

NON-STANDARD ANALYSIS

PATRICK STEVENS

<https://www.patrickstevens.co.uk/misc/NonstandardAnalysis/NonstandardAnalysisPartIII.pdf>

Licence: CC BY-SA <https://creativecommons.org/licenses/by-sa/4.0/>. You are free to share and adapt this work for any purpose, even commercially, as long as you attribute it, indicating which changes were made, without suggesting that the licensor endorses you or your use. You must distribute derivative works under the same license as the original.

Date: Composed from October 2015 to April 2016. Released on 15th June 2016.
Supervised by Dr Thomas Forster for the essay component of Part III of the Cambridge Mathematical Tripos.

CONTENTS

1. What is non-standard analysis?	2
2. Basic definitions	3
3. Derivatives	5
3.1. The chain rule	7
4. Continuity	8
4.1. Uniform continuity	9
4.2. Compactness	9
5. Hypernaturals	10
5.1. Internal sets	10
5.2. Hypersequences	12
5.3. Lattices	13
5.4. Hyper-sums	15
6. Integration	15
7. Series	20
8. More general topological ideas	24
9. Measure theory	27
9.1. Comprehensiveness	29
9.2. Alternative characterisation of Loeb measurability	29
9.3. Lebesgue measure via Loeb measure	32
10. Brownian motion	39
References	43

1. WHAT IS NON-STANDARD ANALYSIS?

Non-standard analysis is the study of a model of the reals in which there are *infinitesimals*: there is some $\varepsilon > 0$ such that, for all “standard” reals x , we have $\varepsilon < x$. We shall follow Pétry [2] in assuming the existence of a “field of hyperreals”: ${}^*\mathbb{R}$, an ordered field which extends \mathbb{R} ’s order structure and inherits its multiplicative structure. The key fact about ${}^*\mathbb{R}$ which we shall also impose is the *Transfer Principle* of Definition 1.1. The truth of the transfer principle follows from Łos’s Theorem (see Theorem 4.3 of Davis [3]). However, this essay aims for a treatment of the analysis involved, rather than getting bogged down in foundational details, so we will not prove it here.

Definition 1.1 (Transfer Principle). *Let ϕ be a first-order sentence in the language of the totally ordered field \mathbb{R} . (That is, ϕ quantifies only over real numbers, not over subsets of the reals.) Then ϕ is true in \mathbb{R} if and only if ${}^*\phi$ is true in ${}^*\mathbb{R}$, where ${}^*\phi$ is obtained by replacing all quantifiers $(\forall x \in \mathbb{R})$ with $(\forall x \in {}^*\mathbb{R})$ and $(\exists x \in \mathbb{R})$ with $(\exists x \in {}^*\mathbb{R})$.*

This version of the transfer principle is not actually the strongest possible; we will return to this in Section 5.1, where we examine certain cases under which we may quantify over sets.

There is a simpler but more imprecise form, which we state here for motivation in the form given by Pétry:

Definition 1.2 (Transfer Principle, imprecise form). *If two (first-order definable) systems of equations are equivalent in \mathbb{R} , then they are equivalent in ${}^*\mathbb{R}$. Moreover, every (first-order definable) function $f : \mathbb{R} \rightarrow \mathbb{R}$ extends to a function ${}^*f : {}^*\mathbb{R} \rightarrow {}^*\mathbb{R}$ such that ${}^*f|_{\mathbb{R}} = f$.*

In the first sections of this essay, whenever we invoke the transfer principle, all the functions and sets we will consider shall be first-order definable. Therefore this formulation suffices for the moment.

Example. To give an example of the transfer principle in action, consider the two equivalent sets

$$\{x \in \mathbb{R} : x^2 > 7\}; \{x \in \mathbb{R} : x > \sqrt{7} \vee -x > \sqrt{7}\}$$

We may define 7 in a first-order way, as $1 + 1 + 1 + 1 + 1 + 1 + 1$ (recalling that we are working in the language of the totally ordered field \mathbb{R}). Similarly, -1 has the following first-order description:

$$(\exists x \in \mathbb{R})(x + 1 = 0)$$

The square-root function may be expressed in a first-order way:

$$(\forall x \geq 0)(\exists y \geq 0)(y^2 = x)$$

so (imprecisely) $\sqrt{\cdot} : x \mapsto \sqrt{x}$ extends to a function ${}^*\sqrt{\cdot} : {}^*\mathbb{R} \rightarrow {}^*\mathbb{R}$, and (formally) ${}^*\sqrt{\cdot}$ is described by

$$(\forall x \in {}^*\mathbb{R})(x \geq 0 \rightarrow (\exists y \in {}^*\mathbb{R})([y \geq 0] \wedge [y \times y = x]))$$

which is a true statement of ${}^*\mathbb{R}$ by the transfer principle.

Then viewing 7 and -1 as elements of ${}^*\mathbb{R}$, the transfer principle states (imprecisely) that

$$\{x \in {}^*\mathbb{R} : x^2 > 7\}; \{x \in {}^*\mathbb{R} : x > {}^*\sqrt{7} \vee -x > {}^*\sqrt{7}\}$$

are equivalent sets in ${}^*\mathbb{R}$. Formally,

$$(\forall x \in \mathbb{R})(x \times x > 7 \leftrightarrow [x > \sqrt{7}] \vee [x < -1 \times \sqrt{7}])$$

and so by transfer,

$$(\forall x \in {}^*\mathbb{R})(x \times x > 7 \leftrightarrow [x > {}^*\sqrt{7}] \vee [x < -1 \times {}^*\sqrt{7}])$$

Therefore, the inequality $x^2 > 7$ has solutions $x > \sqrt{7}$ and $x < -\sqrt{7}$ whether we are working in \mathbb{R} or ${}^*\mathbb{R}$.

Remark (Existence of ${}^*\mathbb{R}$). Why are we justified in assuming the existence of ${}^*\mathbb{R}$? The spirit in which we do so is akin to the fact that we teach arithmetic while perfectly content not to define \mathbb{N} rigorously, but simply to recognise its existence. To obtain an explicit construction, we may use ultrapowers (see Robinson [1]); we will sketch this approach in Section 5.1 of this essay. Alternatively, we may understand the essence

of why a model exists via the Compactness Theorem of first-order logic. Adjoining a symbol ε to the language of the reals, and adding axioms that

$$\{\varepsilon > 0; \varepsilon < \frac{1}{n} : n \in \mathbb{N}\}$$

produces a system which has a model by compactness. Such a model contains an infinitesimal ε .

2. BASIC DEFINITIONS

We consider a non-standard model ${}^*\mathbb{R}$ of the reals. The key difference between ${}^*\mathbb{R}$ and \mathbb{R} is that every finite $r \in {}^*\mathbb{R}$ has a *standard part*, denoted $\text{st}(r)$, and a *non-standard part*. The standard part is a standard real (that is, a member of \mathbb{R}) which is “infinitely close” to r : that is, $r - \text{st}(r)$ is infinitesimal.

We write $r \simeq s$ for the relationship “ $r - s$ is infinitesimal”: explicitly, $r \simeq s$ iff either $r = s$ or for every $\varepsilon \in \mathbb{R}$ with $\varepsilon > 0$, we have

$$\begin{cases} 0 < r - s < \varepsilon & \text{if } r > s \\ 0 < s - r < \varepsilon & \text{if } r < s \end{cases}$$

Definition 2.1. A hyperreal r is finite iff there is a standard real B such that $-B \leq r \leq B$.

Caveat. Several authors draw a distinction between various different flavours of “finite”, “limited”, “appreciable” and similar terms, denoting combinations of “finite and not infinitesimal”, “finite infinitesimal” and so forth. We will not need such distinctions in this essay, so will use simply the word “finite” as in Definition 2.1, but when reading other authors, make sure to look up their terminology.

Theorem 2.1. Every finite hyperreal r has a unique real standard part.

We give the proof here as a simple demonstration of some of the operations on infinitesimals.

Proof. For uniqueness: suppose $x, y \in \mathbb{R}$ are both “infinitely close” to $r \in {}^*\mathbb{R}$. That is, for all $\varepsilon \in \mathbb{R}^{>0}$,

$$|r - x| < \varepsilon, |r - y| < \varepsilon$$

By the triangle inequality,

$$|(r - x) - (r - y)| < 2\varepsilon$$

so $|x - y|$ is less than 2ε for all ε .

(The triangle inequality is true by the transfer principle, but we will omit the easy proof of this fact. We will soon see a fully-worked example of the use of the transfer principle, in Theorem 3.1.)

Since x, y are both standard reals, this means $x - y = 0$.

For existence: if $r \in {}^*\mathbb{R}$ is finite, then there is positive $B \in \mathbb{R}$ such that $-B \leq r \leq B$. Then

$$S := \{x \in \mathbb{R} : r \leq x\}$$

is a set of reals which is nonempty (containing B), and it is bounded below (by $-B$), so it has a greatest lower bound, which we shall optimistically call $\text{st}(r)$ (for “standard”).

By construction, $\text{st}(r)$ is indeed a standard real. We will prove by contradiction that $r - \text{st}(r)$ is infinitesimal.

If $r = \text{st}(r)$ then we are instantly done, so suppose that $|r - \text{st}(r)| > \varepsilon$ for some $\varepsilon \in \mathbb{R}^{>0}$.

If $r > \text{st}(r)$, then $r > \varepsilon + \text{st}(r)$. Then in fact $\text{st}(r) + \frac{\varepsilon}{2}$ is a lower bound for S , contradicting the definition of $\text{st}(r)$ as a *greatest* lower bound.

If $r < \text{st}(r)$, then $r < \text{st}(r) - \varepsilon$, so $\text{st}(r) - \frac{\varepsilon}{2}$ lies in S , contradicting the definition of $\text{st}(r)$ as a lower bound for S . \square

This justifies the definition of $\text{st}(r)$ as *the* standard part of r .

There are many easy results about standard parts, which all have similar patterns of proof; we omit the proofs for brevity, and we may use without comment certain obvious results such as the following.

- $\text{st}(-u) = -\text{st}(u)$
- $\text{st}(u + v) = \text{st}(u) + \text{st}(v)$
- if $u \leq v$, then $\text{st}(u) \leq \text{st}(v)$.

The collection of hyperreals infinitesimally close to hyperreal r is known as the *monad* of r , and the “local properties” we study in Analysis can often be related to the study of behaviour in the monad of a point.

3. DERIVATIVES

As an example of the idea of “local properties may be defined in monads”, we give the following definition of the derivative:

Definition 3.1. *The derivative of $f : \mathbb{R} \rightarrow \mathbb{R}$ at $x \in \mathbb{R}$ is defined to be*

$$f'(x) := \text{st} \left(\frac{{}^*f(x + \delta) - {}^*f(x)}{\delta} \right)$$

for any infinitesimal δ ; f is said to be differentiable at x if this is well-defined.

By way of demonstration, consider the function $f : \mathbb{R} \rightarrow \mathbb{R}$ by $x \mapsto x^n$ (some $n \in \mathbb{N}$). By the (imprecise) transfer principle, this defines a function ${}^*f : {}^*\mathbb{R} \rightarrow {}^*\mathbb{R}$ which coincides with the standard version on standard reals. This being our first theorem proved by non-standard methods, we will walk through it in complete pedantic detail; subsequent proofs will be considerably lighter to aid comprehension.

Theorem 3.1. *If $f : \mathbb{R} \rightarrow \mathbb{R}$ by $x \mapsto x^n$, then $f'(x) = nx^{n-1}$.*

Proof. The function $f : \mathbb{R} \rightarrow \mathbb{R}$ by $x \mapsto x^n$ admits the first-order description

$$(\forall x \in \mathbb{R})(\exists y \in \mathbb{R})(y = x \times x \times \cdots \times x)$$

where there are n terms in the product. (Henceforth we will use the shorthand r^n for $r \times r \times \cdots \times r$, with n terms in the product.) This is a schema of true statements in \mathbb{R} , one for each $n \in \mathbb{N}$.

Therefore, by the transfer principle, for each $n \in \mathbb{N}$ the following sentence is true in ${}^*\mathbb{R}$, so we may use it to define ${}^*f : {}^*\mathbb{R} \rightarrow {}^*\mathbb{R}$:

$$(\forall x \in {}^*\mathbb{R})(\exists y \in {}^*\mathbb{R})(y = x^n)$$

Recall that we wish to show that

$$\text{st} \left(\frac{(x + \delta)^n - x^n}{\delta} \right)$$

is well-defined as δ varies, and equal to nx^{n-1} ; for this, we need the binomial theorem.

We will require the following shorthand: fixing n and m , we understand $\binom{n}{m}$ as being defined by the natural number

$$\frac{n!}{m!(n-m)!}$$

which is expressed as $1 + 1 + \cdots + 1$ for appropriately many terms in the summand.

Now for the binomial theorem itself:

$$(\forall x \in \mathbb{R})(\forall y \in \mathbb{R})((x + y)^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \cdots + \binom{n}{n-1}xy^{n-1} + \binom{n}{n}y^n)$$

This is again a schema of true statements in \mathbb{R} , one for each $n \in \mathbb{N}$, and each of those statements is to be understood as being written out in full with no elided terms. We emphasise that $\binom{n}{m}$ is simply a number: it does not even contain any bound variables, and if we expanded out all of the shorthand, it would appear to be in the form $1 + 1 + \cdots + 1$.

Since the binomial theorem holds in the reals, every statement in that schema must be true in the hyperreals when we transfer to ${}^*\mathbb{R}$:

$$(\forall x \in {}^*\mathbb{R})(\forall y \in {}^*\mathbb{R})((x + y)^n = \binom{n}{0}x^n + \binom{n}{1}x^{n-1}y + \cdots + \binom{n}{n-1}xy^{n-1} + \binom{n}{n}y^n)$$

Notice that behind the scenes, n is the same $1 + 1 + \cdots + 1$ as it always was, so there is no transfer required in the definition of $\binom{n}{m}$.

Therefore, in particular, we have

$$\frac{(x + \delta)^n - x^n}{\delta} = nx^{n-1} + \delta \cdot \binom{n}{2}x^{n-2} + \cdots + \delta^{n-2} \cdot \binom{n}{n-1}x^1 + \delta^{n-1}$$

where δ is any infinitesimal.

The standard part of this is nx^{n-1} , since every subsequent term has a factor of δ . (We are implicitly using that adding infinitesimals does not change standard parts, and multiplying a standard real or an infinitesimal by an infinitesimal results in an infinitesimal.) \square

How about an example where the derivative fails to be defined? The canonical example is, of course, $g(x) = |x|$. This extends to ${}^*g : {}^*\mathbb{R} \rightarrow {}^*\mathbb{R}$; for readability, we will suppress the asterisk on ${}^*|\cdot|$, and henceforth we may use *f without the preamble stating the first-order formula which describes it. We shall take it as read that for any first-order definable function f , *f is defined by the transfer of the formula defining f .

