Logic and Probability
Overview / Probability Logic

Thomas Icard & Krzysztof Mierzewski

August 8, 2022

Overview

Logic as a theory of:

- truth-preserving inference
- consistency
- definability
- proof / deduction
- rationality

...

Probability as a theory of:

- ampliative inference
- learning
- information
- induction
- rationality

...

Some questions and points of contact:

- In what ways might probability be said to extend logic?
- How do probability and various logical systems differ on what they say about rational inference?
- What are sensible ways of discretizing continuous probabilistic models? What do we lose in the process?
- How might probability be a useful tool in elucidating logical phenomena of interest?
- What happens to probability when we impose logical—e.g., computability-theoretic—constraints?

Course Outline

- Day 1: "Probability as Logic" and Landscape of Probability Logics (TI)
- Day 2: Default Reasoning, Acceptance Rules, and the Quantitative/Qualitative Interface (TI/KM)
- Day 3: First Order Probability Logic and 0/1 Law (KM)
- Day 4: Probabilistic Grammars and Programs (TI)
- Day 5: Computable Measure Theory & Applications (KM)

A measurable space is a pair $(W, \mathcal{E})$ with

- W is an arbitrary set
- $\mathcal{E}$ is a σ-algebra over W, i.e., a subset of $\wp(W)$ closed under complement and infinite union.

A probability space is a triple $(W, \mathcal{E}, \mu)$, with $(W, \mathcal{E})$ a measurable space and $\mu : \mathcal{E} \to [0, 1]$ a measure function, satisfying the following two axioms:

1. $\mu(W) = 1$;
2. $\mu(E \cup F) = \mu(E) + \mu(F)$, whenever $E \cap F = \emptyset$.

Suppose we have a propositional logical language $\mathcal{L}$

$$\varphi \quad ::= \quad A \mid B \mid \ldots \mid \varphi \wedge \varphi \mid \neg\varphi$$

We can define a probability $\mathbb{P} : \mathcal{L} \rightarrow [0, 1]$ by requiring

1. $\mathbb{P}(\varphi) = 1$, for any tautology φ ;
2. $\mathbb{P}(\varphi \vee \psi) = \mathbb{P}(\varphi) + \mathbb{P}(\psi)$, whenever $\vDash \neg(\varphi \wedge \psi)$.

Equivalent set of requirements:

1. $\mathbb{P}(\varphi) = 1$ for any tautology ;
2. $\mathbb{P}(\varphi) \leq \mathbb{P}(\psi)$ whenever $\vDash \varphi \rightarrow \psi$;
3. $\mathbb{P}(\varphi) = \mathbb{P}(\varphi \wedge \psi) + \mathbb{P}(\varphi \wedge \neg\psi)$.

It is then easy to show:

- $\mathbb{P}(\varphi) = 0$, for any contradiction φ ;
- $\mathbb{P}(\neg\varphi) = 1 - \mathbb{P}(\varphi)$;
- $\mathbb{P}(\varphi \vee \psi) = \mathbb{P}(\varphi) + \mathbb{P}(\psi) - \mathbb{P}(\varphi \wedge \psi)$;
- A propositional valuation sending atoms to 1 or 0 is a special case of a probability function ;
- A probability on $\mathcal{L}$ gives rise to a standard probability measure over 'world-states', i.e., maximally consistent sets of formulas from $\mathcal{L}$. In fact, any standard probability measure can be obtained this way.

Why these axioms?

Interpretations of Probability

- Frequentist: Probabilities are about 'limiting frequencies' of in-principle repeatable events.
- Propensity: Probabilities are about physical dispositions, or propensities, of events.
- Logical: Probabilities are determined objectively using a logical language and some additional background principles, e.g., of 'symmetry'.
- Bayesian: Probabilities are subjective and reflect an agent's degree of confidence concerning some event.

$\vdots$

> We strive to make judgments as dispassionate, reflective, and wise as possible by a doctrine that shows where and how they intervene and lays bare possible inconsistencies between judgments. There is an instructive analogy between [deductive] logic, which convinces one that acceptance of some opinions as ‘certain’ entails the certainty of others, and the theory of subjective probabilities, which similarly connects uncertain opinions.
>
> —Bruno de Finetti, 1974

De Finetti's Argument

1. Interpret probability assignment as betting odds judged fair. For example, an assignment $\mathbb{P}(A) = 0.2$ means any bet that costs at most $0.2 \times S$, but pays at least S if A turns out to be true, would be judged fair.

