Metrics for Labelled Markov Processes

Josée Desharnais*
Département d'Informatique
Université Laval
Québec, Canada, G1K 7P4
josee.desharnais@ift.ulaval.ca

Vineet Gupta
Stratify Inc.
501 Ellis St.
Mountain View CA 94043, USA
vineet@stratify.com

Radha Jagadeesan†
School of CTI, DePaul University,
243 S. Wabash Avenue
Chicago, Illinois 60604-2287, USA
rjagadeesan@cs.depaul.edu

Prakash Panangaden*
School of Computer Science
McGill University
Montréal, Canada, H3A 2A7
prakash@cs.mcgill.ca

September 18, 2003

Abstract

The notion of process equivalence of probabilistic processes is sensitive to the exact probabilities of transitions. Thus, a slight change in the transition probabilities will result in two equivalent processes being deemed no longer equivalent. This instability is due to the quantitative nature of probabilistic processes. In a situation where the process behaviour has a quantitative aspect there should be a more robust approach to process equivalence. This paper studies a metric between labelled Markov processes. This metric has the property that processes are at zero distance if and only if they are bisimilar. The metric is inspired by earlier work on logics for characterizing bisimulation and is related, in spirit, to the Kantorovich metric.

1 Introduction

Probability, like nondeterminism, is an abstraction mechanism used to hide inessential or unknown details. Statistical mechanics — originated by Boltzmann, Gibbs, Maxwell and others — is the fundamental successful example of the use of the probabilistic abstraction. What makes it successful is that an intractable number of exact mechanical equations are replaced by a much smaller, tractable number of relations between suitable *averages* of mechanical quantities.

Similarly, in our models we use probabilities to average over, and thus abstract away, the effects of a myriad details - some of which may be impossible to observe in practice - that would have made the transition systems determinate. Our investigations are concerned with the development of contextual reasoning principles for concurrent interacting probabilistic systems. Consider the following paradigmatic examples.

*Research supported in part by NSERC and MITACS.
†Research supported by NSF.

Example 1.1 A paper by by Alur et. al. [AJKvO97] analyzes a component of the Lucent Technologies' 5ESS ® telephone switching system that is responsible for detecting malfunctions on the hardware connections between switches. This component responds to alarms generated by another complicated system that is only available as a black-box. A natural model to consider for the black-box is a stochastic one, representing the timing and duration of the alarm by random variables with a given probability distribution. The paper [AJKvO97] shows that the desired properties hold with high probability, showing that the component being analyzed approximates the idealized behavior with sufficient accuracy.

Example 1.2 Consider model-based diagnosis settings. Often information about *failure models* and their associated probabilities is obtained from field studies and studies of manufacturing practices. Failure models can be incorporated by assigning a variable, called the *mode* of the component, to represent the physical state of the component, and associating a failure model with each value of the mode variable. Probabilistic information can be incorporated by letting the mode vary according to the given probability distribution [dKW89]. The diagnostic engine computes the most probable diagnostic hypothesis, given observations about the current state of the system.

These examples illustrate the modes of contextual reasoning that interest us. In the first example, we are interested in exploring whether the analyzed component c can substitute for the idealized behavior i in arbitrary program contexts; i.e. for some context $C[\cdot]$, does $C[c]$ continue to approximate $C[i]$. Similarly, in the second example, we are looking to see the extent to which systems with similar failure behaviors are inter-substitutable. Such a question perforce generalizes the study of congruences elaborated by the theory of concurrency. The theory of concurrency performs a study of "exactly inter-substitutable" processes with temporal behavior. In the probabilistic context, the extant notions of bisimulation (or any process equivalence for that matter) are too sensitive to the probabilities; a slight perturbation of the probabilities would make two systems non-bisimilar. The examples motivate a shift to the study of the more robust notion of "approximately inter-substitutable".

The next example illustrates a deeper interaction of the temporal and probabilistic behavior of processes.

Example 1.3 Consider a producer and a consumer process connected by a buffer, where the producer is, say, a model of a network. Examples of this kind are studied extensively in the performance modelling of systems. In a model of such a system, probability serves to abstract the details of the producer (resp. consumer) process by considering rates of production (resp. consumption) of data based on empirical information. This model can be analyzed to calculate the number of packets lost as a function of the probabilities and the buffer size. The analysis aids in tuning system parameters, e.g. to optimize the buffer size. These studies are often couched in terms of asymptotic/stationary behavior to abstract over the transient behavior associated with system initialization (such as large bursts of communication) evident when the system begins execution.

