Towards a Categorical Account of Conditional Probability

Robert Furber and Bart Jacobs

Institute for Computing and Information Sciences (iCIS),
Radboud University Nijmegen, The Netherlands.
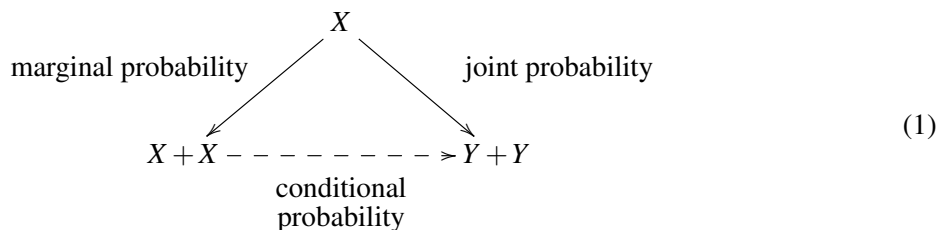Web addresses: `www.cs.ru.nl/~rfurber` and `www.cs.ru.nl/~bart`

This paper presents a categorical account of conditional probability, covering both the classical and the quantum case. Classical conditional probabilities are expressed as a certain "triangle-fill-in" condition, connecting marginal and joint probabilities, in the Kleisli category of the distribution monad. The conditional probabilities are induced by a map together with a predicate (the condition). The latter is a predicate in the logic of effect modules on this Kleisli category.

This same approach can be transferred to the category of C^*-algebras (with positive unital maps), whose predicate logic is also expressed in terms of effect modules. Conditional probabilities can again be expressed via a triangle-fill-in property. In the literature, there are several proposals for what quantum conditional probability should be, and also there are extra difficulties not present in the classical case. At this stage, we only describe quantum systems with classical parametrization.

1 Introduction

In the categorical description of probability theory, several monads play an important role. The main ones are the discrete probability monad $\mathscr{D}$ on the category **Sets** of sets and functions, and the Giry monad $\mathscr{G}$, for continuous probability, on the category **Meas** of measurable spaces and measurable functions. The Kleisli categories of these monads have suitable probabilistic matrices as morphisms, which capture probabilistic transition systems (and Markov chains). Additionally, more recent monads of interest are the expectation monad [8] and the Radon monad [5].

The first contribution of this paper is a categorical reformulation of classical (discrete) conditional probability as a "triangle-fill in" property in the Kleisli category $\mathscr{K}\ell(\mathscr{D})$ of the distribution monad. Abstractly, this fill-in property appears as follows.

$$
\begin{array}{ccc}
 & X & \\
\text{marginal probability} \swarrow & & \searrow \text{joint probability} \\
X + X \dashrightarrow & & Y + Y \\
 & \text{conditional} & \\
 & \text{probability} &
\end{array}
\tag{1}
$$

This diagram incorporates the idea that 'conditional' $\times$ 'marginal' = 'joint'. This idea is illustrated in two examples: first in the simpler non-parametrized case, and later also in parametrized form.

The same idea can be expressed in other Kleisli categories, of the other monads mentioned above. But a more challenging issue is to transfer this approach to the quantum case. This constitutes the main part (and contribution) of this paper. We interpret the above triangle in the opposite of the category of C^*-algebras, with positive unital maps, using effects as predicates. In this quantum case the situation

becomes more subtle, and at this preliminary stage of investigation we only present a non-parametrized example, namely the Elitzur-Vaidman bomb tester [4].

Our work relates to the pre-existing literature as follows. Bub [1] interprets the projection postulate during a measurement as an instance of Bayesian updating of the quantum state. His formulas [1, (21) and (22)] for the special case that $B = B(\mathscr{H})$ (and the state is normal, which is satisfied automatically if $\dim \mathscr{H} < \infty$) agree with our formula (10).

We can see this as follows. Bub has, for $a, b \in \text{Proj}(\mathscr{H})$:

$$\mathbb{P}_\rho(b|a) = tr(\rho' b) \qquad \text{where} \qquad \rho' = \frac{a\rho a}{tr(a\rho a)}.$$

This can be rearranged:

$$\mathbb{P}_\rho(b|a) = tr\left(\frac{a\rho a}{tr(a\rho a)} b\right) = \frac{tr(a\rho a b)}{tr(a\rho a)} = \frac{tr(\rho(aba))}{tr(\rho a^2)} = \frac{tr(\rho(aba))}{tr(\rho a)}.$$

If we reinterpret the ρs as maps $B(\mathscr{H}) \to \mathbb{C}$, i.e. normal states, we get:

$$\mathbb{P}_\rho(b|a) = \frac{\rho(aba)}{\rho(a)}$$

We now see this agrees with (10).

There is also the quantum conditional probability definition of Leifer and Spekkens [10] — expressed graphically in [3]. This work is based on the probabilistic case where, instead of being expressed in terms of probabilities of predicates, the conditional probability is formulated in [10] using a random variable that completely determines the elements of the underlying probability space. This seems to lead to a different formula from ours, but precise comparison is left to future work.

2   Discrete probability, categorically

To describe finite discrete probabilities categorically one uses the distribution monad $\mathscr{D} \colon \textbf{Sets} \to \textbf{Sets}$. It maps a set X to the set $\mathscr{D}(X)$ of probability distributions over X, which we describe as formal finite convex sums:

$$\textstyle\sum_i r_i |x_i\rangle \qquad \text{where} \qquad x_i \in X \text{ and } r_i \in [0,1] \text{ satisfy } \textstyle\sum_i r_i = 1.$$

We use the "ket" notation $|-\rangle$ to distinguish elements $x \in X$ and their occurrences in formal sums. Each function $f \colon X \to Y$ gives a function $\mathscr{D}(f) \colon \mathscr{D}(X) \to \mathscr{D}(Y)$, where:

$$\mathscr{D}(f)\left(\textstyle\sum_i r_i |x_i\rangle\right) = \textstyle\sum_i r_i |f(x_i)\rangle.$$

The unit $\eta \colon X \to \mathscr{D}(X)$ of this distribution monad $\mathscr{D}$ sends $x \in X$ to the singleton/Dirac distribution $\eta(x) = 1|x\rangle$. The multiplication $\mu \colon \mathscr{D}^2(X) \to \mathscr{D}(X)$ is given by:

$$\mu\left(\textstyle\sum_i r_i |\varphi_i\rangle\right) = \textstyle\sum_{i,j}(r_i s_{ij})|x_{ij}\rangle \qquad \text{if} \qquad \varphi_i = \textstyle\sum_j s_{ij}|x_{ij}\rangle.$$

Like for any monad, one can form the Kleisli category $\mathscr{K}\ell(\mathscr{D})$. In this case we get the category of sets and stochastic matrices, as the objects of $\mathscr{K}\ell(\mathscr{D})$ are sets, and its maps $X \to Y$ are functions $X \to \mathscr{D}(Y)$. The unit function $\eta \colon X \to \mathscr{D}(X)$ is then the identity map $X \to X$ in $\mathscr{K}\ell(\mathscr{D})$. Composition

of $f: X \to Y$ and $g: Y \to Z$ in $\mathcal{K}\ell(\mathcal{D})$ yields a map $g \circ f: X \to Z$, which, as a function $X \to \mathcal{D}(Z)$ is given by $g \circ f = \mu \circ \mathcal{D}(g) \circ f$. Explicitly:

$$(g \circ f)(x) = \sum_{i,j}(r_i s_{ij})|z_{ij}\rangle \quad \text{if} \quad f(x) = \sum_i r_i|y_i\rangle \quad \text{and} \quad g(y_i) = \sum_j s_{ij}|z_{ij}\rangle.$$

There is a forgetful functor $\mathcal{K}\ell(\mathcal{D}) \to \mathbf{Sets}$, sending X to $\mathcal{D}(X)$ and f to $\mu \circ \mathcal{D}(f)$. It has a left adjoint $\mathscr{F}: \mathbf{Sets} \to \mathcal{K}\ell(\mathcal{D})$ which is the identity on objects and sends f to $\eta \circ f$.

Products and coproducts of sets, with their projections π_i and coprojections κ_i are written as:

$$X \xleftarrow{\pi_1} X \times Y \xrightarrow{\pi_2} Y \qquad\qquad X \xrightarrow{\kappa_1} X + Y \xleftarrow{\kappa_2} Y$$

There are associated tuples $\langle f,g \rangle: Z \to X \times Y$ and cotuples $[h,k]: X + Y \to Z$. The empty product is a singleton set, typically written as 1, and the empty coproduct is the empty set 0.

