Probability monads as codensity monads

Ruben Van Belle

August 2021

What is a probability distribution?

- A probability distribution on a **finite set** A:

$$p : A \to [0,1] \quad \text{such that } \sum_{a \in A} p(a) = 1.$$

Example: a fair coin.

What is a probability distribution?

- A probability distribution on a **finite set** A:

$$p : A \to [0,1] \quad \text{such that } \sum_{a \in A} p(a) = 1.$$

  Example: a fair coin.

- A probability distribution on a **countable set** A:

$$p : A \to [0,1] \quad \text{such that } \sum_{a \in A} p(a) = 1.$$

  Example: the Poisson distribution.

What is a probability distribution?

- A probability distribution on a **finite set** A:

$$p : A \to [0,1] \quad \text{such that} \sum_{a\in A} p(a) = 1.$$

  Example: a fair coin.

- A probability distribution on a **countable set** A:

$$p : A \to [0,1] \quad \text{such that} \sum_{a\in A} p(a) = 1.$$

  Example: the Poisson distribution.

- Probability distributions on other sets such as $\mathbb{R}$ or $\mathcal{C}([0,\infty),\mathbb{R})$.
  $\to$ We need to use measure theory.

Idea: Using category theory we will construct probability measures as an extension of probability measures on finite or countable spaces.

Codensity monads

A monad on a category C is given by:

- an endofunctor $T : C \to C$,
- a natural transformation $\eta : 1_C \to T$ (the **unit**),
- a natural transformation $\mu : TT \to T$ (the **multiplication**)

such that the following diagrams commute:

$$
\begin{tikzcd}
T \arrow[r, "T\eta"] \arrow[dr, "1_T"'] & TT \arrow[d, "\mu"] & T \arrow[l, "\eta_T"'] \arrow[dl, "1_T"] \\
& T &
\end{tikzcd}
\qquad
\begin{tikzcd}
TTT \arrow[r, "T\mu"] \arrow[d, "\mu_T"'] & TT \arrow[d, "\mu"] \\
TT \arrow[r, "\mu"'] & T
\end{tikzcd}
$$

Suppose we have a functor $R : D \to C$ that has a left adjoint. The adjunction $C \underset{R}{\overset{L}{\rightleftarrows}} D$ induces a monad on C:

- Endofunctor: RL,
- Unit: $\eta : 1_C \to RL$,
- Multiplication: $R\epsilon_L : RLRL \to RL$.

Suppose we have a functor $G : D \to C$ such that the right Kan extension along itself exists. Then G induces a monad (the **codensity monad of** G):

- Endofunctor: $T^G := \mathrm{Ran}_G G$

$$\begin{tikzcd} D \arrow[rr, "G"] \arrow[dr, "G"'] & \arrow[d, phantom, "\Uparrow\gamma"] & C \\ & C \arrow[ur, dashed, "T^G"'] & \end{tikzcd}$$

- Unit:

$$\begin{tikzcd} D \arrow[rr, "G"] \arrow[dr, "G"'] & \arrow[d, phantom, "1_G\Uparrow"] & C \\ & C \arrow[ur, "1_C"'] & \end{tikzcd} = \begin{tikzcd} D \arrow[rr, "G"] \arrow[dr, "G"'] & \arrow[d, phantom, "\Uparrow\gamma"] & C \\ & C \arrow[ur, dashed, "T^G"] \arrow[ur, bend right, "1_C"'] \arrow[ur, phantom, bend right=20, "\eta"] & \end{tikzcd}$$

- Multiplication:

$$\begin{tikzcd} D \arrow[rr, "G"] \arrow[dr, "G"'] & \arrow[d, phantom, "T^G\gamma\Uparrow"] & C \\ & C \arrow[ur, "T^G T^G"'] & \end{tikzcd} = \begin{tikzcd} D \arrow[rr, "G"] \arrow[dr, "G"'] & \arrow[d, phantom, "\Uparrow\gamma"] & C \\ & C \arrow[ur, dashed, "T^G"] \arrow[ur, bend right, "T^G T^G"'] \arrow[ur, phantom, bend right=20, "\mu"] & \end{tikzcd}$$

Examples:

- Let $C \underset{R}{\overset{L}{\rightleftarrows}} D$ be adjoint functors. The codensity monad of R is the monad induced by the adjunction.
- The codensity monad of the inclusion functor $\mathbf{Set}_f \to \mathbf{Set}$ is the ultrafilter monad.

