\mathbb{D}_0$ is isomorphic $\mathbb{P}_0$, as are $\mathbb{D}_1$ and $\mathbb{P}_1$. We may as well give these isomorphisms a name. Let $\text{ext}_0 : \mathbb{P}_0 \to \mathbb{D}_0$ be defined by appending the infinite string of zeros. The inverse operation is $\text{trunc}_0 : \mathbb{D}_0 \to \mathbb{P}_0$ which truncates the trailing zeros. It is unambiguous, because len_0 is unambiguous. Likewise, $\text{ext}_1 : \mathbb{P}_1 \to \mathbb{D}_1$, this time by appending all-ones, and similarly $\text{trunc}_1 : \mathbb{D}_1 \to \mathbb{P}_1$ is defined using len_1.

[1]This set is meant to include the empty string ε, but I'm too lazy to carry this throughout the text, and so only mention it as a pedantic footnote. For this purpose, the definition can be modified to read $\mathbb{P}_0 = \{\cdots\} \cup \varepsilon$, or alternately, when $N = 0$, the restriction $p_N = 1$ is not applied. The goal of the empty string is to make sure that $\overline{0} \in \mathbb{D}_0$ where $\overline{0}$ is the string of all zeros. Similar remarks apply for the $\mathbb{P}_q$ that follow later.

3.3 Eventually periodic sequences

The above constructions generalize to periodic orbits; instead of appending zeros or ones to a prefix, one appends a repeating, periodic string. Again, we fall into a tedious funk of providing notation for the obvious. The definitions that follow are entirely conventional.

Write $\overline{0}$ for the the infinite string of zeros, so that $\overline{0} = (b_k)_{k=1}^{\infty}$ with $b_k = 0$ for all k. Define $\overline{1}$ likewise as the infinite string of ones.

The periodic orbits will be chosen as elements taken out of the set of finite strings

$$\mathbb{S} = 2^{<\omega} = \left\{ q | q = (q_k)_{k=1}^{N} \text{ for } N < \infty \right\}$$

which can be either zero or one-terminated. Clearly, $\mathbb{P}_0 \subset \mathbb{S}$ and $\mathbb{P}_1 \subset \mathbb{S}$ and $\mathbb{S} = \mathbb{P}_0 \cup \mathbb{P}_1$ and $\mathbb{P}_0 \cap \mathbb{P}_1 = \varnothing$, so that the complementation involution $\neg$ partitions $\mathbb{S}$ into two equal parts.

For any $q \in \mathbb{S}$, having length $n = |q|$, one can define a periodic orbit $\overline{q}$ as $\overline{q} = (b_k)_{k=1}^{\infty}$ where each bit is $b_k = q_{k \mod n}$.

This allows $\mathbb{D}_q$ to be defined as the set of ultimately-periodic orbits. Formally, for each $q \in \mathbb{S}$ one has an associated set of eventually-periodic orbits

$$\mathbb{D}_q = \{ p\overline{q} : p \in \mathbb{S} \}$$

This does not quite preserve the definition of $\mathbb{D}_0$ and $\mathbb{D}_1$ from before, although the result is isomorphic. The technical issue is that the prefixes for $\mathbb{D}_0$ and $\mathbb{D}_1$ were taken from $\mathbb{P}_0$ and $\mathbb{P}_1$ and not $\mathbb{S}$. This issue creates difficulties for the definition of a prefix-length function len_q which we would like to have. Similar issues arise if the prefix p just happens to end with q. This allows double-counting, as exactly the same sequence might appear two or more times, with some of the bits being partitioned into the prefix, and some to the cyclic part. On the surface, this does not seem important, but can be a potential source of confusion.

