Probabilistic Böhm Trees and Probabilistic Separation

Thomas Leventis
IRIF, UMR 8243, Univ Paris Diderot, Sorbonne Paris Cité
F-75205 Paris, France
thomas.leventis@irif.fr

Abstract

We study the notion of observational equivalence in the call-by-name probabilistic λ-calculus, where two terms are said observationally equivalent if under any context, their head reductions converge with the same probability. Our goal is to generalise the separation theorem to this probabilistic setting. To do so we define probabilistic Böhm trees and probabilistic Nakajima trees, and we mix the well-known Böhm-out technique with some new techniques to manipulate and separate probability distributions.

CCS Concepts • **Mathematics of computing → Lambda calculus**;

Keywords probabilistic lambda-calculus, Böhm trees, observational equivalence, separation

ACM Reference Format:
Thomas Leventis. 2018. Probabilistic Böhm Trees and Probabilistic Separation. In *LICS '18: LICS '18: 33rd Annual ACM/IEEE Symposium on Logic in Computer Science, July 9–12, 2018, Oxford, United Kingdom.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3209108.3209126

1 Introduction

Given a programming language, one can consider several notions of equality on the programs of this language. The most basic one, the syntactic equality, is useful to study the details of the computations, for instance for complexity purposes. Another one consists in completely hiding the computations and considering the equivalence induced by the computation rules: if P reduces into Q then P and Q are equivalent. A third one is the *observational equivalence* which does not necessarily arise naturally from the construction of our calculus, but which directly translates what we expect from equal programs: given some notions of environment and observation, two programs P and Q are said equal if for any environment E, we observe the same behaviour from $E[P]$ and $E[Q]$.

LICS '18, July 9–12, 2018, Oxford, United Kingdom

ACM ISBN 978-1-4503-5583-4/18/07…$15.00
https://doi.org/10.1145/3209108.3209126

In the pure untyped λ-calculus we usually consider contexts as the environment, and the termination of the reduction as the observation [5]. The equality thus obtained is also known to have a more syntactic characterisation: it corresponds to the *Nakajima tree* (or *Böhm tree modulo infinite extensionality*) equality. The separation theorem (or semi-separation) states that if two terms have different Nakajima trees then they are separable, i.e. they are not observationally equivalent [1]. A stronger result, proven independently by Hyland [7] and Wadsworth [9], states that the Nakajima tree equality and the observational equivalence actually coincide, and that moreover they correspond to the maximal sensible consistent λ-theory, i.e. the maximal non-trivial equality which equates all diverging terms.

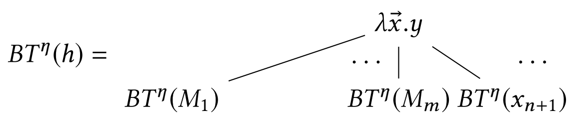
This result holds in the pure λ-calculus, but it does not always hold in its extensions. For instance if we add a non-deterministic choice operator + to the calculus, such that $M + N$ reduces non-deterministically into M or N, then we can still define a notion of Böhm tree, and some weak separation property holds [4], but the correspondence between the non-deterministic Nakajima trees and the observational equivalence fails.

In this paper we are interested in the observational equivalence in the probabilistic λ-calculus. This calculus does not simply involve a non-deterministic choice, but probabilistic choices $+_p$ for all probabilities $p \in [0, 1]$ such that $M +_p N$ reduces into M with probability p and into N with probability $1 - p$. Some characterisation of the observational equivalence using bisimulation exists in the literature [3], but for a call-by-value version of this calculus; the correspondence fails when one consider the call-by-name variant, which we are interested in. There also exists a denotational model for the call-by-name probabilistic λ-calculus using probabilistic coherence spaces [6], which we conjecture characterises the observational equivalence, but this has not been proven.

In this paper we investigate the relation between the probabilistic observational equivalence and a probabilistic version of the Nakajima trees. The notion of Böhm tree is a generalisation of the notion of normal form, such that there is a tree associated to every term. In the deterministic λ-calculus they are defined co-inductively by:

$$T \in \mathcal{BT} := \lambda x_1 \dots x_n.y\ T_1\ \dots\ T_m \mid \Omega$$

To compute the tree of a term, we follow its head reduction; if its reduction terminates then its head normal form gives the first level of its Böhm tree and we then compute the trees of its subterms, and if its reduction does not terminate we simply say its tree is Ω. This notion is not sufficient to characterise the observational equivalence because of its lack of extensionality: the terms y and $\lambda z.y\ z$ have different Böhm trees and yet they are observationally equivalent. This is why Nakajima trees are defined as the infinitely η-expanded versions of the Böhm trees. The Nakajima tree $BT^\eta(h)$ of a head normal form $h = \lambda x_1 \ldots x_n.y\ M_1\ \ldots\ M_m$ is:



where $\lambda\vec{x} = \lambda x_1\ x_2\ \ldots$ describes an infinite number of abstractions. The separation theorem states that if two terms M and N have differents Nakajima trees then there is a context C such that $C[M]$ and $C[N]$ have different behaviours, i.e. one converges while the other diverges. Our goal here is to prove a similar theorem for the probabilistic λ-calculus.

In Sections 2 and 3 we will describe the call-by-name probabilistic λ-calculus and we will define the probabilistic observational equivalence. These are standard definitions and none of them is new.

In Section 4 we define the probabilistic Böhm trees, and we introduce their infinitely extensional variant in Section 5.

The rest of the paper is dedicated to the proof of the separation theorem. In Section 6 we sketch the differences between the deterministic and the probabilistic separations. In Section 7 we define some tools to associate probabilities to Nakajima trees. Finally the Sections 8 and 9 are dedicated to the proofs of Propositions 8.1 and 9.1, which together give the main result of this paper, Theorem 10.1.

Notations

We will write $\mathbb{N}$ for the set of natural numbers, $\mathbb{N}^*$ for the set of non-zero natural numbers, $\mathbb{Z}$ for the set of all integers and $\mathbb{R}$ for the set of real numbers.

Given a reduction relation $\to$, we write $\twoheadrightarrow$ for its reflexive transitive closure.

We call *subprobability distribution* over a set X any function $P : X \to [0, 1]$ such that $\sum_{x \in X} P(x) \leq 1$.

2 The Probabilistic λ-Calculus

The probabilistic λ-calculus is an extension of the basic λ-calculus with probabilistic choice. The *probabilistic λ-terms* over a given set *Var* of variables are defined inductively by:

$$M, N \in \Lambda_+ := x \in Var \mid \lambda x.M \mid M\ N \mid M +_p N, p \in [0, 1].$$

A sum term $M +_p N$ represents a choice and intuitively reduces to M with probability p and to N with probability $1 - p$. We express this with a labelled reduction system: the reduction rules for the sum are $M +_p N \xrightarrow{p} M$ and $M +_p N \xrightarrow{1-p} N$. Then we also write the usual β-reduction

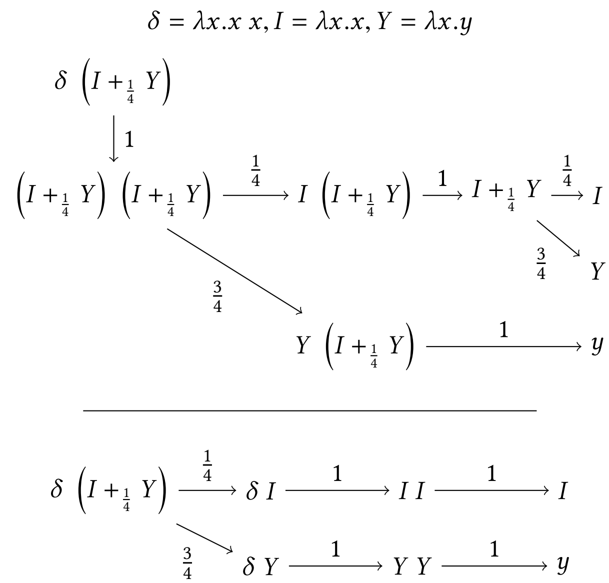


Figure 1. Call-by-name versus call-by-value

Reduction rules:

$$(\lambda x.M)\ N \xrightarrow{1} M\left[{}^{N}/_{x}\right]$$

$$M +_p N \xrightarrow{p} M$$

$$M +_p N \xrightarrow{1-p} N$$

Evaluation contexts:

$$H := \lambda x_1 \ldots x_n.[\ ]\ P_1\ \ldots\ P_m$$

Figure 2. Probabilistic head reduction

as a labelled reduction, with a unique possible reduction with probability 1 for a β-redex: $(\lambda x.M)\ N \xrightarrow{1} M\left[{}^{N}/_{x}\right]$ where $M\left[{}^{N}/_{x}\right]$ is the term obtained by substituting the free occurrences of x in M by N.

Such a non-deterministic reduction requires the specification of a reduction strategy. Indeed if we apply for instance a term which uses its argument twice, such as $\delta = \lambda x.x\ x$, to a choice then using a call-by-name strategy or a call-by-value strategy will yield very different results (see Figure 1). In this paper we are interested in Böhm trees which are defined through head normalisation, so we will consider only the head reduction of terms. The reduction rules are given in Figure 2. To shorten the notations we will sometimes write $\lambda\vec{x}.[\ ]\ \vec{P}$ for the head contexts.

