Bisimulation for Labelled Markov Processes

Richard Blute*
Department of Mathematics
University of Ottawa
rblute@mathstat.uottawa.ca

Josée Desharnais*
School of Computer Science
McGill University
desharna@cs.mcgill.ca

Abbas Edalat†
Department of Computing
Imperial College
A.Edalat@doc.ic.ac.uk

Prakash Panangaden*
BRICS, Aarhus University
and School of Computer Science
McGill University
prakash@cs.mcgill.ca

Abstract

In this paper we introduce a new class of labelled transition systems - Labelled Markov Processes - and define bisimulation for them. Labelled Markov processes are probabilistic labelled transition systems where the state space is not necessarily discrete, it could be the reals, for example. We assume that it is a Polish space (the underlying topological space for a complete separable metric space). The mathematical theory of such systems is completely new from the point of view of the extant literature on probabilistic process algebra; of course, it uses classical ideas from measure theory and Markov process theory. The notion of bisimulation builds on the ideas of Larsen and Skou and of Joyal, Nielsen and Winskel. The main result that we prove is that a notion of bisimulation for Markov processes on Polish spaces, which extends the Larsen-Skou definition for discrete systems, is indeed an equivalence relation. This turns out to be a rather hard mathematical result which, as far as we know, embodies a new result in pure probability theory. This work heavily uses continuous mathematics which is becoming an important part of work on hybrid systems.

1. Introduction

Computer Science has been increasingly expanding its borders to include subjects normally considered part of physics, dynamical systems or control theory, most notably in areas like "hybrid systems". Ideas from continuous mathematics are becoming part of the mathematical toolkit of concurrency theorists. Hillston has pioneered the use of a blend of process algebra and Markov process theory in performance evaluation [14]. Systems like Hytech [13] and the Uppaal system [4] have appeared based on hybrid systems and real-time systems respectively and use process equivalences very fruitfully.

The notion of bisimulation is central to the study of concurrent systems. While there are a bewildering variety of different equivalence relations between processes, bisimulation enjoys some fundamental mathematical properties, most notably its characterization as a fixed-point, which make it the most discussed process equivalence. In the present paper we are not so much concerned with adjudicating between the rival claims of all these relations, but rather, we are concerned with showing how to extend these ideas to the world of continuous state spaces. As we shall see below, new mathematical techniques (from the point of view of extant work in process algebra) have to be incorporated to do this. Once the model and the new mathematical ideas have been assimilated, the whole gamut of process equivalences can be studied and argued about.

¿From an immediate practical point of view, bisimulation can be used to reason about probabilistic, continuous state-space systems (henceforth Markov processes) in the following simple way. One often "discretizes" a continuous system by partitioning the state space into a few equivalence classes. Usually one has some intuition that the resulting discrete system "behaves like" the original continuous system. This can be made precise by our notion of bisimulation. It is also the case that some systems cannot be discretized and, once again, one can formalize what this means via bisimulation.

The present paper develops a notion of *labelled Markov process*. The key technical contribution is the development

*Research supported by NSERC
†Research supported by EPSRC.

of a notion of bisimulation for processes which have continuous state spaces but make discrete temporal steps. These are called discrete-time Markov processes. If the state space is also discrete the phrase "Markov chain" is used. The adjective "Markovian" signifies that the transitions are entirely governed by the present state rather than by the past history of the system. The interaction is governed by "labels" in the manner now familiar from process algebra [15, 25, 26].

In brief, a labelled Markov process is as follows. There is a set of states and a set of labels. The system is in a state at a point in time and moves between states. Which state it moves to is governed by which interaction with the environment is taking place and this is indicated by the labels. The system evolves according to a probabilistic law. If the system interacts with the environment by synchronizing on a label, it makes a transition to a new state governed by a transition probability distribution. So far, this is essentially the model developed by Larsen and Skou [23] in their very important and influential work on probabilistic bisimulation. They specify the transitions by giving, for each label, a probability for going from one state to another. Bisimulation then amounts to matching the moves with matching probabilities as well.

In the case of a continuous state space, however, one cannot just specify transition probabilities from one state to another. In most interesting systems all such transition probabilities would be zero! Instead one must work with probability densities. In so doing, one has to confront the major issues that arose when probability theory was first formalized, such as the existence of subsets for which the notion of probability does not make sense. In the present case we have to introduce a notion of sets for which "probabilities make sense" (i.e. a σ-algebra) and instead of talking about probabilities of going from a state s to another state s', we have to talk about going from a state s to a *set of* states A.

The notion of bisimulation for these systems is a generalization of the definition of Larsen and Skou, which is a compelling, natural notion. Unfortunately this definition cannot be adapted in any simple way to the continuous case. Furthermore, once a reasonable generalization is given it turns out to be a formidable technical problem to even show that bisimulation is an equivalence relation. This is solved by a construction due to the third author and can be read in detail from his home page [**?**]. In fact the construction heavily relies on properties that are not true for measure spaces in general. In the present paper we work with Polish space structure. A Polish space is the topological space underlying a complete, separable metric space, i.e. there is a countable basis for the topology. In the classical study of Markov processes metric ideas end up playing a significant role [27]. In any example of physical interest the spaces will have this Polish structure, indeed they will usually come as metric spaces. Any discrete space is Polish and any of the closed subspaces of $\mathbb{R}^n$ will be Polish as well.

The definition of bisimulation is inspired by the paper of Joyal, Nielsen and Winskel [20] which provides a general categorical view of what bisimulation is in terms of certain special morphisms called *open maps*. It is not straightforward to adapt this to the probabilistic case. For discrete systems this has been done by Cheng and Nielsen [6] using infinitesimals[1]. Unfortunately one still needs to know how to construct pullbacks in the underlying category and for this one has to rely on the basic construction given in [**?**]. It also turns out that our notion of bisimulation is precisely the notion of coalgebra homomorphism [31].

