Boolean-Valued Semantics for the Stochastic $\lambda$-Calculus

Giorgio Bacci
Aalborg University, Denmark

Robert Furber
Aalborg University, Denmark

Dexter Kozen
Cornell University, USA

Radu Mardare
Aalborg University, Denmark

Prakash Panangaden
McGill University, Canada

Dana Scott
Carnegie Mellon University, USA

Abstract

The ordinary untyped $\lambda$-calculus has a set-theoretic model proposed in two related forms by Scott and Plotkin in the 1970s. Recently Scott showed how to introduce probability by extending these models with random variables. However, to reason about correctness and to add further features, it is useful to reinterpret the construction in a higher-order Boolean-valued model involving a measure algebra. We develop the semantics of an extended stochastic $\lambda$-calculus suitable for modeling a simple higher-order probabilistic programming language. We exhibit a number of key equations satisfied by the terms of our language. The terms are interpreted using a continuation-style semantics with an additional argument, an infinite sequence of coin tosses, which serves as a source of randomness. We also introduce a fixpoint operator as a new syntactic construct, as $\beta$-reduction turns out not to be sound for unrestricted terms. Finally, we develop a new notion of equality between terms interpreted in a measure algebra, allowing one to reason about terms that may not be equal almost everywhere. This provides a new framework and reasoning principles for probabilistic programs and their higher-order properties.

Keywords Stochastic $\lambda$-calculus, Boolean-valued models, random variables, denotational semantics.

1 Introduction

Probabilistic programming languages [4–7, 11–13, 21] have become popular recently, sparked by renewed interest in verification and machine learning. The subject began with work by Saheb-Djordjevic [16] on a probabilistic version of LCF and, in an imperative first-order language by [11]. There has been significant recent interest in extending to higher-order functional languages [7, 12, 13, 20]. The higher-order functional paradigm allows one to integrate probability distributions smoothly into the programming language through the probability monad, but finding a cartesian-closed category [9] that can incorporate higher-order features as well as appropriate probabilistic constructions [8] has proven elusive. Only recently [7] has a suitable category been constructed that satisfies all desiderata.

In the present paper, we take an entirely new approach to the semantics of higher-order probabilistic computation. In [18], one of the authors of the present paper, proposed a way of incorporating random variables into a certain kind of model of the untyped $\lambda$-calculus by using the continuity of the $\lambda$-calculus operations modeled by enumeration operators on the powerset of the integers. This set-theoretic model suggests at once incorporating higher types, but to do this requires a nonstandard Boolean-valued interpretation of set theory. Boolean-valued models (see [1] for history and a basic exposition) were employed by Scott [17] to construct models of set theory in order to obtain independence results. The independence of the Continuum Hypothesis was obtained by introducing an arbitrarily large set of real-valued random variables. In this model, the random variables turned out to be the real numbers of the Boolean-valued logic. The continuity of real algebra has an analogue in the continuity of operations in the powerset model. The measure algebra of a standard Borel space, a complete Boolean algebra, is used to bring this idea to fruition. Ordinary logical propositions take elements in this Boolean algebra as truth values, instead of in the simple two-element Boolean algebra.

The ideas behind the present work specify and extend the basic intuitions briefly outlined in [18]. The primary goal here is to develop an equational theory in which equations between stochastic $\lambda$-terms have probabilistic meaning and are valued in a complete Boolean algebra. The intention is to provide reasoning principles for evaluating the equality of $\lambda$-terms under various program transformations.

The notions of equality and invariance are subtle in the presence of probabilities. In the calculus described below, there is a probabilistic choice operator $\otimes$, which captures the idea that a choice is to be made between two terms based on a random process. The source of randomness is called a tossing process: a process that generates a sequence of fair coin tosses, the outcomes of which are used to resolve the probabilistic choices. In general, equality of terms does not mean that identical values are produced, as the final values will depend on the tosses. Instead, we interpret equality statements as elements of a measure algebra formed from the usual measurable sets quotiented by the ideal of negligible sets. Given a tossing process $T$, a pair of closed terms $M, N$ will define a set of tossing sequences where they agree $[M = N]_T$, which is an element of the Boolean algebra. This may be the top element—corresponding to certainty—or something else. However, we would like statements not to be dependent on the specific outcome of a tossing process; rather, we would prefer that truth values of equations be invariant under certain changes in the tossing process. Accordingly, we define a relation $\approx$ on the elements of the Boolean algebra to capture the idea that two truth values of an equation, say $[M = N]_T$ and $[M = N]_{T'}$ for different tossing processes $T$ and $T'$, are related by an automorphism of the Boolean algebra. We write $[M = N]_T \approx [M = N]_{T'}$.
\[M = N \] when this occurs. Many of the equalities that we establish are stated in this way, and the automorphisms relating them are constructed.

A second subtlety is that we often prove results of the form \([M = K]_T \equiv [N = K]_T\), for closed stochastic terms \(M, N\) and closed term \(K\) of the classic untyped \(\lambda\)-calculus (here called a stable term), instead of proving for example that \([M = N]_T\) evaluates to the top element of the Boolean algebra. Here we are using the idea that a tossing process, once it has resolved the choices, makes a term of the stochastic \(\lambda\)-calculus look like an ordinary \(\lambda\)-term. In these cases one cannot prove that \([M = N]_T\) is the top element directly, since this might not be true, but the weaker statement above serves to replace this statement.

Our main contributions are:

1. We develop a use of random variables in this framework in order to identify a rather subtle condition that we call 1-sheeted with respect to \(\alpha\), where \(\alpha\) ranges over \(2^\mathbb{N}, x \in \{0, 1\}^x\), and \(<\) denotes prefix. Let \(B\) be the Boolean \(\sigma\)-algebra of the Cantor topology.

2. We introduce the stochastic \(\lambda\)-calculus by augmenting the ordinary \(\lambda\)-calculus with a probabilistic choice construct. We flesh out the continuation-passing semantics proposed in [18], with the crucial new observation that \(\beta\)-reduction is not sound for all terms with probabilistic choice. To compensate, we need to introduce an explicit fixpoint operator in order to have recursive programs.

3. We develop a Boolean-valued reasoning framework for the stochastic \(\lambda\)-calculus and prove soundness results with respect to the continuation-passing semantics.

The main technical contributions and novelty are in items 1 and 3. In order to obtain the invariance of the semantics, it was necessary to identify a rather subtle condition that we call monolithic. This is the part that required the deepest foray into the technicalities of measure theory. As far as we know, Item 3 is a completely new way of thinking about equational logic. We have only developed the rudiments here for the purposes of the present investigation, but there is clearly a much deeper theory to be explored.

We have not developed an operational semantics or rewrite rules, but have left these investigations for future work. However, because of the restrictions on \(\beta\)-reduction, all probabilistic choices for a term in the argument position must be resolved before applying the function; thus it most resembles a call-by-value strategy, but of course one cannot talk about evaluation strategies in the absence of a reduction system.

2 Standard Probability Spaces

In this section we introduce a few concepts and results regarding standard probability spaces. The concepts of disintegration of a space and of monolithic maps between spaces are essential.

**Definition 2.1.** Given a measurable space \((X, \Sigma)\) and a probability measure \(\mu\) on it, \((X, \Sigma, \mu)\) is a standard probability space\(^1\) if \((X, \Sigma)\) is Borel isomorphic to a Polish space\(^2\) equipped with its Borel algebra\(^3\).