3 Probabilistic Observations

The deterministic case In the deterministic λ-calculus every term is either a head normal form, or it has a unique head redex with a unique possible reduction. Hence we observe two kinds of behaviours: either a term has a unique head normal form reached by a unique reduction path, or it doesn't and its head reduction never terminates.

Similarly probabilistic term may be of three shapes:

$\lambda\vec{x}.y\ \vec{P}$	(head normal form)
$\lambda\vec{x}.(\lambda y.M)\ N\ \vec{P}$	(head β-redex)
$\lambda\vec{x}.(M +_p N)\ \vec{P}$	(head sum)

Every term has at most one head redex, but the head-reduction is no longer unique as we have two reductions rules for the sums. Thus a single probabilistic term may reduce in any number of head normal forms. So what we are interested in is the probability to reach each of them. Let us write $\mathcal{H}$ for the set of all head normal forms. For every term M we define a subprobability distribution over $\mathcal{H}$ describing the probability for M to reduce into each head normal form h, as the sum of the probabilities of all the reductions from M to h:

$$\mathcal{P}(M \twoheadrightarrow h) = \sum_{M=M_0 \xrightarrow{p_1} M_1 \xrightarrow{p_2} \ldots \xrightarrow{p_l} M_l = h} \prod_{k=1}^{l} p_k.$$

Remark: This notation for the reductions is ambiguous. For instance from the term $x +_{\frac{1}{2}} x$ there are two possible reductions and both are written $x +_{\frac{1}{2}} x \xrightarrow{\frac{1}{2}} x$. To distinguish them one could add labels l and r to the reduction to explicit whether we are reducing to the left side or to the right side of a sum. In this paper we will assume that defining probabilities as sums "over all the reductions" from a term is explicit enough rather than introduce heavier notations. Also note that there may be infinitely many normalizing reductions from a single term, so $\mathcal{P}(M \twoheadrightarrow h)$ is defined as a possibly infinite sum, but we can prove that it is actually well-defined and in $[0, 1]$. See for instance [6] for more details on this definition.

We then base the notion of observation on the *convergence probability* $\mathcal{P}_{\Downarrow}(M)$ of a term M, i.e. the total mass of the distribution $\mathcal{P}(M \twoheadrightarrow \_)$:

$$\mathcal{P}_{\Downarrow}(M) = \sum_{h \in \mathcal{H}} \mathcal{P}(M \twoheadrightarrow h) \in [0, 1]$$

Example 3.1. Let $\Omega = \delta\ \delta$ where $\delta = \lambda x.x\ x$. For $M = (x +_{\frac{1}{3}} \Omega) +_{\frac{1}{2}} y$ we have:

$$\mathcal{P}(M \twoheadrightarrow x) = \frac{1}{6}, \mathcal{P}(M \twoheadrightarrow y) = \frac{1}{2}, \mathcal{P}_{\Downarrow}(M) = \frac{2}{3}.$$

Let $\underline{n}$ be the n-th Church numeral, $\underline{\text{succ}}$ be an encoding of the successor, and Θ the usual fixpoint combinator such that $\Theta\ M \xrightarrow{1} \cdot \xrightarrow{1} M\ (\Theta\ M)$ for all M. Let $N = \Theta\ (\lambda n.(\underline{0} +_{\frac{1}{2}} \underline{\text{succ}}))$, we have:

$$\mathcal{P}\left(N \twoheadrightarrow \underline{n}\right) = \frac{1}{2^{n+1}} \qquad \mathcal{P}_{\Downarrow}(N) = 1.$$

Remark: If M is a deterministic term, i.e. a term without probabilistic sum, then either M has a unique head normal form h and $\mathcal{P}(M \twoheadrightarrow h) = 1$ and $\mathcal{P}_{\Downarrow}(M) = 1$, or M is a diverging term and $\mathcal{P}_{\Downarrow}(M) = 0$. So $\mathcal{P}_{\Downarrow}(\_)$ generalises the usual deterministic observation.

Two probabilistic terms are said to be observationnally equivalent if they have the same convergence probability in any context [3, 6].

Definition 3.2. The *observationnal equivalence* $=_{obs}$ is defined by $M =_{obs} N$ iff for every context C:

$$\mathcal{P}_{\Downarrow}(C[M]) = \mathcal{P}_{\Downarrow}(C[N])$$

4 Probabilistic Böhm Trees

The deterministic trees Deterministic Böhm trees are defined as "infinite left normal forms" of λ-terms. The Böhm trees of finite depth $d \in \mathbb{N}$ are described in Figure 3. One can actually define infinite Böhm trees using a co-inductive variant of this definition, but we are mostly interested in the equivalence on terms induced by the Böhm tree equality rather than in the trees themselves, so we will simply say that two terms M and N have the same Böhm tree whenever $BT_d(M) = BT_d(N)$ for all $d \in \mathbb{N}$.

To generalise this construction to probabilistic terms, it is more convenient to distinguish two different layers in such trees. First we define the trees $BT_{d+1}(h)$ for the head normal forms h, and second we define the tree $BT_{d+1}(M)$ by $BT_{d+1}(M) = BT_{d+1}(h)$ if $M \twoheadrightarrow h$, and $BT_{d+1} = \Omega$ if M diverges. It then becomes natural to define the Böhm tree of a probabilistic head normal form in the same way than for the deterministic ones, and the Böhm tree of a term as a subprobability distribution over such trees. We obtain the definition given in Figure 4. Remark that for a term M and $t \in \mathcal{VT}_{d+1}$ we define $PT_{d+1}(M)(t)$ as a possibly infinite sum, but we know that $\mathcal{P}(M \twoheadrightarrow \_)$ is a subprobability distribution over $\mathcal{H}$ so $PT_{d+1}(M)$ is a subprobability distribution as well. Also note that at level 0, $\mathcal{VT}_0 = \emptyset$ so for all M $PT_0(M)$ is the unique function $\emptyset \to [0, 1]$ and the function VT_0 is not defined.

Remark: To better match the notations of the deterministic trees, we could view $PT_d(M)$ as a probability distribution over $\mathcal{VT}_d \cup \{\Omega\}$, with $PT_d(M)(\Omega) = 1 - \sum_{t \in \mathcal{VT}_d} PT_d(M)(t)$.

Example 4.1. Probabilistic Böhm trees are subprobability distributions, so they are not, properly speaking, trees. To keep a graphical, tree-like representation we write distributions as trees with a root + and edges labelled with probabilities. This means a + node may have infinitely many children, and these children are not ordered. The empty distributions will be denoted as a single node + with no children.

$$\mathcal{BT}_0 = \{\Omega\}$$

$$\mathcal{BT}_{d+1} = \{\Omega\} \cup \left\{ \begin{array}{c} \lambda x_1 \ldots x_n.y \\ \diagup \ | \ \ldots \diagdown \\ T_1 \quad T_2 \qquad T_m \end{array} \mid \forall i \leq m, T_i \in \mathcal{BT}_d \right\}$$

$$BT_0(M) = \Omega$$

$$BT_{d+1}(M) = \begin{cases} \begin{array}{c} \lambda \vec{x}.y \\ \diagup \ \ldots \ \diagdown \\ BT_d(P_1) \qquad BT_d(P_m) \end{array} & \text{if } M \twoheadrightarrow \lambda \vec{x}.y\ P_1\ \ldots\ P_m \\ \Omega & \text{otherwise} \end{cases}$$

Figure 3. Finite deterministic Böhm trees [1]

$$\mathcal{PT}_d = \{\text{subprobability distributions over } \mathcal{VT}_d\}$$

$$\mathcal{VT}_0 = \emptyset$$

$$\mathcal{VT}_{d+1} = \left\{ \begin{array}{c} \lambda x_1 \ldots x_n.y \\ \diagup \ | \ \ldots \diagdown \\ T_1 \quad T_2 \qquad T_m \end{array} \mid \forall i \leq m, T_i \in \mathcal{PT}_d \right\}$$

$$PT_{d+1}(M) : t \mapsto \sum_{\substack{h \in \mathcal{H} \text{ s.t.} \\ VT_{d+1}(h) = t}} \mathcal{P}(M \twoheadrightarrow h)$$

$$PT_0(M) : \emptyset \to [0, 1]$$

$$VT_{d+1}(\lambda \vec{x}.y\ P_1\ \ldots\ P_m) = \begin{array}{c} \lambda \vec{x}.y \\ \diagup \ \ldots \ \diagdown \\ PT_d(P_1) \qquad PT_d(P_m) \end{array}$$

Figure 4. Probabilistic Böhm trees

For the deterministic terms $\delta = \lambda x.x\ x$ we have:

$$BT_2(\delta) = \begin{array}{c} \lambda x.x \\ | \\ x \end{array} \text{ and } PT_2(\delta) = \begin{array}{c} + \\ 1\,| \\ \lambda x.x \\ | \\ + \\ 1\,| \\ x \end{array}$$

For $\Omega = \delta\ \delta$, for all $d \in \mathbb{N}$, $BT_d(\Omega) = \Omega$ and $PT_d(\delta)$ is the empty subprobability distribution, represented by a single node $+$.

