MODEL THEORY OF MEASURE SPACES AND PROBABILITY LOGIC

RUTGER KUYPER AND SEBASTIAAN A. TERWIJN

ABSTRACT. We study the model-theoretic aspects of a probability logic suited for talking about measure spaces. This nonclassical logic has a model theory rather different from that of classical predicate logic. In general, not every satisfiable set of sentences has a countable model, but we show that one can always build a model on the unit interval. Also, the probability logic under consideration is not compact. However, using ultraproducts we can prove a compactness theorem for a certain class of weak models.

1. INTRODUCTION

Probability logic is an area in which formal logic is combined with probability theory. There is a large range of possible motivations for doing so, coming from various viewpoints in mathematics, computer science, and philosophy. As in classical logic, the main mathematical challenges in this subject spring from the interplay between the syntax (the logical formalisms) and their semantics (the models). Roughly, the various approaches to the subject (some of which are surveyed in Halpern [8]) fall into two categories: those that consider probability distributions over classes of models (i.e. "probabilities over models"), and those that consider models each with their own probability distribution (i.e. "models with probabilities"). Historically, the majority of the literature on probability logic falls in the first category. Our work below falls into the second category. As examples of work in the first category we can mention the famous papers Carnap [5], Gaifman [7], and Scott and Krauss [18]. (In these papers probabilities are assigned to formulas, which is the same as assigning probabilities to classes of models defined by them. In Howson [11] this work is discussed further, and the proposal is made to move away from formal languages altogether.) However, also probabilistic logics from the area of model checking, such as PCTL and its variants [9], fall in this category. In this context the models are infinite traces of finite state systems, and the probabilities on the transitions of the system define a probability measure on the space of traces.

Examples of work in the second category include H. Friedman's quantifier Q (meaning "almost all" in the sense of measure theory), see Steinhorn [19],

as well as the work of Keisler and Hoover [15]. The logic $L_{A\mathbb{P}}$ from [15] contains probability quantifiers of the form $Px \geqslant r$, where $r \in [0, 1]$, with the intended meaning "true for at least measure r many x". To avoid problems with projections of measurable sets, $L_{A\mathbb{P}}$ does not have the classical existential or universal quantifier. Keisler's Problem 1 [15, p555] asks to "develop a form of $L_{A\mathbb{P}}$ which has the universal quantifier", which, in the presence of classical negation, is of course the same as having the classical existential quantifier.

Jaeger [12] studies an extension of classical predicate logic with probability quantifiers, and also contains a discussion on related work by other authors such as Bacchus and Halpern. Just as the logic studied below, Jaeger's logic is motivated by inductive probabilistic reasoning. Measure quantifiers also naturally occur in descriptive set theory, see e.g. Kechris [14, p114]. We use the behavior of these quantifiers on analytic sets in section 5.

Valiant [22] introduced a logic related to his own model of probabilistic induction, the famous pac-model. Since its introduction in 1984, this model has become one of the cornerstones of computational learning theory, see [13] for an introduction. Though the logic studied below is also motivated by the pac-model, Valiant's setup is completely different from ours. Valiant considers only finite models, subsumed in his definition of a *scene*, and he considers learning certain rules from scenes. Accuracy of learning is defined by probability distributions over scenes, so that this work falls in the first category described above.

In Terwijn [20] a probabilistic logic was introduced that is, like Valiant's logic, related to probabilistic induction and the pac-model. Unlike Valiant's logic however, this logic falls in the models-with-probabilities category. Also, the semantics of this logic is not restricted to finite models, but allows arbitrary measure spaces to serve as models. The language of the logic is identical to that of classical predicate logic, except that the interpretation of the universal quantifier is probabilistic. Since the language carries a parameter ε for the amount of uncertainty with which formulas are evaluated, we will refer to it as ε-logic. This ε-logic is paraconsistent, and combines in a natural way probabilistic universal quantifiers with classical existential ones. This can be seen as a possible answer to Keisler's question quoted above. In section 2 we repeat the definition of ε-logic, and briefly discuss some of its properties. In the subsequent sections, we develop some of its model theory.

The paper is structured as follows. First, in section 2 we introduce ε-logic and discuss some of its basic properties. Next, in section 3 we discuss the appropriate class of models, called ε-models. After that, in section 4 we show that there are satisfiable sentences that are not satisfiable in any countable model, but we also prove a downward Löwenheim-Skolem Theorem for ε-logic in which "countable" is replaced by "of cardinality of the continuum". In section 5 we refine this result and show that every satisfiable sentence is in fact satisfied by a model on the unit interval together with the Lebesgue measure. Section 6 continues with the problem discussed in the preceding section, by discussing what the exact value of the Löwenheim number (the

smallest cardinality for which every satisfiable sentence has a model of that cardinality) is in ε-logic. Next, in section 7 we present a technical result that gives many-one reductions between satisfiability in ε_0-logic and ε_1-logic for different $\varepsilon_0, \varepsilon_1$. Finally, in section 8 we show that compactness fails for ε-logic, but that we can recover a weak notion of compactness using an ultraproduct construction.

Our notation from logic is standard. We denote the natural numbers by ω. For unexplained notions from measure theory, we refer the reader to Bogachev [4], and for notions from descriptive set theory to Kechris [14]. The relevant background for model theory can be found in Hodges [10].

2. ε-LOGIC

In this section, we will repeat the definition of the probabilistic logic from Terwijn [20]. This logic was partly motivated by the idea of what it means to "learn" an ordinary first-order statement φ from a finite amount of data from a model $\mathcal{M}$ of φ, in a way that is similar to learning in Valiant's pac-model [13]. In this setting, atomic data are generated by sampling from an unknown probability distribution $\mathcal{D}$ over $\mathcal{M}$, and the task is to decide with a prescribed amount of certainty whether φ holds in $\mathcal{M}$ or not. On seeing an atomic truth $R(a)$, where R is some relation, one knows with certainty that $\exists x R(x)$, so that the existential quantifier retains its classical interpretation. On the other hand, inducing a universal statement $\forall x R(x)$ can only be done probabilistically. Thus, there is a fundamental asymmetry between the interpretation of the existential quantifier and the interpretation of the universal quantifier. As in the pac-model, it is important that the distribution $\mathcal{D}$ is unknown, which is counterbalanced by the fact that success of the learning task is measured using the same distribution $\mathcal{D}$. (In the pac-setting this is called "distribution-free learning".) In [20] it was shown that ordinary first-order formulas are pac-learnable under the appropriate probabilistic interpretation, given in the definition below. In this paper the focus will be on the model theory of this logic, and no background on pac-learning is required any further.

Definition 2.1. Let $\mathcal{L}$ be a first-order language, possibly containing equality, of a countable signature. Let $\varphi = \varphi(x_1, \ldots, x_n)$ 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 the universe of $\mathcal{M}$. Then we inductively define the notion of ε-*truth*, denoted by $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi$, as follows (where we leave the parameters implicit).

(i) For every atomic formula φ:

$$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi \text{ if } \mathcal{M} \models \varphi.$$

(ii) We treat the logical connectives $\wedge$ and $\vee$ classically, e.g.

$$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi \wedge \psi \text{ if } (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi \text{ and } (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi.$$

(iii) The existential quantifier is treated classically as well:

$$(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \exists x \varphi(x)$$

if there exists an $a \in \mathcal{M}$ such that $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \varphi(a)$.

(iv) The case of negation is split into sub-cases as follows:
  (a) For φ atomic, $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \neg\varphi$ if $(\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon} \varphi$.
  (b) $\neg$ distributes in the classical way over $\wedge$ and $\vee$, e.g.

$$(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \neg(\varphi \wedge \psi) \text{ if } (\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \neg\varphi \vee \neg\psi.$$

  (c) $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \neg\neg\varphi$ if $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \varphi$.
  (d) $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \neg(\varphi \rightarrow \psi)$ if $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \varphi \wedge \neg\psi$.
  (e) $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \neg\exists x \varphi(x)$ if $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \forall x \neg\varphi(x)$.
  (f) $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \neg\forall x \varphi(x)$ if $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \exists x \neg\varphi(x)$.
(v) $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \varphi \rightarrow \psi$ if $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \neg\varphi \vee \psi$.
(vi) Finally, we define $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \forall x \varphi(x)$ if

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

We make some remarks about the definition of ε-truth. Observe that everything in Definition 2.1 is treated classically, except for the interpretation of $\forall x \varphi(x)$ in case (vi). 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, where "many" depends on the error parameter ε. The treatment of negation requires some care, since we no longer have that $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \neg\varphi$ implies that $(\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon} \varphi$ (though the converse still holds, see Terwijn [20, Proposition 3.1]). The clauses for the negation allow us to push the negations down to the atomic formulas.

Note that both $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \forall x \varphi(x)$ and $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon} \exists x \neg\varphi(x)$ can hold, for example if $\varphi(x)$ holds on a set of measure one but not for all x. Thus, the logic defined above is *paraconsistent*. This example also shows that for $\varepsilon = 0$ the notion of ε-truth does not coincide with the classical one. Note also that even though both φ and $\neg\varphi$ may be satisfiable, they cannot both be ε-tautologies, as at most one of them can be true in a model with only one point.

We have chosen to define $\varphi \rightarrow \psi$ as $\neg\varphi \vee \psi$. We note that this is weaker than the classical implication. The classical definition would say that ψ holds in any model where φ holds. Using an atomic inconsistency as falsum, we would thus obtain a classical negation. Since $\exists$ expresses classical existence, we would then also obtain the classical universal quantifier $\forall$, and our logic would become a strong extension of classical predicate logic, which is not what we are after.

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

Definition 2.2. Let $\mathcal{L}$ be a first-order language of a countable signature, possibly containing equality, and let $\varepsilon \in [0, 1]$. Then an ε-*model* $(\mathcal{M}, \mathcal{D})$ for the language $\mathcal{L}$ consists of a classical first-order $\mathcal{L}$-model $\mathcal{M}$ together with

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 (including equality, if it is in $\mathcal{L}$) and all functions of arity n are measurable as functions from $(\mathcal{M}^n, \mathcal{D}^n)$ to $(\mathcal{M}, \mathcal{D})$ (where $\mathcal{D}^n$ denotes the n-fold product measure). In particular, constants are $\mathcal{D}$-measurable.

We remark that condition (2) does not imply condition (1), because even if a set is measurable, its image under a projection need not be measurable. The definition of ε-model is discussed in more detail in section 3.

Definition 2.3. 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 $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi$. Similarly, we say that φ is ε-*satisfiable* if there exists an ε-model $(\mathcal{M}, \mathcal{D})$ such that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi$.

Note that all ε-models are necessarily nonempty because they are probability spaces. From Proposition 2.5 it follows that for $\varepsilon \leqslant \varepsilon'$ and for a suitable class of models, every ε-tautology is an ε'-tautology. In [20] it was proven that the inclusion is strict when $\varepsilon < \varepsilon'$.

Example 2.4. Let Q be a unary predicate. Then $\varphi = \forall x Q(x) \lor \forall x \neg Q(x)$ 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, φ 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.

