
Causal Statistical Decision Problems

Anonymous Author(s)

Affiliation

Address

email

1 Notes on category theoretic probability and string diagrams

Category theoretic treatments of probability theory often start with *probability monads* (for a good overview, see [Jacobs, 2018]). A monad on some category C is a functor $T : C \rightarrow C$ along with natural transformations called the unit $\eta : 1_C \rightarrow T$ and multiplication $\mu : T^2 \rightarrow T$. Roughly, functors are maps between categories that preserve identity and composition structure and natural transformations are "maps" between functors that also preserve composition structure. The monad unit is similar to the identity element of a monoid in that application of the identity followed by multiplication yields the identity transformation. The multiplication transformation is also (roughly speaking) associative.

An example of a probability monad is the discrete probability monad given by the functor $\mathcal{D} : \mathbf{Set} \rightarrow \mathbf{Set}$ which maps a countable set X to the set of functions from $X \rightarrow [0, 1]$ that are probability measures on X , denoted $\mathcal{D}(X)$. \mathcal{D} maps a measurable function f to $\mathcal{D}f : X \rightarrow \mathcal{D}(X)$ given by $\mathcal{D}f : x \mapsto \delta_{f(x)}$. The unit of this monad is the map $\eta_X : X \rightarrow \mathcal{D}(X)$ given by $\eta_X : x \mapsto \delta_x$ (which is equivalent to $\mathcal{D}1_X$) and multiplication is $\mu_X : \mathcal{D}^2(X) \rightarrow \mathcal{D}(X)$ where $\mu_X : \Omega \mapsto \sum_{\phi} \Omega(\phi)\phi$.

For continuous distributions we have the Giry monad on the category \mathbf{Meas} of measurable spaces given by the functor \mathcal{G} which maps a measurable space X to the set of probability measures on X , denoted $\mathcal{G}(X)$. Other elements of the monad (unit, multiplication and map between morphisms) are the "continuous" version of the above.

Of particular interest is the Kleisli category of the monads above. The Kleisli C_T category of a monad T on category C is the category with the same objects and the morphisms $X \rightarrow Y$ in C_T is the set of morphisms $X \rightarrow TY$ in C . Thus the morphisms $X \rightarrow Y$ in the Kleisli category $\mathbf{Set}_{\mathcal{D}}$ are morphisms $X \rightarrow \mathcal{D}(Y)$ in \mathbf{Set} , i.e. stochastic matrices, and in the Kleisli category $\mathbf{Meas}_{\mathcal{G}}$ we have Markov kernels. Composition of arrows in the Kleisli categories correspond to Matrix products and "kernel products" respectively.

Both \mathcal{D} and \mathcal{G} are known to be *commutative* monads, and the Kleisli category of a commutative monad is a symmetric monoidal category.

Diagrams for symmetric monoidal categories consist of wires with arrows, boxes and a couple of special symbols. The identity object (which we identify with the set $\{*\}$) is drawn as nothing at all $\{*\} := \square$ and identity maps are drawn as bare wires:

$$\text{Id}_X := \begin{array}{c} \uparrow \\ X \end{array} \quad (1)$$

We draw Kleisli arrows from the unit (i.e. probability distributions) $\mu : \{*\} \rightarrow X$ as triangles and Kleisli arrows $\kappa : X \rightarrow Y$ (i.e. Markov kernels $X \rightarrow \Delta(\mathcal{Y})$) as boxes. We draw the Kleisli arrow

32 $\mathbb{1}_X : X \rightarrow \{*\}$ (which is unique for each X) as below

$$\mu := \begin{array}{c} \uparrow X \\ \nabla \mu \end{array} \quad \kappa := \begin{array}{c} \uparrow Y \\ \boxed{\kappa} \end{array} \quad (2)$$

33 The product of objects in **Meas** is given by $(X, \mathcal{X}) \cdot (Y, \mathcal{Y}) = (X \times Y, \mathcal{X} \otimes \mathcal{Y})$, which we will
 34 often write as just $X \times Y$. Horizontal juxtaposition of wires indicates this product, and horizontal
 35 juxtaposition also indicates the tensor product of Kleisli arrows. Let $\kappa_1 : X \rightarrow W$ and $\kappa_2 : Y \rightarrow Z$:

$$(X \times Y, \mathcal{X} \otimes \mathcal{Y}) := \begin{array}{c} \uparrow X \quad \uparrow Y \\ \hline \end{array} \quad \kappa_1 \otimes \kappa_2 := \begin{array}{c} \uparrow W \quad \uparrow Z \\ \boxed{\kappa_1} \quad \boxed{\kappa_2} \\ \hline X \quad Y \end{array} \quad (3)$$