The category $\mathcal{K}\ell(\mathcal{D})$ inherits these coproducts $(+,0)$ from $\mathbf{Sets}$, with coprojections $\mathscr{F}(\kappa_i) = \eta \circ \kappa_i$, and cotupling $[f,g]$ as in $\mathbf{Sets}$. The products $(\times, 1)$ from $\mathbf{Sets}$ form a tensor product — not a cartesian product — on $\mathcal{K}\ell(\mathcal{D})$; hence we write $\otimes$ in $\mathcal{K}\ell(\mathcal{D})$ for $\times$. But because the tensor unit 1 is also final in $\mathcal{K}\ell(\mathcal{D})$, since $\mathcal{D}(1) \cong 1$, we have a tensor with projections in $\mathcal{K}\ell(\mathcal{D})$. We shall write $\pi_i: X_1 \otimes X_2 \to X_i$ for the resulting projections in $\mathcal{K}\ell(\mathcal{D})$, which are functions $\mathscr{F}(\pi_i) = \eta \circ \pi_i: X_1 \times X_2 \to \mathcal{D}(X_i)$. This forms the background for the following result. It uses *marginals*, which, for a Kleisli map $f: X \to Y_1 \otimes Y_2$ are obtained by post-composition $\pi_i \circ f = \mathcal{D}(\pi_i) \circ f: X \to Y_i$.

Lemma 1. *In $\mathcal{K}\ell(\mathcal{D})$ there is a bijective correspondence:*

$$\frac{X \xrightarrow{f} Y}{X \xrightarrow{g} X \otimes Y \text{ with } \pi_1 \circ g = \mathrm{id}_X}$$

Proof The condition $\pi_1 \circ g = \mathrm{id}$ means that if $g(x) = \sum_i r_i|(x_i, y_i)\rangle$, then $x_i = x$ for all i. Hence g corresponds to a function $X \to \mathcal{D}(Y)$. $\qquad\qquad\square$

Below we shall write $gr(f): X \to \mathcal{D}(X \times Y)$ for this "graph" map corresponding to $f: X \to \mathcal{D}(Y)$, where, explicitly,

$$gr(f)(x) = \sum_i r_i|(x, y_i)\rangle \quad \text{if} \quad f(x) = \sum_i r_i|y_i\rangle. \tag{2}$$

Now that we have a category $\mathcal{K}\ell(\mathcal{D})$ to model probabilistic transitions, we add a logic to it. Categorically this takes the form of a functor, or indexed category, *Pred*: $\mathcal{K}\ell(\mathcal{D}) \to \mathbf{EMod}^{\mathrm{op}}$, where $\mathbf{EMod}$ is the category of effect modules (see *e.g.* [9]). We briefly explain the relevant definitions.

To start, let $M = (M, \oslash, 0)$ be a partial commutative monoid, where $\oslash$ is a partial operation $M \times M \to M$ that is commutative and associative, in a suitable sense, and has 0 has unit element. One can think of the unit interval $[0,1]$ with addition $+$ and 0. Such an M is called an *effect algebra* if there is a unary "orthocomplement" operation $(-)^\perp: M \to M$ satisfying both:

- $x^\perp \in E$ is the unique element in E with $x \oslash x^\perp = 1$, where $1 = 0^\perp$;

- $x \oslash 1$ is defined only when $x = 0$.

On $[0,1]$ on has $r^\perp = 1 - r$ as orthocomplement.

A morphism of effect algebras $f: M \to N$ is a function between the underlying sets satisfying $f(1) = 1$ and: if $x \oslash y$ is defined, then so is $f(x) \oslash f(y)$, and $f(x \oslash y) = f(x) \oslash f(y)$. This yields a category which we write as $\mathbf{EA}$.

An *effect module* is an effect algebra M with a (total) scalar multiplication $r \bullet x \in M$ for $r \in [0,1]$ and $x \in M$, preserving $\varovee$ in both coordinates separately, satisfying $1 \bullet x = x$, and $r \bullet (s \bullet x) = (r \cdot s) \bullet x$. A map of effect modules is a map of effect algebras that preserves the scalar multiplication. This yields a category **EMod**.

For a set $X \in \mathcal{K}\ell(\mathcal{D})$ we define $Pred(X) = [0,1]^X$, the set of fuzzy predicates on X. There are a true and false predicates, $1 = x \mapsto 1$ and $0 = x \mapsto 0$. For two fuzzy predicates $p, q \in [0,1]^X$ a sum $p \varovee q \in [0,1]^X$ exists if $p(x) + q(x) \leq 1$, for all $x \in X$; then $(p \varovee q)(x) = p(x) + q(x)$. Further, there is an orthocomplement operation $p^{\perp}(x) = 1 - p(x)$. One has, for instance, $p^{\perp\perp} = p$ and $p \varovee p^{\perp} = 1$. There is also a scalar multiplication on fuzzy predicates: for $r \in [0,1]$ one defines $(r \bullet p)(x) = r \cdot p(x)$.

Each Kleisli map $f: X \to Y$ yields a functor $f^{\sharp} = Pred(f): Pred(Y) \to Pred(X)$, which is commonly called substitution; it is given by:

$$f^{\sharp}(q)(x) = \sum_i r_i \cdot q(y_i) \qquad \text{if} \qquad f(x) = \sum_i r_i | y_i \rangle. \tag{3}$$

This $f^{\sharp}$ is a map of effect modules $[0,1]^Y \to [0,1]^X$.

For each set X there is a special predicate $\Omega_X \in Pred(X + X) = [0,1]^{X+X}$, namely $\Omega(\kappa_1 x) = 1$ and $\Omega(\kappa_2 x) = 0$. For each predicate $p \in [0,1]^X$ there is a *characteristic* map $char_p: X \to X + X$ in $\mathcal{K}\ell(\mathcal{D})$ with $char_p^{\sharp}(\Omega) = p$. This characteristic map is defined as convex sum:

$$char_p(x) = p(x)|\kappa_1 x\rangle + p^{\perp}(x)|\kappa_2 x\rangle = p(x)|\kappa_1 x\rangle + (1 - p(x))|\kappa_2 x\rangle. \tag{4}$$

These characteristic maps play an important role below, and are further discussed in [6].

3   Conditional discrete probability

This section reviews classical conditional probability, in the discrete case. A simple example is first described in standard terminology, and then reformulated in categorical form, by using the fuzzy predicate logic *Pred*: $\mathcal{K}\ell(\mathcal{D}) \to$ **EMod**$^{\text{op}}$ over the Kleisli category of the (discrete) probability monad $\mathcal{D}$. The example is extended to "parametrized" form, and again formulated in categorical terms.

Example 2. *In this first illustration we describe a simple situation, involving a set of genders $G = \{M, W\}$ with a distribution $f = \frac{2}{3}|M\rangle + \frac{1}{3}|W\rangle$ of men and women. Assume that the probability of having long hair is $\frac{3}{10}$ for men and $\frac{8}{10}$ for women. More formally this is written as $\mathbb{P}[\ell\,|M] = \frac{3}{10}$ and $\mathbb{P}[\ell\,|W] = \frac{8}{10}$, where ℓ stands for 'long hair'. We now ask ourselves the typical conditional probability question: suppose we see someone with long hair, what is the probability that the person is a man/woman?*

One then proceeds as follows. The joint probabilities are given by:

$$\mathbb{P}[M \wedge \ell] = \frac{2}{3} \cdot \frac{3}{10} = \frac{1}{5} \qquad \mathbb{P}[W \wedge \ell] = \frac{1}{3} \cdot \frac{8}{10} = \frac{4}{15}.$$

And the marginal probability of seeing long hair is:

$$\mathbb{P}[\ell] = \mathbb{P}[M \wedge \ell] + \mathbb{P}[W \wedge \ell] = \frac{1}{5} + \frac{4}{15} = \frac{7}{15}.$$

We then obtain the required conditional probabilities:

$$\mathbb{P}[M\,|\,\ell] = \frac{\mathbb{P}[M \wedge \ell]}{\mathbb{P}[\ell]} = \frac{\frac{1}{5}}{\frac{7}{15}} = \frac{3}{7} \qquad \mathbb{P}[W\,|\,\ell] = \frac{\mathbb{P}[W \wedge \ell]}{\mathbb{P}[\ell]} = \frac{\frac{4}{15}}{\frac{7}{15}} = \frac{4}{7}.$$

By construction we have "conditional · marginal = joint" since $\mathbb{P}[M\,|\,\ell] \cdot \mathbb{P}[\ell] = \mathbb{P}[M \wedge \ell]$, as suggested in (1). It will be elaborated below.

