Probability Logic

Master's Thesis in Mathematics

Author:	Rutger Kuyper
Student number:	0715204
Supervisor:	Sebastiaan Terwijn
Second reader:	Mai Gehrke

Radboud University Nijmegen

Probability Logic

Rutger Kuyper

Contents

Introduction **1**

1 Prerequisites **5**
1.1 Measure Theory 5

2 Probability Logic and PAC-Learning **9**
2.1 The Definition of Probability Logic 9
2.2 Probabilistic Truth and Tautologies 13
2.3 PAC-Learning Probabilistic Sentences 15

3 Earlier Approaches **21**
3.1 Keisler's Probability Logic 21
3.2 Valiant's Robust Logic 24

4 Computability Theory **27**
4.1 Validity Is Π^1_1-Hard 27
4.2 Satisfiability Is Π^0_1-Hard 35

5 Model Theory **39**
5.1 A Downward Löwenheim-Skolem Theorem 39
5.2 Satisfiability and Lebesgue measure 43
5.3 The Löwenheim Number 49
5.4 Compactness 50

Future Research **59**

Bibliography **61**

Index **63**

Introduction

Probability is everywhere. Be it a lottery, goats hiding behind doors or ravens that might not be black, we cannot escape the influence of probability. This influence is also very clear in the theory of learning: if one wants to say something about the *entire* population of some class (for example, that all ravens are black), we can never expect to check all members of the population; the best we can do is *induce* a general statement from our finite observations, making sure to take the unreliability of this induction into account.

Classical mathematical logic does not allow us to express these inductive statements — indeed, the universal quantifier $\forall$ in classical first-order logic expresses that something holds for an entire population, without allowing us to specify the reliability (or probability) of this statement. From this deficiency the field of *probability logic* has sprouted, with many contributions over the years. We will discuss some of these contributions below.

The theory of inductive learning (as described above) has also been the subject of much research. The most notable contribution in this field of machine learning is Valiant's model of Probably Approximately Correct learning (PAC-learning) [19], which has become one of the main paradigms in computational learning theory. We will not discuss this learning model in much detail, but we will be looking at its intersection with probability logic.

The first probability logic that we want to mention comes from unpublished work of Friedman, which is exhibited in Steinhorn [16]. Among the quantifiers he studied is a quantifier expressing that "there exists non-measure 0 many". Since there are not many relations to our current work, we will not discuss this quantifier in more detail.

A probability logic that is closer to the logic which we will discuss in this thesis is the probability logic proposed by Keisler [10]. His models are formed by classical models enriched with a probability measure, so that the classical quantifiers $\forall$ and $\exists$ can be replaced by probability quantifiers expressing that a statement holds with a certain probability. While it is close to ours and we will reuse some of the ideas expressed for Keisler's logic by Keisler [11] and Hoover [7], it is also vastly different — it is not learnable, unlike ours; also, it does not have the classical existential quantifier $\exists$, which our logic does have. Furthermore, our logic has the advantage of using the syntax of classical first-order logic, resulting in a cleaner presentation. Nonetheless, we will use several ideas and constructions originally developed for Keisler's probability logic.

Valiant [20] was the first to introduce a probability logic with PAC-learning properties. However, as can be seen in Section 3.2 of this thesis, his approach is fundamentally different from ours. We therefore do not consider our logic as an extension of his work, but as a different path.

Terwijn [17] also introduced a logic combining probability logic and some sense of PAC-learning. It is this logic that we will be studying in this thesis, exhibiting various model-theoretic, recursion-theoretic and learning-theoretic results. We will present both small and large modifications of earlier results, as well as completely new results.

We will now give a brief outline of this thesis. The first chapter will discuss the various prerequisites needed for this thesis — most notably, it presents the necessary definitions and results from measure theory. Since we need probability measures for the definition of our logic, it is inevitable that we discuss these measures. This chapter only contains some elementary and well-known facts; more specialistic results are introduced when they are needed.

Chapter 2 introduces the probability logic which we will be studying in this thesis. We equip classical models with a probability measure, so that we can give a probabilistic interpretation of the universal quantifier, while keeping the usual (classical) definition of the existential quantifier. We also explain the choices we made for some (possibly controversial) parts of our definition; most notably, we need to work around the well-known problem that projections of measurable sets need not be measurable (which was famously mistaken to be true by Lebesgue). After discussing some elementary results on tautologies in our logic, we return to the issue of learnability and introduce some notion of learning of sentences closely related to PAC-learning. To show that our logic is indeed learnable under this notion, we make slight repairs to an earlier proof of this fact by Terwijn [17].

After this introduction to our logic, in the third chapter we will go back to the logics of Keisler and Valiant. We will present the definitions for both of these logics and compare them to our logic. This chapter serves as a historic overview of some earlier approaches and allows us to put our work into context.

Chapter 4 concerns the computability theory of our logic. Our first result shows that the set of tautologies is Π^1_1-hard; i.e. that it is at least as hard as first-order arithmetic enlarged with universal second-order quantifiers. To prove this, we adapt a proof originally given by Hoover [7] for Keisler's probability logic. In particular, this shows that the set of tautologies is also not computably enumerable; therefore there exists no effective calculus. We expand our argument to show that the universal fragment of arithmetic can be interpreted in the set of satisfiable sentences (where we remark that, unlike in classical logic, the notions of validity and satisfiability are not complementary to each other), which shows that satisfiability in our logic is at least as hard as classical satisfiability.

Finally, in Chapter 5 we obtain some model-theoretic results for our probability logic. The first result, a version of the downwards Löwenheim-Skolem theorem, is obtained through a generalisation of an argument by Hoover [8]. To be more precise, we show that every model is equivalent to a model of cardinality 2^ω (unlike in classical logic, there are sentences which are only probabilistically satisfiable in uncountable models). While *a priori* this tells us nothing about the probability measure on this model, we next present a new result showing that every satisfiable sentence in satisfiable in some model on $[0, 1]$ equipped with the Lebesgue measure. Our final result shows that, while the compactness theorem does not hold for our logic, we still have some weaker notion of compactness. To show this, we use a construction to define a measure on ultraproducts of measure spaces, as originally given in Bageri and Pourmahdian [2].

We finish this thesis by briefly discussing in which directions further research on this logic could be taken. Our specific logic has not been the subject of much research yet and therefore the possible directions are aplenty — for example, one might look into the decidability of fragments of our logic, or one could look at model-theoretic questions such as elimination of quantifiers. Therefore, the open questions are very much like black ravens — we can never expect to research all of them.

1 Prerequisites

1.1 Measure Theory

This thesis relies heavily on measure theory — to be more precise, on probability measures. In this section we will briefly state some basic definitions and results, to be used throughout this thesis. The treatment will be short and mostly without proofs; for more details the reader is referred to a book on measure theory (such as Bogachev [4]).

We begin with the basic definitions.

Definition 1.1.1. Let X be a set. Then a **σ-algebra** over X is a collection $\mathcal{A}$ of subsets of X such that:

1. $\emptyset, X \in A$;
2. If $A \in \mathcal{A}$, then $X \setminus A \in \mathcal{A}$;
3. If $A_0, A_1, \ldots$ is a countable sequence of elements from $\mathcal{A}$, then
$$\bigcup_{i \in \omega} A_i \in \mathcal{A}.$$

For a given collection Γ of subsets of X, we call the least σ-algebra $\mathcal{A}$ containing Γ the **σ-algebra generated by** Γ; this will be denoted by $\sigma(\Gamma)$.

Definition 1.1.2. Let $\mathcal{A}$ be a σ-algebra over X. Then a **measure** on $\mathcal{A}$ is a function $\mu : \mathcal{A} \to [0, \infty)$ such that for every countable sequence $A_0, A_1, \ldots$ of disjoint sets from $\mathcal{A}$ we have

$$\mu(\bigcup_{i \in \omega} A_i) = \sum_{i=0}^{\infty} \mu(A_i).$$

We will call the sets in $\mathcal{A}$ the **μ-measurable sets**. If the intended σ-algebra is clear from the context, we will call μ a measure over X instead of a measure over $\mathcal{A}$.

The tuple $(X, \mathcal{A}, \mu)$ is called a **measure space**

Definition 1.1.3. If μ is a measure over X such that $\mu(X) = 1$, then we call μ a **probability measure** or **probability distribution**, and we will usually denote it by $\mathcal{D}$.

Definition 1.1.4. Let μ be a measure over X. If $f : X \to \mathbb{R}$, we say that f is a **measurable function** if $\{x \mid f(x) < c\}$ is μ-measurable for every $c \in \mathbb{R}$.

More generally, if ν is a measure over Y, then a function $f : X \to Y$ is called a **measurable function** if for every ν-measurable set B the set $f^{-1}(B)$ is μ-measurable.

Now we discuss some specific constructions we will need in this thesis.

Definition 1.1.5. Let $\mathcal{A}$ be a σ-algebra over X and let $\mathcal{A}'$ be a σ-algebra over Y. Then we define the **sum of $\mathcal{A}$ and $\mathcal{A}'$**, denoted as $\mathcal{A} \oplus \mathcal{A}'$, as the σ-algebra $\{A \cup B \mid A \in \mathcal{A}, B \in \mathcal{A}'\}$ over $X \cup Y$.

Furthermore, if we let μ be a measure over $\mathcal{A}$ and let ν be a measure over $\mathcal{A}'$, then we define the **sum of μ and ν**, denoted as $\mu \oplus \nu$, as the measure over $\mathcal{A} \oplus \mathcal{A}'$ and given by

$$(\mu \oplus \nu)(C) := \mu(C \cap X) + \nu(C \cap Y).$$

For our definition of probability logic, we will be needing the product measure. To introduce this measure, we need some form of extension theorem.

Definition 1.1.6. Let X be a set. A **Boolean algebra** $\mathcal{B}$ is defined as in Definition 1.1.1, but with item 3 weakened to finite unions.

A **pre-measure** over a Boolean algebra $\mathcal{B}$ is a $\mu_0 : \mathcal{B} \to [0, \infty)$ defined as in Definition 1.1.2, but with the equality only for sequences such that $\bigcup_{i \in \omega} A_i \in \mathcal{B}$.

Theorem 1.1.7 (Carathéodory's extension theorem). *Every pre-measure μ_0 on a Boolean algebra $\mathcal{B}$ can be uniquely extended to the σ-algebra generated by $\mathcal{B}$.*

Using this extension theorem, we can introduce the product measure.

Definition 1.1.8. Let $\mathcal{A}$ be a σ-algebra over X and let $\mathcal{A}'$ be a σ-algebra over Y. Then we define the **product of $\mathcal{A}$ and $\mathcal{A}'$**, denoted as $\mathcal{A} \otimes \mathcal{A}'$, as the σ-algebra generated by $\{A \times B \mid A \in \mathcal{A}, B \in \mathcal{A}'\}$.

Furthermore, if we let μ be a measure over $\mathcal{A}$ and let ν be a measure over $\mathcal{A}'$, then we define the **product of μ and ν**, denoted as $\mu \otimes \nu$, as the measure over $\mathcal{A} \otimes \mathcal{A}'$ that is the unique extension of

$$(\mu \otimes \nu)(A \times B) := \mu(A)\nu(B).$$

We will denote μ^n for the product of n copies of μ. Observe that, if $\mathcal{D}$ and $\mathcal{E}$ are both probability measures, then so is $\mathcal{D} \otimes \mathcal{E}$.

One of the most well-known theorems from measure theory is Fubini's theorem. Since we do not need any integrals in this thesis, we will only state the part of the theorem about measures.

Theorem 1.1.9 (Fubini's theorem). *Let μ, ν be probability measures over sets X respectively Y. Let $A \subseteq X \times Y$ be $\mu \otimes \nu$-measurable. Then:*

1. *For each $x \in X$, the section*

   $$A_x := \{y \in Y \mid (x, y) \in A\}$$

   is ν-measurable.[1]

2. *The function $x \mapsto \nu(A_x)$ is measurable.*

We now return to the topic of generation of σ-algebras. The definition given above (Definition 1.1.1) is fairly non-constructive in the sense that it is a top-down definition. The next proposition states some facts we will need (in Chapter 5); these will be proven through a more bottom-up approach.

Proposition 1.1.10. *Let Γ be a collection of subsets of X of cardinality κ. Then:*

1. *$\sigma(\Gamma)$ has cardinality at most κ^ω.*

2. *Every $A \in \sigma(Y)$ can be formed using a combination of countable unions, intersections and complements of at most countably many elements from Γ.*

[1] Usually, Fubini's theorem is stated for the *completion* of the product measure. Then this statement only holds for almost all $x \in X$ instead of for all $x \in X$.

Proof. We give an inductive construction of $\sigma(\Gamma)$ over $\aleph_1$, from which these facts will be directly clear.

First, we let $\mathcal{A}_0$ be Γ. For the successor step, we let $\mathcal{A}_{\alpha+1}$ consist of:

- For $A \in \mathcal{A}_\alpha$: $X \setminus A \in \mathcal{A}_{\alpha+1}$.
- For every countable sequence $A_0, A_1, \ldots \in \mathcal{A}_\alpha : \bigcup_{i\in\omega} A_i \in \mathcal{A}_{\alpha+1}$.

Finally, we take unions at limit ordinals.

One can now easily verify that $\sigma(\Gamma)$ is a σ-algebra containing Γ. Both claims can now be directly verified using induction. □

Finally, we define *Borel measures* and *complete* measures.

Definition 1.1.11. Let X be a topological space. The **Borel σ-algebra** is the σ-algebra generated by the open sets of X. If a measure is defined on the Borel σ-algebra, we call it a **Borel measure**. We call the elements of the Borel σ-algebra **Borel sets**.

Definition 1.1.12. A measure μ is called **complete** if every subset of a set having measure 0 is also measurable (and has measure 0). The **completion** of μ is the least complete measure extending μ (in the sense of the smallest σ-algebra).

2

Probability Logic and PAC-Learning

2.1 The Definition of Probability Logic

As is to be expected, we start the main part of this thesis with the necessary definitions. In this thesis, we will be studying the probability logic as originally defined by Terwijn [17].

Our logic is motivated by an idea of what it means to 'learn' a first-order statement. Our language will be that of regular first-order logic. We ask ourselves: what does it mean to learn if such a statement holds in a particular model? In particular, what does it mean to learn a quantifier?

For existential statements we clearly want this to mean that, through observation, we find a witness of such a quantifier. However, for universal statements we can never expect to induce these with full certainty through observation; more precisely, if we take a finite number of *samples* from the model we can only induce the universal quantifier up to a certain inaccuracy — that is, we can induce that it probably holds for *many* elements of the model. This motivates the addition of a *probability distribution* to our model. We can then formulate learning a sentence over an unknown model $\mathcal{M}$ and an unknown probability distribution $\mathcal{D}$ over this model as following:

Given a certain confidence level for the universal quantifiers, can we induce if a statement holds from a finite number of samples taken according to the distribution $\mathcal{D}$?

Following this idea, we introduce the logic that we will be studying in this thesis. Section 2.3 will show that this logic is indeed learnable; namely, that it is PAC-learnable.

Definition 2.1.1. Let $\mathcal{L}$ be a first-order language, possibly containing equality, over an at most countable signature. Let $\varphi = \varphi(x_1, \dots, x_n)$ be a first-order formula in the language $\mathcal{L}$, and let $\varepsilon \in [0,1]$. Furthermore, let $\mathcal{M}$ be a classical first-order model for $\mathcal{M}$ and let $\mathcal{D}$ be a probability measure on $\mathcal{M}$. Finally, let $a_1, \dots, a_n \in \mathcal{M}$. Then we inductively define **ε-truth** $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(\bar{a})$ as follows.

1. For every atomic formula φ:
$$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(\bar{a}) \text{ if } \mathcal{M} \models \varphi(\bar{a}).$$
2. We treat the logical connectives $\wedge$ and $\vee$ classically, e.g.
$$(\mathcal{M}, \mathcal{D}) \models_\varepsilon (\varphi \wedge \psi)(\bar{a}) \text{ if } (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(\bar{a}) \text{ and } (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(\bar{a}).$$
3. The existential quantifier is treated classically as well:
$$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \exists x_{n+1}(\varphi(\bar{a}, x_{n+1}))(\bar{a})$$
if there exists an $a_{n+1} \in \mathcal{M}$ such that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(\bar{a}, a_{n+1})$.
4. The case of negation is split into subcases as follows:
   (a) For φ atomic, $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg\varphi$ if $(\mathcal{M}, \mathcal{D}) \not\models_\varepsilon \varphi(\bar{a})$.
   (b) $\neg$ distributes in the classical way over $\wedge$ and $\vee$, e.g.
   $$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg(\varphi \wedge \psi)(\bar{a}) \text{ if } (\mathcal{M}, \mathcal{D}) \models_\varepsilon (\neg\varphi \vee \neg\psi)(\bar{a}).$$
   (c) $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg\neg\varphi(\bar{a})$ if $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(\bar{a})$.
   (d) $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg(\varphi \rightarrow \psi)(\bar{a})$ if $(\mathcal{M}, \mathcal{D}) \models_\varepsilon (\varphi \wedge \neg\psi)(\bar{a})$.
   (e) $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg\exists x(\varphi(\bar{a}, x))$ if $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \forall x(\neg\varphi(\bar{a}, x))$.
   (f) $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg\forall x(\varphi(\bar{a}, x))$ if $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \exists x(\neg\varphi(\bar{a}, x))$.
5. $(\mathcal{M}, \mathcal{D}) \models_\varepsilon (\varphi \rightarrow \psi)(\bar{a})$ if $(\mathcal{M}, \mathcal{D}) \models_\varepsilon (\neg\varphi \vee \psi)(\bar{a})$.
6. Finally, we define $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \forall x_{n+1}(\varphi(x_{n+1}))(\bar{a})$ if
$$\Pr_{\mathcal{D}}[a_{n+1} \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(\bar{a}, a_{n+1})] \geq 1 - \varepsilon.$$

Definition 2.1.2. Let $\mathcal{L}$ be a first-order language, possibly containing equality, over an at most countable signature. Let $\varepsilon \in [0,1]$. Then an **ε-model** $(\mathcal{M}, \mathcal{D})$ for the language $\mathcal{L}$ consists of a classical first-order model $\mathcal{M}$ for $\mathcal{L}$ and a probability distribution $\mathcal{D}$ over $\mathcal{M}$ such that:

1. For all formulas $\varphi = \varphi(x_1, \ldots, x_n)$ and all $a_1, \ldots, a_{n-1} \in \mathcal{M}$, the set

$$\{a_n \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(a_1, \ldots, a_n)\}$$

is $\mathcal{D}$-measurable (i.e. all definable sets of dimension 1 are measurable).

