A Categorical Basis for Conditional Probability

Bart Jacobs*

Radboud University, Nijmegen, The Netherlands
`bart@cs.ru.nl`

Abstract

This paper identifies several key properties of a monad that allow us to formulate the basics of conditional probability theory, using states for distributions/measures and predicates for events/probability density functions (pdf's). The distribution monad for discrete probability and the Giry monad for continuous probability are leading examples. Our categorical description handles discrete and continuous probability uniformly, and includes: an abstract Fubini theorem with a very simple proof, conditional states satisfying Bayes' rule, and (conditional) non-entwinedness for joint states.

1 Introduction

It is a well-known fact that the very basic aspects of probability theory can be captured via monads. In the early 1980s the Giry monad $\mathcal{G}$ was introduced [8] for *continuous* probability. It acts on the category **Meas** of measurable spaces. The much simpler distribution monad $\mathcal{D}$ on the category of sets can be used for *discrete* probability. Its precise origin is unclear. Just like the powerset monad is used to describe non-deterministic systems, the distribution and Giry monad are used for probabilistic systems, see [3] and [16] for an overview. In such descriptions the Kleisli category $\mathcal{K}\ell(T)$ of the monad T plays a key role, since (non-deterministic / probabilistic / ...) programs are interpreted as maps in this Kleisli category, where the monad T captures the form of computation involved.

Here we extend this categorical description of probability to *conditional* probability. Conditional probability theory deals with probabilities *given* some event or evidence (or predicate, as we shall say). Conditional probability forms the basis for Bayesian analysis and machine learning (see *e.g.* [2]). Conditional reasoning can be quite difficult and un-intuitive, and therefor a clear semantics is important. One of the thoughts underlying this paper — and much other work in category theory — is that a proper abstract presentation makes a topic easier to understand, since it brings the essence to the surface. This abstract presentation is useful as a basis for formal systems for probabilistic reasoning, see [4, 1].

The current work has its origin in an even more abstract theory of effectuses [11, 5] which captures both classical and quantum probability theory. Here we concentrate on the classical 'commutative' case, using a description in terms of monads. Several of the results in this paper also hold in the quantum case, but we ignore this broader perspective. Crucial for our approach is the use of "assert" maps that form the action $X \to X + 1$ associated with a predicate on X. In the present probabilistic setting such actions do not modify the state, unlike in the proper quantum case. Some of the ingredients that are described here have appeared in other places, like [14, 5, 1, 12], but never in a systematic form, dealing with both discrete and continuous (conditional) probability uniformly. The axiomatic categorical approach that we use here resembles to some extent the interface-based approach in Haskell of [19]. But there the emphasis is on computing with conditioning (for Bayesian inference), whereas here we concentrate on generic properties, as part of a 'logic of probability'.

* The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement nr. 320571

Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Among the main achievements we see: the identification of three key properties of monads that allow us to express conditional definitions and results; the abstract analysis of non-entwined states, also in conditional form; the abstract Fubini theorem; the abstract Bayes' rule; basic results about conditional states, like iterated conditioning and marginals.

The paper starts with a gentle introduction to the very basic notions of (discrete) conditional probability theory, as a preparation for the categorical approach. Section 3 describes the properties of monads that are relevant in this setting and also how these properties allow us to describe normalisation and conditioning. Section 4 shows that the non-empty powerset monad $\mathcal{P}_{\bullet}$, the distribution monad $\mathcal{D}$, and the Giry monad $\mathcal{G}$ are instances. Subsequently, Section 5 gives basic results about states (measures/distributions) on product and coproduct carriers $X \times Y$ and $X + Y$. Section 6 concentrates on conditional states, and contains rules for conjunction (Bayes) and iterated conditioning. Finally, Section 7 investigates conditional entwinedness of states.

2 From traditional to categorical probability theory

In this section we sketch how the usual notation and approach in probability theory relate to the categorical approach that we advocate. We do so for the very basic concepts, in the discrete case. Noteworthy is that we generalise events to fuzzy (or soft) predicates, and describe for instance conditional probability for these generalised predicates.

A *discrete distribution* over a 'sample' set A is a weigthed combination of elements of A, where the weights are probabilities from the unit interval $[0,1]$ that add up to 1. Here we only consider finite combinations and write them as:

$$\omega \;=\; r_1|a_1\rangle + \cdots + r_n|a_n\rangle \qquad \text{where} \qquad \begin{cases} a_1,\ldots,a_n \in A \\ r_1,\ldots,r_n \in [0,1] \text{ with } \sum_i r_i = 1. \end{cases} \tag{1}$$

The 'ket' notation $|a\rangle$ is syntactic sugar, used to distinguish elements $a \in A$ from their occurrence in such formal convex sums. We write $\mathcal{D}(A)$ for the set of all such distributions. Distributions are also called states; they express knowledge, in terms of likelihoods of occurrence of elements of A. Notice that such $\omega \in \mathcal{D}(A)$ can be identified with functions $\omega\colon A \to [0,1]$ with finite support $\mathrm{supp}(\omega) = \{a \in A \mid \omega(a) \neq 0\}$ and with $\sum_{a\in A}\omega(a) = 1$. This function-description is often more convenient.

An *event* is a subset $E \subseteq A$ of the sample space. Given a distribution $\omega \in \mathcal{D}(A)$, we can consider the probability of the event E, which is commonly written as $P(E)$ — or as $P_\omega(E)$ to make the distribution ω explicit — and is defined as $P_\omega(E) = \sum_{a\in E}\omega(a)$. These events are traditionally used as predicates on A. We prefer to use a more general 'fuzzy' kind of predicate, namely functions $p\colon A \to [0,1]$. In this discrete case, states (distributions) are predicates, but not the other way around. An event $E \subseteq A$ can be identified with a 'sharp' predicate $A \to [0,1]$, taking values in the subset of booleans $\{0,1\} \subseteq [0,1]$. For such an event E we write $\mathbf{1}_E \in [0,1]^A$ for the associated predicate, given by the indicator function $\mathbf{1}_E$ defined by $\mathbf{1}_E(a) = 1$ if $a \in E$ and $\mathbf{1}_E(a) = 0$ if $a \notin E$.

For a distribution $\omega \in \mathcal{D}(A)$ on a sample space A and a predicate $p \in [0,1]^A$ on the same space we define the validity $\omega \models p$ in $[0,1]$ as:

$$\omega \models p \;\stackrel{\text{def}}{=}\; \textstyle\sum_{a\in A}\omega(a)\cdot p(a) \qquad \text{so that} \qquad \omega \models \mathbf{1}_E \;=\; P_\omega(E). \tag{2}$$

There is a truth predicate $\mathbf{1} = \mathbf{1}_A$ and a falsity predicate $\mathbf{0} = \mathbf{1}_\emptyset$ in $[0,1]^A$, with $\omega \models \mathbf{1} = 1$ and $\omega \models \mathbf{0} = 0$, for each distribution ω. There is also an orthosupplement (negation) $p^\perp$, defined by $p^{\perp}(a) = 1 - p(a)$. Then: $\omega \models p^{\perp} = 1 - (\omega \models p)$. Notice that $(\mathbf{1}_E)^{\perp} = \mathbf{1}_{\neg E}$, wherein $\neg E = \{a \in A \mid a \notin E\}$. The set of predicates $[0,1]^A$ is an *effect module*, see [13, 11, 5]: it has a partial (pointwise) sum $\ovee$, and scalar multiplication $r \cdot p$, for $r \in [0,1]$, satisfying certain properties which are not so relevant here.

We do use an operation $p \,\&\, q$ on $[0,1]^A$, defined by $(p \,\&\, q)(a) = p(a) \cdot q(a)$. This operation is called 'andthen' because it is a form of sequential conjunction. In the current situation it is commutative, but it can be defined more generally and need not be commutative then. It may also be written in modal form as $\langle p? \rangle(q) = p \,\&\, q$. Its De Morgan dual $[p?](q) = \langle p? \rangle(q^{\perp})^{\perp}$ is Reichenbach implication. Notice that for events $E, D \subseteq A$ the corresponding sharp predicates $\mathbf{1}_E, \mathbf{1}_D \in [0,1]^A$ satisfy: $\mathbf{1}_E \,\&\, \mathbf{1}_D = \langle \mathbf{1}_E? \rangle(\mathbf{1}_D) = \mathbf{1}_{E \cap D}$ and $[\mathbf{1}_E?](\mathbf{1}_D) = \mathbf{1}_{E \Rightarrow D}$ where $E \Rightarrow D = \neg E \cup D$ is implication for subsets, as in Boolean algebras.

