Journal of Pure and Applied Algebra 220 (2016) 1229–1251

Contents lists available at ScienceDirect

Journal of Pure and Applied Algebra

www.elsevier.com/locate/jpaa

Codensity and the Giry monad

Tom Avery

School of Mathematics, University of Edinburgh, Edinburgh EH9 3FD, United Kingdom

ARTICLE INFO

Article history:
Received 4 November 2014
Received in revised form 1 July 2015
Available online 3 September 2015
Communicated by G. Rosolini

ABSTRACT

The Giry monad on the category of measurable spaces sends a space to a space of all probability measures on it. There is also a finitely additive Giry monad in which probability measures are replaced by finitely additive probability measures. We give a characterisation of both finitely and countably additive probability measures in terms of integration operators giving a new description of the Giry monads. This is then used to show that the Giry monads arise as the codensity monads of forgetful functors from certain categories of convex sets and affine maps to the category of measurable spaces.

1. Introduction

In general there are many different probability measures on a given measurable space, and the set of all of them can be made into a measurable space in a canonical way. Thus we have a process which turns a space into a new space whose points are the probability measures on the old one; this process is described in categorical language by a monad.

On the other hand, there is a standard categorical machine which turns a functor into a monad, namely the codensity monad of the functor. We show that the monad described above is the output of this machine when it is fed a natural forgetful functor involving certain convex sets. In other words, once we accept the mathematical importance of these convex sets (which may be taken to be all bounded, convex subsets of $\mathbb{R}^n$ together with the set of sequences in the unit interval converging to 0), then the notion of a probability measure is categorically inevitable.

The monad sending a measurable space to its space of probability measures is called the Giry monad, first defined in [7]. There are many variations of this monad; in [7] Giry defines both the monad mentioned above and a similar monad on the category of Polish spaces. In this paper we will be mainly concerned with Giry's monad on measurable spaces (which we refer to simply as the Giry monad), and a modification in which probability measures are replaced by finitely additive probability measures (the finitely additive Giry monad).

E-mail address: Tom.Avery@ed.ac.uk.

http://dx.doi.org/10.1016/j.jpaa.2015.08.017
0022-4049/

Note that there is a similar monad on **Set** that has sometimes been called the finitary Giry monad [6] or the distribution monad [10]. It sends a set to the set of formal convex combinations of its elements, which can be thought of as finitely supported probability measures. The algebras for this monad are abstract "convex spaces", which have been independently discovered and investigated several times, for example in [8,24] and [6]. This *finitary* Giry monad is not to be confused with the *finitely additive* Giry monad, although they behave similarly on finite sets (regarded as discrete measurable spaces).

The Kleisli category of the Giry monad has probability-theoretic significance [22]; it is the category of measurable spaces and "Markov kernels". As a simple example, a finite set (with discrete σ-algebra) equipped with an endomorphism in the Kleisli category of the Giry monad is precisely a discrete time Markov chain. In [5], Doberkat shows that the Eilenberg–Moore category of (the Polish space version of) the Giry monad is the category of continuous convex structures on Polish spaces with continuous affine maps.

The monads described above are examples of a loose family that we may think of as "measure monads"; in each instance, the monad sends a "space" to a space of "measures" on it, where we must interpret space and measure appropriately. Other examples include the ultrafilter monad (to which we shall return shortly), the probabilistic powerdomain [11], the distribution monad [6,10] and the monad defined by Lucyshyn-Wright in [18]. The idea of interpreting monads measure-theoretically has been extensively pursued by Kock in [16] and by Lucyshyn-Wright in [19].

A common theme for all these monads is "double dualisation". For any notion of a measure on a space X, there is a corresponding notion of integration. Integration takes functions $f: X \to R$ from the space to a set R of scalars (usually the reals, positive reals or the unit interval), and returns scalars in R. Such an integration operation can be thought of as an element of

$$\mathrm{Hom}(\mathrm{Hom}(X, R), R)$$

where the inner and outer Hom's must be interpreted appropriately in different contexts. Thus notions of measure are closely related to double dualisation.

In some circumstances the measures can be completely characterised by their integration operators. Perhaps the most well known instance of this phenomenon is the Riesz–Markov–Kakutani representation theorem [12], which says that the space of finite, signed, regular Borel measures on a compact Hausdorff space X is isomorphic to

$$\mathbf{NVS}(\mathbf{Top}(X, \mathbb{R}), \mathbb{R}),$$

(as a normed vector space) where **Top** is the category of topological spaces and continuous maps, and **NVS** is the category of normed vector spaces and bounded linear maps. In Section 3, we give a similar (but easier) characterisation of probability measures in terms of their integration operators, which is a correction of a claim of Sturtz [25], with many parts of the proof appearing there. Sturtz has since issued a corrected version of his paper [26].

Such characterisations might make us hope that there is some general categorical machinery for double dualisation that, when fed an appropriate and relatively simple input, naturally gives rise to measure monads, and the Giry monad in particular. Codensity monads provide such a categorical machine.

Codensity monads were first defined by Kock in [14], and the dual notion was studied independently by Appelgate and Tierney in [1] under the name "model-induced cotriple". Given a functor $U: \mathbb{C} \to \mathcal{M}$, the codensity monad of U (when it exists) is the right Kan extension T^U of U along itself. The universal property of Kan extensions equips T^U with a canonical monad structure. In [17], Leinster describes how the codensity monad can be thought of as a substitute for the monad induced by the adjunction between U and its left adjoint, even when the left adjoint does not exist. In particular, when the left adjoint *does* exist, the codensity monad is the usual monad induced by the adjunction.

Codensity monads can be seen as a form of double dualisation via the end formula,

$$T^U m = \int_{c\in\mathbb{C}} [\mathcal{M}(m, Uc), Uc].$$

At first glance "elements" of this object would appear to be families of integration operators, with the codomain of integration ranging over the objects of $\mathbb{C}$. However, in examples of interest, such a family is determined by its component at a single object i, say, of $\mathbb{C}$. The other objects serve to impose naturality conditions which force the i component to preserve certain algebraic structure which is encoded in the category $\mathbb{C}$. This idea will become clear in the proof of Theorem 5.8.

As observed, for example, in [17], the ultrafilter monad can be viewed as a measure monad in the following way. An ultrafilter on a set X consists of a set of subsets of X; thus it can be thought of as a map from the power set of X to $\{0, 1\}$. Viewing $\{0, 1\}$ as a subset of the unit interval I, it turns out that the functions

$$2^X \to I$$

corresponding to ultrafilters are precisely the finitely additive probability measures taking values in $\{0, 1\}$. This means that the ultrafilter monad is a primitive version of the finitely additive Giry monad.

The ultrafilter monad is the codensity monad of the inclusion of the category of finite sets into the category of sets; this was first proved by Kennison and Gildenhuys in [13] and brought to wider attention by Leinster in [17]. The main theorem of this paper (Theorem 5.8) is an analogous result for the Giry monads, with finite sets replaced by certain convex sets, and with sets replaced by measurable spaces.

Despite the influence of Sturtz's work on the characterisation of probability measures in terms of integration operators mentioned above, our main result (Theorem 5.8) is substantially different from that of [26]. Both seek to exhibit the Giry monad as the codensity monad of a particular functor, however the functors used differ in two significant respects: firstly, Sturtz uses the entire category of convex spaces (as defined in [6]) as the domain of the functor, whereas we will use a small subcategory of this; and secondly, Sturtz incorporates an element of double dualisation into the functor itself, even before taking the codensity monad, whereas we will use a more "direct" forgetful functor.

I am grateful to Tom Leinster for suggesting this topic to work on, for a lot of helpful advice, and for many enlightening discussions.

Conventions: We write **Meas** for the category of measurable spaces (i.e. sets equipped with a σ-algebra of subsets) and measurable maps. We will often refer to measurable spaces by their underlying sets, leaving the σ-algebra implicit.

We write I for the unit interval $[0, 1]$. When viewed as a measurable space, we always equip it with the Borel σ-algebra. Given sets $B \subseteq A$, we write $\chi_B : A \to I$ for the characteristic function of B. If $r \in I$, then $\bar{r} : A \to I$ denotes the constant function with value r.

If A is a set and m is an object of a category $\mathcal{M}$, then $[A, m]$ denotes the A power of m, that is, the product in $\mathcal{M}$ of A copies of m. In particular if $\mathcal{M} = \mathbf{Set}$, then $[A, m] = \mathbf{Set}(A, m)$, the set of functions from A to m.

The integral sign $\int$ has two meanings in this paper: the occurrences in Section 4, and the single occurrence in the introduction, represent the category theoretic notion of an end (see X.5 in [20]). All other instances represent integration with respect to a (possibly only finitely additive) probability measure.

2. The Giry monads

In this section we review some basic definitions relating to finitely additive probability measures. We then define the finitely additive Giry monad, and the Giry monad as a submonad.

