École normale supérieure Paris-Saclay

Stage de Master 1 Hadamard

réalisé à Universität Innsbruck

Free CD Categories

Wendong Liang

Sous la direction de Tobias Fritz

August 19, 2021

Abstract

This project belongs to the newly emerging subject of categorical probability theory and theoretical statistics, which has recently turned out to be a powerful framework for proving some generalizations of classical results in probability and statistics. This suggests that this framework may eventually serve as a new foundation for these subjects, similar to how the scheme-theoretic framework has replaced classical algebraic geometry. In section 1 we summarize some recent work on categorical probability, in the language of Markov categories.

Free objects, such as free groups, play a central role throughout mathematics. While they are defined in terms of an abstract universal property which facilitates working with them to some extent, it is also often useful and important to have explicit descriptions of free objects. In Section 2, we prove such an explicit description in the context of categorical probability. More precisely, it is a combinatorial characterization of free CD categories in terms of cospans of slice categories of hypergraphs. In the context of categorical probability, the first sketch of categories generated by a DAG (directed acyclic graph) is the causal theory by Fong. On the one hand, we formalize the causal theory as the free Markov category, so as to generalize the former from DAG to DAH(directed acyclic hypergraph) and explicit the its set of morphisms; on the other hand, there are future expected applications to causal inference and to implementing categorical probability in terms of computer algebra. In Section 3 we discuss one of its applications in categorical causal inference: generalizing the inflation technique for the causal compatibility problem with latent variables, to the level of Markov categories.

2020 *Mathematics Subject Classification*. 18M30, 18M35, 62A09, 62H22.

Résumé

Ce projet fait partie du nouveau sujet de la théorie des probabilité catégorique et statistiques théoriques, qui s'est récemment révélé être un cadre efficace pour prouver certaines généralisations de résultats classiques en probabilité et en statistiques. Cela suggère que ce cadre pourrait éventuellement servir de nouvelle base pour ces sujets, de la même manière que la théorie des schémas a remplacé la géométrie algébrique classique. Dans la section 1, nous présentons quelques travaux récents sur la probabilité catégorique, dans le langage des catégories de Markov.

Les objets libres, tels que les groupes libres, jouent un rôle central dans les mathématiques. Bien qu'ils soient définis en termes de propriété universelle abstraite, ce qui facilite dans une certaine mesure le travail, il est également souvent utile et important de disposer de descriptions explicites des objets libres. Dans la section 2, nous prouvons une telle description explicite dans le contexte de la probabilité catégorique. Plus précisément, il s'agit d'une caractérisation combinatoire des catégories CD libres en termes de cospans de catégories de tranches d'hypergraphes. Dans le contexte de la probabilité catégorique, la première esquisse de catégories générées par un DAG (graphe orienté acyclique) est la théorie causale de Fong. D'une part, nous formalisons la théorie causale en tant que catégorie de Markov libre, afin de généraliser la première de DAG à DAH (hypergraphe orienté acyclique) et d'expliciter l'ensemble de ses morphismes ; d'autre part, des applications futures sont attendues pour l'inférence causale et l'implémentation de la probabilité catégorique en termes de calcul formel. Dans la section 3, nous discutons de l'une de ses applications en inférence causale catégorique : la généralisation de la technique d'inflation pour le problème de compatibilité causale avec des variables latentes, au niveau de la catégorie de Markov.

Context

My internship of 4 months took place at Innsbruck University, Austria, under the supervision of Tobias Fritz. The goal of the internship is to find some applications of Markov categories in causal inference. Markov categories, named by [Fri20] and implicitly introduced already in [Gol99] and [CJ19], is one of the synthetic ways to generalize the probablity and statistic theory on the level of category theory. A Bayesian network can be canonically seen as a morphism in a certain Markov category, just as the causal theories introduced first by [Fon13]. We thus see that there exists a direct generalization from probabilistic graphical models (including DAH [directed acyclic hypergraphs] that we will focus on in this report) to causal models. However, one question left unthought in his paper is: what are all morphisms in the Markov category freely generated by edges of a given DAG(directed acyclic graph), or even a DAH? This question does not need to be answered to unless we care about some global properties of probabilistic graphical models:

1. Given a probability distribution on observed variables, is it compatible with a given DAG, or more generally a morphism in a theory of Markov categories?
2. Is there a criterion of Markov equivalence on Markov categories, showing that two morphisms encodes the same conditional independencies?

The classical version with latent variables of the first question is solved by the "inflation technique" [WSF19], and the second by a characterization given by [VP13]. However, to solve them in Markov categories, the first (and probably the most technical) thing we need to know is an explicit (combinatorial) characterization of the set of morphisms in the free CD category generated by the given hypergraph, which is exactly the main contribution of this report. In [CJ19] a Markov category is called an affine CD category, so we obtain easily the free Markov category. After that, we formulate the generalization of the inflation technique to solve the first question. The free CD categories, which supposes no axiom in probabilistic context, could have some potential applications in the general category theory.

Acknowledgements

Foremost among these is my supervisor Tobias Fritz, whose depth and breadth of knowledge are astonishing, all the more so since he never loses patience with the trivial questions I have asked. Thanks to his clever intuition and a great amount of time dedicated to our discussion, we have finally found the right way of proving the main theorem of this report. His patience and encouragement helped me also work through the project, and probably finish by publishing an article.

I would like to thank the people who made this internship happen, in particular Frédéric Pascal and the administration of ENS Paris-Saclay and the department of mathematics of Innsbruck University. In addition, I would like to thank the jury of the ENS Paris-Saclay for putting time into reading this document and for listening to my oral presentation in Gif-sur-Yvette.

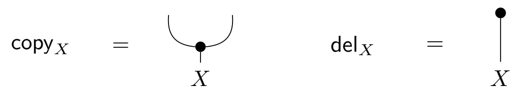
Finally, I want to thank my friends and family who helped me in many ways, in particular Samuel Arsac, with whom I have organized some online seminars for graduate students in the south of Paris working on algebra and logic to share their thought during their internships.

Contents

1 Categorical probability

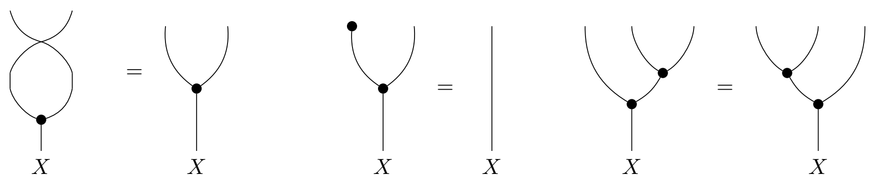
1.1 Markov categories

Definition 1.1. [CJ19] A *CD-category* is a symmetric monoidal category $(\mathcal{C}, \otimes, I)$ with a commutative comonoid copy_X, del_X for each object X, depicted in string diagrams as:



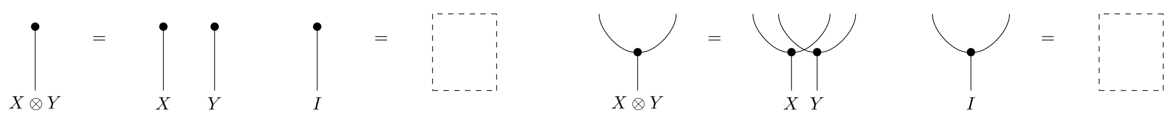
(1)

and satisfying the commutative comonoid equations:



(2)

and compatible with monoidal structure:

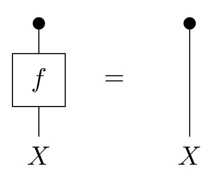


(3)

Here 'CD' stands for Copy/Discard.

Definition 1.2. A *CD-functor* is a symmetric monoidal functor between two CD-categories which preserves the CD structure: for a CD-functor F, it maps copy_c to $\mathsf{copy}_{F(c)}$, and del_c to $\mathsf{del}_{F(c)}$.

Definition 1.3. A *Markov category* is a CD-category that verifies the naturality of all counits, i.e.



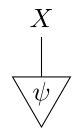
(4)

for every morphism f and every object X.

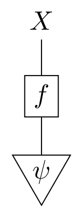
Morphisms in most of the Markov categories are Markov kernels, which explains the name. The comultiplication copy_c represents copying two outcomes without introducing any randomness; the counit del_c is the map which simply discards its input and produces no output.

We generalize the distribution or probability measure of the classical probability theory: distribution in the classical sens is a Markov kernel which takes no input and produces an output, interpreted as a "random element" of the codomain. This motivates the definition that a *distribution* in any

Markov category $\mathcal{C}$ is a morphism $\psi : I \to X$ from the unit object I to other object X, depicted in string diagram notation as

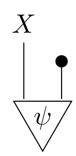


We can now say that the pair (X, ψ) is a *probability space* in $\mathcal{C}$. A *random variable* on (X, ψ) taking values in an object Y can the be defined by a morphism $f : X \to Y$. The distribution of the random variable f is then given by the composition $f\psi : I \to Y$,



so that the pair $(Y, f\psi)$ is also a probability space in $\mathcal{C}$. The formalism of Markov categories unifies the concepts of probability distributions and random variables by considering both of them as instances of the more general concept of Markov kernel.

