Equational reasoning for probabilistic programming Part I: (a) Basic equational logic (b) Metrics

Prakash Panangaden[1]

[1]School of Computer Science
McGill University

Probabilistic Proramming Languages 29th May - 2nd June 2017

Basic ideas

- Equations are at the heart of mathamatical reasoning.
- Reasoning about programs is also based on program equivalences.
- A trinity of ideas: Equationally given algebras, Lawvere theories, Monads on **Set**
- The dawning of the age of quantitative reasoning.
- We want quantitative analogues of algebraic reasoning.
- (Pseudo)metrics instead of equivalence relations.
- Equality indexed by a real number $=_\epsilon$.
- Monads on **Met**.
- Enriched Lawvere theories?

Outline

- Summary of equational logic
- Monads
- Monads and computation
- Metrics for probabilistic systems

Finitary equational theories

- Signature $\Omega = \{(Op_i, n_i) | i = 1 \ldots k\}$
- Terms $t ::== x | Op(t_1, \ldots, t_n)$
- Equations $s = t$
- Axioms, sets of equations Ax
- Deduction $Ax \vdash s = t$
- Usual rules for deduction: equivalence relation, congruence,...
- Theories: set of equations closed under deduction.

Equational deduction rules

- Axiom $Ax \vdash s = t$ if $s = t \in Ax$
- Equivalence

$$\frac{}{Ax \vdash t = t}$$

$$\frac{Ax \vdash s = t, Ax \vdash t = u}{Ax \vdash s = u}$$

$$\frac{Ax \vdash s = t}{Ax \vdash t = s}$$

- Congruence

$$\frac{Ax \vdash t_1 = s_1, \ldots Ax \vdash t_n = s_n}{Ax \vdash Op(t_1, \ldots, t_n) = Op(s_1, \ldots, s_n)}$$

- Substitution

$$\frac{Ax \vdash t = s}{Ax \vdash t[u/x] = s[u/x]}$$

Algebras equationally I

- We assume that that there is one set of "basic things" – one-sorted algebras.
- Fix a set Ω of *operations*, each with a fixed arity $n \in \mathbb{N}$. These include *constants* as arity zero "operations." Such an Ω is called a signature.
- Everything has finite arity.
- As Ω-algebra $\mathcal{A}$ is a set A to interpret the basic sort and, for each operation f of arity n a function $f_{\mathcal{A}} : A^n \longrightarrow A$.

Algebras equtionally II

- Can define homomorphisms and subalgebras easily.
- What about equations that are required to hold?
- Given a set X we define the *term algebra generated by* X, TX
- The elements of X are in TX.
- If $t_1, \ldots, t_n$ are in TX and f has arity n then $f(t_1, \ldots, t_n)$ is in TX.

Algebras from equations I

- Want to write things like $\forall x, y, z; f(x, f(y, z)) = f(f(x, y), z)$.
- X, set of *variables*.
- Let s, t be terms in TX, we say the equation $s = t$ *holds* in an Ω-algebra $\mathcal{A}$ if *for every* homomorphism $h : TX \longrightarrow \mathcal{A}$ we have $h(s) = h(t)$ where, in the latter, $=$ means identity.
- Let S be a set of equations between pairs of terms in TX. We define a *congruence relation* $\sim_S$ on TX in the evident way.

Algebras from equations II

- Easy to check that if $t_1 \sim_S s_1, \ldots, t_n \sim_S s_n$ then $f(t_1, \ldots, t_n) \sim_S f(s_1, \ldots, s_n)$ we can define $f_{\sim_S}$ on $TX/\sim_S$.
- Let $[t]$ be an equivalence class of $\sim_S$; $f_{\sim_S}([t_1], \ldots, [t_n])$ is well defined by $[f(t_1, \ldots, t_n)]$.
- A class of Ω-algebras satisfying a set of equations is called a variety of algebras (not the same as an algebraic variety!).
- When are a set of equations bad? If we can derive $x = y$ from S then the only algebras have one element.

