A Categorical Approach to DIBI Models

Tao Gu[1], Jialu Bao[2], Justin Hsu[2], Alexandra Silva[2], and Fabio Zanasi[1,3]

[1] University College London, UK
[2] Cornell University, NY, USA
[3] University of Bologna, Italy

Abstract. The logic of Dependence and Independence Bunched Implications (DIBI) is a logic to reason about conditional independence (CI); for instance, DIBI formulas can characterise CI in probability distributions and relational databases, using the probabilistic and relational DIBI models, respectively. Despite the similarity of the probabilistic and relational models, a uniform, more abstract account remains unsolved. The laborious case-by-case verification of the frame conditions required for constructing new models also calls for such a treatment. In this paper, we develop an abstract framework for systematically constructing DIBI models, using category theory as the unifying mathematical language. In particular, we use string diagrams – a graphical presentation of monoidal categories – to give a uniform definition of the parallel composition and subkernel relation in DIBI models. Our approach not only generalises known models, but also yields new models of interest and reduces properties of DIBI models to structures in the underlying categories. Furthermore, our categorical framework enables a logical notion of CI, in terms of the satisfaction of specific DIBI formulas. We compare it with string diagrammatic approaches to CI and show that it is an extension of string diagrammatic CI under reasonable conditions.

Keywords: Conditional Independence · Dependence Independence Bunched Implications · String Diagrams · Markov Categories.

1 Introduction

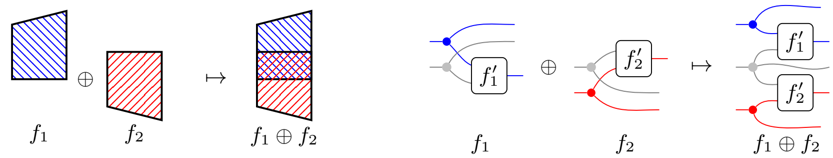
Conditional independence (CI) is a fundamental concept across various research areas, including programming languages [28,6,21], statistics [10], and database theory [1], among others. Although specific definitions may vary, the core idea remains straightforward: events A and B are 'independent' when information about one event does not convey information about the other. Furthermore, events A and B are 'conditionally independent' given event C if, with knowledge of event C, events A and B become independent. Albeit intuitive, reasoning about conditional independence is an intricate task, leading to extensive research efforts aimed at formalising such reasoning [25,14].

For probabilistic programs, an extension of standard programs with constructs to sample from distributions, formal methods for (conditional) independence have emerged as powerful tools for program verification. For instance, Barthe et. al. [6] introduced Probabilistic Separation Logic (PSL) and applied it to formalise several cryptography protocols, where independence of variables guarantees no leakage of information and thus security of the algorithms. A follow-up work from Bao et al. [5] proposed the logic of *Dependence and Independence Bunched Implications* (DIBI), which enhances PSL with the ability to reason about *conditional* independence. Syntactically, DIBI extends the logic of Bunched Implications (BI) [23,27], which is the assertion logic underpinning Separation Logic (SL) [28] and PSL, with a non-commutative conjunction $\fatsemi$ and its adjoints. Semantically, as in BI, the separating conjunction $*$ is interpreted through a partial operation $\oplus$ on states, regarded as the parallel composition. In addition, they define a sequential composition $\odot$ to interpret $P \fatsemi Q$. Informally, $P * Q$ says that P and Q hold in states that can be separated, and $P \fatsemi Q$ expresses a possible dependency of Q on P. Section 3 will review the logic in more details.

Bao et. al [5] introduced two kinds of semantic models for DIBI logic – probabilistic and relational. The probabilistic DIBI models are used to reason about CI of variables in discrete probabilistic computation. The relational DIBI models are designed to express join dependency – a notion of conditional independence in database theory and relational algebra – between variables in relational databases. These two models share many similarities, and the conditions to verify they are models are repetitive. This led Bao et. al. [5] to conjecture a categorical approach to construct abstract DIBI models that induce these concrete models as instances. This would greatly facilitate the construction of new models.

To solve this conjecture, we start with the observation that, in both the probabilistic and relational DIBI models, the states resemble *Markov kernels*: they are maps from input elements to distributions/powersets over output elements. Specifically, the input/output elements are value assignments on a finite set of variables, as an abstraction of program memories or database entries. Such DIBI states can be identified categorically as morphisms in the Kleisli categories associated to the discrete distribution monad (definition 15, for the probabilistic model) or the nonempty powerset monad (definition 16, for the relational model). However, giving a categorical definition for the parallel compositions $\oplus$ is difficult. The previous work [5] gives Figure 1a as a pictorial intuition for the parallel composition. The states are drawn as trapezoids, with the short and long vertical sides representing the input and output domains, respectively. There, given a blue map f_1 and a red map f_2, their parallel composition $f_1 \oplus f_2$ takes as input the union of their inputs. Then, each f_i takes its counterpart in the combined input domain and generates an output. Finally these two outputs are combined to be the output of $f_1 \oplus f_2$. This parallel composition *is partial* because the combination of their outputs is allowed only when the variables overlap in particular ways. This creates a challenge to capture DIBI models categorically because, in a categorical setting, the domains and codomains of DIBI states are objects, and it is not obvious how to define the overlap of objects used in the parallel composition.

Our solution stems from a formalisation of this graphical intuition through *string diagrams*, a pictorial formalism for monoidal categories. String diagrams



(a) Parallel composition depicted in [5]. (b) A diagrammatic representation.

Fig. 1: Intuition for parallel composition.

are widely adopted as intuitive yet mathematically rigorous reasoning tools across different areas of science, see [26] for an overview. We formalise the trapezoids intuition in Figure 1a into string diagrams as in Figure 1b. The maps previously embodied as trapezoids now have a fork shape, with some branches being straight lines and some other branches going through boxes. The boxes represent arbitrary morphisms in the underlying category, and the straight lines represent the identity morphisms. Whereas composition of two DIBI states were hand-waved as two trapezoids tiled together in Figure 1a, with string diagram we can define it precisely: the overlap of the two trapezoids is witnessed by the grey wires, and the composition joins two diagrams side-by-side with the grey wires shared. We will show in Section 4 that this string diagram representation yields DIBI models in any category with enough structure to interpret $-\!\bullet\!\!<$, namely, Markov categories [9,13]. Furthermore, in Section 5, we will derive several concrete DIBI models as instances.

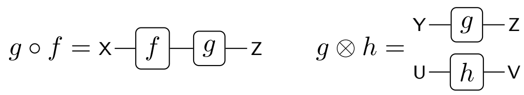
Additionally, our framework enables a comparison between different characterisations of conditional independence (CI). The previous work [5] expresses probabilistic or relational CI in terms of satisfaction of some DIBI formulas. Since we can construct categorical DIBI models based on any Markov categories, we define a logical notion of CI for morphisms in Markov categories as satisfaction of those DIBI formulas. In Section 6, we investigate the relationship between our 'logical' CI and various CI notions based on categorical structures from literature in synthetic statistics [9,13].

Throughout the paper we fix a countably infinite set of variables Var, use $x, y, z, \ldots$ for elements of Var, and use $W, X, Y, \ldots$ for finite subsets of Var.

2 Category Theory Preliminaries

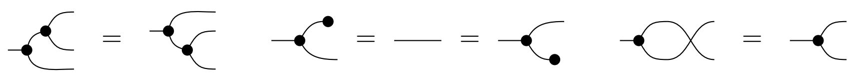
Unless specified, we assume that all monoidal categories we consider are strict and write $dom(f)$ and $cod(f)$ for the domain and codomain of any morphism f. We write $\langle \mathbb{C}, \otimes, \mathsf{I} \rangle$ for a (strict) monoidal category, where $\otimes$ is the monoidal product and I the unit object of $\mathbb{C}$. If is also symmetric, we write $\sigma_{\mathsf{A},\mathsf{B}} : \mathsf{A} \otimes \mathsf{B} \to \mathsf{B} \otimes \mathsf{A}$ for the symmetry natural transformation indexed by objects A and B.

As detailed for instance in [29,26,12], morphisms of symmetric monoidal categories have a graphical presentation as string diagrams, where sequential composition and monoidal product are depicted as concatenation and juxtaposition of diagrams, respectively: given morphisms $f\colon \mathsf{X} \to \mathsf{Y}$, $g\colon \mathsf{Y} \to \mathsf{Z}$, $h\colon \mathsf{U} \to \mathsf{V}$,

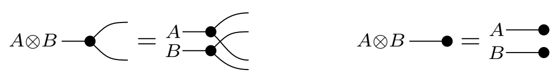


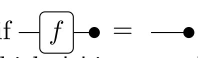
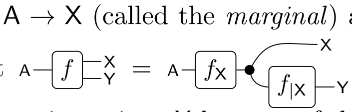
Our convention is to read string diagrams from left to right, and tensor products from top to bottom. We will sometimes omit object labels in the diagrams when they are evident or irrelevant to the context. Symmetries are indicated with the string diagram. We call string diagrams consisting solely of combinations of s *rewirings*: intuitively, they permute the order of the objects.

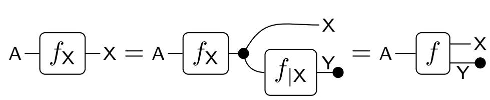
We will need the notion of a Markov category, which suitably generalises categories of probabilistic processes [13]. First, a *copy-delete category* (*CD category*) is a symmetric monoidal category (SMC) $\langle \mathbb{C}, \otimes, \mathsf{I} \rangle$ with 'copy' copy_C and 'delete' del_C morphisms for each object C, drawn diagrammatically as and respectively, that form a commutative comonoid:



Because of the leftmost equation above, we sometimes write a 'trident' for either side of it. Moreover, both copy and del are compatible with the monoidal structure:



We say del *is natural* if  for every morphism f. A *Markov category* is a CD category in which del is natural. A CD category $\mathbb{C}$ *has conditionals* if for each morphism $f\colon \mathsf{A} \to \mathsf{X} \otimes \mathsf{Y}$, there exist (not necessarily unique) morphisms $f_\mathsf{X}\colon \mathsf{A} \to \mathsf{X}$ (called the *marginal*) and $f_{|\mathsf{X}}\colon \mathsf{X} \to \mathsf{Y}$ (called the *conditional*) such that  . When $\mathbb{C}$ is a Markov category, such marginal f_X is unique given X because of the naturality of del:



3 DIBI Logic and its Probabilistic Model

In this section we review the logic of *Dependence and Independence Bunched Implications* (DIBI). For space reasons, we focus on the discrete probabilistic model for DIBI, as introduced in [5]. The interested reader may refer to [5] for the relational model, whose construction follows similar steps.
DIBI formulas are defined inductively as follows:

$$P, Q ::= p \in \mathcal{AP} \mid \top \mid I \mid P \wedge Q \mid P \to Q \mid P * Q \mid P \mathrel{-\!\!*} Q \mid P \fatsemi Q \mid P \multimap Q \mid P \mathrel{\circ\!\!-} Q$$

$$
\begin{array}{llll}
a \oplus b \doteq b \oplus a & (\oplus\text{-}\textsc{Com}) & \exists e \in E\colon a = e \odot a & (\odot\text{-}\textsc{UnitExist}_{\mathrm{L}}) \\
\exists e \in E\colon a = e \oplus a & (\oplus\text{-}\textsc{UnitExist}) & \exists e \in E\colon a = a \odot e & (\odot\text{-}\textsc{UnitExist}_{\mathrm{R}}) \\
(a \oplus b) \oplus c \doteq a \oplus (b \oplus c) & (\oplus\text{-}\textsc{Assoc}) & (a \odot b) \odot c \doteq a \odot (b \odot c) & (\odot\text{-}\textsc{Assoc})
\end{array}
$$

$$
\begin{array}{lr}
e \in E \,\&\, (a \oplus e)\Downarrow \Longrightarrow (a \oplus e) \sqsupseteq a & (\oplus\text{-}\textsc{UnitCoh}) \\
e \in E \,\&\, (a \odot e)\Downarrow \Longrightarrow (a \odot e) \sqsupseteq a & (\odot\text{-}\textsc{UnitCoh}_{\mathrm{R}}) \\
e \in E \,\&\, e' \sqsupseteq e \Longrightarrow e' \in E & (\textsc{UnitClosure}) \\
(a \oplus b)\Downarrow \,\&\, \ a \sqsupseteq a' \,\&\, b \sqsupseteq b' \Longrightarrow (a' \oplus b')\Downarrow \,\&\, (a \oplus b) \sqsupseteq (a' \oplus b') & (\oplus\text{-}\textsc{DownClosed}) \\
(a \odot b)\Downarrow \,\&\, (a \odot b) \sqsubseteq c' \Longrightarrow \exists a', b' \colon a' \sqsupseteq a \,\&\, b' \sqsupseteq b \,\&\, c' = (a' \odot b') & (\odot\text{-}\textsc{UpClosed}) \\
(a_1 \odot a_2) \oplus (b_1 \oplus b_2) \doteq (a_1 \oplus b_1) \odot (a_2 \oplus b_2) & (\textsc{RevExchange})
\end{array}
$$

Fig. 2: DIBI frame conditions (with implicit outermost universal quantifiers), where $\Downarrow$ stands for 'is defined', $\doteq$ means 'equal when either side is defined'.

The additive conjunction $\wedge$ is just the standard Boolean conjunction. The multiplicative conjunction $*$ states that P and Q are independent. Both are already present in BI. DIBI extends BI with the non-commutative conjunction $\fatsemi$[4], where $P \fatsemi Q$ states that Q may depend on P. The operation $-\!\!*$ is adjoint to $*$, and $\multimap$, $\multimapinv$ are adjoints to $\fatsemi$. DIBI formulas are interpreted on DIBI *models*, each consisting of a *DIBI frame* on a set of states A and a *valuation* function $\mathcal{V}\colon \mathcal{AP} \to \mathcal{P}(A)$ that maps an atomic proposition to the set of states on which it is true. While a BI frame is based on a partial commutative monoid [11], a DIBI frame consists of two monoids (one commutative and one not) on the same underlying set, taking care of the two non-additive conjunctions $*$ and $\fatsemi$, respectively.

Definition 1 ([5]). *A* DIBI frame *is a tuple $\mathcal{A} = \langle A, \sqsubseteq, \oplus, \odot, E\rangle$, where A is a set of states, $\sqsubseteq$ is a preorder on A, $E \subseteq A$ are units, and $\oplus, \odot\colon A \times A \to \mathcal{P}(A)$ are partial binary operations*[5]*, satisfying the frame conditions in Figure 2.*

The operations $\odot$ and $\oplus$ are referred to as the sequential and parallel compositions of states. Intuitively, $a \sqsubseteq b$ says that a can be extended to b, and E is the set of states that act as units for these operations.

For capturing conditional independence, atomic propositions $\mathcal{AP}$ have the form $S \triangleright [T]$, for finite sets of variables S, T. Roughly, $S \triangleright [T]$ means the values of variables in T only depend on that of S. We now recall the semantics of DIBI formulas, restricting to the fragment that is necessary for the current work.

---

[4] Not to be confused with the additive context constructor which is also denoted as $\fatsemi$ in the standard BI literature such as [23,27].

[5] Note that, even though $\odot, \oplus$ are also partial in the models considered in [5], they have type $A \times A \to \mathcal{P}(A)$ in that work. This is because the authors obtain completeness of DIBI logic using a method developed by Docherty [11], which only works for the more general type. Because the operations are actually partial rather than non-deterministic, and we are not interested in completeness here, we stick to the more accurate type.

