On Higher-Order Probabilistic Subrecursion

Flavien Breuvart[1], Ugo Dal Lago[1,2](✉), and Agathe Herrou[3]

[1] INRIA, Sophia Antipolis, France
flavien.breuvart@inria.fr
[2] University of Bologna, Bologna, Italy
ugo.dallago@unibo.it
[3] ENS de Lyon, Lyon, France
agathe.herrou@ens-lyon.fr

Abstract. We study the expressive power of subrecursive probabilistic higher-order calculi. More specifically, we show that endowing a very expressive deterministic calculus like Gödel's $\mathbb{T}$ with various forms of probabilistic choice operators may result in calculi which are not equivalent as for the class of distributions they give rise to, although they all guarantee *almost-sure* termination. Along the way, we introduce a probabilistic variation of the classic reducibility technique, and we prove that the simplest form of probabilistic choice leaves the expressive power of $\mathbb{T}$ essentially unaltered. The paper ends with some observations about functional expressivity: expectedly, all the considered calculi represent precisely the functions which $\mathbb{T}$ itself represents.

1 Introduction

Probabilistic models are more and more pervasive in computer science and are among the most powerful modeling tools in many areas like computer vision [20], machine learning [19] and natural language processing [17]. Since the early times of computation theory [8], the very concept of an algorithm has been itself generalised from a purely deterministic process to one in which certain elementary computation steps can have a probabilistic outcome. This has further stimulated research in computation and complexity theory [11], but also in programming languages [21].

Endowing programs with probabilistic primitives (e.g. an operator which models sampling from a distribution) poses a challenge to programming language semantics. Already for a minimal, imperative probabilistic programming language, giving a denotational semantics is nontrivial [16]. When languages also have higher-order constructs, everything becomes even harder [14] to the point of disrupting much of the beautiful theory known in the deterministic case [1]. This has stimulated research on denotational semantics of higher-order probabilistic programming languages, with some surprising positive results coming out recently [4,9].

The authors are partially supported by ANR project 14CE250005 ELICA and ANR project 12IS02001 PACE.

J. Esparza and A.S. Murawski (Eds.): FOSSACS 2017, LNCS 10203, pp. 370–386, 2017.
DOI: 10.1007/978-3-662-54458-7_22

Not much is known about the expressive power of *probabilistic* higher-order calculi, as opposed to the extensive literature on the same subject about *deterministic* calculi (see, e.g. [23,24]). What happens to the class of representable functions if one enriches, say, a deterministic λ-calculus $\mathbb{X}$ with certain probabilistic choice primitives? Are the expressive power or the good properties of $\mathbb{X}$ somehow preserved? These questions have been given answers in the case in which $\mathbb{X}$ is the pure, untyped, λ-calculus [6]: in that case, universality continues to hold, mimicking what happens in Turing machines [22]. But what if $\mathbb{X}$ is one of the many typed λ-calculi ensuring strong normalisation for typed terms [12]?

Let us do a step back, first: when should a higher-order probabilistic program be considered terminating? The question can be given a satisfactory answer being inspired by, e.g., recent works on probabilistic termination in imperative languages and term rewrite systems [2,18]: one could ask the probability of divergence to be 0, i.e., *almost sure termination*, or the stronger *positive almost sure termination*, in which one requires the average number of evaluation steps to be finite. That almost sure termination is a desirable property, even in a probabilistic setting can be seen in the field of languages like Church and Anglican, in which programs are often assumed to be almost surely terminating, e.g. when doing inference by MH algorithms [13].

In this paper, we initiate a study on the expressive power of terminating higher-order calculi, in particular those obtained by endowing Gödel's $\mathbb{T}$ with various forms of probabilistic choice operators. In particular, three operators will be analysed in this paper:

- A binary probabilistic operator $\oplus$ such that for every pair of terms M, N, the term $M \oplus N$ evaluates to either M or N, each with probability $\frac{1}{2}$. This is a rather minimal option which, however, guarantees universality if applied to the untyped λ-calculus [6] (and, more generally, to universal models of computation [22]).
- A combinator $\mathtt{R}$, which evaluates to any natural number $n \geq 0$ with probability $\frac{1}{2^{n+1}}$. This is the natural generalisation of $\oplus$ to sampling from a distribution having *countable* rather than *finite* support. This apparently harmless generalisation (which is absolutely non-problematic in a universal setting) has dramatic consequences in a subrecursive scenario, as we will discover soon.
- A combinator $\mathtt{X}$ such that for every pair of values V, W, the term $\mathtt{X}\langle V, W\rangle$ evaluates to either W or $V(\mathtt{X}\langle V, W\rangle)$, each with probability $\frac{1}{2}$. The operator $\mathtt{X}$ can be seen as a probabilistic variation on $\mathbb{PCF}$'s fixpoint combinator. As such, $\mathtt{X}$ is potentially problematic to termination, giving rise to infinite trees.

This way, various calculi can be obtained, like $\mathbb{T}^{\oplus}$, namely a minimal extension of $\mathbb{T}$, or the full calculus $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$, in which the three operators are all available. In principle, the only obvious fact about the expressive power of the above mentioned operators is that both $\mathtt{R}$ and $\mathtt{X}$ are at least as expressive as $\oplus$: binary choice can be easily expressed by either $\mathtt{R}$ or $\mathtt{X}$. Less obvious but still easy to prove is the equivalence between $\mathtt{R}$ and $\mathtt{X}$ in presence of a recursive operator (see Sect. 3.3). But how about, say, $\mathbb{T}^{\oplus}$ vs. $\mathbb{T}^{\mathtt{R}}$?

Traditionally, the expressiveness of such languages is evaluated by looking at the set of functions $f : \mathbb{N} \rightarrow \mathbb{N}$ defined by typable programs $M : \mathtt{NAT} \rightarrow \mathtt{NAT}$. However, in a probabilistic setting, any program $M : \mathtt{NAT} \rightarrow \mathtt{NAT}$ computes a function from natural numbers to *distributions* of natural numbers. In order to fit usual criteria, we need to fix a notion of observation of which there are at least two, corresponding to two randomised programming paradigms, namely *Las Vegas* and *Monte Carlo* observations. The main question, then, consists in understanding how the obtained classes relate to each other, and to the class of $\mathbb{T}$-representable functions. Along the way, however, we manage to understand how to capture the expressive power of probabilistic calculi *per se*. This paper's contributions can be summarised as follows:

- We first take a look at the full calculus $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$, and prove that it enforces almost-sure termination, namely that the probability of termination of any typable term is 1. This is done by appropriately adapting the well-known reducibility technique [12] to a probabilistic operational semantics. We then observe that while $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$ cannot be *positively* almost surely terminating, $\mathbb{T}^{\oplus}$ indeed is. This already shows that there must be a gap in expressivity. This is done in Sect. 3.
- In Sect. 4, we look more precisely at the expressive power of $\mathbb{T}^{\oplus}$, proving that the mere presence of probabilistic choice does not add much to the expressive power of $\mathbb{T}$: in a sense, probabilistic choice can be "lifted up" to the ambient deterministic calculus.
- We look at other fragments of $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$ and at their expressivity. More specifically, we prove that (the equiexpressive) $\mathbb{T}^{\mathtt{R}}$ and $\mathbb{T}^{\mathtt{X}}$ represent precisely what $\mathbb{T}^{\oplus}$ can do *at the limit*, in a sense which will be made precise in Sect. 3. This result, which is the most challenging, is given in Sect. 5.
- Section 6 is devoted to proving that both for *Monte Carlo* and for *Las Vegas* observations, the class of representable functions of $\mathbb{T}^{\mathtt{R}}$ coincides with the $\mathbb{T}$-representable ones.

Due to lack of space, most proofs are elided. An extended version of this paper with more details is available [3].