Such examples motivate the study of equality notions based on "eventually approximately inter-substitutable" processes.

1.1 Our results

Partial labelled Markov processes are the probabilistic analogs of labelled transition systems; they have state spaces that might be continuous. In this model "internal choice" is modelled probabilistically and the so-called "external choice" is modelled by the indeterminate actions of the

environment. The starting point of our investigation is the study of strong bisimulation for processes. This study was initiated by Larsen and Skou [LS91] for *discrete* processes in a style similar to the queueing theory notion of "lumpability." This theory has been extended to continuous-state spaces and continuous distributions [DEP98, DEP02]. These papers provided a characterization of bisimulation using a negation-free logic $\mathcal{L}$.

In the context of the earlier discussion, we note that probabilistic bisimulation is too "exact" for our purposes — intuitively, two states are bisimilar only if the probabilities of outgoing transitions match exactly, motivating the search for a relaxation of the notion of equivalence of probabilistic processes. Giacalone, Jou and Smolka [GJS90] note that the idea of saying that "processes that are close should have probabilities that are close" does not yield a transitive relation, as illustrated by an example due to van Breugel [vBa]. This leads them to propose that the correct formulation of the "nearness" notion is via a metric.

A metric d is a function that yields a real number distance for each pair of processes. It should satisfy the usual metric conditions: $d(\mathcal{P}, \mathcal{Q}) = 0$ implies $\mathcal{P}$ is bisimilar to $\mathcal{Q}$[1], $d(\mathcal{P}, \mathcal{Q}) = d(\mathcal{Q}, \mathcal{P})$ and $d(\mathcal{P}, \mathcal{R}) \leq d(\mathcal{P}, \mathcal{Q}) + d(\mathcal{Q}, \mathcal{R})$. Inspired by the Kantorovich metric[2] on probability measures [Kan40, Hut81], we demand that d obey a certain "contractivity" property, an idea best conveyed via a concrete example.

Example 1.4 Consider the family of processes $\{\mathcal{P}_\epsilon \mid 0 \leq \epsilon < r\}$ where $\mathcal{P}_\epsilon = a_{r-\epsilon}.\mathcal{Q}$, i.e. $\mathcal{P}_\epsilon$ is the process that does an a with probability $r - \epsilon$ and then behaves like $\mathcal{Q}$. We demand that: $d(\mathcal{P}_{\epsilon_1}, \mathcal{P}_{\epsilon_2}) \leq |\epsilon_1 - \epsilon_2|$. This implies that $\mathcal{P}_\epsilon$ converges to $\mathcal{P}_0$ as ϵ tends to 0.

Metrics on processes The basic intuition behind our metrics is as follows. In view of our earlier results on the logical characterization of bisimulation, we know that if two processes are not bisimilar there will be a formula that distinguishes them. We measure the distance between processes in terms of the smallest formula required to distinguish them. If the formula is very large then only a long sequence of observations will distinguish the processes. This view, as stated, does not take into account the fact that the processes might differ immediately but do so with probabilities that are very close. Thus we need some quantitative analogue of the notion of logical formula. Our technical development of these intuitions is based on an idea expounded by Kozen [Koz85] to generalize logic to handle probabilistic phenomena.

Classical logic	Generalization
Truth values $\{0, 1\}$	Interval $[0, 1]$
Propositional function	Measurable function
State	Measure
The satisfaction relation $\models$	Integration $\int$

Just as the satisfaction relation, $\models$, links states and formulas to give truth values so the integral links measures (generalized states) with measurable functions (generalized formulas) to give real numbers (generalized truth values).

Following these intuitions, we consider a class $\mathcal{F}$ of functions that assign a value in the interval $[0, 1]$ to states of a process. These functions are inspired by the formulas of $\mathcal{L}$ — the result of

[1] Actually this is a pseudo-metric, in a metric we would have to insist that $d(\mathcal{P}, \mathcal{Q}) = 0$ implies that $\mathcal{P} = \mathcal{Q}$. On the bisimulation equivalence classes we have a metric. We will continue to say "metric" in this paper rather than the more accurate "pseudo-metric."