Theorem 3.2. *If $g : \mathbb{R} \rightarrow \mathbb{R}$ by $x \mapsto |x|$, then g' is not defined at 0, but is defined elsewhere and is equal to the sign of x .*

Proof. At 0, the derivative would be

$$\text{st} \left(\frac{|\delta|}{\delta} \right) = {}^* \text{sgn}(\delta)$$

which is not constant as δ varies over the monad of 0, because there are both positive and negative infinitesimals. (If δ is a positive infinitesimal, then $-\delta$ is negative.)

Elsewhere, however, the derivative would be

$$\text{st} \left(\frac{|x + \delta| - |x|}{\delta} \right)$$

which splits into two cases:

$$\begin{cases} +1 = \frac{x + \delta - x}{\delta} & x > 0 \\ -1 = \frac{-(x + \delta) - (-x)}{\delta} & x < 0 \end{cases}$$

This is because $x < 0$ implies $|x + \delta| \leq |x| + |\delta| \simeq |x| < 0$, and similarly for $x > 0$. \square

3.1. The chain rule. We present the chain rule as a simple example of a theorem whose proof is made very neat and tidy by the use of infinitesimals.

Theorem 3.3 (Chain rule). *Let $f, g : \mathbb{R} \rightarrow \mathbb{R}$ be differentiable [in the non-standard sense]. Then $f \circ g$ is differentiable, and $(f \circ g)'(x) = f'(g(x))g'(x)$.*

Proof. Consider

$$X = \text{st} \left(\frac{f(g(x + \delta)) - f(g(x))}{\delta} \right)$$

We have

$$g(x + \delta) = (g'(x) + \varepsilon)\delta + g(x)$$

for some ε infinitesimal (depending on δ), by definition of $g'(x)$ as $\text{st} \left(\frac{g(x + \delta) - g(x)}{\delta} \right)$.

Substituting this into X , obtain

$$X = \text{st} \left(\frac{f[g(x) + \delta(g'(x) + \varepsilon)] - f(g(x))}{\delta} \right)$$

By differentiability of f , have

$$f'(g(x)) = \text{st} \left(\frac{f(g(x) + \gamma) - f(g(x))}{\gamma} \right)$$

for any infinitesimal γ ; so letting $\gamma = \delta(g'(x) + \varepsilon)$, obtain

$$f'(g(x)) = \text{st} \left(\frac{f[g(x) + \delta(g'(x) + \varepsilon)] - f(g(x))}{\delta(g'(x) + \varepsilon)} \right)$$

Factoring out $\text{st}(g'(x) + \varepsilon) = g'(x)$ from the denominator, this is simply $\frac{X}{g'(x)}$, so

$$X = g'(x)f'(g(x))$$

□

Notice how we did not require any bounding of error terms: the infinitesimals “bounded themselves” and simply vanished at the end without further hassle.

The rules of *linearity of the derivative* and the *product rule* follow very similarly.

Theorem 3.4. *Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous on $[a, b]$ and differentiable on (a, b) . If f is increasing, then $f'(x) \geq 0$ for all x .*

Proof. It is a first-order fact that for all $y > x$, have $f(y) \geq f(x)$. Therefore it remains true on moving to the hyperreals.

Then

$$\frac{f(x + \delta) - f(x)}{\delta} \geq 0$$

whenever δ is infinitesimal, so it remains so on taking standard parts. □

Remark. This is a partial converse to the theorem that having a positive derivative means the function is increasing; that fact, however, is a purely standard consequence of the Mean Value Theorem (which is itself a consequence of Rolle’s Theorem, Theorem 5.5), so we will not prove it here. Moreover, the standard proof of the “standard consequence” is elegant; this is an advertisement for using both standard and non-standard methods together.

4. CONTINUITY

Recall the definition of a closed interval in \mathbb{R} :

$$[a, b] = \{x \in \mathbb{R} : a \leq x \leq b\}$$

The transfer principle lets us carry this over to ${}^*\mathbb{R}$: to

$${}^*[a, b] = \{x \in {}^*\mathbb{R} : a \leq x \leq b\}$$

Remark. There are hyperreals *infinitesimally* less than a , which do not appear in ${}^*[a, b]$. Such hyperreals have standard part equal to a and yet are not in ${}^*[a, b]$. This demonstrates that $\text{st}(r) \in X$ does not necessarily imply $r \in {}^*X$.

Definition 4.1 (Continuity). *We say a function $f : [a, b] \rightarrow \mathbb{R}$ is continuous at $x \in [a, b]$ iff ${}^*f(x + \delta) \simeq {}^*f(x)$ for all δ infinitesimal with $x + \delta \in {}^*[a, b]$. That is, *f has a standard part which is constant on the monad of $x \in \mathbb{R}$.*

For illustration, we shall prove here that this definition is equivalent to the ε - δ definition of continuity, but in general we will use non-standard definitions freely throughout this essay. (We will not prove that the non-standard definition of “derivative” is equivalent to the standard one, for instance.)

Theorem 4.1. *The function $f : [a, b] \rightarrow \mathbb{R}$ is ε - δ continuous at x iff it is continuous at x in the non-standard sense.*

To make the proof easier to read, we will continue to suppress the asterisk in the extension $^*|\cdot|$ of the modulus function. This proof may be found as Corollary 7.1.2 in Goldblatt, although it is not very difficult to come up with by oneself.

Proof. Forward direction: let $x \in [a, b]$, and suppose for all ε there is δ such that for all y with $|y - x| < \delta$, have $|f(y) - f(x)| < \varepsilon$. Then letting $\varepsilon_n = \frac{1}{n}$, obtain δ_n such that for all y with $|y - x| < \delta_n$, $|f(y) - f(x)| < \frac{1}{n}$. Transfer this collection of facts (one for each $n \in \mathbb{N}$) to $^*\mathbb{R}$.

But every $y \in ^*\mathbb{R}$ with $y \simeq x$ satisfies $|y - x| < \delta_n$ for all n , so $|^*f(y) - ^*f(x)| < \frac{1}{n}$ for all n . That is, $^*f(y) \simeq ^*f(x)$ for all $y \simeq x$.

Conversely, suppose f is nonstandard-continuous at $x \in \mathbb{R}$. Let $\varepsilon \in \mathbb{R}^{>0}$. For any infinitesimal $\delta > 0$, have that for all $y \in ^*\mathbb{R}$ with $|y - x| < \delta$, $|^*f(y) - ^*f(x)| < \varepsilon$. Therefore (in particular) there is some $\delta \in ^*\mathbb{R}^{>0}$ such that for all $y \in ^*\mathbb{R}$ with $|y - x| < \delta$, $|^*f(y) - ^*f(x)| < \varepsilon$.

By the reverse direction of the transfer principle, the same statement must also hold in \mathbb{R} : there is $\delta \in \mathbb{R}^{>0}$ such that for all $y \in \mathbb{R}$ with $|y - x| < \delta$, have $|f(y) - f(x)| < \varepsilon$. □

This is one of the only places we will use the reverse direction of the transfer principle; Theorem 4.1 was included only to demonstrate the usual character of such proofs. Again, we will focus on the development of analysis results in non-standard analysis rather than on foundational rigour. The reverse of the transfer principle appears usually when showing that standard definitions are equivalent to non-standard ones.

Recall from Definition 3.1 that $f : [a, b] \rightarrow \mathbb{R}$ is *differentiable* at $x \in \mathbb{R}$ iff the quantity

$$\text{st} \left(\frac{f(x + \delta) - f(x)}{\delta} \right)$$

is invariant for infinitesimal δ .

Then it is easy to see that continuity is required for differentiability. Indeed, if $f(y)$ and $f(x)$ are not infinitesimally close even though $y = x + \delta \simeq x$, then $f(x + \delta) - f(x)$ is not infinitesimal, so $\frac{f(x + \delta) - f(x)}{\delta}$ must be infinite.

4.1. Uniform continuity. By definition, $f : [a, b] \rightarrow \mathbb{R}$ is continuous at $x \in \mathbb{R}$ iff $^*f : ^*[a, b] \rightarrow ^*\mathbb{R}$ has invariant standard part when we perturb standard-real $x \in [a, b]$ infinitesimally. If we form the stronger requirement that *f have invariant standard part when we perturb *any hyperreal* $x \in ^*[a, b]$ infinitesimally, then we obtain *uniform continuity*. This fact is highly surprising, but for reasons of space we will not prove it here.

Theorem 4.2. *Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then f is uniformly continuous.*

Proof. Let $x \in {}^*[a, b]$, and $y \simeq x$. Then say $z = \text{st}(x) \in \mathbb{R}$. Certainly $\text{st}(y) = \text{st}(x)$, so $f(x) \simeq f(z)$ by continuity of f at the real z ; and $f(z) \simeq f(y)$ similarly. Hence by transitivity of \simeq , have $f(x) \simeq f(y)$. \square

Example. Notice how this fails for $f : (0, 1] \rightarrow \mathbb{R}$ by $x \mapsto \frac{1}{x}$. Indeed, our choice of x could have been infinitesimal; then $z = \text{st}(x) = 0$ lies outside the domain of f . This formulation makes it obvious what fact we used about the non-standard formulation of a compact set: namely, that the set is closed under taking standard parts.

In a standard setting, the two usual proofs of Theorem 4.2 go via convergence of sequences and Bolzano-Weierstrass, or via the Lebesgue number lemma. This beautifully elegant proof has eliminated an enormous amount of the complexity of the same result in the standard setting, and its simplicity and lack of clutter reveals precisely what “compact” should mean in more general non-standard metric spaces.

4.2. Compactness. Robinson discovered an extremely neat formulation of the notion of compactness.

Definition 4.2 (Compactness). *A set $X \subseteq \mathbb{R}$ is compact iff every $x \in {}^*X$ has some $y \in X$ with $x \simeq y$.*

From this definition, it is easy that the continuous image of a compact set is compact.

Theorem 4.3. *Let $f : X \rightarrow \mathbb{R}$ be continuous and X compact. Then the image of f is compact.*

Proof. Let $y \in {}^*(f(X)) = {}^*f({}^*X)$. We wish to show that there is $x \in f(X)$ with $y \simeq x$.

Say $y = {}^*f(a)$, and let $x = f(\text{st}(a))$. Then ${}^*f(a) \simeq {}^*f(\text{st}(a))$ by continuity of f , so we are immediately done. \square

To obtain a more specific form of Theorem 4.3 without using the Heine-Borel theorem—that is, using closed boundedness directly rather than through compactness—it will be useful to develop the idea of the hypernatural numbers and hyperfinite partitions of a set.

5. HYPERNATURALS

For the moment, we will work within the ultrapower construction of the hyperreals. It is possible to proceed while remaining agnostic about the construction, but the details for this section are much simpler when working with an ultrapower. We will paint the construction in broad strokes here.

5.1. Internal sets.

Motivation. The motivating example for this section is the canonical model of Peano arithmetic embedded in \mathbb{R} : namely, \mathbb{N} . When we pass to ${}^*\mathbb{R}$, we might expect to obtain a non-standard model of Peano arithmetic, playing the same role in ${}^*\mathbb{R}$ as \mathbb{N} does in \mathbb{R} .

The transfer principle tells us, informally, that “from within the model, everything looks like \mathbb{R} ”. We would like a way to capture, for instance, the *Archimedean property*: that for every $r \in \mathbb{R}$, there is $n \in \mathbb{N}$ such that $r < n$. That is, we seek a set of “hypernaturals”, forming a non-standard model of Peano arithmetic embedded in ${}^*\mathbb{R}$, such that every hyperreal r has a hypernatural N bounding it above.

We will require the notion of an “internal” set: intuitively, “a set which exists in the model of ${}^*\mathbb{R}$ ”. For most of our purposes, it is enough to know that sets whose members are precisely those hyperreals satisfying some first-order property are internal (which is proved in Section 11.7 of Goldblatt [4], and is indeed a form of the transfer principle), but we shall be more explicit using the ultrapower construction.

Given a nonprincipal ultrafilter \mathcal{F} on \mathbb{N} , we construct ${}^*\mathbb{R}$ as the collection of sequences of reals, modulo the equivalence relation that $\langle r_n \rangle \sim \langle s_n \rangle$ if and only if $\{n \in \mathbb{N} : r_n = s_n\} \in \mathcal{F}$. The intuition is that the two sequences “agree almost everywhere”. (See Chapter 3 of Goldblatt [4]). We write $[r_n]$ for the equivalence class of the sequence $\langle r_n \rangle$.

Fix a sequence of sets $\langle A_n \rangle$ with each $A_n \subseteq \mathbb{R}$. The *internal set* $[A_n]$ is defined as follows:

$$[r_n] \in [A_n] \text{ if and only if } \{n \in \mathbb{N} : r_n \in A_n\} \in \mathcal{F}.$$

The internal sets are precisely those which may be obtained in this way; they are “the sets which exist in the model”, and they will turn out to be the sets over which we may quantify.

Likewise an *internal function* is obtained from a sequence $\langle f_n \rangle$ of real-valued functions $A_n \rightarrow \mathbb{R}$:

$$[f_n] : [A_n] \rightarrow {}^*\mathbb{R}, [r_n] \mapsto [f_n(r_n)]$$

Definition 5.1. We define the set of hypernaturals to be the set ${}^*\mathbb{N}$ which is the image of the elementary embedding of $\mathbb{N} \subseteq \mathbb{R}$ into ${}^*\mathbb{R}$: that is,

$$[r_n] \in {}^*\mathbb{N} \Leftrightarrow \{n \in \mathbb{N} : r_n \in \mathbb{N}\} \in \mathcal{F}$$

By construction, it is an internal subset of ${}^*\mathbb{R}$.

Remark. There is an extended form of the transfer principle (where second-order statements may be transferred), which is even more difficult to justify than our original statement¹ of Definition 1.1. It allows us to translate between certain second-order statements:

$$(\forall A \subseteq B) \leftrightarrow (\forall A \text{ internal } \subseteq {}^*B)$$

This procedure can be made to result in a formulation which is agnostic with respect to how we constructed the hyperreals, but in this essay we have only sketched a motivation, and have worked within the ultrapower construction. See Chapter 13 of Goldblatt [4] for the details.

Later on, when we treat measure theory, we will make heavy use of this “restricted second-order” transfer principle.

