Domain Theory, Testing and Simulation for Labelled Markov Processes

Franck van Breugel [a,4], Michael Mislove [b,1,3], Joël Ouaknine [c,2], James Worrell [b,1,3]

[a] *York University, Department of Computer Science, 4700 Keele Street, Toronto M3J 1P3, Canada*

[b] *Department of Mathematics, Tulane University, 6823 St Charles Avenue, New Orleans LA 70118, USA*

[c] *Computer Science Department, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh PA 15213, USA*

Abstract

This paper presents a fundamental study of similarity and bisimilarity for *labelled Markov processes*. The main results characterize similarity as a testing preorder and bisimilarity as a testing equivalence. In general, labelled Markov processes are not required to satisfy a finite-branching condition—indeed the state space may be a continuum, with the transitions given by arbitrary probability measures. Nevertheless we show that to characterize bisimilarity it suffices to use finitely-branching labelled trees as tests.

Our results involve an interaction between domain theory and measure theory. One of the main technical contributions is to show that a final object in a suitable category of labelled Markov processes can be constructed by solving a domain equation $D \cong \mathbb{V}(D)^{\mathrm{Act}}$, where $\mathbb{V}$ is the probabilistic powerdomain. Given a labelled Markov process whose state space is an analytic space, bisimilarity arises as the kernel of the unique map to the final labelled Markov process. We also show that the metric for approximate bisimilarity introduced by Desharnais, Gupta, Jagadeesan and Panangaden generates the Lawson topology on the domain D.

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

1 Introduction

It is a notable feature of concurrency theory that there are many different notions of process equivalence. These are often presented in an abstract manner, e.g., using coinduction or domain theory. Ultimately, however, one would like to know that any proposed notion of equivalence has some interpretation in terms of the observable behaviour of a process. One way of formalizing this is via a testing framework [1,5,20]. The idea is to specify an interaction between a tester and the process. The latter is seen as a black box, with hidden internal state, and an interface consisting of buttons by which the tester may control the execution of the process. If the tester cannot distinguish two processes, then they are deemed equivalent. By varying the power of the tester one recovers different equivalences and preorders, e.g., trace equivalence, failures equivalence, simulation, bisimulation, etc. In some cases a testing framework can be used to give a denotational semantics, where the meaning of a process is a function from tests to observations.

This paper presents a testing framework characterizing similarity and bisimilarity for *labelled Markov processes* (**LMP**s). One can view **LMP**s as probabilistic versions of labelled transition systems from concurrency theory, or, alternatively, as indexed collections of discrete-time Markov processes in the sense of classical probability theory. More precisely, a labelled Markov process consists of a measurable space (X, Σ) of states, a family Act of actions, and, for each $a \in \mathrm{Act}$, a transition probability function $\mu_{-,a}$ that, given a state $x \in X$, 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.

Probabilistic models have been studied for quite a while in automata theory and in formal verification, but, for understanding our concerns, the paper of Larsen and Skou [20] is a good starting point. In particular, Larsen and Skou adapted the notion of bisimilarity to discrete probabilistic labelled transition systems. They defined an equivalence relation R on the states of a system to be a bisimulation if related states have exactly matching probabilities of making a transition into any given R-equivalence class. The two main results of [20] characterize bisimilarity as, respectively, equivalence with respect to a probabilistic version of Hennessy-Milner logic, and equivalence with respect to a class of tests similar to those of Abramsky [1] and Bloom and Meyer [5].

Larsen and Skou's probabilistic transition systems are **LMP**s with discrete transition probabilities. Desharnais, Edalat and Panangaden [9] extended the notion of bisimilarity to **LMP**s with arbitrary transition probabilities, and gave a suite of examples motivating the more general model. The main result of [9] is an extension of the logical characterization of bisimilarity to the general setting. In fact, they used a simpler logic than Larsen and Skou—a logic without disjunction. In another paper, Desharnais, Gupta, Jagadeesan and Panangaden [12] gave a logical characterization of similarity of **LMP**s, showing that, in this case, disjunction is essential.

In this paper we generalize the other main result of [20]—the characterization of bisimilarity as a testing equivalence—to the **LMP** model. Our results follow an intriguingly similar pattern to those of [9,12]. In particular, we find that we can simplify the class of tests used by Larsen and Skou to characterize bisimilarity. This validates an intuition of [9] that working with **LMP**s provides the right level of generality for developing the basic theory of probabilistic bisimilarity—even if ultimately one is only interested in discrete systems. Furthermore, and similarly to [12], we find that to characterize similarity, we need to enrich the set of tests with a kind of disjunction. We discuss the parallel between our results and those of [9,12] at greater length in the conclusion.

The tests that we use to characterize bisimilarity are technically just finite trees whose edges are labelled by actions—in other words, traces with branch points. Two states of an **LMP** are bisimilar just in case they pass each test with the same probability. In order to capture similarity, we need to consider a more structured class of trees, where the nodes are labelled with propositional formulas. This provides both a conjunctive and disjunctive mode of combining tests. We show that one state of a process simulates another state just in case it passes each test with a higher probability.

Although we regard the testing results as the highlight of the present paper, they do not appear until Section 8. The main body of this work is concerned with a domain-theoretic analysis of **LMP**s, upon which the results of Section 8 ultimately depend. The central mathematical construction here is the derivation of a final **LMP** as the solution of a domain equation involving the probabilistic powerdomain. The same domain equation was studied in [12], where its status as a universal **LMP** was also described. However, while the construction is the same, we a give a different, functorial justification of the universal property. We also contribute a new result by relating the Lawson topology on the universal **LMP** with the metric for approximate bisimilarity of **LMP**s from [6,10]. In particular, this shows that the class of all **LMP**s is compact with respect to this metric.

Next we give, section by section, a summary of the contents of the paper.

Section 2 presents some preliminary notions from domain theory and measure theory.

In Section 3 we formally introduce **LMP**s and the appropriate morphisms between them: zig-zag maps. While bisimulations could simply be defined to be the kernels of zig-zag maps, following [12] we show that for an **LMP** whose state space is analytic there is a less abstract relational characterization.

After introducing the probabilistic powerdomain $\mathbb{V}(D)$ in Section 4, in Section 5 we investigate the Lawson topology on $\mathbb{V}(D)$, characterizing it as a weak topology in the sense of measure theory. This yields another proof of the result of Jung and Tix [19] that the probabilistic powerdomain of a coherent domain is itself coherent.

In Section 6 we show that the canonical solution of the domain equation $D \cong \mathbb{V}(D)^{\mathrm{Act}}$ can be given the structure of a final **LMP**. The significance of this construction is that we can reduce questions about **LMP**s in general to questions about the domain D—and so take advantage of certain nice properties of D, like Lawson compactness.

In order to study bisimilarity on an **LMP**, Desharnais, Gupta, Jagadeesan and Panangaden [10] introduce a kind of dual space: a certain lattice of measurable functions on the state space. In Section 7, applying the reduction technique alluded to above, we study this class of functions in the case of the final **LMP**. In this case the given functions are all Lawson continuous. Using this observation we show that two states of an **LMP** are bisimilar iff they are indistinguishable by functions in the dual space. This result is the foundation for our main theorems concerning testing. These theorems are proven in Section 8.

2 Preliminaries

In this section we outline some basic definitions and results from domain theory and from measure theory. This is intended as a convenient summary for the reader. A more detailed treatment of the relevant domain theory and measure theory can be found respectively in Gierz et al. [15] and Arveson [4].

2.1 Domain Theory

Let $(P, \sqsubseteq)$ be a poset. Given $A \subseteq P$, we write $\uparrow A$ for the set $\{x \in P \mid (\exists a \in A)\, a \sqsubseteq x\}$; similarly, $\downarrow A$ denotes $\{x \in P \mid (\exists a \in A)\, x \sqsubseteq a\}$. A *directed complete partial order (dcpo)* is a poset P in which each directed set A has a least upper bound, denoted $\sqcup A$. If P is a dcpo, and $x, y \in P$, 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 $\Downarrow y = \{x \in D \mid x \ll y\}$; we say that P is *continuous* if it has a *basis*, i.e., a subset $B \subseteq P$ such that for each $y \in P$, $\Downarrow y \cap B$ is directed with supremum y. We use the term *domain* to mean a continuous dcpo. If a continuous dcpo has a countable basis we say that it is *ω-continuous.*

A subset U of a domain 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 σ_D 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 $\twoheaduparrow x$ where $x \in D$ form a basis for this topology. Given domains 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 $A \subseteq D$, $f(\sqcup A) = \sqcup f(A)$.

In fact the topological and order-theoretic views of a domain are interchangeable. The order on a domain can be recovered from the Scott topology as the *specialization preorder*. Recall that for a topological space X the specialization preorder $\leqslant \subseteq X \times X$ is defined by $x \leqslant y$ iff $x \in \mathrm{Cl}(y)$.

Another topology of interest on a domain D is the *Lawson topology*. This 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 $\{\twoheaduparrow x \setminus \uparrow F \mid x \in D, F \subseteq D \text{ finite}\}$ as a basis. The Lawson topology on a domain is always Hausdorff. A domain that is compact in its Lawson topology is called *coherent*.

2.2 Measure Theory

Recall that a *σ-field* Σ on a set X is a collection of subsets of X containing $\emptyset$ and closed under complements and countable unions. The pair $\langle X, \Sigma\rangle$ is called a *measurable space*. For any collection $\mathcal{C}$ of subsets on X there is a smallest σ-field containing $\mathcal{C}$, written $\sigma(\mathcal{C})$. In case X is a topological space and $\mathcal{C}$ is the class of open subsets, then $\sigma(\mathcal{C})$ is called the *Borel σ-field* on X. One can split the definition of a σ-field into two steps. A collection of subsets of X is called a *π-system* if it closed under finite intersections. A collection of subsets of X closed under countable disjoint unions, complements, and containing the empty set is called a λ-system. The $\pi - \lambda$ theorem [14] states that if $\mathcal{P}$ is a π-system, $\mathcal{L}$ is a λ-system, and $\mathcal{P} \subseteq \mathcal{L}$, then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.