36 Composition of arrows is achieved by “wiring” boxes together. For $\kappa_1 : X \rightarrow Y$ and $\kappa_2 : Y \rightarrow Z$
 37 we have

$$\kappa_1 \kappa_2(x; A) = \int_Y \kappa_2(y; A) \kappa_1(x; dy) := \begin{array}{c} \uparrow Z \\ \boxed{\kappa_2} \\ \hline \boxed{\kappa_1} \\ \hline X \end{array} \quad (4)$$

38 Symmetric monoidal categoris have the following coherence theorem[Selinger, 2010]:

39 **Theorem 1.1** (Coherence (symmetric monoidal)). *A well-formed equation between morphisms in*
 40 *the language of symmetric monoidal categories follows from the axioms of symmetric monoidal*
 41 *categories if and only if it holds, up to isomorphism of diagrams, in the graphical language.*

42 Isomorphism of diagrams for symmetric monoidal categories (somewhat informally) is any planar
 43 deformation of a diagram including deformations that cause wires to cross. We consider a diagram
 44 for a symmetric monoidal category to be well formed only if all wires point upwards.

45 In fact the Kleisli categories of the probability monads above have (for each object) unique *copy*:
 46 $X \rightarrow X \times X$ and *erase*: $X \rightarrow \{*\}$ maps that satisfy the *commutative comonoid axioms* that (thanks
 47 to the coherence theorem above) can be stated graphically. These differ from the copy and erase maps
 48 of *finite product* or *cartesian* categories in that they do not necessarily respect composition of arrows.

$$\text{Erase} = \mathbb{1}_X := \begin{array}{c} * \\ \uparrow \end{array} \quad \text{Copy} = x \mapsto \delta_{x,x} := \begin{array}{c} \swarrow \quad \searrow \\ \uparrow \end{array} \quad (5)$$

$$\begin{array}{c} \swarrow \quad \searrow \\ \uparrow \end{array} = \begin{array}{c} \swarrow \quad \searrow \\ \uparrow \end{array} := \begin{array}{c} \swarrow \quad \searrow \\ \uparrow \end{array} \quad (6)$$

$$\begin{array}{c} * \\ \swarrow \quad \searrow \\ \uparrow \end{array} = \begin{array}{c} \swarrow \quad \searrow \\ \uparrow \end{array} = \begin{array}{c} \uparrow \end{array} \quad (7)$$

$$\begin{array}{c} \swarrow \quad \searrow \\ \uparrow \end{array} = \begin{array}{c} \swarrow \quad \searrow \\ \uparrow \end{array} \quad (8)$$

Finally, $\{*\}$ is a terminal object in the Kleisli categories of either probability monad. This means that the map $X \rightarrow \{*\}$ is unique for all objects X , and as a consequence for all objects X, Y and all $\kappa : X \rightarrow Y$ we have

$$\begin{array}{c} * \\ \boxed{\kappa} \\ \downarrow \\ X \end{array} = \begin{array}{c} * \\ \downarrow \\ X \end{array} \quad (9)$$

This is equivalent to requiring for all $x \in X$ $\int_Y \kappa(x; dy) = 1$. In the case of $\mathbf{Set}_{\mathcal{D}}$, this condition is what differentiates a stochastic matrix from a general positive matrix (which live in a larger category than $\mathbf{Set}_{\mathcal{D}}$).

Thus when manipulating diagrams representing Markov kernels in particular (and, importantly, not more general symmetric monoidal categories) diagram isomorphism also includes applications of 6, 7, 8 and 9.

A particular property of the copy map in $\mathbf{Meas}_{\mathcal{G}}$ (and probably $\mathbf{Set}_{\mathcal{D}}$ as well) is that it commutes with Markov kernels iff the markov kernels are deterministic [Fong, 2013].

1.1 Disintegration and Bayesian inversion

Disintegration is a key operation on probability distributions (equivalently arrows $\{*\} \rightarrow X$) in the categories under discussion. It corresponds to “finding the conditional probability” (though conditional probability is usually formalised in a slightly different way).

