Probability Monads

A motivated treatment of monads and their algebras

Julian Asilis
Advised by Professor Michael Hopkins

A thesis presented in partial fulfillment
of the requirements for the degree of
Bachelor of Arts in Mathematics with Honors.

Harvard University
Cambridge, Massachusetts
March 2020

Acknowledgements

First, I would like to thank Professor Michael Hopkins for serving as a patient and generous advisor, and for suggesting the topic of this thesis. I am truly fortunate to have received such guidance.

In roughly chronological order, I would also like to thank: Eric Peterson for introducing me to the world of pure math (and inspiring me to become math major); James Tao and Vaughan McDonald for their assistance in my most challenging semester at Harvard and the time since, and; Paolo Perrone for generously offering his time to explain probability monads.

I would also like to thank Lobster Fiesta and Lowell's FC for their friendship and support over (nearly) 4 years – it is difficult to imagine being surrounded by a better group of people. Caleb Miller deserves special mention for proofreading this work and for 8 semesters of often asymmetric problem set collaboration.

Lastly, I thank my family for their enduring support, particularly as I unexpectedly complete this at home rather than on campus. Somehow, being stripped of senior spring is a great loss and arriving home early a comfort.

Contents

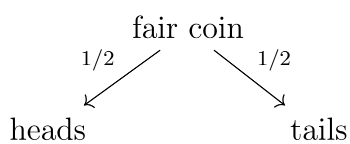
Introduction

This work aims to provide a motivated introduction to the categorical concepts of monads and their algebras, with sights on studying probability monads after outlining the basic categorical theory. The aspiration is that such a target can stimulate what may otherwise be viewed as an uninspired theory – perhaps more importantly, probability monads are themselves compelling and worthy of study.

Motivation and historical context

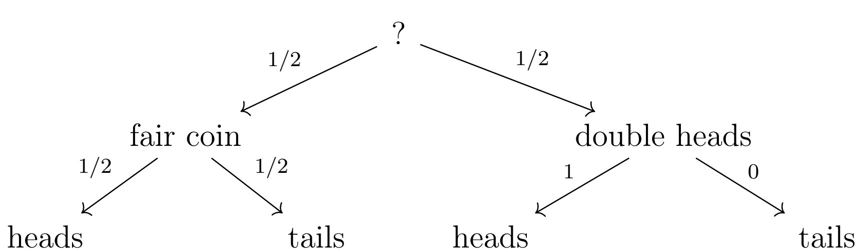
The essential object in probability is that of a probability measure P on a space Ω. For the moment, let us denote the collection of probability measures over Ω by $\Pi(\Omega)$, so that $P \in \Pi(\Omega)$. The importance and ubiquity of probability measures naturally gives rise to the notion of probability measures over probability measures. Such objects admit interpretation as random variables whose laws are themselves random, as are often encountered in Bayesian settings, and form elements of $\Pi(\Pi(\Omega)) = \Pi^2(\Omega)$ in our above notation.

Consider the simplest non-degenerate case $\Omega = \{\text{heads}, \text{tails}\}$ [9]. A fair coin corresponds to a probability distribution P_{fair} over Ω, with $P_{\text{fair}}(\text{heads}) = P_{\text{fair}}(\text{tails}) = 0.5$. One might draw:



(0.0.1)

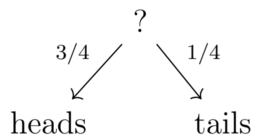
Suppose one also has a double-sided heads coin, likewise determining an element of $\Pi(\Omega)$ via the measure P_{heads} with $P_{\text{heads}}(\text{heads}) = 1, P_{\text{heads}}(\text{tails}) = 0$. What if one were to draw uniformly randomly from the coins, say from a pocket? The setup is depicted as follows:



(0.0.2)

As each coin determines a distribution over Ω, a random law for selecting coins determines a distribution over distributions, i.e. an element of $\Pi^2(\Omega)$. Abstractly, the picture must admit compression – there exists a true probability p of this procedure resulting in a flip of heads, so it is equivalent to the tossing of a single coin which lands heads with probability p.

More constructively, the above tree can be reduced to one of depth 1, an element of $\Pi(X)$, by multiplying through edge weights. A distribution with weight h on heads which itself has weight w should contribute wh to the ultimate probability of heads, a statement much like the law of total probability. Ultimately the tree compresses to:



(0.0.3)

More generally, we have described a map $E : \Pi^2(\Omega) \to \Pi(\Omega)$ which averages distributions over distributions to ordinary distributions. Though its use may not be evident at the moment, there is also a natural map $\Omega \to \Pi(\Omega)$ which builds a measure from a point ω by concentrating all mass at ω, i.e. by outputting the Dirac measure δ_ω at ω.

$$\delta_\omega(A) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

The categorical tool that treats such recursive constructions which, crucially, are equipped with maps for ascending and descending levels (e.g. averaging to descend and Dirac measure to ascend) is that of the *monad*. Outside of category theory, monads have found use in such areas as computer science and probability. Those which carry the interpretation of assigning to a space X a collection of probability measures over X have come to be known as *probability monads*, and form a cornerstone of the field of categorical probability.

Categorical probability has emerged as the effort to apply categorical techniques to the study of probability, measure theory more generally, and mathematical statistics [9]. Though currently less developed than the applications of category theory to such fields as algebraic geometry and topology, categorical probability has witnessed increasing attention in recent decades. In particular, the seeming incongruence between the analytical and categorical viewpoints – which may explain the relative infancy of categorical probability – is precisely a source of inspiration for the subject, the hope being that the translation of a problem from one field to the other can give rise to a useful change in perspective.

The roots of categorical probability can be traced to Lawvere's 1962 seminar notes in which, among other things, he considers categories whose morphisms $\Omega \to \Omega'$ act by assigning distributions over Ω' to elements of Ω [7]. Formally, however, probability monads entered the picture in Giry's seminal 1982 paper, by way of what later become known as the *Giry monad* [3].

Though we do not assume familiariaty with monads – in fact, we assume unfamiliarity – at a high level, the Giry monad and related probability monads take the following form:

- An assignment T sending a space X to a collection of probability measures over X. T is defined on a collection of spaces – be them mere sets, topological spaces, metric spaces obeying some conditions, or other – and is an endomorphism on this collection. For instance, if X is a topological space, then TX need admit a topology as well.

- An assignment T on maps so that a map of spaces $f : X \to Y$ extend to a map of spaces of probability measures $Tf : TX \to TY$.[1] This should respect identities and composition. Usually, this takes the form of a pushforward, i.e. if $P \in TX$ is a probability measure over X, then $Tf \in TY$ is the measure on Y with
$$Tf(A) = P(f^{-1}(A))$$
Note the importance of pre-images respecting disjoint unions in ensuring that Tf indeed be a measure.

- An evaluation map μ for passing from probability measures over probability measures to mere probability measures. In other words, a map $\mu : T(T(-)) \to T(-)$. Usually μ acts by averaging or integrating, as in the coins of (0.0.2).

- Lastly, a map $\eta : (-) \to T(-)$ for passing from a space X to probability measures over X, usually by sending a point x to the Dirac measure over x.

One may note that, having been defined on objects and morphisms in a manner which respects identities and compositions, T appears to be a functor. Similarly, μ resembles a natural transformation $T^2 \Rightarrow T$ and η a natural transformation $1 \Rightarrow T$. Indeed, we will see in 1.2 that the data of a monad consists of an endofunctor T equipped with such natural transformations. All in due course.

Background and notation

Knowledge of category theory at the level of functors and natural transformations, along with basic probability theory, suffices to understand the heart of the work. Familiarity with adjoint functors, presented briefly in Chapter 3, is desirable but not necessary. As is often the case in categorical texts, acquaintance with the basics of algebra (monoids, groups, modules) and topology allows one to enjoy the examples presented but is by no means vital.

With respect to notation, the usual $f : X \to Y$ or $X \xrightarrow{f} Y$ is used to denote a morphism with domain X and codomain Y. When X and Y are understood to reside in a category $\mathcal{C}$, we may write $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$. As a special case, $\alpha : F \Rightarrow G$ denotes

[1]Though technically an abuse of notation, we will come to see that the roles played by T dovetail naturally.

a natural transformation from F to G. Sans Serif is used for categories, as in

Set	sets and functions
Top	topological spaces and continuous maps
Mod_R	(left) R-modules and module homomorphisms
$\mathsf{Vect}_{\Bbbk}$	Mod_R for $R = \Bbbk$ a field
Grp	groups and group homomorphisms
Ab	abelian groups and group homomorphisms, or $\mathsf{Mod}_{\mathbb{Z}}$
$\mathsf{End}(\mathcal{C})$	endofunctors on $\mathcal{C}$ and natural transformations

A brief note on style: we take the view that listing 10 examples which leave all detail to the reader is usually worse than providing a completed example along with a few of the less detailed kind. More generally, we err on the side of the explicit, and strive to accompany each new idea with at least one fully specified example. We take no offense from those who would prefer to skip the more gory of details, but believe nevertheless that it is usually worth doing once.

- In Chapter 1 we motivate the search for the monad, define it, and provide a source of intuition.
- In Chapter 2 we consider algebras over monads and introduce the Eilenberg-Moore and Kleisli categories associated to a monad.
- In Chapter 3 we study how adjunctions induce monads and all monads can be shown to derive from adjunctions, drawing from [10].
- In Chapter 4 we study two probability monads: the Giry and Kantorovich monads.

1. Monads

Two roads lead to the promised land of monads: the first is short, stating the definition with little context and hoping to pick up intuition with example, while muttering something or other about monoids. The second is not as short, making out the landscape by proceeding from classical monoids to monoidal categories to monoid objects and, finally, monads.

We believe that the second path is worth the trouble, and take an unhurried approach to formally arriving at the monad. The hope is that it will appear a natural conclusion to the search for generalization which began with the monoid.

- In 1.1, we innocently generalize the classical monoid of Set and diagnose our shortfalls.
- In 1.2, we introduce the theory of tensor categories, which allows finally for a formal definition of the monad.
- In 1.3, we provide intuition for monads as systems of extending spaces to include generalized elements, and revisit examples from the previous section in a new light.

1.1. As generalizations of monoids

An archetypal algebraic structure is that of the monoid, describing the structure enjoyed by sets equipped with an associative and unital binary operation. Formally, one encounters the following definition in an introductory algebra text:

Definition 1.1. A *monoid* consists of a set M equipped with a *multiplication* $\mu : M \times M \to M$ which admits a *unit* $e \in M$ such that for all $a, b, c \in M$

(i) $\mu(a, \mu(b, c)) = \mu(\mu(a, b), c)$

(ii) $\mu(a, e) = \mu(e, a) = a$.

This uncomplicated understanding of monoids gives rise to a plurality of remarkably rich theories – particularly those whose objects are, or contain, particularly well-behaved monoids, such as groups, rings, and modules – but it clings tightly to its particular context, that of the universe Set.

In particular, how can a topologist – differential or not – come to witness monoids in their field of study, in which mere functions play no role? Perhaps the path to generalization in this case is clear – one should demand that the monoid's multiplication be continuous rather than a meager map of sets – but the ad hoc solution postpones the problem of generalizing and, more importantly, fails in the many abstract settings in which objects and their morphisms do not naturally resemble sets and functions with additional structure.

The first step in the effort to translate the theory of monoids to arbitrary categories, then, is to express a monoid's data and properties using categorical language; that is, via commutative diagrams. Let us examine each condition in turn:

- condition (i) in Definition 1.1 can be seen as corresponding to an equality between maps from M^3 to M which factor through M^2. One first multiplies the leftmost pair of elements in a triple and then multiplies the remaining two elements, while the other begins by multiplying the rightmost pair of elements in a triple. In symbols, with $\mu : M \times M \to M$ an arbitrary map of sets as of yet and 1_M the identity on M, this amounts to:

$$\begin{array}{ccc} M \times M \times M & \xrightarrow{1_M \times \mu} & M \times M \\ {\scriptstyle \mu \times 1_M}\downarrow & & \downarrow{\scriptstyle \mu} \\ M \times M & \xrightarrow{\mu} & M \end{array} \tag{1.1.1}$$

- condition (ii) corresponds to an equality between endomorphisms on M which factor through M^2. The first is defined by adjoining the identity to an element on its left hand side and subsequently multiplying, while the second adjoins the identity on the right hand side. In symbols, and letting $\eta : 1 = \{*\} \to M$ denote the map which picks out the identity $e \in M$, we have:

$$\begin{array}{ccccc} M & \xrightarrow{\eta \times 1_M} & M \times M & \xleftarrow{1_M \times \eta} & M \\ & {\scriptstyle 1_M}\searrow & \downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle 1_M} & \\ & & M & & \end{array} \tag{1.1.2}$$

The fruit of such efforts is the ability to transport the preceding two diagrams to an arbitrary category, thereby demanding that μ and η obey the defining monoidal properties while being appropriate morphisms in the chosen ambient category. In this manner, one (tentatively) generalizes the monoid beyond its birthplace of Set. Note that the generalization appended an additional datum to the definition – the element $e \in M$, formerly a property of the monoid's multiplication, has been substituted for the more general notion of a map $\eta : 1 \to M$. Our aspiration is that at the conclusion of our efforts, it will not be sensible to speak of an 'element' of an arbitrary monoid.

Instantiating the definition in Set should recover the classic monoid of Definition 1.1. Indeed it does: a monoid satisfies (1.1.1) as a consequence of its associativity and satisfies (1.1.2) with $\eta : 1 \to M$ the map which picks out the identity. Conversely, any binary operation which obeys (1.1.1) thereby associates, and which furthermore obeys (1.1.2) has $\eta(*)$ as its identity. By identical reasoning, the generalized monoids in any concrete category are monoids in the classical sense (though the converse need not, and in fact should not, hold).

Instantiating the definition in Top confirms the previous intuition that one need only demand that μ be continuous. In particular, the map η which picks out the unit is automatically continuous in the way all functions out of the singleton are - the only topology admitted by the point is the discrete one. Accordingly, the new theory defines a topological monoid to be a space equipped with the structure of a classical monoid whose multiplication is continuous.

Let us move the picture to an algebraic setting, such as Mod_R. The notion of a monoid structure on a module M is then determined by a multiplication morphism out of the product $M \times M$. This is a somewhat unnatural restriction - the direct product of modules often takes a back seat to the tensor product, which is less restrictive in allowing maps of sets to be homomorphisms.