If $\Sigma = \sigma(\mathcal{C})$ for some countable set $\mathcal{C}$, then we say that Σ is *countably generated*. We say that (X, Σ) is *countably separated* if there is a countable subset $\mathcal{C} \subseteq \Sigma$ such that no two distinct elements of X lie in precisely the same members of $\mathcal{C}$. A topological space is a *Polish* space if it is separable and completely metrizable.

Given a measurable space $\langle X, \Sigma\rangle$, we say that $A \subseteq X$ is (Σ-)measurable if $A \in \Sigma$. If $\langle X', \Sigma'\rangle$ is another measurable space, a function $f\colon X \to X'$ is said to be measurable if $f^{-1}(A) \in \Sigma$ for each $A \in \Sigma'$. Measurable spaces and functions form a category Mes. The limit of a diagram in Mes in obtained by equipping the limit of the underlying diagram in the category of sets with the smallest σ-field structure making all the projections measurable.

A function $\mu\colon \Sigma \to [0,1]$ is a *sub-probability measure* on $\langle X, \Sigma\rangle$ if $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$ for any countable family of pairwise disjoint measurable sets $\{A_n\}$.

3 Labelled Markov Processes

Assume a fixed countable set Act of actions or labels. A labelled Markov process is just an Act-indexed family of Markov processes on the same state space.

Definition 1 *A labelled Markov process (**LMP**) 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 sub-probability measure, and
(2) for all $a \in \mathrm{Act}$ and $A \in \Sigma$, the function $\mu_{-,a}(A) : X \to [0,1]$ is measurable.

This is the so-called reactive model of probabilistic processes. The function $\mu_{-,a}$ describes the reaction of the process to the action a selected by the environment. Given that the process is in state x and action a is selected, $\mu_{x,a}(A)$ is the probability that the process makes a transition to a state in A. Note that we consider *sub*probability measures, i.e., positive measures with total mass no greater than 1. We interpret $1 - \mu_{x,a}(X)$ as the probability of refusing action a in state x. In fact, if every transition measure had mass 1, then all processes would be bisimilar (cf. Definition 3).

An important special case is when the σ-field Σ is taken to be the powerset of X. Then, for all actions a and states x, the sub-probability measure $\mu_{x,a}(\cdot)$ is completely determined by a discrete sub-probability distribution. This case corresponds to the original probabilistic transition system model of Larsen and Skou [20].

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 called a* zig-zag map *if whenever $A' \in \Sigma', x \in X$, and $a \in \mathrm{Act}$, then $\mu_{x,a}(f^{-1}(A')) = \mu'_{f(x),a}(A')$.*

Probabilistic bisimulations (henceforth just bisimulations) are the relational counterparts of zig-zag maps, and can also be seen, in a very precise way, as the probabilistic analogues of the strong bisimulations of Park and Milner [21]. They were first introduced in the discrete case by Larsen and Skou [20]. The notion of bisimulation was extended to **LMP**s in [9,12]. (Though our formulation is slightly different as we explain below.)

Definition 3 *Let $\langle X, \Sigma, \mu\rangle$ be a labelled Markov process and R a reflexive relation on X. For $A \subseteq X$, write $R(A)$ for the image of A under R. We say that R is a* simulation *if it satisfies condition (i) below, and we say that R is a* bisimulation *if it satisfies both conditions (i) and (ii).*

(i) $xRy \Rightarrow (\forall a \in \mathrm{Act})(\forall A \in \Sigma)(A = R(A) \Rightarrow \mu_{x,a}(A) \leqslant \mu_{y,a}(A))$.
(ii) $xRy \Rightarrow (\forall a \in \mathrm{Act})(\mu_{x,a}(X) = \mu_{y,a}(X))$.

We say that two states are (bi)similar if they are related by some (bi)simulation.

The notions of simulation and bisimulation are very close, reflecting the fact that **LMP**s are like deterministic systems. The extra condition $\mu_{x,a}(X) = \mu_{y,a}(X)$ in the definition of bisimulation can be seen as a 'readiness' condition: related states perform given actions with the same probability. It may not be immediately apparent that the notion of bisimulation is symmetric, however this fact is straightforward, as we now show.

Proposition 4 *Suppose R is a bisimulation on a labelled Markov process $\langle X, \Sigma, \mu\rangle$. Then the inverse R^{-1} is also a bisimulation.*

PROOF. Given $x, y \in X$, $A \in \Sigma$ and $a \in \mathrm{Act}$, we have the following chain of implications.

$$\begin{aligned} xR^{-1}y \text{ and } A = R^{-1}(A) &\Rightarrow yRx \text{ and } X \setminus A = R(X \setminus A) \\ &\Rightarrow \mu_{y,a}(X \setminus A) \leqslant \mu_{x,a}(X \setminus A) \\ &\Rightarrow \mu_{x,a}(X) - \mu_{x,a}(X \setminus A) \leqslant \mu_{y,a}(X) - \mu_{y,a}(X \setminus A) \\ &\Rightarrow \mu_{x,a}(A) \leqslant \mu_{y,a}(A). \end{aligned}$$

□

It is straightforward that the relational composition of two bisimulations on $\langle X, \Sigma, \mu\rangle$ is again a bisimulation and that the union of any family of bisimulations is a bisimulation. In particular, there is a largest bisimulation on $\langle X, \Sigma, \mu\rangle$ and it is an equivalence relation. For an equivalence relation R the two criteria in Definition 3 can be compressed into the following more intuitive condition:

$$xRy \Rightarrow (\forall a \in \mathrm{Act})(\forall A \in \Sigma)(A = R(A) \Rightarrow \mu_{x,a}(A) = \mu_{y,a}(A))\,.$$

In words: related states have matching probabilities of jumping into any measurable block of equivalence classes. This is actually the *definition* of bisimulation in [9].

Propositions 5 and 8 below make precise the connection between bisimulations and zig-zag maps. These results are implicit in [9], and our proofs recapitulate arguments from there. The one novelty below is in our use of the existence of a final **LMP** whose state space is a Polish space. This plays a similar role to the countable logic characterizing bisimilarity from [9]. We spell out this small variation in order to make our paper more self-contained.

Proposition 5 *Every bisimulation equivalence is the kernel of a zig-zag map.*

PROOF. Given a measurable space $\langle X, \Sigma \rangle$ and an equivalence relation R on X, let Σ_R be the greatest σ-field on the set of R-equivalence classes X/R such that the quotient map $q\colon X \to X/R$ is measurable. Thus $\Sigma_R = \{E \mid q^{-1}(E) \in \Sigma\}$. Now if $\langle X, \Sigma, \mu \rangle$ is an **LMP** and R is a bisimulation, it is easy to see that

$$\mu_R : X/R \times \mathrm{Act} \times \Sigma_R \to [0, 1]$$

defined by $(\mu_R)_{[x],a}(E) = \mu_{x,a}(q^{-1}(E))$ is well-defined and is the unique transition probability function making q a zig-zag map. $\square$

To prove a converse to Proposition 5 we need to use the following two results about *analytic* measurable spaces. A measurable space is said to be analytic if it is the image of a measurable map from one Polish space to another.

Theorem 6 (Corollary 3.3.1[4]) *Let $f\colon \langle X, \Sigma \rangle \to \langle X', \Sigma' \rangle$ be a surjective measurable map, where $\langle X, \Sigma \rangle$ is analytic and $\langle X', \Sigma' \rangle$ is countably separated. Then $\langle X', \Sigma' \rangle$ is also analytic.*

Theorem 7 (Theorem 3.3.5[4]) *If $\langle X, \Sigma \rangle$ is an analytic measurable space and Σ_0 a countably generated sub-σ-field of Σ that separates points in X (given $x, y \in X$ with $x \neq y$, there exists $A \in \Sigma_0$ with $x \in A$ and $y \notin A$), then $\Sigma_0 = \Sigma$.*

The importance of analycity in the present context was first realized in [9]. We do not know if the result below is true without such an assumption.

Proposition 8 *Given a zig-zag map $f\colon \langle X, \Sigma, \mu \rangle \to \langle X', \Sigma', \mu' \rangle$ with $\langle X, \Sigma \rangle$ an analytic measurable space, the kernel of f is contained in a bisimulation.*

PROOF. By Theorem 22 there is a final **LMP** whose state space is a Polish space. Since the kernel of f is contained in the kernel of the unique zig-zag map from $\langle X, \Sigma, \mu \rangle$ to this final **LMP** we may, without loss of generality, assume that $\langle X', \Sigma' \rangle$ is a Polish space. Let $R \subseteq X \times X$ denote the kernel of f, and $q\colon \langle X, \Sigma \rangle \to \langle X/R, \Sigma_R \rangle$ the quotient map in Mes. It remains to show that R is a bisimulation.

Consider the following two sub-σ-fields $\Sigma_1, \Sigma_2 \subseteq \Sigma$.

$$\Sigma_1 = \{f^{-1}(A) \mid A \in \Sigma'\}$$
$$\Sigma_2 = \{A \in \Sigma \mid A = R(A)\}$$

It is straightforward that $\Sigma_1 \subseteq \Sigma_2 \subseteq \Sigma$. Observe also that $q(\Sigma_1) := \{q(A) \mid A \in \Sigma_1\}$ and $q(\Sigma_2) := \{q(A) \mid A \in \Sigma_2\}$ are both σ-fields on X/R with

$$q(\Sigma_1) \subseteq q(\Sigma_2) \subseteq \Sigma_R\,.$$

But X/R is countably separated, being a subobject of the Polish space X', and so it is an analytic space by Theorem 6. From the fact that Σ' is countably generated and separates points it is readily seen that $q(\Sigma_1)$ is countably generated and separates points in X/R. It follows from Theorem 7 that $q(\Sigma_1) = q(\Sigma_2) = \Sigma_R$ and thence that $\Sigma_1 = \Sigma_2$.

Suppose $x, y \in X$ are chosen such that xRy and $E \subseteq X$ is an R-closed Σ-measurable set. Then $E \in \Sigma_2$ by definition of Σ_2, and so $E \in \Sigma_1$, i.e., there exists $A \in \Sigma'$ with $E = f^{-1}(A)$. Now given $a \in \mathrm{Act}$,

$$\mu_{x,a}(E) = \mu'_{f(x),a}(A) = \mu'_{f(y),a}(A) = \mu_{y,a}(E)\,.$$

□

4 The Probabilistic Powerdomain

We briefly recall some basic definitions and results about valuations and the probabilistic powerdomain. For more details see Jones [18].

