MODEL THEORY OF PROBABILITY SPACES

ALEXANDER BERENSTEIN AND C. WARD HENSON

ABSTRACT. This expository paper treats the model theory of probability spaces using the framework of continuous [0,1]-valued first order logic. The metric structures discussed, which we call probability algebras, are obtained from probability spaces by identifying two measurable sets if they differ by a set of measure zero. The class of probability algebras is axiomatizable in continuous first order logic; we denote its theory by Pr. We show that the existentially closed structures in this class are exactly the ones in which the underlying probability space is atomless. This subclass is also axiomatizable; its theory APA is the model companion of Pr. We show that APA is separably categorical (hence complete), has quantifier elimination, is ω -stable, and has built-in canonical bases, and we give a natural characterization of its independence relation. For general probability algebras, we prove that the set of atoms (enlarged by adding 0) is a definable set, uniformly in models of Pr. We use this fact as a basis for giving a complete treatment of the model theory of arbitrary probability spaces. The core of this paper is an extensive presentation of the main model theoretic properties of APA. We discuss Maharam's structure theorem for probability algebras, and indicate the close connections between the ideas behind it and model theory. We show how probabilistic entropy provides a rank connected to model theoretic forking in probability algebras. In the final section we mention some open problems.

1. Introduction

In this paper we use the continuous version of first order logic to investigate probability spaces (X, \mathcal{B}, μ) . Here \mathcal{B} is a σ -algebra of subsets of X (requiring $\emptyset, X \in \mathcal{B}$) and μ is a σ -additive probability measure on \mathcal{B} . There is a canonical pseudometric d on \mathcal{B} , obtained by taking the distance between sets to be given by $d(A, B) := \mu(A \triangle B)$. (Here \triangle denotes the symmetric difference operation on sets.) This gives rise to a *prestructure*

$$(\mathcal{B}, 0, 1, \cdot^c, \cap, \cup, \mu, d)$$

(which we often write as (\mathcal{B}, μ, d) , regarding \mathcal{B} as a boolean algebra but suppressing the constants and the operations from our notation). We obtain a *structure* in the usual way by turning d into a metric. (Usually we would also need to take the metric completion, but

The authors are grateful to Itaï Ben Yaacov for helpful conversations. Research for this paper was partially supported by NSF grants and by grants from the Simons Foundation (202251 and 422088, to the second author).

the metric quotient is automatically complete here, as we indicate below.) This yields the structure $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$, where $\widehat{\mathcal{B}}$ is the quotient of \mathcal{B} by the equivalence relation $\mu(A \triangle B) = 0$ and $\widehat{\mu}, \widehat{d}$ are the canonical measure and metric induced on $\widehat{\mathcal{B}}$. That is, for each $A \in \mathcal{B}$ and $[A]_{\mu} := \{B \in \mathcal{B} \mid \mu(A \triangle B) = 0\} \in \widehat{\mathcal{B}}$, we have $\widehat{\mu}([A]_{\mu}) = \mu(A)$; similarly $\widehat{d}([A]_{\mu}, [B]_{\mu}) = d(A, B) = \widehat{\mu}([A \triangle B]_{\mu}) = \widehat{\mu}([A]_{\mu} \triangle [B]_{\mu})$.

One sees that $\widehat{\mathcal{B}}$ is a complete (in the sense of order) boolean algebra, $\widehat{\mu}$ is a strictly positive σ -additive measure on $\widehat{\mathcal{B}}$, and \widehat{d} is the metric defined canonically from $\widehat{\mu}$. The metric structures $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$ are the principal objects of study in this paper.

We use standard background from measure theory and analysis (which we summarize in Section 2) and from continuous first order logic. The model theoretic background for this paper comes from [7] and [8], which present the [0, 1]-valued continuous version of first order logic. In Section 3 we give references for some additional concepts and tools from continuous logic that we need here.

The main content of this paper is in Sections 4, 5, 6, and 8. Sections 4 and 5 present the model theory of arbitrary probability spaces in the framework of continuous first order logic. Our results show that everything model theoretic about arbitrary probability spaces can be systematically reduced to the atomless case, which is given a full treatment in Sections 6 and 8.

Atomless probability spaces were studied by Ben Yaacov [1] using the framework of compact abstract theories, with an emphasis on issues around model theoretic stability. In Sections 6 and 8 we study atomless probability spaces in the context of continuous logic and present analogues of results from [1], as well as additional results that are specific to the continuous first order setting. In Section 6 we give axioms for the class of atomless probability algebras, and show that the theory of these structures (denoted by APA) is well behaved from the model theoretic point of view: in particular, it is complete, has quantifier elimination, is separably categorical, and is ω -stable. We characterize (up to equivalence) the induced metric on type spaces of APA. In Section 8 we focus on features of APA that are connected to its stability. Following the work of Ben Yaacov [1], we give an intrinsic characterization of the independence relation of APA, and show that it has built-in canonical bases. We give a direct, elementary proof that APA is strongly finitely based, a fact originally proved in [6] using levely pairs of APA models. We also look at APA from the point of view of Shelah's classification program, and show that APA is non-multidimensional but not unidimensional, using natural translations of those concepts into continuous model theory.

Section 7 is devoted to Maharam's structure theorem for probability algebras and its connections with model theory. In Section 9 we show how model theoretic forking in probability algebras is related to probabilistic entropy. In the last section we identify some open problems that seem worth investigating further.

2. Probability spaces

In this section we present basic information about probability spaces (X, \mathcal{B}, μ) and their measure algebra quotients.

We recall that $A \in \mathcal{B}$ is atomless if for every $B \in \mathcal{B}$ with $B \subseteq A$ and $\mu(B) > 0$, there are $B_1, B_2 \in \mathcal{B}$ such that $B = B_1 \cup B_2$, B_1 and B_2 are disjoint and $\mu(B_1) > 0$, $\mu(B_2) > 0$. We say that (X, \mathcal{B}, μ) is atomless if X is atomless in \mathcal{B} .

We recall that $A \in B$ is an atom if $\mu(A) > 0$, and for every $B \in \mathcal{B}$ with $B \subseteq A$ one has $\mu(B) = 0$ or $\mu(A \setminus B) = 0$. Evidently, if $A_1, A_2 \in \mathcal{B}$ are atoms then either $\mu(A_1 \cap A_2) = 0$ or $\mu(A_1 \triangle A_2) = 0$. Furthermore, there exists a finite or countable family $\mathcal{A} \subseteq \mathcal{B}$ such that each $A \in \mathcal{A}$ is an atom and such that whenever $A \in \mathcal{B}$ is an atom, there exists $A' \in \mathcal{A}$ such that $\mu(A \triangle A') = 0$. The $atomic\ part$ of X is the join (union) of the sets in \mathcal{A} , and its complement is the $atomless\ part$ of X; this partition of X is well defined up to a set of measure X. The atomic part is X in the sense that whenever X is contained in the atomic part of X, then X is (up to a set of measure X) the union of the atoms it contains; equivalently, all atomless subsets of such an X have measure X. Likewise, the atomless part of X is an atomless member of X. We regard X0 as atomic, since it is contained in the atomic part, and X1 is atomless by definition. Further, if X2 is atomless and X3 is atomic, then X4 is atomic, then X5 is atomic, then X6 is atomic, then X6 is atomic, then X8 is atomic, then X9 is atomic ato

We say (A_1, \ldots, A_n) is a partition in \mathcal{B} if $A_i \cap A_j = 0$ whenever $i \neq j$. If, in addition, $A_1 \cup \cdots \cup A_n = B$, then we say (A_1, \ldots, A_n) is a partition of B in \mathcal{B} . Note that we allow A_i to be 0 in such a situation.

We say that $A_1, A_2 \in \mathcal{B}$ determine the same *event*, and write $A_1 \sim_{\mu} A_2$ if the symmetric difference of the sets has μ -measure zero. Clearly \sim_{μ} is an equivalence relation. We denote the equivalence class of $A \in \mathcal{B}$ by $[A]_{\mu}$. The collection of equivalence classes of \mathcal{B} modulo \sim_{μ} is denoted by $\widehat{\mathcal{B}}$. The operations of complement, union and intersection are well defined for events and they make $\widehat{\mathcal{B}}$ a boolean algebra. Moreover, μ induces on $\widehat{\mathcal{B}}$ a σ -additive, strictly positive probability measure $\widehat{\mu}$. As noted above, we denote the canonical metric on $\widehat{\mathcal{B}}$ by \widehat{d} and recall that it is defined by $\widehat{d}([A]_{\mu}, [B]_{\mu}) = d(A, B)$. It is important in this paper that $(\widehat{\mathcal{B}}, \widehat{d})$ is a complete metric space (see the calculation in [15, Lemma 323F]).

We refer to $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$ as the probability algebra of (X, \mathcal{B}, μ) .

2.1. **Notation.** (a) If C is a subset of a boolean algebra, we denote the boolean subalgebra generated by C by $C^{\#}$.

Let (X, \mathcal{B}, μ) be a probability space and let $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$ be its probability algebra.

- (b) If S is a subset of \mathcal{B} , we let $\langle S \rangle$ denote the σ -subalgebra of \mathcal{B} generated by S.
- (c) If S is a subset of $\widehat{\mathcal{B}}$, we let $\langle S \rangle$ denote the \widehat{d} -closure of $S^{\#}$. Note that $\langle S \rangle$ is equal to the σ -subalgebra of $\widehat{\mathcal{B}}$ generated by S, by [15, Lemma 323F], and it is also equal to the \widehat{d} -closed boolean subalgebra generated by S. In other words, $\langle S \rangle$ is \widehat{d} -closed and has $S^{\#}$ as a \widehat{d} -dense subset.
- (d) Throughout this paper we use upper case letters such as A, B for elements of the σ -algebra \mathcal{B} and lower case letters such as a, b for elements of $\widehat{\mathcal{B}}$. If S is a subset of \mathcal{B} , we denote by \widehat{S} the set of events determined by the elements of S; i.e., $\widehat{S} = \{[A]_{\mu} \mid A \in S\} \subseteq \widehat{\mathcal{B}}$.

Whenever $\mathcal{C} \subseteq \mathcal{B}$ is a σ -subalgebra, $(X, \mathcal{C}, \mu | \mathcal{C})$ is a probability space in its own right, and we have defined $\widehat{\mathcal{C}}$ to be the probability algebra of \mathcal{C} , and also (in 2.1(d)) to be a certain subset of $\widehat{\mathcal{B}}$. There is no real ambiguity here; indeed, the inclusion map j of \mathcal{C} into \mathcal{B} induces a measure-preserving boolean isomorphism \widehat{j} (which thus also preserves the metric) between the two versions of $\widehat{\mathcal{C}}$. (The function \widehat{j} maps $[A]_{\mu | \mathcal{C}}$ in the sense of $(X, \mathcal{C}, \mu | \mathcal{C})$ to $[A]_{\mu}$ in the sense of (X, \mathcal{B}, μ) , for each $A \in \mathcal{C}$.) With this identification, $(\widehat{\mathcal{C}}, \widehat{\mu}, \widehat{d})$ is the probability algebra of the probability space $(X, \mathcal{C}, \mu | \mathcal{C})$, and it is (canonically isomorphic to) a substructure of $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$.

In the next result, we record for later use that the converse of the preceding comment is also true.

2.2. **Lemma.** Let (X, \mathcal{B}, μ) be a probability space and let $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$ be its probability algebra. Let S be a subset of $\widehat{\mathcal{B}}$ and consider $\langle S \rangle \subseteq \widehat{\mathcal{B}}$ as in 2.1(c). Let

$$\mathcal{S} := \{ A \in \mathcal{B} \mid [A]_{\mu} \in \langle S \rangle \}.$$

Then S is a σ -subalgebra of B and $\langle S \rangle = \widehat{S}$.

In particular, every substructure $(\langle S \rangle, \mu \upharpoonright \langle S \rangle, d \upharpoonright \langle S \rangle)$ of the probability algebra $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$ of a probability space (X, \mathcal{B}, μ) is (isomorphic to) the probability algebra of a probability space $(X, \mathcal{S}, \mu \upharpoonright \mathcal{S})$, with \mathcal{S} a σ -subalgebra of \mathcal{B} .

Proof. For $A, B \in \mathcal{S} \subseteq \mathcal{B}$ we have $[A \cup B]_{\mu} = [A]_{\mu} \cup [B]_{\mu} \in \langle S \rangle$, so \mathcal{S} is closed under the union operation of \mathcal{B} . Similar calculations show that \mathcal{S} is closed under \cap and \cdot^c . Also, note that \mathcal{S} contains every element of \mathcal{B} that has μ -measure 0.

To finish the proof, we need to consider an increasing sequence (A_n) in \mathcal{S} and show that the union of (A_n) in \mathcal{B} is an element of \mathcal{S} . Given such an (A_n) , the sequence $([A_n]_{\mu}) \subseteq \langle S \rangle$ must be increasing in $\widehat{\mathcal{B}}$, so it converges in the sense of the metric \widehat{d} to an element $[B]_{\mu} \in \langle S \rangle$, with $B \in \mathcal{S}$. This means that $\mu(A_n \triangle B) \to 0$ in \mathcal{B} . Since (A_n) is increasing, this implies $\mu(A_n \setminus B) = 0$ for all n. It also implies $\mu(B \setminus (\cup A_n)) = 0$, and therefore B differs from $\cup A_n$ by a set of μ -measure 0 in \mathcal{B} . Hence $\cup A_n \in \mathcal{S}$.

We need the following familiar special case of the Radon-Nikodym theorem:

2.3. **Theorem.** [12, Theorem 3.8] Let (X, \mathcal{B}, μ) be a probability space, let $\mathcal{C} \subseteq \mathcal{B}$ be a σ subalgebra, and consider $A \in \mathcal{B}$. Then there exists $g \in L_1(X, \mathcal{C}, \mu)$ such that for every $B \in \mathcal{C}$, one has $\int_B g d\mu = \int_B \chi_A d\mu$. The function g is determined by $a = [A]_\mu$ up to equality μ -almost everywhere; it is called the conditional probability of a with respect to \mathcal{C} and we
denote it by $\mathbb{P}(a|\mathcal{C})$ or, equivalently, by $\mathbb{P}(A|\mathcal{C})$. We also refer to g as representing $\mathbb{P}(a|\mathcal{C})$ and $\mathbb{P}(A|\mathcal{C})$.

More generally, for $f \in L_1(X, \mathcal{B}, \mu)$ there exists $\mathbb{E}(f|\mathcal{C}) \in L_1(X, \mathcal{C}, \mu)$ such that for every $B \in \mathcal{C}$, one has $\int_B \mathbb{E}(f|\mathcal{C})d\mu = \int_B f d\mu$. The function $\mathbb{E}(f|\mathcal{C})$ is unique in the sense that the operation $f \mapsto \mathbb{E}(f|\mathcal{C})$ preserves the equivalence relation of equality μ -almost everywhere. The element $\mathbb{E}(f|\mathcal{C})$ is called the conditional expectation of f with respect to \mathcal{C} and we also denote it by $\mathbb{E}_{\mathcal{C}}(f)$.

Note that for any $A \in \mathcal{B}$, the function $\mathbb{P}(A|\mathcal{C})$ must have its values in [0,1] μ -ae.

2.4. **Notation.** Let (X, \mathcal{B}, μ) be a probability space, and consider $a \in \widehat{\mathcal{B}}$. Suppose D is a boolean subalgebra of $\widehat{\mathcal{B}}$ that is closed (with respect to the metric \widehat{d}), and let $\mathcal{D} = \{A \in \mathcal{B} \mid [A]_{\mu} \in D\}$ be the σ -subalgebra discussed in the proof of Lemma 2.2, so $\widehat{\mathcal{D}} = D$. We write $\mathbb{P}(a|D)$ to denote $\mathbb{P}(\chi_A|\mathcal{D})$, where $a = [A]_{\mu}$. Similarly, for $a \in \widehat{\mathcal{B}}$ we write χ_a to denote one of the characteristic functions χ_A where $A \in \mathcal{B}$ and $a = [A]_{\mu}$. Note that if A_1, A_2 are two such sets for a, then $\chi_{A_1} = \chi_{A_2}$ holds μ -almost everywhere, and hence the same is true of $\mathbb{P}(\chi_{A_1}|\mathcal{D})$ and $\mathbb{P}(\chi_{A_2}|\mathcal{D})$.

When we are given a probability algebra (\mathcal{B}, μ, d) without specifying the underlying probability space, and \mathcal{A} is a closed subalgebra of \mathcal{B} , we refer to $\mathbb{P}(b|\mathcal{A})$ as being an \mathcal{A} -measurable function in order to avoid explicitly introducing the probability space representing (\mathcal{B}, μ, d) and the σ -subalgebra representing \mathcal{A} (as described in Lemma 2.2). Similarly we refer to χ_a as being \mathcal{A} -measurable, when $a \in \mathcal{A}$.

As is customary, when $a, b \in \widehat{\mathcal{B}}$, we write $\mathbb{P}(a|D) = \mathbb{P}(b|D)$ to mean that the functions $\mathbb{P}(a|D)$ and $\mathbb{P}(b|D)$ are equal μ -almost everywhere, and we give a similar interpretation to $\mathbb{P}(a|D) \leq \mathbb{P}(b|D)$. Therefore, the associated strict partial ordering $\mathbb{P}(a|D) < \mathbb{P}(b|D)$ (meaning that $\mathbb{P}(a|D) \leq \mathbb{P}(b|D)$ is true while $\mathbb{P}(a|D) = \mathbb{P}(b|D)$ is false) is true if and only if $\mathbb{P}(a|D) \leq \mathbb{P}(b|D)$ holds μ -almost everywhere and $\mathbb{P}(a|D) < \mathbb{P}(b|D)$ holds on a set of positive μ -measure. Similar remarks apply to these relations between other measurable real-valued functions (such as χ_a).

When $\mathcal{E} = \{\emptyset, X\}$ is the trivial subalgebra, and $A \in \mathcal{B}$, then $\mathbb{P}(A|\mathcal{E})$ is the constant function $f(x) := \mu(A)$ for all $x \in X$. Indeed, this f is \mathcal{E} -measurable and $\int_{\mathcal{E}} f d\mu = \mu(A \cap \mathcal{E}) = \mu(A)$

 $\int_E \chi_A d\mu$ for all $E \in \mathcal{E}$, namely for $E = \emptyset$ and E = X. More generally, for finite subalgebras $\mathcal{E} \subseteq \mathcal{B}$ we have the following formula for $\mathbb{P}(A|\mathcal{E})$, which is useful in many places below.

2.5. **Lemma.** Let $\mathcal{E} \subseteq \mathcal{B}$ be a finite subalgebra and $A \in \mathcal{B}$. Suppose E_1, \ldots, E_n are the atoms in \mathcal{E} . Then

$$\mathbb{P}(A|\mathcal{E}) = \sum_{j} \frac{\mu(A \cap E_j)}{\mu(E_j)} \chi_{E_j}.$$

Proof. By additivity of the integral, it suffices to prove that the integral of the displayed function over each atom E_j is equal to $\int_{E_j} \chi_A d\mu$, which equals $\mu(A \cap E_j)$. Since the sets E_j are pairwise disjoint, this is clear.

2.6. Fact. Let (X, \mathcal{B}, μ) be a probability space and let $\mathcal{C} \subseteq \mathcal{B}$ be a σ -subalgebra. The conditional expectation operator $\mathbb{E}_{\mathcal{C}}$ restricted to $L_2(X, \mathcal{B}, \mu)$ is the Hilbert space orthogonal projection of $L_2(X, \mathcal{B}, \mu)$ onto the subspace $L_2(X, \mathcal{C}, \mu)$. (See [10, Proposition 4.2].)

Let $\mathcal{D} \subseteq \mathcal{C} \subseteq \mathcal{B}$ be σ -subalgebras and $A \in \mathcal{B}$. Then $\mathbb{P}(A|\mathcal{D}) = \mathbb{E}_{\mathcal{D}}(\chi_A) = \mathbb{E}_{\mathcal{D}}(\mathbb{E}_{\mathcal{C}}(\chi_A))$ is the orthogonal projection of $\mathbb{P}(A|\mathcal{C}) = \mathbb{E}_{\mathcal{C}}(\chi_A)$ into $L_2(X, \mathcal{D}, \mu)$, so

$$\|\mathbb{P}(A|\mathcal{C})\|_{2}^{2} - \|\mathbb{P}(A|\mathcal{D})\|_{2}^{2} = \|\mathbb{P}(A|\mathcal{C}) - \mathbb{P}(A|\mathcal{D})\|_{2}^{2}$$

Further, since we are working over a probability space, we have $||f||_2 = ||f||_2 ||1||_2 \ge |\langle f, 1 \rangle| = ||f||_1$ for all L_2 functions f, by the Cauchy-Schwartz inequality, and therefore

$$\|\mathbb{P}(A|\mathcal{C}) - \mathbb{P}(A|\mathcal{D})\|_2 \ge \|\mathbb{P}(A|\mathcal{C}) - \mathbb{P}(A|\mathcal{D})\|_1.$$

These give useful quantitative conditions for $\mathbb{P}(A|\mathcal{C}) \neq \mathbb{P}(A|\mathcal{D})$. They are used in proving Remark 8.2, Fact 9.4(5) and Corollary 9.5.

If $\mathcal{E} \subseteq \mathcal{D} \subseteq \mathcal{C} \subseteq \mathcal{B}$ and $A \in \mathcal{B}$, the preceding discussion yields

$$\|\mathbb{P}(A|\mathcal{C}) - \mathbb{P}(A|\mathcal{D})\|_1 \le \|\mathbb{P}(A|\mathcal{C}) - \mathbb{P}(A|\mathcal{D})\|_2 \le \|\mathbb{P}(A|\mathcal{C}) - \mathbb{P}(A|\mathcal{E})\|_2,$$

which can be useful in working with approximations to $\mathbb{P}(A|\mathcal{C})$, as we illustrate next.

2.7. **Lemma.** Let (X, \mathcal{B}, μ) be a probability space, $\mathcal{C} \subseteq \mathcal{B}$ a σ -subalgebra, and $A \in \mathcal{B}$. For each $k \geq 1$ there exists $(E_1, \ldots, E_k) \in \mathcal{C}^k$, a partition of X, such that for any σ -subalgebra \mathcal{D} with $\{E_1, \ldots, E_k\} \subseteq \mathcal{D} \subseteq \mathcal{C}$ one has $\|\mathbb{P}(A|\mathcal{C}) - \mathbb{P}(A|\mathcal{D})\|_1 \leq 1/k$.

Proof. Let $f = \mathbb{P}(\chi_A | \mathcal{C})$, so f is a \mathcal{C} -measurable [0, 1]-valued function. Let I_1, \ldots, I_k be the intervals $I_j = [\frac{j-1}{k}, \frac{j}{k})$ for $j = 1, \ldots, k-1$ and $I_k = [\frac{k-1}{k}, 1]$. So the intervals are pairwise disjoint and their union is [0, 1]. For each j let $E_j = \{x \in X \mid f(x) \in I_j\}$. Then (E_1, \ldots, E_k) is a partition of X in \mathcal{C} . For any sequence (r_1, \ldots, r_k) such that $r_j \in I_j$ for all j, we have $|f - \sum_j r_j \chi_{E_j}| \le 1/k$ pointwise on X, and therefore $||f - \sum_j r_j \chi_{E_j}||_2 \le 1/k$.

Now set $r_j = \frac{\mu(A \cap E_j)}{\mu(E_j)}$. Since f(x) is in I_j for $x \in E_j$, we have $\frac{j-1}{k}\mu(E_j) \le \int_{E_j} f d\mu \le \frac{j}{k}\mu(E_j)$ for all j. Noting that $\int_{E_j} f d\mu = \int_{E_j} \chi_A d\mu = \mu(A \cap E_j)$ we see that $r_j \in I_j$ for all $j = 1, \ldots, k$. It follows using Lemma 2.5 that

$$||f - \mathbb{P}(\chi_A | \{E_1, \dots, E_k\}^{\#})||_2 = ||f - \sum_j \frac{\mu(A \cap E_j)}{\mu(E_j)} \chi_{E_j}||_2 \le 1/k.$$

Further, if \mathcal{D} is any σ -subalgebra of \mathcal{B} that contains $\{E_1, \ldots, E_k\}$, then $||f - \mathbb{P}(\chi_A|\mathcal{D})||_1 \le 1/k$, by the last statement in Fact 2.6, so (E_1, \ldots, E_k) satisfies the stated conditions. \square

Probabilistic independence is very important in this paper. For $A, B \in \mathcal{B}$, we say A and B are (probabilistically) independent if $\mu(A \cap B) = \mu(A)\mu(B)$, and write $A \perp\!\!\!\perp B$. Further, if S, T are subsets of \mathcal{B} , we say S and T are (probabilistically) independent and write $S \perp\!\!\!\!\perp T$ if $A \perp\!\!\!\!\perp B$ holds for every $A \in \langle S \rangle$ and $B \in \langle T \rangle$. Not surprisingly to model theorists, we need a more general version of independence that is relative to a set of parameters:

2.8. **Definition.** If $\mathcal{E} \subseteq \mathcal{B}$ is a σ -subalgebra of \mathcal{B} , we say A and B are (conditionally) independent over \mathcal{E} , and write $A \perp \!\!\! \perp_{\mathcal{E}} B$ if

$$\mathbb{P}(A \cap B|\mathcal{E}) = \mathbb{P}(A|\mathcal{E}) \cdot \mathbb{P}(B|\mathcal{E}).$$

More generally, if S, T, W are subsets of \mathcal{B} , we say S and T are (conditionally) independent over W and write $S \perp \!\!\! \perp_W T$ if $A \perp \!\!\! \perp_{\langle W \rangle} B$ holds for every $A \in \langle S \rangle$ and $B \in \langle T \rangle$.

Note that $\perp\!\!\!\perp_{\langle W \rangle}$ reduces to $\perp\!\!\!\perp$ when $W = \emptyset$, since $\langle \emptyset \rangle = \{\emptyset, X\}$. Also, $S \perp\!\!\!\perp_W T$ if and only if $A \perp\!\!\!\perp_{\langle W \rangle} B$ holds for every $A \in S^{\#}$ and $B \in T^{\#}$.

For us the following characterization of conditional independence is fundamental. As usual in model theory, if Y, Z are sets of parameters, we denote $Y \cup Z$ by YZ.

- 2.9. **Lemma.** If S, T, W are subsets of \mathcal{B} , then the following statements are equivalent:
 - (i) $S \perp \!\!\! \perp_W T$.
 - (ii) $\mathbb{P}(A|\langle WT \rangle) = \mathbb{P}(A|\langle W \rangle)$ for all $A \in S^{\#}$.
 - (iii) $\mathbb{P}(A|\langle WT \rangle)$ is $\langle W \rangle$ -measurable, for all $A \in S^{\#}$.
 - (iv) $\|\mathbb{P}(A|\langle WT\rangle)\|_2 = \|\mathbb{P}(A|\langle W\rangle)\|_2$ for all $A \in S^{\#}$.

Proof. (i) \Leftrightarrow (ii): Apply [18, Theorem 8.9], noting that (ii) is equivalent to the same statement with $S^{\#}$ replaced by $\langle S \rangle$, since $S^{\#}$ is dense in $\langle S \rangle$.

- (ii) \Leftrightarrow (iii): This is immediate.
- (iv) \Leftrightarrow (ii): Let $A \in \mathcal{B}$. Since $\langle W \rangle \subseteq \langle WT \rangle$, Fact 2.6 gives us

$$\|\mathbb{P}(A|\langle WT\rangle)\|_2^2 - \|\mathbb{P}(A|\langle W\rangle)\|_2^2 = \|\mathbb{P}(A|\langle WT\rangle) - \mathbb{P}(A|\langle W\rangle)\|_2^2$$

from which follows

$$\|\mathbb{P}(A|\langle WT\rangle)\|_2 = \|\mathbb{P}(A|\langle W\rangle)\|_2$$
 if and only if $\mathbb{P}(A|\langle WT\rangle) = \mathbb{P}(A|\langle W\rangle)$.

