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

Prakash Panangaden[1]

[1]School of Computer Science
McGill University

Probabilistic Programming 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 $\mathbf{Set}$
- 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 $\mathbf{Met}$.
- 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:

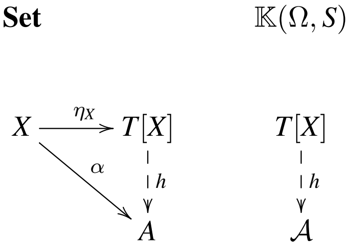
$$\{+_\epsilon | \epsilon \in [0,1]\}$$

- 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!

- 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.

- 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

- $\mathbf{Mes}$: 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.