SOME USEFUL CATEGORIES FOR PROBABILITY THEORY

Contents

1. Introduction. 1
2. The Giry monad and a category of stochastic maps 1
3. The category of probability spaces. 3
4. A first nice category of Bayesian processes. 9
5. The relative entropy of **PrftStat**-morphisms. 11
References 12
References 12

1. Introduction.

2. The Giry monad and a category of stochastic maps

Denote by **Meas** the category of measurable spaces and measurable functions. Central to our category of stochastic maps is the Giry monad, which was introduced by Giry [6] and based on an idea of Lawvere [7].

Definition 2.1 (Stochastic map)**.** A *stochastic map* between two measurable space $f\colon (X,\Sigma) \to (X',\Sigma')$ is a function

$$f\colon X \times \Sigma' \to [0,1]$$

such that

- $f(x,-)\colon \Sigma' \to [0,1]$ is a probability measure for each $x \in X$, and
- $f(-,A')\colon X \to [0,1]$ is a measurable function, with respect to the Borel σ-algebra on $[0,1]$, for each $A' \in \Sigma'$.

Other names found in the literature are Markov kernel or stochastic kernel.

Composition of two stochastic maps

$$(X,\Sigma) \xrightarrow{f} (X',\Sigma') \xrightarrow{g} (X'',\Sigma'')$$

is given by the function

$$f;g\colon X \times \Sigma'' \to [0,1], \quad (x,A'') \mapsto \int_{X'} g(-,A'')df(x,-).$$

This is the $f(x,-)$-expectation of the random variable $g(-,A'')$. Because $0 \le g(-,A'') \le 1$ and $f(x,X') = 1$, this integral takes a value in $[0,1]$. Panangaden in [8, Prop. 3.2] used the monotone convergence theorem to show associativity. A simple calculation shows that the Dirac function $\delta\colon (X,\Sigma) \to (X,\Sigma)$,

$$X \times \Sigma \to [0,1], \quad (x,A) \mapsto \begin{cases} 0, & \text{if } x \notin A, \\ 1, & \text{if } x \in A. \end{cases}$$

serves as the identity.

Definition 2.2 (The category of stochastic maps)**.** Denote by **Stoch** the category of measurable spaces and stochastic maps.

An important class of stochastic maps are the *deterministic* stochastic maps. These are the maps $X \times \Sigma' \to [0,1]$ that only take values 0 or 1. For example, the identities of **Stoch**. However, the main examples of deterministic maps that we care about arise from measurable functions.

Example 2.3 (Deterministic maps in **Stoch**)**.** Given a measurable map $f\colon (X,\Sigma) \to (X',\Sigma')$, denote by δ_f the stochastic map defined by

$$X \times \Sigma' \to [0,1], \quad (x,A') \mapsto \begin{cases} 0 & \text{if } f(x) \notin A', \\ 1 & \text{if } f(x) \in A'. \end{cases}$$

This is measurable in the first variable: the inverse image of a Borel set $B \in [0,1]$ is

$$\begin{cases} \emptyset & \text{if } 0,1 \notin B, \\ f^{-1}(A') & \text{if } 0 \notin B \text{ and } 1 \in B, \\ f^{-1}(X \setminus A') & \text{if } 0 \in B \text{ and } 1 \notin B, \text{ and} \\ X & \text{if } 0,1 \in B. \end{cases}$$

It is clear that δ_f is a probability measure in the second variable. From this point on, we will denote by δ_f the deterministic stochastic map associated to a measurable function f. Not all deterministic maps are obtained this way.

Example 2.4. Suppose that X is uncountable and Σ is the σ-algebra generated by the singletons of the power-set on X. There is a measure m on (X,Σ) assigning no measure to countable sets and a measure of 1 to those sets whose complements are countable. This gives a deterministic map $(\{*\},\{\emptyset,\{*\}\}) \to (X,\Sigma)$

$$\{*\} \times \Sigma \to [0,1], \quad (*,B) \mapsto m(B).$$

This does not arise from any measurable function $\{*\} \to X$ because there are countable sets in Σ containing the image of $*$, and so must have no measure.