Definition 9 *Let (X, τ) be a topological space. A* valuation *on X is a mapping $\mu\colon \tau \to [0,1]$ satisfying:*

- *strictness*
  $\mu\emptyset = 0$
- *monotonicity*
  $U \subseteq V$ *implies* $\mu U \subseteq \mu V$
- *modularity*
  $\mu(U \cup V) + \mu(U \cap V) = \mu U + \mu V$ *for all* U, V.
- *Scott continuity*
  $\mu(\bigcup_{i \in I} U_i) = \sup_{i \in I} \mu U_i$ *for every directed family* $\{U_i\}_{i \in I}$.

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, $r_a \in [0,1]$, 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 $U \in \tau$ and $r \in [0,1]$. The specialization order on $\mathbb{V}X$ with respect to this topology is given by $\mu \sqsubseteq \mu'$ iff $\mu U \leqslant \mu' U$ for all $U \in \tau$. $\mathbb{V}$ 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 [18] 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 10 (Edalat [13]) *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 [19] have shown that if D is a coherent domain then so is $\mathbb{V}D$. In summary we have the following proposition.

Proposition 11 *The endofunctor $\mathbb{V} \colon \mathsf{Top} \to \mathsf{Top}$ preserves the subcategory $\omega\mathsf{Coh}$ of coherent ω-continuous domains and Scott-continuous maps.*

The fact that we define the functor $\mathbb{V}$ on Top rather than just on a category of domains has a payoff later on.

Obviously, valuations bear a close resemblance to measures. In fact, any valuation on a domain D may be uniquely extended to a measure on the Borel σ-field generated by the Scott topology on D [3, Corollary 4.3]. Conversely, any Borel measure on an ω-continuous domain defines a valuation when restricted to the open sets [3, Lemma 2.5]. (ω-continuity is needed here since measures do not in general satisfy the Scott-continuity condition in the definition of valuations.) Henceforth we treat valuations and measures on ω-continuous domains as interchangeable; thus, for instance, we integrate Borel measurable functions against valuations. We also note that on ω-continuous domains the Borel σ-field generated by the Scott topology coincides with the Borel σ-field generated by the Lawson topology.

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

Given an ω-continuous domain D, we define the *weak topology* [5] on $\mathbb{V}D$ to be the weakest topology such that for any Lawson-continuous function $f\colon D \to [0,1]$, the map $\mu \mapsto \int f d\mu$ is continuous. An alternative characterization is that a net of valuations $\langle \mu_\alpha \rangle$ converges to μ in the weak topology iff $\liminf \mu_\alpha O \geqslant \mu O$ for each Lawson-open set O (cf. [22, Thm II.6.1]). Next we show that for a coherent domain D, the Lawson topology on $\mathbb{V}D$ coincides with the weak topology.

Proposition 12 (Jones [18]) *If $\mu \in \mathbb{V}D$ is an arbitrary valuation, then, given a finite set $A \subseteq D$, $\sum_{a\in A} r_a \delta_a \sqsubseteq \mu$ iff $(\forall B \subseteq A) \sum_{a \in B} r_a \leqslant \mu(\uparrow B)$.*

Proposition 13 *Given a finite subset $F \subseteq D$, $0<r<1$ and $\varepsilon>0$, there exists a finite set $\mathcal{G}$ of simple valuations such that for any valuation μ, $\mu(\uparrow F) < r$ implies $\mu \notin \uparrow \mathcal{G}$ and $\mu(\uparrow F) > r + \varepsilon$ implies $\mu \in \uparrow \mathcal{G}$.*

PROOF. Write $F = \{x_1, \ldots, x_n\}$. Let $\delta = \varepsilon/n$ and 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 $\mathcal{G}$ to be the finite set

$$\mathcal{G} = \left\{ \sum_{i=1}^{n} r_i \delta_{x_i} \mid r < \sum_{i=1}^{n} r_i \leqslant 1 \text{ and } \{r_1, \ldots, r_n\} \subseteq \mathrm{Ran}\, f_\delta \right\}.$$

Now suppose that $\mu(\uparrow F) < r$. From the definition of $\mathcal{G}$ one sees that $\nu \in \mathcal{G}$ implies $\nu(\uparrow F) > r$. It immediately follows from Proposition 12 that $\mu \notin \uparrow \mathcal{G}$.

On the other hand, suppose that $\mu(\uparrow F) > r + \varepsilon$. We show that $\mu \in \uparrow \mathcal{G}$. To this end, let $r_i = f_\delta(\mu(\uparrow x_i \setminus \bigcup_{j<i} \uparrow x_j))$ for $i \in \{1, \ldots, n\}$. Now

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

It follows that $\sum_{i=1}^{n} r_i > r$ and so $\sum_{i=1}^{n} r_i \delta_{x_i} \in \mathcal{G}$.

[5] The definite article is a bit misleading here since there is more than one weak topology in the present context. Indeed, both the Scott and Lawson topologies on $\mathbb{V}D$ can be seen as weak topologies.

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

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

□

Proposition 14 *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 13 there is a finite set $\mathcal{G}$ of simple valuations such that $\mu \notin \uparrow \mathcal{G}$ and for all valuations ν, $\nu \notin \uparrow \mathcal{G}$ 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 $\mathbb{V}D \setminus \uparrow \mathcal{G}$.

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 that $\rho = \sum_{a \in A} r_a \delta_a$ 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 15 *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 $U \subseteq D$, and*
(2) $\limsup \mu_\alpha E \leqslant \mu E$ *for all finitely generated upper sets $E \subseteq D$.*

PROOF. Combine Propositions 10 and 14. □

Corollary 16 *If D is Lawson compact, then so is $\mathbb{V}D$ and the weak and Lawson topologies agree on $\mathbb{V}D$.*

PROOF. Recall [22, Thm II.6.4] that the weak topology on the space of Borel measures on a compact Hausdorff space is itself compact. By Corollary 15, 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 [19]. Their proof is purely domain-theoretic and doesn't use the compactness of the weak topology.

6 A Final Labelled Markov Process

In this section we show that one may construct a final labelled Markov process as a fixed point $D \cong \mathbb{V}(D)^{\mathrm{Act}}$ of the probabilistic powerdomain. In order to prove this result it is convenient to use the notion of a *coalgebra of an endofunctor*.

Definition 17 *Let $\mathcal{C}$ be a category and $F\colon \mathcal{C} \to \mathcal{C}$ a functor. An F-*coalgebra *consists of an object C in $\mathcal{C}$ together with an arrow $f\colon C \to FC$ in $\mathcal{C}$. An* F-homomorphism *from an F-coalgebra $\langle C, f\rangle$ to an F-coalgebra $\langle D, g\rangle$ is an arrow $h\colon C \to D$ in $\mathcal{C}$ such that $Fh \circ f = g \circ h$:*

$$\begin{tikzcd} C \arrow[r, "h"] \arrow[d, "f"'] & D \arrow[d, "g"] \\ FC \arrow[r, "Fh"'] & FD \end{tikzcd} \tag{1}$$

F-coalgebras and F-homomorphisms form a category whose final object, if it exists, is called the final F-coalgebra.

Next we recall a standard construction of a final F-coalgebra. Let $\mathcal{C}$ be a category with a final object 1 and with limits of all ω^{op}-chains (i.e., diagrams indexed by the poset ω^{op}). Given an endofunctor $F\colon \mathcal{C} \to \mathcal{C}$ we may form the following ω^{op}-chain

$$1 \xleftarrow{!} F1 \xleftarrow{F!} F^2 1 \xleftarrow{F^2!} F^3 1 \xleftarrow{F^3!} \cdots \tag{2}$$

To be precise, the sequence of objects $F^n 1$ is defined inductively by $F^{n+1}1 = F(F^n 1)$. The unique map $F1 \to 1$ is denoted !, and the maps $F^n!$ are defined inductively by $F^{n+1}! = F(F^n!)$.

We denote the limit cone of the chain (2) by $\{F^\omega 1 \xrightarrow{\pi_n} F^n 1\}_{n<\omega}$. The universal property of this cone entails that there is a unique 'connecting map' $F(F^\omega 1) \xrightarrow{f} F^\omega 1$ such that $\pi_n \cdot f = F\pi_{n-1}$ for each $n < \omega$.

Proposition 18 ([2]) *If the connecting map f is an isomorphism, then $\langle F^\omega 1, f^{-1}\rangle$ is a final F-coalgebra.*

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 an evaluation 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 evaluations p_A are measurable. (In fact this is the smallest σ-field such that integration against any measurable function $g\colon X \to [0,1]$ yields a measurable map $\mathbb{M}X \to [0,1]$.) Next, $\mathbb{M}$ is turned into a functor $\mathsf{Mes} \to \mathsf{Mes}$ by defining

$\mathbb{M}(f)(\mu) = \mu \circ f^{-1}$ for $f \colon X \to Y$ and $\mu \in \mathbb{M}X$. This functor is studied by Giry [16].

Given a labelled Markov process $\langle X, \Sigma, \mu \rangle$, the transition probability function μ may be regarded as a measurable map $X \to \mathbb{M}(X)^{\mathsf{Act}}$, where $(-)^{\mathsf{Act}}$ denotes Act-fold product in Mes. That is, labelled Markov processes are nothing but coalgebras of the endofunctor $\mathbb{M}^{\mathsf{Act}}$ on the category Mes. Furthermore it is easy to verify that the coalgebra homomorphisms are precisely the zig-zag maps.

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. Note in passing that the σ-field underlying $\mathbb{U}D$ is also the Borel σ-field with respect to the Lawson topology on D, and can thus be regarded as the Borel σ-field on a Polish space.

Proposition 19 *$\mathbb{M} \circ \mathbb{U} = \mathbb{U} \circ \mathbb{V}$.*

PROOF. Suppose D is a coherent domain with a countable basis. Since valuations on D in its Scott topology are in one-to-one correspondence with Borel sub-probability measures on $\mathbb{U}(D)$, we have a bijection between the points of the measurable spaces $\mathbb{M}\mathbb{U}(D)$ and $\mathbb{U}\mathbb{V}(D)$. It remains to show that the underlying σ-field structures are the same.