Recall the following definitions.

Definition 2.1. Let (Ω, Σ) be a measurable space and $\pi: \Sigma \to I$ (where I is the unit interval). Suppose

- $\pi(\Omega) = 1$, and
- whenever $A, B \in \Sigma$ are disjoint, we have $\pi(A \cup B) = \pi(A) + \pi(B)$.

Then π is called a **finitely additive probability measure** on (Ω, Σ). Suppose additionally that,

- whenever $A_i \in \Sigma$ are pairwise disjoint for $i \in \mathbb{N}$, we have

$$\pi\left(\bigcup_{i=0}^{\infty} A_i\right) = \sum_{i=0}^{\infty} \pi(A_i).$$

Then π is called a **probability measure** on (Ω, Σ).

The general theory of integration of finitely additive measures, as developed in [3], is quite complex and subtle. There are several definitions of integration; we will be concerned with the D-integral. However, we will only be interested in integrating *measurable, bounded* functions against finitely additive *probability* measures, which makes it possible to simplify the definition considerably. Therefore we will for convenience briefly spell out how the integral is defined in this special case.

Let π be a finitely additive probability measure on Ω. Recall that a function $f: \Omega \to \mathbb{R}$ is **simple** if it is a linear combination of characteristic functions of measurable sets. The integral of a simple function is defined by

$$\int_{\Omega} \left(\sum_{i=1}^{n} a_i \chi_{A_i}\right) \mathrm{d}\pi = \sum_{i=1}^{n} a_i \pi(A_i),$$

and this does not depend on the choice of representation of the function. For an arbitrary measurable, bounded, non-negative function $f: \Omega \to \mathbb{R}$, the integral is defined by

$$\int_{\Omega} f \mathrm{d}\pi = \sup \left\{ \int_{\Omega} f' \,\mathrm{d}\pi \mid f' \text{ is simple and } f' \leq f \right\} \in \mathbb{R},$$

and this extends to functions that may take negative values in a standard way. Note that the fact that f is bounded guarantees that the supremum is finite. The following lemma is easily verified.

Lemma 2.2. *Let $f: \Omega \to \mathbb{R}$ be measurable, bounded and non-negative, and π a finitely additive probability measure on Ω. Then there is a sequence f_n of simple functions converging uniformly to f, and for any such sequence*

$$\int_{\Omega} f_n \,\mathrm{d}\pi \to \int_{\Omega} f \,\mathrm{d}\pi.$$

Moreover, $(f_n)_{n=1}^{\infty}$ can be taken to be a pointwise increasing (or decreasing) sequence. □

Many basic results on integration against probability measures hold true for finitely additive probability measures. In particular, integration is linear and order-preserving (4.4.13(ii) and (vi) in [3]), and the change of variables formula (Lemma 2.4 below) is valid. An important exception is that the monotone convergence theorem (and therefore the dominated convergence theorem) does not hold for finitely additive measures. In fact, the monotone convergence theorem holds if and only if the measure is countably additive.

We now move on to the definitions of the Giry monads.

Definition 2.3. Let Ω be a measurable space. Then $F\Omega$ is defined to be the set of finitely additive probability measures on Ω, equipped with the smallest σ-algebra such that

$$\begin{aligned} \mathrm{ev}_A \colon F\Omega &\to I \\ \pi &\mapsto \pi(A) \end{aligned}$$

is measurable for each measurable $A \subseteq \Omega$. We write $G\Omega \subseteq F\Omega$ for the set of (countably additive) probability measures, and equip it with the subspace σ-algebra.

Let $g\colon \Omega \to \Omega'$ be measurable. Then $Fg\colon F\Omega \to F\Omega'$ is defined by

$$Fg(\pi)(A') = \pi(g^{-1}(A'))$$

for each measurable $A' \subseteq \Omega'$ and $\pi \in F\Omega$. We define $Gg\colon G\Omega \to G\Omega'$ by restricting Fg to $G\Omega$.

We call $Fg(\pi)$ the **push-forward** of π along g, written as $g_*(\pi)$ by some authors. Integration for push-forward measures is described by the **change of variables formula**:

Lemma 2.4. *Let $g\colon \Omega \to \Omega'$ be measurable, $\pi \in F\Omega$ and $f\colon \Omega' \to \mathbb{R}$ be measurable and bounded. Then*

$$\int_\Omega f \circ g \,\mathrm{d}\pi = \int_{\Omega'} f \,\mathrm{d}Fg(\pi).$$

Proof. This is a familiar result for countably additive measures; see for example Chapter VIII Theorem C in [9]. The proof for finitely additive probability measures is identical. □

It is straightforward to check that the above definitions define functors $F, G\colon \mathbf{Meas} \to \mathbf{Meas}$. The following lemma will be used to show that the multiplication of each Giry monad is measurable, and also in Proposition 3.7 below.

Lemma 2.5. *Let $f\colon \Omega \to I$ be measurable. Then the map $\int_\Omega f \,\mathrm{d}(-)\colon F\Omega \to I$ defined by*

$$\pi \mapsto \int_\Omega f \,\mathrm{d}\pi$$

is measurable.

Proof. The inverse image of $[0, r] \subseteq I$ under this map is

$$\left\{ \pi \mid \int_\Omega f \,\mathrm{d}\pi \leq r \right\};$$

we must show that this is measurable. Let f_n be as in Lemma 2.2 and increasing. Then the set above can be written as

$$\bigcap_{n=1}^{\infty} \left\{ \pi \mid \int_{F\Omega} f_n \, \mathrm{d}\pi \leq r \right\}.$$

In the case that $f = \chi_A$ for $A \subseteq \Omega$ measurable, $\int_\Omega f \, \mathrm{d}(-) = \mathrm{ev}_A$ so is measurable by definition. Since integration is linear, and linear combinations of measurable functions are measurable, $\int_\Omega f \, \mathrm{d}(-)$ is also measurable when f is a simple function. Hence, returning to the case of an arbitrary measurable f, each of the sets appearing in the above intersection is measurable, and a countable intersection of measurable sets is measurable. $\square$

We now describe the monad structure on F and G.

Definition 2.6. Let Ω be a measurable space. The natural transformations

$$\eta^{\mathbb{F}} \colon \mathrm{id}_{\mathbf{Meas}} \to F \quad \text{and} \quad \mu^{\mathbb{F}} \colon FF \to F$$

are defined as follows. Let

$$\eta^{\mathbb{F}}_{\Omega}(\omega)(A) = \chi_A(\omega)$$

where $\omega \in \Omega$ and $A \subseteq \Omega$ is measurable, so $\eta^{\mathbb{F}}_{\Omega}(\omega)$ is the *Dirac* or *point measure* at ω. Let

$$\mu^{\mathbb{F}}_{\Omega}(\rho)(A) = \int_{F\Omega} \mathrm{ev}_A \, \mathrm{d}\rho.$$

Here $\rho \in FF\Omega$ is a finitely additive probability measure on $F\Omega$, and $A \subseteq \Omega$ is measurable, so in particular the map $\mathrm{ev}_A \colon F\Omega \to I$ is measurable by definition. Thus integrating it against ρ gives an element of I. The natural transformations

$$\eta^{\mathbb{G}} \colon \mathrm{id}_{\mathbf{Meas}} \to G \quad \text{and} \quad \mu^{\mathbb{G}} \colon GG \to G$$

are defined similarly. It is easy to check that these formulae do define finitely (resp. countably) additive probability measures on Ω.

Let us prove that $\mu^{\mathbb{F}}_{\Omega}$ is measurable. If we take $f = \mathrm{ev}_A$ in Lemma 2.5, then $\int_\Omega f \, \mathrm{d}(-)$ is the composite

$$\mathrm{ev}_A \circ \mu^{\mathbb{F}}_{\Omega} \colon FF\Omega \to F\Omega \to I,$$

hence this composite is measurable. Measurability of $\mu^{\mathbb{F}}_{\Omega}$ follows since the maps ev_A generate the σ-algebra on $F\Omega$. The proof for $\mu^{\mathbb{G}}_{\Omega}$ is similar and measurability of the units is obvious.

Proposition 2.7. *The above definitions give monads* $\mathbb{F} = (F, \eta^{\mathbb{F}}, \mu^{\mathbb{F}})$ *and* $\mathbb{G} = (G, \eta^{\mathbb{G}}, \mu^{\mathbb{G}})$ *on* **Meas**.

Proof. See [7] for $\mathbb{G}$. The proof for $\mathbb{F}$ is similar. Note that Giry invokes the monotone convergence theorem in the proof, however it can be replaced by an instance of Lemma 2.2. $\square$

We call $\mathbb{F}$ the **finitely additive Giry monad** and $\mathbb{G}$ the **Giry monad**.

3. Integration operators

We now turn to the characterisation of finitely and countably additive probability measures in terms of integration operators. This will be used in Section 5 to characterise the Giry monads as codensity monads.

