A bayesian method of cosine-based word2vec bias estimation

Alicja Dobrzeniecka & Rafal Urbaniak
(LoPSE research group, University of Gdansk)

ExpSem2021, ESSLLI

The gist of the project

The gist of the project

Classical probabilism in epistemology (and in applications)

Represent uncertainty by a single probabilistic measure (point estimates).

$$\text{P}(\text{recovery}|\text{treatment}) = .6$$

The gist of the project

Classical probabilism in epistemology (and in applications)

Represent uncertainty by a single probabilistic measure (point estimates).

$$\mathsf{P}(\text{recovery}|\text{treatment}) = .6$$

Imprecise probabilism

Evidence (e.g. expert evidence) sometimes fails to warrant such precision.

$$\mathsf{P}(\text{recovery}|\text{treatment}) \in [.4, .6]$$

The gist of the project

Classical probabilism in epistemology (and in applications)

Represent uncertainty by a single probabilistic measure (point estimates).

$$\mathrm{P}(\mathrm{recovery}|\mathrm{treatment}) = .6$$

Imprecise probabilism

Evidence (e.g. expert evidence) sometimes fails to warrant such precision.

$$\mathrm{P}(\mathrm{recovery}|\mathrm{treatment}) \in [.4, .6]$$

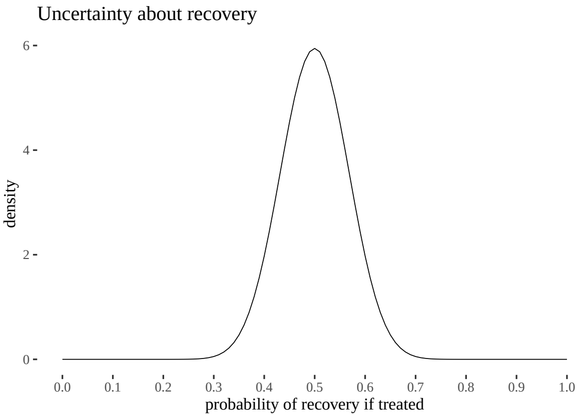
Epistemological difficulties of imprecise probabilism

- hard to model the weight of and combine evidence.
- inertia of agnosticism.
- growth of uncertainty in light of new evidence.

The gist of the project

My novel contribution

- Meet the desiderata motivating imprecision.
- Avoid these difficulties.
- Do so by using higher-order probabilities.



Motivations for imprecision (straighforward considerations)

Motivations for imprecision (straighforward considerations)

Radical lack of information

P(It will snow in Boston on New Year′s Eve 2026) = ?

Motivations for imprecision (straighforward considerations)

Radical lack of information

P(It will snow in Boston on New Year's Eve 2026) $= ?$

Weight of evidence

- You've seen $4/10$ patients recover. P(recovery) $= .4$?

Motivations for imprecision (straighforward considerations)

Radical lack of information

P(It will snow in Boston on New Year′s Eve 2026) = ?

Weight of evidence

- You've seen 4/10 patients recover. P(recovery) = .4?
- You've seen 40/100 patients recover. P(recovery) = .4?

(You could try to use classical statistical tools, such as confidence intervals, but they have their own problems.)

Motivations for imprecision (uniformity and agnosticism)

Agnosticism preservation requirement

Agnosticism about uniquely connected variables should be the same.

$$X \text{ and } Y = X + 2 \text{ or } Z = X^2 \quad \text{for } 10 > X > 0.$$

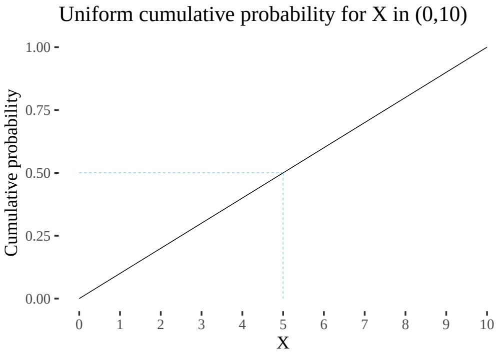
Motivations for imprecision (uniformity and agnosticism)

Agnosticism preservation requirement

Agnosticism about uniquely connected variables should be the same.

$$X \text{ and } Y = X + 2 \text{ or } Z = X^2 \quad \text{for } 10 > X > 0.$$

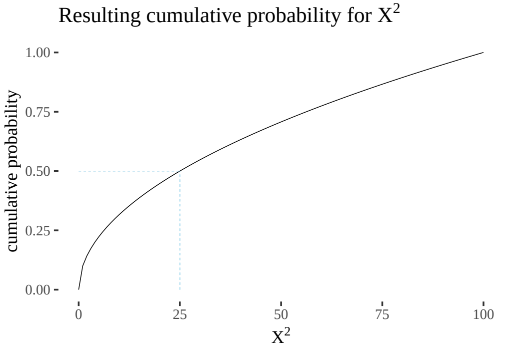
The trouble



$$P(X < 5) = .5 = P(X^2 < 25)$$

Motivations for imprecision (uniformity and agnosticism)