We call two predicates $p, q \in [0,1]^A$ *independent* wrt. a distribution $\omega \in \mathcal{D}(A)$ if:

$$\big(\omega \models p \,\&\, q\big) \;=\; \big(\omega \models p\big) \cdot \big(\omega \models q\big). \tag{3}$$

Notice that for events $E, D \subseteq A$ this yields the usual definition of independence when applied to the corresponding indicator functions: if $\mathbf{1}_E, \mathbf{1}_D$ are independent according to (3) then:

$$P_\omega(E \cap D) \;=\; \big(\omega \models \mathbf{1}_{E\cap D}\big) \;=\; \big(\omega \models \mathbf{1}_E \,\&\, \mathbf{1}_D\big) \;=\; \big(\omega \models \mathbf{1}_E\big) \cdot \big(\omega \models \mathbf{1}_D\big) \;=\; P_\omega(E) \cdot P_\omega(D).$$

For a predicate $p \in [0,1]^A$ and a distribution $\omega \in \mathcal{D}(A)$ with $\omega \models p \neq 0$ we introduce a *conditional distribution* $\omega|_p \in \mathcal{D}(A)$, pronounced as "$\omega$ given p", and defined as:

$$\omega|_p \;=\; \sum_{a \in A} \frac{\omega(a) \cdot p(a)}{\omega \models p} \big|\, a \,\big\rangle. \tag{4}$$

For another predicate $q \in [0,1]^A$ we use the validity $\omega|_p \models q$ as "q, given p" wrt. distribution ω. This specialises to the usual form of conditional probability[1] for sharp predicates:

$$\begin{aligned}\big(\omega|_{\mathbf{1}_E} \models \mathbf{1}_D\big) \;&=\; \sum_{a \in A} \omega|_{\mathbf{1}_E}(a) \cdot \mathbf{1}_D(a) \;=\; \sum_{a\in A} \frac{\omega(a) \cdot \mathbf{1}_E(a) \cdot \mathbf{1}_D(a)}{\omega \models \mathbf{1}_E} \\ &=\; \frac{\sum_{a \in A} \omega(a) \cdot \mathbf{1}_{E\cap D}(a)}{P_\omega(E)} \;=\; \frac{P_\omega(E \cap D)}{P_\omega(E)} \;=\; P_\omega(D \mid E).\end{aligned}$$

As illustration, consider a distribution $\omega = \frac{1}{4}|a\rangle + \frac{1}{3}|b\rangle + \frac{5}{12}|c\rangle$ on a set $A = \{a,b,c\}$, a subset $E = \{a, c\} \subseteq A$ and a predicate $p \in [0,1]^A$ with $p(a) = \frac{1}{2}, p(b) = \frac{1}{4}, p(c) = 1$. Then:

$$\omega|_{\mathbf{1}_E} \;=\; \tfrac{3}{8}|a\rangle + \tfrac{5}{8}|c\rangle \qquad \omega \models p \;=\; \tfrac{5}{8} \qquad \omega|_p \;=\; \tfrac{1}{5}|a\rangle + \tfrac{2}{15}|b\rangle + \tfrac{2}{3}|c\rangle \qquad \omega|_p \models \mathbf{1}_E \;=\; \tfrac{13}{15}.$$

In the remainder of this article we will provide a firm categorical basis for these (generalised) probabilistic definitions. It will allow us to prove basic properties like: $(\omega|_p)|_q = \omega|_{p\&q}$.

3 Monad assumptions

The constructions of the previous section are possible because the sets of (finite discrete) distributions $\mathcal{D}(A)$ carry the structure of a monad, of a particular kind. This section describes

[1] Apart from here, we avoid the notation $D \mid E$, or $q \mid p$, for conditional probability, because it wrongly suggests that '$|$' is an operation on predicates. Instead, we use '$|$' in $\omega|_p$ describing it as a (right) action of predicates p on distributions ω.

such monads in abstract form. We assume that the reader is already familiar with the notion of monad itself, and with the basic concepts and notation in category theory.

Let $T = (T, \eta, \mu)$ be a monad on a category $\mathbf{C}$. We assume that $\mathbf{C}$ is distributive: it has finite products $(\times, 1)$ and coproducts $(+, 0)$, where products distribute over coproducts. We shall write $\mathcal{K}\ell(T)$ for the Kleisli category of the monad T, with objects as in $\mathbf{C}$, and maps $X \to Y$ in $\mathcal{K}\ell(T)$ given by morphisms $X \to T(Y)$ in $\mathbf{C}$. In order to prevent confusion, we use special notation $\bullet$ for composition in $\mathcal{K}\ell(T)$, given by $g \bullet f = \mu \circ T(g) \circ f$. The category $\mathcal{K}\ell(T)$ inherits coproducts $(+, 0)$ from the underlying category $\mathbf{C}$. Each map $f\colon X \to Y$ in $\mathbf{C}$ gives rise to a 'pure' Kleisli map $\langle f \rangle = \eta \circ f\colon X \to T(Y)$. Then $\langle g \circ f \rangle = \langle g \rangle \bullet \langle f \rangle$.

The monad T is called *strong* if there is a strength natural transformation with components $\mathrm{st}_{X,Y}\colon T(X) \times Y \to T(X \times Y)$ commuting appropriately with the unit η and multiplication μ of the monad, and with the 'monoidal' isomorphism $1 \times X \cong X$ and $X \times (Y \times Z) \cong (X \times Y) \times Z$. Given st, we define $\mathrm{st}'\colon X \times T(Y) \to T(X \times Y)$ as $\mathrm{st}' = T(\gamma) \circ \mathrm{st} \circ \gamma$, where $\gamma = \langle \pi_2, \pi_1 \rangle$ is the swap map. The monad is called *commutative* if the two maps $T(X) \times T(Y) \to T(X \times Y)$, given by $\mu \circ T(\mathrm{st}') \circ \mathrm{st}$ and $\mu \circ T(\mathrm{st}) \circ \mathrm{st}'$ are equal; we then write dst for 'double strength' to denote this (single) map. If T is commutative, then $\mathcal{K}\ell(T)$ is symmetric monoidal, with tensor $\otimes$ given on objects as $\times$ and on arrows $f\colon A \to T(B)$ and $g\colon X \to T(Y)$ as:

$$f \otimes g \stackrel{\text{def}}{=} \Big(A \times X \xrightarrow{f \times g} T(B) \times T(Y) \xrightarrow{\text{dst}} T(B \times Y)\Big).$$

The tensor unit in $\mathcal{K}\ell(T)$ is the final object $1 \in \mathbf{C}$ — which need not be final in $\mathcal{K}\ell(T)$.

A monad is called *strongly affine* [12] if it is strong and for all pairs of objects X, Y the diagram on the right is a pullback. Such monads are affine, meaning that $T(1) \cong 1$. This implies that the final object $1 \in \mathbf{C}$ is also final in the Kleisli category $\mathcal{K}\ell(T)$. Since 1 is the unit for the tensor $\otimes$ — assuming that T is commutative — there are projections $\pi_i\colon X_1 \otimes X_2 \to X_i$ for the tensor, given by $X_1 \otimes X_2 \to X_1 \otimes 1 \cong X_1$ and $X_1 \otimes X_2 \to 1 \otimes X_2 \cong X_2$.

$$\begin{array}{ccc} T(X) \times Y & \xrightarrow{\pi_2} & Y \\ {\scriptstyle \mathrm{st}}\downarrow & & \downarrow{\scriptstyle \eta_Y} \\ T(X \times Y) & \xrightarrow{T(\pi_2)} & T(Y) \end{array}$$

We write $2 = 1 + 1$ in the category $\mathbf{C}$. Maps $X \to T(2)$, that is, Kleisli maps $X \to 2$, are called *predicates* on X. Predicates $1 \to 2$ on 1 are also called *scalars*. We have truth, falsity, and orthosupplement predicates (where we write $\kappa_i\colon X_i \to X_1 + X_2$ for the coprojections):

$$\mathbf{1} = \Big(X \xrightarrow{!} 1 \xrightarrow{\eta \circ \kappa_1} T(2)\Big) \qquad \mathbf{0} = \Big(X \xrightarrow{!} 1 \xrightarrow{\eta \circ \kappa_2} T(2)\Big) \qquad p^{\perp} = \Big(X \xrightarrow{p} T(2) \xrightarrow[\cong]{T([\kappa_2, \kappa_1])} T(2)\Big)$$

