Higher-Order Probabilistic Programming

A Tutorial at POPL 2019

Part II

Ugo Dal Lago

(Based on joint work with *Flavien Breuvart*, *Raphaëlle Crubillé*, *Charles Grellois*, *Davide Sangiorgi*,...)

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

Inria informatiques mathématiques

POPL 2019, Lisbon, January 14th

A Foundation for Higher-Order Probabilistic Programming

- We are interested in a better understanding of some crucial questions about higher-order probabilistic programs, e.g.:
  - How could we *formalise* and *prove* programs to have certain desireable **properties**, like being terminating or consuming a bounded amount of resources?
  - How could we prove programs to be **equivalent**, or more generally to be in a certain relation?

A Foundation for Higher-Order Probabilistic Programming

- We are interested in a better understanding of some crucial questions about higher-order probabilistic programs, e.g.:
  - How could we *formalise* and *prove* programs to have certain desireable **properties**, like being terminating or consuming a bounded amount of resources?
  - How could we prove programs to be **equivalent**, or more generally to be in a certain relation?
- We could in principle answer these questions directly in a programming language like OCAML.

A Foundation for Higher-Order Probabilistic Programming

- We are interested in a better understanding of some crucial questions about higher-order probabilistic programs, e.g.:
  - How could we *formalise* and *prove* programs to have certain desireable **properties**, like being terminating or consuming a bounded amount of resources?
  - How could we prove programs to be **equivalent**, or more generally to be in a certain relation?
- We could in principle answer these questions directly in a programming language like OCAML.
- It is methodologically much better to distill some paradigmatic calculi which expose all the essential features, but which are somehow agnostic to many unimportant details.
- We will introduce and study two such calculi:
  - $\mathsf{PCF}_{\oplus}$, a calculus for randomized higher-order programming.
  - $\mathsf{PCF}_{\mathtt{sample},\mathtt{score}}$, a calculus for bayesian programming.

$\mathsf{PCF}_\oplus$: Types, Terms, Values

$$
\begin{array}{llcl}
\textbf{Types} & \tau, \rho & ::= & \textsc{Unit} \mid \textsc{Num} \mid \tau \to \rho \\[1em]
\textbf{Terms} & M, N & ::= & V \mid V\ W \mid \texttt{let}\ M = x\ \texttt{in}\ N \mid M \oplus N \\
 & & & \mid\ \texttt{if}\ V\ \texttt{then}\ M\ \texttt{else}\ N \mid f_n(V_1, \ldots, V_n) \\[1em]
\textbf{Values} & V, W & ::= & \star \mid x \mid r \mid \lambda x.M \mid \texttt{fix}\ x.V
\end{array}
$$

PCF$_\oplus$: Type Assignment Rules

Value Typing Rules

$$\frac{}{\Gamma \vdash \star : \textsc{Unit}}\ \mathsf{S} \qquad \frac{}{\Gamma, x : \tau \vdash x : \tau}\ \mathsf{V} \qquad \frac{}{\Gamma \vdash r : \textsc{Num}}\ \mathsf{R}$$

$$\frac{\Gamma, x : \tau \vdash M : \rho}{\Gamma \vdash \lambda x.M : \tau \to \rho}\ \lambda \qquad \frac{\Gamma, x : \tau \to \rho \vdash M : \tau \to \rho}{\Gamma \vdash \mathtt{fix}\ x.M : \tau \to \rho}\ \mathsf{X}$$