[2] In previous drafts we called this the "Hutchinson metric" but a careful historical search by Franck van Breugel reveals far earlier work by Kantorovich and also Vaserstein.

evaluating these functions at a state corresponds to a quantitative measure of the extent to which the state satisfies a formula of $\mathcal{L}$. The identification of this class of functions is a key contribution of this paper, and motivates a metric d:

$$d(\mathcal{P}, \mathcal{Q}) = \sup\{|f(p_0) - f(q_0)| \mid f \in \mathcal{F}\}.$$

In Section 5, we formalize the above intuitions to define a family of metrics $\{d^c \mid c \in (0, 1]\}$. These metrics support the spectrum of possibilities of relative weighting of the two factors that contribute to the distance between processes: the complexity of the functions distinguishing them versus the amount by which each function distinguishes them. The metric d^1 captures only the differences in the probabilities; probability differences at the first transition are treated on par with probability differences that arise very deep in the evolution of the process. In contrast, d^c for $c < 1$ give more weight to the probability differences that arise earlier in the evolution of the process, i.e. differences identified by simpler functions. As c approaches 0, the future gets discounted more.

As is usual with metrics, the actual numerical values of the metric are less important than properties like the significance of zero distance, relative distance of processes, contractivity and the notion of convergence[3].

Example 1.5 Consider the process $\mathcal{P}$ with two states, and a transition going from the start state to the other state with probability p. Let $\mathcal{Q}$ be a similar process, with the probability q. Then in Section 5, we show that $d^c(\mathcal{P}, \mathcal{Q}) = c|p - q|$. Now if we consider $\mathcal{P}'$ with a new start state, which makes a b transition to $\mathcal{P}$ with probability 1, and similarly $\mathcal{Q}'$ whose start state transitions to $\mathcal{Q}$ on $\mathbf{b}$ with probability 1, then $d^c(\mathcal{P}', \mathcal{Q}') = c^2|p - q|$, showing that the next step is discounted by c.

Each of these metrics agree with bisimulation:

$$d^c(\mathcal{P}, \mathcal{Q}) = 0, \text{ iff } \mathcal{P} \text{ and } \mathcal{Q} \text{ are bisimilar.}$$

For $c < 1$, we show how to compute $d^c(\mathcal{P}, \mathcal{Q})$ to within ϵ.

An "asymptotic" metric on processes. The d^c metric (for $c < 1$) is more heavily influenced by the initial transitions of a process — processes which can be differentiated early are far apart. For each $c \in (0, 1]$, we define a dual metric d^c_∞ (Section 8) on processes to capture the idea that processes are close if they have the same behavior "eventually", thus disregarding their initial behavior. Informally, we proceed as follows. Let $\mathcal{P}$ **after** s stand for the process $\mathcal{P}$ after exhibiting a trace s; of course this is not uniquely defined. Then, the j'th distance d^c_j between $\mathcal{P}, \mathcal{Q}$ after exhibiting traces of length j is defined by

$$\sup\{d^c(\mathcal{P} \textbf{ after } s, \mathcal{Q} \textbf{ after } s) \mid length(s) = j\}$$

where the sup is computed over all possible processes that might result after $\mathcal{P}$ and $\mathcal{Q}$ have exhibited the trace s. The asymptotic distance between $\mathcal{P}, \mathcal{Q}$ is given by the appropriate limit of the d^c_j's:

$$d^c_\infty(\mathcal{P}, \mathcal{Q}) = \limsup_{i \to \infty} \sup_{j > i} d^c_j(\mathcal{P}, \mathcal{Q}).$$

[3]See, however, the recent paper by van Breugel and Worrell [vBW01b] for a contrasting view.

A process algebra of probabilistically determinate processes In order to illustrate the properties of the metrics via concrete examples, we use an algebra of probabilistically determinate processes and a (bounded) buffer example coded in the algebra (Section 7). This process algebra has input and output prefixing, parallel composition and a probabilistic choice combinator. We do not consider hiding since this paper focuses on strong (as opposed to weak) probabilistic bisimulation.

We show that bisimulation is a congruence for all these operations. Furthermore, we generalize the result that bisimulation is a congruence, by showing that process combinators do not increase distance in any of the d^c metrics. Formally, let $d^c(\mathcal{P}_i, \mathcal{Q}_i) = \epsilon_i$. For every n-ary process combinator $C[X_1, \ldots, X_n]$, we have

$$d^c(C(\mathcal{P}_1, \ldots, \mathcal{P}_n), C(\mathcal{Q}_1, \ldots, \mathcal{Q}_n)) \leq \sum_i \epsilon_i.$$

Prefixing and parallel composition combinators do not increase d^c_∞. However, the probabilistic choice combinator is not contractive for d^c_∞.