For objects X, Y in $\mathcal{C}$, morphisms $\psi : I \to X \otimes Y$ correspond to *joint distributions*. In general, such a joint distribution can be marginalized over Y by composing with $\mathrm{id}_X \otimes \mathsf{del}_Y$, depicted in string diagram notation as



Example 1.4. [Fri20] The first example of Markov category we introduce is **FinStoch** the category of finite sets with Markov kernels. A morphism $f : X \to Y$ between finite sets X and Y is a stochastic matrix $(f_{xy})_{x \in X, y \in Y}$, where $f_{xy} \geq 0$ and $\Sigma_y f_{xy} = 1$. Composition is given by matrix multiplication. We use also the notation $f(y|x)$ for $f : X \to Y$, so that the composition is given by the Chapman-Kolmogorov equation:

$$(gf)(z|x) = \sum_y g(z|y) f(y|x) \tag{5}$$

There is an inclusion functor $F : \textbf{FinSet} \to \textbf{FinStoch}$: given a finite set function f,

$$F(f)(y|x) := \begin{cases} 1 & \text{if } y = f(x) \\ 0 & \text{otherwise} \end{cases}$$

The symmetric monoidal structure on **FinStoch** is given by cartesian product on objects, and the tensor product of stochastic matrices on morphisms: for $f : A \to X$ and $g : B \to Y$,

$$(f \otimes g)(xy|ab) := f(x|a) g(y|b) \tag{6}$$

The functor F is faithful: if $F(f) = F(g) : A \to X$, then $F(f)(x|a) = 1$ if and only if $x = f(a)$, thus for every $a \in A$, there is a unique $x \in X$ such that $x = f(a)$, f is uniquely defined.

The image of F is a subcategory in **FinStoch**: the category of finite deterministic maps.

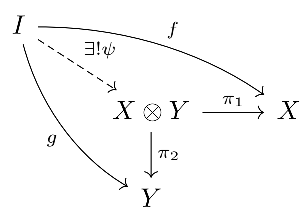
We then define the comonoidal operations on **FinStoch**, which is inherited from **FinSet**:

$$\mathsf{copy}_X(x_1, x_2|x_0) := \begin{cases} 1 & \text{if } x_1 = x_2 = x_0 \\ 0 & \text{otherwise} \end{cases}$$

$$\mathsf{del}_X(|x) := 1$$

Concretely, del_X is the all-ones row vector indexed by the elements of X.

Remark 1.5. The monoidal structure of **FinStoch** is not cartesian.
Suppose **FinStoch** is cartesian, then by the universal property of cartesian product, for I the unit of **FinStoch**, for any $f : I \to X$ and $g : I \to Y$, there exists a unique joint distribution $\psi : I \to X \otimes Y$ making the diagram commute:

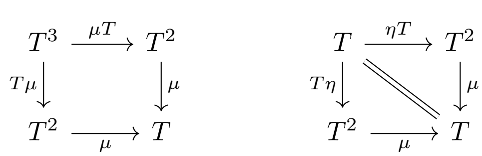


In other words, the joint distribution $I \to X \otimes Y$ is in bijection with the pair of distributions $I \to X$ and $I \to Y$. Whereas in **FinStoch** a joint distribution is not uniquely determined by its marginals. Thus morphisms in every cartesian Markov category are all deterministic (Definition 10.1 of [Fri20]).

1.2 Kleisli categories of monoidal monads

The previous example shows that cartesian categories are not interesting for the categorical probability. One way to construct non-trivial Markov categories is by Kleisli categories of symmetric monoidal affine monads.

Definition 1.6. A *monad* on a category $\mathcal{C}$ consists of an endofunctor $T : \mathcal{C} \to \mathcal{C}$, natural transformations $\eta : 1 \to T$ and $\mu : T^2 \to T$ such that for all objects $X \in \mathcal{C}$, the diagrams



commutes. It is a *symmetric monoidal monad* if it comes equipped with morphisms

$$\nabla_{X,Y} : TX \otimes TY \to T(X \otimes Y)$$

natural in X and Y, which make T a lax symmetric monoidal functor with unit $\eta : I \to TI$, and such that μ and η are monoidal transformations. The latter means that the diagrams

$$\begin{array}{ccccc} T^2X \otimes T^2Y & \xrightarrow{\nabla} & T(TX \otimes TY) & \xrightarrow{T\nabla} & T^2(X \otimes Y) \\ \downarrow{\scriptstyle \mu\otimes\mu} & & & & \downarrow{\scriptstyle \mu} \\ TX \otimes TY & & \xrightarrow{\quad\quad\nabla\quad\quad} & & T(X \otimes Y) \end{array}$$

$$\begin{array}{ccc} & X \otimes Y & \\ {\scriptstyle \eta\otimes\eta}\swarrow & & \searrow{\scriptstyle \eta} \\ TX \otimes TY & \xrightarrow{\quad\nabla\quad} & T(X \otimes Y) \end{array}$$

commute for all $X, Y \in \mathcal{C}$.

A monad is called *affine* if the monoidal unit I is terminal in $\mathcal{C}$ and $TI \cong I$.

Definition 1.7. [nLa21a] Let (T, η, μ) be a monad over a category $\mathcal{C}$. The *Kleisli category* of $\mathcal{C}$ is the category $\mathcal{C}_T$ whose objects are the objects of $\mathcal{C}$, and morphisms $M \to N$ the elements of the hom-set $\mathcal{C}(M, T(N))$ in $\mathcal{C}$, called *Kleisli morphisms*. Composition is given by the *Kleisli composition rule* $g \circ_{Kleisli} f = \mu_P \circ T(g) \circ f$ where $M \xrightarrow{f} N \xrightarrow{g} P$.

Proposition 1.8. [Fri20][Gui80] If (T, μ, η) is a symmetric monoidal monad, then the Kleisli category $\mathcal{K}\ell(T)$ becomes symmetric monoidal, with:

- the tensor product of objects being one of $\mathcal{D}$,
- the tensor product of morphisms represented by $f : A \to TX$ and $g : B \to TY$ being represented by the composite

$$A \otimes B \xrightarrow{f\otimes g} TX \otimes TY \xrightarrow{\nabla} T(X \otimes Y)$$

Moreover, the standard inclusion functor $\mathcal{D} \to \mathcal{K}\ell(T)$ is strict symmetric monoidal.

Corollary 1.9. [Fri20] Let $\mathcal{D}$ be a Markov category and T a symmetric monoidal affine monad on $\mathcal{D}$. Then the Kleisli category $\mathcal{K}\ell(T)$ is again a Markov category in a canonical way.

Proof. By the last statement of Proposition 1.8, we can transport comonoid structures along the inclusion functor: the comonoid multiplication of an object $X \in \mathcal{K}\ell(T)$ is

$$X \xrightarrow{\mathsf{copy}_X} X \otimes X \xrightarrow{\eta} T(X \otimes X)$$

and the counit of X is $X \to TI \cong I$ since the monad is affine.
Moreover, $\mathcal{D}$ is a Markov category, I is thus terminal in $\mathcal{D}$. With the fact that $TI \cong I$, I is also terminal in $\mathcal{K}\ell(T)$. $\square$

1.3 The Giry monad and the construction of Stoch

This section in an example of the previous section. We show how to construct Stoch, the most general category of Markov kernels between measurable spaces by applying the previous section to the *Giry monad*.

Let **Meas** be the category of measurable spaces (X, Σ_X) with measurable maps as morphisms. **Meas** becomes symmetric monoidal equipped with the cartesian product on sets and σ-algebra product. The Giry monad functor $P : \mathbf{Meas} \to \mathbf{Meas}$ takes a measurable space (X, Σ_X) to the set of all probability measures on (X, Σ) with the smallest σ-algebra such that the evaluation map

$$P(X, \Sigma_X) \to [0,1], \quad \mu \mapsto \mathrm{ev}_S(\mu) = \mu(S)$$

is measurable for every $S \in \Sigma_X$. The associated natural transformations of the monad are $1 \to P$ sending a point to its point measure, and $P^2 \to P$ sending a measure on the set of measures to its integral: for Q measure of X, U measurable in X,

$$\mu_X : P^2X \to PX, \quad \mu_X(Q)(U) = \int_{q \in PX} \mathrm{ev}_U(q) dQ = \int_{q \in PX} q(U) dQ$$

The category of Markov kernels between any measurable spaces Stoch is defined as $\mathcal{K}\ell(P)$. Concretely, the objects are measurable spaces, and the morphisms are Markov kernels: f is a morphism from X to Y

$$f : \Sigma_Y \times X \to [0,1], \quad (S, x) \mapsto f(S|x)$$

such that (i) $f(-|x)$ is a probability measure for any $x \in X$, (ii) $f(S|-)$ is measurable for any $S \in \Sigma_Y$. The composition with $g : \Sigma_Y \times X \to [0,1]$ is given by the Chapman-Kolmogorov equation:

$$g \circ f : \Sigma_Z \times X \to [0,1] \quad (g \circ f)(T|x) := \int_{y \in Y} g(T|y) f(dy|x) \tag{7}$$

The comultiplications in Stoch are

$$(\Sigma_X \otimes \Sigma_X) \times X \to [0,1]$$
$$(S \times T, x) \mapsto \begin{cases} 1 & \text{if } x \in A \cap B \\ 0 & \text{otherwise} \end{cases}$$

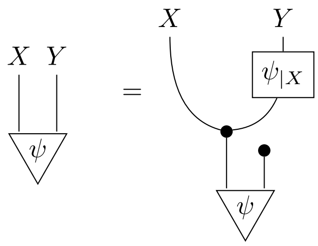
for any generating rectangle $A \times B$ in the product σ-algebra. This corresponds to assigning every $x \in X$ the measure $\delta_{x,x}$ on $X \times X$.

Lemma 1.10. [Fri20] The Giry monad is affine symmetric monoidal.

1.4 Conditionals

Markov categories could have additional properties, some of which correspond to ones in classical probablity theory. We have already generalized some axioms that make a Markov category more similar to categories of Markov kernels. In this section we only introduce the conditionals. All these definitions come from [Fri20].