Term Typing Rules

$$\frac{\Gamma \vdash V : \tau \to \rho \quad \Gamma \vdash W : \tau}{\Gamma \vdash V\ W : \rho}\ \mathtt{@} \qquad \frac{\Gamma \vdash M : \tau \quad \Gamma, x : \tau \vdash N : \rho}{\Gamma \vdash \mathtt{let}\ M = x\ \mathtt{in}\ N : \rho}\ \mathsf{L} \qquad \frac{\Gamma \vdash M : \tau \quad \Gamma \vdash N : \tau}{\Gamma \vdash M \oplus N : \tau}\ \oplus$$

$$\frac{\Gamma \vdash V : \textsc{Num} \quad \Gamma \vdash M : \tau \quad \Gamma \vdash N : \tau}{\Gamma \vdash \mathtt{if}\ V\ \mathtt{then}\ M\ \mathtt{else}\ N : \tau}\ \mathsf{I} \qquad \frac{\Gamma \vdash V_1 : \textsc{Num} \quad \cdots \quad \Gamma \vdash V_n : \textsc{Num}}{\Gamma \vdash f_n(V_1, \ldots, V_n) : \textsc{Num}}\ \mathsf{F}$$

$\mathsf{PCF}_{\oplus}$: Type Assignment Rules

Value Typing Rules

$$\frac{}{\Gamma \vdash \star : \textsc{Unit}}\ \mathsf{S} \qquad \frac{}{\Gamma, x : \tau \vdash x : \tau}\ \mathsf{V} \qquad \frac{}{\Gamma \vdash r : \textsc{Num}}\ \mathsf{R}$$

$$\frac{\Gamma, x : \tau \vdash M : \rho}{\Gamma \vdash \lambda x.M : \tau \to \rho}\ \lambda \qquad \frac{\Gamma, x : \tau \to \rho \vdash M : \tau \to \rho}{\Gamma \vdash \mathtt{fix}\ x.M : \tau \to \rho}\ \mathsf{X}$$

Term Typing Rules

$$\frac{\Gamma \vdash V : \tau \to \rho \quad \Gamma \vdash W : \tau}{\Gamma \vdash V\ W : \rho}\ @ \qquad \frac{\Gamma \vdash M : \tau \quad \Gamma, x : \tau \vdash N : \rho}{\Gamma \vdash \mathtt{let}\ M = x\ \mathtt{in}\ N : \rho}\ \mathsf{L} \qquad \frac{\Gamma \vdash M : \tau \quad \Gamma \vdash N : \tau}{\Gamma \vdash M \oplus N : \tau}\ \oplus$$

$$\frac{\Gamma \vdash V : \textsc{Num} \quad \Gamma \vdash M : \tau \quad \Gamma \vdash N : \tau}{\Gamma \vdash \mathtt{if}\ V\ \mathtt{then}\ M\ \mathtt{else}\ N : \tau}\ \mathsf{I} \qquad \frac{\Gamma \vdash V_1 : \textsc{Num} \quad \cdots \quad \Gamma \vdash V_n : \textsc{Num}}{\Gamma \vdash f_n(V_1, \ldots, V_n) : \textsc{Num}}\ \mathsf{F}$$

- The closed terms of type τ forms a set $\mathbb{CT}_{\tau}$.
- Similarly for values and $\mathbb{CV}_{\tau}$.

Distributions

- Given any set X, a *distribution* on X is a function $\mathcal{D} : X \to \mathbb{R}_{[0,1]}$ such that $\mathcal{D}(x) > 0$ only for denumerably many elements of X and that $\sum_{x \in X} \mathcal{D}(x) \leq 1$.

Distributions

- Given any set X, a *distribution* on X is a function $\mathcal{D} : X \to \mathbb{R}_{[0,1]}$ such that $\mathcal{D}(x) > 0$ only for denumerably many elements of X and that $\sum_{x \in X} \mathcal{D}(x) \leq 1$.
- The *support* of a distribution $\mathcal{D}$ on X is the subset $\mathsf{SUPP}(\mathcal{D})$ of X defined as

$$\mathsf{SUPP}(\mathcal{D}) := \{x \in X \mid \mathcal{D}(x) > 0\}$$

Distributions

- Given any set X, a *distribution* on X is a function $\mathcal{D} : X \to \mathbb{R}_{[0,1]}$ such that $\mathcal{D}(x) > 0$ only for denumerably many elements of X and that $\sum_{x \in X} \mathcal{D}(x) \leq 1$.
- The *support* of a distribution $\mathcal{D}$ on X is the subset $\mathsf{SUPP}(\mathcal{D})$ of X defined as

$$\mathsf{SUPP}(\mathcal{D}) := \{x \in X \mid \mathcal{D}(x) > 0\}$$

- The set of all distributions over X is indicated as $\mathbf{D}(X)$.