For the probabilistic terms $M = x +_{\frac{1}{2}} y\ (z +_{\frac{1}{3}} \Omega)$ and $N = \Theta\ (\lambda y.(x +_{\frac{1}{2}} f\ y))$ we have:

$$PT_2(M) = \begin{array}{c} \frac{1}{2} \ + \ \frac{1}{2} \\ \diagup \quad \diagdown \\ x \qquad y \\ \qquad | \\ \qquad + \\ \qquad \frac{1}{3}\,| \\ \qquad z \end{array} \text{ and } PT_3(N) = \begin{array}{c} \frac{1}{2} \ + \ \frac{1}{2} \\ \diagup \quad \diagdown \\ x \qquad f \\ \qquad | \\ \qquad \frac{1}{2} \ + \ \frac{1}{2} \\ \qquad \diagup \quad \diagdown \\ \qquad x \qquad f \\ \qquad\qquad | \\ \qquad\qquad \frac{1}{2} \ + \ \frac{1}{2} \\ \qquad\qquad \diagup \quad \diagdown \\ \qquad\qquad x \qquad f \\ \qquad\qquad\qquad | \\ \qquad\qquad\qquad + \end{array}$$

5 Probabilistic Nakajima Trees

The deterministic case The Böhm tree equality does not actually characterise the observational equivalence: it lacks extentionality. For instance the terms x and $\lambda y.x\ y$ have different Böhm trees and yet they are observationaly equivalent. The canonical representatives of the observational equivalence are the Nakajima trees, which are a variant of the Böhm trees quotiented by infinite extensionality. Given a head normal form $h = \lambda x_1 \ldots x_n.y\ P_1\ \ldots\ P_m$, the Nakajima tree $BT^{\eta}_{d+1}(h)$ of h is built as a Böhm tree over the infinite η-expansion of h:

$$BT^{\eta}_{d+1}(h) = \begin{array}{c} \lambda x_1 \ldots x_n z_1 \ldots .y \\ \diagup \quad \ldots \ | \quad \diagdown \qquad \ldots \\ BT^{\eta}_d(P_1) \qquad BT^{\eta}_d(P_m) \quad BT^{\eta}_d(z_1) \end{array}$$

Such a definition is easy to read, but as we will manipulate heavy notations in our proofs we will give an alternate definition using lighter (albeit less intuitive) notations by choosing once and for all the names of the abstracted variables. Indeed the above definition uses infinite labels, as Nakajima trees describe infinite η-expansions, and moreover to be accurate we should quotient these trees modulo variable renaming. By fixing in advance the names of the variables, we do not need to explicit them anymore, and we do not need to worry about variable renaming.

We assume given a family $\left(x^d_n\right)_{d \in \mathbb{N}, n \in \mathbb{N}^*}$ of pairwise distinct variables, such that $Var \setminus \{x^d_n \mid d \in \mathbb{N}, n \in \mathbb{N}^*\}$ is still infinite. Then to define the Nakajima tree of depth $d+1$ of a head normal form h (whose free variables are different from the $x^{d'}_n$ for $d' \leq d+1$) we write $h = \lambda x^{d+1}_1 \ldots x^{d+1}_n.y\ P_1\ \ldots\ P_m$ and we define:

$$BT^{\eta}_{d+1}(h) = \begin{array}{c} y \\ \diagup \quad \ldots \diagup \quad \diagdown \qquad \ldots \\ BT^{\eta}_d(P_1) \qquad BT^{\eta}_d(P_m) \quad BT^{\eta}_d\left(x^{d+1}_{n+1}\right) \end{array}$$

For more technical reasons we will also want the labels to describe not only a head variable but also an integer, corresponding to the difference between the number of arguments of a head normal form and its number of abstractions. Given a head normal form $h = \lambda x_1 \ldots x_n.y\ P_1\ \ldots\ P_m$, for all i large enough the i-th child of $BT^{\eta}_{d+1}(h)$ is $BT^{\eta}_d\left(x^d_{i-(m-n)}\right)$. This value $m - n$ is an important piece of information to manipulate a head normal form, and it plays a crucial role in

$$\begin{aligned}
\mathcal{L} &= Var \times \mathbb{Z} \quad \text{(set of labels)} \\
\mathcal{PT}^{\eta}_{d} &= \left\{\text{subprobability distributions over } \mathcal{VT}^{\eta}_{d}\right\} \\
\mathcal{VT}^{\eta}_{0} &= \emptyset \\
\mathcal{VT}^{\eta}_{d+1} &= \left\{((y,s),(T_i)_{i\in\mathbb{N}^*}) \in \mathcal{L} \times \left(\mathcal{PT}^{\eta}_{d}\right)^{\mathbb{N}^*}\right\}
\end{aligned}$$

$$PT^{\eta}_{d+1}(M) : t \mapsto \sum_{\substack{h\in\mathcal{H} \text{ s.t.} \\ VT^{\eta}_{d+1}(h)=t}} \mathcal{P}\left(M \twoheadrightarrow h\right)$$
$$PT^{\eta}_{0}(M) \ : \emptyset \to [0,1]$$

$$VT^{\eta}_{d+1}(\lambda x^{d+1}_1 \ldots x^{d+1}_n.y\ P_1\ \ldots\ P_m) = ((y, m-n), (T_i))$$
$$\text{with } T_i = \begin{cases} PT^{\eta}_{d}(P_i) & \text{if } i \leq m \\ PT^{\eta}_{d}\left(x^{d+1}_{i-m+n}\right) & \text{if } i > m \end{cases}$$

Figure 5. Probabilistic Nakajima trees

the separation proof (cf. Lemma 8.3). Remark that this modification changes the notion of Nakajima tree equality at a fixed finite depth (now we may have $BT^{\eta}_1(h) \neq BT^{\eta}_1(h')$ even if h and h' have the same head variable), but we ultimately have the same general notion of Nakajima tree equality on terms.

In the end we get the definition given in Figure 5.

Definition 5.1. Two probabilistic terms M and N *have the same Nakajima tree*, which we write $M =_{\mathcal{PB}^{\eta}} N$, if $PT^{\eta}_d(M) = PT^{\eta}_d(N)$ for all $d \in \mathbb{N}$.

Proposition 5.2. *For any probabilistic terms M and N, if $PT^{\eta}_d(M) = PT^{\eta}_d(N)$ for some $d \in \mathbb{N}$ then $PT^{\eta}_{d'}(M) = PT^{\eta}_{d'}(N)$ for all $d' \leq d$.*

Proof. By contraposition, we prove by induction on d that if $PT^{\eta}_d(M) \neq PT^{\eta}_d(N)$ then $PT^{\eta}_{d+1}(M) \neq PT^{\eta}_{d+1}(N)$. □

6 Probabilistic Separation

The deterministic separation The separation in the untyped deterministic λ-calculus is a consequence of Lemma 8.3, which lets us substitute terms *only* for the head occurrence of a variable. Given two head normal forms $h = \lambda x_1 \ldots x_n.y\ P_1\ \ldots\ P_m$ and $h' = \lambda x_1 \ldots x_{n'}.y'\ P'_1\ \ldots\ P'_{m'}$, if either $y \neq y'$ or $m-n \neq m'-n'$ (i.e. if $VT^{\eta}_1(h) \neq VT^{\eta}_1(h')$) then the lemma immediately gives a separating context, which for instance substitute x for one head variable and Ω for the other. Otherwise if their trees differ at some deeper position, we can use the lemma to put a projection in head position to dig out this difference.

In the deterministic case we usually separate two terms, and to do so we follow a single branch of their Nakajima trees until we reach a difference. The situation gets more complicated in a quantitative setting. Consider for instance the terms $M = x\ y\ \Omega +_{\frac{1}{2}} x\ \Omega\ y$ and $N = x\ y\ y +_{\frac{1}{2}} x\ \Omega\ \Omega$. At depth 1 both of these terms reduce with probability 1 into x applied to two arguments. At depth 2, for both terms the first argument is y with probability $\frac{1}{2}$. The same goes for the second argument. So we need to look at both arguments at once to separate M and N.

Next let us define:

$$P = x\ u\ v +_{\frac{1}{3}} (x\ v\ \Omega +_{\frac{1}{2}} x\ \Omega\ u)$$
$$Q = x\ \Omega\ v +_{\frac{1}{3}} (x\ v\ u +_{\frac{1}{2}} x\ u\ \Omega).$$

Not only do we need to look at both arguments of x to see a difference between these two terms, but by doing so we will encounter three different terms at each position, u, v and Ω, and we will need to deal with all of them.

The counterpart to this complexity is that we have more freedom as to how we separate terms. In the deterministic case the observation distinguishes two behaviours: convergence and divergence. So we can separate two terms by making one diverge and one converge, but a n-ary separation result does not make sense if we do not refine our definition of observation. On the other hand we observe the convergence probability of probabilistic terms, which ranges over $[0,1]$, so it makes sense to separate several terms at once.

The proof of our separation theorem is a generalisation of the usual proof for the deterministic case. The latter relies on the notion of branch. The key technical result is the *Böhm-out technique*, which is used to build a context which somehow digs out the label at the end of a branch. Then if two trees are different there obviously exists a branch which leads to different labels, and the result follows.

In the probabilistic case branches are insufficient, so in Section 7 we introduce a new notion of *evaluation structures* on which we base our proof. In Section 8 we rely on the same technical result than the Böhm-out technique (Lemma 8.3) to prove that the evaluation of a tree matches the computation of the convergence probability of a term in some particular context. Finally we prove that given finitely many trees, there exists an evaluation structure which separates them.