Consider, e.g., the set \(2^\mathbb{N}\) of infinite binary sequences with the Cantor topology\(^4\), which has, as basic open sets, the sets \(\{x \mid x < a\}\), where \(a\) ranges over \(2^\mathbb{N}, x \in \{0, 1\}^x\), and \(<\) denotes prefix. Let \(B\) be the Boolean \(\sigma\)-algebra of the Cantor topology.

The (fair) coin-flipping probability measure\(^5\) \(P\) on \(B\) is generated by its values on intervals:

\[P(\{x \mid x < a\}) = 2^{-|x|}.\]

The measure space \(\Omega = (2^\mathbb{N}, B, P)\) is a standard probability space that we will use in the rest of this paper.

We consider measure-preserving maps between standard probability spaces. The category we use,\(\text{Meas}\)\(\text{t}\) has maps identified if they are equal almost everywhere, i.e. except for on a null set.

**Definition 2.2.** If \(f_1 : (X_1, \Sigma_1, \mu_1) \to (Z, \Xi, \nu)\), \(j = 1, 2\) are two measure-preserving maps with common codomain, we say \(f_1 \cong f_2\) or \((X_1, f_1) \cong (X_2, f_2)\) if there exists a measure-preserving isomorphism \(i : X_1 \to X_2\) such that \(f_2 \circ i = f_1\), except on a subset of \(X_1\) of measure 0.

There is a measure-preserving Borel isomorphism between any standard probability space \((X, \Sigma, \mu)\) whenever \(\mu\) is atomless (i.e. all singletons have measure 0), and \(\Omega\).

**Definition 2.3.** Let \(f : (X, \Sigma, \mu) \to (Y, \Theta, \nu)\) be a measure-preserving map between standard probability spaces. A set \(S \in \Sigma\) is 1-sheeted with respect to \(f\) if for all \(y \in Y\) we have that \(S \cap f^{-1}(y)\) has at most 1 element. A map \(f\) is monolithic if it has no 1-sheeted sets of positive measure.

Note that \(S\) is 1-sheeted if the restriction of \(f\) to \(S\) is injective. For instance, \(2^\mathbb{N}\) is a 1-sheeted set of measure 1 for the identity function \(i : 2^\mathbb{N} \to 2^\mathbb{N}\), and the set \(\{(a_1) \in 2^\mathbb{N} \mid a_1 = 0\}\) is a 1-sheeted set of measure \(\frac{1}{2}\) for the function \(\text{tail} : 2^\mathbb{N} \to 2^\mathbb{N}\), which takes all but the first element of a sequence to return a sequence.

Next, \(f\) is monolithic whenever all such one-sheeted measurable sets have measure 0. For example, the map \(\text{evens} : 2^\mathbb{N} \to 2^\mathbb{N}\) which takes every second member of a sequence to construct a sequence is monolithic (proven in Section 4), while \(i\) and \(\text{tail}\) are not.

The following theorem provides a useful characterization of the concept of monolithic map\(^6\).

**Theorem 2.4.** Let \(f : (X, \Sigma, \mu) \to (Y, \Theta, \nu)\) be a measure-preserving map of standard probability spaces. Then \(f\) is monolithic iff there exists a standard probability space \((Z, \Xi, \mu)\) such that all points of \(Z\) have \(\xi\)-measure 0, and \((Z \times Y, \pi_2)\) is isomorphic to \((X, f)\); where \(\pi_2\) denotes the second projection.

We conclude this section with a useful result regarding the composition of monolithic maps.

\(^1\)Our concept of standard probability space generalizes Rokhlin’s original one in the sense that we do not insist on \(\mu\) being Lebesgue-complete.

\(^2\)A Polish space is the topological space induced by a complete separable metric space.

\(^3\)The Borel algebra of a topology is the \(\sigma\)-algebra generated by the open sets.

\(^4\)This is the Haar measure on \(2^\mathbb{N}\) as a compact group based on \(\text{mod} 2\) addition.

\(^5\)This concept is formulated in terms of decompositions in [15, §3.1], which are better known as disintegration [2, §452 E].
Lemma 2.5. Let \( f : (X, \Sigma, \mu) \to (Y, \Theta, \nu) \) and \( g : (Y, \Theta, \nu) \to (Z, \Xi, \xi) \) be measure-preserving maps. If \( S \in \Sigma \) is 1-sheeted with respect to \( g \circ f \), then it is 1-sheeted with respect to \( f \). Therefore if \( f \) is monolithic, then \( g \circ f \) is monolithic.

Proof: Let \( S \in \Sigma \) be a 1-sheeted set with respect to \( g \circ f \), i.e. for all \( z \in Z \), \( S \cap f^{-1}(g^{-1}(z)) \) has cardinality at most 1. If \( y \in Y \), then 
\[
S \cap f^{-1}(y) \subseteq S \cap f^{-1}(g^{-1}(g(y)))
\]
so \( f^{-1}(y) \cap S \) has cardinality at most 1, so \( S \) is 1-sheeted with respect to \( f \).

The statement about 1-sheeted sets of positive measure then follows by taking the contrapositive. □

3 Topology and measure of \( \mathcal{P}(\mathbb{N}) \)

Let \( \mathbb{N} \) be the set of natural numbers, \( \mathcal{P}(\mathbb{N}) \) its powerset and \( \mathcal{P}_{fin}(\mathbb{N}) \) the set of finite subsets of \( \mathbb{N} \). We follow [14, 18] to identify some structure on \( \mathbb{N} \) that is relevant to our purpose.

The pairing function puts the set of pairs of natural numbers into a one-to-one correspondence with the positive integers by: \( (m, n) = 2^n(2m+1) \). The function \( \langle \cdot \rangle \) puts finite sequences of natural numbers into a one-to-one correspondence with \( \mathbb{N} \) by nested pairing and this can be used to enumerate finite subsets of \( \mathbb{N} \) using a function that we call \( \text{set} \).

The intuition here is that while enumerating the elements in \( \mathbb{N} \), one can also enumerate the elements of \( \mathcal{P}(\mathbb{N}) \) when exists a recursively enumerable set \( F \in \mathcal{P}(\mathbb{N}) \) such that \( A = F(B) \). Hence, the computable enumeration operators are those given by recursively enumerable sets \( F \).

3.1 The positive topology on \( \mathcal{P}(\mathbb{N}) \)

The positive topology on \( \mathcal{P}(\mathbb{N}) \) is induced by the sets 
\[
Q_n = \{ X | n \in X^* \}.
\]

The continuity of a function \( \Phi : \mathcal{P}(\mathbb{N}) \to \mathcal{P}(\mathbb{N}) \) in the positive topology can be characterized as follows: 
\[
m \in \Phi(X) \text{ iff } \exists n \in \mathbb{N}^+ \text{ s.t. } m \in \Phi(set(n)).
\]

Theorem 3.1. The application operation \( F(X) \) is continuous as a function of two variables on \( \mathcal{P}(\mathbb{N}) \).

Theorem 3.2. For every continuous function \( \Phi : \mathcal{P}(\mathbb{N}) \to \mathcal{P}(\mathbb{N}) \) there is a largest set \( F \) such that for any \( X \in \mathcal{P}(\mathbb{N}) \), \( \Phi(X) = F(X) \), where \( \Phi(X) \) denotes ordinary function application, while \( F(X) \) is application in the set-based model. In fact, \( F \) can be directly defined by \( F = \{ 0 \} \cup \{ (n, m) | m \in \Phi(set(n)) \} \).