Examples

- Monoids, groups, rings, lattices, boolean algebras are all examples.
- Vector spaces have two sorts.
- Fields are annoying because we have to say $x \neq 0$ implies x^{-1} exists. Fields do not form an equational variety.
- Sometimes we need to state conditional equations; these are called *Horn clauses*. Example: cancellative monoids, $x \cdot y = x \cdot z \vdash y = z$.
- Stacks are equationally definable but queues are not.

Example: barycentric algebras (Stone 1949)

- Signature:

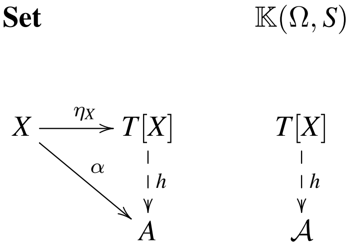
$$\{+_\epsilon | \epsilon \in [0,1]\}$$

- Axioms:

$$(B_1) \vdash t +_1 t' = t$$
$$(B_2) \vdash t +_\epsilon t = t$$
$$(SC) \vdash t +_\epsilon t' = t' +_{1-\epsilon} t$$
$$(SA) \vdash (t +_\epsilon t') +_{\epsilon'} t'' = t +_{\epsilon\epsilon'} (t' +_{\frac{\epsilon' - \epsilon\epsilon'}{1-\epsilon\epsilon'}} t'')$$

Universal properties

- Let $\mathbb{K}(\Omega, S)$ be the collection of algebras satisfying the equations in S. $\mathbb{K}(\Omega, S)$ becomes a category if we take the morphisms to be Ω-homomorphisms.
- Let X be a set of generators. We write $T[X]$ for $TX/\sim_S$. There is a map $\eta_X : X \longrightarrow T[X]$ given by $\eta_X(x) = [x]$.
- Universal property.



Variety theorem

Birkhoff

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

There are analogoues results for algebras defined by Horn clauses: quasivariety theorems.

Example

Consider $\mathbb{Z}_2 \times \mathbb{Z}_2$. It's not a field because, *e.g.* $(1,0) \times (0,1) = (0,0)$. Hence fields cannot be described by equations!

Monads

- Capturing universal algebra categorically.
- Data: (i) Endofunctor $T : \mathcal{C} \longrightarrow \mathcal{C}$, (ii) $\eta : I \longrightarrow T$ natural, and (iii) $\mu : T^2 \longrightarrow T$ also natural.
- Some diagrams are required to commute.

$$\begin{array}{ccc} T^3A & \xrightarrow{\mu_{TA}} & T^2A \\ {\scriptstyle T\mu_A}\downarrow & & \downarrow{\scriptstyle \mu_A} \\ T^2A & \xrightarrow[\mu_A]{} & TA \end{array} \qquad\qquad \begin{array}{ccccc} TA & \xrightarrow{\eta_{TA}} & T^2A & \xleftarrow{T\eta_A} & TA \\ & \searrow\!\!= & \downarrow{\scriptstyle \mu_A} & =\!\!\swarrow & \\ & & TA & & \end{array}$$

- Examples: powerset, “free” constructions *e.g.* monoid, group, the Giry monad.

The Kleisli construction

- From a monad $T : \mathcal{C} \longrightarrow \mathcal{C}$ make a new category: the Kleisli category $\mathcal{C}_T$.
- Objects, the same as those of $\mathcal{C}$.
- Morphisms $f : A \longrightarrow B$ in $\mathcal{C}_T$ are $f : A \longrightarrow TB$ in $\mathcal{C}$.
- Composition? $f : A \longrightarrow TB$ and $g : B \longrightarrow TC$ don't match.
- $f : A \longrightarrow TB$ and $Tg : TB \longrightarrow T^2C$ to match but we are in T^2C.
- Compose with $\mu_C : T^2C \longrightarrow TC$ to get $A \longrightarrow TC$.
- The Kleisli category of the powerset monad is the category of sets and relations.

