Axiomatic First-Order Probability for the Semantic Web

Kathryn Blackmond Laskey
Department of Systems Engineering and Operations Research
Center of Excellence in Command, Control, Communications, Computing and Intelligence
George Mason University

Fifth International Workshop on Uncertainty Reasoning for the Semantic Web
October 2009

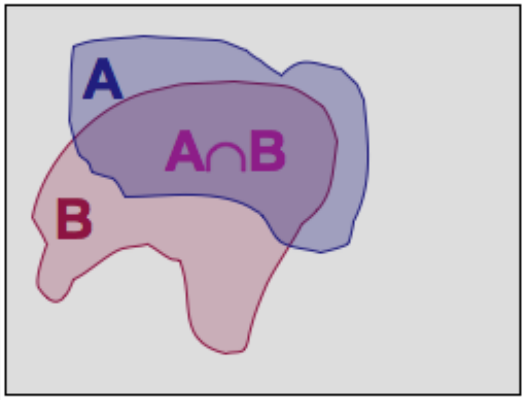
GEORGE MASON UNIVERSITY

Probability

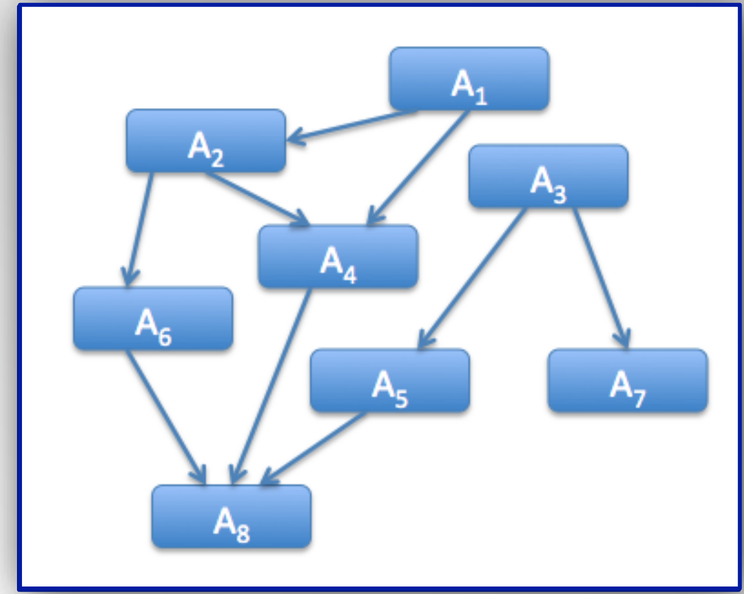
- Uncertainty is ubiquitous
  - *"Prediction is difficult, especially the future."* – Yogi Berra
  - In an open world, attempts to nail down an unambiguous meaning and definite truth-value for every statement are doomed to failure.
- Probability formalizes reasoning under uncertainty
  - *"Symbolic logic is a model [of deductive thought] in much the same way that modern probability theory is a model for situations involving chance and uncertainty."* – Enderton (2001)
  - Probability allows us to draw useful conclusions when our knowledge falls short of certainty
- There is vigorous debate over:
  - Semantics of probability
    - *Classical? Frequency? Propensity? Subjective? Logical?*
  - How to combine probability with classical logic

Mathematical Probability

- Probabilities are assigned to *events*
  - Event represents uncertain outcome
  - Mathematically, events are subsets of a *sample space* Ω
  - (For uncountable Ω, we must restrict events to *measurable* subsets of Ω)
- A *probability measure* $P(\cdot)$ satisfies the following axioms:
  - $P(A) \geq 0$ for all measurable events A
  - $P(\Omega) = 1$
  - If $A_1, A_2, \ldots$ is a sequence of measurable events such that $A_i \cap A_j = \emptyset$ then $\mathrm{P}(A_1 \cup A_2 \cup \cdots) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \cdots$
- The conditional probability of A given B for any two events A and B is defined as a number $\mathrm{P}(A|B)$ satisfying:
  - $\mathrm{P}(A|B)\mathrm{P}(B) = \mathrm{P}(A \cap B)$



Propositional Logic and Probability



- There is a natural way to define probabilities in a propositional language with finitely many sentence symbols
- Each sentence symbol specifies an event
  - Event A corresponding to sentence symbol Q occurs if and only if Q is true
  - $Q_1 \vee \cdots \vee Q_n$ corresponds to $A_1 \cup \cdots \cup A_n$
  - $Q_1 \wedge \cdots \wedge Q_n$ corresponds to $A_1 \cap \cdots \cap A_n$
  - Similarly for the other logical connectives
- We can define a probability measure over truth values of the Q_i
  - Probability measure on truth values of sentence symbols gives rise to probability for each wff
  - Probability measure can be defined consistently and parsimoniously using conditional independence
- This idea can be extended to languages with infinitely many sentence symbols