In fact, the demand that monoids in arbitrary categories have multiplication defined out of the categorical product is a groundless restriction. Tensor products of modules make that much clear. Rather than preserving a peculiarity of Set – the role of products – it would be more appropriate to utilize a general theory of suitably well-behaved binary products between objects in a category. Based in part on the tensor products of Mod_R, it is to this theory of *tensor categories* to which we now turn. This path to generalizing monoids, with slightly more care, will take us to monads as a special case.

1.2. As monoid objects in tensor categories

The theory of tensor categories arises as the study of a more general binary operation on the objects of a category than the categorical product. Informally, it concerns categories whose objects are 'almost' endowed with the structure of a monoid. That the theory centers around near-misses to classical monoids rather than (only) exact matches is a consequence of the additional structure inherited by passing from the elements of a set to the objects of a category.

In particular, elements of a set are either equal or distinct, whereas objects of a category enjoy the notion of proximity or 'essential sameness' witnessed by isomorphisms. As a result, it would be inappropriately restrictive to demand that the binary operation on objects have a unit or associate in the strict sense of Definition 1.1. In the following definition, such properties are permitted to hold up only up to isomorphism. The price one pays for generality, however, is the logical complexity of disciplining such near-misses, witnessed, for instance, in the associahedron of (1.2.2).[2]

Definition 1.2. A *tensor category*, or *monoidal category*, is a category $\mathcal{C}$ equipped with a bifunctor $\otimes : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$, known as the *tensor product* or *monoidal product*, a *unit* object $I \in \mathcal{C}$, and the following natural isomorphisms

(i) A natural isomorphism $\alpha : (-\otimes-)\otimes- \Rightarrow -\otimes(-\otimes-)$, the *associator*, which is natural in its three arguments for all $a, b, c \in \mathcal{C}$, i.e. $\alpha_{a,b,c} : (a\otimes b)\otimes c \to a\otimes(b\otimes c)$

[2]This complexity is effectively lethal to the direct study of weak n-categories for finite $n \geq 4$, in which near-misses are defined in terms of 'higher' near-misses. More explicitly, 1-morphisms associate up to 2-isomorphisms, where 2-morphisms likewise associate up to 3-morphisms, and so on. n-morphisms, meanwhile, associate 'on the nose'; there is nowhere else to go! To witness the effects of the combinatorial explosion, see [14] for a 51-page definition of a tetracategory, or weak 4-category.

(ii) Two natural isomorphisms $\lambda : I \otimes - \Rightarrow 1_{\mathcal{C}}$ and $\rho : - \otimes I \Rightarrow 1_{\mathcal{C}}$, called the *left* and *right unitors*, respectively.

Subject to the following coherence conditions for all $a, b, c, d \in \mathcal{C}$

(a) Commutativity of the triangle

$$\begin{array}{ccc} & a \otimes b & \\ {\scriptstyle \rho_a \otimes 1_b} \nearrow & & \searrow {\scriptstyle 1_a \otimes \lambda_b} \\ (a \otimes I) \otimes b & \xrightarrow{\alpha_{a,I,b}} & a \otimes (I \otimes b) \end{array} \tag{1.2.1}$$

(b) The pentagonal associahedron

$$\begin{array}{ccc} & (a \otimes (b \otimes c)) \otimes d & \\ {\scriptstyle \alpha_{a,b,c} \otimes 1_d} \swarrow & & \searrow {\scriptstyle \alpha_{a,b \otimes c,d}} \\ ((a \otimes b) \otimes c) \otimes d & & a \otimes ((b \otimes c) \otimes d) \\ {\scriptstyle \alpha_{a \otimes b,c,d}} \downarrow & & \downarrow {\scriptstyle 1_a \otimes \alpha_{b,c,d}} \\ (a \otimes b) \otimes (c \otimes d) & \xrightarrow{\alpha_{a,b,c \otimes d}} & a \otimes (b \otimes (c \otimes d)) \end{array} \tag{1.2.2}$$

The picture has been made significantly less compact by weakening the conditions on a monoidal category's tensor product, relative to the classical monoid. In particular, additional data in the form of an associator and two unitors was introduced, as well as the property of explicitly demanding that 4-ary tensor products be defined up to unique associator isomorphisms (tensored with identities, if necessary). The mere presence of an associator isomorphism fails to ensure associativity of higher products owing to the presence of several chains of isomorphisms connecting distinct parenthesizations, which a priori need not coincide. Diagram (1.2.2) displays distinct paths between each parenthesization – defined by traveling opposite sides of the pentagon – and is mandatory in ensuring that distinct identifications between a pair of parenthesizations coincide.

It is not at all obvious that the explicit handling of the 4-ary case suffices to ensure that higher products be defined up to a unique chain of associator isomorphisms. That it does is a result of MacLane, and a central theorem of monoidal categories [8].

Theorem 1.3 (MacLane coherence)**.** *Suppose a monoidal category $\mathcal{C}$ with tensor product $\otimes$ and associator α satisfies associahedron (1.2.2). Then for any functors $F', F : \mathcal{C}^n \to \mathcal{C}$ given by repeated application of $\otimes$, there exists a unique composition of products of associators, their inverses, and identities which serves as a natural isomorphism $F \cong F'$.*

Example 1.4. For $\mathcal{C}$ a category, the category $\mathsf{End}(\mathcal{C})$ of endofunctors and their natural transformations forms a monoidal category. The tensor product is given by composition and the unit object by $1_{\mathcal{C}}$. Its associator and unitors are identities.

Example 1.5. For R a commutative ring, the category Mod_R of R-modules is a monoidal category with $\otimes = \otimes_R$ and $I = R$.

Example 1.6. Any category with finite products is a monoidal category with unit object the terminal object, termed a *cartesian monoidal category.* Dually, categories with finite coproducts are monoidal with unit object the initial object, termed *cocartesian monoidal categories.*

Example 1.4 is a first indication of the ubiquity of monoidal categories, and will serve as the backdrop for the theory of monads. It is also the first example of a strict monoidal category, in which the tensor product associates and has a unit object 'on the nose', rather than up to isomorphism.

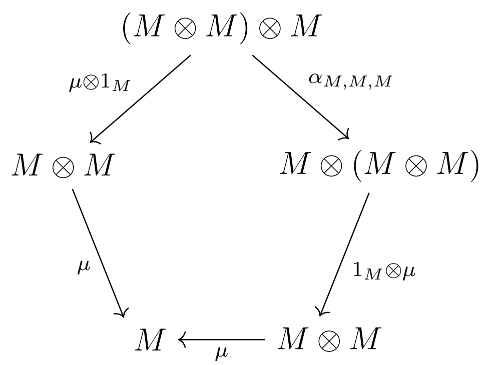
Definition 1.7. A monoidal category whose associator and unitor natural isomorphisms are identities is a *strict* monoidal category.

Remark 1.8. Some authors speak of a dictionary between categorical constructions and classical algebraic objects, defined by passing from a category to isomorphism classes of its objects [2]. It is in this manner that, for instance, a category is a *categorification* of a set. As as consequence of the associators and unitors in Definition 1.2, this dictionary associates monoids to monoidal categories. Note that in strict monoidal categories, one need not pass to isomorphism classes of objects. Abelian categories, which lie beyond the scope of our discussion, correspond to abelian groups in precisely this manner (along with some massaging needed to introduce 'virtual' inverses).

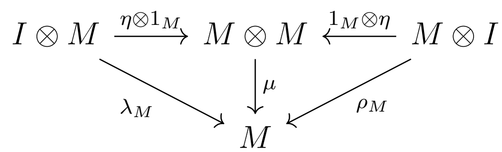
We are now equipped to define the true generalization of a monoid to an arbitrary tensor category.[3] A key observation is that the monoidal structure on the ambient category influences the data of the monoid. In particular, the multiplication μ ought to be out of the tensor product – it is for this reason that we turned to the more general products – and the unit map out of the unit object $\eta : I \to M$. Just as categorical structure permitted laxness with respect to the properties of the tensor product, we should likewise only require that μ 'essentially' associate and be unital.

Definition 1.9. A *monoid*, or *monoid object*, in a monoidal category $\mathcal{C}$ consists of an object M along with a *multiplication* map $\mu : M \otimes M \to M$ and a *unit* $\eta : I \to M$ such that the following diagrams commute.

[3]By Example 1.4, this is more ambitious than our first generalization to categories with finite products.



(1.2.3)



(1.2.4)

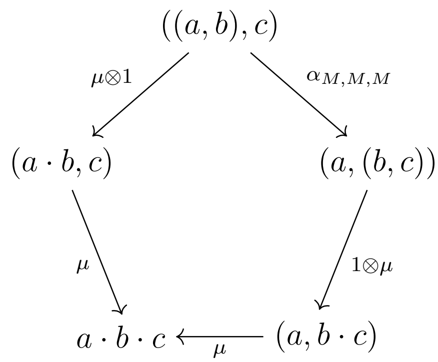
As in Definition 1.2, the pentagon encodes the associativity condition, demanding that multiplication in three entries be defined up to the associator between parenthesizations of $M^{\otimes 3}$. Note that the commutative square (1.1.1) encoding associativity in the classical case would not have sufficed owing to the lax associativity of the tensor product. In particular, the notion of a 'triple' of inputs[4] does not extend in a well-defined manner to the monoidal category – each parenthesization of $M^{\otimes 3}$ demands its own object.

The demands on the unit, however, are witnessed by maps out of a unique binary product, allowing diagram (1.2.4) to closely resemble its analogue from the classical case. Note the importance of the unit morphism η having been defined as a map out of the unit object I.

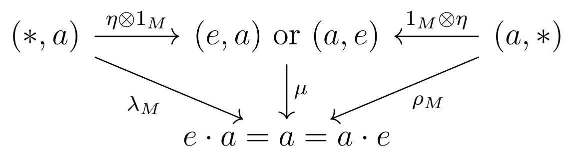
Notation 1.10. The unit object I of a monoidal category and unit $\eta : I \to M$ of a monoid (object) are often each referred to simply as units, creating potential for ambiguity. We reserve the term *unit* for the unit η of a monoid, which will be more central to our study.

In order to be a true generalization, monoid objects should witness classical monoids as a special case. By Example 1.6, Set equipped with the categorical product is a monoidal category with unit $1 = \{*\}$. It has associator $((a, b), c) \mapsto (a, (b, c))$, left unitor $(*, a) \to a$, and right unitor $(a, *) \to a$. To see that a classical monoid is precisely a monoid object in Set, note that (1.2.3) amounts to $(a \cdot b) \cdot c = a \cdot (b \cdot c)$ and (1.2.4) to $e \cdot a = a = a \cdot e$. More explicitly,

[4]Of course, the theory of generalized monoids being developed extends far beyond the concrete cases examined in 1.1, in that multiplication is a morphism which need not resemble a function. The multiplication (and unit) of monads will be of this more abstract type.



(1.2.5)



(1.2.6)

A consequence of the above is that monoid objects in concrete categories are classical monoids.

Notation 1.11. Hereafter, we use *monoid* in the sense of Definition 1.9 – diagrams (1.2.5) and (1.2.6) have earned us the right.

For the last time, we have confirmed the intuition that a monoid in Top is exactly a classical monoid with continuous multiplication.[5] To see intuition fail, consider a monoid object M in the cartesian monoidal category Mon of classical monoids and their homomorphisms. As an object of Mon, M is equipped with its own multiplication, denoted $a \cdot b$, which its product inherits coordinate-wise. By virtue of moreover being a monoid object in the sense of 1.9, it has a multiplication $\mu : M \times M \to M$, which we denote $a \times b$. As μ is a morphism in the category of monoids, it follows that

$$\begin{aligned} \mu(a, b) \cdot \mu(c, d) &= \mu((a, b) \cdot (c, d)) \\ (a \times b) \cdot (c \times d) &= \mu(a \cdot c, b \cdot d) \\ (a \times b) \cdot (c \times d) &= (a \cdot c) \times (b \cdot d) \end{aligned}$$

M's multiplication - as an element of Mon - is unital by definition, while its multiplication - as a monoid object - is unital under (1.2.6). Then by Eckmann-Hilton, the multiplications coincide and furthermore are commutative. Thus, a monoid object in Mon is a commutative monoid, and any commutative monoid becomes a monoid object by 'reusing' its multiplication.

Remark 1.12. On the subject of classical monoids, a source of monoidal categories arises as a converse to Remark 1.8. Just as a category being a categorification of a set allows for one to travel from categories to sets (by considering isomorphism classes of objects), one can pass from a set to a category by constructing the discrete category

[5]Recall that functions out of the terminal space are automatically continuous.

on a set. Similarly, one can pass from a classical monoid $(M, \cdot)$ to a (strict) monoidal category by constructing the discrete category with $a \otimes b := a \cdot b$.

Example 1.13. By Example 1.5, the category Ab of abelian groups, or $\mathbb{Z}$-modules, is a monoidal category when equipped with the tensor product over $\mathbb{Z}$. A ring R is an abelian group equipped with $\mu : R \otimes_{\mathbb{Z}} R \to R$ and $\eta : \mathbb{Z} \to R$ so that the following commute.

$$\begin{array}{ccc} R \otimes_{\mathbb{Z}} R \otimes_{\mathbb{Z}} R & \xrightarrow{1_R \otimes_{\mathbb{Z}} \mu} & R \otimes_{\mathbb{Z}} R \\ {\scriptstyle \mu \otimes_{\mathbb{Z}} 1_R}\downarrow & & \downarrow{\scriptstyle \mu} \\ R \otimes_{\mathbb{Z}} R & \xrightarrow{\mu} & R \end{array} \qquad \begin{array}{ccccc} R & \xrightarrow{\eta \otimes_{\mathbb{Z}} 1_R} & R \otimes_{\mathbb{Z}} R & \xleftarrow{1_R \otimes_{\mathbb{Z}} \eta} & R \\ & {\scriptstyle 1_R}\searrow & \downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle 1_R} & \\ & & R & & \end{array}$$

Thus, rings are monoids in Ab. Note that the pentagon of (1.2.3) reduces to a square because Ab is a strict monoidal category. Likewise, a $\mathbb{k}$-algebra is an object $R \in \mathsf{Vect}_{\mathbb{k}}$ such that the above diagrams commute, with tensor products taken over $\mathbb{k}$. Consequently, $\mathbb{k}$-algebras are monoids in $\mathsf{Vect}_{\mathbb{k}}$.

