Bisimulation for Labeled Markov Processes

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

Josée Desharnais*
School of Computer Science
McGill University
Montreal, Quebec, Canada
desharna@cs.mcgill.ca

Abbas Edalat†
Department of Computing
Imperial College
London, UK
ae@doc.ic.ac.uk

Prakash Panangaden*‡
BRICS
Aarhus University
Aarhus, Denmark
prakash@cs.mcgill.ca

Abstract

In this paper we introduce a new class of labeled transition systems - Labeled Markov Processes - and define bisimulation for them. Labeled Markov processes are probabilistic labeled transition systems where the state space is not necessarily discrete. The state space 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 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 contribution is the development of a notion of bisimulation for Markov processes on Polish spaces, which extends the Larsen-Skou definition for discrete systems. The proof that bisimulation is an equivalence relation turns out to depend on hard mathematical result of Edalat, which, as far as we know, embodies a new result in pure probability theory.

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 such as "hybrid systems". One can look at volume 1066 of Springer-Verlag Lecture Notes In Computer Science [AHS96] for examples of recent work on Hybrid systems. Ideas from continuous mathematics: differential equations, probability theory, stochastic processes, systems theory and others are becoming part of the mathematical toolkit of concurrency theorists. In the area of performance evaluation of computer systems, the use of stochastic Petri nets is quite widespread; one of the early papers is [Sif77]. Hillston has pioneered the use of process algebra in performance evaluation [Hil94]. Systems such as Hytech [HHWT95] and the Uppaal system [BLL$^+$96] have appeared based on hybrid systems and real-time systems respectively and use process equivalences very fruitfully.

*Research supported in part by NSERC.
†Research supported by EPSRC, UK.
‡On Leave from McGill University.

A new ingredient offered by computer science (apart from the manifest idea of computability or effectiveness) to these subjects is *compositionality*. Thus while the theory of stochastic processes [CM65] is concerned with a detailed analysis of the time evolution of systems behaving according to probabilistic laws very little is ever done to analyze the behaviour of coupled systems in a systematic way. Computer scientists have stressed compositionality as a way to attack the formidable intricacy of the systems they have dealt with. For example, in Hillston's work, compositionality is the key contribution of her approach to performance modeling.

The notion of bisimulation is central to the study of concurrent systems. While there are a bewildering variety of different equivalence relations between processes (two-way simulation, trace equivalence, failures equivalence and many more) bisimulation enjoys some fundamental mathematical properties, most notably its characterization as a fixed-point, which make it the most discussed process equivalence. Of course there are many different variants of bisimulation itself! 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 *labeled Markov process* which is meant ultimately to be part of a theory of interacting dynamical systems. The key technical contribution is the development 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 [Hoa85, Mil80, Mil89].

In brief, a labeled Markov process can be described 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. The state which 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 [LS91] 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; this means that both the labels and the probabilities must be the same.

In the case of a continuous state space, however, one cannot simply 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 appendix, we provide a rapid recapitulation of ideas from probability and measure theory. In the present case, we have to introduce a notion of sets for which "probabilities make sense" (i.e. a σ-field) 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, as we shall argue in the next section, is a compelling, natural notion. Unfortunately this definition cannot be adapted in any simple way to the continuous case. One cannot even use their basic terminology for bisimulation because their definition is inextricably wound up in the notion of state-to-state transition probabilities. Furthermore, once the correct generalization is given, it is a formidable technical problem to even show that bisimulation is an equivalence relation. This is solved by a construction due to Edalat which is mainly of interest to analysts. In fact, the construction heavily relies on properties that are not true for measure spaces in general. We have assumed a *polish space* structure. A polish space is a topological space which can be given a metric structure which generates the topology and such that the metric is complete and separable, i.e. there is a countable basis for the topology. In the classical study of Markov processes metric ideas play a significant role [Par67]. 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 $\mathbf{R}^n$ are as well.

The definition of bisimulation is inspired by the paper of Joyal, Nielsen and Winskel [JNW96] 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 [CN95] using infinitesimals[1]. Unfortunately one still needs to know how to construct pullbacks in the underlying category in order to apply their ideas. In fact for labeled Markov processes the best we can do is construct "semi-pullbacks", these allow one to complete a co-span but do not enjoy any universal property. Our notion of bisimulation is corresponds to the notion of coalgebra homomorphism [Rut96].

The rest of this paper is organized as follows. We describe some examples of systems with continuous state spaces in the next section. We then recapitulate the work of Larsen and Skou on discrete systems. The next two sections are concerned with the definition of labeled Markov processes and bisimulation. The last sections discuss related work and conclusions.

2 Two Examples of Processes

We begin with a simple example, more for introducing terminology than for any intrinsic interest. 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 is 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

[1] One can actually do this without talking about infinitesimals [Win96].

behaviour. The more complex internal behaviour is not externally visible. The point of a theory of bisimulation that encompasses such systems is to say when systems are equivalent. Of course this example is already familiar from the non-probabilistic 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. Bisimulation is a very strong notion and there are much weaker notions that are appropriate in various situations.

Now we consider a system which cannot be reduced to a discrete system. There are three labels $\{a, b, c\}$. Suppose that 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, or (c) 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 $f(p_0, p), g(p_0, p)$ and $h(p_0, p)$ respectively, with nontrivial dependence on p_0. In addition, there are two threshold values p_h and p_l. When the pressure rises above p_h the interaction labeled b is disabled, and when the pressure drops below p_l the interaction labeled 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 [LS91]. The systems that they consider will be referred to as *labeled Markov chains* in the present paper.

Definition 3.1 *A* **labeled 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, 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** *labeled Markov chain.*

The sets C_l are the sets of states that *can* do an l-action. If we have partial labeled Markov chains then 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 labeled 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 addition 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, in that model, one looks only at probabilistic branching, equality is precisely the Larsen-Skou notion of bisimulation [BK96].

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 precisely when two processes satisfy the same formulas.

One slight awkwardness in the Larsen-Skou definition is that when one compares two processes one has to combine the state sets and define probabilistic bisimulation on the combined states. This is a minor point but the reader should keep this in mind when reading the proofs below.

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. Bisimulation is most easily thought of in these terms. This presentation will also allow us to compare the theory with the more traditional theory of non-probabilistic processes; see, for example, the handbook article by Winskel and Nielsen [WN95].

In formulating the notion of Markov processes, we need to refine two concepts that were used in the discrete case. First we cannot simply 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-field on the set of states [Ash72, Bil95, Hal74, KT66, Rud66]. A review of the pertinent definitions appears in appendix A.