The Gìry monad I

- **Mes**: objects are sets equipped with a σ-algebra (X, Σ), morphisms $f : (X, \Sigma) \longrightarrow (Y, \Lambda)$ are functions $f : X \longrightarrow Y$ such that $\forall B \in \Lambda, f^{-1}(B) \in \Sigma$.
- $\mathcal{G} : \mathbf{Mes} \longrightarrow \mathbf{Mes}$, $\mathcal{G}(X, \Sigma) = \{p | p$ is a probability measure on $\Sigma\}$.
- For each $A \in \Sigma$, define $e_A : \mathcal{G}(X) \longrightarrow [0, 1]$ by $e_A(p) = p(A)$. Equip $\mathcal{G}(X)$ with the smallest σ-algebra making all the e_A measurable.
- $f : X \longrightarrow Y$, $\mathcal{G}(f) : \mathcal{G}(X) \longrightarrow \mathcal{G}(Y)$ given by $\mathcal{G}(f)(p)(B \in \Lambda) = p(f^{-1}(B))$.

The Giry monad II

- $\eta_X : X \longrightarrow \mathcal{G}(X)$ given by $\eta_X(x) = \delta_x$, where $\delta_x(A) = 1$ if $x \in A$ and 0 if $x \notin A$.
- $\mu_X(Q \in \mathcal{G}^2(X))(A) = \int e_A \mathrm{d}Q$. Averaging over $\mathcal{G}$ using Q.
- Probabilistic analogue of the powerset.

The Kleisli category of $\mathcal{G}$

- Objects: Same as $\mathbf{Mes}$, morphisms from X to Y are measurable functions from X to $\mathcal{G}(Y)$.
- Compose: $h : X \longrightarrow \mathcal{G}(Y)$, $k : Y \longrightarrow \mathcal{G}(Z)$ by the formula: $(k \tilde{\circ} h) = (\mu_Z) \circ (\mathcal{G}(k)) \circ h$ where $\tilde{\circ}$ is the Kleisli composition and $\circ$ is composition in $\mathbf{Mes}$.
- Curry the definition of morphism: $h : X \times \Sigma_Y \longrightarrow [0, 1]$. Markov kernels. We call this category $\mathbf{Ker}$. Probabilistic relations.
- Composition in terms of kernels: $(k \tilde{\circ} h)(x, C \subset Z) = \int k(y, C) h(x, \cdot)$. Relational composition, matrix multiplication.

The Eilenberg-Moore category

- From T we can construct a category of algebras: objects $a : TA \longrightarrow A$
- and morphisms $f : A \longrightarrow B$ such that

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

  commute.
- Many categories of algebras (monoids, groups, rings, lattices) can be reconstructed this way.
- The Kleisli category $=$ the category of "free" algebras.
- We get a monad on $\mathbf{Set}$ from $X \mapsto T[X]$. The Eilenberg-Moore category for this monad is isomorphic to $\mathbb{K}(\Omega, S)$.
- Algebras for a monad $\Leftrightarrow$ Algebras given by equations and operations.

Pseudometrics

- Quantitative analogue of an equivalence relation.
- Space M, (pseudo)metric $d : M \times M \longrightarrow \mathbb{R}^{\geq 0}$
- $d(x,x) = 0$, $d(x,y) = d(y,x)$ and $d(x,z) \leq d(x,y) + d(y,z)$.
- If $d(x,y) = 0$ implies $x = y$ we say d is a **metric.**
- We can define usual notions of convergence, completeness, topology, continuity etc.
- Maps: $f(X,d) \longrightarrow (Y,d')$ are *nonexpansive* $d'(f(x),f(y)) \leq d(x,y)$; automtically continuous
- We define $\mathbf{Met}$: objects metric spaces, morphisms are nonexpansive functions.
- Quantitative equations give monads on $\mathbf{Met}$.

Metrics between probability distributions

Let p, q be probability distributions on (X, d, Σ).

