The probability of non-confluent systems

Alejandro Díaz-Caro

Université Paris Ouest
200 avenue de la République
92001 Nanterre, France

alejandro@diaz-caro.info

Gilles Dowek

INRIA
23 avenue d'Italie, CS 81321
75214 Paris Cedex 13, France

gilles.dowek@inria.fr

INRIA
23 avenue d'Italie, CS 81321
75214 Paris Cedex 13, France

We show how to provide a structure of probability space to the set of execution traces on a non-confluent abstract rewrite system, by defining a variant of a Lebesgue measure on the space of traces. Then, we show how to use this probability space to transform a non-deterministic calculus into a probabilistic one. We use as example λ_+, a recently introduced calculus defined through type isomorphisms.

1 Introduction

Many probabilistic calculi has been developed in the pasts years, e.g. [1, 9, 11, 14, 20]. In particular, the algebraic versions of λ-calculus [5, 24] are extensions to λ-calculus where a linear combination of terms, e.g. $\alpha.\mathbf{r} + \beta.\mathbf{s}$, is also a term. One way to interpret such a linear combination is that it represents a term which is the term $\mathbf{r}$ with probability α, or the term $\mathbf{s}$ with probability β. However, endowing such a calculus with a non-restrictive type system is a challenge [3, 4].

A simpler framework is that of non deterministic calculi which can be seen as algebraic calculi withouth scalars. They have been studied, for instance in [8, 10, 12, 13, 15–18, 21], however moving back from non-determinism to probabilities is not trivial. In this paper we propose, instead of changing these models, to define a probability measure on reductions in non-deterministic systems. In fact, as we shall see, such a probability measure can be defined on any abstract non-deterministic transition systems, or non-confluent abstract rewrite systems (ARS) (cf. [23, Chapter 1]). Our goal is to show that explicit probabilities are not needed in the syntax, and that the simpler non-deterministic calculi are as powerful as the more complicated probabilistic calculi.

Consider for example the following non-confluent ARS

$$\mathbf{a} \to \mathbf{b} \ , \qquad \mathbf{a} \to \mathbf{c} \ , \qquad \mathbf{c} \to \mathbf{d} \ , \qquad \mathbf{c} \to \mathbf{e} \ ,$$

we want to associate a probability to events such as

$$\mathbf{a} \to^* \mathbf{b} \ , \qquad \mathbf{a} \to^* \mathbf{c} \ , \qquad \mathbf{a} \to^* \mathbf{d} \ , \qquad \mathbf{a} \to^* \mathbf{e} \ .$$
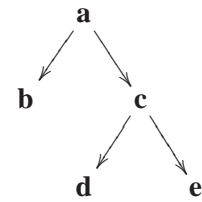
In this example, assuming equiprobability, we have $\mathrm{P}(\mathbf{a} \to^* \mathbf{b}) = \frac{1}{2}$, $\mathrm{P}(\mathbf{a} \to^* \mathbf{c}) = \frac{1}{2}$, $\mathrm{P}(\mathbf{a} \to^* \mathbf{d}) = \frac{1}{4}$, $\mathrm{P}(\mathbf{a} \to^* \mathbf{e}) = \frac{1}{4}$. Notice that these events are not disjoints and that their sum is larger than 1. In particular, $\mathbf{a} \to^* \mathbf{d}$ implies $\mathbf{a} \to^* \mathbf{c}$. Defining the elements of the set Ω of elementary events is not completely straightforward, in particular because we want to make it general enough to also consider infinite cases. For example, in the following system

$$\mathbf{a}_i \to \mathbf{a}_{i+1}, \quad \mathbf{a}_i \to \mathbf{a}'_{i+1} \ ,$$

we naturally would like that $\mathrm{P}(\mathbf{a}_0 \to^* \mathbf{a}_n) = \frac{1}{2^n}$.

Besides defining the elements of the set Ω, we need to define a notion of a measurable subset of Ω and endow such a subset with a probability distribution verifying the Kolmogorov axioms.

Our idea is to follow Lebesgue: define first the probability of rectangles, or boxes, then the probability of any set and finally measurable sets as those that verify Lebesgue's property. Thus besides defining the set Ω, we need to define a subset of $\mathcal{P}(\Omega)$ of boxes.

The first intuition would be to take paths as elements of the set Ω, for instance assigning the probability $\frac{1}{2}$ to the paths $\mathbf{a} \to \mathbf{b}$, $\frac{1}{4}$ to $\mathbf{a} \to \mathbf{c} \to \mathbf{d}$ and $\frac{1}{4}$ to $\mathbf{a} \to \mathbf{c} \to \mathbf{e}$. In fact it seems more convenient to extend such paths to strategies prescribing one reduct for each non normal object. Boxes are then defined as sets of strategies agreeing on a finite domain. We show in this paper that this is sufficient to define a probability space on strategies, consistent with the intuitive probability of events of the form $\mathbf{a} \to^* \mathbf{b}$.

Our study is generic enough to be applicable to several settings, such as automatons, or any other kind of transition systems. We use the nomenclature of abstract rewrite systems, but that of states and transitions could be used as well. Finally, we apply this construction to λ_+ [15, 16].

Plan of the paper. Section 2 introduces the basic concepts of strategies and boxes, it defines the Lebesgue measures. Section 3 proves that the space of strategies forms a probability space. Finally, in section 4 we show how to modify the calculus λ_+ into a probabilistic calculus λ_+^p. Also, we provide an encoding of an algebraic λ-calculus into λ_+^p and, to some extend, the inverse translation.

2   Preliminaries

Let Λ be a set of objects and $\to$ a function from $\Lambda \times \Lambda$ to $\mathbb{N}$ such that for all $\mathbf{a}$ the set $\{\mathbf{b} \mid \to (\mathbf{a}, \mathbf{b}) \neq 0\}$ is finite. We write $\mathbf{a} \to \mathbf{b}$ if $\to (\mathbf{a}, \mathbf{b}) \neq 0$. We allow a term to be written to the same symbol more than once, so its probability increases, e.g. if $\to (\mathbf{a}, \mathbf{b}) = 2$ and $\to (\mathbf{a}, \mathbf{c}) = 1$, then the probability of getting $\mathbf{b}$ will be the double than the probability of getting $\mathbf{c}$. Think for example in a non-deterministic choice between two objects, which happen to be equal, then there would be two ways to get such an object by doing the choice. For a given object $\mathbf{a} \in \Lambda$, we denote by $\rho(\mathbf{a})$ its degree, that is, the number of objects to which it can be rewritten to in one step. Definition 2.1 formalises this.

Definition 2.1 (Degree of an object). $\rho : \Lambda \to \mathbb{N}$ is a function defined by $\rho(\mathbf{a}) = \sum_{\mathbf{b}} \to (\mathbf{a}, \mathbf{b})$.

An object is normal if its degree is 0. We denote by $\Lambda^+ = \{\mathbf{a} \mid \mathbf{a} \in \Lambda \text{ and } \rho(\mathbf{a}) \geq 1\}$ to the set of non-normal objects, that is, objects that can be rewritten to other objects.

A strategy prescribing one reduct for each non-normal object is defined as a function from Λ^+ to Λ (cf. [23, Def. 4.9.1]).

Definition 2.2 (Strategy). A strategy is a total function $f : \Lambda^+ \to \Lambda$ such that $f(\mathbf{a}) = \mathbf{b}$ implies $\mathbf{a} \to \mathbf{b}$. For instance, if $\mathbf{a} \to \mathbf{b}$ and $\mathbf{a} \to \mathbf{b}'$, there are two functions, f and f' assigning different results to $\mathbf{a}$. We denote by Ω the set of all such functions.

A box is a set of strategies agreeing on a finite domain.

Definition 2.3 (Box). A box $B \subseteq \Omega$ is a set of the form $\{f \mid f(\mathbf{a}_1) = \mathbf{a}'_1, \ldots, f(\mathbf{a}_n) = \mathbf{a}'_n\}$ for some objects $\mathbf{a}_i$, $\mathbf{a}'_i$. We write $\mathcal{B}(\Omega)$ the subset of $\mathcal{P}(\Omega)$ containing all the boxes.

Example 2.4. Continuing with the example given at the introduction, $\Lambda^+ = \{\mathbf{a}, \mathbf{c}\}$. Let $f_1(\mathbf{a}) = \mathbf{b}, f_1(\mathbf{c}) = \mathbf{d}$ and $f_2(\mathbf{a}) = \mathbf{b}, f_2(\mathbf{c}) = \mathbf{e}$ be two of the four strategies of Ω. Then the box $\{f \mid f(\mathbf{a}) = \mathbf{b}, f(\mathbf{c}) = \mathbf{d}\} =$

$\{f_1\}$, and the box $\{f \mid f(\mathbf{a}) = \mathbf{b}\}$ is $\{f_1, f_2\}$.

A probability distribution can be defined in term of boxes, and then be extended to arbitrary sets of strategies.

Definition 2.5 (Probability function). Let $\mathtt{p} : \mathcal{B}(\Omega) \to [0,1]$ be a total function defined over boxes as follows. If $B = \{f \mid f(\mathbf{a}_1) = \mathbf{a}'_1, \ldots, f(\mathbf{a}_n) = \mathbf{a}'_n\}$, then

$$\mathtt{p}(B) = \prod_{i=1}^{n} \frac{\to (\mathbf{a}_i, \mathbf{a}'_i)}{\rho(\mathbf{a}_i)} \ .$$

By convention, if no condition is given in B (i.e. $B = \Omega$), we have $n = 0$, and we consider the product of zero elements to be 1, the neutral element of the product.

Then we define the probability measure $\mathtt{P} : \mathcal{P}(\Omega) \to [0,1]$ for arbitrary sets of strategies as follows

$$\mathtt{P}(S) = \begin{cases} 0 & \text{if } S = \emptyset \\ \inf\{\sum_{B \in \mathcal{C}} \mathtt{p}(B) \mid \mathcal{C} \text{ is a countable family of boxes s.t. } S \subseteq \bigcup_{B \in \mathcal{C}} B\} & \text{in other case} \end{cases}$$

Example 2.6. Consider the ARS $\mathbf{a} \to \mathbf{b}$ with multiplicity 2 and $\mathbf{a} \to \mathbf{c}$ with multiplicity 1.

Let B be the box $B = \{f \mid f(\mathbf{a}) = \mathbf{b}\}$. Then we have $\mathtt{p}(B) = \frac{\to(\mathbf{a},\mathbf{b})}{\rho(\mathbf{a})} = \frac{2}{3}$. Intuitively, $\mathtt{P}(B)$ is the same as $\mathtt{p}(B)$ (this will be later formalised in Lemma 3.10), because B is the minimum cover of B, that is, $\{B\}$ is the minimum family of boxes such that B is in its union. Hence $\mathtt{P}(B) = \frac{2}{3}$.
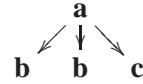
Example 2.7. We continue with the same running example depicted in the introduction. Let $f_1(\mathbf{a}) = \mathbf{b}$, $f_1(\mathbf{c}) = \mathbf{d}$ and $f_3(\mathbf{a}) = \mathbf{c}$, $f_3(\mathbf{c}) = \mathbf{e}$ be two strategies. Then the set $S = \{f_1, f_3\}$ is minimally covered by the boxes $B_1 = \{f_1\} = \{f \mid f(\mathbf{a}) = \mathbf{b}, f(\mathbf{c}) = \mathbf{d}\}$ and $B_2 = \{f_3\} = \{f \mid f(\mathbf{a}) = \mathbf{c}, f(\mathbf{c}) = \mathbf{e}\}$. So we have $P(S) = \mathtt{p}(B_1) + \mathtt{p}(B_2) = \frac{1}{2 \times 2} + \frac{1}{2 \times 2} = \frac{1}{2}$.

