String Diagrams with Factorized Densities

Eli Sennesh
Khoury College of Computer Science
Northeastern University
Boston, Massachusetts, United States of America
sennesh.e@northeastern.edu

Jan-Willem van de Meent
Amsterdam Machine Learning Lab (AMLab)
University of Amsterdam
Amsterdam, the Netherlands
j.w.vandemeent@uva.nl

A growing body of research on probabilistic programs and causal models has highlighted the need to reason compositionally about model classes that extend directed graphical models. Both probabilistic programs and causal models define a joint probability density over a set of random variables, and exhibit sparse structure that can be used to reason about causation and conditional independence. This work builds on recent work on Markov categories of probabilistic mappings to define a category whose morphisms combine a joint density, factorized over each sample space, with a deterministic mapping from samples to return values. This is a step towards closing the gap between recent category-theoretic descriptions of probability measures, and the operational definitions of factorized densities that are commonly employed in probabilistic programming and causal inference.

1 Introduction

Statisticians and machine learners analyze observed data by synthesizing models of those data. These models take a variety of forms, with several of the most widely used being directed graphical models, probabilistic programs, and structural causal models (SCMs). Applications of these frameworks have included cognitive modeling [7, 20], simulation-based inference [9], and model-based planning [12, 21]. Unfortunately, the richer the model class, the weaker the mathematical tools available to reason rigorously about it: SCMs built on linear equations with Gaussian noise admit easy inference, while graphical models have a clear meaning and a wide array of inference algorithms but encode a limited family of models. Probabilistic programs can encode any computably sampleable distribution, but the definition of their densities commonly relies on operational analogies with directed graphical models.

In recent years, category theorists have developed increasingly sophisticated ways to reason diagrammatically about a variety of complex systems. These include (co)parameterized categories of systems that may modify their parameters [5] and hierarchical string diagrams for rewriting higher-order computations [1]. Recent work on Markov categories of probabilistic mappings has provided denotational semantics to probabilistic programs [32, 18], abstract categorical descriptions of conditioning, disintegration, sufficient statistics, conditional independence [8, 13], and generalized causal models [14, 15].

This paper will take a step towards closing the gap between categorical probability and operational practice in probabilistic programming and applied Bayesian statistics. Denotational semantics for probabilistic programs define a measure over return values of a program given its inputs [32, 18]. To reason about inference methods, practitioners need to consider the joint distribution of internal random variables, as well as its density's factorization into conditionals. Section 2 will review basic definitions from probability and measure theory necessary to do so. Section 3 will then develop a category whose morphisms express joint (rather than marginal) distributions with factorized joint densities. Section 4 will show that generalized causal models can factorize these densities and admit interventional and counterfactual queries. Section 5 will work through a pair of examples and summarize the paper's developments.

Appendix A reviews the measure-theoretic concepts employed here; Appendix B reviews parametric and coparametric categories [5]; and Appendix C reviews free copy/delete and Markov categories.

To appear in EPTCS.

Notation The notation $(\mathscr{C}, \otimes, I)$ will range over strict symmetric monoidal categories (SMC's for short). We denote composition as $g \circ f$ or equivalently as $f \mathbin{⨾} g$, write $(X^*, \odot, ())$ for the finite list monoid on X's, and overload $\otimes$ and $\oplus$ for direct products and sums. We draw string diagrams from the top (domain) to the bottom (codomain), showing products from left to right. Given a Markov category $\mathscr{C}$ we will draw deterministic maps in $\mathscr{C}_{det} \subset \mathscr{C}$ (which commute with **copy**) as rectangles and stochastic ones as ellipses/circles. We nest brackets with parentheses $([])$ equivalently.

2 Background: abstract and concrete categorical probability

This section will review the background on which the rest of the paper builds. Categorical probability begins from an abstract notion of nondeterminism: processes with a notion of "independent copies". Categorical probability then refines from a setting in which those nondeterministic processes "happen" whether observed or not, to a refined setting in which processes only "happen" when they affect an observed output. Categories of probability kernels, taking into account the details of measure theory (see Appendix A), will form a concrete instance of the abstract setting.

Definition 1 represents nondeterministic processes abstractly. A copy/delete category is an SMC whose morphisms generate information which can be copied or deleted freely.

Definition 1 (Copy/delete category)**.** *A copy-delete or* CD-category *is an SMC $(\mathscr{C}, \otimes, I)$ in which every object $X \in Ob(\mathscr{C})$ has a commutative comonoid structure* $\mathbf{copy}_X : \mathscr{C}(X, X \otimes X)$ *and* $\mathbf{del}_X : \mathscr{C}(X, I)$ *which commutes with the monoidal product structure.*

Definition 2 then refines the abstract setting of CD categories to require that deleting the only result of a nondeterministic process is equivalent to deleting the process itself.

Definition 2 (Markov category)**.** *A* Markov category *is a semicartesian CD-category $(\mathscr{C}, \otimes, I)$, so that the comonoidal counit is natural ($\forall f : \mathscr{C}(Z, X), f \mathbin{⨾} \mathbf{del}_X = I$) and makes $I \in Ob(\mathscr{C})$ a terminal object.*

Example 1 gives the canonical Markov category, consisting of measurable spaces and maps.

Example 1 (Measurable spaces and functions form a category [33])**.** *Measurable spaces and functions form a Cartesian category* **Meas** *with objects $(X, \Sigma_X) \in Ob(\mathbf{Meas})$ consisting of sets $X \in Ob(\mathbf{Set})$ and their σ-algebras*[1] *Σ_X and morphisms $\mathbf{Meas}((Z, \Sigma_Z), (X, \Sigma_X)) = \left\{ f \in X^Z \mid \forall \sigma_X \in \Sigma_X, f^{-1}(\sigma_X) \in \Sigma_Z \right\}$ consisting of measurable functions between measurable spaces.*

Meas acquires its Markov comonoid structure from its Cartesian structure. Definition 3 below provides the canonical Markov category for measure-theoretic probability.

Definition 3 (Category of measurable spaces and Markov kernels)**.** *The category* $\mathbf{Stoch} = Kl(\mathbb{P})(\mathbf{Meas})$ *of measurable spaces and Markov kernels is the Kleisli category of the Giry monad [17] over* **Meas***, having measurable spaces as objects and Markov kernels (Definition 19) between them as morphisms.*

Much of this paper will require a *strict* Markov category as in Definition 4 below.

Definition 4 (Strict Markov category)**.** *A* strict *Markov category is one whose underlying SMC (with comonoid structure thrown away) is strict monoidal (its associator and unitors are identity).*

Theorem 10.17 in Fritz [13] showed that every Markov category is comonoid equivalent to a strict one, licensing us to work with strictified Markov categories **Meas** and **Stoch** without further concern.

Unless otherwise mentioned, this paper will work with **Meas** and **Stoch** as strict, causal Markov categories[2]. When the ambient category and σ-algebra is clear from context, $f : Z \rightsquigarrow X$ will abbreviate

[1] Collections of "measurable subsets" closed under complements, countable unions, and countable intersections

[2] The latter property is shown in Example 11.35 of Fritz [13]

$f : \mathbf{Stoch}((Z,\Sigma_Z),(X,\Sigma_X))$. In the concrete case of **Stoch**, measurable maps give the deterministic maps $\mathbf{Meas} \simeq \mathbf{Stoch}_{det} \subseteq \mathbf{Stoch}$. While Markov categories provide a compositional setting for nondeterministic processes, Markov kernels in these categories only provide probability measures for their outputs given their inputs. By design, they "forget" (i.e. marginalize over) all intermediate randomness in long chains of composition. Section 3 will build up a novel setting that "remembers" (i.e. does not marginalize over) joint distributions over all intermediate random variables through long chains of composition, and will show when there exist probability densities with respect to the joint distributions thus formed.

3 Joint distributions and densities for string diagrams

Statisticians cannot utilize input-output (parameter to distribution) mappings alone, except for maximum likelihood estimation. Instead, these typically appear as conditional probability distributions in a larger probability model. This larger model necessarily encodes a *joint distribution* over all relevant random variables, both those observed as data and the *latent* variables that give rise to observations. Practical probabilistic reasoning then consists of applying the laws of probability (product law for conjunctions, sum law for disjunctions, marginalization for unconditional events, Bayesian inversion) to numerical *densities* representing the joint distribution. This section will model the algebra of joint probability densities in a novel Markov category **Joint** constructed on the underlying Markov category **Stoch**.

Section 3.1 will first review an abstraction for categories in which morphisms act by "pushing forward" an internal "parameter space" and then instantiate that abstraction on a Markov category to yield a Markov category **Joint** of joint distributions. Section 3.2 will give the conditions for a concrete Markov kernel to admit a density. Section 3.3 will use those preliminaries to define a Markov category whose morphisms generate and push forward a joint probability density.

3.1 Accumulating random variables into joint distributions

Structural graphical models and probabilistic programs separate between the functions and variables they allow into deterministic and random ones [24]. Representing deterministic mechanisms categorically requires assuming that each nondeterministic process consists of a deterministic mechanism and a (potentially conditional) distribution over a random variable. This subsection will exploit "cybernetic" constructions (overviewed in Appendix B) for parameterization of deterministic mechanisms by random inputs and "writing out" of internal joint distributions as coparameters.

Proposition 1 will show the concrete category **Stoch** supports those constructions.

Proposition 1 (**Stoch** forms a symmetric monoidal $\mathscr{M}$-actegory[3])**.** *The concrete category* **Stoch** *forms a symmetric monoidal $\mathscr{M}$-actegory for $\mathscr{M} =$* **Stoch** *and $\mathscr{C} =$* **Stoch**.

