Categorical Stochastic Processes and Likelihood

Dan Shiebler
University of Oxford

April 8, 2020

Abstract

In this work we take a Category Theoretic perspective on the relationship between probabilistic modeling and function approximation. We begin by defining two extensions of function composition to stochastic process subordination: one based on the co-Kleisli category under the comonad $(\Omega \times -)$ and one based on the parameterization of a category with a Lawvere theory. We show how these extensions relate to the category **Stoch** and other Markov Categories.

Next, we apply the **Para** construction to extend stochastic processes to parameterized statistical models and we define a way to compose the likelihood functions of these models. We conclude with a demonstration of how the Maximum Likelihood Estimation procedure defines an identity-on-objects functor from the category of statistical models to the category **Learn**.

Code to accompany this paper can be found on GitHub*.

1 Preliminaries

1.1 Probability Measures, Random Variables and Markov Kernels

A **probability space** is a triplet (Ω, Σ, μ) where (Ω, Σ) is a measurable space and μ is a **probability measure** over (Ω, Σ). That is, μ is a countably additive function over the σ-algebra Σ that returns results in the unit interval $[0, 1]$ such that $\mu(\varnothing) = 0, \mu(\Omega) = 1$. Recall that Σ is a set of subsets of Ω. For some set Ω, we will write $\mathcal{B}(\Omega)$ for the Borel algebra of Ω, or the smallest σ-algebra that contains all open sets.

A **random variable** over a probability space (Ω, Σ, μ) is a measurable function $f : (\Omega, \Sigma) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. In this paper we will assume $\Sigma_{\mathbb{R}}$ is $\mathcal{B}(\mathbb{R})$, or the Borel algebra of $\mathbb{R}$. We will sometimes use the term "random variable" to refer to measurable functions of the form $f : (\Omega, \Sigma) \to (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ as well. These are also called multivariate random variables or random vectors [Wik19]. The pushforward $f_*\mu$ of a random variable $f : \Omega \to \mathbb{R}$ over the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ is a probability measure over $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ defined to be $f_*\mu(\sigma_{\mathbb{R}}) = \mu(f^{-1}(\sigma_{\mathbb{R}}))$.

A **Markov kernel** between the measurable space (A, Σ_A) and the measurable space (B, Σ_B) is a measurable function $\mu : A \times \Sigma_B \to [0, 1]$ such that for all $x_a \in A$,

*https://github.com/dshieble/Categorical_Stochastic_Processes_and_Likelihood

$\mu(x_a, -)$ is a probability measure over (B, Σ_B). That is:

$$\mu_{x_a}(B) = 1 \qquad \mu_{x_a}(\varnothing) = 0$$

For example, a Markov Kernel between $\varnothing$ and the measurable space (A, Σ_A) is just a probability measure over (A, Σ_A).

A **stochastic process** over a probability space (Ω, Σ, μ) is a family of random variables indexed by some set T. That is, we can write a stochastic process as a function $f : \Omega \times T \to \mathbb{R}$. In this paper we will limit our study to stochastic processes that are Borel-measurable in both arguments. We can define the pushforward of the stochastic process f along the probability measure μ to be the Markov Kernel $f_*\mu : T \times \mathcal{B}(\mathbb{R}) \to [0, 1]$ where:

$$f_*\mu_t(\sigma) = f(\_, t)_*\mu(\sigma) = \int_{\omega \in \Omega} \delta_{f(\omega,t)}(\sigma) d\mu$$

1.2 Categories

A central category that we will work off of is the symmetric monoidal category **Meas** of measurable spaces and measurable functions. The objects in **Meas** are pairs (A, Σ_A), where Σ_A is a σ-algebra over A. A morphism from (A, Σ_A) to (B, Σ_B) in **Meas** is a measurable function f such that for any $\sigma_B \in \Sigma_B$, $f^{-1}(\sigma_B) \in \Sigma_A$. The tensor product of the measurable spaces (A, Σ_A) and (B, Σ_B) in **Meas** is the space $(A \times B, \Sigma_A \otimes \Sigma_B)$, where $\Sigma_A \otimes \Sigma_B$ is the product σ-algebra of Σ_A and Σ_B. Note that **Meas** is not cartesian closed. The authors of [HKSY17] introduce a similar category **QBS** that is cartesian closed.

It will sometimes be convenient to work in the following subcategories of **Meas**:

Definition 1.1. $\mathbf{Meas}_{\mathbb{R}}$ *is the subcategory of* **Meas** *where objects are restricted to be Euclidean spaces equipped with their Borel σ-algebras* $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$.

Definition 1.2. **EucMeas** *is the subcategory of* $\mathbf{Meas}_{\mathbb{R}}$ *where morphisms are restricted to be differentiable.*

Another important category that we will consider is **Stoch** [Law62] [Gir82], which has measurable spaces as objects and Markov kernels as morphisms. We define the composition of the Markov kernels $\mu : A \times \Sigma_B \to [0, 1]$ and $\mu' : B \times \Sigma_C \to [0, 1]$ to be the following, where $x_a \in A$ and $x_c \in C$:

$$(\mu' \circ \mu)_{x_a}(\sigma_c) = \int_{x_b \in B} \mu'_{x_b}(\sigma_c) d\mu_{x_a}$$

The identity morphism at (A, Σ_A) is δ where:

$$\delta_{x_a}(\sigma_a) = \begin{cases} 1 & x_a \in \sigma_a \\ 0 & x_a \notin \sigma_a \end{cases}$$

The tensor product of the Markov Kernels $\mu : A \times \Sigma_B \to [0,1]$ and $\mu' : C \times \Sigma_D \to [0,1]$ in **Stoch** is the Markov Kernel:

$$(\mu' \otimes \mu) : (A \times C) \times (\Sigma_B \times \Sigma_D) \to [0,1]$$
$$(\mu' \otimes \mu)_{x_{ac}}(\sigma_{bd}) = \mu_{x_a}(\sigma_b) * \mu_{x_c}(\sigma_d)$$

The objects in **Stoch** are also equipped with a commutative comonoidal structure that is compatible with the monoidal product in **Stoch**. The authors of [Fri19] dub categories with this comonoidal+monoidal structure **Markov Categories**.

Stoch naturally arises as the Kleisli category of the Giry Monad, which is an affine symmetric monoidal monad that sends a measurable space to the space of probability measures over that space [Gir82].

Stoch has many notable subcategories based on slight restrictions of these measurable spaces. For example, the category **FinStoch** consists of finite measurable spaces and Markov Kernels between them. In order to be able to define regular conditional probabilities, the authors of [Fon13] and [CS12] restrict to countably generated measurable spaces (**CGStoch**) and the authors of [FR19] restricts to Polish spaces (**BorelStoch**).

1.2.1 Random Variables and Independence in Stoch

In any categorical presentation of probability, a natural question is how to reason about the notion of independence of random variables [GLS16] [Fra02] [Fri19].

Since **Stoch** is the Kleisli category of **Meas** under the Giry monad [Gir82], we can define an embedding functor from **Meas** into **Stoch** that acts as an identity on objects and sends the measurable function $f : (A, \Sigma_A) \to (B, \Sigma_B)$ to the Dirac Markov kernel δ_f where for any $x_a \in A, \sigma_b \in \Sigma_b$ we have that:

$$\delta_{f(x_a)}(\sigma_b) = \begin{cases} 1 & f(x_a) \in \sigma_b \\ 0 & f(x_a) \notin \sigma_b \end{cases}$$

This formalizes the intuition that Markov Kernels are a generalization of both measurable functions and probability measures, and provides an avenue to directly study random variables and their independence in **Stoch**.

Now say we have a probability space (Ω, Σ, μ) and two real-valued random variables over this space f, f'. We can think of these random variables as morphisms in **Meas** from (Ω, Σ) to $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. We can represent our probability space in **Stoch** as a morphism in **Stoch** between 1 and (Ω, Σ): that is, a Markov kernel $\mu : 1 \times \Sigma \to [0,1]$. We can then represent f and f' with their embeddings into **Stoch**: the Dirac Markov kernels $\delta_f, \delta_{f'}$.

If we compose δ_f and μ in **Stoch**, we form a new probability measure $(\delta_f \circ \mu) : 1 \times \mathcal{B}(\mathbb{R}) \to [0,1]$, which is the pushforward measure of $f_*\mu$. This gives us a hint of how we can reason about the independence or dependence of random variables in **Stoch**.

First, consider the probability measure $(\delta_f \circ \mu) \otimes (\delta_{f'} \circ \mu) : \mathcal{B}(\mathbb{R} \times \mathbb{R}) \to [0,1]$. We write this measure as

$$\left[(\delta_f \circ \mu) \otimes (\delta_{f'} \circ \mu)\right](\sigma \otimes \sigma') = \left[\int_{x\in\mathbb{R}} \delta_{f(x)}(\sigma) d\mu\right]\left[\int_{x\in\mathbb{R}} \delta_{f'(x)}(\sigma') d\mu\right] = f_*\mu(\sigma) * f'_*\mu(\sigma')$$

This is simply the product measure of the probability measures $(\delta_f \circ \mu)$ and $(\delta_{f'} \circ \mu)$. It is completely determined by the marginal distributions of f and f' over the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$, and it is agnostic to the independence or dependence structure of f and f'. The reason for this is that the measure μ is essentially "copied", and the random variables f and f' are not actually compared over the same probability space.

In contrast, consider instead the probability measure $(f \otimes f') \circ cp \circ \mu$, where $cp : \mathbb{R} \to \mathbb{R} \otimes \mathbb{R}$ is the comonoidal copy map at $\mathbb{R}$ in **Stoch** [Fri19]. We write this measure as

$$\left[(f \otimes f') \circ cp \circ \mu\right](\sigma \otimes \sigma') = \left[\int_{x\in\mathbb{R}} \delta_{f(x)}(\sigma)\delta_{f'(x)}(\sigma') d\mu\right] = (f \otimes f')_*\mu(\sigma \otimes \sigma')$$

This is the pushforward of $f \otimes f'$ along μ. That is, it is the probability measure associated with the joint distribution of the random variables f and f'.

Therefore, the random variables f and f' are independent over the probability space $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$ if and only if the probability measures $(\delta_f \circ \mu) \otimes (\delta_{f'} \circ \mu)$ and $(f \otimes f') \circ cp \circ \mu$ are equal.

2 $\mathbf{CoKl}_\Omega$

In [FST17] the authors describe the categories **Euc** and **Para** of Euclidian spaces and differentiable maps between them, and then relate these categories to optimization problems associated with Machine Learning algorithms. However, in practice these optimization problems are defined by first viewing a model as a random variable and selecting parameters to optimize the likelihood of observed data being drawn from the distribution of this random variable. A natural question is therefore whether it is possible to replace the deterministic maps in **Para** with statistical models.