Definition 3.1. Let Ω be a measurable space, and let ϕ be a function

$$\mathbf{Meas}(\Omega, I) \to I.$$

We say that ϕ is a **finitely additive integration operator** on Ω if,

- it is **affine**: $\phi(rf + (1-r)g) = r\phi(f) + (1-r)\phi(g)$ for all $f, g \in \mathbf{Meas}(\Omega, I)$ and $r \in I$, and
- it is **weakly averaging**: $\phi(\bar{r}) = r$ for all $r \in I$.

Recall $\bar{r}$ denotes the constant function with value r. In [25], finitely additive integration operators were called **weakly averaging affine functionals**. We call ϕ an **integration operator** (possibly with the qualification **countably additive** to avoid ambiguity) if, additionally,

- it **respects limits**: if $f_n \in \mathbf{Meas}(\Omega, I)$ is a sequence of measurable functions converging pointwise to 0, then $\phi(f_n)$ converges to 0.

Definition 3.2. Let Ω be a measurable space. Write $S\Omega$ for the set of finitely additive integration operators on Ω and $T\Omega$ for the set of integration operators. Equip $S\Omega$ with the smallest σ-algebra such that

$$\begin{aligned} \mathrm{ev}_f \colon S\Omega &\to I \\ \phi &\mapsto \phi(f) \end{aligned}$$

is measurable for each $f \in \mathbf{Meas}(\Omega, I)$, and define a σ-algebra on $S\Omega$ similarly.

Given $g\colon \Omega \to \Omega'$ in **Meas**, define $Sg\colon S\Omega \to S\Omega'$ by

$$Sg(\phi)(f) = \phi(f \circ g)$$

for $\phi \in S\Omega$ and $f \in \mathbf{Meas}(\Omega, I)$, and define Tg similarly. This makes S and T functors $\mathbf{Meas} \to \mathbf{Meas}$.

The following two lemmas show that (finitely additive) integration operators preserve more structure than the definition suggests, and they will be used often throughout the rest of this paper.

Lemma 3.3. *Let $\phi \in S\Omega$, $f, f' \in \mathbf{Meas}(\Omega, I)$ and $r \in [0, \infty)$. Then*

(i) *if $rf \in \mathbf{Meas}(\Omega, I)$ then $\phi(rf) = r\phi(f)$,*
(ii) *if $f + f' \in \mathbf{Meas}(\Omega, I)$ then $\phi(f + f') = \phi(f) + \phi(f')$, and*
(iii) *if $f \leq f'$ pointwise then $\phi(f) \leq \phi(f')$.*

Proof. (i) follows from the affine property and the fact that $\phi(\bar{0}) = 0$.

(ii) follows from the affine property and

$$\phi(f + f') = 2\phi\left(\frac{1}{2}f + \frac{1}{2}f'\right),$$

which is an instance of (i).

(iii) follows from (ii) applied to $f' = f + (f' - f)$, using the fact that $f' - f \in \mathbf{Meas}(\Omega, I)$ (in particular $\phi(f' - f)$ is defined and is ≥ 0). $\square$

Lemma 3.4. *Let $\phi \in T\Omega$. If $f_n, f \in \mathbf{Meas}(\Omega, I)$ such that $f_n \to f$ pointwise, then $\phi(f_n) \to \phi(f)$.*

Proof. Let

$$g_n(\omega) = \max(0, f_n(\omega) - f(\omega)) \quad \text{and} \quad h_n(\omega) = \max(0, f(\omega) - f_n(\omega)).$$

Then $g_n, h_n \to 0$ pointwise, and

$$f_n + h_n = g_n + f,$$

so the result follows from part (ii) of the previous lemma and the fact that $\phi(g_n) \to 0$ and $\phi(h_n) \to 0$. $\square$

The following lemma allows us to reduce propositions about finitely additive integration operators to special cases involving only simple functions.

Lemma 3.5. *For any $\phi \in S\Omega$ and $f \in \mathbf{Meas}(\Omega, I)$, we have*

$$\begin{aligned}\phi(f) &= \sup\{\phi(g) \mid g \in \mathbf{Meas}(\Omega, I) \textit{ is simple and } g \leq f\} \\ &= \inf\{\phi(g) \mid g \in \mathbf{Meas}(\Omega, I) \textit{ is simple and } g \geq f\}.\end{aligned}$$

Proof. Write $L = \{\phi(g) \mid g \in \mathbf{Meas}(\Omega, I) \text{ is simple and } g \leq f\}$. The inequality $\phi(f) \geq \sup L$ follows from Lemma 3.3(iii). For the other inequality, fix $\varepsilon > 0$ and choose g simple such that $g \leq f$ and $f - g \leq \bar{\varepsilon}$ (this is possible by Lemma 2.2). Then

$$\begin{aligned}\phi(f) &= \phi(g + (f - g)) = \phi(g) + \phi(f - g) \\ &\leq \phi(g) + \phi(\bar{\varepsilon}) = \phi(g) + \varepsilon \\ &\leq \sup L + \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, $\phi(f) \leq \sup L$. The other claim is proved similarly. $\square$

In the next three propositions we establish isomorphisms of functors $F \cong S$ and $G \cong T$, and in the fourth we transfer the monad structure of $\mathbb{F}$ and $\mathbb{G}$ across these isomorphisms. Parts of their proofs are due to Sturtz [25], however Sturtz incorrectly claims that $G \cong S$ rather than $F \cong S$, so we include the proofs here for clarity. Sturtz has since issued a corrected version [26].

Proposition 3.6. *Let Ω be a measurable space, and ϕ a finitely additive integration operator on Ω. Define $\Lambda(\phi)$ by*

$$\Lambda(\phi)(A) = \phi(\chi_A)$$

for $A \subseteq \Omega$ measurable. Then

(i) *Λ is a bijection $S\Omega \cong F\Omega$, and*
(ii) *Λ restricts to a bijection $\Xi\colon T\Omega \cong G\Omega$.*

Proof. (i) It is straightforward to check that $\Lambda(\phi)$ is a finitely additive probability measure.

Given $\pi \in F\Omega$, define $\tilde{\Lambda}(\pi)\colon \mathbf{Meas}(\Omega, I) \to I$ by

$$\tilde{\Lambda}(\pi)(f) = \int_{\Omega} f \, \mathrm{d}\pi$$

for $f \in \mathbf{Meas}(\Omega, I)$. The affine and weakly averaging properties of $\tilde{\Lambda}(\pi)$ are standard properties of integration against finitely additive measures (see 4.4.13(ii) in [3]), so $\tilde{\Lambda}$ does define a function $F\Omega \to S\Omega$.

Now we must show that Λ and $\tilde{\Lambda}$ are inverse to one another. In one direction,

$$\Lambda\tilde{\Lambda}(\pi)(A) = \int_{\Omega} \chi_A \, \mathrm{d}\pi = \pi(A)$$

for $\pi \in F\Omega$ and $A \subseteq \Omega$ measurable, by definition of the integral. In the other, first note that for $\phi \in S\Omega$ and $A \subseteq \Omega$ measurable,

$$\tilde{\Lambda}\Lambda(\phi)(\chi_A) = \int_{\Omega} \chi_A \, \mathrm{d}\Lambda(\phi) = \Lambda(\phi)(A) = \phi(\chi_A).$$

It follows by Lemma 3.3(i) and (ii) that $\tilde{\Lambda}\Lambda(\phi)(g) = \phi(g)$ for all simple $g \in \mathbf{Meas}(\Omega, I)$. Then if $f \in \mathbf{Meas}(\Omega, I)$,

$$\begin{aligned}
\phi(f) &= \sup\{\phi(g) \mid g \in \mathbf{Meas}(\Omega, I) \text{ is simple and } g \le f\} \\
&= \sup\{\tilde{\Lambda}\Lambda(\phi)(g) \mid g \in \mathbf{Meas}(\Omega, I) \text{ is simple and } g \le f\} \\
&= \tilde{\Lambda}\Lambda(\phi)(f),
\end{aligned}$$

by Lemma 3.5.

(ii) Suppose $\pi \in G\Omega$, and $f_n \to 0$ pointwise. Then

$$\int_{\Omega} f_n \, \mathrm{d}\pi \to 0$$

by the dominated convergence theorem. So $\tilde{\Lambda}(\pi) \in T\Omega$.