2. Two Examples of Processes

We begin with a simple example. Imagine a system with two labels $\{a, b\}$. The state space is the real plane, $\mathbf{R}^2$. When the system makes an a-move from state (x_0, y_0) it jumps to (x, y_0) where the probability distribution for x is given by the density $K_\alpha \exp(-\alpha(x - x_0)^2)$, where $K_\alpha = \sqrt{\alpha/\pi}$ is the normalizing factor. When it makes a b-move it jumps from state (x_0, y_0) to (x_0, y) where the distribution of y is given by the density function $K_\beta \exp(-\beta(y - y_0)^2)$. The meaning of these densities is as follows. The probability of jumping from (x_0, y_0) to a state with x-coordinate in the interval $[s, t]$ under an a-move is $\int_s^t K_\alpha \exp(-\alpha(x - x_0)^2)dx$. Note that the probability of jumping to any given point is, of course, 0. In this system the interaction with the environment controls whether the jump is along the x-axis or along the y-axis but the actual extent of the jump is governed by a probability distribution. Interestingly, this system will turn out to be "bisimilar" to a one-state system which can make a or b moves. Thus, from the point of view of an external observer, this system has an extremely simple behaviour. The more complex internal behaviour is not externally visible. One use of a theory of bisimulation that encompasses such systems is to express this idea. Of course this example is already familiar from the nonprobabilistic setting; if there is a system in which all transitions are always enabled it will be bisimilar (in the traditional sense) to a system with one state.

Now we consider a system which cannot be reduced to a discrete system. There are three labels $\{a, b, c\}$. The state space is $\mathbf{R}$. The state gives the pressure of a gaseous mixture in a tank in a chemical plant. The environment can interact by (a) simply measuring the pressure, or (b) it can inject some gas into the tank, (c) or it can pump some gas from the tank. The pressure fluctuates according to some thermodynamic laws depending on the reactions taking place in the tank. With each interaction, the pressure changes according to three different probability density functions, say

[1] One can actually do this without talking about infinitesimals [36].

$f(p_0,p), g(p_0,p)$ and $h(p_0,p)$ respectively, with nontrivial dependence on p_0. There are in addition two threshold values p_h and p_l. When the pressure rises above p_h the interaction labelled b is disabled, and when the pressure drops below p_l the interaction labelled c is disabled. It is tempting to model this as a three state system, with the continuous state space partitioned by the threshold values. Unfortunately one cannot assign unique transition probabilities to these sets of states for any choices of f, g and h; only if very special uniformity conditions are obeyed can one do this.

3. Discrete Probabilistic Systems

In this section we recapitulate the Larsen-Skou definition of probabilistic bisimulation [23]. The systems that they consider will be referred to as labelled Markov chains in the present paper.

Definition 3.1 *A* **labelled Markov chain** *is a quadruple* $(S,\mathcal{L},C_l,P_l)$*, where* S *is a countable set of states,* $\mathcal{L}$ *is a set of labels, and for each* $l \in \mathcal{L}$ *we have a subset* C_l *of* S *and a function,* P_l*, called a* **transition probability matrix***,*

$$P_l : C_l \times S \longrightarrow [0,1]$$

satisfying the normalization condition

$$\forall l \in \mathcal{L}, s \in C_l.\Sigma_{s' \in S} P_l(s,s') = 1.$$

If we have the weaker property

$$\forall l \in \mathcal{L}, s \in C_l.\Sigma_{s' \in S} P_l(s,s') \leq 1$$

we call the system a **partial** *labelled Markov chain.*

The sets C_l are the sets of states that *can* do an l-action. If we have partial labelled Markov chains, we can just dispense with the C_l sets. In what follows we suppress the label set, i.e. we assume a fixed label set given once and for all.

Definition 3.2 *Let* $T = (S, P_l)$ *be a labelled Markov chain. Then a* **probabilistic bisimulation** $\equiv_p$*, is an equivalence on* S *such that, whenever* $s \equiv_p t$*, the following holds:*

$$\forall l \in \mathcal{L}.\forall A \in S/\equiv_p,\ \Sigma_{s' \in A} P_l(s,s') = \Sigma_{s' \in A} P_l(t,s').$$

Two states s *and* t *are said to be* **probabilistically bisimilar** *(*$s \sim_{LS} t$*) in case* (s,t) *is contained in some probabilistic bisimulation.*

Intuitively we can read this as saying that two states are bisimilar if we get the same probability when we add up the transition probabilities to all the states in an equivalence class of bisimilar states. The adding up is crucial – the probabilities are not just another label. The subtlety in the definition is that one has to somehow know what states are probabilistically bisimilar in order to know what the equivalence classes are, which in turn one needs in order to compute the probabilities to match them appropriately. In fact a very natural notion of probabilistic synchronization trees yields a model of a probabilistic version of CCS with both probabilistic branching and nondeterministic branching. If one looks just at probabilistic branching, then equality is precisely the Larsen-Skou notion of bisimulation [3].

The paper by Larsen and Skou does much more than just define bisimulation. They introduce the notion of testing a probabilistic process and associating probabilities with the possible outcomes. They then introduce a notion of testable properties. The link with probabilistic bisimulation is that two processes are probabilistically bisimilar precisely when they agree with the results of all tests. They also introduce a probabilistic modal logic and show that bisimulation holds exactly when two processes satisfy the same formulas.

4. A Category of Markov Processes

A Markov process is a transition system with the property that the transition probabilities depend only on the current state and not on the past history of the process. We will consider systems where there is an interaction with the environment described by a set of labels as in process algebra. For each fixed label the system may undergo a transition governed by a transition probability. One could have a new set of possible states at every instant but, for simplicity, we restrict to a single state space. We will organize the theory in categorical terms with objects being transition systems and morphisms being simulations; this will allow one to compare the theory with the more traditional theory of non-probabilistic processes, see, for example, the handbook article by Winskel and Nielsen [37].

