

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

1 Giorgio Bacci  
2  
3  
4  
5  
6

Aalborg University, Denmark

7 Radu Mardare  
8  
9

Aalborg University, Denmark

Robert Furber  
64

Aalborg University, Denmark  
65

7 Prakash Panangaden  
8  
9

McGill University, Canada  
66

Dexter Kozen  
67

Cornell University, USA  
68

7 Dana Scott  
8  
9

Carnegie Mellon University, USA  
70

## 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. In this paper 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 serve as a source of randomness. The construction of the model requires a subtle measure-theoretic analysis of the space of coin-tossing sequences. 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.

## 1 Introduction

Probabilistic programming languages [4–7, 9–11, 18] have become popular recently, sparked by renewed interest in verification and machine learning. The subject began with an imperative first-order language [9], and there has been significant recent interest in extending to higher-order functional languages [7, 10, 11, 17]. 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 that can incorporate higher-order features as well as appropriate probabilistic constructions 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 [15], the senior author 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 [14] to construct models of set

theory in order to obtain independence results. The independence of the Continuum Hypothesis was obtained by introducing a massive number 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 needed to bring this idea to fruition. Ordinary logical propositions take truth values in this Boolean algebra instead of in the simple two-element Boolean algebra.

The ideas behind the present work were outlined only briefly in [15]. The primary goal here is to develop an equational theory based on these ideas in which equations between stochastic  $\lambda$ -terms have probabilistic meaning. In this paper we flesh out these ideas and provide a continuation-passing semantics for such an equational theory. The system will provide reasoning principles for establishing 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  $\oplus$ , 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  $\mathcal{T}$ , a pair of closed terms  $M, N$  will define a set of tossing sequences where they agree  $\llbracket M = N \rrbracket_{\mathcal{T}}$ , 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  $\llbracket M = N \rrbracket_{\mathcal{T}}$  and  $\llbracket M = N \rrbracket_{\mathcal{T}'}$  for different tossing processes  $\mathcal{T}$  and  $\mathcal{T}'$ , are related by an automorphism of the Boolean algebra. We write  $\llbracket M = N \rrbracket_{\mathcal{T}} \approx \llbracket M = N \rrbracket_{\mathcal{T}'}$  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  $\llbracket M = K \rrbracket_{\mathcal{T}} \approx \llbracket N = K \rrbracket_{\mathcal{T}}$ , 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  $\llbracket M = N \rrbracket_{\mathcal{T}}$  evaluates to the top element of the Boolean algebra. Here we are using the idea that

123 a tossing process, once it has resolved the choices, makes a term  
 124 of the stochastic  $\lambda$ -calculus look like an ordinary  $\lambda$ -term. In these  
 125 cases one cannot prove that  $\llbracket M = N \rrbracket_{\mathcal{T}}$  is the top element directly,  
 126 since this might not be true, but the weaker statement above serves  
 127 to replace this statement.

128 Our main contributions are:

- 130 1. We develop a use of random variables in this framework in order  
 131 to identify a class of tossing processes that can serve as a sources  
 132 of randomness in probabilistic programs. We show that with  
 133 this choice, the semantics is invariant under automorphisms of  
 134 the measure algebra effected by remapping the tossing process.  
 135 This provides a canonical meaning to programs.
- 137 2. We introduce the stochastic  $\lambda$ -calculus by augmenting the ordinary  
 138  $\lambda$ -calculus with a probabilistic choice construct. We flesh  
 139 out the continuation-passing semantics proposed in [15], with  
 140 the crucial new observation that  $\beta$ -reduction is not sound for all  
 141 terms with probabilistic choice. To compensate, we need to in-  
 142 troduce an explicit fixpoint operator in order to have recursive  
 143 programs.
- 144 3. We develop a Boolean-valued reasoning framework for the  
 145 stochastic  $\lambda$ -calculus and prove soundness results with respect  
 146 to the continuation-passing semantics.

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

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

## 165 2 Standard Probability Spaces

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

170 **Definition 2.1.** Given a measurable space  $(X, \Sigma)$  and a probability  
 171 measure  $\mu$  on it,  $(X, \Sigma, \mu)$  is a *standard probability space* iff  $(X, \Sigma)$   
 172 is Borel isomorphic to a Polish space<sup>1</sup> equipped with its Borel  
 173 algebra<sup>2</sup>.

175 Consider, e.g., the set  $2^{\mathbb{N}}$  of infinite binary sequences with the  
 176 Cantor topology<sup>3</sup>, which has, as basic open sets, the sets  $\{\alpha \mid x < \alpha\}$ ,  
 177 where  $\alpha$  ranges over  $2^{\mathbb{N}}$ ,  $x \in \{0, 1\}^*$ , and  $<$  denotes prefix. Let  
 178  $\mathcal{B}$  be the Borel  $\sigma$ -algebra of the Cantor topology.

180 <sup>1</sup>A Polish space is the topological space induced by a complete separable metric space.

181 <sup>2</sup>The Borel algebra of a topology is the  $\sigma$ -algebra generated by the open sets.

182 <sup>3</sup>The space is the topological power of  $\omega$  copies of the discrete space  $2 = \{0, 1\}$ .

184 The (fair) *coin-flipping* probability measure<sup>4</sup>  $P$  on  $\mathcal{B}$  is generated  
 185 by its values on intervals:

$$P(\{\alpha \mid x < \alpha\}) = 2^{-|x|}.$$

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

188 We consider measure-preserving maps between standard probabil-  
 189 ity spaces. The category we use, **Meas**/0 has maps identified if they  
 190 are equal almost everywhere, *i.e.* except for on a null set.

191 **Definition 2.2.** If  $f_j : (X_j, \Sigma_j, \mu_j) \rightarrow (Z, \Xi, \xi)$ ,  $j = 1, 2$  are two  
 192 measure-preserving maps with common codomain, we say  $f_1 \cong f_2$   
 193 or  $(X_1, f_1) \cong (X_2, f_2)$  if there exists a measure-preserving isomor-  
 194 phism  $i : X_1 \rightarrow X_2$  such that  $f_2 \circ i = f_1$ , except on a subset of  $X_1$   
 195 of measure 0.

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

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

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

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

213 The following theorem provides a useful characterization of the  
 214 concept of monolithic map<sup>5</sup>

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

220 We conclude this section with a useful result regarding the compo-  
 221 sition of monolithic maps.

222 **Lemma 2.5.** Let  $f : (X, \Sigma, \mu) \rightarrow (Y, \Theta, \nu)$  and  $g : (Y, \Theta, \nu) \rightarrow$   
 223  $(Z, \Xi, \xi)$  be measure-preserving maps. If  $S \in \Sigma$  is 1-sheeted with  
 224 respect to  $g \circ f$ , then it is 1-sheeted with respect to  $f$ . Therefore iff  
 225  $f$  is monolithic, then  $g \circ f$  is monolithic.

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

$$S \cap f^{-1}(y) \subseteq S \cap f^{-1}(g^{-1}(g(y)))$$

228 so  $f^{-1}(y) \cap S$  has cardinality at most 1, so  $S$  is 1-sheeted with respect  
 229 to  $f$ .

230 <sup>4</sup>This is the Haar measure on  $2^{\mathbb{N}}$  as a compact group based on *mod 2* addition.

231 <sup>5</sup>This concept is formulated in terms of *decompositions* in [13, §3.1], which are better  
 232 known as *disintegrations* [2, §452 E].



random variables  $\xi, \eta \in \mathcal{R}(\Omega)$  can be interpreted as the *measurable event*

$$[\xi = \eta] = \{\omega \in \Omega \mid \forall n \in \xi(\omega).n \in \eta(\omega) \wedge \forall n \in \eta(\omega).n \in \xi(\omega)\},$$

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  $\Omega = (2^{\mathbb{N}}, \mathcal{B}, P)$ , we define the *algebra of events* (also known as the *measure algebra*) as the Boolean algebra

$$\mathcal{B}/_{\text{Null}} = (\mathcal{B}/_{\text{Null}}, \cup, \cap, \sim, \emptyset/_{\text{Null}}, \Omega/_{\text{Null}})$$

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  $\mathcal{B}/_{\text{Null}}$ : 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  $\mathcal{B}/_{\text{Null}}$ .

We call the elements of  $\mathcal{B}/_{\text{Null}}$  events and for  $A \in \mathcal{B}$  we denote its equivalence class by  $A/_{\text{Null}}$ .