Definition 2. *Given a DIBI model* $\langle \mathcal{A}, \mathcal{V} \rangle$, satisfaction $\models_{\mathcal{V}}$ *of DIBI*$_{\{\wedge,*,\fatsemi\}}$*-formulas at* $\mathcal{A}$*-states is inductively defined as follows:*

$$
\begin{array}{lll}
a \models_{\mathcal{V}} I \ \text{ iff } \ a \in E & & a \models_{\mathcal{V}} \top \ \text{ always} \\
a \models_{\mathcal{V}} (A \triangleright [B]) & \text{iff} & a \in \mathcal{V}(A \triangleright [B]) \\
a \models_{\mathcal{V}} P \wedge Q & \text{iff} & a \models_{\mathcal{V}} P \text{ and } a \models_{\mathcal{V}} Q \\
a \models_{\mathcal{V}} P * Q & \text{iff} & \exists b_1, b_2 \text{ such that } b_1 \oplus b_2 \sqsubseteq a,\ b_1 \models_{\mathcal{V}} P,\ b_2 \models_{\mathcal{V}} Q \\
a \models_{\mathcal{V}} P \fatsemi Q & \text{iff} & \exists b_1, b_2 \text{ such that } b_1 \odot b_2 = a,\ b_1 \models_{\mathcal{V}} P,\ b_2 \models_{\mathcal{V}} Q
\end{array}
$$

For a concrete example of DIBI models, we review the probabilistic models. Let Val be a set of values, to which variables in Var are assigned. A *memory over* a finite set of variables X is a function $\mathbf{m} \colon X \to \text{Val}$, and the *memory space* over X is the set of all memories over X, denoted as $\mathbf{M}[X; \text{Val}]$, or $\mathbf{M}[X]$ when Val is clear. Given a memory $\mathbf{m} \in \mathbf{M}[X]$ and a subset $U \subseteq X$, the memory $\mathbf{m}^U : U \to \text{Val}$ is the restriction of $\mathbf{m}$ to the domain U. Given a set S, $\mathcal{D}S$ is the set of discrete distributions over S; that is, for any $\mu \in \mathcal{D}S$, we have $\mu \colon S \to [0,1]$, the support $\text{supp}(\mu) = \{s \in S \mid \mu(s) > 0\}$ is finite, and $\sum_{s \in S} \mu(s) = 1$. To capture conditional independence, we now introduce the notion of *probabilistic input-preserving kernels.*

Definition 3 ([5]). *A* probabilistic input-preserving kernel *(or* probabilistic kernel *for short) is a function* $f \colon \mathbf{M}[X] \to \mathcal{D}\mathbf{M}[Y]$ *satisfying: (i)* $X \subseteq Y$, *(ii)* $\pi_X f = \eta^{\mathcal{D}}_{\mathbf{M}[X]}$. *The set of all probabilistic kernels is denoted* ProbKer.

In words, a probabilistic kernel f maps a memory $\mathbf{m}$ on X to a distribution of memories on $Y \supseteq X$ whose support contains only memories $\mathbf{m}'$ that faithfully extends $\mathbf{m}$ (thus the name 'input-preserving'). Alternatively, f can be seen as a conditional distribution $\Pr(Y \mid X)$ where $Y \supseteq X$, such that $\Pr(Y = B \mid X = A)$ is nonzero only if B restricted to the domain X equals A.

Definition 4 (Probabilistic model, [5]). *The* probabilistic frame $\mathbf{PrFr}[\text{Val}]$ based on Val *is a tuple* $\langle \text{ProbKer}, \sqsubseteq, \oplus, \odot, \text{ProbKer} \rangle$ *where* $\odot, \oplus, \sqsubseteq$ *are defined for arbitrary* $f \colon \mathbf{M}[X] \to \mathcal{D}\mathbf{M}[Y]$ *and* $g \colon \mathbf{M}[Z] \to \mathcal{D}\mathbf{M}[W]$ *as:*

- *the sequential composition* $f \odot g$ *is defined iff* $Y = Z$. *In this case,* $f \odot g$ *is of the form* $\mathbf{M}[X] \to \mathcal{D}\mathbf{M}[W]$, *and given* $\mathbf{m} \in \mathbf{M}[X]$, $(f \odot g)(\mathbf{m})$ *maps* $\mathbf{n} \in \mathbf{M}[W]$ *to* $\sum_{\boldsymbol{\ell} \in \text{supp}(f(\mathbf{m}))} g(\boldsymbol{\ell})(\mathbf{n})$;
- *the parallel composition* $f \oplus g$ *is defined iff* $X \cap Z = Y \cap W$. *In this case,* $f \oplus g$ *is of the form* $\mathbf{M}[X \cup Z] \to \mathcal{D}\mathbf{M}[Y \cup W]$ *such that given* $\boldsymbol{\ell} \in \mathbf{M}[X \cup Z]$ *and* $\mathbf{m} \in \mathbf{M}[Y \cup W]$, *we have* $(f \oplus g)(\boldsymbol{\ell})(\mathbf{m}) = f(\boldsymbol{\ell}^X)(\mathbf{m}^Y) \cdot g(\boldsymbol{\ell}^Z)(\mathbf{m}^W)$;
- *the subkernel relation* $f \sqsubseteq g$ *holds if there exist a finite set of variables* S *and* $h \in \text{ProbKer}$ *such that* $g = \left(f \oplus \eta^{\mathcal{D}}_{\mathbf{M}[S]}\right) \odot h$.

The probabilistic model based on Val *consists of the probabilistic frame* $\mathbf{PrFr}[\text{Val}]$ *and the following* natural valuation $\mathcal{V}_{\text{nat}} \colon \mathcal{AP} \to \mathcal{P}(\text{ProbKer})$*: given* $(S \triangleright [T])$ *and* $f \colon \mathbf{M}[X] \to \mathcal{D}\mathbf{M}[Y]$, $f \in \mathcal{V}_{\text{nat}}(S \triangleright [T])$ *iff there exists a probabilistic kernel* $f' \colon \mathbf{M}[X'] \to \mathcal{D}\mathbf{M}[Y']$ *such that* $f' \sqsubseteq f$, $X' = S$ *and* $T \subseteq Y'$.

We may simply write **PrFr** when the underlying set of values Val is evident.

Next we give examples of probabilistic kernels and how they compose. We will abbreviate a map from a variable x to a value c as c_x and use the ket notation $|\omega\rangle$ for each probabilistic outcome ω.

Example 1. Consider variables x, y, z that take values in $\mathbf{Val} = \{0, 1\}$. We define a map $f\colon \mathbf{M}[\{z\}] \to \mathcal{D}\mathbf{M}[\{x, y, z\}]$ by:

$$f(0_z) = \frac{1}{4}|0_x, 0_y, 0_z\rangle + \frac{1}{4}|0_x, 1_y, 0_z\rangle + \frac{1}{4}|1_y, 0_y, 0_z\rangle + \frac{1}{4}|1_y, 1_y, 0_z\rangle$$
$$f(1_z) = \frac{1}{16}|0_x, 0_y, 1_z\rangle + \frac{3}{16}|0_x, 1_y, 1_z\rangle + \frac{3}{16}|1_y, 0_y, 1_z\rangle + \frac{9}{16}|1_y, 1_y, 1_z\rangle$$

The input memory (coloured) is preserved by f so it is a probabilistic kernel. Then define $g_1\colon \mathbf{M}[\{z\}] \to \mathcal{D}\mathbf{M}[\{x, z\}]$ and $g_2\colon \mathbf{M}[\{z\}] \to \mathcal{D}\mathbf{M}[\{y, z\}]$ as:

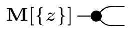
$$g_1(0_z) = \frac{1}{2}|0_x, 0_z\rangle + \frac{1}{2}|1_y, 0_z\rangle \qquad g_1(1_z) = \frac{1}{4}|0_x, 1_z\rangle + \frac{3}{4}|1_y, 1_z\rangle$$
$$g_2(0_z) = \frac{1}{2}|0_y, 0_z\rangle + \frac{1}{2}|1_y, 0_z\rangle \qquad g_2(1_z) = \frac{1}{4}|0_y, 1_z\rangle + \frac{3}{4}|1_y, 1_z\rangle$$

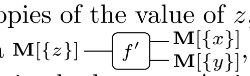
Both g_1 and g_2 are probabilistic kernels as well. The parallel composition $g_1 \oplus g_2$ is defined since $\{z\} \cap \{z\} = \{x, z\} \cap \{y, z\}$; in fact, it is easy to verify that $g_1 \oplus g_2 = f$. Moreover, g_1 and g_2 can be obtained by projecting the output of f on $\{x, z\}$ and $\{y, z\}$, respectively, and we can show $g_1 \sqsubseteq f$ and $g_2 \sqsubseteq f$.

4 DIBI models in Markov categories

In this section we construct more abstract DIBI models based on categorical structures. The starting point of our approach is a categorical characterisation of the concrete probabilistic models given above. In the following, we begin by showing examples of how elements in that model can be reformulated in categorical terms and then formally present our categorical construction of DIBI models.

As we noted in Section 1, the probabilistic DIBI kernels can be identified as morphisms in the Kleisli category for the distribution monad $\mathcal{K}\ell(\mathcal{D})$ (Definition 15); however, not all morphisms in $\mathcal{K}\ell(\mathcal{D})$ are probabilistic DIBI kernels, so we need to define the extra conditions categorically. Let $\mathbb{M}\mathrm{em}\mathbb{P}\mathrm{r}$ be the subcategory of $\mathcal{K}\ell(\mathcal{D})$ where objects are restricted to memory spaces over Val. That is, the objects are memory spaces $\mathbf{m}\colon X \to \mathrm{Val}$, and the morphisms are maps $f : \mathbf{M}[X] \to \mathcal{D}\mathbf{M}[Y]$. Then, probabilistic kernels are exactly those morphisms in the $\mathbb{M}\mathrm{em}\mathbb{P}\mathrm{r}$ that satisfy the input-preserving condition in Definition 3. To visualise the probabilistic kernels, we depict them using string diagrams, which is possible because $\mathbb{M}\mathrm{em}\mathbb{P}\mathrm{r}$ is a subcategory of $\mathcal{K}\ell(\mathcal{D})$ and $\mathcal{K}\ell(\mathcal{D})$ has the monoidal structure. We also observe that the codomain of an input-preserving kernel $f\colon \mathbf{M}[X] \multimap \mathbf{M}[Y]$ can be decomposed as $\mathbf{M}[X] \times \mathbf{M}[Y \setminus X]$, Recall the probabilistic kernel f from Example 1. Since its codomain $\mathbf{M}[\{x, y, z\}]$ can be decomposed as $\mathbf{M}[\{x\}] \times \mathbf{M}[\{y\}] \times \mathbf{M}[\{z\}]$, we can draw it as follows:

Intuitively, 

produces two copies of the value of z, and the values of x and y are computed from that of z via 

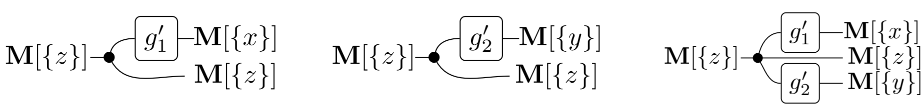
, while the value of z gets preserved through a straight wire in the bottom. As in this example, such copy structure of $\mathcal{K}\ell(\mathcal{D})$ enables us to capture the 'input-preserving' condition of probabilistic kernels generally.

Next we want to express the sequential ($\odot$) and parallel ($\oplus$) compositions of probabilistic kernels categorically. The former is exactly the sequential composition of morphisms in $\mathcal{K}\ell(\mathcal{D})$. The parallel composition, however, is *not* the monoidal product in $\mathcal{K}\ell(\mathcal{D})$, i.e., the Cartesian product. By definition, the monoidal product is total, while the parallel composition is partial. Even when the parallel composition is defined, the types of the resulting morphisms do not match. Suppose that the parallel composition of $f\colon \mathbf{M}[X] \rightharpoondown \mathbf{M}[Y]$ and $g\colon \mathbf{M}[U] \rightharpoondown \mathbf{M}[V]$ is defined, we have

$$f \oplus g\colon \mathbf{M}[X \cup U] \rightharpoondown \mathbf{M}[Y \cup V] \quad g_1 \otimes g_2\colon \mathbf{M}[X] \times \mathbf{M}[U] \rightharpoondown \mathbf{M}[Y] \times \mathbf{M}[V]$$

The key difference is that parallel composition considers a single memory that can be projected into two pieces, while the monoidal product considers the cartesian product of two pieces of memory, no matter if they agree or not on overlapped variables. To combine $\mathbf{M}[X]$ and $\mathbf{M}[U]$ into $\mathbf{M}[X \cup U]$ categorically, we note that for disjoint Z_1, Z_2, $\mathbf{M}[Z_1 \cup Z_2] \cong \mathbf{M}[Z_1] \times \mathbf{M}[Z_2]$, therefore $\mathbf{M}[X \cup U] \cong \mathbf{M}[X \setminus U] \times \mathbf{M}[X \cap U] \times \mathbf{M}[U \setminus X]$. Thus we can illustrate the parallel composition of two probabilistic kernels as in the following example.

Example 2. The probabilistic kernels g_1 and g_2 from Example 1 – seen as $\mathcal{K}\ell(\mathcal{D})$-morphisms – are drawn as the first and second string diagram below respectively.



where $g_1'\colon \mathbf{M}[\{z\}] \rightharpoondown \mathbf{M}[\{x\}]$ and $g_2'\colon \mathbf{M}[\{z\}] \rightharpoondown \mathbf{M}[\{y\}]$ represent the conditional distributions obtained by suitable projections of g_1 and g_2, respectively. The parallel composition $g_1 \oplus g_2$ is given by the rightmost string diagram above.

We omit the formal string diagrammatic definitions here as they can be easily derived from their counterparts in the generic construction of DIBI models, defined later in Definition 7.

Towards that categorical construction of DIBI models, we also want to generalise the concept of memory spaces $\mathbf{M}[X]$, which were customised for reasoning about probabilistic programs and relational databases. We observe that the side conditions of the parallel and sequential compositions are all based on comparing the set of variables in the (co)domains, so they only depend on the X part in $\mathbf{M}[X]$. This motivates us to define DIBI states as morphisms in a category whose objects are made of variables in Definition 5.

To express finite sets of variables and the union of disjoint such sets in a monoidal category, in the following we will represent finite sets of variables as lists. We impose a linear order $\preceq$ on Var such that indexed variables inherit the order of their indices, e.g., $x_1 \preceq x_2 \preceq x_3$. Let $x \prec y$ abbreviate for $x \preceq y$ and $x \neq y$. Then, finite sets of variables can be represented as finite lists of variables ordered by $\prec$, via a translation that we write as $[\![\cdot]\!]$. For instance, $[\![\{x_3, x_1, x_3, x_4\}]\!] = [x_1, x_3, x_4]$.

Throughout the rest of the section we fix a Markov category $\langle \mathbb{C}, \otimes, \mathsf{I} \rangle$, and assignment $\theta\colon \mathrm{Var} \to \mathbf{ob}(\mathbb{C})$ of a $\mathbb{C}$-object to each $x \in \mathrm{Var}$.

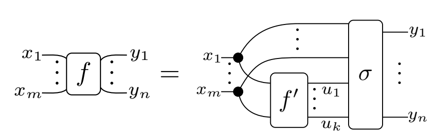
Definition 5. *Let $\mathbb{C}[\theta]$ be the symmetric monoidal category whose objects are finite lists of variables, and morphisms $[x_1, \dots, x_m] \to [y_1, \dots, y_n]$ are $\mathbb{C}$-morphisms $\theta(x_1) \otimes \cdots \otimes \theta(x_m) \to \theta(y_1) \otimes \cdots \otimes \theta(y_n)$. Sequential composition is defined as in $\mathbb{C}$. The identity on $[x_1, \dots, x_m]$ is $id_{\theta x_1 \otimes \cdots \otimes \theta x_m}$. The monoidal product in $\mathbb{C}[\theta]$ – which we also write as $\otimes$ with abuse of notation – is list concatenation on objects, and monoidal product in $\mathbb{C}$ on morphisms.*

Since $\mathbb{C}$ is a Markov category, it follows immediately that $\mathbb{C}[\theta]$ is also a Markov category. Sometimes we restrict ourselves to a uniform assignment θ; that is, for some fixed $\mathsf{C} \in \mathbf{ob}(\mathbb{C})$, $\theta(x) = \mathsf{C}$ for all $x \in \mathrm{Var}$. This is in line with the scenario where a fixed value space Val is used for all variables (see Definition 3). In this case, we write $\mathbb{C}[\theta]$ as $\mathbb{C}[\mathsf{C}]$ to emphasise the uniform value of the assignment. This category can be seen as the full subcategory of $\mathbb{C}$ freely generated by C, but with each occurrence of the generating object named by a variable. The next example shows how the construction in Definition 5 selects morphisms of $\mathcal{K}\ell(\mathcal{D})$ that act on memory spaces, among which we have all the probabilistic kernels.

