Axiomatic First-Order Probability

Kathryn Blackmond Laskey

Department of Systems Engineering and Operations Research
George Mason University, Fairfax VA 22030, USA
klaskey@gmu.edu

Abstract. Most languages for the Semantic Web have their logical basis in some fragment of first-order logic. Thus, integrating first-order logic with probability is fundamental for representing and reasoning with uncertainty in the semantic web. Defining semantics for probability logics presents a dilemma: a logic that assigns a real-valued probability to any first-order sentence cannot be axiomatized and lacks a complete proof theory. This paper develops a first-order axiomatic theory of probability in which probability is formalized as a function mapping Gödel numbers to elements of a real closed field. The resulting logic is fully first-order and recursively axiomatizable, and therefore has a complete proof theory. This gives rise to a plausible reasoning logic with a number of desirable properties: the logic can represent arbitrarily fine-grained degrees of plausibility intermediate between proof and disproof; all mathematical and logical assumptions can be explicitly represented as finite computational structures accessible to automated reasoners; contradictions can be discovered in finite time; and the logic supports learning from observation.

Keywords: First-Order Logic, Probability.

1 Introduction

Logic-based languages have long been recognized as an effective means to represent information clearly, unambiguously, and in a manner that facilitates processing by machines. By far the most common logical basis for Semantic Web languages is classical first-order logic (FOL). This is no accident: its clear syntax, well-understood semantics, and complete proof theory make FOL a natural choice for computational knowledge representation and reasoning. However, FOL lacks a fundamental capability essential for semantically aware systems. As Jeffreys [1] put it, "Traditional or deductive logic admits only three attitudes to any proposition: definite proof, disproof, or blank ignorance." An intelligent reasoner must do more: it must assess the plausibility of uncertain hypotheses, make reasonable choices when the outcome is uncertain, and use observations to improve its representation of the world.

Probability is the unique plausible reasoning calculus that satisfies certain intuitively satisfying axioms of coherent reasoning (e.g., [2]). For this reason, probability has achieved a privileged status for plausible reasoning akin to FOL's privileged status with respect to logical reasoning. The past few decades have given rise to increasingly expressive probability-based languages, as well as a host of

restricted languages designed for scalability. There is increasing interest in probability for semantic web applications [3].

It is often taken for granted that a new kind of logic is needed to capture essential aspects of plausible reasoning: "Ordinary logic seems to be inadequate by itself to cope with problems involving beliefs. In addition a theory of probability is required" [4]. Because of their built-in machinery for reasoning about functions, higher-order logic has been proposed as a natural logical basis for combining probability and logic [5]. On the other hand, its complete proof theory makes first-order logic attractive as a computational logic. Moreover, it is attractive to use the same logic to reason both about the domain itself and about the plausibility of statements about the domain. The question thus arises of whether an adequate formalization of probability is possible within first-order logic itself.

This paper formalizes, within standard first-order logic, a probabilistic logic powerful enough to express uncertainty about arbitrary first-order sentences. By formalizing probability as an axiomatic first-order theory, probabilities can be associated coherently with arbitrary first-order sentences, with no modification of traditional first-order semantics. To stay within axiomatic first-order logic, probabilities are defined not as real numbers, but as elements of a real closed field. An axiom schema is added to the standard axioms of the probability calculus to give "logical teeth" to the idea that probability zero events do not happen. The semantics proposed here connects naturally to algorithmic notions of randomness as proposed by Kolmogorov and Martin-Löf [6], [7], as well as to Dawid's [8] calibration criterion.

2   First-Order Probability

We begin with a first-order language $\mathscr{L}$ used to make assertions about a domain. $\mathscr{L}$ includes the usual logical symbols (variables, logical connectives, universal and existential quantifiers), together with a set of domain-specific predicate, function and constant symbols. Without loss of generality, $\mathscr{L}$ is taken to be a traditional, untyped first-order language.[1] To operationalize the requirement that assertions be expressible as a finite computational structure, a knowledge base (KB) is taken to be an axiomatic theory of $\mathscr{L}$. That is, a KB contains a consistent, recursive set A of sentences of $\mathscr{L}$. The logical consequences of these axioms comprise a recursively enumerable set $T_A =$ Cn(A), called the *theory* of A. Gödel's completeness theorem implies that T_A is equal to the set $\{\sigma : A \vdash \sigma\}$ of sentences provable from A.

In general, a KB may be incomplete – it need not imply a definite truth-value for every sentence. In fact, a sufficiently powerful theory is necessarily incomplete. It is useful for a reasoner to grade the plausibility of propositions it can neither prove nor disprove. Probability is a natural candidate for this purpose.

