10

Quantitative Equational Reasoning

Giorgio Bacci
Aalborg University
Radu Mardare
University of Stratchclyde
Prakash Panangaden
McGill University
Gordon Plotkin
University of Edinburgh

Abstract: Equational logic is central to reasoning about programs. What is the right equational setting for reasoning about probabilistic programs? It has been understood that instead of equivalence relations one should work with (pseudo)metrics in a probabilistic setting. However, it is not clear how this relates to equational reasoning. In recent work the notion of a *quantitative equational logic* was introduced and developed. This retains many of the features of ordinary logic but fits naturally with metric reasoning. The present chapter is an elementry introduction to this topic. In this setting one can define analogues of algebras and free algebras. It turns out that the Kantorovich (Wasserstein) metric emerges as a free construction from a simple quantitative equational theory. We give a couple of examples of quantitative analogues of familiar effects from programming language theory. We do not assume any background in equational logic or advanced category theory.

10.1 Introduction

Equational reasoning is at the heart of mathematics and theoretical computer science. In algebra, we define algebraic structures by giving (mostly) equational axioms. In analysis there are numerous equations linking concepts; of course, inequalities play a major role too but this just highlights the importance of equality. In programming language semantics one has equations capturing notions of behavioural equivalence of programs. The monadic approach, due to Moggi (1991), to incorporating effects in higher-order functional programming has been understood through the work of

[a] From *Foundations of Probabilistic Programming*, edited by Gilles Barthe, Joost-Pieter Katoen and Alexandra Silva published 2020 by Cambridge University Press.

Plotkin and Power (2004), Hyland et al. (2006), and Hyland and Power (2007) and others in terms of operations and equations.

With the emergence of probabilistic programming[1] a new emphasis on quantitative reasoning has become important. One thinks in terms of "how close are two programs?" rather than "are they completely indistinguishable?" This concept is captured by a *metric* and was first advocated in Giacalone et al. (1990). The idea here was that instead of using a behavioural equivalence relation like bisimulation one should use a pseudometric whose kernel is bisimulation. Such a metric was first defined in Desharnais et al. (1999). What we aim to do here is show how a version of equational reasoning, which we call *quantitative equational logic*, captures such metric reasoning principles. This work first appeared in Mardare et al. (2016, 2017) and Bacci et al. (2018).

The most compelling example of a programming language setting where quantitative reasoning is important is probabilistic programming; the subject of this book. While our work is not specifically adapted to this setting it does provide the general framework for such reasoning. In particular, one of the most important ways of comparing probability distributions is the Kantorovich (Wasserstein) metric and, for example, in machine learning it has recently been a source of much attention. In our quantitative equational framework this metric emerges naturally from simple quantitative equations.

The key idea is to introduce equations indexed by positive rational numbers:

$$s =_{\varepsilon} t$$

where s and t are terms of some language and ε is a (presumably small) positive real number. One reads this as "s is within ε of t". Certainly, the relation $=_{\varepsilon}$ is not an equivalence relation: transitivity does not hold, if $s =_{\varepsilon} t$ and $t =_{\varepsilon} u$ then there is no reason to think $s =_{\varepsilon} u$. Indeed, one can only say $s =_{2\varepsilon} u$.

In the usual notion of equational reasoning one has a trinity of ideas: equations, Lawvere theories, and monads on **Set**. The equational presentation of algebras was systematically worked out by universal algebraists. Lawvere showed how to give a cateorical presentation of algebraic theories which freed the subject from some of the awkwardness of dealing with different presentations of the same theory. In the 1950s it was understood that algebras arose as the "algebras of a monad" defined on the category **Set**. Essentially, the action of the monad is to construct the free algebras.

These concepts can be generalized to other settings, see, for example, Robinson (2002). In the present work we have quantitative equations and it turns out that one can get monads on **Met**, for some suitable category of metric spaces.

[1] To a lesser extent, real-time programming as well.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

10.2 Equational Logic

In this section we review the standard familiar concepts of equational logic; where we mean equations in the usual sense of the word. The equality concept is one of the oldest abstract mathematical ideas and is well understood intuitively.

The basic syntax of equational logic starts with a *signature* of symbols

$$\Omega = \{f_j : \kappa_i \mid i \in I\},$$

consisting of a set of *function symbols* (or *operations*) f_i each having associated with it a cardinal number (finite or infinite) κ_i called its *arity*. The arity specifies how many arguments the function symbol takes. Some function symbols may have arity 0; these are the *constants* of the language. We have not restricted the collection of operations to be finite or countable, and one of our main examples will indeed have uncountably many operations. It is possible to consider operations which take infinitely many arguments and we will consider such an example later.

Terms are constructed inductively starting from a fixed countable set X of *variables*, ranged over by $x, y, z, \ldots$. Then the function symbols are applied to the appropriate number of previously constructed terms to give new terms. We can succinctly express the collection of terms through the following grammar:

$$t ::= x \mid f(t_i)_{i\in\kappa}, \qquad \text{for } x \in X \text{ and } f : \kappa \in \Omega$$

The set of terms constructed this way is denoted by $\mathbf{T}_\Omega X$. When the signature of operation symbols Ω is clear from from the context, the set of terms will be simply denoted as $\mathbf{T}X$.

A *substitution* is a function $\sigma : X \to \mathbf{T}X$: it defines what it means to substitute a variable for a term. It can be (homomorphically) extended to a function $\tilde{\sigma} : \mathbf{T}X \to \mathbf{T}X$ over terms as follows:

$$\begin{aligned} \tilde{\sigma}(x) &= \sigma(x) && \text{for } x \in X, \\ \tilde{\sigma}(f(t_i)_{i\in\kappa}) &= f(\tilde{\sigma}(t_i))_{i\in\kappa} && \text{for } f : \kappa \in \Omega. \end{aligned}$$

In what follows we won't make any distinction between the substitution and its extension. We will denote by $\Sigma(X)$ the set of substitutions on $\mathbf{T}X$.

The basic formulas of equational logic are *equations* of the form

$$s = t, \qquad \text{for } s, t \in \mathbf{T}X.$$

There are no quantifiers or logical connectives. We use $\mathcal{E}(\mathbf{T}X)$ to denote the set of equations over $\mathbf{T}X$. Conjunction is implicit when one writes a sequence of equations, but there is no disjunction, nor negation or implication. A *judgement* is an expression of the form

$$\Gamma \vdash \phi,$$

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

where $\Gamma \subseteq \mathcal{E}(\mathbf{T}X)$ is an enumerable set of equations and $\phi \in \mathcal{E}(\mathbf{T}X)$. The judgment $\Gamma \vdash \phi$ is intended to mean that under the assumptions in Γ the equation ϕ holds. We refer to the elements of Γ as the *hypotheses* and to ϕ as the *conclusion* of the judgment; $\mathcal{J}(\mathbf{T}X)$ denotes the collection of judgments on $\mathbf{T}X$.

Judgments are used for reasoning; we now define the important concept of equational theory.

An *equational theory of type* Ω *over* X is a set $\mathcal{U}$ of judgements on $\mathbf{T}X$ such that, for arbitrary $s, t, u \in \mathbf{T}X$ and $\Gamma, \Theta \subseteq \mathcal{E}(\mathbf{T}X)$

$$
\begin{aligned}
&(\mathbf{Refl}) \quad \emptyset \vdash t = t \in \mathcal{U}\,,\\
&(\mathbf{Symm}) \quad \{s = t\} \vdash t = s \in \mathcal{U}\,,\\
&(\mathbf{Trans}) \quad \{s = u, u = t\} \vdash s = t \in \mathcal{U}\,,\\
&(\mathbf{Cong}) \quad \{s_i = t_i \mid i \in \kappa\} \vdash f(s_i)_{i\in\kappa} = f(t_i)_{i\in\kappa} \in \mathcal{U}\,, \quad \text{for any } f\colon \kappa \in \Omega\,,\\
&(\mathbf{Subst}) \quad \text{if } \Gamma \vdash s = t \in \mathcal{U}, \text{ then } \sigma(\Gamma) \vdash \sigma(s) = \sigma(t) \in \mathcal{U}, \text{ for any } \sigma \in \Sigma(X)\,,\\
&(\mathbf{Cut}) \quad \text{if } \Theta \vdash \Gamma \in \mathcal{U} \text{ and } \Theta \vdash s = t \in \mathcal{U}, \text{ then } \Gamma \vdash s = t \in \mathcal{U}\,,\\
&(\mathbf{Assum}) \quad \text{if } s = t \in \Gamma, \text{ then } \Gamma \vdash s = t \in \mathcal{U}\,,
\end{aligned}
$$

where we write $\Gamma \vdash \Theta \in \mathcal{U}$ to mean that $\Gamma \vdash \phi \in \mathcal{U}$ holds for all $\phi \in \Theta$; and $\sigma(\Gamma) = \{\sigma(s) = \sigma(t) \mid s = t \in \Gamma\}$.

The rules (**Refl**), (**Symm**), and (**Trans**) capture the idea that equality is indeed an equivalence relation. The congruence rule (**Cong**) describes how equality interacts with the term-forming operations of the underlying term language. Finally, the substitution rule (**Subst**) states that substitution preserve equality, while (**Cut**) and (**Assum**) are the usual cut and assumption rules of logical reasoning.

A trivial consequence of the cut rule is that $\emptyset \vdash s = t \in \mathcal{U}$ implies that $\Gamma \vdash s = t \in \mathcal{U}$, for any set of equations Γ. In other words, whatever can be proven in a theory $\mathcal{U}$ without using any hypothesis, can also be proven from any set of hypothesis. This is the familiar *weakening* rule.