To avoid this awkwardness, define the prefix set

$$\mathbb{P}_q = \left\{ p | p = (p_k)_{k=1}^{N} \text{ for } N < \infty \text{ and } \text{tail}_{|q|}(p) \neq q \right\}$$

This makes use of the string tail function $b = (b_k)_{k=1}^{n} \mapsto \text{tail}_m(b) = (b_k)_{k=n-m}^{n}$ that removes the leading bits from b, leaving only the final m bits. Clearly, this is only defined if $|b| > m$; otherwise, take the result to be the empty set $\varnothing$. As before, this definition of $\mathbb{P}_q$ is meant to include the zero-length empty string ε among it's elements. This definition of $\mathbb{P}_q$ does preserve the definition of $\mathbb{P}_0$ and $\mathbb{P}_1$ from before; its the same definition.

This allows a definition of $\mathbb{D}_q$ without double-counting:

$$\mathbb{D}_q = \{ p\overline{q} : p \in \mathbb{P}_q \}$$

As before, one can provide explicit isomorphisms $\text{ext}_q : \mathbb{P}_q \to \mathbb{D}_q$. It also allows an unambiguous definition of $\text{trunc}_q : \mathbb{D}_q \to \mathbb{P}_q$ as the leading bits of of $b \in \mathbb{D}_q$ that do not end with q. Similarly,$\text{len}_q : \mathbb{D}_q \to \mathbb{N}$ is length of the leading bits, with all trailing instances of q removed.

As before, the cardinality of $\mathbb{D}_q$ is $\aleph_0$. If the first bit of q is a zero, then we have an unambiguous isomorphism between $\mathbb{D}_0$ and $\mathbb{D}_q$; likewise, if the first bit of q is a one, then we have an unambiguous isomorphism between $\mathbb{D}_1$ and $\mathbb{D}_q$. These isomorphisms commute with trunc_q so that $\mathbb{P}_q$ and $\mathbb{P}_0$ are isomorphic if the first bit of q is a zero, and likewise $\mathbb{P}_q$ and $\mathbb{P}_1$ are isomorphic if the first bit of q is a one.

3.4 Cyclic permutations

By convention, orbits do not have a unique starting place, and so a proper enumeration of orbits would first organize finite binary strings into Lyndon words, then define a collection of prefixes that exclude cyclic permutations of Lyndon words in the tail, and only then define the set of ultimately-periodic sequences for a given orbit. This would mirror the above collection of definitions, modifying them as needed. For now, there does not seem to be any point in writing this out in detail.

3.5 The Vitali quotient space

The above provides more than enough machinery to expand on the sketch of the Vitali quotient space given in the introduction. The space E_0 is will be defined as the quotient space $E_0 = \Delta/\mathbb{D}$, where the cosets are given by all infinite-length binary strings that differ at a finite number of locations. Just a few more very simple definitions are needed to make this idea fully coherent.

The quotient space $E_0 = \Delta/\mathbb{D}$ is to be written with respect to an equivalence relation $\sim$ such that $a \sim b$ iff $a \oplus b \in \mathbb{D}$ where $a \oplus b = (a_k \oplus b_k)_{k=1}^{\infty}$ is the bit-wise compare (alternately, the exclusive-or, xor, or the symmetric difference). That is, $a_k \oplus b_k = 0$ if $a_k = b_k$, else its one. Thus, when writing $E_0 = \Delta/\mathbb{D}$ it is meant $E_0 = \Delta/\sim$ for this equivalence relation $\sim$. Likewise, define $E_1 = \Delta/\mathbb{D}_1$ where $a \sim_1 b$ iff $a \oplus b \in \mathbb{D}_1$, and $E_q = \Delta/\mathbb{D}_q$ where $a \sim_q b$ iff $a \oplus b \in \mathbb{D}_q$.

One can then walk just a bit further down this path, and consider an equivalence relation where two ultimately periodic sequences are equivalent, if the periods are of the same length, and one is a cyclic permutation of the other. That is, $u \approx_p v$ iff $u \in \mathbb{D}_p$ and $v \in \mathbb{D}_q$ and $|p| = |q|$ and p is a cyclic permutation of q. Under this definition, $u \approx_p v$ iff $u \approx_q v$. The wisdom of doing this, without carefully defining sets of cyclically-permuted orbits is somewhat questionable, as one does run the risk of double-counting some of the sequences. The utility of doing this, for the present text, seems to be absent.