Since D is ω-continuous, the Scott topology on D is separable, and we may choose a countable basis $\mathcal{P}$ of Scott-open sets that is closed under finite intersections and finite unions. The set of Borel sub-probability measures on D can be given a σ-field structure in the following ways.

Σ_1 is the smallest σ-field such that p_A is measurable for each Borel set $A \subseteq D$. This is the σ-field underlying $\mathbb{M}\mathbb{U}(D)$.
Σ_2 is the smallest σ-field such that p_A is measurable for each $A \in \mathcal{P}$.
Σ_3 is the Borel σ-field generated by the Scott topology on $\mathbb{V}D$. This is the σ-field underlying $\mathbb{U}\mathbb{V}(D)$.

To complete the proof of the proposition we show that $\Sigma_1 = \Sigma_2 = \Sigma_3$.

- $\Sigma_1 = \Sigma_2$. Clearly $\Sigma_2 \subseteq \Sigma_1$. For the converse, consider

$$\mathcal{L} = \{A \subseteq D \mid p_A \text{ is } \Sigma_2\text{-measurable}\}.$$

  $\mathcal{L}$ is a λ-system, i.e., it is closed under countable disjoint unions, complements and it contains D. Also, by definition of Σ_2, we have that $\mathcal{P}$ is a π-system contained in $\mathcal{L}$. By the $\lambda - \pi$ theorem we have that $\mathcal{L}$ contains the σ-field generated by $\mathcal{P}$; but this is the whole Borel σ-field on D. Thus $\Sigma_1 \subseteq \Sigma_2$ by minimality of Σ_1.

- $\Sigma_2 = \Sigma_3$. Given $A \in \mathcal{P}$, the evaluation map $p_A \colon \mathbb{V}D \to [0,1]$ is Scott continuous and thus Σ_3-measurable. By minimality of Σ_2 it follows that $\Sigma_2 \subseteq \Sigma_3$. Conversely, Σ_2 is generated by sets $\{\mu \mid \mu A > q\}$ for $A \in \mathcal{P}$ and $q \in \mathbb{Q}$. But this is a countable basis for the Scott topology on $\mathbb{V}D$; thus Σ_2 contains all Scott-open sets, and $\Sigma_3 \subseteq \Sigma_2$ by minimality of Σ_3. □

The following proposition collects together some standard facts about limits in Mes and $\omega\mathsf{Coh}$. For this reason we do not give a detailed proof, though we explain the significance of the hypotheses and give pointers to the literature.

Proposition 20 *(i)* $\omega\mathsf{Coh}$ *is closed under countable products of pointed domains.*
(ii) $\omega\mathsf{Coh}$ *is closed under limits of* ω^{op}*-chains where the chain maps are Scott-continuous upper adjoints.*
(iii) $\mathbb{U}$ *preserves the limits in (i) and (ii).*

PROOF. Limits in the category of dcpos and Scott-continuous functions are created by the forgetful functor to the category of sets (via the pointwise order) [15, Proposition IV-4.3]. The full subcategory $\omega\mathsf{Coh}$ is not in general closed under such limits; however it is closed under countable products of *pointed* domains [17, Lemma VII-3.1] and ω^{op}-limits where the bonding maps are Scott-continuous upper adjoints [15, Exercise IV-4.15].

Part (iii) follows from the conjunction of two standard facts. Firstly, the relevant limits in $\omega\mathsf{Coh}$ are also limits in Top, where domains are regarded as topological spaces in their Scott topology. Next, the forgetful functor from Top to Mes preserves countable limits of separable spaces (see, e.g., [22, Theorem 1.10]). □

Starting with the final object 1 of $\omega\mathsf{Coh}$, we construct the chain

$$1 \xleftarrow{\;!\;} \mathbb{V}^{\mathrm{Act}}1 \xleftarrow{\mathbb{V}^{\mathrm{Act}}!} (\mathbb{V}^{\mathrm{Act}})^2 1 \xleftarrow{(\mathbb{V}^{\mathrm{Act}})^2!} (\mathbb{V}^{\mathrm{Act}})^3 1 \xleftarrow{(\mathbb{V}^{\mathrm{Act}})^3!} \cdots \tag{3}$$

and write $\{(\mathbb{V}^{\mathrm{Act}})^\omega 1 \xrightarrow{\pi_n} (\mathbb{V}^{\mathrm{Act}})^n 1\}_{n<\omega}$ for the limit cone. The map $\mathbb{V}^{\mathrm{Act}}1 \xrightarrow{!} 1$ has a lower adjoint since $\mathbb{V}^{\mathrm{Act}}1$ has a least element. Thus each bonding map in (3) has a lower adjoint.

Proposition 21

(i) *The image of (3) under* $\mathbb{U}\colon \omega\mathsf{Coh} \to \mathsf{Mes}$ *is the chain*

$$1 \xleftarrow{\;!\;} \mathbb{M}^{\mathrm{Act}}1 \xleftarrow{\mathbb{M}^{\mathrm{Act}}!} (\mathbb{M}^{\mathrm{Act}})^2 1 \xleftarrow{(\mathbb{M}^{\mathrm{Act}})^2 !} (\mathbb{M}^{\mathrm{Act}})^3 1 \longleftarrow \cdots \tag{4}$$

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

(ii) $\mathbb{U}((\mathbb{V}^{\mathrm{Act}})^{\omega}1) = (\mathbb{M}^{\mathrm{Act}})^{\omega}1$.

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

PROOF. First note that Proposition 19 and 20(iii) imply that $\mathbb{M}^{\mathrm{Act}} \circ \mathbb{U} = \mathbb{U} \circ \mathbb{V}^{\mathrm{Act}}$. Part (i) immediately follows. Next, (ii) follows from (i) and Proposition 20. Finally (iii) follows from (ii) and Proposition 19. □

Theorem 22 *There is a final labelled Markov process whose state space is a Polish space.*

PROOF. The endofunctor $\mathbb{V}^{\mathrm{Act}}\colon \omega\mathsf{Coh} \to \omega\mathsf{Coh}$ is *locally continuous*: i.e., for each pair of objects $D, E \in \omega\mathsf{Coh}$ the action on homsets

$$(\mathbb{V}^{\mathrm{Act}})_{D,E} : \omega\mathsf{Coh}(D, E) \to \omega\mathsf{Coh}(\mathbb{V}(D)^{\mathrm{Act}}, \mathbb{V}(E)^{\mathrm{Act}})$$

is Scott continuous. Thus the fixed-point theorem of Smyth and Plotkin [23] tells us that the connecting map $\mathbb{V}^{\mathrm{Act}}((\mathbb{V}^{\mathrm{Act}})^{\omega}1) \to (\mathbb{V}^{\mathrm{Act}})^{\omega}1$ is an isomorphism. By Proposition 21 (iii) the connecting map $\mathbb{M}^{\mathrm{Act}}(\mathbb{M}^{\mathrm{Act}})^{\omega}1 \to (\mathbb{M}^{\mathrm{Act}})^{\omega}1$ is also an isomorphism. By Proposition 18 the inverse of this last map makes $(\mathbb{M}^{\mathrm{Act}})^{\omega}1$ a final $\mathbb{M}^{\mathrm{Act}}$-coalgebra. Moreover, since $(\mathbb{M}^{\mathrm{Act}})^{\omega}1$ is Lawson compact, and any second countable compact Hausdorff space is metrizable, $(\mathbb{M}^{\mathrm{Act}})^{\omega}1$ is a Polish space. □

Remark 23 *The solution of the domain equation* $D \cong \mathbb{V}(D)^{\mathrm{Act}}$ *has already been considered by Desharnais* et al. *[12]. What is new here is the observation that this domain is final as a 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 Lemma 28.*

7 Functional Expressions and Metrics

In this section we recall the definition of a metric for *approximate bisimilarity* due to Desharnais, Gupta, Jagadeesan and Panangaden [10]. Intuitively the metric measures the behavioural proximity of states of an **LMP**. We show that this metric generates the Lawson topology on the domain $D \cong \mathbb{V}(D)^{\mathrm{Act}}$

from Remark 23. The primary use of the results here is to be found in the analysis of testing in the following section. However, we are also able to deduce some new facts about the metric in and of itself. In particular, we show that the metric induces a compact topology on the space of all **LMP**s, and that this topology is independent of the contraction factor used in the definition of the metric (see below).

Definition 24 *The set* F *of* functional expressions *is given by the grammar*

$$f ::= 1 \mid \min(f_1, f_2) \mid \max(f_1, f_2) \mid \langle a \rangle f \mid f \ominus q$$

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

The syntax for functional expressions is closely related to the modal logic presented below in Equation (12), Section 9. One difference is that the modal connective $\langle a \rangle$ and truncated subtraction replace the single connective $\langle a \rangle_q$. However the intended semantics is quite different.

Fix a constant $0 < c \leqslant 1$. Given a labelled Markov process $\langle X, \Sigma, \mu \rangle$, a functional expression f determines a measurable function $f^c_X \colon X \to [0,1]$ according to the following rules. (We elide the subscript and superscript in f^c_X where no confusion can arise.)

$$\begin{aligned}
1(x) &= 1 \\
\min(f,g)(x) &= \min(f(x), g(x)) \\
\max(f,g)(x) &= \max(f(x), g(x)) \\
(f \ominus q)(x) &= \max(f(x) - q, 0) \\
(\langle a \rangle f)(x) &= c \textstyle\int f d\mu_{x,a}
\end{aligned}$$

In particular, $\langle a \rangle f$ is the composition

$$X \xrightarrow{\mu_{-,a}} \mathbb{M}X \xrightarrow{\int f -} [0,1] \xrightarrow{- \cdot c} [0,1] \; .$$

The left-hand map is measurable by definition of an **LMP**, while the middle map is measurable if f is measurable. Thus $\langle a \rangle f$ is measurable whenever f is measurable.

The interpretation of a functional expression f is relative to the prior choice of the constant c. The role of this constant is to discount observations made at greater modal depth. The interpretation of f is also relative to a particular **LMP**; however we have the following proposition.

Proposition 25 *Suppose* $g \colon \langle X, \Sigma, \mu \rangle \to \langle Y, \Sigma', \mu' \rangle$ *is a zig-zag map. Then for each functional expression* $f \in \mathsf{F}$, $f^c_X = f^c_Y \circ g$.

