On the Taylor Expansion of Probabilistic λ-terms

Ugo Dal Lago
University of Bologna, Italy
INRIA Sophia Antipolis, France
ugo.dallago@unibo.it

Thomas Leventis
INRIA Sophia Antipolis, France
thomas.leventis@ens-lyon.org

Abstract

We generalise Ehrhard and Regnier's Taylor expansion from *pure* to *probabilistic* λ-terms. We prove that the Taylor expansion is adequate when seen as a way to give semantics to probabilistic λ-terms, and that there is a precise correspondence with probabilistic Böhm trees, as introduced by the second author. We prove this adequacy through notions of probabilistic resource terms and explicit Taylor expansion.

2012 ACM Subject Classification Theory of computation → Lambda calculus; Theory of computation → Equational logic and rewriting

Keywords and phrases Probabilistic Lambda-Calculi, Taylor Expansion, Linear Logic

Digital Object Identifier 10.4230/LIPIcs.FSCD.2019.13

Related Version A full version of the paper is available at `http://arxiv.org/abs/1904.09650`.

Funding The authors are partially supported by the ERC Consolidator Grant DLV-818616 DIAPASoN, as well as by the ANR projects 14CE250005 ELICA and 16CE250011 REPAS.

1 Introduction

Linear logic is a proof-theoretical framework which, since its inception [10], has been built around an analogy between on the one hand linearity in the sense of linear algebra, and on the other hand the absence of copying and erasing in cut elimination and higher-order rewriting. This analogy has been pushed forward by Ehrhard and Regnier, who introduced a series of logical and computational frameworks accounting, along the same analogy, for concepts like that of a differential, or the very related one of an approximation. We are implicitly referring to differential λ-calculus [6], to differential linear logic [8], and to the Taylor expansion of ordinary λ-terms [9]. The latter has given rise to an extremely interesting research line, with many deep contributions in the last ten years. Not only the Taylor expansion of pure λ-terms has been shown to be endowed with a well-behaved notion of reduction, but the Böhm tree and Taylor expansion operators are now known to commute [7]. This easily implies that the equational theory (on pure λ-terms) induced by the Taylor expansion coincides with the one induced by Böhm trees.

The Taylor expansion operator is essentially *quantitative*, in that its codomain is not merely the set of resource terms [3, 6], a term syntax for promotion-free differential proofs, but the set of *linear combinations* of those terms, with positive real number coefficients. When enlarging the domain of the operator to account for a more quantitative language, one is naturally lead to consider algebraic λ-calculi, to which giving a clean computational meaning has been proved hard so far [18].

But what about *probabilistic* λ-calculi [11], which have received quite some attention recently (see, e.g. [5, 2, 16]) due to their applicability to randomised computation and bayesian programming? Can the Taylor expansion naturally be generalised to those calculi? This is an interesting question, to which we give the first definite positive answer in this paper. In particular, we show that the Taylor expansion of probabilistic λ-terms is a conservative extension of the well-known one on ordinary λ-terms. In particular, the target can be taken, as usual, as a linear combination of *ordinary* resource terms, i.e., the same kind of structure which Ehrhard and Regnier considered in their work on the Taylor expansion of *pure* λ-terms. We moreover show that the Taylor expansion, as extended to probabilistic λ-terms, continues to enjoy the nice properties it has in the deterministic realm. In particular, it is adequate as a way to give semantics to probabilistic λ-terms, and the equational theory on probabilistic λ-terms induced by Taylor expansion coincides with the one induced by a probabilistic variation on Böhm trees [1]. The latter, noticeably, has been proved to capture observational equivalence, one quotiented modulo η-equivalence [1].

Are we the first ones to embark on the challenge of generalising Taylor's expansion to probabilistic λ-calculi, and in general to effectful calculi? Actually, some steps in this direction have recently been taken. First of all, we need to mention the line of works originated by Tsukada and Ong's paper on rigid resource terms [14]. This has been claimed from the very beginning to be a way to model effects in the resource calculus, but it has also been applied to, among others, probabilistic effects, giving rise to quantitative denotational models [15]. The obtained models are based on species, and are proved to be adequate. The construction being generic, there is no aim at providing a precise comparison between the discriminating power of the obtained theory and, say, observational equivalence: the choice of the underlying effect can in principle have a huge impact on it.

One should also mention Vaux's work on the algebraic λ-calculus [18], where one can build arbitrary linear combinations of terms. He showed a correspondence between Taylor expansion and Böhm trees, but only for terms whose Böhm trees approximants at finite depths are computable in a finite number of steps. This includes all ordinary λ-terms but not all probabilistic ones. More recently Olimpieri and Vaux have studied a Taylor expansion for a non-deterministic λ-calculus [19] corresponding to our notion of *explicit* Taylor expansion (Section 3).

The Probabilistic Taylor Expansion, Informally

The main idea behind building the Taylor expansion of any λ-term M is to describe the dynamics of M by way of *linear approximations* of M. In the realm of the λ-calculus, a linear approximation has traditionally been taken as a *resource term*:

$$s, t \in \Delta := x \mid \lambda x.s \mid \langle s\rangle\ \bar{t} \qquad\qquad \bar{s}, \bar{t} \in \Delta^! := [s_1, \ldots, s_n].$$

A resource term can be seen as a pure λ-term in which applications have the form $\langle s\rangle\ \bar{t}$, where s is a term and $\bar{t}$ is a *multiset* of terms, and in which the result of firing the redex $\langle \lambda x.s\rangle\ \bar{t}$ is the linear combination of all the terms obtained by allocating the resources in $\bar{t}$ to the occurrences of x in s. For instance, one such element in the Taylor expansion of δ is $\lambda x.(\langle x\rangle\ [x])$, where the occurrence of x in head position is provided with only one copy of its argument. If applied to the multiset $[y, z]$, this term would reduce into $\langle y\rangle\ [z] + \langle z\rangle\ [y]$. Similarly, an element in the Taylor expansion of $\delta\ I$ would be $\langle \lambda x.\langle x\rangle\ [x]\rangle\ [I^2]$, which reduces into $2.\langle I\rangle\ [I]$. Another element of the same Taylor expansion is $\langle \lambda x.\langle x\rangle\ [x]\rangle\ [I^3]$, but this one reduces into 0: there is no way to use its resources linearly, i.e., using them without copying and erasing.

The actual Taylor expansion of a term is built by translating any application $M\ N$ into an infinite sum

$$(M\ N)^* = \sum_{n\in\mathbb{N}} \frac{1}{n!}.\langle M^*\rangle\ [(N^*)^n]$$

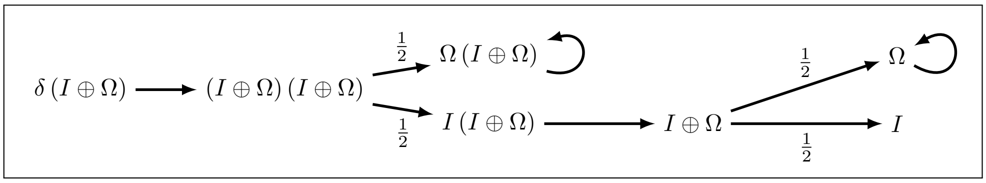

Figure 1 M's Reduction Tree.

and $x^* = x$, $(\lambda x.M)^* = \lambda x.M^*$. Remark that M^* and N^* are linear combinations, but constructors of the resource calculus are multilinear. For instance $(\lambda x.M)^* = \sum_{s\in\Delta} M^*_s.\lambda x.s$. As an example the Taylor expansion of $\delta\ I$ is $\sum_{m,n\in\mathbb{N}} \frac{1}{m!n!}.\langle\lambda x.\langle x\rangle\ [x^m]\rangle\ [I^n]$. Remark that any summand properly reduces only when $n = m+1$, in which case it reduces to $n!.\langle I\rangle\ [I^m]$. In turn $\langle I\rangle\ [I^m]$ reduces properly only when $m = 1$, and the result is I. All the other terms reduce to 0. In the end the Taylor expansion of $\delta\ I$ normalises to $\frac{2!}{1!2!}.I = I$, which is exactly the Taylor expansion of the normal form of $\delta\ I$. More generally every Taylor expansion is normalisable, and the normal form of the Taylor expansion of a term corresponds to the Taylor expansion of its Böhm tree, hence (normal forms of) Taylor expansions yields an interesting model of deterministic λ-calculus.

Now let us consider the probabilistic λ-term $M = \delta(I \oplus \Omega)$, where $\oplus$ is an operator for binary, fair, probabilistic choice, $\delta = \lambda x.xx$, $I = \lambda.x.x$ and $\Omega = \delta\delta$ is a purely diverging, term. As such, M is a term of a minimal, untyped, probabilistic λ-calculus. Evaluation of M is performed leftmost-outermost is as in Figure 1. In particular, the probability of convergence for M is $\frac{1}{4}$. Please observe that two copies of the argument $I \oplus \Omega$ are produced, and that the "rightmost" one is evaluated only when the "leftmost" one converges, i.e. when the probabilistic choice $I \oplus \Omega$ produces I as a result.