An argument that the E_q can be validly called "Vitali sets" was already made in the introduction. A minor expansion of this argument can be made by observing that the cardinality of E_q is $\aleph_1$ and the cardinality of each coset is $\aleph_0$. The latter is a trustworthy conclusion, in part because of the care taken to avoid double-counting in $\mathbb{D}_q$. This establishes that the cosets are each strictly isomorphic to $\mathbb{D}_q$.

3.6 Total orders

In order to have further meaningful conversations about E_q and it's cosets, we would like to have some means of enumerating all of the elements of E_q, ideally placing them into some order, and likewise, a way of enumerating the elements of any given coset in E_q, again, ideally by placing them into some order. To get to there, we have to step back and look at orders on Δ and on $\mathbb{D}_q$.

The Cantor set is totally ordered. That is, for any two $a,b \in \Delta$ one can always determine if $a < b$ by performing a bitwise comparison. Define $a < b$ as true, if there exists an integer N such that $a_k = b_k$ for all $k < N$ and $a_N = 0$ while $b_N = 1$. If there is no such integer, then we say that $a \geq b$. There are several equivalent ways of saying this. One is to say that the Cantor space is metrizable, with metric $g(a,b) = 1/N$ where N is the smallest integer for which $a_N \neq b_N$. Equivalently, it is the length (plus one) of the initial run of zeros in $a \oplus b$. We even know what the greatest and least elements of Δ are: this is easy, they are $\overline{0}$ and $\overline{1}$.

The total order is inherited by the $\mathbb{D}_q$, where, in some sense, it is "even easier", mostly because we know that the elements of $\mathbb{D}_q$ are of (effectively) finite length. Yet, it is also "more difficult", because $\mathbb{D}_q$ does not contain a least element. Write $0^{(n)}$ for the string of n zeros. Then $0^{(n)}1\overline{q} \in \mathbb{D}_q$ for all n. But the total order, as inherited from Δ, implies that $0^{(n)}1\overline{q} < 0^{(m)}1\overline{q}$ whenever $n > m$. Worse,$\mathbb{D}_q$ is not complete, as the limit point of the sequence $0^{(n)}1\overline{q}$ would seem to be $\overline{0}$ but $\overline{0} \notin \mathbb{D}_q$. A different issue is that, in saying "limit point", we assumed that the conventional rules about the limits of sequences should be applied, but we haven't yet firmly stated what the word "limit" means, in the present context. Conventional rules apply when things can be metrized, but absent a metric, it's not clear how something can be close to something else. There are, of course, other definitions of the limits of sequences. Those definitions require the concept of an open set, and the neighborhood of a point. So far, we have not defined either open sets, or neighborhoods. There's no topology. This is not an accident; its intentional. We are staring into the gaping maw of a monster, and so will be a bit more careful.

The primary issue is that the bit-wise-defined order on Δ carries with it some implicit assumptions about distance and convergence. Halting comparisons upon the first bit to mismatch makes the implicit statement that all later bits do not matter. Were they changed in arbitrary ways, one would still have, in some sense, that $0 < |a-b|$, which implies the existence of some sort of metric $|a-b|$. It can be taken to be any decreasing function of of N, as long as it says that bits later than N do not matter for order comparison, once the first mismatch $a_N \neq b_N$ has been found. This has the implicit side-effect of metrizing the space Δ. It also has the implicit effect of defining convergence: some sequence converges to a point whenever more and more bits in the sequence match. This notion can be done without explicit appeal to a metric, yet it unavoidably forces the implicit assumption of one, by saying that points are "closer together", the more bits match.

Is there anything wrong with this? Well, in the conventional sense, no. Writing $|a-b| = 2^{-N}$ just reduces (eventually) to the conventional metric on the reals, the conventional open sets, and so forth, topologizing Δ with the weak topology (the product topology). Yet, the present discussion is attempting to talk about the Vitali set, and so topologizing prematurely threatens to ruin the ability to perform later inference.