7 Evaluation structures

We would like to define some notion of evaluation procedure $\mathcal{PT}^{\eta}_d \to [0,1]$ such that for all $d \in \mathbb{N}$:

- given finitely many pairwise distinct probabilistic Nakajima trees $T_1, \ldots, T_m \in \mathcal{PT}^{\eta}_d$ there is a procedure $\mathcal{E}$ such that $\mathcal{E}(T_1), \ldots, \mathcal{E}(T_m)$ are pairwise distinct;
- given finitely many terms $M_1, \ldots, M_m \in \Lambda_+$ and a procedure $\mathcal{E}$ there is a context C such that $\mathcal{P}_{\Downarrow}(C[M_i]) = \mathcal{E}(PT^{\eta}_d(M_i))$ for all $i \leq m$.

The second requirement is actually too strong. We do not know if it can be achieved, but neither do we need it: given $M_1, \ldots, M_m$ and $\mathcal{E}$ it is enough to find for every $\epsilon > 0$ a context C such that $|\mathcal{P}_{\Downarrow}(C[M_i]) - \mathcal{E}(PT^{\eta}_d(M_i))| \leq \epsilon$ for all $i \leq m$. More precisely we consider the term $\underline{\alpha} = I +_{\alpha} \Omega$, where $I = \lambda x.x$ and Ω is a diverging term, as the "ideal"

$$\mathcal{S} = \left(\mathcal{L}^{\mathcal{S}}, m^{\mathcal{S}}, \left(\varphi_l^{\mathcal{S}}\right)_{l \in \mathcal{L}^{\mathcal{S}}}, (\mathcal{S}_{l,i})_{l \in \mathcal{L}^{\mathcal{S}}, 1 \leq i \leq m^{\mathcal{S}}}\right)$$

where:

- $\mathcal{L}^{\mathcal{S}} \subset_f Var \times \mathbb{Z}$ is a finite set of labels,
- $m^{\mathcal{S}} \in \mathbb{N}$ is an arity,
- $\varphi_l^{\mathcal{S}} \in \mathcal{F}_{m^{\mathcal{S}}}$ is a $m^{\mathcal{S}}$-ary evaluation function,
- $\mathcal{S}_{l,i}$ is an evaluation structure of depth d.

Figure 6. Evaluation structure $\mathcal{S}$ of depth $d + 1$

For $\mathcal{S}$ evaluation structure of depth $d + 1$:

$$\begin{array}{rccc} \mathcal{E}^{\mathcal{S}} & : & \mathcal{PT}^{\eta}_{d+1} & \rightarrow & [0,1] \\ & & T & \mapsto & \sum_{t \in \mathcal{VT}^{\eta}_{d+1}} T(t) \times \mathcal{E}^{\mathcal{S}}_{\mathrm{V}}(t) \end{array}$$

$$\begin{array}{rccc} \mathcal{E}^{\mathcal{S}}_{\mathrm{V}} & : & \mathcal{VT}^{\eta}_{d+1} & \rightarrow & [0,1] \\ & & (l, (T_i)) & \mapsto & \end{array}$$

$$\begin{cases} \varphi_l^{\mathcal{S}} \left(\mathcal{E}^{\mathcal{S}_{l,1}}(T_1), \ldots, \mathcal{E}^{\mathcal{S}_{l,m^{\mathcal{S}}}}(T_{m^{\mathcal{S}}})\right) & \text{if } l \in \mathcal{L}^{\mathcal{S}} \\ 0 & \text{otherwise} \end{cases}$$

Figure 7. Evaluation procedures

representation of a term with convergence probability $\alpha \in [0,1]$, and for $\epsilon > 0$ we say that a term M *ϵ-approximates $\underline{\alpha}$* if:

$$\alpha - \epsilon \leq \mathcal{P}(M \twoheadrightarrow I) \text{ and } \mathcal{P}_{\Downarrow}(M) \leq \alpha + \epsilon$$

i.e. M reduces to I with probability approximately α and diverges with probability approximately $1 - \alpha$ (and its behaviour is unknown for a probability at most 2ϵ). Then we want to prove Proposition 8.1: at any depth $d \in \mathbb{N}$, given $\mathcal{E}$, $M_1,\ldots,M_n$ and $\epsilon > 0$ there is a context C such that $C[M_i]$ ϵ-approximates $\underline{\mathcal{E}\left(PT^{\eta}_d(M_i)\right)}$.

To define such evaluation procedures we first define some functions to manipulate probabilities.

Definition 7.1. The set $\mathcal{F}_n$ of *n-ary evaluation functions* $[0,1]^n \rightarrow [0,1]$ is the set of polynomial functions

$$\varphi : (\alpha_1, \ldots, \alpha_n) \mapsto \sum_{\vec{k} \in \mathcal{K}} a_{\vec{k}} \prod_{i=1}^{n} \alpha_i^{k_i}$$

with $\mathcal{K}$ a finite subset of $\mathbb{N}^n$, such that $a_{\vec{k}} \in [0,1]$ for all $\vec{k} \in \mathcal{K}$ and $\sum_{k \in \mathcal{K}} a_{\vec{k}} \leq 1$.

The degree of an evaluation function φ is:

$$\deg(\varphi) = \sup_{\substack{\vec{k} \in \mathcal{K} \\ a_{\vec{k}} \neq 0}} \sum_{i=1}^{n} k_i$$

The reason we are interested in such functions is that they are continuously represented by probabilistic terms.

Proposition 7.2. *Every evaluation function $\varphi \in \mathcal{F}_n$ is represented by a term F such that for all $\epsilon > 0$, if $M_1,\ldots,M_n$ are terms which $\frac{\epsilon}{2\deg(\varphi)}$-approximate $\underline{\alpha_1},\ldots,\underline{\alpha_n}$ then the term $F\ M_1\ \ldots\ M_n$ ϵ-approximates $\underline{\varphi(\alpha_1, \ldots, \alpha_n)}$.*

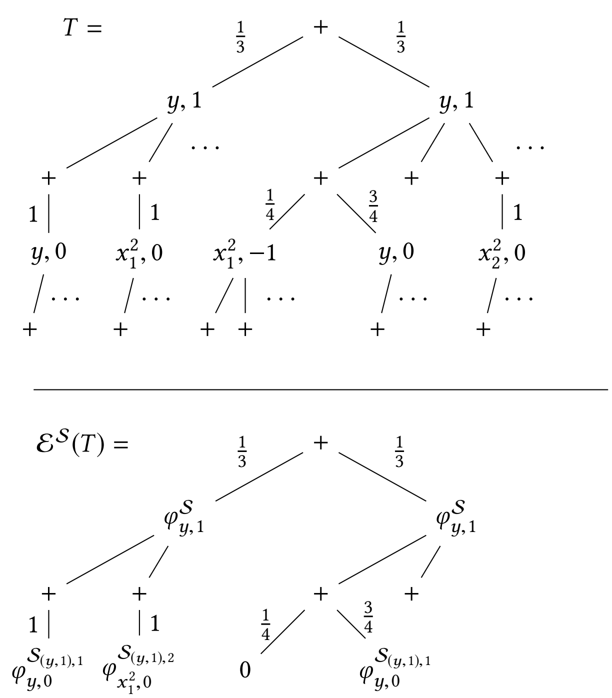


Figure 8. Example: evaluation of a tree

Proof. Remark that $\frac{\epsilon}{2\deg(\varphi)}$ is not defined for functions of degree 0, but those are constant functions which do not use their arguments.

Otherwise observe that if M δ-approximates α and N δ'-approximates β then $M\ N$ $(2\delta + \delta')$-approximates $\alpha\beta$. Then a monomial $a \times \alpha_1 \times \alpha_1 \times \cdots \times \alpha_n \times \alpha_n \times \alpha_n$ can be continuously represented by a sequence of applications $\underline{a}\ (M_1\ (M_1 \ldots (M_n\ (M_n\ M_n)) \ldots))$.

An evaluation function is a subprobability distribution over finitely many monomials. Thus it can be continuously represented by a term F of the form $\lambda x_1 \ldots x_n.F'$ where F' is a sum of terms representing monomials. □

Then we evaluate trees by replacing their nodes by evaluation functions and computing the result. More precisely we define evaluation structures, which detail how to substitute evaluation functions for nodes, in Figure 6 and the corresponding evaluation procedure is defined in Figure 7. Note that we did not define evaluation structures of depth 0, hence for an evaluation structure $\mathcal{S}$ of depth 1 we necessarily have $m^{\mathcal{S}} = 0$.

