Eilenberg-Moore categories of Markov monads

Rob Cornish

Department of Statistics, University of Oxford

Background: Markov categories

Markov categories [CJ19; Fri20] have recently attracted significant interest as a framework for reasoning about probability without needing to contend with low-level measure-theoretic details. Recall that a Markov category is a symmetric monoidal category C such that each object $\mathcal{X}$ in C is equipped with a distinguished commutative comonoid structure with comultiplication $\mathsf{copy}_{\mathcal{X}} : \mathcal{X} \to \mathcal{X} \otimes \mathcal{X}$ and counit $\mathsf{del}_{\mathcal{X}} : \mathcal{X} \to \mathcal{I}$. Markov categories permit purely synthetic reasoning about standard probabilistic concepts, including *conditionals* and *randomness pushback* (see [Fri20]), which we make use of in what follows.

Overview of main results

Theoretical contribution Suppose T is a monad on a Markov category C. We begin with the following theoretical question: does the Eilenberg-Moore category C^T inherit the structure of a Markov category in any way? We show that, for a certain naturally arising class of monads that we term *Markov monads*, the full subcategory $\mathsf{C}^T_{\mathrm{det}} \subseteq \mathsf{C}^T$ consisting of *deterministic* algebras is canonically a Markov category. That is, the objects of $\mathsf{C}^T_{\mathrm{det}}$ are algebras $(\mathcal{X}, \alpha)$ such that α is deterministic. Beyond its utility of our application below, this result is also interesting since determinism is often added as an assumption in various places in the literature to ensure things work as expected [MP23; Cor25], and this gives another example of this phenomenon.

Application Our theoretical results allow us to provide a streamlined and conceptual proof of a key result from the machine learning literature by Bloem-Reddy & Teh [BT20]. In [BT20], the authors consider how to obtain neural network architectures that achieve a given group invariance or equivariance property, which is of interest in many machine learning applications [Bro+17; Bro+21]. Their key result provides sufficient conditions under which the output of a (possibly stochastic) predictive model can be written in a *noise outsourced* form in terms of an equivariant deterministic function and some independent randomness.

The proof given in [BT20] is quite technical, involving many steps and low-level measure theoretic details. In contrast, by using Markov categories, we can give a much more streamlined proof under weaker assumptions. At a high level, we consider the Markov category $\mathsf{C}^{\mathcal{G}} := \mathsf{C}^T_{\mathrm{det}}$ induced by the writer monad $T := \mathcal{G} \otimes -$ of a group $\mathcal{G}$ in $\mathsf{C}_{\mathrm{det}}$. Since $\mathsf{C}^{\mathcal{G}}$ is Markov, we can ask whether it has *conditionals* and admits *randomness pushback*. Whenever it does, we obtain straightforwardly an analogue of the result of [BT20]. The original result is then recovered as the special case with $\mathsf{C} := \mathsf{BorelStoch}$.

Markov functors and Markov monads

Motivation Suppose T is a symmetric colax monoidal monad on a symmetric monoidal category C, with structure maps $\Delta_{\mathcal{X},\mathcal{Y}} : T(\mathcal{X} \otimes \mathcal{Y}) \to \mathcal{X} \otimes \mathcal{Y}$. It is a folkloric result that the Eilenberg-Moore category C^T is then itself canonically symmetric monoidal, where its monoidal product $\otimes_T$ is given by the tensor product of algebras:

$$(\mathcal{X}, \alpha) \otimes_T (\mathcal{Y}, \beta) := (\mathcal{X} \otimes \mathcal{Y}, (\alpha \otimes \beta) \circ \Delta_{\mathcal{X},\mathcal{Y}}). \tag{1}$$

If C is in particular Markov, then to ensure that C^T is Markov, roughly speaking, it is sufficient to ensure that T also respects the copy maps of C in some way.

Markov functors To develop this further, recall that for all symmetric categories C and D, a symmetric colax monoidal functor $F : \mathsf{C} \to \mathsf{D}$ transports every commutative comonoid $(c : \mathcal{X} \to \mathcal{X} \otimes \mathcal{X}, d : \mathcal{X} \to \mathcal{I})$ in C to a commutative comonoid with comultiplication and counit given by

$$F\mathcal{X} \xrightarrow{Fc} F(\mathcal{X} \otimes \mathcal{X}) \xrightarrow{\Delta_{\mathcal{X},\mathcal{X}}} F\mathcal{X} \otimes F\mathcal{X} \qquad\qquad F\mathcal{X} \xrightarrow{Fd} F\mathcal{I} \xrightarrow{\delta} \mathcal{I}, \tag{2}$$

where Δ and δ constitute the colax monoidal structure of F. If C and D are moreover Markov categories, we will then say that F is *Markov* if this procedure also transports the copy maps (i.e. the distinguished commutative comonoids) from C to D. In other words, the following diagram always commutes:

$$\begin{tikzcd} & F\mathcal{X} \arrow[dl, "F(\mathsf{copy}_{\mathcal{X}})"'] \arrow[dr, "\mathsf{copy}_{F\mathcal{X}}"] & \\ F(\mathcal{X} \otimes \mathcal{X}) \arrow[rr, "\Delta_{\mathcal{X},\mathcal{X}}"'] & & F\mathcal{X} \otimes F\mathcal{X} \end{tikzcd} \tag{3}$$