It seems reasonable to define probability as a function mapping each sentence to a real number between zero and 1, in a manner that satisfies the standard identities of probability theory, and so that sentences provable (disprovable) from A are assigned

---

[1] It is well known that a typed first-order logic can be re-expressed via a syntactic transformation as an untyped logic (cf., [7]).

probability 1 (0). This approach, natural as it seems, runs into difficulty. The first roadblock is that in standard first-order logic, arguments of functions must be elements of the domain, not sentences or propositions. The second roadblock is that the theory of the real numbers cannot be fully characterized as an axiomatic first-order theory. Several authors have shown that formalizing probabilities as a real numbers in the unit interval results in a theory that cannot be axiomatized, and that does not admit a complete proof theory (cf., [10], [11]). On the other hand, by abandoning the requirement that probabilities be real-valued, Bacchus [12] developed axiomatic probability logics that have a complete proof theory.

Given the mathematical impossibility of both an axiomatic first-order theory and a function mapping sentences to real numbers, which should be preferred? To answer this question, we step back to first principles, and consider fundamental requirements for a computational probabilistic logic. First, a computational logic should explicitly represent all mathematical and logical assumptions as finite computational structures accessible to an automated reasoner. Second, it should be possible for a reasoner to discover contradictions in a knowledge base, to identify when observations are inconsistent with a theory, and to prove any consequence entailed by a theory. Third, a logic for plausible reasoning must be able to associate measures of plausibility with propositions, to express degrees of plausibility intermediate between proof and disproof, and to do so in a logically coherent manner. All these requirements are met by the proposed formalism. Furthermore, the first two requirements are automatically satisfied if probability is formalized as a traditional first-order axiomatic theory, while the final requirement can be met without demanding either that probability be formalized as a function on sentences, or that probability values be real numbers. Hence, a first-order axiomatic theory is a fundamental requirement, whereas a function from sentences to real numbers is dispensable.

If probability is not a function mapping sentences to real numbers, then what is it? We formalize probability as a function mapping Gödel numbers to elements of a real-closed field (RCF). A RCF is the closest one can come to formalizing the real numbers within first-order logic. The real numbers are uniquely characterized up to isomorphism as an ordered field with the least upper bound property. The ordered field axioms formalize familiar properties of the real and rational numbers: addition, multiplication, additive and multiplicative inverses (hence, subtraction and division), distribution of multiplication over addition, and complete ordering. These axioms can be formalized fully in FOL. The defining property of the real numbers, that every bounded non-empty set of real numbers has a least upper bound, is not a first-order property. In a RCF, the least upper bound property holds for all definable relations. The RCF axioms are sufficient to characterize all first-order properties of the real numbers (cf., [13]). Thus, we assume probabilities are elements of a RCF.

Gödel showed that, given a sufficiently powerful formal system, domain elements (e.g., numbers) can be associated with sentences, formulas, and proofs. This device allows indirect expression of and reasoning about logical notions such as proof and consistency, while complying with FOL's prohibition against direct reference to sentences. Defining probabilities as a mapping from Gödel numbers to elements of a RCF allows us to develop a fully first-order axiomatization of probability. We argue later that our axioms capture the essential requirements for a computational logic of plausible reasoning.

3   The Probability Axioms

The original language $\mathscr{L}$ and axioms A are augmented with additional symbols and axioms to provide the necessary logical apparatus for probabilistic reasoning. The augmented language and axioms are called $\mathscr{L}^*$ and A^*, respectively.

3.1   The Language

The language $\mathscr{L}^*$ has numerical constants 0 and 1; arithmetic ordering predicate $\leq$; arithmetic operators $+$ and $\times$; one-place predicate symbols $\mathcal{R}$ and $\mathcal{N}$ to represent real and natural numbers; and the two-place function symbol $\mathcal{P}$ to represent probability. In addition, there is a predicate $\mathcal{D}$ to represent elements of the domain; and a countable collection $\mathcal{L}_1, \mathcal{L}_2, \ldots$ of labels as names for individuals. If $\mathscr{L}$ already has mathematical symbols and mathematical axioms consistent with our probability axioms, we can make use of the existing logical machinery; otherwise new symbols are added.

3.2   The axioms

Domain axioms. Our first step is to relativize the domain axioms to $\mathcal{D}$. This is achieved through a standard syntactic translation, e.g., $\forall x \; \varphi(x)$ becomes $\forall x \; \mathcal{D}(x){\rightarrow}\varphi(x)$, and the sentence $\exists x \; \varphi(x)$ becomes $\exists x \; \mathcal{D}(x){\wedge}\varphi(x)$ (c.f., [9]).