2 Probabilistic Choice Operators, Informally

Any term of Gödel's $\mathbb{T}$ can be seen as a purely deterministic computational object whose dynamics is finitary, due to the well-known strong normalisation theorem (see, e.g., [12]). In particular, the apparent non-determinism due to multiple redex occurrences is completely harmless because of confluence. In this paper, indeed, we even neglect this problem, and work with a reduction strategy, namely weak call-by-value reduction (keeping in mind that all what we will say also holds in call-by-name). Evaluation of a $\mathbb{T}$-term M of type $\mathtt{NAT}$ can be seen as a finite sequence of terms ending in the normal form $\boldsymbol{n}$ of M (see Fig. 1). More generally, the unique normal form of any $\mathbb{T}$ term M will be denoted as $[\![M]\!]$. Noticeably, $\mathbb{T}$ is computationally very powerful. In particular, the $\mathbb{T}$-representable functions from $\mathbb{N}$ to $\mathbb{N}$ coincide with the functions which are provably total in Peano's arithmetic [12].

As we already mentioned, the most natural way to enrich deterministic calculi and turn them into probabilistic ones consists in endowing their syntax with one or more probabilistic choice operators. Operationally, each of them models the essentially stochastic process of sampling from a distribution and proceeding depending on the outcome. Of course, one has many options here as for *which* of the various operators to grab. The aim of this work is precisely to study to which extent this choice have consequences on the overall expressive power of the underlying calculus.

Suppose, for example, that $\mathbb{T}$ is endowed with the binary probabilistic choice operator $\oplus$ described in the Introduction, whose evaluation corresponds to tossing a fair coin and choosing one of the two arguments accordingly. The presence of $\oplus$ has indeed an impact on the dynamics of the underlying calculus: the evaluation of any term M is not deterministic anymore, but can be modelled as a finitely branching tree (see, e.g., Fig. 3 for such a tree). The fact that all branches of this tree have finite height (and the tree is thus finite) is intuitive, and a proof of it can be given by adapting the well-known reducibility proof of termination for $\mathbb{T}$. In this paper, we in fact prove much more, and establish that $\mathbb{T}^{\oplus}$ can be embedded into $\mathbb{T}$.

If $\oplus$ is replaced by $\mathtt{R}$, the underlying tree is not finitely branching anymore, but, again, there is not (at least apparently) any infinitely long branch, since each of them can somehow be seen as a $\mathbb{T}$ computation (see Fig. 2 for an example). What happens to the expressive power of the obtained calculus? Intuition tells us that the calculus should not be too expressive viz. $\mathbb{T}^{\oplus}$. If $\oplus$ is replaced by $\mathtt{X}$, on the other hand, the underlying tree *is* finitely branching, but its height can be infinite. Actually, $\mathtt{X}$ and $\mathtt{R}$ are easily shown to be equiexpressive in presence of higher-order recursion, as we show in Sect. 3.3. On the other hand, for $\mathtt{R}$ and $\oplus$, no such encoding is available. Nonetheless, $\mathbb{T}^{\mathtt{R}}$ can still be somehow encoded into $\mathbb{T}$ (the embedding being correct only "at the limit") as we will detail in Sect. 5. From this embedding, we can show that applying Monte Carlo or Las Vegas algorithms to $\mathbb{T}^{\oplus,\mathtt{X},\mathtt{R}}$ do not add any expressive power to that $\mathbb{T}$. This is done in Sect. 6.

3 The Full Calculus $\mathbb{T}^{\oplus,\mathrm{R},\mathrm{X}}$

All along this paper, we work with a calculus $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$ whose *terms* are the ones generated by the following grammar:

$$\begin{aligned} M, N, L ::= x \;\big|\; \lambda x.M \;\big|\; M\ N \;\big|\; \langle M, N\rangle \;\big|\; \pi_1 \;\big|\; \pi_2 \\ \;\big|\; \mathtt{rec} \;\big|\; \mathbf{0} \;\big|\; \mathtt{S} \;\big|\; M \oplus N \;\big|\; \mathtt{R} \;\big|\; \mathtt{X}. \end{aligned}$$

Please observe the presence of the usual constructs from the untyped λ-calculus, but also of primitive recursion, constants for natural numbers, pairs, and the three choice operators we have described in the previous sections.

As usual, terms are taken modulo α-equivalence. Terms in which no variable occurs free are said *closed*, and are collected in the set $\mathbb{T}_C^{\oplus,\mathtt{R},\mathtt{X}}$. A *value* is simply a closed term from the following grammar:

$$U, V ::= \lambda x.M \;\Big|\; \langle U, V\rangle \;\Big|\; \pi_1 \;\Big|\; \pi_2 \;\Big|\; \mathtt{rec} \;\Big|\; \mathbf{0} \;\Big|\; \mathtt{S} \;\Big|\; \mathtt{S}\ V \;\Big|\; \mathtt{X}.$$

and the set of values is $\mathbb{T}_V^{\oplus,\mathtt{R},\mathtt{X}}$. *Extended values* are (not necessarily closed) terms generated by the same grammar as values with the addition of variables. Closed terms that are not values are called *reducible* and their set is denoted $\mathbb{T}_R^{\oplus,\mathtt{R},\mathtt{X}}$. The expression $\langle M_1, \ldots, M_n\rangle$ stands for $\langle M_1, \langle M_2, \langle \ldots\rangle\rangle\rangle$. A *context* is a term with a unique hole:

$$C := (\!|\cdot|\!) \;\Big|\; \lambda x.C \;\Big|\; C\ M \;\Big|\; M\ C \;\Big|\; \langle C, M\rangle \;\Big|\; \langle M, C\rangle \;\Big|\; C \oplus M \;\Big|\; M \oplus C.$$

We write $\mathbb{T}_{(\!|\cdot|\!)}^{\oplus,\mathtt{R},\mathtt{X}}$ for the set of all such contexts.

Termination of Gödel's $\mathbb{T}$ is guaranteed by the presence of types, which we also need here. *Types* are expressions generated by the following grammar

$$A, B ::= \mathtt{NAT} \;\Big|\; A \to B \;\Big|\; A \times B.$$

Environmental contexts are expressions of the form $\Gamma = x_1 : A_1, \ldots, x_n : A_n$, while *typing judgments* are of the form $\Gamma \vdash M : A$. *Typing rules* are given in Fig. 5. From now on, only typable terms will be considered. We denote by $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}(A)$ the set of terms of type A, and similarly for $\mathbb{T}_C^{\oplus,\mathtt{R},\mathtt{X}}(A)$ and $\mathbb{T}_V^{\oplus,\mathtt{R},\mathtt{X}}(A)$. We use the shortcut $\boldsymbol{n}$ for values of type $\mathtt{NAT}$: $\mathbf{0}$ is already part of the language of terms, while $\boldsymbol{n+1}$ is simply $\mathtt{S}\ \boldsymbol{n}$.

3.1 Operational Semantics

While evaluating terms in a deterministic calculus ends up in a *value*, the same process leads to a *distribution* of values when performed on terms in a probabilistic calculus. Formalising all this requires some care, but can be done following one of the many definitions from the literature (e.g., [6]).

Given a countable set X, a *distribution* $\mathcal{L}$ *on* X is a probabilistic subdistribution over elements of X:

$$\mathcal{L}, \mathcal{M}, \mathcal{N} \in \mathfrak{D}(X) = \Big\{ f : X \to [0,1] \;\Big|\; \sum_{x \in X} f(x) \leq 1 \Big\}.$$

We are especially concerned with distributions over terms here. In particular, a *distribution of type* A is simply an element of $\mathfrak{D}(\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}(A))$. The set $\mathfrak{D}(\mathbb{T}_V^{\oplus,\mathtt{R},\mathtt{X}})$ is ranged over by metavariables like $\mathcal{U}, \mathcal{V}, \mathcal{W}$. We will use the pointwise order $\leq$ on distributions, which turns them into an ω**CPO**. Moreover, we use the following notation for Dirac's distributions over terms: $\{M\} := \left\{ \begin{matrix} M \mapsto 1 \\ N \mapsto 0 \text{ if } M \neq N \end{matrix} \right\}$. The support of a distribution is indicated as $|\mathcal{M}|$; we also define the reducible and value supports fragments as $|\mathcal{M}|_R := |\mathcal{M}| \cap \mathbb{T}_R^{\oplus,\mathtt{R},\mathtt{X}}$ and $|\mathcal{M}|_V := |\mathcal{M}| \cap \mathbb{T}_V^{\oplus,\mathtt{R},\mathtt{X}}$. Notions like $\mathcal{M}^R$ and $\mathcal{M}^V$ have an obvious and natural meaning: for any $\mathcal{M} \in \mathfrak{D}(X)$ and $Y \subseteq X$, then $\mathcal{M}^Y(x) = \mathcal{M}(x)$ if $x \in \mathbb{T}_Y^{\oplus,\mathtt{R},\mathtt{X}}$ and $\mathcal{M}^Y(x) = 0$ otherwise.