Example 3. Let $\mathbb{C}$ be $\mathcal{K}\ell(\mathcal{D})$, and $\theta\colon \mathrm{Var} \to \mathbf{ob}(\mathcal{K}\ell(\mathcal{D}))$ be the constant function $x \mapsto \mathrm{Val}$ for all $x \in \mathrm{Var}$. Then there is a full and faithful embedding functor $\iota\colon \mathbb{M}\mathrm{em}\mathbb{P}\mathrm{r} \to \mathcal{K}\ell(\mathcal{D})[\theta]$: on objects, given a set X, $\iota(\mathbf{M}[X]) = [\![X]\!]$; on morphisms, given $f\colon \mathbf{M}[X] \to \mathcal{D}\mathbf{M}[Y]$ with $X = \{x_1, \dots, x_m\}$ and $Y = \{y_1, \dots, y_n\}$, its image $\iota(f)\colon X \to Y$ is the composed map $\mathrm{Val}^m \xrightarrow{\cong} \mathbf{M}[X] \xrightarrow{f} \mathcal{D}\mathbf{M}[Y] \xrightarrow{\mathcal{D}\cong} \mathcal{D}\mathrm{Val}^n$, where the isomorphisms are, e.g., $\mathbf{M}[Y] \xrightarrow{\cong} \mathbf{M}[y_1] \times \cdots \times \mathbf{M}[y_n] \xrightarrow{\cong^\theta} \mathrm{Val}^n$, using the valuation $\theta(y_j) = \mathrm{Val}$.

Now that we have abstracted the concept of memory spaces used in concrete DIBI models as objects in Markov categories, the states of DIBI models — the role that in the probabilistic models is filled by probabilistic kernels — takes the form of morphisms in these categories. Next we identify those $\mathbb{C}$-morphisms that constitute the states of DIBI models. We call them the *input-preserving kernels* in $\mathbb{C}$, written $\mathrm{Ker}(\mathbb{C}[\theta])$.

Definition 6. *A $\mathbb{C}[\theta]$-morphism $f\colon [x_1, \dots, x_m] \to [y_1, \dots, y_n]$ is a $\mathbb{C}[\theta]$* input-preserving kernel *(or $\mathbb{C}[\theta]$-*kernel *for short) if $x_1 \prec \cdots \prec x_m$, $y_1 \prec \cdots \prec y_n$, and f can be decomposed as follows, where σ is rewiring:*

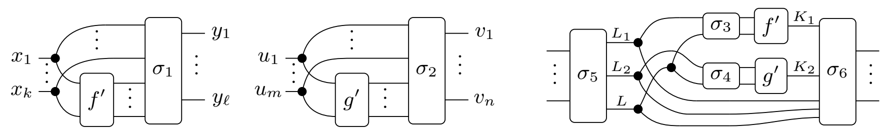


(1)

In words, a $\mathbb{C}[\theta]$-kernel is a morphism whose interfaces are essentially finite sets of variables, such that the input is preserved as part of the output (through the upper leg of those —●⊂s). The map f' in (1) is referred to as the nontrivial part of the input-preserving kernel. It follows from Definition 6 that, for a $\mathbb{C}[\theta]$-kernel, its codomain $[y_1, \ldots, y_n]$ always subsumes its domain $[x_1, \ldots, x_m]$; also, $u_1, \ldots, u_k$ are precisely those y_js that are not among these x_is. Since the (co)domains of $\mathbb{C}[\theta]$-kernel are list presentation of sets, we also write the types of $\mathbb{C}[\theta]$-kernels using the corresponding sets, e.g., in (1), $f\colon \{x_1, \ldots, x_m\} \to \{y_1, \ldots, y_n\}$.

Next we define compositions on input-preserving kernels, generalising what we have seen in Example 2 for the probabilistic models.

Definition 7 (Compositions). *Given arbitrary $\mathbb{C}[\theta]$-kernels $f\colon X \to Y$ and $g\colon U \to V$ as in Figure 3a, their* sequential composition *$f \odot g$ is defined iff $cod(f) = dom(g)$, in which case $f \odot g = g \circ f$. Their* parallel composition *$f \oplus g$ is defined iff $X \cap U = Y \cap V$. Assume $L = [\![X \cap U]\!]$, $L_1 = [\![X \setminus (X \cap U)]\!]$, $L_2 = [\![U \setminus (X \cap U)]\!]$, $K_1 = [\![Y \setminus (Y \cap V)]\!]$, and $K_2 = [\![V \setminus (Y \cap V)]\!]$, then $f \oplus g\colon X \cup U \to Y \cup V$ is defined as in Figure 3b, where all the σ_is are rewiring morphisms for making the input and output variables $\prec$-ordered.*

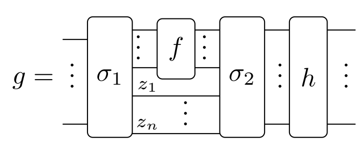


(a) $\mathbb{C}[\theta]$-kernels f and g. (b) $f \oplus g$

Fig. 3: Parallel composition of $\mathbb{C}[\theta]$.

Note here a benefit of the diagrammatic representation: we can easily identify the memory overlap $\mathbf{M}[X \cap Y]$, as it is depicted a separate wire; with traditional syntax, we would need to apply associativity and commutativity to extract it from $\mathbf{M}[X \cup Y]$. It is easy to see that kernels are closed under compositions. Also, for curious readers, we note that $\mathbb{C}[\theta]$-kernels with their parallel compositions form a *partially monoidal category* [3]. Next we define the subkernel relation.

Definition 8 (Subkernel). *Given two $\mathbb{C}[\theta]$-kernels f and g, we say f is a* subkernel *of g – denoted as $f \sqsubseteq g$ – if there exist $z_1, \cdots, z_n \in$* Var*, a $\mathbb{C}[\theta]$-kernel h, and rewiring morphisms σ_1, σ_2 such that g can expressed as on the right-hand side.*



The subkernel relation is transitive and reflexive, which can be shown simply by manipulations of the string diagram. We are finally able to state the main result of this section: $\mathbb{C}[\theta]$-kernels and their compositions form a DIBI frame.

Theorem 1. $\mathbf{Fr}(\mathbb{C}[\theta]) = \langle \mathrm{Ker}(\mathbb{C}[\theta]), \sqsubseteq, \oplus, \odot, \mathrm{Ker}(\mathbb{C}[\theta]) \rangle$ *is a DIBI frame.*

Also, under the natural valuation $\mathcal{V}_{\mathrm{nat}}$, a $\mathbb{C}[\theta]$-kernel $f\colon X \to Y$ satisfies $S \triangleright [T]$ iff there is a subkernel $(f'\colon X' \to Y') \sqsubseteq f$ such that $X' = S$ and $Y' \supseteq T$. Thus:

Corollary 1. $(\mathbf{Fr}(\mathbb{C}[\theta]), \mathcal{V}_{\mathrm{nat}})$ *is a DIBI model.*

We will see in Section 5 how to use this categorical construction to derive a wide range of DIBI models. Moreover, it also helps to extract properties of the underlying category that are essential to a specific feature of a DIBI model. Here is an example.

Proposition 1. *If $\mathbb{C}$ further satisfies that for arbitrary morphisms f, g and object D, $f \otimes \mathsf{del}_{\mathsf{D}} = g \otimes \mathsf{del}_{\mathsf{D}}$ implies $f = g$, then subkernel is unique given its type in the following sense: if $\mathbb{C}[\theta]$-kernels $f_1, f_2 \colon U \to V$ are both subkernels of g, then $f_1 = f_2$.*

Note that the uniqueness of subkernels has been observed already in the context of probabilistic and relational models, see [5, Sect. IV]. Proposition 1 reveals the general conditions under which we have this uniqueness for a wider class of DIBI models. In particular, we will see in Section 5.1 how the probabilistic models meets the conditions of Proposition 1.

5 Examples

In this section we provide concrete instances of the categorical construction in Section 4. The first example concerns the probabilistic DIBI models. The remaining examples are new DIBI models. Some of them have been suggested in the DIBI paper [5], yet not materialised due to the complexity involved in stating each component and verifying the frame conditions. Within our framework, these steps become much easier to perform.

5.1 Probabilistic (and Relational) DIBI Models

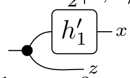
As we sketched in Example 2 and Example 3, the probabilistic DIBI kernels and $\langle \mathbf{Fr}(\mathcal{K}\ell(\mathcal{D})), \mathcal{V}_{\mathrm{nat}} \rangle$ input-preserving kernels correspond to each other. We now formally show that the probabilistic DIBI model in Definition 4 can be recovered from the categorical DIBI model $\langle \mathbf{Fr}(\mathcal{K}\ell(\mathcal{D})), \mathcal{V}_{\mathrm{nat}} \rangle$. Since both models are equipped with the natural valuation $\mathcal{V}_{\mathrm{nat}}$, we focus on the frame part. To make the correspondence precise, we introduce the category of DIBI frames, as hinted in [5, Sect. III].

Definition 9. *In* the category of DIBI frames $\mathbb{D}\mathbb{i}\mathbb{b}\mathbb{i}\mathbb{F}\mathbb{r}$, *objects are DIBI frames; morphisms $f \colon \langle S, \sqsubseteq_S, \oplus_S, \odot_S, E_S \rangle \to \langle T, \sqsubseteq_T, \oplus_T, \odot_T, E_T \rangle$ are functions $f \colon S \to T$ that respect all the relations and partial operations: for arbitrary $s, s' \in S$,*

- *$s \sqsubseteq_S s'$ implies $f(s) \sqsubseteq_T f(s')$;*
- *if $s \star_S s'$ is defined, then $f(s) \star_S f(s')$ is defined, and $f(s) \star_T f(s') = f(s \star_S s')$, for $\star \in \{\oplus, \odot\}$;*
- *$s \in E_S$ implies $f(s) \in E_T$.*

It turns out that the function ι introduced in Example 3 extends to an isomorphism of DIBI frames from $\mathbf{PrFr}[\mathrm{Val}]$ to $\mathbf{Fr}(\mathcal{K}\ell(\mathcal{D})[\mathrm{Val}])$.

Proposition 2. $\mathbf{PrFr}[\mathrm{Val}] \cong \mathbf{Fr}(\mathcal{K}\ell(\mathcal{D})[\mathrm{Val}])$.

Example 4. The probabilistic kernel $g_1 : \mathbf{M}[\{z\}] \to \mathcal{D}\mathbf{M}[\{x, z\}]$ from Example 1 corresponds to the following $\mathcal{K}\ell(\mathcal{D})[\{0,1\}]$-kernel $h_1 : [z] \to [x, z]$ – i.e., a $\mathcal{K}\ell(\mathcal{D})$-morphism $\{0, 1\} \to \{0, 1\}^2$ – where: $0 \mapsto \frac{1}{2}|0, 0\rangle + \frac{1}{2}|1, 0\rangle$, $1 \mapsto \frac{1}{4}|0, 1\rangle + \frac{3}{4}|1, 1\rangle$. Diagrammatically, h_1 is of the form 

, where $h'_1 : [z] \to [x]$ is the map such that $0 \mapsto \frac{1}{2}|0\rangle + \frac{1}{2}|1\rangle$ and $1 \mapsto \frac{1}{4}|0\rangle + \frac{3}{4}|1\rangle$

Similarly, the relational DIBI model from [5] with the value space Val can be shown to be isomorphic to $\mathbf{Fr}(\mathcal{K}\ell(\mathcal{P}_i)[\mathrm{Val}])$, where $\mathcal{P}_i$ is the nonempty powerset monad.

5.2 Stochastic DIBI Models

Using our categorical construction, we can derive a notion of DIBI model for continuous probabilistic (stochastic) processes, not previously considered. This is of interest because, as we will later show in section 6, it allows to capture conditional independence for continuous probability using DIBI formulas. We take as underlying category $\mathbb{Stoch}$ of stochastic processes, defined as the Kleisli category $\mathcal{K}\ell(\mathcal{G})$ for the Giry monad on measurable spaces – see Appendix A for a full definition. Since $\mathcal{G}$ is an affine symmetric monoidal monad, $\mathbb{Stoch}$ is a Markov category [13]. Applying Theorem 1 to $\mathbb{C} = \mathbb{Stoch}$, we get DIBI frames based on stochastic processes.

Proposition 3. *Given an arbitrary map* $\theta : \mathrm{Var} \to \mathbf{ob}(\mathbb{Meas})$*,* $\mathbf{Fr}(\mathbb{Stoch})[\theta]) = \langle \mathrm{Ker}(\mathbb{Stoch}[\theta]), \sqsubseteq, \oplus, \odot, \mathrm{Ker}(\mathbb{Stoch}[\theta]) \rangle$ *is a DIBI frame.*

We call $\mathbf{Fr}(\mathbb{Stoch}[\theta])$ the *stochastic DIBI frame* based on θ and elements in $\mathrm{Ker}(\mathbb{Stoch}[\theta])$ stochastic kernels.

Example 5. We show a representation of the *Box-Muller transformation* using stochastic kernels. Consider θ that maps all variable names to the Borel σ-algebra over reals $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Define stochastic kernels $g_1 : \varnothing \to \{u\}$ and $g_2 : \varnothing \to \{w\}$ – both standing for $\mathbb{Stoch}$-morphisms $(\mathbf{1}, \{\varnothing, \mathbf{1}\}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, or equivalently, a probabilistic measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ – by $g_i(\bullet) = \textsc{Unif}(0, 1)$ for $i = 1, 2$, where $\textsc{Unif}(0, 1)$ is the uniform measure over the interval $(0, 1)$. Such a uniform measure over infinite outcomes is not possible in the discrete probabilistic DIBI model. Define another stochastic kernel $f : \{u, w\} \to \{u, w, x, y\}$ where the value of x, y are determined by the value of u, w:

$$f(u \mapsto v_u, w \mapsto v_w) = \delta_{v_u, v_w, \left(\sqrt{-2 \ln u} \cdot \cos(2\pi w)\right)_x, \left(\sqrt{-2 \ln u} \cdot \sin(2\pi w)\right)_y}.$$

Then $h = (g_1 \oplus g_2) \odot f$ gives a stochastic kernel $\varnothing \to \{u, w, x, y\}$. Box-Muller transformation says that x and y are independent in $h(\langle\rangle)$, despite their seemingly dependence on u and w. We will explain more on this in Example 7.

Comparison with Lilac [21]. Our stochastic DIBI models can be used to reason about independence and conditional probabilities in continuous distributions. A recent work Lilac by Li et al. [21] proposed a BI model for the same goal, yet with some crucial differences in the set-up.

First, the states in Lilac's BI model are probabilistic space fragments of a fixed sample space, and their variables are mathematical random variables that deterministically map elements in the sample space to values. In comparison, we treat variables as names that can be associated to values or distributions. Our stochastic kernels – though not using an ambient sample space – can encode their set-up: we can devise a special variable Ω for 'the sample space,' and deterministic kernels from Ω to other variables encode random variables.

Second, to reason about conditional probabilities, Lilac want probability spaces to be disintegrable with respect to well-behaved random variables. To achieve that, they require probability spaces in their model to be extensible to Borel spaces, since disintegration works nicer in Borel spaces. By working with kernels, which already represents conditional probability spaces, we do not need to impose disintegratability on our DIBI states to reason about conditional probabilities. For instance, while disintegration is not always possible in the category $\mathbb{Stoch}$, we can still construct a DIBI model based on $\mathbb{Stoch}$.

