Algebraic Properties of Conditional Probability

Mika Bohinen

Balliol College

University of Oxford

A thesis submitted for the degree of

MSc in Mathematics and Foundations of Computer Science

Autumn 2024

Acknowledgements

It is not entirely trivial to write an acknowledgements while simultaneously keeping it anonymous, but I will do my best.

This thesis represents the culmination of my mathematical journey so far. The road has not been easy, but it has been helped a great deal by the people I have met along the way.

First and foremost I want to thank my supervisor. It is not an overstatement to say that this thesis would not exist without his generous help and guidance. His enthusiasm for category theory and maths in general is contagious.

I would also like thank the wonderful people I have met while doing the MFoCS degree. This year would not have been anywhere near as rewarding as it has been without them.

Another thank you should be given to the people behind my scholarship for their continued support in my educational and personal journey.

Lastly, but certainly not least, I would like to thank my parents and family for always supporting my mathematical endeavours.

Abstract

The theory of Markov categories is a new synthetic approach to probability theory introduced by Tobias Fritz in 2019. A particular important class of Markov categories are representable such as Stoch, the category of measurable spaces and Markov kernels. We give a self-contained introduction to this framework and prove some new results about the relationship between the existence of conditionals in representable Markov categories and the Beck-Chevalley property on monads.

Contents

List of Figures

Chapter 1

Introduction

The use of category theory has become pervasive in modern mathematics. One alluring explanation for this is that the essence of an object is captured in its relationship to other objects. More informally we have the common adage "you know a person by the company they keep". Category theory has been especially successful in fields such as topology, algebraic geometry, commutative algebra, and homological algebra, to name a few. However, it is much less used in analysis where we often care more about the internal structure of the objects we study.

Recently there has been an increasing interest in trying to apply category to analytical fields such as probability theory. The synthetic framework of Markov categories is the most recent take on categorical probability theory. It was first introduced by Fritz in [5]. Some of the precursors to that paper is work done by Golubtsov ([9], [10], [11]) as well as by Cho and Jacobs [2].

The goal of this thesis is to relate the existence of conditionals in Markov categories with properties of a certain monad that "represent" it. A secondary goal is that it can serve as self-contained introduction to Markov categories. Someone who has taken a first course in category theory should be able to gain a solid understanding of Markov categories and how they are used.

More specifically, this thesis contains the following new results:

- Proposition 4.14 explicitly constructs the equivalence between the partial evaluation relation and second-order stochastic dominance. Other source (see e.g. [7]) mentions it vaguely in passing, but we make it explicit here.

- Lemma 5.5 shows that the multiplication of a monad on an a.s.-compatibly representable Markov category with conditionals is weakly Cartesian.

- Theorem 5.6 is the main theorem of the thesis and extends Lemma 5.5 to show that the monad of an a.s.-compatibly representable Markov category with conditionals and that satisfies the equalizer principle is BC.

- Corollary 5.7 applies Theorem 5.6 to show that the Giry monad on $\mathsf{BorelStoch}$ is BC.

- Corollary 5.9 uses Theorem 5.6 to show that the second-order stochastic dominance relation on $\mathsf{C}(\Theta, A)$ is transitive and hence $\mathsf{C}(\Theta, A) \in \mathsf{Ord}$ for all $\Theta, A \in \mathsf{C}$.

Outline

The thesis is organized as follows:

- Chapter 1 is the introduction.

- In Chapter 2 we give a brief introduction to the graphical formalism of string diagrams.

- In Chapter 3 we present a mainly self-contained introduction to Markov categories and their properties.

- In Chapter 4 we introduce partial evaluations and relate them to representable Markov categories.

- In Chapter 5 we present our new findings on the relationship between the existence of conditionals and the BC property on monads.

Prerequisites

We will assume basic familiarity with category theory at the level of [18]. You should at the very least be comfortable with monads and not be frightened by the sentence "a monad is a monoid in the category of endofunctors". It is also helpful to know something about bicategories, but this is not crucial.

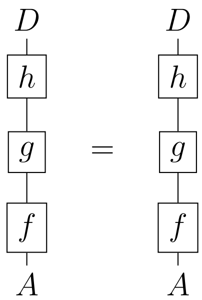
Chapter 2

A Briefer on String Diagrams for SMCs

A big benefit of graphical formalisms is that they often hide much of the complicated bookkeeping, enabling us to focus on the most essential parts. Consider for example associativity of composition in a category: given morphisms $f : A \to B$, $g : B \to C$, and $h : C \to D$, it is the statement that

$$h \circ (g \circ f) = (h \circ g) \circ f. \tag{2.1}$$

In the graphical formalism of string diagrams (which will shortly be properly introduced) this is translated into the trivial statement



(2.2)

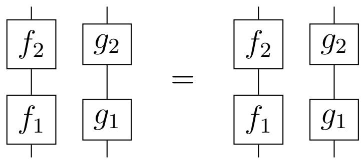
The graphical notation is justified precisely because composition is associativity. It is the same reason we can also write

$$h \circ g \circ f \tag{2.3}$$

without worrying about whether the resulting morphism is well defined. So, what is the big difference? Why not just use the notation in Equation 2.3? You certainly can, but the point is that it is convenient to have a formalism which reflects the underlying features of the things you work with. To further illustrate this, consider the interchange law as it applies to symmetric monoidal categories. It states that composing morphisms horizontally and then vertically is the same as composing them vertically and then horizontally. The equation is

$$(f_1 \circ f_2) \otimes (g_1 \circ g_2) = (f_1 \otimes g_1) \circ (f_2 \otimes g_2). \tag{2.4}$$

There is no way to do the same trick here as in the case of associativity. You are forced to pick between one of the two ways of writing it, and there is no canonical choice. On the other hand, the string diagrammatic version of Equation 2.4 is given by the following trivial looking statement



(2.5)

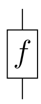
Using string diagrams the two sides of Equation 2.4 become notationally equivalent. We are no longer forced into making an arbitrary choice, the formalism does this bookkeeping internally. This bookkeeping feature has a cascading effect. Proofs which would be extremely tedious to write down with the traditional approach tend to become much more pleasant to work with as string diagrams. The author admits that this is a subjective viewpoint, but hopes that the remainder of this thesis might serve as an example of this claim.

2.1 The Formalism

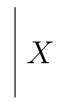
For the remainder of this section, let $(\mathsf{C}, \otimes, I)$ be a symmetric monoidal category. It is enlightening to think of C as a "theory of resources". Following Coecke et al. [3] we think of the objects $X, Y \in \mathsf{C}$ as resources and morphisms of the form $f : X \to Y$ as a way of transforming the resource X into the resource Y.

Definition 2.1. *Let $f : X \to Y$ and $g : X \otimes Y \to A \otimes B$ be morphisms in C. We draw f as follows*

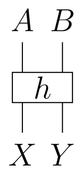
When the domain and codomain is clear from the context we draw f as



When f is the identity on X, we draw it as the line



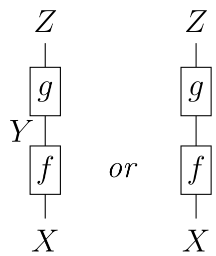
The morphism g is drawn as follows



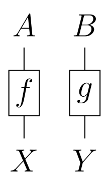
This way of drawing morphisms lends itself well to the resource theoretic understanding of symmetric monoidal categories. We can think of the whole morphisms as an assembly line that gets resources along the input wire and sends the finished product along the output wire.

Now, if we have two assembly lines, we can either connect the end of one to the beginning of the other, or put them next to each other. Of course, connecting the output and input forces the transformed resource of one machine to be the input of the other. More formally we have the following definitions.

Definition 2.2. *Let $f : X \to Y$ and $g : Y \to Z$ be two morphisms in C. We draw their composition $g \circ f : X \to Z$ as the following diagram*



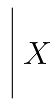
Definition 2.3. *Let $f : X \to A$ and $g : Y \to B$ be morphisms in C. We draw their* ***parallel composition*** *$f \otimes g : X \otimes Y \to A \otimes B$ as*



The first definition is akin to putting one assembly line after the other, while the second definition is akin to putting them next to each other.

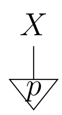
Take care to notice the difference between the two ways of drawing a morphism $X \otimes Y \to A \otimes B$ presented in the above definitions. Parallel composition is supposed to highlight the independence of the two morphisms. That is to say, the two assembly lines do not depend on each other. On the other hand, the morphism g in Definition 2.1 is akin to having an assembly line with one machine that has two inputs and two outputs.

From the resource theoretic point of view, the monoidal unit is the same as the "empty resource". The semantics of the monoidal product in $X \otimes Y$ is that you have both resource X and resource Y. Seeing as $X \otimes I = X = I \otimes X$ it makes sense that we should not draw the monoidal unit. Hence

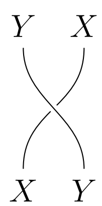


represents both X, $I \otimes X$, and $X \otimes I$. In cases where we do want to highlight the monoidal unit, such as in a discarding map $X \to I$ we draw it as a dotted line, i.e.,

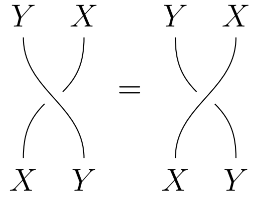
Of particular interest are maps of the form $p : I \to X$, also referred to as states. We draw these as



Finally, the braiding of the monoidal category is represented by the following diagram



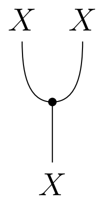
In a symmetric monoidal category, the braiding is its own inverse and hence we have the equation



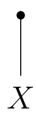
2.2 Commutative Comonoids

An important notion we will need for later is that of a commutative comonoid. Given an object X in a monoidal category $(C, \otimes, I)$, a **commutative comonoid** structure on X consists of the following things: maps $X \to X \otimes X$ and $X \to I$ called the **comultiplication** and **counit**, or **copy** and **discard**. For the purpose of Markov categories we denote them by `copy` $: X \to X \otimes X$ and `del` $: X \to I$.

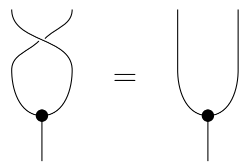
In terms of string diagrams, we represent these maps as follows:

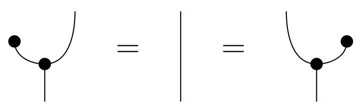
`copy` $=$  ,

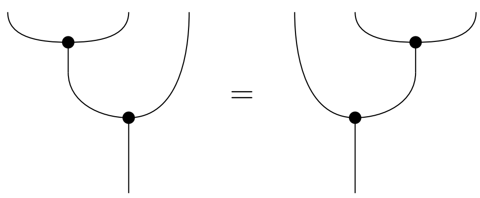
and

`del` $=$ 

These maps need to satisfy the following equations (commutativity, counitality, and coassociativity)

 (2.6)

 (2.7)

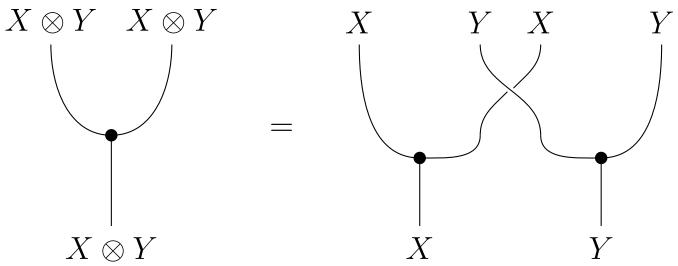
 (2.8)

2.3 Copy-Discard Categories

Definition 2.4. *A **copy-discard** category is a symmetric monoidal category where*

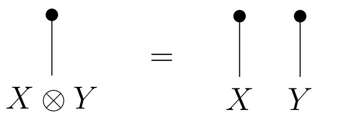
- *every object is equipped with a distinguished commutative comonoid structure,*
- *the comonoid structure is compatible with the tensor product in the following ways:*

For all objects $X, Y \in C$ we have

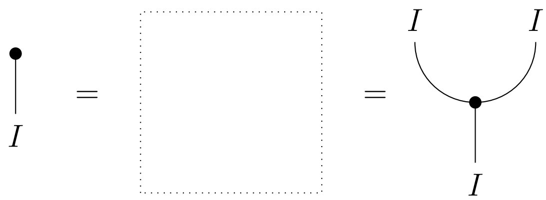


(2.9)

We also have the conditions that



(2.10)



(2.11)

Example 2.5. *Take* Set *with its usual Cartesian structure. Let the copy map be the diagonal*

$$\begin{aligned}\Delta_X : X &\to X \times X \\ x &\mapsto (x, x)\end{aligned}$$

and the discard map be the unique map

$$\begin{aligned}! : X &\to \{*\} \\ x &\mapsto *\end{aligned}$$

Disregarding some details about the technical implementation of Set *(i.e., how you choose to represent things like tuples in terms of sets) the equations for commutativity, counitality, and coassociativity are all trivial. Similarly the compatibility of this comonoid structure with the monoidal product is straightforward, but tedious, to verify.*

Generalizing the above example one gets the following result.

Proposition 2.6. *Every Cartesian symmetric monoidal category has a canonical interpretation as a CD category.*

Proof. Let $(\mathsf{D}, \times, I)$ be a Cartesian symmetric monoidal category. Define the copy map componentwise by

$$\texttt{copy}_X := (1_X, 1_X) : X \to X \times X. \tag{2.12}$$

Now, as D is Cartesian we must have that I is the empty product and hence is terminal. Thus, the discard map is componentwise defined to be

$$\texttt{del}_X := \,! : X \to I. \tag{2.13}$$

Commutativity, counitality, and coassociativity, as well as compatibility with the monoidal product, all follow from the universal properties of the product, $\times$, and the terminal object, I. $\square$

Chapter 3

Markov Categories

Having introduced the language of string diagrams we can now give a self-contained introduction to the theory of Markov categories. We have favoured an example driven approach and tried to include examples where they might help build intuition. The main sources for this chapter is [5] and [7] which the reader is invited to look at for a more technical treatment.

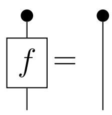
3.1 Basic Notions

Definition 3.1. *A **Markov category** is a CD category* $(\mathsf{C}, \otimes, I)$ *where any of the following equivalent conditions hold*

(i) the monoidal unit I *is a terminal object;*

(ii) the discard maps $\texttt{del} : X \to I$ *are part of a natural transformation* $\texttt{del} : 1_{\mathsf{C}} \Longrightarrow I$*;*

(iii) for every $f : X \to Y$ *in* C *the following equality holds:*



(3.1)

Lemma 3.2. *The conditions presented in Definition 3.1 are equivalent.*

Proof. (i) $\Longrightarrow$ (ii): Assume that the monoidal unit I is terminal. We must show that for all $f : X \to Y$ in C, the following diagram is commutative

$$\begin{tikzcd} X \arrow[r, "f"] \arrow[d, "\texttt{del}_Y"'] & Y \arrow[d, "\texttt{del}_Y"] \\ I \arrow[r, "1_I"'] & I \end{tikzcd}$$

This follows immediately from the assumed uniqueness of the map $\mathtt{del}_Y : Y \to I$.

(ii) $\Longrightarrow$ (iii): This is just a restatement in diagrams of the naturality of $\mathtt{del}$.

(iii) $\Longrightarrow$ (i): Let $f : X \to I$ be a morphism. Equation 3.1 tells us that $\mathtt{del}_I \circ f = \mathtt{del}_X$. However, by Equation 2.11 we have that $\mathtt{del}_I = 1_I$ and hence it follows that $f = \mathtt{del}_X$.

□

Thinking in terms of resource theories, Equation 3.1 tells us that, in a Markov category, transforming a resource and then discarding it is the same as just discarding it. That is to say, there is only one way of discarding something.

Example 3.3. *As the one point set $\{*\}$ is a terminal object in* Set*, the CD structure from Example 2.5 makes* Set *into a Markov category.*

Example 3.4. *Consider the full subcategory* FinSet *of* Set*. Letting the monoidal structure again be determined by the ussual Cartesian product of sets on object we get a Markov structure on* FinSet *via the same copy and discard maps as in the previous example.*

Example 3.5. *Take* $\mathsf{FinStoch}$*, the category of finite sets with Markov kernels as morphisms. That is to say, if $f : X \to Y$ is a morphism in* $\mathsf{FinStoch}$ *then f is a stochastic matrix $(f_{xy})_{x\in X, y\in Y}$ where f_{xy} is interpreted as the probability of outcome $y \in Y$ given the input $x \in X$. Composition of morphisms $f : X \to Y$ and $g : Y \to Z$ amounts to doing matrix multiplication on the stochastic matrices. Using the notation $f(y|x)$ for f_{xy} we have that*

$$(gf)(z|x) = \sum_{y\in Y} g(z|y)f(y|x) \tag{3.2}$$

which one can also recognize as the Chapman-Kolmogorov equations.