- Total variation $tv(p,q) = \sup_{E\in\Sigma} |p(E) - q(E)|$.
- Kantorovich: $\kappa(p,q) = \sup_f |\int f\mathrm{d}p - \int f\mathrm{d}q|$ where f is nonexpansive.
- A *coupling* π between p, q is a distribution on $X \times X$ such that the marginals of π are p, q. Write $\mathcal{C}(p,q)$ for the space of couplings.
- Kantorovich: $\kappa(p,q) = \inf_{\mathcal{C}(p,q)} \int_{X\times X} d(x,y)\mathrm{d}\pi(x,y)$. Kantorovich-Rubinshtein duality.
- Wasserstein: $W^{(l)}(p,q) = \inf_{\mathcal{C}(p,q)} [\int_{X\times X} d(x,y)^l \mathrm{d}\pi(x,y)]^{1/l}$. $l = 1$ gives Kantorovich.
- $W^{(l)}(\delta_x, \delta_y) = d(x,y)$.

Markov processes

- Basic operational semantics for probabilistic programming languages.
- $(S, \Sigma, \mathcal{A}, \forall a \in \mathcal{A} \tau_a : X \times \Sigma \longrightarrow [0, 1]$.
- τ_a are Markov kernels.

Bisimulation

- Let R be an equivalence relation. R is a bisimulation if: $s\ R\ t$ if $(\forall\ a)$:

  $$\tau_a(s, C) = \tau_a(t, C)$$

  where C is a measurable union of R-equivalence classes.
- We say R is *a* bisimulation relation.
- s, t are bisimilar if there is a bisimulation relating them.
- There is a maximum bisimulation relation.

A metric-based approximate viewpoint

- Move from equality between processes to distances between processes (Jou and Smolka 1990).
- There is a logical characterization of bisimulation.
- If two states are not bisimilar then some formula distinguishes them.
- If the *smallest* formula separating two states is "big" the states are "close."
- We can define a pseudometric such that distance is zero iff the states are bisimilar.

Metric "bisimulation"

- d is a metric-bisimulation if: $d(s,t) < \epsilon \Rightarrow$:

$$\kappa(\tau(s,\cdot),\tau(t,\cdot)) < \epsilon$$

- The required canonical metric on processes is the least such: ie. the distances are the least possible.
- Thm: *Canonical least metric exists*.
- Uses basic fixed-point theory on the complete lattice of pseudometrics.

Real-valued modal logic I

- Develop a real-valued "modal logic" based on the analogy:

Kozen's analogy

Program Logic	Probabilistic Logic
State s	Distribution μ
Formula ϕ	Random Variable f
Satisfaction $s \models \phi$	$\int f \mathrm{d}\mu$

- Define a metric based on how closely the random variables agree.
- Thm: d coincides with the fixed-point definition of metric-bisimulation.

Equational reasoning for probabilistic programming Part II: Quantitative equational logic

Prakash Panangaden[1]

[1]School of Computer Science
McGill University

Probabilistic Proramming Languages 29th May - 2nd June 2017

The basic idea

- Approximate equations: $s =_\varepsilon t$, s is within ε of t.
- Definitely not an equivalence relation;
- it defines a *uniformity* (but we won't stress this point of view).
- Quantitative analogue of equational reasoning.
- completeness results, universality of free algebras, Birkhoff-like variety theorem, monads ....

History

- Moggi 1988: How to incorporate “effects” into denotational semantics?
- (Strong) Monads!
- Plotkin, Power (and then many others): view effects algebraically. Monads are given by operations and equations.
- Categorically: equational presentations are Lawvere theories (but we won’t talk about them here either).
- A monad of great interest: Lawvere (1964) The category of probabilistic mappings.
- Giry (1981): monad on measure spaces and also on Polish spaces.

Metrics

- Probabilistic reasoning requires measure theory but,
- measure theory works best on Polish spaces (topological space underlying separable complete metric spaces).
- Metric ideas present in semantics from the start: Jaco de Bakker's school.
- Mardare, P., Plotkin (2016): Develop the theory of effects in a metric setting (motivated by probability).
- Algebras will come with metric structure and quantitative equational theories will define monads on $\mathbf{Met}$.