Applying the quantifier "for all $A \in S^{\#}$ " yields the desired equivalence.

The next result shows that several different definitions of $\perp \!\!\! \perp$ that one finds in the literature are equivalent.

- 2.10. Corollary. If S, T, W are subsets of \mathcal{B} , then the following statements are equivalent:
 - (i) $S \perp \!\!\! \perp_W T$.
 - (ii) $S \perp \!\!\!\perp_W \langle WT \rangle$
 - (iii) $\langle WS \rangle \perp \!\!\! \perp_W \langle WT \rangle$
- *Proof.* (ii) \Rightarrow (i) is clear. Assume now that $S \perp \!\!\! \perp_W T$ holds and prove (ii). By Lemma 2.9(ii) we have that $\mathbb{P}(A|\langle WT\rangle) = \mathbb{P}(A|\langle W\rangle)$ for all $A \in S^{\#}$ and thus $S \perp \!\!\! \perp_W \langle WT\rangle$ holds, again using Lemma 2.9 part (ii). The proof of (ii) \Leftrightarrow (iii) is similar, since the definition of independence is a symmetric condition on the left and right families.
- 2.11. **Remark.** When $W \subseteq \mathcal{B}$ is finite, we have the following simple characterization of $\underline{\perp}$. Namely, $S \underline{\perp}_W T$ if and only if $\mu(A \cap B \cap C)\mu(C) = \mu(A \cap C)\mu(B \cap C)$ for every $A \in S^{\#}$, and $B \in T^{\#}$ and every atom $C \in W^{\#}$.

This is easily proved by comparing coefficients in the expressions for $\mathbb{P}(A \cap B|W^{\#})$ and $\mathbb{P}(A|W^{\#}) \cdot \mathbb{P}(B|W^{\#})$ given by Lemma 2.5.

2.12. **Notation.** Suppose C, D, E are subsets of $\widehat{\mathcal{B}}$, and S, T, W are subsets of \mathcal{B} such that $\langle C \rangle = \langle \widehat{S} \rangle$, $\langle D \rangle = \langle \widehat{T} \rangle$, and $\langle E \rangle = \langle \widehat{W} \rangle$. We write $C \perp \!\!\!\!\perp_E D$ to mean the same as $S \perp \!\!\!\!\perp_W T$.

Next we prove a Lemma that will be used in the proof of Theorem 8.1.

2.13. Lemma (Extension). Let (X, \mathcal{B}, μ) be a probability space. Let \mathcal{A} be a finite subalgebra of \mathcal{B} , with atoms A_1, \ldots, A_m , and let $\mathcal{C} \subseteq \mathcal{D}$ be closed subalgebras of \mathcal{B} . Then there exists a probability space (X', \mathcal{B}', μ') and a boolean, measure-preserving embedding $B \mapsto B'$ of \mathcal{B} into \mathcal{B}' , together with a finite subalgebra \mathcal{E} of \mathcal{B}' whose atoms E_1, \ldots, E_m satisfy $\mathbb{P}(E_j|\mathcal{C}') = \mathbb{P}(A'_j|\mathcal{C}')$ for all $i = 1, \ldots, m$ and $\mathcal{E} \perp_{\mathcal{C}'} \mathcal{D}'$. (Here for $\mathcal{Z} = \mathcal{C}$ or \mathcal{D} we write \mathcal{Z}' for $\{B' \mid B \in \mathcal{Z}\}$.)

Proof. Let $([0,1], \mathcal{F}, \lambda)$ be the Lebesgue measure space on [0,1] and take $(X \times [0,1], \mathcal{B}', \mu')$ to be the product measure space, with $\mathcal{B}' = \mathcal{B} \otimes \mathcal{F}$ and $\mu' = \mu \otimes \lambda$.

For each $B \in \mathcal{B}$, let $B' := B \times [0,1]$. The correspondence $B \mapsto B'$ is obviously a boolean, measure-preserving embedding of \mathcal{B} into \mathcal{B}' . Also, for any \mathcal{B} -measurable function $f: X \to [0,1]$, we let f' denote the \mathcal{B}' -measurable function $(x,y) \mapsto f(x)$. We see easily that

the embedding preserves integration; namely for any $B \in \mathcal{B}$ and \mathcal{B} -measurable $f: X \to [0, 1]$ we have by Fubini's Theorem $\int_{B'} f' d(\mu \otimes \lambda) = \int_{B} \left(\int_{[0,1]} f' d\lambda \right) d\mu = \int_{B} f d\mu$. We also note that $\mathbb{P}(B'|\mathcal{C}') = \mathbb{P}(B|\mathcal{C})'$ for any $B \in \mathcal{B}$ and σ -subalgebra \mathcal{C} of \mathcal{B} . (Using Lemma 2.7 it suffices to prove this when \mathcal{C} is finite; this is done by applying Lemma 2.5. Note that if C_1, \ldots, C_k are the atoms of \mathcal{C} , then C'_1, \ldots, C'_k are the atoms of \mathcal{C}' , and for each $j = 1, \ldots, k$ we have $(\mu \otimes \lambda)(B' \cap C'_j) = \mu(B \cap C_j), (\mu \otimes \lambda)(C'_j) = \mu(C_j)$, and $\chi_{C'_j} = (\chi_{C_j})'$.)

For each $i=1,\ldots,m$, let $f_i:=\mathbb{P}(A_i|\mathcal{C})$ and note that since A_1,\ldots,A_m is a partition of 1, we have $f_1(x)+\cdots+f_m(x)=1$ on X μ -a.e.. Now let $E_1=\{(x,y)\in X\times[0,1]:0\leq y\leq f_1(x)\}$ and for $1< i\leq m$ let $E_i=\{(x,y)\in X\times[0,1]:f_1(x)+\cdots+f_{i-1}(x)< y\leq f_1(x)+\cdots+f_i(x))\}$. The sets $\{E_i\}_{i\leq m}$ are \mathcal{B}' -measurable and pairwise disjoint, and $X'=X\times[0,1]=\bigcup_{i\leq m}E_i$ (except possibly for a set of μ' -measure zero).

Note that for any $B \in \mathcal{B}$ and i = 1, ..., m, we have

$$(\mu \otimes \lambda)(B' \cap E_i) = \int_{B'} \chi_{E_i} d(\mu \otimes \lambda) = \int_{B} \left(\int_{[0,1]} \chi_{E_i} d\lambda \right) d\mu = \int_{B} f_i d\mu$$

using the definitions and Fubini's Theorem. If $C \in \mathcal{C}$, this gives

$$\int_{C'} \chi_{E_i} d(\mu \otimes \lambda) = \int_{C} f_i d\mu = \int_{C'} \mathbb{P}(A_i | \mathcal{C})' d(\mu \otimes \lambda) = \int_{C'} \mathbb{P}(A_i' | \mathcal{C}') d(\mu \otimes \lambda)$$

and therefore $\mathbb{P}(E_i|\mathcal{C}') = \mathbb{P}(A_i'|\mathcal{C}')$ for all i = 1, ..., m. For $D \in \mathcal{D}$ we get

$$(\mu \otimes \lambda)(D' \cap E_i) = \int_{D'} f'_i d(\mu \otimes \lambda) = \int_{D'} \mathbb{P}(A_i | \mathcal{C})' d(\mu \otimes \lambda) = \int_{D'} \mathbb{P}(A'_i | \mathcal{C}') d(\mu \otimes \lambda),$$

so $\mathbb{P}(E_i | \mathcal{D}') = \mathbb{P}(E_i | \mathcal{C}')$ for all $i = 1, \dots, m$. Therefore $\mathcal{E} \perp \!\!\!\perp_{\mathcal{C}'} \mathcal{D}'$.

3. Some continuous model theory

In this paper we use the setting of continuous logic to discuss the model theory of probability algebras. The fundamental ideas of continuous logic are presented in [7, 8]. We assume familiarity with the material in these sources, and often use it without specific reference.

In addition, we need some background concerning *metric imaginaries* in continuous logic, and their role in some topics within stability theory, especially when dealing with canonical parameters for definable predicates and definability of types. Here we give pointers to published sources for this background, and a very brief summary of the topics we use.

There is some treatment of imaginary sorts (i.e., of *interpretations*) in our key references [7, 8]. In [7, Section 11] only finitary imaginaries are presented; these are quotients of finite products of sorts modulo a definable pseudometric.

However, in our Section 5 and later in the paper, in connection with certain concepts in stability theory, we need more general, *infinitary* imaginaries. These are quotients of the product of a countably infinite family of sorts modulo a definable pseudometric; they are

connected to the existing structure by their projection maps onto the sorts from which they come. A central example of these imaginaries is given by *canonical parameters* for a definable predicate relative to the (possibly infinite) sequence of parameters used in defining it. These are treated in detail in [8, Section 5].

Given a continuous theory T, the many sorted theory obtained by adding to T all possible metric imaginary sorts is called the *meq expansion of* T, and it is denoted T^{meq} . Likewise, given $\mathcal{M} \models T$, the corresponding expansion of \mathcal{M} to a model of T^{meq} is denoted \mathcal{M}^{meq} . Presentations of the full construction of T^{meq} and some of its properties are in [13, Section 3.3], and in [5, Section 1].

In Section 8 we also use concepts and tools from stability theory in the setting of continuous model theory. Many of these, including canonical parameters for formulas, definability of types, and canonical bases for stationary types, are developed in [8, Sections 7 and 8]. Beyond these, we use concepts such as types being parallel, the parallelism class of a stationary type, Morley sequences, orthogonal types, and non-multidimensional theories. While these concepts lack a thorough exposition in the continuous model theory literature, it is not difficult to formulate and understand them based on how they are treated in the main references for stability theory in classical model theory, especially given the tools provided in [8]. An example needed here of such a fact is that the canonical base of a stationary type is contained in the meq definable closure of a Morley sequence of that type. For this material in the classical discrete setting, we follow closely the presentation in [11].

4. The model theory of probability spaces

We deal here with structures of the form

$$\mathcal{M} = (\widehat{\mathcal{B}}, 0, 1, \cdot^c, \cap, \cup, \widehat{\mu}, \widehat{d})$$

where $(\widehat{\mathcal{B}}, \widehat{\mu})$ is the probability algebra of a probability space (X, \mathcal{B}, μ) , 0 is the event corresponding to \emptyset and 1 is the event corresponding to X; \cdot^c is the complement operation and \cap, \cup are the intersection and union operations on $\widehat{\mathcal{B}}$; and \widehat{d} is the canonical metric on $\widehat{\mathcal{B}}$ (defined for $a, b \in \widehat{\mathcal{B}}$ by $\widehat{d}(a, b) = \widehat{\mu}(a \triangle b)$). The predicates, namely $\widehat{\mu}$ and \widehat{d} , take their values in the interval [0, 1]. The modulus of uniform continuity for the unary operation \cdot^c and the unary predicate $\widehat{\mu}$ is given by $\Delta(\epsilon) = \epsilon$; for the binary operations \cap and \cup the modulus is given by $\Delta(\epsilon) = \epsilon/2$.

For the rest of this paper we take L^{pr} to be the continuous signature indicated in the previous paragraph.

Note that every probability algebra of a probability space is indeed an L^{pr} -structure. (This requires, in particular, that it is complete as a metric space, which we noted in Section 2.)

4.1. **Notation.** For any L^{pr} -prestructure \mathcal{M} and $a \in M$, we write a^{-1} for a^{c} and a^{+1} for a.

The following L^{pr} -conditions are easily seen to be true in every probability algebra of a probability space.

- (1) Boolean algebra axioms:
 - Each of the usual axioms for a boolean algebra is the \forall -closure of an equation between terms (see [17, p.38]) and thus it can be expressed in continuous logic as a condition. For example, the axiom $\forall x \forall y (x \cup y = y \cup x)$ is equivalent to $\sup_x \sup_y \left(d(x \cup y, y \cup x) \right) = 0$.
- (2) Measure axioms:

```
\mu(0) = 0 \text{ and } \mu(1) = 1
\sup_{x} \sup_{y} \left( \mu(x \cap y) \doteq \mu(x) \right) = 0
\sup_{x} \sup_{y} \left( \mu(x) \doteq \mu(x \cup y) \right) = 0
\sup_{x} \sup_{y} \left| \left( \mu(x) \doteq \mu(x \cap y) - \left( \mu(x \cup y) \doteq \mu(y) \right) \right| = 0
The last three axioms express that \mu(x \cup y) + \mu(x \cap y) = \mu(y) + \mu(x) for all x, y.
```

(3) Connections between d and μ : $\sup_{x} \sup_{y} |d(x,y) - \mu(x\Delta y)| = 0 \text{ where } x\Delta y \text{ denotes the boolean term giving the symmetric difference: } x\Delta y = (x \cap y^c) \cup (x^c \cap y).$

We denote the set of L^{pr} -conditions above by Pr.

- 4.2. **Notation.** For the rest of this paper we use the notation $\mathcal{M} = (\mathcal{B}, \mu, d)$ for a general model of Pr. We take care that this notation is not confused with our usual notation (X, \mathcal{B}, μ) for a probability space and $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$ for its associated probability algebra.
- 4.3. **Theorem.** The models $\mathcal{M} = (\mathcal{B}, \mu, d)$ of Pr are exactly the (abstract) probability algebras. That is, \mathcal{B} is a σ -order complete Boolean algebra, and μ is a strictly positive, σ -additive probability measure on \mathcal{B} ; further, d is defined on \mathcal{B} by $d(a, b) := \mu(a \triangle b)$.

Proof. If $\mathcal{M} = (\mathcal{B}, \mu, d)$ is indeed a probability algebra as described in the statement, then it is clear that it satisfies all conditions in Pr. Moreover, the metric space (\mathcal{B}, d) is complete, as shown by the calculation in [15, Lemma 323F].

Conversely, suppose \mathcal{M} is a model of Pr. It is clear from the axioms that \mathcal{M} consists of a boolean algebra \mathcal{B} with a finitely additive probability measure μ such that \mathcal{B} is a complete metric space under the metric $d(a,b) = \mu(a \triangle b)$. Moreover, μ must be continuous on \mathcal{B} with respect to d; indeed, μ is 1-Lipschitz with respect to d, as is dictated by the signature L^{pr} .

Any increasing sequence in \mathcal{B} is necessarily a Cauchy sequence with respect to d, so it converges. This and the continuity of μ ensure that \mathcal{B} is σ -order complete as a boolean algebra and μ is σ -additive on \mathcal{B} .

It follows from Theorem 4.3 that the models of Pr are (up to isomorphism) exactly the probability algebras of probability spaces. This is proved in [15, Theorem 321J]; a key ingredient in the proof of that result is the Loomis-Sikorski representation theorem for σ -order complete boolean algebras; see [15, Theorem 314M]. In Theorem 4.5 we give a proof of this fact about the models of Pr using tools from model theory.

4.4. **Example.** Suppose \mathcal{A} is a boolean algebra and μ is a finitely additive probability measure on \mathcal{A} . We may define a distance d on \mathcal{A} in the familiar way, by setting d(a,b) equal to $\mu(a\triangle b)$, where \triangle denotes the symmetric difference in \mathcal{A} . Then (\mathcal{A}, μ, d) is an L^{pr} -prestructure, and it satisfies all of the axioms of Pr. Therefore we may obtain a model $(\widehat{\mathcal{A}}, \widehat{\mu}, \widehat{d})$ of Pr by first taking the quotient of (\mathcal{A}, μ, d) by the ideal of elements of μ -measure 0, and then taking the metric completion of the resulting quotient (as discussed in the middle of [7, pages 329-331]).

For those readers who are familiar with Abraham Robinson's nonstandard analysis (NSA), we note how this construction relates to the Loeb measure construction [20], which has been one of the most important tools for applications of NSA. For that construction, we begin with \mathcal{A} being an internal boolean algebra of subsets of an internal set X, and μ being obtained from an internal finitely additive *[0,1]-valued measure ν on \mathcal{A} , by taking $\mu(a)$ to be the standard part of $\nu(a)$ for each $a \in \mathcal{A}$. Let $(\widehat{\mathcal{A}}, \widehat{\mu}, \widehat{d})$ be constructed as above from (\mathcal{A}, μ, d) as in the preceding paragraph. In that setting, the quotient algebra of \mathcal{A} by the ideal of μ -null sets is already complete with respect to the quotient metric obtained from d, owing to the assumption of ω_1 -saturation that is part of the basic NSA framework. Moreover, the saturation assumption also implies that μ has a natural and unique extension to a σ -additive probability measure on the σ -algebra of subsets of X that is generated by \mathcal{A} . The resulting probability space has $(\widehat{\mathcal{A}}, \widehat{\mu}, \widehat{d})$ as its probability algebra. See [19, Section II.2] and [22, Section 2.1] for elementary discussions of the Loeb construction and its basic properties.

The metric ultraproduct of a family of probability algebras of probability spaces is an example of the Loeb construction. In that case, the internal measure space is the discrete ultraproduct of the family of probability spaces.

This approach gives an alternative way of proving that every model of Pr is the probability algebra of some probability space, as we show next.

- 4.5. **Theorem.** Let \mathcal{M} be a L^{pr} -structure. The following are equivalent:
- (1) \mathcal{M} is a model of Pr.
- (2) \mathcal{M} is isomorphic to the probability algebra of a probability space.
- *Proof.* (2) \Rightarrow (1): See the first paragraph of the proof of Theorem 4.3.
- $(1) \Rightarrow (2)$: Let \mathcal{M} be a model of Pr. Let I be the set of all finite subsets of M. For each $\tau \in I$, let \mathcal{M}_{τ} be the subalgebra $\tau^{\#}$ of \mathcal{M} , which is finite. Each \mathcal{M}_{τ} is the probability

algebra of a finite probability space $(X_{\tau}, \mathcal{A}_{\tau}, \mu_{\tau})$. Here X_{τ} is the set of atoms in \mathcal{M}_{τ} , \mathcal{A}_{τ} is the boolean algebra of all subsets of X_{τ} , and $\mu_{\tau}(\{a\}) = \mu(a)$ for each element a of X_{τ} .

There exists an ultrafilter U on I such that for each $a \in M$ the set $\{\tau \in I \mid a \in \tau\}$ is an element of U. As discussed in the preceding example, the U-ultraproduct of the family $(\mathcal{M}_{\tau} \mid \tau \in I)$ is the probability algebra of a probability space, by the Loeb measure construction. Moreover, \mathcal{M} is isomorphic to a substructure of this ultraproduct; the embedding maps $a \in M$ to the equivalence class of the family $(a_{\tau} \mid \tau \in I)$ where we define a_{τ} as follows: (i) if $a \notin \tau^{\#}$ we take $a_{\tau} = 0$; (ii) if $a \in \tau^{\#}$, we take a_{τ} to be the subset of X_{τ} consisting of all atoms of \mathcal{M}_{τ} that are contained in a (so a is the join of a_{τ} in \mathcal{M}).

Therefore we have embedded \mathcal{M} into the probability algebra of a probability space. The proof is completed by applying Lemma 2.2.

In the rest of this section we aim to discuss elementary equivalence of probability algebras and to characterize (axiomatize) the complete extensions of Pr. This depends on studying the definability in continuous logic of the set of atoms (and some related sets) in models of Pr. (See [7, Section 9] for a discussion of definable predicates and definable sets.)

In the rest of this section \mathcal{M} denotes a model of Pr, with underlying boolean algebra \mathcal{B} , measure μ and metric d. We let $A_1^{\mathcal{M}}$ denote the set of atoms of \mathcal{B} together with 0. Note that for each r > 0 there are only finitely many $a \in A_1^{\mathcal{M}}$ such that $\mu(a) \geq r$. Therefore $A_1^{\mathcal{M}}$ is finite or countable; the join (union) of $A_1^{\mathcal{M}}$ is therefore in \mathcal{B} and provides a measurable splitting of 1 in \mathcal{B} between its atomic and atomless parts, either of which may be 0. Also, $A_1^{\mathcal{M}}$ is a closed set with respect to the metric d.

We consider the following formulas in the signature of Pr:

$$\chi(x) := \inf_{y} |\mu(x \cap y) - \mu(x \cap y^{c})|$$

$$\psi(x) := \mu(x) \div \chi(x)$$

$$\varphi_{1}(x) := \inf_{z} (d(x, z) \dotplus \psi(z))$$

$$\theta(x) := \sup_{y} \inf_{z} |\mu(x \cap y \cap z) - \mu(x \cap y \cap z^{c})|$$

To understand the meanings of these formulas in models of Pr, the next result is needed. The elementary argument needed for the proof is given in [16, Section 41, Theorem A].

4.6. **Lemma.** Suppose $\mathcal{M} = (\mathcal{B}, \mu, d) \models Pr$. If $b \in \mathcal{B}$ is atomless, then for every $\delta > 0$ there is a partition of 1 in \mathcal{B} , say $u = (u_1, \ldots, u_n)$, such that $\mu(b \cap u_i) \leq \delta$ for all $i = 1, \ldots, n$.

Proof. It is sufficient to prove the result assuming b = 1 in \mathcal{B} . By the downward Löwenheim-Skolem Theorem, \mathcal{M} has a separable elementary substructure \mathcal{M}' , which is necessarily also atomless, and it obviously suffices to prove the Lemma for \mathcal{M}' . Suppose \mathcal{M}' is based on the

algebra \mathcal{B}' , which is a closed subalgebra of \mathcal{B} , and the predicates of \mathcal{M}' are the restrictions of μ and d to \mathcal{B}' . By separability of \mathcal{M}' , we may take $(\mathcal{A}_n \mid n \geq 1)$ to be an increasing family of finite boolean subalgebras of \mathcal{B}' such that $\bigcup (\mathcal{A}_n \mid n \geq 1)$ is a dense subset of \mathcal{B}' . For each $n \geq 1$, let π_n be the partition of 1 in \mathcal{A}_n that consists of the atoms of \mathcal{A}_n . The argument for Theorem A in [16, Section 41] shows that these partitions satisfy the conclusion of the Lemma.

4.7. **Proposition.** Let $\mathcal{M} = (\mathcal{B}, \mu, d) \models Pr \text{ and } b \in \mathcal{B}$.

- (a) If b is atomless, then $\chi^{\mathcal{M}}(b) = 0$.
- (b) b is an atom or 0 if and only if $\chi^{\mathcal{M}}(b) = \mu(b)$.
- (c) If b is not atomless and a is an atom of largest measure contained in b, then $\chi^{\mathcal{M}}(b) \leq \mu(a)$.
- (d) dist $(b, A_1^{\mathcal{M}}) = \varphi_1^{\mathcal{M}}(b)$.
- (e) b is atomless in \mathcal{B} if and only if $\theta^{\mathcal{M}}(b) = 0$.

Proof. (a) Fix $\delta > 0$ and use Lemma 4.6 to obtain a partition of 1 in \mathcal{B} , say $u = (u_1, \ldots, u_n)$ such that $\mu(b \cap u_i) \leq \delta$ for all $i = 1, \ldots, n$. Let $a_i = b \cap u_i$ for all i, so $b = a_1 \cup \cdots \cup a_n$. There exists i such that $\mu(a_1 \cup \cdots \cup a_i) \leq \frac{1}{2}\mu(b) \leq \mu(a_1 \cup \cdots \cup a_{i+1})$. Then $y = a_1 \cup \cdots \cup a_i$ witnesses $\chi^{\mathcal{M}}(b) \leq \delta$.

(b) If b is an atom and a is arbitrary, then one of the events $b \cap a, b \cap a^c$ equals b and the other is 0. In that case $|\mu(b \cap a) - \mu(b \cap a^c)| = \mu(b)$ for all a, so indeed $\chi^{\mathcal{M}}(b) = \mu(b)$.

If b is not an atom, there exists $a \in \mathcal{B}$ such that $\mu(b) > \mu(b \cap a) > 0$ and $\mu(b) > \mu(b \cap a^c) > 0$, from which it follows that $|\mu(b \cap a) - \mu(b \cap a^c)| < \mu(b)$.

- (c) Suppose b is not atomless and let a_1, a_2, \ldots be a listing of all the (finitely or countably many) distinct atoms of \mathcal{B} contained in b, arranged so that $\mu(a_1) \geq \mu(a_2) \geq \ldots$. Take $u \subseteq b$ to be the union of all a_j such that j is odd and $v \subseteq b$ to be the union of all a_j such that j is even. Then u, v are disjoint and $b \setminus (u \cup v)$ is atomless. One checks easily that $\chi^{\mathcal{M}}(b) \leq \mu(u) \mu(v) \leq \mu(a_1)$.
- (d) The key idea is this: if b is atomless, then $\operatorname{dist}(b, A_1^{\mathcal{M}}) = \mu(b)$; if b is not atomless and a is an atom of largest measure contained in b, then $\operatorname{dist}(b, A_1^{\mathcal{M}}) = \mu(b) \mu(a)$.

Therefore, from (a) and (c) we conclude that $\operatorname{dist}(b, A_1^{\mathcal{M}}) \leq \psi^{\mathcal{M}}(b)$ for all b. From (b) we see that $\psi^{\mathcal{M}}(b) = 0$ when b is an atom, and therefore $A_1^{\mathcal{M}}$ is the zeroset of $\psi^{\mathcal{M}}$. This makes it clear that $\operatorname{dist}(b, A_1^{\mathcal{M}}) \geq \varphi_1^{\mathcal{M}}(b)$. Conversely, for every b we have

$$\varphi_1^{\mathcal{M}}(b) \ge \inf_z(d(b,z) \dotplus \operatorname{dist}(z,A_1^{\mathcal{M}})) \ge \operatorname{dist}(b,A_1^{\mathcal{M}})$$

which completes the proof.

(e) This follows from (a) and (b). Note that $\theta(b) = 0$ is equivalent to saying $\chi(u) = 0$ holds for every $u \leq b$.

Proposition 4.7(d) shows that $A_1^{\mathcal{M}}$ is a definable set, uniformly in all models \mathcal{M} of Pr. (See [7, Definition 9.16].) It is useful to introduce for each n > 1 the further set

$$A_n^{\mathcal{M}} = \{x_1 \cup \dots \cup x_n \mid x_1, \dots, x_n \in A_1^{\mathcal{M}}\}.$$

Note that $A_1^{\mathcal{M}} \subseteq A_2^{\mathcal{M}} \subseteq \cdots \subseteq A_n^{\mathcal{M}} \subseteq \ldots$ Using [7, Theorem 9.17] and the definability of the set $A_1^{\mathcal{M}}$, we may conclude that $A_n^{\mathcal{M}}$ is a definable set in all models \mathcal{M} of Pr, for all $n \geq 1$. Indeed, as we show next, the distance to $A_n^{\mathcal{M}}$ is given explicitly by the following formula in the signature of Pr (where we define the formulas for n > 1 by induction on n):

$$\varphi_n(x) = \inf_{w} (\varphi_{n-1}(x \cap w) \dotplus \varphi_1(x \cap w^c))$$

4.8. **Proposition.** Let $\mathcal{M} = (\mathcal{B}, \mu, d) \models Pr \text{ and } a \in \mathcal{B}$. Then for each $n \geq 1$

$$\operatorname{dist}(a, A_n^{\mathcal{M}}) = \varphi_n^{\mathcal{M}}(a).$$

Proof. Let $a \in \mathcal{B}$ and let a_1, a_2, \ldots be a listing of all distinct atoms contained in a, arranged so that $\mu(a_1) \geq \mu(a_2) \geq \ldots$, and extended to an infinite sequence by taking $a_k = 0$ for larger k, if necessary. We note that $\operatorname{dist}(a, A_n^{\mathcal{M}}) = d(a, w)$ where $w = a_1 \cup \cdots \cup a_n$, and for this w we have $\operatorname{dist}(a, w) = \mu(a) - \mu(a_1 \cup \cdots \cup a_n) = \mu(a) - (\mu(a_1) + \cdots + \mu(a_n))$.