As syntactic sugar, we use integral notations to manipulate distributions, *i.e.*, for any family of distributions $(\mathcal{N}_M)_{M \in \mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}} : \mathfrak{D}(\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}})^{\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}}$, the expression $\int_{\mathcal{M}} \mathcal{N}_M.dM$ stands for $\sum_{M \in \mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}} \mathcal{M}(M) \cdot \mathcal{N}_M$ (by abuse of notation, we may define $\mathcal{N}_M$ only for $M \in |\mathcal{M}|$, since the others are not used anyway). The notation can be easily adapted, e.g., to families of real numbers $(p_M)_{M \in \mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}}$ and to other kinds of distributions. We indicate as $C(\!|\mathcal{M}|\!)$ the push-forward distribution $\int_{\mathcal{M}} \{C(\!|M|\!)\} dM$ induced by a context C, and as $\sum \mathcal{M}$ the norm $\int_{\mathcal{M}} 1 dM$ of $\mathcal{M}$. Remark, finally, that we have the useful equality $\mathcal{M} = \int_{\mathcal{M}} \{M\} dM$.

Reduction rules of $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$ are given by Fig. 6. For reasons of simplicity, the relation $\to$ indicates both a subset of $\mathbb{T}_C^{\oplus,\mathtt{R},\mathtt{X}} \times \mathfrak{D}(\mathbb{T}_C^{\oplus,\mathtt{R},\mathtt{X}})$ *and* a relation on $\mathfrak{D}(\mathbb{T}_C^{\oplus,\mathtt{R},\mathtt{X}}) \times \mathfrak{D}(\mathbb{T}_C^{\oplus,\mathtt{R},\mathtt{X}})$. Notice that the reduction $\to$ is deterministic. We can easily define $\to^n$ as the n^{th} exponentiation of $\to$ and $\to^*$ as the reflexive and transitive closure of $\to$ taking the latter as a relation on distributions. In probabilistic systems, we might want to consider infinite reductions such as the ones induced by $\mathtt{X}\langle(\lambda x.x), \mathbf{0}\rangle$, which reduces to $\{\mathbf{0}\}$, but in an infinite number of steps. Remark that for any value V, and whenever $\mathcal{M} \to \mathcal{N}$, it holds that $\mathcal{M}(V) \leq \mathcal{N}(V)$. As a consequence, we can proceed as follows:

Definition 1. *Let M be a term and let $(\mathcal{M}_n)_{n \in \mathbb{N}}$ be the unique distribution family such that $M \to^n \mathcal{M}_n$. The* evaluation *of M is the value distribution*

$$[\![M]\!] \; := \; \{V \mapsto \lim_{n \to \infty} \mathcal{M}_n(V)\} \; \in \; \mathfrak{D}(\mathbb{T}_V^{\oplus,\mathtt{R},\mathtt{X}}).$$

The success *of M is its probability of normalisation, which is formally defined as the norm of its evaluation, i.e., $Succ(M) := \sum [\![M]\!]$. $\mathcal{M}_n^{\Delta V}$ stands for $\{V \mapsto \mathcal{M}_n(V) - \mathcal{M}_{n-1}(V)\}$, the distributions of values reachable in exactly n steps. The* average reduction length *from M is then*

$$[M] \; := \; \sum_n \left(n \cdot \sum \mathcal{M}_n^{\Delta V} \right) \; \in \; \mathbb{N} \cup \{+\infty\}$$

Notice that, by Rule (r-$\in$), the evaluation is continuous: $[\![\mathcal{M}]\!] = \int_{\mathcal{M}} [\![M]\!] dM$. Any closed term M of type $\mathtt{NAT} \to \mathtt{NAT}$ *represents* a function $g : \mathbb{N} \to \mathfrak{D}(\mathbb{N})$ iff for every n, m it holds that $g(n)(m) = [\![M\ \boldsymbol{n}]\!](\boldsymbol{m})$.

3.2 Almost-Sure Termination

We now have all the necessary ingredients to specify a quite powerful notion of probabilistic computation. When, precisely, should such a process be considered *terminating*? Do all probabilistic branches (see Figs. 1, 2, 3 and 4) need to be finite? Or should we stay more liberal? The literature on the subject is pointing to the notion of *almost-sure termination*: a probabilistic computation should be considered terminating if the set of infinite computation branches, although not necessarily empty, has null probability [10,15,18]. This has the following incarnation in our setting:

$$M \to \cdots \to \boldsymbol{n}$$

Fig. 1. A reduction in $\mathbb{T}$

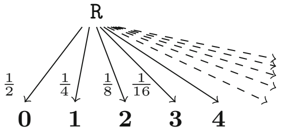


Fig. 2. A reduction in $\mathbb{T}^{\mathtt{R}}$

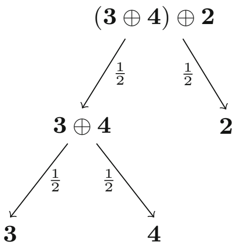

$(\mathbf{3} \oplus \mathbf{4}) \oplus \mathbf{2}$
$\frac{1}{2}$
$\frac{1}{2}$
$\mathbf{3} \oplus \mathbf{4}$
2
$\frac{1}{2}$
$\frac{1}{2}$
3
4

Fig. 3. A reduction in $\mathbb{T}^{\oplus}$

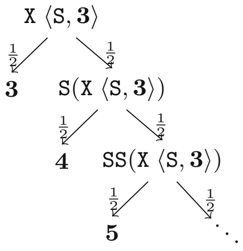

$\mathtt{X}\ \langle \mathtt{S}, \mathbf{3} \rangle$
$\frac{1}{2}$
$\frac{1}{2}$
3
$\mathtt{S}(\mathtt{X}\ \langle \mathtt{S}, \mathbf{3} \rangle)$
$\frac{1}{2}$
$\frac{1}{2}$
4
$\mathtt{SS}(\mathtt{X}\ \langle \mathtt{S}, \mathbf{3} \rangle)$
$\frac{1}{2}$
$\frac{1}{2}$
5

Fig. 4. A reduction in $\mathbb{T}^{\mathtt{X}}$

$$\frac{}{\Gamma, x : A \vdash x : A} \qquad \frac{\Gamma, x : A \vdash M : B}{\Gamma \vdash \lambda x.M : A \to B} \qquad \frac{\Gamma \vdash M : A \to B \qquad \Gamma \vdash N : A}{\Gamma \vdash M\ N : B}$$

$$\frac{}{\Gamma \vdash \mathbf{0} : \mathtt{NAT}} \qquad \frac{}{\Gamma \vdash \mathtt{S} : \mathtt{NAT} \to \mathtt{NAT}} \qquad \frac{}{\Gamma \vdash \mathtt{rec} : A \times (\mathtt{NAT} \to A \to A) \times \mathtt{NAT} \to A}$$

$$\frac{\Gamma \vdash M : A \qquad \Gamma \vdash N : B}{\Gamma \vdash \langle M, N \rangle : A \times B} \qquad \frac{}{\Gamma \vdash \pi_1 : (A \times B) \to A} \qquad \frac{}{\Gamma \vdash \pi_2 : (A \times B) \to B}$$

$$\frac{\Gamma \vdash M : A \qquad \Gamma \vdash N : A}{\Gamma \vdash M \oplus N : A} \qquad \frac{}{\Gamma \vdash \mathtt{R} : \mathtt{NAT}} \qquad \frac{}{\Gamma \vdash \mathtt{X} : (A \to A) \times A \to A}$$