3.2 Random Variables on \( \Omega \)

Let \( \Omega = (\mathbb{N}, \mathcal{B}, P) \) denote the standard probability space of infinite binary sequences, with \( \mathcal{B} \) the Borel-algebra of the Cantor topology, and with the coin-flipping probability measure \( P \), as defined in §2. The following theorem relates \( \Omega \) to the reals and allows us to use \( \Omega \) as the "source of randomness" when we define random variables.

Theorem 3.3. The measure spaces \( \Omega \) and \( [0, 1] \) with Lebesgue measure, restricted to Borel sets, are Borel isomorphic.

We view \( \mathcal{P}(\mathbb{N}) \) as a measurable space with the Borel-algebra of the positive topology on \( \mathcal{P}(\mathbb{N}) \). This allows us to define random variables on \( \Omega = (\mathbb{N}, \mathcal{B}, P) \) as the measurable functions 
\[
\xi : \Omega \to \mathcal{P}(\mathbb{N}).
\]

Let \( \mathcal{R}(\Omega) \) denote the set of these \( \mathcal{P}(\mathbb{N}) \)-valued random variables on \( \Omega \). Given a random variable \( \xi \), we obtain a measure on \( \mathcal{P}(\mathbb{N}) \) by...
A family of random variables is independent if the induced measures are independent.

On \( R(\Omega) \) we define a few functions. Firstly, we associate to any \( X \in \mathcal{P}(\Omega) \) a random variable \( X \) defined by \( X(\omega) = X \) for any \( \omega \in \Omega \). Secondly, since application, \( \lambda \)-abstraction, Pair, \( \text{Fst} \) and \( \text{Snd} \) are all continuous functions on \( \mathcal{P}(\Omega) \), they can be canonically extended to \( R(\Omega) \). In [18] it is emphasized that \( R(\Omega) \) is a (non-extensional) model for the untyped \( \lambda \)-calculus. Moreover, an equation between two random variables \( \xi, \eta \in R(\Omega) \) can be interpreted as the measurable event
\[
[\xi = \eta] = \{ \omega \in \Omega \mid \forall n \in \xi(\omega), n \in \eta(\omega) \land \forall n \in \eta(\omega), n \in \xi(\omega) \},
\]
which is a Borel set in \( \mathcal{B} \), and this motivates the study of the algebra of events of \( \Omega \) defined below.

### 3.3 The Algebra of Events

Given the probability space \( \Omega = (\mathbb{2}^\mathbb{N}, \mathcal{B}, P) \), we define the algebra of events (also known as the measure algebra) as the Boolean algebra
\[
\mathcal{B}/\mathcal{N}ull = (\mathcal{B}/\mathcal{N}ull, \cup, \cap, \sim, \emptyset/\mathcal{N}ull, \Omega/\mathcal{N}ull)
\]
which is the quotient algebra of the \( \sigma \)-algebra \( \mathcal{B} \) modulo the \( \sigma \)-ideal of Borel sets of \( P \)-measure zero. Observe that we do not get more expressive if we consider the \( P \)-Lebesgue completion of \( \mathcal{B}/\mathcal{N}ull \); because every Lebesgue measurable set differs from a Borel set by a null set, the measure algebra of Lebesgue measurable sets modulo \( P \)-null sets is isomorphic to \( \mathcal{B}/\mathcal{N}ull \).

We call the elements of \( \mathcal{B}/\mathcal{N}ull \) events and for \( A \in \mathcal{B} \) we denote its equivalence class by \( A/\mathcal{N}ull \).

**Theorem 3.4.** \( (\mathcal{B}/\mathcal{N}ull, \cup, \cap, \sim, \emptyset/\mathcal{N}ull, \Omega/\mathcal{N}ull) \) is a \( \sigma \)-complete Boolean algebra in which any family of pairwise disjoint elements is countable (i.e. it satisfies the countable chain condition). Therefore it is a complete Boolean algebra.

\( \mathcal{B}/\mathcal{N}ull \) plays a central role in the semantics of the stochastic \( \lambda \)-calculus.

### 4 Tossing processes

A key ingredient in any probabilistic programming language is the source of randomness. As in [18], this is taken to be a random variable which uses an infinite sequence of independent fair coin tosses to resolve the random choice. The semantics should not depend on the vagaries of a particular sequence; accordingly, we aim to prove a property that shows that the semantics should be independent, in a suitable sense, of the coin tosses that occur. This is where the notion of monolithic function becomes important. We call the special random variables that we use tossing processes. This section is devoted to the properties of tossing processes.

#### 4.1 Independent coin sequences

We need to move between sequences of coin tosses and subsets of \( \mathbb{N} \) by using some appropriate coding and decoding functions. Concretely, we can a Borel-measurable map \( \text{pack} : \mathbb{2}^\mathbb{N} \to \mathcal{P}(\Omega) \) which encodes a sequence as a set in a way that can be easily inverted and its inverse, called \( \text{unpack} \), is Borel-measurable as well.

Moreover, these can be defined so that they properly relate the well-known operations on sequences \( \text{head} : \mathbb{2}^\mathbb{N} \to \mathbb{2}^\mathbb{N} \) and \( \text{tail} : \mathbb{2}^\mathbb{N} \to \mathbb{2}^\mathbb{N} \), given by \( \text{head}(a) = a_0 \) and \( \text{tail}(a)n = a_{n+1} \), to the functions \( \text{Fst} \) and \( \text{Snd} \) defined on sets in Section 3 (i.e., on constant random variables) as stated in the following lemma.

**Lemma 4.1.** Let \( a \in \mathbb{2}^\mathbb{N} \) and \( \alpha \in \{0, 1\} \).

1. \( \{\alpha\} = \text{Fst}(\text{pack}(a)) \) iff \( \text{head}(a) = \alpha \).
2. \( \text{Snd}(\text{pack}(a)) = \text{pack}(\text{tail}(a)) \).

Let \( \Omega = (\mathbb{2}^\mathbb{N}, \mathcal{B}, P) \) be the probability space defined in § 3.2.

**Definition 4.2.** A coin flip is a random variable that has the form \( F : \Omega \to \{0, 1\} \). A coin flip is fair whenever \( P(F^{-1}(\{0\})) = 1/2 \).

An independent sequence of coin tosses (ICS) is a random variable \( \mathcal{T} : \Omega \to \mathcal{P}(\Omega) \) such that \( \text{Fst}(\mathcal{T}) \) is a fair coin flip and \( \text{Snd}(\mathcal{T}) \) is another ICS—with the successive flips all mutually independent.

Note that an ICS is a \( \{0, 1\} \)-valued map, i.e., specialized to take values in the image of \( \text{pack} \). Every ICS \( \mathcal{T} : \Omega \to \mathcal{P}(\Omega) \) is of the form \( \mathcal{T} = \text{pack} \circ \mathcal{T} \) for some (Borel) measurable map \( \mathcal{T} : \Omega \to \mathbb{2}^\mathbb{N} \).

By Lemma 4.1, \( \text{Fst} \circ \text{pack} \circ \mathcal{T} = \text{head} \circ \mathcal{T} \), so the condition of \( \text{Fst}(\mathcal{T}) \) being a fair coin is that for \( i \in \{0, 1\} \),
\[
P(\text{head} \circ \mathcal{T}^{-1}(i)) = \frac{1}{2}.
\]

That \( \text{Snd}(\mathcal{T}) \) is an ICS implies that for all \( n \in \mathbb{N} \) and \( i \in \{0, 1\} \),
\[
P(\text{ev}_n \circ \mathcal{T}^{-1}(i)) = \frac{1}{2},
\]
where \( \text{ev}_n : \mathbb{2}^\mathbb{N} \to \{0, 1\} \) for \( n \in \mathbb{N} \), is defined by \( \text{ev}_n(a) = a_n \).