Fortunately, the lexicographic order is available. For this, recall that $\text{trunc}_q : \mathbb{D}_q \to \mathbb{P}_q$ is an isomorphism, and the lexicographic order is entirely unproblematic for $\mathbb{P}_q$, as long as shorter prefixes precede longer ones, and that only then is the bitwise compare performed. Write this order as $<^{\text{lex}}$. Just to be tediously precise, $a <^{\text{lex}} b$ iff $|a| < |b|$ or if $|a| = |b|$ and $a <^{\text{bit}} b$ with $<^{\text{bit}}$ being the earlier-defined bitwise ordering. By isomorphism, $<^{\text{lex}}$ is well-defined on both $\mathbb{P}_q$ and $\mathbb{D}_q$. It can be extended to all of Δ, if the length function $\text{len}_q(b) = |b|_q$ is used, so that the length is measured only after removing a trailing $\overline{q}$, if any. There will only be a countable number of $b \in \Delta$ with a trailing $\overline{q}$; these are precisely the strings in $\mathbb{D}_q$. For all the other (uncountably many) infinite-length strings, use the bitwise compare $<^{\text{bit}}$ just as before. This extends $<_q^{\text{lex}}$ to all of Δ. The subscript q is used once again to remind us that len_q is to be used for measuring length for the lexicographic sorting. The ordering $<_q^{\text{lex}}$ has the interesting property that it places all of $\mathbb{D}_q$ before any string not in $\mathbb{D}_q$.

The ordering $<_q^{\text{lex}}$ is well-founded on $\mathbb{D}_q$, in that the minimum element $\overline{q} \in \mathbb{D}_q$ and all other elements in $\mathbb{D}_q$ are no more than a finite number of steps above $\mathbb{D}_q$. Specifically, they are all len_q steps away, and, of course $\text{len}_q\overline{q} = 0$. By contrast, this ordering is not well-founded on Δ; all elements in Δ but not in $\mathbb{D}_q$ are at least a countably infinite number of steps away, and usually more. The bitwise order $<^{\text{bit}}$ is not well-founded; in particular, it places an uncountably infinite number of elements underneath $\overline{q}$.

3.7 Orders on the Vitali set

Neither of the orders discussed above extend easily or naturally to either E_q or any of it's cosets. Consider first the case of some coset $\gamma \in E_q$. It has a countable number of elements, but the bit-order $<^{\text{bit}}$ cannot be used to find a least element, since the order may converge to a limit point outside of the coset. The lexicographic order $<_q^{\text{lex}}$ cannot be applied, since all strings appear to be of infinite length to len_q. From the point of view of len_q, the strings in γ seem to be "encrypted": they are elements of $\mathbb{D}_q$ that have been xor'ed with some unknown $u \in \gamma$. That is to say, if one were to select some specific $u \in \gamma$ (thus applying the axiom of choice), then all other $v \in \gamma$ could be "decrypted" by computing $u \oplus v \in \mathbb{D}_q$. Each selection of $u \in \gamma$ provides a distinct mapping $u : \gamma \to \mathbb{D}_q$ given by $u : v \mapsto u \oplus v$. There are a countable number of such mappings, they are clearly all inequivalent; the points in the coset can be distinguished from one another, but they cannot be labeled without selecting an order, and selecting an order requires selecting a distinguished $u \in \gamma$, bringing us back to the axiom of choice. There doesn't seem to be any function, natural or unnatural, that provides a distinguished element $u \in \gamma$.

There is another, distinct possibility. Suppose the quotient was being constructed "algorithmically", one point at a time. Then, perhaps, the very first point to be added to a coset could serve as the distinguished label u. Later points do not even have to be added, since $\gamma = u \oplus \mathbb{D}_q$ and so simply having one distinguished point in the coset is enough to define the entire coset. Is such an "algorithm" achievable? In the narrow sense, no, since clearly conventional finite algorithms cannot enumerate uncountable sets. Let's ignore this minor inconvenience for a moment, and assume it was possible, for some suitably defined uncountably-long runtime. The goal of such an algorithm is to enumerate the cosets; for that to happen, it would seem that there needs to be a way of enumerating the elements of Δ first. The total order provided by $<^{\text{bit}}$ cannot be used. The first element is the least element $\overline{0}$, but what is the next element? Can we find some way of iterating to the next element?