Extending the Taylor expansion to probabilistic terms seems straightforward, a natural candidate for the Taylor expansion of $M \oplus N$ being just $\frac{1}{2}.M^* + \frac{1}{2}.N^*$. When computing the Taylor expansion of M we will find expressions such as $\langle\lambda x.\langle x\rangle\ [x]\rangle\ [(\frac{1}{2}.I + \frac{1}{2}.\Omega^*)^2]$, i.e. $\frac{1}{4}.\langle\lambda x.\langle x\rangle\ [x]\rangle\ [I^2] + \frac{1}{4}.\langle\lambda x.\langle x\rangle\ [x]\rangle\ [\Omega^2] + \frac{1}{2}.\langle\lambda x.\langle x\rangle\ [x]\rangle\ [I,\Omega]$. For non-trivial reasons, the Taylor expansion of any diverging term normalises to 0, so just like in our previous example, the only element in M^* which does not reduce to 0 is $\langle\lambda x.\langle x\rangle\ [x]\rangle\ [I^2]$. The difference is that this time it appears with a coefficient $\frac{1}{1!2!}\frac{1}{4}$, so M^* normalises to $\frac{1}{4}.I$. Please notice how this is once again the "normal form" of the original term M.

The goal of this paper is to show that the correspondence between Taylor expansions and Böhm trees is preserved when we add probabilistic choices to the calculus. Unfortunately although the final result is the same as in the deterministic setting, the known proof techniques fail. To begin with not all (infinite) linear combinations of resource term are normalisable. Normal forms of deterministic Taylor expansions always exist because such expansions are *uniform*: all terms in their support have the same shape (the support of $(\lambda x.M)^*$ only contains abstractions, etc.) and this property is known to ensure normalisability. This property does not hold for *probabilistic* expansions: $(x \oplus \lambda y.y)^*$ contains both the variable x and the abstraction $\lambda y.y$. Thus we need to find a different way to prove that probabilistic Taylor expansions normalise. We proceed by using an intermediate notion of Taylor expansion *with explicit choices*, which enjoys uniformity and share other properties with deterministic Taylor expansion, and we then transpose directly the results on this intermediate construction to the "natural" probabilistic Taylor expansion.

Layout of the Paper

In Section 2 we introduce the resource calculus with explicit choices and show it enjoys the same properties as the usual deterministic resource calculus. In Section 3, we define the explicit Taylor expansion from probabilistic λ-terms. These constructions have an interest in themselves and they have been independently studied by Olimpieri and Vaux [19] for a non-deterministic calculus but in this paper they are just an intermediate step towards proving our main results. Definitionally, the crux of the paper is Section 4, in which the Taylor expansion of a probabilistic λ-term is made to produce *ordinary* resource terms. The relation between Böhm trees and Taylor expansions is investigated in Section 5 and Section 6.

Notations

We write $\mathbb{N}$ for the set of natural numbers and $\mathbb{R}^+$ for the set of nonnegative real numbers. Given a set A, we write $\mathbb{R}^+\langle A\rangle$ for the set of families of positive real numbers indexed by elements in A. We write such families as linear combinations: an element $S \in \mathbb{R}^+\langle A\rangle$ is a sum $S = \sum_{a\in A} S_a.a$, with $S_a \in \mathbb{R}^+$. The support of a family $S \in \mathbb{R}^+\langle A\rangle$ is $\mathrm{supp}(S) = \{a \in A \mid S_a > 0\}$. We write $\mathbb{R}^+[A]$ for those families $S \in \mathbb{R}^+\langle A\rangle$ such that $\mathrm{supp}(S)$ is finite. Given $a \in A$ we often write a for $1.a \in \mathbb{R}^+\langle A\rangle$ unless we want to emphasise the difference between the two expressions. We also define *finite multisets* over A as functions $m : A \to \mathbb{N}$ such that $m(a) \neq 0$ for finitely many $a \in A$. We use the notation $[a_1, \dots, a_n]$ to describe the multiset m such that $m(a)$ is the number of indices $i \leq n$ such that $a_i = a$.

2 Probabilistic Resource Calculus

In this section, we describe the theory of resource terms with explicit choices, for the purpose of extending many of the properties of resource terms to the probabilistic case. All this has an interest in itself, but here this is mainly useful as a way to render certain proofs about the Taylor Expansion easier (see Section 3 for more details). For this reason we try to give the reader a clear understanding of this calculus and of why these definitions and properties are useful, without focusing on the actual proofs. These are straightforward generalisations of those for deterministic resource terms [9] and can be found in an extended version of this paper [4]. The same results have recently been given for a non-deterministic calculus [19] by Olimpieri and Vaux.

2.1 The Basics

▶ **Definition 1.** *The sets of* ***probabilistic simple resource terms*** $\Delta_\oplus$ *and of* ***probabilistic simple resource poly-terms*** $\Delta^!_\oplus$ *over a set of variables* $\mathcal{V}$ *are defined by mutual induction as follows:*

$$s, t \in \Delta_\oplus := x \mid \lambda x.s \mid \langle s\rangle\ \bar{t} \mid s \oplus_p \bullet \mid \bullet \oplus_p s \qquad\qquad \bar{s}, \bar{t} \in \Delta^!_\oplus := [s_1, \dots, s_n]$$

where p ranges over $[0, 1]$. We call ***finite probabilistic resource terms*** *the finite linear combinations of resource terms in $\mathbb{R}^+[\Delta_\oplus]$, and* ***finite probabilistic resource poly-terms*** *the finite linear combinations of resource poly-terms in $\mathbb{R}^+[\Delta^!_\oplus]$. We extend the constructors of simple (poly-)terms to (poly-)terms by linearity, e.g., if $S \in \mathbb{R}^+[\Delta_\oplus]$ then $\lambda x.S$ is defined as the poly-term such that $(\lambda x.S)_{\lambda x.s} = S_s$ and $(\lambda x.S)_t = 0$ if t is not an abstraction.*

Some consecutive abstractions $\lambda x_1. \dots \lambda x_n.s$ will be indicated as $\lambda x_1 \dots x_n.s$, or even as $\lambda\vec{x}.s$. Similarly, to describe many successive applications $\langle\langle\langle M\rangle\ N_1\rangle\ \dots\rangle\ N_k$, we use a single pair of brackets and we write $\langle M\rangle\ N_1\ \dots\ N_k$.

We write $\Delta_\oplus^{(!)}$ for $\Delta_\oplus \cup \Delta_\oplus^!$, which is ranged over by metavariables like σ, τ. Note that intuitively $\Delta_\oplus^{(!)}$ should stand for either $\Delta_\oplus$ or $\Delta_\oplus^!$, not their union. For instance we will prove some properties for finite linear combinations in $\mathbb{R}^+[\Delta_\oplus^{(!)}]$, but the only relevant linear combinations are the actual (poly-)terms in $\mathbb{R}^+[\Delta_\oplus]$ or $\mathbb{R}^+[\Delta_\oplus^!]$. Yet this distinction is technically irrelevant, and all our results hold if we define $\Delta_\oplus^{(!)}$ as a union.

The reason why linear combinations over such elements are dubbed *terms* will be clear once we describe the operational semantics of the resource calculus. The main point of the resource calculus is to allow functions to use their argument arbitrarily many times and yet remain entirely linear, which is achieved by taking multisets as arguments: if a function uses its argument n times then it needs to receive n resources as argument and use each of them linearly. This idea has two consequences. First, an application can fail if a function is not given exactly as many arguments as it needs, as it would need either to duplicate or to discard some of them. Second, the result of a valid application is often not unique: a function can choose how to allocate the different resources to the different calls to its argument, and different choices may lead to different results. Both these features are treated using linear combinations: a failed application results in 0 (i.e. the trivial linear combination) and a successful one yields the sum of all its possible outcomes.

▶ **Definition 2.** *We define the substitution of $\bar{t} \in \Delta_\oplus^!$ for $x \in \mathcal{V}$ in $s \in \Delta_\oplus$ by:*

$$\delta_x s \cdot [t_1, \ldots, t_n] = \begin{cases} 0 \text{ if } s \text{ does not have exactly } n \text{ free occurences of } x \\ \sum_{\rho \in \mathfrak{S}_n} s[t_{\rho(1)}/x_1, \ldots, t_{\rho(n)}/x_n] \in \mathbb{R}^+[\Delta_\oplus^{(!)}] \text{ otherwise} \end{cases}$$

where $x_1, \ldots, x_n$ are the free occurrences of x in s and $\mathfrak{S}_n$ is the set of permutations over $\{1, \ldots, n\}$.