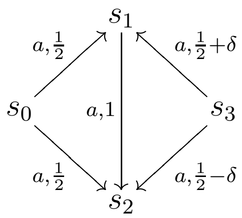
PROOF. The proof is by a straightforward induction on the structure of $f \in \mathsf{F}$. $\square$

Given an **LMP** $\langle X, \Sigma, \mu \rangle$, Desharnais et al. [10] defined a metric [6] d^c_X on the state space X by

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

It is shown in [10] 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^c_X is the same for any value of c in the open interval $(0, 1)$.

Example 26 *In the labelled Markov process below, $d^c_X(s_0, s_3) = c^2\delta$. The two states are bisimilar just in case $\delta = 0$.*



Now consider the domain $D \cong \mathbb{V}(D)^{\mathrm{Act}}$ from Remark 23 qua labelled Markov process; denote the transition probability function by μ.

Proposition 27 *For any $f \in \mathsf{F}$, the induced map $f : D \to [0, 1]$ is monotone and Lawson continuous.*

PROOF. The proof is by induction on $f \in \mathsf{F}$. The only non-trivial case is $f \equiv \langle a \rangle g$; then $f : D \to [0, 1]$ is given by the composite

$$D \xrightarrow{\mu} \mathbb{V}(D)^{\mathrm{Act}} \xrightarrow{\pi_a} \mathbb{V}D \xrightarrow{\int g d-} [0, 1] \tag{5}$$

Note that each map above is Lawson continuous—the last one by the induction hypothesis and Corollary 16. $\square$

Define a preorder $\preccurlyeq$ on D by

$$x \preccurlyeq y \text{ iff } f(x) \leqslant f(y) \text{ for all } f \in \mathsf{F}.$$

[6] Strictly speaking we should say that d^c_X is a pseudometric, since distinct states may have distance 0.

Since each functional expression gets interpreted as a monotone function, $x \sqsubseteq y$ implies $x \preccurlyeq y$. Theorem 29 asserts that the converse also holds. In order to prove this result we need the following lemma.

Note that in the lemma 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 $y \in U$).

Lemma 28 *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. Let $K = \{x_1, \ldots, x_m\} \subseteq U$ and $z \in D \setminus U$ be given. For each $i \in \{1, \ldots, m\}$, since $x_i \not\preccurlyeq z$, there exists $g_i \in \mathsf{F}$ such that $g_i(x_i) > g_i(z)$. Since F is closed under truncated subtraction, and each g_i is Lawson continuous, we may, without loss of generality, assume that $g_i(x_i) > 0$ and g_i is identically zero on a Lawson-open neighbourhood of z. Moreover, if we set $g_z = \max_i g_i$, then $g_z \in \mathsf{F}$ is identically zero in a Lawson-open neighbourhood of z and is bounded away from 0 on $\uparrow K$. Such a function g_z can be exhibited for any $z \in D \setminus U$.

Since $D \setminus U$ is Lawson compact (being Lawson closed) we can pick $z_1, \ldots, z_m \in D \setminus U$ such that $f = \min_j g_{z_j}$ 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 $\mu_{x,a}(U) \leqslant \mu_{y,a}(U)$. $\square$

Theorem 29 *The order on D coincides with $\preccurlyeq$.*

PROOF. Let σ_D 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 $\iota x = x$.

$$\begin{tikzcd} \langle D, \sigma_D \rangle \arrow[r, "\mu"] \arrow[d, "\iota"'] & \mathbb{V}\langle D, \sigma_D \rangle^{\mathrm{Act}} \arrow[d, "\mathbb{V}\iota^{\mathrm{Act}}"] \\ \langle D, \tau \rangle \arrow[r, dashed, "\mu'"'] & \mathbb{V}\langle D, \tau \rangle^{\mathrm{Act}} \end{tikzcd} \quad (6)$$

Since ι is a bijection there is a unique function μ' making the above diagram commute in the category of sets.

Recall that the topology on $\mathbb{V}\langle D, \tau\rangle$ is generated by sub-basic opens of the form $\{\nu \mid \nu U > r\}$ for $U \in \tau$ and $0<r<1$. The inverse image of such a set under μ' is Scott open by the Scott continuity of μ and is $\preccurlyeq$-upper by Lemma 28. Thus μ' is a continuous map and yields a $\mathbb{V}^{\mathrm{Act}}$-coalgebra structure on $\langle D, \tau\rangle$.

The finality of the $\mathbb{V}^{\mathrm{Act}}$-coalgebra $\langle\langle D, \sigma_D\rangle, \mu\rangle$, as indicated in Remark 23, implies that ι has a continuous left inverse, and is thus a homeomorphism. Hence, for each $y \in D$, the Scott-closed set $\downarrow y$ is τ-closed, and thus $\preccurlyeq$-lower. Thus $x \preccurlyeq y$ implies $x \sqsubseteq y$. $\square$

Corollary 30 (Theorem 4.10[11]) *Let $\langle X, \Sigma, \mu\rangle$ be a labelled Markov process with X an analytic space. Denote by $\backsim$ the bisimilarity relation on X. Then $x \backsim y$ iff $f^c_X(x) = f^c_X(y)$ for all functional expressions $f \in \mathsf{F}$.*

PROOF. Let g denote the unique zig-zag map from $\langle X, \Sigma, \mu\rangle$ to the final **LMP**, i.e., the domain D from Remark 23. Then

$$\begin{aligned} x \backsim y &\Leftrightarrow g(x) = g(y) \quad \text{by Propositions 5 and 8} \\ &\Leftrightarrow f^c_D(g(x)) = f^c_D(g(y)) \text{ for all } f \in \mathsf{F}, \text{ by Theorem 29} \\ &\Leftrightarrow f^c_X(x) = f^c_X(y) \text{ for all } f \in \mathsf{F}, \text{ by Proposition 25.} \end{aligned}$$

$\square$

Remark 31 *Corollary 30 has already appeared as [11, Theorem 4.10]. The proof there is quite different. Among other things it relies on a modal logic characterizing bisimilarity from [9], a translation between functional expressions and formulas of the modal logic, and an approximation scheme for recovering an arbitrary* **LMP** *as the join of a chain of finite-state approximants. These last two points are discussed at greater length in Section 9. We should add that [11] also proves that given an* **LMP** *$\langle X, \Sigma, \mu\rangle$, $x \in X$ is simulated by $y \in X$ just in case $f^c_X(x) \leqslant f^c_X(y)$ for all functional expressions f.*

Since we view the domain D as a labelled Markov process, we can consider the metric d^c_D as defined in Section 3. We will need the following result.

Proposition 32 (Lemma 4.6[10]) *Suppose $0<c<1$ and* Act *is finite. Then given $\varepsilon > 0$, there exists finite $\mathsf{F}' \subseteq \mathsf{F}$ such that for all $x, y \in D$*

$$0 \leqslant d^c_D(x, y) - \sup_{f \in \mathsf{F}'} |f^c_D(x) - f^c_D(y)| < \varepsilon.$$

Theorem 33 *For $0<c<1$ and finite* Act *the Lawson topology on D is induced by d^c_D.*

PROOF. The Lawson topology on D is compact. By Theorem 29, d_D^c is a metric (not just as pseduometric), and so it induces a Hausdorff topology. Thus it suffices to show that the Lawson topology is finer than the topology induced by d_D^c. Now if $x_n \to x$ in the Lawson topology, then $f(x_n) \to f(x)$ for each $f \in \mathsf{F}$, since each functional expression is interpreted as a Lawson-continuous map. Now, by Proposition 32, $d_D^c(x_n, x) \to 0$ as $n \to \infty$. $\square$

Remark 34 *Both hypotheses in the above theorem are necessary. In particular it is shown in [10] that the topology induced by d_X^c differs for $c < 1$ and $c = 1$.*

We defined a metric d_X^c for each labelled Markov process X. However, if one thinks of a labelled Markov process $X = \langle X, \Sigma, \mu \rangle$ as being equipped with a distinguished (initial) state s_X, then one can define a metric d^c on the class $\mathcal{LMP}$ of all labelled Markov processes by

$$d^c(X, Y) = \sup_{f \in \mathsf{F}} |f_X^c(s_X) - f_Y^c(s_Y)|.$$

Corollary 35 *For $0 < c < 1$ the topology on $\mathcal{LMP}$ induced by d^c is compact and independent of the value of c.*

PROOF. Consider the function $\mathcal{LMP} \to D$ mapping a labelled Markov process X to the image of the distinguished state s_X under the unique zig-zag map $X \to D$. By Proposition 25 this map is an isometry (i.e., a distance preserving map) $\langle \mathcal{LMP}, d^c \rangle \to \langle D, d_D^c \rangle$. Furthermore this map it is clearly surjective. The stated results now easily follow from Theorem 33. $\square$

8 Testing

In this section we characterize similarity on an **LMP** as a testing preorder, and bisimilarity as a testing equivalence. The testing formalism we use is that set forth by Larsen and Skou [20]. (See also Abramsky [1] and Bloom and Meyer [5] for similar formalisms.) 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.

A typical intuition is that a process is a black box whose interface to the outside world includes a button for each action $a \in \mathrm{Act}$. The most basic kind of test is to try and press one of the buttons: either the button will go down and the process will make an invisible state change (corresponding to a labelled transition), or the button doesn't go down (corresponding to a refusal). An important question arises as to which mechanisms are allowed to combine the basic button-pushing experiments. Here, following Larsen and Skou, we suppose that the tester can save and restore the state of the process at any time. Or rather we make the equivalent assumption that the tester can make multiple copies of the process in order to experiment independently on one copy at a time. The facility of copying or replicating processes is crucial in capturing branching-time equivalences like bisimilarity.

Definition 36 *The test language* T_0 *is given by the grammar*

$$t ::= 1 \mid at \mid t \cdot t$$

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

The term 1 represents the test that does nothing but successfully terminate. The term at represents the test: press button a, and in case of success proceed with test t. We usually abbreviate $a1$ to just a. Finally, $t_1 \cdot t_2$ specifies the test: make two copies of (the current state of) the process, perform the test t_i on the i-th copy, and record success in case both sub-tests succeed.