Giry monad

Let X be a measurable space.

- Define

$$\mathcal{G}X := \{\mathbb{P} \mid \mathbb{P} \text{ is a probability measure on } X\}$$

$$\Sigma_{\mathcal{G}X} := \sigma(\mathrm{ev}_A \mid A \text{ is a measurable subset of } X).$$

  Here $\mathrm{ev}_A : \mathcal{G}X \to [0,1]$ is defined by $\mathbb{P} \mapsto \mathbb{P}(A)$.
  Let $f : X \to Y$ be a measurable map. There is a measurable map $\mathcal{G}f : \mathcal{G}X \to \mathcal{G}Y$ defined by

$$\mathbb{P} \mapsto \mathbb{P} \circ f^{-1}.$$

  $\to$ We obtain a functor $\mathcal{G} : \mathbf{Mble} \to \mathbf{Mble}$.

- There is a measurable map $\eta_X : X \to \mathcal{G}X$ defined by

$$x \mapsto \delta_x.$$

$\to$ This induces a natural transformation $\eta : 1_{\mathbf{Mble}} \to \mathcal{G}$.

- There is a measurable map $\mu_X : \mathcal{G}\mathcal{G}X \to \mathcal{G}X$ defined by

$$\mu_X(\mathbf{P})(A) := \int_{\mathcal{G}X} \lambda(A)\mathbf{P}(\mathrm{d}\lambda)$$

for all $\mathbf{P} \in \mathcal{G}\mathcal{G}X$ and measurable subsets A of X.
$\to$ This induces a natural transformation $\mu : \mathcal{G}\mathcal{G} \to \mathcal{G}$.

Proposition (Giry)

The triple $(\mathcal{G}, \eta, \mu)$ is a monad (the **Giry monad).**

The Giry monad as codensity monad

Let A be a countable set. The measurable subspace

$$\{(p_a)_a \in [0,1]^A \mid \sum_{a \in A} p_a = 1\} \subseteq [0,1]^A$$

is denoted by GA.
Every map of countable sets $f : A \to B$ induces a measurable map $Gf : GA \to GB$ defined by

$$(p_a)_a \mapsto \left(\sum_{a \in f^{-1}(b)} p_a \right)_b .$$

This gives us a functor $G : \mathbf{Set}_c \to \mathbf{Mble}$.

Theorem

The codensity monad of G is the Giry monad.

Proof.

We will show that

$$\mathcal{G}X \cong (\mathrm{Ran}_G G)(X) \cong \lim(X \downarrow G \xrightarrow{U} \mathbf{Set}_c \xrightarrow{G} \mathbf{Mble}).$$

For a measurable map $f : X \to GA$ define $p_f : \mathcal{G}X \to GA$ by

$$\mathbb{P} \mapsto \left(\int_X f_a \mathrm{d}\mathbb{P} \right)_a .$$

Then $(\mathcal{G}X, (p_f)_f)$ forms a cone over the diagram. $\square$

Theorem

The codensity monad of G is the Giry monad.

Proof.

Suppose we have another cone $(Y, (q_f)_f)$ over the diagram.

- For $f \in \mathbf{Mble}(X, [0,1])$ define $\hat{f} : X \to G\mathbf{2}$ by $\hat{f}(x) := (1 - f(x), f(x))$.
- For $y \in Y$ define $I_y : \mathbf{Mble}(X, [0,1]) \to [0,1]$ by

$$f \mapsto q_{\hat{f}}(y)_1$$

$\square$

Theorem

The codensity monad of G is the Giry monad.

Proof.

Let a countable collection $(f_a)_{a\in A}$ such that $f := \sum_{a\in A} f_a \leq 1$.
Claim: $I_y(f) = \sum_{a\in A} I_y(f_a)$ Define the following:

- $B := A \cup \{*\}$
- Let $s : B \to \mathbf{2}$ be the map such that $s(b) = 0$ if and only if $b = *$.
- For $a \in A$ let $s_a : B \to \mathbf{2}$ be the map such that $s(b) = 1$ if and only if $b = a$.
- Let $h : X \to GB$ be the map such that $h(x)_a = f_a(x)$ for all $a \in A$ and $h(x)_* = 1 - f(x)$.

$\square$

Theorem

The codensity monad of G is the Giry monad.

Proof.