Our first step will be to replace the morphisms in **Euc** with stochastic processes, or indexed families of random variables. Let's start with the following definition:

Definition 2.1. *For some category* $\mathbf{C}$ *and object* Ω *in* $\mathbf{C}$, $\mathbf{CoKl}_\Omega(\mathbf{C})$ *is the co-Kleisli category of* $\mathbf{C}$ *under the co-monad* $(\Omega \times -)$

Now say Ω is a Euclidean space and we have a probability space $(\Omega, \mathcal{B}(\Omega), \mu)$. Then we can consider $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$ to be a category of stochastic processes over the probability space $(\Omega, \mathcal{B}(\Omega), \mu)$. This category has Euclidean spaces as objects, and the morphisms between $\mathbb{R}^a$ and $\mathbb{R}^b$ are Borel-measurable functions of the form $f : \Omega \times \mathbb{R}^a \to \mathbb{R}^b$. Note that $\mathbf{CoKl}_\Omega(\mathbf{EucMeas})$ is the subcategory of $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$ restricted to differentiable morphisms.

In $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$, the composition of $f : \Omega \times \mathbb{R}^a \to \mathbb{R}^b$ and $f' : \Omega \times \mathbb{R}^b \to \mathbb{R}^c$ is:

$$(f' \circ f) : \Omega \times \mathbb{R}^a \to \mathbb{R}^c$$
$$(f' \circ f)(\omega, x_a) = f'(\omega, f(\omega, x_a))$$

And the tensor of $f : \Omega \times \mathbb{R}^a \to \mathbb{R}^b$ and $f' : \Omega \times \mathbb{R}^c \to \mathbb{R}^d$ is:

$$(f' \otimes f) : \Omega \times \mathbb{R}^a \times \mathbb{R}^c \to \mathbb{R}^b \times \mathbb{R}^d$$
$$(f' \otimes f)(\omega, x_a \oplus x_c) = f(\omega, x_a) \oplus f'(\omega, x_c)$$

One important thing to note is that ω is reused when we compose or tensor f and f'. That is, there is a "single source of randomness" in $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$.

Definition 2.2. *For any $\omega \in \Omega$, the **realization functor** $R_\omega : \mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R}) \to \mathbf{Meas}_\mathbb{R}$ associated with ω is an identity-on-objects monoidal functor that maps the stochastic process $f : \Omega \times \mathbb{R}^a \to \mathbb{R}^b$ to the function $f(\omega, \_) : \mathbb{R}^a \to \mathbb{R}^b$ in $\mathbf{Meas}_\mathbb{R}$.*

Now consider the Borel-measurable function $r : \Omega \to \mathbf{Cat}[\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R}), \mathbf{Meas}_\mathbb{R}]$ that maps $\omega \in \Omega$ to R_ω. The pushforward $r_*\mu$ is therefore a probability measure over the space of functors from $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$ to $\mathbf{Meas}_\mathbb{R}$.

Example 2.1. *A **Levy Process** is a one-dimensional stochastic process $f : \Omega \times \mathbb{R} \to \mathbb{R}$ such that:*

- *For $t_d > t_c > t_b > t_a \in \mathbb{R}$, the random variables $f(\_, t_b) - f(\_, t_a)$ and $f(\_, t_d) - f(\_, t_c)$ are independent.*
- *For $t_b > t_a \in \mathbb{R}$, the random variables $f(\_, t_b) - f(\_, t_a)$ and $f(\_, t_{b-a})$ have the same distribution.*
- *For any $\omega \in \Omega$ the function $f(\omega, \_)$ is continuous.*

A subordinator is a non-decreasing Levy Process. That is, for any fixed $\omega \in \Omega$ the function $f(\omega, \_)$ is non-decreasing. Now say f and g are subordinators. By [Lal07] we have that $g \circ f$ is a Levy Process. Since both f and g are non-decreasing, for $t_2 > t_1$ we have for any fixed $\omega \in \Omega$ that

$$g(\omega, f(\omega, t_2)) > g(\omega, f(\omega, t_1))$$

Therefore, $g \circ f$ is a subordinator as well. Since the identity arrow on $\mathbb{R}$ is a subordinator, we get a monoid at $\mathbb{R}$.

2.1 Random Variables and Transformations

Consider the category of random variables and transformations. The objects in this category are pairs $(\mathbb{R}^a, f)$ where f is a $\mathbb{R}^a$-valued random variable over the probability space $(\Omega, \mathcal{B}(\Omega), \mu)$, and the morphisms between the random variables f and f' are random variable transformations, or functions h such that $h(f(\omega)) = f'(\omega)$.

Note that random variable transformations are distinct from measure-preserving transformations, which exist between random variables with the same distribution. Even if $f'(\omega) = h(f(\omega))$, it is possible that the distributions of f and f' are distinct. Similarly, even if f and f' have identical distributions, there will not be any random variable transformation between them if f' is not a function of f (e.g. if f and f' are independent and non-degenerate).

This category is a subcategory of the coslice category $1/\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$. A random variable transformation is simply a $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$ arrow that ignores its first argument.

2.2 Stochastic Processes in Stoch

We can define a mapping $\mathcal{E} : \mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R}) \to \mathbf{Stoch}$ that is similar to the embedding functor from **Meas** into **Stoch** that we introduced in section 1. This mapping acts as the identity on objects and maps the stochastic process $f : \Omega \times \mathbb{R}^a \to \mathbb{R}^b$ to the Markov Kernel:

$$\mathcal{E}f : (\Omega \otimes \mathbb{R}^a, \mathcal{B}(\Omega) \otimes \mathcal{B}(\mathbb{R}^a)) \to (\Omega \otimes \mathbb{R}^b, \mathcal{B}(\Omega) \otimes \mathcal{B}(\mathbb{R}^b))$$
$$\mathcal{E}f_{(\omega,x_a)}(\sigma \otimes \sigma_a) = \delta_{(\omega,f(\omega,x_a))}(\sigma \otimes \sigma_b)$$

That is, we can view $\mathcal{E}f_{(\omega,x_a)}$ as a Markov kernel representation of the stochastic process f.

Proposition 1. *$\mathcal{E}$ is a functor*

Proof in Appendix

2.3 Independence and Dependence in $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$

Since all of the stochastic processes in $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$ are defined over the same probability space and both composition and tensor use a "single source of randomness", there is a major difference between how $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$ and **Stoch** represent independence and dependence.

Namely, every random variable and stochastic process in $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$ is defined over the same probability space $(\Omega, \mathcal{B}(\Omega), \mu)$. Therefore, given the arrows $f : \Omega \times \mathbb{R}^a \to \mathbb{R}^b$ and $f' : \Omega \times \mathbb{R}^c \to \mathbb{R}^d$ in $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$ and the vector $x_a \in \mathbb{R}^a, x_c \in \mathbb{R}^c$, the random variables $f(\_, x_a)$ and $f'(\_, x_c)$ may be either dependent or independent.

In order to see how this differs from the situation in **Stoch**, let's recall that the pushforward of the stochastic process $f : \Omega \times \mathbb{R}^a \to \mathbb{R}^b$ along μ is a mapping from $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$ to **Stoch** such that:

$$f_*\mu : \mathbb{R}^a \times \mathcal{B}(\mathbb{R}^b) \to [0, 1]$$
$$f_*\mu_{x_a}(\sigma_b) = f(\_, x_a)_*\mu(\sigma_b) = \int_{\omega\in\Omega} \delta_{f(\omega,x_a)}(\sigma_b) d\mu$$

However, this mapping does not form a functor. We see that for $f : \Omega \times \mathbb{R}^a \to \mathbb{R}^b$, $f' : \Omega \times \mathbb{R}^b \to \mathbb{R}^c$, and $x_a \in \mathbb{R}^a$:

$$\left(f' \circ f\right)_* \mu_{x_a}(\_) = \int_{x_b\in\mathbb{R}^b} \int_{\omega\in\Omega} \delta_{f'(\omega,x_b)}(\_) \ d\delta_{f(\omega,x_a)}(x_b) \ d\mu$$

Whereas:

$$\left[f'_*\mu \circ f_*\mu\right]_{x_a}(\_) = \int_{x_b\in\mathbb{R}^b} \left(\int_{\omega\in\Omega} \delta_{f'(\omega,x_b)}(\_) d\mu\right) \left(\int_{\omega\in\Omega} d\delta_{f(\omega,x_a)}(x_b) d\mu\right)$$

These probability measures are not necessarily equivalent if the random variables $f'(\_, x_b), x_b \in \mathbb{R}^b$ are not independent of the random variable $f(\_, x_a)$.

The reason for this mismatch comes down to the nature of tensor and composition in **Stoch**. We can think of a Markov Kernel in **Stoch** as being implicitly associated with mutually independent random variables. The fact that the composition or tensor of arrows in the category involves the creation of a new probability space enforces this independence property.

We can slightly modify $\mathbf{CoKl}_\Omega(\mathbf{Meas}_\mathbb{R})$ to define a new category of stochastic processes that exhibit this independence behavior.

3 $\mathbf{Para}_{\Omega^*}(\mathbf{Meas})$

3.1 A generalization of Para

We will begin by slightly generalizing the **Para** construction from [Gav19], which is itself a generalization of **Para** from [FST17].

Consider a small symmetric monoidal category $\mathbf{C}$ and a category $\mathbf{D}$ such that there exists a faithful identity-on-objects monoidal functor $\iota : \mathbf{D} \hookrightarrow \mathbf{C}$. That is, we can think of $\mathbf{D}$ as a subcategory of $\mathbf{C}$. Then write $(- \otimes A) \circ \iota : \mathbf{D} \hookrightarrow \mathbf{C}$ to denote the functor that sends the object A' in $\mathbf{D}$ to $A' \otimes A$ in $\mathbf{C}$ and write $c_B : \mathbf{D} \to \mathbf{C}$ for the constant functor that sends all objects in $\mathbf{D}$ to B.

Definition 3.1. *For a small symmetric monoidal category $\mathbf{C}$ and a category $\mathbf{D}$ such that there exists a faithful identity-on-objects monoidal functor $\iota : \mathbf{D} \hookrightarrow \mathbf{C}$, $\mathbf{Para}_\mathbf{D}(\mathbf{C})$ is the category with the same objects as $\mathbf{C}$ and the homset $\mathbf{Para}_\mathbf{D}(\mathbf{C})[A, B]$ equal to the set of objects in the comma category $(- \otimes A) \circ \iota \downarrow c_B$ such that composition and tensor work the same way as in [Gav19].*

That is, the morphisms between A and B in $\mathbf{Para}_\mathbf{D}(\mathbf{C})$ are morphisms of the form $P \otimes A \to B$ in $\mathbf{C}$, where P is an object in $\mathbf{D}$. The composition of the arrows $f : P \otimes A \to B$ and $f : Q \otimes B \to C$ is:

$$P \otimes Q \otimes A \xrightarrow{\simeq} P \otimes (Q \otimes A) \xrightarrow{id_P \otimes f} P \otimes B \xrightarrow{g} C$$

And the tensor of arrows $f : P \otimes A \to B$ and $f : Q \otimes C \to D$ is:

$$P \otimes Q \otimes A \otimes B \xrightarrow{id_P \otimes swap_{(Q,A)} \otimes id_B} P \otimes A \otimes Q \otimes B \xrightarrow{\simeq} (P \otimes A) \otimes (Q \otimes B) \xrightarrow{f \otimes f'} C \otimes D$$

Note that for any object A, the identity arrow $id_A : 1 \otimes A \to A$ in $\mathbf{Para}_\mathbf{D}(\mathbf{C})$ is $\_ \otimes id'_A$, where $id'_A : A \to A$ is the identity in $\mathbf{C}$. Note also that the monoidal unit, 1, is the same in $\mathbf{C}$ and $\mathbf{Para}_\mathbf{D}(\mathbf{C})$. Clearly, $\mathbf{Para}_\mathbf{C}(\mathbf{C})$ is equivalent to $\mathbf{Para}(\mathbf{C})$ from [Gav19].