**Theorem 3.4.**  $(\mathcal{B}/_{\text{Null}}, \cup, \cap, \sim, \emptyset/_{\text{Null}}, \Omega/_{\text{Null}})$  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.

$\mathcal{B}/_{\text{Null}}$  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 [15], 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  $\text{pack} : 2^{\mathbb{N}} \rightarrow \mathcal{P}(\mathbb{N})$  which encodes a sequence as a set in a way that can be easily inverted and its inverse, called **unpack**, is Borel-measurable as well.

Moreover, these can be defined so that they properly relate the well-known operations on sequences  $\text{head} : 2^{\mathbb{N}} \rightarrow 2$  and  $\text{tail} : 2^{\mathbb{N}} \rightarrow 2^{\mathbb{N}}$ , 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<sup>6</sup>.

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

<sup>6</sup>A detailed construction of a pair of such coding/decoding functions can be found in the appendix.

$$1. \{\alpha\} = \text{Fst}(\text{pack}(a)) \text{ iff } \text{head}(a) = \alpha.$$

$$2. \text{Snd}(\text{pack}(a)) = \text{pack}(\text{tail}(a)).$$

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

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

An *independent sequence of coin tosses* (ICS) is a random variable  $\mathcal{T} : \Omega \rightarrow \mathcal{P}(\mathbb{N})$  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  $\mathcal{T} : \Omega \rightarrow \mathcal{P}(\mathbb{N})$  is of the form  $\mathcal{T} = \text{pack} \circ T$  for some (Borel) measurable map  $T : \Omega \rightarrow 2^{\mathbb{N}}$ . By Lemma 4.1,  $\text{Fst} \circ \text{pack} \circ T = \text{head} \circ T$ , so the condition of  $\text{Fst}(\mathcal{T})$  being a fair coin is that for  $i \in \{0, 1\}$ ,

$$P((\text{head} \circ T)^{-1}(i)) = \frac{1}{2},$$

That  $\text{Snd}(\mathcal{T})$  is an ICS implies that for all  $n \in \mathbb{N}$  and  $i \in \{0, 1\}$ ,

$$P((\text{ev}_n \circ T)^{-1}(i)) = \frac{1}{2},$$

where  $\text{ev}_n : 2^{\mathbb{N}} \rightarrow \{0, 1\}$  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  $\{0, 1\}$ ,

$$P\left(\bigcap_{i=1}^n (\text{ev}(m_i) \circ T)^{-1}(b_i)\right) = 2^{-n}.$$

This means that the image measure  $T_*(P) = P \circ T^{-1}$  agrees with the standard coin-flipping measure on  $2^{\mathbb{N}}$  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  $2^{\mathbb{N}}$ ,  $T_*(P)$  and the coin-flipping measure agree on all Borel sets [19, Lemma 1.6], so we have that  $T$  is measure-preserving from  $\Omega$  to  $2^{\mathbb{N}}$ .

Conversely, for any measure-preserving map  $T : \Omega \rightarrow 2^{\mathbb{N}}$ ,  $\text{pack} \circ T$  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  $\text{pack} \circ T$ , for some measure-preserving map  $T : \Omega \rightarrow \Omega$ .*

Now we focus on another important function for our discussion,

$$\text{evens} : 2^{\mathbb{N}} \rightarrow 2^{\mathbb{N}},$$

defined by  $\text{evens}(a)(n) = a_{2n}$ .

**Theorem 4.4.** *A measure-preserving map  $T : 2^{\mathbb{N}} \rightarrow 2^{\mathbb{N}}$  is monolithic iff  $T \cong \text{evens}$ .*

*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  $(2^{\mathbb{N}} \times 2^{\mathbb{N}}, \pi_2) \cong (2^{\mathbb{N}}, \text{evens})$ . 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.  $\square$

489 **4.2 Tossing Processes** 550490 We are ready to define the concept of tossing process; the crucial  
491 point is to insist on a monolithic function.492 **Definition 4.5** (Tossing Process). A *tossing process* is an independent  
493 sequence of coin tosses  $\mathcal{T} = \text{pack} \circ T$ , where  $T : 2^{\mathbb{N}} \rightarrow 2^{\mathbb{N}}$  is  
494 a monolithic measure-preserving map.495 We denote by  $\text{Toss}$  the set of tossing processes. The next theorem  
496 states that all tossing processes are the same up to a measure-  
497 preserving automorphism of the measure algebra. Let  $\text{Aut}(\Omega)$  de-  
498 note the set of measure-preserving automorphisms on  $\Omega$ .499 **Theorem 4.6** (Representation Theorem). *For any two tossing pro-  
500 cesses  $\mathcal{S}$  and  $\mathcal{T}$  there exists a measure-preserving automorphism  
501  $\alpha : \Omega \rightarrow \Omega \in \text{Aut}(\Omega)$  such that  $\mathcal{T} = \mathcal{S} \circ \alpha$  except for on a subset of  
502  $\Omega$  of measure 0.*503 **4.3 Tossing Process Operators** 563504 Before concluding this section, we show a few useful operators on  
505  $\mathcal{R}(\Omega)$  that are closed on tossing processes.

506 Let

507 
$$(\cdot)^e, (\cdot)^o, \text{Swap} : \text{Toss} \rightarrow \mathcal{R}(\Omega),$$
 567

508 defined for an arbitrary tossing process  $\mathcal{T} = \text{pack} \circ T$  as follows.

509 
$$\mathcal{T}^e = \text{pack} \circ \text{evens} \circ T$$
 568

510 
$$\mathcal{T}^o = \text{pack} \circ \text{odds} \circ T$$
 569

511 
$$\text{Swap}(\mathcal{T}) = \text{pack} \circ \text{swap} \circ T,$$
 570

512 where  $\text{odds}, \text{swap} : \Omega \rightarrow \Omega$  are defined by  $\text{odds}(a)(n) = a_{2n+1}$  and  
513  $\text{swap}(a)(2n) = a_{2n+1}, \text{swap}(a)(2n+1) = a_{2n}.$  571514 **Lemma 4.7.** *If  $\mathcal{T}$  is a tossing process, then  $\text{Snd}(\mathcal{T}), \mathcal{T}^e, \mathcal{T}^o$  and  
515  $\text{Swap}(\mathcal{T})$  are tossing processes as well.* 572527 **5 Stochastic  $\lambda$ -Calculus** 580528 In this section we introduce the stochastic  $\lambda$ -calculus. The syntax  
529 of the stochastic  $\lambda$ -calculus over a set  $V \ni x$  of variables extends  
530 the syntax of the classical untyped  $\lambda$ -calculus with a (binary) prob-  
531 abilistic choice operator  $\oplus$  on  $\lambda$ -terms, and a fixpoint combinator  
532  $\mu x.M$ :

533 
$$M := x \mid \lambda x.M \mid MM \mid \mu x.M \mid M \oplus M.$$
 573

534 **Notation:** In what follows we will call terms without any occur-  
535 rence of  $\oplus$  *stable terms*; however, a stable term may contain the  
536 fixpoint operator. And, as usual, we use  $M\{N/x\}$  to denote the  
537 substitution of the variable  $x$  by the term  $N$  in  $M.$  574538 Unlike in the classical  $\lambda$ -calculus, in the stochastic  $\lambda$ -calculus the fix-  
539 point combinator cannot be defined from the other operators, as we  
540 will demonstrate later. For the development of the fixpoint operator,  
541 it is useful to define the *unfolding* of a recursive term.542 Given  $\mu x.M$ , its *unfolding* is the sequence of terms  $M^0, M^1, \dots, M^n, \dots$   
543 defined inductively as follows.