A key ingredient in the theory is the transition probability function.

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. The transition probability is really a *conditional probability*; it gives the probability of the system being in one of the states of the set A after the transition, *given* that it was in the state x before the transition. In general the transition probabilities could depend on time, in the sense that the transition probability could be different at every step (but still independent of past history); we are looking at the time-independent case.

We will work with *sub-probability* functions; i.e. with functions where $T(x, X) \leq 1$ rather than $T(x, X) = 1$. The mathematical results go through in this extended case and the resulting categories often are nicer, but the stochastic systems studied in the literature are usually only the very special version where $T(x, X)$ is either 1 or 0. In fact what is often done is that a state x with no possibility of making a transition is modeled 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 modeling 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. It is taken as a fundamental observable in process algebra that a process rejects or accepts an action.

The key mathematical construction requires a Polish space structure on the set of states. Thus instead of imposing an arbitrary σ-field structure on the set of states, we will require that the set of states be a Polish space and the σ-field be the Borel algebra generated by the topology.

Definition 4.2 *A* **partial labeled 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-field 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 **labeled Markov process**. *We have a partial labeled Markov process as above and a predicate* **Can** *on $S \times \mathcal{L}$ such that for every $(x, l) \in$* **Can** *we have $k_l(x, X) = 1$ and for every $(x, l) \notin$* **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 (S, Σ, k_l) for partial labeled 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 "labeled Markov chain" rather than "discrete, labeled, Markov process".

In a (partial), labeled, Markov *chain* the set of states is countable, and one can easily define a metric so that each point is an open set and the space is complete and obviously separable, the Borel σ-field is then 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 following way $k_l(s, A) \stackrel{\text{def}}{=} \sum_{s' \in A} P_l(s, s')$. 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 labeled Markov chain. The partial notion has been used by Cheng and Nielsen [CN95] to give an open maps [JNW96] presentation of bisimulation.

In order to define a category of Markov processes, we define simulation morphisms between processes.

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

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

Suppose that the system being simulated can make a transition from the state s to the set of states A with a certain probability, say p. Then the simulating system will simulate it in the following sense: the transition probability from $f(s)$ to any measurable set B such that $f(A) \subseteq B$ is greater than p. This is equivalent to the statement in the definition. We cannot directly talk about the transition probability from $f(s)$ to $f(A)$ since the latter may not be measurable.

One could have defined simulation to be continuous, however, the simulations play a secondary role in the present paper and we will simply require them to be measurable [2] We do at least require measurability for the definition to make sense. If f were not measurable we would not be guaranteed that $f^{-1}(A')$ is measurable and the definitions would be very awkward.

This notion extends the standard notion of simulation of labeled transition systems [WN95] in the following way. Given a partial labeled Markov chain (S, P_l), we can define a labeled transition system (lts) with the same label set as follows. We take the same set of states S and we define a labeled transition relation $\rightarrow \subseteq (S \times \mathcal{L} \times S)$ by $(s, l, s') \in \rightarrow \iff P_l(s, s') > 0$. Given two labeled transition systems, $(S_1, \rightarrow_1)$ and $(S_2, \rightarrow_2)$, a function $f : S_1 \longrightarrow S_2$ is a simulation morphism if $\forall s \in S_1 . s \xrightarrow{l} s' \Rightarrow f(s) \xrightarrow{l} f(s')$. We cannot do this for Markov processes because we can easily have systems where all the point-to-point transition probabilities are zero but the Markov process is nontrivial because the transition probabilities are nonzero to "larger" sets.

Proposition 4.4 *Given two partial labeled Markov chains, a simulation morphism between them is also a simulation morphism between the associated labeled transition systems.*

Proof (sketch). Suppose that we have two partial labeled Markov chains (S, Σ, k_l) and (S', Σ', k'_l) with f a simulation morphism from S to S'. Now suppose that in the associated lts the transition $s_1 \xrightarrow{l} s_2$ is possible. This means that $k_l(s_1, \{s_2\}) > 0$. Since f is a morphism we must have that $k'_l(f(s_1), \{f(s_2)\}) \geq k_l(s, f^{-1}(f(s_2))) \geq k_l(s_1, \{s_2\}) > 0$; hence in the lts $s_1 \xrightarrow{l} s_2$ is possible. ∎

5 Bisimulation for Markov Processes

The definition of bisimulation is very heavily influenced by the ideas of Joyal, Nielsen and Winskel [JNW96]. 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".

What one can prove for ordinary labeled transition systems is that if we take paths to be labeled paths in the usual sense then the open maps are morphisms that satisfy a condition called the "zigzag" condition[3]. 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 then 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 partial labeled Markov chains. For partial labeled Markov

[2] In older texts, such as Halmos [Hal74] or Rudin [Rud66] measurable is defined to mean that the inverse image of an *open* set is measurable. This means that the composite of two measurable functions need not be measurable. Our definitions are standard and, of course with this definition the composite of two measurable functions is measurable.

[3] The name arises from modal logic, see, for example, [Pop94].

processes it will be our definition of bisimulation. From now on, we assume that all systems are partial and we will stop writing the adjective "partial" explicitly.

Definition 5.1 *The objects of the category* **LMP** *are labeled Markov processes, having* $\mathcal{L}$ *as set of labels, with simulations as the morphisms. The category of labeled, Markov chains is written* **LMC** *and is the full subcategory of* **LMP** *that includes only the labeled, 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. $\forall l \in \mathcal{L}, s \in S, A' \in \Sigma', \;\; k_l(s, f^{-1}(A')) = k'_l(f(s), A')$.

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 labeled Markov process by its set of states. Following Joyal, Nielsen and Winskel ([JNW96]), we define bisimulation as the existence of a span of zigzag morphisms.

Definition 5.3 *Let* T *and* T' *be two labeled Markov processes.* T *is* probabilistically bisimilar to T' *(written* $T \sim T'$*) if there is a span of zigzag morphisms between them, i.e. there exists a labeled Markov process* S *and zigzag morphisms* f *and* g *such that*

$$\begin{tikzcd} & S \arrow[dl, "f"'] \arrow[dr, "g"] & \\ T & & T' \end{tikzcd}$$

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 [AM89, Rut95, Rut96] as well. We can view a labeled Markov process as a coalgebra of a suitable functor; in fact it is a functor introduced by Giry [Gir81] 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, these are precisely our zigzag morphisms.

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, it probably isn't true for probabilistic transition systems without some assumption.