2. All relations of arity n are $\mathcal{D}^n$-measurable and all functions of arity n are measurable as functions from $(\mathcal{M}^n, \mathcal{D}^n)$ to $(\mathcal{M}, \mathcal{D})$ (in particular, constants are $\mathcal{D}$-measurable).

Definition 2.1.3. We will say that a sentence φ is an **ε-tautology** or is **ε-valid** (notation: $\models_\varepsilon \varphi$) if for all ε-models $(\mathcal{M}, \mathcal{D})$ it holds that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi$. Furthermore, we will say that φ is a **probabilistic tautology** if it is ε-valid for all $\varepsilon \in [0, 1]$.

Similarly, we call φ **ε-satisfiable** if there exists an ε-model $(\mathcal{M}, \mathcal{D})$ such that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi$. Finally, φ is called **probabilistically satisfiable** if there exists an $\varepsilon \in [0, 1)$ such that φ is ε-satisfiable.

We make some remarks on the truth definition.

1. Observe that everything is treated classically, except for the interpretation of $\forall x(\varphi(x))$ in case 6. Instead of saying that we have $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(a)$ for **all** elements $a \in \mathcal{M}$, we merely say that it holds for 'many' of the elements. This makes precise the notion described above.

2. The definition of negation might need some explanation. Classically, our definition of negation would be the same as defining $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg\varphi$ if $(\mathcal{M}, \mathcal{D}) \not\models_\varepsilon \varphi$. However, in our case this is different. Since, if we were to choose this alternative definition, we would get that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg\forall x(\varphi(x))$ holds if and only if we have that $\Pr_{\mathcal{D}}[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(a)] > \varepsilon$. This is stronger than a mere existential statement. Aside from the fact that this changes the intended meaning of $\exists$, it would also be impossible to learn this universal quantifier using only a finite sample. Therefore, our alternative negation transforming a $\forall$ into an $\exists$ and vice versa is more suited for our purposes.

3. Note that both $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \forall x(\varphi(x))$ and $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \exists x(\neg\varphi(x))$ might hold; for example, φ might hold on a set of measure one and simultaneously there could be a counterexample of measure 0. Thus, the logic defined above is *paraconsistent*. However, even

though both φ and $\neg\varphi$ might be satisfiable, they cannot both be tautologies; this can be seen by observing that only one of them can be true in a model on one point.

4. We choose to interpret $\varphi \to \psi$ as $\neg\varphi \vee \psi$. This can be seen as expressing that $(\mathcal{M}, \mathcal{D}) \not\models_\varepsilon \neg\varphi$ implies that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi$ holds. This is weaker than the classical definition of saying that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi$ implies $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi$; namely, because of the paraconsistency, the fact that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi$ holds does not necessarily imply that $(\mathcal{M}, \mathcal{D}) \not\models_\varepsilon \neg\varphi$. We really do not want the classical definition; for example, if one were to take for ψ a contradiction (such as $\exists x(R(x) \wedge \neg R(x))$) this would introduce the classical $\forall x(\varphi(x))$ as $\exists x(\neg\varphi(x)) \to \psi$. For reasons explained above, we do not want this.

5. The case for $\varepsilon = 1$ is pathological; for example, all universal statements are always true. We will therefore often exclude this case.

6. We remark that it is not enough to require just the relations to be measurable instead of all definable sets; because, even if a set is measurable, its image under a projection need not be measurable. This was famously mistaken to be true by Lebesgue, but there are examples showing that a set can be measurable while its images under projections are not.

7. Our definition slightly differs from the original definition in Terwijn [17]. We require more sets to be measurable in our ε-models than in the original definition, where the measurability condition was moved to the truth definition. However, we need this stronger requirement on our models to be able to prove anything worthwhile — in fact, a stronger requirement is already necessary for most proofs published in earlier papers.

   An alternative (weaker) possibility would be not to require the relations to be measurable. This is, however, less natural, as discussed in Kuyper and Terwijn [13]. Nonetheless, we remark that we use this property in only one theorem (Theorem 5.2.10); all other theorems hold without it. We even have that compactness fails with this requirement on relations (Theorem 5.4.1), while compactness does hold if we do not impose it (Theorem 5.4.11).

Example 2.1.4. Let Q be a unary predicate. Then the sentence φ defined by $\forall x(Q(x)) \vee \forall x(\neg Q(x))$ is a $\frac{1}{2}$-tautology; namely, in every $\frac{1}{2}$-model, either the set on which Q holds or its complement has measure at least $\frac{1}{2}$. However, it is not an ε-tautology for $\varepsilon < \frac{1}{2}$.

Furthermore, both φ and $\neg\varphi$ are classically satisfiable and hence ε-satisfiable for every ε; in particular we see that φ can be an ε-tautology while simultaneously $\neg\varphi$ is ε-satisfiable (cf. item 3 above).

In many of our proofs, we will want our formulas to be in prenex normal form, so that we can state our arguments more concisely. That we can bring every formula into prenex normal form is therefore expressed in the next proposition.

Proposition 2.1.5 (Terwijn [17]). *Every formula φ is semantically equivalent to a formula φ' in prenex normal form; that is, φ' satisfies $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi \Leftrightarrow (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi'$ for all $\varepsilon \in [0, 1]$ and all ε-models $(\mathcal{M}, \mathcal{D})$.*

Proof. First, use Definition 2.1.1 to replace every implication by a disjunction and a negation. Furthermore, we can use it to push all negations inward. Next, we pull al quantifiers outside. Since the interpretation of $\exists$ is classical, we already know that we can pull it outside. For the universal quantifier, we need to check that

$$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi \vee \forall x(\psi(x)) \Leftrightarrow (\mathcal{M}, \mathcal{D}) \models_\varepsilon \forall x(\varphi \vee \psi(x))$$

and

$$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi \wedge \forall x(\psi(x)) \Leftrightarrow (\mathcal{M}, \mathcal{D}) \models_\varepsilon \forall x(\varphi \wedge \psi(x)),$$

where we assume that x is not free in φ. For the first one, we have:

$$\begin{aligned}(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi \vee \forall x(\psi(x)) &\Leftrightarrow (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi \text{ or } \Pr_{\mathcal{D}}[a \in \mathcal{M} \mid \psi(a)] \geq 1 - \varepsilon \\ &\Leftrightarrow \Pr_{\mathcal{D}}[a \in \mathcal{M} \mid \varphi \vee \psi(a)] \geq 1 - \varepsilon \\ &\Leftrightarrow (\mathcal{M}, \mathcal{D}) \models_\varepsilon \forall x(\varphi \vee \psi(x)).\end{aligned}$$

The other statement can be checked analogously. □

2.2 Probabilistic Truth and Tautologies

As an introduction to the subject, it is interesting to compare our notion of truth to the classical case and the intuitionistic case, for which we will use the next proposition.

Proposition 2.2.1 (Terwijn [17]). *Let $\varphi = \varphi(x)$ and let $(\mathcal{M}, \mathcal{D})$ be an ε-model. Then the sets $Y := \{a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(a)\}$ and $N := \{a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg\varphi(a)\}$ satisfy $Y \cup N = \mathcal{M}$, but not necessarily $Y \cap N = \emptyset$.*

Proof. We prove this by induction on the structure of φ. The only non-classical and therefore non-trivial case is that of the quantifiers; since negation exchanges the universal and existential quantifier we may suppose our quantifier to be universal. Therefore, suppose that $\varphi(x) = \forall y(\psi(x, y))$ and suppose that $a \notin Y$. Then

$$\Pr_{\mathcal{D}}[b \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a, b)] < 1 - \varepsilon,$$

and therefore the induction hypothesis tells us that

$$\Pr_{\mathcal{D}}[b \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg\psi(a, b)] > \varepsilon.$$

In particular, there exists a $b \in \mathcal{M}$ such that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon (\neg\psi(a, b))$. But this exactly says that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg\varphi(a)$, so $a \in N$. □

Using this proposition, we can make a nice comparison to classical and intuitionistic logic, as announced above. In the classical case, the sets Y and N are disjoint and satisfy $Y \cup N = \mathcal{M}$. In intuitionistic logic, Y and N are still disjoint but do not necessarily satisfy $Y \cup N = \mathcal{M}$. In our probabilistic interpretation, we do have $Y \cup N$, but instead the sets Y and N need not be disjoint (because of the paraconsistency discussed above).

We will now turn our attention towards the ε-tautologies. The 0-tautologies form an interesting first case to discuss. First observe that 0-truth does not coincide with classical truth: it only says that a proposition holds almost everywhere (i.e. on a set of measure 1), but it allows counterexamples of measure 0. However, we do have that the 0-tautologies coincide with the classical tautologies, as expressed in the next proposition.

Proposition 2.2.2 (Terwijn [17]). *The* 0*-tautologies coincide with the classical tautologies.*

Proof. (Sketch, for more details see Terwijn [17].) That every classical tautology is a 0-tautology is directly verified from the definitions. For the converse, observe that if φ is not a classical tautology, then there exists a countable countermodel $\mathcal{M}$. We can define a measure $\mathcal{D}$ on this model giving every point non-zero measure. One can then verify that $(\mathcal{M}, \mathcal{D}) \not\models_\varepsilon \varphi$. □

We can also express a relation between the ε-tautologies and the ε'-tautologies for different ε and ε'. The original proof in Terwijn [17] does not take the measurability condition we imposed on the definable sets of our ε-models into account, but (as mentioned in the same paper) we can easily modify the proof to respect this condition.

Theorem 2.2.3 (Terwijn [17]). *For all $0 \leq \varepsilon < \varepsilon' \leq 1$, the ε-tautologies are strictly included in the ε'-tautologies.*

Proof. The inclusion follows directly from Definition 2.1.1, since case 6 gets weaker if ε becomes bigger. For the strictness, first observe that we may assume $\varepsilon' < 1$, since otherwise we are in a degenerate case. Furthermore, since the rationals are dense in the reals we may assume ε and ε' to be rational.

Thus, let $\varepsilon' = 1 - \frac{m}{n}$. We introduce new unary predicates $X_1, \ldots, X_n$ and define the sentence n-split by

$$\forall x((X_1(x) \vee \ldots \vee X_n(x)) \wedge \bigwedge_{1 \leq i < j \leq n} \neg(X_i(x) \wedge X_j(x)))$$

Then, since the formula n-split is universal, if a probabilistic model $(\mathcal{M}, \mathcal{D})$ satisfies $(\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon'} \neg n\text{-split}$, we have that n-split holds classically. We now let φ be the sentence

$$\neg n\text{-split} \vee \bigvee_{i_1, \ldots, i_m} \forall x \left(\bigvee_{1 \leq j \leq m} X_{i_j}(x) \right).$$

Then, if $(\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon'} \neg n\text{-split}$, we see that $\mathcal{M}$ splits into n disjoint parts. By taking the m largest of these, we see that $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon'} \varphi$ and φ is therefore an ε'-tautology.

However, if we take the model $(\mathcal{M}, \mathcal{D})$ based on the set $\{1, \ldots, n\}$ where we let each X_i hold exactly on i and we let $\mathcal{D}$ be the uniform distribution, we see that $(\mathcal{M}, \mathcal{D}) \not\models_\varepsilon \varphi$. Thus, we see that φ is not an ε-tautology. □

2.3 PAC-Learning Probabilistic Sentences

We now return to our initial motivation for introducing this logic: namely, that we can induce facts accurately from a finite sample. We follow the approach from Terwijn [17]. However, the proof given there is partially incorrect; we will discuss the problem below and we will give a corrected proof.

Before we can say something useful, we will need to introduce the notion of a *sampling oracle*.

Definition 2.3.1. A **sampling oracle** $\mathrm{EX}(\mathcal{M}, \mathcal{D})$ for an ε-model $(\mathcal{M}, \mathcal{D})$ is an oracle which, when called upon, draws a random element from $\mathcal{M}$ according to $\mathcal{D}$. Furthermore, when supplied with a sample of elements and an atomic formula, the oracle tells us if the atomic formula holds for this particular sample (observe that atomic formulas may contain constants and function symbols).

In our definition of learning below we will have two parameters: an error parameter ε (corresponding to the ε in our logic) and a confidence parameter δ. We will usually take these to be small numbers from $(0, 1]$.

Definition 2.3.2. A (probabilistic) algorithm L **PAC-learns a sentence** φ if L, given an error parameter $\varepsilon > 0$ and a confidence parameter $\delta > 0$, for every unknown ε-model $(\mathcal{M}, \mathcal{D})$, and with access to a sampling oracle $\mathrm{EX}(\mathcal{M}, \mathcal{D})$, L outputs one of the possibilities $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi$ or $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg\varphi$ such that with probability at least $1 - \delta$ the output is correct; that is, if L draws a sample of size n, then

$$\Pr_{\mathcal{D}^n}[s \in \mathcal{M}^n \mid L \text{ gives a correct answer when supplied with } s] \geq 1 - \delta.$$

Remember that *both* possible outputs can be correct, as discussed below Definition 2.1.1.

This definition is inspired by, but not the same as, Valiant's [19] much-studied definition of Probably Approximately Correct learning (PAC-learning). PAC-learning concerns forming, with high probability, a hypothesis that approximates a concept. More precisely, let C be a set of subsets of some set X (called the *concepts*). We consider algorithms that have access to a sampling oracle $\mathrm{EX}(c, \mathcal{D})$ that, for an arbitrary concept $c \in C$ and an arbitrary probability measure on X, gives us samples from X according to $\mathcal{D}$ and tells us if $x \in c$.

We now say that C is *PAC-learnable* if there exists some algorithm $L(\varepsilon, \delta)$ polynomial in $\frac{1}{\varepsilon}$ and $\frac{1}{\delta}$, using $\mathrm{EX}(c, \mathcal{D})$, such that for all $\varepsilon, \delta > 0$, every unknown probability distribution $\mathcal{D}$ on X and every unknown concept $c \in C$ the algorithm outputs with probability at least $1 - \delta$ a *hypothesis* $h \in C$ such that $\Pr_{\mathcal{D}}[h \triangle c] < \varepsilon$.

Clearly, the δ in our definition can be linked to the Valiant's δ and stands for the *Probably*. With regard to the *Approximately* represented by the ε, our definitions do not connect as clearly. However, in both definitions the ε clearly represents some notion of approximation — be it approximate truth or approximation in measure. Therefore, there is a clear connection between our probability logic and PAC-learning.

To make this connection more precise, we turn towards the following theorem from Terwijn [17].

Theorem (incorrect). *There exists an algorithm L that PAC-learns any sentence φ.*[1] *If φ has n quantifiers, L takes a sample of size* $\left(\frac{1}{\varepsilon}\ln\frac{n}{\delta}\right)^n$.

While the algorithm described in this paper (and the first statement of the theorem) is correct, the proof of its correctness is incorrect. The last part of the proof contains incorrect reasoning with contrapositives in a probabilistic environment. The lower bound on the sample size is incorrect, as we will show after explaining the algorithm. Fortunately, we can fix these problems if we take a slightly larger sample, as we will demonstrate next. We wish to emphasise that the lower bound remains polynomial in $\frac{1}{\varepsilon}$ and $\frac{1}{\delta}$.

Theorem 2.3.3. *There exists an algorithm L that PAC-learns any sentence φ. If φ has n quantifiers, L takes a sample of size at most* $\left(\frac{1}{\varepsilon^2}\frac{1}{\delta}(2n)!\right)^{n+1}$. *In particular, L is polynomial in $\frac{1}{\varepsilon}$ and $\frac{1}{\delta}$.*

Proof. We use the same algorithm as in Terwijn [17], but slightly alter the proof. The main idea behind this algorithm is to decide existential quantifiers by taking a large enough sample and looking for a witness in it, and likewise one can decide universal quantifiers by taking a large enough sample and looking if it contains any counterexamples. For a given, fixed number of quantifiers one can then iterate this idea and compute how large a sample one needs to take to make the decision accurately.

More precisely, consider any sentence φ. By Proposition 2.1.5, we may assume φ to be in prenex normal form, say

$$\varphi = Qx_1Q'x_2Qx_3Q'x_4\dots Q''X_n\psi(x_1,\dots,x_n).$$

Let $m \in \omega$; this will represent the number of values we sample per quantifier. Consider the algorithm defined as follows:

[1] This first part of the statement is correct; it is the second part that is incorrect.