Definition 37 *Given a labelled Markov process* $\langle X, \Sigma, \mu\rangle$, *we define an indexed family* $\{P(-,t)\}_{t\in\mathsf{T}_0}$ *of real-valued random variables on* $\langle X, \Sigma\rangle$ *by*

$$\begin{aligned} P(x,1) &= 1 \\ P(x,at) &= \textstyle\int P(-,t)d\mu_{x,a} \\ P(x,t_1 \cdot t_2) &= P(x,t_1) \cdot P(x,t_2) \end{aligned}$$

Intuitively $P(x,t)$ *is the probability that state* x *passes test* t.

The following simple example motivates the inclusion of the branching construct in T_0.

Example 38 *Consider the labelled Markov process* $\langle X, \Sigma, \mu\rangle$ *over label set* $\mathrm{Act} = \{a, b\}$ *depicted below.*

$$\begin{tikzcd}
x_0 \arrow[d, "{a[1]}"] & & y_0 \arrow[dl, "{a[\frac{1}{2}]}"'] \arrow[dr, "{a[\frac{1}{2}]}"] & \\
x_1 \arrow[d, "{b[\frac{1}{2}]}"] & y_1 & & y_2 \arrow[d, "{b[1]}"] \\
x_2 & & & y_3
\end{tikzcd}$$

It is readily verified that $P(x_0,t) = P(y_0,t)$ *for any test* t *with no branching, i.e., for any trace* t. *However* x_0 *is not bisimilar to* y_0. *This is witnessed by the test* $t \equiv a(b \cdot b)$, *since* $P(x_0,t) = 1/4$ *while* $P(y_0,t) = 1/2$.

Theorem 39 *Let* $\langle X, \Sigma, \mu\rangle$ *be a labelled Markov process. Then* $x, y \in X$ *are bisimilar just in case* $P(x,t) = P(y,t)$ *for each test* $t \in \mathsf{T}_0$.

PROOF. Consider the free real vector space $V = \{\sum \lambda_i t_i \mid \lambda_i \in \mathbb{R},\, t_i \in \mathsf{T}_0\}$ over T_0. The binary product map on T_0 has a unique extension to a bilinear map $V \times V \to V$. Furthermore, $P\colon X \times \mathsf{T}_0 \to \mathbb{R}$ has a unique extension to a function $P\colon X \times V \to \mathbb{R}$ that is linear in its second argument.

$P(-, v)$ is a bounded real-valued function on X for each $v \in V$. Furthermore, the pointwise product of $P(-, v_1)$ and $P(-, v_2)$ is just $P(-, v_1 \cdot v_2)$. Let $\mathcal{A}$ denote the closure of the family of the functions $P(-, v)$ in the Banach algebra of all bounded real-valued functions on X equipped with the supremum norm. Then $\mathcal{A}$ is a closed sub-algebra, i.e., it is closed under sums, scalar multiplication and (pointwise) products. Now it is well-known that any such sub-algebra is also closed under (pointwise) binary minima and maxima (see Johnstone [17]). We recall the argument for the reader's convenience.

It is enough to show that $f \in \mathcal{A}$ implies $|f| \in \mathcal{A}$ since

$$\max(f, g) = \tfrac{1}{2}(f + g) + \tfrac{1}{2}|f - g|\,.$$

Without loss of generality, since $\mathcal{A}$ is closed under scalar multiplication, we may suppose that $-1 \leqslant f \leqslant 1$. Let $g = 1 - f^2$; then $0 \leqslant g \leqslant 1$, and

$$\begin{aligned} |f| &= \sqrt{f^2} = \sqrt{1 - g} \\ &= 1 - \tfrac{1}{2}g - \tfrac{1}{8}g^2 - \cdots - \tfrac{1\cdot 3 \cdots (2n-3)}{2^n n!} g^n - \cdots \end{aligned}$$

But this sum converges uniformly; thus $|f| \in \mathcal{A}$, and $\mathcal{A}$ is closed under pointwise binary minima and maxima.

Furthermore, given $a \in \mathrm{Act}$ and $v = \sum \lambda_i t_i \in V$, let $v' = \sum \lambda_i\, a t_i$. Then, by linearity of the integral, the function $x \mapsto \int P(-, v) d\mu_{x,a}$ is just $P(-, v')$. Thus $\mathcal{A}$ contains the interpretations of all functional expressions $f \in \mathsf{F}$.

Now suppose $x, y \in X$ are such that $P(x, t) = P(y, t)$ for all $t \in \mathsf{T}_0$. Then $P(x, v) = P(y, v)$ for all all $v \in V$. Thus $f(x) = f(y)$ for all functional expressions $f \in \mathsf{F}$, and x and y are bisimilar by Corollary 30. $\square$

Theorem 39 generalizes and simplifies a result of Larsen and Skou [20, Theorem 6.5]. The generalization is that Larsen and Skou's result only applied to discrete probabilistic transition systems satisfying the *minimal deviation assumption*. This last condition says that there is a fixed $\varepsilon > 0$ such that any transition probability $\mu_{x,a}(\{y\})$ is an integer multiple of ε. Theorem 39 simplifies [20, Theorem 6.5] in that the test language T_0 contains no negative observations or failures. We explain this point in more detail in Appendix A.

Given the fact that bisimilarity on an **LMP** is just mutual similarity, one might conjecture that $x \in X$ is simulated by $y \in X$ just in case $P(x,t) \leqslant P(y,t)$ for all $t \in \mathsf{T}_0$. However the following example shows that this is not the case.

Example 40 *Consider the process from Example 38. It is readily verified that $P(x_0,t) \leqslant P(y_0,t)$ for all $t \in \mathsf{T}_0$. However x_0 is not simulated by y_0. In particular, x_1 is only simulated by y_2, but the probability of moving from x_0 to x_1 is greater than the probability of moving from y_0 to y_2.*

There is no hope of using the elements of V to characterize similarity, since V contains negative scalar multiples of tests—so the functions $P(-,v)$ are not monotone with respect to the similarity preorder. On the other hand, if we were to restrict attention to the cone V_+ of *positive* linear combinations of elements of T_0, then in the example above we would still have $P(x_0,v) \leqslant P(y_0,v)$ for all $v \in V_+$. Nevertheless the solution we outline below does follow the general idea of using a 'monotone' subset of V as a test language.

One can think of the test $t \equiv t_1 \cdot t_2$ as a conjunction, in that t succeeds if each of its components succeeds. In order to capture similarity the idea is to consider more general truth-functional ways of combining tests.

Definition 41 *For each $n \in \mathbb{N}$, the set $\mathrm{Fma}(n)$ of propositional formulas on variables $p_1,\ldots,p_n$ is generated by the syntax*

$$\varphi ::= \top \mid p_i \mid \varphi \vee \varphi \mid \varphi \wedge \varphi\,.$$

Under the standard Boolean semantics, each $\varphi \in \mathrm{Fma}(n)$ is interpreted as a function $\varphi_{\mathbb{B}} \colon \mathbb{B}^n \to \mathbb{B}$, where $\mathbb{B} = \{\mathrm{false},\mathrm{true}\}$. We also consider a real-valued semantics where $\varphi \in \mathrm{Fma}(n)$ is interpreted as a function $\varphi_{\mathbb{R}} \colon [0,1]^n \to [0,1]$. Given $r_1,\ldots,r_n \in [0,1]$, consider n independently distributed Booelan-valued Bernoulli random variables $X_1,\ldots,X_n$, where X_i takes value true with probability r_i. We define

$$\varphi_{\mathbb{R}}(r_1,\ldots,r_n) = P(\varphi_{\mathbb{B}}(X_1,\ldots,X_n) = \mathrm{true}).$$

Definition 42 *The test language T_1 is given by the grammar*

$$t ::= at \mid \varphi(t_1,\ldots,t_n) \quad [\varphi \in \mathrm{Fma}(n)]\,.$$

Given a labelled Markov process $\langle X,\Sigma,\mu\rangle$ and $x \in X$, we extend the definition of the function $P(x,-)$ from T_0 to T_1 by

$$P(x,\varphi(t_1,\ldots,t_n)) = \varphi_{\mathbb{R}}(P(x,t_1),\ldots,P(x,t_n))\,.$$

A test $t \in \mathsf{T}_1$ can be viewed as a tree whose edges are labelled with elements of Act and such that an n-way branching node is labelled by an element of

Fma(n). Intuitively, the test $t \equiv \varphi(t_1, \ldots, t_n)$ is implemented as follows. Make n copies of the current state of the process; run test t_i on the i-th copy; record success for t if φ is true under the (Boolean) valuation $v \in \mathbb{B}^n$ given by $v_i =$ true iff t_i succeeds.

If $\varphi \equiv p_1 \vee p_2$, we abbreviate $\varphi(t_1, t_2)$ to $t_1 \vee t_2$. This test succeeds iff either of the disjuncts succeeds. The notation $t_1 \wedge t_2$ is interpreted similarly. Both notations should be employed with care since neither of these operations is idempotent. In fact, $t_1 \wedge t_2$ exactly corresponds to the test $t_1 \cdot t_2$ from the language T_0.

Theorem 43 *Let $\langle X, \Sigma, \mu \rangle$ be a labelled Markov process. Then $x \in X$ is simulated by $y \in X$ iff $P(x, t) \leqslant P(y, t)$ for all tests $t \in \mathsf{T}_1$.*

Example 44 *Recall the process from Example 38 and consider the test $t \equiv a(b \vee b)$. Then $P(x_0, t) = 3/4$ while $P(y_0, t) = 1/2$. Thus t witnesses the fact that x_0 is not simulated by y_0.*

The rest of this section is devoted to a proof of Theorem 43. This proof, which is inspired by [20, Theorem 6.5], has a statistical flavour and is strikingly different from that of Theorem 39. However, we believe that an alternative proof using the technology of compact pospaces may be possible (see [17]).

Definition 45 *Let $\langle X, \Sigma, \mu \rangle$ be a labelled Markov process. Recall that each functional expression $f \in \mathsf{F}$ defines a function $X \to [0, 1]$ (again, take $c = 1$). Given $f \in \mathsf{F}$, $0 \leqslant \alpha < \beta \leqslant 1$ and $\varepsilon > 0$, we say that $t \in \mathsf{T}_1$ is a test for $(f, \alpha, \beta, \varepsilon)$ if for all $x \in X$,*

Whenever $f(x) \geqslant \beta$ then $P(x, t) \geqslant 1 - \varepsilon$;
Whenever $f(x) \leqslant \alpha$ then $P(x, t) \leqslant \varepsilon$.