Proof. Any SMC $\mathscr{C}$ forms a symmetric monoidal $\mathscr{M}$-actegory for $\mathscr{M} = \mathscr{C}$ with the product functor $\mathscr{M} \bullet \mathscr{C} = \mathscr{C} \times \mathscr{C}$ from the product category. Any Markov category is also an SMC, and so **Stoch** suffices. □

In this trivial case, Definition 25 simplifies so that Definition 5 will form an SMC.

Definition 5 (Symmetric monoidal parametric categories)**.** *Given a strict SMC $(\mathscr{C},\otimes,I)$, the symmetric monoidal* parametric (bi)category $\mathbf{Para}_{\otimes}(\mathscr{C})$ *has as objects those of $\mathscr{C}$ and as morphisms the pairs* $\mathbf{Para}_{\otimes}(\mathscr{C})(A,B) = \{(M,k) \in Ob(\mathscr{C}) \times \mathscr{C}(M \otimes A,B)\}$. *Composition for the two parameterized morphisms (M,k) :* $\mathbf{Para}_{\otimes}(\mathscr{C})(A,B)$ *and (M',k') :* $\mathbf{Para}_{\otimes}(\mathscr{C})(B,C)$ *consists of $(M' \otimes M, k' \circ (id_{M'} \otimes k))$; identities on objects A consist of (I, id_A); and* $(\mathbf{Para}_{\otimes}(\mathscr{C}),\otimes,I)$ *inherits its monoidal structure from $\mathscr{C}$*[4].

[3] Definition 24 in Appendix B

[4] see Proposition 6

$\mathbf{Para}_\otimes(\mathbf{Stoch})$ will suffice for Definition 6 to model a Markov kernel over a joint distribution. The jointly random *residual* $(M, \Sigma_M) \in Ob(\mathbf{Stoch})$ will parameterize the deterministic map k.

Definition 6 (Joint Markov kernel)**.** *A* joint Markov kernel *is a pair of a Markov kernel with a deterministic mapping parameterized by that Markov kernel, up to permutation of residual components*

$$\mathbf{Joint}(Z,X) := \{(f,[M,k]) : \mathbf{Stoch}(Z,M) \times \mathbf{Para}_\otimes(\mathbf{Stoch}_{det})(Z,X)\}.$$

As implied by the hom-set notation above, joint Markov kernels will form a category of nondeterministic processes. Since the residuals of joint distributions only contribute to downstream processes through their local outputs, Theorem 1 will show this to be a copy/delete category.

Theorem 1 (Joint Markov kernels form a copy/delete category)**.** **Joint** *is a strict copy/delete category having* $Ob(\mathbf{Joint}) = Ob(\mathbf{Stoch})$ *and joint Markov kernels as morphisms.*

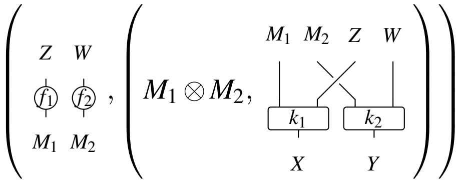
Proof. **Joint** must admit the typical requirements of a category as well as deterministic, copy-delete symmetric monoidal structure. We can demonstrate the necessary deterministic structure by exhibiting joint kernels $(I,k) : \mathbf{Para}_\otimes(\mathbf{Stoch}_{det})(Z,X) \implies ([I,k], \mathbf{del}_Z) : \mathbf{Joint}(Z,X)$ for any noiseless causal mechanism. Setting $k = \mathbf{copy}_X$ or $k = \mathbf{del}_Z$ yields the necessary copy and delete maps. Setting $k = \mathbf{swap}_{Z \otimes X}$ gives the necessary symmetry of the monoidal product. It remains to show that **Joint** has a monoidal product over morphisms and that its hom-sets are closed under composition.

Given two joint Markov kernels $(f_1,[M_1,k_1]) : \mathbf{Joint}(Z,X)$ and $(f_2,[M_2,k_2]) : \mathbf{Joint}(W,Y)$, their monoidal product is formed by pairing their causal mechanisms and noise distributions

$$(f_1,[M_1,k_1]) \otimes (f_2,[M_2,k_2]) := (f_1 \otimes_{\mathbf{Stoch}} f_2, [M_1,k_1] \otimes_{\mathbf{Para}_\otimes(\mathbf{Stoch}_{det})} [M_2,k_2]) : \mathbf{Joint}(Z \otimes W, X \otimes Y).$$

Composing two joint Markov kernels $(f_1,[M_1,k_1]) : \mathbf{Joint}(Z,X)$ and $(f_2,[M_2,k_2]) : \mathbf{Joint}(X,Y)$ along their intermediate object involves composing their parametric maps and taking a conditional product of their stochastic kernels to form the composite joint distribution

$$(f_1,[M_1,k_1]) \fatsemi (f_2,[M_2,k_2]) :=$$



$$: \mathbf{Joint}(Z,Y). \tag{1}$$

□

Anything called a *joint* Markov kernel ought to expose its internal joint distribution in a structured way. Definition 7 will link the composition of joint distributions to the cybernetics literature.

Definition 7 (Symmetric monoidal coparametric categories[5])**.** *Given a strict SMC* $(\mathscr{C}, \otimes, I)$*, the symmetric monoidal* coparametric (bi)category $\mathbf{CoPara}_\otimes(\mathscr{C})$ *has as objects those of* $\mathscr{C}$ *and as morphisms the pairs* $\mathbf{CoPara}_\otimes(\mathscr{C})(A,B) = \{(M,k) \in Ob(\mathscr{C}) \times \mathscr{C}(A, M \otimes B)\}$ *of a residual object and a morphism from* A *to* $M \otimes B$*. Composition for the morphisms* $(M,k) : \mathbf{CoPara}_\otimes(\mathscr{C})(A,B)$ *and* $(M',k') : \mathbf{CoPara}_\otimes(\mathscr{C})(B,C)$ *consists of* $(M' \otimes M, (id_M \otimes k') \circ k))$*; identities on objects* A *consist of* (I, id_A)*; and* $(\mathbf{CoPara}_\otimes(\mathscr{C}), \otimes, I)$ *inherits its monoidal structure from* $\mathscr{C}$[6].

Joint serves to work with joint distributions compositionally rather than marginalizing them out. Theorem 2 will show how mapping from $\mathbf{Joint} \to \mathbf{CoPara}_\otimes(\mathbf{Stoch})$ exposes the full joint distribution.

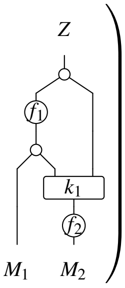
[5] See Definition 26 for the more general case

[6] See Proposition 6

Theorem 2 (Joint Markov kernels coparameterize joint distributions)**.** *There exists a full, identity-on-objects Markov functor* $[\![\cdot]\!] : \mathbf{Joint} \to \mathbf{CoPara}_{\otimes}(\mathbf{Stoch})$ *which maps the residual of a joint Markov kernel in* **Joint** *onto the residual of its image in* $\mathbf{CoPara}_{\otimes}(\mathbf{Stoch})$.

Proof. The required functor sends morphisms $[\![\cdot]\!] : \mathbf{Joint}(Z,X) \to \mathbf{CoPara}_{\otimes}(\mathbf{Stoch})(Z,X)$ to coparameterized Markov kernels whose codomain is the joint distribution over the residual and the output

$$[\![(f,[M,k])]\!] = \left(M, \quad \right).$$



This functor is trivially full, since any morphism $f : \mathbf{CoPara}_{\otimes}(\mathbf{Stoch})(Z,X)$ embeds trivially into **Joint** by setting the corresponding deterministic $k = id_{M \otimes X}$. It is not faithful: multiple "divisions of labor" between f and k can yield the same Markov kernel in $\mathbf{CoPara}_{\otimes}(\mathbf{Stoch})$. □

Corollary 3 will give the trivial extension of marginalizing over the residual.

Corollary 3 (Marginalizing a joint Markov kernel's residual yields a Markov kernel)**.** *There exists a full, identity-on-objects functor* $J : \mathbf{Joint} \to \mathbf{Stoch}$.

Proof. The required functor J just applies $[\![\cdot]\!]$ and then forgets the residual by composition with $\mathbf{del}_M$: its action on morphisms is $J((f,[M,k])) = [\![(f,[M,k])]\!] \,\mathring{,}\, (\mathbf{del}_M \otimes id_X)$. □

This subsection has considered arbitrary, unstructured joint distributions **Joint**. Section 3.2 will examine the special case in which the residual object is a standard Borel space and the conditional distribution into it meets the necessary conditions to admit a probability density.

3.2 Base measures and densities over standard Borel spaces

Applied probability typically works not with probability measures but with probability densities, functions over a finite-dimensional sample space giving the "derivative" of a probability measure at a point. However, probability densities only exist for measures that meet the conditions of the Radon-Nikodymn Theorem, and only relative to a specified base measure over the sample space. This section will restrict the residual objects or internal noises of joint Markov kernels to standard Borel sample spaces admitting probability densities, and then show that this restriction still admits a broad class of joint Markov kernels.

Definition 8 provides a suitable ambient category for base measures.

Definition 8 (Category of measure spaces)**.** *The* category of measure spaces $\mathbb{M}$ *has as objects the measure spaces* (X, Σ_X, μ) *(Definition 22) and as morphisms the measure-preserving maps*

$$\mathbb{M}((Z,\Sigma_Z,\mu_Z),(X,\Sigma_X,\mu_X)) = \left\{ f : \mathbf{Meas}((Z,\Sigma_Z),(X,\Sigma_X)) \mid \forall \sigma_X \in \Sigma_X, \mu_Z(f^{-1}(\sigma_X)) = \mu_X(\sigma_X) \right\}.$$

Applications typically deal with probability densities over finite-dimensional Euclidean spaces and countable sets. In **Meas**, these can be characterized by the standard Borel spaces $\mathbf{Sbs} \subset \mathbf{Meas}$, which are unique for each cardinality up to uncountability. Assigning these their canonical base measures will provide a suitable setting of measure spaces for characterizing densities.

