Probability Sheaves and the Giry Monad*

Alex Simpson

Faculty of Mathematics and Physics, University of Ljubljana, Slovenia
Alex.Simpson@fmf.uni-lj.si

Abstract

I introduce the notion of *probability sheaf*, which is a mathematical structure capturing the relationship between probabilistic concepts (such as *random variable*) and sample spaces. Various probability-theoretic notions can be (re)formulated in terms of category-theoretic structure on the category of probability sheaves. As a main example, I consider the Giry monad, which, in its original formulation, constructs spaces of probability measures. I show that the Giry monad generalises to the category of probability sheaves, where it turns out to have a simple, purely category-theoretic definition.

1998 ACM Subject Classification F.3.2 Semantics of Programming Languages

Keywords and phrases Random variable, conditional independence, category theory, sheaves, Giry monad

Digital Object Identifier 10.4230/LIPIcs.CALCO.2017.1

Category Invited Talk

1 Introduction

This article provides technical notes for an invited talk. While these notes present an outline of the main thread of mathematical content of the talk, they do not include the motivation and non-technical discussion that will be given in the talk. Moreover, the talk will assume only basic category theory and probability theory as background, whereas these notes presuppose quite a bit more. The notes also omit the secondary thread of the talk, which concerns parallels between the technical development presented here and one presentation of the *Schanuel topos*, see, e.g., [2], which, in another guise, is well known in computer science as the category of *nominal sets* [3].

2 Sheaves of random variables

Traditionally, a *random variable* is a measurable function from a probability space Ω (the *sample space*) to the measurable space in which the random variable takes its values. In most uses of random variables, however, the sample space plays only an auxiliary role. It serves mainly as a convenient device for manipulating joint probability distributions over all random variables under consideration. The precise identity of the space Ω itself is irrelevant. Indeed, on the contrary, probabilistic notions, such as random variable, enjoy an invariance property under *change of sample space*, which has been argued by Tao to be a characterising

* This research was carried out at the University of Edinburgh, supported by the John Templeton Foundation (grant no. 39465 "Randomness via Information Independence"), and at the University of Ljubljana, supported by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF (Award No. FA9550-14-1-0096) and by the Slovenian Research Agency (research core funding No. P1–0294).

feature of legitimate probabilistic concepts [5]. Tao's principle of invariance under change of sample space is an important guiding principle behind this talk. However, our mathematical formulation will differ from his.

The appropriate notion of change of sample space will be formulated in terms of a category of sample spaces. We avoid mathematically pathological sample spaces, by restricting to *Polish probability spaces* (also known as *standard Borel spaces*).[1]

Recall that a *Polish space* is a topological space that is metrisable as a complete separable metric space. A *Polish probability space* is given by a Polish space Ω together with a probability measure P_Ω on its σ-algebra $\mathcal{B}(\Omega)$ of Borel sets. Henceforth when we say *sample space* we mean Polish probability space. We define the category $\mathbb{P}$ of sample spaces to have:

- *objects:* Polish probability spaces;
- *morphisms from Ω' to Ω:* Borel-measurable functions $q\colon \Omega' \to \Omega$ that are *measure preserving* (i.e., for every Borel $B \subseteq \Omega$, it holds that $P_\Omega(B) = P_{\Omega'}(q^{-1}(B))$).

(In [5], Tao admits arbitrary probability spaces, but restricts maps to *surjective* measure-preserving measurable functions.)

Let A be any Polish space, and Ω any sample space. We define the set $\underline{\mathrm{RV}}(A)(\Omega)$ of A-valued random variables with sample space Ω by:

$$\underline{\mathrm{RV}}(A)(\Omega) \;=\; \{X : \Omega \to A \mid X \text{ is Borel measurable}\}/{=_{\mathrm{a.e.}}} \;,$$

where $=_{\mathrm{a.e.}}$ is the equivalence relation of almost everywhere equality.