Organization of this paper The rest of this paper is organized as follows. First, in Section 2, we review the notions of process and probabilistic bisimulation and associated results to make the paper self-contained. We next present (Section 4) an alternate way to study processes using real-valued functions and show that this view presents an alternate characterization of probabilistic bisimulation. In Section 5, we define a family of metrics and illustrate with various examples. The following section 7 describes a process algebra of probabilistically determinate processes. We conclude with a section 8 on the asymptotic metric.

2 Background

This section on background recalls definitions from previous work on partial labelled Markov processes [BDEP97, DEP98, LS91] and sets up the basic notations and framework for the rest of the paper.

We define discrete and continuous processes separately. A reader interested only in the discrete case can safely skip the section on continuous systems though the proofs are usually carried out for the general case of continuous processes. The first section recalls the definition of labelled Markov chains and bisimulation for them. We then give the extension of these definitions to continuous state-space systems and finally we recall the logic and the logical characterization of bisimulation.

2.1 Labelled Markov chains

Definition 2.1 *A* labelled Markov chain *with a label set $\mathcal{A}$ is a structure $(S, s_0, \{\tau_a \mid a \in \mathcal{A}\})$, where S is a countable set of states, s_0 is the start state, and $\forall a \in \mathcal{A}. \tau_a : S \times S \longrightarrow [0,1]$ is a transition function such that $\forall s \in S. \sum_t \tau_a(s,t) \leq 1$.*

A labelled Markov chain is finite if S is finite. There is no finite branching restriction; $\tau_a(s,t)$ can be non-zero for countably many t's. τ_a is extended to a function $S \times \mathcal{P}(S) \longrightarrow [0,1]$ by defining: $\tau_a(s,X) = \sum_{t \in X} \tau_a(s,t)$. We will assume a fixed finite set of labels $\mathcal{A}$. We will often use the expression *finite process* to mean labelled Markov chain.

We could have alternatively presented a labelled Markov chain as a structure $(S, \mu, \{\tau_a \mid a \in \mathcal{A}\})$ where μ is an initial distribution on S. Given a labelled Markov chain with initial distribution μ, one can construct an essentially equivalent[4]labelled Markov chain $\mathcal{S}'$ with initial state s'_0 as follows.

[4]We do *not* mean bisimilar.

$S' = S \cup \{s'_0\}$ where s'_0 is a new state not in S; $\tau'_a(s', t') = \tau_a(s', t')$ if $s', t' \in S$; $\tau'_a(s', s'_0) = 0$, and $\tau'_a(s'_0, t') = \sum \tau_a(s', t')\mu(s')$.

We will freely move between the notions of initial state and initial distribution. For example, when a transition on label l occurs in a labelled Markov chain $\mathcal{S}$, there is a new initial distribution given by $\mu'(t) = \sum \tau_a(s, t) \times \mu(s)$.

We recall the definition of bisimulation on labelled Markov chains from [LS91]. This notion captures the idea that processes are equivalent if they react exactly the same way to all external interactions in terms of accepting or rejecting actions. Thus we do not see the "internal dynamics", i.e. the state transitions but we do see whether an action is accepted or rejected and with what probability.

Definition 2.2 *A **bisimulation** between two processes $\mathcal{S}$ and $\mathcal{S}'$ is an equivalence relation R, on $S \uplus S'$ such that whenever two states $s \in S$ and $s' \in S'$ are R-related, then for any label a and any R-equivalence class of states $C \subseteq S \cup S'$, $\tau_a(s, C) = \tau_a(s', C)$.*

Two states are bisimilar if they are related by a bisimulation relation. We say that $\mathcal{S}$ and $\mathcal{S}'$ are bisimilar if their initial states are.

2.2 Continuous processes

We now turn to the general case of continuous processes. A labelled Markov process is a labelled Markov chain with a continuous state space.

The extension to continuous state systems introduces some measure-theoretic subtleties. For instance, we cannot ask for the transition probability to any set of states — we need to restrict ourselves to measurable sets. In fact we need to assume metric space structure on the state space. The classical theory of Markov processes is typically carried out in the setting of Polish spaces rather than on abstract measure spaces. We work with *analytic spaces* which generalize Polish spaces.