A *state* of $X \in \mathbf{C}$ is a map $\omega\colon 1 \to T(X)$, that is, a Kleisli map $\omega\colon 1 \to X$. A state of the form $\omega\colon 1 \to X \otimes Y$ may be called a 'joint' or a 'bipartite' state. Its *marginals* are obtained by post-composition with projections, in $\mathcal{K}\ell(T)$, for which we use notation ω_i defined as:

$$\omega_1 \stackrel{\text{def}}{=} \Big(1 \xrightarrow{\omega} X \otimes Y \xrightarrow{\pi_1} X\Big) \qquad\qquad \omega_2 \stackrel{\text{def}}{=} \Big(1 \xrightarrow{\omega} X \otimes Y \xrightarrow{\pi_2} Y\Big) \tag{5}$$

For a state $\omega\colon 1 \to X$ and a predicate $p\colon X \to 2$ we define the validity $\omega \models p$, as a scalar $1 \to 2$, simply by Kleisli composition:

$$\omega \models p \stackrel{\text{def}}{=} p \bullet \omega \;:\; 1 \longrightarrow X \longrightarrow 2. \tag{6}$$

We shall see soon that the earlier description (2) is a special case. For a Kleisli map $f\colon X \to T(Y)$, a state $\omega\colon 1 \to T(X)$ and a predicate $q\colon Y \to T(2)$ we write:

$$f_*(\omega) = f \bullet \omega \quad \text{and} \quad f^*(q) = q \bullet f \quad \text{such that} \quad f_*(\omega) \models q = q \bullet f \bullet \omega = \omega \models f^*(q). \tag{7}$$

For a predicate $p\colon X \to T(2)$ we define associated *instrument* and *assert* maps $\mathrm{instr}_p\colon X \to T(X+X)$ and $\mathrm{asrt}_p\colon X \to T(X+1)$ as the following composite in the underlying category:

$$\begin{array}{ccc} X & \xrightarrow{\quad\mathrm{instr}_p\quad} & T(X+X) \\ {\scriptstyle\langle p,\mathrm{id}\rangle}\downarrow & & \uparrow{\scriptstyle \cong\; T(\pi_2+\pi_2)} \\ T(2)\times X \xrightarrow{\mathrm{st}} T\big(2\times X\big) & \xrightarrow{\cong} & T\big((1\times X)+(1\times X)\big) \end{array} \qquad \begin{array}{ccc} X & \xrightarrow{\quad\mathrm{asrt}_p\quad} & T(X+1) \\ {\scriptstyle \mathrm{instr}_p}\searrow & & \nearrow{\scriptstyle T(\mathrm{id}+!)} \\ & T(X+X) & \end{array} \tag{8}$$

In [12] it is shown, for a strongly affine monad T, that (a) instruments are side-effect-free, in the sense that $T(\nabla)\circ \mathrm{instr}_p = \eta$, where $\nabla = [\mathrm{id},\mathrm{id}]$ is the codiagonal; and (b) these instruments give a bijective correspondence between predicates $p\colon X \to T(2)$ and maps $f\colon X \to T(X+X)$ satisfying $T(\nabla)\circ f = \eta$. It is not hard to see that $(!+!) \bullet \mathrm{instr}_p = (!+\mathrm{id}) \bullet \mathrm{asrt}_p = p$, and that $\mathrm{instr}_r = r = \mathrm{asrt}_r$ for a scalar $r\colon 1 \to T(1+1)$. Examples of these assert maps will be given in Section 4.

We use the assert maps for three constructions, namely sequential conjunction $p \,\&\, q$, normalisation, and conditional states $\omega|_p$. The 'andthen' operations & is defined as:

$$p \,\&\, q \;=\; [q,\kappa_2] \bullet \mathrm{asrt}_p \;:\; X \longrightarrow X+1 \longrightarrow 2. \tag{9}$$

In [12] it is shown that & is commutative — since the monad T is commutative — and has the truth predicate $\mathbf{1}$ as neutral element: $p \,\&\, \mathbf{1} = p = \mathbf{1} \,\&\, p$.

A 'partial' Kleisli map of the form $f\colon X \to T(Y+1)$ is called *nowhere zero*, if its domain predicate $f \bullet \omega\colon 1 \to Y+1$ is not non-zero for each state $\omega\colon 1 \to X$. We say that the monad T is *normalising*, or admits *normalisation*, if for each nowhere zero $f\colon X \to Y+1$ in $\mathcal{K}\ell(T)$ there is a unique map $\mathrm{nrm}(f)\colon X \to Y$, making the diagram on the right commute in the Kleisli category $\mathcal{K}\ell(T)$. Notice that $\mathrm{nrm}(\mathrm{asrt}_p) = \mathrm{id}$, if $p \neq \mathbf{0}$. This description of normalisation is a (probabilistic) generalisation of the formulation from [14] that is only given for 'substates' $1 \to Y+1$.

$$\begin{array}{ccc} X & \xrightarrow{\quad f\quad} & Y+1 \\ {\scriptstyle \mathrm{asrt}_{\mathrm{dp}(f)}}\searrow & & \nearrow{\scriptstyle \mathrm{nrm}(f)+\mathrm{id}} \\ & X+1 & \end{array}$$

We summarise the properties that we will be using.

▶ **Definition 1.** We call a monad 'CSAN' if it is Commuative, Strongly Affine, and Normalising. In this situation we assume that the underlying category is distributive.

A *conditional state* $\omega|_p$ is defined for a state $\omega\colon 1 \to X$ and a predicate $p\colon X \to 2$ satisfying $\omega \models p \neq \mathbf{0} = \kappa_2\colon 1 \to 1+1$. The definition uses normalisation and assert in:

$$\omega|_p \stackrel{\mathrm{def}}{=} \mathrm{nrm}\Big(1 \xrightarrow{\omega} X \xrightarrow{\mathrm{asrt}_p} X+1\Big) \quad \text{unique with} \quad \begin{array}{ccc} 1 \xrightarrow{\omega} X & \xrightarrow{\mathrm{asrt}_p} & X+1 \\ {\scriptstyle \omega\models p}\searrow & & \nearrow{\scriptstyle \omega|_p+\mathrm{id}} \\ & 1+1 & \end{array} \tag{10}$$

The composition inside this normalisation operation $\mathrm{nrm}(-)$ is Kleisli composition.

The above abstract descriptions of states and predicates suffice to develop the basics of the theory of conditional probability. Maps in the underlying category can be seen as random (or stochastic) variables, translating between sample spaces. They don't play an explicit role in the present account.

▶ Remark. We have introduced the abbreviation 'CSAN' for certain monads in Definition 1. We refrain from using a more meaningful name, like 'probabilistic' for such monads. The reason is that the term 'probabilistic' is better used for a CSAN monad whose Kleisli category is an effectus (see [11, 5]). This additional requirement guarantees that collections of predicates $X \to 2$ form an effect module. But since we do not use such logical structure here, we prefer to keep things as simple as possible and use the *ad hoc* name CSAN. As we shall see in the next section, the non-empty powerset monad is an example of a CSAN monad, but calling it a probabilistic monad is hardly appropriate.

4 Monad examples

This section describes our main examples of CSAN monads: non-empty powerset $\mathcal{P}_\bullet$, distribution $\mathcal{D}$, and Giry $\mathcal{G}$. Probabilistic powerdomains are omitted, due to lack of space.

4.1 The non-empty powerset monad

The ordinary powerset monad $\mathcal{P}$ is not affine, since $\mathcal{P}(1) = 2$. Instead we use the non-empty powerset monad $\mathcal{P}_\bullet$ on the category **Sets** of sets and functions. It is a well-known example of a monad, with Kleisli composition $g \bullet f$ of $f \colon A \to \mathcal{P}_\bullet(B)$ and $g \colon B \to \mathcal{P}_\bullet(C)$ given by $(g \bullet f)(a) = \bigcup\{g(b) \mid b \in f(a)\}$. It is not hard to see that $\mathcal{P}_\bullet$ is strongly affine.