Given an equational theory $\mathcal{U}$ and a set $S \subseteq \mathcal{U}$, we say that S is a set of axioms for $\mathcal{U}$, or S axiomatizes $\mathcal{U}$, if $\mathcal{U}$ is the smallest equational theory that contains S. An equational theory $\mathcal{U}$ is *inconsistent* if $\emptyset \vdash x = y \in \mathcal{U}$ for two distinct variables $x, y \in X$; $\mathcal{U}$ is *consistent* if it is not inconsistent. From the substitution rule, inconsistency implies that *every* equation is derivable.

10.2.1 Algebra

Equational logic is intimately tied to algebra. For most of the familiar algebraic structures one sees, the basic definition is given in terms of equations (although, there are a few notable exceptions). To describe the equations characterising algebraic structures like a group or monoid, e.g., the associativity equation, one starts with a set of variables X and signature of operations. This gives the *term algebra* over which the equational properties are described. For example, for monoids one can use the signature

$$\{e\colon 0, \quad \cdot\colon 2\}$$

consisting of a 0-arity function symbol e (*i.e.*, a constant) for the identity element and an 2-arity function symbol "$\cdot$" (typically used as an infix operator). The terms for this language look like $x \cdot (y \cdot z)$, $e \cdot x, \ldots$ [2].

The properties are spelled out as equations. For monoids, these are

$$e \cdot x = x\,, \qquad x \cdot e = x\,, \qquad (x \cdot y) \cdot z = x \cdot (y \cdot z)\,,$$

for $x, y, z \in X$ variables.

Given a Ω-*algebra*, *i.e.*, an algebraic structure over the signature Ω, such as a monoid, we need to explain what it means for it to *satisfy* an equation, or more generally a judgement.

A particular instance of an Ω-algebra $\mathcal{A} = (A, \Omega_{\mathcal{A}})$ consists of a set A, called the *carrier*, containing the elements of the algebra, and a collection $\Omega_{\mathcal{A}}$ of *interpretations* for each function symbol in Ω. If $f\colon a \in \Omega$ is a function symbol of arity κ, then its interpretation is a function $f_{\mathcal{A}}\colon A^{\kappa} \longrightarrow A$, where A^{κ} is the κ-fold cartesian product of A; for a constant symbol this corresponds to select a designated element of A. Thus a particular monoid $\mathcal{M}$ will be described by giving a set M of its elements, a designated element $e_{\mathcal{M}} \in M$ to stand for e, and a binary operation $\cdot_{\mathcal{M}}\colon M \times M \longrightarrow M$.

Given the notion of algebra we can define a *subalgebra*. A subalgebra of $\mathcal{A}$ is another algebra with the same signature and whose elements form a subset of A. Given two algebras $\mathcal{A} = (A, \Omega_{\mathcal{A}})$, $\mathcal{B} = (B, \Omega_{\mathcal{B}})$ of the same signature, we can define a *homomorphism* h to be a set-theoretic function from A to B, which preserves the operations of the signature:

$$h(f_{\mathcal{A}}(a_i)_{i\in\kappa}) = f_{\mathcal{B}}(h(a_i))_{i\in\kappa}\,, \qquad \text{for all } f\colon \kappa \in \Omega\,,$$

where the equality symbol appearing in this equation means **identity** between elements in B.

If we fix a set of variables X and a signature Ω, the *term algebra* has the set of terms $\mathbf{T}X$ build over X as carrier and interpretation for the function symbols $f\colon \kappa \in \Omega$ canonically given by

$$(t_i)_{i\in\kappa} \in (\mathbf{T}X)^{\kappa} \mapsto f(t_i)_{i\in\kappa} \in \mathbf{T}X\,.$$

We denote as $\mathbf{T}X$ for this structure as well as its the underlying set of terms. Consider a Ω-algebra $\mathcal{A} = (A, \Omega_{\mathcal{A}})$ and a assignment function $\iota\colon X \longrightarrow A$ from X to A,

[2] Actually, it is even more common to leave it out altogether and indicate the operation by mere juxtaposition.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

interpreting variables as elements in $\mathcal{A}$; this extends inductively to a function, also written ι, from $\mathbf{T}X$ to A:

$$\begin{aligned} \iota(x) &= \sigma(x) && \text{for } x \in X\,, \\ \iota(f(t_i)_{i\in\kappa}) &= f_{\mathcal{A}}(\iota(t_i))_{i\in\kappa} && \text{for } f\colon \kappa \in \Omega\,. \end{aligned}$$

It is immediate from this definition that ι is a homomorphism from $\mathbf{T}X$ to $\mathcal{A}$. It should also be clear that *every* homomorphism from $\mathbf{T}X$ to $\mathcal{A}$ arises in this way; we just have to restrict the homomorphism to X.

Definition 10.1. We say that a judgement $\Gamma \vdash s = t$ in $\mathcal{J}(\mathbf{T}X)$ is **satisfied** by $\mathcal{A}$, written $\mathcal{A} \models (\Gamma \vdash s = t)$, if for every assignment $\iota\colon X \to A$ the following implication holds:

$$\big(\text{for all } (s' = t') \in \Gamma,\ \iota(s') = \iota(t')\big) \text{ implies } \iota(s) = \iota(t)\,.$$

The "term algebra" that we have defined so far is not really a proper algebra of the type we have in mind because it does not satisfy the equations that define the class of algebraic structures. We now repair this by an appropriate quotient construction.

Let S be a set of judgements and $\mathcal{U}_S$ the smallest equational theory that contains S. We define a relation between terms s, t by

$$s \sim_S t\,, \text{ if } (\emptyset \vdash s = t) \in \mathcal{U}_S\,.$$

Since $\mathcal{U}_S$ is closed under the rules of equational logic, in other words, it is closed under reflexivity, symmetry, transitivity, substitution, and congruence; this gives us a congruence[3] relation on $\mathbf{T}X$. We write $[t]$ for the equivalence class of t with respect to the congruence $\sim_S$. The quotient set $\mathbf{T}X/_{\sim_S}$ is the collection of such equivalence classes. It will be the underlying set of the term algebra. We define an interpretation, written f_S, for an operator $f\colon \kappa \in \Omega$ of the signature on $\mathbf{T}X/_{\sim_S}$ as follows:

$$f_S([t_i])_{i\in\kappa} = [f(t_i)_{i\in\kappa}]\,.$$

This is well defined precisely because $\sim_S$ is a congruence. The set $\mathbf{T}X/_{\sim_S}$ is now a Ω-algebra; we denote it by $\mathbf{T}_S[X]$, or simply $\mathbf{T}[X]$ when S is clear from the context. It should be clear that by its construction, the algebra $\mathbf{T}_S[X]$ satisfies all the judgements in S (and $\mathcal{U}_S$).

Examples of familiar algebras that can be presented purely equationally are semigroups, monoids, groups, rings, lattices, and boolean algebras. Vector spaces have two sorts of elements, but the theory described above can readily be extended to this case and thus we include vector spaces as equationally defined algebras. Stacks as used in computer science are another familiar example. Some algebraic structures require a strictly more powerful construct, namely *Horn clauses* in

[3] We always implicitly include the notion of equivalence relation when we say "congruence."

their axiomatizations. These are judgements with a nonempty set of hypothesis representing a condition that must hold. For example, right-cancellative monoids are required to satisfy the judgement

$$\{x \cdot z = y \cdot z\} \vdash x = y\,.$$

A familiar example of a right-cancellative monoid is the set of words over an alphabet with the operation being concatenation of words.

Fields are not an example of equationally defined class of algebras. One of the field axioms says: "if $x \neq 0$ then there exist an element x^{-1} such that $x \cdot x^{-1} = x^{-1} \cdot x = 1$". Here 1 is the multiplicative identity element. It is clearly not an equation because of the side condition. It is not obvious to see that one cannot replace this with a bona-fide equation.

A very interesting example of an algebra that we will extensively discuss in the quantitative setting are *barycentric algebras*. The signature of barycentric algebras has uncountably many binary operations

$$\{+_e : 2 \mid e \in [0,1]\}$$

satisfying the following equations, due to Stone,

$$\begin{aligned}
&(\mathbf{B1}) && \emptyset \vdash x +_1 y = x\,,\\
&(\mathbf{B2}) && \emptyset \vdash x +_e x = x\,,\\
&(\mathbf{SC}) && \emptyset \vdash x +_e y = y +_{1-e} x\,,\\
&(\mathbf{SA}) && \emptyset \vdash (x +_{e_1} y) +_{e_2} z = x +_{e_1 e_2} (y +_{\frac{e_2 - e_1 e_2}{1 - e_1 e_2}} z)\,, \qquad \text{for } e_1, e_2 \in (0,1)\,,
\end{aligned}$$

axiomatizing the notion of convex combination of a pair of elements. Any convex subset of a real vector space satisfies these axioms. Barycentric algebras can be axiomatized in other ways. For example, instead of binary convex combinations one can introduce n-ary convex combinations for all $n \in \mathbb{N}$. One of the most important examples of a barycentric algebra is the set of probability measures on a finite set or indeed on more complicated spaces. In Section 10.3.1 we review some probability theory on metric spaces.

10.2.2 General Results

Given an equational theory $\mathcal{U}$ and an algebra $\mathcal{A}$ over the same signature, we write $\mathcal{A} \models \mathcal{U}$ to mean that $\mathcal{A}$ satisfies all the judgements in $\mathcal{U}$. As is usual in model theory, we write $\mathcal{U} \models (\Gamma \vdash \phi)$ to mean that the judgement $\Gamma \vdash \phi$ is satisfied by any algebra $\mathcal{A}$ for which $\mathcal{A} \models \mathcal{U}$ holds.