The condition of independence implies that for any finite increasing sequence \( (m_i)_{i=1}^N \) in \( \mathbb{N} \), and finite sequence \( (b_i)_{i=1}^n \) in \( \{0, 1\} \),
\[
P\left(\bigcap_{i=1}^n (\text{ev}_i(m_i) \circ \mathcal{T}^{-1}(b_i))\right) = 2^{-n}.
\]

This means that the image measure \( T_\mathcal{T}(P) = P \circ \mathcal{T}^{-1} \) agrees with the standard coin-flipping measure on \( \mathbb{2}^\mathbb{N} \) on basic clopens of Cantor topology. As basic clopens form a \( \pi \)-system (they are closed under finite intersections) and generate the Borel sets of \( \mathbb{2}^\mathbb{N} \), \( T_\mathcal{T}(P) \) and the coin-flipping measure agree on all Borel sets [22, Lemma 1.6], so we have that \( T \) is measure-preserving from \( \Omega \) to \( \mathbb{2}^\mathbb{N} \).

Conversely, for any measure-preserving map \( \mathcal{T} : \Omega \to \mathbb{2}^\mathbb{N} \), \( \text{pack} \circ \mathcal{T} \) will be an independent sequence of coin tosses.

These provide the following characterization for ICS.

**Theorem 4.3 (Characterization).** Independent coin sequences are exactly the maps of the form \( \text{pack} \circ \mathcal{T} \), for some measure-preserving map \( \mathcal{T} : \Omega \to \Omega \).

Now we focus on another important function for our discussion, \( \text{evens} : \mathbb{2}^\mathbb{N} \to \mathbb{2}^\mathbb{N} \), defined by \( \text{evens}(a)(n) = a_{2n} \).

\(^1\text{A detailed construction of a pair of such coding/decoding functions can be found in the appendix.}\)
Theorem 4.4. A measure-preserving map \( T : 2^\mathbb{N} \to 2^\mathbb{N} \) is monolithic iff \( T \equiv \text{evens} \).

Proof: Because \( \mathbb{N} \cong \mathbb{N} + \mathbb{N} \) by mapping the odd numbers to the first part and the even numbers to the second part, we have \( (2^\mathbb{N} \times 2^\mathbb{N}, \pi_2) \equiv (2^\mathbb{N}, \text{evens}) \). As any atomless standard probability space is isomorphic to \( 2^\mathbb{N} \), applying Theorem 2.4 with \( X \) and \( Y \) specialized to \( 2^\mathbb{N} \), we get the desired equivalence. \( \square \)

4.2 Tossing Processes

We are ready to define the concept of tossing process; the crucial point is to insist on a monolithic function.

Definition 4.5 (Tossing Process). A tossing process is an independent sequence of coin tosses \( T = \text{pack} \circ T \), where \( T : 2^\mathbb{N} \to 2^\mathbb{N} \) is a monolithic measure-preserving map.

We denote by \( \text{Toss} \) the set of tossing processes. The next theorem states that all tossing processes are the same up to a measure-preserving automorphism of the measure algebra. Let \( \text{Aut}(\Omega) \) denote the set of measure-preserving automorphisms on \( \Omega \).

Theorem 4.6 (Representation Theorem). For any two tossing processes \( S \) and \( T \) there exists a measure-preserving automorphism \( \alpha : \Omega \to \Omega \in \text{Aut}(\Omega) \) such that \( T = S \circ \alpha \) except for a subset of \( \Omega \) of measure \( 0 \).

4.3 Tossing Process Operators

Before concluding this section, we show a few useful operators on \( \mathcal{R}(\Omega) \) that are closed on tossing processes.

Let \( (\cdot)^T, (\cdot)^o \). Swap : \text{Toss} \to \mathcal{R}(\Omega) \) be defined for an arbitrary tossing process \( T \) as follows.

\[
T^T = \text{pack} \circ \text{evens} \circ T \\
T^o = \text{pack} \circ \text{odds} \circ T \\
\text{Swap}(T) = \text{pack} \circ \text{swap} \circ T,
\]

where \( \text{odds}, \text{swap} : \Omega \to \Omega \) are defined by \( \text{odds}(x) = 2a + 1 \) and \( \text{swap}(x) = a \), for \( a \in \mathbb{N} \).

Lemma 4.7. If \( T \) is a tossing process, then \( \text{Snd}(T), T^T, T^o \) and \( \text{Swap}(T) \) are tossing processes as well.

5 Stochastic \( \lambda \)-Calculus

In this section we introduce the stochastic \( \lambda \)-calculus. The syntax of the stochastic \( \lambda \)-calculus over a set \( V \) of variables extends the syntax of the classical untyped \( \lambda \)-calculus with a (binary) probabilistic choice operator \( @ \) on \( \lambda \)-terms, and a fixpoint combinator \( \mu \times M \);

\[
M ::= x | \lambda x.M | M M | \mu x.M | M @ M.
\]

Notation: In what follows we will call terms without any occurrence of \( @ \) stable terms; however, a stable term may contain the fixpoint operator. And, as usual, we use \( M(N/x) \) to denote the substitution of the variable \( x \) by the term \( N \) in \( M \).

Unlike in the classical \( \lambda \)-calculus, in the stochastic \( \lambda \)-calculus the fixpoint combinator cannot be defined from the other operators, as we will demonstrate later. For the development of the fixpoint operator, it is useful to define the unfolding of a recursive term.

Given \( \mu x. M \), its unfolding is the sequence of terms \( M^0, M^1, \ldots \) defined inductively as follows.

\[
M^0 = (\lambda x. M)(\lambda x. M), \quad \text{and for arbitrary } n, M^{n+1} = (\lambda x. M)M^n.
\]

6 Probabilistic Continuation Semantics

The continuation semantics for the stochastic \( \lambda \)-calculus interprets a \( \lambda \)-term relative to an environment giving values to the free variables, a continuation giving a subsequent computation, and a tossing process \( T \) used to resolve probabilistic choices.

We know that the set \( \mathcal{R}(\Omega) \) of random variables with the pointwise order forms a domain such that \( [\mathcal{R}(\Omega) \to \mathcal{R}(\Omega)] \) is a continuous retract of \( \mathcal{R}(\Omega) \), where \( [\mathcal{R}(\Omega) \to \mathcal{R}(\Omega)] \) is the space of Scott-continuous functions. We write this explicitly by introducing the functions

\[
\phi : \mathcal{R}(\Omega) \to [\mathcal{R}(\Omega) \to \mathcal{R}(\Omega)], \quad \psi : [\mathcal{R}(\Omega) \to \mathcal{R}(\Omega)] \to \mathcal{R}(\Omega).
\]

It is useful to also define the direct and the continuation-passing semantics for stable \( \lambda \)-terms. We use \( x \) to range over variables, \( E \) to range over arbitrary environments, \( C \) to range over arbitrary continuations, and \( T \) to range over arbitrary tossing processes. As before, the \( n \)-th unfolding of the term \( \mu x. M \) is denoted by \( M^n \).

