Discrete Probabilistic Programming

M2 MPRI 2023-2024

Guillaume Baudart guillaume.baudart@inria.fr
Guillaume Geoffroy guillaume.geoffroy@irif.fr
Xavier Rival xavier.rival@ens.fr
Christine Tasson christine.tasson@isae-supaero.fr

Discrete Probabilistic Programming

In the preceding lectures:

- Inference is intractable in general
- First order semantics of discrete and continuous PPL.

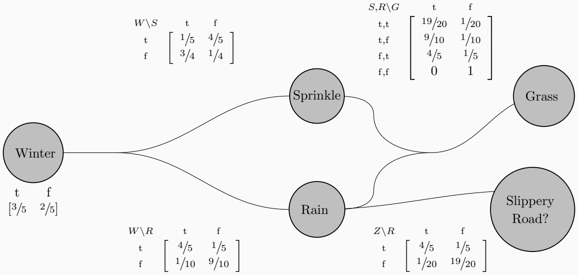
In this lecture: Discrete Setting

- Under some conditions, exact inference is possible : Bayesian Network
- Semantics of Higher-Order CBV Discrete Probabilistic Lambda-Calculus
- Exam 2021: Probabilistic Hoare Logic

Bayesian Networks

Bayseian Network - Introduction

Example



Purpose: compact representation of the joint distribution $\mathbb{P}(G, Z, S, R, W)$

- Variables X and their sample sets (or carrier or web) $|X| = \{\mathtt{t}, \mathtt{f}\}$
- Conditional Probability Tables (CPT): $\mathbb{P}(X|\mathtt{Pa}(X)) : |\mathtt{Pa}(X)| \times |X| \to \mathbb{R}^+$
- Dependence to parents $\mathtt{Pa}(X)$ in the DAG and independence to other variables.

The function mass $\mathbb{P}(G = g, Z = z, S = s, R = r, W = w)$ has dimension 2^5.

The CPTs: $\mathbb{P}(G = g|S = s, R = r)$: 2^3 and the others $3 * 2^2 + 2$

Bayesian Network - Definition

Definition A bayesian network is given by

- A DAG
- Labeled by variables and conditional probability tables (CPT)

The graph defines Dependence and Independence conditions

- Parents: $\mathtt{Pa}(G) = \{S, R\}$ and $\mathtt{Pa}(S) = W$
- The probability of X given variables depends only on $\mathtt{Pa}(X)$:

$$\mathbb{P}(X|\mathtt{Vars}) = \mathbb{P}(X|\mathtt{Pa}(X))$$

Bayesian Network - Properties

Joint distribution from conditional probability tables

$$\mathbb{P}(Z, G, R, S, W) = \mathbb{P}(Z|R)\,\mathbb{P}(G|S, R)\,\mathbb{P}(S|W)\,\mathbb{P}(R|W)\,\mathbb{P}(W)$$

- **Conditional probability**

$$\mathbb{P}(Z, G, R, S, W) = \mathbb{P}(Z|G, S, R, W)\ \mathbb{P}(G, S, R, W)$$

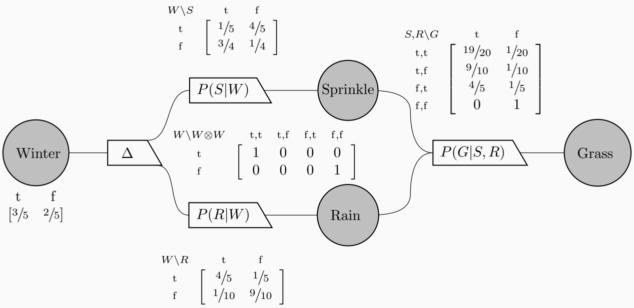
- **Chain Rule**

$$\mathbb{P}(Z, G, R, S, W) = \mathbb{P}(Z|G, S, R, W)\,\mathbb{P}(G|S, R, W)\,\mathbb{P}(S|R, W)\,\mathbb{P}(R|W)\,\mathbb{P}(W)$$

- **Dependence**

$$\mathbb{P}(Z|G, S, R, W) = \mathbb{P}(Z|R) \quad \mathbb{P}(G|S, R, W) = \mathbb{P}(G|S, R) \quad \mathbb{P}(S|R, W) = \mathbb{P}(S|W)$$