▶ **Example 3.** A basic example is $\delta_x(\langle x \rangle\ [x]) \cdot [y, z] = \langle y \rangle\ [z] + \langle z \rangle\ [y]$: there are two occurrences of x in $\langle x \rangle\ [x]$, so there are two ways to substitute $[y, z]$ for them. Remark that we also have $\delta_x[x, x] \cdot [y, z] = [y, z] + [z, y] = 2.[y, z]$: the two occurrences of x are not as clearly distinguished as in the first example but they still count as different occurrences. Similarly $\delta_x(\langle x \rangle\ [x]) \cdot [y, y] = 2.\langle y \rangle\ [y]$ and $\delta_x[x, x] \cdot [y, y] = 2.[y, y]$: there are two distinct occurrences of y, so there are two ways to allocate them. As another example, please consider $\delta_x(\lambda x.x) \cdot [y] = \delta_x(\langle x \rangle\ [x]) \cdot [y] = 0$: the substitution fails if the number of resources does not match the number of free occurrences of the substituted variable.

The operational semantics of the deterministic resource calculus [9] is usually given as a single rule of β-reduction. In the probabilistic setting, we also need rules to make choices commute with head contexts.

▶ **Definition 4.** *The reductions $\to_\beta$ and $\to_\oplus$ are defined from $\Delta_\oplus^{(!)}$ to $\mathbb{R}^+[\Delta_\oplus^{(!)}]$ by:*

$$\langle \lambda x.s \rangle\ \bar{t} \to_\beta \delta_x s \cdot \bar{t}$$

$$\lambda x.(s \oplus_p \bullet) \to_\oplus \lambda x.s \oplus_p \bullet \qquad \lambda x.(\bullet \oplus_p s) \to_\oplus \bullet \oplus_p \lambda x.s$$

$$\langle s \oplus_p \bullet \rangle\ \bar{t} \to_\oplus \langle s \rangle\ \bar{t} \oplus_p \bullet \qquad \langle \bullet \oplus_p s \rangle\ \bar{t} \to_\oplus \bullet \oplus_p \langle s \rangle\ \bar{t}$$

extended under arbitrary contexts. We simply write $\to$ for $\to_\beta \cup \to_\oplus$. Reduction can be extended to finite terms in the following way: if $S \in \mathbb{R}^+[\Delta_\oplus^{(!)}]$, $S_\sigma > 0$ and $\sigma \to T$ then $S \to S - S_\sigma.\sigma + S_\sigma T$.

As the resource calculus does not allow any duplication, and β-reduction erases some constructors, it naturally decreases the size of the involved simple terms. Consequently, β-reduction is strongly normalising. This result can be extended to the whole reduction $\to$, which is also confluent.

▶ **Proposition 5.** *The reduction* $\to$ *is confluent and strongly normalising on* $\mathbb{R}^+[\Delta_\oplus^{(!)}]$*. Given* $S \in \mathbb{R}^+[\Delta_\oplus^{(!)}]$ *we write* $\mathrm{nf}(S)$ *for its unique normal form for* $\to$*, and given* $\sigma \in \Delta_\oplus^{(!)}$ *we write* $\mathrm{nf}(\sigma)$ *for* $\mathrm{nf}(1.\sigma)$*.*

2.2 Infinite Terms

So far we only worked with finite terms but to fully express the operational behaviour of a λ-term in the resource λ-calculus, which is the purpose of the Taylor expansion, we need infinite ones. We can extend the constructors of the calculus to $\mathbb{R}^+\langle\Delta_\oplus^{(!)}\rangle$ by linearity and generalise the reduction relation $\to$, but Proposition 5 fails. Indeed let $I_0 = I = \lambda x.x$ and $I_{n+1} = \langle I_n\rangle\ [I]$. For $n \in \mathbb{N}$, let $S = \sum_{n\in\mathbb{N}} I_n$. Then, for all $n \in \mathbb{N}$ the term I_n normalises in n steps and S *does not* normalise in a finite number of reduction steps. A simple solution to this problem is to define the "normal form" of an infinite term by normalising each of its components: we can set $\mathrm{nf}(S) = \sum_{\sigma\in\Delta_\oplus^{(!)}} S_\sigma \mathrm{nf}(\sigma)$. But then another problem arises. In our previous example, we have $\mathrm{nf}(I_n) = I$ for all $n \in \mathbb{N}$, thus we would have $\mathrm{nf}(S) = \sum_{n\in\mathbb{N}} I$, which is not an element of $\mathbb{R}^+\langle\Delta_\oplus^{(!)}\rangle$ as the coefficient of I is infinite. Still we can use this pointwise normalisation if we consider terms with a particular property, called *uniformity.*

▶ **Definition 6.** *The* ***coherence relation*** $\frown$ *on* $\Delta_\oplus^{(!)}$ *is defined by:*

$$\frac{}{x \frown x} \qquad \frac{s \frown s'}{\lambda x.s \frown \lambda x.s'} \qquad \frac{s \frown s' \qquad \bar{t} \frown \bar{t'}}{\langle s\rangle\ \bar{t} \frown \langle s'\rangle\ \bar{t'}} \qquad \frac{s \frown s'}{s \oplus_p \bullet \frown s' \oplus_p \bullet} \qquad \frac{s \frown s'}{\bullet \oplus_p s \frown \bullet \oplus_p s'}$$

$$\frac{s \frown s \qquad t \frown t}{s \oplus_p \bullet \frown \bullet \oplus_p t} \qquad \frac{s \frown s \qquad t \frown t}{\bullet \oplus_p t \frown s \oplus_p \bullet} \qquad \frac{\forall i,j \le m+n, s_i \frown s_j}{[s_1,\dots,s_m] \frown [s_{m+1},\dots,s_{m+n}]}$$

For $S, S' \in \Delta_\oplus^{(!)}$ *we write* $S \frown S'$ *when for all* $\sigma,\sigma' \in \mathrm{supp}(S)\cup\mathrm{supp}(S')$*,* $\sigma \frown \sigma'$*. A simple (poly-)term* $\sigma \in \Delta_\oplus^{(!)}$ *is called* ***uniform*** *if* $\sigma \frown \sigma$*, and a term* $S \in \mathbb{R}^+\langle\Delta_\oplus^{(!)}\rangle$ *is called* ***uniform*** *if* $S \frown S$*.*

Observe that the term S defined above is not uniform: I_0 is an abstraction whereas for any $n \in \mathbb{N}$, I_{n+1} is an application, hence I_0 and I_{n+1} are not coherent. More generally I_m and I_n are not coherent whenever $m \neq n$. We can change the definition of I_0 to get a uniform term: let $I'_0 = \langle I\rangle\ []$, $I'_{n+1} = \langle I'_n\rangle\ [I]$ for $n \in \mathbb{N}$, and $S' = \sum_{n\in\mathbb{N}} I'_n$. Then I'_{n+1} reduces into I'_n, just like I_{n+1} reduces into I_n, but I'_0 is not a normal form as it reduces into 0. Thus the uniform term S' has a normal form $\mathrm{nf}(S') = \sum_{n\in\mathbb{N}} \mathrm{nf}(I'_n) = 0$.

▶ Remark 7. In the rules for $s \oplus_p \bullet \frown \bullet \oplus_p t$ and $\bullet \oplus_p t \frown s \oplus_p \bullet$ we require $s \frown s$ and $t \frown t$ to ensure that whenever $\sigma \frown \tau$, the simple (poly-)terms σ and τ are necessarily uniform. This is not crucial as we usually consider uniform (poly-)terms (whose support only contains uniform simple (poly-)terms), and indeed in [19] the non-deterministic terms $s \oplus \bullet$ and $\bullet \oplus t$ are always considered coherent. We only add this requirements to simplify inductive reasoning on $\frown$.

What makes coherence and uniformity interesting is that if two coherent terms S and S' have disjoint supports, then all of their reducts, and in particular their normal forms, have disjoint supports. Then any element in the support of $\mathrm{nf}(S + S')$ comes either from $\mathrm{nf}(S)$ or from $\mathrm{nf}(S')$, but it cannot come from both.

▶ **Proposition 8.** *Given* $S, S' \in \mathbb{R}^+[\Delta_\oplus^{(!)}]$*, if* $S \frown S'$ *then* $\mathrm{nf}(S) \frown \mathrm{nf}(S')$*. If moreover* $\mathrm{supp}(S) \cap \mathrm{supp}(S') = \emptyset$ *then* $\mathrm{supp}(\mathrm{nf}(S)) \cap \mathrm{supp}(\mathrm{nf}(S')) = \emptyset$*.*

This immediately implies that pointwise reduction of infinite uniform terms is well defined, as both complete left reducts and normal forms of distinct but coherent simple (poly-)terms have disjoint supports.

