Part E

Logics of Topology and Analysis

This part of the book is devoted to logics which presuppose different kinds of structures than those underlying first-order logic and its extensions so far dealt with in Parts B, C and D.

Chapter XIV is about logics where the underlying structure is a probability space, a structure with a countably additive probability measure. In addition to the usual propositional operations, the basic form of quantification is given by allowing formulas

$$(Px \geq r)\phi(x),$$

which means that the probability of the set $\{x: \phi(x)\}$ is at least r. Structures take the form of probability spaces with countably additive measures. To have a successful theory here a number of changes in perspective must be made. In the first place, one must arrange things so that all definable sets are measurable. As a result, the logics considered here are not closed under the usual quantifiers $\forall$ and $\exists$. Consequently, these logics do not contain first-order logic, nor do they satisfy all the assumptions on logics given in the general definition. They also have model-theoretic properties that have no first-order analogue, like the Law of Large Numbers.

While the lack of ordinary quantifiers entail a loss in expressive power, we can compensate for that, in part, by the use of countable conjunctions and disjunctions, as in $\mathscr{L}_{\omega_1\omega}$, since such operations preserve measurability (due to countable additivity of probability measures). Expressed in terms of admissible sets, one finds the appropriate forms of completeness and compactness results. Interestingly, there is also an analogue of the Robinson consistency property, which fails for $\mathscr{L}_{\omega_1\omega}$. This chapter should be read after reading the relevant sections of Chapter VIII.

In his retiring address as president of the Association for Symbolic Logic in 1972, Abraham Robinson (Robinson [1973]) asked what logic for topological structures was the analogue of first-order logic for algebraic structures. Chapter XV presents the work that has gone into this problem. Obviously the structures to be considered are of the form $(\mathfrak{M}, \tau)$, where τ is a topology on the domain of the first-order structure $\mathfrak{M}$. Examples include topological space, topological groups, and topological fields. It has taken a lot of effort to arrive at what appears to be the right answer to Robinson's question. The chapter begins by describing three of the logics for such structures that have been studied: $\mathscr{L}(I^n)$, logic with the interior operator, $\mathscr{L}^{\text{mon}}$, a version of monadic second-order logic but where the set quantification is taken to be only over open sets, and a sublogic of this, $\mathscr{L}^t$, where such second-order quantifiers are restricted in a certain way. The logic $\mathscr{L}^t$ is stronger than $\mathscr{L}(I^n)$ but weaker than $\mathscr{L}^{\text{mon}}$. Chapter XV presents results and arguments to support the claim that $\mathscr{L}^t$ is the solution to Robinson's problem by being the "right" analogue of first-order logic for topological logic. Unlike $\mathscr{L}^{\text{mon}}$, $\mathscr{L}^t$ (and *a fortiori*, its sublogic $\mathscr{L}(I^n)$) is compact, has the Löwenheim–Skolem property and has a completeness theorem. However, unlike $\mathscr{L}(I^n)$, $\mathscr{L}^t$ allows one to express continuity, surely a desirable property for the logic of topology.

The logic $\mathscr{L}^t$ also satisfies the interpolation property, a result which leads to a persuasive analogue of Lindstrom's theorem: *$\mathscr{L}^t$ is the strongest logic for topological structures which is compact and has the Löwenheim–Skolem property.* The chapter concludes with some applications of the theory to specific topological theories, including the theory of abelian Hausdorff groups, the theory of the complex numbers as a topological field and topological vector spaces. (The reader may find this chapter is rather dense, but it repays study.)

Chapter XVI presents some previously unpublished work on the logic of Borel structures, due largely to Harvey Friedman. Friedman's basic idea is that while there are some very pathological sets and relations of real numbers, the collection of Borel sets and relations is much better behaved. Why not restrict attention to structures on the reals that are Borel and study the resulting logic? A *Borel structure* is one whose domain is a Borel subset of R and whose relations and functions are all Borel. Given a logic $\mathscr{L}$, a structure is *totally Borel* for $\mathscr{L}$ if all relations definable using $\mathscr{L}$-formulas are Borel.

Thus, whereas an essential feature of the other logics discussed in this part is the structures they consider are richer, the logics studied here are richer in that their structures are constrained to be totally Borel. The chapter applies the notion to two different logics, $\mathscr{L}(Q, Q_m)$ and $\mathscr{L}(Q, Q_c)$ where Q is "there exist uncountably many," Q_m is "there exist a set not of measure 0" and Q_c is "there is a set which is not meager." For example,

$$Q_m x Q_m y \phi(x, y) \leftrightarrow Q_m y Q_m x \phi(x, y)$$

expresses a version of the Fubini theorem, which is true of all totally Borel structures for $\mathscr{L}(Q_m)$. The main results of Chapter XVI are abstract and concrete completeness theorems for the logics $\mathscr{L}(Q, Q_m)$ and $\mathscr{L}(Q, Q_c)$ relative to the collection of totally Borel structures. These logics are less well known but seem very interesting in their potential applications and because they represent a really different direction in the study of extended logics.

Chapter XIV
Probability Quantifiers

by H. J. KEISLER

In this chapter we develop logics appropriate for probability structures, these being first-order structures endowed with a probability measure on the universe. We consider logics having the property that in every probability structure, every definable set is measurable. The price for this is high: The logics do not have the ordinary quantifiers $\forall x$ and $\exists x$. Instead, they have probability quantifiers and countable conjuctions. The main probability logic $L_{\mathbb{A}P}$ satisfies the Barwise completeness and compactness theorems, but does not satisfy finitary compactness. In spite of this, however, this logic does possess the Robinson consistency property. And it also has model-theoretic properties with no first-order analog, such as the law of large numbers, a principle that is presented in Section 3. In Section 4 we will study logics for richer structures with conditional expectations. This development will lead to a model theory which is closely tied to current research in stochastic processes and which has applications to stochastic differential equations.

1. Logic with Probability Quantifiers

In this section we will introduce the logic $L_{\mathbb{A}P}$, which is quite similar to the infinitary logic $L_{\mathbb{A}}$ except that instead of the ordinary quantifiers $(\forall x)$ and $(\exists x)$, the logic $L_{\mathbb{A}P}$ possesses the probability quantifiers $(Px \geq r)$. A structure for this logic is a first-order structure with a (countably additive) probability measure on the universe, such that each relation is measurable. The formula

$$(Px \geq r)\varphi(x)$$

means that the set $\{x \mid \varphi(x)\}$ has probability at least r. Axioms and rules of inference appropriate to our investigation will be presented in this section. The following sections will then examine the subject in more detail.

1.1. Syntax

1.1.1 Convention. We will assume throughout this chapter that $\mathbb{A}$ is an admissible set (possibly with urelements) such that $\omega \in \mathbb{A}$, and each $a \in \mathbb{A}$ is countable (that is, $\mathbb{A} \subseteq \mathrm{HC}$, where HC is the set of hereditarily countable sets).

We refer the reader to Chapter VIII of this volume for a detailed treatment of admissible sets and the infinitary logic $L_{\mathbb{A}}$. Briefly, however, we note that the set of formulas of $L_{\mathbb{A}}$ is the set of all expressions in $\mathbb{A}$ that are built from atomic formulas using negation $\neg$, finite or infinite conjuction, and the quantifier $(\forall x)$.

1.1.2 Definition. We will assume throughout our exposition that L is a countable $\mathbb{A}$-recursive set of finitary relation and constant symbols (no function symbols). The logic $L_{\mathbb{A}P}$ has the following logical symbols:

(a) A countable list of individual variables v_n, for $n \in \mathbb{N}$.
(b) The connectives $\neg$ and $\bigwedge$.
(c) The quantifiers $(P\bar{x} \geq r)$, where $\bar{x} = \langle x_1, \ldots, x_n \rangle$ is a tuple of distinct variables and $r \in \mathbb{A} \cap [0, 1]$.
(d) The equality symbol $=$ (optional).

1.1.3 Definition. The set of *formulas* of $L_{\mathbb{A}P}$ is the least set such that:

(a) Each atomic formula of first-order logic is a formula of $L_{\mathbb{A}P}$.
(b) If φ is a formula of $L_{\mathbb{A}P}$, then $\neg\varphi$ is a formula of $L_{\mathbb{A}P}$.
(c) If $\Phi \in \mathbb{A}$ is a set of formulas of $L_{\mathbb{A}P}$ with only finitely many free variables, then $\Lambda\Phi$ is a formula of $L_{\mathbb{A}P}$;
(d) If φ is a formula of $L_{\mathbb{A}P}$ and $(P\bar{x} \geq r)$ is a quantifier of $L_{\mathbb{A}P}$, then $(P\bar{x} \geq r)\varphi$ is a formula of $L_{\mathbb{A}P}$.

It is understood that the formulas are constructed set theoretically so that $L_{\mathbb{A}P} \subseteq \mathbb{A}$. We denote $L_{\mathbb{A}P}$ where $\mathbb{A} = \mathrm{HC}$ by $L_{\omega_1 P}$. Thus,

$$L_{\mathbb{A}P} = \mathbb{A} \cap L_{\omega_1 P}.$$

The notions of free and bound variables are defined as usual, with the quantifier $(P\bar{x} \geq r)$ binding all the variables in the tuple $\bar{x}$.

The equality relation plays only a minor role in the logic $L_{\mathbb{A}P}$, a fact which stems from the absence of the universal quantifier and of function symbols.

1.1.4 Definition. It is convenient ot use the following abbreviations in $L_{\mathbb{A}P}$:

(i) $(P\bar{x} < r)\varphi$ for $\neg\,(P\bar{x} \geq r)\varphi$.
(ii) $(P\bar{x} \leq r)\varphi$ for $(P\bar{x} \geq 1 - r)\neg\,\varphi$.
(iii) $(P\bar{x} > r)\varphi$ for $\neg\,(P\bar{x} \geq 1 - r)\neg\,\varphi$.
(iv) $\bigvee_{\varphi \in \Phi} \varphi$ for $\neg \bigwedge_{\varphi \in \Phi} \neg\,\varphi$.
(v) The finitary connectives $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$ are defined as usual.

The quantifier $(P\bar{x} \geq 1)$ is a weak analog of $(\forall\bar{x})$, while $(P\bar{x} > 0)$ is a strong analog of $(\exists\bar{x})$. In principle, it would be possible to make do with the one-variable probability quantifier $(Px \geq r)$ alone and introduce the n-variable quantifier $(P\bar{x} \geq r)$ as an abbreviation. However, this abbreviation would be quite complicated, and it is simpler to include $(P\bar{x} \geq r)$ explicitly in the language.

1.2. Probability Models

We will begin with some very basic notions from probability theory. First, a *finitely additive probability space* is a triple $\langle M, S, \mu \rangle$ where S is a field of subsets of M, $\mu: S \to [0, 1]$, $\mu(M) = 1$, and for $X, Y \in S$,

$$\mu(X \cup Y) = \mu(X - Y) + \mu(Y - X) + \mu(X \cap Y).$$

The sets $X \in S$ are *μ-measurable*, and μ is called a *finitely additive probability measure* on M. Next, we say that $\langle M, S, \mu \rangle$ is a *probability space* if, in addition, S is a σ-field and μ is countably additive; that is, whenever $X_0 \subseteq X_1 \subseteq \cdots$ in S, then

$$\mu\left(\bigcup_n X_n\right) = \lim_n \mu(X_n).$$

In this case, μ is said to be a *probability measure* on M. We emphasize that "probability measure" without an adjective will mean "countably additive probability measure."

A set X is said to be a *null set* of μ if there is a $Y \supseteq X$ with $\mu(Y) = 0$. The *product* of two probability spaces $\langle M, S, \mu \rangle$ and $\langle N, T, \nu \rangle$ is the probability space

$$\langle M \times N, S \otimes T, \mu \otimes \nu \rangle,$$

where $S \otimes T$ is the σ-algebra generated by the set of measurable rectangles $X \times Y$, with $X \in S$, $Y \in T$ and where

$$(\mu \otimes \nu)(X \times Y) = \mu(X) \cdot \nu(Y).$$

The n-fold product space is denoted by (M^n, S^n, μ^n).

In general, the diagonal set

$$\{\langle x, x \rangle : x \in M\}$$

is not μ^2-measurable. However, if each singleton is measurable, then there is a canonical way to extend the product measure to the diagonal. In the case that every singleton has measure zero, the diagonal is given measure zero. In general, however, only countably many singletons have positive measure, and the measure of the diagonal is the sum of the squares of the measures of the singletons.

1.2.1 Definition. Let $\langle M, S, \mu \rangle$ be a probability space such that each singleton is measurable. Then, for each $n \in \mathbb{N}$, we have that $\langle M^n, S^{(n)}, \mu^{(n)} \rangle$ is the *probability space* such that $S^{(n)}$ is the σ-algebra generated by the measurable rectangles and the diagonal sets

$$D_{ij} = \{\bar{x} \in M^n : x_i = x_j\},$$

and $\mu^{(n)}$ is the unique extension of μ^n to $S^{(n)}$ such that

$$\mu^{(n)}(D_{ij}) = \sum_{x \in M} \mu(\{x\})^2.$$

In the sequel, we will use $A \Delta B$ for the symmetric difference of the sets A and B. The above ideas clear, we will now consider

1.2.2 Proposition. *If $\langle M, S, \mu \rangle$ is a probability space such that each singleton is measurable, then the measure $\mu^{(n)}$ on $S^{(n)}$ given in Definition* 1.2.1 *exists and is unique. Moreover, for each set $X \in S^{(n)}$ there is a μ^n-measurable set U such that $\mu^{(n)}(X \Delta U)$* $= 0$.

Proof. We will give the proof in the case $n = 2$. Here, $S^{(2)}$ is the set of all sets $X \subseteq M^2$ of the form

$$X = (Y \cap D_{12}) \cup (Z - D_{12}), \qquad Y, Z \in S^2.$$

Let $\nu: S^{(2)} \to [0, 1]$ be defined by

$$\nu(X) = \sum_{\langle x, x \rangle \in Y} \mu(\{x\})^2 + \mu^2(Z) - \sum_{\langle x, x \rangle \in Z} \mu(\{x\})^2.$$

Then $\nu = \mu^{(2)}$ is the unique countably additive probability measure on $S^{(2)}$ which extends μ^2 and satisfies

$$(1) \qquad \nu(D_{12}) = \sum_{x \in M} \mu(\{x\})^2.$$

Finally, let $E = \{\langle x, x \rangle \in D_{12} : \mu(\{x\}) > 0\}$ and let $U = (Y \cap E) \cup (Z - E)$. Since E is countable and each singleton is measurable, E and hence U are μ^2-measurable. Also, $\nu(D_{12} - E) = 0$, and $X \Delta U \subseteq D_{12} - E$, whence, we have that $\nu(X \Delta U) = 0$. ☐

We are now ready to define a probability structure for L, where L is a set of n_i-placed relation symbols R_i, for $i \in I$, and constant symbols c_j, for $j \in J$.

1.2.3 Definition. A *probability structure for* L is a structure

$$\mathscr{M} = \langle M, R_i^{\mathscr{M}}, c_j^{\mathscr{M}}, \mu \rangle_{i \in I, j \in J},$$

where μ is a (countably additive) probability measure on M such that each singleton is measurable, each $R_i^{\mathscr{M}}$ is $\mu^{(n_i)}$-measurable, and each $c_j^{\mathscr{M}} \in M$.

1.2.4 Theorem. Let $\mathscr{M}$ be a probability structure for L. The *satisfaction relation* $\mathscr{M} \models \varphi[\bar{a}]$, for $\varphi(\bar{x}) \in L_{\mathbb{A}P}$ and $\bar{a}$ in M, is defined recursively exactly as for $L_{\mathbb{A}}$ except for the following quantifier clause:

$$\mathscr{M} \models (P\bar{y} \geq r)\varphi(\bar{x}, \bar{y})[\bar{a}] \quad \text{iff} \quad \{\bar{b} \in M^n : \mathscr{M} \models \varphi[\bar{a}, \bar{b}]\}$$
$$\text{is } \mu^{(n)}\text{-measurable and has measure at least } r.$$

Moreover, $\mathscr{M}$ is a *model* of a sentence φ if $\mathscr{M} \models \varphi$.

1.2.5 Theorem. *For each probability structure $\mathscr{M}$, formula $\varphi(\bar{x}, \bar{y}) \in L_{\mathbb{A}P}$, and tuple $\bar{a}$ in M, the set $\{\bar{b} \in M^n : \mathscr{M} \models \varphi[\bar{a}, \bar{b}]\}$ is $\mu^{(n)}$-measurable.* □

This theorem is needed to show that the satisfaction relation has the intended meaning for $L_{\mathbb{A}P}$, and its proof follows easily by induction from a "diagonal" form of the Fubini theorem. A function $f: M \to \mathbb{R}$ is *μ-measurable* if $f^{-1}(-\infty, r]$ is μ-measurable for each $r \in \mathbb{R}$.

1.2.6 Fubini Theorem. *Let μ be a probability measure such that every singleton is measurable, and let $B \subseteq M^{m+n}$ be $\mu^{(m+n)}$-measurable. Then*

(i) *Each section $B_{\bar{x}} = \{\bar{y} \in M^n : \bar{x}\bar{y} \in B\}$ is $\mu^{(n)}$-measurable.*
(ii) *The function $f(\bar{x}) = \mu^{(n)}(B_{\bar{x}})$ is $\mu^{(m)}$-measurable.*
(iii) *We have $\mu^{(m+n)}(B) = \int f(\bar{x})\, d\mu^{(m)}$.* □

The proof here is exactly like the proof of the usual Fubini theorem for product measures. Theorem 1.2.5 would fail if we were to include both the universal quantifier and the probability quantifiers in the language, because projections of measurable sets need not themselves be measurable.

The model-theoretic notions of isomorphism, $L_{\mathbb{A}P}$-equivalence, and $L_{\mathbb{A}P}$-elementary substructure are defined as one would expect, and are respectively written as $\mathscr{M} \cong \mathscr{N}$, $\mathscr{M} \equiv_{\mathbb{A}P} \mathscr{N}$, and $\mathscr{M} \prec_{\mathbb{A}P} \mathscr{N}$.

1.3. Examples

The following examples of sentences of $L_{\mathbb{A}P}$ indicate the expressive power of the language.

(1) "There is a countable set of measure one" is expressed by:

$$(Px \geq 1)(Py > 0)x = y.$$

(2) "There are no point masses" (that is, there are no singletons of positive measure) is expressed by:

$$(Px \geq 1)(Py \geq 1)x \neq y.$$

Every model of this last sentence is uncountable. In the class of structures with no point masses, every sentence of $L_{\mathbb{A}P}$ with equality is equivalent to the sentence without equality that is obtained by replacing $v_n = v_n$ by "true," and $v_m = v_n$ by "false" if $m \neq n$.

