Stochastic Neural Network Symmetrisation in Markov Categories

Rob Cornish
Department of Statistics, University of Oxford

Extended abstract

Introductory talk

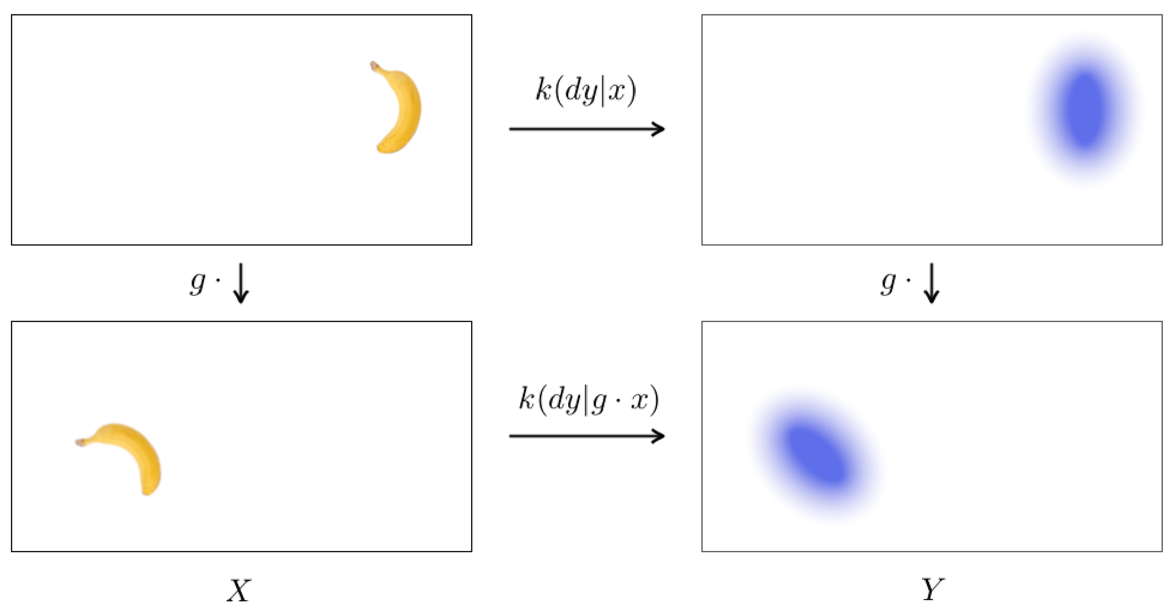
Full paper

UNIVERSITY OF OXFORD

TLDR: Recent interest in procedures for symmetrisation:

Unconstrained $f : X \to Y$ ⟶ Procedure ⟶ Equivariant $\mathbf{sym}(f) : X \to Y$

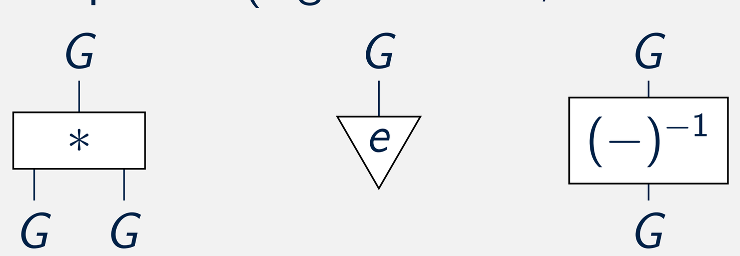
We fully characterise the space of such procedures. This extends existing methods, including to stochastic equivariance:



We do this using the framework of Markov categories.

Definition

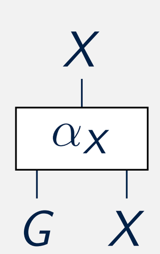
A group in a Markov category $\mathbf{C}$ is an object G (e.g. a set, measurable set, etc.) equipped with morphisms (e.g. functions, Markov kernels, etc.).



satisfying the usual axioms (unitality, associativity, inversion)

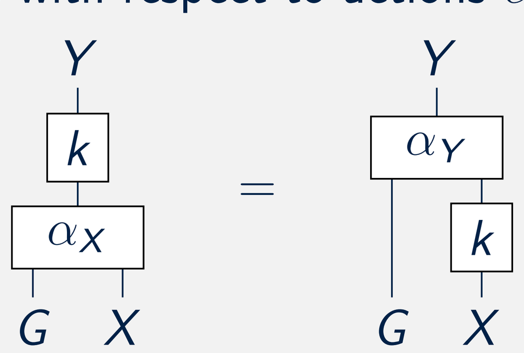
Definition

An action of G on X is a (unital and associative) morphism



Definition

$k : X \to Y$ is equivariant with respect to actions α_X and α_Y if



When k is a function, this says $k(g \cdot x) = g \cdot k(x)$ for all $g \in G$ and $x \in X$