Distributions

- Given any set X, a *distribution* on X is a function $\mathcal{D} : X \to \mathbb{R}_{[0,1]}$ such that $\mathcal{D}(x) > 0$ only for denumerably many elements of X and that $\sum_{x \in X} \mathcal{D}(x) \leq 1$.
- The *support* of a distribution $\mathcal{D}$ on X is the subset $\mathsf{SUPP}(\mathcal{D})$ of X defined as

$$\mathsf{SUPP}(\mathcal{D}) := \{x \in X \mid \mathcal{D}(x) > 0\}$$

- The set of all distributions over X is indicated as $\mathbf{D}(X)$.
- We indicate the distribution assigning probability 1 to the element $x \in X$ and 0 to any other element of X as $\delta(x)$.

Distributions

- Given any set X, a *distribution* on X is a function $\mathcal{D} : X \to \mathbb{R}_{[0,1]}$ such that $\mathcal{D}(x) > 0$ only for denumerably many elements of X and that $\sum_{x \in X} \mathcal{D}(x) \leq 1$.
- The *support* of a distribution $\mathcal{D}$ on X is the subset $\mathsf{SUPP}(\mathcal{D})$ of X defined as

$$\mathsf{SUPP}(\mathcal{D}) := \{x \in X \mid \mathcal{D}(x) > 0\}$$

- The set of all distributions over X is indicated as $\mathbf{D}(X)$.
- We indicate the distribution assigning probability 1 to the element $x \in X$ and 0 to any other element of X as $\delta(x)$.
- Given a distribution $\mathcal{D}$ on X, its *sum* $\sum \mathcal{D}$ is simply $\sum_{x \in X} \mathcal{D}(x)$.

PCF$_\oplus$: Reduction

One-Step Reduction

$$
\begin{aligned}
(\lambda x.M)V &\to \delta(M[V/x]) \\
\texttt{let } V = x \texttt{ in } M &\to \delta(M[V/x]) \\
\texttt{if } 0 \texttt{ then } M \texttt{ else } N &\to \delta(M) \\
\texttt{if } r \texttt{ then } M \texttt{ else } N &\to \delta(N) \text{ if } r \neq 0 \\
M \oplus N &\to \left\{ M : \frac{1}{2}, N : \frac{1}{2} \right\} \\
f(r_1, \cdots, r_n) &\to \delta(f^*(r_1 \ldots, r_n))
\end{aligned}
$$

$$
\frac{M \to \{L_i : p_i\}_{i \in I}}{\texttt{let } M = x \texttt{ in } N \to \{\texttt{let } L_i = x \texttt{ in } N : p_i\}_{i \in I}}
$$

$\mathsf{PCF}_\oplus$: Reduction

Step-Indexed Reduction

$$\frac{}{M \Rightarrow_0 \emptyset} \qquad \frac{}{V \Rightarrow_1 \delta(V)} \qquad \frac{}{V \Rightarrow_{n+1} \emptyset} \qquad \frac{M \to \mathcal{D} \quad \forall N \in \mathsf{SUPP}(\mathcal{D}).N \Rightarrow_n \mathcal{E}_N}{M \Rightarrow_{n+1} \sum_{N \in \mathsf{SUPP}(\mathcal{D})} \mathcal{D}(N) \cdot \mathcal{E}_N}$$

$\mathsf{PCF}_{\oplus}$: Some Easy Meta-Theorems

Lemma

If $M \Rightarrow_n \mathcal{D}$, then $\mathsf{SUPP}(\mathcal{D})$ is a finite set.