Integer arithmetic. We require enough integer arithmetic to allow Gödel numbering and reasoning about provability. The following axioms, together with the RCF axioms defined below (which apply to natural numbers by virtue of inclusion) serve this purpose:

   N1. $\forall x \; \mathcal{N}(x){\rightarrow}\mathcal{R}(x)$
   N2. $\mathcal{N}(0)$
   N3. $\forall x \; \mathcal{N}(x){\rightarrow}\mathcal{N}(x{+}1)$
   N4. $\forall x \; \forall y \; \mathcal{N}(x) \wedge \mathcal{N}(y) \rightarrow ((x < y{+}1) \rightarrow (x \leq y))$
   N5. $\forall x \; \mathcal{N}(x){\rightarrow}\neg(x < 0)$
   N6. All universal closures of formulas $\mathcal{N}(x){\rightarrow}(\varphi(0){\wedge}\forall x(\varphi(x) \rightarrow \varphi(x{+}1))) \rightarrow$
   $\forall x \; \varphi(x)$, where $\varphi(x)$ has x (and possibly other variables) free.

Real closed field axioms. As described above, probabilities are formalized as elements of a real closed field (RCF), the first-order theory of the real numbers. Axioms for a real closed field can be found in [13]. Note that by virtue of being natural numbers, the constants 0 and 1 are also real numbers.

   R1. *Additive and multiplicative closure*: $\forall x \; \forall y \; \mathcal{R}(x){\wedge}\mathcal{R}(y) \rightarrow \mathcal{R}(x{+}y) \wedge \mathcal{R}(x \cdot y)$
   R2. *Commutativity*: $\forall x \; \forall y \; \mathcal{R}(x){\wedge}\mathcal{R}(y) \rightarrow (x{+}y = y{+}x) \wedge (x \cdot y = y \cdot x)$
   R3. *Associativity*: $\forall x \forall y \forall z \; \mathcal{R}(x){\wedge}\mathcal{R}(y){\wedge}\mathcal{R}(z) \rightarrow (((x{+}y) + z = x + (y + z))$
   $\wedge \; ((x \cdot y) \cdot z = x \cdot (y \cdot z)))$
   R4. *Identity*: $\mathcal{R}(0) \wedge \mathcal{R}(1) \wedge 0{\neq}1 \wedge (\; \forall x \; \mathcal{R}(x) \rightarrow ((x{+}0 = x) \wedge (x \cdot 1 = x))\;)$
   R5. *Inverses*: $\forall x \; \mathcal{R}(x) \rightarrow (\; \exists y \; (x + y = 0) \wedge (x{\neq}0 \rightarrow \exists y \; (xz = 1)\;)\;)$
   R6. *Distributive Law*: $\forall x \forall y \forall z \; \mathcal{R}(x){\wedge}\mathcal{R}(y){\wedge}\mathcal{R}(z) \rightarrow (x \cdot (y + z) = (x \cdot y) + (x \cdot z))$

R7. *Total order*: $\forall x \forall y \forall z \; \mathcal{R}(x) \wedge \mathcal{R}(y) \wedge \mathcal{R}(z) \rightarrow ((x \le y \vee y \le x) \wedge (x \le y \wedge y \le x \rightarrow x=y) \wedge (x \le y \wedge y \le z \rightarrow x \le z))$

R8. *Agreement of ordering with field operations*: : $\forall x \forall y \forall z \; \mathcal{R}(x) \wedge \mathcal{R}(y) \wedge \mathcal{R}(z) \rightarrow ((x \le y \rightarrow x+z \le y+z) \wedge ((0 \le x \wedge 0 \le y) \rightarrow 0 \le x \cdot y))$

R9. *First-order closure*: The following axiom schema holds for all one-place formulas $\varphi(x)$: $\forall x \; (\varphi(x) \rightarrow \mathcal{R}(x)) \wedge \exists x \; \varphi(x) \wedge \exists y \; (\mathcal{R}(y) \wedge \forall x \; (\varphi(x) \rightarrow x \le y))$
$\rightarrow \exists y \; (\mathcal{R}(y) \wedge \forall x \; (\varphi(x) \rightarrow x \le y) \wedge \forall z \; (\mathcal{R}(z) \wedge \forall x \; (\varphi(x) \rightarrow x \le z) \leftrightarrow y \le z)$

Axiom schema R9 is the first-order "image" of the least upper bound axiom. It states that if $\varphi(x)$ represents a non-empty subset of the real numbers and $\varphi(x)$ has a real upper bound, then $\varphi(x)$ has a real least upper bound. Tarski [14] showed that the theory of real closed real fields can be characterized as an ordered field in which every element has a square root and every polynomial of odd degree has a root. R9 covers not only relations definable in the language of the real numbers, but also any real relation definable in $\mathscr{L}^*$. Thus, the above axioms are stronger than the standard RCF axioms.

Probability axioms. Good [4] stresses that probability is properly a two-place function P(E|H), taken to mean the probability that would be assigned to the proposition E if the proposition H were known to be true. Good introduces the symbol H^* to denote "the usual assumptions of logic and pure mathematics," which must be taken as given in all probability assessments. He makes no attempt to decide exactly what should be assumed as part of H^*, and says it is conceivable that H^* cannot be expressed in a finite number of words. Because our concern is reasoning by computational agents, we depart from Good and insist that the underlying assumptions be formalized as a first-order axiomatic theory. We require that H^* be expressed as a finite computational structure, with an effective procedure for generating the axioms explicitly, and an effective procedure for checking whether any given sentence is an axiom. In particular, we assume that H^* includes the axioms N1-N5, R1-R9, and P1-P6 (below).

