Hyperfinite Probability Theory and Stochastic Analysis within Edward Nelson's Internal Set Theory

Revised Version (1. Februar 2011)Diplomarbeit

des
Department Mathematik
Friedrich Alexander Universität
Erlangen-Nürnberg

Thomas Runge
geb. 20. September 1981 in Nürnberg

Betreuer	Prof. Dr. Günter Leugering
	Prof. Dr. Ulrich Rüde
	Dr. Wolfgang Degen
Angemeldet	29. Juli 2010
Abgegeben	5. August 2010

Erklärung:

Ich versichere, dass ich die Arbeit ohne fremde Hilfe und ohne Benutzung anderer als der angegebenen Quellen angefertigt habe und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Ausführungen, die wörtlich oder sinngemäß übernommen wurden, sind als solche gekennzeichnet.

Erlangen, 5. August 2010 .....

Contents

Chapter 1

Introduction

The Differential and Integral Calculus was originally based on the Infinitesimal Calculus (independently) introduced by GOTTFRIED LEIBNIZ and ISAAC NEWTON. The problem with their Infinitesimal Calculi was that the given descriptions all turned out to be inconsistent.

From the modern point of view the main defect was that mathematicians at that time had no precise formal language. LEIBNIZ suggested a field containing every real number and additionally some elements that are greater than 0, but smaller than every real number greater than 0. This field was expected to have the same properties as the field of real numbers – which is obviously a contradiction, if one wants to preserve absolutely every definable property. For example, we know that no proper field extension of $\mathbb{R}$ can satisfy the Archimedean property.

The first solution of these contradictions was given by KARL WEIERSTRASS who strictly replaced the use of infinitesimals by limits of sequences. Definitions and proofs became longer, more technical and less intuitive, but the differential and integral calculus he introduced was free of inconsistencies and hence became the only accepted method for about one century, until the 1960s.

Independently of this problem the introduction of formal languages by GOTTLOB FREGE gave a possibility to distinguish in general between the statements (about the real numbers) that should hold (in a field containing real numbers and infinitesimals) and those statements that cannot be expected to hold.

The first step towards a consistent version of an infinitesimal calculus was ALBERT THORALF SKOLEM's essay [Sk34]. In this essay SKOLEM constructed a model of Arithmetic that was not isomorphic to the natural numbers, but contained elements that were greater than any number in the usual (or standard) model.

In 1961 ABRAHAM ROBINSON published the first essays about Nonstandard Analysis. ROBINSON used a precise formal language and an analogon to SKOLEM's nonstandard model of Arithmetic to construct a field extension $\mathbb{R} \subset \mathbb{R}^*$ in a way, that a statement about the set of real numbers, $\mathbb{R}$, is true if and only if a

corresponding statement about the set of *hyper*real numbers, $\mathbb{R}^*$, is true. For example, because the Archimedean property is true for $\mathbb{R}$, there is a corresponding property that is true for $\mathbb{R}^*$, but the natural numbers of the original statement have to be replaced by the set of *hyper*natural numbers (that is the set that SKOLEM constructed for his model).

The great logician KURT GÖDEL said in 1973 that *there are good reasons to believe that the nonstandard analysis, in some version or other, will be the analysis of the future.* One alternative version was given in 1977 by EDWARD NELSON in his paper [Ne77]. He used an idea of ROBINSON's approach, namely the differentiation between standard and nonstandard entities to create a system of axioms that allows one to find infinitesimally small *real* numbers (without a field extension), and hence allows to use an infinitesimal calculus as Newton suggested without model theoretical tools. Since we have a system of axioms we can to a great extent argue in a mere syntactic way, like most mathematicians do in their everyday work.

Beside NELSON's *Internal Set Theory* there are other approaches to an Alternative or Nonstandard Set Theory, e.g. KAREL HRBACEK's theory that can handle external sets in a better way – NELSON just forbids their existence (as sets in the sense of his set theory), but uses them anyway by referring to an arbitrary model within a metatheory (see appendix A).

In this work we will use NELSON's *Internal Set Theory* since this system of axioms has the advantage of being simpler and more intuitive than the several other versions of Nonstandard Set Theory. Within this context we will study stochastic processes with the help of nonstandard finite sets, as ROBERT M. ANDERSON or HOWARD JEROME KEISLER did in ROBINSON's model.

1.1 Structure and Outline

In chapter 2 we will introduce *Internal Set Theory* and define basic concepts like the $\approx$-relation and hyperfinite sets. Since we want to concentrate on Hyperfinite Stochastical Analysis we will not give many examples of how to use infinitesimals in Differential or Integral Calculus.

In chapter 3 we will turn to more practical applications of the defined concepts by using hyperfinite sets and infinitesimal and infinite real numbers to construct a 'radically elementary probability theory'[1] using nonstandard finite measure and probability spaces as models of continuous phenomena. We will construct the Lebesgue integral in a nonstandard way and will be able to formulate nonstandard

[1] So the title of a publication of NELSON, [Ne87].

versions of the central limit theorem and the laws of large numbers.

In chapter 4 we will construct a nonstandard version of the Itô integral, using stochastic processes on (nonstandard) finite probability spaces; with this section we want to demonstrate the power of NELSON's *Internal Set Theory* (even without external sets, and therefore without Loeb spaces).

In the first appendix we give a formulation of the use of external sets within *Internal Set Theory* and formulate the theorem of Loeb. We will also compare the construction of Loeb spaces with (our notion of) standardizations of measure spaces.

The second appendix completes the thesis by constructing of a model of IST and giving a proof of the conservativity of IST relative to ZFC.

Chapter 2

Internal Set Theory

In this chapter we will describe the axiomatic approach of NELSON's *Internal Set Theory* (*IST*). The goal is a short introduction for readers who don't know much about this theory; it is oriented to preparing theorems we will need later. We will not give many examples of how to use nonstandard methods for proving analytic theorems in a smoother way.

Readers who are interested in such examples may study e.g. [Ne77] where several such applications are discussed.

2.1 Axiomatic Description

The language of *IST* contains besides logical symbols (including equality $=$) two relation symbols: the binary relation $\in$ as used in ZFC and a new unary relation S that we will interpret as '...is standard'. We will call a formula that does not use the symbol S an $\mathcal{L}_{\in}$-formula or sometimes (as NELSON does) an *internal formula*; if we want to underline that a formula is using S (in a non-trivial way) we call it a (proper) $\mathcal{L}_{\{\in,S\}}$-formula or sometimes an *external formula*.

The axioms of *IST* as described in [Ne77] are the axioms of ZFC (where the axiom schemata only use $\mathcal{L}_{\in}$-formulas) plus three axiom schemata that we will describe here. This extension of ZFC is conservative[1] as Nelson proved in his paper,[2] a fortiori it is consistent, if ZFC is consistent.

We will use some abbreviations for quantifiers that should be self-explanatory: $\forall^S x(\Phi)$, $\exists^S x(\Phi)$, $\forall^{FIN} x(\Phi)$ and $\exists^{FIN} x(\Phi)$ mean $\forall x(S(x) \to \Phi)$, $\exists x(S(x) \wedge \Phi)$,

[1] An extension T of a theory S is called conservative, if every statement in the language of S that can be proved in T can be proved in S itself. This means that we can not expect T to prove new S-theorems, but some of the proofs in T may be shorter or easier to understand than the 'classical' proofs in S.

[2] See B.3.1 for an alternative proof, given in [CK90].

$\forall x(FIN(x) \to \Phi)$ and $\exists x(FIN(x) \wedge \Phi)$, respectivaly.

Abbreviations in common use like $\forall x \in X(\Phi)$ will be used without further explanation.

The symbol $FIN(x)$ for 'x is finite' is defined as: There is an injection into a proper initial segment of the natural numbers; please note that since we use the axiom of choice in this work any other (reasonable) definition of finiteness is equivalent to this one.

We will denote logical formulas by Greek capital letters like $\Phi, \Psi, \dots$ with an index if necessary. If we write only the letter we allow occurrencies of arbitrary free variables. If we want to highlight the occurrence of certain free variables we use the notation $\Phi(v_1, \dots, v_n)$ and understand that the free variables in Φ are any of the variables $v_1, \dots, v_n$. If there may be other free variables we will write $\Phi(v_1, \dots, v_n, \dots)$. As customary we denote by $\Phi(t, \dots)$ the formula that arises by replacing any occurrence of the free variable v in $\Phi(v, \dots)$ by the term t (understandig that v is free for t).

2.1.1 The Transfer principle

Perhaps the primary merit of ROBINSON was his transfer principle, that states that a theorem about the real universe is true if and only if a corresponding theorem about the hyperreal universe is true. The theory IST transforms this basic idea into an axiom schema.

Transfer principle: For every $\mathcal{L}_\in$-formula $\Phi(v_0, v_1, \dots, v_n)$:

$$\forall^S t_1, \dots, t_n(\forall^S x\ \Phi(x, t_1, \dots, t_n) \Rightarrow \forall x\ \Phi(x, t_1, \dots, t_n)) \qquad \text{(T)}$$

The other direction $\Leftarrow$ is true by pure logic. A first example of a consequence of **(T)** is the following.

Theorem 2.1.1 (Internal Definition Principle). If an entity x can be described uniquely by an $\mathcal{L}_\in$-formula using only standard parameters, then x is standard.

Proof. Directly from **(T)** and its converse direction we can conclude for any $\mathcal{L}_\in$-formula Φ that

$$\forall^S t_1, \dots, t_n(\exists^S x \Phi(x, t_1, \dots, t_n) \Leftrightarrow \exists x \Phi(x, t_1, \dots, t_n))$$

Since we can describe x uniquely by an internal formula Ψ we get the claim together with the ZFC-statement $\exists x(\Psi(x, \dots) \wedge \forall y : \Psi(y, \dots) \Rightarrow x = y)$. □

Directly from this theorem we learn that the sets $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$, $\aleph_{17}$, etc. are standard, as well as the (real) numbers $0, 1, 2, \ldots, \pi, e, \sqrt{2}, \ldots$, and so on. Moreover: If f is standard and $x \in dom(f)$ is standard then $f(x)$ is standard, what opens a wide area of standard entities like powersets $\mathcal{P}(X)$, sets of maps X^Y or intersections $X \cap Y$ and unions $X \cup Y$ of any standard X and Y.

Another application of **(T)** is a possibility of proving $\mathcal{L}_\in$-theorems of the form $\forall x : \Phi(x, \ldots)$ containing only standard parameter. We will see that we can formulate some basic concepts (like continuity of real functions) in a more intuitive way for standard entities. Then we can use this more intuitive formulations to prove $\forall^S x : \Phi(x, \ldots)$ and by **(T)** this proves $\forall x : \Phi(x, \ldots)$.

2.1.2 The Idealization principle

The first steps into an Infinitesimal Calculus as LEIBNIZ and others imagined it used some fuzzy ideas of infinitesimal numbers that are greater than 0, but smaller than any positive real number. Obviously this formulation is contradictory (if the infinitesimal numbers should be real numbers) and hence the use of infinitesimals was replaced by WEIERSTRASS by the use of null sequences, which is indeed far less intuitive.

In *IST* we want to define the notion of infinitesimal numbers with the help of the standard real numbers as real numbers that are greater than 0, but smaller than any positive *standard* real number. The idealization principle states the existence of such 'ideal' elements in a more general formulation.

Idealization principle: For every $\mathcal{L}_\in$-formula $\Phi(x, y, \ldots)$ (where the variable z does not occur in Φ):

$$\forall^{S,FIN} z \exists x \forall y \in z \Phi(x, y, \ldots) \Leftrightarrow \exists x \forall^S y \Phi(x, y, \ldots) \qquad \text{(I)}$$

Example. If we use the statement $x \in \mathbb{R} \wedge (y \in \mathbb{R} \wedge y > 0 \rightarrow 0 < x < y)$ the left-hand side is surely true, e.g. we set $x := \frac{min\{y \in z | y > 0\}}{2}$ (or chose an arbitrary x if there are no positive real numbers in z). Now the right-hand side states the existence of infinitesimal real numbers as we described them above.

The definition of infinitesimal numbers together with some conclusions will be given in section 2.2. $\diamond$

By **I** we get two fundamental statements about the standard elements of the IST set universe.

Theorem 2.1.2. A set X is standard and finite if and only if every element $x \in X$ is standard.

Proof. Use the idealization principle with the formula $x \in X \wedge x \neq y$. Then the right-hand side (of **(I)**) is equivalent to $\exists x \in X \neg S(x)$. We use idealization with the negation of this statement:

$$\forall x \in X : S(x) \Leftrightarrow \exists^{S,FIN} z \forall x \exists y \in z (x \notin X \vee x = y) \Leftrightarrow \exists^{S,FIN} z : X \subseteq z$$

If X is a standard finite set, then the right-hand side is true (take $z := X$), and hence every element of X is standard.

Conversely, if every element of X is standard, then there is a standard and finite set $z \supseteq X$. But every element of the standard finite set $\mathcal{P}(z)$ is standard (and of course finite), hence X is standard and finite. □

Corollary 2.1.3. Every infinite set contains nonstandard elements.

In particular there are nonstandard natural and real numbers. We will see some properties of these numbers in section 2.2.

Theorem 2.1.4. There is a finite set F that contains all standard entities.[3]

Proof. The statement $\forall^{S,FIN} z \exists F \forall x \in z (x \in F \wedge FIN(F))$ says only that there is a finite superset for every standard and finite set. That is of course true.

By idealization we get the claim. □

Such a set F is finite, but it is not standard (if it were standard, then it would, by transfer, contain any set). Furthermore there is no smallest set that contains all standard entites, because if there were such a set we had two possible cases: Either F were exactly the set of all standard elements, but then F were standard, since any of its elements were standard, but that cannot be; or we could find an element $x \in F$ that was nonstandard, but then the set $F \setminus \{x\}$ was a smaller set that would still contain any standard elements, a contradiction to the assumption that F was the smallest one.

Such sets F are – even if not uniquely definable – important for our work, as the following frequently used argument demonstrates: Let F be a finite set that contains all standard elements and $\Phi(x, \dots)$ an $\mathcal{L}_{\in}$-formula containing only standard parameters. Then if we can prove that $\forall x \in F(\Phi(x, \dots))$ we know in particular $\forall^S x(\Phi(x, \dots))$, and then we can use **(T)** to get $\forall x(\Phi(x, \dots))$.

[3]

2.1.3 The Standardization principle

Recall that we are allowed to use only $\mathcal{L}_{\in}$-formulas in the original ZFC-axioms. One well-known axiom schema is the axiom schema of separation that allows us to build the set of all $x \in y$ with $\Phi(x, \dots)$ for any ZFC-statement Φ.

It is important that we have not a separation axiom for arbitrary $\mathcal{L}_{\{\in,S\}}$-formulas, because if we had we could for example build the set of all standard natural numbers N, that was, by theorem 2.1.2 a standard finite set. Then we could define a maximal natural number $n := max(N)$ that were by theorem 2.1.1 also standard – but if it is a standard natural number then $n + 1 > max(N)$ were a standard natural number with $n + 1 \notin N$.

The idea of Standardization is to formulate a corresponding axiom schema for arbitrary formulas.

Standardization principle: For every $\mathcal{L}_{\{\in,S\}}$-formula $\Phi(z, \dots)$ (where the free variable y does not occur in Φ):

$$\forall^S x \exists^S y \forall^S z (z \in y \Leftrightarrow (z \in x \wedge \Phi(z, \dots))) \tag{S}$$

Now we can define standard sets with the help of any $\mathcal{L}_{\{\in,S\}}$-formula, but we know only that the defining statement holds for the standard elements. Since the existence of such a set is stated by **(S)** and the uniqueness follows directly from **(T)** we can write $\mathcal{S}_{z \in X}(\Phi(z, \dots))$ for this set.

Theorem 2.1.5. If for $x \in \mathbb{R}$ there is a standard number $r > |x|$, then there is a unique standard number ${}^\circ x$ such that $|x - {}^\circ x| < s$ for every standard real number $s > 0$.

Proof. Let $A(x) = \mathcal{S}_{z \in \mathbb{R}}(z < x) \subseteq \mathbb{R}$. Since $r > |x|$ we know that $A(x)$ is limited from above and not the empty set, because $-r \in A(x)$, and hence there is a real number ${}^\circ x = sup(A(x))$. By the Internal Definition Principle (2.1.1) we know that ${}^\circ x$ is a standard number.

Suppose that ${}^\circ x \in A(x)$, and let $s > 0$ be a standard real number. Then ${}^\circ x - s < {}^\circ x < x$; on the other hand ${}^\circ x + s > {}^\circ x$ is as a standard number not an element of $A(x)$, hence ${}^\circ x + s \geq x$. Now we have seen that for any standard real number $s > 0$ we have ${}^\circ x - s \leq x \leq {}^\circ x + s$, that was the claim.

The case ${}^\circ x \notin A(x)$ is proved analogue. □

As we have seen in the proof we can use this axiom schema for set definitions using external formulas; we also have seen that we don't know much about the

nonstandard elements: Since every standard number less then $^\circ x$ is an element of the standard set $A(x)$ and no standard number greater then $^\circ x$ is an element of the standard set $A(x)$ we know by **(T)** that $(-\infty, {}^\circ x) \subseteq A(x)$ and $({}^\circ x, \infty) \subseteq \mathbb{R}\backslash A(x)$. In the case ${}^\circ x \notin A(x)$ we have $x \leq {}^\circ x = inf(A(x))$, and that means $x \in A(x)$ – even if we defined $A(x)$ by the formula $z < x$, which is for $z = x$ false.