Other measure-theoretic probabilistic DIBI models. The category $\mathbb{Stoch}$ is not the only Markov category for measure-theoretic probability. Another choice is $\mathbb{BorelStoch}$, a subcategory of $\mathbb{Stoch}$ obtained by restricting to standard Borel spaces as objects. It has some nice properties that $\mathbb{Stoch}$ does not satisfy, such as having conditionals as mentioned above. $\mathbb{BorelStoch}$ is also a Markov category and we can easily instantiate a DIBI model.

Proposition 4. *Given any map* $\theta\colon \mathrm{Var} \to \mathbf{ob}(\mathbb{BorelStoch})$*,* $\mathbf{Fr}(\mathbb{BorelStoch}[\theta])$ *defined as* $\langle \mathrm{Ker}(\mathbb{BorelStoch}[\theta]), \sqsubseteq, \oplus, \odot, \mathrm{Ker}(\mathbb{BorelStoch}[\theta])\rangle$ *is a DIBI frame.*

The study of measure theory is also intertwined with topology, and another category for measure-theoretic probability is the Kleisli category of the *Radon monad* $\mathcal{R}$ based on the category of compact Hausdorff spaces $\mathbb{CHaus}$ and continuous maps (cf. Appendix A), which we will denote as $\mathcal{K}\ell_{\mathbb{CHaus}}(\mathcal{R})$. $\mathcal{K}\ell_{\mathbb{CHaus}}(\mathcal{R})$ is also a Markov category [13], so Theorem 1 applies.

Proposition 5. *Given any map* $\theta : \mathrm{Var} \to \mathbf{ob}(\mathcal{K}\ell_{\mathbb{CHaus}}(\mathcal{R}))$*,* $\mathbf{Fr}(\mathcal{K}\ell_{\mathbb{CHaus}}(\mathcal{R})[\theta])$ *defined as* $\langle \mathrm{Ker}(\mathcal{K}\ell_{\mathbb{CHaus}}(\mathcal{R})[\theta]), \sqsubseteq, \oplus, \odot, \mathrm{Ker}(\mathcal{K}\ell_{\mathbb{CHaus}}(\mathcal{R})[\theta])\rangle$ *is a DIBI frame.*

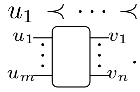
A measure-theoretic Markov category not formed as Kleisli categories is the Gaussian probability category $\mathbb{Gauss}$ [13]. Its objects are natural numbers, and a morphism $n \to m$ is a tuple (M, σ^2, μ) representing the function $f : \mathbb{R}^n \to \mathbb{R}^m$ with $f(v) = M \cdot v + \xi$, where ξ is the Gaussian noise with mean μ and covariance matrix σ^2. Its monoidal product is addition $+$ on the objects and vector concatenation on morphisms. $\mathbb{Gauss}$ differs from $\mathbb{Stoch}, \mathbb{BorelStoch}$ and $\mathcal{K}\ell_{\mathbb{CHaus}}(\mathcal{R})$ in that it does not arise as the Kleisli category associated to some monad. But since it is a Markov category, we can again instantiate DIBI models based on $\mathbb{Gauss}$.

Proposition 6. *Given any map* $\theta : \mathrm{Var} \to \mathbf{ob}(\mathbb{Gauss})$*,* $\mathbf{Fr}(\mathbb{Gauss}[\theta])$ *defined as* $\langle \mathrm{Ker}(\mathbb{Gauss}[\theta]), \sqsubseteq, \oplus, \odot, \mathrm{Ker}(\mathbb{Gauss}[\theta]) \rangle$ *is a DIBI frame.*

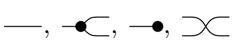
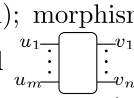
5.3 Syntactic DIBI Models

The DIBI models defined so far all have kernels defined by some processes over memory spaces. It is worth considering a different flavour: purely formal, syntactically generated DIBI models. We start by defining the underlying category.

Definition 10. $\mathbb{SynVar}$ *is the Markov category freely generated as follows:*

- *the generating objects are variables in* Var*;*
- *there is exactly one generating morphism of type* $[u_1, \dots, u_m] \to [v_1, \dots, v_n]$ *for distinct variables* $u_1 \prec \cdots \prec u_m$ *and* $v_1 \prec \cdots \prec v_n$*, written as string diagrams of the form* 

.

In words, $\mathbb{SynVar}$-objects are finite lists of variables (without the requirements of duplicate-free or $\preceq$-ordered); morphisms are diagrams freely concatenated using , and 

, taken equivalence class modulo the Markov category equations. The syntactic DIBI frame is based on the category $\mathbb{SynVar}[id]$, where $id\colon \mathrm{Var} \to \mathbf{ob}(\mathbb{SynVar})$ is the identity function.

Proposition 7. $\mathbf{SynFr} = \langle \mathrm{Ker}(\mathbb{SynVar}[id]), \sqsubseteq, \oplus, \odot, \mathrm{Ker}(\mathbb{SynVar}[id]) \rangle$ *is a DIBI frame.*

Equipped with the natural valuation $\mathcal{V}_{\mathrm{nat}}$, one obtains a DIBI model $\langle \mathbf{SynFr}, \mathcal{V}_{\mathrm{nat}} \rangle$. We postpone an example of $\mathbb{SynVar}[id]$-kernels till Section 6, Example 9, in which $\mathbb{SynVar}[id]$-kernels are used to distinguish two notions of conditional independence in Markov categories.

An interesting question for future work is how to extend the syntactic DIBI model to a term model. Typically being initial objects in categories of models, term models play an important role in proving completeness and defining categorical semantics for formal systems, including algebraic theories [20], logics [30] (e.g., Lindenbaum–Tarski algebras) and type theories [18,17]. A term model for DIBI could lead to a sound and complete axiomatisation of the specific version of DIBI logic in this paper, whose atomic propositions take the form of $S \triangleright [T]$.

6 Conditional independence

DIBI logic is designed for reasoning about CI. The prior work [5] shows that, conditional independence in the discrete probabilistic models and join dependency in the relational models can be characterised by the same class of DIBI formulas. Generalising this result, in this section we define a notion of CI on $\mathbb{C}[\theta]$-kernels based on formula satisfaction. Since $\mathbb{C}[\theta]$ is a Markov category, we can compare our logical notion of CI with existing categorical definitions of CI in Markov categories [9,13].

Fix a Markov category $\mathbb{C}$ and a map $\theta\colon \mathrm{Var} \to \mathbf{ob}(\mathbb{C})$. We define CI in the DIBI model $\langle \mathbf{Fr}(\mathbb{C}[\theta]), \mathcal{V}_{\mathrm{nat}} \rangle$.

Definition 11 (Conditional Independence). *For any mutually disjoint finite sets of variables W, X, Y, U, X and Y are* DIBI conditionally independent given *W in a $\mathbb{C}[\theta]$-kernel*[6] *$f\colon \varnothing \to W \cup X \cup Y \cup U$ if*

$$f \vDash_{\mathcal{V}_{\mathrm{nat}}} (\varnothing \rhd [W]) \mathbin{⨟} ((W \rhd [X]) * (W \rhd [Y])). \tag{2}$$

In this case, we write $X \perp\!\!\!\perp_{\mathrm{L}} Y \mid W$.

Let us unfold what (2) means. Under the natural valuation $\mathcal{V}_{\mathrm{nat}}$, an atomic proposition of shape $S \rhd [T]$ encodes the dependence of T on S: formally, a $\mathbb{C}[\theta]$-kernel $f\colon X \to Y$ satisfies $S \rhd [T]$ iff f contains some subkernel $f'\colon S \to Y'$ such that $T \subseteq Y'$. So the formula (2) requires that f is a kernel with empty domain that can be decomposed as $f \sqsupseteq f_0 \odot (f_1 \oplus f_2)$, where f_0 determines the value on W, f_1 determines the value on X given the value on W, f_2 determines the value on Y given the value on W, and f_1 and f_2 do so independently of each other.

We illustrate the formula with examples in the discrete probabilistic DIBI model and the stochastic DIBI model.

Example 6. In the setting of Example 1, consider the probabilistic kernel $h\colon \mathbf{M}[\varnothing] \to \mathcal{D}\mathbf{M}[\{x,y,z\}]$ representing the following distribution:

$$\begin{aligned}\mu = &\frac{1}{8}|0_x,0_y,0_z\rangle + \frac{1}{8}|0_x,1_y,0_z\rangle + \frac{1}{8}|1_y,0_y,0_z\rangle + \frac{1}{8}|1_y,1_y,0_z\rangle \\ &+ \frac{1}{32}|0_x,0_y,1_z\rangle + \frac{3}{32}|0_x,1_y,1_z\rangle + \frac{3}{32}|1_y,0_y,1_z\rangle + \frac{9}{32}|1_y,1_y,1_z\rangle\end{aligned}$$

Then $h \vDash_{\mathcal{V}_{\mathrm{nat}}} (\varnothing \rhd [\{z\}]) ⨟ ((\{z\} \rhd [\{z,x\}]) * (\{z\} \rhd [\{z,y\}]))$, because $h = h_0 \odot f = h_0 \odot (f_1 \oplus f_2)$, where h_0 denotes the uniform distribution $\frac{1}{2}|0_z\rangle + \frac{1}{2}|1_z\rangle$.

Example 7. Define g_1, g_2, f, h as in Example 5. We want to assert that variables x and y are independent in the distribution constructed by Box-Muller Transform. Independence is a special case of conditional independence in which the set of conditioned variables is empty. Thus, the goal is to assert $(\varnothing \rhd [\varnothing]) ⨟ ((\varnothing \rhd [\{x\}]) * (\varnothing \rhd [\{y\}]))$ – equivalently, $(\varnothing \rhd [\{x\}]) * (\varnothing \rhd [\{y\}])$.

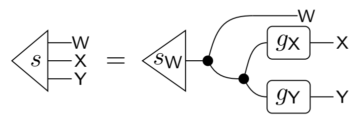
Define $h_1 : \varnothing \to \{x\}$ and $h_2 : \varnothing \to \{y\}$ both as the standard normal distribution $\mathcal{N}(0,1)$. Clearly $h_1 \vDash_{\mathcal{V}_{\mathrm{nat}}} \varnothing \rhd [\{x\}]$ and $h_2 \vDash_{\mathcal{V}_{\mathrm{nat}}} \varnothing \rhd [\{y\}]$. Moreover, some non-trivial calculations would show that $(h_1 \oplus h_2) \sqsubseteq h$, and consequently we have $h \vDash_{\mathcal{V}_{\mathrm{nat}}} (\varnothing \rhd [\{x\}]) * (\varnothing \rhd [\{y\}])$ by definition.

Since the categorical DIBI models are based on Markov categories, we compare our logical notion of CI on kernels with the canonical notion of CI in Markov

[6] Note that $\mathbb{C}[\theta]$-kernels with domain $\varnothing$ are not to be thought of as maps with empty domains. For instance, $\mathcal{K}\ell(\mathcal{D})[\theta]$-kernels of the form $\varnothing \to \{x,y\}$ corresponds to $\mathcal{K}\ell(\mathcal{D})$-morphisms $\mathbf{1} \rightharpoonup \theta(x) \times \theta(y)$, which denote distributions over x, y.

categories, which defines CI as decomposability of morphisms. In Definitions 12 to 14, a Markov category $\mathbb{X}$ is fixed.

Definition 12. *An $\mathbb{X}$-morphism $s\colon \mathsf{I} \to \mathsf{W} \otimes \mathsf{X} \otimes \mathsf{Y}$ displays the conditional independence of X and Y given W if there exist $\mathbb{X}$-morphisms $s_{\mathsf{W}}\colon \mathsf{I} \to \mathsf{W}$, $g_{\mathsf{X}}\colon \mathsf{W} \to \mathsf{X}$, $g_{\mathsf{Y}}\colon \mathsf{W} \to \mathsf{Y}$ such that equation on the right holds. We write this as $\mathsf{X} \perp \mathsf{Y} \mid \mathsf{W}$.*



In the context of DIBI models, Definition 12 restricts to stating the conditional independence of X and Y given W in $\mathbb{C}[\theta]$-kernels of the form $\varnothing \to W \cup X \cup Y$. In particular, no extra variable in the kernel's codomain is allowed.

Example 8. We show an example of this notion of CI in the Markov category $\mathbb{G}\mathrm{auss}$. Consider morphism $s : \varnothing \to \{w, x, y\}$ specified by the tuple

$$\left(!, \sigma^2 = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix}, \mu = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}\right), \text{where ! denotes the trivial map from empty domain.}$$

That is, s takes a length 0 vector and generates a length 3 vector, holding the values of w, x and y, with the normal distribution $\mathcal{N}(\mu, \sigma^2)$. This s can be decomposed as in Definition 12 with $s_w = (!, 0, 1)$, $g_x = (1, 0, 1)$, and $g_y = (1, 0, 1)$: composing s_2, g_x and g_y as in Definition 12, we get $\mathbb{E}(w) = \mathbb{E}(\xi_w) = 0$, $\mathbb{E}(x) = \mathbb{E}(w + \xi_x) = 0 + 0 = 0$, and similarly, $\mathbb{E}(y) = \mathbb{E}(w + \xi_y) = 0$, justifying the noise's mean μ being a zero vector. For the covariance matrix, let $v = (w, x, y) - (\mathbb{E}(w), \mathbb{E}(x), \mathbb{E}(y))$. Then $\sigma^2 = \mathbb{E}(v \cdot v^T) = \mathbb{E}((w, x, y) \cdot (w, x, y)^T)$, and one may show that σ^2 is equal to the matrix above.

Proposition 8. *For any $\mathbb{C}[\theta]$-kernel $s\colon \varnothing \to W \cup X \cup Y$ where W, X, Y are mutually disjoint, $X \perp Y \mid W$ iff $X \perp\!\!\!\perp_{\mathrm{L}} Y \mid W$.*

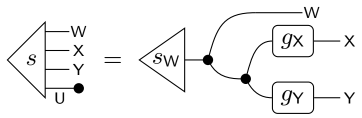
In order to extend Proposition 8 to the scenario in Definition 11 where a kernel f might contain some U in its codomain that does not appear in the CI statement, we need to modify the notion of CI from Definition 12 – referred to as plain CI – to allow objects that do not appear in the CI statement to occur in the codomain of s. We suggest two sensible extensions.

Definition 13. *Given an $\mathbb{X}$-morphism $s\colon \mathsf{I} \to \mathsf{W} \otimes \mathsf{X} \otimes \mathsf{Y} \otimes \mathsf{U}$,*

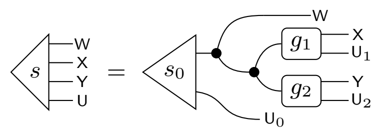
s **displays Markov CI** *if there exist $\mathbb{X}$-morphisms $s_{\mathsf{W}}, g_{\mathsf{X}}, g_{\mathsf{Y}}$ such that 4a holds. We write this as $\mathsf{X} \perp_{\mathrm{M}} \mathsf{Y} \mid \mathsf{W}$.*

s **displays superset CI** *if there exist $\mathbb{X}$-morphisms s_0, g_1, g_2 such that 4b holds. We write this as $\mathsf{X} \perp_{\mathrm{S}} \mathsf{Y} \mid \mathsf{W}$.*

These two notions differ regarding to the treatment of the extra object U. In Figure 4a, we project out the extra object U and reduce the situation to that of Definition 12. In Figure 4b, U is kept and passed along through s_0, g_1, g_2. Clearly, both reduce to Definition 12 when no such U appears. We can now state that DIBI CI coincides with Markov CI, but is weaker than superset CI.



(a) Markov CI.



(b) superset CI.

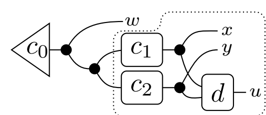
Fig. 4: Two possible extension of plain CI.