The celebrated Birkhoff completeness theorem relates the semantic notion of

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

satisfiability to deducibility (see, for example p. 95 of Burris and Sankappanavar, 1981).

Theorem 10.2 (Completeness)**.** *$\mathcal{U} \models (\Gamma \vdash \phi)$, if and only if, $(\Gamma \vdash \phi) \in \mathcal{U}$.*

The proof is by construction of a suitable universal model: the algebra $\mathbf{T}_{\mathcal{U}}[X]$ over the quotient $\mathbf{T}X/_{\sim_{\mathcal{U}}}$ introduced in Section 10.2.1. We will see that an analogous result holds in the quantitative case.

Let $\mathbf{K}(\Omega, \mathcal{U})$ denote the collection of Ω-algebras satisfying all the judgements in $\mathcal{U}$, *i.e.*, $\mathcal{A} \in \mathbf{K}(\Omega, \mathcal{U})$ if $\mathcal{A} \models \mathcal{U}$; $\mathbf{K}(\Omega, \mathcal{U})$ becomes a category if we take the morphisms to be Ω-homomorphisms. If we don't need to emphasize which signature we use will simply write $\mathbf{K}(\mathcal{U})$.

Let X be a set of variables. We have seen that $\mathbf{T}_{\mathcal{U}}[X]$ is an algebra in $\mathbf{K}(\mathcal{U})$. There is a map $\eta_X : X \to \mathbf{T}_{\mathcal{U}}[X]$ given by $\eta_X(x) = [x]$, which is *universal* in the following sense:

$$\text{in } \mathbf{Set} \qquad\qquad \text{in } \mathbf{K}(\mathcal{U})$$

$$\begin{tikzcd} X \arrow[r, "\eta_X"] \arrow[dr, "\alpha"'] & \mathbf{T}_{\mathcal{U}}[X] \arrow[d, "h"] \\ & A \end{tikzcd} \qquad\qquad \begin{tikzcd} \mathbf{T}_{\mathcal{U}}[X] \arrow[d, dashed, "h"] \\ \mathcal{A} \end{tikzcd}$$

for any algebra $\mathcal{A} \in \mathbf{K}(\mathcal{U})$ and function α from X to the underlying set A of $\mathcal{A}$, there exists a *unique* algebra homomorphism $h \colon \mathbf{T}_{\mathcal{U}}[X] \to \mathcal{A}$ such that $h \circ \eta_X = \alpha$. In other words any set theoretic function can be uniquely extended to an algebra homomorphism. This makes $\mathbf{T}_{\mathcal{U}}[X]$ the *free algebra* in $\mathbf{K}(\mathcal{U})$ generated from X.

The construction of $\mathbf{T}_{\mathcal{U}}[X]$ from the set X is functorial and such a functor is left adjoint to the forgetful functor from $\mathbf{K}(\mathcal{U})$ to $\mathbf{Set}$. As usual with an adjunction, one gets a monad: the *term monad* $(\mathbf{T}_{\mathcal{U}}, \eta, \mu)$ on the category $\mathbf{Set}$, with unit $\eta \colon Id \Rightarrow \mathbf{T}_{\mathcal{U}}$ assigning a variable x to the equivalence class $[x]$ of terms provably equal in $\mathcal{U}$, and multiplication $\mu \colon \mathbf{T}^2_{\mathcal{U}} \Rightarrow \mathbf{T}_{\mathcal{U}}$ expressing term composition up to provable equivalence in $\mathcal{U}$.

Moreover, the (Eilenberg-Moore) algebras for the monad $\mathbf{T}_{\mathcal{U}}$ are in one-to-one correspondence with the algebras in $\mathbf{K}(\mathcal{U})$; actually this correspondence is and isomorphism of categories.

The collection of algebras defined by a set of equations is called a *variety of algebras*[4]. A famous theorem, also due to Birkhoff (1935) gives conditions under which a collection of algebras can be a variety.

Theorem 10.3. *A collection of algebras is a variety of algebras if and only if it is closed under homomorphic images, subalgebras, and products.*

[4] Please do not confuse this with the notion of algebraic variety which means something completely different.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

There are analogous results for algebras defined by Horn clauses: these are called quasi-variety theorems. Consider $\mathbb{Z}_2 \times \mathbb{Z}_2$. It is not a field because, *e.g.* $(1, 0) \times (0, 1) = (0, 0)$; fields are not supposed to have zero-divisors. Hence fields cannot be described by equations.

There is a quantitative analogue of this theorem (see Mardare et al., 2017) but we will not discuss it in this article as it is rather more technical than is appropriate for the present chapter.

10.3 Background

We assume that basic concepts of metric and topological spaces are well-known to the reader.

Definition 10.4. A**pseudometric** on a set X is a function $d: X \times X \to [0, \infty)$ satisfying

$$\begin{aligned}
\forall x \in X, &\qquad d(x, x) = 0, \\
\forall x, y \in X, &\qquad d(x, y) = d(y, x), \\
\forall x, y, z \in X, &\qquad d(x, y) \leq d(x, z) + d(z, y).
\end{aligned}$$

Note that we do *not* require $d(x, y) = 0$ implies that $x = y$; if we impose this condition we get what is usually called a *metric*. In a pseudometric one can have distinct points at 0 distance. The relation of being at zero distance is easily seen to be an equivalence relation called the *kernel* of the pseudometric. If we take the quotient of the underlying space by the kernel there is a natural metric defined on the equivalence classes which will satisfy the additional axiom above. The concepts of induced topology, convergence, continuity, completeness all work equally well with pseudometrics as with metrics.

Let (X, d_X), (Y, d_Y) be two (pseudo)metric spaces. A function f from to X to Y is *non-expansive* if for all $x, x' \in X$, $d_X(x, x') \geq d_Y(f(x), f(x'))$.

Metric spaces with the non-expansive maps between them form a category, usually called **Met**. Although **Met** has finite products and, more generally, finite limits, it does not have countable products nor binary coproducts. A simple way to recover completeness of the category, is to work with *extended* metric spaces: these are spaces where the metric may take on infinite values.

We define **EMet** to be the category where the objects are extended metric spaces and the morphisms are non-expansiveness maps. In **EMet** the product of a collection of spaces, $\{(X_i, d_i)\}_{i \in I}$ as the cartesian product of the individual spaces $\prod_{i \in I} X_i$ and the metric between two points $(x_i)_{i \in I}, (y_i)_{i \in I}$ is $\sup_{i \in I} d_i(x_i, y_i)$. This supremum may, of course, be infinite; this is fine in **EMet** but not in **Met**. Coproducts is **EMet** are defined by taking the coproduct as sets, *i.e.*, the disjoint union; the distance between

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

two points in the same component is just whatever it is in the original space, while the distance between two points in different components is ∞. Whenever one has an extended metric space (X, d), one can define an equivalence relation $\sim$ by $x \sim y$ iff $d(x, y) < \infty$. The equivalence classes are called components and in fact the original space is just the coproduct of the components. One can extend many standard results about ordinary metric spaces by using this decomposition. However, some things require extra caution: for example, Banach's fixed point theorem requires some care.

10.3.1 Measure Theory

Excellent sources for background on measure theory and probability are Billingsley (1995) or Dudley (1989). A quicker introduction is in Panangaden (2009).

It is a fact that on many familiar spaces, like $\mathbb{R}$, one cannot define a sensible measure on all the sets. For example, on the real line one would like a measure that that co-incides with the concept of length on intervals; but there is no such measure defined on *all* subsets of the real line. Accordingly, we have to choose "nice" families of sets on which one can hope to do measure theory properly. Being able to take countable unions and sums is the key.

Definition 10.5. A **σ-algebra** on a set X is a family of subsets of X which includes X itself and which is closed under complementation and countable unions.

A set equipped with a σ-algebra is called a *measurable space*. Given a topological space, we can define the σ-algebra generated by the open sets (or, equivalently, by the closed sets). Here, when we say that a σ-algebra is *generated* by some family of sets, say $\mathcal{F}$, we mean the smallest σ-algebra containing $\mathcal{F}$; which always exists and is unique. When the σ-algebra is generated by a topology it is usually referred to as *Borel σ-algebra*.

Definition 10.6. Given a σ-algebra (X, Σ), a **(subprobability) measure** on X is a ($[0, 1]$-valued) $[0, \infty]$-valued set function, μ, defined on Σ such that

- $\mu(\emptyset) = 0$,
- for a countable collection of pairwise disjoint sets, $\{A_i \mid i \in I\}$, in Σ, we require

$$\mu(\bigcup_{i \in I} A_i) = \sum_{i \in I} \mu(A_i).$$

In addition, for probability measures we require $\mu(X) = 1$, while for subprobability measures we require $\mu(X) \leq 1$.

There is a unique measure that one can construct defined on the Borel algebra of the real line which coincides with the notion of length of an interval: this is called *Lebesgue* measure.

It is worth clarifying how the word "measurable" is used in the literature. Given a σ-field Σ on a set X one says "measurable set" for a member of Σ. Suppose that one has a measure μ. One can have the following situation. There can be sets of measure zero which contain non-measurable subsets. Because these sets are not measurable one cannot say that they have measure zero. This happens with Lebesgue measure on the Borel sets in the real line, for example. There is a "completion" procedure[5] which produces a larger σ-algebra and an extension of the original measure in such a way that all subsets of sets of measure zero are measurable and have measure zero. The completion works by adding to the σ-algebra all sets X such that there exist Y, Z measurable sets with $Y \subseteq X \subseteq Z$ and with Y and Z having the same measure. When applied to the Borel subsets of the real line we get a much bigger σ-algebra called the Lebesgue measurable sets. One often uses the phrase "measurable set" to mean a set which belongs to the completed σ-field rather than the original σ-field.