The stage has been set for a compact, if abstruse, definition of a monad.

Definition 1.14. A *monad* on $\mathcal{C}$ is a monoid in $\mathsf{End}(\mathcal{C})$.

Let us be slightly more explicit: the morphisms of $\mathsf{End}(\mathcal{C})$ are natural transformations and its tensor product is composition of functors, so that the data of a monad T consists of natural transformations $\mu : T^2 \Rightarrow T$ and $\eta : 1_{\mathcal{C}} \Rightarrow T$. The properties μ and η are subject to are simplified by $\mathsf{End}(\mathcal{C})$ being a strict monoidal category, as functors associate and have a unit 'on the nose'. However, the action of $\mathsf{End}(\mathcal{C})$'s tensor product on morphisms is slightly convoluted. Let $(\alpha, \beta) : A \times B \to A' \times B'$ be a morphism in the product category $\mathsf{End}(\mathcal{C}) \times \mathsf{End}(\mathcal{C})$, meaning $\alpha : A \Rightarrow A'$ and $\beta : B \Rightarrow B'$ are natural transformations. In order to construct the components of a transformation $AB \Rightarrow A'B$ from α and β, one must pass through AB'. In particular, for $c, d \in \mathcal{C}$ and $f : c \to d$ one has

$$\begin{array}{ccc} ABc & \xrightarrow{AB(f)} & ABd \\ {\scriptstyle A(\beta_c)}\downarrow & & \downarrow{\scriptstyle A(\beta_d)} \\ AB'c & \xrightarrow{AB'(f)} & AB'd \\ {\scriptstyle \alpha_{B'c}}\downarrow & & \downarrow{\scriptstyle \alpha_{B'd}} \\ A'B'c & \xrightarrow{A'B'(f)} & A'B'(d) \end{array} \tag{1.2.7}$$

The uppermost square commutes before application of A by naturality of β, so it commutes by functoriality of A. The bottommost square commutes by naturality of α, and we conclude that $c \mapsto \alpha_{B'c} \circ A(\beta_c)$ defines a natural transformation $AB \Rightarrow A'B'$.

Fortunately, we need only consider the images of $(1_T, \mu) : (T, T^2) \to (T, T)$ and $(\mu, 1_T) : (T^2, T) \to (T, T)$ in order to understand the constraints on the multiplication of a monad. By the preceding, the components of the image of $(1_T, \mu)$ take the form

$$c \mapsto 1_{T(Tc)} \circ T(\mu_c) = T(\mu_c) \tag{1.2.8}$$

while the components of the image of $(\mu, 1_T)$ take the form

$$c \mapsto \mu_{Tc} \circ T^2(1_{Tc}) = \mu_{Tc} \tag{1.2.9}$$

Thus, the pentagon (1.2.3) encoding the associativity of a monad reduces to a square like that in (1.1.1) – there's only one T^3 – of the form:

$$\begin{array}{ccc} T^3 & \overset{T\mu}{\Longrightarrow} & T^2 \\ {\scriptstyle \mu T}\Downarrow & & \Downarrow{\scriptstyle \mu} \\ T^2 & \overset{\mu}{\Longrightarrow} & T \end{array} \tag{1.2.10}$$

with the notation $T\mu$ and μT justified (and made precise) in (1.2.8) and (1.2.9). In words, μ must extend uniquely to a natural transformation $T^3 \Rightarrow T$, either by applying T to the components of μ or applying μ to the image of T.

The condition on the unit $\eta : 1_{\mathcal{C}} \Rightarrow T$ is simpler to make sense of: as the unitors in $\mathsf{End}(\mathcal{C})$ are identities, the upper corners of (1.2.4) can be identified with the monad. One is left

$$\begin{array}{ccccc} T & \overset{\eta T}{\Longrightarrow} & T^2 & \overset{T\eta}{\Longleftarrow} & T \\ & {\scriptstyle 1_T}\searrow & \Downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle 1_T} & \\ & & T & & \end{array} \quad = \quad \begin{array}{ccc} T & \overset{\eta T}{\Longrightarrow} & T^2 \\ {\scriptstyle T\eta}\Downarrow & \searrow & \Downarrow{\scriptstyle \mu} \\ T^2 & \overset{\mu}{\Longrightarrow} & T \end{array} \tag{1.2.11}$$

We are now justified in presenting an equivalent, sometimes more popular, definition for the monad.

Definition 1.15. A *monad* on a category $\mathcal{C}$ is an endofunctor T on $\mathcal{C}$ equipped with natural transformations $\mu : T^2 \Rightarrow T$ and $\eta : 1_{\mathcal{C}} \Rightarrow T$ such that diagrams (1.2.10) and (1.2.11) commute.

The advantage of Definition 1.15 is accessibility, having avoided monoidal categories, which bears the cost of opacity, having obscured the origins of the relevant diagrams. With the basic theory of monoidal categories behind us, we make use of the more elementary definition when verifying monads – which has the benefit of having flushed out the consequences of $\mathsf{End}(\mathcal{C})$ being strictly monoidal and the particularities of the action of its tensor product on morphisms – while always keeping the humble beginnings of classical monoids in mind.

Example 1.16. The power set functor $\mathcal{P} \in \mathsf{End}(\mathsf{Set})$ is a monad with unit sending an element to the singleton containing it $\eta : X \to \mathcal{P}X, x \mapsto \{x\}$ and multiplication which takes the union across a subset of subsets of X to form a subset of X, i.e. $\mu : \mathcal{P}^2X \to \mathcal{P}X, S \mapsto \bigcup_{R\in S} \bigcup_{r\in R} r$. η is natural because the image $\mathcal{P}f$ of $f : X \to Y$ acts on subsets by acting on each of their elements by f. In particular, $\mathcal{P}f(\{x\}) = \{f(x)\}$.

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle \eta_X}\downarrow & & \downarrow{\scriptstyle \eta_Y} \\ \mathcal{P}X & \xrightarrow{\mathcal{P}f} & \mathcal{P}Y \end{array} \quad \leftrightsquigarrow \quad \begin{array}{ccc} x & \longmapsto & f(x) \\ \downarrow & & \downarrow \\ \{x\} & \longmapsto & \{f(x)\} \end{array}$$

Likewise for μ; $\mathcal{P}\mathcal{P}f$ applies $\mathcal{P}f$ to each of its elements – subsets of X – and $\mathcal{P}f$ acts element-wise by f. Naturality amounts to the fact that direct images commute with unions.

$$\begin{array}{ccc} \mathcal{P}^2X & \xrightarrow{\mathcal{P}^2f} & \mathcal{P}^2Y \\ {\scriptstyle \mu}\downarrow & & \downarrow{\scriptstyle \mu} \\ \mathcal{P}X & \xrightarrow{\mathcal{P}f} & \mathcal{P}Y \end{array} \quad \leftrightsquigarrow \quad \begin{array}{ccc} \{\{x\},\{y,z\}\} & \longmapsto & \{\{f(x)\},\{f(y),f(z)\}\} \\ \downarrow & & \downarrow \\ \{x,y,z\} & \longmapsto & \{f(x),f(y),f(z)\} \end{array} \tag{1.2.12}$$

The associativity condition on μ amounts to the claim that an element of $\mathcal{P}^3X$ can be assembled into an element of $\mathcal{P}X$ by starting with either 'layer' of subsets.

$$\begin{array}{ccc} \mathcal{P}^3 & \xRightarrow{\mathcal{P}\mu} & \mathcal{P}^2 \\ {\scriptstyle \mu\mathcal{P}}\Downarrow & & \Downarrow{\scriptstyle \mu} \\ \mathcal{P}^2 & \xRightarrow{\mu} & \mathcal{P} \end{array} \tag{1.2.13}$$

For instance, $S = \{\{\{a,b\},\{b,c\}\},\{\{d,e\}\}\} \in \mathcal{P}^3X$ sees two roads to $\mathcal{P}X$:

(i) Take the union of each element of S, via $\mathcal{P}\mu$, so as to arrive at $\{\{a,b,c\},\{d,e\}\}$, and union once more to settle at $\{a,b,c,d,e\}$, or

(ii) Take the union across elements of S, via $\mu\mathcal{P}$, so as to arrive at $\{\{a,b\},\{b,c\},\{d,e\}\}$, and union once more to settle at $\{a,b,c,d,e\}$.

That these methods coincide corresponds to commutativity of (1.2.13). Now let $Y \in \mathcal{P}X$; naturality of η results from $\{Y\}$ and $\{\{y\} | y \in Y\}$ both being 'unwrapped' by μ to Y.

$$\begin{array}{ccccc} \mathcal{P} & \xRightarrow{\eta\mathcal{P}} & \mathcal{P}^2 & \xLeftarrow{\mathcal{P}\eta} & \mathcal{P} \\ & {\scriptstyle 1_{\mathcal{P}}}\searrow & \Downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle 1_{\mathcal{P}}} & \\ & & \mathcal{P} & & \end{array} \tag{1.2.14}$$

For instance, the path through μ starting at the top right acts as

$$\{a,b,c\} \longmapsto \{\{a\},\{b\},\{c\}\} \longmapsto \{a,b,c\}$$

and the path through μ starting at the top left acts as

$$\{a,b,c\} \longmapsto \{\{a,b,c\}\} \longmapsto \{a,b,c\}$$

thereby completing the example.

Demonstrating that an object of $\mathsf{End}(\mathcal{C})$ is a monad can be an involved task, worthy of doing at least once. In what follows, we provide detail at a level slightly less gruesome, all the while encouraging the interested – or skeptical – reader to iron out the details.

Example 1.17. The *discrete distribution monad* $\mathcal{D}$ on Set sends a set X to the set of finitely supported functions $\phi : X \to [0,1]$ such that $\sum \phi(x) = 1$, where the sum is taken over the support of ϕ [6]. In words, $\mathcal{D}X$ consists of weights across X which sum to 1 and for which cofinitely many weights are zero.

A useful notational tool is to identify $\phi \in \mathcal{D}X$ with the expression $\sum \phi(x)[x]$. The brackets in $[x]$ serve as syntactic sugar which distinguishes an element of X from its position in the description of $\phi \in \mathcal{D}X$. For instance, when $X = \{a, b, c\}$, the map $\phi : a, b, c \mapsto \frac{1}{3}$ encodes uniform weights on the elements of X, and has the syntactic representation $\frac{1}{3}[a]+\frac{1}{3}[b]+\frac{1}{3}[c]$ as an element of $\mathcal{D}X$. $\mathcal{D}$ acts on morphisms by extending them linearly, so that

$$\mathcal{D}(f)(\sum_i w_i[x_i]) = \sum_i w_i[f(x_i)]$$

for $f : X \to Y$. Note that $\mathrm{supp}(\mathcal{D}(f))$ is finite as a consequence of $\mathrm{supp}(f)$ being finite. Multiplication $\mu : \mathcal{D}^2X \to \mathcal{D}X$ transforms weights on weights on X into weights on X by averaging, while the unit $\eta : X \to \mathcal{D}X$ concentrates full weight on its input, i.e.

$$\eta(x) = 1[x] \qquad \mu(\Omega)(x) = \sum_{\phi \in \mathrm{supp}(\Omega)} \Omega(\phi) \cdot \phi(x)$$

Naturality of μ is a consequence of the fact that relabeling elements and averaging weights commute. Formally, consider the lower left path of the following:

$$\begin{array}{ccc} \mathcal{D}^2X & \xrightarrow{\mathcal{D}^2 f} & \mathcal{D}^2Y \\ {\scriptstyle\mu}\downarrow & & \downarrow{\scriptstyle\mu} \\ \mathcal{D}X & \xrightarrow{\mathcal{D}f} & \mathcal{D}Y \end{array} \tag{1.2.15}$$

$\Omega \in \mathcal{D}^2X$ maps to $\mu(\Omega) = \sum_x w_x[x]$, where $w_x = \sum_\phi \Omega(\phi)\phi(x)$. Passing through $\mathcal{D}f$ yields $\mathcal{D}f(\mu(\Omega)) = \sum_x w_x[f(x)]$. The upper right path sends Ω to $\mathcal{D}^2f(\Omega)$ defined by

$$\mathcal{D}^2f(\Omega) = \mathcal{D}^2f(\sum_\phi \Omega(\phi)[\phi]) = \sum_\phi \Omega(\phi)[\mathcal{D}g(\phi)]$$

Passing through μ yields $\sum_x \sum_\phi \Omega(\phi)\phi(x)[f(x)] = \sum_x w_x[f(x)]$, as desired. Unitarity of η corresponds to the fact that weights on X, say $\phi \in \mathcal{D}X$, can be extended to weights on weights on X and averaged back to weights on X in two ways:

(i) Give ϕ itself weight one, so that $\mathcal{D}\eta(\phi) = 1[\phi]$, and average to $1 \cdot \phi = \phi$, or

(ii) Give each $x \in X$ weight 1 and weight such a distribution with $\phi(x)$, so that $\eta\mathcal{D}(\phi) = \sum_x \phi(x)[1[x]]$. Average to $\sum_x \phi(x)[x] = \phi$.

That these coincide completes the example.

Having established the basic theory, we depart from formalism in the following section in an effort to attach an intuition – or several – to monads, which will pay dividends particularly as we steer a course toward probability monads. For the moment, we can pride ourselves in being on the right side of the joke definition sometimes exchanged between computer scientists [4]:

> "A monad is just a monoid in the category of endofunctors, what's the problem?"

That is, we are on the side of those who don't laugh at all.

1.3. As spaces of generalized elements

An interpretation of monads which can lend intuition is to consider them as procedures for *extending spaces* to include generalized elements and appropriately generalized functions [9]. In particular, the data of a monad T is then seen as consisting of:

- The assignment on objects $X \mapsto TX$, with TX thought of as an extension of X inhabited by *generalized elements*
- For each ordinary map $f : X \to Y$, an extended map $Tf : TX \to TY$ which accepts input of the generalized kind and likewise returns generalized output.

In the case of the discrete distribution monad $\mathcal{D}$, a discrete distribution on a set X can be thought of as a generalized element which draws from several ordinary elements.

The multiplication map $\mu : TTX \to TX$ of a monad *simplifies* or *evaluates* twice generalized elements to once generalized elements. In the case of $\mathcal{D}$, a distribution on distributions can be reduced to an ordinary distribution by averaging, precisely as in the coins of (0.0.3).