Theorem 5.4 *Consider the following diagram*

$$\begin{tikzcd} S_1 \arrow[dr, "f_1"] & & S_2 \arrow[dl, "f_2"'] \\ & S & \end{tikzcd}$$

where f_1 and f_2 are zigzags. Then we can find an object U and zigzag morphisms g_1 and g_2 such that the following diagram commutes:

$$\begin{tikzcd}
 & U \arrow[dl, "g_1"'] \arrow[dr, "g_2"] & \\
S_1 \arrow[dr, "f_1"'] & & S_2 \arrow[dl, "f_2"] \\
 & S &
\end{tikzcd}$$

This is not a pullback because it does not have the universal property, it is not even a weak pullback. We refer to it as the *semipullback* construction. This proof rests on a technical result which involves quite intricate issues in measure theory; this result will be in a forthcoming paper of Edalat. The appendix contains the background definitions. The important consequence of this theorem is the following.

Corollary 5.5 *Probabilistic bisimulation is an equivalence relation.*

Here it is worth discussing why the theorem isn't trivial. It is natural to mimic the pullback construction in **Set**, thus, we take U to be the collection of all pairs (s_1, s_2) from $S_1 \times S_2$ such that $f_1(s_1) = f_2(s_2)$, and g_1 and g_2 to be the obvious projections, π_1 and π_2. The set U inherits a σ-field from the product σ-field on $S_1 \times S_2$. Now we have to define a transition probability $h : U \times \Sigma_U \longrightarrow [0,1]$. The requirement that π_1 and π_2 are zigzags forces us to define $h((s_1,s_2),(A_1 \times S_2) \cap U) = k_1(s_1, A_2)$ and similarly for π_2. Unfortunately, sets of the form $A_1 \times S_2$ and $S_1 \times A_2$ are not enough to determine a measure on all of Σ_U. Roughly speaking we need the values on all the rectangles not just on the "strips". Worse yet, given a set of the form $(A_1 \times A_2) \cap U$ we do not have A_1 and A_2 unambiguously determined; there could be two unrelated sets A_1' and A_2' with

$$(A_1 \times A_2) \cap U = (A_1' \times A_2') \cap U$$

so we cannot really use A_1 and A_2 to define h.

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 considered 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 $S_1 \times S_2$. This finesses the ambiguities of the naive approach. Actually carrying out the details of the proof involves some tricky analytic arguments to show measurability and is of less interest from the logical point of view.

We close this section with some examples. The first example is technically trivial but conceptually crucial.

Example 5.6 *We let the label set be the one element set. Consider a system (S, Σ, k) with S an arbitrarily complicated states space and Σ a σ-field generated by some Polish space structure on S. For example, S could be $\mathbf{R}$, the reals with the Borel algebra. We define the transition function,*

$k(s,A)$ in any manner we please subject only to the conditions of the definition of a transition function and to the condition that $\forall s \in S.k(s,S) = 1$; i.e. for every s, the distribution $k(s,\cdot)$ is a probability measure. Consider the trivial labeled Markov chain with just one label, one state and one transition from the state to itself with probability 1. These two systems are bisimilar!

This example allows us to clarify the discussion in the introduction. 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 modeling *interaction* and all such systems are indeed trivial from the point of view of interaction. In order to get nontrivial examples, one has to consider systems with richer label sets, and which are not always capable of making transitions with every label. Note also that many continuous systems are bisimilar to discrete ones. If we know this we can use the (presumably simpler) discrete system when reasoning about composite systems. Without our framework this would not be possible.

Example 5.7 *Consider the labeled Markov chain over a one-element label set $(\{a,b,c\},P)$ where the transition matrix, P is given by $P(a,b) = P(a,c) = 1/2$ with all other entries being zero. Consider another system over the same label set with 4 states $\{w,x,y,z\}$ and with transition matrix Q given by $Q(w,x) = Q(w,y) = Q(w,z) = 1/3$. These two systems are bisimilar. One can easily verify this by constructing the obvious "product" system and checking that the projections are zigzags. On the other hand there is no zigzag morphism between the two systems. These systems are small enough to check this by hand.*

The last example shows that bisimulation really must be defined in terms of spans, i.e. it genuinely is a relation.

Example 5.8 *The next example illustrates a continuous system. Consider the labeled 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 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 factor of $\lambda/2$ 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. Note that every positive state has a different probability density for jumping to a negative state. 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 Discrete Systems Revisited

In this section, we reconsider discrete systems (Markov chains) from the point of view of the bisimulation notion that we have defined for general partial labeled Markov processes. We show that the Larsen-Skou definition coincides with the "span of zigzags" definition. First, we have to say what it means for two Markov chains to be Larsen-Skou bisimilar, since the Larsen-Skou definition involves states of a single process rather than states of two different processes, and so doesn't apply without an appropriate interpretation. Roughly speaking, if there were an initial state in every Markov chain, then we would say that two Markov chains are Larsen-Skou bisimilar if and only if in a system "combining" them, their initial states are Larsen-Skou bisimilar. We have chosen not to equip Markov chains with initial states, instead we will use the following as a definition of Larsen-Skou bisimulation between two Markov processes.

Definition 6.1 Let (S_1, P_l^1) and (S_2, P_l^2) be two Markov chains and (T, H_l) their disjoint union, that is, T is the disjoint union of the two sets S_1 and S_2, and for $s,t \in T$, $H_l(s,t) = P_l^i(s,t)$ if s,t are both in S_i $(i=1,2)$ and zero otherwise. We will say that $S_1 \sim_{LS} S_2$ if in T there is a probabilistic bisimulation $\equiv_p$ such that for every state s_i of S_i there is a state s_j of S_j such that $s_i \equiv_p s_j$ $(i,j \in \{1,2\}, i \neq j)$.

We assume that all systems are partial in this section. Recall that we can always define a topology in which all the sets are open and hence all functions are continuous. Since the space is countable, the topology is clearly Polish. Thus, we can forget about the topology in this section but we keep in mind that discrete systems are a special case of the formalism of the previous section.

Proposition 6.2 *Let* (S, P_l) *and* (S', P_l') *be two labeled Markov chains.* $(S, P_l) \sim_{LS} (S', P_l')'$ *if and only if there exists a span of zigzag morphisms* f *and* f'*:*

$$\begin{tikzcd} & U \arrow[dl, "f"'] \arrow[dr, "f'"] & \\ S & & S' \end{tikzcd}$$

Proof. We write k_l and k'_l for the two transition functions induced by the transition matrices P_l and P'_l respectively.