▶ **Corollary 9.** *If $S \in \mathbb{R}^+\langle\Delta_\oplus^{(!)}\rangle$ is uniform then $\sum_{\sigma\in\Delta_\oplus^{(!)}} S_\sigma \mathrm{nf}(\sigma)$ is in $\mathbb{R}^+\langle\Delta_\oplus^{(!)}\rangle$. We write $\mathrm{nf}(S)$ for this sum.*

2.3 Regular Terms

The deterministic Taylor expansion associates to any λ-term a *uniform* term, and explicit choices are adopted precisely for the sake of preserving this property in the probabilistic case. Taylor expansions have another important property: they are entirely defined by their support. If a simple term s is in the support of the Taylor expansion of a λ-term M, then its coefficient is the inverse of its *multinomial coefficient*, which does not depend on M. Moreover this property is preserved by normalisation. Using explicit choices enforces this result in the probabilistic case, as well.

▶ **Definition 10.** *For any $\sigma \in \Delta_\oplus^{(!)}$ we define the **multinomial coefficient** $\mathrm{m}(\sigma) \in \mathbb{N}$ by:*

$$\mathrm{m}(x) = 1 \qquad \mathrm{m}(\langle s\rangle\ \bar{t}) = \mathrm{m}(s)\mathrm{m}(\bar{t})$$

$$\mathrm{m}(\lambda x.s) = \mathrm{m}(s \oplus_p \bullet) = \mathrm{m}(\bullet \oplus_p s) = \mathrm{m}(s) \qquad \mathrm{m}(\bar{s}) = \prod_{u\in\Delta_\oplus} \bar{s}(u)! \cdot \mathrm{m}(u)^{\bar{s}(u)}$$

where $\bar{s}(u)$ is the multiplicity of u in $\bar{s}$.

▶ **Definition 11.** *A uniform term $S \in \mathbb{R}^+\langle\Delta_\oplus^{(!)}\rangle$ is called **regular** if for all $\sigma \in \mathrm{supp}(S)$, $S_\sigma = \frac{1}{\mathrm{m}(\sigma)}$.*

Multinomial coefficients correspond to the number of permutations of multisets which preserve the description of simple (poly-)terms. For instance, given variables $x_1, \dots, x_n \in \mathcal{V}$, the coefficient $\mathrm{m}([x_1, \dots, x_n])$ is exactly the number of permutations $\rho \in \mathfrak{S}_n$ such that $(x_{\rho(1)}, \dots, x_{\rho(n)}) = (x_1, \dots, x_n)$. For a more precise interpretation of multinomial coefficients see [9] or [14]. Due to their relation with permutations in multisets, these coefficients appear naturally when we perform substitutions.

▶ **Theorem 12.** *For any $\sigma \in \Delta_\oplus^{(!)}$ uniform, for $x \in \mathcal{V}$, $\bar{t} \in \Delta_\oplus^!$ and $u \in \mathrm{supp}(\delta_x\sigma \cdot \bar{t})$, we have: $(\delta_x\sigma \cdot \bar{t})_u = \frac{\mathrm{m}(\bar{t})\mathrm{m}(\sigma)}{\mathrm{m}(u)}$.*

This theorem ensures that a regular β-redex $\frac{1}{\mathrm{m}(\langle\lambda x.s\rangle\ \bar{t})}.\langle\lambda x.s\rangle\ \bar{t}$ reduces into a regular term. More generally, the theorem is the key step towards proving that regular (poly-)terms always normalise to regular (poly-)terms.

▶ **Theorem 13.** *If $S \in \mathbb{R}^+\langle\Delta_\oplus^{(!)}\rangle$ is regular then $\mathrm{nf}(S)$ is regular.*

2.4 Regularity and the Exponential

The regularity of terms is preserved by the constructors of simple resource terms.

▶ **Proposition 14.** *For all $x \in \mathcal{V}$, $S \in \mathbb{R}^+\langle\Delta_\oplus\rangle$ regular and $\overline{T} \in \mathbb{R}^+\langle\Delta_\oplus^!\rangle$ regular, the terms $1.x$, $\lambda x.S$, $S \oplus_p \bullet$, $\bullet \oplus_p S$ and $\langle S\rangle\ \overline{T}$ are regular.*

One may expect a similar result for poly-terms: if $S_1,\dots,S_n$ in $\mathbb{R}^+\langle\Delta_\oplus\rangle$ are regular then $[S_1,\dots,S_n]$ is regular. However, this is not the case: $1.x$ is regular and yet $1.[x,x]$ is not. Indeed nontrivial coefficients appear in $\mathrm{m}(\sigma)$ precisely when σ contains simple poly-terms with multiplicities greater than 1, so the regular sum with the same support as $[S_1,\dots,S_n]$ has no simple description. A natural way to build regular poly-terms from regular terms is to use the following construction.

▶ **Definition 15.** *The* ***exponential*** *of $S \in \mathbb{R}^+\langle\Delta_\oplus\rangle$ is $!S = \sum_{n\in\mathbb{N}} \frac{1}{n!}[S^n] \in \mathbb{R}^+\langle\Delta^!_\oplus\rangle$, where $[S^n]$ stands for the poly-term $[S,\dots,S]$ with n copies of S.*

▶ **Proposition 16.** *If $S \in \mathbb{R}^+\langle\Delta_\oplus\rangle$ is regular then $!S$ is regular.*

Proof. The key point is that the number of sequences $(s_1,\dots,s_n)$ which describe a given simple poly-term $\overline{s} = [s_1,\dots,s_n]$ is exactly $\frac{n!}{\prod_{u\in\Delta_\oplus}\overline{s}(u)!}$. ◀

With these results, we have all the ingredients we need to translate (probabilistic) λ-terms into regular terms: variables and abstractions of regular terms are regular, and we can define an application between regular terms following Girard's call-by-name translation of intuitionistic logic into linear logic [10]: S applied to T is $\langle S\rangle\ !T$.

3 Explicit Probabilistic Taylor Expansion

This section is devoted to defining and studying the *Taylor expansion with explicit choices*, or *explicit Taylor expansion*, of probabilistic λ-terms. It is named as such because its target is the set of probabilistic resource terms, as defined in the previous section, rather than the usual ones. This is *not* the main contribution of this paper, but an intermediate step in the study of Taylor expansion as defined in Section 4.

3.1 The Definition

Probabilistic λ-terms are λ-terms enriched with a probabilistic choice operator.

▶ **Definition 17.** *The set of* ***probabilistic λ-terms*** *Λ^+ is:*

$$M, N \in \Lambda^+ := x \mid \lambda x.M \mid M\ \ N \mid M \oplus_p N$$

▶ **Definition 18.** *The* ***explicit Taylor expansion*** *$M^\oplus$ is defined inductively as follows:*

$$x^\oplus = x \qquad (M\ \ N)^\oplus = \langle M^\oplus\rangle\ !N^\oplus = \sum_{n\in\mathbb{N}} \frac{1}{n!}\langle M^\oplus\rangle\ [(N^\oplus)^n]$$

$$(\lambda x.M)^\oplus = \lambda x.M^\oplus \qquad (M \oplus_p N)^\oplus = (M^\oplus \oplus_p \bullet) + (\bullet \oplus_p N^\oplus)$$

The results from the previous section immediately imply that Taylor expansions are regular resource terms and that they are normalisable.

▶ **Proposition 19.** *For all $M \in \Lambda^+$, the explicit Taylor expansion $M^\oplus$ is uniform and regular.*

Proof. This is a direct consequence of Proposition 14 and Proposition 16. ◀

▶ **Corollary 20.** *Every explicit Taylor expansion $M^\oplus$ has a normal form $\mathrm{nf}(M^\oplus)$, which we call the* ***explicit Taylor normal form*** *of M, and which is regular.*

Proof. This is given by Theorem 13. ◀

3.2 Probabilistic Reduction

In the literature, the probabilistic λ-calculus is usually endowed with a labelled transition relation $\xrightarrow{p}$ describing a probabilistic reduction process, where a choice $M \oplus_p N$ reduces to M with probability p and to N with probability $1-p$. Here to emphasise the correspondence between such a reduction and the constructors $s \oplus_p \bullet$ and $\bullet \oplus_p t$ of the resource calculus we rather use labels l, p and r, p to explicit whether we reduce to the left-hand side or the right-hand side of a choice $\oplus_p$. Since we were mostly interested in normalisation in the resource calculus, we will only consider a big-step operational semantics for the probabilistic λ-calculus.

