Self-Referential Probabilities

And a Kripkean semantics

Catrin Campbell-Moore

Munich Center for Mathematical Philosophy

Bridges 2
September 2015

Supported by

Alexander von Humboldt
Stiftung/Foundation

Introduction

Languages that can talk about probability
where probabilistic liars are expressible.

PrLiar:

The probability of this very sentence is less than $1/2$.

or

Alice will get a promotion just if she believe she'll get the promotion to degree less than $1/2$.

These languages are interesting because they're expressively rich
But such probabilistic liars can cause problems. . .

Outline

Outline

Why such languages?

Consider formal languages that can express:

- "Annie believes to degree 0.99 that Billy believes to degree $1/2$ that the coin will land heads."

$$\mathsf{P}^{\mathrm{A}}_{=0.99}\ulcorner \mathsf{P}^{\mathrm{B}}_{=1/2}\ulcorner H\urcorner\urcorner$$

- "Every sentence has non-negative probability."

$$\forall\ulcorner\varphi\urcorner(\mathsf{P}^{\mathrm{A}}_{\geqslant 0}\ulcorner\varphi\urcorner)$$

- "Alice's degree of belief in this very sentence is less than $1/2$."

$$\mathit{PrLiar} \leftrightarrow \mathsf{P}^{\mathrm{A}}_{<1/2}\ulcorner \mathit{PrLiar}\urcorner$$

What is probability?

Probability is some $\mathrm{p} : \mathsf{Sent}_L \to \mathbb{R}$ with:

- $\mathrm{p}(\varphi) \geqslant 0$ for all φ,
- $\mathrm{p}(q \vee \neg q) = 1$,
- $\mathrm{p}(\varphi \vee \psi) = \mathrm{p}(\varphi) + \mathrm{p}(\psi)$ for φ and ψ logically incompatible.

Many possible applications of the probability notion. E.g.

- Subjective probability, degrees of belief of an agent,
- Objective chance,
- Evidential support,
- "Semantic probability",
- ...

I focus on subjective probability.
But work can apply to many probability notions.

Connection to the liar

Liar:

"This very sentence is not true"

Leads to contradictions under basic assumptions about truth.

PrLiar leads to contradictions between the axioms of probability and some seemingly harmless principles.

E.g. Introspection:

The agent is certain about her own degrees of belief.

$$\mathsf{P}_{<1/2}\ulcorner\varphi\urcorner \implies \mathsf{P}_{=1}\ulcorner\mathsf{P}_{<1/2}\ulcorner\varphi\urcorner\urcorner$$

and Deference, e.g. Lewis's Principal Principle:

The agent A defers to B.

$$\mathsf{P}^{\mathrm{A}}(\ulcorner\varphi\urcorner \mid \ulcorner\mathsf{P}^{\mathrm{B}}_{<1/2}\ulcorner\varphi\urcorner\urcorner) < 1/2$$

Important questions:

"How can one develop a formal semantics for this language?"

- Allows us to determine consistencies and inconsistencies.
- Should also develop a corresponding axiomatic theory.

"To what degree should a rational agent believe such sentences?"

- Are agents who do best from an accuracy or Dutch book perspective representable in the proposed semantics?
- Caie (2013) argued that usual arguments for probabilistic coherence, e.g. accuracy, need to be re-considered. (Campbell-Moore, ta; Konek and Levinstein, ms).

Outline

What we need in the underlying models

Want to develop a semantics that can model agents' degrees of belief.

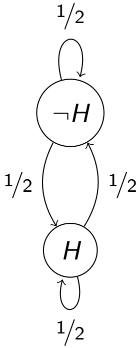
Need to include facts about the probability directly into the model.

We will use a possible world structures.

Underlying models

Probabilistic modal structures are used in economics and game theory to model interacting agents' beliefs about one another.