Thus, if test t succeeds on state x, then with high confidence we can assert that $f(x) > \alpha$. On the other hand, if t fails on state x, then with high confidence we can assert that $f(x) < \beta$.

Lemma 46 *Let $\langle X, \Sigma, \mu \rangle$ be a labelled Markov process. Then for any $f \in \mathsf{F}$, $0 \leqslant \alpha < \beta \leqslant 1$ and $\varepsilon > 0$, there is a test t for $(f, \alpha, \beta, \varepsilon)$.*

PROOF. The proof proceeds by induction on $f \in \mathsf{F}$. The cases $f \equiv 1$ and $f \equiv g \ominus q$ are straightforward and we omit them.

Case $f \equiv \min(f_1, f_2)$: By induction, let t_i be a test for $(f_i, \alpha, \beta, \varepsilon/2)$ for $i = 1, 2$.

Then we take $t \equiv t_1 \wedge t_2$ as a test for $(f, \alpha, \beta, \varepsilon)$. Now

$$\begin{aligned}
\min(f_1, f_2)(x) \geqslant \beta &\Rightarrow f_1(x) \geqslant \beta \;\text{ and }\; f_2(x) \geqslant \beta \\
&\Rightarrow P(x, t_1) \geqslant 1 - \varepsilon/2 \;\text{ and }\; P(x, t_2) \geqslant 1 - \varepsilon/2 \\
&\Rightarrow P(x, t) \geqslant 1 - \varepsilon,
\end{aligned}$$

and

$$\begin{aligned}
\min(f_1, f_2)(x) \leqslant \alpha &\Rightarrow f_1(x) \leqslant \alpha \;\text{ or }\; f_2(x) \leqslant \alpha \\
&\Rightarrow P(x, t_1) \leqslant \varepsilon/2 \;\text{ or }\; P(x, t_2) \leqslant \varepsilon/2 \\
&\Rightarrow P(x, t) \leqslant \varepsilon/2.
\end{aligned}$$

Case $f \equiv \max(f_1, f_2)$: Let t_i be a test for $(f_i, \alpha, \beta, \varepsilon/2)$ for $i = 1, 2$. Then we take $t \equiv t_1 \vee t_2$ as a test for $(f, \alpha, \beta, \varepsilon)$. The justification is similar to the case above.

Case $f \equiv \langle a \rangle g$: Pick $n \in \mathbb{N}$ and $\varepsilon' > 0$. By the induction hypothesis, for $1 \leqslant i \leqslant n$ we have a test t_i for $(g, \frac{i-1}{n}, \frac{i}{n}, \varepsilon')$. Pick $\varphi \in \mathrm{Fma}(n)$ such that

$$\varphi_{\mathbb{B}}(p_1, \ldots, p_n) = \text{true iff } \tfrac{1}{n}|\{i \mid p_i = \text{true}\}| \geqslant \tfrac{\beta+\alpha}{2}\,.$$

The rest of the proof is a calculation to show that for suitably large n and small ε', $t \equiv \varphi(at_1, \ldots, at_n)$ can be used as a test for $(f, \alpha, \beta, \varepsilon)$.

Fix $x \in X$. Let $\theta_1, \ldots, \theta_n$ be independent $\{0,1\}$-valued Bernoulli random variables, where $\theta_i = 1$ with probability $P(x, at_i)$. Furthermore, define $\theta = (1/n)\sum_{i=1}^{n} \theta_i$. Thus $P(x, t) = P(\theta \geqslant \frac{\beta+\alpha}{2})$.

The induction hypothesis is that for $1 \leqslant i \leqslant n$

$$g(y) \geqslant \tfrac{i}{n} \Rightarrow P(y, t_i) \geqslant 1 - \varepsilon' \tag{7}$$

$$g(y) \leqslant \tfrac{i-1}{n} \Rightarrow P(y, t_i) \leqslant \varepsilon'\,. \tag{8}$$

We estimate $P(x, at_i)$ by conditioning on the value of g using (7) and (8).

$$(1 - \varepsilon')\mu_{x,a}\left\{g \geqslant \frac{i}{n}\right\} \leqslant P(x, at_i) \leqslant \mu_{x,a}\left\{g > \frac{i-1}{n}\right\} + \varepsilon'.$$

Since $E[\theta] = \frac{1}{n}\sum_{i=1}^{n} P(x, at_i)$, it follows that

$$\frac{(1 - \varepsilon')}{n}\sum_{i=1}^{n} \mu_{x,a}\left\{g \geqslant \frac{i}{n}\right\} \leqslant E[\theta] \leqslant \frac{1}{n}\sum_{i=1}^{n} \mu_{x,a}\left\{g > \frac{i-1}{n}\right\} + \varepsilon'.$$

Whence, by a straightforward manipulation of terms in the summation,

$$(1-\varepsilon')\sum_{i=1}^{n}\frac{i}{n}\mu_{x,a}\left\{\frac{i}{n}\leqslant g<\frac{i+1}{n}\right\}\leqslant E[\theta]\leqslant\sum_{i=1}^{n}\frac{i}{n}\mu_{x,a}\left\{\frac{i-1}{n}<g\leqslant\frac{i}{n}\right\}+\varepsilon'$$

Thus we can choose ε' small enough and n large enough to ensure that

$$|E[\theta]-\textstyle\int g d\mu_{x,a}|<\frac{\beta-\alpha}{4}\,. \tag{9}$$

Since $V[\theta]=(1/n^2)\sum_{i=1}^{n}V[\theta_i]\leqslant 1/n$, by Chebyshev's inequality [14] for large n it holds that

$$P\left\{|\theta-E[\theta]|\leqslant\tfrac{\beta-\alpha}{4}\right\}\geqslant 1-\varepsilon. \tag{10}$$

It is straightforward that the choice of ε' and n required to make (9) and (10) true can be made independently of $x\in X$. Now

$$\begin{aligned}(\langle a\rangle g)(x)\geqslant\beta &\Rightarrow \textstyle\int g d\mu_{x,a}\geqslant\beta \quad \text{by definition of } \langle a\rangle g\\ &\Rightarrow E[\theta]\geqslant\tfrac{3\beta+\alpha}{4} \quad \text{by (9)}\\ &\Rightarrow P\left(\theta\geqslant\tfrac{\beta+\alpha}{2}\right)\geqslant 1-\varepsilon \quad \text{by (10)}\\ &\Rightarrow P(x,t)\geqslant 1-\varepsilon.\end{aligned}$$

Similarly it follows that $(\langle a\rangle g)(x)\leqslant\alpha\Rightarrow P(x,t)\leqslant\varepsilon$. $\Box$

Theorem 43 now follows from Lemma 46 using the characterization of simulation in terms of functional expressions from Remark 31.

9 Conclusion and Related Work

The theme of this paper has been the use of domain-theoretic and coalgebraic techniques to analyze labelled Markov processes. These systems generalize the discrete labelled probabilistic processes investigated by Larsen and Skou [20]. Our main results extend and simplify the work of Larsen and Skou on the connection between probabilistic bisimulation and testing. The direction of this generalization, and the ideas and techniques we use, are mainly inspired by the work of Desharnais, Edalat, Gupta, Jagadeesan and Panandagen [9,10,12]. In particular, as we now explain, there are several interesting parallels between the results reported here and their work on the logical characterization of bisimilarity.

A central result of Larsen and Skou [20] was a logical characterization of bisimilarity for discrete **LMP**s satisfying the minimum deviation assumption.

The formulas in their logic were generated by the grammar

$$\varphi ::= \top \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \langle a \rangle_q \varphi \mid \Delta_a \tag{11}$$

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

This is a probabilistic version of Henessey-Milner logic [21]. The semantics is given by a satisfaction relation $\vDash$ between states of a labelled Markov process and formulas. In particular, one has $x \vDash \langle a \rangle_q \varphi$ if the probability that x makes an a-labelled transition to the set of states satisfying φ exceeds q. Also $x \vDash \Delta_a$ just in case no a-transition is possible from x. This logic characterizes bisimilarity in the sense that states satisfy the same formulas just in case they are bisimilar.

In generalizing the result of Larsen and Skou beyond the discrete case, Desharnais et al. [9] realized that an even simpler logic, generated by the grammar

$$\varphi ::= \top \mid \varphi \wedge \varphi \mid \langle a \rangle_q \varphi \,, \tag{12}$$

is sufficient to characterize bisimilarity *for all* **LMP***s*. This is reflected in our observation that negative observations, or failures, are not needed to test for bisimilarity. Indeed the grammar for the smaller logic is very similar in form to the grammar for tests in Definition 36. The one significant difference is that in the grammar for tests the modalities are not indexed with numbers. Of course, the semantics of tests is completely different, with, in particular, an arithmetic interpretation of conjunction as multiplication.

It was later shown in [12] that the logic (12) is inadequate to characterize similarity: one needs to include disjunction. Again, this is reminiscent of the observation that the test language in Definition 36 doesn't characterize similarity, and that one needs to use the more general test language T_1 from Definition 42.

We would also like to clarify the relationship between parts of this work and the paper [12] on approximating **LMP**s. That work features the same domain equation $D \cong \mathbb{V}(D)^{\mathrm{Act}}$ appearing in the present paper; furthermore, the authors exhibit a two-stage construction for interpreting an arbitrary **LMP** in D. In the first stage they show how to interpret a finite-state **LMP** as an element of D. The second stage utilizes a method for unfolding and discretizing an arbitrary **LMP** $X = \langle X, \Sigma, \mu \rangle$ into finite-state approximants. In fact they produce a sequence of finite approximants, which is a chain in the simulation order, and such that any formula satisfied by X is also satisfied by one of the finite approximants. Then they define the interpretation of X in the domain D to be the join of the interpretations of its finite approximants. Using their results on the logical characterization of bisimilarity they show that each **LMP** is bisimilar to its interpretation in D. It follows that their domain-theoretic semantics is the same as our final semantics.

As far as we are aware, it was de Vink and Rutten [24] who were the first to study probabilistic transition systems as coalgebras. However, since they work with ultrametric spaces, their results only apply in the discrete setting, not to arbitrary **LMP**s. It was also noted in [9] that **LMP**s are coalgebras of the Giry functor, although this observation was not developed there.

