An Intrinsic Characterization of Approximate Probabilistic Bisimilarity

Franck van Breugel[1], Michael Mislove[2], Joël Ouaknine[3] and James Worrell[2]

[1] York University, Department of Computer Science
4700 Keele Street, Toronto, M3J 1P3, Canada
[2] Tulane University, Department of Mathematics
6823 St Charles Avenue, New Orleans LA 70118, USA
[3] Computer Science Department, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh PA 15213, USA

Abstract. In previous work we have investigated a notion of approximate bisimilarity for labelled Markov processes. We argued that such a notion is more realistic and more feasible to compute than (exact) bisimilarity. The main technical tool used in the underlying theory was the Hutchinson metric on probability measures. This paper gives a more fundamental characterization of approximate bisimilarity in terms of the notion of (exact) similarity. In particular, we show that the topology of approximate bisimilarity is the Lawson topology with respect to the simulation preorder. To complement this abstract characterization we give a statistical account of similarity, and by extension, of approximate bisimilarity, in terms of the process testing formalism of Larsen and Skou.

1 Introduction

Labelled Markov processes provide a simple operational model of reactive probabilistic systems. A *labelled Markov process* consists of a measurable space (X, Σ) of states, a family Act of actions, and a transition probability function $\mu_{-,-}$ that, given a state $x \in X$ and an action $a \in \mathrm{Act}$, yields the probability $\mu_{x,a}(A)$ that the next state of the process will be in the measurable set $A \in \Sigma$ after performing action a in state x. These systems are a generalization of the probabilistic labelled transition systems with discrete distributions considered by Larsen and Skou [16].

The basic notion of process equivalence in concurrency is bisimilarity. This notion, due to Park [18], asserts that processes are *bisimilar* iff any action by

[1] Supported by the Natural Sciences and Engineering Research Council of Canada.
[2] Supported by the US National Science Foundation and the US Office of Naval Research (ONR).
[3] Supported by ONR contract N00014-95-1-0520, Defense Advanced Research Project Agency and the Army Research Office under contract DAAD19-01-1-0485.

either can be matched with the same action by the other, and the resulting processes are also bisimilar. Larsen and Skou adapted the notion of bisimilarity to discrete probabilistic systems, by defining an equivalence relation R on states to be a bisimulation if related states have *exactly matching* probabilities of making transitions into any R-equivalence class. Later the theory of probabilistic bisimilarity was extended beyond the discrete setting by Edalat, Desharnais and Panangaden [8]. From quite early on, however, it was realized that for probabilistic systems a notion of approximate bisimilarity might prove more appropriate than a notion of exact bisimilarity. One advantage of such a notion is that it is more informative: one can say that two processes are almost bisimilar, even though they do not behave exactly the same. More fundamentally, one could even argue that the idea of exact bisimilarity is meaningless if the probabilities appearing in the model of a system are approximations based on statistical data, or if the algorithm used to calculate bisimilarity is not based on exact arithmetic.

Desharnais, Gupta, Jagadeesan and Panangaden [9] formalized a notion of approximate bisimilarity by defining a metric[1] on the class of labelled Markov processes. Intuitively the smaller the distance between two processes, the more alike their behaviour; in particular, they showed that states are at zero distance just in case they are bisimilar. The original definition of the metric in [9] was stated through a real-valued semantics for a variation of Larsen and Skou's probabilistic modal logic [16]. Later it was shown how to give a coinductive definition of this metric using the Hutchinson metric on probability measures [4]. Using this characterization [5] gave an algorithm based on linear programming to approximate the distance between the states of a finite labelled Markov process.

The fact that zero distance coincides with bisimilarity can be regarded as a sanity check on the definition of the metric. The papers [9, 4] also feature a number of examples showing how processes with similar transition probabilities are close to one another. A more precise account of how the metric captures approximate bisimilarity is given in [6], where it is shown that convergence in the metric can be characterized in terms of the convergence of observable behaviour; the latter is formalized by Larsen and Skou's process testing formalism [16]. As Di Pierro, Hankin and Wiklicky [20] argue, such an account is vital if one wants to use the metric to generalize the formulations of probabilistic non-interference based on bisimilarity.

Both of the above mentioned characterizations of the metric for approximate bisimilarity are based on the idea of defining a distance between measures by integration against a certain class of functions, which is a standard approach from functional analysis. But it is reasonable to seek an intrinsic characterization of approximate bisimilarity that does not rely on auxiliary notions such as integration. In this paper we give such a characterization. We show that the topology induced by the metric described above coincides with the Lawson topology on the domain that arises by endowing the class of labelled Markov processes with the probabilistic *simulation* preorder. The characterization is intrinisic: the Lawson

[1] Strictly speaking, a pseudometric since distinct processes can have distance zero.

topology is defined solely in terms of the order on the domain. For this reason, we view this characterization as more fundamental than the existing ones.

Our results are based on a simple interaction between domain theory and measure theory. This is captured in Corollary 2 which shows that the Lawson topology on the probabilistic powerdomain of a coherent domain agrees with the weak topology on the family of subprobability measures on the underlying coherent domain, itself endowed with the Lawson topology. A simple corollary of this result is that the probabilistic powerdomain of a coherent domain is again coherent, a result first proved by Jung and Tix [15] using purely domain-theoretic techniques.

We use the coincidence of the Lawson and weak topologies to analyze a recursively defined domain D of probabilistic processes first studied by Desharnais *et al.* [10]. The key property of the domain D is that it is equivalent (as a preordered class) to the class of all labelled Markov processes equipped with the simulation preorder. The proof of this result in [10] makes use of a discretization construction, which shows how an arbitrary labelled Markov process can be recovered as the limit of a chain of finite state approximations. In this paper, we give a more abstract proof: we use the coincidence of the Lawson and weak topologies to show that the domain D has a universal property: namely, it is final in a category of labelled Markov processes.

A minor theme of the present paper is to extend the characterization of approximate bisimilarity in terms of the testing formalism of Larsen and Skou [16]. We show that bisimilarity can be characterized as testing equivalence, where one records only positive observations of tests. On the other hand, characterizing similarity requires one also to record negative observations, i.e., refusals of actions.