In many proofs it will be convenient to work with formulas in prenex normal form. We may assume that formulas are in this form by the following:

Proposition 2.5 (Terwijn [20]). *Every formula φ is semantically equivalent to a formula φ' in prenex normal form; i.e. $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi \Leftrightarrow (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi'$ for all $\varepsilon \in [0, 1]$ and all ε-models $(\mathcal{M}, \mathcal{D})$.*

3. ε-Models

Our definition of ε-model slightly differs from the original definition in [20]. We require more sets to be measurable in our ε-models than in the original definition, where the measurability condition was included in the truth definition. However, we need this stronger requirement on our models to be able to prove anything worthwhile, in fact, (1) in Definition 2.2 is already implicit in most proofs published in earlier papers.

We now discuss condition (2) in Definition 2.2. This is a natural assumption: When we are talking about probabilities over certain predicates we may as well require that all such probabilities exist, even if in some cases this would not be necessary. To illustrate this point we give an example of what can happen without it.

Example 3.1. The following example is based on the famous argument of Sierpinski showing that under the continuum hypothesis CH there are unmeasurable subsets of the real plane. Let $\mathcal{D}$ be a measure on the domain ω_1 defined by

$$\mathcal{D}(A) = \begin{cases} 1 & \text{if } A = \omega_1 \text{ with the exception of at} \\ & \text{most countably many elements,} \\ 0 & \text{if A is countable.} \end{cases}$$

It is easy to check that $\mathcal{D}$ is a probability measure. Let $<$ be the usual order relation on ω_1. Then we have

$$(\omega_1, \mathcal{D}) \models_0 \forall x \forall y (x < y)$$

since for every $x \in \omega_1$ the vertical section $\{y \mid x < y\}$ has $\mathcal{D}$-measure 1. Similarly,

$$(\omega_1, \mathcal{D}) \not\models_0 \forall y \forall x (x < y)$$

since for every $y \in \omega_1$ the horizontal section $\{x \mid x < y\}$ has $\mathcal{D}$-measure 0. Note that the relation $\{(x,y) \in \omega_1{}^2 \mid x < y\}$ is not $\mathcal{D}^2$-measurable: Since all its vertical sections $\{y \mid x < y\}$ have $\mathcal{D}$-measure 1, and all its horizontal sections $\{x \mid x < y\}$ have $\mathcal{D}$-measure 0, $\mathcal{D}^2$-measurability of the relation $<$ would contradict Fubini's theorem. That in general universal quantifiers do not commute under a probabilistic interpretation was already remarked in Keisler [15]. In fact, it is easy to give a three-element example of a model $\mathcal{M}$ such that $(\mathcal{M}, \mathcal{D}) \models_{\frac{1}{3}} \forall x \forall y R(x, y)$ but not $(\mathcal{M}, \mathcal{D}) \models_{\frac{1}{3}} \forall y \forall x R(x, y)$. So in this respect condition (2) does not help anyway.

We point out that the choice to impose condition (2) or not *does* make a difference for the resulting probability logic: Let

$$\mathsf{lin} = \forall x \forall y (x \leqslant y \vee y \leqslant x) \wedge \forall x \forall y \forall z (x \leqslant y \wedge y \leqslant z \rightarrow x \leqslant z)$$

be the sentence saying that $\leqslant$ is a total preorder (not necessarily antisymmetric).

Proposition 3.2. *For every $\varepsilon > 0$, the sentence*

$$\varphi = \neg \mathsf{lin} \vee \exists x \forall y (y \leqslant x)$$

is an ε-tautology if and only if we impose condition (2) in Definition 2.2.

Proof. Let $\varepsilon > 0$. Without (2) we can construct a countermodel for φ as follows. Consider ω_1 with the measure $\mathcal{D}$ from Example 3.1. Then since every initial segment of ω_1 has measure 0, clearly $(\omega_1, \mathcal{D}) \not\models_\varepsilon \varphi$.

Now suppose (2) holds. Let $(\mathcal{M}, \mathcal{D})$ be an ε-model with $\varepsilon > 0$. We show that φ is an ε-tautology. When $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \neg \mathsf{lin}$ then we are done. When $(\mathcal{M}, \mathcal{D}) \not\models_\varepsilon \neg \mathsf{lin}$ then we have classically $\mathcal{M} \models \mathsf{lin}$ so $\leqslant$ really is a linear preorder in $\mathcal{M}$. Suppose that

$$(3) \qquad\qquad \forall x \; \Pr_{\mathcal{D}}\big[y \in \mathcal{M} \mid x \leqslant y\big] > \varepsilon.$$

Suppose further that

$$\forall x_0 \exists y \geqslant x_0 \Pr_{\mathcal{D}}\big[[x_0, y]\big] \geqslant \frac{1}{2}\varepsilon$$

where $[x_0, y]$ is the interval between x_0 and y in $\mathcal{M}$. We may assume that for every $y \in \mathcal{M}$ there exists a $z \in \mathcal{M}$ with $z \not\leqslant y$; otherwise y is clearly a maximal element and we are done. Then we also have

$$\forall x_0 \exists y > x_0 \Pr_{\mathcal{D}}\big[[x_0, y]\big] \geqslant \frac{1}{2}\varepsilon$$

(where $y > x_0$ denotes $x_0 \leqslant y \wedge y \not\leqslant x_0$). But then we can find infinitely many intervals $[y_0, y_1], [y_1, y_2], \ldots$ with $y_i < y_{i+1}$ of measure at least $\frac{1}{2}\varepsilon$, which are disjoint by the transitivity of $\leqslant$. This is a contradiction. So, choose x_0 such that $\forall y \geqslant x_0 \; \Pr_{\mathcal{D}}\big[[x_0, y]\big] < \frac{1}{2}\varepsilon$ and consider the set

(4) $$\big\{(x, y) \in \mathcal{M} \times \mathcal{M} \mid x, y \geqslant x_0 \wedge x \leqslant y\big\},$$

i.e. the restriction of the relation $\leqslant$ to elements greater than x_0. Then, similarly as in Example 3.1, all vertical sections of (4) have measure $> \varepsilon$ and all horizontal sections have measure $< \frac{1}{2}\varepsilon$, so by Fubini's theorem the set (4) is not $\mathcal{D}^2$-measurable. But then, since (4) is the intersection of sets defined using $\leqslant$, the relation $\leqslant$ itself is not measurable, contradicting (2). So (3) is false, and hence there is an $x \in \mathcal{M}$ such that at least $1 - \varepsilon$ of the weight is to the left of x. Hence $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi$. □

After these considerations we take the standpoint that it is natural to assume the measurability condition (2) in Definition 2.2. As we will see below, for the discussion of compactness it is useful to consider a weaker notion of ε-model, where we drop the condition (2) from the definition:

Definition 3.3. 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*.

4. A Downward Löwenheim-Skolem Theorem

In this section, we will prove a downward Löwenheim-Skolem theorem for ε-logic. We will see that it is not always possible to push the cardinality of a model down to being countable, as in classical logic. In many ways, countable ε-models are analogous to *finite* classical models, as exemplified by the following result:

Theorem 4.1. (Terwijn [21]) *Let φ be a sentence. Then the following are equivalent:*

(i) *φ classically holds in all finite classical models,*
(ii) *$(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi$ for all $\varepsilon > 0$ and for all ε-models $(\mathcal{M}, \mathcal{D})$ such that $\mathcal{M}$ is countable.*

Definition 4.2. 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 $\mu(A_B) = \nu(B)$.

To motivate this definition, let us first consider the special case where $A_B = B$ for every $B \in \mathcal{B}$. In this case, we have that $\nu(B) = \mu(B)$ for every $B \in \mathcal{B}$. In other words, ν is just the restriction of the measure ν to the μ-measurable set N. However, requiring N to be a measurable subset of M is too restrictive for the constructions below. Our definition of submeasure also allows us to restrict μ to certain non-μ-measurable sets N, by allowing us some freedom in the choice of the set A_B. The precise method of constructing such submeasures $\mathcal{N}$ will become clear in the proof of Theorem 4.6.

Definition 4.3. 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 $\mathcal{M}$ in the classical sense,
- $\mathcal{E}$ is a submeasure of $\mathcal{D}$.

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

Definition 4.4. 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 $\varphi = \varphi(x_1, \ldots, x_n)$ and sequences $a_1, \ldots, a_n \in \mathcal{N}$ we have:

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

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

The next example shows that there are satisfiable sentences without any countable model.

Example 4.5. 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, φ does not have any countable 0-models. Namely, if $(\mathcal{M}, \mathcal{D}) \models_0 \varphi$ 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. Since $x \in B_x$, the set $\bigcup_{x \in \mathcal{M}} B_x$ equals $\mathcal{M}$, and therefore has measure 1. But if $\mathcal{M}$ is countable it is also the union of countable many sets of measure 0 and hence has measure 0, a contradiction.

Note also that φ is *finitely* ε-satisfiable (i.e. ε-satisfiable with a finite model) for every $\varepsilon > 0$.

Using the reduction from Theorem 7.1, we now also find for every rational $\varepsilon \in [0, 1)$ a sentence φ_ε which is only ε-satisfiable in uncountable models.

Example 4.5 shows that we cannot always find countable elementary submodels. However, we can find such submodels of cardinality 2^ω, as we will show next.

Theorem 4.6. (Downward Löwenheim-Skolem theorem for ε-logic) *Let $\mathcal{L}$ be a countable first-order language, possibly containing equality, but 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 an ε-model*

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

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

Proof. We start by fixing some model-theoretic notation. For basics about types we refer the reader to Hodges [10]. For an element $x \in \mathcal{M}$, let $\mathrm{tp}(x/\mathcal{M})$ denote the complete 1-type of x over $\mathcal{M}$, i.e. the set of all formulas $\varphi(z)$ in one free variable and with parameters from $\mathcal{M}$ such that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(x)$. Clearly the relation $\mathrm{tp}(x/\mathcal{M}) = \mathrm{tp}(y/\mathcal{M})$ defines an equivalence relation on $\mathcal{M}$. The idea is to construct $\mathcal{N}$ by picking one element from every equivalence class of a finer equivalence relation. We will show that we can do this in such a way that we need at most 2^ω many points. Furthermore, this finer equivalence relation induces a natural submeasure $\mathcal{E}$ of $\mathcal{D}$ on $\mathcal{N}$, which will turn $(\mathcal{N}, \mathcal{E})$ into the desired ε-submodel.

Let $R = R(x_1, \ldots, x_n)$ be a relation. By Definition 2.2 we have that $R^{\mathcal{M}}$ is a $\mathcal{D}^n$-measurable set. We can view the construction of the product σ-algebra on $\mathcal{M}^n$ as an inductive process over the countable ordinals: we start with n-fold products or boxes of $\mathcal{D}$-measurable edges, for successor ordinals $\alpha + 1$ we take countable unions and intersections of elements from α (we could also take complements, but this is not necessary) and for limit ordinals λ we take the union of sets constructed in steps $< \lambda$. Now the product σ-algebra is exactly the union of all the sets constructed in this way.

In particular, we see from this construction 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 Δ_R consisting of the $\mathcal{D}$-measurable sets occurring as edges of Cartesian products in this expression. Let Δ be $\bigcup_R \Delta_R$ (where the union includes equality, if it is in the language) together with $\{c^{\mathcal{M}}\}$ for every constant c.

