Equivariant stochastic neural networks in Markov categories

Rob Cornish

Department of Statistics, University of Oxford

In many machine learning problems, it is useful to have a neural network that is equivariant with respect to a group action. That is, for some group G acting on some input and output spaces X and Y of interest, we would like a neural network $f : X \to Y$ that satisfies

$$f(g \cdot x) = g \cdot f(x) \qquad \text{for all } x \in X \text{ and } g \in G. \tag{1}$$

Such constraints naturally arise in many applications involving some geometric structure, such as computer vision, or scientific problems where the data involved are known to follow certain symmetries. For example, a common task in computer vision involves segmenting a point cloud according to the different objects it contains. If a neural network is used for this, it is desirable that it is equivariant with respect to the group of rigid-body transformations acting on the point cloud. Otherwise, the predictions of the network would depend on how the point cloud is oriented, which intuitively should not matter for the task at hand.

However, most off-the-shelf neural networks are not equivariant. Even after training on data that involves symmetries (such as point clouds), typically it will hold that $f(g \cdot x) \neq g \cdot f(x)$, possibly to a large degree. This can hamper the performance of the network and affect its ability to generalise, and so an active research area considers how to develop neural networks that are equivariant by design.

Intrinsic equivariance vs. symmetrisation Broadly speaking, there are two major approaches to obtaining equivariant neural networks. A significant body of work has focussed on *intrinsic equivariance*, which imposes certain constraints on individual layers of a neural network to ensure that the network as a whole is equivariant [CW16; RSP17; FWW21]. In contrast, a recent line of work may be described as *symmetrisation*, which takes an existing unconstrained neural network and modifies it in some way to become equivariant [Mur+19; Pun+22; Kab+23; Kim+23]. These approaches are attractive as they can be used with arbitrary neural network architectures, which reduces implementation complexity and allows for the use of powerful existing models.

Stochastic equivariance In this work, we consider a more general problem than (1), where we additionally allow our neural network to depend on some randomness, so that its outputs are stochastic. Such models are of interest in applications including generative modelling and reinforcement learning, and in situations where uncertainty quantification is required. Roughly speaking, we would like a model whose distribution of outputs is equivariant *across repeated executions*, rather than (say) at any single one. We can formalise this by considering f to be a *Markov kernel*, i.e. a measurable function of the form $X \to PY$, where PY denotes the set of distributions on Y equipped with some suitable σ-algebra. The equivariance condition (1) remains the same, except that G now acts on PY via the pushforward of its underlying action on Y.

Equivariance in Markov categories We use *Markov categories* [CJ19; Fri20] as a framework for reasoning about stochastically equivariant neural networks. At a high level, given a Markov category C, our approach is to consider groups and group actions internal to the subcategory of deterministic morphisms $\mathsf{C}_{\mathrm{det}}$. Since $\mathsf{C}_{\mathrm{det}}$ is cartesian monoidal [Fri20], we can recover standard results about group actions and homomorphisms via the usual arguments. However, we also naturally obtain a notion of equivariance that makes sense in the whole of C, i.e. including for morphisms that are not deterministic. Given a group G internal to C, this allows us to define a Markov category C^G, whose objects are objects of C equipped with a G-action, and whose morphisms are morphisms of C that are appropriately equivariant.

Stochastic symmetrisation We then consider the problem of symmetrising a stochastic neural network along a group homomorphism $\varphi : H \to G$, where H and G are groups in C. We formulate this as follows. Any such φ gives rise to a functor $\mathsf{Res}_\varphi : \mathsf{C}^G \to \mathsf{C}^H$ that maps a G-object to its *restriction* via φ. Our goal is to obtain functions of the following form:

$$\mathsf{C}^H(\mathsf{Res}_\varphi X, \mathsf{Res}_\varphi Y) \longrightarrow \mathsf{C}^G(X, Y) \qquad \text{where } X \text{ and } Y \text{ are } G\text{-objects.} \tag{2}$$

Notice that this maps H-equivariant morphisms to G-equivariant ones, which is exactly what is desired of a symmetrisation procedure (although previous work has not framed the problem in this way).