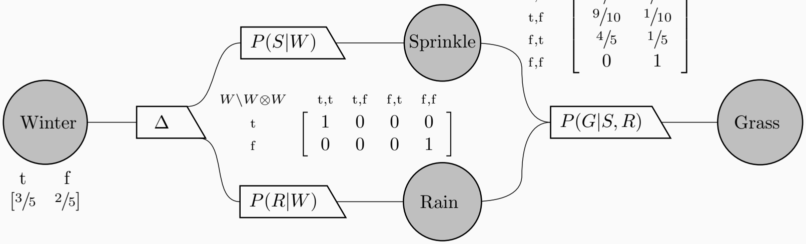
Bayesian Network - Example



Compute $\mathbb{P}(G)$ using:

- **Joint distribution:** $\mathbb{P}(G, R, S, W) = \mathbb{P}(G|S, R)\ \mathbb{P}(S|W)\ \mathbb{P}(R|W)\ \mathbb{P}(W)$
- **Marginal** $\mathbb{P}(G) = \displaystyle\sum_{(r,s,w)\in|R|\times|S|\times|W|} \mathbb{P}(G, R, S, W)$

Bayesian Network - Example



$$\begin{array}{c|cc} & \text{t} & \text{f} \\ \hline & \left[\begin{matrix} 3/5 & 2/5 \end{matrix}\right] \end{array}$$

$$\begin{array}{cc} W \backslash S & \begin{matrix} \text{t} & \text{f} \end{matrix} \\ \begin{matrix} \text{t} \\ \text{f} \end{matrix} & \left[\begin{matrix} 1/5 & 4/5 \\ 3/4 & 1/4 \end{matrix}\right] \end{array}$$

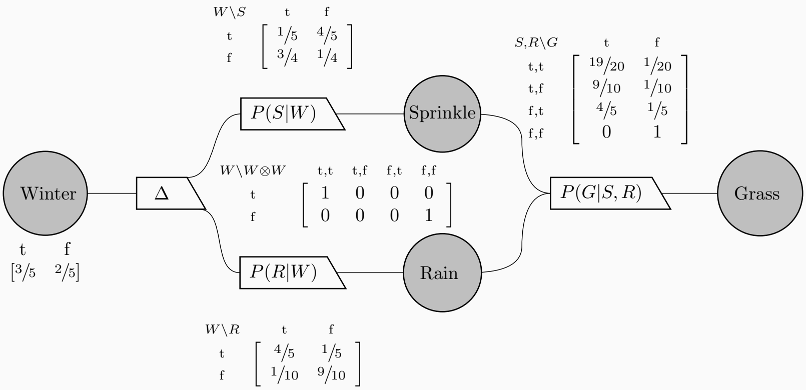
$$\begin{array}{cc} W \backslash W \otimes W & \begin{matrix} \text{t,t} & \text{t,f} & \text{f,t} & \text{f,f} \end{matrix} \\ \begin{matrix} \text{t} \\ \text{f} \end{matrix} & \left[\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{matrix}\right] \end{array}$$

$$\begin{array}{cc} W \backslash R & \begin{matrix} \text{t} & \text{f} \end{matrix} \\ \begin{matrix} \text{t} \\ \text{f} \end{matrix} & \left[\begin{matrix} 4/5 & 1/5 \\ 1/10 & 9/10 \end{matrix}\right] \end{array}$$

$$\begin{array}{cc} S,R \backslash G & \begin{matrix} \text{t} & \text{f} \end{matrix} \\ \begin{matrix} \text{t,t} \\ \text{t,f} \\ \text{f,t} \\ \text{f,f} \end{matrix} & \left[\begin{matrix} 19/20 & 1/20 \\ 9/10 & 1/10 \\ 4/5 & 1/5 \\ 0 & 1 \end{matrix}\right] \end{array}$$

$$p(S) = \left(\sum_{a \in \{\text{t,f}\}} P(S|W)_{a,b} \cdot p(W)_a \right)_{b \in \{\text{t,f}\}}$$

