Extended Abstract Track

Neural Network Symmetrisation in Concrete Settings

Editors: List of editors' names

Abstract

Cornish (2024) recently gave a general theory of neural network *symmetrisation* in the abstract context of Markov categories. We describe these results in familiar concrete settings.

Keywords: Equivariance, Symmetrisation, Stochastic

1. Introduction

It is often useful to ensure that a neural network $f : X \to Y$ is *equivariant* with respect to the actions of some group. Recently there has been interest in doing so via *symmetrisation* techniques (Yarotsky, 2018). Roughly speaking, these approaches take $f \coloneqq \mathsf{sym}(f_0)$ where $f_0 : X \to Y$ is some neural network that is *not* equivariant, and sym is a function that maps non-equivariant neural networks to equivariant ones.

A variety of different choices of sym have been considered in the literature to-date, including Janossy pooling (Murphy et al., 2019), frame averaging (Puny et al., 2022), canonicalisation (Kaba et al., 2023), and probabilistic averaging (Kim et al., 2023; Dym et al., 2024). Recently, Cornish (2024) gave a general theory of symmetrisation that characterises all possible choices of sym, recovering these previous techniques as special cases, and extending to the novel setting of *stochastic* neural networks f, which had not previously been considered. This framework also streamlines the presentation of compositional and recursive usage of these techniques, and encompasses a range of complex situations such as noncompact translation groups and semidirect products in a uniform way.

The results of Cornish (2024) were developed in terms of *Markov categories* (Fritz, 2020), which provide a high-level algebraic framework for reasoning about probability in an intuitive yet rigorous way. However, Markov categories are currently not widely known in the machine learning community, and so in this paper we present special cases of Cornish (2024) in more familiar concrete settings. We begin in Section 2 with the purely deterministic case considered by Kaba et al. (2023), and then extend to include randomness in Section 3.

We will assume knowledge of only the basic definition of a category, provided for completeness in Appendix A. Given a category C, we will denote by $\mathsf{C}(X, Y)$ the set of morphisms $X \to Y$ in C. We also denote the category of sets and functions by Set.

2. Deterministic symmetrisation

The category of G-sets For every group G, there is a category Set^G. Each object of this category is a *G-set*, or in other words a set X equipped with a group action $\alpha_X : G \times X \to X$. We will usually write $\alpha_X(g, x)$ simply as $g \cdot x$ when α_X is clear from context. The morphisms $f : X \to Y$ in Set^G are then functions that are *equivariant* with respect to α_X and α_Y, so that $f(g \cdot x) = g \cdot f(x)$ holds for all $x \in X$ and $g \in G$. *Invariant* functions are the special case where α_Y is the trivial action with no effect, in which case we have $f(g \cdot x) = f(x)$.

Given two G-sets X and Y, their *product* is always another G-set that we denote by $X \otimes Y$. This is simply the cartesian product of X and Y equipped with the *diagonal* action $\alpha_{X \otimes Y}$ defined as $g \cdot (x, y) \coloneqq (g \cdot x, g \cdot y)$.

Symmetrisation procedures Suppose $H \subseteq G$ is a subgroup. Given any G-set X, we can always obtain an H-set RX with the same underlying set as X, and whose H-action is obtained via *restriction*, so that $\alpha_{RX}(h, x) \coloneqq \alpha_X(h, x)$. This allows us to define a *symmetrisation procedure* concretely as any function of the form:

$$\mathsf{Set}^H(RX, RY) \xrightarrow{\mathsf{sym}} \mathsf{Set}^G(X, Y). \tag{1}$$

Notice that such a sym takes a function that is equivariant only with respect to the subgroup H and "upgrades" it to become equivariant with respect to the whole group G. If H is the trivial subgroup consisting of just the identity element, then H-equivariance always holds trivially, and sym may take as input any arbitrary function $X \to Y$. The latter case has received the majority of attention in the symmetrisation literature, although Section 3.3 of Kaba et al. (2023) considers the general setup here.