In formulating the notion of Markov processes we need to refine two concepts that were used in the discrete case. First, we cannot just define transition probabilities between states; except in rare cases such transition probabilities are zero. We have to define transition probabilities between a state and a set of states. Second, we cannot define transition probabilities to any arbitrary set of states; we need to identify a family of sets for which transition probabilities can be sensibly defined. These are the *measurable sets*. Thus in addition to specifying a set of states we need to specify a σ-algebra on the set of states [2, **?**].

Definition 4.1 *A* transition probability function *on a measurable space* (X,Σ) *is a function* $T : X \times \Sigma \longrightarrow [0,1]$ *such that for each fixed* $x \in X$*, the set function* $T(x,\cdot)$ *is a (sub)probability measure and for each fixed* $A \in \Sigma$ *the function* $T(\cdot,A)$ *is a measurable function.*

One interprets $T(x, A)$ as the probability of the system starting in state x making a transition into one of the states in A.

It will turn out to be convenient to work with *sub-probability* functions; i.e. with functions where $T(x, X) \leq 1$ rather than $T(x, X) = 1$. In fact what is often done is that a state x with no possibility of making a transition is modelled by having a transition back to itself. For questions concerning which states will eventually be reached (the bulk of the analysis in the traditional literature) this is convenient. If, however, we are modelling the interactions that the system has with its environment, it is essential that we make a distinction between a state which can make a transition and one which cannot. The key mathematical construction uses Polish-space structure on the set of states. Thus instead of imposing an arbitrary σ-algebra structure on the set of states we will require that the set of states be a Polish space and the σ-algebra be the Borel algebra generated by the topology.

Definition 4.2 *A* **partial, labelled, Markov process** *with label set $\mathcal{L}$ is a structure $(S, \Sigma, \{k_l \mid l \in \mathcal{L}\})$, where S is the set of states, which is assumed to be a Polish space, and Σ is the Borel σ-algebra on S, and*

$$\forall l \in \mathcal{L}, k_l : S \times \Sigma \longrightarrow [0, 1]$$

is a transition sub-probability function. We are usually interested in the following special case called a **labelled Markov process**. *We have a partial, labelled, Markov process as above and a predicate* **Can** *on $S \times \mathcal{L}$ such that for every $(x, l) \in \mathbf{Can}$ we have $k_l(x, X) = 1$ and for every $(x, l) \notin \mathbf{Can}$ we have $k_l(x, X) = 0$.*

We will fix the label set to be some $\mathcal{L}$ once and for all. The resulting theory is not seriously restricted by this. We will write just (S, Σ, k_l) for partial, labelled, Markov processes instead of the more precise $(S, \Sigma, \forall l \in \mathcal{L}.k_l)$. In case we are talking about discrete systems we will use the phrase "labelled Markov chain" rather than "discrete, labelled, Markov process".

In a (partial), labelled, Markov *chain* the set of states is countable, the σ-algebra is the entire powerset and the transition probabilities are given by a $\mathcal{L}$-indexed family of functions $\forall l \in \mathcal{L}.P_l : S \times S \longrightarrow [0, 1]$ satisfying the conditions required of a (sub)probability distribution. From this presentation we can construct the k_l in the evident way. We use the phrase "transition function" for an object of type $S \times \Sigma \longrightarrow [0, 1]$ and "transition matrix" for an object of type $S \times S \longrightarrow [0, 1]$. A probabilistic transition system as defined by Larsen and Skou is precisely a labelled Markov chain.

We define simulation morphisms between processes. Intuitively a simulation says that a simulating process can make all the transitions of the simulated process with greater probability than in the process being simulated.

Definition 4.3 *A* **simulation morphism** *f between two partial, labelled, Markov processes, (S, Σ, k_l) and (S', Σ', k'_l) is a measurable function $f : (S, \Sigma) \longrightarrow (S', \Sigma')$ such that*

$$\forall l \in \mathcal{L}.\forall x \in S.\forall \sigma' \in \Sigma'.k_l(x, f^{-1}(\sigma')) \leq k'_l(f(x), \sigma').$$

We could have defined simulations to be continuous but in this paper general simulations are very secondary so we do not need them to be continuous. Without requiring at least measurability the definition would not make sense since $f^{-1}(\sigma')$ would not necessarily be measurable[2] For discrete systems, this notion extends the standard notion of simulation of labelled transition systems [37] in a straightforward way.

5. Bisimulation for Markov Processes

The definition of bisimulation is very heavily influenced by the ideas of Joyal, Nielsen and Winskel [20]. The idea is to identify a class of special systems called "observations" or "observable paths" or better still "observable path shapes", and to define bisimulation as a relation satisfying a kind of path lifting property, the so-called "open maps". We will not follow this programme in detail. We have not, in fact, tried very hard to give a path-lifting account of Larsen-Skou bisimulation in the continuous case, since until now we have focused on the technical problem of proving that bisimulation is indeed an equivalence.

For ordinary labelled transition systems, if we take paths to be labelled transition sequences, then the open maps are so-called "zigzag" morphisms. We essentially want to say that there is a "zigzag relation". One can talk about relations by talking about **spans**. A span in any category between an object S_1 and another object S_2 is a third object T together with morphisms from T to both S_1 and S_2. One can think of this in the category of **Sets** as viewing a relation as a set of ordered pairs with the morphisms being the projections. Bisimulation is defined to hold between two systems if they are connected by a span of zigzags.

In our case the zigzag condition is easy to state and it is easy to see that it corresponds to Larsen-Skou bisimulation in the case of labelled Markov chains. For partial, labelled Markov processes it will be our definition of bisimulation. From now on we will assume all systems to be partial and not add the adjective "partial" explicitly.

[2] Measurable means that the inverse image of measurable sets is measurable. Older books, e.g Halmos or Rudin [12, 29] use a different notion of measurable according to which measurability is not compositional. Most modern books [**?**, 24] follow our conventions.

Definition 5.1 *The objects of the category* **LMP** *are labelled Markov processes, having $\mathcal{L}$ as set of labels, with simulations as the morphisms. The category of labelled, Markov chains is written* **LMC** *and is the full subcategory of* **LMP** *that includes only the labelled, Markov chains as objects.*

The key concept is the following.

