A probabilistic lambda calculus - Some preliminary investigations

Ugo Dal Lago, Margherita Zorzi

Università di Bologna, Università di Verona

June, 9-11, 2010, Torino

Introduction: Λ_P

- We present some results about probabilistic type-free λ-calculus Λ_P. We will focus on the operational behavior of lambda terms, giving basic results about operational semantics and reduction strategies.
- The syntax of Λ_P is a simply extension of the syntax of lambda calculus with a choice operator $\oplus$:

$$x \mid \lambda x.M \mid M(N) \mid M \oplus N$$

where $M \oplus N$ rewrites to M or to N with the same probability $1/2$.

State of Art

- Probabilistic functional paradigm has been taken into account in last years. The related investigations are oriented to programming languages.
  "Probability distributions are useful for expressing the meanings of probabilistic languages, which support formal modeling of and reasoning about uncertainty ... " (Ramsey-Pfeffer).
- Moreover, in general, the *monadic approach* is considered.
- In the literature, there is still a lack of investigation about foundational results.

State of Art: three fundamental papers

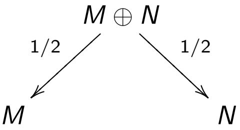
Our investigation is inspired, in several aspects, by the following works:

- `Call-by-Value, Call-by-Name and the` λ`-calculus`
  Plotkin, 1974
- `Nondeterministic Extensions of Untyped` λ`-calculus`
  De'Liguoro & Piperno, 1995
- `Coinductive big step semantics`
  Leroy & Grall 2008

Syntax of Λ_P

$$x \mid \lambda x.M \mid M(N) \mid M \oplus N$$

where



- A term is a value iff it is a variable or a lambda-abstraction.
- We have considered and defined both Call-by-Value and Call-by-Name strategies; here we will focus our attention on CbV one.

Leftmost reduction (CbyV)

We define a probabilistic version of *leftmost reduction*, between Λ_P and Λ_P^*, where Λ_P^* is the set of the sequences $\overline{M} \equiv M_1, \ldots, M_n$ of lambda terms.

Some rules:

$$(\lambda x.M)V \rightarrow_{\mathsf{v}} M[V/x]$$
$$V \oplus W \rightarrow_{\mathsf{v}} V, W$$
$$M \oplus N \rightarrow_{\mathsf{v}} \overline{L} \oplus N \qquad \text{if } M \rightarrow_{\mathsf{v}} \overline{L}$$

where $V, W \in \mathsf{Val}$ and $\overline{L} \oplus N$ is $L \oplus N_1, \ldots, L \oplus N_n$ whenever $\overline{L}$ is $L_1, \ldots, L_n$

Note: we are working in a *weak* setting.

Distributions

Definition

(i) A *probabilistic distribution* is a function $\mathscr{D} : \mathsf{Val} \rightarrow \mathbb{R}_{[0,1]}$ for which there exists a set $Q \subseteq \mathsf{Val}$ with $\mathscr{D}(V) = 0$ except when $V \in Q$ and moreover $\sum_{v \in V} \mathscr{D}(V) = 1$.

(ii) A *probabilistic sub-distribution* is defined as probabilistic distribution but $\sum_{v \in V} \mathscr{D}(V) \leq 1$.

where Val is the set of values in Λ_P.

Distributions

- In the following, we will use distributions of observable results in the definition of semantics.
- Informally, in Λ_P a computation of a term M reduces in a distribution of possible observables: in the evaluation we take into account the outputs of all probabilistic choices, i.e. all the possible results of a computation from $M \in \Lambda_P$.
- We will use the linear notation $\{V_1^{p_1}, \ldots, V_k^{p_k}\}$ to denote the probabilistic distribution and sub-distribution attributing p_i to V_i (for every i).

Towards operational semantics

- We would to deal with both finite and infinite computations.
- We would like to capture, in an elegant way, the operational behavior of all kind of computations. Some terms are, in fact, "critical".

Example

$$G \equiv Y\,\overbrace{(\lambda f.\lambda x.\ \text{if } (0 \oplus 1) \text{ then } x \text{ else } f(\text{succ}(x)))}^{R}$$

$$G(0) = Y(R)(0) = R(YR)(0) \to (\ \text{if } (0 \oplus 1) \text{ then } 0 \text{ else } (YR)(\text{succ}(0))) \to \dots$$

then $G(0)$ evaluates to the distribution $\{0^{1/2}, 1^{1/4}, 2^{1/8}, \dots\} = \{n^{\frac{1}{2^{(n+1)}}}\}_{n\in\mathbb{N}}$

- `Co-induction` is an useful instrument in order to characterize divergence and infinite computations.

Small-step semantics...

