Probabilistic Programming Languages

M2 MPRI 2021-2022

Guillaume Baudart (guillaume.baudart@ens.fr)
Christine Tasson (christine.tasson@lip6.fr)

Semantics of Probabilistic Programming

Continuous probability

Probabilistic PCF - Discrete Probability

Syntax

$$M, N, P := \underbrace{x \mid \lambda x\, M \mid (M)\, N \mid (M, N)}_{\lambda\text{-calculus}} \mid \underbrace{\texttt{fix}\, M}_{\text{Recursion}}$$

$$\mid \underbrace{0 \mid \texttt{succ}\, M}_{\text{Arithmetics}} \mid \underbrace{\texttt{true} \mid \texttt{false} \mid \texttt{if}\, M\, \texttt{then}\, N\, \texttt{else}\, P}_{\text{Conditionnal}}$$

$$\mid \underbrace{\texttt{let x} = \texttt{sample}(\texttt{bernoulli}\ p)\ \texttt{in}\ M}_{\text{Discrete Probability}}\ \forall p \in [0,1]$$

Types

$$\frac{\Gamma, x : \texttt{bool} \vdash M : B}{\Gamma \vdash \texttt{let x} = \texttt{sample}(\texttt{bernoulli}\ p)\ \texttt{in}\ M : B}$$

Semantics

Operational : $M \xrightarrow{p} M'$ **Denotational :** $[\![\Gamma \vdash M : A]\!] : [\![\Gamma]\!] \to [\![A]\!]$

Probabilistic Coherent Spaces

Soundness : $[\![\Gamma \vdash M : A]\!] = \sum_{M'} \textbf{Proba}\,(M, M')\,[\![\Gamma \vdash M' : A]\!]$.

Probabilistic PCF - Continuous Probability

Syntax

$$M, N, P := \underbrace{x \mid \lambda x\, M \mid (M)\, N \mid (M, N)}_{\lambda\text{-calculus}} \mid \underbrace{\texttt{fix}\, M}_{\text{Recursion}}$$

$$\mid \underbrace{\underline{n}, \mid \texttt{succ}\, M}_{\text{Arithmetics}} \mid \underbrace{\texttt{true} \mid \texttt{false} \mid \texttt{if}\, M\, \texttt{then}\, N\, \texttt{else}\, P}_{\text{Conditionnal}}$$

$$\mid \underbrace{\texttt{Dirac M} \mid \texttt{Uniform a : 0. b : 1.} \mid \texttt{let x = sample M in N}}_{\text{Continuous Probability}}$$

$$\mid \underline{r} \mid \underline{f}(M_1, \ldots, M_n)\ \forall r \text{ float } \forall f : \mathbb{R}^n \to \mathbb{R} \text{ measurable}$$

Types

$$\frac{}{\underline{r} : \texttt{float} \vdash \texttt{Dirac r} : \texttt{float dist}} \qquad \frac{}{\vdash \texttt{Uniform a : 0. b : 1.} : \texttt{float dist}}$$

$$\frac{\Gamma, x : A \vdash M : B \qquad \Gamma \vdash N : A\ \texttt{dist}}{\Gamma \vdash \texttt{let x = sample } N \texttt{ in } M : B} \qquad \frac{\Gamma \vdash M_i : \texttt{float dist}}{\Gamma, \underline{f} : \texttt{float}^n \to \texttt{float} \vdash f(M_1, \ldots, M_n) : \texttt{float}^\texttt{n}\ \texttt{dist}}$$

Operational Semantics : The evaluation of a program is a markov process described by the probability of reduction from M to N.

Operational Semantics - Discrete Probability

If $\vdash M : \mathtt{nat}$, then $\mathbf{Proba}^{\infty}(M, \_)$ is the discrete distribution over $\mathbb{N}$ computed by M.

$$\mathbf{Proba}(\mathtt{sample\ Bernoulli\ p}, \underline{0}) = \tfrac{1}{2}$$

Transition Matrix : $\mathbf{Proba}(M, M')$ a stochastic matrix indexed by terms.

$$\mathbf{Proba}(M, M') = \begin{cases} p & \text{if } M \xrightarrow{p} M' \\ 1 & \text{if } M \text{ normal and } M = M' \\ 0 & \text{otherwise.} \end{cases}$$

Iterated Transition Matrix :