Given an equivalence class $[X] \in \underline{\mathrm{RV}}(A)(\Omega)$ and a map $q\colon \Omega' \to \Omega$, we write $[X].q$ for the equivalence class $[X \circ q] \in \underline{\mathrm{RV}}(A)(\Omega')$ obtained by composition. This is a well-defined operation on equivalence classes. Henceforth, we shall use such explicit equivalence-class notation only where necessary to avoid confusion. In most cases, we shall write X both for the function $X : \Omega \to A$, and for its equivalence class.

▶ **Proposition 1.** *The above data defines a functor $\underline{\mathrm{RV}}(A)\colon \mathbb{P}^{\mathsf{op}} \to$ **Set**; i.e., $\underline{\mathrm{RV}}(A)$ is a presheaf on $\mathbb{P}$.*

We remark that there is an obvious alternative presheaf of random variables, in which the quotienting modulo almost everywhere equality is not performed. Neverthless, the quotiented presheaf seems the more significant of the two. For example, important probabilistic constructions, such as conditional expectation, only define random variables up to almost everywhere equality. Furthermore, the next property, which is fundamental to the ensuing technical development, forces the quotiented presheaf upon us.

▶ **Proposition 2.** *The presheaf $\underline{\mathrm{RV}}(A)$ is **separated** in the sense that, for any $X, X' \in \underline{\mathrm{RV}}(A)(\Omega)$ and map $q\colon \Omega' \to \Omega$, if $X.q = X'.q$ then $X = X'$.*

This follows from the fact that the image of q in Ω has measure 1 in the completion of P_Ω (it is measurable because it is an analytic set).

The theorem below gives an important strengthening of the proposition above. Suppose we have $q\colon \Omega' \to \Omega$ and $Y \in \underline{\mathrm{RV}}(A)(\Omega')$. We say that Y is *q-invariant* if, for every parallel pair $p, p'\colon \Omega'' \to \Omega'$ for which $q \circ p = q \circ p'$, it holds that $Y.p = Y.p'$.

▶ **Theorem 3.** *The presheaf $\underline{\mathrm{RV}}(A)$ is a **sheaf** in the sense that, for every $q\colon \Omega' \to \Omega$ and q-invariant $Y \in \underline{\mathrm{RV}}(A)(\Omega')$, there exists a unique $X \in \underline{\mathrm{RV}}(A)(\Omega)$ such that $Y = X.q$.*

---

[1] Our development can be adapted to use other classes of space, such as *analytic spaces*, or *standard probability spaces*.

We briefly state the intuitive reading of the material in this section. As in [5], one can view a map $q\colon \Omega' \to \Omega$ as presenting Ω' as an *extension* of the sample space Ω, in which there is additional room for probabilistic variation. Indeed, any $\omega \in \Omega$ in the image of q is expanded by q to the non-empty fibre $q^{-1}(\omega)$ of points in Ω' above it. And since the image of q has measure 1 in Ω, such expansions occur almost everywhere.

Given the above interpretation, Proposition 1 formalises the preservation properties enjoyed by random variables under extension of sample space. This is consistent with the thesis of Tao [5], which may be reformulated as stating that legitimate probabilistic concepts form presheaves. The sheaf property of Theorem 3 also states a property that is desirable, in general, of probabilistic concepts. Suppose we have $q\colon \Omega' \to \Omega$, understood as presenting Ω' as an extension of the sample space Ω. Then the property of $Y \in \underline{\mathrm{RV}}(A)(\Omega')$ being q-invariant asserts that Y does not exploit any of the additional scope for variation available in Ω' beyond that which is already present in Ω. The sheaf property of Theorem 3 then asserts that, in this situation, a (uniquely determined) version of Y is available directly on the sample space Ω itself. Such a restriction from a larger sample space Ω' to a smaller one Ω, in cases in which the additional variation in Ω' is ignored, is a natural property to require of probabilistic concepts in general. Motivated by this, we postulate that probabilistic notions should form *sheaves*, in the general sense introduced in the next section.