The trouble



Motivations for imprecision (expert evidence)

Weather forecasters

Al:	P(It will rain tomorrow) $= .6$
Bert:	P(It will rain tomorrow) $= .8$

Motivations for imprecision (expert evidence)

Weather forecasters

Al:	P(It will rain tomorrow) $= .6$
Bert:	P(It will rain tomorrow) $= .8$

Linear pooling

P(It will rain tomorrow) $= \alpha \times .6 + \beta \times .8$ (*where* $\alpha + \beta = 1$.)

Motivations for imprecision (expert evidence)

Weather forecasters

Al:	P(It will rain tomorrow) $= .6$
Bert:	P(It will rain tomorrow) $= .8$

Linear pooling

P(It will rain tomorrow) $= \alpha \times .6 + \beta \times .8$ (*where* $\alpha + \beta = 1$.)

Mathematical impossibilities

- Experts's agreement on independence is not preserved by linear pooling. (**???**)
- You can't at the same time hold the following: (**???**)

$$\mathrm{P}(A = B) < 1 \tag{1}$$

$$\mathrm{P}(r|A = a) = a \tag{2}$$

$$\mathrm{P}(r|B = b) = b \tag{3}$$

$$\forall a, b\, \mathrm{P}(r|A = a, B = b) = \alpha a + \beta b \tag{4}$$

Imprecise probabilism and the motivations

The main claim

Represent uncertainty by a set of probability measures.
(**???**, **???**, **???**, **???**, **???**, **???**, **???**; Walley, 1991)

Imprecise probabilism and the motivations

The main claim

Represent uncertainty by a set of probability measures.
(**???**, **???**, **???**, **???**, **???**, **???**, **???**; Walley, 1991)

Snowing in Boston?

Say, (.5,1).
(**???**)

Imprecise probabilism and the motivations

The main claim

Represent uncertainty by a set of probability measures.
(**???**, **???**, **???**, **???**, **???**, **???**, **???**; Walley, 1991)

Snowing in Boston?

Say, (.5,1).
(**???**)

Weight of evidence?

The more evidence you have, the more narrow the interval.
(**???**, **???**, **???**)

Imprecise probabilism and the motivations

The main claim

Represent uncertainty by a set of probability measures.
(**???**, **???**, **???**, **???**, **???**, **???**, **???**; Walley, 1991)

Snowing in Boston?

Say, (.5,1).
(**???**)

Weight of evidence?

The more evidence you have, the more narrow the interval.
(**???**, **???**, **???**)

Agnosticisim?

[0,1] interval instead of a single uniform distribution.

Imprecise probabilism and the motivations

The main claim

Represent uncertainty by a set of probability measures.
(**???**, **???**, **???**, **???**, **???**, **???**, **???**; Walley, 1991)

Snowing in Boston?

Say, (.5,1).
(**???**)

Weight of evidence?

The more evidence you have, the more narrow the interval.
(**???**, **???**, **???**)

Agnosticisim?

[0,1] interval instead of a single uniform distribution.

Multiple sources?

Just put all those probabilities in one set.

Epistemological challenges

Epistemological challenges

Radical uncertainty and weight

Say one experiment observed 40/100 and another 8/15 recoveries.
Simply putting the frequencies in a set won't do justice to the evidence.
(**???**)

Epistemological challenges

Radical uncertainty and weight

Say one experiment observed 40/100 and another 8/15 recoveries. Simply putting the frequencies in a set won't do justice to the evidence.
(**???**)

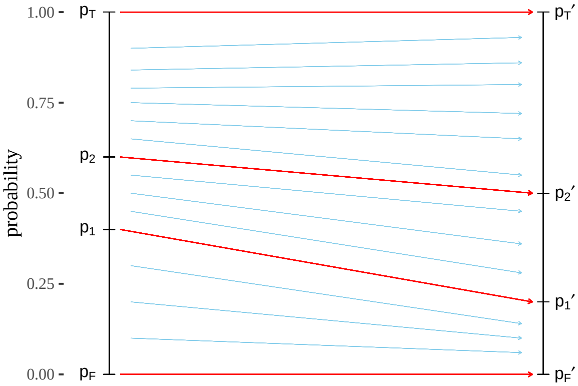
Dilation

Whenever your roomate goes to the track he flips a fair coin. If it comes up heads he bets on Speedy. Otherwise he places no bets. He comes home from the track grinning. You're now more uncertain about P(heads) than before.
(**???**, **???**, **???**, **???**, **???**, **???**; Walley, 1991)

Epistemological challenges

Belief inertia

Say you're fully agnostic about H: $P(H) = [0, 1]$.