Although we focus on (1), Definition 4.4 of Cornish (2024) formulates symmetrisation slightly more generally in terms of a homomorphism $\varphi : H \to G$. The case of (1) is then recovered by considering a subgroup inclusion $\varphi : H \hookrightarrow G$. As we explain in Appendix B, this approach can be more convenient for describing the composition of multiple symmetrisation procedures in sequence, which allows for "building up" complex equivariance properties in a structured way.

A general characterisation The following result allows us to characterise all symmetrisation procedures (1). Here G/H denotes the set of *H-cosets*, so that $G/H \coloneqq \{[g] \mid g \in G\}$, where $[g] \coloneqq \{gh \mid h \in H\}$. This set always comes equipped with the G-action $g \cdot [g'] \coloneqq [gg']$, and hence may always be regarded as an object in Set^G.

Theorem 1 *For all choices of the various components involved, there is a bijection*

$$\mathsf{Set}^H(RX, RY) \xrightarrow{\cong} \mathsf{Set}^G(G/H \otimes X, Y) \tag{2}$$

that sends $f \mapsto f^\sharp$ defined as $f^\sharp([g], x) \coloneqq g \cdot f(g^{-1} \cdot x)$.

This follows from Theorem 5.1 of Cornish (2024). We also give a self-contained proof in Appendix C, where we explain how this result arises very naturally in the context of category theory. It now follows that the following two steps always constitute a symmetrisation procedure for every choice of the function func shown:

$$\mathsf{Set}^H(RX, RY) \xrightarrow{\cong} \mathsf{Set}^G(G/H \otimes X, Y) \xrightarrow{\mathsf{func}} \mathsf{Set}^G(X, Y), \tag{3}$$

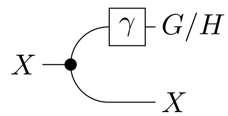
where the first arrow denotes (2). Moreover, since (2) is a bijection, *every* symmetrisation procedure can be obtained in this way for some choice of func. In other words, (3) fully characterises all possible functions of the desired form (1).

Precomposition The question now is, how can we obtain func? If we want a "general purpose" strategy that works without further assumptions, then there is only one obvious choice. This is namely the mapping that sends $f^\sharp : G/H \otimes X \to Y$ in Set^G to the composition

$$X \xrightarrow{\omega} G/H \otimes X \xrightarrow{f^\sharp} Y \tag{4}$$

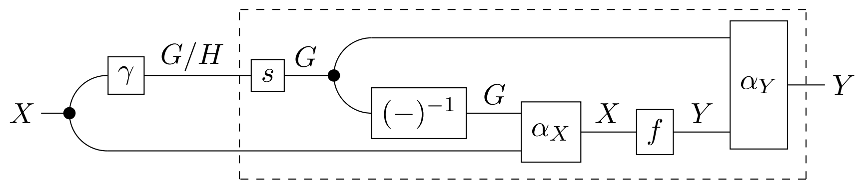
where ω is any choice of morphism $X \to G/H \otimes X$ in Set^G. In other words, $\mathsf{func}(f^\sharp) := f^\sharp \circ \omega$. It follows directly from the fact that Set^G is closed under composition that func here is a function of the type required in (3).

A reasonable condition for a symmetrisation procedure sym to satisfy is *stability*: it should be the case that, if f is already G-equivariant, then $\mathsf{sym}(f) = f$. When func is obtained via precomposition as in (4), this holds if and only if $\omega : X \to G/H \otimes X$ can be written as $\omega(x) = (\gamma(x), x)$ for some $\gamma : X \to G/H$ in Set^G (Cornish, 2024, Proposition 5.1). In other words, ω must have the following "shape" (read from left to right):



(5)

End-to-end procedure With func obtained in this way, the overall symmetrisation procedure (3) maps $f : RX \to RY$ in Set^H to the following morphism $X \to Y$ in Set^G:



(6)

Here the dashed box denotes $f^\sharp$ obtained from (2), with $s : G/H \to G$ any choice of *coset representative*, so that $s([g]) \in [g]$. Likewise, $(-)^{-1}$ just inverts its input, sending $g \mapsto g^{-1}$. Note that (5) is simply "plugged in" to the dashed box as per the precomposition step (4).