Now suppose $\phi \in T\Omega$, and that $(A_n)_{n=1}^{\infty}$ is a disjoint family of measurable sets. Write

$$B_k = \bigcup_{n=k}^{\infty} A_k$$

and $A = B_1$. Then $\chi_{B_k} \to 0$ pointwise, so $\Lambda(\phi)(B_k) \to 0$. But

$$\Lambda(\phi)(A) = \sum_{n=1}^{k-1} \Lambda(\phi)(A_n) + \Lambda(\phi)(B_k) \to \sum_{n=1}^{\infty} \Lambda(\phi)(A_n)$$

as $k \to \infty$, so $\Lambda(\phi) \in G\Omega$ as required. $\square$

Proposition 3.7. *The bijections Λ and $\tilde{\Lambda}$ are measurable (and hence so are Ξ and $\tilde{\Xi}$).*

Proof. The σ-algebra on $F\Omega$ is generated by the maps

$$\mathrm{ev}_A : F\Omega \to I$$

for measurable A, so in order to show that Λ is measurable it is sufficient to show that each $\mathrm{ev}_A \circ \Lambda$ is measurable. But the diagram

$$\begin{tikzcd} S\Omega \arrow[r, "\Lambda"] \arrow[dr, "\mathrm{ev}_{\chi_A}"'] & F\Omega \arrow[d, "\mathrm{ev}_A"] \\ & I \end{tikzcd}$$

commutes, and $\mathrm{ev}_{\chi_A} : S\Omega \to I$ is measurable by definition of the σ-algebra on $T\Omega$.

To show that $\tilde{\Lambda}$ is measurable we must show that $\mathrm{ev}_f \circ \tilde{\Lambda}$ is measurable for each $f \in \mathbf{Meas}(\Omega, I)$. But

$$\mathrm{ev}_f \circ \tilde{\Lambda}(\pi) = \int_\Omega f \, \mathrm{d}\pi,$$

So the composite is measurable by Lemma 2.5. □

Proposition 3.8. *The maps $\Lambda : S\Omega \to F\Omega$ and $\Xi : T\Omega \to G\Omega$ are natural in Ω.*

Proof. A straightforward verification, or see Theorem 4.4 in [25]. □

Thus Λ and Ξ are natural isomorphisms $S \cong F$ and $T \cong G$. Since F and G carry monad structures, there are unique monad structures on S and T making Λ and Ξ into morphisms of monads, giving an alternative description of the Giry monads.

Proposition 3.9. *The monad structure $\mathbb{S} = (S, \eta^{\mathbb{S}}, \mu^{\mathbb{S}})$ on S induced by Λ is given by*

$$\eta^{\mathbb{S}}_\Omega(\omega)(f) = f(\omega)$$

for $\Omega \in \mathbf{Meas}$, $\omega \in \Omega$ and $f \in \mathbf{Meas}(\Omega, I)$, and

$$\mu^{\mathbb{S}}_\Omega(\psi)(f) = \psi(\mathrm{ev}_f),$$

for $\psi \in SS\Omega$. Similarly for $\Xi : G \cong T$.

The second of these expressions deserves some explanation. Here $\psi \in SS\Omega$ is an affine and weakly averaging function

$$\psi : \mathbf{Meas}(S\Omega, I) \to I.$$

The elements of $S\Omega$ are functions $\mathbf{Meas}(\Omega, I) \to I$ and $f \in \mathbf{Meas}(\Omega, I)$, so we have $\mathrm{ev}_f : S\Omega \to I$, and this is measurable. Therefore ψ can be applied to ev_f, yielding an element of I.

Proof of Proposition 3.9. We know that the unit and multiplication of the induced monad structure on S make the diagrams

$$\begin{tikzcd} S\Omega \arrow[rr, "\Lambda_\Omega"] & & F\Omega \\ & \Omega \arrow[ul, "\eta^{\mathbb{S}}_\Omega"] \arrow[ur, "\eta^{\mathbb{F}}_\Omega"'] & \end{tikzcd}$$

and

$$\begin{tikzcd} SS\Omega \arrow[r, "\Lambda_{S\Omega}"] \arrow[d, "\mu^{\mathbb{S}}_\Omega"'] & FS\Omega \arrow[r, "F\Lambda_\Omega"] & FF\Omega \arrow[d, "\mu^{\mathbb{F}}_\Omega"] \\ S\Omega \arrow[rr, "\Lambda_\Omega"'] & & F\Omega \end{tikzcd}$$

commute. Therefore we have $\eta^{\mathbb{S}}_\Omega = \tilde{\Lambda}_\Omega \circ \eta^{\mathbb{F}}_\Omega$ and $\mu^{\mathbb{S}}_\Omega = \tilde{\Lambda}_\Omega \circ \mu^{\mathbb{F}}_\Omega \circ F\Lambda_\Omega \circ \Lambda_{S\Omega}$.

If $\omega \in \Omega$ and $f \in \mathbf{Meas}(\Omega, I)$ then

$$\begin{aligned} \eta^{\mathbb{S}}_\Omega(\omega)(f) &= \tilde{\Lambda}_\Omega \circ \eta^{\mathbb{F}}_\Omega(\omega)(f) \\ &= \int_\Omega f \,\mathrm{d}(\eta^{\mathbb{F}}_\Omega(\omega)) \\ &= f(\omega). \end{aligned}$$

Now suppose $\psi \in SS\Omega$. Then, if $A \subseteq \Omega$ is measurable, we have

$$\begin{aligned} \mu^{\mathbb{S}}_\Omega(\psi)(\chi_A) &= \mu^{\mathbb{F}}_\Omega \circ F\Lambda_\Omega \circ \Lambda_{S\Omega}(\psi)(A) \\ &= \int_{F\Omega} \mathrm{ev}_A \,\mathrm{d}(F\Lambda_\Omega \circ \Lambda_{S\Omega}(\psi)) \\ &= \int_{S\Omega} \mathrm{ev}_A \circ \Lambda_\Omega \,\mathrm{d}(\Lambda_{S\Omega}(\psi)) \\ &= \int_{S\Omega} \mathrm{ev}_{\chi_A} \,\mathrm{d}(\Lambda_{S\Omega}(\psi)) \\ &= \tilde{\Lambda}_{S\Omega} \circ \Lambda_{S\Omega}(\psi)(\mathrm{ev}_{\chi_A}) \\ &= \psi(\mathrm{ev}_{\chi_A}). \end{aligned}$$

Note that if $g = \sum_{i=1}^n a_i \chi_{A_i}$ is a simple function in $\mathbf{Meas}(\Omega, I)$ then $\mathrm{ev}_g = \sum_{i=1}^n a_i \,\mathrm{ev}_{\chi_{A_i}}$ as elements of $\mathbf{Meas}(S\Omega, I)$, so

$$\mu^{\mathbb{S}}_\Omega(\psi)(g) = \sum_{i=1}^n a_i \mu^{\mathbb{S}}_\Omega(\psi)(\chi_{A_i}) = \sum_{i=1}^n a_i \psi(\mathrm{ev}_{\chi_{A_i}}) = \psi(\mathrm{ev}_g).$$

Now if $f \in \mathbf{Meas}(\Omega, I)$, we have

$$\begin{aligned} \mu^{\mathbb{S}}_\Omega(\psi)(f) &= \sup\{\mu^{\mathbb{S}}_\Omega(\psi)(g) \mid g \in \mathbf{Meas}(\Omega, I) \text{ is simple and } g \leq f\} \\ &= \sup\{\psi(\mathrm{ev}_g) \mid g \in \mathbf{Meas}(\Omega, I) \text{ is simple and } g \leq f\} \\ &\leq \psi(\mathrm{ev}_f) \end{aligned}$$

$$\begin{aligned}
&\le \inf\{\psi(\mathrm{ev}_g) \mid g \in \mathbf{Meas}(\Omega, I) \text{ is simple and } g \ge f\} \\
&= \inf\{\mu^{\mathbb{S}}_\Omega(\psi)(g) \mid g \in \mathbf{Meas}(\Omega, I) \text{ is simple and } g \ge f\} \\
&= \mu^{\mathbb{S}}_\Omega(\psi)(f),
\end{aligned}$$

where the first and last equalities are Lemma 3.5, and the inequalities are due to the facts that if $f \le g$ then $\mathrm{ev}_f \le \mathrm{ev}_g$ and that ψ is order-preserving. Hence $\mu^{\mathbb{S}}_\Omega(\psi)(f) = \psi(\mathrm{ev}_f)$ as required. The proof for Ξ is similar. □

4. Review of codensity monads

In this section we review the basics of codensity monads. A more thorough introduction can be found in [17]. The main purpose of this section, besides a review of the definitions, is to obtain a description of codensity monads that will make it easy to establish an isomorphism of monads between the codensity monads defined in Section 5 and the monads defined in Section 3. This description is given by Equations (1), (2) and (3).