¹For which, recall, we appealed to the power of Los’s Theorem.

The definition of “internal” here was very much dependent on the implementation of ${}^*\mathbb{R}$ as an ultrapower. However, it may be viewed more generally; this is the start of “Internal Set Theory”. We will elide such details.

Definition 5.2 (Transfer Principle, extended). *Let ϕ be a first-order sentence in the language of the totally ordered field \mathbb{R} , which is additionally allowed to mention (finitely many) specific sets $A_n \subseteq \mathbb{R}$, and is allowed to contain quantifiers $(\forall A \subseteq \mathbb{R})$ and $(\exists A \subseteq \mathbb{R})$. Then ϕ is true in \mathbb{R} if and only if ${}^*\phi$ is true in ${}^*\mathbb{R}$, where ${}^*\phi$ is obtained by*

- (1) replacing all quantifiers $(\forall x \in \mathbb{R})$ with $(\forall x \in {}^*\mathbb{R})$ and $(\exists x \in \mathbb{R})$ with $(\exists x \in {}^*\mathbb{R})$;
- (2) replacing each A_n with *A_n ;
- (3) replacing all quantifiers $(\forall A \subseteq \mathbb{R})$ with $(\forall A \text{ internal } \subseteq {}^*\mathbb{R})$ and likewise $(\exists A \subseteq \mathbb{R})$ with $(\exists A \text{ internal } \subseteq {}^*\mathbb{R})$.

By transfer, hypernaturals have several obvious properties:

- They are closed under addition.
- There are no hypernaturals between m and $m + 1$, for m hypernatural.
- For every $x \in {}^*\mathbb{R}$, have $x \in {}^*[\lfloor x \rfloor, \lfloor x \rfloor + 1]$.

Every hypernatural is either standard natural, or bigger than all standard naturals; this can be seen from the (transferred) fact that subtracting 1 from a nonzero hypernatural yields another hypernatural, and the (transferred) fact that all hypernaturals are nonnegative.

Example. It is a second-order expressible fact in the reals that every nonempty finite set has a least element:

$$(\forall A \subseteq \mathbb{R})((\exists n \in \mathbb{N})(A \text{ has } n \text{ elements and } n > 0) \Rightarrow (\exists x \in A)(\forall y \in A)(x \leq y))$$

Here, we use a shorthand:

$$(A \text{ has } n \text{ elements}) \leftrightarrow (\exists \text{ a bijective function } \{x \in \mathbb{N} : x < n\} \rightarrow A)$$

The transfer principle from Definition 1.1 does not apply. But our amended second-order version does apply, to give us

$$(\forall A \text{ internal } \subseteq {}^*\mathbb{R})((\exists n \in {}^*\mathbb{N})(A \text{ has } n \text{ elements and } n > 0) \Rightarrow (\exists x \in A)(\forall y \in A)(x \leq y))$$

where

$$(A \text{ has } n \text{ elements}) \leftrightarrow (\exists \text{ a bijective internal function } \{x \in {}^*\mathbb{N} : x < n\} \rightarrow A)$$

That is, internal hyperfinite sets have least elements.

In fact, if we work in the ultrapower construction, the minimal element of the internal set $[A_n]$ is $[a_n]$, where for each $n \in \mathbb{N}$, a_n is the least element of A_n . (We will not prove this.)

Additionally, if m is a hypernatural and x_1, x_2, \dots, x_m is an internal set of hyperreals, then there is a maximum x_j .

5.2. Hypersequences. We may define *hypersequences*, by means of transferring a function $f : \mathbb{N} \rightarrow \mathbb{R}$ to a function ${}^*f : {}^*\mathbb{N} \rightarrow {}^*\mathbb{R}$. (We will assume that we can always extend f , whether or not the transfer principle applies to our particular choice of f . This requirement is justifiable; it is studied in a little more detail in Section 9.1.) The notion of a hypersequence leads to a very neat characterisation of the convergence of a sequence.

Definition 5.3. *If $f : \mathbb{N} \rightarrow \mathbb{R}$, have $(f(n))_{n \in \mathbb{N}} \rightarrow x \in \mathbb{R}$ if and only if ${}^*f(k) \simeq x$ for all k infinite. (Recall that *f maps from ${}^*\mathbb{N}$ to ${}^*\mathbb{R}$.)*

Of course, from *within* the model, the statement “for all infinite k ” is not meaningful, because the set of infinite naturals is not internal.

Theorem 5.1. *The set of infinite hypernaturals is not internal.*

Proof. Every non-empty internal set of hypernaturals has a smallest element. This is Exercise 1 from Section 12.4 of Goldblatt [4], and it follows by transferring the true statement of the well-ordering of the naturals:

$$(\forall A \subseteq \mathbb{N})(\exists n \in A)(\forall m \in A)(n \leq m)$$

to

$$(\forall A \text{ internal } \subseteq {}^*\mathbb{N})(\exists n \in A)(\forall m \in A)(n \leq m)$$

But there is no smallest infinite hypernatural, because subtracting 1 from any infinite hypernatural yields another infinite hypernatural. \square

Theorem 5.2. *The non-standard and the standard definitions of convergence coincide: if (and only if) $(f(n)) \rightarrow x$ in the non-standard sense of Definition 5.3, then for every $\varepsilon \in \mathbb{R}^{>0}$ there is $N \in \mathbb{N}$ such that for all $n > N$, $|f(n) - x| < \varepsilon$.*

Proof. If the sequence $(f(n))$ converges to x in the non-standard sense, then for every $\varepsilon \in \mathbb{R}^{>0}$ we have all infinite $N \in {}^*\mathbb{N}$ satisfying

$$\text{for all } n > N, |{}^*f(n) - x| < \varepsilon$$

Therefore it must be the case that some $N \in \mathbb{N}$ has this same property transferred to the reals:

$$\text{for all } n > N, |f(n) - x| < \varepsilon$$

since if not, this would be an internal property that distinguished between finite N and infinite N ; this contradicts Theorem 5.1.

Conversely, suppose the sequence $(f(n))$ converges to x in the standard sense. Fix some $\varepsilon \in \mathbb{R}^{>0}$. Then there is some particular N_ε - say, for the sake of argument, $N_\varepsilon = 10$ - such that all larger $n \in \mathbb{N}$ have $|f(n) - x| < \varepsilon$.

This property transfers to the hyperreals: for all $n \in {}^*\mathbb{N}$ with $n > N_\varepsilon$ (which, for the sake of argument, is 10), we have

$$|{}^*f(n) - x| < \varepsilon$$

In particular, all infinite hypernaturals n satisfy $|*f(n) - x| < \varepsilon$.

Finally, allowing ε to vary over all positive reals, we discover that for all infinite $n \in {}^*\mathbb{N}$, it is the case that $*f(n) \simeq x$. \square

5.3. Lattices. Hypernaturals yield a succinct proof of the Intermediate Value Theorem, by means of considering a “hyperfinite lattice” of points on the real line. We examine the behaviour on that lattice of a \mathbb{R} -continuous function when it is extended to ${}^*\mathbb{R}$, and gain access to the nice properties of hyperfinite sets. The following proof is from Pétry [2], where it appears as Theorem 25.

Theorem 5.3. *Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous with $f(a) < 0 < f(b)$. Then there is c such that $f(c) = 0$.*

Proof. Extend f to $*f$, and consider the “hyperpartition” of $[a, b]$

$$S = \left\{ \frac{b-a}{m}k + a : 0 \leq k \leq m \right\}$$

for m a fixed infinite hypernatural.

Construct

$$T = \{ *f(x) : x \in S, *f(x) > 0 \}$$

Then $\bar{k} := \min\{x \in S : *f(x) \in T\}$ has

$$\text{st } *f(\bar{k}) = 0$$

since certainly it is ≥ 0 by definition as (the standard part of) a member of a set T of positive numbers, while if it were strictly greater than 0 then we could subtract $\frac{b-a}{m}$ from \bar{k} to obtain another member κ of S such that $*f(\kappa) \in T$, contradicting minimality of \bar{k} . \bar{k} does exist, because it is the minimum of a hyperfinite set. \square

(This proof can be converted into another non-standard one which does not use hypernaturals, instead using the least upper bound property of the standard reals.)

We are now ready to revisit Theorem 4.3 in a more concrete form.

Theorem 5.4. *Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then f is bounded and attains its bounds.*

Proof. Extend f to $*f$.

Then for m a hyperfinite integer,

$$S = \left\{ \frac{b-a}{m}k + a : 0 \leq k \leq m \right\}$$

is a hyperfinite set and so $*f$ attains a maximum and a minimum on that set; say at m_+, m_- respectively. Since $[a, b]$ is bounded, we must have $*[a, b]$ bounded, and so m_+ and m_- are finite.

Now, if $*f(m_+)$ is infinite, then $\text{st}(*f(m_+)) = f(\text{st}(m_+))$ would be infinite too (by continuity), which is a contradiction unless $\text{st}(m_+)$ lies outside $[a, b]$. But this can never

happen because $[a, b]$ is closed: since $m_+ \geq a$, we must have $\text{st}(m_+) \geq \text{st}(a) = a$, and likewise since $m_+ \leq b$, we must have $\text{st}(m_+) \leq \text{st}(b) = b$.

Therefore ${}^*f(m_+)$ is finite; so *f is bounded above on S and attains its bound, at m_+ . Hence by continuity f is bounded above and attains its bound at $\text{st}(m_+)$, which we have already shown is in the domain $[a, b]$.

Exactly the same argument shows that f is bounded below and attains its bound. \square

Remark. The proof of Theorem 5.4 is so short and concise that it becomes extremely clear what its real content is: namely, that closed bounded intervals are compact (in the sense of Definition 4.2). Most of the proof is simply showing that m_+ and m_- must have standard parts in the domain of f , which is precisely what it means for $[a, b]$ to be compact.

We move on to the related Rolle's Theorem.

Theorem 5.5 (Rolle's Theorem). *Let $f : [a, b] \rightarrow \mathbb{R}$ be differentiable on (a, b) and continuous on $[a, b]$. Suppose that $f(a) = f(b)$. Then there is $c \in (a, b)$ such that $f'(c) = 0$.*

Proof. The following proof is from Section 8.5 of [4]. Since f is continuous on a closed bounded interval, it is bounded and attains its bounds (Theorem 5.4). If $f(x) = f(a) = f(b)$ for all x , then we are done: f is constant and so has zero derivative everywhere. Otherwise, without loss of generality, f attains a global maximum at $x \notin \{a, b\}$, say. (If in fact $f(a)$ is a global maximum, then consider $-f$ instead.)

Now, it is a first-order fact that $f(y) \leq f(x)$ for every $y \in [a, b]$; so by transfer, x maximises *f .

Therefore

$$\frac{f(x + \varepsilon) - f(x)}{\varepsilon} \leq 0 \leq \frac{f(x + \delta) - f(x)}{\delta}$$

for any $\varepsilon > 0, \delta < 0$ both infinitesimal.

Since the left-hand side and right-hand side are both infinitesimally close to each other, on taking standard parts we obtain that $f'(x)$ is both nonnegative and nonpositive, so it must be 0. \square

The Mean Value Theorem and then Taylor's theorem can be proved in an ε - δ free way from Rolle's theorem (for example, as in Theorem 4 of Chapter 11 in Spivak [5]).

5.4. Hyper-sums. Continuing the theme of the hypernaturals, we investigate *hyper-sums*.

Let $S(n) = \sum_{i=0}^n f(i)$, for $f : \mathbb{N} \rightarrow \mathbb{R}$. This function S extends to a function ${}^*S : {}^*\mathbb{N} \rightarrow {}^*\mathbb{R}$, since we may express S in a transferrable way as

$$S(0) = f(0); (\forall n \in \mathbb{N}^{>0})(S(n) = S(n-1) + f(n))$$

and then (using our second-order transfer principle) show that there is only one *S which satisfies the transferred condition.

The transfer principle yields properties such as

$$(1) \quad \left| \sum_{k=0}^m \lambda_k u_k \right| \leq (\max_k |u_k|) \sum_{k=0}^m |\lambda_k|$$

where for clarity we have implicitly suppressed the asterisk on the \sum symbol.

6. INTEGRATION

The definition of the Riemann integral is made especially comprehensible by the notion of the hyper-sum. Taking the standard method of approximating an integral by the area of rectangles, and then taking the number of rectangles to be larger and larger and eventually infinite (i.e. hyperfinite), is a highly intuitive idea.

The “obvious” choice of the definition of $\int_a^b f$ would be

$$\text{st} \left(\sum_{k=0}^{m-1} f(x_k)(x_{k+1} - x_k) \right)$$

where m is some hyperfinite integer, and

$$x_k = \frac{b-a}{m}k + a$$

However, a little thought suggests that some rather pathological functions would thereby be considered to have very wrong integrals, because we can only ever sample f at some fixed points; letting our function misbehave away from those points would yield counterintuitive results.

The correct refinement is as follows.

Definition 6.1. *The integral of $f : [a, b] \rightarrow \mathbb{R}$ is the following expression, if it is well-defined:*

$$\int_a^b f = \text{st} \left(\sum_{k=0}^{m-1} f(\phi(x_k, x_{k+1}))(x_{k+1} - x_k) \right)$$

where ϕ is any function $[a, b]^2 \rightarrow \mathbb{R}$ such that $r \leq \phi(r, s) \leq s$, and $(x_{k+1} - x_k) = \frac{1}{m}$, and m is an infinite hypernatural.

This captures the idea that our m sampling points are allowed to vary their position slightly in their respective intervals.

We say a function $f : [a, b] \rightarrow \mathbb{R}$ is *integrable* if $\int_a^b f$ is well-defined as hyperfinite m varies and for all choices of ϕ .

Showing that a function is integrable is approximately as difficult using the non-standard definition as it is using the standard definition (as a limit of a Riemann sum taken over smaller and smaller dissections). However, the non-standard definition is perhaps conceptually a little simpler, because it lacks a limiting process.

To illustrate the process of showing a function is integrable, we prove that continuous functions are integrable. The proof here, rendered down from Pétry’s ([2], Section 12.5), is very similar to a standard proof.

Theorem 6.1. *Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then f is integrable.*

Proof. We need to show that given two different discretisations $(x_k)_{k=0}^m$ and $(y_k)_{k=0}^n$, and two “nearness” functions ϕ and ψ , the following is true:

$$\sum_{k=0}^{m-1} {}^*f({}^*\phi(x_k, x_{k+1}))(x_{k+1} - x_k) \simeq \sum_{k=0}^{n-1} {}^*f({}^*\psi(y_k, y_{k+1}))(y_{k+1} - y_k)$$

Consider the more general discretisation given by taking the union of the x_i and y_i : label this list $(w_i)_{i=0}^l$. (Note that this is no longer “uniform”: the w_i do not necessarily have equal intervals between them, although the x_i and y_i did.) We will suppress the asterisk on ${}^*[\alpha, \beta]$ henceforth.