Definition 1.11. [CJ19] Let $\mathcal{C}$ be a Markov category. We say that $\mathcal{C}$ has *conditional distribution* if for every distribution $\psi : I \to X \otimes Y$, there is $\psi_{|X} : X \to Y$ such that



(8)

Example 1.12. [Fri20] In FinStoch, this is equivalent to

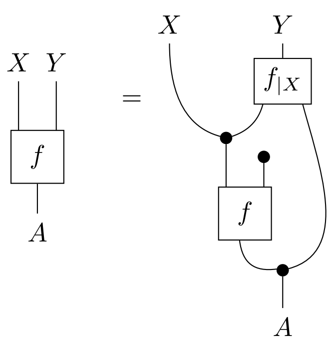
$$\psi(x, y) = \psi_{|X}(y|x)\psi(x) \tag{9}$$

In Stoch, a conditional distribution is a Markov kernel $\psi_{|X} : (X, \Sigma_X) \to (Y, \Sigma_Y)$ satisfying

$$\psi(S \times T) = \int_{x \in S} \psi_{|X}(T|x)\psi(dx) \tag{10}$$

for all $S \in \Sigma_X$ and $T \in \Sigma_Y$. Such a $\psi_{|X}$ is called a product regular conditional probability, and is known to exist whenever (Y, Σ_Y) is a Borel space, but not every space in Stoch. Thus the category of Borel spaces and Markov kernels BorelStoch has conditionals, but Stoch does not.

Definition 1.13. Let $\mathcal{C}$ be a Markov category. We say that $\mathcal{C}$ has *conditionals* if for every morphism $f : A \to X \otimes Y$, there is $f_{|X} : X \otimes A \times Y$ such that



(11)

In FinStoch, this is equivalent to

$$f(x, y|a) = f_{|X}(y|x, a)f(x|a) \tag{12}$$

1.5 Categorical Bayesian networks

Let us recall the definition of Bayesian network:

Definition 1.14. [Fon13] Let $G = (V, A, s, t)$ be a DAG (directed acyclic graph), for each $v \in V$ let X_v be a measurable space, and let P be a joint probability measure on $\prod_{v \in V} X_v$. We say that G

and P are *compatible* if there exists an ancestral ordering of the elements of V such that there exist conditionals such that

$$P(A_1 \times A_2 \times \ldots \times A_n) = \int_{A_1} \int_{A_2} \cdots \int_{A_n} P_{X_n|\mathrm{pa}(X_n)} \ldots P_{X_2|\mathrm{pa}(X_2)} P_{X_1}$$

The first attempt of generalize Bayesian network to the category theory seems to be in [Fon13], where Fong constructed a category named *causal theory* associated to a causal structure, i.e. a DAG, with objects in the free monoid V^* [1] and two types of generating morphisms: the comonoid structure and edges in DAG. In fact the causal theory he constructed is a free CD category generated by a DAG.

However, two things still need to be explored:

- the free CD category could be generated by a DAH(directed acyclic hypergraph) which is more general than a DAG. The only case where DAG could not be valid is when the Markov category does not have conditionals, which is the case for Stoch, and for some categories used in quantum cryptography.
- To solve the two questions mentioned in the beginning of the report, we need an explicit formulation of an arbitrary morphism in the free CD-category, so that we could take an arbitrary morphism without loss of generality.

In the next section, we will give a combinatorial characterization of free CD category generated by a DAH, which in our formulation, is called a *monoidal theory*.

2 Free Markov Categories

Definition 2.1. [BGK+16] The *category of finite directed hypergraphs* **Hyp** is the functor category $\mathbf{FinSet}^{\mathbf{I}}$ where $\mathbf{I}$ has an object pairs of natural numbers $(k, l) \in \mathbb{N} \times \mathbb{N}$ together with one extra object $*$; For each $k, l \in \mathbb{N}$, there are $k + l$ arrows from (k, l) to $*$, denoted respectively by in (input) and out (output), i.e. $(\mathrm{in}_1, \ldots, \mathrm{in}_k, \mathrm{out}_1, \ldots, \mathrm{out}_l)$.

Intuitively, an object G of $\mathbf{FinSet}^{\mathbf{I}}$ is the hypergraph with set of nodes $G(*)$ and, for each k, l, $G(k, l)$ is the set of hyperedges with k ordered sources and l ordered targets. For each G_1, G_2 in $\mathbf{FinSet}^{\mathbf{I}}$, the morphisms are natural transformations $p = (p_{k,l})_{k,l}$: for each k, l, the diagram

$$\begin{array}{ccc} G_1((k,l)) & \xrightarrow{e \mapsto p_{k,l}(e)} & G_2((k,l)) \\ {\scriptstyle G_1^{k,l}(i_j)}\downarrow & & \downarrow{\scriptstyle G_2^{k,l}(i_j)} \\ G_1(*) & \xrightarrow[n \mapsto p_{k,l}(n)]{} & G_2(*) \end{array}$$

commutes.

[1] In [Fon13] Fong was wrong when he said that the set of objects is $V^{\mathbb{N}}$. In fact, $V^{\mathbb{N}}$ is isomorphic to the free commutative monoid on V, in which all swaps $c \otimes c \to c \otimes c$ is $\mathrm{id}_{c \otimes c}$, which is not what we want.

Intuitively this diagram means that the j-th interface of the hyperedge e in G_1 is sent to the node that is exactly the j-th interface of the hyperedge $p(e)$ in G_2.

In the following we denote $s(e)$ as the source nodes of the edge e, and $t(e)$ the target nodes.

A hypergraph is said to be *acyclic* if for any sequence $(e_1, \ldots, e_k) \in \Sigma^k$, if $t(e_i) \cap s(e_{i+1}) \neq \emptyset$ $\forall i = 1, \ldots, k-1$, then $t(e_k) \cap s(e_1) = \emptyset$.

A hypergraph is said to be *discrete* if there is no hyperedge in it.

In the following, we use $E(G)$ to denote the set of all edges in G, to simplfy the notion $\bigcup_{k,l} G((k,l))$.

Definition 2.2. [Zan17] A *monoidal theory* is a tuple (Σ, C) of a totally ordered signature Σ and a totally ordered finite set C of colors. Elements of Σ are operations $o : w \to v$ with arity w and coarity v, with $w, v \in C^*$.

If Σ, C are finite, (Σ, C) can be seen as an object G of **Hyp**, with $G(*) = C$, $G((k,l))$ is the set of operations in Σ of k arities and l coarities.

To construct the free CD and free Markov category generated by (Σ, C), and explicit its morphisms and compositions, it is important to consider it in a more general category and find its decomposition up to isomorphism. The general category we choose is the category of props (product and permutation category[HR15]).

Definition 2.3. Let C be a set. A *C-colored prop* is a strict symmetric monoidal category (SMC) where the set of objects is C^* and the monoidal product $\oplus$ on objects is word concatenation. (Details are in 1.1 of [HR15].) All the C-colored props form a category $\mathbf{Prop}_C$, with morphisms the identity-on-objects strict symmetric monoidal functors.

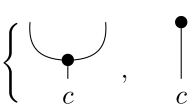
Definition 2.4. *hyp* is the forgetful functor **CDCat** $\to$ **SMC** $\to$ **Hyp**, where **SMC** is the category of symmetric monoidal categories. In **CDCat** $\to$ **SMC**, *hyp* ignores all the comonoids in the CD-category; in **SMC** $\to$ **Hyp**, *hyp* maps an SMC to its underlying hypergraph.

Example 2.5. $\mathbf{P}_{\Sigma,C}$ is defined as the *free C-colored prop* generated by (Σ, C):

Given α in $\mathbf{Hom}_{\mathbf{Hyp}}(\Sigma, hyp(\mathcal{D}))$, there is a unique $F \in \mathbf{Hom}_{\mathbf{Prop}_C}(\mathbf{P}_\Sigma, \mathcal{D})$ which maps every generator morphism e to $F(e)$.

Concretely, $\mathbf{P}_{\Sigma,C}$ has arrows the Σ-terms quotiented by the laws of SMC. Σ-terms are freely obtained by composing edges in Σ, a *unit id* : $c \to c$ for each $c \in C$ and a *symmetry* $\sigma_{c,d} : cd \to dc$ for each $c, d \in C$, by sequential (;) and parallel ($\otimes$) composition.

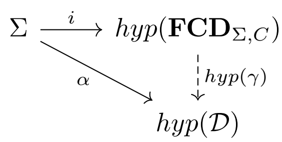
Example 2.6. The C-colored prop $\mathbf{CD}_C$ of *CD C-comonoids* is freely generated by the set of operations Σ_C^{CD}, where each $c \in C$ is equipped with a comultiplication copy_c and a counit del_c:

$\Sigma_C^{CD} = \Big\{$  $\Big| \; c \in C \Big\}$

quotiented by equations of commutative comonoid 2

When C is a singleton, C-colored props are just called props, i.e. SMCs with objects the natural numbers where the monoidal product is addition on objects.

Definition 2.7. The *free CD-category* $\mathbf{FCD}_{\Sigma,C}$ over a monoidal theory (Σ, C) is defined by the universal property: Given a CD-category $\mathcal{D}$ whose monoid of objects is $\mathcal{C}^*$, for every identity-on-nodes hypergraph homomorphism $\alpha : \Sigma \to hyp(\mathcal{D})$ with $\mathcal{D}$ a CD-category, there is a unique identity-on-objects strict CD-functor γ in **CDCat** (the category of CD-categories) that makes the following diagram commute.



Proposition 2.8. There is a C-colored prop isomorphism $\mathbf{P}_\Sigma + \mathbf{CD}_C \cong \mathbf{FCD}_{\Sigma,C}$. It extends to an isomorphism of CD-categories: thus $\mathbf{P}_\Sigma + \mathbf{CD}_C$ is the free CD-category on (Σ, C).

Proof. With the notations in the definition 2.7, let $\mathcal{D}$ be a CD-category. We need to prove the following natural isomorphisms, with the first identity-on-nodes and the second identity-on-object.

1. $\mathbf{Hom}_{\mathbf{Hyp}}(\Sigma, hyp(\mathcal{D})) \cong \mathbf{Hom}_{\mathbf{Prop}_C}(\mathbf{P}_\Sigma, \mathcal{D})$ is immediate by definition of $\mathbf{P}_{\Sigma,C}$.