Fig. 5. Typing rules.

$$\frac{}{(\lambda x.M)\ V \to \left\{M[V/x]\right\}}\ (r\text{-}\beta) \qquad \frac{M \to \mathcal{M}}{M\ V \to \mathcal{M}\ V}\ (r\text{-}@L)$$

$$\frac{N \to \mathcal{N}}{M\ N \to M\ \mathcal{N}}\ (r\text{-}@R) \qquad \frac{M \to \mathcal{M}}{\langle M, N \rangle \to \langle \mathcal{M}, N \rangle}\ (r\text{-}\langle\cdot\rangle L) \qquad \frac{M \to \mathcal{M}}{\langle V, M \rangle \to \langle V, \mathcal{M} \rangle}\ (r\text{-}\langle\cdot\rangle R)$$

$$\frac{}{\mathtt{rec}\langle U, V, \mathbf{0} \rangle \to \left\{U\right\}}\ (r\text{-}\mathtt{rec}\mathbf{0}) \qquad \frac{}{\mathtt{rec}\langle U, V, \mathtt{S}\ \boldsymbol{n} \rangle \to \left\{V\ \boldsymbol{n}\ (\mathtt{rec}\langle U, V, \boldsymbol{n} \rangle)\right\}}\ (r\text{-}\mathtt{recS})$$

$$\frac{}{\pi_1\ \langle V, U \rangle \to \left\{V\right\}}\ (r\text{-}\pi_1) \qquad \frac{}{\pi_2\ \langle V, U \rangle \to \left\{U\right\}}\ (r\text{-}\pi_2) \qquad \frac{}{\mathtt{R} \to \left\{\boldsymbol{n} \mapsto \frac{1}{2^{n+1}}\right\}_{n \in \mathbb{N}}}\ (r\text{-}\mathtt{R})$$

$$\frac{}{M \oplus N \to \left\{\begin{matrix} M \mapsto \frac{1}{2} \\ N \mapsto \frac{1}{2} \end{matrix}\right\}}\ (r\text{-}\oplus) \qquad \frac{}{\mathtt{X}\langle V, W \rangle \to \left\{\begin{matrix} V\ (\mathtt{X}\ \langle V, W \rangle) \mapsto \frac{1}{2} \\ W \mapsto \frac{1}{2} \end{matrix}\right\}}\ (r\text{-}\mathtt{X})$$

$$\frac{\forall M \in |\mathcal{M}|_R,\ M \to \mathcal{N}_M \qquad \forall V \in |\mathcal{M}|_V,\ \mathcal{N}_V = \{V\}}{\mathcal{M} \to \int_{\mathcal{M}} \mathcal{N}_M.dM}\ (r\text{-}\in)$$

Fig. 6. Operational semantics.

Definition 2. *A term M is said to be* almost-surely terminating *(AST) iff $Succ(M) = 1$.*

This section is concerned with proving that $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$ indeed guarantees almost-sure termination. This will be done by adapting Girard-Tait's reducibility technique.

The following is a crucial intermediate step towards Theorem 1, the main result of this section.

Lemma 1. *For any M, N, it holds that $[\![M\ N]\!] = [\![[\![M]\!]\ [\![N]\!]]\!]$. In particular, if the application $M\ N$ is almost-surely terminating, so are M and N.*

Theorem 1. *The full system $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$ is almost-surely terminating (AST),* i.e.,

$$\forall M \in \mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}, \quad Succ(M) = 1.$$

Proof. The proof[1] is based on the notion of a *reducible* term which is given as follows by induction on the structure of types:

$$\begin{aligned} Red_{\mathtt{NAT}} &:= \left\{M \in \mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}(\mathtt{NAT}) \mid M \text{ is AST}\right\}; \\ Red_{A\to B} &:= \{M \mid \forall V \in Red_A \cap \mathbb{T}_V^{\oplus,\mathtt{R},\mathtt{X}},\ (M\ V) \in Red_B\}; \\ Red_{A\times B} &:= \{M \mid (\pi_1\ M) \in Red_A,\ (\pi_2\ M) \in Red_B\}. \end{aligned}$$

Then we can observe that:

- *The reducibility candidates over Red_A are $\to$-saturated:* by induction on A we can indeed show that if $M \to \mathcal{M}$ then $|\mathcal{M}| \subseteq Red_A$ iff $M \in Red_A$.
- *The reducibility candidates over Red_A are precisely the AST terms M such that $[\![M]\!] \subseteq Red_A$:* this goes by induction on A. Trivial for $A = \mathtt{NAT}$. Let $M \in Red_{B\to C}$: remark that there is a value $V \in Red_B$, thus $(M\ V) \in Red_C$ and $(M\ V)$ is AST by IH; using Lemma 1 we get M AST and it is easy to see that if $U \in |[\![M]\!]|$ then $U \in |\mathcal{M}|$ for some $M \to^* \mathcal{M}$ so that $U \in Red_{B\to C}$ by saturation. Conversely, let M be AST with $|[\![M]\!]| \subseteq Red_{B\to C}$ and let $V \in Red_B$ be a value: by IH, for any $U \in |[\![M]\!]| \subseteq Red_{B\to C}$ we have $(U\ V)$ AST with an evaluation supported by elements of Red_C; by Lemma 1 $[\![M\ V]\!] = [\![[\![M]\!]\ V]\!]$ meaning that $(M\ V)$ is AST and has an evaluation supported by elements of Red_C, so that we can conclude by IH. Similar for products.
- *Every term M such that $x_1 : A_1, \ldots, x_n : A_n \vdash M : B$ is a candidate in the sense that if $V_i \in Red_{A_i}$ for every $1 \leq i \leq n$, then $M[V_1/x_1, \ldots, V_n/x_n] \in Red_B$:* by induction on the type derivation. The only difficult cases are those for the application and for $\mathtt{X}$ (the one for $\mathtt{rec}$ is just an induction on its third argument).
  - We need to show that if $M \in Red_{A\to B}$ and $N \in Red_A$ then $(M\ N) \in Red_B$. But since $N \in Red_A$, this means that it is AST and for every $V \in |[\![N]\!]|$, $(M\ V) \in Red_B$. In particular, by Lemma 1, we have $[\![M\ N]\!] = [\![M\ \ [\![N]\!]]\!]$ so that $(M\ N)$ is AST and $|[\![M\ N]\!]| \subseteq \bigcup_{V\in|[\![N]\!]|} |[\![M\ V]\!]| \subseteq Red_B$.

[1] Another proof of almost sure termination using reducibility candidate can be found in [25].

- We need to show that for any value $U \in Red_{A\to A}$ and $V \in Red\, A$ if holds that $(\mathtt{X}\ \langle U, V\rangle) \in Red_A$. By an easy induction on n, $(U^n\ V) \in Red_A$. Moreover, by an easy induction on n we have $[\![\mathtt{X}\ \langle U, V\rangle]\!] = \frac{1}{2^{n+1}}[\![U^n\ (\mathtt{X}\ \langle U, V\rangle)]\!] + \sum_{i\leq n} \frac{1}{2^{i+1}}[\![U^i\ V]\!]$ so that at the limit $[\![\mathtt{X}\ \langle U, V\rangle]\!] = \sum_{i\in\mathbb{N}} \frac{1}{2^{i+1}}[\![U^i\ V]\!]$. We can then conclude that $(\mathtt{X}\ \langle U, V\rangle)$ is AST (since each of the $(U^i\ V) \in Red_B$ are AST and $\sum_i \frac{1}{2^{i+1}} = 1$) and that $|[\![M\ N]\!]| = \bigcup_i |[\![U^i\ V]\!]| \subseteq Red_A$.