1. First, take a **sample for** φ: that is, for each sampled value of x_i separately, sample m values for x_{i+1} from $\mathcal{M}$ according to $\mathcal{D}$. Thus, in total we take a sample of size $m + m^2 + \cdots + m^n$ — namely, we have m values for x_1, for each of these values we have m values for x_2 (for a total of m^2), and so on. Call this sample S.

2. Determine if S **satisfies** φ. We define this inductively as follows:

   - S satisfies $\psi(s_1, \ldots, s_n)$ if $\mathcal{M} \models \psi(s_0, \ldots, s_n)$.
   - If $Q = \exists$, then S satisfies φ if there exists a value s for x_1 in S such that the corresponding sample for $x_2, \ldots, x_n$ satisfies $\forall x_2 \exists x_3 \ldots Q'' X_n \psi(x_1, \ldots, x_n)$.
   - If $Q = \forall$, then S satisfies φ if for all values s for x_1 in S we have that the corresponding sample for $x_2, \ldots, x_n$ satisfies $\exists x_2 \forall x_2 \ldots Q'' X_n \psi(x_1, \ldots, x_n)$.

3. If S satisfies φ, output φ; otherwise, output $\neg\varphi$.

We will henceforth assume that $Q = \exists$ and $Q' = \forall$, since otherwise we can perform the same decision procedure on $\neg\varphi$. In order to prove the probable correctness of this algorithm, we will first prove the following claim.
Claim: If $m \geq \frac{1}{\varepsilon} \ln \frac{1}{\delta}$, then the algorithm described above gives a correct output with probability at least $(1-\delta)^{m^n}$.

We prove this claim by induction on n.

- n=0: In the base case, all atomic truths about atomic formulas are given, from which we can easily decide which of φ and $\neg\varphi$ holds with correctness 1.

- n+1: We distinguish three cases: either $(\mathcal{M}, \mathcal{D}) \not\models_\varepsilon \varphi$ (and then $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg\varphi$ by Proposition 2.2.1), $(\mathcal{M}, \mathcal{D}) \not\models \neg\varphi$, or $(\mathcal{M}, \mathcal{D})$ models both of them. Write $\varphi = \exists x_1(\varphi'(x_1))$.

  – $(\mathcal{M}, \mathcal{D}) \not\models_\varepsilon \varphi$: then we have (classically) for every element $a \in \mathcal{M}$ that $(\mathcal{M}, \mathcal{D}) \not\models_\varepsilon \varphi'(a)$. So, by induction hypothesis we have that for every sampled value s for x_1 with probability at least $(1-\delta)^{m^{n-1}}$ that the corresponding sample for $x_2, \ldots, x_n$ does not satisfy φ'. So, with probability at least

$$\left((1-\delta)^{m^{n-1}}\right)^m = (1-\delta)^{m^n}$$

we have that none of the m sampled values for x_1 work, i.e. that S does not satisfy φ; so with probability at least $(1-\delta)^{m^n}$ the (correct) answer $\neg\varphi$ is outputted.

- $(\mathcal{M},\mathcal{D}) \not\models_\varepsilon \neg\varphi$: Then we know that

$$\Pr_{\mathcal{D}}[a \in \mathcal{M} \mid (\mathcal{M},\mathcal{D}) \models_\varepsilon \neg\varphi'(a)] < 1-\varepsilon.$$

So the probability that all of the sampled values for x_1 are in this set is

$$< (1-\varepsilon)^m \leq (e^{-\varepsilon})^m \leq \delta$$

where the last inequality follows per assumption on m. So, with probability $\geq 1-\delta$ we have for at least one of the values s for x_1 that $(\mathcal{M},\mathcal{D}) \not\models_\varepsilon \neg\varphi'(s)$. But in that case we have, by induction hypothesis, that the corresponding sample for $x_2,\ldots,x_n$ satisfies φ' with probability at least $(1-\delta)^{m^{n-1}}$. Therefore, we see that with probability at least

$$(1-\delta)^{m^{n-1}}(1-\delta) \geq (1-\delta)^{m^n}$$

the sample satisfies φ and therefore the correct answer φ is outputted.

- If $(\mathcal{M},\mathcal{D})$ models both φ and $\neg\varphi$, then both answers are correct so this case is trivial.

This proves the claim. To finish the proof, let $m \geq \frac{1}{\varepsilon^2}\frac{1}{\delta}(2n)!$. For this fixed m, we can repeat the above proof with $\frac{\delta}{m^n}$ instead of δ. To this end, we need to show that $m \geq \frac{1}{\varepsilon}\ln\frac{m^n}{\delta}$. But we have:

$$\begin{aligned} e^{\varepsilon m} &> \frac{(\varepsilon m)^{2n}}{(2n)!} \geq \frac{\left(\frac{1}{\varepsilon}\frac{1}{\delta}(2n)!\right)^n(\varepsilon m)^n}{(2n)!} = \frac{1}{\varepsilon}\frac{1}{\delta}\left(\frac{1}{\varepsilon}\frac{1}{\delta}(2n)!\right)^{n-1}(\varepsilon m)^n \\ &\geq \frac{1}{\varepsilon}\frac{1}{\delta}\left(\frac{1}{\varepsilon}\right)^{n-1}(\varepsilon m)^n = \frac{1}{\delta}m^n. \end{aligned}$$

So, we see that for this m the algorithm gives a correct output with probability at least $(1-\frac{\delta}{m^n})^{m^n}$. However, for every $k \in \omega$ we have that $(1-\frac{\delta}{k})^k \geq 1-\delta$, as can be seen from the binomial or Taylor expansion. Thus, for $m \geq \frac{1}{\varepsilon^2}\frac{1}{\delta}(2n)!$ we have that the algorithm gives a correct output with probability at least $1-\delta$, which finishes our proof. □

As promised above, we will now give a counterexample to the lower bound as originally given in Terwijn [17].

Example 2.3.4. Let $\varepsilon := \frac{1}{3}$, $\delta := \frac{1}{2}$ and $\varphi := \exists x \forall y(R(y))$. Consider the model $\mathcal{M}$ on $\{1, \dots, 450\}$ where R holds on $\{1, \dots, 299\}$ and $\mathcal{D}$ is the uniform distribution on $\mathcal{N}$.

Observe that, in this case, $5 > \frac{1}{\varepsilon} \ln \frac{n}{\delta} = 3 \ln 4$. So, we can take $m = 5$. For fixed $x \in \mathcal{M}$, we have that $\Pr_{\mathcal{D}}[y \in \mathcal{M} \mid R(y)] = \frac{299}{450}$. So the probability that a sample of size 5 for y satisfies R is $\left(\frac{299}{450}\right)^5$. Therefore, the probability that our sample for φ does not satisfy φ, which is the same as saying that all of our 5 samples of size 5 for y do not satisfy $\forall y(R(y))$, is equal to

$$\left(1 - \left(\frac{299}{450}\right)^5\right)^5 \approx 0.4998 < \frac{1}{2}.$$

Thus, the probability that our sample does satisfy φ is $> \frac{1}{2}$. However, then the algorithm outputs the incorrect answer φ with probability $> \frac{1}{2}$.

3 Earlier Approaches

3.1 Keisler's Probability Logic

The first known work on first-order models equipped with probability measures was done by H. Friedman, who studied a quantifier expressing that "there exists non-measure 0 many". He did not publish his work, but it is nonetheless exhibited in Steinhorn [16]. Since there are not many relations to our current work, we will not discuss this quantifier in more detail.

The first probability logic we will compare ours to is the one introduced by Keisler [10]. A slightly more recent survey of this logic can be found in Keisler [11], which will be our main reference for this section.

Keisler introduces his logic as an extension to $\mathcal{L}_{\omega_1\omega}$, in which countable conjunctions and disjunctions are allowed. In Keisler [11] and Hoover [7], the fragment with finite conjunctions and disjunctions is also discussed. It is this fragment that we will compare our logic to. We begin by introducing the formulas, which (unlike in our logic) are not the formulas of first-order logic.

Definition 3.1.1. The set of **formulas** of Keisler's probability logic $\mathcal{L}_{\omega P}$ is the least set such that:

1. Each atomic formula (of first-order logic) is a formula of $\mathcal{L}_{\omega P}$.

2. If φ is a formula of $\mathcal{L}_{\omega P}$, then $\neg\varphi$ is a formula of $\mathcal{L}_{\omega P}$.

3. If φ, ψ are formulas of $\mathcal{L}_{\omega P}$, then $\varphi \wedge \psi$, $\varphi \vee \psi$ and $\varphi \to \psi$ are formulas of $\mathcal{L}_{\omega P}$.

4. If $\varphi(\bar{x})$ is a formula of $\mathcal{L}_{\omega P}$ and $r \in [0, 1]$, then $(P\bar{x} \geq r)\varphi$ is a formula of $\mathcal{L}_{\omega P}$.

Definition 3.1.2. A **probability model** $(\mathcal{M}, \mathcal{D})$ consists of a first-order model $\mathcal{M}$ and a probability measure $\mathcal{D}$ such that all relations, functions and constants are measurable (cf. Definition 2.1.2).[1]

Definition 3.1.3. Let $\varphi = \varphi(x_1, \ldots, x_n)$ be a formula of $\mathcal{L}_{\omega P}$, let $(\mathcal{M}, \mathcal{D})$ be a probability model and $a_1, \ldots, a_n \in \mathcal{M}$. Then we inductively define **Keisler probabilistic truth** $(\mathcal{M}, \mathcal{D}) \models_K \varphi$ as follows.

1. Atomic formulas and logical connectives are treated as in first-order logic, e.g. $(\mathcal{M}, \mathcal{D}) \models_K \neg\varphi(\bar{a})$ if $(\mathcal{M}, \mathcal{D}) \not\models_K \varphi(\bar{a})$.
2. $(\mathcal{M}, \mathcal{D}) \models_K (P\bar{y} \geq r)\varphi(\bar{a}, \bar{y})$ if
$$\Pr_{\mathcal{D}^m}[b_1, \ldots, b_m \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_K \varphi(\bar{a}, \bar{b})] \geq r.$$

Definition 3.1.4. We will say that a sentence φ of $\mathcal{L}_{\omega P}$ is a **Keisler-tautology** or is **Keisler-valid** (notation: $\models_K \varphi$) if for all probability models $(\mathcal{M}, \mathcal{D})$ it holds that $(\mathcal{M}, \mathcal{D}) \models_K \varphi$.

Similarly, we call φ **Keisler-satisfiable** if there exists a probability model $(\mathcal{M}, \mathcal{D})$ such that $(\mathcal{M}, \mathcal{D}) \models_K \varphi$.

The first remark we make is that Keisler has the classical negation, unlike our (weaker) negation. Therefore, his logic is not suitable for learning processes — unlike ours, as discussed in Section 2.1.

Secondly, he allows the probability to vary per quantifier — while in our logic, ε is fixed for the entire formula. At first sight, this makes Keisler's probability logic stronger; however, he does not have a classical existential quantifier, which we do have. Quite remarkably, we will show in Section 4.1 that ε-validity is at least as computationally hard as Keisler-validity (in fact, we present a method to interpret variable ε' within ε-validity).

We also remark that the definition of probability model coincides with our definition of ε-model, in the sense that if all relations are measurable, then one can verify that all $\mathcal{L}_{\omega P}$-definable sets are also measurable (see Keisler [11], Theorem 1.2.5). For us it is not enough to require just the relations to be measurable, since this does not guarantee

[1] Keisler's definition of probability model does not completely coincide with our definition of an ε-model. In particular, he requires all singletons to be measurable and extends the product measure with diagonal sets, to make sure the equality relation is always measurable. In this paper, however, we will study his logic as if he used our definition of a probability model — the differences are small and this way we avoid having to treat more measure theory just to talk about $\mathcal{L}_{\omega P}$.

the image of projections to be measurable — since Keisler does not have the classical $\exists$, this is not a problem in $\mathcal{L}_{\omega P}$.

The reader has probably also observed that Keisler allows us to quantify over multiple variables at once and take the probability over the product measure. While both Keisler and Hoover reason that this is only to get a clean notation and that the n-ary quantifier can be defined using the unary quantifier, this only seems to hold for the infinitary case with countable conjunctions and disjunctions, unlike the case we study. Nonetheless, they also treat $\mathcal{L}_{\omega P}$ as having this n-ary quantifier, which is a further strengthening of the logic when compared to ours. The expressive strength of $\mathcal{L}_{\omega P}$ is demonstrated below.

Example 3.1.5. Let Q be a binary predicate and let

$$\varphi := (Px \geq \frac{2}{3})\neg((Py \geq 1)\neg Q(x,y)).$$

Since $\mathcal{L}_{\omega P}$ has the classical negation, φ expresses that the set of x such that $Q(x,y)$ occurs with strictly positive probability over y has measure at least $\frac{2}{3}$.

We cannot express this in our language, since we do not have the classical negation and can therefore not express that something occurs with probability > 0 (indeed, if this were the case, we would be unable to distinguish between something holding almost everywhere and its negation holding on a set of positive measure, so we would lose our PAC-learnability).

Surprisingly, we can recycle a lot of the ideas used in the proofs of various properties of $\mathcal{L}_{\omega P}$, with varying amounts of modifications. For example, the proof of the downwards Löwenheim-Skolem theorem in Section 5.1 uses part of the construction from Keisler [11], while our proof of the complexity of ε-validity in Section 4.1 uses ideas from a construction by Hoover [7].

Keisler and Hoover also study what they call *graded probability models*. Instead of requiring the n-ary relations to be $\mathcal{D}^n$-measurable, they require them to be measurable in some extension of $\mathcal{D}^n$ satisfying some extra properties (for example, that every section is $\mathcal{D}$-measurable, which for $\mathcal{D}^n$ would follow from Fubini's theorem). Graded probability models occur most eminently in the study of a completeness theorem for $\mathcal{L}_{\omega P}$ in which it is used as an intermediate step. We could also study graded probability models in our probability logic, but have decided not to do so, since we consider ε-models to be the most natural choice and did not pursue a proof-theoretic completeness theorem.

3.2 Valiant's Robust Logic

In [20], Valiant introduces a logic, which he calls a *robust logic*, that is also inspired by PAC-learning. We will present his logic in a slightly different way from the original presentation in [20], in order to more clearly illustrate the connections with and differences from our logic. We begin by introducing the models, which he calls *scenes*.

Definition 3.2.1. A **scene** over a finite set $\mathcal{A}$ for a set of relations $R = \{R_1, \ldots, R_t\}$ is a first-order model for these relations over the set $\mathcal{A}$. We denote $\Pi_{\mathcal{A},R}$ for all scenes over $\mathcal{A}$ and R.

Thus, Valiant only considers finite models. This shows a clear contrast with our approach: we consider models of arbitrary cardinality and restrict ourselves when we need to. Something else that is different from our approach is that he does not consider probability distributions over the model. Instead, he fixes a set $\mathcal{A}$ and assumes a probability distribution over the scenes $\Pi_{\mathcal{A},R}$. We will return to this later, after we have given the complete definition of the system.

We now move to the formulas, which he calls *rules*.

Definition 3.2.2. The set of **rules** of Valiant's logic are of the form:

$$\forall x_1, \ldots, x_s[f(e_1(R_{i_1}), \ldots, e_k(R_{i_k})) \equiv R_{i_0}(x_1, \ldots, x_s)]$$

where:

- Each $e_i(R_{i_1})$ is a first-order formula containing only quantifiers and the relation R_{i_1}, which only has free variables among $x_1, \ldots, x_s$.
- f represents a Boolean function $f : \{0,1\}^k \to \{0,1\}$.

The idea of such a rule is that it expresses that some expression involving a given Boolean function f is equivalent to the relation R_{i_0} for *many* of the scenes over some fixed $\mathcal{A}$, according to some given probability distribution $\mathcal{D}$ on $\Pi_{\mathcal{A},R}$. An easy example of a rule is

$$\exists y(R_1(y, x)) \vee \forall y_1 \exists y_2(R_2(y_1, y_2, x)) \equiv R_3(x)$$

(where the function f is disjunction).

Observe that Valiant only allows relations in his language, while we do not have any restrictions (except in theorems where this is necessary, but we do not restrict ourselves until we need to).

Finally, we introduce when a rule holds. More precisely, we say what it means for a rule to be *ε-accurate*.

Definition 3.2.3. Let $\mathcal{A}$ be a fixed finite set. Let q be a rule with free variables $x_1, \ldots, x_s$, let $\sigma \in \Pi_{\mathcal{A},R}$ be a scene and let $a_1, \ldots, a_s \in \mathcal{A}$. We define the function $lhs(q, \sigma, a_1, \ldots, a_s)$ by letting it have value 1 if the left-hand side of the equivalence in q holds when evaluated in the scene σ and with $a_1, \ldots, a_s$ substituted for the variables $x_1, \ldots, x_s$, and value 0 otherwise. We define $rhs(q, \sigma, s_1, \ldots, a_s)$ similarly for the right-hand side.

We now define the **false positive** and **false negative** functions err^+ and err^- by letting $err^+(q, \sigma, a_1, \ldots, a_s)$ have value 1 if we have $lhs(q, \sigma, a_1, \ldots, a_s) = 1$ but $rhs(q, \sigma, a_1, \ldots, a_s) = 0$, and defining err^- with the roles of lhs and rhs exchanged.