Then $\mathrm{P}(H|E) = [0, 1]$, no matter what the evidence says.
(**???**,(**???**), **???**, **???**, **???**, **???**)

Epistemological challenges

Mystery coins

- A fair draw of one of two biased coins, with biases $.4, and .6$.

Epistemological challenges

Mystery coins

- A fair draw of one of two biased coins, with biases $.4, and .6$.
- Imprecise reaction: $\{.4, .6\}$.

Epistemological challenges

Mystery coins

- A fair draw of one of two biased coins, with biases $.4, and .6$.
- Imprecise reaction: $\{.4, .6\}$.
- Precise probabilism: expected bias of .5.

Epistemological challenges

Mystery coins

- A fair draw of one of two biased coins, with biases $.4, and .6$.
- Imprecise reaction: $\{.4, .6\}$.
- Precise probabilism: expected bias of .5.

Question

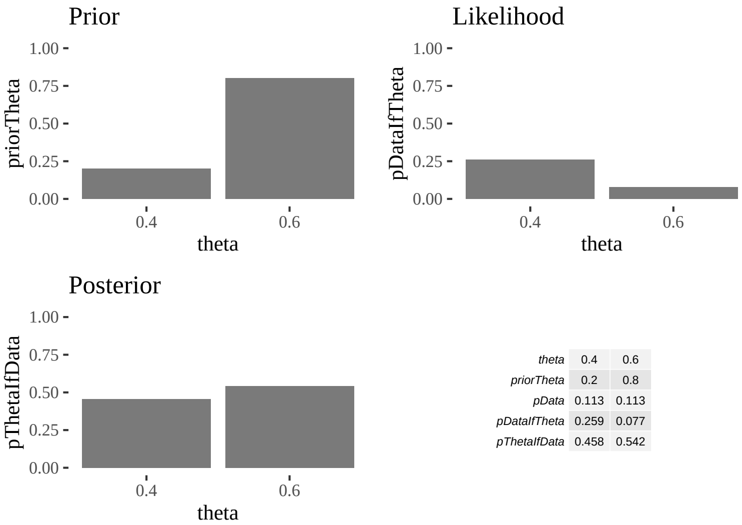
As far as accuracy is concerned, why prefer the imprecise way?
Say the actual bias is .4. A first stab:

$$\text{inaccuracy(imprecise)} = \frac{0 + .2}{2} = .1$$
$$\text{inaccuracy(precise)} = .1$$

(**???**, **???**; Schoenfield, 2017)

Higher-order approach

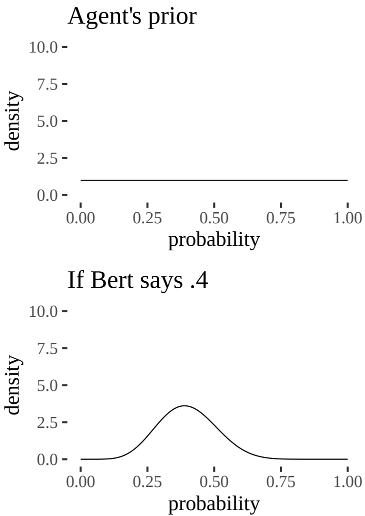
The mystery coin with (.2, .8) and 1 head in five tosses



theta	0.4	0.6
priorTheta	0.2	0.8
pData	0.113	0.113
pDataIfTheta	0.259	0.077
pThetaIfData	0.458	0.542

Higher-order approach

Two experts, uniform prior



The key conjecture

With higher-order methods, it will be possible to:

- match the representation of uncertainty with the evidence,
- combine various sources of uncertainty without running into the known problems,
- use existing mathematics to make sense of accuracy in such contexts,
- capture reasoning with radical uncertainty without dilation and belief inertia.

Tasks (discussion)

- Relate this representation to weight of evidence.
- Go beyond grid approximation and beta representation.
- Represent agnosticism so that:
  - agnosticism preservation holds, and
  - inertia fails.
- Represent opinion aggregation more generally, compare to known limitations and features of already studied opinion pooling methods.
- Investigate dilation-like phenomena in this context.
- Study known distance functions between probabilistic measures as candidates for key elements of accuracy measures.
- Explain away the apparent contrast between evidence-grounding and accuracy considerations (Easwaran & Fitelson, 2015).

Thank you for your attention!

References

Easwaran, K., & Fitelson, B. (2015). Accuracy, coherence, and evidence. In T. S. Gendler & J. Hawthorne (Eds.), *Oxford studies in epistemology* (Vol. 5). Oxford University Press.

Schoenfield, M. (2017). The Accuracy and Rationality of Imprecise Credences: The Accuracy and Rationality of Imprecise Credences. *Noûs*, *51*(4), 667–685. https://doi.org/10.1111/nous.12105

Walley, P. (1991). *Statistical reasoning with imprecise probabilities*. Chapman; Hall London.