Proposition 2. *Say $\mathbf{C}$ and $\mathbf{C}'$ are small symmetric monoidal categories with a monoidal functor $F : \mathbf{C} \to \mathbf{C}'$ between them. Say $\mathbf{D}$ is symmetric monoidal category such that there exists a faithful monoidal functor $\iota : \mathbf{D} \hookrightarrow \mathbf{C}$ and that the image of $F \circ \iota$ is a subcategory $\mathbf{D}'$ of $\mathbf{C}'$. Then the map $F' : \mathbf{Para}_\mathbf{D}(\mathbf{C}) \to \mathbf{Para}_{\mathbf{D}'}(\mathbf{C}')$ that applies the same actions on objects and arrows as F is a monoidal functor.*

Proof in appendix

Proposition 3. *Say* $\mathbf{D}, \mathbf{D}', \mathbf{C}$ *are small symmetric monoidal categories and* $\iota : \mathbf{D} \hookrightarrow \mathbf{C}, \iota' : \mathbf{D}' \hookrightarrow \mathbf{C}$ *are faithful identity-on-objects functors. Then* $\mathbf{Para}_{\mathbf{D}}(\mathbf{Para}_{\mathbf{D}'}(\mathbf{C})) \simeq \mathbf{Para}_{\mathbf{D}'}(\mathbf{Para}_{\mathbf{D}}(\mathbf{C}))$

Proof in appendix

3.2 Lawvere Parameterization

Definition 3.2. *Say* $\mathbf{C}$ *is a cartesian monoidal category,* Ω^* *is a Lawvere theory with generating object* Ω*, and* ι *is a faithful identity-on-objects functor* $\iota : \Omega^* \hookrightarrow \mathbf{C}$*. Then* $\mathbf{Para}_{\Omega^*}(\mathbf{C})$ *is a* ***Lawvere parameterization*** *of* $\mathbf{C}$*.*

Let's first note that if $\mathbf{C}$ is an additive category, then by [AAK] and section 7 of [Tab11] $\mathbf{Para}_{\Omega^*}(\mathbf{C})$ is the category $(\mathbf{C}^{op}/\Omega \otimes -)^{op}$ where $\mathbf{C}^{op}/\Omega \otimes -$ is the Orbit category of the automorphism $(\Omega \otimes -)$. This implies that the opposite projection functor $\pi^{op} : \mathbf{C} \to \mathbf{Para}_{\Omega^*}(\mathbf{C})$ is symmetric monoidal.

Now for any cartesian monoidal category $\mathbf{C}$ with a Lawvere parameterization $\mathbf{Para}_{\Omega^*}(\mathbf{C})$ we can define a mapping $Copy_\Omega$ to the co-Kleisli category of $(\Omega \otimes -)$ on $\mathbf{C}$. This mapping acts as the identity-on-objects and sends the arrow $f : \Omega^n \otimes A \to B$ in $\mathbf{Para}_{\Omega^*}(\mathbf{C})$ to the following arrow in the co-Kleisli category of $(\Omega \otimes -)$ on $\mathbf{C}$:

$$f \circ (cp_\Omega^{n-1} \otimes id_A^{\mathbf{C}}) : \Omega \otimes A \to B$$

For clarity, $id_A^{\mathbf{C}}$ is the identity arrow on A in $\mathbf{C}$, cp_Ω is the copy map $\Omega \to \Omega \otimes \Omega$ in $\mathbf{C}$, $cp_\Omega^{n-1} : \Omega \to \Omega \otimes \Omega \otimes \cdots \otimes \Omega$ is the repeated application of this map $n-1$ times and cp_Ω^0 is the identity on Ω.

Proposition 4. *$Copy_\Omega$ is a full identity-on-objects monoidal functor*

Proof in Appendix

3.3 A Category of Parameterized Measurable Maps

Say we have a Lawvere theory Ω^* with generating object $(\Omega, \mathcal{B}(\Omega))$, a faithful identity-on-objects functor $\iota : \Omega^* \hookrightarrow \mathbf{Meas}$, and a probability space $(\Omega, \mathcal{B}(\Omega), \mu)$. Then for any $(\Omega^n, \mathcal{B}(\Omega^n)) \in \Omega^*$, we can create the probability space $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$ where μ^n is the product measure $\mu^n(\omega_1, \omega_2, \cdots, \omega_n) = \mu(\omega_1) * \mu(\omega_2) * \cdots * \mu(\omega_n)$.

Now consider the Lawvere parameterization $\mathbf{Para}_{\Omega^*}(\mathbf{Meas})$. Intuitively, $\mathbf{Para}_{\Omega^*}(\mathbf{Meas})$ allows us to reason about probabilistic relationships in terms of measurable functions rather than probability measures. We can make this probabilistic intuition more formal.

First, $\mathbf{Para}_{\Omega^*}(\mathbf{Meas})$ behaves similarly to a category of Markov Kernels and we can show the following:

Proposition 5. $\mathbf{Para}_{\Omega^*}(\mathbf{Meas})$ *is a Markov Category from [Fri19]*

Proof in Appendix

Next, we can interpret the arrows in $\mathbf{Para}_{\Omega^*}(\mathbf{Meas})[A, \mathbb{R}]$ as stochastic processes with an index set of A. That is, given some arrow $f : \Omega^n \otimes A \to \mathbb{R}$ and $x_a \in A$, the measurable function $f(\_, x_a)$ is a random variable over the probability space $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$. The pushforward of this random variable $f(\_, x_a)_*\mu(-)$ is then a probability measure over the space $(\mathbb{R}^b, \mathcal{B}(\mathbb{R}^b))$.

In general, we can extend this pushforward procedure to define a mapping between parameterized famililies of measurable maps and Markov Kernels. Given some $f : \Omega^n \otimes A \to B$ we can define

$$Push_\mu f : A \otimes \Sigma_B \to [0, 1]$$
$$Push_\mu(f(x_a, \sigma_b)) = f(\_, x_a)_*\mu^n(\sigma_b) = \int_{\omega_n \in \Omega^n} \delta_{f(\omega, x_a)}(\sigma_b) d\mu^n$$

Proposition 6. *The mapping $Push_\mu$ that takes a parameterized family of measurable maps $f : \Omega^n \otimes A \to B$ to the Markov Kernel $f_*\mu^n(-)$ is an identity-on-objects monoidal functor from $\mathbf{Para}_{\Omega^*}(\mathbf{Meas})$ to* **Stoch**

Proof in appendix.

Now say Ω is a Euclidian space and consider the Lawvere parameterization $\mathbf{Para}_{\Omega^*}(\mathbf{EucMeas})$. By Proposition 4, we have an identity-on-objects functor, $Copy_\Omega$, from $\mathbf{Para}_{\Omega^*}(\mathbf{EucMeas})$ to $\mathbf{CoKl}_\Omega(\mathbf{EucMeas})$.

Let's drill deeper into this relationship. We can view an arrow of the form $f : \Omega^n \otimes \mathbb{R}^a \to \mathbb{R}^b$ in $\mathbf{Para}_{\Omega^*}(\mathbf{EucMeas})$ as a stochastic process over $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$. However, unlike in $\mathbf{CoKl}_\Omega(\mathbf{EucMeas})$, if we compose or tensor f with another arrow in $\mathbf{Para}_{\Omega^*}(\mathbf{EucMeas})$, we do not get another stochastic process over $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$. Instead, we get a stochastic process over some other probability space. Intuitively, we can think of the stochastic processes in $\mathbf{Para}_{\Omega^*}(\mathbf{EucMeas})$ as being defined over different, non-interacting probability spaces.

4 Parameterized Statistical Models and Likelihood

4.1 $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas}))$

Given some μ, any stochastic process $f : \Omega^n \otimes \mathbb{R}^a \to \mathbb{R}^b$ in $\mathbf{Para}_{\Omega^*}(\mathbf{EucMeas})$ and the corresponding Markov kernel in **Stoch** define a stochastic relationship between values in $\mathbb{R}^a$ and $\mathbb{R}^b$. A parametric statistical model is a parametrized family of such relationships. For example, consider a univariate linear regression model l

$$l : \Omega^n \otimes \mathbb{R}^3 \otimes \mathbb{R} \to \mathbb{R}$$
$$l(\omega_n, [a, b, s^2], x) = x * a + b + f_{\mathcal{N}(0,s^2)}(\omega_n)$$

Where $f_{\mathcal{N}(0,s^2)}$ is a normally distributed random variable with variance s^2. Any value $[a, b, s] \in \mathbb{R}^3$ defines the stochastic process, or $\mathbf{Para}_{\Omega^*}(\mathbf{EucMeas})$ arrow

$$l(\_, [a, b, s], \_) : \Omega^n \otimes \mathbb{R} \to \mathbb{R}$$

For any input value $x \in \mathbb{R}$, the function $l(\_, [a, b, s], x)$ is then a random variable over the probability space $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$. Like with any ordinary univariate linear regression model, this random variable is normally distributed on the real line.

We can define a category of such models by applying the **Para** construction from [Gav19] and [FST17] to $\mathbf{Para}_{\Omega^*}(\mathbf{EucMeas})$. For convenience, we will work in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas}))$. This is possible because by proposition 3 we have:

$$\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})) \simeq \mathbf{Para}(\mathbf{Para}_{\Omega^*}(\mathbf{EucMeas}))$$

Note that the statistical models that appear as arrows in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas}))$ are both discriminative and frequentist. That is, we fix both the parameters and input values and only probabilistically model the output value. In particular, the homset $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas}))[\mathbb{R}, \mathbb{R}]$ includes our linear regression model above.

4.2 $\mathbf{Para}(Push_\mu)$

By proposition 2 we have that there is a functor $\mathbf{Para}(Push_\mu) : \mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})) \to \mathbf{Para}(\mathbf{Stoch})$ that has the same actions on objects and morphisms as $Push_\mu$.