This monad $\mathcal{P}_\bullet$ is commutative, with double strength $\mathrm{dst} \colon \mathcal{P}_\bullet(A) \times \mathcal{P}_\bullet(B) \to \mathcal{P}_\bullet(A \times B)$ given by $\mathrm{dst}(U, V) = U \times V$. The most interesting property is normalisation. So let $f \colon A \to \mathcal{P}_\bullet(B+1)$ be nowhere zero, where $A+1 = A \cup \{*\}$. Since each $f(a)$ is non-empty, and $f(a) \neq \{*\}$, there must be at least one element $b \in B$ in $f(a)$. Hence $\mathrm{nrm}(f)(a) = f(a) - \{*\}$ obtained by removing $*$, if present, is non-empty, giving a function $\mathrm{nrm}(f) \colon A \to \mathcal{P}_\bullet(B)$.

The set $2 = 1 + 1$ is commonly identified with $\{0, 1\}$. The power set $\mathcal{P}_\bullet(2)$ has three elements, namely falsity $\{0\}$, truth $\{1\}$, and a third, combined option $\{0, 1\}$. For a predicate $p \colon A \to \mathcal{P}_\bullet(2)$ and a state $U \in \mathcal{P}_\bullet(A)$ we thus have $U \models p = \bigcup\{p(a) \mid a \in U\}$ and $U|_p = \{a \in U \mid 1 \in p(a)\}$, where in the latter case we assume $U \models p \neq \{0\}$, so there is at least one $a \in U$ with $1 \in p(a)$.

4.2 The distribution monad

The elements of the set $\mathcal{D}(A)$ are the (finite discrete) probability distributions on the set A, as described earlier in (1). This is forms a well-known monad, see *e.g.* [10] for details. The (Kleisli) composition of maps $f \colon A \to \mathcal{D}(B)$ and $g \colon B \to \mathcal{D}(C)$ is given by:

$$(g \bullet f)(a) = \sum_{c \in C} \big(\textstyle\sum_{b \in B} f(a)(b) \cdot g(b)(c)\big)\big|c\big\rangle.$$

Notice that the outer sum is a formal one, whereas the inner one is a proper sum in the unit interval $[0, 1]$. These Kleisli maps form stochastic matrices, which are composed via matrix multiplication. Given a Kleisli map $f \colon A \to \mathcal{D}(B)$, a state $\omega \in \mathcal{D}(A)$ and a predicate $p \in [0, 1]^B$ we get, according to (7):

$$f_*(\omega) = \sum_{b \in B} \big(\textstyle\sum_{a \in A} f(a)(b) \cdot \omega(a)\big)\big|b\big\rangle \quad \text{and} \quad f^*(p)(a) = \sum_{b \in B} f(a)(b) \cdot p(b). \tag{11}$$

The distribution monad $\mathcal{D}$ is strongly affine, as shown in [12], and commutative via the map $\mathrm{dst} \colon \mathcal{D}(A) \times \mathcal{D}(B) \to \mathcal{D}(A \times B)$ given by $\mathrm{dst}(\phi, \psi)(a, b) = \phi(a) \cdot \psi(b)$. We show that it also admits normalisation. Assume we have $f \colon A \to \mathcal{D}(B + 1)$ which is nowhere zero. This means $f(a)(*) \neq 1$ for each $a \in A$. Hence we can define its normalisation $\mathrm{nrm}(f) \colon A \to \mathcal{D}(B)$ as:

$$\mathrm{nrm}(f)(a) = \sum_{b \in B} \frac{f(a)(b)}{1 - f(a)(*)}\big|b\big\rangle.$$

Thus, for instance, for $\phi = \frac{1}{3}|u\rangle + \frac{1}{2}|v\rangle + \frac{1}{6}|*\rangle \in \mathcal{D}(\{u, v\} + 1)$ we get $\mathrm{nrm}(\phi) = \frac{2}{5}|u\rangle + \frac{3}{5}|v\rangle$.

We notice that $\mathcal{D}(1) \cong 1$ and $\mathcal{D}(2) \cong [0,1]$. Hence a predicate $p\colon A \to \mathcal{D}(2)$ may be identified with a fuzzy predicate $p \in [0,1]^A$ as in Section 2. Scalars $1 \to \mathcal{D}(2)$ correspond to probabilities in $[0,1]$. The assert map (8) for a predicate $p \in [0,1]^A$ is the map $\mathrm{asrt}_p\colon A \to \mathcal{D}(A+1)$ determined as: $\mathrm{asrt}_p(a) = p(a)|\,a\,\rangle + (1-p(a))|\,*\,\rangle$. The 'andthen' operation $p \,\&\, q$ is given by multiplication: $(p \,\&\, q)(a) = p(a) \cdot q(a)$. For a state/distribution $\omega \in \mathcal{D}(A)$ the conditional state $\omega|_p$ is defined as in (4).

4.3 The Giry monad

The Giry monad $\mathcal{G}$ is defined on the category **Meas** of measurable spaces, with measurable functions between them, see [8] or [9] (whose notation we follow). It sends a measurable space $A = (A, \Sigma_A)$ to the set $\mathcal{G}(A)$ of probabilty measures $\omega\colon \Sigma_A \to [0,1]$, where Σ_A is the σ-algebra of measurable subsets of A. The latter map ω is a countably additive function with $\omega(A) = 1$. Kleisli composition of $f\colon A \to \mathcal{G}(B)$ and $g\colon B \to \mathcal{G}(C)$ is given by integration:

$$\big(g \bullet f\big)(a)(N) \;=\; \int g(-)(N)\,\mathrm{d}f(a) \quad \text{for } a \in A \text{ and } N \in \Sigma_C, \text{ with } g(-)(N)\colon B \to [0,1].$$

A Kleisli map $f\colon A \to \mathcal{G}(B)$, a state $\omega \in \mathcal{G}(A)$ and a predicate $p\colon B \to [0,1]$ give, as in (7):

$$f_*(\omega)(N) \;=\; \int f(-)(N)\,\mathrm{d}\omega \qquad \text{and} \qquad f^*(p)(a) \;=\; \int p\,\mathrm{d}f(a) \tag{12}$$

This Giry monad is also strongly affine, see [12], and commutative via the map $\mathrm{dst}\colon \mathcal{G}(A) \times \mathcal{G}(B) \to \mathcal{G}(A \times B)$ determined by $\mathrm{dst}(\phi,\psi)(M \times N) = \phi(M) \cdot \psi(N)$, where $M \in \Sigma_A, N \in \Sigma_B$. Normalisation of $f\colon A \to \mathcal{G}(B+1)$ is done in the following way, when f is nowhere zero, that is $f(a)(\{*\}) \neq 1$, or equivalently, $f(a)(B) \neq 0$. We then define a measurable function $\mathrm{nrm}(f)\colon A \to \mathcal{G}(B)$ via measures $\mathrm{nrm}(f)(a)\colon \Sigma_B \to [0,1]$ via $\mathrm{nrm}(f)(a)(N) = \frac{f(a)(N)}{f(a)(B)}$.