Let $\mathbb{C}$ be a small category, $\mathcal{M}$ a complete, locally small category, and $U\colon \mathbb{C} \to \mathcal{M}$ a functor. Then the right Kan extension of U along itself always exists; it consists of a functor $T^U\colon \mathcal{M} \to \mathcal{M}$ and a natural transformation $\kappa^U\colon T^U U \to U$, which are defined by the following universal property: if $H\colon \mathcal{M} \to \mathcal{M}$ and $\lambda\colon HU \to U$, then there is a unique $\tau\colon H \to T^U$ such that

$$\begin{tikzcd}
\mathbb{C} \arrow[r, "U"] \arrow[dr, "U"'] \arrow[dr, phantom, shift left=4, "\Swarrow\kappa^U"] & \mathcal{M} \arrow[d, bend right=40, "T^U"'] \arrow[d, bend left=40, "H"] \arrow[d, phantom, "\Leftarrow\tau"] \\
& \mathcal{M}
\end{tikzcd}
=
\begin{tikzcd}
\mathbb{C} \arrow[r, "U"] \arrow[dr, "U"'] \arrow[dr, phantom, shift left=4, "\Swarrow\lambda"] & \mathcal{M} \arrow[d, "H"] \\
& \mathcal{M}.
\end{tikzcd}$$

We make T^U the endofunctor part of a monad $\mathbb{T}^U = (T^U, \eta^U, \mu^U)$, called the **codensity monad of** G, defining η^U and μ^U using the universal property of T^U as follows:

$$\begin{tikzcd}
\mathbb{C} \arrow[r, "U"] \arrow[dr, "U"'] \arrow[dr, phantom, shift left=4, "\Swarrow\kappa^U"] & \mathcal{M} \arrow[d, bend right=40, "T^U"'] \arrow[d, bend left=40, "\mathrm{id}_{\mathcal{M}}"] \arrow[d, phantom, "\Leftarrow\eta^U"] \\
& \mathcal{M}
\end{tikzcd}
=
\begin{tikzcd}
\mathbb{C} \arrow[r, "U"] \arrow[dr, "U"'] \arrow[dr, phantom, shift left=4, "//"] & \mathcal{M} \arrow[d, equal] \\
& \mathcal{M}
\end{tikzcd}$$

and

$$\begin{tikzcd}
\mathbb{C} \arrow[r, "U"] \arrow[ddr, "U"'] \arrow[ddr, phantom, shift left=4, "\Swarrow\kappa^U"] & \mathcal{M} \arrow[dd, "T^U"] \arrow[dr, "T^U"] & \\
& & \mathcal{M} \arrow[dl, "T^U"] \arrow[l, phantom, "\Leftarrow\mu^U"] \\
& \mathcal{M} &
\end{tikzcd}
=
\begin{tikzcd}
\mathbb{C} \arrow[r, "U"] \arrow[dr, "U"] \arrow[ddr, "U"'] \arrow[dr, phantom, shift left=4, "\Swarrow\kappa^U"] \arrow[ddr, phantom, shift left=3, "\Swarrow\kappa^U"] & \mathcal{M} \arrow[d, "T^U"] \\
& \mathcal{M} \arrow[d, "T^U"] \\
& \mathcal{M}.
\end{tikzcd}$$

The fact that these maps satisfy the monad axioms follows from the uniqueness part of the universal property.

It will also be useful to have a more explicit description of η^U and μ^U. The end formula for right Kan extensions ([20] X.4) gives

$$T^U m \cong \int_{c\in\mathbb{C}} [\mathcal{M}(m, Uc), Uc].$$

Let

$$\mathrm{ev}_f\colon \int_{c\in\mathbb{C}} [\mathcal{M}(m, Uc), Uc] \to Uc$$

be the canonical limit projection (where $f\colon m \to Uc$ in $\mathcal{M}$).

If $g\colon m \to m'$, then $T^U g$ is defined to be the unique morphism making

$$\begin{tikzcd} T^U m \arrow[d, "T^U g"'] \arrow[dr, "\mathrm{ev}_{f\circ g}"] & \\ T^U m' \arrow[r, "\mathrm{ev}_f"'] & Uc \end{tikzcd}$$

commute for each $f\colon m' \to Uc$.

We will now describe the functor and monad structure of T^U in terms of generalised elements. Recall that if $m \in \mathcal{M}$, a **generalised element** e **with shape** $s \in \mathcal{M}$, or s**-element**, of m is simply a morphism $e\colon s \to m$, and we write

$$e \in_s m.$$

The shape s has also been called the **stage of definition** of e, for example in [15].

Any morphism $g\colon m \to m'$ defines a function (also denoted g) mapping s-elements of m to s-elements of m': if $e \in_s m$, then

$$g(e) = g \circ e \in_s m'.$$

Furthermore, a consequence of the Yoneda lemma is that any such function defined on generalised elements corresponds to a unique morphism $m \to m'$ (provided it is natural in s). This provides a convenient way of describing morphisms in $\mathcal{M}$. Note that

$$\begin{aligned}\mathcal{M}(s, T^U m) &\cong \mathcal{M}\left(s, \int_{c\in\mathbb{C}} [\mathcal{M}(m, Uc), Uc]\right)\\ &\cong \int_{c\in\mathbb{C}} [\mathcal{M}(m, Uc), \mathcal{M}(s, Uc)],\end{aligned}$$

so, by the nature of limits in **Set**, an s-element of $T^U m$ can be thought of as a family of functions (natural in c) that map morphisms $m \to Uc$ to s-elements of Uc. Given $\alpha \in_s T^U m$, and $f\colon m \to Uc$,

$$\mathrm{ev}_f(\alpha) = \alpha_c(f) \in_s Uc.$$

Thus, in terms of generalised elements, the functor T^U is defined by

$$((T^U g)(\alpha))_c(f) = \mathrm{ev}_f \circ (T^U g)(\alpha) = \mathrm{ev}_{f\circ g}(\alpha) = \alpha_c(f \circ g), \tag{1}$$

where $g\colon m \to m'$ and $f\colon m' \to Uc$. In [14], Kock describes η^U and μ^U in terms of the equations

$$\mathrm{ev}_f \circ \eta^U_m = f \quad \text{and} \quad \mathrm{ev}_f \circ \mu^U_m = \mathrm{ev}_{\mathrm{ev}_f}$$

for each $c \in \mathbb{C}$ and $f \in \mathcal{M}(m, Uc)$. Translating these into generalised element notation, η^U and μ^U are defined by

$$(\eta^U_m(e))_c(f) = \mathrm{ev}_f \circ \eta^U_m(e) = f(e) \tag{2}$$

and

$$(\mu^U_m(\beta))_c(f) = \mathrm{ev}_f \circ \mu^U_m(\beta) = \mathrm{ev}_{\mathrm{ev}_f}(\beta) = \beta_c(\mathrm{ev}_f) \tag{3}$$

where $e \in_s m$ and $\beta \in_s T^U T^U m$.

5. Probability measures via codensity

We will now show that the finitely additive Giry monad and the Giry monad arise as codensity monads.

Definition 5.1. A **convex set** is a convex subset of a real vector space. That is, $c \subseteq V$ is a convex set if for all $x, y \in c$ and $r \in I$ we have $rx + (1-r)y \in c$. We write

$$x +_r y = rx + (1-r)y.$$

If c, c' are convex sets, then an **affine map** $h\colon c \to c'$ is a function such that

$$h(x +_r y) = h(x) +_r h(y)$$

for all $x, y \in c$ and $r \in I$.

There is a more abstract notion of a convex space, investigated in [6], namely an algebra for the distribution monad mentioned in the introduction. These more general convex spaces are used by Sturtz in [25]. However, all the convex spaces we will be concerned with are convex subsets of vector spaces, so we omit the more general definition.

We choose the term “affine map” rather than “convex map” to avoid confusion with the notion of a “convex function” (a real-valued function with convex epigraph). This is potentially ambiguous: the term affine is already used for a map between vector spaces that preserves *affine* combinations (i.e. linear combinations of the form $rx + (1-r)y$ where $r \in \mathbb{R}$) rather than just *convex* combinations (those for which $r \in I$). However it is easily seen that a map preserving convex combinations also preserves whatever affine combinations exist in the domain. Moreover, we have the following useful result:

Lemma 5.2. *Let c and c' be convex subsets of real vector spaces V and V', and let $h\colon c \to c'$ be an affine map. Then h has a unique affine extension* $\mathrm{aff}(c) \to \mathrm{aff}(c')$*, where*

$$\mathrm{aff}(c) = \{ru + (1-r)v \mid u, v \in c \text{ and } r \in \mathbb{R}\}$$

is the ***affine span*** *of c in V.*

Proof. Define

$$h(ru + (1-r)v) = rh(u) + (1-r)h(v).$$

It is straightforward to check that this is well-defined and affine. □

The following corollary will be used in the proof of Proposition 5.11.

Corollary 5.3. *Let $\phi\colon \mathbf{Meas}(\Omega, I) \to I$ be a finitely additive integration operator. Then ϕ has a unique linear extension*

$$\phi\colon \mathbf{Meas}_b(\Omega, \mathbb{R}) \to \mathbb{R},$$

where $\mathbf{Meas}_b(\Omega, \mathbb{R})$ denotes the vector space of bounded measurable maps $\Omega \to \mathbb{R}$.