To prove the Lemma, we argue by induction on $n \ge 1$. The n = 1 case is Proposition 4.7(d). Taking n > 1, it remains to prove the induction step from n - 1 to n.

First note that if we take $w = a_1 \cup \cdots \cup a_{n-1}$ we have $\varphi_{n-1}^{\mathcal{M}}(a \cap w) = 0$ (by the induction hypothesis) and $\varphi_1^{\mathcal{M}}(a \cap w^c) = \mu(a \cap w^c) - \mu(a_n)$ (by Proposition 4.7(d)). Therefore, for this w we have

$$\varphi_{n-1}^{\mathcal{M}}(a \cap w) + \varphi_1^{\mathcal{M}}(a \cap w^c) = \mu(a \cap w^c) - \mu(a_n)$$

$$= (\mu(a) - \mu(a_1 \cup \dots \cup a_{n-1})) - \mu(a_n)$$

$$= \mu(a) - (\mu(a_1) + \dots + \mu(a_{n-1}) + \mu(a_n))$$

$$= \operatorname{dist}(a, A_n^{\mathcal{M}}).$$

To finish the argument, it suffices to prove that for any other w we have

$$\left(\varphi_{n-1}^{\mathcal{M}}(a\cap w) \dotplus \varphi_1^{\mathcal{M}}(a\cap w^c)\right) \ge \mu(a) - (\mu(a_1) + \dots + \mu(a_n)).$$

So fix $w \in M$ and let a_1^1, a_2^1, \ldots be a listing of all distinct atoms contained in $a \cap w$, arranged so that $\mu(a_1^1) \geq \mu(a_2^1) \geq \ldots$, and extended to an infinite sequence by taking $a_k^1 = 0$ if necessary. Also, let a_1^2 be one of the the largest atoms contained in $a \cap w^c$ (which can be 0). Note that the nonzero elements among $a_1^2, a_1^1, \ldots, a_{n-1}^1$ are distinct atoms, and all are $\leq a$. By the induction hypothesis, $\varphi_{n-1}^{\mathcal{M}}(a \cap w) = \mu(a \cap w) - \sum_{i=1}^{n-1} \mu(a_i^1)$ and by Proposition

4.7(d)
$$\varphi_1^{\mathcal{M}}(a \cap w^c) = \mu(a \cap w^c) - \mu(a_1^2)$$
. Thus

$$\varphi_{n-1}^{\mathcal{M}}(a \cap w) \dotplus \varphi_{1}^{\mathcal{M}}(a \cap w^{c}) = \left(\mu(a \cap w) - \sum_{i=1}^{n-1} \mu(a_{i}^{1})\right) + \left(\mu(a \cap w^{c}) - \mu(a_{1}^{2})\right)$$
$$= \mu(a) - \left(\sum_{i=1}^{n-1} \mu(a_{i}^{1}) + \mu(a_{1}^{2})\right)$$

The smallest possible value of this last expression occurs when $a_2^1, a_1^1, \ldots, a_{n-1}^1$ have the largest possible measures, which happens when the sequence $\mu(a_2^1), \mu(a_1^1), \ldots, \mu(a_{n-1}^1)$ is a permutation of $\mu(a_1), \ldots, \mu(a_n)$.

For a similar treatment of atoms in the setting of random variable structures see [3, Lemma 2.16].

4.9. **Remark.** Let $\mathcal{M} = (\mathcal{B}, \mu, d) \models Pr$ and $a \in \mathcal{B}$. Propositions 4.7(d) and 4.8 make it clear that

$$\mu(a) \ge \varphi_1^{\mathcal{M}}(a) \ge \varphi_2^{\mathcal{M}}(a) \ge \dots \ge \varphi_n^{\mathcal{M}}(a) \ge \dots$$

4.10. **Notation.** Let $\mathcal{M} = (\mathcal{B}, \mu, d) \models Pr$ and $a \in \mathcal{B}$; let a_1, a_2, \ldots be a listing of all distinct atoms contained in a, arranged so that $\mu(a_1) \geq \mu(a_2) \geq \ldots$, and extended to an infinite sequence by taking $a_k = 0$ for larger k, if necessary. For each $n \geq 1$, we refer to $\mu(a_n)$ as the n^{th} largest measure of an atom contained in a, and we denote this number as $at_n^{\mathcal{M}}(a)$.

Note that the nonzero elements of $(a_n \mid n \in \mathbb{N})$ are distinct, whereas the measure values $(\mu(a_n) \mid n \in \mathbb{N})$ may contain repetitions.

4.11. Corollary. For each $n \geq 1$, the predicate at_n is definable in all models of Pr. Indeed, if $\mathcal{M} = (\mathcal{B}, \mu, d) \models Pr$ and $a \in \mathcal{B}$, then

$$at_1^{\mathcal{M}}(a) = \mu(a) \div \varphi_1^{\mathcal{M}}(a)$$

and for each n > 1

$$at_n^{\mathcal{M}}(a) = \varphi_{n-1}^{\mathcal{M}}(a) \div \varphi_n^{\mathcal{M}}(a).$$

Proof. This is immediate from Proposition 4.8.

We now consider an extension by definitions of Pr obtained by adding unary predicate symbols $(P_n \mid n \ge 1)$ to the signature and by adding as axioms the conditions

$$\sup_{x} |P_1(x) - (\mu(x) \div \varphi_1(x))| = 0$$

and for n > 1

$$\sup_{x} |P_n(x) - (\varphi_{n-1}(x) \div \varphi_n(x))| = 0.$$

This extension of Pr is denoted by Pr^* . Note that each model \mathcal{M} of Pr has a unique expansion, which we denote by \mathcal{M}^* , that is a model of Pr^* . This expansion is given by interpreting each P_n so that, for each $a \in \mathcal{M}$, one takes $P_n^{\mathcal{M}^*}(a)$ to be the n^{th} largest measure of an atom contained in a.

4.12. **Notation.** Suppose $\mathcal{M} = (\mathcal{B}, \mu, d) \models Pr$. Fix $n \geq 1$ and consider any tuple $a = (a_1, \ldots, a_n) \in M^n = \mathcal{B}^n$. Let $e = (e_1, \ldots, e_{2^n})$ be the partition of 1 in the boolean algebra \mathcal{B} generated by a_1, \ldots, a_n . By this we mean that the elements of e are all possible intersections of the form $a_1^{k_1} \cap \cdots \cap a_n^{k_n}$, where each k_i comes from $\{-1, +1\}$, and we list these intersections in order according to lexicographic order on the tuples of superscripts k_1, \ldots, k_n . We refer to e as the partition of 1 in \mathcal{B} associated to e. Further, note that when (a_1, \ldots, a_n) and (e_1, \ldots, e_{2^n}) are as above, then each e is the union of the coordinates e of e that are intersections e in which e in which e in the correspondence between coordinates of e in the coordinates e in the coordinates of e in the coordinates e in the coo

For simplicity of notation, we write a^s for $a_1^{k_1} \cap \cdots \cap a_n^{k_n}$ when $s = (k_1, \ldots, k_n)$ is an arbitrary element of $\{-1, +1\}^n$ and $a = (a_1, \ldots, a_n) \in M^n$. Likewise we write x^s for the boolean term $x_1^{k_1} \cap \cdots \cap x_n^{k_n}$ when $s = (k_1, \ldots, k_n)$ and x stands for the tuple (x_1, \ldots, x_n) of variables. As indicated above, the identity

$$x_i = \bigcup_{s} (x^s \mid s = (k_1, \dots, k_n) \text{ and } k_i = +1)$$

is true in all models of Pr. Frequently when we use this notation, as here, we omit the standard specifications that $x = (x_1, \ldots, x_n)$ and $s = (k_1, \ldots, k_n) \in \{-1, +1\}^n$. In particular, we view k_i as a function of s when $s \in \{-1, +1\}^n$. When it is needed, we list the elements of $\{-1, +1\}^n$ in lexicographical order.

4.13. **Remark.** For future use we note that for every L^{pr} -formula $\varphi(x)$ there exists an L^{pr} -formula $\psi(y_s \mid s \in \{-1, +1\}^n)$, such that $\varphi(x)$ is equivalent to $\psi(x^s \mid s \in \{-1, +1\}^n)$ in all models of Pr. Indeed, it suffices to take $\psi(y_s \mid s \in \{-1, +1\}^n)$ to be the result of substituting the boolean term $\bigcup (y_s \mid s = (k_1, \ldots, k_n) \in \{-1, +1\}^n$ and $k_i = +1)$ for the variable x_i in $\varphi(x)$, for $i = 1, \ldots, n$.

Further, for every L^{pr} -formula $\varphi(x)$ there exists an L^{pr} -formula $\psi(x)$ such that $\psi(x)$ is Pr-equivalent to $\varphi(x)$ and every atomic formula occurring in $\psi(x)$ is of the form $\mu(x^s)$ for some $s \in \{-1, +1\}^n$. Moreover, $\psi(x)$ can be chosen so that it is obtained from such atomic formulas using the restricted connectives $0, 1, t \mapsto t/2$, and $(t, u) \mapsto t \div u$.

(Proof: A general atomic formula $\alpha(x)$ in L^{pr} can be taken to be one of the form $\mu(t(x))$ where t(x) is a boolean term in x. If $\alpha(x) = d(t_1(x), t_2(x))$, then $\alpha(x)$ can be replaced by

 $\mu(t_1(x)\Delta t_2(x))$. For each such t(x) there is a subset $S \subseteq \{-1, +1\}$ such that the equation $t(x) = \bigcup (x^s \mid s \in S)$ is true in all models of Pr. (If S is empty, then t(x) = 0 is true in all models of Pr.) Moreover, $\mu(t(x)) = \widehat{\sum}(\mu(x^s) \mid s \in S)$ is true in all models of Pr, where by $\widehat{\sum}$ we mean the connective $(u_1, \ldots, u_{2^n}) \mapsto \min(\sum (u_s \mid s \in S), 1)$. For the "Moreover" statement, including treatment of the connectives $\widehat{\sum}$, see [7, Chapter 6].)

A consequence of the preceding observation is that when C is a subalgebra of a model \mathcal{M} of Pr, then for any $a \in M^n$ the type $\operatorname{tp}_{\mathcal{M}}(a/C)$ is determined by the values $\psi^{\mathcal{M}}(a)$ of L^{pr} -formulas $\psi(x)$ over C in which all atomic formulas are of the form $\mu(x^s \cap c)$ for some $c \in C$. (As above, $x = (x_1, \ldots, x_n)$ and $s \in \{-1, +1\}^n$.)

4.14. **Theorem.** The theory Pr^* admits quantifier elimination.

Proof. We use [8, Theorem 4.16], so we need to show that Pr^* has the back-and-forth property given in [8, Definition 4.15]. Therefore, consider two ω -saturated models \mathcal{M}^* , \mathcal{N}^* of Pr^* and tuples (a_1, \ldots, a_n) in \mathcal{M}^* ; (b_1, \ldots, b_n) in \mathcal{N}^* such that the quantifier-free type of (a_1, \ldots, a_n) in \mathcal{M}^* is the same as the quantifier-free type of (b_1, \ldots, b_n) in \mathcal{N}^* . Given any u in \mathcal{M}^* we need to find v in \mathcal{N}^* such that (a_1, \ldots, a_n, u) and (b_1, \ldots, b_n, v) have the same quantifier-free type in the language of Pr^* . It suffices to do this for the case in which (a_1, \ldots, a_n) and (b_1, \ldots, b_n) are partitions of 1, by the discussion in 4.12.

Using Lemma 4.6 and the fact that \mathcal{M} is ω -saturated, for each atomless $c \in M$ and 0 < r < 1 there exists $a \le c$ in M such that $\mu(a) = r\mu(c)$, and hence $\mu(c \cap a^c) = (\mu(c) - r)\mu(c)$. Indeed, x = a can be taken to satisfy all of the conditions $|\mu(x) - r\mu(c)| \le \delta$ for $\delta > 0$, which we just showed were finitely satisfiable in \mathcal{M} . This is used in the next paragraph.

In the assumed situation we know that for each j = 1, ..., n we have $\mu(a_j) = \mu(b_j)$ and, for all $k \geq 1$, we also have $P_k^{\mathcal{M}^*}(a_j) = P_k^{\mathcal{N}^*}(b_j)$. For each j, let a_j^0 be the atomic part of a_j (i.e., the union of the atoms of \mathcal{M} that are $\leq a_j$), so $a_j^1 := a_j \cap (a_j^0)^c$ is the atomless part of a_j . Define b_j^0, b_j^1 from b_j similarly. Our assumptions yield that $\mu(a_j^0) = \mu(b_j^0)$ (and indeed, that the atoms below a_j and b_j are in a bijective, measure-preserving correspondence). Hence also $\mu(a_j^1) = \mu(b_j^1)$.

Take any u in \mathcal{M}^* and fix $j=1,\ldots,n$. Define $v_j^0 \leq b_j^0$ to be the union of the atoms below b_j that correspond to atoms below $a_j \cap u$. Further, choose $v_j^1 \leq b_j^1$ so that $\mu(v_j^1) = \mu(a_j \cap u) - \mu(v_j^0)$, and let $v_j = v_j^0 \cup v_j^1$. We obtain

$$\mu(v_j) = \mu(a_j \cap u)$$
, and
$$P_k^{\mathcal{N}^*}(v_j) = P_k^{\mathcal{M}^*}(a_j \cap u) \text{ for all } k \ge 1.$$

Then let $v = v_1 \cup \cdots \cup v_n$, and note that $v_j = b_j \cap v$ for all j. It follows that the quantifier-free type of (a_1, \ldots, a_n, u) in \mathcal{M}^* is the same as the quantifier-free type of (b_1, \ldots, b_n, v) in \mathcal{N}^* , as desired.

Theorem 4.14 allows us to characterize (and axiomatize) the complete extensions of Pr.

4.15. **Definition.** For any $\mathcal{M} = (\mathcal{B}, \mu, d) \models Pr$, let $\Phi^{\mathcal{M}}$ denote the sequence $(at_n^{\mathcal{M}}(1) \mid n \geq 1)$, which lists the sizes of the atoms of \mathcal{B} in decreasing order (and then has a tail of 0s if there are only finitely many atoms in \mathcal{B}).

Note that the range of the operator Φ consists of all the sequences $(t_n \mid n \geq 1)$ such that $1 \geq t_1 \geq t_2 \geq \cdots \geq 0$ and $\sum_{n=1}^{\infty} t_n \leq 1$.

4.16. Corollary. For models \mathcal{M}, \mathcal{N} of Pr, we have that $\mathcal{M} \equiv \mathcal{N}$ if and only if $\Phi^{\mathcal{M}} = \Phi^{\mathcal{N}}$. Therefore, any complete extension T of Pr (in the same signature) can be axiomatized by adding to Pr the conditions $\varphi_1(1) = 1 - t_1$ and $\varphi_{n-1}(1) \div \varphi_n(1) = t_n$ (for n > 1), where $(t_n \mid n \geq 1)$ is the common value of $\Phi^{\mathcal{M}}$ for $\mathcal{M} \models T$.

Proof. Let \mathcal{M}, \mathcal{N} be models of Pr such that $\Phi^{\mathcal{M}} = \Phi^{\mathcal{N}}$. From the definition of the operator Φ we see that 1 has the same quantifier-free type in \mathcal{M}^* as in \mathcal{N}^* . Theorem 4.14 yields that $\mathcal{M}^* \equiv \mathcal{N}^*$, from which it follows that $\mathcal{M} \equiv \mathcal{N}$. The converse and the rest of the Corollary follow using the definition of Φ .

4.17. Corollary. Every completion T of Pr (in L^{pr}) is separably categorical, and the unique separable model of T is strongly ω -homogeneous.

Proof. Let T be a completion of Pr and let \mathcal{M} be a separable model of T. A strengthening of Lemma 4.6 that is proved in [16, Section 41] says that the key property used in the proof of Theorem 4.14 is actually true in all models, without assuming they are ω -saturated. That is, for each atomless $c \in \mathcal{M}$ and 0 < r < 1 there exists $a \le c$ in \mathcal{M} such that $\mu(a) = r\mu(c)$, and hence $\mu(c \cap a^c) = (\mu(c) - r)\mu(c)$.

So the proof of Theorem 4.14 not only shows that Pr^* admits quantifier elimination, but shows further that for each $\mathcal{M}^* \models Pr^*$ and every (a_1, \ldots, a_n) in \mathcal{M}^* , every 1-type over (a_1, \ldots, a_n) for the theory of \mathcal{M}^* is realized in \mathcal{M}^* . In other words, every model of Pr^* is ω -saturated. It follows trivially from the definition that the same is true of every structure $(\mathcal{M}, a_1, \ldots, a_n)$ where \mathcal{M} is a model of Pr and $a_1, \ldots, a_n \in \mathcal{M}$.

It is routine to show that if $\mathcal{M}_1, \mathcal{M}_2$ are ω -saturated separable metric structures for the same language, and $\mathcal{M}_1, \mathcal{M}_2$ are elementarily equivalent, then \mathcal{M}_1 and \mathcal{M}_2 are isomorphic. One uses the usual inductive back-and-forth argument to produce an elementary bijection $f: S_1 \to S_2$, where S_i is a dense subset of M_i for both values of i. Then f extends to a map on M_1 that is an isomorphism from \mathcal{M}_1 onto \mathcal{M}_2 . The proof of the corollary is completed by applying this construction to models of the form $(\mathcal{M}, a_1, \ldots, a_n)$ discussed above. \square

4.18. **Remark.** Corollary 4.17 yields the following well known fact due to Carathéodory: if (X, \mathcal{B}, μ) and (Y, \mathcal{C}, ν) are atomless, countably generated measure spaces with $\mu(X) =$

- $\nu(Y) < \infty$, then the measured algebras of (X, \mathcal{B}, μ) and (Y, \mathcal{C}, ν) are isomorphic. Proof: Without loss of generality, we may take $\mu(X) = \nu(Y) = 1$. In that case, the measured algebras of these two probability spaces are atomless, separable models of Pr. Using Theorem 4.14 we see that these probability algebras are elementarily equivalent (since in an atomless probability algebra the predicates interpreting P_n are identically 0) and hence by Corollary 4.17 we get the desired result. For proofs in analysis see [16, Section 41] and [23, Theorem 4, p. 399].
- 4.19. Corollary. Let T be any complete L^{pr} -theory that extends Pr and let $(t_n \mid n \geq 1)$ be the common value of $\Phi^{\mathcal{M}}$ for $\mathcal{M} \models T$.
- (1) If $\sum_{n=1}^{\infty} t_n = 1$, then T has a unique model, which consists of an atomic probability algebra having atoms $(a_n \mid n \geq 1 \text{ and } t_n > 0)$ with $\mu(a_n) = t_n$ for all n.
- (2) If $\sum_{n=1}^{\infty} t_n < 1$, then the models of T are exactly the probability algebras with atoms as described in (1) together with an atomless part of measure $1 \sum_{n=1}^{\infty} t_n$.

Proof. These statements are immediate from Corollaries 4.16 and 4.17. \Box

- 4.20. **Remark.** Let \mathcal{M} be any model of Pr. From the previous results it follows that $\operatorname{acl}_{\mathcal{M}}(\emptyset)$ is the σ -subalgebra generated by the atoms in \mathcal{M} .
- 4.21. **Remark.** If $\mathcal{M} \models Pr$, let $a_0 \in M$ be the join of the atoms of \mathcal{M} , and let $a_1 = a_0^c$. Further, let $\mathcal{A}_i = \{a \in M \mid a \leq a_i\}$ for each i = 0, 1. Thus \mathcal{A}_0 is the set of atomic elements of \mathcal{M} and \mathcal{A}_1 is the set of atomics elements, and a_1 is the largest atomics element. The partition $\{a_0, a_1\}$ of 1 is important because it splits every element a of M into its atomic part $a \cap a_0$ and its atomics part $a \cap a_1$. We note the following facts concerning the definability of these elements and sets:
- (i) Relative to all models \mathcal{M} of Pr: \mathcal{A}_0 is not a zeroset; \mathcal{A}_1 is a zeroset but is not a definable set; a_0, a_1 are not definable elements.
- (ii) Fix a complete extension T of Pr. Relative to all models of T: a_0, a_1 are definable elements and $\mathcal{A}_0, \mathcal{A}_1$ are definable sets.
- Proof. (i) First we show \mathcal{A}_0 is not a zeroset over Pr. Suppose otherwise, so there exists a definable predicate R(x) over Pr such that for all $\mathcal{M} \models Pr$ and all $a \in \mathcal{M}$, we have $R^{\mathcal{M}}(a) = 0$ if and only if a is an atomic element in \mathcal{M} . For each $n \geq 1$, let \mathcal{M}_n be the probability algebra of the probability space having n points, each of which has measure 1/n, and let \mathcal{M} be the metric ultraproduct of $(\mathcal{M}_n \mid n \geq 1)$ with respect to a nonprincipal ultrafilter. Then 1 is atomic in every \mathcal{M}_n while it is not atomic in \mathcal{M} ; indeed, \mathcal{M} is atomless, so 0 is its only atomic element. This means $R^{\mathcal{M}}(1) \neq 0$ whereas $R^{\mathcal{M}_n}(1) = 0$ for all n. This violates the Fundamental Theorem of ultraproducts for the definable predicate R. (See Theorem 5.4 and the discussion of extensions by definition in Section 9 in [7].)

Proposition 4.7(e) shows that \mathcal{A}_1 is the zeroset of the formula $\theta(x)$ in all models of Pr. Further, \mathcal{A}_1 cannot be a definable set uniformly in all models of Pr, since otherwise there would be a definable predicate R(x) over Pr such that for all $\mathcal{M} \models Pr$ and all $a \in M$

$$R^{\mathcal{M}}(a) = \sup\{\mu(a \cap y) \mid y \in \mathcal{A}_1\}.$$

But then the zeroset of R(x) would be A_0 in all models of Pr, which we just proved is not possible.

Since $a_0^c = a_1$, they are either both definable or both undefinable. We work with a_0 . If a_0 were definable over Pr, then the operation $x \mapsto a_0 \cap x$ would be definable, so its image, which is \mathcal{A}_0 , would be a definable set, but it isn't.

(ii) Consider a complete extension T of Pr and let \mathcal{U} be an ω_1 -universal domain for T. By Corollary 4.19, we see that a_0 and a_1 are fixed by every automorphism of \mathcal{U} . Using [7, Exercise 10.7 or Theorem 9.32] it follows that a_0 and a_1 are each definable uniformly in all models of T. Finally, note that for each i=0,1, the set \mathcal{A}_i is the image of the definable function f_i defined by $f_i(x) = x \cap a_i$. Using [7, Theorem 9.17], we infer that \mathcal{A}_i is a definable set uniformly in all models of T.

4.22. **Exercise.** By Remark 4.9, for any $\mathcal{M} \models Pr$ and any $a \in M$, we have that $(\varphi_n^{\mathcal{M}}(a)) \mid n \geq 1$) is a decreasing sequence from [0,1], so it converges, and its limit must be the distance from a to the set of atomic elements of \mathcal{M} , which is the closure of $\cup_n A_n^{\mathcal{M}}$. Show that if we restrict attention to models of a completion T of Pr, this convergence is uniform (in \mathcal{M} as well as a), so its limit is a T-definable predicate. This gives an alternative proof of Remark 4.21(ii).

5. Random variables

Here we discuss how to represent the space RV of [0, 1]-valued random variables (modulo pointwise equality a.e.), equipped with the L_1 -distance, as a metric imaginary sort for the theory Pr. We focus on the measure theoretic aspects of the matter, and avoid many technical details of the meq construction, for which we refer to various articles for the details (see Section 3). An interpretation of random variables in atomless probability algebras was originally given by Ben Yaacov in [1] in the CAT setting, and extended to an interpretation of RV in Pr by him in [3]. In [1] certain other classes of random variables are also treated, and the approach used here can easily be extended to apply to them.

When \mathcal{M} is an arbitrary model of Pr, we let (X, \mathcal{B}, μ) denote a probability space whose probability algebra $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$ is (isomorphic to) \mathcal{M} . In much of this section we argue in (X, \mathcal{B}, μ) using measure theory.

We denote by $RV = RV(\mathcal{B}, \mu) := L_1(\mathcal{B}, \mu; [0, 1])$ the space of all \mathcal{B} -measurable functions $f \colon X \to [0, 1]$. We equip this space with the L_1 (pseudo)metric, for which the distance between f and g is $||f - g||_1 = \int_X |f - g| d\mu$. As done here for probability spaces, we consider the quotient metric space obtained by identifying two random variables if they are equal pointwise μ -a.e., equipped with the distance induced by $||f - g||_1$ (which we call the L_1 -distance).

The goal of this section is to explain how this quotient can be seen as a metric imaginary sort for the model $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$ of Pr. One value of doing so is that it allows seeing $\mathbb{P}(a|\mathcal{A})$ as an imaginary in $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})^{\text{meq}}$, for any $a \in \widehat{\mathcal{B}}$ and any closed subalgebra \mathcal{A} of $\widehat{\mathcal{B}}$.

For $n \geq 1$, consider the subset $RV_n = RV_n(\mathcal{B}, \mu)$ of $L_1(\mathcal{B}, \mu; [0, 1])$ consisting of those functions that can be written as $\sum_{i=1}^n \frac{i}{n} \chi_{E_i}$ where $E = (E_1, \dots, E_n)$ is a partition of X from \mathcal{B} . Below we denote $\sum_{i=1}^n \frac{i}{n} \chi_{E_i}$ by f_E . On $RV_n(\mathcal{B}, \mu)$ we take the L_1 -distance. If $f_E \in RV_m$ and $f_F \in RV_n$, we have f_E, f_F both in $L_1(\mathcal{B}, \mu; [0, 1])$, so we can compare them, compute the L_1 -distance between them, etc., even if $m \neq n$.

Of course we may identify RV_n with the set of $E = (E_1, \ldots, E_n)$ that are partitions of X in \mathcal{B} . For two such partitions E, F, we define $\rho_n(E, F)$ to be the L_1 -distance;

(A)
$$\rho_n(E, F) := \|f_E - f_F\|_1 = \left\| \sum_{i=1}^n \frac{i}{n} \chi_{E_i} - \sum_{i=1}^n \frac{i}{n} \chi_{F_i} \right\|_1$$
$$= \frac{1}{n} \sum_{i \neq j} |i - j| \mu(E_i \cap F_j).$$

By analyzing the expressions in (A), we obtain (in the following Lemma) a Lipschitz equivalence between $\rho_n(E, F)$ and $d_P(E, F) := \frac{1}{2} \sum_{i=1}^n d(E_i, F_i) = \frac{1}{2} \sum_{i=1}^n \mu(E_i \triangle F_i)$. Note that when n = 2 we have $d_P(E, F) = \frac{1}{2} (d(E_1, F_1) + d(E_1^c, F_1^c)) = d(E_1, F_1)$, since $d(E_1^c, F_1^c) = d(E_1, F_1)$; this explains the factor $\frac{1}{2}$. Also, d_P is equivalent to the usual pseudometric $d(E, F) := \max_i d(E_i, F_i)$, since

$$\frac{1}{2}d(E,F) \le d_P(E,F) \le \frac{n}{2}d(E,F)$$

holds for all $E, F \in RV_n$.