We now reformulate this example in categorical form. The distribution $f = \frac{2}{3}|M\rangle + \frac{1}{3}|W\rangle$ corresponds to a map $f\colon 1 \to G$ in the Kleisli category $\mathscr{K}\ell(\mathscr{D})$, where $1 = \{0\}$ is the final (singleton) set and $G = \{M, W\}$ is the two-element set of genders. In this correspondence we identify f with the value $f(0) \in \mathscr{D}(G)$, for the sole element $0 \in 1$. The likelihood of having long hair corresponds to a fuzzy predicate $\ell \in Pred(G) = [0,1]^G$ on the set G, given by $\ell(M) = \frac{3}{10}$, $\ell(W) = \frac{8}{10}$. The associated characteristic map $char_\ell\colon G \to G + G$ in $\mathscr{K}\ell(\mathscr{D})$ is, according to (4):

$$char_\ell(M) = \tfrac{3}{10}|\kappa_1 M\rangle + \tfrac{7}{10}|\kappa_2 M\rangle \qquad char_\ell(W) = \tfrac{8}{10}|\kappa_1 W\rangle + \tfrac{2}{10}|\kappa_2 W\rangle.$$

The left κ_1-option in the coproduct $G + G$ thus captures the probability that the predicate (in this case ℓ) is true, and the right κ_2-option is for false. The composite map $char_\ell \circ f\colon 1 \to G + G$ in the Kleisli category now describes the *joint* probability:

$$
\begin{aligned}
char_\ell \circ f &= \textstyle\sum_{z \in G+G} \left(\sum_{g \in G} f(g) \cdot char_\ell(g)(z) \right) |z\rangle \\
&= \tfrac{2}{3} \cdot \tfrac{3}{10}|\kappa_1 M\rangle + \tfrac{2}{3} \cdot \tfrac{7}{10}|\kappa_2 M\rangle + \tfrac{1}{3} \cdot \tfrac{8}{10}|\kappa_1 W\rangle + \tfrac{1}{3} \cdot \tfrac{2}{10}|\kappa_2 W\rangle \\
&= \tfrac{1}{5}|\kappa_1 M\rangle + \tfrac{7}{15}|\kappa_2 M\rangle + \tfrac{4}{15}|\kappa_1 W\rangle + \tfrac{1}{15}|\kappa_2 W\rangle \\
&= \mathbb{P}[M \wedge \ell]|\kappa_1 M\rangle + \mathbb{P}[M \wedge \ell^\perp]|\kappa_2 M\rangle + \mathbb{P}[W \wedge \ell]|\kappa_1 W\rangle + \mathbb{P}[W \wedge \ell^\perp]|\kappa_2 W\rangle.
\end{aligned}
$$

The substituted predicate $f^\sharp(\ell) \in Pred(1) = [0,1]^1 \cong [0,1]$, defined in (3), gives the *marginal* probability $\Pr[\ell] \in [0,1]$:

$$f^\sharp(\ell) = \textstyle\sum_{g \in G} f(g) \cdot \ell(g) = f(M) \cdot \ell(M) + f(W) \cdot \ell(W) = \tfrac{2}{3} \cdot \tfrac{3}{10} + \tfrac{1}{3} \cdot \tfrac{8}{10} = \tfrac{7}{15}.$$

The conditional probabilities can be organized into two maps $f|\ell, f|\ell^\perp \colon 1 \to G$ in $\mathscr{K}\ell(\mathscr{D})$, namely:

$$
\begin{aligned}
f|\ell &= \tfrac{1}{\mathbb{P}[\ell]}\left(\mathbb{P}[M \wedge \ell]|M\rangle + \mathbb{P}[W \wedge \ell]|W\rangle \right) = \tfrac{15}{7}\left(\tfrac{1}{5}|M\rangle + \tfrac{4}{15}|W\rangle \right) = \tfrac{3}{7}|M\rangle + \tfrac{4}{7}|W\rangle \\
f|\ell^\perp &= \tfrac{1}{\mathbb{P}[\ell^\perp]}\left(\mathbb{P}[M \wedge \ell^\perp]|M\rangle + \mathbb{P}[W \wedge \ell^\perp]|W\rangle \right) = \tfrac{15}{8}\left(\tfrac{7}{15}|M\rangle + \tfrac{1}{15}|W\rangle \right) = \tfrac{7}{8}|M\rangle + \tfrac{1}{8}|W\rangle.
\end{aligned}
$$

The first distribution $f|\ell$ gives the probabilities for men and women under the assumption that you see long hair; similarly, $f|\ell^\perp$ gives these probabilities if you do not see long hair.

The final observation is that these two maps $f|\ell$ and $f|\ell^\perp$ make the following triangle in the Kleisli category $\mathscr{K}\ell(\mathscr{D})$ commute, like in pattern (1):

$$
\begin{array}{ccc}
 & 1 & \\
{\scriptstyle char_{f^\sharp(\ell)}}\swarrow & & \searrow{\scriptstyle char_\ell \circ f} \\
1 + 1 \xrightarrow[\ (f|\ell) + (f|\ell^\perp)\ ]{} & & G + G
\end{array}
\tag{5}
$$

This simple hair example is "non-parametrized", in the sense that in the above triangle we have the final/singleton set 1 at the top. More generally, we can start with a Kleisli map $f\colon X \to Y$ and predicate on $X \otimes Y$.

Example 3. *Suppose we now have two different countries A, B which have different gender distributions and different distributions of long and short hair. We will use $C = \{A, B\}$ as the set of countries, with*

given gender distributions captured by a Kleisli map $f: C \to \mathcal{D}(G)$, where $G = \{M, W\}$ is the set of genders like in Example 2:

$$f(A) = \tfrac{9}{20}|M\rangle + \tfrac{11}{20}|W\rangle \qquad f(B) = \tfrac{1}{2}|M\rangle + \tfrac{1}{2}|W\rangle.$$

The probabilities of having long hair depend on both C and G and are already given in some way, formalised via predicate $L \in \mathrm{Pred}(C \otimes G)$ with:

$$L(A, M) = \tfrac{1}{10} \qquad L(B, M) = \tfrac{2}{10} \qquad L(A, W) = \tfrac{8}{10} \qquad L(B, W) = \tfrac{9}{10}.$$

When instead of a proper set C we had a trivial (singleton) set 1, getting the joint probability distribution was a simple matter of composition in $\mathcal{K}\ell(\mathcal{D})$. However, we now have $f: C \to G$ but $\mathrm{char}_L: C \otimes G \to C \otimes G + C \otimes G$, so that is out of the question. To solve this we will define a map $j = f \wedge L: C \to G + G$ as the composite of the following maps, using the definition of gr from equation (2):

$$
\begin{array}{ccc}
C & \xrightarrow{\quad j = f \wedge L \quad} & G + G \\
\mathrm{gr}(f) \downarrow & & \uparrow \mathscr{F}(\pi_2 + \pi_2) \\
C \otimes G & \xrightarrow{\quad \mathrm{char}_L \quad} & C \otimes G + C \otimes G
\end{array}
\qquad (6)
$$

This produces the correct joint probability. First the case of $A \in C$:

$$
\begin{aligned}
j(A) &= \big(\mathcal{D}(\pi_2 + \pi_2) \circ (\mathrm{char}_L \circ \mathrm{gr}(f))\big)(A) \\
&= \big(\mathcal{D}(\pi_2 + \pi_2) \circ \mathrm{char}_L^{\$}\big)\big(\textstyle\sum_{g \in G} f(A)(g)|A, g\rangle\big) \\
&= \mathcal{D}(\pi_2 + \pi_2)\big(\textstyle\sum_{g \in G} L(A, g) \cdot f(A)(g)|\kappa_1(A, g)\rangle + L^{\perp}(A, g) \cdot f(A)(g)|\kappa_2(A, g)\rangle\big) \\
&= \mathcal{D}(\pi_2 + \pi_2)\big(\tfrac{1}{10} \cdot \tfrac{9}{20}|\kappa_1(A, M)\rangle + \tfrac{9}{10} \cdot \tfrac{9}{20}|\kappa_2(A, M)\rangle \\
&\qquad\qquad + \tfrac{8}{10} \cdot \tfrac{11}{20}|\kappa_1(A, W)\rangle + \tfrac{2}{10} \cdot \tfrac{11}{20}|\kappa_2(A, W)\rangle\big) \\
&= \tfrac{9}{200}|\kappa_1 M\rangle + \tfrac{81}{200}|\kappa_2 M\rangle + \tfrac{88}{200}|\kappa_1 W\rangle + \tfrac{22}{200}|\kappa_2 W\rangle.
\end{aligned}
$$

Similarly one obtains the distribution $j(B) \in \mathcal{D}(G + G)$, namely:

$$j(B) = \tfrac{2}{20}|\kappa_1 M\rangle + \tfrac{8}{20}|\kappa_2 M\rangle + \tfrac{9}{20}|\kappa_1 W\rangle + \tfrac{1}{20}|\kappa_2 W\rangle.$$

We now calculate the marginal probability of L, getting rid of the dependence on G. To do this, we compute $\mathrm{gr}(f)^{\sharp}(L) \in \mathrm{Pred}(C)$ as:

$$
\begin{aligned}
\mathrm{gr}(f)^{\sharp}(L)(A) &= \textstyle\sum_{(c, g) \in C \times G} \mathrm{gr}(f)(A)(c, g) \cdot L(c, g) \\
&= \tfrac{9}{20} \cdot L(A, M) + \tfrac{11}{20} \cdot L(A, W) = \tfrac{9}{20} \cdot \tfrac{1}{10} + \tfrac{11}{20} \cdot \tfrac{8}{10} = \tfrac{97}{200}.
\end{aligned}
$$

In the same way, $\mathrm{gr}(f)^{\sharp}(L)(B) = \tfrac{11}{20}$. And obviously,

$$\mathrm{gr}(f)^{\sharp}(L^{\perp})(A) = \big(\mathrm{gr}(f)^{\sharp}(L)\big)^{\perp}(A) = 1 - \mathrm{gr}(f)^{\sharp}(L)(A) = \tfrac{103}{200},$$

and similarly $\mathrm{gr}(f)^{\sharp}(L^{\perp})(B) = \tfrac{9}{20}$.