3 The category of probability sheaves

Let $F\colon \mathbb{P}^{\mathsf{op}} \to \mathbf{Set}$ be any presheaf. Given any $q\colon \Omega' \to \Omega$ and $y \in F(\Omega')$, we say that y is *q-invariant* if, for every parallel pair $p, p'\colon \Omega'' \to \Omega'$ for which $q \circ p = q \circ p'$, it holds that $y.p = y.p'$. We say that F is a *sheaf* if, for every $q\colon \Omega' \to \Omega$ and q-invariant $y \in F(\Omega')$, there exists a unique $x \in F(\Omega)$ such that $y = x.q$. Because our sheaves are defined over a category of sample spaces, we call them *probability sheaves*.

The category $\mathit{Sh}(\mathbb{P})$ is defined as the full subcategory of $\mathbb{P}$-presheaves on sheaves; i.e., objects are sheaves and morphisms are natural transformations.

In order to understand the structure of $\mathit{Sh}(\mathbb{P})$, we need to look more deeply at the structure of $\mathbb{P}$ itself. Consider a commuting square

$$\begin{array}{ccc} \Omega' & \xrightarrow{\;r_2\;} & \Omega_2 \\ {\scriptstyle r_1}\big\downarrow & & \big\downarrow{\scriptstyle q_2} \\ \Omega_1 & \xrightarrow[\;q_1\;]{} & \Omega \end{array} \tag{1}$$

and write r for the resulting map $r = q_1 \circ r_1 = q_2 \circ r_2$. We say that the commuting square is *independent* if r_1 and r_2 are conditionally independent relative to r, where all maps are considered as random variables over sample space Ω'.

▶ **Theorem 4.** *Every cospan* $\Omega_1 \xrightarrow{q_1} \Omega \xleftarrow{q_2} \Omega_2$ *in* $\mathbb{P}$ *completes to a commuting square*

$$\begin{array}{ccc} \Omega_1 \otimes_\Omega \Omega_2 & \xrightarrow{\;p_2\;} & \Omega_2 \\ {\scriptstyle p_1}\big\downarrow & & \big\downarrow{\scriptstyle q_2} \\ \Omega_1 & \xrightarrow[\;q_1\;]{} & \Omega \end{array} \tag{2}$$

which enjoys the following characterisation as a universal independent square.

1. *The commuting square (2) is independent; and*
2. *for every independent square (1) completing* $\Omega_1 \xrightarrow{q_1} \Omega \xleftarrow{q_2} \Omega_2$*, there exists a unique map* $\Omega' \xrightarrow{s} \Omega_1 \otimes_\Omega \Omega_2$ *such that* $p_1 \circ s = r_1$ *and* $p_2 \circ s = r_2$*.*

In the proof of the theorem, the Polish space $\Omega_1 \otimes_\Omega \Omega_2$ has the expected set-theoretic pullback

$$\Omega_1 \otimes_\Omega \Omega_2 \;=\; \{(\omega_1, \omega_2) \mid q_1(\omega_1) = q_2(\omega_2)\}$$

as its underlying set, and the probability measure is constructed by integrating fibrewise product measures of the regular conditional probabilities on the fibres of p_1 and p_2 over the conditioning space Ω. For more details, see [4], where this structure is axiomatised as a *local independent product* on $\mathbb{P}$.

Theorem 4 implies *a fortiori* that every cospan in $\mathbb{P}$ completes to a commuting square. This property is known as the *right Ore condition*. It is equivalent to saying that the *dense coverage* on $\mathbb{P}$ is generated by singleton covers, i.e., that the $\mathbb{P}$ carries an *atomic coverage*.[2] Given this, it is immediate from our definition of sheaf, that our sheaves are simply the atomic sheaves over $\mathbb{P}$, and thus $\mathit{Sh}(\mathbb{P})$ is an atomic Grothendieck topos.

The full statement of Theorem 4 strengthens the right Ore condition with a universal property related to conditional independence. Essential use will be made of this strengthening in Section 5 below. The notion of independent square that is associated with this strengthening also permits the following alternative characterisation of the sheaf property.