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

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

Then we can check that points in E_a are equivalent, in the sense that for all $x, y \in E_a$ we have that $\mathrm{tp}(x/\mathcal{M}) = \mathrm{tp}(y/\mathcal{M})$. Namely, first we check that for any n-ary relation R and $z_1, \ldots, z_{n-1} \in \mathcal{M}$, $R^{\mathcal{M}}(x, z_1, \ldots, z_{n-1}) \Leftrightarrow R^{\mathcal{M}}(y, z_1, \ldots, z_{n-1})$. This follows by induction on the construction of $R^{\mathcal{M}}$ from Δ. The equivalence for arbitrary formulas from the 1-types then follows by induction over formulas in prenex normal form.

From each nonempty E_a, pick one point x_a, and define

$$\mathcal{N} = X \cup \{x_a \mid a \in 2^\omega\}.$$

Clearly, $\mathcal{N}$ has cardinality at most 2^ω. Finally, for each $\mathcal{D}$-measurable A such that

(5) $$\forall a \in 2^\omega \forall x, y \in E_a(x \in A \Leftrightarrow y \in A),$$

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

First, observe that $\mathcal{E}$ is well-defined. Namely, let $A \neq C$ be $\mathcal{D}$-measurable sets satisfying (5), say $x \in A$ and $x \notin C$. Let $a \in 2^\omega$ be such that $x \in E_a$. Then $x_a \in A$, but $x_a \notin C$. So $A \cap \mathcal{N} \neq C \cap \mathcal{N}$. Also, $\mathcal{E}$ is a probability measure since $\mathcal{D}$ is.

Next, we prove that $(\mathcal{N}, \mathcal{E}) \prec_\varepsilon (\mathcal{M}, \mathcal{D})$. It is clear that $\mathcal{N}$ is a submodel of $\mathcal{M}$ and that $\mathcal{E}$ is a submeasure of $\mathcal{D}$. We prove that $(\mathcal{N}, \mathcal{E})$ is an elementary ε-submodel. We use formula-induction on formulas in prenex normal form 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

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

For propositional formulas, this 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 $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(x, b_1, \ldots, b_n)$, and let $a \in 2^\omega$ be such that $x \in E_a$. Then, as explained above, x and x_a are equivalent, so we also have $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(x_a, b_1, \ldots, b_n)$. Using the induction hypothesis, we therefore find $(\mathcal{N}, \mathcal{E}) \models_\varepsilon \psi(x_a, b_1, \ldots, b_n)$. Since $x_a \in \mathcal{N}$ this implies that $(\mathcal{N}, \mathcal{E}) \models_\varepsilon \exists x \psi(x, b_1, \ldots, b_n)$.

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

$$B = \{x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(x, b_1, \ldots, b_n)\},$$
$$C = \{x \in \mathcal{N} \mid (\mathcal{N}, \mathcal{E}) \models_\varepsilon \psi(x, b_1, \ldots, b_n)\}.$$

Then by induction hypothesis we have

$$C = \{x \in \mathcal{N} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(x, b_1, \ldots, b_n)\}$$
$$= B \cap \mathcal{N}.$$

From this and the fact that B satisfies (5), we see that $\mathcal{E}(C) = \mathcal{D}(B)$, and hence

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

This concludes the induction.

It remains to check that $(\mathcal{N}, \mathcal{E})$ is an ε-model (see Definition 2.2). For every formula $\varphi = \varphi(x_1, \ldots, x_n)$ and every sequence $a_1, \ldots, a_{n-1} \in \mathcal{N}$ we have

$$B_\varphi := \{a_n \in \mathcal{N} \mid (\mathcal{N}, \mathcal{E}) \models_\varepsilon \varphi(a_1, \ldots, a_n)\}$$
$$= \{a_n \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(a_1, \ldots, a_n)\} \cap \mathcal{N}$$

and since the right-hand side is the intersection of $\mathcal{N}$ and a $\mathcal{D}$-measurable set satisfying (5), it follows that B_φ is $\mathcal{E}$-measurable. That relations in $\mathcal{N}$ are measurable follows directly from the construction; for constants c use the fact that $\{c\} \in \Delta$ 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 4.7. The proof given above uses the full measurability condition (2) from Definition 2.2. We remark that one can also prove the theorem without using that the relations are measurable, by following the proof of Keisler [15, Theorem 2.4.4]; in that case the language is also allowed to contain function symbols. However, we need the proof above to be able to derive Theorem 5.8 below.

By varying ε, we can easily see that in fact the following strengthening of Theorem 4.6 holds.

Theorem 4.8. (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 an ε-model $(\mathcal{N}, \mathcal{E})$ such that $(\mathcal{N}, \mathcal{E}) \prec_\varepsilon (\mathcal{M}, \mathcal{D})$ for all $\varepsilon \in A$, such that $X \subseteq \mathcal{N}$ and such that $\mathcal{N}$ is of cardinality at most 2^ω.*

Proof. This can be shown using the same proof as for Theorem 4.6.   □

5. Satisfiability and Lebesgue measure

The construction from the proof of Theorem 4.6 produces an unknown probability measure on 2^ω. However, we can say a bit more about the σ-algebra of measurable sets of $\mathcal{E}$ in that proof: for example, that it is countably generated. We will use this and other facts to show that every ε-satisfiable set Γ 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 Γ; the new model will in general satisfy more sentences.[1]

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

Proposition 5.1. *Let $\mathcal{M}$ be a first-order model that is a Polish space, and let $\mathcal{D}_0$ be a Borel probability measure on $\mathcal{M}$ such that all relations and functions are $\mathcal{D}_0^n$-measurable. Then all definable sets are analytic. In particular, if we let $\mathcal{D}$ be the completion of $\mathcal{D}_0$, then $(\mathcal{M}, \mathcal{D})$ is an ε-model for every $\varepsilon \in [0,1]$.*

Proof. 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, using

---

[1]This can happen even if Γ is already complete, i.e. if for every sentence φ at least one of $\varphi \in \Gamma$ and $\neg\varphi \in \Gamma$ holds: because of the paraconsistency of the logic, it could happen that both φ and $\neg\varphi$ hold in our new model, while only one of them is in Γ.

induction over the number of quantifiers of a formula in prenex normal form (see Proposition 2.5). Clearly, this holds for propositional formulas. For the existential quantifier, use that projections of analytic sets are analytic (which is clear from the definition of an analytic set, see e.g. Kechris [14, Definition 14.1]), and for the universal quantifier, this fact is expressed by the Kondô-Tugué theorem (see Kechris [14, Theorem 29.26]).

Since all definable sets are analytic, in particular the definable sets of dimension 1 are analytic and hence $\mathcal{D}$-measurable (see e.g. Bogachev [4, Theorem 1.10.5]). This proves the second claim (see Definition 2.2). $\qquad\square$

Proposition 5.2. *Let $\mathcal{L}$ be a countable first-order language not containing function symbols. Let Γ be an ε-satisfiable set of sentences. Then there exists an ε-model $(\mathcal{M}, \mathcal{D})$ on 2^ω which ε-satisfies Γ such that $\mathcal{D}$ is the completion of a Borel measure. Furthermore, the relations in $\mathcal{M}$ can be chosen to be Borel.*

Proof. Fix an ε-model ε-satisfying all sentences from Γ and apply Theorem 4.6 (with $X = \emptyset$) to find a model $(\mathcal{N}, \mathcal{E})$. Let $\Delta = \{B_0, B_1, \dots\}$ and E_a be as in the proof of Theorem 4.6, that is,

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

By construction of $\mathcal{N}$, each E_a contains at most one point of $\mathcal{N}$, namely x_a. So, the function $\pi : \mathcal{N} \to 2^\omega$ mapping each $x_a \in \mathcal{N}$ to a is injective.[2]

Now, define the 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 (different from equality). Write $R^\mathcal{N}$ as an infinitary expression using countable unions and intersections of Cartesian products of $\mathcal{E}$-measurable sets from $\{B \cap \mathcal{N} \mid B \in \Delta_R\}$ (see the definition of Δ_R and $\mathcal{E}$ in the proof of Theorem 4.6); 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 probability measure $\mathcal{D}_0$ on the Borel sets of $\mathcal{M} = 2^\omega$ by

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

Let $\mathcal{D}$ be the completion of $\mathcal{D}_0$. Then Proposition 5.1 tells us that $(\mathcal{M}, \mathcal{D})$ is an ε-model. Now it is easy to see that for all propositional formulas $\varphi(x_1, \dots, x_n)$ and all $x_1, \dots, x_n \in \mathcal{N}$ we have that $\mathcal{N} \models \varphi(x_1, \dots, x_n)$ if and only if $\mathcal{M} \models \varphi(\pi(x_1), \dots, \pi(x_n))$. For the atomic formulas not using equality, this follows from the definition of the relations, and for equality this follows from the injectivity of π. For general formulas in prenex normal

---

[2]The idea of sending each x_a to $a \in 2^\omega$ also appears in Bogachev [4, Theorem 9.4.7], albeit in a different context. However, there only the case in which the function π is also surjective is discussed, the nonsurjective case being irrelevant in that context.

form, we can now easily prove the implication from left to right, i.e. that $(\mathcal{N}, \mathcal{E}) \models_\varepsilon \varphi(x_1, \ldots, x_n)$ implies that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(\pi(x_1), \ldots, \pi(x_n))$. Note that for universal quantifiers, going from $\mathcal{N}$ to $\mathcal{M}$, the set on which a formula holds can only increase in measure, and for existential quantifiers, there are at least as many witnesses in $\mathcal{M}$ as in $\mathcal{N}$. It follows from this that in particular $(\mathcal{M}, \mathcal{D})$ satisfies Γ. □

Next, we show that we can eliminate atoms.

Definition 5.3. Let μ be a measure and let x be a measurable singleton. We say that x is an *atom* of μ if $\mu(\{x\}) > 0$. The measure μ is called *atomless* if it does not have any atoms.

Often a different notion of atom is used in the literature, in which an atom is a measurable set A of strictly positive measure such that every measurable subset of A either has measure 0 or the same measure as A. However, in Polish spaces such as 2^ω these two notions coincide, as can be seen in e.g. Aliprantis and Border [1, Lemma 12.18].

Note that there are at most countably many atoms, since there can only be finitely many atoms of measure $\geqslant \frac{1}{n}$ for every $n \in \omega$.

Definition 5.4. 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*, denoted by $(\mathcal{M}, \mathcal{D}) \equiv_\varepsilon (\mathcal{N}, \mathcal{E})$, if for all sentences φ we have

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

Lemma 5.5. *Let $\mathcal{L}$ be a first-order language not containing equality or function symbols. Let $(\mathcal{M}, \mathcal{D})$ be an ε-model such that $\mathcal{D}$ is the completion of a Borel measure $\mathcal{D}_0$. Then there exists an atomless ε-model $(\mathcal{N}, \mathcal{E})$ such that $\mathcal{E}$ is the completion of a Borel measure $\mathcal{E}_0$ and $(\mathcal{N}, \mathcal{E}) \equiv_\varepsilon (\mathcal{M}, \mathcal{D})$. Furthermore, if $\mathcal{M}$ is Polish, then so is $\mathcal{N}$.*

Proof. We first show how to eliminate a single atom of $\mathcal{D}_0$. Let x_0 be an atom, say of measure r. Let $\mathcal{N}$ be the disjoint union of $\mathcal{M}$ and $[0, r]$. We define a new measure $\mathcal{E}_0$ on $\mathcal{N}$ by setting, for each $\mathcal{D}_0$-measurable $B \subseteq \mathcal{M}$ and Borel $C \subseteq [0, r]$,

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

where μ denotes the Lebesgue measure, restricted to Borel sets. Then clearly, x_0 is no longer an atom (since it now has measure zero). The interpretation of constants in $\mathcal{N}$ is simply defined by $c^\mathcal{N} = c^\mathcal{M}$.