Now we can define the Lebesgue measure in terms of the given probability measure.

Definition 2.8 (Measurable). Let A be an element of $\mathcal{P}(\Omega)$, we write $A^{\sim}$ for the complement of A, that is $\Omega \setminus A$. The set A is Lebesgue measurable if $\forall S \in \mathcal{P}(\Omega)$, we have $\mathtt{P}(S) = \mathtt{P}(S \cap A) + \mathtt{P}(S \cap A^{\sim})$.

We define $\mathcal{A} = \{A \mid A \text{ is measurable}\}$.

3   A probability space of strategies

The aim of this section is to prove that $(\Omega, \mathcal{A}, \mathtt{P})$ is a probability space. That is, the sample space Ω (the set of all possible strategies), the set of events $\mathcal{A}$, which is the set of the Lebesgue measurable sets of

strategies, and the probability measure P, form a probability space. Our proof follows [7]. We proceed by proving that this triplet satisfies the Kolmogorov axioms, that is the probability of any event is between 0 and 1, the probability of Ω is 1, and the probability of any countable sequence of pairwise disjoint (that is incompatible) events, is the sum of their probabilities. In order to do so, we need first to prove several properties.

Lemma 3.1 establishes several known properties of Lebesgue measurable sets.

Lemma 3.1.

1. *Let $A \in \mathcal{A}$ and $S \in \mathcal{P}(\Omega)$. If $A \cap S = \emptyset$, then $P(A \cup S) = P(A) + P(S)$.*

2. *Let $A_1, A_2 \in \mathcal{A}$. If $A_1 \subseteq A_2$, then $P(A_1) \leq P(A_2)$.*

3. *$\emptyset$, the empty set, is Lebesgue measurable.*

4. *A is Lebesgue measurable if and only if $A^\sim$ is Lebesgue measurable.*

5. *If A_1, A_2 are Lebesgue measurable, then $A_1 \cup A_2$ is Lebesgue measurable.*

Proof.

1. $P(A \cup S) = P((A \cup S) \cap A) + P((A \cup S) \cap A^\sim) = P(A) + P(S)$.

2. First notice that by definition, $P(S) \geq 0$ for any $S \in \mathcal{P}(\Omega)$. Hence, $P(A_2) = P(A_2 \cap A_1) + P(A_2 \cap A_1^\sim) = P(A_1) + P(A_2 \cap A_1^\sim) \geq P(A_1)$.

3. Notice that $P(\emptyset) = 0$. Then, $\forall S \in \mathcal{P}(\Omega)$, $P(S) = P(S \cap \emptyset) + P(S \cap \Omega)$, so $\emptyset$ is Lebesgue measurable.

4. Let A be Lebesgue measurable, then $\forall S \subseteq \Omega$, $P(S) = P(S \cap A) + P(S \cap A^\sim) = P(S \cap A^\sim) + P(S \cap A^{\sim\sim})$, so $A^\sim$ is Lebesgue measurable.

5. Let A_1, A_2 be Lebesgue measurable, then $\forall S \subseteq \Omega$, we have

$$P(S) = P(S \cap A_1) + P(S \cap A_1^\sim) \quad (1) \qquad \text{and} \qquad P(S) = P(S \cap A_2) + P(S \cap A_2^\sim) \quad (2)$$

From set theory $$S \cap (A_1 \cup A_2) = S \cap (A_1 \cup (A_1^\sim \cap A_2)) = (S \cap A_1) \cup (S \cap A_1^\sim \cap A_2) \quad (3)$$

Using $S \cap A_1^\sim$ for S in (2) gives

$$P(S \cap A_1^\sim) = P(S \cap A_1^\sim \cap A_2) + P(S \cap A_1^\sim \cap A_2^\sim) = P(S \cap A_1^\sim \cap A_2) + P(S \cap (A_1 \cup A_2)^\sim) \quad (4)$$

From (3), using items 1 and 2, we have $P(S \cap (A_1 \cup A_2)) = P(S \cap A_1) + P(S \cap A_1^\sim \cap A_2)$. Adding $P(S \cap (A_1 \cup A_2)^\sim)$ to both sides gives $P(S \cap (A_1 \cup A_2) + P(S \cap (A_1 \cup A_2)^\sim) = P(S \cap A_1) + P(S \cap A_1^\sim \cap A_2) + P(S \cap (A_1 \cup A_2)^\sim)$ Using (4) and (1) we obtain $P(S \cap (A_1 \cup A_2)) + P(S \cap (A_1 \cup A_2)^\sim) = P(S \cap A_1) + P(S \cap A_1^\sim) = P(S)$. □

The concept of algebra (Definition 3.2) gives a closure property of subsets. As a corollary of the Lemma 3.1 we can show that the set $\mathcal{A}$ of Lebesgue measurable sets form an algebra (Corollary 3.3).

Definition 3.2 (Algebra). Let X be a set. We say that a set $\mathbb{A} \in \mathcal{P}(X)$ is an algebra over X if for all $A, B \in \mathbb{A}$, $A \cup B$, $A^\sim$ and X itself are also in $\mathbb{A}$.

Corollary 3.3. *$\mathcal{A}$ is an algebra over Ω.*

Proof. $\mathcal{A} \in \mathcal{P}(\Omega)$. Let $A, B \in \mathcal{A}$, then by Lemma 3.1(5), $A \cup B \in \mathcal{A}$. By Lemma 3.1(4), $A^\sim \in \mathcal{A}$. Finally, by Lemma 3.1(3) and (4), $\Omega \in \mathcal{A}$. □

Moreover, we can show that A is a σ-algebra, that is an algebra, completed to include countably infinite operations. Definition 3.4 formalises it.

Definition 3.4 (σ-algebra). Let X be a set. We say that a set $\Sigma \in \mathcal{P}(X)$ is a σ-algebra over X if it is an algebra and it is closed under countable unions, that is, if $A_1, A_2, A_3, \ldots$ are in Σ, then so is $\bigcup A_i$.

Theorem 3.7 states that the set $\mathcal{A}$ of Lebesgue measurable sets is a σ-algebra. We need to prove two properties of Lebesgue measurable sets first (Lemmas 3.5 and 3.6).

Lemma 3.5. *Let $S \subseteq \Omega$ and $A_1, \ldots, A_n \in \mathcal{A}$ be a disjoint family. Then*

$$P\left(S \cap \left(\bigcup_{i=1}^{n} A_i\right)\right) = \sum_{i=1}^{n} P(S \cap A_i) \ .$$

Proof. We proceed by induction on n. If $n = 1$ it is trivial. Assume it is true for $n-1$. Notice that

$$S \cap \left(\bigcup_{i=1}^{n} A_i\right) \cap A_n = S \cap A_n \quad (5) \qquad \text{and} \qquad S \cap \left(\bigcup_{i=1}^{n} A_i\right) \cap A_n^{\sim} = S \cap \left(\bigcup_{i=1}^{n-1} A_i\right) \quad (6)$$

Equation (5) is clear, and (6) follows since $(\bigcup_{i=1}^{n} A_i) \cap A_n^{\sim} = \bigcup_{i=1}^{n} (A_i \cap A_n^{\sim}) = (\bigcup_{i=1}^{n-1} (A_i \cap A_n^{\sim})) \cup (A_n \cap A_n^{\sim}) = \bigcup_{i=1}^{n-1} (A_i \cap A_n^{\sim})$.

Thus, since A_n is measurable, we have that

$$P\left(S \cap \left(\bigcup_{i=1}^{n} A_i\right)\right) = P\left(S \cap \left(\bigcup_{i=1}^{n} A_i\right) \cap A_n\right) + P\left(S \cap \left(\bigcup_{i=1}^{n} A_i\right) \cap A_n^{\sim}\right)$$

and from (5) and (6) this is equal to $P(S \cap A_n) + P\left(S \cap \left(\bigcup_{i=1}^{n-1} A_i\right)\right)$, which by the induction hypothesis is equal to $\sum_{i=1}^{n} P(S \cap A_i)$. □

Lemma 3.6. *Let $S_1, S_2, \cdots \subseteq \Omega$. Then*

$$P\left(\bigcup_{i=1}^{\infty} S_i\right) \leq \sum_{i=1}^{\infty} P(S_i) \ .$$

Proof. If $P(S_i) = \infty$ for some i, then we are finished. Therefore, assume $P(S_i) < \infty$ for each $i \in \mathbb{N}$.

Without lost of generality, assume $S_i \neq \emptyset$, for all i. Indeed, since $P(\emptyset) = 0$, an empty set would not add anything to any side of the equation. For a given $\varepsilon > 0$ and i, there is a sequence $\{B_{ij} \mid i = 1, \ldots, j = 1, \ldots\}$ of boxes such that $S_i \subseteq \bigcup_{j=1}^{\infty} B_{ij}$ and $\sum_{j=1}^{\infty} p(B_{ij}) < P(S_i) + 2^{-i}\varepsilon$, by the definition of P. Now, $\#\{B_{ij} \mid i, j\} \leq \aleph_0$ and $\bigcup_{i=1}^{\infty} S_i \subseteq \bigcup_{i=1}^{\infty} \bigcup_{j=1}^{\infty} B_{ij}$. Therefore, using the definition of P,

$$P\left(\bigcup_{i=1}^{\infty} S_i\right) \leq \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p(B_{ij}) \leq \sum_{i=1}^{\infty} P(S_i) + \varepsilon \sum_{i=1}^{\infty} \frac{1}{2^i} = \sum_{i=1}^{\infty} P(S_i) + \varepsilon$$

Since this is true for each ε, the lemma holds. □

Using these properties, we can prove that $\mathcal{A}$ is a σ-algebra (Theorem 3.7).

Theorem 3.7. *$\mathcal{A}$ is a σ-algebra over Ω.*

Proof. By Corollary 3.3, $\mathcal{A}$ is an algebra. We only have to prove that $\mathcal{A}$ is closed under any countable unions. That is, if $B_1, B_2, \dots \in \mathcal{A}$, then $\bigcup_{i=1}^{\infty} B_i \in \mathcal{A}$. Since $\mathcal{A}$ is an algebra (Corollary 3.3), there is a disjoint family $A_1, A_2, \dots \in \mathcal{A}$ such that $A = \bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i$. For example, we can take $A_1 = B_1$, $A_2 = B_2 \setminus B_1, A_3 = B_3 \setminus (B_1 \cup B_2), \dots$. Let $C_n = \bigcup_{i=1}^{n} A_i$, so $C_n \in \mathcal{A}$ again using that $\mathcal{A}$ is an algebra. Also notice that $A^{\sim} \subseteq C_n^{\sim}$ because $C_n \subseteq A$.