However, the Radon-Nikodym Theorem requires that the sample space admit not only a measure but a σ-finite (Definition 20) base measure. Proposition 2 and Proposition 3 will therefore characterize the algebraic operations under which σ-finite measure spaces are closed. Proposition 2 below will characterize the base measures for joint probability densities.

Proposition 2 (σ-finite measure spaces have finite direct products)**.** *Let $I \in Ob(\mathbf{FinSet})$ be a set and let there be an I-indexed family of σ-finite measure spaces $(X_i,\Sigma_{X_i},\mu_{X_i})_{i\in I} \in Ob(\mathbb{M})$. Then there exists a σ-finite* direct product *measure space $\bigotimes_{i\in I}(X_i,\Sigma_{X_i},\mu_{X_i}) = (X,\Sigma_X,\mu_X)$.*

Proof. The product $\bigotimes_{i\in I}(X_i,\Sigma_i) \in Ob(\mathbf{Meas})$ exists thanks to **Meas** being Cartesian, so that the resulting set is that of Cartesian products and the σ-algebra is also that of Cartesian products. Letting π_i be the projection indexed by $i \in I$ of a Cartesian product, we write the σ-finite product measure (which exists and is unique when $(X_i,\Sigma_{X_i},\mu_{X_i})$ are σ-finite [33][7]) as $\mu_X(\sigma_X) = \prod_{i\in I}\mu_{X_i}(\{\pi_i(x) : x \in \sigma_X\})$, yielding the direct product $(\bigotimes_{i\in I}X_i, \bigotimes_{i\in I}\Sigma_{X_i}, \mu_X) \in Ob(\mathbb{M})$. □

The reader can check that the direct product of measure spaces does not form a categorical product: the pairing required to witness the universal property will not be measure-preserving, with intervals of different lengths in the real line providing a counterexample.

Proposition 3 will then characterize the base measures for mixture probability densities.

Proposition 3 (σ-finite measure spaces have countable direct sums [11][8])**.** *Let $I \in Ob(\mathbf{Set})$ be a countable set and $(X_i,\Sigma_{X_i},\mu_{X_i})_{i\in I} \in Ob(\mathbb{M})$ be a family of σ-finite measure spaces indexed by I. Then there exists a σ-finite* direct sum *measure space $\bigoplus_{i\in I}(X_i,\Sigma_{X_i},\mu_{X_i}) \in Ob(\mathbb{M})$.*

Proof. The direct sum $\bigoplus_{i\in I}(X_i,\Sigma_{X_i},\mu_{X_i}) = (X,\Sigma_X,\mu_X) \in Ob(\mathbb{M})$ of the indexed family consists of the set $X = \bigcup_{i\in I}(X_i \times \{i\})$, the σ-algebra $\Sigma_X = \{\sigma_X : \sigma_X \subseteq X, \forall i \in I, \{x : (x,i) \in \sigma_X\} \in \Sigma_{X_i}\}$, and the sum measure $\mu_X(\sigma_X) = \sum_{i\in I}\mu_{X_i}(\{x : (x,i) \in \sigma_X\})$. □

The reader can check that the direct sum of measure spaces does not form a categorical coproduct: the copairing required to witness the universal property will not be measure-preserving.

The above propositions characterized the algebra of σ-finite measure spaces, which thus now requires base cases. Restricting our attention to the standard Borel spaces, we can take the singleton set $(I,\mathscr{B}(I),\mu_\#)$ equipped with the counting measure $\mu_\#$ and the real line $(\mathbb{R},\mathscr{B}(\mathbb{R}),\lambda)$ with the Lebesgue measure as those base cases. An n-fold or countable direct sum of the singleton set gives finite and countable discrete measure spaces, whose counting measure is σ-finite, while an n-fold product of the real line gives the Euclidean spaces, whose n-dimensional Lebesgue measures are σ-finite. Definition 9 will therefore formally give the class of measure spaces suitable for forming probability densities.

Definition 9 (σ-finite standard Borel measure space)**.** *The subcategory $\mathbb{M}_{\mathscr{B}} \subset \mathbb{M}$ restricts the category of measure spaces to the σ-finite* standard Borel *measure spaces freely generated by finite direct products $\otimes$ (Proposition 2) and countable direct sums $\oplus$ (Proposition 3) of the counting-measured singleton space $(\mathbb{1},\mathscr{B}(\mathbb{1}),\mu_\#)$ and the Lebesgue-measured reals $(\mathbb{R},\mathscr{B}(\mathbb{R}),\lambda)$.*

Definition 9 covers the most common sample spaces and their base measures, as instances of a more general construction assigning base measures to finite-dimensional manifolds as sample spaces for probability densities [27]. The above only allows finite products, since the product-of-Lebesgues measure on the Hilbert cube $\mathbb{R}^{\mathbb{N}}$ (via the Borel isomorphism $\mathbb{R} \simeq [0,1]$) fails to be σ-finite [2]. The rest of the paper will therefore work with measure spaces $\mathbb{M}_{\mathscr{B}}$, whose isomorphisms preserve base measures.

[7] Definition 1.7.4, page 161

[8] 214L, page 38

Having a class of measure spaces suitable for stating probability densities with respect to count, length, area, volume, etc., Definition 10 gives the class of Markov kernels which will admit densities.

Definition 10 (Density kernel). *A* standard Borel density kernel *is a σ-finite (Definition 20) Markov kernel $f : Z \rightsquigarrow X$ whose codomain forms a σ-finite standard Borel measure space $(X, \Sigma_X, \mu_X) \in Ob(\mathbb{M}_{\mathscr{B}})$ and which is absolutely continuous $\forall z, f(z) \ll \mu_X$ with respect to the base measure μ_X*

$$\mathbf{Dens}((Z,\Sigma_Z),(X,\Sigma_X)) := \{(f,\mu_X) : Z \rightsquigarrow X \times \mathbb{M}(X) \mid (X,\Sigma_X,\mu_X) \in Ob(\mathbb{M}_{\mathscr{B}}), \forall z \in Z, f(z) \ll \mu_x\}.$$

Probability (and arbitrary measure) densities $p(x \mid z)$ also admit an alternative interpretation as measure kernels $Z \times X \times \Sigma_I \to [0,\infty]$ whose integration under the base measure yields the normalizing constant. Proposition 4 verifies that density kernels in fact admit probability densities.

Theorem 4 (Density kernels admit densities). *Every density kernel $(f,\mu_X) : Z \rightsquigarrow X \times \mathbb{M}(X)$ (Definition 10) into a standard Borel measure space admits a density with respect to the base measure μ_X.*

Proof. σ-finiteness of the kernel f and the base measure μ_X, plus absolute continuity, give the necessary conditions for the classical Radon-Nikodym theorem: a Radon-Nikodym derivative therefore exists

$$\frac{df(z)}{d\mu_X} : \mathbf{Meas}(X, \mathbb{R}_{\geq 0}) \qquad\qquad f(z)(\sigma_X) = \int_{x \in \sigma_X} \frac{df(z)}{d\mu_X}(x)\ \mu_X(dx).$$

The Radon-Nikodym derivative is the measure-theoretic notion of a probability density function

$$\frac{df}{d\mu_X} : \mathbf{Meas}(Z \times X, \mathbb{R}_{\geq 0}), \qquad p_f(\cdot \mid \cdot) : \mathbf{Meas}(X \times Z, \mathbb{R}_{\geq 0}), \qquad p_f(x \mid z) := \frac{df(z)}{d\mu_X}(x).$$

The conditions on density kernels are therefore sufficient to yield probability densities. □

Despite the hom-set notation used for convenience, density kernels do not form a category: identity Markov kernels are Dirac delta measures that only admit densities in discrete spaces. They do, however, support all compositional structure under which the resulting base measure still indexes a standard Borel measure space. Definition 11 lays the foundation for this structure.

Definition 11 (Precomposition of a density kernel). *Given a density kernel $(f,\mu_X) : Z \rightsquigarrow X \times \mathbb{M}(X)$ and a Markov kernel $h : W \rightsquigarrow Z$, their* precomposition *is $(f,\mu_X) \circ_{\mathbf{Dens}} h = (f \circ_{\mathbf{Stoch}} h, \mu_X)$.*

The above precomposition gives a definition for the composition of two density kernels: given (f,μ_X) and (g,μ_Y) their composite will just be $(g \circ f, \mu_Y)$. The existence of precomposition supports a product and coproduct algebra of density kernels, as expected based on the probability algebra itself.