We introduce into $\mathscr{L}^*$ the two-place function symbol $\mathcal{P}$. The value of $\mathcal{P}$ represents a meaningful probability whenever the following conditions are met   (*i*) the first argument of $\mathcal{P}$ is the Gödel number $\#\sigma$ of a sentence σ of $\mathscr{L}^*$; (*ii*) the second argument of $\mathcal{P}$ is the Gödel number $\#\varphi(x)$ of a one-place open formula $\varphi(x)$ defining a relation representable in $T_{A^*} = \mathrm{Cn}(A^*)$; and (*iii*) the relation represented by $\varphi(x)$ contains the Gödel numbers of all axioms in A^*. The formula $\varphi(x)$ is used to represent the set of sentences whose Gödel numbers satisfy φ. $\mathcal{P}(\#\sigma, \#\varphi(x))$ represents the probability of σ, given that all sentences in the set represented by $\varphi(x)$ are true. Condition (*iii*) says that the domain axioms and probability axioms are taken as given. For readability, we write $\mathcal{P}(\sigma \mid \varphi)$ for $\mathcal{P}(\#\sigma, \#\varphi(x))$ and $\mathcal{P}(\sigma \mid \tau, \varphi)$ for $\mathcal{P}(\#\sigma, \#\psi(x))$, where $\#\psi(x)$ represents the union of the relation defined by φ and $\{\#\tau\}$. That is, $\mathcal{P}(\sigma \mid \tau, \varphi)$ represents the likelihood of σ under the assumption that τ and all sentences in the set represented by $\varphi(x)$ are true.

With this preamble, we now present the probability axioms. The axioms are stated informally for readability; stating them formally is straightforward. The axioms are universally quantified over (Gödel numbers of) sentences σ and τ, and formulas φ.

P1. $0 \le \mathcal{P}(\sigma \mid \varphi) \le 1$.

P2. If $A^* \vdash \sigma$, then $\mathcal{P}(\sigma \mid A^*) = 1$.

P3. If $\mathcal{P}(\sigma \wedge \tau \mid \varphi) = 0$, then $\mathcal{P}(\sigma \vee \tau \mid \varphi) = \mathcal{P}(\sigma \mid \varphi) + \mathcal{P}(\tau \mid \varphi)$.

P4. $\mathcal{P}(\sigma \wedge \tau \mid \varphi) = \mathcal{P}(\sigma \mid \tau, \varphi) \times \mathcal{P}(\tau \mid \varphi)$

P5. If $\sigma \leftrightarrow \tau$, then $\mathcal{P}(\sigma \mid \varphi) = \mathcal{P}(\tau \mid \varphi)$, and $\mathcal{P}(\gamma \mid \sigma, A^*) = \mathcal{P}(\gamma \mid \tau, A^*)$ for all γ.

The first three axioms are the usual axioms for finitely additive probability. P4 formalizes Bayesian conditioning. P5 is taken from Good [4], and formalizes the notion that logically equivalent propositions should be interchangeable with regard to rational degrees of belief.

Some authors (including Good and de Finetti) regard finite additivity as sufficient to formalize rational degrees of belief. Other authors consider countable additivity to be essential. Because countable additivity is typically taken for granted in applications, we regard it as essential. However, full formalization of countable additivity is not possible within FOL, because FOL cannot express the notion of an arbitrary infinite sequence of Gödel numbers. To formalize countable additivity, we adopt a condition introduced by Gaifman [15]. Gaifman's condition can be formalized as a first-order axiom schema. Informally, it is stated as:

P6. $\mathcal{P}(\forall x\ \psi(x) \mid \varphi)$ is the supremum of the values $\mathsf{P}(\psi(\kappa_1) \vee \cdots \vee \psi(\kappa_n) \mid \varphi)$, for all finite conjunctions $\psi(\kappa_1) \vee \cdots \vee \psi(\kappa_n)$ of sentences, formed by substituting constant terms of $\mathcal{L}^*$ into $\psi(x)$.

The constants κ_i may be constants of the original language $\mathcal{L}$, numerical constants (0 or 1), or label constants (one of the $\mathcal{L}_i$). The label constants provide enough constants to cover individuals that might not otherwise be referenced explicitly.

3.3 Terminology

The following definitions provide some necessary terminology.

Definition 1: Let $\mathcal{L}$ be a first-order language. A *p*-language $\mathcal{L}^*$ for $\mathcal{L}$ is a language that augments $\mathcal{L}$ with symbols for domain elements, real and natural number arithmetic, and probability, as described in Section 3.1 above. A *p*-theory T_{A^*} for an axiomatic theory T_A in $\mathcal{L}$ augments the axioms A of T_A as described in Section 3.2 above, to include: (*i*) axioms relativizing axioms in A to elements of the original domain; (*ii*) axioms N1-N5, R-1R9, and P1-P6; and (*iii*) additional axioms defining a domain-specific probabilistic theory. A *p*-theory T_{A^*} containing only N1-N5, RCF, and P1-P6, with no domain axioms, is called the *base p*-theory for $\mathcal{L}$.