As we don't know anything about the nonstandard elements of the resulting set (at least without further considerations), $\mathcal{S}_{z\in X}(\Phi)$ could sometimes have some unexpecteded properties.

Example. Let n be a nonstandard natural number and let $S_0 := \mathcal{S}_{i\in\mathbb{N}}(i \leq n)$ and $S_1 := \mathcal{S}_{i\in\mathbb{N}}(i > n)$.

Since a natural number in ZFC is the set of all smaller natural numbers we get that a natural number is standard if and only if all smaller natural numbers are standard. In particular, there is no standard natural number that is greater than n. Hence we know that S_0 contains all standard natural numbers, but S_1 contains no standard natural numbers. Since both are standard sets we get $S_0 = \mathbb{N}$ and $S_1 = \emptyset$, because the only standard subset of $\mathbb{N}$ that contains all standard natural numbers is $\mathbb{N}$ itself by **(T)**, and a standard subset of $\mathbb{N}$ that contains no standard numbers cannot contain any elements.

In this example we used an $\mathcal{L}_\in$-formula to define the sets S_0 and S_1, hence we can also use the separation axiom schema to get a set $S_3 = \{i \in \mathbb{N} | i \leq n\}$ This set is obviously nonstandard. The notation $\{0, 1, 2, \ldots, n\}$ is a common abbreviation for sets of this form. We will use it to denote finite sets, but we should be aware that nonstandard finite sets have some interesting properties – more about nonstandard finite sets in section 2.2.2. $\diamond$

Theorem 2.1.6. Let A be a set, such that there is a standard set $B \supseteq A$. Then there is a standard set $\mathcal{S}(A)$ such that for any standard x:

$$x \in A \Leftrightarrow x \in \mathcal{S}(A)$$

Proof. We can use **(S)** with the statement $z \in A$ to set

$$\mathcal{S}(A) := \mathcal{S}_{x\in B}(x \in A)$$

$\square$

Definition 1. The set $\mathcal{S}(A)$ from theorem 2.1.6 is called the *standardization of* A

2.2 Basics of Nonstandard Analysis

Now that we know the axioms of *Internal Set Theory* we can work out some elementary facts about the Infinitesimal Calculus. First of all we have to give the exact definitions of some concepts we mentionend in the last section.

Definition 2. A real number x is

- an *infinitesimal* or is *infinitely small*, if $|x| < s$ for any standard number $s > 0$.
- *infinitely large* or *infinite* for short, if $|x| > s$ for any standard number s.
- *finite* or *limited by a standard number*, if $|x| < s$ for some standard number s.

The existence of infinitesimals and infinite numbers follows directly from **(I)**, because the infinite set $\mathbb{R}$ contains a nonstandard element x by theorem 2.1.2 and if x is infinite we have with x^{-1} an infinitesimal. Otherwise we have by theorem 2.1.5 a standard number $^\circ x$ with $x - {}^\circ x$ being infinitesimal and $(x - {}^\circ x)^{-1}$ being infinite.

In Analysis there are some common abbreviations for limits of sequences; since this abbreviations are very intuitive we should use them in our context for real numbers, too.

Notation. For real numbers x, y we write $x \approx y$, if $x - y$ is infinitesimal.

For infinitely large numbers x we write $x \approx \pm\infty$.[4]

For infinitesimals we write $x \approx 0$; for strictly positive infinitesimals we write $x \approx +0$, for strictly negative infinitesimals we write $x \approx -0$, for infinitesimals that are unequal 0 we write $x \approx \pm 0$.

If $x \approx y$ and $x > y$ we write $x \approx y + 0$; for $x < y$ we write $x \approx y - 0$.

The following often needed calculation rules are easy to see and will not be proved:

Lemma 2.2.1. Let $\varepsilon, \delta, r \in \mathbb{R}$ such that $\varepsilon, \delta \approx 0$ and r is finite.

[4]This notation can be visualized with the help of the Alexandroff extension $\mathbb{R} \cup \{\infty\}$ which is homeomorphic to a circle. The map $\iota : x \mapsto \frac{1}{x}$ for $x \in \mathbb{R}$ and $\iota(0) = \infty$ maps infinitesimals to infinite numbers and vice versa, preserving the sign of x; in this image we can say $x \approx sgn(x)\infty$, if x is infinitely large.

1. $\varepsilon + \delta \approx \varepsilon\delta \approx r\varepsilon \approx 0$
2. If $\varepsilon \neq 0$, then $\frac{r}{\varepsilon} \approx \pm\infty$

Definition 3. If $x \in \mathbb{R}$ is finite, then the uniquely defined standard number ${}^\circ x \approx x$ from theorem 2.1.5 is called the *standard part* of x.

If $x \approx \pm\infty$ we will write ${}^\circ x = \pm\infty$.

The notion of the standard part can be generalized in many ways, obviously for real vector spaces, but also in general for topological spaces. To define this general standard part we would need external sets, as described in the appendix, so we will not define it here.

Lemma 2.2.2. The following calculation rules for finite x, y are easy to see:

1. $x \pm y = ({}^\circ x + (x - {}^\circ x)) \pm ({}^\circ y + (y - {}^\circ y)) \approx {}^\circ x \pm {}^\circ y$, and because this is standard: ${}^\circ(x \pm y) = {}^\circ x \pm {}^\circ y$
2. $xy = ({}^\circ x + (x - {}^\circ x))({}^\circ y + (y - {}^\circ y)) \approx {}^\circ x {}^\circ y$, and hence ${}^\circ(xy) = {}^\circ x {}^\circ y$
3. If $y \not\approx 0$ then y^{-1} is finite and ${}^\circ(xy^{-1}) = {}^\circ x {}^\circ(y^{-1})$
4. In particular: If $y \not\approx 0$ we have ${}^\circ(y^{-1}) = ({}^\circ y)^{-1}$

2.2.1 The Infinitesimal Calculus

We mentioned before that some of the concepts of LEIBNIZ' and NEWTON's Infinitesimal Calculus were more intuitive than the descriptions by limits of sequences given later. In *Internal Set Theory* these concepts are equivalent for standard entities.

Lemma 2.2.3 (Overflow and Underflow). Let $A \subseteq \mathbb{N}$ and **either**

1. there is a standard number $n \in A$ such that for all standard $i \geq n$ we have $i \in A$ (overflow), **or**
2. there is a nonstandard number $m \in A$ such that for all nonstandard $i \leq m$ we have $i \in A$ (underflow)

then there is a nonstandard number m or a standard number n, respectivaly such that

$$\{n, n+1, \ldots, m-1, m\} \subseteq A$$

Proof. Assume that we find a standard number $n \in A$ so, that for all standard $i \geq n$ we have $i \in A$. If the set $B := \{i \in \mathbb{N} \mid i \geq n \wedge i \notin A\}$ is empty, then A contains every nonstandard number, and hence we are finished.

If $B \neq \emptyset$ we find a natural number $m := min(B) - 1$. This number m is nonstandard by assumption and by the construction of B we see

$$\{n, n+1, \ldots, m-1, m\} \subseteq A$$

The other claim is proved similarly, using $\{i \in \mathbb{N} \mid i < m \wedge i \notin A\}$. □

Lemma 2.2.4 (Infinitesimal Overflow and Underflow). Let $A \subseteq \mathbb{R}$ and **either**

1. there is a finite number $a \in A$ such that for all $x \approx a + 0$ we have $x \in A$ (infinitesimal overflow), **or**
2. there is a finite number $b \in A$ such that for all $x \approx b - 0$ we have $x \in A$ (infinitesimal underflow)

then there is a number b or a, respectivaly and $a \not\approx b$ such that

$$[a, b] \subseteq A$$

And we find a standard intervall $[c, d] \subseteq [a, b]$.

Proof. The existence of the interval $[a, b]$ is similarly to the proof of theorem 2.2.3 using the infimum of the set $\{x \in \mathbb{R} \mid x > a \wedge x \notin A\}$ or the supremum of the set $\{x \in \mathbb{R} \mid x < b \wedge x \notin A\}$, respectivaly

Since $a \not\approx b$ we have ${}^\circ a \neq {}^\circ b$ and so the standard number $\frac{{}^\circ b - {}^\circ a}{2}$ is in $[a, b]$. Then we find a standard number $r < \frac{b-a}{2}$ such that

$$[\frac{{}^\circ b - {}^\circ a}{2} - r, \frac{{}^\circ b - {}^\circ a}{2} + r] \subseteq [a, b]$$

□

Theorem 2.2.5. Let $(s_n)_{n\in\mathbb{N}}$ be a standard sequence of elements from some metric space. The limit $lim_{n\to\infty} s_n$ exists if and only if there is a standard element x such that for all $n \approx \infty$ we have $s_n \approx x$.

In this case we have $x = lim_{n\to\infty} s_n$.

Proof. Assume that $x = lim_{n\to\infty} s_n$ exists. Then we know by definition that for any $r > 0$ we find $N(r)$ such that for all $n > N(r)$ we get $|x - s_n| < r$. With **(T)** we get that $N(r)$ is standard, if r is standard; that means in particular that for all $n \approx \infty$ and standard $r > 0$ that $|x - s_n| < r$, i.e. $s_n \approx x$.

For the other direction we assume that we find x with $s_n \approx x$ for all $n \approx \infty$. That is, for all standard $r > 0$ we have for all $n \approx \infty$ that $|s_n - x| < r$. By underflow we find a standard number $N(r)$ as needed. □

Theorem 2.2.6. A standard function $f : (a, b) \to \mathbb{R}$ is continuous, if and only if for all standard x and all y

$$x \approx y \Rightarrow f(x) \approx f(y)$$

f is uniformly continuous, if and only if the statement holds for all x and y.

Proof. Let f be continuous in a standard point x, i.e. for any $r > 0$ we find $s > 0$ such that for any y with $|x - y| < s$ we have $|f(x) - f(y)| < r$. By **(T)** we see that s is standard, if f, x and r are standard, so we can conlude: If $x \approx y$ then we can choose for any standard r a standard s with $|x - y| < s$ and $|f(x) - f(y)| < r$. Since $|x - y| < s$ is true for all standard s we get $f(x) \approx f(y)$.

If f is uniformly continuous we find for any r a number s (without any use of x) so that the given argument holds even if we do not assume x to be standard.

For the other direction: Assume that for all standard x and all $y \approx x$ we have $f(x) \approx f(y)$, then we know for any standard $r > 0$ that any $s \approx +0$ fulfills the claim. We can conclude that for all standard x and standard $r > 0$ we find $s > 0$ as wanted, so by **(T)** we get the claim for any $r > 0$.

If the constraint of x being standard is not needed, then the argumentation holds even if we don't know x, hence we get the uniformly continuity. □

Directly from theorem 2.2.5 we get the following famous result.

Corollary 2.2.7. A standard function f is differentiable in any standard point $x \in \mathbb{R}$, if and only if there is a standard number $f'(x)$ such that for any infinitesimal $\varepsilon \approx \pm 0$ we have

$$f'(x) \approx \frac{f(x + \varepsilon) - f(x)}{\varepsilon}$$

Example. The function $f(x) = x^2$ is continuous: Let x be standard and $y \approx x$, i.e. $\varepsilon := y - x \approx 0$. Then:

$$y^2 = (x+\varepsilon)^2 = x^2 + 2x\varepsilon + \varepsilon^2 \approx x^2$$

But f is not uniformly continuous, because if $\varepsilon \approx \pm 0$ and $x := \varepsilon^{-1}$ we get $2x\varepsilon = 2 \not\approx 0$.

The function f is also differentiable. Let again x be any standard point and $\varepsilon \approx \pm 0$

$$\frac{f(x+\varepsilon) - f(x)}{\varepsilon} = \frac{x^2 + 2x\varepsilon + \varepsilon^2 - x^2}{\varepsilon} = 2x + \varepsilon \approx 2x = f'(x)$$

So we have $f'(x) = 2x$ for any $x \in \mathbb{R}$. $\diamond$

2.2.2 Hyperfinite Sets and Hyperfinite Maps

In ROBINSON's approach there is an original and an extended universe and a map $*$ connecting both universes. The natural numbers of the original universe (i.e. the (external) set of all standard natural numbers) are mapped into the set of hypernatural numbers. Since ROBINSON's theory can also translate statements about the original universe into statements about the extended universe the concept of a finite set can be translated into a similar concept using hypernatural numbers instead of natural numbers. This translated concept is called hyperfiniteness or $*$-finiteness (note that the $*$-map in Robinsonean Analysis is not the map $*$ from section A.1).

With the help of hyperfinite sets one can reduce many continuous concepts to finite analogs. The most famous examples hold within IST, as we will see, so we should call the corresponding sets of IST 'hyperfinite' for historical reasons, even if there are some differences. In Robinsonean Analysis e.g. every hyperfinite set is finite (if the original natural numbers are identified with their embedding into the extended universe), whereas in our context it is the other way round: Any nonstandard finite set is of course finite.

Definition 4. A finite set X is *hyperfinite*, if $|X| \approx \infty$, i.e. $|X|$ is a nonstandard natural numer.

If $\mathbf{X}$ is an infinite standard set we say X is a *hyperfinite version* of $\mathbf{X}$ if X is hyperfinite, $X \subset \mathbf{X}$ and $\mathcal{S}(X) = \mathbf{X}$, i.e. X contains the same standard elements as $\mathbf{X}$.

Note that not every nonstandard finite set is hyperfinite: If x is a nonstandard entity, then $\{x\}$ is finite and by theorem 2.1.2 nonstandard, but with $|\{x\}| = 1$ it is not hyperfinite.

Definition 5. Let $R \subset \mathbb{R}$ be any finite set and $x \in R$. If $x < max(R)$ we set the *increment* of x (in R) as $\Delta_R x := min\{y \in R | y > x\} - x$ and $\Delta_R max(R) := 0$.

If there is no confusion about the hyperfinite set R we will write Δx instead of $\Delta_R x$.

Nonstandard finite sets are finite, so they have any property finite sets have. On the other hand we can formulate some further properties within IST that need getting used to. For example: A hyperfinite set $\{0, 1, 2, \ldots, m\}$ contains any standard number, i.e. it contains 0 and for any standard number it contains its successor. Reading this the first time, one is tempted to think of infinity, but we cannot describe this kind of infinity within IST (recall that we find a hyperfinite set that contains all standard elements).

Such discrepancies can be studied with the help of the 'external point of view': Within IST a hyperfinite set is finite, but within the external theory it is even uncountable (see theorem A.1.1).

Definition 6. A map $f : X \to Y$ is called *hyperfinite*, if its domain X is hyperfinite.

We will especially be interested in real hyperfinite maps, i.e. $f : T \to \mathbb{R}$ where $T \subseteq \mathbb{R}$ is a hyperfinite version of an interval.

If we have such a function f we can find a (unique) standard partial function $\mathbf{f} : \mathcal{S}(T) \to \mathbb{R}$ by **(S)** such that for all standard values $x \in T$ we have $f(x) \approx \mathbf{f}(x)$.

$\mathbf{f}$ becomes a total function, if and only if $f(x) \not\approx \infty$ for all standard x, because the value $\mathbf{f}(x)$ is standard, so it must be finite. If this is not the case, $\mathbf{f}$ must not be defined for such x.

Definition 7. Let $f : T \to \mathbb{R}$ be a hyperfinite map with $f(x) \not\approx \infty$ for all standard $x \in T$. We call the standard function $\mathbf{f} : \mathcal{S}(T) \to \mathbb{R}$ with $\mathbf{f}(x) \approx f(x)$ for all standard $x \in T$ the *standardization* of f.

Let $g : \mathbf{T} \to \mathbb{R}$ be a standard function. We say $f : T \to \mathbb{R}$ is a *hyperfinite version* of g, if $\mathbf{T} = \mathcal{S}(T)$ and g is the standardization of f.

The existence of the standardization is discussed above, its uniqueness follows from the uniqueness of the standardization of a set.

Hyperfinite versions of g exist for any hyperfinite version T of $\mathbf{T}$, since the restriction $g|_T$ is a hyperfinite version of g.

As we have seen in section 2.1.3 we don't know anything about $\mathbf{f}(x)$ for nonstandard $x \in \mathcal{S}(T)$. This seems to be a problem for analysing global properties of $\mathbf{f}$ like continuity; we can solve this problem with the help of **(T)**.

One example is the following lemma.

Lemma 2.2.8. Let T be a hyperfinite version of $[a, b]$ and $f : T \to \mathbb{R}$ a hyperfinite map with

$$f(x) \approx f(y)$$

for all standard $x \in T$ and any $y \in T$ with $x \approx y$.

Then the standardization $\mathbf{f}(x)$ is continuous.

If this holds for all $x \in T$ (not only for standard x) then $\mathbf{f}$ is uniformly continuous.

Proof. Let $r > 0$ be standard. Then $|f(x) - f(y)| < r$ is true for all $y \in T$ with $x \approx y$. By infinitesimal overflow (theorem 2.2.4) we find some standard number $s > 0$ such that $|f(x)-f(y)| < r$ holds for all $y \in T$ with $|x-y| < s$.

Since this holds for any $y \in T$ it is especially fulfilled by any standard $y \in T$ with $|x - y| < s$. Now we can replace f by $\mathbf{f}$ and $[a, b]$ instead of T to get the statement

$$\forall^S x, y \in [a, b]\, \forall^S r \in (0, \infty)\, \exists^S s \in (0, \infty) : |x-y| < s \Rightarrow |\mathbf{f}(x)-\mathbf{f}(y)| < r$$

so we can use **(T)** to get the claim.