2. Assume that fair gambles do not become collectively unfair upon collection into a joint gamble.

3. Show that $\mathbb{P} : \mathcal{L} \to [0, 1]$ is consistent with the axioms if and only if no system of bets with odds licensed by $\mathbb{P}$ results in a sure loss.

In other words, the axioms can be interpreted as consistency constraints on betting odds. (See also Howson 2007.)

Related Arguments

- Cox's Theorem: Axioms fall out of basic (logical) consistency postulates on real-number-valued "plausibility assignments" (Cox, Jaynes, etc.).
- Accuracy Dominance: Any violation of the axioms results in probability assignments that could be strictly more accurate (Joyce, Leitgeb & Pettigrew, etc.).

A different way of construing probability as logic—also pioneered by de Finetti—is to interpret the probability function as representing purely qualitative, comparative judgments:

"E is more likely than F"
"E is at least as likely as F"
"E and F are equally likely"

What is the logic of such comparative judgments?

What kind of logic would we expect if such judgments were derived from some probability measure?

Definition

Call $(\mathcal{E}, \succeq)$ a de Finetti order (de Finetti 1937) if it satisfies:

- Positivity:
$$E \succeq \varnothing$$
- Non-triviality:
$$\varnothing \not\succeq W$$
- Totality:
$$E \succeq F \text{ or } F \succeq E$$
- Quasi-additivity: Whenever $(E \cup F) \cap G = \varnothing$,
$$E \succeq F \quad \Leftrightarrow \quad E \cup G \succeq F \cup G \,.$$

Agreement

Does every de Finetti order $(\mathcal{E}, \succeq)$ admit of an agreeing probability measure? That is, a measure μ such that

$$E \succeq F \quad \Leftrightarrow \quad \mu(E) \geq \mu(F) \text{ ?}$$

Notation

Given an order $(\mathcal{E}, \succeq)$ let us write $E \succ F$ just in case $E \succeq F$ but not $F \succeq E$. Agreement requires $E \succ F \Rightarrow \mu(E) > \mu(F)$.

Example (Kraft, Pratt, & Seidenberg, 1959)

Let $W = \{a, b, c, d, e\}$:

$$\{d\} \succ \{a, c\} \qquad \{b, c\} \succ \{a, d\} \qquad \{a, e\} \succ \{c, d\}$$

$$\{a, c, d\} \succ \{b, e\}$$

Fact

$(\wp(W), \succeq)$ admits no agreeing probability measure.

$$\begin{aligned} \mu(\{d\}) &> \mu(\{a, c\}) \\ \mu(\{b, c\}) &> \mu(\{a, d\}) \\ \mu(\{a, e\}) &> \mu(\{c, d\}) \end{aligned}$$

Example (Kraft, Pratt, & Seidenberg, 1959)

Let $W = \{a, b, c, d, e\}$:

$$\{d\} \succ \{a, c\} \qquad \{b, c\} \succ \{a, d\} \qquad \{a, e\} \succ \{c, d\}$$

$$\{a, c, d\} \succ \{b, e\}$$

Fact

$(\wp(W), \succeq)$ admits no agreeing probability measure.

$$\begin{aligned} \mu(\{d\}) &> \mu(\{a\}) + \mu(\{c\}) \\ \mu(\{b\}) + \mu(\{c\}) &> \mu(\{a\}) + \mu(\{d\}) \\ \mu(\{a\}) + \mu(\{e\}) &> \mu(\{c\}) + \mu(\{d\}) \end{aligned}$$

Example (Kraft, Pratt, & Seidenberg, 1959)

Let $W = \{a, b, c, d, e\}$:

$$\{d\} \succ \{a, c\} \qquad \{b, c\} \succ \{a, d\} \qquad \{a, e\} \succ \{c, d\}$$

$$\{a, c, d\} \succ \{b, e\}$$

Fact

$(\wp(W), \succeq)$ admits no agreeing probability measure.