We have commutative diagrams:

$$\begin{tikzcd} X \arrow[r, "h"] \arrow[dr, "\hat{f}"'] & GB \arrow[d, "Gs"] \\ & G\mathbf{2} \end{tikzcd} \Rightarrow \begin{tikzcd} Y \arrow[r, "q_h"] \arrow[dr, "q_{\hat{f}}"'] & GB \arrow[d, "Gs"] \\ & G\mathbf{2} \end{tikzcd}$$

For every a in A we have the commutative diagrams:

$$\begin{tikzcd} X \arrow[r, "h"] \arrow[dr, "\hat{f}_a"'] & GB \arrow[d, "Gs_a"] \\ & G\mathbf{2} \end{tikzcd} \Rightarrow \begin{tikzcd} Y \arrow[r, "q_h"] \arrow[dr, "q_{\hat{f}_a}"'] & GB \arrow[d, "Gs_a"] \\ & G\mathbf{2} \end{tikzcd}$$

□

Theorem

The codensity monad of G is the Giry monad.

Proof.

We find that

$$I_y(f) = q_{\hat{f}}(y)_1 = (Gs \circ q_h(y))_1 = \sum_{a \in A} q_h(y)_a$$

and

$$I_y(f_a) = q_{\hat{f}_a}(y)_1 = (Gs_a \circ q_h(y))_1 = q_h(y)_a.$$

Therefore it follows that $I_y(f) = \sum_{a \in A} I_y(f_a)$. $\square$

Theorem

The codensity monad of G is the Giry monad.

Proof.

The map I_y has the following properties:

- For a countable collection $(f_a)_a$ such that $f := \sum_{a\in A} f_a \leq 1$ we have that

$$I_y(f) = \sum_{a\in A} I_y(f_a).$$

- We have that $I_y(1) = 1$.

This implies that there exists a unique probability measure $\mathbb{P}_y$ on X such that $I_y(f) = \int_X f \mathrm{d}\mathbb{P}_y$ for all $f \in \mathbf{Mble}(X, [0, 1])$.
The assignment $y \mapsto \mathbb{P}_y$ defines a measurable map $q : Y \to \mathcal{G}X$. This is morphism of cones and every other morphism of cones has to be equal q. $\square$

Remarks

- Let $i : \mathbf{Set}_f \to \mathbf{Set}_c$ be the inclusion functor. The codensity monad of Gi is a probability monad of finitely additive probability measures.
- Integration operators vs. probability measures:
  $\to$ In the proof we can see that $\mathrm{Ran}_G G(X)$ is the space of all integration operators on X, which we then identified with probability measures.
  $\to$ Let $j : \mathbf{Set}_c \to \mathbf{Mble}$ be the functor that sends a countable set A to the measurable space $(A, \mathcal{P}(A))$. Then $\mathrm{Ran}_j G(X)$ would be a more direct construction of the space of probability measures on X.
- The Giry monad is the codensity monad of the inclusion of the category of convex sets in **Mble** (Avery).

Radon monad

Let X be a compact Hausdorff space. A Borel probability measure $\mathbb{P}$ on X is called a Radon probability measure if for every Borel measurable subset A we have that $\mathbb{P}(A) = \sup\{\mathbb{P}(K) \mid K \subseteq A \text{ and } K \text{ is compact}\}$.

Radon monad

Let X be a compact Hausdorff space. A Borel probability measure $\mathbb{P}$ on X is called a Radon probability measure if for every Borel measurable subset A we have that $\mathbb{P}(A) = \sup\{\mathbb{P}(K) \mid K \subseteq A \text{ and } K \text{ is compact}\}$.

- Define

$$\mathcal{R}X := \{\mathbb{P} \mid \mathbb{P} \text{ is a Radon probability measure on } X\}$$

$$\mathcal{T}_{\mathcal{R}X} := \tau(\mathrm{ev}_f \mid f \in \mathbf{CH}(X, [0,1]))$$

  Here $\mathrm{ev}_f : \mathcal{R}X \to [0,1]$ is defined by $\mathbb{P} \mapsto \int_X f \mathrm{d}\mathbb{P}$.
  This defines a functor $\mathcal{R} : \mathbf{CH} \to \mathbf{CH}$.

Radon monad

Let X be a compact Hausdorff space. A Borel probability measure $\mathbb{P}$ on X is called a Radon probability measure if for every Borel measurable subset A we have that $\mathbb{P}(A) = \sup\{\mathbb{P}(K) \mid K \subseteq A \text{ and } K \text{ is compact}\}$.