Each $[w_i, w_{i+1}]$ lies fully within some $[x_j, x_{j+1}]$, by construction of the w_i , so

$$[x_k, x_{k+1}] = [w_{i_k}, w_{i_k+1}] \cup \dots \cup [w_{r_k-1}, w_{r_k}]$$

for some i_k, r_k .

Therefore, suppressing the asterisk on $\phi(x_k, x_{k+1})$, we may un-telescope the sum:

$${}^*f(\phi(x_k, x_{k+1}))(x_{k+1} - x_k) = \sum_{j=i_k}^{r_k-1} {}^*f(\phi(x_k, x_{k+1}))(w_{j+1} - w_j)$$

so

$$\sum_{k=0}^{m-1} {}^*f(\phi(x_k, x_{k+1}))(x_{k+1} - x_k) = \sum_{k=0}^{m-1} \sum_{j=i_k}^{r_k-1} {}^*f(\phi(x_k, x_{k+1}))(w_{j+1} - w_j)$$

Now, since $x_k \simeq x_{k+1}$ and w_j, w_{j+1} are both in $[x_k, x_{k+1}]$, we have

$$\phi(x_k, x_{k+1}) \simeq w_j \simeq w_{j+1}$$

Hence in fact the right-hand side is an expression for a Riemann sum with (w_i) as a dissection, although recall that the dissection does not have equal intervals between successive points.

Relabelling the sum on the right-hand side, for some $(c_k)_{k=0}^l$ (which, if we were so inclined, we could express in terms of ϕ and the x_i), we have

$$\sum_{k=0}^{m-1} {}^*f(\phi(x_k, x_{k+1}))(x_{k+1} - x_k) = \sum_{k=0}^{l-1} {}^*f(c_k)(w_{k+1} - w_k)$$

Symmetrically,

$$\sum_{k=0}^{n-1} {}^*f(\psi(y_k, y_{k+1}))(y_{k+1} - y_k) = \sum_{k=0}^{l-1} {}^*f(d_k)(w_{k+1} - w_k)$$

for some $(d_k)_{k=0}^m$ with each $c_k \simeq d_k$. Notice that the upper limit of this sum is indeed the same $l - 1$ as before, because the dissection is taken over the same sequence (w_i) .

Taking the modulus of the difference of the two expressions, obtain

$$\left| \sum_{k=0}^{l-1} ({}^*f(c_k) - {}^*f(d_k))(w_{k+1} - w_k) \right|$$

But f is continuous on a compact set, so is uniformly continuous (see Section 4.1); so ${}^*f(c_k) - {}^*f(d_k)$ is infinitesimal for all k , because $c_k \simeq d_k$.

By equation 1 in Section 5.4, for every $\varepsilon \in \mathbb{R}^{>0}$ we have

$$\left| \sum_{k=0}^{l-1} ({}^*f(c_k) - {}^*f(d_k))(w_{k+1} - w_k) \right| \leq \varepsilon \sum_{k=0}^{l-1} |w_{k+1} - w_k|$$

Therefore the two expressions for the integral are indeed infinitesimally close, since

$$\sum_{k=0}^{l-1} |w_{k+1} - w_k| = \sum_{k=0}^{l-1} (w_{k+1} - w_k) = b - a$$

□

Before we introduce the link between integration and differentiation (the Fundamental Theorem of Calculus), we first require a lemma, which is Pétry's Theorem 32.

Theorem 6.2 (Integral mean value theorem). *Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous. Then there is a real $u \in [a, b]$ such that*

$$\int_a^b f(x)dx = (b - a)f(u)$$

Proof. Since f is continuous on a closed bounded interval, it attains its bounds; say $f(c) \leq f(x) \leq f(d)$ for all x .

Now, $(b - a)f(c) \leq S(\delta) \leq (b - a)f(d)$ for any Riemann sum $S(\delta)$ with box-width δ , so

$$f(c) \leq \frac{1}{b - a} \int_a^b f(x)dx \leq f(d)$$

Then we are done by the Intermediate Value Theorem: there is $u \in [c, d]$ such that

$$f(u) = \frac{1}{b - a} \int_a^b f(x)dx$$

□

Theorem 6.3 (Fundamental Theorem of Calculus, first part). *Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous, and $x_0 \in [a, b]$. Then the function*

$$g : x \mapsto \int_{x_0}^x f(t)dt$$

is differentiable on (a, b) with derivative $g'(x) = f(x)$. That is, an antiderivative of f is given by the integral.

Proof. Let δ be infinitesimal. Then

$$\frac{{}^*g(x + \delta) - {}^*g(x)}{\delta} = \frac{1}{\delta} \int_x^{x+\delta} f(t)dt$$

By transferring the integral mean value theorem (Theorem 6.2), there is $w \in {}^*[x, x+\delta]$ such that

$$\frac{1}{\delta} \int_x^{x+\delta} f(t) dt = {}^*f(w)$$

On taking standard parts and using continuity, we obtain

$$\text{st} \left(\frac{{}^*g(x+\delta) - {}^*g(x)}{\delta} \right) = \text{st}({}^*f(w)) = f(x)$$

□

The first part of the FTC told us how to find an antiderivative by integrating. There is a second part to the FTC, which will tell us how to integrate f in terms of a known antiderivative. To prove it, we shall require a standard theorem on antiderivatives.

Theorem 6.4. *Antiderivatives are unique up to the addition of a constant. That is, if $H_1, H_2 : [a, b] \rightarrow \mathbb{R}$ satisfy $H_i' = f$, then $H_1 = H_2 + k$ for some constant k .*

Proof. Consider $H_1 - H_2 : [a, b] \rightarrow \mathbb{R}$. This function has derivative 0 at all points. By the Mean Value Theorem, this means $H_1 - H_2$ is constant. (This follows from the remark after Theorem 3.4, by which we deduce that $H_1 - H_2$ is both nondecreasing and nonincreasing.) □

Theorem 6.5 (FTC, second part). *Let $f : [a, b] \rightarrow \mathbb{R}$ be continuous, and suppose $F : [a, b] \rightarrow \mathbb{R}$ is an antiderivative for f (so $F' = f$). Then*

$$\int_a^b f = F(b) - F(a)$$

Proof. Break the integral up at the point x_0 , and define

$$G(x) = \int_{x_0}^x f$$

By the first part of the FTC (Theorem 6.3), G is an antiderivative for f ; so since F and G are both antiderivatives, they differ by a constant (Theorem 6.4):

$$F(x) + c = G(x)$$

Therefore

$$\int_a^b f = \int_a^{x_0} f + \int_{x_0}^b f = -G(a) + G(b) = F(b) - F(a) + c - c$$

where the transformation

$$\int_a^b f = \int_a^{x_0} f + \int_{x_0}^b f$$

follows from splitting up the Riemann sum. (Because the integral is well-defined, we are free to choose a convenient Riemann sum where x_0 is a point of the discretisation, so that the sum splits up perfectly into two chunks.) □

Remark. One consequence of this theorem is that if f' is continuous between a and b , then its integral is $f(b) - f(a)$.

Definition 6.2 (Improper integral). We write $\int_a^\infty f(x)dx$ for the expression

$$\text{st} \left(\int_a^M f(x)dx \right)$$

where M is an arbitrary infinite positive hyperreal, if that expression is finite and well-defined as M varies.

Example. When does $\int_a^\infty x^\theta dx$ exist? Assuming $\theta \neq -1$, the integral is

$$\text{st} \left(\int_a^M x^\theta dx \right) = \text{st} \left(\left[\frac{x^{\theta+1}}{\theta+1} \right]_a^M \right) = \text{st} \left(\frac{M^{\theta+1}}{\theta+1} - \frac{a^{\theta+1}}{\theta+1} \right)$$

This is well-defined if and only if $\theta + 1 < 0$, in which case its value is $\frac{-a^{\theta+1}}{\theta+1}$.

If instead $\theta = -1$, the integral is $\text{st}(\log(M) - \log(a))$, which is infinite, so the integral is not defined in this case either.

7. SERIES

Intuitively, since series are simply infinite sums, it should be the case that hyperfinite sums express series neatly. This turns out to be the case. We are still assuming throughout that we may extend standard sequences to non-standard sequences; see further discussion in Section 9.1.

Definition 7.1. Let $c : \mathbb{N} \rightarrow \mathbb{R}$ by $k \mapsto c_k$. We say

$$\sum_{i=0}^{\infty} c_k$$

converges iff

$$\text{st} \left(\sum_{k=0}^m {}^*c_k \right)$$

is finite and is well-defined as m varies over all infinite hypernaturals in ${}^*\mathbb{N}$.

Two of the standard examples of infinite sums are

$$\sum_{k=1}^{\infty} \frac{1}{2^k}$$

and

$$\sum_{k=1}^{\infty} 1$$

The former, of course, converges. In this instance, it is a fact that

$$\sum_{k=1}^m \frac{1}{2^k} = 1 - 2^{-m}$$

and this fact is true in the limited second-order sense of Section 5.1, so it remains true when we transfer to infinite m ; in particular, then 2^{-m} is infinitesimal, so the standard part of our resulting sum is simply 1.

For the latter sum (that is, the sum of infinitely many of the constant 1), we have

$$\sum_{k=1}^m 1 = m$$

which has infinite standard part when m is infinite; so in this instance, as expected, the sum fails to converge.

To understand convergence, the following result is very useful; it basically states that a sequence is Cauchy if and only if it converges, in the specific case that the sequence is a sequence of partial sums of a series.

Theorem 7.1 (Cauchy's criterion). $\sum_{k=1}^{\infty} a_k$ converges if and only if

$$\sum_{k=m}^n {}^*a_k$$

is infinitesimal for all $m \leq n$ infinitely large.

Proof. This proof is from Pétry [2], section 19.2, where it appears as Theorem 53.

Suppose $\sum_{k=1}^{\infty} a_k$ converges to a . Then

$$\sum_{k=m}^n {}^*a_k = \sum_{k=1}^n {}^*a_k - \sum_{k=1}^{m-1} {}^*a_k \simeq a - a = 0$$

Conversely, suppose

$$\sum_{k=m}^n {}^*a_k$$

is infinitesimal for all $m \leq n$ infinite. We need $\sum_{k=1}^m {}^*a_k$ to have the same, finite, standard part as $\sum_{k=1}^n {}^*a_k$ for all $m < n$ infinite.

Clearly they have the same standard part if both sums are finite, because when we subtract them, we obtain

$$\sum_{k=m+1}^n {}^*a_k$$

which is infinitesimal by assumption. So it remains to show that they are indeed finite.

Let *n be some fixed infinite hypernatural (which we decorate with the asterisk as a cue for the fact that is infinite). For all infinite $m < {}^*n$, we have that

$$\left| \sum_{k=m}^{{}^*n} {}^*a_k \right| < 1$$

(In fact, we have much more: we have that the left-hand side is infinitesimal.)

Since the property

$$P(m) = \left[(m < {}^*n) \rightarrow \left(\left| \sum_{k=m}^{{}^*n} {}^*a_k \right| < 1 \right) \right]$$

is internal, it cannot suffice by itself as a means of distinguishing between finite and infinite integers m . (Indeed, no such means exists, by Theorem 5.1.) So there must be a finite natural p such that $P(p)$ holds:

$$\left| \sum_{k=p}^{{}^*n} {}^*a_k \right| < 1$$

Therefore $\sum_{k=1}^{{}^*n} {}^*a_k$ is finite, being the sum of two finite quantities

$$\sum_{k=1}^{p-1} {}^*a_k + \sum_{k=p}^{{}^*n} {}^*a_k$$

□

Example (The harmonic series).

$$\sum_{k=m+1}^{2m} \frac{1}{k} \geq m \times \frac{1}{2m} = \frac{1}{2}$$

which is not infinitesimal when m is infinite, so Cauchy's criterion fails for the harmonic series. (This is basically the usual proof by Cauchy condensation.)

Theorem 7.2. *Absolute convergence implies convergence.*

Proof. Because

$$\left| \sum_{k=m}^n a_k \right| \leq \sum_{k=m}^n |a_k|$$

(for n, m either infinite or standard integers), the result is immediately clear from Cauchy's criterion. □

The comparison test likewise follows by transferring the fact that

$$\sum_{k=m}^n a_k \leq \sum_{k=m}^n b_k$$

whenever $a_k \leq b_k$ for all $k \in [m, n]$.

Theorem 7.3 (Ratio test). *Let a_k be a sequence of reals.*

(1) *If there is some real L such that for all infinitely large m ,*

$$\left| \frac{a_{m+1}}{a_m} \right| \leq L < 1$$

then $\sum_{k=1}^{\infty} a_k$ converges absolutely.

(2) *If instead for all infinite m we have*

$$\left| \frac{a_{m+1}}{a_m} \right| \geq 1$$

then the sum diverges.

Proof. We prove only the first of these, since the second is almost identical in proof. Notice that we are omitting the asterisk of $^*a_{m+1}$, to make a more readable theorem statement.

Using the idea from Theorem 7.1 that “for all $m > m_0$, have $\left| \frac{a_{m+1}}{a_m} \right| \leq L$ ” cannot be a way to distinguish those m_0 which are infinite from those m_0 which are finite, there must be some m_0 finite with

$$\left| \frac{a_{m+1}}{a_m} \right| \leq L$$

for all $m > m_0$. Then we simply proceed by comparison with the convergent geometric series with common ratio L . \square

Notice that this easily implies the standard statement of the ratio test, because if

$$\frac{a_{m+1}}{a_m} \rightarrow L < 1$$

as $m \rightarrow \infty$, then for all infinitely large m , have

$$\frac{a_{m+1}}{a_m} \simeq L < \frac{1+L}{2} < 1$$

Remark. The non-standard formulation of the ratio test is a little more clumsy in its proof than the standard version, because it essentially states that “it is enough to pass to the standard version” and then proves the standard version by comparison. However, the non-standard version has the aesthetic benefit of being free of any explicit limits.

Theorem 7.4 (Alternating series test). *Let (a_n) be a sequence of positive reals which are decreasing to 0. Then*

$$\sum_{k=1}^{\infty} (-1)^k a_k$$

converges.

Proof. This proof is derived from Theorem 59 of Pétry [2].