The act of obtaining a deterministic stochastic map from a measurable function is functorial. That is, there is an identity-on-objects functor

$$\delta\colon \mathbf{Meas} \to \mathbf{Stoch} \tag{1}$$

given by $f \mapsto \delta_f$. As tempting as it is, do not think of **Meas** as a subcategory of **Stoch** because δ is not faithful as illustrated by the next example.

Example 2.5. Consider a measurable space $(X, \{\emptyset, X\})$ for any set X. Any set function $f\colon * \to X$ induces a measurable function $f\colon (*, \{\emptyset, *\}) \to (X, \{\emptyset, X\})$. Hitting this with δ gives us the stochastic map

$$\delta_f\colon * \times \{\emptyset, X\} \to [0,1]$$

sending $\emptyset \mapsto 0$ and $X \mapsto 1$. Notice that f plays no role in the definition of δ_f. It turns out that $\delta_f = \delta_g$ for any pair of measurable functions $f, g\colon (*, \{\emptyset, *\}) \to (X, \{\emptyset, X\})$.

Indeed, all indiscrete measurable spaces are become isomorphic under δ, as are measurable maps between them.

This next proposition and its immediate corollary tell us more about how δ treats measurable functions on indiscrete measure spaces.

Proposition 2.6. *For any set X, the image of a measurable function*

$$f\colon (*, \{\emptyset, *\}) \to (X, \{\emptyset, X\})$$

under δ is an isomorphism.

Proof. A simple calculation will show that $\delta_f^{-1} = \delta_g$ where g is the measurable function $(X, \{\emptyset, X\}) \to (*, \{\emptyset, *\})$. □

Corollary 2.7. *All indiscrete measurable spaces are isomorphic in* **Stoch**.

Corollary 2.8. *Any indiscrete measurable space are terminal in* **Stoch**.

Proof. Any stochastic map $X \times \{\emptyset, *\} \to [0,1]$ is forced to assign $(x, \varepsilon) \mapsto 0$ and $(x, *) \mapsto 1$ for every $x \in X$. This is sufficient because of 2.7. □

Let's learn a bit more about isomorphisms in **Stoch**. We'll first learn that, in general, stochastic maps do not have much respect for the structure of measurable spaces. Indeed, isomorphisms in **Stoch** are quite restrictive.

Proposition 2.9 ([5, Prop. 2.16])**.** *Let $k\colon (X, \Sigma) \to (X', \Sigma')$ be an isomorphism in* **Stoch**. *Then k is deterministic.*

Now, we will move towards the next category of interest which will continue the trend and use **Stoch** in our construction. To be sure, we will actually focus on a particular subcategory of **Stoch** to ensure this category is sufficiently nice.

3. The category of probability spaces.

Denote the **Stoch**-object $(*, \{\emptyset, *\})$ by 1. Think for a second about a stochastic map $1 \to (X, \Sigma)$. This is a function $* \times \Sigma \to [0,1]$ that is measurable in the first variable and a probability measure in the second. Since there is only a single element in the first factor, we are really dealing with a probability measure $\Sigma \to [0,1]$. In fact, we will take the definition of a *probability measure space* to be an arrow $1 \to (X, \Sigma)$ in **Stoch**. This leads us to our next category.

Definition 3.1. The category of probability spaces and probability measure preserving functions **Prob** is the under-category $1/\mathbf{Stoch}$.

The category we really care about, however, is not **Prob** itself, but is closely related. But, for the desired Bayesian properties to be present, we will need to appropriately restrict to certain types of probability spaces.

3.1. **Restricting to a convenient subcategory.** Let's start with a *joint probability measure*, that is, a probability measure $j\colon 1 \to (X \times X', \Sigma \otimes \Sigma')$ on a product space. From this we obtain the *marginal probability measures* $\delta_\pi j$ and $\delta_{\pi'} j$ on (X, Σ) and (X', Σ'), respectively, where π and π' are the **Meas**-projections from the product $(X \times X', \Sigma \otimes \Sigma')$.