First-Order ✓

- Second-order Logic
  - \+ Specify probability density functions directly
  - \+ Represent probabilities with real numbers
  - − No completeness or compactness theorem (valid sentence may not be provable)
  - − Even logicians do not agree on semantics (or on whether it is meaningful to quantify over *all* functions and relations)
- First-order Logic
  - − Cannot refer *directly* to properties, functions or sentences
  - \+ Represent arbitrarily fine-grained degrees of plausibility
  - \+ Completeness and compactness theorem (valid sentences are provable)
  - \+ Well-understood and universally accepted semantics
  - \+ Can refer *indirectly* to sentences (via Gödel numbers)
  - \+ Most SW languages are based on a subset of first-order logic
- Second-order logic with *general semantics*
  - Second-order syntax with first-order model theory
  - Talk about sentences *and* retain other benefits of FOL

Axiomatic First-Order Theory

- Represent knowledge explicitly as finite computational structure
- Contradictions and entailments can be detected in finite time
- Consequences are effectively enumerable
- ***Can we formalize probability as a first-order axiomatic theory?***

No-Go Results (*i*)

- Gaifman (1964) assigned probabilities to sentences of a first-order language
  - Extend original language $\mathscr{L}$ to a new language $\mathscr{L}^*$ with additional individual constants to cover all objects in the domain
  - Assign probability measure to quantifier-free sentences of $\mathscr{L}^*$
  - Extend to probability measure on all sentences via *Gaifman's condition:*

    $\mathsf{P}(\forall x\ \psi(x))$ *is supremum of* $\mathsf{P}(\psi(\kappa_1) \vee \cdots \vee \psi(\kappa_n))$, *for all finite conjunctions* $\psi(\kappa_1) \vee \cdots \vee \psi(\kappa_n)$ *of sentences, formed by substituting constant terms of the extended language* $\mathscr{L}^*$ *into* $\psi(x)$
  - *Measure-model* semantics defines a probability measure on possible worlds
- Gaifman and Snir (1982) studied definability of probabilities and tests for satisfaction
  - Refer to sentences indirectly via Gödel numbers
  - Semantics restricts mathematical sublanguage to intended interpretation on natural numbers; therefore:
    - Probabilities are not definable on mathematical sublanguage
    - All definable probability functions on empirical sublanguage are "dogmatic" (assign probability zero to some satisfiable sentence)

No-Go Results (*ii*)

- Bacchus criticized Gaifman's approach because it "fail[s] to address some of the main concerns of AI"
  - Cannot represent assertions *about* probabilities, e.g.:
    - "The false positive probability is less than 0.05"
    - "Rain is more likely today than it was yesterday."
- Abadi and Halpern (1994) examined first-order logics that can reason both with and about probability
  - "…first-order …language for reasoning about probabilities ought to have easily comprehensible syntax and semantics.
  - "Ideally, the validity problem would not be worse than for first-order logic, and we would have a complete axiomatization…"
  - *But* "…as long as [the language] is sufficiently rich, the validity problem for first-order reasoning about probability is wildly undecidable."
  - No complete axiomatization is possible
- Undecidability results apply even if probabilities are restricted to rational numbers

Addressing the Roadblocks

- Probabilities are usually formalized as real numbers, and real numbers cannot be axiomatized in a first-order theory
  - Real numbers = ordered field + least upper bound axiom
  - Least upper bound axiom refers to *all bounded subsets* of the real numbers
  - We can formulate a first-order least upper bound axiom that applies to all *definable* bounded subsets
  - This is the theory of *real closed fields*
- FOL cannot refer to sentences
  - We can refer indirectly to sentences via their Gödel numbers
- We cannot define a "truth function" on the natural numbers
  - Any definable first-order probability function must be uncertain about some statements about the natural numbers

Desirable Features of Probability Logic

- Express statements about domain and about probabilities
- Express arbitrarily fine-grained degrees of likelihood
- Define a probability for every sentence in the language
- Define non-dogmatic distributions
- Condition explicitly on all background knowledge (mathematical, logical, domain)
- Discover any contradiction in finite time
- Support learning from observation
- Deal appropriately with infinite limits

All these can be achieved by formalizing probability as an axiomatic first-order theory

The Axioms: Real Closed Field

R1: *Additive and multiplicative closure*: For all x and y, $\mathcal{R}(x)$ and $\mathcal{R}(y)$ imply $\mathcal{R}(x+y)$ and $\mathcal{R}(x \cdot y)$.