If x does not need to be standard the uniformly continuity is proved in a similar way. □

Theorem 2.2.9. A standard real function f is (Riemann) integrable on the standard interval $[a, b]$, if and only if the *Riemann sum*

$$\sum_{x \in H} f(x)\Delta x$$

has nearly the same value for every hyperfinite version H of $[a, b]$.

If this holds we get the Riemann integral of f by

$$\int_a^b f(x)dx = {}^\circ \sum_{x \in H} f(x)\Delta x$$

Proof. For any standard sequence of partitions with $max\Delta x \to 0$ we find a limit of the corresponding Riemann sums, if and only if f is integrable (by definition of the Riemann integral), and also if and only if for any partition with infinitesimal $max\Delta x$ we are infinitely close to the same standard value (theorem 2.2.5). □

In a similar way we will construct the Lebesgue integral in chapter 3 and the Itô integral in chapter 4.

The last example of how to use hyperfinite sets is the theorem of PEANO about the solvability of differential equations. The interesting fact about this proof is that – different to the classical proof – the theorem of ARZELÁ-ASCOLI is not needed.

Theorem 2.2.10. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be standard, continuous and bounded and $x_0 \in \mathbb{R}$. Then there is a differentiable function $\xi : \mathbb{R} \to \mathbb{R}$ with

$$\xi(0) = x_0 \text{ and for all } t \in \mathbb{R} : \xi'(t) = f(t, \xi(t))$$

Proof. Without loss of generality let f and x_0 be standard. Let R be any hyperfinite version of $\mathbb{R}$.

Set ξ inductively as $\xi(0) = x_0$ and

$$\xi(t + \Delta t) = \xi(t) + f(t, \xi(t))\Delta t$$

for $t > 0$ and similar for $t < 0$.

Then by theorem 2.2.8 the standardization $\hat{\xi}(t)$ is continuous (using that f is bounded to get that $f(t, \xi(t))\varepsilon \approx 0$ for any $\varepsilon \approx 0$) and for any standard $s, t \in \mathbb{R}$ we have

$$\hat{\xi}(t) - \hat{\xi}(s) \approx \sum_{r \in R \cap [s,t]} f(r, \hat{\xi}(r))\Delta r$$

Since f is continuous the value $\hat{\xi}(t) - \hat{\xi}(s)$ does not depend on the choice of R, so that we get from theorem 2.2.9 that

$$\hat{\xi}(t) - \hat{\xi}(s) \approx \int_s^t f(r, \hat{\xi}(r))dr$$

and that means that $\hat{\xi}$ is differentiable with

$$\hat{\xi}'(t) = f(r, \hat{\xi}(t))$$

□

Chapter 3

Measure and Probability Theory

In this chapter we want to get a short overview over the basics of probability theory using nonstandard methods to model continuous phenomena. With the help of the methods introduced in section 2.2.2 we will reduce measure and probability theory to hyperfinite probability spaces.

In chapter 4 we will take a further step by reducing stochastic processes to hyperfinite maps; we will then be able to get the Itô integral for standardizations similar to the Riemann integral in theorem 2.2.9.

3.1 Measure and Probability Spaces

Definition 8. A *(finite) measure space* is a pair $\langle \Omega, \mu \rangle$ where Ω is a finite (nonempty) set and $\mu : \mathcal{P}(\Omega) \to [0, \infty)$ with $\mu(\Omega) > 0$ is an additive function, i.e.

$$\mu(A \cup B) = \mu(A) + \mu(B)$$

for any disjoint $A, B \subseteq \Omega$. The map μ is called a *measure* on Ω.

A measure with $\mu(\Omega) = 1$ is a *probability distribution*; then we call $\langle \Omega, \mu \rangle$ a *probability space* and the subsets of Ω *events*.

Notation. Since most of our measure and probability spaces will be finite (or standardizations of finite ones, see section 3.1.1) we will often write only 'measure space' instead of 'finite measure space'.

Theorem 3.1.1. Any measure μ on a finite set Ω is given by $\mu(A) = \sum_{\omega \in A} m(\omega)$ where $m : \Omega \to [0, \infty]$ is a map and there is at least one $\omega_0 \in \Omega$ with $m(\omega_0) > 0$.

Such a map m is called a *weight function*.

Proof. The statement is clear by setting $m(\omega) = \mu(\omega)$. □

This theorem is useful to define finite measure spaces by a weight function. As a similar concept we will introduce the density functions for infinite measure spaces, but a density function does not need to exist.

Definition 9. Let $\langle \Omega, P \rangle$ be a probability space and $C \subseteq \Omega$ with $P(C) > 0$. The *conditional probability* is

$$P_C(A) = \frac{P(A \cap C)}{P(C)}$$

for any $A \subseteq \Omega$.

We will often write $P(A \mid C) := P_C(A)$.

A natural question about definition 9 is the following: Given some $C \subseteq \Omega$ with $P(C) > 0$ is there a nontrivial possibility to find some set $A \subseteq \Omega$ with $P(A) = P(A \mid C)$?

Example. Let $\langle \Omega_i, P_i \rangle_{0 \leq i \leq n}$ be probability spaces ($n \in \mathbb{N}$). Then we can define the *product space* $\langle \Omega, P \rangle$ where

$$\Omega = \Omega_0 \times \cdots \times \Omega_n$$

and

$$P(\omega) = \prod_{0 \leq i \leq n} P_i(\omega_i)$$

for any $\omega \subset \Omega$.

Now let $n = 2$ and $A_i \subseteq \Omega_i$ for $i \in \{1, 2\}$. Then

$$\begin{aligned} P((A_1 \times \Omega_2) \cap (\Omega_1 \times A_2)) &= P(A_1 \times A_2) \\ &= P_1(A_1) P_2(A_2) = P(A_1 \times \Omega_2) P(\Omega_1 \times A_2) \end{aligned}$$

◇

Definition 10. Events $A, B \subseteq \Omega$ with $P(A)P(B) = P(A \cap B)$ are called *(stochastically) independent*.

If $A, B \subseteq \Omega$ are independent and $P(B) > 0$ we see that

$$P(A) = \frac{P(A \cap B)}{P(B)} = P(A \mid B)$$

3.1.1 Standardized Measures

We mentioned that we are interested in continuous phenomena, but defined only finite measure spaces, yet. The tool to model continuous phenomena will be the standardization, but before we can formulate standardized measures in definition 13 we have to prepare some notions.

Remark. This section is a continuation and perhaps an improvement of Nelson's idea of using nonstandard finite concepts to describe infinite standard phenomena.

Our concept of standardization is thought to be an alternative to the theorem of Loeb. With the help of this concept we can construct measure spaces from hyperfinite measure spaces in a similar way as we can do with the help of Loeb spaces. A closer comparison can be found in the appendix (section A.2).

Definition 11. A *standardizable measure space* is a pair $\langle \mathbf{\Omega}, \mu \rangle$ where $\mathbf{\Omega}$ is an infinite standard set and μ a map of measures $\mu_\Omega : \mathcal{P}(\Omega) \to [0, \infty)$ for any hyperfinite version Ω of $\mathbf{\Omega}$ such that $\mu_\Omega(\Omega)$ is nearly equal for all Ω.

The notion 'standardizable' will be justified in definition 13.

Definition 12. Let Ω be any set. A set $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ is a *σ-algebra* if

1. $\Omega \in \mathcal{A}$
2. For any $A \in \mathcal{A}$ we have $\Omega \setminus A \in \mathcal{A}$
3. For any countable sequence $A_1, A_2, \cdots \in \mathcal{A}$ we have $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$

A map $\mu : \mathcal{A} \to [0, \infty) \cup \{\infty\}$ is *σ-additive* if for any countable sequence $A_1, A_2, \cdots \in \mathcal{A}$ of pairwise disjoint sets

$$\mu(\bigcup_{i \in \mathbb{N}} A_i) = \sum_{i \in \mathbb{N}} \mu(A_i)$$

A *(general) measure space* is a triple $\langle \Omega, \mathcal{A}, \mu \rangle$ where Ω is a set, $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ a σ-algebra and $\mu : \mathcal{A} \to [0, \infty]$ a σ-additive map with $\mu(\Omega) > 0$.

The general definition of a measure space is sometimes useful, but as mentioned before most of our measure spaces will be finite, so that we will not everytime write explicitly that it is finite. The only important class of infinite measure spaces in our context is the class of standardized measure spaces.

Theorem 3.1.2. Let $\langle \mathbf{\Omega}, \mu \rangle$ be a standardizable measure space. Then there is a unique standard σ-algebra $\mathcal{A} \subseteq \mathcal{P}(\mathbf{\Omega})$ and a unique σ-additive standard map $\hat{\mu} : \mathcal{A} \to [0, \infty]$ satisfying for any standard $A \in \mathcal{A}$

$$\hat{\mu}(A) \approx \mu_\Omega(A \cap \Omega) \text{ and } \hat{\mu}(\Omega \setminus A) \approx \mu_\Omega(\Omega \setminus A)$$

for all hyperfinite versions Ω of $\mathbf{\Omega}$.

Proof. Let μ be the standard function satisfying the condition and $\mathcal{A} := dom(\mu)$ (using **(S)**). We have to show that $\mathcal{A}$ is a σ-algebra and $\hat{\mu}$ a measure. Since Ω is measurable by definition and if A is measurable than so is $\Omega \setminus A$. So the only thing to show is that μ is σ-additive, since then the limit of countable sequences have to be in $\mathcal{A}$ by definition.

Let $(A_i)_{i \in \mathbb{N}}$ be a standard sequence of disjoint sets of $\mathcal{A}$ and $A = \bigcup_{i \in \mathbb{N}} A_i$. Since $(\hat{\mu}(\bigcup_{j<i} A_j))_{i \in \mathbb{N}}$ is increasing and limited by $\hat{\mu}(\mathbf{\Omega})$ it has a (standard) limit l.

For any standard number $i \in \mathbb{N}$ we have

$$l - \hat{\mu}(\bigcup_{i \leq j} A_j) = \hat{\mu}(\bigcup_{j<i} A_j) = \sum_{j<i} \hat{\mu}(A_j)$$

so we see with $\hat{\mu}(\bigcup_{i \leq j} A_j) \to 0$ for $i \to \infty$ that

$$\sum_{i \in \mathbb{N}} \hat{\mu}(A_i) = l = \hat{\mu}(\bigcup_{i \in \mathbb{N}} A_i)$$

So we see that A has a measure $\hat{\mu}(A)$, hence $A = \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$. □

Definition 13. Let $\langle \mathbf{\Omega}, \mu \rangle$ be a standardizable measure space. The standard measure space $\langle \mathbf{\Omega}, \mathcal{A}, \hat{\mu} \rangle$ from theorem 3.1.2 is the *standardized measure space* corresponding to $\langle \mathbf{\Omega}, \mu \rangle$.

Elements of $\mathcal{A}$ are called *($\hat{\mu}$-) measurable sets*.

Theorem 3.1.3. Any standardized measure space $\langle \mathbf{\Omega}, \mathcal{A}, \hat{\mu} \rangle$ is complete, i.e. if we find for $A \subseteq \mathbf{\Omega}$ some $B \in \mathcal{A}$ with $A \subseteq B$ and $\hat{\mu}(B) = 0$, then $A \in \mathcal{A}$.

Proof. Let Ω be any hyperfinite version of $\mathbf{\Omega}$. Since μ_Ω is a measure we have $\mu_\Omega(A) \leq \mu_\Omega(B)$ and since $\mu_\Omega(B) \approx 0$ we have $\mu_\Omega(A) \approx 0$.

This means that $A \in \mathcal{A}$ and $\hat{\mu}(A) = 0$. □

Example. For any hyperfinite version R of $\mathbb{R}$ set $\mu_R(A) = \frac{|A \cap R|}{|R|}$ for any $A \subseteq \mathbb{R}$. This measure is called the counting measure on R.

Let R and L be hyperfinite versions of $\mathbb{R}$ with $|R|^2 \leq |L|$. For every standard interval $I \subset \mathbb{R}$ with $\emptyset \neq I \neq \mathbb{R}$ we have $\mu_{R \cup (L \cap A)}(A) \approx 1$, but $\mu_{R \cup (L \setminus A)}(A) \approx 0$.

We can conclude that $\mathcal{A}$ contains no intervals (besides $\mathbb{R}$ itself). This measure space is obviously unsuitable for a probybility theory of subsets of $\mathbb{R}$. $\diamond$

In Robinsonean analysis (e.g. [Cu83]) the Lebesgue measure is constructed by the counting measure on special hyperfinite sets, so-called Lebesgue samples. In this approach we want to avoid the use of special well-behaved sets, so we need a more general construction of the Lebesgue measure.

Remark. The following example seems to be a new construction of the Lebesgue measure; I could not find a similar construction in the literature.

Example. For any hyperfinite version R of $\mathbb{R}$ let $\lambda_R(x) = \Delta x$ for any $x \in R$. The standardized measure $\hat{\lambda}$ is the *Lebesgue measure*.

1. Let $[a, b] \subseteq \mathbb{R}$ be a standard interval, then
$$\hat{\lambda}([a, b]) \approx \sum_{x \in R \cap [a,b]} \Delta x \approx b - a$$
independent of the choice of R. Since $\mathcal{A}$ is a σ-algebra that contains any real interval and hence any *Borel set*, i.e. a set in the σ-algebra generated by the intervals.
2. Let $t \in \mathbb{R}$ and $A \subseteq \mathbb{R}$ both be standard. For any hyperfinite version R of $\mathbb{R}$ the translated set $R_t := \{r + t \mid r \in R\}$ is another hyperfinite version of $\mathbb{R}$, because $r \in R$ is standard if and only if $r + t$ is standard. So we have
$$\hat{\lambda}(A) \approx \sum_{x \in A \cap R} \Delta x = \sum_{x \in A_t \cap R_t} \Delta x$$
and since this holds for any t we see that $A_t \in \mathcal{A}$ and $\hat{\lambda}(A_t) = \hat{\lambda}(A)$.
3. Let $V \subset [0, 1]$ be a standard Vitali set[1] and for any $q \in [-1, 1] \cap \mathbb{Q}$ let V_q be the translated set as in 2.

[1] A Vitali sets is constructed as follows: Let $x \sim y :\Leftrightarrow x - y \in \mathbb{Q}$ and V a system of representants of $\mathbb{R}/\sim$, i.e. V contains one and only one element of any equivalence class of $\sim$. The existence of a Vitali set can be formulated and proved within ZFC, so we find even a standard Vitali set by **(T)**.

We have $\bigcup_{q\in[-1,1]\cap\mathbb{Q}} V_q \subseteq [-1,2]$ and hence $\sum_{q\in[-1,1]\cap\mathbb{Q}} \hat{\lambda}(V) \leq 3$ because of the invariance under translation; this is only possible if $\hat{\lambda}(V) = 0$.

Let now $x \in [0,1]$ and let $x_V \in V$ be the representant of the equivalence class of x. Then $x - x_V \in \mathbb{Q}$ and $-1 \leq x - x_V \leq 1$, so $x \in V_{x-x_V}$, that means we have $[0,1] \subseteq \bigcup_{q\in[-1,1]\cap\mathbb{Q}} V_q$ and hence $1 \leq \sum_{k\in\mathbb{N}} \hat{\lambda}(V)$.

This is a contradiction, so V must not be measurable.

We see that $\hat{\lambda}$ is invariant under translations, and since $\hat{\lambda}([0,1]) = 1$ we have indeed the classical Lebesgue measure. ◇

Definition 14. Let $\langle \mathbf{\Omega}, \mathcal{A}, \hat{\mu}\rangle$, $\langle \mathbf{F}, \mathcal{F}, \hat{\nu}\rangle$ be standardized measure spaces and $\mathbf{F}$ a field.

A function $f : \mathbf{\Omega} \to \mathbf{F}$ is *measurable*, if for any set $B \in \mathcal{F}$ the pre-image is measurable, i.e.

$$f^{-1}(B) = \{\omega \in \Omega \mid f(\omega) \in B\} \in \mathcal{A}$$

If $\hat{\mu}$ is a probability measure we call the measurable functions *random variables*.

Notation. If we define some measurable function $f : \mathbf{\Omega} \to \mathbb{R}$ we use the Lebesgue measure on $\mathbb{R}$, unless we define another measure.

Theorem 3.1.4. Let $\langle \mathbf{\Omega}, \mathcal{A}, \hat{\mu}\rangle$ be a standardized measure space. The set $\mathcal{L}^0 \subseteq \mathbb{R}^\Omega$ of measurable functions is an $\mathbb{R}$-vector space.

Proof. Since the set of all functions $\mathbf{\Omega} \to \mathbb{R}$ is an $\mathbb{R}$-vector space and the skalar multiplication is clear the only problem is the sum of two measurable functions.

Let $f, g \in \mathcal{L}^0$ be standard and R any hyperfinite version of $\mathbb{R}$. Then for any standard measurable $A \subseteq \mathbb{R}$

$$(f+g)^{-1}(A) = \bigcup_{x\in f(\mathbf{\Omega})} g^{-1}(\{a - x \mid a \in A\})$$

and as a union of measurable sets this is measurable. □

Notation. We use capital letters $X, Y, \ldots$ for random variables and will also use some usual abbreviations:

$$\begin{aligned} P(X = x) &:= P(\{\omega \in \Omega \mid X(\omega) = x\}) \\ P(X \leq x) &:= P(\{\omega \in \Omega \mid X(\omega) \leq x\}) \\ P(X \in A) &:= P(\{\omega \in \Omega \mid X(\omega) \in A\}) \\ P(X \approx x) &:= P(\{\omega \in \Omega \mid {}^\circ X = {}^\circ x\}) \\ &\vdots \end{aligned}$$

Definition 15. Random variables X, Y are called *(stochastically) independent*, if for any measurable $A \subseteq \mathbb{R}$ the events $\{X \in A\}$ and $\{Y \in A\}$ are independent, i.e. $P(X = x)P(Y = y) = P(X = x \wedge Y = y)$.