The relationship between joint measure j and its marginals, which we'll denote p, p', is not symmetric. While knowing j is sufficient to tell you the marginals, we cannot, in general, determine j from the marginals alone. The extra information contained in j is hidden in some sort of relationship between p and p'. Namely, this relationship is given by a **Prob**-morphism between the marginals. That is a diagram

$$\begin{tikzcd} & 1 \arrow[dl, "p"'] \arrow[dr, "p'"] & \\ (X, \Sigma) \arrow[rr, "f"] & & (X', \Sigma') \end{tikzcd}$$

induces a joint measure $j\colon 1 \to (X \times X', \Sigma \otimes \Sigma')$ whose marginals are p and p'. This justifies us changing viewpoints between joint probability measures (as favored in [1]) and morphisms in **Prob** (as we will do).

Now, let's start to restrict out attention to a select class of probability spaces. We will be content to remain unmotivated as to the reason for this restriction for the time being. It will become clear when we introduce the category **Stat** of Bayesian processes in Section 4.

Definition 3.2 (Perfect Probability Measures)**.** A probability space (X, Σ, p) is *perfect* if for any measurable function $f\colon X \to \mathbb{R}$, there is a Borel set $E \subseteq f(X)$ such that $p(*, f^{-1}(E)) = p(*, X) = 1$. A family of measures $\{p_i\colon 1 \to (X, \Sigma)\}$ on X is called *equiperfect* if, given f as before, there is a single Borel set E such that $p_i(*, f^{-1}(E)) = p_i(*, X) = 1$.

The rough idea of a perfect probability measure space (X, Σ, p) is that, whenever p is measurably pushed-forward via $f\colon X \to \mathbb{R}$ to a measure on $\mathbb{R}$, then every Σ-set is p-approximately a Borel set $f^{-1}(B)$.

The class of perfect probability measures is quite large. It is not difficult to show that any probability space on a discrete measure is perfect. In particular, any finite probability space is perfect. Additionally, we obtain a large collection of important examples, namely Radon probability spaces.

Example 3.3. Recall that a Radon probability measure p on a Hausdorff topological space X with the Borel σ-algebra is one such that, for any Borel set A, $p(A)$ is the supremum of $p(K)$ over compact sets K contained in A. Radon probability spaces are perfect. Indeed, by Lusin's theorem (see any measure theory text), there is a closed set E_n such that $p(X \setminus E_n) < \frac{1}{2n}$ and $f|_{E_n}$ is continuous. Let K_n be a compact set contained in E_n such that $p(X \setminus K_n) < \frac{1}{2n}$. Then $f(K_n)$ is compact and $p(X \setminus f^{-1}(f(K_n))) < \frac{1}{n}$. Denoting $A \coloneqq \cup_n f(K_n)$, we have that A is Borel and $p(X \setminus f^{-1}(A)) = 0$.

In particular, any probability measure on the Borel σ-algebra of a Polish space is a Radon measure.

The theorem (discussed in [2, Thm. 2.2]) below collects some nice properties of perfect probability measures whose proofs can be found in [3] and [9].

Theorem 3.4. *Let (X, Σ) and (X', Σ') be measurable spaces.*

(a) Any probability measure with values in $\{0, 1\}$ is perfect.
(b) Given probability measures $p, q \colon 1 \to (X, \Sigma)$, such that $q \ll p$, then p perfect implies q perfect.
(c) Any restriction of a perfect probability measure to a σ-subalgebra is perfect.
(d) Pushforwards of perfect probability measure along measurable functions are perfect.
(e) A joint probability is perfect if and only if its marginals are.
(f) Let $f \colon (X, \Sigma) \to (X', \Sigma')$ be a stochastic map that is measurable in the first factor and let p be a perfect probability measure on (X, Σ). Then the assignment

$$A' \mapsto \int_X f(x, A')dp$$

is a perfect probability measure if and only if $\{f(x, -) : x \in X\}$ is equiperfect p-almost everywhere.