$\mathsf{PCF}_{\oplus}$: Some Easy Meta-Theorems

Lemma

If $M \Rightarrow_n \mathcal{D}$, then $\mathsf{SUPP}(\mathcal{D})$ is a finite set.

Proposition (Progress)

For every $M \in \mathbb{CT}_\tau$, either M is a value or there is $\mathcal{D}$ with $M \to \mathcal{D}$

$\mathsf{PCF}_{\oplus}$: Some Easy Meta-Theorems

Lemma

If $M \Rightarrow_n \mathcal{D}$, then $\mathsf{SUPP}(\mathcal{D})$ is a finite set.

Proposition (Progress)

For every $M \in \mathbb{CT}_\tau$, either M is a value or there is $\mathcal{D}$ with $M \to \mathcal{D}$

Proposition (Subject Reduction)

For every $M \in \mathbb{CT}_\tau$ and for every $n \in \mathbb{N}$, if $M \to \mathcal{D}$ and $M \Rightarrow_n \mathcal{E}$, then $\mathcal{D} \in \mathbf{D}(\mathbb{CT}_\tau)$ and $\mathcal{E} \in \mathbf{D}(\mathbb{CV}_\tau)$.

$\mathsf{PCF}_{\oplus}$: Some Easy Meta-Theorems

Lemma

If $M \Rightarrow_n \mathcal{D}$, then $\mathsf{SUPP}(\mathcal{D})$ is a finite set.

Proposition (Progress)

For every $M \in \mathbb{CT}_\tau$, either M is a value or there is $\mathcal{D}$ with $M \to \mathcal{D}$

Proposition (Subject Reduction)

For every $M \in \mathbb{CT}_\tau$ and for every $n \in \mathbb{N}$, if $M \to \mathcal{D}$ and $M \Rightarrow_n \mathcal{E}$, then $\mathcal{D} \in \mathbf{D}(\mathbb{CT}_\tau)$ and $\mathcal{E} \in \mathbf{D}(\mathbb{CV}_\tau)$.

Corollary

For every $M \in \mathbb{CT}_\tau$ and for every $n \in \mathbb{N}$, there is exactly one distribution $\mathcal{D}_n$ such that $M \Rightarrow_n \mathcal{D}_n$. We will write $\langle M \rangle_n$ for such a distribution.

$\mathsf{PCF}_{\oplus}$: The Operational Semantics of a Term

- Given two distributions $\mathcal{D}, \mathcal{E} \in \mathbf{D}(X)$, we write $\mathcal{D} \leq \mathcal{E}$ iff $\mathcal{D}(x) \leq \mathcal{E}(x)$ for every $x \in X$. This relation endows $\mathbf{D}(X)$ with the structure of a partial order, which is actually an $\omega\mathbf{CPO}$:
- Given a closed term $M \in \mathbb{CT}_{\tau}$, the *operational semantics* of M is defined to be the distribution $\langle M \rangle \in \mathbb{CV}_{\tau}$ defined as $\sum_{n \in \mathbb{N}} \langle M \rangle_n$.

$\mathsf{PCF}_{\oplus}$: The Operational Semantics of a Term

- Given two distributions $\mathcal{D}, \mathcal{E} \in \mathbf{D}(X)$, we write $\mathcal{D} \leq \mathcal{E}$ iff $\mathcal{D}(x) \leq \mathcal{E}(x)$ for every $x \in X$. This relation endows $\mathbf{D}(X)$ with the structure of a partial order, which is actually an $\omega\mathbf{CPO}$:
- Given a closed term $M \in \mathbb{CT}_\tau$, the *operational semantics* of M is defined to be the distribution $\langle M \rangle \in \mathbb{CV}_\tau$ defined as $\sum_{n\in\mathbb{N}} \langle M \rangle_n$.