In more traditional notation, letting $h := s \circ \gamma$, the result (6) maps $x \in X$ to the value

$$h(x) \cdot f(h(x)^{-1} \cdot x),$$

which recovers the canonicalisation architecture of Kaba et al. (2023) (see their Theorem 3.1). The account given here therefore provides a theoretical explanation of how this architecture arises. Additionally, as we describe next, this same story can now be generalised beyond Set to incorporate *stochasticity*, which is useful for many practical applications.

3. Stochastic symmetrisation

Markov kernels Every component considered above (e.g. f and γ in (6)) was a deterministic function. In this section, we now allow these to depend on some additional randomness. We do so by formalising these components as *Markov kernels* instead of functions. We gloss over some mathematical details here, giving additional background in Appendix D instead.

A Markov kernel $k : X \to Y$ models a *conditional distribution* or *stochastic function* that, when given an input $x \in X$, produces a random output in Y with distribution $k(dy|x)$. For most Markov kernels of interest, $k(dy|x)$ is obtained as the distribution of some random variable $f(x, \boldsymbol{U})$, where $\boldsymbol{U}$ is a random variable taking values in some space U, and $f : X \times U \to Y$ is a (deterministic) function. By letting $\boldsymbol{U}$ be constant, every deterministic function $X \to Y$ can be regarded as a Markov kernel that happens to be deterministic. Markov kernels can be composed and so form a category called Stoch (see Appendix D).

Stochastic equivariance Suppose G is a group acting on X and Y. It is natural to say that a Markov kernel $k : X \to Y$ is *equivariant* if for all $g \in G$ and $x \in X$ we have

$$k(dy|g \cdot x) = (g \cdot k)(dy|x), \tag{7}$$

where the right-hand side denotes the *pushforward* of $k(dy|x)$, i.e. the distribution of $g \cdot \boldsymbol{Y}$ when $\boldsymbol{Y} \sim k(dy|x)$. For k obtained from f and $\boldsymbol{U}$ as above, this condition says $f(g \cdot x, \boldsymbol{U}) \overset{\mathrm{d}}{=} g \cdot f(x, \boldsymbol{U})$, where $\overset{\mathrm{d}}{=}$ is equality in distribution. By Proposition 3.1 of Cornish (2024), this also recovers the usual notion of equivariance for conditional *densities* $p(y|x)$, which says that $p(g \cdot y|g \cdot x) = p(y|x)$ always holds (e.g. (Xu et al., 2022; Hoogeboom et al., 2022)).

Stochastic symmetrisation We now consider how to obtain equivariant Markov kernels via symmetrisation. The key point of Cornish (2024) is that the developments in Section 2 can be generalised beyond Set to any *Markov category*. It turns out that Stoch is also a Markov category (Fritz, 2020, Section 4), which gives rise to a framework for stochastic symmetrisation directly as a special case. We sketch this now.

Like in Set, every group G gives rise to a category Stoch^G whose morphisms are G-equivariant Markov kernels. Given a subgroup $H \subseteq G$, we can define a *symmetrisation procedure* as in (1), with Set replaced by Stoch. Under typical circumstances, Theorem 5.1 of Cornish (2024) then yields a bijection analogous to (2), also with Set replaced by Stoch. This gives rise to a recipe for symmetrising Markov kernels analogous to (3):

$$\mathsf{Stoch}^H(RX, RY) \xrightarrow{\cong} \mathsf{Stoch}^G(G/H \otimes X, Y) \xrightarrow{\text{Precompose by (5)}} \mathsf{Stoch}^G(X, Y), \tag{8}$$

where now $\gamma : X \to G/H$ is a morphism in Stoch^G, hence a G-equivariant Markov kernel. Again, this is the only obvious approach that works without further assumptions.