Finally, we define the **error** $e_D(q)$ of a rule q according to the distribution $\mathcal{D}$ on $\Pi_{\mathcal{A},R}$ by

$$\max_{a_1,\ldots,a_s} \sum_{\sigma \in \Pi_{\mathcal{A},R}} \Pr_{\mathcal{D}}[\sigma](err^+(q, \sigma, a_1, \ldots, a_s) + err^-(q, \sigma, a_1, \ldots, a_s)).$$

Now a rule q is called **ε-accurate** in the distribution D if $e_D(q) \leq \varepsilon$.

This shows that Valiant's logic is vastly different from ours: Valiant has a probability distribution over the models instead of over the elements of a fixed model. Furthermore, instead of talking about the approximate truth of statements, it talks about the approximate truth of rules.

Nonetheless, there are parallels — most notably, Valiant's logic has a PAC-learning property, although it is different from ours. In [20] Valiant shows that one can PAC-learn the function f (with suitable restrictions on the class of functions considered); that is, given a rule with an unknown Boolean function f and an unknown distribution $\mathcal{D}$ over the scenes, there exists a probabilistic algorithm that with high probability gives us an approximation f' such that the rule q with f' substituted for f is ε-accurate in $\mathcal{D}$.

So, the logic given above is also motivated by PAC-learning, but instead of trying to learn if a statement holds, it is about learning a Boolean function. While the motivation of our logic might be closer to Valiant than to Keisler, it should be clear that our logic is more like $\mathcal{L}_{\omega P}$ than it is like Valiant's logic.

4

Computability Theory

Assumption *In this chapter, we will only be looking at models* $(\mathcal{M}, \mathcal{D})$ *that are* ε*-models for all rational* ε*. For example, in this chapter we will call* φ *an* ε*-tautology if and only if for all* $(\mathcal{M}, \mathcal{D})$ *which are an* ε'*-model for all rational* ε' *we have that* $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi$*.*

4.1 Validity Is Π^1_1-Hard

In this section we will show that for every rational ε, the set of ε-tautologies is Π^1_1-hard.[1] This strengthens a result by Terwijn [18], in which it was shown that probability logic is Σ^0_1-hard.

We will first by showing that for different rational $\varepsilon_0, \varepsilon_1 \in (0, 1)$ the ε_0-tautologies many-one reduce to the ε_1-tautologies. We will begin with reducing to bigger ε_1. To do this, we refine the argument by Terwijn [18], where it is shown that the 0-tautologies many-one reduce to the ε-tautologies for $\varepsilon \in [0, 1)$.

Proposition 4.1.1. *Let* $\mathcal{L}$ *be a first-order language not containing equality. Then, for every rational* $0 \leq \varepsilon_0 \leq \varepsilon_1 < 1$*, the* ε_0 *tautologies many-one reduce to the* ε_1 *tautologies.*

Proof. Let $\frac{m}{n} = \frac{1-\varepsilon_1}{1-\varepsilon_0}$. Let φ be a formula in prenex normal form. We use induction over the structure of φ to define a many-one reduction f. For propositional formulas and existential quantifiers, there is nothing to be done and we use the identity map.

Next, we consider the universal quantifiers. Let $\varphi = \forall x(\psi(x))$. Our idea will be to introduce new unary predicates, so we can strengthen

[1] Π^1_1 consists of those sets $A \subseteq \omega$ such that A is definable by a Π^1_1-formula; that is, a formula $\forall Q(\varphi(Q, n))$ where Q is a second-order predicate variable and φ is a first-order formula.

the universal quantifier. We will make these predicates split the model into disjoint parts; if we split this into just the right amount of parts (in this case, n), then we can choose m of these parts to get just the right amount of strengthening.

So, we introduce new unary predicates $X_1, \dots, X_n$. We define the sentence n-split by

$$\forall x((X_1(x) \vee \ldots \vee X_n(x)) \wedge \bigwedge_{1 \le i < j \le n} \neg(X_i(x) \wedge X_j(x)))$$

Now, we define $f(\varphi)$ to be the formula

$$\neg n\text{-split} \vee \bigvee_{i_1,\dots,i_m} \forall x((X_{i_1}(x) \vee \cdots \vee X_{i_m}(x)) \wedge f(\psi)(x))$$

where the disjunction is over all subsets of size m from $\{1, \dots, n\}$. Thus, $f(\varphi)$ expresses that one can pick m of the n parts and that $\psi(x)$ holds often enough when restricted to these parts of the model.

We claim:

$$\varphi \text{ is an } \varepsilon_0\text{-tautology if and only if } f(\varphi) \text{ is an } \varepsilon_1\text{-tautology.}$$

Namely, first assume that φ is an ε_0-tautology and also assume that $(\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon_1} \neg n$-$\varphi$-split. Then the universal quantifiers hold classically, so $(\mathcal{M}, \mathcal{D})$ splits into n disjoint parts. We prove, by induction, that $f(\varphi)$ is an ε_1-tautology.

The only non-trivial case is the case for the universal quantifier. So, let $\varphi = \forall x(\psi(x))$. Then, we can find $i_1, \dots, i_m$ such that $X_{i_1}, \dots, X_{i_m}$ cover at least $\frac{m}{n}$ of the part of the model on which

$$(\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} \psi(x)$$

holds (by taking the m largest ones). Furthermore, we find (by induction hypothesis) that

$$\{a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} \psi(a)\} \subseteq \{a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} f(\psi)(a)\}.$$

Because φ is assumed to be an ε_0-tautology the left-hand side has measure $\geq 1 - \varepsilon_0$, so we find that

$$\Pr_{\mathcal{D}}[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} (X_{i_1}(a) \vee \cdots \vee X_{i_m}(a)) \wedge f(\psi)(a)] \geq \frac{m}{n}(1 - \varepsilon_0) = 1 - \varepsilon_1.$$

Therefore, $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} f(\varphi)$ and we thus find that $f(\varphi)$ is an ε_1-tautology.

Conversely, suppose $(\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon_0} \varphi$. Let $(\mathcal{M}', \mathcal{D}')$ be the model consisting of n copies $M_1, \ldots, M_n$ of $(\mathcal{M}, \mathcal{D})$, each with weight $\frac{1}{n}$. That is, $\mathcal{D}'$ is the sum of n copies of $\frac{1}{n}\mathcal{D}$ (see Definition 1.1.5). Relations are defined just as on $\mathcal{M}$; that is, for an n-ary relation R we define $R(a_1, \ldots, a_n)$ by viewing all a_i as elements of $\mathcal{M}$. The same holds for the arguments of functions; for the codomain we choose an arbitrary copy. Finally, we let each X_i be true exactly on the copy M_i.

Then we see that $(\mathcal{M}', \mathcal{D}') \not\models \neg n$-split. Assume $(\mathcal{M}', \mathcal{D}') \models_{\varepsilon_1} f(\varphi)$, we will show that this implies $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} \varphi$ and thus leads to a contradiction. Again, the only interesting case is the universal one. Thus, assume $\varphi = \forall x(\psi(x))$. Fix $i_1, \ldots, i_m$ such that

$$(\mathcal{M}', \mathcal{D}') \models_{\varepsilon_1} \forall x((X_{i_1}(x) \vee \cdots \vee X_{i_m}(x)) \wedge f(\psi)(x)).$$

Then, using the induction hypothesis, it is easily verified that

$$\begin{aligned}
&\Pr_{\mathcal{D}}[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} \psi(a)] \\
&= \frac{n}{m} \Pr_{\mathcal{D}}[a \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} (X_{i_1}(a) \vee \cdots \vee X_{i_m}(a)) \wedge f(\psi)(a)] \\
&\geq \frac{n}{m}(1 - \varepsilon_1) = 1 - \varepsilon_0
\end{aligned}$$

and we therefore see that $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} f(\varphi)$, which leads to the desired contradiction. Thus, $(\mathcal{M}', \mathcal{D}') \not\models_{\varepsilon_0} f(\varphi)$ so $f(\varphi)$ is not an ε_1-tautology, as desired. □

Next, we consider the reduction in the other direction.

Proposition 4.1.2. *Let $\mathcal{L}$ be a first-order language not containing equality. Then, for every rational $0 < \varepsilon_1 \leq \varepsilon_0 \leq 1$, the ε_0 tautologies many-one reduce to the ε_1 tautologies.*

Proof. Let $\frac{m}{n} = \frac{\varepsilon_0 - \varepsilon_1}{\varepsilon_0}$. Let φ be a formula in prenex normal form. We again use induction over the structure of φ to define a many-one reduction f. We only change the case for the universal quantifier. Let $\varphi = \forall x(\psi(x))$. Our idea will again be to introduce new unary predicates, but this time so we can weaken the universal quantifier.

Again introduce new unary predicates $X_1, \ldots, X_n$. We define $f(\varphi)$ to be the formula:

$$\neg n\text{-split} \vee \bigvee_{i_1, \ldots, i_m} \forall x(X_{i_1}(x) \vee \cdots \vee X_{i_m}(x) \vee f(\psi)(x))$$

where the disjunction is again over all subsets of size m from $\{1, \ldots, n\}$.

We claim:

φ is an ε_0-tautology if and only if $f(\varphi)$ is an ε_1-tautology.

Namely, first assume that φ is an ε_0-tautology and also assume that $(\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon_1} \neg n$-$\varphi$-split. The only non-trivial case is again the case for the universal quantifier. So, let $\varphi = \forall x(\psi(x))$. Then, we can find $i_1, \ldots, i_m$ such that $X_{i_1}, \ldots, X_{i_m}$ cover at least $\frac{m}{n}$ of the part of the model on which

$$(\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon_0} \psi(x)$$

holds. Because φ is assumed to be an ε_0-tautology we therefore find that

$$\Pr_{\mathcal{D}}[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} X_{i_1}(a) \vee \cdots \vee X_{i_m}(a) \vee f(\psi)(a)] \geq (1 - \varepsilon_0) + \varepsilon_0 \cdot \frac{m}{n} = 1 - \varepsilon_1.$$

So, $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} f(\varphi)$ and we therefore find that $f(\varphi)$ is an ε_1-tautology.

Conversely, suppose $(\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon_0} \varphi$. Let $(\mathcal{M}', \mathcal{D}')$ be the model consisting of n copies $M_1, \ldots, M_n$ of $(\mathcal{M}, \mathcal{D})$, as above. Assume that $(\mathcal{M}', \mathcal{D}') \models_{\varepsilon_1} f(\varphi)$, we will show that this leads to a contradiction. Again, the only interesting case is the universal one. Thus, assume $\varphi = \forall x(\psi(x))$. Fix $i_1, \ldots, i_m$ such that

$$(\mathcal{M}', \mathcal{D}') \models_{\varepsilon_1} \forall x(X_{i_1}(x) \vee \cdots \vee X_{i_m}(x) \vee f(\psi)(x)).$$

Then, using the induction hypothesis, it is easily verified that

$$\begin{aligned}&\Pr_{\mathcal{D}}[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} \psi(a)] \\ &= \Pr_{\mathcal{D}}[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} X_{i_1}(a) \vee \cdots \vee X_{i_m}(a) \vee f(\psi)(a)] \\ &\quad - \frac{m}{n}(1 - \Pr_{\mathcal{D}}[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} \psi(a)])\end{aligned}$$

and a straightforward calculation then shows that

$$\Pr_{\mathcal{D}}[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} \psi(a)] \geq 1 - \varepsilon_0,$$

which leads to the desired contradiction. □

Combining the two propositions above, we come to the following conclusion.

Theorem 4.1.3. *Let $\mathcal{L}$ be a first-order language not containing equality. Then, for rational $\varepsilon_0, \varepsilon_1 \in (0,1)$, the ε_0-tautologies many-one reduce to the ε_1-tautologies.*

Remark 4.1.4. Observe that we can perform these reductions per quantifier. In particular, we can talk about what it means for a formula with variable ε (for each quantifier separately) to be a tautology. This way, we get something like Keisler's probability logic (see Section 3.1); however, remember that we still have our non-classical negation (unlike Keisler).

To show that the set of ε-tautologies is indeed Π^1_1-hard, we adapt a proof by Hoover [7] which shows that $\mathcal{L}_{\omega P}$ is Π^1_1-complete. To do this, he shows that we can define the natural numbers within probability logic. We will present our proof after the next definition.

Definition 4.1.5. Let φ be a formula in prenex normal form and N a unary predicate. Then φ^N, or φ **relativised to** N, is the formula where each $\forall x(\psi(x))$ is replaced by $\forall x(N(x) \to \psi(x))$ and each $\exists x(\psi(x))$ is replaced by $\exists x(N(x) \wedge \psi(x))$.

Proposition 4.1.6. *There exists finite theories T, T' in the language with constant symbol 0, unary relations $N(x)$, binary relations $x \leq y$, $x = y$,[2] $S(x) = y$ and $R(x,y)$, and ternary relations $x + y = z$ and $x \cdot y = z$ such that, if we let f be the reduction from 0-tautologies to $\frac{1}{2}$-tautologies from Proposition 4.1.1, for every first-order formula $\varphi(n)$, containing a new predicate symbol Q, the following are equivalent:*

1. $\models_{\frac{1}{2}} f(\neg(\bigwedge T)) \vee \neg(\bigwedge T') \vee f(\neg\varphi(S^n(0)))$;

2. $\mathbb{N} \models \forall Q(\neg\varphi(Q, n))$.

Proof. We give an adaptation of the proof by Hoover [7] for $\mathcal{L}_{\omega P}$. We first remark that for every formula ψ we have that $(\mathcal{M}, \mathcal{D}) \not\models_0 \neg\psi$ if and only if all universal quantifiers hold classically and all existential quantifiers hold on a set of positive measure. Likewise, $(\mathcal{M}, \mathcal{D}) \not\models_{\frac{1}{2}} \neg\psi$ holds if and only if all universal quantifiers hold classically and all existential quantifiers hold on a set of measure strictly greater than $\frac{1}{2}$.

Inspired by this, we form the theories T and T'. T consists of Robinson's Q relativised to N, axioms specifying that the arithmetical

[2] Here we do not mean true equality, but rather a binary relation that we will use to represent equality.

relations almost only hold on N, and some special axioms (the last four of the enumeration below). Thus, we put the following axioms in T:

For all equality axioms ψ: ψ^N. We now give the axioms for the successor function:

$$\forall x \forall y (S(x) = y \rightarrow (N(x) \wedge N(y)))$$
$$(\forall x \exists y (S(x) = y))^N$$
$$(\forall x \forall y \forall u \forall v ((S(x) = y \wedge S(u) = v \wedge x = u) \rightarrow y = v))^N$$
$$(\forall x (\neg S(x) = 0))^N$$
$$(\forall x (x = 0 \vee \exists y (S(y) = x)))^N.$$[3]

In the axioms below, we will leisurely denote by $\psi(S(x))$ the formula $\forall y (S(x) = y \rightarrow \psi(y))$. Next up is the ordering:

$$(\forall x \forall y (x \leq y \rightarrow (N(x) \wedge N(y))))$$
$$(\forall x (x \leq x))^N$$
$$(\forall x \forall y (x \leq y \rightarrow y \leq x))^N$$
$$(\forall x \forall y \forall z (x \leq y \rightarrow y \leq z \rightarrow x \leq z))^N$$
$$(\forall x \forall y (x \leq y \vee y \leq x))^N$$
$$(\forall x (0 \leq x))^N$$
$$(\forall x \forall y (x \leq S(x) \wedge (x \leq y \wedge y \leq S(x)) \rightarrow (x = y \vee S(x) = y)))^N.$$

We proceed with the inductive definitions of + and $\cdot$:

$$(\forall x \forall y \forall z (x + y = z \rightarrow (N(x) \wedge N(y) \wedge N(z))))$$
$$(\forall x (x + 0 = x))^N$$
$$(\forall x \forall y (x + S(y) = S(x + y)))^N$$
$$(\forall x \forall y \forall z (x \cdot y = z \rightarrow (N(x) \wedge N(y) \wedge N(z))))^N$$
$$(\forall x (x \cdot 0 = 0))^N$$
$$(\forall x \forall y (x \cdot S(y) = (x \cdot y) + x))^N.$$

We also want to guarantee that N has positive weight:

$$\exists x (N(x))$$

Finally, we introduce a predicate R. This predicate is meant to function as a sort of 'padding'. The goal of this predicate is to force the measure of a point $S^n(0)$ to be equal to the measure of $\{x \mid N(x) \wedge x > S^n(0)\}$ (the precise use will be made clear in the proof below):

$$(\forall x \forall y(\neg R(x, y)))^N.$$

The last two axioms will be in T' instead of in T, since the idea is that this will be evaluated for $\varepsilon = \frac{1}{2}$ while the rest will be evaluated for $\varepsilon = 0$:

$$\forall x(N(x) \to \exists y(R(x, y) \vee x = y))$$
$$\forall x(N(x) \to \exists y(\neg(R(x, y) \vee x < y)).$$

We will now show that these axioms indeed to what we promised. First, let $n \in \omega$ and assume $\mathbb{N} \not\models \forall Q(\neg\varphi(Q, n))$. Fix a predicate Q such that $\mathbb{N} \not\models \neg\varphi(Q, n)$. Now take the model $M := \omega \times \{0, 1\}$ to be the disjoint union of two copies of ω, where we define $S, +, \cdot, \leq, 0$ as usual on $\omega \times \{0\}$ and undefined elsewhere. Let

$$N := \omega \times \{0\} \text{ and } R := \{((a, 0), (b, 1)) \mid \mu k[2^{k+1} > 3^{a+1}] \neq b\}.$$