5.1. **Lemma.** For all $E, F \in RV_n$ we have

$$\frac{1}{n}d_P(E,F) \le \rho_n(E,F) = ||f_E - f_F||_1 \le d_P(E,F).$$

Proof. First note that for each i the family $(E_i \cap F_j \mid j \neq i)$ is a partition of $E_i \setminus F_i$, and the same with E and F interchanged. Therefore

$$\sum_{i=1}^{n} \mu(F_i \setminus E_i) = \sum_{i \neq j} \mu(E_i \cap F_j) = \sum_{i=1}^{n} \mu(E_i \setminus F_i),$$

from which follows

$$\sum_{i \neq j} \mu(E_i \cap F_j) = \frac{1}{2} \sum_{i=1}^n \mu(E_i \triangle F_i) = d_P(E, F).$$

Then we get the desired inequality using (A) together with

$$\frac{1}{n}\sum_{i\neq j}\mu(E_i\cap F_j)\leq \frac{1}{n}\sum_{i\neq j}|i-j|\mu(E_i\cap F_j)\leq \sum_{i\neq j}\mu(E_i\cap F_j).$$

Write $\widehat{RV}_n = \widehat{RV}_n(\mathcal{B}, \mu)$ for the image of RV_n under the quotient map from \mathcal{B} onto $\widehat{\mathcal{B}}$. If \mathcal{M} is the model $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$, we also denote \widehat{RV}_n by $\widehat{RV}_n^{\mathcal{M}}$, and note that it satisfies

$$\widehat{RV}_n^{\mathcal{M}} = \{(e_1, \dots, e_n) \in M^n \mid (e_1, \dots, e_n) \text{ is a partition of 1 in } \mathcal{M}\}.$$

We put on $\widehat{RV}_n^{\mathcal{M}}$ the pseudometric $\widehat{\rho}_n^{\mathcal{M}}$ obtained canonically from ρ_n on RV_n ; that is, for $e, f \in \widehat{RV}_n^{\mathcal{M}}$, we have

(B)
$$\widehat{\rho}_n^{\mathcal{M}}(e, f) = \sum_i \sum_j \left| \frac{i}{n} - \frac{j}{n} \right| \mu(e_i \cap f_j)$$

$$= \frac{1}{n} \sum_i \sum_j \left| i - j \right| \mu(e_i \cap f_j).$$

5.2. **Lemma.** Let $\mathcal{M} \models Pr$. Then $\widehat{\rho}_n^{\mathcal{M}}$ is a complete metric on $\widehat{RV}_n^{\mathcal{M}}$. Also, $\widehat{RV}_n^{\mathcal{M}}$ is a definable set, and $\widehat{\rho}_n^{\mathcal{M}}$ is a definable predicate on $\widehat{RV}_n^{\mathcal{M}}$, uniformly in all models of Pr.

Proof. Obviously $\widehat{\rho}_n^{\mathcal{M}}$ is a pseudometric. The fact that it is a metric follows from Lemma 5.1 and the definition of d_P on $RV_n^{\mathcal{M}}$, which implies that $d_P(E,F) = 0$ iff $\mu(E_i \triangle F_i)$ for all $i = 1, \ldots, n$.

To show that $\widehat{RV}_n^{\mathcal{M}}$ is uniformly a definable subset of M^n , it is sufficient to show that it is the image of a definable function on a definable set. To do this, consider the function defined on M^{n-1} by $(a_1, \ldots, a_{n-1}) \mapsto (e_1, \ldots, e_n)$ where $e_1 = a_1, e_j = a_j \cap a_1^c \cap \cdots \cap a_{j-1}^c$ for $2 \leq j \leq n-1$, and $e_n = a_1^c \cap \cdots \cap a_{n-1}^c$. Obviously this is a definable function, since the coordinates e_j are given by boolean terms. Note that $e_i \cap e_j = 0$ whenever $i \neq j$. Moreover, by induction on j < n we can show $a_1 \cup \cdots \cup a_j = e_1 \cup \cdots \cup e_j$. Therefore (e_1, \ldots, e_{n-1}) is a partition of $a_1 \cup \cdots \cup a_{n-1}$. Since $e_n = (a_1 \cup \cdots \cup a_{n-1})^c$, we have that (e_1, \ldots, e_n) is always

a partition of 1 in \mathcal{M} . To see that this map is surjective onto $\widehat{RV}_n^{\mathcal{M}}$, note that whenever (e_1, \ldots, e_n) is a partition of 1, then $(e_1, \ldots, e_{n-1}) \mapsto (e_1, \ldots, e_n)$

Equation (B) shows that $\widehat{\rho}_n^{\mathcal{M}}$ is a definable predicate, uniformly on all models \mathcal{M} of Pr.

It remains to show that $\widehat{RV}_n^{\mathcal{M}}$ is complete. Because $M = \widehat{\mathcal{B}}$ is complete with respect to the metric $d(a,b) := \widehat{\mu}(a\triangle b)$ and $\widehat{RV}_n^{\mathcal{M}} \subseteq (\widehat{\mathcal{B}})^n$ is closed, we see that $\widehat{RV}_n^{\mathcal{M}}$ is complete with respect to the metric $d_P(e,f) = \frac{1}{2} \sum_{i=1}^n d(e_i,f_i)$. Therefore $\widehat{RV}_n^{\mathcal{M}}$ is complete with respect to $\widehat{\rho}_n^{\mathcal{M}}$, since $\widehat{\rho}_n^{\mathcal{M}}$ is uniformly equivalent to d_P by Lemma 5.1.

5.3. **Remark.** Similar reasoning to that in the preceding proof shows that the map $E \mapsto f_E$ for $E \in RV_n^{\mathcal{M}}$ is an isometric map from $RV_n^{\mathcal{M}}$ into $L_1(\mathcal{B}, \mu; [0, 1])$ whose range is the collection of all random variables whose values are in $\{1/n, \ldots, n/n\}$. (Here isometric means with respect to $\rho_n^{\mathcal{M}}$ and the L_1 -distance.) Hence this map induces an isometry from $\widehat{RV}_n^{\mathcal{M}}$ onto the set of = a.e. equivalence classes of those random variables.

Fix $\mathcal{M} \models Pr$ and, as above, let (X, \mathcal{B}, μ) be a probability space whose probability algebra is (isomorphic to) \mathcal{M} . We denote by $\widehat{RV}^{\mathcal{M}}$ the inverse limit of the spaces $(\widehat{RV}_{2^n}^{\mathcal{M}} \mid n \geq 1)$ equipped with a suitable family of maps $\widehat{\pi} : \widehat{RV}_{2^{n+1}}^{\mathcal{M}} \to \widehat{RV}_{2^n}^{\mathcal{M}}$ that we now define. For $(e_1, \dots, e_{2^{n+1}}) \in \widehat{RV}_{2^{n+1}}^{\mathcal{M}}$ we set

$$\widehat{\pi}(e_1,\ldots,e_{2^{n+1}}) := (e_1 \cup e_2,\ldots,e_{2^{n+1}-1} \cup e_{2^{n+1}}).$$

This clearly makes $\widehat{\pi} : \widehat{RV}_{2^{n+1}}^{\mathcal{M}} \to \widehat{RV}_{2^n}^{\mathcal{M}}$ a definable map, uniformly for all models of Pr. Note that an element of $\widehat{RV}^{\mathcal{M}}$ is given by a sequence $(e(n))_n = (e(n) \mid n \geq 1)$ where $e(n) = (e_1(n), \dots, e_{2^n}(n)) \in \widehat{RV}_{2^n}^{\mathcal{M}}$ and $\widehat{\pi}(e(n+1)) = e(n)$ for all $n \geq 1$. In what follows, we refer to such a sequence as *coherent*.

On $\widehat{RV}^{\mathcal{M}}$ we want to define the inverse limit pseudometric, denoted by $\widehat{\rho}^{\mathcal{M}}$, by

$$\widehat{\rho}^{\mathcal{M}}((e(n))_n, (f(n))_n) := \lim_{n} \widehat{\rho}_n(e(n), f(n))$$

for any elements $(e(n))_n$, $(f(n))_n$ of $\widehat{RV}^{\mathcal{M}}$. The fact that the limit in this definition of $\widehat{\rho}^{\mathcal{M}}$ exists (with a rate of convergence that can be taken to be uniform over all sequences in $\widehat{RV}^{\mathcal{M}}$ and all $\mathcal{M} \models Pr$)) and that the resulting quotient corresponds to $L_1(\mathcal{B}, \mu; [0, 1])$ modulo the L_1 -distance follows from the results in Lemma 5.5 below. For proving such results it is useful to pull the objects involved back to the probability space (X, \mathcal{B}, μ) of which \mathcal{M} is the probability algebra.

We write π for the corresponding maps from $RV_{2^{n+1}}$ to RV_{2^n} ; namely

$$\pi(E_1,\ldots,E_{2^{n+1}}):=(E_1\cup E_2,\ldots,E_{2^{n+1}-1}\cup E_{2^{n+1}}).$$

Via the quotient map from \mathcal{B} to $M = \widehat{\mathcal{B}}$, the inverse limit $\widehat{RV}^{\mathcal{M}}$ described above corresponds to the inverse limit of the spaces $(RV_{2^n}(\mathcal{B},\mu) \mid n \geq 1)$ equipped with their L_1 -pseudometrics and the connecting maps $\pi \colon RV_{2^{n+1}} \to RV_{2^n}$. An element of this inverse limit consists of a sequence $(E(n) \mid n \geq 1)$ such that $E(n) \in RV_{2^n}$ and $\pi(E(n+1)) = E(n)$, for all $n \geq 1$. As above, we refer to such a sequence as *coherent*.

- 5.4. Fact. It is evident that the quotient map from \mathcal{B} to $\widehat{\mathcal{B}}$ induces a map from coherent sequences $(E(n) \mid n \geq 1)$ of measurable partitions of X to coherent sequences $(\widehat{E}(n) \mid n \geq 1)$ of partitions of 1 in \mathcal{M} , where $\widehat{E}(n) := (\widehat{E_1(n)}, \dots, \widehat{E_{2^n}(n)})$ for all $n \geq 1$. In fact, this map is surjective. That is, suppose $(e(n) \mid n \geq 1)$ is a coherent sequence in $\widehat{RV}^{\mathcal{M}}$, with $e(n) = (e_1(n), \dots, e_{2^n}(n))$ for all $n \geq 1$. It is easy to show, working by induction on n, that there exists a coherent sequence $(E(n) \mid n \geq 1)$ with $E(n) \in RV_{2^n}$ for all $n \in I$ such that $e_j(n) = \widehat{E_j(n)}$ for all $1 \leq j \leq 2^n$ and all $n \geq 1$.
- 5.5. **Lemma.** (a) For all $m \ge 1$ and $E \in RV_{2m}$, we have $f_E \le f_{\pi(E)} \le f_E + \frac{1}{m}$ pointwise; therefore $||f_E f_{\pi(E)}||_1 \le \frac{1}{m}$.
 - (b) For every coherent sequence $(E(n) \mid n \geq 1)$, with $E(n) \in RV_{2^n}$ for all $n \geq 1$, the sequence $(f_{E(n)} \mid n \geq 1)$ is monotone decreasing pointwise and has $||f_{E(n+1)} f_{E(n)}||_1 \leq 2^{-n}$ for all $n \geq 1$. Therefore $(f_{E(n)} \mid n \geq 1)$ converges to its pointwise infimum, in L_1 -distance, in $L_1(\mathcal{B}, \mu; [0, 1])$, and does so with a rate of convergence that can be taken to be uniform over all sequences in $RV^{\mathcal{M}}$ and all probability spaces (X, \mathcal{B}, μ) .
 - (c) For every $f \in RV = L_1(\mathcal{B}, \mu; [0, 1])$ there is a coherent sequence $(E(n) \mid n \geq 1)$, with $E(n) \in RV_{2^n}$ for all $n \geq 1$ such that $(f_{E(n)} \mid n \geq 1)$ converges to f in L_1 -distance.
- *Proof.* (a) This follows immediately from the definitions.
- (b) From the definition of $f_{E(n)}$ we have $0 \le f_{E(n)}(x) \le 1$ for all $x \in X$. Further, (a) implies that $\chi_X f_{E(n)}$ is pointwise monotone increasing on X and that it converges in L_1 . Since χ_X is integrable, the Monotone Convergence Theorem implies that $(\chi_X f_{E(n)} \mid n \ge 1)$ converges in L_1 to its a.e. pointwise supremum, and therefore $(f_{E(n)} \mid n \ge 1)$ converges in L_1 to its a.e. pointwise infimum. Uniformity of the rate of convergence follows from the uniformity of the estimates in (a).
- (c) For each $n \geq 1$, consider the dyadic intervals I_1, \ldots, I_{2^n} defined by $I_1 := [0, 2^{-n}]$ and for $1 < j \leq 2^n$ by $I_j := ((j-1)2^{-n}, j2^{-n}]$. Let $E(n) \in RV_{2^n}$ be defined by $E(n) := (E_1(n), \ldots, E_{2^n}(n))$ where $E_j(n) := f^{-1}(I_j)$ for $j = 1, \ldots, 2^n$. Note that the sequence $(E(n)|n \geq 1)$ is coherent. Let $C \subseteq B$ be the σ -subalgebra generated by $\{E_j(n) \mid n \geq 1 \text{ and } 1 \leq j \leq 2^n\}$; clearly C is the smallest σ -subalgebra of B such that f is C-measurable.

In particular, $\mathbb{E}(f|\mathcal{C}) = f$. The proof of Lemma 2.7 shows that $(f_{E(n)} \mid n \geq 1)$ converges to f relative to the L_1 -distance. The rate of convergence is as indicated in (b).

5.6. Corollary. The inverse system $(\widehat{RV}_{2^n}^{\mathcal{M}}, \widehat{\rho}_{2^n}^{\mathcal{M}})_{(n\geq 1)}$ equipped with the maps $\widehat{\pi} : \widehat{RV}_{2^{n+1}}^{\mathcal{M}} \to \widehat{RV}_{2^n}^{\mathcal{M}}$ has an inverse limit pseudometric space $(\widehat{RV}^{\mathcal{M}}, \widehat{\rho}^{\mathcal{M}})$ for every model $\mathcal{M} = (\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$ of Pr. Its metric quotient corresponds to $L_1(\mathcal{B}, \mu; [0, 1])$ modulo the L_1 -distance. Moreover, this quotient is a metric imaginary sort for \mathcal{M} , uniformly over all models \mathcal{M} of Pr.

Proof. The first two sentences follow from Lemma 5.5. For the third sentence we apply [5, Lemma 1.5].

5.7. **Remark.** Note that although the spaces $\widehat{RV}_n^{\mathcal{M}}$ are metric spaces (as shown in the proof of Lemma 5.2), the inverse limit distance $\widehat{\rho}^{\mathcal{M}}$ is not a metric. That is, there exist distinct coherent sequences $(e(n))_n, (f(n))_n$ such that $\lim_n \widehat{\rho}_n^{\mathcal{M}}(e(n), f(n)) = 0$. Therefore, to obtain the imaginary sort described in Corollary 5.6, it is necessary to form the metric quotient. However, in contrast to the general case of metric imaginary sorts, the metric space quotient of $(\widehat{RV}^{\mathcal{M}}, \widehat{\rho}^{\mathcal{M}})$ is complete no matter which model \mathcal{M} of Pr is being considered. (In general one needs to take the metric completion of such a quotient for some models.) Indeed, by the Riesz-Fischer Theorem, the quotient of $L_1(\mathcal{B}, \mu)$ modulo the L_1 -distance is complete, no matter which probability space (X, \mathcal{B}, μ) is considered, and the image of $L_1(\mathcal{B}, \mu; [0, 1])$ in that quotient is a norm-closed subset. (See [23, Theorem 6.6, pp. 124–125].)

We next discuss some definable operations on the metric imaginary sort just described; they correspond to the operations taken to be basic (or proved to be definable) in [3] and thus show that our imaginary sort does indeed provide a model of the theory RV. (See [3, Lemma 2.13] for a discussion of the corresponding operations on models of RV.)

Consider any continuous function $\theta \colon [0,1]^m \to [0,1]$. This function induces an operation on $L_1(\mathcal{B}, \mu; [0,1])$ by composition, which we also denote by θ . Namely, we define

$$\theta(f_1,\ldots,f_m)(x) := \theta(f_1(x),\ldots,f_m(x))$$

for all $x \in X$, where $f_1, \ldots, f_m \colon X \to [0, 1]$ are \mathcal{B} -measurable .

We show below that θ induces a definable operation from \widehat{RV}^m to \widehat{RV} . Let us here restrict its domain to $RV_n(\mathcal{B}, \mu)$. We consider the case m = 2 to reduce the complexity of notation. For $E, F \in RV_n$ we have

$$\theta(f_E, f_F) = \sum_{i=1}^n \sum_{j=1}^n \theta(\frac{i}{n}, \frac{j}{n}) \chi_{E_i \cap F_j},$$

which induces a definable function of the events of (E, F), uniformly over all probability spaces.

Moreover, the restrictions of θ to the spaces $\widehat{RV}_n(\mathcal{B}, \mu)$ converge to a definable function on their inverse limit. Again restricting attention to the case m=2, suppose $(E(n))_n$ and $(F(n))_n$ are two coherent families representing elements of the inverse limit, and suppose $f, g \in L_1(\mathcal{B}, \mu; [0, 1])$ are their limits: $f = \lim_n f_{E(n)}$ and $g = \lim_n f_{F(n)}$. Then $\theta(f_{E(n)}, f_{F(n)})$ converges in L_1 -distance to $\theta(f, g)$, and does so at a uniform rate which is determined by the modulus of uniform continuity of θ and the exponential rates of convergence of $\lim_n f_{E(n)}$ and $\lim_n f_{F(n)}$, as given by Lemma 5.5(b). We leave details to the reader.

It is a general fact that every automorphism τ of a metric structure has a unique extension to an automorphism of its meq-expansion. We illustrate this in the present context: given $\mathcal{M} \models Pr$ the probability algebra of (X, \mathcal{B}, μ) as above, consider an automorphism τ of \mathcal{M} . Since $\widehat{RV}_n^{\mathcal{M}}$ is a definable subset of M^{2^n} , we see that τ induces a natural bijection of $\widehat{RV}_n^{\mathcal{M}}$ onto itself, by the coordinatewise action, namely $e = (e_1, \dots, e_{2^n}) \mapsto \tau(e) = (\tau(e_1), \dots, \tau(e_{2^n}))$. Furthermore, if $(e(n) \mid n \geq 1)$ is a coherent sequence in the inverse limit of the spaces $\widehat{RV}_n^{\mathcal{M}}$, convergent to the element $[f]_{\mu}$ of the quotient of $\widehat{RV}^{\mathcal{M}}$ modulo the L_1 -distance, then $(\tau(e(n)) \mid n \geq 1)$ is also coherent; the image of $[f]_{\mu}$ under the desired extension of τ is defined to be the (equivalence class of the) limit of $(\tau(e(n)))_n$.

Another way of looking at this extension process concerns the situation where $a \in \widehat{\mathcal{B}}$ and C is a closed subalgebra of $\widehat{\mathcal{B}}$, and $f \in L_1(\mathcal{B}, \mu; [0, 1])$ is a C-measurable random variable representing $\mathbb{P}(a|C)$. When τ is an automorphism of \mathcal{M} , then $\tau(f)$ as defined above represents $\mathbb{P}(\tau(a)|\tau(C))$, as can be shown by a routine argument based on the details presented in this section. Indeed, this fact is easy to show when C is finite (use Lemma 2.5) and the general case follows using the proof approach for Lemma 2.7 by taking limits.

Finally, we use the operations θ defined above to prove a definability relationship between each random variable $f \in RV(\mathcal{B}, \mu)$ and the smallest σ -subalgebra of \mathcal{B} with respect to which f is measurable, which we will denote by $\sigma(f)$. Note that $\sigma(f)$ is generated as a σ -subalgebra by the measurable sets of the form $f^{-1}(r, 1]$ for $r \in [0, 1]$. This implies that $dcl^{meq}(\sigma(f)) = dcl^{meq}(\{f^{-1}(r, 1] \mid r \in [0, 1]\})$.

5.8. **Lemma.** Let $\mathcal{M} \models Pr$ be the probability algebra of the probability space (X, \mathcal{B}, μ) . For every $f \in RV(\mathcal{B}, \mu)$ we have $\operatorname{dcl}^{\operatorname{meq}}(f) = \operatorname{dcl}^{\operatorname{meq}}(\sigma(f))$.

Proof. First we note that the proof of 5.5(c) shows that $f \in dcl^{meq}(\sigma(f))$.

Thus it remains to show $f^{-1}(r,1] \in \operatorname{dcl}^{\operatorname{meq}}(f)$ for every $r \in [0,1]$. For r=1 this is trivial. Fix $0 \le r < 1$ and for each $n \ge \frac{1}{1-r}$ let $\theta_n \colon [0,1] \to [0,1]$ be the continuous function defined by: $\theta_n(t) = 0$ when $0 \le t \le r$, $\theta_n(t) = n(t-r)$ when $r < t < r + \frac{1}{n}$, and $\theta_n(t) = 1$ for $r + \frac{1}{n} \le t \le 1$. For each $t \in [0,1]$ the sequence $(\theta_n(t))_n$ is monotone increasing in n, and its supremum is the function with values 0 for $t \le r$ and 1 for t > r. Therefore the

sequence $(\theta_n(f))_n$ converges in L_1 to the characteristic function of $f^{-1}(r,1]$, which shows that $f^{-1}(r,1] \in \operatorname{dcl}^{\operatorname{meq}}(f)$, as desired.

6. Atomless probability spaces

By Proposition 4.7(e), the fact that a model of Pr is atomless is expressed by the condition $\theta(1) = 0$, which is equivalent in Pr to the condition

(A)
$$\sup_{x} \inf_{y} |\mu(x \cap y) - \mu(x \cap y^{c})| = 0.$$

We denote by APA the set of axioms Pr together with (A).

6.1. Corollary. Let \mathcal{M} be an L^{pr} -structure. Then \mathcal{M} is a model of APA if and only if \mathcal{M} is isomorphic to the probability algebra of an atomless probability space.

Proof. Immediate from the discussion above.

The main purpose of this section is to give a basic model theoretic analysis of APA. Many of the results correspond to things about atomless probability algebras that were proved by Ben Yaacov in the framework of compact abstract theories [1]. We bring these results into the setting of continuous first order logic and give proofs expressed in familiar language of measure theory and analysis.

In terms of the invariants introduced in 4.15, a model \mathcal{M} of Pr is a model of APA if and only if $\Phi^{\mathcal{M}}$ consists of the constant sequence with every entry equal to 0. Therefore, using results in Section 4 we get the following basic properties of APA:

6.2. Corollary. The theory APA admits quantifier elimination, is separably categorical, and is complete, and the unique separable model of APA is strongly ω -homogeneous. Further, APA is the model companion of Pr.

Proof. Let APA^* denote the theory of all structures \mathcal{M}^* , where \mathcal{M} is a model of APA (i.e., \mathcal{M} is an atomless model of Pr). By Theorem 4.14 we have that APA^* admits quantifier elimination. However, in models of APA^* , all of the extra predicates P_n have the trivial value 0, and thus can be eliminated from any formula. That is, every formula is equivalent in APA^* to a quantifier-free formula in the language of Pr. It follows that APA admits quantifier elimination. Separable categoricity of APA and strong ω -homogeneity of the separable model follow from Corollary 4.17. Completeness of APA follows from separable categoricity and also from quantifier elimination (because every model of APA contains the trivial probability algebra $\{0,1\}$ as a substructure).

Since APA admits quantifier elimination, it is model complete; also, it is an extension of Pr. Therefore, to show that APA is the model companion of Pr it remains only to show that

every model of Pr has an extension that is a model of APA. Let $\mathcal{M} \models Pr$ be the measured algebra of the probability space (X, \mathcal{B}, μ) and let (Y, \mathcal{C}, ν) be any atomless probability space. Then the product measure space $(X \times Y, \mathcal{B} \otimes \mathcal{C}, \mu \otimes \nu)$ is atomless and it can be seen as an extension of (X, \mathcal{B}, μ) by the embedding that takes $B \in \mathcal{B}$ to $B \times Y \in \mathcal{B} \otimes \mathcal{C}$. Therefore the probability algebra of $(X \times Y, \mathcal{B} \otimes \mathcal{C}, \mu \otimes \nu)$ is a model of APA into which \mathcal{M} can be embedded.

6.3. **Remark.** Let \mathcal{M} be the unique separable model of APA, and let $a = (a_1, \ldots, a_n), b = (b_1, \ldots, b_n) \in M^n$. From Corollary 6.2, if $\operatorname{tp}(a) = \operatorname{tp}(b)$, then there exists an automorphism σ of \mathcal{M} such that $\sigma(a_i) = b_i$ for all $i = 1, \ldots, n$. However, this result gives no information about the behavior of σ on the rest of \mathcal{M} . In [9] we proved a stronger form of homogeneity for \mathcal{M} . In order to state that result, we need to bring in another natural metric on M^n , defined by

$$d_P(a,b) := \frac{1}{2} \sum_s d(a^s, b^s).$$

(The version of this distance that is defined on measurable partitions of 1 in a probability space was used in Section 5. See Lemma 5.1. It also appears below in Corollary 6.13.) Here we are using the notation introduced in Notation 4.12 for the partitions of 1 associated to the tuples a and b, and s ranges over $\{-1,+1\}^n$. Note that when $a,b \in M$ are single elements, $d_P(a,b) = d(a,b)$, since $\mu(a^c \triangle b^c) = \mu(a \triangle b)$; this is the reason for the $\frac{1}{2}$ factor in the definition of d_P . Further, when a and b are partitions of 1, then $d_P(a,b)$ agrees with the definition of $d_P(a,b)$ given just before Lemma 5.1, since a^s, b^s will both be 0 unless s contains exactly one occurrence of +1; further, if the unique occurrence of +1 is at place i, then $a^s = a_i$ and $b^s = b_i$.

The homogeneity result from [9] is the following:

Lemma (5.6 in [9]). Let $a, b \in M^n$ be tuples with $\operatorname{tp}(a) = \operatorname{tp}(b)$. Then there is an automorphism σ of \mathcal{M} such that $\sigma(a_i) = b_i$ for all i and, for every finite tuple $c \in M^k$, we have

$$d_P(ac, b\sigma(c)) = d_P(a, b).$$

In particular, for every $c \in M$ we have $d(\sigma(c), c) \leq d_P(a, b)$.

The following lemma appears in [1, Section 2.1], in the framework of compact abstract theories. To make our paper more self-contained and because our setting is different, and in order to make clear the elementary tools from analysis from which these facts can be derived, we give complete proofs.

6.4. **Lemma.** Let $\mathcal{M} = (\mathcal{B}, \mu, d) \models APA$. Let $C \subseteq M = \mathcal{B}$ and $a, b \in M^n$. Recall that $\langle C \rangle$ is the σ -subalgebra of \mathcal{B} generated by C. The following conditions are equivalent:

- (1) $\operatorname{tp}(a/C) = \operatorname{tp}(b/C)$;
- (2) $\mu(a^s \cap c) = \mu(b^s \cap c)$ for all $s = (k_1, \dots, k_n) \in \{-1, +1\}^n$ and all $c \in \langle C \rangle$;
- (3) $\mathbb{P}(a^s|\langle C \rangle) = \mathbb{P}(b^s|\langle C \rangle)$ for all $s = (k_1, \dots, k_n) \in \{-1, +1\}^n$.

Proof. (1) \Leftrightarrow (2): First we deal with the case $C = \emptyset$. Only the right to left direction needs to be proved. By QE for APA (Corollary 6.2), showing $\operatorname{tp}(a_1, \ldots, a_n) = \operatorname{tp}(b_1, \ldots, b_n)$ is equivalent to proving $\mu(t(a_1, \ldots, a_n)) = \mu(t(b_1, \ldots, b_n))$ holds for every boolean term $t(x_1, \ldots, x_n)$. By the discussion in 4.12, this holds whenever $\mu(a^s) = \mu(b^s)$ for all $s \in \{-1, +1\}$.