2. $\mathbf{Hom}_{\mathbf{Prop}_C}(\mathbf{P}_\Sigma, \mathcal{D}) \cong \mathbf{Hom}_{\mathbf{CDCat}_C}(\mathbf{P}_\Sigma + \mathbf{CD}_C, \mathcal{D})$ where all functors preserve the identity-on-object CD-structure, i.e. they are identity-on-object CD-functors.
   This is equivalent to say that $\mathbf{Hom}_{\mathbf{Prop}_C}(\mathbf{P}_\Sigma, \mathcal{D}) \cong \mathbf{Hom}_{\mathbf{CDCat}_C}(\mathbf{P}_\Sigma + \mathbf{CD}_C, \mathcal{D})$ with $\mathbf{CDCat}_C$ the category of CD-categories with set of objects C^*.
   To prove that, given $F \in \mathbf{Hom}_{\mathbf{Prop}_C}(\mathbf{P}_\Sigma, \mathcal{D})$, the comonoid structure of every color $c \in C$ is only in $\mathbf{CD}_c$. We define F' with $F'|_{\mathbf{P}_\Sigma} = F$, and $F'(\mathsf{copy}_c) = \mathsf{copy}_c$, $F'(\mathsf{del}_c) = \mathsf{del}_c$. Then $F \mapsto F'$ is a natural bijection.

□

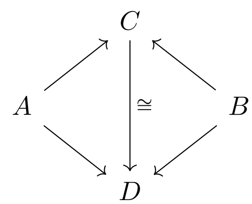
The goal of this section is to give a combinatorial description of the free CD-category and the free Markov category. In a way similar to [Zan17], we consider the monoidal theory (Σ, C) with Σ, C finite as an object of **Hyp**, with every node appearing only once. Then the labelling in a free CD-category can be given formally by the slice category $\mathbf{Hyp}/(\Sigma, C)$, for which we shall use the notation $\mathbf{Hyp}_{\Sigma,C}$. This definition ensures that a Σ-operation $o : w \to v$ labels a hyperedge only when the label of its input nodes forms the word w, and output v.

Definition 2.9. The *input* of G is defined by $\mathrm{In}(G) = \{n \in G(*) | \forall e \in E(G), n \notin \mathrm{out}(e)\}$.
The *output* of G is defined by $\mathrm{Out}(G) = \{n \in G(*) | \forall e \in E(G), n \notin \mathrm{in}(e)\}$.

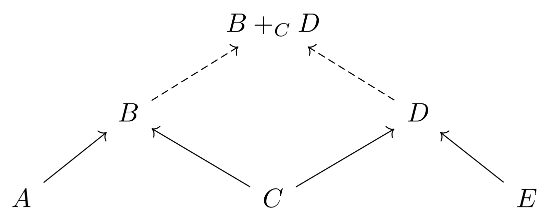
Definition 2.10. [BGK+16] The *indegree* $\mathsf{indeg}_G(v)$ of a node v in hypergraph G is the number of pairs (h, i) where h is a hyperedge with v its i-th target. Similarly, the *outdegree* $\mathsf{outdeg}_G(v)$ of v is the number of pairs (h, j) where h is an hyperedge with v its j-th source.

Definition 2.11. [nLa21c] In any category $\mathcal{C}$, a *span* from x to y is a diagram of the form $x \leftarrow s \to y$. A span in $\mathcal{C}^{op}$ is called a cospan in $\mathcal{C}$.

Definition 2.12. [nLa21c] Let $\mathcal{C}$ be a category that has pushouts. The category of cospans of $\mathcal{C}$ has objects the same as in $\mathcal{C}$, and morphisms up to 2-isomorphism: $A \to C \leftarrow B$ and $A \to D \leftarrow B$ are identified if there exists an isomorphism making the diagram commute:



The composition of morphisms $A \to B \leftarrow C$ and $C \to D \leftarrow E$ is the pushout diagram:



Definition 2.13. [nLa21b] The *slice category* $\mathcal{C}/c$ of category $\mathcal{C}$ over c has

- objects that are all arrows $f \in \mathcal{C}$ such that $cod(f) = c$, and
- morphisms $g : X \to X' \in \mathcal{C}$ from $f : X \to c$ to $f' : X' \to c$ such that $f' \circ g = f$.

Definition 2.14. Fix a set $\{x_i\}_{i \in \mathbb{N}}$ totally ordered by $\mathbb{N}$. The *category of CD termgraphs* $\mathbf{CDTerm}_{\Sigma,C}$ is the restriction of the category of cospans of the slice category $\mathbf{Hyp}_{\Sigma,C}$ on discrete objects whose set of nodes is $\{1, \dots n\}$ for $n \in \mathbb{N}$. such that every morphism has an acyclic hypergraph and is left-monogamous: for every morphism $m \xrightarrow{p} G \xleftarrow{q} n$, p is monomorphism and, for all nodes v of G,

$$\text{the indegree } \mathsf{indeg}_G(v) \text{ is } \begin{cases} 0 & \text{if } v \in \mathrm{Im}(p) \\ 1 & \text{otherwise} \end{cases}$$

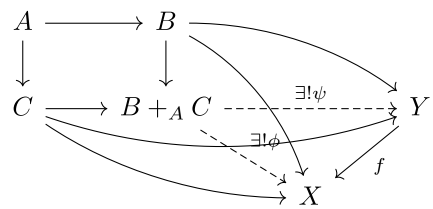
The idea of left-monogamicity comes from [BGK+16], where they use the *monogamicity* for a different motivation.

We should show that $\mathbf{CDTerm}_{\Sigma,C}$ is a subcategory of $\mathsf{cospan}(\mathbf{Hyp}_{\Sigma,C})$:

The identity morphism is the empty cospan.

Lemma 2.15. If a category $\mathcal{C}$ has pushouts, X is an object in $\mathcal{C}$, then pushouts of the slice category $\mathcal{C}/X$ are the unique morphisms from pushouts in $\mathcal{C}$ to X $\forall X \in \mathcal{C}$.

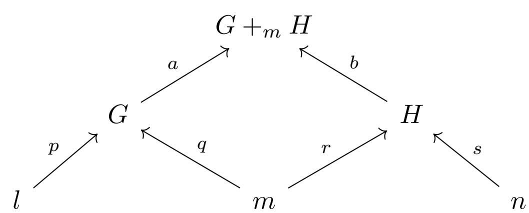
Proof. Let A, B, C be in $\mathcal{D}$, and let $B +_A C$ be the pushout of $A \to B$ and $A \to C$, with morphisms in $\mathcal{C}$. It suffices to prove that ψ satisfies the universal property of pushout in $\mathcal{C}/X$.



By the universal property of $B +_A C$ in $\mathcal{C}$, there is a unique arrow ϕ in $\mathbf{Hom}(\mathcal{C})$ such that the pushout diagram commute. Given a morphism $f : Y \to X$ in $\mathbf{Hom}(\mathcal{C})$ with Y object in $\mathcal{C}$, by the universal property $B +_A C$ in $\mathcal{C}$, there is a unique arrow $\psi : B +_A C \to Y$ such that the diagram of $(B, C, B +_A C, Y)$ commutes. Thus ψ is the unique morphism such that $\phi = f \circ \psi$ and that the whole diagram commutes. □

By Lemma 2.15, it suffices to calculate the pushout of q, r in the single-colored case, then the C-colored case follows immediately.

Since **Hyp** is a presheaf category, a pushout of hypergraphs consists of pushouts on the underlying nodes and hyperedges and the unique choice of two maps that complete the pushout diagram:



Left-monogamicity of $G +_m H$:

Without loss of generality, we can suppose that p and r are inclusions of $\mathrm{In}(G)$ and $\mathrm{In}(H)$. a is mono since r is. So $a \circ p$ is mono.

For $v \in (G +_m H)(*)$, if $v \in \mathrm{Im}(a \circ p) = \mathrm{Im(p)}$, then $\mathsf{indeg}_G(v) = 0$.

- If $v \notin \mathrm{Im}(r)$, then no node in H is identified with v. $\mathsf{indeg}_{G+H}(v) = 0$.
- Otherwise, $\mathsf{indeg}_H(v) = 0$ since H is left-monogamous.

Thus $\mathsf{indeg}_{G+H}(v) = 0$.

If $v \notin \mathrm{Im}(a \circ p) = \mathrm{Im(p)}$,

- If $v \in G(*)$, then $\mathsf{indeg}_G(v) = 1$. With the same reasoning as the previous case, $\mathsf{indeg}_{G+H}(v) = 1$.
- Otherwise, $v \in H(*) \setminus G(*)$, so $v \notin m = \mathrm{In}$ which is the intersection of $G(*)$ and $H(*)$. Since H is left-monogamous, $\mathsf{indeg}_H(v) = 1$, thus $\mathsf{indeg}_{G+H}(v) = 1$.

Acyclicity of $G +_m H$:

Suppose there is a sequence of nodes $(v_1, \ldots, v_k)$ and a sequence of hyperedges $(f_1, \ldots, f_k)$ such that

$$v_i \in s(f_i), v_i \in t(f_{i-1}) \ \forall i \in \{2, \ldots, k\}$$

with $v_1 \in t(f_k), v_1 \in s(f_k)$.

Upon resetting the f_1 of the sequence, suppose $f_k \in E(G)$, $f_1 \in E(H)$, and thus there exists $i \in \{2, \dots, k\}$ such that $f_i \in E(H)$, $f_{i+1} \in E(G)$. There exists thus a $v_{i+1} \in s(f_{i+1}) \cap t(f_i) \subseteq m$.