Any ordinary function between finite sets can also be thought of as a Markov kernel. In this case, all entries $f(y|x)$ are either 0 or 1. In this way we get an inclusion $\mathsf{FinSet} \hookrightarrow \mathsf{FinStoch}$*.*

Now, the monoidal structure on $\mathsf{FinStoch}$ *is the usual Cartesian product on objects, while it is the tensor product of matrices on morphisms. More explicitly, for $f : A \to X$ and $g : B \to Y$ we have that*

$$(f \otimes g)(x, y|a, b) = f(x|a)g(y|b). \tag{3.3}$$

The copy map $\texttt{copy}_X : X \to X \times X$ *is specified by*

$$\texttt{copy}_X(x_1, x_2|x_0) = \begin{cases} 1, & \text{if } x_1 = x_2 = x_0 \\ 0, & \text{else.} \end{cases} \tag{3.4}$$

Similarly, the discard map $\texttt{del}_X : X \to I$ *is a vector with all ones, i.e.,*

$$\texttt{del}_X(|x) = 1. \tag{3.5}$$

While it is possible to extend $\mathsf{FinStoch}$ to arbitrary measurable spaces there are some technical problems with this approach. Consider the following question: is a measure that assigns either 0 or 1 to each event (i.e., a measurable set) in a measure space $(X, \mathcal{A})$ a Dirac delta at a unique point? A positive answer to this is useful when trying to specify a broader stochastic category.

Definition 3.6 ([15])**.** *A measurable space* X *is called* ***sober*** *if every zero-one measure on* X *is the Dirac delta over a unique point. We denote the category of sober measurable spaces and measurable functions by* $\mathsf{SoberMeas}$*.*

Example 3.7. *The category* $\mathsf{SoberMeas}$ *is a Cartesian symmetric monoidal category, and hence has a canonical CD structure. As the one point sober measurable space* $\{*\}$ *is terminal it follows that* $\mathsf{SoberMeas}$ *is a Markov category.*

Example 3.8. *The category* Stoch *can intuitively be seen as an extension of* $\mathsf{FinStoch}$ *to* $\mathsf{SoberMeas}$*. More specifically,* Stoch *is specified via the following data*

- *Objects are sober measurable spaces, i.e., pairs* $(X, \mathcal{A})$ *where* $X \in \mathsf{Set}$ *and* $\mathcal{A}$ *is a sigma-algebra on* X *subject to the constraint of being sober.*

- *Morphisms* $(X, \mathcal{A}) \to (Y, \mathcal{B})$ *are Markov kernels of entries* $k(B|x)$*, for* $x \in X$ *and* $B \in \mathcal{B}$*. That is to say,* $k(-|x)$ *is a probability measure on* $(Y, \mathcal{B})$ *for all* $x \in X$*.*

- *The identity* $(X, \mathcal{A}) \to (X, \mathcal{A})$ *is given by the Dirac delta, i.e.,*

$$\delta(A|x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{else.} \end{cases} \tag{3.6}$$

  Importantly, the fact that we are working with sober measurable spaces ensures us that the identity is well defined.

- *The composition of two kernels* $k : (X, \mathcal{A}) \to (Y, \mathcal{B})$ *and* $h : (Y, \mathcal{B}) \to (Z, \mathcal{C})$ *is given by taking the Lebesgue integral, i.e.,*

$$(hk)(C|x) = \int_Y h(C|y)k(dy|x) \tag{3.7}$$

for all $x \in X$ *and* $C \in \mathcal{C}$*. The equation above is sometimes called the Chapman-Kolmogorov equation.*

The monoidal and copy-discard structure on Stoch *is an extension of the one on* FinStoch*. On objects, the monoidal product is the product of measurable spaces while on morphisms it is the tensor product of Markov kernels. More explicitly, given* $f : (A, \mathcal{A}) \to (X, \mathcal{C})$ *and* $g : (B, \mathcal{B}) \to (Y, \mathcal{D})$ *we have that*

$$(f \otimes g)(C \times D|a, b) = f(C|a)g(D|b) \tag{3.8}$$

for all $C \in \mathcal{C}$*,* $D \in \mathcal{D}$*,* $a \in A$*, and* $b \in B$*.*

The copy and discard maps also take on analogous forms as the ones in Equation 3.4 and Equation 3.5. That is to say,

$$\texttt{copy}_X(A \times B|x) = \begin{cases} 1, & \text{if } x \in A \times B \\ 0, & \text{else} \end{cases} \tag{3.9}$$

and

$$\texttt{del}_X(|x) = 1. \tag{3.10}$$

An important feature that Stoch lacks is having conditionals. To remedy this we need to focus on standard Borel spaces.

Definition 3.9. *[16] Let* $(X, \mathcal{A})$ *be a measurable space. We say that* X *is a* ***standard Borel space*** *if it can be written as a Polish space with its Borel sigma-algebra. The category of standard Borel spaces and measurable functions between them is denoted by* BorelMeas*.*

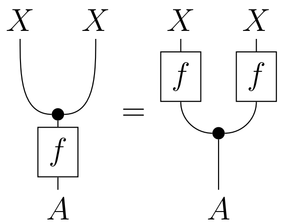
Remark 3.10. *A topological space is called a* ***Polish space****, also known as a completely separably metrizable space, if the topology can be metrized by a complete and separable metric.*

Example 3.11. *The Markov category* BorelStoch *is defined to be the full subcategory of* Stoch *where the objects are standard Borel spaces.*

The attentive reader might have noticed a pattern in the previous examples: we started with a "deterministic category", e.g., FinSet, SoberMeas, BorelMeas, and then we "added" randomness to the morphisms and got FinStoch, Stoch, and BorelStoch respectively. This is not by accident and we will later see how this ties into the notion of probability monads and representability.[1]

We now move on to discussing some important properties of Markov categories that we will make heavy use of later.

Definition 3.12 ([7])**.** *Let* C *be a Markov category. A morphism* $f : A \to X$ *in* C *is said to be* ***deterministic*** *if the following equality holds*



(3.11)

The subcategory of C *that consists of only deterministic morphisms is denoted by* C_{det}*.*

Example 3.13. *In the case of* FinStoch *we have that* $\mathsf{FinStoch}_{det} \cong \mathsf{FinSet}$*. To see why, first note that the equation in 3.11 spelled out in* FinStoch *becomes*

$$\sum_{x_0 \in X} \boldsymbol{copy}_X(x_1, x_2|x_0) f(x_0|a) = \sum_{(a_1,a_2)\in A\times A} f(x_1|a_1) f(x_2|a_2) \boldsymbol{copy}_A(a_1, a_2|a) \quad (3.12)$$

The only contributing summands in the left hand side above is the ones where $x_1 = x_2 = x_0$*. Similarly, the only contributing summands in the right hand side above is the ones where* $a_1 = a_2 = a$*. Using* x *for* $x_1 = x_2$ *the above equation simplifies to*

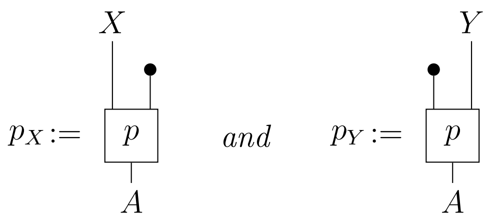
$$f(x|a) = f(x|a)f(x|a) \quad (3.13)$$

and hence we must have that $f(x|a) = 0$ *or* $f(x|a) = 1$ *which is how an ordinary function looks like when we represent it as a stochastic map. Hence, deterministic morphisms in* FinStoch *is the same thing as morphisms in* FinSet*.*

In traditional probability theory we can take the marginal distribution of a collection of random variables of a probability distribution. The analogous concept in Markov categories is that of marginalization.

[1] The category FinStoch is technically not representable, but is very close to being and only fails because the space of probability distributions on a finite set is not itself finite.

Definition 3.14. *Let $p : A \to X \otimes Y$ be a joint morphism in a Markov category* C*. The **marginal morphisms** $p_X : A \to X$ and $p_Y : A \to Y$ are defined as follows:*



(3.14)

Example 3.15. *Consider* FinStoch *and a state $p : \{*\} \to X \times Y$. The marginal morphisms then recover the usual notion of a marginal distribution, i.e., we have*

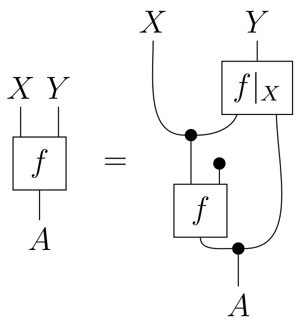
$$p_X(x|) = \sum_{y \in Y} p(x, y|) \tag{3.15}$$

and

$$p_Y(y|) = \sum_{x \in X} p(x, y|). \tag{3.16}$$

One of the most important concepts of this thesis is that of conditionals which we now define.

Definition 3.16. *Let $f : A \to X \otimes Y$ be a joint morphism in a Markov category* C*. A map $f|_X : X \otimes A \to Y$ is called a **conditional** of f with respect to X if the equation*



(3.17)

holds.

Example 3.17. *We claim that* FinStoch *has conditionals. To see this, note that Equation 3.17 takes the following form in* FinStoch

$$f(x, y|a) = f_X(x|a) f|_X(y|a, x). \tag{3.18}$$

Assuming $f_X(x|a) > 0$ this can be rewritten as

$$f|_X(y|a, x) = \frac{f(x, y|a)}{f_X(x|a)}. \tag{3.19}$$

Using this as the definition of $f|_X$ then gives us the Markov kernel we are looking for.

The following example is based on [5, Example 11.3]

Example 3.18. *Does* Stoch *have conditionals? We stated earlier that this is not the case. In fact, this was the reason why we introduced* BorelStoch*. Let us therefore see why* Stoch *does not have conditionals.*

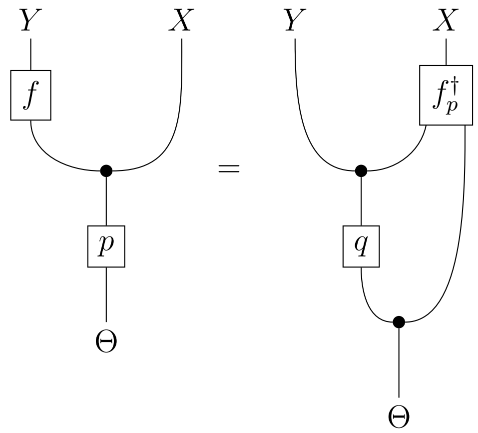
First of all, given a probability measure $\psi : \{*\} \to X \otimes Y$*, a conditional probability distribution* $\psi|_X : X \to Y$ *would, by [5, Lemma 4.2] and Equation 3.17, have to satisfy*

$$\psi(S \times T|) = \int_{s \in S} \psi|_X(T|x)\psi(dx) \tag{3.20}$$

for T *a measurable set in* X *and* S *a measurable set in* Y*. Such a* $\psi|_X$ *is called a* ***product regular conditional probability****, and has been proven to exist whenever* Y *is standard Borel, but not in general [5]. Thus,* Stoch *does not have conditionals, but* BorelStoch *does.*

Closely related to the concept of conditionals is that of a **Bayesian inverse**.

Definition 3.19. *Let* $p : \Theta \to X$ *and* $f : X \to Y$ *be morphism in* C*. A* ***Bayesian inverse*** *of* f *with respect to* p *is a morphism* $f_p^\dagger : \Theta \otimes Y \to X$ *such that the following equation holds*



(3.21)

where $q = p \circ f$*. When* p *is clear from the context we shall drop the subscript* p *from* $f_p^\dagger$ *and just write* $f^\dagger$*.*

Remark 3.20. *A Bayesian inverse is a special case of a conditional. More specifically, it is the conditional of the left hand side in Equation 3.21 with respect to* Y*.*

The intuition behind Equation 3.21 is that is a more general version of Bayes theorem. To see why this is the case, consider the following concrete example.

Example 3.21. *In* FinStoch *the defining equation for a conditional becomes*

$$q(y)f^{\dagger}(x|y) = p(x)f(y|x). \tag{3.22}$$

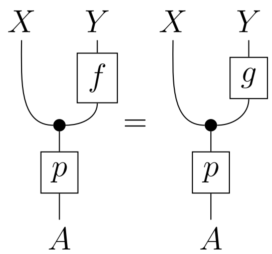
Using the naive approach we can talk about "the probability" of an event with the notation $P(-)$. We can then rewrite the above equation as

$$P(y)P(x|y) = P(x)P(y|x) \tag{3.23}$$

which is Bayes formula on the nose.

As Fritz et al. note in [7], conditionals are generally not unique. However, they are unique up to almost sure equality [5].

Definition 3.22 ([7])**.** *Let $p : A \to X$ be a morphism and $f, g : X \to Y$ two parallel morphisms. We say that f and g are p-almost surely equal, denoted $f =_{p\text{-a.s.}} g$, if we have*



(3.24)

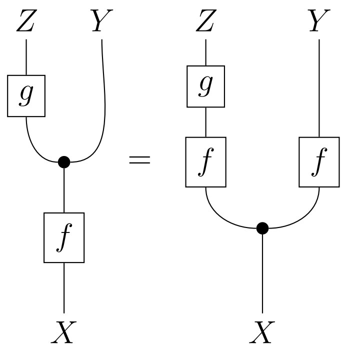
Example 3.23 ([7])**.** *In* BorelStoch*, Definition 3.22 gives the standard notion of almost surely equality. More specifically, given Markov kernels $f, g : (X, \Sigma_A) \to (Y, \Sigma_B)$ and a probability measure $\nu : I \to X$, the relation $f =_{\nu\text{-a.s.}} g$ Equation 3.24 becomes the condition*

$$\int_S f(T|x)\nu(dx) = \int_S g(T|x)\nu(dx), \tag{3.25}$$

for all $S \in \Sigma_X$ and $T \in \Sigma_Y$. This is the same as saying that $f(T|-)$ and $g(T|-)$ are ν-almost everywhere equal for all T.

A useful property of Markov categories with conditionals is positivity.

Definition 3.24 ([5])**.** *We say that a Markov category* C *is* ***positive*** *if whenever $f : X \to Y$ and $g : Y \to Z$ are such that $g \circ f$ is deterministic, then*



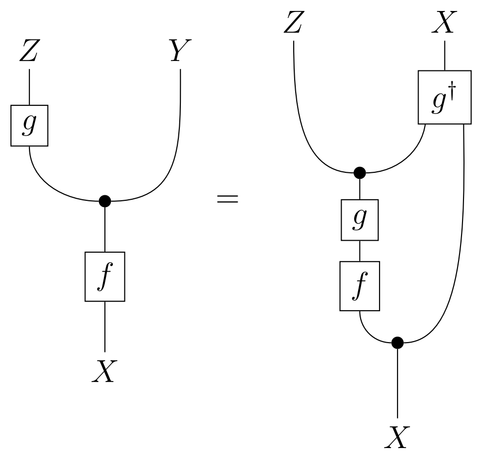
(3.26)

Remark 3.25. *If one instead requires Equation 3.26 to hold p-a.s. for some $p : \Theta \to X$ then one says that* C *is relative positive.*

Proposition 3.26 ([5])**.** *Markov categories with conditionals are positive.*

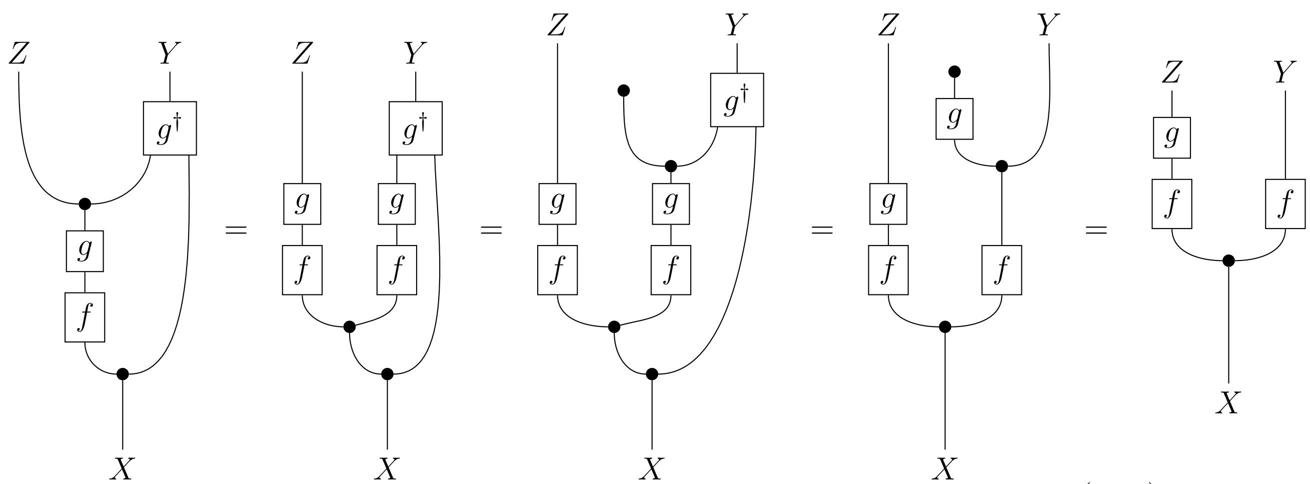
Proof. The following proof is adapted from [5].