Definition 3.5. The subcategory **PStoch** of **Stoch** has as objects, the countably generated measurable spaces and as arrows, stochastic maps that form a family of equiperfect probability measures in the second variable.

Composition in **PStoch** is well-defined: given a **PStoch**-maps $f \colon (X, \Sigma) \to (Y, \Sigma')$, $g \colon (Y, \Sigma') \to (Z, \Sigma'')$, a measurable map $h \colon Z \to \mathbb{R}$, and a Borel set $E \subseteq h(Z)$ satisfying the equiperfect condition of g, then

$$g \circ f(x, C) = \int_Y g(-, h^{-1}(E))df(x, -) = f(x, Y) = 1.$$

EDIT BETWEEN HERE AND NUMBER 1

To talk about probability measures in our new context, we can define **PrftProb** to be the under-category $1/$**PrftStoch**.

The reason we only consider countably generated measurable spaces is for the following theorems.

Theorem 3.6. *While in the context of the category* **PrftProb***, if given a joint probability* $j \colon 1 \to (X \times X', \Sigma \otimes \Sigma')$ *with marginals* p *and* p'*, there is an arrow* f *such that*

$$\begin{tikzcd} & 1 \arrow[dl, "p"'] \arrow[dr, "p'"] & \\ (X,\Sigma) & & (X',\Sigma') \arrow[ll, dashed, "f"] \end{tikzcd}$$

and $j(A \times A') = \int_{A'} f(-, A) dp'(*, -)$*. Moreover,* f *is unique* p'*-almost everywhere.*

In fact, we can use this theorem to show that this f actually factors through the product, as shown in the next theorem.

Theorem 3.7. *While in the context of the category* **PrftProb***, if given a joint probability* $j \colon 1 \to (X \times X', \Sigma \otimes \Sigma')$ *with marginals* p *and* p'*, there are arrows* f *and* g *such that*

$$\begin{tikzcd} & 1 \arrow[dl, bend right, "p"'] \arrow[d, "j"] \arrow[dr, bend left, "p'"] & \\ & (X \times X', \Sigma \otimes \Sigma') \arrow[dl, "\delta_\pi"'] \arrow[dr, "\delta_{\pi'}"] & \\ (X,\Sigma) \arrow[ur, dashed, bend right=10, "f"'] & & (X',\Sigma') \arrow[ul, dashed, bend left=10, "g"] \end{tikzcd}$$

The statement and proofs of these theorems can be found in [3] and [4]. In [4, Theorem 6], Faden uses a very different language to talk about these things. In particular, he describes Theorem **??** by showing an equivalence, in the case of countably generated measurable spaces, between a perfect probability measure and, what he calls, having the product regular conditional probability property. In [4, Theorem 1], Faden shows that having the product regular conditional probability property implies the existence of the map f from Theorem **??**. Hopefully, it's now clear why we required **PrftProb** to have countably generated measurable spaces.

Without diverging too much from our topic, let's say a couple of words about regular conditional probability. When working with finite spaces, we define conditional probability, that is the probability of x given y, to be

$$p(x|y) = p(x \cap y)/p(y).$$

This is called Bayes formula. Of course, it only makes sense when $p(y) \neq 0$, but this is fine since you'd only ask for $p(x|y)$ when $p(y) \neq 0$. The infinite case is more subtle, particularly when doing probability theory over arbitrary measurable spaces. Here, singletons will typically have no measure, rendering this definition nearly useless. Regular conditional probability measures allow us to attach a meaningful value to events, even if the given set has

no measure. However, regular conditional probabilities do not always exist, and so we need to work in a context where we are sure that they do exist.
NUMBER ONE

3.2. **A variation on the Giry monad.** The functor δ (see equation (1)) is important for more than just providing a wealth of **Stoch**-morphisms. It is half of the adjunction from which the Giry monad arises!