For some arrow $f : \Omega^n \otimes \mathbb{R}^p \otimes \mathbb{R}^a \to \mathbb{R}^b$ in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas}))$, we can think of the Markov kernel $\mathbf{Para}(Push_\mu)f$ as a function that accepts an input value $x_a \in \mathbb{R}^a$ and a parameter vector $x_p \in \mathbb{R}^p$ and returns the conditional probability measure defined by the parametric statistical model f. For example, in our univariate linear regression model l, the probability measure $\mathbf{Para}(Push_\mu)l([a, b, s], x, \_) : \mathcal{B}(\mathbb{R}) \to [0, 1]$ is defined as:

$$\mathbf{Para}(Push_\mu)l_{[a,b,s],x}(\sigma) = \int_{y \in \sigma} \frac{1}{s\sqrt{2\pi}} e^{-\frac{(y-(a*x+b))^2}{2s^2}} dy$$

4.3 CondLikelihood

Abstractly, a conditional likelihood is a function that accepts a data point (x_a, x_b) and a parameter value x_p and returns some notion of the "likelihood" of observing the output value x_b given the parameter value x_p and the input value x_a. We can define a monoidal category **CondLikelihood** of these abstract conditional likelihood functions.

Definition 4.1. *In the monoidal category* **CondLikelihood**, *Euclidian spaces are objects and the morphisms between $\mathbb{R}^a$ and $\mathbb{R}^b$ are either identity arrows or Borel-measurable functions of the form $L : \mathbb{R}^p \otimes \mathbb{R}^a \otimes \mathbb{R}^b \to \mathbb{R}$. The composition of $L : \mathbb{R}^p \otimes \mathbb{R}^a \otimes \mathbb{R}^b \to \mathbb{R}, L' : \mathbb{R}^q \otimes \mathbb{R}^b \otimes \mathbb{R}^c \to \mathbb{R}$ is:*

$$(L' \circ L) : \mathbb{R}^{q+p} \otimes \mathbb{R}^a \otimes \mathbb{R}^c \to \mathbb{R}$$

$$(L' \circ L)(x_{qp}, x_a, x_c) = \int_{x_b \in \mathbb{R}^b} L'(x_q, x_b, x_c) * L(x_p, x_a, x_b) dx_b$$

and the tensor of $L : \mathbb{R}^p \otimes \mathbb{R}^a \otimes \mathbb{R}^b \to \mathbb{R}, L' : \mathbb{R}^q \otimes \mathbb{R}^c \otimes \mathbb{R}^d \to \mathbb{R}$ is:

$$(L' \otimes L) : \mathbb{R}^{q+p} \otimes \mathbb{R}^{c+a} \otimes \mathbb{R}^{d+b} \to \mathbb{R}$$

$$(L' \otimes L)(x_{qp}, x_{ca}, x_{db}) = L'(x_q, x_c, x_d) * L(x_p, x_a, x_b)$$

We can think of the identity arrows in **CondLikelihood** as instantiations of the Dirac delta "function" δ. That is, δ is the operator such that:

$$(\delta \circ f)(x_{qp}, x_a, x_b) = \int_{x'_b \in \mathbb{R}^b} \delta(\_, x_b, x'_b) * f(x_p, x_a, x'_b) dx_b = f(x_p, x_a, x_b)$$

We can formally characterize δ as a generalized function [Wei04], but we will not go this route for simplicity.

In practice, conditional likelihoods are associated with statistical models. The value of the conditional likelihood function associated with the statistical model $f : \Omega^n \otimes \mathbb{R}^p \otimes \mathbb{R}^a \to \mathbb{R}^b$ over the probability space $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$ at x_p, x_a, x_b is the value of the Radon-Nikodym derivative of the pushforward of the random variable $f(\_, x_p, x_a)$ evaluated at x_p (if it exists). That is, the likelihood function for the parametric statistical model $f : \Omega^n \otimes \mathbb{R}^p \otimes \mathbb{R}^a \to \mathbb{R}^b$ is:

$$L(x_p, x_a, x_b) = \frac{d\ f(\_, x_p, x_a)_* \mu}{d\lambda^b}(x_b)$$

Note that λ^b is the Lebesgue measure over $\mathbb{R}^b$. In our univariate linear regression model that we introduced in section 4.1, the likelihood function of l is:

$$L([a, b, s], x, y) = \frac{1}{s\sqrt{2\pi}} e^{-\frac{(y-(a*x+b))^2}{2s^2}}$$

Next, let us define $\mathcal{R}_\mu$ to be the set of $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas}))$ arrows $f : \Omega^n \otimes \mathbb{R}^p \otimes \mathbb{R}^a \to \mathbb{R}^b$ such that for any $x_a \in \mathbb{R}^a, x_p \in \mathbb{R}^p$, the measurable function $\frac{d\ f(\_, x_a, x_p)_* \mu}{\lambda^b}$ exists, where λ^b is the Lebesgue measure over $\mathbb{R}^b$. We can therefore define $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$ to be the monoidal category generated by $\mathcal{R}_\mu$ and the identity arrows in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas}))$. Note that $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$ is trivially a subcategory of $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas}))$.

Proposition 7. *Any non-identity arrow* $f : \Omega^n \otimes \mathbb{R}^p \otimes \mathbb{R}^a \to \mathbb{R}^b$ *in* $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$ *is in* $\mathcal{R}_\mu$.

Proof in Appendix

We can now define the mapping $\mathcal{RN}_\mu : \mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu} \to \mathbf{CondLikelihood}$ that acts as the identity on objects, sends the identity arrows in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$ to the identity arrows in **CondLikelihood** by construction and sends any non-identity morphism $f : \Omega^n \otimes \mathbb{R}^p \otimes \mathbb{R}^a \to \mathbb{R}^b$ to the function

$$\mathcal{RN}_\mu f : \mathbb{R}^p \otimes \mathbb{R}^a \otimes \mathbb{R}^b \to \mathbb{R}$$
$$\mathcal{RN}_\mu f(x_p, x_a, x_b) = \frac{df(\_, x_p, x_a)_* \mu}{\lambda^b}(x_b)$$

Note that Proposition 7 implies that this function exists.

Proposition 8. $\mathcal{RN}_\mu$ *is a monoidal functor.*

Proof in Appendix

4.4 $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{N}_\mu}$

Definition 4.2. *A* ***Gaussian-preserving transformation*** $T : \mathbb{R}^a \to \mathbb{R}^b$ *is a function such that for any random variable f that is distributed according to a multivariate Gaussian distribution over $\mathbb{R}^a$, the random variable $T(f(\_))$ has a multivariate Gaussian distribution over $\mathbb{R}^b$ and we have:*

$$\int_{\omega \in \Omega} T(f(\omega))d\mu = T\left(\int_{\omega \in \Omega} f(\omega)d\mu\right)$$

For example, any linear function is Gaussian-preserving.

Now for some probability space $(\Omega, \mathcal{B}(\Omega), \mu)$, we can construct a set of $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas}))$-arrows $\mathcal{N}_\mu$ such that if f has the signature $f : \Omega^n \otimes \mathbb{R}^a \otimes \mathbb{R}^p \to \mathbb{R}^b$, then we can write f in the form:

$$f(\omega_n, x_a, x_p) = T(x_a, x_p) + G(\omega_n)$$

Where $T : \mathbb{R}^{a+p} \to \mathbb{R}^b$ is a Gaussian-preserving transformation, and $G : \Omega^n \to \mathbb{R}^b$ is a multivariate normal random variable over the probability space $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$. Note that this includes our univariate linear regression model l, as well as the identity arrow, since constant distributions are multivariate normal with variance 0.

We can therefore define $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{N}_\mu}$ to be the monoidal category generated by $\mathcal{N}_\mu$. Note that $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{N}_\mu}$ is a subcategory of $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$.

Proposition 9. *Given any arrow $f : \Omega^n \otimes \mathbb{R}^a \otimes \mathbb{R}^p \to \mathbb{R}^b$ in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{N}_\mu}$ and $x_a \in \mathbb{R}^a, x_p \in \mathbb{R}^p$, $f(\_, x_a, x_p)$ is a multivariate normal random variable with respect to the probability space $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$.*

Proof in Appendix.

Proposition 10. *For any two arrows $f' : \Omega^m \otimes \mathbb{R}^p \otimes \mathbb{R}^a \to \mathbb{R}^b$ and $f : \Omega^n \otimes \mathbb{R}^q \otimes \mathbb{R}^b \to \mathbb{R}^c$ in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{N}_\mu}$, we have that:*

$$\int_{\omega_{mn} \in \Omega^{m+n}} f'(\omega_m, f(\omega_n, x_a, x_p), x_q)d\mu^{m+n} = \int_{\omega_m \in \Omega^m} f'\left(\omega_m, \int_{\omega_n \in \Omega^n} f(\omega_n, x_a, x_p)d\mu^n, x_q\right) d\mu^m$$

Proof in Appendix.

5 Optimization

When we "train" a parameterized statistical model $f : \Omega^n \times \mathbb{R}^p \times \mathbb{R}^a \to \mathbb{R}^b$ in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$ using the **maximum likelihood** method, we choose some $x_p \in \mathbb{R}^p$ that maximizes a function derived from $\mathcal{R}\mathcal{N}_\mu f$. Our trained model is then the $\mathbf{Para}_{\Omega^*}(\mathbf{EucMeas})$-arrow $f(\_, x_p, \_) : \Omega^n \times \mathbb{R}^a \to \mathbb{R}^b$.

For example, say we have the arrow $f : \Omega^n \times \mathbb{R}^p \times \mathbb{R}^a \to \mathbb{R}^b$ in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$ and a probability measure τ over $(\mathbb{R}^a \times \mathbb{R}^b, \mathcal{B}(\mathbb{R}^a \times \mathbb{R}^b))$. The **maximum expected**

likelihood estimator for f with respect to τ is the vector $x_p \in \mathbb{R}^p$ that maximizes the function

$$L(x_p) = \int_{(x_a,x_b)\in\mathbb{R}^a\times\mathbb{R}^b} \frac{df(\_, x_p, x_a)_*\mu^n}{\lambda^b}(x_b)d\tau$$

That is, the maximum expected likelihood estimator for f with respect to τ is the vector x_p that maximizes the expected value of $\frac{df(\_,x_p,x_a)_*\mu^n}{\lambda^b}(x_b)$ over τ. For each $x_a \in \mathbb{R}^a$, we can think of this estimator as driving the distributions $f(\_, x_p, x_a)_*\mu^n : \mathcal{B}(\mathbb{R}^b) \to [0, 1]$ towards the distribution that τ defines over $\mathbb{R}^b$ conditioned on x_a. More specifically, x_p minimizes the weighted sum of the KL-divergences between these distributions, where the weight of the distribution determined by x_a is determined by the likelihood of x_a under the marginal distribution that τ defines over $\mathbb{R}^a$ [Mur12].