Since C_n is measurable, take any $S \subseteq \Omega$ and, using Lemma 3.1(2), we can calculate $P(S) = P(S \cap C_n) + P(S \cap C_n^{\sim}) \geq P(S \cap C_n) + P(S \cap A^{\sim})$. Since $P(S \cap C_n) = P(S \cap (\bigcup_{i=1}^{n} A_i))$, using Lemma 3.5, we obtain $P(S) \geq \sum_{i=1}^{n} P(S \cap A_i) + P(S \cap A^{\sim})$ and, since the left-hand side is independent of n, $P(S) \geq \sum_{i=1}^{\infty} P(S \cap A_i) + P(S \cap A^{\sim})$. Thus, by Lemma 3.6, $P(S) \geq P(S \cap (\bigcup_{i=1}^{\infty} A_i)) + P(S \cap A^{\sim}) = P(S \cap A) + P(S \cap A^{\sim})$.

For the converse inequality, notice that $S = (S \cap A) \cup (S \cap A^{\sim})$, so using Lemma 3.6 we have $P(S) = P((S \cap A) \cup (S \cap A^{\sim})) \leq P(S \cap A) + P(S \cap A^{\sim})$. Hence, $A \in \mathcal{A}$. $\qquad\square$

As intuited in Example 2.6, the probability of a box B is $p(B)$. Lemma 3.10 formalises it. Before proving this lemma, we need two auxiliary ones (Lemmas 3.8 and 3.9). For short, we use the notation $B \cap \mathbf{a} = \mathbf{b}$ for $B \cap \{f \mid f(\mathbf{a}) = \mathbf{b}\}$.

Lemma 3.8. *Let $N \subseteq \mathbb{N}$ and for all $i \in N$, let $B, B_i \subseteq \Omega$ be boxes s.t. $B \subseteq \bigcup_{i \in N} B_i$ and $p(B) > \sum_{i \in N} p(B_i)$. Then for every object $\mathbf{a}$, there exists an object $\mathbf{b}$ such that, $p(B \cap \mathbf{a} = \mathbf{b}) > \sum_{i \in N} p(B_i \cap \mathbf{a} = \mathbf{b})$.*

Proof. Let $\mathbf{a} \to \mathbf{b}_i$, with $i = 1, \dots, n$. Hence notice that $p(B) = \sum_{j=1}^{n} p(B \cap \mathbf{a} = \mathbf{b}_j)$, and this happens for any B. Then, from $p(B) > \sum_{i \in N} p(B_i)$, we have $\sum_{j=1}^{n} p(B \cap \mathbf{a} = \mathbf{b}_j) > \sum_{i \in N} \sum_{j=1}^{n} p(B_i \cap \mathbf{a} = \mathbf{b}_j)$ $= \sum_{j=1}^{n} \sum_{i \in N} p(B_i \cap \mathbf{a} = \mathbf{b}_j)$. Therefore, there must be at least one h such that $p(B \cap \mathbf{a} = \mathbf{b}_h) > \sum_{i \in N} p(B_i \cap \mathbf{a} = \mathbf{b}_h)$. $\qquad\square$

Lemma 3.9. *Let $N \subseteq \mathbb{N}$ and for all $i \in N$, let $B, B_i \subseteq \Omega$ be boxes s.t. $B \subseteq \bigcup_{i \in N} B_i$ and $p(B) > \sum_{i \in N} p(B_i)$. Then for all family $\{\mathbf{a}_j\}$ of objects, there exists a family $\{\mathbf{b}_j\}$ such that, for every k, $p(B \cap \mathbf{a}_1 = \mathbf{b}_1 \cap \dots \cap \mathbf{a}_k = \mathbf{b}_k) > \sum_{i \in N} p(B_i \cap \mathbf{a}_1 = \mathbf{b}_1 \cap \dots \cap \mathbf{a}_k = \mathbf{b}_k)$.*

Proof. We proceed by induction on k. For $k = 1$, use Lemma 3.8. By the induction hypothesis, we have $p(B \cap \mathbf{a}_1 = \mathbf{b}_1 \cap \dots \cap \mathbf{a}_{k-1} = \mathbf{b}_{k-1}) > \sum_{i \in N} p(B_i \cap \mathbf{a}_1 = \mathbf{b}_1 \cap \dots \cap \mathbf{a}_{k-1} = \mathbf{b}_{k-1})$. We conclude by Lemma 3.8. $\qquad\square$

Lemma 3.10. *Let $B \subseteq \Omega$ be a box, then $P(B) = p(B)$.*

Proof. Let $B = \{f \mid f(\mathbf{a}_1) = \mathbf{a}_1', \dots, f(\mathbf{a}_n) = \mathbf{a}_n'\}$. Since $B \subseteq B$, by definition of P, we have $P(B) \leq p(B)$. We must prove $p(B) \leq P(B) = \inf\{\sum_{i \in N} p(B_i) \mid B \subseteq \bigcup_{i \in N} B_i\}$. In other words, we must prove that $B \subseteq \bigcup_{i \in N} B_i$ implies $p(B) \leq \sum_{i \in N} p(B_i)$. We proceed by induction on n.

- If $n = 0$, $p(B) = 1$. Notice that, without restrictions in B, $B = \Omega$. We prove this case by contradiction. Let $p(F) > \sum_{i \in N} p(B_i)$. Then by Lemma 3.9, there exists g such that for all k,

$$p(\mathbf{a}_1 = g(\mathbf{a}_1) \cap \dots \cap \mathbf{a}_k = g(\mathbf{a}_k)) > \sum_{i \in N} p(B_i \cap \mathbf{a}_1 = g(\mathbf{a}_1) \cap \dots \cap \mathbf{a}_k = g(\mathbf{a}_k)) \qquad (7)$$

  Since $g \in \Omega \subseteq \bigcup_{i \in N} B_i$, there exists j such that $g \in B_j$. Let B_j be defined with constraints on objects $\mathbf{a}_{j_1}, \dots, \mathbf{a}_{j_q}$. Let $k = q$ and from equation (7),

$$p(\mathbf{a}_1 = g(\mathbf{a}_1) \cap \dots \cap \mathbf{a}_q = g(\mathbf{a}_q)) > \sum_{i \in N} p(B_i \cap \mathbf{a}_1 = g(\mathbf{a}_1) \cap \dots \cap \mathbf{a}_q = g(\mathbf{a}_q)) \qquad (8)$$

We know that $p(\mathbf{a}_1 = g(\mathbf{a}_1) \cap \cdots \cap \mathbf{a}_q = g(\mathbf{a}_q)) = \prod_{h=1}^{q} \frac{\rightarrow(\mathbf{a}_h, g(\mathbf{a}_h))}{\rho(\mathbf{a}_h)}$, and since $g \in B_j$, we know that this is also equal to $p(B_j \cap \mathbf{a}_1 = g(\mathbf{a}_1) \cap \cdots \cap \mathbf{a}_q = g(\mathbf{a}_q))$. Hence equation (8) leads to a contradiction.

- Consider the case $n - 1$. Let $B' = \{f \mid \exists g \in B \text{ s.t. } \forall \mathbf{a} \neq \mathbf{a}_n, f(\mathbf{a}) = g(\mathbf{a})\}$. Then if $B' \subseteq \bigcup_{i \in N'} B'_i$ we have $p(B') \leq \sum_{i \in N'} p(B'_i)$. Notice that either $B'_i = B_i$ or B_i has a constraint on $\mathbf{a}_n$ and so $\frac{\rightarrow(\mathbf{a}_n, g(\mathbf{a}_n))}{\rho(\mathbf{a}_n)} p(B'_i) = p(B_i)$. In any case, $\frac{\rightarrow(\mathbf{a}_n, g(\mathbf{a}_n))}{\rho(\mathbf{a}_n)} p(B'_i) \leq p(B_i)$. Then $p(B) = \frac{\rightarrow(\mathbf{a}_n, g(\mathbf{a}_n))}{\rho(\mathbf{a}_n)} p(B') \leq \sum_{i \in N'} \frac{\rightarrow(\mathbf{a}_n, g(\mathbf{a}_n))}{\rho(\mathbf{a}_n)} p(B'_i) \leq \sum_{i \in N'} p(B_i)$. $\square$

Theorem 3.11 (Space of strategies). $(\Omega, \mathcal{A}, \mathrm{P})$ *is a probability space.*

Proof. We prove it satisfies the Kolmogorov axioms.

1$^{\text{st}}$ axiom: $\forall A \in \mathcal{A}, 0 \leq \mathrm{P}(A) \leq 1$.

Since P is defined as an inf of sums of p, and p is always positive, so P cannot be negative. By the second Kolmogorov axiom $\mathrm{P}(\Omega) = 1$. Notice that A is measurable and $A \subseteq \Omega$, so $1 = \mathrm{P}(\Omega) = \mathrm{P}(\Omega \cap A) + \mathrm{P}(\Omega \setminus A) = \mathrm{P}(A) + \mathrm{P}(\Omega \setminus A)$, hence $1 - \mathrm{P}(\Omega \setminus A) = \mathrm{P}(A)$. Since P is not negative, $\mathrm{P}(A) \leq 1$.

2$^{\text{nd}}$ axiom: $\mathrm{P}(\Omega) = 1$.

Notice that Ω is the box including all the functions. Hence, there is no condition on the functions and so $n = 0$. Then $p(\Omega) = 1$. By Lemma 3.10, $\mathrm{P}(\Omega) = p(\Omega) = 1$.