▶ **Definition 21.** ***Head contexts*** *are contexts of the form $\lambda\vec{x}.[\,]\,\vec{P}$, and are indicated with the metavariable H.* ***Head normal forms*** *are terms of the form $H[y]$. We write* hnf *for the set of all head normal forms. We now define a formal system deriving judgements in the form $\rho \vdash M \twoheadrightarrow h$ where $M \in \Lambda^+$, $h \in$ hnf and ρ is a finite sequence of elements in $\{\mathrm{l},\mathrm{r}\} \times [0,1]$:*

$$\frac{}{\epsilon \vdash h \twoheadrightarrow h} \qquad \frac{\rho \vdash H[M\left[{}^{N}/_{x}\right]] \twoheadrightarrow h}{\rho \vdash H[(\lambda x.M)N] \twoheadrightarrow h} \qquad \frac{\rho \vdash H[M] \twoheadrightarrow h}{(\mathrm{l},p)\cdot\rho \vdash H[M \oplus_p N] \twoheadrightarrow h} \qquad \frac{\rho \vdash H[N] \twoheadrightarrow h}{(\mathrm{r},p)\cdot\rho \vdash H[M \oplus_p N] \twoheadrightarrow h}$$

where ϵ is the empty sequence and $(\ell,p)\cdot(\rho_1,\ldots,\rho_n) = ((\ell,p),\rho_1,\ldots,\rho_n)$ for $\ell \in \{\mathrm{l},\mathrm{r}\}$.

▶ **Lemma 22.** *For all $M \in \Lambda^+$ and ρ there is at most one $h \in$ hnf such that $\rho \vdash M \twoheadrightarrow h$.*

An interesting property of explicit Taylor expansion is that the explicit Taylor normal form of a term M is precisely given by the explicit Taylor normal forms of the head normal forms h of M, as well as the sequences of choices ρ such that $\rho \vdash M \twoheadrightarrow h$.

▶ **Definition 23.** *Given a sequence of choices ρ and $s \in \Delta_\oplus$ we define $\rho \cdot s \in \Delta_\oplus$ by induction on the length of ρ by:*

$$\epsilon \cdot s = s \qquad ((\mathrm{l},p)\cdot\rho)\cdot s = (\rho\cdot s) \oplus_p \bullet \qquad ((\mathrm{r},p)\cdot\rho)\cdot s = \bullet \oplus_p (\rho \cdot s)$$

We extend this definition to $\mathbb{R}^+\langle\Delta_\oplus\rangle$ by linearity.

▶ **Theorem 24.** *Given any $M \in \Lambda^+$,*

$$\mathrm{nf}(M^\oplus) = \sum_{h\in\mathrm{hnf}} \sum_{\rho\vdash M\twoheadrightarrow h} \rho\cdot\mathrm{nf}(h^\oplus).$$

Proof. First observe that these resource terms are regular: Corollary 20 states that $\mathrm{nf}(M^\oplus)$ and the $\mathrm{nf}(h^\oplus)$ are regular (so the $\rho\cdot\mathrm{nf}(h^\oplus)$ are regular too), and if $\rho \vdash M \twoheadrightarrow h$ and $\rho' \vdash M \twoheadrightarrow h'$ then either $\rho = \rho'$ and by Lemma 22 $h = h'$, or $\rho \neq \rho'$ and then $\rho\cdot\mathrm{nf}(h^\oplus)$ and $\rho'\cdot\mathrm{nf}(h'^\oplus)$ have disjoint supports and we can show they are coherent. Thus we only need to prove that these terms have the same supports. On one hand we can prove that for any $s \in \mathrm{supp}(M^\oplus)$ and any $t \in \mathrm{supp}(\mathrm{nf}(s))$ there exist ρ and h such that $\rho \vdash M \twoheadrightarrow h$ and $t \in \mathrm{supp}(\rho\cdot\mathrm{nf}(h^\oplus))$, by reasoning by induction on the size of s. When s has a head β-redex we need to check that in general if $u \in \mathrm{supp}(U^\oplus)$ and $\overline{v} \in \mathrm{supp}(!V^\oplus)$ then $\mathrm{supp}(\delta_x u \cdot \overline{v}) \subset \mathrm{supp}((U\left[{}^{V}/_{x}\right])^\oplus)$, which is immediate by induction on U. On the other hand if $\rho \vdash M \twoheadrightarrow h$ we prove that for any $t \in \mathrm{supp}(\mathrm{nf}(h^\oplus))$ we have $\rho\cdot t \in \mathrm{supp}(\mathrm{nf}(M^\oplus))$, by induction on the proof of $\rho \vdash M \twoheadrightarrow h$. ◀

4 Generic Taylor Expansion of Probabilistic λ-terms

4.1 Barycentric Semantics of Choices

The explicit probabilistic Taylor expansion is satisfactory in that it is an extension of deterministic Taylor expansion which preserves its most important properties: it is regular and so are its normal forms. But while deterministic Taylor normal forms are well known to correspond to Böhm trees [7], explicit Taylor normal forms are not such a good denotational semantics for probabilistic λ-calculus, as they take the exact choices made during the reduction into account. For instance the terms $x \oplus_{\frac{1}{2}} y$ and $y \oplus_{\frac{1}{2}} x$ have *distinct* explicit Taylor normal forms while one could expect them to have *the same semantics*. More precisely we expect any model of the probabilistic λ-calculus to interpret probabilistic choices as a *barycentric sum* respecting the following equivalence.

▶ **Definition 25.** *The* ***barycentric equivalence*** $\equiv_{\mathrm{bar}}$ *is the least congruence on* Λ^+ *such that for all* $M, N, P \in \Lambda^+$ *and* $p, q \in [0,1]$*:*

$$M \oplus_p N \equiv_{\mathrm{bar}} N \oplus_{1-p} M \qquad\qquad M \oplus_p M \equiv_{\mathrm{bar}} M$$
$$(M \oplus_p N) \oplus_q P \equiv_{\mathrm{bar}} M \oplus_{pq} (N \oplus_{\frac{q(1-p)}{1-pq}} P) \text{ if } pq \neq 1 \qquad\qquad M \oplus_1 N \equiv_{\mathrm{bar}} M$$

We want a notion of Taylor expansion M^* such that if $M \equiv_{\mathrm{bar}} N$ then $M^* = N^*$. This is easy to achieve, as the resource calculus stemmed precisely from quantitative models of the λ-calculus, and resource terms are linear combinations.

▶ **Definition 26.** *The sets of* ***simple resource terms*** Δ *and of* ***simple resource poly-terms*** $\Delta^!$ *are:*

$$s, t \in \Delta := x \mid \lambda x.s \mid \langle s \rangle\ \bar{t} \qquad\qquad \bar{s}, \bar{t} \in \Delta^! := [s_1, \ldots, s_n]$$

The set of ***resource terms*** *is* $\mathbb{R}^+\langle \Delta \rangle$ *and the set of* ***resource poly-terms*** *is* $\mathbb{R}^+\langle \Delta^!_\oplus \rangle$.

▶ **Definition 27.** *The* ***Taylor expansion*** $M^* \in \mathbb{R}^+\langle \Delta \rangle$ *of a term* $M \in \Lambda^+$ *is defined inductively as follows:*

$$x^* = x \qquad\qquad (M\ N)^* = \sum_{n \in \mathbb{N}} \frac{1}{n!} \langle M^* \rangle\ [(N^*)^n]$$
$$(\lambda x.M)^* = \lambda x.M^* \qquad\qquad (M \oplus_p N)^* = pM^* + (1-p)N^*$$

The definition of the Taylor expansion of a probabilistic choice immediately gives the expected property.

▶ **Proposition 28.** *If* $M \equiv_{\mathrm{bar}} N$ *then* $M^* = N^*$.

4.2 Normalisation

Unfortunately, these Taylor expansions lack all the good properties of explicit expansions: they are not entirely defined by their support, and those supports are not uniform, so we do not even know if such Taylor expansions admit normal forms. But there is actually a close relationship between explicit and non explicit Taylor expansions which can be used to recover our most important results. Indeed, switching from the explicit Taylor expansion to the Taylor expansion simply amounts to using coefficients instead of explicit choices.