- Naturality of evaluation corresponds to the fact that generalized functions commute with evaluation
- The diagrammatic constraint on μ states that three-time generalized elements can be unambiguously evaluated to generalized elements, either by first using generalized evaluation $T\mu$ or first applying usual evaluation to generalized inputs $T\mu$ (and, in each case, applying ordinary evaluation μ afterward).

The unit η associated to a monad serves to witnesses original elements as (usually degenerate) instances of generalized elements. In the case of $\mathcal{D}$, the elements of a set do not carry the structure of a distribution but do embed naturally as deterministic distributions, i.e. as Dirac distributions with full weight on a single element. Moreoever,

- Naturality of η corresponds to the property that Tf must coincide with f on 'original' or 'old' elements:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle\eta}\downarrow & & \downarrow{\scriptstyle\eta} \\ TX & \xrightarrow{Tf} & TY \end{array} \tag{1.3.1}$$

  In the case of $\mathcal{D}$, this amounted to $Tf(1[x]) = 1[f(x)] = \eta(f(x))$.

- The diagram constraining η corresponds to the fact that a generalized element made twice general – usually in a degenerate manner – will evaluate to itself. In $\mathcal{D}$, a distribution transformed into a distribution over distributions – either with a single deterministic 'outer' distribution or several deterministic 'inner' distributions – averages to itself.

An additional interpretation of monads is as a system of extending input to include a class of *formal expressions* and maps between formal expressions. For instance, a mere set S can be extended to FS, the set of formal sums over S with integral coefficients, so that $2s+r \in FX$. Crucially, a sum carries little structure – $2s+r$ cannot be simplified further – but sums of sums admit an evaluation, so that $(2s+r)+3(r+t)$ reduce to $2s+4r+3t$. Additionally, any element $s \in S$ embeds as a degenerate formal expression in FS, of the form $1s$ or simply s. Lastly, a map of sets $S \to R$ lifts to a map of formal expressions $FS \to FR$ by extending linearly, so that

$$Tf(\alpha_1 s_1 + \cdots + \alpha_n s_n) = \alpha_1 f(s_1) + \cdots + \alpha_n f(s_n) \in FR$$

Indeed, we have described the *free abelian group monad* on Set, which sends a set S to the set of finite formal $\mathbb{Z}$-linear combinations over S. This is not unlike the discrete distribution monad $\mathcal{D}$, and the verifications of their monadicity closely resemble each another. A second pass at the details – with a newfound interpretation of monads – is worth the trouble.

- Naturality of the unit $\eta : 1 \Rightarrow T$ ensures that the output of an original element coincide with its output as a formal expression.

$$\begin{array}{ccc} S & \xrightarrow{f} & R \\ \downarrow{\scriptstyle \eta_S} & & \downarrow{\scriptstyle \eta_R} \\ TS & \xrightarrow{Tf} & TR \end{array} \quad \rightsquigarrow \quad \begin{array}{ccc} s & \longmapsto & f(s) \\ \downarrow & & \downarrow \\ 1s & \longmapsto & 1f(s) \end{array}$$

- Naturality of evaluation $\mu : T^2 \Rightarrow T$ – which, informally, collects like terms – corresponds to its commuting with linear functions.

$$\begin{array}{ccc} T^2S & \xrightarrow{T^2f} & T^2R \\ \downarrow{\scriptstyle \mu_S} & & \downarrow{\scriptstyle \mu_R} \\ TS & \xrightarrow{Tf} & TR \end{array}$$

  For instance, $3(2s+r)+1(2r) \in T^2X$ can first have its like terms collected and subsequently have $f : S \to R$ applied to its constituent terms, as in

$$3(2s+r)+1(2r) \longmapsto 6s+5r \longmapsto 6f(s)+5f(r)$$

  or it can first have $f : X \to Y$ applied to the constituent terms of its constituent terms and then have its terms collected, as in

$$3(2s+r)+1(2r) \longmapsto 3(2f(s)+f(r))+1(2f(r)) \longmapsto 6f(s)+5f(r)$$

- Unitarity of η corresponds to the fact that a formal expression turned into a two-layed formal expression – with an outer coefficient of 1 or several inner coefficients of 1 – will evalaute to itself.

$$
\begin{array}{ccc}
T & \xRightarrow{\eta T} & T^2 \\
{\scriptstyle T\eta}\Downarrow & \searrow\!\!\!\!\!= & \Downarrow{\scriptstyle \mu} \\
T^2 & \xRightarrow[\mu]{} & T
\end{array}
\qquad \rightsquigarrow \qquad
\begin{array}{ccc}
2r + 3s & \longmapsto & 1(2r + 3s) \\
\downarrow & & \downarrow \\
2(1r) + 3(1s) & \longmapsto & 2r + 3s
\end{array}
$$

- Finally, associativity of μ expresses the fact that three-layered formal expressions admit a single evaluation to formal expressions, i.e.

$$
\begin{array}{ccc}
T^3 & \xRightarrow{T\mu} & T^2 \\
{\scriptstyle \mu T}\Downarrow & & \Downarrow{\scriptstyle \mu} \\
T^2 & \xRightarrow[\mu]{} & T
\end{array}
\qquad \rightsquigarrow \qquad
\begin{array}{ccc}
((r + s) + (s + t)) & \longmapsto & (r + s) + (s + t) \\
\downarrow & & \downarrow \\
(r + 2s + t) & \longmapsto & r + 2s + t
\end{array}
$$

A formal expression can be considered a particular form of generalized expression, so the second interpretation of monads may, strictly speaking, offer no more than the first. Yet they have each their own flavors, and the flexibility to alternate between stories can be a great tool.

In fact, the free abelian group monad belongs to a larger class of *free R-module monads* which naturally admit interpretation as creating spaces of formal expressions. Rather than assigning to a set X the collection of $\mathbb{Z}$-linear combinations of its elements, it assigns the R-linear combinations, for R a ring. The unit and multiplication are defined using the structure of R, as above. As special cases, taking $R = \mathbb{Z}$ recovers the free abelian group monad and $R = \Bbbk$ the *free vector space monad.*

2. Algebras

We have seen how a monad T on a category $\mathcal{C}$ can be understood as returning generalized versions of its inputs, which witness an algebraic structure borne by the unit and multiplication maps η and μ. As monads are endofunctors on a category, such generalized spaces are in fact peers to the 'original' or 'classical' objects contrasted in 1.3.

In the spirit of generalization, it is natural to ask which objects of $\mathcal{C}$ are 'already generalized', enjoying the interplay with μ and η provided to the image of T. Of course, the image of T obeys these properties by design – in a sense, we search precisely for a generalization of the image of T. In the case of the discrete distribution monad, the question amounts to: which sets behave as though they were distributions over sets? Stated slightly differently, which sets consist of elements which appear naturally as weighted averages across a smaller set? Convexity may come to mind, and this turns out to be on the mark.

- In 2.1 we generalize the niceties granted to the image of T and arrive at the *algebras* of a monad. We confirm that this is a generalization by observing that elements in the image of T are special cases referred to as *free algebras*.
- In 2.2 we consider the categories of algebras and free algebras associated to a monad, which set the stage for central results in Chapter 3.

2.1. (Free) Algebras

We look to generalize the algebraic structure afforded to objects TX in the image of a monad T by considering objects in the ambient category $\mathcal{C}$ which – in the language of 1.3 – carry their own analogue to 'evaluation'. In order to ensure that the evaluation have the appropriate structure, the most naive approach is to:

(i) Consider the diagram encoding TX's interaction with T and μ, the structure being generalized:

$$\begin{array}{ccc} T^3X & \xrightarrow{T\mu} & T^2X \\ {\scriptstyle \mu T}\downarrow & & \downarrow{\scriptstyle \mu} \\ T^2X & \xrightarrow{\mu} & TX \end{array} \tag{2.1.1}$$

(ii) Strike out each appearance of TX, replacing it with a more general object $A \in \mathcal{C}$, so as to arrive at

$$\begin{array}{ccc} T^2A & \xrightarrow{T\mu} & TA \\ {\scriptstyle \mu T}\downarrow & & \downarrow{\scriptstyle \mu} \\ TA & \xrightarrow{\mu} & A \end{array} \tag{2.1.2}$$

Note that (2.1.2) is ill-posed in its current form; $\mu : TA \to A$ need not be defined. In fact, none of the morphisms need align with their claimed (co)domains.

(iii) Replace each $\mu : TA \to A$ with an analogue to the evaluation map which is suited to A, denoted $a : TA \to A$.

$$\begin{array}{ccc} T^2A & \xrightarrow{T\mu} & TA \\ {\scriptstyle \mu T}\downarrow & & \downarrow{\scriptstyle a} \\ TA & \xrightarrow{a} & A \end{array} \tag{2.1.3}$$

(iv) Finally, replace $T\mu$ with Ta – so that the map indeed be from T^2A to TA – and replace μT with μ, again so that the map agree with its source and target.

$$\begin{array}{ccc} T^2A & \xrightarrow{Ta} & TA \\ {\scriptstyle \mu}\downarrow & & \downarrow{\scriptstyle a} \\ TA & \xrightarrow{a} & A \end{array} \tag{2.1.4}$$

Note the interpretation: an object A equipped with a map $a : TA \to A$ satisfying (2.1.4) is a space which behaves as if it were generalized, in that:

(i) Generalized elements over A (via T), can be evaluated by a, as any object twice generalized by T can be evaluated by μ

(ii) The evaluation a particular to A is compatible with μ in the only manner sensible – evaluation $T^2A \to A$ beginning with μ or with Ta coincide.

In order to mimic the algebraic structure afforded to the image of T, we should also demand that A and its associated evaluation map a behave well with respect to η. Recall the constraint on μ's interaction with η, reproduced below, which had the interpretation of once generalized elements evaluating to themselves after being made twice generalized by the unit (either post- or pre-composed with T).

$$\begin{array}{ccccc} TX & \xrightarrow{\eta T} & T^2X & \xleftarrow{T\eta} & TX \\ & {\scriptstyle 1_{TX}}\searrow & \downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle 1_{TX}} & \\ & & TX & & \end{array} \tag{2.1.5}$$

Replacing TX with A, as above, produces

$$\begin{array}{ccccc} A & \xrightarrow{\eta T} & TA & \xleftarrow{T\eta} & A \\ & {\scriptstyle 1_A}\searrow & \downarrow{\scriptstyle \mu} & \swarrow{\scriptstyle 1_A} & \\ & & A & & \end{array} \tag{2.1.6}$$

However, η gives rise to only one map from A to TA (that is, η itself applied to A in the usual way). In shifting the setup of (2.1.5) down by an application of T – i.e. 'undoing' one instance of T by passing from TX to A – the choice in constructing a map from $T \Rightarrow T^2$ using η has been eliminated. There is only one map $A \to TA$, given by η itself. As above, $a : TA \to A$ takes the role of μ. We are left with a triangle, rather than a pyramid, of the form

$$\begin{array}{ccc} A & \xrightarrow{\eta_A} & TA \\ & \searrow^{1_A} & \downarrow a \\ & & A \end{array} \tag{2.1.7}$$

which captures the fact that 'old' elements of A evaluate to themselves after embedding in TA.

Diagrams (2.1.4) and (2.1.7) are the fruit of an innocent effort aimed at classifying the objects which, in addition to the image of T, enjoy an algebraic structure vis-à-vis μ and η. Our simple-minded approach turns out to be the righteous one, landing us at the definition of an algebra of T.

Definition 2.1. An *algebra* of a monad T on $\mathcal{C}$, or $T-$*algebra*, consists of an object $A \in \mathcal{C}$ and a map $a : TA \to A$ such that (2.1.4) and (2.1.7) commute.

A common first task in the study of algebras over monads is to verify that elements in the image of T are indeed algebras when equipped with the standard multiplication μ, i.e. that (TX, μ_X) forms an algebra for any $X \in \mathcal{C}$. In our treatment, however, this arises as an immediate consequence of our path to defining an algebra – by starting with the properties afforded to the image of T and weakening conditions using appropriate substitutions. Nevertheless, such algebras deserve their own name.

Definition 2.2. A *free algebra* of T is an algebra of the form (TX, μ_X) for some $X \in \mathcal{C}$.

Example 2.3. Free algebras of the discrete distribution monad $\mathcal{D}$ are discrete distributions over sets with the usual monadic multiplication of averaging. Commutativity of (2.1.4) and (2.1.7) is immediate from monadicity of $\mathcal{D}$. What of algebras over $\mathcal{D}$ which are not free? Recall that an algebra A over $\mathcal{D}$ admits its own analogue to averaging $a : TA \to A$ of distributions over itself which is compatible with $\mathcal{D}$'s averaging μ and with the embedding $\alpha \to 1[\alpha]$, $\alpha \in A$. A first candidate for A is to suppose that such expressions as $\frac{1}{2}[\alpha_1] + \frac{1}{4}[\alpha_2] + \frac{1}{4}[\alpha_3]$ have a natural meaning 'baked into' A – in other words, that A be convex. This turns out to be well-founded, and indeed to classify the algebras of $\mathcal{D}$; going any further requires a definition.

Definition 2.4. A *convex set* consists of a set X equipped with a ternary operation $\langle -, -, - \rangle : [0,1] \times X \times X \to X$ such that, for all $x, y, z \in X$, $r \in [0,1]$ and $r + (1-r)s \neq 0$,

(i) $\langle r, x, x \rangle = x$
(ii) $\langle 0, x, y \rangle = y$
(iii) $\langle r, x, y \rangle = \langle 1-r, y, x \rangle$
(iv) $\langle r, x, \langle s, y, z \rangle \rangle = \langle r + (1-r)s, \langle \frac{r}{r+(1-r)s}, x, y \rangle, z \rangle$

Informally, $\langle r, x, y \rangle$ has the interpretation of returning the element of X with a proportion r of weight on x and $(1-r)$ of weight on y. That $\langle r, x, y \rangle$ specifies an element of

X for all $r \in [0,1]$ recovers the interpretation of a convex set as containing the geodesics between its pairs of points. Conditions (i) - (iii) are sanity checks and (iv) plays a role not unlike associativity, ensuring that a triple of points (x, y, z) can be averaged either by first averaging (y, z), as one of the left hand side, or by first averaging (x, y), as on the right hand side. We are prepared to classify the algebras of $\mathcal{D}$.