Perhaps this can be provided by using the ergodic properties of the Bernoulli shift applied to the Cantor set. The Bernoulli shift is the shift operator $T : \Delta \to \Delta$ acting on individual points as $T : (b_k)_{k=1}^{\infty} \mapsto (b_k)_{k=2}^{\infty}$. The shift is ergodic in all the conventional definitions of ergodicity. For the present purposes, it can be employed as a generator of candidate cosets. Select a point $b \in \Delta$ "at random". This has two issues: what do we mean by "select"? The axiom of choice, I guess. What do we mean by "at random"? Choosing from the uniform distribution on Δ. But defining such a distribution requires defining a measure on Δ, which requires defining Borel sets on Δ, which requires topologizing Δ sufficiently to define the open sets that will become elements of Σ_1^0 of the Borel hierarchy. All this can be done; the required topologization is provided by the bitwise metric given by the first miscompare, when comparing two bit-strings. This provides the conventional weak topology on Δ, and the rest flows downhill.

With these tools, one proceeds to generate sequences of candidate points. The algorithm is as follows: select a point $p \in \Delta$ and then iterate to obtain other points $T^k p$. Then, if $\left(T^{k+1}p\right) \oplus \left(T^k p\right) \notin \mathbb{D}_q$, write $u^{(k+1)} = T^{k+1}p$ as the new, unique label for a coset $\gamma = \gamma^{(u)}$. Due to the nature of the uniform distribution on Δ, it will almost certainly be the case that the randomly selected point $p \in \Delta$ is ergodic. In the present case, this means that it will almost certainly be the case that $T^{m+n}p \oplus T^n p \notin \mathbb{D}_q$ for all non-negative integers m,n. This is a slightly stronger statement than the conventional definition of ergodicity, so I suppose I should prove that it is true. For now, I assume that it is true, and that proving it would be another tedious exercise. At any rate, running this loop once will generate at most a countable number of cosets $u^{(k)}$. It still required the axiom of choice; only that the burden was shifted from choosing base-points $u \in \gamma$ for each coset $\gamma \in E_q$, to choosing an initial random $p \in \Delta$. No matter. Only a countable number of cosets $u^{(k)}$ were generated, and thus the loop has to be run again, starting with a new randomly-chosen $p \in \Delta$, *ad infinitum*, for an uncountable number of iterations of the outer loop. This time, taking care to verify that each new coset is distinct from all the previously generated ones. Of course, this will be the case almost certainly. The check can be avoided by instead verifying that $p \oplus p_\alpha \notin \mathbb{D}_q$ for all previously selected points p_α. At any rate, we've described a process that requires algorithms that somehow run for uncountably-long periods of time, and the algo still has not magically evaded the axiom of choice. This is a good time to just give up.

The above does reveal a minor curiosity, though: if the above algo is terminated in finite time, then the resulting set of $u^{(k)}$ are uniformly distributed across the unit interval. This follows from the random draw of p from a uniform distribution. As this finite set of $u^{(k)}$ can be given the counting measure, one must conclude that the non-measurability of the Vitali set must come from repeating the construction into the limit. But of course! This will be reviewed in the next section.

What about E_q? Given that each coset $\gamma \in E_q$ carries a unique, distinguished label $u \in \gamma \subset \Delta$, it should be clear that all of the cosets γ can be totally ordered by using the bitwise total order $<^{\text{bit}}$ on the collection of u.