This concludes the proof. □

Almost-sure termination could however be seen as too weak a property: there is no guarantee about the average computation length. For this reason, another stronger notion is often considered, namely *positive* almost-sure termination:

Definition 3. *A term M is said to be* positively almost-surely terminating *(or* PAST*) iff the average reduction length $[M]$ is finite.*

Gödel's $\mathbb{T}$, when paired with $\mathtt{R}$, is combinatorially too powerful to guarantee positive almost sure termination:

Theorem 2. $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$ *is not positively almost-surely terminating.*

Proof. The naive exponential function applied to $\mathtt{R}$ is computing, with probability $\frac{1}{2^{n+1}}$ the number 2^{n+1} in time 2^{n+1}. This is already a counterexample, because it clearly has infinite average termination time. □

3.3 On Fragments of $\mathbb{T}^{\oplus,\mathrm{R},\mathrm{X}}$: A Roadmap

The calculus $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$ contains at least four fragments, namely Gödel's $\mathbb{T}$ and the three fragments $\mathbb{T}^{\oplus}$, $\mathbb{T}^{\mathtt{R}}$ and $\mathbb{T}^{\mathtt{X}}$ corresponding to the three probabilistic choice operators we consider. It is then natural to ask how these fragments relate to each other as for their respective expressive power. At the end of this paper, we will have a very clear picture in front of us.

The first result we can give is the equivalence between the apparently dual fragments $\mathbb{T}^{\mathtt{R}}$ and $\mathbb{T}^{\mathtt{X}}$. The embeddings are in fact quite simple:

Proposition 1. $\mathbb{T}^{\mathtt{R}}$ *and* $\mathbb{T}^{\mathtt{X}}$ *are both equiexpressive with* $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$*.*

Proof. The calculus $\mathbb{T}^{\mathtt{R}}$ embeds the full system $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$ via the encoding:[2]

$$M \oplus N := \mathtt{rec}\langle \lambda z.N, \lambda xyz.M, \mathtt{R}\rangle \mathbf{0}; \qquad \mathtt{X} := \lambda xy.\mathtt{rec}\langle y, \lambda z.x, \mathtt{R}\rangle.$$

The fragment $\mathbb{T}^{\mathtt{X}}$ embeds the full system $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$ via the encoding:

$$M \oplus N := \mathtt{X}\langle \lambda xy.M, \lambda y.N\rangle\ \mathbf{0}; \qquad \mathtt{R} := \mathtt{X}\langle \mathtt{S}, \mathbf{0}\rangle.$$

In both cases, the embedding is compositional and preserves types. That the two embeddings are correct can be proved easily, see [3]. □

[2] Notice that the dummy abstractions on z and the $\mathbf{0}$ at the end ensure the correct reduction order by making $\lambda z.N$ a value.

Notice how simulating $\mathtt{X}$ by $\mathtt{R}$ requires the presence of recursion, while the converse is not true. The implications of this fact are intriguing, but lie outside the scope of this work.

In the following, we will no longer consider $\mathbb{T}^{\mathtt{X}}$ nor $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$ but only $\mathbb{T}^{\mathtt{R}}$, keeping in mind that all these are equiexpressive due to Proposition 1. The rest of this paper, thus, will be concerned with understanding the relative expressive power of the three fragments $\mathbb{T}$, $\mathbb{T}^{\oplus}$, and $\mathbb{T}^{\mathtt{R}}$. Can any of the (obvious) strict *syntactical* inclusions between them be turned into a strict *semantic* inclusion? Are the three systems equiexpressive?

In order to compare probabilistic calculi to deterministic ones, several options are available. The most common one is to consider notions of observations over the probabilistic outputs; this will be the purpose of Sect. 6. In this section, we will look at whether it is possible to deterministically represent the distributions computed by the probabilistic calculus at hand. We say that the distribution $\mathcal{M} \in \mathfrak{D}(\mathbb{N})$ is *finitely represented* by[3] $f : \mathbb{N} \to \mathbb{B}$, if there exists a q such that for every $k \geq q$ it holds that $f(k) = 0$ and

$$\mathcal{M} = \{\boldsymbol{k} \mapsto f(k)\}.$$

Moreover, the definition can be extended to families of distributions $(\mathcal{M}_n)_n$ by requiring the existence of $f : \mathbb{N} \times \mathbb{N} \to \mathbb{B}$, $q : \mathbb{N} \to \mathbb{N}$ such that for all $k \geq q(n)$, $f(n,k) = 0$ and

$$\forall n, \quad \mathcal{M}_n = \ \{\boldsymbol{k} \mapsto f(n,k)\}.$$

In this case, we say that the representation is *parameterized.*

We will see in Sect. 4 that the distributions computed by $\mathbb{T}^{\oplus}$ are exactly the (parametrically) finitely representable by $\mathbb{T}$ terms. In $\mathbb{T}^{\mathtt{R}}$, however, distributions are more complex (infinite, non-rational). That is why only a characterisation in terms of approximations is possible. More specifically, a distribution $\mathcal{M} \in \mathfrak{D}(\mathbb{N})$ is said to be *functionally represented* by two functions $f : \mathbb{N} \times \mathbb{N} \to \mathbb{B}$ and $g : \mathbb{N} \to \mathbb{N}$ iff for every $n \in \mathbb{N}$ and for every $k \geq g(n)$ it holds that $f(n,k) = 0$ and

$$\sum_{k \in \mathbb{N}} \Big| \ \mathcal{M}(k) \ - \ f(n,k) \ \Big| \ \leq \ \frac{1}{n}.$$

In other words, the distribution $\mathcal{M}$ can be approximated arbitrarily well, and uniformly, by finitely representable ones. Similarly, we can define a parameterised version of this definition at first order.

In Sect. 5, we show that distributions generated by $\mathbb{T}^{\mathtt{R}}$ terms are indeed uniform limits over those of $\mathbb{T}^{\oplus}$; using our result on $\mathbb{T}^{\oplus}$ this give their (parametric) functional representability in the deterministic $\mathbb{T}$.

[3] Here $\mathbb{B}$ stands for the set of dyadic numbers, i.e. rationals in the form $\frac{n}{2^m}$ (where $m, n \in \mathbb{N}$) and $\mathtt{BIN}$ for their representation in system $\mathbb{T}$, encoded as pairs of natural numbers.

4 Binary Probabilistic Choice

This section is concerned with two theoretical results on the expressive power of $\mathbb{T}^{\oplus}$. The main feature of $\mathbb{T}^{\oplus}$ is that its terms are *positively* almost surely terminating. This is a corollary of the following theorem (whose proof [3] proceeds again by reducibility).

Theorem 3. *For any term $M \in \mathbb{T}^{\oplus}$, $M \to^{*} [\![M]\!]$.*

Now, if $M \to^{n} [\![M]\!]$, then $[M]$ can be at most n since the distribution $\mathcal{M}_{m}^{\Delta V}$ of values reachable in exactly m steps (see Definition 1) will be 0 for every $m > n$. But this means that typable terms normalise in finite time:

Corollary 1. *Any term $M \in \mathbb{T}^{\oplus}$ is positively almost-surely terminating.*

But this is not the only consequence. In fact, the finiteness of $[\![M]\!]$ and the fact that $\mathbb{T}^{\oplus}$ is sufficiently expressive allow for a finite representation of $\mathbb{T}^{\oplus}$-distributions by $\mathbb{T}$-definable functions. To prove it, let us consider an extension of $\mathbb{T}$ with a single memory-cell c of type NAT. This memory-cell is used to store some "random coins" simulating probabilistic choices. The operator $\oplus$ can be encoded as follows:

$$(M \oplus N)^{*} \quad := \quad \texttt{if}\ (\texttt{mod}_2\, c)\ \texttt{then}\ (c{:=}\texttt{div}_2\, c\, ; M^{*})\ \texttt{else}\ (c{:=}\texttt{div}_2\, c\, ; N^{*})$$

Notice that conditionals and modulo arithmetic are easily implementable in $\mathbb{T}$. From Theorem 3, we know that for any $M \in \mathbb{T}^{\oplus}(\texttt{NAT})$, there is $n \in \mathbb{N}$ such that $M \to^{n} [\![M]\!]$, and since the evaluation of M can thus involve at most n successive probabilistic choices, we have that

$$[\![M]\!](\boldsymbol{k}) = \frac{\#\{m < 2^{n} \mid k = [\![c{:=}\boldsymbol{m}\, ; M^{*}]\!]\}}{2^{n}}.$$

By way of a state-passing transformation, we can enforce $(c{:=}\boldsymbol{m}\, ; M^{*})$ into a term of $\mathbb{T}$. But then, the whole $\#\{m{<}2^{n} \mid k{=}[\![c{:=}\boldsymbol{m}\, ; M^{*}]\!]\}$ can be represented as a $\mathbb{T}$-term $k : \mathbb{N} \vdash N : \mathbb{N}$ which finitely represents the distribution $[\![M]\!]$.