3.1.2 The Lebesgue Integral

As an example for the use of this hyperfinite measure theory we can construct the Lebesgue integral for standardized measure spaces. We will see that the integral is nearly equal to a sum over a hyperfinite version of the integration set.

Let $\langle \mathbf{\Omega}, \mu \rangle$ be any standardizable measure space and $\langle \mathbb{R}, \lambda \rangle$ a standardizable measure space on $\mathbb{R}$.

Definition 16. The *Lebesgue sum* of a function $f : \mathbf{\Omega} \to \mathbb{R}$ on a measurable set $A \subseteq \mathbf{\Omega}$ and with respect to a hyperfinite version Ω of $\mathbf{\Omega}$ is defined as

$$\sum_{x \in A \cap \Omega} f(x)\mu_\Omega(x)$$

The *(Lebesgue) integral* of f on a measurable set $A \subseteq \mathbf{\Omega}$ and with respect to the measure $\hat{\mu}$ is the standard operator $\int$ satisfying

$$\int_A f\hat{\mu} \approx \sum_{x \in A \cap \Omega} f(x)\mu_\Omega(x)$$

for all standard measurable sets A and all standard functions f (if the Lebesgue sums have nearly the same value for any Ω).

The set of *quasi-integrable functions* $\mathcal{L}^1_q$ is the set of all functions f such that there are positive measurable functions g, h with $f = g - h$. The set of *integrable functions* $\mathcal{L}^1$ is the subset of $\mathcal{L}^1_q$ of all functions f with $\int_{\mathbf{\Omega}} |f|\hat{\mu} < \infty$.

Notation. We replace the commonly used notation $\int_A f d\hat{\mu}$ by $\int_A f\hat{\mu}$, because contrary to the Riemann integral we have no 'differential' dx as integrator, but a measure. It is more intuitive to think about the point-wise product $f\hat{\mu}(x) = f(x)\hat{\mu}(x)$.

As we will see in corollary 3.1.6 there is another reason for this notation: If $\hat{\lambda}$ is the Lebesgue measure we have $\int_A f\hat{\lambda} = \int_A f(x)dx$, i.e. the Lebesgue integral equals the Riemann integral. With our notation this is more intuitive, since we can think of the differential dx as the standardization of the measure given by the weights Δx, so that we see the equation directly by (the **undefined** formula) $\hat{\lambda}(x) = dx$.

Theorem 3.1.5. For all $f \in \mathcal{L}_q^1$ the Lebesgue integral is defined.

For $f \in \mathcal{L}^1$ the intregral is not $\pm\infty$.

Proof. We prove this claim in three steps according to the classical construction of the Lebesgue integral. Let Ω be any hyperfinite version of $\mathbf{\Omega}$.

1. Assume $f \in \mathcal{L}_q^1$ is a positive (i.e. $f(x) \geq 0$ for all x) *simple function* (i.e. we find some pairwise disjoint sets $A_1, \ldots, A_n \subseteq A$ with $\bigcup_{1\leq i\leq n} A_i = A$ and $f(A_i) = \{a_i\}$ for all i). Then the Lebesgue sum equals
$$\sum_{x\in A\cap\Omega} f(x)\mu_\Omega(x) = \sum_{1\leq i\leq n} a_i\mu_\Omega(A_i)$$
So the integral exists (but may be ∞).

2. Assume that $f \in \mathcal{L}_q^1$ is positive and measurable. Then we find some sequence f_i of simple functions with $f_i(x) \leq f_{i+1}(x) \leq f(x)$ for any i and x and $lim_{i\to\infty} f_i = f$ pointwise, and then
$$\begin{aligned}\sum_{x\in A\cap\Omega} f(x)\mu_\Omega(x) &= \sum_{x\in A\cap\Omega} lim_{i\to\infty} f_i(x)\mu_\Omega(x)\\ &= lim_{i\to\infty} \sum_{x\in A\cap\Omega} f_i(x)\mu_\Omega(x)\\ &\approx lim_{i\to\infty} \int_A f_i\hat{\mu}\end{aligned}$$
Since the sequence f_i is pointwise monotonic increasing we find a limit (but it may be ∞).

3. Now let $f = g - h \in \mathcal{L}_q^1$ with g, h being positive, measurable functions. The Lebesgue sum
$$\sum_{x\in A\cap\Omega} f(x)\mu_\Omega(x) = \sum_{x\in A\cap\Omega} g(x)\mu_\Omega(x) - \sum_{x\in A\cap\Omega} h(X)\mu_\Omega(x)$$

is nearly equal for all Ω, since g and h are integrable. Note that we can set $\int_A f\hat{\mu}$ even if $\int_A g\hat{\mu} = \int_A h\hat{\mu} = \infty$, since both Lebesgue sums are (infinite) real numbers and so their difference is well defined.

Let $f = g - h \in \mathcal{L}^1$ and g, h positive, measurable functions. With

$$\int_{\mathbf{\Omega}} |f|\hat{\mu} = \int_{\mathbf{\Omega}} g + h\hat{\mu} = \int_{\mathbf{\Omega}} g\hat{\mu} + \int_{\mathbf{\Omega}} h\hat{\mu} < \infty$$

we see directly $\int_{\mathbf{\Omega}} f\hat{\mu} \in \mathbb{R}$. □

Remark. In the proof we used

$$\sum_{x\in A\cap\Omega} lim_{i\to\infty} f_i(x)\mu_\Omega(x) = lim_{i\to\infty} \sum_{x\in A\cap\Omega} f_i(x)\mu_\Omega(x)$$

This is not a spectacular transformation because the sum has only finitely many summands, but if we look at the standardization we get the equation

$$\int_A lim_{i\to\infty} f_i\hat{\mu} = lim_{i\to\infty} \int_A f_i\hat{\mu}$$

This is the monotone convergence theorem – with this construction an absolutely natural and easy to see property of the Lebesgue integral.

Corollary 3.1.6. Let $\hat{\lambda}$ be the Lebesgue measure on $\mathbb{R}$, $A \in \mathcal{A}$ standard and $f : A \to \mathbb{R}$ a standard integrable function, then

$$\int_A f\hat{\lambda} = {}^\circ \sum_{x\in A\cap R} f(x)\Delta x = \int_A f(x)dx$$

for any hyperfinite version R of $\mathbb{R}$.

Theorem 3.1.7. $\mathcal{L}^1$ is a vector subspace of $\mathcal{L}^0$ and the Lebesgue integral is a linear form on it.

Proof. For $f, g \in \mathcal{L}^1$ we find positive, measurable functions f_1, f_2, g_1, g_2 with $f = f_1 - f_2$ and $g = g_1 - g_2$, resp. Then

$$f + g = (f_1 - f_2) + (g_1 - g_2) = (f_1 + g_1) - (f_2 + g_2)$$

The functions $f_1 + g_1$ and $f_2 + g_2$ are positive and measurable, because the measurable sets are a vector space (lemma 3.1.4), so the sum of integrable functions is again integrable.

If $f = f_1 - f_2 \in \mathcal{L}^1$ and $a \in R$ we have obviously $af = af_1 - af_2 \in \mathcal{L}^1$ independent of the sign of a.

The linearity of the integral follows directly from its definition. □

From the Lebesgue integral together with theorem 3.1.1 we get the usual description of standardized probability distributions by a density function.

Definition 17. A standardized probability distribution $\mathbf{P}$ on $\mathbb{R}$ has a *density function*, if there is a function $f : \mathbb{R} \to [0, 1]$ with $\mathbf{P}(A) = \int_A f(x)dx$ for any measurable $A \subseteq \mathbf{\Omega}$.

Directly from the definition of the integral we get the following:

Corollary 3.1.8. If a probability distribution $\mathbf{P}$ has a density function f then we have for any hyperfinite version R of $\mathbb{R}$

$$\mathbf{P}(A) \approx \sum_{x \in R \cap A} f(x)\Delta x$$

In other words: P_R is described by the weight function $p_R : R \to [0, 1]$ given by $p_R(x) = f(x)\Delta x$.

Note that we can describe any probability distribution by such weight functions on the hyperfinite versions of R of $\mathbb{R}$. The interesting fact of the corollary is that if we have a density function then all the weight functions have the same easy to handle form $f(x)\Delta x$ independent of R.

Definition 18. Let $\langle \mathbf{\Omega}, \mathcal{A}, \hat{\mu} \rangle$ be a standardized measure space. For any $n \in \mathbb{N}$ let $\mathcal{L}^n \subseteq \mathcal{L}^{n-1}$ be the vector space of functions f with

$$\int_{\mathbf{\Omega}} |f|^n \hat{\mu} \neq \infty$$

We call $\mathcal{L}^n$ the set of n-times integrable functions.

3.1.3 Mean and Variance

Definition 19. Let $\langle \Omega, P \rangle$ be a finite probability space and $X \in \mathcal{L}^1$ a random variable. The *mean* or *expected value* of X is defined as

$$E(X) := \sum_{\omega \in \Omega} X(\omega)P(\omega)$$

If $\Omega \subset \mathbb{R}$ the *mean* of the probability distribution P is $E(P) = E(id)$.

Theorem 3.1.9. Let $\langle \mathbf{\Omega}, \mathcal{A}, \mathbf{P} \rangle$ be a standardized probability space and $X \in \mathcal{L}^1$ be a standard random variable. Then we have for all hyperfinite versions Ω of $\mathbf{\Omega}$

$$E_\Omega(X) \approx \int_{\mathbf{\Omega}} X\mathbf{P}$$

Proof. Directly from the definition of the integral. □

Definition 20. Let $\langle \mathbf{\Omega}, \mathcal{A}, \mathbf{P} \rangle$ be a standardized probability space and $X \in \mathcal{L}^1$ a standard random variable. The *mean* or *expected value* of X is

$$E(X) := {}^\circ E_\Omega(X)$$

for any hyperfinite version Ω of $\mathbf{\Omega}$.

Theorem 3.1.10. Let $\langle \mathbf{\Omega}, P \rangle$ be a standardizable probability space and Ω a hyperfinite version of $\mathbf{\Omega}$.

1. The expected value is a linear map, i.e. for any $r \in \mathbb{R}$ and any random variables $X, Y : \Omega \to \mathbb{R}$

$$E(rX + Y) = rE(X) + E(Y)$$

2. If X, Y are independent random variables we have

$$E(XY) = E(X)E(Y)$$

Proof. 1. Follows directly from the linearity of the sum.

2. This can be calculated without difficulties:

$$\begin{aligned} E(XY) &= \sum_{\langle x,y\rangle \in X(\Omega)\times Y(\Omega)} xyP(X = x \wedge Y = y) \\ &= \sum_{x\in X(\Omega)} \sum_{y\in Y(\Omega)} xP(X = x)yP(Y = y) \\ &= \sum_{x\in X(\Omega)} xE(Y)P(X = x) = E(X)E(Y) \end{aligned}$$

□

Example. 1. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ with the counting measure P. Then

$$E(P) = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = \frac{7}{2}$$

2. Let $N := \{0, 1, 2 \dots n-1\} \subset \mathbb{N}$ and $\beta_p : N \to [0,1]$ be the *binomial distribution* defined by the weight function $\beta_p(i) = \binom{n}{i} p^i (1-p)^{n-i}$. This is a probability distribution, because

$$\sum_{i \in N} \beta_p(i) = \sum_{i \in N} \binom{n}{i} p^i (1-p)^{n-i} = (p + 1 - p)^n = 1$$

The mean of the binomial distribution is

$$\begin{aligned} E(\beta_p) &= \sum_{i \in N} \binom{n}{i} i p^i (1-p)^{n-i} \\ &= 0 + np \sum_{i \in N \setminus \{0\}} \binom{n-1}{i-1} p^{i-1} (1-p)^{n-1-i+1} \\ &= np \sum_{i \in N \setminus \{n-1\}} \beta_{n-1}(i) = np \end{aligned}$$

◇

Definition 21. For $X \in \mathcal{L}^2$ the *variance* is

$$Var(X) = E((X - E(X))^2)$$

If $\Omega \subset \mathbb{R}$ the *variance* of a probability distribution P on Ω is $Var(P) = Var(id)$.

Lemma 3.1.11. $Var(X) = E(X^2) - E(X)^2$

Proof. $E((X - E(X))^2) = E(X^2 - 2XE(X) + E(X)^2)$
$= E(X^2) - 2E(X)E(X) + E(X)^2 = E(X^2) - E(X^2)$ □

Example. We calculate the variances of the distributions studied in the last example.

1. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ with the counting measure P. Then

$$Var(P) = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) - \frac{49}{4} = \frac{35}{4}$$

2. Let $N := \{0, 1, 2 \dots n-1\} \subset \mathbb{N}$ and β_p the binomaial distribution. Then the variance is

$$\begin{aligned} Var(\beta_p) &= \sum_{i \in N} \binom{n}{i} i^2 p^i (1-p)^{n-i} - n^2 p^2 \\ &= np(\sum_{i \in N \setminus \{0\}} \binom{n-1}{i-1} (i-1+1) p^{i-1} (1-p)^{n-1-i+1} - np) \\ &= np(E(\beta'_p) + 1 - np) = np((n-1)p + 1 - np) \\ &= np(1-p) \end{aligned}$$

◇

Besides mean and variance we will need the conditional mean. To formulate this concept we need to restrict our random variables to smaller σ-algebras.

Example. Let $\Omega = \{1, 2, 3, 4, 5, 6\}$ with the counting measure. The set

$$\mathcal{F} := \{\emptyset, \{1, 2, 3\}, \{4, 5, 6\}, \Omega\} \subset \mathcal{P}(\Omega)$$

is a σ-algebra.

An $\mathcal{F}$-measurable random variable is a function $X : \Omega \to \mathbb{R}$ that is measurable with respect to $\mathcal{F}$. Since $\{X(1)\}$ is a Lebesgue nullset it is measurable and we have $X^{-1}(\{X(1)\}) \in \mathcal{F} \setminus \{\emptyset\}$ and so $\{1, 2, 3\} \subseteq X^{-1}(\{X(1)\})$, i.e.

$$X(1) = X(2) = X(3)$$

Analogue we see that

$$X(4) = X(5) = X(6)$$

Obviously every $\mathcal{F}$-measurable function is a random variable. So the restriction to a sub-σ-algebra is a stronger condition. ◇

Definition 22. Let $\langle \Omega, P \rangle$ be a finite probability space, $\mathcal{F} \subseteq \mathcal{P}(\Omega)$ a sub-σ-algebra of $\mathcal{A}$ and $X : \Omega \to \mathbb{R}$ a random variable.

An $\mathcal{F}$-measurable function $E^{\mathcal{F}}(X) : \Omega \to \mathbb{R}$ with

$$\sum_{\omega \in F} E^{\mathcal{F}}(X)(\omega) P(\omega) = \sum_{\omega \in F} X(\omega) P(\omega)$$

for all $F \in \mathcal{F}$ is called a *conditional mean* of X.

Theorem 3.1.12. Let X be a random variable and U, V conditional means of X with respect to the sub-σ-algebra $\mathcal{F}$. Then $U(\omega) = V(\omega)$ almost sure.

Proof. Since $\mathcal{P}(\Omega)$ is finite we find $F \in \mathcal{F}$ that have no subsets in $\mathcal{F} \setminus \{\emptyset\}$. As we have seen in the example above U and V are constant on F, and if $P(F) > 0$ we see that $U(\omega) = V(\omega)$ for any $\omega \in F$. □

Theorem 3.1.13. If $\langle \mathbf{\Omega}, \mathcal{A}, \mathbf{P} \rangle$ is a standardized measure space, $\mathcal{F} \subseteq \mathcal{A}$ and X a standard random variable. Then the conditional means of X are nearly equal for all hyperfinite versions Ω of $\mathbf{\Omega}$. We define the conditional mean of X as

$$E^{\mathcal{F}}(X) := {}^{\circ}E^{\mathcal{F}}_{\Omega}(X)$$

Proof. For each $F \in \mathcal{F} \subseteq \mathcal{A}$ we have

$$\sum_{\omega \in F} X(\omega)P(\omega) \approx \int_F XP$$

□

3.2 Large numbers

Part of the stochastic research is the question of how to approximate finite distributions with a large population by an infinite distribution. The result is lots of approximation theorems like the theorem of MOIVRE-LAPLACE, POISSON's approximation theorem, BERNOULLI's law of large numbers, and so on.

In the nonstandard context we can formulate these approximation theorems to describe the connection between standardizable hyperfinite measures and their standardizations. Then we can easily 'switch' between these distributions: If it is comfortable we use some hyperfinite version of a measure space (e.g. some proofs become much easier if we have a finite probability space), but if it is more comfortable to use the infinite version we can do this (e.g. by (approximately) calculating certain measures).

For most applied mathematicians, physicists and economists this seems to be a usual interpretation of approximation statements anyway: If some numbers are very small or very large (compared to others) they are 'nearly 0' or 'nearly infinite'. Though this assumption is still somewhat creepy IST is at least a mathematical environment in which one can do clean work with such an assumption.