(3) The reader can check that no two of the sentences

$$(Px \geq \tfrac{1}{2})(Py \geq \tfrac{1}{2})R(x, y),$$

$$(Py \geq \tfrac{1}{2})(Px \geq \tfrac{1}{2})R(x, y);$$

and

$$P(xy \geq \tfrac{1}{4})R(x, y),$$

are equivalent. (Consider structures with three elements of measure $\frac{1}{3}$.)

A measurable function $X: M \to \mathbb{R}$ is sometimes called a *random variable*. By the Fubini theorem, each binary relation $R(x, y)$ in a probability structure $\mathscr{M}$ induces the random variable $X(u) = \mu\{v \mid R(u, v)\}$. In the following examples, let the language L have binary relation symbols R, R_n, $n \in \mathbb{N}$, and denote the corresponding random variables by X, X_n, $n \in \mathbb{N}$.

(4) The condition $X(u) \geq r$ is expressed by:

$$(Pv \geq r)R(u, v).$$

(5) $|X_1(u) - X_2(u)| \leq r$ is expressed by:

$$\bigwedge_{q \in \mathbb{Q}} (X_1(u) \geq q \to X_2(u) \geq q - r) \wedge (X_2(u) \geq q \to X_1(u) \geq q - r).$$

(6) $X_n \to X$ almost surely (a.s.) is expressed by:

$$(Pu \geq 1) \bigwedge_n \bigvee_m \bigwedge_{k \geq m} |X_k(u) - X(u)| \leq \frac{1}{n}.$$

(7) $X_n \to X$ in probability is expressed by:

$$\bigwedge_n \bigvee_m \bigwedge_{k \geq m} \left(Pu \geq 1 - \frac{1}{n}\right) |X_k(u) - X(u)| \leq \frac{1}{n}.$$

(8) X_1 and X_2 have the same distribution is expressed by

$$\bigwedge_{q \in \mathbb{Q}} \bigwedge_{r \in \mathbb{Q}} (Pu \geq r)(X_1(u) \geq q) \leftrightarrow (Pu \geq r)(X_2(u) \geq q)$$

(9) X_1 and X_2 are independent is expressed by

$$\bigwedge_{q, r \in \mathbb{Q}} \bigwedge_{a, b \in \mathbb{Q}} (Pu \geq a)X_1(u) \geq q \wedge (Pu \geq b)X_2(u) \geq r \to (Pu \geq ab)(X_1(u) \geq q \wedge X_2(u) \geq r),$$

and similarly with $(Pu \leq)$ in place of $(Pu \geq)$.

(10) $1/X(u)$ is integrable is expressed by

$$\neg \bigwedge_{m \in \mathbb{N}} \bigvee_{\substack{s_1 + \cdots + s_n \geq m \\ s_i \in \mathbb{Q}}} \bigwedge_{k=1}^{n} (Pu \geq s_k)|X(u)| \leq \frac{1}{k}.$$

1.4. Proof Theory

$L_{\mathbb{A}P}$ has the following set of axioms, where $\varphi \in L_{\mathbb{A}P}$ and $r, s \in \mathbb{A} \cap [0, 1]$. All but the last axiom B4 are in Hoover [1978a, b].

1.4.1 Definition. The Axioms for weak $L_{\mathbb{A}P}$ are as follows:

A1. All axioms of $L_{\mathbb{A}}$ without quantifiers.
A2. Monotonicity:

$$(P\bar{x} \geq r)\varphi \rightarrow (P\bar{x} \geq s)\varphi, \quad \text{where} \quad r \geq s.$$

A3. $(P\bar{x} \geq r)\varphi(\bar{x}) \rightarrow (P\bar{y} \geq r)\varphi(\bar{y})$.
A4. $(P\bar{x} \geq 0)\varphi$.
A5. Finite additivity:
(i) $(P\bar{x} \leq r)\varphi \wedge (P\bar{x} \leq s)\psi \rightarrow ((P\bar{x} \leq r + s)(\varphi \vee \psi))$;
(ii) $(P\bar{x} \geq r)\varphi \wedge (P\bar{x} \geq s)\psi \wedge (P\bar{x} \leq 0)(\varphi \wedge \psi) \rightarrow (Px \geq r + s)(\varphi \vee \psi)$.
A6. The Archimedean property:

$$(P\bar{x} > r)\varphi \leftrightarrow \bigvee_{n \in \mathbb{N}} \left(P\bar{x} \geq r + \frac{1}{n}\right)\varphi$$

1.4.2 Definition. The axioms for (full) $L_{\mathbb{A}P}$ consist of the axioms for weak $L_{\mathbb{A}P}$ plus:

B1. Countable additivity:

$$\bigwedge_{\Psi \subseteq \Phi} (P\bar{x} \geq r) \bigwedge \Psi \rightarrow (P\bar{x} \geq r) \bigwedge \Phi,$$

where Ψ ranges over the finite subsets of Φ.
B2. Symmetry:

$$(Px_1 \cdots x_n \geq r)\varphi \leftrightarrow (Px_{\pi 1} \cdots x_{\pi n} \geq r)\varphi,$$

where π is a permutation of $\{1, \ldots, n\}$.
B3. Product independence:

$$(P\bar{x} \geq r)(P\bar{y} \geq s)\varphi \rightarrow (P\bar{x}\bar{y} \geq rs)\varphi,$$

provided all variables in $\bar{x}, \bar{y}$ are distinct.

B4. Product measurability: For each $r < 1$,

$$(P\bar{x} \geq 1)(P\bar{y} > 0)(P\bar{z} \geq r)(\varphi(\bar{x}\bar{z}) \leftrightarrow \varphi(\bar{y}\bar{z})),$$

provided all variables in $\bar{x}$, $\bar{y}$, $\bar{z}$ are distinct.

The central purpose of Axiom B4 is to guarantee that $\varphi(\bar{x}, \bar{y})$ can be approximated by a finite union of measurable rectangles. It is obviously valid if $\varphi(\bar{x}, \bar{z})$ is a "rectangle" $\psi(\bar{x}) \wedge \theta(\bar{z})$. We will see later on that it is valid in general (the Soundness Theorem).

1.4.3 Definition. The Rules of Inference for $L_{\mathbb{A}P}$ are as follows:

R1. *Modus Ponens*:

$$\varphi, \varphi \rightarrow \psi \vdash \psi.$$

R2. Conjunction:

$$\{\varphi \rightarrow \psi \mid \psi \in \Psi\} \vdash \varphi \rightarrow \bigwedge \Psi.$$

R3. Generalization:

$$\varphi \rightarrow \psi(\bar{x}) \vdash \varphi \rightarrow (P\bar{x} \geq 1)\psi(\bar{x}),$$

provided $\bar{x}$ is not free in φ.

1.4.4 Definition. The notion of a *deduction* of a formula ψ from a set of sentences Φ, and the expressions

$$\Phi \vdash \psi, \quad \vdash \psi, \quad \Phi \models \psi, \quad \models \psi$$

are defined in the usual way. A *theorem* of $L_{\mathbb{A}P}$ is a sentence ψ such that $\vdash \psi$.

1.4.5 Deduction Theorem. *In either $L_{\mathbb{A}P}$ or weak $L_{\mathbb{A}P}$, if ψ is a sentence and $\Phi \cup \{\psi\} \vdash \theta$, then $\Phi \vdash \psi \rightarrow \theta$.*

1.4.6 Proposition. *The following are theorems of $L_{\mathbb{A}P}$, and their proofs do not require use of Axiom* B4:

(i) $(P\bar{x} \leq 1)\varphi$.
(ii) $(P\bar{x} > r) \bigvee \Phi \leftrightarrow \bigvee_{\Psi \subseteq \Phi} (P\bar{x} > r) \bigvee \Psi$, *where Ψ is finite.*
(iii) $(P\bar{x} \geq r)\varphi \leftrightarrow \bigwedge_n (Px \geq r - 1/n)\varphi$.
(iv) $(P\bar{x} \leq a)\varphi(\bar{x}) \wedge (P\bar{y} \leq b)\psi(\bar{y}) \rightarrow (P\bar{x}\bar{y} \leq ab)(\varphi(\bar{x}) \wedge \psi(\bar{y}))$.
(v) $(P\bar{x} \geq r)\varphi(\bar{x}) \rightarrow (P\bar{x}\bar{y} \geq r)\varphi(\bar{x})$.
(vi) $(P\bar{x}\bar{y} \geq a + b - ab)\varphi \rightarrow (P\bar{x} \geq a)(P\bar{y} \geq b)\varphi$.

Taking $a = b = 1$:

(vii) $(P\bar{x} \geq 1)(P\bar{y} \geq 1)\varphi \leftrightarrow (P\bar{x}\bar{y} \geq 1)\varphi$ ☐

1.4.7 Theorem (Soundness Theorem). *Any set Φ of sentences of $L_{\mathbb{A}P}$ which has a model is consistent.*

Outline of Proof. As usual, to prove the soundness theorem it suffices to show that each axiom is valid and the rules of inference preserve validity. The only difficulty lies in checking the validity of the product measurability axiom, (Axiom B4). In view of part (iii) of Proposition 1.4.6, it suffices to show that for each $q, r < 1$,

$$(P\bar{x} \geq q)(P\bar{y} > 0)(P\bar{z} \geq r)(\varphi(\bar{x}\bar{z}) \leftrightarrow \varphi(\bar{y}\bar{z}))$$

is valid. This can be proven by use of the Fubini theorem, Proposition 1.2.2, and the direction (ii) implies (i) of the following lemma. ☐

1.4.8 Lemma. *Let μ, ν, and λ be probability measures on M, N, and $M \times N$ such that $\mu \otimes \nu \subseteq \lambda$. Let U be λ-measurable. The following are equivalent:*

(i) *For every $\varepsilon > 0$, there is a finite union B of $\mu \otimes \nu$-measurable rectangles such that $\lambda(U \,\Delta\, B) < \varepsilon$.*
(ii) *There is a $\mu \otimes \nu$-measurable set C with $\lambda(U \,\Delta\, C) = 0$.*

Idea of Proof. From (i) to (ii), we may take C to be a limit of the B's. We then use the monotone class theorem to show that for each $\mu \otimes \nu$-measurable U, (i) holds. It then follows at once that (ii) implies (i). ☐

Remark. D. Hoover has pointed out the curious fact that the logic $L_{\mathbb{A}P}$ is equivalent to the richer logic on L with ord($\mathbb{A}$) variables, which allows formulas with $\mathbb{A}$-finitely many free variables and quantifiers $(P\bar{x} \geq r)$ over $\mathbb{A}$-finite sequences $\bar{x}$. The axioms and rules are as before with the additional scheme

$$(Px_1x_2 \ldots \geq r)\varphi(x_{i_1} \ldots x_{i_n}) \leftrightarrow (Px_{j_1} \ldots x_{j_n} \geq r)\varphi(x_{j_1} \ldots x_{j_n}),$$

where none of the other x_j's are free in φ. It can be shown by the logical monotone class theorem (Keisler [1977c]) that every sentence of the richer logic is equivalent to a sentence of $L_{\mathbb{A}P}$. The situation is radically different, however, when universal quantifiers are present, since well-ordering is definable in $L_{\omega_1\omega_1}$.

1.5. Weak Models

We will now begin working toward the completeness theorem for $L_{\mathbb{A}P}$. To this end, we first examine

1.5.1 Definition. A *weak structure for $L_{\mathbb{A}P}$* is a structure

$$\mathscr{M} = \langle M, R_i^{\mathscr{M}}, c_j^{\mathscr{M}}, \mu_n \rangle_{i \in I, j \in J, n \in \mathbb{N}}$$

such that each μ_n is a finitely additive probability measure on M^n with each singleton measurable, and (with the natural definition of satisfaction) the set

$$\{\bar{b} \in M^n : \mathscr{M} \models \varphi[\bar{a}, \bar{b}]\}$$

is μ_n-measurable for each $\varphi(\bar{x}, \bar{y}) \in L_{\mathbb{A}P}$ and $\bar{a}$ in M.

By Theorem 1.2.5, every probability structure $\mathscr{M}$ induces a weak structure for $L_{\omega_1 P}$ with $\mu_n = \mu^{(n)}$.

1.5.2 Weak Soundness Theorem. *Let Φ be a set of sentences of $L_{\mathbb{A}P}$. If Φ has a weak model, then Φ is consistent in weak $L_{\mathbb{A}P}$.*

1.5.3 Weak Completeness Theorem (Hoover [1978b]). *Let $\mathbb{A}$ be countable. If Φ is consistent in weak $L_{\mathbb{A}P}$, then Φ has a weak model.*

Sketch of Proof. Let C be a countable set of new constants, and let $K = L \cup C$. By a Henkin construction, Φ can be extended to a maximal weak $K_{\mathbb{A}P}$-consistent set Γ of sentences with the following witness properties:

(1) If $\Phi \subseteq \Gamma$ and $\bigwedge\Phi \in K_{\mathbb{A}P}$, then $\bigwedge\Phi \in \Gamma$;
(2) If $\varphi(\bar{c}) \in \Gamma$ for all $\bar{c}$ in C, then $(P\bar{x} \geq 1)\varphi(\bar{x}) \in \Gamma$.

Let C' be the set of constants of K. Γ induces a first-order structure

$$\mathscr{M}_0 = \langle M, R_j^{\mathscr{M}}, c^{\mathscr{M}} \rangle_{i \in I,\, c \in C'},$$

for K in the usual way, and $M = \{c^{\mathscr{M}} \mid c \in C'\}$. Define μ_n by

$$\mu_n\{\bar{c}^{\mathscr{M}} \mid \varphi(\bar{c}, \bar{d}) \in \Gamma\} = \sup\{r \mid (P\bar{x} \geq r)\varphi(\bar{x}, \bar{d}) \in \Gamma\}$$

for each $\varphi(\bar{x}, \bar{y})$ and $\bar{d}$. Axioms A1 through A5 insure that μ_n is well defined and is a finitely additive probability measure. This gives us a weak structure $\mathscr{M}$. Axiom A6 (in the dual form of Proposition 1.4.6(iii)) insures that the above supremum is always attained, and it follows by induction that $\mathscr{M} \models \Gamma$; and, hence, $\mathscr{M} \models \Phi$. ☐

1.6. Atomic and Countable Models

In this section we will dispose of the degenerate case in which there is a countable set of measure one; that is, we will consider the situation in which

$$(*) \qquad (Px \geq 1)(Py > 0)x = y$$

holds. Notice that the last axiom of $L_{\mathbb{A}P}$, namely Axiom B4, is provable from (*) and the other axioms.

Assume first that L has no constant symbols.

1.6.1 Definition. Let $\mathscr{M}$ be a probability structure. An element $a \in M$ is an *atom* if $\{a\}$ has positive measure. $\mathscr{M}$ is *atomic* if every element is an atom.

We list some easy facts in the next proposition.

1.6.2 Proposition. (i) *If $\mathscr{M}$ is atomic, then $\mathscr{M}$ is countable.*

(ii) *If $\mathscr{M}$ is countable, then $\mathscr{M}$ satisfies* $(*)$.

(iii) *If $\mathscr{M}$ satisfies $(*)$, then there is a unique atomic model $\mathscr{N}$ such that $\mathscr{N} \prec_{\mathbb{A}P} \mathscr{M}$.*

(iv) *If $\mathscr{M}$ and $\mathscr{N}$ are atomic and $L_{\mathbb{A}P}$-equivalent, then they are isomorphic.*

(v) *If $\mathscr{M}$ is atomic, then for every formula $\varphi(x\bar{y})$ of $L_{\mathbb{A}P}$ and $\bar{b}$ in M, we have*

$$\mathscr{M} \models (\forall x)\varphi(x\bar{b}) \leftrightarrow (Px \geq 1)\varphi(x\bar{b});$$

and

$$\mathscr{M} \models (\exists x)\varphi(x\bar{b}) \leftrightarrow (Px > 0)\varphi(x\bar{b}). \quad \square$$

Part (v) of the proposition shows that in atomic structures the ordinary quantifiers can be defined in terms of probability quantifiers.

1.6.3 Theorem (Completeness Theorem for Atomic Models). *A countable set of sentences Φ of $L_{\mathbb{A}P}$ has an atomic model if and only if $\Phi \cup \{(*)\}$ is consistent in $L_{\mathbb{A}P}$.*

Sketch of Proof. We may take $\mathbb{A}$ to be countable. Suppose $\Phi \cup \{(*)\}$ is consistent in $L_{\mathbb{A}P}$. Then it has a weak model $\mathscr{M}_0$ in which all theorems of $L_{\mathbb{A}P}$ hold. From Section 1.4, for each m, the following are deducible from $(*)$ in $L_{\mathbb{A}P}$:

$$(Px \geq 1) \bigvee_n \left(P_1 \geq \frac{1}{n}\right) x = y_y;$$

$$\left(Px > 1 + \frac{1}{m}\right) \bigvee_n \left(Py \geq \frac{1}{n}\right) x = y;$$

and

$$\bigvee_n \left(Px > 1 - \frac{1}{m}\right)\left(Py \geq \frac{1}{n}\right) x = y.$$

It follows that M_0 has a finite subset of μ_1-measure greater than $1 - 1/m$. Thus μ_1 can be extended to a probability measure μ defined on all subsets of M_0 by

$$\mu(X) = \sup\{\mu_1(Y): Y \subseteq X, Y \text{ finite}\},$$

forming a probability structure $\mathscr{M} \equiv_{\mathbb{A}P} \mathscr{M}_0$. The atomic model $\mathscr{N} \prec_{\mathbb{A}P} \mathscr{M}$ is the required model of Φ. $\square$

We now consider the general case where L has constant symbols. Define $\mathscr{M}$ to be *atomic* if every element of $\mathscr{M}$ is either of positive measure or equal to a constant symbol. With this definition, all the results remain true except for part (v) of Proposition 1.6.2. If the set of constant symbols is A-finite, we can still define the ordinary quantifiers in terms of probability quantifiers in an atomic structure $\mathscr{M}$ by

$$\mathscr{M} \models \forall x\varphi(x\bar{b}) \leftrightarrow (Px \geq 1)\varphi(x\bar{b}) \wedge \bigwedge_{j \in J} \varphi(c_j\bar{b});$$

and

$$\mathscr{M} \models \exists x\varphi(x\bar{b}) \leftrightarrow (Px > 0)\varphi(x\bar{b}) \vee \bigvee_{j \in J} \varphi(c_j\bar{b}).$$

2. *Completeness Theorems*

The main result of this section, to be given in Section 2.3, states that the set of axioms given in Section 1 is complete. As a preliminary result, in Section 2.2 we prove a completeness theorem for $L_{\mathbb{A}P}$ without using the axiom B4. However, this is done for a wider class of models. The chief difficulty is the construction of a countably additive probability structure from a finitely additive one. The key to getting past this difficulty is the Loeb measure construction from non-standard analysis, and this we examine in the following discussion.