2 Labelled Markov Processes

Let us assume a fixed set Act of actions. For ease of exposition we suppose that Act is finite, but all our results hold in case it is countable.

Definition 1. *A* labelled Markov process *is a triple $\langle X, \Sigma, \mu \rangle$ consisting of a set X of states, a σ-field Σ on X, and a transition probability function $\mu : X \times \mathrm{Act} \times \Sigma \to [0,1]$ such that*

1. *for all $x \in X$ and $a \in \mathrm{Act}$, the function $\mu_{x,a}(\cdot) : \Sigma \to [0,1]$ is a subprobability measure, and*
2. *for all $a \in \mathrm{Act}$ and $A \in \Sigma$, the function $\mu_{-,a}(A) : X \to [0,1]$ is measurable.*

The function $\mu_{-,a}$ describes the reaction of the process to the action a selected by the environment. This represents a reactive model of probabilistic processes. Given that the process is in state x and reacts to the action a chosen by the environment, $\mu_{x,a}(A)$ is the probability that the process makes a transition to a state in the set of states A. Note that we consider *sub*probability measures, i.e. positive measures with total mass no greater than 1, to allow for the possibility that the process may refuse an action. The probability of refusal of the action a given the process is in state x is $1 - \mu_{x,a}(X)$.

An important special case is when the σ-field Σ is the powerset of X and, for all actions a and states x, the subprobability measure $\mu_{x,a}(\cdot)$ is completely determined by a discrete subprobability distribution. This case corresponds to the original probabilistic transition system model of Larsen and Skou [16].

A natural notion of a map between labelled Markov processes is given in:

Definition 2. *Given labelled Markov processes $\langle X, \Sigma, \mu\rangle$ and $\langle X', \Sigma', \mu'\rangle$, a measurable function $f\colon X \to X'$ is a* zigzag map *if for all $x \in X$, $a \in \mathrm{Act}$ and $A' \in \Sigma'$, $\mu_{x,a}(f^{-1}(A')) = \mu'_{f(x),a}(A)$.*

Probabilistic bisimulations (henceforth just bisimulations) were first introduced in the discrete case by Larsen and Skou [16]. They are the relational counterpart of zigzag maps and can also be seen, in a very precise way, as the probabilistic analogues of the strong bisimulations of Park and Milner [17]. The definition of bisimulation was extended to labelled Markov processes in [8, 10].

Definition 3. *Let $\langle X, \Sigma, \mu\rangle$ be a labelled Markov process. A reflexive relation R on X is a* simulation *if whenever $x\,R\,y$, then for all $a \in \mathrm{Act}$ and all R-closed $A \in \Sigma$, $\mu_{x,a}(A) \leqslant \mu_{y,a}(A)$. A set A is* R-closed *if $x \in A$ and $x\,R\,y$ imply $y \in A$. We say that R is a* bisimulation *if, in addition, whenever $x\,R\,y$ then $\mu_{x,a}(X) = \mu_{y,a}(X)$. Two states are* bisimilar *if they are related by some bisimulation.*

The notions of simulation and bisimulation are very close in the probabilistic case. The extra condition $\mu_{x,a}(X) = \mu_{y,a}(X)$ in the definition of bisimulation allows one to show that if R is a bisimulation, then the inverse relation R^{-1} is also a bisimulation. It follows that the union of all bisimulations is an equivalence relation R such that xRy implies $\mu_{x,a}(A) = \mu_{y,a}(A)$ for all $a \in \mathrm{Act}$ and measurable R-closed $A \subseteq X$. This equality, which entails infinite precision, is the source of the fragility in the definition of bisimilarity. This motivates the idea of defining a notion of approximate bisimilarity.

2.1 A Metric for Approximate Bisimilarity

We recall a variant of Larsen and Skou's probabilistic modal logic [16], and a real-valued semantics due to Desharnais *et al.* [9]. The set of formulas of probabilistic modal logic (PML), denoted $\mathcal{F}$, is given by the following grammar:

$$f ::= \top \mid f \wedge f \mid f \vee f \mid \langle a\rangle f \mid f \dot{-} q$$

where $a \in \mathrm{Act}$ and $q \in [0,1] \cap \mathbb{Q}$.

The modal connective $\langle a\rangle$ and truncated subtraction replace a single connective $\langle a\rangle_q$ in Larsen and Skou's presentation.

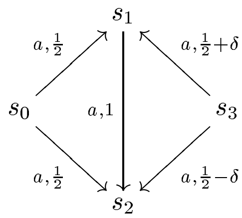
Fix a constant $0 < c < 1$ once and for all. Given a labelled Markov process $\langle X, \Sigma, \mu\rangle$, a formula f determines a measurable function $f\colon X \to [0,1]$ according to the following rules. $\top$ is interpreted as the constant function 1, conjunction and disjunction are interpreted as max and min respectively, truncated subtraction is defined in the obvious manner, and $(\langle a\rangle f)(x) = c\int f d\mu_{x,a}$ for each $a \in \mathrm{Act}$. Thus the interpretation of a formula f depends on c. The role of this constant is to discount observations made at greater modal depth.

Given a labelled Markov process $\langle X, \Sigma, \mu\rangle$, one defines a metric d on X by

$$d(x,y) = \sup_{f\in\mathcal{F}} |f(x) - f(y)| .$$

It is shown in [9] that zero distance in this metric coincides with bisimilarity. Roughly speaking, the smaller the distance between states, the closer their behaviour. The exact distance between two states depends on the value of c, but one consequence of our results is that the topology induced by the metric d is independent of the original choice of c.

Example 1. In the labelled Markov process below, $d(s_0, s_3) = c^2\delta$. The two states are bisimilar just in case $\delta = 0$.



3 Domain Theory

A *directed subset* $A \subseteq D$ of a poset D is one for which every finite subset of A has an upper bound in A, and a *directed complete partial order (dcpo)* is a poset D in which each directed set A has a least upper bound, denoted $\sqcup A$. If D is a dcpo, and $x, y \in D$, then we write $x \ll y$ if each directed subset $A \subseteq D$ with $y \sqsubseteq \sqcup A$ satisfies $\uparrow x \cap A \neq \emptyset$. We then say x *is way-below* y. Let $\mathord{\downarrow\!\!\downarrow} y = \{x \in D \mid x \ll y\}$; we say that D is *continuous* if it has a *basis*, i.e., a subset $B \subseteq D$ such that for each $y \in D$, $\mathord{\downarrow\!\!\downarrow} y \cap B$ is directed with supremum y. We use the term *domain* to mean a continuous dcpo.

