Self-Referential Probabilities
And a Kripkean semantics

Catrin Campbell-Moore

Munich Center for Mathematical Philosophy

Bridges 2
September 2015

Supported by
Introduction

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

*PrLiar*:

The probability of this very sentence is less than $\frac{1}{2}$.

or

*Alice will get a promotion just if she believe she’ll get the promotion to degree less than $\frac{1}{2}$.*

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

Introduction

Self-referential probabilities
  Why are they interesting?
  Some problems

Semantics
  The setup
  A challenge

Kripkean semantics
  The semantics
  An axiomatic theory
  Avoiding inconsistencies

Conclusions
Outline

Introduction

Self-referential probabilities

Why are they interesting?
Some problems

Semantics

The setup
A challenge

Kripkean semantics

The semantics
An axiomatic theory
Avoiding inconsistencies

Conclusions
Consider formal languages that can express:

- “Annie believes to degree 0.99 that Billy believes to degree \( \frac{1}{2} \) that the coin will land heads.”

\[
P_A \equiv 0.99 \quad P_B \equiv \frac{1}{2} \quad \neg H_\perp
\]

- “Every sentence has non-negative probability.”

\[
\forall \neg \varphi (P_A \geq 0 \quad \neg \varphi_\perp)
\]

- “Alice’s degree of belief in this very sentence is less than \( \frac{1}{2} \).”

\[
PrLiar \leftrightarrow P_A \equiv \frac{1}{2} \quad \neg PrLiar_\perp
\]
What is probability?

Probability is some $p : \text{Sent}_L \rightarrow \mathbb{R}$ with:

- $p(\varphi) \geq 0$ for all $\varphi$,
- $p(q \lor \neg q) = 1$,
- $p(\varphi \lor \psi) = p(\varphi) + p(\psi)$ for $\varphi$ and $\psi$ 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.

\[ P_{\frac{1}{2}} \neg \varphi \Rightarrow P_{=1} \neg P_{\frac{1}{2}} \neg \varphi \]

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

The agent A defers to B.

\[ P_A(\neg \varphi | \neg P_{\frac{1}{2}} \neg \varphi) < \frac{1}{2} \]
Some problems

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

Introduction

Self-referential probabilities
  Why are they interesting?
  Some problems

Semantics
  The setup
  A challenge

Kripkean semantics
  The semantics
  An axiomatic theory
  Avoiding inconsistencies

Conclusions
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 $M(w)$ 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 $w$” $d_{w,v}^A$:
  - $\sum_{v \in W} d_{w,v}^A = 1$
  - More generally, $m_w$ probability measure over $W$. 

\[ \begin{array}{c}
\text{1/2} \\
\text{1/2} \\
\text{1/2}
\end{array} \]

\[ \begin{array}{c}
\neg H \\
H \\
\end{array} \]
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:

- \( w_H \models H \),
- \( w_{\neg H} \models \neg H \).
- \( w_H \models P_{1/2} \neg H \).

Following this intuition, try to define:

\[
 w \models P_{=r} \neg \varphi \iff \sum_{v \models \varphi} d_{w,v} = r \\
\iff m_w \{ v \mid v \models \varphi \} = r
\]
This definition sometimes doesn’t work

\[ PrLiar \leftrightarrow P_{< 1/2} \neg PrLiar \]

- \( w_0 \models PrLiar \implies \sum_{v \models PrLiar} d_{w_0, v} = 1 \)
- \( \implies w_0 \models P_{= 1} \neg PrLiar \)
- \( \implies w_0 \not\models P_{< 1/2} \neg PrLiar \)
- \( \implies w_0 \not\models PrLiar \)

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

$$p(\forall x \in N \varphi(x)) = \lim_{n} p(\varphi(\overline{0}) \land \ldots \land \varphi(\overline{n})).$$

Which is not possible because of:

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

$$\gamma \leftrightarrow \neg \forall n \in N \left( P=1 \right)^{n+1}$$

(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

  - 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

Introduction

Self-referential probabilities
  Why are they interesting?
  Some problems

Semantics
  The setup
  A challenge

Kripkean semantics
  The semantics
  An axiomatic theory
  Avoiding inconsistencies

Conclusions
The Setup

Language with predicates $P_{\geq}$ and $T$.

In fact $P_{\geq}^A$ for each agent $A$.

Sentences like

$$P_{\geq} r \varphi \quad \text{the probability of } \varphi \text{ is } \geq r$$

$$T \varphi \quad \varphi \text{ is true}$$

Or actually $P_{\geq}(\varphi, r)$

Other probability clauses are defined.

- $P_{\geq} r \varphi := \exists s > r (P_{\geq} s \varphi)$
  
  Actually: $P_{\geq} (t', t) := \exists x > t (P_{\geq} (t', x))$

- $P_{\leq} r \varphi := P_{\geq} (1 - r) \neg \varphi$
- $P < r \varphi := P_{\geq} (1 - r) \neg \varphi$
- $P = r \varphi := P_{\geq} r \varphi \land P_{\leq} r \varphi$
The Semantics

Generalise Kripke (1975).

- Formalising the process of evaluating a sentence $\varphi$.
- To evaluate $T^\downarrow \varphi \downarrow$ one first needs to evaluate $\varphi$.
- E.g.:
  
  First evaluate $0 = 0$ positively.
  Then evaluate $T^\downarrow 0 = 0 \downarrow$ positively.
  And $T^\downarrow 0 = 0 \downarrow \lor$ Liar positively.
  Then $T^\downarrow T^\downarrow 0 = 0 \downarrow \downarrow$ 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 $P_{\geq r}^\downarrow \varphi \downarrow$ we first evaluate $\varphi$ in different possible states of affairs.