Definition 3.8 (Giry monad)**.** The *Giry monad* $G\colon \mathbf{Meas} \to \mathbf{Meas}$ maps a measurable space (X, Σ) to the set GX of all probability measures on (X, Σ) paired with the smallest σ-algebra so that the evaluation maps

$$\mathrm{ev}_A\colon GX \to [0,1], \quad m \mapsto m(A)$$

are measurable for every $A \in \Sigma$. To turn G into a monad, we equip it with unit

$$\eta_{(X,\Sigma)}\colon (X,\Sigma) \to G(X,\Sigma),$$

that sends x to its point measure and the multiplication

$$\mu_{(X,\Sigma)}\colon GG(X,\Sigma) \to G(X,\Sigma), \quad \mu_{(X,\Sigma)}(m)(A) = \int_{GX} \mathrm{ev}_A dm$$

for each probability measure $m\colon G(X,\Sigma) \to [0,1]$.

Along with introducing this monad, Giry [6] also showed that **Stoch** is the Kleisli category of the Giry monad. This is well known enough, so we will instead discuss a slight variation: the perfect Giry monad. To construct the perfect Giry monad, we will adjust the adjoint pair making the Giry monad.

The first functor of our adjoint pair is simply the evident restriction of δ, which we will still denote in the same manner.

Lemma 3.9. *There is an identity on objects functor $\delta\colon$ **CGMeas** $\to$ **PStoch** given $f \mapsto \delta_f$.*

Proof. This follows from Theorem 3.4 parts (a) and (f). □

Lemma 3.10. *Define $\varepsilon\colon$ **PStoch** $\to$ **CGMeas** as follows. For any measurable space (X,Σ), let $\varepsilon(X,\Sigma)$ be the measurable space consisting of the set $\varepsilon(X)$ of all perfect probability measures on (X,Σ) equipped with the weakest σ-algebra so that the evaluation map*

$$\mathrm{ev}_A\colon \varepsilon(X) \to [0,1], \quad p \mapsto p(A)$$

is measurable for every $A \in \Sigma$. Given a perfect stochastic map $f\colon (X,\Sigma) \to (X',\Sigma')$, define $\varepsilon f\colon \varepsilon(X,\Sigma) \to \varepsilon(X',\Sigma')$ by $p \mapsto f \circ p$. Then ε is a functor.

Proof. First, we show that $\varepsilon(X,\Sigma)$ is countably generated. Recall that the Borel σ-algebra on $[0,1]$ is generated by sets $[0,a]$, for rational points a in the unit interval. Let $\{A_i\}_I$ be a countable collection generating Σ. Then $\varepsilon(\Sigma)$ is generated by $\{\mathrm{ev}_{A_i}^{-1}([0,a])\}$ as i runs through I and a through rationals in $[0,1]$.

Next, we show that εf is well-defined and measurable. The former holds because $f \circ p$ is a **PStoch**-morphism. To show that latter, start by taking a generator $\mathrm{ev}_B^{-1}[0,a]$ of $\varepsilon\Sigma'$. This generator consists of all perfect probability measures $p' \colon \Sigma' \to [0,1]$ such that $p'(B) \leq a$. Then $(\varepsilon f)^{-1}(\mathrm{ev}_B^{-1}[0,a])$ consists of all probability measures $p \colon \Sigma \to [0,1]$ such that

$$(\varepsilon f)(p)(B) = f \circ p(B) = \int_X f(-,B)dp \leq a.$$

Because $f(-,B)$ is measurable, choose a non-decreasing sequence f_n of simple functions converging uniformly to $f(-,B)$. Then

$$(\varepsilon f)^{-1}(\mathrm{ev}_B^{-1}[0,a]) = \bigcap_{n=1}^{\infty} \left\{ p : \int_X f_n dp \leq a \right\}. \tag{2}$$

Observe that for any $A \in \Sigma$, the function

$$\int_X \chi_A d(-) \colon \varepsilon(X,\Sigma) \to [0,1], \quad p \mapsto \int_X \chi_A dp = p(A)$$

is exactly the evaluation map ev_A, hence measurable by construction. It follows that each $\int_X f_n d(-)$ is a linear combination of the evaluation functions, hence is measurable. Therefore (2) is a countable intersection of measurable functions and so belongs to Σ.