$$\begin{aligned} & \mu(\{d\}) + \mu(\{b\}) + \mu(\{c\}) + \mu(\{a\}) + \mu(\{e\}) \\ > \; & \mu(\{a\}) + \mu(\{c\}) + \mu(\{a\}) + \mu(\{d\}) + \mu(\{c\}) + \mu(\{d\}) \end{aligned}$$

Example (Kraft, Pratt, & Seidenberg, 1959)

Let $W = \{a, b, c, d, e\}$:

$$\{d\} \succ \{a, c\} \quad \{b, c\} \succ \{a, d\} \quad \{a, e\} \succ \{c, d\}$$

$$\{a, c, d\} \succ \{b, e\}$$

Fact

$(\wp(W), \succeq)$ admits no agreeing probability measure.

$$\mu(\{b\}) + \mu(\{e\}) \;\; > \;\; \mu(\{a\}) + \mu(\{c\}) + \mu(\{d\})$$

Example (Kraft, Pratt, & Seidenberg, 1959)

Let $W = \{a, b, c, d, e\}$:

$$\{d\} \succ \{a, c\} \quad \{b, c\} \succ \{a, d\} \quad \{a, e\} \succ \{c, d\}$$

$$\{a, c, d\} \succ \{b, e\}$$

Fact

$(\wp(W), \succeq)$ admits no agreeing probability measure.

$$\mu(\{b, e\}) > \mu(\{a, c, d\})$$

World Cup

- Denmark is more likely to win than either of Argentina or China.
- One of Argentina or England is more likely to win than China or Denmark.
- One of Brazil or China is more likely than one of Argentina or Denmark.
- One of Argentina, China, or Denmark is more likely than Brazil or England.

For finite sequences of events $E_0, \ldots, E_n$ and $F_0, \ldots, F_n$, write

$$(E_0, \ldots, E_n) =_0 (F_0, \ldots, E_n)$$

(the sequences are balanced) if for all $w \in W$,

$$|\{i : w \in E_i\}| \ = \ |\{i : w \in F_i\}| \, .$$

Definition (Kraft et al. 1959, Scott 1964)

$(\mathcal{E}, \succeq)$ satisfies Finite Cancellation (FC) if for all balanced sequences $E_0, \ldots, E_n$ and $F_0, \ldots, F_n$, if $F_i \succeq E_i$ for $i < n$, then

$$E_n \succeq F_n \, .$$

Fact

If $(\mathcal{E}, \succeq)$ is probabilistically representable, then it satisfies FC.

Proof.

Let μ agree with $\succeq$, and $(E_0, \dots, E_n) =_0 (F_0, \dots, F_n)$. Then

$$\sum_{i \leq n} \sum_{w \in E_i} \mu(\{w\}) = \sum_{i \leq n} \sum_{w \in F_i} \mu(\{w\})$$

Since μ is additive, this means

$$\sum_{i \leq n} \mu(E_i) = \sum_{i \leq n} \mu(F_i) \tag{1}$$

If $\mu(F_i) \geq \mu(E_i)$, for $i < n$, then by (1) we must have $\mu(E_n) \geq \mu(F_n)$, and hence $E_n \succeq F_n$. □

Theorem (Scott 1964)

If $(\mathcal{E}, \succeq)$ is a de Finetti order that satisfies FC, it is probabilistically representable.

Proof Sketch.

Consider the vector space generated by linear combinations of indicator functions $\mathbf{1}_E$ for $E \in \mathcal{E}$. Let Γ be the set of pairs $\gamma = E \succeq F$, and let $\overline{\gamma} = \mathbf{1}_E - \mathbf{1}_F$. Let Σ be the set of pairs $\sigma = E \not\succeq F$, and let $\overline{\sigma} = \mathbf{1}_E - \mathbf{1}_F$. Define

$$\mathcal{G} = cone(\{\overline{\gamma} : \gamma \in \Gamma\}) \qquad \mathcal{S} = cone(\{\overline{\sigma} : \sigma \in \Sigma\}) \ .$$

Using the axioms and invoking a separation theorem, one can show there is a vector $\mathbf{v}$ such that

$$E \succeq F \quad \Leftrightarrow \quad \mathbf{v} \cdot (\mathbf{1}_E - \mathbf{1}_F) \geq 0 \ .$$

Proof Sketch.