Definition 5.2 *A morphism f from (S, Σ, k_l) to (S', Σ', k'_l) is* a zigzag morphism *if it satisfies the properties:*

1. *f is surjective;*
2. *f is continuous*
3. *$\forall l \in \mathcal{L}, s \in S, \sigma' \in \Sigma', \quad k_l(s, f^{-1}(\sigma')) = k'_l(f(s), \sigma')$.*

Asking f to be surjective allows us to avoid introducing initial states and worrying about reachable states. Note that we are now taking the topological structure seriously and requiring zigzag morphisms to be continuous. One can immediately check that the identity morphism is a zigzag.

We often refer to a labelled Markov process by its set of states. Following Joyal, Nielsen and Winskel ([20]) we define bisimulation as the existence of a span of zigzag morphisms.

Definition 5.3 *Let S and S' be two labelled Markov processes. S is* probabilistically bisimilar to *S' (written $S \sim S'$) if there is a span of zigzag morphisms between them, i.e. there exists a labelled Markov process U and zigzag morphisms $f : U \longrightarrow S$ and $g : U \longrightarrow S'$.*

Notice that if there is a zigzag morphism between two systems, they are bisimilar since the identity is a zigzag morphism.

It is interesting to note that we can take a coalgebraic view of bisimulation [1, 30, 31] as well. We can view a labelled Markov process as a coalgebra of a suitable functor; in fact it is a functor introduced by Giry [11] in order to define a monad on **Mes** analogous to the powerset monad. From this point of view, bisimulation is a span of coalgebra homomorphisms. But if one checks what this means it turns out that these are precisely zigzag morphisms.

The following proposition shows that we are extending the Larsen-Skou definition. One needs to interpret this appropriately since the Larsen-Skou definition applies to the states of a single system rather than between the states of two different systems. We have to therefore define a correspondence between the two notions. The full paper makes this precise.

Proposition 5.4 *Two discrete systems are bisimilar in the sense of Larsen and Skou iff there is a span of zigzags between them.*

We want bisimulation to be an equivalence, so we need to prove transitivity of the existence of span, since it is obviously reflexive and symmetric. Proving transitivity presents formidable difficulties. In particular, bisimulation probably isn't transitive for probabilistic transition systems without some assumption like the assumption of Polish structure on the set of states. The proof works only for Polish spaces [3]. The following theorem immediately implies that bisimulation is an equivalence relation in our setting.

Theorem 5.5 *[**?**] In the category where the objects are labelled Markov processes and the morphisms are zigzags, pullbacks exist.*

Corollary 5.6 *Bisimulation defined as above (i.e. with Polish space structure on the sets of states and with zigzag morphisms assumed to be continuous) on labelled Markov processes is an equivalence relation.*

We have three transition systems, we refer to them briefly as S_1 and S_2 and S and two zigzags $f : S_1 \longrightarrow S$ and $g : S_2 \longrightarrow S$. We want to construct another system U with zigzags to S_1, S_2 which makes the square commute. We have no choice but to define the set of states of U as the subset of $S_1 \times S_2$ where f and g agree. Then the task is to define a suitable transition probability h which makes the projections onto S_1, S_2 into zigzags. In the full paper we explain why naive approaches fail [5]. The actual construction uses two key ideas. The first is to forget about U for the moment and define a transition probability on $S_1 \times S_2$ but instead of just taking the product system we condition the probabilities on the transitions agreeing when their images are looked at in S. Now we can define the transition probability on "rectangles" and extend to all of the space. The conditioning automatically makes the measure live just on U rather than on all of $S_1 \times S_2$.

Proof outlined for theorem 5.5. It suffices to consider just a single label so we suppress the labels for this proof. In this proof when we say "object" we mean labelled Markov process and when we say "morphism" we mean a zigzag morphism, which means that the function is continuous. Three objects

$$S_1 = (S_1, \Sigma_1, k_1 : S_1 \times \Sigma_1 \to [0, 1]),$$

$$S_2 = (S_2, \Sigma_2, k_2 : S_2 \times \Sigma_2 \to [0, 1]),$$

$$S = (S, \Sigma, k : S \times \Sigma \to [0, 1])$$

and morphisms $f_1 : S_1 \to S$ and $f_2 : S_2 \to S$ are given.

[3] We have recently extended this to analytic spaces

Let $U = \{(s_1, s_2) \in S_1 \times S_2 \mid f_1(s_1) = f_2(s_2)\}$ equipped with the subspace topology of the product topology on $S_1 \times S_2$. The Borel σ-algebra Σ_U on U is generated by the set $\{(\sigma_1 \times \sigma_2) \cap U \mid \sigma_1 \in \Sigma_1, \sigma_2 \in \Sigma_2\}$. Let $\pi_1 : U \to S_1$ and $\pi_2 : U \to S_2$ be the projection maps. Since f_1 and f_2 are surjective, U is not empty. We want to construct $h : U \times \Sigma_U \to [0,1]$ so that (U, Σ_U, h) is an object and $\pi_1 : U \to S_1$ and $\pi_2 : U \to S_2$ are morphisms. For an element x of a set X, an element y of a set Y, subsets $A \subseteq X$, $B \subseteq Y$ and a function $f : X \to Y$, we sometimes write x instead of $\{x\}$, fx instead of $f(x)$, fA instead of $f(A)$, and $f^{-1}B$ instead of $f^{-1}(B)$.

We fix $(s_1, s_2) \in U$, i.e. $s_1 \in S_1$ and $s_2 \in S_2$ with $f_1 s_1 = f_2 s_2$, throughout this proof. The index i always takes the values 1 and 2. The variable ω_i always runs through S_i whereas the variable s runs through S.