A subset U of a dcpo D is *Scott-open* if it is an upper set (i.e., $U = \uparrow U$) and for each directed set $A \subseteq D$, if $\sqcup A \in U$ then $A \cap U \neq \emptyset$. The collection ΣD of all Scott-open subsets of D is called the *Scott topology* on D. If D is continuous, then the Scott topology on D is locally compact, and the sets $\mathord{\uparrow\!\!\uparrow} x$ where $x \in D$ form a basis for this topology. Given dcpos D and E, a function $f \colon D \to E$ is continuous with respect to the Scott topologies on D and E iff it is monotone and preserves directed suprema: for each directed $A \subseteq D$, $f(\sqcup A) = \sqcup f(A)$.

Another topology of interest on a continuous dcpo D is the *Lawson topology*. This topology is the join of the Scott topology and the *lower interval topology*, where the latter is generated by sub-basic open sets of the form $D \setminus \uparrow x$. Thus, the

Lawson topology has the family $\{\Uparrow x \setminus \uparrow F \mid x \in D, F \subseteq D \text{ finite}\}$ as a basis. The Lawson topology on a domain is always Hausdorff. A domain which is compact in its Lawson topology is called *coherent*.

4 The Probabilistic Powerdomain

We briefly recall some basic definitions and results about valuations and the probabilistic powerdomain.

Definition 4. *Let (X, Ω) be a topological space. A* valuation *on X is a mapping $\mu\colon \Omega \to [0,1]$ satisfying:*

1. $\mu\emptyset = 0$.
2. $U \subseteq V \Rightarrow \mu U \leqslant \mu V$.
3. $\mu(U \cup V) + \mu(U \cap V) = \mu U + \mu V, \;\; U, V \in \Omega$

Departing from standard practice, we also require that a valuation is Scott continuous as a map $(\Omega, \subseteq) \to ([0,1], \leqslant)$.

Each element $x \in X$ gives rise to a valuation δ_x defined by $\delta_x(U) = 1$ if $x \in U$, and $\delta_x(U) = 0$ otherwise. A *simple valuation* has the form $\sum_{a\in A} r_a \delta_a$ where A is a finite subset of X, $r_a \geqslant 0$, and $\sum_{a\in A} r_a \leqslant 1$.

We write $\mathbb{V}X$ for the space whose points are valuations on X, and whose topology is generated by sub-basic open sets of the form $\{\mu \mid \mu U > r\}$, where $U \in \Omega$ and $r \in [0,1]$. The specialization order[2] on $\mathbb{V}X$ with respect to this topology is given by $\mu \sqsubseteq \mu'$ iff $\mu U \leqslant \mu' U$ for all $U \in \Omega$. $\mathbb{V}$ extends to an endofunctor on Top – the category of topological spaces and continuous maps – by defining $\mathbb{V}(f)(\mu) = \mu \circ f^{-1}$ for a continuous map f.

Suppose D is a domain regarded as a topological space in its Scott topology. Jones [14] has shown that the specialization order defines a domain structure on $\mathbb{V}D$, with the set of simple valuations forming a basis. Furthermore, it follows from the following proposition that the topology on $\mathbb{V}D$ is actually the Scott topology with respect to the pointwise order on valuations.

Proposition 1 (Edalat [11]). *A net $\langle \mu_\alpha \rangle$ converges to μ in the Scott topology on $\mathbb{V}D$ iff* $\liminf \mu_\alpha U \geqslant \mu U$ *for all Scott open $U \subseteq D$.*

Finally, Jung and Tix [15] have shown that if D is a coherent domain then so is $\mathbb{V}D$. In summary we have the following proposition.

Proposition 2. *The endofunctor $\mathbb{V}\colon \mathsf{Top} \to \mathsf{Top}$ preserves the subcategory $\omega\mathsf{Coh}$ of coherent domains with countable bases equipped with their Scott topologies.*

The fact that we define the functor $\mathbb{V}$ over Top rather than just considering the probabilistic powerdomain as a construction on domains has a payoff later on.

[2] The specialization preorder on a topological space is defined by $x \sqsubseteq y$ iff, for every open set U, $x \in U$ implies $y \in U$. It is a partial order precisely when the space is T_0.

Obviously, valuations bear a close resemblance to measures. In fact, any valuation on a coherent domain D may be uniquely extended to a measure on Borel σ-field generated by the Scott topology (equivalently by the Lawson topology) on D [2]. Thus we may consider the so-called *weak topology* on $\mathbb{V}D$. This is the weakest topology such that for each Lawson continuous function $f\colon D \to [0,1]$, $\Phi_f(\mu) = \int f d\mu$ defines a continuous function $\Phi_f\colon \mathbb{V}D \to [0,1]$. Alternatively, it may be characterized by saying that a net of valuations $\langle \mu_\alpha \rangle$ converges to μ iff $\liminf \mu_\alpha O \geqslant \mu O$ for each Lawson open set O (cf. [19, Thm II.6.1]). We emphasize that the weak topology on $\mathbb{V}D$ is defined with respect to the Lawson topology on D.

5 The Lawson Topology on $\mathbb{V}D$

In this section we show that for a coherent domain D, the Lawson topology on $\mathbb{V}D$ coincides with the weak topology.