Suppose Res_φ has a left adjoint Ext_φ. This corresponds to the usual notion of *extension* of a group action along a homomorphism. Then we obtain a procedure of the form (2) very naturally via the following steps:

$$\mathsf{C}^H(\mathsf{Res}_\varphi X, \mathsf{Res}_\varphi Y) \xrightarrow{\text{Apply adjunction}} \mathsf{C}^G(\mathsf{Ext}_\varphi \mathsf{Res}_\varphi X, Y) \xrightarrow{\text{Precompose}} \mathsf{C}^G(X, Y). \tag{3}$$

Here in the second step we precompose by any G-equivariant morphism of the form $X \to \mathsf{Ext}_\varphi \mathsf{Res}_\varphi X$ in C^G. This accords with existing symmetriation procedures, which typically require some neural network that is already equivariant. However, the idea is that this can be much smaller and simpler than the "backbone" neural network we are symmetrising, which only needs to be H-equivariant. In the extreme, taking H to be the trivial group means the "backbone" network is completely constrained, and thereby potentially very expressive. Empirically, this leads to stronger predictive performance overall while maintaining the desired equivariance properties.

Theoretical results For this approach to be viable, we require that a left adjoint Ext_φ exists. We give a useful sufficient condition for this as follows:

Theorem 1. *A left adjoint to* Res_φ *exists if* C *has coequalisers of the form*

$$G \otimes X \underset{\mathsf{triv}}{\overset{\mathsf{act}}{\rightrightarrows}} X \tag{4}$$

where act *is any action of* G *on* X *and* triv *is the trivial action, and if moreover these coequalisers are preserved by the functor* $- \otimes Y$ *for every* Y *in* C.

We would also like a concrete Markov category C that actually satisfies this condition. We do not know whether the category Stoch of measurable spaces and Markov kernels does: while it does have coequalisers of the form (4), it is not clear that these are always preserved by the functor $- \otimes Y$. However, we do have the following, which is adequate for our purposes since neural networks are almost invariably continuous:

Theorem 2. *The Markov category* $\mathsf{TopStoch}$ *of topological spaces and continuous Markov kernels [FPR21] satisfies the conditions of the previous theorem.*

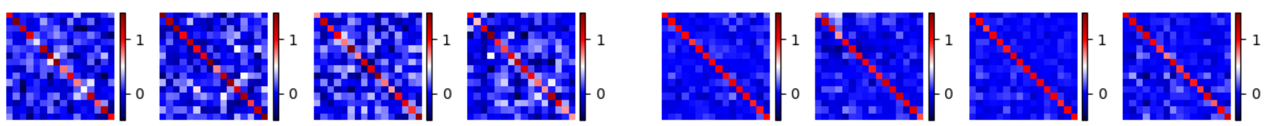
Practical methodology To make this approach practical, we need to be able actually to compute the steps in (3). For this, we show that there is an isomorphism

$$\mathsf{Ext}_\varphi \mathsf{Res}_\varphi X \cong G/H \otimes X$$

where G/H is the coequaliser of $\mathsf{act}, \mathsf{triv} : H \otimes G \rightrightarrows G$ when H acts on G by right-multiplication (through φ) by the inverse. This recovers a standard result that appears for example in equivariant algebraic topology [MC96], and leads to a straightforward procedure for computing (3) in practice. End-to-end, the resulting methodology recovers existing approaches to symmetrisation under various configurations, and moreover generalises these immediately to the stochastic setting.

Applications We show how our procedure can be instantiated for various groups and actions of interest, including compact groups, products, and semidirect products, all of which can be considered abstractly in a general Markov category C. We also provide specific concrete examples in $\mathsf{TopStoch}$, including for translation groups, compact matrix groups, the Euclidean groups, and the general linear group. Equivariance with respect to the general linear group does not appear to have been considered in the literature previously, possibly because it is quite complex to work with (being e.g. noncompact).