2.1. The Loeb Measure

We assume once and for all that we have an ω_1-saturated non-standard universe

$$*: \langle V_\omega(U), \in\rangle \to \langle V_\omega(*U), \in\rangle,$$

where U is a set of urelements large enough for our purposes (see Keisler [1976] or Loeb [1979a] for the details). For $r \in {}^*\mathbb{R}$, ${}^\circ r$ denotes the standard part of r. We will briefly state the definition and main facts about the Loeb measure. They are due to Loeb [1975].

2.1.1 Definition. Let M be an internal set in $V_\omega(*U)$ and let $\langle M, S, \mu\rangle$ be an internal $*$-finitely additive probability space. (Thus, μ and S are internal and $\mu: S \to {}^*[0, 1]$.) The *Loeb measure* of μ is the unique (countably additive) probability space $\langle M, \sigma(S), \hat{\mu}\rangle$ such that:

(i) $\sigma(S)$ is the σ-algebra generated by S.
(ii) $\hat{\mu}(X) = {}^\circ\mu(X)$ for all $X \in S$.

2.1.2 Theorem. *The Loeb measure exists and is unique.*

Proof. Use ω_1-saturation and the Caratheodory extension theorem. ☐

2.1.3 Theorem. *Let $X \in \sigma(S)$. Then,*

(i) *for each $n \in \mathbb{N}$, there exist $Y, Z \in S$ such that $Y \subseteq X \subseteq Z$ and $\mu(Z - Y) < 1/n$;*
(ii) *there exists $Y \in S$ such that $\hat{\mu}(X \,\Delta\, Y) = 0$.*

Proof. Part (i) of the theorem uses the monotone class theorem, and part (ii) follows from part (i) by ω_1-saturation. ☐

Intuitively, part (i) says that every Loeb measurable set can be approximated above and below by internal measurable sets.

2.2. Graded Probability Models

A graded probability structure is a generalization of a probability structure in which the diagonal product $\mu^{(n)}$ is replaced by any probability measure on M^n which satisfies the Fubini theorem. We will show that the set of axioms for $L_{\mathbb{A}P}$ without axiom B4 is sound and complete for these structures.

2.2.1 Definition. A *graded probability structure* for L is a structure

$$\mathscr{M} = \langle M, R_j^{\mathscr{M}}, c_j^{\mathscr{M}}, \mu_n \rangle_{i \in I, j \in J, n \in \mathbb{N}}$$

such that:

(a) Each μ_n is a (countably additive) probability measure on M^n.
(b) Each n-placed relation $R_i^{\mathscr{M}}$ is μ_n-measurable, and the identity relation is μ_2-measurable.
(c) If B is μ_m-measurable, then $B \times M^n$ is μ_{m+n}-measurable.
(d) The symmetry property holds; that is, each μ_n is preserved under permutations of $\{1, \ldots, n\}$.
(e) $\langle \mu_n | n \in \mathbb{N} \rangle$ has the *Fubini property*: If B is μ_{m+n}-measurable, then
 (1) For each $\bar{x} \in M^n$, the section $B_{\bar{x}} = \{\bar{y} | B(\bar{x}, \bar{y})\}$ *is* μ_n-measurable.
 (2) The function $f(\bar{x}) = \mu_n(B_{\bar{x}})$ is μ_m-measurable.
 (3) $\int f(\bar{x}) d\mu_m = \mu_{m+n}(B)$.

2.2.2 Proposition. (i) *If $\mathscr{M}$ is a probability structure, then*

$$\langle M, R_i^{\mathscr{M}}, c_j^{\mathscr{M}}, \mu^{(n)} \rangle_{n \in \mathbb{N}}$$

is a graded probability structure.
(ii) *Every graded probability structure is a weak structure for $L_{\omega_1 P}$.* ☐

2.2.3 Proposition. *In a graded probability structure $\mathscr{M}$, μ_n is an extension of $\mu_1^{(n)}$.* □

An important example of a graded probability structure arises from the Loeb measure construction.

2.2.4 Theorem (Keisler [1977b]). *Let M be a $*$-finite set. For each n, let ν_n be the internal probability measure on M^n giving each element the same weight (the counting measure), and let $\mu_n = \hat{\nu}_n$ be the Loeb measure of ν_n. Then $\langle \mu_n | n \in \mathbb{N} \rangle$ has the Fubini property. Hence, if each n-ary relation of $\mathscr{M}$ is μ_n-measurable, $\mathscr{M}$ is a graded probability structure.* □

The following example of Hoover provides a graded probability structure which is not $L_{\mathbb{A}P}$-equivalent to any ordinary probability structure.

2.2.5 Example (White Noise). Let H be an infinite $*$-finite set, let $M = {}^*\mathscr{P}(H)$ be the set of all internal subsets of H, and take μ_n as in Theorem 2.2.4. Let $f: M \to H$ be an internal function partitioning M into H equal parts. Let $R(x, y)$ iff $f(x) \in y$. Then R is internal and hence μ_2-measurable.

If $f(a) \neq f(b)$, then the sets $R(a, v)$ and $R(b, v)$ are independent; similarly, for $f(a_1), \ldots, f(a_n)$ distinct. This suggests the name "white noise." Thus,

$$\mathscr{M} \models (Px \geq 1)(Py \geq 1)(Pz \leq \tfrac{1}{2})R(xz) \leftrightarrow R(yz).$$

But then,

$$\mathscr{M} \models (Px \leq 0)(Py > 0)(Pz > \tfrac{1}{2})R(xz) \leftrightarrow R(yz).$$

Thus axiom B4 fails in $\mathscr{M}$. In fact, R is not measurable in the completion of $\mu_1^{(2)}$.

2.2.6 Definition. By *graded $L_{\mathbb{A}P}$* we mean $L_{\mathbb{A}P}$ with all the axioms except for the product measurability axiom B4.

One may check that all axioms except axiom B4 hold in all graded probability structures.

2.2.7 Theorem (Graded Soundness Theorem). *Every set of sentences of $L_{\mathbb{A}P}$ which has a graded model is consistent in graded $L_{\mathbb{A}P}$.* □

2.2.8 Theorem (Graded Completeness Theorem by Hoover [1978b]). *Every countable set Φ of sentences which is consistent in graded $L_{\mathbb{A}P}$ has a graded model.*

Sketch of Proof. Let $\mathbb{A}$ be countable, and assume L has countably many constants not appearing in Φ. From the proof of the weak completeness theorem, Φ has a weak model

$$\mathscr{M} = \langle M, R_i, c_j, \mu_n \rangle_{i \in I, j \in J, n \in \mathbb{N}}$$

such that $\mathscr{M}$ satisfies each theorem of graded $L_{\mathbb{A}P}$, $M = \{c_j | j \in J\}$, and the domain of each μ_n is the set of $L_{\mathbb{A}P}$-definable subsets of M^n. Form the internal structure

$$^*\mathscr{M} = \langle ^*M, ^*R_j, ^*c_j, ^*\mu_n \rangle_{i \in ^*I, j \in ^*J, n \in ^*\mathbb{N}}.$$

Let

$$\hat{\mathscr{M}} = \langle ^*M, ^*R_j, ^*c_j, \hat{\mu}_n \rangle_{i \in I, j \in J, n \in \mathbb{N}},$$

where $\hat{\mu}_n$ is the Loeb measure of μ_n. By Theorem 2.1.3, every $\hat{\mu}_n$- measurable set can be approximated above and below by $*$-definable sets in n variables. Using this fact and axioms B2 and B3 in $\mathscr{M}$, it can be shown that $\hat{\mathscr{M}}$ is a graded probability structure. An induction on formulas will show that $\hat{\mathscr{M}}$ is $L_{\mathbb{A}P}$-equivalent to $\mathscr{M}$. Axiom B1 is used in the $\wedge$ step, and axioms B2 and B3 in the quantifier step. Therefore, $\hat{\mathscr{M}} \models \Phi$. □

Remark. The graded soundness and completeness theorems hold with little change if L has function symbols, and graded probability structures are defined so that the interpretation of every atomic formula in n variables is μ_n-measurable. This is done in Hoover [1978b].

2.3. *The Main Completeness Result*

We are now ready to prove the completeness theorem for $L_{\mathbb{A}P}$. The results of this section, including the completeness theorem, are new. We make use of axiom B4 by way of the following lemma.

2.3.1 Lemma (Rectangle Approximation Lemma). *Let $\mathscr{M}$ be a graded probability structure satisfying every theorem of $L_{\mathbb{A}P}$. Then for each $\varepsilon > 0$ and formula $\varphi(\bar{y})$ of $L_{\mathbb{A}P}$ there are finitely many formulas $\psi_{ij}(\bar{x}y_j)$, where $i = 1, \ldots, m$, and $j = 1, \ldots, n$, such that*

$$\mathscr{M} \models (P\bar{x} > 0)(P\bar{y} > 1 - \varepsilon)(\varphi(\bar{y}) \leftrightarrow \bigvee_{i=1}^{m} \bigwedge_{j=1}^{n} \psi_{ij}(\bar{x}y_j)). \ \square$$

The lemma asserts that any definable set $\varphi(\bar{y})$ in $\mathscr{M}$ can be approximated within ε by a finite union of definable rectangles, uniformly in parameters $\bar{x}$ from a set of positive measure. The proof is rather technical, and axiom B4 is used n times.

2.3.2 Definition. Let $\mathscr{M}$ and $\mathscr{N}$ be graded probability structures. We say that $\mathscr{M} = \mathscr{N}$ *almost surely*, (in symbols, $\mathscr{M} = \mathscr{N}$ a.s.) if $\mathscr{M}$ and $\mathscr{N}$ have the same universe, constants, and measures, and if for each atomic formula $\varphi(\bar{x})$ of $L_{\mathbb{A}P}$,

$$\mathscr{M} \models \varphi[\bar{a}] \quad \text{iff} \quad \mathscr{N} \models \varphi[\bar{a}]$$

for μ_n-almost all $\bar{a}$.

2.3.3 Lemma. *If $\mathscr{M} = \mathscr{N}$ a.s. then $\mathscr{M}$ and $\mathscr{N}$ are $L_{\mathbb{A}P}$-equivalent. Also for each formula $\varphi(\bar{x})$ of $L_{\mathbb{A}P}$,*

$$\mathscr{M} \models \varphi[\bar{a}] \quad \textit{iff} \quad \mathscr{N} \models \varphi[\bar{a}]$$

for μ_n-almost all $\bar{a}$.

Proof. The proof here is by induction on φ. ☐

The following theorem is the last step needed for the completeness theorem. The proof of this result would break down if we were to allow function symbols in L.

2.3.4 Theorem. *Let $\mathscr{M}$ be a graded probability structure satisfying every theorem of $L_{\mathbb{A}P}$, and let $\mu = \mu_1$. Then there is a graded probability structure $\mathscr{N}$ such that $\mathscr{M} = \mathscr{N}$ a.s., and each relation $R_i^{\mathscr{N}}$ is $\mu^{(n)}$-measurable. Thus, $\mathscr{N}$ induces an ordinary probability structure.*

Sketch of Proof. By the Rectangle Approximation Lemma, for each $\varepsilon > 0$ and $L_{\mathbb{A}P}$-definable set $U \subseteq M^n$ in $\mathscr{M}$, there is a finite union B of μ^n-measurable rectangles such that $\mu_n(B \mathbin{\Delta} U) < \varepsilon$. Then, by Lemma 1.4.8, there is a μ^n-measurable set C such that $\mu_n(C \mathbin{\Delta} U) = 0$. By patching diagonals together, we find that for each $i \in I$, there is a $\mu^{(n)}$-measurable relation $R_i^{\mathscr{N}}$ such that $\mathscr{M} = \mathscr{N}$ a.s. ☐

2.3.5. Theorem (Completeness Theorem for $L_{\mathbb{A}P}$). *Every countable consistent set Φ of sentences of $L_{\mathbb{A}P}$ has a model.*

Proof. The proof of this result is by the Graded Completeness Theorem, Theorem 2.3.4, and Lemma 2.3.3. ☐

By the usual $L_{\mathbb{A}}$ arguments (as given in Chapter IX), we obtain Barwise-type results. Similar results for graded $L_{\mathbb{A}P}$ are given in Keisler [1977b].

2.3.6 Theorem (Barwise Completeness Theorem). *The set of valid sentences of* $L_{\mathbb{A}P}$ is Σ *on* $\mathbb{A}$.

2.3.7 Theorem (Barwise Compactness Theorem). *Let $\mathbb{A}$ be countable and let Φ be a set of sentences of $L_{\mathbb{A}P}$. If Φ is Σ on $\mathbb{A}$ and every $\mathbb{A}$-finite $\Psi \subseteq \Phi$ has a model, then Φ has a model.*

2.4. Hanf and Löwenheim Numbers

We have seen in Section 1 that the sentence stating that $\mathscr{M}$ is atomless has no countable models. Thus, the Löwenheim number of $L_{\mathbb{A}P}$ is at least ω_1. On the other hand, given any probability structure $\mathscr{M}$, we can obtain $L_{\omega_1 P}$-equivalent structures of arbitrarily large cardinality by adding a set of new elements of measure zero. Thus, the Hanf number is ω but for a trivial reason. When considering cardinalities, we should restrict our attention to *reasonable* structures.

2.4.1 Definition. A probability structure $\mathscr{M}$ is *reasonable* if every set of measure one has the same cardinal as M. The *reasonable* Löwenheim or Hanf number of $L_{\mathbb{A}P}$ is obtained by restricting to reasonable probability structures.

2.4.2 Proposition. *A reasonable structure is countable if and only if the set of atoms has measure one.* $\square$

Let $\mathscr{M}$ and $\mathscr{N}$ be probability structures for L. Notice that if $\mathscr{M} \prec_{\mathbb{A}P} \mathscr{N}$, then the first-order part of $\mathscr{M}$ is a substructure of the first-order part of $\mathscr{N}$ but is not necessarily an elementary substructure in the sense of $L_{\omega\omega}$.

2.4.3 Proposition. *Every probability structure $\mathscr{M}$ has a reasonable $L_{\mathbb{A}P}$-elementary substructure $\mathscr{N}$ such that $\mu(N) = 1$ and ν is the restriction of μ to N. The cardinal of N is unique.* $\square$

The following theorem and corollary are new.

2.4.4 Theorem (Downward Löwenheim–Skolem Theorem). *Let $\mathscr{M}$ be a reasonable probability structure of power at least λ, where $\lambda = \lambda^{\omega}$. Then, for every set $X \subseteq M$ of power $\leq \lambda$, $\mathscr{M}$ has a reasonable $L_{\omega_1 P}$-elementary substructure $\mathscr{N}$ of power λ with $X \subseteq N$.*

Proof. Let $X \subseteq X_0 \subseteq M$ where X_0 has power λ and contains all constants. Form a chain X_α, $\alpha < \lambda$ of subsets of M of power λ such that for every Borel combination B of sets $L_{\omega_1 P}$-definable in $\mathscr{M}$ with parameters in X_α,

(1) if $B \neq \varnothing$ then $B \cap X_{\alpha+1} \neq \varnothing$;
(2) if $\mu(B) = 1$ then $|B \cap X_{\alpha+1}| = \lambda$.

Take unions at limit α. Form the structure $\mathscr{N}$, with $N = \bigcup_\alpha X_\alpha$ and $\nu(B \cap N) = \mu(B)$, for each Borel combination B of sets $L_{\omega_1 P}$-definable with parameters in N. Then $\mathscr{N}$ is as required. $\square$

2.4.5 Corollary. *Let λ be the reasonable Löwenheim number for $L_{\mathbb{A}P}$. Then,*

(i) $\omega_1 \leq \lambda \leq 2^{\omega}$;
(ii) *Martin's axiom implies $\lambda = 2^{\omega}$.*

Proof. As to the argument for (i), we note that Theorem 2.4.4 shows that $\lambda \leq 2^{\omega}$. In order to prove (ii), we let φ say that $R_n(x)$, $n \in \mathbb{N}$ are independent sets of probability $\frac{1}{2}$. By Martin's axiom, every subset of $2^{\mathbb{N}}$ of power $< 2^{\omega}$ has Lebesgue measure zero, and it thus follows that φ has no reasonable model of power $< 2^{\omega}$. $\square$

2.4.6 Theorem (Hoover [1978b]). *Every uncountable reasonable probability structure $\mathscr{M}$ has reasonable $L_{\omega_1 P}$-elementary extensions of arbitrarily large cardinality.*

Sketch of Proof. Working in a κ^+-saturated universe, form ${}^*\mathscr{M}$ and use the Loeb process to get a graded structure $\hat{\mathscr{M}} \succ_{\omega_1 P} \mathscr{M}$ and probability structure $\mathscr{N} \succ_{\omega_1 P} \mathscr{M}$.

By Proposition 2.4.2, $\mathscr{M}$ has an atomless set of positive measure ε. By κ^+-saturation, every internal set in $^*M$ of measure $> 1 - \varepsilon/2$ has power at least κ^+, so every Loeb measurable set of measure one has power $\geq \kappa^+$. □

2.4.7 Corollary. *$L_{\mathbb{A}P}$ has reasonable Hanf number ω_1.* □

2.5. Random Variables

In this section we will consider structures with random variables instead of relations. From the examples of Section 3.1 we saw that structures with random variables are of interest in probability theory. In general, one could consider random variables with values in a Polish space. We will restrict our discussion here to random variables with values in $\mathbb{R}$ and will use a language $L = \{X_i, c_j | i \in I, j \in J\}$.

2.5.1 Definition. An *n-fold random variable* on a probability space $\langle M, S, \mu\rangle$ is a $\mu^{(n)}$-measurable function $X: M^n \to \mathbb{R}$. A *random variable structure for L* is a structure

$$\mathscr{M} = \langle M, X_i^{\mathscr{M}}, c_j^{\mathscr{M}}, \mu\rangle_{i \in I, j \in J},$$

where μ is a probability measure on M, $X_i^{\mathscr{M}}$ is an n_i-fold random variable, and $c_j \in M$, and each

2.5.2 Definition. The *auxiliary language* of L is the language L' which has the same constant symbols c_j of L but has new relation symbols $[X_i(\bar{u}) \geq r]$, and $[X_i(\bar{u}) \leq r]$, for each $i \in I$ and $r \in \mathbb{Q}$.

Each random variable structure $\mathscr{M}$ for L induces a probability structure $\mathscr{M}'$ for L', where $[X_j(\bar{u}) \geq r]$ is interpreted in the natural way.