> $\mathbf{Proba}^{k}(M, N)$ is the probability that M reduces to N in at most k steps.
> $\mathbf{Proba}^{\infty}(M, N)$ when N is normal is the probability that M reduces to N in any number of steps

Adequation Lemma : If M closed term of type $\mathtt{nat}$,
then for every n, $[\![M]\!]_n = \mathbf{Proba}^{\infty}(M, \underline{n})$

Operational Semantics - Continuous Probability

If $\vdash M : \texttt{float}$, then $\textbf{Proba}^{\infty}(M, \_)$ is the continuous distribution over $\mathbb{R}$ computed by M. $\textbf{Proba}(\texttt{sample Uniform a} : 0.\ \texttt{b} : 1., U) = \int_{x \in U} dx$

The probability to observe U after at most one reduction step applied to M is **Proba**(M , U)

Proba : $\Lambda^{\Gamma \vdash A} \times \Sigma_{\Lambda^{\Gamma \vdash A}} \to \mathbb{R}^+$ is a stochastic **Kernel**, i.e :

for all $M \in \Lambda^{\Gamma \vdash A}$, $\textbf{Proba}(M, \_)$ is a measure ;

for all $U \in \Sigma_{\Lambda^{\Gamma \vdash A}}$, $\textbf{Proba}(\_, U)$ is a measurable function.

$\textbf{Proba}^{\infty}(M, U)$ is the probability to observe U after any steps.

Operational Semantics - Continuous Probability

If $\vdash M : \texttt{float}$, then $\mathbf{Proba}^{\infty}(M, \_)$ is the continuous distribution over $\mathbb{R}$ computed by M. $\mathbf{Proba}(\texttt{sample Uniform a} : \texttt{0. b} : \texttt{1.}, U) = \int_{x \in U} dx$

The probability to observe U after at most one reduction step applied to M is **Proba**(M , U)

$\Lambda^{\Gamma \vdash A}$: the set of terms M s.t. $\Gamma \vdash M : A$.

$\mathbf{Proba} : \Lambda^{\Gamma \vdash A} \times \Sigma_{\Lambda^{\Gamma \vdash A}} \to \mathbb{R}^+$ is a stochastic **Kernel**, i.e :

for all $M \in \Lambda^{\Gamma \vdash A}$, $\mathbf{Proba}(M, \_)$ is a measure ;

for all $U \in \Sigma_{\Lambda^{\Gamma \vdash A}}$, $\mathbf{Proba}(\_, U)$ is a measurable function.

$\mathbf{Proba}^{\infty}(M, U)$ is the probability to observe U after any steps.

Operational Semantics - Continuous Probability

If $\vdash M : \texttt{float}$, then $\mathbf{Proba}^{\infty}(M, \_)$ is the continuous distribution over $\mathbb{R}$ computed by M. $\mathbf{Proba}(\texttt{sample Uniform a} : 0.\ \texttt{b} : 1., U) = \int_{x \in U} dx$

The probability to observe U after at most one reduction step applied to M is **Proba**(M , U)

$\Lambda^{\Gamma \vdash A}$: the set of terms M s.t. $\Gamma \vdash M : A$.

$\Sigma_{\Lambda^{\Gamma \vdash A}}$, i.e. U is measurable : $\forall n, \forall S,\ \{\vec{r}\ s.t.\ S\underline{\vec{r}} \in U\}$ meas. in $\mathbb{R}^n$

Proba : $\Lambda^{\Gamma \vdash A} \times \Sigma_{\Lambda^{\Gamma \vdash A}} \to \mathbb{R}^+$ is a stochastic **Kernel**, i.e :

for all $M \in \Lambda^{\Gamma \vdash A}$, $\mathbf{Proba}(M, \_)$ is a measure ;

for all $U \in \Sigma_{\Lambda^{\Gamma \vdash A}}$, $\mathbf{Proba}(\_, U)$ is a measurable function.

$\mathbf{Proba}^{\infty}(M, U)$ is the probability to observe U after any steps.

Operational Semantics - Continuous Probability

If $\vdash M : \texttt{float}$, then $\mathbf{Proba}^{\infty}(M, \_)$ is the continuous distribution over $\mathbb{R}$ computed by M. $\mathbf{Proba}(\texttt{sample Uniform a} : 0.\ \texttt{b} : 1., U) = \int_{x \in U} dx$

> The probability to observe U after at most one reduction step applied to M is **Proba**(M , U)

$\Lambda^{\Gamma \vdash A}$: the set of terms M s.t. $\Gamma \vdash M : A$.