3$^{\text{rd}}$ axiom: Any countable sequence of pairwise disjoint (i.e. incompatible) events $A_1, A_2 \cdots \in \mathcal{A}$, satisfies $\mathrm{P}(A_1 \cup A_2 \ldots) = \sum_{i=1}^{\infty} \mathrm{P}(A_i)$.
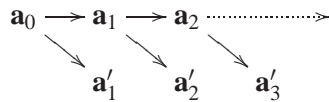
Let $\emptyset \neq I \subsetneq \mathbb{N}$. Since the sets A_i are in $\mathcal{A}$, consider $n \in \mathbb{N} \setminus I$ and we have

$$\mathrm{P}\left(\bigcup_{i \in \mathbb{N} \setminus I} A_i\right) = \mathrm{P}\left(\left(\bigcup_{i \in \mathbb{N} \setminus I} A_i\right) \cap A_n\right) + \mathrm{P}\left(\left(\bigcup_{i \in \mathbb{N} \setminus I} A_i\right) \cap A_n^{\sim}\right)$$

Notice that $\left(\bigcup_{i \in \mathbb{N} \setminus I} A_i\right) \cap A_n = A_n$ and since the A_i's are pairwise disjoint $\left(\bigcup_{i \in \mathbb{N} \setminus I} A_i\right) \cap A_n^{\sim} = \bigcup_{i \in \mathbb{N} \setminus (I \cup \{n\})} A_i$. Therefore, considering that this is valid for any I and $n \notin I$, we have

$$\mathrm{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \mathrm{P}(A_1) + \mathrm{P}\left(\bigcup_{i=2}^{\infty} A_i\right) = \mathrm{P}(A_1) + \mathrm{P}(A_2) + \mathrm{P}\left(\bigcup_{i=3}^{\infty} A_i\right) = \cdots = \sum_{i=1}^{\infty} \mathrm{P}(A_i). \quad \square$$

Example 3.12. Consider the non-strongly-normalising non-confluent rewrite system described in the introduction $\mathbf{a}_i \to \mathbf{a}_{i+1}, \quad \mathbf{a}_i \to \mathbf{a}'_{i+1}$, where each reduction is equiprobable and each symbol is different from each other. It can be depicted as follows.

$$\mathbf{a}_0 \longrightarrow \mathbf{a}_1 \longrightarrow \mathbf{a}_2 \cdots\cdots\cdots\cdots\cdots\cdots\cdots\rightarrow$$
$$\searrow \qquad \searrow \qquad \searrow$$
$$\mathbf{a}'_1 \qquad \mathbf{a}'_2 \qquad \mathbf{a}'_3$$

The probability that this rewrite system stops after exactly n steps, starting from term $\mathbf{a}_0$ is $\mathrm{P}(B)$, with $B = \{f \mid f(\mathbf{a}_0) = \mathbf{a}_1, \ldots f(\mathbf{a}_{n-2}) = \mathbf{a}_{n-1} \text{ and } f(\mathbf{a}_{n-1}) = \mathbf{a}'_n\})$, and since B is a box, by Lemma 3.10 it is the same to $\mathrm{P}(B) = p(B) = \dfrac{1}{\rho(\mathbf{a}_0) \ldots \rho(\mathbf{a}_{n-1})} = \dfrac{1}{2^n}$.

The probability of stopping at the step n or before, starting at any point before a_{n-1}, is just the probability of the box $\{f \mid f(\mathbf{a}_{n-1}) = \mathbf{a}'_n\}$, which is $\frac{1}{2}$.

The probability of stopping at the step n or m, starting at any point before $\mathbf{a}_{n-1}$ and $\mathbf{a}_{m-1}$ is the probability of the union of two boxes, however they are not independent events (its intersection is not empty). Hence let $B_1 = \{f \mid \mathbf{a}_{n-1} = \mathbf{a}'_n\}$ and $B_2 = \{f \mid \mathbf{a}_{m-1} = \mathbf{a}'_m\}$. The probability $\mathsf{P}(B_1 \cup B_2) = \mathsf{P}((B_1 \setminus B_2) \cup B_2) = \mathsf{P}(B_1 \setminus B_2) \cup \mathsf{P}(B_2) = \mathsf{P}(\{f \mid \mathbf{a}_{n-1} = \mathbf{a}'_n, \mathbf{a}_{m-1} = \mathbf{a}'_m) + \mathsf{P}(B_2) = \frac{1}{4} + \frac{1}{2} = \frac{3}{4}$.

Finally, the probability of not stopping at all, is the probability of the set $S = \{f \mid f(\mathbf{a}_i) = \mathbf{a}_{i+1}$ for $i \in \mathbb{N}\}$, which is not a box, since there is an infinite number of conditions. It is easy to check that we need an infinite number of boxes to cover such a set, however we can chose boxes as small as we want (that is, with a big number of conditions), which makes the infimum of their sums to be 0, and so the probability of not stopping is, as expected, 0.

In other words, $\mathsf{P}(S) \le \{f \mid f(\mathbf{a}_i) = \mathbf{a}_{i+1}, i \in [0,n]\} = \frac{1}{2^n}$, for any n. Hence when n tends to ∞, $\mathsf{P}(S)$ tends to 0.

4 Transforming a non-deterministic into a probabilistic calculus

4.1 The calculus λ_+

In [15, 16] we have introduced a non-deterministic calculus called λ_+, which is a simplification of an earlier probabilistic calculus by keeping non-determinism but removing explicit probabilities. Now we can transform this calculus into a probabilistic one.

The full calculus is depicted in Table 1. Typing judgements are of the form $\mathbf{r} : A$. A term $\mathbf{r}$ is typable if there exists a type A such that $\mathbf{r} : A$. Following [19, 22], we use a presentation of typed lambda-calculus without contexts and where each variable occurrence is labelled by its type, such as. $\lambda x^A.x^A$ or $\lambda x^A.y^B$. We sometimes omit the labels when they are clear from the context and write, for example, $\lambda x^A.x$ for $\lambda x^A.x^A$. We use different letters for different variables and the type system forbids terms such as $\lambda x^A.x^B$ when A and B are different, by imposing preconditions to when the typing rules apply. Let $S = \{x_1^{A_1}, \ldots, x_n^{A_n}\}$ be a set of declarations, we write S^f when this set is functional, that is when $x_i = x_j$ implies $A_i = A_j$. For example $\{x^A, y^{A \Rightarrow B}\}^f$, but not $\{x^A, x^{A \Rightarrow B}\}^f$. Typing rules have the following structure:

$$[\text{Preconditions}] \frac{\text{Hypotheses}}{\text{Derived judgement}} \ (\text{Rule name})$$

The α-conversion and the sets $FV(\mathbf{r})$ of free variables of $\mathbf{r}$ and $FV(A)$ of free variables of A are defined as usual in the λ-calculus (cf. [6, §2.1]). For example $FV(x^A y^B) = \{x^A, y^B\}$. We say that a term $\mathbf{r}$ is closed whenever $FV(\mathbf{r}) = \emptyset$. If $FV(\mathbf{r}) = \{x_1^{A_1}, \ldots, x_n^{A_n}\}$, we write $\Gamma(\mathbf{r}) = \{A_1, \ldots, A_n\}$. $FV(\{A_1, \ldots, A_n\})$ is defined by $\bigcup_{i=1}^{n} FV(A_i)$. Given two terms $\mathbf{r}$ and $\mathbf{s}$ we denote by $\mathbf{r}[\mathbf{s}/x]$ the term obtained by simultaneously substituting the term $\mathbf{s}$ for all the free occurrences of x in $\mathbf{r}$, subject to the usual proviso about renaming bound variables in $\mathbf{r}$ to avoid capture of the free variables of $\mathbf{s}$. Analogously $A[B/X]$ denotes the substitution of the type B for all the free occurrences of X in A, and $\mathbf{r}[B/X]$ the substitution in $\mathbf{r}$. For example, $(x^A)[B/Y] = x^{(A[B/Y])}$, $(\lambda x^A.\mathbf{r})[B/X] = \lambda x^{A[B/X]}.\mathbf{r}[B/X]$ and $(\pi_A(\mathbf{r}))[B/X] = \pi_{A[B/X]}(\mathbf{r}[B/X])$. Simultaneous substitutions are defined in the same way. Finally, terms and types are considered up to α-conversion.

Each term of the language has a main type associated, which can be obtained from the type annotations, and other types induced by the type equivalences.

The operational semantics of λ_+ is also given in Table 1, where there are two distinct relations between terms: a symmetric relation $\rightleftarrows$ and a reduction relation $\hookrightarrow$. We write $\rightleftarrows^*$ and $\hookrightarrow^*$ for the transitive and reflexive closures of $\rightleftarrows$ and $\hookrightarrow$ respectively. In particular, notice that $\rightleftarrows^*$ is an equivalence relation. We just write $\rightarrow$ when we do not want to make the distinction between these relations. We write $n.\mathbf{r}$ in λ_+ as a shorthand for $\underbrace{\mathbf{r}+\cdots+\mathbf{r}}_{n\text{ times}}$.

This calculus has a non-deterministic projector. Indeed, the rule "If $\mathbf{r}:A$, then $\pi_A(\mathbf{r}+\mathbf{s})\hookrightarrow\mathbf{r}$" is not-deterministic because the symbol $+$ is commutative, so if $\mathbf{s}:A$, this rule can produce either $\mathbf{r}$ or $\mathbf{s}$ non-deterministically. In any case, both reducts are valid proofs of A, and so the proof system is consistent. Refer to [15] for details.

Grammar of types and terms

$$A,B,C,\ldots \;::=\; X \mid A \Rightarrow B \mid A \wedge B \mid \forall X.A \ .$$

$$\mathbf{r},\mathbf{s},\mathbf{t} \;::=\; x^A \mid \lambda x^A.\mathbf{r} \mid \mathbf{rs} \mid \mathbf{r}+\mathbf{s} \mid \pi_A(\mathbf{r}) \mid \Lambda X.\mathbf{r} \mid \mathbf{r}\{A\} \ .$$

Equivalence between types

$$A\wedge B \equiv B\wedge A \ , \qquad (A\wedge B)\wedge C \equiv A\wedge(B\wedge C) \ , \qquad A\Rightarrow(B\wedge C) \equiv (A\Rightarrow B)\wedge(A\Rightarrow C) \ .$$

Rewriting system

Symmetric relation:

$$\mathbf{r}+\mathbf{s}\rightleftarrows\mathbf{s}+\mathbf{r} \ , \qquad (\mathbf{r}+\mathbf{s})\mathbf{t}\rightleftarrows\mathbf{rt}+\mathbf{st} \ , \qquad \text{If } \mathbf{r}:A\Rightarrow(B\wedge C), \text{ then}$$
$$(\mathbf{r}+\mathbf{s})+\mathbf{t}\rightleftarrows\mathbf{r}+(\mathbf{s}+\mathbf{t}) \ , \qquad \lambda x^A.(\mathbf{r}+\mathbf{s})\rightleftarrows\lambda x^A.\mathbf{r}+\lambda x^A.\mathbf{s} \ , \qquad \pi_{A\Rightarrow B}(\mathbf{r})\mathbf{s}\rightleftarrows\pi_B(\mathbf{rs}) \ .$$

Reductions:

$$(\lambda x^A.\mathbf{r})\,\mathbf{s}\hookrightarrow\mathbf{r}[\mathbf{s}/x] \ , \qquad (\Lambda X.\mathbf{r})\{A\}\hookrightarrow\mathbf{r}[A/X] \ , \qquad \text{If } \mathbf{r}:A, \text{ then } \pi_A(\mathbf{r}+\mathbf{s})\hookrightarrow\mathbf{r} \ .$$

Typing system

$$[A\equiv B]\frac{\mathbf{r}:A}{\mathbf{r}:B}\;(\equiv) \qquad \frac{}{x^A:A}\;(ax) \qquad [(FV(\mathbf{r})\cup\{x^A\})^f]\frac{\mathbf{r}:B}{\lambda x^A.\mathbf{r}:A\Rightarrow B}\;(\Rightarrow_i) \qquad [FV(\mathbf{rs})^f]\frac{\mathbf{r}:A\Rightarrow B \quad \mathbf{s}:A}{\mathbf{rs}:B}\;(\Rightarrow_e)$$

$$[FV(\mathbf{r}+\mathbf{s})^f]\frac{\mathbf{r}:A \quad \mathbf{s}:B}{\mathbf{r}+\mathbf{s}:A\wedge B}\;(\wedge_i) \qquad \frac{\mathbf{r}:A\wedge B}{\pi_A(\mathbf{r}):A}\;(\wedge_e) \qquad [X\notin FV(\Gamma(\mathbf{r}))]\frac{\mathbf{r}:A}{\Lambda X.\mathbf{r}:\forall X.A}\;(\forall_i) \qquad \frac{\mathbf{r}:\forall X.A}{\mathbf{r}\{B\}:A[B/X]}\;(\forall_e)$$

Table 1: The non-deterministic calculus λ_+

4.2   From non-determinism to probabilities (or from λ_+ to λ_+^p)

Consider the following example (cf. [16, Example 5]). Two possible reduction paths can be fired from $(\Lambda X.(\pi_A(x^A+y^X)))\{A\}$: Reducing first the projection, $(\Lambda X.x^A)\{A\}\hookrightarrow x^A$, or reducing first the beta $\pi_A(x^A+y^A)\hookrightarrow x^A$. The former path is deterministic and will always reduce to x^A, on the contrary, the latter can non-deterministically chose between x^A and y^A. However, in both cases a proof of A is obtained.