Let m, n be infinite hypernaturals. The sum

$$S = a_m - a_{m+1} + a_{m+2} - \cdots + (-1)^{n-m} a_n$$

is precisely one of the two following:

$$a_m - (a_{m+1} - a_{m+2}) - \cdots - (a_{n-2} - a_{n-1}) - a_n$$

or

$$a_m - (a_{m+1} - a_{m+2}) - \cdots - (a_{n-3} - a_{n-2}) - (a_{n-1} - a_n)$$

depending on the parity of n . Because the a_i are decreasing, the result must be less than a_m in either case, since each bracketed term is nonnegative. If we omit the first term, we obtain a quantity

$$-S' := S - a_m = -(a_{m+1} - a_{m+2} + \cdots + (-1)^{n-m+1}a_n)$$

where S' is less than a_{m+1} , so $-S'$ is greater than $-a_{m+1}$. Hence in fact $S \geq a_m - a_{m+1} \geq 0$, so S is nonnegative.

Therefore

$$S = \left| \sum_{k=m}^n (-1)^k a_k \right| \leq a_m$$

Since (a_n) converges to 0, for infinite M we have $a_M \simeq 0$, so $\sum_{k=M}^N (-1)^k a_k$ is infinitesimal (being bounded in modulus by a_M). We are therefore done by Cauchy's criterion (Theorem 7.1). \square

We omit the integral test for convergence, because it is a routine application of similar ideas to the above; it may be found in Pétry, section 19.3 [2].

8. MORE GENERAL TOPOLOGICAL IDEAS

In this section, we will discuss how the more abstract ideas of topology can be viewed in the non-standard setting of ${}^*\mathbb{R}^n$, the product of n copies of ${}^*\mathbb{R}$. Although we will not prove that the transfer principle holds between \mathbb{R}^n and ${}^*\mathbb{R}^n$, it is intuitive that $({}^*\mathbb{R})^n$ has some of the properties we would like from ${}^*(\mathbb{R}^n)$: given any infinitesimal $\varepsilon \in {}^*\mathbb{R}$, we can obtain infinitesimals $\hat{\varepsilon} \in ({}^*\mathbb{R})^n$ in any direction by simply multiplying by any unit vector, while given any infinitesimal $\hat{\varepsilon} \in ({}^*\mathbb{R})^n$ we can take its length to get an infinitesimal in ${}^*\mathbb{R}$. (Length, of course, is defined by transferring the usual Euclidean distance.) In fact, it is the case that the monad of a point (x, y) in a general product space is equal to the product of the monads; see Section III.1 of Hurd and Loeb [6].

To be clear, then, two elements of ${}^*\mathbb{R}^n$ are infinitesimally close iff their norm is infinitesimal; equivalently, if and only if their coordinates are pointwise infinitesimally close. We define standard parts of vectors pointwise.

Definition 8.1. *A subset $X \subseteq \mathbb{R}^n$ is open if and only if, for every $x \in X$, the monad of x lies entirely in *X .*

Remark (Equivalence to the standard definitions). This non-standard definition is easily implied by the standard definition when stated as being a union of open balls. It implies the standard definition when stated as “every point has a neighbourhood within the set”: if x has monad entirely within *X , then there is $\varepsilon > 0$ (for instance, any infinitesimal ε)

such that all $y \in {}^*X$, with $|y - x| < \varepsilon$, lie within *X . This property transfers to \mathbb{R}^n by the reverse direction of the transfer principle.

Definition 8.2. *A subset $X \subseteq \mathbb{R}^n$ is closed if (and only if) its complement is open. That is, X is closed if and only if, whenever $y \in \mathbb{R}^n \setminus X$, the monad of y is entirely outside *X .*

Theorem 8.1. *Let $X \subseteq \mathbb{R}^n$. Then X is closed if and only if every convergent sequence in X has its limit point within X .*

Proof. Let X be closed, and let $(x_i)_{i=1}^{\infty}$ be a sequence in X , tending to x .

If $x \notin X$, then $x \in \mathbb{R}^n \setminus X$, so the monad of x lies entirely outside *X . But (x_i) converges to x , so all infinite hypernaturals m have $x_m \simeq x$, and this is a contradiction because all $x_n \in X \subseteq {}^*X$ for n finite. Since no internal property can distinguish finite hypernaturals from infinite hypernaturals, it must be the case that some infinite hypernatural M has $x_M \in {}^*X$.

Conversely, let X be not closed. Then $\mathbb{R}^n \setminus X$ is not open, so there is $y \in \mathbb{R}^n \setminus X$ such that some $r \in {}^*\mathbb{R}^n$ has $\text{st}(r) = y$ but $r \notin {}^*(\mathbb{R}^n \setminus X)$.

We claim that y is a limit point of X . Indeed, fix some specific $\varepsilon \in \mathbb{R}^{\geq 0}$. Then the following statement is true:

$$(\exists r \in {}^*\mathbb{R}^n)(r \in {}^*X \wedge |y - r| < \varepsilon)$$

so by the second-order version of transfer (Definition 5.2),

$$(\exists r \in \mathbb{R}^n)(r \in X \wedge |y - r| < \varepsilon)$$

Denote such an r by x_ε .

Finally releasing ε , $(x_{1/n})$ converges to x in the ε - δ sense, and therefore in the non-standard sense (by Theorem 5.2). \square

Remark. The proof of Theorem 8.1 is somewhat clumsy: it requires explicit uses of real $\varepsilon \in \mathbb{R}$. From a certain point of view, the proof contains two uses of the transfer principle: once to generate the sequence $(x_{1/n})$ which converges in the ε - δ sense, and once in the statement that ε - δ convergence is equivalent to convergence in the sense of Definition 5.3. (This latter usage is hidden away in the proof of Theorem 5.2.)

Theorem 8.2 (Robinson's theorem). *Let $X \subseteq \mathbb{R}^n$. Then the following are equivalent:*

- (1) *X is compact in the non-standard sense that every point $x \in {}^*X$ has a standard $r_x \in X$ to which it is infinitesimally close.*
- (2) *X is compact in the standard sense that every open cover has a finite sub-cover.*

Proof. (2) \Rightarrow (1): This direction of the proof is from Theorem 4.1.13 of Robinson [1].

Suppose $x \in {}^*X$ has the property that no $r \in X$ has $r \simeq x$. The idea is to transfer the fact that “there is a finite collection of points which together are near to every point”, for a contradiction.

For each $r \in X$ we can find a ball B_r around r , of positive standard-real radius ε_r , such that *B_r does not contain x (indeed, if not, then r would be infinitesimally near to x).

This collection of balls B_r forms an open cover of X , so it has a finite subcover; but each ball B_r can be specified in a first-order way as $\{x \in X : d(x, r) < \varepsilon_r\}$, so we have the internal statement that

$$X = B_{r_1} \cup B_{r_2} \cup \cdots \cup B_{r_n}$$

By transfer, this must be true of *X too:

$${}^*X = {}^*B_{r_1} \cup \cdots \cup {}^*B_{r_n}$$

which is a contradiction because we built the B_r such that no *B_r contained x .

(1) \Rightarrow (2): This direction of the proof is taken from Goldblatt [4], where it is given in Section 10.3. The proof is beautiful, but has the weakness that it does not extend to arbitrary topological spaces, because it relies on the presence of the rationals. (The result is true in general.)

Suppose $X \subseteq \mathbb{R}^n$ is not compact in the standard sense. Since the product of compact spaces is compact, there must be some coordinate i such that the i th projection $\pi_i(X)$ is not compact in the standard sense, so it is enough to work with $\pi_i(X) = X' \subseteq \mathbb{R}$. If we can deduce that there is some $x' \in X'$ whose standard part is not in ${}^*X'$, then we can lift it up to any point $x \in X$ with i th coordinate equal to x' ; the standard part of x then does not lie in *X .

To reiterate, then, we are working with $X' \subseteq \mathbb{R}$ which is not compact in the standard sense. Let $(U_i)_{i=1}^\infty$ be an open cover of X' without any finite subcover. We tweak each U_i into an open interval whose endpoints are rational.

Every $r \in X'$ lies within some U_{i_r} , say. Since U_{i_r} is open, it contains an interval $C_r := (p_r, q_r)$ with rational endpoints, such that the interval contains r . This creates an open cover

$$\mathcal{C} := \langle C_r : r \in X' \rangle$$

such that every C_r is entirely contained within some U_{i_r} .

But there are only countably many such intervals, so we may in fact enumerate the open cover \mathcal{C} : say as

$$\mathcal{C} = \langle (p_n, q_n) : n \in \mathbb{N} \rangle$$

Certainly \mathcal{C} covers X' , because every $r \in X'$ lies in its C_r , and C_r is included in the enumeration. It has no finite subcover, because each C_r is contained entirely within U_{i_r} , so that would imply a finite subcover from the U_i .

But now it is true that

$$(\forall k \in \mathbb{N})(\exists x \in X')(\forall n \in \mathbb{N})[n \leq k \Rightarrow x \notin (p_n, q_n)]$$

which precisely states that for all k ,

$$X' \not\subseteq (p_1, q_1) \cup \cdots \cup (p_k, q_k)$$

This is a statement which transfers to the hyperreals. Fix some infinite hypernatural K (it does not matter which), and let $x \in {}^*(X')$ be such that for all $n \in {}^*\mathbb{N}$ with $n \leq K$, we have $x \notin {}^*(p_n, q_n)$. Then for all finite n , we have $x \notin {}^*(p_n, q_n)$.

Finally, we claim that x is our hyperreal in ${}^*(X')$ which has its standard part not contained in X' , thereby witnessing that X' is not compact in the non-standard sense. Indeed, any $r \in X'$ is contained within some (p_n, q_n) , so if $x \simeq r$ then $p_n < x < q_n$, which would contradict the previous paragraph. \square

Theorem 8.3 (Heine-Borel). *The compact sets in \mathbb{R}^n are precisely the closed bounded sets.*

Proof. Let $X \subseteq \mathbb{R}^n$ be closed and bounded. Then take a point $x \in {}^*X$. x is finite, because the statement that X is bounded is a restricted second-order statement in the sense of Definition 5.2, so it remains true of *X by transfer. Therefore x has a standard part $\text{st}(x) \simeq x$, which we claim lies in X .

Indeed, if $\text{st}(x)$ were not in X , then the monad of $\text{st}(x)$ would lie outside *X (since X is closed), and in particular x would not be in *X .

Conversely, let $X \subseteq \mathbb{R}^n$ be compact. Then X is bounded: if X were unbounded, then *X would contain an infinite $x \in {}^*\mathbb{R}^n$ (since if not, we could distinguish infinite hyperreals from finite ones by comparing them with members of *X), and so no member of $X \subseteq \mathbb{R}^n$ could be infinitesimally close to x .

X is closed: let $y \in \mathbb{R}^n \setminus X$. We wish to show that the monad of y is entirely outside *X . If the monad of y contained an element x of *X , then by compactness, there would be a member of \mathbb{R}^n which lay in X , infinitesimally close to x . The only possible such standard member of \mathbb{R}^n is y , so in fact y must lie in X after all. \square

9. MEASURE THEORY

These final sections will address some meatier ideas, with the two goals of formulating Lebesgue measure and Brownian motion in the language of infinitesimals. For Lebesgue measure, we primarily use Goldblatt [4], Chapter 16. For the applications to Brownian motion, we use Hurd and Loeb [6], section IV.6.

Recall the following definitions of several objects fundamental to the study of (standard) measure theory:

Definition 9.1 (σ -algebra). *A collection \mathcal{A} of subsets of set S is a σ -algebra if it contains the empty set and is closed under countable unions, complements, and symmetric differences.*

If \mathcal{A} is only closed under finite unions (and complements and symmetric differences), it is called a ring of sets.

Definition 9.2 (Measure). *Let \mathcal{A} be a σ -algebra. A function $\mu : \mathcal{A} \rightarrow \mathbb{R}^{\geq 0} \cup \{\infty\}$ is a measure if $\mu(\emptyset) = 0$ and it is countably additive: whenever $(A_n)_{n=1}^{\infty}$ is a sequence of pairwise-disjoint members of \mathcal{A} , we have*

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n)$$

If \mathcal{A} is instead merely a ring of sets, μ is a measure if instead it is countably additive over all sequences of pairwise-disjoint members of \mathcal{A} whose union is in \mathcal{A} .

It is a classical fact that any measure on a ring of sets \mathcal{A} may be extended to a measure on a σ -algebra $\sigma(\mathcal{A})$, the intersection of all σ -algebras containing \mathcal{A} ; this result is known as the Carathéodory extension theorem. The theorem takes a measure μ on \mathcal{A} and outputs a measure called the *outer measure* μ^+ on $\sigma(\mathcal{A})$. The exact construction of the outer measure does not concern us, but it coincides with μ on members of \mathcal{A} . The outer measure need not be even a finitely additive set function on the power set $\mathcal{P}(S)$, though it is countably additive on $\sigma(\mathcal{A}) \subseteq \mathcal{P}(S)$; those subsets of S on which μ^+ is guaranteed to behave additively are known as the *measurable sets*, which we now define.

Definition 9.3 (Measurable set; see Halmos [7], §11). *Given a measure μ on ring of sets $\mathcal{A} \subseteq \mathcal{P}(S)$, we say that a set $B \subseteq S$ is μ^+ -measurable if, for every $E \in \mathcal{A}$, we have*

$$\mu^+(E) = \mu^+(E \cap B) + \mu^+(E \setminus B)$$

where μ^+ is the outer measure on $\sigma(\mathcal{A})$. That is, B “splits every $E \in \mathcal{A}$ in a way that is additive with respect to μ^+ ”.

It is a fact that every member of \mathcal{A} is μ^+ -measurable.

Recall the definition of the Lebesgue measure on \mathbb{R} :

Definition 9.4 (Lebesgue measure). *The Lebesgue measure on \mathbb{R} is the measure λ , on the σ -algebra generated by the open sets of the Euclidean topology (that is, the σ -algebra whose members are the Borel sets), such that*

$$\lambda([a, b]) = b - a$$

for all reals $a \leq b$.

Remark. This does indeed specify a measure, by the Carathéodory extension theorem. It is in fact the *unique* measure such that $\lambda([a, b]) = b - a$; this is Theorem A of §13 in [7].

We will consider measures constructed by applying Carathéodory’s extension theorem to measures μ_L of the following form, as in Goldblatt [4], Section 16.5. Let \mathcal{A} be an internal ring of subsets of $S \subseteq {}^*\mathbb{R}$. Let

$$\mu : \mathcal{A} \rightarrow {}^*\mathbb{R}^{\geq 0} \cup \{\infty\}$$

be any finitely-additive function. Define μ_L (for “ μ -Loeb”) by $\mu_L(A) = \text{st}(\mu(A))$ if $\mu(A)$ is finite, and ∞ if $\mu(A)$ is infinite (as a hyperreal) or is the literal value ∞ .