544 
$$M_0 = (\lambda x.xx)(\lambda x.xx), \text{ and for arbitrary } n, M^{n+1} = (\lambda x.M)M^n.$$
 575

549 **6 Probabilistic Continuation Semantics** 550551 The continuation semantics for the stochastic  $\lambda$ -calculus interprets  
552 a  $\lambda$ -term relative to an *environment* giving values to the free vari-  
553 ables, a *continuation* giving a subsequent computation, and a *tossing*  
554 *process*  $\mathcal{T}$  used to resolve probabilistic choices.555 We know that the set  $\mathcal{R}(\Omega)$  of random variables with the point-  
556 wise order forms a domain such that  $[\mathcal{R}(\Omega) \rightarrow \mathcal{R}(\Omega)]$  is a con-  
557 tinuous retract of  $\mathcal{R}(\Omega)$ , where  $[\mathcal{R}(\Omega) \rightarrow \mathcal{R}(\Omega)]$  is the space of  
558 Scott-continuous functions. We write this explicitly by introducing  
559 the functions

560 
$$\phi : \mathcal{R}(\Omega) \rightarrow [\mathcal{R}(\Omega) \rightarrow \mathcal{R}(\Omega)] \text{ and } \psi : [\mathcal{R}(\Omega) \rightarrow \mathcal{R}(\Omega)] \rightarrow \mathcal{R}(\Omega).$$
 561

562 It is useful to also define the direct and the continuation-passing  
563 semantics for stable  $\lambda$ -terms. We use  $x$  to range over variables,  $E$  to  
564 range over arbitrary environments,  $C$  to range over arbitrary contin-  
565 uations, and  $\mathcal{T}$  to range over arbitrary tossing processes. As before,  
566 the  $n$ -th unfolding of the term  $\mu x.M$  is denoted by  $M^n.$  567568 Let  $\langle \cdot \rangle$  denote the direct semantics and  $\langle\langle \cdot \rangle \rangle$  the continuation-passing  
569 semantics for stable  $\lambda$ -terms. In addition, for stochastic terms, we  
570 denote by  $\langle|\cdot|\rangle$  the continuation-passing semantics augmented with  
571 a tossing process.

572 
$$\begin{aligned} \langle \cdot \rangle &: \text{Term} \rightarrow \text{Env} \rightarrow \mathcal{R}(\Omega) & 573 \\ \langle\langle \cdot \rangle \rangle &: \text{Term} \rightarrow \text{Env} \rightarrow \text{Cont} \rightarrow \mathcal{R}(\Omega) & 574 \\ \langle|\cdot|\rangle &: \text{Term} \rightarrow \text{Env} \rightarrow \text{Cont} \rightarrow \text{Toss} \rightarrow \mathcal{R}(\Omega). & 575 \end{aligned}$$

576 where

577 
$$\text{Env} = \text{Var} \rightarrow \mathcal{R}(\Omega) \quad \text{Cont} = [\mathcal{R}(\Omega) \rightarrow \mathcal{R}(\Omega)]$$
 578

579 The **direct semantics** is

580 
$$\begin{aligned} \langle x \rangle E &= E(x) & 581 \\ \langle MN \rangle E &= \phi(\langle M \rangle E)(\langle N \rangle E) & 582 \\ \langle \lambda x . M \rangle E &= \psi(\lambda v . \langle M \rangle (E\{v/x\})) & 583 \\ \langle \mu x . M \rangle E &= \sup_n \langle M^n \rangle E. & 584 \end{aligned}$$

585 With **continuations**, define

586 
$$\begin{aligned} \langle x \rangle EC &= C(E(x)) & 587 \\ \langle\langle MN \rangle \rangle EC &= \langle\langle M \rangle \rangle E(\lambda a . \langle\langle N \rangle \rangle E(\lambda b . C(\phi(a)b))) & 588 \\ \langle\langle \lambda x . M \rangle \rangle EC &= C(\psi(\lambda v . \langle M \rangle (E\{v/x\})(\lambda u . u))) & 589 \\ \langle\langle \mu x . M \rangle \rangle EC &= \sup_n \langle\langle M^n \rangle \rangle EC. & 590 \end{aligned}$$

591 The **probabilistic continuation** is defined as follows.

592 
$$\begin{aligned} \langle|x|\rangle ECT &= C(E(x)) & 593 \\ \langle|M|\rangle ECT &= \langle|M|\rangle E(\lambda a . \langle|N|\rangle E(\lambda b . C(\phi(a)b))\mathcal{T}^e)\mathcal{T}^o & 594 \\ \langle|\lambda x . M|\rangle ECT &= C(\psi(\lambda v . \langle|M|\rangle (E\{v/x\})(\lambda u . u))\mathcal{T}) & 595 \\ \langle|M \oplus N|\rangle ECT &= \text{Test}(\text{Fst}(\mathcal{T}))(\langle|M|\rangle EC(\text{Snd}(\mathcal{T})))(\langle|N|\rangle EC(\text{Snd}(\mathcal{T}))) & 596 \\ \langle|\mu x . M|\rangle ECT &= \sup_n \langle|M^n|\rangle ECT & 597 \end{aligned}$$

598 The relation between the three semantics for stable terms is stated  
599 in the following proposition.600 **Proposition 6.1.** *If  $M$  is a stable term, then for an arbitrary environ-  
601 ment  $E$ , an arbitrary continuation process  $C$ , and an arbitrary  
602 tossing process  $\mathcal{T}$ ,*

603 
$$C(\langle M \rangle E) = \langle\langle M \rangle \rangle EC = \langle|M|\rangle ECT.$$
 604

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  $\mathcal{T}, \mathcal{T}'$ ,*

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

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  $\mathcal{T}$ , 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.  $\langle (\lambda x.M)N \rangle ECT = \sup_n \langle M^n N \rangle ECT$ ;

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

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

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  $\mathcal{T}, \mathcal{T}'$ , there exists a measure-preserving automorphism  $\alpha : \Omega \rightarrow \Omega \in \text{Aut}(\Omega)$  such that*

$$\{ \omega \in \Omega \mid \langle M \rangle ECT(\omega) = \langle M \rangle EC(\mathcal{T}' \circ \alpha)(\omega) \} / \text{Null} = \Omega / \text{Null}.$$

## 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  $\mathcal{B}/\text{Null}$  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 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  $\mathcal{T}$ , let

$$\llbracket M = N \rrbracket_{\mathcal{T}} = \{ \omega \in \Omega \mid \langle M \rangle \emptyset id \mathcal{T}(\omega) = \langle N \rangle \emptyset id \mathcal{T}(\omega) \} / \text{Null}.$$

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

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

$$A \approx A'$$

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

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

$$\alpha^{-1}(\llbracket M = N \rrbracket_{\mathcal{T}}) = \llbracket M = N \rrbracket_{\mathcal{T}'},$$

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

$$\llbracket M = N \rrbracket_{\mathcal{T}} \approx \llbracket M = N \rrbracket_{\mathcal{T}'}.$$

$$\begin{aligned} \text{Proof. } \llbracket M = N \rrbracket_{\mathcal{T}'} &= \{ \omega \in \Omega \mid \langle M \rangle \emptyset id \mathcal{T}'(\omega) = \langle N \rangle \emptyset id \mathcal{T}'(\omega) \} / \text{Null} \\ &= \{ \omega \in \Omega \mid \langle M \rangle \emptyset id(\mathcal{T} \circ \alpha)(\omega) = \langle N \rangle \emptyset id(\mathcal{T} \circ \alpha)(\omega) \} / \text{Null}. \end{aligned}$$

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

$$\begin{aligned} &\{ \omega \in \Omega \mid \langle M \rangle \emptyset id \mathcal{T}(\alpha(\omega)) = \langle N \rangle \emptyset id \mathcal{T}(\alpha(\omega)) \} / \text{Null} \\ &= \alpha^{-1}(\{ \omega \in \Omega \mid \langle M \rangle \emptyset id \mathcal{T}(\omega) = \langle N \rangle \emptyset id \mathcal{T}(\omega) \} / \text{Null}) \\ &= \alpha^{-1}(\llbracket M = N \rrbracket_{\mathcal{T}}). \end{aligned}$$

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 Lemma 6.1.