We want to formulate POISSON's approximation theorem as an example of such a hyperfinite approximation theorem. Then we will formulate and prove the hyperfinite version of the central limit theorem and a version of the law of large

numbers (NELSON's weak law of large numbers). Note that the original theorems claim similar statements using limits of variables instead of infinite numbers. By theorem 2.2.5 we see that these nonstandard versions are equivalent to the classical versions for standard sequences.

Theorem 3.2.1 (POISSON**'s Approximation Theorem).** Let $\lambda \in \mathbb{R}$, $\lambda > 0$ be standard, N any hyperfinite version of $\mathbb{N}$ and $P_N : N \to [0,1]$ be the *binomial distribution* given by $\beta_p(i) = \binom{|N|}{i} p^i (1-p)^{|N|-i}$ with $p = \frac{\lambda}{|N|}$. Then the standardization of $\langle \mathbb{N}, P \rangle$ exists, any event $A \subseteq \mathbb{N}$ is measurable and $E(\mathbf{P}) = Var(\mathbf{P}) = \lambda$.

This standardized distribution is called the *Poisson distribution*.

Proof. Let $n = |N|$ and $i \in N$ be standard.

$$\begin{aligned} \beta_{\frac{\lambda}{n}}(i) &= \binom{n}{i} (\frac{\lambda}{n})^i (1 - \frac{\lambda}{n})^{n-i} \\ &= \frac{n(n-1)\dots(n-k+1)}{n^i} \frac{\lambda^i}{k!} (1 - \frac{\lambda}{n})^{n-i} \\ &\approx \frac{\lambda^i}{k!} e^{-\lambda} \end{aligned}$$

We see that elementary events $\{i\}$ are measurable, so every subset of $\mathbb{N}$ is measurable.

With the help of mean and variance of the binomial distribution (see examples in section 3.1.3)

$$\begin{aligned} E(\mathbf{P}) &\approx E(\beta_{\frac{\lambda}{|N|}}) = |N| \frac{\lambda}{|N|} = \lambda \\ Var(\mathbf{P}) &\approx Var(\beta_{\frac{\lambda}{|N|}}) = |N| (\frac{\lambda}{|N|})(1 - \frac{\lambda}{|N|}) \approx \lambda \end{aligned}$$

□

3.2.1 Central Limit Theorem

Definition 23. Let $P, P_1, P_2, \dots : \Omega \to \mathbb{C}$ be probability distributions. We say the sequence $P_1, P_2, \dots$ *converges weakly* to P, if for all bounded, continuous functions $X : \Omega \to \mathbb{R}$

$$E_{P_i}(X) \to E_P(X) \text{ for } i \to \infty$$

A sequence of random variables $X_1, X_2, \dots$ *converges in distribution*, if their probability distributions converge weakly.

By lemma 2.2.5 we get that a standard sequence $P_1, P_2, \ldots$ converges weakly to P, if and only if

$$E_{P_N}(X) \approx E_P(X)$$

for any bounded, continuous function X and any $N \approx \infty$.

Definition 24. Let $\Omega \subseteq \mathbb{R}$ and $X : \Omega \to \mathbb{R}$ be a random variable. The function

$$\chi_X : t \mapsto E(e^{itX})$$

is the *characteristic function* of X.

Theorem 3.2.2 (LEVY's continuity theorem, special case). A sequence of random variables $(X_i)_{i\in\mathbb{N}}$ over a finite probability space converges in distribution to a random variable X, if and only if the sequence of their characteristic functions converges pointwise to the characteristic function χ_X.

In general LEVY's continuity theorem holds for any general probability spaces (not only finite ones), but for the proof of the hyperfinite version of the central limit theorem given here this special case is enough.

Proof. Let $P, P_1, P_2, \ldots$ be the probability distributions of $X, X_1, X_2, \ldots$, resp. The characteristic function of X_i is

$$\chi_{X_i}(t) = E_{P_i}(e^{itX_i}) = \sum_{\omega\in\Omega} e^{itX_i(\omega)} P_i(\omega)$$

Without loss of generality let $(X_i)_{i\in\mathbb{N}}$ be standard.

If $X_i \to X$ in distribution we have immediately that for any standard t and $N \approx \infty$

$$\sum_{\omega\in\Omega} e^{itX_N(\omega)} P_N(\omega) \approx \sum_{\omega\in\Omega} e^{itX(\omega)} P(\omega)$$

If the sequence of the characteristic functions converge this equation holds for any t, and then we have

$$\sum_{x\in X(\Omega)} e^{itx} P(X = x) \approx \sum_{x\in X_i(\Omega)} e^{itx} P(X_i = x)$$

Now we have that $X_i \to X$ in distribution, because we can chose t in a way that the values of e^{itx} are mutually different for $x \in X(\Omega)$. □

Theorem 3.2.3 (Central Limit Theorem). Let $X_1, X_2, \ldots$ be independent, identically distributed random variables on a finite probability space with mean μ and variance $v < \infty$. Then for $N \approx \infty$ we have for any real number x

$$P(\frac{1}{\sqrt{N}}(X_1 + \cdots + X_N) \leq x) \approx \mathcal{N}_\mu^v(\leq x)$$

where $\mathcal{N}_\mu^v$ is the probability distribution on $\mathbb{R}$ given by the density function $t \mapsto \frac{1}{\sqrt{2\pi v}} e^{\frac{-(t-\mu)^2}{2v}}$. This distribution is called the *normal distribution*.

Proof. Without loss of generality let $\mu = 0$ and $v = 1$ and X_i standard random variables. The characteristic function of $S_N = \frac{1}{\sqrt{N}}(X_1 + \cdots + X_N)$ is given by (using some properties of the mean value from theorem 3.1.10 and the definition of the characteristic function)

$$\chi_{S_N}(t) = \chi_{\frac{1}{\sqrt{N}}(X_1+\cdots+X_N)}(t) = \chi_{X_1}(\frac{t}{\sqrt{N}})^N$$

We can use the theorem of Taylor to describe χ_{X_1}; for that we need the following values:

$$\begin{aligned}
\chi_{X_1}(0) &= 1 \\
\chi'_{X_1}(0) &= E(e^{i0X_1} iX_1) = iE(X_1) = 0 \\
\chi''_{X_1}(0) &= E(e^{i0X_1} i^2 X_1^2) = -E(X^2) = Var(X) = 1
\end{aligned}$$

Now we see that

$$\chi_{X_1}(t) = 1 + 0 + \frac{t^2}{2} + R(t)$$

where $R(t) < t^3$. Now it is easy to see that for standard t and $N \approx \infty$

$$\chi_{X_1}(\frac{t}{\sqrt{N}})^N = (1 + \frac{t^2}{2N} + R(\frac{t}{\sqrt{N}}))^N \approx (1 + \frac{t^2}{2N})^N \approx e^{-\frac{t^2}{2}}$$

With the density function from the claim we get exactly this characteristic function:

$$\begin{aligned}
E(e^{itX}) &= \int_{\mathbb{R}} e^{itx} P(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{itx} e^{\frac{-x^2}{2}} dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{\frac{1}{2}(-x^2+2itx+t^2-t^2)} dx = \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}} \int_{\mathbb{R}} e^{\frac{1}{2}(ix+t)^2} dx \\
&= e^{\frac{-t^2}{2}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{\frac{x^2}{2}} dx = e^{\frac{-t^2}{2}}
\end{aligned}$$

using Cauchy's integral theorem and – without proof – that the integral over the density function of the normal distribution equals 1.

The characteristic function describes the probability distribution uniquely, so the claim is proved. □

Corollary 3.2.4. The central limit theorem holds also, if the random variables are defined on a standardized probability space.

Proof. By definition of standardized probability spaces. □

3.2.2 The Laws of Large Numbers

The formulations and proofs of this section is close to [Ne77].

Lemma 3.2.5 (CHEBYSHEV's inequality). Let $\langle \Omega, P \rangle$ be a probability space, X a random variable and $\lambda \in \mathbb{R}, \lambda > 0$. Then

$$P(X \geq \lambda) \leq \frac{E(X^2)}{\lambda^2}$$

Proof. $E(X^2) \geq \sum_{X(\omega) \geq \lambda} X(\omega)^2 P(\omega) \geq \lambda^2 P(X \geq \lambda)$ □

Definition 25. Let $\langle \Omega, P \rangle$ be a probability space. A property Φ holds *almost sure*, if for any standard number $r > 0$ we find an event A_r with $P(A_r) < r$ and Φ holding for any $\omega \in \Omega \setminus A_r$.

Theorem 3.2.6 (NELSON's Weak Law of Large Numbers). Let $X_1, X_2, \ldots$ be independent random variables with the same probability distribution with mean μ and limited variance v.

Then for any natural number $n \approx \infty$ we have almost sure

$$\frac{1}{n}(X_1 + \cdots + X_n) \approx \mu$$

Proof. Let $n \approx \infty$ be a natural number, $r > 0$ a standard real number and $A_{n,r}$ the set of all λ with $P(\frac{1}{n}(X_1 + \cdots + X_n) - \mu \geq \lambda) \leq r$.

$A_{n,r}$ contains any standard λ, because for any standard λ we have by Chebyshev's inequality

$$\begin{aligned} P(\frac{1}{n}(X_1 + \cdots + X_n) - \mu \geq \lambda) &\leq \frac{E((\frac{1}{n}(X_1 + \cdots + X_n) - \mu)^2)}{\lambda^2} \\ &= \frac{Var(\frac{1}{n}(X_1 + \cdots + X_n))}{\lambda^2} \\ &= \frac{\frac{n}{n^2} v}{\lambda^2} \approx 0 < r \end{aligned}$$

Thus, $A_{n,r}$ contains some infinitesimal $\varepsilon_{n,r}$ by Underflow (see 2.2.3), so that the claim $\frac{1}{n}(X_1 + \cdots + X_n) \approx \mu$ holds almost sure, i.e. with probability approximately 1. □

Remark. In the weak law of large numbers the condition holds for any $n \approx \infty$ almost sure, i.e. the constructed sets $A_{n,r}$ may depend on n.

We can formulate the strong law of large numbers as the same property

$$\frac{1}{n}(X_1 + \cdots + X_n) \approx \mu$$

but it holds almost sure for any $n \approx \infty$, i.e. the sets of exceptions A_r depends only on r, not on n.

For conditions under that the strong law of large numbers hold (and its proof) see [Ne77].

Chapter 4

Hyperfinite Stochastic Analysis

In this final chapter we want to use the hyperfinite probability spaces introduced in chapter 3 to study stochastic processes, i.e. random functions. The goal of this chapter is the construction of a stochastic integral in the form of the Lebesgue integral constructed in section 3.1.2.

4.1 Stochastic processes

Definition 26. A *stochastic process* is a map $X : T \to \mathbb{R}^\Omega$, where $T \subseteq \mathbb{R}$ is the *timeline* and every X_t is a random variable on a pobability space $\langle \Omega, \mathcal{A}, P \rangle$.

A *sample path* of a stochastic process X is a function $\xi : T \to \mathbb{R}$ with $\xi(t) = X_t(\omega)$ for some (fixed) $\omega \in \Omega$.

We can study sample paths with the help of the probability space $\langle \Omega, \mathcal{A}, P \rangle$. To use our hyperfinite methods in stochastic analysis we have to define hyperfinite versions and standardizations of stochastic processes.

The most often used examples of timelines are $[0, \infty)$ ('continuous time') and $\mathbb{N}$ ('discrete time').

Definition 27. Let $\mathbf{T} \subseteq \mathbb{R}$ be an infinite standard set. A *standardizable stochastic process* $\mathbf{X}$ maps any hyperfinite version T of $\mathbf{T}$ to a stochastic process $X^{(T)}$ defined on a hyperfinite probability space $\langle \Omega_T, P_T \rangle$, such that for any standard $t \in \mathbf{T}$ and standard measurable set $A \subseteq \mathbb{R}$ the values $P_T(X_t^{(T)} \in A)$ are nearly equal for all T.

Theorem 4.1.1. Let $\mathbf{T} \subseteq \mathbb{R}$ be a standard set and $\mathbf{X}$ a standardizable stochastic process on $\mathbf{T}$. Then there is a standard probability space $\langle \mathbf{\Omega}, \mathcal{A}, \mathbf{P} \rangle$ and a

standard stochastic process $\mathbf{X} : \mathbf{T} \to \mathbb{R}^{\mathbf{\Omega}}$ such that for any standard $t \in \mathbf{T}$, standard measurable set $A \subseteq \mathbb{R}$ and any hyperfinite version T of $\mathbf{T}$

$$P(\mathbf{X}_t \in A) \approx P_T(X_t^{(T)} \in A)$$

Proof. For any standard $t \in \mathbf{T}$ we have a family of random variables $X_t^{(T)}$. By **(S)** we find a standard random variable $\mathbf{X}_t$ with $P(\mathbf{X}_t \in A) \approx P_T(X_t^{(T)} \in A)$ for any standard measurable set A and any hyperfinite version T of $\mathbf{T}$.

Again by **(S)** we find a standard stochastic process $\mathbf{X}$ that maps any standard $t \in T$ to the random variable $\mathbf{X}_t$ defined above.

Since there is a probability space $\langle \mathbf{\Omega}, \mathcal{A}, \mathbf{P} \rangle$ that is the range of every random variable $\mathbf{X}_t$ this probability space has to be standard (as the range of the standard random variables X_t with $t \in T$ being standard). □

For the rest of the chapter we assume that we have a given standard probability space $\langle \mathbf{\Omega}, \mathcal{A}, \mathbf{P} \rangle$ and for hyperfinite versions T of any timeline $\mathbf{T}$ hyperfinite probability spaces $\langle \Omega_T, P_T \rangle$. All random processes are defined on this probability spaces.

Definition 28. A stochastic process is called a *standardized stochastic process*, if it is the standardization of a standardizable stochastic process.

Example. Let $\mathbf{T} = [0, \infty)$, T any hyperfinite version of $\mathbf{T}$ and $\Omega_T := \{-1, 1\}^T$ with P_T the counting measure and

$$W_t^{(T)}(\omega) = \sum_{s<t} \omega_s \sqrt{\Delta s}$$

be the *random walk* on T. The class of random walks on the hyperfinite versions of $\mathbf{T}$ is standardizable.

Since $W_t^{(T)}(\omega)$ is the sum of independent random variables $R_s = \omega_s \sqrt{\Delta s}$ with mean 0 and variance Δs the sum $W_t^{(T)} = \sum_{s<t} R_s$ has mean 0 and variance t and by the central limit theorem we have

$$P_T(\sum_{s<t} R_s \in A) \approx \mathcal{N}_0^t(A)$$

for any standard measurable A and t, independent from the choice of T. ◇

4.1.1 Brownian motion

We have seen in an example that the class of random walks on the hyperfinite versions of $[0, \infty)$ is standardizable. The corresponding standardized stochastic process plays an importand role for the stochastic integral we want to construct.

Definition 29. A stochastic process W is a *Brownian motion*[1] or a *Wiener process*,[2] if

1. $W_0(\omega) = 0$ almost sure.
2. For $s, t \in [0, \infty)$ with $s < t$ we have for all measurable sets A
$$P(W_t(\omega) - W_s(\omega) \in A) \approx \mathcal{N}_0^{t-s}(A)$$
3. The sample paths $W(\omega) : [0, \infty) \to \mathbb{R}$ are continuous almost sure

In classical probability theory the existence of a Brownian motion is not obvious. One possibility of the existence proof is to show the convergence of random walks on partitions of $[0, \infty)$ with mesh $max\Delta t \to 0$ (see e.g. [Ba02]). In this nonstandard theory we can directly see that the standardization of the class of random walks is a Brownian motion.

Theorem 4.1.2. Let $\mathbf{T}$ be a timeline and W the class of random walks on its hyperfinite versions.

Proof. For any hyperfinite version T of $\mathbf{T}$:

1. $W_0^{(T)} = \sum_{s<0} \omega_s \sqrt{\Delta s} = 0$ is clear
2. $\sum_{s \leq r < t} \omega_s \sqrt{\Delta r}$ is normal distributed with mean 0 and variance $t - s$ by the central limit theorem
3. For each hyperfinite version T of $[0, \infty)$ and any standard $\langle t, \omega \rangle \in T \times \Omega_T$ and every $s \in T$ with $s \approx t$ we have $|W_t^{(T)}(\omega) - W_s^{(T)}(\omega)| \leq \sqrt{|s - t|} \approx 0$ so the sample paths are continuous by theorem 2.2.8. $\square$

The main difference to the classical theory is that we can imagine the Brownian motion with the help of its hyperfinite versions and we will replace a Brownian motion by a random walk on a hyperfinite version of $[0, \infty)$ whenever this is more comfortable.

To justify this method we will need the following theorem.

[1] Named after the botanist ROBERT BROWN (1773-1858) who described the physical phenomenon of random movement of particles in a liquid.

[2] Named after the mathematician NORBERT WIENER (1894-1964), who gave a first mathematical description of the physical phenomenon of Brownian motion.

Theorem 4.1.3. Let $\mathbf{T} = [0, \infty)$, T be a hyperfinite version of $\mathbf{T}$, $\mathbf{W}$ a Brownian motion on $\mathbf{T}$ and $W^{(T)}$ the random walk on T.