$\Sigma_{\Lambda^{\Gamma \vdash A}}$, i.e. U is measurable : $\forall n, \forall S,\ \{\vec{r}\ s.t.\ S\underline{\vec{r}} \in U\}$ meas. in $\mathbb{R}^n$

> **Proba** : $\Lambda^{\Gamma \vdash A} \times \Sigma_{\Lambda^{\Gamma \vdash A}} \to \mathbb{R}^+$ is a stochastic **Kernel**, i.e :
>
> for all $M \in \Lambda^{\Gamma \vdash A}$, **Proba**$(M, \_)$ is a measure ;
>
> for all $U \in \Sigma_{\Lambda^{\Gamma \vdash A}}$, **Proba**$(\_, U)$ is a measurable function.

Measurable sets and kernels constitute the category Kern.

> **Proba**$^{\infty}(M, U)$ is the probability to observe U after any steps.

Examples : Distributions

```
sample = sample Uniform a:0. b:1.
```

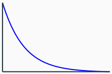
The Bernoulli distribution takes the value 1 with probability p and the value 0 with probability $1 - p$.

$$p\delta_1 + (1-p)\delta_0$$

$\texttt{bernoulli}\ p ::= \texttt{let}\, x{=}\texttt{sample}\,\texttt{in}\, x \leq p$

tests if `sample` draws a value within $[0, p]$.

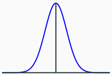
The exponential distribution is specified by its density e^{-x}.



$\mathrm{exp} ::= \texttt{let}\, x{=}\texttt{sample}\,\texttt{in}\, -\log(x)$

by the inversion sampling method.

The standard normal distribution defined by its density $\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{1}{2}x^2}$.



$\texttt{gauss} ::= \texttt{let}\, x{=}\texttt{sample}\,\texttt{in}$

$\texttt{let}\, y{=}\texttt{sample}\,\texttt{in}\, \sqrt{-2\log(x)}\, \cos(2\pi y)$

by the Box Muller method.

Denotational Semantics - First order

Discrete : **PCOH**

For $\vdash M : \texttt{nat}$,
$[\![M]\!]$ a distribution over $\mathbb{N}$

For $\vdash \underline{n} : \texttt{nat}$,
$[\![\underline{n}]\!]_p = \delta_{p,n}$

For $\vdash \texttt{Bernoulli p} : \texttt{bool}$,

$[\![\texttt{Bern p}]\!]_{\texttt{b}} = \frac{1}{2}\delta_{\texttt{true,b}} + \frac{1}{2}\delta_{\texttt{false,b}}$

For $\vdash N : \texttt{nat}, \vdash P : A, \vdash Q : A$,

$$[\![\textbf{if } N \textbf{ then } P \textbf{ else } Q]\!]_a = [\![N]\!]_{\texttt{true}}[\![P]\!]_a + [\![N]\!]_{\texttt{false}}[\![Q]\!]_a$$

$$[\![\texttt{let}\, x{=}N \,\texttt{in}\, P]\!]_a = \sum_{n=0}^{\infty} [\![N]\!]_n \widehat{[\![P]\!]}(n)_a$$

Denotational Semantics - First order

Discrete : **PCOH**

For $\vdash M : \mathtt{nat}$,
$[\![M]\!]$ a distribution over $\mathbb{N}$

For $\vdash \underline{n} : \mathtt{nat}$,
$[\![\underline{n}]\!]_p = \delta_{p,n}$

For $\vdash \mathtt{Bernoulli\ p} : \mathtt{bool}$,
$$[\![\mathtt{Bern\ p}]\!]_{\mathtt{b}} = \frac{1}{2}\delta_{\mathtt{true},\mathtt{b}} + \frac{1}{2}\delta_{\mathtt{false},\mathtt{b}}$$

For $\vdash N : \mathtt{nat}, \vdash P : A, \vdash Q : A$,
$$[\![\textbf{if } N \textbf{ then } P \textbf{ else } Q]\!]_a = [\![N]\!]_{\mathtt{true}}[\![P]\!]_a + [\![N]\!]_{\mathtt{false}}[\![Q]\!]_a$$

$$[\![\mathtt{let}\, x{=}N\, \mathtt{in}\, P]\!]_a = \sum_{n=0}^{\infty} [\![N]\!]_n \widehat{[\![P]\!]}(n)_a$$