$\Leftarrow$: We first show that if $S \xrightarrow{f} S'$, where f is a zigzag morphism, then $S \sim_{LS} S'$. Let (t, H_l) be the disjoint union of (S, P_l) and (S', P'_l), and h_l be the transition function induced by the transition matrix H_l. Now f defines the following equivalence relation, R, on T:

$$s_1 R s_2 \iff (s_1 = s_2) \vee (f(s_1) = s_2) \vee (f(s_1) = f(s_2)).$$

The equivalence classes are of the form $\{s'\} \cup f^{-1}(s')$ for each $s' \in S'$; thus each equivalence class can be represented uniquely by an element of S'. Let $l \in \mathcal{L}, s_1, s_2 \in T$ such that $s_1 R s_2$, and choose any $t' \in S'$ that represents the equivalence class $\{t'\} \cup f^{-1}(t')$. We want to show that $h_l(s_1, -)$ and $h_l(s_2, -)$ agree on $\{t'\} \cup f^{-1}(t')$.

First assume $s_1 \in S$ and $s_2 \in S'$, meaning that $f(s_1) = s_2$. Then

$$\begin{aligned}
h_l(s_1, \{t'\} \cup f^{-1}(t')) &= k_l(s_1, f^{-1}(t')) \ \text{ since } s_1 \in S \\
&= k'_l(f(s_1), \{t'\}) \ \text{ since } f \text{ is a zigzag morphism} \\
&= h_l(s_2, \{t'\} \cup f^{-1}(t')),
\end{aligned}$$

which is precisely the condition for Larsen-Skou bisimulation. Now if s_1 and s_2 are both in S (and still R-related), they have the same image, so $s_1 \, R \, f(s_1)$ and $f(s_1) = f(s_2)$ and $f(s_2) \, R \, s_2$ and we can simply apply the above calculation. Finally we have the trivial case where s_1 and s_2 are both in S'. They are then equal and we are done since, f being a surjective function, every state of S is R-equivalent to a state of S' are vice versa. Since we know that Larsen-Skou bisimulation is an equivalence relation it follows that whenever we have a span of zigzags connecting two labeled Markov chains they are Larsen-Skou bisimilar.

$\Rightarrow$: Assume $(S, P_l) \sim_{LS} (S', P'_l)$, with $\equiv_p$ the probabilistic bisimulation over (T, H_l) the disjoint union of (S, P_l) and (S', P'_l). We need to construct a span of zigzag morphisms

$$\begin{tikzcd}
 & U \arrow[dl, "f"'] \arrow[dr, "f'"] & \\
S & & S'
\end{tikzcd}$$

To do this, let $U = (U, (Q_l)_{l \in \mathcal{L}})$ where $U = \{(s, s') \in S \times S' : s \sim_{LS} s' \text{ in } T\}$ and where the transition matrix Q is given by, for $l \in \mathcal{L}$,

$$Q_l((s, s'), (t, t')) = \frac{P_l(s, t) \, P'_l(s', t')}{h_l(s, [t]_{\sim_{LS}})}$$

where $[t]_{\sim_{LS}}$ denotes the equivalence class containing t in T, and h_l is the transition function induced by the transition matrix H_l. Since $s \sim_{LS} s'$ and $t \sim_{LS} t'$, we have by definition of $\sim_{LS}$ that $h_l(s, [t]_{\sim_{LS}}) = h_l(s', [t]_{\sim_{LS}}) = h_l(s', [t']_{\sim_{LS}})$.

To prove that U is a partial labeled Markov chain, we need that for any $(s, s') \in U$,

$$\sum_{(t,t') \in U} Q_l((s, s'), (t, t')) \leq 1.$$

This will follow from the proof that we have zigzag morphisms from U to S and S'. As morphisms $f : U \to S$ and $f' : U \to S'$, we simply take the left and right projections which are surjective by definition of Larsen-Skou probabilistic bisimulation. We prove that they are zigzag morphisms. We write q_l for the transition function derived from the transition matrix Q_l. First note that

$$\forall t \in S. f^{-1}(t) = \{t\} \times (S' \cap [t]_{\sim_{LS}}).$$

For any $l \in \mathcal{L}, (s, s') \in U, t \in S$, we have

$$\begin{aligned}
q_l((s,s'), f^{-1}(t)) &= \sum_{t' \in S' \cap [t]_{\sim_{LS}}} Q_l((s,s'),(t,t')) \\
&= \sum_{t' \in S' \cap [t]_{\sim_{LS}}} \frac{P_l(s,t)\, P'_l(s',t')}{h_l(s,[t]_{\sim_{LS}})} \\
&= \frac{P_l(s,t)}{h_l(s',[t]_{\sim_{LS}})} \sum_{t' \in S' \cap [t]_{\sim_{LS}}} P'_l(s',t') \\
&= \frac{P_l(s,t)}{k'_l(s,[t]_{\sim_{LS}} \cap S)} k'_l(s',[t]_{\sim_{LS}} \cap S') \\
&= P_l(s,t) = k_l(f(s,s'),\{t\}).
\end{aligned}$$

and f is thus a zigzag morphism. The same argument applies to f'. ∎

7 Towards A Modal Logic for Bisimulation

In this section we begin the study of logics that characterize bisimulation. For discrete systems we can characterize bisimulation in terms of a simple probabilistic version of Henessy-Milner logic. The continuous case is not covered and we hope to address it later. What we can show in the continuous case is that if two systems are bisimilar they will satisfy the same formulas of an appropriate modal logic; the converse is not attempted here.

We follow the plan in the treatment of Larsen and Skou [LS91] but our proof is different. The result we have, unlike that of Larsen and Skou, does not depend on a minimum branching assumption. This is quite a surprise. The point is that the probabilistic systems we are considering, without explicit nondeterminism, resemble deterministic systems quite closely, rather than nondeterministic systems. In the latter case - as is well-known [Mil90] - without a finite-branching assumption the result is false, unless the logic has infinitary conjunctions.

Of course we would really like to have a logical characterization for the continuous state-space systems that we are concerned with in the present paper but this seems to require new techniques. We give a simple counterexample showing the problem. In an earlier version [BDEP97] of this paper we incorrectly asserted that this characterization worked if only one of the systems is discrete.

We take as the syntax the following formulas:

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

where a is an action from the fixed set of actions $\mathcal{L}$ and q is a rational number. Given a labeled 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 $[\![\phi]\!]_S$. This logic differs from that of Larsen and Skou in that it uses explicit negation.