Proposition 2.5. *The algebras of the discrete distribution monad $\mathcal{D}$ are exactly convex sets* [5].

Proof sketch. Given an algebra (A, a) over $\mathcal{D}$, the ternary operation on A can be defined using a, i.e.

$$\langle r, x, y\rangle = a(r[x] + (1-r)[y])$$

Making use of commutativity of (2.1.7), which amounts to $a(1[x]) = x$, we have

$$\begin{aligned}\langle r, x, x\rangle &= a(r[x] + (1-r)[x]) \\ &= a(1[x]) \\ &= x \\ \langle 0, x, y\rangle &= a(0[x] + 1[y]) \\ &= y \\ \langle r, x, y\rangle &= a(r[x] + (1-r)[y]) \\ &= a((1-r)[y] + r[x]) \\ &= \langle 1-r, y, x\rangle\end{aligned}$$

The proof of (iv) from 2.4 is more involved and makes use of commutativity of (2.1.4). The more demanding direction is, given a convex set C, to extend its ternary operation to an analogue c to averaging for arbitrary finite distributions. The idea is to proceed inductively.

$$c(w_1[x_1] + \cdots + w_n[x_n]) = \begin{cases} x_1 & w_1 = 1 \\ \langle w_1, x_1, c(\frac{w_2}{1-w_1}[x_2] + \cdots + \frac{w_n}{1-w_1}[x_n])\rangle & w_1 < 1 \end{cases} \tag{2.1.8}$$

One then checks that c is well-defined with respect to re-ordering of the x_i. Then c satisfies (2.1.7) by design, i.e. $c(1[x]) = x$, and (2.1.4) with some work. That these procedures are inverse to one another completes the proof. □

Remark 2.6. Roughly speaking, convex sets admit averaging of pairs of elements, while algebras of $\mathcal{D}$ admit averaging of finite collections of elements – that they coincide is a testament to the strength of the conditions imposed on convex sets. In particular, it was crucial that the extension from pairs to finite sets witnessed in (2.1.8) be well-defined and, moreoever, well-behaved.

Remark 2.7. How can the free algebras C over $\mathcal{D}$ be detected at the level of convex sets? They are exactly those which admit a uniquely generating set, i.e. some $G = \{g_i\} \subseteq C$ such that any element of C appears as $c(\sum w_i[g_i])$ for unique w_i, with c defined as in (2.1.8). Any such C appears as TG and, conversely, any TX has generating set X.

Example 2.8. The generalization of free algebras to algebras is sometimes fruitless – consider, for instance, the free vector space monad. It can be shown that its algebras are vector spaces; as all vector spaces are free modules over a choice of basis, the only algebras over the free vector space monad are the free algebras.

Where do algebras fit into the picture described in 1.3? They admit interpretation as objects in which generalized elements (or formal expressions) can already be evaluated. Convex sets admit their own averaging, without first passing to formal averages, and vector spaces likewise admit evaluation of sums, rather than just sums of sums. If monads are ways of encoding algebraic structure in objects, its algebras are those objects which didn't need the help.

2.2. The Eilenberg-Moore and Kleisli categories

Having formalized the theory of algebras, it is natural to ask what form structure-preserving morphisms between such algebras should take. As the structure on an algebra (A, a) is determined by its evaluation map $a : TA \to A$, a morphism of algebras $f : A \to B$ should respect the evident evaluation maps.

Definition 2.9. A morphism $f : (A, a) \to (B, b)$ of T-algebras is a map $f : A \to B$ in $\mathcal{C}$ so that the following diagram commutes.

$$\begin{array}{ccc} TA & \xrightarrow{Tf} & TB \\ {\scriptstyle a}\downarrow & & \downarrow{\scriptstyle b} \\ A & \xrightarrow{f} & B \end{array}$$

Endowing morphisms of T-algebras with the composition in $\mathcal{C}$, and recalling that rectangles commute when their constituent squares do, the category of T-algebras and their morphisms takes shape.

Definition 2.10. The *Eilenberg-Moore category* $\mathcal{C}^T$ of a monad T over $\mathcal{C}$ is the category of T-algebras and morphisms of T-algebras.

Extending the work from 2.5 to morphisms yields the description of the Eilenberg-Moore category for the discrete distribution monad.

Corollary 2.11. *$\mathsf{Set}^{\mathcal{D}}$ is isomorphic to the category of convex sets and affine functions $f : X \to Y$ satisfying $f(\langle r, x, y\rangle) = \langle r, f(x), f(y)\rangle$* [5].

Though arriving at the Eilenberg-Moore category of a monad came as a natural consequence of examining its algebras, the value of this discovery is postponed to Chapter 3, where the Eilenberg-Moore category – alongside the Kleisli category – plays a crucial role in connecting monads to adjunctions.

Just as the study of algebras led to special consideration of – and a name for – free algebras, we turn now to the category of free algebras over a monad, which earns its own name.

Definition 2.12. The *Kleisli category* $\mathcal{C}_T$ of a monad T over $\mathcal{C}$ is the full subcategory of the Eilenberg-Moore category $\mathcal{C}^T$ consisting of free algebras.

The above definition of the Kleisli category, though compact, is somewhat unconventional, often supplanted by the following definition as $\mathcal{C}$ with 'T-shifted' morphisms.

Definition 2.13. The *Kleisli category* $\mathcal{C}_T$ of a monad T over $\mathcal{C}$ is the category with

- objects equal to those of $\mathcal{C}$
- $\mathrm{Hom}_{\mathcal{C}_T}(A, B) = \mathrm{Hom}_{\mathcal{C}}(A, TB)$

In order to disambiguate hom-sets, we denote $f \in \mathrm{Hom}_{\mathcal{C}_T}(A, B)$ as $f : A \rightsquigarrow B$. Identities and composition make use of the structure of T:

- $\eta_A : A \rightsquigarrow A$ is the identity on A in $\mathcal{C}_T$
- The composition $g \circ_k f$ of $f : A \rightsquigarrow B$, $g : B \rightsquigarrow C$ is defined

$$A \xrightarrow{f} TB \xrightarrow{Tg} T^2C \xrightarrow{\mu_c} TC$$

That η_A serves as the identity in the Kleisli category is a consequence of its naturality and unitarity. For $g : Z \rightsquigarrow A$ and $f : A \rightsquigarrow B$ one has

$$\begin{aligned} \eta_A \circ_k g &= Z \xrightarrow{g} TA \xrightarrow{T\eta_A} T^2A \xrightarrow{\mu_A} TA \\ &= Z \xrightarrow{g} TA && \text{(unitarity of } \eta) \end{aligned}$$

and

$$\begin{aligned} f \circ \eta_A &= A \xrightarrow{\eta_A} TA \xrightarrow{Tf} T^2B \xrightarrow{\mu_B} TB \\ &= A \xrightarrow{f} TB \xrightarrow{\eta_{TB}} T^2B \xrightarrow{\mu_B} TB && \text{(naturality of } \eta) \\ &= A \xrightarrow{f} TB && \text{(unitarity of } \eta) \end{aligned}$$

Associativity of composition is slightly more involved. For $h : Y \rightsquigarrow G$ and f, g as in the proof of unitarity, one has:

$$\begin{aligned} (f \circ_k g) \circ_k h &= \mu_B \circ T(f \circ_k g) \circ h \\ &= \mu_B \circ T(\mu_B \circ Tf \circ g) \circ h \\ &= \mu_B \circ T\mu_B \circ T^2 f \circ Tg \circ h \\ &= \mu_B \circ \mu_{TB} \circ T^2 f \circ Tg \circ h && \text{(associativity of } \mu) \\ &= \mu_B \circ Tf \circ \mu_A \circ Tg \circ h && \text{(naturality of } \mu) \\ &= \mu_B \circ Tf \circ (g \circ_k h) \\ &= f \circ_k (g \circ_k h) \end{aligned}$$

For the moment we cave to convention and adhere to the more involved Definition 2.13, turning to example in order to observe T-shifted morphisms and their composition.

Example 2.14. The Kleisli category $\mathsf{Set}_{\mathcal{P}}$ of the power set monad consists of sets equipped with $\mathcal{P}$-shifted maps $X \rightsquigarrow Y$ represented by functions $X \to \mathcal{P}Y$. In particular, a Kleisli morphism $X \rightsquigarrow Y$ is exactly a binary relation between X and Y under the identification between a function $f : X \to \mathcal{P}Y$ and (nearly) its graph $\{(x, y) : x \in X, y \in f(x)\}$. Kleisli composition

$$g \circ_k f = X \xrightarrow{f} \mathcal{P}Y \xrightarrow{Tg} \mathcal{P}^2 Z \xrightarrow{\mu_Z} \mathcal{P}Z$$

applies f to X, then g element-wise to the image of f, and takes unions over the resulting collection of subsets of Z. In symbols,

$$(g \circ_k f)(x) = \{z \in Z : \exists y \in f(X),\ z \in f(y)\}$$

This coincides with the category Rel of sets equipped with binary relations. In particular, a morphism in Rel from A to B is exactly a relation, i.e. a subset R of $A \times B$. Writing $a \sim_R b$ if $(a, b) \in R$, composition in Rel

$$A \xrightarrow{R} B \xrightarrow{S} C$$

is exactly transitivity of the relation, i.e. $a \sim_{S \circ R} c \iff a \sim_R b \wedge b \sim_S c$ for some $b \in B$. Or, in set notation, $(a, c) \in S \circ R$ exactly when $(a, b) \in R$ and $(b, c) \in S$ for some $b \in B$. This coincides with Kleisli composition in $\mathsf{Set}_{\mathcal{P}}$, completing the isomorphism.

It remains to show that Definitions 2.12 and 2.13 are equivalent, a claim which is neither obvious nor intuitive. We approach it, and see the merits of C^T and C_T, in Chapter 3.

3. Adjunctions

Consider an opposing pair of functors $F : \mathcal{C} \rightleftarrows \mathcal{D} : G$. The degree of structure satisfied by F and G is determined by how strictly they 'undo' each other – as measured by how closely the composites FG and GF resemble the evident identities. When F and G exhibit the most structure, FG identically equals $1_\mathcal{D}$ and likewise $GF = 1_\mathcal{C}$, giving rise to an *isomorphism* of categories. Yet this is an overly restrictive condition to ask of F and G; its view of functors is too granular. When equality is relaxed to natural isomorphism – $FG \cong 1_\mathcal{D}$ and $GF \cong 1_\mathcal{C}$ – one obtains an *equivalence* of categories, a relationship of greater importance in category theory, which is nevertheless strong.[6]

Can the demands on F and G be further relaxed? Following the path from isomorphism to equivalence, one can require only the existence of natural transformations – not necessarily natural isomorphisms – connecting the composites of F and G with the appropriate identities. In symbols, that there be $\eta : 1_\mathcal{C} \Rightarrow GF$ and $\epsilon : FG \Rightarrow 1_\mathcal{D}$. That one transformation be out of the identity and the other into the identity is crucial in order to introduce compatibility constraints between η and ϵ. In particular, one ought demand that η and ϵ behave like inverses in some manner – their mere presence says little about the relationship between F and G. Notably, this has the effect of, for the first time, destroying the symmetry between F and G. When all is said and done, there will be a handedness to the definition of an adjunction.

An obstruction to constraining η and ϵ is that they share neither source nor target; this can be remedied by making use F or G, with some subtlety as to when. In particular, $F\eta$ yields a transformation $F \Rightarrow FGF$, while ϵF - with $(\epsilon F)_c = \epsilon_{Fc}$ - yields a transformation $FGF \Rightarrow F$. Likewise, one has $\eta G : G \Rightarrow GFG$ and $G\epsilon : GFG \Rightarrow G$. We are equipped to demand that η and ϵ behave like inverses, up to translation.

Definition 3.1. An oppposing pair of functors $F : \mathcal{C} \rightleftarrows \mathcal{D} : G$ assemble into an *adjunction* when there exist natural transformations $\eta : 1_\mathcal{C} \Rightarrow GF$ and $\epsilon : FG \Rightarrow 1_D$ such that the following triangle identities commute, in which case F is *left adjoint* to G and G is *right adjoint* to F.

$$\begin{array}{ccc} F & \overset{F\eta}{\Longrightarrow} & FGF \\ & {\scriptstyle 1_F} \searrow\!\!\searrow & \Downarrow {\scriptstyle \epsilon F} \\ & & F \end{array} \qquad\qquad \begin{array}{ccc} G & \overset{\eta G}{\Longrightarrow} & GFG \\ & {\scriptstyle 1_G} \searrow\!\!\searrow & \Downarrow {\scriptstyle G\epsilon} \\ & & G \end{array} \tag{3.0.1}$$

[6]Compare with the choice to demand that the tensor product in a monoidal category associate and be unital up to isomorphism, rather than 'on the nose'.

In the above, η is denoted the *unit* of the adjunction and ϵ the *counit*. One writes $F \dashv G$ to denote that F is left adjoint to G. In addition to Definition 3.1, adjunctions admit an equivalent, and perhaps more concrete, characterization by way of Hom-sets.

Definition 3.2. A functor $F : \mathcal{C} \to \mathcal{D}$ is *left adjoint* to $G : \mathcal{D} \to \mathcal{C}$ if there exists an isomorphism of sets

$$\mathrm{Hom}_{\mathcal{D}}(Fc, d) \cong \mathrm{Hom}_{\mathcal{C}}(c, Gd)$$

for all $c \in \mathcal{C}$ and $d \in \mathcal{D}$ which is natural in both variables. That is, the isomorphisms assemble into natural transformations $\mathrm{Hom}_{\mathcal{D}}(F-, d) \cong \mathrm{Hom}_{\mathcal{C}}(-, Gd)$ and $\mathrm{Hom}_{\mathcal{D}}(Fc, -) \cong \mathrm{Hom}_{\mathcal{C}}(c, G-)$.

Notably, one passes from Definition 3.2 to Definition 3.1 by applying the isomorphism on Hom-sets to identities. In particular, the component of the unit $\eta : 1_{\mathcal{C}} \Rightarrow GF$ at c is defined to be the image – or the *transpose* – of the identity 1_{Fc} under the isomorphism. Similarly, the component of the counit $\epsilon : FG \Rightarrow 1_{\mathcal{D}}$ at d is defined to be the transpose of 1_{Gd}. We will see that Definition 3.2 lends itself to detecting adjunctions in nature, while 3.1 will be of great use when monads enter the picture.