For $i = 1, 2$, we have the probability measures $k_i(s_i, -) : \Sigma_i \to [0,1]$ on the space (S_i, Σ_i). Also $\Sigma_i' = \{f_i^{-1}\sigma \mid \sigma \in \Sigma\} \subseteq \Sigma_i$ is a sub-σ-algebra of Σ_i ($i = 1, 2$). We therefore have, for a given $\sigma_i \in \Sigma_i$, the conditional probability distribution $P((s_i, \sigma_i)|\Sigma_i') : S_i \to [0,1]$ of the probability measure $k_i(s_i, -)$ given the sub-σ-algebra $\Sigma_i' \subseteq \Sigma_i$. Since S_i is a Polish space we can assume [9, Theorem 10.2.2] that $P((s_i, \sigma_i)|\Sigma_i')(-)$ is a *regular* conditional probability distribution, i.e.

(i) $P((s_i, \sigma_i)|\Sigma_i') : S_i \to [0,1]$ is Σ_i' measurable and integrable.

(ii) For all $\gamma_i' \in \Sigma_i'$, we have

$$\int_{\gamma_i'} P((s_i, \sigma_i)|\Sigma_i')(\omega_i) k_i(s_i, d\omega_i) = k_i(s_i, \sigma_i \cap \gamma_i').$$

(iii) For $k_i(s_i, -)$-almost all $\omega_i \in S_i$, $P((s_i, -)|\Sigma_i')(\omega_i) : \Sigma_i \to [0,1]$ is a probability measure on S_i.

A regular conditional probability distribution is unique up to a set of measure zero, i.e. any two functions satisfying the above three properties are equal for $k_i(s_i, -)$-almost all $\omega_i \in S_i$.

One should think of $P((s_i, \sigma_i)|\Sigma_i')(\omega_i)$ as the probability that s_i makes a transition to σ_i given that s_i makes a transition to $f_i^{-1} f_i \omega_i$, or equivalently as the probability that s_i makes a transition to σ_i given that $f_i s_i$ makes a transition to $f_i \omega_i$.

By a standard method we can obtain the conditional probability distribution

$$P((s_i, \sigma_i)|\Sigma) : S \to [0,1].$$

Here, $P((s_i, \sigma_i)|\Sigma)(s)$ gives the probability that s_i makes a transition to σ_i given that s_i makes a transition to some ω_i with $f_i(\omega_i) = s$ or equivalently the probability that s_i makes a transition to σ_i given that $f_i s_i$ makes a transition to s.

In order to define $h((s_1, s_2), -) : \Sigma_U \to [0,1]$ we first define a probability measure $g((s_1, s_2), -) : \Sigma_{S_1 \times S_2} \to [0,1]$ on the product space $(S_1 \times S_2, \Sigma_{S_1 \times S_2})$ where $\Sigma_{S_1 \times S_2}$ is the Borel σ-algebra of $S_1 \times S_2$. We do this by taking the product of the two regular conditional probability densities and integrating with respect to k. By using the conditional probability densities we are ensuring that the transitions are being paired together only if their images agree. It turns out that $g((s_1, s_2), -)$ is supported on U, i.e. $g((s_1, s_2), U^c) = 0$. We will finally define $h((s_1, s_2), \alpha \cap U) = g((s_1, s_2), \alpha)$. This proof used some subtle facts about Borel sets and continuous functions from Polish spaces to Hausdorff spaces [10].

We can now complete the proof by checking the following statement. $(U, \Sigma_U, h : U \times \Sigma_U \to [0,1])$ is an object of the category and $\pi_i : U \to S_i$ are morphisms. ∎

Example 5.7 The first example is a reexamination of the first example from section 2. We let the label set be the one element set. Consider a system (S, Σ, k) with S an arbitrarily complicated state space with a Polish space structure and Σ the Borel σ-algebra on S. We define the $k(s, \sigma)$ in any manner we please, consistent with the definition and subject only to the condition that $\forall s \in S. k(s, S) = 1$. Consider the trivial labelled Markov chain with just one label, one state and one transition from the state to itself with probability 1. These two systems are bisimilar!

All of conventional stochastic process theory is described by systems like the first system above. From our point of view they are trivial. This is to be expected, as we are modelling *interaction* and all such systems are indeed trivial from the point of view of interaction; they can always make an a-step with probability 1.

Example 5.8 Consider the labelled Markov process, over the trivial label set, defined as follows $S = (\mathbf{R}, \mathcal{B}, k)$, i.e. the states are real numbers, the measurable sets are the Borel sets and the transition function is defined on intervals (and then extended to arbitrary Borel sets) as follows:

$$k(x, [r, s]) = \begin{cases} \lambda/2 \int_r^s e^{-\lambda|x-y|} dy & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where the constant $\lambda/2$ is chosen to make k be 1 on the whole space. Intuitively this is a system where a particle makes random jumps with probability exponentially distributed with the length. However, there is an "absorbing wall" at the point $x = 0$ so that if the system jumps to the left of this point it gets stuck there. Now consider the system $U = (\mathbf{R}^2, \mathcal{B}^2, h)$ defined as

$$h((x, y), [r, s] \times [p, q]) = k(x, [r, s]) P([p, q]),$$

where P is some arbitrary probability measure over $\mathbf{R}$. This system should behave "observably" just like the first system because, roughly speaking, the first coordinate behaves just like the first system and the second system has trivial dynamics, i.e. it is bisimilar to the one-state, one-transition system. Indeed these two systems are bisimilar with the projection from the second to the first being a zigzag.

The next example illustrates a possible objection to our definition.

Example 5.9 Suppose that we have two systems, $(\mathbf{R}, \mathcal{B}, \{a, b\}, k_l)$ and $(\mathbf{R}, \mathcal{B}, \{a, b\}, h_l)$. In the first system we have the following transitions

$$k_a(x, S) = \mu(S \cap [x - 0.5, x + 0.5])$$

and

$$k_b(x, S) = \mu(S \cap [x + 0.5, x + 1.5])$$

where S is a Borel set and μ is Lebesgue measure. For the other system we have

$$h_a(x, S) = \begin{cases} \mu(S \cap [x - 0.5, x + 0.5]) & \text{if } x \text{ is irrational} \\ 0 & \text{if } x \text{ is rational.} \end{cases}$$

the b transitions are the same as those for the first system. These two systems are not bisimilar by our definition. The first one is bisimilar to the trivial one-state system with both a and b enabled all the time while the second one is has states in which a gets disabled. However, the probability of landing in one of these states is 0. Thus, in some sense, the difference is visible only on a set of probability 0. Should they be distinguished?