Definition 2: Let $\mathcal{L}$ be a first-order language; let $\mathcal{L}^*$ be a *p*-language for $\mathcal{L}$. An axiomatic theory T_{A^*} of $\mathcal{L}^*$ is *probabilistically complete* if it assigns a unique probability $\mathcal{P}(\sigma \mid A^*)$ to every sentence σ of $\mathcal{L}$. That is, T_{A^*} is probabilistically complete if for every sentence σ there is a unique real number p_σ such that $T_{A^*} \vdash \mathcal{P}(\sigma \mid A^*) = p_\sigma$.

A probabilistically complete *p*-theory assigns a single RCF element to each sentence. Incomplete *p*-theories give rise to interval probabilities. Some writers have advocated founding the theory of probability on interval rather than point-valued probabilities (e.g., [4]). The possibility of incomplete *p*-theories is attractive when the KB designer is not able to specify a probability for every sentence. With the advent of first-order languages based on graphical probability models, it is now possible to

define probabilistically complete *p*-theories suitable for many interesting problems, to develop workable knowledge engineering procedures for specifying *p*-theories, and to devise tractable inference and learning algorithms for *p*-theories.

Definition 3: Let $\mathscr{L}$ be a first-order language; let *T* be a theory of $\mathscr{L}$; and let $\mathscr{L}^*$ be a *p*-language for $\mathscr{L}$. An axiomatic theory T_{A*} of $\mathscr{L}^*$ *corresponds* to *T* if $T \vdash \sigma$ implies $T_{A*} \vdash \mathcal{P}(\sigma \,|\, A^*) = 1$ for any sentence σ of $\mathscr{L}$. T_{A*} *corresponds strongly* to *T* if $T \vdash \sigma$ if and only if $T_{A*} \vdash \mathcal{P}(\sigma \,|\, A^*) = 1$.

Clearly, if the axioms *A* of T_A are included among the axioms A^* of T_{A*}, then T_{A*} corresponds to T_A. In general, augmenting *A* with N1-N5, RCF, and P1-P6, will not determine a unique *p*-theory. If it is assumed that *A* incorporates all objective, incontrovertibly true domain knowledge, then adding probabilistic axioms to complete a *p*-theory brings subjectivity into the KB.

Of course, in actual applications, it is rarely the case that all logical axioms are incontrovertibly true assertions. More realistically, some axioms will be highly questionable; others, though quite useful, may be downright false. Axioms in real KBs are carefully engineered to be "good enough for the task." A great deal of subjective judgment goes into developing a "good enough" KB. In short, the logical axioms of a KB are often as subjective as the probabilities, and sometimes more so.

Definition 4: Let $\mathscr{L}$ be a first-order language, $\mathscr{L}^*$ a *p*-language for $\mathscr{L}$, and T_{A*} a probabilistically complete axiomatic theory of $\mathscr{L}^*$. Let $\varphi(x,y)$ be a formula of $\mathscr{L}^*$ that functionally represents a recursive sequence of Gödel numbers of sentences of $\mathscr{L}^*$ (i.e., for each natural number *n*, there is exactly one sentence σ_n such that $\varphi(n, \#\sigma_n)$ is provable from A^*). We say the sequence $\sigma_1, \sigma_2, \ldots$ of sentences is *negligible* if for every RCF element $u > 0$ there is a natural number *n* such that $\mathcal{P}(\sigma_1 \wedge \cdots \wedge \sigma_n \,|\, A^*) < u$. The sequence $\sigma_1, \sigma_2, \ldots$ is *certain* if for every RCF element $u > 0$ there is a natural number *n* such that $\mathcal{P}(\sigma_1 \wedge \cdots \wedge \sigma_n \,|\, A^*) > 1 - u$.

A negligible sequence is vanishingly improbable. That is, the probabilities of its finite-length leading segments tend to zero as their lengths increase without bound. Clearly, any sequence containing a zero-probability sentence is negligible. We can define negligible or certain individual sentences or finite-length sequences of sentences in the obvious way, by appending infinitely many copies of a tautology to the end of the sequence. An individual sentence is certain if it has probability 1 and negligible if it has probability zero.

Defintion 5: Let $\mathscr{L}$ be a first-order language; let $\mathscr{L}^*$ be a *p*-language for $\mathscr{L}$; let T^* be a theory of $\mathscr{L}^*$. The *core* of T^* is the set of sentences $\{\sigma : \sigma$ is certain under $T^*\}$.

4 Semantics

This section defines semantics for *p*-theories.