Let C be a Markov category with conditionals. We then have



(3.27)

and the determinism assumption further gives



(3.28)

as desired. ∎

3.2 Probability Monads

We previously described how we could "add randomness" to a category in which all morphisms are deterministic. A formal way to do this is through monads, and the type of monads we will look at can be colloquially referred to as probability monads. We will assume the reader is familiar with the usual notion of a monad and build

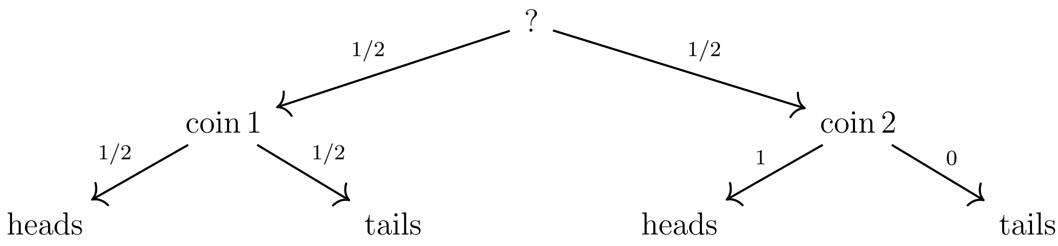


Figure 3.1: A random choice between a normal coin and one with heads on both sides.

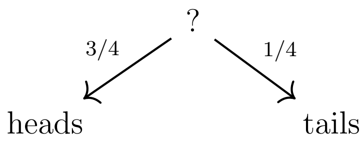


Figure 3.2: The probability distribution obtained by summing over paths in Figure 3.1.

upon this by describing this new class of monads. Those unfamiliar with monads should refer to classic sources such as [13], and [18], or the excellent treatment by [1].

The intuition behind probability monads is best grasped through concrete examples. Thus, imagine the following scenario: your friend presents you with two coins. One coin is regular while the other one has heads on both sides. You choose a coin at random and then flip it. The whole situation can be pictured as in Figure 3.1. What we have here is a probability distribution of probability distributions and we can write it as

$$q = \frac{1}{2}\,\mathrm{coin1} + \frac{1}{2}\,\mathrm{coin2} \tag{3.29}$$

where

$$\begin{aligned} \mathrm{coin1} &= \frac{1}{2}\,\mathrm{heads} + \frac{1}{2}\,\mathrm{tails} \\ \mathrm{coin2} &= 1\,\mathrm{heads} + 0\,\mathrm{tails}. \end{aligned}$$

However, we can simplify things by averaging probabilities to get the new probability distribution

$$p = \frac{3}{4}\,\mathrm{heads} + \frac{1}{4}\,\mathrm{tails}. \tag{3.30}$$

This new distribution is pictured in Figure 3.2.

Summarizing the above:

1. We started with a set of outcomes, in this case $X = \{\mathrm{heads}, \mathrm{tails}\}$.

2. We then created a new set DX which consisted of distributions on X and considered two specific elements, coin1 and coin2.

3. In the choice between coin1 and coin2 we had a distribution of distributions, i.e., an element of DDX.

4. We obtained an element of DX from DDX through averaging.

The attentive reader should recognize this as a monad in disguise.

Definition 3.27. *Consider* Set *and define the* ***distribution monad****, also known as the* ***finitary Giry monad****,* (D, E, δ) *on it as follows:*

- *Given* $X \in \mathsf{Set}$*,* DX *is defined to be the set of functions* $p : X \to [0, 1]$ *such that*

  (i) $p(x) \neq 0$ *for only finitely many* $x \in X$*;*

  (ii) $\sum_{x \in X} p(x) = 1$.

  We refer to elements of DX *as* ***finite distributions****.*

- *Given* $f : X \to Y$*,* $Df : DX \to DY$ *is defined to be the pushforward, i.e., for* $p \in DX$ *we have*

$$Pf(p)(y) = \sum_{x \in f^{-1}(y)} p(x). \tag{3.31}$$

- *The multiplication* $E : PP \Longrightarrow P$ *is defined componentwise by averaging. More specifically, given an element* $\xi \in DDX$ *we have*

$$E_X(\xi)(x) = \sum_{p \in DX} p(x)\xi(p). \tag{3.32}$$

  Note that this sum is finite since by definition ξ *is only non-zero for finitely many* $p \in DX$*.*

- *The unit* $\delta : 1_{\mathsf{Set}} \Longrightarrow P$ *is defined componentwise for* $X \in \mathsf{Set}$ *by*

$$\delta_X(x) = \delta_x \tag{3.33}$$

  where

$$\delta_x(y) := \begin{cases} 1, & \text{if } y = x, \\ 0, & \text{else.} \end{cases} \tag{3.34}$$

It is straightforward to verify that E and δ satisfy the usual monad laws and we leave the details to the interested reader.

The distribution monad is a prototypical probability monad, and we can use it to characterize how a probability monad should look like more generally. For a symmetric monoidal category $(\mathsf{D}, \otimes, I)$ we generally say that a monad (P, μ, δ) on C is a **probability monad** if

1. D is Cartesian, making it into a deterministic Markov category;
2. on objects P assigns a "collection of probability measures" PX for each $X \in \mathsf{D}$;
3. on morphisms P acts like the pushforward of measures;
4. the multiplication μ reduces measures over measures to measures;
5. the unit δ assigns the Dirac delta measure componentwise;

This is not a formal definition and is intended more to work as guideline for ascertaining whether you are dealing with something that can be considered a probability monad. To help solidify this let us take a look at another probability monad.

Definition 3.28. *Consider the category* $\mathsf{SoberMeas}$*. The* ***Giry monad*** (G, μ, δ) *on* $\mathsf{SoberMeas}$ *is defined as follows:*

- *Given* $X \in \mathsf{SoberMeas}$*,* GX *is the space of probability of measures on* X *assigned with the* σ*-algebra generated by the set of all evaluation maps*

$$\begin{aligned} \mathit{ev}_U : GX &\to [0,1] \\ P &\mapsto P(U), \end{aligned}$$

  where U *ranges over all measurable sets in* X*.*

- *Given a measurable function* $f : X \to Y$*,* $Gf : GX \to GY$ *is the pushforward of measures, i.e., for* $P \in GX$ *we have that*

$$Gf(P) = P \circ f^{-1}. \tag{3.35}$$

- *The multiplication is defined componentwise by*

$$\mu_X(Q)(U) = \int_{q \in GX} \mathit{ev}_U(q)\, dQ \tag{3.36}$$

  where $Q \in GGX$ *and* U *is a measurable set on* X*.*

- *The unit is defined componentwise to be the Dirac measure, i.e.,*

$$\delta_X(x) = \delta_x \tag{3.37}$$

where δ_x is the Dirac measure on $x \in X$.

Remark 3.29. *One can also define the Giry monad as an endofunctor P on* $\mathsf{BorelMeas}$. *For $X \in$ $\mathsf{BorelMeas}$, PX is the space of probability measures on the Borel subsets of X. Moreover, PX is equipped with the weakest topology that make the integration map*

$$\tau \mapsto \int_X f \, d\tau \tag{3.38}$$

continuous for any bounded, continuous, real valued function $f : X \to \mathbb{R}$. The multiplication $\mu : PP \Longrightarrow P$ is componentwise defined by

$$\mu_X(M)(A) = \int_{\tau \in PX} \tau(A) M(d\tau) \tag{3.39}$$

where $M \in PPX$ and A is a Borel subset of X. The unit $\delta : 1_{\mathsf{BorelMeas}} \Longrightarrow P$ is the same as for the Giry monad on $\mathsf{SoberMeas}$.

3.3 Representable Markov Categories

The importance of probability monads is that they give rise to representable Markov categories via the Kleisli construction. For example, $Kl(G) = \mathsf{Stoch}$ where G is the Giry monad on $\mathsf{SoberMeas}$, and similarly, $Kl(P) = \mathsf{BorelStoch}$ for P the Giry monad on $\mathsf{BorelMeas}$. Equally important is the other direction and we will give a classification for when a Markov category is representable.

We first need some terminology on monoidal functors and monads.

Definition 3.30 ([17])**.** *A lax monoidal functor $(\mathsf{C}, \otimes, I_\mathsf{C}) \to (\mathsf{D}, \otimes, I_\mathsf{D})$ is a triple (F, η, ∇) such that:*

(i) $F : \mathsf{C} \to \mathsf{D}$ *is a functor;*

(ii) The unit $\eta : I_\mathsf{D} \to FI_\mathsf{C}$ *is a morphism of* D*;*

(iii) The composition $\nabla : F(-) \otimes F(-) \Longrightarrow F(- \otimes -)$ *is a natural transformation of functors* $\mathsf{C} \times \mathsf{C} \to \mathsf{D}$*;*

(iv) The following associativity diagram commutes for every $X, Y, Z \in \mathsf{C}$:

$$\begin{tikzcd}
(FX \otimes FY) \otimes FZ \arrow[r, "\cong"] \arrow[d, "\nabla_{X,Y} \otimes 1_{FZ}"'] & FX \otimes (FY \otimes FZ) \arrow[d, "1_{FX} \otimes \nabla_{Y,Z}"] \\
F(X \otimes Y) \otimes FZ \arrow[d, "\nabla_{X \otimes Y, Z}"'] & FX \otimes F(Y \otimes Z) \arrow[d, "\nabla_{X, Y \otimes Z}"] \\
F((X \otimes Y) \otimes Z) \arrow[r, "\cong"'] & F(X \otimes (Y \otimes Z))
\end{tikzcd}$$

(v) The following unitality diagrams commute for every $X \in \mathsf{C}$:

$$\begin{tikzcd}
I_{\mathsf{D}} \otimes FX \arrow[r, "\eta \otimes 1_{FX}"] \arrow[d, "\cong"'] & FI_{\mathsf{C}} \otimes FX \arrow[d, "\nabla_{I_{\mathsf{C}}, X}"] \\
FX & F(I_{\mathsf{C}} \otimes X) \arrow[l, "\cong"]
\end{tikzcd}
\qquad
\begin{tikzcd}
FX \otimes I_{\mathsf{D}} \arrow[r, "1_{FX} \otimes \eta"] \arrow[d, "\cong"'] & FX \otimes FI_{\mathsf{C}} \arrow[d, "\nabla_{X, I_{\mathsf{C}}}"] \\
FX & F(X \otimes I_{\mathsf{C}}) \arrow[l, "\cong"]
\end{tikzcd}$$

We say that (F, η, ∇) is also ***symmetric****, or* ***braided*** *if C is symmetric and the multiplication commutes with the braiding:*

$$\begin{tikzcd}
FX \otimes FY \arrow[r, "\cong"] \arrow[d, "\nabla_{X,Y}"'] & FY \otimes FX \arrow[d, "\nabla_{Y,X}"] \\
F(X \otimes Y) \arrow[r, "\cong"'] & F(Y \otimes X)
\end{tikzcd}$$

Definition 3.31 ([17])**.** *Let (F, η_F, ∇_F) and (G, η_G, ∇_G) be lax monoidal functors of type $(\mathsf{C}, \otimes, I_{\mathsf{C}}) \to (\mathsf{D}, \otimes, I_{\mathsf{D}})$. A* ***lax monoidal natural transformation****, or just a* ***monoidal natural transformation*** *(when it is clear from the context), is a natural transformation $\alpha : F \implies G$ which is compatible with the unit and multiplication. More specifically, for all $X, Y \in \mathsf{C}$ the following diagrams commute*

$$\begin{tikzcd}
I_{\mathsf{D}} \arrow[r, "\eta_F"] \arrow[dr, "\eta_G"'] & FI_{\mathsf{C}} \arrow[d, "\alpha_{I_{\mathsf{C}}}"] \\
& GI_{\mathsf{C}}
\end{tikzcd}
\qquad
\begin{tikzcd}
FX \otimes FY \arrow[r, "\nabla_F"] \arrow[d, "\alpha_X \otimes \alpha_Y"'] & F(X \otimes Y) \arrow[d, "\alpha_{X \otimes Y}"] \\
GX \otimes GY \arrow[r, "\nabla_G"'] & G(X \otimes Y)
\end{tikzcd}$$

Definition 3.32. *A* ***monoidal monad*** *is a monad in the bicategory of monoidal categories, lax monoidal functors, and monoidal transformations. Moreover, if C is a symmetric monoidal category, then a monoidal monad on C is* ***symmetric*** *if the underlying monoidal monad functor F is a symmetric monoidal functor.*

The importance of the above conditions is the following result.

Proposition 3.33 ([5]). *Let (P, μ, δ) be a symmetric monoidal monad on some symmetric monoidal category $(D, \otimes, I)$. Then $Kl(P)$ is a symmetric monoidal category, with:*

- *the same monoidal product as the one in D;*
- *the tensor product of morphisms represented by $f : A \to PX$ and $g : B \to PY$ being represented by the composite*

$$A \otimes B \xrightarrow{f \otimes g} PX \otimes PY \xrightarrow{\nabla} P(X \otimes Y). \tag{3.40}$$

Moreover, the inclusion $\mathsf{D} \to Kl(P)$ is strict symmetric monoidal.

Proof. See [5, Proposition 3.1]. □

The final statement about the inclusion implies that if $X \in \mathsf{D}$ has a distinguished comonoid structure, then so does $X \in Kl(P)$.

Definition 3.34. *A monad (P, μ, δ) on D is said to be* ***affine*** *if $PI \cong I$.*

Thus, if P is affine and I is terminal, then $PI \in Kl(D)$ is also terminal and we get the following result.

Corollary 3.35 ([5]). *Let (P, μ, δ) be a symmetric monoidal affine monad on a Markov category D. Then the Kleisli category $Kl(P)$ is again a Markov category in a canonical way.*

Proof. Follows immediately from Proposition 3.33 and the fact that $PI \cong I$ is terminal. More specifically, the copy map for an object $X \in Kl(P)$ is given as the composite

$$X \xrightarrow{\mathsf{copy}_X} X \otimes X \xrightarrow{\delta} P(X \otimes X). \tag{3.41}$$

□

It can be shown, although it is rather tedious, that the Giry monad is symmetric monoidal affine (see e.g. [5, Lemma 4.1]). The same is true for a lot of the commonly encountered probability monads. This ensures us that the Kleisli categories are Markov categories and we can use all the tools developed so far.

Now, the previous paragraphs tell the story where we start with a deterministic category, add randomness via a probability monad, and then end up with a Markov category via the Kleisli construction. So, what about the other direction? Suppose we start with a Markov category C. When is it the case that there is a category

D and a monad P on D such that $\mathsf{C} \cong Kl(P)$, i.e., such that C is "representable"? To answer this question we will use the following result which intuitively states that the deterministic morphisms of a representable Markov category $Kl(P)$ is exactly the underlying category D.

Proposition 3.36 ([7])**.** *Let* D *be a Cartesian monoidal category and* (P, μ, δ) *an affine symmetric monoidal monad on* D*. Then the canonical functor* $\mathsf{D} \to Kl(P)_{det}$ *is an isomorphism of categories if and only if the diagram*

$$\begin{tikzcd} X \arrow[r, "\delta"] \arrow[d, "{(\delta,\delta)}"'] \arrow[dr, phantom, "\lrcorner", very near start] & PX \arrow[d, "{P(1_X,1_X)}"] \\ PX \times PX \arrow[r, "\nabla"'] & P(X \times X) \end{tikzcd} \tag{3.42}$$

is a pullback for every $X \in \mathsf{D}$*.*

Proof. The following proof is due to Fritz et al. [7].