Hence, the non-determinism is present not only due to the projector, but also by a combination of not defining a reduction strategy and the polymorphism, which can turn a deterministic projection into a non-deterministic one. We want to associate a probability to the second case, that is, to the non-deterministic projector (the π reduction). With this aim, we consider the following ARS, called $\lambda_+^\downarrow$. The closed normal terms of λ_+ are objects of $\lambda_+^\downarrow$. If $\mathbf{r}_1,\ldots,\mathbf{r}_n$ are objects, then it is also an object. The function $\rightarrow$ is

given by the relations $\rightleftarrows$ and $\hookrightarrow$. In particular, if $\mathbf{r} : A$, then $\pi_A(\mathbf{r}+\mathbf{r}) \to \mathbf{r}$, with multiplicity 2, i.e. $\to (\pi_A(\mathbf{r}+\mathbf{r}),\mathbf{r}) = 2$.

Theorem 4.1. *Let* $(\Omega,\mathcal{A},\mathrm{P})$ *be a probability space over* $\lambda_+^\downarrow$. *Let* $B_{\mathbf{r}_i} = \{f \mid f(\pi_A(\sum_{j=1}^n m_j.\mathbf{r}_j)) = \mathbf{r}_i\}$ *be a box. Then* $\mathrm{P}(B_{\mathbf{r}_i}) = \frac{m_i}{\sum_{j=1}^n m_j}$.

Proof. Notice that $\rho(\pi_A(\sum_{i=1}^n m_i.\mathbf{r}_i)) = \sum_{\mathbf{r}} \to (\pi_A(\sum_{i=1}^n m_i.\mathbf{r}_i),\mathbf{r}) = \sharp[\overbrace{\mathbf{r}_1,\cdots,\mathbf{r}_1}^{m_1\ \text{times}},\ldots,\overbrace{\mathbf{r}_n,\cdots,\mathbf{r}_n}^{m_n\ \text{times}}] = \sum_{j=1}^n m_j$
And $\to (\pi_A(\sum_{i=1}^n m_i.\mathbf{r}_i),\mathbf{r}_i) = m_i$. Hence, $\mathrm{P}(B_{\mathbf{r}_i}) = \mathrm{p}(B_{\mathbf{r}_i}) = \frac{m_i}{\sum_{j=1}^n m_j}$. $\qquad\square$

Definition 4.2 (The probabilistic calculus λ_+^p)**.** Let λ_+^p be the language of Table 1, with the following modification:
Replace rule "If $\mathbf{r} : A$, then $\pi_A(\mathbf{r}+\mathbf{s}) \hookrightarrow \mathbf{r}$" by
"For $i = 1,\ldots,n$, let $\mathbf{r}_i : A$ and $\mathbf{s} \not{/} A$, be closed normal terms. Then

$$\pi_A(\sum_{i=1}^n m_i.\mathbf{r}_i + \mathbf{s}) \hookrightarrow \mathbf{r}_i \qquad \text{with probability } \frac{m_i}{\sum_{j=1}^n m_j}\text{"} \ .$$

Remark 4.3. Notice that by Theorem 4.1 the probabilistic reduction is well defined.

4.3   The calculus $\mathbf{Alg}_F^p$

The calculus Alg_F^p is inspired from [5, 24]. We restrict the algebraic calculus to only have probabilistic superpositions, and we type it with a simple extension of System F (cf. [2, Def. 5.1]). The grammar of terms ensures that the linear combinations of terms are probability distributions, however the type system allows typing pseudo-terms, that is, terms that are not probability distributions. A term in this language, is a term produced by the grammar of terms, and typed. The full calculus is depicted in Table 2.

4.4   From $\mathbf{Alg}_F^p$ to λ_+^p

We give a translation from the probabilistic calculus Alg_F^p, including scalars, to the probabilistic calculus λ_+^p.

$$[\![x^A]\!] = x^A \qquad\qquad [\![\mathbf{rs}]\!] = [\![\mathbf{r}]\!][\![\mathbf{s}]\!] \qquad\qquad [\![\mathbf{r}\{A\}]\!] = [\![\mathbf{r}]\!]\{A\}$$
$$[\![\lambda x^A.\mathbf{r}]\!] = \lambda x^A.[\![\mathbf{r}]\!] \qquad [\![\Lambda X.\mathbf{r}]\!] = \Lambda X.[\![\mathbf{r}]\!] \qquad [\![\sum_{i=1}^n \frac{n_i}{d_i}.\mathbf{r}_i]\!] = \pi_A(\sum_{i=1}^n m_i.[\![\mathbf{r}_i]\!])$$
$$\text{where } \mathbf{r}_i : A, d_i \in \mathbb{N}^*, m_i = n_i(\prod_{\substack{k=1 \\ k \neq i}}^n d_k), \text{ for } i = 1,\ldots,n.$$

Example 4.4. Let $\mathbf{r} : A$, $\mathbf{t} : A$ and $\mathbf{s} : A$. $[\![\frac{3}{4}.\mathbf{r}+\frac{1}{8}.\mathbf{t}+\frac{1}{8}.\mathbf{s}]\!] = \pi_A(192.[\![\mathbf{r}]\!]+32.[\![\mathbf{t}]\!]+32.[\![\mathbf{s}]\!])$. By Theorem 4.1, this last term reduces to $[\![\mathbf{r}]\!]$ with probability $\frac{192}{192+32+32} = \frac{3}{4}$, to $[\![\mathbf{t}]\!]$ with probability $\frac{32}{192+32+32} = \frac{1}{8}$, and to $[\![\mathbf{s}]\!]$ with probability $\frac{32}{192+32+32} = \frac{1}{8}$.

Lemma 4.5.

  1. $[\![\mathbf{r}]\!][A/X] = [\![\mathbf{r}[A/X]]\!]$. $\qquad\qquad\qquad$ 2. $[\![\mathbf{r}]\!][[\![\mathbf{s}]\!]/x] = [\![\mathbf{r}[\mathbf{s}/x]]\!]$.

Proof.

  1. We proceed by induction on $\mathbf{r}$.
     - Let $\mathbf{r} = x^B$. $[\![x^B]\!][A/X] = x^B[A/X] = x^{B[A/X]} = [\![x^{B[A/X]}]\!] = [\![x^B[A/X]]\!]$.

<div style="border:1px solid black">

Grammar of types

$$A, B, C, \ldots ::= \quad X \mid A \Rightarrow B \mid \forall X.A \ .$$

Grammar of pseudo-terms

$$\mathbf{r}, \mathbf{s}, \mathbf{t} ::= \quad x^A \mid \lambda x^A.\mathbf{r} \mid \mathbf{rs} \mid \Lambda X.\mathbf{r} \mid \mathbf{r}\{A\} \mid p.\mathbf{r} \mid \mathbf{r} + \mathbf{s}$$

Grammar of terms

$$\mathbf{r}, \mathbf{s}, \mathbf{t} ::= \quad x^A \mid \lambda x^A.\mathbf{r} \mid \mathbf{rs} \mid \Lambda X.\mathbf{r} \mid \mathbf{r}\{A\} \mid \sum_{i=1}^{n} p_i.\mathbf{r}_i \qquad \text{with} \begin{cases} n > 0, \\ p_i \in \mathbb{Q}(0,1] \text{ and} \\ \sum_{i=1}^{n} p_i = 1 \end{cases}$$

Rewriting system
Symmetric relation:

$$\mathbf{r} + \mathbf{s} \rightleftarrows \mathbf{s} + \mathbf{r} \ , \qquad\qquad (\mathbf{r} + \mathbf{s})\mathbf{t} \rightleftarrows \mathbf{rt} + \mathbf{st} \ , \qquad\qquad 1.\mathbf{r} \rightleftarrows \mathbf{r} \ .$$
$$(\mathbf{r} + \mathbf{s}) + \mathbf{t} \rightleftarrows \mathbf{r} + (\mathbf{s} + \mathbf{t}) \ , \qquad \lambda x^A.(\mathbf{r} + \mathbf{s}) \rightleftarrows \lambda x^A.\mathbf{r} + \lambda x^A.\mathbf{s} \ ,$$

Reductions:

Beta	Elementary	Factorisation
$(\lambda x^A.\mathbf{r}) \, \mathbf{s} \hookrightarrow \mathbf{r}[\mathbf{s}/x]$,	$p.q.\mathbf{r} \hookrightarrow pq.\mathbf{r}$,	$p.\mathbf{r} + q.\mathbf{r} \hookrightarrow (p+q).\mathbf{r}$.
$(\Lambda X.\mathbf{r})\{A\} \hookrightarrow \mathbf{r}[A/X]$,	$p.(\mathbf{r} + \mathbf{s}) \hookrightarrow p.\mathbf{r} + p.\mathbf{s}$,	

Typing system

$$\frac{}{x^A : A} \, (ax) \qquad [(FV(\mathbf{r}) \cup \{x^A\})^f] \frac{\mathbf{r} : B}{\lambda x^A.\mathbf{r} : A \Rightarrow B} \, (\Rightarrow_i) \qquad [FV(\mathbf{rs})^f] \frac{\mathbf{r} : A \Rightarrow B \quad \mathbf{s} : A}{\mathbf{rs} : B} \, (\Rightarrow_e)$$

$$[FV(\mathbf{r}+\mathbf{s})^f] \frac{\mathbf{r} : A \quad \mathbf{s} : A}{\mathbf{r} + \mathbf{s} : A} \, (+_i) \qquad \frac{\mathbf{r} : A}{p.\mathbf{r} : A} \, (p_i) \qquad [X \notin FV(\Gamma(\mathbf{r}))] \frac{\mathbf{r} : A}{\Lambda X.\mathbf{r} : \forall X.A} \, (\forall_i) \qquad \frac{\mathbf{r} : \forall X.A}{\mathbf{r}\{B\} : A[B/X]} \, (\forall_e)$$

</div>

Table 2: The algebraic calculus Alg_F^p.