Let \( (\cdot) \) denote the direct semantics and \( (\cdot)' \) the continuation-passing semantics for stable \( \lambda \)-terms. In addition, for stochastic terms, we denote by \( (\cdot)_n \) the continuation-passing semantics augmented with a tossing process.

\[
(\cdot) : \text{Term} \to \text{Env} \to \mathcal{R}(\Omega) \\
(\cdot)' : \text{Term} \to \text{Env} \to \text{Cont} \to \mathcal{R}(\Omega) \\
(\cdot)_n : \text{Term} \to \text{Env} \to \text{Cont} \to \text{Toss} \to \mathcal{R}(\Omega),
\]

where

\[
\text{Env} = \text{Var} \to \mathcal{R}(\Omega) \quad \text{Cont} = [\mathcal{R}(\Omega) \to \mathcal{R}(\Omega)]
\]

The direct semantics is

\[
(\lambda x. M) \to \phi(\lambda x. \lambda y. E)(\lambda x. M)E(\psi y/x))
\]

\[
(\mu x. M) \to \sup_n (M^n)E.
\]

With continuations, define

\[
(\lambda x. M)EC = C((\lambda x. M)E)
\]

\[
(\mu x. M)EC = \sup_n (M^n)E.
\]

The probabilistic continuation is defined as follows.

\[
(\lambda x. M)ECT = C((\lambda x. M)E(\lambda x. C(\phi (a) b))T^T)
\]

\[
(\mu x. M)ECT = \sup_n (M^n)ECT.
\]

The probability distribution is defined as follows.

\[
(\mu x. M)ECT = \sup_n (M^n)ECT.
\]
The relation between the three semantics for stable terms is stated in the following proposition.

**Proposition 6.1.** If $M$ is a stable term, then for an arbitrary environment $E$, an arbitrary continuation process $C$, and an arbitrary tossing process $T$,  

$$C(\langle M \rangle E) = \langle M \rangle EC = \langle M \rangle ECT .$$

A corollary of this lemma is that if a closed program has a value, then its value is the same for all tossing processes.

**Corollary 6.2** (Absoluteness I). If $M$ is a stable term, then for an arbitrary environment $E$, continuation process $C$, and tossing processes $T, T'$,  

$$\langle M \rangle ECT = \langle M \rangle ECT' .$$

We conclude this section with two useful lemmas.

**Lemma 6.3.** For any stochastic $\lambda$-terms $M, N$, any arbitrary environment $E$, any arbitrary continuation $C$, and any arbitrary tossing process $T$, the following statements hold, where $M^n$ denote the $n$-unfolding of $\mu x.M$.

1. $\langle \lambda y.(\mu x.M) \rangle ECT = \sup_n \langle \lambda y.M^n \rangle ECT$;
2. $\langle N(\mu x.M) \rangle ECT = \sup_n \langle NM^n \rangle ECT$;
3. $\langle (\lambda x.M)N \rangle ECT = \sup_n \langle M^N \rangle ECT$.

**Lemma 6.4.** For any stochastic $\lambda$-terms $M, N$, any arbitrary environment $E$, any arbitrary continuation $C$, and any arbitrary tossing process $T$,  

$$\langle (\lambda x.M)N \rangle ECT = \langle M \rangle E((\langle N \rangle E(\lambda w.w)T^{\tau} / x))CT^{\tau} .$$

We conclude this section by presenting a direct consequence of Theorem 4.6.

**Theorem 6.5** (Absoluteness II). Given a term $M$, for an arbitrary environment $E$, an arbitrary continuation process $C$, and arbitrary tossing processes $T, T'$, there exists a measure-preserving automorphism $\alpha : \Omega \rightarrow \Omega$ in $\Omega = \Aut(\Omega)$ such that  

$$\{ \omega \in \Omega | \langle M \rangle ECT(\omega) = \langle N \rangle ECT'(\omega) \}/Null = \Omega/Null .$$

7 A Boolean-Valued Model

The Boolean-valued model gives a novel interpretation of equality. Equalities of closed terms, when interpreted over $\mathcal{R}(\Omega)$, are associated with events in $\mathcal{B}/\Null$ up to a measure-preserving automorphism of $\Omega$. Since we are working with closed terms we evaluate terms in the empty environment and with the the identity continuation.

Let $\emptyset$ denote the empty environment and $id = \lambda x.x$ denote the identity continuation.

**Definition 7.1.** For arbitrary closed terms $M, N$, and tossing process $T$, let  

$$[M = N]_T = \{ \omega \in \Omega | \langle M \rangle \emptyset idT(\omega) = \langle N \rangle \emptyset idT(\omega) \}/Null .$$

Note that $[M = N]_T \in \mathcal{B}/\Null$ and that this value depends directly on the tossing process $T$. However, since the tossing processes are all equal up to a measure-preserving automorphism, as proven in Theorem 4.6, $[M = N]_T$ is unique up to an automorphism of $\mathcal{B}/\Null$.

In what follows, for arbitrary $A, A' \in \mathcal{B}/\Null$, we write  

$$A \simeq A'$$

if there exists a measure-preserving automorphism $\sigma$ of $\mathcal{B}/\Null$ such that $\sigma(A) = A'$.

**Theorem 7.2** (Absoluteness III). For arbitrary closed terms $M, N$, and arbitrary tossing processes $T$ and $T'$, where $\alpha \in \Aut(\Omega)$,  

$$\alpha^{-1}([M = N]_T) = [M = N]_{T'},$$

where $\alpha^{-1}$ is the set-theoretical inverse of $\alpha$, hence an automorphism of $\mathcal{B}/\Null$. Consequently,  

$$[M = N]_T \simeq [M = N]_{T'} .$$

**Proof.**  

Since $\emptyset$ and $id$ are constant, this set is further equal to  

$$\{ \omega \in \Omega | \langle M \rangle \emptyset idT'(\alpha(\omega)) = \langle N \rangle \emptyset idT'(\alpha(\omega)) \}/Null = \alpha^{-1}(\{ \omega \in \Omega | \langle M \rangle \emptyset idT(\omega) = \langle N \rangle \emptyset idT(\omega) \}/Null) .$$

This last theorem suggests that in what follows we can use any tossing process to evaluate the equality between closed programs, since the result is in any case unique up to an automorphism of the measure algebra.

8 Sound Equations

In this section we establish a series of sound equations that provide basic reasoning principles for our stochastic $\lambda$-calculus. These equations are by no means complete, but they do describe several basic facts about $\mathcal{R}(\Omega)$. We also show how the usual equations for $\alpha$-reduction and $\beta$-reduction are generalized. Note that, for stable closed terms, we have the soundness of $\alpha$-reduction and $\beta$-reduction for the model $\mathcal{R}(\Omega)$ of ordinary $\lambda$-calculus from Proposition 6.1.

In what follows, all the terms in expressions of the form $[M = N]_T$ are implicitly assumed to be closed terms.

The first result shows how one can substitute terms in equations with terms that are equal almost everywhere.

**Theorem 8.1** (Substitution). If $M$ and $N$ are closed terms such that $[M = N]_T = \Omega/\Null$, then for any closed stable term $K$,  

$$[M = K]_T = [N = K]_T .$$

**Proof.** We have that  

$$[M = N]_T = \{ \omega \in \Omega | \langle M \rangle \emptyset idT(\omega) = \langle N \rangle \emptyset idT(\omega) \}/Null .$$
Boolean-Valued Semantics for the Stochastic λ-Calculus

This means that \([M = N]_T = \Omega/Null\) implies that \(\langle M \rangle 0 idT \) and \(\langle N \rangle 0 idT\) are equal almost everywhere. But then,
\[
[M = K]_T = \{\omega \in \Omega \mid \langle M \rangle 0 idT(\omega) = \langle K \rangle 0 idT(\omega)\}/Null = \{\omega \in \Omega \mid \langle N \rangle 0 idT(\omega) = \langle K \rangle 0 idT(\omega)\}/Null = \{ N = K \}_T. \quad \square
\]

There are many interesting properties that one can prove in this setting. We begin by observing that both \(\alpha\)-reduction and \(\beta\)-reduction for stable terms (i.e. terms without any occurrence of \(\oplus\)) hold, as direct consequences of the fact that \(R(\Omega)\) is a model of the usual untyped \(\lambda\)-calculus [18]. However, \(\alpha\)-reduction also holds for our stochastic \(\lambda\)-calculus; this follows from the probabilistic continuation semantics of \(\lambda\)-terms introduced in Section 6.