Assume that the diagram described above is a pullback. We only need to prove that $\mathsf{D} \to Kl(P)_{\mathrm{det}}$ is fully faithful (seeing as it is the identity on objects). To see that it is faithful, suppose $\delta_Y \circ f = \delta_Y \circ g$ for some $f, g : X \to Y$. Now, ∇ has a left inverse given by the canonical map

$$\Delta : P(X \times X) \to PX \times PX \tag{3.43}$$

which is induced from the Cartesian monoidal structure on D. Thus, ∇ is a monomorphism as it has a left inverse. Now, it is a known fact that monomorphisms are stable under pullbacks and hence it follows that $\delta : X \to PX$ is a monomorphism as well. We therefore have that $f = g$ as desired.

Next, to show that $\mathsf{D} \to Kl(P)_{\mathrm{det}}$ is full, let $f^\# : A \to PX$ be the representative of a deterministic morphism $f : A \to X$ in $Kl(P)$. The determinism assumption can be shown to equivalent (by a straightforward unfolding of definitions) to commutativity of the diagram

$$\begin{tikzcd} A \arrow[r, "f^\#"] \arrow[d, "{(f^\#,f^\#)}"'] & PX \arrow[d, "{P(1_X,1_X)}"] \\ PX \times PX \arrow[r, "\nabla"'] & P(X \times X) \end{tikzcd} \tag{3.44}$$

We can now use the pullback assumption to get a unique map $\tilde{f} : A \to X$ such that

$$\begin{tikzcd}
A \arrow[dr, dashed, "\tilde{f}"] \arrow[drr, bend left, "f^{\#}"] \arrow[ddr, bend right, "(f^{\#}, f^{\#})"'] & & \\
& X \arrow[r, "\delta"] \arrow[d, "(\delta,\delta)"'] & PX \arrow[d, "P(1_X, 1_X)"] \\
& PX \times PX \arrow[r, "\nabla"'] & P(X \times X)
\end{tikzcd} \tag{3.45}$$

is commutative. However, $\tilde{f}$ is exactly the map we need as the commutativity implies $\delta \circ \tilde{f} = f^{\#}$.

For the other direction, suppose that $\mathsf{D} \to Kl(P)_{\det}$ is an isomorphism of categories. Let $f^{\#} : A \to PX$ and $g : A \to PX \times PX$ be such that

$$\begin{tikzcd}
A \arrow[drr, bend left, "f^{\#}"] \arrow[ddr, bend right, "g"'] & & \\
& X \arrow[r, "\delta"] \arrow[d, "(\delta,\delta)"'] & PX \arrow[d, "P(1_X, 1_X)"] \\
& PX \times PX \arrow[r, "\nabla"'] & P(X \times X)
\end{tikzcd} \tag{3.46}$$

is commutative. The commutativity of this diagram then forces $g = (f^{\#}, f^{\#})$. Now, the morphisms $f^{\#} : A \to PX$ represents a morphism $f : A \to X$ in $Kl(P)_{\det}$. Via the isomorphism $\mathsf{D} \to Kl(P)_{\det}$ we then get a unique morphism $\tilde{f} : A \to X$ in D such that $\delta \circ \tilde{f} = f^{\#}$. We have thus found the desired morphism. $\square$

So, back to the question at hand: when does a Markov category arise as the Kleisli category of a probability monad? To gain some intuition for this it might be helpful to assume that such a monad exist and see what we can learn about C. What we learn might then help us discover the correct assumptions that we need to make on C.

Thus, assume $\mathsf{C} \cong Kl(P)$ for some probability monad P on a symmetric monoidal category D. This implies that

$$\mathsf{C}(X, Y) \cong Kl(P)(X, Y) = \mathsf{D}(X, PY). \tag{3.47}$$

Using Proposition 3.36 we then get that

$$\mathsf{D}(X, PY) \cong Kl(P)_{\det}(X, PY) \cong \mathsf{C}_{\det}(X, PY). \tag{3.48}$$

In particular, this means that we have natural bijections

$$C(X, Y) \cong \mathsf{C}_{\det}(X, PY). \tag{3.49}$$

For example, a Markov kernel $f : X \to Y$ can be thought of as being represented by a deterministic map $f^{\#} : X \to PY$.

From the standard theory of adjunctions we can then conclude that the inclusion functor $\mathsf{C}_{\text{det}} \to \mathsf{C}$ has a right adjoint $\mathsf{C} \to \mathsf{C}_{\text{det}}$ which we will also denote by P. The naturality of Equation 3.3 tells us how this $P : \mathsf{C} \to \mathsf{C}_{\text{det}}$ acts on morphisms: a morphism $f : X \to Y$ in C which is represented by $f^{\#} : X \to PY$ gets taken to

$$PX \xrightarrow{Pf^{\#}} PPY \xrightarrow{\mu} PY \tag{3.50}$$

where μ is the multiplication of the monad. This might seem like a circular definition, but the P in Equation 3.50 is the original $P : \mathsf{D} \to \mathsf{D}$ and not the new $P : \mathsf{C} \to \mathsf{C}_{\text{det}}$. The reader might complain about some abuse of notation here, but the point is that the new $P : \mathsf{C} \to \mathsf{C}_{\text{det}}$ is really the old $P : \mathsf{D} \to \mathsf{D}$ in disguise.

The bijection from Equation also gives us the unit and counit of the adjunction $\mathsf{C}_{\text{det}} \rightleftarrows \mathsf{C}$. More specifically, we have that the unit $\delta : 1_{\mathsf{C}_{\text{det}}} \Longrightarrow P$ is componentwise the map that assigns delta distributions. On the other hand, the counit $\mathtt{samp} : P \Longrightarrow 1_{\mathsf{C}}$ is componentwise represented by the identity map $1_{PX} : PX \to PX$. Using our previous notation we therefore have that

$$(\mathtt{samp}_X)^{\#} = 1_{PX}. \tag{3.51}$$

The reason for labeling the counit with the label samp is that it componentwise acts as a sampling map. That is to say, given a state $p : I \to PX$, composing with $\mathtt{samp}_X$ gives a state which is represented by the composition

$$I \xrightarrow{p^{\#}} PPX \xrightarrow{P(\mathtt{samp}_X^{\#})} PPX \xrightarrow{\mu_X} PX. \tag{3.52}$$

Remembering that $(\mathtt{samp}_X)^{\#} = 1_{PX}$, this is really the map

$$I \xrightarrow{p^{\#}} PPX \xrightarrow{\mu_X} PX. \tag{3.53}$$

In other words, applying samp to a state p behaves the same as sampling that state. Based on these observations we therefore make the following more general definition.

Definition 3.37 ([7]). *Let C be a Markov category and $X \in \mathsf{C}$ an object. A **distribution object** for X is an object PX together with a morphism $\mathtt{samp}_X : PX \to X$ so that the induced map*

$$\mathtt{samp}_X \circ - : \mathsf{C}_{det}(A, PX) \to C(A, X) \tag{3.54}$$

is a bijection for all $A \in \mathsf{C}$. We denote the inverse of this map by

$$(-)^{\#} : \mathsf{C}(A, X) \to \mathsf{C}_{det}(A, PX) \tag{3.55}$$

which also explains our use of the $^{\#}$ symbol previously.

Using this notation we also have a more concrete description (other than just being the unit of an adjunction) of the abstract delta distribution, namely

$$\delta_X := (1_X)^{\#}. \tag{3.56}$$

In other words, it is the unique deterministic morphism $\delta_X : X \to PX$ such that

$$\mathtt{samp}_X \circ \delta_X = 1_X. \tag{3.57}$$

We can now properly define what it means for a Markov category to be representable.

Definition 3.38 ([7])**.** *Let C be a Markov category. We say that C is **representable** if every object has a distribution object. The corresponding right adjoint functor $P : \mathsf{C} \to \mathsf{C}_{det}$ is called the **distribution functor** for C.*

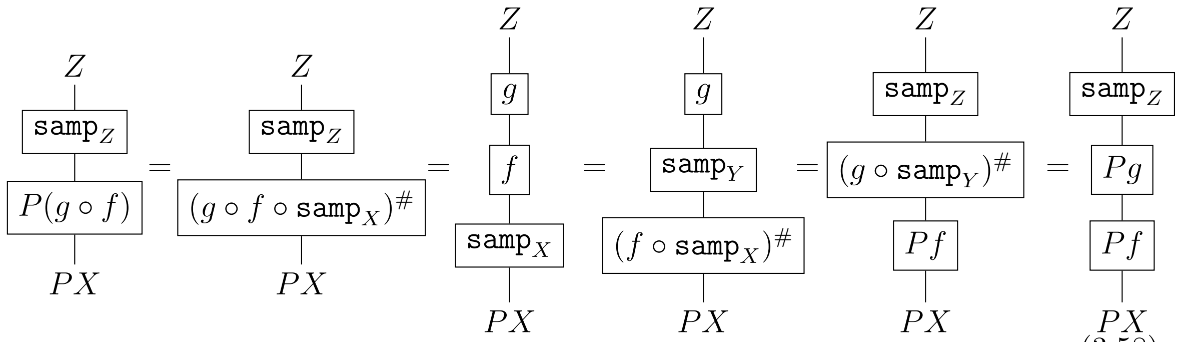
Inherent in the above definition is the claim that the distribution objects assemble together to give a right adjoint $\mathsf{C} \to \mathsf{C}_{\mathrm{det}}$. More concretely, we have the following lemma.

Lemma 3.39 ([7])**.** *Let C be a Markov category in which every object $X \in \mathsf{C}$ has a distribution object PX. The assignment $X \mapsto PX$ is the object part of a functor $P : \mathsf{C} \to \mathsf{C}_{det}$ which is the right adjoint to the inclusion functor $\mathsf{C}_{det} \to \mathsf{C}$, and with the counit of the adjunction being the natural transformation which componentwise is the sampling map of the distribution object.*

Proof. As Fritz et al. mentions [7] this is part of the standard theory on adjunctions, but we include a proof nonetheless in the hopes that the reader might find it enlightning.

We must first specify how P acts on morphisms. As is usual in category theory, there is really only one sensible way to define it: P takes a morphism $f : X \to Y$ in C to $(f \circ \mathtt{samp}_X)^{\#} : PX \to PY$ in $\mathsf{C}_{\mathrm{det}}$. To see that this assignment is functorial, let

$g : Y \to Z$ be another morphism in C. We then have that



(3.58)

and since $\texttt{samp}_Z$ is a bijection, we must have that

$$P(g \circ f) = Pg \circ Pf \tag{3.59}$$

showing that P indeed is a functor. We will skip the step of showing that the δ_X's assemble together to a natural transformation $\delta : 1_{\mathsf{C}_{\text{det}}} \Longrightarrow P$ and likewise that the $\texttt{samp}_X$'s assemble together to a natural transformation $\texttt{samp} : P \Longrightarrow 1_{\mathsf{C}}$ (we have again done some abuse of notation by using the same notation P for functors which are technically different but in reality act the same). The proof is just an unwrapping of definitions and checking that everything work as expected.

The final step is showing that δ and $\texttt{samp}$ satisfy the triangle identities which componentwise is the requirements

$$\texttt{samp}_X \circ \delta_X = 1_X \tag{3.60}$$
$$P\texttt{samp}_X \circ \delta_{PX} = 1_{PX}. \tag{3.61}$$

The first equation is immediate from the definition of δ_X. The second equation is also immediate after an unrolling of definitions. More explicitly, applying $\texttt{samp}_X$ to the left hand side, we have

$$\begin{aligned} \texttt{samp}_X \circ P\texttt{samp}_X \circ \delta_{PX} &= \texttt{samp}_X \circ \texttt{samp}_{PX}\delta_{PX} \\ &= \texttt{samp}_X, \end{aligned}$$

where we used the naturality of $\texttt{samp}$. Since $\texttt{samp}_X \circ -$ is a bijection we must have that

$$P\texttt{samp}_X \circ \delta_{PX} = 1_{PX} \tag{3.62}$$

as desired. □

Remark 3.40. *A subtle point which might not have been clear in the proof above is that δ is only natural when considering P on C_{det}. As soon as you consider all of C, it is no longer natural. This can, as Fritz et al. [7] mentions, initially be confusing. For the interested reader, this problem is discussed at length in [14].*

All this leads us to the following theorem which completely classifies representable Markov categories in terms of the existence of a specific kind of monad on the deterministic subcategory.

Theorem 3.41 ([7])**.** *Let C be a Markov category. Then the following statements are equivalent:*

(i) C is representable;

(ii) there is an affine symmetric monoidal monad P on C_{det} such that

- *the diagram in 3.42 is a pullback for every $X \in \mathsf{C}$;*
- *the identity functor on C_{det} extends to an isomorphism of Markov categories $\mathsf{C} \cong Kl(P)$.*

Proof. See [7, Theorem 3.19]. □

Remark 3.42. *The monad on C_{det} is given by $(P, P\texttt{samp}, \delta)$ where P is taken to be the composition of the inclusion functor $\mathsf{C}_{det} \to \mathsf{C}$ with the $P : \mathsf{C} \to \mathsf{C}_{det}$ defined earlier.*

3.4 Other Notions of Representability

While representable Markov categories are nice, they do not tell the whole story. Consider again the natural bijection

$$\mathsf{C}_{\mathrm{det}}(A, PX) \cong \mathsf{C}(A, X). \tag{3.63}$$

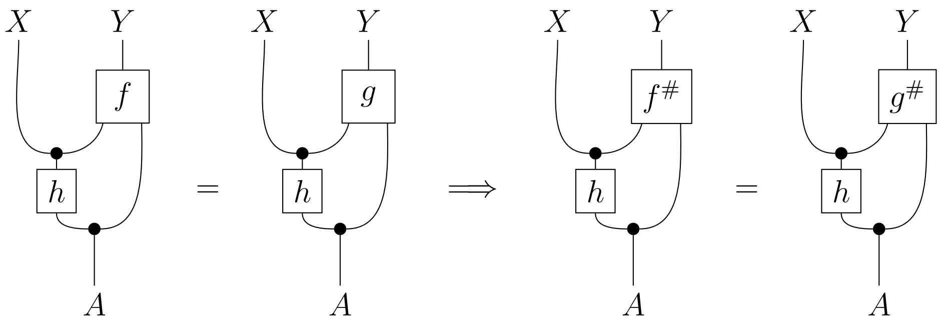
Is it the case that this bijection respects almost sure equality? Suppose for example that we have two deterministic morphisms $f^{\#}, g^{\#} : A \to PX$ which are almost surely equal with respect to a morphism $p : \Theta \to A$. In this case, the resulting morphisms $f, g : A \to X$ are also almost surely equal with respect to Θ as f and g are obtained by applying `samp` to $f^{\#}$ and $g^{\#}$. The other direction is not so clear: are the deterministic versions of two almost surely equal morphisms $A \to X$ still almost surely equal? Fritz et al. proves in [7, Example 3.26] that this is indeed not always the case. This motivates the following definition.

Definition 3.43 ([7])**.** *Let* C *be a Markov category. We say that* C *is* ***a.s.-compatibly representable*** *if it is representable and for any morphism* $p : \Theta \to A$*, the natural bijection from Equation 3.63 respects almost sure equality. That is to say, for all* $f, g : A \to X$*, we have*

$$f^{\#} =_{p\text{-}a.s} g^{\#} \iff f =_{p\text{-}a.s.} g. \tag{3.64}$$

In later proofs we shall make use of another equivalent characterization of a.s.-compatibly representable Markov categories.

Definition 3.44 ([7])**.** *Let* C *be a representable Markov category. It is said to satisfy the* ***sampling cancellation property*** *if, for any three morphisms* $f, g : X \otimes A \to Y$ *and* $h : A \to X$*, the following implication holds*



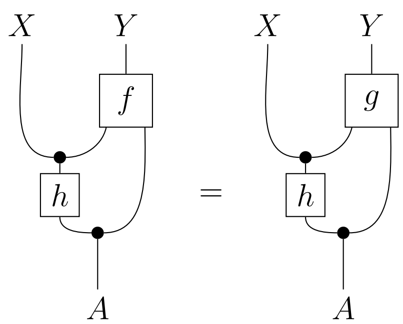
(3.65)

Remark 3.45. *Remembering that* $f = \texttt{samp} \circ f^{\#}$ *for all morphisms in* C*, the condition above is named so because it amounts to cancelling the* $\texttt{samp}$ *from* f *and* g*.*

We then have the following classification of a.s.-compatibly representable Markov categories.