A famous class of adjunctions is that of free $\dashv$ forgetful adjunctions. Forgetful functors $U : B \to A$ usually send an object $b \in B$ to the object $a \in A$ which 'underlies' b and enjoys less structure. Any left adjoint F to a forgetful functor earns the name 'free', and it usually enriches an object a with additional structure in a choiceless manner.

Example 3.3. The forgetful functor $U : \mathsf{Top} \to \mathsf{Set}$ sends a topological space X to the set UX underlying it and a continuous map $f : X \to Y$ to the function $Uf : UX \to UY$ underlying it. By Definition 3.2, the free functor $F : \mathsf{Set} \to \mathsf{Top}$ admits a natural isomorphism

$$\mathrm{Hom}_{\mathsf{Top}}(FX, Y) \cong \mathrm{Hom}_{\mathsf{Set}}(X, UY)$$

meaning continuous maps out of FX correspond to functions out of X. This is achieved by the discrete topology on X, for which functions out of X are automatically continuous. We made use of this fact when noting that monoids in Top are exactly classical monoids with continuous multiplication. So the free topological space on a set X is the space X equipped with the discrete topology.

In this case, U admits a right adjoint G as well. Again making use of Definition 3.2, we see that such a functor would assign a space GX to X so that functions into X correspond naturally to continuous maps into GX. This is satisfied by the indiscrete topology, as any function into an indiscrete space is obligated to be continuous.

Example 3.4. The forgetful functor $U : \mathsf{Vect}_{\Bbbk} \to \mathsf{Set}$ sends a vector space to its underlying set and a linear map to its underlying function. The free vector space on a set X is the vector space with basis X, as a linear map out of such a space is exactly a function out of X. More generally, the free R-module on a set X is $FX = \oplus_X R$, as a homomorphism out of $\oplus_X R$ can be identified with a function of X, and vice versa.

Consider the actions of the unit and counit in the previous example: $\eta_X : X \to UF(X)$ is a function from X to the set of $\Bbbk$-linear combinations on X. By the earlier remark on extracting the (co)unit of an adjunction from the natural bijection between Hom-sets,

$$\mathrm{Hom}_{\mathsf{Vect}}(FX, V) \cong \mathrm{Hom}_{\mathsf{Set}}(X, UV)$$

we have that η_X is the transpose of 1_{FX}. Since our bijection sends a linear map to its image on basis vectors, we have that η_X sends an element of X to itself, thought of as a basis vector. On the other hand, $\epsilon_V : FU(V) \to V$ is the transpose of 1_{UV}; that is, the linear function $FU(V) \to V$ which sends basis vectors of $FU(V)$ (vectors in V) to themselves, as elements of V. Thus, on arbitrary vectors in $FU(V)$, which are $\Bbbk$-linear combinations of vectors in V, it has the effect of evaluation.

Thus, the unit embeds an element as a degenerate case of an element with more structure, and the counit evaluates formal $\Bbbk$-linear combinations of vectors into vectors. This smells of monads, and indeed it should – we have arrived at the free vector space monad. In fact, the free vector space monad is precisely UF, and the unit of the adjuction coincides with the unit of the monad. The counit, as a linear map rather than (just) a function, requires some massaging in order to serve as the monad's multiplication, but it captures the relevant notion of evaluation.

This is no coincidence. Taking $R = \mathbb{Z}$, rather than $R = \Bbbk$, in the free $\dashv$ forgetful adjunction on Mod_R gives rise to the free abelian group monad. In fact, the relationship between adjunctions and monads extends far beyond the free $\dashv$ forgetful adjunction and forms the subject of this chapter.

- In 3.1 we show that all adjunctions give rise to monads, by post-composing the right adjoint with the left.
- In 3.2 we use the Kleisli and Eilenberg-Moore categories to show, remarkably, that all monads arise in this way. We also characterize the Kleisli and Eilenberg-Moore categories as universal among the adjunctions giving rise to a given monad.

3.1. From adjunctions to monads

The first step in examining the intimate relationship between monads and adjunctions is to generalize the behavior of the free $\dashv$ forgetful adjunction in generating monads. In particular, we have seen in $U : \mathsf{Mod}_R \rightleftarrows \mathsf{Set} : F$ with $F \dashv U$ that UF gave rise to the free R-module monad, which sends a set X to the collection of finite formal R-linear combinations over X and extends functions on X linearly to the free R-module over X. The unit $\eta : 1_{\mathsf{Set}} \Rightarrow UF$ from the adjunction dovetailed with the unit of the monad, sending an element of X to itself as a degenerate R-linear. The multiplication μ of the monad, however, which evaluates R-linear combinations over R-linear combinations into R-linear combinations, did not coincide exactly with the counit $\epsilon : FU \Rightarrow 1_{\mathsf{Mod}_R}$, which evaluates a formal sum of elements in an R-module to its true sum. Translating ϵ using F and U, however, so as to arrive at $U\epsilon F : UFUF \Rightarrow UF$, yields the function – not R-module homomorphism – which plays the role of monadic multiplication by evaluating R-linear combinations of R-linear combinations.

Proposition 3.5 shows that this procedure is not an artifact of the free $\dashv$ forgetful adjunction but rather a property of all adjuncions: post-composing a right adjoint with its left always produces a monad.

Proposition 3.5. *An adjunction $F \dashv U$ with $F : \mathcal{C} \rightleftarrows \mathcal{D} : U$ and $\eta : 1_{\mathcal{C}} \Rightarrow UF$, $\epsilon : FU \Rightarrow 1_{\mathcal{D}}$ gives rise to a monad on $\mathcal{C}$ with:*

- *endofunctor $T = UF$*
- *unit $\eta : 1_{\mathcal{C}} \Rightarrow UF$*
- *multiplication $\mu = U\epsilon F : UFUF \Rightarrow UF$*

Proof. The unitarity and commutativity axioms on T take the form

$$\begin{array}{ccccc} UF & \xRightarrow{\eta UF} & UFUF & \xLeftarrow{UF\eta} & UF \\ & {\scriptstyle 1_{UF}}\searrow & \Downarrow{\scriptstyle U\epsilon F} & \swarrow{\scriptstyle 1_{UF}} & \\ & & UF & & \end{array} \qquad \begin{array}{ccc} UFUFUF & \xRightarrow{UFU\epsilon F} & UFUF \\ {\scriptstyle U\epsilon FUF}\Downarrow & & \Downarrow{\scriptstyle U\epsilon F} \\ UFUF & \xRightarrow[U\epsilon F]{} & UF \end{array} \tag{3.1.1}$$

The unitarity triangles commute by the triangle identities demanded of η and ϵ in Definition 3.1 (and pre- or post-composition of F or U). The associativity square commutes by naturality of $U\epsilon : UFU \Rightarrow F$. In particular, setting $UFU = X, Y = F$, and $\epsilon F = f$, the square amounts to

$$\begin{array}{ccc} UFU(X) & \xRightarrow{UFU(f)} & UFU(Y) \\ {\scriptstyle U\epsilon(X)}\Downarrow & & \Downarrow{\scriptstyle U\epsilon(Y)} \\ U(X) & \xRightarrow[U(f)]{} & U(Y) \end{array}$$

which expresses precisely naturality of $U\epsilon$ at $f : X \to Y$. □

Example 3.6. We are acquainted with the *free R-module monad*, which arises from the adjunction $U : \mathsf{Mod}_R \rightleftarrows \mathsf{Set} : F$ and specializes to the free vector space monad when $R = \Bbbk$ and the free abelian group monad when $R = \mathbb{Z}$. Similarly, the *free group monad* arises from the free $\dashv$ forgetful adjunction $U : \mathsf{Grp} \rightleftarrows \mathsf{Set} : F$. It sends a set X to the collection of finite words with letters in X and X^{-1} – the set of formal inverses to elements in X – subject to identifications which, for instance, identify xx^{-1} with the empty word.

Example 3.7. The *maybe monad* in computer science arises from the free $\dashv$ forgetful adjunction between pointed sets and ordinary sets. The category Set_* of pointed sets consists of sets with a distinguished point and functions which preserve such points. The forgetful functor $U : \mathsf{Set}_* \to \mathsf{Set}$ sends a pointed set to its underlying set, forgetting which element takes the role of the distinguished point, and sends functions between pointed sets to their underlying functions between sets.

The free pointed set on an ordinary set X is $FX = X \cup \{X\}$, also denoted X_+, the set X along with a new distinguished point $\{X\}$. A function $f : X \to Y$ between ordinary sets extends to a function $Ff : X_+ \to Y_+$ by sending the added basepoint to the added basepoint. This defines an adjuction, as the natural bijection

$$\mathrm{Hom}_{\mathsf{Set}_*}(X_+, P) \cong \mathrm{Hom}_{\mathsf{Set}}(X, UP)$$

is defined by restricting a map of pointed sets $X_+ \to P$ to a function $X \to UP$, i.e. by forgetting the action on the point $\{X\}$. Injectivity amounts to the statement that a map of pointed sets out of X_+ is determined by its action on X. The unit $\eta : 1 \Rightarrow UF$ is given by the inclusion $X \hookrightarrow X_+$ and the monad's multiplication $\mu : (X_+)_+ \to X_+$ acts as the identity on X and sends both added points to the single added point in X_+. By Proposition 3.5 - or direct means - the following diagrams commute, confirming monadicity of the maybe monad.

$$\begin{array}{ccccc} X_+ & \xrightarrow{\eta_{X_+}} & (X_+)_+ & \xleftarrow{(\eta_X)_+} & X_+ \\ & {\scriptstyle 1_{X_+}} \searrow & \downarrow{\scriptstyle \mu_X} & \swarrow {\scriptstyle 1_{X_+}} & \\ & & X_+ & & \end{array} \qquad \begin{CD} ((X_+)_+)_+ @>{(\mu_X)_+}>> (X_+)_+ \\ @V{\mu_{X_+}}VV @VV{\mu_X}V \\ (X_+)_+ @>{\mu_X}>> X_+ \end{CD} \tag{3.1.2}$$

3.2. From monads to adjunctions

We have seen in Proposition 3.5 that any adjunction is the source of a monad on its left adjoint's domain. It is natural to wonder whether one can travel in the opposite direction, constructing an adjunction which witnesses a given monad. Remarkably, this indeed the case, a fact which makes use of the Kleisli and Eilenberg-Moore categories [10].

Proposition 3.8. *Let T be a monad on $\mathcal{C}$ with unit η and multiplication μ. There is an adjunction $F^T : \mathcal{C} \rightleftarrows \mathcal{C}^T : U^T$ between $\mathcal{C}$ and the Eilenberg-Moore category of T which witnesses T as its induced monad.*

Proof. Recall that $\mathcal{C}^T$ consists of algebras over T equipped with morphisms between algebras which respect their evaluation maps, i.e.

$$\begin{CD} TA @>{Tf}>> TB \\ @V{a}VV @VV{b}V \\ A @>{f}>> B \end{CD}$$

Then there is a forgetful functor $\mathcal{C}^T \to \mathcal{C}$ which we take to be U^T. In the other direction, $F^T : \mathcal{C} \to \mathcal{C}^T$ sends an object A to its free algebra TA with evaluation map $\mu_A : T^2A \to TA$. Note that morphisms in $\mathcal{C}$ map to morphisms of T-algebras in $\mathcal{C}^T$ by naturality of μ.

To see that F^T is left adjoint to U^T, we appeal to Definition 3.1. Since $U^T F^T = T$, we may take the monadic unit $\eta : 1_{\mathcal{C}} \Rightarrow T$ to be the unit of the adjunction. The counit $\epsilon : F^T U^T \Rightarrow 1_{\mathcal{C}^T}$ is defined at an algebra of $\mathcal{C}^T$ using its evaluation map. In particular, the component $\epsilon_{(A,a)}$ of the counit at the algebra (A, a) is precisely $a : (TA, \mu_A) \to (A, a)$. That this be a morphism of T-algebras is identically the content of axiom (2.1.4) demanded of T-algebras.

$$\begin{array}{ccc} T^2A & \xrightarrow{Ta} & TA \\ {\scriptstyle \mu_A}\downarrow & & \downarrow{\scriptstyle a} \\ TA & \xrightarrow{a} & A \end{array}$$

Since $U^T \epsilon F^T(A) = U^T \epsilon_{(F^T A, \mu_A)} = U^T \mu_A = \mu_A$, the monad induced by $F^T \dashv U^T$ is precisely (T, η, μ). It remains only to show F^T is indeed left adjoint to U^T with η and ϵ as above.

$$\begin{array}{ccc} F^T & \xRightarrow{F^T\eta} & F^T U^T F^T \\ & {\scriptstyle 1_{F^T}}\searrow & \Downarrow{\scriptstyle \epsilon F^T} \\ & & F \end{array} \qquad \begin{array}{ccc} U^T & \xRightarrow{\eta U^T} & U^T F^T U^T \\ & {\scriptstyle 1_{U^T}}\searrow & \Downarrow{\scriptstyle U^T\epsilon} \\ & & U^T \end{array} \tag{3.2.1}$$

Composing the left diagram of (3.2.1) with U^T, and recalling $U^T F^T = T$, yields

$$\begin{array}{ccc} T & \xRightarrow{T\eta} & T^2 \\ & {\scriptstyle 1_T}\searrow & \Downarrow{\scriptstyle \mu} \\ & & T \end{array}$$

which commutes by unitarity of η. Since U^T is faithful, the original diagram commutes. The right diagram of (3.2.1) can likewise be reduced to a unitarity condition on η demanded by T. $\square$

Note that the adjunction of Proposition 3.8 is simply a free $\dashv$ forgetful adjunction between $\mathcal{C}$, endowed with the monad T, and the category of T-algebras. An analogous result furthermore holds for the Kleisli category C_T associated to T, defined for the moment as $\mathcal{C}$ with T-shifted morphisms, i.e. $\mathrm{Hom}_{\mathcal{C}_T}(A, B) = \mathrm{Hom}_{\mathcal{C}}(A, TB)$. We will see shortly, however, that this definition coincides with that of the Kleisli category as the full subcategory of $\mathcal{C}^T$ consisting of free algebras.