Now say instead we have some dataset of samples $S = (x_{a_1}, x_{b_1}), (x_{a_2}, x_{b_2}), ..., (x_{a_n}, x_{b_n})$ and the arrow $f : \Omega^n \times \mathbb{R}^p \times \mathbb{R}^a \to \mathbb{R}^b$ in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$. The **maximum likelihood estimator** for f with respect to this dataset is the vector $x_p \in \mathbb{R}^p$ that maximizes the function

$$L(x_p) = \sum_i^n log \frac{df(\_, x_p, x_a)_*\mu^n}{\lambda^b}(x_b)$$

Next, for any $j \leq b$, the jth component of f is the function $f_j : \Omega^n \times \mathbb{R}^p \times \mathbb{R}^a \to \mathbb{R}$ and the marginal likelihood of this component for some sample (x_a, x_b) is:

$$L_j(x_p) = \frac{df(\_, x_p, x_a)_{j*}\mu^n}{\lambda}(x_{b_j})$$

The maximum log-marginal likelihood estimator for f with respect to this dataset is then the vector $x_p \in \mathbb{R}^p$ that maximizes the function:

$$L_m(x_p) = \sum_i^n \sum_j^m log \frac{df(\_, x_p, x_a)_{j*}\mu^n}{\lambda}(x_{b_j})$$

Note that $L_m(x_p) = L(x_p)$ when the random variables $f(\_, x_p, x_a)_j$ are mutually independent.

5.1 Marginal Likelihood Factorization Category

Let's write $E_{\mu^n}[f(\_, x_p, x_a)_j]$ for the expectation of the real-valued random variable $f(\_, x_p, x_a)_j$ over the probability space $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$ and define $f^0(\_, x_p, x_a)_j$ to be the centered version of $f(\_, x_p, x_a)_j$:

$$E_{\mu^n}[f(\_, x_p, x_a)_j] = \int_{\omega\in\Omega} f(\omega, x_p, x_a)_j \; d\mu^n$$

$$f^0(\_, x_p, x_a)_j = f(\_, x_p, x_a)_j - E_{\mu^n}[f(\_, x_p, x_a)_j]$$

Definition 5.1. *A subcategory C of* $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$ *is a* ***Marginal Likelihood Factorization Category*** *if it satisfies the following properties*

First, there exist some measurable function $\epsilon_1 : (\Omega \to \mathbb{R}) \to \mathbb{R}$, non-negative measurable function $\epsilon_2 : (\Omega \to \mathbb{R}) \to \mathbb{R}$, and differentiable function with invertible derivative $er : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ such that we have the following for any f in D:

$$log \frac{df(\_, x_p, x_a)_{j*}\mu^n}{\lambda}(x_{b_j}) = \epsilon_1(f^0(\_, x_p, x_a)_j) - \epsilon_2(f^0(\_, x_p, x_a)_j) * er\left(E_{\mu^n}[f(\_, x_p, x_a)_j], x_{b_j}\right)$$

*We will refer to er as the **marginal error function** of C.*

*Next, for any two arrows $f' : \Omega^m \times \mathbb{R}^q \times \mathbb{R}^b \to \mathbb{R}^c$ and $f : \Omega^n \times \mathbb{R}^p \times \mathbb{R}^a \to \mathbb{R}^b$ in C, we have the following, which we will refer to as the **Expectation Composition** condition:*

$$\int_{\omega_{mn} \in \Omega^{m+n}} f'(\omega_m, x_q, f(\omega_n, x_p, x_a))d\mu^{m+n} = \int_{\omega_m \in \Omega^m} f'\left(\omega_m, x_q, \int_{\omega_n \in \Omega^n} f(\omega_n, x_p, x_a)d\mu^n\right) d\mu^m$$

Proposition 11. $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{N}_\mu}$ *is a Marginal Likelihood Factorization Category*

First, let's note that the Expectation Composition condition is satisfied by Proposition 10. Next, consider some $f : \Omega^n \times \mathbb{R}^p \times \mathbb{R}^a \to \mathbb{R}^b$ in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{N}_\mu}$, and let's note that for any $x_p \in \mathbb{R}^p, x_a \in \mathbb{R}^a, j \leq b$, the random variable $f(\_, x_p, x_a)_j$ is univariate normal. Therefore, we can write the following, where s is the standard deviation of $f(\omega, x_p, x_a)_j$:

$$\begin{aligned} log\ \frac{df(\_, x_p, x_a)_{j*}\mu^n}{\lambda}(x_{b_j}) &= \\ log\ \frac{1}{s\sqrt{2\pi}} e^{-\left(\frac{x_{b_j} - E_{\mu^n}[f(\_, x_p, x_a)_j]}{4s}\right)^2} &= \\ -\frac{log(2\pi s^2)}{2} - \frac{1}{2s^2}\left(x_{b_j} - E_{\mu^n}[f(\_, x_p, x_a)_j]\right)^2 & \end{aligned}$$

Therefore $er(a, b) = (a - b)^2$

□

Now for any arrow $f : \Omega^n \times \mathbb{R}^p \times \mathbb{R}^a \to \mathbb{R}^b$ in a Marginal Likelihood Factorization Category C, let's note that $E_{\mu^n}[f(\_, x_p, x_a)_j]$ must be differentiable. We can therefore define a mapping $Exp_C : C \to \mathbf{Para}$ that acts as the identity on objects and sends the arrow $f : \Omega^n \times \mathbb{R}^p \times \mathbb{R}^a \to \mathbb{R}^b$ to $E_{\mu^n}[f(\_, x_p, x_a)_j]$. This mapping is functorial by the Expectation Composition condition.

This then implies that for any Marginal Likelihood Factorization Category we can define a functor $L_{er} \circ Exp_C : C \to Learn$, where L_{er} is the Backpropagation functor from [FST17] under the marginal error function er of C. For example, this functor sends arrows in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{N}_\mu}$ to learning algorithms that use mean square error to train.

6 Future Work

- In Section 4 we describe the maximum expected likelihood estimator for a $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$ arrow $f : \Omega^n \times \mathbb{R}^p \times \mathbb{R}^a \to \mathbb{R}^b$ and a probability measure τ over $(\mathbb{R}^a \times \mathbb{R}^b, \mathcal{B}(\mathbb{R}^a \times \mathbb{R}^b))$. A next step is to categorify this construction, potentially extending it into a functor from $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$ into **Euc**.

- It is unclear whether marginal likelihood factorization categories are the most general class of subcategories of $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$ for which the backpropagation functor is defined. A next step would be to relax the restrictions on these categories or prove that they are necessary.

References

[AAK] AAK. It looks so cokleisli, but it's not. what is it? MathOverflow. URL:https://mathoverflow.net/q/153550 (version: 2014-01-04).

[CS12] Jared Culbertson and Kirk Sturtz. A categorical foundation for Bayesian probability, 2012.

[Fon13] Brendan Fong. Causal theories: A categorical perspective on Bayesian networks. *arXiv preprint arXiv:1301.6201*, 2013.

[FR19] Tobias Fritz and Eigil Fjeldgren Rischel. The zero-one laws of Kolmogorov and Hewitt–Savage in categorical probability, 2019.

[Fra02] Uwe Franz. What is stochastic independence?, 2002.

[Fri19] Tobias Fritz. A synthetic approach to Markov kernels, conditional independence, and theorems on sufficient statistics. *arXiv preprint arXiv:1908.07021*, 2019.

[FST17] Brendan Fong, David I. Spivak, and Rémy Tuyéras. Backprop as functor: A compositional perspective on supervised learning, 2017.

[Gav19] Bruno Gavranovic. Compositional deep learning. *CoRR*, abs/1907.08292, 2019.

[Gir82] Michele Giry. A categorical approach to probability theory. In *Categorical aspects of topology and analysis*, pages 68–85. Springer, 1982.

[GLS16] Malte Gerhold, Stephanie Lachs, and Michael Schürmann. Categorial Lévy processes, 2016.

[HKSY17] Chris Heunen, Ohad Kammar, Sam Staton, and Hongseok Yang. A convenient category for higher-order probability theory. *CoRR*, abs/1701.02547, 2017.

[Lal07] STEVEN P Lalley. Lévy processes, stable processes, and subordinators. 2007.

[Law62] F.W. Lawvere. The category of probabilistic mappings. 1962.

[Mur12] Kevin Patrick Murphy. *Machine Learning: A Probabilistic Perspective*. MIT Press, 2012.

[Tab11] Goncalo Tabuada. Chow motives versus non-commutative motives, 2011.

[Wei04] Eric W Weisstein. Delta function. *delta*, 29:30, 2004.

[Wik19] Wikipedia contributors. Multivariate random variable — Wikipedia, the free encyclopedia, 2019. [Online; accessed 11-March-2020].

7 Appendix

7.1 Proof of Proposition 1

$$
\begin{aligned}
f : \Omega \otimes \mathbb{R}^a \to \mathbb{R}^b \qquad f' : \Omega \otimes \mathbb{R}^b \to \mathbb{R}^c \\
(\mathcal{E}f' \circ \mathcal{E}f) : (\Omega \otimes \mathbb{R}^a, \mathcal{B}(\Omega) \otimes \mathcal{B}(\mathbb{R}^a)) \to (\Omega \otimes \mathbb{R}^c, \mathcal{B}(\Omega) \otimes \mathcal{B}(\mathbb{R}^c)) \\
(\mathcal{E}f' \circ \mathcal{E}f)_{(\omega, x_a)}(\sigma \otimes \sigma_c) = \int_{(\omega', x_b) \in \Omega \otimes \mathbb{R}^b} \mathcal{E}f'_{(\omega', x_b)}(\sigma \otimes \sigma_c) d\mathcal{E}f_{(\omega, x_a)} = \\
\int_{(\omega', x_b) \in \Omega \otimes \mathbb{R}^b} \delta_{(\omega', f'(\omega', x_b))}(\sigma \otimes \sigma_c) d\delta_{(\omega, f(\omega, x_a))} = \\
\delta_{(\omega, f'(\omega, f(\omega, x_a)))}(\sigma \otimes \sigma_c) = \\
\delta_{(\omega, (f' \circ f)(\omega, x_a))}(\sigma \otimes \sigma_c) = \\
\mathcal{E}(f' \circ f)_{(\omega, x_a)}(\sigma \otimes \sigma_c)
\end{aligned}
$$

7.2 Proof of Proposition 2

Since F' trivially preserves the identity and monoidal unit, we need to show that F' respects composition and tensor.