In what follows, all the terms in expressions of the form  $\llbracket M = N \rrbracket_{\mathcal{T}}$  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  $\llbracket M = N \rrbracket_{\mathcal{T}} = \Omega / \text{Null}$ , then for any closed stable term  $K$ ,*

$$\llbracket M = K \rrbracket_{\mathcal{T}} = \llbracket N = K \rrbracket_{\mathcal{T}}.$$

*Proof.* We have that

$$\llbracket M = N \rrbracket_{\mathcal{T}} = \{ \omega \in \Omega \mid \langle M \rangle \emptyset id \mathcal{T}(\omega) = \langle N \rangle \emptyset id \mathcal{T}(\omega) \} / \text{Null}.$$

This means that  $\llbracket M = N \rrbracket_{\mathcal{T}} = \Omega / \text{Null}$  implies that  $\langle M \rangle \emptyset id \mathcal{T}$  and  $\langle N \rangle \emptyset id \mathcal{T}$  are equal almost everywhere. But then,

$$\begin{aligned} \llbracket M = K \rrbracket_{\mathcal{T}} &= \{ \omega \in \Omega \mid \langle M \rangle \emptyset id \mathcal{T}(\omega) = \langle K \rangle \emptyset id \mathcal{T}(\omega) \} / \text{Null} \\ &= \{ \omega \in \Omega \mid \langle N \rangle \emptyset id \mathcal{T}(\omega) = \langle K \rangle \emptyset id \mathcal{T}(\omega) \} / \text{Null} = \llbracket N = K \rrbracket_{\mathcal{T}} \end{aligned}$$

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  $\mathcal{R}(\Omega)$  is a model of the usual untyped  $\lambda$ -calculus [15]. 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*

$$\llbracket \lambda x. M = \lambda y. M\{y/x\} \rrbracket_{\mathcal{T}} = \Omega_{/Null}.$$

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,*

$$\llbracket (\lambda x. M)(N) = M\{N/x\} \rrbracket_{\mathcal{T}} = \Omega_{/Null}.$$

Next we prove a series of results regarding the properties of the probabilistic choice operator. It is useful to start with the following lemma.

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

$$\llbracket M_1 \oplus M_2 = K \rrbracket_{\mathcal{T}} \approx \llbracket N_1 \oplus N_2 = K \rrbracket_{\mathcal{T}}.$$

*Proof.* The key observation is that there exists a  $j \in \{1, 2\}$  such that

$$\llbracket M_1 \oplus M_2 \rrbracket \emptyset id \mathcal{T} = \llbracket M_j \rrbracket \emptyset id \text{Snd}(\mathcal{T})$$

and at the same time

$$\llbracket N_1 \oplus N_2 \rrbracket \emptyset id \mathcal{T} = \llbracket N_j \rrbracket \emptyset id \text{Snd}(\mathcal{T}).$$

Also, since  $K$  is stable,  $\llbracket K \rrbracket \emptyset id \mathcal{T} = \llbracket K \rrbracket \emptyset id \text{Snd}(\mathcal{T})$ . Hence,

$$\begin{aligned} \llbracket M_1 \oplus M_2 = K \rrbracket_{\mathcal{T}} \\ = \{ \omega \in \Omega \mid \llbracket M_1 \oplus M_2 \rrbracket \emptyset id \mathcal{T}(\omega) = \llbracket K \rrbracket \emptyset id \mathcal{T}(\omega) \} / Null \\ = \{ \omega \in \Omega \mid \llbracket M_j \rrbracket \emptyset id \text{Snd}(\mathcal{T})(\omega) = \llbracket K \rrbracket \emptyset id \text{Snd}(\mathcal{T})(\omega) \} / Null \\ = \llbracket M_j = K \rrbracket_{\text{Snd}(\mathcal{T})}. \end{aligned}$$

Similarly,

$$\llbracket N_1 \oplus N_2 = K \rrbracket_{\mathcal{T}} = \llbracket N_j = K \rrbracket_{\text{Snd}(\mathcal{T})}.$$

Now we use the hypothesis that guarantees that

$$\llbracket M_j = K \rrbracket_{\text{Snd}(\mathcal{T})} \approx \llbracket N_j = K \rrbracket_{\text{Snd}(\mathcal{T})}.$$

□

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*

$$\llbracket M \oplus N = K \rrbracket_{\mathcal{T}} \approx \llbracket N \oplus M = K \rrbracket_{\mathcal{T}}.$$

*Proof.* The map  $\text{neg} : 2^{\mathbb{N}} \rightarrow 2^{\mathbb{N}}$ , where  $\text{neg}(a)(0) = 1 - a(0)$  and  $\text{neg}(a)(i) = a(i)$  for  $i > 0$  is measurable and measure-preserving. The tossing process  $\mathcal{T}$  is of the form  $\text{pack} \circ T$  for some measure-preserving monolithic map  $T : \Omega \rightarrow 2^{\mathbb{N}}$ . By Lemma 2.5,  $\mathcal{T}' = \text{pack} \circ \text{neg} \circ T$  is also a tossing process, and so by Theorem 4.6 there

exists a measure-preserving automorphism  $\alpha : \Omega \rightarrow \Omega$  such that  $\mathcal{T}' = \alpha \circ \mathcal{T}$ . We start by observing that

$$\begin{aligned} \llbracket M \oplus N \rrbracket \emptyset id \mathcal{T} \\ = \text{Test}(\text{Fst}(\mathcal{T}))(\llbracket M \rrbracket \emptyset id \text{Snd}(\mathcal{T}))(\llbracket N \rrbracket \emptyset id \text{Snd}(\mathcal{T})) \\ = \text{Test}(\text{Fst}(\mathcal{T}))(\llbracket N \rrbracket \emptyset id \text{Snd}(\mathcal{T}))(\llbracket M \rrbracket \emptyset id \text{Snd}(\mathcal{T})) \\ = \llbracket N \oplus M \rrbracket \emptyset id \mathcal{T}' \\ = \llbracket N \oplus M \rrbracket \emptyset id (\mathcal{T} \circ \alpha), \end{aligned}$$

using the fact that  $\text{Snd}(\mathcal{T}') = \text{Snd}(\mathcal{T})$ , since neg only affects the first part. So

$$\begin{aligned} \llbracket M \oplus N = K \rrbracket_{\mathcal{T}} \\ = \{ \omega \in \Omega \mid \llbracket M \oplus N \rrbracket \emptyset id \mathcal{T}(\omega) = \llbracket K \rrbracket \emptyset id \mathcal{T}(\omega) \} / Null \\ = \{ \omega \in \Omega \mid \llbracket N \oplus M \rrbracket \emptyset id \mathcal{T}(\alpha(\omega)) = \llbracket K \rrbracket \emptyset id \mathcal{T}(\omega) \} / Null \\ = \{ \omega \in \Omega \mid \llbracket N \oplus M \rrbracket \emptyset id \mathcal{T}(\alpha(\omega)) = \llbracket K \rrbracket \emptyset id \mathcal{T}(\alpha(\omega)) \} / Null \\ = \alpha^{-1}(\{ \omega \in \Omega \mid \llbracket N \oplus M \rrbracket \emptyset id \mathcal{T}(\omega) = \llbracket K \rrbracket \emptyset id \mathcal{T}(\omega) \}) / Null \\ = \alpha^{-1}(\llbracket N \oplus M = K \rrbracket), \end{aligned}$$

where the third equality holds because  $\llbracket K \rrbracket$  does not depend on  $\omega$  since it is stable. □

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*

$$\llbracket M \oplus M = K \rrbracket_{\mathcal{T}} \approx \llbracket M = K \rrbracket_{\mathcal{T}}.$$

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*

$$\llbracket (M_1 \oplus M_2)(N) = K \rrbracket_{\mathcal{T}} \approx \llbracket M_1 N \oplus M_2 N = K \rrbracket_{\mathcal{T}}.$$

*Proof.* The semantics of application gives us

$$\llbracket (M_1 \oplus M_2)(N) \rrbracket \emptyset id \mathcal{T} = \llbracket M_1 \oplus M_2 \rrbracket \emptyset (\lambda x. \llbracket N \rrbracket \emptyset (\lambda y. (xy id)) \mathcal{T}^e) \mathcal{T}^o.$$

Assume that  $\text{Test}(\text{Fst}(\mathcal{T}))$ , applied to a particular  $\omega \in \Omega$  chooses  $M_i$  for some  $i \in \{1, 2\}$ , in the continuation semantics of  $M_1 \oplus M_2$ .

