Toposes for probabilistic programming

Ulrik Buchholtz

CIRM 2686, Feb. 3, 2022

Toposes for probabilistic programming

Where I'm coming from:

- Work on homotopy type theory and synthetic homotopy theory (see you at CIRM 2689!),
- I like to explore applications to other kinds of synthetic mathematics, and synthetic measure/probability theory seems like an intriguing target.

What's in this talk:

- Brief note on synthetic mathematics in general (and dependent type theory) and how it might connect with (the semantics of) probabilistic programs.
- A survey of some approaches in the literature and some properties of the corresponding toposes.
- Some questions, both technical and methodological.

What's not in this talk:

- Any original work.
- Any new answers/theorems.

But let me know if you have some answers or you'd like to explore the suggested connections!
Or if I missed any important developments!

Dependent type theory in brief

Martin-Löf type theory is:

A typed programming language

Key ingredients: Dependent types $(a : A) \times B(a)$ and $(a : A) \to B(a)$ and universes $\mathcal{U}$.

A foundation for mathematics

Key idea: proposition as types.

Homotopy type theory (HoTT) makes it much better for both!

Univalence: A canonical map $(A =_{\mathcal{U}} B) \to (A \simeq B)$ is an equivalence. Allows us to think of the universe as really containing (small) types, not codes thereof.

Systems: Agda, Coq, Idris, Lean (*MathLib*), ...

Synthetic mathematics

Basic idea: The systems have many models, and some may make reasoning and programming in certain domains much easier (DSLs):

- synthetic homotopy theory (HoTT) – has models in any $(\infty, 1)$-topos (Shulman 2019);
- synthetic euclidean topology (Shulman 2018);
- synthetic differential geometry (SDG); basic model: smooth $(\infty, 1)$-topos, well-adapted model: Dubuc topos using germ-determined C^∞-rings, where we have the object D of dual numbers, so that the tangent bundle of X is X^D.
- synthetic differential topology (SDT);
- synthetic domain theory;
- synthetic algebraic geometry;
- synthetic measure theory (A. Kock 2012);
- ...

In each case, we can extend the basic language with new objects, constructions and axioms suitable for the domain.

A recipe for toposes

Here's a way to get many toposes:

1. Start with a category that captures the first-order structure of interest (e.g., Meas, Prob, CHaus, $\omega\mathrm{Cpo}$, Top, Smooth, ...).
2. Cut down to a small subcategory, $\mathbb{C}$ (e.g., Sbs, Sps, $\mathrm{CHaus}_{<\kappa}$, retracts of $\mathrm{Pow}(\omega)$, $\mathrm{CartSp}_{\mathrm{top}}$, $\mathrm{CartSp}_{\mathrm{smooth}}$, ...).
3. Put a Grothendieck topology J on $\mathbb{C}$ telling you how objects are glued together.
4. Consider the Grothendieck topos of sheaves of sets (or better: ∞-groupoids) on $\mathbb{C}$: contravariant functors $X : \mathbb{C}^{\mathrm{op}} \to \mathsf{Set}$ satisfying a gluing condition (as in the def. of quasi-Borel spaces this morning).

Internal language

Each topos models both Martin-Löf type theory and higher order logic. Each ∞-topos models HoTT (MLTT plus univalent universes plus higher quotient constructions).

Smooth sets and automatic differentiation

Huot, Staton, and Vákár 2020 considered diffeological spaces for the purposes of verifying automatic differentiation, these embed in the cohesive $(\infty, 1)$-topos of smooth ∞-groupoids, $\mathrm{Sh}(\mathrm{CartSp}_{\mathrm{smooth}})$. This also embeds Δ-generated topological spaces.

Maybe it would be worth-while to connect automatic differentiation (AD) to other models, such as the well-adapted models of SDG?

Synthetic measure theory

Many settings in which we have something like a monad of distributions (Jacobs 2018):

- Sets, discrete distributions.
- Measurable spaces, Giry monad.
- Domains, probabilistic powerdomain.
- Compact Hausdorff spaces, Radon monad.
- 1-Bounded metric spaces, Kantorovich monad.

Question

Can we capture the common logic of these in the internal language of a topos built from these?

Will probably have to consider both modal and linear aspects of type theory!

Example: Quasi-Borel spaces

Recall from Heunen, Kammar, Staton, and Yang 2017:

Definition

A quasi-Borel space is a pair (X, M_X) of a set X and a set of functions $M_X \subseteq (\mathbb{R} \to X)$ containing constant functions, closed under precomposition with Lebesgue measurable functions $\mathbb{R} \to \mathbb{R}$, and closed under disjoint Borel gluing.

Theorem

The category Qbs *of quasi-Borel spaces is equivalent to the concrete sheaves in the topos* $\mathrm{Sh}(\mathrm{Sbs})$*.*

Question

Hence Qbs *is a quasi-topos. Can we extend the monad of measures to all of* $\mathrm{Sh}_1(\mathrm{Sbs})$*? Or even* $\mathrm{Sh}_\infty(\mathrm{Sbs})$

Probability sheaves

From A. Simpson 2017:

Definition

A standard probability space is a pair (Ω, μ_Ω) of a standard Borel space Ω with a probability measure μ_Ω. These assemble into a category Sps whose morphisms are measure-preserving measurable functions.

Theorem (Simpson)

$\mathrm{Sh}(\mathrm{Sps})$ *is a boolean topos satisfying dependent choice, and it has a strong monad of probability measures.*

Note the similarity with the Solovay model of set theory, where all subsets of $\mathbb{R}$ are measurable.

