Analysis of Probabilistic Systems

Boot camp Lecture 3: Probabilistic relations

Prakash Panangaden[1]

[1]School of Computer Science
McGill University

Fall 2016, Simons Institute

Outline

Ordinary binary relations

- X, Y, Z sets and R, S relations.
- $R : X \longrightarrow Y$, $S : Y \longrightarrow Z$; $R \subseteq X \times Y$, $S \subseteq Y \times Z$.
- Compose $S \circ R \subseteq X \times Z$ given by $x(S \circ R)z$ iff $\exists y \in Y, xRy \wedge ySz$.
- Composition is associative.
- Rich algebraic structure: Kleene algebra, a compact closed category, the Kleisli category of the powerset monad.

The category $\mathbf{Rel}$

- Objects: sets, an arrow X to Y is a binary relation $R \subseteq X \times Y$.
- Composition is relational composition. The identity on X is $I_X = \{\langle x, x\rangle : x \in X\}$; also often written as Δ.
- Shares many properties of the category of finite-dimensional vector spaces: tensor structure, transpose, trace, *etc.*
- Any relation $R \subseteq X \times Y$ corresponds uniquely to a function $f_R : X \longrightarrow \mathcal{P}(Y)$: $f_R(x) = \{y : xRy\}$.
- Any function $f : X \longrightarrow \mathcal{P}(Y)$ defines a binary relation.
- Any function $f : X \longrightarrow \mathcal{P}(Y)$ can be *lifted* to a function $f^\dagger : \mathcal{P}(X) \longrightarrow \mathcal{P}(Y)$ where
- $f^\dagger(A \subseteq X) = \bigcup_{a \in A} f(a)$. We usually write just $f(A)$.
- Composition of these functions is exactly the same as relational composition.
- Intimate relation between $\mathbf{Set}$, $\mathbf{Rel}$ and the powerset operation.

Powerset categorically

- Recall $\mathbf{Set}$ the category of sets and functions.
- We have a *functor* $\mathcal{P}() : \mathbf{Set} \longrightarrow \mathbf{Set}$ which $X \mapsto \mathcal{P}(X)$ and
- $f : X \longrightarrow Y \mapsto \mathcal{P}(f)$ defined by $\mathcal{P}(f)(A \subseteq X) = f(A)$.
- This has extra structure: $\forall X, \exists \eta_X : X \longrightarrow \mathcal{P}(X)$ given by $\eta_X(x) = \{x\}$.
- $\forall X, \exists \mu_X : \mathcal{P}(\mathcal{P}(X)) \longrightarrow \mathcal{P}(X)$ given by union.
- These maps are very “natural”!

Naturality

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

Monads

- $\mathcal{C}$ a category, $T : \mathcal{C} \longrightarrow \mathcal{C}$ an endofunctor.
- Natural transformations: $\eta : I \longrightarrow T$ and $\mu : T^2 \longrightarrow T$ satisfying
- $$\begin{array}{ccc} T^3(X) & \xrightarrow{T\mu_X} & T^2(X) \\ {\scriptstyle \mu_{T(X)}}\downarrow & & \downarrow{\scriptstyle \mu_X} \\ T^2(X) & \xrightarrow[\mu_X]{} & T(X) \end{array} \quad \text{and} \quad \begin{array}{ccc} T(X) & \xrightarrow{T(\eta_X)} & T^2(X) \\ {\scriptstyle \eta_{T(X)}}\downarrow & \diagdown\!\!\diagdown & \downarrow{\scriptstyle \mu_X} \\ T^2(X) & \xrightarrow[\mu_X]{} & T(X) \end{array}$$

The Kleisli category of a monad

- Monads are the main ingredient in the categorical treatment of universal algebra. The Eilenberg-Moore category.
- But it will take us too far afield to do that.
- Given a monad $T : \mathcal{C} \longrightarrow \mathcal{C}$ we define a new category $\mathcal{C}_T$: objects are the same as in $\mathcal{C}$. An arrow from X to Y in $\mathcal{C}_T$ is an arrow of $\mathcal{C}$ from X to TY.
- Composition? $f : X \longrightarrow TY$, $g : Y \longrightarrow TZ$.
- Apply T to g to get $Tg : TX \longrightarrow T^2Y$, compose with f to get $Tg \circ f : X \longrightarrow T^2Z$.
- Now compose with $\mu_Z : T^2Z \longrightarrow TZ$ to get an arrow $X \longrightarrow TZ$.
- Doing this with $\mathcal{P}(\cdot)$ is exactly what we did before and gives relational composition.
- Thus, the Kleisli category of the powerset monad is $\mathbf{Rel}$.