Now consider arbitrary C. From the discussion before Theorem 2.3, we see $C \subseteq \langle C \rangle \subseteq \operatorname{dcl}(C)$, so each type over $\langle C \rangle$ is determined by its restriction to a type over C. Since $\langle C \rangle$ is closed under boolean combinations, the equivalence of (1) and (2) follows immediately from the first part of this proof.

(2) \Leftrightarrow (3): By Corollary 6.1, there is an atomless probability space (X, \mathcal{A}, ν) whose probability algebra is (\mathcal{B}, μ, d) . Using Lemma 2.2 there exists a σ -subalgebra \mathcal{D} of \mathcal{A} for which $\widehat{\mathcal{D}} = \langle C \rangle$. According to 2.4, for any $a \in \mathcal{B} = \widehat{\mathcal{A}}$ we have $\mathbb{P}(a|\langle C \rangle) = \mathbb{P}(a|\mathcal{D})$. That is, (3) is equivalent to the version of (3) in which we have replaced $\langle C \rangle$ by \mathcal{D} .

The equivalence of (2) and this new version of (3) follows from Theorem 2.3; specifically, from the fact that for any $a = [A]_{\mu} \in \widehat{\mathcal{A}}$, the function $\mathbb{P}(a|\mathcal{D})$ is \mathcal{D} -measurable and is determined, up to equality μ -almost everywhere, by the values of $\int_E \chi_A d\mu = \mu(A \cap E)$ as E ranges over \mathcal{D} .

6.5. **Remark.** The preceding result takes an especially simple form when $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ are partitions of 1 in \mathcal{M} . For example, condition (2) reduces to $\mu(a_i \cap c) = \mu(b_i \cap c)$ for all $i = 1, \ldots, n$ and all $c \in \langle C \rangle$. This means in particular that when $a = (a_1, \ldots, a_n)$ is a partition of 1, the *n*-type of *a* over *C* is determined by the 1-types $\operatorname{tp}(a_i/C)$ for $i = 1, \ldots, n$.

Lemma 6.4 characterizes equality of types. Next we complete the description of the type space $S_n(C)$. First we need a definition:

- 6.6. **Definition.** Let $\mathcal{M} = (\mathcal{B}, \mu, d) \models APA$ and $C \subseteq M = \mathcal{B}$. An additive functional on $\langle C \rangle$ is a finitely additive function $\lambda \colon \langle C \rangle \to [0, 1]$.
- 6.7. **Note.** Let $\mathcal{M} = (\mathcal{B}, \mu, d) \models APA$. Let $C \subseteq M = \mathcal{B}$ and $a \in M^n$. In 6.4(2), each of the functions $\lambda^s \colon \langle C \rangle \to [0, 1]$ defined by $c \mapsto \mu(a^s \cap c)$ is an additive functional on $\langle C \rangle$ with $\lambda^s(1) = \mu(a^s)$. Furthermore, for every $c \in \langle C \rangle$ we have $\sum \{\lambda^s(c) \mid s \in \{-1, +1\}^n\} = \mu(c)$, since $(a^s \cap c \mid s \in \{-1, +1\}^n)$ is a partition of c. Our next result shows that any such family of additive functionals arises from a type in $S_n(\langle C \rangle)$.

6.8. **Lemma.** Let $\mathcal{M} = (\mathcal{B}, \mu, d) \models APA$ and $C \subseteq M = \mathcal{B}$. Assume \mathcal{M} is κ -saturated, where $\kappa > \operatorname{card}(\langle C \rangle)$. Let $(\lambda^s \mid s \in \{-1, +1\}^n)$ be a family of additive functionals on $\langle C \rangle$ such that $\sum \{\lambda^s(c) \mid s \in \{-1, +1\}^n\} = \mu(c)$ for all $c \in \langle C \rangle$. Then there exists $a = (a_1, \ldots, a_n) \in M^n$ such that for every $s \in \{-1, +1\}^n$ and $c \in \langle C \rangle$ we have $\mu(a^s \cap c) = \lambda^s(c)$. Moreover, $\operatorname{tp}(a/C)$ is determined by these conditions.

Proof. Let $x = (x_1, \ldots, x_n)$ be a tuple of distinct variables, and let $\Sigma(x)$ be the set of all conditions of the form $|\mu(x^s \cap c) - \lambda^s(c)| = 0$ as c varies over $\langle C \rangle$ and s varies over $\{-1, +1\}^n$. We must show that Σ is satisfiable in \mathcal{M} , and by saturation it suffices to show that $\Sigma(x)$ is finitely satisfiable. So let F be any finite subset of $\langle C \rangle$ and let f_1, \ldots, f_k be the atoms of $F^\#$. For each $i = 1, \ldots, n$, let $(a_{i,s} \mid s \in \{-1, +1\}^n)$ be a partition of f_i in \mathcal{B} such that $\mu(a_{i,s}) = \lambda^s(f_i)$ for all s. This is possible because $\mu(f_i) = \sum \{\lambda^s(f_i) \mid s \in \{-1, +1\}^n\}$. Note that the family $(a_{i,s})$ is a partition of 1 in \mathcal{B} . Finally, for each i set $a_i = \cup \{a_{i,s} \mid s_i = +1\}$, and set $a = (a_1, \ldots, a_n)$. An easy calculation shows that for all i, s we have $a^s \cap f_i = a_{i,s}$ and hence $\mu(a^s \cap f_i) = \lambda^s(f_i)$. Additivity implies that $\mu(a^s \cap f) = \lambda^s(f)$ for every $f \in F$. This shows that $\Sigma(x)$ is finitely satisfiable in \mathcal{M} and completes the proof (when combined with Lemma 6.4 to provide uniqueness).

6.9. **Remark.** The preceding Lemma takes an especially simple form for 1-types over C. Namely, suppose λ is an additive functional on $\langle C \rangle$ that satisfies $\lambda(c) \leq \mu(c)$ for all $c \in \langle C \rangle$. Define λ' on $\langle C \rangle$ by $\lambda'(c) = \mu(c) - \lambda(c)$. Then λ' is also an additive functional on $\langle C \rangle$, and the pair λ, λ' satisfies the assumptions in Lemma 6.8. Therefore, λ determines a 1-type $p_{\lambda} \in S_1(C)$, and every element of $S_1(C)$ can be described in this way. Specifically, a realizes p_{λ} if and only if $\mu(a \cap c) = \lambda(c)$ for all $c \in \langle C \rangle$ (since this implies $\mu(a^c \cap c) = \lambda'(c)$). This observation is especially useful when considering n-types of partititions of 1, as discussed in Remark 6.5. We also use this description of 1-types in discussing the model theoretic content of Maharam's Lemma in Section 7 (Lemma 7.16).

An equivalent approach to 1-types over C in terms of $\langle C \rangle$ -measurable, [0,1]-valued functions (i.e., random variables), corresponding to clause (3) in Lemma 6.4, is the following: let f be any $\langle C \rangle$ -measurable, [0,1]-valued function. For $\mathcal{M} \models APA$ and $C \subseteq M$, the condition $\mathbb{P}(a|\langle C \rangle) = f$ on $a \in M$ is type-definable over C; indeed, it precisely determines $\operatorname{tp}(a/C)$. To see this, consider λ_f defined for $c \in \langle C \rangle$ by $\lambda_f(c) := \int_c f \, d\mu$. Then λ_f is an additive functional on $\langle C \rangle$ that satisfies $\lambda(c) \leq \mu(c)$ for all $c \in \langle C \rangle$. Therefore, as discussed in the preceding paragraph, λ_f exactly determines a 1-type in $S_1(C)$. Moreover, the condition $\mu(a \cap c) = \lambda_f(c)$ for all $c \in \langle C \rangle$ is equivalent to $\mathbb{P}(a|\langle C \rangle) = f$.

6.10. **Lemma.** Let $\mathcal{M} = (\mathcal{B}, \mu, d) \models APA$ and $C \subseteq M$. Then $dcl(C) = acl(C) = \langle C \rangle$.

Proof. Recall that $\langle C \rangle$ is the σ -subalgebra of \mathcal{B} generated by C. As noted at the beginning of the previous proof, $C \subseteq \langle C \rangle \subseteq \operatorname{dcl}(C)$. Therefore, to complete the proof it suffices to prove $\operatorname{acl}(C) \subseteq \langle C \rangle$.

Now let $a \in M \setminus \langle C \rangle$. To show $a \notin \operatorname{acl}(C)$, by [7, Exercise 10.8] it suffices to prove that for some $\mathcal{N} \succeq \mathcal{M}$, there is a realization of $\operatorname{tp}(a/C)$ in \mathcal{N} that is not in \mathcal{M} . Let \mathcal{M} be the probability algebra of the probability space (X, \mathcal{A}, ν) . Consider the standard probability space $([0, 1], \mathcal{L}, m)$ of Lebesgue measure and form the product space $(X \times [0, 1], \mathcal{A} \otimes \mathcal{L}, \nu \otimes m)$; let \mathcal{N} be the probability algebra of this product space. There is a canonical embedding $J: \mathcal{A} \to \mathcal{A} \otimes \mathcal{L}$ defined by $J(A) = A \times [0, 1]$ for $A \in \mathcal{A}$; this map gives rise to an embedding \widehat{J} of \mathcal{M} into \mathcal{N} . Since APA admits quantifier elimination, \widehat{J} is an elementary embedding. As in the previous proof, there is a σ -subalgebra \mathcal{D} of \mathcal{A} for which $\widehat{\mathcal{D}} = \langle C \rangle$. Since $\mathbb{P}(a|\mathcal{D})$ is \mathcal{D} -measurable, the set $A' = \{(x,s) \in X \times [0,1] \mid s \leq \mathbb{P}(a|\mathcal{D})(x)\}$ is $\mathcal{D} \otimes \mathcal{L}$ -measurable. We let $a' = [A']_{\nu \otimes m} \in \widehat{\mathcal{D}} \otimes \mathcal{L} \subseteq \widehat{\mathcal{A}} \otimes \mathcal{L} = N$. We complete the proof by showing that a' is not in the image of \mathcal{M} under \widehat{J} and that a' is a realization of $\operatorname{tp}(a/C)$ in \mathcal{N} .

For the first of these statements, we note that $0 < \mathbb{P}(a|\mathcal{D}) < 1$ holds on a set of positive measure. Otherwise $\mathbb{P}(a|\mathcal{D}) = \chi_B$ for some $B \in \mathcal{D}$; this would imply $a = [B]_{\nu} \in \langle C \rangle$, which would contradict our assumptions. It follows that A' is not of the form $A \times [0,1]$ where $A \in \mathcal{A}$, and thus a' is not in the image of \mathcal{M} under \widehat{J} .

Finally, let $\mathcal{D}' = \{B \times [0,1] \mid B \in \mathcal{D}\} = J(\mathcal{D})$, so \mathcal{D}' is a σ -algebra and $\widehat{\mathcal{D}'} = \widehat{J}(\langle C \rangle) = \langle \widehat{J}(C) \rangle$. Fubini's Theorem shows that $\mathbb{P}(a'|\mathcal{D}') = \mathbb{P}(\widehat{J}(a)|\mathcal{D}')$, which implies by Lemma 6.4 that $\operatorname{tp}_{\mathcal{N}}(a'/\widehat{J}(C)) = \operatorname{tp}_{\mathcal{M}}(a/C)$. (Here we mean, of course, that the parameters in $\widehat{J}(C)$ are identified with those in C via the bijection \widehat{J} .)

In several results in the rest of this section it is convenient to work in a κ -universal domain for APA, where κ is uncountable. For the rest of the section we denote such a model of APA as \mathcal{U} . Recall that a subset C of U is called small if $card(C) < \kappa$. In this situation, every type in $S_n(C)$ is realized in \mathcal{U} . Furthermore, \mathcal{U} is strongly κ -homogeneous; *i.e.*, every elementary map between small subsets of U extends to an automorphism of \mathcal{U} . (In applications, \mathcal{U} and the size of κ may need to be changed in order to insure that specific parameter sets are small.)

Recall that the metric d on \mathcal{U} yields an *induced metric* on each space of types (see [7, Section 8]), as follows: when $C \subseteq U$ is small and p,q are n-types over C, the distance between p and q is defined by

$$d(p,q) = \inf\{\max_{1 \le i \le n} d(a_i, b_i) : (a_1, \dots, a_n) \models p, (b_1, \dots, b_n) \models q\}.$$

The next result provides an explicit formula for the induced metric on types of partitions of 1 in atomless probability algebras. (When the parameter set C is empty, this formula occurs as (6.2) in the proof of [25, Lemma 6.3].)

6.11. **Theorem.** Let $C \subseteq U$ be small and let $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ be partitions of 1 in \mathcal{U} . Then

$$d(\operatorname{tp}(a/C),\operatorname{tp}(b/C)) = \max_{1 < i < n} \|\mathbb{P}(a_i|\langle C \rangle) - \mathbb{P}(b_i|\langle C \rangle)\|_1$$

where $\| \|_1$ is the L_1 -norm.

Moreover, there exists $b' = (b'_1, \ldots, b'_n)$, a partition of 1 in U, such that $\operatorname{tp}(b'/C) = \operatorname{tp}(b/C)$ and for all $i = 1, \ldots, n$

$$d(a_i, b_i') = \|\mathbb{P}(a_i | \langle C \rangle) - \mathbb{P}(b_i | \langle C \rangle)\|_1.$$

Proof. Replacing C by $C^{\#}$ (which is still small) we may assume throughout this proof that C is a boolean subalgebra of \mathcal{U} . Since $C \subseteq C^{\#} \subseteq \operatorname{dcl}(C)$, this does not change the types being considered nor the distance between them. Furthermore, it is obvious that $\langle C^{\#} \rangle = \langle C \rangle$, so the right side of the equality to be proved is also not changed by this move.

We begin the proof by noting that

$$\|\mathbb{P}(u|\langle C\rangle) - \mathbb{P}(v|\langle C\rangle)\|_1 \le \mu(u\triangle v)$$

for any $u, v \in U$. Indeed, linearity of the conditional expectation yields

$$\|\mathbb{P}(u|\langle C\rangle) - \mathbb{P}(v|\langle C\rangle)\|_1 = \|\mathbb{P}(u \setminus v|\langle C\rangle) - \mathbb{P}(v \setminus u|\langle C\rangle)\|_1 \le \mu(u \triangle v)$$

where the last step uses the triangle inequality for the L_1 -norm and the fact that $\|\mathbb{P}(w|C)\|_1 = \mu(w)$ for any $w \in U$. By Lemma 6.4, $\|\mathbb{P}(u|\langle C \rangle) - \mathbb{P}(v|\langle C \rangle)\|_1$ only depends on $\operatorname{tp}(u/C)$ and $\operatorname{tp}(v/C)$. Fixing $i \in \{1, \ldots, n\}$ and letting u, v range over realizations of $\operatorname{tp}(a_i/C)$, $\operatorname{tp}(b_i/C)$ respectively, and taking the infimums, we obtain

$$\|\mathbb{P}(a_i|\langle C\rangle) - \mathbb{P}(b_i|\langle C\rangle)\|_1 \le d(\operatorname{tp}(a_i/C), \operatorname{tp}(b_i/C)).$$

Taking the maximum over i yields

$$\max_{1 \le i \le n} \| \mathbb{P}(a_i | \langle C \rangle) - \mathbb{P}(b_i | \langle C \rangle) \|_1 \le d(\operatorname{tp}(a/C), \operatorname{tp}(b/C)).$$

Therefore it remains to show

$$d(\operatorname{tp}(a/C), \operatorname{tp}(b/C)) \le \max_{1 \le i \le n} \|\mathbb{P}(a_i|\langle C \rangle) - \mathbb{P}(b_i|\langle C \rangle)\|_1,$$

given that $a, b \in U^n$ are partitions of 1. We do this in the remainder of the proof.

We first prove this inequality when $C = \emptyset$, noting that the right side of the last inequality is equal to $\max_{1 \le i \le n} |\mu(a_i) - \mu(b_i)|$ in this situation.

Let $I = \{i \mid \mu(a_i) \geq \mu(b_i)\}$ and $J = \{i \mid \mu(a_i) < \mu(b_i)\}$. Note that $I \neq \emptyset$; also, we may assume $J \neq \emptyset$, since otherwise $\operatorname{tp}(a) = \operatorname{tp}(b)$ and so the inequality to be proved is trivial. Since \mathcal{U} is atomless, we may choose $b_i' \leq a_i$ in U satisfying $\mu(b_i') = \mu(b_i)$, for each $i \in I$. For each such i, let $u_i = a_i \setminus b_i'$, and set $u = \bigcup \{u_i \mid i \in I\}$. Note that $\mu(u) = \sum \{\mu(a_i) - \mu(b_i) \mid i \in I\}$. Because (a_1, \ldots, a_n) and (b_1, \ldots, b_n) are partitions of 1 in \mathcal{U} , it follows that $\mu(u) = \sum \{\mu(b_j) - \mu(a_j) \mid j \in J\}$. Hence we may partition u into $\{u_j \mid j \in J\} \subseteq U$ such that $\mu(u_j) = \mu(b_j) - \mu(a_j)$ for all $j \in J$. Finally, for $j \in J$ we set $b_j' = a_j \cup u_j$ and note that (b_1', \ldots, b_n') is a measurable partition of 1 in \mathcal{U} satisfying $\mu(b_j') = \mu(b_j)$ for all $j \in J$; in other words, (b_1', \ldots, b_n') realizes the same type as (b_1, \ldots, b_n) . Moreover, for all $j \in J$ we have

$$\mu(a_j \triangle b_j') = \mu(u_j) = |\mu(a_j) - \mu(b_j')| = |\mu(a_j) - \mu(b_j)|$$

which justifies the desired inequality.

Now assume that C is a finite boolean subalgebra of \mathcal{U} and let the atoms of C be c_1, \ldots, c_p . For each $i \leq n$ and $j \leq p$, let $a_{ij} = a_i \cap c_j$ and $b_{ij} = b_i \cap c_j$. We argue as in the previous paragraph within each c_j . This yields b'_{ij} for $i \leq n, j \leq p$ with the following properties: (a) $\mu(b'_{ij}) = \mu(b_{ij})$ for all i, j; (b) for each $j \leq p$, the tuple $(b'_{1j}, \ldots, b'_{nj})$ is a partition of c_j ; and (c) $\mu(a_{ij} \triangle b'_{ij}) = |\mu(a_{ij}) - \mu(b_{ij})|$ for all i, j. For each $i \leq n$, let $b'_i = \bigcup_{j \leq m} b'_{ij}$. Then $\operatorname{tp}(b'_1, \ldots, b'_n/C) = \operatorname{tp}(b_1, \ldots, b_n/C)$ and

$$\mu(a_i \triangle b_i') = \sum_{j \le m} |\mu(a_{ij}) - \mu(b_{ij})| = \sum_{j \le m} |\mu(a_i \cap c_j) - \mu(b_i \cap c_j)| = \|\mathbb{P}(a_i | \langle C \rangle) - \mathbb{P}(b_i | \langle C \rangle)\|_1.$$

Finally, consider a general algebra C. For each $k \geq 1$, use Lemma 2.7 applied to C and $a_1, \ldots, a_n, b_1, \ldots, b_n$ to obtain a finite subalgebra $C_k \subseteq C$ such that for all closed subalgebras $D \subseteq C$ that contain C_k we have $\|\mathbb{P}(u|C) - \mathbb{P}(u|D)\|_1 \leq 1/k$ for all $u = a_i$ and $u = b_i$ with $1 \leq i \leq n$. We may assume $C_k \subseteq C_{k+1}$ for all $k \geq 1$.

Further, we may use properties of type spaces to enlarge each C_k to ensure additionally for $k \geq 1$ that

$$|d(\operatorname{tp}(a/C),\operatorname{tp}(b/C)) - d(\operatorname{tp}(a/C_k),\operatorname{tp}(b/C_k))| \le 1/k.$$

Indeed, note that if $E \subseteq D \subseteq C$, then $d(\operatorname{tp}(a/E), \operatorname{tp}(b/E)) \leq d(\operatorname{tp}(a/D), \operatorname{tp}(b/D)) \leq d(\operatorname{tp}(a/C), \operatorname{tp}(b/C))$. Moreover, $d(\operatorname{tp}(a/C), \operatorname{tp}(b/C))$ is the supremum of $d(\operatorname{tp}(a/D), \operatorname{tp}(b/D))$ as D varies over finite subsets of C. (Otherwise there would exist $r < d(\operatorname{tp}(a/C), \operatorname{tp}(b/C))$ such that $d(\operatorname{tp}(a/D), \operatorname{tp}(b/D)) \leq r$ for all finite $D \subseteq C$. Thus the following set of conditions would be finitely satisfiable in \mathcal{U} :

$$\Sigma := \{ \varphi(x) = 0 \mid \varphi(x) \in \operatorname{tp}(a/C) \} \cup \{ \psi(y) = 0 \mid \psi(x) \in \operatorname{tp}(b/C) \} \cup \{ \max_i \ d(x_i, y_i) \leq r \}.$$

Since C is a small set, we may choose x = a' and y = b' that realize Σ in \mathcal{U} . But then we would have $\operatorname{tp}(a'/C) = \operatorname{tp}(a/C), \operatorname{tp}(b'/C) = \operatorname{tp}(b/C),$ and $d(a',b') \leq r < d(\operatorname{tp}(a/C),\operatorname{tp}(b/C)),$ which is impossible.)

Putting these two arguments together, we have an increasing family $(C_k \mid k \geq 1)$ of finite subalgebras of C such that for all $k \geq 1$

$$\|\mathbb{P}(u|C) - \mathbb{P}(u|C_k)\|_1 \le 1/k$$

for all $u = a_i$ and $u = b_i$ with $1 \le i \le n$, and

$$|d(\operatorname{tp}(a/C),\operatorname{tp}(b/C)) - d(\operatorname{tp}(a/C_k),\operatorname{tp}(b/C_k))| \le 1/k.$$

From what is proved earlier for n-types over finite algebras, for all $k \geq 1$ we have

$$d(\operatorname{tp}(a/C_k),\operatorname{tp}(b/C_k) = \max_{1 \le i \le n} \|\mathbb{P}(a_i|\langle C_k \rangle) - \mathbb{P}(b_i|\langle C_k \rangle)\|_1.$$

Taking limits as $k \to \infty$ yields

$$d(\operatorname{tp}(a/C),\operatorname{tp}(b/C)) = \max_{1 \le i \le n} \|\mathbb{P}(a_i|\langle C \rangle) - \mathbb{P}(b_i|\langle C \rangle)\|_1,$$

completing the proof.

6.12. Corollary. Let $C \subseteq U$ be small and let a, b be elements of U. Then

$$d(\operatorname{tp}(a/C),\operatorname{tp}(b/C)) = \|\mathbb{P}(a|\langle C\rangle) - \mathbb{P}(b|\langle C\rangle)\|_{1}.$$

Proof. Apply the preceding Lemma to (a, a^c) and (b, b^c) , and use the fact that each side of the equation to be proved is unchanged if we replace a, b by a^c, b^c

The definition of the metric d_P on U^n that is given in Remark 6.3 says for $a, b \in U^n$

$$d_P(a,b) := \frac{1}{2} \sum_s d(a^s, b^s),$$

where s ranges over $\{-1, +1\}$. Following the established pattern, we can define d_P on $S_n(C)$ by

$$d_P(p,q) := \inf\{d_P(a,b) \mid a \models p \text{ and } b \models q\}.$$

From Theorem 6.11 we get immediately

6.13. Corollary. Let $C \subseteq U$ be small and let $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$ be in U^n . Then

$$d_P(\operatorname{tp}(a/C), \operatorname{tp}(b/C)) = \frac{1}{2}\inf\{\sum_s \|\mathbb{P}(a^s|\langle C\rangle) - \mathbb{P}(b^s|\langle C\rangle)\|_1 \mid a \models p \text{ and } b \models q\}.$$

where $\|\cdot\|_1$ is the L_1 -norm.

Proof. This follows from the "Moreover" statement in Theorem 6.11. \square

Moving beyond types of partitions of 1, we now discuss the induced metric on the full type space $S_n(C)$ for APA. For $r \geq 1$, let $S_r^*(C)$ denote the space of r-types for APA that are realized by partitions (a_1, \ldots, a_r) of 1 in the κ -universal domain \mathcal{U} for APA, where C is a small subset of U. Theorem 6.11 gives an explicit formula for the induced metric on $S_r^*(C)$. Since $S_r^*(C)$ is a proper, metrically closed subset of the full space of r-types $S_r(C)$, this does not immediately characterize the metric on all of $S_r(C)$. However, by looking at types for APA in the right way, and taking $r = 2^n$, we can use this lemma to characterize the induced metric on $S_n(C)$ up to equivalence of metrics, which is enough for most purposes.

To accomplish this, consider the map $\Pi_n \colon S_n(C) \to S_{2^n}^*(C)$ on types that is induced by mapping the type of an arbitrary n-tuple (a_1, \ldots, a_n) to the type of its associated partition $(a^s \mid s \in \{-1, +1\})$ (as discussed in 4.12). Since APA admits quantifier elimination, Π_n is a bijection from $S_n(C)$ onto $S_{2^n}^*(C)$. The discussion in Remark 4.13 shows that Π_n is also a homeomorphism for the (logic) topologies. In what follows, we often drop the subscript n when doing so will not cause confusion.

6.14. **Lemma.** Let $C \subseteq U$ be small and let $p, q \in S_n(C)$. Then

$$(2^{-n+1}) \cdot d_n(p,q) \le d_{2^n}(\Pi_n(p), \Pi_n(q)) \le n \cdot d_n(p,q)$$

where d_n, d_{2^n} denote the induced metrics on the type spaces $S_n(C), S_{2^n}(C)$ respectively (usually denoted simply by d, but here given a subscript to indicate the type space on which the metric is defined).

Proof. This uses an easy calculation based on the description of the bijection between n-tuples (a_1, \ldots, a_n) and partitions $(a^s \mid s \in \{-1, +1\})$ that is given in 4.12.

Thus Π_n is a bi-Lipschitz homeomorphism from $S_n(C)$ onto $S_{2^n}^*(C)$ with respect to the two induced metrics. Since an explicit formula for the induced metric on $S_{2^n}^*(C)$ is given by Theorem 6.11, this gives us considerable information about the induced metric topology on all of $S_n(C)$.

Note that this observation strengthens Lemma 6.4.

We next prove that the theory APA is stable; indeed, we simply count types, and show that APA is ω -stable:

6.15. **Proposition** (Prop. 4.4, [1]). The theory APA is ω -stable.

Proof. We may take \mathcal{U} to be the probability algebra of an atomless probability space (X, \mathcal{A}, ν) . Let $C \subseteq U$ be countable. For each $a \in C$ chose a set $A_a \in \mathcal{A}$ satisfying $a = [A_a]_{\nu}$ and let \mathcal{C} be the boolean subalgebra of \mathcal{A} generated by $\{A_a \mid a \in C\}$. Then \mathcal{C} is countable and $\widehat{\mathcal{C}} = C^{\#}$.

Let $\mathcal{S}(C)$ be the set of \mathcal{C} -measurable simple functions with coefficients in $\mathbb{Q} \cap [0,1]$, and let

$$\mathcal{F} = \{ \operatorname{tp}(a/C) \mid \mathbb{P}(a|\mathcal{C}) \in \mathcal{S}(C) \}.$$

Then \mathcal{F} is a countable set of types. By Lemmas 2.7 and 6.12, \mathcal{F} is a metrically dense subset of the space of 1-types over C.