Define a function $\pi : \mathcal{N} \to \mathcal{M}$ by

$$\pi(x) = \begin{cases} x & \text{if } x \in M, \\ x_0 & \text{if } x \in [0, r]. \end{cases}$$

We now define the relations on $\mathcal{N}$ by letting

$$R^\mathcal{N}(x_1, \ldots, x_n) \Longleftrightarrow R^\mathcal{M}(\pi(x_1), \ldots, \pi(x_n))$$

for every relation R. To simplify notation, in the following we write $\vec{x} = x_1, \ldots, x_n$ and $\pi(\vec{x}) = \pi(x_1), \ldots, \pi(x_n)$.

Let $\mathcal{E}$ be the completion of $\mathcal{E}_0$. To show that $(\mathcal{N}, \mathcal{E}) \equiv_\varepsilon (\mathcal{M}, \mathcal{D})$ we prove the stronger assertion that for all formulas φ and for all $\vec{x} \in \mathcal{N}$,

$$(6) \qquad (\mathcal{N}, \mathcal{E}) \models_\varepsilon \varphi(\vec{x}) \Longleftrightarrow (\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(\pi(\vec{x})),$$

so that this holds in particular for all sentences φ. We prove (6) by formula-induction on φ in prenex normal form. We only prove the universal case; the other cases are easy. So, let $\varphi = \forall y \psi(y, \vec{x})$. To prove (6) it suffices to show that

$$\Pr_{\mathcal{E}}\big[b \in \mathcal{N} \mid (\mathcal{N}, \mathcal{E}) \models_\varepsilon \psi(b, \vec{x})\big] = \Pr_{\mathcal{D}}\big[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a, \pi(\vec{x}))\big]$$

for all $\vec{x} \in \mathcal{N}$. Now by induction hypothesis,

$$\Pr_{\mathcal{E}}\big[b \in \mathcal{N} \mid (\mathcal{N}, \mathcal{E}) \models_\varepsilon \psi(b, \vec{x})\big]$$
$$= \Pr_{\mathcal{E}}\big[b \in \mathcal{N} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(\pi(b), \pi(\vec{x}))\big]$$
$$= \Pr_{\mathcal{E}}\big[b \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(\pi(b), \pi(\vec{x}))\big]$$
$$\qquad\qquad\qquad + \Pr_{\mathcal{E}}\big[b \in [0, r] \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(\pi(b), \pi(\vec{x}))\big]$$
$$= \Pr_{\mathcal{D}}\big[a \in \mathcal{M} \setminus \{x_0\} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a, \pi(\vec{x}))\big]$$
$$\qquad\qquad\qquad + \Pr_{\mathcal{D}}\big[x_0 \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(x_0, \pi(\vec{x}))\big]$$
$$= \Pr_{\mathcal{D}}\big[a \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_\varepsilon \psi(a, \pi(\vec{x}))\big]$$

Finally, using the remark below Definition 5.3, we can iterate this construction a countable number of times, eliminating the atoms one by one. It should be clear that the limit model exists and satisfies the theorem.   □

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

Definition 5.6. Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be measure spaces. $(X, \mathcal{A}, \mu)$ 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 $\mu \circ f^{-1} = \nu$.

Theorem 5.7. (Kechris [14, Theorem 17.41]) *Let $\mathcal{D}$ be an atomless Borel probability measure on a Polish space. Then it is isomorphic to $[0, 1]$ with the Lebesgue measure restricted to Borel sets.*

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

Theorem 5.8. *Let $\mathcal{L}$ be a countable first-order language not containing equality or function symbols. Let Γ be an ε-satisfiable set of sentences. Then there exists an ε-model on $[0, 1]$ with the Lebesgue measure which ε-satisfies Γ. Furthermore, all relations in the new ε-model can be chosen to be Borel.*

Proof. By Proposition 5.2 we know that Γ has an ε-model $(\mathcal{M}, \mathcal{D})$ on 2^ω with $\mathcal{D}$ the completion of a Borel measure and with Borel relations. By Lemma 5.5 we may assume that $\mathcal{M}$ is atomless. By Theorem 5.7, $(\mathcal{M}, \mathcal{D})$ is isomorphic to $[0, 1]$ with the Lebesgue measure, by a Borel isomorphism. Note that isomorphisms preserve ε-truth, so the induced model on $[0, 1]$ is ε-elementary equivalent to $\mathcal{M}$. $\qquad\square$

Note that Theorem 5.8 fails for languages with equality, since then a sentence such as $\exists x \forall y (x = y)$ expresses that there is an atom of measure at least $1 - \varepsilon$. For languages with equality we have the following more elaborate result:

Theorem 5.9. *Let $\mathcal{L}$ be a countable first-order language not containing function symbols. Let Γ be an ε-satisfiable set of sentences. Then there exists an ε-model $(\mathcal{M}, \mathcal{D})$ which ε-satisfies Γ such that:*

(i) *$\mathcal{M}$ is based on $[0, r] \cup X$ for some $r \in [0, 1]$ and a countable set X,*
(ii) *$\mathcal{D}$ is the Lebesgue measure on $[0, r]$,*
(iii) *all relations in $\mathcal{M}$ are Borel.*

Proof. As in the proof of Theorem 5.8, by Proposition 5.2 we know that Γ has an ε-model $(\mathcal{N}, \mathcal{E})$ on 2^ω with $\mathcal{E}$ the completion of a Borel measure and with Borel relations. We can separate $\mathcal{N}$ into a countable set of atoms X and an atomless part Y. Let r be the $\mathcal{E}$-measure of the atomless part. If $r = 0$, then observe that we can apply the classical Downwards Löwenheim-Skolem Theorem to find a countable submodel $\mathcal{M}$ of $\mathcal{N}$ containing X, and then $(\mathcal{M}, \mathcal{E} \restriction \mathcal{M})$ can be easily verified to satisfy the requirements of the theorem. If $r > 0$, then by Theorem 5.7 we know that $(Y, \frac{1}{r}\mathcal{E})$ is isomorphic to $[0, 1]$ with the Lebesgue measure, so $(Y, \mathcal{E})$ is isomorphic to $[0, r]$ with the Lebesgue measure. As in Theorem 5.8, this induces an ε-elementary equivalent model on $[0, r]$ together with the atoms X. $\qquad\square$

6. The Löwenheim Number

At this point we may ask ourselves how tight Theorem 4.6 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, let λ_ε be the Löwenheim number 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 parallels Corollary 2.4.5 in Keisler [15]. MA is Martin's axiom from set theory, see Kunen [16].

Theorem 6.1. *Let $\varepsilon \in [0, 1)$ be rational. For the Löwenheim number λ_ε of ε-logic we have*

(i) *$\aleph_1 \leqslant \lambda_\varepsilon \leqslant 2^{\aleph_0}$,*
(ii) *If Martin's axiom MA holds then $\lambda_\varepsilon = 2^{\aleph_0}$.*

Proof. The first part was already proven above, in Example 4.5 and Theorem 4.6. For the second part, assume that MA holds. Let φ be the sentence from Example 4.5. 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, 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 it has measure 0 (see Fremlin [6, p127]). But then

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

a contradiction.   □

Theorem 6.1 shows that we cannot prove in the standard set-theoretic framework of ZFC such statements as $\lambda_\varepsilon = \aleph_1$, because this is independent of ZFC. Namely under CH we have that $\lambda_\varepsilon = \aleph_1$ by item (i), and under MA we have $\lambda_\varepsilon = 2^{\aleph_0}$ by item (ii), and MA is consistent with $2^{\aleph_0} > \aleph_1$, see Kunen [16, p278]. (Note that this does not exclude the possibility that $\lambda_\varepsilon = 2^{\aleph_0}$ could be provable within ZFC.) So in this sense Theorem 4.6 is optimal.

7. REDUCTIONS

In this section we discuss many-one reductions between the sets of ε-satisfiable formulas for various ε. Recall that a many-one reduction between two sets A, B of formulas is a computable function f such that for all formulas φ we have that $\varphi \in A$ if and only if $f(\varphi) \in B$. The reductions we present below are useful, e.g. in our discussion of compactness in section 8, and also for complexity issues not discussed here.

In what follows we will need to talk about satisfiability of formulas rather than sentences. We will call a formula $\varphi(x_1, \ldots, x_n)$ ε-*satisfiable* if there is an ε-model $(\mathcal{M}, \mathcal{D})$ and elements $a_1, \ldots, a_n \in \mathcal{M}$ such that $(\mathcal{M}, \mathcal{D}) \models_\varepsilon \varphi(a_1, \ldots, a_n)$.

Theorem 7.1. *Let $\mathcal{L}$ be a countable first-order language not containing function symbols or equality. Then, for all rationals $0 \leqslant \varepsilon_0 \leqslant \varepsilon_1 < 1$, there exists a language $\mathcal{L}'$ (also not containing function symbols or equality) such that ε_0-satisfiability in $\mathcal{L}$ many-one reduces to ε_1-satisfiability in $\mathcal{L}'$.*

Proof. We can choose integers $a > 0$, n, and $m \leqslant n$ so that $\varepsilon_0 = 1 - \frac{a}{m}$ and $\varepsilon_1 = 1 - \frac{a}{n}$, and hence $\frac{m}{n} = \frac{1-\varepsilon_1}{1-\varepsilon_0}$. Let $\varphi(y_1, \ldots, y_k)$ be a formula in prenex normal form (see Proposition 2.5). For simplicity we write $\vec{y} = y_1, \ldots, y_k$. Also, for a function π we let $\pi(\vec{y})$ denote the vector $\pi(y_1), \ldots, \pi(y_k)$. We use formula-induction to define a computable function f such that for every formula φ,

(7) φ is ε_0-satisfiable if and only if $f(\varphi)$ is ε_1-satisfiable.

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 $\varphi = \forall x \psi(\vec{y}, x)$. The idea is to introduce new unary predicates, that can be used to vary the strength of the universal quantifier. We will make these predicates split the model into disjoint parts. If we split it into

just the right number of parts (in this case n), then we can choose m of these parts to get just the right strength.

So, we introduce new unary predicates $X_1, \ldots, X_n$. For $1 \leqslant i \leqslant n$, define

$$Y_i(x) = X_i(x) \wedge \bigwedge_{j \neq i} \neg X_j(x).$$

Then the predicates Y_i define disjoint sets in any model.

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

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

where $\#I$ denotes the cardinality of I. Then one can verify that in any model, if 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-valid. Conversely, if a-n-split holds, then the sets Y_i all have measure $\frac{1}{n}$ by Lemma 7.2 below. 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

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

where the conjunction is over all subsets of size m from $\{1, \ldots, n\}$. (It will be clear from the construction that $f(\psi)$ has the same arity as ψ.) Thus, $f(\varphi)$ expresses that for any choice of m of the n parts, $f(\psi)(x)$ holds often enough when restricted to the resulting part of the model.