Let  $\mathcal{T}' = \text{Snd}(\text{Swap}(\mathcal{T}))$ . The previous term is further equal to

$$\begin{aligned} \llbracket M_i \rrbracket \emptyset (\lambda x. \llbracket N \rrbracket \emptyset (\lambda y. (xy id)) \mathcal{T}^e) (\text{Snd}(\mathcal{T}^o)) \\ = \llbracket M_i N \rrbracket \emptyset id (\text{Snd}(\mathcal{T}')) \\ = \llbracket M_1 N \oplus M_2 N \rrbracket \emptyset id \mathcal{T}'. \end{aligned}$$

In the last line we have used the fact that if  $\text{Test}(\text{Fst}(\mathcal{T}))$ , when applied to a particular  $\omega \in \Omega$  in the semantics of  $M_1 \oplus M_2$ , chooses  $M_i$ , then the same test applied to the same  $\omega$  in the semantics of  $M_1 N \oplus M_2 N$  will choose  $M_i N$  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  $\mathcal{T}' = \mathcal{T} \circ \alpha$ . Hence, we have

$$\begin{aligned} \llbracket (M_1 \oplus M_2)(N) = K \rrbracket_{\mathcal{T}} \\ = \{ \omega \in \Omega \mid \llbracket (M_1 \oplus M_2)N \rrbracket \emptyset id \mathcal{T}(\omega) = \llbracket K \rrbracket \emptyset id \mathcal{T}(\omega) \} / Null \\ = \{ \omega \in \Omega \mid \llbracket M_1 N \oplus M_2 N \rrbracket \emptyset id \mathcal{T}'(\omega) = \llbracket K \rrbracket \emptyset id \mathcal{T}(\omega) \} / Null \\ \text{Since } K \text{ is stable, } \llbracket K \rrbracket \emptyset id \mathcal{T}(\omega) = \llbracket K \rrbracket \emptyset id (\mathcal{T} \circ \alpha)(\omega) \text{ and the previous set is equal to} \\ = \{ \omega \in \Omega \mid \llbracket M_1 N \oplus M_2 N \rrbracket \emptyset id \mathcal{T}'(\omega) = \llbracket K \rrbracket \emptyset id \mathcal{T}'(\omega) \} / Null \end{aligned}$$

$$\begin{aligned}
&= \alpha^{-1}(\{\omega \in \Omega \mid (M_1 N \oplus M_2 N) \emptyset id\mathcal{T}(\omega) = (K) \emptyset id\mathcal{T}(\omega)\} / \text{Null}) \\
&= \alpha^{-1}(\llbracket M_1 N \oplus M_2 N = K \rrbracket_{\mathcal{T}}).
\end{aligned}$$

□

**Theorem 8.8** (Right-distributivity of  $\oplus$  w.r.t. application). *If  $K$  is a stable term, then we have*

$$\llbracket N(M_1 \oplus M_2) = K \rrbracket_{\mathcal{T}} \approx \llbracket NM_1 \oplus NM_2 = K \rrbracket_{\mathcal{T}}.$$

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

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*

$$\llbracket (M_1 \oplus M_2) \oplus (N_1 \oplus N_2) = K \rrbracket_{\mathcal{T}} \approx \llbracket (M_1 \oplus N_1) \oplus (M_2 \oplus N_2) = K \rrbracket_{\mathcal{T}}.$$

By exploiting the idempotence of  $\oplus$  and the entropic equality, one can derive  $\oplus$ -distributivity.

**Theorem 8.10** ( $\oplus$ -distributivity). *If  $K$  is a stable term, then we have*

1.  $\llbracket N \oplus (M_1 \oplus M_2) = K \rrbracket_{\mathcal{T}} \approx \llbracket (N \oplus M_1) \oplus (N \oplus M_2) = K \rrbracket_{\mathcal{T}}$ ;
2.  $\llbracket (M_1 \oplus M_2) \oplus N = K \rrbracket_{\mathcal{T}} \approx \llbracket (M_1 \oplus N) \oplus (M_2 \oplus N) = K \rrbracket_{\mathcal{T}}$ .

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.  $\oplus$ ).

$$\llbracket \lambda x.(M_1 \oplus M_2) = \lambda x.M_1 \oplus \lambda x.M_2 \rrbracket_{\mathcal{T}} = \Omega / \text{Null}.$$

*Proof.* From the semantics of  $\lambda$ -terms we get

$$\llbracket \lambda x.(M_1 \oplus M_2) \rrbracket \emptyset id\mathcal{T} = id(\psi(\lambda v. (M_1 \oplus M_2)(\emptyset\{v/x\})(\lambda u. u)\mathcal{T})).$$

Assume that  $\text{Test}(\text{Fst}(\mathcal{T}))$ , applied to a particular  $\omega \in \Omega$  chooses  $M_i$  for some  $i \in \{1, 2\}$ , when applied in the continuation semantics of  $M_1 \oplus M_2$ . Then, the previous term is equal to

$$\begin{aligned}
id(\psi(\lambda v. (M_i)(\emptyset\{v/x\})(\lambda u. u)(\text{Snd}(\mathcal{T})))) &= \llbracket \lambda x.M_i \rrbracket \emptyset id(\text{Snd}(\mathcal{T})) \\
&= \llbracket \lambda x.M_1 \oplus \lambda x.M_2 \rrbracket \emptyset id\mathcal{T}.
\end{aligned}$$

In the last line we have used the fact that if  $\text{Test}(\text{Fst}(\mathcal{T}))$ , when applied to a particular  $\omega \in \Omega$  in the semantics of  $M_1 \oplus M_2$ , chooses  $M_i$ , then the same test applied to the same  $\omega$  in the semantics of  $\lambda x.M_1 \oplus \lambda x.M_2$  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 close terms, then*

$$\llbracket ((\lambda x.\lambda y.M)N_1)N_2 = ((\lambda y.\lambda x.M)N_2)N_1 \rrbracket_{\mathcal{T}} = \Omega / \text{Null}.$$

At this point we are ready to prove the soundness of some equations involving the fixpoint operators.

**Theorem 8.13** (Recursive application).

$$\llbracket \mu x.M = (\lambda x.M)(\mu x.M) \rrbracket_{\mathcal{T}} = \Omega / \text{Null}.$$

*Proof.* We know that, for each  $n \geq 0$ ,  $M^{n+1} = (\lambda x.M)M^n$ . Hence, for arbitrary  $E, C$  and  $\mathcal{T}$ ,

$$\llbracket M^{n+1} \rrbracket ECT = \llbracket (\lambda x.M)M^n \rrbracket ECT.$$

Lemma 6.4 applied to this equality gives us further that

$$\llbracket M^{n+1} \rrbracket ECT = \llbracket M \rrbracket E\{\llbracket M^n \rrbracket E(\lambda w.w)\mathcal{T}^e/x\}CT^o.$$

Hence,  $\sup_n \llbracket M^{n+1} \rrbracket ECT = \sup_n \llbracket M \rrbracket E\{\llbracket M^n \rrbracket E(\lambda w.w)\mathcal{T}^e/x\}CT^o$  and using Scott continuity we get

$$\sup_n \llbracket M^{n+1} \rrbracket ECT = \llbracket M \rrbracket E\{\sup_n \llbracket M^n \rrbracket E(\lambda w.w)\mathcal{T}^e/x\}CT^o.$$

Since  $M$  is continuous, hence monotonic, the above is equivalent to

$$\sup_n \llbracket M^n \rrbracket ECT = \llbracket M \rrbracket E\{\sup_n \llbracket M^n \rrbracket E(\lambda w.w)\mathcal{T}^e/x\}CT^o,$$

or equivalently,  $\llbracket \mu x.M \rrbracket ECT = \llbracket M \rrbracket E\{\llbracket \mu x.M \rrbracket E(\lambda w.w)\mathcal{T}^e/x\}CT^o$  and again applying Lemma 6.4,  $\llbracket \mu x.M \rrbracket ECT = \llbracket (\lambda x.M)(\mu x.M) \rrbracket ECT$ . In particular, we also have

$$\llbracket \mu x.M \rrbracket \emptyset id\mathcal{T} = \llbracket (\lambda x.M)(\mu x.M) \rrbracket \emptyset id\mathcal{T}.$$

□

**Theorem 8.14** (Recursive choice).

$$\llbracket \mu x.(x \oplus M) \equiv \mu x.M \rrbracket_{\mathcal{T}} = \Omega / \text{Null}.$$

*Proof.* Consider arbitrary  $E, C$  and  $\mathcal{T}$ . For each  $\omega \in \Omega$ , the sequence  $\llbracket M^i \rrbracket ECT(\omega)$  is increasing and its limit is  $\llbracket \mu x.M \rrbracket ECT(\omega)$ .