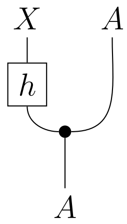
Proposition 3.46 ([7])**.** *Let* C *be a representable Markov category. Then* C *is a.s.-compatibly representable if and only it satisfies the sampling cancellation property.*

Proof. Assume that C is a.s.-compatibly representable and let $f, g : X \otimes A \to Y$ and $h : A \to X$ be such that

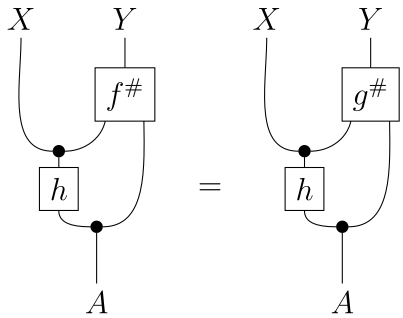


(3.66)

If in Equation 3.24 we set p to be the morphism

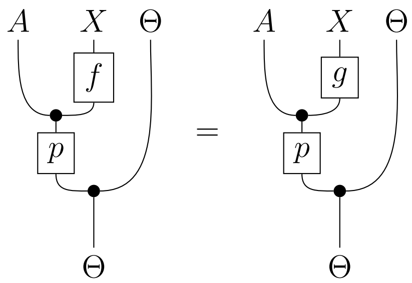
$p :=$  (3.67)

then it follows that

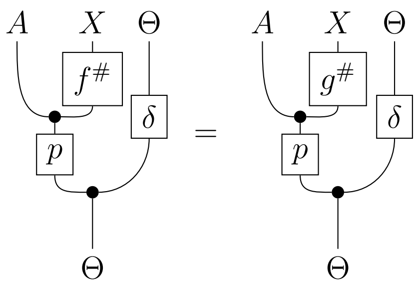
 (3.68)

as desired.

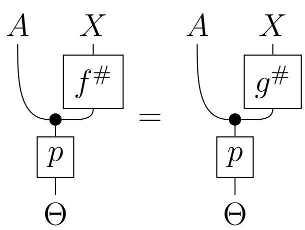
Now, suppose conversely that C satisfies the sampling cancellation property. Let $f, g : A \to X$ be morphisms and assume $f =_{p\text{-a.s.}} g$ for some $p : \Theta \to A$. We need to show that this implies $f^{\#} =_{p\text{-a.s.}} g^{\#}$. By assumption we have that

 (3.69)

This is an instance of the left side of the implication in Equation 3.65, substituting with appropriate maps. By assumption we must therefore have that

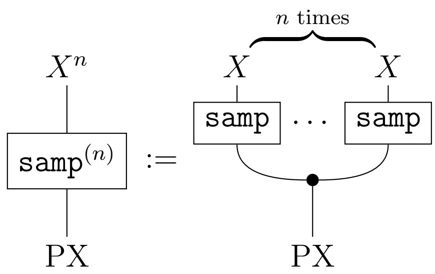
 (3.70)

Marginalizing we therefore have

 (3.71)

which was what we wanted to show. $\square$

In traditional probability theory, we can distinguish between different probability distributions by iterated sampling. For example, if you take samples from the two normal distributions $\mathcal{N}(0,1)$ and $\mathcal{N}(1,1)$ then you will eventually start to see that the samples cannot be taken from the same distribution. In representable Markov categories we have the `samp` map that we can use to get a similar notion. More specifically, we have a notion of **iterated sampling** which is defined as the morphism



(3.72)

Suppose we have a states $p, q : I \to X$ in C. This is the same as having deterministic states $p^\#, q^\# : I \to PX$. Taking samples in ordinary probability theory should then intuitively correspond to post-composing $p^\#$ and $q^\#$ with $\mathtt{samp}^{(n)}$ in C. Moreover, in traditional probability theory, if we always get the same thing from p and q no matter how many times we sample, we would conclude that they must be the same distribution. This corresponds to the statement that if $\mathtt{samp}^{(n)} f = \mathtt{samp}^{(n)} g$ for all $n \in \mathbb{N}$ then $f = g$. In other words, the morphisms $\{\mathtt{samp}^{(n)}\}_{n \in \mathbb{N}}$ are jointly monic.

Definition 3.47. *Let C be a representable Markov category. We say that C is **observationally representable** if the morphisms $\{\mathtt{samp}^{(n)}\}_{n\in\mathbb{N}}$ are jointly monic, i.e., for all $n \in \mathbb{N}$, and all morphisms $f, g : A \to PX$ we have that if $\mathtt{samp}^{(n)} f = \mathtt{samp}^{(n)} g$, then $f = g$.*

It turns out that being observationally representable is equivalent with being "a.s.-compatibly observationally representable".

Proposition 3.48 ([6]). *Let C be a representable Markov category. The following are equivalent*

(i) the morphisms $\{\mathtt{samp}^{(n)}\}_{n\in\mathbb{N}}$ are jointly monic;

(ii) the morphisms $\{\mathtt{samp}^{(n)}\}$ are jointly monic modulo a.s. equality, i.e., for all $p : A \to X$, and $f, g : X \to PY$, we have

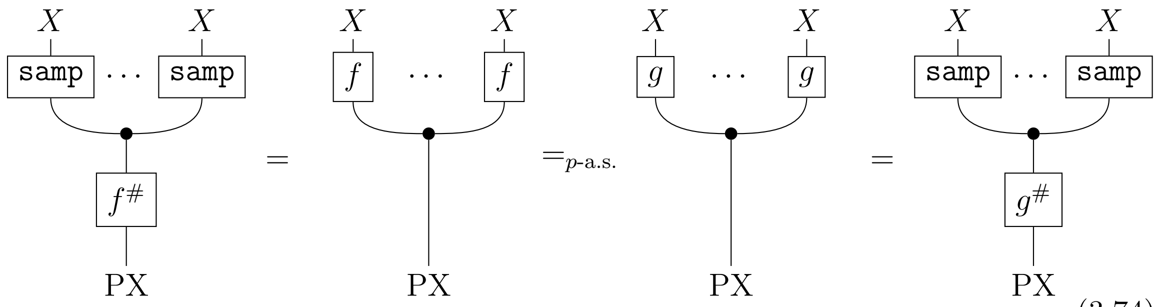
$$\mathtt{samp}^{(n)} f =_{p\text{-a.s.}} \mathtt{samp}^{(n)} g \quad \forall n \in \mathbb{N} \Longrightarrow f =_{p\text{-a.s.}} g. \tag{3.73}$$

Proof. See [6, Proposition A.4.2]. □

We can now place observationally representable into the hierarchy of the different notions of representability.

Proposition 3.49 ([6])**.** *Every observationally representable Markov category is also a.s.-compatibly representable.*

Proof. Let $f, g : A \to X$ be such that $f =_{p\text{-a.s.}} g$. Then, as $f^{\#}$ and $g^{\#}$ are deterministic we have



(3.74)

where the middle p-a.s. equality follows from [5, Lemma 13.5]. Observationally representability then implies that $f^{\#} =_{p\text{-a.s.}} g^{\#}$. □

Chapter 4

Partial Evaluations

We now turn our attention to partial evaluation and see how they fit into the framework of Markov categories. In fact, we will see that partial evaluations correspond exactly to the second-order dominance relation on Markov categories.

Now, the starting point for partial evaluations is the observation that while a formal expression like $1+2+3$ can be totally evaluated to 6 it can also be partially evaluated to $3+3$ or $1+5$. Keeping in mind this remark from Hyland et. al. [12], *a monad is like a consistent choice of spaces of formal expression in a signature*, it makes sense to formalize partial evaluations in the theory of monads.

4.1 General Theory

We want to be able to relate partial evaluations to the dominance relation on $\mathsf{C}(\Theta, X)$ for all $\Theta, X \in \mathsf{C}$. With this in mind we will give a very general treatment of partial evaluations, but the reader should keep in mind that intuition is best gotten when thinking about the case $\Theta = I$.

Definition 4.1. *Let C be a category and $X \in \mathsf{C}$. An S-shaped* ***generalized element*** *of X is a morphism $p : S \to X$ for some $S \in \mathsf{C}$. By abuse of notation we will also write $p \in X$ when p is a generalized element.*

Example 4.2. *Putting $\mathsf{C} = \mathsf{Set}$ and $S = \{*\}$ recovers the usual notion of elements in a set.*

Definition 4.3. *Let (T, μ, η) be a monad on some category C, and $A \in \mathsf{C}$. An S-shaped* ***generalized formal expression*** *on A is an S-shaped generalized element $p \in TA$ for some $S \in \mathsf{C}$.*

A partial evaluation of $1+2+3$ only makes sense when we also have a notion of "total evaluation". The way to formalize total evaluations of formal expressions is through T-algebras for some monad T.

Definition 4.4. *Let C be a category and (T, μ, η) a monad on it. A T-**algebra** is a pair (A, e) where $A \in \mathsf{C}$ and $e : TA \to A$ such that the following diagrams commute*

$$\begin{tikzcd} TTA \arrow[r, "Te"] \arrow[d, "\mu"'] & TA \arrow[d, "e"] \\ TA \arrow[r, "e"'] & A \end{tikzcd} \qquad \begin{tikzcd} A \arrow[r, "\eta_A"] \arrow[dr, "1_A"'] & TA \arrow[d, "e"] \\ & A. \end{tikzcd}$$

A morphism of T-algebras $(A, e) \to (B, d)$ is a morphism $f : A \to B$ in C such that the diagram

$$\begin{tikzcd} TA \arrow[r, "Tf"] \arrow[d, "e"'] & TB \arrow[d, "d"] \\ A \arrow[r, "f"'] & B \end{tikzcd}$$

*is commutative. The category whose objects are T-algebras and morphisms are morphisms of T-algebras is denoted by C^T and is referred to as an **Eilenberg-Moore category**.*

Many of the usual constructions in algebra can be formalized this way. For example, Abelian groups are the same thing as $\mathbb{N}$-algebras where $\mathbb{N}$ is the commutative monoid monad. Similarly, groups, rings, modules, etc. can all be instantiated as algebras of monads.

Now, if we have an algebra, then it makes sense to talk about the result of a total evaluation.

Definition 4.5 ([8])**.** *Let (A, e) be a T-algebra. Given a generalized formal expression $p \in TA$, the **result** of p is the generalized element $e \circ p \in A$.*

Having defined the necessary terminology we can now give proper meaning to partial evaluations.

Definition 4.6. *Let $p, q \in TA$ be S-shaped generalized formal expressions on a T-algebra (A, e). A **partial evaluation** from p into q is an S-shaped generalized element $k \in TTA$ such that the following diagram commutes*

$$\begin{tikzcd} & S \arrow[dl, "p"'] \arrow[d, "k"] \arrow[dr, "q"] & \\ TA & TTA \arrow[l, "\mu"] \arrow[r, "Te"'] & TA. \end{tikzcd}$$

Thus, partial evaluation is a relation on $\mathsf{C}(S, TA)$ and we can ask ourselves about what properties this relation has: is it reflexive, transitive, something else?

For example, suppose that we have a partial evaluation from p to q and from q to r, does this imply that there exist a partial evaluation from p to r? In other words, can we compose partial evaluations? To answer this we need to introduce some terminology.

Definition 4.7. *We say that a commutative diagram*

$$\begin{tikzcd} A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[d, "m"] \\ C \arrow[r, "n"'] & D \end{tikzcd}$$

is a ***weak pullback*** *if, given $p : S \to C$ and $q : S \to B$ such that $n \circ p = m \circ q$ then there exists $r : S \to A$ such that*

$$\begin{tikzcd} S \arrow[drr, bend left, "q"] \arrow[ddr, bend right, "p"'] \arrow[dr, dashed, "r"] & & \\ & A \arrow[r, "f"] \arrow[d, "g"'] & B \arrow[d, "m"] \\ & C \arrow[r, "n"'] & D \end{tikzcd}$$

is commutative.

Remark 4.8. *This is almost the same as a pullback except that we have dropped the uniqueness condition.*

Definition 4.9. *Let (T, μ, η) be a monad on some category C. We say that μ is* ***weakly Cartesian*** *if the diagram*

$$\begin{tikzcd} TTX \arrow[r, "TTf"] \arrow[d, "\mu_X"'] & TTY \arrow[d, "\mu_Y"] \\ TX \arrow[r, "Tf"'] & TY \end{tikzcd}$$

is a weak pullback for all $X, Y \in \mathsf{C}$ and $f : X \to Y$.

Definition 4.10 ([4])**.** *Let (T, μ, η) be a monad on some category C. We say that T is* ***Beck-Chevalley*** *(BC for short) if μ is weakly Cartesian and T preserves weak pullback.*

Proposition 4.11 ([8])**.** *Let (T, μ, η) be a monad on a category C, $S \in \mathsf{C}$, and $A \in \mathsf{C}^T$. Then $\mathsf{C}(S, TA)$ equipped with the partial evaluation relation $\to$ has the following properties*

- *Reflexivity: for each $s \in \mathsf{C}(S, TA)$, $s \to s$;*
- *Confluence: for all $s, t, u \in \mathsf{C}(S, TA)$ such that*

$$\begin{tikzcd} & s \arrow[dl] \arrow[dr] & \\ t & & u \end{tikzcd}$$

there exists $z \in \mathsf{C}(S, TA)$ such that

$$\begin{tikzcd} & s \arrow[dl] \arrow[dr] & \\ t \arrow[dr] & & u \arrow[dl] \\ & z. & \end{tikzcd}$$

More specifically, z can be obtained by $\eta \circ e \circ t = \eta \circ e \circ u$.

- *If T is BC then $\to$ is transitive.*

Proof. See [8, Proposition 4.1] $\square$

We now introduce a couple of examples of monads on Set and look at the partial evaluations that are induced by them.

Example 4.12 ([8])**.** *Let G be a monoid in Set and consider the $(G \times -, \mu, \eta)$ monad defined as follows:*

- *Given $X \in \mathsf{Set}$, $G \times X$ is the Cartesian product of G and X.*
- *Given $f : X \to Y$, $G \times f : G \times X \to G \times Y$ is defined by*

$$G \times f = 1_G \times f. \tag{4.1}$$

- *The multiplication is defined by*

$$\mu_X(g, h, x) = (gh, x) \tag{4.2}$$

for $X \in \mathsf{Set}$ and $(g, h, x) \in G \times G \times X$.