Definition 2.3 *A **labelled Markov process** $\mathcal{S}$ with label set $\mathcal{A}$ is a structure $(S, s_0, \Sigma, \{\tau_a \mid a \in \mathcal{A}\})$, where S is the set of states, s_0 is the initial state, and Σ is the Borel σ-field on S, and*

$$\forall a \in \mathcal{A}, \tau_a : S \times \Sigma \longrightarrow [0, 1]$$

is a transition sub-probability function, i.e., the set function $\tau_a(s, \cdot)$ is a (sub-)probability measure for each fixed $s \in S$, and for each fixed $X \in \Sigma$ the function $\tau(\cdot, X)$ is a measurable function.

One interprets $\tau(s, X)$ as the probability of the process starting in state s making a transition into one of the states in X. The transition probability is a *conditional probability*; it gives the probability of the process being in one of the states of the set X after the transition, *given* that it was in the state s 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 always consider the time-independent case.

We will work with *sub-probability* functions; i.e. with functions where $\tau(s, S) \leq 1$ rather than $\tau(s, S) = 1$. The mathematical results go through in this extended case. We view processes where the transition functions are only sub-probabilities as being *partially defined*. The stochastic systems studied in the literature are usually only the very special version where $\tau(s, S)$ is either 1 or 0. We call such processes *total* and the general processes are called *partial*. We capture the idea that an action is rejected by setting $\tau(s, S)$ to be 0. We will fix the label set to be $\mathcal{A}$ once and for all. The resulting theory is not seriously restricted by this. We will write (S, s_0, Σ, τ) for labelled Markov processes, instead of the more precise $(S, s_0, \Sigma, \{\tau_a \mid a \in \mathcal{A}\})$.

Example 2.4 We give a simple example, taken from [DEP02], to illustrate the ideas. Consider a process with two labels $\{a, b\}$. The state space is the real plane, $\mathbb{R}^2$. When the process 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 process 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. If there were just a single label we would have an ordinary (time-independent) Markov process.

The fundamental process equivalence that we consider is *strong probabilistic bisimulation* or just "bisimulation" for the present paper. Probabilistic bisimulation means matching the moves and probabilities — thus each system must be able to make the same transitions with the same probabilities as the other. The definition that we use is an adaptation of the definition presented in the previous section. In an earlier paper [BDEP97] we had introduced a version of this definition based on categorical ideas but in the present paper we use a version much closer in form to that of Larsen and Skou that we introduced in [DGJP00]. We also recapitulate the result on logical characterization of bisimulation.

Let R be a relation on a set S. We say a set $X \subseteq S$ is R-closed if $R(X) = \{t | \exists s \in X, sRt\}$ is a subset of X. If R is reflexive, this becomes $R(X) = X$.

Definition 2.5 *A **bisimulation relation** between two labelled Markov processes $\mathcal{S} = (S, s_0, \Sigma, \tau)$ and $\mathcal{S}' = (S', s_0', \Sigma', \tau')$ is an equivalence relation R on $S \uplus S'$ such that, for $s \in S$ and $s' \in S'$, with sRs', for every R-closed set $X \subseteq S \uplus S'$ such that $X \cap S \in \Sigma$ and $X \cap S' \in \Sigma'$, we have*

$$\tau_a(s, X \cap S) = \tau_a'(s', X \cap S')$$

for every $a \in \mathcal{A}$. Two states are bisimilar if they are related by a bisimulation relation. We say that $\mathcal{S}$ and $\mathcal{S}'$ are bisimilar if their initial states are.

The intuition behind this definition is that the relation R relates those states that can be "lumped" together. Bisimulation is obviously reflexive and symmetric. The logical characterization of bisimulation shows that it is transitive.

2.3 Logical Characterization of Bisimulation

One can define a simple modal logic and prove that two states are bisimilar if and only if they satisfy exactly the same formulas. Indeed for finite-state processes one can decide whether two states are bisimilar and effectively construct a distinguishing formula in case they are not [DEP98, DEP02].