Consider now the unfoldings of the term  $\mu x.(x \oplus M)$ . 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  $(x \oplus M)^k$ , where sibling's nodes are connected by  $\oplus$ . For instance, we have that

$$(x \oplus M)^0 = M^0,$$

$$(x \oplus M)^1 = M^0 \oplus M^1,$$

$$(x \oplus M)^2 = (M^0 \oplus M^1) \oplus (M^1 \oplus M^2),$$

$$(x \oplus M)^3 = ((M^0 \oplus M^1) \oplus (M^1 \oplus M^2)) \oplus ((M^1 \oplus M^2) \oplus (M^2 \oplus M^3)), \text{ etc.}$$

For each  $\omega \in \Omega$  and each  $i \in \mathbb{N}$  there exists  $j \leq i$  such that

$$\llbracket (x \oplus M)^i \rrbracket ECT(\omega) = \llbracket M^j \rrbracket EC(\text{Snd}^i(\mathcal{T}))(\omega).$$

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  $(x \oplus M)^i$ ) is exactly the term  $M^j$  satisfying the previous equation.

$((\llbracket (x \oplus M)^i \rrbracket ECT(\omega))_{i \in \mathbb{N}}$  increasingly converges to  $\llbracket \mu x.M \rrbracket ECT(\omega)$ , 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

$$S = \{\omega \in \Omega \mid \omega = v0^*, w \in \{0, 1\}^*\}.$$

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*

$$\llbracket (\lambda x.M)(N) = M\{N/x\} \rrbracket_{\mathcal{T}} \neq \Omega / \text{Null}.$$

916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976

977 *Proof.* We exploit the results in Theorems 8.5 - 8.3 to derive a contradiction from the assumption that for all terms  $M, N$  we have that  $\llbracket (\lambda x. M)(N) = M\{N/x\} \rrbracket_{\mathcal{T}} = \Omega/_{\text{Null}}$ .

978 Consider the following terms

$$981 \quad \top = \lambda x. \lambda y. x, \perp = \lambda x. \lambda y. y, \mathbf{xor} = \lambda x. \lambda y. x(y \perp \top)(y \top \perp), N = \top \oplus \perp. \\ 982$$

983 Then, consider  $M = (\lambda x. \mathbf{xor} xx)N$ . On one hand we have that all 984 the following equations have the value  $\Omega/_{\text{Null}}$  when evaluated for 985  $\mathcal{T}$ , due to Theorem 8.1.

$$986 \quad M = \mathbf{xor} NN = N(N \perp \top)(N \top \perp)$$

988 Hence, for any closed stable term  $K$ ,

$$989 \quad \llbracket M = K \rrbracket_{\mathcal{T}} \approx \llbracket (\top(N \perp \top)(N \top \perp)) \oplus (\perp(N \perp \top)(N \top \perp) = K) \rrbracket_{\mathcal{T}} \\ 990 \quad \approx \llbracket (N \perp \top) \oplus (N \top \perp) = K \rrbracket_{\mathcal{T}} \approx \llbracket N \perp \top = K \rrbracket_{\mathcal{T}} \\ 991 \quad \llbracket (\top \perp \top) \oplus (\perp \top \perp) = K \rrbracket_{\mathcal{T}} \approx \llbracket \perp \oplus \top = K \rrbracket_{\mathcal{T}} \approx \llbracket \top \oplus \perp = K \rrbracket_{\mathcal{T}}.$$

993 On the other hand, we have that for any closed stable term  $K$ :

$$994 \quad \llbracket M = K \rrbracket_{\mathcal{T}} \approx \llbracket (\lambda x. \mathbf{xor} xx)\top \oplus (\lambda x. \mathbf{xor} xx)\perp = K \rrbracket_{\mathcal{T}} \\ 995 \quad \approx \llbracket (\mathbf{xor} \top \top) \oplus (\mathbf{xor} \perp \perp) = K \rrbracket_{\mathcal{T}} \approx \llbracket \perp \oplus \perp = K \rrbracket_{\mathcal{T}} \approx \llbracket \perp = K \rrbracket_{\mathcal{T}}.$$

997 Putting together these two sequences of automorphic elements we 998 get that for any closed stable term  $K$ ,

$$999 \quad \llbracket \top \oplus \perp = K \rrbracket_{\mathcal{T}} \approx \llbracket \perp = K \rrbracket_{\mathcal{T}}.$$

1000 Since  $\perp$  is a closed stable term, this last equation that

$$1002 \quad \llbracket \top \oplus \perp = \perp \rrbracket_{\mathcal{T}} = \Omega/_{\text{Null}},$$

1003 meaning  $\mathcal{R}(\Omega)$  is a degenerate model, i.e. a singleton; this is a 1004 contradiction.  $\square$

1006 We can, however, have a stronger version of  $\beta$ -reduction than the 1007 one stated in Theorem 8.3.

1009 **Theorem 8.16** (Extended  $\beta$ -reduction). *If  $\lambda x. M_1, \lambda x. M_2$  are closed 1010 terms and  $N$  is a stable closed term such that for any stable closed 1011 term  $K$  and any  $i \in \{1, 2\}$  we have that*

$$1012 \quad \llbracket (\lambda x. M_i)N = K \rrbracket_{\mathcal{T}} \approx \llbracket M_i\{N/x\} = K \rrbracket_{\mathcal{T}},$$

1013 *then for any stable closed term  $K$ ,*

$$1015 \quad \llbracket (\lambda x. (M_1 \oplus M_2))N = K \rrbracket_{\mathcal{T}} \approx \llbracket M_1\{N/x\} \oplus M_2\{N/x\} = K \rrbracket_{\mathcal{T}}.$$

1017 *Proof.* Note that

$$1018 \quad \llbracket (\lambda x. (M_1 \oplus M_2))N \rrbracket \emptyset id \mathcal{T} \\ 1019 \quad = \llbracket (\lambda x. (M_1 \oplus M_2)) \rrbracket \emptyset (\lambda a. \llbracket N \rrbracket \emptyset (\lambda b. id(\phi(a)b)) \mathcal{T}^e) T^o$$

1020 and applying  $\lambda$ -distributivity,

$$1022 \quad = \llbracket (\lambda x. M_1 \oplus \lambda x. M_2) \rrbracket \emptyset (\lambda a. \llbracket N \rrbracket \emptyset (\lambda b. id(\phi(a)b)) \mathcal{T}^e) T^o$$

1024 now we solve the probabilistic choice and get some  $j \in \{1, 2\}$  such 1025 that

$$1026 \quad = \llbracket (\lambda x. M_j) \rrbracket \emptyset (\lambda a. \llbracket N \rrbracket \emptyset (\lambda b. id(\phi(a)b)) \mathcal{T}^e) \text{Snd}(T^o) \\ 1027 \quad \quad \quad \llbracket (\lambda x. M_j)N \rrbracket \emptyset id \mathcal{T}',$$

1028 where  $\mathcal{T}' = \text{Swap}(\text{Snd}(\mathcal{T}))$ . Hence, we have that

$$1029 \quad \llbracket (\lambda x. (M_1 \oplus M_2))N = K \rrbracket_{\mathcal{T}} \\ 1030 \quad = \{\omega \in \Omega \mid \llbracket (\lambda x. M_1 \oplus \lambda x. M_2)N \rrbracket \emptyset id \mathcal{T}(\omega) = \llbracket K \rrbracket \emptyset id \mathcal{T}(\omega)\} /_{\text{Null}} \\ 1031 \quad = \{\omega \in \Omega \mid \llbracket (\lambda x. M_j)N \rrbracket \emptyset id \mathcal{T}'(\omega) = \llbracket K \rrbracket \emptyset id \mathcal{T}(\omega)\} /_{\text{Null}}.$$

1033 Since  $K$  is stable,  $\llbracket K \rrbracket \emptyset id \mathcal{T} = \llbracket K \rrbracket \emptyset id \mathcal{T}'$ , so

$$1034 \quad = \{\omega \in \Omega \mid \llbracket (\lambda x. M_j)N \rrbracket \emptyset id \mathcal{T}'(\omega) = \llbracket K \rrbracket \emptyset id \mathcal{T}'(\omega)\} /_{\text{Null}} \\ 1035 \quad = \llbracket (\lambda x. M_j)N = K \rrbracket_{\mathcal{T}} \approx \llbracket M_j\{N/x\} = K \rrbracket_{\mathcal{T}}.$$

1037 Similarly we get that

$$1038 \quad \llbracket M_1\{N/x\} \oplus M_2\{N/x\} = K \rrbracket_{\mathcal{T}} \approx \llbracket M_j\{N/x\} = K \rrbracket_{\mathcal{T}'},$$

1039 for some tossing  $\mathcal{T}'$ . Further, Theorema 7.2 ensures us that

$$1040 \quad \llbracket M_j\{N/x\} = K \rrbracket_{\mathcal{T}'} \approx \llbracket M_j\{N/x\} = K \rrbracket_{\mathcal{T}''},$$

1041 which concludes our proof, since  $\approx$  is transitive.  $\square$

## 9 Generating Random Numbers

1042 In this section we present a small example of programming in 1043 the stochastic  $\lambda$ -calculus and use the semantics to argue for the 1044 correctness of the program behavior. Our program takes a Church 1045 numeral  $n$  and produces a random Church numeral from 0 to  $2^n - 1$  1046 with equal probability.