With that done, we can now work out the conditional probabilities of a man or a woman given the country and that they had long hair. In other words, we are looking for a pair of maps $f\,|\,L, f\,|\,L^\perp \colon C \to G$ in $\mathscr{K}\ell(\mathscr{D})$ to fill in the following triangle:

$$
\begin{array}{ccc}
 & C & \\
{}^{char_{gr(f)^\sharp(L)}}\nearrow & & \searrow{}^{j=(\pi_2+\pi_2)\circ char_L\circ gr(f)} \\
C+C & \xrightarrow{\ (f\,|\,L)+(f\,|\,L^\perp)\ } & G+G
\end{array}
\tag{7}
$$

As distribution $f\,|\,L(A) \in \mathscr{D}(G+G)$ we take:

$$
f\,|\,L(A) \;=\; \frac{j(A)\circ\kappa_1}{gr(f)^\sharp(L)(A)} \;=\; \tfrac{200}{97}\cdot\Big(\tfrac{9}{200}|M\rangle + \tfrac{88}{200}|W\rangle\Big) \;=\; \tfrac{9}{97}|M\rangle + \tfrac{88}{97}|W\rangle.
$$

This can be read as: in country A, if we see someone with long hair, the probability of this person being male (resp. female) is $\tfrac{9}{97}$ (resp. $\tfrac{88}{97}$). In the same way one gets $f\,|\,L(B) = \tfrac{2}{11}|M\rangle + \tfrac{9}{11}|W\rangle$. And in the two "negated" cases:

$$
f\,|\,L^\perp(A) \;=\; \tfrac{81}{103}|M\rangle + \tfrac{22}{103}|W\rangle \qquad f\,|\,L^\perp(B) \;=\; \tfrac{8}{9}|M\rangle + \tfrac{1}{9}|W\rangle.
$$

With these definitions it is easy to see that diagram (7) commutes.

Having completed an example, we move on to the general case for probabilities.

Theorem 4. *For a morphism $f \colon X \to Y$ in the Kleisli category $\mathscr{K}\ell(\mathscr{D})$ of the distribution monad $\mathscr{D}$, and for a predicate $\phi \in [0,1]^{X\times Y}$, there are conditional probability maps $f\,|\,\phi, f\,|\,\phi^\perp \colon X \to Y$ in $\mathscr{K}\ell(\mathscr{D})$ making the following triangle commute.*

$$
\begin{array}{ccc}
 & X & \\
{}^{char_{gr(f)^\sharp(\phi)}}\nearrow & & \searrow{}^{f\wedge\phi=(\pi_2+\pi_2)\circ char_\phi\circ gr(f)} \\
X+X & \xrightarrow{\ (f\,|\,\phi)+(f\,|\,\phi^\perp)\ } & Y+Y
\end{array}
$$

If $gr(f)^\sharp(\phi)(x) \in (0,1)$, for each $x \in X$, then both these maps $f\,|\,\phi$ and $f\,|\,\phi^\perp$ are uniquely determined.

Proof We define the functions $f\,|\,\phi, f\,|\,\phi^\perp \colon X \to \mathscr{D}(Y)$ on $x \in X$ as:

$$
(f\,|\,\phi)(x) = \sum_y \frac{(f\wedge\phi)(x)(\kappa_1 y)}{gr(f)^\sharp(\phi)(x)}|y\rangle \qquad (f\,|\,\phi^\perp)(x) = \sum_y \frac{(f\wedge\phi)(x)(\kappa_2 y)}{1-gr(f)^\sharp(\phi)(x)}|y\rangle.
$$

If $gr(f)^\sharp(\phi)(x) \in \{0,1\}$, we choose an arbitrary distribution instead.　$\square$

Here we have formulated conditional probability with respect to a single formula ϕ. It can be generalized to n-tests, which are sequences of formulas $\phi_1, \ldots, \phi_n$ with $\phi_1 \oslash \cdots \oslash \phi_n = 1$ (see also [6]). Then one gets n corresponding conditional maps $f\,|\,\phi_i$. In the situation of the above theorem we actually use a 2-test, given by ϕ and $\phi^\perp$.

4   Conditional probability for C^*-algebras

We shall write $\mathbf{Cstar}_{\mathrm{PU}}$ for the category of unital C^*-algebras (over the complex numbers $\mathbb{C}$) with positive unital maps, and $\mathbf{Cstar}_{\mathrm{MIU}} \hookrightarrow \mathbf{Cstar}_{\mathrm{PU}}$ for the subcategory where maps preserve multiplication (M), involution (I), and unit (U); such maps are usually called *-homomorphisms. In the present setting we assume all C^*-algebras have a unit 1. We write $\mathscr{Z}(A) \hookrightarrow A$ for the center of a C^*-algebra A, defined as usual as $\mathscr{Z}(A) = \{a \in A \mid \forall_{b \in A}.\, ab = ba\}$. This center forms a commutative (sub) C^*-algebra. Obviously, A itself is commutative iff $A = \mathscr{Z}(A)$.

We remark at this point that any map in $\mathbf{Cstar}_{\mathrm{PU}}$, when considered as a map of Banach spaces, is of norm 1. This is [11, corollary 1]. This is equivalent to saying that any positive unital map also preserves the norm.

The category $\mathbf{Cstar}_{\mathrm{PU}}$ has finite products, via direct sums $\oplus$ of vector spaces (*i.e.* cartesian products of the underlying sets). The operations are used pointwise. There are also tensor products of C^*-algebras. These are described in more detail in [12, section IV.4], but we outline them here. The C^*-tensors for two C^*-algebras A and B are obtained by taking the usual tensor of underlying vector spaces $A \otimes B$, defining a *-algebra structure as follows:

$$(a_1 \otimes b_1)(a_2 \otimes b_2) = (a_1 a_2) \otimes (b_1 b_2) \qquad (a \otimes b)^* = a^* \otimes b^*.$$

One then obtains a C^*-algebra by introducing a C^*-norm compatible with the *-algebra structure, and taking the completion. There are minimal and maximal, or injective and projective C^*-norms, but if A or B is finite dimensional these coincide [13, chapter XV, 1.4 - 1.6]. Therefore, as a simplifying assumption, all C^*-algebras in this paper will be taken as finite dimensional unless explicitly stated otherwise.

The positive cone of $A \otimes B$ contains the positive elements according to the multiplication and involution, *i.e.* $a \in A \otimes B$ is positive if $a = b^* b$ for some other element b. We note at this point that this cone is larger than the cone obtained by taking sums of elements $a \otimes b$ with $a \in A$ and $b \in B$ both positive. The effect of this is that no C^*-tensor is a functor on $\mathbf{Cstar}_{\mathrm{PU}}$. The maps that can be tensored are called completely positive and form a non-full subcategory $\mathbf{Cstar}_{\mathrm{cPU}} \hookrightarrow \mathbf{Cstar}_{\mathrm{PU}}$ with the same objects (see [12, section IV.3, and proposition IV.4.23]).

Since $A \otimes B$ is the completion of the algebraic tensor of A and B, the span of elements of the form $a \otimes b$ is dense, and in fact in our finite dimensional case $A \otimes B$ is just the span of such elements. We can define coprojections $\kappa_i : A_i \to A_1 \otimes A_2$ as follows:

$$\kappa_1(a) = a \otimes 1 \qquad\qquad\qquad \kappa_2(a) = 1 \otimes a,$$

where 1 is the unit of the C^*-algebra. It is simple to see these are MIU maps, and therefore in $\mathbf{Cstar}_{\mathrm{PU}}$.