Theorem 5 (Density kernels admit products and coproducts). *Density kernels have products $(f,\mu_X) \otimes (g,\mu_Y)$ and coproducts $(f,\mu_X) \oplus (g,\mu_Y)$, witnessed by a pairing and copairing.*

Proof. Any two density kernels $(f,\mu_X) : \mathbf{Dens}((Z,\Sigma_Z),(X,\Sigma_X))$ and $(g,\mu_Y) : \mathbf{Dens}((Z,\Sigma_Z),(Y,\Sigma_Y))$ admit a pairing via precomposition with copying and the product measure space $(X,\Sigma_X,\mu_X) \otimes (Y,\Sigma_Y,\mu_Y) = (X \times Y, \Sigma_X \times \Sigma_Y, \mu_X \otimes \mu_Y) \in Ob(\mathbb{M}_{\mathscr{B}})$

$$(\mathbf{copy}_Z \mathbin{⨟} (f \otimes g), \mu_x \otimes \mu_Y) : \mathbf{Dens}((Z,\Sigma_Z),(X,\Sigma_X) \otimes (Y,\Sigma_Y)).$$

Any two density kernels $(f,\mu_Y) : \mathbf{Dens}((Z,\Sigma_Z),(Y,\Sigma_Y))$ and $(g,\mu_Y) : \mathbf{Dens}((X,\Sigma_X),(Y,\Sigma_Y))$ also admit a copairing $\binom{(f,\mu_Y)}{(g,\mu_Y)} = \left(\binom{f}{g}, \mu_Y\right)$ via the copairing of their Markov kernels in **Stoch**. □

The above theorems demonstrate that density kernels represent probability densities compositionally. However, density kernels do not admit post-composition with arbitrary Markov kernels. Section 3.3 will remedy this issue by applying density kernels to generate the residuals in joint Markov kernels.

3.3 Joint densities over joint distributions

Density kernels are not closed under pushforwards, and they do not form a category. **Joint** cannot apply directly to them. Definition 12 therefore gives an appropriate definition for joint density kernels.

Definition 12 (Joint density kernel)**.** *A* joint density kernel *between objects $Z,X \in Ob(\mathbf{Stoch})$ is a pair of a density kernel into $(M,\Sigma_M,\mu_M) \in Ob(\mathbb{M}_{\mathscr{B}})$ with a deterministic map parameterized by the residual*

$$\partial\mathbf{Joint}(Z,X) := \{((f,\mu_M),[M,k]) : \mathbf{Dens}(Z,M) \times \mathbf{Para}_{\otimes}(\mathbf{Stoch}_{det})(Z,X) \mid (M,\Sigma_M,\mu_M) \in Ob(\mathbb{M}_{\mathscr{B}})\}.$$

Hom-set notation once again implies these kernels form a category, which in fact they will. First, Corollary 6 shows density kernels are closed under the joint distribution construction of Equation 1.

Corollary 6 (Density kernels admit joint distributions as conditional products)**.** *Given a density kernel $(f_1,\mu_{M_1}) : \mathbf{Dens}(Z,M_1)$, a measurable map $k_1 : \mathbf{Stoch}_{det}(Z \otimes M_1, X)$, and a density kernel $(f_2,\mu_{M_2}) : \mathbf{Dens}(X,M_2)$, composing them according to the diagram in Equation 1 forms a joint density kernel*

$$\left(\mathbf{copy}_Z \fatsemi ((\mathbf{copy}_{M_1} \circ f_1) \otimes id_Z) \fatsemi (id_{M_1} \otimes (f_2 \circ k_1)), \mu_{M_1} \otimes \mu_{M_2}\right) : \mathbf{Dens}(Z, M_1 \otimes M_2).$$

Theorem 7 will show that joint density kernels form a category, and characterize them as joint Markov kernels with the extra data of a base measure on the residual.

Theorem 7 (Joint density kernels form a category)**.** *Joint density kernels $\partial\mathbf{Joint}$ form a wide subcategory of the restriction $\mathbf{Joint}_{\mathbf{BorelStoch}}$ of* **Joint** *to standard Borel Markov kernels in* **BorelStoch***.*

Proof. First we show the joint density kernels form a subcategory, then show that subcategory is wide.

Corollary 6 shows that density kernels are closed under the composition of **Joint** (Equation 1), and so along with the obvious identity morphisms and associativity law they form a category. Theorem 5 shows that this category inherits the product and coproduct structure of **Joint**. The structure morphisms in **Joint** all have the unit I for their residual, which admits a trivial density as a finite standard Borel space; $\partial\mathbf{Joint}$ therefore inherits the copy/delete structure of **Joint**. This implies $\partial\mathbf{Joint} \subset \mathbf{Joint}_{\mathbf{BorelStoch}}$.

Objects and structure morphisms are inherited from **Joint**, so the subcategory is wide. □

The theorem above gives a copy/delete categorical structure for joint density kernels, whose base and probability measures will be σ-finite (Definition 20) as conditions for Radon-Nikodym. There is then a precise class of measures formed by pushing forward a σ-finite measure [34]: the s-finite measures (Definition 23). Proposition 4 shows that such s-finite measure kernels form a copy/delete category.

Proposition 4 (s-finite measure kernels form a CD-category [8][9])**.** *s-finite measure kernels (Definition 23) between measurable spaces form a CD-category* **sfKrn** *with $Ob(\mathbf{sfKrn}) = Ob(\mathbf{Meas})$ and hom-sets given by $\mathbf{sfKrn}((Z,\Sigma_Z),(X,\Sigma_X)) = \{f : Z \times \Sigma_X \to [0,\infty] \mid \forall z, f(z) \text{ is } s\text{-finite}\}$.*

sfKrn only forms a copy/delete category, not a Markov category, since different measure kernels may have different normalizing constants, including an infinite normalizing constant. Corollary 8 shows that restricting to probability kernels forms a Markov category.

Corollary 8 (s-finite probability kernels form a Markov category)**.** *The s-finite* probability *kernels f : $\mathbf{sfKrn}((Z,\Sigma_Z),(X,\Sigma_X))$, for which $\forall z \in Z, f(z,X) = 1$, form a Markov category* **sfStoch** $\subset$ **Stoch***.*

Proof. The restriction of all kernels to normalize to measure 1 renders every map $\mathbf{del}_Z$ unique, making I a terminal object and the resulting subcategory **sfStoch** a Markov category. □

[9] Example 7.2

Having a categorical setting capturing the Markov kernels used in computable applications, the remainder of this paper will interpret morphisms in $\partial\mathbf{Joint}$ into s-finite Markov kernels $\mathbf{sfStoch}(Z,X)$ with densities $\mathbf{sfKrn}(Z\otimes X,I)$. Theorem 9 shows that the joint Markov kernels of $\partial\mathbf{Joint}$ are s-finite and admit densities jointly measurable in the parameter and the residual.

Theorem 9 (Joint density kernels give s-finite probability kernels and densities)**.** *Joint density kernels* $(f,[M,k]):\partial\mathbf{Joint}((Z,\Sigma_Z),(X,\Sigma_X))$ *admit probability kernels* $p:\mathbf{sfStoch}((Z,\Sigma_Z),(X,\Sigma_X))$ *marginalizing out their randomness and probability densities* $p_f(\cdot\mid\cdot):\mathbf{sfKrn}((Z,\Sigma_Z)\otimes(M,\Sigma_M),I)$*.*

Proof. Any density kernel $f:\mathbf{Dens}((Z,\Sigma_Z),(M,\Sigma_M))$ gives a σ-finite probability measure and any $(M,k):\mathbf{Para}_{\otimes}(\mathbf{Stoch}_{det})(Z,X)$ pushes it forward. Every pushforward of a σ-finite Markov kernel is s-finite (Proposition 5), so $\partial\mathbf{Joint}$ consists entirely of s-finite joint Markov kernels. Being s-finite, joint density kernels admit the required probability kernels $p:\mathbf{sfStoch}((Z,\Sigma_Z),(X,\Sigma_X))$ with $p(z,\sigma_X)=f(z,k(z)^{-1}(\sigma_X))$ and densities $p_f(\cdot\mid z):M\times\Sigma_I\to[0,\infty]$ measurable in z and m. Proposition 4 defines these as the Radon-Nikodym derivative $p_f(m\mid z)(\{*\})=\frac{df(z)}{d\mu_M}(m)$. □

Theorem 7 and Theorem 9 finally gives a desirable categorical setting: one which supports composition, products, and coproducts as a copy/delete category should, while decomposing into a deterministic causal mechanism applied to a random variable with a joint density as a structural causal model should. Section 4 will put together the machinery in this section with existing work on factorizing string diagrams syntactically to interpret those factorizations as generalizing directed graphical models.

4 Diagrams as causal factorizations of joint distributions and densities

This section demonstrates that string diagrams with factorized densities support the full "ladder of causation" [25] as probabilistic models: factorized distributions, interventions, and counterfactual queries. Section 3 presented the $\partial\mathbf{Joint}$ construction for building up joint densities while still expressing arbitrary pushforward measures over them. Reasoning about directed graphical models or probabilistic programs compositionally requires providing a graphical syntax interpretable into $\partial\mathbf{Joint}$. Recent work [14, 15] treated a combinatorial syntax of string diagrams as generalized causal models. This section first reviews the definitions of a generalized causal model and its factorization of a Markov kernel, then applies that syntax to this paper's novel constructions. Doing so will enable show that via generalized causal models, joint density kernels admit factorization of their densities (Theorem 10), interventional distributions (Theorem 11), and counterfactual distributions (Theorem 12).

Generalized causal models [14] provide several advantages over causal Bayesian networks as a representation of causal structure in probability models. They allow for global inputs to and outputs from a causal model, making explicit the interface necessary to reason compositionally about causal structures. It also makes explicit the grouping of "nodes" (in the underlying graph or hypergraph) into Markov kernels, clarifying how the joint distribution decomposes into random variables and causal mechanisms.

Definition 13 will now describe a generalized causal model.