We've shown that ε is well-defined. It remains to show that it is actually a functor. Composition follows from associativity of stochastic maps:

$$\varepsilon(gf)(p) = (gf)p = g(fp) = \varepsilon(g)\varepsilon(f)(p).$$

And identities are preserved: $\varepsilon(\mathrm{id})(p) = \int_X \mathrm{id}(-,-)dp$ is the measure that maps A to $\int_X \mathrm{id}(-,A)dp = \int_A dp = p(A)$. □

Lemma 3.11. *There are natural transformations*

(a) $\eta \colon 1 \to \varepsilon\delta$ *made of measurable functions*

$$\eta_{(X,\Sigma)} \colon (X,\Sigma) \to \varepsilon(X,\Sigma)$$

assigning to each $x \in X$, the probability measure $A \mapsto \chi_A(x)$, where χ_A is the characteristic function of A; and

(b) $\mathrm{ev} \colon \delta\varepsilon \to 1$ *made of stochastic maps*

$$\mathrm{ev}_{(X,\Sigma)} \colon \varepsilon(X,\Sigma) \to (X,\Sigma)$$

assigning $(p,A) \mapsto p(A)$.

Theorem 3.12. *The data $(\delta,\varepsilon,\eta,\mathrm{ev})$ form an adjunction.*

Proof. Consider the composite

$$\varepsilon(X,\Sigma) \xrightarrow{\eta_{(X,\Sigma)}} \varepsilon\varepsilon(X,\Sigma) \xrightarrow{\varepsilon(\mathrm{ev}_{(X,\Sigma)})} \varepsilon(X,\Sigma). \tag{3}$$

Denote by p' the image of a perfect probability measure p under $\eta_{(X,\Sigma)}$. That is, p' is the **PStoch**-map $p' \colon 1 \to \varepsilon(X,\Sigma)$ given by $(*,B) \mapsto \chi_B(p)$.

Then $\varepsilon(\mathrm{ev})(p')$ is the **PStoch**-composite $\mathrm{ev} \circ p'$ given by

$$(*, A) \mapsto \int_{\varepsilon X} \mathrm{ev}(-, A) dp'(*, -) = p(A)$$

Hence (3) is the identity.

Next, the composite

$$(X, \Sigma) \xrightarrow{\delta(\eta_{(X,\Sigma)})} \varepsilon(X, \Sigma) \xrightarrow{\mathrm{ev}_{(X,\Sigma)}} (X, \Sigma). \tag{4}$$

is the **PStoch**-map $X \times \Sigma \to [0, 1]$ given by

$$(x, A) \mapsto \int_{\varepsilon X} \mathrm{ev}_{(X,\Sigma)}(-, A) d\delta(\eta_{(X,\Sigma)})(x, -) = \chi_A(x).$$

Hence (4) is the identity. □

Definition 3.13. The *perfect Giry monad* $G\colon \mathbf{CGMeas} \to \mathbf{CGMeas}$ is given by the adunction $(\delta, \varepsilon, \eta, \mathrm{ev})$.

Theorem 3.14. *The Kleisli category* $\mathbf{K}(G)$ *for the perfect Giry monad* G *is equivalent to* **PStoch**.