Definition 10.7. A function $f\colon (X, \Sigma_X) \to (Y, \Sigma_Y)$ between measurable spaces is said to be **measurable** if $\forall B \in \Sigma_Y.\, f^{-1}(B) \in \Sigma_X$.

A very important class of spaces are the ones that come from metrics.

Definition 10.8. A **Polish** space is the topological space underlying a complete, separable metric space; *i.e.*, it has a countable dense subset.

Note that completeness is a metric concept but being Polish is a topological concept. A space like $(0, 1)$ is not complete in its usual metric, however, it is homeomorphic to the whole real line which is complete in its usual metric; thus, $(0, 1)$ is a Polish space.

10.3.2 The Giry Monad

A very important monad that arises in probabilistic semantics is the Giry monad described in Giry (1981). The idea was originally due to Lawvere (1962), who described a category of probabilistic mappings. Later Giry described the monad from which Lawvere's category emerges as the Kleisli category.

The underlying category is **Mes**: the objects are measurable spaces (X, Σ) and the morphisms $f\colon (X, \Sigma) \to (Y, \Lambda)$ are measurable functions. The monad is an endofunctor: $\mathcal{G}\colon \mathbf{Mes} \to \mathbf{Mes}$. The explicit definition is, on objects

$$\mathcal{G}(X, \Sigma) = \{p \mid p \text{ is a probability measure on } \Sigma\}.$$

We need to equip the set $\mathcal{G}(X, \Sigma)$ with a σ-algebra structure. For each $A \in \Sigma$, define $e_A\colon \mathcal{G}(X, \Sigma) \to [0, 1]$ by $e_A(p) = p(A)$. We equip $\mathcal{G}(X, \Sigma)$ with the smallest σ-algebra making all the e_A measurable.

[5] This is an unfortunate name because it gives the mistaken impression that the result cannot be further extended.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

The action of $\mathcal{G}$ on morphisms $f\colon (X,\Sigma) \to (Y,\Lambda)$ is given by

$$\mathcal{G}(f)\colon \mathcal{G}(X,\Sigma) \to \mathcal{G}(Y,\Lambda) \ : \qquad \mathcal{G}(f)(p)(B \in \Lambda) = p(f^{-1}(B))\,.$$

Here p is a probability measure on (X,Σ). The effect of the functor is to push forward measures through the measurable function f.

Now we can define the monad structure as follows: $\eta_X\colon X \to \mathcal{G}(X)$ is given by $\eta_X(x) = \delta_x$, where $\delta_x(A) = 1$ if $x \in A$ and 0 if $x \notin A$. The monad multiplication is

$$\mu_X(Q \in \mathcal{G}^2(X))(A) = \int e_A \mathrm{d}Q\,.$$

One can think of this as an "averaging" over all the measures in $\mathcal{G}(X)$ where we use Q as the weight for the averaging process.

Now we can present the Kleisli category as follows. The objects are the same as **Mes**; the morphisms from (X,Σ) to (Y,Λ) are measurable functions from (X,Σ) to $\mathcal{G}(Y,\Lambda)$. Kleisli composition of $h\colon X \to \mathcal{G}(Y)$ and $k\colon Y \to \mathcal{G}(Z)$ is given by the formula:

$$(k \,\tilde{\circ}\, h) = \mu_Z \circ \mathcal{G}(k) \circ h\,,$$

where $\tilde{\circ}$ denotes the Kleisli composition and $\circ$ is composition in **Mes**.

If we curry the definition of Kleisli morphism from (X,Σ) to (Y,Λ) as $h\colon X\times\Lambda \to [0,1]$ we get what are called *Markov kernels*. We call this category **Ker**. One can think of these as the probabilistic analogue of relations just as ordinary relations are the Kleisli category of the powerset monad. Kleisli composition can be written in terms of kernels:

$$(k \,\tilde{\circ}\, h)(x,C) = \int_{y\in Y} k(y,C)\mathrm{d}h(x,\cdot)\,,$$

for $x \in X$ and C a measurable subset in Z. In this form the analogy with relational composition is much clearer. One can also see this as the analogue of matrix multiplication; if the spaces were finite sets the kernels would be matrices and this composition formula would be matrix multiplication.

10.3.3 Metrics between Probability Distributions

Let p,q be probability distributions on a metric space (X,d) equipped with its Borel σ-algebra Σ. There are a number of important metrics one can place on the space of probability distributions $\mathcal{G}(X,\Sigma)$.

The most basic is the *total variation metric*

$$TV(p,q) = \sup_{E\in\Sigma} |p(E) - q(E)|\,.$$

This measures how much p and q disagree on particular measurable sets.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

A more subtle metric is the Kantorovich metric which measures how different are the integrals defined by the two measures being compared. The definition is

$$\mathbb{K}(p,q) = \sup_f \left| \int f \mathrm{d}p - \int f \mathrm{d}q \right|,$$

with supremum ranging over bounded and non-expansive (which implies continuous) $[0,1]$-valued functions f on X. If we allowed any measurable functions then we could exaggerate the value of the function being integrated on sets where the measures disagree and obtain an infinite sup every time. One can think of the total variation metric as a variant of the Kantorovich metric by considering only indicator functions. However, this is not quite right, as indicator functions are far from non-expansive.

There is an entirely different way of thinking of the Kantorovich metric in terms of transport theory. One thinks of the probability distribution as a "pile of sand" on the space X. Then one needs to move some sand around to change the shape of the pile from p to q. Moving a certain amount of sand has a cost associated with it: this cost is measured by the distance that one has to move the sand. In order to describe a specific "plan" for moving sand we introduce a measure on the product space $X \times X$. A *coupling* π between p, q is a probability distribution on $X \times X$ such that the marginals of π are p, q; in other words we have

$$\pi(A \times X) = p(A) \qquad \text{and} \qquad \pi(X \times B) = q(B).$$

Such a coupling describes a transport plan: $\pi(A \times B)$ describes how much of the probability mass was moved from A to B.

We write $C(p,q)$ for the space of couplings. Then we have the following theorem called *Kantorovich–Rubinstein duality*

$$\mathbb{K}(p,q) = \inf_{\pi \in C(p,q)} \int d(x,y)\mathrm{d}\pi(x,y).$$

In other words the same metric is given by the cost of the minimum-cost transport plan. The right hand side can also be taken to be the definition of the metric. This is usually *incorrectly* called the *Wasserstein metric*[6]

A small variation of the Kantorovich metric can be obtained as follows:

$$W^{(m)}(p,q) = \inf_{\pi \in C(p,q)} \left[\int d(x,y)^m \mathrm{d}\pi(x,y)\right]^{1/m}.$$

If we take $m = 1$ we get the usual Kantorovich metric. A fundamental fact is that

$$W^{(m)}(\delta_x, \delta_y) = d(x,y).$$

[6] Both versions of the metric were invented by Kantorovich. Years later Wasserstein used it in a minor way. Perhaps the fact that Kantorovich used the letter W in his paper added to the confusion. It is also called the "earth movers' distance" by people in the computer vision community and the Hutchinson metric by researchers working on fractals.

This says that the original space is *isometrically embedded* in the space of probability measures.

10.3.4 Markov Processes

Markov processes provide the basic operational semantics for probabilistic programming languages. A L-labelled Markov process (see Panangaden, 2009) is a quadruple:

$$(X, \Sigma, L, (\tau_a \colon X \times \Sigma \to [0,1])_{a \in L}),$$

where the τ_a are Markov kernels. One thinks of a labelled Markov process as a probabilistic labelled transition system with a state space that may be a general measurable space or a Polish space.

One can define a notion of bisimulation as was done by Larsen and Skou (1991) and later extended to the continuous case in Desharnais et al. (2002).

Definition 10.9. An equivalence relation $R \subseteq X \times X$ on the state space of a Markov Process as above is a **bisimulation** if whenever $x \; R \; y$, then

$$\text{for all } a \in L, \quad \tau_a(x, C) = \tau_a(y, C)$$

where C is a measurable union of R-equivalence classes.

Two states x, y are bisimilar if there is *some* bisimulation relation relating them. There is a maximum bisimulation relation which we call simply bisimulation. There is a logical characterization of bisimulation proved in Desharnais et al. (1998, 2002, 2003).

Giacalone et al. (1990) suggested that one move from equality between processes to distances between processes. In Desharnais et al. (1999, 2004) a pseudometric was defined whose kernel was bisimulation. If two states are not bisimilar then some formula distinguishes them. The idea of the metric is: if the *smallest* formula separating two states is "big" the states are "close." Later Worrell and van Breugel (2001) developed a fixed-point definition of the metric and showed how ideas from transport theory could be used to compute the metric more efficiently.

10.4 Quantitative Equational Logic

As we mentioned in the introduction, the basic idea is to introduce approximate equations of the form: $s =_\varepsilon t$, which we understand to mean that s is within ε of t. Clearly, the phrase "within ε" is redolent of a metric but the theory has to be developed to the point where it becomes clear that it is indeed a metric in the precise technical sense.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

At the outset it should be clear that, whatever else it might be, the binary relation denoted by $=_\varepsilon$ is not an equivalence relation. If we have $s =_\varepsilon t$ and $t =_\varepsilon u$ there is no reason to expect $s =_\varepsilon u$; indeed one might expect something more like $s =_{2\varepsilon} u$. The family of relations $\{=_\varepsilon \mid \varepsilon \in [0, \infty]\}$ defines a structure called a *uniformity* but we will not stress this aspect here. We need to formalize what it means to reason with the symbol $=_\varepsilon$ and see that it really corresponds to a quantitative analogue of equational reasoning. In order to do this we will state analogues of the results one has for ordinary equational logic: completeness results, universality of free algebras, Birkhoff-like variety theorem and monads arising from free algebras.