(It is most natural to consider categories of C^*-algebras in opposite form. For instance, in [5] it is shown that the opposite $(\mathbf{CCstar}_{\mathrm{PU}})^{\mathrm{op}}$ of the category of commutative C^*-algebras with positive unital maps is equivalent to a Kleisli category, namely that of the "Radon" monad on compact Hausdorff spaces. This restricts to an equivalence between finite-dimensional commutative C^*-algebras and the subcategory $\mathscr{K}\ell_{\mathbb{N}}(\mathscr{D}) \hookrightarrow \mathscr{K}\ell(\mathscr{D})$ with natural numbers as objects. In opposite form, $(\mathbf{Cstar}_{\mathrm{cPU}})^{\mathrm{op}}$ has similar structure to the Kleisli category $\mathscr{K}\ell(\mathscr{D})$ used in the previous section, namely finite coproducts and tensors with projections.)

We are working towards a C^*-algebraic analogue of Lemma 1. But this requires some lemmas of its own. The following result is based on theorem 1 of [14].

Lemma 5. *If $f : A \to B$ is a map in $\mathbf{Cstar}_{MIU}$.*

(i) The algebra B is a bimodule of A under the left and right multiplications:

$$a \cdot b = f(a)b \qquad\qquad b \cdot a = bf(a).$$

*(ii) If a **Cstar**$_{PU}$ map $g : B \to A$ is a retraction of f, i.e. $g \circ f = \mathrm{id}_A$, then g is a map of bimodules:*

$$a_1 g(b) a_2 = g(a_1 \cdot b \cdot a_2) = g(f(a_1)bf(a_2)).$$

Proof For the first point, the unit and multiplication properties follow easily from those of f. For the second point we notice that, since f has a left inverse, it is a split monic and therefore is isomorphic to its image $f(A)$, a subalgebra of B. Then $f \circ g$ is a positive unital projection onto $f(A)$, and is therefore a projection of norm 1 in the sense of [14]. Applying [14, theorem 1, part 2] we have that $f \circ g$ is a bimodule map. Thus, if $f(a_1), f(a_2) \in f(A)$, and $b \in B$, then:

$$(f \circ g)\big(f(a_1)bf(a_2)\big) = f(a_1)(f \circ g)(b)f(a_2) = f(a_1 g(b) a_2),$$

the latter because f is a MIU-map. Applying the injectivity of f, we have, as required:

$$g(f(a_1)bf(a_2)) = a_1 g(b) a_2. \qquad\qquad \square$$

Lemma 6. *If A is a C^*-algebra, multiplication of an element by an element of the centre $\mathscr{Z}(A)$ is a MIU map $\mu : A \otimes \mathscr{Z}(A) \to A$.*

Proof Here is the definition of μ:

$$\mu\left(\textstyle\sum_i a_i \otimes z_i\right) = \textstyle\sum_i a_i z_i.$$

Since the multiplication is bilinear, this is well-defined. To show it preserves multiplication, it suffices to show it does so on basic tensors. We start with $\mu((a \otimes z)(b \otimes w)) = abzw$. Since z commutes with b, we can rearrange this to get $azbw = \mu(a \otimes z)\mu(b \otimes w)$. The preservation of involution and unit are routine arguments. $\qquad\square$

The following is the analogue of Lemma 1.

Lemma 7. *In **Cstar**$_{PU}$ there is a bijective correspondence:*

$$\frac{B \xrightarrow{\;f\;} \mathscr{Z}(A)}{A \otimes B \xrightarrow{\;\;g\;\;} A \text{ positive and unital, with } g \circ \kappa_1 = \mathrm{id}_A}$$

Of course, when A is commutative, the '$\mathscr{Z}$' can be dropped.

Like before we shall write $\mathrm{gr}(f) : A \otimes B \to A$ for the map corresponding to $f : B \to \mathscr{Z}(A)$, where $\mathrm{gr}(f) = \mu \circ (\mathrm{id}_A \otimes f)$, or on elements $\mathrm{gr}(f)(a \otimes b) = a \cdot f(b) = f(b) \cdot a$.

Proof Given $f : B \to \mathscr{Z}(A)$, in **Cstar**$_{PU}$, since $\mathscr{Z}(A)$ is commutative we can use [12, corollary IV.3.5] to show it is in **Cstar**$_{cPU}$, and hence $\mathrm{id}_A \otimes f$ is positive. It is unital, and hence in **Cstar**$_{PU}$ because $(\mathrm{id}_A \otimes f)(1 \otimes 1) = 1 \otimes f(1) = 1 \otimes 1$. By lemma 6, μ is in **Cstar**$_{MIU}$ and hence in **Cstar**$_{PU}$, so $\mu \circ (\mathrm{id}_A \otimes f)$ has the right type. To see it is a left inverse for κ_1:

$$(\mu \circ (\mathrm{id}_A \otimes f) \circ \kappa_1)(a) = \mu((\mathrm{id}_A \otimes f)(a \otimes 1)) = \mu(a \otimes f(1)) = \mu(a \otimes 1) = a.$$

as required.

Conversely, if $g : A \otimes B \to A$ is a map such that $g \circ \kappa_1 = \mathrm{id}_A$, then since κ_1 is an MIU map, Lemma 5 shows that g is a bimodule map. We take f to be $g \circ \kappa_2 : B \to A$. This appears at first to have the wrong type. However, if $a \in A$ and $b \in B$

$$af(b) = ag(1 \otimes b) = g(a \otimes 1 \cdot 1 \otimes b) = g(a \otimes b) = g(1 \otimes b \cdot a \otimes 1) = g(1 \otimes b)a = f(b)a.$$

Hence $f(b) \in \mathscr{Z}(A)$. It is left to the reader to check that the correspondences we have described are each other's inverses. $\square$

Also for C^*-algebras there is a logic $(\mathbf{Cstar}_{\mathrm{PU}})^{\mathrm{op}} \to \mathbf{EMod}^{\mathrm{op}}$ of effect modules. For each C^*-algebra A, its "effects" $[0,1]_A = \{a \in A \mid 0 \leq a \leq 1\}$ form an effect module. The sum $e \oslash d$ exists and is equal to $e + d$ if $e + d \leq 1$. The orthocomplement of e is $e^\perp = 1 - e$. Each positive unital map $f : A \to B$ forms an effect module map $f^\sharp : [0,1]_A \to [0,1]_B$ by restriction. Since each such map is determined by what it does on positive elements, we have a full and faithful functor $\mathbf{Cstar}_{\mathrm{PU}} \to \mathbf{EMod}$, see [5] for more details.

The product $\times$ of C^*-algebras forms a coproduct $+$ in $(\mathbf{Cstar}_{\mathrm{PU}})^{\mathrm{op}}$. When we work in this opposite category, we shall thus use the coproduct notation. There is a special effect $\Omega = (1,0) \in [0,1]_{A+A} = ([0,1]_A)^2$. For each effect $e \in [0,1]_A$ there is a choice of characteristic map, in the non-commutative case. In the category $(\mathbf{Cstar}_{\mathrm{PU}})^{\mathrm{op}}$ one can define:

$$A \xrightarrow{\ char_e\ } A + A \qquad \text{as} \qquad (a,a') \longmapsto \sqrt{e} \cdot a \cdot \sqrt{e} + \sqrt{(1-e)} \cdot a' \cdot \sqrt{(1-e)}$$

We have that $char_e^\sharp(\Omega) = char_e(\Omega) = e$, and $char_e^\sharp(\Omega^\perp) = char_e(\Omega^\perp) = e^\perp$. This property replaces that of [7], being a section of ∇_A, which is not satisfied in the non-commutative case.