- *The unit is defined by*
$$\eta_X(x) = (1, x) \tag{4.3}$$
where 1 *is the unit of the monoid.*

A $G\times -$*-algebra is then the same thing as a set* X *together with a* G *action* $G\times X \to X$*.*

Now, let (g, x) *and* (h, y) *be elements of* $G \times X$*. Then* (h, y) *is a partial evaluation of* (g, x) *if there is an element* (h, ℓ, x) *such that* $h\ell = g$ *and* $\ell x = y$*. Loosely speaking this means that* (h, y) *is further along in the orbit compared to* (g, x)*. One can show that this monad is BC and hence the partial evaluation order is transitive. Loosely speaking this is the trivial statement that if* (k, z) *is further along in the orbit than* (ℓ, y) *and* (ℓ, y) *is further along than* (g, x) *then* (k, z) *is further along than* (g, x) *too.*

Example 4.13. *Consider the* ***list monad*** $(\mathsf{List}, \boldsymbol{join}, \boldsymbol{ret})$ *defined as follows*

- *Given* $S \in \mathsf{Set}$, $\mathsf{List}(S) = \coprod_{n\in\mathbb{N}} S^{\times n}$*. Thus, a list is a sequence* $(a_1, \ldots, a_n)$ *for some* $n \geq 0$ *and* $a_1, \ldots, a_n \in S$ *and where the empty list* $()$ *corresponds to* $n = 0$*.*

- *Given* $f : X \to Y$, $\mathsf{List}\, f : \mathsf{List}\, X \to \mathsf{List}\, Y$ *is the pushforward of lists, i.e.,*
$$\mathsf{List}\, f = \coprod_{n\in\mathbb{N}} f^{\times n} \tag{4.4}$$
so that
$$\mathsf{List}\, f((a_1, \ldots, a_n)) = (f(a_1), \ldots, f(a_n)). \tag{4.5}$$

- *The multiplication reduces lists of lists to lists by concatenation*
$$\boldsymbol{join}_S\left(\begin{pmatrix}(a_{11}, \ldots, a_{1n_1}),\\ \vdots \\ (a_{k1}, \ldots, a_{kn_k}),\end{pmatrix}\right) = (a_{11}, \ldots, a_{1n_1}, \ldots, a_{k1}, \ldots, a_{kn_k}). \tag{4.6}$$

- *The unit sends an element* $s \in S$ *to the list containing that element, i.e.,*
$$\boldsymbol{ret}_S(s) = (s). \tag{4.7}$$

A List*-algebra is then a map* $e : \mathsf{List}\, S \to S$ *which is the same thing as specifying a monoid structure on* S *(the* List*-algebra axioms gives the necessary monoid axioms). Thus, a partial evaluation from a list* $(a_1, \ldots, a_n)$ *to a list* $(b_1, \ldots, b_m)$ *corresponds to a list of lists*
$$\begin{pmatrix}(c_{11}, \ldots, c_{1n_1}),\\ \vdots \\ (c_{k1}, \ldots, c_{kn_k}),\end{pmatrix}$$

such that

$$(a_1, \ldots, a_n) = (c_{11}, \ldots, c_{1n_1}, \ldots, c_{k1}, \ldots, c_{kn_k}) \tag{4.8}$$

and

$$(b_1, \ldots, b_m) = (c_{11} \ldots c_{1n_1}, \ldots, c_{k1} \ldots c_{kn_k}). \tag{4.9}$$

In other words, it corresponds to partitioning $(a_1, \ldots, a_n)$ into sublists, evaluating these sublists, and then putting the result back together. From this intuitive description it should also be clear that partial evaluation is transitive for List*. We leave the details of this to the interested reader.*

4.2 Partial Evaluations in Representable Markov Categories

Every representable Markov category C gives rise to an affine symmetric monoidal monad on $\mathsf{C}_{\mathrm{det}}$ through Theorem 3.41. It is therefore a valid question to ask what the partial evaluation relation of this monad looks like. Thus, suppose we have a P-algebra $e : PA \to A$ in $\mathsf{C}_{\mathrm{det}}$. Using that

$$\mathsf{C}_{\mathrm{det}}(\Theta, PA) \cong \mathsf{C}(\Theta, A) \tag{4.10}$$

we see that the partial evaluation relation correspond to a certain relation on $\mathsf{C}(\Theta, A)$. More specifically, if $p, q : \Theta \to A$ in C then $p \leq q$ in the relation on $\mathsf{C}(\Theta, A)$ means that there exists $k : \Theta \to PA$ such that the diagram

$$\begin{tikzcd} & \Theta \arrow[dl, "p"'] \arrow[d, "k"] \arrow[dr, "q"] & \\ A & PA \arrow[l, "\mathtt{samp}_A"] \arrow[r, "e"'] & A \end{tikzcd} \tag{4.11}$$

is commutative in C. We have the following result.

Proposition 4.14. *The isomorphism*

$$\mathsf{C}_{det}(\Theta, PA) \cong \mathsf{C}(\Theta, A) \tag{4.12}$$

is monotone in both directions where the order on $\mathsf{C}_{det}(\Theta, PA)$ is given by the partial evaluation order and the order on $\mathsf{C}(\Theta, A)$ is the one described above.

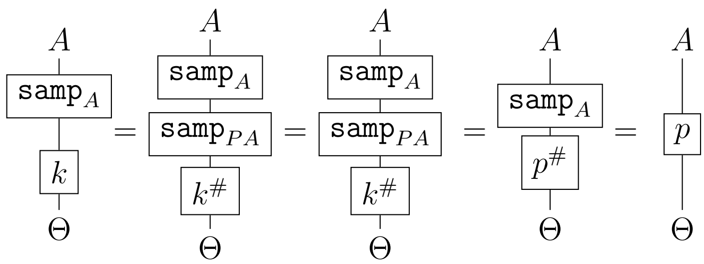
Proof. We must show that $\mathtt{samp}\circ -$ and $(-)^\#$ are order preserving. Thus, let $p^\#, q^\# \in PA$ and assume $p^\#$ can be partially evaluated to $q^\#$. We want to show that this implies that $p \le q$, i.e., that there exists $k \in PA$ such that

$$\begin{tikzcd} & \Theta \arrow[dl, "p"'] \arrow[d, "k"] \arrow[dr, "q"] & \\ A & PA \arrow[l, "\mathtt{samp}_A"] \arrow[r, "e"'] & A \end{tikzcd} \tag{4.13}$$

is commutative. By assumption there must exist a $k^\# \in PPA$ such that

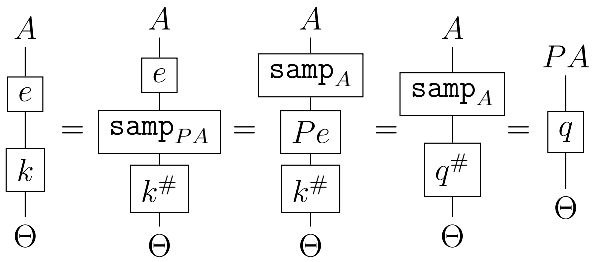
$$\begin{tikzcd} & \Theta \arrow[dl, "p^\#"'] \arrow[d, "k^\#"] \arrow[dr, "q^\#"] & \\ PA & PPA \arrow[l, "P\mathtt{samp}_A"] \arrow[r, "Pe"'] & PA \end{tikzcd} \tag{4.14}$$

is commutative. We let this be our candidate k and have that



(4.15)

as desired. We also have



(4.16)

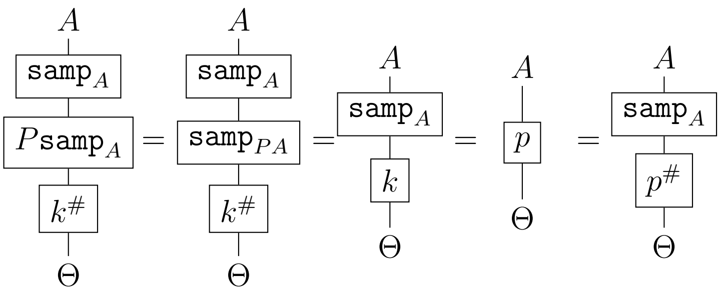
as desired. Thus the diagram in (4.13) is commutative.

Conversely, suppose that $p \le q$. Let $k^\#$ be the deterministic counterpart to the k in

$$\begin{tikzcd} & \Theta \arrow[dl, "p"'] \arrow[d, "k"] \arrow[dr, "q"] & \\ A & PA \arrow[l, "\mathtt{samp}_A"] \arrow[r, "e"'] & A \end{tikzcd} \tag{4.17}$$

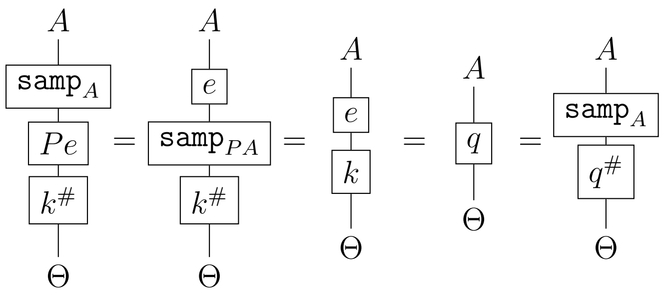
which makes everything commute. Now, remember that $\mathtt{samp}_A \circ - : \mathsf{C}_{\mathrm{det}}(\Theta, PA) \to$

$\mathsf{C}(\Theta, A)$ is a bijection. Thus, from



(4.18)

it follows that $P\mathtt{samp}_A \circ k^\# = p^\#$ as desired. Similarly



(4.19)

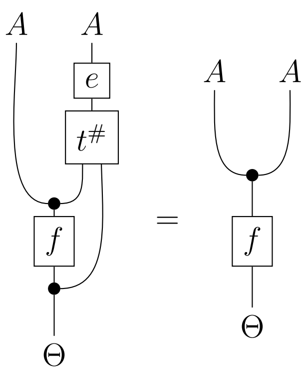
which implies $Pe \circ k^\# = q^\#$ as desired. □

Definition 4.15. *The relation on* $\mathsf{C}(\Theta, X)$ *described above is called the* ***second-order dominance relation****.*

Remark 4.16. *Putting* $\Theta = I$ *in* $\mathsf{BorelStoch}$ *restores the usual notion of stochastic dominance. For the details of this we refer the reader to [7] and also [17, Chapter 4].*

Now, the second-order dominance relation is closely related to the concept of dilations.

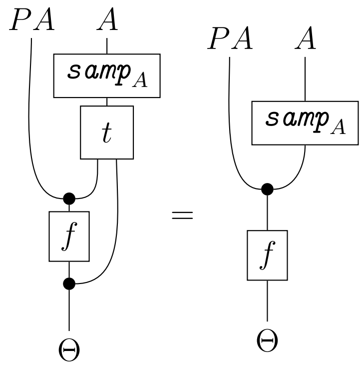
Definition 4.17 ([7])**.** *Let* C *be a representable Markov category,* $e : PA \to A$ *an algebra of the monad* $(P, P\mathtt{samp}, \delta)$ *on* C_{det}*, and* $f : \Theta \to A$ *a morphism in* C*. We say that a morphism* $t : A \otimes \Theta \to A$ *is an* f***-dilation*** *(with respect to the* P*-algebra* e*) if the following equality holds for* t



(4.20)

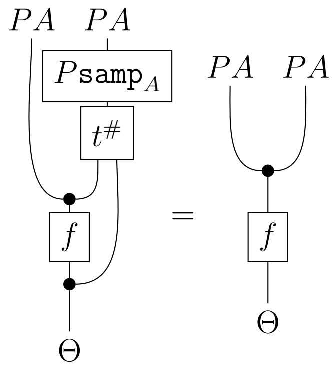
When showing the relationship between second-order stochastic dominance and dilations it is helpful to assume that the category is not only representable, but also a.s.-compatibly representable.

Lemma 4.18 ([7])**.** *Let C be an a.s.-compatibly representable Markov category. Consider the free algebra $P\mathtt{samp}_X : PPA \to PA$ of the monad $(P, P\mathtt{samp}, \delta)$ on C_{det} and a morphism $f : \Theta \to PA$. Then, a morphism $t : A \otimes \Theta \to PA$ is an f-dilation if and only if the following equality holds*



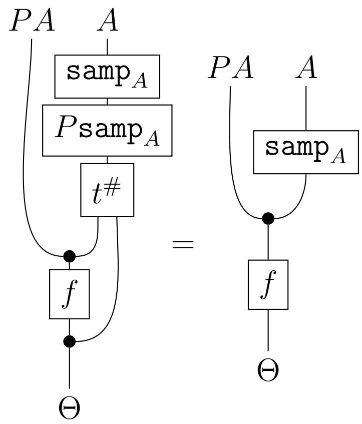
(4.21)

Proof. Assume t is an f-dilation. We then have that



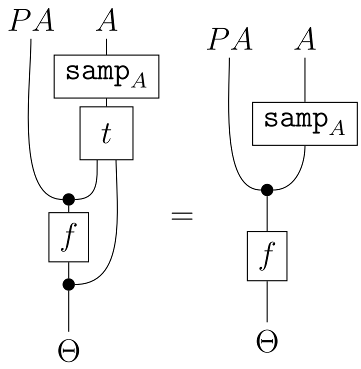
(4.22)

Then, using the sampling cancellation property (we can do this as $P\mathtt{samp}_A$ is deterministic) we have that the above equation holds if and only if



(4.23)

and since $P\mathtt{samp}_A \circ \mathtt{samp}_{PA} = \mathtt{samp}_A \circ \mathtt{samp}_{PA}$ this holds if and only if



(4.24)

as desired. ☐

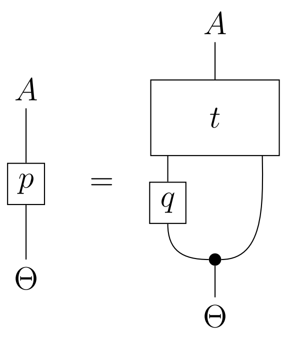
Proposition 4.19. *[7] Let C be an a.s.-compatibly representable Markov category which has conditionals. Given any P-algebra $e : PA \to A$ in C_{det} and any morphisms $p, q : \Theta \to A$ in C, the following conditions are equivalent*

(i) The inequality $p \leq q$ holds in $\mathsf{C}(\Theta, A)$, i.e., there exists a morphism $k : \Theta \to PA$ such that

$$\begin{tikzcd} & \Theta \arrow[dl, "p"'] \arrow[d, "k"] \arrow[dr, "q"] & \\ A & PA \arrow[l, "\mathtt{samp}_A"] \arrow[r, "e"'] & A \end{tikzcd} \tag{4.25}$$

commutes.

(ii) There exist a q-dilation $t : A \otimes \Theta \to A$ which converts q to p. That is to say, the following equality holds



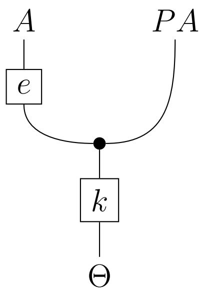
(4.26)

Proof. This is an adaptation of a proof from [7].

(i) $\implies$ (ii): Let $k \in PA$ be such that

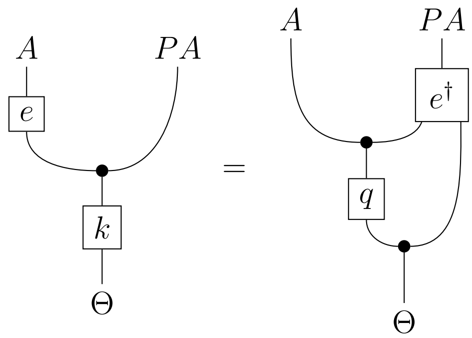
$$\begin{tikzcd} & \Theta \arrow[dl, "p"'] \arrow[d, "k"] \arrow[dr, "q"] & \\ A & PA \arrow[l, "\mathtt{samp}_A"] \arrow[r, "e"'] & A \end{tikzcd} \tag{4.27}$$

commutes. Let $t = \texttt{samp}_A \circ e^\dagger$ where $e^\dagger$ is the Bayesian inverse of e in



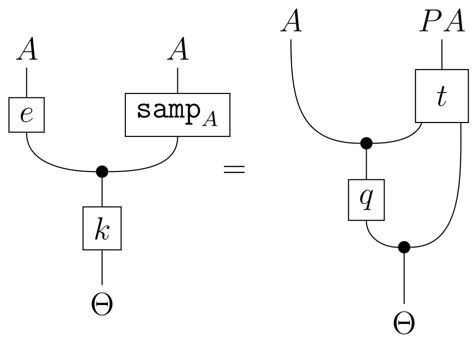
(4.28)

Then, the defining equation of the Bayesian inverse gives



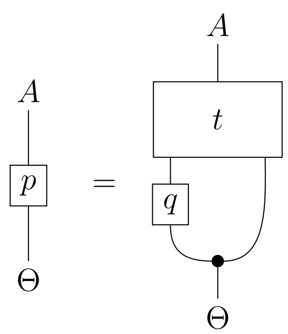
(4.29)

from which it follows that



(4.30)

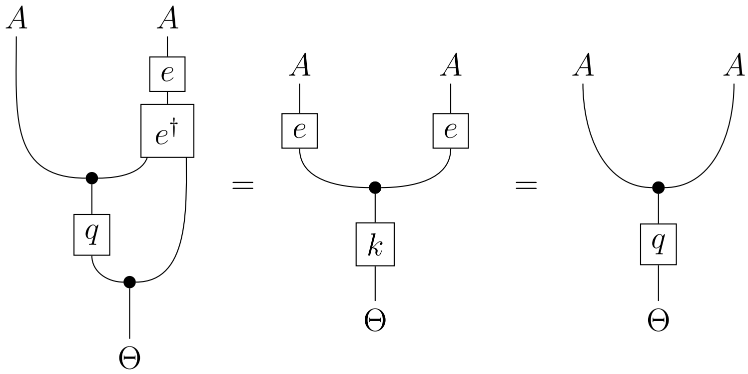
Marginalization then gives



(4.31)