2.5.3 Definition. We will use the following abbreviations:

$$[X(\bar{u}) > r] \quad \text{for} \quad \neg[X(\bar{u}) \geq r],$$

$$[X(\bar{u}) < r] \quad \text{for} \quad \neg[X(\bar{u}) \geq r].$$

2.5.4 Definition. The *language* $L_{\mathbb{A}P}(\mathbb{R})$ has the same set of formulas as $L'_{\mathbb{A}P}$. It has all the axioms and rules of inference of $L'_{\mathbb{A}P}$, plus four new axioms, where $r, s \in \mathbb{Q}$:

C1. $[X(\bar{u}) \geq r] \to [X(\bar{u}) \geq s]$, where $r \geq s$;
C2. $[X_i(\bar{u}) > r] \leftrightarrow \bigvee_n [X_i(\bar{u}) \geq r + 1/n]$;
C3. $[X_i(\bar{u}) \geq r] \leftrightarrow \bigwedge_n [X_i(\bar{u}) \geq r - 1/n]$;
C4. $\bigvee_n ([X_i(\bar{u}) \geq -n] \wedge [X_i(\bar{u}) \leq n])$, and each singleton is measurable.

2.5.5 Theorem (Soundness and Completeness Theorem for $L_{\mathbb{A}P}(\mathbb{R})$). *A countable set Φ of sentences of $L_{\mathbb{A}P}(\mathbb{R})$ has a random variable model if and only if it is consistent in $L_{\mathbb{A}P}(\mathbb{R})$.*

Proof. The soundness is easy. Suppose Φ is consistent. Let Ψ be the set of sentences of the form $(P\bar{v} \geq 1)\psi$, where ψ is one of the axioms C_1 through C_4. Then, $\Phi \cup \Psi$ is consistent and has a probability model $\mathscr{M}'$. Make $\mathscr{M}'$ into a random variable model $\mathscr{M}$ by defining

$$X^{\mathscr{M}}(\bar{a}) = \sup\{r \in \mathbb{Q} \mid \mathscr{M}' \models [X(\bar{a}) \geq r]\}.$$

Use Ψ to show that $X^{\mathscr{M}}$ is almost surely finite and uniquely defined. ☐

2.6. Finitary Probability Logic

We will now discuss the situation when ω is not an element of $\mathbb{A}$, so that each formula of $L_{\mathbb{A}P}$ is finite. We will assume that the rationals are defined in such a way that $\mathbb{Q} \subseteq \mathbb{A}$, so $L_{\mathbb{A}P}$ has at least the quantifiers $(P\bar{x} \geq r)$, $r \in \mathbb{Q} \cap [0, 1]$. By throwing additional reals into $\mathbb{A}$ as urelements, we can obtain more probability quantifiers. When $\omega \notin \mathbb{A}$, the infinitary axiom B1 and the infinite conjunction rule R2 become finite. However, the other infinitary axiom A6 is outside the language $L_{\mathbb{A}P}$ and must be replaced by a new infinite rule of inference, a rule which is due to Hoover [1978a].

2.6.1 Definition. The rule of inference for finitary $L_{\mathbb{A}P}$ is given by

$$\{\psi \to (P\bar{x} \geq r)(P\bar{y} \geq s - 1/n)\varphi \mid n \in \mathbb{N}\} \vdash \psi \to (P\bar{x} \geq r)(P\bar{y} \geq s)\varphi.$$

With this new rule of inference, the weak, graded, and full completeness theorems hold for the finitary case $\omega \notin \mathbb{A}$. Hoover [1978b] has shown that when $\mathbb{A} = \mathbb{HF}$, the set of valid sentences of $L_{\mathbb{A}P}$ is complete Π^1_1 and thus not recursively enumerable. This was done by interpreting the standard model of number theory in a finite theory of $L_{\mathbb{A}P}$. The compactness theorem fails for $L_{\mathbb{A}P}$, so that some infinitary rule of inference is needed.

2.6.2 Example. Let Φ be the set of sentences containing $(Px > 0)R(x)$, and $(Px \leq 1/n)R(x)$, for $n = 1, 2, \ldots$. Then every finite subset of Φ has a model, but Φ itself has no model.

However, there is a compactness theorem for special sentences, which we will state for $L_{\omega_1 P}$.

2.6.3 Definition. The set of *universal conjunctive formulas* of $L_{\omega_1 P}$ is the least set containing all quantifier-free formulas and closed under arbitrary $\bigwedge$, finite $\vee$, and the quantifiers $(P\bar{x} \geq r)$.

2.6.4 Theorem (Finite Compactness Theorem (see Hoover [1978b])). *Let Φ be a set of universal conjunctive sentences of $L_{\mathbb{A}P}$. If every finite subset of Φ has a graded model, then Φ has a graded model.*

Proof. Suppose each finite subset $\Psi \subseteq \Phi$ has a model $\mathscr{M}_\Psi$. Take an ultraproduct $^*\mathscr{M}$ of the $\mathscr{M}_\Psi$'s such that, for each $\varphi \in \Phi$, almost every $\mathscr{M}_\Psi$ satisfies φ. Form a graded probability structure $\hat{\mathscr{M}}$ from $^*\mathscr{M}$ by the Loeb construction. Then, by induction show that every universal conjunctive formula holding in almost all $\mathscr{M}_\Psi$ holds in $\hat{\mathscr{M}}$ also. □

The above proof, of course, does not work for probability models, because axiom B4 is not universal conjunctive.

2.6.5 Example. Let Φ be the set of universal conjunctive sentences

$$(Px \geq 1)\left(Py \geq 1 - \frac{1}{n}\right)(Pz \geq \tfrac{1}{2})\neg(R(xz) \leftrightarrow R(yz)),$$

where $n \in \mathbb{N}$. Each finite subset of Φ has a (finite) probability model. However, Φ implies the white noise sentence

$$(Px \geq 1)(Py \geq 1)(Pz \leq \tfrac{1}{2})(R(xz) \leftrightarrow R(yz))$$

of Example 2.2.3. Thus, Φ has no probability model.

However, if in Theorem 2.6.4 every instance of axiom B4 is deducible from Φ in graded $L_{\mathbb{A}P}$, then Φ does have a probability model.

2.7. *Probabilities on Sentences of $L_{\omega_1\omega}$*

We can easily generalize our treatment of $L_{\mathbb{A}P}$ to two-sorted logic. It is more interesting that there is a mixed two-sorted logic which has probability quantifiers on one sort and ordinary quantifiers on the other sort. In this mixed two-sorted logic, we can study models which assign probabilities to sentences of $L_{\mathbb{A}}$. We will use $x, y, \ldots$ for the first sort of variables, and $s, t, \ldots$ for the second.

2.7.1 Definition. $L_{\mathbb{A}}(P, \forall)$ is the two-sorted logic which has probability quantifiers $(P\bar{x} \geq r)$ on the first sort and the universal quantifier $(\forall t)$ on the second sort. *Probability structures for* $L(P, \forall)$ have the form

$$\mathscr{M} = \langle M, T, R_i, c_j, \mu \rangle_{i \in I, j \in J},$$

where μ is a probability measure on M, and $R_i(\bar{x}; \bar{t})$ is $\mu^{(n)}$-measurable for each $\bar{t}$ in T.

If T is countable, there is no difficulty in defining the satisfaction relation in $\mathscr{M}$, with the usual clause for $(\forall t)$. This is the case which is needed for the completeness theorem.

There is also a definition of satisfaction which applies to any probability structure for $L(P, \forall)$, as introduced by Gaifman [1964] and extended by Krauss–Scott [1966]. The idea underlying this development is to assign, for each $\varphi(\bar{x}; \bar{t})$ and $\bar{b}$ in T, an element $\varphi(\bar{x}; \bar{b})^{\mathcal{M}}$ of the measure algebra of $\mu^{(n)}$ modulo the null sets. The $\forall$ clause is

$$(\forall t)\varphi(\bar{x}; \bar{b})^{\mathcal{M}} = \inf\{\varphi(\bar{x}; \bar{b}, c)^{\mathcal{M}} \mid c \in T\},$$

taking inf in the measure algebra. This coincides with the natural definition of satisfaction when T is countable, but not when T is uncountable.

2.7.2 Definition. The axioms for $L_{\mathbb{A}}(P, \forall)$ consist of all axiom schemes for $L_{\mathbb{A}P}$ and $L_{\mathbb{A}}$, with quantifiers on the appropriate sort, plus the new axiom

$$(P\bar{x} \geq r)(\forall t)\varphi(\bar{x}; t) \leftrightarrow \bigwedge_{n} (\forall t_1) \ldots (\forall t_n)(P\bar{x} \geq r) \bigwedge_{k=1}^{n} \varphi(\bar{x}; t_k).$$

The rules of inference for $L_{\mathbb{A}}(P, \forall)$ are the natural combination of rules for $L_{\mathbb{A}P}$ and $L_{\mathbb{A}}$.

2.7.3 Theorem (Soundness Theorem). *Every set Φ of sentences of $L_{\mathbb{A}}(P, \forall)$ which has a model is consistent.* □

2.7.4 Theorem (Completeness Theorem). *Every countable consistent set of sentences of $L_{\mathbb{A}}(P, \forall)$ has a model $\mathcal{M}$ with T countable.*

Proof. Form a countable weak model. Then keep the second sort fixed while using the method of Sections 2.2 and 2.3 to extend the first sort to a probability model. □

The following simpler logic is of particular interest.

2.7.5 Definition. Let L be a first-order language with variables $t_0, t_1, \ldots$ and relation symbols $R(\bar{t})$. By *L-probability logic* we mean the two-sorted logic $L'_{\mathbb{HC}}(P, \forall)$ which has only one variable x of the new sort and where L' is formed by replacing each relation $R(\bar{t})$ of L by $R(x; \bar{t})$.

L-probability logic is a logic which assigns probabilities to sentences of $L_{\omega_1\omega} = L_{\mathbb{HC}}$. Its model theory was studied by Krauss–Scott [1966]. A probability structure

$$\mathcal{M} = \langle M, T, R_i, \mu \rangle$$

for L-probability logic may be regarded as an indexed family $\langle \mathcal{M}_x \mid x \in M \rangle$ of first-order structures $\mathcal{M}_x$ for L each with universe T, together with a probability measure μ on M such that each $\{x \mid R_i(x; \bar{t})\}$ is measurable.

2.7.6 Definition. A *probability on* $L_{\omega_1\omega}$ is a function μ from sentences of $L_{\omega_1\omega}$ to $[0, 1]$ which is countably additive with respect to $\bigvee$, $\neg$ and such that each valid sentence has measure one.

Each structure $\mathscr{M}$ for L-probability logic induces the probability $\mu^{\mathscr{M}}$ on $L_{\omega_1\omega}$ given by

$$\mu^{\mathscr{M}}(\varphi) \geq r \quad \text{iff} \quad \mathscr{M} \models (Px \geq r)\varphi.$$

The axioms and rules for L- probability logic are like those for $L_{\mathbb{HC}}(P, \forall)$ except that axioms A3, B2, B3, and B4 disappear. The soundness and completeness theorems still hold and have easier proofs which avoid graded structures.

The following completeness theorem was proved by Krauss–Scott [1966], extending results of Gaifman [1964] and Łoś [1963]. Although it does not follow from Theorem 2.7.4, it can be proven by a similar argument.

2.7.7 Theorem. *Let μ be a probability on $L_{\omega_1\omega}$ which assigns* 0 *or* 1 *to pure equality sentences. For every countable set $\Psi \subseteq L_{\omega_1\omega}$, there is a structure $\mathscr{M}$ for L-probability logic such that $\mu^{\mathscr{M}}$ agrees with μ on Ψ.* ☐

Other work on probabilities of sentences can be found in Havranek [1975], Fenstad [1967], Fagin [1976], Compton [1980], Lynch [1980], Gaifman–Snir [1982], and Krauss [1969].

The logic $L_{\mathbb{A}P}$ should be compared with the logic $L(Q_m)$ of H. Friedman, which is discussed in Chapter XVI. This logic also has models with measures as well as both the classical quantifier $(\forall x)$ and the measure quantifier $(Q_m x)$ which has the same interpretation as our $(Px > 0)$. In order to have both quantifiers, one must pay the price of restricting attention to those structures in which every definable set is Borel (the absolutely Borel structures). A similar treatment of logic with both quantifiers $\forall x$ and $(P\bar{x} \geq r)$ for absolutely Borel structures should be possible and interesting.

3. *Model Theory*

In this section, we will develop the model theory of the logic $L_{\mathbb{A}P}$. In Section 3.1 we state a model-theoretic form of the law of large numbers, showing that every probability structure is "approximated" by almost every sequence of finite substructures. This result is used in Section 3.2 to prove the existence and uniqueness of hyperfinite models, which play the role for $L_{\mathbb{A}P}$ that saturated models play in classical model theory. These models are used in Section 3.3 to prove the Robinson consistency and Craig interpolation theorems for $L_{\mathbb{A}P}$. The section concludes with the development of integrals, which eliminate quantifiers from $L_{\mathbb{A}P}$ in a manner analogous to Skolem functions in classical logic.

3.1. Laws of Large Numbers

The results in this section hold for all graded probability structures. First, we have

3.1.1 Definition. A *finite universal formula of* $L_{\mathbb{A}P}$ is a formula of the form

$$(P\bar{x}_1 \geq r_1) \ldots (P\bar{x}_n \geq r_n)\psi,$$

where ψ is a finite quantifier-free formula of L. A *finite existential formula* of $L_{\mathbb{A}P}$ is a formula of the form

$$(P\bar{x}_1 > r_1) \ldots (P\bar{x}_n > r_n)\psi,$$

where ψ is as before.

Note that since $\neg(P\bar{x} > r)\psi$ is equivalent to $(P\bar{x} \geq 1 - r) \neg \psi$, the negation of a finite existential formula is equivalent to a finite universal formula, and vice versa. We shall see that the laws of large numbers for $L_{\mathbb{A}P}$ deal with finite existential sentences. To state them, however, we need the notion of a finite sample of $\mathscr{M}$.

3.1.2 Definition. Let $\mathscr{M}$ be a graded probability structure for L, and let $\bar{a}_k = \langle a_1, \ldots, a_k \rangle \in M^k$ be a k-tuple of elements of M. The *finite sample* $\mathscr{M}(\bar{a}_k)$ is the probability structure whose universe is the union of $\{a_1, \ldots, a_k\}$ and the constants (if any) of $\mathscr{M}$, whose first-order part is a substructure of $\mathscr{M}$, and whose measure v is given by

$$v(S) = |\{m \leq k \,|\, a_m \in S\}|/k.$$

Thus, the finite set $\{a_1, \ldots, a_k\}$ has measure one in $\mathscr{M}(\bar{a}_k)$, and the measure of a singleton $\{a\}$ is $1/k$ times the number of occurrences of a in the sequence $\bar{a}_k$.

3.1.3 Theorem. *Let $\mathscr{M}$ be a graded probability structure for L with measures μ_n, and let φ be a finite existential sentence of $L_{\mathbb{A}P}$ such that $\mathscr{M} \models \varphi$.*

(i) *Weak Law of Large Numbers for $L_{\mathbb{A}P}$*:

$$\lim_{k \to \infty} \mu_k\{\bar{a}_k \in M^k \,|\, \mathscr{M}(\bar{a}_k) \models \varphi\} = 1.$$

(ii) *Strong Law of Large Numbers for $L_{\mathbb{A}P}$. Let $\mu_{\mathbb{N}}$ be the completion of the measure on $M^{\mathbb{N}}$ determined by the μ_n. Then, for $\mu_{\mathbb{N}}$ almost all sequences $\bar{a} \in M^{\mathbb{N}}$, $\mathscr{M}(\bar{a}_k) \models \varphi$ for all but finitely many $k \in N$.* □

The above theorem is a reformulation of Lemma 6.13 in Keisler [1977b]. In the special case in which φ has the form $(Px > r)\psi(x)$, the result follows directly from the weak and strong laws of large numbers in probability theory. Hoover has pointed out that the case in which φ has the form $(P\bar{x} > r)\psi(\bar{x})$ can be proved by the same argument as the proof of the strong law in probability theory, using the martingale convergence theorem. The general case uses an induction on the number of quantifiers.

3.1.4 Theorem (Normal Form Theorem (Hoover [1982])). *Every formula $\varphi(\bar{x})$ of graded $L_{\omega_1 P}$ is equivalent to a countable boolean combination of formulas of the form $(P\bar{y} \geq r)\psi(\bar{x}\bar{y})$, where $\psi(\bar{x}\bar{y})$ is a finite conjunction of atomic formulas of L.*

Proof. By a prenex normal form argument, it can be shown that every formula of graded $L_{\omega_1 P}$ is equivalent to a countable boolean combination of finite universal formulas (with the same free variables). By the Weak Law of Large Numbers, each statement below implies the next, where ψ is a finite quantifier-free formula.

(1) $$\mathscr{M} \models (P\bar{x} \geq r)(P\bar{y} \geq s)\psi$$

(2) $$\bigwedge_n \mathscr{M} \models \left(P\bar{x} > r - \frac{1}{n}\right)\left(P\bar{y} > s - \frac{1}{n}\right)\psi$$

(3) $$\bigwedge_n \lim_{k\to\infty} \mu_k\left\{\bar{a}_k \mid \mathscr{M}(\bar{a}_k) \models \left(P\bar{x} > r - \frac{1}{n}\right)\left(P\bar{y} > s - \frac{1}{n}\right)\psi\right\} = 1$$

(4) $$\bigwedge_n \lim_{k\to\infty} \mu_k\left\{\bar{a}_k \mid \mathscr{M}(\bar{a}_k) \models \left(P\bar{x} \geq r - \frac{1}{n}\right)\left(P\bar{y} \geq s - \frac{1}{n}\right)\psi\right\} > 0$$

(5) $$\bigwedge_n \mathscr{M} \models \left(P\bar{x} \geq r - \frac{1}{n}\right)\left(P\bar{y} \geq s - \frac{1}{n}\right)\psi$$

(6) $$\mathscr{M} \models (P\bar{x} \geq r)(P\bar{y} \geq s)\psi$$

Hence, these statements are equivalent. Each property

$$\mathscr{M}(\bar{z}_k) \models \left(P\bar{x} \geq r - \frac{1}{n}\right)\left(P\bar{y} \geq s - \frac{1}{n}\right)\psi$$

is expressible in $\mathscr{M}$ by a finite quantifier-free formula $\theta(\bar{z}_k)$ of L. It follows, then, that each formula is equivalent to a countable Boolean combination of formulas of the form $(P\bar{z} \geq t)\theta$, where θ is finite and quantifier-free. Finally, we reduce to the case in which θ is a conjunction of atomic formulas using the probability rules

$$P(\neg\varphi) = 1 - P(\varphi),$$

and

$$P(\varphi \vee \psi) = P(\varphi) + P(\psi) - P(\varphi \wedge \psi). \quad \square$$

3.1.5 Corollary. *Let $\mathscr{M}$ and $\mathscr{N}$ be graded probability structures for L. The following are equivalent.*

(a) $\mathscr{M} \equiv_{\omega_1 P} \mathscr{N}$.
(b) $\mathscr{M} \equiv_{\mathbb{A}P} \mathscr{N}$.
(c) $\mathscr{M} \models \varphi$ *iff* $\mathscr{N} \models \varphi$ *for each sentence* φ *of* $L_{\mathbb{A}P}$ *in the normal form of Theorem* 3.1.4. □

The following consequence characterizes $L_{\mathbb{A}P}$ equivalence in terms of truth values in finite samples. It has no analog in first-order logic.