Composition

Say $f : P \times A \to B, g : Q \times B \to C$ are arrows in $\mathbf{Para}_{\mathbf{D}}(\mathbf{C})$. Then they are also arrows in $\mathbf{C}$, and their $\mathbf{Para}_{\mathbf{D}}(\mathbf{C})$ composition $g \circ_{\mathbf{Para}_{\mathbf{D}}(\mathbf{C})} f$ is equal to $g \circ_C (id \otimes_C f)$. Therefore, $F'(g \circ_{\mathbf{Para}_{\mathbf{D}}(\mathbf{C})} f) = F'g \circ_{\mathbf{Para}_{\mathbf{D}'}(\mathbf{C}'))} F'f$ if the following diagram commutes:

$$
\begin{tikzcd}
FQ \otimes FP \otimes FA \arrow[r, "\mu_F"] \arrow[d, "Fid_Q \otimes Ff"] & F(Q \otimes P \otimes A) \arrow[d, "F(id_Q \otimes f)"] \\
FQ \otimes FB \arrow[d, "Fg"] & F(Q \otimes B) \arrow[dl, "Fg"] \\
FC &
\end{tikzcd}
$$

Since F is a monoidal functor from $\mathbf{C}$ to $\mathbf{C}'$, it is clear that this commutes.

Tensor Composition

Say $f : P \times A \to B, g : Q \times C \to D$ are arrows in $\mathbf{Para}_{\mathbf{D}}(\mathbf{C})$. Then they are also arrows in $\mathbf{C}$, and their $\mathbf{Para}_{\mathbf{D}}(\mathbf{C})$ tensor $f \otimes_{\mathbf{Para}_{\mathbf{D}}(\mathbf{C})} g$ is equal to $(f \otimes_{\mathbf{C}} g) \circ (id \otimes swap \otimes id)$. Therefore, $F'(f \otimes_{\mathbf{Para}_{\mathbf{D}}(\mathbf{C})} g) = F'f \otimes_{\mathbf{Para}_{\mathbf{D}'}(\mathbf{C}'))} F'g$ if the following diagram commutes:

$$
\begin{tikzcd}
FP \otimes FQ \otimes FA \otimes FC \arrow[rr, "\mu_F"] \arrow[d, "Fid_P \otimes Fswap_{Q,A} \otimes Fid_C"] & & F(P \otimes Q \otimes A \otimes C) \arrow[d, "F(id_P \otimes swap_{Q,A} \otimes id_C)"] \\
FP \otimes FA \otimes FQ \otimes FC \arrow[d, "Ff \otimes Fg"] & & F(P \otimes A \otimes Q \otimes C) \arrow[dll, "F(f \otimes g)"] \\
FC & &
\end{tikzcd}
$$

Since F is a monoidal functor from $\mathbf{C}$ to $\mathbf{C}'$, it is clear that this commutes.

□

7.3 Proof of Proposition 3

Consider the map $F : \mathbf{Para_D}(\mathbf{Para_{D'}}(\mathbf{C})) \to \mathbf{Para_{D'}}(\mathbf{Para_D}(\mathbf{C}))$ that acts as the identity on objects and maps the morphism $f : P_D \otimes P_{D'} \otimes A \to B$ to $(f \circ (swap \otimes id)) : P_{D'} \otimes P_D \otimes A \to B$. It is clear that this map is a bijection, since $(swap \otimes id) \circ (swap \otimes id) = id \otimes id$. We show that this map is a monoidal functor.

Let's quickly note that this map preserves monoidal unit since it acts as the identity on objects and the monoidal unit is the same between $\mathbf{Para_D}(\mathbf{Para_{D'}}(\mathbf{C}))$ and $\mathbf{Para_{D'}}(\mathbf{Para_D}(\mathbf{C}))$. It also trivially preserves any identity arrow $id_A : 1 \otimes 1 \otimes A \to A$.

Composition

Say $f : P_D \otimes P_{D'} \otimes A \to B, g : Q_D \otimes Q_{D'} \otimes B \to C$ are arrows in $\mathbf{Para_D}(\mathbf{Para_{D'}}(\mathbf{C}))$. Then we have:

$$
\begin{array}{r}
F(g \circ_{\mathbf{Para_D}(\mathbf{Para_{D'}}(\mathbf{C}))} f) = \\
(g \circ_{\mathbf{Para_D}(\mathbf{Para_{D'}}(\mathbf{C}))} f) \circ (swap \otimes id) = \\
g \circ (id_{Q_D} \otimes id_{Q_{D'}} \otimes f) \circ (swap \otimes id) = \\
g \circ (swap \otimes id) \circ (swap \otimes id) \circ (id_{Q_D} \otimes id_{Q_{D'}} \otimes f) \circ (swap \otimes id) = \\
Fg \circ (swap \otimes id) \circ (id_{Q_D} \otimes id_{Q_{D'}} \otimes f) \circ (swap \otimes id) = \\
Fg \circ_{\mathbf{Para_{D'}}(\mathbf{Para_D}(\mathbf{C}))} f \circ (swap \otimes id) = \\
Fg \circ_{\mathbf{Para_{D'}}(\mathbf{Para_D}(\mathbf{C}))} Ff
\end{array}
$$

Tensor

Say $f : P_D \times P_{D'} \times A \to B, g : Q_D \times Q_{D'} \times C \to D$ are arrows in $\mathbf{Para_D}(\mathbf{Para_{D'}}(\mathbf{C}))$. Then we have:

$$
\begin{array}{r}
F(f \otimes_{\mathbf{Para_D}(\mathbf{Para_{D'}}(\mathbf{C}))} g) = \\
(f \otimes_{\mathbf{Para_D}(\mathbf{Para_{D'}}(\mathbf{C}))} g) \circ (swap \otimes id_{A \otimes C}) = \\
(f \otimes_{\mathbf{C}} g) \circ (id_{P_{D'}} \otimes swap \otimes id_C) \circ (id \otimes swap \otimes id) \circ (swap \otimes id_{A \otimes C}) = \\
(f \otimes_C g) \circ (swap \otimes id_A \otimes swap \otimes id_C) \circ (id \otimes swap \otimes id) \circ (id \otimes swap \otimes id_{A \otimes C}) = \\
(Ff \otimes_C Fg) \circ (id \otimes swap \otimes id) \circ (id \otimes swap \otimes id) = \\
Ff \otimes_{\mathbf{Para_{D'}}(\mathbf{Para_D}(\mathbf{C}))} Fg
\end{array}
$$

□

7.4 Proof of Proposition 4

7.4.1 Full

Consider any A, B in $\mathbf{C}$ and any arrow $f : \Omega \times A \to B$ in the co-Kleisli category of $\mathbf{C}$ under $(\Omega \times -)$. Then f is also an arrow in $\mathbf{Para}_{\Omega^*}(\mathbf{C})$ and $Copy_\Omega$ maps f to f. Therefore $Copy_\Omega$ is full.

7.4.2 Identity

Since the id_A arrow in $\mathbf{Para}_{\Omega^*}(\mathbf{C})$ is of the form $1 \otimes A \to A$, $Copy_\Omega$ maps it to the arrow $id_A \circ_{\mathbf{C}} (cp^0_\Omega \otimes id^{\mathbf{C}}_A) = id_A$

7.4.3 Composition

Say $f : \Omega^m \times A \to B$ and $f' : \Omega^n \times B \to C$ are arrows in $\mathbf{Para}_{\Omega^*}(\mathbf{C})$. For any $\omega \in \Omega$ and $x_a \in A$ we have:

$$
\begin{aligned}
(Copy_\Omega f' \circ Copy_\Omega f) =\\
f' \circ (cp^{n-1}_\Omega \otimes id^{\mathbf{C}}_B) \circ f \circ (cp^{m-1}_\Omega \otimes id^{\mathbf{C}}_A) =\\
f' \circ_{\mathbf{C}} (cp^{n-1}_\Omega \otimes id^{\mathbf{C}}_B) \circ_{\mathbf{C}} (id^{\mathbf{C}}_\Omega \otimes f) \circ_{\mathbf{C}} (id^{\mathbf{C}}_\Omega \otimes cp^{m-1}_\Omega \otimes id^{\mathbf{C}}_A) \circ_{\mathbf{C}} (cp_\Omega \otimes id^{\mathbf{C}}_A) =\\
f' \circ_{\mathbf{C}} (cp^{n-1}_\Omega \otimes id^{\mathbf{C}}_B) \circ_{\mathbf{C}} (id^{\mathbf{C}}_\Omega \otimes f) \circ_{\mathbf{C}} (cp^{m}_\Omega \otimes id^{\mathbf{C}}_A) =\\
f' \circ_{\mathbf{C}} (id^{\mathbf{C}}_{\Omega^n} \otimes f) \circ_{\mathbf{C}} (cp^{n+m-1}_\Omega \otimes id^{\mathbf{C}}_A) =\\
(f' \circ f) \circ_{\mathbf{C}} (cp^{n+m-1}_\Omega \otimes id^{\mathbf{C}}_A) =\\
Copy_\Omega f' \circ f
\end{aligned}
$$

7.4.4 Tensor

Say $f : \Omega^m \times A \to B$ and $f' : \Omega^n \times C \to D$ are arrows in $\mathbf{Para}_{\Omega^*}(\mathbf{C})$. For any $\omega \in \Omega$ and $x_a \in A, x_c \in C$ we have:

$$
\begin{aligned}
(Copy_\Omega f' \otimes Copy_\Omega f) =\\
\left[f' \circ (cp^{n-1}_\Omega \otimes id^{\mathbf{C}}_C)\right] \otimes \left[f \circ (cp^{m-1}_\Omega \otimes id^{\mathbf{C}}_A)\right] =\\
(f' \otimes f) \circ_{\mathbf{C}} (cp^{n-1}_\Omega \otimes id^{\mathbf{C}}_C \otimes cp^{m-1}_\Omega \otimes id^{\mathbf{C}}_A) \circ_{\mathbf{C}} (swap_{(\Omega^2, C)} \otimes id^{\mathbf{C}}_A) \circ_{\mathbf{C}} (cp \otimes id^{\mathbf{C}}_C \otimes id^{\mathbf{C}}_A) =\\
(f' \otimes f) \circ_{\mathbf{C}} (cp^{n+m-1}_\Omega \otimes id^{\mathbf{C}}_C \otimes id^{\mathbf{C}}_A) =\\
Copy_\Omega f' \otimes f
\end{aligned}
$$

□

7.5 Proof of Proposition 5

Definition 7.1. *A Markov category C is a symmetric monoidal category in which every object $X \in C$ is equipped with a commutative comonoid structure given by a comultiplication $copy_X : X \to X \otimes X$ and a counit $del_X :: X \to I$, satisfying the commutative comonoid equations, compatibility with the monoidal structure and naturality of del_X, as well as $del_I = id_I$.*

Comonoid Structure: Defined similar to page 5 in this and page 9 in this

$$
\begin{aligned}
cp : 1 \times A \to A \times A\\
cp(\_, v) = v \oplus v\\
dc : 1 \times A \to 1\\
dc(\_, v) = \_
\end{aligned}
$$