The symmetrised Markov kernel $\mathsf{sym}(k)$ obtained from (8) has the same form as (6), with f replaced by k. However, since the components involved are now Markov kernels, the interpretation is different: given an input $x \in X$, we now *sample* from $\mathsf{sym}(k)(dy|x)$ via

$$\boldsymbol{\Gamma} \sim \gamma(d\mathfrak{g}|x) \qquad \boldsymbol{G} \sim s(dg|\boldsymbol{\Gamma}) \qquad \boldsymbol{Y} \sim k(dy|\boldsymbol{G}^{-1} \cdot x) \qquad \text{return } \boldsymbol{G} \cdot \boldsymbol{Y}.$$

Here s again amounts to a choice of coset representatives (Cornish, 2024, Remark 5.2). If G is compact, we may take $(s \circ \gamma)(dg|x)$ to be the Haar measure on G (Cornish, 2024, Example 6.1). When k is an *unconditional* distribution on Y, so that $k(dy|x)$ does not depend on x, this recovers the symmetrisation approach in Section 4 of Gelberg et al. (2024).

Averaging The symmetrisation procedure sym just described produces a Markov kernel $\mathsf{sym}(k)$ that is in general stochastic. When $Y = \mathbb{R}^d$, we can obtain a deterministic result by computing $\mathsf{ave}(\mathsf{sym}(k))$, where $\mathsf{ave}(m)$ deterministically outputs the expected value $\int y\, m(dy|x)$ when given the input $x \in X$. If G acts on Y linearly, then $\mathsf{ave}(\mathsf{sym}(k))$ is also G-equivariant (Cornish, 2024, Proposition 4.1). For H the trivial subgroup, this recovers the method of Kim et al. (2023), and by implication other related methods such as Janossy pooling (Murphy et al., 2019) and frame averaging (Puny et al., 2022).

However, ave can be a costly operation to compute, especially in high dimensions, and may not even be defined if Y is not a convex space like $\mathbb{R}^d$. By working with the stochastic condition (7) directly, the symmetrisation approach of Cornish (2024) avoids these issues while still incorporating symmetry constraints into the architecture of the resulting model.

References

Rob Cornish. Stochastic neural network symmetrisation in markov categories, 2024. URL https://arxiv.org/abs/2406.11814.

Nadav Dym, Hannah Lawrence, and Jonathan W. Siegel. Equivariant frames and the impossibility of continuous canonicalization. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 12228–12267. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/dym24a.html.

Tobias Fritz. A synthetic approach to markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics*, 370:107239, 8 2020. doi: 10.1016/j.aim.2020.107239. URL https://doi.org/10.1016%2Fj.aim.2020.107239.

Yoav Gelberg, Tycho F. A. van der Ouderaa, Mark van der Wilk, and Yarin Gal. Variational inference failures under model symmetries: Permutation invariant posteriors for bayesian neural networks, 2024. URL https://arxiv.org/abs/2408.05496.

Emiel Hoogeboom, Víctor Garcia Satorras, Clément Vignac, and Max Welling. Equivariant diffusion for molecule generation in 3D. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 8867–8887. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/hoogeboom22a.html.

Sékou-Oumar Kaba, Arnab Kumar Mondal, Yan Zhang, Yoshua Bengio, and Siamak Ravanbakhsh. Equivariance with learned canonicalization functions. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 15546–15566. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/kaba23a.html.

Jinwoo Kim, Dat Nguyen, Ayhan Suleymanzade, Hyeokjun An, and Seunghoon Hong. Learning probabilistic symmetrization for architecture agnostic equivariance. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine, editors, *Advances in Neural Information Processing Systems*, volume 36, pages 18582–18612. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/3b5c7c9c5c7bd77eb73d0baec7a07165-Paper-Conference.pdf.

J.P. May. *Equivariant Homotopy and Cohomology Theory: Dedicated to the Memory of Robert J. Piacenza*. Regional conference series in mathematics. American Mathematical Society, 1996. ISBN 9780821803196. URL https://books.google.co.uk/books?id=KOcZYVxkQO8C.

Ryan L. Murphy, Balasubramaniam Srinivasan, Vinayak Rao, and Bruno Ribeiro. Janossy pooling: Learning deep permutation-invariant functions for variable-size inputs. In *In-*

ternational Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=BJluy2RcFm.

Paolo Perrone. *Starting Category Theory*. WORLD SCIENTIFIC, October 2023. ISBN 9789811286018. doi: 10.1142/13670. URL http://dx.doi.org/10.1142/13670.