3.1.6 Theorem. *Let $\mathscr{M}$ and $\mathscr{N}$ be graded probability structures for L. The following are equivalent.*

(a) *$\mathscr{M}$ and $\mathscr{N}$ are $L_{\mathbb{A}P}$-equivalent.*
(b) *For every sentence φ of $L_{\mathbb{A}P}$ and $k \in \mathbb{N}$, we have*

$$\mu_k\{\bar{a}_k \mid \mathscr{M}(\bar{a}_k) \models \varphi\} = \nu_k\{\bar{b}_k \mid \mathscr{N}(\bar{b}_k) \models \varphi\}.$$

That is, φ has the same probability in the set of k-element samples of $\mathscr{M}$ as in the set of k-element samples of $\mathscr{N}$.
(c) *For each sentence φ of $L_{\mathbb{A}P}$,*

$$\lim_{k\to\infty} \mu_k\{\bar{a}_k \mid \mathscr{M}(\bar{a}_k) \models \varphi\} = 1$$

if and only if

$$\lim_{k\to\infty} \nu_k\{\bar{b}_k \mid \mathscr{N}(\bar{b}_k) \models \varphi\} = 1.$$

That is, φ has large probability in large finite samples in $\mathscr{M}$ iff it does in $\mathscr{N}$.

Proof. We give a proof using Hoover's normal form theorem. The result can also be proved directly from the Weak Law of Large Numbers for $L_{\mathbb{A}P}$. Now, (a) implies (b), because for each k and ψ there is a formula $\psi(v_1, \ldots, v_k)$ of $L_{\mathbb{A}P}$ which says that a k-element sample satisfies φ. It is trivial that (b) implies (c). Assume then that (c) holds, and let $\varphi(\bar{x})$ be a finite quantifier-free formula. Suppose that $\mathscr{M} \models (P\bar{x} \geq r)\varphi(\bar{x})$, and let $s < r$. By the Weak Law of Large Numbers, we have

$$\lim_{k\to\infty} \mu_k\{\bar{a}_k \mid \mathscr{M}(\bar{a}_k) \models (P\bar{x} > s)\varphi\} = 1.$$

By (c), the same holds in $\mathscr{N}$. Applying the Weak Law again, we thus have $\mathscr{N} \models (P\bar{x} \geq s)\varphi$. Since this holds for all $s < r$, $\mathscr{N} \models (P\bar{x} \geq r)\varphi$. It follows from the Normal Form Theorem that $\mathscr{M} \equiv_{\omega_1 P} \mathscr{N}$. □

3.2. Hyperfinite Models

We will assume throughout this section that L has only finitely many constant symbols. We have seen in Example 1.3.2 that the sentence

$$(Px \geq 1)(Py \geq 1)x \neq y,$$

stating that $\mathscr{M}$ is atomless, has no countable models. In this section, we prove an analogue of the Löwenheim–Skolem theorem for atomless probability structures, but with infinite $*$-finite numbers in place of infinite cardinals. We will show that, for each atomless structure $\mathscr{M}$ and infinite $*$-finite number H, there is an essentially unique hyperfinite probability structure $\mathscr{N} \equiv_{\omega_1 P} \mathscr{M}$ of $*$-cardinal H. We will use a fixed ω_1-saturated nonstandard universe.

3.2.1 Definition. A (uniform) *finite probability structure* is a probability structure $\mathscr{M}$ whose universe M is finite and whose measure is the counting measure $\mu(Y) = |Y|/|M|$. A *$*$-finite probability structure* is a finite probability structure in the sense of the nonstandard universe. A *hyperfinite probability structure* is a probability structure $\mathscr{M}$ such that the universe M is an infinite $*$-finite set and μ is the Loeb measure determined by the $*$-counting measure on M. A *hyperfinite graded structure* is a graded probability structure whose universe M is an infinite $*$-finite set and each μ_n is the Loeb measure determined by the $*$-counting measure on M^n.

3.2.2 Proposition. *Every hyperfinite probability structure or graded structure is atomless.* ☐

Here is a reformulation of Proposition 2.2.4.

3.2.3 Proposition. *Let $\mathscr{M}_0$ be a first-order structure such that the universe M is an infinite $*$-finite set and each relation of $\mathscr{M}_0$ is Loeb measurable with respect to the $*$-counting measure on M^n. Then there is a unique hyperfinite graded structure with first-order part $\mathscr{M}_0$.* ☐

We will now introduce an important relation between hyperfinite and $*$-finite structures, called a *lifting*.

3.2.4 Definition. Let $\mathscr{M}$ be a hyperfinite graded structure. A *lifting* of $\mathscr{M}$ is a $*$-finite probability structure $\mathscr{N}$ such that $\mathscr{N}$ has the same universe and constants as $\mathscr{M}$, and for each atomic formula $\varphi(\bar{x})$, the set

$$\{\bar{a} \mid \mathscr{M} \models \varphi[\bar{a}] \quad \text{iff} \quad \mathscr{N} \models \varphi[\bar{a}]\}$$

has μ_n-measure one. By a *lifting* of a hyperfinite probability structure $\mathscr{M}$ we mean a lifting of the unique hyperfinite graded structure $\mathscr{M}'$ which has the same first-order part as $\mathscr{M}$.

3.2.5 Lemma. (i) *Every infinite $*$-finite probability structure is a lifting of some hyperfinite graded structure.*
(ii) *Every hyperfinite graded structure has a lifting.*
(iii) *If $\mathscr{M}$ and $\mathscr{N}$ are hyperfinite graded structures with a common lifting, then $\mathscr{M} = \mathscr{N}$ a.s. and $\mathscr{M} \equiv_{\omega_1 P} \mathscr{N}$.*

Proof. The argument for part (i) follows by Proposition 3.2.3. The argument for part (ii) follows by Theorem 2.1.3. And the argument for part (iii) follows by Lemma 2.3.3. □

3.2.6 Theorem (Existence Theorem for Hyperfinite Models (Keisler [1977b])). *Let $\mathscr{N}$ be an atomless probability structure for L, and let M be an infinite $*$-finite set. Then there exists a hyperfinite probability structure $\mathscr{M}$ with universe M which is $L_{\mathbb{A}P}$-equivalent to $\mathscr{N}$.*

Proof. Assume first that L has no constant symbols. Let S be the set of all infinite sequences $\bar{a}$ of elements of $\mathscr{N}$ such that for every finite existential sentence φ of $L_{\mathbb{A}P}$, if $\mathscr{N} \models \varphi$ then $\mathscr{N}(\bar{a}_k) \models \varphi$, for all but finitely many k. By the Strong Law of Large Numbers, $\nu^{\mathbb{N}}(S) = 1$. Since $\mathscr{N}$ is atomless, $\nu^{\mathbb{N}}$ almost every sequence $\bar{a}$ is one-to-one; and, hence, each $\mathscr{N}(\bar{a}_k)$ is a uniform finite probability structure. Thus, there exists $\bar{a} \in S$ such that $\bar{a}$ is one-to-one. Let K be an infinite hyperinteger. Then $\mathscr{N}(\bar{a}_K)$ is a $*$-finite probability structure of $*$-cardinal K and is a lifting of a hyperfinite graded structure $\mathscr{M}'$. Since $\bar{a} \in S$, for each finite quantifier-free sentence $\psi(\bar{x})$ and each r, $\mathscr{N} \models (P\bar{x} > r)\psi$ implies $\mathscr{M}' \models (P\bar{x} \geq r)\psi$. It follows then that, for each ψ and r, $\mathscr{N} \models (P\bar{x} \geq r)\psi$ iff $\mathscr{M}' \models (P\bar{x} \geq r)\psi$. By the Normal Form Theorem, $\mathscr{M}'$ is $L_{\mathbb{A}P}$-equivalent to $\mathscr{N}$. By Theorem 2.3.4, there is a hyperfinite probability structure $\mathscr{M}$ with $\mathscr{M} = \mathscr{M}'$ a.s. Then $\mathscr{M}$ is $L_{\mathbb{A}P}$-equivalent to $\mathscr{N}$.

The case in which L has finitely many constants is the same except that the measure on $*\mathscr{N}(\bar{a}_K)$ is slightly different from the counting measure since constants have measure zero. □

The Existence Theorem also holds for graded probability structures, with the same proof. For $L_{\mathbb{A}P}$ without equality, the existence theorem holds even without the hypothesis that $\mathscr{N}$ is atomless.

3.2.7 Definition. Let $\mathscr{M}$ and $\mathscr{N}$ be probability structures. An *almost sure isomorphism from $\mathscr{M}$* to $\mathscr{N}$ (in symbols, $h\colon \mathscr{M} \cong \mathscr{N}$ a.s.) is a bijection $h\colon M \to N$ such that

(a) h is an isomorphism on the probability spaces, $h\colon \langle M, \mu\rangle \cong \langle N, \nu\rangle$;
(b) for each atomic formula $\varphi(\bar{x})$,

$$\mathscr{M} \models \varphi[\bar{a}] \quad \text{iff} \quad \mathscr{N} \models \varphi[h\bar{a}]$$

almost surely in $\mu^{(n)}$.

3.2.8 Lemma. *Suppose $h: \mathscr{M} \cong \mathscr{N}$ a.s., then*

(i) *for each formula $\varphi(\bar{x})$ of $L_{\mathbb{A}P}$,*

$$\mathscr{M} \models \varphi[\bar{a}] \quad \text{iff} \quad \mathscr{N} \models \varphi[h\bar{a}]$$

almost surely in $\mu^{(n)}$;

(ii) *$\mathscr{M}$ and $\mathscr{N}$ are $L_{\mathbb{A}P}$-equivalent.*

Proof. The proof follows by induction on φ. ☐

The following result is new.

3.2.9 Theorem (Uniqueness Theorem for Hyperfinite Models). *Let $\mathscr{M}$ and $\mathscr{N}$ be hyperfinite probability structures with the same universe M. The following are equivalent:*

(a) *$\mathscr{M}$ and $\mathscr{N}$ are $L_{\mathbb{A}P}$-equivalent.*
(b) *There is an $h: \mathscr{M} \cong \mathscr{N}$ a.s.*
(c) *There is an internal h such that $h: \mathscr{M} \cong \mathscr{N}$ a.s.*

Idea of Proof. We assume that (a) holds and prove that (c) must hold also. Note that any internal bijection preserves measure. Consider an n-tuple of atomic formulas $\varphi_1(\bar{y}), \ldots, \varphi_n(\bar{y})$ of L and let $\varepsilon > 0$. Using the Rectangle Approximation Lemma (Lemma 2.3.1), one can find a bijection $h_0: M \to M$ such that

$$\mu^{(m)}\left(\bigcap_{k=1}^{n} \{\bar{a} \mid \mathscr{M} \models \varphi_k[\bar{a}] \quad \text{iff} \quad \mathscr{N} \models \varphi_k[h_0\bar{a}]\}\right) \geq 1 - \varepsilon.$$

The idea is to use Theorem 2.1.3 in choosing an h_0 which approximately preserves each coordinate of the rectangles which approximate φ_k. Now let $\hat{\mathscr{M}}, \hat{\mathscr{N}}$ be liftings of $\mathscr{M}, \mathscr{N}$. By ω_1-saturation, we can find an internal bijection h so that for all atomic $\varphi(\bar{y})$ and all real $\varepsilon > 0$,

$$\mu^{(m)}(\{\bar{a} \mid \hat{\mathscr{M}} \models \varphi[\bar{a}] \quad \text{iff} \quad \hat{\mathscr{N}} \models \varphi[h\bar{a}]\}) \geq 1 - \varepsilon.$$

It follows then that $h: \mathscr{M} \cong \mathscr{N}$ a.s., and thus (c) holds. ☐

As a consequence of the preceding, we obtain a "soft" characterization of the $L_{\mathbb{A}P}$-equivalence relation, namely

3.2.10 Corollary. *Let $\approx$ be an equivalence relation on the atomless probability structures for L such that:*

(a) *If $\mathscr{M} \cong \mathscr{N}$ a.s., then $\mathscr{M} \approx \mathscr{N}$.*
(b) *For each $\mathscr{M}$ and each infinite $*$-finite set H, there is a hyperfinite probability structure $\mathscr{N}$ with universe H such that $\mathscr{M} \approx \mathscr{N}$.*
(c) *If $\mathscr{M} \approx \mathscr{N}$, then $\mathscr{M} \equiv_{\mathbb{A}P} \mathscr{N}$.*

Then $\approx$ is the relation $\equiv_{\mathbb{A}P}$.

3.2.11 Example (D. Hoover, unpublished). This example shows that the uniqueness theorem (Theorem 3.2.9) is false for hyperfinite graded models. Let M be a hyperfinite set of the form $M = A \cup B \cup C \cup D$ where A, B, C, D are disjoint sets with $*$-cardinalities

$$|A| = \tfrac{1}{2}|M|, \qquad |B| = \tfrac{1}{4}|M|, \qquad |C| = |D| = \tfrac{1}{8}|M|.$$

Let f be an internal bijection from C to D. By using an exponential form of Chebyshev's inequality, it can be shown that there is an internal binary relation $R \subseteq A \times (B \cup C \cup D)$ such that:

(1) For all $y \in B \cup C \cup D$,

$$\mu\{x \mid R(x, y)\} = \tfrac{1}{4}.$$

(2) For all $y \in C$,

$$\{x \mid R(x, y)\} = \{x \mid R(x, fy)\}.$$

(3) For all $y, z \in B \cup C \cup D$ with $z \neq y$, $z \neq fy$,

$$\mu\{x \mid R(x, y) \wedge R(x, z)\} = \tfrac{1}{8}.$$

Let $\mathscr{M}$ and $\mathscr{N}$ be the graded hyperfinite structures with first-order parts $\mathscr{M}_0 = \langle M, B, R\rangle$, and $\mathscr{N}_0 = \langle M, C \cup D, R\rangle$. The reader can check that $\mathscr{M}$ and $\mathscr{N}$ are $L_{\omega_1 P}$-equivalent but for any internal bijection h on M which maps B onto $C \cup D$, the set

$$\{(x, y) \mid R(x, y) \quad \text{iff} \quad R(hx, hy)\}$$

has measure at most $\tfrac{31}{32}$.

A weak uniqueness theorem for hyperfinite graded models is given in Keisler [1977, p. 34].

3.3. Robinson Consistency and Craig Interpolation

The results of this section are due to Hoover [1978b]. The hyperfinite structures play the same role that saturated structures play in first-order model theory.

3.3.1 Theorem (Robinson Consistency Theorem for $L_{\mathbb{A}P}$). *Let L^1 and L^2 be two languages with $L^0 = L^1 \cap L^2$. Let $\mathscr{M}^1$, $\mathscr{M}^2$ be probability structures for L^1, L^2 respectively whose reducts $\mathscr{M}^1 \upharpoonright L^0$, $\mathscr{M}^2 \upharpoonright L^0$ are $L^0_{\mathbb{A}P}$-equivalent. Then there exists a probability structure $\mathscr{N}$ for $L^1 \cup L^2$ such that*

$$\mathscr{N} \upharpoonright L^1 \equiv_{L^1_{\mathbb{A}P}} \mathscr{M}^1 \quad \textit{and} \quad \mathscr{N} \upharpoonright L^2 \equiv_{L^2_{\mathbb{A}P}} \mathscr{M}^2.$$

Proof. We give the proof when L^1 and L^2 have no constants and $\mathscr{M}^1$, $\mathscr{M}^2$ are atomless. The general case will follow by adding a new relation symbol for each atomic formula, and working with the atomless parts. By Theorem 3.2.6, we may take $\mathscr{M}^1$ and $\mathscr{M}^2$ to be hyperfinite probability structures with the same universe M. By Theorem 3.2.9 there is an internal bijection $h: \mathscr{M}^1 \upharpoonright L^0 \cong \mathscr{M}^2 \upharpoonright L^0$ a.s. Renaming elements, we can take h to be the identity map. By changing the L^0 relations of $\mathscr{M}^2$ on a set of measure zero, we get $\mathscr{M}^1 \upharpoonright L^0 = \mathscr{M}^2 \upharpoonright L^0$. Let $\mathscr{N}$ be the common expansion of $\mathscr{M}^1$ and $\mathscr{M}^2$. □

3.3.2 Theorem (Craig Interpolation Theorem for $L_{\mathbb{A}P}$). *Let $L^0 = L^1 \cap L^2$ and let $\varphi^1 \in L^1_{\mathbb{A}P}$, and $\varphi^2 \in L^2_{\mathbb{A}P}$ be sentences such that $\models \varphi^1 \to \varphi^2$. Then there is a sentence $\varphi^0 \in L^0_{\mathbb{A}P}$ such that $\models \varphi^1 \to \varphi^0$, and $\models \varphi^0 \to \varphi^2$.*

Proof. Suppose there is no such φ^0. By a Henkin construction, there then are weak models $\mathscr{M}^1$ of φ^1 and $\mathscr{M}^2$ of $\neg\varphi^2$ for $L^1_{\mathbb{A}P}$ and $L^2_{\mathbb{A}P}$ such that $\mathscr{M}^1 \upharpoonright L^0$ and $\mathscr{M}^2 \upharpoonright L^0$ are $L^0_{\mathbb{A}P}$-equivalent, and all the axioms of $L^1_{\mathbb{A}P}$, $L^2_{\mathbb{A}P}$ are valid. By the completeness proof, we then obtain probability models $\mathscr{N}^1$ of φ^1, $\mathscr{N}$ of $\neg\varphi^2$, where $\mathscr{N}^1 \upharpoonright L^0$ and $\mathscr{N}^2 \upharpoonright L^0$ are $L^0_{\mathbb{A}P}$-equivalent. By Robinson consistency, $\varphi^1 \wedge \neg\varphi^2$ has a model—contradicting $\models \varphi^1 \to \varphi^2$. □

Since the compactness theorem fails for $L_{\mathbb{A}P}$, we cannot apply the general fact that Robinson consistency and compactness implies Craig interpolation. A separate Henkin construction is thus needed. Mundici, in Chapter VIII, showed that for many logics, Robinson consistency implies compactness. The logic $L_{\mathbb{A}P}$ fails to satisfy several of his hypotheses, including closure under universal quantification and under disjoint unions.

Hoover [198?] has recently proved the following stronger interpolation theorem, thus improving an earlier result which appeared in Hoover [1982].