Example 7.3. Let us consider the tree $T \in \mathcal{PT}^{\eta}_2$ given by $T = PT^{\eta}_2\left(y\ y +_{\frac{1}{3}} \left(\lambda x_1^2.y\ \left(\lambda x_1^1.x_1^2 +_{\frac{1}{4}} y\right)\ \Omega +_{\frac{1}{2}} \Omega\right)\right)$. Let $\mathcal{S}$ be the evaluation structure of depth 2 such that:

- $\mathcal{L}^{\mathcal{S}} = \{(y,1),(z,3)\}$,

- $m^{\mathcal{S}} = 2$,
- $\varphi^{\mathcal{S}}_{y,1} : (\alpha, \beta) \mapsto \frac{1}{2}\alpha + \frac{1}{2}\alpha\beta$ and $\varphi^{\mathcal{S}}_{z,3} : (\alpha, \beta) \mapsto \frac{1}{5}$,
- $\mathcal{L}^{\mathcal{S}_{(y,1),1}} = \{(y, 0)\}$ and $\mathcal{L}^{\mathcal{S}_{(y,1),2}} = \{(x_1^2, 0)\}$,
- $m^{\mathcal{S}_{(y,1),1}} = m^{\mathcal{S}_{(y,1),2}} = 0$,
- $\varphi^{\mathcal{S}_{(y,1),1}}_{y,0} : () \mapsto \frac{1}{3}$ and $\varphi^{\mathcal{S}_{(y,1),2}}_{x_1^2,0} : () \mapsto 1$.

Remark that we should technically define $\mathcal{S}_{(z,3),1}$ and $\mathcal{S}_{(z,3),2}$, but these are irrelevant as $\varphi^{\mathcal{S}}_{z,3}$ does not use its arguments.

The evaluation $\mathcal{E}^{\mathcal{S}}(T)$ is described by the tree in Figure 8. We get

$$\begin{aligned}\mathcal{E}^{\mathcal{S}}(T) = &\ \frac{1}{3} \times \varphi^{\mathcal{S}}_{y,1}\left(\varphi^{\mathcal{S}_{(y,1),1}}_{y,0}(), \varphi^{\mathcal{S}_{(y,1),2}}_{x_1^2,0}()\right) \\ &+ \frac{1}{3} \times \varphi^{\mathcal{S}}_{y,1}\left(\frac{3}{4}\varphi^{\mathcal{S}_{(y,1),1}}_{y,0}(), 0\right) \\ = &\ \frac{11}{72}\end{aligned}$$

8 Terms, Trees and Evaluation

The first step to our separation result is to show that the evaluation of the tree of a term can be performed by a context.

Proposition 8.1. *Let $\mathcal{S}$ be an evaluation structure of depth $d+1$. Given finitely many terms $M_1,\ldots,M_m$, for all $\epsilon > 0$ there is a context C such that for all $i \le m$, $C[M_i]$ ϵ-approximates $\mathcal{E}^{\mathcal{S}}\left(PT^{\eta}_{d}(M_i)\right)$.*

Given Proposition 7.2 this result is just a generalisation of the Böhm-out technique, and we use the same proof techniques as in [7]. So the contexts we will build will be mostly deterministic: we will only use the probabilistic sum to represent evaluation functions, while the replacement of the nodes of the trees by those functions will be entirely deterministic.

The main difficulty here is that different occurrences of the same variable may receive different evaluation functions. For instance if we consider Example 7.3, to evaluate the term $y\ y +_{\frac{1}{3}} \left(\lambda x_1^2.y\ \left(\lambda x_1^1.x_1^2 +_{\frac{1}{4}} y\right)\ \Omega +_{\frac{1}{2}} \Omega\right)$ we cannot directly substitute a term representing $(\alpha, \beta) \mapsto \frac{1}{2}\alpha + \frac{1}{2}\alpha\beta$ to some occurrences of y and another term representing $() \mapsto \frac{1}{3}$ to the others.

To replace different occurrences of the same variable by different terms we use *permutators*, i.e. terms of the form $R_r = \lambda x_1 \ldots x_{r+1}.x_{r+1}\ x_1\ \ldots\ x_{r+1}$. Note that permutators are usually defined as terms of the form $\lambda x_1 \ldots x_{r+1}.x_{r+1}\ x_1\ \ldots\ x_r$, but our version slightly simplifies the notations (although it doesn't actually permute its arguments anymore). We will write substitutions as functions $\sigma : Var \to \Lambda_+$ such that $\sigma(x) \neq x$ for finitely many $x \in Var$, which we generalise to Λ_+ by induction:

- $\sigma(\lambda x.M) = \lambda x.\sigma'(M)$ with $\sigma'(y) = \begin{cases} x & \text{if } y = x \\ \sigma(y) & \text{if } y \neq x \end{cases}$;
- $\sigma(M\ N) = \sigma(M)\ \sigma(N)$;
- $\sigma(M +_p N) = \sigma(M) +_p \sigma(N)$.

We are interested in substitutions which only use permutators, but we are interested in three other properties as well: we want to be sure every variable we are interested in is actually replaced by a permutator; we want those permutators to reach far enough (if we substitute R_r for x in $x\ P_1\ \ldots\ P_m$, we want to have $r > m$ to be able to choose the term which will be fetched by the permutator); and we want different variables to be replaced by different permutators. So for any finite set of variables $Y \subset_f Var$ and any $\delta \in \mathbb{N}$ we define $\Sigma_{Y,\delta}$ as the set of substitutions σ such that:

- for all $y \in Y$, $\sigma(y) \neq y$;
- for all $x \in Var$, if $\sigma(x) \neq x$ then $\sigma(x) = R_{r_x}$ for some $r_x \ge \delta$;
- for all $x \neq x'$ with $\sigma(x) \neq x$ and $\sigma(x') \neq x'$, we have $|r_x - r_{x'}| \ge \delta$.

Remark that if $Y \subset Y'$ and $\delta \le \delta'$ then $\Sigma_{Y,\delta} \supset \Sigma_{Y',\delta'}$.

These substitutions are very convenient to manipulate terms: under any such substitution with sufficiently strong constraints, we can read the whole structure of a term just by giving it some arguments. In particular we will prove the following detailed version of Proposition 8.1.

Proposition 8.2. *Let $\mathcal{S}$ be an evaluation structure of depth $d+1$. Given finitely many terms $M_1,\ldots,M_m$, for all $\epsilon > 0$ there exists Y and δ such that for all $\sigma \in \Sigma_{Y,\delta}$, there are terms $L_1,\ldots,L_k$ such that for all $i \le m$, $\sigma(M_i)\ L_1\ \ldots\ L_k$ ϵ-approximates $\mathcal{E}^{\mathcal{S}}\left(PT^{\eta}_{d}(M_i)\right)$.*

Remark: For any substitution σ there is a context C such that for all M, $C[M] \twoheadrightarrow \sigma(M)$, so Proposition 8.1 is indeed a direct corollary of Proposition 8.2.

The key lemma to prove this proposition is that such contexts can substitute an arbitrary terms for the head variable of a head normal form without disturbing the global structure of the term.

Lemma 8.3. *Given finitely many head normal forms $h_i = \lambda x_1 \ldots x_{n_i}.y_i\ P_{i,1}\ \ldots\ P_{i,m_i}$, given any arity $m \in \mathbb{N}$, there exist Y and δ such that for all $\sigma \in \Sigma_{Y,\delta}$, given any family of terms $(F_l)_{l\in\mathcal{L}}$ there are terms $L_1,\ldots,L_k$ such that:*

$$\sigma(h_i)\ L_1\ \ldots\ L_k \twoheadrightarrow F_{l_i}\ \sigma\left(P'_{i,1}\right)\ \ldots\ \sigma\left(P'_{i,m}\right)$$

where $l_i = (y_i, m_i - n_i)$ *and* $P'_{i,j} = \begin{cases} P_{i,j} & \text{if } j \le m_i \\ x^d_{j-m_i+n_i} & \text{if } j > m_i \end{cases}$.

Proof. Given the h_is and the arity m:

- we choose Y such that $y_i \in Y$ for all i,
- we choose some $n \in \mathbb{N}$ such that $n \ge n_i$ for all i and $n + m_i - n_i \ge m$ for all i,
- and we choose δ such that $\delta > n + m_i - n_i$ for all i and $\delta > |(m_i - n_i) - (m_{i'} - n_{i'})|$ for all i and i'.

Then given $\sigma \in \Sigma_{Y,\delta}$ and the family (F_l):

- we define r_i as the integer such that $\sigma(y_i) = R_{r_i}$,

- we define $k = \max_i(r_i - (n + m_i - n_i))$,
- for all i we define $L_{r_i-(n+m_i-n_i)} = \lambda z_1 \dots z_{n+m_i-n_i+k}.F_{l_i}\ z_1\ \dots\ z_m$,
- and we choose arbitrarily other terms L_j to define all the $L_1,\dots,L_k$.

Remark that if $r_i - (n + m_i - n_i) = r_{i'} - (n + m_{i'} - n_{i'})$ then necessarily $y_i = y_{i'}$, otherwise we would have $|r_i - r_{i'}| \geq \delta$ (by definition of $\Sigma_{Y,\delta}$) and $\delta > |(m_i - n_i) - (m_{i'} - n_{i'})|$ (by construction). Then we also have $r_i = r_{i'}$, and $m_i - n_i = m_{i'} - n_{i'}$, hence $l_i = l_{i'}$. So the definition of the terms $L_{r_i-(n+m_i-n_i)}$ is sound.

Now for all i we have:

$$\sigma(h_i)\ \sigma(x_1^d)\ \dots\ \sigma(x_n^d)\ L_1\ \dots\ L_k \twoheadrightarrow F_{l_i}\ \sigma(P'_{i,1})\ \dots\ \sigma(P'_{i,m})$$

□

The previous lemma helps us deal with head normal forms, but we want to simulate the evaluation of arbitrary terms. We need to show that the behaviour of a term in our evaluation contexts is given by the behaviour of its head normal forms.