Definition 13 (Generalized causal model [15])**.** *A* generalized causal model φ *over* $\Sigma\in\mathbf{FinHyp}$[10] *is a string diagram* $p\to\mathrm{dom}(\tau)\leftarrow q:\mathbf{FreeMarkov}_{\Sigma}(n,m)$ *for* $n,m\in\mathbb{N}$ *with a bijection* q *on wires.*

Any generalized causal model $p\to\mathrm{dom}(\tau)\leftarrow q$ is equivalent to a morphism [14]

$$\varphi:\mathbf{FreeMarkov}_{\Sigma}\left(\bigotimes_{i=1}^{n}\tau(p(i)),\bigotimes_{j=1}^{m}\tau(q(j))\right).$$

[10] see Appendix C

Definition 14 will capture factorization of a Markov kernel by a generalized causal model; Fritz and Klinger [14] called it causal compatibility in their Definition 11.

Definition 14 (Factorization of a Markov kernel by a causal model [14])**.** *A* factorization (f, φ, F) *in* **Stoch** *consists of a morphism with decomposed domain and codomain* $f : \mathbf{Stoch}\left(\bigotimes_{i=1}^{n} D_i, \bigotimes_{j=1}^{m} C_j\right)$*, a causal model* $\varphi : \mathbf{FreeMarkov}_{\Sigma}(n,m)$*, and a strict Markov functor* $F : \mathbf{FreeMarkov}_{\Sigma} \to \mathbf{Stoch}$ *such that* $f = F(\varphi)$*,* $\forall i \in [1..n], D_i = F(\mathrm{dom}(\varphi)_i)$*, and* $\forall j \in [1..m], C_j = F(\mathrm{cod}(\varphi)_j)$*.*

The joint density kernels $\partial\mathbf{Joint}(Z,X)$ have an important difference from the simple Markov kernels factorized by generalized causal models in Definition 14: the density to factorize is not over $x \in X$ but over the extra structure of the residual $m \in M$. This subsection will show how to add this extra structure to a factorization, then show how to access that structure to show that generalized causal models over joint density kernels support causal inference as such: interventions and counterfactual reasoning.

Definition 15 will require a factorization to label each box's residual to apply to joint Markov kernels.

Definition 15 (Joint factorization functor)**.** *A* joint *factorization functor for a signature* $\Sigma \in \mathbf{FinHyp}$ *is a labeling of boxes with residual wires* $r : B(\Sigma) \to W(\Sigma)^*$ *and a strict Markov functor* $F : \mathbf{FreeMarkov}_{\Sigma} \to \mathbf{Joint}$ *respecting* $\forall b \in B(\Sigma), F(b) = ([\bigotimes_{w \in r(b)} F(w), k], f) : \mathbf{Joint}(F(\mathrm{dom}(b)), F(\mathrm{cod}(b)))$*.*

Joint factorizations label residuals in the signature and also map to joint density kernels. Theorem 10 shows they factorize the implied joint density of a causal model.

Theorem 10 (Joint density kernels admit factorized densities)**.** *Given a signature* $\Sigma \in \mathbf{FinHyp}$*, a strict Markov functor* $F : \mathbf{FreeMarkov}_{\Sigma} \to \partial\mathbf{Joint}$ *gives a joint density* $p_f(\cdot \mid \cdot \in F(\mathrm{dom}(\varphi)))$ *for every causal model* $\varphi : \mathbf{FreeMarkov}_{\Sigma}(n,m)$*.*

Proof. Definition 15 requires for any sub-diagram $\varphi' \subseteq \varphi$ there will be some $F(\varphi') = (f, [M,k])$. Theorem 9 then gives a density over the residual, while the functoriality of F and Corollary 6 together imply that products of individual joint-densities yield the complete joint density. □

Theorem 11 then shows that by assigning boxes optional points in their codomains, joint factorizations also admit interventional distributions.

Theorem 11 (Joint factorizations admit interventional distributions)**.** *Consider a joint factorization* (f, φ, F) *over a signature* Σ*. Then any intervention* $\mathbf{do} : \prod_{b:B(\Sigma)} I \oplus \mathscr{C}_{det}(I, F(\mathrm{cod}(b))$ *induces a functor* $\mathrm{Int} : \mathbf{FreeMarkov}_{\Sigma} \to \mathbf{Joint}$ *and an interventional distribution* $\mathrm{Int}(\varphi)$*.*

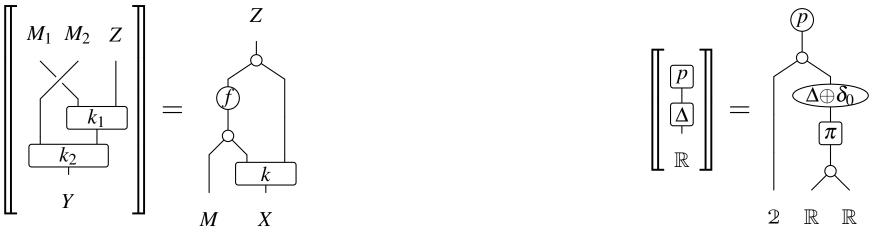
Proof. Any single-box free string diagram has an image $F(\langle b \rangle)$. We define the required functor $\mathrm{Int} : \mathbf{FreeMarkov}_{\Sigma} \to \mathbf{Joint}$ by extension of a hypergraph morphism $\alpha : \Sigma \to \mathrm{hyp}(\mathbf{Joint})$ following Fritz and Liang [15] (see their Remark 4.3). α will be identity on wires and intervene on boxes

$$\begin{aligned} &\alpha(b) : B(\Sigma) \to B(\mathrm{hyp}(\mathbf{Joint})) \\ &\alpha(b) = \begin{cases} \mathrm{hyp}(([I, \mathbf{del}_{\mathrm{dom}(b)}], \mathbf{del}_{\mathrm{dom}(b)} \fatsemi x)) & \mathbf{do}(b) = \mathbf{inr}(x) \\ \mathrm{hyp}(F(\langle b \rangle)) & \mathbf{do}(b) = \mathbf{inl}(I) \end{cases}. \end{aligned}$$

□

Finally, Theorem 12 employs similar reasoning to model counterfactual queries over jointly factorized causal models, given fixed values for random variables and an intervention.

Theorem 12 (Joint factorizations give counterfactuals)**.** *Consider a signature* $\Sigma \in \mathbf{FinHyp}$ *and a joint factorization* (f, φ, F)*. Then any intervention* $\mathbf{do} : \prod_{b:B(\Sigma)} I \oplus \mathscr{C}_{det}(I, F(\mathrm{cod}(b))$ *and any assignment* $U : B(\Sigma) \to [0,1]$ *of uniform random variates to boxes induces a functor* $\mathrm{If} : \mathbf{FreeMarkov}_{\Sigma} \to \mathbf{Joint}$ *and a counterfactual distribution* $\mathrm{If}(\varphi)$*.*



(a) The wiring diagram in ∂**Joint** of a mixture model between a delta and a Gaussian, and its image in **Stoch** with a coproduct projection

(b) A Markov kernel in ∂**Joint** projecting a sample from the uniform circle through a linear transformation, and its image in **Stoch**

Figure 1: Example joint density kernels (Definition 12): a mixture model between a constant and a Gaussian distribution depending on a coin flip (*left*) and a Markov kernel projecting a random angle onto a parametrically skewed ellipse (*right*). The $[\![\cdot]\!]$ functor (Corollary 3) maps into **Stoch**.

Proof. We work as above, but this time explicitly consider the structure of the image $F(\langle b \rangle) = (f, [M,k])$. f gives a standard Borel probability measure, so the Randomization Lemma [3] demonstrates an isomorphism of f with a pushforward $f(\cdot) \simeq g(\cdot,z)_*(U)(du)$ of the uniform distribution $U(du)$ by a deterministic map $g(\cdot,z)$. Our hypergraph morphism utilizes that fact

$$\alpha(b) = \begin{cases} \mathrm{hyp}(([I, \mathbf{del}_{\mathrm{dom}(b)}], \mathbf{del}_{\mathrm{dom}(b)} \mathbin{⨾} x)) & \mathbf{do}(b) = \mathbf{inr}(x) \\ \mathrm{hyp}(\delta_{U(b)}(g(b,\cdot))) & \mathbf{do}(b) = \mathbf{inl}(I), F(\langle b \rangle)_2 \simeq \int_{u \in [0,1]} g(u,\cdot)\, U(du) \end{cases}. \qquad \square$$

Together, Theorems 10, 11, and 12 demonstrate that joint density kernels, jointly factorized by a generalized causal model, support the properties that have made directed graphical models so widely useful. With these theorems as "sanity checks", Section 5 will summarize the paper's overall contributions, give some worked examples applying ∂**Joint**, and discuss future work.

5 Discussion

This paper started from the existing work on copy/delete categories, Markov categories, and the factorization of morphisms in those categories by generalized causal models. From there, Section 3 constructed a novel Markov category **Joint** whose morphisms keep internal track of the joint distribution they denote, defined a subcategory ∂**Joint** $\subset$ **Joint** whose morphisms support only joint densities over standard Borel spaces as their internal distributions. Section 4 then demonstrated that **Joint** supports factorization by generalized causal models, that these factorize joint densities ∂**Joint**, and that these support the interventional and counterfactual reasoning necessary for causal inference. This section will discuss some short worked examples of using ∂**Joint** for real probability models (Section 5.1), and then move on to speculate what future work could spring from the paper's developments (Section 5.2).

5.1 Worked examples

The previous sections have focused on formalism. Section 3 defined a Markov category ∂**Joint** of joint density kernels in **Stoch** (rather than the typical restriction to **FinStoch**) whose residuals (by construction) admit probability densities. Section 4 then established that the generalized causal models recently described in the categorical probability literature can indeed apply to ∂**Joint** morphisms, factorizing their joint densities and providing for causal reasoning. This subsection will apply the ∂**Joint** formalism to the models shown in Figure 1, taken from Wu et al [36] and Radul and Alexeev [27].