Definition (Probabilistic Modal Structure (finite))



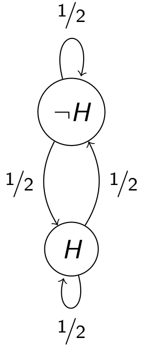
- W a set, called "possible worlds".
- For each $w \in W$, a model $\mathbf{M}(w)$ of the language without probability and truth
  - We assume true arithmetic $\mathbb{N}$.
- For each agent A and worlds w and v, a "degree of accessibility of v from w" $d^{\mathrm{A}}_{w,v}$
  - $\sum_{v \in W} d^{\mathrm{A}}_{w,v} = 1$
  - More generally, m_w probability measure over W.

What does this presuppose?

Assumes that the agents are probabilistically coherent
and that they have common knowledge of probabilistic coherency.

So we're modelling rational agents.

The desired semantics definition (which won't work)



For example we want to be able to say:

- $w_H \models H$,
- $w_{\neg H} \models \neg H$.
- $w_H \models \mathrm{P}_{=1/2}\ulcorner H \urcorner$.

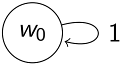
Following this intuition, try to define:

$$w \models \mathrm{P}_{=r}\ulcorner \varphi \urcorner \iff \sum_{v \models \varphi} d_{w,v} = r$$

$$\iff m_w\{v | v \models \varphi\} = r$$

This definition sometimes doesn't work

$PrLiar \leftrightarrow \mathrm{P}_{<1/2}\ulcorner PrLiar \urcorner$



$$
\begin{aligned}
w_0 \models PrLiar &\implies \sum_{v \models PrLiar} d_{w_0,v} = 1 \\
&\implies w_0 \models \mathrm{P}_{=1}\ulcorner PrLiar \urcorner \\
&\implies w_0 \not\models \mathrm{P}_{<1/2}\ulcorner PrLiar \urcorner \\
&\implies w_0 \not\models PrLiar
\end{aligned}
$$

And similarly $w_0 \not\models PrLiar \implies w_0 \models PrLiar$.

So this leads to contradictions.

The definition *often* doesn't work

Similar problems for all finite probabilistic modal structures.
(For infinite: at least whenever the m_w are all countably additive)

Otherwise the agent would have fully introspected certainty of being probabilistic and satisfying

$$\mathrm{p}(\forall x \in N \varphi(x)) = \lim_n \mathrm{p}(\varphi(\overline{0}) \wedge \ldots \wedge \varphi(\overline{n}))).$$

Which is not possible because of:

I do not have fully introspected certainty in this very sentence.

$$\gamma \leftrightarrow \neg \forall n \in N \overbrace{\mathrm{P}_{=1} \ulcorner \mathrm{P}_{=1} \ulcorner \ldots \mathrm{P}_{=1} \ulcorner \gamma \urcorner \urcorner \urcorner}^{n+1}$$

(McGee, 1985; Halbach et al., 2003)

What to do?

Liar paradox leads to similar challenges.

There has been work developing semantics and theories of truth.
And some generalisations of these for modal predicates.
We can generalise these.

Options for semantics

- **Kripke-style semantics,** (Kripke, 1975).
  (Halbach and Welch, 2009; Stern, 2014b).
  - **Based on Strong-Kleene logic,** (Campbell-Moore, 2015)
  - Based on supervaluational logic.
- Revision theory, (Gupta and Belnap, 1993).
  (Leitgeb, 2012; Campbell-Moore, ms; Horsten, ms)

Outline

The Setup

Language with predicates $\mathrm{P}_{\geqslant}$ and T.

In fact $\mathrm{P}^{\mathrm{A}}_{\geqslant}$ for each agent A.

Sentences like

$$\begin{array}{ll} \mathrm{P}_{\geqslant r}\ulcorner\varphi\urcorner & \text{the probability of } \varphi \text{ is } \geqslant r \\ \mathrm{T}\ulcorner\varphi\urcorner & \varphi \text{ is true} \end{array}$$

Or actually $\mathrm{P}_{\geqslant}(\ulcorner\varphi\urcorner, \ulcorner r\urcorner)$

Other probability clauses are defined.