- Let $\mathbf{r} = \lambda x^B.\mathbf{t}$. $[\![\lambda x^B.\mathbf{t}]\!][A/X] = \lambda x^B.[\![\mathbf{t}]\!][A/X] = \lambda x^{B[A/X]}.[\![\mathbf{t}]\!][A/X] \stackrel{IH}{=} \lambda x^{B[A/X]}.[\![\mathbf{t}[A/X]]\!] = [\![\lambda x^{B[A/X]}.\mathbf{t}[A/X]]\!] = [\![(\lambda x^B.\mathbf{t})[A/X]]\!]$.

- Let $\mathbf{r} = \mathbf{t}_1\mathbf{t}_2$. $[\![\mathbf{t}_1\mathbf{t}_2]\!][A/X] = [\![\mathbf{t}_1]\!][A/X][\![\mathbf{t}_2]\!][A/X] \stackrel{IH}{=} [\![\mathbf{t}_1[A/X]]\!][\![\mathbf{t}_2[A/X]]\!] = [\![\mathbf{t}_1[A/X]\mathbf{t}_2[A/X]]\!] = [\![(\mathbf{t}_1\mathbf{t}_2)[A/X]]\!]$.

- Let $\mathbf{r} = \Lambda Y.\mathbf{t}$, with $Y \notin FV(A)$. $[\![\Lambda Y.\mathbf{t}]\!][A/X] = \Lambda Y.[\![\mathbf{t}]\!][A/X] \stackrel{IH}{=} \Lambda Y.[\![\mathbf{t}[A/X]]\!] = [\![\Lambda Y.\mathbf{t}[A/X]]\!] = [\![(\Lambda Y.\mathbf{t})[A/X]]\!]$.

- Let $\mathbf{r} = \mathbf{t}\{B\}$. $[\![\mathbf{t}\{B\}]\!][A/X] = [\![\mathbf{t}]\!]\{B\}[A/X] = [\![\mathbf{t}]\!][A/X]\{B[A/X]\} \stackrel{IH}{=} [\![\mathbf{t}[A/X]]\!]\{B[A/X]\} = [\![\mathbf{t}[A/X]\{B[A/X]\}]\!] = [\![(\mathbf{t}\{B\})[A/X]]\!]$.

- Let $\mathbf{r} = \sum_{i=1}^{n} \frac{n_i}{d_i}.\mathbf{r}_i$. $[\![\sum_{i=1}^{n} \frac{n_i}{d_i}.\mathbf{r}_i]\!][A/X] = \pi_A \left(\sum_{i=1}^{n} m_i.[\![\mathbf{r}_i]\!]\right)[A/X] = \pi_A \left(\sum_{i=1}^{n} m_i.[\![\mathbf{r}_i]\!][A/X]\right) \stackrel{IH}{=} \pi_A \left(\sum_{i=1}^{n} m_i[\![\mathbf{r}_i[A/X]]\!]\right) = [\![\sum_{i=1}^{n} \frac{n_i}{d_i}.\mathbf{r}_i[A/X]]\!] = [\![(\sum_{i=1}^{n} \frac{n_i}{d_i}.\mathbf{r}_i)[A/X]]\!]$.

2. We proceed by induction on $\mathbf{r}$.

- Let $\mathbf{r} = x^A$. $[\![x^A]\!][[\![\mathbf{s}]\!]/x] = x^A[[\![\mathbf{s}]\!]/x] = [\![\mathbf{s}]\!] = [\![x^A[\mathbf{s}/x]]\!]$.
- Let $\mathbf{r} = y^A$, $[\![y^A]\!][[\![\mathbf{s}]\!]/x] = y^A[[\![\mathbf{s}]\!]/x] = y^A = [\![y^A]\!] = [\![y^A[\mathbf{s}/x]]\!]$.
- Let $\mathbf{r} = \lambda y^B.\mathbf{t}$. $[\![\lambda y^B.\mathbf{t}]\!][[\![\mathbf{s}]\!]/x] = \lambda y^B.[\![\mathbf{t}]\!][[\![\mathbf{s}]\!]/x] \stackrel{IH}{=}$
$$\lambda y^B.[\![\mathbf{t}[\mathbf{s}/x]]\!] = [\![\lambda y^B.\mathbf{t}[\mathbf{s}/x]]\!] = [\![(\lambda y^B.\mathbf{t})[\mathbf{s}/x]]\!].$$

- Let $\mathbf{r} = \mathbf{t}_1\mathbf{t}_2$. $[\![\mathbf{t}_1\mathbf{t}_2]\!][[\![\mathbf{s}]\!]/x] = [\![\mathbf{t}_1]\!][[\![\mathbf{s}]\!]/x][\![\mathbf{t}_2]\!][[\![\mathbf{s}]\!]/x] \overset{IH}{=} [\![\mathbf{t}_1[\mathbf{s}/x]]\!][\![\mathbf{t}_2[\mathbf{s}/x]]\!] = [\![\mathbf{t}_1[\mathbf{s}/x]\mathbf{t}_2[\mathbf{s}/x]]\!] = [\![(\mathbf{t}_1\mathbf{t}_2)[\mathbf{s}/x]]\!]$.

- Let $\mathbf{r} = \Lambda X.\mathbf{t}$, $[\![\Lambda X.\mathbf{t}]\!][[\![\mathbf{s}]\!]/x] = \Lambda X.[\![\mathbf{t}]\!][[\![\mathbf{s}]\!]/x] \overset{IH}{=} \Lambda X.[\![\mathbf{t}[\mathbf{s}/x]]\!] = [\![\Lambda X.\mathbf{t}[\mathbf{s}/x]]\!] = [\![(\Lambda X.\mathbf{t})[\mathbf{s}/x]]\!]$.

- Let $\mathbf{r} = \mathbf{t}\{B\}$. Let $FV(\mathbf{s}) = \vec{X}$ and $\vec{Y}$ be a set of free variables such that $\mathbf{s}[\vec{Y}/\vec{X}][\vec{X}/\vec{Y}] = \mathbf{s}$.
  Then, $[\![\mathbf{t}\{B\}]\!][[\![\mathbf{s}]\!]/x] = [\![\mathbf{t}]\!]\{B\}[[\![\mathbf{s}]\!]/x] = [\![\mathbf{t}]\!][[\![\mathbf{s}[\vec{Y}/\vec{X}]/x]\!]\{B\}[\vec{X}/\vec{Y}] \overset{IH}{=} [\![\mathbf{t}[\mathbf{s}[\vec{Y}/\vec{X}]/x]]\!]\{B\}[\vec{X}/\vec{Y}]$
  $= [\![\mathbf{t}[\mathbf{s}[\vec{Y}/\vec{X}]/x]\{B\}]\!][\vec{X}/\vec{Y}] \overset{item\ 1}{=} [\![\mathbf{t}[\mathbf{s}[\vec{Y}/\vec{X}]/x]\{B\}[\vec{X}/\vec{Y}]]\!] = [\![(\mathbf{t}\{B\})[\mathbf{s}/x]]\!]$.

- Let $\mathbf{r} = \sum_{i=1}^n \frac{n_i}{d_i}.\mathbf{r}_i$. $[\![\sum_{i=1}^n \frac{n_i}{d_i}.\mathbf{r}_i]\!][A/X] = \pi_A\left(\sum_{i=1}^n m_i.[\![\mathbf{r}_i]\!]\right)[A/X] = \pi_A\left(\sum_{i=1}^n m_i.[\![\mathbf{r}_i]\!][A/X]\right) \overset{IH}{=}$
  $\pi_A\left(\sum_{i=1}^n m_i[\![\mathbf{r}_i[A/X]]\!]\right) = [\![\sum_{i=1}^n \frac{n_i}{d_i}.\mathbf{r}_i[A/X]]\!] = [\![(\sum_{i=1}^n \frac{n_i}{d_i}.\mathbf{r}_i)[A/X]]\!]$. $\qquad\qquad\square$

Theorem 4.6. *If $\mathbf{r} \to^* \sum_{i=1}^n p_i.\mathbf{t}_i$, with $\mathbf{t}_i$ in Alg_F^p, with $\sum_{i=1}^n p_i = 1$ and $[\![\mathbf{t}_i]\!] \to^* \mathbf{s}_i$, then $[\![\mathbf{r}]\!] \to^* \mathbf{s}_i$ with probability $p_i\left(\sum_{j=1}^n p_j\right)^{-1}$ in λ_+^p.*

Proof. Let $\mathbf{r} : A$ in Alg_F^p. For $i = 1, \ldots, n$, assume $p_i = \dfrac{n_i}{d_i}$ with $n_i, d_i \in \mathbb{N}^*$. We proceed by a case analysis on the last reduction step to reach $\sum_{i=1}^n p_i.\mathbf{t}_i$.

- If $\mathbf{r} = \sum_{i=1}^n p_i.\mathbf{t}_i$, then $\pi_A(\sum_{i=1}^n (\prod_{\substack{k=1 \\ k \neq i}}^n d_k n_i).[\![\mathbf{t}_i]\!]) \to^* \pi_A(\sum_{i=1}^n (\prod_{\substack{k=1 \\ k \neq i}}^n d_k n_i).\mathbf{s}_i')$ By Theorem 4.1, this
  term reduces in one step to $\mathbf{s}_i'$ with probability

$$
\frac{\prod_{\substack{k=1 \\ k \neq i}}^n d_k n_i}{\sum_{i=1}^n \left(\prod_{\substack{k=1 \\ k \neq i}}^n d_k n_i\right)} = \left(\frac{\frac{n_i}{d_i}}{\sum_{i=1}^n \frac{n_i}{d_i}}\right) \cdot \left(\frac{\prod_{k=1}^n d_k}{\prod_{k=1}^n d_k}\right) = p_i \left(\sum_{j=1}^n p_j\right)^{-1} .
$$

- Consider $1.\mathbf{r} \rightleftarrows \mathbf{r}$, with $\mathbf{r} = \sum_{i=1}^n p_i.\mathbf{t}_i$. We have, $[\![1.\mathbf{r}]\!] = \pi_A(1.[\![\mathbf{r}]\!]) \to^* \pi_A(1.\mathbf{s})$, which reduces with probability one to $\mathbf{s}$. Notice that $\mathbf{s}$ is a reduct of $[\![\sum_{i=1}^n p_i.\mathbf{t}_i]\!] = \pi_A(\sum_{i=1}^n m_i.[\![\mathbf{t}_1]\!])$. We conclude with Theorem 4.1.