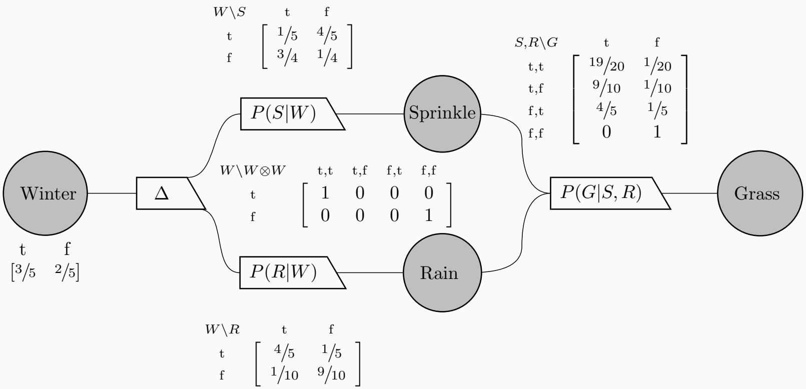
Bayesian Network - Example



$$
\begin{aligned}
p(W)\ P(S|W) &= p(S) \\
p(W)\ P(R|W) &= p(R)
\end{aligned}
$$

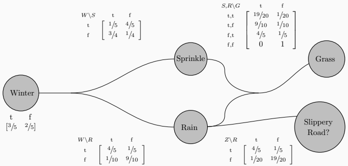
and $(p(S)\otimes p(R))\ P(G|S,R) = p(G)$

Bayesian Network - Example



$$p(W)\ \Delta\ (P(S|W) \otimes P(R|W))\ P(G|S,R)\ =\ p(G)$$

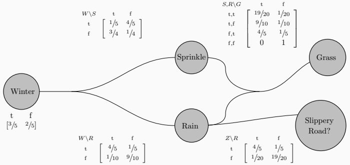
Bayesian Network - Computing Marginals



Using Chain Rule naive methods with 2^5 operations

$$\begin{aligned}
\mathbb{P}(Z, G) &= \sum_{w,s,r} \mathbb{P}(Z, G, S, R, W) \\
&= \sum_{w,s,r} \mathbb{P}(Z|R)\,\mathbb{P}(G|S, R)\,\mathbb{P}(S|W)\,\mathbb{P}(R|W)\,\mathbb{P}(W)
\end{aligned}$$

Bayesian Network - Computing Marginals



Using Chain Rule naive methods with 2^5 operations

$$\begin{aligned}\mathbb{P}(Z,G) &= \sum_{w,s,r} \mathbb{P}(Z,G,S,R,W) \\ &= \sum_{w,s,r} \mathbb{P}(Z|R)\,\mathbb{P}(G|S,R)\,\mathbb{P}(S|W)\,\mathbb{P}(R|W)\,\mathbb{P}(W)\end{aligned}$$

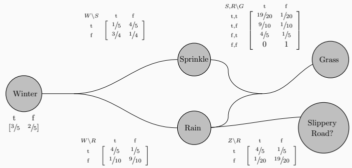
Using *Memoization* and *Factorization* Optimized methods using 2^4 operations.

$$\mathbb{P}(Z,G) = \sum_{r} \mathbb{P}(Z|R) \left(\sum_{s} \mathbb{P}(G|S,R) \left(\sum_{w} \mathbb{P}(S|W)\,\mathbb{P}(R|W)\,\mathbb{P}(W) \right) \right)$$

Variable Elimination

An algorithm for exact inference, based on a chosen order to eliminate variables (i.e. marginalize): in the following factorization, the order is W then S then R.

$$\mathbb{P}(Z,G) = \sum_r \mathbb{P}(Z|R)\left(\sum_s \mathbb{P}(G|S,R)\left(\sum_w \mathbb{P}(S|W)\mathbb{P}(R|W)\mathbb{P}(W)\right)\right)$$

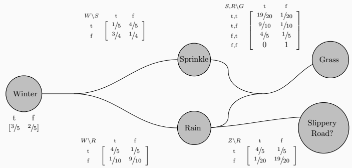


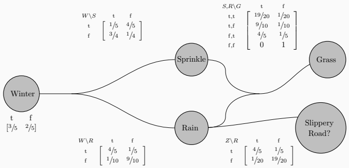
- Different orders have different perfomances
- Choosing an optimal order is known as NP-hard (see Cooper, The computational complexity of probabilistic inference using bayesian belief networks, Arti cial Intelligence 42 (1990), no. 2, 393–405)
- Many heuristic depending on the structure of the graph

Variable Elimination

In a BN, variable elimination corresponds to the replacement of a subgraph with a **factor**.

$$
\begin{aligned}
\mathbb{P}(Z,G) &= \sum_r \mathbb{P}(Z|R)\left(\sum_s \mathbb{P}(G|S,R)\left(\sum_w \mathbb{P}(S|W)\,\mathbb{P}(R|W)\,\mathbb{P}(W)\right)\right)\\
&= \sum_r \mathbb{P}(Z|R)\left(\sum_s \mathbb{P}(G|S,R)m_W(S,R)\right)\\
&= \sum_r \mathbb{P}(Z|R)\ m_S(R,G)\\
&= m_R(Z,G)
\end{aligned}
$$