Then μ_L extends to a measure μ_L^+ on $\sigma(\mathcal{A})$. We say a set $B \subseteq {}^*\mathbb{R}$ is *Loeb measurable* if it is measurable with respect to μ_L^+ , in the sense of Definition 9.3.

The Loeb measure construction essentially lets us specify a hyperreal size for each member of an internal collection of subsets of ${}^*\mathbb{R}$, and pull that back into a true real-valued ordinary measure on that internal collection.

Example. Take $S = {}^*\mathbb{R}^{\geq 0}$. Let \mathcal{A} be the set of singletons from ${}^*\mathbb{R}^{\geq 0}$.

Then $\sigma(\mathcal{A})$ is the collection of countable sets and cocountable sets (that is, those whose complements are countable), since this collection is closed under taking countable unions, complements, and symmetric differences.

Define $\mu : \{^*a\} \mapsto 1$, the counting measure. Then μ_L^+ is the measure which takes a set $A \subseteq {}^*\mathbb{R}$, and returns the following:

- If A has cardinality n , where n is finite, then μ_L^+ returns n ;
- Otherwise μ_L^+ returns ∞ (that is, in the case that A is infinite or hyperfinite).

So, for example, $\mu_L^+(\{0, 1, \dots, N\}) = \infty$, where N is an infinite hypernatural. Notice that μ_L^+ need not be internal: the above μ_L^+ is capable of distinguishing between finite and infinite hypernaturals.

9.1. Comprehensiveness. We shall work in a *sequentially comprehensive* system, as defined in Section 15.4 of Goldblatt [4]. That is, one in which any function $f : \mathbb{N} \rightarrow B$ from \mathbb{N} to an internal set $B \subseteq {}^*\mathbb{R}$ extends to a function ${}^*f : {}^*\mathbb{N} \rightarrow B$. Alternatively stated, any sequence $(s_n)_{n=1}^{\infty}$ of elements of internal set $B \subseteq {}^*\mathbb{R}$ will extend to an internal hypersequence $({}^*s_n)_{n \in {}^*\mathbb{N}}$. (We implicitly used this property earlier when dealing with hypersequences; it comes free of charge through the transfer principle when f admits a first-order description, for instance.)

The ultrapower construction always creates a sequentially comprehensive system, for what Goldblatt calls “intricate” reasons which we will not cover here.

The upshot will be as follows. Given an internal sequence of sets $(A_n)_{n \in \mathbb{N}}$, and a list of properties (P_n) such that A_m satisfies property P_n for all $m \geq n$, we can extend the sequence (A_n) to a hypersequence. For N an infinite hypernatural, A_N will then have property P_i for all finite naturals i . That is, we will have constructed an object which has all these properties simultaneously.

This construction is akin in spirit to taking an intersection of nested sets with some property, to obtain an object which has all the properties.

9.2. Alternative characterisation of Loeb measurability. Recall the classical fact that a subset of $[0, 1]$ is Lebesgue measurable if and only if it can be approximated arbitrarily well by finite unions of intervals, and its complement can also be approximated arbitrarily well by finite unions of intervals. (See the proof of Theorem 4.3a in [8], for instance.)

With this in mind, we prove the following characterisation, from Section 16.6 of Goldblatt [4]. The goal here is to show that Lebesgue measure is a (rather natural) example of a Loeb measure.

Definition 9.5 (μ -approximability). *Let \mathcal{A} be an internal ring of subsets of $S \subseteq {}^*\mathbb{R}$, and $\mu : \mathcal{A} \rightarrow {}^*\mathbb{R}^{\geq 0} \cup \{\infty\}$ a finitely-additive function.*

We say that $B \subseteq S$ is μ -approximable if, for every $\varepsilon \in \mathbb{R}^{>0}$, there are sets $C_\varepsilon, D_\varepsilon \in \mathcal{A}$ with $\mu_L(D_\varepsilon \setminus C_\varepsilon) < \varepsilon$ and $C_\varepsilon \subseteq B \subseteq D_\varepsilon$.

Lemma 9.1. *We can approximate any μ -approximable set $B \subseteq {}^*\mathbb{R}$ by a member of \mathcal{A} , in the following sense: there is some $A \in \mathcal{A}$ such that the symmetric difference $A \Delta B$ has μ_L^+ -measure zero.*

Proof. Take a sequence of nested $\frac{1}{n}$ -approximations $C_n \subseteq B \subseteq D_n$, and extend each $\langle C_n : n \in \mathbb{N} \rangle, \langle D_n : n \in \mathbb{N} \rangle$ to a hypersequence of $\frac{1}{N}$ -approximations for N ranging over the hypernaturals. (This is justified by the remark on sequential comprehensiveness.)

Let the hypersequences be $\langle C_n : n \in {}^*\mathbb{N} \rangle$ and $\langle D_n : n \in {}^*\mathbb{N} \rangle$ where $C_n \subseteq B \subseteq D_n$ for all $n \in {}^*\mathbb{N}$. Then if we fix any (necessarily finite) $k \in \mathbb{N}$, it is an internal true statement that for every $n \in {}^*\mathbb{N}$, if $n \leq k$ then $C_n \subseteq D_k \subseteq D_n$ (by nestedness).

Since all finite naturals k satisfy that property, it must be the case that some infinite hypernatural K does too (because otherwise this would be a property that distinguished between finite and infinite naturals, contradicting Theorem 5.1): there is K such that for all $n \in {}^*\mathbb{N}$,

$$n \leq K \Rightarrow C_n \subseteq D_K \subseteq D_n$$

Then D_K is our desired member of \mathcal{A} such that $\mu_L^+(D_K \Delta B) = 0$.

Indeed,

$$D_K \Delta B = (D_K \setminus B) \cup (B \setminus D_K) \subseteq D_n \setminus C_n$$

for all $n \in \mathbb{N}$. The right-hand side has μ_L^+ -measure less than $\frac{1}{n}$; so we obtain that $\mu_L^+(D_K \Delta B)$ is a standard real which is less than all $\frac{1}{n}$.

That is, $\mu_L^+(D_K \Delta B) = 0$. □

In fact, there is a very concrete definition of the Loeb measure μ_L^+ of a set $B \subseteq S \subseteq {}^*\mathbb{R}$ which does not appear in the ring of sets $\mathcal{A} \subseteq \mathcal{P}(S)$. This concreteness comes at the cost of working with arbitrary $\varepsilon > 0$, but it will turn out to be the key ingredient allowing us to move freely between Loeb measure and Lebesgue measure. Morally speaking, Lebesgue measure is a statement about “those sets we may approximate by nice sets”, and approximation is easiest with an ε -related treatment; this is the motivation for the following lemma about Loeb measure.

Lemma 9.2. *Suppose $B \subseteq {}^*\mathbb{R}$ is Loeb measurable with finite Loeb measure. Then*

$$\mu_L^+(B) = \inf\{\mu_L(A) : A \in \mathcal{A}, B \subseteq A\} = \sup\{\mu_L(A) : A \in \mathcal{A}, A \subseteq B\}$$

Proof. We will assume the first equality; its proof requires manipulating the specific construction of the measure μ_L^+ according to the Carathéodory extension theorem. This

is not particularly difficult, but it requires some “grubby details”. The proof may be found² as Lemma 16.5.1 of Goldblatt [4].

To prove the second equality, then, we will show that

$$\mu_L^+(B) = \sup\{\mu_L(A) : A \in \mathcal{A}, A \subseteq B\}$$

Let $\varepsilon \in \mathbb{R}^{>0}$. We need to find $A_\varepsilon \in \mathcal{A}$ with $A_\varepsilon \subseteq B$ and

$$\mu_L^+(B) < \mu_L(A_\varepsilon) + \varepsilon$$

Now, we stipulated that B had finite Loeb measure, so (by the first part of the lemma)

$$\mu_L^+(B) = \inf\{\mu_L(A) : A \in \mathcal{A}, B \subseteq A\} < \infty$$

That is, there is some $A \in \mathcal{A}$ with $B \subseteq A$ and $\mu_L(A) < \infty$.

We can therefore use the first part of the lemma again, applied to $A \setminus B$. (Drawing Venn diagrams will elucidate this section.)

Since $\mu_L(A)$ is finite, we have

$$\mu_L^+(A \setminus B) = \mu_L(A) - \mu_L^+(B)$$

which is again finite, so by the first part of the lemma, we can approximate it: let $C \in \mathcal{A}$ be such that $A \setminus B \subseteq C$ and

$$\mu_L(C) < \mu_L^+(A \setminus B) + \varepsilon$$

Now, $A \setminus C$ is a set-difference of members of \mathcal{A} , so it lies in \mathcal{A} ; it is also a subset of B , since $A \setminus B \subseteq C$ so $A \setminus C \subseteq A \setminus (A \setminus B) \subseteq B$.

Then

$$C \supseteq (A \setminus B) \sqcup (B \setminus [A \setminus C])$$

(where \sqcup is a disjoint union), and each of those terms is Loeb measurable, so

$$\mu_L(C) \geq \mu_L^+(A \setminus B) + \mu_L^+(B \setminus [A \setminus C])$$

We chose C such that $\mu_L(C) < \mu_L^+(A \setminus B) + \varepsilon$. Therefore

$$\mu_L^+(B \setminus [A \setminus C]) < \varepsilon$$

so

$$\mu_L^+(B) < \mu_L^+(A \setminus C) + \varepsilon$$

Finally, we have already noted that $A \setminus C$ lies in \mathcal{A} , so the proof is complete: set $A_\varepsilon = A \setminus C$. \square

Theorem 9.3 (Alternative definition of Loeb measurability). *Let $B \subseteq {}^*\mathbb{R}$.*

- (1) *Suppose B is μ -approximable. Then B is Loeb measurable.*
- (2) *Suppose B is Loeb measurable with $\mu_L(B) \neq \infty$. Then B is μ -approximable.*

²Be warned that Goldblatt abuses the notation $\mu_L(B)$ to mean $\mu_L^+(B)$ when B is Loeb measurable.

Proof. (1): We sketch this direction. Suppose B is μ -approximable. Recall that μ_L^+ is the measure obtained by extending μ_L to a measure on $\sigma(\mathcal{A})$.

Then we can find a set $A \in \mathcal{A}$ such that $\mu_L^+(A \Delta B) = 0$ (for Δ the symmetric difference), as guaranteed by Lemma 9.1.

We now need to show that B itself is Loeb measurable: that is, it is measurable in the sense of Definition 9.3 with respect to μ_L^+ .

That is, for any $E \in \mathcal{A}$, we need B to split E additively with respect to μ_L^+ : we need

$$\mu_L^+(E) = \mu_L^+(E \cap B) + \mu_L^+(E \setminus B)$$

Recall from Definition 9.3 that every member of \mathcal{A} is automatically μ_L^+ -measurable; so the equality holds if we replace B with A throughout. But we have defined A to be “almost equal” to B in the sense of Lemma 9.1, and so by a simple argument (elucidated by a Venn diagram), the equality holds with B as well. (For full details, see Lemma 16.6.3 of [4]; what remains of the argument is mere unenlightening algebra. The essence is that “ A is extremely close to B from the point of view of μ_L^+ ”.)

(2): Let $\varepsilon \in \mathbb{R}^{>0}$. We need to show that we can find approximating sets $C_\varepsilon, D_\varepsilon \in \mathcal{A}$ with

$$\mu_L(D_\varepsilon \setminus C_\varepsilon) < \varepsilon \text{ and } C_\varepsilon \subseteq B \subseteq D_\varepsilon$$

But this is the content of Lemma 9.2:

$$\mu_L^+(B) = \inf\{\mu_L(A) : A \in \mathcal{A}, B \subseteq A\} = \sup\{\mu_L(A) : A \in \mathcal{A}, A \subseteq B\}$$

so we can find $C_\varepsilon \in \mathcal{A}$ with $C_\varepsilon \subseteq B$ such that

$$\mu_L(C_\varepsilon) \geq \mu_L^+(B) - \frac{\varepsilon}{2}$$

and we can similarly find $D_\varepsilon \supseteq B$ such that

$$\mu_L(D_\varepsilon) \leq \mu_L^+(B) + \frac{\varepsilon}{2}$$

from which the result follows immediately. \square

9.3. Lebesgue measure via Loeb measure. It turns out that Lebesgue measure can be defined in a very natural way as a Loeb measure, by “assigning a weight to an infinitesimal-width lattice on \mathbb{R} ”, as in Section 16.8 of Goldblatt [4].

Definition 9.6 (Loeb measure defining the Lebesgue measure). *Fix an infinite hypernatural N , and define a lattice*

$$S = \left\{ \frac{k}{N} : k \in {}^*\mathbb{Z}, -N^2 \leq k \leq N^2 \right\}$$

Define $\mathcal{P}_I(S)$ to be the collection of internal subsets of S , so each of its members is hyperfinite (indeed, of hyperfinite size less than or equal to $2N^2 + 1$). It is an internal algebra, and $\mu : \mathcal{P}_I(S) \rightarrow {}^\mathbb{R}^{\geq 0}$ given by*

$$\mu(A) = \frac{|A|}{N}$$

defines a finitely additive function suitable for creating a Loeb measure.

Specifically,

$$\mu_L(A) = \text{st} \left(\frac{|A|}{N} \right)$$

if $\frac{|A|}{N}$ is finite; and $\mu_L(A)$ takes the literal value ∞ otherwise. μ_L^+ is the measure on $\sigma(\mathcal{P}_I(S))$ whose existence is guaranteed by the Carathéodory extension theorem.

Remark. Notice that μ_L^+ is a measure on $\sigma(\mathcal{P}_I(S))$, not on $\mathcal{P}_I(S)$. Since $\sigma(\mathcal{P}_I(S))$ may contain some non-internal subsets of S , this means that while μ_L is restricted only to internal sets, μ_L^+ may be able to measure some external sets as well. An example is the finite hyperreals, which are given by

$$\bigcup_{n \in \mathbb{N}} {}^*(-n, n) \cap S$$

This is a countable union of measurable sets, so it is measurable; but it cannot be specified internally. The union is taken over a non-internal set.

Now we wish to show that, in some sense, the above Loeb measure coincides with the Lebesgue measure λ on \mathbb{R} .