Quantitative equations

- Signature Ω, variables X we get terms $\mathbb{T}X$.
- Quantitative equations: $\mathcal{V}(\mathbb{T}X)$:

$$s =_\varepsilon t, \quad s, t \in \mathbb{T}X, \quad \varepsilon \in \mathbb{Q} \cap [0, 1]$$

- A substitution σ is a map $X \longrightarrow \mathbb{T}X$; we write $\Sigma(X)$ for the set of substitutions.
- Any σ extends to a map $\mathbb{T}X \longrightarrow \mathbb{T}X$.
- Quantitative inferences: $\mathcal{E}(\mathbb{T}X) = \mathcal{P}_{\text{fin}}(\mathcal{V}(\mathbb{T}X)) \times \mathcal{V}(\mathbb{T}X)$

$$\{s_1 =_{\varepsilon_1} t_1, \ldots, s_n =_{\varepsilon_n} t_n\} \vdash s =_\varepsilon t$$

Deducibility relations

(Refl) $\emptyset \vdash t =_0 t$

(Symm) $\{t =_\varepsilon s\} \vdash s =_\varepsilon t$.

(Triang) $\{t =_\varepsilon s, s =_{\varepsilon'} u\} \vdash t =_{\varepsilon+\varepsilon'} u$.

(Max) For $e' > 0$, $\{t =_\varepsilon s\} \vdash t =_{\varepsilon+\varepsilon'} s$.

(Arch) For all $\varepsilon \geq 0$, $\{t =_{\varepsilon'} s \mid \varepsilon' > \varepsilon\} \vdash t =_\varepsilon s$. Infinitary!

(NExp) For $f : n \in \Omega$,
$\{t_1 =_\varepsilon s_1, \ldots, t_n =_\varepsilon s_n\} \vdash f(t_1, ..t_i, ..t_n) =_\varepsilon f(s_1, ..s_i, ..s_n)$

(Subst) If $\sigma \in \Sigma(X)$, $\Gamma \vdash t =_\varepsilon s$ implies $\sigma(\Gamma) \vdash \sigma(t) =_\varepsilon \sigma(s)$.

(Cut) If $\Gamma \vdash \phi$ for all $\phi \in \Gamma'$ and $\Gamma' \vdash \psi$, then $\Gamma \vdash \psi$.

(Assumpt) If $\phi \in \Gamma$, then $\Gamma \vdash \phi$.

Quantitative equational theories

- Given $S \subset \mathcal{E}(\mathbb{T}X)$, $\vdash_S$: smallest deducibility relation containing S.
- Equational theory: $\mathcal{U} = \ \vdash_S \bigcap \mathcal{E}(\mathbb{T}X)$.

Quantitative algebras

- Ω: signature; $\mathcal{A} = (A, d)$,
  A an Ω-algebra and (A, d) a metric space.
- All functions in Ω are nonexpansive.
- Morphisms are Ω-algebra homomorphisms that are nonexpansive.
- $\mathbb{T}X$ is an Ω-algebra. $\sigma : \mathbb{T}X \longrightarrow A$, Ω-homomorphism.
- (A, d) **satisfies** $\{s_i =_{\varepsilon_i} t_i / i = 1, \ldots, n\} \vdash s =_\varepsilon t$ if

$$\forall \sigma,\ d(\sigma(s_i), \sigma(t_i)) \leq \varepsilon_i,\ i = 1, \ldots, n$$

implies

$$d(\sigma(s), \sigma(t)) \leq \varepsilon.$$

- We write $\{s_i =_{\varepsilon_i} t_i / i = 1, \ldots, n\} \models_{\mathcal{A}} s =_\varepsilon t$.
- We write $\mathbb{K}(\mathcal{U}, \Omega)$ for the algebras satisfying $\mathcal{U}$.

A metric on $\mathbb{T}X$

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

- Why not use the following?

$$d^{\mathcal{U}}(s,t) = \inf\{\varepsilon \mid \forall V \in \mathcal{P}_f(\mathcal{V}(X)), V \vdash s =_{\varepsilon} t \in \mathcal{U}\}$$

- They are the same!
- The (pseudo)metric can take on infinite values.
- The kernel is a congruence for Ω.
- If we take the quotient we get an (extended) metric space.
- The resulting algebra is in $\mathbb{K}(\Omega, \mathcal{U})$.
- We can do this for any set M of generators and produce a “free” quantitative algebra.

