Compositional Imprecise Probability

Jack Liell-Cock [1] Sam Staton [1]

[1] University of Oxford

AI-Fun with ELLIS
13th November, 2024

Preprint

arXiv > cs > arXiv:2405.09391

Search... All fields Search

Help | Advanced Search

Computer Science > Programming Languages

[Submitted on 15 May 2024 (v1), last revised 29 Oct 2024 (this version, v2)]

Compositional imprecise probability

Jack Liell-Cock, Sam Staton

Imprecise probability is concerned with uncertainty about which probability distributions to use. It has applications in robust statistics and machine learning.
We look at programming language models for imprecise probability. Our desiderata are that we would like our model to support all kinds of composition, categorical and monoidal; in other words, guided by dataflow diagrams. Another equivalent perspective is that we would like a model of synthetic probability in the sense of Markov categories.
Imprecise probability can be modelled in various ways, with the leading monad-based approach using convex sets of probability distributions. This model is not fully compositional because the monad involved is not commutative, meaning it does not have a proper monoidal structure. In this work, we provide a new fully compositional account. The key idea is to name the non-deterministic choices. To manage the renamings and disjointness of names, we use graded monads. We show that the resulting compositional model is maximal and relate it with the earlier monadic approach, proving that we obtain tighter bounds on the uncertainty.

Comments: Conditionally accepted for POPL 2025
Subjects: **Programming Languages (cs.PL)**; Logic in Computer Science (cs.LO); Category Theory (math.CT); Probability (math.PR)
Cite as: arXiv:2405.09391 **[cs.PL]**
(or arXiv:2405.09391v2 **[cs.PL]** for this version)
https://doi.org/10.48550/arXiv.2405.09391

Submission history

From: Sam Staton [view email]
[v1] Wed, 15 May 2024 14:49:13 UTC (31 KB)
[v2] Tue, 29 Oct 2024 21:31:20 UTC (91 KB)

Access Paper:

- View PDF
- TeX Source
- Other Formats

Current browse context:
cs.PL
< prev | next >
new | recent | 2024-05
Change to browse by:
cs
cs.LO
math
math.CT
math.PR

References & Citations
- NASA ADS
- Google Scholar
- Semantic Scholar

1 blog link (what is this?)

Export BibTeX Citation

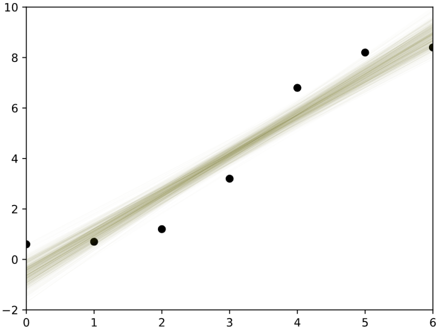
Bookmark

Context

Probabilistic programming: [S. Dash et al. 2023]

- Programming languages designed to encode statistical models
- Commands for sampling from distributions and conditioning

```haskell
regress :: Double -> Prob (a -> Double) -> [(a, Double)] -> Meas (a -> Double)
regress sigma prior dataset =
  do
    f <- sample prior
    forM_ dataset (\(x, y) -> score $ normalPdf (f x) sigma y)
    return f
```



Overview

This work:

- Design a fully compositional probabilistic programming language with Bernoulli and Knightian uncertainty
- Using a graded perspective and named Knightian choices
- **Theorem 1:** This gives a refined bound on uncertainty
- **Theorem 2:** It is maximal among compositional accounts

Overview

This work:

- Design a fully compositional probabilistic programming language with Bernoulli and Knightian uncertainty
- Using a graded perspective and named Knightian choices
- **Theorem 1:** This gives a refined bound on uncertainty
- **Theorem 2:** It is maximal among compositional accounts

Not this work:

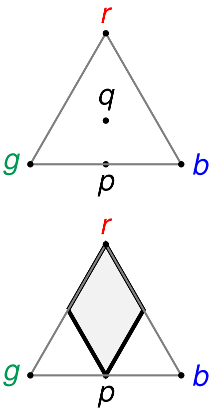
- There is a broader interest in combining non-determinism and probability [Swaraj Dash and Staton 2021; Swaraj Dash and Staton 2020; Dahlqvist et al. 2018; Keimel et al. 2017; Swaraj Dash 2024; Kozen et al. 2023; Varacca et al. 2006; Jacobs 2021]
- Our focus is in the setting of imprecise probability