Proof. Define an identity on objects functor $F\colon \mathbf{PStoch} \to \mathbf{K}(G)$ by sending a stochastic map $f \in \mathbf{PStoch}((X, \Sigma), (X', \Sigma'))$, that is, $f\colon X \times \Sigma' \to [0, 1]$, to the measurable map

$$Ff\colon (X, \Sigma) \to G(X', \Sigma'), \quad x \mapsto f(x).$$

This functor is certainly essentially surjective and it is straightforward to check that it is faithful. It remains to show that F is full. Let $g\colon (X, \Sigma) \to G(X', \Sigma')$ be measurable. Define a stochastic map $\hat{g}\colon X \times \Sigma' \to [0, 1]$ by $\hat{g}(-, -) = g(-)(-)$. Given any $B \in \Sigma'$, $\hat{g}(-, B) = \mathrm{ev}_B \circ f$, so measurable (recall ev_B is measurable by construction). For each x, $\hat{g}(x, -) = g(x)(-)$ is perfect, but we require that the collection $\{\hat{g}(x, -) : x \in X\}$ is equiperfect. So, given a measurable function $h\colon X' \to \mathbb{R}$, for each $x \in X$ there is a Borel set E_x contained in $h(X')$ so that $\hat{g}(x, h^{-1}(E_x) = 1$. In case X is countable, then $E = \cup_x E_x$ is Borel and $\hat{g}(x, h^{-1}(E)) = 1$. If X is uncountable, then ... □

EDIT BELOW HERE

4. A first nice category of Bayesian processes.

Baez and Fritz [2] defined a category they called **FinStat** to which they could associate a relative entropy. Our goal is the same, but we will try to make our context work for us and go about defining our category a bit differently than they did. To justify this approach, we'll first introduce some theorems.

Theorem 4.1 ([3, Section 3.1])**.** *Given a diagram*

$$1 \xrightarrow{p} (X, \Sigma) \xrightarrow{f} (X', \Sigma')$$

in **PrftProb***, there exists a joint measure* $j\colon 1 \to (X \times X', \Sigma \otimes \Sigma')$ *uniquely determined by*

$$j(A \times A') = \int_A f(-, A')dp(*, -).$$

The marginal $p'\colon 1 \to (X', \Sigma')$ *of* j *equals* $p; f$*. Moreover, both* j *and* p' *are perfect.*

Definition 4.2. We define a category **PrftStat** with objects, countably generated perfect probability measures $p\colon 1 \to (X, \Sigma)$ and arrows, not necessarily commuting triangles

$$\begin{tikzcd} & 1 \arrow[dl, "p"'] \arrow[dr, "q'"] & \\ (X,\Sigma) \arrow[rr, "f"] & & (X',\Sigma') \end{tikzcd}$$

in **PrftProb** where q' is absolutely continuous with respect to $p; f$. Note, the motivation to use notation q' instead of q or p' will become apparent soon enough.

Consider a morphism f as described Definition 4.2. By Theorem 4.1, we get the following information for free:

- there exists a perfect probability measure $p'\colon 1 \to (X', \Sigma')$ such that $p' = p; f$,
- there exists a **PerfProb** map $g\colon (X', \Sigma') \to (X, \Sigma)$, unique p'-almost everywhere, such that $p = p'; g$, and
- there exists a perfect probability measure $q\colon 1 \to (X, \Sigma)$ such that $q = q'; g$ and $q \ll p$.

Diagrammatically, a morphism is

$$\begin{tikzcd} & 1 \arrow[dl, "p"] \arrow[dr, "q'"'] \arrow[dl, dashed, bend right, "q"'] \arrow[dr, dashed, bend left, "p'"] & \\ (X,\Sigma) \arrow[rr, "f"] & & (X',\Sigma') \arrow[ll, dashed, bend left, "g"] \end{tikzcd}$$

such that $q' \ll p'$. It follows that $p = p'; g = p; f; g$ and $q \ll p$. The solid arrows are specified information and the dashed arrows follow from our context.