We will now prove claim (7) above. For the implication from left to right, we will prove the following strengthening:

For every formula $\varphi(\vec{y})$, if φ is ε_0-satisfied in some ε_0-model $(\mathcal{M}, \mathcal{D})$, then there exists an ε_1-model $(\mathcal{N}, \mathcal{E})$ together with a measure-preserving surjective measurable function $\pi : \mathcal{N} \to \mathcal{M}$ (i.e. for all $\mathcal{D}$-measurable A we have that $\mathcal{E}(\pi^{-1}(A)) = \mathcal{D}(A)$) such that for all $\vec{y} \in \mathcal{N}$ we have that

$$(\mathcal{N}, \mathcal{E}) \models_{\varepsilon_1} f(\varphi)(\vec{y}) \text{ if and only if } (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} \varphi(\pi(\vec{y})).$$

We prove this by formula-induction over the formulas in prenex normal form. For propositional formulas, there is nothing to be done (we can simply take the models to be equal and π the identity). For the existential quantifier, let $\varphi = \exists x \psi(x)$ and apply the induction hypothesis to ψ to find a model $(\mathcal{N}, \mathcal{E})$ and a mapping π. Then we can take the same model and mapping for φ, as easily follows from the fact that π is surjective.

Next, we consider the universal quantifier. Suppose $\varphi = \forall x \psi(\vec{y}, x)$ is ε_0-satisfied in $(\mathcal{M}, \mathcal{D})$. Use the induction hypothesis to find a model $(\mathcal{N}, \mathcal{E})$ and a measure-preserving surjective measurable function $\pi : \mathcal{N} \to \mathcal{M}$ such that for all $\vec{y}, x \in \mathcal{N}$ we have that

$$(\mathcal{N}, \mathcal{E}) \models_{\varepsilon_1} f(\psi)(\vec{y}, x) \text{ if and only if } (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} \psi(\pi(\vec{y}), \pi(x)).$$

Now form the model $(\mathcal{N}', \mathcal{E}')$ consisting of n copies $\mathcal{N}_1, \ldots, \mathcal{N}_n$ of $(\mathcal{N}, \mathcal{E})$, each with weight $\frac{1}{n}$. That is, $\mathcal{E}'$ is the sum of n copies of $\frac{1}{n} \mathcal{E}$. Let $\pi' : \mathcal{N}' \to \mathcal{M}$ be the composition of the projection map $\sigma : \mathcal{N}' \to \mathcal{N}$ with π. Relations

in $\mathcal{N}'$ are defined just as on $\mathcal{N}$, that is, for a t-ary relation R we define $R^{\mathcal{N}'}(x_1,\ldots,x_t)$ by $R^{\mathcal{N}}(\sigma(x_1),\ldots,\sigma(x_t))$. Observe that this is the same as defining $R^{\mathcal{N}'}(x_1,\ldots,x_t)$ by $R^{\mathcal{M}}(\pi'(x_1),\ldots,\pi'(x_t))$. We interpret constants $c^{\mathcal{N}'}$ by embedding $c^{\mathcal{N}}$ into the first copy $\mathcal{N}_1$. Finally, we let each X_i be true exactly on the copy $\mathcal{N}_i$.

Then π' is clearly surjective. To show that it is measure-preserving, it is enough to show that σ is measure-preserving. If A is $\mathcal{E}$-measurable, then $\sigma^{-1}(A)$ consists of n disjoint copies of A, each having measure $\frac{1}{n}\mathcal{E}(A)$, so $\pi^{-1}(A)$ has $\mathcal{E}'$-measure exactly $\mathcal{E}(A)$.

Now, since $(\mathcal{N}',\mathcal{E}')$ satisfies a-n-split, we see that

$$(8) \qquad (\mathcal{N}',\mathcal{E}') \models_{\varepsilon_1} f(\psi)(\vec{y})$$

is equivalent to the statement that for all $1 \leqslant i_1 < \cdots < i_m \leqslant n$ we have

$$(9) \ \Pr_{\mathcal{E}'}\big[x \in \mathcal{N}' \mid (\mathcal{N}',\mathcal{E}') \models_{\varepsilon_1} (Y_{i_1}(x) \vee \cdots \vee Y_{i_m}(x)) \wedge f(\psi)(\vec{y},x)\big] \geqslant 1 - \varepsilon_1.$$

By Lemma 7.3 below we have that $(\mathcal{N}',\mathcal{E}') \models_{\varepsilon_1} f(\psi)(\vec{y},x)$ if and only if $(\mathcal{N},\mathcal{E}) \models_{\varepsilon_1} f(\psi)(\sigma(\vec{y}),\sigma(x))$. In particular, we see for every $1 \leqslant i \leqslant n$ that

$$(10) \qquad \Pr_{\mathcal{E}'}\big[x \in \mathcal{N}' \mid (\mathcal{N}',\mathcal{E}') \models_{\varepsilon_1} Y_i(x) \wedge f(\psi)(\vec{y},x)\big] =$$

$$\frac{1}{n}\Pr_{\mathcal{E}}\big[x \in \mathcal{N} \mid (\mathcal{N},\mathcal{E}) \models_{\varepsilon_1} f(\psi)(\sigma(\vec{y}),x)\big].$$

It follows that (9) is equivalent to

$$\frac{m}{n}\Pr_{\mathcal{E}}\big[x \in \mathcal{N} \mid (\mathcal{N},\mathcal{E}) \models_{\varepsilon_1} f(\psi)(\sigma(\vec{y}),x)\big] \geqslant 1 - \varepsilon_1.$$

The induction hypothesis tells us that this is equivalent to

$$\frac{m}{n}\Pr_{\mathcal{E}}\big[x \in \mathcal{N} \mid (\mathcal{M},\mathcal{D}) \models_{\varepsilon_0} \psi(\pi(\sigma(\vec{y})),\pi(x))\big] \geqslant 1 - \varepsilon_1$$

and since π is surjective and measure-preserving, this is the same as

$$\Pr_{\mathcal{D}}\big[x \in \mathcal{M} \mid (\mathcal{M},\mathcal{D}) \models_{\varepsilon_0} \psi(\pi(\sigma(\vec{y})),x)\big] \geqslant \frac{n}{m}(1-\varepsilon_1) = 1 - \varepsilon_0.$$

This proves that $(\mathcal{N}',\mathcal{E}') \models_{\varepsilon_1} f(\varphi)(\vec{y})$ if and only if $(\mathcal{M},\mathcal{D}) \models_{\varepsilon_0} \varphi(\pi'(\vec{y}))$.

We have not yet explained why $(\mathcal{N}',\mathcal{E}')$ is actually an ε_1-model. However, by Theorem 5.8 we may assume the relations on the original model $\mathcal{M}$ to be Borel, and it is easily seen that our construction of successively making copies keeps the relations Borel. So, from Proposition 5.1 we see that the models $(\mathcal{N},\mathcal{E})$ and $(\mathcal{N}',\mathcal{E}')$ are in fact ε-models for *every* ε.

To prove the right to left direction of (7) we will use induction to prove the following stronger statement:

If $(\mathcal{M},\mathcal{D})$ is an ε_1-model and $\vec{y} \in \mathcal{M}$ are such that $(\mathcal{M},\mathcal{D}) \models_{\varepsilon_1} f(\varphi)(\vec{y})$, then we also have $(\mathcal{M},\mathcal{D}) \models_{\varepsilon_0} \varphi(\vec{y})$.[3]

In particular, if $f(\varphi)$ is ε_1-satisfiable, then φ is ε_0-satisfiable. The only interesting case is the universal case, so let $\varphi = \forall x \psi(\vec{y},x)$. Let $\vec{y} \in \mathcal{M}$

---

[3]As explained above, we may assume $(\mathcal{M},\mathcal{D})$ to be an ε_0-model.

be such that $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} f(\varphi)(\vec{y})$. Assume, towards a contradiction, that $(\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon_0} \varphi(\vec{y})$. Then

$$\Pr_{\mathcal{D}}\big[x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon_0} \psi(\vec{y}, x)\big] > \varepsilon_0$$

and by the induction hypothesis we have

(11) $$\Pr_{\mathcal{D}}\big[x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon_1} f(\psi)(\vec{y}, x)\big] > \varepsilon_0.$$

But 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

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

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

Lemma 7.2. *Let $a, n \in \omega$, $a > 0$, and let $\mathcal{D}$ be a probability measure. Let $Y_1, \ldots, Y_n$ be disjoint $\mathcal{D}$-measurable sets such that for all subsets $I \subseteq \{1, \ldots, n\}$ of cardinality a the union $\bigcup_{i \in I} Y_i$ has measure at least $\frac{a}{n}$. Then all Y_i have measure exactly $\frac{1}{n}$.*

Proof. Assume there exists $1 \leqslant i \leqslant n$ such that Y_i has measure $< \frac{1}{n}$. Determine a sets Y_i with minimal measure, say with indices from the set I. Then, by assumption, $\mathcal{D}(\bigcup_{i \in I} Y_i) \geqslant \frac{a}{n}$. 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, $\mathcal{D}(\bigcup_{i \notin I} Y_i) \leqslant \frac{n-a}{n}$, so there is a Y_j with $j \notin I$ having measure $\leqslant \frac{1}{n}$. 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 $\frac{1}{n}$. $\qquad\square$

Lemma 7.3. *Let $(\mathcal{N}', \mathcal{E}')$ and $(\mathcal{N}, \mathcal{E})$ be as in the proof of Theorem 7.1 above. Then for every formula $\zeta(x_1, \ldots, x_t)$ in the language of $\mathcal{M}$, for every $\varepsilon \in [0, 1]$ and all $x_1, \ldots, x_t \in \mathcal{N}'$ we have that $(\mathcal{N}', \mathcal{E}') \models_{\varepsilon} \zeta(x_1, \ldots, x_t)$ if and only if $(\mathcal{N}, \mathcal{E}) \models_{\varepsilon} \zeta(\sigma(x_1), \ldots, \sigma(x_t))$.*

Proof. By induction on the structure of the formulas in prenex normal form. The base case holds by definition of the relations in $\mathcal{N}'$. The only interesting induction step is the one for the universal quantifier. So, let $\zeta = \forall x_0 \zeta'(x_0, \ldots, x_t)$ and let $x_1, \ldots, x_t \in \mathcal{N}'$. Using the induction hypothesis, we find that the set $A = \{x_0 \in \mathcal{N}' \mid (\mathcal{N}', \mathcal{E}') \models_{\varepsilon} \zeta'(x_0, \ldots, x_t)\}$ is equal to the set $\{x_0 \in \mathcal{N}' \mid (\mathcal{N}, \mathcal{E}) \models_{\varepsilon} \zeta'(\sigma(x_0), \ldots, \sigma(x_t))\}$, which consists of n disjoint copies of the set $B = \{x_0 \in \mathcal{N} \mid (\mathcal{N}, \mathcal{E}) \models_{\varepsilon} \zeta'(x_0, \sigma(x_1), \ldots, \sigma(x_t))\}$; denote the copy of B living inside $\mathcal{N}_i$ by B_i. Then

$$\mathcal{D}(A) = \sum_{i=1}^{n} \mathcal{E}'(B_i) = \sum_{i=1}^{n} \tfrac{1}{n} \mathcal{E}(B) = \mathcal{E}(B)$$

from which we directly see that $(\mathcal{N}', \mathcal{E}') \models_{\varepsilon} \zeta(x_1, \ldots, x_t)$ if and only if $(\mathcal{N}, \mathcal{E}) \models_{\varepsilon} \zeta(\sigma(x_1), \ldots, \sigma(x_t))$. $\qquad\square$

Next we show that we have a similar reduction in the other direction, i.e. from a bigger ε_0 to a smaller ε_1.