As before we assume that there is a fixed set of "actions" $\mathcal{A}$. The logic is called $\mathcal{L}$ and has the following syntax:

$$\mathcal{L} := \mathsf{T} \mid \phi_1 \wedge \phi_2 \mid \langle a \rangle_q \phi$$

where a is an action and q is a rational number. This is the basic logic with which we establish the logical characterization. We introduce also a logic with disjunction, $\mathcal{L}_\vee$, and its infinite version, $\mathcal{L}_{\bigvee}$:

$$\mathcal{L}_\vee := \mathcal{L} \mid \phi_1 \vee \phi_2$$

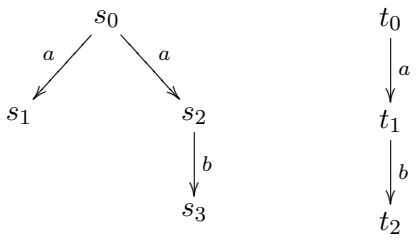


Figure 1: Two processes which cannot be distinguished without negation in HML.

$$\mathcal{L}_{\bigvee} := \mathcal{L} \mid \bigvee_{i=1}^{\infty} \phi_i.$$

Given a labelled Markov process $\mathcal{S} = (S, s_0, \Sigma, \tau)$ we write $s \models_{\mathcal{S}} \phi$ to mean that the state s satisfies the formula ϕ in $\mathcal{S}$, and we write $[\![\phi]\!]_{\mathcal{S}}$ for the set of states that satisfy ϕ. The definition of the relation $\models$ is given by induction on formulas – and one can prove along the way that $[\![\phi]\!]_{\mathcal{S}}$ is always in Σ. The definition is obvious for the propositional constant T, conjunction and disjunction. We say $s \models \langle a \rangle_q \phi$ if and only if $\tau_a(s, [\![\phi]\!]_{\mathcal{S}}) > q$. In other words, the process in state s can make an a-move to a state, that satisfies ϕ, with probability strictly greater than q[5]. We often omit the subscript in $[\![\phi]\!]_{\mathcal{S}}$ and $\models_{\mathcal{S}}$ when no confusion can arise.

The following example helps to illustrate some of the key aspects of the logic.

Example 2.6 (Example from [DEP98]) Consider the processes shown in Figure 1. They are both nonprobabilistic processes. It is well known that they cannot be distinguished by a negation-free formula of Hennessy-Milner logic; the process on the left satisfies $\langle a \rangle \neg \langle b \rangle \mathsf{T}$ while the process on the right does not. However, for no assignment of probabilities are the two processes going to be bisimilar. Suppose that the two a-labelled branches of the left hand process are given probabilities p and q, assume that the b-labelled transitions have probability 1. Now if the right hand process has its a-labelled transition given a probability anything other than $p + q$, say $r > p + q$ we can immediately distinguish the two processes by the formula $\langle a \rangle_{p+q} \mathsf{T}$ which will not be satisfied by the left hand process. If $r = p + q$ then we can use the formula $\langle a \rangle_{r'} \langle b \rangle_{1/2} \mathsf{T}$, where $q < r' < r$. The left hand process cannot satisfy this formula but the right hand one does unless $p = 0$ in which case the processes are bisimilar.

The logic that Larsen and Skou used in [LS91] has more constructs than $\mathcal{L}$, including disjunction and some negative constructs. They show that for finitely branching systems[6], two *states* of the same process are bisimilar if and only if they satisfy the same formulas of their logic.

The main theorem relating the logic and bisimulation is the following. This was proved in [DEP98, DEP02] for the categorical presentation of bisimulation and in [DGJP00] for the relational presentation.

Theorem 2.7 *Let (S, s_0, Σ, τ) be a labelled Markov process. Two states $s, s' \in S$ are bisimilar if and only if they satisfy the same formulas of $\mathcal{L}$.*

A corollary to this theorem is that bisimulation is an equivalence relation.

[5] In our earlier work we had used $\geq$ instead of $>$.

[6] They actually use a stronger property, the "minimum deviation condition" which uniformly bounds the degree of branching everywhere.

The fact that a logic without negation and without infinitary conjunction is sufficient for processes with infinite branching was somewhat of a surprise based on what we expect from the nonprobabilistic case. It is even more surprising that this logical characterization goes through even in the continuous case.

We now prove that every formula satisfied in a state of a countable process is witnessed by a finite sub-process.

Definition 2.8 *$\mathcal{S} = (S, s_0, \Sigma, \tau)$ is a sub-process of $\mathcal{S}' = (S', s_0', \Sigma', \tau')$ if $(S, \Sigma) \subseteq (S', \Sigma')$ (this means that the inclusion map $S \subseteq S'$ is measurable), $s_0 = s_0'$ and for every $a \in \mathcal{A}$, $s \in S$, $X \in \Sigma$ we have $\tau_a(s, X) \leq \tau_a'(s, X)$.*

Thus, a sub-process has fewer states and lower probabilities than the original process.