Proposition 3. *[Jones [14]] Suppose $\mu \in \mathbb{V}D$ is an arbitrary valuation, then $\sum_{a\in A} r_a \delta_a \sqsubseteq \mu$ iff $(\forall B \subseteq A) \sum_{a\in B} r_a \leqslant \mu(\uparrow B)$.*

Proposition 4. *Let $F = \{x_1, ..., x_n\} \subseteq D$, $\mu \in \mathbb{V}D$ and $\varepsilon > 0$ be given. Then there exists a finite set $\mathcal{K}$ of simple valuations such that $\mu \notin \uparrow\mathcal{K}$ but if $\nu \in \mathbb{V}D$ satisfies $\nu(\uparrow F) > \mu(\uparrow F) + \varepsilon$ then $\nu \in \uparrow\mathcal{K}$.*

Proof. Let $\delta = \varepsilon/n$. Define $f_\delta\colon [0,1] \to [0,1]$ by $f_\delta(x) = \max\{m\delta \mid m\delta \ll x, m \in \mathbb{N}\}$. Next we define $\mathcal{K}$ to be the finite set

$$\mathcal{K} = \left\{ \sum_{i=1}^{n} r_i \delta_{x_i} \mid \mu(\uparrow F) < \sum_{i=1}^{n} r_i \leqslant 1 \text{ and } \{r_1, ..., r_n\} \subseteq \operatorname{Ran} f_\delta \right\}.$$

From Proposition 3 we immediately deduce $\mu \notin \uparrow\mathcal{K}$. Now given $\nu \in \mathbb{V}D$ with $\nu(\uparrow F) > \mu(\uparrow F) + \varepsilon$, we set $r_i = f_\delta(\nu(\uparrow x_i \setminus \bigcup_{j<i} \uparrow x_j))$ for $i \in \{1, ..., n\}$. First we verify that $\sum_{i=1}^{m} r_i \delta_{x_i} \in \mathcal{K}$. Now

$$\begin{aligned}
\nu(\uparrow F) - \sum_{i=1}^{n} r_i &= \nu(\uparrow F) - \sum_{i=1}^{n} f_\delta(\nu(\uparrow x_i \setminus \bigcup_{j<i} \uparrow x_j)) \\
&= \sum_{i=1}^{n} \left(\nu(\uparrow x_i \setminus \bigcup_{j<i} \uparrow x_j) - f_\delta(\nu(\uparrow x_i \setminus \bigcup_{j<i} \uparrow x_j)) \right) \\
&< n\delta = \varepsilon.
\end{aligned}$$

It follows that $\sum_{i=1}^{n} r_i > \mu(\uparrow F)$, thus $\sum_{i=1}^{n} r_i \delta_{x_i} \in \mathcal{K}$.

Finally, we observe that $\sum_{i=1}^{n} r_i \delta_{x_i} \sqsubseteq \nu$ since, if $B \subseteq \{1, ..., n\}$, then

$$\sum_{i\in B} r_i = \sum_{i\in B} f_\delta(\nu(\uparrow x_i \setminus \bigcup_{j<i} \uparrow x_j)) \leqslant \sum_{i\in B} \nu(\uparrow x_i \setminus \bigcup_{j<i} \uparrow x_j) \leqslant \nu(\uparrow B).$$

Proposition 5. *A net $\langle \mu_\alpha \rangle$ converges to μ in the lower interval topology on $\mathbb{V}D$ iff $\limsup \mu_\alpha E \leqslant \mu E$ for all finitely generated upper sets E.*

Proof. Suppose $\mu_\alpha \to \mu$. Let $E = \uparrow F$, where F is finite, and suppose $\varepsilon > 0$ is given. Then by Proposition 4 there is a finite set $\mathcal{K}$ of simple valuations such that $\mu \notin \uparrow \mathcal{K}$ and for all valuations ν, $\nu \notin \uparrow \mathcal{K}$ implies $\nu E \leqslant \mu E + \varepsilon$. Then we conclude that $\limsup \mu_\alpha E \leqslant \mu E + \varepsilon$ since the net μ_α is eventually in the open set $\mathbb{V}D \setminus \uparrow \mathcal{K}$.

Conversely, suppose $\mu_\alpha \not\to \mu$. Then μ has a sub-basic open neighbourhood $\mathbb{V}D \setminus \uparrow \rho$ such that some subnet μ_β never enters this neighbourhood. We can assume $\rho = \sum_{a \in A} r_a \delta_a$ is a simple valuation. Since $\rho \not\sqsubseteq \mu$ there exists $B \subseteq A$ such that $\sum_{a \in B} r_a > \mu(\uparrow B)$. But $\mu_\beta(\uparrow B) \geqslant \sum_{a \in B} r_a > \mu(\uparrow B)$ for all β. Thus $\limsup \mu_\alpha(\uparrow B) > \mu(\uparrow B)$. □

Corollary 1. *Let $\langle \mu_\alpha \rangle$ be a net in $\mathbb{V}D$. Then $\langle \mu_\alpha \rangle$ converges to μ in the Lawson topology on $\mathbb{V}D$ iff*

1. $\liminf \mu_\alpha U \geqslant \mu U$ *for all Scott open $U \subseteq D$.*
2. $\limsup \mu_\alpha E \leqslant \mu E$ *for all finitely generated upper sets $E \subseteq D$.*

Proof. Combine Propositions 1 and 5. □

Corollary 2. *If D is Lawson compact, then so is $\mathbb{V}D$ and the weak and Lawson topologies agree on $\mathbb{V}D$.*

Proof. Recall [19, Thm II.6.4] that the weak topology on the space of Borel measures on a compact space is itself compact. By Corollary 1, the Lawson topology on $\mathbb{V}D$ is coarser than the weak topology. But it is a standard fact that if a compact topology is finer than a Hausdorff topology then the two must coincide.

The Lawson compactness of $\mathbb{V}D$ was first proved by Jung and Tix in [15]. Their proof is purely domain theoretic and doesn't use the compactness of the weak topology.

6 A Final Labelled Markov Process

In a previous paper [4] we used the Hutchinson metric on probability measures to construct a final object in the category of labelled Markov processes and zigzag maps. Here we show that one may also construct a final labelled Markov process as a fixed point D of the probabilistic powerdomain. As we mentioned in the introduction, the significance of this result is that D can be used to represent the class of all labelled Markov processes in the simulation preorder.

Given a measurable space $X = \langle X, \Sigma \rangle$, we write $\mathbb{M}X$ for the set of subprobability measures on X. For each measurable subset $A \subseteq X$ we have a projection function $p_A \colon \mathbb{M}X \to [0, 1]$ sending μ to μA. We take $\mathbb{M}X$ to be a measurable space by giving it the smallest σ-field such that all the projections p_A are measurable. Next, $\mathbb{M}$ is turned into a functor $\mathsf{Mes} \to \mathsf{Mes}$, where Mes denotes the category of measure spaces and measurable maps, by defining $\mathbb{M}(f)(\mu) = \mu \circ f^{-1}$ for $f \colon X \to Y$ and $\mu \in \mathbb{M}X$; see Giry [12] for details.