- We define an inductive binary relation $\Rightarrow$ between Λ_P and sub-distributions.

$$\frac{}{M \Rightarrow \emptyset} \qquad \frac{}{V \Rightarrow \{V^1\}}$$

$$\frac{M \rightarrow_{\mathsf{v}} \overline{N} \quad N_i \Rightarrow \mathscr{D}_i}{M \Rightarrow \sum_{i=1}^{n} \frac{1}{n} \mathscr{D}_i}$$

$\mathscr{S}(M)$ is the sub-distribution $\sup_{M \Rightarrow \mathscr{D}} \mathscr{D}$.

... and small steps semantics for divergence

- We define also a coinductive binary relation $\Rightarrow^{\infty}_{p}$ between Λ_P and $\mathbb{R}_{[0,1]}$ capturing divergence.

$$\frac{}{V \Rightarrow^{\infty}_{0}}$$

$$\frac{M \rightarrow_{\mathsf{v}} \overline{N} \qquad N_i \Rightarrow^{\infty}_{p_i}}{M \Rightarrow^{\infty}_{\sum_{i=1}^{n} \frac{1}{n} p_i}}$$

The relation $\Rightarrow^{\infty}_{p}$ is related to our definition of *observational equivalence* $M \approx N$: $C[M] \Rightarrow^{\infty}_{p}$ iff $C[N] \Rightarrow^{\infty}_{p}$.

Big-step semantics rules

There is a coinductively defined binary relation $\Downarrow$ between Λ_P and sub-distributions:

$$\overline{\overline{V \Downarrow \{V\}}}$$

$$\frac{M \Downarrow \{\lambda x_i.M_i^{p_i}\}_i \quad N \Downarrow \{V_j^{q_j}\}_j \quad (M_i[V_j/x_i] \Downarrow \{W_{ijk}^{r_{ijk}}\})_{i,j}}{MN \Downarrow \{W_{ijk}^{p_i q_j r_{ijk}}\}_{i,j,k}}$$

$$\frac{M \Downarrow \{V_i^{p_i}\} \quad N \Downarrow \{W^{q_j}\}}{M \oplus N \Downarrow \{V_i^{\frac{1}{2}p_i \sum_j q_j}, W_j^{\frac{1}{2}q_j \sum_i p_i}\}_{i,j}}$$

$\mathscr{B}(M)$ is the sub-distribution $\inf_{M \Downarrow \mathscr{D}} \mathscr{D}$.

Relating different semantics: Convergence and Divergence in Small-step

- The following theorem relates Small-step semantics and divergence relation...

Theorem

For each $M \in \Lambda_P$, $M \Rightarrow_p^\infty$ *iff* $\mathscr{S}(M) = \mathscr{D} = \{v_1^{p_1}, \ldots, v_k^{p_k} \ldots\}$ *with* $\sum_i p_i = q$ *and* $p + q = 1$.

Relating different semantics: Small-step vs Big-step

- ... and the following one states the relationship between Small-step and Big-step semantics:

Theorem

For each $M \in \Lambda_P$, $\mathscr{B}(M) = \mathscr{S}(M)$.

Simulation CbyV-CbyN

Let us consider the following example (Selinger, 2004):

Example

$M \equiv (\lambda x.x \text{ plus } x).((\lambda xy.x) \oplus (\lambda xy.y))$, where plus is a xor-like function.
Choosing a call-by name strategy, the computation gives the distribution $\{(\lambda xy.x)^{1/2}, (\lambda xy.y)^{1/2}\}$ as result , whereas call-by name strategy gives $(\lambda xy.y)$ with probability 1.

CbyV and CbyN are not equivalent strategies, not even on ground values.

- In general, the order in which probabilistic choices and duplications are performed is relevant.

Simulation CbyV-CbyN

- We are able to extend Plotkin's simulation between Call-by-Value and Call-by-Name in the case of finite computations, following the same CPS approach.
- We define recursively the term simulation maps $[\cdot]$, which sends terms in the Call-by-Value language $\mathscr{L}_{\mathsf{v}}$, into terms in the call Call-by-Name language $\mathscr{L}_{\mathsf{n}}$.
  For example, for the choice, the map is defined as

$$[M \oplus N] = (\lambda\epsilon.[M](\lambda\alpha.[N](\lambda\beta.(\alpha \oplus \beta))\epsilon)$$

Simulation CbyV-CbyN

The following theorem holds:

Theorem (Simulation)

For each closed term M, $\boldsymbol{\Psi}(\mathscr{B}_{\mathsf{v}}(M)) = \mathscr{B}_{\mathsf{n}}([M](\lambda x.x))$

where $\boldsymbol{\Psi}$ is a function which maps distributions of values in $\mathscr{L}_{\mathsf{v}}$ into distributions of values in $\mathscr{L}_{\mathsf{n}}$.

- We proved also the reverse simulations CbyN$\mapsto$CbyV.

Future directions

- Simulation for infinite computations
- Probabilistic Böhm Trees
- Separability for Λ_P
- Relationship between algebraic distance and separability (?)

Grazie!

Merci!

Thanks!

Vic toor e smursma