6. Towards A Modal Logic for Bisimulation

In this section we begin the study of logics that characterize bisimulation. For discrete systems, indeed even if only one system is discrete, we can define a simple logic which plays the role of Hennessy-Milner logic. In practice one often wants to know that a continuous system can be discretized, i.e. is bisimilar to some discrete system, and this fact will be useful to determine that. For two continuous systems the situations appear to involve subtle problems of topology and measure theory.

We follow the treatment of Larsen and Skou [23] closely in terms of the definition of the logic but not in terms of proofs. The key difference is that we use negation. Showing that a logic characterizes bisimulation for general Polish systems appears to involve nontrivial measure-theoretic technicalities and is the subject of future work.

We take as the syntax the following formulas:

$$T | \neg\phi | \phi_1 \wedge \phi_2 | \langle a \rangle_q \phi$$

where a is an action from the fixed set of actions $\mathcal{L}$ and q is a rational number. Given a labelled Markov process (S, Σ, k_a) we write $s \models \phi$ to mean that the state s satisfies the formula ϕ. The definition of the relation $\models$ is given by induction on formulas. The definition is obvious for the propositional constants and connectives. We say $s \models \langle a \rangle_q \phi$ iff $\exists A \in \Sigma. \forall s' \in A. s' \models \phi \wedge k_a(s, A) \geq q$. In other words, s can make an a-move to a state that satisfies ϕ with probability greater than q. We write $[\![\phi]\!]$ for the set $\{s \in S | s \models \phi\}$. If we have two systems, say S, S', in mind and we want to distinguish them we write $[\![\phi]\!]_S$.

The first proposition below says that sets of states definable by formulas are always measurable.

Proposition 6.1 *For all formulas ϕ, we have $[\![\phi]\!] \in \Sigma$.*

The next proposition links zigzag morphisms with formulas in the logic.

Proposition 6.2 *If f is a zigzag morphism from S to S' then for all formulas ϕ,*

$$s \models \phi \iff f(s) \models \phi.$$

¿From this we get a simple corollary, but we need to define bisimilarity of states first. We say $s \in S$ and $s' \in S'$ are bisimilar if S and S' are bisimilar and for *some* span of morphisms $f : T \longrightarrow S$ and $f' : T \longrightarrow S'$ there exists $t \in T$ with $f(t) = s$ and $f'(t) = s'$. It can be shown, using existence of pullbacks, that bisimilarity of states is an equivalence relation.

Corollary 6.3 *If s and s' are bisimilar states of S and S' respectively then they satisfy the same formulas.*

Now one can show that the transition probabilities to definable sets are determined completely by the formulas, independently of the system. Suppose that (S, Σ, k_a) and (S', Σ', k'_a) are two systems. We say that the two systems satisfy *all the same formulas* if $\forall s \in S \exists s' \in S'$ such that s and s' satisfy all the same formulas and the same with s and s' interchanged. We write $S \bowtie S'$ if this is the case.

Proposition 6.4 *Suppose that $(S, \Sigma, k_a) \bowtie (S', \Sigma', k'_a)$ then for all formulas ϕ and all pairs (s, s') such that s and s' satisfy all the same formulas, we have $k_a(s, [\![\phi]\!]_S) = k'_a(s', [\![\phi]\!]_{S'})$.*

Now we would like to show that if two systems satisfy all the same formulas they must be bisimilar. Instead of defining a span of zigzags directly, we can define a cospan and use the pullback property to infer that a span must exist. Given two systems $(S, \Sigma, k_a) \bowtie (S', \Sigma', k'_a)$ we first construct a system, (T, Σ_T, j_a), called the *direct sum* of S and S', as follows. We set $T = S \uplus S'$ with the evident σ-algebra. We define the transition probabilities as follows: $j_a(s, A \uplus A') = k_a(s, A)$ if $s \in S$ and

$j_a(s', A \uplus A') = k'_a(s', A')$ if $s' \in S'$ where $A \in \Sigma$ and $A' \in \Sigma'$. There are the evident canonical injections ι, ι' which are *not* zigzags because they are not surjective. Now we define an equivalence relation on the states of T by saying that two states are equivalent iff they satisfy all the same formulas. We write $s \approx s'$. We define a quotient system (V, Σ_V, h_a) as follows. The states of V are the equivalence classes and the topology is the greatest one making the canonical surjection $r : T \longrightarrow V$ continuous. Because T has to be discrete it is a Polish space. The composites $\iota \circ r$ and $\iota' \circ r$ are measurable and surjective, henceforth we call them q and q' respectively. To see that q, q' are surjective we recall that an equivalence class must, by hypothesis, include members of both S and S'. If we can define h_a so as to make q, q' both zigzag we will be done.

Proposition 6.5 *If we have $(S, \Sigma, k_a) \bowtie (S', \Sigma', k'_a)$ and $s \approx s'$ then for all $t \in T$, defined as above, we have $k_a(s, [t] \cap S) = k'_a(s', [t] \cap S')$.*

Now we can define h_a as follows. We only need to specify the point to point transition probabilities since, in the discrete case, these determine all the transition probabilities. We set $h_a([s], [t]) \stackrel{\text{def}}{=} k_a(s, q^{-1}([t]))$. Clearly the representative of the $[s]$ equivalence class does not matter. Furthermore by proposition 6.5, we can see that $h_a([s], [t]) = k'_a(s', {q'}^{-1}([t]))$ and the morphisms are now zigzags. Thus we have proved the following result.

Proposition 6.6 *If we have two systems and one of them is discrete then they are bisimilar iff they obey the same formulas.*

Note that this implies the result for discrete systems without using the minimum deviation assumption used by Larsen and Skou.