3.3.3 Theorem (Almost Sure Interpolation Theorem). *Let $L^0 = L^1 \cap L^2$ and suppose the symbols of L^0 have a well-ordering in $\mathbb{A}$. Let $\varepsilon > 0$ and let $\varphi^1(\bar{v}) \in L^1_{\mathbb{A}P}$, and $\varphi^2(\bar{v}) \in L^2_{\mathbb{A}P}$ be formulas such that*

$$\models (P\bar{v} \geq 1 - \varepsilon)(\varphi(\bar{v}) \to \psi(\bar{v})).$$

Then, for every $\delta > \varepsilon^{1/4} + \varepsilon^{1/2}$, there is a formula $\theta(\bar{v}) \in L^0_{\mathbb{A}P}$ such that

$$\models (P\bar{v} \geq 1 - \delta)(\varphi(\bar{v}) \to \theta(\bar{v})) \quad \text{and} \quad \models (P\bar{v} \geq 1 - \delta)(\theta(\bar{v}) \to \psi(\bar{v})).$$

Moreover, if

$$\models (P\bar{v} \geq 1)(\varphi(\bar{v}) \to \psi(\bar{v})),$$

then there is a formula $\theta(\bar{v}) \in L^0_{\mathbb{A}P}$ such that

$$\models (P\bar{v} \geq 1)(\varphi(\bar{v}) \to \theta(\bar{v})) \quad \text{and} \quad \models (P\bar{v} \geq 1)(\theta(\bar{v}) \to \psi(\bar{v})). \ \square$$

Hoover proved each of the results Sections 3.2.1–3.2.3 for graded probability structures, and the results for probability structures follow. His proof of the Robinson consistency theorem was somewhat more difficult, because the uniqueness theorem for graded hyperfinite structures is false.

Additional model-theoretic results for $L_{\mathbb{A}P}$ are in Keisler [1977b] and Hoover [1982]. Hoover [1981] gives some applications to probability theory. Kaufmann [1978a] in his thesis gave a back-and-forth criterion which is *sufficient* for two graded structures to be $L_{\mathbb{A}P}$-equivalent, and *necessary* for two hyperfinite graded structures to be $L_{\mathbb{A}P}$-equivalent.

3.4. Logic with Integrals

Properties of random variables are usually easier to express using integrals rather than probability quantifiers. We will now introduce a logic $L_{\mathbb{A}\int}$ (from Keisler [1977]), which is equivalent to $L_{\mathbb{A}P}$. It has no quantifiers, although it does have an integral operator which builds terms with bound variables. Indeed, the logics $L_{\mathbb{A}P}$ and $L_{\mathbb{A}\int}$ correspond to two alternative approaches to integration theory: Lebesgue measure theory and the Daniell integral. The completeness proof for $L_{\mathbb{A}P}$ used Loeb's construction of a measure by non-standard analysis, while the completeness proof for $L_{\mathbb{A}\int}$ will use the construction of the Daniell integral as given in Loeb [1982].

Given a relation $R(\bar{x})$, the *indicator function* $\mathbf{1}(R(\bar{x}))$ is defined by

$$\mathbf{1}(R(\bar{x})) = \begin{cases} 1 & \text{if } R(\bar{x}) \text{ is true,} \\ 0 & \text{if } R(\bar{x}) \text{ is false.} \end{cases}$$

The language $L_{\mathbb{A}\int}$ will have atomic terms interpreted as the indicator functions of the atomic formulas of L, and more complex terms will be built from these by applying continuous real functions and integration. The atomic formulas of $L_{\mathbb{A}\int}$ will be inequalities between terms.

3.4.1 Definition. Let L be an $\mathbb{A}$-recursive set of finitary relation and constant symbols. For each atomic formula

$$R(\bar{x}) \quad \text{or} \quad x = y$$

of the first-order logic L, $L_{\mathbb{A}\int}$ has an *atomic term*

$$\mathbf{1}(R(\bar{x})) \quad \text{or} \quad \mathbf{1}(x = y).$$

The set of *terms* of $L_{\mathbb{A}\int}$ is the least set such that:

(a) Every atomic term is a term.
(b) If τ is a term and x is an individual variable, then $\int \tau \, dx$ is a term.
(c) Each real number $r \in \mathbb{A} \cap \mathbb{R}$ is a term.
(d) If $\tau_1, \ldots, \tau_n$ are terms and F belongs to the set $C_{\mathbb{A}}(\mathbb{R}^n)$ of continuous functions $F: \mathbb{R}^n \to \mathbb{R}$ such that $F \upharpoonright \mathbb{Q}^n \in \mathbb{A}$, then $\mathbf{F}(\tau_1, \ldots, \tau_n)$ is a term.

Clause (c) is just the special case of Clause (d) when $n = 0$. We will usually identify the function F and the corresponding logical symbol $\mathbf{F}$. Note that individual variables and constants are not terms of $L_{\mathbb{A}\int}$. The notion of *free and bound* variables in a term are defined as usual, with the integral $\int \tau\, dx$ binding x. A *closed term* is a term with no free variable.

3.4.2 Definition. The set of *formulas of* $L_{\mathbb{A}\int}$ is the least set such that:

(a) For each term τ of $L_{\mathbb{A}\int}$, $\tau \geq 0$ is an atomic formula.
(b) If φ is a formula, so is $\neg\varphi$.
(c) If Φ is a set of formulas with finitely many free variables and $\Phi \in \mathbb{A}$, then $\bigwedge\Phi$ is a formula.

A *sentence* is a formula with no free variables. The structures for $L_{\mathbb{A}\int}$ are the same as the structures for $L_{\mathbb{A}P}$, namely, the probability structures for L.

3.4.3 Definition. Let $\mathscr{M}$ be a probability structure for L. The *value* $\tau(\bar{a})^{\mathscr{M}}$ of a term $\tau(\bar{v})$ of $L_{\mathbb{A}\int}$ in $\mathscr{M}$ at a tuple $\bar{a}$ in M is defined by:

(a) $\mathbf{1}(\varphi(\bar{a}))^{\mathscr{M}} = \begin{cases} 1 & \text{if } \mathscr{M} \models \varphi[\bar{a}], \\ 0 & \text{if } \mathscr{M} \models \neg\varphi[\bar{a}] \end{cases}$

for each atomic formula $\varphi(\bar{v})$ of $L_{\omega\omega}$.
(b) $(\int \tau(x, \bar{a})\, dx)^{\mathscr{M}} = \int \tau(b, \bar{a})^{\mathscr{M}}\, d\mu(b)$.
(c) $r^{\mathscr{M}} = r$.
(d) $\mathbf{F}(\tau_1, \ldots, \tau_n)(\bar{a})^{\mathscr{M}} = F(\tau_1(\bar{a})^{\mathscr{M}}, \ldots, \tau_n(\bar{a})^{\mathscr{M}})$.

Since each atomic term has values in $\{0, 1\}$, by induction we see that each function $\tau(\bar{a})^{\mathscr{M}}$ is bounded and $\mu^{(k)}$-measurable. In particular, the integral in Part (b) exists and is finite. The *satisfaction relation* $\mathscr{M} \models \varphi[\bar{a}]$ for $L_{\mathbb{A}\int}$ is defined in the natural way, with the atomic formula rule

$$\mathscr{M} \models (\tau(\bar{u}) \geq 0)[\bar{a}] \quad \text{iff} \quad \tau(\bar{a})^{\mathscr{M}} \geq 0.$$

3.5. Completeness Theorem with Integrals

3.5.1 Definition. The axioms for $L_{\mathbb{A}\int}$, where σ, τ are terms, r, s are elements of $\mathbb{A} \cap \mathbb{R}$, and x, y are individual variables are:

D1. All axiom schemes for $L_{\mathbb{A}}$ without quantifiers, with $\mathbf{1}(x = y) = 1$ in place of $x = y$.
D2. For each atomic term τ, we have $\tau = 0 \vee \tau = 1$.
D3. *Order axioms.* Using abbreviations, we have
  (i) $r \geq r$.
  (ii) $\tau \geq r \rightarrow \tau \geq s$, when $r \geq s$.
D4. For each rational closed rectangle $S \subseteq \mathbb{R}^m$, and each $F \in C_{\mathbb{A}}(\mathbb{R}^m)$ with image $F(S) = [a, b]$, we have

$$\langle \tau_1, \ldots, \tau_m \rangle \in S \rightarrow \mathbf{F}(\tau_1, \ldots, \tau_M) \in [a, b].$$

D5. *Integral axioms.*

(i) $\int r\, dx = r$.

(ii) $\int \tau(x)\, dx = \int \tau(y)\, dy$.

(iii) $\iint \tau(x, y)\, dx\, dy = \iint \tau(x, y)\, dy\, dx$.

(iv) $\int (r \cdot \sigma + s \cdot \tau)\, dx = r \cdot \int \sigma\, dx + s \cdot \int \tau\, dx$.

D6. *Archimedean axiom.*

$$\tau > r \leftrightarrow \bigvee_n \tau \geq r + \frac{1}{n}.$$

D7. *Product measurability.* For each $m \in \mathbb{N}$, we have

$$\bigvee_k \int F_k\left(1 - \int F_m\left(\int |\tau(\bar{x}, \bar{z}) - \tau(\bar{y}, \bar{z})|\, d\bar{z}\right) d\bar{y}\right) d\bar{x} \geq 1 - \frac{1}{m},$$

where $d\bar{x}$ is $dx_1 \ldots dx_n$, and

$$F_k(u) = \begin{cases} 0 & \text{if } u \leq 1/k, \\ \text{linear for} & 1/k \leq u \leq 2/k, \\ 1 & \text{if } u \geq 2/k. \end{cases}$$

This is essentially a translation of the axiom B4 for $L_{\mathbb{A}P}$.

3.5.2 Definition. The rules of inference for $L_{\mathbb{A}\int}$ are:

S1. *Modus ponens*: $\varphi, \varphi \to \psi \vdash \psi$.

S2. *Conjunction*: $\{\varphi \to \psi \mid \psi \in \Psi\} \vdash \varphi \to \bigwedge\Psi$.

S3. *Generalization*: $\varphi \to (\tau(x) \geq 0) \vdash \varphi \to (\int \tau(x)\, dx \geq 0)$, where x is not free in φ.

This set of axioms was motivated by the thesis of Rodenhausen [1982].

3.5.3 Theorem (Soundness Theorem for $L_{\mathbb{A}\int}$). *Any set Φ of sentences for $L_{\mathbb{A}\int}$ which has a model is consistent.* ☐

3.5.4 Theorem (Completeness Theorem for $L_{\mathbb{A}\int}$). *A countable consistent set Φ of sentences of $L_{\mathbb{A}\int}$ has a model.* ☐

Idea of Proof. As in the case of $L_{\mathbb{A}P}$, the main steps are to prove a weak completeness theorem, and then use a construction from non-standard analysis to obtain a graded model of Φ. This done, the product measurability axiom is then used to get a model of Φ.

A *weak structure for* $L_{\mathbb{A}\int}$ is a structure

$$\mathcal{M} = \langle M, R_i, c_j, I \rangle,$$

where $\langle M, R_i, c_j \rangle$ is a first-order structure and I is a positive linear real function on the set of terms of $L_{\mathbb{A}\int}$ with at most one free variable x and parameters from M. That is,

(1) $I(r) = r$.
(2) $I(r \cdot \sigma + s \cdot \tau) = r \cdot I(\sigma) + s \cdot I(\tau)$.
(3) If $\tau(b, \bar{a})^{\mathcal{M}} \geq 0$, for all $b \in M$, then $I(\tau(x, \bar{a})) \geq 0$.

The recursive definition of $\tau(\bar{a})^{\mathcal{M}}$ is the same as for ordinary probability structures but with the integral clause

$$\left(\int \tau(x, \bar{a})\, dx\right)^{\mathcal{M}} = I(\tau(x, \bar{a})).$$

A Henkin argument is used to construct a weak model of Φ in which each axiom of $L_{\mathbb{A}\int}$ is valid. Then the internal structure $\mathcal{M}$ is formed in the non-standard universe. The Daniell integral construction of Loeb [1984] produces a probability measure μ on $^*M$ such that for each $*$-term $\tau(x)$, the standard part of $^*I(\tau)$ is the integral $\int {}^\circ\tau(b)^{\mathcal{M}}\, d\mu(b)$. Define measures μ_n on $^*M^n$ using iterated integrals. This yields a graded model of Φ,

$$\hat{\mathcal{M}} = \langle {}^*M, {}^*R_i, {}^*c_j, \mu_n \rangle,$$

which satisfies the produce measurability axiom almost everywhere. Finally, Theorem 2.3.4 is used, as before, to obtain a probability model $\mathcal{N}$ of Φ. ☐

3.6. Conservative Extension Theorem

We will now show that the logics $L_{\mathbb{A}P}$ and $L_{\mathbb{A}\int}$ are equivalent in a strong sense. This is done by considering their common extension $L_{\mathbb{A}P\int}$.

3.6.1 Definition. $L_{\mathbb{A}P\int}$ is the language which has all the symbols and formation rules of $L_{\mathbb{A}P}$ and $L_{\mathbb{A}\int}$. The satisfaction relation in probability structures is defined as before.

3.6.2 Theorem. *$L_{\mathbb{A}P\int}$ is a conservative definitional extension of both $L_{\mathbb{A}P}$ and $L_{\mathbb{A}\int}$. That is*:

(i) (Conservative): *For any sentence φ in $L_{\mathbb{A}P}$, we have $L_{\mathbb{A}P\int} \models \varphi$ iff $L_{\mathbb{A}P} \models \varphi$. And, for any φ in $L_{\mathbb{A}\int}$, we have $L_{\mathbb{A}P\int} \models \varphi$ iff $L_{\mathbb{A}\int} \models \varphi$.*
(ii) (Definitional): *For each $\varphi(\bar{v})$ in $L_{\mathbb{A}P\int}$, there are $\psi(\bar{v})$ in $L_{\mathbb{A}P}$ and $\theta(\bar{v})$ in $L_{\mathbb{A}\int}$ such that $L_{\mathbb{A}P\int} \models \varphi(\bar{v}) \leftrightarrow \psi(\bar{v})$, and $L_{\mathbb{A}P\int} \models \varphi(\bar{v}) \leftrightarrow \theta(\bar{v})$.*

Remark. The $L_{\mathbb{A}\int}$ case is given in Keisler [1977b] and the $L_{\mathbb{A}P}$ case in Hoover [1978b]. In his work Hoover also gave an axiom set and completeness theorem for $L_{\mathbb{A}P\int}$.

Idea of Proof of Theorem 3.6.2. The proof of part (i) is trivial. Concerning the proof of part (ii), interpretations

$$f: L_{\mathbb{A}P\int} \to L_{\mathbb{A}P} \quad \text{and} \quad g: L_{\mathbb{A}P\int} \to L_{\mathbb{A}\int}$$

can be defined first for atomic formulas $\tau \geq r$ by induction on terms τ, and then by induction on formulas. The idea is to formalize the definitions of integral in terms of measure and vice versa. It is important in this result that $\omega \in \mathbb{A}$, so that the appropriate limits can be expressed in $L_{\mathbb{A}\int}$ and $L_{\mathbb{A}P}$. The finitary analogs of $L_{\mathbb{A}P}$ and $L_{\mathbb{A}\int}$ do not seem to be equivalent. □

This theorem often allows one to convert a theorem about $L_{\mathbb{A}P}$ to a similar theorem about $L_{\mathbb{A}\int}$, and vice versa.

3.6.3 Corollary (Keisler [1977b] and Hoover [1978b]). *Let $\mathcal{M}$ and $\mathcal{N}$ be probability structures for L. The following are equivalent:*

(a) $\mathcal{M} \equiv_{\mathbb{A}P} \mathcal{N}$.
(b) $\mathcal{M} \equiv_{\mathbb{A}\int} \mathcal{N}$.
(c) *For each closed term τ of $L_{\mathbb{A}\int}$, $\tau^{\mathcal{M}} = \tau^{\mathcal{N}}$.*

Proof. The proof of this result follows from the conservative extension and normal form theorems. □

The Barwise completeness and compactness theorems also hold for $L_{\mathbb{A}\int}$. For these one must check that the interpretation functions f and g in the proof of Theorem 3.6.2 are $\mathbb{A}$-recursive.

3.6.4 Theorem (Finite Compactness Theorem (Keisler [1977b])). *Let Φ be a set of sentences of $L_{\mathbb{A}\int}$ of the form $\tau \in [r, s]$. If every finite subset of Φ has a graded model, then Φ has a graded model.*

Proof. The proof follows by an ultraproduct construction. □

The Strong Law of Large Numbers takes a particularly nice form for $L_{\mathbb{A}\int}$.

3.6.5 Theorem (Strong Law of Large Numbers for $L_{\mathbb{A}\int}$). *For any (graded) probability structure $\mathcal{M}$ and term τ with no variables in $L_{\mathbb{A}\int}$,*

$$\lim_{k \to \infty} \tau^{\mathcal{M}(\bar{a}_k)} = \tau^{\mathcal{M}}$$

for $\mu_{\mathbb{N}}$-almost all sequences $\bar{a} \in M^{\mathbb{N}}$.

3.6.6 Definition. When the product measurability axiom, (Axiom D7), is omitted, we obtain the logic *graded* $L_{\mathbb{A}\int}$. Satisfaction in graded probability structures is defined in the natural way.

All the results in Sections 3.5 and 3.6 hold for graded $L_{\mathbb{A}\int}$ and graded $L_{\mathbb{A}P}$.

4. *Logic with Conditional Expectation Operators*

The logic $L_{\mathbb{A}P}$ is not rich enough to express many basic notions from probability theory, notions such as martingale, Markov process, and stopping time. The missing ingredient here is the concept of conditional expectation. In this section, we will develop an enriched language in which these notions can be expressed. It is easier to work with logic having integral operators rather than with probability quantifiers when we add conditional expectations.

4.1. Random Variables and Integrals

We first prepare to extend our logic by introducing a random variable form of $L_{\mathbb{A}\int}$ which is equivalent to the random variable logic $L_{\mathbb{A}P}(\mathbb{R})$ of Section 2.5. In $L_{\mathbb{A}\int}$, each term $\tau(\bar{v})$ is interpreted by an n-fold random variable $\tau^{\mathcal{M}}[\bar{a}]$, and the atomic terms have values in $\{0, 1\}$. In the new logic $L_{\mathbb{A}\int}(\mathbb{R})$ the atomic terms are allowed to have values in $\mathbb{R}$. Let L be the language $L = \{X_i, c_j | i \in I, j \in J\}$.

4.1.1 Definition. The logic $L_{\mathbb{A}\int}(\mathbb{R})$ has atomic terms

$$[X_i(\bar{u}) \upharpoonright r], \qquad \mathbf{1}(u_1 = u_2),$$

where $\bar{u}$ is a tuple of constants or variables and $r \in \mathbb{Q}^+$. The set of *terms* and *formulas* of $L_{\mathbb{A}\int}(\mathbb{R})$ is defined exactly as for $L_{\mathbb{A}\int}$.

The structures for $L_{\mathbb{A}\int}(\mathbb{R})$ are the random variable structures

$$\mathcal{M} = \langle M, X_i^{\mathcal{M}}, c_j^{\mathcal{M}}, \mu \rangle$$

as defined in Section 2.5.