We have $\mathcal{G}(1) \cong 1$ and $\mathcal{G}(2) \cong [0,1]$, so that predicates $p\colon A \to 2$ in $\mathcal{K}\ell(\mathcal{G})$ can be identified with measurable functions $p\colon A \to [0,1]$. Scalars $1 \to \mathcal{G}(2)$ correspond to elements of $[0,1]$. The assert map $\mathrm{asrt}_p\colon A \to \mathcal{G}(A+1)$ is given by:

$$\mathrm{asrt}_p(a)(M) \;=\; p(a) \cdot \mathbf{1}_M, \;\; \text{for } M \in \Sigma_A \qquad \text{and} \qquad \mathrm{asrt}_p(a)(\{*\}) \;=\; 1 - p(a).$$

Again, andthen & is given by multiplication, but conditioning is more interesting: for a state/measure $\omega \in \mathcal{G}(A)$ and a predicate $p\colon A \to [0,1]$ we have:

$$\omega \models p \;=\; p \bullet \omega \;=\; \int p\,\mathrm{d}\omega. \tag{13}$$

Similarly, for $M \in \Sigma_A$,

$$\big(\mathrm{asrt}_p \bullet \omega\big)(M) \;=\; \int \mathrm{asrt}_p(-)(M)\,\mathrm{d}\omega \;=\; \int p(-) \cdot \mathbf{1}_M\,\mathrm{d}\omega \;=\; \int_M p\,\mathrm{d}\omega.$$

Now we see that the conditional state/measure $\omega|_p\colon \Sigma_A \to [0,1]$ is given by:

$$\omega|_p(M) \;=\; \frac{\int_M p\,\mathrm{d}\omega}{\int p\,\mathrm{d}\omega} \qquad \text{so that} \qquad \omega|_{\mathbf{1}_N}(M) \;=\; \frac{\int_M \mathbf{1}_N\,\mathrm{d}\omega}{\int \mathbf{1}_N\,\mathrm{d}\omega} \;=\; \frac{\int \mathbf{1}_{M\cap N}\,\mathrm{d}\omega}{\omega(N)} \;=\; \frac{\omega(M\cap N)}{\omega(N)}.$$

▶ Remark. The conditional state/measure $\omega|_p \in \mathcal{G}(A)$ defined above gives for a measurable subset $M \in \Sigma_A$ the (normalised) probability determined by the integral $\int_M p\,\mathrm{d}\omega$, describing the surface under p on M. This is precisely what a 'probability density function' (pdf) does. If $\phi = \omega|_p$, then the function p is called the Radon-Nikodym derivative of ϕ, written as $\frac{d\phi}{d\omega}$. In our case the predicate p is $[0,1]$-valued, but this is not really a restriction since we can always scale a bounded function $A \to \mathbb{R}_{\geq 0}$ to fit in $A \to [0,1]$.

5 States of compound objects

In the remainder of this paper we fix a commutative, strongly affine, normalisable (CSAN) monad T on a category $\mathbf{C}$. Its Kleisli category $\mathcal{K}\ell(T)$ is symmetric monoidal, with tensor $\otimes$ given by cartesian product $\times$ on objects, and with final object 1 as tensor unit. In this section we first consider 'joint' states $1 \to X \otimes Y$ in $\mathcal{K}\ell(T)$ on (tensor) products, and then on states $1 \to X + Y$ on coproducts. Recall from (5) that we write $\omega_i = \pi_i \bullet \omega$ for the marginals of $\omega\colon 1 \to X \otimes Y$. For states $\phi\colon 1 \to X$ and $\psi\colon 1 \to Y$ we write $\phi \otimes \psi\colon 1 \to X \otimes Y$ for the 'product' state, and $\phi \otimes \mathrm{id}\colon Y \to X \otimes Y$ and $\mathrm{id} \otimes \psi\colon X \to X \otimes Y$ for the obvious maps.

▶ **Definition 2.** A state $\omega\colon 1 \to X \otimes Y$ in $\mathcal{K}\ell(T)$ is called *non-entwined* if it is the tensor $\omega = \omega_1 \otimes \omega_2$ of its marginals $\omega_i = \pi_i \bullet \omega$. A state is called *entwined* otherwise.

More explicitly, a state ω is non-entwined if the diagram on the right commutes (in the Kleisli category). This notion of entwinedness corresponds to entanglement in the quantum world. Sometimes it is called dependence, but that may create confusion because dependence is best seen as a property of predicates. We choose to use new words 'entwinedness' and 'non-entwinedness', with a precise meaning, as described above, since we formulate this notion generically, wrt. a monad T.

$$\begin{array}{ccc} 1 & \xrightarrow{\quad\omega\quad} & X \otimes Y \\ {\scriptstyle\cong}\downarrow & & \uparrow{\scriptstyle\pi_1 \otimes \pi_2} \\ 1 \otimes 1 & \xrightarrow[\omega\otimes\omega]{} & (X \otimes Y) \otimes (X \otimes Y) \end{array}$$

Notice that a 'product' state of the form $\phi \otimes \psi\colon 1 \to X \otimes Y$ is non-entwined by construction. We shall describe what (dis)entwinedness means for our running examples from Section 4.

▶ **Example 3.** Marginalisation for the non-empty powerset monad $\mathcal{P}_\bullet$ of $\omega \in \mathcal{P}_\bullet(A \times B)$ is obtained as $\omega_1 = \mathcal{P}_\bullet(\pi_1)(\omega) = \{a \in A \mid \exists b \in B.\,(a,b) \in \omega\}$. An example of a non-entwined state for $\mathcal{P}_\bullet$ with sets $X = \{1,2\}$ and $Y = \{a,b\}$ is the non-empty subset $\{(1,a)\} = \{1\} \otimes \{a\} \in \mathcal{P}_\bullet(X \times Y)$. An entwined state is $\omega = \{(1,b),(2,a)\}$, since its marginals are $\omega_1 = \{1,2\}$ and $\omega_2 = \{a,b\}$, but $\omega \neq \omega_1 \otimes \omega_2 = \{(1,a),(1,b),(2,a),(2,b)\}$.

For the distribution monad $\mathcal{D}$ the marginals of $\omega \in \mathcal{D}(A \times B)$ are computed as:

$$\omega_1 \;=\; \sum_a \big(\textstyle\sum_b \omega(a,b)\big)\big|\,a\,\big\rangle \;\in\; \mathcal{D}(A) \qquad \text{and} \qquad \omega_2 \;=\; \displaystyle\sum_b \big(\textstyle\sum_a \omega(a,b)\big)\big|\,b\,\big\rangle \;\in\; \mathcal{D}(B).$$

We have a non-entwined example in $\mathcal{D}(X \times Y)$, for $X = \{1,2\}, Y = \{a,b\}$ like before:

$$\tfrac{1}{5}|1,a\rangle + \tfrac{2}{5}|2,a\rangle + \tfrac{2}{15}|1,b\rangle + \tfrac{4}{15}|2,b\rangle \;=\; \big(\tfrac{1}{3}|1\rangle + \tfrac{2}{3}|2\rangle\big) \otimes \big(\tfrac{3}{5}|a\rangle + \tfrac{2}{5}|b\rangle\big).$$

An entwined example is $\omega = \frac{1}{2}|1,b\rangle + \frac{1}{2}|2,a\rangle$ with marginals $\omega_1 = \frac{1}{2}|1\rangle + \frac{1}{2}|2\rangle$ and $\omega_2 = \frac{1}{2}|a\rangle + \frac{1}{2}|b\rangle$, which give $\omega_1 \otimes \omega_2 = \frac{1}{4}|1,a\rangle + \frac{1}{4}|2,a\rangle + \frac{1}{4}|1,b\rangle + \frac{1}{4}|2,b\rangle \neq \omega$.

In general, one can prove that a state $\omega = r_1|1,a\rangle + r_2|2,a\rangle + r_3|1,b\rangle + r_4|2,b\rangle$, where $r_1 + r_2 + r_3 + r_4 = 1$, is non-entwined if and only if $r_1 r_4 = r_2 r_3$. This fact also holds in the quantum case, see *e.g.* [15, §1.5].

For a state/measure $\omega \in \mathcal{G}(A \times B)$ one calculates the (first) marginal $\omega_1\colon \Sigma_A \to [0,1]$ as $\omega_1(M) = \mathcal{G}(\pi_1)(\omega)(M) = \omega\big(\pi_1^{-1}(M)\big) = \omega(M \times Y)$. This ω is thus non-entwined, that is, the product $\omega = \omega_1 \otimes \omega_2$ of its marginals, if and only if for each $M \in \Sigma_A$ and $N \in \Sigma_B$ we have $\omega(M \times N) = \omega(M \times Y) \cdot \omega(X \times N)$.

The next result gives an abstract version of what is called the Fubini (or Fubini-Tonelli) Theorem in the theory of integration. Here the proof is extremely simple, and only uses the monoidal structure. This shows the power of abstraction.