7. Maharam's Theorem

Maharam's Theorem is a structure theorem for probability algebras. It says that a model $\mathcal{M} = (\mathcal{B}, \mu, d)$ of Pr is determined up to isomorphism by the information $\Phi^{\mathcal{M}}$ given in Section 4 about the atomic part of \mathcal{B} together with a countable set $\mathcal{K}^{\mathcal{M}}$ of infinite cardinal numbers and a function $\Psi \colon \mathcal{K}^{\mathcal{M}} \to (0,1]$ whose sum equals the μ -measure of the atomless part of \mathcal{B} . Note that 1 is atomic in \mathcal{B} if and only if $\mathcal{K}^{\mathcal{M}} = \emptyset$. In general, we know that the atomic part of \mathcal{M} is determined up to isomorphism by $\Phi^{\mathcal{M}}$, as is the measure of the atomless part of \mathcal{B} . (See Corollary 4.19.) Therefore we may focus our attention on the atomless part of \mathcal{M} . When it is nonzero, it can be considered as a model of APA by rescaling the measure and the metric. That is, to prove Maharam's Theorem, we may focus on models of APA.

In this section we give a full discussion of Maharam's Theorem for models of APA, to make clear the ways in which its proof resonates with ideas from model theory.

- 7.1. **Definition.** Let $\mathcal{M} = (\mathcal{B}, \mu, d) \models APA$ and $0 \neq b \in \mathcal{B}$. Define $\mathcal{B} \upharpoonright b$ to be the ideal of all $a \leq b$ in \mathcal{B} .
- 7.2. **Note.** Since we require $b \neq 0$ in the preceding definition, we may regard $\mathcal{B} \upharpoonright b$ as a boolean algebra; the interpretations of \cap and \cup as well as of 0 are inherited from \mathcal{B} , while 1 is interpreted as b and the complement operation is taken to be $a \mapsto a^c \cap b$. Note that with this understanding of the structure of $\mathcal{B} \upharpoonright b$, the map $a \mapsto a \cap b$ is a boolean morphism from \mathcal{B} onto $\mathcal{B} \upharpoonright b$. We equip $\mathcal{B} \upharpoonright b$ with the measure and distance obtained from \mathcal{M} by restriction to $\mathcal{B} \upharpoonright b$; for convenience we continue to denote these restrictions by μ and d.

It is clear that $(\mathcal{B} \upharpoonright b, \mu, d)$ is a measured algebra, and that it becomes a model of APA if we rescale μ and d appropriately (namely, multiply by $1/\mu(b)$). We systematically use this point of view below.

7.3. **Notation.** Unless otherwise specified, in the rest of this section we take $\mathcal{M} = (\mathcal{B}, \mu, d)$ to be a model of APA. When we refer to the *density* of a subset of \mathcal{B} , we mean the metric density.

A key quantity for the arguments behind Maharam's Theorem is the density of $\mathcal{B} \upharpoonright b$; for brevity we also refer to this as the *density of b*. When b is atomless, this density is an infinite cardinal number.

7.4. **Definition.** For $b \in \mathcal{B}$, we say $\mathcal{B} \upharpoonright b$ is homogeneous and (alternatively) b is homogeneous if $b \neq 0$ and $\mathcal{B} \upharpoonright a$ has the same density as $\mathcal{B} \upharpoonright b$ for every $0 \neq a \leq b$.

We are now in position to define the *Maharam invariants* $(\mathcal{K}^{\mathcal{M}}, \Psi^{\mathcal{M}})$ for a model $\mathcal{M} = (\mathcal{B}, \mu, d)$ of APA.

7.5. **Definition.** Define $\mathcal{K}^{\mathcal{M}}$ to be the set of all infinite cardinal numbers κ for which there exists $b \in \mathcal{B}$ such that b is homogeneous and the density of $\mathcal{B} \upharpoonright b$ is κ . For each $\kappa \in \mathcal{K}^{\mathcal{M}}$ define

$$\Psi^{\mathcal{M}}(\kappa) := \sup\{\mu(b) \mid b \text{ is homogeneous and the density of } \mathcal{B} \upharpoonright b = \kappa\}.$$

We call $b \in \mathcal{B}$ maximal homogeneous if b is homogeneous and $\mu(b) = \Psi^{\mathcal{M}}(\kappa)$, where $\kappa =$ density of $\mathcal{B} \upharpoonright b$. We call \mathcal{B} homogeneous if 1 is homogeneous in \mathcal{B} .

We say \mathcal{M} realizes its Maharam invariants if there exists a family $(b_{\kappa} \mid \kappa \in \mathcal{K}^{\mathcal{M}})$ of pairwise disjoint maximal homogeneous elements of \mathcal{B} such that b_{κ} has density κ for every $\kappa \in \mathcal{K}^{\mathcal{M}}$ and $\sum \{\mu(b_{\kappa}) \mid \kappa \in \mathcal{K}^{\mathcal{M}}\}$ exists and equals 1.

We show below that every model \mathcal{M} of APA realizes its Maharam invariants. In particular, this means that $\mathcal{K}^{\mathcal{M}}$ is nonempty and countable.

- 7.6. **Note.** If $\mathcal{M} \models APA$ realizes its Maharam invariants, then the density of \mathcal{M} is the supremum of $\mathcal{K}^{\mathcal{M}}$ (taken in the cardinal numbers).
- 7.7. **Example.** Obviously the unique separable model \mathcal{M} is homogeneous of density \aleph_0 .

Let (X, \mathcal{B}, μ) be any countably generated, atomless probability space, and let κ be any uncountable cardinal number. Let \mathcal{A} be the probability algebra of the product space X^{κ} with the product probability measure obtained by taking μ as the measure on each factor. Then \mathcal{A} is homogeneous and has density κ .

Proof. Let \mathcal{S} be a countable dense subset of \mathcal{B} . For each $\alpha < \kappa$ let π_{α} be the coordinate projection from X^{κ} onto X. The σ -algebra of product-measurable subsets of X^{κ} is generated by the sets of the form $\pi_{\alpha}^{-1}(Q)$, where $\alpha < \kappa$ and $Q \in \mathcal{S}$. Therefore \mathcal{A} has density at most κ . Also, if $Q \in \mathcal{B}$ has $\mu(Q) = r \in (0,1)$ and α, β are distinct, then $d(\pi_{\alpha}^{-1}(Q), \pi_{\beta}^{-1}(Q)) = 2r(1-r) > 0$, so \mathcal{A} has density at least κ .

If V is any product-measurable subset of X^{κ} , then V only depends on countably many ordinals $\alpha < \kappa$, in the sense that there is a countable set S of such ordinals such that for any $u, v \colon \kappa \to X$, if $u \in V$ and $u(\alpha) = v(\alpha)$ for all $\alpha \in S$, then also $v \in V$. When V, S satisfy this condition, we say V depends only on the coordinates in S. (Note that the collection of product measurable $V \subseteq X^{\kappa}$ that only depend on countably many $\alpha < \kappa$ is a σ -algebra, and it contains all sets of the form $\pi_{\alpha}^{-1}(Q)$, where $\alpha < \kappa$ and $Q \in \mathcal{S}$.)

A variant of the argument in the first paragraph shows that the restriction of \mathcal{A} to the event determined by any product-measurable set V also has density equal to κ . (Just work on the coordinates in $\kappa \setminus S$, where S is countable and V depends only on the coordinates in S.) Therefore \mathcal{A} is homogeneous of density κ .

7.8. **Remark.** It is now clear that for every nonempty countable set \mathcal{K} of infinite cardinal numbers and every function $\Psi \colon \mathcal{K} \to (0,1]$ whose sum equals 1, we can construct an atomless probability space (X, \mathcal{B}, μ) whose probability algebra \mathcal{M} realizes its Maharam invariants and such that $\mathcal{K}^{\mathcal{M}} = \mathcal{K}$ and $\Psi^{\mathcal{M}} = \Psi$. For each $\kappa \in \mathcal{K}$, let $(X_{\kappa}, \mathcal{B}_{\kappa}, \mu_{\kappa})$ be a probability space whose probability algebra is homogeneous of density κ ; take the sets X_{κ} to be pairwise disjoint. For each $\kappa \in \mathcal{K}$, let μ'_{κ} be $\Psi(\kappa)\mu$. Then take X to be the union of $(X_{\kappa} \mid \kappa \in \mathcal{K})$ and let \mathcal{B} be the σ -algebra of subsets of X generated by $\bigcup \{\mathcal{B}_{\kappa} \mid \kappa \in \mathcal{K}\}$. Note that each $Q \in \mathcal{B}$ is equal to $\bigcup \{Q \cap X_{\kappa} \mid \kappa \in \mathcal{K}\}$, and set $\mu(Q) := \sum \{\mu'_{\kappa}(Q \cap X_{\kappa}) \mid \kappa \in \mathcal{K}\}$. Then it is clear that (X, \mathcal{B}, μ) is an atomless probability space and that its probability algebra \mathcal{M} satisfies $(\mathcal{K}^{\mathcal{M}}, \Psi^{\mathcal{M}}) = (\mathcal{K}, \mathcal{M})$.

Next we state a lemma giving properties of homogeneous elements.

- 7.9. Lemma. Let $\mathcal{M} = (\mathcal{B}, \mu, d) \models APA$.
 - (a) If b_1, b_2 are homogeneous elements of \mathcal{B} , and if b_1, b_2 have different densities, then $b_1 \cap b_2 = 0$.
 - (b) If b_n is a homogeneous element of \mathcal{B} for $n \geq 1$, and the density of b_n is κ for all n, then $b = \bigcup \{b_n \mid n \geq 1\}$ is also homogeneous in \mathcal{B} and b has density κ .
 - (c) If there exists a homogeneous element $b \in \mathcal{B}$ of density κ , then there exists a maximal homogeneous element b' such that b' also has density κ .
 - (d) If b' is a maximal homogeneous element of density κ , then every homogeneous element b of density κ satisfies b < b'.

Proof. Left as exercises for the reader.

7.10. **Proposition.** Every model $\mathcal{M} = (\mathcal{B}, \mu, d)$ of APA realizes its system of Maharam invariants.

Proof. We refer to the items in Lemma 7.9 by their letters. Let $\mathcal{K}^{\mathcal{M}}$ be defined as in Definition 7.5. Note that $\mathcal{K}^{\mathcal{M}}$ is nonempty, since taking $0 \neq b \in \mathcal{B}$ such that b has the least possible density implies that $\mathcal{B} \upharpoonright b$ is homogeneous. For each $\kappa \in \mathcal{K}^{\mathcal{M}}$, let b_{κ} be a maximal homogeneous element of \mathcal{B} that has density κ , which exists by (c). By (a) the elements $(b_{\kappa} \mid \kappa \in \mathcal{K}^{\mathcal{M}})$ are pairwise disjoint in \mathcal{B} and by (d) we have $\mu(b_{\kappa}) = \Psi^{\mathcal{M}}(\kappa)$ for every κ . Note that this implies that $\mathcal{K}^{\mathcal{M}}$ is countable.

It remains to show that $\sum \{\mu(b_{\kappa}) \mid \kappa \in \mathcal{K}^{\mathcal{M}}\} = 1$. If not, let b be the complement in \mathcal{B} of $\cup \{b_{\kappa} \mid \kappa \in \mathcal{K}^{\mathcal{M}}\}$, so b > 0. Let b' be a nonzero element of $\mathcal{B} \upharpoonright b$ of least possible density, so $\mathcal{B} \upharpoonright b'$ is homogeneous. If κ is the density of b', then $\kappa \in \mathcal{K}^{\mathcal{M}}$ by definition, and we have that $b' \cap b_{\kappa} = 0$. This contradicts the maximality of b_{κ} .

7.11. **Note.** It remains to show that a model \mathcal{M} of APA is determined up to isomorphism by its Maharam invariants. Evidently it suffices to prove the special case that when \mathcal{M}, \mathcal{N} are homogeneous models and have the same density, then $\mathcal{M} \cong \mathcal{N}$. Indeed, if \mathcal{M} is any model of APA and the family $(b_{\kappa} \mid \kappa \in \mathcal{K}^{\mathcal{M}})$ witnesses that \mathcal{M} realizes its Maharam invariants (as in the proof of Proposition 7.10), then the isomorphism type of each $\mathcal{B} \upharpoonright b_{\kappa}$ (as a measured algebra) would be determined by κ and $\mu(b_{\kappa}) = \Psi^{\mathcal{M}}(\kappa)$. The isomorphism type of \mathcal{M} is easily reconstructed from this data, since \mathcal{K} is countable, the elements $(b_{\kappa} \mid \kappa \in \mathcal{K}^{\mathcal{M}})$ are pairwise disjoint, and $\sum \{\mu(b_{\kappa}) \mid \kappa \in \mathcal{K}^{\mathcal{M}}\} = 1$.

A similar discussion applies to arbitrary models $\mathcal{M} = (\mathcal{B}, \mu, d)$ of Pr. In this case the necessary decomposition of \mathcal{B} consists of a family $(b_i \mid i \in I)$ of elements of \mathcal{B} and a family $(\kappa_i \mid i \in I)$ of cardinal numbers satisfying the following conditions: (i) the elements b_i are pairwise disjoint and nonzero; (ii) $\sum \{\mu(b_i) \mid i \in I\} = 1$; (iii) if κ_i is finite, it equals 1 and b_i is an atom in \mathcal{B} ; (iv) if κ_i is infinite, then b_i is a maximal homogeneous component of the atomless part of \mathcal{B} of density κ_i ; and (v) if κ_i, κ_j are infinite with $i \neq j$, they are distinct. As we show now, the additional information needed to determine \mathcal{M} up to isomorphism is the family $(\mu(b_i) \mid i \in I)$ of real numbers, which all come from (0, 1] and whose sum is 1.

What remains to be proved is that every homogeneous model of APA is determined up to isomorphism by its density. It is in this proof where model theoretic ideas come into play, as we explain next. Indeed, the homogeneous models of APA are the same as the saturated models (i.e., the models that have density κ and are κ -saturated, for some κ). To make this connection precise requires the introduction of the following notion.

- 7.12. **Definition.** Let (\mathcal{B}, μ, d) be a probability algebra and let \mathcal{A} be a σ -subalgebra of \mathcal{B} . A non-zero element $b \in \mathcal{B}$ is called an *atom relative to* \mathcal{A} if for all $b' \leq b$ in \mathcal{B} there is $a \in \mathcal{A}$ such that $b' = a \cap b$. We say that \mathcal{B} is *atomless over* \mathcal{A} if no nonzero element $b \in \mathcal{B}$ is an atom relative to \mathcal{A} .
- 7.13. **Remark.** Consider the setting of Definition 7.12 and let b be a nonzero element of \mathcal{B} . Then b is an atom relative to \mathcal{A} if and only if $\mathbb{P}(b' \mid \mathcal{A})$ is equal to a restriction of $\mathbb{P}(b \mid \mathcal{A})$, for every $b' \leq b$ in \mathcal{B} . Here we are considering each $\mathbb{P}(\cdot \mid \mathcal{A})$ as a μ -ae equivalence class of \mathcal{A} -measurable [0, 1]-valued functions, and "restriction" means to multiply by the characteristic function of an \mathcal{A} -measurable set. (See Notation 2.4.)

Note that if \mathcal{B} is a probability algebra and $0 \neq b \in \mathcal{B}$, then b is an atom in \mathcal{B} if and only if b is an atom relative to the trivial subalgebra $\{0,1\}$.

7.14. **Lemma.** Suppose $\mathcal{M} = (\mathcal{B}, \mu, d)$ is a homogeneous model of APA and its density is κ , and $\mathcal{N} = (\mathcal{A}, \mu, d)$ is a substructure of \mathcal{M} of density $< \kappa$. Then \mathcal{B} is atomless over \mathcal{A} .

Proof. For each nonzero $b \in \mathcal{B}$, the density of $\mathcal{B} \upharpoonright b$ is κ , whereas the density of $\{a \cap b \mid a \in \mathcal{A}\}$ is at most the density of \mathcal{A} .

7.15. **Lemma.** Suppose $\mathcal{M} = (\mathcal{B}, \mu, d)$ is a model of APA, and $\mathcal{N} = (\mathcal{A}, \mu, d)$ is a substructure of \mathcal{M} . If \mathcal{B} is atomless over \mathcal{A} , then \mathcal{B} is atomless over $\langle \mathcal{A} \cup F \rangle$ for every finite set $F \subseteq \mathcal{B}$.

Proof. Using induction, it suffices to consider the case $F = \{b\}$. We prove the contrapositive. Suppose there is a nonzero b' in \mathcal{B} that is an atom relative to $\langle \mathcal{A} \cup \{b\} \rangle$. We show that $b' \cap b$ is either 0 or an atom relative to \mathcal{A} , and the same for $b' \cap b^c$. Since $b' \neq 0$, at least one of them must be an atom relative to \mathcal{A} .

Consider $b'' \leq b' \cap b$ (the case of $b' \cap b^c$ is similar). Note that $b'' \leq b$, so $b'' \cap b^c = 0$. Since b' is an atom relative to $\langle \mathcal{A} \cup \{b\} \rangle$, there exists $x \in \langle \mathcal{A} \cup \{b\} \rangle$ with $b'' = x \cap b'$. There exist $a_1, a_2 \in \mathcal{A}$ such that $x = (a_1 \cap b) \cup (a_2 \cap b^c)$, and therefore

$$b'' = x \cap b' = (a_1 \cap b \cap b') \cup (a_2 \cap b^c \cap b') = a_1 \cap (b' \cap b).$$

The last equality is because b'' and $a_2 \cap b^c \cap b'$ are disjoint, so $a_2 \cap b^c \cap b' = 0$. It follows that $b' \cap b$ is either 0 or an atom relative to \mathcal{A} .

7.16. **Lemma** (Maharam's lemma). Let $\mathcal{M} = (\mathcal{B}, \mu, d)$ be a model of APA and let $\mathcal{N} = (\mathcal{A}, \mu, d)$ be a substructure of \mathcal{M} . If \mathcal{B} is atomless over \mathcal{A} , then \mathcal{M} realizes every n-type over \mathcal{A} .

Proof. Using Lemma 7.15 and the fact that it allows us to realize n-types over \mathcal{A} "coordinate by coordinate", it suffices to prove the result for 1-types. Remark 6.9 implies that proving \mathcal{M} realizes every 1-type over \mathcal{A} is equivalent to proving the following statement:

Suppose $\lambda \colon \mathcal{A} \to [0,1]$ is an additive functional over \mathcal{A} such that $\lambda(a) \leq \mu(a)$ holds for every $a \in \mathcal{A}$. Then there exists $b \in \mathcal{B}$ such that $\lambda(a) = \mu(a \cap b)$ for every $a \in \mathcal{A}$.

A proof of exactly this statement is given as Lemma 3.2 in Fremlin's chapter [14] on measure algebras, and also as Lemma 331B in volume 3 [15] of his multi-volume treatise on measure theory.

7.17. Corollary. Every homogeneous model of APA is determined up to isomorphism by its density.

Proof. Suppose $\mathcal{M} = (\mathcal{B}, \mu, d) \models APA$ has density κ and is homogeneous. Since \mathcal{B} is homogeneous, it is atomless over $\langle C \rangle$ for every $C \subseteq \mathcal{B}$ with $\operatorname{card}(C) < \kappa$. By Lemma 7.16, \mathcal{M} realizes every n-type over C for every such C. That is, \mathcal{M} is a κ -saturated model of APA and it has density κ . Using the standard back-and-forth argument from model theory, any two such models are isomorphic.

Finally, we have Maharam's Theorem, which characterizes the structure of all probability algebras up to isomorphism.

7.18. **Theorem.** Every model \mathcal{M} of Pr is determined up to isomorphism by its invariants $\Phi^{\mathcal{M}}$ for the atomic part and its Maharam invariants $(\mathcal{K}^{\mathcal{M}}, \Psi^{\mathcal{M}})$ for the atomless part.

Proof. The definition of the Maharam invariants for general probability algebras is in the first paragraph of this section; the definition of $\Phi^{\mathcal{M}}$ is in Section 4. The proof of the Theorem is given above, with the key result being Corollary 7.17, which handles the maximal homogeneous components of the atomless part of \mathcal{M} . Note 7.11 indicates how the structure of \mathcal{M} is determined by what these invariants say about its component parts.

Note that for each infinite cardinal κ , we identified the κ -saturated model of APA of density character κ as the Maharam homogenous model of density κ . More information on κ -saturated and κ -homogeneous models of APA can be found in [26].

The following characterization of the "atomless over" property is often useful:

- 7.19. **Proposition.** Let $\mathcal{M} = (\mathcal{B}, \mu, d)$ be a model of APA and let \mathcal{A} be a σ -subalgebra of \mathcal{B} . The following are equivalent.
 - (a) \mathcal{B} is atomless over \mathcal{A} .
 - (b) For an infinite set of positive integers n, there is in \mathcal{B} a partition of 1, say $u = (u_1, \ldots, u_n)$, such that $\mu(a \cap u_i) = \frac{1}{n}\mu(a)$ for all $a \in \mathcal{A}$ and $i = 1, \ldots, n$. (In other words, each u_i satisfies $u_i \perp \!\!\! \perp \mathcal{A}$ and has measure $\frac{1}{n}$.)
 - (c) There is an atomless σ -subalgebra \mathcal{C} of \mathcal{B} such that $\mathcal{A} \perp \!\!\! \perp \mathcal{C}$.
- Proof. (a) \Rightarrow (c): We build inductively a sequence $\{C_n\}_{n\geq 1}$ of finite subalgebras of \mathcal{B} such that for all $n\geq 2$ we have $C_n \perp \!\!\! \perp (\mathcal{A} \cup (\bigcup_{i< n} C_i))$ and C_n is generated by a partition of 1, say (u_1,\ldots,u_n) , such that $\mu(u_i)=\frac{1}{n}$ for all $i=1,\ldots,n$. We take $C_1=\{0,1\}$. Assume we have built $\{C_i\}_{i< n}$. By Lemma 7.15 the algebra \mathcal{B} is atomless over $(\mathcal{A} \cup (\bigcup_{i< n} C_i))^{\#}$. The existence of C_n follows from Lemma 7.16, since we can describe the properties of (u_1,\ldots,u_n) by formulas over $(\mathcal{A} \cup (\bigcup_{i< n} C_i))^{\#}$. Now let C be the σ -algebra generated by $\bigcup_{i>1} C_i$.
- (c) \Rightarrow (b): This is immediate, since for any n there exists a partition of 1, say (u_1, \ldots, u_n) , in \mathcal{C} with $\mu(u_i) = \frac{1}{n}$ for all i, and $u_i \perp \!\!\! \perp \mathcal{A}$ automatically for all i.

(b) \Rightarrow (a): Let $0 \neq b \in \mathcal{B}$. Let (u_1, \ldots, u_n) be a partition of 1 in \mathcal{B} as in (b) such that $\frac{1}{n} \leq \frac{1}{2}\mu(b)$. Note that

$$\mathbb{P}(u_i \cap b \mid \mathcal{A}) \leq \mathbb{P}(u_i | \mathcal{A}) \leq \frac{1}{n} \leq \frac{1}{2}\mu(b).$$

Therefore, for some i we have that on a set of positive μ -measure

$$0 < \mathbb{P}(u_i \cap b \mid \mathcal{A}) < \mathbb{P}(b \mid \mathcal{A}),$$

which means that $u_i \cap b$ is not of the form $a \cap b$ with $a \in \mathcal{A}$. Thus b is not an atom relative to \mathcal{A} . (See Remark 7.13.)

8. Stability of APA

In this section we continue our study of the theory APA, concentrating on stability-theoretic properties. Throughout this section we work in a κ -universal domain for APA, which is denoted by \mathcal{U} , with underlying set U. A subset of U is *small* if its cardinality is $< \kappa$. Unless otherwise specified, we take parameter sets always to be small subsets of U. We adjust κ as needed for specific models of APA to be substructures of \mathcal{U} .

This section uses background on stability, forking, definitions of types, and canonical bases that can be found in [6] and [8].

Since APA is stable, by Proposition 6.15, we have the relation of model theoretic independence, denoted $C \downarrow_E D$, defined for small sets $C, D, E \subseteq U$ by:

 $C \downarrow_E D$ if and only if $\operatorname{tp}(a/DE)$ does not fork over E for all finite tuples a from $\langle C \rangle$. Our next result is that model theoretic independence is exactly the same as probabilistic independence, from which we also get a quantitative criterion for non-forking in APA. The corresponding result in the CAT setting was proved in [1, Theorem 2.10]. Our proof uses the same general approach, with details based on properties of conditional expectation and \bot that are discussed in Section 2.

The argument follows a familiar pattern: prove that in models of APA the relation \perp satisfies invariance, symmetry, finite character, transitivity, extension, and local character, and also that types of tuples over arbitrary sets are stationary. From this one gets that APA is stable and that \downarrow is the same as \perp (see [7, Theorem 14.14]). Throughout the proof we use the results from Lemma 2.9.

8.1. **Theorem.** Let $C, D, E \subseteq U$ be small. Then

$$C \underset{E}{\bigcup} D$$
 if and only if $C \underset{E}{\coprod} D$.

Consequently, for every $c = (c_1, \ldots, c_n) \in U^n$, we have that $\operatorname{tp}(c/DE)$ does not fork over E if and only if

$$\mathbb{P}(c^s|\langle DE \rangle) = \mathbb{P}(c^s|\langle E \rangle)$$

for all
$$s = (k_1, \dots, k_m) \in \{-1, +1\}^n$$
.

Proof. Let a be a finite tuple from U, and let A, C, D, E be small subsets of U. We prove each of the conditions invariance, symmetry, finite character, transitivity, extension, and local character, and also that types of tuples over arbitrary sets are stationary. (As we verify each condition, we make clear what it means.)

Invariance, Symmetry, and Finite Character: it is obvious from Definition 2.8 that the relation $C \perp\!\!\!\perp_E D$ is invariant under automorphisms of \mathcal{U} and equivalent to $D \perp\!\!\!\perp_E C$. Finite character requires that $C \perp\!\!\!\perp_E D$ holds if and only if $c \perp\!\!\!\perp_E D$ holds for every finite tuple c from C. This follows from the definition using the disjoint additivity of the conditional expectation operators (over $\langle D \rangle$ and over $\langle DE \rangle$). Indeed, if c_1, \ldots, c_k are the atoms in $c^{\#}$, then $c \perp\!\!\!\perp_E D$ iff for every $j = 1, \ldots, k$ we have $c_j \perp\!\!\!\perp_E D$.

Transitivity: This condition says that $a \perp\!\!\!\perp_E CDE$ if and only if $a \perp\!\!\!\perp_{CE} CDE$ and $a \perp\!\!\!\perp_E CE$. By Lemma 2.9((i) \Leftrightarrow (ii)), this statement is equivalent to the statement $\mathbb{P}(a|\langle CDE\rangle) = \mathbb{P}(a|\langle E\rangle)$ if and only if $\mathbb{P}(a|\langle CDE\rangle) = \mathbb{P}(a|\langle CE\rangle)$ and $\mathbb{P}(a|\langle CE\rangle) = \mathbb{P}(a|\langle E\rangle)$, which is true by Fact 2.6. Applying this for a ranging over $A^{\#}$ proves transitivity for $A \perp\!\!\!\perp_E CD$, using finite character.