Omri Puny, Matan Atzmon, Edward J. Smith, Ishan Misra, Aditya Grover, Heli Ben-Hamu, and Yaron Lipman. Frame averaging for invariant and equivariant network design. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=zIUyj55nXR.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2023. URL https://arxiv.org/abs/1706.03762.

Minkai Xu, Lantao Yu, Yang Song, Chence Shi, Stefano Ermon, and Jian Tang. Geodiff: A geometric diffusion model for molecular conformation generation. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=PzcvxEMzvQC.

Dmitry Yarotsky. Universal approximations of invariant maps by neural networks, 2018.

Appendix A. Basic category theory

A highly accessible introduction to category theory can be found in Perrone (2023). We provide only the basic definition of a category here, which is all we require in the main text. A *category* consists of a collection of *objects* and a collection of *morphisms*. Each morphism f has two associated objects X and Y referred to as its *domain* and *codomain* respectively, which we denote by $f : X \to Y$. Given morphisms $f : X \to Y$ and $g : Y \to Z$, we can obtain their *composition* $g \circ f$, which is a morphism $X \to Z$. In addition, for every object X, there is an identity morphism $\mathrm{id}_X : X \to X$ which interacts with composition in the obvious way. Precisely, it holds that $f \circ \mathrm{id}_X = f$ and $\mathrm{id}_X \circ h = h$ for all $f : X \to Y$ and $h : Z \to X$.

Example Perhaps the most familiar example of a category is Set, whose objects $X, Y, \ldots$ are sets, and whose morphisms $X \to Y$ are just functions $f : X \to Y$. We can always compose pairs of functions whose domain and codomain match, and moreover every set admits an identity function, so that Set is indeed a category.

Appendix B. Symmetrising along a homomorphism

Cornish (2024) formulates symmetrisation procedures in slightly more general terms than (1). In particular, their starting point is an arbitrary group homomorphism $\varphi : H \to G$. As for the subgroup approach described above, any G-set X again yields an H-set $R_\varphi X$ with the same underlying set as X, and its H-action defined as

$$\alpha_{R_\varphi X}(h, x) \coloneqq \alpha_X(\varphi(h), x).$$

This always satisfies the axioms of an action since φ is a homomorphism. By Definition 4.1 of Cornish (2024), a symmetrisation procedure in Set is then any function of the form

$$\mathsf{Set}^H(R_\varphi X, R_\varphi Y) \xrightarrow{\mathsf{sym}} \mathsf{Set}^G(X, Y). \tag{9}$$

The case for a subgroup $H \subseteq G$ from (1) can then be recovered by letting φ be the subgroup inclusion $H \hookrightarrow G$.

In practice, most homomorphisms φ of interest seem to correspond to a subgroup inclusion in some way, at least in spirit. However, symmetrisation "along a homomorphism" like this can be cleaner to talk about for more complex use-cases. For example, given groups H and K, there is an obvious homomorphism

$$H \to K \times H$$

that sends $h \mapsto (e_K, h)$, where $e_K \in K$ is the identity element. With (9) we can talk about symmetrising along this homomorphism directly, whereas we would otherwise need to define the subgroup

$$\{(e_K, h) \mid h \in H\} \subseteq H \times K,$$

which is of course isomorphic to H, but somewhat more unwieldy to write down.

The homomorphism approach is particularly convenient for describing the composition of multiple symmetrisation procedures in sequence (see Section 4.5 of Cornish (2024)). For example, for 3D point cloud data, it is often of interest to obtain a model that is $S_n \times E(3)$ equivariant, where S_n is the symmetric group and $E(3)$ is the Euclidean group (see e.g. (Kim et al., 2023, Section 3.2)). One approach here would be to symmetrise in sequence along the obvious homomorphisms

$$S_n \longrightarrow O(3) \times S_n \longrightarrow E(3) \times S_n, \tag{10}$$

where $O(3)$ is the orthogonal group. This would start with a model that is S_n-equivariant (such as a transformer without positional encodings (Vaswani et al., 2023)), then upgrade this to become also $O(3)$-equivariant, and finally upgrade this again to become fully $E(3)$-equivariant. The diagram (10) summarises this process directly, whereas this again becomes more unwieldy to write down in terms of subgroups.