Empirical results We have implemented our methodology and obtained promising empirical results. A flavour of these is given in Figure 1, which shows the result of applying our methodology to learn the function $(-)^{-1}$ that maps an invertible matrix to its inverse. This is equivariant with respect to the action of the orthogonal group by left-multiplication and right-multiplication by the transpose,[1] since for orthogonal U we have $(UB)^{-1} = B^{-1}U^T$. Figure 1 shows the results produced by a neural network before and after symmetrisation using our method for a 16-dimensional matrix input. We have also obtained similarly positive results compared with other baseline procedures for obtaining equivariance, both intrinsically and via symmetrisation.



(a) f is an unconstrained neural network. (b) f is symmetrised using our method.

Figure 1: Values of $A\,f(A)$ obtained for four random 16×16 matrices A. Ideally each value should be close to the identity matrix, i.e. blue with a red line down the diagonal.

[1] More generally, $(-)^{-1}$ is equivariant with respect to the action of the two copies of the full general linear group acting via left- and right-multiplication, although the procedure becomes more complicated.

References

[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.

[CW16] Taco Cohen and Max Welling. "Group Equivariant Convolutional Networks". In: *Proceedings of The 33rd International Conference on Machine Learning*. Ed. by Maria Florina Balcan and Kilian Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, 20–22 Jun 2016, pp. 2990–2999. URL: https://proceedings.mlr.press/v48/cohenc16.html.

[FPR21] Tobias Fritz, Paolo Perrone, and Sharwin Rezagholi. "Probability, valuations, hyperspace: Three monads on top and the support as a morphism". In: *Mathematical Structures in Computer Science* 31.8 (Sept. 2021), pp. 850–897. ISSN: 1469-8072. DOI: 10.1017/s0960129521000414. URL: http://dx.doi.org/10.1017/S0960129521000414.

[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.

[FWW21] Marc Finzi, Max Welling, and Andrew Gordon Gordon Wilson. "A Practical Method for Constructing Equivariant Multilayer Perceptrons for Arbitrary Matrix Groups". In: *Proceedings of the 38th International Conference on Machine Learning*. Ed. by Marina Meila and Tong Zhang. Vol. 139. Proceedings of Machine Learning Research. PMLR, 18–24 Jul 2021, pp. 3318–3328. URL: https://proceedings.mlr.press/v139/finzi21a.html.

[Kab+23] Sékou-Oumar Kaba et al. "Equivariance with Learned Canonicalization Functions". In: *Proceedings of the 40th International Conference on Machine Learning*. Ed. by Andreas Krause et al. Vol. 202. Proceedings of Machine Learning Research. PMLR, 23–29 Jul 2023, pp. 15546–15566. URL: https://proceedings.mlr.press/v202/kaba23a.html.

[Kim+23] Jinwoo Kim et al. "Learning Probabilistic Symmetrization for Architecture Agnostic Equivariance". In: *Advances in Neural Information Processing Systems*. Ed. by A. Oh et al. Vol. 36. Curran Associates, Inc., 2023, pp. 18582–18612. URL: https://proceedings.neurips.cc/paper_files/paper/2023/file/3b5c7c9c5c7bd77eb73d0baec7a07165-Paper-Conference.pdf.

[MC96] J Peter May and Michael Cole. *Equivariant homotopy and cohomology theory: Dedicated to the memory of Robert J. Piacenza*. 91. American Mathematical Soc., 1996.

[Mur+19] Ryan L. Murphy et al. "Janossy Pooling: Learning Deep Permutation-Invariant Functions for Variable-Size Inputs". In: *International Conference on Learning Representations*. 2019. URL: https://openreview.net/forum?id=BJluy2RcFm.

[Pun+22] Omri Puny et al. "Frame Averaging for Invariant and Equivariant Network Design". In: *International Conference on Learning Representations*. 2022. URL: https://openreview.net/forum?id=zIUyj55nXR.

[RSP17] Siamak Ravanbakhsh, Jeff Schneider, and Barnabás Póczos. "Equivariance Through Parameter-Sharing". In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. PMLR, June 2017, pp. 2892–2901. URL: https://proceedings.mlr.press/v70/ravanbakhsh17a.html.