Since μ_L is a function on subsets of the lattice $S \subset {}^*\mathbb{R}$, one might imagine that the following procedure is required to find the “Loeb measure” of a set $B \subseteq \mathbb{R}$ (recalling that, strictly speaking, only subsets of ${}^*\mathbb{R}$ can have Loeb measure):

- (1) Transfer B to the hyperreals;
- (2) Intersect *B with S ;
- (3) Take the Loeb measure of the resulting set.

However, this procedure fails to give the desired answer for the set \mathbb{Q} , which we know to have Lebesgue measure 0. Indeed, ${}^*\mathbb{Q}$ contains S , so it must have infinite Loeb measure under this (faulty) scheme, rather than the 0 we would like if our measure resembles Lebesgue measure. Additionally, if B is not an internal subset of \mathbb{R} , then we cannot necessarily transfer it to ${}^*\mathbb{R}$ in the first place.

In keeping with the idea of “approximate with an infinitesimal mesh” (as opposed to “project onto an infinitesimal mesh”), we instead aim to show that

$$\lambda(B) = \mu_L^+(\text{app}[B])$$

where $\text{app}[B]$, the “approximation of B ”, is defined³ to be

$$\text{app}[B] = \{s \in S : s \text{ is finite and } \text{st}(s) \in B\}$$

Definition 9.7. Given $B \subseteq \mathbb{R}$, we will say B is “Loeb measurable” (with quotation marks) if $\text{app}[B]$ is Loeb measurable, with respect to μ_L^+ . We will say B has “Loeb measure” m (with quotation marks) if $\text{app}[B]$ has Loeb measure m .

This notation is not standard.

³Goldblatt uses notation equivalent to $\text{st}^{-1}(\cdot)$ for $\text{app}[\cdot]$ in Section 16.8 of [4].

Remark. Notice that $\text{app}[B]$ is not necessarily an internal set, because it relies on the predicate “ s is finite”. In fact, $\text{app}[B]$ might not even be Loeb measurable. This happens, for instance, whenever B is not Lebesgue measurable. We will prove this as Theorem 9.6.

Example. In the case $B = \mathbb{R}$, we have

$$\text{app}[B] = \{s \in S : s \text{ is finite}\}$$

which is definitely not an internal set. Nonetheless, $\text{app}[B]$ is still Loeb measurable, because it is a countable (albeit non-internal) union (in ${}^*\mathbb{R}$) of Loeb measurable sets:

$$\bigcup_{n \in \mathbb{N}} (S \cap {}^*(-n, n))$$

Each of the terms in the union is an internal subset of the lattice S , and so is Loeb measurable.

Example (The Cantor set). Throughout this example, we will use the notation $0.a_1a_2\dots$ to denote ternary expansion. Recall that the Cantor set is defined as

$$\mathcal{C} = \{x \in [0, 1] : x \text{ contains no } 1 \text{ in its ternary expansion}\}$$

It is well-known to have Lebesgue measure 0 despite being uncountable.

What is $\text{app}[\mathcal{C}]$?

It is an internal fact of \mathbb{R} that every real in $[0, 1]$ has a ternary expansion. Therefore it is true also in ${}^*\mathbb{R}$: every hyperreal in ${}^*[0, 1]$ has an expansion of the form

$$0.(a_1a_2\dots a_n a_{n+1}\dots)(\dots a_{M-1}a_M a_{M+1}\dots)\dots$$

where M is an example of an infinite hypernatural number, and (to ensure uniqueness of the expansion) we choose the a_i so that it is never the case that from some point on, all the a_j are equal to 2.

The standard part of this number x is precisely

$$0.a_1a_2\dots$$

the “standard part” of its base-3 expansion, where we truncate any infinite-index places.

So, a hyperreal $s \in S$ lies in $\text{app}[\mathcal{C}]$ if and only if the “standard part” of its base-3 expansion consists only of the base-3 digits 0 and 2.

But how many of these are there? Recalling that N is the hypernatural denominator of every element of our lattice $S \subseteq {}^*\mathbb{R}$, we can pick $N = 3^P$ (where P is hypernatural) in such a way that all the elements $\frac{k}{N} \in S$ have hyper-terminating base-3 expansion:

$$0.(a_1a_2\dots)(\dots a_{P-1}a_P)$$

Then

$$\text{app}[\mathcal{C}] = \{(0.a_1a_2\dots)(\dots a_{P-1}a_P) : a_i \neq 1 \text{ for each } i \in \mathbb{N}\}$$

This set is not *a priori* internal, because we built it using the “ $i \in \mathbb{N}$ ” clause. But it is the following (external) nested intersection of Loeb measurable sets:

$$\text{app}[\mathcal{C}] = \bigcap_{n \in \mathbb{N}} \{(0.a_1a_2\dots)(\dots a_{P-1}a_P) : a_i \neq 1 \text{ for each } i < n\}$$

Being a countable intersection of Loeb measurable sets, it is measurable according to the extension μ_L^+ , and its measure is the limit of the measures of the components.

The component $\{(0.a_1a_2\dots)(\dots a_{P-1}a_P) : a_i \neq 1 \text{ for each } i < n\}$ has hyperfinite size

$$2^n 3^{P-n} = \left(\frac{2}{3}\right)^n 3^P$$

where we note that since every such ternary expansion is hyper-terminating, it is automatically “acceptable” in that it is not of the form $0.a_1a_2\dots a_{K-1}a_K2222\dots$ for any (finite or infinite) hypernatural K .

So the “Loeb measure” of \mathcal{C} is

$$\mu_L^+(\text{app}[\mathcal{C}]) = \text{st}\left(\frac{|\text{app}[\mathcal{C}]|}{3^P}\right) = \text{st}\left(\lim_{n \rightarrow \infty} \left[\frac{2}{3}\right]^n\right) = 0$$

Theorem 9.4. *Lebesgue measurable sets are “Loeb measurable”. That is,*

- (1) *The “Loeb measure” of $(a, b) \subset \mathbb{R}$ exists, and is $b - a$.*
- (2) *More generally, if $B \subseteq \mathbb{R}$ is Lebesgue measurable, then it is “Loeb measurable”, and its “Loeb measure” is equal to its Lebesgue measure.*

Proof. (1):

$$\text{app}[(a, b)] = \{s \in S : a < \text{st}(s) < b\} = \bigcup_{n \in \mathbb{N}^{\geq 1}} \left[S \cap^*(a + \frac{1}{n}, b - \frac{1}{n}) \right]$$

This is a nested countable union of internal sets, so its Loeb measure exists and is equal to the limit of the Loeb measures of the individual $S \cap^*(a + \frac{1}{n}, b - \frac{1}{n})$.

But $S \cap^*(a + \frac{1}{n}, b - \frac{1}{n})$ is hyperfinite, because it is an internally specified subset of a hyperfinite set; so it has greatest and least elements g and l respectively, and it is easy to see that $g \simeq b - \frac{1}{n}$ and $l \simeq a + \frac{1}{n}$.

Therefore

$$S \cap^*(a + \frac{1}{n}, b - \frac{1}{n}) = \left\{ \frac{K}{N}, \frac{K+1}{N}, \dots, \frac{L}{N} \right\}$$

for some K, L hypernaturals, where $\frac{K}{N} = l$ and $\frac{L}{N} = g$. The internal cardinality of this set is the hypernatural $L - K + 1$, so the Loeb measure is

$$\text{st}\left(\frac{|S \cap^*(a + \frac{1}{n}, b - \frac{1}{n})|}{N}\right) = \text{st}\left(\frac{L - K + 1}{N}\right) = \text{st}\left(g - l + \frac{1}{N}\right) = b - a - \frac{2}{n}$$

By taking the limit as $n \rightarrow \infty$, we obtain $\text{app}[(a, b)] = b - a$.

(2): This section of the proof will involve working with epsilons, as is characteristic of showing that standard definitions are equivalent to non-standard ones.

Note that $\text{app}[\emptyset] = \emptyset$, so the “Loeb measure” of \emptyset coincides with the Lebesgue measure.

The first part of this theorem states that the “Loeb measure” agrees with Lebesgue measure on a basis of the Borel σ -algebra on \mathbb{R} . There is a uniqueness theorem for measures, which forces the “Loeb measure” and the Lebesgue measure to agree on every Borel set (since they already agree on a basis of the σ -algebra). This is the content of Lemma 16.4.1 of [4].

Now, every Lebesgue measurable set B can be approximated by Borel sets in an arbitrarily fine way: for every ε there are Borel sets C_ε and D_ε such that

$$C_\varepsilon \subseteq B \subseteq D_\varepsilon$$

with $\lambda(B \setminus C_\varepsilon) < \varepsilon$, and similarly $\lambda(D_\varepsilon \setminus B) < \varepsilon$. (Recall this fact from the preamble to Section 9.2 of this essay, as the alternative characterisation of Lebesgue measurability.)

Since C_ε and D_ε are Borel, they are “Loeb measurable”. Certainly

$$\text{app}[C_\varepsilon] \subseteq \text{app}[B] \subseteq \text{app}[D_\varepsilon]$$

but C_ε has $\lambda(C_\varepsilon) = \mu_L^+(C_\varepsilon)$ by the fact that λ and μ_L^+ agree on Borel sets; similarly $\lambda(D_\varepsilon) = \mu_L^+(D_\varepsilon)$.

Therefore $\text{app}[B]$ is μ -approximable (recall Definition 9.5)—strictly speaking, by taking a μ -approximation of C_ε to be the smaller set and a μ -approximation of D_ε to be the larger set, as per Lemma 9.1—so by Theorem 9.3, B is “Loeb measurable”.

Its “Loeb measure” is sandwiched between $\lambda(C_\varepsilon)$ and $\lambda(D_\varepsilon)$ for all $\varepsilon \in \mathbb{R}^{>0}$, and the sets C_ε and D_ε are at most 2ε apart in Lebesgue measure. Therefore the “Loeb measure” of B is equal to the limit of the $\lambda(C_\varepsilon)$ (and also equal to the limit of the $\lambda(D_\varepsilon)$), which is $\lambda(B)$. \square

We now prove the other direction of the equivalence, using Theorem 9.3 on μ -approximability so that we may ensure we are always dealing concretely with “nice” sets. Again the alternative characterisation of Lebesgue measurability (from the preamble to Section 9.2) will come into play: Lebesgue measurable sets are those which may be approximated arbitrarily well by Borel sets.

First, we require a technical lemma, which is Theorem 11.13.1 of Goldblatt [4]. It is technical in the sense that it is potentially sensitive to the explicit construction of the hyperreals; it is a fact of the ultrapower construction. If we wished to remain agnostic about the construction of the hyperreals, it would be necessary to insist that we were working in a *countably saturated* system—a requirement which is satisfied by the ultrapower construction—which is to say that the intersection of any decreasing sequence of nonempty internal sets is nonempty. Goldblatt (in Theorem 11.10.1 of [4]) calls this a “delicate” fact about the ultrapower; we will not prove it here.

Lemma 9.5. *Let X be an internal subset of ${}^*\mathbb{R}$. Then $\{\text{st}(x) : x \in X\}$ is closed as a subset of \mathbb{R} .*

Proof. Let $r \in \mathbb{R}$ be a limit point of $\{\text{st}(x) : x \in X\}$; say r is the limit of the sequence

$$(\text{st}(x_i))_{i \in \mathbb{N}}$$

We need $r = \text{st}(y)$ for some $y \in X$.

By omitting some terms of the sequence if necessary, we may pick each x_i so that

$${}^*|r - x_i| < \frac{1}{i}$$

That is,

$$x_i \in X \cap {}^*\left(r - \frac{1}{i}, r + \frac{1}{i}\right)$$

Given countable saturation, there is some $y \in X$ lying in all the $X \cap {}^*(r - \frac{1}{n}, r + \frac{1}{n})$; then $\text{st}(y) = r$ as required, since for all $n \in \mathbb{N}$, we have

$$|r - y| < \frac{1}{n}$$

□

Theorem 9.6. *Let $B \subseteq \mathbb{R}$ be “Loeb measurable” (that is, $\text{app}[B]$ is Loeb measurable). Then B is Lebesgue measurable, and $\lambda(B)$ is equal to $\mu_L^+(\text{app}[B])$.*

Proof. Let B be “Loeb measurable”. It is enough to show that B is Lebesgue measurable; then Theorem 9.4 tells us that its “Loeb measure” is equal to its Lebesgue measure.

There are two cases: B ’s “Loeb measure” is either finite or the literal value ∞ .

If it is finite, then by Theorem 9.3, $\text{app}[B]$ is μ -approximable. (Recall Definitions 9.5 and 9.6: $\mathcal{A} = \mathcal{P}_I(S)$ is the collection of internal subsets of the lattice S , and B has the property that for every $\varepsilon \in \mathbb{R}^{>0}$, there are sets $C, D \in \mathcal{P}_I(S)$ such that $\mu_L(D \setminus C) < \varepsilon$ and $C \subseteq \text{app}[B] \subseteq D$.)

Let $\varepsilon \in \mathbb{R}^{>0}$, and pick ${}^*C, {}^*D \in \mathcal{A}$ approximating $\text{app}[B]$ to within ε . (We label them *C and *D as a cue to the fact that these are internal subsets of the hyperreal lattice.)

Now, we would like to use *C to build a set of reals which approximates B from below in the Lebesgue sense. There are two possible sets to take: the set of all standard parts of elements of *C , or C itself. (These are in general not necessarily the same set: consider ${}^*(0, 1)$, which contains an infinitesimal $\varepsilon > 0$ whose standard part is 0, even though 0 does not lie in $(0, 1)$.) It turns out that the correct set is the set

$$C_\varepsilon := \{\text{st}(c) : c \in {}^*C\}$$

By Lemma 9.5, C_ε is closed; that makes it Borel, as is required for the alternative definition of Lebesgue measurability.

Similarly the set D_ε is open and hence Borel, where

$$D_\varepsilon := \mathbb{R} \setminus \{\text{st}(d) : d \in S \setminus {}^*D\}$$

(This definition is a little cunning; it is designed to maximise the symmetry between C_ε and D_ε . It reflects the idea of “cover the set and its complement” from the preamble before Definition 9.5 of μ -approximability.)