Continuous : **KERN**

For $\vdash M : \mathtt{float}$,
$[\![M]\!]$ a measure over $\mathbb{R}$

For $\vdash \underline{r} : \mathtt{float}$,
$[\![\underline{r}]\!](U) = \delta_r(U)$

For $\vdash \mathtt{Uniform\ a : 0.\ b : 1. : float}$,
$$[\![\mathtt{Unif\ 0.\ 1.}]\!](\mathtt{U}) = \int_{\mathtt{x}\in\mathtt{U}} \mathtt{dx}$$

For $\vdash R : \mathtt{float}, \vdash P, Q : A$,
$$[\![\textbf{if } R \textbf{ then } P \textbf{ else } Q]\!](U) = [\![R]\!](\{\mathtt{true}\})[\![\mathtt{P}]\!](\mathtt{U}) + [\![\mathtt{R}]\!](\{\mathtt{false}\})[\![\mathtt{Q}]\!](\mathtt{U})$$

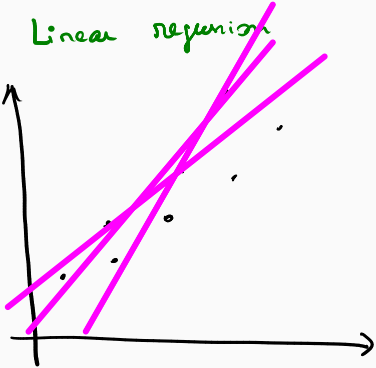
$$[\![\mathtt{let}\, x{=}R\, \mathtt{in}\, P]\!](U) = \int [\![R]\!](dr)[\![P]\!](\delta_r)(U)$$

Denotational Semantics - Higer-Order

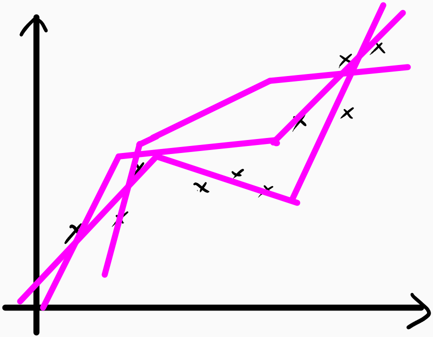
The Program Synthesis problem

Given a data set in R^2 can we build the probabilistic program $\vdash M :$ `float -> float` that generated this data set ? We need a distribution over `float -> float`.

Linear regression



Piecewise linear Regression



Denotational Semantics - Higher-Order problem

Théorème (Aumann' 61)

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

By contradiction, assume for all Z, Y measurable space, Z^Y measurable and eval measurable:

$$\begin{tikzcd} X \times Y \arrow[d, "\Lambda h \times id"'] \arrow[dr, bend left, "h"] & \\ \mathbb{R}^Y \times Y \arrow[r, "eval"'] & \mathbb{R} \end{tikzcd}$$

$h : x, y \mapsto \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$

Assume $X = \mathbb{R}, \mathcal{P}(\mathbb{R})$ and $Y = \mathbb{R}, \mathcal{C}(\mathbb{R})$

all subsets of $\mathbb{R}$

countable or cocountable subsets of $\mathbb{R}$

Then $X \times Y = \mathbb{R} \times \mathbb{R}, \mathcal{P}(\mathbb{R}) \otimes \mathcal{C}(\mathbb{R})$

σ-algebra generated by $U \times V$ by
- complement
- countable unions
- countable intersections

Denotational Semantics - Higher-Order problem

Théorème (Aumann' 61)

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

$$\begin{tikzcd} X \times Y \arrow[d, "\Lambda h \times id"'] \arrow[dr, bend left] & \\ \mathbb{R}^Y \times Y \arrow[r, "ev"'] & \mathbb{R} \end{tikzcd}$$

$h : x, y \mapsto \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$

- $\Lambda h : X \to \mathbb{R}^Y$ is measurable: $\forall A \in \sigma(\mathbb{R}^Y)$, $\Lambda h^{-1}(A) \in \mathcal{P}(\mathbb{R})$!