▶ **Theorem 4** (Fubini). *For states $\phi\colon 1 \to X$ and $\psi\colon 1 \to Y$, and a predicate $p\colon X \otimes Y \to 2$ in $\mathcal{K}\ell(T)$ the validity of p in the product state $\phi \otimes \psi\colon 1 \to X \otimes Y$ can be computed via both the component states separately:*

$$\phi \models (\mathrm{id} \otimes \psi)^*(p) \;=\; \phi \otimes \psi \models p \;=\; \psi \models (\phi \otimes \mathrm{id})^*(p). \tag{14}$$

Proof. We simply have:

$$\begin{aligned} \phi \models (\mathrm{id} \otimes \psi)^*(p) \;=\; p \bullet (\mathrm{id} \otimes \psi) \bullet \phi &= p \bullet (\phi \otimes \psi) \\ &= (\phi \otimes \psi) \models p \\ &= p \bullet (\phi \otimes \mathrm{id}) \bullet \psi \;=\; \psi \models (\phi \otimes \mathrm{id})^*(p). \end{aligned}$$ ◀

For the distribution monad $\mathcal{D}$ the Fubini equations (14) amount to:

$$\sum_x \phi(x) \cdot \Big(\sum_y \psi(y) \cdot p(x,y)\Big) \;=\; \sum_{x,y} \phi(x) \cdot \psi(y) \cdot p(x,y) \;=\; \sum_y \psi(y) \cdot \Big(\sum_x \phi(x) \cdot p(x,y)\Big).$$

For the Giry monad $\mathcal{G}$ the equations (14) take the familiar 'Fubini' form, via (12) and (13):

$$\int \Big(\int p \,\mathrm{d}(\mathrm{id} \otimes \psi)\Big)\,\mathrm{d}\phi \;=\; \int p\,\mathrm{d}(\phi \otimes \psi) \;=\; \int \Big(\int p\,\mathrm{d}(\phi \otimes \mathrm{id})\Big)\,\mathrm{d}\psi.$$

We should point out again that we use integration only for functions to $[0,1]$, so that our form of Fubini is more restricted than usual, see *e.g.* [18, Thm. 13.8], [17, Thm. 9.4.1] or [16, Thm. 3.16].

So far we have studied states of tensor products $X \otimes Y$. Next we consider states of coproducts $X + Y$. We shall describe one result about validity for such states. We additionally require that the Kleisli category $\mathcal{K}\ell(T)$ our CSAN monad T is an effectus. We do not explain what this means here, and refer to [11, 5] for more details. Specifically we use that the collection of predicates is an effect module, with (partial) sum $\varovee$ and convex sums. Less abstractly, one can read $T = \mathcal{D}$ or $T = \mathcal{G}$ in this result, so that the scalars are in $[0,1]$.

We use that a state $\omega\colon 1 \to X + Y$ of a coproduct gives rise to two substates $\rhd_1 \bullet \omega\colon 1 \to X + 1$ and $\rhd_2 \bullet \omega\colon 1 \to Y + 1$, where $\rhd_1 = (\mathrm{id} + !)\colon X + Y \to X + 1$ and $\rhd_2 = [! \circ \kappa_2, \kappa_1]\colon X + Y \to Y + 1$ are called 'partial projections' (in [11, 5]). These two substates can be normalised to ordinary states $1 \to X$ and $1 \to Y$.

▶ **Theorem 5.** *Let $\omega\colon 1 \to X + Y$ be state and $p\colon X \to 2$ and $q\colon Y \to 2$ be predicates in the Kleisli category of a CSAN monad whose Kleisli category is an effectus. Then:*

$$\omega \models [p,q] \;=\; r \cdot \big(\mathrm{nrm}(\rhd_1 \bullet \omega) \models p\big) \varovee r^{\perp} \cdot \big(\mathrm{nrm}(\rhd_2 \bullet \omega) \models q\big),$$

where $r = (! + !) \bullet \omega\colon 1 \to 2$ is a scalar used in the above convex sum, with $r^{\perp} = 1 - r$.

Proof. We use that the coproduct predicate $[p,q]$ can be written as sum $[p,\mathbf{0}] \varovee [\mathbf{0},q]$, see [11, §6]. Write $\phi = \rhd_1 \bullet \omega = (\mathrm{id} + !) \circ \omega\colon 1 \to X + 1$. The normalisation $\mathrm{nrm}(\phi)$ satisfies by construction:

$$\phi \;=\; (\mathrm{nrm}(\phi) + \mathrm{id}) \bullet (! + \mathrm{id}) \bullet \phi \;=\; (\mathrm{nrm}(\phi) + \mathrm{id}) \bullet (! + !) \bullet \omega \;=\; (\mathrm{nrm}(\phi) + \mathrm{id}) \bullet r.$$

Hence:

$$\begin{aligned} \omega \models [p,\mathbf{0}] \;&=\; [p, \kappa_2 \bullet !] \bullet \omega \;=\; [p,\kappa_2] \bullet (\mathrm{id} + !) \bullet \omega \;=\; [p,\kappa_2] \bullet \phi \;=\; [p,\kappa_2] \bullet (\mathrm{nrm}(\phi) + \mathrm{id}) \bullet r \\ &=\; \big[p \bullet \mathrm{nrm}(\phi), \kappa_2\big] \bullet r \;=\; \big[\mathrm{nrm}(\phi) \models p, \kappa_2\big] \bullet r \;=\; r \cdot \big(\mathrm{nrm}(\rhd_1 \bullet \omega) \models p\big). \end{aligned}$$

In a similar one obtains $\omega \models [p, \mathbf{0}] = r^{\perp} \cdot (\mathrm{nrm}(\rhd_2 \bullet \omega) \models q)$. Putting things together we get:

$$\begin{aligned}\omega \models [p, q] \;=\; \omega \models ([p, \mathbf{0}] \ovee [\mathbf{0}, q]) \;&=\; (\omega \models [p, \mathbf{0}]) \ovee (\omega \models [\mathbf{0}, q]) \\ &=\; r \cdot \big(\mathrm{nrm}(\rhd_1 \bullet \omega) \models p\big) \ovee r^{\perp} \cdot \big(\mathrm{nrm}(\rhd_2 \bullet \omega) \models q\big). \quad \blacktriangleleft\end{aligned}$$

This result occurs as [9, Lem. 12] for the Giry monad.

6 Conditional states

In this section we take a closer look at the conditional states $\omega|_p$ as defined in (10). We first show how to obtain Bayes' rule in this setting — assuming a CSAN monad in the background.

▶ **Proposition 6.** *Assume the validity $\omega \models p$ is non-zero, for a state ω and predicate p. Then:*

$$\big(\omega|_p \models q\big) \cdot \big(\omega \models p\big) \;=\; (\omega \models p \,\&\, q). \tag{15}$$

By commutativity of & we obtain Bayes' rule, formulated here without division as:

$$\big(\omega|_p \models q\big) \cdot \big(\omega \models p\big) \;=\; (\omega \models p \,\&\, q) \;=\; (\omega \models q \,\&\, p) \;=\; \big(\omega|_q \models p\big) \cdot \big(\omega \models q\big).$$

(Commutativity of & fails in a quantum setting, so that this derivation does not work there.)

Proof. Recall from (10) that $\omega|_p$ is defined as the normalisation $\omega|_p = \mathrm{nrm}(\mathrm{asrt}_p \bullet \omega)$, satisfying $(\omega|_p + \mathrm{id}) \bullet (\omega \models p) = \mathrm{asrt}_p \bullet \omega$. Hence:

$$\begin{aligned}(\omega|_p \models q) \cdot (\omega \models p) \;&=\; [(\omega|_p \models q), \kappa_2] \bullet (\omega \models p) \\ &=\; [q \bullet \omega|_p, \kappa_2] \bullet (\omega \models p) \\ &=\; [q, \kappa_2] \bullet (\omega|_p + \mathrm{id}) \bullet (\omega \models p) \\ &=\; [q, \kappa_2] \bullet \mathrm{asrt}_p \bullet \omega \;=\; (p \,\&\, q) \bullet \omega \;=\; \omega \models (p \,\&\, q). \quad \blacktriangleleft\end{aligned}$$

Applying the above conditional validity equation (15) to the Giry monad for continuous probability gives a formula for integration with a conditional measure $\omega|_p$, namely:

$$\int q \,\mathrm{d}(\omega|_p) \;=\; \frac{\int p \cdot q \,\mathrm{d}\omega}{\int p \,\mathrm{d}\omega}. \tag{16}$$

The next result says that the mapping $(-)|_p$ is an action of predicates on states, using the monoid structure $(\&, \mathbf{1})$ on predicates.