Lemma 8.4. *Given a term M, a substitution σ and terms $L_1,\dots,L_k$, for every head normal form h_0 we have*

$$\mathcal{P}\left(\sigma(M)\ \vec{L} \twoheadrightarrow h_0\right) = \sum_{h\in\mathcal{H}} \mathcal{P}\left(M \twoheadrightarrow h\right) \mathcal{P}\left(\sigma(h)\ \vec{L} \twoheadrightarrow h_0\right)$$

$$\mathcal{P}_{\Downarrow}(\sigma(M)\ \vec{L}) = \sum_{h\in\mathcal{H}} \mathcal{P}\left(M \twoheadrightarrow h\right) \mathcal{P}_{\Downarrow}(\sigma(h)\ \vec{L})$$

Proof. For $n \in \mathbb{N}$ let us write $\mathcal{P}\left(M \twoheadrightarrow^n h\right)$ for the probability for M to reduce in h in exactly n steps, we have can prove by induction on n that:

$$\mathcal{P}\left(\sigma(M)\ \vec{L} \twoheadrightarrow^n h_0\right) = \sum_{l+l'=n} \sum_{h\in\mathcal{H}} \mathcal{P}\left(M \twoheadrightarrow^l h\right) \mathcal{P}\left(\sigma(h)\ \vec{L} \twoheadrightarrow^{l'} h_0\right)$$

□

With these two lemmas we can prove Proposition 8.2. We reason by induction on the evaluation structure $\mathcal{S}$.

First we reason on finitely many head normal forms $h_i = \lambda x_1^{d+1} \dots x_{n_i}^{d+1}.y\ P_{i,1}\ \dots\ P_{i,m_i}$. Let $l_i = (y_i, m_i - n_i)$, if $l_i \in \mathcal{L}^{\mathcal{S}}$ we have

$$\mathcal{E}_{\mathrm{V}}^{\mathcal{S}}(VT_{d+1}^{\eta}(h_i)) = \varphi_{l_i}^{\mathcal{S}}\left(\mathcal{E}^{\mathcal{S}_{l_i,1}}\left(PT_d^{\eta}(P'_{i,1})\right),\dots,\mathcal{E}^{\mathcal{S}_{l_i,m^{\mathcal{S}}}}\left(PT_d^{\eta}(P'_{i,m^{\mathcal{S}}})\right)\right)$$

where the $P'_{i,j}$s are given as usual by the η-expansion of h_i. We can assume wlog that for all i, we do have $l_i \in \mathcal{L}^{\mathcal{S}}$, with possibly $\varphi_{l_i}^{\mathcal{S}} : (\alpha_1,\dots,\alpha_{m^{\mathcal{S}}}) \mapsto 0$.

Under strong enough constraints Y and δ we can both use Lemma 8.3 on the h_is with the arity $m^{\mathcal{S}}$, and use the induction hypothesis on the structures $\mathcal{S}_{l,j}$, with the terms $P'_{i,j}$ such that $l_i = l$ and with a precision $\frac{\epsilon}{2\deg(\varphi_l^{\mathcal{S}})}$. Then for all $\sigma \in \Sigma_{Y,\delta}$:

- the induction hypothesis gives terms $\vec{L}_{l,j}$ such that $\sigma(P'_{i,j})\ \vec{L}_{l_i,j}$ $\frac{\epsilon}{2\deg(\varphi_l^{\mathcal{S}})}$-approximates $\mathcal{E}^{\mathcal{S}_{l_i,j}}(PT_d^{\eta}(P'_{i,j}))$;
- Proposition 7.2 gives terms F_l which continuously represent the functions $\varphi_l^{\mathcal{S}}$;
- Lemma 8.3 gives a sequence of terms $\vec{L}$ such that
$$\sigma(h_i)\ \vec{L} \twoheadrightarrow G_{l_i}\ \sigma(P'_{i,1})\ \dots\ \sigma(P'_{i,m^{\mathcal{S}}})$$
with $G_l = \lambda z_1 \dots z_{m^{\mathcal{S}}}.F_l\ (z_1\ \vec{L}_{l,1})\ \dots\ (z_{m^{\mathcal{S}}}\ \vec{L}_{l,m_l^{\mathcal{S}}})$.

We get that $\sigma(h_i)\ \vec{L}$ ϵ-approximates $\mathcal{E}_{\mathrm{V}}^{\mathcal{S}}(VT_{d+1}^{\eta}(h_i))$.

Now given finitely many arbitrary terms $M_1,\dots,M_n$, for all i we have

$$\mathcal{E}^{\mathcal{S}}(PT_{d+1}^{\eta}(M_i)) = \sum_{t\in\mathcal{VT}_{d+1}^{\eta}} PT_{d+1}^{\eta}(M_i)(t) \times \mathcal{E}_{\mathrm{V}}^{\mathcal{S}}(t) = \sum_{h\in\mathcal{H}} \mathcal{P}\left(M_i \twoheadrightarrow h\right) \times \mathcal{E}_{\mathrm{V}}^{\mathcal{S}}(VT_{d+1}^{\eta}(h))$$

Let us pick finitely many head normal forms h_j such that for all i, $\mathcal{P}_{\Downarrow}(M_i) - \epsilon \leq \sum_j \mathcal{P}\left(M_i \twoheadrightarrow h_j\right) \leq \mathcal{P}_{\Downarrow}(M_i)$. We just proved there are Y and δ such that for all $\sigma \in \Sigma_{Y,\delta}$ we can find terms $\vec{L}$ such that for all j, $\sigma(h_j)\ \vec{L}$ ϵ-approximates $\mathcal{E}_{\mathrm{V}}^{\mathcal{S}}(VT_{d+1}^{\eta}(h_j))$. Then using Lemma 8.4 we have:

$$\begin{aligned}\mathcal{P}\left(\sigma(M_i)\ \vec{L} \twoheadrightarrow I\right) &= \sum_{h\in\mathcal{H}} \mathcal{P}\left(M_i \twoheadrightarrow h\right) \mathcal{P}\left(\sigma(h)\ \vec{L} \twoheadrightarrow I\right)\\ &\geq \sum_j \mathcal{P}\left(M_i \twoheadrightarrow h_j\right) \mathcal{P}\left(\sigma(h_j)\ \vec{L} \twoheadrightarrow I\right)\\ &\geq \sum_j \mathcal{P}\left(M_i \twoheadrightarrow h_j\right) \mathcal{E}_{\mathrm{V}}^{\mathcal{S}}(VT_{d+1}^{\eta}(h_j)) - \epsilon\\ &\geq \mathcal{E}^{\mathcal{S}}(PT_{d+1}^{\eta}(M_i)) - 2\epsilon\end{aligned}$$

and similarly $\mathcal{P}_{\Downarrow}(\sigma(M_i)\ \vec{L}) \leq \mathcal{E}^{\mathcal{S}}(PT_{d+1}^{\eta}(M_i)) + 2\epsilon$. Hence for all i, the term $\sigma(M_i)\ \vec{L}$ 2ϵ-approximates $\mathcal{E}^{\mathcal{S}}(PT_{d+1}^{\eta}(M_i))$.

This concludes the proof of Proposition 8.2.

9 Tree Separation

Now that we have a notion of evaluation procedure on Nakajima trees which we know we can simulate on terms, we can put terms aside and work solely with trees. We want to prove the following proposition.

Proposition 9.1. *Given finitely many trees $T_1,\dots,T_n$ there is an evaluation structure $\mathcal{S}$ such that for all i and i',*

$$\mathcal{E}^{\mathcal{S}}(T_i) = \mathcal{E}^{\mathcal{S}}(T_{i'}) \text{ if and only if } T_i = T_{i'}$$

We prove this by induction on the depth of the trees. Let $T_1,\dots,T_n$ in $\mathcal{PT}_{d+1}^{\eta}$, which wlog we assume pairwise distinct. Our goal is to build an evaluation structure:

$$\mathcal{S} = \left(\mathcal{L}^{\mathcal{S}}, m^{\mathcal{S}}, \left(\varphi_l^{\mathcal{S}}\right)_{l\in\mathcal{L}^{\mathcal{S}}}, (\mathcal{S}_{l,i})_{l\in\mathcal{L}^{\mathcal{S}},1\leq i\leq m^{\mathcal{S}}}\right)$$

We will split the evaluation function $\varphi_l^{\mathcal{S}}$ into three functions: we will have $\varphi_l^{\mathcal{S}} = \varphi_0 \circ \varphi_l \circ \varphi_+$ with

- $\varphi_0 : [0,1]^{m^{\mathcal{S}}} \to [0,1]$;
- $\varphi_l : [0,1] \to [0,1]$ for all $l \in \mathcal{L}^{\mathcal{S}}$;
- $\varphi_+ : [0,1] \to [0,1]$.