Consider the vector space generated by linear combinations of indicator functions $\mathbf{1}_E$ for $E \in \mathcal{E}$. Let Γ be the set of pairs $\gamma = E \succeq F$, and let $\overline{\gamma} = \mathbf{1}_E - \mathbf{1}_F$. Let Σ be the set of pairs $\sigma = E \not\succeq F$, and let $\overline{\sigma} = \mathbf{1}_E - \mathbf{1}_F$. Define

$$\mathcal{G} = cone(\{\overline{\gamma} : \gamma \in \Gamma\}) \qquad \mathcal{S} = cone(\{\overline{\sigma} : \sigma \in \Sigma\}) \ .$$

Using the axioms and invoking a separation theorem, one can show there is a vector $\mathbf{v}$ such that

$$E \succeq F \quad \Leftrightarrow \quad \mathbf{v} \cdot (\mathbf{1}_E - \mathbf{1}_F) \geq 0 \ .$$

Let $v(E) = \mathbf{v} \cdot \mathbf{1}_E$ $\left[\text{note } v(E) \geq 0 \text{ for all } E \in \mathcal{E}\right]$ and define

$$\mu(E) = \frac{v(E)}{v(W)} \ .$$

Then μ is a probability measure that agrees with $\succeq$. □

One can couch all of this in a modal logical setting.

$$\varphi \quad ::= \quad A \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \succsim \varphi)$$

Natural models are triples $\mathcal{M} = \langle W, \mu, V \rangle$ such that $\mu : \wp(W) \to [0, 1]$ is a probability function. Crucial clause:

$$\mathcal{M}, w \vDash \varphi \succsim \psi \quad \text{iff} \quad \mu([\![\varphi]\!]^{\mathcal{M}}) \geq \mu([\![\psi]\!]^{\mathcal{M}}).$$

Theorem (Segerberg 1971, Gärdenfors 1975)

The complete logic of probability measure models is given by boolean tautologies, *modus ponens*, and the following:

From φ infer $(\varphi \succsim \top)$

$$\Big((\varphi_1 \rightarrow \varphi_2) \succsim \top \wedge (\psi_2 \rightarrow \psi_1) \succsim \top\Big) \rightarrow \Big((\varphi_1 \succsim \psi_1) \rightarrow (\varphi_2 \succsim \psi_2)\Big)$$

$$\varphi \succsim \bot$$

$$\neg(\bot \succsim \top)$$

$$(\varphi \succsim \psi) \vee (\psi \succsim \varphi)$$

$$\varphi_1 \dots \varphi_n \mathbb{E} \psi_1 \dots \psi_n \rightarrow \Big(\big(\bigwedge_{i<n} (\varphi_i \succsim \psi_i)\big) \rightarrow (\psi_n \succsim \varphi_n)\Big)$$

An alternative (Kraft et al. 1959, Burgess 2010): add to de Finetti's quasi-additivity a polarization rule.

$$\text{From } (\alpha \wedge A) \approx (\alpha \wedge \neg A) \rightarrow \varphi \text{ infer } \varphi.$$

Argument for soundness: if $\neg\varphi$ is satisfiable, show it is also satisfiable together with $(\alpha \wedge A) \approx (\alpha \wedge \neg A)$ by "duplicating" the extension of α (where A is fresh).

What happens if we add addition over probability terms?

$$\mathbf{P}(\varphi) \approx \mathbf{P}(\varphi \wedge \psi) + \mathbf{P}(\varphi \wedge \neg\psi)$$

$$\mathbf{P}(\varphi) \approx \mathbf{P}(\varphi \land \psi) + \mathbf{P}(\varphi \land \neg\psi)$$

$$a + (b + c) \approx (a + b) + c$$

$$a + b \approx b + a$$

$$a + 0 \approx a$$

$$(a + e \succsim c + f \land b + f \succsim d + e) \to a + b \succsim c + d$$

$$(a + b \succsim c + d \land d \succsim b) \to a \succsim c$$

Theorem

1. The additive system is finitely axiomatizable; there is no finite axiomatization for the purely comparative system.
2. Moreover, both systems are decidable in NP-time.
3. Both admit models in (natural or) rational numbers.