1047 **Functions for Church Numerals and Booleans.** In the following, 1048 we use mathematical symbols as the names of lambda terms, 1049 for ease of notation. Note that this means arithmetical expressions 1050 are in (forward) Polish notation.

1051 Recall that a Church numeral for the number  $n \in \mathbb{N}$  is a function 1052 of two arguments  $f$  and  $x$ , returning  $f$  applied  $n$  times to  $x$ . We 1053 use the well-known encodings of the arithmetic operations  $\text{succ}$ , 1054  $+$  and  $\times$ .

1055 **Picking a Random Number from 0 to  $2^n - 1$ .** The following 1056 stochastic  $\lambda$ -term is the key of our encoding.

$$1057 \quad \text{rand} = \lambda n. n(\lambda x. ((\times 2 x) \oplus (\text{succ}(\times 2 x))))0.$$

1058 In intuitive terms, the program starts with a number equal to 0 and 1059 flips a fair coin  $n$  times, either doubling the number or doubling 1060 and adding one, depending on the outcome.

1061 The following statement, if demonstrated, proves that the program 1062 has the desired behavior. It exemplifies how our deduction principles 1063 can be applied.

1064 **Statement:** For all Church numerals  $n$  and tossing processes  $\mathcal{T}$ ,

$$1065 \quad \bigvee_{i=0}^{2^n-1} \llbracket \text{rand } n = i \rrbracket_{\mathcal{T}} = \Omega/_{\text{Null}}$$

1066 and for all  $0 \leq i, j \leq 2^n - 1$

$$1067 \quad \llbracket \text{rand } n = i \rrbracket_{\mathcal{T}} \approx \llbracket \text{rand } n = j \rrbracket_{\mathcal{T}}.$$

1068 Therefore, for each  $0 \leq i \leq 2^n - 1$ ,  $\llbracket \text{rand } n = i \rrbracket_{\mathcal{T}}$  are of equal 1069 probabilities and summing to 1, i.e. of probability  $2^{-n}$ .

1070 **Sketch:** We do not provide a detailed proof of this statement, that 1071 would require further developments of the deduction principles. 1072 Instead, we sketch below how such a proof shall be organized.

1073 **Induction on  $n$ :** the inductive hypothesis we need is that

$$1074 \quad \bigvee_{i=0}^{\infty} \llbracket \text{rand } n = i \rrbracket_{\mathcal{T}} = \Omega/_{\text{Null}}, \text{ that for all } i \geq 2^n, \text{ we have}$$

$$1075 \quad \llbracket \text{rand } n = i \rrbracket_{\mathcal{T}} = \emptyset/_{\text{Null}}, \text{ and for all } 0 \leq i, j \leq 2^n - 1, \text{ we have}$$

$$1076 \quad \llbracket \text{rand } n = i \rrbracket_{\mathcal{T}} \equiv \llbracket \text{rand } n = j \rrbracket_{\mathcal{T}}.$$

1077 We start with the base case,  $n = 0$ . Then

$$1078 \quad \text{rand } \emptyset = \emptyset(\lambda x. ((\times 2 x) \oplus (\text{succ}(\times 2 x))))0 = \emptyset.$$

1079 We have  $\llbracket \text{rand } 0 = 0 \rrbracket = \Omega/_{\text{Null}}$ . This shows the three facts we 1080 want, because  $0 \leq i \leq 2^0 - 1 = 0$  implies  $i = 0$ .

1099 For the inductive step, we temporarily introduce the name  
 1100  
 1101

$$f' = (\lambda x.((\times 2 x) \oplus (\text{succ}(\times 2 x)))),$$

1102 for the latter part of the definition of `rand`, excluding the final 0.  
 1103 We start by re-expressing `rand` ( $\text{succ } n$ ) in terms of `rand`  $n$ .

$$\begin{aligned} \text{rand}(\text{succ } n) &= (\text{succ } n)f'0 = f'(\text{succ } n)0 = f'(\text{rand } n) \\ &= (\lambda x.((\times 2 x) \oplus (\text{succ}(\times 2 x))))(\text{rand } n) \\ &= ((\lambda x. \times 2 x) \oplus (\lambda x.\text{succ}(\times 2 x)))(\text{rand } n) \\ &= ((\lambda x. \times 2 x)(\text{rand } n)) \oplus ((\lambda x.\text{succ}(\times 2 x))(\text{rand } n)). \end{aligned}$$

1110 We first have to prove that  $\bigvee_{i=0}^{\infty} [\![\text{rand}(\text{succ } n) = i]\!]_{\mathcal{T}} = \Omega/\text{Null}$ .  
 1111

1112 By the inductive hypothesis, we have  $\bigvee_{i=0}^{\infty} [\![\text{rand } n = i]\!]_{\mathcal{T}} = \Omega/\text{Null}$ .  
 1113

1114 By well-definedness of function application, it holds that  
 1115

$$\begin{aligned} \bigvee_{i=0}^{\infty} [\![\lambda x. \times 2 x](\text{rand } n) = i]\!]_{\mathcal{T}} &= \Omega/\text{Null}, \\ \bigvee_{i=0}^{\infty} [\![\lambda x.\text{succ}(\times 2 x)](\text{rand } n) = i]\!]_{\mathcal{T}} &= \Omega/\text{Null}. \end{aligned}$$

1116 so, because of the properties of  $\oplus$  (Theorems 8.5,8.10),  
 1117

$$\bigvee_{i=0}^{\infty} [\![\text{rand}(\text{succ } n) = i]\!]_{\mathcal{T}} = \Omega/\text{Null}.$$

1118 For the second part of the inductive hypothesis, that  $i \geq 2^{n+1}$  implies  $[\![\text{rand}(\text{succ } n) = i]\!]_{\mathcal{T}} = \emptyset/\text{Null}$ , 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.

1119 Then  $i = 2i'$  for some integer  $i'$ . By the inductive hypothesis  
 1120  $[\![\text{rand } n = i']\!]_{\mathcal{T}} = \emptyset/\text{Null}$ , so by well-definedness and the fact  
 1121 that  $(\times 2i') = i$ , we have  $[\![\text{rand}(\text{succ } n) = i]\!]_{\mathcal{T}} = \emptyset/\text{Null}$ .

1122 Finally, we need to show that for all  $0 \leq i, j \leq 2^{n+1} - 1$  we have  
 1123  $[\![\text{rand}(\text{succ } n) = i]\!]_{\mathcal{T}} \approx [\![\text{rand}(\text{succ } n) = j]\!]_{\mathcal{T}}$ .  
 1124

1125 We have a four-way case split according to whether  $i$  and  $j$  are odd  
 1126 or even. We treat the case where  $i$  and  $j$  are even, as the other three  
 1127 cases are similar. We have that there are  $i', j'$  such that  $i = 2i'$  and  
 1128  $j = 2j'$ . Hence,  
 1129

$$\begin{aligned} &[\![\text{rand}(\text{succ } n) = i]\!]_{\mathcal{T}} \\ &= [\![\lambda x. \times 2 x](\text{rand } n) \oplus ((\lambda x.\text{succ}(\times 2 x))(\text{rand } n)) = (\lambda x. \times 2 x)i'\!]_{\mathcal{T}} \\ &\approx [\![\lambda x. \times 2 x](\text{rand } n) = (\lambda x. \times 2 x)i']_{\mathcal{T}} \\ &= [\![\text{rand } n = i']\!]_{\mathcal{T}} = [\![\text{rand } n = j']\!]_{\mathcal{T}} = [\![\text{rand}(\text{succ } n) = j]\!]_{\mathcal{T}}. \end{aligned}$$

1130 This shall complete the proof.  
 1131

## 10 Conclusions

1132 We see this paper as the beginning of an investigation into ran-  
 1133 dom processes at higher type. There are many things to investi-  
 1134 gate:  
 1135

- 1136 • We need to understand how this relates to more categorical  
 1137 approaches [7] to probability theory at higher type by con-  
 1138 structing a suitable cartesian closed category of quasi-Borel  
 1139 spaces.
- 1140 • We need to develop a deeper understanding of the Boolean-  
 1141 valued reasoning principles that we have used here.