Lemma 2.9 *Let $\mathcal{P}$ be a labelled Markov chain, $p \in P$ and $\phi \in \mathcal{L}_\vee$ such that $p \models_{\mathcal{P}} \phi$. Then there exists a finite sub-process of $\mathcal{P}$, $\mathcal{Q}_\phi^p$, such that $p \in Q_\phi^p$, and $p \models_{\mathcal{Q}_\phi^p} \phi$.*

Proof The proof is by induction on ϕ. For T, the one state process containing p suffices. For $\phi = \phi_1 \wedge \phi_2$, we take the union of the finite processes, $\mathcal{Q}_{\phi_1}^p, \mathcal{Q}_{\phi_2}^p$ given by the induction hypothesis, which ensures that $p \models_{\mathcal{Q}_\phi^p} \phi_1 \wedge \phi_2$. For disjunction, $\bigvee_{i=1}^\infty \phi_i$, we take $\mathcal{Q}_{\phi_1}^p$ (or any other $\mathcal{Q}_{\phi_i}^p$).

Let $p \models_{\mathcal{P}} \langle a \rangle_r \psi$. Then, since $\tau_a(p, [\![\psi]\!]_{\mathcal{P}}) > r$, there is a finite subset $U = \{p_1, \ldots, p_n\} \subseteq [\![\psi]\!]_{\mathcal{P}}$, such that $\tau_a(p, U) > r$. The required finite process, $\mathcal{Q}_{\langle a \rangle_r . \psi}^p$ is now constructed by taking the unions of the finite processes, $\mathcal{Q}_\psi^{p_1}, \ldots, \mathcal{Q}_\psi^{p_n}$, adding state p and the transitions from p to p_i for $i = 1 \ldots n$. ∎

2.4 Simulation and Strict Simulation

This discussion of this subsection is taken from our paper on approximation [DGJP00, DGJP03]. The notion of simulation is the natural one-directional version of the definition of bisimulation. Normally, the fact that the definition of bisimulation is coinductive means that, in general, two-way simulation is not bisimulation. However, in the case of our reactive systems, two-way simulation *is* bisimulation; this is in contrast with the usual situation with indeterminate processes. Furthermore, when we make contact with domain-theoretic ideas the notion of simulation will correspond to the domain ordering. Thus when we say $\mathcal{S}$ *approximates* $\mathcal{S}'$ we mean that $\mathcal{S}'$ simulates $\mathcal{S}$. We also introduce a concept called *strict simulation* which will correspond to the "way-below" relation.

Our definition of simulation follows [Des99b, DGJP03].

Definition 2.10 *Let $\mathcal{S} = (S, i, \Sigma, \tau)$ be a labelled Markov process. A reflexive and transitive relation (a preorder) R on S is a **simulation** if whenever sRs', with $s, s' \in S$, we have that for all $a \in \mathcal{A}$ and every R-closed measurable set X, $\tau_a(s, X) \leq \tau_a(s', X)$. We say s is simulated by s' if sRs' for some simulation relation R.*

*R is a **strict simulation** if there is an $\epsilon > 0$ such that for all R-closed $X \in \Sigma$, we have $\tau_a(s, X) < \tau_a(s', X) - \epsilon$ whenever $\tau_a(s, X) > 0$. If we wish to emphasize the role of ϵ we will say that R is an ϵ-**strict** simulation, and we will write R_ϵ. We say s is simulated (strictly simulated) by s' if sRs' for some simulation (resp. strict simulation) relation R.*

Let $\mathcal{S} = (S, i, \Sigma, \tau)$ and $\mathcal{S}' = (S', i', \Sigma', \tau')$ labelled Markov processes. $\mathcal{S}$ is simulated (strictly simulated) by $\mathcal{S}$ if there is a simulation (resp. strict simulation) relation on some process $\mathcal{U}$ of which $\mathcal{S}$ and $\mathcal{S}'$ are direct summands, relating i and i' in $\mathcal{U}$.

Alternately, simulation on the states of a labelled Markov process can be viewed as the maximum fixed point of the following (monotone) functional G on the lattice of preorders on $(S \times S, \subseteq)$, defined as follows:

$$s\, G(R)\, t \text{ if for all } a \in \mathcal{A}, \text{ for all } R\text{-closed } C \in \Sigma, \tau_a(s, C) = \tau_a(t, C)$$

We do not require $\mathcal{U}$ to be exactly $\mathcal{S} + \mathcal{S}'$ but rather a sum of a number of processes, of which are $\mathcal{S}$ and $\mathcal{S}'$. The proof of transitivity of simulation (resp. strict simulation) follows the transitivity proof for bisimulation.