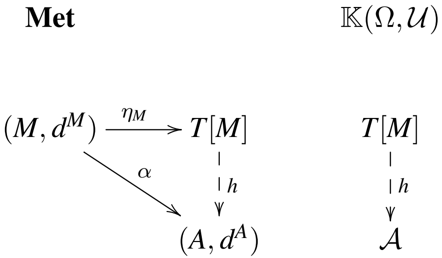
Completeness

$\forall \mathcal{A} \in \mathbb{K}(\mathcal{U}, \Omega)$, $\Gamma \models_{\mathcal{A}} \phi$ if and only if $[\Gamma \vdash \phi] \in \mathcal{U}$.

- Analogue of the usual completeness theorem for equational logic.
- Right to left is by definition.
- Left to right is by a model construction.
- The proof needs to deal with quantitative aspects and uses the archimedean property.

Free construction from a metric space

- Starting from a **metric space** (M, d) we can define $\mathbb{T}M$ by adding constants for each $m \in M$
- and axioms $\emptyset \vdash m =_e n$ for every rational e such that $d(m, n) \leq e$.
- Call this extended signature Ω_M and the extended theory $\mathcal{U}_M$.
- Any algebra in $\mathbb{K}(\mathcal{U}_M, \mathcal{U}_M)$ can be viewed as an algebra in $\mathbb{K}(\Omega, \mathcal{U})$ by forgetting about the interpretation of the constants from M.
- Given any $\alpha : M \longrightarrow A$ non-expansive we can turn $\mathcal{A} = (A, d)$ into an algebra in $\mathbb{K}(\Omega_M, \mathcal{U}_M)$ by interpreting each $m \in M$ as $\alpha(m) \in A$.



$\mathcal{U}_M$ is consistent if and only if the map η_M is an isometry.

We have a monad on **Met**.

Birkhoff Variety Theorems

- Three kinds of equations: (a) unconditional equations
- (b) basic equations : assumptions of the form $x =_\varepsilon y$, x, y variables.
- (c) Horn clauses, assumptions may involve terms.
- Usual variety theorem says: a class of algebras is equationally definable if and only it it is closed under products, homomorphic images and subalgebras.
- We have to consider a new kind of closure property.

Reflexive homomorphisms

- A $\mathfrak{c}$-reflexive homomorphism f between QA's $\mathcal{A}, \mathcal{B}$, where $\mathfrak{c}$ is a cardinal number, is a homomorphism with the property that for any subset $B' \subset B$ with $|B'| < \mathfrak{c}$, there is a subset $A' \subset A$ with $f(A') = B'$ and f restricted to A' is an *isometry*.
- If $\mathcal{U}$ is an unconditional theory then $\mathbb{K}(\Omega, \mathcal{U})$ is closed under homomorphic images.
- If $\mathcal{U}$ is a basic equational theory with every conditional equation having only finitely many assumptions then $\mathbb{K}(\Omega, \mathcal{U})$ is closed inder $\aleph_0$-reflexive homomorphisms.
- If $\mathcal{U}$ is a basic equational theory then $\mathbb{K}(\Omega, \mathcal{U})$ is closed inder $\aleph_1$-reflexive homomorphisms.
- A $\mathfrak{c}$-variety is a class of algebras closed under products, subalgebras and $\mathfrak{c}$-reflexive homomorphisms.
- A $\mathfrak{c}$-equational class is a class of algebras defined by $\mathfrak{c}$-basic conditional equations.

The main theorem

$\mathcal{K}$ is a $\mathfrak{c}$-variety if and only if it is a $\mathfrak{c}$-basic equational class.

- $\mathcal{K}$ is an unconditional equational class iff it is a variety.
- $\mathcal{K}$ is a finitary-basic equational class iff it is an $\aleph_0$-variety.
- $\mathcal{K}$ is a basic equational class iff it is an $\aleph_1$-variety.