In the long version of this paper [3], a stronger result is proved, namely that for any functional $M \in \mathbb{T}^{\oplus}(\texttt{NAT}{\to}\texttt{NAT})$, there are terms $M_{\downarrow} \in \mathbb{T}(\texttt{NAT}{\to}\texttt{NAT}{\to}\texttt{NAT})$ and $M_{\#} \in \mathbb{T}(\texttt{NAT}{\to}\texttt{NAT})$ such that for all $n \in \mathbb{N}$:

$$[\![M\ \boldsymbol{n}]\!](\boldsymbol{k}) \quad = \quad \frac{\#\{m < 2^{[\![M_{\#}\ \boldsymbol{n}]\!]} \mid k = [\![M_{\downarrow}\ \boldsymbol{n}\ \boldsymbol{m}]\!]\}}{2^{[\![M_{\#}\ \boldsymbol{n}]\!]}}.$$

The supplementary difficulty, here, comes from the bound $M_{\#}$ that have to be computed dynamically as it depends on its argument $\boldsymbol{n}$.

As a consequence:

Theorem 4. *Distributions generated by $\mathbb{T}^{\oplus}$-terms are precisely those which can be finitely generated by parameterized $\mathbb{T}$-functionals;* i.e., *for any term $M : \texttt{NAT} \to \texttt{NAT}$, there are two $\mathbb{T}$-functionals $f : (\mathbb{N} \times \mathbb{N}) \to \mathbb{B}$ and $q : \mathbb{N} \to \mathbb{N}$ such that for all n:*

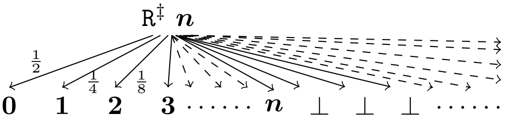
$$[\![M\boldsymbol{n}]\!] = \{\boldsymbol{k} \mapsto f(n,k) \mid k \leq q(n)\}.$$

5 Countable Probabilistic Choice

In this section, we show that $\mathbb{T}^{\oplus}$ approximates $\mathbb{T}^{\mathtt{R}}$: for any term $M \in \mathbb{T}^{\mathtt{R}}(\mathtt{NAT})$, there is a term $N \in \mathbb{T}^{\oplus}(\mathtt{NAT} \to \mathtt{NAT})$ that represents a sequence approximating M uniformly. We will here make strong use of the fact that M has type NAT. This is a natural drawback when we understand that the encoding $(\cdot)^{\dagger}$ on which the result above is based is not direct, but goes through yet another state passing style transformation. Nonetheless, everything can be lifted easily to the first order, achieving the parameterisation of our theorem.

The basic idea behind the embedding $(\cdot)^{\dagger}$ is to mimic any instance of the operator R in the source term by some term $\mathbf{0} \oplus (1 \oplus (\cdots (\boldsymbol{n} \oplus \bot) \cdots))$, where n is *sufficiently large*, and $\bot$ is an arbitrary value of type NAT. Of course, the semantics of this term is *not* the same as that of R, due to the presence of $\bot$; however, n will be chosen sufficiently large for the difference to be negligible. Notice, moreover, that this term can be generalized into the following parametric form $\mathtt{R}^{\ddagger} := \lambda x.\mathtt{rec}\ \langle \bot, (\lambda x.\mathtt{S} \oplus (\lambda y.\mathbf{0})), x\rangle$.

Once $\mathtt{R}^{\ddagger}$ is available, a natural candidate for the encoding $(\cdot)^{\dagger}$ would be to consider something like $M^{\ddagger} := \lambda z.M[(\mathtt{R}^{\ddagger}\ z)/\mathtt{R}]$. In the underlying execution tree, $(M^{\ddagger}\ \boldsymbol{n})$ correctly simulates the first n branches of R (which has infinite arity), but truncates the rest with garbage terms $\bot$:


$\mathtt{R}^{\ddagger}\ \boldsymbol{n}$
$\frac{1}{2}$
$\frac{1}{4}$
$\frac{1}{8}$
$\mathbf{0}$ $\mathbf{1}$ $\mathbf{2}$ $\mathbf{3}$ $\cdots\cdots$ $\boldsymbol{n}$ $\bot$ $\bot$ $\bot$ $\cdots\cdots$

The question is whether the remaining untruncated tree has a "sufficient weight", i.e., whether there is a minimal bound to the probability to stay in this untruncated tree. However, in general $(\cdot)^{\ddagger}$ fails on this point, not achieving to approximate M uniformly. In fact, this probability is basically $(1 - \frac{1}{2^n})^d$ where d is its depth. Since in general the depth of the untruncated tree can grow very rapidly on n in a powerful system like $\mathbb{T}$, there is no hope for this transformation to perform a uniform approximation.

The solution we are using is to have the precision m of $\mathbf{0} \oplus (\mathbf{1} \oplus (\cdots (\boldsymbol{m} \oplus \bot) \cdots))$ to *dynamically grow* along the computation. More specifically, in the approximants $M^{\dagger}\ \boldsymbol{n}$, the growing speed of m will increase with n: in the n-th approximant $M^{\dagger}\ \boldsymbol{n}$, the operator R will be simulated as $\mathbf{0} \oplus (1 \oplus (\cdots (\boldsymbol{m} \oplus \bot) \cdots))$ *and*, somehow, m will be updated to $m + n$. Why does it work? Simply because even for an (hypothetical) infinite and complete execution tree of M, we would stay inside the n^{th} untruncated tree with probability $\prod_{k \geq 0}(1 - \frac{1}{2^{k*n}})$ which is asymptotically above $(1 - \frac{1}{n})$.

Implementing this scheme in $\mathbb{T}^{\oplus}$ requires a feature which is not available (but which can be encoded), namely ground-type references. We then prefer to show that the just described scheme can be realised in an intermediate language called $\mathbb{T}^{\bar{\mathtt{R}}}$, whose operational semantics is formulated not on *terms*, but rather on triples in the form (M, m, n), where M is the term currently being evaluated,

m is the current approximation threshold value, and n is the value of which m is incremented whenever $\mathtt{R}$ is simulated. The operational semantics is standard, except for the following rule:

$$\frac{}{(\bar{\mathtt{R}}, m, n) \to \left\{(k, m{+}n, n) \mapsto \frac{1}{2^{k+1}} \mid k < m\right\}}\,(r\text{-}\bar{\mathtt{R}})$$

Notice how this operator behaves similarly to $\mathtt{R}$ with the exception that it fails when drawing too big of a number (*i.e.*, bigger that the fist state m). Notice that the failure is represented by the fact that the resulting distribution does not necessarily sum to 1. The intermediate language $\mathbb{T}^{\bar{\mathtt{R}}}$ is able to approximate $\mathbb{T}^{\mathtt{R}}$ at every order (Theorem 5 below). Moreover, the two memory cells can be shown to be expressible in $\mathbb{T}^{\oplus}$, again by way of a continuation-passing transformation. Crucially, the initial value of n can be passed as an argument to the encoded term.