Figure 1a shows a generative model in which we detect fake coins by placing an even number of coins on a well-calibrated balance. The presence of a fake coin, whose weight deviates from the others, will tip the balance away from the neutral position. p determines whether the a fake coin is present, which in turn determines whether the balance position is distributed according to a Gaussian $\Delta \sim \mathscr{N}(1, 0.5)$ or according to a Dirac measure $\Delta \sim \delta_0$. The joint distribution shown on the right-hand side of the equation admits a density with respect to the standard Borel measure space $(\mathbb{2}, \mathscr{B}(\mathbb{2}), \mu_{\#}) \otimes ((\mathbb{R}, \mathscr{B}(\mathbb{R}), \lambda) \oplus (\mathbb{1}, \mathscr{B}(\mathbb{1}), \mu_{\#}))$, whereas the marginal on $\mathbb{R}$ lacks a density for the Lebesgue measure λ.

Figure 1b shows the example from Radul and Alexeev [27] in which a sample from $U(0, 2\pi)$ is projected onto a non-isotropic ellipse. Those authors calculate a probability density on the ellipse via the projection's Jacobian. Figure 1b shows the two components of a ∂**Joint** morphism: how the uniformly random angle U and a linear transformation $\mathbb{R}^{2\times 2}$ parameterize the the geometric projection k. The equation shows how $[\![\cdot]\!]$ maps the single box in ∂**Joint** (left) to the Markov kernel in **Stoch** (right).

The two examples in Figure 1 both show how the ∂**Joint** construction can compactly encode complex, parameterized joint probability densities linked by deterministic causal mechanisms. Section 5.2 will discuss potential future work extending this paper's construction and conclude.

5.2 Future work and conclusion

This paper's mathematical constructions could generalize or be strengthened in a number of ways. It would be desirable to obtain a category in which Markov kernels admit common-sense densities without having to separate into a density over a standard Borel space and a pushforward through a deterministic map; the Lebesgue decomposition of arbitrary measures into mutually singular absolutely-continuous, diffuse, and atomic portions suggests a possible route to that goal. Up to a normalization constant, every reference measure in $\mathbb{M}$ is a Hausdorff measure. This suggests densities could be obtained by considering manifolds, standardizing on the Hausdorff measure as Radul and Alexeev [27] suggest, and then defining density kernels on that foundation. Finally, Definition 9 forms an endofunctor in the category of measure spaces whose algebras and coalgebras may prove of interest. For example, recent work by Dash [10] explored defining probability measures on quasi-Borel spaces as pushforwards of a uniform distribution on the Hilbert cube, an element of the endofunctor's terminal coalgebra.

Future work can go in a number of directions to unify the formalisms of applied probabilistic reasoning. Instantiating this paper's constructions in a Markov category in which all randomness arises from an independent noise source would transform any causal factorization of a joint (density) kernel into a structural causal model [24], unifying causal Bayes nets with structural equation models. In the application area of probabilistic programming, this paper has only described "first-order" probabilistic programming languages lacking general stochastic recursion [22], corresponding to non-closed Markov categories. A combinatorial syntax for hierarchical string diagrams [1] would extend our reasoning in this paper to the closed Markov categories such as **QBS** [18] that provide denotations for higher-order probabilistic programming languages. We intend to extend this paper's formalism to categorify Sequential Monte Carlo methods [23] for generalized causal models of unnormalized distributions. We aim to apply the ∂**Joint** construction alongside recent work on unique name generation [28] to model heterogeneous tracing in probabilistic programming. Recent work on free string diagrams [35] has also suggested ways to map from free string diagrams to free diagrams of optics; equipping joint density kernels with optic structure would follow up on the work of Smithe [31] and Schauer [29].

Acknowledgements We would like to thank the anonymous reviewers for their feedback and advice in refining the paper for camera-ready. We would also like to thank the ACT 2023 program chairs for their careful shepherding of the review process. We thank Tobias Fritz, Luke Ong, Sam Staton, and Matthijs Vákár for laying the categorical foundations of s-finite Markov kernels. Finally, we would like to extensively thank Alex Lew for early discussions and cooperation on preliminary work to this paper.

References

[1] Mario Alvarez-picallo, Dan Ghica, David Sprunger & Fabio Zanasi (2022): *Rewriting for Monoidal Closed Categories*. In: *7th International Conference on Formal Structures for Computation and Deduction (FSCD 2022)*, 228, Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany, pp. 29:1–29:0, doi:10.4230/LIPIcs.FSCD.2022.29.

[2] Richard Baker (1991): *"Lebesgue measure" on $\mathbb{R}^\infty$*. *Proceedings of the American Mathematical Society* 113(4), pp. 1023–1029.

[3] V. I. Bogachev (2007): *Measure theory*. Springer, Berlin; New York.

[4] Filippo Bonchi, Fabio Gadducci, Aleks Kissinger, Paweł Sobociński & Fabio Zanasi (2016): *Rewriting modulo symmetric monoidal structure*. In: *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*, ACM, New York NY USA, p. 710–719, doi:10.1145/2933575.2935316. Available at `https://dl.acm.org/doi/10.1145/2933575.2935316`.

[5] Matteo Capucci, Bruno Gavranović, Jules Hedges & Eigil Fjeldgren Rischel (2021): *Towards foundations of categorical cybernetics*. In: *Applied Category Theory Conference (ACT 2021)*, EPTCS, pp. 235–248. Available at `http://arxiv.org/abs/2105.06332`.

[6] Matteo Capucci & Bruno Gavranović (2022): *Actegories for the Working Amthematician*.

[7] Nick Chater, Joshua B Tenenbaum & Alan Yuille (2006): *Probabilistic models of cognition: Conceptual foundations*. *Trends in cognitive sciences* 10(7), pp. 287–291.

[8] Kenta Cho & Bart Jacobs (2019): *Disintegration and Bayesian inversion via string diagrams*. *Mathematical Structures in Computer Science* 29(7), pp. 938–971.

[9] Kyle Cranmer, Johann Brehmer & Gilles Louppe (2020): *The frontier of simulation-based inference*. *Proceedings of the National Academy of Sciences* 117(48), pp. 30055–30062.

[10] Swaraj Dash, Younesse Kaddar, Hugo Paquet & Sam Staton (2023): *Affine monads and lazy structures for bayesian programming*. *Proceedings of the ACM on Programming Languages* 7(POPL), pp. 1338–1368.

[11] David H. Fremlin (2010): *Measure theory. 2: Broad foundations*, 2. ed edition. Torres Fremlin, Colchester.

[12] Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck & Giovanni Pezzulo (2017): *Active inference: a process theory*. *Neural computation* 29(1), pp. 1–49.

[13] Tobias Fritz (2020): *A synthetic approach to Markov kernels, conditional independence and theorems on sufficient statistics*. *Advances in Mathematics* 370, p. 107239.

[14] Tobias Fritz & Andreas Klingler (2023): *The d-Separation Criterion in Categorical Probability*. *Journal of Machine Learning Research* 24(46), pp. 1–49.

[15] Tobias Fritz & Wendong Liang (2023): *Free gs-Monoidal Categories and Free Markov Categories*. *Applied Categorical Structures* 31(2), p. 21, doi:10.1007/s10485-023-09717-0.

[16] Giorgio Gallo, Giustino Longo, Stefano Pallottino & Sang Nguyen (1993): *Directed hypergraphs and applications*. *Discrete Applied Mathematics* 42(2–3), p. 177–201, doi:10.1016/0166-218X(93)90045-P.

[17] Michèle Giry (1982): *A categorical approach to probability theory*. In B. Banaschewski, editor: *Categorical Aspects of Topology and Analysis*, Springer Berlin Heidelberg, Berlin, Heidelberg, p. 68–85.

[18] Chris Heunen, Ohad Kammar, Sam Staton & Hongseok Yang (2017): *A convenient category for higher-order probability theory*. In: *Proceedings - Symposium on Logic in Computer Science*, pp. 1–12, doi:10.1109/LICS.2017.8005137. ArXiv: 1701.02547 Citation Key: Heunen2017 ISSN: 10436871.

[19] Kiyosi Itô et al. (1984): *An Introduction to Probability Theory*. Cambridge University Press.

[20] Brenden M Lake, Tomer D Ullman, Joshua B Tenenbaum & Samuel J Gershman (2017): *Building machines that learn and think like people*. *Behavioral and brain sciences* 40, p. e253.

[21] Sergey Levine (2018): *Reinforcement learning and control as probabilistic inference: Tutorial and review*. *arXiv preprint arXiv:1805.00909*.

[22] Jan-Willem van de Meent, Brooks Paige, Hongseok Yang & Frank Wood (2018): *An introduction to probabilistic programming*. *arXiv preprint arXiv:1809.10756*.

[23] Christian A Naesseth, Fredrik Lindsten & Thomas B Schon (2019): *Elements of Sequential Monte Carlo*. *Foundations and Trends in Machine Learning* 12(3), pp. 187–306.

[24] Judea Pearl (2012): *The causal foundations of structural equation modeling*. *Handbook of structural equation modeling*, pp. 68–91.

[25] Judea Pearl & Dana Mackenzie (2018): *The book of why: the new science of cause and effect*. Basic books.

[26] Paolo Perrone (2019): *Notes on Category Theory with examples from basic mathematics*. *arXiv preprint arXiv:1912.10642*.

[27] Alexey Radul & Boris Alexeev (2021): *The Base Measure Problem and its Solution*. In: *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS) 2021*, 130, Proceedings of Machine Learning Research, San Diego, California, p. 3583–3591.

[28] Marcin Sabok, Sam Staton, Dario Stein & Michael Wolman (2021): *Probabilistic programming semantics for name generation*. *Proceedings of the ACM on Programming Languages* 5(POPL), pp. 1–29.