Equational reasoning for probabilistic programming
Part III: The Kantorovich metric and cousins

Prakash Panangaden[1]

[1] School of Computer Science
McGill University

Probabilistic Proramming Languages 29th May - 2nd June 2017

Barycentric algebras again

- $\Omega = \{+_e : 2|e \in [0, 1]\}$; uncountably many operations!
- $(\mathbf{B1}) \quad \emptyset \vdash x +_1 y =_0 x$
- $(\mathbf{B2}) \quad \emptyset \vdash x +_e x =_0 x$
- $(\mathbf{SC}) \quad \emptyset \vdash x +_e y =_0 y +_{1-e} x$
- $(\mathbf{SA})(x +_{e_1} y) +_{e_2} z =_0 x +_{e_1 e_2} (y +_{\frac{e_2 - e_1 e_2}{1 - e_1 e_2}} z)$ where $e_1, e_2 \in (0, 1)$
- $(\mathbf{LI}) \quad x +_e z =_\varepsilon y +_e z$ where $e \leqslant \varepsilon \in \mathbb{Q} \cap [0, 1]$
- The last equation uses one of the new indexed equations in a nontrivial way.
- We call it the *left-invariant* axiom scheme; LIB algebras for short.
- What does this axiomatize?
- The total variation metric on probability distributions.

Total variation metric

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

- It measures the size of the set on which p, q disagree the most.
- There is a duality theorem that gives it as a minimum rather than a maximum.

Couplings

- Let $\mathcal{B}(M,\Sigma)$ be the Borel measures on a metric space M with Borel algera Σ.
- We have a product space $M \times M$ with product σ-algebra $\Sigma \otimes \Sigma$ and Borel measures $\mathcal{B}(M \times M, \Sigma \otimes \Sigma)$.
- Given probability measures p, q a *coupling* is a probability measure ω on $(M \times M, \Sigma \otimes \Sigma)$ such that for all $E \in \Sigma$:

$$\omega(E \times M) = p(E) \quad \text{and} \quad \omega(M \times E) = q(E).$$

- $\mathcal{C}(p,q)$ is the set of couplings for (p,q).
- Write Δ for the diagonal in $M \times M$.
- TV duality: $TV(p,q) = \min\{\omega(\Delta^c) | \omega \in \mathcal{C}(p,q)\}$; min is attained.
- Convex combinations of couplings are couplings.
- Splitting lemma: If p, q are Borel probability measures on M and $e = T(p,q)$. There are p', q', r such that

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

Freely generated LIB algebra

- We know there is a freely generated LIB algebra from a metric space M. What is it concretely?
- Let $\Pi[M]$ be the LIB algebra obtained by taking the *finitely-supported* probability measures on M and interpreting $+_e$ as convex combination.
- We endow it with the total-variation metric to make it a quantitative algebra.
- Theorem: $\Pi[M] \in \mathbb{K}(\mathcal{B}, \mathcal{U}^{LI})$.
- Use convexity and splitting lemma to show LI and Nexp.
- Theorem: $\Pi[M]$ is the free algebra generated by M.
- Use the embedding of convex spaces into vector spaces (Stone 49).
- The axioms give rise to the total-variation metric.

Interpolative barycentric algebras

- Same signature as barycentric algebras.
- Axioms (B1), (B2), (SC), (SA); drop (LI).
- $(\mathbf{IB}_p)$