We let $Q(a, 0)$ hold if $\mathbb{N} \models Q(n)$. Finally, define $\mathcal{D}$ by

$$\mathcal{D}(a, 0) = \mathcal{D}(a, 1) := \frac{1}{3^{a+1}}.$$

Since all points have positive measure, it is now directly verified that

$$(\mathcal{M}, \mathcal{D}) \not\models_0 \neg(\bigwedge T) \vee \neg\varphi(S^n(0)).$$

Furthermore, if we let $a \in \omega$ and denote b for $\mu k[2^{k+1} > 3^{a+1}]$ then we have that

$$\begin{aligned} &\Pr_{\mathcal{D}}[y \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models R((a, 0), y) \vee (a, 0) = y] \\ &= \frac{1}{2} - \frac{1}{2^{b+1}} + \frac{1}{3^{a+1}} \\ &> \frac{1}{2} \end{aligned}$$

[3] We do not strictly need this last axiom, but we have added it anyway so that all axioms of Robinson's Q are in T.

while we also have that

$$\begin{aligned}
&\Pr_{\mathcal{D}}[y \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\frac{1}{2}} \neg(R((a,0),y) \vee (a,0) < y)] \\
&= 1 - (\frac{1}{2} - \frac{1}{2^{b+1}} + \frac{1}{2 \cdot 3^{a+1}}) \\
&> \frac{1}{2}
\end{aligned}$$

so we see that $(\mathcal{M}, \mathcal{D}) \not\models_{\frac{1}{2}} \neg(\bigwedge T')$. But then we see from (the proof of) Theorem 4.1.1 that there is a model $(\mathcal{N}, \mathcal{E})$ such that

$$(\mathcal{N}, \mathcal{E}) \not\models_{\frac{1}{2}} f(\neg(\bigwedge T)) \vee \neg(\bigwedge T') \vee f(\neg\varphi(S^n(0))).$$

Conversely, assume that statement 1 does not hold. Again, from (the proof of) Theorem 4.1.1 we see that

$$(\mathcal{M}, \mathcal{D}) \not\models_0 \neg(\bigwedge T) \vee \neg\varphi(S^n(0)) \text{ and } (\mathcal{M}, \mathcal{D}) \not\models_{\frac{1}{2}} \neg(\bigwedge T').$$

From the three axioms involving R, we see that for every $m \in \mathcal{M}$ with $\mathcal{M} \models N(m)$:

$$\begin{aligned}
\Pr_{\mathcal{D}}[a \in \mathcal{M} \mid \mathcal{M} \models m = a] &> \frac{1}{2} - \Pr_{\mathcal{D}}[a \in \mathcal{M} \mid \mathcal{M} \models R(m,a)] \\
&> \Pr_{\mathcal{D}}[a \in \mathcal{M} \mid \mathcal{M} \models m < a].
\end{aligned}$$

But then we inductively see that

$$\Pr_{\mathcal{D}}[\{0, \ldots, S^k(0)\}] > \left(1 - \frac{1}{2^{k+1}}\right) \Pr_{\mathcal{D}}[N]$$

which shows that all weight of N rests on $X := \{S^n(0) \mid n \in \omega\}$. From the discussion at the beginning of this proof, we therefore easily that $(\mathcal{M}, \mathcal{D}) \not\models_0 \neg\varphi(S^n(0))$ implies that also $(\mathcal{M}\restriction X, \mathcal{D}\restriction X) \not\models_0 \neg\varphi^X(n)$.

However, we can directly verify that $\mathcal{M}\restriction X$ is isomorphic to the standard natural numbers $\mathbb{N} = (\omega, S, +, \cdot, \leq)$. So, by transferring the predicate Q from $\mathcal{M}$ to $\mathbb{N}$ we find that indeed $\mathbb{N} \not\models \forall Q(\neg\varphi(Q, n))$. □

Putting this together, we reach our conclusion.

Theorem 4.1.7. *For rational $\varepsilon \in (0,1)$, the set of ε-tautologies is Π^1_1-hard.*

In particular, we see that the ε-tautologies are not computably enumerable. This also implies that there is no effective calculus for our logic.

4.2 Satisfiability Is Π_1^0-Hard

In Section 4.1 above, it was shown that the set of ε-valid sentences is Π_1^1-hard for rational ε. Here, we will study the complexity of the satisfiability problem. Observe that, unlike in the classical case, satisfiability is not complementary to validity (classically, φ is satisfiable if and only if $\neg\varphi$ is not a tautology).

Example 4.2.1. The sentence $\varphi = \exists x(R(x)) \wedge \forall x(\neg R(x))$ is 0-satisfiable, but its negation $\forall x(\neg R(x)) \vee \exists x(R(x))$ is a 0-tautology (indeed, it is even a classical tautology).

As for validity, we first present some reductions between different ε. We will assume the reader has already read Section 4.1 and will therefore not present the proofs in full detail, since most of the ideas are the same as in the proof for validity.

Proposition 4.2.2. *Let $\mathcal{L}$ be a first-order language not containing equality. Then, for every rational $0 \leq \varepsilon_0 \leq \varepsilon_1 < 1$, ε_0-satisfiability many-one reduces to ε_1-satisfiability.*

Proof. Let $\varepsilon_1 = 1 - \frac{a}{n}$ and let $\frac{m}{n} = \frac{1-\varepsilon_1}{1-\varepsilon_0}$. Let φ be a formula in prenex normal form. As before, we use induction over the structure of φ to define a many-one reduction f, observing we only need to consider universal quantifiers.

Therefore, let $\varphi = \forall x(\psi(x))$. Again introduce new unary predicates $X_1, \ldots, X_n$. As before, we want a formula like n-split, but need to do slightly more work to make this work for satisfiability. For $1 \leq i \leq n$, define

$$Y_i(x) := X_i(x) \wedge \bigwedge_{1 \leq j \leq n, j \neq i} \neg X_j(x).$$

Then the Y_i define disjoint sets.

We now define the sentence a-n-split by:

$$\bigwedge_{I \subseteq \{1,\ldots,n\}, \#I = a} \forall y \left(\bigvee_{i \in I} Y_i(y) \right).$$

Then one can verify that, if all of the sets X_i are disjoint sets of measure exactly $\frac{1}{n}$ (and hence the same holds for the Y_i), then a-n-split is ε_1-true. Conversely, we claim that if a-n-split holds, then the sets Y_i all have measure $\frac{1}{n}$.

Namely, assume there is a set of measure strictly less than $\frac{1}{n}$. Determine a sets Y_i with minimal measure; say with indices from the set I. Then, since a-n-split holds, $\Pr_{\mathcal{D}}[\bigcup_{i\in I} Y_i] \geq \frac{a}{n}$. But at least one of the Y_i with $i \in I$ has measure strictly less than $\frac{1}{n}$, so also one of them needs to have measure strictly greater than $\frac{1}{n}$. However, $\Pr_{\mathcal{D}}[\bigcup_{i\notin I} Y_i] \leq \frac{n-a}{n}$, so there is a Y_j with $j \notin I$ having measure $\leq \frac{1}{n}$. This contradicts the minimality. So, all sets Y_i have measure at least $\frac{1}{n}$ and since they are disjoint they therefore have measure exactly $\frac{1}{n}$.

In particular we see that, if a-n-split holds, then the Y_i together disjointly cover a set of measure 1.

Now define $f(\varphi)$ to be the formula

$$a\text{-}n\text{-split} \wedge \bigwedge_{i_1,\ldots,i_m} \forall x((Y_{i_1}(x) \vee \cdots \vee Y_{i_m}(x)) \wedge f(\psi)(x))$$

where the conjunction is over all subsets of size m from $\{1,\ldots,n\}$.

If φ is ε_0-satisfiable by some model $(\mathcal{M},\mathcal{D})$, we can form the model consisting of n copies, each with weight $\frac{1}{n}$ (as in the case for tautologies) and check that this model ε_1-satisfies $f(\varphi)$. Conversely, assume $f(\varphi)$ is ε_1-satisfied by some model $(\mathcal{M},\mathcal{D})$, then we can inductively show that this same model also ε_0-satisfies φ. Namely, assume it does not. By induction hypothesis, we then have:

$$\Pr_{\mathcal{D}}[x \in \mathcal{M} \mid (\mathcal{M},\mathcal{D}) \models_{\varepsilon_1} f(\psi)(x)] < 1 - \varepsilon_0.$$

But by taking those m of the Y_i (say $Y_{i_1},\ldots,Y_{i_m}$) which have the smallest intersection with this set, we find that

$$\Pr_{\mathcal{D}}[x \in \mathcal{M} \mid (\mathcal{M},\mathcal{D}) \models_{\varepsilon_1} (Y_{i_1} \vee \cdots \vee Y_{i_m}) \wedge f(\psi)(x)] < \frac{m}{n}(1-\varepsilon_0) = 1-\varepsilon_1$$

which contradicts our choice of $(\mathcal{M},\mathcal{D})$. □

Proposition 4.2.3. *Let $\mathcal{L}$ be a first-order language not containing equality. Then, for every rational $0 < \varepsilon_1 \leq \varepsilon_0 \leq 1$, ε_0-satisfiability many-one reduces to ε_1-satisfiability.*

Proof. Let $\varepsilon_1 = 1 - \frac{a}{n}$ and let $\frac{m}{n} = \frac{\varepsilon_0 - \varepsilon_1}{\varepsilon_0}$. Again, we only consider the non-trivial case where φ is a universal formula $\forall x(\psi(x))$. We define $f(\varphi)$ to be the formula:

$$a\text{-}n\text{-split} \wedge \bigwedge_{i_1,\ldots,i_m} \forall x(Y_{i_1}(x) \vee \cdots \vee Y_{i_m}(x) \vee f(\psi)(x))$$

where the conjunction is again over all subsets of size m from $\{1, \ldots, n\}$. As above, one can check that if $(\mathcal{M}, \mathcal{D})$ ε_0-satisfies φ, that we can take a model consisting of n copies which will ε_1-satisfy $f(\varphi)$. Conversely, assume that $(\mathcal{M}, \mathcal{D})$ ε_1-satisfies $f(\varphi)$ and assume that

$$\Pr_{\mathcal{D}}[x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} f(\psi)(x)] < 1 - \varepsilon_0.$$

By taking those m of the Y_i (say $Y_{i_1}, \ldots, Y_{i_m}$) which have the largest intersection with this set, we find that

$$\begin{aligned}
&\Pr_{\mathcal{D}}[x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} Y_{i_1} \vee \cdots \vee Y_{i_m} \vee f(\psi)(x)] \\
&\leq \Pr_{\mathcal{D}}[x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} f(\psi)(x)] \\
&\quad + \frac{m}{n}(1 - \Pr_{\mathcal{D}}[x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} f(\psi)(x)])
\end{aligned}$$

and a straightforward calculation then shows that

$$\Pr_{\mathcal{D}}[x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} \psi(x, b_1, \ldots, b_t)] < 1 - \varepsilon_1,$$

a contradiction. So, $(\mathcal{M}, \mathcal{D})$ also ε_0-satisfies φ. □

Theorem 4.2.4. *Let $\mathcal{L}$ be a first-order language not containing equality. Then, for rational $\varepsilon_0, \varepsilon_1 \in (0,1)$, ε_0-satisfiability many-one reduces to ε_1-satisfiability.*

Now we can interpret arithmetic, like we did for tautologies. However, this time we can only interpret a weaker fragment, namely the universal fragment.

Proposition 4.2.5. *There exists finite theories T, T' in the language with constant symbol 0, unary relations $N(x)$, binary relations $x \leq y$, $x = y$,[4] $S(x) = y$ and $R(x,y)$, and ternary relations $x + y = z$ and $x \cdot y = z$ such that, if we let f be the reduction from 0-satisfiability to $\frac{1}{2}$-satisfiability from Proposition 4.2.2, for every universal formula $\varphi = \varphi(n)$, the following are equivalent:*

1. $f(\bigwedge T) \wedge (\bigwedge T') \wedge \varphi(S^n(0))$ *is $\frac{1}{2}$-satisfiable;*
2. $\mathbb{N} \models \varphi(n)$.

[4]Here we do not mean true equality, but rather a binary relation that we will use to represent equality.

Proof. This is proven in a similar way as Proposition 4.1.6. We take T as in that proposition, but remove $\exists x(N(x))$ from it. We add the formulas $\forall x(N(x))$ and $\forall x(N(x))$ to T' (defining that N has weight exactly $\frac{1}{2}$), and replace the formulas

$$\forall x(N(x) \to \exists y(R(x,y) \vee x = y))$$
$$\forall x(N(x) \to \exists y(\neg(R(x,y) \vee x < y))$$

by

$$\forall x(N(x) \wedge \forall y(R(x,y) \vee x = y))$$
$$\forall x(N(x) \wedge \forall y(\neg(R(x,y) \vee x < y)).$$

One can then follow the proof of Proposition 4.1.6 with three minor modifications:

1. The measure $\mathcal{D}$ on $\omega \times \{0,1\}$ becomes
$$\mathcal{D}(a,0) = \mathcal{D}(a,1) := \frac{1}{2^{n+2}}.$$
2. We let
$$R := \{((a,0),(b,1)) \mid a \neq b\}.$$
3. All strict inequalities become non-strict inequalities. □

From this, we reach our conclusion.

Theorem 4.2.6. *For rational $\varepsilon \in (0,1)$, the set of ε-satisfiable sentences is Π^0_1-hard.*

Proof. Use the reduction from the previous proposition, combined with the fact that the universal fragment of arithmetic is Π^0_1-hard (for this last statement, see for example Odifreddi [15, p368]). □

5

Model Theory

5.1 A Downward Löwenheim-Skolem Theorem

In this section, we will be prove a downward Löwenheim-Skolem Theorem for probability logic. We begin with the necessary definitions.

Definition 5.1.1. We will call a measure ν on a σ-algebra $\mathcal{B}$ of subsets of N a **submeasure** of a measure μ on a σ-algebra $\mathcal{A}$ of subsets of some set $M \supseteq N$ if for every $B \in \mathcal{B}$ there exists an $A_B \in \mathcal{A}$ such that $B = A_B \cap N$ and $\mu(A_B) = \nu(B)$.

Definition 5.1.2. An **ε-submodel** of an ε-model $(\mathcal{M}, \mathcal{D})$ is an ε-model $(\mathcal{N}, \mathcal{E})$ over the same language such that:

- $\mathcal{N}$ is a submodel of $\mathcal{M}$,
- $\mathcal{E}$ is a submeasure of $\mathcal{D}$.

We will denote this by $(\mathcal{N}, \mathcal{E}) \subset_\varepsilon (\mathcal{M}, \mathcal{D})$.

Definition 5.1.3. An **elementary ε-submodel** of an ε-model $(\mathcal{M}, \mathcal{D})$ is an ε-submodel $(\mathcal{N}, \mathcal{E})$ such that, for all formulas $\varphi = \varphi(x_1, \dots, x_n)$ and sequences $a_1, \dots, a_n \in \mathcal{N}$ we have:

$$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(a_1, \dots, a_n) \Leftrightarrow (\mathcal{N}, \mathcal{E}) \models_\varepsilon \varphi(a_1, \dots, a_n).$$

We will denote this by $(\mathcal{N}, \mathcal{E}) \prec_\varepsilon (\mathcal{M}, \mathcal{D})$.

Before proving a downward Löwenheim-Skolem Theorem, we first remark that we cannot hope to prove an exact analogue of the theorem in classical logic. This is because in the next example we will show that there are sentences which are satisfied by some uncountable model, but not by any countable model.

Example 5.1.4. Let $\varphi := \forall x \forall y (R(x,y) \wedge \neg R(x,x))$. Then φ is 0-satisfiable; for example, take the unit interval $[0,1]$ equipped with the Lebesgue measure and take $R(x,y)$ to be $x \neq y$.

However, it does not have any countable 0-models. Namely, if we have $(\mathcal{M}, \mathcal{D}) \models_0 \varphi$, then for almost every $x \in \mathcal{M}$ the set

$$B_x := \{y \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_0 \neg R(x,y) \vee R(x,x)\}$$

has measure zero. However, $x \in B_x$, so on one hand the set $\bigcup_{x \in \mathcal{M}} B_x$ contains almost every x and therefore has measure 1, but it is also the union of countable many sets of measure 0 and hence has measure 0, a contradiction.

Using the reduction from Proposition 4.2.2, we now also find for every rational $\varepsilon \in [0,1)$ a sentence φ_ε which is only ε-satisfiable in uncountable models.

The example has shown that we cannot always find countable elementary submodels. However, we can find such submodels of cardinality 2^ω, as we will show next. Our proof is inspired by the very briefly stated proof in Keisler [11]. We have made various modifications to be able to apply the construction to our logic and to be able to extract more information from the construction afterwards.

Theorem 5.1.5 (Downward Löwenheim-Skolem theorem for probability logic). *Let $\mathcal{L}$ be a countable first-order language, possibly containing equality, over a signature not containing function symbols. Let $(\mathcal{M}, \mathcal{D})$ be an ε-model and let $X \subseteq \mathcal{M}$ be of cardinality at most 2^ω. Then there exists*

$$(\mathcal{N}, \mathcal{E}) \prec_\varepsilon (\mathcal{M}, \mathcal{D})$$

such that $X \subseteq \mathcal{N}$ and $\mathcal{N}$ is of cardinality at most 2^ω.