Variable Elimination for marginal inference - Message Passing Algorithm

1. Compute the distance from the query variables
2. Factors are morphisms/CPT labelling edges (output edges are labelled with 1)
3. Variables are types/Nodes in the graph
4. Propagate Messages following distance order of edges:
   - Message from variable to factors: $\nu_{x \to f} = \prod_{g \in \mathtt{ne}(x) \setminus f} \mu_{g \to x}$
   - Message from factor to variable: $\mu_{f \to x} = \sum_{X_f \setminus x} \prod_{y \in \mathtt{ne}(f) \setminus x} \nu_{y \to f}$

Higher-Order Syntax and Semantics for PPL Discrete case

A Call-by-value probabilistic λ-calculus

Ground types: $b ::= \texttt{unit} \mid \texttt{bool} \mid \texttt{int} \mid b \times b$

Types: $t ::= b\ \texttt{dist} \mid t \otimes t \mid t \to t$

Constants: $c ::= () \mid \texttt{true} \mid \texttt{false} \mid \texttt{n} \mid (c, c) \mid f(c, \dots, c)$ where f is a ground operator

Values: $v ::= \texttt{Dirac}\ c \mid \lambda x.e$

Terms: $e ::= \texttt{Dirac c} \mid \texttt{Bernoulli r} \mid \texttt{Dice n}$
$x \mid (e, e) \mid \lambda x.e \mid (e)e \mid \text{let } x = \text{ Sample } e \text{ in } e$

Contexts: $G ::= 1 \mid G_0, x : t$ where $x \notin G_0$

$\boxed{G \vdash e : t}$ Type System

$$\frac{}{G \vdash () : \text{unit}} \qquad \frac{}{G \vdash \underline{\texttt{true}} : \text{bool}} \qquad \frac{}{G \vdash \underline{\texttt{false}} : \text{bool}} \qquad \frac{n \in \mathbb{N}}{G \vdash \underline{n} : \text{int}} \qquad \frac{f \in b_1 \times \cdots \times b_n \to b'}{G \vdash f(\vec{c}) : b'}$$

$$\frac{r \in \mathbb{R}}{G \vdash \text{Bernoulli } r : \text{bool dist}} \qquad \frac{n \in \mathbb{N}}{G \vdash^P \text{Dice } n : \text{int dist}} \qquad \frac{G \vdash c : b}{G \vdash \text{Dirac}(c) : b \text{ dist}}$$

$$\frac{G(x) = t}{G \vdash x : t} \qquad \frac{G \vdash e_1 : t_1 \qquad G \vdash e_2 : t_2}{G \vdash (e_1,\ e_2) : t_1 \otimes t_2}$$

$$\frac{G + [x : t_1] \vdash e : t_2}{G \vdash \lambda x.e : t_1 \to t_2} \qquad \frac{G \vdash e_2 : t_1 \to t_2 \qquad G \vdash e_1 : t_1}{G \vdash (e_2)e_1 : t_2} \qquad \frac{G \vdash e_1 : b \text{ dist} \qquad G + [x : b] \vdash e_2 : t_2}{G \vdash \text{let } x = \text{ Sample } e_1 \text{ in } e_2 : t_2}$$

Operational Semantics (discrete case) Labeled Transition System (LTS)

$$\frac{}{\text{sample}(\text{Bernoulli } r) \xrightarrow{r} \underline{\texttt{true}}}$$

$$\frac{}{\text{sample}(\text{Bernoulli } r) \xrightarrow{1-r} \underline{\texttt{false}}}$$

$$\frac{\forall i \in \{1, \ldots, n\}}{\text{sample}(\text{Dice } n) \xrightarrow{\frac{1}{n}} \underline{i}}$$

$$\frac{}{\text{sample}(\text{Dirac } c) \xrightarrow{1} c}$$

$$\frac{}{\text{let } x = c \text{ in } e_2 \xrightarrow{1} e_2[\text{Dirac c}/x]}$$

$$\frac{e_1 \xrightarrow{r} e_1'}{\text{let } x = e_1 \text{ in } e_2 \xrightarrow{r} \text{let } x = e_1' \text{ in } e_2}$$