4.1.2 Definition. The *value* $\tau(\bar{a})^{\mathcal{M}}$ of a term $\tau(\bar{v})$ of $L_{\mathbb{A}\int}(\mathbb{R})$ in a random variable structure $\mathcal{M}$ is defined as for $L_{\mathbb{A}\int}$ except for the following new rule for atomic terms:

$$[X_i(\bar{a}) \upharpoonright r]^{\mathcal{M}} = \begin{cases} r & \text{if } X_i^{\mathcal{M}}(\bar{a}) \geq r, \\ -r & \text{if } X_i^{\mathcal{M}}(\bar{a}) \leq -r, \\ X_i^{\mathcal{M}}(\bar{a}) & \text{otherwise.} \end{cases}$$

Thus, $[X_i(\bar{a}) \upharpoonright r]^{\mathcal{M}}$ is equal to $X_i^{\mathcal{M}}(\bar{a})$ truncated at r. The reason the atomic terms are truncated is so that each term will be interpreted by a bounded, and hence integrable, random variable.

4.1.3 Definition. *The axioms and rules of inference of* $L_{\mathbb{A}\int}(\mathbb{R})$ are exactly the same as for $L_{\mathbb{A}\int}$ except that the atomic term axiom, (Axiom D2), is replaced by the following list of axioms, where $\bar{u}$ is an n-tuple of constants or variables.

E1. $\mathbf{1}(u = v) = 0 \vee \mathbf{1}(u = v) = 1.$
E2. $[X_i(\bar{u}) \upharpoonright s] = \min(s, \max(-s, [X(\bar{u}) \upharpoonright r]))$ when $0 \le s \le r$.
E3. $\bigvee_k (|[X_i(\bar{u}) \upharpoonright k + 1]| \le k)$.

This says that $X_i(\bar{u})$ is finite.

E4. For each $m \in \mathbb{N}$,

$$\bigvee_k \int |[X_i(\bar{u}) \upharpoonright k + 1] - [X_i(\bar{u}) \upharpoonright k]| \, d\bar{u} \le \frac{1}{m}.$$

(The probability that $|X_i(\bar{u})| \ge k$ approaches zero as $k \to \infty$.)

We state the main facts without proof.

4.1.4. Theorem (Soundness and Completeness Theorem). *A countable set* Φ *of sentences of* $L_{\mathbb{A}\int}(\mathbb{R})$ *has a model if and only if it is consistent.* ▯

4.1.5 Theorem. *The logics* $L_{\mathbb{A}\int}(\mathbb{R})$ *and* $L_{\mathbb{A}P}(\mathbb{R})$ *for random variable structures have a common conservative definitional extension.* ▯

In other words, $L_{\mathbb{A}\int}(\mathbb{R})$ and $L_{\mathbb{A}P}(\mathbb{R})$ are equivalent logics. Those logics may be generalized to study random variables with values in a Polish space $\mathbb{S}$ instead of in $\mathbb{R}$. The only changes needed are in the definition and axioms for atomic formulas of $L_{\mathbb{A}P}(\mathbb{S})$ and atomic terms of $L_{\mathbb{A}\int}(\mathbb{S})$.

4.2. Conditional Expectations

We will introduce the logic $L_{\mathbb{A}E}(\mathbb{R})$ by adding a conditional expectation operator E to the logic $L_{\mathbb{A}\int}(\mathbb{R})$. A structure for $L_{\mathbb{A}E}(\mathbb{R})$ has the form $(\mathscr{M}, \mathscr{F})$ where $\mathscr{M}$ is a random variable structure and $\mathscr{F}$ is a σ-algebra of measurable subsets of M. We first review the notion of conditional expectation.

4.2.1 Definition. Let $\langle M, S, \mu \rangle$ be a probability space, let $\mathscr{F}$ be a σ-subalgebra of S, and let $g: M \to \mathbb{R}$ be bounded and measurable. A *conditional expectation of* g *with respect to* $\mathscr{F}$ is an $\mathscr{F}$-measurable function $h: M \to \mathbb{R}$ such that for all $B \in \mathscr{F}$, $\int_B g \, d\mu = \int_B h \, d\mu$. It is denoted by $h = E[g | \mathscr{F}]$, or $h(x) = E[g(\cdot) | \mathscr{F}](x)$.

4.2.2 Proposition. (i) *The conditional expectation* $h(x) = E[g(\cdot) | \mathscr{F}](x)$ *exists and is almost surely unique in the sense that any two such functions are equal except on a null set.* ▯

This is a standard consequence of the Radon–Nikodym theorem. Here, h is the Radon–Nikodym derivative of the measure $\nu(B) = \int_B g\, d\mu$, for $B \in \mathscr{F}$, with respect to $\mu \restriction \mathscr{F}$.

We now introduce the logic $L_{\mathbb{A}E}(\mathbb{R})$.

4.2.3 Definition. The logic $L_{\mathbb{A}E}(\mathbb{R})$ has all the formation rules of $L_{\mathbb{A}\int}(\mathbb{R})$ plus the *a* term-builder, the *conditional expectation operator*: If $\tau(u, \bar{v})$ is a term and u, w are individual variables (with u not in $\bar{v}$), then

$$E[\tau(u, \bar{v})\,|\,u](w)$$

is a term in which the occurrences of u are bound and w is free.

This logic has not been considered before in the literature.

4.2.4 Definition. We will use the abbreviation:

$$E[\tau(u, \bar{v})\,|\,u] \quad \text{for} \quad E[\tau(u, \bar{v})\,|\,u](u).$$

Thus, u is free in $E[\tau(u, \bar{v})\,|\,u]$.

The values of a term of $L_{\mathbb{A}E}(\mathbb{R})$ are only almost surely unique in a structure. Here are the details.

4.2.5 Definition. A *conditional expectation structure* for L is a pair $\mathscr{M} = (\mathscr{M}_0, \mathscr{F})$, where $\mathscr{M}_0$ is a random variable structure and $\mathscr{F}$ is a σ-field of μ-measurable sets.

4.2.6 Definition. An *interpretation* of $L_{\mathbb{A}E}(\mathbb{R})$ in a conditional expectation structure $\mathscr{M}$ assigns to each term $\tau(u_1, \ldots, u_n)$ a $\mu^{(n)}$-measurable function $\tau^{\mathscr{M}}: M^n \to \mathbb{R}$ such that

(a) The clauses of the definition of $\tau(\bar{a})^{\mathscr{M}}$ for $L_{\mathbb{A}\int}(\mathbb{R})$ hold.
(b) $(E[\tau(\bar{a}, v)\,|\,v](b))^{\mathscr{M}}$ is $\mu^{(n)} \times \mathscr{F}$-measurable and, for each $\bar{a} \in M^n$,

$$(E[\tau(\bar{a}, v)\,|\,v](b))^{\mathscr{M}} = E[\tau(\bar{a}, \cdot)^{\mathscr{M}}\,|\,\mathscr{F}](b)$$

for μ-almost all b.

4.2.7 Lemma. *For every conditional expectation structure $\mathscr{M}$ for L, there exists an interpretation of $L_{\mathbb{A}E}(\mathbb{R})$ in $\mathscr{M}$, and two interpretations agree almost surely on each term. The values of closed terms and sentences in $\mathscr{M}$ are the same for all interpretations.* □

4.2.8 Definition. The logic $L_{\mathbb{A}E}(\mathbb{R})$ has all the axiom schemes and rules of inference of $L_{\mathbb{A}\int}(\mathbb{R})$ as well as:

F1. $E[\tau(x, \bar{v})\,|\,x](w) = E[\tau(y, \bar{v})\,|\,y](w)$ where x and y do not occur in $\bar{v}$.
F2. $\int E[\sigma(u)\,|\,u] \cdot \tau(u)\, du = \int E[\sigma(u)\,|\,u] \cdot E[\tau(u)\,|\,u]\, du$. This formalizes the definition of conditional expectation.

4.2.9 Definition. The *bound* $\|\tau\|$ of a term τ of $L_{\mathbb{A}E}$ is defined by:

(i) $\|[X_i(\bar{u}) \restriction r]\| = r$.
(ii) $\|\mathbf{1}(x = y)\| = 1$.
(iii) $\|\int \tau \, dx\| = \|\tau\|$.
(iv) $\|E[\tau | u](w)\| = \|\tau\|$.
(v) $\|F(\tau_1, \ldots, \tau_n)\| = \sup\{|F(s_1, \ldots, s_n)| \, | \, |s_i| \leq \| \tau_i \|\}$.
(vi) $\|r\| = r$.

4.2.10 Lemma. $|\tau(\bar{a})^{\mathscr{M}}| \leq \|\tau\|$. ☐

4.2.11 Theorem (Soundness and Completeness Theorem for $L_{\mathbb{A}E}(\mathbb{R})$). *A countable set of sentences of $L_{\mathbb{A}E}(\mathbb{R})$ has a model if and only if it is consistent.*

Proof. The proof of soundness is easy. Let Φ be consistent. Form a new language $K \supseteq L$ by adding a new random variable symbol $X_\tau(\bar{v})$ for each term $\tau(\bar{v})$ of $L_{\mathbb{A}E}(\mathbb{R})$ of the form $E[\sigma | u](w)$. Each such term $\tau(\bar{v})$ translates to the atomic term $[X_\tau(\bar{v}) \restriction n]$ where $n \geq \|\tau\|$; and, hence, each term and sentence of $L_{\mathbb{A}E}(\mathbb{R})$ has a translation in $K_{\mathbb{A}\int}(\mathbb{R})$. Let Ψ be the theory in $K_{\mathbb{A}\int}(\mathbb{R})$ consisting of: all translations of sentences of Φ, all translations of theorems of $L_{\mathbb{A}E}(\mathbb{R})$, and

$$(P\bar{v} \geq 1)[X_\tau(\bar{v}) \restriction r] = [X_\tau(\bar{v}) \restriction s], \qquad r, s \geq \|\tau\|.$$

Ψ is consistent in $K_{\mathbb{A}\int}(\mathbb{R})$ and has a random variable model $\mathscr{M}$. Let $\mathscr{F}$ be the σ-algebra on M generated by the sets

$$\{d \in M | X_\tau(\bar{c}, d)^{\mathscr{M}} \geq r\},$$

where $\tau(\bar{u}, w)$ has the form $E[\sigma(\bar{u}, v) | v](w)$ and $\bar{c}$ is in M. Let $\mathscr{M}_0$ be the reduct of $\mathscr{M}$ to L and $\mathscr{N} = (\mathscr{M}_0, \mathscr{F})$. Using the axioms, it can be shown by induction on τ that if τ has translation σ, then $\tau(\bar{a})^{\mathscr{N}} = \sigma^{\mathscr{M}}(\bar{a})$ is an interpretation of $L_{\mathbb{A}E}(\mathbb{R})$ in $\mathscr{N}$, and thus $\mathscr{N}$ is a model of Φ. ☐

Our treatment of $L_{\mathbb{A}E}(\mathbb{R})$ can be readily extended to logics with two or more conditional expectation operators and to logics with conditional expectation operators on n variables. A case of particular interest is two operators E_1 and E_2 where one σ-algebra is to be contained in another. The author's student, S. Fajardo, has proved the following.

4.2.12 Theorem. *Let Φ be a countable set of sentences in $L_{\mathbb{A}E}(\mathbb{R})$ with two conditional expectation operators. Φ has a model $\mathscr{M} = (\mathscr{M}_0, \mathscr{F}_1, \mathscr{F}_2)$, with $\mathscr{F}_1 \subseteq \mathscr{F}_2$ if and only if Φ is consistent in $L_{\mathbb{A}E}$ with the additional axiom scheme*

$$E_1[\tau(u, \bar{v}) | u] = E_2[E_1[\tau(u, \bar{v}) | u] | u]. \quad ☐$$

4.3. Adapted Probability Logic

We now consider a special case of the logic $L_{\mathbb{A}E}(\mathbb{R})$, a case that is appropriate for the study of stochastic processes. Throughout this section we will assume that μ is a probability measure on M and β is the Borel measure on $[0, 1]$. By a (continuous time) *stochastic process* we mean a $(\mu \otimes \beta)$-measurable function

$$X: M \times [0, 1] \to \mathbb{R}.$$

In probability theory, the evolution of a stochastic process over time is studied by means of an adapted probability space as defined below.

4.3.1 Definition. If $B \subseteq M \times [0, 1]$ and $t \in [0, 1]$, the *section* B_t is the set $B_t = \{w \in M \mid \langle w, t\rangle \in B\}$.

4.3.2 Definition. An *adapted* (*probability*) *space* (or stochastic base) is a structure

$$\mathscr{S} = \langle M, \mu, \mathscr{F}_t\rangle_{t \in [0,1]},$$

where:

(a) μ is a probability measure on M.
(b) Each $\mathscr{F}_t$ is a σ-algebra of μ-measurable subsets of M.
(c) For each $t \in [0, 1]$, $\mathscr{F}_t = \bigcap_{s>t} \mathscr{F}_s$, that is, $\mathscr{F}_t$ is increasing and right continuous.

The family of σ-algebras $\langle \mathscr{F}_t \mid t \in [0, 1]\rangle$ is called the *filtration* of $\mathscr{S}$. Intuitively, M is the set of possible states of the world, and a set $B \subseteq M$ belongs to $\mathscr{F}_t$ if B is an event whose outcome is determined at or before time t.

Adapted spaces have been extensively studied in the literature (see, for example, Dellacherie–Meyer [1981], or Metivier–Pellaumail [1980]).

4.3.3 Definition. Let L be a set of "stochastic process" symbols X_i, $i \in I$. An *adapted* (*probability*) *structure* for L is a structure

$$\mathscr{M} = \langle M, X_i, \mu, \mathscr{F}_t\rangle_{i \in I, t \in [0,1]},$$

such that $\langle M, \mu, \mathscr{F}_t\rangle$ is an adapted space and each $X_i^{\mathscr{M}}: M \times [0, 1] \to \mathbb{R}$ is a stochastic process on $\langle M, \mu\rangle$.

4.3.4 Definition. The *adapted probability logic* $L_{\mathbb{A}\,\mathrm{ad}}(\mathbb{R})$, or more briefly L_{ad}, is a two-sorted form of $L_{\mathbb{A}E}(\mathbb{R})$ with just one variable w of the first sort, and countably many variables $t_1, t_2, \ldots$ of the second sort, called *time variables*. The non-logical symbols of L are stochastic process symbols X_i with just one argument place of each sort. L_{ad} has no equality symbol. This logic was introduced in Keisler [1979] and has been studied further in Rodenhausen [1982].

4.3.5 Definition. The *terms* of L_{ad} are as follows, where s, t are time variables.

(a) For each $r \in \mathbb{Q}^+$, $[X_i(w, t)\upharpoonright r]$ is an atomic term.
(b) Each time variable t is a term.
(c) Each real $r \in \mathbb{A} \cap \mathbb{R}$ is a term.
(d) If τ is a term, so are

$$\int \tau \, dw, \quad \int \tau \, ds, \quad E[\tau \,|\, s](w, t).$$

(e) If $\tau_1, \ldots, \tau_n$ are terms and $F \in C_{\mathbb{A}}(\mathbb{R}^n)$, then $F(\tau_1, \ldots, \tau_n)$ is a term.

For each term τ, $\tau \geq 0$ is an atomic formula, and formulas are closed under $\neg$ and $\wedge$.

4.3.6 Definition. The adapted structure

$$\mathcal{M} = \langle M, X_i, \mu, \mathcal{F}_t \rangle$$

for L is identified with the two-sorted conditional expectation structure

$$\mathcal{M} = \langle M, [0, 1], X_i^{\mathcal{M}}, \mu, \beta, \mathcal{F} \rangle.$$

Here, β is Borel measure on $[0, 1]$, and $\mathcal{F}$ is the σ-algebra on $M \times [0, 1]$ generated by the set of $(\mu \otimes \beta)$-measurable sets B such that for each t, $B_t \in \mathcal{F}_t$ and $B_t = \bigcap_{s>t} B_s$. $\mathcal{F}$ is called the *optional σ-algebra.*

4.3.7 Definition. The notion of an interpretation $\tau^{\mathcal{M}}$ of a term $\tau(w, \bar{t})$ in an adapted structure $\mathcal{M}$ is defined as in Definition 4.2.6 for $L_{\mathbb{A}E}(\mathbb{R})$, but with the following stronger clause for the conditional expectation operator.

For each term $\tau(w, s, \bar{b})$ with n parameters $\bar{b}$ from $[0, 1]$, (b1) through (b3) hold:

(b1) $(E[\tau(w, s, \bar{b}) \,|\, s](w, a))^{\mathcal{M}}$ *is* $\mathcal{F} \otimes \beta^n$-measurable.
(b2) For each $\bar{b}$, $(E[\tau(w, s, \bar{b}) \,|\, s](w, a))^{\mathcal{M}} = E[\tau(\cdot, \cdot, \bar{b})^{\mathcal{M}} \,|\, \mathcal{F}](w, a)$ $(\mu \otimes \beta)$-almost surely.
(b3) For each $\bar{b}$ and $a \in [0, 1]$, $(E[\tau(w, s, \bar{b}) \,|\, s](w, a))^{\mathcal{M}} = E[\tau(\cdot, a, \bar{b})^{\mathcal{M}} \,|\, \mathcal{F}_a](w)$ μ-almost surely.

4.3.8 Lemma. *Every adapted structure $\mathcal{M}$ has an interpretation $\tau \mapsto \tau^{\mathcal{M}}$. For each term $\tau(w, \bar{s})$ and all tuples $\bar{a}$ in $[0, 1]$, any two interpretations agree at $\tau(w, \bar{a})$ for μ-almost all w. In particular, if w is not free in $\tau(\bar{a})$, then any two interpretations in $\mathcal{M}$ agree at $\tau(\bar{a})$ for all $\bar{a}$ in $[0, 1]$.*

Idea of Proof. The main difficulty here lies in proving the existence of an interpretation by induction on τ, at the conditional expectation step. We use the fact that for any random variable $f(w)$, $E[f(\cdot) \,|\, \mathcal{F}_t](w)$ has a right continuous version, and any right continuous process is measurable in the optional σ-algebra $\mathcal{F}$ (see Dellacherie–Meyer [1981]). This done, we then show that $E[g(\cdot, t) \,|\, \mathcal{F}_t](w)$ is $\mathcal{F}$-measurable by applying the monotone class theorem. ∎

Remark. In case L is the empty language, the adapted structures for L are just the adapted spaces. In this case, the value of each term $\tau^{\mathscr{M}}(w, \bar{a})$ depends only on $\bar{a}$ and not on w or $\mathscr{M}$.