Given a distribution $\mu : \{*\} \rightarrow X \otimes Y$, a disintegration $c : X \rightarrow Y$ is a Markov kernel that satisfies

$$\begin{array}{c} X \quad Y \\ \downarrow \quad \downarrow \\ \mu \end{array} = \begin{array}{c} X \quad Y \\ \downarrow \quad \downarrow \\ \boxed{c} \\ \downarrow \\ * \\ \downarrow \\ \mu \end{array} \quad (10)$$

Disintegrations always exist in $\mathbf{Set}_{\mathcal{D}}$ but not in $\mathbf{Meas}_{\mathcal{G}}$. They do exist in the latter if we restrict ourselves to standard measurable spaces. If c_1 and c_2 are disintegrations $X \rightarrow Y$ of μ , they are equal μ -A.S. In fact, this equality can be strengthened somewhat - they are equal almost surely with respect to any distribution that shares the “ X -marginal” of μ .

Given $\sigma : \{*\} \rightarrow X$ and a channel $c : X \rightarrow Y$, a Bayesian inversion of (σ, c) is a channel $d : Y \rightarrow X$ such that

$$\begin{array}{c} X \quad Y \\ \downarrow \quad \downarrow \\ \boxed{c} \\ \downarrow \\ \sigma \end{array} = \begin{array}{c} X \quad Y \\ \downarrow \quad \downarrow \\ \boxed{d} \\ \downarrow \\ \boxed{c} \\ \downarrow \\ \sigma \end{array} \quad (11)$$

We can obtain disintegrations from Bayesian inversions and vice-versa.

Clerc et al. [2017] offer an alternative view of Bayesian inversion which they claim doesn’t depend on standard measurability conditions, but there is a step in their proof I didn’t follow.

1.2 Generalisations

Cho and Jacobs [2019] make use of a larger “CD” category by dropping 9. I’m not completely clear whether you end up with arrows being “Markov kernels for general measures” or something else (can we have negative arrows?). This allows for the introduction of “observables” or “effects” of the form

$$\begin{array}{c} \triangle \\ \downarrow \\ f \end{array}$$

79 Jacobs et al. [2019] make use of an embedding of \mathbf{Set}_D in $\mathbf{Mat}(\mathbb{R}^+)$ with morphisms all positive
80 matrices (I’m not totally clear on the objects, or how they are self-dual - this doesn’t seem to be
81 exactly the same as the category of finite dimensional vector spaces). This latter category is compact
82 closed, which - informally speaking - supports the same diagrams as symmetric monoidal categories
83 with the addition of “upside down” wires.

84 1.3 Key questions for Causal Theories

85 **generalised disintegrations** : Of key importance to our work is generalising the notion of disinte-
86 gration (and possibly Bayesian inversion) to general kernels $X \rightarrow Y$ rather than restricting ourselves
87 to probability distributions $\{*\} \rightarrow Y$.

88 Given $\kappa : D \rightarrow X \times Y$, a kernel $c : D \times X \rightarrow Y$ is a *generalised disintegration* (“g-disintegration”)
89 of κ if the following holds:

90 **Theorem 1.2.** For all $\kappa : D \rightarrow X \times Y$, if D is countable and $X \times Y$ is standard measurable, a
91 g-disintegration of κ exists.

92 *Proof.* For all $y \in D$ we have a disintegration $c_y : X \rightarrow Y$ of $\delta_y \kappa$ by standard measurability of
93 $X \times Y$. Define $c : D \times X \rightarrow Y$ by $c : (y, x) \mapsto c_y(x)$. Clearly, $c(y, x)$ is a probability distribution
94 on Y for all $(y, x) \in D \times X$. It remains to show $c(\cdot)^{-1}(B)$ is measurable for all $B \in \mathcal{B}([0, 1])$. But
95 $c(\cdot)^{-1}(B) = \bigcap_{y \in D} c_y(\cdot)^{-1}(B)$. The right hand side is measurable by measurability of $c_y(\cdot)^{-1}(B)$
96 and the properties of a σ -algebra. \square

97 **Conjecture:** This can be generalised to any κ that is determined by its values on a countable set of
98 points along with some notion of continuity. This seems likely to be true. In a more general setting, I
99 think I could find a counterexample, but the converse also seems unlikely.

100 We can also defined *generalised conditional independences*. In fact, there are two separate definitions
101 of independence here $\perp\!\!\!\perp$ $\perp\!\!\!\perp$

102 Generalised disintegrations facilitate the following construction of a “graphical model”:

103 Suppose we have two causal theories, \mathcal{T}^* and \mathcal{T} both with signature $E \times D \rightarrow E$, and \mathcal{T} is a decision
104 randomised version of \mathcal{T}^* (i.e. $\mathcal{T} = \{(\lambda \kappa, \mu) | (\kappa, \mu) \in \mathcal{T}^*\}$ for some $\lambda : D \rightarrow D$). We will construct
105 a graphical model from \mathcal{T}^* and \mathcal{T} in three steps:

106 First, we assume *reproducibility* in the stronger theory \mathcal{T}^* . That is, for all $(\kappa, \mu) \in \mathcal{T}^*$ we suppose
107 there exists $\gamma \in \Delta(D)$ such that $\gamma \kappa = \mu$.