$$\frac{}{(\lambda x.e_2)v \xrightarrow{1} e_2[v/x]}$$

$$\frac{e_2 \xrightarrow{r} e_2'}{(e_2)e_1 \xrightarrow{r} (e_2')e_1}$$

$$\frac{e_1 \xrightarrow{r} e_1'}{(e_2)e_1 \xrightarrow{r} (e_2)e_1'}$$

Denotational semantics of Higher Order

Discrete Probabilistic Programming

Model of Linear Logic

SMCC - Symmetric Monoidal Closed Category

- Category : $\mathcal{C}$ objects + morphisms, identity and associative composition
- Symmetric Monoidal : $1, A \otimes B$ + unit, associative and commutative laws
- Closed: $A \multimap B, \Lambda, \mathrm{ev}$ + currying and evaluation are in tightly related, $\dfrac{A \otimes B \to C}{A \to B \multimap C}$.

Comonad

- $! : \mathcal{C} \to \mathcal{C}$ a functor (action on objects and morphisms)
- counit $\epsilon_A : !A \to A$ and comultiplication $\delta_A : !A \to !!A$ + diagrams.

Monoidal Strength:

- natural isomorphisms $!T \xrightarrow{\sim} 1$ and $!(A \times B) \xrightarrow{\sim} !A \otimes !B$ + coherence diagrams

Commutative Comonoid

- weakening $w_A : !A \to 1$ and contraction $c_A : !A \to !A \otimes !A$ + coherence diagrams

Call-By-Value in models of LL

Interpretation of types

- Ground types: $Z^* = !Z$
- Function types: $(A \to B)^* = !(A^* \multimap B^*)$
- Types A^* are preceded by !, thus

$$\begin{array}{ccccc} A^* & \xrightarrow{\delta} & !A^* & & \\ A^* & \xrightarrow{c} & A^* \otimes A^* & \qquad A^* & \xrightarrow{w} \; 1 \end{array}$$

- Contexts: $(A_1, \ldots, A_n)^* = {A_1}^* \otimes \cdots \otimes {A_n}^*$

Interpretation of terms: $G \vdash e : A$ is interpreted as a morphism $e^* : G^* \to A^*$

See call-by-name, call-by-value, call-by-need, and the Linear Lambda Calculus, Maraist & al.

Semantics of CBV in LL

Interpretation of terms: $G \vdash e : A$ is interpreted as a morphism $e^* : G^* \to A^*$

Variable:

$$G, x : t \vdash x : t \qquad G^* \otimes t^* \xrightarrow{w_{G^*} \otimes \mathrm{id}} 1 \otimes t^* \xrightarrow{\sim} t^*$$

Abstraction

$$\frac{G, x : t_1 \vdash e : t_2}{G \vdash \lambda x.e : t_1 \to t_2} \qquad \frac{G^* \otimes {t_1}^* \xrightarrow{e^*} {t_2}^*}{G^* \xrightarrow{\Lambda e^*} {t_1}^* \multimap {t_2}^*}$$

$$G^* \xrightarrow{\delta} !G^* \xrightarrow{!\Lambda e^*} !({t_1}^* \multimap {t_2}^*) = {t_1 \to t_2}^*$$

Application

$$\frac{G \vdash e_2 : t_1 \to t_2 \qquad G \vdash e_1 : t_1}{G \vdash (e_2)e_1 : t_1}$$

$$G^* \xrightarrow{e_2} !({t_1}^* \multimap {t_2}^*) \xrightarrow{\epsilon} {t_1}^* \multimap {t_2}^* \qquad G^* \xrightarrow{e_1} {t_1}^*$$

$$G^* \xrightarrow{c} G^* \otimes G^* \to ({t_1}^* \multimap {t_2}^*) \otimes {t_1}^* \xrightarrow{ev} {t_2}^*$$

Probabilistic Coherence Spaces

A model of LL for discrete probability

Vincent Danos, Thomas Ehrhard, On probabilistic coherence spaces, 2011

PCOH

The category of Probabilistic Coherent spaces - Pcoh

- **Object:** $(|A|, \mathrm{P}(A))$ with $|A|$ a set and $\mathrm{P}(A) \subset (\mathbb{R}^+)^{|A|}$ such that

$$\mathrm{P}(A) = \mathrm{P}(A)^{\perp\perp} \text{ where } \mathsf{P}^{\perp} = \{x \in (\mathbb{R}^+)^{|A|} \mid \forall x' \in \mathsf{P}\ \langle x, x' \rangle = \sum_{a \in |A|} x_a x' a \leq 1\}$$