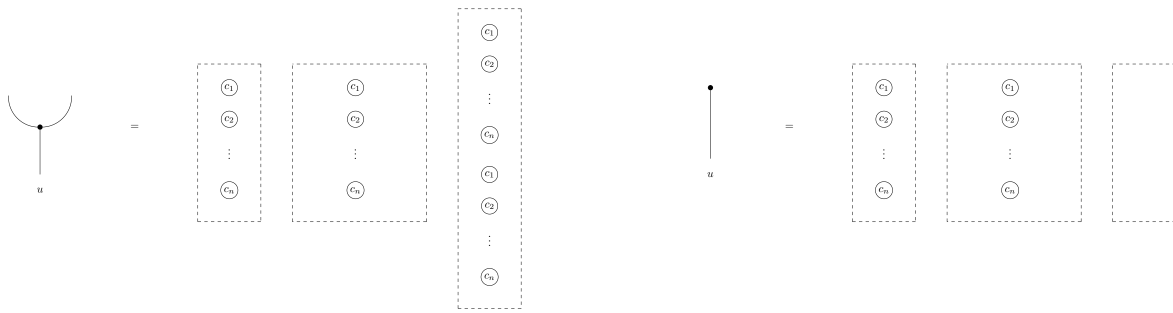
By the definition of left-monogamicity of H, $\mathsf{indeg}_H(v) = 0$, contradiction with $v_i \in t(f_i)$.

We have proven that the composition by pushout in $\mathbf{CDTerm}_{\Sigma,C}$ preserves left-monogamicity and acyclicity, thus $\mathbf{CDTerm}_{\Sigma,C}$ is a subcategory of $\mathsf{cospan}(\mathbf{Hyp}_{\Sigma,C})$.

The monoidal product in $\mathbf{CDTerm}_{\Sigma,C}$ is:

$(w \to G \leftarrow v) \otimes (u \to H \leftarrow t) = (w \otimes u \to G \amalg H \leftarrow v \otimes t)$ where $\otimes$ on objects is word concatenation.

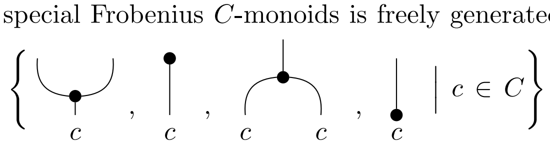
The CD structure for every object in $\mathbf{CDTerm}_{\Sigma,C}$, i.e. for $u = c_1 c_2 \dots c_n$:

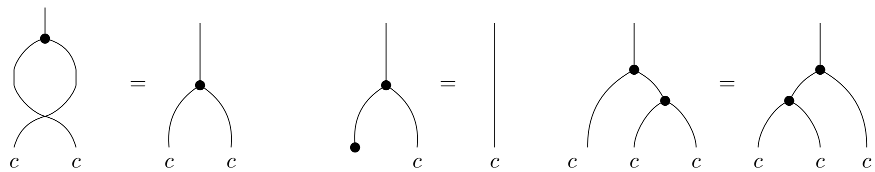


Now we arrive at the most important result of this report:

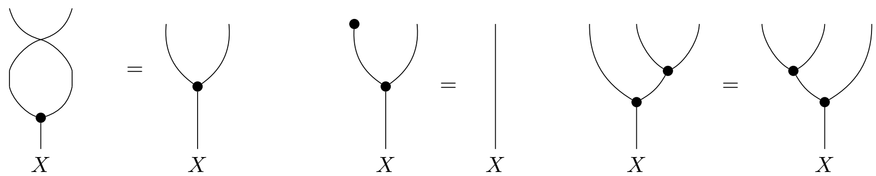
Theorem 2.16. $\mathbf{CDTerm}_{\Sigma,C}$ is the free CD category over a monoidal theory (Σ, C).

Before proving this theorem, we need several lemmas and propositions. The idea is to characterize $\mathbf{CDTerm}_{\Sigma,C}$ as a subcategory of Frobenius termgraphs $\mathbf{FTerm}_{\Sigma,C}$ [Zan17].

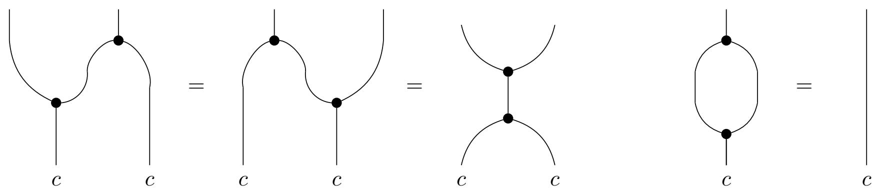
Definition 2.17. The C-colored prop $\mathbf{Frob}_C$ of special Frobenius C-monoids is freely generated by the monoidal theory $(\Sigma^C_{\mathbf{Frob}}, C)$, where $\Sigma^C_{\mathbf{Frob}} =$  and quotiented by equations:



(13)



(14)



(15)

Definition 2.18. [Zan17] Fix a set $\{x_i\}_{i\in\mathbb{N}}$ totally ordered by $\mathbb{N}$. The *category of Frobenius termgraphs* $\mathbf{FTerm}_{\Sigma,C}$ is defined as the restriction of the category of cospans of the slice category $\mathbf{Hyp}_{\Sigma,C}$ on discrete objects whose set of nodes is $\{1,\dots n\}$ for $n\in\mathbb{N}$.

Theorem 2.19. [Zan17] $\mathbf{P}_{\Sigma,C}+\mathbf{Frob}_C \to \mathbf{FTerm}_{\Sigma,C}$ is an isomorphism of C-colored props.

Lemma 2.20. Let $m \xrightarrow{f} G \xleftarrow{g} n$ be a left-monogamous directed acyclic cospan in $\mathbf{FTerm}_{\Sigma,C}$, and let L be a sub-hypergraph. Then L extends to a cospan $i\to L\leftarrow j$ such that

$$m \xrightarrow{f} G \xleftarrow{n} = (m\to C_1 \leftarrow i+k)\,; \begin{matrix}(i\to L\leftarrow i+k)\\ \otimes \\ (k \xrightarrow{id} k \xleftarrow{id} k)\end{matrix} ;(j+k\to C_2\leftarrow n) \tag{16}$$

where the cospans are left-monogamous directed acyclic.

Lemma 2.21. The functor $I:\mathbf{FCD}_{\Sigma,C}\to\mathbf{FTerm}_{\Sigma,C}$ as the copairing of functors $\mathrm{Id}_{\mathbf{P}_\Sigma}$ and $i: \mathbf{CD}_C\to\mathbf{Frob}_C$ is faithful.

Proof. $\mathbf{CD}_C$ can be decomposed as a coproduct in $\mathbf{Prop}_C$, namely $\sum_{c\in C}\mathbf{CD}_c$. Then by Proposition 2.8, $\mathbf{FCD}_{\Sigma,C}$ is the coproduct of $\mathbf{P}_\Sigma$ and $\sum_{c\in C}\mathbf{CD}_c$. By Theorem 2.19, $\mathbf{FTerm}_{\Sigma,C}$ is the coproduct of $\mathbf{P}_\Sigma$ and $\sum_{c\in C}\mathbf{Frob}_c$. The functor i is identity on $\mathbf{P}_\Sigma$ and $\sum_{c\in C}\mathbf{CD}_c$, so identity on their coproduct, thus faithful. □

Proposition 2.22. A morphism $n\xrightarrow{f} G\xleftarrow{g} m$ in $\mathbf{FTerm}_{\Sigma,C}$ is in the image of I if and only if it is left-monogamous directed acyclic, i.e. in $\mathbf{CDTerm}_{\Sigma,C}$.

The proof of Theorem 2.16 is then immediate: I is faithful and surjective, $\mathbf{FCD}_{\Sigma,C}$ is then isomorphic to $\mathrm{Im}(I)$, which equals $\mathbf{CDTerm}_{\Sigma,C}$ by Proposition 2.22. Since $\mathbf{CDTerm}_{\Sigma,C}$ is a subcategory of $\mathbf{FTerm}_{\Sigma,C}$ which is identified with a C-colored prop, $\mathbf{CDTerm}_{\Sigma,C}$ also is identified in the same way.

Definition 2.23. The *free Markov category* $\mathbf{FM}_{\Sigma,C}$ over a monoidal theory (Σ, C) is defined by the universal property: Given a Markov category $\mathcal{D}$ whose monoid of objects is $\mathcal{C}^*$, for every identity-on-nodes hypergraph homomorphism $\alpha : \Sigma \to hyp(\mathcal{D})$ with $\mathcal{D}$ a Markov category, there is a unique identity-on-objects strict CD-functor γ in **MarkovCat** (the category of Markov categories) that makes the following diagram commute.

$$\begin{array}{ccc} \Sigma & \xrightarrow{\ i\ } & hyp(\mathbf{FM}_{\Sigma,C}) \\ & {\scriptstyle\alpha}\searrow & \Big\downarrow {\scriptstyle hyp(\gamma)} \\ & & hyp(\mathcal{D}) \end{array}$$

hyp is a forgetful functor **MarkovCat** $\to$ **SMC** $\to$ **Hyp**, where **SMC** is the category of symmetric monoidal categories. In **MarkovCat** $\to$ **SMC**, *hyp* ignores all the comonoids in the Markov category; in **SMC** $\to$ **Hyp**, *hyp* maps an SMC to its underlying hypergraph.

Corollary 2.24. The free Markov category $\mathbf{FM}_{\Sigma,C}$ over a monoidal theory (Σ, C) is $\mathbf{CDTerm}_{\Sigma,C}$ quotiented by the naturality of the counits (4).

3 The Inflation Technique in Markov Categories

We show a first application of free Markov categories in categorical probability: generalizing the inflation technique [WSF19] for causal inference. A problem in causal inference is to determine if a given probability distribution on observed variables is compatible with a DAG causal structure. The inflation technique in Markov categories generalizes the classical one on three aspects:

- The causal structure is generalized to a DAH (directed acyclic hypergraph), compared with a DAG in the classical case.
- The conditional factorization may not exist in certain Markov categories, whereas it is always assumed in the classical case.
- The probability distribution in the compatibility problem is generalized to a conditional distribution.

We leave the readers to find the inflation technique for Bayesian network in [WSF19]. Here we generalize the idea to Markov categories.

Definition 3.1. Given a monoidal theory (Σ, C) with Σ acyclic, C as the node set of the given hypergraph, a morphism ϕ is constructed as follows: $G = (\Sigma, C)$ can be identified with an object in $\mathbf{Hyp}_{\Sigma,C}$, by choosing an arbitrary total order on the node set, and the discrete hypergraphs on two legs are $\mathrm{In}(G)$ and $\mathrm{Out}(G)$ with the total order induced from the node set. We denote such $\phi = (\Sigma, C)$.