Note that this is more general than the notion of a Markov functor given in Definition 10.14 of [Fri20], which takes F to be strong monoidal (so that $\Delta_{\mathcal{X},\mathcal{X}}$ is an isomorphism). We then have the following result (extending Lemma 1.15 of [Fri20]):

Theorem 1. *Every Markov functor $F : \mathsf{C} \to \mathsf{D}$ restricts to a colax monoidal functor $F_{\mathrm{det}} : \mathsf{C}_{\mathrm{det}} \to \mathsf{D}_{\mathrm{det}}$.*

In other words, F maps deterministic morphisms to deterministic morphisms, and its colax monoidal structure maps $\Delta_{\mathcal{X},\mathcal{Y}}$ are all deterministic. Since $\mathsf{C}_{\mathrm{det}}$ and $\mathsf{D}_{\mathrm{det}}$ are both cartesian, each $\Delta_{\mathcal{X},\mathcal{Y}}$ on F is then in fact uniquely determined (as is standard to show).

We could not prove an exact converse to Theorem 1, but could show the following instead:

Proposition 2. *Suppose C and D are Markov categories and $F : \mathsf{C} \to \mathsf{D}$ is a functor that restricts to a functor $F_{\mathrm{det}} : \mathsf{C}_{\mathrm{det}} \to \mathsf{D}_{\mathrm{det}}$, and let Δ be the unique colax monoidal structure for F_{det}. Then this is also a colax monoidal structure for F,* provided *the naturality condition*

$$\begin{tikzcd} F(\mathcal{X} \otimes \mathcal{Y}) \arrow[r, "F(f \otimes g)"] \arrow[d, "\Delta_{\mathcal{X},\mathcal{Y}}"] & F(\mathcal{X}' \otimes \mathcal{Y}') \arrow[d, "\Delta_{\mathcal{X}',\mathcal{Y}'}"] \\ F\mathcal{X} \otimes F\mathcal{Y} \arrow[r, "Ff \otimes Fg"] & F\mathcal{X}' \otimes F\mathcal{Y}' \end{tikzcd} \tag{4}$$

is satisfied for all f and g in C (as opposed to just $\mathsf{C}_{\mathrm{det}}$).

Markov monads With this definition, we can then define a *Markov monad* to be a Markov functor $T : \mathsf{C} \to \mathsf{C}$ together with colax monoidal natural transformations η and μ for its unit and multiplication. (It is possible to show that the colax monoidal condition holds if and only if each component of η and μ is deterministic.) We then have the following result (with $\mathsf{C}^T_{\mathrm{det}}$ as defined earlier):

Theorem 3. *If T is a Markov monad on C, then $\mathsf{C}^T_{\mathrm{det}}$ is canonically a Markov category with symmetric monoidal structure obtained via (1), and $\mathsf{copy}_{(\mathcal{X},\alpha)} := \mathsf{copy}_{\mathcal{X}}$ and $\mathsf{del}_{(\mathcal{X},\alpha)} := \mathsf{del}_{\mathcal{X}}$ for each object $(\mathcal{X},\alpha)$ in $\mathsf{C}^T_{\mathrm{det}}$.*

A fairly straightforward consequence of this is the following: if C is a Markov category, and $\mathcal{G}$ is a group internal to $\mathsf{C}_{\mathrm{det}}$, then $\mathsf{C}^{\mathcal{G}} := \mathsf{C}^T_{\mathrm{det}}$ is also a Markov category, where we denote by T the writer monad $\mathcal{G} \otimes -$.

Application: probabilistic neural network symmetries

Theorem 3 provides a streamlined proof of (a more general version of) Theorem 9 of [BT20], the main result of that paper. To state this result, suppose $\mathcal{G}$ is a group acting on standard Borel spaces $\mathcal{X}$ and $\mathcal{Y}$, and X and Y are random elements of $\mathcal{X}$ and $\mathcal{Y}$ respectively. The idea is that X and Y represent the input and output of a (possibly stochastic) neural network. In this context, [BT20] show that (under some additional regularity conditions), if $(g \cdot X, g \cdot Y) \overset{\mathrm{d}}{=} (X, Y)$ holds for all $g \in \mathcal{G}$, then we can write

$$(X, Y) \overset{\mathrm{d}}{=} (X, f(X, \eta)), \tag{5}$$

where $\eta \sim \mathrm{Uniform}(0, 1)$ is independent of X, and it moreover holds that $f(g \cdot X, \eta) \overset{\mathrm{d}}{=} g \cdot f(X, \eta)$. Here $\overset{\mathrm{d}}{=}$ denotes equality in distribution.[1]

Our approach is to reduce this statement to the following result. (Notice that here we rely crucially on our Theorem 3, since without knowing whether $\mathsf{C}^{\mathcal{G}}$ is Markov, we cannot even talk about conditionals or randomness pushback in the first place.)