Then $\mathbf{W}$ and the standardization of $W^{(T)}$ are *equivalent*, i.e. we have for standard $t \in \mathbf{T}$ and standard measurable $A \subseteq \mathbb{R}$

$$P(\mathbf{W}_t \in A) \approx P(W_t^{(T)} \in A)$$

Proof. Both probabilities are nearly normal distributed (see the central limit theorem), i.e.

$$P(\mathbf{W}_t \in A) \approx \mathcal{N}_0^t(A) \approx P(W_t^{(T)} \in A)$$

□

4.2 Stochastic Differential Equations

In classical analysis we can define functions indirectly with the help of differential equations, e.g. as a solution of an ordinary differential equation of order 1

$$f'(x) = F(x, f(x))$$

Differential equations are a powerful tool for the mathematical modelling of various systems, as long as there are no random influences. Problems occur if one wants to model a random influenced system (e.g. quantum mechanics) or a system with so many influences that there are only statistical statements about its behaviour available (e.g. the stock market).

In Stochastic Analysis we want to describe stochastic processes with the help of (random) differential equations, so we need to construct some kind of random differentiation operator.

4.2.1 The Stochastic Integral

In this section let $\mathbf{T} = [0, \infty)$ be the timeline of all stochastic processes.

Definition 30. Let $\mathbf{W}$ a Brownian motion and $\mathbf{X}$ a stochastic process.

For any hyperfinite version T of $\mathbf{T}$ the *Itô sum*[3] of $\mathbf{X}$ over $[a, b] \subset \mathbf{T}$ with respect to T and $\mathbf{W}$ is the random variable

$$\sum_{t \in T \cap [a,b]} \mathbf{X}_t \Delta W_t^{(T)}$$

[3]Named after the Japanese mathematician KIYOSHI ITÔ (1915-2008) who invented a Stochastic Calculus with the help of this stochastic integral, and thus solved the question for a stochastic version of the theory of differential equations.

where $\Delta W_t^{(T)} := W_{t+\Delta t}^{(T)} - W_t^{(T)} = \pm\sqrt{\Delta t}$.

The stochastic process $\mathbf{X}$ is *(Itô) integrable* over $[a, b]$ with respect to $\mathbf{W}$, if the class of its Itô sums is standardizable. Then we define the *(Itô) integral* of $\mathbf{X}$ over $[a, b]$ with respect to $\mathbf{W}$ as the standard operator such that for all standard intervals $[a, b] \subseteq \mathbf{T}$ we have for all hyperfinite versions T of $\mathbf{T}$

$$\int_a^b \mathbf{X} d\mathbf{W} \approx \sum_{t \in T \cap [a,b]} \mathbf{X}_t \Delta W_t^{(T)}$$

Remark.

1. The Itô integral is not a real number (as the Riemann or Lebesgue integral would be), but a random variable.

2. The Itô sum can be defined ω-pathwise, i.e. for every $\omega \in \Omega_T$ we define the sum.

   Since the sample paths of a Brownian motion are almost sure nowhere differentiable the formal symbol $d\mathbf{W}$ is not definable, and for that reason the integral cannot be directly defined ω-pathwise. In the classical theory we have to show that the Itô sums converge in probability for $max\Delta t \to 0$; in our context we have this directly by **(S)**.

3. If $\mathbf{X}^{(1)}$, $\mathbf{X}^{(2)}$ are equivalent stochastic processes (i.e. for all measurable A we have $\mathbf{P}(\mathbf{X}^{(1)} \in A) = \mathbf{P}(\mathbf{X}^{(2)} \in A)$) and $\mathbf{W}^{(1)}$, $\mathbf{W}^{(2)}$ are equivalent Brownian motions the Itô sums are obviously equivalent.

   Since the Itô integral is the standardization of the Itô sums it is easy to see that the integrals $\int_0^t \mathbf{X}^{(1)} d\mathbf{W}^{(1)}$ and $\int_0^t \mathbf{X}^{(2)} d\mathbf{W}^{(2)}$ are equivalent.[4]

Example. Let $\mathbf{W}$ be a Brownian motion. The integral of $\mathbf{W}$ with respect to itself exists. Let W be a random walk corresponding to $\mathbf{W}$ and $\omega \in \Omega$ and $t \in \mathbf{T}$ be standard.

With the help of the binomial equation

$$W_t^2 + 2W_t \Delta W_t + (\Delta W_t)^2 = (W_t + \Delta W_t)^2 = W_{t+\Delta t}^2$$

and the identity

$$(\Delta W_t)^2 = (\pm\sqrt{\Delta t})^2 = \Delta t$$

[4]The integrals are even indistinguishable, i.e. $\mathbf{P}(\int_0^t \mathbf{X}^{(1)} d\mathbf{W}^{(1)} - \int_0^t \mathbf{X}^{(2)} d\mathbf{W}^{(2)}) = 1$.

we get

$$\begin{aligned}\sum_{s\in T\cap[0,t]} W_s\Delta W_s &= \frac{1}{2}\sum_{s\in T\cap[0,t]}(W^2_{s+\Delta s} - W_s^2 - (\Delta W_s)^2)\\ &= \frac{1}{2}(W_{t+\Delta t} - W_0 - \sum_{s\in T\cap[0,t]}\Delta s)\\ &\approx \frac{1}{2}(W_t^2 - t)\end{aligned}$$

This is independent of the choice of T, so that $\mathbf{W}$ is integrable with respect to itself and

$$\int_0^t \mathbf{W}d\mathbf{W} = \frac{1}{2}\mathbf{W}_t^2 - \frac{t}{2}$$

This example demonstrates that the integrand and the integrator do not need to be stochastically independent. ◇

An interesting step in the last proof is that we have used $(\Delta W_t)^2 = \Delta t$. In the classical theory this could be read as '$(d\mathbf{W})^2 = dt$', but this equation has no definable meaning and had to be transcripted into terms of limits.

4.2.2 The Lemma of Itô

Definition 31. A stochastic process $\mathbf{X}$ is an *Itô process*, if there is a Brownian motion $\mathbf{W}$, a random variable $\mathbf{X}_0$ and functions $\mu, \sigma : \mathbb{R}^2 \to \mathbb{R}$ such that

$$\mathbf{X}_t = \mathbf{X}_0 + \int_0^t \mu(s, \mathbf{X}_s)ds + \int_0^t \sigma(s, \mathbf{X}_s)d\mathbf{W}_s$$

where the first integral is a Lebesgue integral using the Lebesgue measure and the second integral is an Itô integral.

μ is called the *drift* and σ the *diffusion*.

As a more general definition we could use two-dimensional stochastic processes $\mu, \sigma : \mathbb{R}^2 \to \mathbb{R}^{\mathbf{\Omega}}$ as drift and diffusion.

Similar to theorem 2.2.10 we can prove the existence of an Itô process whenever μ and σ are continuous:

Theorem 4.2.1. Let $\mu, \sigma : \mathbb{R}^2 \to \mathbb{R}$ be continuous and $\mathbf{X}_0$ any random variable. Then there is an Itô process $\mathbf{X}_t$ with

$$\mathbf{X}_t = \mathbf{X}_0 + \int_0^t \mu(s, \mathbf{X}_s)ds + \int_0^t \sigma(s, \mathbf{X}_s)d\mathbf{W}_s$$

Proof. Let T be any hyperfinite version of $\mathbf{T}$ and X, W the corresponding hyperfinite versions of $\mathbf{X}, \mathbf{W}$, resp.

For each $t \geq 0$ and $\omega \in \Omega_T$ set

$$X_{t+\Delta t}(\omega) := X_t(\omega) + \mu(t, X_t(\omega))\Delta t + \sigma(t, X_t(\omega))\Delta W_t(\omega)$$

The standardization of X_t is obviously a stochastic process with the properties we want. □

Notation. We will often write such an integral equation formal as a differential equation

$$d\mathbf{X}_t = \mathbf{U}_t dt + \mathbf{V}_t d\mathbf{W}_t$$

though this is only a notation. $\mathbf{X}$ will almost sure be nowhere differentiable (like the Brownian motion).

This notation is justified by the following fact: For each hyperfinite version of the timeline and X and t being standard we have

$$\begin{aligned}
\Delta X_t &= X_{t+\Delta t} - X_t \\
&\approx X_0 + \sum_{s \leq t} U_s \Delta s + \sum_{s \leq t} V_s \Delta W_s - X_0 - \sum_{s<t} U_s \Delta s - \sum_{s<t} V_s \Delta W_s \\
&= U_t \Delta t + V_t \Delta W_t
\end{aligned}$$

So we see again that we can interpret d as a notation for the standardization of Δ as before in the definition of the Riemann integral as the Lebesgue integral with respect to the Lebesgue measure given by $\lambda_T(x) = \Delta x$.

Finally we will prove the one-dimensional special case of Itô's lemma. This proof is similar to the proof of [An76], but somewhat shorter, because Anderson proves a more general version.

Theorem 4.2.2 (ITÔ's lemma). Let $h : \mathbb{R}^2 \to \mathbb{R}$ be continuous differentiable in the first argument and two-times continuous differentiable in the second one, and $\mathbf{X}$ an Itô process satisfying $d\mathbf{X}_t = \mu(t, \mathbf{X}_t)dt + \sigma(t, \mathbf{X}_t)d\mathbf{W}_t$. Then $h(t, \mathbf{X}_t)$ is an Itô process and

$$\begin{aligned}
dh(t, \mathbf{X}_t) &= \dot{h}(t, \mathbf{X}_t)dt + h'(t, \mathbf{X}_t)d\mathbf{X}_t + \frac{1}{2}\sigma(t, \mathbf{X}_t)^2 h''(t, \mathbf{X}_t)dt \\
&= (\dot{h}(t, \mathbf{X}_t) + h'(t, \mathbf{X}_t)\mu(t, \mathbf{X}_t) + \frac{1}{2}\sigma(t, \mathbf{X}_t)^2 h''(t, \mathbf{X}_t))dt \\
&\quad + h'(t, \mathbf{X}_t)\sigma(t, \mathbf{X}_t)d\mathbf{W}_t
\end{aligned}$$

where $\dot{h}$ is the (first) derivative in the first argument, h' is the first derivative in the second argument and h'' the second derivative in the second argument.

Proof. Let T be any hyperfinite version of the timeline, and X and $t \in T$ be standard.

We will need the identity

$$\begin{aligned}\frac{(\Delta X_t)^2}{\Delta t} &\approx \frac{(\mu(t,\mathbf{X}_t)\Delta t + \sigma(t,\mathbf{X}_t)\Delta W_t)^2}{\Delta t} \\ &= \mu(t,\mathbf{X}_t)^2\Delta t + 2\mu(t,\mathbf{X}_t)\sigma(t,\mathbf{X}_t)\Delta W_t + \frac{(\sigma(t,\mathbf{X}_t)\Delta W_t)^2}{\Delta t} \\ &\approx \frac{(\pm\sigma(t,\mathbf{X}_t)\sqrt{\Delta t})^2}{\Delta t} = \sigma(t,\mathbf{X}_t)^2\end{aligned}$$

That means $(\Delta X_t)^2 \approx \sigma(t,\mathbf{X}_t)^2\Delta t$ up to an error ε with $\frac{\varepsilon}{\Delta t} \approx 0$. With the help of Taylor's theorem we see

$$\begin{aligned}h(t,x) &= h(t_0,x_0) + \dot{h}(t_0,x_0)(t-t_0) \\ &\quad + h'(t_0,x_0)(x-x_0) + \frac{1}{2}h''(t_0,x_0)(x-x_0)^2 + \varepsilon(t,x)\end{aligned}$$

where $\varepsilon(t,x)$ is an error function continuous in $\langle 0,0\rangle$. Now we can calculate (using $t_0 = s$, $t = s+\Delta s$, $x_0 = X_s$, $x = X_{s+\Delta s}$ and $\frac{\varepsilon(s,X_s)}{ds} \approx 0$)

$$\begin{aligned}h(t,X_t) - h(0,X_0) &= \sum_{s<t}(h(s+\Delta s, X_{s+\Delta s}) - h(s,X_s)) \\ &= \sum_{s<t}\dot{h}(s,X_s)\Delta s + h'(s,X_s)\Delta X_s \\ &\quad + \frac{1}{2}h''(s,X_s)(\Delta X_s)^2 + \varepsilon(s,X_s) \\ &\approx \sum_{s<t}\dot{h}(s,X_s)\Delta s + h'(s,X_s)\Delta X_s \\ &\quad + \frac{1}{2}h''(s,X_s)\sigma(s,X_s)^2\Delta s\end{aligned}$$

So we get the claim by

$$\begin{aligned}\Delta h(t,X_t) &= h(t+\Delta t, X_{t+\Delta t}) - h(t,X_t) \\ &\approx \dot{h}(t,X_t)\Delta t + h'(t,X_t)\Delta X_t + \frac{1}{2}h''(t,X_t)\sigma(t,X_t)^2\Delta t \\ &\approx \dot{h}(t,X_t)\Delta t + \frac{1}{2}h''(t,X_t)\sigma(t,X_t)^2\Delta t \\ &\quad + h'(t,X_t)(\mu(t,X_t)\Delta t + \sigma(t,X_t)\Delta W_t) \\ &\approx (\dot{h}(t,X_t) + h'(t,X_t)\mu(t,X_t) + \frac{1}{2}h''(t,X_t)\sigma(t,X_t)^2)\Delta t \\ &\quad + h'(X_t)\sigma(t,X_t)\Delta W_t\end{aligned}$$

□

Example. Let $\mu, \sigma \in \mathbb{R}$ and $\mathbf{S}_0$ a random variable. Then the *geometric Brownian motion*

$$\mathbf{S}_t = \mathbf{S}_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma \mathbf{W}_t}$$

is a solution of the stochastic differential equation

$$d\mathbf{S}_t = \mu \mathbf{S}_t dt + \sigma \mathbf{S}_t d\mathbf{W}_t$$

with (random) initial value $\mathbf{S}_0$.

Proof. We use Itô's lemma and any hyperfinite version T of $\mathbf{T}$ to get

$$\begin{aligned} \Delta S_t &= S_0 e^{(\mu - \frac{\sigma^2}{2})t + \sigma W_t} \left((\mu - \frac{\sigma^2}{2})\Delta t + \sigma \Delta W_t + \frac{1}{2}\sigma^2 \Delta t \right) \\ &= S_t \left(\mu \Delta t + \sigma \Delta W_t \right) \\ &= \mu S_t \Delta t + \sigma S_t \Delta W_t \end{aligned}$$

This Itô process is used in the *Black & Scholes-formula* for price valuations of (European style) stock market options. ◇

Final Remark. This Hyperfinite Probability Theory should be generalizable in the usual ways, e.g. generalization of the integrands and integrators of the Itô integral, multidimensional Stochastic Analysis, study of complex random variables, and so on.

The present thesis aims to demonstrate that hyperfinite measure spaces are definable within *Internal Set Theory*, even without the use of external sets, and did this by the construction of the Itô integral and giving a proof of Itô's lemma.

Appendix A

External Sets

An important tool of ROBINSON's Nonstandard Mathematics are external sets. In [Ne77] NELSON is not enough perspicuous on how to use external sets within (an extension of) IST. It seems he wants to work with another set theory, that is connected to IST with the help of a map $*$ (as the two superstructures are within ROBINSON's theory). In section A.1 I try to formulate this part of [Ne77] in a clearer way.

We have seen in this thesis (by the example of Stochastic Analysis) that one can use IST to use nonstandard methods for applications in measure theory. Surprisingly we can find the opinion that this is not possible within IST, e.g. [CK90], p. 287: 'In practice, internal set theory has been adequate for certain areas of Robinsonean analysis [...], but **inadequate for** others (e.g. **probability theory**, [...]).' A similar opinion can be read in the preface of [KR04].

The only argument for this claim is that one cannot construct the so-called Loeb spaces within this theory. In sections A.2 and A.3 I will demonstrate that it is possible to formulate the theorem of Loeb – but Loeb spaces are uninteresting, because we are not interested in an external Loeb space, but we want to construct another (standard) measure space within IST (as the standardized measure space).

A.1 The $*$-map and the σ-map

In [Ne77] the theory of external sets and the internal set theory are connected by a map $*$ that maps any entity of IST to an entity of a model of IST. In this section we will try to formalize this vague idea with the help of the *internal theory* IST as described in this paper and the *external theory* $ZFC+$inaccessible from appendix B.

By theorem B.2.2 we find a natural model $\mathbb{S} = \langle S, \in \rangle$ of ZFC and a model

$\mathbb{I} = \langle I, e, S\rangle$ of the internal theory in the external theory, and $\mathbb{I}$ is an elementary extension of $\mathbb{S}$, i.e. every theorem of ZFC or IST is true within $\mathbb{S}$ or $\mathbb{I}$, resp, i.e. if we restrict the quantifiers to S or I, resp.

Definition 32. Let $*$ be the following map

$$\begin{aligned} I &\to \mathcal{P}(I) \\ X &\mapsto X^* := \{x \in I \mid xeX\} \end{aligned}$$

Example. We can define the set of real numbers within IST, so we find a set $\mathbb{R} \in I$ that represents this set. In the external theory we talk about the model of IST where we find a set $\{x \in I \mid xe\mathbb{R}\} = \mathbb{R}^*$.

This is not the usual set of real numbers of ZFC, because this set of real numbers had to be a subset of S and we proved that $\mathbb{R} \setminus S \neq \emptyset$. Since we get $\mathbb{R}$ in the same way as we get the real numbers in ZFC, the ZFC-real numbers have to be the restrict of our set to its standard elements $\mathbb{R}^\sigma := \mathbb{R}^* \cap S$.