[29] Moritz Schauer & Frank van der Meulen (2023): *Compositionality in algorithms for smoothing*. *arXiv preprint arXiv:2303.13865*.

[30] Adam Ścibior, Ohad Kammar, Matthijs Vákár, Sam Staton, Hongseok Yang, Yufei Cai, Klaus Ostermann, Sean K. Moss, Chris Heunen & Zoubin Ghahramani (2017): *Denotational Validation of Higher-Order Bayesian Inference*. *Proc. ACM Program. Lang.* 2(POPL), doi:10.1145/3158148. Available at `https://doi.org/10.1145/3158148`.

[31] Toby St Clere Smithe (2020): *Bayesian updates compose optically*. *arXiv preprint arXiv:2006.01631*.

[32] Sam Staton (2017): *Commutative Semantics for Probabilistic Programming*, p. 855–879. *Lecture Notes in Computer Science* 10201, Springer Berlin Heidelberg, Berlin, Heidelberg, doi:10.1007/978-3-662-54434-1_32. Available at `https://link.springer.com/10.1007/978-3-662-54434-1_32`.

[33] Terence Tao (2011): *An introduction to measure theory*. *Graduate studies in mathematics* 126, American Mathematical Society, Providence, R.I.

[34] Matthijs Vákár & Luke Ong (2018): *On S-Finite Measures and Kernels*. Available at `http://arxiv.org/abs/1810.01837`. ArXiv:1810.01837 [math].

[35] Paul Wilson & Fabio Zanasi (2023): *Data-Parallel Algorithms for String Diagrams*. arXiv:2305.01041.

[36] Yi Wu, Siddharth Srivastava, Nicholas Hay, Simon Du & Stuart Russell (2018): *Discrete-Continuous Mixtures in Probabilistic Programming: Generalized Semantics and Inference Algorithms*. In: *Proceedings of the 35th International Conference on Machine Learning*, PMLR, p. 5343–5352. Available at `https://proceedings.mlr.press/v80/wu18f.html`.

A Measure theory background

Measure theory studies ways of assigning a "size" to a set (beyond its cardinality); these can include count, length, volume, and probability. Definition 16 begins with a nice class of measurable spaces.

Definition 16 (Standard Borel space)**.** *Let* $(X,T_X) \in Ob(\mathbf{Top})$ *be a separable complete metric space or homeomorphic to one. Equipping X with its Borel* σ*-algebra* $\mathscr{B}(X)$ *generated by complements, countable unions, and countable intersections of open subsets* $U \in T$ *yields a* standard Borel space $(X,\mathscr{B}(X)) \in Ob(\mathbf{Sbs})$*, which is also a measurable space since* $\mathbf{Sbs} \subset \mathbf{Meas}$*.*

The paper uses standard Borel spaces as a basis for its category of measure spaces (Definition 9). Example 2 is such a space.

Example 2 (The unit interval)**.** *The closed unit interval* $[0,1]$ *with its Borel* σ*-algebra of open sets* $\mathscr{B}(0,1)$ *forms a standard Borel space* $([0,1],\mathscr{B}(0,1))$*.*

Having a category of measurable spaces and some nice examples, Definition 17 formally defines what it means to assign a "size" to a measurable set.

Definition 17 (Measure)**.** *A* measure $\mu : \mathbb{M}(Z)$ *on a measurable space* $(Z,\Sigma_Z) \in Ob(\mathbf{Meas})$ *is a function* $\mu : \Sigma_Z \to [0,\infty]$ *that is null on the empty set (*$\mu(\emptyset) = 0$*) and countably additive over pairwise disjoint sets*

$$\frac{\{\sigma_k \in \Sigma_Z\}_{k\in\mathbb{N}} \quad \forall k \in \mathbb{N}, n \in \mathbb{N}, n \neq k \implies \sigma_k \cap \sigma_n = \emptyset}{\mu\left(\bigcup_{k\in\mathbb{N}} \sigma_k\right) = \sum_{k\in\mathbb{N}} \mu(\sigma_k)}$$

Reasoning compositionally about measure requires a class of maps between a domain and a codomain that form measures. The Giry monad [17] sends a measurable space (X,Σ_X) to its space of measures $\mathbb{M}(X)$ and probability measures $\mathbb{P}(X) \subset \mathbb{M}(X)$. Definition 18 defines maps into those spaces, treating the domain as a parameter space for a measure over the codomain.

Definition 18 (Measure kernel)**.** *A* measure kernel *between two measurable spaces* $(Z,\Sigma_Z),(X,\Sigma_X) \in Ob(\mathbf{Meas})$ *is a function* $f : Z \times \Sigma_X \to [0,\infty]$ *such that* $\forall z \in Z, f(z,\cdot) : \mathbb{M}(X)$ *is a measure and* $\forall \sigma_X \in \Sigma_X, f(\cdot,\sigma_X) : \mathbf{Meas}((Z,\Sigma_Z),([0,\infty],\Sigma_{[0,\infty]}))$ *is measurable.*

Measure kernels serve both to define Markov kernels below, and to form a broader class of copy/delete categories, which in Theorem 9 are seen to admit probability densities as morphisms. Definition 19 specializes to measure kernels yielding only normalized probability measures.

Definition 19 (Markov kernel)**.** *A* Markov kernel *is a measure kernel* $f : Z \times \Sigma_X \to [0,\infty]$ *whose measure is a probability measure so that* $\forall z \in Z, f(z,\cdot) : \mathbb{P}(X)$ *and* $\forall z \in Z, f(z,X) = 1$*.*

The Giry monad, restricted to probability spaces, yields Markov kernels as its Kleisli morphisms $\mathbf{Meas}((Z,\Sigma_Z),\mathbb{M}(X))$, forming the main category of Markov kernels in this paper (**Stoch**, Definition 3). Describing densities categorically then requires invoking the Radon-Nikodym Theorem, which determines when probability measures have densities. The next two definitions give the Theorem's conditions, which must be satisfied for a density to exist.

Definition 20 will formalize the condition that both the base measure and a probability measure consist of sums over countable partitions of the sample space.

Definition 20 (σ-finite measure kernel)**.** *A* σ-finite measure kernel $f : Z \times \Sigma_X \to [0,\infty]$ *is a measure kernel which at every parameter* $z \in Z$ *splits its codomain into countably many measurable sets* $X = \bigcup_{n\in\mathbb{N}} X_n \in \Sigma_X$*, each of which has finite measure* $f(z)(X_n) < \infty$*.*

Definition 21 will now formalize the further requirement that for a probability measure to admit a density function, it must have only the same null-sets as the underlying base measure.

Definition 21 (Absolute continuity)**.** *One σ-finite measure kernel $f : Z \times \Sigma_X \to [0,\infty]$ is* absolutely continuous *($f \ll g$) with respect to another σ-finite measure kernel over the same codomain $g : Y \times \Sigma_X \to [0,\infty]$ when $\forall z \in Z, y \in Y, \sigma_X \in \Sigma_X, g(y)(\sigma_X) = 0 \implies f(z)(\sigma_X) = 0$.*

The conditions in Definition 20 and Definition 21 are necessary and sufficient for the existence of a probability density via the Radon-Nikodym Theorem, as used in density kernels in Definition 10. Density kernels use measure *spaces* as their codomains: these group together the desired topology, dimensionality, and base measure. Definition 22 below formally defines measure spaces, which the paper uses in the specific form of standard Borel measure spaces (Definition 9).

Definition 22 (Measure space)**.** *A* measure space *is a pair $((X,\Sigma_X),\mu)$ of a measurable space $(X,\Sigma_X) \in Ob(\mathbf{Meas})$ with a measure $\mu : \mathbb{M}(X)$ on that space.*

The measure spaces just defined form objects in a category which Definition 8 describes. Passing from the category of measurable spaces **Meas** to the category of measure spaces $\mathscr{M}$ requires the resulting morphisms to respect the chosen measure, so that measurable sets do not “grow” or “shrink”.

Having given the conditions for densities to exist, the paper passes from density kernels to joint density kernels. Definition 23 will give a class of Markov kernels encompassing all those in this paper, particularly joint density kernels.

Definition 23 (s-finite measure kernel)**.** *An s-*finite measure kernel *$f : Z \times \Sigma_X \to [0,\infty]$ is a measure kernel (as in Definition 18 above) which decomposes into a sum of finite kernels $f = \sum_{n\in\mathbb{N}} f_n$ such that $\forall n \in \mathbb{N}, f_n : Z \times \Sigma_X \to [0,\infty]$ and $\forall n \in \mathbb{N}, \exists r_n \in \mathbb{R}_{\geq 0}, \forall z \in Z, f_n(z,X) \leq r_n$.*

Proposition 5 will demonstrate that the class of s-finite kernels (Definition 23) includes all pushforwards of σ-finite kernels, and therefore the pushforwards of all measure kernels admitting densities.

Proposition 5 (s-finite kernels are pushforwards of σ-finite kernels [34, 32])**.** *A measure kernel $f : Z \times \Sigma_X \to [0,\infty]$ is s-finite if and only if it is a pushforward $f = \mathbf{copy}_Z \mathbin{⨟} (p \otimes id_Z) \mathbin{⨟} k$ of a σ-finite measure kernel p through a deterministic k.*

The above proposition includes trivial pushforwards, so every σ-finite (Definition 20) measure kernel is s-finite (Definition 23) but not the other way around.

B Parametric and coparametric categories

This section will review the definitions of parametric and coparametric (bi)categories, first given in the categorical cybernetics literature [5]. For the sake of rigor, the reader can also see a recent review on actegories [6]. As a starting point, Definition 24 will describe how a symmetric monoidal category (SMC) can “act upon” another category functorially.