Definition 5. *Let $\mathcal{C}$ be a category and $F : \mathcal{C} \to \mathcal{C}$ a functor. An F-*coalgebra *consists of an object C in $\mathcal{C}$ together with an arrow $f : C \to FC$ in $\mathcal{C}$. An F-*homomorphism *from F-coalgebra $\langle C, f\rangle$ to F-coalgebra $\langle D, g\rangle$ is an arrow $h : C \to D$ in $\mathcal{C}$ such that $Fh \circ f = g \circ h$. F-coalgebras and F-homomorphisms form a category whose final object, if it exists, is called the* final F-coalgebra.

Given a labelled Markov process $\langle X, \Sigma, \mu\rangle$, μ may be regarded as a measurable map $X \to \mathbb{M}(X)^{\mathrm{Act}}$. That is, labelled Markov processes are nothing but coalgebras of the endofunctor $\mathbb{M}(-)^{\mathrm{Act}}$ on the category Mes. Furthermore the coalgebra homomorphisms in this case are just the zigzag maps, cf. [8].

Next, we relate the functor $\mathbb{M}$ to the probabilistic powerdomain functor $\mathbb{V}$. To mediate between domains and measure spaces we introduce the forgetful functor $\mathbb{U}\colon \omega\mathsf{Coh} \to \mathsf{Mes}$ which maps a coherent domain to the Borel measurable space generated by the Scott topology (equivalently by the Lawson topology).

Proposition 6. *The forgetful functor $\mathbb{U}\colon \omega\mathsf{Coh} \to \mathsf{Mes}$ satisfies $\mathbb{M} \circ \mathbb{U} = \mathbb{U} \circ \mathbb{V}$.*

Proof. (Sketch) Given a coherent domain D with countable basis, since valuations on the Scott topology on D are in 1-1 correspondence with Borel measures on $\mathbb{U}(D)$, we have a bijection between the points of the measurable spaces $\mathbb{MU}(D)$ and $\mathbb{UV}(D)$. That this bijection is an isomorphism of measurable spaces follows from the coincidence of the Lawson and weak topologies and the unique structure theorem[3]. $\square$

The following proposition is a straightforward adaptation of [19, Thm I.1.10].

Proposition 7. *The forgetful functor $\mathbb{U}\colon \omega\mathsf{Coh} \to \mathsf{Mes}$ preserves limits of ω^{op}-chains.*

Starting with the final object of $\omega\mathsf{Coh}$, we construct the chain

$$1 \xleftarrow{!} \mathbb{V}1 \xleftarrow{\mathbb{V}!} \mathbb{V}^2 1 \xleftarrow{\mathbb{V}^2 !} \mathbb{V}^3 1 \xleftarrow{\mathbb{V}^3 !} \cdots \tag{1}$$

by iterating the functor $\mathbb{V}$. Writing $\{\mathbb{V}^n 1 \xleftarrow{\pi_n} \mathbb{V}^\omega 1\}_{n<\omega}$ for the limit cone of this chain, there is a unique 'connecting' map $\mathbb{V}^\omega 1 \leftarrow \mathbb{V}\mathbb{V}^\omega 1$ whose composition with π_n gives $\mathbb{V}\pi_n$.

Proposition 8. (i) *The image of (1) under the forgetful functor $\mathbb{U}\colon \omega\mathsf{Coh} \to \mathsf{Mes}$ is equal to the chain*

$$1 \xleftarrow{!} \mathbb{M}1 \xleftarrow{\mathbb{M}!} \mathbb{M}^2 1 \xleftarrow{\mathbb{M}^2 !} \mathbb{M}^3 1 \longleftarrow \cdots \tag{2}$$

similarly obtained by iterating the functor $\mathbb{M}$.

(ii) *The forgetful functor $\mathbb{U}\colon \omega\mathsf{Coh} \to \mathsf{Mes}$ maps $\mathbb{V}^\omega 1$ to $\mathbb{M}^\omega 1$.*

(iii) *The image of the connecting map $\mathbb{V}^\omega 1 \leftarrow \mathbb{V}(\mathbb{V}^\omega 1)$ under $\mathbb{U}$ is the connecting map $\mathbb{M}^\omega 1 \leftarrow \mathbb{M}(\mathbb{M}^\omega 1)$.*

---

[3] In a Polish space any sub σ-field of the Borel σ-field which is countably generated and separates points is equal to the whole Borel σ-field [3].

Proof. (i) follows from Proposition 6; then (ii) follows from (i) and Proposition 7. Finally (iii) follows from (ii) and Proposition 6. ◻

Theorem 1. *The greatest fixed point of the functor* $\mathbb{V}(-)^{\mathrm{Act}}$ *can be given the structure of a final labelled Markov process.*

Proof. For simplicity we prove the theorem for the case that Act is a singleton. Since $\mathbb{V}$ restricts to a locally continuous functor on $\omega\mathsf{Coh}$, the fixed point theorem of Smyth and Plotkin [21] tells us that the connecting map $\mathbb{V}^\omega 1 \leftarrow \mathbb{V}(\mathbb{V}^\omega 1)$ is an isomorphism. It follows from Proposition 8 (iii), that the connecting map $\mathbb{M}^\omega 1 \leftarrow \mathbb{M}(\mathbb{M}^\omega 1)$ is also an isomorphism. The inverse of this last map gives $\mathbb{M}^\omega 1$ the structure of an $\mathbb{M}$-coalgebra. That this coalgebra is final follows from a simple categorical argument, cf. [1]. ◻

Remark 1. Desharnais *et al.* [10] consider the solution of the domain equation $D \cong \mathbb{V}(D)^{\mathrm{Act}}$. Theorem 1 shows that D can be given the structure of a final labelled Markov process. By similar reasoning, D in its Scott topology, can be given the structure of a final coalgebra of the endofunctor $\mathbb{V}(-)^{\mathrm{Act}}$ on Top. We exploit this last observation in Proposition 9.