Extension: We need to show that for all small subsets A, C, D of U, there is a copy E of A over C such that $E \perp\!\!\!\perp_C D$. By a "copy" we mean that there is a bijection f from A onto E such that for every $a \in A^n$, every L^{pr} -formula $\varphi(x;y)$, and every finite tuple c from C, the values of $\varphi(f(a);c)$ and $\varphi(a;c)$ in \mathcal{U} are equal. (In short: $\operatorname{tp}(A/C) = \operatorname{tp}(E/C)$, with f giving the correspondence between enumerations of A and E. Otherwise said, f is an elementary map over C, from A onto E.) Thus the statement that E is a copy of A over C is expressed by a family of L^{pr} -conditions in $\operatorname{card}(A)$ many variables and in parameters from C. The same is true of the condition $E \perp\!\!\!\!\perp_C D$, except that the parameters come from $C \cup D$. (Namely, for each $a = (a_1, \ldots, a_m) \in A^m$ and the corresponding $e = (f(a_1), \ldots, f(a_m))$, for each $u = (u_1, \ldots, u_n) \in D$ and for each $s \in \{-1, +1\}^m$ and $t \in \{-1, +1\}^n$, we require $\mathbb{P}(e^s \cap u^t | \langle C \rangle) = \mathbb{P}(a^s | \langle C \rangle) \cdot \mathbb{P}(u^t | \langle C \rangle)$. By the second paragraph of Remark 6.9, this is a type-definable condition over CD, since $\mathbb{P}(a^s | \langle C \rangle) \cdot \mathbb{P}(u^t | \langle C \rangle)$ is $\langle C \rangle$ -measurable. Note that taking $u^t = 1$, this independence condition already implies $\mathbb{P}(e^s | \langle C \rangle) = \mathbb{P}(a^s | \langle C \rangle)$), which is equivalent to $\operatorname{tp}(e/C) = \operatorname{tp}(a/C)$ by Lemma 6.4.)

Since \mathcal{U} is κ -saturated and the conditions discussed above involve $< \kappa$ many formulas, it suffices to show that this set of conditions is finitely satisfiable in \mathcal{U} . In particular, we may assume A is finite.

The rest of the argument is based on Lemma 2.13. Let (X, \mathcal{B}, μ) be a probability space whose probability algebra $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$ is a small elementary substructure of \mathcal{U} that contains ACD. Lemma 2.13 yields a probability space (X', \mathcal{B}', μ) and a map $B \mapsto B'$ that is a

measure-preserving boolean embedding of \mathcal{B} into \mathcal{B}' . The construction used in proving Lemma 2.13 ensures that the probability algebra of (X', \mathcal{B}', μ') is small. Using the fact that \mathcal{U} is κ -saturated and strongly κ -homogeneous, as well as the fact that APA has QE, we may realize $(\widehat{\mathcal{B}}', \widehat{\mu}', \widehat{d}')$ as a substructure of \mathcal{U} , and ensure that the induced embedding of $(\widehat{\mathcal{B}}, \widehat{\mu}, \widehat{d})$ into $(\widehat{\mathcal{B}}', \widehat{\mu}', \widehat{d}')$ is an inclusion. Therefore we obtain in $\widehat{\mathcal{B}}'$ a copy E of A over C such that $E \perp \!\!\!\perp_C D$, as desired.

Local Character: We need to show that there exists a countable set $C' \subseteq C$ such that $a \perp\!\!\!\perp_{C'} C$. Let C' be any countable set such that $\mathbb{P}(a|\langle C \rangle)$ is $\langle C' \rangle$ -measurable. Then $\mathbb{P}(a|\langle C' \rangle) = \mathbb{P}(a|\langle C \rangle)$, which implies $a \perp\!\!\!\perp_{C'} C$ by Lemma 2.9((i) \Leftrightarrow (iii)).

Stationarity of Types: We need to show that if $a \perp \!\!\! \perp_E D$, then $\operatorname{tp}(a/DE)$ is uniquely determined by $\operatorname{tp}(a/E)$. So assume $a,b \in U^n$ satisfy $a \perp \!\!\! \perp_E D$, $b \perp \!\!\! \perp_E D$, and $\operatorname{tp}(a/E) = \operatorname{tp}(b/E)$. So for each $s \in \{-1,+1\}^n$ we have $a^s \perp \!\!\! \perp_E D$ and $b^s \perp \!\!\! \perp_E D$ by Definition 2.8, and $\operatorname{tp}(a^s/E) = \operatorname{tp}(b^s/E)$ by Lemma 6.4. It follows for all s that

$$d(\operatorname{tp}(a^s/DE), \operatorname{tp}(b^s/DE)) = \|\mathbb{P}(a^s|\langle DE \rangle) - \mathbb{P}(b^s|\langle DE \rangle)\|_1$$
$$= \|\mathbb{P}(a^s|\langle E \rangle) - \mathbb{P}(b^s|\langle E \rangle)\|_1 = d(\operatorname{tp}(a^s/E), \operatorname{tp}(b^s/E)) = 0$$

by Corollary 6.12 and Lemma 2.9. Therefore tp(a/DE) = tp(b/DE) by Lemma 6.4.

8.2. **Remark.** Since APA is ω -stable (see Proposition 6.15), it follows that APA is also superstable, by [7, Remark 14.8]. In fact, APA has a property that is analogous, in the continuous logic setting, to the classical property of being superstable of finite SU-rank. To see this, take $\epsilon > 0$ and small sets $D \subseteq C \subseteq U$. Say that $\operatorname{tp}(a/C)$ ϵ -forks over D if $d(\operatorname{tp}(a/C), \operatorname{tp}(a'/C)) \ge \epsilon$, where $\operatorname{tp}(a'/C)$ is the (unique) non-forking extension of $\operatorname{tp}(a/D)$. Let $SU_{\epsilon}(\operatorname{tp}(a/D))$ be the foundation rank of $\operatorname{tp}(a/D)$ for this relation of ϵ -forking. Then for any $\epsilon > 0$, $a \in U$ and small $D \subseteq U$, it can be shown using Fact 2.6, and Theorems 6.11 and 8.1 that $SU_{\epsilon}(\operatorname{tp}(a/D))$ is at most $(1/\epsilon)^2$.

The next result shows that APA has built-in canonical bases. Before getting into the details, we provide some intuition about the connection between canonical bases and conditional probabilities. Consider the case where x is a single variable and $p(x) = \operatorname{tp}(a/C)$, where C is a small closed subalgebra of \mathcal{U} . Let $(a_i)_{i \in \mathbb{N}}$ be a Morley sequence in p and consider $(\chi_{a_i})_i$ as elements of $L^2(U,\mu)$. Let H be the Hilbert subspace corresponding to $L^2(C,\mu)$, that is, the collection of elements of $L^2(U,\mu)$ that are C-measurable, and let P_H be the orthogonal projection operator from $L^2(U,\mu)$ onto $L^2(C,\mu)$. Then, for each $i \in \mathbb{N}$, we may write $\chi_{a_i} = P_H(\chi_{a_i}) + v_i$ where $\{v_i\}_i$ are pairwise orthogonal and they all have the same norm. Then the sequence of averages $\left(\sum_{i=1}^n \frac{P_H(a_i)+v_i}{n}\right)_n$ converges (in $L^2(U,\mu)$) to $P_H(a_i) = \mathbb{P}(a|C)$, so $\mathbb{P}(a|C) \in \operatorname{dcl}(\{a_i\}_i)$. A similar computation can be carried out using

any Morley sequence in a type parallel to p, so $\mathbb{P}(a|C)$ belongs to the definable closure of the parallelism class of p and thus $\mathbb{P}(a|C) \in \operatorname{dcl}^{\operatorname{meq}}(Cb(p))$.

On the other hand, we would like to know the information that $\mathbb{P}(a|C)$ provides at the level of definability of types for p. By Lemma 6.4, to understand Cb(p), it is enough to find the p-definitions for the formulas $\psi_1(x,y) = \mu(x \cap y)$ and $\psi_2(x,y) = \mu(x^c \cap y) = \mu(y) \div \mu(x \cap y)$. Note that for any $c \in C$, we have

$$\psi_1^p(x,c) = \mu(a \cap c) = \int_c \mathbb{P}(a|C) \, d\mu \text{ and}$$
$$\psi_2^p(x,c) = \mu(c) - \int_c \mathbb{P}(a|C) \, d\mu.$$

Thus from $\mathbb{P}(a|C)$ we recover the *p*-definitions of the formulas $\psi_1(x,y)$ and $\psi_2(x,y)$, and so we recover Cb(p).

(Our approach uses the fact that that for APA we have proved that all types over small algebraically closed sets are stationary; indeed, stationarity for types over all sets follows from Theorem 8.1 and Lemma 6.4.)

8.3. **Theorem** (Prop. 4.5, [1]). Let $C \subseteq U$ be small and let $a = (a_1, \ldots, a_n) \in U^n$. Further, let D be the smallest σ -subalgebra of \mathcal{U} such that $\mathbb{P}(a^s|\langle C \rangle)$ is D-measurable for all $s = (k_1, \ldots, k_n) \in \{-1, +1\}^n$, so $D \subseteq \langle C \rangle$. Then D is a canonical base for $\operatorname{tp}(a/C)$.

Proof. See Notation 2.4 for some background that we use here, especially for what we mean precisely by D-measurability of $\mathbb{P}(a^s|\langle C\rangle)$ for σ -subalgebras $D\subseteq \langle C\rangle$ of \mathcal{U} .

Let $a = (a_1, \ldots, a_n) \in U^n$ and $p = \operatorname{tp}(a/C)$, and let τ be any automorphism of \mathcal{U} . We must prove that $\tau(p) := \operatorname{tp}(\tau(a)/\tau(C))$ is parallel to p (that is, that they have a common non-forking extension) if and only if τ fixes D pointwise.

First assume that $\tau(p)$ is parallel to p. Hence there is a type q over $\langle C \cup \tau(C) \rangle$ such that q extends p and $\tau(p)$, and also that q does not fork over $\langle C \rangle$, and q does not fork over $\langle \tau(C) \rangle = \tau(\langle C \rangle)$. Let $b = (b_1, \ldots, b_n) \models q$.

By Lemma 6.4 and Theorem 8.1, and the stated properties of q, we have

$$\mathbb{P}(\tau(a^s)|\tau(\langle C\rangle)) = \mathbb{P}(b^s|\tau(\langle C\rangle)) = \mathbb{P}(b^s|\langle C\cup\tau(C)\rangle) = \mathbb{P}(b^s|\langle C\rangle) = \mathbb{P}(a^s|\langle C\rangle)$$

for all $s \in \{-1, +1\}^n$. As discussed before Lemma 5.8, $\tau(\mathbb{P}(a^s|\langle C \rangle)) = \mathbb{P}(\tau(a^s)|\tau(\langle C \rangle))$, so $\tau(\mathbb{P}(a^s|\langle C \rangle)) = \mathbb{P}(a^s|\langle C \rangle)$, by the preceding calculation. Therefore, applying Lemma 5.8 to the representatives of $\mathbb{P}(a^s|\langle C \rangle)$ for each $s \in \{-1, +1\}^n$, we conclude $\tau(u) = u$ for every $u \in D$, as needed to be shown.

Conversely, assume that τ fixes D pointwise. We know $\mathbb{P}(a^s|\langle C\rangle) = \mathbb{P}(a^s|D)$ for all $s \in \{-1, +1\}^n$, so by Theorem 8.1 we get $a \downarrow_D C$. By Invariance for \downarrow , we also have

 $\tau(a) \downarrow_D \tau(C)$. Using Extension for \downarrow we get $p', q' \in S_n(C \cup \tau(C))$ such that p' is a nonforking extension of $p = \operatorname{tp}(a/C)$ and q' is a non-forking extension of $\tau(p) = \operatorname{tp}(\tau(a)/\tau(C))$. By Transitivity for \downarrow , it follows that both p' and q' are non-forking over D, so by Stationarity for \downarrow and the fact that $\operatorname{tp}(\tau(a)/D) = \operatorname{tp}(a/D)$, we conclude that p' = q'. It follows that p and $\tau(p)$ are parallel.

In [1], the perspective on canonical bases is the same as the one we use here. The proof of Prop. 4.5 in [1] shows that if E is any closed algebra $\subseteq U$ over which the type does not fork, and it is minimal with this property, then E coincides with D. Another approach to canonical bases can be found in [2], where Ben Yaacov shows that a better way of dealing with these objects is by introducing a sort for [0,1]-valued random variables associated to the corresponding probability space. It turns out that one can identify the canonical base of $\operatorname{tp}(a/C)$ with $\mathbb{P}(a|\mathcal{C})$ (a [0,1]-valued random variable) in a uniform way in order to construct uniform canonical bases (see [2, Definition 1.1 and Corollary 2.3]) a process which requires imaginaries for APA (see [2, Corollary 2.5]).

The fact that types have canonical bases in the home sort gives some information about elimination of imaginaries for APA. Namely, APA has weak elimination of metric imaginaries, which means that for every element a of an imaginary sort, there exists a subset A of the home sort such that $acl^{meq}(a) = dcl^{meq}(A)$. (See [4, Defn. 1.5].)

8.4. Corollary. The theory APA has weak elimination of metric imaginaries.

Proof. This follows from the fact that APA is stable and that it has canonical bases in the home sort (Theorem 8.3), together with [4, Fact 1.6].

We turn now to another property that is related to canonical bases, namely being strongly finitely based (SFB). Let $C \subseteq U$ be a small set, $a = (a_1, \ldots, a_n) \in U^n$, $x = (x_1, \ldots, x_n)$, and $p(x) = \operatorname{tp}(a/C)$. The type p(x) is stationary. For $\varphi(x;y)$ an L^{pr} -formula, with $y = (y_1, \ldots, y_k)$, let $d_x^p \varphi(y)$ be a φ -definition for p(x), which exists because APA is stable. Its main property is that for every $c \in C^k$, the p(x)-value of the formula $\varphi(x;c)$ equals the value of $d_x^p \varphi(c)$ in \mathcal{U}^{meq} . The φ -definition for p(x) can be constructed from a Morley sequence $(a^i)_{i \in \mathbb{N}}$ in any type parallel to p(x) over C, by defining it as the average value of $\varphi(x;c)$ along the Morley sequence:

$$d_x^p \varphi(c) = \lim_{k \to \infty} \frac{\sum_{i=1}^k \varphi(a^i; c)}{k}$$

This definition depends only on the parallelism class of p and not on the specific Morley sequence under consideration. So $d_x^p \varphi(y)$ is an $(L^{pr})^{\text{meq}}$ -formula in which a parameter from $dcl^{\text{meq}}(C)$ occurs.

Now consider $p, q \in S_n(C)$, with φ -definitions $d_x^p \varphi(y)$, $d_x^q \varphi(y)$ respectively. Another way to measure how much p, q differ is by considering the pseudometrics

$$d_{\varphi}(p,q) := \sup_{c \in C^k} |d_x^p \varphi(c) - d_x^q \varphi(c)|$$

as $\varphi(x;y)$ ranges over all L^{pr} -formulas in variables (x;y), with x fixed and $y=(y_1,\ldots,y_k)$ any finite sequence of parameter variables. These pseudometrics define a uniform structure on $S_n(C)$, which we denote by \mathcal{V}_{Cb} . The topology induced by \mathcal{V}_{Cb} on $S_n(C)$ is denoted by τ_{Cb} .

With the topology just described, we have three natural topologies on $S_n(C)$, namely, the logic topology $\tau_{\mathcal{L}}$, the metric topology τ_d and now the canonical base topology τ_{Cb} .

8.5. **Remark.** A more transparent way of evaluating $d_{\varphi}(p,q)$ comes from simply using what the p- and q-definitions express. Suppose $a \models p$ and $b \models q$, Then

$$d_{\varphi}(p,q) := \sup_{c \in C^k} |\varphi^{\mathcal{U}}(a;c) - \varphi^{\mathcal{U}}(b;c)|.$$

In the setting of APA, since types over the set C are stationary, we have $\tau_d \subseteq \tau_{Cb} \subseteq \tau_{\mathcal{L}}$ (see [6, Lemma 1.5]). When C is finite, adding names for the elements of C preserves \aleph_0 -categoricity over APA, and thus the three topologies are identical. On the other hand, when we take a sufficiently large set of parameters, for example if $\langle C \rangle$ is atomless and thus the universe of a model of APA, then $\tau_d \subsetneq \tau_{\mathcal{L}}$ (by the continuous Ryll-Nardzewski Theorem). It is natural to ask for more precise information for how the topology τ_{Cb} relates in general to the other two topologies.

8.6. **Definition** ([6]). A stable theory T is strongly finitely based (SFB) if for every $\mathcal{M} \models T$ and every n, the topologies τ_d and τ_{Cb} agree on $S_n(M)$.

Applications of the SFB property can be found in [6], where the concept was introduced as a continuous analogue of a strong version of the notion of being 1-based.

In that paper, the theory of lovely pairs was used to prove that APA has SFB. Below we give a direct proof using Theorem 6.11 and Lemma 6.14.

8.7. **Definition.** Let $C \subseteq U$ be a small subalgebra. Let $n \ge 1$ and $x = (x_1, \ldots, x_n)$ a tuple of distinct variables, and y a variable not occurring in x. Let $\varphi_s(x;y) := \mu(x^s \cap y)$ for each $s \in \{-1,1\}^n$. Then we define d_{Cb} on $S_n(C)$ by $d_{Cb}(p,q) := \max_s d_{\varphi_s}(p,q)$.

Obviously d_{Cb} is a pseudometric; Proposition 8.9 below shows that $d_{Cb}(p,q) = 0$ implies d(p,q) = 0, so d_{Cb} is in fact a metric.

¹The reason for including these details is to justify that our topology τ_{Cb} is the same as the topology introduced in [6]. See Remark 8.5 below for a more elementary formula for d_{φ} .

8.8. **Lemma.** Let $C \subseteq U$ be a small subalgebra. Then \mathcal{V}_{Cb} contains the uniform structure induced by d_{Cb} and is contained in the uniform structure induced by d.

Proof. As $\{\varphi_s\}_{s\in\{-1,+1\}^n}$ are L^{pr} -formulas, the set $\{(p,q)\in S_n(C)^2\mid d_{Cb}(p,q)\leq\epsilon\}$ is in \mathcal{V}_{Cb} for every $\epsilon>0$.

On the other hand, any L^{pr} -formula $\varphi(x; y_1, \ldots, y_k)$ with parameters from C is uniformly continuous with respect to d (see [7, Theorem 3.5]). Therefore, all formulas $\varphi(x; c_1, \ldots, c_k)$ have the same modulus of uniform continuity with respect to d. By [7, Proposition 2.8] it follows that d_{φ} is uniformly continuous with respect to d, with the same modulus.

8.9. **Proposition.** The theory APA is SFB; that is, the topologies τ_{Cb} and τ_d coincide over any set C of parameters. Indeed, the metrics d_{Cb} and d both induce the uniform structure \mathcal{V}_{Cb} on $S_n(C)$.

Proof. Let $C \subseteq U$ be small. Together with Lemma 8.8, it suffices to show that d is uniformly continuous with respect to d_{Cb} on $S_n(C)$. Fix $p, q \in S_n(C)$ and let $a, b \in U^n$ satisfy $a \models p$ and $b \models q$.

Recall from Theorem 6.11 and Lemma 6.14 that

$$\left(2^{-n+1}\right) \cdot d(p,q) \le d(\Pi_n(p), \Pi_n(q)) = \max_{s} \|\mathbb{P}(a^s | \langle C \rangle) - \mathbb{P}(b^s | \langle C \rangle)\|_1.$$

Fix $s \in \{-1, +1\}^n$ and $\epsilon \in (0, 1]$. Let k satisfy $k - 1 \le 1/\epsilon < k$, so $1/k < \epsilon$ and $k \le \frac{1}{\epsilon} + 1$. Let u_1, \ldots, u_k be a partition of 1 in $\langle C \rangle$ obtained using Lemma 2.7 applied to a^s and k. This ensures that for any closed subalgebra E of $\langle C \rangle$ that contains $\{u_1, \ldots, u_k\}$ we have $\|\mathbb{P}(a^s|\langle C \rangle) - \mathbb{P}(a^s|E)\|_1 \le 1/k$.

Similarly, let v_1, \ldots, v_k be a partition of 1 in $\langle C \rangle$ such that for any closed subalgebra E of $\langle C \rangle$ that contains $\{v_1, \ldots, v_k\}$ we have $\|\mathbb{P}(b^s|\langle C \rangle) - \mathbb{P}(b^s|E)\|_1 \leq 1/k$.

Then $E := \{u_1, \ldots, u_k, v_1, \ldots, v_k\}^{\#}$ is a finite subalgebra of $\langle C \rangle$ with at most k^2 atoms (namely all the intersections $u_i \cap v_j$) such that both $\|\mathbb{P}(a^s|\langle C \rangle) - \mathbb{P}(a^s|E)\|_1 \leq 1/k$ and $\|\mathbb{P}(b^s|\langle C \rangle) - \mathbb{P}(a^s|E)\|_1 \leq 1/k$. Let the atoms of E be e_1, \ldots, e_N , so $N \leq k^2$.

Using the triangle inequality for $\|\cdot\|_1$ we get

$$\|\mathbb{P}(a^s|\langle C\rangle) - \mathbb{P}(b^s|\langle C\rangle)\|_1 \le \|\mathbb{P}(a^s|E) - \mathbb{P}(b^s|E)\|_1 + 2\epsilon.$$

The proof is completed by the following estimate:

$$\|\mathbb{P}(a^{s}|E) - \mathbb{P}(b^{s}|E)\|_{1} = \|\sum_{j=1}^{N} \frac{\mu(a^{s} \cap e_{j})}{\mu(e_{j})} \chi_{e_{j}} - \sum_{j=1}^{N} \frac{\mu(b^{s} \cap e_{j})}{\mu(e_{j})} \chi_{e_{j}}\|_{1}$$

$$\stackrel{\star}{=} \sum_{j=1}^{N} \|\frac{\mu(a^{s} \cap e_{j}) - \mu(b^{s} \cap e_{j})}{\mu(e_{j})} \chi_{e_{j}}\|_{1} = \sum_{j=1}^{N} |\mu(a^{s} \cap e_{j}) - \mu(b^{s} \cap e_{j})|$$

$$= N \cdot d_{\varphi_{s}}(p, q) \leq (\frac{1}{\epsilon} + 1)^{2} \cdot d_{\varphi_{s}}(p, q).$$

(The equality (\star) holds because the different χ_{e_i} are disjointly supported.) Therefore

$$(2^{-n+1}) \cdot d(p,q) \le (\frac{1}{\epsilon} + 1)^2 \cdot d_{Cb}(p,q) + 2\epsilon,$$

so $d_{Cb}(p,q) < \epsilon^3$ implies $d(p,q) < (2^{n+2}) \cdot \epsilon$, showing that d is uniformly continuous relative to d_{Cb} .

When combined with Lemma 8.8, this argument completes the proof that d and d_{Cb} induce the same uniform structure on $S_n(C)$ (namely \mathcal{V}_{Cb}). It follows that the topologies τ_{Cb} and τ_d coincide, over any set C of parameters.

In the rest of this section we make a few connections with Shelah's classification program for models of classical first order theories, and we offer some speculative suggestions about how some aspects of the program might be carried into continuous model theory.

Shelah describes in [24] what one would hope for from a structure theorem versus nonstructure theorem that allows one to versus prevents one from completely classifying the
models of a complete theory T. This distinction is expressed in terms of invariants that
determine the models of T up to isomorphism. The invariants that come into the picture at
the lowest level of complexity (ordered by depth, which is in general any ordinal number),
are defined as follows. (When we assign an invariant to a model \mathcal{M} , it should only depend
on the isomorphism type of \mathcal{M} .) An invariant of depth 0 for models of T is an assignment
to each M of a cardinal $\leq \lambda = \operatorname{card}(M)$. An invariant of depth 1 for models of T is an
assignment to each \mathcal{M} of a set \mathcal{K} of cardinals $\leq \lambda = \operatorname{card}(M)$ together with a family of $\leq 2^{\aleph_0}$ many functions from the set \mathcal{K} to the set of cardinals $\leq \lambda$. (Frequently one ignores
models of cardinality $< \kappa$, for some infinite κ .) For example, if T is uncountably categorical
and its language is countable, the models of T have invariants of depth 0, namely to the
uncountable model \mathcal{M} is assigned $\operatorname{card}(M)$. Shelah's thesis is that T has a structure theory
iff there is an ordinal α and invariants (or sets of invariants) of depth α that determine every
model of T up to isomorphism.

The Maharam invariants for models of APA fit into this framework, with a twist that is not surprising, given that the setting has changed from classical model theory to its

continuous counterpart. Namely, invariants coming from the interval [0,1] come into the picture. Consider $\mathcal{M} \models APA$ of density λ . The invariant $\mathcal{K}^{\mathcal{M}}$ is a nonempty countable set of cardinal numbers that satisfies $\sup \mathcal{K}^{\mathcal{M}} = \lambda$; the additional invariant $\Phi^{\mathcal{M}}$ is a function from $\mathcal{K}^{\mathcal{M}}$ to (0,1] such that $\sum \{\Phi(\kappa) \mid \kappa \in \mathcal{K}^{\mathcal{M}}\} = 1$. This feels analogous to Shelah's invariants of depth 1, with the function $\Phi^{\mathcal{M}}$ as an additional feature.

Recall that for every model \mathcal{M} of APA there exists a family $(b_{\kappa} \mid \kappa \in \mathcal{K}^{\mathcal{M}})$ that witnesses the invariants $(\mathcal{K}^{\mathcal{M}}, \Phi^{\mathcal{M}})$ in the sense that each b_{κ} is maximal homogeneous of density κ and $\mu(b_{\kappa}) = \Phi^{\mathcal{M}}(\kappa)$ for each κ . (This implies that the elements b_{κ} are pairwise disjoint and their union is 1.) (See Section 7.)

Note that exactly as in Shelah's framework, the Maharam invariants can be used to calculate $I(\lambda, APA)$, which here is defined to be the number of models (up to isomorphism) of APA having density character λ . We know $I(\aleph_0, APA) = 1$. If $1 \le n < \omega$ and we are considering models of density \aleph_n , there are 2^n many choices for $\mathcal{K}^{\mathcal{M}}$ (it must contain \aleph_n) and 2^{\aleph_0} many choices of $\Phi^{\mathcal{M}}$ on each choice of $\mathcal{K}^{\mathcal{M}}$, except for the case $\mathcal{K}_{\mathcal{M}} = {\aleph_n}$, where $\Phi(\aleph_n) = 1$ is required. (This last choice is the invariant of the unique homogeneous model of density \aleph_n .) Hence $I(\aleph_n, APA) = 2^{\aleph_0}$. For an ordinal $\gamma \ge \omega$, a similar calculation shows that $I(\aleph_{\gamma}, APA) = (\operatorname{card}(\gamma))^{\aleph_0}$.

This analogy makes it seem likely that APA can be placed somewhere in a Shelah-style classification framework for ω -stable continuous theories with a countable language. To begin exploring this possibility, we introduce possible definitions of notions like *unidimensional* and *non-multidimensional* into the continuous logic setting, and explore the extent to which they apply to APA. Mostly they are taken directly from the classical first order discrete case, as presented in [11].

- 8.10. **Notation.** Let T be a stable theory and let $\mathcal{V} \models T$ be a κ -universal domain. Let $B \subseteq A \subset V$ be small and let $p \in S_n(B)$. When p is stationary, we let $p \upharpoonright A$ denote the unique type in $S_n(A)$ that is a non-forking extension of p.
- 8.11. **Definition.** Let T be a stable theory and let $\mathcal{V} \models T$ be a κ -universal domain. Let $A \subset V$ be small and let $p, q \in S(A) := \bigcup_n S_n(A)$. We say p, q are almost orthogonal and write $p \perp^a q$ if for all $b \models p$ and all $c \models q$ we have $b \downarrow_A c$. Given stationary types $p \in S(B), q \in S(C)$, with $B, C \subseteq V$ small, we say that p and q are orthogonal and write $p \perp q$ if for all small sets A with $B \cup C \subseteq A \subseteq V$, we have $(p \uparrow A) \perp^a (q \uparrow A)$.

We say a theory T is unidimensional if $p \not\perp q$ whenever p, q are non-algebraic stationary types over small sets of parameters $\subseteq V$.

Whenever p is a stationary type, we write $p \perp \emptyset$ if $p \perp q$ for all $q \in S(\operatorname{acl}^{\operatorname{meq}}(\emptyset))$.

For superstable T, we say T is bounded or non-multidimensional if every non-algebraic stationary type p satisfies $p \not\perp \emptyset$.