**Theorem 8.2 (\(\alpha\)-reduction).** If \(M\) is a term without free occurrences of \(y\), then
\[
[L x . M = L x . M(y/x)]_T = \Omega/Null.
\]

Now we state \(\beta\)-reduction only for stable terms. Later in this section we will show that an unrestricted version of \(\beta\)-reduction is not possible, but that we have, however, some extensions that involve terms that might be not stable.

**Theorem 8.3 (\(\beta\)-reduction).** Let \(N\) and \(M\) be stable terms. Then,
\[
[\langle L x . M \rangle(N) = M[N/x]]_T = \Omega/Null.
\]

Next we prove a series of results regarding the properties of the probabilistic choice operator. We start with the following lemma.

**Lemma 8.4.** Let \(M_1, M_2, N_1, N_2\) be closed terms and \(T\) an arbitrary tossing process. If for any stable closed term \(K\) and each \(i \in \{1, 2\}\) we have that \([M_i = K]_{\text{Snd}(T)} \approx [N_i = K]_{\text{Snd}(T)}\), then for any stable closed term \(K\),
\[
[M_1 \oplus M_2 = K]_T \approx [N_1 \oplus N_2 = K]_T.
\]

**Proof.** The key observation is that there exists a \(j \in \{1, 2\}\) such that \([M_j \oplus M_j = K]_{\text{Snd}(T)}\) and at the same time
\[
[N_j \oplus N_j = K]_{\text{Snd}(T)}.
\]

Also, since \(K\) is stable, \([K]_{\text{Snd}(T)} = [K]_{\text{Snd}(T)}\). Hence,
\[
[M_1 \oplus M_2]_{\text{Snd}(T)} = \{\omega \in \Omega \mid \langle M_1 \oplus M_2 \rangle 0 idT(\omega) = \langle K \rangle 0 idT(\omega)\}/Null
\]
\[
[N_1 \oplus N_2]_{\text{Snd}(T)} = \{\omega \in \Omega \mid \langle M_1 \oplus M_2 \rangle 0 idT(\omega) = \langle K \rangle 0 idT(\omega)\}/Null
\]
\[
[M_j = K]_{\text{Snd}(T)} \approx [N_j = K]_{\text{Snd}(T)}.
\]

Similarly, \([N_1 \oplus N_2 = K]_T \approx [N_1 \oplus N_2 = K]_T\). Now the hypothesis guarantees that \([M_j = K]_{\text{Snd}(T)} \approx [N_j = K]_{\text{Snd}(T)}\). \(\square\)

A first fundamental property of probabilistic choice is a kind of commutativity, stated in the following axiom. Note that one cannot assert commutativity naively; one has to talk in terms of a stable term obtained by resolving all the choices.

**Theorem 8.5 (\(\oplus\)-commutativity).** If \(K\) is a stable term, then we have
\[
[M \oplus N = K]_T \approx [N \oplus M = K]_T.
\]

**Proof.** The map \(\neg : 2^N \rightarrow 2^N\) where \(\neg(\omega) = 1 - \omega(0)\) and \(\neg(\omega)(i) = 1 - \omega(i)\) for \(i > 0\) is measurable and measure-preserving. The tossing process \(T\) is of the form \(\text{pack} \circ T\) for some measure-preserving monolithic map \(T : \Omega \rightarrow 2^N\). By Lemma 2.5, \(T' = \text{pack} \circ \neg \circ T\) is also a tossing process, and so by Theorem 4.6 there exists a measure-preserving automorphism \(\alpha : \Omega \rightarrow \Omega\) such that \(T' = \alpha \circ T\). We start by observing that
\[
[M @ N]_T = \text{Test}(\text{Fst}(T))((M) \oplus \text{id} \text{(Snd}(T))))((N) \oplus \text{id} \text{(Snd}(T))))
\]
\[
= \text{Test}(\text{Fst}(T'))(((M) \oplus \text{id}(\text{Snd}(T))))((N) \oplus \text{id}(\text{Snd}(T))))
\]
\[
= [M @ N]_T \oplus id T'
\]
\[
= [N @ M]_T \oplus id (T \circ \alpha),
\]
using the fact that \(\text{Snd}(T') = \text{Snd}(T)\), since \(\neg\) only affects the first part. So
\[
[M \oplus N = K]_T
\]
\[
= \{\omega \in \Omega \mid \langle M @ N \rangle 0 idT(\omega) = \langle K \rangle 0 idT(\omega)\}/Null
\]
\[
= \{\omega \in \Omega \mid \langle N @ M \rangle 0 idT(\omega) = \langle K \rangle 0 idT(\omega)\}/Null
\]
\[
= \alpha^{-1}(\{\omega \in \Omega \mid \langle N @ M \rangle 0 idT(\omega) = \langle K \rangle 0 idT(\omega)\}/Null
\]
where the third equality holds because \([K]\) does not depend on \(\omega\) since it is stable. \(\square\)

Next we state that \(\oplus\) is idempotent in the same sense as in the previous theorem.

**Theorem 8.6 (\(\oplus\)-idempotence).** If \(K\) is a stable term, then we have
\[
[M @ M = K]_T \approx [M = K]_T.
\]

The next two theorems state that \(\oplus\) is distributive to the left and to the right with respect to application.

**Theorem 8.7 (Left-distributivity of \(\oplus\) w.r.t. application).** If \(K\) is a stable term, then we have
\[
[[M_1 @ M_2]_T(N) = K]_T \approx [M_1 @ M_2 @ N]_T = K]_T.
\]

**Proof.** The semantics of application gives us
\[
(M_1 @ M_2)(N)0 idT = (M_1 @ M_2)0(\lambda x. N)(0(\lambda y. (x y) id))T' \circ T^\alpha,
\]
Assume that \(\text{Test}(\text{Fst}(T))\), applied to a particular \(\omega \in \Omega\) chooses \(M_i\) for some \(i \in \{1, 2\}\), in the continuation semantics of \(M_1 @ M_2\).

Let \(T' = \text{Snd}(\text{Swap}(T))\). The previous term is further equal to
\[
\langle M_1 \rangle 0(\lambda x. N)(0(\lambda y. (x y) id))T' \circ (\text{Snd}(T')) \approx (M_1 N)0 idT(Snd(T'))
\]
\[
= (M_1 N @ M_2)0 idT'.
\]

In the last line we have used the fact that if \(\text{Test}(\text{Fst}(T))\), when applied to a particular \(\omega \in \Omega\) in the semantics of \(M_1 @ M_2\), chooses \(M_i\), then the same test applied to the same \(\omega\) in the semantics of \(M_1 N @ M_2 N\) will choose \(M_i\) for the same \(i \in \{1, 2\}\).