Theorem 2. *Given the* $\mathbb{C}[\theta]$*-kernel* $f\colon \varnothing \to W \cup X \cup Y \cup U$ *from Definition 11,*

1. f *satisfies* $X \perp_{\mathrm{M}} Y \mid W$ *if and only if it satisfies* $X \perp\!\!\!\perp_{\mathrm{L}} Y \mid W$*;*
2. *if* f *satisfies* $X \perp_{\mathrm{S}} Y \mid Z$*, then it also satisfies* $X \perp\!\!\!\perp_{\mathrm{L}} Y \mid Z$*.*

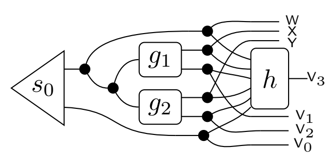
The proof of Item 1 is in Appendix E. Item 2 follows from Item 1 and that $\mathsf{X} \perp_{\mathrm{S}} \mathsf{Y} \mid \mathsf{W}$ implies $\mathsf{X} \perp_{\mathrm{M}} \mathsf{Y} \mid \mathsf{W}$: one can simply apply $-\!\!\bullet_{\mathsf{U}}$ on both sides of 4b and obtain Figure 4a via naturality if $-\!\!\bullet$. The converse of Item 2 does not hold in general, as demonstrated by the following example.

Example 9. Consider the syntactic DIBI model $\langle \mathbf{SynFr}, \mathcal{V}_{\mathrm{nat}} \rangle$ from Section 5.3. Define the $\mathbb{SynVar}[id]$-kernel f as on the right-hand side, where c_0, c_1, c_2, d are all generating morphisms, i.e., not further decomposable. Then f satisfies the DIBI CI $x \perp\!\!\!\perp_{\mathrm{L}} y \mid w$, but not the superset CI $x \perp_{\mathrm{S}} y \mid w$: one cannot rewrite the diagram in the dotted box into a juxtaposition of two diagrams with output wires containing x and y, respectively; in other words, it cannot be rewritten as the style in Figure 4b.

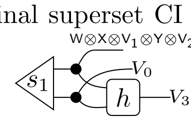


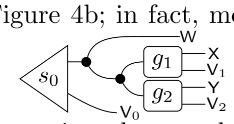
Example 9 gives some hint at how to weaken the superset CI to match DIBI CI: there one needs to allow some morphism d following the morphism which witnesses $x \perp_{\mathrm{S}} y \mid z$. We formalise this idea and show the resulting notion is indeed equivalent to both Markov and DIBI CI.

Definition 14. *An* $\mathbb{X}$*-morphism* $s\colon \mathsf{I} \to \mathsf{W} \otimes \mathsf{X} \otimes \mathsf{Y} \otimes \mathsf{U}$ *displays the extended superset conditional independence – denoted as* $\mathsf{X} \perp_{\mathrm{S+}} \mathsf{Y} \mid \mathsf{W}$ *– if there exist* $\mathbb{X}$*-morphisms* s_0, g_1, g_2, h *such that* s *can be decomposed as on the right-hand side.*



(3)

Compared with Figure 4b, here one allows an extra morphism h to appear after the original superset CI diagrams in Figure 4b; in fact, modulo rewiring, (3) is exactly 

, where $s_1 =$ 

. One intuitive way to think of the extended superset CI is to view the morphisms as certain computational processes [24]: X and Y are independent given W in s if s could be obtained via a computation in which X and Y are computed independently from W (using g_1 and g_2 in (3) respectively), after which some further computation may apply (for which stands the h part in (3)).

Proposition 9. *In Markov categories with conditionals, extended superset CI and Markov CI are equivalent. Therefore, in the context of Theorem 2, suppose further that $\mathbb{C}$ has conditionals, then the three notions of CI – DIBI CI, Markov CI, and extended superset CI – coincide.*

7 Conclusion

In this paper we provided a general recipe to construct models for DIBI logic, generalising the previously studied probabilistic and relational models. We adopted string diagrams to best visualise the 'input-preserving' property that characterises the states in the models, as well as the compositions and subkernel relations, whose definition would be quite convoluted in non-diagrammatic syntax. Then we derived various new classes DIBI models of interest. Also, we abstractly define a notion of conditional independence in terms of DIBI formulas. As our approach is based on Markov categories, we were then able to compare it with other definitions of CI proposed in the literature.

There are many promising directions for future work. On the logic side, DIBI logic – interpreted in the probabilistic models – was designed to be the assertion logic of Conditional Probabilistic Separation Logic (CPSL). Our categorical construction of a wide class of DIBI models suggests a generalisation of CPSL to obtain program logics in various scenarios beyond probabilistic programs, in the spirit of Moggi [22].

The notion of CI we propose can be straightforwardly generalised from Markov categories to copy-delete categories (see Section 2). This would allow us to encompass models such as relations with bag semantics in databases [8,16], subprobability measures [19]. However, to the best of the authors' knowledge, Proposition 8 fails for generic CD categories. Hence, finding appropriate notions of CI in this more general setting remains an open question.

From a categorical perspective, the definition of the category $\mathbb{C}[\theta]$ deserves further exploration, from at least two angles. First, the $\mathbb{C}[\theta]$-morphisms may be seen as a "bundle" of the images of some syntactic categories of variables and renaming (similar to $\mathbb{SynVar}$ from Section 5.3) under suitable functors – usually referred to as 'models' in functorial semantics. We would like to makes the connection with functorial semantics rigorous in terms of the categorical structures involved. [20,7]. Second, while the current work represents finite sets of variables with finite $\preceq$-ordered lists without duplicate, towards a more principled treatment, it is worth exploring the adoption of nominal string diagrams, a diagrammatic calculus for variables and renaming [3,4,2].

References

1. Aho, A.V., Beeri, C., Ullman, J.D.: The theory of joins in relational databases. ACM Transactions on Database Systems (TODS) **4**(3), 297–314 (1979). https://doi.org/10.1145/320083.320091

2. Balco, S.: Display calculi and nominal string diagrams. Ph.D. thesis, University of Leicester (2020)
3. Balco, S., Kurz, A.: Partially monoidal categories and the algebra of simultaneous substitutions (2018), available at https://gdlyrttnap.pl/resources/papers/syco1.pdf
4. Balco, S., Kurz, A.: Nominal String Diagrams. In: Roggenbach, M., Sokolova, A. (eds.) 8th Conference on Algebra and Coalgebra in Computer Science (CALCO 2019). Leibniz International Proceedings in Informatics (LIPIcs), vol. 139, pp. 18:1–18:20. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany (2019). https://doi.org/10.4230/LIPIcs.CALCO.2019.18
5. Bao, J., Docherty, S., Hsu, J., Silva, A.: A bunched logic for conditional independence. In: 2021 36th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS). pp. 1–14. IEEE (2021). https://doi.org/10.1109/LICS52264.2021.9470712
6. Barthe, G., Hsu, J., Liao, K.: A probabilistic separation logic. Proceedings of the ACM on Programming Languages **4**(POPL), 1–30 (2019). https://doi.org/10.1145/3371123
7. Bonchi, F., Sobociński, P., Zanasi, F.: Deconstructing lawvere with distributive laws. Journal of logical and algebraic methods in programming **95**, 128–146 (2018). https://doi.org/10.1016/j.jlamp.2017.12.002
8. Chaudhuri, S., Vardi, M.Y.: Optimization of real conjunctive queries. In: Proceedings of the twelfth ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems. pp. 59–70 (1993). https://doi.org/10.1145/153850.153856
9. Cho, K., Jacobs, B.: Disintegration and Bayesian inversion via string diagrams. Mathematical Structures in Computer Science **29**(7), 938–971 (2019). https://doi.org/10.1017/S0960129518000488
10. Dawid, A.P.: Conditional independence in statistical theory. Journal of the Royal Statistical Society: Series B (Methodological) **41**(1), 1–15 (1979). https://doi.org/10.1111/j.2517-6161.1979.tb01052.x
11. Docherty, S.: Bunched logics: a uniform approach. Ph.D. thesis, UCL (University College London) (2019)
12. Fong, B., Spivak, D.I.: An invitation to applied category theory: seven sketches in compositionality. Cambridge University Press (2019). https://doi.org/10.1017/9781108668804
13. Fritz, T.: A synthetic approach to markov kernels, conditional independence and theorems on sufficient statistics. Advances in Mathematics **370**, 107239 (Aug 2020). https://doi.org/10.1016/j.aim.2020.107239
14. Glymour, C., Zhang, K., Spirtes, P.: Review of causal discovery methods based on graphical models. Frontiers in genetics **10**, 524 (2019). https://doi.org/10.3389/fgene.2019.00524
15. Golubtsov, P.V.: Monoidal kleisli category as a background for information transformers theory. Inf. Process. **2**(1), 62–84 (2002)
16. Green, T.J.: Bag Semantics, pp. 201–206. Springer US, Boston, MA (2009). https://doi.org/10.1007/978-0-387-39940-9_979
17. Hofmann, M.: On the interpretation of type theory in locally cartesian closed categories. In: Pacholski, L., Tiuryn, J. (eds.) Computer Science Logic. pp. 427–441. Springer Berlin Heidelberg, Berlin, Heidelberg (1995). https://doi.org/10.1007/BFb0022273
18. Jacobs, B.: Semantics of weakening and contraction. Annals of pure and applied logic **69**(1), 73–106 (1994). https://doi.org/10.1016/0168-0072(94)90020-5

19. Klenke, A.: Probability theory: a comprehensive course. Springer Science & Business Media (2013). https://doi.org/10.1007/978-1-84800-048-3
20. Lawvere, F.W.: Functorial semantics of algebraic theories. Proceedings of the National Academy of Sciences **50**(5), 869–872 (1963). https://doi.org/10.1073/pnas.50.5.869
21. Li, J.M., Ahmed, A., Holtzen, S.: Lilac: a modal separation logic for conditional probability. Proceedings of the ACM on Programming Languages **7**(PLDI), 148–171 (2023). https://doi.org/10.1145/3591226
22. Moggi, E.: Notions of computation and monads. Information and computation **93**(1), 55–92 (1991). https://doi.org/10.1016/0890-5401(91)90052-4
23. O'Hearn, P.W., Pym, D.J.: The logic of bunched implications. Bulletin of Symbolic Logic **5**(2), 215–244 (1999). https://doi.org/10.2307/421090
24. Pavlovic, D.: Programs as diagrams: From categorical computability to computable categories. Springer Nature (2023). https://doi.org/10.1007/978-3-031-34827-3
25. Pearl, J.: Causality. Cambridge university press (2009). https://doi.org/10.1017/CBO9780511803161
26. Piedeleu, R., Zanasi, F.: An introduction to string diagrams for computer scientists. CoRR **abs/2305.08768** (2023)
27. Pym, D.J.: The semantics and proof theory of the logic of bunched implications, vol. 26. Springer Science & Business Media (2013). https://doi.org/10.1007/978-94-017-0091-7
28. Reynolds, J.C.: Separation logic: A logic for shared mutable data structures. In: Proceedings 17th Annual IEEE Symposium on Logic in Computer Science. pp. 55–74. IEEE (2002), https://dl.acm.org/doi/10.5555/645683.664578
29. Selinger, P.: A survey of graphical languages for monoidal categories. In: New structures for physics, pp. 289–355. Springer (2010). https://doi.org/10.1007/978-3-642-12821-9_4
30. Tarski, A.: Logic, semantics, metamathematics: papers from 1923 to 1938. Hackett Publishing (1983). https://doi.org/10.2307/2216869

A Background on Monads

We first recall the basic definition of monads. We refer to [13, Sect. 3] for an overview of the material in this section. An endofunctor $\mathcal{T} \colon \mathbb{C} \to \mathbb{C}$ is a *monad* if it has a unit $\eta^{\mathcal{T}} : 1_{\mathbb{C}} \to \mathcal{T}$ and a multiplication $\mu^{\mathcal{T}} : \mathcal{T}^2 \to \mathcal{T}$ natural transformations satisfying certain compatibility conditions. Every monad $\mathcal{T} \colon \mathbb{C} \to \mathbb{C}$ induces a Kleisli category $\mathcal{K}\ell(\mathcal{T})$, whose objects are exactly $\mathbb{C}$-objects, and morphisms $\mathsf{X} \to \mathsf{Y}$ are $\mathbb{C}$ morphisms of type $\mathsf{X} \to \mathcal{T}\mathsf{Y}$, with compositions of $f \colon \mathsf{X} \to \mathcal{T}\mathsf{Y}$ and $g \colon \mathsf{Y} \to \mathcal{T}\mathsf{Z}$ given by $\mathsf{X} \xrightarrow{f} \mathcal{T}\mathsf{Y} \xrightarrow{\mathcal{T}g} \mathcal{T}\mathcal{T}\mathsf{Z} \xrightarrow{\mu^{\mathcal{T}}_{\mathsf{Z}}} \mathcal{T}\mathsf{Z}$. We will write the morphisms in $\mathcal{K}\ell(\mathcal{T})$ as $\mathsf{X} \rightarrowtriangle \mathsf{Y}$ to distinguish them from their counterpart $\mathsf{X} \to \mathcal{T}\mathsf{Y}$ in $\mathbb{C}$. Importantly, if $\mathbb{C}$ is a SMC and $\mathcal{T}$ is a commutative monad, then $\mathcal{K}\ell(\mathcal{T})$ is also an SMC [18]. If $\mathcal{T}$ is affine symmetric monoidal, then $\mathcal{K}\ell(\mathcal{T})$ is a Markov category [15,9].

In the remainder of this section, we recall the monads used in this paper: the distribution monad $\mathcal{D}$, the powerset monad $\mathcal{P}$, the Giry monad $\mathcal{G}$, and the Radon monad $\mathcal{R}$.

Definition 15 (Discrete Distribution Monad). *The discrete distribution monad $\mathcal{D}$ is an endofunctor on* Set*. It maps a countable set X to the set of distributions over X, i.e., the set containing all functions μ over X is satisfying $\sum_{x \in X} \mu(x) = 1$, and maps a function $f : X \to Y$ to $\mathcal{D}(f) : \mathcal{D}(X) \to \mathcal{D}(Y)$, given by $\mathcal{D}(f)(\mu)(y) := \sum_{f(x)=y} \mu(x)$.*

For the monadic structure, define the unit η by $\eta_X(x) := \delta_x$, where δ_x denotes the Dirac distribution on x: for any $y \in X$, we have $\delta_x(y) = 1$ if $y = x$, otherwise $\delta_x(y) = 0$. Further, define $\mathsf{bind} \colon \mathcal{D}(X) \to (X \to \mathcal{D}(Y)) \to \mathcal{D}(Y)$ by $\mathsf{bind}(\mu)(f)(y) := \sum_{p \in \mathcal{D}(Y)} \mathcal{D}(f)(\mu)(p) \cdot p(y)$.

Definition 16 (Powerset monad). *The powerset monad $\mathcal{P}$ is an endofunctor on* Set*. It maps every set to the set of its subsets $\mathcal{P}(X) = \{U \mid U \subseteq X\}$. We define $\eta_X \colon X \to \mathcal{P}(X)$ mapping each $x \in X$ to the singleton $\{x\}$, and $\mathsf{bind} \colon \mathcal{P}(X) \to (X \to \mathcal{P}(Y)) \to \mathcal{P}(Y)$ by $\mathsf{bind}(U)(f) := \cup\{y \mid \exists x \in U. f(x) = y\}$.*

The next monad is defined on the category $\mathbb{M}\mathrm{eas}$ of measurable spaces, which consists of measurable spaces (A, Σ_A) as objects, and measurable functions as morphisms. $\mathbb{M}\mathrm{eas}$ is a monoidal category, where the monoidal product on objects and morphisms are given by the product of measurable spaces and measurable functions, respectively. In particular, the monoidal unit consists of the singleton measurable space $(\mathbf{1} = \{\bullet\}, \{\varnothing, \mathbf{1}\})$.