Theorem 7.4. *Let $\mathcal{L}$ be a countable first-order language not containing function symbols or equality. Then, for all rationals $0 < \varepsilon_1 \leqslant \varepsilon_0 \leqslant 1$, there exists a language $\mathcal{L}'$ (also not containing function symbols or equality) such that ε_0-satisfiability in $\mathcal{L}$ many-one reduces to ε_1-satisfiability in $\mathcal{L}'$.*

Proof. The proof is very similar to that of Theorem 7.1. There are two main differences: the choice of the integers a, n and m, and a small difference in the construction of f (as we shall see below). We can choose integers a, n and m such that $\varepsilon_1 = 1 - \frac{a}{n}$ and $\frac{m}{n} = \frac{\varepsilon_0 - \varepsilon_1}{\varepsilon_0}$. The case $a = 0$ is trivial, so we may assume that $a > 0$. We construct a many-one reduction f such that for all formulas φ,

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

Again, we only consider the nontrivial case where φ is a universal formula $\forall x \psi(\vec{y}, x)$. We define $f(\varphi)$ to be the formula

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

where the conjunction is over all subsets of size m from $\{1, \ldots, n\}$. So, the essential change from the proof of Theorem 7.1 is that the conjunction of $Y_{i_1}(x) \vee \cdots \vee Y_{i_m}(x)$ and $f(\psi)(\vec{y}, x)$ has become a disjunction.

The remainder of the proof is now almost the same as for Theorem 7.1. In the proof for the implication from left to right, follow the proof up to equation (8), i.e.

$$(\mathcal{N}', \mathcal{E}') \models_{\varepsilon_1} f(\varphi)(\vec{y}).$$

Again this is equivalent to the statement that for all $1 \leqslant i_1 < \cdots < i_m \leqslant n$ we have

$$(12) \quad \Pr_{\mathcal{E}'}\big[x \in \mathcal{N}' \mid (\mathcal{N}', \mathcal{E}') \models_{\varepsilon_1} Y_{i_1}(x) \vee \cdots \vee Y_{i_m}(x) \vee f(\psi)(\vec{y}, x) \big] \geqslant 1 - \varepsilon_1.$$

Similar as before, using Lemma 7.3, we find that this is equivalent to

$$\frac{m}{n} + \frac{n - m}{n} \Pr_{\mathcal{E}}\big[x \in \mathcal{N} \mid (\mathcal{N}, \mathcal{E}) \models_{\varepsilon_1} f(\psi)(\sigma(\vec{y}), x) \big] \geqslant 1 - \varepsilon_1.$$

Again, using the induction hypothesis and the fact that π is measure-preserving we find that this is equivalent to

$$\Pr_{\mathcal{D}}\big[x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} \psi(\pi(\sigma(\vec{y})), x) \big]$$

$$\geqslant \frac{n}{n - m}\left(1 - \varepsilon_1 - \frac{m}{n} \right) = \frac{\varepsilon_0}{\varepsilon_1}\left(\frac{-\varepsilon_0 \varepsilon_1 + \varepsilon_1}{\varepsilon_0} \right) = 1 - \varepsilon_0.$$

This proves that $(\mathcal{N}', \mathcal{E}') \models_{\varepsilon_1} f(\varphi)(\vec{y})$ if and only if $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon_0} \varphi(\vec{y})$.

For the converse implication, we also need to slightly alter the proof of Theorem 7.1. Assuming that $(\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} f(\varphi)(\vec{y})$, follow the proof up to equation (11), where we obtain

$$(13) \qquad \Pr_{\mathcal{D}}\big[x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \not\models_{\varepsilon_1} f(\psi)(\vec{y}, x) \big] > \varepsilon_0.$$

Define

$$\eta = \Pr_{\mathcal{D}}\big[x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} f(\psi)(\vec{y}, x)\big]$$

and take those m of the Y_i (say $Y_{i_1}, \ldots, Y_{i_m}$) which have the smallest intersection with this set. Note that by (13) we have $\eta < 1 - \varepsilon_0$. Then we find that

$$\Pr_{\mathcal{D}}\big[x \in \mathcal{M} \mid (\mathcal{M}, \mathcal{D}) \models_{\varepsilon_1} Y_{i_1} \vee \cdots \vee Y_{i_m} \vee f(\psi)(\vec{y}, x)\big]$$

$$\leqslant \frac{m}{n} + \Big(1 - \frac{m}{n}\Big)\eta = \frac{\varepsilon_0 - \varepsilon_1}{\varepsilon_0} + \frac{\varepsilon_1}{\varepsilon_0}\eta$$

$$< \frac{\varepsilon_0 - \varepsilon_1}{\varepsilon_0} + \frac{\varepsilon_1}{\varepsilon_0}(1 - \varepsilon_0) = 1 - \varepsilon_1.$$

which contradicts our choice of $(\mathcal{M}, \mathcal{D})$. Note that the last inequality holds because $\varepsilon_1 \neq 0$. $\qquad\square$

8. Compactness

We start this section with a negative result, namely that in general ε-logic is not compact. This results holds for rational ε different from 0 and 1. The case $\varepsilon = 1$ is pathological, and the case $\varepsilon = 0$ will be discussed elsewhere.

First, we prove a technical lemma.

Lemma 8.1. *Let $(\mathcal{M}, \mathcal{D})$ be an ε-model, let $R(x, y)$ 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 $y' \in C_y$:*

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

Proof. It suffices to 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 show that we can approximate the horizontal sections of $R^{\mathcal{M}}$ in a suitable way, namely that there exist a finite number $U_1, \ldots, U_n$ and $V_1, \ldots, V_n$ of $\mathcal{D}$-measurable sets such that:

$$(14) \quad \Pr_{\mathcal{D}}\Big[y \in \mathcal{M} \,\Big|\, \Pr_{\mathcal{D}}\Big[u \in \mathcal{M} \mid (u, y) \in R^{\mathcal{M}} \triangle \Big(\bigcup_{i=1}^{n} U_i \times V_i\Big)\Big] \leqslant \frac{\delta}{2}\Big] \geqslant 1 - \delta'.$$

(Here $\triangle$ denotes the symmetric difference.) To show that this is possible, determine U_i and V_i such that

$$(15) \qquad\qquad \Pr_{\mathcal{D}}\Big[R^{\mathcal{M}} \triangle \Big(\bigcup_{i=1}^{n} U_i \times V_i\Big)\Big] < \frac{\delta\delta'}{2}.$$

Observe that such an approximation exists: this obviously holds for relations $R^{\mathcal{M}}$ that are a rectangle $U \times V$, and the existence of an approximation is preserved under countable unions and complements. (This is usually part of a proof of Fubini's theorem.)

Now assume that (14) does not hold for these U_i and V_i. Then we have

$$\Pr_{\mathcal{D}}\Big[y \in \mathcal{M} \,\Big|\, \Pr_{\mathcal{D}}\Big[u \in \mathcal{M} \mid (u,y) \in R^{\mathcal{M}} \triangle \Big(\bigcup_{i=1}^{n} U_i \times V_i\Big)\Big] > \frac{\delta}{2}\Big] > \delta'.$$

But then we see that

$$\Pr_{\mathcal{D}}\Big[R^{\mathcal{M}} \triangle \Big(\bigcup_{i=1}^{n} U_i \times V_i\Big)\Big] \geqslant \frac{\delta \delta'}{2}$$

by taking the integral, contradicting (15).

We now show that for almost all y there exists a set C_y of strictly positive measure such that for all $y' \in C_y$ we have that

$$(16) \qquad \Pr_{\mathcal{D}}\Big[u \in \mathcal{M} \mid (u,y) \in \bigcup_{i=1}^{n} U_i \times V_i \leftrightarrow (u,y') \in \bigcup_{i=1}^{n} U_i \times V_i\Big] = 1.$$

Note that the V_i induce a partition of $\mathcal{M}$ into at most 2^n many disjoint parts Y_j (by choosing for each $i \leqslant n$ 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 $C_y = Y_j$. Then it is clear that (16) holds.

Call the set of measure $1 - \delta'$ elements y from (14) $\mathcal{M}'$. The same argument used to prove (16) can be applied to $\mathcal{M}'$, using the partition $Y_j \cap \mathcal{M}'$ of $\mathcal{M}'$. For $y \in \mathcal{M}'$ this gives the same conclusion (16), but with the extra property that $C_y \subseteq \mathcal{M}'$.

Finally, the lemma follows by combining (14) and (16): For every $y \in \mathcal{M}'$, by (14) its sections with $R^{\mathcal{M}}$ and the approximation differ by at most $\frac{\delta}{2}$. By (16) and the previous remark there is $C_y \subseteq \mathcal{M}'$ of positive measure such that for every $y' \in C_y$ the sections of y and y' with the approximation agree almost everywhere. Again by (14) the sections of y' with $R^{\mathcal{M}}$ and the approximation differ by at most $\frac{\delta}{2}$. Hence the sections of y and y' with $R^{\mathcal{M}}$ differ by at most $\frac{\delta}{2} + \frac{\delta}{2}$.   □

Theorem 8.2. *For every rational $\varepsilon \in (0,1)$, ε-logic is not compact, i.e. there exists a countable set Γ of sentences such that each finite subset of Γ is ε-satisfiable, but Γ itself is not ε-satisfiable.*

Proof. The example we use is adapted from Keisler [15, Example 2.6.5]. Let R be a binary relation. Using the reductions from Theorem 7.1 (observing, from the proof of that theorem, 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 (i.e. measure 1 many), there exists a set A_y of measure at least $1 - \frac{1}{n}$ such that 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 $\frac{1}{2}$).

Then each φ_n has a finite ε-model, as illustrated in Figure 8.1 for $n = 4$: 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 colored grey. If we now take for each A_y exactly those three intervals of length $\frac{1}{4}$ of which y is not an element, we can directly verify that φ_n holds.
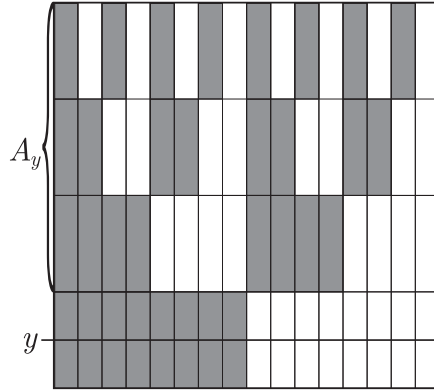

FIGURE 8.1. A model for φ_4 on $[0, 1]$.