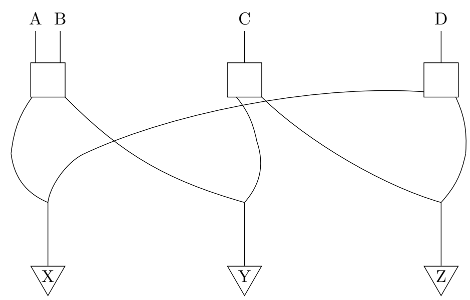
In the following, we suppose ϕ is left-monogamous. In Markov category, we define compatibility as a functorial property:

Definition 3.2. A morphism f in a Markov category $\mathcal{C}$ with codomain $(A_i)_{i\in I}$ is *compatible* with $\phi = (\Sigma, C)$ with codomain $(B_j)_{j\in I}$ if there exists a CD functor $F : \mathbf{FM}_{\Sigma,C} \to \mathcal{C}$ such that $F(\phi) = f$ and $F(B_i) = A_i$. We say that F is associated to f.

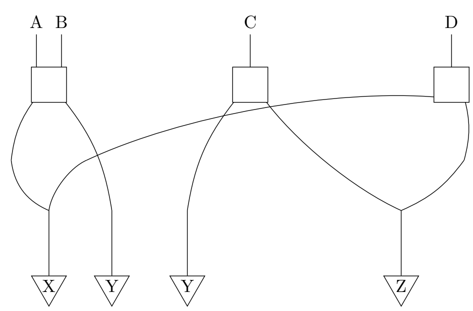
The key property of the inflation technique, which is (7) in [WSF19], becomes the following:

Remark 3.3. If f is compatible with $\phi = (\Sigma, C)$, with F the associated functor of f, then for any morphism ψ in $\mathbf{FM}_{\Sigma,C}$, for any subsequence of objects $S \subseteq cod(\phi)$ such that $\phi_S = \psi_S$ (marginalization over S), then we have $f_{F(S)} = F(\psi_S)$.
We often use its contrapositive:
If there exists ψ morphism in $\mathbf{FM}_{\Sigma,C}$ and a subsequence $S \subseteq cod(\phi)$ such that $\phi_S = \psi_S$, and we have for any functor F associated to f, $f_{F(S)} \neq F(\psi_S)$, then f is not compatible with ϕ.

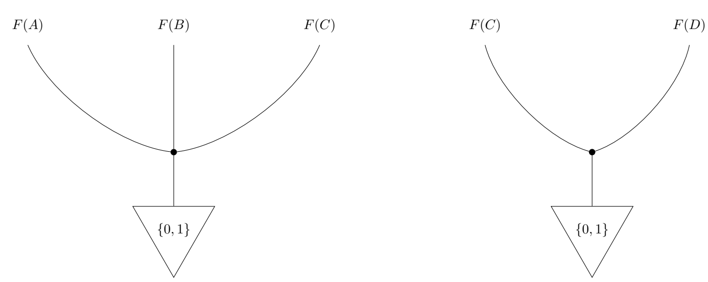
Example 3.4. Given $\phi : \ I \to A \otimes B \otimes C \otimes D$:



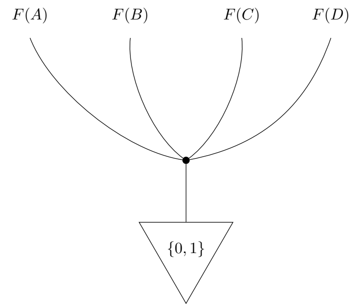
We ask if $f = \frac{[0000]+[1111]}{2}$ morphism in **FinStoch** is compatible with ϕ.
Consider ψ in $\mathbf{FM}_{\Sigma,C}$:



We will show that f is not compatible with ϕ using Remark 3.3. Consider $S_1 = \{A, B, C\}$, $S_2 = \{C, D\}$, we see that $\phi_{S_i} = \psi_{S_i}$ $i = 1, 2$. Thus for any functor F associated to f, $F(\psi_{S_1}) = f_{F(S_1)} = \frac{[000]+[111]}{2}$, $F(\psi_{S_2}) = f_{F(S_2)} = \frac{[00]+[11]}{2}$.
Thus $F(A), F(B), F(C)$ are in perfect correlation, and so are $F(C), F(D)$:



Then there exists a morphism



However, ψ displays the independance of AB and C. (cf. Definition 12.1 and Remark 12.3 of [Fri20]), contradiction.

4 Conclusion

Despite the complicated pandemic situation, I have fortunately got the opportunity of a fully on-site internship at Innsbruck University. Before studying the free CD category, I have also tried two subjects on April and May without success, but the knowledge and experience I learned from that is no less than the current one. The uncertainty of difficulty itself is a crucial part of the research and no exercise of a textbook could replace that. I feel also that in some domains, the truthfulness is highly correlated to beauty and simplicity: if a proof is full of unnecessarily complicated definitions, then it probably is not optimal. This faith has helped us find the most elegant way to prove the combinatorial characterization of free CD categories.

During the internship, we conjectured a combinatorial characterization of free CD category over a monoidal theory, and proved it. We showed one of its application in categorical causal inference, i.e. the inflation technique in Markov categories which solves the causal compatibility problem. However, we have only given one simple example of inflation, and we expect to give more examples in the category of relations, in which a morphism is not a Markov kernel but indicating the possibility instead of probability. We also expect that there is a necessary and sufficient condition for the causal compatibility using inflation technique, which in the classical case appears in [NW20].

We did not answer the second problem in the Context of this report, i.e. the criteria of Markov equivalence of morphisms in a Markov category, but we expect that this could be solved using the free Markov categories.

A Omitted proofs

Proof of Lemma 2.20. The construction of the factorization of cospan is the same as Lemma 3.11 in [BGK$^+$16], whereas L in our case is not necessarily convex.

Let C_1 be the smallest sub-hypergraph containing $\mathrm{In}(G)$ and every hyperedge which is not in L, but has a path to it. Let C_2 be the smallest sub-hypergraph containing $\mathrm{Im}(g)$ such that $C_1 \cup L \cup C_2 = G$.

By construction, C_1 and L share no hyperedges. If C_2 shared a hyperedge with C_1 or L, then there would exist a smaller C_2' such that $C_1 \cup L \cup C_2' = G$, so C_2 shares no hyperedge with either C_1 or L. Hence, the three sub-hypergraphs only overlap on nodes. Now let:

$$\begin{aligned} i &:= C_1(*) \cap L(*) \\ j &:= C_2(*) \cap L(*) \\ k &:= (C_1(*) \cap C_2(*)) \setminus L(*) \end{aligned}$$

It follows that

$$\begin{aligned} k + i &= C_1(*) \cap (L(*) \cup C_2(*)) \\ k + j &= C_2(*) \cap (L(*) \cup C_1(*)) \end{aligned}$$

Now define the following cospans, where arrows are all inclusion except the last right leg:

$$\begin{gathered} m \xrightarrow{p} C_1 \xleftarrow{q} k + i \\ i \xrightarrow{r} L \xleftarrow{s} j \\ k + j \xrightarrow{t} C_2 \xleftarrow{u} n \end{gathered}$$

Then (16) is computed as the colimit of the following diagram:

$$m \rightarrow C_1 \leftarrow i + k \rightarrow L + k \leftarrow j + k \rightarrow C_2 \leftarrow n$$

The two spans of $i + k$ and $j + k$ identify those nodes from G which occur in more than one sub-hypergraph, so this amounts to simply taking the union:

$$\left(m \rightarrow C_1 \cup L \cup C_2 \leftarrow n \right) \quad = \quad \left(m \rightarrow G \leftarrow n \right)$$

C_1, C_2 and L are all directed acyclic because G is. So it only remains to show that each of these cospans is left-monogamous.

C_1 is left-monogamous: for each $v \in C_1(*)$, if $v \in \mathrm{Im}(p) = \mathrm{Im}(a \circ p)$, then $\mathsf{indeg}_G(v) = 0$. For each $v \notin \mathrm{Im}(a \circ p)$, $\mathsf{indeg}_G(v) = 1$, there exists only one $(h, i) \in E(G) \times \mathbb{N}$ such that $v = t_i(h)$. If h is not in $E(C_1)$, then v does not appear in C_1 since $v = t_i(h)$ can not be in $\mathrm{In}(L + k)$. Hence $h \in E(C_1)$. $\mathsf{indeg}_{C_1}(v) = 1$.

C_2 is left-monogamous: for each $v \in \mathrm{Im}(t) = k + j$, there are two cases:

1. $v \in k$, in particular $v \in C_1(*)$. By definition of C_1, there are two possibilities:

   (a) $v \in \mathrm{Im}(p)$, then $\mathsf{indeg}_G(v) = 0$. Thus $\mathsf{indeg}_{C_2}(v) \leq \mathsf{indeg}_G(v) = 0$.

(b) $\exists f \in E(C_1)$ such that $v \in t(f)$ and f has a path to L. Then the only output appearance of v is in C_1. Thus $\mathsf{indeg}_{C_2}(v) = 0$.

2. $v \in j$. Suppose $\mathsf{indeg}_{C_2}(v) = 1$, i.e. $\exists f \in E(C_2)$ such that $v \in t(f)$, thus f has a path to v. By definition of C_1, $f \in E(C_1)$, contradiction.

   Hence $\mathsf{indeg}_{C_2}(v) = 0$.

For each $v \in C_2(*) \setminus \mathrm{Im}(u) = C_2(*) \setminus (L(*) \cup C_1(*))$, in particular $v \notin C_1(*)$, thus $v \notin \mathrm{Im}(p)$. Thus $\mathsf{indeg}_G(v) = 1$, i.e. $\exists f \in E(G)$ such that $v \in t(f)$. If $f \in E(C_1) \amalg E(L)$, then $v \in C_1(*) \cup L(*)$, contradiction. Thus $f \in E(C_2)$, $\mathsf{indeg}_{C_2}(v) = 1$.