- $\mathrm{P}_{>r}\ulcorner\varphi\urcorner := \exists s > r(\mathrm{P}_{\geqslant s}\ulcorner\varphi\urcorner)$

  Actually: $\mathrm{P}_{>}(t', t) := \exists x \succ t(\mathrm{P}_{\geqslant}(t', x))$
- $\mathrm{P}_{\leqslant r}\ulcorner\varphi\urcorner := \mathrm{P}_{\geqslant 1-r}\ulcorner\neg\varphi\urcorner$
- $\mathrm{P}_{<r}\ulcorner\varphi\urcorner := \mathrm{P}_{>1-r}\ulcorner\neg\varphi\urcorner$
- $\mathrm{P}_{=r}\ulcorner\varphi\urcorner := \mathrm{P}_{\geqslant r}\ulcorner\varphi\urcorner \wedge \mathrm{P}_{\leqslant r}\ulcorner\varphi\urcorner$

The Semantics

Generalise Kripke (1975).

- Formalising the process of evaluating a sentence φ.
- To evaluate $\mathrm{T}\ulcorner\varphi\urcorner$ one first needs to evaluate φ.
- E.g.:

  First evaluate $0 = 0$ positively.
  Then evaluate $\mathrm{T}\ulcorner 0 = 0\urcorner$ positively.
  And $\mathrm{T}\ulcorner 0 = 0\urcorner \vee \mathit{Liar}$ positively.
  Then $\mathrm{T}\ulcorner\mathrm{T}\ulcorner 0 = 0\urcorner\urcorner$ positively.
  And keep going. . .

- Until the evaluation process doesn't lead to anything new.

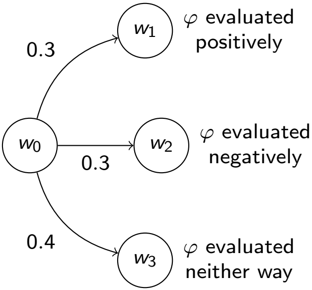
  Called a *fixed point*. These are the good evaluations.

To evaluate $\mathrm{P}_{\geqslant r}\ulcorner\varphi\urcorner$ we first evaluate φ in different possible states of affairs.

Use a *Probabilistic Modal Structure*

How to evaluate probability

Start with some evaluation.
How should we evaluate $\mathrm{P}_{\geqslant r}\ulcorner\varphi\urcorner$ at w_0 in the next stage?



If φ were evaluated at w_3:	Probability of φ would be:
positively	0.7
negatively	0.3

At next stage we can evaluate:

- $\mathrm{P}_{\geqslant 0.3}\ulcorner\varphi\urcorner$ positively,
- $\mathrm{P}_{\geqslant 0.8}\ulcorner\varphi\urcorner$ negatively,
- $\mathrm{P}_{\geqslant 0.5}\ulcorner\varphi\urcorner$ neither way.

Understanding the evaluations

These evaluations give ranges for the probability of φ

$$\underline{\mathbf{p}}^w(\varphi) := \sum_{\substack{\varphi \text{ evaluated} \\ \text{positively}}} d_{w,v} \qquad \overline{\mathbf{p}}^w(\varphi) := 1 - \sum_{\substack{\varphi \text{ evaluated} \\ \text{negatively}}} d_{w,v}$$

Also at the fixed point evaluations.

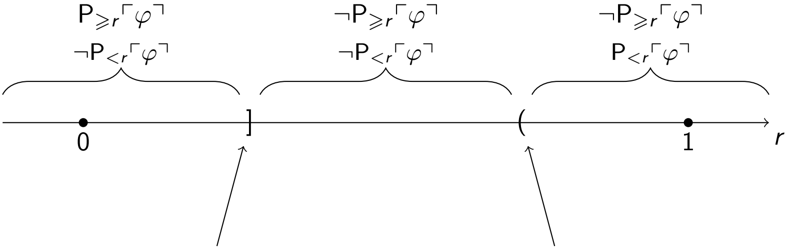
Axiomatisation

We give axioms that allow us to reason about these models syntactically.

Using classical logic.
Consider an *induced model*.