$$
\begin{array}{llll}
\textbf{Term Contexts} & C_{\mathbb{T}}, D_{\mathbb{T}} & ::= & C_{\mathbb{V}} \;\Big|\; [\cdot] \;\Big|\; C_{\mathbb{V}}\, V \;\Big|\; V\, C_{\mathbb{V}} \\
 & & & \Big|\; \mathtt{let}\ C_{\mathbb{T}} = x\ \mathtt{in}\ N \;\Big|\; \mathtt{let}\ M = x\ \mathtt{in}\ C_{\mathbb{T}} \;\Big|\; C_{\mathbb{T}} \oplus D_{\mathbb{T}} \\
 & & & \Big|\; \mathtt{if}\ V\ \mathtt{then}\ C_{\mathbb{T}}\ \mathtt{else}\ D_{\mathbb{T}} \\
\textbf{Value Contexts} & C_{\mathbb{V}}, D_{\mathbb{V}} & ::= & \lambda x.C_{\mathbb{T}} \;\Big|\; \mathtt{fix}\ x.C_{\mathbb{V}}
\end{array}
$$

$\mathsf{PCF}_\oplus$: The Operational Semantics of a Term

- Given two distributions $\mathcal{D}, \mathcal{E} \in \mathbf{D}(X)$, we write $\mathcal{D} \leq \mathcal{E}$ iff $\mathcal{D}(x) \leq \mathcal{E}(x)$ for every $x \in X$. This relation endows $\mathbf{D}(X)$ with the structure of a partial order, which is actually an ω**CPO**:
- Given a closed term $M \in \mathbb{CT}_\tau$, the *operational semantics* of M is defined to be the distribution $\langle M \rangle \in \mathbb{CV}_\tau$ defined as $\sum_{n\in\mathbb{N}} \langle M \rangle_n$.

$$
\begin{array}{llcl}
\textbf{Term Contexts} & C_{\mathbb{T}}, D_{\mathbb{T}} & ::= & C_{\mathbb{V}} \;\Big|\; [\cdot] \;\Big|\; C_{\mathbb{V}}\, V \;\Big|\; V\, C_{\mathbb{V}} \\
& & & \Big|\; \mathtt{let}\ C_{\mathbb{T}} = x\ \mathtt{in}\ N \;\Big|\; \mathtt{let}\ M = x\ \mathtt{in}\ C_{\mathbb{T}} \;\Big|\; C_{\mathbb{T}} \oplus D_{\mathbb{T}} \\
& & & \Big|\; \mathtt{if}\ V\ \mathtt{then}\ C_{\mathbb{T}}\ \mathtt{else}\ D_{\mathbb{T}} \\
\textbf{Value Contexts} & C_{\mathbb{V}}, D_{\mathbb{V}} & ::= & \lambda x.C_{\mathbb{T}} \;\Big|\; \mathtt{fix}\ x.C_{\mathbb{V}}
\end{array}
$$

- Given two terms M, N such that $\Gamma \vdash M : \tau$ and $\Gamma \vdash N : \tau$, we say that M and N are (Γ, τ)-*equivalent*, and we write $M \equiv_\Gamma^\tau N$ iff whenever $\emptyset \vdash C[\Gamma \vdash \cdot : \tau] : \textsc{Unit}$, it holds that $\sum \langle C[M] \rangle = \sum \langle C[N] \rangle$.

$\mathsf{PCF}_{\oplus}$: from Equivalences to Metrics?