10.4.1 Quantitative Equations

We begin by following as closely as possible the presentation of ordinary equational logic. We have a signature Ω and a set of variables X; in the usual inductive way we get terms denoted by $\mathbf{T}X$. A *quantitative equation* over these terms is of the form:

$$s =_\varepsilon t\,, \qquad\qquad \text{for } s, t \in \mathbf{T}X \text{ and } \varepsilon \in \mathbb{Q}_+\,.$$

We use $\mathcal{I}(\mathbf{T}X)$ to denote the set of quantitative equations over $\mathbf{T}X$. Note that $=_0$ represents ordinary equality $=$, and consequently, $\mathcal{E}(\mathbf{T}X) \subseteq \mathcal{I}(\mathbf{T}X)$.

Let $\mathcal{Q}(\mathbf{T}X)$ be the class of *quantitative judgments* on $\mathbf{T}X$, which are expressions of the form

$$\Gamma \vdash \phi\,,$$

with as hypotheses is an enumerable set $\Gamma \subseteq \mathcal{I}(\mathbf{T}X)$ of quantitative equations and a quantitative equation $\phi \in \mathcal{I}(\mathbf{T}X)$ as conclusion. Since we are identifying $=$ with $=_0$, we observe that $\mathcal{J}(\mathbf{T}X) \subseteq \mathcal{Q}(\mathbf{T}X)$.

Quantitative equations and quantitative judgments are used for reasoning, and to this end we define the concept of quantitative equational theory, which, as might be expected, will generalize the classical equational theory, in the sense that $=_0$ is ordinary term equality. However, for $\varepsilon \neq 0$, $=_\varepsilon$ is not an equivalence: the transitivity rule has to be replaced by a rule, (**Triang**) encoding the triangle inequality. We will also have an infinitary rule, (**Cont**), that reflects the density of rational numbers within the reals.

A *quantitative equational theory of type* Ω *over* X is a set $\mathcal{U}$ of quantitative judgements on $\mathbf{T}X$ such that for arbitrary terms $s, t, u \in \mathbf{T}X$, set of quantitative

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

judgements $\Gamma, \Theta \subseteq \mathcal{I}(\mathbf{T}X)$, and positive rationals $\varepsilon, \varepsilon' \in \mathbb{Q}_+$

$$
\begin{array}{rl}
(\textbf{Refl}) & \emptyset \vdash t =_0 t \in \mathcal{U}\,, \\
(\textbf{Symm}) & \{s =_\varepsilon t\} \vdash t =_\varepsilon s \in \mathcal{U}\,, \\
(\textbf{Triang}) & \{s =_\varepsilon u, u =_{\varepsilon'} t\} \vdash s =_{\varepsilon+\varepsilon'} t \in \mathcal{U}\,, \\
(\textbf{Max}) & \{s = t\} \vdash s =_{\varepsilon+\varepsilon'} t \in \mathcal{U}\,, \\
(\textbf{NExp}) & \{s_i =_\varepsilon t_i \mid i \in \kappa\} \vdash f(s_i)_{i\in\kappa} =_\varepsilon f(t_i)_{i\in\kappa} \in \mathcal{U}\,, \quad \text{for any } f \colon \kappa \in \Omega\,, \\
(\textbf{Cont}) & \{s =_{\varepsilon'} t \mid \varepsilon' > \varepsilon\} \vdash s =_\varepsilon t \in \mathcal{U}\,, \\
(\textbf{Subst}) & \text{if } \Gamma \vdash s =_\varepsilon t \in \mathcal{U}, \text{ then } \sigma(\Gamma) \vdash \sigma(s) =_\varepsilon \sigma(t) \in \mathcal{U}, \text{ for any } \sigma \in \Sigma(X)\,, \\
(\textbf{Cut}) & \text{if } \Theta \vdash \Gamma \in \mathcal{U} \text{ and } \Theta \vdash s =_\varepsilon t \in \mathcal{U}, \text{ then } \Gamma \vdash s =_\varepsilon t \in \mathcal{U}\,, \\
(\textbf{Assum}) & \text{if } s =_\varepsilon t \in \Gamma, \text{ then } \Gamma \vdash s =_\varepsilon t \in \mathcal{U}\,,
\end{array}
$$

Given a quantitative equational theory $\mathcal{U}$ and a set $S \subseteq \mathcal{U}$, we say, as in the classical case, that S is a set of axioms for $\mathcal{U}$, or S axiomatizes $\mathcal{U}$, if $\mathcal{U}$ is the smallest quantitative equational theory that contains S. A quantitative equational theory $\mathcal{U}$ over $\mathbf{T}X$ is *inconsistent* if $\emptyset \vdash x =_0 y \in \mathcal{U}$, where $x, y \in X$ are two distinct variables; $\mathcal{U}$ is *consistent* if it is not inconsistent.

10.4.2 Quantitative Algebras

Now that we have quantitative equations we can turn to defining quantitative analogues of the concept of algebra. Essentially, one combines the algebraic structure from Section 10.2.1 with the concept of a metric space.

A *quantitative* Ω*-algebra* $\mathcal{A} = (A, d, \Omega_\mathcal{A})$ consists of an extended metric space (A, d) and a collection $\Omega_\mathcal{A}$ of non-expansive interpretations for each operation symbol in Ω. If $f \colon a \in \Omega$ is a function symbol of arity κ, then its interpretation is a non-expansive function $f_\mathcal{A} \colon A^\kappa \to A$, where A^κ is the κ-fold cartesian product[7] of the metric space A.

An *homomorphism* from $\mathcal{A} = (A, d_A, \Omega_\mathcal{A})$ to $\mathcal{B} = (B, d_B, \Omega_\mathcal{B})$ is a non-expansive homomorphism of Ω-algebras from $(A, \Omega_\mathcal{A})$ to $(B, \Omega_\mathcal{B})$.

Fixed a set of variables X and a signature Ω, we would like to define the quantitative analogue of the term algebra, but to do so we don't yet have a metric on $\mathbf{T}X$. To do that, we need to explain what it means for an algebra to satisfy a judgement.

Definition 10.10. We say that a quantitative Ω-algebra $\mathcal{A}$ **satisfies** a quantitative judgement $\Gamma \vdash s =_\varepsilon t$ in $\mathcal{Q}(\mathbf{T}X)$, written $\mathcal{A} \models (\Gamma \vdash s =_\varepsilon t)$, if for every assignment $\iota \colon X \to A$ the following implication holds:

$$\big(\text{for all } (s' =_{\varepsilon'} t') \in \Gamma,\ d(\iota(s'), \iota(t')) \leq \varepsilon'\big) \text{ implies } d(\iota(s), \iota(t)) \leq \varepsilon\,.$$

[7] Note that extended metric spaces have all small products; this is not the case for metric spaces.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

For a quantitative equational theory $\mathcal{U}$, we write $\mathcal{A} \models \mathcal{U}$ to mean that $\mathcal{A}$ satisfies all the judgements in $\mathcal{U}$. We write $\mathbf{K}(\mathcal{U}, \Omega)$ for the collection of Ω-algebras satisfying $\mathcal{U}$, or simply $\mathbf{K}(\Omega)$ when the signature is clear.

We can now define a metric on $\mathbf{T}X$ over the quantitative theory $\mathcal{U}$:

$$d^{\mathcal{U}}(s,t) = \inf\{\varepsilon \mid \emptyset \vdash s =_\varepsilon t \in \mathcal{U}\}\,.$$

The idea is that we look at the equations we can derive with the smallest possible ε. We allow only special judgements with empty set of hypotheses. Why not using the following?

$$d^{\mathcal{U}}(s,t) = \inf\{\varepsilon \mid \forall\Gamma \subseteq \mathcal{I}(X), \Gamma \vdash s =_\varepsilon t \in \mathcal{U}\}\,.$$

It turns out that it defines exactly the same metric. Two things are to be noted: first we only have a pseudometric and second, the metric can take on infinite values. To get a proper quantitative algebra on $\mathbf{T}_{\mathcal{U}}[X]$, we have to do the analogue of what we did in the case of ordinary equations: quotient by a suitable equivalence relation. The kernel of the pseudometric is a congruence for Ω. If we take the quotient we get an extended metric space.

We call the resulting quantitative algebra on $\mathbf{T}_{\mathcal{U}}[X]$, the *quantitative term algebra* generated from X; by construction is in $\mathbf{K}(\mathcal{U})$.

10.4.3 General Results

In this section we describe the quantitative analogues of the results mentioned in Section 10.2.2. The first is completeness which was proved in Mardare et al. (2016).

Theorem 10.11 (Completeness)**.** *$\mathcal{U} \models (\Gamma \vdash \phi)$, if and only if, $(\Gamma \vdash \phi) \in \mathcal{U}$.*

This is the analogue of the usual completeness theorem for equational logic. From the right to the left is by definition. The reverse direction is also a model construction argument as in the ordinary case but the proof needs to deal with quantitative aspects and uses the infinitary limit rule (**Cont**) in a crucial way.