Proposition 3.9. *Let (T, η, μ) be a monad on $\mathcal{C}$. There is an adjunction $F_T : \mathcal{C} \rightleftarrows \mathcal{C}_T : U_T$ between $\mathcal{C}$ and the Kleisli category of T which witnesses T as its induced monad.*

Proof. Define F_T to act as the identity on objects and send $f : A \to B$ in $\mathcal{C}$ to the Kleisli morphism $F_T f : A \rightsquigarrow B$ with

$$F_T f = A \xrightarrow{f} B \xrightarrow{\eta_B} TB$$

We saw in Definition 2.13 that $\eta_A : A \rightsquigarrow A$ is the identity on A in $\mathcal{C}_T$, so F_T preserves identities. To see that it respects composition, let $g : A \to B$ and $f : B \to C$.

$$\begin{aligned}
F_T(f) \circ_k F_T(g) &= (\eta_C \circ f) \circ_k (\eta_B \circ g) \\
&= \mu_C \circ T(\eta_C \circ f) \circ (\eta_B \circ g) \\
&= \mu_C \circ T\eta_C \circ Tf \circ \eta_B \circ g \\
&= \mu_C \circ T\eta_C \circ \eta_C \circ f \circ g && \text{(naturality of } \eta) \\
&= \eta_c \circ f \circ g && \text{(unitarity of } \eta) \\
&= F_T(f \circ g)
\end{aligned}$$

Now set U_T to act as T on objects and to carry $A \rightsquigarrow B$, represented by $f : A \to TB$ in $\mathcal{C}$, to

$$U_T f = TA \xrightarrow{Tf} T^2 B \xrightarrow{\mu_B} TB$$

Then U_T carries $\eta_A : A \rightsquigarrow A$ to $\mu_A \circ T\eta_A = 1_{TA}$. Let $B \rightsquigarrow TC$ be represented by $g : B \to TC$ in $\mathcal{C}$ and f as above.

$$\begin{aligned}
U_T(g \circ_k f) &= U_T(\mu_C \circ Tg \circ f) \\
&= \mu_C \circ T(\mu_C \circ Tg \circ f) \\
&= \mu_C \circ T\mu_C \circ T^2 g \circ Tf \\
&= \mu_C \circ \mu_{TC} \circ T^2 g \circ Tf && \text{(associativity of } \mu) \\
&= \mu_C \circ Tg \circ \mu_B \circ Tf && \text{(naturality of } \mu) \\
&= U_T(g) \circ U_T(f)
\end{aligned}$$

That F_T be left adjoint to U_T is an immediate consequence of the definition of hom-sets in the Kleisli category. In particular,

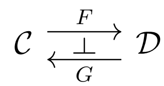
$$\mathrm{Hom}_{\mathcal{C}_T}(F_T A, B) = \mathrm{Hom}_{\mathcal{C}_T}(A, B) \cong \mathrm{Hom}_{\mathcal{C}}(A, TB) = \mathrm{Hom}_{\mathcal{C}}(A, U_T B)$$

The proof is completed by observing that $U_T F_T = T$. □

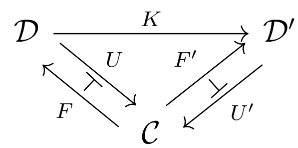
Remark 3.10. Propositions 3.8 and 3.9 justify a view of monads as byproducts of adjunctions. Indeed some adopt such a view, characterizing the monad as the 'trace' of an adjunction or the 'shadow' cast by an adjunction on its left adjoint's domain [10, 11]. This seems somewhat like the characterization of limits as terminal objects – recall that limits are terminal objects in the category of cones over a diagram and a terminal object is a limit over the empty diagram. That is to say, it is a correct but somewhat unnatural presentation. Just as it often feels stilted to describe a limit by defining a suitably convoluted category and observing its terminal object, it is difficult to imagine the discovery of most monads of interest by way of examining adjunctions.

Given two solutions to the problem of witnessing an adjunction which induces a given monad, by Eilenberg-Moore and by Kleisli, it is natural to wonder in what manner, if any, the solutions are related. As is often the case, resolution of this question demands consideration of the greater landscape – it is to the world of adjunctions inducing a given monad to which we now turn.

Definition 3.11. For (T, η, μ) a monad on $\mathcal{C}$, the category Adj_T of adjunctions over T has adjunctions $F \dashv U$ inducing (T, η, μ) as objects, i.e.



A morphism between adjunctions is a functor between the domains of the right adjoints which commutes with the right and left adjoints, i.e.



with $KF = F'$ and $U'K = U$.

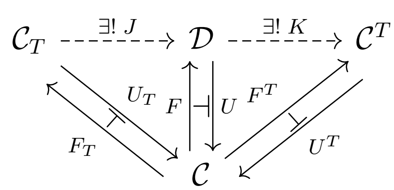
Note that all adjunctions in Adj_T share units – recall that the unit of an adjunction and the monad it induces coincide – whereas counits may differ. Simple consequences whose proofs we omit are that:

(i) morphisms of adjunctions commute with counits, i.e. $K\epsilon = \epsilon' K$, and

(ii) morphisms of adjunctions respect transpositions, i.e. the transpose of a morphism $f : c \to Ud = U'Kd$ in $\mathcal{D}'$ is the image under K of its tranpose in $\mathcal{D}$.

We are equipped to characterize the Kleisli and Elienberg-Moore categories as extremal objects in Adj_T.

Proposition 3.12. *Fix a monad (T, η, μ) on $\mathcal{C}$. The Kleisli category $\mathcal{C}_T$ of T is initial in Adj_T and the Eilenberg-Moore category $\mathcal{C}^T$ is terminal. Diagrammatically, for any adjunction $F \dashv U$ inducing T, there exist unique $J : \mathcal{C}_T \to \mathcal{D}$ and $K : \mathcal{D} \to \mathcal{C}^T$ such that the following commutes.*



Proof. We observe that J and K are uniquely (and functorially) defined by the above constraints. Since $JF_T = F$ and F_T acts as the identity on objects, J must coincide with F on objects. And since J commutes with transposes, Jf is obligated to be the transpose of f under $F \dashv U$.

As $U^T Kd = Ud$, and U^T is the identity on objects, $Kd = Ud$ equipped with an evaluation map as a T-algebra. By the proof of Proposition 3.8, the evaluation map a of an algebra (A, a) is precisely the component of the counit of $F^T \dashv U^T$ at (A, a). Thus a is the map $(TA, \mu_A) \to (A, a)$ which is the transpose of $1_A = 1_{U^T(A,a)}$. Since K commutes with transposes, we must then define $Kd = (Ud, U\epsilon_d)$. On morphisms, K acts simply as U, again a consequence of $U^T K = U$.

We leave functoriality of J and K as an exercise. ☐

Then, by the above, there exists a canonical functor from the Kleisli category of a monad to the Eilenberg-Moore category which commutes with their respective free and forgetful functors. Notably, it succeeds in identifying the working Definition 2.12 of the Kleisli category with 2.13 – $\mathcal{C}$ with T-shifted morphisms is equivalent to the full subcategory of $\mathcal{C}^T$ whose objects are the free T-algebras.

Corollary 3.13. *The canonical functor $K : \mathcal{C}_T \to \mathcal{C}^T$ in Adj_T is fully faithful with image consisting of free T-algebras.*

Proof. By the proof of Proposition 3.12, $Kc = (Tc, \mu_C)$. In particular, $Tc = U_T c$ and $\mu_c = U_T \epsilon_c^{C_T}$ for ϵ^{C_T} the counit of the Kleisli adjunction. To see that K is fully faithful, recall that its action on hom-sets

$$\mathrm{Hom}_{\mathcal{C}_T}(c, c') \longrightarrow \mathrm{Hom}_{\mathcal{C}^T}(Kc, Kc') = \mathcal{C}^T((Tc, \mu_c), (Tc', \mu_{c'}))$$

commutes with the isomorphisms which identify both sets with $\mathrm{Hom}_{\mathcal{C}}(c, Tc')$. □

Having developed the basic machinery of monads, we look to bring life to the subject by way of categorical probability, a modern world in which probability monads take center stage. It is in these spirits that we turn to Chapter 4.

4. Probability Monads

Categorical probability has emerged as the effort to apply categorical techniques to the study of probability, measure theory more generally, and mathematical statistics [9]. In the introduction, we considered the silhouette of a probability monad, that central tool of categorical probability which assigns to a space X a collection of probability measures over X.

Despite having become literate in the theory of monads in the time since, the picture cannot be made much crisper – there is no formal definition for a probability monad. Nevertheless, our background in monads merits a second pass at the framework:

- A probability monad T assigns to $X \in \mathcal{C}$ a collection TX of probability measures over X. In those settings in which elements of $\mathcal{C}$ carry structure beyond that of a set – say a topology or a metric – TX must thus be endowed with such structure as well.
- T acts on morphisms, so that a map $f : X \to Y$ extend to a map $Tf : TX \to TY$, most often via the pushforward of measures.
- $\mu : T^2 \Rightarrow T$ serves to reduce measures over measures to mere measures, usually via averaging. An archetypal example is the compression of (0.0.2) to (0.0.3), in which a random choice of a fair coin and a double-sided heads coin was seen to amount to a single coin with 3/4 weight on heads. Formally, μ customarily takes the form of integration.
- $\eta : 1 \Rightarrow T$ extracts a probability measure over a space X from each of its elements x. The natural choice is the Dirac measure at x, which outputs 1 if a measurable set contains x and 0 otherwise.

The profile may bring to mind a familiar figure, that of the discrete distribution monad $\mathcal{D}$ on Set. Indeed it should; $\mathcal{D}$ is an elemental probability monad which has served as an object of our study in slight disguise.

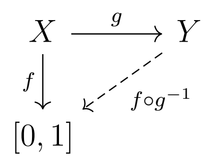
- In 4.1 we revisit the discrete distribution monad, casting it in a new light and making use of tools from Chapters 2 and 3.
- In 4.2 we turn to the Giry monad, examine its Kleisli category, and briefly introduce the Kantorovich monad.

4.1. An old friend

Recall the discrete distribution monad $\mathcal{D}$ on Set, defined in 1.17 so as to assign to X the set of finitely supported functions $f : X \to [0, 1]$ with $\sum f(x) = 1$. $\mathcal{D}X$ now admits interpretation as a collection of probability measures on X endowed with the discrete

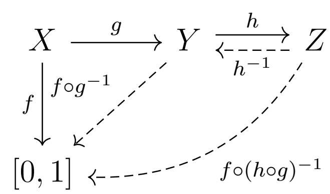
σ-algebra, so that all its subsets be measurable. Formally, $f \in \mathcal{D}X$ gives rise to the probability measure P_f on X with $P_f(S) = \sum f(s)$ for all $S \subseteq X$, the sum taken over the elements of S which lie in the support of f.[7]

On morphisms $\mathcal{D}$ was defined so that for $g : X \to Y$ and $f \in \mathcal{D}X$, one have $(\mathcal{D}g)f = f \circ g^{-1}$.



Strictly speaking, f is – as in the definition of P_f – being extended to act on subsets of X via intersection with $\mathrm{supp}(f)$ followed by pointwise sum. Intuitively, each $x \in X$ sends its weight, defined as $f(x)$, to its image under g, so that the weight on a point $y \in Y$ come out to $f \circ g^{-1}(y)$. This is an instance of the pushforward of measures mentioned previously.

That $\mathcal{D}$ preserves identities amounts to $f \circ 1_X^{-1} = f$; that it commute with composition is a consequence of the fact that the pre-image of a composition $g \circ h$ may be taken all at once or in two steps (by pre-imaging under h and then under g). Diagrammatically, the dashed arrows below commute.



$\mu : \mathcal{D}^2 \Rightarrow \mathcal{D}$ is defined to average a probability distribution over probability distributions, as in the fair and two-headed coins. Formally, for $Q \in \mathcal{D}^2X$ a probability measure over $\mathcal{D}X$ equipped with the discrete σ-algebra, one has:

$$\mu(Q)(S) = \sum_{P \in \mathrm{supp}(Q)} Q(P) \cdot P(S)$$

Here $\mathrm{supp}(Q)$ refers to those finitely many distributions in $\mathcal{D}X$ to which Q assigns nonzero probability. In words, $\mu(Q_X)$ is defined – as a probability measure on X – to assign to $S \subseteq X$ a weight according to the following rule: a distribution P in $\mathcal{D}X$ assigns a weight of $P(S)$ to S, and is itself scaled by $Q(P)$ so as so ultimately contribute a weight of $Q(P) \cdot P(S)$ to S. Note the similarity to the process of multiplying through edge weights in (0.0.2), an instance of μ being applied to an element of $\mathcal{D}^2(\{\text{heads, tails}\})$.

The Dirac measure defines the unit $\eta : 1_{\mathsf{Set}} \Rightarrow \mathcal{D}$, with $x \in X$ mapped to δ_x and, for $S \subseteq X$:

$$\delta_x(S) = \begin{cases} 1 & x \in S \\ 0 & x \notin S \end{cases}.$$

[7]Restricting the sum is necessary in the event that S be uncountably infinite. If S is countable, then finiteness of the support of f assures convergence of the sum; alternatively, monotone converge suffices, as the sum is bounded by 1.

Naturality of η amounts to $\delta_x \circ f^{-1} = \delta_{f(x)}$.

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ {\scriptstyle\eta}\downarrow & & \downarrow{\scriptstyle\eta} \\ \mathcal{D}X & \xrightarrow{\mathcal{D}f} & \mathcal{D}Y \end{array} \quad \leftrightsquigarrow \quad \begin{array}{ccc} x & \longmapsto & f(x) \\ \downarrow & & \downarrow \\ \delta_x & \longmapsto & \delta_{f(x)} \end{array}$$

In words, a set's pre-image under f contains x exactly when the set contains $f(x)$.[8] Unitarity of η corresponds to the fact that a distribution $P \in \Pi(X)$ can be made an element of $\Pi^2(X)$ by two means, each of which average back to P under μ:

- Via $\mathcal{D}\eta$: apply η to each element $x \in X$, and assign a weight of $P(x)$ to the resulting distribution δ_x.
- Via $\eta\mathcal{D}$: Apply η to P itself, resulting in the element of $\Pi^2(X)$ which places full weight on P.