Measurable spaces as a category

- **Mes**: objects (X, Σ) measurable spaces, arrows measurable functions.
- In many ways, just like **Set**: products, coproducts, pushouts etc.
- In fact it is a topological concrete category over **Set** so is complete and cocomplete.
- What about adding probability measures?
- **Prob**: objects (X, Σ, P) and arrows are measurable and *measure preserving*.
- $f : (X, \Sigma, P) \longrightarrow (Y, \Lambda, Q)$ satisfies $Q = P \circ f^{-1}$.
- This category does not have products.
- Many variations are possible.
- Slogan: put the probabilities into the arrows.

The Lawvere-Gíry monad I

- The underlying category is $\mathbf{Mes}$.
- $\Gamma : \mathbf{Mes} \longrightarrow \mathbf{Mes}$ defined by $\Gamma(X, \Sigma) = \{\nu : \nu$ a prob. measure on $\Sigma\}$.
- But $\Gamma(X)$ needs a σ-algebra structure on it too.
- $\forall A \in \Sigma$, define $ev_A : \Gamma(X) \longrightarrow [0, 1]$ by $ev_A(\nu) = \nu(A)$.
- Endow $\Gamma(X)$ with the *smallest* σ-algebra making all the p_A measurable.
- Given $f : (X, \Sigma) \longrightarrow (Y, \Lambda)$ define $\Gamma(f)(\nu) = \nu \circ f^{-1}$.
- $\eta_X : X \longrightarrow \Gamma(X)$ is $x \mapsto \delta_x$
- This is like a “probabilistic powerset” with a singleton embedding.
- But what is the analogue of union?

The Lawvere-Gíry monad II

- We need $\mu_X : \Gamma(\Gamma(X)) \longrightarrow \Gamma(X)$.
- Let $\Omega \in \Gamma(\Gamma(X))$ *i.e.* a measure on $\Gamma(X)$. $\mu_X(\Omega)$ has to be a measure on (X, Σ).
- It has to assign a number to any measurable set $A \in \Sigma$.
- Recall that $ev_A : \Gamma(X) \longrightarrow [0, 1]$ is a measurable function.
- $\mu_X(\Omega)(A) = \int_{\Gamma(X)} ev_A \mathrm{d}\Omega$.
- Checking the various equations is readily done with the monotone convergence theorem.

The Kleisli category – probabilistic relations

- The objects are measurable spaces (X, Σ).
- A map h from (X, Σ) to (Y, Λ) is a measurable function from X to $\Gamma(Y)$.
- $h : X \longrightarrow (\Lambda \longrightarrow [0, 1])$ or, currying, we get
- $h : X \times \Lambda \longrightarrow [0, 1]$. Write it as $h(x, B)$ for $x \in X$ and $B \in \Lambda$.
- $h(\cdot, B)$ is measurable and $h(x, \cdot)$ is a probability measure.
- In fact these are precisely disintegrations, aka regular conditional probability densities aka Markov kernels.
- Composition $h : X \longrightarrow Y, k : Y \longrightarrow Z$ and $(k \circ h)(x, C) = \int_Y k(y, C) \mathrm{d}h(x, \cdot)$.
- This is the analogue of relational composition with integration replacing $\exists$.
- In finite discrete spaces this is just the formula for matrix multiplication.
- We will call this category **SRel** for "stochastic relations."

Syntax

Kozen's Language

$$S ::== x_i := f(\vec{x}) | S_1; S_2 | \textit{if } \mathbf{B} \textit{ then } S_1 \textit{ else } S_2 | \textit{while } \mathbf{B} \textit{ do } S.$$

- There are a fixed set of variables $\vec{x}$ taking values in a measurable space (X, Σ_X).
- f is a measurable function.
- B is a measurable subset.

Outline of the semantics

- State transformer semantics: distribution (measure) transformer semantics.
- Meaning of statements: Markov kernels *i.e.* **SRel** morphisms.
- The only subtle part: how to give fixed-point semantics to the while loop?