Definition

Given a group G in a Markov category $\mathbf{C}$, obtain a Markov category $\mathbf{C}^G$:

- Objects are pairs (X, α_X), where α_X is an action of G on X
- Morphisms $(X, \alpha_X) \to (Y, \alpha_Y)$ are equivariant w.r.t. α_X and α_Y

Definition (for today)

A symmetrisation procedure is a function **sym** of the following form

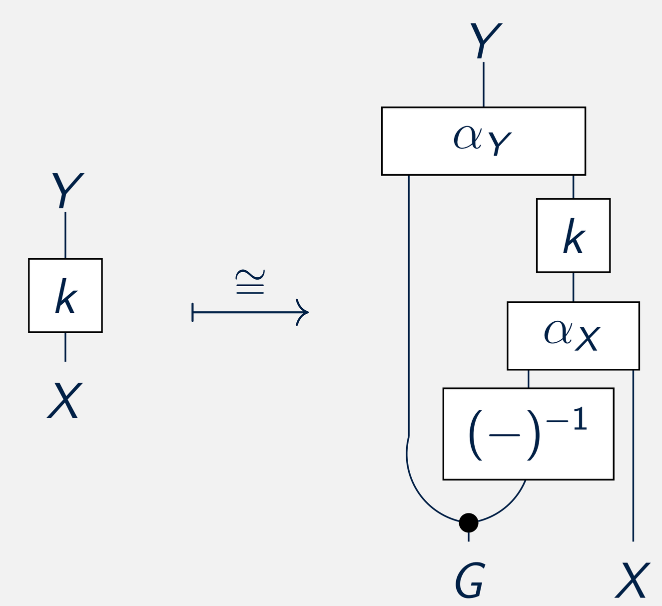
$$\underbrace{\mathbf{C}(X, Y)}_{\text{Morphisms } X \to Y \text{ in } \mathbf{C}} \xrightarrow{\mathbf{sym}} \mathbf{C}^G((X, \alpha_X), (Y, \alpha_Y))$$

Theorem

There is always a bijection

$$\mathbf{C}(X, Y) \xrightarrow{\cong} \mathbf{C}^G((G, *) \otimes (X, \alpha_X), (Y, \alpha_Y))$$

defined as follows:



Arises from a left adjoint to forgetful functor $U(X, \alpha_X) := X$, i.e.

$$\begin{tikzcd} \mathbf{C} \arrow[r, shift left, "F"] \arrow[r, phantom, "\perp"] & \mathbf{C}^G \arrow[l, shift left, "U"] \end{tikzcd}$$

Can generalise to partially equivariant models by replacing U with

$$R_\varphi : \mathbf{C}^G \to \mathbf{C}^H$$

the restriction functor of a homomorphism $\varphi : H \to G$

Corollary

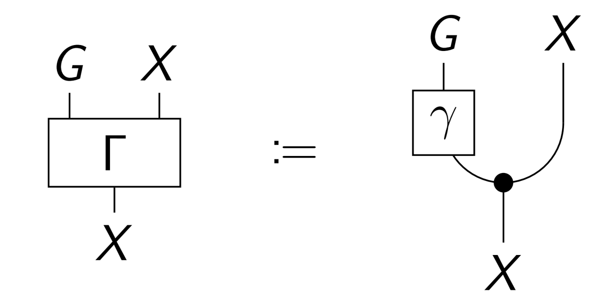
Every symmetrisation procedure $\mathbf{C}(X, Y) \xrightarrow{\mathbf{sym}} \mathbf{C}^G((X, \alpha_X), (Y, \alpha_Y))$ *can be expressed as a composition of the form*

$$\mathbf{C}(X, Y) \xrightarrow{\cong} \mathbf{C}^G((G, *) \otimes (X, \alpha_X), (Y, \alpha_Y)) \longrightarrow \mathbf{C}^G((X, \alpha_X), (Y, \alpha_Y)).$$

Only natural choice for second step is precomposition, i.e. $k \mapsto k \circ \Gamma$ as in