For any $M \in \mathbb{T}^{\mathtt{R}}$ we denote $M^* := M[\bar{\mathtt{R}}/\mathtt{R}]$. We say that $(M, m, n) \in \mathbb{T}^{\bar{\mathtt{R}}}$ if $m, n \in \mathbb{N}$ and $M = N^*$ for some $N \in \mathbb{T}^{\mathtt{R}}$. Similarly, $\mathfrak{D}(\mathbb{T}^{\bar{\mathtt{R}}})$ is the set of probabilistic distributions over $\mathbb{T}^{\bar{\mathtt{R}}} \times \mathbb{N}^2$, *i.e.*, over the terms plus states.

For any m and n, the behaviour of M and (M^*, m, n) are similar, except that (M^*, m, n) will "fail" more often. In other words, all $(M^*, m, n)_{m,n\in\mathbb{N}}$ somehow approximate M from below:

Lemma 2. *For any $M \in \mathbb{T}^{\mathtt{R}}$ and any $m, n \in \mathbb{N}$, $[\![M]\!] \succeq [\![M^*, m, n]\!]$, i.e., for every $V \in \mathbb{T}^{\mathtt{R}}_V$, we have*

$$[\![M]\!](V) \geq \sum_{p,q} [\![M^*, m, n]\!](V^*, p, q).$$

Proof. By an easy induction, one can show that for any $\mathcal{M} \in \mathfrak{D}(\mathbb{T}^{\mathtt{R}})$ and $\mathcal{N} \in \mathfrak{D}(\mathbb{T}^{\bar{\mathtt{R}}})$ if $\mathcal{M} \succeq \mathcal{N}$, $\mathcal{M} \to \mathcal{L}$ and $\mathcal{N} \to \mathcal{P}$, then $\mathcal{L} \succeq \mathcal{P}$. This ordering is then preserved at the limit so that we get our result. □

In fact, the probability of "failure" of any $(M, m, n)_{m,n\in\mathbb{N}}$ can be upper-bounded explicitly. More precisely, we can find an infinite product underapproximating the success rate of (M, m, n) by reasoning inductively over the execution $(M, m, n) \to^* [\![(M, m, n)]\!]$, which is possible because of the PAST.

Lemma 3. *For any $M \in \mathbb{T}^{\bar{\mathtt{R}}}$ and any $m, n \geq 1$*

$$Succ(M, m, n) \geq \prod_{k \geq 0} \Big(1 - \frac{1}{2^{m+kn}}\Big).$$

Proof. We denote $\#(m, n) := \prod_{k\geq 0} \Big(1 - \frac{1}{2^{m+kn}}\Big)$ and $\#\mathcal{M} := \int_{\mathcal{M}} \#(m, n) dM\, dm dn$. Remark that for any M and any m, n, if $(M, m, n) \to \mathcal{M}$ then $\mathcal{M}$ is either of the form $\{(N, m, n)\}$ or $\{(N_i, m{+}n, n) \mapsto \frac{1}{2^{i+1}} \mid i < m\}$ for some N of $(N_i)_{i\leq m}$.

Thus we have that if $(M, m, n) \to \mathcal{N}$ then $\#\mathcal{N} = \#(m, n)$ and that if $\mathcal{M} \to \mathcal{N}$ then $\#\mathcal{N} = \#\mathcal{M}$. In particular, since $(M, m, n) \to^* [\![M, m, n]\!]$ we can conclude

$$Succ(M, m, n) \;=\; \int_{[\![M,m,n]\!]} 1 \; dM dm dn \;\geq\; \#[\![M]\!] \;=\; \#(m, n) \;=\; \prod_{k \geq 0} \Big(1 - \frac{1}{2^{m+kn}} \Big).$$

□

This gives us an analytic lower bound to the success rate of (M, m, n). However, it is not obvious that this infinite product is an interesting bound, it is not even clear that it can be different from 0. This is why we will further under-approximate this infinite product to get a simpler expression whenever $m = n$:

Lemma 4. *For any $M \in \mathbb{T}^{\bar{\mathsf{R}}}$ and any $n \geq 4$*

$$Succ(M, n, n) \quad \geq \quad 1 - \frac{1}{n}.$$

Proof. By Lemma 3 we have that $Succ(M, n, n) \geq \prod_{k\geq 1} \big(1 - \frac{1}{2^{kn}}\big)$ which is above the product $\prod_{k \geq 1} \big(1 - \frac{1}{n^2 k^2}\big)$ whenever $n \geq 4$. This infinite product has been shown by Euler to be equal to $\frac{sin(\frac{\pi}{n})}{\frac{\pi}{n}}$. By an easy numerical analysis we then obtain that $\frac{sin(\frac{\pi}{n})}{\frac{\pi}{n}} \geq 1 - \frac{1}{n}$. □

This lemma can be restated by saying that the probability of "failure" of (M^*, n, n), *i.e.* the difference between $[\![M^*, n, n]\!]$ and $[\![M]\!]$, is bounded by $\frac{1}{n}$. With this we then get our first theorem, which is the uniform approximability of elements of $\mathbb{T}^{\mathsf{R}}$ by those of $\mathbb{T}^{\bar{\mathsf{R}}}$:

Theorem 5. *For any $M \in \mathbb{T}^{\mathsf{R}}$ and any $n \in \mathbb{N}$,*

$$\sum_V \Big| \; [\![M]\!](V) \; - \; \Sigma_{m',n'} [\![M^*, n, n]\!](V^*, m', n') \; \Big| \quad \leq \quad \frac{1}{n}.$$

Proof. By Lemma 2, for each V the difference is positive, thus we can remove the absolute value and distribute the sum. We conclude by using the fact that $Succ(M) = 1$ and $Succ(M^*, n, n) \geq 1 - \frac{1}{n}$. □

The second theorem, *i.e.*, the uniform approximability of ground elements of $\mathbb{T}^{\mathsf{R}}$ by those of $\mathbb{T}^{\oplus}$, follows immediately:

Theorem 6. *Distributions in $\mathbb{T}^{\mathsf{R}}(\mathtt{NAT})$ can be approximated by $\mathbb{T}^{\oplus}$-distributions (which are finitely $\mathbb{T}$-representable), i.e., for any $M \in \mathbb{T}^{\mathsf{R}}(\mathtt{NAT})$, there is $M^{\dagger} \in \mathbb{T}^{\oplus}(\mathtt{NAT} \to \mathtt{NAT})$ such that:*

$$\forall n, \quad \sum_{k} \Big| \; [\![M]\!](\boldsymbol{k}) \; - \; [\![M^{\dagger} \; \boldsymbol{n}]\!](\boldsymbol{k}) \; \Big| \quad \leq \quad \frac{1}{n}.$$

Moreover:

- *the encoding is parameterisable, in the sense that for all* $M \in \mathbb{T}^{\mathtt{R}}(\mathtt{NAT} \to \mathtt{NAT})$*, there is* $M^{\wr} \in \mathbb{T}^{\oplus}(\mathtt{NAT} \to \mathtt{NAT} \to \mathtt{NAT})$ *such that* $[\![(M\ \boldsymbol{m})^{\dagger}]\!] = [\![M^{\wr}\ \boldsymbol{m}]\!]$ *for all* $m \in \mathbb{N}$*;*
- *the encoding is such that* $[\![M]\!](\boldsymbol{k}) \leq [\![M^{\dagger}\ \boldsymbol{n}]\!](\boldsymbol{k})$ *only if* $k = 0$*.*

Proof. It is clear that in an extension of $\mathbb{T}^{\oplus}$ with two global memory cells m, n and with an exception monad, the $\bar{\mathtt{R}}$ operator can be encoded by the term $\bar{\mathtt{R}} := \mathtt{rec}\ \langle \bot, (\lambda x.\mathtt{S} \oplus (\lambda y.\mathbf{0})), (m :=!m{+}!n)\rangle$, where $\bot$ is raising an error/exception and where $m :=!m{+}!n$ is returning the value of m before changing the memory cell to $m + n$. We can conclude by referring to the usual state passing style encoding of exceptions and state-monads into $\mathbb{T}$ (and thus into $\mathbb{T}^{\oplus}$). □

6 Subrecursion

If one wishes to define $\mathbb{T}^{\oplus}$-definable or $\mathbb{T}^{\mathtt{R}}$-definable functions as a set of ordinary set-theoretic functions (say from $\mathbb{N}$ to $\mathbb{N}$), it is necessary to collapse the random output into a deterministic one. As already acknowledged by the complexity community, there are at least two reasonable ways to do so: by using either Monte Carlo or Las Vegas observations.