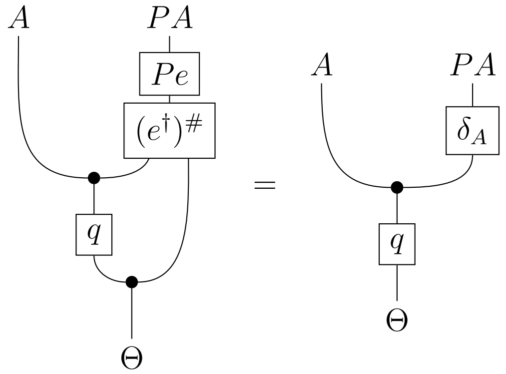
showing Equation 4.26. Now, to see that t is a q-dilation, note that Equation 4.29

gives



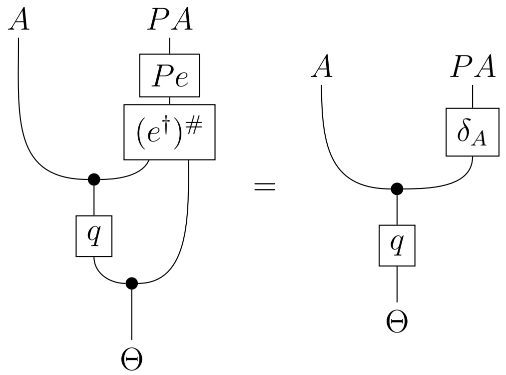
(4.32)

where we used the assumption that e is deterministic. Then, applying the sampling cancellation property gives



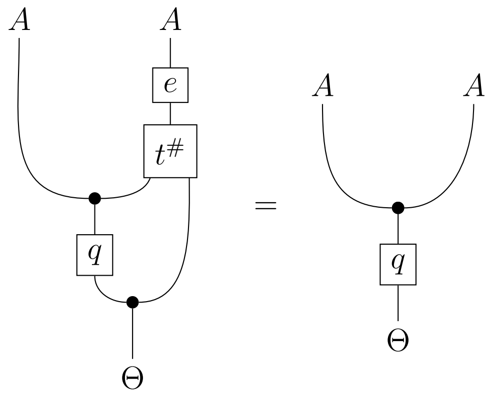
(4.33)

Keeping in mind that $t^\# = P\mathtt{samp}_A \circ (e^\dagger)^\#$ we have



(4.34)

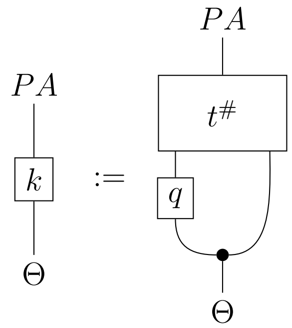
Applying $1_A \otimes e$ to both sides and keeping in mind that e is a P-algebra then gives



(4.35)

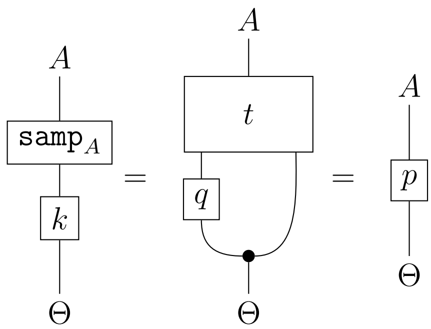
showing that t is a q-dilation.

(ii) $\implies$ (i): Let $t : A \otimes \Theta \to A$ be a q-dilation such that Equation 4.26 holds. Define $k \in PA$ by



(4.36)

Applying $\mathtt{samp}_A$ then gives



(4.37)

where the second equalities follow from the fact that t converts q to p. On the other hand, applying e to Equation 4.36 gives $e \circ k = q$ by the assumption that t is a q-dilation with respect to e. Thus

$$\begin{tikzcd} & \Theta \arrow[dl, "p"'] \arrow[d, "k"] \arrow[dr, "q"] & \\ A & PA \arrow[l, "\mathtt{samp}_A"] \arrow[r, "e"'] & A \end{tikzcd} \tag{4.38}$$

commutes as desired. □

Summarizing we have the following three-way equivalence between partial evaluations, second-order stochastic dominance and dilations.

Corollary 4.20. *Let C be an a.s.-compatibly representable Markov category with conditionals. Given $p, q \in \mathsf{C}(\Theta, A)$, the following are equivalent:*

(i) $p \leq q$ in the pre-order on $\mathsf{C}(\Theta, A)$;

(ii) there exists a partial evaluation from $p^{\#}$ to $q^{\#}$;

(iii) there exists a q-dilation $t : A \otimes \Theta \to A$ which converts q to p.

Chapter 5

Relationship Between Conditionals and BC Monads

In this final chapter we relate the existence of conditionals with the BC property (introduced in Definition 4.10) on monads. More specifically, we show how the underlying monad of an a.s.-compatibly representable Markov category C that has conditionals, and satisfies something known as the equalizer principle, is BC. This lets us conclude that the second-order stochastic dominance relation is transitive in such Markov categories and gives us a functor

$$\mathsf{C}(\Theta, -) : C^P_{\text{det}} \to \mathsf{Ord} \tag{5.1}$$

for all $\Theta \in \mathsf{C}$.

5.1 Conditionals Implies BC

Before stating the main theorem we need some technical lemmas and terminology.

Definition 5.1 ([6]). *A Markov category* C *is said to satisfy the* ***equalizer principle*** *if:*

(i) Equalizers in C_{det} *exist.*

(ii) For every equalizer diagram

$$E \xrightarrow{\mathsf{eq}} X \underset{g}{\overset{f}{\rightrightarrows}} Y \tag{5.2}$$

in C_{det}*, every* $p : A \to X$ *in* C *satisfying* $f =_p g$ *factors uniquely across* eq*.*

Lemma 5.2 ([6]). $\mathsf{BorelStoch}$ *satisfies the equalizer principle.*

Proof. See [6, Proposition 3.5.4]. $\square$

Lemma 5.3. *Let $\mathcal{C}$ be a category with pullbacks. Suppose $f : X \to Z$ and $g : Y \to Z$ are two morphisms in $\mathcal{C}$. With*

$$\begin{tikzcd} X \times_Z Y \arrow[r, "g^*f"] \arrow[d, "f^*g"'] \arrow[dr, phantom, "\lrcorner", very near start] & Y \arrow[d, "g"] \\ X \arrow[r, "f"'] & Z \end{tikzcd}$$

being the corresponding pullback diagram we have that the following diagram is an equalizer diagram

$$X \times_Z Y \xrightarrow{(f^*g, g^*f)} X \times Y \underset{g \circ \pi_2}{\overset{f \circ \pi_1}{\rightrightarrows}} Z.$$

where π_1 and π_2 are the projection maps $X \times Y \to X$ and $X \times Y \to Y$ respectively.

Proof. Let $(p, q) : A \to X \times Y$ be a map such that

$$f \circ p = g \circ q. \tag{5.3}$$

Using the universal property of the pullback we then have a unique map $r : A \to X \times_Z Y$ such that

$$f^*g \circ r = p \tag{5.4}$$

and

$$g^*f \circ r = q. \tag{5.5}$$

However, this means that the following diagram is commutative

$$\begin{tikzcd} X \times_Z Y \arrow[r, "{(f^*g,g^*f)}"] & X \times Y \arrow[r, shift left, "f\circ\pi_1"] \arrow[r, shift right, "g\circ\pi_2"'] & Z \\ A. \arrow[u, dashed, "r"] \arrow[ur, "{(p,q)}"'] & & \end{tikzcd}$$

Moreover, having another map $r' : A \to X \times_Z Y$ such that

$$\begin{tikzcd} X \times_Z Y \arrow[r, "{(f^*g,g^*f)}"] & X \times Y \arrow[r, shift left, "f\circ\pi_1"] \arrow[r, shift right, "g\circ\pi_2"'] & Z \\ A. \arrow[u, dashed, "r'"] \arrow[ur, "{(p,q)}"'] & & \end{tikzcd}$$

is commutative, we must have that $r' = r$ by the uniqueness property of the pullback. Thus, we have shown that $X \times_Z Y$ is the equalizer of $f \circ \pi_1 : X \times Y \to Z$ and $g \circ \pi_2 : X \times Y \to Z$. $\square$

Lemma 5.4. *Let* C *be a category with pullbacks and* (P, μ, δ) *a monad on* C*. If* P *turns pullbacks into weak pullbacks then* P *preserves weak pullbacks.*

Proof. Consider a weak pullback diagram

$$\begin{tikzcd} W \arrow[r, "g^*f"] \arrow[d, "f^*g"'] & Y \arrow[d, "g"] \\ X \arrow[r, "f"'] & Z \end{tikzcd} \tag{5.6}$$

in C. We want to show that

$$\begin{tikzcd} PW \arrow[r, "P(g^*f)"] \arrow[d, "Pf({}^*g)"'] & PY \arrow[d, "Pg"] \\ PX \arrow[r, "Pf"'] & PZ \end{tikzcd} \tag{5.7}$$

is a weak pullback. Thus, suppose we have maps $p : A \to PX$ and $q : A \to PY$ such that

$$\begin{tikzcd} A \arrow[drr, bend left, "q"] \arrow[ddr, bend right, "p"'] & & \\ & PW \arrow[r, "P(g^*f)"] \arrow[d, "P(f^*g)"'] & PY \arrow[d, "Pg"] \\ & PX \arrow[r, "Pf"'] & PZ \end{tikzcd} \tag{5.8}$$

commutes. By assumption we have that

$$\begin{tikzcd} P(X \times_Z Y) \arrow[r, "P\pi_2"] \arrow[d, "P\pi_1"'] & PY \arrow[d, "Pg"] \\ PX \arrow[r, "Pf"'] & PZ \end{tikzcd} \tag{5.9}$$

is a weak pullback. Thus there exists $\tilde{r} : A \to P(X \times_Z Y)$ such that

$$\begin{tikzcd} A \arrow[drr, bend left, "q"] \arrow[ddr, bend right, "p"'] \arrow[dr, dashed, "\tilde{r}"] & & \\ & P(X \times_Z Y) \arrow[r, "P\pi_2"] \arrow[d, "P\pi_1"'] & PY \arrow[d, "Pg"] \\ & PX \arrow[r, "Pf"'] & PZ \end{tikzcd} \tag{5.10}$$

commutes. Now, there is also a canonical map $s : X \times_Z Y \to W$ such that

$$
\begin{tikzcd}
A \arrow[dr, "\tilde{r}"] \arrow[drrr, bend left, "q"] \arrow[ddrr, bend right, "p"'] & & & \\
& P(X \times_Z Y) \arrow[dr, "Ps"] \arrow[urr, phantom] \arrow[rr, bend left, "P\pi_2", to=3-4] \arrow[ddr, bend right, "P\pi_1"', to=4-3] & & \\
& & PW \arrow[r, "f^*g"] \arrow[d, "g^*f"'] & PY \arrow[d, "Pg"] \\
& & PX \arrow[r, "Pf"'] & PZ
\end{tikzcd}
\tag{5.11}
$$

commutes. Defining $r := Ps \circ \tilde{r}$ then gives the desired map making

$$
\begin{tikzcd}
A \arrow[dr, dashed, "r"] \arrow[drr, bend left, "q"] \arrow[ddr, bend right, "p"'] & & \\
& PW \arrow[r, "f^*g"] \arrow[d, "g^*f"'] & PY \arrow[d, "Pg"] \\
& PX \arrow[r, "Pf"'] & PZ
\end{tikzcd}
\tag{5.12}
$$

commute. $\square$

Lemma 5.5. *Let* C *be a representable Markov category. If* C *has conditionals then* μ *of the underlying monad* (P, μ, δ) *on* C_{det} *is weakly Cartesian.*

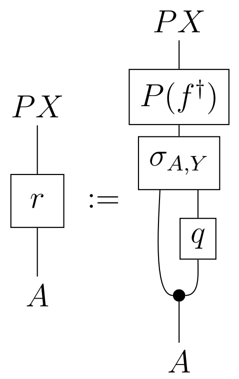
Proof. Let $f : X \to Y$, $p^\# : A \to PX$, and $q^\# : A \to PPY$ be morphisms in C_{det} such that the following diagram is commutative.

$$
\begin{tikzcd}
A \arrow[drr, bend left, "q^\#"] \arrow[ddr, bend right, "p^\#"'] & & \\
& PPX \arrow[r, "PPf"] \arrow[d, "\mu"'] & PPY \arrow[d, "\mu"] \\
& PX \arrow[r, "Pf"'] & PY.
\end{tikzcd}
$$

We need to find a map $r^\# : A \to PPX$ in C_{det} such that

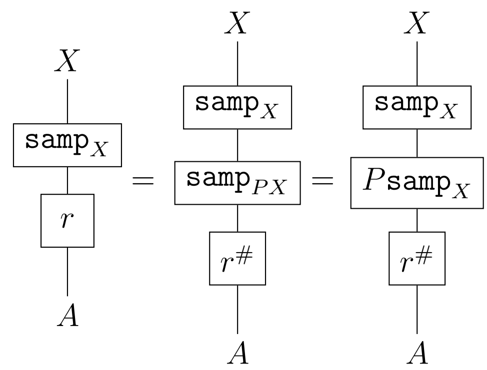
$$
\begin{tikzcd}
A \arrow[dr, dashed, "r^\#"] \arrow[drr, bend left, "q^\#"] \arrow[ddr, bend right, "p^\#"'] & & \\
& PPX \arrow[r, "PPf"] \arrow[d, "\mu"'] & PPY \arrow[d, "\mu"] \\
& PX \arrow[r, "Pf"'] & PY
\end{tikzcd}
$$

is commutative. Consider therefore the following morphism in C



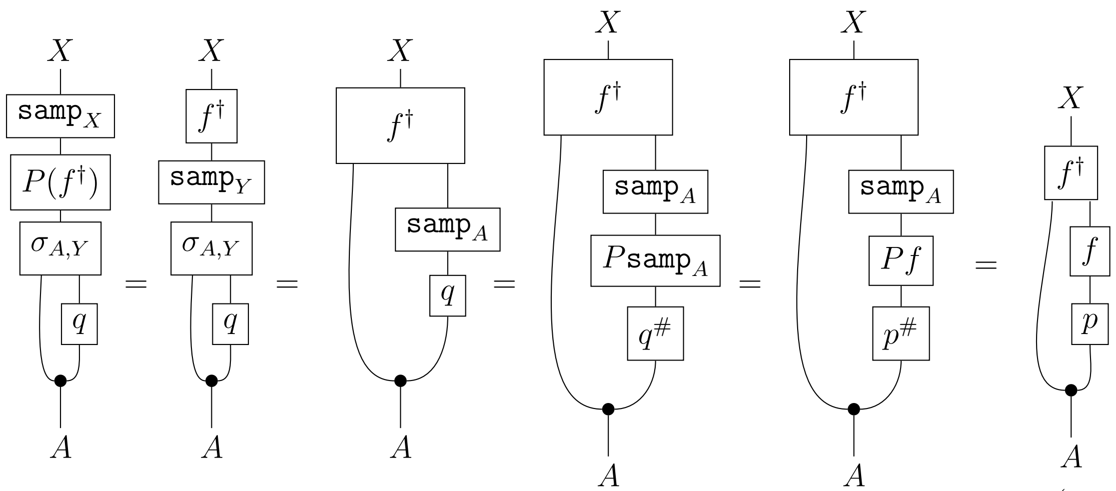
(5.13)

where $f^\dagger$ is the Bayesian inverse of f with respect to p and $\sigma_{A,Y} : A \otimes PY \to P(A \otimes Y)$ is the left strength of P. Now, applying $\mathtt{samp}_X$ to the left hand side of Equation 5.13 we have that



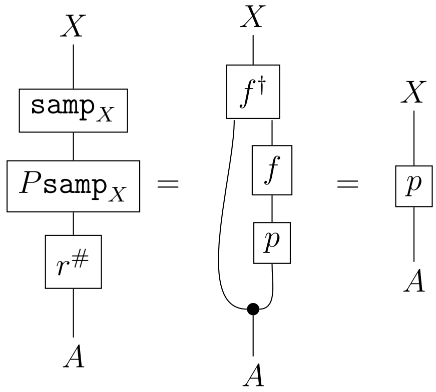
(5.14)

Similarly, applying $\mathtt{samp}_X$ to the right hand side gives



(5.15)

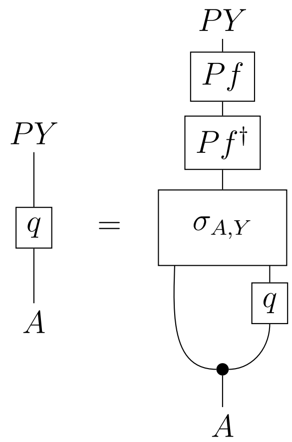
Combining this with Lemma gives



(5.16)

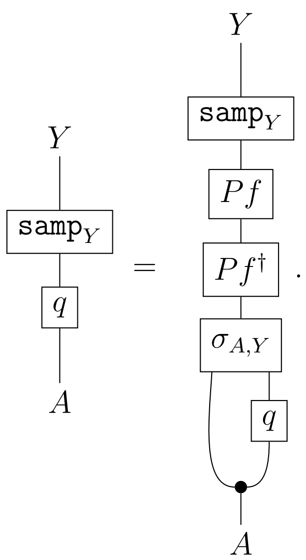
Applying the sampling cancellation property then gives that $p^{\#} = P\texttt{samp} \circ r^{\#}$ which shows commutativity of the left triangle (remember that $\mu = P\texttt{samp}$).