R2: *Commutativity of addition and multiplication*: For all x and y, $\mathcal{R}(x)$ and $\mathcal{R}(y)$ imply $x+y = y+x$ and $x \cdot y = y \cdot x$.

R3: *Associativity of addition and multiplication*: For all x, y, and z, $\mathcal{R}(x)$, $\mathcal{R}(y)$ and $\mathcal{R}(z)$ imply $(x+y) + z = x + (y + z)$ and $(x \cdot y) \cdot z = x \cdot (y \cdot z)$.

R4: *Additive and multiplicative identity*: $\mathcal{R}(0)$ and $\mathcal{R}(1)$ and $0 \neq 1$ and for all x, $\mathcal{R}(x)$ implies $x+0 = x$ and $x \cdot 1 = x$.

R5: *Additive and multiplicative inverses*: For all x, $\mathcal{R}(x)$ implies there exists y such that $x + y = 0$. For all x, $\mathcal{R}(x)$ and $x \neq 0$ implies there exists z such that $xz = 1$.

R6: *Distributivity of multiplication over addition*: For all x, y, and z, $\mathcal{R}(x)$, $\mathcal{R}(y)$ and $\mathcal{R}(z)$ imply $x \cdot (y + z) = (x \cdot y) + (x \cdot z)$.

R7: *Total ordering*: For all x, y, and z, $\mathcal{R}(x)$, $\mathcal{R}(y)$ and $\mathcal{R}(z)$ imply ($x \leq y$ or $y \leq x$) and (if $x \leq y$ and $y \leq x$ then $x=y$ and (if $x \leq y$ and $y \leq z$ then $x \leq z$).

R8: *Agreement of ordering with field operations*: For all x, y, and z, $\mathcal{R}(x)$, $\mathcal{R}(y)$ and $\mathcal{R}(z)$ imply if ($x \leq y$ then $x+z \leq y+z$) and (if $0 \leq x$ and $0 \leq y$ then $0 \leq x \cdot y$).

R9: *First-order closure*: The following axiom schema holds for all one-place formulas $\varphi(x)$:

$$\forall x\, (\varphi(x) \rightarrow \mathcal{R}(x)) \wedge \exists x\, \varphi(x) \wedge \exists y\, (\mathcal{R}(y) \wedge \forall x\, (\varphi(x) \rightarrow x \leq y)\,) \rightarrow$$
$$\exists y\, (\mathcal{R}(y) \wedge \forall x\, (\varphi(x) \rightarrow x \leq y) \wedge \forall z\, (\mathcal{R}(z) \wedge \forall x\, (\varphi(x) \rightarrow x \leq z) \;\leftrightarrow y \leq z)).$$

The Axioms: Natural Numbers

Integer arithmetic. The following axioms, together with the real closed field axioms, provide enough power for Gödel numbering and reasoning about provability:

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. Induction axiom schema: 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.

The Axioms: Probability

- Probability axioms are stated informally but can be formalized as first-order axioms.
- Refer to sentence σ indirectly through its Gödel number $\#\sigma$.
- $\mathcal{P}(\#\sigma, \#\varphi(x))$ represents probability of σ given sentences represented by formula $\varphi(x)$.
- Suppress Gödel numbers for readability, e.g., $\mathcal{P}(\sigma \mid \varphi)$.
- A^* represents mathematical and domain axioms.
- Probability axioms are universally quantified over (Gödel numbers of) sentences σ and τ, and formulas φ.

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

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 $\mathcal{P}(\gamma \mid \sigma, \varphi) = \mathcal{P}(\gamma \mid \tau, \varphi)$ for all sentences γ.

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

Semantics

- Standard first-order model theoretic semantics applies.
- A model (or *possible world*) consists of:
  - A domain D
  - A function on D^n for each n-ary function symbol
  - A subset of D^n for each n-ary predicate symbol
  - An element of D for each constant symbol

  such that every axiom of A^* is true in the model.
- *Certainty restriction*: Without affecting any probabilities, we can add an axiom schema concluding the negation of a sentence that provably has probability zero.
- *Measure models*. If a probability is defined for every sentence then there is a unique measure model; otherwise there is a set of measure models.

…for the Semantic Web

- First-order languages provide well-known advantages, e.g.
  - Explicit finite computational representation
  - Complete proof theory
  - Compatibility with SW languages
- Can translate to second-order language with general semantics
- Provides unified semantics for a variety of probability languages making different expressivity / tractability tradeoffs

Thank You!