An interesting problem, suggested by the development in Section 8, would be to realize the final **LMP** as the Gelfand-Naimark dual of an equationally presented C*-algebra. The idea would be to take the free vector space V in Section 8 and quotient by a suitable set of equations to get a commutative algebra. An issue that is as yet unresolved is how to define a suitable norm in order to get a C*-algebra. We conjecture that this can be done, and moreover that the final **LMP** can be recovered as the space of characters of the resulting algebra.

Acknowledgements

We thank the anonymous referees for numerous suggestions for improving the presentation of the paper. In particular, their comments helped us find a more transparent formulation of the testing results in Section 8.

References

[1] S. Abramsky. Observation equivalence as a testing equivalence. *Theoretical Computer Science*, 53:225–241, 1987.

[2] J. Adámek and V. Koubek. On the greatest fixed point of a set functor, *Theoretical Computer Science*, 150:57–75, 1995

[3] M. Alvarez-Manilla, A. Edalat, and N. Saheb-Djahromi. An extension result for continuous valuations. *Journal of the London Mathematical Society*, 61(2):629–640, 2000.

[4] W. Averson. *An Invitation to C*-Algebras.* Springer-Verlag, 1976.

[5] B. Bloom and A. Meyer. Experimenting with process equivalence. *Theoretical Computer Science*, 101:223-237, 1992.

[6] F. van Breugel and J. Worrell. An Algorithm for Quantitative Verification of Probabilistic Transition Systems. In *Proc. 12th International Conference on Concurrency Theory*, volume 2154 of *LNCS*, Springer-Verlag, 2001.

[7] F. van Breugel, S. Shalit and J. Worrell. Testing Labelled Markov Processes. In *Proc. 29th International Colloquium on Automata, Languages and Programming*, volume 2380 of *LNCS*, Springer-Verlag 2002.

[8] F. van Breugel, M. Mislove, J. Ouaknine and J. Worrell. An Intrinsic Characterization of Approximate Probabilistic Bisimilarity. In *Proc. Foundations of Software Science and Computation Structures (FOSSACS 03)*, volume 2620 of *LNCS*, Springer-Verlag, 2003.

[9] J. Desharnais, A. Edalat and P. Panangaden. Bisimulation for Labelled Markov Processes. *Information and Computation*, 179(2):163–193, 2002.

[10] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for Labeled Markov Processes. In *Proc. 10th International Conference on Concurrency Theory*, volume 1664 of *LNCS*, Springer-Verlag, 1999.

[11] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for Labeled Markov Systems. Accepted to *Theoretical Computer Science*, 2004.

[12] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Approximating Labeled Markov Processes. *Information and Computation*, 184(1):160–200, 2003.

[13] A. Edalat. When Scott is weak at the top. *Mathematical Structures in Computer Science*, 7:401–417, 1997.

[14] G.A. Edgar. *Integral, Probability, and Fractal Measures*. Springer-Verlag, 1998.

[15] G. Gierz, K. Hofmann, K. Keimel, J. Lawson, M. Mislove, D. Scott. *Continuous Lattices and Domains*, Cambridge, 2003.

[16] M. Giry. A Categorical Approach to Probability Theory. In *Proc. International Conference on Categorical Aspects of Topology and Analysis*, volume 915 of *Lecture Notes in Mathematics*, Springer-Verlag, 1981.

[17] P. Johnstone. *Stone Spaces*. Cambridge University Press, 1982.

[18] C. Jones. *Probabilistic nondeterminism*, PhD Thesis, Univ. of Edinburgh, 1990.

[19] A. Jung and R. Tix. The Troublesome Probabilistic Powerdomain. In *Third Workshop on Computation and Approximation, Proceedings. Electronic Notes in Theoretical Computer Science*, vol 13, 1998.

[20] K.G. Larsen and A. Skou. Bisimulation through Probabilistic Testing. *Information and Computation*, 94(1):1–28, 1991.

[21] R. Milner. *Communication and Concurrency.* Prentice Hall, 1989.

[22] K.R. Parthasarathy. *Probability Measures on Metric Spaces.* Academic Press, 1967.

[23] M. Smyth and G. Plotkin. The Category Theoretic Solution of Recursive Domain Equations, *SIAM Journal of Computing*, 11(4):761–783, 1982.

[24] E.P. de Vink and J.J.M.M. Rutten. Bisimulation for Probabilistic Transition Systems: a Coalgebraic Approach. *Theoretical Computer Science*, 221(1/2):271-293, June 1999.

A Tests and Observations

Below we recall the testing formalism used by Larsen and Skou [20] to characterize probabilistic bisimilarity on discrete systems. This framework was also used in an earlier version of this paper [8]. We show that the test languages T_0 and T_1, introduced in Definitions 36 and 42 respectively, correspond to two fragments of Larsen and Skou's language.

In fact, the set T_{LS} of tests introduced by Larsen and Skou is almost exactly the same as T_0. The only difference in the syntax is that in T_{LS} tupling plays the role of multiplication. However, rather than considering only that a test may succeed or fail, Larsen and Skou associate to each test $t \in \mathsf{T}_{\mathsf{LS}}$ a set O_t of observations. In this way they account for the fact that some branches of t may succeed while others may fail.

Definition 47 ([20]) *The test language* T_{LS} *is given by the grammar*

$$t ::= 1 \mid at \mid \langle t_1, \ldots, t_n \rangle$$

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

For $t \in \mathsf{T}_{\mathsf{LS}}$ *the set of observations* O_t *is defined by*

$$\begin{aligned} O_1 &= \{1\} \\ O_{at} &= \{a^{\times}\} \cup \{ae \mid e \in O_t\} \\ O_{\langle t_1,\ldots,t_n\rangle} &= O_{t_1} \times \cdots \times O_{t_n} \,. \end{aligned}$$

The only observation of the test 1 is success—which is again denoted 1. An observation of at is either failure of a, denoted $a^{\times}$, or success of a followed by observation $e \in O_t$, denoted ae. An observation of a tuple test $\langle t_1, \ldots, t_n \rangle$ consists of a tuple $\langle e_1, \ldots, e_n \rangle$, where e_i is an observation of t_i.

Thus O_t is a set of mutually exclusive and exhaustive observations that might arise when test t is performed. Given an **LMP** $\langle X, \Sigma, \mu \rangle$, each state $x \in X$ induces a probability distribution $P_t(x, -)$ on O_t according to the following rules.

$$
\begin{aligned}
P_1(x,1) &= 1\\
P_{at}(x,ae) &= \textstyle\int P_t(-,e)d\mu_{x,a}\\
P_{at}(x,a^\times) &= 1-\mu_{x,a}(X)\\
P_{\langle t_1,\ldots,t_n\rangle}(x,\langle e_1,\ldots,e_n\rangle) &= P_{t_1}(x,e_1)\cdots P_{t_n}(x,e_n)
\end{aligned}
$$

Thus $P_t(x,e)$ is the probability of making observation e when test t is run in state x. Given $E \subseteq O_t$ we write $P_t(x,E) = \sum_{e\in E} P_t(x,e)$, i.e., the probability of observing some result in E. Larsen and Skou [20, Theorem 6.5] showed that in a discrete **LMP** satisfying the minimal deviation assumption, two states x and y are bisimilar just in case $P_t(x,E) = P_t(y,E)$ for all tests $t \in \mathsf{T}_{\mathsf{LS}}$ and $E \subseteq O_t$.

Next we show how to intepret the language T_1 in T_{LS}.

Proposition 48 *For each test $t \in \mathsf{T}_1$ there is a test $t' \in \mathsf{T}_{\mathsf{LS}}$ and a set of observations $E \subseteq O_{t'}$ such that, for any **LMP** $\langle X,\Sigma,\mu\rangle$ and $x \in X$, $P(x,t) = P_{t'}(x,E)$.*

PROOF. The proof is by induction on $t \in \mathsf{T}_1$. The base case $\top \in \mathsf{T}_1$ is trivial. Consider now the test at. By induction there exists $t' \in \mathsf{T}_{\mathsf{LS}}$ and $E \subseteq O_{t'}$ such that $P(x,t) = P_{t'}(x,E)$ for all $x \in X$. Then, by linearity of the integral, we get that $P(x,at) = P_{at'}(x,aE)$ for all $x \in X$, where $aE = \{ae \mid e \in E\}$. Finally, suppose $t \equiv \varphi(t_1,\ldots,t_n)$, and, by induction, let $t'_i \in \mathsf{T}_{\mathsf{LS}}$ and $E_i \subseteq O_{t'_i}$ be such that $P(x,t_i) = P_{t'_i}(x,E_i)$ for all $x \in X$. Write $t' \equiv \langle t'_1,\ldots,t'_n\rangle$ and define $E \subseteq O_{t'}$ by

$$E = \{\langle e_1,\ldots,e_n\rangle \mid \varphi_{\mathbb{B}}(e_1 \in E_1,\ldots,e_n \in E_n) = \text{true}\}\,.$$

When test t'_i is run in state x, the probability of making an observation in E_i is $P_{t'_i}(x,E_i)$. We conclude that $P_{t'}(x,E) = \varphi_{\mathbb{R}}(P_{t'_1}(x,E_1),\ldots,P_{t'_n}(x,E_n))$ (cf. Definition 41). Now

$$
\begin{aligned}
P(x,t) &= \varphi_{\mathbb{R}}(P(x,t_1),\ldots,P(x,t_n))\\
&= \varphi_{\mathbb{R}}(P_{t'_1}(x,E_1),\ldots,P_{t'_n}(x,E_n))\\
&= P_{t'}(x,E)\,.
\end{aligned}
$$

Example 49 *Corresponding to the test $a(b\vee b)$ in T_1 is the test $a\langle b,b\rangle$ in T_{LS} with set of observations $E = \{a\langle b,b\rangle, a\langle b^\times,b\rangle, a\langle b,b^\times\rangle\}$.*

Remark 50 *Given a test $t \in \mathsf{T}_0$, the corresponding test $t' \in \mathsf{T}_{\mathsf{LS}}$ is obtained by a trivial syntactic replacement of multiplication by tupling. Futhermore the associated set of observations $E \subseteq O_{t'}$ is just the singleton $\{t'\}$, i.e., the observation that all parts of the test succeed.*