  is well defined: $\Lambda(h)(x) : Y \to \mathbb{R}$ measurable

  $y \mapsto \begin{cases} 1 & x = y \\ 0 & \text{otherwise} \end{cases}$

  $\Lambda(h)(x)^{-1}(\{1\}) = \{x\}$ countable $\in \mathcal{C}(\mathbb{R})$

  $\Lambda(h)(x)^{-1}(\{0\}) = \mathbb{R} \setminus \{x\}$ co countable $\in \mathcal{C}(\mathbb{R})$

- We assumed that ev is measurable, hence $h = ev \circ (\Lambda h \times id)$ is measurable

Denotational Semantics - Higher-Order problem

Théorème (Aumann' 61)

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

If h is measurable then so is $h^{-1}(\{1\}) = \{(x,x) \mid x \in \mathbb{R}\} \notin \mathcal{P}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ NO

$$h : x, y \mapsto \begin{cases} 1 & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$$

$X \times Y = \mathbb{R} \times \mathbb{R}$, $\mathcal{P}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$

σ-algebra generated by $U \times V$ by
- complement
- countable unions
- countable intersections

(P) If $W \in \mathcal{P}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R})$ then $\exists (b_n) \in \mathbb{R}^{\mathbb{N}}$ st if $\underline{(x,y) \in W}$ and $y \notin b_n$ then $\forall z \notin b_n$, $\underline{(x,z) \in W}$

$\Delta = \{(x,x) \mid x \in \mathbb{R}\}$ does not satisfy P: $\forall (b_n) \in \mathbb{R}^{\mathbb{N}}$, $\forall y \notin b_n$, $(y,y) \in \Delta$ $\exists z \neq y$ $(y,z) \in \Delta$!

If $U \in \mathcal{P}(\mathbb{R})$, $C \in \mathcal{B}(\mathbb{R})$, then $U \times C$ satisfies P | If W_n satisfy P, then $\cup W_n$ and $\cap W_n$ satisfy P.

Semantics of Probabilistic Programming

Measurable Stable Cones

Measurable Stable Cones - Definition

CSTAB is a **CCC** based on Selinger's **cones** (dcpos with the order induced by addition and a convex structure).

Objects are cones and measurable spaces

Morphisms are stable and measurable functions

1. **Complete cones** (convex dcpos with the order induced by addition) with Scott continuous functions
   However, the category is cartesian but not closed.
2. Complete cones and **Stable functions** (∞-non-decreasing functions) is a CCC.
   However, not every stable function is measurable.
3. **Measurable Cones** (complete cones with **measurable tests**). Measurable paths pass measurable tests and Measurable functions preserve measurable paths.
   CSTAB is a CCC with measurability included !

Results

The category **CSTAB** is a CCC and a model of Real PPCF.

Invariance of the semantics

$$[\![M]\!]_{\Gamma \vdash A} = \int_{\Lambda^{\Gamma \vdash A}} [\![t]\!]_{\Gamma \vdash A} \, \mathbf{Proba}(M, dt)$$

Adequacy

$$[\![M]\!]_{\vdash \mathtt{real}}(U) = \mathbf{Proba}^{\infty}(M, U)$$

Conservative extension of PCOH

If X and Y are probabilistic coherence spaces and $f : X \to Y$ is a stable measurable map, then f is a map of probabilistic coherence spaces. (Crubille 2018).

Step 1 : Complete Cones

A Cone P is analogous to a real normed vector space, except that scalars are $\mathbb{R}^+$ and the norm $\|\_\|_P : P \to \mathbb{R}^+$ satisfies :

$$x+y=0 \Rightarrow x,y=0, \qquad \|x+x'\|_P \leq \|x\|_P + \|x'\|_P, \qquad \|\alpha x\|_P = \alpha\|x\|_P$$
$$x+y=x+y' \Rightarrow y=y', \qquad \|x\|_P = 0 \Rightarrow x=0, \qquad \|x\|_P \leq \|x+x'\|_P$$

The Unit Ball is the set $\mathcal{B}P = \{x \in P \mid \|x\|_P \leq 1\}$.

Order $x \leq_P x'$ if there is a $y \in P$ such that $x' = x+y$. This unique y is denoted as $y = x' - x$.

A Complete Cone is s.t. any non-decreasing $(x_n)_{n\in\mathbb{N}}$ of $\mathcal{B}P$ has a lub and $\|\sup_{n\in\mathbb{N}} x_n\|_P = \sup_{n\in\mathbb{N}} \|x_n\|_P$.