- Consider $\left(\sum_{i=m+1}^n p_i.\mathbf{t}_i\right) + \left(\sum_{i=1}^m p_i.\mathbf{t}_i\right) \rightleftarrows \sum_{i=1}^n p_i.\mathbf{t}_i$, with $1 \leq m < n$. Since $\mathbf{r} : A$, then each $\mathbf{t}_i : A$. We have,

$$
[\![\sum_{i=m+1}^n p_i.\mathbf{t}_i + \sum_{i=1}^m p_i.\mathbf{t}_i]\!] = \pi_A\left(\sum_{i=m+1}^n m_i.[\![\mathbf{t}_i]\!] + \sum_{i=1}^m m_i.[\![\mathbf{t}_i]\!]\right) \rightleftarrows \pi_A\left(\sum_{i=1}^n m_i.[\![\mathbf{t}_i]\!]\right) .
$$

  where $m_i = \prod_{\substack{k=1 \\ k \neq i}}^n d_k n_i$. We conclude with Theorem 4.1.

- Consider $\lambda x^A.(\mathbf{r} + \mathbf{s}) \rightleftarrows \lambda x^A.\mathbf{r} + \lambda x^A.\mathbf{s}$. We have $[\![\lambda x^A.(\mathbf{r} + \mathbf{s})]\!] = \lambda x^A.([\![\mathbf{r} + \mathbf{s}]\!]) = \lambda x^A.\pi_A([\![\mathbf{r}]\!] + [\![\mathbf{s}]\!]) \to^* \lambda x^A.\pi_A(\mathbf{r}' + \mathbf{s}')$ By Theorem 4.1, $\lambda x^A.\pi_A(\mathbf{r}' + \mathbf{s}')$ reduces to $\lambda x^A.\mathbf{r}'$ (which is a reduct of $[\![\lambda x^A.\mathbf{r}]\!] = \lambda x^A.[\![\mathbf{r}]\!]$), with probability $\frac{1}{2}$, and to $\lambda x^A.\mathbf{s}'$ (which is a reduct of $[\![\lambda x^A.\mathbf{s}]\!] = \lambda x^A.[\![\mathbf{s}]\!]$), with probability $\frac{1}{2}$.

- Consider $(\lambda x^A.\mathbf{r})\ \mathbf{s} \hookrightarrow \mathbf{r}[\mathbf{s}/x]$, with $\mathbf{r}[\mathbf{s}/x] = \sum_{i=1}^n p_i.\mathbf{t}_i$. Then $[\![(\lambda x^A.\mathbf{r})\ \mathbf{s}]\!] = (\lambda x^A.[\![\mathbf{r}]\!])\ [\![\mathbf{s}]\!] \hookrightarrow [\![\mathbf{r}]\!][[\![\mathbf{s}]\!]/x]$ which, by Lemma 4.5(2), is equal to $[\![\mathbf{r}[\mathbf{s}/x]]\!] = [\![\sum_{i=1}^n p_i.\mathbf{t}_i]\!]$ and this, by definition is equal to $\pi_A\left(\sum_{i=1}^n (\prod_{\substack{k=1 \\ k \neq i}}^n d_k n_i).[\![\mathbf{t}_i]\!]\right)$. We conclude with Theorem 4.1.

- Consider $(\Lambda X.\mathbf{r})\{A\} \hookrightarrow \mathbf{r}[A/X]$, with $\mathbf{r}[A/X] = \sum_{i=1}^n p_i.\mathbf{t}_i$. Then, $[\![\Lambda X.\mathbf{r}\{A\}]\!] = \Lambda X.[\![\mathbf{r}]\!]\{A\} \hookrightarrow [\![\mathbf{r}]\!][A/X]$, which by Lemma 4.5(1) is equal to $[\![\mathbf{r}[A/X]]\!] = [\![\sum_{i=1}^n p_i.\mathbf{t}_i]\!] = \pi_A\left(\sum_{i=1}^n (\prod_{\substack{k=1 \\ k \neq i}}^n d_k n_i).[\![\mathbf{t}_i]\!]\right)$.
  We conclude with Theorem 4.1.

- Consider $p.q.\mathbf{r} \hookrightarrow pq.\mathbf{r}$. Let $p.q.\mathbf{r} : A$. Since $pq.\mathbf{r} = \sum_{i=1}^{n} p_i.\mathbf{t}_i$ with $\sum_{i=1}^{n} p_i = 1$, we have $n = 1$ and $pq = p_1 = 1$. Also, since $p.q.\mathbf{r}$ is a term, $p = q = 1$. So, we have $[\![1.1.\mathbf{r}]\!] = \pi_A(1.[\![1.\mathbf{r}]\!]) = \pi_A(1.\pi_A(1.[\![\mathbf{r}]\!])) \to^* \pi_A(1.\pi_A(1.\mathbf{s}))$ Notice that this term reduces with probability 1 to $\pi_A(1.\mathbf{s})$, which is a reduct of $\pi_A(1.[\![\mathbf{r}]\!]) = [\![1.\mathbf{r}]\!]$.

- Consider $p.(\mathbf{r}_1 + \mathbf{r}_2) \hookrightarrow p.\mathbf{r}_1 + p.\mathbf{r}_2$. Since $p.\mathbf{r}_1 + p.\mathbf{r}_2 = \sum_{i=1}^{n} p_i.\mathbf{t}_i$, with $\sum_{i=1}^{n} p_i = 1$, we have $n = 2$ and $p = \frac{1}{2}$, however in such case $\frac{1}{2}.(\mathbf{r}_1 + \mathbf{r}_2)$ is a pseudo-term, not a term.

- Consider $p.\mathbf{r} + q.\mathbf{r} \hookrightarrow (p+q).\mathbf{r}$. Since $(p+q).\mathbf{r} = \sum_{i=1}^{n} p_i.\mathbf{t}_i$, with $\sum_{i=1}^{n} p_i = 1$, we have $n = 1$ and $p + q = 1$. Let $p = \frac{m}{d}$, then $q = \frac{d-m}{d}$. So, $[\![p.\mathbf{r} + q.\mathbf{r}]\!] = \pi_A(dm.[\![\mathbf{r}]\!] + (d(d-m)).[\![\mathbf{r}]\!]) \hookrightarrow \pi_A(d^2.[\![\mathbf{r}]\!])$, which reduces with probability 1 to $\mathbf{s}$, where $[\![\mathbf{r}]\!] \to^* \mathbf{s}$.

- Contextual rules are straightfoward.   □

4.5   Back from λ_+^p to $\mathbf{Alg}_F^p$

The inverse translation is given by

$$
\begin{aligned}
(\!|x^A|\!) &= x^A & (\!|\mathbf{rs}|\!) &= (\!|\mathbf{r}|\!)(\!|\mathbf{s}|\!) & (\!|\mathbf{r}\{A\}|\!) &= (\!|\mathbf{r}|\!)\{A\} \\
(\!|\lambda x^A.\mathbf{r}|\!) &= \lambda x^A.(\!|\mathbf{r}|\!) & (\!|\Lambda X.\mathbf{r}|\!) &= \Lambda X.(\!|\mathbf{r}|\!) & (\!|\mathbf{r}+\mathbf{s}|\!) &= (\!|\mathbf{r}|\!) + (\!|\mathbf{s}|\!)
\end{aligned}
$$

If $\pi_A(\mathbf{t}) \hookrightarrow \mathbf{s}_i$ with probability p_i, for $i = 1,\ldots,n$, $(\!|\pi_A(\mathbf{t})|\!) = \sum_{i=1}^{n} p_i.(\!|\mathbf{s}_i|\!)$

Remark 4.7. This translation does not admit translating a term of the form $\pi_A(\mathbf{t})$ in normal form. Moreover, let Π be the rule "$\pi_{A \Rightarrow B}(\mathbf{r})\mathbf{s} \rightleftarrows \pi_B(\mathbf{rs})$ with $\mathbf{r} : A \Rightarrow (B \wedge C)$", then the translation keep reductions, except for the one using rule Π, as expressed in Theorem 4.9.

Lemma 4.8.

1. $(\!|\mathbf{r}|\!)[A/X] = (\!|\mathbf{r}[A/X]|\!)$ 2. $(\!|\mathbf{r}|\!)[(\!|\mathbf{s}|\!)/x] = (\!|\mathbf{r}[\mathbf{s}/x]|\!)$

Proof. Both items follow by induction on $\mathbf{r}$. Cases x^B, $\lambda x^B.\mathbf{t}$, $\mathbf{t}_1\mathbf{t}_2$, $\Lambda Y.\mathbf{t}$ and $\mathbf{t}\{B\}$ are analogous to those in proof of Lemma 4.5. Hence we only need to verify the case $\pi_B(\mathbf{t})$, when $\mathbf{r} \hookrightarrow \mathbf{r}_i$ with probability p_i, for $i = 1,\ldots,n$.

1. $(\!|\pi_B(\mathbf{t})|\!)[A/X] = (\sum_{i=1}^{n} p_i.(\!|\mathbf{r}_i|\!))[A/X] = \sum_{i=1}^{n} p_i.(\!|\mathbf{r}_i|\!)[A/X]$, which by the induction hypothesis, is equal to $\sum_{i=1}^{n} p_i.(\!|\mathbf{r}_i[A/X]|\!) = (\!|\pi_{B[A/X]}(\mathbf{t}[A/X])|\!) = (\!|(\pi_B(\mathbf{t}))[A/X]|\!)$.

2. $(\!|\pi_B(\mathbf{t})|\!)[\mathbf{s}/x] = (\sum_{i=1}^{n} p_i.(\!|\mathbf{r}_i|\!))[\mathbf{s}/x] = \sum_{i=1}^{n} p_i.(\!|\mathbf{r}_i|\!)[\mathbf{s}/x]$, which by the induction hypothesis, is equal to $\sum_{i=1}^{n} p_i.(\!|\mathbf{r}_i[\mathbf{s}/x]|\!) = (\!|\pi_B(\mathbf{t}[\mathbf{s}/x])|\!) = (\!|(\pi_B(\mathbf{t}))[\mathbf{s}/x]|\!)$.   □

Theorem 4.9. *Let* $\mathbf{r}, \mathbf{s}, \mathbf{s}_i$ *in* λ_+^p.

- *If* $\mathbf{r} \rightleftarrows \mathbf{s}$, *then* $(\!|\mathbf{r}|\!) \rightleftarrows (\!|\mathbf{s}|\!)$.

- *If* $\mathbf{r} \hookrightarrow \mathbf{s}$, *with probability* 1, *then* $(\!|\mathbf{r}|\!) \hookrightarrow (\!|\mathbf{s}|\!)$, *except if the reduction is done by rule* Π.

- *If* $\mathbf{r} \hookrightarrow \mathbf{s}_i$ *with probability* p_i, *for* $i = 1,\ldots,n$, *then* $(\!|\mathbf{r}|\!) = \sum_{i=1}^{n} p_i.(\!|\mathbf{s}_i|\!)$.

Proof. Case by case analysis.

- Consider $\mathbf{r} + \mathbf{s} \rightleftarrows \mathbf{s} + \mathbf{r}$. Notice that $(\!|\mathbf{r}+\mathbf{s}|\!) = (\!|\mathbf{r}|\!) + (\!|\mathbf{s}|\!) \rightleftarrows (\!|\mathbf{s}|\!) + (\!|\mathbf{r}|\!) = (\!|\mathbf{s}+\mathbf{r}|\!)$.