▶ **Theorem 5.** *The following are equivalent for a presheaf* $F\colon \mathbb{P}^{\mathsf{op}} \to \mathbf{Set}$*.*

1. *F is a sheaf.*
2. *F maps every independent square in $\mathbb{P}$ to a pullback square in* $\mathbf{Set}$*. (Note that, due to contravariance, for an independent square (1), it is $F(\Omega)$ that is the apex of the resulting pullback square.)*

4 The $\underline{\mathrm{RV}}$ functor

We saw in Section 2 that, for every Polish space A, it holds that $\underline{\mathrm{RV}}(A)$ is a sheaf. In this short section we establish that $\underline{\mathrm{RV}}$ defines a faithful functor from a category of Polish spaces to $\mathit{Sh}(\mathbb{P})$.

A function $f\colon A' \to A$, between two Polish spaces, is said to be *universally measurable* if, for every Borel probability measure $P_{A'}$ on A', and every Borel subset $B \subset A$, it holds that $f^{-1}B$ is measurable in the $P_{A'}$ -completion of $\mathcal{B}(A')$. Every Borel-measurable function is trivially universally measurable, but the converse does not hold. Universally measurable functions are closed under composition. (This is not immediate from the definition.) We write $\mathbf{Pol}_{\mathrm{um}}$ for the category with Polish spaces as objects and universally measurable functions as morphisms.

▶ **Proposition 6.** *The mapping $A \mapsto \underline{\mathrm{RV}}(A)$ extends to a faithful functor $\underline{\mathrm{RV}}\colon \mathbf{Pol}_{um} \to \mathit{Sh}(\mathbb{P})$ whose action on morphisms defined as follows. For every universally measurable $f\colon A' \to A$, sample space Ω and $X \in \underline{\mathrm{RV}}(A)(\Omega)$, define $\underline{\mathrm{RV}}(f)(\Omega)(X) = f \circ X$.*

[2] We adopt the terminology of [2] where, in particular, *coverage* is used as a synonym for *Grothendieck topology*.

The functor $\underline{\mathrm{RV}}$ is far from full. For example, the following morphism from $\underline{\mathrm{RV}}(\mathbf{2})$ to $\underline{\mathrm{RV}}(\mathbf{2})$, where $\mathbf{2} = \{0, 1\}$, is not in the image of $\underline{\mathrm{RV}}$.

$$X \colon \Omega \to \{0,1\} \;\;\mapsto\;\; \omega \in \Omega \;\mapsto\; \begin{cases} 0 & \text{if } \mathbf{P}(X = 0) = 1 \text{ or } \mathbf{P}(X = 1) = 1 \\ 1 & \text{otherwise} \ . \end{cases}$$

In general, morphisms from $\underline{\mathrm{RV}}(A)$ to $\underline{\mathrm{RV}}(A')$ can exploit statistical properties of the random variable given as an argument, whereas morphisms in the image of $\underline{\mathrm{RV}}$ have no such capacity.

5 The Giry monad

Giry's classic paper [1] defines two monads of spaces of probability measures. The first is a monad on the category of all measurable spaces. The second is a monad on the category of continuous maps between Polish spaces. The latter construction is defined as follows. It is standard that the set of all Borel probability measures on a Polish space A itself forms a Polish space under the weak topology on probability measures. We write $\mathcal{M}(A)$ for this Polish space of probability measures. In [1], the mapping $A \mapsto \mathcal{M}(A)$ is shown to extend to a monad on the category of continuous maps between Polish spaces. In fact, it can be shown to extend further to a monad on the category $\mathbf{Pol}_{\mathrm{um}}$ of universally measurable maps between Polish spaces. We end this note by showing that it extends beyond this to a monad $\underline{\mathcal{M}}$ on the whole of $\mathit{Sh}(\mathbb{P})$.