Bounded covering $\forall a \in |A|\ (\exists x \in \mathrm{P}(A),\ x_a \neq 0)$ and $(\exists p \in \mathbb{R}^+,\ \forall x\ x_a < p)$

Examples $[\![\tau]\!] = (|\tau|, \mathrm{P}(\tau))$

$$|\texttt{unit dist}| = \{*\} \qquad \mathrm{P}(\texttt{unit dist}) = [0,1]$$

$$|\texttt{int dist}| = \mathbb{N} \qquad \mathrm{P}(\texttt{int dist}) = \{(x_n) | \sum x_n \leq 1\}$$

$$|\texttt{bool dist}| = \{\texttt{t},\ \texttt{f}\} \qquad \mathrm{P}(\texttt{bool dist}) = \{(x_{\texttt{t}}, x_{\texttt{f}}) \mid x_{\texttt{t}} + x_{\texttt{f}} \leq 1\}$$

$$|A \times B| = |A| \uplus |B| \qquad \mathrm{P}(A \times B) = \{(x_i)_{i \in |A| \uplus |B|} \mid (x_i)_{i \in |A|} \in \mathrm{P}(A), (x_i)_{i \in |B|} \in \mathrm{P}(B)\}$$

$$[\![\text{Bernoulli p}]\!] = (p, 1-p) \qquad [\![\text{Dice n}]\!] = (\tfrac{1}{n}, \ldots, \tfrac{1}{n}, 0, \ldots, 0, \ldots)$$

PCOH and Linear coherent maps is a model of simply typed lambda-calculus

The linear category of Probabilistic Coherence Spaces

- **Object:** $(|A|, \mathrm{P}(A))$ with $|A|$ a set and $\mathrm{P}(A) \subset (\mathbb{R}^+)^{|A|}$ set of functions
- **Morphism:** $f : (|A|, \mathrm{P}(A)) \to (|B|, \mathrm{P}(B))$ a matrix $(f_{(a,b)})$ indexed by $|A| \times |B|$ such that

$$\forall x \in \mathrm{P}(A), f(x) : b \mapsto \sum_{a \in |A|} x(a) f_{a,b} \in \mathrm{P}(B)$$

Examples

- $f : [\![\text{bool dist}]\!] \to [\![\text{bool dist}]\!]$ such that $f = \begin{array}{ccc} W \backslash S & \mathrm{t} & \mathrm{f} \\ \begin{array}{c} \mathrm{t} \\ \mathrm{f} \end{array} & \multicolumn{2}{c}{\begin{bmatrix} 1/5 & 4/5 \\ 3/4 & 1/4 \end{bmatrix}} \end{array}$

PCOH and Linear coherent maps is a model of Linear Logic

The linear category of Probabilistic Coherence Spaces

- **Object:** $(|A|, \mathrm{P}(A))$ with $|A|$ a set and $\mathrm{P}(A) \subset (\mathbb{R}^+)^{|A|}$ set of functions
- **Morphism:** $f : (|A|, \mathrm{P}(A)) \to (|B|, \mathrm{P}(B))$ such that $f \cdot \mathrm{P}(A) \subseteq \mathrm{P}(B)$

Tensor product

- $|X \otimes Y| = |X| \times |Y|$
- $\mathrm{P}(X \otimes Y) = \{x \otimes y \mid x \in \mathrm{P}(X),\ y \in \mathrm{P}(Y)\}^{\perp\perp}$ where $(x \otimes y) : (a, b) \mapsto x(a)y(b)$

Examples of morphisms in PCOH:

- Duplication: $\Delta : [\![\text{b dist}]\!] \to [\![\text{b dist} \otimes \text{b dist}]\!]$ such that $\Delta(x) = x \otimes x$
- Marginalization: $\text{proj} : [\![\text{b dist}]\!] \otimes [\![\text{b' dist}]\!] \to [\![\text{b dist}]\!]$ such that $\text{proj}(x \otimes y) = x$

Exponential

- $|!X| = \mathcal{M}_{\text{fin}}(|X|)$
- $\mathrm{P}(!X) = \{x^! \mid x \in \mathrm{P}(X)\}^{\perp\perp}$ where $x^! : m \mapsto \prod_{a \in m} x(a)^{m(a)}$