Given any tree $T \in \mathcal{PT}^{\eta}_{d+1}$ we isolate the role of each element of $\mathcal{S}$ in the evaluation of T using the following subprobability distributions:

$$\widehat{T}(l, U_1, \dots, U_{m^{\mathcal{S}}}) = \sum_{(U_j)_{j>m^{\mathcal{S}}}} T(l, (U_j)_{j\in\mathbb{N}^*})$$

$$\widetilde{T}(l, \alpha_1, \dots, \alpha_{m^{\mathcal{S}}}) = \sum_{\substack{U_1,\dots,U_{m^{\mathcal{S}}} \text{ s.t.} \\ \forall j, \mathcal{E}^{\mathcal{S}_{l,j}}(U_j)=\alpha_j}} \widehat{T}(l, U_1, \dots, U_{m^{\mathcal{S}}})$$

$$\overline{T}^1(l, \beta) = \sum_{\substack{\vec{\alpha} \text{ s.t.} \\ \varphi_0(\vec{\alpha})=\beta}} \widetilde{T}(l, \vec{\alpha})$$

$$\overline{T}^2(\gamma) = \sum_{l\in\mathcal{L}^{\mathcal{S}}} \sum_{\substack{\beta \text{ s.t.} \\ \varphi_l(\beta)=\gamma}} \overline{T}^1(l, \beta)$$

where l ranges over $\mathcal{L}^{\mathcal{S}}$, the U_js over $\mathcal{PT}^{\eta}_{d}$ and the α_js, β and γ over $[0,1]$.

At the end we get:

$$\mathcal{E}^{\mathcal{S}}(T) = \sum_{\gamma} \overline{T}^2(\gamma)\varphi_+(\gamma)$$

We want to choose $\mathcal{L}^{\mathcal{S}}$, $m^{\mathcal{S}}$, the structures $\mathcal{S}_{l,j}$ and the functions φ_0, φ_l and φ_+ to keep the $\widehat{T_i}$, $\widetilde{T_i}$, $\overline{T_i}^1$, $\overline{T_i}^2$ and ultimately the $\mathcal{E}^{\mathcal{S}}(T_i)$ pairwise distinct. We know there are some trees $t_k = (l_k, (U_{k,j})_{j\in\mathbb{N}^*}) \in \mathcal{VT}^{\eta}_{d+1}$ such that for all $i \neq i'$ we have $T_i(t_k) \neq T_{i'}(t_k)$ for some k, and we want to preserve this difference. Let us define:

$$\begin{aligned}
\widehat{t_k} &= (l_k, U_{k,1}, \dots, U_{k,m^{\mathcal{S}}}) \\
\widetilde{t_k} &= (l_k, \mathcal{E}^{\mathcal{S}_{l_k,1}}(U_{k,1}), \dots, \mathcal{E}^{\mathcal{S}_{l_k,m^{\mathcal{S}}}}(U_{k,m^{\mathcal{S}}})) \\
\overline{t_k}^1 &= (l, \varphi_0(\mathcal{E}^{\mathcal{S}_{l_k,1}}(U_{k,1}), \dots, \mathcal{E}^{\mathcal{S}_{l_k,m^{\mathcal{S}}}}(U_{k,m^{\mathcal{S}}}))) \\
\overline{t_k}^2 &= \varphi_l(\varphi_0(\mathcal{E}^{\mathcal{S}_{l_k,1}}(U_{k,1}), \dots, \mathcal{E}^{\mathcal{S}_{l_k,m^{\mathcal{S}}}}(U_{k,m^{\mathcal{S}}})))
\end{aligned}$$

For all $i \neq i'$ we want to have some k such that $\widehat{T_i}(\widehat{t_k}) \neq \widehat{T_{i'}}(\widehat{t_k})$, $\widetilde{T_i}(\widetilde{t_k}) \neq \widetilde{T_{i'}}(\widetilde{t_k})$, $\overline{T_i}^1(\overline{t_k}^1) \neq \overline{T_{i'}}^1(\overline{t_k}^1)$ and $\overline{T_i}^2(\overline{t_k}^2) \neq \overline{T_{i'}}^2(\overline{t_k}^2)$.

The labels $\mathcal{L}^{\mathcal{S}}$. We simply choose $\mathcal{L}^{\mathcal{S}} = \cup_k\{l_k\}$.

The arity $m^{\mathcal{S}}$. For any $T \in \mathcal{VT}^{\eta}_{d+1}$ and $t = (l, (U_j)) \in \mathcal{VT}^{\eta}_{d+1}$ we have:

$$T(t) = \lim_{m\to\infty} \sum_{(U'_j)_{j>m}} T(l, (U_1, \dots, U_m, U'_{m+1}, \dots))$$

So if we choose $m^{\mathcal{S}}$ large enough then we have $\widehat{T_i}(\widehat{t_k}) \neq \widehat{T_{i'}}(\widehat{t_k})$ whenever $T_i(t_k) \neq T_{i'}(t_k)$.

The structures $\mathcal{S}_{l,j}$. We can find finite sets $\mathcal{U}_j \subset_f \mathcal{VT}^{\eta}_{d+1}$ for $j \leq m^{\mathcal{S}}$ such that:

- $U_{k,j} \in \mathcal{U}_j$ for all j and k;
- and for all $i \neq i'$, if $\widehat{T_i}(\widehat{t_k}) \neq \widehat{T_{i'}}(\widehat{t_k})$ then:

$$\sum_{\substack{U_1,\dots,U_{m^{\mathcal{S}}} \text{ s.t.} \\ \exists j, U_j \notin \mathcal{U}_j}} \widehat{T_i}(l_k, U_1, \dots, U_{m^{\mathcal{S}}}) < |\widehat{T_i}(\widehat{t_k}) - \widehat{T_{i'}}(\widehat{t_k})|$$

Then for each $l \in \mathcal{L}^{\mathcal{S}}$ and $j \leq m^{\mathcal{S}}$ the induction hypothesis applied to the finite set of trees $\mathcal{U}_j$ gives a structure $\mathcal{S}_{l,j}$. For all k, for all $l \in \mathcal{L}^{\mathcal{S}}$ and all $U_1, \dots, U_{m^{\mathcal{S}}} \in \mathcal{VT}^{\eta}_{d+1}$ we may have $\widetilde{t_k} = (l, \mathcal{E}^{\mathcal{S}_{l,1}}(U_1), \dots, \mathcal{E}^{\mathcal{S}_{l,m^{\mathcal{S}}}}(U_{m^{\mathcal{S}}}))$ only if $(U_1, \dots, U_{m^{\mathcal{S}}}) \notin \mathcal{U}_1 \times \cdots \times \mathcal{U}_{m^{\mathcal{S}}}$. Hence if $\widehat{T_i}(\widehat{t_k}) \neq \widehat{T_{i'}}(\widehat{t_k})$ then $\widetilde{T_i}(\widetilde{t_k}) \neq \widetilde{T_{i'}}(\widetilde{t_k})$.

The function φ_0. To define φ_0 and keep the $\overline{T_i}^1$s pairwise distinct we proceed in the same way as for the $\mathcal{S}_{l,j}$s. We have the following lemma, somehow similar to Proposition 9.1.

Lemma 9.2. *Given any finite set $\mathcal{A} \subset_f [0,1]^m$, there exists an evaluation function $\varphi \in \mathcal{F}_m$ such that for all $\vec{\alpha} \neq \vec{\alpha'}$ in $\mathcal{A}$, $\varphi(\vec{\alpha}) \neq \varphi(\vec{\alpha'})$.*

Proof. By induction on m. For $m = 0$ $[0,1]^0$ has at most one element. Otherwise given $\mathcal{A} \subset_f [0,1]^{m+1}$, using the induction hypothesis we build $\varphi \in \mathcal{F}_{m+1}$ such that, for all $\vec{\alpha} = (\alpha_1, \dots, \alpha_{m+1}) \neq (\alpha'_1, \dots, \alpha'_{m+1}) = \vec{\alpha'}$ in $\mathcal{A}$, if $(\alpha_1, \dots, \alpha_m) \neq (\alpha'_1, \dots, \alpha'_m)$ then $\varphi(\vec{\alpha}) \neq \varphi(\vec{\alpha'})$. Let $\pi_{m+1} : [0,1]^{m+1} \to [0,1]$ be the $m+1$-th projection. Then for all $\vec{\alpha} \neq \vec{\alpha'} \in \mathcal{A}$ we have either $\varphi(\vec{\alpha}) \neq \varphi(\vec{\alpha'})$ or $\pi_{m+1}(\vec{\alpha}) \neq \pi_{m+1}(\vec{\alpha'})$. Let δ be the minimum of the $|\varphi(\vec{\alpha}) - \varphi(\vec{\alpha'})| > 0$, the function $\varphi' : \vec{\alpha} \mapsto \left(1 - \frac{\delta}{2}\right)\varphi(\vec{\alpha}) + \frac{\delta}{2}\pi_{m+1}(\vec{\alpha})$ has the wanted property. ◻

Now we can find a finite set $\mathcal{A} \subset_f [0,1]^{m^{\mathcal{S}}}$ such that

- $\left(\mathcal{E}^{\mathcal{S}_{l_k,1}}(U_{k,1}), \dots, \mathcal{E}^{\mathcal{S}_{l_k,m^{\mathcal{S}}}}(U_{k,m^{\mathcal{S}}})\right) \in \mathcal{A}$ for all k;
- and for all $i \neq i'$, if $\widetilde{T_i}(\widetilde{t_k}) \neq \widetilde{T_{i'}}(\widetilde{t_k})$ then:

$$\sum_{(\alpha_1,\dots,\alpha_{m^{\mathcal{S}}})\notin\mathcal{A}} \widetilde{T_i}(l_k, \alpha_1, \dots, \alpha_{m^{\mathcal{S}}}) < |\widetilde{T_i}(\widetilde{t_k}) - \widetilde{T_{i'}}(\widetilde{t_k})|$$

Then we can use the previous lemma on $\mathcal{A}$ to get φ_0 such that the $\overline{T_i}^1$s are pairwise distinct.