7. Related Work

There has been a substantial amount of work on probabilistic transition systems and their associated equivalences. As far as we are aware, none of them have looked at bisimulation for continuous state spaces.

The starting point of work in the area of probabilistic semantics are the fundamental papers of Saheb-Djahromi [32, 33] and of Kozen [21, 22]. These are concerned with domain theory and programming languages rather than with process equivalences but they both introduced nontrivial measure-theoretic ideas. Kozen also noticed a very interesting duality between state-transformer semantics as described by stochastic kernels and a probabilistic predicate-transformer semantics in which programs are seen as inducing linear continuous maps on the Banach algebra of bounded measurable functions.

The one paper with an abstract categorical approach to stochastic processes is by Giry [11] but she studies categorical constructions rather than process equivalences. In particular she shows that the stochastic kernels (conditional probability distributions) that we use to define transition probabilities arise as the Kleisli category of a monad, which is a natural generalization of the powerset monad to the probabilistic case. If we recall that the category **Rel** of sets and relations is the Kleisli category of the powerset monad, we see that the stochastic kernels can reasonably be viewed as the probabilistic analogues of relations. This makes the analogy between labelled Markov processes and ordinary transitions quite striking.

Our investigations began with a study of the pioneering paper of Larsen and Skou [23] which gave the first compelling analysis of probabilistic bisimulation. The work of Joyal, Nielsen and Winskel [20] and of Cheng and Nielsen [6] provided vital clues. Following Joyal et. al. we define probabilistic bisimulation as spans of zigzags.

¿From the point of view of applications there have been a number of very interesting results. The most interesting work, in our opinion, is the work of Jane Hillston [14] on developing a process algebra for performance evaluation. Her work is not comparable to ours at present because she works with continuous-time Markov chains, i.e. with continuous time and a discrete state space. Nevertheless it is our hope that as our own work evolves we will make contact with her approach.

The group at Oxford has been developing an extensive theory of probabilistic systems; see the collection of reports available from the web [28]. The focus has been on equational laws satisfied by processes. From the semantic point of view they have extensively developed and enriched Kozen's [22] predicate-transformer view. They have looked at continuous state-space systems and have incorporated nondeterminism in their framework.

Finally we mention a sample of the large amount of work on various notions of testing [8, 7, 18], simulations [34], specification [17], model checking [16], generative processes [35, 19] and other variations which, for example, allow one to have both probability and nondeterminism. It should be possible to develop analogous notions for a variety of continuous state-space systems once the mathematical foundations are in place. With the present work the mathematical apparatus needed is in place though of course a significant amount of work remains to determine which models and equivalences are useful in practice.

8. Conclusions

In this paper we have two main contributions. First, we develop a notion of labelled Markov process as a model of interacting Markov processes. The state space may have complex structure but the interactions are simple synchronizations. Unlike the classical theory of Markov processes, which asks about the evolution of the system in state space, we study behavioural equivalences from the point of view of observing interactions. This kind of system is exactly what is studied in much of hybrid systems – simple finite-state controllers interacting with a complex dynamical system. Of course the stochastic versions of such systems have not been studied very much as yet.

Secondly we develop a notion of bisimulation for probabilistic processes which applies to systems where the state space can be continuous and which extends the definition of Larsen and Skou [23] for the discrete case. The essential limitation, met in all examples known in physics, control theory or computing, is that the state space be equipped with the structure of a Polish space. The fundamental technical contribution is the proof that this indeed is an equivalence relation. This required a hard proof which can be seen as a theorem of Markov processes in its own right. In particular no previous study of probabilistic bisimulation has had to deal with these technical questions. We expect that future studies of such systems will use either this technique or ones very like it. A very preliminary result is that we have a logical characterization of bisimulation which works even when one of the systems is not discrete and does not use the minimum deviation assumption. This mildly extends the work of Larsen and Skou.

The important questions to answer for the future are:

1. The notion of bisimulation in the probabilistic case can be, and should be, debated. For example suppose there are two systems which have different behaviours but these differences are only visible with probability 0, do we want to say that they are or are not bisimilar. In the present paper we would say that they are not bisimilar.

2. What is the general logical characterization of bisimulation? We might have to adopt the point of view that formulas are measurable functions and that the satisfaction "relation" has to be generalized to mean integration. Kozen has already used such ideas [22] in his study of probabilistic PDL.

3. What is the right notion of testing and can we exhibit bisimulation as a testing equivalence? The bisimulation relation that we introduce in this paper is, like Larsen and Skou's, sensitive to small changes in the probabilities. This makes the development of the appropriate testing notions important in order to discuss what one might plausibly test for in a continuous state space setting. In particular it should allow us to initiate a serious study of stability of systems.

4. What is the right notion of simulation? What will happen to simulation and bisimulation in the presence of nondeterminism? These questions have been studied extensively in the discrete case.

5. What are the right calculi for such systems? This is completely wide open at this point.

6. Can we extend these ideas to continuous time? This certainly should be possible and really bring this work into contact with extant work on hybrid systems.

Acknowledgements Abbas Edalat would like to thank Terry Lyons for helpful discussions and EPSRC(UK) for support. Prakash Panangaden would like to thank Samson Abramsky, Luca Cattani, Jørgen Hoffman-Jørgensen, Alan Jeffrey, Jan Rutten, Ian Stark, Colin Stirling and Glynn Winskel for helpful discussions and BRICS for its hospitality during the course of this work. Richard Blute, Josée Desharnais and Prakash Panangaden thank NSERC (Canada) for its support. Richard Blute and Prakash Panangaden acknowledge the support of an Ontario-Québec Bilateral Exchange grant.

References

[1] P. Aczel and N. Mendler. A final-coalgebra theorem. In *Category Theory and Computer Science*, Lecture Notes In Computer Science, pages 357–365, 1989.

[2] R. B. Ash. *Real Analysis and Probability*. Academic Press, 1972.