Example of Complete Cones

- $\mathrm{Meas}(X)$ with X a measurable space.
- $\widehat{\mathcal{X}} = \{u \in (\mathbb{R}^+)^{|\mathcal{X}|} \mid \exists\epsilon > 0\ \epsilon u \in \mathbf{PCOH}\mathcal{X}\}$ if $\mathcal{X} \in \mathbf{PCOH}$.

Step 2 : Stable functions

The category of **complete cones** and **Scott-continuous** functions is not cartesian closed as *currying* fails to be *non-decreasing*.

A function $f : \mathcal{B}P \to Q$ is **n-non-decreasing function** if :

$n = 0$ and f is non-decreasing

$n > 0$ and $\forall u \in \mathcal{B}P$, $\Delta f(x; u) = f(x+u) - f(x)$ is $(n-1)$-non-decreasing in x.

A function is **stable** if it is Scott-continuous and ∞-non-decreasing, i.e. n-non-decreasing for all $n \in \mathbb{N}$.

Complete cones and **stable** functions constitute a **CCC**.

Weak Parallel Or

wpor : $[0,1] \times [0,1] \to [0,1]$ given as $\mathrm{wpor}(s,t) = s + t - st$ is Scott-continuous, but not Stable. Its currying is not Scott-continuous.

Step 3 : The Measurability Problem

Type $\texttt{real}$ is interpreted as $[\![\texttt{real}]\!] = \text{Meas}(\mathbb{R})$,
Closed term $\vdash M : \texttt{real}$ as a measure μ and
Term $x : \texttt{real} \vdash N : \texttt{real}$ as a stable $f : \text{Meas}(\mathbb{R}) \to \text{Meas}(\mathbb{R})$.

Operational semantics

$$\forall r, \text{ s.t. } M \to r, \ \texttt{let}\, x{=}M\, \texttt{in}\, N \to N\{r/x\}$$

By **Soundness**

$$[\![\texttt{let}\, x{=}M\, \texttt{in}\, N]\!] = \int_{\mathbb{R}} (\ f \ \circ \ \delta \ )(r) \ \ \mu\,(dr)$$

Step 3 : The Measurability Problem

Type `real` is interpreted as $[\![\texttt{real}]\!] = \mathrm{Meas}(\mathbb{R})$,
Closed term $\vdash M : \texttt{real}$ as a measure μ and
Term $x : \texttt{real} \vdash N : \texttt{real}$ as a stable $f : \mathrm{Meas}(\mathbb{R}) \to \mathrm{Meas}(\mathbb{R})$.

Operational semantics

$$\forall r,\ \text{s.t. } M \to r,\ \texttt{let}\, x{=}M\, \texttt{in}\, N \to N\{r/x\}$$

By **Soundness**

$$[\![\texttt{let}\, x{=}M\, \texttt{in}\, N]\!] = \int_{\mathbb{R}} (\ f\ \circ\ \delta\ )(r)\ \ \mu\,(dr)$$

$[\![N]\!]$

Step 3 : The Measurability Problem

Type $\texttt{real}$ is interpreted as $\llbracket \texttt{real} \rrbracket = \mathrm{Meas}(\mathbb{R})$,
Closed term $\vdash M : \texttt{real}$ as a measure μ and
Term $x : \texttt{real} \vdash N : \texttt{real}$ as a stable $f : \mathrm{Meas}(\mathbb{R}) \to \mathrm{Meas}(\mathbb{R})$.

Operational semantics

$$\forall r,\ \text{s.t. } M \to r,\ \texttt{let}\, x{=}M\, \texttt{in}\, N \to N\{r/x\}$$

By **Soundness**

$$\llbracket \texttt{let}\, x{=}M\, \texttt{in}\, N \rrbracket = \int_{\mathbb{R}} (\ f \ \circ \ \delta\ )(r) \ \ \mu\,(dr)$$

$\llbracket N \rrbracket$ Dirac measure

Step 3 : The Measurability Problem

Type `real` is interpreted as $[\![\texttt{real}]\!] = \mathsf{Meas}(\mathbb{R})$,
Closed term $\vdash M : \texttt{real}$ as a measure μ and
Term $x : \texttt{real} \vdash N : \texttt{real}$ as a stable $f : \mathsf{Meas}(\mathbb{R}) \to \mathsf{Meas}(\mathbb{R})$.