Just as in the ordinary case the construction of the term algebra provides us with free algebra. The difference this time is that we start from an *extended metric space* instead of just a set. Starting from an extended metric space (M,d) and a quantitative theory $\mathcal{U}$, we can construct the free quantitative Ω-algebra $\mathbf{T}_{\mathcal{U}}[X]$ generated from (M,d), by adding constants for each $m \in M$ and the judgements $\emptyset \vdash m =_\varepsilon n$ to the generating quantitative theory $\mathcal{U}$, for every rational $\varepsilon \in \mathbb{Q}_+$ such that $d(m,n) \le \varepsilon$. Call this extended signature Ω_M and the extended theory $\mathcal{U}_M$. Clearly, any algebra in $\mathbf{K}(\Omega_M, \mathcal{U}_M)$ can be viewed as an algebra in $\mathbf{K}(\Omega, \mathcal{U})$ by forgetting the interpretations of the additional constants from M.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

Again, we have a non-expansive map $\eta_M : (M, d) \to (\mathbf{T}_{\mathcal{U}}[M], d^{\mathcal{U}_M})$, defined as $\eta_M(m) = [m]$, which is *universal* in the following sense:

$$\text{in } \mathbf{EMet} \qquad\qquad \text{in } \mathbf{K}(\Omega, \mathcal{U})$$

$$\begin{tikzcd} M \arrow[r, "\eta_M"] \arrow[dr, "\alpha"'] & \mathbf{T}_{\mathcal{U}}[M] \arrow[d, "h"] \\ & A \end{tikzcd} \qquad \begin{tikzcd} \mathbf{T}_{\mathcal{U}}[M] \arrow[d, dashed, "h"] \\ \mathcal{A} \end{tikzcd}$$

for any quantitative algebra $\mathcal{A} \in \mathbf{K}(\Omega, \mathcal{U})$ and function α from M to the underlying set A of $\mathcal{A}$, there exists a *unique* quantitative algebra homomorphism $h : \mathbf{T}_{\mathcal{U}}[M] \to \mathcal{A}$ such that $h \circ \eta_M = \alpha$. In other words, $\mathbf{T}_{\mathcal{U}}[M]$ is the *free algebra* in $\mathbf{K}(\Omega, \mathcal{U})$ generated from the space M.

The construction of $\mathbf{T}_{\mathcal{U}}[M]$ from the space (M, d) is functorial and gives the left adjoint to the forgetful functor from $\mathbf{K}(\mathcal{U})$ to **EMet**, the category of extended metric spaces and non-expansive maps. As usual, this gives rise to a monad on **EMet**, namely, the *quantitative term monad* $(\mathbf{T}_{\mathcal{U}}, \eta, \mu)$ with unit and multiplication defined as in the equational case.

Differently from the equational case, the (Eilenberg-Moore) algebras for the monad $\mathbf{T}_{\mathcal{U}}$ are not always in one-to-one correspondence with the algebras in $\mathbf{K}(\mathcal{U})$. However, the isomorphism of categories is recovered in the case the quantitative theory $\mathcal{U}$ is *basic*, *i.e.*, generated by judgements of the form

$$\{x_i =_{\varepsilon_i} y_i \mid i \in I\} \vdash s =_{\varepsilon} t$$

where x_i, y_i are variables in X (see Bacci et al., 2018).

10.5 Examples

The subject as we have presented it so far may seem like generalization for its own sake. In fact there are compelling examples that drove this investigation and these examples come from the world of probabilistic programming.

10.5.1 Axiomatizing the Total Variation Metric

First we return to the example of barycentric algebras from the end of Section 10.2.1. This time we present it as a quantitative algebra. Recall that the signature of barycentric algebras has uncountably many binary operations

$$\{+_e : 2 \mid e \in [0, 1]\},$$

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

satisfying the equations below (here given in the form of quantitative judgements) expressing that $+_e$ is the convex combination of pair of elements

(B1) $\emptyset \vdash x +_1 y =_0 x$,

(B2) $\emptyset \vdash x +_e x =_0 x$,

(SC) $\emptyset \vdash x +_e y =_0 y +_{1-e} x$,

(SA) $\emptyset \vdash (x +_{e_1} y) +_{e_2} z =_0 x +_{e_1 e_2} (y +_{\frac{e_2 - e_1 e_2}{1 - e_1 e_2}} z)$, for $e_1, e_2 \in (0, 1)$,

To the above equations, which just use ordinary equality $=_0$, we add a new quantitative equation schema

$$(\mathbf{LI}) \quad \emptyset \vdash x +_e z =_\varepsilon y +_e z, \qquad \text{for all } e \leq \varepsilon \in \mathbb{Q} \cap [0, 1],$$

called the *left-invariant axiom schema*. Here we are using a nontrivial instance of a quantitative equation.

The barycentric algebras that satisfy (**LI**) are called *left-invariant barycentric algebras* or LIB algebras for short. Denote by $\mathcal{U}_{LI}$ the quantitative equational theory generated form the axioms above. Clearly, the objects in $\mathbf{K}(\mathcal{U}_{LI})$ are exactly the LIB algebras.

If one were to draw a picture of what this means it would violate one's geometric intuition; *it is not meant to be understood in terms of euclidean distance in the plane*. What does this axiomatize? Remarkably, this axiomatizes the total variation metric on probability distributions. This is striking because no mention was made of probability in the above axiomatization and of all the metrics that one can imagine there is nothing in the (**LI**) axiom schema that suggests the total variation metric. Here we sketch the ideas, a detailed proof can be found in Mardare et al. (2016).

We know from the general theory that there is a freely generated LIB algebra from an extended metric space (M, d). What is it concretely? Let us return to this question after constructing a specific LIB algebra.

We recall the definition of the total variation metric from Section 10.3.3:

$$TV(p, q) = \sup_{E \in \Sigma} |p(E) - q(E)|.$$

Here p and q are probability distributions on (M, d) with Borel σ-algebra Σ. There is a beautiful duality theorem for the total variation metric just as there is for the Kantorovich metric (see Lindvall, 2002) which is based on the notion of coupling (see Section 10.3.3 for the definition):

$$TV(p, q) = \min\{\pi(\neq) \mid \pi \in C(p, q)\}.$$

where $C(p, q)$ denote the space of couplings and $\neq$ is the inequality relation on M. Implicit in this statement is the claim that the minimum is attained.

It is easy to see that a convex combination of couplings is a coupling, hence

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

$C(p,q)$ can be turned into a barycentric algebra. Moreover, one can prove (see Mardare et al., 2016) that the following splitting lemma holds:

Lemma 10.12. *If p,q are Borel probability distributions and $e = TV(p,q)$, then there are Borel probability distributions p',q',r such that*

$$p = ep' + (1-e)r \qquad \textit{and} \qquad q = eq' + (1-e)r\,.$$

With these tools in hand we investigate the space of Borel probability distributions on (M,d).

Let $\Pi[M]$ be the barycentric algebra obtained by taking the *finitely-supported* probability distributions on M and interpreting $+_e$ as convex combination; it is easy to verify that the barycentric axioms hold. Then we endow this algebra of distributions with the total-variation metric to make it a quantitative algebra. Using the convexity property of $C(p,q)$ one can prove the following theorem.

Theorem 10.13. $\Pi[M] \in \mathbf{K}(\mathcal{U}_{LI})$.

Moreover, by using the splitting lemma we can prove:

Theorem 10.14. $\Pi[M]$ *is the free algebra generated from M in* $\mathbf{K}(\mathcal{U}_{LI})$.

Since free algebras are unique up to isomorphism, $\Pi[M]$ and the term algebra $\mathbf{T}_{\mathcal{U}_{LI}}[M]$ generated over the left-invariant barycentric axioms are essentially the same algebra. In this sense, we say that the axioms of LIB algebras give rise to the total-variation metric.

10.5.2 Interpolative Barycentric Algebras

We consider a (seemingly) slight variation of the above construction. We have the same signature as barycentric algebras: we keep the axioms (**B1**), (**B2**), (**SC**), (**SA**) but we drop (**LI**). Instead we add the following quantitative equation schema

$$(\mathbf{IB}_m) \;\; \{x =_{\varepsilon_1} y, x' =_{\varepsilon_2} y'\} \vdash x +_e x' =_\delta y +_e y'\,,$$

for all $\delta \in \mathbb{Q}_+$ such that

$$(e\varepsilon_1^m + (1-e)\varepsilon_2^m)^{1/m} \leq \delta\,.$$

Note that now we have assumptions in the equation so this axiom is a judgment with a nonempty left-hand side. We call this axiom ($\mathbf{IB}_m$), which stands for *interpolative barycentric* and the m is a numerical parameter. The barycentric algebras satisfying ($\mathbf{IB}_m$) are called *interpolative barycentric algebras* or IB algebras for short.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

To better understand the axiom ($\mathbf{IB}_m$), it is more illuminating to look at the special case where $m = 1$:

$$\{x =_{\varepsilon_1} y, x' =_{\varepsilon_2} y'\} \vdash x +_e x' =_\delta y +_e y', \qquad \text{where } e\varepsilon_1 + (1-e)\varepsilon_2 \leq \delta\,.$$

We can illustrate this with the picture shown in Figure 10.1.