7 A Metric for the Lawson Topology

Now consider the domain D from Remark 1 qua labelled Markov process; denote the transition probability function by μ. For any formula $f \in \mathcal{F}$, the induced map $f\colon D \to [0,1]$ is monotone and Lawson continuous. This follows by induction on $f \in \mathcal{F}$ using the coincidence of the Lawson and weak topologies on $\mathbb{V}D$. We define a preorder $\preccurlyeq$ on D by $x \preccurlyeq y$ iff $f(x) \leqslant f(y)$ for all $f \in \mathcal{F}$. Since each formula gets interpreted as a monotone function on D it holds that $x \sqsubseteq y$ implies $x \preccurlyeq y$. The theorem below asserts that the converse also holds.

Theorem 2. *The order on* D *coincides with* $\preccurlyeq$*.*

Desharnais *et al.* [10] have proven a corresponding version of Theorem 2 in which formulas have the usual Boolean semantics. In fact, one can deduce Theorem 2 from this result and another result of the same authors [9, Corollary 3.8] which relates the Boolean and real valued semantics for the logic in the case of finite labelled Markov processes. However, we include a direct topological proof (below) as a nice application of the Lawson = weak coincidence, and because we will need to use this theorem later.

Note that in the following proposition we distinguish between an upper set $V \subseteq D$, and a $\preccurlyeq$-upper set $U \subseteq D$ ($x \in U$ and $x \preccurlyeq y$ implies $y \in U$).

Proposition 9. *If* $a \in \mathrm{Act}$, $x \preccurlyeq y$ *and* $U \subseteq D$ *is Scott open and* $\preccurlyeq$*-upper, then* $\mu_{x,a}(U) \leqslant \mu_{y,a}(U)$.

Proof. Since U is the countable union of sets of the form $\uparrow K$ for finite subsets K of U, it suffices to show that $\mu_{x,a}(\uparrow K) \leq \mu_{y,a}(\uparrow K)$ for all finite subsets K of U.

Let $K = \{x_1, \dots, x_n\} \subseteq U$ and $z \in D \setminus U$ be given. For each $j \in \{1, ..., n\}$, since $x_j \not\preccurlyeq z$, there exists a formula $g_j \in \mathcal{F}$ such that $g_j(x_j) > g_j(z)$. Since $\mathcal{F}$ is closed under truncated subtraction, and each g_j is Lawson continuous, we may, without loss of generality, assume that $g_j(x_j) > 0$ and g_j is identically zero in a Lawson open neighbourhood of z.

If we set $g = \max_j g_j$, then $g \in \mathcal{F}$ is identically zero in a Lawson open neighbourhood of z and is bounded away from zero on $\uparrow K$. Since $D \setminus U$ is Lawson compact (being Lawson closed) and $\mathcal{F}$ is closed under finite minima, we obtain $f \in \mathcal{F}$ such that f is identically zero on $D \setminus U$ and is bounded away from zero on $\uparrow K$ by, say, $r > 0$. Finally, setting $h = \min(f, r)$, we get

$$\mu_{x,a}(\uparrow K) \leqslant \frac{1}{r} \int h \, d\mu_{x,a} \leqslant \frac{1}{r} \int h \, d\mu_{y,a} \leqslant \mu_{y,a}(U),$$

where the middle inequality follows from $(\langle a \rangle h)(x) \leqslant (\langle a \rangle h)(y)$.

Since U is the (countable) directed union of sets of the form $\uparrow K$ for finite $K \subseteq U$, it follows that $\mu_{x,a}(U) \leqslant \mu_{y,a}(U)$. $\square$

We can now complete the proof of Theorem 2. Let ΣD denote the Scott topology on D and τ the topology of Scott open $\preccurlyeq$-upper sets. Consider the following diagram, where ι is the continuous map given by $\iota x = x$.

$$\begin{tikzcd}
(D, \Sigma D) \arrow[r, "\mu"] \arrow[d, "\iota"'] & \mathbb{V}(D, \Sigma D)^{\mathrm{Act}} \arrow[d, "\mathbb{V}\iota^{\mathrm{Act}}"] \\
(D, \tau) \arrow[r, dashed] & \mathbb{V}(D, \tau)^{\mathrm{Act}}
\end{tikzcd} \tag{3}$$

All the solid maps are bijections, so there is a unique dotted arrow making the diagram commute in the category of sets. The inverse image of a sub-basic open in $\mathbb{V}\langle D, \tau\rangle$ under the dotted arrow is τ-open by Proposition 9. By the finality of $\langle D, \mu\rangle$ qua $\mathbb{V}(-)^{\mathrm{Act}}$-coalgebra, ι has a continuous left inverse, and is thus a $\mathbb{V}^{\mathrm{Act}}$-homomorphism. Hence, for each $y \in D$, the Scott closed set $\downarrow y$ is τ-closed, and thus $\preccurlyeq$-lower. Thus $x \preccurlyeq y$ implies $x \sqsubseteq y$. $\square$

Since we view D as a labelled Markov process, we can consider the metric d on D as defined in Section 2.

Theorem 3. *The Lawson topology on D is induced by d.*

Proof. Since the Lawson topology on D is compact, and, by Theorem 2, the topology induced by d is Hausdorff, it suffices to show that the Lawson topology is finer. Now if $x_n \to x$ in the Lawson topology, then $f(x_n) \to f(x)$ for each $f \in \mathcal{F}$, since each formula gets interpreted as a Lawson continuous map. But d may be uniformly approximated on D to any given tolerance by looking at a finite set of formulas, cf. [6, Lemma 3]. (This lemma crucially uses the assumption $c < 1$ from the definition of d.) Thus $d(x_n, x) \to 0$ as $n \to \infty$. $\square$

8 Testing

Our aim in this section is to characterize the order on the domain D as a testing preorder. The testing formalism we use is that set forth by Larsen and Skou [16]; the idea is to specify an interaction between an experimenter and a process. The way a process responds to the various kinds of tests determines a simple and intuitive behavioural semantics.

Definition 6. *The set of* tests $t \in \mathcal{T}$ *is defined according to the grammar*

$$t ::= \omega \mid a.t \mid (t, \ldots, t),$$

where $a \in \mathrm{Act}$.