▶ **Definition 29.** *Given any $\sigma \in \Delta_\oplus^{(!)}$ we define $|\sigma| \in \Delta^{(!)}$ and a probability $\mathcal{P}(\sigma)$ as follows:*

$$\begin{aligned} |x| &= x & \mathcal{P}(x) &= 1 \\ |\lambda x.s| &= \lambda x.|s| & \mathcal{P}(\lambda x.s) &= \mathcal{P}(s) \\ |\langle s\rangle\ \bar{t}| &= \langle |s|\rangle\ |\bar{t}| & \mathcal{P}(\langle s\rangle\ \bar{t}) &= \mathcal{P}(s)\mathcal{P}(\bar{t}) \\ |s \oplus_p \bullet| &= |s| & \mathcal{P}(s \oplus_p \bullet) &= p\mathcal{P}(s) \\ |\bullet \oplus_p s| &= |s| & \mathcal{P}(\bullet \oplus_p s) &= (1-p)\mathcal{P}(s) \\ |[s_1,\dots,s_n]| &= [|s_1|,\dots,|s_n|] & \mathcal{P}([s_1,\dots,s_n]) &= \prod_{i=1}^{n} \mathcal{P}(s_i) \end{aligned}$$

To any probabilistic resource (poly-)term $S \in \mathbb{R}^+\langle\Delta_\oplus^{(!)}\rangle$ one could associate the resource term $\sum_{\sigma\in\Delta_\oplus^{(!)}} S_\sigma \mathcal{P}(\sigma).|\sigma|$. But just like with normalisation, infinite coefficients may appear. For instance, removing the choices from $S = \sum((x \oplus_1 \bullet)\dots) \oplus_1 \bullet$ (the sum of all simple terms with x under any number of left choices) could give x an infinite coefficient. Fortunately, we do not get any infinite coefficient if we work with regular terms.

▶ **Proposition 30.** *For any $\mathcal{S} \subset \Delta_\oplus^{(!)}$ such that for all $\sigma, \sigma' \in \mathcal{S}$, $\sigma \supset\subset \sigma'$ and $|\sigma| = |\sigma'|$ we have $\sum_{\sigma\in\mathcal{S}} \mathcal{P}(\sigma) \leq 1$.*

▶ **Corollary 31.** *For all $S \in \mathbb{R}^+\langle\Delta_\oplus^{(!)}\rangle$ regular, $\sum_{\sigma\in\Delta_\oplus^{(!)}} S_\sigma \mathcal{P}(\sigma).|\sigma|$ is in $\mathbb{R}^+\langle\Delta^{(!)}\rangle$.*

In particular, we can apply this process to explicit Taylor expansions *and to their normal forms*. It is easy to see that we associate to every explicit Taylor expansion the corresponding Taylor expansion, but more interestingly erasing choices commutes with normalisation.

▶ **Proposition 32.** *For any $M \in \Lambda^+$:*

$$\sum_{s\in\Delta_\oplus} M_s^\oplus \mathcal{P}(s).|s| = M^* \qquad \sum_{t\in\Delta_\oplus} \mathrm{nf}(M^\oplus)_t \mathcal{P}(t).|t| = \sum_{s\in\Delta} M_s^*.\mathrm{nf}(s)$$

hence $\sum_{s\in\Delta} M_s^*.\mathrm{nf}(s)$ is well defined. We denote it by $\mathrm{nf}(M^*)$ and we call it the ***Taylor normal form*** *of M.*

Proof. The key point is that $\mathrm{nf}(|\sigma|) = |\mathrm{nf}(\sigma)|$ and for any $\tau \in \mathrm{supp}(\mathrm{nf}(\sigma))$, $\mathcal{P}(\tau) = \mathcal{P}(\sigma)$. ◀

4.3 Adequacy

The behaviour of a probabilistic λ-term is usually described as a (sub-)probability distribution over the possible results of its evaluation. In particular, the *observable* behaviour of a term is its *convergence probability*, i.e. the probability for its computation to terminate [11, 5]. To show that the Taylor expansion gives a meaningful semantics we will prove it is *adequate*, i.e. it does not equate terms which are not observationally equivalent. We can actually show a more refined result, given as a Corollary of Theorem 24: the Taylor normal form of a term is given by the Taylor normal forms of its head normal forms.

▶ **Definition 33.** *The any sequence of choices ρ we associate a probability $\mathcal{P}(\rho)$ by:*

$$\mathcal{P}(\epsilon) = 1 \qquad \mathcal{P}((\mathrm{l}, p) :: \rho) = p\mathcal{P}(\rho) \qquad \mathcal{P}((\mathrm{r}, p) :: \rho) = (1-p)\mathcal{P}(\rho)$$

The probability $\mathcal{P}(M \twoheadrightarrow h)$ for $M \in \Lambda^+$ to reduce into a head normal form h and its ***convergence probability*** *$\mathcal{P}_\Downarrow(M)$ are defined as follows:*

$$\mathcal{P}(M \twoheadrightarrow h) := \sum_{\rho\vdash M\twoheadrightarrow h} \mathcal{P}(\rho) \qquad \mathcal{P}_\Downarrow(M) = \sum_{h\in\mathrm{hnf}} \mathcal{P}(M \twoheadrightarrow h).$$

▶ **Proposition 34.** *For $M \in \Lambda^+$ we have:*

$$\mathrm{nf}(M^*) = \sum_{h \in \mathrm{hnf}} \mathcal{P}\left(M \twoheadrightarrow h\right) \mathrm{nf}(h^*).$$

Proof. This is given by Proposition 32 and Theorem 24. Observe that for any ρ and $s \in \mathrm{nf}(h^{\oplus})$ we have $\mathcal{P}(\rho \cdot s) = \mathcal{P}(\rho)\mathcal{P}(s)$ and $|\rho \cdot s| = |s|$. ◀

The adequacy follows immediately.

▶ **Proposition 35.** *If $\mathrm{nf}(M^*) = \mathrm{nf}(N^*)$ then for all context C, $\mathcal{P}_{\Downarrow}(C[M]) = \mathcal{P}_{\Downarrow}(C[N])$, i.e. M and N are contextually equivalent.*

Proof. First the convergence probability of a term M is exactly the sum of the coefficients $\mathrm{nf}(M^*)_{\lambda\vec{x}.y\,[\,]\ldots[\,]}$. Second if $\mathrm{nf}(M^*) = \mathrm{nf}(N^*)$ then $\mathrm{nf}(C[M]^*) = \mathrm{nf}(C[N]^*)$ for all C. ◀

5 On the Taylor Expansion and Böhm Trees

5.1 A Commutation Theorem

Deterministic Taylor normal forms are an adequate semantics for the probabilistic λ-calculus, but more precisely they are known to correspond to Böhm trees [7]. We are now able to show that this result extends to the probabilistic case.

▶ **Definition 36.** *The sets of **probabilistic Böhm trees** $\mathcal{PT}_d$ and of **probabilistic value trees** $\mathcal{VT}_d$ for $d \in \mathbb{N}$ are defined inductively by induction on the **depth** d:*

$$\mathcal{PT}_0 = \{\perp : \emptyset \to [0,1]\} \qquad \mathcal{VT}_0 = \emptyset$$
$$\mathcal{PT}_{d+1} = \mathbf{D}(\mathcal{VT}_{d+1}) \qquad \mathcal{VT}_{d+1} = \{\lambda\vec{x}.y\ \mathbf{T}_1\ \cdots\ \mathbf{T}_m \mid \mathbf{T}_1, \ldots, \mathbf{T}_m \in \mathcal{PT}_d\}$$

where $\mathbf{D}(X)$ is the set of countable-support subprobability distributions on any set X, $\perp$ is the only subprobability distribution over the empty set, i.e. over $\mathcal{VT}_0$.

▶ **Definition 37.** *We define $PT_d(M)$ for $M \in \Lambda^+$ and $d \geq 0$, and $VT_d(h)$ for $h \in \mathrm{hnf}$ and $d \geq 1$ by induction on the depth d as follows: $PT_0(M)$ is the unique function $\emptyset \to [0,1]$ and*

$$PT_{d+1}(M) = \mathbf{t} \mapsto \sum_{h \in VT_{d+1}^{-1}(\mathbf{t})} \mathcal{P}\left(M \twoheadrightarrow h\right)$$
$$VT_{d+1}(\lambda\vec{x}.y\ M_1\ \ldots\ M_m) = \lambda\vec{x}.y\ PT_d(M_1)\ \ldots\ PT_d(M_m)$$

Intuitively the Böhm tree of a term M would be some limit of its finite Böhm approximants $PT_d(M)$. To avoid making the structure of Böhm trees of infinite depth explicit, we simply write $PT(M)$ for the sequence $(PT_d(M))_{d\in\mathbb{N}}$. In particular we say that M and N have *the same Böhm tree* iff $PT_d(M) = PT_d(N)$ for every $d \in \mathbb{N}$.