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

Proposition 7.1 *For all formulas ϕ, we have $[\![\phi]\!] \in \Sigma$.*

Proof. We proceed by structural induction on ϕ. The base cases corresponding to T is trivial since $S \in \Sigma$. The propositional connectives are trivial because a σ-field is closed under countable union and complementation, by definition. Finally we have $[\![\langle a \rangle_q \phi]\!] = k_a(\cdot, [\![\phi]\!])^{-1}([q, 1]) \in \Sigma$. To justify this first note that, by hypothesis, $[\![\phi]\!] \in \Sigma$ so $k_a(s, [\![\phi]\!])$ is meaningful. Secondly, k_a is a measurable function in its first argument and finally intervals are Borel. ∎

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

Proposition 7.2 *If f is a zigzag morphism from S to S' then for all formulas ϕ,*

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

Proof. We proceed by structural induction on ϕ. The only nontrivial cases correspond to the modal formulas. We proceed as follows; $s \models \langle a \rangle_q \phi$ means $k_a(s, [\![\phi]\!]_S) \geq q$. By the inductive hypothesis we have that $f^{-1}([\![\phi]\!]_{S'}) = [\![\phi]\!]_S$. Thus we have

$$q \leq k_a(s, [\![\phi]\!]_S) = k_a(s, f^{-1}([\![\phi]\!]_{S'})) = k'_a(f(s), [\![\phi]\!]_{S'}).$$

But the last equality means $f(s) \models \langle a \rangle_q \phi$. ∎

From this we get the immediate corollary, but first we need to define bisimilarity of states. 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 $f'(t) = s'$. Now this defines an equivalence relation. In anticipation of this we write, $s \sim s'$ when two states are bisimilar. To check the claim suppose that for $s \in S$, $s' \in S'$ and $s'' \in S''$ we have $s \sim s'$ and $s' \sim s''$. Then we have $f : T \longrightarrow S$ and $f' : T \longrightarrow S'$ and $g : T' \longrightarrow S'$ and $g' : T' \longrightarrow S''$ with the required properties and we can use the pullback $(h : U \longrightarrow T, k : U \longrightarrow T')$ of $f' : T \longrightarrow S'$ and $g : T' \longrightarrow S'$ to obtain $p : U \longrightarrow S$ and $q : U \longrightarrow S''$. In order to show that $s \sim s''$ we proceed as follows. By assumption there exists $t \in T$ and $t' \in T'$ such that $f(t) = s$, $f'(t) = s'$, $g(t') = s'$ and $g'(t') = s''$. Now $(t, t') \in U$ since $f'(t) = s'$ and $g(t') = s'$. Therefore, putting $p = f \circ h$ and $q = g' \circ k$ we have $p((t, t')) = f(t) = s$ and $q((t, t')) = g'(t') = s''$. Thus $s \sim s''$.

Corollary 7.3 *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 they 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 7.4 *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 $k_a(s, [\![\phi]\!]_S) = k'_a(s', [\![\phi]\!]_{S'})$.*

Proof. Suppose that the equation does not hold. Then, say, for some ϕ, $k_a(s, [\![\phi]\!]_S) < k'_a(s', [\![\phi]\!]_{S'})$. Now we choose a rational number q between these values. Now it follows that $s' \models \langle a \rangle_q \phi$ but $s \not\models \langle a \rangle_q \phi$, which contradicts the assumption that s and s' satisfy all the same formulas. ∎

Now we would like to show that if two systems satisfy all the same formulas they must be bisimilar. This is - as should be expected - much harder. We will only succeed in the case that the systems are discrete. Instead of defining a span of zigzags directly, we can define a cospan and use the semipullback 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 topology and σ-field. Clearly T is Polish. 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. The set V is a Polish space since it is discrete. Furthermore since all the spaces involved are discrete all maps are continuous - this is what fails if one of the systems is continuous.

The composites $r \circ \iota$ and $r \circ \iota'$ are continuous 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 7.5 *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 $k_a(s, [t] \cap S) = k'_a(s', [t] \cap S')$.*

Proof. Let $\{\phi_i | i \in \mathbf{N}\}$ be an enumeration of the formulas satisfied by the members of $[t]$. If s is any state such that $s \bowtie t$, s must satisfy any formula satisfied by t, so s is in the intersection. Conversely, if $s \not\bowtie t$, there is a formula ϕ that distinguishes s and t, hence either $s \notin [\![\phi]\!]$ and $t \in [\![\phi]\!]$ or $s \in [\![\phi]\!]$ and $t \notin [\![\phi]\!]$; since the latter is equivalent to saying $s \notin [\![\neg\phi]\!]$ and $t \in [\![\neg\phi]\!]$, we have in both cases that s is not in the intersection of the ϕ_i's.

Let $\forall i \in \mathbf{N}. \psi_i \stackrel{\text{def}}{=} \bigwedge_{l=1}^{i} \phi_l$. We show that $[t] = \cap_{i=1}^{\infty} [\![\psi_i]\!]$. The sets $[\![\psi_i]\!]_S$ and $[\![\psi_i]\!]_{S'}$ are both measurable for all i and decrease as i increases. Furthermore, these decreasing chains converge to $[t] \cap S$ and $[t] \cap S'$ respectively. By the monotone convergence property of measures [Rud66] we have

$$k_a(s, [t] \cap S) = \lim_{i \longrightarrow \infty} k_a(s, [\![\psi_i]\!]_S),$$

by the previous proposition we have

$$= \lim_{i \longrightarrow \infty} k'_a(s', [\![\psi_i]\!]_{S'}) = k'_a(s', [t] \cap S').$$

∎

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. Now 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 7.5, we can see that $h_a([s], [t]) = k'_a(s', q'^{-1}([t]))$. The morphisms are now zigzags. Thus we have proved the following proposition

Proposition 7.6 *If we have two discrete systems then they are bisimilar iff they obey the same formulas.*

We close with the following simple counterexample showing that if one of the systems is continuous then the surjection onto the quotient is not necessarily continuous. Let $S = [-1, 1]$ with

$$k(x, -) = \begin{cases} \delta_x \text{ if} x \in [-1, 0) \\ \lambda\delta_x \text{ if } x \in [0, 1] \end{cases}$$

where δ_x is the Dirac measure based at x and λ is a fixed constant with $0 \leq \lambda < 1$.