Contents

Outline

Imprecise Probability

- **Probability** = point in simplex
- **Imprecise probability** = convex set of points



Applications

Wider applicataions:

- Statistical robustness [Walley 1991; Huber 1981],
- Economics [Amarante et al. 2007; Battigalli, Cerreia-Vioglio, et al. 2015; Battigalli, Corrao, et al. 2019; Cerreia-Vioglio et al. 2023; Vicig 2008; Pavlidis et al. 2011]
- Engineering [Ferson et al. 2003; Aslett et al. 2022; Vasile 2021; Krpelík et al. 2021]

In machine learning:

- Bayesian learning [Caprio et al. 2024]
- Reinforcement learning [Oren et al. 2024; Zanger et al. 2023]
- Conformal prediction [Stutz et al. 2023; Javanmardi et al. 2024]
- Foundations of safe AI [Dalrymple 2024; Dalrymple et al. 2024]

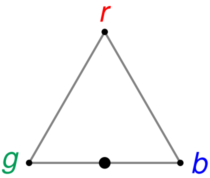
A First Language

Our prototype language for imprecise probability will have:

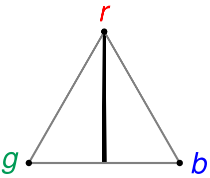
- If-then-else statements;
  - if b then t else u
- Let bindings;
  - $x \leftarrow t \; ; \; u$
- Two commands returning booleans:
  - ***bernoulli***: a fair Bernoulli choice;
  - ***knight***: a Knightian choice.

Examples

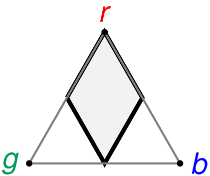
$z \leftarrow$ ***bernoulli*** ;
if z then *g* else *b*



$x \leftarrow$ ***knight*** ; $z \leftarrow$ ***bernoulli*** ;
if z then (if x then *r* else *g*)
else (if x then *r* else *b*)



$x \leftarrow$ ***knight*** ; $y \leftarrow$ ***knight*** ; $z \leftarrow$ ***bernoulli*** ;
if z then (if x then *r* else *g*)
else (if y then *r* else *b*)



Outline

Desiderata

Desideratum (1)

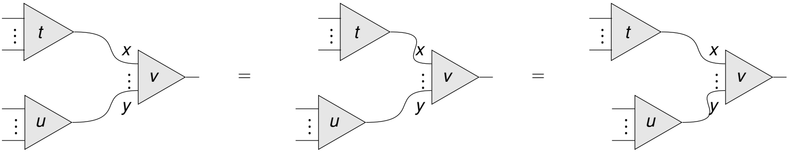
The language should be commutative:

$$x \leftarrow t \; ; \; y \leftarrow u \; ; \; v \quad = \quad y \leftarrow u \; ; \; x \leftarrow t \; ; \; v$$

for $x \notin \mathrm{free}(u)$ and $y \notin \mathrm{free}(t)$; and affine:

$$x \leftarrow t \; ; \; u \quad = \quad u$$

for $x \notin \mathrm{free}(u)$.



Desiderata

Desideratum (2)

Standard equational reasoning about if/then/else should apply:

$$\text{if } b \text{ then } (x \leftarrow t \;;\; u) \text{ else } (x \leftarrow t \;;\; v)$$
$$=$$
$$x \leftarrow t \;;\; \text{if } b \text{ then } u \text{ else } v$$

for $x \notin \mathrm{free}(b)$.

Outline

Monads

- A *monad* is a mathematical abstraction that captures computational effects in programming languages [Moggi 1989; Moggi 1991]
- Consider programs as functions into computational contexts rather than pure values

$$[X \Rightarrow Y] \qquad \rightsquigarrow \qquad [X \Rightarrow T(Y)]$$

- Additional structure for composition (satisfying some laws...) is required

$$\begin{aligned} [X \Rightarrow Y] &\to [X \Rightarrow T(Y)] \\ [X \Rightarrow T(Y)] \times [Y \Rightarrow T(Z)] &\to [X \Rightarrow T(Z)] \\ [X \Rightarrow T(Y)] \times [X' \Rightarrow T(Y')] &\to [X \times X' \Rightarrow T(Y \times Y')] \end{aligned}$$

Example: Non-determinism

- Non-determinism is given by the (non-empty) powerset monad
- Rather than a single result, programs can return a set of results

$$
\begin{array}{ccc}
[X \Rightarrow Y] & \rightsquigarrow & [X \Rightarrow \mathcal{P}(Y)] \\
[x \mapsto y] & \rightsquigarrow & [x \mapsto \{y_1, y_2, y_3\}]
\end{array}
$$