[3] C. Baier and M. Kwiatkowska. Domain equations for probabilistic processes. Available from URL http://www.cs.bham.ac.uk/ mzk/, March 1996.

[4] J. Bengtsson, K. G. Larsen, F. Larsson, P. Pettersson, and W. Yi. Uppaal: A tool suite for automatic verification of real-time systems. In R. Alur, T. Henzinger, and E. Sontag, editors, *Hybrid Systems III*, number 1066 in Lecture Notes In Computer Science, pages 232–243, 1996.

[5] R. Blute, J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labelled markov processes. Forthcoming BRICS report, available from `www.daimi.aau.dk/~prakash/pubs.html`, 1996.

[6] A. Cheng and M. Nielsen. Open maps at work. In P. S. Thiagarajan, editor, *Proceedings of the 15th Annual Conference on Foundations of Software Technology and Theoretical Computer Science*, volume 1026 of *Lecture Notes In Computer Science*. Springer-Verlag, 1995.

[7] I. Christoff. Testing equivalences and fully abstract models for probabilistic processes. In *CONCUR 90*, number 458 in Lecture Notes In Computer Science. Springer-Verlag, 1990.

[8] R. Cleaveland, S. Smolka, and A. Zwarico. Testing preorders for probabilistic processes. In *Proceedings of the International Colloquium On Automata Languages And Programming 1992*, number 623 in Lecture Notes In Computer Science. Springer-Verlag, 1992.

[9] R. M. Dudley. *Real Analysis and Probability*. Wadsworth and Brookes/Cole, 1989.

[10] H. Federer. *Geometric Measure Theory*. Springer-Verlag, 1969.

[11] M. Giry. A categorical approach to probability theory. In B. Banaschewski, editor, *Categorical Aspects of Topology and Analysis*, number 915 in Lecture Notes In Mathematics, pages 68–85. Springer-Verlag, 1981.

[12] P. Halmos. *Measure Theory*. Number 18 in Graduate Texts in Mathematics. Springer-Verlag, 1974. Originally published in 1950.

[13] T. Henzinger, P.-H. Ho, and H. Wong-Toi. Hytech: the next generation. In *Proceedings of the 16th Annual Real-time Systems Symposium*, pages 55–65. IEEE Computer Society Press, 1995.

[14] J. Hillston. *A Compositional Approach to Performance Modelling*. PhD thesis, University of Edinburgh, 1994. To be published as a Distinguished Dissertation by Cambridge University Press.

[15] C. A. R. Hoare. *Communicating Sequential Processes*. Series in Computer Science. Prentice-Hall International, London, 1985.

[16] M. Huth and M. Kwiatkowska. On probabilistic model checking. Technical Report CSR-96-15, University of Birmingham, 1996. available from http://www.cs.bham.ac.uk/ mzk/.

[17] B. Jonsson and K. Larsen. Specification and refinement of probabilistic processes. In *Proceedings of the 6th Annual IEEE Symposium On Logic In Computer Science*, 1991.

[18] B. Jonsson and W. Yi. Compositional testing preorders for probabilistic processes. In *Proceedings of the 10th Annual IEEE Symposium On Logic In Computer Science*, pages 431–441, 1995.

[19] C. Jou and S. Smolka. Equivalences, congruences and complete axiomatizations for probabilistic processes. In *CONCUR 90*, number 458 in Lecture Notes In Computer Science, 1990.

[20] A. Joyal, M. Nielsen, and G. Winskel. Bisimulation from open maps. *Information and Computation*, 127(2):164–185, 1996.

[21] D. Kozen. Semantics of probabilistic programs. *Journal of Computer and Systems Sciences*, 22:328–350, 1981.

[22] D. Kozen. A probabilistic PDL. *Journal of Computer and Systems Sciences*, 30(2):162–178, 1985.

[23] K. G. Larsen and A. Skou. Bisimulation through probablistic testing. *Information and Computation*, 94:1–28, 1991.

[24] P. Malliavin. *Integration and Probability*. Number 157 in Graduate Texts in Mathematics. Springer-Verlag, 1995. French edition appeared in 1993.

[25] R. Milner. *A Calculus for Communicating Systems*, volume 92 of *Lecture Notes in Computer Science*. Springer-Verlag, 1980.

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

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

[28] Probabilistic systems group, collected reports. Available from `www.comlab.ox.ac.uk` in the directory `/oucl/groups/probs/bibliography.html`.

[29] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 1966.

[30] J. J. M. M. Rutten. A calculus of transition systems (towards universal coalgebra). In A. Ponse, M. de Rijke, and Y. Venema, editors, *Modal Logic and Process Algebra, a bisimulation perspective*, number 53 in CSLI Lecture Notes, 1995. Available electronically from `www.cwi.nl/~janr`.

[31] J. J. M. M. Rutten. Universal coalgebra: a theory of systems. Technical Report CS-R9652, CWI AMsterdam, 1996. Available from URL `www.cwi.nl/~janr/papers/`.

[32] N. Saheb-Djahromi. Probabilistic LCF. In *Mathematical Foundations Of Computer Science*, number 64 in Lecture Notes In Computer Science. Springer-Verlag, 1978.

[33] N. Saheb-Djahromi. Cpos of measures for nondeterminism. *Theoretical Computer Science*, 12(1):19–37, 1980.

[34] R. Segala and N. Lynch. Probabilistic simulations for probabilistic processes. In B. Jonsson and J. Parrow, editors, *Proceedings of CONCUR94*, number 836 in Lecture Notes In Computer Science, pages 481–496. Springer-Verlag, 1994.

[35] R. van Glabbeek, S. Smolka, B. Steffen, and C. Tofts. Reactive generative and stratified models for probabilistic processes. In *Proceedings of the 5th Annual IEEE Symposium On Logic In Computer Science*, 1990.

[36] G. Winskel. Unpublished notes. Private Communication, 1996.

[37] G. Winskel and M. Nielsen. *Handbook of Logic in Computer Science*, volume 4, chapter Models for Concurrency. Oxford University Press, 1995.