Operational semantics

$$\forall r,\ \text{s.t. } M \to r,\ \texttt{let}\, x{=}M\,\texttt{in}\, N \to N\{r/x\}$$

By **Soundness**

$$[\![\texttt{let}\, x{=}M\,\texttt{in}\, N]\!] = \int_{\mathbb{R}} (\underset{\substack{\uparrow\\ [\![N]\!]}}{f} \circ \underset{\substack{\uparrow\\ \text{Dirac measure}}}{\delta})(r)\ \underset{\substack{\uparrow\\ [\![M]\!]}}{\mu}\,(dr)$$

Step 3 : The Measurability Problem

Type $\texttt{real}$ is interpreted as $[\![\texttt{real}]\!] = \mathrm{Meas}(\mathbb{R})$,
Closed term $\vdash M : \texttt{real}$ as a measure μ and
Term $x : \texttt{real} \vdash N : \texttt{real}$ as a stable $f : \mathrm{Meas}(\mathbb{R}) \to \mathrm{Meas}(\mathbb{R})$.

Operational semantics

$$\forall r,\ \text{s.t. } M \to r,\ \texttt{let}\, x{=}M\, \texttt{in}\, N \to N\{r/x\}$$

By **Soundness**

$$[\![\texttt{let}\, x{=}M\, \texttt{in}\, N]\!] = \int_{\mathbb{R}} (\ f\ \circ\ \delta\ )(r)\ \ \mu\ (dr)$$

Thus $f \circ \delta$ needs to be measurable.

There are non measurable stable functions
We need to equip every cone with a notion of measurability

Step 3 : Measurability tests

Measurability tests of $\mathrm{Meas}(\mathbb{R})$ are given by measurable sets of $\mathbb{R}$:

$$\forall U \subseteq \mathbb{R} \text{ measurable}, \ \epsilon_U \in \mathrm{Meas}(\mathbb{R})' : \mu \mapsto \mu(U)$$

For needs of CCC, we parameterized measurable tests of a cone :

Measurable Cone

A cone P with a collection $(\mathrm{M}^n(P))_{n\in\mathbb{N}}$ with $\mathrm{M}^n(P) \subseteq (P')^{\mathbb{R}^n}$ s.t. :

$$0 \in \mathrm{M}^n(P), \quad \ell \in \mathrm{M}^n(P) \text{ and } h : \mathbb{R}^p \to \mathbb{R}^n \Rightarrow \ell \circ h \in \mathrm{M}^p(P)$$

$$\ell \in \mathrm{M}^n(P) \text{ and } x \in P \Rightarrow \left\{ \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^+ \\ \mathrm{Vect}\, r & \mapsto & \ell(\mathrm{Vect}\, r)(x) \end{array} \right. \quad \text{measurable.}$$

Measurable Tests, Paths and Functions

CSTAB is the category of complete and measurable cones with stable and measurable functions.

Let P and Q be measurable and complete cones :

Measurable Test : $\mathrm{M}^n(P) \subseteq (P')^{\mathbb{R}^n}$
Measurable Path : $\mathrm{Path}^n(P) \subseteq P^{\mathbb{R}^n}$ the set of bounded $\gamma : \mathbb{R}^n \to P$
such that $\ell * \gamma : \mathbb{R}^{k+n} \to \mathbb{R}^+$ is measurable with

$$\ell * \gamma : (\mathrm{Vect}\, r, \mathrm{Vect}\, s) \mapsto \ell(\mathrm{Vect}\, r)(\gamma(\mathrm{Vect}\, s))$$

Measurable Functions : Stable functions $f : P \to Q$ such that :

$$\forall n \in \mathbb{N},\ \forall \gamma \in \mathrm{Path}^n_1(P), \quad f \circ \gamma \in \mathrm{Path}^n(Q)$$

If X is a measurable space, then $\mathrm{Meas}(X)$ is equipped with :
$\mathrm{M}^n(X) = \{\epsilon_U : \mathbb{R}^n \to \mathrm{Meas}(X)' \text{ s.t. } \epsilon_U(\mathrm{Vect}\, r)(\mu) = \mu(U),\ U \text{ meas.}\}$
$\mathrm{Path}^n_1(P)$ is the set of stochastic kernels from $\mathbb{R}^n$ to X.