- Additional structure given by singletons, unions and products

$$
[x \mapsto y] \to [x \mapsto \{y\}]
$$

$$
[x_1 \mapsto \{y_1, y_2\}] \times \begin{bmatrix} y_1 \mapsto \{z_1, z_2\} \\ y_2 \mapsto \{z_2, z_3\} \end{bmatrix} \to [x_1 \mapsto \{z_1, z_2, z_3\}]
$$

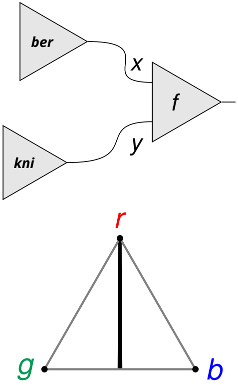
$$
[x \Rightarrow \{y_1\}] \times [x' \mapsto \{y'_1, y'_2\}] \to [(x, x') \mapsto \{(y_1, y'_1), (y_1, y'_2)\}]
$$

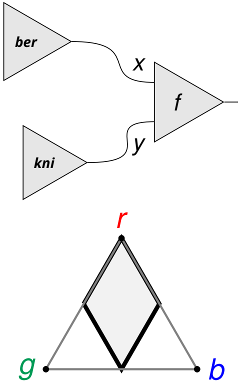
Outline

The Problem: Convex powerset doesn't work

bernoulli interpreted as $\left\{ \begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix} \right\}$

knight interpreted as $\left\{ \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right\}$

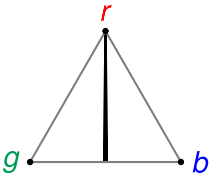


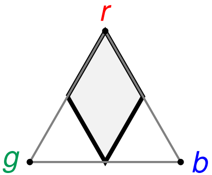


$$
\begin{aligned}
f(x,y) = \text{if } x \text{ then } &(\text{ if } y \text{ then } r \text{ else } g\,) \\
\text{else } &(\text{ if } y \text{ then } r \text{ else } b\,)
\end{aligned}
$$

The Problem

Theorem: Any semantic model that satisfies our desiderata cannot distinguish the following convex sets of distributions.



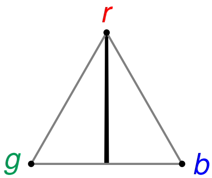


$z \leftarrow$ ***bernoulli*** ;
if z then (if $x \leftarrow$ ***knight*** ; x then *r* else *g*)
else (if $x \leftarrow$ ***knight*** ; x then *r* else *b*)

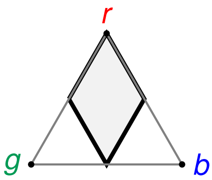
Outline

The Solution: Named Knightian choices

$z \leftarrow$ ***bernoulli*** ;
if z then (if $x \leftarrow$ ***knight***(a_1) ; x then return r else return g)
else (if $x \leftarrow$ ***knight***(a_1) ; x then return r else return b)



$z \leftarrow$ ***bernoulli*** ;
if z then (if $x \leftarrow$ ***knight***(a_1) ; x then return r else return g)
else (if $x \leftarrow$ ***knight***(a_2) ; x then return r else return b)



The Solution: Named Knightian choices

- Reader monad transformer of finite distributions monad

$$T_{2^A}(X) = [2^A \Rightarrow D(X)]$$

The Solution: Named Knightian choices

- Reader monad transformer of finite distributions monad

$$T_{2^A}(X) = [2^A \Rightarrow D(X)]$$

- **Bernoulli** choices given by *distributions*

The Solution: Named Knightian choices

- Reader monad transformer of finite distributions monad

$$T_{2^A}(X) = [2^A \Rightarrow D(X)]$$

- **Bernoulli** choices given by *distributions*
- **Knightian** choices given by *reading*