▶ **Lemma 7.** *Assuming the relevant constructions are defined we have:*

$$(\omega|_p)|_q \;=\; \omega|_{p\&q} \qquad \text{and} \qquad \omega|_{\mathbf{1}} \;=\; \omega.$$

Moreover, $f \circ (\omega|_{q \circ f}) = (f \circ \omega)|_q$ for a (pure) map f in the underlying category. ◀

Since & is commutative, we obtain that conditioning can be exchanged: $(\omega|_p)|_q = (\omega|_q)|_p$. We conclude with a result combining marginalisation, product states, and conditioning. It uses Fubini, and can be read as a 'conditional' extension of Fubini.

▶ **Theorem 8.** *Let $\phi \colon 1 \to X$ and $\psi \colon 1 \to Y$ be two states, with a predicate $p \colon X \times Y \to 2$. The marginalisation $\big((\phi \otimes \psi)|_p\big)_1$ equals $\phi|_{(\mathrm{id} \otimes \psi)^*(p)}$.*

Proof. We apply uniqueness of normalisation to get the required result, via Fubini:

$$\begin{array}{rcl}\big(((\phi \otimes \psi)|_p)_1 + \mathrm{id}\big) \bullet (\phi \models (\mathrm{id} \otimes \psi)^*(p)) & = & (\pi_1 + \mathrm{id}) \bullet ((\phi \otimes \psi)|_p + \mathrm{id}) \bullet (\phi \otimes \psi \models p) \\ & \stackrel{(10)}{=} & (\pi_1 + \mathrm{id}) \bullet \mathrm{asrt}_p \bullet (\mathrm{id} \otimes \psi) \bullet \phi \\ & \stackrel{(*)}{=} & \mathrm{asrt}_{(\mathrm{id} \otimes \psi)^*(p)} \bullet \phi \\ & \stackrel{(10)}{=} & \big(\phi|_{(\mathrm{id} \otimes \psi)^*(p)} + \mathrm{id}\big) \bullet (\phi \models (\mathrm{id} \otimes \psi)^*(p)).\end{array}$$

The marked equation $\stackrel{(*)}{=}$ is obtained by unraveling the definition of assert maps. ◀

One can prove additional properties, like $(\phi|_p) \otimes (\psi|_q) = (\phi \otimes \psi)|_{\pi_1^*(p) \& \pi_2^*(q)}$, for predicates $p\colon X \to 2$ and $q\colon Y \to 2$. The details will appear later.

7 Conditional non-entwinedness

We first need some notation. For a Kleisli map $f\colon Y \to T(X)$ we write, like in [7], $\mathrm{gr}(f)\colon Y \to T(X \times Y)$ for the 'graph' map $\mathrm{gr}(f) = \mathrm{st} \circ \langle f, \mathrm{id}\rangle$. We shall need the Kleisli extension $\mathrm{gr}(f)_* = \mu \circ T(\mathrm{gr}(f))\colon T(Y) \to T(X \times Y)$ below. For the monads $\mathcal{D}$ and $\mathcal{G}$ these extension maps are given, respectively, by:

$$\mathrm{gr}(f)_*(\psi) = \sum_{x,y} f(y)(x) \cdot \psi(y)\big|x, y\big\rangle \quad \text{and} \quad \mathrm{gr}(f)_*(\psi)(M \times N) = \int_N f(-)(M)\,\mathrm{d}\psi. \tag{17}$$

This $\mathrm{gr}(f)_*(\psi)$ is the *joint* probability induced by the *conditional* probability f and state ψ.

The following description captures at an abstract categorical level what has been formulated for the Giry monad in [6].

▶ **Definition 9.** **1.** Let $\omega\colon 1 \to X \otimes Y$ be a 'joint' or 'bipartity' state. A *conditional* for ω is a (Kleisli) map $f\colon Y \to T(X)$ such that the following diagram commutes in $\mathcal{K}\!\ell(T)$.

$$\begin{array}{ccc} 1 & \xrightarrow{\;\omega\;} & X \otimes Y \\ {\scriptstyle\omega}\downarrow & & \uparrow{\scriptstyle\mathrm{gr}(f)} \\ X \otimes Y & \xrightarrow[\;\pi_2\;]{} & Y \end{array}$$

This says that ω can be reconstructed from its marginal $\omega_2 = \pi_2 \bullet \omega$ in a functional way.

2. A 'tripartite' state $\omega\colon 1 \to (X_1 \otimes X_2) \otimes Y$ is called *conditionally non-entwined* over Y if its marginal bipartite states $(\pi_i \otimes \mathrm{id}) \bullet \omega\colon 1 \to X_i \otimes Y$ have conditionals $f_i\colon Y \to X_i$ with:

$$\begin{array}{ccc} 1 & \xrightarrow{\;\omega\;} & (X_1 \otimes X_2) \otimes Y \\ {\scriptstyle\omega}\downarrow & & \uparrow{\scriptstyle\mathrm{gr}(\langle f_1, f_2\rangle)} \\ (X_1 \otimes X_2) \otimes Y & \xrightarrow[\;\pi_2\;]{} & Y \end{array}$$

Before illustrating these notions with examples, we mention some basic results.

▶ **Lemma 10.** *1. A state $1 \to X_1 \otimes X_2$ is non-entwined (as in Definition 2) iff it is conditionally non-entwined over the final object 1, as map $1 \to X_1 \otimes X_2 \cong (X_1 \otimes X_2) \otimes 1$.*

2. Given a state $\psi\colon 1 \to Y$ and two maps $f_i\colon Y \to X_i$, then the state $\omega = \mathrm{gr}(\langle f_1, f_2\rangle) \bullet \psi\colon 1 \to (X_1 \otimes X_2) \otimes Y$ is conditionally non-entwined over Y. ◀

The situation in the second point is typical for Bayesian networks where each subgraph $\swarrow\!\!\nearrow\!\!\searrow$ gives rise to a conditionally non-entwined distribution/measure on the product of the (three) nodes, over the node at the top — see also [6].

We shall give examples for the distribution and the Giry monad. For a joint distribution $\omega \in \mathcal{D}(X \times Y)$ there is a canonical conditional $f\colon Y \to \mathcal{D}(X)$, namely:

$$f(y) \;=\; \sum_{x\in X} \frac{\omega(x,y)}{\omega_2(y)}\big|\,x\,\big\rangle \qquad \text{where } \omega_2 \in \mathcal{D}(Y) \text{ is the marginal.} \tag{18}$$

This is well-defined if the set Y is the support of the marginal ω_2.

Here is a concrete example of a conditionally non-entwined state. Consider sets $A = \{a, a^\perp\}, B = \{b, b^\perp\}, C = \{c, c^\perp\}$, and the tripartite distribution $\omega \in \mathcal{D}((A \times B) \times C)$,

$$\begin{aligned}\omega \;=\; & \tfrac{2}{120}|\,abc\,\rangle + \tfrac{12}{120}|\,abc^\perp\,\rangle + \tfrac{4}{120}|\,ab^\perp c\,\rangle + \tfrac{3}{120}|\,ab^\perp c^\perp\,\rangle \\ & + \tfrac{8}{120}|\,a^\perp bc\,\rangle + \tfrac{60}{120}|\,a^\perp bc^\perp\,\rangle + \tfrac{16}{120}|\,a^\perp b^\perp c\,\rangle + \tfrac{15}{120}|\,a^\perp b^\perp c^\perp\,\rangle.\end{aligned}$$

The bipartite states $\omega_1 = \mathcal{D}(\pi_1 \times \mathrm{id})(\omega) \in \mathcal{D}(A \times C)$ and $\omega_2 = \mathcal{D}(\pi_2 \times \mathrm{id})(\omega) \in \mathcal{D}(B \times C)$, together with $\phi = \mathcal{D}(\pi_2)(\omega) \in \mathcal{D}(C)$, are:

$$\begin{aligned}\omega_1 &= \tfrac{6}{120}|\,ac\,\rangle + \tfrac{15}{120}|\,ac^\perp\,\rangle + \tfrac{24}{120}|\,a^\perp c\,\rangle + \tfrac{75}{120}|\,a^\perp c^\perp\,\rangle \\ \omega_2 &= \tfrac{10}{120}|\,bc\,\rangle + \tfrac{72}{120}|\,bc^\perp\,\rangle + \tfrac{20}{120}|\,b^\perp c\,\rangle + \tfrac{18}{120}|\,b^\perp c^\perp\,\rangle\end{aligned} \qquad \phi \;=\; \tfrac{30}{120}|\,c\,\rangle + \tfrac{90}{120}|\,c^\perp\,\rangle$$

The associated conditional maps $f_1\colon C \to \mathcal{D}(A)$ and $f_2\colon C \to \mathcal{D}(B)$ are $f_1(c) = \frac{1}{5}|\,a\,\rangle + \frac{4}{5}|\,a^\perp\,\rangle, f_1(c^\perp) = \frac{1}{6}|\,a\,\rangle + \frac{5}{6}|\,a^\perp\,\rangle$, and $f_2(c) = \frac{1}{3}|\,b\,\rangle + \frac{2}{3}|\,b^\perp\,\rangle, f_2(c^\perp) = \frac{4}{5}|\,b\,\rangle + \frac{1}{5}|\,b^\perp\,\rangle$. Then indeed: $\mathrm{gr}(f_i)_*(\phi) = \omega_i$, for $i = 1, 2$, using (17). Moreover, $\mathrm{gr}(\langle f_1, f_2\rangle)_*(\phi) = \omega$. This shows that the state/distribution ω is conditionally non-entwined over C.