The classical induced model

For f a fixed point evaluation define induced model $\mathrm{IM}_{\mathfrak{M}}[w,f] \models$:



$$\underline{\mathbf{p}}^{w}(\varphi) := \sum_{\substack{\varphi \text{ evaluated} \\ \text{positively} \\ \text{by f(v)}}} d_{w,v}$$

$$\overline{\mathbf{p}}^{w}(\varphi) := 1 - \sum_{\substack{\varphi \text{ evaluated} \\ \text{negatively} \\ \text{by f(v)}}} d_{w,v}$$

ProbKFC

Informally written
$\underline{\mathbf{p}}\ulcorner\varphi\urcorner := \sup\{r \mid \mathrm{P}_{\geqslant r}\ulcorner\varphi\urcorner\}$

- Some basic facts to get things working. E.g.:
  - $\forall a \exists x (\mathrm{P}_{\geqslant a} x \to \mathrm{Rat}(a))$
  - $\mathrm{P}_{\geqslant r}\ulcorner\varphi\urcorner \leftrightarrow \forall s < r \mathrm{P}_{\geqslant s}\ulcorner\varphi\urcorner$
- KFC extended to include probability. E.g.:
  - $\mathrm{T}\ulcorner\varphi \vee \psi\urcorner \leftrightarrow (\mathrm{T}\ulcorner\varphi\urcorner \vee \mathrm{T}\ulcorner\psi\urcorner)$
  - $\mathrm{T}\ulcorner\neg(\varphi \vee \psi)\urcorner \leftrightarrow (\mathrm{T}\ulcorner\neg\varphi\urcorner \wedge \mathrm{T}\ulcorner\neg\psi\urcorner)$
  - $\mathrm{T}\ulcorner\mathrm{P}_{\geqslant r}\ulcorner\varphi\urcorner\urcorner \leftrightarrow \mathrm{P}_{\geqslant r}\ulcorner\varphi\urcorner$
  - $\mathrm{T}\ulcorner\neg\mathrm{P}_{\geqslant r}\ulcorner\varphi\urcorner\urcorner \leftrightarrow \mathrm{P}_{< r}\ulcorner\varphi\urcorner$
  - $\neg(\mathrm{T}\ulcorner\varphi\urcorner \wedge \mathrm{T}\ulcorner\neg\varphi\urcorner)$
- Axioms which say that P acts like a probability over the logic inside T:
  - $\underline{\mathbf{p}}\ulcorner 0 = 0\urcorner = 1$
  - $\underline{\mathbf{p}}\ulcorner 0 = 1\urcorner = 0$
  - $\underline{\mathbf{p}}\ulcorner\varphi\urcorner + \underline{\mathbf{p}}\ulcorner\psi\urcorner = \underline{\mathbf{p}}\ulcorner\varphi \wedge \psi\urcorner + \underline{\mathbf{p}}\ulcorner\varphi \vee \psi\urcorner$
  - $\dfrac{\mathrm{T}\ulcorner\varphi\urcorner \to \mathrm{T}\ulcorner\psi\urcorner}{\underline{\mathbf{p}}\ulcorner\varphi\urcorner \leqslant \underline{\mathbf{p}}\ulcorner\psi\urcorner}$

These axioms are complete(ish)

This is complete with the ω-rule $\dfrac{\varphi(\bar{0}) \qquad \varphi(\bar{1}) \qquad \ldots}{\forall x \varphi(x)}$

Theorem

$\Gamma \vdash^{\omega}_{\mathrm{ProbKFC}} \varphi$ if and only if for each probabilistic modal structure $\mathfrak{M}$, consistent fixed point evaluation f, and $w \in W$,

$$\mathrm{IM}_{\mathfrak{M}}[w, f] \models \Gamma \implies \mathrm{IM}_{\mathfrak{M}}[w, f] \models \varphi$$

Proof.