3.8 The Vitali set measurability paradox

It is time to recap the core argument for the measurability of the Vitali set. One starts with the Cantor space Δ and assigns a total measure of one to it. The quotient $E_q = \Delta/\mathbb{D}_q$ shatters it into an uncountable number of cosets (cardinality $\aleph_1$), each labeled by a unique string $u \in \Delta$. The construction forces that each such label is distinct: for all label pairs u, v one has that $u \oplus v \notin \mathbb{D}_q$. This allows an entire coset to be exhibited as $u \oplus \mathbb{D}_q$. Since $\mathbb{D}_q$ is countable (i.e. having cardinality $\aleph_0$), it can be indexed with integers. The indexing can be made explicit: for each $d(m,n) \in \mathbb{D}_q$ just define $k = m + 2^n$. Recall how these elements were defined: the length $\text{len}_q d(m,n) = n$, and there are exactly 2^n finite strings of length exactly n; these are labeled with $0 \le m < 2^n$. These are in one-to-one correspondence with the dyadic rationals $(2m+1)/2^{n+1}$

Each distinct label is assigned to a distinct Vitali set V_k, so that

$$V_{q;k} = \bigcup_{u \in E_q} u \oplus d(m,n)$$

is a set of disjoint points. Disjoint, simply because we've never created any topology, so disjoint by default. Each set is also pair-wise disjoint: $V_{q;j} \cap V_{q;k} = \varnothing$ for each pair of integers $j \neq k$.

By construction

$$\Delta = \bigcup_{k=0}^{\infty} V_{q;k}$$

as the point-wise union. Every point in Δ was carefully accounted for, in this construction. More specifically, each and every possible infinite string of binary digits $b \in \Delta$ has been assigned to one and only one $V_{q;k}$. There are no extras, there are no duplicates or double-counting, and nothing has been missed or forgotten.

The conventional argument proves the non-measurability of each $V_{q;k}$ with an argument by contradiction. Assume there exists a measure μ such that $\mu(\Delta) = 1$ and $\mu(V_{q;k}) = \varepsilon_k > 0$. Since all $V_{q;k}$ are isomorphic to one-another, they must all be of the same size: the measures must all be equal: $\varepsilon_k = \varepsilon$. One then concludes that

$$\begin{aligned} 1 =& \mu(\Delta) \\ =& \mu\left(\bigcup_{k=0}^{\infty} V_{q;k}\right) \\ =& \sum_{k=0}^{\infty} \mu(V_{q;k}) \\ =& \sum_{k=0}^{\infty} \varepsilon_k \\ =& \omega\varepsilon \end{aligned}$$

Clearly, there is no real number ε that can preserve this identity; thus, one concludes that each individual Vitali set $V_{q;k}$ is unmeasurable.

Written in this way, it is also clear that nonstandard analysis, using hyperreal numbers, avoids this ugly fate: just set $\varepsilon = 1/\omega$ as the infinitesimal, and the problem goes away. Of course, this opens up a can of worms: what are the hyperreals, and what are their properties? How do they behave? Is this a legitimate and valid construction? As I have no desire to write a textbook on the hyperreals, the short answer is, yes, everything is just fine. Here, ε is just an infinitesimal; it behaves just like any other number, and doesn't present any particular challenges. One can go farther, and use the surreals as well, although this is not strictly called for, at this point. Again, there's no particular problem, here.

There is also an alternative interpretation, given in the next section.

3.9 Scaling and the Renormalization Group

An earlier section attempted to provide an algorithmic construction of the Vitali set, albeit with algorithms that might take an uncountably long time to run. The construction founders on technical details, but not before noting that a finite approximation can be achieved. The finite approximation can be used as a stand-in for the infinite limit. Increasing the size/length of the finite approximation by a factor of α causes assorted quantities to scale as a function of α. These can be renormalized, by rewriting newly scaled quantities in terms of the old. This can be done arbitrarily, thus presenting the idea of a renormalization group. Using the renormalization group allows one to always work with finite quantities, which behave exactly as they should, even if the limit was taken. Thus, one does not have to deal with the "actual infinite", but only with the "potential infinite". This is just a trick, a sleight of hand, but appears to be consistent. Or rather, should be consistent; consistency would need to be proved. At the time of writing, I see no reason to doubt the validity and consistency; I'm not expecting dragons here; everything is finite, everything scales.

The procedure is simple, to the point of silliness; the big words above are used only to shock the reader into taking a more principled approach.