A tiny variation of Gìry's monad

- Instead of Γ we define $\Pi : (X, \Sigma) = \{\nu | \nu(X) \leq 1\}$.
- Other structure defined exactly as in Γ.
- We will use $\mathbf{SRel}$ to stand for the Kleisli category of Π.
- Why make this tiny change?

Partially additive structure

- Now we can "add" **SRel** morphisms!
- Not always, the sum may exceed 1, but we can define *summable families* which may even be countaby infinite.
- The homsets of **SRel** form *partially additive monoids*.
- The sums can be rearranged at will (partition-associativity).
- Limit property: If F is a countable family in which every *finite* subfamily is summable then F is summable.
- In the category **SRel**, the sums interact properly with composition.
- If $\{f_i \mid i \in \mathbb{N}\}$ is a countable set of morphisms from X to Y and there is a morphism $f : X \longrightarrow (Y + Y + \ldots)$ such that when projected onto the X's we get the f_i, then the family is summable.

Why this structure matters

Arbib and Manes

Given a partially additive category $\mathcal{C}$ and $f : X \longrightarrow X + Y$ we can find a unique pair $f_! : X \longrightarrow X$ and $f_2 : X \longrightarrow Y$ such that $f = \iota_1 \circ f_1 + \iota_2 \circ f_2$. Furthermore, there is a morphism $f^* : X \longrightarrow Y$ given by

$$f^* = \sum_{n=0}^{\infty} f_2 \circ f_1^n.$$

The theorem says that the family $f_2 \circ f_1^n$ is summable.
It is the *iterate* of f.

Semantics of Kozen's Language I

- Statements are $\mathbf{SRel}$ morphisms of type $(X^n, \Sigma^n) \longrightarrow (X^n, \Sigma^n)$.
- **Assignment**: $x := f(\vec{x})$

$$[\![x_i := f(\vec{x})]\!](\vec{x}, \vec{A}) = \delta(x_1, A_1) \ldots \delta(x_{i-1}, A_{i-1}) \delta(f(\vec{x}), A_i) \delta(x_{i+1}, A_{i+1}) \ldots$$

- **Sequential Composition**: $S_1; S_2$

$$[\![S_1; S_2]\!] = [\![S_2]\!] \circ [\![S_1]\!]$$

  where the composition on the right hand side is the composition in $\mathbf{SRel}$.
- **Conditionals**: $\mathit{if}\ \mathbf{B}\ \mathit{then}\ S_1\ \mathit{else}\ S_2$

$$[\![\mathit{if}\ \mathbf{B}\ \mathit{then}\ S_1\ \mathit{else}\ S_2]\!](\vec{x}, \vec{A}) = \delta(\vec{x}, \mathbf{B})[\![S_1]\!](\vec{x}, \vec{A}) + \delta(\vec{x}, \mathbf{B}^c)[\![S_2]\!](\vec{x}, \vec{A})$$

Semantics of Kozen's Language II

While Loops: *while* $\mathbf{B}$ *do* S

$$[\![while\ \mathbf{B}\ do\ S]\!] = h^*$$

where we are using the $*$ in $\mathbf{SRel}$ and the morphism

$$h : (X^n, \Sigma^n) \longrightarrow (X^n, \Sigma^n) + (X^n, \Sigma^n)$$

is given by

$$h(\vec{x}, \vec{A_1} \uplus \vec{A_2}) = \delta(\vec{x}, \mathbf{B})[\![S]\!](\vec{x}, \vec{A_1}) + \delta(\vec{x}, \mathbf{B}^c)\delta(\vec{x}, \vec{A_2}).$$

Weakest precondition semantics

- We can construct a category of probabilistic predicate transformers: $\mathbf{SPT}$.
- Objects are measurable spaces.
- Given (X, Σ_X) we can construct the (Banach) space of bounded measurable functions on X (the "predicates") $\mathcal{F}(X)$.
- A morphism $X \longrightarrow Y$ in $\mathbf{SPT}$ is a bounded (continuous) linear map from $\mathcal{F}(X)$ to $\mathcal{F}(Y)$.
- $$\mathbf{SPT} \simeq \mathbf{SRel}^{op}.$$
- This gives us the structure needed for a $\mathbf{wp}$ semantics.