- The definition of contexual equivalence asks that $\sum\langle C[M]\rangle = \sum\langle C[N]\rangle$ for every context C.
- But what if $\sum\langle C[M]\rangle$ and $\sum\langle C[N]\rangle$ are very close, without being really equal to each other?
- It makes sense to generalize contextual equivalence to a notion of **distance**:

$$\delta^{\Gamma,\tau}(M,N) = \sup_{\emptyset\vdash C[\Gamma\vdash\cdot:\tau]:\textsc{Unit}} \left|\sum\langle C[M]\rangle - \sum\langle C[N]\rangle\right|.$$

- For every Γ,τ, $\delta^{\Gamma,\tau}$ is indeed a *pseudo-metric*:

$$\delta^{\Gamma,\tau}(M,M) = 0$$
$$\delta^{\Gamma,\tau}(M,N) = \delta^{\Gamma,\tau}(N,M)$$
$$\delta^{\Gamma,\tau}(M,L) \leq \delta^{\Gamma,\tau}(M,N) + \delta^{\Gamma,\tau}(N,L)$$

Termination in a Probabilistic Setting

- Let M be any closed term. We say that M is **almost surely terminating** if $\sum\langle M\rangle = 1$, namely if its probability of convergence is 1.

Termination in a Probabilistic Setting

- Let M be any closed term. We say that M is **almost surely terminating** if $\sum\langle M\rangle = 1$, namely if its probability of convergence is 1.
- **Example**:

$$GEO := (\texttt{fix}\ f.\lambda x.x \oplus (\texttt{let}\ succ_1(x) = y\ \texttt{in}\ f\ y))0$$

Termination in a Probabilistic Setting

- Let M be any closed term. We say that M is **almost surely terminating** if $\sum\langle M\rangle = 1$, namely if its probability of convergence is 1.
- **Example**:

$$GEO := (\texttt{fix}\ f.\lambda x.x \oplus (\texttt{let}\ succ_1(x) = y\ \texttt{in}\ f\ y))0$$

- The **expected evaluation length** of any closed term as follows:

$$ExLen(M) := \sum_{m=0}^{\infty}\left(1 - \sum_{n=0}^{m}\sum\langle M\rangle_n\right)$$

Let M be any closed term. We say that M is **positively almost surely terminating** if $ExLen(M) < +\infty$.

Termination in a Probabilistic Setting

- Let M be any closed term. We say that M is **almost surely terminating** if $\sum\langle M\rangle = 1$, namely if its probability of convergence is 1.
- **Example**:

$$GEO := (\mathtt{fix}\ f.\lambda x.x \oplus (\mathtt{let}\ succ_1(x) = y\ \mathtt{in}\ f\ y))0$$

- The **expected evaluation length** of any closed term as follows:

$$ExLen(M) := \sum_{m=0}^{\infty}\left(1 - \sum_{n=0}^{m}\sum\langle M\rangle_n\right)$$

Let M be any closed term. We say that M is **positively almost surely terminating** if $ExLen(M) < +\infty$.

Lemma

Every positively almost-surely terminating term is almost-surely terminating.

Variations on PCF$_\oplus$

- **Pree, Untyped** rather than applied, typed.
  - **Terms**: $M \ ::= \ x \ \big| \ \lambda M. \ \big| \ MM \ \big| \ M \oplus M;$
  - **Values**: $V \ ::= \ \lambda M.;$
  - **One-Step Reduction**:

$$(\lambda x.M)V \to \delta(M[V/x]) \qquad\qquad M \oplus N \to \left\{M : \frac{1}{2}, N : \frac{1}{2}\right\}$$

$$\frac{M \to \{L_i : p_i\}_{i \in I}}{MN \to \{L_i N : p_i\}_{i \in I}} \qquad \frac{M \to \{L_i : p_i\}_{i \in I}}{VM \to \{VL_i : p_i\}_{i \in I}}$$

  - The obtained calculus will be referred to as $\Lambda_\oplus$.