Ibeling, Icard, Mierzewski, and Mossé, Probing the Qualitative-Quantitative Distinction in Probability Logics. Manuscript.

$$A \big| B \succeq C \big| D$$

$$A \perp\!\!\!\perp B$$

$$H \big| E \succ H$$

$$\alpha|\beta \gtrsim \gamma|\delta$$

$$\alpha \perp\!\!\!\perp \beta$$

$$\alpha|\beta \succ \alpha$$

Example

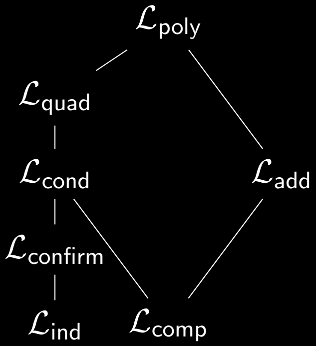
$$(\alpha \wedge \beta) \approx \neg(\alpha \wedge \beta)$$
$$\alpha|\beta \approx \beta$$

Any probability model will have $\mu([\![\beta]\!]) = 1/\sqrt{2}$.

We could also allow explicit multiplication, just as we previously added addition.

$$\mathbf{P}(\alpha)^3 + 5 \cdot \mathbf{P}(\beta)^2 \succsim \mathbf{P}(\gamma) - \mathbf{P}(\theta)\mathbf{P}(\beta)$$

An Expressive Hierarchy



Ibeling, Icard, Mierzewski, and Mossé, Probing the Qualitative-Quantitative Distinction in Probability Logics. Manuscript.

The polynomial system

Add to the axioms of additive probability logic:

$$a \cdot (b \cdot c) \approx (a \cdot b) \cdot c$$

$$a \cdot b \approx b \cdot a$$

$$a \cdot 0 \approx 0$$

$$a \cdot 1 \approx a$$

$$c \succ 0 \to (a \cdot c \gtrsim b \cdot c \leftrightarrow a \gtrsim b)$$

$$a \cdot (b + c) \approx a \cdot b + a \cdot c$$

$$a \gtrsim b \wedge c \gtrsim d \to a \cdot c + b \cdot d \gtrsim a \cdot d + b \cdot c$$

Completeness by Positivstellensatz (Krivine 1964).

Complexity

ETR is the class of all sentences of the form

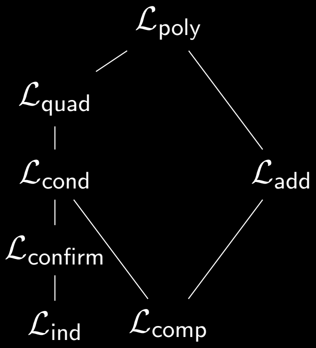
$$\exists x_1 \ldots \exists x_n \varphi,$$

with φ quantifier-free in the language of first-order arithmetic.

$\exists\mathbb{R}$ is the complexity class for ETR. NP $\subseteq \exists\mathbb{R} \subseteq$ PSPACE.

Theorem (Ibeling, Icard, Mierzewski & Mossé)

Satisfiability for the polynomial probability calculus is $\exists\mathbb{R}$-complete. So is it for all other (even minimally) multiplicative languages: comparative conditionals, independence, confirmation, etc.


$\mathcal{L}_{\text{poly}}$
$\mathcal{L}_{\text{quad}}$
$\mathcal{L}_{\text{cond}}$
$\mathcal{L}_{\text{add}}$
$\mathcal{L}_{\text{confirm}}$
$\mathcal{L}_{\text{ind}}$
$\mathcal{L}_{\text{comp}}$

Conclusion and Look Ahead

- Probability can be seen as an axiomatic subject. This already brings in issues central to logic.
- On one way of thinking about justification for the probability axioms, the operative notion is consistency, on a par with ordinary deductive logic.
- Devising probabilistic logical languages allows us to study probabilistic reasoning in explicitly logical terms, manifesting a rich landscape of systems.
- Next time we will continue on the qualitative/quantitative distinction, especially as it relates to important aspects of reasoning (default inference, acceptance, etc.).