The Solution: Named Knightian choices

- Reader monad transformer of finite distributions monad

$$T_{2^A}(X) = [2^A \Rightarrow D(X)]$$

- **Bernoulli** choices given by *distributions*
- **Knightian** choices given by *reading*
- We generalise the Knightian choices 2^A to arbitrary sets B

$$T_B(X) = [B \Rightarrow D(X)]$$

The Solution: Named Knightian choices

- Reader monad transformer of finite distributions monad

$$T_{2^A}(X) = [2^A \Rightarrow D(X)]$$

- **Bernoulli** choices given by *distributions*
- **Knightian** choices given by *reading*
- We generalise the Knightian choices 2^A to arbitrary sets B

$$T_B(X) = [B \Rightarrow D(X)]$$

- Convex powerset recovered by pushing forward all possible distributions on B

$$[\![t]\!]_B = \{p \gg\!=_D t \mid p \in D(B)\} \in \mathrm{CP}(X).$$

A graded monad

Definition

T is the $\mathbf{FinStoch}_{\mathrm{Surj}}$-graded version of:

$$T_A(X) = [A \Rightarrow D(X)]$$

T supports finite probability and finite non-determinism:

- ***bernoulli*** $\in T_1(2)$ given by $\begin{pmatrix} 0.5 \\ 0.5 \end{pmatrix}$
- ***knight*** $\in T_2(2)$ given by $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$

Function and monoidal composition use independent non-deterministic branches:

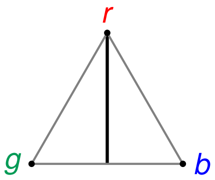
$$\circ_{A,B} : [X \Rightarrow T_A(Y)] \times [Y \Rightarrow T_B(Z)] \to [X \Rightarrow T_{A \times B}(Z)]$$

$$\otimes_{A,B} : [X \Rightarrow T_A(Y)] \times [X' \Rightarrow T_B(Y')] \to [X \times X' \Rightarrow T_{A \times B}(Y \times Y')]$$

Outline

Threorem 1: Improved Bounds

$$\begin{array}{c} \color{red}{r} : \\ \color{green}{g} : \\ \color{blue}{b} : \end{array} \begin{pmatrix} 1 & 0 \\ 0 & 0.5 \\ 0 & 0.5 \end{pmatrix} \longrightarrow$$



- There is a mapping $[\![-]\!] : T_A(X) \to \mathrm{CP}(X)$ by taking the convex hull of the image
- This induces an 'op-lax functor':

$$R : [X \Rightarrow T_A(Y)] \to [X \Rightarrow \mathrm{CP}(Y)]$$

- That is, composition in our framework gives tighter bounds on the *Knightian uncertainty* than composition with CP

$$R(g \circ f) \subseteq R(g) \circ R(f)$$

Threorem 2: Maximality

- The mapping into convex powersets is not injective
- There is ”more stuff” in our theory than in convex powersets
- So can we quotient our theory any further?
- Not without:
  - Breaking connection with convex subsets (and imprecise probability); or
  - Compromising the compositional structure

Conclusion

This work:

- Using graded perspective and naming Knightian choices we give a fully compositional account of Bernoulli and Knightian uncertainty together
- **Theorem 1:** This gives a refined bound on uncertainty
- **Theorem 2:** It is maximal among compositional accounts

Future work:

- Iteration and infinite dimensional structures
- Function spaces via quasi-Borel spaces
- Implementation and approximation of bounds

References I

Amarante, Massimiliano and Fabio Maccheroni (2007). "When an Event Makes a Difference". In: *Uncertainty and Risk: Mental, Formal, Experimental Representations*. Ed. by Mohammed Abdellaoui et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 27–32. ISBN: 978-3-540-48935-1. DOI: 10.1007/978-3-540-48935-1_3.

Aslett, Louis, Frank Coolen, and Jasper Bock (Jan. 2022). *Uncertainty in Engineering Introduction to Methods and Applications: Introduction to Methods and Applications*. ISBN: 978-3-030-83639-9. DOI: 10.1007/978-3-030-83640-5.

Battigalli, Pierpaolo, Simone Cerreia-Vioglio, et al. (Feb. 2015). "Self-Confirming Equilibrium and Model Uncertainty". In: *American Economic Review* 105.2, pp. 646–77. DOI: 10.1257/aer.20130930.

References II

- Battigalli, Pierpaolo, Roberto Corrao, and Martin Dufwenberg (2019). "Incorporating belief-dependent motivation in games". In: *Journal of Economic Behavior & Organization* 167, pp. 185–218. ISSN: 0167-2681. DOI: 10.1016/j.jebo.2019.04.009.
- Caprio, Michele et al. (2024). *Credal Bayesian Deep Learning*. arXiv: 2302.09656 [cs.LG].
- Cerreia-Vioglio, Simone, Roberto Corrao, and Giacomo Lanzani (July 2023). "Dynamic Opinion Aggregation: Long-run Stability and Disagreement". In: *The Review of Economic Studies* 91.3, pp. 1406–1447. ISSN: 0034-6527. DOI: 10.1093/restud/rdad072.
- Dahlqvist, Fredrik, Louis Parlant, and Alexandra Silva (2018). "Layer by layer: composing monads". In: *Proc. ICTAC 2018*.