Clearly there are two equivalence classes corresponding to the subsets $[-1, 0)$ and $[0, 1]$. So we have that the set of equivalence classes is given by $S' = \{A, B\}$, where A represents $[-1, 0)$ and B represents $[0, 1]$. If we require the quotient map $q : S \to S'$ be continuous then the open subsets of S' are precisely S', $\emptyset$ and $\{A\}$ which does not give a Polish space. Conversely if we made the quotient space Polish - which we can do since it is discrete - the map onto the quotient is not continuous.

8 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 [SD78, SD80] and of Kozen [Koz81, Koz85]. 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 first paper with an abstract categorical approach to stochastic processes is by Giry [Gir81], but she studies categorical constructions rather than process equivalences. Of course her work, inspired originally by Lawvere, does provide some of the mathematical underpinnings of our ideas. 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 labeled Markov processes and ordinary transitions quite striking. For a more detailed discussion of generalized notions of relations see [ABP97].

The starting point of our investigations was the pioneering paper of Larsen and Skou [LS91] which gave the first compelling analysis of probabilistic bisimulation. As we have mentioned before, their work is couched in terms of state-to-state transition probabilities so there were both conceptual as well as technical obstacles to overcome. The work of Joyal, Nielsen and Winskel [JNW96] and of Cheng and Nielsen [CN95] 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 [Hil94] 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 [Pro]. The focus has been on equational laws satisfied by processes. From the semantic point of view, they have extensively developed and enriched Kozen's [Koz85] predicate-transformer view. They have considered continuous state-space systems and have incorporated nondeterminism in their framework. As with the earlier work of Kozen [Koz81, Koz85], stochastic kernels play an important role.

Finally we mention a sample of the large amount of work on various notions of testing [CSZ92, Chr90, JY95], simulations [SL94], specification [JL91], model checking [HK96], generative processes [vGSST90] 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.

9 Conclusions

In this paper we have two main contributions. First, we develop a notion of labeled 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. 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 is indeed an equivalence relation. This requires a fairly difficult 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.

The secondary results of this paper are

1. that we have shown how this notion of bisimulation extends the definition of Larsen and Skou [LS91] for the discrete case.

2. and that we have a logical characterization of bisimulation which works even when the systems are discrete and *does not use the minimum deviation assumption* but at the price of using negation in the logic.

The important questions to answer for the future are the following.

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? It might be necessary 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 [Koz85] in his study of probabilistic PDL.

3. What is the correct 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 correct 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 appropriate 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 would really bring this work into contact with extant work on hybrid systems.

Acknowledgments

Abbas Edalat would like to thank Terry Lyons for helpful discussions and EPSRC (UK) for support. Prakash Panangaden would like to thank Samson Abramsky, Christel Baier, 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.

A A Review of Measure Theory

For completeness, we give the basic definitions in this section. We only discuss positive measures. We assume, however, that the reader knows the basic ideas of measure theory and probability as expounded in, for example "Probability and Measure" by Billingsley [Bil95] or "Real Analysis and Probability" by Ash [Ash72] or the book with the same title by Dudley [Dud89] or "Introduction to Measure and Probability" by Kingman and Taylor [KT66].

The basic concept is that of a σ-field on a set. These are the sets for which a notion of "measure" or "size" can be defined.

Definition A.1 *A* **σ-field** *on a set X is a family of subsets of X which includes X itself and which is closed under complementation and countable unions.*

A set equipped with a σ-field is called a *measurable space*. For simple situations, we can usually take the σ-field to be all the subsets but when we get to sets such as the real line (or closed intervals of the real line) we cannot any longer work with this σ-field because one cannot consistently define probabilities (measures) sensibly on all subsets.

Fact A.2 *The intersection of an arbitrary collection of σ-fields is a σ-field.*

Thus, given an arbitrary collection of subsets of a set we can define the σ-field generated by the collection. For example, given a topological space $(X, \mathcal{T})$, we can define the σ-field, often written

$\mathcal{B}$, generated by the open sets (or, equivalently, by the closed sets). This is usually called the *Borel algebra.*

The notion of *measure* generalizes the concept of "length of an interval" to arbitrary σ-fields.

Definition A.3 *Given a σ-field (X, Σ), a* **(subprobability) measure** *on X is a ($[0,1]$-valued) $[0,\infty]$-valued set function, μ, defined on Σ such that*

- $\mu(\emptyset) = 0$,
- *for a pairwise disjoint, countable collection of sets, $\{A_i | i \in I\}$, in Σ, we require*

$$\mu(\bigcup_{i\in I} A_i) = \sum_{i\in I} \mu(A_i).$$

In addition, for probability measures we require $\mu(X) = 1$.

The second property above is called *countable additivity.* A *probability space* is a set X equipped with a σ-field Σ and a measure P such that $P(X) = 1$.

Example A.4 *The key example in the subject is provided by Lebesgue measure on the real line. Let* $\mathbf{R}$ *stand for the real numbers. The elements of the σ-field generated by the open intervals are called the Borel sets. There is a measure defined on these sets which on intervals agrees with the intuitive notion of length. This is the Lebesgue measure. It may be calculated as follows. Suppose that A is a Borel set. We take a family of intervals, $\mathcal{J}$, (not necessarily pairwise disjoint) which covers A. We define the length of $\mathcal{J}$ as the sum of the lengths of all the intervals in $\mathcal{J}$. Now we take the inf of the lengths over all $\mathcal{J}$ which cover A. One can verify that single points have measure zero and that any countable set has measure zero. The Cantor set gives an example of an uncountable set of measure zero. The set of all irrationals between 0 and 1 has measure 1. This measure is translation invariant and is in fact the unique translation-invariant measure which assigns 1 to the interval $[0,1]$. The fact that makes this entire theory so complicated is the existence of non-measurable sets.*

It is worth clarifying how the word "measurable" is used in the literature. Given a σ-field Σ on a set X one says "measurable set" for a member of Σ. Suppose that one has a measure μ. One can have the following situation. There can be sets of measure zero which contain non-measurable subsets. Because these sets are not measurable one cannot say that they have measure zero. This happens with Lebesgue measure on the Borel sets in the real line, for example. There is a "completion" procedure[4] which produces a larger σ-field and an extension of the original measure in such a way that all subsets of sets of measure zero are measurable and have measure zero. When applied to the Borel subsets of the real line we get the so called Lebesgue measurable sets. One often uses the phrase "measurable set" to mean a set which belongs to the completed σ-field rather than the original σ-field. In the detailed proof of the next section we say "Borel set" when we mean set that is generated by the topology and "measurable set" when we mean a set in the completed σ-field.