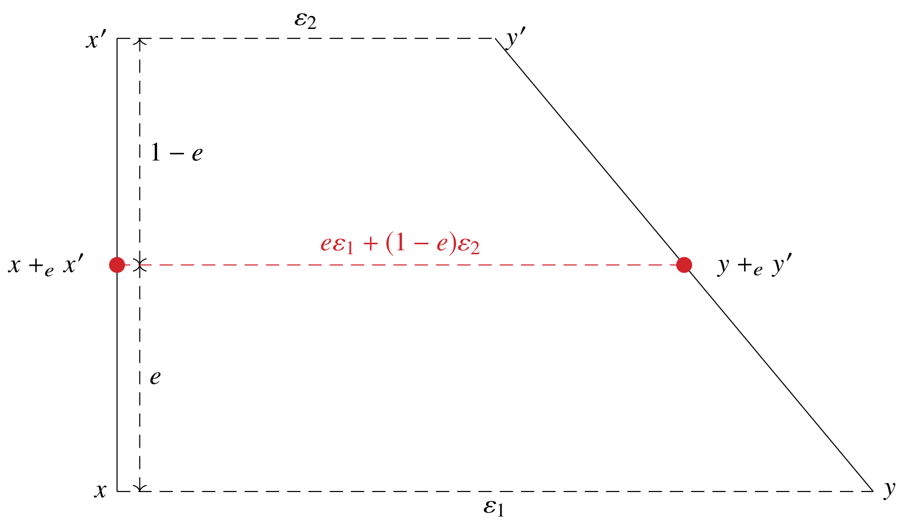

Figure 10.1 The interpolative axiom

We can ask the same questions as we asked for the LIB algebras. What are free IB algebras? We start with an extended metric space (M, d) and consider finitely-supported Borel distributions on it, and interpret them as a barycentric algebra as before. We endow it with the m-Kantorovich metric (see Section 10.3.3) and show that we get an IB algebra. This uses the definition of the $W^{(m)}$ metrics as an inf and convexity of couplings. Again, we can prove a splitting lemma for this case and show that the space of finitely-supported probability distributions with the m-Kantorovich metric is the free IB algebra. The arguments are similar to, but more involved than, the total variation case (see Mardare et al., 2016 for more details).

In fact one can do more. The finitely-supported measures are *weakly dense* in the space of all Borel probability measures. One can show that the space of all Borel probability measures on an extended metric space (M, d), call it $\mathcal{G}_m(M)$, endowed with the $W^{(m)}$ metric gives an IB algebra. One can show that if one constructs the free algebra from (M, d) and then performs *Cauchy completion* one gets a quantitative algebra isomorphic to $\mathcal{G}_m(M)$ by exploiting the weak denseness of the finitely-supported measures.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

10.5.3 Quantitative Exceptions

In this and the next two subsections we discuss quantitative analogues of the well-known work by Plotkin and Power (2001, 2002, 2003, 2004).

The simplest example of an equational theory of effects is given by the algebraic theory of exceptions. We fix a set E of exceptions. For a given set of exception E, the signature is given by a nullary operation symbol $\mathsf{raise}_e : 0$ for each exception $e \in E$:

$$\Omega_E = \{\mathsf{raise}_e : 0 \mid e \in E\}\,.$$

The theory is simply the trivial one, that is the one that contains only identities $t = t$ between terms constructed over the signature.

The induced monad on **Set**, called the *exception monad*, maps a set A to the set $A + E$, the disjoint union of sets A and E.

In the quantitative case one is allowed to view the set of exceptions as an extended metric space with metric measuring the distance between exceptions. This interpretation can be useful, for example, in scenarios where exceptions carry the time-stamps of the moment they have been thrown. In this way one can compare program implementations by measuring the frequency of which exception are thrown.

For (E, d_E), an extended metric space of exceptions, we define the quantitative equational theory of exceptions over E by taking the same signature as above, namely Ω_E, and adding to the theory the quantitative equations

$$\emptyset \vdash \mathsf{raise}_{e_1} =_\varepsilon \mathsf{raise}_{e_2}\,, \qquad \text{for } \varepsilon \geq d_E(e_1, e_2)$$

for any pair of exceptions $e_1, e_2 \in E$ and positive rational ε. The role of this axiom is to lift to the set of terms the underlying metric of E.

The monad T_E on **EMet** induced by this quantitative equational theory is the one that maps an extended metric space M to the extended metric space $M + E$, i.e., the disjoint sum of the extended metric spaces M and E. This example is, admittedly, a trivial extension of the non-metric case.

10.5.4 Quantitative Interactive Input/Output

For representing interactive input and output using equational theories of effects, we typically assume a countable alphabet I of inputs and a set O of outputs; for a signature we take an operation symbol input of arity $|I|$ and a unary operation symbol output_o, for each output symbol $o \in O$

$$\Omega_{I/O} = \{\mathsf{input} : |I|\} \cup \{\mathsf{output}_o : 1 \mid o \in O\}\,.$$

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

The meaning behind the operation symbols is that $\mathsf{input}(t)_i$ represents a computation that waits for user's input and proceeds as t, if the user's entered input is i; while $\mathsf{output}_o(t)$ represents a computation that outputs o and proceeds as t. For example, given a mapping $f \colon I \to O$ from inputs to outputs, the term

$$\mathsf{input}(\mathsf{output}_{f(i)}(\mathsf{output}_{f(i)}(t)))_i\,, \qquad \text{for all } i \in I$$

represents a computation that waits for the user's input i, repeats the output $f(i)$ twice, and then proceeds as t. Above, the term $\mathsf{input}(t_i)_i$ abbreviates the countably branching term

$$\mathsf{input}(t_{i_1}, t_{i_2}, \dots)\,,$$

where $i_1, i_2, \dots$ is an enumeration of input alphabet I.

The equational theory for interactive input/output is given by the trivial theory over the signature $\Omega_{I/O}$. The **Set**-monad $T_{I/O}$ for interactive I/O corresponding to this equational theory is the free monad on the signature functor $\Omega_{I/O}(Y) = Y^I + (O \times Y)$, which is given by the least fixed point

$$T_{I/O}(X) = \mu Y.(Y^I + (O \times Y) + X)\,.$$

Now we consider the situation where the difference between the output symbols produced is measured by a metric. For example we may produce output streams and there are natural metrics between streams. We assume that (O, d_O) is a metric space of outputs and we define a quantitative equational theory to capture interactive input/output effects.

Recall that the general theory for quantitative equations requires every operation symbol to satisfy the following axiom of non-expansiveness:

$$\{x_i =_\varepsilon y_i \mid i \in I\} \vdash \mathsf{input}(x_i)_i =_\varepsilon \mathsf{input}(y_i)_i\,,$$
$$\{x =_\varepsilon y\} \vdash \mathsf{output}_o(x) =_\varepsilon \mathsf{output}_o(y) \qquad \text{for all } o \in O\,.$$

In order to obtain a quantitative theory of interactive input/output effects able to reflect the difference of two computations producing sequences of outputs symbols, in addition to the above quantitative equations we require the theory to have the following axioms:

$$\{x =_\varepsilon y\} \vdash \mathsf{output}_{o_1}(x) =_\delta \mathsf{output}_{o_2}(y)\,, \qquad \text{for } \delta \geq \max(\varepsilon, d_O(o_1, o_2))\,,$$

for each pair $o_1, o_2 \in O$ of outputs symbols and positive rationals ε, δ.

As a consequence the theory will also contain the quantitative equation

$$\emptyset \vdash \mathsf{output}_{a_1,\dots,a_n}(x) =_\delta \mathsf{output}_{b_1,\dots,b_n}(x)\,, \qquad \text{for } \delta \geq \max_{i=1}^{n} d_O(a_i, b_i)\,,$$

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

where $\mathsf{output}_{o_1,\dots,o_n}(t)$ abbreviates the term

$$\mathsf{output}_{o_1}(\mathsf{output}_{o_2}(\mathsf{output}_{o_3}(\dots\mathsf{output}_{o_n}(t)))),$$

representing a computation printing the word $o_1 \cdots o_n$ and proceeding as t. Hence the difference of printing two words of the same length is quantified as the maximal point-wise distance between their characters. There are, of course, other variations one can imagine.

This quantitative equational theory induces a monad $T_{I/O}$ for interactive input/output determined as the following least fixed point on **EMet**

$$T_{I/O}(X) = \mu Y.(Y^I + (O \times Y) + X).$$

10.5.5 Quantitative Side-Effects (State Monad)

To describe state with a finite set L of locations and a countable metric space (V, d_V) of data values, we take a signature containing an operation symbol lookup_l of arity $|V|$ for each location $l \in L$, and a unary operation symbol $\mathsf{update}_{l,v}$ for each location $l \in L$ and data value $v \in V$.

$$\Omega_{\mathsf{State}} = \{\mathsf{lookup}_l : |V| \mid l \in L\} \cup \{\mathsf{update}_{l,v} : 1 \mid l \in L \text{ and } v \in V\}.$$

The term $\mathsf{lookup}_l(t)_v$ represents a computation that looks up the contents of location l and proceeds as t if the stored value is v. The term $\mathsf{update}_{l,v}(t)$ represents a computation that updates the location l with v and proceeds as t. For example, the term

$$\mathsf{lookup}_{l_1}(\mathsf{update}_{l_2,v}(t))_v \qquad \text{for all } v \in V$$

represents a computation that copies the contents of l_1 into the location l_2 and proceeds as t. Note that, as for the case of the input operation in Section 10.5.4, the term $\mathsf{lookup}_l(t_{v_i})_{v_i}$ is an abbreviation for the countably branching term

$$\mathsf{lookup}_l(t_{v_1}, t_{v_2}, \dots),$$

where $v_1, v_2, \dots$ is an enumeration of the data values in V.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

The quantitative theory of side-effects is given by the following axioms

$$\begin{aligned}
&\emptyset \vdash \mathsf{lookup}_l(\mathsf{update}_{l,v}(x))_v =_0 x\,,\\
&\emptyset \vdash \mathsf{lookup}_l(\mathsf{lookup}_l(x)_{v_2})_{v_1} =_0 \mathsf{lookup}_l(y)_{v_1}\,,\\
&\emptyset \vdash \mathsf{update}_{l,v_1}(\mathsf{lookup}_l(x)_{v_2}) =_0 \mathsf{update}_{l,v_1}(y)\,,\\
&\emptyset \vdash \mathsf{update}_{l,v_1}(\mathsf{update}_{l,v_2}(x)) =_0 \mathsf{update}_{l,v_2}(x)\,,\\
&\emptyset \vdash \mathsf{lookup}_{l_1}(\mathsf{lookup}_{l_2}(x)_{v_2})_{v_1} =_0 \mathsf{lookup}_{l_2}(\mathsf{lookup}_{l_1}(x)_{v_1})_{v_2}\,,\\
&\emptyset \vdash \mathsf{update}_{l_1,v_1}(\mathsf{lookup}_{l_2}(x)_{v_2}) =_0 \mathsf{lookup}_{l_2}(\mathsf{update}_{l_1,v_1}(x))_{v_2}\,,\\
&\emptyset \vdash \mathsf{update}_{l_1,v_1}(\mathsf{update}_{l_2,v_2}(x)) =_0 \mathsf{update}_{l_2,v_2}(\mathsf{update}_{l_1,v_1}(x))\,,
\end{aligned}$$

$$\{x =_\varepsilon y\} \vdash \mathsf{update}_{l,v_1}(x) =_\delta \mathsf{update}_{l,v_2}(y)\,, \qquad \text{for } \delta \geq \max(\varepsilon, d_V(v_1, v_2))\,,$$

where in the above, the locations l_1, l_2 are assumed to be distinct: $l_1 \neq l_2$.