We have

$$C_\varepsilon \subseteq \{\text{st}(b) : b \in \text{app}[B]\} = B \subseteq D_\varepsilon$$

by taking standard parts of the true-by-definition

$${}^*C \subseteq \text{app}[B] \subseteq {}^*D$$

We just need $\lambda(D_\varepsilon) - \lambda(C_\varepsilon)$ to be small. But we have just proved (in Theorem 9.4) that Lebesgue measurable sets have “Loeb measure” equal to their Lebesgue measure, and we already know C_ε and D_ε are Borel and hence Lebesgue measurable, so

$$\lambda(D_\varepsilon) - \lambda(C_\varepsilon) = \mu_L^+(\text{app}[D_\varepsilon]) - \mu_L^+(\text{app}[C_\varepsilon])$$

Finally, $\mu_L^+(*C) \leq \mu_L^+(\text{app}[C_\varepsilon])$ because $*C \subseteq \text{app}[C_\varepsilon]$. Indeed, to recap the definitions,

$$*C \in \mathcal{P}_I(S) \text{ and } C_\varepsilon = \{\text{st}(c) : c \in *C\}$$

so $c \in *C$ implies $\text{st}(c) \in C_\varepsilon$ which means $c \in \{s \in S : \text{st}(s) \in C_\varepsilon\} = \text{app}[C_\varepsilon]$.

And likewise $\mu_L^+(\text{app}[D_\varepsilon]) \leq \mu_L^+(*D)$ because $\text{app}[D_\varepsilon] \subseteq *D$. Indeed, if $s \in \text{app}[D_\varepsilon]$ then $\text{st}(s) \in D_\varepsilon$ so

$$\text{st}(s) \in \mathbb{R} \setminus \{\text{st}(d) : d \in S \setminus *D\}$$

That is, $s \in *D$.

Therefore

$$\lambda(D_\varepsilon) - \lambda(C_\varepsilon) \leq \mu_L(*D) - \mu_L(*C) < 2\varepsilon$$

so we have shown that B may be approximated arbitrarily well by Borel sets, so it is Lebesgue measurable. This completes the first case.

If instead B 's ‘‘Loeb measure’’ is the literal ∞ , then we may express $B \subseteq \mathbb{R}$ as the following union:

$$\bigcup_{n \in \mathbb{N}} B \cap (-n, n)$$

We can show that B itself is Lebesgue measurable by showing that each $B \cap (-n, n)$ is Lebesgue measurable.

But note that

$$\text{app}[B \cap (-n, n)] = \text{app}[B] \cap \text{app}[(-n, n)]$$

so that set is Loeb measurable, being the intersection of two Loeb measurable sets. It has finite Loeb measure, since

$$\mu_L^+(\text{app}[B \cap (-n, n)]) \leq \mu_L^+(\text{app}[(-n, n)]) = 2n$$

(using Theorem 9.4 to deduce that the ‘‘Loeb measure’’ of $(-n, n)$ is equal to its Lebesgue measure).

Therefore we are done by the first case, since $B \cap (-n, n)$ is a ‘‘Loeb measurable’’ set with finite ‘‘Loeb measure’’. \square

Remark. The following remark is more in a handwaving motivational vein. Very loosely, the lattice S can be viewed as a hyperfinite probability space, and the function μ on it can be viewed as simply a constant (infinite) multiple of a probability measure. (A probability measure would have to have denominator $2N^2 + 1$ rather than N , to make the total probability 1 rather than infinite.) Then intuitively every point of a set $B \subset \mathbb{R}$ is assigned *infinitesimal* measure, rather than the absolute zero measure assigned by the standard theory. This is a powerful point in favour of the non-standard approach to measure theory: now it is only literally impossible events that are assigned literally zero measure, and all other events are assigned some positive, possibly infinitesimal, measure. (Of course, on taking standard parts, producing μ_L and therefore bringing the results back into the realm of standard probability theory, we recover the possibility that a non-impossible event may have zero measure.)

10. BROWNIAN MOTION

Our primary reference for this section is Hurd and Loeb [6], where we use Chapter IV.6.

Brownian motion is a model of the motion of a small particle (such as a speck of pollen) suspended in a fluid (such as stationary air). It is physically observable with a microscope, and it occurs because molecules of the surrounding fluid collide with the particle in a random way, causing random changes of both course and speed.

Definition 10.1 (Brownian motion). *Consider the random position X_t of a particle on the real line at time t , where t varies between 0 and 1. Let Ω be the state space of the variables X_t . We say the $[0, 1]$ -indexed collection of random variables $\langle X_t : 0 \leq t \leq 1 \rangle$ is a Brownian motion if:*

- (1) $X_0 = 0$. That is, the particle starts at the origin.
- (2) Given any sequence of nonempty intervals

$$[s_1, t_1], [s_2, t_2], \dots, [s_n, t_n]$$

where $s_1 < t_1 \leq s_2 < t_2 < \dots \leq s_n < t_n$, we have the random variables

$$X_{t_1} - X_{s_1}, \dots, X_{t_n} - X_{s_n}$$

all independent. That is, the particle's overall movement during any time period is not affected by its overall movement during any other time period.

- (3) If $s < t$, then

$$\mathbb{P}(\{\omega \in \Omega : X_t(\omega) - X_s(\omega) \leq \alpha\}) = \Phi\left(\frac{\alpha}{\sqrt{t-s}}\right)$$

for Φ the Gaussian integral

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-u^2/2} du$$

That is, if the particle's position at time s is known, then its position at time t is normally distributed.

A non-standard way to obtain a Brownian motion is derived from considering a random walk.

Let $\Omega^{(n)}$ be the space of all sequences $\omega = (\omega_1, \omega_2, \dots, \omega_n)$, where each $\omega_i = \pm 1$. Let a particle move by a distance of $1/\sqrt{n}$ every time-step $t_k = k/n$, in the direction indicated by ω_k .

Then the position of the particle at time t following walk-sequence ω is given by

$$(2) \quad \chi(t, \omega) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \omega_i$$

The reason for the distance per step being $1/\sqrt{n}$ is so that the resulting walk has the right variance to satisfy the normal distribution requirement that is item (3) in Definition 10.1; this is in anticipation of $\langle \text{st}(\chi(t, \cdot)) : t \in [0, 1] \rangle$ being a Brownian motion when we let n , the length of the sequences ω , be an infinite hypernatural.

Theorem 10.1. *Let χ be defined as in Equation 2. Let $X_0 = \text{st}(\chi(0, \cdot))$, a random variable on state space $\Omega^{(N)}$, where N is an infinite hypernatural. Then X_0 is the constant 0.*

Proof. This is immediate: the sum defining X_0 is the empty sum. \square

Theorem 10.2. *Let χ be defined as in Equation 2. Let $X_t := \text{st}(\chi(t, \cdot))$ be $[0, 1]$ -indexed random variables on states $\Omega^{(N)}$. Then for any $s_1 < t_1 \leq s_2 < t_2$, the random variables*

$$X_{t_1} - X_{s_1}, X_{t_2} - X_{s_2}$$

are independent.

Proof.

$$X_{t_1}(\omega) - X_{s_1}(\omega) = \text{st}(\chi(t_1, \omega) - \chi(s_1, \omega)) = \text{st} \left(\frac{1}{\sqrt{N}} \sum_{i=[Ns_1]+1}^{[Nt_1]} \omega_i \right)$$

$$X_{t_2}(\omega) - X_{s_2}(\omega) = \text{st}(\chi(t_2, \omega) - \chi(s_2, \omega)) = \text{st} \left(\frac{1}{\sqrt{N}} \sum_{i=[Ns_2]+1}^{[Nt_2]} \omega_i \right)$$

The two random variables are therefore clearly independent: their values are defined by sums ranging over disjoint sections of the input ω . \square

This proof easily generalises to the second property of Definition 10.1 with n rather than two almost-disjoint intervals: in each interval, the sum is being taken over disjoint regions of the input ω .

To prove the third property of Definition 10.1, we will need a non-standard analogue of the Central Limit Theorem.

Definition 10.2 (**-independence of random variables*). *Let $(X_n)_{n \in {}^*\mathbb{N}}$ be an internal sequence of random variables. We say they are *-independent if, given any internal M -subtuple $(X_{n_i})_{i=1}^M$, and given any internal M -tuple $(\alpha_i)_{i=1}^M$ of reals, we have*

$$\mathbb{P}(\{\omega \in \Omega : X_1(\omega) < \alpha_1, \dots, X_M(\omega) < \alpha_M\}) = \prod_{i=1}^M \mathbb{P}(\{\omega \in \Omega : X_i(\omega) < \alpha_i\})$$

That is, "all hyperfinite subcollections are independent".

Lemma 10.3 (Non-standard Central Limit Theorem). *Let $(X_n)_{n \in {}^*\mathbb{N}}$ be an internal $*$ -independent identically distributed sequence of random variables. Suppose the mean of each random variable is 0 and the variance of each is 1. Then for any infinite hypernatural M and any $\alpha \in {}^*\mathbb{R}$, have*

$$\mathbb{P} \left(\left\{ \omega \in \Omega : \frac{1}{\sqrt{M}} \sum_{n=1}^M X_n(\omega) \leq \alpha \right\} \right) \simeq {}^*\Phi(\alpha)$$

where we recall that Φ is the cumulative density of the normal distribution, as in condition 3 of Definition 10.1.

We join Hurd and Loeb in omitting the proof of this lemma; it is a fairly short but rather unenlightening consequence of the transfer principle applied to the standard Central Limit Theorem. The proof may be found as Theorem 21 from Anderson [9].

Theorem 10.4. *Property 3 of Definition 10.1 holds for the hyperfinite random walk. That is, if χ is defined as in Equation 2, and $X_t = \text{st}(\chi(t, \cdot))$, then*

$$\mathbb{P}(\{\omega \in \Omega^{(N)} : X_t(\omega) - X_s(\omega) \leq \alpha\}) = \Phi \left(\frac{\alpha}{\sqrt{t-s}} \right)$$

Proof. The left-hand side is

$$\mathbb{P} \left(\left\{ \omega \in \Omega^{(n)} : \text{st} \left(\frac{1}{\sqrt{N}} \sum_{i=[Ns]+1}^{[Nt]} \omega_i \right) \leq \alpha \right\} \right)$$

In order to apply the non-standard Central Limit Theorem, we need to manipulate this into a form which has a sum of random variables, rather than the *standard part* of a sum of random variables.

We can convert standard parts of sums into simple sums by passing to a limiting process: the left-hand side is precisely

$$\lim_{r \rightarrow \infty} \mathbb{P} \left(\left\{ \omega \in \Omega^{(n)} : \frac{1}{\sqrt{N}} \sum_{i=[Ns]+1}^{[Nt]} \omega_i \leq \alpha + \frac{1}{r} \right\} \right)$$

Now, to get this into precisely the form of the non-standard Central Limit Theorem, we must rewrite the sum so that its indices are from 1 to some constant. Letting

$$T = [Nt] - [Ns]$$

have

$$\lim_{r \rightarrow \infty} \mathbb{P} \left(\left\{ \omega \in \Omega^{(n)} : \frac{1}{\sqrt{T}} \sum_{i=1}^T \omega_{i+[Ns]} \leq \left(\alpha + \frac{1}{r} \right) \frac{\sqrt{N}}{\sqrt{T}} \right\} \right)$$

Let us assume for the moment that the collection of $Y_i(\omega) := \omega_{i+[Ns]}$ is $*$ -independent. Certainly they all have mean 0, variance 1, and are identically distributed.

By the non-standard Central Limit Theorem, each term of the limit is infinitesimally close to

$$*\Phi\left(\left(\alpha + \frac{1}{r}\right) \frac{\sqrt{N}}{\sqrt{T}}\right)$$

But Φ is uniformly continuous because it is increasing, bounded and continuous. (Recall the non-standard definition of uniform continuity from Section 4.1: infinitesimally perturbing x induces only an infinitesimal perturbation in $*\Phi(x)$.) Therefore

$$\text{st } *\Phi\left(\left(\alpha + \frac{1}{r}\right) \frac{\sqrt{N}}{\sqrt{T}}\right) = \Phi\left(\text{st}\left(\left(\alpha + \frac{1}{r}\right) \frac{\sqrt{N}}{\sqrt{T}}\right)\right)$$

Recall that

$$T = \lfloor Nt \rfloor - \lfloor Ns \rfloor$$

so

$$\frac{\sqrt{N}}{\sqrt{T}} = \frac{\sqrt{N}}{\sqrt{\lfloor Nt \rfloor - \lfloor Ns \rfloor}} \simeq \frac{1}{\sqrt{t-s}}$$

because

$$\lim_{n \rightarrow \infty} \frac{n}{\lfloor nt \rfloor - \lfloor ns \rfloor} = \frac{1}{t-s}$$

Therefore each term of our limit is infinitesimally close to

$$\Phi\left(\left(\alpha + \frac{1}{r}\right) \text{st}\left(\frac{\sqrt{N}}{\sqrt{T}}\right)\right) = \Phi\left(\left(\alpha + \frac{1}{r}\right) \frac{1}{\sqrt{t-s}}\right)$$

Taking the limit as $r \rightarrow \infty$, obtain

$$\Phi\left(\frac{\alpha}{\sqrt{t-s}}\right)$$

exactly as required.

We still need to show that the collection of $Y_i(\omega) := \omega_{i+\lfloor Ns \rfloor}$ is $*$ -independent, so as to justify the use of the non-standard Central Limit Theorem and complete the proof. But this is immediate: each Y_i inspects a different part of the input. □

Together, Theorems 10.1, 10.2 and 10.4 prove that the non-standard “random walk” approach yields a Brownian motion as defined in Definition 10.1.

Remark. In fact, it is possible to prove that almost all of these Brownian motions are continuous: for almost all fixed ω , it is the case that $t \mapsto X_t(\omega)$ is continuous. This is Theorem 6.13 of Hurd and Loeb [6].

That is, this scheme of creating Brownian motions almost always creates “physically realistic” motions, in the sense that the paths are continuous.

REFERENCES

- [1] Abraham Robinson, *Non-standard analysis*, Princeton Landmarks in Mathematics
- [2] André Pétry, *Analyse Infinitésimale: une présentation non standard*, first edition, Céfal
- [3] Isaac Davis, *An Introduction to Nonstandard Analysis*,
www.math.uchicago.edu/~may/VIGRE/VIGRE2009/REUPapers/Davis.pdf
- [4] Robert Goldblatt, *Lectures on the Hyperreals: an Introduction to Nonstandard Analysis*, Springer, Graduate Texts in Mathematics
- [5] Michael Spivak, *Calculus*, third edition, Cambridge University Press
- [6] A. E. Hurd and P. A. Loeb, *An Introduction to Nonstandard Real Analysis*, Academic Press Inc
- [7] Paul R Halmos, *Measure Theory*, Springer, Graduate Texts in Mathematics
- [8] J. H. Williamson, *Lebesgue Integration*, Courier Corporation
- [9] Robert M. Anderson, *A non-standard representation for Brownian motion and \hat{I} to integration*, Israel Journal of Mathematics, Vol. 25, 1976