Lemma 8. *If a, b are positive elements of a C^*-algebra, then aba is positive.*

Proof Since b is positive, $b = p^*p$ for some p. So

$$aba = ap^*pa = a^*p^*pa = (pa)^*pa$$

and aba is positive. (We have used that all positive elements are self-adjoint.) $\square$

Corollary 9. *The map $char_e$ is positive and unital.*

Proof Effects e and $e^\perp = 1 - e$ are positive, and so are their square roots. Hence the previous lemma makes $char_e$ positive. The proof of unitality is straightforward. $\square$

We can now give a proof that this definition of characteristic map, for commutative C^*-algebras, coincides with the monadic definition for the Radon monad under the equivalence $\mathscr{K}\ell(\mathscr{R}) \simeq \mathbf{CCstar}_{\mathrm{PU}}$ from [5, Theorem 2]. In both cases we can start with a compact Hausdorff space X, and take the corresponding C^*-algebra to be $C(X)$, the C^*-algebra of continuous functions $X \to \mathbb{C}$. A predicate is a continuous map to the unit interval, $e \in \mathbf{CHaus}(X, [0,1])$. We have two possible characteristic maps

$$char_e : C(X) \times C(X) \to C(X)$$

and

$$char_e' : X \to \mathscr{R}(X + X)$$

which, following (4), is defined as

$$char_e'(x) = e(x)\delta_{\kappa_1 x} + (1 - e(x))\delta_{\kappa_2 x},$$

where the δs are Dirac delta measures. This may equivalently be defined, given a function $f \in C(X + X)$, as

$$char_e'(x)(f) = e(x) \cdot f(\kappa_1 x) + (1 - e(x)) \cdot f(\kappa_2 x).$$

Theorem 10. *Under the equivalence $\mathscr{C}_\mathscr{R} : \mathscr{K}\ell(\mathscr{R}) \xrightarrow{\sim} \mathbf{CCstar}_{PU}$ from [5, Theorem 2], char_e coincides with char'_e, which is to say, given $a_1, a_2 \in C(X)$*

$$\mathscr{C}_\mathscr{R}(\mathrm{char}'_e)([a_1, a_2]) = \mathrm{char}_e(a_1, a_2).$$

Proof. Consider the right hand side. We have that

$$\mathrm{char}_e(a_1, a_2) = \sqrt{e}a_1\sqrt{e} + \sqrt{1-e}a_2\sqrt{1-e},$$

which by commutativity can be rewritten as

$$\mathrm{char}_e(a_1, a_2) = ea_1 + (1-e)a_2.$$

Now let $x \in X$, and we can see

$$
\begin{aligned}
\mathscr{C}_\mathscr{R}(\mathrm{char}'_e)([a_1, a_2])(x) &= \mathrm{char}'_e(x)([a_1, a_2]) \\
&= e(x) \cdot [a_1, a_2](\kappa_1 x) + (1 - e(x)) \cdot [a_1, a_2](\kappa_2 x) \\
&= e(x) \cdot a_1(x) + (1 - e(x)) \cdot a_2(x) \\
&= (ea_1 + (1-e)a_2)(x)
\end{aligned}
$$

and so $\mathscr{C}_\mathscr{R}(\mathrm{char}'_e)([a_1, a_2]) = \mathrm{char}_e(a_1, a_2)$ as required. $\qquad\square$

We are now in a position to describe a setting for conditional probability for C^*-algebras. In order to maximize the analogy with the situation in the previous section — involving the Kleisli category $\mathscr{K}\ell(\mathscr{D})$ — we work in the opposite category $(\mathbf{Cstar}_{PU})^{\mathrm{op}}$. There tensors have projections π_i.

Assume we have a map $f \colon \mathscr{L}(A) \to B$ and an effect $e \in [0,1]_{A \otimes B}$. Then we can form the marginal and total probability maps as follows.

- Via the graph map $gr(f) \colon A \to A \otimes B$ obtained in Lemma 7 we can substitute and get $gr(f)^\sharp(e) \in [0,1]_A$ and form the characteristic map $\mathrm{char}_{gr(f)^\sharp(e)} \colon A \to A + A$.

- We can also form the joint probability $f \wedge e$ as the composition, in $(\mathbf{Cstar}_{PU})^{\mathrm{op}}$:

$$f \wedge e = \left(A \xrightarrow{gr(f)} A \otimes B \xrightarrow{\mathrm{char}_e} (A \otimes B) + (A \otimes B) \xrightarrow{\pi_2 + \pi_2} B + B \right)$$

The conditional probability maps $f \mid e, f \mid e^\perp \colon A \to B$ in $(\mathbf{Cstar}_{PU})^{\mathrm{op}}$ then fit in the triangle:

$$
\begin{array}{ccc}
 & A & \\
\mathrm{char}_{gr(f)^\sharp(e)} \swarrow & & \searrow f \wedge e \\
A + A & \xrightarrow{\;\;(f \mid e) + (f \mid e^\perp)\;\;} & B + B
\end{array}
\tag{8}
$$

Theorem 11. *If $gr(f)^\sharp(e)$ and $gr(f)^\sharp(e^\perp)$ are invertible, then the maps $f \mid e$ and $f \mid e^\perp$ exist and are unique. The formulas for each are:*

$$f \mid e(b) = \frac{1}{\sqrt{gr(f)^\sharp(e)}} \cdot (f \wedge e)(b, 0) \cdot \frac{1}{\sqrt{gr(f)^\sharp(e)}}$$

$$f \mid e^\perp(b) = \frac{1}{\sqrt{gr(f)^\sharp(1-e)}} \cdot (f \wedge e)(0, b) \cdot \frac{1}{\sqrt{gr(f)^\sharp(1-e)}}$$

Proof The proof has three steps. First we show that these maps are in $\mathbf{Cstar}_{\mathrm{PU}}$. Then we show they make the diagram commute. Finally, we show they are the unique such maps.

But first, we remark that for any positive invertible element a of a C^*-algebra A, the spectrum of a is a closed subset of $(0,\infty)$ and so we may use continuous functional calculus (see [12, definition I.4.7]) to take $\frac{1}{\sqrt{a}}$, so the positive square root of a is invertible.

These maps can be seen to be unital because the inverse square roots on either side cancel with the square roots.

To show that they are positive, let a be a positive element of A. Then:

$$f\,|\,e(a) \;=\; \frac{1}{\sqrt{\mathrm{gr}(f)^\sharp(e)}} \cdot \mathrm{gr}(f)^\sharp\big(char_e\big((\kappa_2 \times \kappa_2)(a,0)\big)\big) \cdot \frac{1}{\sqrt{\mathrm{gr}(f)^\sharp(e)}}.$$

If we show that $(\kappa_2 \times \kappa_2)(a,0)$ is positive, then it will follow from corollary 9, lemma 7 and lemma 8 that $f\,|\,e(a)$ is positive. Since $(\kappa_2 \times \kappa_2)(a,0) = (1 \otimes a,0)$ and a is positive, it can be written $a = b^*b$, so that we have:

$$(1 \otimes b^*b,0) = ((1 \otimes b^*)(1 \otimes b),0) = ((1 \otimes b)^*(1 \otimes b),0),$$

which is positive. Thus $f\,|\,e(a)$ is positive. The case of $f\,|\,e^\perp$ is similar.

To show that these maps $f\,|\,e,\ f\,|\,e^\perp$ make the diagram (8) commute, let $(b_1,b_2) \in B \times B$, where we are reconsidering the diagram in $\mathbf{Cstar}_{\mathrm{PU}}$. Then

$$
\begin{aligned}
&char_{\mathrm{gr}(f)^\sharp(e)}\big((f\,|\,e \times f\,|\,e^\perp)(b_1,b_2)\big)\\
&= char_{\mathrm{gr}(f)^\sharp(e)}\big((f\,|\,e)(b_1),(f\,|\,e^\perp)(b_2)\big)\\
&= char_{\mathrm{gr}(f)^\sharp(e)}\left(\frac{1}{\sqrt{\mathrm{gr}(f)^\sharp(e)}} \cdot (f \wedge e)(b_1,0) \cdot \frac{1}{\sqrt{\mathrm{gr}(f)^\sharp(e)}}, \frac{1}{\sqrt{\mathrm{gr}(f)^\sharp(1-e)}} \cdot (f \wedge e)(0,b_2) \cdot \frac{1}{\sqrt{\mathrm{gr}(f)^\sharp(1-e)}} \right)\\
&= (f \wedge e)(b_1,0) + (f \wedge e)(0,b_2)\\
&= (f \wedge e)((b_1,0) + (0,b_2))\\
&= (f \wedge e)(b_1,b_2)
\end{aligned}
$$

To show the uniqueness, suppose we have g_1,g_2 such that $char_{\mathrm{gr}(f)^\sharp(e)} \circ (g_1 \times g_2) = f \wedge e$, and let $(b_1,b_2) \in B \times B$. Then if $b_1 \in B$ we have:

$$char_{\mathrm{gr}(f)^\sharp(e)}\big((g_1 \times g_2)(b_1,0)\big) \;=\; (f \wedge e)(b_1,0).$$

Rearranging the left hand side, we get:

$$char_{\mathrm{gr}(f)^\sharp(e)}(g_1(b_1),0) \;=\; \sqrt{\mathrm{gr}(f)^\sharp(e)} \cdot g_1(b_1) \cdot \sqrt{\mathrm{gr}(f)^\sharp(e)} + 0 \;=\; (f \wedge e)(b_1,0).$$