Coassociativity (2.2 in this):

$$\begin{aligned}
(id \otimes cp) \circ cp \circ f(\omega_n, x_a) &= \\
(id \otimes cp) \circ (f(\omega_n, x_a) \otimes f(\omega_n, x_a)) &= \\
(f(\omega_n, x_a) \otimes (f(\omega_n, x_a) \otimes f(\omega_n, x_a))) &= \\
((f(\omega, x_a) \otimes f(\omega_n, x_a)) \otimes f(\omega_n, x_a)) &= \\
(cp \circ f(\omega_n, x_a)) \otimes (id \circ f(\omega_n, x_a)) &= \\
(cp \otimes id) \circ (f(\omega_n, x_a) \otimes f(\omega_n, x_a)) &= \\
(cp \otimes id) \circ cp \circ f(\omega_n, x_a) &
\end{aligned}$$

Coidentity (2.3 in this):

$$\begin{aligned}
(dc \otimes id) \circ cp \circ f(\omega_n, x_a) &= \\
(dc \otimes id) \circ (f(\omega_n, x_a) \otimes f(\omega_n, x_a)) &= \\
1 \otimes f(\omega_n, x_a) &= \\
f(\omega_n, x_a) \otimes 1 &= \\
(id \circ f(\omega_n, x_a)) \otimes (dc \circ f(\omega_n, x_a)) &= \\
(id \otimes dc) \circ cp \circ f(\omega_n, x_a) &
\end{aligned}$$

2.4.a of this

$$\begin{aligned}
dc \circ (f(\omega_n, x_a) \otimes f'(\omega_m, x_b)) &= \\
1 &= \\
1 \otimes 1 &= \\
(dc \circ f(\omega_n, x_a)) \otimes (dc \circ f'(\omega_m, x_b)) &
\end{aligned}$$

2.4.b of this

$$\begin{aligned}
cp \circ (f(\omega_n, x_a) \otimes f'(\omega_m, x_b)) &= \\
(f(\omega_n, x_a) \otimes f'(\omega_m, x_b)) \otimes (f(\omega_n, x_a) \otimes f'(\omega_m, x_b)) &= \\
f(\omega_n, x_a) \otimes (f'(\omega_m, x_b) \otimes f(\omega_n, x_a)) \otimes f'(\omega_m, x_b) &= \\
f(\omega_n, x_a) \otimes (swap \circ (f(\omega_n, x_a) \otimes f'(\omega_m, x_b))) \otimes f'(\omega_m, x_b) &= \\
(id \otimes swap \otimes id) \circ \left[f(\omega_n, x_a) \otimes f(\omega_n, x_a) \otimes f'(\omega_m, x_b) \otimes f'(\omega_m, x_b)\right] &= \\
(id \otimes swap \otimes id) \circ \left[(cp \circ f'(\omega_n, x_a)) \otimes (cp \circ f'(\omega_m, x_b))\right] &
\end{aligned}$$

2.5 of this

$$dc \circ f(\omega_n, x_a) = 1 = dc \circ (f' \circ f)(\omega_n, x_a)$$

7.6 Proof of Proposition 6

7.6.1 Composition

$$
\begin{aligned}
f : \Omega^m \times A \to B \qquad f' : \Omega^n \times B \to C \\
Push_\mu \left(f' \circ f\right) : A \times \mathcal{B}(C) \to [0,1] \\
Push_\mu \left(f' \circ f\right)_{x_a} (\sigma_c) = \\
\int_{(\omega_n,\omega_m)\in\Omega^n\times\Omega^m} \delta_{(f'\circ f)(\omega_n,\omega_m,x_a)}(\sigma_c) d\mu^{n+m} = \\
\int_{\omega_n\in\Omega^n}\int_{\omega_m\in\Omega^m} \delta_{(f'\circ f)(\omega_n,\omega_m,x_a)}(\sigma_c) d\mu^n d\mu^m = \\
\int_{\omega_n\in\Omega^n}\int_{\omega_m\in\Omega^m} \delta_{(f'(\omega_n,f(\omega_m,x_a))}(\sigma_c) d\mu^n d\mu^m = \\
\int_{\omega_n\in\Omega^n}\int_{\omega_m\in\Omega^m}\int_{x_b\in B} \delta_{f'(\omega_n,x_b)}(\sigma_c) d\delta_{f(\omega_m,x_a)}(x_b) d\mu^n d\mu^m = \\
\int_{x_b\in B}\int_{\omega_n\in\Omega^n}\int_{\omega_m\in\Omega^m} \delta_{f'(\omega_n,x_b)}(\sigma_c) d\delta_{f(\omega_m,x_a)}(x_b) d\mu^n d\mu^m = \\
\int_{x_b\in B}\left[\int_{\omega_n\in\Omega^n} \delta_{f'(\omega_n,x_b)}(\sigma_c) d\mu^n\right]\left[\int_{\omega_m\in\Omega^m} d\delta_{f(\omega_m,x_a)}(x_b) d\mu^m\right] = \\
\int_{x_b\in B} [Push_\mu f']_{x_b}(\sigma_c) d[Push_\mu f]_{x_a}(x_b) = \\
(Push_\mu f' \circ Push_\mu f)_{x_a}(\sigma_c)
\end{aligned}
$$

7.6.2 Tensor

:

$$
\begin{aligned}
f : \Omega^m \times A \to B \qquad f' : \Omega^n \times C \to D \\
Push_\mu \left(f' \otimes f\right) : (A \times C) \times \mathcal{B}(B \times D) \to [0,1] \\
Push_\mu \left(f' \otimes f\right)_{x_{ca}} (\sigma_{db}) = \\
\int_{\omega_m\in\Omega^m}\int_{\omega_n\in\Omega^n} \delta_{(f'\otimes f)(\omega_n,\omega_m,x_a,x_c)}(\sigma_{db}) d\mu^m d\mu^n = \\
\int_{\omega_m\in\Omega^m}\int_{\omega_n\in\Omega^n} \delta_{(f'(\omega_m,x_c)\otimes f(\omega_n,x_a)}(\sigma_{db}) d\mu^m d\mu^n = \\
\int_{\omega_m\in\Omega^m}\int_{\omega_n\in\Omega^n} \delta_{f'(\omega_m,x_c)}(\sigma_d)\delta_{f(\omega_n,x_a))}(\sigma_b) d\mu^m d\mu^n = \\
\int_{\omega_m\in\Omega^m} \delta_{f'(\omega_m,x_c)}(\sigma_d) d\mu^m \int_{\omega_n\in\Omega^n} \delta_{f(\omega_n,x_a))}(\sigma_b) d\mu^n = \\
(Push_\mu f')_{x_c}(\sigma_d) * (Push_\mu f)_{x_a}(\sigma_b) = \\
(Push_\mu f' \otimes Push_\mu f)_{x_a,x_c}(\sigma_{db})
\end{aligned}
$$

7.7 Proof of Proposition 7

7.7.1 Composition

Say $f : \Omega^n \otimes \mathbb{R}^a \otimes \mathbb{R}^p \to \mathbb{R}^b$ and $f' : \Omega^m \otimes \mathbb{R}^b \otimes \mathbb{R}^q \to \mathbb{R}^c$ are arrows in $\mathcal{R}_\mu$. We can show that for all $x_a \in \mathbb{R}^a, x_p \in \mathbb{R}^p, x_q \in \mathbb{R}^q$ we can write $(f' \circ f)_*\mu(x_p, x_a, x_q)(\sigma_c)$ in the form $\int_{x_c\in\sigma_c} g(x_c) d\lambda^c$ where

λ^c is the Lebesgue measure over $\mathbb{R}^c$:

$$\begin{aligned}
&(f'\circ f)_*\mu(x_p,x_a,x_q)(\sigma_c) = \\
&\int_{x_b\in\mathbb{R}^b} f'(-,x_q,x_b)_*\mu(\sigma_c) \quad df(-,x_p,x_a)_*\mu(x_b) = \\
&\int_{x_b\in\mathbb{R}^b}\left[\int_{x_c\in\sigma_c}\frac{df'(-,x_q,x_b)_*\mu^m}{d\lambda^c}(x_c)d\lambda^c\right]\left[\frac{df(-,x_p,x_a)_*\mu}{d\lambda^b}(x_b)d\lambda^b\right] = \\
&\int_{x_c\in\sigma_c}\left[\int_{x_b\in\mathbb{R}^b}\frac{df'(-,x_q,x_b)_*\mu^m}{d\lambda^c}(x_c) * \frac{df(-,x_p,x_a)_*\mu}{d\lambda^b}(x_b)d\lambda^b\right]d\lambda^c
\end{aligned}$$

7.7.2 Tensor

Say $f:\Omega^n\otimes\mathbb{R}^p\otimes\mathbb{R}^a\to\mathbb{R}^b$ and $f':\Omega^m\otimes\mathbb{R}^q\otimes\mathbb{R}^c\to\mathbb{R}^d$ are arrows in $\mathcal{R}_\mu$. We can show that for all $x_a\in\mathbb{R}^a, x_c\in\mathbb{R}^c, x_p\in\mathbb{R}^p, x_q\in\mathbb{R}^q$ we can write $(f'\otimes f)_*\mu_{(x_q,x_p,x_c,x_a)}(\sigma_d\times\sigma_b)$ in the form $\int_{x_{db}\in\sigma_d\times\sigma_b} g(x_{db})d\lambda^{d+b}$ where λ^{d+b} is the Lebesgue measure over $\mathbb{R}^{d+b}$:

$$\begin{aligned}
&(f'\otimes f)_*\mu_{(x_q,x_p,x_c,x_a)}(\sigma_d\times\sigma_b) = \\
&f'(-,x_q,x_c)_*\mu(\sigma_d) * f(-,x_p,x_a)_*\mu(\sigma_b) = \\
&\left[\int_{x_d\in\sigma_d}\frac{df'(-,x_q,x_c)_*\mu}{d\lambda^d}(x_d)d\lambda^d\right] * \left[\int_{x_b\in\sigma_b}\frac{df(-,x_p,x_a)_*\mu}{d\lambda^b}(x_b)d\lambda^b\right] = \\
&\int_{x'_{db}\in\sigma_d\times\sigma_b}\left[\frac{df'(-,x_q,x_c)_*\mu}{d\lambda^d}(x_d) * \frac{df(-,x_p,x_a)_*\mu}{d\lambda^b}(x_b)\right]d\lambda^{d+b}
\end{aligned}$$

□

7.8 Proof of Proposition 8

Let's begin by noting that $\mathcal{RN}_\mu$ preserves the identity by construction.