Since \(\text{Snd}\) and \(\text{Swap}\) both preserve tossing processes (Lemma 4.7), there exists an automorphism \(\alpha \in \text{Aut}(\Omega)\) such that \(T' = T \circ \alpha\). Hence, we have
\[
[[M_1 @ M_2](N) = K]_T
\]
\[
= \{\omega \in \Omega \mid \langle M_1 @ M_2 \rangle(N)0 idT(\omega) = \langle K \rangle 0 idT(\omega)\}/Null
\]

LICS ‘18, July 9–12, 2018, Oxford, United Kingdom
\[
\{ \omega \in \Omega \mid \langle M_1 N \circ M_2 N \rangle 0 id T' (\omega) = \langle K \rangle 0 id T (\omega) \}/Null.
\]
Since \( K \) is stable, \( \langle K \rangle 0 id T (\omega) = \langle K \rangle 0 id (T' \circ \alpha (\omega)) \) and the previous set is equal to
\[
\{ \omega \in \Omega \mid \langle M_1 N \circ M_2 N \rangle 0 id T' (\omega) = \langle K \rangle 0 id T (\omega) \}/Null
= \alpha^{-1} (\{ \omega \in \Omega \mid \langle M_1 N \circ M_2 N \rangle 0 id T (\omega) = \langle K \rangle 0 id T (\omega) \}/Null)
= \alpha^{-1} (\{ M_1 N \circ M_2 N = K \}) T .
\]

**Theorem 8.8** (Right-distributivity of \( \circ \) w.r.t. application). If \( K \) is a stable term, then we have
\[
[ N (M_1 \circ M_2) = K ] T \approx [ NM_1 \circ NM_2 = K ] T .
\]

**Proof.** The proof is similar to the one for left-distributivity, except that instead of \( T' \) we use \( T'' = Swap (Snd (T)) \).

Using Lemma 8.4, we can prove the entropic equality equation.

**Theorem 8.9** (Entropic equality). If \( K \) is a stable term, then we have
\[
[ (M_1 \circ M_2) \circ (N_1 \circ N_2) = K ] T \approx [ (M_1 \circ N_1) \circ (M_2 \circ N_2) = K ] T .
\]

By exploiting the idempotence of \( \circ \) and the entropic equality, one can derive \( \circ \)-distributivity.

**Theorem 8.10** (\( \circ \)-distributivity). If \( K \) is a stable term, then we have
1. \( [ N \circ (M_1 \circ M_2) = K ] T \approx [ (N \circ M_1) \circ (N \circ M_2) = K ] T ; \)
2. \( [ (M_1 \circ M_2) \circ N = K ] T \approx [ (M_1 \circ N) \circ (M_2 \circ N) = K ] T . \)

There are also some equalities between terms that are much stronger; these are equalities that hold almost everywhere which means that they are interpreted as the top element of the Boolean algebra. One such equation is \( \lambda \)-distributivity.

**Theorem 8.11** (\( \lambda \)-distributivity w.r.t. \( \circ \)).
\[
[ \lambda x . (M_1 \circ M_2) = \lambda x . M_1 \circ \lambda x . M_2 ] T = \Omega / Null .
\]

**Proof.** From the semantics of \( \lambda \)-terms we get
\[
\langle \lambda x . (M_1 \circ M_2) \rangle 0 id T = id (\psi (\lambda v . \langle M_1 \circ M_2 \rangle (0(v/x))) (\lambda v . u) T) .
\]
Assume that Test (Fst (T)), applied to a particular \( \omega \in \Omega \), chooses \( M_i \) for some \( i \in \{1, 2\} \), when applied in the continuation semantics of \( M_1 \circ M_2 \). Then, the previous term is equal to
\[
id (\psi (\lambda v . \langle M_1 \rangle (0(v/x))) (\lambda v . u) (Snd (T))) = (\lambda x . M_1) \circ \lambda x . M_2 ) 0 id T .
\]
In the last line we have used the fact that if Test (Fst (T)), when applied to a particular \( \omega \in \Omega \) in the semantics of \( M_1 \circ M_2 \), chooses \( M_i \), then the same test applied to the same \( \omega \) in the semantics of \( \lambda x . M_1 \circ \lambda x . M_2 \) will choose \( \lambda x . M_i \) for the same \( i \in \{1, 2\} \).

**Theorem 8.12** (Order of applications). If \( N_1, N_2 \) are two stable closed terms, then
\[
[ ((\lambda x . \lambda y . M) N_1) N_2 = ((\lambda y . \lambda y . M) N_2) N_1 ] T = \Omega / Null .
\]
At this point we are ready to prove the soundness of some equations involving the fixpoint operators.

**Theorem 8.13** (Recursive application).
\[
[ \mu x . M = (\lambda x . M) (\mu x . M)] T = \Omega / Null .
\]

**Proof.** We know that, for each \( n \geq 0, M^{n+1} = (\lambda x . M)^{n} \). Hence, for arbitrary \( E, C \) and \( T \),
\[
(\mu x . M)^{n+1} E C T = (\langle \lambda x . M \rangle^{n}) E C T .
\]
Lemma 6.4 applied to this equality gives us further that
\[
(\mu x . M)^{n+1} E C T = \langle M \rangle E ((\mu x . M) E (\lambda w . w) T^E / x) C T .
\]
Hence, sup \( n (\mu x . M)^{n+1} E C T = sup \_n (\mu x . M) E (\lambda w . w) T^E / x) C T ^o \) and using Scott continuity we get
\[
sup \_n (\mu x . M)^{n+1} E C T = \langle M \rangle E (sup \_n (\mu x . M) E (\lambda w . w) T^E / x) C T ^o .
\]
Since \( M \) is continuous, hence monotonic, the above is equivalent to
\[
sup \_n (\mu x . M)^{n+1} E C T = \langle M \rangle E (sup \_n (\mu x . M) E (\lambda w . w) T^E / x) C T ^o ,
\]
or equivalently, \( \mu x . M)^{n+1} E C T = \langle M \rangle E (\mu x . M) E (\lambda w . w) T^E / x) C T ^o \) and again applying Lemma 6.4, \( (\mu x . M)^{n+1} E C T = (\langle \lambda x . M \rangle (\mu x . M)) E C T . \)
In particular, we also have
\[
[ \mu x . M ] 0 id T = (\langle \lambda x . M \rangle (\mu x . M)) 0 id T .
\]

**Theorem 8.14** (Recursive choice).
\[
[ \langle \mu x . M \rangle (\mu x . M) ] T = \Omega / Null .
\]

**Proof.** Consider arbitrary \( E, C \) and \( T \). For each \( \omega \in \Omega \), the sequence \( (\mu x . M)^{n+1} E C T (\omega) \) is increasing and its limit is \( \mu x . M E C T (\omega) \).
Consider now the unfoldings of the term \( \mu x . (x \circ M) \). They can be represented as a balanced tree structure, where the nodes are the \( M_i \) and row at depth \( k \) represents the syntax of \( (x \circ M)^{k} \), where sibling’s nodes are connected by \( \circ \). For instance, we have that
\[
(x \circ M)^0 = M^0, \quad (x \circ M)^1 = M^0 \circ M^1, \quad (x \circ M)^2 = (M^0 \circ M^0) \circ (M^1 \circ M^2), \quad (x \circ M)^3 = ((M^0 \circ M^1) \circ (M^1 \circ M^2) \circ (M^2 \circ M^3)) .
\]
For each \( \omega \in \Omega \) and each \( i \in \mathbb{N} \) there exists \( j \leq i \) such that
\[
\langle (x \circ M)^{i+1} E C T (\omega) = (\mu x . M)^j E C (Snd (T)) (\omega) \rangle .
\]
In fact, each \( \omega \in \Omega \) represents a path in this syntactic tree (e.g., 0 chooses the left branch and 1 chooses the right branch) and the intersection of this path with the \( i \)-th level of the graph (representing \( (x \circ M)^{i} \)) is exactly the term \( M_i \) satisfying the previous equation.
\[
((x \circ M)^{i+1} E C T (\omega))_{i \in \mathbb{N}} \text{ increasingly converges to } \mu x . M E C T (\omega),
\]
except for those \( \omega \) for which the sequence stabilises, i.e., for those \( \omega \in \Omega \) for which the corresponding path in the syntactic tree always chooses the left branch after a certain level. This is the set
\[
S = \{ \omega \in \Omega | \omega = v o^w, w \in \{0, 1\}^* \}.
\]
Obviously \( S \) is a null set, and this concludes our proof.