Proof. Regarding $\mathbf{Meas}(\Omega, I)$ as a convex set in $\mathbf{Meas}(\Omega, \mathbb{R})$, we have

$$\mathrm{aff}(\mathbf{Meas}(\Omega, I)) = \mathbf{Meas}_b(\Omega, \mathbb{R}),$$

and $\mathrm{aff}(I) = \mathbb{R}$ as a subset of $\mathbb{R}$, so there is a unique affine extension by the previous lemma. Moreover, since ϕ preserves 0, the extension is in fact linear. □

The domain categories of the functors whose codensity monads we will prove to be the Giry monads are both full subcategories of the category of convex sets.

Definition 5.4.

(i) Let c_0 be the vector space of real sequences converging to 0, and let $d_0 \subseteq c_0$ be the (convex) set of sequences in c_0 contained entirely in I. We will occasionally mention the sup-norm on c_0 defined by

$$\|x\|_\infty = \sup_n |x_n|.$$

(ii) Let $\mathbb{C}$ be the category whose objects are all finite powers of I (including $1 = I^0$) and all affine maps between them.

(iii) Let $\mathbb{D}$ be the category whose objects are all finite powers of I, together with d_0, and all affine maps between them.

Proposition 5.5.

(i) *Every affine $h\colon I^n \to I$ is of the form*

$$h(x) = a_0 + \sum_{i=1}^{n} a_i x_i$$

for some $a_i \in \mathbb{R}$.

(ii) *Every affine $h\colon d_0 \to I$ is of the form*

$$h(x) = a_0 + \sum_{i=1}^{\infty} a_i x_i$$

for some $a_i \in \mathbb{R}$ with $\sum_{i=1}^{\infty} |a_i| < \infty$.

Proof. (i) By Lemma 5.2, there is a unique extension of h to an affine map $\mathbb{R}^n \to \mathbb{R}$. But any affine map between vector spaces can be written as a linear map followed by a translation of the codomain. The general form of such a map $\mathbb{R}^n \to \mathbb{R}$ is as claimed.

(ii) As in (i), h has a unique affine extension $c_0 \to \mathbb{R}$ (since $\mathrm{aff}(d_0) = c_0$), and this can be written as a linear map followed by a translation; write h' for the linear part. We claim that h' is continuous with respect to the sup-norm on c_0:

For subsets A and B of a vector space, write

$$A - B = \{a - b \mid a \in A, b \in B\}$$

Then

$$h'(d_0 - d_0) = h'(d_0) - h'(d_0) \subseteq I - I = [-1, 1],$$

but $d_0 - d_0$ is the unit ball in c_0 with the sup-norm, and h' maps it into a bounded set, so h' is continuous.

But a continuous linear functional on c_0 is of the form

$$x \mapsto \sum_{i=1}^{\infty} a_i x_i$$

for some $a_i \in \mathbb{R}$ with $\sum_{i=1}^{\infty} |a_i| < \infty$ (this fact is a common exercise in courses on functional analysis; see for example Exercise 1 in Chapter 3 of [4]). So h is as claimed. $\square$

Every object of $\mathbb{D}$ can be given a measurable space structure as a subspace of a product of copies of I (recall that I is always given the Borel σ-algebra).

Proposition 5.6. *All the maps in $\mathbb{D}$ are measurable.*

Proof. A map $c \to c'$ in $\mathbb{D}$ is measurable if and only if its composite with each projection $c' \to I$ is measurable, so it is sufficient to show that affine maps $h: c \to I$ are measurable.

If $c = I^n$, then h is of the form described in Proposition 5.5(i), and is measurable, since all the basic arithmetic operations are.

Suppose $c = d_0$ and h is of the form described in Proposition 5.5(ii).

Now, consider the topology on d_0 as a subset of c_0 with the sup-norm. A basic open set for this topology is of the form

$$U = d_0 \cap \prod_{i=1}^{\infty} (x_i - \varepsilon, x_i + \varepsilon)$$

for some $x \in d_0$ and $\varepsilon > 0$. The σ-algebra on d_0 is generated by sets of the form $\prod_{i=1}^{\infty} A_i$, where $A_i = I$ for all but one i, say i_0, and A_{i_0} is measurable. Clearly a basic open set can be written as a countable intersection of such sets, so is measurable. On the other hand, d_0 is a separable metric space (a countable dense set is given by the sequences of rationals that are eventually 0), and therefore second countable, by a standard exercise in topology (e.g. Exercise 2.23 in [23]). Moreover, the countable base we obtain is contained in the original base, and so consists of measurable sets. Hence every open subset is a countable union of measurable sets, so is measurable, and it follows that a norm-continuous function $d_0 \to I$ is measurable. But h is the composite of a continuous linear functional on c_0 and a translation of $\mathbb{R}$, so is continuous, and hence measurable. $\square$

Corollary 5.7. *There are natural forgetful functors $U: \mathbb{C} \to$ **Meas** and $V: \mathbb{D} \to$ **Meas**.* $\square$

We can now state the main theorem of this paper:

Theorem 5.8.

(i) *The codensity monad* $\mathbb{T}^U$ *of* $U: \mathbb{C} \to \mathbf{Meas}$ *is isomorphic to the finitely additive Giry monad.*
(ii) *The codensity monad* $\mathbb{T}^V$ *of* $V: \mathbb{D} \to \mathbf{Meas}$ *is isomorphic to the Giry monad.*

The proof will follow shortly, but first let us make some general observations about the measurable space $T^U\Omega$. We saw in Section 4 that an s-element of $T^U\Omega$ is a family α of functions

$$\alpha_c: \mathbf{Meas}(\Omega, Uc) \to \mathbf{Meas}(s, Uc),$$

natural in c. In particular, an *ordinary* element of $T^U\Omega$ (which is the same as a generalised element of shape 1) is a natural family of functions

$$\alpha_c: \mathbf{Meas}(\Omega, Uc) \to Uc.$$

The σ-algebra on $T^U\Omega$ is the smallest such that

$$\mathrm{ev}_f: T^U\Omega \to Uc$$

is measurable, for each $f \in \mathbf{Meas}(\Omega, Uc)$.

Lemma 5.9. *Let* $\alpha \in T^U\Omega$*. Then*

$$\alpha_{I^n} = (\alpha_I)^n: \mathbf{Meas}(\Omega, I^n) \cong \mathbf{Meas}(\Omega, I)^n \to I^n.$$

The same is true if $\alpha \in T^V\Omega$*, and then* α_{d_0} *is also obtained by applying* α_I *componentwise.*

Proof. Let

$$f = (f_1, \ldots, f_n) \in \mathbf{Meas}(\Omega, I^n)$$

By commutativity of

$$\begin{tikzcd} \mathbf{Meas}(\Omega, I^n) \arrow[r, "\alpha_{I^n}"] \arrow[d, "(\pi_i)_*"'] & I^n \arrow[d, "\pi_i"] \\ \mathbf{Meas}(\Omega, I) \arrow[r, "\alpha_I"'] & I \end{tikzcd}$$

we have $(\alpha_{I^n}(f))_i = \alpha_I(f_i)$, as required. The proof for d_0 is similar. $\square$

Proof of Theorem 5.8. (i) We will establish a bijection between $T^U\Omega$ and $S\Omega$, where S is as in Definition 3.2. Given $\alpha \in T^U\Omega$ we claim that

$$\alpha_I: \mathbf{Meas}(\Omega, I) \to I,$$

is affine and weakly averaging, i.e. an element of $S\Omega$. Suppose $f, g \in \mathbf{Meas}(\Omega, I)$, and $r \in I$. The map

$$+_r: I^2 \to I$$

is affine, so

$$\begin{tikzcd} \mathbf{Meas}(\Omega, I^2) \arrow[r, "\alpha_I^2"] \arrow[d, "(+_r)_*"'] & I^2 \arrow[d, "+_r"] \\ \mathbf{Meas}(\Omega, I) \arrow[r, "\alpha_I"'] & I \end{tikzcd}$$

commutes, and following (f, g) around this diagram yields

$$\alpha_I(f) +_r \alpha_I(g) = \alpha_I(f +_r g),$$

so α_I is affine. The fact that for $r \in I$, we have $\alpha_I(\bar{r}) = r$ follows from commutativity of

$$\begin{tikzcd} \mathbf{Meas}(\Omega, 1) \arrow[r] \arrow[d, "r_*"'] & 1 \arrow[d, "r"] \\ \mathbf{Meas}(\Omega, I) \arrow[r, "\alpha_I"'] & I. \end{tikzcd}$$

So α_I is a finitely additive integration operator. Now suppose $\phi \in S\Omega$ is a finitely additive integration operator. Define

$$\alpha_{I^n} \colon \mathbf{Meas}(\Omega, I^n) \to I^n$$

by

$$\alpha_{I^n}(f_1, \dots, f_n) = (\phi(f_1), \dots, \phi(f_n)).$$

We must check that this is natural with respect to all maps in $\mathbb{C}$. Since any function into I^n is determined by its composites with the projections, it is sufficient to check naturality with respect to maps with codomain I. Suppose $h \colon I^n \to I$ is of the form

$$h(x) = a_0 + \sum_{i=1}^{n} a_i x_i$$

from Proposition 5.5. Then if $f \in \mathbf{Meas}(\Omega, I^n)$,

$$\begin{aligned} h \circ \alpha_{I^n}(f) &= h(\phi(f_1), \dots, \phi(f_n)) \\ &= a_0 + \sum_{i=1}^{n} a_i \phi(f_i) \\ &= \phi\left(a_0 + \sum_{i=1}^{n} a_i f_i\right) \\ &= \phi(h \circ f) \\ &= \alpha_I \circ h_*(f) \end{aligned}$$

as required (where the third equality comes from implicitly identifying ϕ with its linear extension from Corollary 5.3, and the weakly averaging property). It is immediate that these assignments

$$T^U\Omega \to S\Omega \quad \text{and} \quad S\Omega \to T^U\Omega$$

are inverse to each other, and measurable.