- Define

$$\mathcal{R}X := \{\mathbb{P} \mid \mathbb{P} \text{ is a Radon probability measure on } X\}$$

$$\mathcal{T}_{\mathcal{R}X} := \tau(\mathrm{ev}_f \mid f \in \mathbf{CH}(X, [0,1]))$$

  Here $\mathrm{ev}_f : \mathcal{R}X \to [0,1]$ is defined by $\mathbb{P} \mapsto \int_X f \mathrm{d}\mathbb{P}$.
  This defines a functor $\mathcal{R} : \mathbf{CH} \to \mathbf{CH}$.
  Remarks:
  $\to$ It can be shown that $\mathcal{R}X$ is a compact Hausdorff space.
  $\to$ The pushforward of a Radon probability measure along a continuous function is again a Radon probabiliy measure.
  $\to$ Note that $\mathcal{T}_{\mathcal{R}X}$ is the topology of weak convergence of probability measures.

- The assignment $x \mapsto \delta_x$ defines a natural transformation $\eta : 1_{\mathbf{CH}} \to \mathcal{R}$.
- For $\mathbf{P} \in \mathcal{R}\mathcal{R}X$ we have that

$$\mu_X(\mathbf{P}) := \int_{\mathcal{R}X} \lambda(\cdot)\mathbf{P}(\mathrm{d}\lambda)$$

  is a Radon probability measure on X. The assignment $\mathbf{P} \mapsto \mu_X(\mathbf{P})$ defines a natural transformation $\mu : \mathcal{R}\mathcal{R} \to \mathcal{R}$.

- The assignment $x \mapsto \delta_x$ defines a natural transformation $\eta : 1_{\mathbf{CH}} \to \mathcal{R}$.
- For $\mathbf{P} \in \mathcal{R}\mathcal{R}X$ we have that

$$\mu_X(\mathbf{P}) := \int_{\mathcal{R}X} \lambda(\cdot)\mathbf{P}(\mathrm{d}\lambda)$$

is a Radon probability measure on X. The assignment $\mathbf{P} \mapsto \mu_X(\mathbf{P})$ defines a natural transformation $\mu : \mathcal{R}\mathcal{R} \to \mathcal{R}$.

Proposition (Swirszcz)

The triple $(\mathcal{R}, \eta, \mu)$ is a monad (the **Radon monad).**

Radon monad as codensity monad

Let A be a finite set. The (compact) subspace

$$\{(p_a)_a \in [0,1]^A \mid \sum_{a \in A} p_a = 1\} \subseteq [0,1]^A$$

is denoted by GA.
This gives us a functor $G : \mathbf{Set}_f \to \mathbf{CH}$.

Radon monad as codensity monad

Let A be a finite set. The (compact) subspace

$$\{(p_a)_a \in [0,1]^A \mid \sum_{a\in A} p_a = 1\} \subseteq [0,1]^A$$

is denoted by GA.
This gives us a functor $G : \mathbf{Set}_f \to \mathbf{CH}$.

Theorem

The codensity monad of G is the Radon monad.

Proof.

Let X be a compact Hausdorff space.

$$\begin{aligned} T^G(X) &\cong \{I : \mathbf{CH}(X,[0,1]) \to [0,1] \mid I(f+g) = I(f) + I(g) \text{ and } I(1) = 1\} \\ &\cong \{I : \mathbf{CH}(X,\mathbb{R}) \to \mathbb{R} \mid I \text{ is positive and linear and } I(1) = 1\} \\ &\cong \mathcal{R}X \end{aligned}$$

In the last step we used the Riesz-Markov representation theorem. □

bounded Lipschitz monad

Let X be a compact metric space.

- Define

$$TX := \{\mathbb{P} \mid \mathbb{P} \text{ is a Radon probability measure on } X\}$$

$$d_{bl}(\mathbb{P}, \mathbb{Q}) := \sup\left\{\left|\int_X f\mathrm{d}\mathbb{P} - \int_X f\mathrm{d}\mathbb{Q}\right| \mid f : X \to [0,1] \text{ is a 1-Lipschitz function}\right\}$$

  This is a metric on TX (the **bounded Lipschitz metric**).The metric space TX is compact.
  We have a functor T : **CompMet** $\to$ **CompMet**. (Where **CompMet** is the category of compact metric spaces and 1-Lipschitz functions.)

bounded Lipschitz monad

Let X be a compact metric space.