The definition of the Taylor expansion can easily be generalised to finite-depth Böhm trees. We simply define $\mathbf{T}^*$ for $\mathbf{T} \in \mathcal{PT}_d$ and $\mathbf{t}^*$ for $\mathbf{t} \in \mathcal{VT}_{d+1}$ by:

$$\mathbf{T}^* = \sum_{\mathbf{t} \in \mathcal{VT}_d} \mathbf{T}(\mathbf{t})\mathbf{t}^* \qquad (\lambda\vec{x}.y\ \mathbf{T}_1\ \ldots\ \mathbf{T}_m)^* = \lambda\vec{x}.\langle y\rangle\ !\mathbf{T}_1^* \ldots !\mathbf{T}_m^*$$

We extend this definition to infinite Böhm trees as follows: if $s \in \Delta$ contains at most d_s layers of nested multisets then for any $M \in \Lambda^+$, $PT_d(M)^*_s = PT_{d_s}(M)^*_s$ for all $d \geq d_s$, so $PT(M)^*_s$ can be taken as $PT_{d_s}(M)^*_s$. Then the Taylor normal form of a term is exactly the Taylor expansion of its Böhm tree.

▶ **Theorem 38.** *For all $M \in \Lambda^+$, $\mathrm{nf}(M^*) = (PT(M))^*$.*

Proof. We prove $\mathrm{nf}(M^*)_s = (PT(M))^*_s$ by induction on d_s, using to Proposition 34. ◀

This theorem is important but it does not actually prove the correspondence between Böhm trees and Taylor expansions: we still do not know if Taylor expansion is injective on Böhm trees. In the deterministic case this is simple to prove: to every deterministic Böhm tree $\mathbf{T}$ of depth d we can associate a simple resource term $s_{\mathbf{T}}$ such that for all $M \in \Lambda$, $BT_d(M) = \mathbf{T}$ iff $s_{\mathbf{T}} \in \mathrm{supp}(\mathrm{nf}(M^*))$ (by associating $\lambda\vec{x}.\langle y\rangle\ [s_{\mathbf{T}_1}] \ldots [s_{\mathbf{T}_m}]$ to $\lambda\vec{x}.y\,\mathbf{T}_1 \ldots \mathbf{T}_m$). This works because ordinary Böhm trees are not quantitative, thus the quantitative part of their Taylor expansions (the coefficients) is irrelevant. The situation is more complicated in the probabilistic case, as Taylor expansions are no longer defined solely by their supports. The rest of this article is devoted to proving injectivity for the probabilistic Taylor expansion.

5.2 Böhm Tests

In order to better understand coefficients in probabilistic Taylor expansions and to get our injectivity property, we use a notion of *testing* coming from the literature on labelled Markov decision processes [17].

▶ **Definition 39** (Böhm Tests). *The classes of* Böhm term tests *(BTTs) and* Böhm hnf tests *(BHTs) are given as follows, by mutual induction:*

$$T, U ::= \omega \mid T \wedge U \mid \mathsf{ev}(t) \qquad\qquad t, u ::= \omega \mid t \wedge u \mid (\lambda x_1.\cdots.\lambda x_n.y)(T^1, \ldots, T^m)$$

The probability of success of a BTT T on a term M and the probability of success of a BHT t on an head-normal-form h, indicated as $\mathsf{Pr}(T, M)$ and $\mathsf{Pr}(t, h)$ respectively, are defined as follows:

$$\mathsf{Pr}(T \wedge U, M) = \mathsf{Pr}(T, M) \cdot \mathsf{Pr}(U, M); \qquad \mathsf{Pr}(\omega, M) = \mathsf{Pr}(\omega, h) = 1;$$
$$\mathsf{Pr}(t \wedge u, h) = \mathsf{Pr}(t, h) \cdot \mathsf{Pr}(u, h); \qquad \mathsf{Pr}(\mathsf{ev}(t), M) = \sum_{h \in \mathrm{hnf}} \mathcal{P}\,(M \twoheadrightarrow h) \cdot \mathsf{Pr}(t, h);$$
$$\mathsf{Pr}((\lambda x_1.\cdots.\lambda x_n.y)(T^1, \ldots, T^m), \lambda x_1 \cdots .\lambda x_n.yM_1 \cdots M_m) = \Pi_{i=1}^m \mathsf{Pr}(T^i, M_i);$$
$$\mathsf{Pr}((\lambda x_1.\cdots.\lambda x_n.y)(T^1, \ldots, T^m), h) = 0, \textit{otherwise}$$

The following is the first step towards proving the main result of this paper, as it characterises Böhm tree equality as equality of families of real numbers.

▶ **Theorem 40.** *Two terms M and N have the same Böhm trees iff for every BTT T it holds that $\mathsf{Pr}(M, T) = \mathsf{Pr}(N, T)$.*

A detailed proof of Theorem 40 can be found in the Extended Version of this paper [4]. Let us briefly discuss how the proof goes. The starting point is a result due to van Breugel et al. [17], which establishes a precise correspondence between bisimilarity and testing in a probabilistic scenario: two states s and s' of any labelled Markov decision processes (satisfying certain natural conditions) are bisimilar *iff* the probabilities of any test T to succeed in s and in s' are the same, and tests are defined inductively as follows:

$$T, U ::= \omega \mid T \wedge U \mid a(T)$$

where a is an action of the underlying labelled Markov decision process. Given that Böhm trees can be naturally presented coinductively, most of the involved work has already been done. What remains to be proved, then, is that the result above also holds for transition systems in which firing an action a brings the system into k distinct states, thus capturing the kind of tree-like evolution typical of Böhm trees. This can be done by translating any such tree-like transition system into a linear one in such a way that bisimilarity and testing-equivalence remain unaltered.

To achieve our main result we now need to relate the probability of success of tests to the coefficients of Taylor normal forms. This is the purpose of Section 6.

6 Implementing Tests as Resource Terms

There is a very tight correspondence between simple resource terms and Böhm tests, but this correspondence does not hold for *all* Böhm tests. Simple resource terms can be seen as a particular class of Böhm tests.

▶ **Definition 41.** *The classes of* resource Böhm term tests *(rBTTs) and* resource Böhm hnf tests *(rBHTs) are given as follows, by mutual induction:*

$$T, U ::= \omega \mid T \wedge U \mid \mathsf{ev}(t) \qquad t ::= (\lambda x_1.\cdots.\lambda x_n.y)(T^1, \ldots, T^m)$$

▶ **Definition 42.** *For every rBTT T we define a simple poly-term $\overline{s}_T$ and for every rBHT t we define a simple term s_t in the following way:*

$$\overline{s}_\omega = [\,] \quad \overline{s}_{T\wedge U} = \overline{s}_T \cdot \overline{s}_U \quad \overline{s}_{\mathsf{ev}(t)} = [s_t] \quad s_{(\lambda\vec{x}.y)(T^1,\ldots,T^m)} = \lambda\vec{x}.\langle y\rangle\ \overline{s}_{T^1}\ \ldots\ \overline{s}_{T^m}$$

The similarity between simple resource terms and resource Böhm tests is more than structural: the probability of success of a resource Böhm test is actually given by a coefficient in the Taylor normal form.

▶ **Proposition 43.**
1. *For every rBTT T and $M \in \Lambda^+$,* $!\mathrm{nf}(M^*)_{\overline{s}_T} = \frac{\mathsf{Pr}(T,M)}{\mathrm{m}(\overline{s}_{T_t})}$.
2. *For every rBHT t and $h \in \mathrm{hnf}$,* $\mathrm{nf}(h^*)_{s_t} = \frac{\mathsf{Pr}(t,h)}{\mathrm{m}(s_{T_h})}$.

Proof. We reason by induction on tests. Observe that these can be considered modulo commutativity and associativity of the conjunction and modulo $\omega \wedge T \simeq T$: these equivalences preserve both the results of testing and the associated simple resource (poly-)terms. Then every rBTT is equivalent either to ω or to a conjunction $T = \mathsf{ev}(t_1) \wedge \cdots \wedge \mathsf{ev}(t_k)$. In the first case we always have $!\mathrm{nf}(M^*)_{[\,]} = 1$. In the second case just like in the proof of regularity of the exponential (Proposition 16) for any $M \in \Lambda^+$ we have $!\mathrm{nf}(M^*)_{\overline{s}_T} = \frac{1}{\prod_{u\in\Delta} \overline{s}_T(u)!} \prod_{i=1}^k \mathrm{nf}(M^*)_{s_{t_i}}$. To conclude we want to show that $\mathrm{nf}(M^*)_{s_{t_i}} = \frac{\mathsf{Pr}(\mathsf{ev}(t_i),M)}{\mathrm{m}(s_{t_i})}$ for all $i \leq k$. We have by definition $\mathsf{Pr}(\mathsf{ev}(t_i), M) = \sum_{h\in\mathrm{hnf}} \mathcal{P}\,(M \twoheadrightarrow h) \cdot \mathsf{Pr}(t_i, h)$, and Proposition 34 gives $\mathrm{nf}(M^*)_{s_{t_i}} = \sum_{h\in\mathrm{hnf}} \mathcal{P}\,(M \twoheadrightarrow h) \cdot \mathrm{nf}(h^*)_{s_{t_i}}$, so we conclude by induction hypothesis on t_i. Now given a rBHT $t = (\lambda\vec{x}.y)(T^1, \ldots, T^m)$ and $h \in \mathrm{hnf}$ we have either $\mathrm{nf}(h^*)_{s_t} = \prod_{i=1}^m !\mathrm{nf}(M_i^*)_{\overline{s}_{T^i}}$ and $\mathsf{Pr}(t,h) = \prod_{i=1}^m \mathsf{Pr}(T^i, M_i)$ if h is of the form $\lambda\vec{x}.y\ M_1\ \ldots\ M_m$, in which case we conclude by induction hypothesis, or $\mathrm{nf}(h^*)_{s_t} = \mathsf{Pr}(t,h) = 0$ otherwise. ◀

With this result, we completely characterise Taylor normal forms by resource Böhm tests.