To see that these bijections are natural and respect the monad structures on T^U and S, we must establish the commutativity of certain diagrams. Recall that the functor and monad structures of S are defined in Definition 3.2 and Proposition 3.9 respectively. A description of the relevant structure on T^U is given by Equations (1), (2) and (3) of Section 4 with $s = 1$, so that these become statements about ordinary, rather than generalised elements. From these facts, and recalling that the (unnamed) bijection $T^U\Omega \cong S\Omega$ is given by sending $\alpha \in T^U\Omega$ to α_I, it is straightforward to check that the relevant diagrams commute.

(ii) Let $\alpha \in T^V\Omega$. As before, α_I is affine and weakly averaging; now we show it respects limits. Suppose $f_n\colon \Omega \to I$ is a sequence of measurable functions converging pointwise to 0. Then f defines an element of $\mathbf{Meas}(\Omega, d_0)$. By Lemma 5.9,

$$(\alpha_{d_0}(f))_i = \alpha_I(f_i),$$

so since $\alpha_{d_0}(f) \in d_0$, we must have $\alpha_I(f_i) \to 0$.

Now suppose ϕ is an integration operator. Let α_{I^n} be defined as in (i), and define $\alpha_{d_0}(f)_i = \phi(f_i)$ for $f \in \mathbf{Meas}(\Omega, d_0)$. The fact that ϕ preserves limits of sequences converging to 0 means that α_{d_0} does map into d_0. Once again we only need to check that α is natural with respect to maps with codomain I, and for maps out of I^n this is as before.

Suppose $h\colon d_0 \to I$ is affine, say

$$h(x) = a_0 + \sum_{i=1}^{\infty} a_i x_i.$$

Then

$$\begin{aligned} h \circ \alpha_{d_0}(f) &= h\left((\phi(f_i))_{i=1}^{\infty}\right) \\ &= a_0 + \sum_{i=1}^{\infty} a_i \phi(f_i) \\ &= \lim_{N\to\infty} \left(a_0 + \sum_{i=1}^{N} a_i \phi(f_i)\right) \\ &= \lim_{N\to\infty} \phi\left(a_0 + \sum_{i=1}^{N} a_i f_i\right) \\ &= \phi\left(a_0 + \sum_{i=1}^{\infty} a_i f_i\right) \\ &= \phi(h \circ f) \\ &= \alpha_I \circ h_*(f), \end{aligned}$$

as required. Here we have again implicitly used the linear extension of ϕ from Corollary 5.3, and also the result that ϕ preserves all limits (Lemma 3.4). As in (i), the remainder of the proof is a series of straightforward checks. □

Note that in the preceding proof we only made use of the objects 1, I, I^2, and in part (ii), d_0. Thus we could have taken $\mathbb{C}$ and $\mathbb{D}$ to be the categories with just these objects and affine maps between them. In fact, even more is true:

Proposition 5.10. *Let M and N be the monoids of affine endomorphisms of I^2 and d_0 respectively. Then*

(i) *The codensity monad $\mathbb{T}^M$ of the action of M on I^2 in* **Meas** *is the finitely additive Giry monad.*
(ii) *The codensity monad $\mathbb{T}^N$ of the action of N on d_0 in* **Meas** *is the Giry monad.*

Recall that an action of a monoid on an object of a category is essentially the same as a functor from the monoid (regarded as a category with one object) to the category, so it makes sense to talk about the codensity monad of an action.

Proof of Proposition 5.10. We will prove (ii); (i) is similar. It is clear from Theorem 5.8 that an integration operator on Ω will define an element of $T^N\Omega$. Given $\alpha \in T^N\Omega$, which we regard as a function $\mathbf{Meas}(\Omega, d_0) \to d_0$ that commutes with affine endomorphisms of d_0, we must construct an integration operator. Define

$$\begin{aligned}
\iota_0 &: 1 \to d_0, && \text{an arbitrary map,}\\
\iota_1 &: I \to d_0, && x \mapsto (x, 0, 0, \ldots),\\
\iota_2 &: I^2 \to d_0, && (x_1, x_2) \mapsto (x_1, x_2, 0, \ldots),\\
\pi'_i &: d_0 \to d_0, && (x_1, x_2, \ldots) \mapsto (x_i, 0, \ldots),\\
+'_r &: d_0 \to d_0, && (x_1, x_2, \ldots) \mapsto (x_1 +_r x_2, 0 \ldots),\\
r' &: d_0 \to d_0, && (x_1, x_2, \ldots) \mapsto (r, 0, \ldots)
\end{aligned}$$

(where $r \in I$), and let ϕ be the composite

$$\mathbf{Meas}(\Omega, I) \xrightarrow{(\iota_1)_*} \mathbf{Meas}(\Omega, d_0) \xrightarrow{\alpha} d_0 \xrightarrow{\pi_1} I.$$

Then α is obtained by applying ϕ componentwise, by the commutativity of

$$\begin{tikzcd}
\mathbf{Meas}(\Omega, d_0) \arrow[r, equal] \arrow[d, "(\pi_i)_*"'] & \mathbf{Meas}(\Omega, d_0) \arrow[r, "\alpha"] \arrow[d, "(\pi'_i)_*"'] & d_0 \arrow[r, equal] \arrow[d, "\pi'_i"'] & d_0 \arrow[d, "\pi_i"] \\
\mathbf{Meas}(\Omega, I) \arrow[r, "(\iota_1)_*"'] & \mathbf{Meas}(\Omega, d_0) \arrow[r, "\alpha"'] & d_0 \arrow[r, "\pi_1"'] & I,
\end{tikzcd}$$

for each i. In particular, since $\alpha(f) \in d_0$, it follows that ϕ respects limits. The affine and weakly averaging properties of ϕ follow from the commutativity of

$$\begin{tikzcd}
\mathbf{Meas}(\Omega, I^2) \arrow[r, "(\iota_2)_*"'] \arrow[d, "(+_r)_*"'] \arrow[rrr, bend left, "\phi^2"] & \mathbf{Meas}(\Omega, d_0) \arrow[r, "\alpha"'] \arrow[d, "(+'_r)_*"'] & d_0 \arrow[r, "(\pi_1, \pi_2)"'] \arrow[d, "+'_r"'] & I^2 \arrow[d, "+_r"] \\
\mathbf{Meas}(\Omega, I) \arrow[r, "(\iota_1)_*"] \arrow[rrr, bend right, "\phi"'] & \mathbf{Meas}(\Omega, d_0) \arrow[r, "\alpha"] & d_0 \arrow[r, "\pi_1"] & I
\end{tikzcd}$$

and

$$\begin{tikzcd}
\mathbf{Meas}(\Omega,1) \arrow[r, "(\iota_0)_*"'] \arrow[d, "r_*"'] \arrow[rrr, bend left, "\phi^0"] & \mathbf{Meas}(\Omega,d_0) \arrow[r, "\alpha"'] \arrow[d, "r'_*"'] & d_0 \arrow[r] \arrow[d, "r'"'] & 1 \arrow[d, "r"] \\
\mathbf{Meas}(\Omega,I) \arrow[r, "(\iota_1)_*"] \arrow[rrr, bend right, "\phi"'] & \mathbf{Meas}(\Omega,d_0) \arrow[r, "\alpha"] & d_0 \arrow[r, "\pi_1"] & I
\end{tikzcd}$$

respectively. □

The preceding proposition gives categories of convex sets that are in some sense minimal (although not uniquely so) such that the codensity monads of their inclusions into **Meas** are the Giry monads. It is natural to ask how large a category of convex sets (or even convex spaces in the sense of [6]) can be and still give rise to the Giry monad. We have not answered this question precisely, but the following proposition at least gives a class of convex sets that can be included in the domain category without altering the codensity monad.