The functions φ_l. We write $\mathcal{L}^{\mathcal{S}} = \{l^0, \dots, l^{p-1}\}$ where the l^rs are pairwise distinct (which is not necessarily the case for the l_rs). Then we define $\varphi_{l^r} : \beta \mapsto \frac{r}{p} + \frac{1}{2p}\beta$, so that for all $l, l' \in \mathcal{L}^{\mathcal{S}}$ and $\beta, \beta' \in [0,1]$ we have $\varphi_l(\beta) = \varphi_{l'}(\beta')$ iff $l = l'$ and $\beta = \beta'$. Thus in particular for all i and k we have $\overline{T_i}^2(\overline{t_k}^2) = \overline{T_i}^1(\overline{t_k}^1)$.

The function φ_+. The key observation here is that any subprobability distribution τ over $[0,1]$ is such that its moments $\sum_{\gamma\in[0,1]} \tau(\gamma)\gamma^r$ for $r \in \mathbb{N}$ are all in $[0,1]$: according to [2], Theorem 30.1, such distributions are entirely characterised by their moments. In other words, as for all $i \neq i'$ we know that $\overline{T_i}^2 \neq \overline{T_{i'}}^2$, then there is necessarily some $r_{i,i'} \in \mathbb{N}$ such that $\sum_{\gamma \in [0,1]} \overline{T_i}^2(\gamma)\gamma^{r_{i,i'}} \neq \sum_{\gamma \in [0,1]} \overline{T_{i'}}^2(\gamma)\gamma^{r_{i,i'}}$.

Using the same technique as in the proof of Lemma 9.2 we can build $\varphi_+ \in \mathcal{F}_1$ as a linear combination of the functions $\gamma \mapsto \gamma^{r_{i,i'}}$ for $i \neq i'$ in such a way that the $\sum_\gamma \overline{T}^2(\gamma)\varphi_+(\gamma)$ are pairwise distinct.

This concludes the proof of Proposition 9.1.

10 Main result

The results proven in the Sections 8 and 9 immediately give us our main theorem.

Theorem 10.1. *For all $M, N \in \Lambda_+$:*

$$M =_{obs} N \Rightarrow M =_{\mathcal{PB}^\eta} N$$

Proof. If $M \neq_{\mathcal{PB}^\eta} N$ then $PT_d^\eta(M) \neq PT_d^\eta(N)$ for some $d \in \mathbb{N}$. According to Proposition 9.1 there is an evaluation structure $\mathcal{S}$ of depth d such that $\mathcal{E}^\mathcal{S}(PT_d^\eta(M)) \neq \mathcal{E}^\mathcal{S}(PT_d^\eta(N))$. Then let $\epsilon > 0$ be such that $\epsilon < \frac{1}{2}|\mathcal{E}^\mathcal{S}(PT_d^\eta(M)) - \mathcal{E}^\mathcal{S}(PT_d^\eta(N))|$, Proposition 8.1 gives a context C such that $|\mathcal{P}_\Downarrow(C[M]) - \mathcal{E}^\mathcal{S}(PT_d^\eta(M))| < \epsilon$ and $|\mathcal{P}_\Downarrow(C[N]) - \mathcal{E}^\mathcal{S}(PT_d^\eta(N))| < \epsilon$, hence $\mathcal{P}_\Downarrow(C[M]) \neq \mathcal{P}_\Downarrow(C[N])$ and $M \neq_{obs} N$. □

When we defined the observational equivalence we arbitrarily chose to consider the convergence probability as the only observational behaviour, and this theorem only holds for this particular notion of observation. But we can actually prove a more complete result if we make full use of Proposition 8.1 and the precise definition of ϵ-approximation of $\underline{\alpha}$.

Theorem 10.2. *Given any set $H \subset \mathcal{H}$ of head normal forms, for all $M, N \in \Lambda_+$, if for every context C:*

$$\sum_{h \in H} \mathcal{P}\left(C[M] \twoheadrightarrow h\right) = \sum_{h \in H} \mathcal{P}\left(C[N] \twoheadrightarrow h\right)$$

then $M =_{\mathcal{PB}^\eta} N$.

Proof. If $M \neq_{\mathcal{PB}^\eta} N$ then Propositions 9.1 and 8.1 give a context C such that for instance $\mathcal{P}_\Downarrow(C[M]) < \mathcal{P}\left(C[N] \twoheadrightarrow I\right)$ (where $I = \lambda x.x$). Then let $h_0 \in H$, from Lemma 8.4 we can derive $\mathcal{P}_\Downarrow(C[M]\, h_0) \leq \mathcal{P}_\Downarrow(C[M])$, hence:

$$\sum_{h \in H} \mathcal{P}\left(C[M]\, h_0 \twoheadrightarrow h\right) \leq \mathcal{P}_\Downarrow(C[M]\, h_0) \leq \mathcal{P}_\Downarrow(C[M])$$

and $\mathcal{P}\left(C[N] \twoheadrightarrow I\right) \leq \mathcal{P}\left(C[N]\, h_0 \twoheadrightarrow h_0\right)$, hence:

$$\mathcal{P}\left(C[N] \twoheadrightarrow I\right) \leq \mathcal{P}\left(C[N]\, h_0 \twoheadrightarrow h_0\right) \leq \sum_{h \in H} \mathcal{P}\left(C[N]\, h_0 \twoheadrightarrow h\right)$$

So $\sum_{h \in H} \mathcal{P}\left(C[M]\, h_0 \twoheadrightarrow h\right) \neq \sum_{h \in H} \mathcal{P}\left(C[N]\, h_0 \twoheadrightarrow h\right)$. □

11 Conclusion

To characterise the observational equivalence in the call-by-name probabilistic λ-calculus, we introduced a notion of probabilistic Böhm trees, which we then refined to define their infinitely extensional variant, the probabilistic Nakajima trees. We proved that observationally equivalent probabilistic terms have the same Nakajima trees.

This separation result is a first step in the generalisation of the Hyland and Wadsworth theorem. To prove that the probabilistic observational equivalence and the probabilistic Nakajima tree equality actually coincide, we can observe that given two terms M and N, if $M =_{\mathcal{PB}^\eta} N$ then in particular $PT_1^\eta(M) = PT_1^\eta(N)$, and:

$$\mathcal{P}_\Downarrow(M) = \sum_{t \in \mathcal{VT}_1^\eta} PT_1^\eta(M)(t) = \sum_{t \in \mathcal{VT}_1^\eta} PT_1^\eta(N)(t) = \mathcal{P}_\Downarrow(N)$$

So all we need to prove is that if $M =_{\mathcal{PB}^\eta} N$ then $C[M] =_{\mathcal{PB}^\eta} C[N]$ for all C. This is not a result one could expect in a calculus where the reduction strategy is so important. This is even more true if we look at the last part of the Hyland and Wadsworth theorem: a λ-theory is an equivalence $\simeq$ on terms which respects the reduction rules (if $M \to N$ then $M \simeq N$) and which is a congruence (if $M \simeq N$ then $C[M] \simeq C[N]$ for all C), so the very notion of theory implies that we allow reduction under arbitrary contexts. To tackle this problem we defined a deterministic and contextual operational semantics for the probabilistic λ-calculus [8], in which we proved that the Hyland and Wadsworth theorem does generalise to a probabilistic setting.

References

[1] H. P. Barendregt, *The lambda calculus. Its syntax and semantics* Studies in logic and foundations of mathematics, vol. 103. North-Holland Publishing Company, Amsterdam, New York, and Oxford, 1981, xiv + 615 pp.

[2] P. Billingsley, *Probability and Measure*, third edition Wiley Series in Probability and Mathematical Statistics, 1995.

[3] R. Crubillé and U. Dal Lago, *On Probabilistic Applicative Bisimulation and Call-by-Value Lambda-Calculi* Programming Languages and Systems, 23rd European Symposium on Programming, Proceedings, volume 8410 of Lecture Notes in Computer Science, pages 209-228. Springer, 2014.

[4] U. de'Liguoro and A. Piperno, *Nondeterministic Extensions of Untyped λ-Calculus* Information and Computation, 1995.

[5] M. Dezani-Ciancaglini and E. Giovannett, *From Böhm's Theorem to Observational Equivalences: an Informal Account* Electronic Notes in Theoretical Computer Science, 2001.

[6] T. Ehrhard, M. Pagani and C. Tasson, *The Computational Meaning of Probabilistic Coherence Spaces* Logic in Computer Science (LICS), Jun 2011, Toronto, ON, Canada. pp.87 - 96, 2011.

[7] M. Hyland, *A syntactic characterisation of the Equality in some Models of the Lambda Calculus* Journal of the London Mathematical Society, 12, 1976, 361-370.

[8] T. Leventis, *Probabilistic λ-Theories* Aix-Marseille Université D.Phil. Thesis, 2016.

[9] C. P. Wadsworth, *The relation between computational and denotational properties for Scott's d_∞-models of the lambda-calculus* SIAM Journal on Computing, 5(3):488–521, 1976.