Note that both $\mathbb{R}^*$ and $\mathbb{R}^\sigma$ have the same properties in the language $\mathcal{L}_\in$, if we interpret $\in$ with E in the first case. ◇

Definition 33. Let σ be the map

$$\begin{aligned} I &\to S \\ X &\mapsto X^\sigma := X^* \cap S \end{aligned}$$

Example. In the external theory we may define sets in any way we are used to, e.g. let the monad of $x \in \mathbb{R}^*$ be

$$\mu(x) := \{y \in \mathbb{R}^* \mid y \approx x\}$$

As we have seen in the paper $|\mu(x) \cap S| \leq 1$, so we can define the standard part $^\circ$ as an external map

$$\begin{aligned} \mathbb{R}^* &\to \mathbb{R}^\sigma \cup \{\pm\infty\} \\ x &\mapsto \begin{cases} {}^\circ x \in \mu(x) \cap S \\ {}^\circ x := sgn(x)\infty \end{cases} \begin{matrix} \text{if } |\mu(x) \cap S| = 1 \\ \text{else} \end{matrix} \end{aligned}$$

In the internal theory this map cannot be formulated, because its domain is an external set, and this would be illegal set formation. ◇

The use of these maps is that we can change the set theory in which we talk about the entities of IST. To avoid confusion NELSON defines a *set* as an entity of the internal theory and an *external set* as an entity of the external theory. An external set is called *internal*, if it is the $*$-image of a set and it is *standard*, if it is the $*$-image of a standard set.

In the external theory we can talk about external sets with a different point of view, as the following examples demonstrates.

Theorem A.1.1. Let X be a hyperfinite set. Then X^* is uncountable.

Proof. Without loss of generality let $X = \{0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\} \subseteq [0,1]$ (because every finite set is bijective to such a set). In the external theory we can define a function $f : X^* \to [0,1]^*$ by $f(x) = x$ if $[x, x+\Delta x]^\sigma = \emptyset$ and $f(x) \in [x, x+\Delta x]^\sigma$ else. Note that the external set $[x, x+\Delta x]^\sigma$ has at most one element, since the interval $[x, x+\Delta x]$ has infinitesimal length. Obviously f is injective.

On the other hand we have by construction of f any standard element of the interval $[0,1]^*$ in $f(X^*)$, i.e. we can find an (external) injection $\iota : [0,1]^\sigma \to X^*$ by $\iota(x) := f^{-1}(x)$, so X^* is uncountable. □

A.2 Where are the Loeb spaces in IST?

An important tool of measure theory within Robinsonean Analysis is the construction of measure spaces with the help of hyperfinite spaces. In chapter 3 we use this idea to construct probability measures as standardizations of finite probability spaces.

As above we will use the notion of 'measure space' for measure spaces within the internal theory. If we talk about measure spaces of the external theory we will use the term 'external measure space'.

The problem with measure theory within Robinsonean Analysis is that the external version $\langle X, \mathcal{A}, \mu\rangle^* = \langle X^*, \mathcal{A}^*, \mu^*\rangle$[1] of a measure space $\langle X, \mathcal{A}, \mu\rangle$ is not an external meassure space, since $\mathcal{A}^*$ does not need to be an external σ-algebra:

Example. Let $X = \mathbb{R}$ and $\mathcal{A} = \sigma_{algebra}(\{(a,b) \mid a,b \in \mathbb{R}, a < a\})$ be the Borel sets (where $\sigma_{algebra}(X)$ is the smallest σ-algebra containing X). Then $\bigcup_{i\in\mathbb{N}^*} [-\frac{1}{n}, \frac{1}{n}]^* = \{0\}^* \in \mathcal{A}^*$, but $\bigcup_{i\in\mathbb{N}^\sigma} [-\frac{1}{n}, \frac{1}{n}]^* = \mu(0) \notin \mathcal{A}^*$. ◇

The solution to this difficulty was given by PETER A. LOEB in [Lo75] with the following construction.

[1]The triple is an entity of ZFC, so if we use an external triple we will get the triple of the corresponding external sets.

Theorem A.2.1 (Theorem of Loeb). Let $\langle X^*, \mathcal{A}^*, \mu^* \rangle$ be an internal measure space, and set the map ${}^\circ\nu : \mathcal{A}^* \to [0,\infty]^\sigma$ by ${}^\circ\nu(A) = {}^\circ(\nu(A))$.

Then there is a unique $\lambda : \sigma_{algebra}(\mathcal{A}^*) \to [0,\infty]^\sigma$ with $\lambda(A) = {}^\circ\nu(A)$ for $A \in \mathcal{A}^*$; λ is given by

$$\lambda(B) = inf\{{}^\circ\nu(A) \mid A \in \mathcal{A}^* \wedge B \subseteq A\}$$

Proof. This proof is a translation of the proof given in [Cu83] into the external theory of IST.

Existence. Let λ be the defined external set function and $(B_n)_{n\in\mathbb{N}^\sigma}$ an external sequence of mutually disjoint subsets of $\sigma_{algebra}(\mathcal{A}^*)$.

$$\sum_{n\in\mathbb{N}^\sigma} \lambda(B_n) \leq \lambda(\bigcup_{n\in\mathbb{N}^\sigma} B_n)$$

is clear, since any superset of $\bigcup_{n\in\mathbb{N}^\sigma} B_n$ is a superset of any of the (mutually disjoint) B_n.

If $\sum_{n\in\mathbb{N}^\sigma} \lambda(B_n) = \infty$ we are done.

Otherwise let $s := \sum_{n\in\mathbb{N}^\sigma} \lambda(B_n) \in \mathbb{R}^\sigma$ and $\varepsilon > 0$. Choose an internal sequence $(A_n^*)_{n\in\mathbb{N}^*}$ with $\nu(A_n^*) < \lambda(B_n) + 2^{-n}\varepsilon$ (using **(S)**). Then we have for any $m \in \mathbb{N}^\sigma$

$$\nu(\bigcup_{n\leq m} A_n^*) < \sum_{n\leq m} (\lambda(B_n) + 2^{-n}\varepsilon) < s + \varepsilon$$

and thus by lemma 2.2.3 (overflow) for some $M \in \mathbb{N}^* \setminus \mathbb{N}^\sigma$

$$\nu(\bigcup_{n\leq M} A_n^*) < s + \varepsilon$$

For any $\varepsilon > 0$ we find a superset $\bigcup_{n\leq M} A_n^* \in \mathcal{A}^*$ of $\bigcup_{n\in\mathbb{N}^\sigma} B_n$ with measure less than $s + \varepsilon$, so we have

$$\bigcup_{n\in\mathbb{N}^\sigma} B_n \leq s = \sum_{n\in\mathbb{N}^\sigma} \lambda(B_n)$$

Thus λ is σ-additive, and it is obviously an external measure extending ${}^\circ\nu$.

Uniqueness. Assume λ is the measure given by the inf-definition and λ' is another measure on $\sigma_{algebra}(\mathcal{A}^*)$ extending ${}^\circ\nu$, and let $A \in \sigma_{algebra}(\mathcal{A}^*)$.

If we find some $B \in \mathcal{A}^*$ with $B \subseteq A$ and ${}^\circ\nu(B) = \infty$ then $\lambda(A) = \infty = \lambda'(A)$.

If we find no such B then there is a sequence $(B_n)_{n\in\mathbb{N}}$ in $\mathcal{A}^*$ with $A = \bigcup_{n\in\mathbb{N}} B_n$, because A is an element of the σ-algebra generated by $\mathcal{A}^*$. Let $B' := \bigcup_{n\in\mathbb{N}} B_n$ and $b := {}^\circ\nu(B')$, note that $b < \infty$.

Now we have

$$\lambda'(A) \leq \lambda(A)$$

since $\lambda'(A) \leq {}^\circ\nu(A')$ for any $A \subseteq A' \in \sigma_{algebra}(\mathcal{A}^*)$, and thus for the infimum of those values.

On the other hand we have for the same reason $\lambda'(B'\setminus A) \leq \lambda(B'\setminus A)$, but that implies

$$\lambda'(A) = b - \lambda'(B' \setminus A) \geq b - \lambda(B' \setminus A) = \lambda(A)$$

So that we see that $\lambda = \lambda'$. □

Definition 34. Let $\langle X^*, \mathcal{A}^*, \mu^*\rangle$ be an internal measure space and λ the measure extension from theorem A.2.1. Then the *Loeb space* of $\langle X^*, \mathcal{A}^*, \mu^*\rangle$ is the external measure space $\langle X^*, L(\mathcal{A}^*), \nu_L^*\rangle$ where

1. $L(\mathcal{A}^*)$ is the completion of $\sigma_{algebra}(\mathcal{A}^*)$ with respect to λ, i.e. the σ-algebra generated by $\sigma_{algebra}(\mathcal{A}^*)$ and the λ-nullsets.
2. ν_L^* is the extension of λ to $L(\mathcal{A}^*)$.

ν_L^* is called the *Loeb measure* associated with ν^*.

A.3 A comparison of Loeb spaces and standardized measure spaces

The difference between Loeb spaces and standardized measure spaces consists in the fact that the construction of Loeb is neccessary to extend an internal measure space $\langle X^*, \mathcal{A}^*, \mu^*\rangle$ to an external measure space $\langle X^*, L(\mathcal{A}^*), \mu_L^*\rangle$ so that the restriction to a subset $Y^\sigma \subseteq X^*$ is a new external measure space: The Loeb measure can be used to study phenomena of the original universe by **'pushing down'** the measure to the standard universe $\mathbb{S}$.

Example. Let $F = \{0, \frac{1}{H}, \ldots, \frac{H-1}{H}, 1\}$ be a finite set with $H \approx +0$, and ν be the counting measure. Then ν_L^* is an external measure on F^* and we can show that F^* implies the Lebesgue measure on $[0,1]^\sigma$ by setting $\lambda(A) = \nu(\{h \in F \mid {}^\circ h \in A\})$.

Note that $\{h \in F \mid {}^\circ h \in A\}$ does not need to be Loeb measurable for any subset A of $[0,1]^\sigma$ (see [Cu83], §4(1) for more details). ◇

In IST we use the different idea of **'pushing up'** the measures of the hyperfinite measure spaces $\langle X^*, \mathcal{P}(X)^*, \mu_X^* \rangle$ to a standard measure space $\langle \mathbf{X}^*, \mathcal{A}^*, \hat{\mu}^* \rangle$ where $X^\sigma = \mathbf{X}^\sigma$ for any hyperfinite version X^*. Then we can study (standard) phenomena of the internal universe having in mind that $\mathbb{S}$ and $\mathbb{I}$ are elementary equivalent, i.e. the theories of $\mathbb{S}$ and $\mathbb{I}$ are equal (see definition 44).

Example. Let X^* be any hyperfinite version of $[0,1]^*$ and μ_X be the measure generated by the weight function

$$m_X(x) := \Delta x = min\{y \in X^* \mid y > x\} - x$$

for $x \neq max(X^*)$ and $m_X(max(X^*)) = 0$. Then we have seen in section 3.1.1 that the standardization of this measure spaces (on any hyperfinite version X^* of $[0,1]^*$) is the Lebesgue measure on $[0,1]^*$.

Since $\mathbb{S}$ and $\mathbb{I}$ are elementary equivalent we may study the internal properties (using only the symbol $\in$, not the symbol S) of this measure on $[0,1]^*$ to get the classical theory of the Lebesgue measure. $\diamond$

In this examples we see that the Loeb space construction is allowed to concentrate on special (well-behaved) sets. The cost of this advantage is a loss of intuitiveness, since we don't know anything about the standard elements. This leads e.g. to terms like the *lifting* $F : X^* \to \mathbb{R}^*$ of a function $f : X^\sigma \to \mathbb{R}^\sigma$ with ${}^\circ F(x) = f({}^\circ x)$ for all $x \in X^*$.

In my opinion the use of standardizable measures solves these technical problems and thus removes the loss of intuitiveness, since we can imagine any hyperfinite version to get the idea of the construction (e.g. to interpret dx formally as the standardization of Δx on any hyperfinite version of $\mathbb{R}$) and yet we have always the property that X contains any standard element that we might be interested in.

Appendix B

How to get a model of IST

In [CK90] we find a possibility of how to expand certain ZFC-models to an IST-model. Since we need some definitions and lemmata from set theory and model theory this section will be somewhat technical. However, everything we need will be defined. The proofs of existence of an IST-model (section B.2) and of the conservativeness (section B.3) reproduce the arguments in [CK90] completing the outlines given there.

In the following we will work within an extension of ZFC. To get a natural model of ZFC (see below) we use the additional assumption that there is a strongly inaccessible cardinal number (see definition 40).

B.1 Tools from Set Theory and Model Theory

Notation. When we talk about models we will use mathematical bold face letters to denote the model ($\mathbb{A}, \mathbb{B}, \mathbb{C}, ...$) and the corresponding roman letters for its set of individuals ($A, B, C, ...$).

Definition 35. A set X is *transitive*, if $Y \subseteq X$ for any $Y \in X$.

An *ordinal number* or just an *ordinal* is a transitive set that contains only transitive elements.

We will write ON for the class of all ordinal numbers.

The *successor* $s : ON \to ON$ maps any ordinal ξ to $s(\xi) = \xi \cup \{\xi\}$.

ON is ordered by $\in$, so we will sometimes write $\alpha < \beta$ instead of $\alpha \in \beta$.

This order is even a well-order, since there is no descending chain $\xi_1 > \xi_2 > \ldots$ by the axiom of regularity. In particular we can conclude that s maps ξ to the smallest ordinal greater than ξ.

As the first ordinals we get $0 := \emptyset, 1 := \{\emptyset\}, 2 := \{\emptyset, \{\emptyset\}\}, ..., s(n) := n \cup \{n\}, ...$ and as the union of all these numbers we get the smallest infinite ordinal number ω, then we get $s(\omega) = \omega \cup \{\omega\}, s(s(\omega)) = \omega \cup \{\omega, \{\omega\}\}$ and so on. The next limit ordinal is $\omega + \omega$.

For practical reasons we regard 0 not as a limit ordinal.

Definition 36. We can define addition, multiplication and exponentiation on the ordinal numbers as usually by transfinite induction, where α is a fixed ordinal, β is an arbitrary ordinal and λ is a limit ordinal.

	1	2	3
A	$\alpha + 0 = \alpha$	$\alpha + s(\beta) = s(\alpha + \beta)$	$\alpha + \lambda = \bigcup_{\xi<\lambda} \alpha + \xi$
M	$\alpha \cdot 0 = 0$	$\alpha \cdot s(\beta) = \alpha \cdot \beta + \alpha$	$\alpha \cdot \lambda = \bigcup_{\xi<\lambda} \alpha \cdot \xi$
E	$\alpha^0 = 1$	$\alpha^{s(\beta)} = \alpha^\beta \cdot \alpha$	$\alpha^\lambda = \bigcup_{\xi<\lambda} \alpha^\xi$

From **A2** with $\beta = 0$ we see that $\alpha + 1 = s(\alpha)$.

Although these operations are the usual operations for finite numbers we have to be careful with infinite ordinal numbers, e.g. addition and multiplication are not commutative:

$$2 + \omega = \bigcup_{n<\omega} (2 + n) = \omega \neq \omega \cup \{\omega, \{\omega\}\} = \omega + 1 + 1 = \omega + 2$$

In particular we see by $2 + \omega = \omega$ that some ordinal numbers can absorb others.

Definition 37. Two sets A, B are *equipollent*, if we find a bijection between them; we write $A \sim B$.

The ordinal number ω has the property that it is not equipollent to any smaller ordinal number, but obviously $\omega + 17 \sim \omega$, $\omega \cdot 2 \sim \omega$ ($\mathbb{Z}$ is countable) and even $\omega^2 \sim \omega$ ($\mathbb{Q}$ is countable). This leads to the definition of the cardinal numbers.

Definition 38. A *cardinal number* or just a *cardinal* is an ordinal number that is not equipollent to any predecessor.

Lemma B.1.1. Any set X is equipollent to a unique cardinal number κ.

Proof. With the help of the axiom of choice we find a function $c : \mathcal{P}(X) \to X$ with $c(\emptyset) = \emptyset$ and $c(Y) \in Y$ otherwise. Then we construct by induction a map $\iota : ON \to X$ by transfinite induction as

$$\iota(\alpha) = c(X \setminus \bigcup_{\xi\in\alpha} \iota(\xi))$$

There is an ordinal α such that $\bigcup_{\xi\in\alpha} \iota(\xi) = X$, because otherwise we had an injection $ON \to X$, but ON is a proper class.[1] and $\iota(\xi) = \emptyset$ for all $\xi > \alpha$. The smallest ordinal that is equipollent to α is the cardinal κ.

The uniqueness of κ follows directly from $\kappa \sim X$. □

The proved fact, that every set is bijective to an ordinal number is called the *well order principle*, since the constructed bijection $N \to X$ implies a well-order on X. The well-order principle is even equivalent to the axiom of choice, but we won't need the other direction of the statement here.

Definition 39. For any set X the unique cardinal κ with $X \sim \kappa$ from lemma B.1.1 is called the *power of* X. We write $|X| := \kappa$.

With the help of the power of sets we can define addition, multiplication and exponentiation for cardinal numbers as follows:

$$\begin{aligned} |A| + |B| &:= |(\{0\} \times A) \cup (\{1\} \times B)| \\ |A| \cdot |B| &:= |A \times B| \\ |A|^{|B|} &:= |A^B| \end{aligned}$$

We don't want to prove the independence of the choice of A and B.