Completeness via a canonical model construction

- $W := \left\{ w \subseteq \mathrm{Sent}_{\mathrm{P,T}} \;\middle|\; \begin{array}{l} w \text{ is maximally finitely } \vdash^{\omega}_{\mathrm{ProbKFC}}\text{-consistent} \\ \text{and closed under } \omega\text{-rule} \end{array} \right\}$
- $\mathbf{M}(w) \models \varphi \iff \varphi \in w$
- φ evaluated positively by $f(w)$ iff $\mathrm{T}\ulcorner\varphi\urcorner \in w$
- Find $d_{w,v}$ with $\sum_{\mathrm{T}\ulcorner\varphi\urcorner \in v} d_{w,v} = \sup\{r \mid \mathrm{P}_{\geqslant r}\ulcorner\varphi\urcorner \in w\}$

$\square$

Reformulating introspection

Express introspection by

$$\mathrm{T}\ulcorner \mathrm{P}_{<1/2}\ulcorner \varphi \urcorner\urcorner \rightarrow \mathrm{P}_{=1}\ulcorner \mathrm{P}_{<1/2}\ulcorner \varphi \urcorner\urcorner$$

instead of

$$\mathrm{P}_{<1/2}\ulcorner \varphi \urcorner \rightarrow \mathrm{P}_{=1}\ulcorner \mathrm{P}_{<1/2}\ulcorner \varphi \urcorner\urcorner.$$

This is consistent and is satisfied in exactly the $\mathrm{IM}_{\mathfrak{M}}[w, f]$ where the operator

$$\mathbb{P}_{<1/2}\varphi \rightarrow \mathbb{P}_{=1}\mathbb{P}_{<1/2}\varphi$$

is satisfied.

General strategy (Stern, 2014a)

Outline

Conclusions

- Want expressively rich languages, with probabilistic liars.
- Probabilistic liars can lead to challenges. E.g.:
  - Undesirable conflicts with other principles.
  - Challenges in determining rational requirements of agents.
  - Challenges in giving semantics.
- I developed a semantics by applying Kripke's theory of truth to probabilistic modal structures.
  - "Problematic" sentences are generally assigned ranges for probability.
  - Can now understand these languages better
    - E.g. see how to express introspection.
- We can also give a (sort of) complete axiomatisation allowing us to reason about these languages axiomatically.

Thanks!

References I

Michael Caie. Rational probabilistic incoherence. *Philosophical Review*, 122(4):527–575, 2013.

Catrin Campbell-Moore. How to express self-referential probabilities. *Review of Symbolic Logic*, 2015. http://dx.doi.org/10.1017/S1755020315000118.

Catrin Campbell-Moore. The revision theory of probability. ms.

Catrin Campbell-Moore. Rational probabilistic incoherence? A reply to Michael Caie. *Philosophical Review*, ta. Forthcoming.

Anil Gupta and Nuel Belnap. The revision theory of truth. *MIT Press, Cambridge*, 1(99):3, 1993.

Volker Halbach and Philip Welch. Necessities and necessary truths: A prolegomenon to the use of modal logic in the analysis of intensional notions. *Mind*, 118(469):71–100, 2009.

References II

Volker Halbach, Hannes Leitgeb, and Philip Welch. Possible-worlds semantics for modal notions conceived as predicates. *Journal of Philosophical Logic*, 32:179–222, 2003.

Leon Horsten. On revising probability and truth. ms.

Jason Konek and Ben Levinstein. The foundations of epistemic decision theory. Accessed Oct 31. 2014., ms.

S. Kripke. Outline of a theory of truth. *The journal of philosophy*, 72(19):690–716, 1975.

Hannes Leitgeb. From type-free truth to type-free probability. In Greg Restall and Gillian Russel, editors, *New Waves in Philosophical Logic edited by Restall and Russell*, pages 84–94. Palgrave Macmillan, 2012.

References III

Vann McGee. How truthlike can a predicate be? A negative result. *Journal of Philosophical Logic*, 14(4):399–410, 1985.

Johannes Stern. Modality and axiomatic theories of truth I: Friedman-Sheard. *The Review of Symbolic Logic*, 7(02): 273–298, 2014a.

Johannes Stern. Modality and axiomatic theories of truth II: Kripke-Feferman. *The Review of Symbolic Logic*, 7(02): 299–318, 2014b.