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

Prakash Panangaden[1]

[1]School of Computer Science
McGill University

Probabilistic Programming 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

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

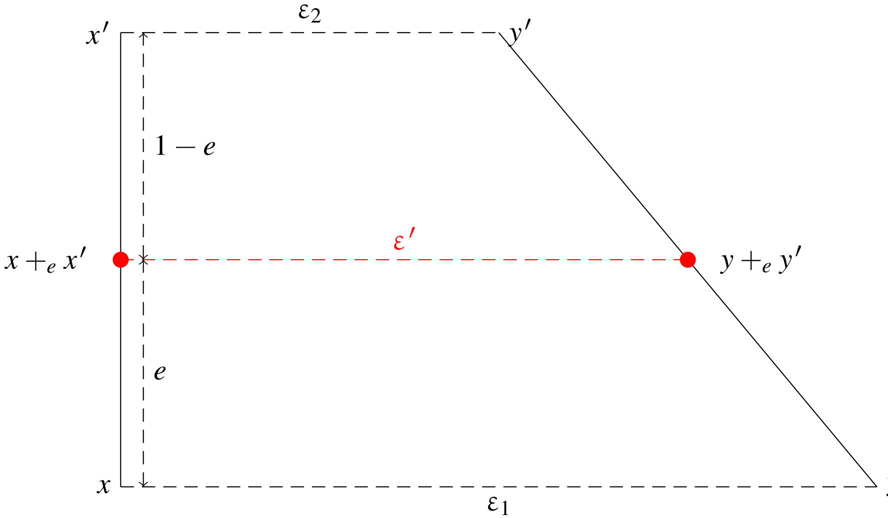
Freely generated LIB algebra

- 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:

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

- 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.