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, Lemma 8.1 tells us that in any ε-model with a binary relation R, 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 necessarily $A_y \cap C_y = \emptyset$ this shows that A_y cannot have measure 1. □

Next, we will present an ultraproduct construction that allows us to partially recover compactness, which is due to Hoover and described in Keisler [15]. This construction uses the Loeb measure from nonstandard analysis, which is due to Loeb [17]. The same construction as in Keisler is also described (for a different logic) in Bageri and Pourmahdian [2], however, there the Loeb measure is not explicitly mentioned and the only appearance of nonstandard analysis is in taking the standard part of some element. Below we will describe the construction without resorting to nonstandard analysis. To be able to define the measure, we need the notion of a *limit over an ultrafilter*. This notion corresponds to taking the standard part of a nonstandard real number.

We refer the reader to Hodges [10] for an explanation of the notions of ultrafilter and ultraproduct.

Definition 8.3. 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 $\mathcal{U}$, or a $\mathcal{U}$-*limit*, 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 8.4. *Limits over ultrafilters are unique. Furthermore, for any bounded sequence and every ultrafilter $\mathcal{U}$, the limit of the sequence over $\mathcal{U}$ exists.*

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

$$\left\{i \in \omega \mid |a_i - r_0| < \tfrac{1}{2}|r_0 - r_1|\right\}$$

and

$$\left\{i \in \omega \mid |a_i - r_1| < \tfrac{1}{2}|r_0 - r_1|\right\}$$

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

Now, assume the sequence $a_0, a_1, \ldots$ 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

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

or

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

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

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

Using these limits over ultrafilters, we show how to define a probability measure on an ultraproduct of measure spaces. As mentioned above, this construction is essentially due to Loeb [17], but we describe it on ultraproducts instead of using nonstandard analysis.

Definition 8.5. Let $\mathcal{U}$ be a nonprincipal ultrafilter on ω and let $\mathcal{D}_0, \mathcal{D}_1, \ldots$ be a sequence of finitely additive probability measures over sets $X_0, X_1, \ldots$. Then we let $\prod_{i \in \omega} X_i / \mathcal{U}$ denote the ultraproduct, and for a sequence $a = a_0, a_1, \ldots$ with $a_i \in X_i$ we let $[a]$ denote the element of $\prod_{i \in \omega} X_i / \mathcal{U}$ corresponding to the equivalence class of the sequence $a_0, a_1, \ldots$.

For each sequence $A = A_0, A_1, \ldots$ with each A_i a $\mathcal{D}_i$-measurable set we will call the set

$$[A] = \left\{[a] \in \prod_{i \in \omega} X_i / \mathcal{U} \mid \{i \in \omega \mid a_i \in A_i\} \in \mathcal{U}\right\}$$

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

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

Note that $\mathcal{E}$ is a σ-additive measure, even if the $\mathcal{D}_i$ are only finitely additive.

Proposition 8.6. *The ultraproduct measure exists and is well-defined.*

Proof. We need to verify that $\mathcal{E}$, as defined on the Boolean algebra of basic measurable sets, satisfies the conditions of Carathéodory's extension theorem (see e.g. Bogachev [4, Theorem 1.5.6]). Thus, we need to show that, for any disjoint sequence $[A^0], [A^1], \ldots$ of nonempty basic measurable sets such that $\bigcup_{j\in\omega}[A^j]$ is a basic measurable set, we have that

$$\Pr_{\mathcal{E}}\Big(\bigcup_{j\in\omega}[A^j]\Big) = \sum_{j\in\omega}\Pr_{\mathcal{E}}([A^j]).$$

In fact, we will show that if the $[A^j]$ are disjoint and nonempty, then $\bigcup_{j\in\omega}[A^j]$ is never a basic measurable set.

Namely, let $[A^j]$ be as above and assume $\bigcup_{j\in\omega}[A^j]$ is a basic measurable set. Without loss of generality, we may assume that $\bigcup_{j\in\omega}[A^j] = \prod_{i\in\omega} X_i/\mathcal{U}$. We will construct an element of $\prod_{i\in\omega} X_i/\mathcal{U}$ which is not in $\bigcup_{j\in\omega}[A^j]$, which is a contradiction. Observe that, because all the $[A^j]$ are disjoint and nonempty, the set $\bigcup_{j=0}^n [A^j]$ will always be a proper subset of $\prod_{i\in\omega} X_i/\mathcal{U}$. So, for each $n \in \omega$, fix $[x^n] \notin \bigcup_{j=0}^n [A^j]$. For every $m \in \omega$, let $I_m \in \mathcal{U}$ be the set $\{i \in \omega \mid x_i^m \notin \bigcup_{j=0}^m A_i^j\}$. Furthermore, let $k_n \in \omega$ be the biggest $m \leqslant n$ such that $n \in I_m$, and let it be 0 if no such $m \leqslant n$ exists.

Now define x_i as $x_i^{k_i}$. We claim that $[x] \notin \bigcup_{j\in\omega}[A^j]$. Namely, let $m \in \omega$. Then, for every $n \geqslant m$ with $n \in I_m$ we have $k_n \geqslant m$, so we see that $x_n \notin \bigcup_{j=0}^m A_n^j$. In particular, we see that $x_n \notin A_n^m$ for every $n \in I_m \setminus \{0, 1, \ldots, n-1\}$. However, since $\mathcal{U}$ is nonprincipal, this set is in $\mathcal{U}$, so we see $[x] \notin [A^m]$. □

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 will see that we merely know that all definable subsets of $\mathcal{M}$ of arity 1 are measurable. We thus only obtain a *weak ε-model* (see Definition 3.3). To even achieve this, we need to extend the measure to all subsets of the model first, at the cost of moving to a finitely additive measure. The final measure that we obtain is still σ-additive though.

Theorem 8.7. (Tarski) *Every finitely additive measure $\mathcal{D}$ on a Boolean algebra of subsets of X can be extended to a finitely additive measure $\mathcal{D}'$ on the power set $\mathcal{P}(X)$.*

Proof. See Birkhoff [3, p. 185]. □

Definition 8.8. Let $\varepsilon \in [0,1]$, let $\mathcal{U}$ be a nonprincipal ultrafilter over ω and let $(\mathcal{M}_0, \mathcal{D}_0), (\mathcal{M}_1, \mathcal{D}_1), \ldots$ be a sequence of finitely additive weak ε-models, where each $\mathcal{D}_i$ is defined on all of $\mathcal{P}(\mathcal{M}_i)$ (e.g. using Theorem 8.7). 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 $R(x^1, \ldots, x^n)$ we

define the relation on $\prod_{i \in \omega}(\mathcal{M}_i, \mathcal{D}_i)/\mathcal{U}$ by

$$R([a^1], \ldots, [a^n]) \Leftrightarrow \left\{ i \in \omega \mid R^{\mathcal{M}_i}(a_i^1, \ldots, a_i^n) \right\} \in \mathcal{U},$$

and we interpret function symbols $f(x_1, \ldots, x_n)$ by

$$f([a^1], \ldots, [a^n]) = [f^{\mathcal{M}_0}(a_0^1, \ldots, a_0^n), f^{\mathcal{M}_1}(a_1^1, \ldots, a_1^n), \ldots].$$

In particular, constants c are interpreted as

$$c = [c^{\mathcal{M}_0}, c^{\mathcal{M}_1}, \ldots].$$

We can now show that a variant of the fundamental theorem of ultraproducts, or Łoś's theorem, holds for this kind of model.

Theorem 8.9. *For every formula $\varphi(x^1, \ldots, x^n)$ and every sequence of elements $[a^1], \ldots, [a^n] \in \prod_{i \in \omega}(\mathcal{M}_i, \mathcal{D}_i)/\mathcal{U}$, the following are equivalent:*
  (i) *$\prod_{i \in \omega}(\mathcal{M}_i, \mathcal{D}_i)/\mathcal{U} \models_\varepsilon \varphi([a^1], \ldots, [a^n])$,*
  (ii) *for all $\varepsilon' > \varepsilon$, $\{i \in \omega \mid (\mathcal{M}_i, \mathcal{D}_i) \models_{\varepsilon'} \varphi(a_i^1, \ldots, a_i^n)\} \in \mathcal{U}$,*[4]
  (iii) *there exists a sequence $\varepsilon_0, \varepsilon_1, \ldots$ with $\mathcal{U}$-limit ε such that $\{i \in \omega \mid (\mathcal{M}_i, \mathcal{D}_i) \models_{\varepsilon_i} \varphi(a_i^1, \ldots, a_i^n)\} \in \mathcal{U}$.*

In particular, if $\{i \in \omega \mid (\mathcal{M}_i, \mathcal{D}_i) \models_\varepsilon \varphi(a_i^1, \ldots, a_i^n)\} \in \mathcal{U}$, then we have $\prod_{i \in \omega}(\mathcal{M}_i, \mathcal{D}_i)/\mathcal{U} \models_\varepsilon \varphi([a^1], \ldots, [a^n])$.

Proof. Before we begin with the proof we note that for any formula ψ, if $\varepsilon_0 \leqslant \varepsilon_1$ and $(\mathcal{M}_i, \mathcal{D}_i) \models_{\varepsilon_0} \psi$, then also $(\mathcal{M}_i, \mathcal{D}_i) \models_{\varepsilon_1} \psi$. This can be directly shown using induction over formulas in prenex normal form, together with the fact that we have assumed every subset of $\mathcal{M}_i$ to be $\mathcal{D}_i$-measurable (see Definition 8.8).

We first prove the equivalence of (ii) and (iii). If (ii) holds, let

$$\delta_i = \inf \left\{ \varepsilon' > \varepsilon \mid (\mathcal{M}_i, \mathcal{D}_i) \models_{\varepsilon'} \varphi([a^1], \ldots, [a^n]) \right\}$$

if this set is nonempty, and 0 otherwise. Note that this set is $\mathcal{U}$-almost always nonempty, as can be seen by applying (ii) with an arbitrary $\varepsilon' > \varepsilon$. Furthermore, (ii) also tells us that the sequence δ_i converges to ε with respect to $\mathcal{U}$. Now, the sequence $\frac{1}{2^i}$ converges to 0, so that the sequence $\varepsilon_i = \min(\delta_i + \frac{1}{2^i}, 1)$ also has $\mathcal{U}$-limit ε. By definition of δ_i, we now have that (iii) holds for the sequence $\varepsilon_0, \varepsilon_1, \ldots$.

Conversely, assume that (iii) holds and fix $\varepsilon' > \varepsilon$. Because the sequence $\varepsilon_0, \varepsilon_1, \ldots$ has $\mathcal{U}$-limit ε, we know that $\{i \in \omega \mid |\varepsilon_i - \varepsilon| < \varepsilon' - \varepsilon\} \in \mathcal{U}$. Using (iii) we therefore see that also

$$\{i \in \omega \mid |\varepsilon_i - \varepsilon| < \varepsilon' - \varepsilon\} \cap \{i \in \omega \mid (\mathcal{M}_i, \mathcal{D}_i) \models_{\varepsilon_i} \varphi(a_i^1, \ldots, a_i^n)\} \in \mathcal{U}$$

and using the observation above we directly see that this set is contained in $\{i \in \omega \mid (\mathcal{M}_i, \mathcal{D}_i) \models_{\varepsilon'} \varphi(a_i^1, \ldots, a_i^n)\}$, so the latter set is also in $\mathcal{U}$ and we therefore see that (ii) holds.