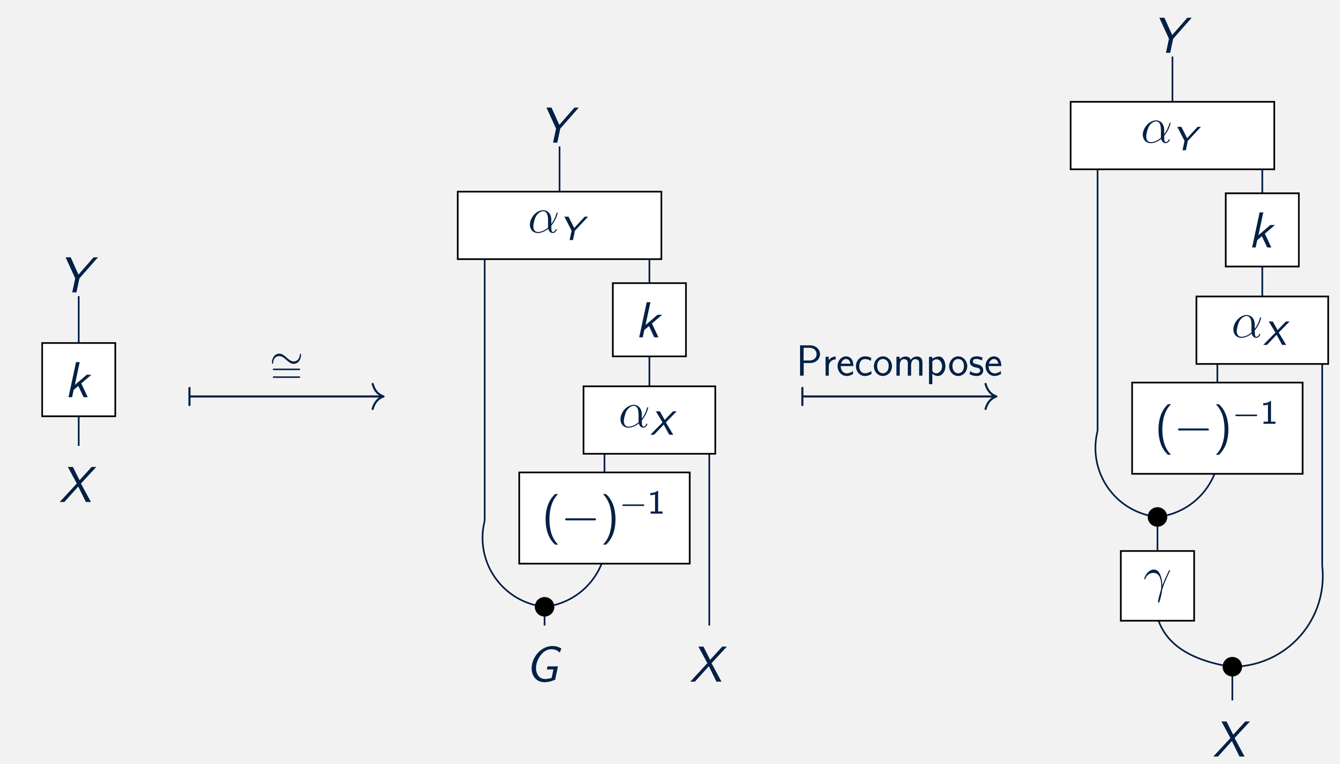
$$(X, \alpha_X) \xrightarrow{\Gamma} (G, *) \otimes (X, \alpha_X) \xrightarrow{k} (Y, \alpha_Y)$$

For example, an attractive choice is



Algorithm

Given a suitable $\gamma : X \to G$, overall procedure may now be computed as follows:



Corollary (Canonicalisation [Kaba et al., 2023])

If $k : X \to Y$ is any function, and $\gamma : X \to G$ is an equivariant function, then the following defines an equivariant function $X \to Y$:

$$\gamma(x) \cdot k(\gamma(x)^{-1} \cdot x)$$

Corollary (Stochastic symmetrisation)

If $k : X \to Y$ is any Markov kernel, and $\gamma : X \to G$ is an equivariant Markov kernel, then the following defines an equivariant Markov kernel $X \to Y$:

$$\mathbf{G} \sim \gamma(dg|x) \qquad \mathbf{Y} \sim k(dy|\mathbf{G}^{-1} \cdot x) \qquad \text{return } \mathbf{G} \cdot \mathbf{Y}$$

Other methods (e.g. [Puny et al., 2022, Kim et al., 2023]) obtained via a further averaging step:

$$k(dy|x) \xmapsto{\mathbf{ave}} \int y \, k(dy|x)$$

provided Y is convex and its action is linear