Proposition 5.11. *Let $\mathbb{C}'$ be the category of compact, convex subsets of $\mathbb{R}^n$ (where n can vary) with affine maps between them and let $\mathbb{D}'$ be similar but with d_0 adjoined. Then the codensity monads of the forgetful functors $U'\colon \mathbb{C}' \to$ **Meas** and $V'\colon \mathbb{D}' \to$ **Meas** are the finitely additive Giry monad and the Giry monad respectively.*

Proof. An element of $T^{U'}\Omega$ is a family of functions $\mathbf{Meas}(\Omega, U'c') \to U'c'$ natural in $c' \in \mathbb{C}'$. Since $\mathbb{C} \subseteq \mathbb{C}'$, and U is the restriction of U', such a family restricts to a family $\mathbf{Meas}(\Omega, Uc) \to Uc$ natural in $c \in \mathbb{C}$, that is, an element of $T^U\Omega \cong S\Omega$. Therefore we just have to check that every element of $T^U\Omega$ has a unique extension to an element of $T^{U'}\Omega$. Similarly for V and V'.

Suppose ϕ is a finitely additive integration operator on Ω and c a compact convex subset of $\mathbb{R}^n$. Write ϕ also for the unique linear extension of ϕ to

$$\mathbf{Meas}_b(\Omega, \mathbb{R}) \to \mathbb{R},$$

which exists by Corollary 5.3. We can define

$$\alpha_c\colon \mathbf{Meas}(\Omega, c) \to \mathbb{R}^n$$

by applying ϕ in each coordinate. We will now show that

(i) If $h\colon \mathbb{R}^n \to \mathbb{R}^m$ is affine then

$$\begin{tikzcd}
\mathbf{Meas}(\Omega,c) \arrow[r, "\alpha_c"] \arrow[d, "h_*"'] & \mathbb{R}^n \arrow[d, "h"] \\
\mathbf{Meas}(\Omega,h(c)) \arrow[r, "\alpha_{h(c)}"'] & \mathbb{R}^m
\end{tikzcd}$$

commutes, and

(ii) If $f\colon \Omega \to c$ is measurable then $\alpha_c(f) \in c$ (this is presumably known but we were unable to find a reference).

(i) Since $\mathbb{R}^m$ is a power of $\mathbb{R}$ it is sufficient to consider $h\colon \mathbb{R}^n \to \mathbb{R}$. Such an h is of the form

$$(x_1, \ldots, x_n) \mapsto a_0 + a_1 x_1 + \ldots a_n x_n,$$

and the fact that α commutes with such maps follows from linearity and the weakly averaging property of ϕ.

(ii) Let $f \in \mathbf{Meas}(\Omega, c)$, and suppose for a contradiction that $\alpha_c(f) \notin c$. By applying an affine change of coordinates, which we may do without loss of generality using (i), we may assume that $\alpha_c(f) = 0 \notin c$. Then by the separating hyperplane theorem (see for example Corollary 2.4 in Chapter 3 of [21]) there is a linear functional $h\colon \mathbb{R}^n \to \mathbb{R}$ and $\varepsilon > 0$ such that $h(x) > \varepsilon$ for all $x \in c$. By (i), we have

$$\phi(h \circ f) = h(\alpha_c(f)) = 0$$

But $h \circ f > \bar{\varepsilon}$, and so, since ϕ is order-preserving and weakly averaging, $\phi(h \circ f) \geq \varepsilon$. This is a contradiction, completing the proof of (ii).

From (ii), we have maps $\alpha_c\colon \mathbf{Meas}(\Omega, c) \to c$, all that remains is to check that they commute with all affine maps $c \to c'$. As usual, since c' is a subset of a power of $\mathbb{R}$, it is sufficient to check commutativity of all diagrams

$$\begin{tikzcd} \mathbf{Meas}(\Omega, c) \arrow[r, "\alpha_c"] \arrow[d, "h_*"'] & c \arrow[d, "h"] \\ \mathbf{Meas}_b(\Omega, \mathbb{R}) \arrow[r, "\phi"'] & \mathbb{R} \end{tikzcd}$$

for affine $h\colon c \to \mathbb{R}$, where $c \subseteq \mathbb{R}^n$ is compact and convex. By Lemma 5.2, h has an affine extension $\mathrm{aff}(c) \to \mathbb{R}$ which we shall also write as h; we will extend this to an affine map $\mathbb{R}^n \to \mathbb{R}$ as follows. Choose $x_0 \in \mathrm{aff}(c)$, and write $L = \{x \in \mathbb{R}^n \mid x + x_0 \in \mathrm{aff}(c)\}$. Then L is a *linear* subspace of $\mathbb{R}^n$, and the map $l\colon L \to \mathbb{R}$ defined by

$$l(x) = h(x + x_0) - h(x_0)$$

is linear, so l has a linear extension $l'\colon \mathbb{R}^n \to \mathbb{R}$. Let

$$h'(x) = l'(x - x_0) + h(x_0)$$

for $x \in \mathbb{R}^n$; then h' is the desired affine extension of h. The result follows by the same argument as in (i) above, with h' in place of h. □

References

[1] H. Appelgate, M. Tierney, Categories with models, in: B. Eckmann, M. Tierney (Eds.), Seminar on Triples and Categorical Homology Theory, in: Lecture Notes in Mathematics, vol. 80, Springer-Verlag, 1969, pp. 156–224. Reprinted in [2].
[2] H. Appelgate, M. Tierney, Categories with models, Repr. Theory Appl. Categ. 18 (2008) 1–303.
[3] K.P.S. Bhaskara Rao, M. Bhaskara Rao, Theory of Charges: A Study of Finitely Additive Measures, Academic Press, 1983.
[4] B. Bollobás, Linear Analysis, 2nd edition, Cambridge University Press, 1999.
[5] E.-E. Doberkat, Eilenberg–Moore algebras of stochastic relations, Inf. Comput. 204 (12) (2006) 1756–1781.
[6] T. Fritz, Convex spaces I: definition and examples, arXiv:0903.5522, 2009.
[7] M. Giry, A categorical approach to probability theory, in: Categorical Aspects of Topology and Analysis, in: Lecture Notes in Mathematics, vol. 915, Springer, 1982.
[8] S.P. Gudder, A general theory of convexity, Rend. Semin. Mat. Fis. Milano 49 (1) (1979) 89–96.

[9] P.R. Halmos, Measure Theory, D. van Nostrand Company, 1950.
[10] B. Jacobs, Probabilities, distribution monads and convex categories, Theor. Comput. Sci. 412 (28) (2011) 3323–3336.
[11] C. Jones, G. Plotkin, A probabilistic powerdomain of evaluations, in: Proceedings of the Fourth Annual Symposium on Logic in Computer Science, Jun 1989, pp. 186–195.
[12] S. Kakutani, Concrete representation of abstract (m)-spaces (a characterization of the space of continuous functions), Ann. Math. 42 (4) (1941) 994–1024.
[13] J.F. Kennison, D. Gildenhuys, Equational completion, model induced triples and pro-objects, J. Pure Appl. Algebra 1 (4) (1971) 317–346.
[14] A. Kock, Continuous Yoneda representation of a small category, Aarhus University, preprint, 1966.
[15] A. Kock, Synthetic Differential Geometry, London Mathematical Society Lecture Note Series, vol. 333, Cambridge University Press, 2006.
[16] A. Kock, Commutative monads as a theory of distributions, Theory Appl. Categ. 26 (4) (2012) 97–131.
[17] T. Leinster, Codensity and the ultrafilter monad, Theory Appl. Categ. 28 (13) (2013) 332–370.
[18] R.B.B. Lucyshyn-Wright, Algebraic theory of vector-valued integration, Adv. Math. 230 (2012) 552–576.
[19] R.B.B. Lucyshyn-Wright, Riesz–Schwartz extensive quantities and vector-valued integration in closed categories, PhD thesis, York University, Canada, 2013.
[20] S. Mac Lane, Categories for the Working Mathematician, Graduate Texts in Mathematics, vol. 5, Springer, 1971.
[21] O.L. Mangasarian, Nonlinear Programming, Classics in Applied Mathematics, vol. 10, Society for Industrial and Applied Mathematics, 1994.
[22] P. Panangaden, The category of Markov kernels, Electron. Notes Theor. Comput. Sci. 22 (1999) 171–187.
[23] W. Rudin, Priciples of Mathematical Analysis, McGraw-Hill, 1953.
[24] M.H. Stone, Postulates for the barycentric calculus, Ann. Mat. Pura Appl. 29 (1) (1949) 25–30.
[25] K. Sturtz, The Giry monad as a codensity monad, arXiv:1406.6030v2, 2014, updated in [26].
[26] K. Sturtz, Categorical probability theory, arXiv:1406.6030, 2015.