L is left-monogamous: for $v \in L(*)$, if $\mathsf{indeg}_L(v) = 0$:

1. $\mathsf{indeg}_G(v) = 0$, then $v \in \mathrm{In}(G) \subseteq C_1(*)$;

2. $\mathsf{indeg}_G(v) = 1$, then $\exists f \in E(C_1) \amalg E(C_2)$ such that $v \in t(f)$. Since $v \in L(*)$, by definition of C_1, $f \in E(C_1)$, hence $v \in C_1(*)$.

In both cases, $v \in L(*) \cap C_1(*) = i$.
Conversely, if $v \in i$, then $v \in C_1(*)$. By definition of C_1, there are two possibilities:

1. $v \in \mathrm{Im}(p)$: suppose $\mathsf{indeg}_L(v) = 1$, then $\exists f \in E(L)$ such that $v \in t(f)$, thus $\mathsf{indeg}_G(v) = 1$, $v \notin \mathrm{Im}(a \circ p) = \mathrm{Im}(p)$. Hence $\mathsf{indeg}_L(v) = 0$.

2. $\exists g \in E(C_1)$ such that $v \in t(g)$. Then the only output appearance of v is taken by a hyperedge in $E(C_1)$. Thus there is no edge in $E(L)$ that has v as an output. $\mathsf{indeg}_L(v) = 0$.

□

Proof of Proposition 2.22. The forward direction: by Lemma 2.21, I is faithful, thus the image of I is a subcategory. Every generator in $\mathbf{FCD}_{\Sigma,C}$ maps to a directed acyclic and left monogamous cospan: $\forall e \in \Sigma$, $\forall c \in C$,

$$I(e) = s(e) \to e \leftarrow t(e)$$
$$I(\mathsf{copy}_c) = c \to c \leftarrow c \otimes c$$
$$I(\mathsf{del}_c) = c \to \emptyset \leftarrow \emptyset$$

By the functoriality of I, the image of I is contained in $\mathbf{CDTerm}_{\Sigma,C}$.

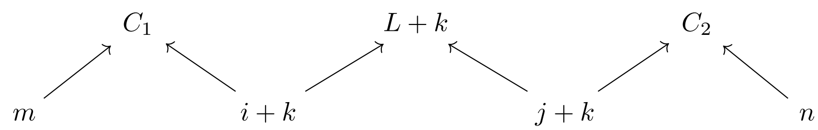
The converse direction: Induction on the number of hyperedges in G:

If G does not contain any hyperedge and is left-monogamous acyclic, then the cospan is $n \xrightarrow{\mathrm{id}} n \xleftarrow{p} m$ up to equivalence of cospans, which is a morphism in $\mathbf{RCDTerm}_C$(cf. Definition B.1). It suffices to show that $\mathbf{RCDTerm}_C$ is a subcategory of $I(\mathbf{CD}_C)$: By Lemma B.5, $n \xrightarrow{\mathrm{id}} n \xleftarrow{p} m$ can be factorized into a monotone function f and a permutation σ. The preimage of f by I is given by 17. The preimage of σ is itself. Hence $n \xrightarrow{\mathrm{id}} n \xleftarrow{p} m$ is in the image of I.

If G contains only one edge e, then there exists $\sigma \in I(\mathbf{RP}_C)$, $f \in I(\mathbf{RM}_C)$ (cf. Lemma B.5) such that up to equivalence, $n \xrightarrow{\text{id}} n \xleftarrow{p} m = I(\sigma; e; f)$. f is the parallel composition of copy and del.

Suppose for every hypergraph G with $|E(G)| = k$, $n \to G \leftarrow m$ is in $\mathrm{Im}(I)$. For a hypergraph G with $|E(G)| = k + 1$:

Choose an arbitrary hyperedge e of G. By Lemma 2.20, $n \xrightarrow{f} G \xleftarrow{g} m$ is factorized into



with C_1, L, C_2 defined in Lemma 2.20, and all copsans left-monogamous acyclic.

By the inductive hypothesis, these cospans are in $\mathrm{Im}(I)$. By the functoriality of I, $n \xrightarrow{f} G \xleftarrow{g} m$ is also in $\mathrm{Im}(I)$.

□

B Factorization of CD Termgraph

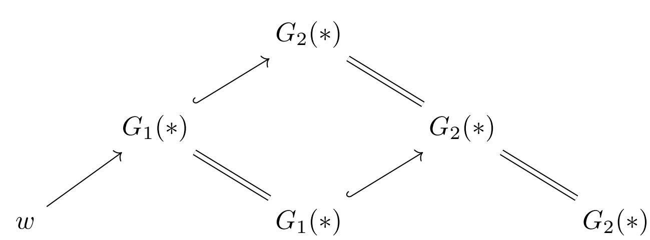
During the research, we tried another method to prove the combinatorial characterization of free CD-category. Although we have not completed the proof using that method, the factorization that it uses is still of value of itself. And some of the results are used in the previous proof.

Definition B.1. (and Proposition) The category $\mathbf{LCDTerm}_{\Sigma,C}$ is the subcategory of $\mathbf{CDTerm}_{\Sigma,C}$ such that the right leg is a permutation on $G(*)$.

The category $\mathbf{RCDTerm}_C$ is the subcategory of $\mathbf{CDTerm}_{\Sigma,C}$ such that every morphism is in the form $u = u \leftarrow v$.

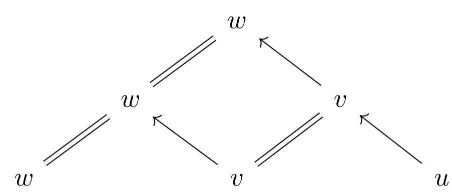
Proof. We need to prove that $\mathbf{LCDTerm}_{\Sigma,C}$ and $\mathbf{RCDTerm}_C$ are subcategories of $\mathbf{CDTerm}_{\Sigma,C}$. For the cospan category, the composition is the pushout diagram. Since $\mathbf{Hyp}_{\Sigma,C}$ is a presheaf category, a pushout of hypergraphs consists of pushouts on the underlying nodes and hyperedges and the unique choice of two maps that complete the pushout diagram.

In $\mathbf{LCDTerm}_{\Sigma,C}$, given $w \to G_1 \leftarrow G_1(*)$ and $G_1(*) \to G_2 \leftarrow G_2(*)$, the pushout of nodes is



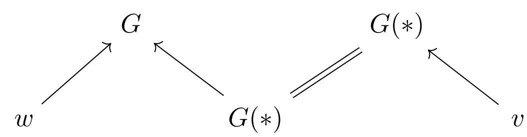
The pushout of edges is just the disjoint union of edges.

In $\mathbf{RCDTerm}_C$, given $G_1(*) = G_1(*) \leftarrow v$, $v = v \leftarrow u$, the pushout is



∎

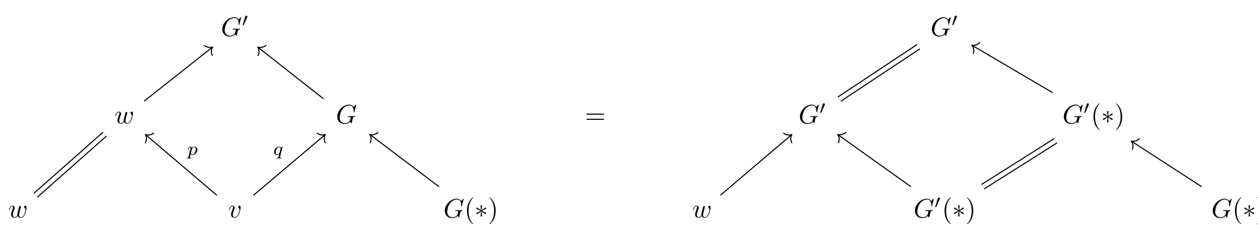
Proposition B.2. Every morphism $w \to G \leftarrow v$ in $\mathbf{CDTerm}_{\Sigma,C}$ has a unique factorization in $(\mathbf{LCDTerm}_{\Sigma,C}, \mathbf{RCDTerm}_C)$ up to permutations in $G(*)$:



Proof. The total order on $G(*)$ is fixed, and the pushout of the diagram is G. ∎

Proposition B.3. Given a composition of a pair of morphism $(\mathbf{RCDTerm}_C, \mathbf{LCDTerm}_{\Sigma,C})$ $w = w \xleftarrow{p} v \xrightarrow{q} G \leftarrow G(*)$, it is equal to a composition of a pair of morphism in $(\mathbf{LCDTerm}_{\Sigma,C}, \mathbf{RCDTerm}_C)$ $w \to G' \leftarrow G'(*) = G'(*) \leftarrow G(*)$, where G' is the disjoint union of G and w where $p(n) \sim q(n)\ \forall n$.

Proof.

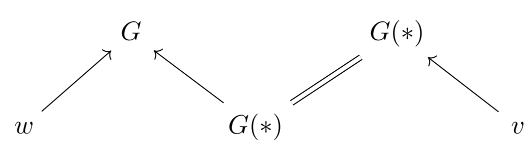


∎

Theorem B.4. $\mathbf{FCD}_{\Sigma,C} = \mathbf{CDTerm}_{\Sigma,C}$.

To prove this theorem, we first construct the γ given an α.

Given a morphism $w \to G \leftarrow v$, by Proposition B.2, we have a unique factorization up to permutations in $G(*)$



We define $\gamma(w \to G \leftarrow v)$ as the composition $\gamma(G(*) = G(*) \leftarrow v) \circ \gamma(w \to G \leftarrow G(*))$.

We define the parallel composition $\gamma((w \to G \leftarrow v) \otimes (a \to H \leftarrow b)) = \gamma(w \oplus a \to G \amalg H \leftarrow v \oplus b)$, with the numbering of nodes of the second cospan added by $|G(*)|$.