What about morphisms in **PrftStat** such that the diagram

$$\begin{tikzcd} & 1 \arrow[dl, "p"'] \arrow[dr, "q'"] & \\ (X,\Sigma) \arrow[rr, "f"] & & (X',\Sigma') \end{tikzcd}$$

commutes? It is easy to see that the collection of such morphisms gives a subcategory **OptPrftStat**, where the **Opt-** stands for optimal. What is meant by optimal will be explained later. In this case, we get a commuting diagram

$$\begin{tikzcd} & 1 \arrow[dl, "p=q"'] \arrow[dr, "p'=q'"] & \\ (X,\Sigma) \arrow[rr, shift left, "f"] & & (X',\Sigma') \arrow[ll, dashed, shift left, "g"] \end{tikzcd}$$

The title of this section may still be a mystery at this point. In the appendix, we recall the idea of a Bayesian process. Observe that here, (X, Σ) plays the role of the hypothesis, (X', Σ') the evidence, p plays the role of the prior data and q the posterior data. Every **PrftStat**-morphism is like a single iteration in a Bayesian updating process. Then an optimal morphism, that is an arrow in **OptPrftStat** is one in which the best initial hypothesis is chosen, the new evidence does not affect the hypothesis.

5. The relative entropy of **PrftStat**-morphisms.

Given probability measures q, p on a set X, where $q \ll p$, define the *relative entropy* or *Kullback-Leibler divergence* from p to q by

$$S(q,p) = \int_X \log\left(\frac{dq}{dp}\right)\frac{dq}{dp}dp,$$

where $\frac{dp}{dq}$ is the Radon-Nikodym derivative. This value is an attempt to measure the difference between the probability measures. For us, it will measure the amount of information gained by updating from a prior probability to a posterior probability via a Bayesian process.

First, note this doesn't always exist for any two probability measures: we need the Radon-Nikodym derivative. Also, this is not a metric: there is no symmetry, for instance.

However, to every **PrftStat**-morphism

$$\begin{tikzcd} & 1 \arrow[dl, "p"] \arrow[dl, dashed, bend right, "q"'] \arrow[dr, "q'"'] \arrow[dr, dashed, bend left, "p'"] & \\ (X,\Sigma) \arrow[rr, shift left, "f"] & & (X',\Sigma') \arrow[ll, dashed, shift left, "g"] \end{tikzcd}$$

we can associate a relative entropy $S(q, p)$ to q with respect to p. Because all probability spaces are countably generated and $q \ll p$, we satisfy the requisite conditions for the Radon-Nikodym derivative to exist.

Observe that because $p = q$ on **OptPrftStat**-morphisms, we have that $dq/dp = 1$. Hence $S(q, p) = 0$. This is the sense in which these morphisms are optimal. There is no new information to be gained by updating your probability measure from p to q.

References

[1] Samson Abramsky, Richard Blute, and Prakash Panangaden. Nuclear and trace ideals in tensored $*$-categories. *J. Pure Appl. Algebra*, 143(1-3):3–47, 1999. Special volume on the occasion of the 60th birthday of Professor Michael Barr (Montreal, QC, 1997).
[2] Jared Culbertson and Kirk Sturtz. A categorical foundation for Bayesian probability. *Appl. Categ. Structures*, 22(4):647–662, 2014.
[3] Arnold M. Faden. The existence of regular conditional probabilities: necessary and sufficient conditions. *Ann. Probab.*, 13(1):288–298, 1985.
[4] Gerald B. Folland. *Real analysis*. Pure and Applied Mathematics (New York). John Wiley & Sons, Inc., New York, second edition, 1999. Modern techniques and their applications, A Wiley-Interscience Publication.
[5] Brendan Fong. Causal theories: A categorical perspective on bayesian networks. *arXiv preprint arXiv:1301.6201*, 2013.
[6] Michèle Giry. A categorical approach to probability theory. In *Categorical aspects of topology and analysis (Ottawa, Ont., 1980)*, volume 915 of *Lecture Notes in Math.*, pages 68–85. Springer, Berlin-New York, 1982.
[7] F William Lawvere. The category of probabilistic mappings. *preprint*, 1962.
[8] Prakash Panangaden. Probabilistic relations. *SCHOOL OF COMPUTER SCIENCE RESEARCH REPORTS-UNIVERSITY OF BIRMINGHAM CSR*, pages 59–74, 1998.
[9] Robert H Rodine. Perfect probability measures and regular conditional probabilities. *The Annals of Mathematical Statistics*, pages 1273–1278, 1966.