Proof. The idea of the proof is to select one point from every *equivalent* part of the model: that is, a subset $Y \subseteq \mathcal{M}$ such that for all $x, y \in Y$ we have for all formulas $\varphi = \varphi(x)$ that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(x)$ holds if and only if $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(y)$ holds. We will show that we can do this in such a way that we need at most 2^ω many points.

Let $R = R(x_1, \ldots, x_n)$ be a relation. By Definition 2.1.2, we see that the set $R^{\mathcal{M}}$ is a $\mathcal{D}^n$-measurable set. Thus, Definition 1.1.8 and Proposition 1.1.10 tell us that $R^{\mathcal{M}}$ can be formed using countable unions and intersections of Cartesian products of at most countably many $\mathcal{D}$-measurable sets. This expression need not be unique — so, for each relation R, pick one such expression t and form the set Γ_R consisting of the $\mathcal{D}$-measurable sets occurring as edges of Cartesian products in this expression. Let Γ be $\bigcup \Gamma_R$ together with $\{c\}$ for each constant c.

Since Γ is countable, we can fix an enumeration $B_0, B_1, \ldots$ of it. For each $a \in 2^\omega$, we define:

$$E_a := \bigcap_{a_i=1} B_i \cap \bigcap_{a_i=0} (\mathcal{M} \setminus B_i).$$

Then the E_a are directly verified to be equivalent parts, in the sense that for every formula $\varphi = \varphi(x)$ and for all $x, y \in E_a$ we have that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(x)$ holds if and only if $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(y)$ holds. Namely, if R is a box $B_0 \times B_1$ of two $\mathcal{D}$-measurable sets, then we can verify this fact using induction over the structure of formulas; furthermore, we can also use induction over the formulas to check that this property is preserved under unions and complements.

From each non-empty E_a, pick one point x_a. Now let $\mathcal{N}$ be the set $X \cup \{x_a \mid a \in 2^\omega\}$. Clearly, $\mathcal{N}$ then has cardinality at most 2^ω.

Finally, for each $\mathcal{D}$-measurable B such that

$$\forall a \in 2^\omega \forall x, y \in E_a (x \in B \Leftrightarrow y \in B), \tag{5.1}$$

we let $\mathcal{E}(B \cap \mathcal{N}) := \mathcal{D}(B)$. We claim: $(\mathcal{N}, \mathcal{E})$ (with relations restricted to $\mathcal{N}$) satisfies the required properties.

First, observe that $\mathcal{E}$ is well-defined. Namely, let $B \neq C$ be $\mathcal{D}$-measurable sets satisfying (5.1). Without loss of generality we may assume there is some $x \in B$ with $x \notin C$. Let $a \in 2^\omega$ be such that $x \in E_a$. Then $x_a \in B$, but $x_a \notin C$. So $B \cap \mathcal{N} \neq C \cap \mathcal{N}$.

Next, we prove that $(\mathcal{N}, \mathcal{E}) \prec_\varepsilon (\mathcal{M}, \mathcal{D})$. We let φ be a formula in prenex normal form (using Proposition 2.1.5). We use induction over the number of quantifiers to show that, for all sequences $b_1, \ldots, b_n \in \mathcal{N}$ and for every formula $\varphi = \varphi(x_1, \ldots, x_n)$, we have

$$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(b_1, \ldots, b_n) \Leftrightarrow (\mathcal{N}, \mathcal{E}) \models_\varepsilon \varphi(b_1, \ldots, b_n).$$

The base case is clear.

For the existential case, observe that

$$(\mathcal{N}, \mathcal{E}) \models_\varepsilon \exists x(\psi(x, b_1, \ldots, b_n))$$

clearly implies that this also holds in $(\mathcal{M}, \mathcal{D})$. For the converse, assume

$$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \exists x(\psi(x, b_1, \ldots, b_n)).$$

Let $x \in \mathcal{M}$ be such that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(x, b_1, \dots, b_n)$, and let $a \in 2^\omega$ be such that $x \in E_a$. Then (since each E_a is an equivalent part, as described above) we also have $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(x_a, b_1, \dots, b_n)$. Using the induction hypothesis, we therefore find $(\mathcal{N}, \mathcal{E}) \models_\varepsilon \psi(x_a, b_1, \dots, b_n)$. Since $x_a \in \mathcal{N}$ this implies that $(\mathcal{N}, \mathcal{E}) \models_\varepsilon \exists x(\psi(x, b_1, \dots, b_n))$.

For the universal case, let $\varphi = \forall x(\psi(x, x_1, \dots, x_n))$. Let

$$B := \{x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(x, b_1, \dots, b_n)\}.$$

Then we we find that

$$\begin{aligned} C &:= \{x \in \mathcal{N} \mid (\mathcal{N}, \mathcal{E}) \models_\varepsilon \psi(x, b_1, \dots, b_n)\} \\ &= \{x \in \mathcal{N} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(x, b_1, \dots, b_n)\} \\ &= \{x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(x, b_1, \dots, b_n)\} \cap \mathcal{N} \\ &= B \cap \mathcal{N}. \end{aligned}$$

From this, we see that $\mathcal{E}(C) = \mathcal{D}(B)$, and therefore we see

$$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \forall x(\psi(x, b_1, \dots, b_n)) \Leftrightarrow (\mathcal{N}, \mathcal{E}) \models_\varepsilon \forall x(\psi(x, b_1, \dots, b_n)).$$

This finishes our induction.

Finally, we remark that we can now easily see that $(\mathcal{N}, \mathcal{E})$ is an ε-model (cf. Definition 2.1.2). For every formula $\varphi = \varphi(x_1, \dots, x_n)$ and every sequence $a_1, \dots, a_{n-1} \in \mathcal{N}$ we have:

$$\begin{aligned} B_\varphi &:= \{a_n \in \mathcal{N} \mid (\mathcal{N}, \mathcal{E}) \models_\varepsilon \varphi(a_1, \dots, a_n)\} \\ &= \{a_n \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(a_1, \dots, a_n)\} \cap \mathcal{N} \end{aligned}$$

and since the right-hand side is the intersection of a $\mathcal{D}$-measurable set and $\mathcal{N}$, it is easily verified that B_φ is $\mathcal{E}$-measurable. That relations are measurable follows directly from the construction; for constants c use the fact that $\{c\} \in \Gamma$ and therefore there exists an $a \in 2^\omega$ such that $E_a = \{c\}$.

Thus, we see that $(\mathcal{N}, \mathcal{E})$ is an elementary ε-submodel of $(\mathcal{M}, \mathcal{D})$. □

Remark 5.1.6. The proof given above uses the full measurability condition on ε-models (cf. Section 2.1, remark 7). We remark that we can also prove the theorem without using that the relations are measurable, by more closely following the proof of Keisler [11]. However, we need our proof to be able to derive Theorem 5.2.10 below.

In fact, by varying ε, we can easily see that the following strengthening also holds.

Theorem 5.1.7 (Downward Löwenheim-Skolem Theorem for variable ε). *Let $\mathcal{L}$ be a first-order language as above. Let $A \subseteq [0,1]$, let $(\mathcal{M}, \mathcal{D})$ be an ε-model for all $\varepsilon \in A$ and let $X \subseteq M$ be of cardinality at most 2^ω. Then there exists $(\mathcal{N}, \mathcal{E})$ such that $(\mathcal{N}, \mathcal{E}) \prec_\varepsilon (\mathcal{M}, \mathcal{D})$ for all $\varepsilon \in A$.*

Proof. Using the same proof as for Theorem 5.1.5. □

5.2 Satisfiability and Lebesgue measure

The construction from Theorem 5.1.5 gives us an unknown probability measure on 2^ω. However, we can say something more about the σ-algebra of measurable sets of $\mathcal{E}$ above: for example, that it is countably generated. We will use this and other facts to show that every ε-satisfiable set T of sentences has an ε-model on $[0,1]$ equipped with the Lebesgue measure. This model need not be equivalent to the original model satisfying the sentences in T — the Lebesgue model will generally satisfy more formulas.

We cannot directly show that the measure space is isomorphic to the Lebesgue measure on $[0,1]$ — we need to make some modifications to the model first. As a first step, we show that each set ε-satisfiable set T of sentences is satisfied in some Borel measure on the Cantor set 2^ω (equipped with the usual Euclidean topology). For this, we first need an auxiliary theorem.

Theorem 5.2.1. *Let $\mathcal{D}_0$ be a Borel probability measure and let $\mathcal{M}$ be a first-order model such that all relations and functions are $\mathcal{D}_0^n$-measurable*[1]*. Let $\mathcal{D}$ be the completion of $\mathcal{D}_0$. Then $(\mathcal{M}, \mathcal{D})$ is an ε-probability model for every $\varepsilon \in [0,1]$.*

Proof. What remains to be proven is that all definable sets of dimension 1 are $\mathcal{D}$-measurable (see Definition 2.1.2).

Since every relation is $\mathcal{D}_0^n$-measurable, it is in particular Borel and therefore analytic. We now verify that every definable set is analytic,

[1] Observe that this is strictly stronger than saying that they should be Borel, since the product of Borel measures might not be a Borel measure, see e.g. Bogachev [5, Section 7.6].

using induction over the number of quantifiers in prenex normal form (cf. Proposition 2.1.5). Clearly, this holds for propositional formulas. For the existential quantifier, use that projections of analytic sets are analytic (see e.g. Bogachev [4, Corollary 1.10.9]), and for the universal quantifier, this fact is exactly expressed by the Kondo-Tugue theorem (see e.g. Kechris [9, Theorem 29.26]).

In particular, we see that every definable set of dimension 1 is analytic. But then it is measurable in the completion $\mathcal{D}$ of $\mathcal{D}_0$ (see e.g. Bogachev [4, Theorem 1.10.5]). □

Proposition 5.2.2. *Let $\mathcal{L}$ be a countable first-order language not containing equality or function symbols. Let T be an ε-satisfiable set of sentences. Then there exists an ε-model $(\mathcal{M}, \mathcal{D})$ on 2^ω which ε-satisfies T such that $\mathcal{D}$ is the completion of a Borel measure. Furthermore, all relations are Borel.*

Proof. Fix a model ε-satisfying all sentences from T and apply Theorem 5.1.5 (with $X = \emptyset$) to find a model $(\mathcal{N}, \mathcal{E})$. Let $\Gamma = \{B_0, B_1, \dots\}$ and E_a be as in the proof of Theorem 5.1.5, i.e.:

$$E_a := \bigcap_{a_i=1} B_i \cap \bigcap_{a_i=0} (X \setminus B_i).$$

Then, per construction of $\mathcal{N}$, each such E_a contains at most one point (namely x_a). So, the function $\pi : \mathcal{N} \to 2^\omega$ sending each $x_a \in \mathcal{N}$ to a is injective.

Now, define subsets $C_n \subseteq 2^\omega$ by

$$C_n := \{a \in 2^\omega \mid a_n = 1\}.$$

Then $\{C_n \mid n \in \omega\}$ generate the Borel σ-algebra of 2^ω and we have $\pi^{-1}(C_n) = B_n$. Thus, C_n can be seen as an enlargement of B_n.

Next, let $R(x_1, \dots, x_n)$ be an n-ary relation. Write $R^{\mathcal{N}}$ as an expression using countable unions and intersections of Cartesian products of $\mathcal{E}$-measurable sets from Γ_R (cf. the definition of Γ_R in the proof of Theorem 5.1.5); say as the expression $t(B_0, B_1, \dots)$. Then we define $R^{\mathcal{M}}$ by $t(C_0, C_1, \dots)$. Furthermore, we define each constant $c^{\mathcal{M}}$ to be $\pi(c^{\mathcal{N}})$.

Finally, define a Borel probability measure $\mathcal{D}_0$ on 2^ω by

$$\mathcal{D}_0 := \mathcal{E} \circ \pi^{-1}.$$

Let $\mathcal{D}$ be the completion of $\mathcal{D}_0$. Then the previous theorem tells us that $(\mathcal{M}, \mathcal{D})$ is an ε-model. But, since we only enlarged the sets on which formulas hold, we can now directly verify (using induction over the structure of a formula φ in prenex normal form) that every formula $\varphi \in T$ is ε-satisfied in $(\mathcal{M}, \mathcal{D})$: namely, for universal quantifiers we observe that the set on which a formula holds can only have larger measure, and for existential quantifiers we remark that we only have added more witnesses, not removed any. □

The idea of sending each x_a to $a \in 2^\omega$ was originally inspired by the proof of Bogachev [5, Theorem 9.4.7], albeit in a different context. However, he only discussed the case in which the function π is also surjective (the non-surjective case is irrelevant in his context).

Next, we show that we can eliminate *atoms*.

Definition 5.2.3. Let μ be a measure and let A be a μ-measurable set. We say that A is an **atom** of the measure μ if $\mu(A) > 0$ and every measurable subset $B \subseteq A$ has measure either 0 or $\mu(A)$.

The measure μ is called **atomless** if it does not have any atoms.

We remark that a measure has at most countably many *inequivalent* atoms, where two atoms A, B are called inequivalent if $\mu(A \triangle B) > 0$. Namely, if $\mu(A \triangle B) > 0$, then $\mu(A \cap B) = 0$, and it is easily verified that there can be at most countably many of such almost-disjoint sets of positive measure (indeed, there can be at most n almost-disjoint sets of measure $\geq \frac{1}{n}$).

Lemma 5.2.4. *If μ is a Borel measure on a second countable Hausdorff space X, then μ is atomless if and only if there are no singletons of strictly positive measure.*

Proof. We follow the proof from Aliprantis and Border [1, Lemma 12.18]. The implication from left to right is trivial. For the converse, assume μ has some atom A. Fix a countable base $V_0, V_1, \dots$ for the topology on X. Let $I := \{i \in \omega \mid \mu(A \cap V_i) = 0\}$. Finally, let $B := A \setminus (\bigcup_{i \in I} V_i)$. Then B has positive measure (equal to A). We claim: B is a singleton.

Since, assume $a, b \in B$ are two distinct points. Determine disjoint open sets V_i, V_j such that $a \in V_i$ and $b \in V_j$. Since $\mu(A \cap V_i) = 0$ implies $a \notin B$, we see that $\mu(A \cap V_i) = \mu(A)$. Analogously, we see $\mu(A \cap V_j) = \mu(A)$. However, since V_i and V_j are disjoint we then have:

$$\mu(A) \geq \mu(A \cap V_i) + \mu(A \cap V_j) = 2\mu(A),$$

a contradiction. □

Definition 5.2.5. Let $(\mathcal{M}, \mathcal{D})$ and $(\mathcal{N}, \mathcal{E})$ be two ε-models over the same language. Then we say that $(\mathcal{M}, \mathcal{D})$ and $(\mathcal{N}, \mathcal{E})$ are **ε-elementary equivalent** (notation: $(\mathcal{M}, \mathcal{D}) \equiv_\varepsilon (\mathcal{N}, \mathcal{E})$) if for all sentences φ we have:

$$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi \Leftrightarrow (\mathcal{N}, \mathcal{E}) \models_\varepsilon \varphi.$$

Lemma 5.2.6. *Let $\mathcal{L}$ be a first-order language not containing equality, function symbols or constants. Let $(\mathcal{M}, \mathcal{D})$ be an ε-model, such that $\mathcal{M}$ is a second countable Hausdorff space and $\mathcal{D}$ is the completion of a Borel measure $\mathcal{D}_0$. Then there exists an atomless model $(\mathcal{N}, \mathcal{E})$ such that $\mathcal{E}$ is the completion of a Borel measure $\mathcal{E}_0$ and $(\mathcal{N}, \mathcal{E}) \equiv_\varepsilon (\mathcal{M}, \mathcal{D})$.*

Proof. We first show how to eliminate a single atom of $\mathcal{D}_0$. By Lemma 5.2.4, we may assume it to be a singleton x_0; say of measure r. We define a new measure $\mathcal{E}_0$ on the disjoint union of $\mathcal{M} \setminus \{x_0\}$ and $[0, r]$ by setting, for each $\mathcal{D}_0$-measurable $B \subseteq \mathcal{M}$ and Borel $C \subseteq [0, r]$ (where we let λ denote the Lebesgue measure, restricted to Borel sets):

$$\mathcal{E}_0(B \cup C) = \mathcal{D}_0(B \setminus \{x_0\}) + \lambda(C).$$

Then clearly, x_0 is no longer an atom (since it now has measure zero). We show how to define the unary relations on $\mathcal{M} \cup [0, r]$; the general case is done in the same way. We let $R(x)$ hold if:

- $x \in \mathcal{M}$ and $(\mathcal{M}, \mathcal{D}) \models_\varepsilon R(x)$, or
- $x \in [0, r]$ and $(\mathcal{M}, \mathcal{D}) \models_\varepsilon R(x_0)$.