Semantics of Probabilistic Programming

Quasi Borel Spaces

(Ohad Kammar Tutorial)

Measurable Stable Cones - Definition

Quasi Borel Space $X = (|X|, \mathcal{R}(X))$ such that

Random elements : $\mathcal{R}(X) \subset \mathbb{R} \to |X|$
Constants : if $x \in |X|$, then $\lambda r.x \in \mathcal{R}(X)$
Precomposition : if $\alpha \in \mathcal{R}(X)$ and $\varphi : \mathbb{R} \to \mathbb{R}$ measurable, then $\varphi \circ \alpha \in \mathcal{R}(X)$.
Recombination : if $\alpha \in \mathcal{R}(X)^{\mathbb{N}}$ and $\mathbb{R} = \uplus A_n$ and A_n measurable, then $\lambda r.\alpha_n(r)$ (if $r \in A_n$) $\in \mathbb{R}(X)$

Examples

Reals : $(\mathbb{R}, \mathcal{M}eas(\mathbb{R}, \mathbb{R}))$
Discrete QBS : $(|X|, \sigma - \text{simple}(\mathbb{R}, |X|))$
Indiscrete QBS : $(|X|, |X|^{\mathbb{R}})$

Measurable Stable Cones - Definition

Quasi Borel Space $X = (|X|, \mathcal{R}(X))$ such that

Random elements : $\mathcal{R}(X) \subset \mathbb{R} \to |X|$
Constants : if $x \in |X|$, then $\lambda r.x \in \mathcal{R}(X)$
Precomposition : if $\alpha \in \mathcal{R}(X)$ and $\varphi : \mathbb{R} \to \mathbb{R}$ measurable, then $\varphi \circ \alpha \in \mathcal{R}(X)$.
Recombination : if $\alpha \in \mathcal{R}(X)^{\mathbb{N}}$ and $\mathbb{R} = \uplus A_n$ and A_n measurable, then $\lambda r.\alpha_n(r)$ (if $r \in A_n$) $\in \mathbb{R}(X)$

Morphism $f : X \to Y$

Function $f : |X| \to |Y|$ such that
If $\alpha \in \mathcal{R}(X)$, then $f \circ \alpha \in R(Y)$

Measurable Stable Cones - Properties

QBS is a category

Cartesian : $|X \times Y| = |X| \times |Y|$ and
$\mathcal{R}(X \times Y) = \{\lambda r.(\alpha(r), \beta(r)) \mid \alpha \in \mathcal{R}(X), \beta \in \mathcal{R}(Y)$
Closed : $\left|Y^X\right| = \mathbf{QBS}(X, Y)$ and
$\mathcal{R}(Y^X) = \{\alpha \mid \lambda(r, x).\alpha(r)(x) \in \mathbf{QBS}(\mathbb{R} \times X \to Y)\}$
Limits : Coproducts, Quotients, ... as in Sets

QBS is a **conservative extension** of Standard Borel Sets

One uniform distribution is sufficient to generate all probability measures on Borel spaces.
if $\vdash d$: `X dist`, then there is $\alpha \in \mathcal{R}(X)$ such that
`sample d` $\sim$ `let r = sample uniform a:0. b:1. in` α(`r`)

Measure μ on a QBS is a borel space Σ, a random element $\alpha \in \mathcal{R}(X)$ and a measure on Σ. If $f : X \to \mathbb{R}^+$, then its integral with respect to μ :

$$\int_X \mu f = \int_\Sigma \mu(dr)(f(\alpha(r))$$

Take home

Semantics for (discrete) probabilistic programs

The operational semantics of continuous probability using kernels

The category **Meas** is not a CCC.

The Measurable Cones solution.

The Quasi Borel Spaces solution.

Both are sound models of probabilistic higher order programs

References

Semantics of probabilistic programs by Dexter Kozen

A convenient category for higher-order probability theory by Chris Heunen, Ohad Kammar, Sam Staton, Hongseok Yang.

Probabilistic Coherent Spaces by V. Danos and T. Ehrhard

Measurable Cones and Stable, Measurable Functions Ehrhard, et al.

Borel structures for function spaces by Aumann

A lambda-calculus foundation for universal probabilistic programming. by Johannes Borgström, Ugo Dal Lago, Andrew D. Gordon, and Marcin Szymczak.