The most basic kind of test, denoted ω, does nothing but successfully terminate. $a.t$ specifies the test: see if the process is willing to perform the action a, and in case of success proceed with the test t. Finally, $(t_1, \ldots, t_n)$ specifies the test: make n copies of (the current state of) the process and perform the test t_i on the i-th copy for each i. This facility of copying or replicating processes is crucial in capturing branching-time equivalences like bisimilarity. We usually omit to write ω in non-trivial tests.

Definition 7. *To each test t we associate a set O_t of* possible observations *as follows.*

$$O_\omega = \{\omega^{\checkmark}\} \qquad O_{a.t} = \{a^{\times}\} \cup \{a^{\checkmark}e \mid e \in O_t\} \qquad O_{(t_1,\ldots,t_n)} = O_{t_1} \times \ldots \times O_{t_n}.$$

The only observation of the test ω is successful termination, $\omega^{\checkmark}$. Upon performing $a.t$ one possibility, denoted by $a^{\times}$, is that the a-action fails (and so the test terminates unsuccessfully). Otherwise, the a-action succeeds and we proceed to observe e by running t in the next state; this is denoted $a^{\checkmark}e$. Finally an observation of the test $(t_1, ..., t_n)$ is a tuple $(e_1, ..., e_n)$ where each e_i is an observation of t_i.

Definition 8. *For a given test t, each state x of a labelled Markov process $\langle X, \Sigma, \mu\rangle$ induces a probability distribution $P_{t,x}$ on O_t. The definition of $P_{t,x}$ is by structural induction on t as follows.*

$$P_{\omega,x}(\omega^{\checkmark}) = 1, \;\; P_{a.t,x}(a^{\times}) = 1 - \mu_{a,x}(X), \;\; P_{a.t,x}(a^{\checkmark}e) = \int(\lambda y.P_{t,y}(e))d\mu_{a,x}$$
$$P_{(t_1,\ldots,t_n),x}(e_1, \ldots, e_n) = \prod_{i=1}^{n} P_{t_i,x}(e_i).$$

The following theorem, proved in an earlier paper [6], shows how bisimilarity may be characterized using the testing framework outlined above. This generalizes a result of Larsen and Skou from discrete probabilistic transition systems satisfying the minimal deviation assumption[4] to labelled Markov processes.

[4] A discrete labelled Markov process $\langle X, \Sigma, \mu\rangle$ satisfies the minimal deviation assumption if the set $\{\, \mu_{x,a}(y) \mid x, y \in X \,\}$ is finite for each $a \in \mathrm{Act}$.

Theorem 4. *Let $\langle X, \Sigma, \mu \rangle$ be a labelled Markov process. Then $x, y \in X$ are bisimilar if and only if $P_{t,x}(E) = P_{t,y}(E)$ for each test t and $E \subseteq O_t$, where $P_{t,x}(E) = \sum_{e \in E} P_{t,x}(e)$.*

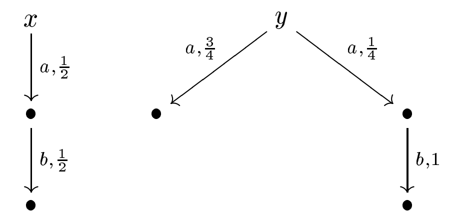
In fact the statement of Theorem 4 can be sharpened somewhat, as we now explain. For each test t there is a distinguished observation, denoted $t^{\checkmark}$, representing complete success – no action is refused. For instance, if $t = a.(b, c)$ then the completely successful observation is $a^{\checkmark}(b^{\checkmark}, c^{\checkmark})$.

Theorem 5. *Let $\langle X, \Sigma, \mu \rangle$ be a labelled Markov process. Then $x, y \in X$ are bisimilar iff $P_{t,x}(t^{\checkmark}) = P_{t,y}(t^{\checkmark})$ for all tests t.*

The idea is that for any test t and $E \subseteq O_t$, the probability of observing E can be expressed in terms of the probabilities of making completely successful observations on all the different 'subtests' of t using the principle of inclusion-exclusion. For example, if $t = a.(b, c)$; then the probability of observing $a^{\checkmark}(b^{\checkmark}, c^{\times})$ in state x is equal to $P_{t_1,x}(t_1^{\checkmark}) - P_{t,x}(t^{\checkmark})$ where $t_1 = a.b$.

Given Theorem 5 one might conjecture that x is simulated by y if and only if $P_{t,x}(t^{\checkmark}) \leqslant P_{t,y}(t^{\checkmark})$ for all tests t. However, the following example shows that to characterize simulation one really needs negative observations.

Example 2. Consider the labelled Markov process $\langle X, \Sigma, \mu \rangle$ depicted below, with distinguished states x and y and label set Act $= \{a, b\}$.



It is readily verified that $P_{t,x}(t^{\checkmark}) \leqslant P_{t,y}(t^{\checkmark})$ for all tests t. However x is not simulated by y. Indeed, consider the test $t = a.(b, b)$ with

$$E = \{a^{\checkmark}(b^{\times}, b^{\checkmark}), a^{\checkmark}(b^{\checkmark}, b^{\times}), a^{\checkmark}(b^{\checkmark}, b^{\checkmark})\}.$$

If x were simulated by y, then it would follow from Theorem 6 that $P_{t,x}(E) \leqslant P_{t,y}(E)$. But it is easy to calculate that $P_{t,x}(E) = 3/8$ and $P_{t,y}(E) = 1/4$; thus E witnesses the fact that x is not simulated by y.

The example above motivates the following definition. For each test t we define a partial order $\leqslant$ on the set of observations O_t as follows.

1. $a^{\times} \leqslant a^{\checkmark} e$
2. $a^{\checkmark} e \leqslant a^{\checkmark} e'$ if $e \leqslant e'$
3. $(e_1, ..., e_n) \leqslant (e'_1, ..., e'_n)$ if $e_i \leqslant e'_i$ for $i \in \{1, ..., n\}$.