Standard first-order semantics. The logic set forth in this paper is a standard, untyped first-order logic. As such, it can be given standard first-order model-theoretic semantics.

A first-order structure for a theory in $\mathscr{L}^*$ is a pair (D, m), where *D* is a non-empty set called the domain of interpretation, and *m* assigns to each function, constant, and

relation symbol of $\mathcal{L}^*$ a function, constant or relation of the correct arity on D. A structure (D, m) implies a truth-value for each sentence of $\mathcal{L}^*$. (D, m) is called a *model* of T^* if every sentence in T^* is true in (D, m). Models of T^* are sometimes called *possible worlds* for T^*. A sentence σ is implied by A^* if it is true in all models of A^*, and satisfiable if it has a model.

Standard semantics for *p*-theories gives rise to a seeming conflict between logical truth and probabilistic certainty: there may be logically possible sentences that have probability zero. For example, we might represent successive tosses of a symmetric die as independent and identically distributed with probability 1/6 of landing on each face. In a hypothetical infinite sequence of tosses, the frequency of tosses that land on, say, the number 2 is certain to be 1/6. Any sequence of outcomes that does not have a limiting frequency of 1/6 is negligible, in the sense of D4. Nevertheless, *every* sequence of outcomes is logically possible. Standard first-order semantics cannot distinguish between typical realizations of this probabilistic process (i.e., "random looking" sequences with limiting frequency 1/6) and highly atypical realizations (e.g., sequences that have the incorrect limiting frequency, or exhibit some other unusual regularity, such as a 2 on every sixth toss). Suppose the sentence σ_1 asserts that the limiting frequency is 1/6 that the die lands on 2; the sentence σ_2 asserts that every toss comes up 2; and the sentence σ_3 asserts that a 2 occurs on the first two tosses. The sentence σ_1 has probability 1; σ_2 has probability zero; and σ_3 has probability 1/36. None of these sentences is either implied by or inconsistent with the logical axioms. Traditional first-order semantics seems to provide no way to differentiate among them. For this reason, many authors have considered standard first-order semantics inadequate for probabilistic theories, and have turned to alternative semantics.

Measure models. A common approach to giving semantics to probabilistic logics is through a probability measure on structures. Using results from measure theory, Gaifman [15] showed that a coherent probability assignment to quantifier-free sentences can be extended to a countably additive probability measure on a σ-algebra of subsets of $\{(D, m)\}$, the set of all structures on $D = \{\mathcal{L}_1, \mathcal{L}_2, \ldots\}$.[2] In measure model semantics, a sentence is assigned probability equal to the measure of the set of models of the sentence.

Just as traditional first-order semantics is defined in terms of set theory, measure model semantics is defined in terms of measure theory. Measure theory is the branch of real analysis used to formalize probability. In measure model semantics, it is generally taken for granted that probabilities are interpreted as real numbers. As noted above, the semantic condition that probabilities must be interpreted as real numbers results in a non-axiomatizable logic that lacks a complete proof theory. On the other hand, axiomatic set theory provides sufficient mathematical machinery to prove the standard results of measure theory. If A^* contains set theory axioms, then the results of measure theory hold in all models of A^*. Therefore, a probabilistically complete *p*-theory that includes set theory axioms has a unique measure model. A probabilistically incomplete *p*-theory has a family of measure models, one for each

---

[2] Gaifman's domain of interpretation included the constants of the original language as well as the added constants; the original constants are then self-interpreted. This requires that no two constants be equal, an assumption we do not make.

distribution consistent with T_{A*}. Any measure model assigns probability zero to the set of models of negligible sequences.

Under measure model semantics, the logical and probabilistic aspects of a theory remain semantically distinct. Provable sentences are true in all ordinary models of T_{A*}, and hence have probability 1 in any measure model of T_{A*}. Unsatisfiable sentences are false in all models and hence have probability zero in any measure model. Other than P1-P6, probabilities for other sentences are unrestricted. In particular, a sentence may provably have probability 1 and yet be false in some models.

Certainty restriction semantics. If a sentence has probability 1, then conditioning on the sentence does not change its probability or the probability of any other sentence. Furthermore, because there are only countably many sentences, we can condition on *all* certain sentences – the core of the *p*-theory – without changing any probabilities. We can use this fact to rule out negligible sentences as models of a *p*-theory. Note that a *p*-theory T_{A*} has enough logical machinery to define a provability predicate. We can thus introduce an axiom schema that infers σ from $T_{A*} \vdash \mathcal{P}(\sigma \mid A^*)$ = 1. Adding this axiom schema excludes provably negligible sentences as models of T_{A*}. We call this axiom schema the *certainty restriction*. Adding the certainty restriction schema to T_{A*} reduces the set of models of T_{A*} without changing either the probability of any sentence or any of the measure models consistent with T_{A*}.