Appendix C. Proof of Theorem 1

Proof It follows from the assumption that f is H-equivariant that $f^\sharp$ is well-defined, since we have

$$\begin{aligned}
f^\sharp([gh], x) &= (gh) \cdot f((gh)^{-1} \cdot x) \\
&= (gh) \cdot f(h^{-1} \cdot g^{-1} \cdot x) \\
&= (gh) \cdot h^{-1} \cdot f(g^{-1} \cdot x) \\
&= g \cdot f(g^{-1} \cdot x) \\
&= f^\sharp([g], x).
\end{aligned}$$

Now, given $f^\sharp$, we can recover f since

$$f^\sharp([e], x) = e \cdot f(e^{-1} \cdot x) = f(x),$$

where $e \in G$ is the identity element. This shows (2) is injective. On the other hand, given any $h : G/H \otimes X \to Y$ in Set^G, letting $f(x) \coloneqq h([e], x)$, we recover $f^\sharp = h$ since

$$\begin{aligned} f^\sharp([g], x) &= g \cdot f(g^{-1} \cdot x) \\ &= g \cdot h([e], g^{-1} \cdot x) \\ &= h(g \cdot [e], g \cdot g^{-1} \cdot x) \\ &= h([g], x), \end{aligned}$$

where we use the assumption that h is G-equivariant in the third step. This shows (2) is surjective, hence a bijection. ∎

Categorical explanation This result arises very naturally from the perspective of category theory. The idea is that R is actually part of a *functor* with a left adjoint E as shown:

$$\begin{tikzcd} \mathsf{Set}^H \arrow[r, shift left, "E"] \arrow[r, phantom, "\perp"] & \mathsf{Set}^G. \arrow[l, shift left, "R"] \end{tikzcd}$$

This is a classical result (see e.g. (May, 1996, (1.4))). In particular, the existence of this adjunction means that

$$\mathsf{Set}^H(RX, RY) \cong \mathsf{Set}^G(ERX, Y). \tag{11}$$

It is also classical to show that there is an isomorphism of G-sets

$$ERX \cong G/H \otimes X$$

(see e.g. (May, 1996, (1.6))). Substituting this into (11) yields the desired bijection

$$\mathsf{Set}^H(RX, RY) \cong \mathsf{Set}^G(G/H \otimes X, Y).$$

The proof of Theorem 1 simply describes in more detail exactly how this bijection can actually be computed.

Appendix D. Technical details around Markov kernels

Definition Technically, a Markov kernel $k : X \to Y$ is a function of the form

$$k : \Sigma_Y \times X \to [0, 1], \tag{12}$$

where X and Y are measurable spaces, such that the function $x \mapsto k(B|x)$ is measurable for every $B \in \Sigma_Y$, and the function $B \mapsto k(B|x)$ is a probability measure for every $x \in X$. Here Σ_Y denotes the σ-algebra associated to Y.

The category Stoch Markov kernels give rise to a category called Stoch. Formally, this has measurable spaces as its objects. Given two measurable spaces X and Y, the morphisms $X \to Y$ are then just the Markov kernels of this form. Each identity kernel $\mathrm{id}_X : X \to X$ is obtained as

$$\mathrm{id}_X(dy|x) \coloneqq \delta_x(dy),$$

where the right-hand side denotes the Dirac measure at $x \in X$. Composition is performed via the *Chapman-Kolmogorov formula*: given Markov kernels $k : X \to Y$ and $m : Y \to Z$, we obtain $m \circ k : X \to Z$ with

$$(m \circ k)(B|x) \coloneqq \int m(B|y)\, k(dy|x)$$

for all $B \in \Sigma_Z$ and $x \in X$. Intuitively, to sample from $(m \circ k)(dz|x)$, we just sample from k and m in sequence as follows:

$$\boldsymbol{Y} \sim k(dy|x) \qquad \boldsymbol{Z} \sim m(dz|\boldsymbol{Y}) \qquad \text{return } \boldsymbol{Z}.$$

For a more detailed overview of Stoch, see e.g. Section 4 of Fritz (2020).