As the careful reader may have foreseen, the finite (parametric) representation of $\mathbb{T}^{\oplus}$-distributions into $\mathbb{T}$ is collapsing both observations into $\mathbb{T}$-definable functions. One only need to explore the finite representation, the resulting process suffers from an exponential blow-up, which is easily absorbed by $\mathbb{T}$, in which all elementary functions (and much more than that!) can be expressed.

Theorem 7 (Monte Carlo). *Let* $f : \mathbb{N} \to \mathbb{N}$ *and* $M : \mathtt{NAT} \to \mathtt{NAT}$ *a* $\mathbb{T}^{\mathtt{R}}$*-term such that* $(M\ \boldsymbol{m})$ *evaluates into* $f(m)$ *with probability* $p \geq \frac{1}{2} + \frac{1}{g(m)}$ *for a* $\mathbb{T}$*-definable function* g*. Then* f *is* $\mathbb{T}$*-definable.*

Theorem 8 (Las Vegas). *Let* $f : \mathbb{N} \to \mathbb{N}$ *and* $M : \mathtt{NAT} \to \mathtt{NAT}$ *a* $\mathbb{T}^{\mathtt{R}}$*-term such that* $(M\ \boldsymbol{m})$ *evaluate either to* $\mathbf{0}$ *(representing a failure) or to* $\boldsymbol{f(m) + 1}$*, the later happening with probability* $p \geq \frac{1}{g(m)}$ *for some* $\mathbb{T}$*-definable function* g*. Then* f *is* $\mathbb{T}$*-definable.*

7 Conclusions

This paper is concerned with the impact of adding various forms of probabilistic choice operators to a higher-order subrecursive calculus in the style of Gödel's $\mathbb{T}$. The presented results help in understanding the relative expressive power of various calculi which can be obtained this way, by showing some separation and equivalence results.

The probabilistic choice operators we consider here are just examples of how one can turn a deterministic calculus like $\mathbb{T}$ into a probabilistic model of computation. The expressiveness of $\mathbb{T}^{\oplus,\mathtt{R},\mathtt{X}}$ is sufficient to encode most reasonable probabilistic operators, but what can we say about their own expressive power? For example, what about a ternary operator in which either of the first two operators is chosen with a probability *which depends* on the value of the third operator? A general theory of probabilistic choice operators and of their expressive power is still lacking.

Another research direction to which this paper hints at consists in studying the logical and proof-theoretical implications of endowing a calculus like $\mathbb{T}$ with probabilistic choice operators. What is even more exciting, however, is the application of the ideas presented here to polynomial time computation. This would allow to go towards a characterization of expected polynomial time computation, thus greatly improving on the existing works on the implicit complexity of probabilistic systems [5,7], which only deals with worst-case execution time. The authors are currently engaged in that.

References

1. Barendregt, H.P.: The Lambda Calculus, Its Syntax and Semantics. Studies in Logic and the Foundations of Mathematics. North-Holland, Amsterdam (1981)
2. Bournez, O., Garnier, F.: Proving positive almost-sure termination. In: Giesl, J. (ed.) RTA 2005. LNCS, vol. 3467, pp. 323–337. Springer, Heidelberg (2005). doi:10.1007/978-3-540-32033-3_24
3. Breuvart, F., Dal Lago, U., Herrou, A.: On probabilistic subrecursion (long version) (2016). http://arxiv.org/abs/1701.04786
4. Crubillé, R., Dal Lago, U.: On probabilistic applicative bisimulation and call-by-value λ-calculi. In: Shao, Z. (ed.) ESOP 2014. LNCS, vol. 8410, pp. 209–228. Springer, Heidelberg (2014). doi:10.1007/978-3-642-54833-8_12
5. Dal Lago, U., Toldin, P.P.: A higher-order characterization of probabilistic polynomial time. Inf. Comput. **241**, 114–141 (2015)
6. Dal Lago, U., Zorzi, M.: Probabilistic operational semantics for the lambda calculus. RAIRO - Theor. Inf. Appl. **46**(3), 413–450 (2012)
7. Dal Lago, U., Zuppiroli, S., Gabbrielli, M.: Probabilistic recursion theory and implicit computational complexity. Sci. Ann. Comp. Sci. **24**(2), 177–216 (2014)
8. De Leeuw, K., Moore, E.F., Shannon, C.E., Shapiro, N.: Computability by probabilistic machines. Automata Stud. **34**, 183–198 (1956)
9. Ehrhard, T., Pagani, M., Tasson, C.: Probabilistic coherence spaces are fully abstract for probabilistic PCF. In: Sewell, P. (ed.) Proceedings of POPL. ACM (2014)
10. Ferrer Fioriti, L.M., Hermanns, H.: Probabilistic termination: soundness, completeness, and compositionality. In: Proceedings of POPL, pp. 489–501 (2015)
11. Gill, J.: Computational complexity of probabilistic turing machines. SIAM J. Comput. **6**(4), 675–695 (1977)
12. Girard, J.-Y., Taylor, P., Lafont, Y.: Proofs and Types. Cambridge University Press, Cambridge (1989)
13. Goodman, N.D., Mansinghka, V.K., Roy, D.M., Bonawitz, K., Tenenbaum, J.B.: Church: a language for generative models. In: UAI, pp. 220–229 (2008)

14. Jung, A., Tix, R.: The troublesome probabilistic powerdomain. Electr. Notes Theor. Comput. Sci. **13**, 70–91 (1998)
15. Kaminski, B.L., Katoen, J.-P.: On the hardness of almost–sure termination. In: Italiano, G.F., Pighizzini, G., Sannella, D.T. (eds.) MFCS 2015. LNCS, vol. 9234, pp. 307–318. Springer, Heidelberg (2015). doi:10.1007/978-3-662-48057-1_24
16. Kozen, D.: Semantics of probabilistic programs. J. Comput. Syst. Sci. **22**(3), 328–350 (1981)
17. Manning, C.D., Schütze, H.: Foundations of Statistical Natural Language Processing, vol. 999. MIT Press, Cambridge (1999)
18. McIver, A., Morgan, C.: Abstraction, Refinement and Proof for Probabilistic Systems. Monographs in Computer Science. Springer, Heidelberg (2005)
19. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo (1988)
20. Prince, S.J.D.: Computer Vision: Models, Learning, and Inference. Cambridge University Press, New York (2012)
21. Saheb-Djahromi, N.: Probabilistic LCF. In: Winkowski, J. (ed.) MFCS 1978. LNCS, vol. 64, pp. 442–451. Springer, Heidelberg (1978). doi:10.1007/3-540-08921-7_92
22. Santos, E.S.: Probabilistic turing machines and computability. Proc. Am. Math. Soc. **22**(3), 704–710 (1969)
23. Sørensen, M.H., Urzyczyn, P.: Lectures on the Curry-Howard Isomorphism. Elsevier Science Inc., New York (2006)
24. Statman, R.: The typed lambda-calculus is not elementary recursive. Theor. Comput. Sci. **9**, 73–81 (1979)
25. Staton, S., Yang, H., Heunen, C., Kammar, O., Wood, F.: Semantics for probabilistic programming: higher-order functions, continuous distributions, and soft constraints. In: Proceedings of LICS, pp. 525–534 (2016)