- Define

$$TX := \{\mathbb{P} \mid \mathbb{P} \text{ is a Radon probability measure on } X\}$$

$$d_{bl}(\mathbb{P}, \mathbb{Q}) := \sup \left\{ \left| \int_X f \mathrm{d}\mathbb{P} - \int_X f \mathrm{d}\mathbb{Q} \right| \mid f : X \to [0,1] \text{ is a 1-Lipschitz function} \right\}$$

  This is a metric on TX (the **bounded Lipschitz metric**).The metric space TX is compact.
  We have a functor T : **CompMet** $\to$ **CompMet**. (Where **CompMet** is the category of compact metric spaces and 1-Lipschitz functions.)
- We can define a unit η and a multiplication μ as before.

bounded Lipschitz monad

Let X be a compact metric space.

- Define

$$TX := \{\mathbb{P} \mid \mathbb{P} \text{ is a Radon probability measure on } X\}$$

$$d_{bl}(\mathbb{P}, \mathbb{Q}) := \sup \left\{ \left| \int_X f \mathrm{d}\mathbb{P} - \int_X f \mathrm{d}\mathbb{Q} \right| \mid f : X \to [0,1] \text{ is a 1-Lipschitz function} \right\}$$

  This is a metric on TX (the **bounded Lipschitz metric**).The metric space TX is compact.
  We have a functor T : **CompMet** $\to$ **CompMet**. (Where **CompMet** is the category of compact metric spaces and 1-Lipschitz functions.)
- We can define a unit η and a multiplication μ as before.

Proposition

The triple (T, η, μ) is a monad (the **bounded Lipschitz monad**).

bounded Lipschitz monad as codensity monad

Let A be a finite set and define $GA := \{(p_a)_a \in [0,1]^A \mid \sum_{a \in A} p_a = 1\}$. For $p, q \in GA$ define

$$d_{GA}(p,q) := \sup \left\{ \left| \sum_{a \in A'} p_a - \sum_{a \in A'} q_a \right| \mid A' \subseteq A \right\}.$$

The map d_{GA} is a metric on GA and the obtained metric space is compact. We have functor $G : \mathbf{Set}_f \to \mathbf{CompMet}$.

bounded Lipschitz monad as codensity monad

Let A be a finite set and define $GA := \{(p_a)_a \in [0,1]^A \mid \sum_{a \in A} p_a = 1\}$. For $p, q \in GA$ define

$$d_{GA}(p,q) := \sup \left\{ \left| \sum_{a \in A'} p_a - \sum_{a \in A'} q_a \right| \mid A' \subseteq A \right\}.$$

The map d_{GA} is a metric on GA and the obtained metric space is compact. We have functor $G : \mathbf{Set}_f \to \mathbf{CompMet}$.

Theorem

The codensity monad of G is the bounded Lipschitz monad.

Proof.

The proof is the same as for compact Hausdorff space but we need to verify that every map in the proof is 1-Lipschitz. □

bounded Lipschitz monad as codensity monad

Let A be a finite set and define $GA := \{(p_a)_a \in [0,1]^A \mid \sum_{a\in A} p_a = 1\}$. For $p, q \in GA$ define

$$d_{GA}(p,q) := \sup\left\{\left|\sum_{a\in A'} p_a - \sum_{a\in A'} q_a\right| \mid A' \subseteq A\right\}.$$

The map d_{GA} is a metric on GA and the obtained metric space is compact. We have functor $G : \mathbf{Set}_f \to \mathbf{CompMet}$.

Theorem

The codensity monad of G is the bounded Lipschitz monad.

Proof.

The proof is the same as for compact Hausdorff space but we need to verify that every map in the proof is 1-Lipschitz. □

Remark: We could also have used the category of compact metric spaces and Lipschitz maps. Then TX would be isomorphic to the Kantorovich space.

Baire monad

Let C be the category of countable sets and maps $f : A \to B$ such that $f^{-1}(B')$ is finite or cofinite for every finite set B' of B.
For a countable set A denote GA for the subspace

$$\{(p_a)_a \in [0,1]^A \mid \sum_{a \in A} p_a = 1\} \subseteq [0,1]^A.$$

We have a functor $G : C \to \mathbf{Top}$.

Theorem

The codensity monad of G is a the monad of probability Baire measures.