Proposition 4. *Suppose C is a Markov category with conditionals and randomness pushback, and $\mathcal{G}$ is a group in $\mathsf{C}_{\mathrm{det}}$. Then $\mathsf{C}^{\mathcal{G}}$ has conditionals and randomness pushback.*

This allows us to generalise the result of [BT20] as follows. Let $p : \mathcal{I} \to \mathcal{X} \otimes \mathcal{Y}$ denote the joint distribution of (X, Y), and $\mathsf{act}_{\mathcal{X}}$ and $\mathsf{act}_{\mathcal{Y}}$ the actions on $\mathcal{X}$ and $\mathcal{Y}$ respectively. The condition $(g \cdot X, g \cdot Y) \overset{\mathrm{d}}{=} (X, Y)$ then says that p is a morphism $(\mathcal{I}, \mathsf{act}_{\mathcal{I}}) \to (\mathcal{X}, \alpha_{\mathcal{X}}) \otimes (\mathcal{Y}, \alpha_{\mathcal{Y}})$ in $\mathsf{BorelStoch}^{\mathcal{G}}$. By Proposition 4, $\mathsf{BorelStoch}^{\mathcal{G}}$ has conditionals (since $\mathsf{BorelStoch}$ does), and so p admits a conditional $k : (\mathcal{X}, \mathsf{act}_{\mathcal{X}}) \to (\mathcal{Y}, \mathsf{act}_{\mathcal{Y}})$, which is hence automatically equivariant with respect to $\mathsf{act}_{\mathcal{X}}$ and $\mathsf{act}_{\mathcal{Y}}$ just by definition of $\mathsf{C}^{\mathcal{G}}$. In a similar way, $\mathsf{BorelStoch}^{\mathcal{G}}$ has randomness pushback, which allows us to write $k(dy|x)$ as the distribution of $f(x, \eta)$ for some deterministic function f and independent $\eta \sim \mathrm{Uniform}(0, 1)$. This then gives directly the desired result (5).

This same argument can be given inside a general Markov category C using string diagrams (although at the expense of more space than we have available here).

[1] [BT20] give a version that replaces the distributional equalities involving f with almost sure ones. This can be obtained by a straightforward extension of what we describe here.

References

[Bro+17] Michael M. Bronstein et al. "Geometric Deep Learning: Going beyond Euclidean data". In: *IEEE Signal Processing Magazine* 34.4 (2017), pp. 18–42. DOI: 10.1109/MSP.2017.2693418.

[Bro+21] Michael M. Bronstein et al. *Geometric Deep Learning: Grids, Groups, Graphs, Geodesics, and Gauges*. 2021. arXiv: 2104.13478 [cs.LG]. URL: https://arxiv.org/abs/2104.13478.

[BT20] Benjamin Bloem-Reddy and Yee Whye Teh. "Probabilistic symmetries and invariant neural networks". In: *The Journal of Machine Learning Research* 21.1 (2020), pp. 3535–3595.

[CJ19] Kenta Cho and Bart Jacobs. "Disintegration and Bayesian inversion via string diagrams". In: *Mathematical Structures in Computer Science* 29.7 (Mar. 2019), pp. 938–971. DOI: 10.1017/s0960129518000488. URL: https://doi.org/10.1017%2Fs0960129518000488.

[Cor25] Rob Cornish. *Stochastic Neural Network Symmetrisation in Markov Categories*. 2025. arXiv: 2406.11814 [stat.ML]. URL: https://arxiv.org/abs/2406.11814.

[Fri20] Tobias Fritz. "A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics". In: *Advances in Mathematics* 370 (Aug. 2020), p. 107239. DOI: 10.1016/j.aim.2020.107239. URL: https://doi.org/10.1016%2Fj.aim.2020.107239.

[MP23] Sean Moss and Paolo Perrone. "A category-theoretic proof of the ergodic decomposition theorem". In: *Ergodic Theory and Dynamical Systems* 43.12 (Feb. 2023), pp. 4166–4192. ISSN: 1469-4417. DOI: 10.1017/etds.2023.6. URL: http://dx.doi.org/10.1017/etds.2023.6.