Definition 24 ($\mathscr{M}$-actegory)**.** *Consider a symmetric monoidal category $(\mathscr{M}, J, \odot)$ and a category $\mathscr{C}$. An $\mathscr{M}$-actegory is a pair of the two with a functor $\bullet : \mathscr{M} \times \mathscr{C} \to \mathscr{C}$ from the product category and natural transformations $\varepsilon : J \bullet X \simeq X$ and $\delta : (M \bullet N) \bullet X = M \bullet (N \bullet X)$.*

Definition 25 will then apply the actegory concept to define a bicategory whose morphisms accumulate parameters in the course of composition.

Definition 25 (Parametric categories [5])**.** *Given an $\mathscr{M}$-actegory $\mathscr{C}$, the* parametric (bi)category $\mathbf{Para}_\bullet(\mathscr{C})$ *has as objects those of $\mathscr{C}$ and as morphisms the pairs $\mathbf{Para}_\bullet(\mathscr{C})(A,B) = \{(M,k) \in Ob(\mathscr{M}) \times \mathscr{C}(M \bullet A, B)\}$. Composition for morphisms $(M,k) : \mathbf{Para}_\bullet(\mathscr{C})(A,B)$ and $(M',k') : \mathbf{Para}_\bullet(\mathscr{C})(B,C)$ consists of $(M' \odot M, k' \circ (id_{M'} \bullet k)))$ while identities on objects A consist of (I, id_A).*

Parametric (bi)categories of course have a dual, definable as $\mathbf{Para}_{\bullet}(\mathscr{C}^{op})^{op}$. Definition 26 will describe this category, whose morphisms admit "coparameters" accumulate extra elements of the codomain.

Definition 26 (Coparametric categories [5])**.** *Given an $\mathscr{M}$-actegory $\mathscr{C}$, the* coparametric category $\mathbf{CoPara}_{\bullet}(\mathscr{C})$ *has as objects those of $\mathscr{C}$ and as morphisms $\mathbf{CoPara}_{\bullet}(\mathscr{C})(A,B)$ the pairs $(M,f) \in Ob(\mathscr{M}) \times \mathscr{C}(A, M \bullet B)$. Composition for $(M,f) : \mathbf{CoPara}_{\bullet}(\mathscr{C})(A,B)$ and $(M',g) : \mathbf{CoPara}_{\bullet}(\mathscr{C})(B,C)$ consists of $(M \odot M', (id_M \bullet g) \circ f))$ while identities on objects A consist of (I, id_A).*

The coparametric category construction generalizes the idea of a writer monad to more than one object, and represents morphisms that "log" or "leave behind" a cumulative effect. Definition 27 will describe symmetric monoidality for the $\mathscr{M}$-actegory on $\mathscr{C}$ when $(\mathscr{C}, \otimes I)$ is symmetric monoidal.

Definition 27 (Symmetric monoidal $\mathscr{M}$-actegory)**.** *A* symmetric monoidal $\mathscr{M}$-actegory *is an $\mathscr{M}$-actegory $\mathscr{C}$ equipped with a symmetric monoidal structure and a natural isomorphism $\kappa_{M,X,Y} : M \bullet (X \otimes Y) \simeq X \otimes (M \bullet Y)$, satisfying coherence laws similar to those of a costrong comonad.*

Finally, Proposition 6 will demonstrate that given a symmetric monoidal actegory as in Definition 27, the constructions above admit symmetric monoidal structure themselves.

Proposition 6 (Parametric and coparametric categories admit monoidal structure [6][11])**.** *Given a symmetric monoidal $\mathscr{M}$-actegory $(\mathscr{C}, \otimes, I)$, the parametric bicategory $\mathbf{Para}_{\bullet}(\mathscr{C})$ and coparametric bicategory $\mathbf{CoPara}_{\bullet}(\mathscr{C})$ form symmetric monoidal bicategories $(\mathbf{Para}_{\bullet}(\mathscr{C}), \otimes, I)$ and $(\mathbf{CoPara}_{\bullet}(\mathscr{C}), \otimes, I)$.*

C Free copy/delete and Markov categories

Generalized causal models [14] employ hypergraphs, which "flip" the status of nodes and edges relative to ordinary graphs: "hypernodes" are drawn as wires and "hyperedges" connecting them as boxes. These hypergraphs represent string diagrams combinatorially; restricting hypergraphs to conditions matching certain kinds of categories defines "free" categories of those kinds. This subsection will build up free copy/delete and Markov categories with generalized causal models as morphisms.

Definition 28 defines hypergraphs via sets [16]; Bonchi et al [4] provides categorical intuition.

Definition 28 (Hypergraph)**.** *A* hypergraph *is a 4-tuple $(W, B, \text{dom}, \text{cod})$ consisting of a set of vertices, nodes, or "wires" W; a set of hyperedges or "boxes" B; a function* $\text{dom} : B \to W^*$ *assigning a domain to each box; and a function* $\text{cod} : B \to W^*$ *assigning a codomain to each box.*

We abuse notation and write individual boxes $b \in B : \text{dom}(b) \to \text{cod}(b)$.

Definition 29 specifies relabelings of one hypergraph's wires and boxes with those of another.

Definition 29 (Hypergraph morphism)**.** *Given hypergraphs G, H, a* hypergraph morphism $\alpha : G \to H$ *is a pair of functions assigning wires to wires and boxes to boxes, the latter respecting the former*

$$\mathbf{Hyp}(G,H) := \left\{ (\alpha_W, \alpha_B) \in W(H)^{W(G)} \times B(H)^{B(G)} \mid \forall b \in B(G), \alpha_B(b) : \alpha_W(\text{dom}(b)) \to \alpha_W(\text{cod}(b)) \right\}.$$

As implied by the hom-set notation, hypergraphs and their morphisms form a category **Hyp** [4], and our application will employ the full subcategory **FinHyp** in which W and B both have finite cardinality. Finally, a hypergraph H is *discrete* when $B(H) = \emptyset$; $\underline{n}$ denotes a discrete hypergraph with $n \in \mathbb{N}$ wires. Any monoidal category has a (potentially infinite) underlying hypergraph, which we denote $\text{hyp}(\cdot) : \mathbf{MonCat} \to \mathbf{Hyp}$ following Fritz and Liang [15].

[11] Example 5.1.8

Often a finite hypergraph $\Sigma \in \mathbf{FinHyp}$ denotes the generating objects and morphisms of a free monoidal category, or the primitive types and functions of a domain-specific programming language. We call such a finite hypergraph a *monoidal signature*. Definition 30 formally defines the copy/delete category freely generated by a signature Σ, which Definition 31 will restrict to free Markov categories.

Definition 30 (Free copy/delete category for the signature Σ [15])**.** *The* free CD category $\mathbf{FreeCD}_\Sigma$ *for* $\Sigma \in \mathbf{FinHyp}$ *is a subcategory* $\mathbf{FreeCD}_\Sigma \subseteq \mathbf{cospan}(\mathbf{FinHyp}/\Sigma)$ *where*

- *Objects are the pairs* $(n, \sigma) \in \mathbb{N} \times \underline{n} \to \Sigma$ *assigning outer wires of a string diagram to wires in* Σ*;*
- *Morphisms are isomorphism classes of cospans, given combinatorially*

$$\mathbf{FreeCD}_\Sigma((n,\sigma_n),(m,\sigma_m)) = \\ \{p \to \mathrm{dom}(\tau) \leftarrow q \in \mathbf{FinHyp}(\underline{n}, \mathrm{dom}(\tau)) \times Ob(\mathbf{FinHyp}/\Sigma) \times \mathbf{FinHyp}(\underline{m}, \mathrm{dom}(\tau))\},$$

such that $\tau : G \to \Sigma \in Ob(\mathbf{FinHyp}/\Sigma)$ *is a hypergraph morphism from an acyclic G and every wire* $w \in W(G)$ *has at most one "starting place" as the diagram's input or a box's output*

$$|p^{-1}(w)| + \sum_{b \in B(G)} \sum_{w' \in \mathrm{cod}(b)} \mathbb{I}[w' = w] \leq 1.$$

Intuitively, a morphism in $\mathbf{FreeCD}_\Sigma$ is syntax specifying a string diagram with no looping or merging wires, whose boxes and wires are labeled by Σ. Definition 31 passes to the free Markov category $\mathbf{FreeMarkov}_\Sigma$ just by syntactically enforcing the naturality of $\mathbf{del}_Z$.

Definition 31 (Free Markov category for the signature Σ)**.** *The* free Markov category $\mathbf{FreeMarkov}_\Sigma$ *for* $\Sigma \in \mathbf{FinHyp}$ *is the wide subcategory of* $\mathbf{FreeCD}_\Sigma$ *restricted to morphisms in which every output from every box connects to somewhere else*

$$\mathrm{connects}(w, G, q) := \mathbb{I}[\exists b \in B(G) : w \in \mathrm{cod}(b) \implies q^{-1}(w) \neq \emptyset \vee \exists b' \in B(G) : w \in \mathrm{dom}(b')]$$
$$\mathbf{FreeMarkov}_\Sigma(n,m) := \{p \to \mathrm{dom}(\tau) \leftarrow q \in \mathbf{FreeCD}_\Sigma(n,m) \mid \forall w \in W(\mathrm{dom}(\tau)), \mathrm{connects}(w, \mathrm{dom}(\tau), q)\},$$

and with composition redefined to syntactically enforce this by iterating the deletion of discarded boxes to a fixed-point after composition in $\mathbf{FreeCD}_\Sigma$.