The cartesian closed category of Probabilistic Coherence spaces

The category of Probabilistic Coherence spaces and analytic maps

- **Object:** $(|A|, \mathrm{P}(A))$ with $|A|$ a set and $\mathrm{P}(A) \subset (\mathbb{R}^+)^{|A|}$ set of functions
- **Morphism:** $f : (|A|, \mathrm{P}(A)) \to (|B|, \mathrm{P}(B))$ a matrix $(f_{(m,b)})$ indexed by $\mathcal{M}_{\mathrm{fin}}(|A|) \times |B|$ such that

$$\forall x \in \mathrm{P}(A), f(x) : b \mapsto \sum_{m \in \mathcal{M}_{\mathrm{fin}}(|A|)} \prod_{a \in m} x(a)^{m(a)} f_{m,b} \in \mathrm{P}(B)$$

Examples

- $f : [\![unit]\!] \to [\![unit]\!]$ such that $\forall x \in [0,1],\ f \cdot x = \sum_n f_n x^n \in [0,1]$
- $f : [\![bool]\!] \to [\![unit]\!]$ such that $f_{(\mathtt{true}^n, *)} = 1$ otherwise $f_{m,*} = 0$, then $f \cdot (p, 1-p) = \sum_n p^n$ and $f \cdot (1,0) = 0$.

PCOH and analytic maps is a CCC (the Kleisli Category of !) CPO enriched.

It is a model of PCF = CBN lambda calculus and fixpoints

Results on Probabilistic Coherent Spaces

Compositionality

$$[\![(e)e_2]\!]_b = [\![e]\!]([\![e_2]\!])_b = \sum_m [\![e]\!]_{m,b} \prod_{a \in m} [\![e_2]\!]_a^{m(a)}$$

Invariance of the semantics

$$[\![e]\!] = \sum_{e_2} \mathbf{Proba}(e, e_2)[\![e_2]\!]$$

Adequacy Lemma

$$\text{if } \vdash e : \mathtt{nat}, \text{ then } \mathbf{Proba}^{\infty}(e, \underline{n}) = [\![e]\!]_n$$

Full Abstraction at ground type nat

$$[\![e]\!]_1 = [\![e]\!]_2 \text{ if and only if } \mathbf{Proba}^{\infty}(C[e_1], n) \overset{\forall C[\,]\forall n}{=} \mathbf{Proba}^{\infty}(C[e_2], n)$$

Examples: Compute the semantics


```
fun (d:bool dist) →
  fix fun (f:unit → unit) →
        let x = sample(d) in
        if x then f(d) else ()
```

```
let rec f (d: bool dist)
      : unit =
        let x = sample(d) in
        if x then f(d) else ()
```

- **By definition:** $[\![G]\!](d) \in [0,1]$
- **By adequation:** if $G = \text{fix (fun } f \to e)$ then $[\![G]\!] = [\![e[G/f]]\!]$ (from adequation)
- **Let semantics:** $[\![\text{let } x = \text{ sample}(d) \text{ in } e']\!] = \sum_a [\![d]\!]_a [\![e'[a/x]]\!]$
- **Computation:** $[\![G]\!](d) = d_{\texttt{true}}[\![G]\!](d) + d_{\texttt{false}}$ thus $[\![G]\!](d) = \dfrac{d_{\texttt{false}}}{1 - d_{\texttt{true}}}$

Examples: Compute the semantics


```
fix fun (b : bool) →
  let x = sample(Bernoulli p) in
  if x  then (if x then b else true)
        else (if x then false else b)

```

$$b = p(b + \delta_{\texttt{true}}) + (1-p)(\delta_{\texttt{false}} + b)$$

$$b_{\texttt{true}} = b_{\texttt{false}} = 0$$

```
fix fun (bb : bool) →
  let x = (Bernoulli p) in
  if x  then (if x then bb else true)
        else (if x then false else bb)

```

$$bb = p(p\ b + (1-p)\delta_{\texttt{true}}) + (1-p)(p\delta_{\texttt{false}} + (1-p)\ b)$$

$$bb_{\texttt{true}} = bb_{\texttt{false}} = \frac{p}{1+p}$$

Why you should care or not on Probabilistic Coherence Spaces

- ✓ Interpretation of probabilistic programs of discrete type (int or real) are PCOH maps, with good analytic Properties

  Probabilistic Stable Functions on Discrete Cones are Power Series. R. Crubille, 2018

- ✓ Concrete Sandbox for getting intuitions on probabilistic programs

- Non definability: $\mathrm{Pcoh}(\mathrm{bool}, 1) = \left\{ Q \in (\mathbb{R}^+)^{\mathcal{M}_{\mathrm{fin}}(\mathtt{t},\mathtt{f})} \mid Q_{[\mathtt{t}^n,\mathtt{f}^m]} \le \frac{(n+m)^{n+m}}{n^n m^m} \right\}$ but the greatest we can get is $\{\!|e|\!\} \le \frac{(n+m)!}{n!m!}$

```
fix fun x → if x then if x then f(x) else () else if x then () else f(x)
```