Quantum probability

Long line of work by Butterfield, Döring, Caspers, Heunen, Isham, Landsman, Spitters, Wolters (et al.) on interpreting quantum theory as 'classical physics' internal to suitable topos.
We start with either a C^*-algebra or a von Neumann algebra A and consider the poset of commutative subalgebras.

In this case, it's not so clear what is a good 'gros topos'!

Synthetic physics

And Hilbert's sixth problem. See work by Döring and Isham 2010 and Schreiber 2017 and others.

Questions:

Some overlap with proposal of Bidlingmaier, Faissole, and Spitters 2019.

- How do we move across the analytic/synthetic spectrum while getting the best of both worlds? (Informally? And formally in a proof assistant; via formalized model constructions?)
- Does probabilistic semantics benefit from recent work on condensed mathematics or the pyknotic topos, Clausen and Scholze 2020; Barwick and Haine 2019
- Are higher types useful in probabilistic reasoning? (Distributions on ∞-groupoids?)
- Can this help make sense of symmetry considerations in the choice of priors?
- Which models can be developed constructively and thus indirectly provide a semantics of probabilistic programs?
- Is there a connection to experimental mathematics?
- ...

References: I

- Barwick, Clark and Peter Haine (2019). *Pyknotic objects, I. Basic notions*. arXiv: 1904.09966 [math.AG].
- Bidlingmaier, Martin E., Florian Faissole, and Bas Spitters (2019). "Synthetic topology in Homotopy Type Theory for probabilistic programming". In: *CoRR* abs/1912.07339. arXiv: 1912.07339. URL: http://arxiv.org/abs/1912.07339.
- Clausen, Dustin and Peter Scholze (2020). *Masterclass in condensed mathematics*. URL: https://www.math.ku.dk/english/calendar/events/condensed-mathematics/.
- Coquand, Thierry and Bas Spitters (2009). "Integrals and valuations". In: *J. Log. Anal.* 1, Paper 3, 22. DOI: 10.4115/jla.2009.1.3.
- Döring, A. and C. Isham (2010). ""What is a Thing?": Topos Theory in the Foundations of Physics". In: *Lecture Notes in Physics*, 753–937. ISSN: 1616-6361. DOI: 10.1007/978-3-642-12821-9_13. URL: http://dx.doi.org/10.1007/978-3-642-12821-9_13.
- Forré, Patrick (2021). *Quasi-Measurable Spaces*. arXiv: 2109.11631 [math.PR].
- Fritz, Tobias and Paolo Perrone (2019). "A probability monad as the colimit of spaces of finite samples". In: *Theory Appl. Categ.* 34, Paper No. 7, 170–220. URL: http://www.tac.mta.ca/tac/volumes/34/7/34-07.pdf.

References: II

- Furber, Robert, Radu Mardare, Prakash Panangaden, and Dana S. Scott (2021). "Interpreting Lambda Calculus in Domain-Valued Random Variables". In: *CoRR* abs/2112.06339. arXiv: 2112.06339.
- Heunen, Chris, Ohad Kammar, Sam Staton, and Hongseok Yang (2017). "A convenient category for higher-order probability theory". In: *2017 32nd Annual ACM/IEEE Symposium on Logic in Computer Science (LICS)*. IEEE, p. 12.
- Huot, Mathieu, Sam Staton, and Matthijs Vákár (2020). "Correctness of automatic differentiation via diffeologies and categorical gluing". In: *Foundations of software science and computation structures*. Vol. 12077. Lecture Notes in Comput. Sci. Springer, Cham, pp. 319–338. DOI: 10.1007/978-3-030-45231-5_17.
- Huot, Mathieu, Sam Staton, and Matthijs Vákár (2020). "Correctness of Automatic Differentiation via Diffeologies and Categorical Gluing". In: *CoRR* abs/2001.02209. arXiv: 2001.02209.
- Jacobs, Bart (2018). "From probability monads to commutative effectuses". In: *J. Log. Algebr. Methods Program.* 94, pp. 200–237. ISSN: 2352-2208. DOI: 10.1016/j.jlamp.2016.11.006.

References: III

- Kock, Anders (2012). "Commutative monads as a theory of distributions". In: *Theory Appl. Categ.* 26, No. 4, 97–131.
- — (2021). "New methods for old spaces: synthetic differential geometry". In: *New spaces in mathematics*. Cambridge Univ. Press, Cambridge, pp. 83–116. arXiv: 1610.00286.
- Schreiber, Urs (2017). *Differential cohomology in a cohesive infinity-topos*. URL: https://ncatlab.org/schreiber/files/dcct170811.pdf.
- Shulman, Michael (2018). "Brouwer's fixed-point theorem in real-cohesive homotopy type theory". In: *Math. Structures Comput. Sci.* 28.6, pp. 856–941. DOI: 10.1017/S0960129517000147.
- — (2019). *All $(\infty, 1)$-toposes have strict univalent universes*. arXiv: 1904.07004 [math.AT].
- Simpson, Alex (2017). "Probability sheaves and the Giry monad". In: *7th Conference on Algebra and Coalgebra in Computer Science*. Vol. 72. LIPIcs. Leibniz Int. Proc. Inform. Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, Art. No. 1, 6.
- Tjur, Tue (1980). *Probability based on Radon measures*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester, pp. xi+232. ISBN: 0-471-27824-6.

References: IV

Ścibior, Adam, Ohad Kammar, Matthijs Vákár, Sam Staton, Hongseok Yang, Yufei Cai, Klaus Ostermann, Sean K. Moss, Chris Heunen, and Zoubin Ghahramani (2017). "Denotational Validation of Higher-Order Bayesian Inference". In: *Proc. ACM Program. Lang.* 2.POPL. DOI: 10.1145/3158148.