7.8.1 Composition

Say $f:\Omega^n\otimes\mathbb{R}^p\otimes\mathbb{R}^a\to\mathbb{R}^b$ and $f':\Omega^m\otimes\mathbb{R}^q\otimes\mathbb{R}^b\to\mathbb{R}^c$ are non-identity arrows in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$:

$$\begin{aligned}
&\mathcal{RN}_\mu(f'\circ f):\mathbb{R}^{q+p}\otimes\mathbb{R}^a\otimes\mathbb{R}^c\to\mathbb{R} \\
&\mathcal{RN}_\mu(f'\circ f)(x_{qp},x_a,x_c) = \\
&\frac{d\int_{x_b\in\mathbb{R}^b}(f'\circ f)(-,x_{qp},x_a)_*\mu^{n+m}}{d\lambda^c}(x_c) = \\
&\frac{d\int_{x_b\in\mathbb{R}^b} f'(-,x_q,x_b)_*\mu^m(-) \quad df(-,x_p,x_a)_*\mu^n(x_b)}{d\lambda^c}(x_c) = \\
&\frac{d\int_{x_b\in\mathbb{R}^b}\int_{x'_c\in(-_c)}\frac{df'(-,x_q,x_b)_*\mu^m}{d\lambda^c}(x'_c)d\lambda^c \quad df(-,x_p,x_a)_*\mu^n(x_b)}{d\lambda^c}(x_c) = \\
&\frac{d\int_{x_b\in\mathbb{R}^b}\int_{x'_c\in(-_c)}\frac{df'(-,x_q,x_b)_*\mu^m}{d\lambda^c}(x'_c)d\lambda^c \quad \frac{df(-,x_p,x_a)_*\mu^n}{d\lambda^b}(x_b)d\lambda^b}{d\lambda^c}(x_c) = \\
&\frac{d\int_{x'_c\in(-_c)}\left[\int_{x_b\in\mathbb{R}^b}\frac{df'(-,x_q,x_b)_*\mu^m}{d\lambda^c}(x'_c) \quad \frac{df(-,x_p,x_a)_*\mu^n}{d\lambda^b}(x_b)d\lambda^b\right]d\lambda^c}{d\lambda^c}(x_c) = \\
&\int_{x_b\in\mathbb{R}^b}\frac{df'(-,x_q,x_b)_*\mu^m}{d\lambda^c}(x_c) \quad \frac{df(-,x_p,x_a)_*\mu^n}{d\lambda^b}(x_b)d\lambda^b = \\
&(\mathcal{RN}_\mu f'\circ\mathcal{RN}_\mu f)(x_{qp},x_a,x_c)
\end{aligned}$$

7.8.2 Tensor

Say $f : \Omega^n \otimes \mathbb{R}^a \to \mathbb{R}^b$ and $f' : \Omega^m \otimes \mathbb{R}^c \to \mathbb{R}^d$ are non-identity arrows in $\mathbf{Para}_{\Omega^*}(\mathbf{Para}(\mathbf{EucMeas})))_{\mathcal{R}_\mu}$:

$$\begin{aligned}
\mathcal{RN}_\mu(f' \otimes f) : \mathbb{R}^{q+p} \otimes \mathbb{R}^{c+a} \otimes \mathbb{R}^{d+b} \to \mathbb{R} \\
\mathcal{RN}_\mu(f' \otimes f)(x_{qp}, x_{ca}, x_{db}) = \\
\frac{d\left[(f' \otimes f)(-, x_{qp}, x_{ca})_* \mu^{n+m}\right]}{d\lambda^{d+b}}(x_{db}) = \\
\frac{d\left[f'(-, x_q, x_c)_* \mu^m(-) * f(-, x_p, x_a)_* \mu^n(-)\right]}{d\lambda^{d+b}}(x_{db}) = \\
\frac{d\left[\left(\int_{x'_d \in (-)_d} \frac{df'(-,x_q,x_c)_*\mu^m}{d\lambda^d}(x'_d) d\lambda^d\right) * \left(\int_{x'_b \in (-)_b} \frac{df(-,x_p,x_a)_*\mu^n}{d\lambda^b}(x'_b) d\lambda^b\right)\right]}{d\lambda^{d+b}}(x_{db}) = \\
\frac{d\left[\int_{x'_{db} \in (-)_{db}} \left(\frac{df'(-,x_q,x_c)_*\mu^m}{d\lambda^d}(x'_d)\right) * \left(\frac{df(-,x_p,x_a)_*\mu^n}{d\lambda^b}(x'_b)\right) d\lambda^{d+b}\right]}{d\lambda^{d+b}}(x_{db}) = \\
\left(\frac{df'(-, x_q, x_c)_* \mu^m}{d\lambda^d}(x_d)\right) * \left(\frac{df(-, x_p, x_a)_* \mu^n}{d\lambda^b}(x_b)\right) = \\
(\mathcal{RN}_\mu f' \otimes \mathcal{RN}_\mu f)(x_{qp}, x_{ca}, x_{db})
\end{aligned}$$

$\square$

7.9 Proof of Proposition 9

First, let's note that for any m, the pushforward distribution of the random variable $f : \Omega^n \to \mathbb{R}^a$ over the probability space $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$ is equivalent to the pushforward distribution of the random variable $f^e(\omega^n, \omega^m) = f(\omega^n)$ over $(\Omega^{n+m}, \mathcal{B}(\Omega^{n+m}), \mu^{n+m})$.

$$\begin{aligned}
f^e_* \mu^{n+m}(\sigma_a) = \\
\int_{\omega_{nm} \in \Omega^{n+m}} \delta_{f^e(\omega_n, \omega_m)}(\sigma_a) d\mu^{n+m} = \\
\int_{\omega_n \in \Omega^n} \int_{\omega_m \in \Omega^m} \delta_{f^e(\omega_n, \omega_m)}(\sigma_a) d\mu^n d\mu^m = \\
\int_{\omega_n \in \Omega^n} \delta_{f(\omega_n)}(\sigma_a) d\mu^n \int_{\omega_m \in \Omega^m} d\mu^m = \\
f_* \mu^n(\sigma_a)
\end{aligned}$$

Next, we note that for any $x_a \in \mathbb{R}^a, x_p \in \mathbb{R}^p$ arrow $f : \Omega^n \otimes \mathbb{R}^p \otimes \mathbb{R}^a \to \mathbb{R}^b \in \mathcal{N}_\mu$, $f(-, x_p, x_a)$ is multivariate normal over the probability space $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$. This follows from the fact that $f(\omega_n, x_p, x_a) = T(x_p, x_a) + G(\omega_n)$ where $T(x_p, x_a)$ is a constant and $G(-)$ is multivariate normal over $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$.

Next, we show that for any pair of arrows $f : \Omega^n \otimes \mathbb{R}^p \otimes \mathbb{R}^a \to \mathbb{R}^b, f' : \Omega^m \otimes \mathbb{R}^q \otimes \mathbb{R}^b \to \mathbb{R}^c$ and $x_a \in \mathbb{R}^a, x_p \in \mathbb{R}^p, x_q \in \mathbb{R}^q$ such that $f(-, x_p, x_a)$ is multivariate normal over $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$, the random variable $(f' \circ f)(-, x_q, x_p, x_a)$ is multivariate normal over $(\Omega^{n+m}, \mathcal{B}(\Omega^{n+m}), \mu^{n+m})$.

$$(f' \circ f)(\omega_{nm}, x_q, x_p, x_a) = f'(\omega_n, x_q, f(\omega_m, x_p, x_a)) = \frac{T'(x_q, f(\omega_m, x_p, x_a)) + G'(\omega_n)}{1 + v'}$$

Since x_q is constant and $f(-, x_p, x_a)$ is multivariate normal over $(\Omega^m, \mathcal{B}(\Omega^m), \mu^m)$ by our note above we have that $f^e(\omega_n, \omega_m, x_p, x_a) = f(\omega_m, x_p, x_a)$ is multivariate normal over $(\Omega^{n+m}, \mathcal{B}(\Omega^{n+m}), \mu^{n+m})$, and $T^e(x_q, f^e(\omega_n, \omega_m, x_q, x_p, x_a))$ is as well. Similarly, $G'^e(\omega_n, \omega_m) = G'(\omega_n)$ is also multivariate normal and independent of $T^e(x_q, f^e(\omega_n, \omega_m, x_q, x_p, x_a))$ over $(\Omega^{n+m}, \mathcal{B}(\Omega^{n+m}), \mu^{n+m})$.

Therefore, we can write

$$(f' \circ f)(\omega_{nm}, x_q, x_p, x_a) = \frac{T'(x_q, f(\omega_m, x_p, x_a)) + G'(\omega_n)}{1 + v'} = \frac{T'(x_q, f^e(\omega_n, \omega_m, x_p, x_a)) + G'^e(\omega_n, \omega_m)}{1 + v'}$$

Since this is a sum of independent normally distributed random variables, $(f' \circ f)(-, x_q, x_p, x_a)$ is multivariate normal over $(\Omega^{n+m}, \mathcal{B}(\Omega^{n+m}), \mu^{n+m})$.

Finally, we show that for any pair of arrows $f : \Omega^n \otimes \mathbb{R}^p \otimes \mathbb{R}^a \to \mathbb{R}^b, f' : \Omega^m \otimes \mathbb{R}^q \otimes \mathbb{R}^c \to \mathbb{R}^d$ and $x_a \in \mathbb{R}^a, x_c \in \mathbb{R}^c, x_q \in \mathbb{R}^q, x_p \in \mathbb{R}^p$ such that f' is multivariate normal over $(\Omega^m, \mathcal{B}(\Omega^m), \mu^m)$ and $f(-, x_p, x_a)$ is multivariate normal over $(\Omega^n, \mathcal{B}(\Omega^n), \mu^n)$, the random variable $(f' \otimes f)(-, x_q, x_p, x_c, x_a)$ is multivariate normal over $(\Omega^{n+m}, \mathcal{B}(\Omega^{n+m}), \mu^{n+m})$. This follows from the fact that both $f'^e(-, x_q, x_c)$ and $f^e(-, x_p, x_a)$ are independent and multivariate normal over $(\Omega^{n+m}, \mathcal{B}(\Omega^{n+m}), \mu^{n+m})$ $\square$

7.10 Proof of Proposition 10

$$\begin{aligned}
\int_{\omega_{mn} \in \Omega^{m+n}} f'(\omega_m, x_q, f(\omega_n, x_p, x_a)) d\mu^{m+n} &= \\
\int_{\omega_{mn} \in \Omega^{m+n}} T'(x_q, f(\omega_n, x_p, x_a)) + G'(\omega_m) d\mu^{m+n} &= \\
\int_{\omega_n \in \Omega^n} T'(x_q, f(\omega_n, x_p, x_a)) d\mu^n + \int_{\omega_m \in \Omega^m} G'(\omega_m) d\mu^m &= \\
T'\left(x_q, \int_{\omega_n \in \Omega^n} f(\omega_n, x_p, x_a) d\mu^n\right) + \int_{\omega_m \in \Omega^m} G'(\omega_m) d\mu^m &= \\
\int_{\omega_m \in \Omega^m} f'\left(x_q, \omega_m, \int_{\omega_n \in \Omega^n} f(\omega_n, x_p, x_a) d\mu^n\right) d\mu^m &
\end{aligned}$$