It works like so. A source of random binary digits was identified. In the present case, a pseudo-random generator will do; it generates bits of sufficient quality that they offer no impediment to numerical algorithms. These are used to generate a sequence of values u that are interpreted as the indicator elements of the Vitali cosets. That they are numerically distinct can be readily checked. But what does this really mean? Suppose one has two random 32-bit integers u, v Computing the xor $u \oplus v$ is trivial: the result is some other 32-bit integer. The theory insists that one must have $u \oplus v \notin \mathbb{D}_0$, where, for now, take $q = 0$ just to keep the discussion simple. Clearly $u \oplus v$ is just some 32-bit integer; it is clearly in $\mathbb{D}_0$ simply by rescaling by 2^{32}. This would seem to violate the base requirement of the construction. Well, not really. Set $N = 32$ as the physical scale, but set $M < N$ as the "computational scale". For the present example, $M = 22$ seems like a reasonably adequate scale to work with. Then, the requirement that $u \oplus v \notin \mathbb{D}_0$ translates into the idea that the ten bits 23 to 32 are not zero. That is, all bits from M to N lie in the "scaling continuum". If we have a number x and discover that bits 23-32 are all zero, we conclude that $x \in \mathbb{D}_0$, else assume $x \notin \mathbb{D}_0$. Since there are 10 bits between 23 and 32, the chance that a random number is dyadic becomes one in $2^{10} \approx 10^3$. That is, the chance of being wrong, and of accidentally misclassifying a real number as a dyadic rational is about 1 in a thousand. The chance of mis-classifying in the opposite direction is zero: any the dyadic rational will have bits 23-32 all zero, and there is no chance of accidentally getting that wrong.

Consider now the act of simulating the Vitali set. With the above parameter choices, we are allowed a total of $2^N \approx 4 \times 10^9$ distinct real numbers, and so we conclude that $\omega = N = 32$ at this computational scale. Not very large! If exactly the same representational system is used to represent the dyadic rationals, then we can have $2^M \approx 4 \times 10^6$ distinct dyadics. This follows only because dyadics smaller than 2^{-M} cannot be distinguished from reals. Put differently, the size of $\mathbb{D}_0$ is 2^M. Since the size of $\mathbb{D}_0$ is supposed to be ω, we have a clashing estimate that $\omega = 2^M \gg N$. What is the size of each Vitali set? By construction, it is of size $\varepsilon = 1/|\mathbb{D}_0| = 2^{-M}$. From the nonstandard analysis, we concluded that $\omega\varepsilon = 1$ and so again $\omega = 2^M$.

The above considerations indicate that there are two independent scaling factors: $\mathfrak{c} = 2^N$ and $\omega = 2^M$ and that it is a mistake to conflate $\log_2 \mathfrak{c}$ with ω. To have renormalization work correctly, these need to be tracked distinctly. In the scaling limit, we want to arrive at $\mathfrak{c} = 2^\omega$ as the limit, and so, for $N = 32$, setting $M = 5$ would preserve the scaling. For $N = 64$,one has $M = 6$ which really shows just how small a "set of measure zero" really is, in gut-sense terms. For practical calculations, there is no particular reason to adhere to the scaling limit; who wants to count up to only 32 or 64? It's OK to have "too many dyadics" with respect to the real numbers, as long as one is aware of this, and adjusts scales appropriately.

To complete the presentation above, it would be appropriate to describe the actual group that carries sets from one scale to another. This can be done either by working directly with the construction above, or by attempting block renormalization on both the Vitali set, and the Borel measure at the same time. This will not be done here.

4 Conclusion

The construction of the Vitali set was articulated in two different ways. The introduction provided a simple abstract but general construction. It was enough to present the general idea, but lacked the mathematical rigor and precision to turn it into anything but a sketch of the general idea. The second construction resorted to a specific and concrete example: the Bernoulli shift on the Cantor space. This allowed most of the notions to be made precise. In particular, it was useful for highlighting exactly why the axiom of choice is required to construct the Vitali set. Several next steps are possible; an obvious one is to repeat the construction, this time taking the effort to found all statements within the language of effective descriptive set theory, so as to highlight all of the boundaries, and, in particular, where each of the specific statements show up in the Borel hierarchy.