The first four equations describe the behaviour of operations on a single location: the first one says that updating a location with its current contents has no effect; the second one that the state does not change between two consecutive lookups; the third one that the state is determined immediately after an update; and the fourth one that the second update overwrites the first one. The next three ordinary equations state that operations on different locations commute. The last equation, which is also the only truly quantitative one in the above list, states that the difference between side-effects depends on the distance of the values observed point-wise in each location.

The monad on **EMet** induced by the above axioms maps an extended metric space M to $(S \times M)^S$, where $S = V^L$.

Remark 10.15. If we took an *infinite* set L of locations, the induced monad would not be the standard one for state. Since the elements of the free model are built inductively from operations and represent computations that only update a finite number of locations at a time. In contrast, the elements of the standard monad represent computations that can perform an arbitrary modification of the state.

10.6 Conclusions

This chapter introduces a new approach to approximate reasoning. Metrics for probabilistic processes have been investigated for nearly twenty years by Desharnais et al. (1999, 2004) and van Breugel and Worrell (2001b,a) and of course the deBakker school has emphasized metric ideas in semantics for decades. Logics for reasoning quantitatively have essentially been modal logics that were particularly crafted for probabilistic systems but a *generic* way of capturing the notion of approximate equality has been missing.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

The approach described in this chapter is just a beginning. We hope that the striking emergence of the Kantorovich metric as a free algebra from a fairly simple equational theory is a foretaste of what might be expected in the future. From the programming point of view we have just presented very simple obvious extensions to quantitative theories of effects. We are actively investigating a more comprehensive theory of effects specifically for probabilistic programming languages. In recent work (Bacci et al., 2018) we have shown how one can combine different monads to obtain, for example, an equational characterization of Markov processes.

The theory presented here has a number of restrictions introduced for ease of exposition. For example, nonexpansiveness can certainly be weakened. We know that we only require nonexpansiveness in each argument separately. However, we expect that yet weaker conditions are possible, perhaps at the price of complicating the underlying theory.

A number of other directions for future research are: (i) developing a quantitative term rewriting theory that meshes with quantitative equational logic, (ii) understanding better how much the bounds degrade as one manipulates sequences of equations and (iii) algorithms based on quantitative equations. To elaborate point (ii): in ordinary equational logic, a long series of equations comes without cost but in quantitative equational logic a long series of quantitative equational manipulations may well cause the ε's appearing to get larger and larger to the point of being uninformative. It would be useful to get a handle on the "ergonomics"[8] of quantitative equational reasoning.

Acknowledgements

We have benefitted from many discussions with colleagues. We mention William Boshuk, Florence Clerc, Vincent Danos, Nathanaël Fijalkow, Marcelo Fiore, Jeremy Gibbons, Bartek Klin, Dexter Kozen, Alexander Kurz, Bart Jacobs, Paul Levy, Michael Mislove, Marcin Sabok, Dana Scott and Alexandra Silva. This research has been supported by a grant from NSERC (Panangaden) and the Danish Research Council (Mardare, Bacci).

References

Bacci, Giorgio, Mardare, Radu, Panangaden, Prakash, and Plotkin, Gordon D. 2018. An Algebraic Theory of Markov Processes. Pages 679–688 of: *Proceedings of the 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2018, Oxford, UK, July 09-12, 2018*.

Billingsley, P. 1995. *Probability and Measure*. Wiley-Interscience.

[8] We thank Jeremy Gibbons for bringing this point to our attention as well as coining this phrase.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

Birkhoff, G. 1935. On the structure of abstract algebras. *Proc. Cambridge Philos. Soc.*, **31**, 433–454.

Burris, Stanley, and Sankappanavar, Hanamantagouda P. 1981. *A course in universal algebra*. Graduate Texts in Mathematics, vol. 78. Springer-Verlag.

Desharnais, J., Edalat, A., and Panangaden, P. 1998. A Logical Characterization of Bisimulation for Labelled Markov Processes. Pages 478–489 of: *proceedings of the 13th IEEE Symposium On Logic In Computer Science, Indianapolis*. IEEE Press.

Desharnais, J., Gupta, V., Jagadeesan, R., and Panangaden, P. 1999. Metrics for Labeled Markov Systems. In: *Proceedings of CONCUR99*. Lecture Notes in Computer Science, no. 1664. Springer-Verlag.

Desharnais, J., Edalat, A., and Panangaden, P. 2002. Bisimulation for Labeled Markov Processes. *Information and Computation*, **179**(2), 163–193.

Desharnais, J., Gupta, V., Jagadeesan, R., and Panangaden, P. 2003. Approximating Labeled Markov Processes. *Information and Computation*, **184**(1), 160–200.

Desharnais, Josée, Gupta, Vineet, Jagadeesan, Radhakrishnan, and Panangaden, Prakash. 2004. A metric for labelled Markov processes. *Theoretical Computer Science*, **318**(3), 323–354.

Dudley, R. M. 1989. *Real Analysis and Probability*. Wadsworth and Brookes/Cole.

Giacalone, A., Jou, C., and Smolka, S. 1990. Algebraic Reasoning for Probabilistic Concurrent Systems. In: *Proceedings of the Working Conference on Programming Concepts and Methods*. IFIP TC2.

Giry, M. 1981. A Categorical Approach to Probability Theory. Pages 68–85 of: Banaschewski, B. (ed), *Categorical Aspects of Topology and Analysis*. Lecture Notes In Mathematics, no. 915. Springer-Verlag.

Hyland, Martin, and Power, John. 2007. The Category Theoretic Understanding of Universal Algebra: Lawvere Theories and Monads. *Electronic Notes in Theor. Comp. Sci.*, **172**, 437–458.

Hyland, Martin, Plotkin, Gordon, and Power, John. 2006. Combining effects: Sum and tensor. *Theoretical Computer Science*, **357**(1), 70–99.

Larsen, K. G., and Skou, A. 1991. Bisimulation through Probablistic Testing. *Information and Computation*, **94**, 1–28.

Lawvere, F. W. 1962. *The category of probabilistic mappings*. Unpublished typescript. Available at
`ncatlab.org/nlab/files/lawvereprobability1962.pdf`.

Lindvall, Torgny. 2002. *Lectures on the coupling method*. Courier Corporation.

Mardare, Radu, Panangaden, Prakash, and Plotkin, Gordon. 2016. Quantitative algebraic reasoning. Pages 700–709 of: *Proceedings of the 31st Annual ACM-IEEE Symposium on Logic in Computer Science*.

Mardare, Radu, Panangaden, Prakash, and Plotkin, Gordon. 2017. On the axiomatizability of quantitative algebras. In: *Proceedings of the 32nd Annual ACM-IEEE Symposium on Logic in Computer Science*.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press

Moggi, Eugenio. 1991. Notions of computation and monads. *Information and computation*, **93**(1), 55–92.

Panangaden, Prakash. 2009. *Labelled Markov Processes*. Imperial College Press.

Plotkin, Gordon, and Power, John. 2001. Semantics for algebraic operations. *Electronic Notes in Theoretical Computer Science*, **45**, 332–345.

Plotkin, Gordon, and Power, John. 2002. Notions of computation determine monads. Pages 342–356 of: *Foundations of Software Science and Computation Structures*. Springer.

Plotkin, Gordon, and Power, John. 2003. Algebraic operations and generic effects. *Applied Categorical Structures*, **11**(1), 69–94.

Plotkin, Gordon, and Power, John. 2004. Computational effects and operations: An overview. *Electronic Notes in Theoretical Computer Science*, **73**, 149–163.

Robinson, Edmund. 2002. Variations on Algebra:Monadicity and generalisations of equational theories. *Formal Aspects of Computing*, **13**, 308–326.

van Breugel, Franck. 2001. An introduction to metric semantics: operational and denotational models for programming and specification languages. *Theoretical Computer Science*, **258**(1), 1–98.

van Breugel, Franck, and Worrell, James. 2001a. An Algorithm for Quantitative Verification of Probabilistic Systems. Pages 336–350 of: Larsen, K. G., and Nielsen, M. (eds), *Proceedings of CONCUR'01*. Lecture Notes In Computer Science, no. 2154. Springer-Verlag.

van Breugel, Franck, and Worrell, James. 2001b. Towards Quantitative Verification of Probabilistic Systems. In: *Proceedings of the Twenty-eighth International Colloquium on Automata, Languages and Programming*. Springer-Verlag.

https://doi.org/10.1017/9781108770750.011 Published online by Cambridge University Press