Why you should care or not on Probabilistic Coherence Spaces

- It is not computable and thus cannot be used to implement inference
if $\vdash^P M : \tau$ and $\{\!|M|\!\} \in \mathrm{P}\,(\tau)$ then $\vdash^D$ infer $M : \tau\mathit{dist}$ and $[\![\text{infer } M]\!]$ is a subprobability distribution over τ.

$$[\![\text{infer } M]\!] = \frac{\{\!|M|\!\}}{\sum_{a \in |\tau|} \{\!|M|\!\}_a}$$

Scaling exact inference for discrete probabilistic programs. Holtzen & al. 2020

Meas is not an SMCC

Borel structures for function spaces, Aumann, 1961

Denotational Semantics for continuous probability - Higher-Order problem

Meas is a symmetric monoidal category

- Meas:
  - Objects: Measurable spaces $|X|, \mathfrak{S}(X)$
  - Morphisms: Measurable functions
- Symmetric Monoidal:
  - Unit: $|1| = \{*\}$
  - Monoidal product:
    - $|X \otimes Y| = |X| \times |Y|$
    - $\mathfrak{S}(X \otimes Y) = \mathfrak{S}(X) \otimes \mathfrak{S}(Y)$ is generated by squared $U \times V$ where $U \in \mathfrak{S}(X)$ and $V \in \mathfrak{S}(Y)$ using countable intersections, unions and complement.

Theorem (Aumann' 61)

There is no σ-algebra on $\mathbb{R}^{\mathbb{R}}$ such that **eval** $: \mathbb{R}^{\mathbb{R}} \otimes \mathbb{R} \to \mathbb{R}$ is measurable.

The category of measurable space and measurable functions is symmetric monoidal but not closed.

By contradiction: Assume Meas is an SMCC: $\forall X, Y, \mathrm{ev} : Y^X \otimes X \to Y$ is measurable.

Measurable spaces X is $\mathbb{R}$ with $\Sigma_X = \mathcal{P}(X)$ any parts and Y is $\mathbb{R}$ with the Σ-algebra generated by countable and cocountable subsets (closed by complement and countable unions and intersections).

Diagonal function: $h : \begin{cases} (\mathbb{R} \times \mathbb{R}, \mathcal{P}(\mathbb{R}) \otimes \mathcal{C}(\mathbb{R})) & \to \{0, 1\} \\ (x, y) & \mapsto 1 \text{ if } x = y \\ & \mapsto 0 \text{ otherwise} \end{cases},$

$\Lambda(h) : (\mathbb{R}, \mathcal{P}(\mathbb{R})) \to (\{0, 1\}^{\mathbb{R}}, \Sigma_{2^Y})$ is **measurable**

$h = \mathrm{ev} \circ \Lambda(h)$ is **measurable** as the composition of measurable functions

$\Delta = \{(x, y) \in \mathbb{R}^2 \mid x = y\} = h^{-1}(1)$ is measurable in $\mathcal{P}(\mathbb{R}) \otimes \mathcal{C}(\mathbb{R})$.

The category of measurable spaces and measurable functions is not closed.

By contradiction: Assume Meas is an SMCC: $\forall X, Y, \mathrm{ev} : Y^X \otimes X \to Y$ is measurable.

We deduced that $\Delta = \{(x,y) \in \mathbb{R}^2 \mid x = y\} = h^{-1}(\{1\})$ is **measurable** in $\mathcal{P}(\mathbb{R}) \otimes \mathcal{C}(\mathbb{R})$.

If $W \in \mathcal{P}(\mathbb{R}) \otimes \mathcal{C}(\mathbb{R})$, then there is $B \subseteq \mathbb{R}$ countable such that:

if there is $(x,y) \in W$ such that $y \notin B$, then $\forall z \notin B$, $(x,z) \in W$.

Proof: satisfied by basic measurable subsets and closed by countable intersection and unions

Consequence: Δ satisfies this property: there is B countable such there is $(x,x) \in \Delta$ and $x \notin B$. Yet, for any $z \notin B$ and $z \neq x$, $(x,z) \notin \delta$.

CONTRADICTION

(with many thanks to Ohad Kammar and Thomas Ehrhard)