1142 • It would be very interesting to develop suitable dependently-  
 1143 typed versions of a stochastic  $\lambda$ -calculus; indeed this was one  
 1144 of the main motivations of [15].  
 1145

1146 • The relation between invariance results as we have used them  
 1147 and exchangeability and symmetry principles in probability the-  
 1148 ory (see, for example, [8]) need to be understood better.  
 1149

## Acknowledgments

1150 This research was supported by a DFF Danish research grant, by  
 1151 a grant from NSERC (Canada), and by a grant from the National  
 1152 Science Foundation (USA). We gratefully acknowledge the support  
 1153 of the Simons Institute Logical Structures in Computation Pro-  
 1154 gram in Fall 2016. We thank Ugo Dal Lago, Cameron Freer, Marco  
 1155 Gaobardi, Chris Heunen, Alex Simpson and Sam Staton for useful  
 1156 discussions.  
 1157

## References

- [1] John L. Bell. 1985. *Boolean-Valued Models and Independence Proofs in Set Theory* (2nd ed.). Number 12 in Oxford Logic Guides. Oxford University Press.
- [2] David H. Fremlin. 2003. Measure Theory, Volume 4. <http://www.essex.ac.uk/~maths/people/fremlin/mt.htm> (2003).
- [3] R. M. Friedberg and H. Rogers. 1959. Reducibility and Completeness for sets of integers. *Mathematical Logic Quarterly* 5 (1959), 117–125.
- [4] Noah Goodman, Vikash Mansinghka, Daniel Roy, Keith Bonawitz, and Joshua Tenenbaum. 2008. Church: a language for generative models. In *Proceedings of the 24th Conference on Uncertainty in Artificial Intelligence*. 220–229.
- [5] Andrew D. Gordon, Thore Graepel, Nicolas Rolland, Claudio V. Russo, Johannes Borgström, and John Guiver. 2014. Tabular: a schema-driven probabilistic programming language. In *Proceedings of POPL '14, San Diego, CA, USA, January 20-21, 2014*. 321–334.
- [6] V. Gupta, R. Jagadeesan, and P. Panangaden. 1999. Stochastic Processes as Concurrent Constraint Programs. In *Proceedings of the 26th Proceedings Of The Annual ACM Symposium On Principles Of Programming Languages*. 189–202.
- [7] Chris Heunen, Ohad Kammar, Sam Staton, and Hongseok Yang. 2017. A convenient category for higher-order probability theory. In *Proceedings of the Thirty-second Annual ACM-IEEE Symposium on Logic in Computer Science*.
- [8] Olav Kallenberg. 2006. *Probabilistic symmetries and invariance principles*. Springer Science and Business Media.
- [9] D. Kozen. 1981. Semantics of Probabilistic Programs. *Journal of Computer and Systems Sciences* 22 (1981), 328–350.
- [10] Ugo Dal Lago, Davide Sangiorgi, and Michele Alberti. 2014. On Coinductive Equivalences for Higher-order Probabilistic Functional Programs. In *Proceedings of the 41st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages (POPL '14)*. ACM, New York, NY, USA, 297–308.
- [11] Norman Ramsey and Avi Pfeffer. 2002. Stochastic lambda calculus and monads of probability distributions. In *The 29th SIGPLAN-SIGACT Symposium on Principles of Programming Languages*. 154–165.
- [12] H. Rogers. 1967. *Theory of Recursive Functions and Effective Computability*. McGraw-Hill.
- [13] Vladimir A. Rohlin. 1952. On the Fundamental Ideas of Measure Theory. *Amer. Math. Soc. Transl.* 71 (1952).
- [14] Dana Scott. 1967. A Proof of the Independence of the Continuum Hypothesis. *Mathematical Systems Theory* 1, 2 (1967), 89–111.
- [15] D. Scott. 2014. Stochastic Lambda Calculi: An extended Abstract. *Journal of Applied Logic* 12 (2014), 369–376.
- [16] S. Smolka, P. Kumar, N. Foster, and D. Kozen, and A. Silva. 2017. Cantor meets Scott: Domain-theoretic foundations for probabilistic network programming. In *Proceedings of the 44th ACM SIGPLAN Symp. Principles of Programming Languages (POPL '17)*. ACM, 557–571. ACM SIGPLAN Notices - POPL '17, Volume 52 Issue 1.
- [17] Sam Staton, Hongseok Yang, Frank Wood, Chris Heunen, and Ohad Kammar. 2017. Semantics for probabilistic programming: higher-order functions, continuous distributions, and soft constraints. In *Proceedings of the 31st Annual ACM-IEEE Symposium On Logic In Computer Science*. 525–534.
- [18] David Tolpin, Jan-Willem van de Meent, Hongseok Yang, and Frank Wood. 2016. Design and implementation of probabilistic programming language anglican. In *Proceedings of the 28th Symposium on the Implementation and Application of Functional Programming Languages*. ACM, 6.
- [19] David Williams. 1991. *Probability with Martingales*. Cambridge University Press.

## 1221 A The functions pack and unpack 1282

1222 For a better connection with the mainstream of measure theory, we 1283  
 1223 consider how to pack and unpack sequences in  $2^{\mathbb{N}}$ . In aid of this, 1284  
 1224 we consider two infinite sequences in  $\mathbb{N}$ : 1285  
 1225

$$1226 s_n^0 = 2^n - 1 \text{ and } s_n^1 = 3 \cdot 2^n - 1. \quad 1286$$

1227 These sequences solve the recurrence relations: 1287

$$1229 \{s_0^0 = 0; s_n^0 = 2s_{n-1}^0 + 1\}_{n \geq 1} \quad \text{and} \quad \{s_0^1 = 2; s_n^1 = 2s_{n-1}^1 + 1\}_{n \geq 1}. \quad 1288$$

1230 Therefore each sequence is even iff  $n = 0$ . Moreover,  $s_j^i = s_{j'}^{i'}$  implies 1289  
 1231  $i = i'$  and  $j = j'$ . 1290

1232 We define a family of maps  $\text{unpack}_n : \mathcal{P}(\mathbb{N}) \rightarrow 2$  by 1291

$$1233 \text{unpack}_n(S) = \begin{cases} 0 & \text{if } s_n^0 \in S \\ 1 & \text{otherwise} \end{cases}, \quad 1292$$

1234 and two maps  $\text{unpack} : \mathcal{P}(\mathbb{N}) \rightarrow 2^{\mathbb{N}}$  and  $\text{pack} : 2^{\mathbb{N}} \rightarrow \mathcal{P}(\mathbb{N})$  as 1293  
 1235 follows: 1294

$$1236 \text{unpack}(S)(n) = \text{unpack}_n(S) \text{ pack}(a) = \bigcup_{i=0}^{\infty} \{s_i^{a_i}\}. \quad 1295$$

1237 It is not difficult to verify that the maps  $\text{unpack}_n$ ,  $\text{unpack}$  and  $\text{pack}$  1296  
 1238 are (Borel) measurable, and that for all  $a \in 2^{\mathbb{N}}$ , we have 1297

$$1239 \text{unpack}(\text{pack}(a))(n) = a_n. \quad 1298$$

1240	1301
1241	1302
1242	1303
1243	1304
1244	1305
1245	1306
1246	1307
1247	1308
1248	1309
1249	1310
1250	1311
1251	1312
1252	1313
1253	1314
1254	1315
1255	1316
1256	1317
1257	1318
1258	1319
1259	1320
1260	1321
1261	1322
1262	1323
1263	1324
1264	1325
1265	1326
1266	1327
1267	1328
1268	1329
1269	1330
1270	1331
1271	1332
1272	1333
1273	1334
1274	1335
1275	1336
1276	1337
1277	1338
1278	1339
1279	1340
1280	1341
1281	1342