By invertibility of $\sqrt{\mathrm{gr}(f)^\sharp(e)}$, we have that:

$$g_1(b_1) \;=\; \frac{1}{\sqrt{\mathrm{gr}(f)^\sharp(e)}} \cdot (f \wedge e)(b_1,0) \cdot \frac{1}{\sqrt{\mathrm{gr}(f)^\sharp(e)}},$$

as required. The g_2 case is similar. □

For ease of application in the example in the next section, we specialize quantum conditional probability to when there is no parametrization, taking $A = \mathbb{C}$ in Theorem 11. Then f is a state, considered as

a map $B \to \mathbb{C}$ in **Cstar**$_{\text{PU}}$. We can use the isomorphism of $B \otimes \mathbb{C} \cong B$ to view e as a predicate on B. Then $f \wedge e = f \circ char_e$, much like in (5). This means diagram (8) becomes in $(\textbf{Cstar}_{\text{PU}})^{\text{op}}$:

$$
\begin{array}{ccc}
 & \mathbb{C} & \\
 {}^{char_{f^{\sharp}(e)}}\swarrow & & \searrow^{char_e \circ f} \\
 \mathbb{C} + \mathbb{C} & \xrightarrow{\quad (f|e) + (f|e^{\perp}) \quad} & B + B
\end{array}
\tag{9}
$$

Corollary 12. *For a state* $f : B \to \mathbb{C}$ *and an effect* $e \in [0,1]_B$*, if* $f^{\sharp}(e) = f(e) \neq 0, 1$ *then the conditional states* $f|e, f|e^{\perp}$ *in* (9) *exist and are unique, and can be given by the formulas:*

$$
f|e(b) = \frac{f(\sqrt{e}b\sqrt{e})}{f(e)} \qquad f|e^{\perp}(b) = \frac{f(\sqrt{1-e}b\sqrt{1-e})}{f(1-e)}.
\tag{10}
$$

Proof Since $\mathbb{C}$ is a field, 0 is the only non-invertible element. Since $f^{\sharp}(1-e) = 1 - f^{\sharp}(e)$ as f is unit-preserving and linear, $f^{\sharp}(e) \neq 0, 1$ implies that $f^{\sharp}(e)$ and $f^{\sharp}(1-e)$ are invertible. We then apply Theorem 11 and use the commutativity of $\mathbb{C}$. $\square$

Since this definition of conditional probability applies to effects, not just projections, it in fact works as a definition of conditional expectation for postiive operators less than or equal to 1. As such, it may be related to the definition of conditional expectation given in [2]. However, we have used C^*-algebras here and a non-commutative version of the definition in that paper is more naturally formulated in the setting of W^*-algebras, so we leave relating the two to future work.

5 Example

As an example, we use the bomb tester of [4]. Suppose some bombs exist that explode if a single photon is absorbed by a detector attached to it. However, some of these bombs are duds, and the photon passes through the detector unaltered, failing to explode the bomb, if this is the case. We want to find out which of the bombs are which. If we try to test a bomb to see if it explodes, we seemingly can only keep the bomb if it turns out to be a dud, as the bomb will explode if tested with a photon, the smallest amount of light that we could use. However it is shown in [4] that this is not the case, and a bomb tester can be built. We reformulate this to use our framework for quantum conditional probability.

The set-up is similar to a Mach-Zehnder interferometer, as observed in [4]. A photon passes through a semi-silvered mirror, where the bomb is in the path of one branch, the photon is reflected from two mirrors to hit a second semi-silvered mirror, after which there are two detectors. This can be seen in figure 5. We represent the system with the following C^*-algebra:

$$
A = A_E \otimes A_P \otimes A_B = C(\{\text{L},\text{D}\}) \otimes B(\ell^2(\{\uparrow,\rightarrow,\emptyset\})) \otimes B(\ell^2(\{0,1\})).
$$

The status of the bomb being Live or a Dud is treated as classical, the direction or absence of a photon is represented by a 3-dimensional Hilbert space and the state of the bomb as unexploded or exploded is treated as a 2-dimensional Hilbert space. We use the shortened names A_E, A_P and A_B for these algebras, the letters standing for explosivity, photon, and bomb respectively. All together, the C^*-algebra is $2 \times 3^2 \times 2^2 = 72$-dimensional.

The mirrors (semi-silvered or fully silvered) act only on A_P. They are maps of the form $U^* \cdot - \cdot U$ for U a unitary from $\ell^2(\{\uparrow, \rightarrow, \emptyset\})$ to itself. On basis vectors, the semi-silvered mirrors' unitaries, U_S, are:

$$| \rightarrow \rangle \mapsto \tfrac{1}{\sqrt{2}} | \rightarrow \rangle + \tfrac{1}{\sqrt{2}} | \uparrow \rangle \qquad | \uparrow \rangle \mapsto \tfrac{1}{\sqrt{2}} | \rightarrow \rangle - \tfrac{1}{\sqrt{2}} | \uparrow \rangle \qquad | \emptyset \rangle \mapsto | \emptyset \rangle.$$

And the fully silvered mirrors' unitaries, U_F, are:

$$| \rightarrow \rangle \mapsto | \uparrow \rangle \qquad | \uparrow \rangle \mapsto | \rightarrow \rangle \qquad | \emptyset \rangle \mapsto | \emptyset \rangle.$$

The reader may verify that these are unitary and that $U_S U_F U_S | \rightarrow \rangle = | \rightarrow \rangle$, so that in the absence of a bomb the photon always comes out to the right. Already, a stark difference is apparent from what would happen if the semi-silvered mirrors acted probabilistically.
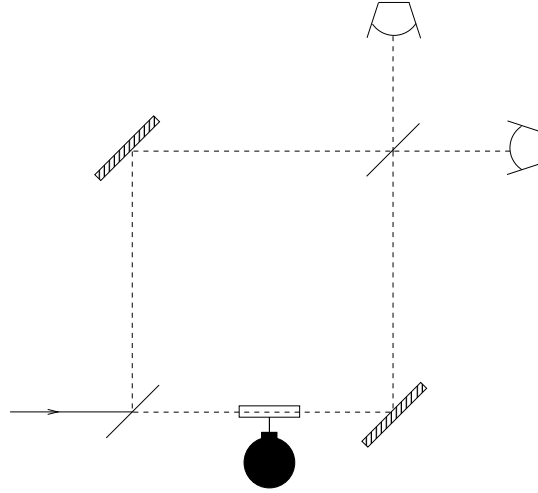

Figure 1: The bomb tester

The explosion of the bomb can be represented as a unitary U_B in $A_P \otimes A_B$. To do this, we allow the bomb to spontaneously unexplode, emitting a rightward photon. This does not affect the results as the map is never evaluated in this state.

$$| \uparrow 0 \rangle \mapsto | \uparrow 0 \rangle \qquad | \uparrow 1 \rangle \mapsto | \uparrow 1 \rangle \qquad | \rightarrow 0 \rangle \mapsto | \emptyset 1 \rangle \quad \text{(bomb explodes)}$$
$$| \rightarrow 1 \rangle \mapsto | \rightarrow 1 \rangle \qquad | \emptyset 0 \rangle \mapsto | \emptyset 0 \rangle \qquad | \emptyset 1 \rangle \mapsto | \rightarrow 0 \rangle \quad \text{(bomb unexplodes)}.$$

We can then describe the dynamics of the exploding bomb on $A = A_E \otimes A_P \otimes A_B$. For ease of use later, we actually use the Schrödinger picture. We have the states $\delta_x \in \mathbf{Cstar}_{\mathrm{PU}}(A_E, \mathbb{C}) \cong \mathscr{D}(\{L, D\})$, for $x \in \{L, D\}$. It is the usual delta measure, which is a map $A_E \rightarrow \mathbb{C}$. For a state $\rho : A_P \otimes A_B \rightarrow \mathbb{C}$ the δ_x determine the dynamics, as in:

$$\delta_L \otimes \rho \mapsto \delta_L \otimes \rho(U_B^* \cdot - \cdot U_B) \qquad \delta_D \otimes \rho \mapsto \delta_D \otimes \rho.$$

As we can see, whether the bomb can explode or not depends on whether we have an L or D state in the first component of A.

The way conditional probability is supposed to work is that for $x \in A$ considered to be a random variable, we have some $f : A \to \mathbb{C}$ such that $f(x) = \mathbb{E}(x)$, and $f | e(x) = \mathbb{E}(x | e = 1)$, as in diagram (9). To work out f, we start off with the initial state:

$$f_0 = \left(A_E \otimes (A_P \otimes A_B) \xrightarrow{\;(\frac{1}{2}\delta_D + \frac{1}{2}\delta_L) \otimes \langle \to 0| \; - \; |\to 0\rangle\;} \mathbb{C} \otimes \mathbb{C} \cong \mathbb{C} \right).$$

In other words, we start with an even probability of a live bomb or a dud, and with the photon moving to the right, before it hits the first mirror.