A rational agent makes no practical distinction between propositions with probability one and those provable from its knowledge base. Many texts use limiting frequencies (as well as other certain events) to define the meaning of probability statements (e.g., that "fair die" means that the limiting frequency of tosses landing on each face is 1/6). This suggests that the certainty restriction captures some aspect of the intuitive semantics of probability as it is commonly applied and understood.

Strong probabilistic semantics. The certainty restriction and the conditioning restriction can be formulated in first-order logic. This means that these conditions can be imposed as satisfaction criteria for *p*-theories without any change to first-order semantics. However, these conditions cannot capture the stronger semantic notion that infinite-length negligible sequences should not occur in models of a probabilistic theory. Strong probabilistic semantics requires that no model of a *p*-theory may contain all sentences in a negligible sequence of sentences.

Each negligible sequence can be identified with an *effectively null* binary string, as defined by Martin-Löf [6]. Martin-Löf randomness has been studied extensively (c.f., [7]), and is popular as a characterization of what it means for a sequence to be a typical realization of a probability distribution on sequences. Dawid's [8] calibration criterion is closely related to Martin-Löf randomness: the set of uncalibrated sequences for a given probability distribution is effectively null for that distribution.

If the axioms A^* of our *p*-theory are strong enough to formalize measure theory, then we can prove that the set of negligible sequences has probability zero under the measure model for A^*. Thus, negligible sequences can be excluded as models of A^* without changing any probabilities. However, excluding the negligible sentences as models means abandoning traditional first-order semantics, because the proposition that a sequence is negligible cannot be formalized as a recursive first-order axiom schema. As a consequence, there is no complete proof system for probabilistic logic with strong probabilistic semantics.

Frequency probability. Some authors have argued that fundamentally different kinds of probability are required to formalize different metaphysical notions such as subjective degrees of belief, long-run frequencies, physical randomness, or algorithmic randomness. Others argue that a single kind of probability is adequate for all these metaphysical positions. Good ([4], [16]) and Barnett [17] discuss the different viewpoints on this issue.

Because the same mathematics is applied to reason about all these kinds of probability, and a proliferation of different logics complicates knowledge representation and knowledge interchange, it seems reasonable to investigate whether a single computational logic might be applicable to different notions of probability.

We have argued above that *p*-theories can represent subjective degrees of belief about propositions that can neither be proven nor disproven. We argue that *p*-theories can represent long-run frequencies and physical randomness. The basic idea derives from a theorem of de Finetti [18] stating that an infinitely exchangeable sequence of events is mathematically equivalent to one a frequentist or proponent of physical propensity would model as independent and identically distributed (iid) given an unknown parameter, together with a subjective probability distribution on values of the parameter. To the frequentist, the parameter corresponds to the unknown long-run frequency. To the propensity theorist, the parameter corresponds to the unknown propensity. To the strict subjectivist, the parameter is a modeling fiction that provides a parsimonious representation for an exchangeable sequence.

Infinitely exchangeable sequences in *p*-theories are represented as first-order axiom schemas stating that different orderings of finite-length initial segments of a sequence are equally probable. *P*-theories can also represent iid propositions with unknown parameters. Incomplete *p*-theories can be used if no subjective distribution is available (perhaps due to philosophical aversion to subjective probability) for the unknown parameter. Thus, frequency and propensity probability as well as degree of belief probability can be represented with *p*-theories.

Summary: Semantics for p-theories. With no change to standard first-order semantics, a complete *p*-theory assigns probabilities consistently to sentences of $\mathscr{L}^*$ such that the axioms A^* have probability 1. By including the certainty restriction axiom schema in A^*, we can identify probability 1 with provability from A^*. If A^* includes set theory axioms, there is a unique measure model (probability measure over models) for each complete *p*-theory. Requiring that all models of a *p*-theory be non-negligible in the sense of Martin-Löf would take us out of the realm of first-order model theory.

5 Learning and Dogmatism

An attractive feature of probability theory is its inbuilt support for learning from observation. We can add any new non-negligible axiom to a *p*-theory, and Bayesian conditioning can be used to obtain an updated *p*-theory with the evidence as an axiom. As new observations accrue, we obtain a sequence of *p*-theories, each containing additional axioms representing new information obtained since the previous *p*-theory in the sequence.

Fundamental to the scientific attitude is lack of dogmatism. In our context, non-dogmatism means assigning probability zero only to propositions known incontrovertibly to be false. A dogmatic theory will be overturned if it makes definite empirical predictions that turn out to be false. However, a theory can be dogmatic without ever being proven false. For example, if a theory starts out certain that a coin is fair, it can never learn that the coin is biased, no matter how many trials are observed. A theory that allows for bias will eventually become convinced that a biased coin is biased, even if it begins with a high likelihood that the coin is fair.