Variations on PCF$_\oplus$

- **Pree, Untyped** rather than applied, typed.
  - **Terms**: $M \ ::= \ x \ \Big| \ \lambda M. \ \Big| \ MM \ \Big| \ M \oplus M;$
  - **Values**: $V \ ::= \ \lambda M.;$
  - **One-Step Reduction**:

$$(\lambda x.M)V \to \delta(M[V/x]) \qquad M \oplus N \to \left\{M : \frac{1}{2}, N : \frac{1}{2}\right\}$$

$$\frac{M \to \{L_i : p_i\}_{i\in I}}{MN \to \{L_i N : p_i\}_{i\in I}} \qquad \frac{M \to \{L_i : p_i\}_{i\in I}}{VM \to \{VL_i : p_i\}_{i\in I}}$$

  - The obtained calculus will be referred to as $\Lambda_\oplus$.
- **CBN** rathen than CBV.
  - **One-Step Reduction**:

$$(\lambda x.M)N \to \delta(M[N/x]) \qquad M \oplus N \to \left\{M : \frac{1}{2}, N : \frac{1}{2}\right\}$$

$$\frac{M \to \{L_i : p_i\}_{i\in I}}{MN \to \{L_i N : p_i\}_{i\in I}}$$

From Randomised Algorithms to Bayesian Programming

- Binary probabilistic choice is perfectly adequate to model randomised computation.

From Randomised Algorithms to Bayesian Programming

- Binary probabilistic choice is perfectly adequate to model randomised computation.

Theorem

The class of computable probabilistic functions coincides with the class of probabilistic functions computable by $\mathsf{PCF}^{\mathbb{N}}_{\oplus}$.

From Randomised Algorithms to Bayesian Programming

- Binary probabilistic choice is perfectly adequate to model randomised computation.

Theorem

The class of computable probabilistic functions coincides with the class of probabilistic functions computable by $\mathsf{PCF}^{\mathbb{N}}_{\oplus}$.

- In recent years, starting from the pioneering works on languages like CHURCH, ANGLICAN, or HANSEI, functional programs have been also employed as means to represent probabilistic *models* rather than *algorithms*.
- The languages above can be modeled [Staton2017] as λ-calculi endowed with two new operators:
  - `sample`, modeling sampling from the uniform distribution on $[0, 1]$.
  - `score`, which takes a positive real number r as a parameter, and modify the *weight* of the current probabilistic branch by multiplying it by r.

$\mathsf{PCF}_{\mathtt{sample},\mathtt{score}}$: Terms, Typing Rules, and Reduction

Terms $\quad M, N ::= \mathtt{sample} \mid \mathtt{score}(V).$

$\mathsf{PCF}_{\texttt{sample},\texttt{score}}$: Terms, Typing Rules, and Reduction

Terms $\quad M, N ::= \texttt{sample} \mid \texttt{score}(V).$

Typing Rules

$$\frac{}{\Gamma \vdash \texttt{sample} : \textsc{Num}}\ \mathsf{A} \qquad \frac{\Gamma \vdash V : \textsc{Num}}{\Gamma \vdash \texttt{score}(V) : \textsc{Unit}}\ \mathsf{C}$$

$\mathsf{PCF}_{\texttt{sample},\texttt{score}}$: Terms, Typing Rules, and Reduction

Terms $\qquad M, N ::= \texttt{sample} \mid \texttt{score}(V).$

Typing Rules $\qquad \dfrac{}{\Gamma \vdash \texttt{sample} : \textsc{Num}}\ \mathsf{A} \qquad \dfrac{\Gamma \vdash V : \textsc{Num}}{\Gamma \vdash \texttt{score}(V) : \textsc{Unit}}\ \mathsf{C}$

- One needs to switch from distributions to *measures*, and assume the underlying set, namely $\mathbb{R}$ to have the structure of a measurable space
- Adapting the rule for `let`-terms naturally leads to

$$\frac{M \to \mu}{\texttt{let}\ M = x\ \texttt{in}\ N \to \texttt{let}\ \mu = x\ \texttt{in}\ N}$$

  where $\texttt{let}\ \mu = x\ \texttt{in}\ N$ should itself be a measure.
- The key rule in step-indexed reduction needs to be adapted:

$$\frac{M \to \mu \quad \forall N \in \mathsf{SUPP}(\mu). N \Rightarrow_n \sigma_N}{M \Rightarrow_{n+1} A \mapsto \int \sigma_N(A)\mu(dN)}$$

The Operational Meaning of Bayesian Terms

- It can well be that $\langle M \rangle = \mu$, where μ sums to something *strictly higher* than 1.