Definition 17 (Giry Monad). *The* giry monad *$\mathcal{G}$ maps a measurable space (X, Σ_X) to another measurable space $(\mathcal{G}(X), \Sigma_{\mathcal{G}(X)})$, where $\mathcal{G}(X)$ is the set of probability measures over X, and the σ-algebra $\Sigma_{\mathcal{G}(X)}$ is the coarsest σ-algebra over $\mathcal{G}(X)$ making the evaluation function $ev_A : \mathcal{G}(X) \to [0, 1]$, defined by $ev_A(\nu) = \nu(A)$, measurable for any $A \in \Sigma_X$. For each measurable function $f : X \mapsto Y$, $\mathcal{G}f : \mathcal{G}X \to \mathcal{G}Y$ is defined by $(\mathcal{G}f)(\nu)(B) = \nu(f^{-1}(B))$ for $B \in \Sigma_Y$. For the monadic structure, define the unit η by $\eta_X(x) = \delta_x$; define the bind operator $\mathsf{bind}_{X,Y} : \mathcal{G}X \to ((X \to \mathcal{G}Y) \to \mathcal{G}Y)$ by $\mathsf{bind}(\nu)(f)(B) = \int_X f(X)(B) d\nu$ for $B \in \Sigma_{\mathcal{G}Y}$.*

Definition 18 (Radon Monad). *The Radon monad $\mathcal{R}$ is a measure monad on the category of compact Hausdorff spaces. If maps a compact Hausdorff space X to the set of Radom measures μ on X such that $\mu(X) \leq 1$. It maps a continuous map between compact Hausdorff spaces $f : X \to Y$ to the push-forward measure $\mathcal{R}(f) : \mathcal{R}X \to \mathcal{R}Y$ given by $\mathcal{D}(f)(\mu)(y) := \mu(f^{-1}(y))$.*

For the monadic structure: we define the unit η to take a point $x \in X$ to the diract distribution δ_x solely supported at x. We also define the bind operator $\mathsf{bind}_{X,Y} : \mathcal{R}X \to ((X \to \mathcal{R}Y) \to \mathcal{R}Y)$ by $\mathsf{bind}(\nu)(f)(B) = \int_X f(X)(B) d\nu$.

The category of stochastic processses $\mathbb{Stoch}$ is the Kleisli category of the Giry monad $\mathcal{G}$. It is helpful to explicate its structure.

Definition 19. *The symmetric monoidal category of stochastic processses $\mathbb{Stoch}$ has the following components:*

- *objects are measurable spaces (A, Σ_A);*
- *morphisms $(A, \Sigma_A) \to (B, \Sigma_B)$ are maps $f \colon \Sigma_B \times A \to [0, 1]$ satisfying: for arbitrary $T \in \Sigma_B$, $f(T, -) \colon A \to [0, 1]$ is measurable, and for arbitrary $a \in A$, $f(-, a) \colon \Sigma_B \to [0, 1]$ is a probability measure;*
- *compositions of $f \colon (A, \Sigma_A) \to (B, \Sigma_B)$ and $g \colon (B, \Sigma_B) \to (C, \Sigma_C)$ is the map $g \circ f \colon \Sigma_C \times A \to [0, 1]$ mapping (U, a) to $\int_{b \in B} g(U, b) \cdot f(db, a)$;*
- *the identity morphism id on (A, Σ_A) maps $(S, a) \in \Sigma_A \times A$ to 1 if $a \in S$, and to 0 if $a \notin S$;*
- *the monoidal product $\otimes$ acts on objects as $(A, \Sigma_A) \otimes (B, \Sigma_B) = (A \times B, \Sigma_A \otimes \Sigma_B)$, where $\Sigma_A \otimes \Sigma_B$ is the smallest sigma-algebra containing $\Sigma_A \times \Sigma_B$);*
- *the monoidal product $\otimes$ acts on morphisms to obtain product measures. That is, $(U, V) \in \Sigma_B \times \Sigma_D$ as follows: given $f \colon (A, \Sigma_A) \to (B, \Sigma_B)$ and $g \colon (C, \Sigma_C) \to (D, \Sigma_D)$, $f \otimes g \colon \Sigma_B \otimes \Sigma_D \times A \times C \to [0, 1]$ maps (U, V, a, c) to $f(U, a) \cdot g(V, b)$.*

B Omitted Proofs from Section 4

This section contains the missing proof of statements in Section 4, as well as some useful properties of the $\mathbb{C}[\theta]$-kernels and the DIBI model. We stay with the setting in Section 4 for $\mathbb{C}$, Var, and θ.

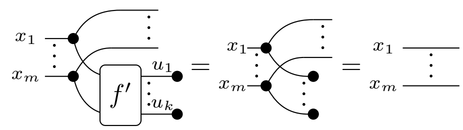
Proposition 10. *If $\mathbb{C}$ is a Markov category, then $\mathbb{C}[\theta]$ is also Markovian. If Markov cateogry $\mathbb{C}$ has conditionals, then so does $\mathbb{C}[\theta]$.*

Proof. Both follow immediately from the construction of $\mathbb{C}[\theta]$: note that $\mathbb{C}[\theta]$-morphisms are $\mathbb{C}$-morphisms. □

The following observation says that, in $\mathbb{C}[\theta]$-kernels, if one forgets the new variables in the output, then one get exactly identity on the input variables.

Proposition 11. *For an arbitrary $\mathbb{C}[\theta]$-kernel $f \colon X \to Y$, $(id_X \otimes \mathsf{del}_{Y \setminus X}) \circ f = id_X$.*

Proof. This is an immediate consequence of the Markovian property:

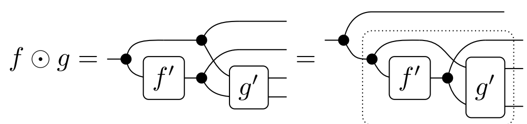


where we assume $X = \{x_1, \ldots, x_m\}$, $Y \setminus X = \{u_1, \ldots, u_k\}$. ☐

The class of $\mathbb{C}[\theta]$-kernels is closed under both parallel and sequential compositions.

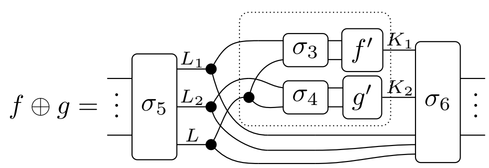
Proposition 12. *For arbitrary* $\mathbb{C}[\theta]$*-kernels* $f\colon X \to Y$ *and* $g\colon U \to V$*, whenever* $f \star g$ *is defined, the result* $f \star g$ *is also an* $\mathbb{C}[\theta]$*-kernel, for* $\star \in \{\oplus, \odot\}$*.*

Proof. Suppose $f \odot g$ is defined, then spelling out the definition, $Y = U$,



therefore $f \odot g$ is an input-preserving kernel.

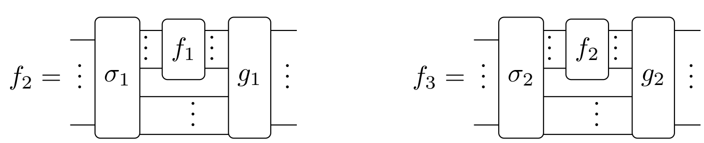
Suppose $f \oplus g$ is defined. That is, $X \cap U = Y \cap V$. Follow the notation in Definition 7, $f \oplus g$ is also an input-preserving kernel:



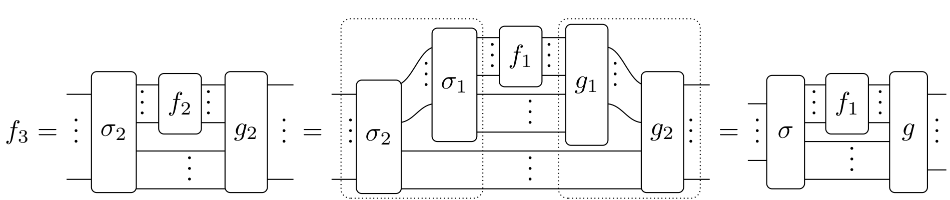
The morphisms inside the dotted square play the role of the nontrivial part of the input-preserving kernel. ☐

Proposition 13. *The subkernel relation* $\sqsubseteq$ *on* $\mathbb{C}[\theta]$*-kernels (Definition 8) is a preorder.*

Proof. In other words, we prove $\sqsubseteq$ is reflexive and transitive. The diagram in Definition 8 is trivial when spelled out for witnessing $f \sqsubseteq f$. Suppose $f_1 \sqsubseteq f_2$ and $f_2 \sqsubseteq f_3$, then they are witnessed by:



Then $f_1 \sqsubseteq f_3$ is witnessed by the following reasoning:

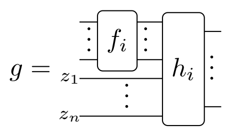


where rewiring σ and $\mathbb{C}[\theta]$-kernel g are obtained by the morphisms in the two dotted areas, respectively. $\square$

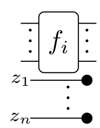
We are now ready to provide a proof of Proposition 1. We first restate the proposition.

Proposition 1. *If $\mathbb{C}$ further satisfies that for arbitrary morphisms f, g and object D, $f \otimes \mathsf{del}_\mathsf{D} = g \otimes \mathsf{del}_\mathsf{D}$ implies $f = g$, then subkernel is unique given its type in the following sense: if $\mathbb{C}[\theta]$-kernels $f_1, f_2 \colon U \to V$ are both subkernels of g, then $f_1 = f_2$.*

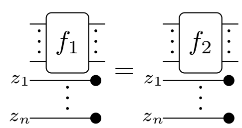
Proof (Proof of Proposition 1). Suppose $f_i \sqsubseteq g$ is witnessed by the following diagrams, where $i = 1, 2$:



where we don't worry about the rewiring morphisms, and h_1, h_2 are some $\mathbb{C}[\theta]$-kernels. Discarding the z_is via $\mathsf{del} z_i$, we obtain



Therefore,



By the assumption on $\mathbb{C}$, this entails $f_1 = f_2$. $\square$

Next, we prove the central result Theorem 1. We first present two lemmas that are useful in verifying the DIBI frame conditions.

Lemma 1. *For any $\mathbb{C}[\theta]$-kernel $f \colon X \to Y$, and any $U \subseteq_f \mathrm{Var}$ such that $U \cap X = U \cap Y = \varnothing$, by viewing id_U as a (trivial) $\mathbb{C}[\theta]$-kernel, $f \oplus id_U$ is defined, and $f \oplus id_U = \sigma_2 \circ (f \otimes id_U) \circ \sigma_1$, for some rewiring σ_1 and σ_2.*

Proof. That $f \oplus id_U$ is defined follows immediately from the assumption that $X \cap U = Y \cap U (= \varnothing)$. Then $f \oplus id_U = \sigma_2 \circ (f \otimes id_U) \circ \sigma_1$ follows immediately by the definition of parallel composition. $\square$

Lemma 2. *Two $\mathbb{C}[\theta]$-kernels f and g satisfy $f \sqsubseteq g$ if and only if there exist a set of variables U and another $\mathbb{C}[\theta]$-kernel h such that $g = (f \oplus id_U) \odot h$.*

Proof. This follows immediately from Lemma 1, by noticing that taking the parallel composition with id_U – when viewed as a trivial kernel – is nothing but the monoidal product with id_U, modulo some rewiring. $\square$

Now we are ready to check all the DIBI frame conditions for the syntactic DIBI frame whose states are input-preserving diagrams in the freely generated Markov category of string diagrams. We first restate Theorem 1.

Theorem 1. $\mathbf{Fr}(\mathbb{C}[\theta]) = \langle \mathrm{Ker}(\mathbb{C}[\theta]), \sqsubseteq, \oplus, \odot, \mathrm{Ker}(\mathbb{C}[\theta])\rangle$ *is a DIBI frame.*

Proof (Proof of Theorem 1). We verify all the frame conditions in Figure 2 for the $\mathbb{C}[\theta]$-kernels. We omit the references to Lemma 1 and Lemma 2 when using them.

1. ($\oplus$-Com). Note that the definition of $\oplus$ over $\mathbb{C}[\theta]$-kernels does not rely on the specific order of the two kernels, thus commutativity holds; in other words, commutativity of $\oplus$ holds by definition.
2. ($\oplus$-UnitExist). Given an arbitrary kernel f, show there exists a kernel e such that $f \oplus e = f$ holds. In fact, let $e = id_{[\,]}$ (where $[\,]$ is the empty list), which is trivially input-preserving, hence is a $\mathbb{C}[\theta]$-kernel. $f \oplus e$ is defined: $dom(f) \cap dom(id_{[\,]}) = \varnothing = cod(f) \cap cod(id_{[\,]})$. Moreover, $f \oplus id_{[\,]} = f$.
3. ($\oplus$-Assoc). Given $\mathbb{C}[\theta]$-kernels f, g, h, suppose $f \oplus g$ and $(f \oplus g) \oplus h$ are both defined. We show that $g \oplus h$ and $f \oplus (g \oplus h)$ are also defined; moreover, $f \oplus (g \oplus h) = (f \oplus g) \oplus h$. The converse holds by a similar argument, hence omitted.
   Let's first show that they are defined. By definition, to verify $g \oplus h$ is defined amounts to showing $dom(g) \cap dom(h) = cod(g) \cap cod(h)$. Note that, by $f \oplus g$ is defined, we have $cod(g) = cod(f \oplus g) \setminus (cod(f) \setminus (cod(f) \cap cod(g)))$. Therefore we can calculate $cod(g) \cap cod(h)$ as:

$$\begin{aligned}
cod(g) \cap cod(h) &= (cod(f \oplus g) \setminus (cod(f) \setminus (cod(f) \cap cod(g)))) \cap cod(h) \\
&= (cod(f \oplus g) \cap cod(h)) \setminus (cod(f) \setminus (cod(f) \cap cod(g))) \\
&= (dom(f \oplus g) \cap dom(h)) \setminus (cod(f) \setminus (cod(f) \cap cod(g))) \\
&= (dom(f \oplus g) \setminus (cod(f) \setminus (cod(f) \cap cod(g)))) \cap dom(h) \\
&= (dom(f \oplus g) \setminus (cod(f) \setminus (dom(f) \cap dom(g)))) \cap dom(h) \\
&= ((dom(f \oplus g) \setminus cod(f)) \cup (dom(f \oplus g) \cap (dom(f) \cap dom(g)))) \cap dom(h) \\
&= ((dom(f \oplus g) \setminus dom(f)) \cup (dom(f) \cap dom(g))) \cap dom(h) \\
&= (((dom(f) \cup dom(g)) \setminus dom(f)) \cup (dom(f) \cap dom(g))) \cap dom(h) \\
&= ((dom(g) \setminus dom(f)) \cup (dom(f) \cap dom(g))) \cap dom(h) \\
&= dom(g) \cap dom(h)
\end{aligned}$$

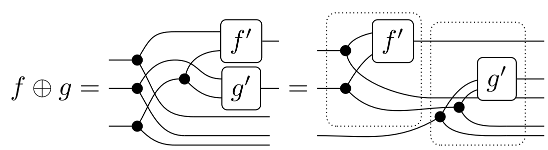
   To show that $f \oplus (g \oplus h)$ is defined, we first note that a similar argument as that above shows $dom(f) \cap dom(h) = cod(f) \cap cod(h)$. Then, $dom(f) \cap dom(g \oplus h) = cod(f) \cap cod(g \oplus h)$ follows immediately:

$$\begin{aligned}
cod(f) \cap cod(g \oplus h) &= cod(f) \cap (cod(g) \cup cod(h)) \\
&= (cod(f) \cap cod(g)) \cup (cod(f) \cap cod(h)) \\
&= (dom(f) \cap dom(g)) \cup (dom(f) \cap dom(h))
\end{aligned}$$

$$= dom(f) \cap (dom(g) \cup dom(h))$$
$$= dom(f) \cap dom(g \oplus h)$$

The equivalence of $(f \oplus g) \oplus h$ and $f \oplus (g \oplus h)$ follows immediately from their diagrammatic presentation.