Then we show how f_0 changes under the dynamics. To save space, we replace the ket of the state with an ellipsis $(\dots)$:

$$\text{first mirror} \quad (\frac{1}{2}\delta_D + \frac{1}{2}\delta_L) \otimes (\frac{1}{\sqrt{2}}\langle \to 0| + \frac{1}{\sqrt{2}}\langle \uparrow 0|) \; - \; |\cdots\rangle$$

$$\text{light hits bomb} \quad \frac{1}{2}\delta_D \otimes (\frac{1}{\sqrt{2}}\langle \to 0| + \frac{1}{\sqrt{2}}\langle \uparrow 0|) \; - \; |\cdots\rangle + \frac{1}{2}\delta_L \otimes (\frac{1}{\sqrt{2}}\langle \emptyset 1| + \frac{1}{\sqrt{2}}\langle \uparrow 0|) \; - \; |\cdots\rangle$$

$$\text{opaque mirrors} \quad \frac{1}{2}\delta_D \otimes (\frac{1}{\sqrt{2}}\langle \uparrow 0| + \frac{1}{\sqrt{2}}\langle \to 0|) \; - \; |\cdots\rangle + \frac{1}{2}\delta_L \otimes (\frac{1}{\sqrt{2}}\langle \emptyset 1| + \frac{1}{\sqrt{2}}\langle \to 0|) \; - \; |\cdots\rangle$$

$$\text{last mirror} \quad \frac{1}{2}\delta_D \otimes \langle \to 0| \; - \; |\to 0\rangle + \frac{1}{2}\delta_L \otimes (\frac{1}{\sqrt{2}}\langle \emptyset 1| + \frac{1}{2}\langle \to 0| + \frac{1}{2}\langle \uparrow 0|) \; - \; |\cdots\rangle$$

We shall write f for this last state $A \to \mathbb{C}$. Now that it is fixed, consider the situation in which the bomb did not explode and the photon was detected going up. This is captured by the following effect.

$$e = 1_{A_E} \otimes |\uparrow 0\rangle\langle \uparrow 0| \in [0,1]_A.$$

We wish to calculate the probability that the bomb is a dud, given e, *i.e.* given that the bomb did not explode and the photon was detected going up. In symbols this is $\mathbb{P}(\text{Dud} | e)$, *i.e.* $\mathbb{E}(\chi_D \otimes 1_{A_P \otimes A_B} | e)$, where $\chi_D \in C(\{L,D\})$ is the obvious indicator function. In triangle diagram (9) we wish to calculate the conditional state $f | e$ with input event $b = \chi_D \otimes 1_{A_P \otimes A_B} \in A$.

We apply the formula (10) for $f | e(b) \in \mathbb{C}$. To do this, we first calculate $f^\sharp(e) = f(e)$, using the abbreviation $|\psi\rangle = \frac{1}{\sqrt{2}}|\emptyset 1\rangle + \frac{1}{2}|\to 0\rangle + \frac{1}{2}|\uparrow 0\rangle$. First $f^\sharp(e)$:

$$f(e) = f(1_{A_E} \otimes |\uparrow 0\rangle\langle \uparrow 0|) = \frac{1}{2}\delta_D(1) \otimes \langle \to 0|\uparrow 0\rangle\langle \uparrow 0|\to 0\rangle + \frac{1}{2}\delta_L(1) \otimes \langle \psi|\uparrow 0\rangle\langle \uparrow 0|\psi\rangle$$
$$= 0 + \frac{1}{2}(\frac{1}{2}\langle \uparrow 0|\uparrow 0\rangle\langle \uparrow 0|\uparrow 0\rangle\frac{1}{2}) = \frac{1}{8}.$$

Since the effect e is a projection, it is its own positive square root. Therefore we have:

$$char_e(b,0) = \sqrt{e}b\sqrt{e} = ebe = (1_{A_E} \otimes |\uparrow 0\rangle\langle \uparrow 0|)(\chi_D \otimes 1_{A_P \otimes A_B})(1_{A_E} \otimes |\uparrow 0\rangle\langle \uparrow 0|) = \chi_D \otimes |\uparrow 0\rangle\langle \uparrow 0|.$$

We may now substitute all of these values into (10) and get $f | e(b)$:

$$f | e(b) = \frac{f(\chi_D \otimes |\uparrow 0\rangle\langle \uparrow 0|)}{\frac{1}{8}} = 8\left(\frac{1}{2}\langle \to 0|\uparrow 0\rangle\langle \uparrow 0|\to 0\rangle + 0 \right) = 0.$$

Thus if an upward-moving photon is detected and the bomb did not explode, the probability that it is a dud is 0, and it must be live. This gives a way to get live bombs without exploding them. Note that this contradicts a commonly stated notion about quantum mechanics, that one cannot observe something without affecting it[1], as in this case we have a way to use quantum mechanics to observe something without affecting it in a way that we would have had to do classically.

---

[1]This is intended to refer to the projection that occurs in a measurement

6   Conclusions

In this paper we have given a categorical formulation of conditional probability. It involves a triangle-fill property, where the condition is a predicate from an associated predicate logic, formalized via an indexed category of effect modules. It is shown that this formulation gives the familiar classical notion of conditional probability, when interpreted in the Kleisli category of the distribution monad.

Next, the formulation can also be used in a quantum setting, given by the category of finite-dimensional C^*-algebras. We have presented a general "parametrized" formulation, but our main example, the bomb tester, only involves the non-parametrized case. Further clarification is needed, in this general parametrized case, also in relation to other approaches in the literature. Our approach has the advantage that it is based on a general categorical scheme, that can be instantiated in various settings.

Acknowledgements

References

[1] J. Bub (2007): *Quantum Probabilities as Degrees of Belief.* Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics 38(2), pp. 232 – 254, doi:10.1016/j.shpsb.2006.09.002.

[2] P. Chaput, V. Danos, P. Panangaden & G. Plotkin (2009): *Approximating Markov Processes by Averaging.* In S. Albers, A. Marchetti-Spaccamela, Y. Matias, S. Nikoletseas & W. Thomas, editors: *Automata, Languages and Programming, Lecture Notes in Computer Science* 5556, Springer Berlin Heidelberg, pp. 127–138, doi:10.1007/978-3-642-02930-1_11.

[3] B. Coecke & R. Spekkens (2012): *Picturing classical and quantum Bayesian inference.* Synthese 186(3), pp. 651–696, doi:10.1007/s11229-011-9917-5.

[4] A. Elitzur & L. Vaidman (1993): *Quantum mechanical interaction-free measurements.* Foundations of Physics 23(7), pp. 987–997, doi:10.1007/BF00736012.

[5] R. Furber & B. Jacobs (2013): *From Kleisli Categories to Commutative C^*-Algebras: Probabilistic Gelfand Duality.* In R. Heckel & S. Milius, editors: *Algebra and Coalgebra in Computer Science, Lecture Notes in Computer Science* 8089, Springer Berlin Heidelberg, pp. 141–157, doi:10.1007/978-3-642-40206-7_12.

[6] B. Jacobs (2013): *On Block Structures in Quantum Computation.* Electronic Notes in Theoretical Computer Science 298(0), pp. 233 – 255, doi:10.1016/j.entcs.2013.09.016. Proceedings of the Twenty-ninth Conference on the Mathematical Foundations of Programming Semantics, {MFPS} {XXIX}.

[7] B. Jacobs (2015): *New Directions in Categorical Logic for Classical, Probabilistic and Quantum Logic.* Logical Methods in Computer Science 11(3).

[8] B. Jacobs & J. Mandemaker (2012): *The Expectation Monad in Quantum Foundations.* In B. Jacobs, P. Selinger & B. Spitters, editors: *Quantum Physics and Logic (QPL) 2011, Elect. Proc. in Theor. Comp. Sci.* 95, pp. 143–182, doi:10.4204/EPTCS.95.12.

[9] B. Jacobs & J. Mandemaker (2016, to appear): *Relating Operator Spaces via Adjunctions.* In J. Chubb Reimann, V. Harizanov & A. Eskandarian, editors: *Logic and Algebraic Structures in Quantum Computing and Information*, Lect. Notes in Logic, Cambridge Univ. Press. Available at `http://arxiv.org/abs/1201.1272`.

[10] M. Leifer & R. Spekkens (2013): *Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference.* Phys. Rev. A 88(5), p. 052130.

[11] B. Russo & H.A. Dye (1966): *A Note on Unitary Operators in C*-algebras. Duke Math. J.* 33, pp. 413–416, doi:10.1215/S0012-7094-66-03346-1.

[12] Masamichi Takesaki (2002): *Theory of Operator Algebra, volume I. Encyclopedia of Mathematical Sciences* 124, Springer Verlag, doi:10.1007/978-1-4612-6188-9.

[13] Masamichi Takesaki (2003): *Theory of Operator Algebras, volume III. Encyclopedia of Mathematical Sciences* 127, Springer Verlag, doi:10.1007/978-3-662-10453-8.

[14] Jun Tomiyama (1957): *On the Projection of Norm One in W*-algebras. Proceedings of the Japan Academy* 33(10), pp. 608–612, doi:10.3792/pja/1195524885.