The analogue of the formula (18) for continuous probability (with the Giry monad) is much more difficult, see *e.g.* [6, Prop. 3.3] and [16, Prop. 6.7]. The existence of a conditional f as in (18) is a consequence of the Radon-Nikodym Theorem. We conclude with a general result describing a conditional for a joint state of the form $(\phi\otimes\psi)|_p$. This is useful in practice, since it applies to the common situation where p is a probability density function on $\mathbb{R}^2$, see Remark 4.3, and ϕ, ψ are the Lebesgue measure on $\mathbb{R}$, with: $(\phi\otimes\psi)|_p(M \times N) = \frac{\int_{M\times N} p \,\mathrm{d}\phi\otimes\psi}{\int p\,\mathrm{d}\phi\otimes\psi}$.

▶ **Proposition 11.** *Let $\phi\colon 1 \to X$ and $\psi\colon 1 \to Y$ be states with a predicate $p\colon X \otimes Y \to 2$ such that the predicate $(\phi \otimes \mathrm{id})^*(p) = p \bullet (\phi \otimes \mathrm{id})\colon Y \to 2$ is non-zero. A conditional for the conditional state $\omega = (\phi \otimes \psi)|_p$ is then given by the map $\pi_1 \bullet \mathrm{nrm}(f)\colon Y \to X$ obtained as normalisation of the composite $f = \mathrm{asrt}_p \bullet (\phi \otimes \mathrm{id})\colon Y \to X \otimes Y \to (X \otimes Y) + 1$.*

Proof. The normalisation $\mathrm{nrm}(f)\colon Y \to X \otimes Y$ satisfies $(\mathrm{nrm}(f) + \mathrm{id}) \bullet \mathrm{asrt}_{(\phi\otimes\mathrm{id})^*(p)} = f$. We now prove the equation $\mathrm{gr}(\pi_1 \bullet \mathrm{nrm}(f)) \bullet \pi_2 \bullet \omega = \omega$ from Definition 9 (1). Since $\omega = (\phi \otimes \psi)|_p$ is a conditional state, we can use the uniqueness from (10) in:

$$\begin{aligned}&\big((\mathrm{gr}(\mathrm{nrm}(f)) \bullet \pi_2 \bullet \omega) + \mathrm{id}\big) \bullet (\phi \otimes \psi \models p) \\ &\quad= (\mathrm{gr}(\mathrm{nrm}(f)) + \mathrm{id}) \bullet (\psi|_{(\phi\otimes\mathrm{id})^*(p)} + \mathrm{id}) \bullet (\psi \models (\phi \otimes \mathrm{id})^*(p)) \quad \text{by Theorem 8 and 4} \\ &\quad\overset{(10)}{=} (\mathrm{gr}(\mathrm{nrm}(f)) + \mathrm{id}) \bullet \mathrm{asrt}_{(\phi\otimes\mathrm{id})^*(p)} \bullet \psi \\ &\quad\overset{(*)}{=} (\mathrm{nrm}(f) + \mathrm{id}) \bullet \mathrm{asrt}_{(\phi\otimes\mathrm{id})^*(p)} \bullet \psi \\ &\quad= f \bullet \psi \;=\; \mathrm{asrt}_p \bullet (\phi \otimes \psi) \overset{(10)}{=} \big((\phi \otimes \psi)|_p + \mathrm{id}\big) \bullet (\phi \otimes \psi \models p) \;=\; \big(\omega + \mathrm{id}\big) \bullet (\phi \otimes \psi \models p).\end{aligned}$$

The marked equation $\overset{(*)}{=}$ follows from the fact that $\pi_2 \bullet \mathrm{nrm}(f) = \mathrm{nrm}\big((\pi_2 + \mathrm{id}) \bullet \mathrm{asrt}_p \bullet (\phi \otimes \mathrm{id})\big) = \mathrm{nrm}(\mathrm{asrt}_{(\phi\otimes\mathrm{id})^*(p)}) = \mathrm{id}$. ◀

References

1 R. Adams and B. Jacobs. A type theory for probabilistic and Bayesian reasoning. See `arxiv.org/abs/1511.09230`, 2015.

2 D. Barber. *Bayesian Reasoning and Machine Learning*. Cambridge Univ. Press, 2012.

3 F. Bartels, A. Sokolova, and E. de Vink. A hierarchy of probabilistic system types. *Theoretical Computer Science*, 327(1-2):3–22, 2004.

4 J. Borgström, A.D. Gordon, M. Greenberg, J. Margetson, and J. Van Gael. Measure transformer semantics for Bayesian machine learning. *Logical Methods in Comp. Sci.*, 9(3):1–39, 2013.

5 K. Cho, B. Jacobs, A. Westerbaan, and B. Westerbaan. An introduction to effectus theory. see `arxiv.org/abs/1512.05813`, 2015.

6 B. Fong. Causal theories: A categorical perspective on Bayesian networks. Master's thesis, Univ. of Oxford, 2012. see `arxiv.org/abs/1301.6201`.

7 R. Furber and B. Jacobs. Towards a categorical account of conditional probability. In C. Heunen, P. Selinger, and J. Vicary, editors, *Quantum Physics and Logic (QPL) 2015*, number 195 in Elect. Proc. in Theor. Comp. Sci., pages 179–195, 2015.

8 M. Giry. A categorical approach to probability theory. In B. Banaschewski, editor, *Categorical Aspects of Topology and Analysis*, number 915 in Lect. Notes Math., pages 68–85. Springer, Berlin, 1982.

9 B. Jacobs. Measurable spaces and their effect logic. In *Logic in Computer Science*. IEEE, Computer Science Press, 2013.

10 B. Jacobs. Introduction to coalgebra. Towards mathematics of states and observations. Cambridge Univ. Press, to appear, 2015.

11 B. Jacobs. New directions in categorical logic, for classical, probabilistic and quantum logic. *Logical Methods in Comp. Sci.*, 11(3):1–76, 2015.

12 B. Jacobs. Affine monads and side-effect-freeness. CMCS'16, to appear, 2016.

13 B. Jacobs and J. Mandemaker. The expectation monad in quantum foundations. In B. Jacobs, P. Selinger, and B. Spitters, editors, *Quantum Physics and Logic (QPL) 2011*, number 95 in Elect. Proc. in Theor. Comp. Sci., pages 143–182, 2012.

14 B. Jacobs, B. Westerbaan, and A. Westerbaan. States of convex sets. In A. Pitts, editor, *Foundations of Software Science and Computation Structures*, number 9034 in Lect. Notes Comp. Sci., pages 87–101. Springer, Berlin, 2015.

15 N.D. Mermin. *Quantum Computer Science: An Introduction*. Cambridge Univ. Press, 2007.

16 P. Panangaden. *Labelled Markov Processes*. Imperial College Press, 2009.

17 J.S. Rosenthal. *A First Look at Rigorous Probability Theory*. World Scientific, Singapore, 2^{nd} edition, 2006.

18 R. Schilling. *Measures, Integrals and Martingales*. Cambridge Univ. Press, 2006.

19 A. Ścibior, Z. Ghahramani, and A. Gordon. Practical probabilistic programming with monads. In *Proc. 2015 ACM SIGPLAN Symp. on Haskell*, pages 165–176. ACM, 2015.