That $(\mathcal{N}, \mathcal{E}) \equiv_\varepsilon (\mathcal{M}, \mathcal{D})$ is verified by induction on the structure of a formula φ. We prove the universal case; the other cases are similar. So, let $\varphi = \forall x(\psi(x))$. We need to show that

$$\Pr_{\mathcal{D}}[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a)] \geq 1 - \varepsilon$$

if and only if

$$\Pr_{\mathcal{E}}[a \in \mathcal{N} \mid (\mathcal{N}, \mathcal{E}) \models_\varepsilon \psi(a)] \geq 1 - \varepsilon,$$

which, using the induction hypothesis, is equivalent to

$$\Pr_{\mathcal{E}}[a \in \mathcal{N} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a)] \geq 1 - \varepsilon.$$

We either have

$$x_0 \notin \{a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a)\}$$

and then

$$\begin{aligned}
&\Pr_{\mathcal{D}}[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a)] \\
=&\Pr_{\mathcal{D}}[\{a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a)\} \setminus \{x_0\}] \\
=&\Pr_{\mathcal{E}}[a \in \mathcal{N} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a)],
\end{aligned}$$

or x_0 is in the set mentioned above and then

$$\begin{aligned}
&\Pr_{\mathcal{D}}[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a)] \\
=&\Pr_{\mathcal{D}}[\{a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a)\} \setminus \{x_0\}] + r \\
=&\Pr_{\mathcal{D}}[\{a \in \mathcal{N} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a)\} \setminus \{x_0\}] + \lambda[0, r] \\
=&\Pr_{\mathcal{E}}[a \in \mathcal{N} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a)].
\end{aligned}$$

This is exactly what we needed to show.

Finally, using the remark below Definition 5.2.3 above, we can iterate this construction and eliminate the atoms one by one. □

The final definitions we need are the important notions of *measure isomorphism* and *measure isomorphism mod 0.*

Definition 5.2.7. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces. We will say that $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are **isomorphic** if there exists an **isomorphism** from X to Y; that is, a bijection $f : X \to Y$ such that $f(\mathcal{A}) = \mathcal{B}$ and $\mu \circ f^{-1} = \nu$.

The measure spaces $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ are called **isomorphic mod 0** if there exist $A \in \mathcal{A}$, $B \in \mathcal{B}$ with $\mu(A) = \nu(B) = 0$ such that the measure spaces restricted to respectively $X \setminus A$ and $Y \setminus B$ are isomorphic.

The next theorem shows the connection between Borel measures on 2^ω and the Lebesgue measure.

Theorem 5.2.8 (Bogachev [5, Theorem 9.2.2]). *Let $\mathcal{D}$ be an atomless Borel probability measure on a complete separable metric space.*

Then it is isomorphic mod 0 to $[0,1]$ *with the Lebesgue measure restricted to Borel sets.*

Furthermore, x *is in the domain of this isomorphism if and only if it is in the topological support of* $\mathcal{D}$ *(i.e. every open set containing* x *has strictly positive measure).*

We will combine this with the following result concerning isomorphisms mod 0, showing that in our context it does not hurt to have isomorphisms mod 0 instead of true isomorphisms.

Lemma 5.2.9. *Let* $\mathcal{L}$ *be a first-order language not containing equality, function symbols or constants. Let* $(\mathcal{M}, \mathcal{D})$ *be an* ε*-model of cardinality at most* 2^ω *such that* $\mathcal{D}$ *is isomorphic mod 0 to* $[0,1]$ *with the Lebesgue measure. Then there exists an* ε*-model based on* $[0,1]$ *with the Lebesgue measure which is* ε*-elementary equivalent to* $(\mathcal{M}, \mathcal{D})$*.*

Proof. Fix sets A, B such that $\mathcal{D}(A) = \lambda(B) = 0$ and fix an isomorphism $f : [0,1] \setminus B \to \mathcal{M} \setminus A$.

Our idea is to transform f into a surjective mapping g from $[0,1]$ to $\mathcal{M}$ such that g^{-1} is measure-preserving (that is, for every $\mathcal{D}$-measurable set U we have $\lambda(g^{-1}(U)) = \mathcal{D}(U)$). We will then be able to use g^{-1} to define the relations.

To construct such a mapping, let $C \subseteq \left(\frac{1}{2}, 1\right]$ be a copy of the Cantor set and fix a function $\alpha : C \to \mathcal{M}$ such that A is contained in the image of α. We will use α to attain the points not attained by f. Also fix a point $x_0 \in [0,1] \setminus B$. Now define $g : [0,1] \to \mathcal{M}$ by

$$g(x) := \begin{cases} f(2x) & \text{if } x \in \left[0, \frac{1}{2}\right] \text{ and } 2x \notin B \\ f(x_0) & \text{if } x \in \left[0, \frac{1}{2}\right] \text{ and } 2x \in B \\ f(2x-1) & \text{if } x \in \left(\frac{1}{2}, 1\right] \setminus C \text{ and } 2x-1 \notin B \\ f(x_0) & \text{if } x \in \left(\frac{1}{2}, 1\right] \setminus C \text{ and } 2x-1 \in B \\ \alpha(x) & \text{if } x \in U. \end{cases}$$

Then all points in $\mathcal{M} \setminus A$ are attained by g on $\left[0, \frac{1}{2}\right]$ and all points in A are attained by g on C, so g is surjective. That g^{-1} is measure-preserving can be seen by observing that $g^{-1}(U) \cap \left[0, \frac{1}{2}\right]$ and $g^{-1}(U) \cap \left(\frac{1}{2}, 1\right]$ both have Lebesgue-measure exactly $\frac{1}{2}\mathcal{D}(U)$.

Now define the relations on $[0,1]$ by letting $R^{\mathcal{N}}(x_1, \ldots, x_n)$ hold if $R^{\mathcal{M}}(g(x_1), \ldots, g(x_n))$ holds. That the relations are measurable can be directly seen induction over the generation of a relation in the original model (see Proposition 1.1.10); for the base case we remark that we distinguish the various possibilities for g in every variable and thus get a union of 5^n boxes with Lebesgue-measurable edges (where we use that every subset of C is Lebesgue-measurable). It is then easily verified that indeed $(\mathcal{N}, \mathcal{E}) \equiv_\varepsilon (\mathcal{M}, \mathcal{D})$.

We remark that, if A happens to be a subset of 2^ω, we can take α to be the canonical function mapping C to 2^ω. If furthermore A and B are Borel and f is an isomorphism of the Borel measures, it is easily seen that if the relations on $\mathcal{M}$ were Borel sets, then the relations on $\mathcal{N}$ are also Borel. □

Putting together everything we have found, we reach the theorem announced at the beginning of this section.

Theorem 5.2.10. *Let $\mathcal{L}$ be a countable first-order language not containing equality, function symbols or constants. Let T be an ε-satisfiable set of sentences. Then there exists an ε-model on $[0,1]$ with the Lebesgue measure which ε-satisfies T. Furthermore, all relations are Borel.*

Proof. This follows from Proposition 5.2.2, Lemma 5.2.6, Theorem 5.2.8 and Lemma 5.2.9. □

As remarked above (Remark 5.1.6), this theorem uses the full measurability condition. We expect that we really need this condition, since otherwise the sections $R(a, x_2)$ of a binary relation $R(x_1, x_2)$ could form uncountably many different sets, which makes it a problem to show that the measure is countably generated — which is certainly a necessary requirement.

5.3 The Löwenheim Number

The next question we ask ourselves is how tight Theorem 5.1.5 is. The **Löwenheim number** of a logic is the smallest cardinal λ such that every satisfiable sentence has a model of cardinality at most λ. For every ε, we can study the Löwenheim number λ_ε of ε-logic; i.e. the smallest cardinal such that every ε-satisfiable sentence has an ε-model of cardinality at most λ_ε. The next theorem tells us what happens if we assume Martin's axiom MA (see e.g. Kunen [12]), which is equiconsistent with ZFC.

Theorem 5.3.1. *Let $\varepsilon \in [0,1)$ be rational. For the Löwenheim number λ_ε of ε-logic we have*

1. $\aleph_1 \leq \lambda_\varepsilon \leq 2^{\aleph_0}$*,*

2. If Martin's axiom MA holds then $\lambda_\varepsilon = 2^{\aleph_0}$*.*

Proof. The first statement was proven above (in Example 5.1.4 and Theorem 5.1.5). For the second, assume MA holds. We use a proof inspired by Keisler [11]. Let φ be the sentence from Example 5.1.4. Let $\kappa < 2^\omega$ and assume φ has a model of cardinality κ. We remark that any model of φ has to be atomless.

Therefore, if we now use the construction from Proposition 5.2.2, we find a model $(\mathcal{M}, \mathcal{D})$ which ε-satisfies φ and where $\mathcal{D}$ is the completion of an atomless Borel measure. Furthermore, if we let π and $(\mathcal{N}, \mathcal{E})$ be as in the proof of this proposition, the set $\pi(\mathcal{N})$ is a set of cardinality at most κ, so by MA, see Fremlin [6, p127], it has measure 0. But then

$$\mathcal{E}(\mathcal{N}) = \mathcal{D} \circ \pi(\mathcal{N}) = 0,$$

a contradiction. □

So, Theorem 5.1.5 is optimal in the sense that we cannot prove that $\lambda_\varepsilon < 2^{\aleph_0}$ in ZFC. In particular, since MA is consistent with $2^{\aleph_0} > \aleph_1$ (Kunen, [12, p278]), we cannot prove $\lambda_\varepsilon = \aleph_1$. Furthermore, since $2^{\aleph_0} = \aleph_1$ is also consistent with ZFC, $\lambda_\varepsilon = \aleph_1$ is *independent* of ZFC. This, however, does not eliminate the possibility that we might be able to prove $\lambda_\varepsilon = 2^{\aleph_0}$ within ZFC.

5.4 Compactness

We start this section with a negative result: we show that compactness does not hold for our logic, as the next example (taken from Keisler [11, Example 2.6.5]) shows.

Theorem 5.4.1. *For rational* $\varepsilon \in (0, 1)$*, compactness does not hold; i.e. there exists a countable set* Γ *of formulas such that each finite subset is* ε*-satisfiable, but* Γ *is not* ε*-satisfiable.*

Proof. Let R be a binary relation. Using the reductions from Proposition 4.2.2 (observing, from the proof of that proposition, that we can apply the reduction per quantifier), we can form a sentence φ_n such that φ_n is ε-satisfiable if and only if there is a model satisfying:
For almost all y[2]*, there exists a set* A_y *of measure at least* $1-\frac{1}{n}$ *such that*

[2]For measure 1 many.

for all $y' \in A_y$ the sets $B_y := \{u \mid R(u,y)\}$ and $B_{y'} = \{u \mid R(u,y')\}$ both have measure $\frac{1}{2}$, while $B_y \cap B_{y'}$ has measure $\frac{1}{4}$ (in other words, the two sets are independent sets of measure $\frac{1}{2}$)[3].

Then each φ_n has a finite ε-model, as illustrated below: for each x (displayed on the horizontal axis) we let $R(x,y)$ hold exactly for those y (displayed on the vertical axis) where the box has been coloured grey. If we now take for each A_y exactly those three intervals of length $\frac{1}{4}$ in which y is not, we can directly verify that φ_n holds.

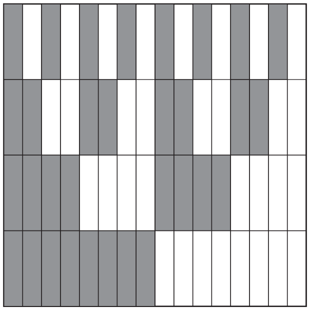

Figure 5.1: A model for φ_4 on $[0,1]$.

However, the set $\{\varphi_n \mid n \in \omega\}$ has no ε-model. Namely, for such a model, we would have that for almost all y, there exists a set A_y of measure 1 such that for all $y' \in A_y$ the sets B_y and $B_{y'}$ (defined above) are independent sets of measure $\frac{1}{2}$.

Clearly, such a model would need to be atomless and therefore cannot be countable. But then we would have uncountably many of such independent sets B_y. Intuitively, this contradicts the fact that R is measurable in the product measure and can therefore be formed using countable unions and countable intersections of Cartesian products.

More formally, the next lemma tells us that for almost all y there exists a set C_y of strictly positive measure such that for all $y' \in C_y$ the sets B_y and $B_{y'}$ agree on a set of measure at least $\frac{7}{8}$; since we can check that $A_y \cap C_y = \emptyset$ this shows that A_y cannot have measure 1. $\square$

Lemma 5.4.2. *Let $(\mathcal{M}, \mathcal{D})$ be an ε-model, let $R = R(x,y)$ be a binary relation and let $\delta > 0$. Then for almost all y there exists a set C_y of strictly positive measure such that for all $y' \in C_y$:*

$$\Pr_{\mathcal{D}}[u \in \mathcal{M} \mid R(u,y) \leftrightarrow R(u,y')] \geq 1 - \delta.$$

[3] That is, $\Pr_{\mathcal{D}}[B_y \cap B_{y'}] = \Pr_{\mathcal{D}}[B_y] \cdot \Pr_{\mathcal{D}}[B_{y'}]$.

Proof. Clearly, it is enough if we show this not for almost all y, but instead show that for all $\delta' > 0$ this holds for at least $\mathcal{D}$-measure $1 - \delta'$ many y.

We first remark that the set $R^{\mathcal{M}}$ can be approximated using finite unions of rectangles $U \times V$ of $\mathcal{D}$-measurable sets; i.e. there exist $\mathcal{D}$-measurable sets U_i, V_i such that:

$$\Pr_{\mathcal{D}}\left[R^{\mathcal{M}} \triangle (\bigcup_{i=1}^{n} U_i \times V_i)\right] < \delta\delta'.$$

This fact can be easily checked by observing that it holds for rectangles $\varphi = U \times V$ and that it is preserved under countable unions and complements (it is usually part of a proof of Fubini's theorem).

Now the V_i induce a partition of $\mathcal{M}$ into at most 2^n many disjoint parts Y_j (by choosing for each $i \leq n$ either V_i or its complement, and intersecting these) . But each such Y_j has either measure zero (so we can ignore it), or Y_j has strictly positive measure and for all $y \in Y_j$ we can take $C_y := Y_j$. One can then verify that for all $y' \in C_y$ and all $u \in \mathcal{M}$ we have $(u, y) \in \bigcup_{i=1}^{n} U_i \times V_i$ if and only if $(u, y') \in \bigcup_{i=1}^{n} U_i \times V_i$. The requested statement about $R^{\mathcal{M}}$ then easily follows from how we approximated $R^{\mathcal{M}}$. □

Next, we will present an ultraproduct-construction that allows us to partially recover compactness, which is due to Hoover and described in Keisler [11]. This construction uses the Loeb measure from non-standard analysis, which is due to Loeb [14]. The same construction as in Keisler is also described in [2] (for a different logic); however, there the Loeb measure is not explicitly mentioned and the only appearance of non-standard analysis is in taking the standard part of some element. Below we will describe the construction without resorting to non-standard analysis. To be able to define the measure, we need the notion of a *limit over an ultrafilter.*

Definition 5.4.3. Let $\mathcal{U}$ be an ultrafilter over ω and let $a_0, a_1, \dots \in \mathbb{R}$. Then a **limit** of the sequence $a_0, a_1, \dots$ over the ultrafilter $\mathcal{U}$ is an $r \in \mathbb{R}$ such that for all $\varepsilon > 0$ we have $\{i \in \omega \mid |a_i - r| < \varepsilon\} \in \mathcal{U}$. We will denote such a limit by $\lim_{\mathcal{U}} a_i$.

Proposition 5.4.4. *Limits over ultrafilters are unique. Furthermore, if $a_0, a_1, \dots$ is a bounded sequence, then for every ultrafilter $\mathcal{U}$ the limit over $\mathcal{U}$ exists.*

Proof. First assume we have an ultrafilter $\mathcal{U}$ over ω and a sequence $a_0, a_1, \dots \in \mathbb{R}$ which has two distinct limits r_0 and r_1. Then the sets

$$\{i \in \omega \mid |a_i - r_0| < |r_0 - r_1|\}$$

and

$$\{i \in \omega \mid |a_i - r_1| < |r_0 - r_1|\}$$

are disjoint elements of $\mathcal{U}$; so, $\emptyset \in \mathcal{U}$, which contradicts $\mathcal{U}$ being a proper filter.

Now, assume the sequence $a_0, a_1, \dots$ is bounded; without loss of generality we may assume that it is a sequence in $[0, 1]$. We will inductively define a decreasing chain $[b_n, c_n]$ of intervals such that for all $n \in \omega$ we have $\{i \in \omega \mid a_i \in [b_n, c_n]\} \in \mathcal{U}$.

First we let $[a_0, b_0] = [0, 1]$. Next, if $\{i \in \omega \mid a_i \in [b_n, c_n]\} \in \mathcal{U}$, then either

$$\left\{i \in \omega \mid a_i \in \left[b_n, \frac{b_n + c_n}{2}\right]\right\} \in \mathcal{U}$$

or

$$\left\{i \in \omega \mid a_i \in \left[\frac{b_n + c_n}{2}, c_n\right]\right\} \in \mathcal{U}.$$

Choose one of these two intervals to be $[b_{n+1}, c_{n+1}]$.

Now there exists a unique point $r \in \bigcap_{n\in\omega}[b_n, c_n]$; it is easily verified that this is the limit of the sequence. □

Using these limits over ultrafilters, we show how to define a probability measure on an ultraproduct of probability measures.