Now we return to APA. For types over \emptyset the picture is very simple:

8.12. **Proposition.** If $p \in S_m(\emptyset)$ and $q \in S_n(\emptyset)$ are non-algebraic, then p and q are not almost orthogonal.

Proof. It suffices to consider types of partitions of 1, since any n-tuple is interdefinable with its associated partition of 1.

Consider two partitions of 1, say $a = (a_1, \ldots, a_m)$ and $b = (b_1, \ldots, b_n)$ in \mathcal{U} , and suppose that neither type $\operatorname{tp}(a), \operatorname{tp}(b)$ is algebraic (here this means that at least one a_i and at least one b_j are distinct from 0 and 1).

Let \mathcal{M} be the probability algebra of the standard Lebesgue space ([0, 1], \mathcal{B} , μ), in which all n-types over \emptyset can be realized. Realize $\operatorname{tp}(a)$ in \mathcal{M} by a sequence of pairwise disjoint intervals (I_1, \ldots, I_m) , where each interval is of the form [r, s) for r < s in [0, 1] and $\sup I_k = \min I_{k+1}$ for every $k = 1, \ldots, m-1$. Then the union of the intervals I_i is [0, 1). Let (J_1, \ldots, J_n) be a similar sequence of intervals that realizes $\operatorname{tp}(b)$.

Choose i, j least so that $0 < \mu(a_i) < 1$ and $0 < \mu(b_j) < 1$. By the choice of i, j we have $I_i = [0, \mu(a_i))$ and $J_j = [0, \mu(b_j))$.

Then $\mu(I_i \cap J_j) = \min(\mu(a_i), \mu(b_j)) > \mu(a_i)\mu(b_j) = \mu(I_i)\mu(J_j)$. Therefore we have non-independent realizations of the types $\operatorname{tp}(a), \operatorname{tp}(b)$. This shows that no pair of non-algebraic types over \emptyset is almost orthogonal.

Once we allow parameters, elements may be supported over disjoint sets and we obtain more freedom:

8.13. **Proposition.** For any $e \in U$ with $e \notin \{0,1\}$ there are non-algebraic $p,q \in S_1(\{e\})$ which are orthogonal.

Proof. Let $e \in U$ have $0 < \mu(e) < 1$ and let $E = \{0, 1, e, e^c\}$. Let $a, b \in U$ satisfy $a \le e$, $\mu(a) = \mu(e)/2$, $b \le e^c$, and $\mu(b) = \mu(e^c)/2$. Consider $p = \operatorname{tp}(a/E)$ and $q = \operatorname{tp}(b/E)$. Claim 1. $p \perp^a q$.

We work in the measure algebra $(\widehat{\mathcal{B}}, \mu, d)$ associated to the standard Lebesgue space $([0, 1], \mathcal{B}, d)$, which is \aleph_0 -saturated, and may assume that $(\widehat{\mathcal{B}}, \mu, d)$ is an elementary substructure of \mathcal{U} . Let $r = \mu(e)$. Since the type of an element is determined by its measure, we take e to be (the equivalence class of) [0, r) and e^c to be (the equivalence class of) [r, 1]. Let $a \leq e$ have $\mu(a) = r/2$, so $a \models p$; similarly let $b \leq e^c$ have $\mu(b) = (1 - r)/2$, so $b \models q$.

Note that $\mathbb{P}(a|E) = \frac{1}{2}\chi_e$ and $\mathbb{P}(a^c|E) = \frac{1}{2}\chi_e + \chi_{e^c}$. Since $b \leq e^c$, we have $b \cap a = \emptyset$, $b^c \cap a = e \cap a$ and $\mathbb{P}(a|(Eb)^\#) = \frac{1}{2}\chi_e$. Similarly, $\mathbb{P}(a^c|(Eb)^\#) = \frac{1}{2}\chi_e + \chi_{e^c}$. By Theorem 8.1, we have $a \downarrow_E b$, and thus Claim 1 is proved.

Claim 2. $p \perp q$.

Now let a small closed subalgebra $F \subseteq U$ have $E \subseteq F$. Choose $a_F, b_F \in U$ with $a_F \models p \upharpoonright F$, $b_F \models q \upharpoonright F$; then as before $a_F \leq e$ and by Theorem 8.1, we have $\mathbb{P}(a_F|F) = \frac{1}{2}\chi_e$, $\mathbb{P}(a_F^c|F) = \frac{1}{2}\chi_e + \chi_{e^c}$. Likewise $b_F \leq e^c$ and we get $\mathbb{P}(a_F|(Fb_F)^\#) = \frac{1}{2}\chi_e = \mathbb{P}(a_F|F)$, $\mathbb{P}(a_F^c|(Fb_F)^\#) = \frac{1}{2}\chi_e + \chi_{e^c} = \mathbb{P}(a_F^c|F)$. Using Theorem 8.1 we conclude $a_F \downarrow_F b_F$, as desired.

From the previous proposition we get:

8.14. Corollary. The theory APA is not unidimensional.

Maharam's Theorem (7.18) provides a countable set of cardinals $\mathcal{K}^{\mathcal{M}}$ that helps classify a given model of APA up to isomorphism. In the classical first order setting, the existence of two or more distinct classifying cardinals is related to the existence of orthogonal types. In the example that follows, we illustrate this phenomenon in the setting of continuous model theory, for APA.

8.15. **Example.** Let $(\mathcal{B}, \mu, d) \models APA$ and assume $b_1, b_2 \in \mathcal{B}$ are homogeneous elements of different density. Let C be the algebra generated by $\{b_1, b_2\}$. Let $a_1, a_2 \in \mathcal{B}$ satisfy $0 < a_i < b_i$ for i = 1, 2. Then $\operatorname{tp}(a_1/C) \perp \operatorname{tp}(a_2/C)$.

More generally, assume that $b_1, b_2 \in \mathcal{B}$ are disjoint elements and C is an algebra containing these elements. Also assume we are given $a_1, a_2 \in \mathcal{B}$ nonalgebraic over C with $a_i < b_i$ for i = 1, 2. Then $\operatorname{tp}(a_1/C) \perp \operatorname{tp}(a_2/C)$.

Proof. We start with the first statement. Since b_1, b_2 are homogeneous elements of different density we have $b_1 \cap b_2 = \emptyset$ and thus $C = \{0, 1, b_1, b_2, (b_1 \cup b_2)^c\}$. We first show $a_1 \downarrow_C a_2$.

We see $\mathbb{P}(a_1|C) = \frac{\mu(a_1)}{\mu(b_1)}\chi_{b_1}$ and $\mathbb{P}(a_1^c|C) = \frac{\mu(b_1)-\mu(a_1)}{\mu(b_1)}\chi_{b_1} + \chi_{b_1^c}$. On the other hand we have $a_2 \leq b_2$ and b_2 is disjoint from b_1 , so $\mathbb{P}(a_1|(Ca_2)^{\#}) = \frac{\mu(a_1)}{\mu(b_1)}\chi_{b_1} = \mathbb{P}(a_1|Ca_2)$ and $\mathbb{P}(a_1^c|(Cb_2)^{\#}) = \frac{\mu(b_1)-\mu(a_1)}{\mu(b_1)}\chi_{b_1} + \chi_{b_1^c} = \mathbb{P}(a_1^c|C)$. Therefore we have $a_1 \downarrow_C a_2$. Claim. $\operatorname{tp}(a_1/C) \perp \operatorname{tp}(a_2/C)$.

Let $F \subseteq U$ be a small closed subalgebra with $C \subseteq F$ and let $p
mathcal{T} F$ and $q
mathcal{T} F$ be the nonforking extensions to F of $p = \operatorname{tp}(a_1/C)$ and $q = \operatorname{tp}(a_2/C)$ respectively. Choose $a_{1F}, a_{2F} \in U$ with $a_{1F} \models p
mathcal{T} F$, $a_{2F} \models q
mathcal{T} F$. By Theorem 8.1, we have $\mathbb{P}(a_{1F}|F) = \frac{\mu(a_1)}{\mu(b_1)} \chi_{b_1}$ and $\mathbb{P}(a_{1F}^c|C) = \frac{\mu(b_1) - \mu(a_1)}{\mu(b_1)} \chi_{b_1} + \chi_{b_1^c}$. Since $a_{2F} \leq b_2$ and b_1 and b_2 are disjoint, we get $\mathbb{P}(a_{1F}|(Fa_{2F})^\#) = \frac{\mu(a_1)}{\mu(b_1)} \chi_{b_1} = \mathbb{P}(a_{1F}|F)$ and $\mathbb{P}(a_{1F}^c|(Fa_{2F})^\#) = \frac{\mu(b_1) - \mu(a_1)}{\mu(b_1)} \chi_{b_1} + \chi_{b_1^c} = \mathbb{P}(a_{1F}^c|F)$.

Using Theorem 8.1 we conclude $a_{1F} \downarrow_F a_{2F}$, as desired.

The more general statement has a similar proof and we leave the details to the reader. \Box

We need the following easy result.

8.16. **Observation.** Let $0 < \delta < \frac{1}{2}$ and let $\delta \le r \le 1 - \delta$. Then $\min\{\frac{1}{2}, r\} - \frac{1}{2}r \ge \frac{\delta}{2}$.

Proof. Assume first that $\min\{\frac{1}{2}, r\} = \frac{1}{2}$. Then $\min\{\frac{1}{2}, r\} - \frac{1}{2}r = \frac{1}{2}(1 - r) \ge \frac{\delta}{2}$. On the other hand, if $\min\{\frac{1}{2}, r\} = r$, then $\min\{\frac{1}{2}, r\} - \frac{1}{2}r = \frac{1}{2}r \ge \frac{\delta}{2}$.

8.17. **Proposition.** The theory APA is nonmultidimensional.

Proof. Let $a \in U$ with $\mu(a) = 1/2$ and let $q = \operatorname{tp}(a/\emptyset)$. We will show any non-algebraic type over any set is non-orthogonal to q. Let $D \subset U$ be a small closed subalgebra of \mathcal{U} and let $b = (b_1, \ldots, b_k)$ be a partition of 1 such that $p = \operatorname{tp}(b/D)$ is not algebraic. We may assume without loss of generality that $b_1 \notin D$ and we may work with $p_1 = \operatorname{tp}(b_1/D)$ instead of p. Below we will show that $p_1 \not \perp^a q \not \mid D$.

Since $b_1 \notin D$, there is $u \in D$ of positive measure and $\delta > 0$ such that

(*)
$$\delta \leq \mathbb{P}(b_1|D)(x) \leq 1 - \delta \text{ for } \mu\text{-almost every } x \in u.$$

We may assume \mathcal{U} is the probability algebra associated to a probability space (X, \mathcal{B}, μ) . Let $([0,1], \mathcal{C}, \lambda)$ be a standard atomless Lebesgue space and work in the probability algebra of the space $(X \times [0,1], \mathcal{B} \otimes \mathcal{C}, \mu \otimes \lambda)$, identifying each $v \in D$ with $v' = v \times [0,1]$ as done in the proof of Lemma 2.13.

Let $b'_1 = \{(x, y) \in X \times [0, 1] : 0 \le y \le \mathbb{P}(b_1|D)(x)\}$. Just as in the proof of Lemma 2.13, we have that $b'_1 \models p_1$. Also let $a' = X \times [0, 1/2] = \{(x, y) \in X \times [0, 1] : 0 \le y \le \frac{1}{2}\}$, which is a realization of $q \nmid D$ since it has measure 1/2 and is independent from all elements of U.

We will prove $b'_1 \not\perp_D a'$ using Theorem 8.1 and Definition 2.8.

Note that $\mathbb{P}(b_1'|D)\mathbb{P}(a'|D) = \frac{1}{2}\mathbb{P}(b_1'|D)$. On the other hand,

$$b'_1 \cap a' = \{(x, y) \in X \times [0, 1] : 0 \le y \le \min(\mathbb{P}(b_1|D)(x), \frac{1}{2})\}.$$

For almost every $x \in u$, we get

$$\mathbb{P}(b_1' \cap a'|D)(x) - \frac{1}{2}\mathbb{P}(b_1'|D)(x) \ge \delta/2$$

using (*) and Observation 8.16.

Since u has positive measure, we get

$$\int \left| \mathbb{P}(b_1' \cap a'|D) - \mathbb{P}(a'|D)\mathbb{P}(b_1'|D) \right| d(\mu \otimes \lambda) \ge \mu(u)\delta/2 > 0$$

and thus $b'_1 \not\downarrow _D a'$ as desired.

9. Ranks obtained from entropy

In this section we discuss the definition and main properties of entropy, following [27, Chapter 4] and [10, Chapter 4], to bring out the connection with model theoretic aspects of APA. The results given in Fact 9.4 and Corollary 9.5 show how entropy provides a rank that is closely connected to model theoretic forking.

Let (X, \mathcal{B}, μ) be an atomless probability space.

9.1. **Definition.** Let \mathcal{A} be a finite subalgebra of \mathcal{B} with atoms $\{A_1, \ldots, A_k\}$. Let \mathcal{C} be a σ -subalgebra of \mathcal{B} . Then the *entropy of* \mathcal{A} *given* \mathcal{C} is

$$H(\mathcal{A}/\mathcal{C}) = -\int \sum_{1 \le i \le k} \mathbb{P}(A_i|\mathcal{C}) \ln(\mathbb{P}(A_i|\mathcal{C})) d\mu$$

We write H(A) for $H(A/\{\emptyset, X\})$. If A and C are σ -algebras, we denote by $A \vee C$ the σ -algebra generated by A and C.

9.2. **Definition.** A continuous real-valued function F with domain [a, b] is convex if

$$F(tx_1 + (1-t)x_2) \le tF(x_1) + (1-t)F(x_2)$$

for all choices of x_1, x_2 in [a, b] and all $t \in [0, 1]$.

Note that if $F: [a, b] \to \mathbb{R}$ is continuous and is twice differentiable on (a, b), and if F''(x) > 0 for all x in (a, b), then F is convex.

9.3. Fact ([10] Proposition 4.4). Let \mathcal{E} be a σ -subalgebra of \mathcal{B} . Let $F:[a,b] \to \mathbb{R}$ be a continuous convex function, where $0 \le a \le b < \infty$. Then

$$F(\mathbb{E}_{\mathcal{E}}(f)) \le \mathbb{E}_{\mathcal{E}}(F(f))$$

for each $f \in L_1(X, \mathcal{B}, \mu)$ with $f(X) \subseteq [a, b]$.

- 9.4. Fact. Let \mathcal{A} , \mathcal{C} be finite subalgebras of \mathcal{B} and let \mathcal{D} , \mathcal{E} be σ -subalgebras of \mathcal{B} such that $\mathcal{E} \subseteq \mathcal{D}$. Let $\{a_1, a_2, \ldots, a_n\}$ be the atoms in \mathcal{A} . Then:
 - (1) $H(A \vee C/\mathcal{E}) = H(A/\mathcal{E}) + H(C/A \vee \mathcal{E}).$
 - $(2) \ \mathcal{A} \subseteq \mathcal{C} \implies H(\mathcal{A}/\mathcal{E}) \leq H(\mathcal{C}/\mathcal{E}).$
 - (3) $H(\mathcal{A}/\mathcal{E}) \ge H(\mathcal{A}/\mathcal{D})$.
 - (4) If τ is an automorphism of $(\mathcal{B}, \mu.d)$, then $H(\tau(\mathcal{A})/\tau(\mathcal{E})) = H(\mathcal{A}/\mathcal{E})$.
 - (5) $H(\mathcal{A}/\mathcal{E}) H(\mathcal{A}/\mathcal{D}) \ge \frac{1}{2} \sum_{j=1}^{n} (\|\mathbb{E}_{\mathcal{D}}(\chi_{a_j})\|_2^2 \|\mathbb{E}_{\mathcal{E}}(\chi_{a_j})\|_2^2) \ge 0.$ Moreover, $H(\mathcal{A}/\mathcal{E}) = H(\mathcal{A}/\mathcal{D})$ iff \mathcal{A} is independent from \mathcal{D} over \mathcal{E} .

Proof. The first four properties are proved in [27, Section 4.3]. Note that (5) implies (3), and (4) is obvious from the definition. Throughout the argument, we let a be any element of U; we will apply the results for a ranging over the set of atoms $\{a_1, \ldots, a_n\}$.

Fact 2.6 shows that

(A)
$$\|\mathbb{E}_{\mathcal{D}}(\chi_a)\|_2^2 - \|\mathbb{E}_{\mathcal{E}}(\chi_a)\|_2^2 = \|\mathbb{E}_{\mathcal{D}}(\chi_a) - \mathbb{E}_{\mathcal{E}}(\chi_a)\|_2^2 \ge 0.$$

Applying this for $a \in \{a_1, \ldots, a_n\}$ proves the second inequality in (5).

Next we prove the first inequality in (5).

Consider $F(x) = 2x \ln(x) - x^2$ restricted to [0,1]. We have F''(x) = 2/x - 2 > 0 for $x \in (0,1)$. Applying Fact 9.3 for this F and $f = \mathbb{E}_{\mathcal{D}}(\chi_a)$ we get

$$2\mathbb{E}_{\mathcal{E}}(\chi_a)\ln(\mathbb{E}_{\mathcal{E}}(\chi_a)) - \mathbb{E}_{\mathcal{E}}(\chi_a)^2 \le 2\mathbb{E}_{\mathcal{E}}(\mathbb{E}_{\mathcal{D}}(\chi_a)\ln(\mathbb{E}_{\mathcal{D}}(\chi_a))) - \mathbb{E}_{\mathcal{E}}(\mathbb{E}_{\mathcal{D}}(\chi_a)^2).$$

Integrating and moving the terms in the preceding inequality yields

(B)
$$\left(-\int \mathbb{P}(a|\mathcal{E}) \ln(\mathbb{P}(a|\mathcal{E})) d\mu \right) - \left(-\int \mathbb{P}(a|\mathcal{D}) \ln(\mathbb{P}(a|\mathcal{D})) d\mu \right)$$

$$\geq \frac{1}{2} \left(\|\mathbb{E}_{\mathcal{D}}(\chi_a)\|_2^2 - \|\mathbb{E}_{\mathcal{E}}(\chi_a)\|_2^2 \right).$$

Summing the terms in (B) over $a \in \{a_1, \ldots, a_n\}$ yields the first inequality in (5).

Finally, we prove the "Moreover" statement. We know \mathcal{A} is independent from \mathcal{D} over \mathcal{E} iff $\mathbb{E}_{\mathcal{D}}(\chi_{a_j}) = \mathbb{E}_{\mathcal{E}}(\chi_{a_j})$ for all $j = 1 \dots, n$ (by Theorem 8.1), and the latter implies $H(\mathcal{A}/\mathcal{E}) = H(\mathcal{A}/\mathcal{D})$, by definition of entropy. On the other hand, by the inequalities in (5), we see $H(\mathcal{A}/\mathcal{E}) = H(\mathcal{A}/\mathcal{D})$ implies $\|\mathbb{E}_{\mathcal{D}}(\chi_{a_j})\|_2^2 = \|\mathbb{E}_{\mathcal{E}}(\chi_{a_j})\|_2^2$ for all $j = 1, \dots, n$, which in turn implies $\mathbb{E}_{\mathcal{D}}(\chi_{a_j}) = \mathbb{E}_{\mathcal{E}}(\chi_{a_j})$ for all $j = 1 \dots, n$ by statement (A) applied to $a = a_1, \dots, a_n$. \square

Fact 9.4(5) provides a connection between forking and change of entropy. It has a quantitative aspect that we record here. Recall that for $a = (a_1, \ldots, a_n)$ from $\widehat{\mathcal{B}}$ and $D \subseteq C \subseteq \widehat{\mathcal{B}}$, we say $\operatorname{tp}(a/C)$ ϵ -forks over D if $d(\operatorname{tp}(a/C), \operatorname{tp}(a/D) \upharpoonright C) > \epsilon$ where $\operatorname{tp}(a/D) \upharpoonright C$ is the unique non-forking extension of $\operatorname{tp}(a/D)$ to C. (See Remark 8.2.)

9.5. Corollary. Let \mathcal{A} be a finite subalgebra of \mathcal{B} and let \mathcal{E} , \mathcal{D} be σ -subalgebras of \mathcal{B} such that $\mathcal{E} \subseteq \mathcal{D}$. Let $\{a_1, a_2, \ldots, a_n\}$ be the events corresponding to the atoms in \mathcal{A} , D the set of events associated to \mathcal{D} and E the set of events associated to \mathcal{E} . If $\operatorname{tp}((a_1, \ldots, a_n)/D)$ ϵ -forks over E, then $H(\mathcal{A}/\mathcal{E}) > H(\mathcal{A}/\mathcal{D}) + \epsilon^2/2$.

Proof. Assume that $\operatorname{tp}(a_1, \ldots, a_n/D)$ ϵ -forks over E. Then by Theorems 6.11 and 8.1, for some $j = 1, \ldots, n$ we have $\|\mathbb{E}_{\mathcal{D}}(a_j) - \mathbb{E}_{\mathcal{E}}(a_j)\|_1 > \epsilon$. By Fact 2.6 this implies $\|\mathbb{E}_{\mathcal{D}}(a_j) - \mathbb{E}_{\mathcal{E}}(a_j)\|_2 > \epsilon$ and thus $\|\mathbb{E}_{\mathcal{D}}(a_j) - \mathbb{E}_{\mathcal{E}}(a_j)\|_2^2 > \epsilon^2$. Then we get $H(\mathcal{A}/\mathcal{E}) > H(\mathcal{A}/\mathcal{D}) + \epsilon^2/2$ by the inequality in Fact 9.4(5).

10. Some problems

In this final section we briefly indicate a few problems that seem interesting and worth investigation.

(P1) Give an explicit formula for the induced distance between types in $S_n(C)$ for APA.

$$d(p,q) = \inf\{\max_{1 \le i \le n} d(a_i, b_i) : (a_1, \dots, a_n) \models p, (b_1, \dots, b_n) \models q\}.$$

(P2) Provide a thorough analysis of the imaginary sorts for APA.

(P3) Complete the model theoretic background behind a generalization to continuous model theory of Shelah's classification theory for superstable theories. (Some first steps for this as applied to APA were discussed at the end of Section 8.). In particular, study appropriate versions of properties such as DOP (dimensional order property) and OTOP (omitting types order property) in the continuous setting and prove a dichotomy theorem relating a small bound for $I(\lambda, T)$ to when T is superstable and has neither DOP nor OTOP, along the lines of [24, Theorem 2.3].

There is also the possibility of proving the equivalence, for continuous theories, between uncountable categoricity and being both ω -stable and unidimensional, as is true for classical first order theories.

(P4) Consider two existentially closed actions of the free group F_k on the unique separable model \mathcal{M} of APA, where $2 \leq k \in \mathbb{N} \cup \{\omega\}$. Are they approximately isomorphic?

In the joint paper [9] of the authors with Ibarlucía, the class of existentially closed actions by a family of automorphisms of \mathcal{M} is axiomatized, and some concrete examples of such actions are given that are approximately isomorphic but not isomorphic. The answer is known to be positive when k = 1. (See [7, Remark 18.9]).

References

- [1] Itaï Ben Yaacov, Schroedinger's cat, Israel Journal of Mathematics 153, 2006, 157–191.
- [2] _____, On uniform canonical bases in L_p lattices and other metric structures, Journal of Logic and Analysis 4, 2012, paper 12, 30pp.
- [3] ______, On theories of random variables, Israel Journal of Mathematics 194, 2013, 957–1012.
- [4] _____, On a Roelcke-precompact Polish group that cannot act transitively on a complete metric space, *Israel Journal of Mathematics* 224, 2018, 105–132.
- [5] ______, Star sorts, Lelek fans, and the reconstruction of non-ℵ₀-categorical theories in continuous logic, to appear in *Model Theory*, 23 pages, arxiv:2203.02184v3.
- [6] Itaï Ben Yaacov, Alexander Berenstein, and C. Ward Henson, Almost indiscernible sequences and convergence of canonical bases, *Journal of Symbolic Logic* 79, 2014, 460–484.
- [7] Itaï Ben Yaacov, Alexander Berenstein, C. Ward Henson, and Alexander Usvyatsov, Model theory for metric structures, in *Model Theory with Applications to Algebra and Analysis*, Vol. II, eds. Z. Chatzidakis, D. Macpherson, A. Pillay, and A. Wilkie, Lecture Notes series of the London Mathematical Society, No. 350, Cambridge University Press, 2008, 315–427.
- [8] Itaï Ben Yaacov and Alexander Usvyatsov, Continuous first order logic and local stability, *Transactions of the American Mathematical Society* 362, 2010, 5213–5259.
- [9] Alexander Berenstein, C. Ward Henson, and Tomás Ibarlucía, Existentially closed measure-preserving actions of free groups, submitted, arXiv:2203.10178.
- [10] James R. Brown, Ergodic Theory and Topological Dynamics, Academic Press, New York, 1976.
- [11] Steven Buechler, Essential Stability Theory, Springer-Verlag, Berlin, Heidelberg, 1996.
- [12] Gerald B. Folland, Real Analysis, John Wiley and Sons, 1984.

- [13] Ilijas Farah, Bradd Hart, Martino Lupini, Leonel Robert, Aaron Tikuisis, Alessandro Vignati, and Wilhelm Winter, *Model Theory of C*-algebras*, Memoirs of the AMS No. 1324 (2021).
- [14] D. H. Fremlin, Measure algebras, in *Handbook of Boolean Algebras*, vol. 3, North-Holland, 1989, 877–980.
- [15] ______, Measure Algebras, vol. 3 of Measure Theory, Torres Fremlin, 2003-04; for information see http://www.essex.ac.uk/maths/staff/fremlin/mt.htm.
- [16] Paul A. Halmos, Measure Theory, Van Nostrand, 1950.
- [17] Wilfrid Hodges, Model Theory, Cambridge University Press, 1993.
- [18] Olav Kallenberg, Foundations of Modern Probability, 3rd edition, Springer (2021).
- [19] Tom Lindstrøm, An invitation to nonstandard analysis, in *Nonstandard Analysis and its Applications*, ed. Nigel Cutland, London Mathematical Society, Student Texts, No. 10, Cambridge University Press, 1988, 1–105.
- [20] Peter A. Loeb, Conversion from nonstandard to standard measure spaces and applications in probability theory, *Transactions of the American Mathematical Society* 211, 1975, 113–122.
- [21] Dorothy Maharam, On homogeneous measure algebras, Proceedings of the National Academy of Sciences USA 108, 1942, 108–111.
- [22] David Ross, Loeb measure and probability, in Nonstandard Analysis: Theory and Practice (eds. L. O. Arkeryd, C. W. Henson, and N. J. Cutland), NATO Advanced Study Institutes Series C, Vol. 493, Kluwer Academic Publishers, 1997, 91–120.
- [23] H. L. Royden, Real Analysis third edition, Prentice Hall, 1988.
- [24] Saharon Shelah, Classification of first order structures which have a structure theorem, Transactions of the American Mathematical Society 12, 1985, 227–232
- [25] Paul Shields, *The Theory of Bernoulli Shifts*, Chicago Lectures in Mathematics, The University of Chicago Press, 1973.
- [26] Shichang Song, Saturated structures from probability theory, Journal of the Korean Mathematical Society, 53,2016, 315–329.
- [27] Peter Walters, An Introduction to Ergodic Theory, Springer Verlag, 1982.

ALEXANDER BERENSTEIN; UNIVERSIDAD DE LOS ANDES,

Cra 1 Nº 18A-12, Edificio H, Bogotá, Colombia

URL: http://www.matematicas.uniandes.edu.co/~aberenst

C. WARD HENSON; UNIVERSITY OF ILLINOIS, URBANA-CHAMPAIGN;

Urbana, Illinois 61801, USA

URL: https://faculty.math.illinois.edu/~henson