Perhaps surprisingly, the monad $\underline{\mathcal{M}}$ can be given a purely category-theoretic definition. For a presheaf $F \colon \mathbb{P}^{\mathsf{op}} \to \mathbf{Set}$ and sample space Ω, define

$$\underline{\mathcal{M}}F\,(\Omega) \;=\; \int^{\Omega'} \mathbb{P}(\Omega', \Omega) \times F(\Omega') \ . \tag{3}$$

Here we use a coend formula for convenience of notation. Nevertheless, since the inside expression is purely contravariant in Ω', the definition simply finds the colimit in $\mathbf{Set}$ of the contravariant functor $\Omega' \mapsto \mathbb{P}(\Omega', \Omega) \times F(\Omega')$.

Because the parameter Ω in (3) is covariant, $\underline{\mathcal{M}}F$ defines a covariant functor from $\mathbb{P}$ to $\mathbf{Set}$. More interestingly, for our purposes, it turns out that $\underline{\mathcal{M}}F$ also defines a contravariant functor; i.e., $\underline{\mathcal{M}}F$ carries the structure of a presheaf. The construction of this presheaf structure exploits *local independent products* in $\mathbb{P}$, as formulated in Theorem 4.

Specifically, suppose we have an equivalence class $[(r, x)] \in \underline{\mathcal{M}}F\,(\Omega)$, where $r \colon \Omega'' \to \Omega$ and $x \in F(\Omega'')$, and a map $q \colon \Omega' \to \Omega$. We need to define $[(r, x)].q \in \underline{\mathcal{M}}F\,(\Omega')$. Consider the local independent product diagram below.

$$\begin{array}{ccc} \Omega' \otimes_\Omega \Omega'' & \xrightarrow{\;p''\;} & \Omega'' \\ {\scriptstyle p'}\downarrow & & \downarrow{\scriptstyle r} \\ \Omega' & \xrightarrow[\;q\;]{} & \Omega \end{array}$$

Then define

$$[(r,\, x)].q \;=\; [(p',\, x.p'')] \ ,$$

which is indeed well-defined on equivalence classes. Henceforth, we take this contravariant action of $\underline{\mathcal{M}}F$ as basic. Of course it needs to be shown that this action is functorial; i.e., that it defines a presheaf. This property is subsumed under the first item of the theorem below.

▶ **Theorem 7.**

1. *For every presheaf F, it holds that $\underline{\mathcal{M}}F$ is a sheaf.*
2. *The operation $F \mapsto \underline{\mathcal{M}}F$ defines a functor $\underline{\mathcal{M}}$ from $Psh(\mathbb{P})$ to $Sh(\mathbb{P})$.*
3. *The induced endofunctor $\underline{\mathcal{M}}$ on $Sh(\mathbb{P})$ carries the structure of a strong monad.*

Our main result states that $\underline{\mathcal{M}}$ is indeed an extension of $\mathcal{M}$ to the whole of $Sh(\mathbb{P})$.

▶ **Theorem 8.** *There is a natural isomorphism $\underline{\mathrm{RV}}\,\mathcal{M} \cong \underline{\mathcal{M}}\,\underline{\mathrm{RV}}$ which exhibits the functor $\underline{\mathrm{RV}}\colon \mathbf{Pol}_{um} \to Sh(\mathbb{P})$ as strong-monad preserving.*

References

1 M. Giry. A categorical approach to probability theory. In *Categorical Aspects of Topology and Analysis*, pages 68–85. Springer-Verlag, 1982.

2 P. T. Johnstone. *Sketches of an Elephant: A Topos Theory Compendium*. Oxford University Press, 2002.

3 A. Pitts. *Nominal Sets: Names and Symmetry in Computer Science*. Cambridge University Press, 2013.

4 A. Simpson. Category-theoretic structure for independence and conditional independence. In *Proceedings of MFPS*, 2017.

5 T. Tao. A review of probability theory. Note0 in 254A – random matrices, 2010. URL: `https://terrytao.wordpress.com/2010/01/01/254a-notes-0-a-review-of-probability-theory/`.