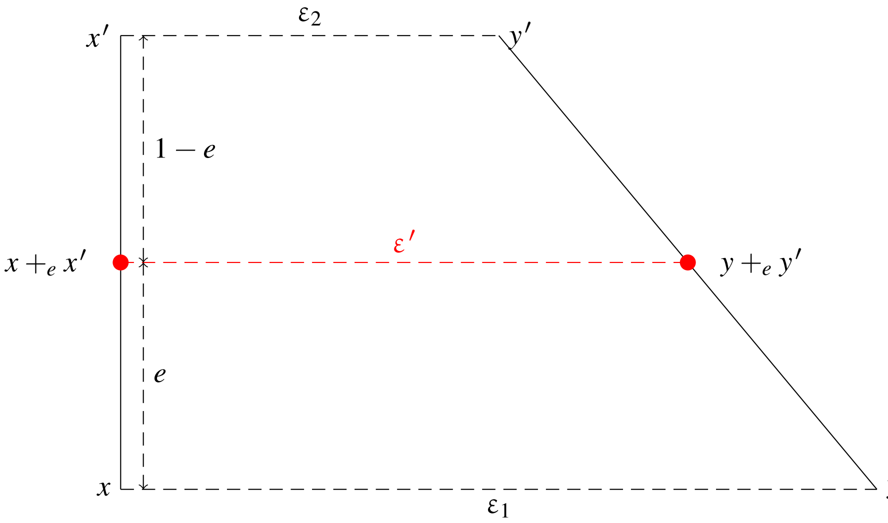
$$\{x =_{\varepsilon_1} y, x' =_{\varepsilon_2} y'\} \vdash x +_e x' =_\delta y +_e y',$$

  where $(e\varepsilon_1^p + (1-e)\varepsilon_2^p)^{1/p} \leqslant \delta$.
- Now we need assumptions in the equation.
- If $p = 1$ we get

$$\{x =_{\varepsilon_1} y, x' =_{\varepsilon_2} y'\} \vdash x +_e x' =_\delta y +_e y',$$

  where $e\varepsilon_1 + (1-e)\varepsilon_2 \leqslant \delta$.

Picture of IB_1



Kantorovich-Wasserstein metric

Let (M,d) be a complete separable metric space and $p \geqslant 1$.

Wasserstein-p metric

$$W_d^p(\mu,\nu) = \inf\left\{ \left[\int_{M\times M} d^p(x,y)\mathrm{d}\omega\right]^{1/p} \middle| \omega \in \mathcal{C}(\mu,\nu) \right\}$$

Kantorovich

$$K_d(\mu,\nu) = \sup\left\{ \left| \int f\mathrm{d}\mu - \int f\mathrm{d}\nu \right| \right\}$$

Duality

$$K_d(\mu,\nu) = \min\left\{ \left[\int_{M\times M} d(x,y)\mathrm{d}\omega\right] \middle| \omega \in \mathcal{C}(\mu,\nu) \right\}$$

Finitary case

- We take the finitely supported measures on M and interpret it as a barycentric algebra as before.
- We give it the Wasserstein metric and show that we get an IB algebra.
- This uses the definition of the W_d^p metrics as an inf and convexity of couplings.
- We prove a splitting lemma for this case and show that we get the free algebra by similar, but more involved arguments.
- How do we lift it to the continuous case?

Weak convergence

- Suppose we have a sequence of measures $\{\mu_i | i \in I\}$. What does it mean to converge?
- For a "suitable" class of functions:

$$\int f \mathrm{d}\mu_i \rightarrow \int f \mathrm{d}\mu.$$

- For Kantorovich use contractive functions; for Wasserstein use a class of functions whose growth is controlled by d and p.
- The Wasserstein metrics give the topology of weak convergence.
- The finitely supported probability measures are *dense* in the space of all probability measures.

Complete separable metric spaces

- A separable metric space has a countable dense subset.
- Define $\Delta[M]$ to be the space of all Borel probability measures on a complete separable metric space. We give it the W_d^p metric and interpret $+_e$ as convex combination.
- This gives an IB algebra.
- If we construct the term algebra $\mathbb{T}[M]$ as before and *complete it* we get an algebra isomorphic to $\Delta[M]$.
- In this case we get a monad on $\mathbf{CSMet_1}$: complete separable 1-bounded metric spaces.

Conclusions

- Quantitative equations give a handle on otherwise arcane things like the Wasserstein metrics.
- Other examples: Hausdorff metric, pointed barycentric algebras.
- To do; many more examples:
    - Markov processes
    - Choquet capacities and games
    - quantitative theory of effects
    - quantitative equational axioms for probabilistic programming languages.