Semantics for probabilistic programming

Florence Clerc, Vincent Danos, Fredrik Dahlqvist, Ilias Garnier

November 1, 2016

The design of operational and denotational models for programming languages has historically been a rather post-hoc endeavour, justified by practical concerns such as the validation of program transformations. If one rather views algorithms as "the idiom of modern science"[1], then formal semantics is the precondition for models *as programs* to be sound objects of mathematical scrutiny. Dually, the structures naturally arising in the mathematical universe of the domain of interest should inform the design of programming languages. In this extended abstract, we take a look from this perspective to *probabilistic programming* and see how our work on on a category-theoretical approach to measure theory and statistics is relevant to this blooming field. We focus on two recent contributions, namely *pointless Bayesian inversion* and *natural transformations in probability and statistics*.

1 Pointless Bayesian inversion

Following Kozen [8] and Moggi [10], terms in probabilistic programs can be given a denotational semantics as arrows in the Kleisli category of the Giry monad $(\mathsf{G}, \mu, \delta)$ [7], i.e. *kernels*. Mathematical models of probabilistic programming have all relied on this pointful, *kernel-centric* view–including for the key operation in Bayesian learning, namely *Bayesian inversion*. We argue that a *pointless*, operator-based approach to Bayesian inversion is both more general, simpler and offers a more structured view of Bayesian machine learning.

Culbertson & Sturtz [2] present Bayesian inversion as a mapping associating each pair consisting of a *prior* p and a *likelihood* f with $(p, f) \in \mathsf{G}(H) \times (H \to \mathsf{G}(D))$ to a pair $(q, f^\dagger) \in \mathsf{G}(D) \times (D \to \mathsf{G}(H))$ consisting of the *posterior* $f^\dagger$ and the *marginal likelihood* q (which is just the pushforward of p through f). The kernel $f^\dagger$ is characterised as the q-almost surely unique one satisfying an infinitary generalisation of Bayes' law. We identify the following problems with this approach:

- it relies on the disintegration theorem, hence on rather strong topological assumptions on the underlying spaces;
- it is *unstructured*: the map from (p, f) to $(q, f^\dagger)$ is not even a function, as $f^\dagger$ is only defined q-almost surely.

[1] http://www.cs.princeton.edu/~chazelle/pubs/algorithm.html

In [5], we establish that this operation can be recast as the familiar operation of *adjunction* of operators. This is formalised in a categeory of ω-complete cones following earlier work [1, 11]. We construct a functorial representation from a category of typed kernels to a category of linear operators (viewing these kernels as generalised predicate transformers, as in [9]). We show that pointful inversion coincides with adjunction whenever it is defined (e.g. on standard Borel spaces). This move presents the following advantages:

1. no topological assumptions are required;
2. adjunction is *structured*: it corresponds to an equivalence of categories between so-called *abstract Markov kernels* and *Markov operators* [1].

Besides being a ground for denotational semantics, we foresee that this pointless setting will be amenable to the application of the approximation techniques for Markov processes developed in [1] to Bayesian learning.

2 Natural transformations in probability and statistics

Modelling effects functorially, as in Moggi's approach to computational lambda calculi, makes the study of natural transformations between those functors a useful source of potential programming primitives. In [6, 3] we constructed a general theorem (called the *Machine*) allowing to reduce the proof of existence of natural transformations between a wide family of endofunctors of the category of Polish spaces to their components at finite spaces. These transformations include the monadic data of the Giry functor, the iid distribution, the normalisation of a nonzero measure to a probability, the Poisson point process, the Dirichlet process and last but not least, the de Finetti representation theorem. Further, one of our results in [3] called *rigidity* gives a sufficient condition for there to be at most *one* natural transformation between two well-chosen functors. We believe the *Machine* is relevant to probabilistic programming, as it eases the exploration of potential primitives to be added in existing or future languages. On this matter, we derived in [4] some results which establish the basis for a compositional language of natural transformations seen as robustly parameterised probabilistic models. Further, we expect that our techniques can address questions of computability of such natural stochastic processes by reducing them to the finite-dimensional case.

3 Conclusion

Ongoing work includes importing more results of measure theory in our framework, so as to bridge the Machine with the cone-theoretic developments of Sec. 1. Our long term plan is to obtain a cohesive category-theoretical view of measure theory and statistics, where both semantics and a sound theory of approximation for probabilistic models coexist.

References

[1] Philippe Chaput, Vincent Danos, Prakash Panangaden, and Gordon Plotkin. Approximating Markov Processes by averaging. *Journal of the ACM*, 61(1), January 2014. 45 pages.

[2] Jared Culbertson and Kirk Sturtz. A categorical foundation for Bayesian probability. *Applied Categorical Structures*, pages 1–16, 2012.

[3] F. Dahlqvist, V. Danos, and I. Garnier. Giry and the machine. Feb 2016. To appear in the proceedings of MFPS XXXII.

[4] Fredrik Dahlqvist, Vincent Danos, and Ilias Garnier. Robustly Parameterised Higher-Order Probabilistic Models. In Josée Desharnais and Radha Jagadeesan, editors, *27th International Conference on Concurrency Theory (CONCUR 2016)*, volume 59 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 23:1–23:15, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[5] Fredrik Dahlqvist, Vincent Danos, Ilias Garnier, and Ohad Kammar. Bayesian Inversion by Omega-Complete Cone Duality (Invited Paper). In Josée Desharnais and Radha Jagadeesan, editors, *27th International Conference on Concurrency Theory (CONCUR 2016)*, volume 59 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 1:1–1:15, Dagstuhl, Germany, 2016. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.

[6] V. Danos and I. Garnier. Dirichlet is natural. *Electronic Notes in Theoretical Computer Science*, 319:137 – 164, 2015. MFPS XXXI.

[7] M. Giry. A categorical approach to probability theory. In *Categorical Aspects of Topology and Analysis*, number 915 in Lecture Notes In Math., pages 68–85. Springer-Verlag, 1981.

[8] Dexter Kozen. Semantics of probabilistic programs. *J. Comput. Syst. Sci.*, 22:328–350, 1981.

[9] Dexter Kozen. A probabilistic pdl. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, STOC '83, pages 291–297, New York, NY, USA, 1983. ACM.

[10] Eugenio Moggi. Notions of computation and monads. *Information and Computation*, 93(1):55 – 92, 1991.

[11] Peter Selinger. Towards a semantics for higher-order quantum computation. In *Proceedings of the 2nd International Workshop on Quantum Programming Languages, TUCS General Publication*, volume 33, pages 127–143, 2004.