Suppose γ is functorial on respectively $\mathbf{LCDTerm}_{\Sigma,C}$ and $\mathbf{RCDTerm}_C$, then γ is functorial on $\mathbf{LCDTerm}_{\Sigma,C}$, because

$$\begin{array}{ccccccccccccc}
 & G_1 & & & G_1(*) & & & G_2 & & & G_2(*) & & \\
 \nearrow & & \nwarrow & /\!/ & & \nwarrow & \nearrow & & \nwarrow & /\!/ & & \nwarrow & \\
w & & & G_1(*) & & & v & & & G_2(*) & & & u
\end{array}$$

is equal to (by Proposition B.3)

$$\begin{array}{ccccccccccccc}
 & G_1 & & & G' & & & G'(*) & & & G_2(*) & & \\
 \nearrow & & \nwarrow & \nearrow & & \nwarrow & /\!/ & & \nwarrow & /\!/ & & \nwarrow & \\
w & & & G_1(*) & & & G'(*) & & & G_2(*) & & & u
\end{array}$$

where G' is the pushout of $v \to G_1(*)$ and $v \to G_2$.

We have again a morphism in $\mathbf{LCDTerm}_{\Sigma,C}$ followed by a morphism in $\mathbf{RCDTerm}_C$. Thus γ is functorial on $\mathbf{CDTerm}_{\Sigma,C}$.

The only things we need to prove are the well-defineness and functoriality of γ on respectively $\mathbf{LCDTerm}_{\Sigma,C}$ and $\mathbf{RCDTerm}_C$, and the universal property in Definition 2.7.

First, we define γ on $\mathbf{RCDTerm}_C$: Since every morphism in $\mathbf{RCDTerm}_C$ can be identified with its right leg,

$$\begin{aligned}
\gamma|_{\mathbf{RCDTerm}_C} : (\mathbf{FinOrd}/C)^{op} &\longrightarrow \mathcal{D} \\
(\underline{n}, p) &\longmapsto \bigotimes_{i \in \underline{n}} p(i)
\end{aligned}$$

And given a morphism $f : (\underline{m}, q) \longrightarrow (\underline{n}, p)$,

$$\gamma(f) := \sigma \circ \left(\bigotimes_{i \in \underline{n}} \begin{cases} \mathsf{copy}(k)_{p(i)} & \text{if } k = |f^{-1}(i)| > 1 \\ \mathsf{id}_{p(i)} & \text{if } k = |f^{-1}(i)| = 1 \\ \mathsf{del}_{p(i)} & \text{if } k = |f^{-1}(i)| = 0 \end{cases} \right) \tag{17}$$

The tensor product is in the order of $\underline{n}$. σ is a permutation on codomain of the morphism in the parentheses such that $\forall i \in \underline{n}$, $f^{-1}(i)$ is exactly the places for $p(i)$ in codomain of $\gamma(f)$. In other words, σ could be any permutation whose orbits are the fibers $f^{-1}(i)$ $\forall i$: for all such σ_1, σ_2, $g\sigma_1^{-1}\sigma_2 = g$.

$\gamma(f)$ is well-defined because $\mathsf{copy}(k)_{p(i)}$ commute with σ, which is a permutation whose orbits are the fibers $f^{-1}(i)$ $\forall i$.

In fact, we have defined γ on a factorization system ($\mathbf{RM}_C$, $\mathbf{RP}_C$):

Lemma B.5. $\mathbf{RCDTerm}_C$ has a factorization system ($\mathbf{RM}_C$, $\mathbf{RP}_C$) where

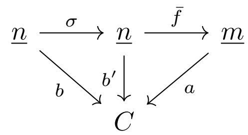
$\mathbf{RM}_C$ is the category of monotone functions in $(\mathbf{FinOrd}/C)^{op}$,

$\mathbf{RP}_C$ is the category of permutations in $(\mathbf{FinOrd}/C)^{op}$.

Proof. Given a morphism

$$\begin{array}{ccc}
\underline{n} & \xrightarrow{f} & \underline{m} \\
 {\scriptstyle b}\searrow & & \swarrow{\scriptstyle a} \\
 & C &
\end{array}$$

there exist (non unique) σ in $\mathbf{RP}_C$, $\bar{f}$ in $\mathbf{RM}_C$ such that $f = \sigma \circ \bar{f}$ and

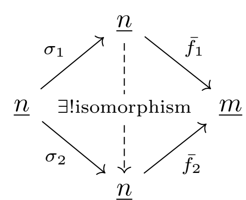


commutes. $b' := b \circ \sigma^{-1}$. $\bar{f}$ is constructed as follows:

For every $i \in \underline{m}$, put $f^{-1}(i)$(with the total order of $dom(f)$) in $dom(\bar{f}) = \underline{n}$ such that $dom(\bar{f})$ is the union of the ordered sequence of $(f^{-1}(i))_{i \in \underline{m}}$.

σ is the permutation of $\underline{n}$ that makes $dom(f)$ into $dom(\bar{f})$.

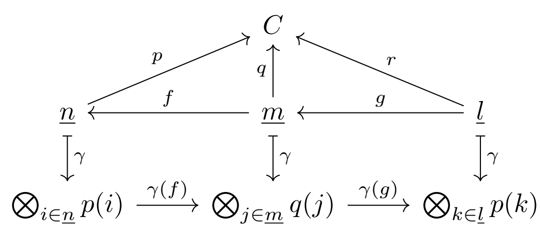
To prove that this forms a factorization system, it suffices to observe that the factorization is unique up to unique isomorphism:



□

Lemma B.6. γ is functorial on $\mathbf{RM}_C$ and $\mathbf{RP}_C$.

Proof. On $\mathbf{RM}_C$:



$$\gamma(g) \circ \gamma(f) = \left(\bigotimes_{j \in \underline{m}} \begin{cases} \mathsf{copy}(k)_{q(j)} & \text{if } k = |g^{-1}(j)| > 1 \\ \mathsf{id}_{q(j)} & \text{if } k = |g^{-1}(j)| = 1 \\ \mathsf{del}_{q(j)} & \text{if } k = |g^{-1}(j)| = 0 \end{cases} \right) \circ \left(\bigotimes_{i \in \underline{n}} \begin{cases} \mathsf{copy}(k)_{p(i)} & \text{if } k = |f^{-1}(i)| > 1 \\ \mathsf{id}_{p(i)} & \text{if } k = |f^{-1}(i)| = 1 \\ \mathsf{del}_{p(i)} & \text{if } k = |f^{-1}(i)| = 0 \end{cases} \right)$$

$$= \left(\bigotimes_{j \in \underline{m}} \begin{cases} \mathsf{copy}(k)_{p \circ f(j)} & \text{if } k = |g^{-1}(j)| > 1 \\ \mathsf{id}_{p \circ f(j)} & \text{if } k = |g^{-1}(j)| = 1 \\ \mathsf{del}_{p \circ f(j)} & \text{if } k = |g^{-1}(j)| = 0 \end{cases} \right) \circ \left(\bigotimes_{i \in \underline{n}} \begin{cases} \mathsf{copy}(k)_{p(i)} & \text{if } k = |f^{-1}(i)| > 1 \\ \mathsf{id}_{p(i)} & \text{if } k = |f^{-1}(i)| = 1 \\ \mathsf{del}_{p(i)} & \text{if } k = |f^{-1}(i)| = 0 \end{cases} \right)$$

$$= \bigotimes_{i \in \underline{n}} \begin{cases} \mathsf{copy}(k)_{p(i)} & \text{if } k = |g^{-1} \circ f^{-1}(i)| > 1 \\ \mathsf{id}_{p(i)} & \text{if } k = |g^{-1} \circ f^{-1}(i)| = 1 \\ \mathsf{del}_{p(i)} & \text{if } k = |g^{-1} \circ f^{-1}(i)| = 0 \end{cases}$$

$$= \gamma(f \circ g)$$

On $\mathbf{RP}_C$, γ is the identity, thus functorial. □

References

[BGK+16] Filippo Bonchi, Fabio Gadducci, Aleks Kissinger, Paweł Sobociński, and Fabio Zanasi. Rewriting modulo symmetric monoidal structure. In *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 710–719, 2016.

[CJ19] Kenta Cho and Bart Jacobs. Disintegration and bayesian inversion via string diagrams. *Mathematical Structures in Computer Science*, 29(7):938–971, 2019.

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

[Fri20] Tobias Fritz. A synthetic approach to markov kernels, conditional independence and theorems on sufficient statistics. *Advances in Mathematics*, 370:107239, 2020.

[Gol99] Petr Viktorovich Golubtsov. Axiomatic description of categories of information transformers. *Problemy Peredachi Informatsii*, 35(3):80–98, 1999.

[Gui80] René Guitart. Tenseurs et machines. *Cahiers de topologie et géométrie différentielle catégoriques*, 21(1):5–62, 1980.

[HR15] Philip Hackney and Marcy Robertson. On the category of props. *Applied Categorical Structures*, 23(4):543–573, 2015.

[nLa21a] nLab authors. Kleisli category. `http://ncatlab.org/nlab/show/Kleisli%20category`, August 2021. Revision 27.

[nLa21b] nLab authors. over category. `http://ncatlab.org/nlab/show/over%20category`, July 2021. Revision 42.

[nLa21c] nLab authors. span. `http://ncatlab.org/nlab/show/span`, August 2021. Revision 56.

[NW20] Miguel Navascués and Elie Wolfe. The inflation technique completely solves the causal compatibility problem. *Journal of Causal Inference*, 8(1):70–91, 2020.

[VP13] Tom S Verma and Judea Pearl. On the equivalence of causal models. *arXiv preprint arXiv:1304.1108*, 2013.

[WSF19] Elie Wolfe, Robert W Spekkens, and Tobias Fritz. The inflation technique for causal inference with latent variables. *Journal of Causal Inference*, 7(2), 2019.

[Zan17] Fabio Zanasi. Rewriting in free hypergraph categories. *Electronic Proceedings in Theoretical Computer Science*, 263:16–30, Dec 2017.