The results we have proven so far allow us to say more about \( \beta \)-reduction. Firstly we prove that an unrestricted version of \( \beta \)-reduction cannot be stated for the stochastic \( \lambda \)-calculus.

**Theorem 8.15.** There exist terms \( M, N \) (which are not stable) such that
\[
[ (\lambda x . M) (N/x) ] T \neq \Omega / Null .
\]
We can, however, have a stronger version of contradiction from the assumption that for all terms \( M, N \) we have that \([\lambda x.M](N) = M[N/x] \). For some tossing \( T \), Further, Theorema 7.2 ensures us that \([M_j(N/x) = K]_T \), which concludes our proof, since \( \approx \) is transitive.

### 9 Generating Random Numbers

In this section we present a small example of programming in stochastic \( \lambda \)-calculus and use the semantics to argue for the correctness of the program behavior. Our program takes a Church numeral \( n \) and produces a random Church numeral from 0 to \( 2^n - 1 \) with equal probability.

**Functions for Church Numerals and Booleans.** In the following, we use mathematical symbols as the names of lambda terms, for ease of notation. Note that this means arithmetical expressions are in (forward) Polish notation.

Recall that a Church numeral for the number \( n \in \mathbb{N} \) is a function of two arguments \( f \) and \( x \), returning \( f \) applied \( n \) times to \( x \). We use the well-known encodings of the arithmetic operations \( \text{succ} \), \( + \) and \( \times \).

**Picking a Random Number from 0 to 2\(^n\) − 1.** The following stochastic \( \lambda \)-term is the key of our encoding.

\[
r \text{and} = \lambda n.(\lambda x.(\text{succ}(x \times 2)))(\text{succ}(x \times 2)) \Theta.
\]

In intuitive terms, the program starts with a number equal to 0 and flips a fair coin \( n \) times, either doubling the number or doubling and adding one, depending on the outcome.

The following statement, if demonstrated, proves that the program has the desired behavior. It exemplifies how our deduction principles can be applied.

**Statement:** For all Church numerals \( n \) and tossing processes \( T \),

\[
\forall_{i=0}^{2^n-1} [r \text{and } i]_T = \Omega/Null
\]

and for all \( 0 \leq i, j \leq 2^n - 1 \)

\[
[r \text{and } n = i]_T \approx [r \text{and } n = j]_T.
\]

Therefore, for each \( 0 \leq i \leq 2^n - 1 \), \([r \text{and } n = i]_T \) are of equal probabilities and summing to 1, i.e. of probability \( 2^{-n} \).

**Sketch:** We do not provide a detailed proof of this statement, that would require further developments of the deduction principles. Instead, we sketch below how such a proof shall be organized.

Induction on \( n \): the inductive hypothesis we need is that

\[
\forall_{i=0}^n [r \text{and } n = i]_T = \Omega/Null, \text{ that for all } i \geq 2^n, \text{ we have }
\]

\[
[r \text{and } n = i]_T = \emptyset/Null. \text{ and for all } 0 \leq i, j \leq 2^n - 1, \text{ we have }
\]

\[
[r \text{and } n = i]_T \approx [r \text{and } n = j]_T.
\]

We start with the base case, \( n = 0 \). Then

\[
r \text{and } 0 = \emptyset (\lambda x.(x \times 2) \Theta(\text{succ}(x \times 2)))(\Theta) = \emptyset.
\]

We have \([r \text{and } n = 0] = \Omega/Null\). This shows the three facts we want, because \( 0 \leq i \leq 2^0 - 1 = 0 \) implies \( i = 0 \).
For the inductive step, we temporarily introduce the name $$f' = (\lambda x. ((x \times 2) \oplus (\text{succ}(x \times 2)))$$,
for the latter part of the definition of rand, excluding the final @. We start by re-expressing rand and succ n in terms of rand n.

$$\text{rand}(\text{succ } n) = (\text{succ } n)f' = f'(n \cdot f' @) = f'(\text{rand } n)$$

Let $$\text{rand}(\text{succ } n) = (\text{succ } n)f'$$

By the inductive hypothesis, we have

For the second part of the inductive hypothesis, that $$i \geq 2^{n+1}$$ implies $$\text{rand} (\text{succ } n) - i \cdot f' = \emptyset$$, we make a case split depending on whether i is even or odd. As both cases are similar, we only show the case where i is even. Then $$i = 2i'$$ for some integer $$i'$$. By the inductive hypothesis $$\text{rand } n = i' \cdot f'$$, and so by well-definedness and the fact that $$(x \times 2) = i$$, we have $$\text{rand} (\text{succ } n) = i' \cdot f' = \emptyset$$.

Finally, we need to show that for all $$0 \leq i, j \leq 2^{n+1} - 1$$ we have $$\text{rand}(\text{succ } n) = i \cdot f' \approx j \cdot f'$$.

Let $$i', j'$$ such that $$i = 2i'$$ and $$j = 2j'$$. Hence,

$$\text{rand}(\text{succ } n) = i' \cdot f' = j' \cdot f'$$.

References

A The functions pack and unpack

For a better connection with the mainstream of measure theory, we consider how to pack and unpack sequences in $2^\mathbb{N}$. In aid of this, we consider two infinite sequences in $\mathbb{N}$:

$s_0^n = 2^n - 1$ and $s_1^n = 3 \cdot 2^n - 1$.

These sequences solve the recurrence relations:

$$\begin{align*}
(s_0^0 = 0; & \quad s_0^n = 2s_0^{n-1} + 1)_{n \geq 1} \\
(s_1^0 = 2; & \quad s_1^n = 2s_1^{n-1} + 1)_{n \geq 1}.
\end{align*}$$

Therefore each sequence is even iff $n = 0$. Moreover, $s_j^i = s_{j'}^{i'}$ implies $i = i'$ and $j = j'$.

We define a family of maps $\text{unpack}_n : \mathcal{P}(\mathbb{N}) \to 2$ by

$$\text{unpack}_n(S) = \begin{cases} 
0 & \text{if } s_0^n \in S \\
1 & \text{otherwise}
\end{cases},$$

and two maps $\text{unpack} : \mathcal{P}(\mathbb{N}) \to 2^\mathbb{N}$ and $\text{pack} : 2^\mathbb{N} \to \mathcal{P}(\mathbb{N})$ as follows:

$$\text{unpack}(S)(n) = \text{unpack}_n(S) \quad \text{and} \quad \text{pack}(a) = \bigcup_{i=0}^{\infty} \{s_i^n\}.$$ 

It is not difficult to verify that the maps $\text{unpack}_n$, $\text{unpack}$ and $\text{pack}$ are (Borel) measurable, and that for all $a \in 2^\mathbb{N}$, we have

$$\text{unpack}(\text{pack}(a))(n) = a_n.$$