For commutativity of the right triangle, the sampling cancellation property implies that it is sufficient to show that



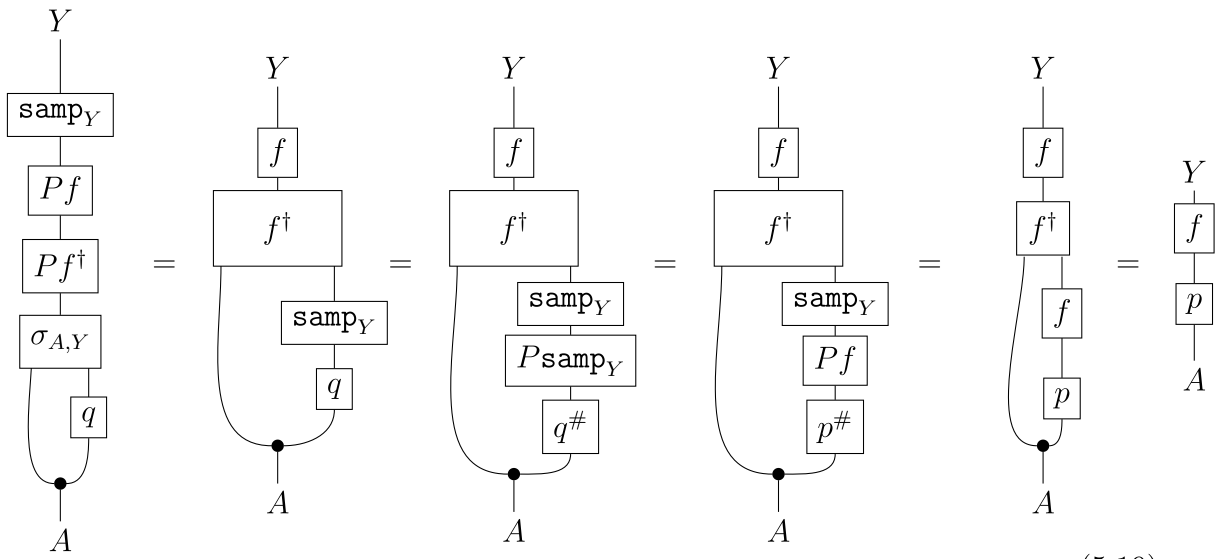
(5.17)

Now, by a second usage of the sampling cancellation property (and marginalization) it suffices to show that.



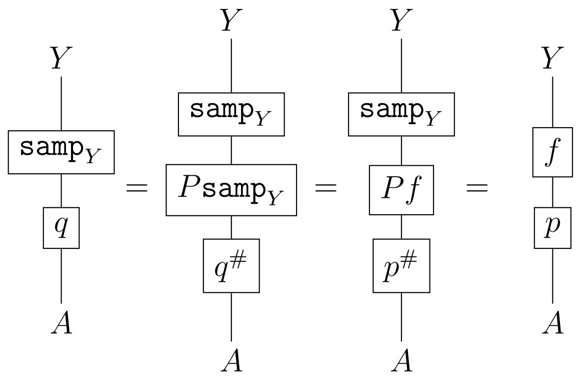
(5.18)

The right hand side is equal to



(5.19)

while the left hand side gives



(5.20)

showing that Equation 5.17 holds and hence μ is weakly Cartesian. □

Theorem 5.6. *Let* C *be an a.s-compatibly representable Markov category with pullbacks. If* C *satisfies the equalizer principle and has conditionals, then the underlying monad* (P, μ, δ) *on* C_{det} *is BC.*

Proof. Using Lemma 5.5 we know that μ is weakly Cartesian. Hence all that remains to show is that P preserves weak pullbacks.

By Lemma 5.4 it suffices to show that P turns pullbacks into weak pullbacks. Thus, suppose we have a pullback

$$\begin{tikzcd} X \times_Z Y \arrow[r, "g^*f"] \arrow[d, "f^*g"'] & Y \arrow[d, "g"] \\ X \arrow[r, "f"'] & Z. \end{tikzcd}$$

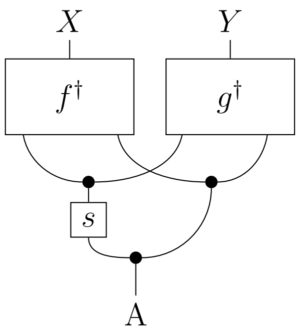
in $\mathsf{C}_{\mathrm{det}}$. We want to show that

$$\begin{tikzcd} P(X \times_Z Y) \arrow[r, "P(g^* f)"] \arrow[d, "P(f^* g)"'] & PY \arrow[d, "Pg"] \\ PX \arrow[r, "Pf"'] & PZ. \end{tikzcd}$$

is a weak pullback in $\mathsf{C}_{\mathrm{det}}$. Suppose therefore that we have maps $p^\# : A \to PX$ and $q^\# : A \to PY$ such that the following diagram is commutative

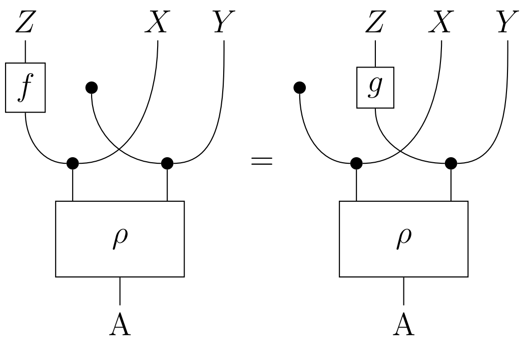
$$\begin{tikzcd} A \arrow[drr, bend left, "q^\#"] \arrow[ddr, bend right, "p^\#"'] & & \\ & P(X \times_Z Y) \arrow[r, "P(g^* f)"] \arrow[d, "P(f^* g)"'] & PY \arrow[d, "Pg"] \\ & PX \arrow[r, "Pf"'] & PZ. \end{tikzcd}$$

Now, define $\rho : A \to X \times Y$ by

$\rho :=$ 

(5.21)

where s is the common composition $f \circ p = g \circ q$ in $\mathcal{C}$. The equalizer principle and Lemma 5.3 tells us that if

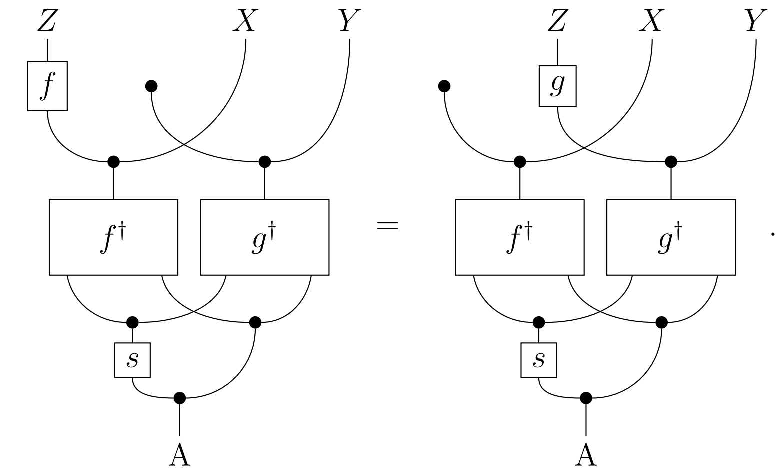


(5.22)

then there exists a map $r : A \to X \times_Z Y$ in $\mathcal{C}$ such that

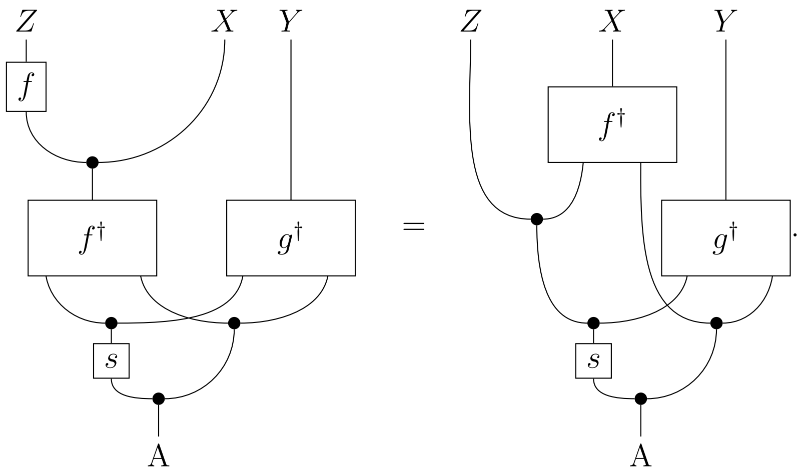
$$\rho = (f^* g, g^* f) \circ r \tag{5.23}$$

in $\mathcal{C}$, which is exactly what we need. Expanding ρ in Equation 5.22 we get the requirement



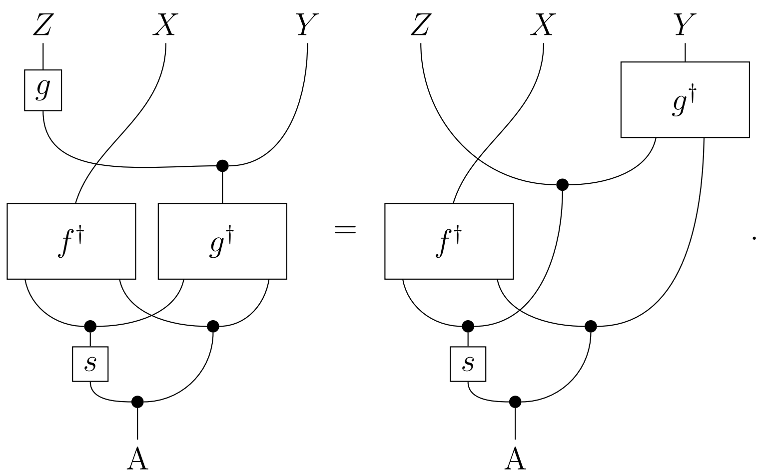
(5.24)

Now, using relative positivity of $\mathcal{C}$ and the fact that $f \circ f^\dagger$ and $g \circ g^\dagger$ are s-a.s. deterministic, we have that



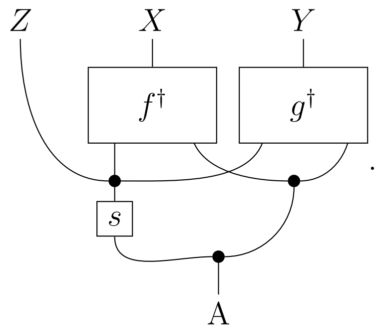
(5.25)

Similarly,



(5.26)

Using the associativity of the copy map, the right hand side in Equation 5.25 becomes



. (5.27)

Doing the same for the right hand side in Equation 5.26 also gives the exact same result and hence we can conclude that the equality in Equation 5.22 indeed holds. Thus, using the equalizer principle there must exist a map $r : A \to X \times_Z Y$ in $\mathcal{C}$ such that

$$\rho = (f^* g, g^* f) \circ r \tag{5.28}$$

in $\mathcal{C}$. The map $\hat{r}^\# : A \to P(X \times_Z Y)$ in $\mathcal{C}_{\text{det}}$ is therefore the desired map which makes

$$\begin{tikzcd}
A \arrow[dr, dotted, "r^\#"] \arrow[drr, bend left, "q^\#"] \arrow[ddr, bend right, "p^\#"'] & & \\
& P(X \times_Z Y) \arrow[r, "P(g^* f)"] \arrow[d, "P(f^* g)"'] & PY \arrow[d, "Pg"] \\
& PX \arrow[r, "Pf"'] & PZ
\end{tikzcd}$$

into a commutative diagram. □

As far as we know the following result is new.

Corollary 5.7. *The Giry monad on* $\mathsf{BorelStoch}$ *is BC.*

Proof. Follows immediately from Lemma 5.2 and Theorem 5.6. □

Remark 5.8. *There exists at least one paper [19] which claims the opposite of the above corollary. However, in [19] the author confuses equalizers for pullbacks and wrongly states that the pullback is empty.*

In light of Proposition 4.11 we also have the following property on the second-order stochastic dominance relation.

Corollary 5.9. *Let* C *be an a.s.-compatibly representable Markov category which has conditionals and satisfies the equalizer principle. Then the second-order dominance relation on* $\mathsf{C}(\Theta, X)$ *is transitive for all* $\Theta \in \mathsf{C}$ *and* $X \in \mathsf{C}^P_{det}$.

Proof. Follows immediately by Proposition 4.11, 4.20, and Theorem 5.6. ☐

In particular, this means that $\mathsf{C}(\Theta, X)$ is a pre-order and that the assignment

$$\begin{aligned} \mathsf{C}(\Theta, -) : \mathsf{C}^P_{\text{det}} &\to \mathsf{Ord} \\ (A, e) &\mapsto \mathsf{C}(\Theta, A) \\ f : (A, e) \to (B, d) &\mapsto f \circ - : \mathsf{C}(\Theta, A) \to \mathsf{C}(\Theta, B) \end{aligned}$$

is a functor. The only non-trivial thing to show is that $f \circ -$ is monotone. However, this is a straight forward proof and we leave the details to the interested reader.

As far as we know this functor has not been studied in great detail and we do not know what properties it possesses. A particular problem which should have a definite yes or no answer is the following.

Problem 5.10. *Is the functor* $\mathsf{C}(\Theta, -) : \mathsf{C}^P_{det} \to \mathsf{Ord}$ *a Grothendieck bifibration?*

Bibliography

[1] Julian Asilis. Probability monads, March 2020.

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

[3] Bob Coecke, Tobias Fritz, and Robert W. Spekkens. A mathematical theory of resources. *Information and Computation*, 250:59–86, October 2016.

[4] Carmen Constantin, Paolo Perrone, Tobias Fritz, and Brandon T. Shapiro. Partial evaluations and the compositional structure of the bar construction. Theory Appl. Categ. 39(11), 322–364 (2023).

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

[6] Tobias Fritz, Tomáš Gonda, Antonio Lorenzin, Paolo Perrone, and Dario Stein. Absolute continuity, supports and idempotent splitting in categorical probability.

[7] Tobias Fritz, Tomáš Gonda, Paolo Perrone, and Eigil Fjeldgren Rischel. Representable markov categories and comparison of statistical experiments in categorical probability. *Theoretical Computer Science*, 961:113896.

[8] Tobias Fritz and Paolo Perrone. Monads, partial evaluations, and rewriting.

[9] P. V. Golubtsov and S. S. Moskaliuk. Method of additional structures on the objects of a monoidal kleisli category as a background for information transformers theory, 2002.

[10] Peter Golubtsov. Axiomatic description of categories of information transformers. *Problems of Information Transmission*, 35:259–274, 1999.

[11] Peter Golubtsov. Information transformers: Category-theoretical structure, informativeness, decision-making problems. 2004.

[12] Martin Hyland and John Power. The category theoretic understanding of universal algebra: Lawvere theories and monads. *Electronic Notes in Theoretical Computer Science*, 172:437–458.

[13] Saunders MacLane. *Categories for the Working Mathematician*. Springer New York, NY, September 1998.

[14] Sean Moss and Paolo Perrone. Probability monads with submonads of deterministic states. In *Proceedings of the 37th Annual ACM/IEEE Symposium on Logic in Computer Science*, LICS '22. ACM, August 2022.

[15] nLab authors. sober measurable space. `http://ncatlab.org/nlab/show/sober+measurable+space`, July 2024. revision 6.

[16] nLab authors. standard borel space. `http://ncatlab.org/nlab/show/standard+Borel+space`, March 2024. revision 4.

[17] Paolo Perrone. *Categorical Probability and Stochastic Dominance in Metric Spaces*. PhD thesis, University of Leipzig, July 2018.

[18] Emily Riehl. *Category Theory in Context*. Dover Publications.

[19] Ignacio D. Viglizzo. Final sequences and final coalgebras for measurable spaces. In José Luiz Fiadeiro, Neil Harman, Markus Roggenbach, and Jan Rutten, editors, *Algebra and Coalgebra in Computer Science*, pages 395–407, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.