▶ **Corollary 44.** *Two terms M and N have the same Taylor normal form iff for every rBTT T it holds that* $\mathsf{Pr}(M,T) = \mathsf{Pr}(N,T)$.

Proof. Simply observe that every simple resource term in normal form is equal to s_T for some resource Böhm test T. ◀

Thanks to Theorem 40 and Corollary 44 both Böhm tree equality and Taylor normal form equality are characterised by tests. They still leave a gap in our reasoning, as not all Böhm tests are resource Böhm tests. This difference is not just cosmetic: $\mathsf{ev}(\omega)$ is a valid Böhm test which computes the convergence probability of any λ-term, which cannot be done using only resource Böhm tests. More precisely this cannot be done using a *single* Böhm test. To fill the gap between Böhm tests and resource Böhm tests we observe that any of the former can be simulated by a *family* of resource Böhm tests.

▶ **Proposition 45.** *For every BTT T there is a family $(T_i)_{i\in I}$ of rBTTs of arbitrary size (possibly empty, possibly infinite) such that for all λ-term M we have $\mathsf{Pr}(T,M) = \sum_{i\in I}\mathsf{Pr}(T_i,M)$.*

Proof. We prove this, as well as the corresponding result for BHTs, by induction on the size of tests. In the case of BTTs, the result is simply given by induction hypothesis. To the BTT ω we associate the single-element family (w), to $T \wedge U$ we associate $(T_i \wedge U_j)_{i\in I, j\in J}$ where $(T_i)_{i\in I}$ and $(U_j)_{j\in J}$ are given by induction hypothesis on T and U, and to $\mathsf{ev}(t)$ we associate $(\mathsf{ev}(t_i))_{i\in I}$. The interesting part of the proof is on BHTs, where we want to remove two constructors. Modulo commutativity and associativity of the conjunction and the equivalence $\omega \wedge T \simeq T$, every BHT is either ω or of the form $(\lambda x_1 ... x_{n_1}.y_1)(T_1^1,\dots,T_1^{m_1}) \wedge \dots \wedge (\lambda x_1 ... x_{n_k}.y_k)(T_k^1,\dots,T_k^{m_k})$ with $k \geq 1$. In the first case to ω we associate the family $((\lambda x_1 \dots x_n.y)(\omega^m))_{m,n\in\mathbb{N},y\in\mathcal{V}}$ where ω^m denotes the sequence $\omega,\dots,\omega$ of length m. In the second case if $m_i \neq m_j$, $n_i \neq n_j$ or $y_i \neq y_j$ for some $i,j \leq k$ then the result of the test is always 0, which is simulated by the empty family of rBHTs. Otherwise let $m = m_1$, $n = n_1$ and $y = y_1$, the test is equivalent to $(\lambda x_1 \dots x_n.y)(T_1^1 \wedge \dots \wedge T_k^1,\dots,T_1^m \wedge \dots \wedge T_k^m)$. We apply the induction hypothesis to the BTTs $T_1^i \wedge \dots \wedge T_k^i$ to get families $(U_j^i)_{j\in J_i}$ and we associate the family $((\lambda x_1 \dots x_n.y)(U_{j_1}^1,\dots,U_{j_m}^m))_{j_1\in J_1,\dots,j_m\in J_m}$ to the original BHT. ◀

▶ **Corollary 46.** *Given two terms M and N, for every BTT T it holds that $\mathsf{Pr}(M,T) = \mathsf{Pr}(N,T)$ iff for every rBTT T it holds that $\mathsf{Pr}(M,T) = \mathsf{Pr}(N,T)$.*

We can now state the main result of this paper.

▶ **Theorem 47.** *Two terms have the same Böhm trees iff their Taylor expansions have the same normal forms.*

Proof. The result follows from Theorem 40, Corollary 46 and Corollary 44. ◀

7 Conclusion

In this paper, we attack the problem of extending the Taylor Expansion construction to the probabilistic λ-calculus, at the same time preserving its nice properties. What we find remarkable about the defined notion of Taylor expansion is that its codomain is the set of *ordinary* resource terms, and that the equivalence induced by the Taylor expansion is precisely the one induced by Böhm trees [13]. The latter, not admitting η, is strictly included in contextual equivalence.

Among the many questions this work leaves open, we could cite the extension of the proposed definition to call-by-value reduction, along the lines of [12], and a formal comparison between the notion of equivalence introduced here and the the one from [15] in which, however, the target language is not the one of ordinary resource terms, but one specifically designed around probabilistic effects.

References

1 H.P. Barendregt. *The Lambda Calculus: Its Syntax and Semantics*. Studies in Logic and the Foundations of Mathematics. Elsevier Science, 1984.

2 Johannes Borgström, Ugo Dal Lago, Andrew D. Gordon, and Marcin Szymczak. A lambda-calculus foundation for universal probabilistic programming. In *Proc. of ICFP 2016*, pages 33–46, 2016.

3 Gérard Boudol. The Lambda-Calculus with Multiplicities. Technical Report 2025, INRIA Sophia-Antipolis, 1993.

4 Ugo Dal Lago and Thomas Leventis. On the Taylor Expansion of Probabilistic Lambda Terms (Long Version), 2019. `arXiv:1904.09650`.

5 Thomas Ehrhard, Michele Pagani, and Christine Tasson. Full Abstraction for Probabilistic PCF. *J. ACM*, 65(4):23:1–23:44, 2018.

6 Thomas Ehrhard and Laurent Regnier. The differential lambda-calculus. *Theor. Comput. Sci.*, 309(1-3):1–41, 2003.

7 Thomas Ehrhard and Laurent Regnier. Böhm Trees, Krivine's Machine and the Taylor Expansion of Lambda-Terms. In *Proc. of CIE 2006*, pages 186–197, 2006.

8 Thomas Ehrhard and Laurent Regnier. Differential interaction nets. *Theor. Comput. Sci.*, 364(2):166–195, 2006.

9 Thomas Ehrhard and Laurent Regnier. Uniformity and the Taylor expansion of ordinary lambda-terms. *Theor. Comput. Sci.*, 403(2-3):347–372, 2008.

10 Jean-Yves Girard. Linear Logic. *Theor. Comput. Sci.*, 50:1–102, 1987.

11 Claire Jones and Gordon D. Plotkin. A Probabilistic Powerdomain of Evaluations. In *Proc. of LICS 1989*, pages 186–195, 1989.

12 Emma Kerinec, Giulio Manzonetto, and Michele Pagani. Revisiting Call-by-value Bohm trees in light of their Taylor expansion, 2018. `arXiv:1809.02659`.

13 Thomas Leventis. Probabilistic Böhm Trees and Probabilistic Separation. In *Proc. of LICS 2018*, pages 649–658, 2018.

14 Takeshi Tsukada, Kazuyuki Asada, and C.-H. Luke Ong. Generalised species of rigid resource terms. In *Proc. of LICS 2017*, pages 1–12, 2017.

15 Takeshi Tsukada, Kazuyuki Asada, and C.-H. Luke Ong. Species, Profunctors and Taylor Expansion Weighted by SMCC: A Unified Framework for Modelling Nondeterministic, Probabilistic and Quantum Programs. In *Proc. of LICS 2018*, pages 889–898, 2018.

16 Matthijs Vákár, Ohad Kammar, and Sam Staton. A domain theory for statistical probabilistic programming. *PACMPL*, 3(POPL):36:1–36:29, 2019.

17 Franck van Breugel, Michael W. Mislove, Joël Ouaknine, and James Worrell. Domain theory, testing and simulation for labelled Markov processes. *Theor. Comput. Sci.*, 333(1-2):171–197, 2005.

18 Lionel Vaux. The algebraic lambda calculus. *Mathematical Structures in Computer Science*, 19(5):1029–1059, 2009.

19 Lionel Vaux Auclair and Federico Olimpieri. On the Taylor expansion of λ-terms and the groupoid structure of their rigid approximants. Informal proc. of TLLA 2018, 2018.