4. ($\odot$-UNITEXIST$_\mathrm{L}$). We show that for arbitrary $\mathbb{C}[\theta]$-kernel f, there exists a $\mathbb{C}[\theta]$-kernel e_L such that $e_L \odot f = f$. Suppose $dom(f) = \bar{x}$, then simply define $e_L := id_{\bar{x}}$, which is also a $\mathbb{C}[\theta]$-kernel kernel; moreover, it satisfies $e_L \odot f = f \circ id_{\bar{x}} = f$.
5. ($\odot$-UNITEXIST$_\mathrm{R}$). Similar to the ($\odot$-UNITEXIST$_\mathrm{L}$) case. Suppose $cod(f) = \bar{y}$, then let $e_R = id_{\bar{y}}$. e_R satisfies $f \odot e_R = id_{\bar{y}} \circ f = f$.
6. ($\odot$-ASSOC). Note that the sequential operator $\odot$ on $\mathbb{C}[\theta]$-kernels is exactly the sequential composition in the category $\mathbb{C}[\theta]$, which is associative by definition.
7. ($\oplus$-UNITCOH). For arbitrary $\mathbb{C}[\theta]$-kernels f and g such that $f \oplus g$ is defined, we show $f \oplus g \sqsupseteq f$. Following the assumption in Definition 7, $f \oplus g$ is of the following form, where we omit the rewiring for simplicity:

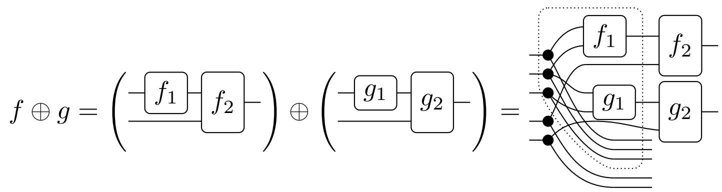


The two parts in dotted boxes are f and $id_{cod(f)} \otimes g$ (which is also a $\mathbb{C}[\theta]$-kernel), respectively. This witnesses $(f \oplus g) \sqsupseteq f$.
8. ($\odot$-UNITCOH$_\mathrm{R}$). We show that for arbitrary $\mathbb{C}[\theta]$-kernels f, g, if $f \odot g$ is defined, then $f \odot g \sqsupseteq f$. Given the assumption, $f \odot g = (f \otimes id_{[\,]}) \odot g$, which witnesses that $f \odot g \sqsupseteq f$.
9. (UNITCLOSURE). Since the set E is the set of all syntactic input-preserving kernels in the current DIBI frame, this condition is trivially satisfied.
10. ($\oplus$-DOWNCLOSED). Suppose for two $\mathbb{C}[\theta]$-kernels of the form $f \colon X \to Y$ and $g \colon U \to V$, $f \oplus g$ is defined, and there are two subkernels $f_1 \sqsubseteq f$, $g_1 \sqsubseteq g$, where $f_1 \colon X_1 \to Y_1$ and $g_1 \colon U_1 \to V_1$. The goal is to that $f_1 \oplus g_1$ is defined, and is a subkernel of $f \oplus g$. To see $f_1 \oplus g_1$ is defined, since $cod(f_1) \cap dom(f) = dom(f_1)$ and $cod(g_1) \cap dom(g) = dom(g_1)$, we have:

$$\begin{aligned} dom(f_1) \cap dom(g_1) &= (cod(f_1) \cap dom(f)) \cap (cod(g_1) \cap dom(g)) \\ &= (cod(f_1) \cap cod(g_1)) \cap (dom(f) \cap dom(g)) \\ &= (cod(f_1) \cap cod(g_1)) \cap (cod(f) \cap cod(g)) \\ &= (cod(f_1) \cap cod(f)) \cap (cod(g_1) \cap cod(g)) \\ &= cod(f_1) \cap cod(g_1) \end{aligned}$$

Next, we show the subkernel relation $f_1 \oplus g_1 \sqsubseteq f \oplus g$. By Lemma 2, we can assume that $f = (f_1 \oplus id_S) \odot f_2$, $g = (g_1 \oplus id_T) \odot g_2$, where f_2, g_2 are also

$\mathbb{C}[\theta]$-kernels. Then, diagrammatically we have:

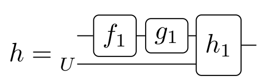


Notice that the diagrams in the dotted circle is precisely $f_1 \oplus g_1$, therefore $f_1 \oplus g_1 \sqsubseteq f \oplus g$.
We can also derive the desired property using some other frame conditions as follows:

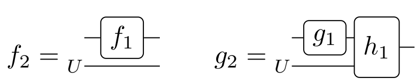
$$\begin{aligned}
f \oplus g &= ((f_1 \oplus id_S) \odot f_2) \oplus ((g_1 \oplus id_T) \odot g_2) \\
&= ((f_1 \oplus id_S) \oplus (g_1 \oplus id_T)) \odot (f_2 \oplus g_2) && (\textsc{RevExchange}) \\
&= (f_1 \oplus g_1) \oplus (id_S \oplus id_T)) \odot (f_2 \oplus g_2) && (\oplus\text{-}\textsc{Assoc}), (\oplus\text{-}\textsc{Com}) \\
&= ((f_1 \oplus g_1) \oplus (id_{(S \cup T)}) \odot (f_2 \oplus g_2) \\
&\sqsupseteq f_1 \oplus g_1 && (\text{Def. } 8)
\end{aligned}$$

Crucially, the proof below of (RevExchange), ($\oplus$-Assoc), and ($\oplus$-Com) does not rely on ($\oplus$-DownClosed).

11. ($\odot$-UpClosed). Given kernels f_1, g_1, h such that $f_1 \odot g_1$ is defined and $f_1 \odot g_1 \sqsubseteq h$, we show that there exist $f_2 \sqsupseteq f_1$, $g_2 \sqsupseteq g_1$, such that $f_2 \odot g_2$ is well-defined and is exactly h. By definition, $f_1 \odot g_1 \sqsubseteq h$ means that there exist a set of variables U and a kernel h_1 such that:



We simply define



It is obvious then that $f_2 \sqsupseteq f_1$ and $g_2 \sqsupseteq g_1$; moreover, $f_2 \odot g_2 = h$.

12. (RevExchange). For arbitrary $\mathbb{C}[\theta]$-kernels f_1, f_2, g_1, g_2, if $(f_1 \odot f_2) \oplus (g_1 \odot g_2)$ is defined, then $f_1 \oplus g_1$ and $f_2 \oplus g_2$ are defined as well, and $(f_1 \odot f_2) \oplus (g_1 \odot g_2) = (f_1 \oplus g_1) \odot (f_2 \oplus g_2)$.
We first verify that $f_1 \oplus g_1$ and $f_2 \oplus g_2$ are defined. For $f_1 \oplus g_1$, we need to show that $dom(f_1) \cap dom(g_1) = cod(f_1) \cap cod(g_1)$. Half of the equation is 'free' given the 'non-decreasing' nature of the kernels, namely $dom(f_1) \cap dom(g_1) \subseteq cod(f_1) \cap cod(g_1)$. So it suffices to prove the other inclusion. Note that $dom(f_1) \cap dom(g_1) = dom(f_1 \odot f_2) \cap dom(g_1 \odot g_2)$, so it suffices to show that $cod(f_1) \cap cod(g_1) \subseteq dom(f_1 \odot f_2) \cap dom(g_1 \odot g_2)$, as follows:

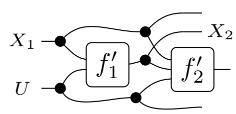
$$cod(f_1) \cap cod(g_1) = dom(f_2) \cap dom(g_2) \qquad (\text{Definition of } \odot)$$

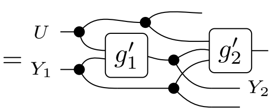
$$
\begin{aligned}
&\subseteq cod(f_2) \cap cod(g_2) && \text{(Definition of kernels)}\\
&= cod(f_1 \odot f_2) \cap cod(g_1 \odot g_2) && \text{(Definition of } \odot\text{)}\\
&= dom(f_1 \odot f_2) \cap dom(g_1 \odot g_2) && \text{(}(f_1 \odot f_2) \oplus (g_1 \odot g_2) \text{ is defined)}
\end{aligned}
$$

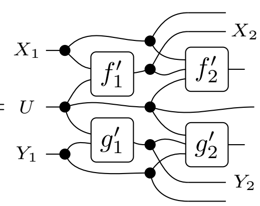
A similar argument confirms that $f_2 \oplus g_2$ is defined.
Next we show that $(f_1 \oplus g_1) \odot (f_2 \oplus g_2)$ is defined, namely $cod(f_1 \oplus g_1) = dom(f_2 \oplus g_2)$, as follows:

$$
\begin{aligned}
cod(f_1 \oplus g_1) &= cod(f_1) \cup cod(g_1)\\
&= dom(f_2) \cup dom(g_2)\\
&= dom(f_2 \oplus g_2)
\end{aligned}
$$

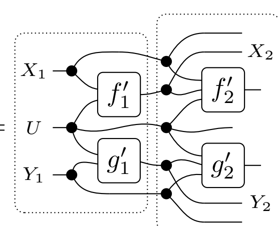
Finally, for the equivalence $(f_1 \oplus g_1) \odot (f_2 \oplus g_2) = (f_1 \odot f_2) \oplus (g_1 \odot g_2)$ we draw their diagrams:

$f_1 \odot f_2 =$ 

$g_1 \odot g_2 =$ 

$(f_1 \odot f_2) \oplus (g_1 \odot g_2) =$ 

where $dom(f_1) = X_1 \cup U$, $dom(g_1) = Y_1 \cup U$, and $dom(f_1) \cap dom(g_1) = U$. This is exactly $(f_1 \oplus g_1) \odot (f_2 \oplus g_2)$, as shown in the diagram below, where the two diagrams inside the dotted circles are $f_1 \oplus g_1$ and $f_2 \oplus g_2$, respectively:

$(f_1 \oplus g_1) \odot (f_2 \oplus g_2) =$ 

This completes the verification that $\langle \mathrm{Ker}(\mathbb{C}[\theta]), \sqsubseteq, \oplus, \odot, \mathrm{Ker}(\mathbb{C}[\theta]) \rangle$ satisfies all DIBI frame conditions, and therefore it is a DIBI frame. $\square$

C Omitted Proofs from Section 5

We recall Proposition 2, which states the isomorphism between the probabilistic DIBI frames and the categorical DIBI frames based on $\mathcal{K}\ell(\mathcal{D})$.

Proposition 2. $\mathbf{PrFr}[\mathrm{Val}] \cong \mathbf{Fr}(\mathcal{K}\ell(\mathcal{D})[\mathrm{Val}])$.

Proof (Proof of Proposition 2). Recall the embedding : $\mathbb{MemPr} \to \mathcal{K}\ell(\mathcal{D})[\theta]$ from Example 3. We show that the restriction of such ι to probabilistic kernels – a subclass of $\mathbb{MemPr}$-objects – which we also write ι with an abuse of notation witness the desired isomorphism of DIBI frames. First, ι establish a bijection between probabilistic kernels of the form $\mathbf{M}[X] \to \mathcal{D}\mathbf{M}[Y]$ and $\mathcal{K}\ell(\mathcal{D})[\mathrm{Val}]$-kernel of the form $[\![X]\!] \to [\![Y]\!]$ – recall that $[\![\cdot]\!]$ denotes the list representation of finite sets of variables. For simplicity we assume $X = \{x_1, \dots, x_m\}$, $Y = \{y_1, \dots, y_n\}$. Given a probabilistic kernel $f \colon \mathbf{M}[X] \to \mathcal{D}\mathbf{M}[Y]$, its image $\iota(f) \colon [x_1, \dots, x_m] \to [y_1, \dots, y_n]$ is a $\mathcal{K}\ell(\mathcal{D})$-morphism $\mathrm{Val}^m \rightarrow\!\!\!\!\!\triangleright\ \mathrm{Val}^n$ obtained by the composition:

$$\mathrm{Val}^m \xrightarrow{\cong} \mathbf{M}[X] \xrightarrow{f} \mathcal{D}\mathbf{M}[Y] \xrightarrow{\mathcal{D}\cong} \mathcal{D}\mathrm{Val}^n$$

where $\cong$ is the isomorphism $\mathbf{M}[Y] \xrightarrow{\cong} \mathbf{M}[y_1] \times \cdots \times \mathbf{M}[y_n] \xrightarrow{\cong} \mathrm{Val}^n$. It satisfies Definition 6 immediately by the input-preserving conditions (see Definition 3) of the probabilistic kernel f.

Before verifying that ι respects the structure on DIBI frames, for convenience we introduce the notion of *combination* of memories: two memories $\mathbf{m} \in \mathbf{M}[X]$ and $\mathbf{n} \in \mathbf{M}[Y]$ are combinable if $X \cap Y = \varnothing$; in this case, their combination $\mathbf{m} \uplus \mathbf{n}$ is a memory in $\mathbf{M}[X \cup Y]$, such that

$$\mathbf{m} \uplus \mathbf{n} \colon (u \in X \cup Y) \mapsto \begin{cases} \mathbf{m}(u) & u \in X \\ \mathbf{n}(u) & v \in Y \end{cases}$$

Now we show ι respects both compositions. The sequential composition in probabilistic kernels is obviously the Kleisli composition, as already observed in [5]. So we focus on the parallel composition. Consider the $\mathcal{K}\ell(\mathcal{D})[\mathrm{Val}]$-kernels f, g from Definition 7. Their counterpart probabilistic kernels $\iota^{-1}(f) \colon \mathbf{M}[X] \to \mathcal{D}\mathbf{M}[Y]$ and $\iota^{-1}(g) \colon \mathbf{M}[U] \to \mathcal{D}\mathbf{M}[V]$ does the following:

$$\iota^{-1}(f) \colon (\mathbf{m} \in \mathbf{M}[X]) \mapsto \sum_{\mathbf{n} \in \mathbf{M}[Y \setminus X]} \iota^{-1}(f')(\mathbf{m})(\mathbf{n}) |\mathbf{m} \uplus \mathbf{n}\rangle$$
$$\iota^{-1}(g) \colon (\mathbf{m} \in \mathbf{M}[U]) \mapsto \sum_{\mathbf{n} \in \mathbf{M}[V \setminus U]} \iota^{-1}(g')(\mathbf{m})(\mathbf{n}) |\mathbf{m} \uplus \mathbf{n}\rangle$$

where $f' \colon X \to (Y \setminus X)$ and $g' \colon U \to (V \setminus U)$ are the nontrivial parts of the kernels f and g, respectively (see Definition 6). The parallel composition of these two probabilistic kernels is as follows, according to Definition 3. Given $\mathbf{m} \in \mathbf{M}[X \cup U]$, $\mathbf{n} \in \mathbf{M}[Y \cup V]$,

$$\begin{aligned} \iota^{-1}(f) \oplus \iota^{-1}(g)(\mathbf{m})(\mathbf{n}) &= \iota^{-1}(f)(\mathbf{m}^X)(\mathbf{n}^Y) \cdot \iota^{-1}(g)(\mathbf{m}^U)(\mathbf{n}^V) \\ &= \iota^{-1}(f')(\mathbf{m}^X)(\mathbf{n}^{Y \setminus X}) \cdot \iota^{-1}(g')(\mathbf{m}^U)(\mathbf{n}^{V \setminus U}) \end{aligned}$$

The probabilistic kernel counterpart of $f \oplus g$ is:

$$\begin{aligned} &\iota^{-1}(f \oplus g) \colon (\mathbf{m} \in \mathbf{M}[X \cup U]) \\ &\qquad \mapsto \sum_{\boldsymbol{\ell}_1 \in \mathbf{M}[Y \setminus X], \boldsymbol{\ell}_2 \in \mathbf{M}[V \setminus U]} \iota^{-1}(f')(\mathbf{m}^X)(\boldsymbol{\ell}_1) \cdot \iota^{-1}(g')(\mathbf{m}^U)(\boldsymbol{\ell}_2) |\mathbf{m} \uplus \boldsymbol{\ell}_1 \uplus \boldsymbol{\ell}_2\rangle \end{aligned}$$

That is, for arbitrary $\mathbf{m} \in \mathbf{M}[X \cup U]$ and $\mathbf{n} \in \mathbf{M}[Y \cup V]$,

$$\iota^{-1}(f \oplus g)(\mathbf{m})(\mathbf{n}) = \iota^{-1}(f')(\mathbf{m}^X)(\mathbf{n}^{Y \setminus X}) \cdot \iota^{-1}(g')(\mathbf{m}^U)(\mathbf{n}^{V \setminus U})$$

Therefore $\iota^{-1}(f) \oplus \iota^{-1}(g) = \iota^{-1}(f \oplus g)$.