Theorem 6. *Let $\langle X, \Sigma, \mu \rangle$ be a labelled Markov process. Then $x \in X$ is simulated by $y \in X$ iff $P_{t,x}(E) \leqslant P_{t,y}(E)$ for all tests t and upper sets $E \subseteq O_t$.*

We stress that the Theorem characterizes simulation in terms of the measure of *upper sets* E. The 'only if' direction in the above theorem follows from a straightforward induction on tests. The proof of the 'if' direction relies on the definition and lemma below. The idea behind Definition 9 is that one can determine the approximate value of a PML formula in a state x by testing x. This is inspired by [16, Theorem 8.4] where Larsen and Skou show how to determine the truth or falsity of a PML formula using testing. Our approach differs in two respects. Firstly, since we restrict our attention to the positive fragment of the logic it suffices to consider upward closed sets of observations. Also, since we interpret formulas as real-valued functions we can test for the approximate truth value of a formula. It is this last fact that allows us to dispense with the minimal deviation assumption and more generally the assumption of the discreteness of the state space.

Definition 9. *Let $\langle X, \Sigma, \mu \rangle$ be a labelled Markov process. Let $f \in \mathcal{F}$, $0 \leqslant \alpha < \beta \leqslant 1$ and $\delta > 0$. Then there exists a $t \in \mathcal{T}$ and $E \subseteq O_t$ such that for all $x \in X$,*

- *whenever $f(x) \geq \beta$ then $P_{t,x}(E) \geq 1 - \delta$ and*
- *whenever $f(x) \leq \alpha$ then $P_{t,x}(E) \leq \delta$.*

In this case, we say that t is a test for $\langle f, \alpha, \beta \rangle$ with evidence set E and significance level δ.

Thus, if we run t in state x and observe $e \in E$ then with high confidence we can assert that $f(x) > \alpha$. On the other hand, if we observe $e \notin E$ then with high confidence we can assert that $f(x) < \beta$.

Lemma 1. *Let $\langle X, \Sigma, \mu \rangle$ be a labelled Markov process. Then for any $f \in \mathcal{F}$, $0 \leqslant \alpha < \beta \leqslant 1$ and $\delta > 0$, there is a test t for (f, α, β) with level of significance δ and whose associated evidence set $E \subseteq O_t$ is upward closed.*

A proof of Lemma 1 may be found in an appendix to a fuller version of this paper [7]. The lemma implies that if $P_{t,x}(E) \leqslant P_{t,y}(E)$ for all tests t and upper sets $E \subseteq O_t$, then $f(x) \leqslant f(y)$ for all PML formulas f. It follows from Theorem 2 that x is simulated by y. This completes the proof of the 'if' direction of Theorem 6.

9 Summary and Future Work

The theme of this paper has been the use of domain-theoretic and coalgebraic techniques to analyze labelled Markov systems. These systems, which generalize the discrete labelled probabilistic processes investigated by Larsen and Skou [16], have been studied by Desharnais *et al* [8–10] and in earlier papers by some of the authors of this paper [4–6]. In part, we use domain theory to replace more traditional functional-analytic techniques in earlier papers.

In future, we intend to apply our domain theoretic approach in the more general setting of processes which feature both nondeterministic and probabilistic choice. We believe such a model will be useful in a number of areas, including for example in the analysis of leak rates in covert channels that arise in the study of non-interference.

References

1. Adámek, J. and V. Koubek: On the greatest fixed point of a set functor. Theoretical Computer Science, 150 (1995) 57–75.
2. Alvarez-Manilla, M., A. Edalat, and N. Saheb-Djahromi. An extension result for continuous valuations. Journal of the London Mathematical Society, 61 (2000) 629–640.
3. Averson, W.: An Invitation to C*-Algebras, Springer-Verlag, 1976.
4. van Breugel, F. and J. Worrell: Towards quantitative verification of probabilistic transition system, Lecture Notes in Computer Science, 2076 (2001), Springer-Verlag, 421–432.
5. van Breugel, F. and J. Worrell: An algorithm for quantitative verification of probabilistic transition systems, Lecture Notes in Computer Science, 2154 (2001), Springer-Verlag, 336–350.
6. van Breugel, F., S. Shalit and J. Worrell: Testing labelled Markov processes, Lecture Notes in Computer Science, 2380 (2002), Springer-Verlag, 537–548.
7. van Breugel, F., M. Mislove, J. Ouaknine and J. Worrell: An intrinsic characterization of approximate probabilistic bisimilarity, Report CS-2003-01, York University, 2003.
8. Desharnais, J., A. Edalat and P. Panangaden: A logical characterization of bisimulation for labelled Markov processes, In *Proc. 13th IEEE Symposium on Logic in Computer Science* (1998), IEEE Press, 478–487.
9. Desharnais, J., V. Gupta, R. Jagadeesan and P. Panangaden: Metrics for labeled Markov systems, Lecture Notes in Computer Science, 1664 (1999), Springer-Verlag, 258–273.
10. Desharnais, J., V. Gupta, R. Jagadeesan and P. Panangaden: Approximating labeled Markov processes. In *Proc. 15th Symposium on Logic in Computer Science*, (2000), IEEE Press, 95–106.
11. Edalat, A.: When Scott is weak at the top, Mathematical Structures in Computer Science, 7 (1997), 401–417.
12. Giry, M.: A categorical approach to probability theory, Lecture Notes in Mathematics, 915 (1981), Springer-Verlag, 68–85.
13. Heckmann, R.: Spaces of valuations, Annals of the New York Academy of Sciences, 806 (1996), New York Academy of Sciences, 174–200.
14. Jones, C.: Probabilistic nondeterminism, PhD Thesis, Univ. of Edinburgh, 1990.
15. Jung, A. and R. Tix: The troublesome probabilistic powerdomain. Electronic Notes in Theoretical Computer Science, 13 (1998), Elsevier.
16. Larsen, K.G. and A. Skou: Bisimulation through probabilistic testing: Information and Computation, 94 (1991), Academic Press, 1–28.
17. Milner, R.: Communication and Concurrency, Prentice Hall, 1989.
18. Park, D.: Concurrency and automata on infinite sequences. Lecture Notes in Computer Science, 104 (1981), Springer-Verlag, 167–183.
19. Parthasarathy, K.R.: Probability Measures on Metric Spaces, Academic Press, 1967.
20. Di Pierro, A., C. Hankin, and H. Wiklicky, Approximate non-interference, In *Proc. 15th IEEE Computer Security Foundations Workshop* (2002), IEEE Press, 3-17.
21. Smyth, M.B. and G.D. Plotkin: The category theoretic solution of recursive domain equations, SIAM Journal of Computing, 11 (1982), 761–783.