---

[4]Here we also consider $\varepsilon' > 1$, which is interpreted in the same way as in Definition 2.1. Of course, this is not necessary if $\varepsilon < 1$, since ε'-truth is equivalent to 1-truth when $\varepsilon' > 1$. However, this way $\varepsilon = 1$ is also included.

Next, we simultaneously show the equivalence of (i) with (ii) and (iii) using induction over formulas φ in prenex normal form. For propositional formulas, this proceeds in the same way as the classical proof. For the existential case, we use formulation (iii): using the induction hypothesis, we know that

$$\prod_{i\in\omega}(\mathscr{M}_i,\mathscr{D}_i)/\mathscr{U}\models_\varepsilon \exists x\psi([a^1],\ldots,[a^n],x)$$

is equivalent to saying that there exists an $[a^{n+1}]\in\prod_{i\in\omega}(\mathscr{M}_i,\mathscr{D}_i)$ and a sequence $\varepsilon_0,\varepsilon_1,\ldots$ with $\mathscr{U}$-limit ε such that for $\mathscr{U}$-almost all $i\in\omega$ we have that $(\mathscr{M}_i,\mathscr{D}_i)\models_{\varepsilon_i}\psi(a_i^1,\ldots,a_i^n,a_i^{n+1})$, which is in turn equivalent to saying that for the same sequence $\varepsilon_0,\varepsilon_1,\ldots$ we have for $\mathscr{U}$-almost all $i\in\omega$ that $(\mathscr{M}_i,\mathscr{D}_i)\models_{\varepsilon_i}\exists x\psi(a_i^1,\ldots,a_i^n,x)$.

Finally, consider the universal case. By definition,

$$\prod_{i\in\omega}(\mathscr{M}_i,\mathscr{D}_i)/\mathscr{U}\models_\varepsilon \forall x\psi([a^1],\ldots,[a^n],x)$$

is equivalent to

$$\Pr_{\mathscr{E}}\left(\left\{[a^{n+1}]\,\Big|\,\prod_{i\in\omega}(\mathscr{M}_i,\mathscr{D}_i)/\mathscr{U}\models_\varepsilon \psi([a^1],\ldots,[a^{n+1}])\right\}\right)\geqslant 1-\varepsilon.$$

By induction hypothesis, we know that this is equivalent to

$$\Pr_{\mathscr{E}}\left(\bigcap_{\varepsilon'>\varepsilon}\left\{[a^{n+1}]\,\Big|\,\{i\in\omega\mid(\mathscr{M}_i,\mathscr{D}_i)\models_{\varepsilon'}\psi(a_i^1,\ldots,a_i^{n+1})\}\in\mathscr{U}\right\}\right)\geqslant 1-\varepsilon.$$

Because we can restrict ourselves to the (countable) intersection of rational ε', this is the same as having for all $\varepsilon'>\varepsilon$ that

$$(17)\quad \Pr_{\mathscr{E}}\left(\left\{[a^{n+1}]\,\Big|\,\{i\in\omega\mid(\mathscr{M}_i,\mathscr{D}_i)\models_{\varepsilon'}\psi(a_i^1,\ldots,a_i^{n+1})\}\in\mathscr{U}\right\}\right)\geqslant 1-\varepsilon.$$

Observe that the set in (17) is precisely the basic measurable set

$$\left[\{a_i^{n+1}\in\mathscr{M}_i\mid(\mathscr{M}_i,\mathscr{D}_i)\models_{\varepsilon'}\psi(a_i^1,\ldots,a_i^{n+1})\}\right],$$

where we once again use that every subset of $\mathscr{M}_i$ is $\mathscr{D}_i$-measurable. So, using the definition of the ultraproduct measure (Definition 8.5), (17) is equivalent to having for every $\delta>0$ that

$$\left\{i\in\omega\,\Big|\,\Pr_{\mathscr{D}_i}\left(\{a_i^{n+1}\in\mathscr{M}_i\mid(\mathscr{M}_i,\mathscr{D}_i)\models_{\varepsilon'}\psi(a_i^1,\ldots,a_i^{n+1})\}\right)\geqslant 1-\varepsilon-\delta\right\}\in\mathscr{U}.$$

But this holds for all $\varepsilon'>\varepsilon$ and all $\delta>0$ if and only if we have for all $\varepsilon'>\varepsilon$ that

$$\left\{i\in\omega\,\Big|\,\Pr_{\mathscr{D}_i}\left(\{a_i^{n+1}\in\mathscr{M}_i\mid(\mathscr{M}_i,\mathscr{D}_i)\models_{\varepsilon'}\psi(a_i^1,\ldots,a_i^{n+1})\}\right)\geqslant 1-\varepsilon'\right\}\in\mathscr{U}.$$

which is in turn equivalent to having for all $\varepsilon'>\varepsilon$ that

$$\{i\in\omega\mid(\mathscr{M}_i,\mathscr{D}_i)\models_{\varepsilon'}\forall x\psi(a_i^1,\ldots,a_i^n,x)\}\in\mathscr{U}.$$

This completes the induction. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Corollary 8.10. *The ultraproduct is a weak ε-model.*

Proof. For every formula $\varphi = \varphi(x_1, \ldots, x_n)$ and parameters $[a^1], \ldots, [a^{n-1}]$ in $\prod_{i \in \omega}(\mathcal{M}_i, \mathcal{D}_i)/\mathcal{U}$, Theorem 8.9 tells us that the subset of $\prod_{i \in \omega}(\mathcal{M}_i, \mathcal{D}_i)/\mathcal{U}$ defined by φ and the parameters $[a^1], \ldots [a^{n-1}]$ is exactly

$$\bigcap_{\varepsilon' > \varepsilon, \varepsilon' \in \mathbb{Q}} \left\{ [a^n] \in \prod_{i \in \omega}(\mathcal{M}_i, \mathcal{D}_i)/\mathcal{U} \mid \{i \in \omega \mid (\mathcal{M}_i, \mathcal{D}_i) \models_{\varepsilon'} \varphi(a_i^1, \ldots, a_i^n)\} \in \mathcal{U} \right\}$$

which is a countable intersection of basic measurable sets, and therefore measurable. $\qquad\square$

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 8.2 above.

Theorem 8.11. (Weak compactness theorem) *Let Γ be a countable set of sentences such that each finite subset is satisfied in a weak ε-model. Then there exists a weak ε-model satisfying Γ.*

Proof. Let $A_0, A_1, \ldots$ be an enumeration of the finite subsets of Γ. For each A_i, fix a weak ε-model $(\mathcal{M}_i, \mathcal{D}_i)$ satisfying all formulas from A_i. Then the filter on ω generated by

$$\left\{ \{i \in \omega \mid (\mathcal{M}_i, \mathcal{D}_i) \models_\varepsilon \varphi\} \mid \varphi \in \Gamma \right\}$$

is a proper filter, so we can use the ultrafilter lemma (see e.g. Hodges [10, Theorem 6.2.1]) to determine an ultrafilter $\mathcal{U}$ on ω containing this filter. If $\mathcal{U}$ is principal, then there exists an $n \in \omega$ with $\{n\} \in \mathcal{U}$ and therefore $(\mathcal{M}_n, \mathcal{D}_n)$ satisfies Γ. Otherwise we form the ultraproduct (where we note that we may assume every subset of $\mathcal{M}_i$ to be $\mathcal{D}_i$-measurable by Theorem 8.7, provided we only assume that $\mathcal{D}_i$ is finitely additive). It then follows from Theorem 8.9 and Corollary 8.10 that this ultraproduct is a weak ε-model that satisfies every $\varphi \in \Gamma$. $\qquad\square$

9. Future research

There are various open questions about the model theory of ε-logic, of which we want to mention a few. One open question was already mentioned in section 6: what is the Löwenheim number λ_ε of ε-logic?

Another interesting question is if Craig interpolation holds for ε-logic, and similarly for related properties such as Beth definability and Robinson consistency. The study of these is especially interesting, because even though one can easily derive one from the other in classical logic, those proofs often use compactness in an essential way. Since our logic is not compact, it could be the case that some of these properties hold for ε-logic, while others do

not. For a discussion of these properties for various extensions of first-order logic see for example Väänänen [23].

References

[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 (2009) 465–492.

[3] G. Birkhoff, *Lattice theory*, revised edition, American Mathematical Society, 1948.

[4] V. I. Bogachev, *Measure theory*, Springer, 2007.

[5] R. Carnap, *On inductive logic*, Philosophy of Science 12 (1945) 72–97.

[6] D. H. Fremlin, *Consequences of Martin's axiom*, Cambridge University Press, 1984.

[7] H. Gaifman, *Concerning measures in first order calculi*, Israel Journal of Mathematics 2 (1964) 1–18.

[8] J. Y. Halpern, *An analysis of first-order logics of probability*, Proceedings of 11th International Joint Conference on Artificial Intelligence (IJCAI-89) (1989) 1375–1381.

[9] H. Hansson and B. Jonsson, *A logic for reasoning about time and reliability*, Formal Aspects of Computing 6 (1994) 512–535.

[10] W. Hodges, *Model theory*, Cambridge University Press, 1993.

[11] C. Howson, *Can logic be combined with probability? Probably*, Journal of applied logic 7(2) (2009) 177–187.

[12] M. Jaeger, *A logic for inductive probabilistic reasoning*, Synthese 144(2) (2005) 181–248.

[13] M. J. Kearns, U. V. Vazirani, *An introduction to computational learning theory*, MIT Press, 1994.

[14] A. S. Kechris, *Classical descriptive set theory*, Springer, 1995.

[15] H. J. Keisler, *Probability quantifiers*, in: J. Barwise and S. Feferman (eds.), Model-Theoretic Logics, Springer-Verlag 1985, 509–556.

[16] K. Kunen, *Set theory: An introduction to independence proofs*, North-Holland, 1983.

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

[18] D. Scott and P. Krauss, *Assigning probabilities to logical formulas*, in: J. Hintikka and P. Suppes (eds.), Aspects of Inductive Logic, North Holland (1966) 219–264.

[19] C. I. Steinhorn, *Borel structures and measure and category logics*, in: J. Barwise and S. Feferman (eds.), Model-Theoretic Logics, Springer-Verlag 1985, 579–596.

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

[21] S. A. Terwijn, *Decidability and undecidability in probability logic*, Proceedings of Logical Foundations of Computer Science, Lecture Notes in Computer Science 5407, 441-450, Springer, 2009.

[22] L. G. Valiant, *Robust logics*, Artificial Intelligence 117 (2000) 231–253.

[23] J. Väänänen, *The Craig interpolation theorem in abstract model theory*, Synthese 164 (2008) 401–420.

(Rutger Kuyper) Radboud University Nijmegen, Department of Mathematics, P.O. Box 9010, 6500 GL Nijmegen, the Netherlands.
    E-mail address: r.kuyper@math.ru.nl

(Sebastiaan A. Terwijn) Radboud University Nijmegen, Department of Mathematics P.O. Box 9010, 6500 GL Nijmegen, the Netherlands.
    E-mail address: terwijn@math.ru.nl