References III

Dalrymple, David ‘davidad’ (Jan. 2024). *Safeguarded AI: constructing safety by design*. ARIA Programme Thesis. https://www.aria.org.uk/wp-content/uploads/2024/01/ARIA-Safeguarded-AI-Programme-Thesis-V1.pdf.

Dalrymple, David ‘davidad’ et al. (2024). *Towards Guaranteed Safe AI: A Framework for Ensuring Robust and Reliable AI Systems*. arXiv: 2405.06624 [cs.AI].

Dash, S. et al. (2023). “Affine monads and lazy structures for Bayesian programming”. In: *Proc. POPL 2023*.

Dash, Swaraj (2024). “A Monadic Theory of Point Processes”. PhD thesis. University of Oxford.

Dash, Swaraj and Sam Staton (2020). “A monad for probabilistic point processes”. In: *Proc. ACT 2020*.

— (2021). “Monads for Measurable Queries in Probabilistic Databases”. In: *Proc. MFPS 2021*.

References IV

- Ferson, Scott et al. (2003). *Constructing Probability Boxes and Dempster-Shafer Structures*. Tech. rep. Technical Report SAND2002-4015. Sandia National Laboratiories.
- Huber, P J (1981). *Robust statistics*. Wiley.
- Jacobs, Bart (2021). “From multisets over distributions to distributions over multisets”. In: *Proc. LICS 2021*.
- Javanmardi, Alireza, David Stutz, and Eyke Hüllermeier (2024). *Conformalized Credal Set Predictors*. arXiv: `2402.10723 [stat.ML]`.
- Keimel, Klaus and Gordon D. Plotkin (2017). “Mixed powerdomains for probability and nondeterminism”. In: *Log. Methods Comput. Sci.* 13.
- Kozen, Dexter and Alexandra Silva (2023). *Multisets and Distributions*. arxiv:2301.10812.

References V

- Krpelík, Danda and Tathagata Basu (Jan. 2021). "Introduction to Imprecise Probabilities". In: chap. 13, pp. 35–79. ISBN: 978-3-030-60165-2. DOI: 10.1007/978-3-030-60166-9_2.
- Moggi, E. (1989). "Computational Lambda-Calculus and Monads". In: *Proceedings of the Fourth Annual Symposium on Logic in Computer Science*. Pacific Grove, California, USA: IEEE Press, pp. 14–23. ISBN: 0818619546.
- — (1991). "Notions of computation and monads". eng. In: *Information and computation* 93.1. Place: San Diego, CA Publisher: Elsevier Inc, pp. 55–92. ISSN: 0890-5401. DOI: 10.1016/0890-5401(91)90052-4.
- Oren, Yaniv et al. (2024). *Epistemic Monte Carlo Tree Search*. arXiv: 2210.13455 [cs.LG].

- Pavlidis, E.G. et al. (Aug. 2011). "Nonlinear dynamics in economics and finance and unit root testing". In: *The European Journal of Finance* 2011. DOI: 10.1080/1351847X.2011.607006.
- Stutz, David et al. (2023). *Conformal prediction under ambiguous ground truth*. arXiv: 2307.09302 [cs.LG].
- Varacca, D and G Winskel (2006). "Distributing probability over non-determinism". In: *Math. Structures Comput. Sci.*, pp. 87–113.
- Vasile, Massimiliano (July 2021). *Optimization Under Uncertainty with Applications to Aerospace Engineering*. ISBN: 978-3-030-60165-2.
- Vicig, Paolo (Sept. 2008). "Financial Risk Measurement with Imprecise Probabilities". In: *Int. J. Approx. Reasoning* 49.1, pp. 159–174. ISSN: 0888-613X. DOI: 10.1016/j.ijar.2007.06.009.

References VII

Walley, Peter (1991). *Statistical Reasoning with Imprecise Probabilities*. Chapman and Hall.

Zanger, Moritz A., Wendelin Böhmer, and Matthijs T. J. Spaan (2023). *Diverse Projection Ensembles for Distributional Reinforcement Learning*. arXiv: 2306.07124 [cs.LG].