If we begin with an axiomatic theory T_A of $\mathcal{L}$, and assume that the axioms A represent a set of sentences known incontrovertibly to be true (whether by definition or by meticulous empirical observation), we would like to be able to represent a p-theory that assigns probability zero to exactly those sentences that can be disproven from A. We say such a p-theory corresponds *non-dogmatically* to T_A. There are many reasons, including tractability, convenience of specification, economy of communication, and the like, that we might choose to represent and reason with a dogmatic theory, as long as it is judged "good enough" for the task at hand. But as a matter of principle, we want the capability to represent a non-dogmatic theory, even if reasoning with it is impractical.

A result of Gaifman and Snir [11] would seem to doom any hope of finding a non-dogmatic theory. They proved that, under measure-model semantics, every axiomatizable theory is dogmatic. On the other hand, Laskey [19] described how to specify a non-dogmatic p-theory corresponding to any consistent, finitely axiomatizable first-order theory. This apparent inconsistency is resolved by noting that Gaifman and Snir assume that all true sentences of natural number arithmetic are base axioms of the logical language, which is therefore not axiomatizable. Gaifman and Snir's no-go result does not apply to axiomatic first-order probability logics. An inevitable price of the axiomatic first-order approach, implied by Tarski's undefinability theorem, is that any complete p-theory must assign probability intermediate between 0 and 1 to some sentences in the language of arithmetic. This is natural if we view probability as a degree of provability. No first-order axiom system can prove all true sentences of arithmetic; therefore, no axiomatic first-order probability logic can assign probability 1 to all true sentences of arithmetic.

6 Conclusion

As probabilistic languages find increasing application to the Semantic Web (e.g., [20]), there is a need for a logical foundation that integrates traditional SW logics with probability logic. The logic presented here formalizes probability as a standard axiomatic first-order theory, with no alteration to traditional first-order model theoretic semantics. Advantages of this approach are the ability to represent p-theories as finite computational structures amenable to machine processing, the availability of a complete proof system, and compatibility with semantics of traditional logic-based languages. While this paper focuses on expressive power of the logic, SW applications require tractability and scalability. Future work will consider tractable restrictions of the logic, as well as fast approximate inference methods.

Acknowledgments

References

1. Jeffreys, H. Theory of Probability. (3rd ed) Oxford University Press (1961)
2. Savage, L.J. The Foundations of Statistics. Wiley, New York (1954)
3. Costa, P. C. G., Laskey, K.B. and Lukasiewicz, T. Uncertainty Representation and Reasoning in the Semantic Web, in Semantic Web Engineering in the Knowledge Society, Idea Information Science (2008)
4. Good, I. J. Probability and the Weighing of Evidence. London: Charles Griffin and Co (1950)
5. Ng, K. S. and Lloyd, J. W. Probabilistic Reasoning in a Classical Logic. Journal of Applied Logic, vol. 7, no. 2, pp. 218-238 (2009).
6. Martin-Lof, P. The Definition of Random Sequences. Information and Control vol. 9, pp. 602-619 (1966)
7. Li, M. and Vitányi, P. An Introduction to Kolmogorov Complexity and Its Applications (2nd ed). Springer (1997)
8. Dawid, A. P. Calibration-Based Empirical Probability. The Annals of Statistics vol. 13, no. 4, pp. 1251-1274 (1985)
9. Enderton, H.B. A Mathematical Introduction to Logic: Harcourt Academic Press (2001)
10. Halpern, J.Y. An Analysis of First-Order Logics of Probability. Artificial Intelligence, vol. 46, pp. 311-50 (1991)
11. Gaifman, H. and M. Snir. Probabilities over Rich Languages, Journal of Symbolic Logic, vol. 47, pp. 495-548 (1982)
12. Bacchus, F. Representing and Reasoning with Probabilistic Knowledge: A Logical Approach to Probabilities. Boston, MA, MIT Press (1990)
13. Shoenfeld, J. R. Mathematical Logic. Association for Symbolic Logic (1967)
14. Tarski, A. A Decision Method for Elementary Algebra and Geometry. Univ. of California Press (1951)
15. Gaifman, H. Concerning Measures in First-Order Calculi. Israel Journal of Mathematics. vol. 2, pp. 1-18 (1964)
16. Good, I.J. Good Thinking: The Foundations of Probability and Its Applications. University of Minnesota Press (1983)
17. Barnett, V. Comparative Statistical Inference. (3rd ed.) Wiley, New York (1999)
18. de Finetti, Bruno. Theory of Probability: A Critical Introductory Treatment. Wiley, New York (1974, originally published in 1934).
19. Laskey, K.B. MEBN: A Language for First-Order Bayesian Knowledge Bases. Artificial Intelligence, vol. 172, no. 2-3, pp. 140-178 (2008)
20. Carvalho, R., Laskey, K.B., Costa, P., Ladeira, M., Santos, L., and Matsumoto, S. UnBBayes: Modeling Uncertainty for Plausible Reasoning in the Semantic Web. In: V. Kordic (ed.) Semantic Web, IN-TECH Publishing, ISBN: 978-953-7619-33-6 (in press)