Functions between measurable spaces are called measurable functions.

Definition A.5 *A function $f : (X, \Sigma_X) \longrightarrow (Y, \Sigma_Y)$ between measurable spaces is said to be* **measurable** *if $\forall B \in \Sigma_Y . f^{-1}(B) \in \Sigma_X$.*

[4] This is a terrible name because it gives the mistaken impression that the result cannot be further extended.

This resembles the topological notion of continuity. The fundamental difference is that limits of sequences of measurable functions is measurable; the analogous property certainly does not hold for continuity. The most basic measurable functions are the *characteristic functions.* Given a measurable subset, say A, of X, the characteristic function of A, written χ_A, is a function from X to $\mathbf{R}$ such that $\chi_A(x) = 1$ if x is in A and $\chi_A(x) = 0$ if x is not in A.

Measurable spaces and measurable functions form a category called **Mes**. This category has countable coproducts and all finite limits (in particular products and equalizers).

A key result in classical measure theory is the Radon-Nikodym theorem. We write $\mu << \nu$ if $\nu(A) = 0 \Rightarrow \mu(A) = 0$ for any measurable set A. We say that μ is *absolutely continuous* with respect to ν. The Radon-Nikodym says that if the measures are finite[5], then there is a measurable function h such that for any measurable set A we have

$$\mu(A) = \int_A h d\nu.$$

Furthermore this function is unique *almost everywhere*; in other words, if there is another function h' with the same property, it can only differ from h on a set of ν measure 0. This theorem gives a notion of differentiating one measure by another. One often calls h the *Radon-Nikodym derivative* of μ with respect to ν.

An important application of the Radon-Nikodym theorem is the construction of *conditional probability distributions.* Conditional probabilities relate probabilistic information with definite information and are the key to probabilistic reasoning. In the discrete case the conditional probability can be defined as follows

$$P(A|B) \stackrel{\text{def}}{=} \frac{P(A \cap B)}{P(B)}.$$

This should be read as "the probability of A being true *given that* B is true." Of course, this makes sense only if $P(B) \neq 0$. In the continuous case most of the probabilities are 0, so conditional probabilities must be defined more subtly. We will present the concept in three stages.

Suppose that we have a probability space (X, Σ, P). Recall that this describes a random mechanism which produces results $x \in X$, distributed according to P. Now fix some measurable set B. Suppose that the observer *knows* that x lies in B. Then, from the point of view of this observer, the probability that x lies in some other measurable set, say A, is $P(A|B)$. The observer's knowledge has restricted the sample space but the probability distribution has not changed in any absolute sense. This is the first stage in our presentation and essentially reviews the discrete case.

For the second stage we generalize the above situation slightly. Let $\{B_i | i \in I\}$ be a partition of X by sets in Σ, and let Σ' be the σ-field generated by this partition. Now imagine that the observer performs an experiment which allows him to determine to which member of the collection $\{B_i | i \in I\}$ the point x belongs. This is the same as determining the members of Σ' to which x belongs. Now after performing the experiment, the conditional probability estimates for x being in A are given by $Q(x) =_{df} P(A|B_i)$ if $x \in B_i$. We write $P(A|\Sigma')(\cdot) : X \longrightarrow [0, 1]$ as the conditional probability of A given Σ' and define it to be equal to Q if $P(B_i) \neq 0$. If $P(B_i)$ is 0 we give an arbitrary value to $P(A|\Sigma')(x)$ for $x \in B_i$. Thus there may be different *versions* of the conditional probability density but they differ on sets of zero probability.

In the final stage we define conditional probability density by viewing Σ' as an arbitrary sub σ-field of Σ. We wish to know the following kind of information. Suppose that we do an experiment

[5] Actually a weaker condition suffices, but the present result is all one needs for probability.

and find out in which subset of Σ' a point lies; now we want to know how to estimate probabilities that the point lies in A. We can define a new probability measure V by $V(B) = P(A \cap B)$ for $B \in \Sigma'$. Now using the Radon-Nikodym theorem we can define a conditional probability density function $P(A|\Sigma') : X \longrightarrow [0,1]$ with the properties:

1. $P(A|\Sigma')(\cdot)$ is measurable with respect to Σ' and integrable as well;
2. for a set B in Σ' we have

$$\int_B P(A|\Sigma')(x)dP(x) = P(A \cap B).$$

This density replaces the naive conditional probabilities of the discrete case.

Stochastic processes are, in fact, usually defined in terms of *random variables* and the conditional probabilities are set up accordingly. The conditional probability distributions that we have defined above are very special cases. We now introduce the basic concepts which should allow the reader to begin a study of the literature. The picture that one typically has based on experience of labeled transition systems is quite far from the formalism used in the literature on stochastic processes. Most importantly, the latter is couched in terms of random variables on an abstract probability space rather than as state transitions on a state space.

Definition A.6 *A* ***random variable*** *f on a probability space (X, Σ_X, P) is a measurable function $f : (X, \Sigma_X) \longrightarrow (Y, \Sigma_Y)$.*

Often, in the probability literature, one defines random variables to be real valued. Given a random variable as above we obtain a probability measure on Y by $P_f(B \in \Sigma_Y) = P(f^{-1}(B))$. It is now natural to talk about a conditional probability $P(y \in Y, A \in \Sigma_X)$ satisfying the condition

$$P(f^{-1}(B) \cap A) = \int_B P(y, A)dP_f(y).$$

The Radon-Nikodym theorem implies that such a conditional probability exists.

One can ask whether the notion of conditioning with respect to a sub-σ-field is possible in the setting of random variables. The following definition is naturally suggested.

Definition A.7 *Let $f : (X, \Sigma_X) \longrightarrow (Y, \Sigma_Y)$ be a random variable and let $\mathcal{G}$ be a sub-σ-field of Σ_X. A function $Q : X \times \Sigma_Y \longrightarrow [0,1]$ is called a* ***regular conditional probability*** *for f given $\mathcal{G}$ iff*

1. *$Q(x, \cdot)$ is a probability measure for each fixed x*
2. *for each fixed $B \in \Sigma_Y$, the function $Q(\cdot, B)$ is $\mathcal{G}$-measurable and*
3. *for each $G \in \mathcal{G}$ we have the equation*

$$\int_G Q(x, B)dP(x) = P(f^{-1}(B) \cap G).$$

Finally we have the following concept where there are two random variables and a regular conditional probability distribution relating them.