- Consider $(\mathbf{r}+\mathbf{s})+\mathbf{t} \rightleftarrows \mathbf{r}+(\mathbf{s}+\mathbf{t})$. Notice that $(\!|(\mathbf{r}+\mathbf{s})+\mathbf{t}|\!) = ((\!|\mathbf{r}|\!) + (\!|\mathbf{s}|\!)) + (\!|\mathbf{t}|\!) \rightleftarrows (\!|\mathbf{r}|\!) + ((\!|\mathbf{s}|\!) + (\!|\mathbf{t}|\!)) = (\!|\mathbf{r}+(\mathbf{s}+\mathbf{t})|\!)$.

- Consider $(\mathbf{r}+\mathbf{s})\mathbf{t} \rightleftarrows \mathbf{rt} + \mathbf{st}$. Notice that $(\!|(\mathbf{r}+\mathbf{s})\mathbf{t}|\!) = ((\!|\mathbf{r}|\!) + (\!|\mathbf{s}|\!))(\!|\mathbf{t}|\!) \rightleftarrows (\!|\mathbf{r}|\!)(\!|\mathbf{t}|\!) + (\!|\mathbf{s}|\!)(\!|\mathbf{t}|\!) = (\!|\mathbf{rt}+\mathbf{st}|\!)$.

- Consider $\lambda x^A.(\mathbf{r}+\mathbf{s}) \rightleftarrows \lambda x^A.\mathbf{r} + \lambda x^A.\mathbf{s}$. Notice that $(\!|\lambda x^A.(\mathbf{r}+\mathbf{s})|\!) = \lambda x^A.((\!|\mathbf{r}|\!) + (\!|\mathbf{s}|\!)) \rightleftarrows \lambda x^A.(\!|\mathbf{r}|\!) + \lambda x^A.(\!|\mathbf{s}|\!) = (\!|\lambda x^A.\mathbf{r} + \lambda x^A.\mathbf{s}|\!)$.

- Consider $(\lambda x^A.\mathbf{r})\mathbf{s} \hookrightarrow \mathbf{r}[\mathbf{s}/x]$. Notice that $(\!|(\lambda x^A.\mathbf{r})\mathbf{s}|\!) = (\lambda x^A.(\!|\mathbf{r}|\!))(\!|\mathbf{s}|\!) \hookrightarrow (\!|\mathbf{r}|\!)[(\!|\mathbf{s}|\!)/x]$, and this, by Lemma 4.8(2), is equal to $(\!|\mathbf{r}[\mathbf{s}/x]|\!)$.

- Consider $(\Lambda X.\mathbf{r})\{A\} \hookrightarrow \mathbf{r}[A/X]$. Notice that $(\!|(\Lambda X.\mathbf{r})\{A\}|\!) = \Lambda X.(\!|\mathbf{r}|\!)[A/X] \hookrightarrow (\!|\mathbf{r}|\!)[A/X]$, and this, by Lemma 4.8(1), is equal to $(\!|\mathbf{r}[A/X]|\!)$.

- Consider $\pi_A(\sum_{i=1}^n m_i.\mathbf{r}_i + \mathbf{s}) \hookrightarrow \mathbf{r}_i$ with probability $\frac{m_i}{\sum_{j=1}^n m_j}$, where $\mathbf{r}_i : A$ and $\mathbf{s} \not{:} A$ are closed normal terms. Notice that, by definition, $(\!|\pi_A(\sum_{i=1}^n m_i.\mathbf{r}_i + \mathbf{s})|\!) = \sum_{i=1}^n \frac{m_i}{\sum_{j=1}^n m_j}.(\!|\mathbf{r}_i|\!)$. $\square$

5   Conclusion

In this paper we have defined a probability space on the execution traces of non-confluent abstract rewrite systems. We define a sample space on strategies deciding the rewrite to apply at each state (cf. Definition 2.2).

Our main motivation has been to be able to use this probability space in non-deterministic calculi, hence being able to encode a probability superposition of the kind $\alpha.\mathbf{t} + \beta.\mathbf{r}$, with $\alpha + \beta = 1$, as a term having probability α of rewriting to $\mathbf{t}$ and probability β of rewriting to $\mathbf{r}$. As an example, we provided such an encoding from an algebraic calculus into a non-deterministic calculus.

References

[1] Suzana Andova (1999): *Process Algebra with Probabilistic Choice*. In Joost-Pieter Katoen, editor: *Formal Methods for Real-Time and Probabilistic Systems*, Lecture Notes in Computer Science 1601, pp. 111–129, doi:10.1007/3-540-48778-6_7.

[2] Pablo Arrighi & Alejandro Díaz-Caro (2012): *A System F Accounting for Scalars*. Logical Methods in Computer Science 8(1):11, doi:10.2168/LMCS-8(1:11).

[3] Pablo Arrighi, Alejandro Díaz-Caro & Benoît Valiron (2012): *A Type System for the Vectorial Aspects of the Linear-Algebraic Lambda-Calculus*. In Elham Kashefi, Jean Krivine & Femke van Raamsdonk, editors: *Developments in Computational Models*, Electronic Proceedings in Theoretical Computer Science 88, pp. 1–15, doi:10.4204/EPTCS.88.1.

[4] Pablo Arrighi, Alejandro Díaz-Caro & Benoît Valiron (2013): *The Vectorial Lambda-Calculus*. Available at http://www.arxiv.org/abs/1308.1138. (Under review).

[5] Pablo Arrighi & Gilles Dowek (2008): *Linear-algebraic λ-calculus: higher-order, encodings, and confluence*. In Andrei Voronkov, editor: *Rewriting Techniques and Applications*, Lecture Notes in Computer Science 5117, pp. 17–31, doi:10.1007/978-3-540-70590-1_2. Available at http://www.arxiv.org/abs/quant-ph/0612199.

[6] Henk Barendregt (1984): *The Lambda Calculus: Its Syntax and Semantics*. North-Holland, Amsterdam.

[7] John J. Benedetto & Wojciech Czaja (2009): *Integration and Modern Analysis*, chapter Lebesgue Measure and General Measure Theory. Birkhiäuser Advanced Texts Basler Lehrbücher, Birkhäuser Boston. doi:10.1007/978-0-8176-4656-1.

[8] Gérard Boudol (1994): *Lambda-Calculi for (Strict) Parallel Functions*. Information and Computation 108(1), pp. 51–127, doi:10.1006/inco.1994.1003.

[9] Olivier Bournez & Mathieu Hoyrup (2003): *Rewriting Logic and Probabilities*. In Robert Nieuwenhuis, editor: *Rewriting Techniques and Applications*, Lecture Notes in Computer Science 2706, pp. 61–75, doi:10.1007/3-540-44881-0_6. Available at http://hal.inria.fr/inria-00099620.

[10] Antonio Bucciarelli, Thomas Ehrhard & Giulio Manzonetto (2012): *A Relational Semantics for Parallelism and Non-Determinism in a Functional Setting*. Annals of Pure and Applied Logic 163(7), pp. 918–934, doi:10.1016/j.apal.2011.09.008. Available at http://hal.inria.fr/inria-00628887.

[11] Ugo Dal Lago & Margherita Zorzi (2012): *Probabilistic operational semantics for the lambda calculus*. RAIRO - Theoretical Informatics and Applications 46(03), pp. 413–450, doi:10.1051/ita/2012012. Available at http://www.arxiv.org/abs/1104.0195.

[12] Ugo de'Liguoro & Adolfo Piperno (1995): *Non Deterministic Extensions of Untyped λ-calculus*. Information and Computation 122(2), pp. 149–177, doi:10.1006/inco.1995.1145.

[13] Mariangiola Dezani-Ciancaglini, Ugo de'Liguoro & Adolfo Piperno (1998): *A filter model for concurrent λ-calculus*. SIAM Journal on Computing 27(5), pp. 1376–1419, doi:10.1137/S0097539794275860.

[14] Alessandra Di Pierro, Chris Hankin & Herbert Wiklicky (2005): *Probabilistic λ-Calculus and Quantitative Program Analysis*. Journal of Logic and Computation 15(2), pp. 159–179, doi:10.1093/logcom/exi008.

[15] Alejandro Díaz-Caro & Gilles Dowek (2013): *Non determinism through type isomorphism*. In Delia Kesner & Petrucio Viana, editors: *Logical and Semantic Frameworks, with Applications*, Electronic Proceedings in Theoretical Computer Science 113, pp. 137–144, doi:10.4204/EPTCS.113.13.

[16] Alejandro Díaz-Caro & Gilles Dowek (2013): *Normalisation of a non-deterministic type isomorphic λ-calculus*. Available at http://www.arxiv.org/abs/1306.5089. (Under review).

[17] Alejandro Díaz-Caro, Giulio Manzonetto & Michele Pagani (2013): *Call-by-value non-determinism in a linear logic type discipline*. In Sergei Artemov & Anil Nerode, editors: *Logical Foundations of Computer Science*, Lecture Notes in Computer Science 7734, pp. 164–178, doi:10.1007/978-3-642-35722-0_12.

[18] Alejandro Díaz-Caro & Barbara Petit (2012): *Linearity in the non-deterministic call-by-value setting*. In Luke Ong & Ruy de Queiroz, editors: *Logic, Language, Information and Computation*, Lecture Notes in Computer Science 7456, pp. 216–231, doi:10.1007/978-3-642-32621-9_16. Available at http://www.arxiv.org/abs/1011.3542.

[19] Herman Geuvers, Robbert Krebbers, James McKinna & Freek Wiedijk (2010): *Pure Type Systems without Explicit Contexts*. In Karl Crary & Marino Miculan, editors: *Logical Frameworks and Meta-languages: Theory and Practice*, Electronic Proceedings in Theoretical Computer Science 34, pp. 53–67, doi:10.4204/EPTCS.34.6.

[20] Oltea Mihaela Herescu & Catuscia Palamidessi (2000): *Probabilistic Asynchronous π-Calculus*. In Jerzy Tiuryn, editor: *Foundations of Software Science and Computation Structures*, Lecture Notes in Computer Science 1784, pp. 146–160, doi:10.1007/3-540-46432-8_10. Available at http://www.arxiv.org/abs/cs/0109002.

[21] Michele Pagani & Simona Ronchi Della Rocca (2010): *Linearity, non-determinism and solvability*. Fundamental Informaticae 103(1–4), pp. 173–202, doi:10.3233/FI-2010-324.

[22] Jonghyun Park, Jeongbong Seo, Sungwoo Park & Gyesik Lee (2013): *Mechanizing Metatheory without Typing Contexts*. Journal of Automated Reasoning, doi:10.1007/s10817-013-9287-4.

[23] TeReSe (2003): *Term Rewriting Systems*. Cambridge Tracts in Theoretical Computer Science 55, Cambridge University Press.

[24] Lionel Vaux (2009): *The algebraic lambda calculus*. Mathematical Structures in Computer Science 19(5), pp. 1029–1059, doi:10.1017/S0960129509990089. Available at http://hal.archives-ouvertes.fr/hal-00379750.