Use a Probabilistic Modal Structure
How to evaluate probability

Start with some evaluation.
How should we evaluate $P_{\geq r} \neg \varphi$ at $w_0$ in the next stage?

If $\varphi$ were evaluated at $w_3$: would be:

<table>
<thead>
<tr>
<th>Evaluation</th>
<th>Probability of $\varphi$</th>
</tr>
</thead>
<tbody>
<tr>
<td>positively</td>
<td>0.7</td>
</tr>
<tr>
<td>negatively</td>
<td>0.3</td>
</tr>
</tbody>
</table>

At next stage we can evaluate:

- $P_{\geq 0.3} \neg \varphi$ positively,
- $P_{\geq 0.8} \neg \varphi$ negatively,
- $P_{\geq 0.5} \neg \varphi$ neither way.
Understanding the evaluations

These evaluations give ranges for the probability of $\varphi$

\[
P \geq r \left\langle \varphi \right\rangle \text{ evaluated positively at next stage}
\]

\[
P \geq r \left\langle \varphi \right\rangle \text{ evaluated negatively at next stage}
\]

\[
p^w(\varphi) := \sum_{\varphi \text{ evaluated positively}} d_{w,v}
\]

\[
\overline{p}^w(\varphi) := 1 - \sum_{\varphi \text{ evaluated negatively}} d_{w,v}
\]

Also at the fixed point evaluations.
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 \( IM_{\mathbb{M}}[w, f] \models: \)

\[
\begin{align*}
P \geq r & \models \varphi \neg \\
\neg P & \geq r \models \varphi \neg \\
\neg P & < r \models \varphi \neg \\
P & < r \models \varphi \neg
\end{align*}
\]

\[
\begin{align*}
\mathbf{p}^w(\varphi) & := \sum_{\varphi \text{ evaluated positively by } f(v)} d_{w,v} \\
\overline{\mathbf{p}}^w(\varphi) & := 1 - \sum_{\varphi \text{ evaluated negatively by } f(v)} d_{w,v}
\end{align*}
\]
An axiomatic theory

ProbKFC

- Some basic facts to get things working. E.g.:
  - $\forall a \exists x (P \geq_a x \rightarrow \text{Rat}(a))$
  - $P \geq r \phi \iff \forall s < r P \geq s \phi$

- KFC extended to include probability. E.g.:
  - $T \phi \vee \psi \iff (T \phi \vee T \psi)$
  - $T \neg (\phi \vee \psi) \iff (T \neg \phi \land T \neg \psi)$
  - $T \leq r \phi \iff P \geq r \phi$
  - $T \neg \leq r \phi \iff P < r \phi$
  - $\neg (T \phi \land T \neg \phi)$

- Axioms which say that $P$ acts like a probability over the logic inside $T$:
  - $p^\phi 0 = 0 \_ = 1$
  - $p^\phi 0 = 1 \_ = 0$
  - $p^\phi \phi + p^\phi \psi = p^\phi \phi \land \psi + p^\phi \phi \lor \psi$
  - $T \phi \rightarrow T \psi$
  - $p^\phi \phi \leq p^\phi \psi$
An axiomatic theory

**These axioms are complete(ish)**

This is complete with the $\omega$-rule

$$
\frac{\varphi(0) \quad \varphi(1) \quad \ldots}{\forall x \varphi(x)}
$$

**Theorem**

$$
\Gamma \vdash_{\text{ProbKFC}} \varphi \text{ if and only if for each probabilistic modal structure } M, \text{ consistent fixed point evaluation } f, \text{ and } w \in W,
$$

$$
\text{IM}_M[w, f] \models \Gamma \implies \text{IM}_M[w, f] \models \varphi
$$

**Proof.**

Completeness via a canonical model construction

- $W := \left\{ w \subseteq \text{Sent}_{P,T} \mid w \text{ is maximally finitely } \vdash_{\text{ProbKFC}} \text{-consistent and closed under } \omega\text{-rule} \right\}$
- $M(w) \models \varphi \iff \varphi \in w$
- $\varphi$ evaluated positively by $f(w)$ iff $T^\Gamma \varphi \in w$
- Find $d_{w,v}$ with $\sum_{\Gamma \varphi \in v} d_{w,v} = \sup\{ r \mid P \geq r \Gamma \varphi \in w \}$
Avoiding inconsistencies

Reformulating introspection

Express introspection by

\[ T \models P < \frac{1}{2} \models \varphi \models \Rightarrow P \models P = 1 \models P < \frac{1}{2} \models \varphi \models \]

instead of

\[ P < \frac{1}{2} \models \varphi \models \Rightarrow P \models P = 1 \models P < \frac{1}{2} \models \varphi \models . \]

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

\[ P < \frac{1}{2} \varphi \Rightarrow P \models P < \frac{1}{2} \varphi \]

is satisfied.

General strategy (Stern, 2014a)
Outline

Introduction

Self-referential probabilities
  Why are they interesting?
  Some problems

Semantics
  The setup
  A challenge

Kripkean semantics
  The semantics
  An axiomatic theory
  Avoiding inconsistencies

Conclusions
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


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


Leon Horsten. On revising probability and truth. ms.