The Operational Meaning of Bayesian Terms

- It can well be that $\langle M \rangle = \mu$, where μ sums to something *strictly higher* than 1.
- How should we interpret $\mu(A)$ for a measurable set of terms A?
- We need to *normalize* μ!

The Operational Meaning of Bayesian Terms

- It can well be that $\langle M \rangle = \mu$, where μ sums to something *strictly higher* than 1.
- How should we interpret $\mu(A)$ for a measurable set of terms A?
- We need to *normalize* μ!
- Explicitly building a normalized version of μ (if it exists) is the goal of so-called *inference algorithms*.

The Operational Meaning of Bayesian Terms

- It can well be that $\langle M \rangle = \mu$, where μ sums to something *strictly higher* than 1.
- How should we interpret $\mu(A)$ for a measurable set of terms A?
- We need to *normalize* μ!
- Explicitly building a normalized version of μ (if it exists) is the goal of so-called *inference algorithms*.
- The operational semantics we have just introduced, called **distribution-based** is thus just an *idealized* form of semantics.
- An *executable* semantics can be given in the form of **sampling-based semantics**.

Sampling-Based Semantics (1)

$$\begin{aligned}
\langle (\lambda x.M)V, s\rangle &\xrightarrow{1}\!\!\!\!\rightarrow \langle M[V/x], s\rangle \\
\langle \texttt{let}\ V = x\ \texttt{in}\ M, s\rangle &\xrightarrow{1}\!\!\!\!\rightarrow \langle M[V/x], s\rangle \\
\langle \texttt{if}\ 0\ \texttt{then}\ M\ \texttt{else}\ N, s\rangle &\xrightarrow{1}\!\!\!\!\rightarrow \langle M, s\rangle \\
\langle \texttt{if}\ r\ \texttt{then}\ M\ \texttt{else}\ N, s\rangle &\xrightarrow{1}\!\!\!\!\rightarrow \langle N, s\rangle \text{ if } r \neq 0 \\
\langle \texttt{sample}, r :: s\rangle &\xrightarrow{1}\!\!\!\!\rightarrow \langle r, s\rangle \\
\langle \texttt{score}(r), s\rangle &\xrightarrow{r}\!\!\!\!\rightarrow \langle \star, s\rangle \\
\langle f(r_1, \cdots, r_n), s\rangle &\xrightarrow{1}\!\!\!\!\rightarrow \langle f^*(r_1 \ldots, r_n), s\rangle
\end{aligned}$$

$$\frac{\langle M, s\rangle \xrightarrow{r}\!\!\!\!\rightarrow \langle L, t\rangle}{\langle \texttt{let}\ M = x\ \texttt{in}\ N, s\rangle \xrightarrow{r}\!\!\!\!\rightarrow \langle \texttt{let}\ L = x\ \texttt{in}\ N, s\rangle}$$

Sampling-Based Semantics (2)

$$\frac{}{\langle V, s\rangle \overset{1}{\Rrightarrow} \langle V, s\rangle} \qquad \frac{\langle M, s\rangle \overset{r}{\twoheadrightarrow} \langle N, t\rangle \quad \langle N, t\rangle \overset{s}{\Rrightarrow} \langle L, u\rangle}{\langle M, s\rangle \overset{r\cdot s}{\Rrightarrow} \langle L, u\rangle}$$

Thank You!

Questions?