Definition 40. The *cofinality* $cof(\kappa)$ of a cardinal κ is the smallest ordinal α such that there is a family of sets $(A_i)_{i<\alpha}$ with $|A_i| < \kappa$ and $\bigcup_\alpha A_i = \kappa$.

κ is *regular*, if $cof(\kappa) = \kappa$ and *singular* otherwise.

A cardinal number $\kappa > \omega$ is *(strongly) inaccessible*, if κ is regular and for all cardinals $\alpha < \kappa$ we have $2^\alpha < \kappa$.

Definition 41. We set recursively for any ordinal α and for limit ordinals λ the *regular sets*

$$R_0 := \emptyset; R_{\alpha+1} := \mathcal{P}(R_\alpha); R_\lambda := \bigcup_{\xi<\lambda} R_\xi$$

And the $\beth$-sequence (read: 'beth')

$$\beth_0 = \omega; \beth_{\alpha+1} := 2^{\beth_\alpha}; \beth_\lambda := \bigcup_{\xi<\lambda} \beth_\xi$$

[1] If ON was not a proper class it had to be a set. Then it contained only transitive elements and was transitive itself, so that we had the contradiction $ON \in ON$ by definition of the ordinal numbers.

Note that regular sets are transitive, i.e. if $X \in R_\xi$ we have $X \subseteq R_\xi$; in particular, $X \in Y \in R_\xi$ implies $X \in R_\xi$.

Lemma B.1.2. For any ordinal ξ we have

$$R_{\omega+\xi} \sim \beth_\xi$$

Proof. By transfinite induction (using the axiom of choice):

1. $R_\omega = \omega = \beth_0$ by definition.
2. For any ordinal α with $R_{\omega+\alpha} \sim \beth_\alpha$:
$$R_{\omega+\alpha+1} = \mathcal{P}(R_{\omega+\alpha}) \sim 2^{R_{\omega+\alpha}} \sim 2^{\beth_\alpha} = \beth_{\alpha+1}$$
3. For any limit ordinal λ with $R_{\omega+\xi} \sim \beth_\xi$ for all $\xi < \lambda$:
$$R_{\omega+\lambda} = \bigcup_{\xi<\omega+\lambda} R_\xi = \bigcup_{\xi<\lambda} R_{\omega+\xi} \overset{\text{AC}}{\sim} \bigcup_{\xi<\lambda} \beth_\xi = \beth_\lambda$$

□

We will use the following fact without proof: If ξ is an inaccessible cardinal number, then $\langle R_\xi, \in\rangle$ is a model of ZFC; such a model (using the 'real' $\in$-relation on a regular set) is called *natural*.

B.2 A model of IST

Definition 42. A consistent set of formulas $\Sigma(v)$ (in a given language $\mathcal{L}$) with v being the only free variable is called a *type*.

A model $\mathbb{A}$ *realizes* a type $\Sigma(v)$, if there is some element $a \in A$ such that any statement of $\Sigma[a]$ (replacing any occurrence of the free variable v by a in any formula of $\Sigma(v)$) holds in $\mathbb{A}$.

Let $\mathcal{L}$ be a language and X a set. Then we define $\mathcal{L}_X = \mathcal{L} \cup \{c_x \mid x \in X\}$, where any of the c_x is a new constant.

Definition 43. Let $\kappa \geq \omega$ be a cardinal. A model $\mathbb{A}$ is κ-*saturated*, if for every $X \subseteq A$ with fewer than κ elements, $\mathbb{A}$ realizes any type $\Sigma(v)$ in the language $\mathcal{L}_X$, if $\Sigma(v)$ is consistent with the theory of $\mathbb{A}$ in the language $\mathcal{L}_X$.[2]

$\mathbb{A}$ is *saturated*, if it is $|A|$-saturated.

[2]The theory of $\mathbb{A}$ is the set of all true statements about $\mathbb{A}$ in the language $\mathcal{L}$. Since we may use constants for any element of X in $\Sigma(v)$ we need consistency with respect to a theory of $\mathbb{A}$ that uses this constants, too, i.e. a theory of $\mathbb{A}$ in the language $\mathcal{L}_X$.

Definition 44. A model $\mathbb{B}$ is an *extension* of $\mathbb{A}$, if $A \subseteq B$ and for any relation $R \subseteq A^n$ exists $R' \subseteq B^n$ such that $R = R' \cap A^n$.[3]

$\mathbb{B}$ is an *elementary extension* of $\mathbb{A}$, if for any formula $\Phi(v_1, \ldots v_n)$ and any n-tuple $a_1, \ldots, a_n \in A$ the statement $\Phi(a_1, \ldots, a_n)$ holds in $\mathbb{A}$ if and only if it holds in $\mathbb{B}$.

Note that a model extension is in general not elementary, since we will find statements using an all-quantifier in the theory of $\mathbb{A}$, but the quantifier will range over the larger set of individuals B in the theory of $\mathbb{B}$.

Lemma B.2.1. Let $\mathbb{A}$ be a model and κ a cardinal with $\omega \leq |A| \leq 2^\kappa$ and the language of the theory has less than κ symbols. Then there is a κ^+-saturated elementary extension of $\mathbb{A}$ of power 2^κ.

In the lemma the symbol κ^+ denotes the smallest cardinal greater than κ.

Proof. We define inductively elementary extensions $\mathbb{B}_\xi$ of $\mathbb{A}$ for $0 \leq \xi \leq 2^\kappa$ with $|B_\xi| \leq 2^\kappa$.

1. $\mathbb{B}_0 := \mathbb{A}$ is by definition an elementary extension of $\mathbb{A}$.
2. For any $0 \leq \xi < 2^\kappa$: The number of subsets $X \subseteq B_\xi$ with $|X| = \kappa$ is at least $|B_\xi|$.

   Reason: Assume $|B_\xi| = 2^\kappa$. The number of subsets of power κ is with the binomial theorem $\binom{B_\xi}{\kappa} \leq 2^\kappa$.

   If $|B_\xi| < 2^\kappa$ the number of subsets of power κ could only get smaller.
3. For any $0 \leq \xi < 2^\kappa$: Given $X \subseteq B_\xi$ with $|X| = \kappa$ we find 2^κ types $\Sigma(v)$ in $\mathcal{L}_X$.

   Reason: Since $|\mathcal{L}| \leq \kappa$ we have $|\mathcal{L}_X| = 2\kappa \sim \kappa$. Since any formula can only contain finitely many constants the total number of formulas in $\mathcal{L}_X$ is $[\kappa]^{<\omega}\omega \sim \kappa$, where $[\kappa]^{<\omega}$ denotes the power of the set of subsets of κ with power less than ω.

   The set of all types in $\mathcal{L}_X$ is a subset of the set of all formulas, hence we have 2^κ types.
4. For any $0 \leq \xi < 2^\kappa$, any $X \subseteq B_\xi$ with $|X| = \kappa$ and any type $\Sigma(v)$ in $\mathcal{L}_X$ that is consistent with the theory of $\mathbb{B}_\xi$ in $\mathcal{L}_X$ we introduce a new constant $c_{X,\Sigma}$.

   Note: By the estimations given above we introduce at least $2 \cdot 2^\kappa$ new symbols, so that $|B_\xi \cup \{c_{X,\Sigma}\}| \leq 3 \cdot 2^\kappa \sim 2^\kappa$.

[3] For this definition we assume that any function $f : A^n \to A$ is actually a relation $R_f \subseteq A^{n+1}$ with $R_f(a_1, \ldots, a_{n+1}) :\Leftrightarrow a_{n+1} = f(a_1, \ldots, a_n)$.

5. For any $0 \leq \xi < 2^\kappa$ we find an elementary model extension $\mathbb{B}_{\xi+1}$ of power 2^κ for the theory of $\mathbb{B}_\xi$, in which all statements $\Sigma(c_{X,\Sigma})$ hold.

   Reason: For any $X \subseteq A$ and any type $\Sigma(v)$ the existence of $c_{X,\Sigma}$ is consistent to the theory of $\mathbb{B}_\xi$ in the language $\mathcal{L}_X$ and thus also in the language $\mathcal{L}_{B_\xi}$.

   The theory $T = Th_{\mathcal{L}_{B_\xi}}(\mathbb{B}_\xi) \cup \bigcup_{X,\Sigma} \Sigma(c_{X,\Sigma})$ is consistent and has an infinite model, because $\mathbb{B}_\xi$ is infinite; so we find a model $\mathbb{B}_{\xi+1}$ of T of power 2^κ by the theorem of LÖWENHEIM/SKOLEM/TARSKI.

   Since the theory of $\mathbb{B}_{\xi+1}$ is formulated in the language $\mathcal{L}_{B_\xi}$ we have $B_\xi \subseteq B_{\xi+1}$.

   Of course, $\mathbb{B}_{\xi+1}$ is also a model of the reduct of T to the language $\mathcal{L}$.

6. For any limit ordinal λ set $\mathbb{B}_\lambda := \bigcup_{\xi<\lambda} \mathbb{B}_\xi$. The power of $\mathbb{B}_\lambda$ is 2^κ, since $B_\xi \subseteq B_{\xi+1}$ for any ξ.

Any $\mathbb{B}_\xi$ is by construction an elementary extension of all models $\mathbb{B}_\nu$ for $\nu < \xi$ and hence of $\mathbb{A}$. Finally set $\mathbb{B} := \mathbb{B}_{2^\kappa}$.

Now let $X \subseteq B$ be of power κ and $\Sigma(v)$ be a type in the language $\mathcal{L}_X$. Then there is some $\xi < 2^\kappa$ such that $X \subseteq B_\xi$, because $cof(2^\kappa) > \kappa$ (by KÖNIG's theorem) and thus $\bigcup X \not\sim 2^\kappa$.

By the construction above $\Sigma(v)$ is realized in $\mathbb{B}_{\xi+1}$, and because $\mathbb{B}$ is an elementary extension of $\mathbb{B}_{\xi+1}$ it is also realized in $\mathbb{B}$; that means $\mathbb{B}$ is κ-saturated. □

Theorem B.2.2. Let ξ be a limit ordinal, $\mathbb{A} = \langle R_{\omega+\xi}, \in\rangle \subseteq \langle B, E\rangle$ an $(\beth_\xi)^+$-saturated elementary extension. Then $\mathbb{B} = \langle B, E, R_{\omega+\xi}\rangle$ satisfies **(I)**, **(S)** and **(T)** (see section 2.1), if we interpret $\in$ by E and S by $R_{\omega+\xi}$.

Proof. The existence of $\mathbb{B}$ follows from lemma B.2.1. We have to check that the axioms **(I)**, **(S)**, **(T)** hold in $\mathbb{B}$.

T. Let $\Phi(v_0, v_1, \ldots, v_n)$ be any formula of the language of $\mathbb{A}$, $a_1, \ldots, a_n \in A$ and asume

$$\forall x \in R_{\omega+\xi}\ \Phi(x, a_1, \ldots, a_n)$$

holds in $\mathbb{B}$. Since $\mathbb{B}$ is an elementary extension of $\mathbb{A}$ we get immediately

$$\forall x\ \Phi(x, t_1, \ldots, t_n))$$

holding in $\mathbb{B}$.

I. Let $\Phi(x, y, \dots)$ be a formula in the language of $\mathbb{A}$ where the variable z does not occur. Assume

$$\forall^{FIN} z \in R_{\omega+\xi} \exists x \forall y \in z \Phi(x, y, \dots)$$

holds in $\mathbb{B}$. Now let $F \subseteq R_{\omega+\xi}$ be the set of all sets z that are finite with respect to the theory of $\mathbb{B}$, i.e. that are equipollent to some n with $nE\omega$.[4] We have $|F| \leq |R_{\omega+\xi}| = \beth_\xi$ by lemma B.1.2. In $\mathcal{L}_F$ we can fomrulate the type

$$\Sigma(x) := \{\forall y \in c_z \Phi(x, y, \dots)\}$$

where $c_z \in \mathcal{L}_F$ are constants for $z \in F$. By assumption $\Sigma(x)$ is consistent to the theory of $\mathbb{B}$, and because $\mathbb{B}$ is $(\beth_\xi)^+$-saturated we find some element $c_\Sigma \in B$ that fulfils all statements in $\Sigma[c_\Sigma]$.
Since it is clear that $\bigcup_{z \in F} z = R_{\omega+\xi}$ the element c_Σ has the property

$$\forall y \in R_{\omega+\xi} \Phi(c_\Sigma, y, \dots)$$

and that implies the consequence of **(I)**

$$\exists x \forall y \in R_{\omega+\xi} \Phi(x, y, \dots)$$

For the other direction we use that $R_{\omega+\xi}$ is transitive: Assume that

$$\exists x \forall y \in R_{\omega+\xi} \Phi(x, y, \dots)$$

holds in $\mathbb{B}$. For any $z \in R_{\omega+\xi}$ we have by transitivity also $z \subseteq R_{\omega+\xi}$ and hence

$$\exists x \forall y \in z \Phi(x, y, \dots)$$

holds for any $z \in R_{\omega+\xi}$. The restriction to the $\mathbb{B}$-finite subsets (see above)

$$\forall^{FIN} z \in R_{\omega+\xi} \exists x \forall y \in z \Phi(x, y, \dots)$$

is what we had to show.

S. Let $\Phi(z, \dots)$ be any formula in the language of $\mathbb{A}$ where the free variables x, y do not occur, and $x \in R_{\omega+\xi}$. Then the set

$$y = \{z \in x \mid \Phi(z, \dots) \text{ holds in } \mathbb{B}\}$$

is a subset of x and thus belongs to $R_{\omega+\xi}$. □

[4] Those elements z do not need to be finite in the sense of theory of $\mathbb{A}$; we defined finiteness as z is bijective to a proper initial segment of ω; but since $\mathbb{B}$ is not a natural model we could find additional elements n with $nE\omega$, but $n \notin \omega$. Since in the natural model any $n \in \omega$ is an initial segment of ω this holds also in the elementary model extension $\mathbb{B}$ for all $nE\omega$.

Corollary B.2.3. Let ξ be an inaccessible cardinal number and $\mathbb{A} = \langle R_\xi, \in \rangle$, then the model $\mathbb{B}' = \langle B, E, R_\xi \rangle$ from theorem B.2.2 is a model of IST.

Proof. Since $\mathbb{A}$ is a natural model of ZFC and $\omega + \xi = \xi$ the conditions of theorem B.2.2 are satisfied. □

B.3 Conservativity

Finally we can easily prove that IST is conservative over ZFC.

Theorem B.3.1. Let φ be a $\mathcal{L}_\in$-statement that is provable in IST. Then φ is provable in ZFC.

Proof. Since φ is provable in IST we can find a finite subset $A \subseteq ZFC$ such that φ is provable from A and **(I)**, **(S)** and **(T)**.

If we assume that $\neg\varphi$ is consistent to A we can find a natural model $\langle R_{\omega+\xi}, \in \rangle$ for $A \cup \{\neg\varphi\}$ by the reflection principle and then by theorem B.2.2 a model $\langle B, E, R_{\omega+\xi} \rangle$ satisfying $A \cup \{\neg\varphi,$ **(I)**, **(S)**, **(T)**$\}$. That is impossible, because this set of statements is inconsistent (we can prove $\varphi \wedge \neg\varphi$ from it).

Hence the assumption must be wrong, that is we can not find any model of ZFC that makes $\neg\varphi$ true. We can conclude that φ is true in any model and hence it is provable in ZFC. □

Bibliography

Papers about Nonstandard Analysis

[An76] R. M. ANDERSON: A Nonstandard Representation for Brownian Motion and Itô Integration; Israel Journal of Math. **25**, 1976 (pp 15-46).

[Cu83] N. J. CUTLAND: Nonstandard Measure Theory And Its Applications; Bull. London Math. Soc. **15**, 1976 (pp 529-589).

[Lo75] P. A. LOEB: Conversion from Nonstandard to Standard Measure Spaces and Applications in Probability Theory; Bull. American Math. Soc. **211**, 1975 (pp 113-122).

[Ne77] E. NELSON: Internal Set Theory: A New Approach to Nonstandard Analysis; Bull. American Math. Soc. **83**, 1977 (pp 1165-1198).

[Sk34] A. T. SKOLEM: Über die Nicht-Charakterisierbarkeit der Zahlenreihe mittels endlich oder unendlich vieler Aussagen mit ausschließlich Zahlenvariablen; Fundam. Math. **23**, 1934 (pp 150-161).

Textbooks about Nonstandard Analysis

[Go98] R. GOLDBLATT: Lectures on the Hyperreals. An Introduction to Nonstandard Analysis; Springer, 1998.

[KR04] KANOVEI/REEKEN: Nonstandard Analysis, Axiomatically; Springer, 2004.

[Ke76] H.J. KEISLER: Foundations of Infinitesimal Calculus; Boston, 1976.

[Ne87] E. NELSON: Radically Elementary Probability Theory; Princeton Press, 1987.

[Ro76] A. ROBINSON: Nonstandard Analysis; North-Holland, 1974.

Other textbooks used

[Ba02] H. BAUER: Wahrscheinlichkeitstheorie (5. Auflage); de Gruyter Lehrbuch, 2002.

[CK90] CHANG/KEISLER: Model Theory (3nd ed); Springer, 1990.

[KS98] I. KARATZAS/S. E. SHREVE: Brownian Motion and Stochastic Calculus (2nd ed); Springer, 1998.

[Kl06] A. KLENKE: Wahrscheinlichkeitstheorie; Springer, 2006.

[Sc09] K.D. SCHMIDT: Maß und Wahrscheinlichkeit; Springer, 2009.

Index