Definition 5.4.5. Let $\mathcal{U}$ be an ultrafilter on ω and let $(\mathcal{A}_i, \mathcal{D}_i)$ be a sequence of measure spaces over sets X_i. For each sequence $A_0, A_1, \dots$ with $A_i \in \mathcal{A}_i$ we will call the set

$$[A_i] := \left\{[a_0, a_1, \dots] \in \prod_{i\in\omega} X_i/\mathcal{U} \mid \{i \in \omega \mid a_i \in A_i\} \in \mathcal{U}\right\}$$

a **basic measurable set**. If we let Γ be the collection of all basic measurable sets, then we define the **ultraproduct measure** to be the unique measure $\mathcal{E}$ on $\sigma(\Gamma)$ such that for all basic measurable sets:

$$\Pr_{\mathcal{E}}([A_i]) = \lim_{\mathcal{U}} \Pr_{\mathcal{D}_i}(A_i).$$

Proposition 5.4.6. *The ultraproduct measure exists and is well-defined.*

Proof. We need to verify that $\mathcal{E}$ as defined on the set Γ of basic measurable sets is a pre-measure, so that we can apply Carathéodory's extension theorem (Theorem 1.1.7). We will show that for any decreasing sequence $[A_i^0] \supseteq [A_i^1] \supseteq \ldots$ such that $\bigcap_{j\in\omega}[A_i^j] = \emptyset$ we have that $\lim_{j\to\infty} \Pr_{\mathcal{E}}([A_i^j]) = 0$, which is equivalent to $\mathcal{E}$ being a pre-measure.

So, assume this limit is not 0. By taking a subsequence we may assume that there exists an $\varepsilon > 0$ such that for all $j \in \omega$ we have

$$\Pr_{\mathcal{E}}([A_i^j]) \geq \varepsilon$$

or equivalently

$$\{i \in \omega \mid \Pr_{\mathcal{D}_i}(A_i^j) \geq \varepsilon\} \in \mathcal{U}.$$

Let $I := \{i \in \omega \mid \Pr_{\mathcal{D}_i}(A_i^0) \geq \varepsilon\}$. Now observe that for all $j \in \omega$ we have

$$\{i \in \omega \mid A_i^j \supseteq A_i^{j+1}\} \in \mathcal{U}.$$

But then, we also have for all $j \in \omega$ that

$$S_j := \{i \in \omega \mid A_i^0 \supseteq A_i^j\} \cap \{i \in \omega \mid \Pr_{\mathcal{D}_i}(A_i^j) \geq \varepsilon\} \in \mathcal{U}.$$

Now define sets B_i^j by

$$B_i^j := \begin{cases} A_i^0 & \text{if } j = 0 \\ A_i^j & \text{if } j > 0 \text{ and } i \in S_j \\ A_i^k & \text{if } j > 0,\ i \notin S_j \text{ and } k = \mu n[i \in S_n]. \end{cases}$$

Then we have for all $i, j \in \omega$ that $B_i^j \supseteq B_i^{j+1}$ and for all $i \in I$, $j \in \omega$ that $\Pr_{\mathcal{D}_i}(B_i^j) \geq \varepsilon$. Furthermore, we can verify that for all $j \in \omega$ we have that $[B_i^j] = [A_i^j]$; therefore we have in particular that $\bigcap_{j\in\omega}[B_i^j] = \emptyset$. Since $I \in \mathcal{U}$, we now see that there must exist an $i \in I$ such that $\bigcap_{j\in\omega} B_i^j = \emptyset$, which contradicts $\Pr_{\mathcal{D}_i}(B_i^j) \geq \varepsilon$. □

We can now define a model on the ultraproduct in the usual way; however, we cannot guarantee that this is an ε-model, since we merely know that all definable subsets of $\mathcal{M}$ of arity 1 are measurable (cf. Section 2.1, remark 7). This is made precise in the next definition.

Definition 5.4.7. If $\mathcal{M}$ is a first-order model and $\mathcal{D}$ is a probability measure on $\mathcal{M}$ such that for all formulas $\varphi = \varphi(x_1, \ldots, x_n)$ and all $a_1, \ldots, a_{n-1} \in \mathcal{M}$, the set

$$\{a_n \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(a_1, \ldots, a_n)\}$$

is $\mathcal{D}$-measurable, we say that $(\mathcal{M}, \mathcal{D})$ is a **weak ε-model**.

Clearly, all ε-models are also weak ε-models. However, because we do not require relations and functions to be measurable, the converse need not hold.

We now proceed with the definition of the model-structure on the ultraproduct.

Definition 5.4.8. Let $\varepsilon \in [0,1]$, let $\mathcal{U}$ be an ultrafilter over ω and let $(\mathcal{M}_0, \mathcal{D}_0), (\mathcal{M}_1, \mathcal{D}_1), \dots$ be a sequence of weak ε-models. We then define the **ultraproduct** of this sequence, which we will denote by $\prod_{i\in\omega}(\mathcal{M}_i, \mathcal{D}_i)/\mathcal{U}$, to be the classical ultraproduct of the models $\mathcal{M}_i$, equipped with the ultraproduct measure.

More precisely, we define it to be the model having as universe $\prod_{i\in\omega} \mathcal{M}_i/\mathcal{U}$, where for each relation $R(x^1, \dots, x^n)$ we define the relation on $\prod_{i\in\omega}(\mathcal{M}_i, \mathcal{D}_i)/\mathcal{U}$ by

$$R([a_i^1], \dots, [a_i^n]) \Leftrightarrow \{i \in \omega \mid (\mathcal{M}_i, \mathcal{D}_i) \models R(a_i^1, \dots, a_i^n)\} \in \mathcal{U},$$

and we define function symbols $f(x_1, \dots, x_n)$ by

$$f([a_i^1], \dots, [a_i^n]) := [f(a_0^1, \dots, a_0^n), f(a_1^1, \dots, a_1^n), \dots];$$

in particular for constants c we define

$$c := [c, c, \dots].$$

Finally, we take as measure the ultraproduct measure.

We can now show that the fundamental theorem of ultraproducts, or Łoś's theorem, holds for this definition.

Theorem 5.4.9 (Łoś's theorem for probability logic). *For every formula $\varphi(x^1, \dots, x^n)$ and every sequence of elements $[a_i^1], \dots, [a_i^n] \in \prod_{i\in\omega}(\mathcal{M}_i, \mathcal{D}_i)/\mathcal{U}$, the following are equivalent:*

1. $\prod_{i\in\omega}(\mathcal{M}_i, \mathcal{D}_i)/\mathcal{U} \models_\varepsilon \varphi([a_i^1], \dots, [a_i^n])$,
2. $\{i \in \omega \mid (\mathcal{M}_i, \mathcal{D}_i) \models_\varepsilon \varphi(a_i^1, \dots, a_i^n)\} \in \mathcal{U}$.

Proof. We may assume φ to be in prenex normal form by Proposition 2.1.5. We now use induction over the structure of φ. The only case that is different from the classical case is the universal case. Thus, let $\varphi = \forall x^{n+1}(\psi(x^1, \dots, x^{n+1})$.

By definition,

$$\prod_{i\in\omega}(\mathcal{M}_i,\mathcal{D}_i)/\mathcal{U}\models_\varepsilon \varphi([a_i^1],\ldots,[a_i^n])$$

is equivalent to

$$\Pr_{\mathcal{E}}([a_i^{n+1}]\in\prod_{i\in\omega}\mathcal{M}_i/\mathcal{U}\mid\prod_{i\in\omega}(\mathcal{M}_i,\mathcal{D}_i)/\mathcal{U}\models\psi([a_i^1],\ldots,[a_i^{n+1}]))\geq 1-\varepsilon.$$

By induction hypothesis, we know that this is equivalent to

$$\Pr_{\mathcal{E}}([a_i^{n+1}]\in\prod_{i\in\omega}\mathcal{M}_i/\mathcal{U}\mid\{i\in\omega\mid(\mathcal{M}_i,\mathcal{D}_i)\models_\varepsilon\psi(a_i^1,\ldots,a_i^{n+1})\}\in\mathcal{U})$$
$$\geq 1-\varepsilon. \tag{5.2}$$

Observe that this set is exactly the basic measurable set

$$[\{a_i^{n+1}\in\mathcal{M}_i\mid(\mathcal{M}_i,\mathcal{D}_i)\models_\varepsilon\psi(a_i^1,\ldots,a_i^{n+1})\}].$$

So, equation (5.2) can be verified to be equivalent to

$$\{i\in\omega\mid\Pr_{\mathcal{E}}(a_i^{n+1}\in\mathcal{M}_i\mid(\mathcal{M}_i,\mathcal{D}_i)\models_\varepsilon\psi(a_i^1,\ldots,a_i^{n+1}))\geq 1-\varepsilon\}\in\mathcal{U}$$

which is of course equivalent to

$$\{i\in\omega\mid(\mathcal{M}_i,\mathcal{D}_i)\models_\varepsilon\varphi(a_i^1,\ldots,a_i^n)\}\in\mathcal{U}.$$

□

Corollary 5.4.10. *The ultraproduct is a weak ε-model.*

Proof. For every formula $\varphi=\varphi(x)$, Łoś's theorem tells us that the subset of $\prod_{i\in\omega}(\mathcal{M}_i,\mathcal{D}_i)/\mathcal{U}$ defined by φ is exactly the basic measurable set given by the subsets of $(\mathcal{M}_i,\mathcal{D}_i)$ defined by φ. □

We remark that this construction, in general, does not yield an ε-model. For example, if we have a binary relation $R(x^1,x^2)$ and on each model $(\mathcal{M}_i,\mathcal{D}_i)$ the relation R consists of the union of two 'boxes' $(X_i\times Y_i)\cup(U_i\times V_i)$, then we would need an uncountable union of boxes of basic measurable sets to form R in the ultraproduct model. This is, of course, not an allowed operation on σ-algebras.

A more formal argument showing that the ultraproduct construction does not necessarily yield ε-models is that this construction allows us to prove a weak compactness result in the usual way. If this would always yield an ε-model, this would contradict Theorem 5.4.1 above.

Theorem 5.4.11 (Weak compactness theorem). *Let T be an at most countable set of sentences such that each finite subset is satisfied in a weak ε-model. Then there exists a weak ε-model $(\mathcal{M}, \mathcal{D})$ satisfying T.*

Proof. Let $A_0, A_1, \dots$ be an enumeration of the finite subsets of T. For each A_i, fix a weak ε-model $(\mathcal{M}_i, \mathcal{D}_i)$ satisfying all formulas from A_i. Use the ultrafilter theorem to determine an ultrafilter $\mathcal{U}$ on ω such that for all $\varphi \in T$ we have

$$\{i \in \omega \mid (\mathcal{M}_i, \mathcal{D}_i) \models_\varepsilon \varphi\} \in \mathcal{U}.$$

Now form the ultraproduct; by Theorem 5.4.9 we can directly verify that this ultraproduct satisfies all $\varphi \in T$. $\square$

Future Research

Since the probability logic that we studied has only recently been introduced, there are a lot of open questions remaining and the research could be taken in many directions. We briefly state some of them.

1. *What is the computational complexity of validity and satisfiability?* In Chapter 4 we found lower bounds for the complexity of these problems. It is very likely that, using the Löwenheim-Skolem results from Chapter 5 (Theorem 5.1.5 and Theorem 5.2.10), we could find an upper bound of Π^2_1 for validity and Δ^2_0 for satisfiability (although this would take a lot of effort to arithmetize the necessary parts of measure theory). We would like to know the exact complexity of validity and satisfiability.

2. *Further development of model theory.* So far, we have only looked at the most elementary model-theoretic theorems. There are many more theorems which are worth further investigation: for example, we have not touched on questions of axiomatisability, elimination of quantifiers or categoricity. Of course, it is also to be expected that we will find new theorems with no analogue in classical logic.

3. *Decidability of fragments.* We have proven that both validity and satisfiability are undecidable. However, these problems are decidable for certain fragments of probability logic. We wish to study these fragments and show for which ones validity and satisfiability are decidable, and for which ones they are not.

Bibliography

[1] C. D. Aliprantis and K. C. Border. *Infinite Dimensional Analysis: A Hitchhiker's Guide*. Springer, 2006.

[2] S.-M. Bageri and M. Pourmahdian. The logic of integration. *Archive for Mathematical Logic*, 48:465–492, 2009.

[3] J. Barwise and S. Feferman, editors. *Model-Theoretic Logics*, volume 8 of *Perspectives in Mathematical Logic*. Springer-Verlag, 1985.

[4] V. I. Bogachev. *Measure Theory*, volume 1. Springer, 2007.

[5] V. I. Bogachev. *Measure Theory*, volume 2. Springer, 2007.

[6] D. H. Fremlin. *Consequences of Martin's axiom*, volume 84 of *Cambridge tracts in mathematics*. Cambridge University Press, 1984.

[7] D. N. Hoover. Probability logic. *Annals of Mathematical Logic*, 14:287–313, 1978.

[8] D. N. Hoover. A probabilistic interpolation theorem. *The Journal of Symbolic Logic*, 50(3):708–713, 1985.

[9] A. S. Kechris. *Classical Descriptive Set Theory*. Springer-Verlag, 1995.

[10] H. J. Keisler. Hyperfinite model theory. In R. O. Gandy and J. M. E. Hyland, editors, *Logic Colloquium 76*, volume 87 of *Studies in Logic and the Foundations of Mathematics*, pages 5–110. North-Holland Publishing Company, 1977.

[11] H. J. Keisler. Probability quantifiers. In Barwise and Feferman [3].

[12] K. Kunen. *Set Theory: An Introduction to Independence Proofs*. Elsevier, 1980.

[13] R. Kuyper and S. A. Terwijn. Model-theoretic aspects of probability logic. In preparation.

[14] P. A. Loeb. Conversion from nonstandard to standard measure spaces and applications in probability theory. *Transactions of the American Mathematical Society*, 211:113–122, 1975.

[15] P. Odifreddi. *Classical Recursion Theory*, volume 125 of *Studies in Logic and the Foundations of Mathematics*. North-Holland, 1989.

[16] C. I. Steinhorn. Borel structures and measure and category logics. In Barwise and Feferman [3].

[17] S. A. Terwijn. Probabilistic logic and induction. *Journal of Logic and Computation*, 15(4):507–515, 2005.

[18] S. A. Terwijn. Decidability and undecidability in probability logic. In *Proceedings of Logical Foundations of Computer Science*, volume 5407 of *Lecture Notes in Computer Science*, pages 441–450. Springer, 2009.

[19] L. G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

[20] L. G. Valiant. Robust logics. *Artificial Intelligence*, 117:231–253, 2000.

Index

$(\mathcal{M},\mathcal{D}) \models_K$, 22
Π^0_1, 38
Π^1_1, 27, 34
ε-elementary equivalent, 46
ε-model
    weak, 55
$\equiv_\varepsilon$, 46
$\mathcal{L}_{\omega P}$, 21–23
$\lim_{\mathcal{U}}$, 52
$\models_K$, 22
$\models_\varepsilon$, 11
$(\mathcal{M},\mathcal{D}) \models_\varepsilon$, 10
μ^n, 7
$\oplus$, 6
$\otimes$, 7
φ^N, 31
$\prod_{i\in\omega}(\mathcal{M}_i,\mathcal{D}_i)/\mathcal{U}$, 55
$\sigma(\Gamma)$, 5, 7
$\subset_\varepsilon$, 39

algebra
    σ, 5
    Boolean, 6
almost everywhere, 14
atom, 45
atomless measure, 45

Boolean algebra, 6
Borel
    σ-algebra, 8
    measure, 8
    set, 8

Carathéodory's extension theorem, 6
complete measure, 8
completion of a measure, 8

distribution, probability, 6
downward Löwenheim-Skolem theorem, 39–43

elementary ε-submodel, 39
elementary equivalent, ε-, 46
ε-model, 10
ε-satisfiable, 11
ε-submodel
    elementary, 39
ε-tautology, 11
ε-truth, 10
ε-valid, 11
equivalent, ε-elementary, 46
extension theorem
    Carathéodory's, 6

Fubini's theorem, 7
function, measurable, 6

isomorphic measure spaces, 47
    mod 0, 47

Keisler's probability logic, 21–23

limit over ultrafilter, 52
Löwenheim number, 49

Löwenheim-Skolem theorem
  downward, 39–43

measurable function, 6
measurable set, 6
measure, 5
  atomless, 45
  Borel, 8
  complete, 8
  completion, 8
  pre-, 6
  probability, 6
  product, 7
  space, 6
    isomorphism, 47
    isomorphism mod 0, 47
  sum, 6
  ultraproduct, 53
model
  ε-model, 10

normal form, prenex, 13
number, Löwenheim, 49

PAC-learning, 16
paraconsistency, 11
pre-measure, 6
prenex normal form, 13
probabilistic tautology, 11
probabilistically satisfiable, 11
probability
  distribution, 6
  measure, 6
product
  of σ-algebras, 7
  of measures, 7

relativisation, 31

sampling oracle, 16
satisfiable
  ε-satisfiable, 11
  probabilistically, 11
σ-algebra, 5
  Borel, 8
  generated by a set, 5, 7
  product, 7
  sum, 6
space, measure, 6
submeasure, 39
submodel
  ε-submodel, 39
  elementary ε-submodel, 39
sum
  of σ-algebras, 6
  of measures, 6

tautology
  ε-tautology, 11
  classical, 14
  probabilistic, 11
theorem
  Carathéodory's extension, 6
  Fubini, 7
  Löwenheim-Skolem
    downward, 39–43
truth
  ε-, 10

ultrafilter, limit over, 52
ultraproduct, 55
ultraproduct measure, 53

Valiant's Robust Logic, 24–25
valid, ε-, 11

weak ε-model, 55