108
109 Second, we will assume certain *generalised conditional independences* hold for the stronger theory
110 \mathcal{T}^* (we have not defined these, but they are the obvious generalisation of standard conditional
111 independence lifted to g-disintegrations). Because we’re constructing a graphical model, we will
112 assume these are a “DAG-compatible” set, though we are under no obligation to do so. I conjecture
113 we can illustrate these independences graphically. Suppose we have random variables $X : E \rightarrow X$,
114 $Y : E \rightarrow Y$ and $Z : E \rightarrow Z$, and we assume we have at least the generalised CIs implied by the

I don’t think reproducibility is quite the right assumption, but it is good enough for now

115 following diagram for all $(\kappa, \mu) \in \mathcal{T}^*$:

$$\kappa = \begin{array}{c} \begin{array}{c} Z \\ \downarrow \\ \boxed{c_{Z|Y}} \\ \swarrow \quad \searrow \\ Y \quad \quad X \\ \swarrow \quad \searrow \\ \boxed{c_{Y|X}} \\ \swarrow \quad \searrow \\ X \quad \quad Y \\ \swarrow \quad \searrow \\ ** \\ \boxed{\kappa} \\ \downarrow \\ E \end{array} \end{array} \quad (13)$$

116 The above diagram is typed incorrectly, but we can always construct a kernel κ_{XYZ} that maps to
 117 $X \times Y \times Z$.

118 References

- 119 Kenta Cho and Bart Jacobs. Disintegration and Bayesian inversion via string diagrams.
 120 *Mathematical Structures in Computer Science*, 29(7):938–971, August 2019. ISSN
 121 0960-1295, 1469-8072. doi: 10.1017/S0960129518000488. URL [https://www.
 122 cambridge.org/core/journals/mathematical-structures-in-computer-science/
 123 article/disintegration-and-bayesian-inversion-via-string-diagrams/
 124 0581C747DB5793756FE135C70B3B6D51](https://www.cambridge.org/core/journals/mathematical-structures-in-computer-science/article/disintegration-and-bayesian-inversion-via-string-diagrams/0581C747DB5793756FE135C70B3B6D51).
- 125 Florence Clerc, Fredrik Dahlqvist, Vincent Danos, and Ilias Garnier. Pointless learn-
 126 ing. *20th International Conference on Foundations of Software Science and Compu-
 127 tation Structures (FoSSaCS 2017)*, March 2017. doi: 10.1007/978-3-662-54458-7_
 128 21. URL [https://www.research.ed.ac.uk/portal/en/publications/
 129 pointless-learning\(694fb610-69c5-469c-9793-825df4f8ddec\).html](https://www.research.ed.ac.uk/portal/en/publications/pointless-learning(694fb610-69c5-469c-9793-825df4f8ddec).html).
- 130 Brendan Fong. Causal Theories: A Categorical Perspective on Bayesian Networks. *arXiv:1301.6201*
 131 *[math]*, January 2013. URL <http://arxiv.org/abs/1301.6201>. arXiv: 1301.6201.
- 132 Bart Jacobs. From probability monads to commutative effectuses. *Journal of Logical and*
 133 *Algebraic Methods in Programming*, 94:200–237, January 2018. ISSN 2352-2208. doi:
 134 10.1016/j.jlamp.2016.11.006. URL [http://www.sciencedirect.com/science/article/
 135 pii/S2352220816301122](http://www.sciencedirect.com/science/article/pii/S2352220816301122).
- 136 Bart Jacobs, Aleks Kissinger, and Fabio Zanasi. Causal Inference by String Diagram Surgery. In
 137 Mikołaj Bojańczyk and Alex Simpson, editors, *Foundations of Software Science and Computation*
 138 *Structures*, Lecture Notes in Computer Science, pages 313–329. Springer International Publishing,
 139 2019. ISBN 978-3-030-17127-8.
- 140 Peter Selinger. A survey of graphical languages for monoidal categories. *arXiv:0908.3347 [math]*,
 141 813:289–355, 2010. doi: 10.1007/978-3-642-12821-9_4. URL [http://arxiv.org/abs/0908.
 142 3347](http://arxiv.org/abs/0908.3347). arXiv: 0908.3347.