4.4. Examples

As an indication of the expressive power of the logic L_{ad}, we formalize some central notions from the theory of stochastic processes. In each example, the process $\tau^{\mathscr{M}}$ has the stated property if and only if the formula holds for all s, t in $[0, 1]$. We use the abbreviations

$$E[\tau \mid s] \quad \text{for} \quad E[\tau \mid s](w, s),$$

and

$$\sigma(s) = \tau(s) \text{ a.s.} \quad \text{for} \quad \int |\sigma(s) - \tau(s)| \, dw \leq 0.$$

(1) $\sigma(s)$ is a version of $\tau(s)$: $\sigma(s) = \tau(s)$ a.s.
(2) $\tau(s)$ is adapted: $\tau(s) = E[\tau(s) \mid s]$ a.s.
(3) $\tau(s)$ is a martingale: $s \leq t \rightarrow \tau(s) = E[\tau(t) \mid s]$ a.s.; recall that s and t are terms of L_{ad}.
(4) $\tau(s)$ is a submartingale: $\tau(s) \leq E[\tau(t) \mid s]$ a.s.
(5) $\tau(s)$ is Markov process with continuous transition function $F \mapsto T_F$ (a Feller process): For each $F \in C_{\mathbb{A}}(\mathbb{R})$,

$$s \leq t \rightarrow E[F(\tau(t)) \mid s] = T_F(s, t, \tau(s)) \text{ a.s.}$$

(6) $\tau(w)$ is a stopping time: $\min(\tau, s) = E[\min(\tau, s) \mid s]$ a.s.
(7) X is a Brownian motion (X is not bounded, so this would have to be modified to fit within the language L_{ad}):
   (a) X is a martingale
   (b) $s = 0 \rightarrow X(s) = 0$ a.s.
   (c) $s \leq t \rightarrow E[(X(t) - X(s))^2 \mid s] = t - s$ a.s.
   (d) $s \leq t \rightarrow E[F(X(t) - X(s)) \mid s] = \int F(X(t) - X(s)) \, dw$ a.s.; that is, $X(t) - X(s)$ is independent of $\mathscr{F}_s$.

4.5. Axioms and Completeness

4.5.1 Definition. The logic L_{ad} has all the axiom schemes and rules of inference for two-sorted $L_{\mathbb{A}E}(\mathbb{R})$ (with only one variable of the first sort, and E applied to one variable of each sort) as well as:

G1. For any $F \in C_{\mathbb{A}}([0, 1])$ with $\int_0^1 F(x) \, dx = r$,

$$\int F(t) \, dt = r.$$

G2. $s \leq t \to E[\tau|s] = E[E[\tau|s]|t]$; that is, $s \leq t$ implies $\mathscr{F}_s \subseteq \mathscr{F}_t$.
G3. $\bigwedge_m \bigvee_n \iiint |\tau(w, s) - \tau(w, t)| \cdot \max(0, 1 - |s - t| \cdot n)\, ds\, dt\, dw \leq 1/m \cdot n$. That is, τ is $(\mu \otimes \beta)$-measurable. Intuitively, on a small diagonal strip $\{\langle w, s, t\rangle : |s - t| \leq 1/n\}$, $\tau(w, s)$ is usually close to $\tau(w, t)$.

This set of axioms is essentially due to Rodenhausen [1982].

4.5.2 Theorem (Soundness and Completeness Theorem for L_{ad} (Rodenhausen [1982])). *A countable set Φ of sentences of L_{ad} has a model if and only if it is consistent.* ☐

The proof of Rodenhausen is direct and quite long. A fairly short alternative proof can be given using the completeness theorem for the two-sorted logic $L_{\mathbb{A}E}(\mathbb{R})$. The idea is to add an extra stochastic process symbol $I(t)$ to L to represent the term t. A two-sorted model $\mathscr{M}$ for $L_{\mathbb{A}E}$ is made into an adapted model by using $I^{\mathscr{M}}$ to replace the second universe of $\mathscr{M}$ by [0, 1]. The extra axioms G1 through G3 are needed at that point.

The Barwise completeness and compactness theorems, and the finite compactness theorem, carry over to L_{ad}.

4.6. *Elementary Equivalence in Adapted Probability Logic*

There are two natural notions of elementary equivalence in L_{ad}.

4.6.1 Definition. Let $\mathscr{M}$ and $\mathscr{N}$ be adapted probability structures for L.

(i) $\mathscr{M}$ and $\mathscr{N}$ are *weakly L_{ad}-equivalent*, in symbols,

$$\mathscr{M} \equiv^{w} \mathscr{N},$$

if $\mathscr{M}$ and $\mathscr{N}$ satisfy the same sentences of L_{ad}.

(ii) $\mathscr{M}$ and $\mathscr{N}$ are *strongly L_{ad}-equivalent*,

$$\mathscr{M} \equiv^{s} \mathscr{N},$$

if for each tuple $\bar{a}$ in [0, 1] and formula $\varphi(\bar{t})$ of L_{ad} in which w is not free,

$$\mathscr{M} \models \varphi[\bar{a}] \quad \text{iff} \quad \mathscr{N} \models \varphi[\bar{a}].$$

4.6.2 Proposition. *Any two adapted spaces (adapted structures for $L = \varnothing$) are strongly L_{ad}-equivalent.* ☐

The strong L_{ad}-equivalence relation is more important than weak L_{ad}-equivalence, because each adapted structure has the same second universe [0, 1]. Each notion in Example 4.4 is preserved under strong L_{ad}-equivalence but not under weak L_{ad}-equivalence. Following are some useful characterizations of these relations.

4.6.3 Proposition (Hoover and Keisler [1984]). *The following ar equivalent:*

(a) $\mathscr{M} \equiv^w \mathscr{N}$.
(b) *There is a set* $T \subseteq [0, 1]$ *of measure one such that for each* $\bar{a}$ *in* T *and formula* $\varphi(s)$ *of* L_{ad}, $\mathscr{M} \models \varphi[\bar{a}]$ *iff* $\mathscr{N} \models \varphi[\bar{a}]$.
(c) *For each term* $\tau(\bar{s})$ *of* L_{ad} *with no integrals over time variables,* $\tau(\bar{a})^{\mathscr{M}} = \tau(\bar{a})^{\mathscr{N}}$, *for almost all* $\bar{a}$ *in* $[0, 1]$. ☐

4.6.4 Proposition. *The following are equivalent:*

(a) $\mathscr{M} \equiv^s \mathscr{N}$.
(b) *For each term* $\tau(\bar{s})$ *with no integrals over time variables, and all* $\bar{a}$ *in* $[0, 1]$, $\tau(\bar{a})^{\mathscr{M}} = \tau(\bar{a})^{\mathscr{N}}$. ☐

The function $\tau(\bar{a}) \mapsto \tau(\bar{a})^{\mathscr{M}}$ is called the *adapted distribution* of $\mathscr{M}$ and it is analogous to the distribution of a random variable. Most stochastic processes which arise naturally are right continuous (in t for almost all w). For right continuous processes the two notions of L_{ad}-equivalence coincide.

4.6.5 Theorem (Hoover and Keisler [1984]). *If* $\mathscr{M} \equiv^w \mathscr{N}$ *and each stochastic process* $X_i^{\mathscr{M}}$ *and* $X_i^{\mathscr{N}}$ *is right continuous, then* $\mathscr{M} \equiv^s \mathscr{N}$. ☐

Brownian motion plays a central role in the study of stochastic processes. The following result shows that the L_{ad}-theory of independent Brownian motions is complete.

4.6.6 Theorem (Keisler [1984]). *Let* $\mathscr{M}$ *and* $\mathscr{N}$ *be adapted structures for* L *whose stochastic processes are mutually independent Brownian notions. Then* $\mathscr{M} \equiv^s \mathscr{N}$. ☐

4.7. Robinson Consistency and Craig Interpolation

The results of this section are all from the paper Hoover and Keisler [1982], a paper which studies L_{ad}-equivalence and which gives its applications to the theory of stochastic processes. The following notion corresponds to saturated structures in first-order model theory, except that stochastic processes take the place of both relations and constants.

4.7.1 Definition. An adapted space

$$\mathscr{S} = \langle M, \mu, \mathscr{F}_t \rangle$$

is *saturated* if whenever $L^1 \subseteq L^2$, $\mathscr{M}^1$ is an expansion of $\mathscr{S}$ to L^1, $\mathscr{N}^1 \equiv^s \mathscr{M}^1$, and $\mathscr{N}^2$ is an expansion of $\mathscr{N}^1$ to L^2, there exists an expansion $\mathscr{M}^2$ of $\mathscr{M}^1$ to L^2, such that $\mathcal{N}^2 \equiv^s \mathcal{M}^2$. The space $\mathcal{S}$ is *weakly saturated* if the above condition holds with weak L_{ad}-equivalence instead of strong L_{ad}-equivalence.

4.7.2 Proposition. *Every saturated adapted space $\mathcal{S}$ is universal; that is, for every adapted structure $\mathcal{N}$, there is an expansion $\mathcal{M}$ of $\mathcal{S}$ with $\mathcal{M} \equiv^s \mathcal{N}$. Furthermore, every weakly saturated adapted space is weakly universal.*

Proof. Take $L^1 = \emptyset$. ◻

4.7.3 Proposition (Hoover and Keisler [1984]). *Every saturated adapted space is weakly saturated.* ◻

4.7.4 Definition. An *adapted Loeb space* is an adapted space

$$\langle M, \mu, \mathcal{F}_t \rangle_{t \in [0,1]}$$

such that for some internal $*$-adapted space

$$\langle M, \nu, \mathcal{G}_s \rangle_{s \in {}^*[0,1]},$$

with universe M, μ is the completion of the Loeb measure of ν and $\mathcal{F}_t$ is the σ-algebra generated by

$$\bigcup_{{}^\circ s = t} \mathcal{G}_s \cup (\text{null sets of } \mu).$$

The following theorem is the main result in Hoover and Keisler [1984].

4.7.5 Theorem. *Every adapted Loeb space which admits a Brownian motion is saturated.* ◻

Remark. Anderson [1976] constructed an adapted Loeb space which admits a Brownian motion. Hence, saturated adapted probability spaces exist.

4.7.6 Theorem (Robinson Consistency Theorem for L_{ad}). *Let $L^0 = L^1 \cap L^2$, and let $\mathcal{M}^1$, $\mathcal{M}^2$ be adapted structures for L^1 and L^2 such that $\mathcal{M}^1 \restriction L^0 \equiv^s \mathcal{M}^2 \restriction L^0$. Then there is an adapted structure $\mathcal{N}$ for $L^1 \cup L^2$ such that $\mathcal{N} \restriction L^1 \equiv^s \mathcal{M}^1$, and $\mathcal{N} \restriction L^2 \equiv^s \mathcal{M}^2$. A similar result holds for weak L_{ad}-equivalence.*

Proof. There is an adapted structure $\mathcal{N}^1 \equiv^s \mathcal{M}^1$ on any saturated space. Then $\mathcal{N}^1 \restriction L^0 \equiv^s \mathcal{M}^2 \restriction L^0$; so, by saturation, there is an expansion $\mathcal{N}^2$ of $\mathcal{N}^1 \restriction L^0$ with $\mathcal{N}^2 \equiv^s \mathcal{M}^2$. Let $\mathcal{N}$ be the common expansion of $\mathcal{N}^1$ and $\mathcal{N}^2$. ◻

4.7.7 Theorem. *The Craig interpolation theorem holds for L_{ad}, with or without time constants from* [0, 1].

Proof. As in Theorem 3.6.2 for $L_{\mathbb{A}P}$, we first use a Henkin construction and then apply Robinson consistency. ☐

Following is a characterization of $\mathscr{M} \equiv^s \mathscr{N}$ as a coarsest equivalence relation in the style of soft model theory.

4.7.8 Theorem (Hoover). *Let $\approx$ be an equivalence relation on adapted structures for L with the following properties for all adapted structures $\mathscr{M}$, $\mathscr{N}$ for L:*

(a) *If $L^0 \subseteq L$ and $\mathscr{M} \approx \mathscr{N}$, then $\mathscr{M} \restriction L^0 \approx \mathscr{N} \restriction L^0$.*
(b) *If $\mathscr{M} \approx \mathscr{N}$, then for each term $\tau(w, \bar{t})$ with no integral or conditional expectation operators and all $\bar{a}$ in $[0, 1]$, $(\int \tau(w, \bar{a})\, dw)^{\mathscr{M}} = (\int \tau(w, \bar{a})\, dw)^{\mathscr{N}}$.*
(c) *If $X_i^{\mathscr{M}}$ is a martingale and $\mathscr{M} \approx \mathscr{N}$, then $X_i^{\mathscr{N}}$ is a martingale.*
(d) *The relation $\approx$ has the Robinson consistency property.*

Then $\mathscr{M} \approx \mathscr{N}$ implies $\mathscr{M} \equiv^s \mathscr{N}$. ☐

Aldous [198?] introduced the notion of *synonymous* adapted structures. $\mathscr{M}$ and $\mathscr{N}$ are *synonymous* if $\tau(\bar{a})^{\mathscr{M}} = \tau(\bar{a})^{\mathscr{N}}$, for each term $\tau(\bar{v})$ with at most one conditional expectation operator and each $\bar{a}$ in $[0, 1]$. He showed that each property in Section 4.4 is preserved under synonymity. In Hoover–Keisler [1982] there is an example of two synonymous adapted structures which are not weakly L_{ad}-equivalent. It follows from Theorem 4.7.8 that the Robinson consistency property fails for synonymity.

A theory of hyperfinite adapted structures has been developed in Keisler [1979] and Rodenhausen [1982] with results that parallel those on hyperfinite probability structures in Section 3.5, for both $\equiv^w$ and $\equiv^s$.

The adapted Loeb structures have a number of applications to standard probability theory, this is particularly true of existence theorems for stochastic differential equations where the richness of the space is necessary. See Cutland [1982], Hoover–Perkins [1983a, b], Keisler [1984], Kosciuck [1982], T. Lindstrom [1980a–d], and Perkins [1982].

Our treatment of adapted probability logic can be extended in several ways such as the following:

(a) The optional σ-algebra $\mathscr{F}$ may be replaced by any σ-algebra $\mathscr{G} \supseteq \mathscr{F}$ of $(\mu \otimes \beta)$-measurable sets such that for each $U \in \mathscr{G}$ and $t \in [0, 1]$, $U_t \in \mathscr{F}_t$. Each interpretation in $(\mathscr{M}, \mathscr{F})$ is then an interpretation in $(\mathscr{M}, \mathscr{G})$, and hence $(\mathscr{M}, \mathscr{F}) \equiv (\mathscr{M}, \mathscr{G})$.
(b) The language L has constant time symbols c_r, $r \in \mathbb{A} \cap [0, 1]$, which occur in place of time variables (only finitely many in a single formula). The additional axiom scheme is

$$\text{G4.}\quad c_r = r.$$

(c) The time variables range over $[0, \infty)$ instead of $[0, 1]$. Changes must be made since β is no longer a probability measure.

5. Open Questions and Research Problems

Following is a list of questions and problems which suggest some fruitful areas of research with respect to some of the notions and relationships that were examined in this chapter.

Problem 1. Develop a form of $L_{\mathbb{A}P}$ which has the universal quantifier $(\forall x)$.

Three ways to add $(\forall x)$ so that the satisfaction relation behaves properly are:

(a) Restrict to absolutely Borel structures as indicated at the end of Section 2.
(b) Add $(\forall x)$ to $L_{\mathbb{A}P}$ with the restriction that no universal quantifier may occur within the scope of a probability quantifier.
(c) Add $(\forall x)$ to $L_{\mathbb{A}\int}$ with no restrictions.

None of our major proofs carry over to these logics, because the Loeb measure construction does not preserve truth values involving $(\forall x)$.

Problem 2. Develop a logic with $(\forall x)$ and quantifiers for inner measure at least r and outer measure at least r.

Since inner and outer measure are defined for all subsets of M, there is no difficulty in defining the satisfaction relation.

Problem 3. Study a logic such as $L_{\mathbb{A}P}$ for structures with infinite measures instead of probability measures.

Problem 4. Study a logic such as $L_{\mathbb{A}P}$ for structures with two measures (and corresponding quantifiers). Obtain completeness theorems for structures with two measures μ, ν such that:

(a) μ is orthogonal to ν.
(b) μ is absolutely continuous with respect to ν.

Problem 5. Define hyperfine conditional expectation structures appropriately and prove an existence and uniqueness theorem for $L_{\mathbb{A}E}$.

Problem 6. Does $L_{\mathbb{A}E}$ have the Robinson consistency and/or the Craig interpolation property?

Problem 7. Extend the results for adapted probability logic to allow universal quantifiers $(\forall t)$ for the second sort $[0, 1]$.

Problem 8. Study various operations on probability structures from the viewpoint of the logics examined in this chapter.

A small beginning for $L_{\mathbb{A}\int}$ is in Keisler [1977b].

Problem 9. The results on graded $L_{\mathbb{A}P}$ carry over without difficulty when L has function symbols (Hoover [1978b]). Do the results on $L_{\mathbb{A}P}$ carry over when L has function symbols?

The difficulty lies in the proof of Theorem 2.3.4.

Problem 10. Reexamine abstract model theory in the light of logics such as $L_{\mathbb{A}P}$.

The hypotheses for a logic in the enriched abstract model theory of Mundici, Chapter VIII, fail badly for $L_{\mathbb{A}P}$ and the other logics of this chapter. Mundici proved (under the set-theoretic assumption $\natural$) that every logic with relativization which has the Robinson consistency property is compact. Since $L_{\mathbb{A}P}$ is not compact, it follows that no extension of $L_{\mathbb{A}P}$ which is a logic with relativization in the sense of Mundici has the Robinson consistency property.

The logic $L_{\mathbb{A}P}$ does not have universal quantifiers and does not allow function symbols. The relativization property holds only for relativizing to a set of positive measure. Moreover, there does not seem to be a way to make the class of probability structures into a semantic domain in the sense of Mundici. Closure under strict expansion fails. The natural notions of isomorphic embedding which come to mind fail to satisfy either factorization, or existence and closure under disjoint union.

An essential characteristic of $L_{\mathbb{A}P}$ is that sets of measure zero are unimportant. It appears that to prove that Robinson consistency implies compactness, constructions are needed which make sets of measure zero important.

Problem 11. Is there any equivalence relation $\approx$ on adapted structures which satisfies conditions (a)–(d) of Theorem 4.7.8, is strictly finer than $\equiv^s$, and is strictly coarser than $\cong$? Here $h: \mathscr{M} \cong \mathscr{N}$ means that h sends μ to v modulo null sets, and for all t, $h(\mathscr{F}_t) = \mathscr{G}_t$ modulo null sets, and $X_i(w, t) = X_i(hw, t)$ for μ-almost all w.

Added in proof: Problems 3, 4, 5, and 6 were solved while this article was in press. M. Rašković solved Problem 3 in the forthcoming paper "Model Theory for $L_{\mathbb{A}M}$ Logic". M. Rašković and R. Zivaljevič solved Problem 4 in "Barwise Completeness for Biprobability Logics". S. Fajardo will publish affirmative solutions to Problems 5 and 6 in "Probability Logic with Conditional Expectation".