Finally, as the subkernel relation is defined in terms of the sequential and parallel compositions in the same way for both the probabilistic kernels and $\mathcal{K}\ell(\mathcal{D})[\text{Val}]$ kernels, ι also respects the subkernel relation. Therefore we can conclude that $\mathbf{PrFr}[\text{Val}]$ and $\mathbf{Fr}(\mathcal{K}\ell(\mathcal{D})[\text{Val}])$ are isomorphic as DIBI frames. □

D Partially Monoidal Categories

In this section we prove that in the DIBI frames constructed via our recipe, the parallel composition forms a partially monoidal structure. We formalise the result using the notion of partially monoidal categories [3].

A *partial monoid* $\langle A, e, \diamond, D \rangle$ consists of a domain $D \subseteq A \times A$, a binary operation $\diamond \colon D \to A$ whose (partial) unit is $e \in D$, such that the monoidal equations hold whenever defined: $(a \diamond b) \diamond c \doteq a \diamond (b \diamond c)$ and $e \diamond a = a = a \diamond e$, where $\doteq$ stands for 'equal when either side is defined'.

Definition 20 ([3]). *A* (strict) partially monoidal category (PMC) *consists of:*

- *a small category $\mathbb{C}$ with sets of objects $\mathbf{ob}(\mathbb{C})$ and sets of morphisms $\mathbf{mor}(\mathbb{C})$;*
- *partial monoids $\langle \mathbf{ob}(\mathbb{C}), \mathsf{E}_0, \otimes_0, D_0 \rangle$ and $\langle \mathbf{mor}(\mathbb{C}), \mathsf{E}_1, \otimes_1, D_1 \rangle$, such that $\mathbb{D}$ with objects from D_0 and morphisms from D_1 is a subcategory of $\mathbb{C} \times \mathbb{C}$;*
- *the operator $\otimes$ defined as $\otimes_0$ on objects and $\otimes_1$ on morphisms forms a functor $\mathbb{D} \to \mathbb{C}$.*

Intuitively, a PMC is a category with compatible partial monoid structures on both the sets of objects and morphisms.

We familiarise the notion of PMC with the probabilistic DIBI models.

Proposition 14. *$\langle \mathbb{PrKern}, \oplus, id_{\mathbf{M}[\varnothing]} \rangle$ is a partially monoidal category, where $\mathbb{PrKern}$ is the category of probabilistic kernels viewed as a subcategory of $\mathcal{K}\ell(\mathcal{D})$.*

Proof. Let us define the partial monoids on the set of objects and morphisms, respectively.

On objects, the partial monoid $\langle \mathbf{ob}(\mathbb{PrKern}), \mathsf{E}_0, \otimes_0, D_0 \rangle$ is indeed total: $D_0 = \mathbf{ob}(\mathbb{PrKern}) \times \mathbf{ob}(\mathbb{PrKern})$; given $\mathbf{M}[X]$ and $\mathbf{M}[Y]$, $\mathbf{M}[X] \otimes_0 \mathbf{M}[Y] = \mathbf{M}[X \cup Y]$; $\mathsf{E}_0 = \mathbf{M}[\varnothing]$.

On morphisms, define $D_1 \subseteq \mathbf{mor}(\mathbb{PrKern}) \times \mathbf{mor}(\mathbb{PrKern})$ to consist of precisely those probabilistic kernels that are parallelly composable; that is, a pair of morphisms $(f \colon \mathbf{M}[X] \to \mathcal{D}\mathbf{M}[Y], g \colon \mathbf{M}[U] \to \mathcal{D}\mathbf{M}[V]) \in D_1$ iff $X \cap U = Y \cap V$. In this case, $f \otimes_1 g$ is $f \oplus g$. The unit E_1 is $id_\varnothing$.

It remains to verify that $\langle D_0, D_1 \rangle$ forms a subcategory of $\mathbb{PrKern} \times \mathbb{PrKern}$ – denoted $\mathbb{D}$, and $\oplus$ is a functor $\mathbb{D} \to \mathbb{PrKern}$.

For the former, note that if $f_1 \oplus g_1$ and $f_2 \oplus g_2$ are both defined, $f_1 \odot f_2$ and $g_1 \odot g_2$ are both defined, then $(f_1 \odot f_2) \oplus (g_1 \odot g_2)$ is also defined by

(RevExchange). Also, the identity morphisms are present in D_1. Therefore $\langle D_0, D_1\rangle$ forms a subcategory $\mathbb{D}$ of $\mathbb{PrKern}$.

For the latter, the functoriality spells out as: given arbitrary $(f_1, g_1), (f_2, g_2) \in \mathbf{ob}(\mathbb{D}) = D_1 \times D_1$ that are sequentially composable, $(g_1 \circ f_1) \oplus (g_2 \circ f_2) = (g_1 \oplus g_2) \circ (f_1 \oplus f_2)$. This is guaranteed by (RevExchange).

Therefore $\langle \mathbb{PrKern}, \oplus, id_{\mathbf{M}[\varnothing]}\rangle$ is a PMC. □

Proposition 15. $\langle \mathbb{Ker}(\mathbb{C}[\theta]), \oplus, id_{[\,]}\rangle$ *is a partially monoidal category.*

Proof. The proof is similar to that of Proposition 14, which is a concrete case of the current proposition.

We define the partial monoids on the set of objects and morphisms as follows.

For objects, its partial monoid structure $\langle D_0, \otimes_0, \mathsf{E}_0\rangle$ is total: D_0 is simply $\mathbf{ob}(\mathbb{Ker}(\mathbb{C}[\theta])) \times \mathbf{ob}(\mathbb{Ker}(\mathbb{C}[\theta]))$, namely pairs of list representation of finite sets of variables; given $(L, K) \in D_0$, where $L = [\![X]\!]$ and $K = [\![Y]\!]$, $L \otimes_0 K = [\![X \cup Y]\!]$; $\mathsf{E}_0 = [\,]$.

For morphisms, its partial monoid structure $\langle D_1, \otimes_1, \mathsf{E}_1\rangle$ is: $D_1 = \{(f, g) \in \mathbf{ob}(\mathbb{Ker}(\mathbb{C}[\theta])) \times \mathbf{ob}(\mathbb{Ker}(\mathbb{C}[\theta])) \mid f \oplus g \text{ is defined}\}$; given $(f, g) \in D_1$, $f \otimes_1 g = f \oplus g$; $\mathsf{E}_2 = id_{[\,]}$.

With a bit abuse of notation, we shall simply denote both $\otimes_0$ and $\otimes_1$ simply by their corresponding operation $\oplus$.

Next, that $\langle D_0, D_1\rangle$ forms a subcategory $\mathbb{D}$ of $\mathbb{Ker}(\mathbb{C}[\theta]) \times \mathbb{Ker}(\mathbb{C}[\theta])$ and $\oplus\colon \mathbb{D} \to \mathbb{Ker}(\mathbb{C}[\theta])$ is a functor follows from the frame conditions – in particular (RevExchange). □

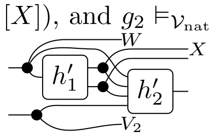
E Omitted Proofs from Section 6

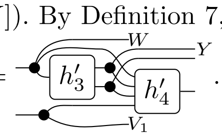
We fix a Markov category $\mathbb{C}$ and a choice function $\theta\colon \mathrm{Var} \to \mathbf{ob}(\mathbb{C})$.

Theorem 2. *Given the* $\mathbb{C}[\theta]$*-kernel* $f\colon \varnothing \to W \cup X \cup Y \cup U$ *from Definition 11,*

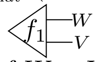
1. f *satisfies* $X \perp_{\mathrm{M}} Y \mid W$ *if and only if it satisfies* $X \perp\!\!\!\perp_{\mathrm{L}} Y \mid W$*;*
2. *if* f *satisfies* $X \perp_{\mathrm{S}} Y \mid Z$*, then it also satisfies* $X \perp\!\!\!\perp_{\mathrm{L}} Y \mid Z$*.*

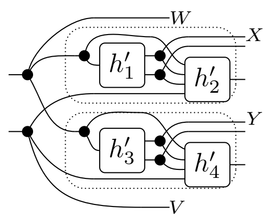
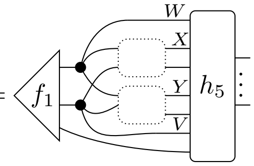
Proof. We prove the first point only. The proof follows immediately by spelling out the definition of $f \vDash_{\mathcal{V}_{\mathrm{nat}}} (\varnothing \triangleright [W]) \fatsemi ((W \triangleright [X]) * (W \triangleright [Y]))$. According to Definition 2, it means there exist $\mathbb{C}[\theta]$-kernels f_1, f_2 such that $f = f_1 \odot f_2$, $f_1 \vDash_{\mathcal{V}_{\mathrm{nat}}} (\varnothing \triangleright [W])$, and $f_2 \vDash_{\mathcal{V}_{\mathrm{nat}}} ((W \triangleright [X]) * (W \triangleright [Y]))$.

$f_2 \vDash_{\mathcal{V}_{\mathrm{nat}}} ((W \triangleright [X]) * (W \triangleright [Y]))$ means there exist $\mathbb{C}[\theta]$-kernels g_1, g_2 such that $f_2 \sqsupseteq g_1 \oplus g_2$, $g_1 \vDash_{\mathcal{V}_{\mathrm{nat}}} (W \triangleright [X])$, and $g_2 \vDash_{\mathcal{V}_{\mathrm{nat}}} (W \triangleright [Y])$. By Definition 7, Definition 8, and $\mathcal{V}_{\mathrm{nat}}$, infer $g_1 =$ 

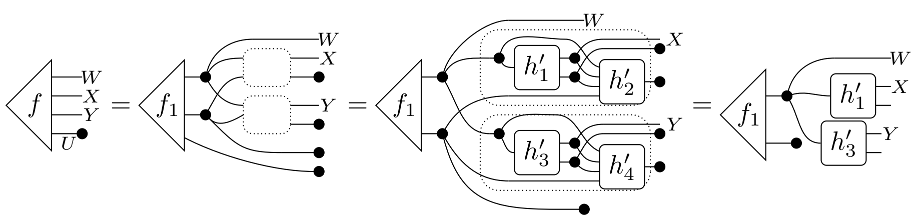
and $g_2 =$ 

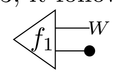
.

$f_1 \vDash_{\mathcal{V}_{\mathrm{nat}}} (\varnothing \triangleright [W])$ means, according to $\mathcal{V}_{\mathrm{nat}}$ and Definition 7, f_1 is of the the form  . Given the form of g_1, g_2, and f_2, such V – which is already a subset of $W \cup X \cup Y \cup U$ by definition – must satisfy $V \subseteq U$. So far we have:

$f_2 \sqsupseteq g_1 \oplus g_2 =$  , therefore $f =$ 

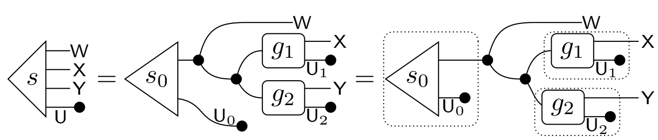
where the two morphisms in the dotted boxes (represented only by dotted boxes later for simplicity) play the role of g_1 and g_2 in Definition 13.13, respectively. Delete U in the output of f by postcomposing del_U, one gets:



Compare the resulting diagram with Definition 13.13, it follows that f displays Markov CI, namely $X \perp_{\mathrm{M}} Y \,|\, W$; in particular, here  plays the role of s_{W} in Figure 4a. □

Lemma 3. *In a Markov category $\mathbb{X}$, superset CI implies Markov CI.*

Proof. Given an $\mathbb{X}$-morphism $s\colon \mathsf{I} \to \mathsf{W} \otimes \mathsf{X} \otimes \mathsf{Y} \otimes \mathsf{U}$ satisfying $\mathsf{X} \perp_{\mathrm{S}} \mathsf{Y} \,|\, \mathsf{W}$ – i.e., it can be decomposed as in Figure 4b, we show that is also satisfies $\mathsf{X} \perp_{\mathrm{M}} \mathsf{Y} \,|\, \mathsf{W}$ – i.e., it can be expressed as in Figure 4a.



where the three diagrams in the dotted circles play the role of s_{W}, g_{X}, and g_{Y} in Figure 4a, respectively. □

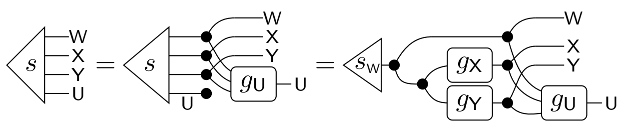
Proposition 8. *For any $\mathbb{C}[\theta]$-kernel $s\colon \varnothing \to W \cup X \cup Y$ where W, X, Y are mutually disjoint, $X \perp Y \,|\, W$ iff $X \perp\!\!\!\perp_{\mathrm{L}} Y \,|\, W$.*

Proof. Note that when $U = \varnothing$, both the superset CI statement $X \perp_{\mathrm{S}} Y \,|\, W$ and the Markov CI statement $X \perp_{\mathrm{M}} Y \,|\, W$ reduce to the plain CI statement $X \perp Y \,|\, W$. So the statement follows immediately from Theorem 2 and Lemma 3. □

Proposition 9. *In Markov categories with conditionals, extended superset CI and Markov CI are equivalent. Therefore, in the context of Theorem 2, suppose further that* $\mathbb{C}$ *has conditionals, then the three notions of CI – DIBI CI, Markov CI, and extended superset CI – coincide.*

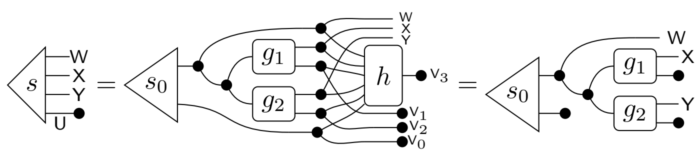
Proof. For the first part, we stick with the setting in Definition 13, and further assume that $\mathbb{X}$ has conditionals.

Suppose s satisfies $\mathsf{X} \perp_{\mathrm{M}} \mathsf{Y} \mid \mathsf{W}$, namely decomposition as in Figure 4a holds. Then, since $\mathbb{X}$ has conditionals, there exist $g_{\mathsf{U}} \colon \mathsf{W} \otimes \mathsf{X} \otimes \mathsf{Y}$ such that:



The last diagram witness the extended superset CI $\mathsf{X} \perp_{\mathrm{S+}} \mathsf{Y} \mid \mathsf{W}$.

Suppose s satisfies $\mathsf{X} \perp_{\mathrm{S+}} \mathsf{Y} \mid \mathsf{W}$; i.e., s can be decomposed as (3). Then, deleting the U part (where $\mathsf{V}_0 \otimes \mathsf{V}_1 \otimes \mathsf{V}_2 \otimes \mathsf{V}_3 = \mathsf{U}$), we get:



This precisely says that s satisfies Markov CI $\mathsf{X} \perp_{\mathrm{M}} \mathsf{Y} \mid \mathsf{W}$.

For the second part of the statement, note that $\mathbb{C}$ has conditionals implies that $\mathbb{C}[\theta]$ also has conditionals (Proposition 10). Then it follows immediately from the first half of the current statement together with Theorem 2. □