Formally, for $x \in X$,

$$\mu(\mathcal{D}\eta P)(x) = (\mathcal{D}\eta P)(\delta_x) \cdot \delta_x(x) = P(x) \cdot 1$$

and

$$\mu(\eta\mathcal{D}P)(x) = (\eta\mathcal{D}P)(P) \cdot P(x) = 1 \cdot P(x).$$

In the case $\Omega = \{\text{heads}, \text{tails}\}$, one has:

$$\begin{array}{ccc} \mathcal{D}\Omega & \xrightarrow{\eta\mathcal{D}} & \mathcal{D}^2\Omega \\ {\scriptstyle\mathcal{D}\eta}\downarrow & & \downarrow{\scriptstyle\mu} \\ \mathcal{D}^2\Omega & \xrightarrow{\mu} & \mathcal{D}\Omega \end{array} \quad \leftrightsquigarrow \quad \begin{array}{ccc} \frac{1}{2}\text{heads} + \frac{1}{2}\text{tails} & \longmapsto & \delta_{\frac{1}{2}\text{heads}+\frac{1}{2}\text{tails}} \\ \downarrow & & \downarrow \\ \frac{1}{2}\delta_{\text{heads}} + \frac{1}{2}\delta_{\text{tails}} & \longmapsto & \frac{1}{2}\text{heads} + \frac{1}{2}\text{tails} \end{array}$$

Lastly, associativity of μ amounts to the fact that an element of $\mathcal{D}^3(X)$ – a distribution over distributions over distributions – can be averaged beginning with the outermost two dinner distributions or the innermost distributions. Averaging the resulting element of $\mathcal{D}^2(X)$ yields the same distribution over Ω in either case.

Our vocabulary has expanded since our last pass at $\mathcal{D}$. In particular, we may inquire as to the structure of the Kleisli category $\mathsf{Set}_{\mathcal{D}}$ associated to $\mathcal{D}$. It is the category in which:

- Objects are sets X
- Morphisms $X \rightsquigarrow Y$ are functions $X \to \mathcal{D}Y$ assigning to each element of X a distribution over Y
- The identity $X \rightsquigarrow X$ assigns the Dirac measure δ_x to each $x \in X$
- Composition of $f : X \rightsquigarrow Y$ and $g : Y \rightsquigarrow Z$ is given by

$$g \circ_k f = \mu \circ \mathcal{D}g \circ f.$$

[8]This heuristic made use of a correspondence between the Dirac measure δ_x and the characteristic function χ_x, a sometimes useful trick.

Intuitively, morphisms in $\mathsf{Set}_{\mathcal{D}}$ play the role of transition probabilities. The output of $x \in X$ under a map $X \rightsquigarrow Y$ records the random law describing where one 'lands' in Y after leaving x. Kleisli composition then acts as usual composition of transition probabilities. For $X \rightsquigarrow Y \rightsquigarrow Z$, the composition $X \rightsquigarrow Z$ records the probability of landing in $z \in Z$ when beginning in $x \in X$ – given by the sum across $y \in Y$ of the probability of landing in y after x times the probability of landing in z after y.

We saw in 2.11, meanwhile, that the Eilenberg-Moore category $\mathsf{Set}^{\mathcal{D}}$ is isomorphic to the category of convex sets and affine functions. Though this may appear to lack the natural probabilistic interpretation of the Kleisli category and its stochastic maps, it is in fact the first sign of a profound relationship in categorical probability. That is, the algebras of probability monads tend to look convex, and convex spaces are those which - roughly speaking - admit weighted sums of their elements with coefficients which sum to 1, as witnessed in the proof of Proposition 2.5.

The conclusion is that the algebras of probability monads tend to be those spaces which permit a notion of expected value, arguably the most important operation in probability. It suggests that expectation is the ultimate well-behaved rule for evaluating a distribution over a space into an element of the space, an affirming and far-reaching result. As we proceed in the study of probability monads, we should be watchful for this relationship, one of the key discoveries of categorical probability.

4.2. The Giry and Kantorovich monads

The probability monad of greatest historical importance is unambiguously that of the Giry monad, introduced by Giry in 1982 as a pair of related monads [3]. We focus on the monad defined on Mes, the category of measurable spaces, rather than the monad on Pol, the category of Polish spaces. In each case, the monads act nearly identically, and the monad on Mes is arguably more fundamental, while capturing the essential idea.

As a brief reminder, the category Mes is inhabited by measurable spaces $(\Omega, \mathcal{F}_\Omega)$, i.e. a pair of a set Ω and an associated σ-algebra $\mathcal{F}_\Omega$. We opt to suppress a measurable space's σ-algebra, so that Ω denote the set underlying a measurable space, its σ-algebra understood to be $\mathcal{F}_\Omega$. An element of $\mathcal{F}_\Omega$ is referred to as a measurable set, and the morphisms in Mes are functions which reflect measurable sets. More explicitly, a measurable function $f : \Omega \to \Omega'$ is a function such that $f^{-1}(B') \in \mathcal{F}_\Omega$ for all $B' \in \mathcal{F}_{\Omega'}$.

Definition 4.1. The *Giry monad* (Π, η, μ) on Mes is defined as follows:

- On objects, Π assigns to Ω the set $\Pi(\Omega)$ of all probability measures on Ω. It is equipped with the coarsest σ-algebra such that the following evaluation maps are measurable for all $B \in \mathcal{F}_\Omega$

$$\begin{aligned} \Pi(\Omega) &\longrightarrow [0,1] \\ P &\longmapsto P(B) \end{aligned}$$

- On morphisms, Π acts to pushforward measures, i.e. for $f : \Omega \to \Omega'$, one has

$$\Pi(f) : \Pi(\Omega) \longrightarrow \Pi(\Omega')$$
$$P \longmapsto P \circ f^{-1}$$

Note that $P(f^{-1}(B))$ is defined for $B \in \mathcal{F}_{\Omega'}$ precisely because $f^{-1}(B) \in \mathcal{F}_\Omega$, by measurability of f.[9]

- The unit $\eta : 1_{\mathsf{Mes}} \Rightarrow \Pi$ assigns to $\omega \in \Omega$ the Dirac measure δ_ω
- The multiplication $\mu : \Pi^2 \Rightarrow \Pi$ acts by integration; for $Q \in \Pi^2(\Omega)$ and $B \in \mathcal{F}_\Omega$, $\mu(Q)$ assigns measure to B by

$$\mu(Q)(B) = \int\limits_{P \in \Pi(X)} P(B) \mathrm{d}Q$$

The integral is defined because evaluation maps are measurable out of $\Pi(\Omega)$.

This is likely a case in which a disciplined examination of the details of Π's monadicity is not worth the trouble, especially after two passes at the discrete distribution monad. Naturality and associativity of μ are the only non-trivial arguments, and follow without much difficulty from the following lemma [3].

Lemma 4.2. *The following hold for $f : \Omega \to \Omega'$, $P \in \Pi(\Omega)$, $Q \in \Pi^2(\Omega)$, and bounded $\theta' : \Omega' \to \mathbb{R}$.*

(i) $\int \theta' d\Pi(f)(P) = \int \theta' \circ f \, dP$

(ii) $\xi'_\theta : \Pi(\Omega') \to \mathbb{R}$ *defined by* $\xi_\theta(P) = \int \theta' dP$ *is measurable*

(iii) $\int \theta \, d\mu(Q) = \int \xi_\theta \, dQ$

Proof. When θ is a characteristic function χ_B, the result holds by definition. By linearity of integration, it thus holds for simple functions (i.e. $\mathbb{R}$-linear combinations of characteristic functions). By monotone convergence, it finally holds for arbitrary θ, which can be witnessed as the limit of a monotone sequence of simple functions. □

Thus, the discrete distribution monad admits an extreme generalization to the Giry monad, which places no restrictions whatsoever on the probability measures in $\Pi(\Omega)$. It is somewhat remarkable that such an assignment indeed forms a monad. What of the Kleisli and Eilenberg-Moore categories associated to Π? As with the discrete distribution monad, the Kleisli category of Π has transition probabilities as morphisms, this time without restriction on σ-algebra or support. In particular, a transition probability from Ω to Ω' is defined formally as a function $t : \Omega \times \mathcal{F}_{\Omega'} \to [0, 1]$ such that:

(i) $t(\omega, -)$ is a probability measure on Ω' for all $\omega \in \Omega$

[9] In our mind, this is the raison d'être for the definition of measurable function: one would like for a morphism between measurable spaces to extend to a transfer of measures (either from domain to codomain or vice versa). The natural options are pushforward or pullback - the pullback is hopeless, because direct images almost never commute with disjoint unions. In order for the pushforward to work, the map on underlying sets must have reflected measurable subsets, and one arrives at the definition.

(ii) $t(-, B')$ is measurable for all $B' \in \mathcal{F}_{\Omega'}$

where $t(\omega, B')$ is interpreted as recording the probability that one will land in B', having departed from ω. A morphism $f : \Omega \rightsquigarrow \Omega'$ in Mes_Π determines a transition probability from Ω to Ω' by defining $t_f(\omega, B') := f(\omega)(B')$. Then

(i) $t_f(\omega, -) = f(x)(-)$ is a probability measure on Ω' precisely because $f : \Omega \rightsquigarrow \Omega'$ corresponds to a morphism $\Omega \to \Pi(\Omega')$

(ii) $t_f(-, B') = f(-)(B')$ is measurable because it is the composition of f with the function $\Pi(\Omega') \to [0, 1]$ which evaluates a measure at B'; f is measurable by assumption and the evaluation map by design of the σ-algebra on $\Pi(\Omega')$

In fact, the Kleisli morphisms of Mes_Π attain all transition probabilities in this manner, and the compositions in each setting coincide. An immediate consequence of associativity of composition in Kleisli categories is thus that composition of transition probabilities associates, a result which would otherwise be non-trivial [3].

The Eilenberg-Moore category Mes^Π, however, is more difficult to study, as it involves the task of classifying the Giry monad's algebras. A recent paper claimed to have demonstrated an equivalence between Mes^Π and the category of convex measurable spaces but has since been proven incorrect [12, 1]. As of yet, such a result escapes categorical probabilists. Nevertheless, the Giry monad has been the subject of active mathematical research and, furthermore, found application in such areas as theoretical machine learning and functional programming [13].

The Kantorovich monad belongs to the younger generation of probability monads and, though less approachable than the Giry monad, offers an example of a probability monad on a category endowed with more structure than Mes [9]. In particular, the Kantorovich monad acts on CMet, the category of complete metric spaces and short, or 1-Lipschitz, maps.

Definition 4.3. The *Kantorovich monad* P on CMet is as follows:

- For $X \in \mathsf{CMet}$, PX is a subset of probability measures on X made a complete metric space under the *Wasserstein* or *earth mover's* distance, i.e.

$$d_{PX}(p, q) := \inf_{r \in \Gamma(p,q)} \int_{X \times X} d_X(x, y) dr(x, y)$$

  where $\Gamma(p, q)$ denotes probability measures on $X \times X$ with marginals p and q

- On morphisms, P acts via the pushforward
- The Dirac embedding serves as the unit η, and taking the expected distribution of a distribution over distributions serves as the multiplication μ, as in the Giry monad.

For the sake of simplicity, we suppress the conditions imposed on the measures in PX; roughly speaking, the measures are required to be well-behaved with respect to the metric on X and to have expected values with respect to Lipschitz functions. The restriction to short maps in CMet, meanwhile, ensures that the measures of PX have finite distance between one another.

Consider now the Wasserstein or earth mover's distance, used to endow PX with the structure of a metric space. For $p, q \in PX$, their distance carries the interpretation of the work needed to move the mass of one distribution to that of the other. In the integrand, $dr(x, y)$ can be seen to measure the amount of mass being interchanged between p and q, and $d_X(x, y)$ the distance it must travel. Perhaps more intuitively, for p and q thought of as piles of sand over X, the earth mover's distance records the work needed to exchange the grains constituting each pile, explaining its name.

Critically, the earth mover's distance makes use of the metric on X, thereby recording information as to its topology when evaluated on measures over X. Contrast with, for instance, the total variation distance of probability measures

$$\delta(p, q) = \sup_{B \in \mathcal{F}_X} |p(B) - q(B)|$$

which disregards the topology on X. We conclude with two remarks on the Kantorovich monad:

(i) It admits a purely formal characterization of its multiplication, allowing one to make use of integration without integrating. This presents advantages primarily in infinite-dimensional settings.

(ii) It enjoys a classification of its algebras as closed convex subsets of Banach spaces. In particular, the relationship between the algebras of a probability monad and expectation is once again witnessed, by way of convexity.

REFERENCES

[1] Crhak, Tomas. "On functors from category of Giry algebras to category of convex spaces." arXiv preprint arXiv:1804.01345 (2018).

[2] Etingof, Pavel, et al. *Tensor categories.* Vol. 205. American Mathematical Soc., 2016.

[3] Giry, Michele. "A categorical approach to probability theory." Categorical aspects of topology and analysis. Springer, Berlin, Heidelberg, 1982. 68-85.

[4] Iry, James. “A Brief, Incomplete, and Mostly Wrong History of Programming Languages.” 7 May 2009, http://james-iry.blogspot.com/2009/05/brief-incomplete-and-mostly-wrong.html.

[5] Jacobs, Bart. "Convexity, duality and effects." IFIP International Conference on Theoretical Computer Science. Springer, Berlin, Heidelberg, 2010.

[6] Jacobs, Bart. "From probability monads to commutative effectuses." Journal of logical and algebraic methods in programming 94 (2018): 200-237.

[7] Lawvere, Francis William. “The Category of Probabilistic Mappings.” 1962, https://ncatlab.org/nlab/files/lawvereprobability1962.pdf.

[8] MacLane, Saunders. "Natural associativity and commutativity." Rice Institute Pamphlet-Rice University Studies 49.4 (1963).

[9] Perrone, Paolo. "Categorical Probability and Stochastic Dominance in Metric Spaces." (2018).

[10] Riehl, Emily. *Category theory in context.* Courier Dover Publications, 2017.

[11] Sankar, Anirudh. “Monads and Algebraic Structures.” 1 Sept. 2012, http://math.uchicago.edu/ may/REU2012/REUPapers/Sankar.pdf.

[12] Sturtz, Kirk. "The factorization of the Giry monad." Advances in Mathematics 340 (2018): 76-105.

[13] Tassarotti, Joseph, Jean-Baptiste Tristan, and Koundinya Vajjha. "A Formal Proof of PAC Learnability for Decision Stumps." *arXiv preprint arXiv:1911.00385* (2019).

[14] Trimble, Todd. “Notes on Tetracategories.” Oct. 2006, http://math.ucr.edu/home/baez/trimble/tetracategories.html.