Definition A.8 *Suppose we are given a probability space $(\Omega, \mathcal{F}, P)$, and measurable spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$ with random variables $x : \Omega \longrightarrow S$ and $y : \Omega \longrightarrow T$. A* ***regular conditional probability distribution*** *of x given y is a function, written $P_x^y(\cdot|\cdot) : \mathcal{S} \times T \longrightarrow [0,1]$ such that*

1. $P^y_x(\cdot|t)$ *is a probability measure on* S*,*

2. $P^y_x(A|\cdot)$ *is* $\mathcal{T}$*-measurable for all* $A \in \mathcal{S}$*,*

3. $P(x^{-1}(A) \cap y^{-1}(B)) = \int_B P^y_x(A|t)P_y(dt)$ *for all* $A \in \mathcal{S}$ *and all* $B \in \mathcal{T}$*.*

Now a stochastic process on a given probability space $(\Omega, \mathcal{F}, P)$ is just a sequence of random variables $x_n : \Omega \longrightarrow \mathcal{S}$. One can think of the indexing set as time. In general, the distribution of one of the x_n could depend on all the previous x_i, but in a Markov process it only depends on the last one. Thus for a Markov for each transition it is the pair (x_{n-1}, x_n) which is relevant. In order to recover the notion of transition probability one uses the concepts of the last paragraph. In general, one gets transition probabilities that may change in time, however we look at the *stationary* case, where these transition probabilities are time independent and hence can be described by a single conditional probability distribution. It is important to realize that the probability space Ω is not the state space, it is better thought of as the space of "paths" or "trajectories" which the system undergoes.

References

[ABP97] S. Abramsky, R. Blute, and P. Panangaden. Nuclear and trace ideals in tensor-$*$ categories. Available from `www.daimi.aau.dk/~prakash/` or from `www-acaps.cs.mcgill.ca/~prakash/`, 1997.

[AHS96] R. Alur, T. Henzinger, and E. Sontag, editors. *Hybrid Systems III*, number 1066 in Lecture Notes in Computer Science. Springer-Verlag, 1996.

[AM89] P. Aczel and N. Mendler. A final-coalgebra theorem. In *Category Theory and Computer Science*, Lecture Notes In Computer Science, pages 357–365, 1989.

[Arv76] W. Arveson. *An Invitation to C^*-Algebra*. Springer-Verlag, 1976.

[Ash72] R. B. Ash. *Real Analysis and Probability*. Academic Press, 1972.

[BDEP97] Richard Blute, Josée Desharnais, Abbas Edalat, and Prakash Panangaden. Bisimulation for labelled markov processes. In *Proceedings of the Twelfth IEEE Symposium On Logic In Computer Science, Warsaw, Poland.*, 1997.

[Bil95] P. Billingsley. *Probability and Measure*. Wiley-Interscience, 1995.

[BK96] C. Baier and M. Kwiatkowska. Domain equations for probabilistic processes. Available from URL http://www.cs.bham.ac.uk/ mzk/, March 1996.

[BLL+96] 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.

[Chr90] 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.

[CM65] D. R. Cox and H. D. Miller. *The Theory of Stochastic Processes.* Chapman and Hall, 1965.

[CN95] 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.

[CSZ92] 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.

[Dud89] R. M. Dudley. *Real Analysis and Probability.* Wadsworth and Brookes/Cole, 1989.

[Eda] A. Edalat. Semi-pullbacks in categories of markov processes. Available from `http://theory.doc.ic.ac.uk:80/people/Edalat/Markov.ps.Z` .

[Gir81] Michèle 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.

[Hal74] P. Halmos. *Measure Theory.* Number 18 in Graduate Texts in Mathematics. Springer-Verlag, 1974. Originally published in 1950.

[HHWT95] 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.

[Hil94] Jane Hillston. *A Compositional Approach to Performance Modelling.* PhD thesis, University of Edinburgh, 1994. To be published as a Distinguished Dissertation by Cambridge University Press.

[HK96] 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/.

[Hoa85] C. A. R. Hoare. *Communicating Sequential Processes.* Series in Computer Science. Prentice-Hall International, London, 1985.

[JL91] 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.

[JNW96] A. Joyal, M. Nielsen, and G. Winskel. Bisimulation from open maps. *Information and Computation*, 127(2):164–185, 1996.

[JY95] 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.

[Koz81] D. Kozen. Semantics of probabilistic programs. *Journal of Computer and Systems Sciences*, 22:328–350, 1981.

[Koz85] D. Kozen. A probabilistic PDL. *Journal of Computer and Systems Sciences*, 30(2):162–178, 1985.

[KT66] J. F. C. Kingman and S. J. Taylor. *Introduction to Measure and Probability.* Cambridge University Press, 1966.

[LS91] K. G. Larsen and A. Skou. Bisimulation through probablistic testing. *Information and Computation*, 94:1–28, 1991.

[Mil80] R. Milner. *A Calculus for Communicating Systems*, volume 92 of *Lecture Notes in Computer Science.* Springer-Verlag, 1980.

[Mil89] R. Milner. *Communication and Concurrency.* Prentice-Hall, 1989.

[Mil90] Robin Milner. *Handbook of Theoretical Computer Science: Volume B*, chapter Operational and Algebraic Senmantics of Concurrent Processes, pages 1201–1242. MIT Press, 1990.

[Par67] K. R. Parthasarathy. *Probability Measures on Metric Spaces.* Academic Press, 1967.

[Pop94] Sally Popkorn. *First Steps in Modal Logic.* Cambridge University Press, 1994.

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

[Rud66] W. Rudin. *Real and Complex Analysis.* McGraw-Hill, 1966.

[Rut95] 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`.

[Rut96] 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/`.

[SD78] N. Saheb-Djahromi. Probabilistic LCF. In *Mathematical Foundations Of Computer Science*, number 64 in Lecture Notes In Computer Science. Springer-Verlag, 1978.

[SD80] N. Saheb-Djahromi. Cpos of measures for nondeterminism. *Theoretical Computer Science*, 12(1):19–37, 1980.

[Sif77] J. Sifakis. Use of petri nets for performance evaluation. In H. Beilner and E. Gelenbe, editors, *Measuring, Modeling and Evaluating Computer Systems*, pages 75–93. North-Holland, 1977.

[SL94] 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.

[vGSST90] 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.

[Win96] G. Winskel. Unpublished notes. Private Communication, 1996.

[WN95] G. Winskel and M. Nielsen. *Handbook of Logic in Computer Science*, volume 4, chapter Models for Concurrency. Oxford University Press, 1995.