A Categorical Foundation for Bayesian Probability (Culbertson and Sturtz, [2014])

Overview.

The authors of this paper develop a category theoretic framework for reasoning about decision rules over probability measures and the relationship between prior and posterior probability distributions.

The authors primarily work in two categories, each of which have countably generated measurable spaces as objects (tuples of the space X and the σ-algebra Σ_x):

(1) M_{cg}, where the morphisms between (X, Σ_X) and (Y, Σ_Y) are measurable functions $f : X \to Y$.
(2) $\mathcal{P}$, where the morphisms between (X, Σ_X) and (Y, Σ_Y) are functions $f : X \times \Sigma_Y \to [0, 1]$, such that f varies measurably in X and is a family of perfect probability measures in Σ_Y.

The authors describe the following two functors between these categories:

(1) $\delta : M_{cg} \to \mathcal{P}$, which acts as the identity on objects and maps the measurable function f to the deterministic $\mathcal{P}$-arrow:

$$\delta_f(x, B) = 1 \; if \; f(x) \in B \; else \; 0$$

(2) $\mathcal{P} : \mathcal{P} \to M_{cg}$, which maps the measurable space X to the set of perfect probability measures on X, $\mathcal{P}X$. $\mathcal{P}$ maps the $\mathcal{P}$-arrow $f : X \to Y$ to a measurable function $\mathcal{P}f : \mathcal{P}X \to \mathcal{P}Y$, where we define $\mathcal{P}f$ as the composition of $P : 1 \to X$ and f in $\mathcal{P}$.

Then the adjunction $\delta \dashv \mathcal{P}$ defines a monad T (the Giry monad) over M_{cg}. The unit $\nu : X \to TX$ of this monad maps a measurable space to the set of probability measures over this space, and a T-algebra $\alpha : TX \to X$ serves as a "decision rule" over this monad that collapses a probability measure on X into a value in X. This allows us to reason about a category of decision rules, where decisions are defined by monad algebras.

The authors also use this framework to define the basic constructs of Bayesian Inference. They begin by defining the hypothesis H and data D as objects in $\mathcal{P}$. Then they define the prior probability as a morphism $P_H : 1 \to H$ (where 1 is the terminal object in $\mathcal{P}$, which is isomorphic to the one-element set). They also define the inference map $I : D \to H$, sampling distribution $S : H \to D$, and measurement map $\mu : 1 \to D$. They show how we can use these tools to compute posterior probabilities with the composition of measurement and inference $I \circ \mu : 1 \to H$.

Comments.

This paper is relatively clear to follow, and the authors' general construction of inference maps from prior probabilities and a sampling distribution is definitely exciting. However, given the authors' claim that their paper lays a "Categorical Foundation for Bayesian Probability," this work is quite light on categorical structure. For example, the authors' description of how we can update prior probabilities in an iterative manner is not particularly rigorous and does not naturally fit into their framework. Their construction for Bayesian inference is simply a set of morphisms in $\mathcal{P}$, and they do not define any sort of categorical construct to represent the appearance of new data, such as the evolution of time or different "states" of the inference system. That said, there are several ways that the authors could formalize this aspect of their construction. For example, they might be able to define endofunctors to represent state transitions or the evolution of time.

Moreover, the authors' exploration of the category of decision rules seems particularly promising, even though they only begin to develop it in this paper. By representing decisions as monad algebras, they permit a broad perspective on decision algorithms. Machine Learning lies at the border of decision theory and probability, and this work may help move us towards a unified theory of intelligent systems.

REFERENCES

[1] Culbertson, J. and Sturtz, K. (2014). A categorical foundation for bayesian probability. *Applied Categorical Structures*, 22(4):647–662.