The logic $\mathcal{L}_\bigvee$ - since it does not have negation - satisfies a basic monotonicity property with respect to simulation.

Proposition 2.11 *If s is simulated (or strictly simulated) by s', then for all formulas $\phi \in \mathcal{L}_\bigvee$, $s \models \phi$ implies $s' \models \phi$.*

The converse of this result is also true and was proven in [DGJP00].

3 Approximating Labelled Markov Processes

In a recent paper [DGJP00, DGJP03] we developed a theory of approximation of labelled Markov processes. Using this theory we can extend our results about the metric for the discrete case reported in [DGJP99] to the continuous case. Defining the metrics for processes that have continuous state spaces does not require the approximation theory, however showing certain properties of a class of functions (that they characterize bisimulation) does use the approximation theory to lift the result from the discrete case to the continuous case. Readers who are not interested in all the fine points of the continuous case can skip this section without jeopardizing their understanding of the rest of the paper.

In [DGJP00, DGJP03] we give an explicit concrete construction of the approximants and also a deeper domain-theoretic analysis of the notion of the approximants. In the present summary we will entirely skip the domain theory and talk about the explicit construction only. In a short section after the introduction of the metric we show that these approximants converge to the original process being approximated in the metrics of the present paper. In the approximants paper we justified the notion of approximation by establishing that the approximants form a directed set in a suitable ordering and the sup of the approximants gives back the labelled Markov process being approximated.

The key tool in our analysis is the construction of some approximants via an "unfolding" construction. As the approximation is refined there are more and more transitions possible. There are two parameters to the approximation, one is a natural number n, and the other is a positive rational ϵ. The number n gives the number of successive transitions possible from the start state. The number ϵ measures the accuracy with which the probabilities approximate the transition probabilities of the original process.

Given a labelled Markov process $\mathcal{S} = (S, i, \Sigma, \tau)$, an integer n and a rational number $\epsilon > 0$, we define $\mathcal{S}(n, \epsilon)$ to be an n-step unfolding approximation of $\mathcal{S}$. Its state-space is divided into $n + 1$ levels which are numbered $0, 1, \ldots, n$. At each level, say n, the states of the approximant are the elements of a partition of S; these partitions correspond to the equivalence classes corresponding to the level n approximation to bisimulation. The initial state of $\mathcal{S}(n, \epsilon)$ is at level n and transitions only occur between a state of one level to a state of one lower level. Thus, in particular, states of level 0 have no outgoing transitions. In the following we omit the curly brackets around singletons.

Definition 3.1 *Let (S, i, Σ, τ) be a labelled Markov process, $n \in \mathbf{N}$ and ϵ a positive rational. We denote the finite-state approximation by $\mathcal{S}(n, \epsilon) = (P, p_0, \rho)$ where P is a subset of $\Sigma \times \{0, \ldots, n\}$. It is defined as follows, for $n \in \mathbf{N}$ and $\epsilon > 0$. $\mathcal{S}(n, \epsilon)$ has $n+1$ levels. States are defined by induction on their level. Level 0 has one state $(S, 0)$. Now, given the sets from level l, we define states of level $l+1$ as follows. Suppose that there are m states at level l, we partition the interval $[0, 1]$ into intervals of size ϵ/m. Let $(B_j)_{j \in I}$ stand for this partition; i.e. for $\{\{0\}, (0, \epsilon/m], (\epsilon/m, 2\epsilon/m], \ldots\}$. States of level $l+1$ are obtained by the partition of S that is generated by the sets $\tau_a(\cdot, C)^{-1}(B_j)$, for every set C corresponding to state at level l and every label $a \in \{a_1, \ldots, a_n\}$, $i \in I$. Thus, if a set X is in this partition of S, $(X, l+1)$ is a state of level $l+1$. Transitions can happen from a state of level $l+1$ to a state of level l, and the transition probability function is given by*

$$\rho_a((X,k),(B,l)) = \begin{cases} \inf_{t \in X} \tau_a(t, B)\big) & \text{if } k = l+1, \\ 0 & \text{otherwise.} \end{cases}$$

The initial state p_0 of $\mathcal{S}(n, \epsilon)$ is the unique state (X, n) such that X contains i, the initial state of $\mathcal{S}$.

If $B = \cup B_j$, is a (finite and disjoint) union of sets at the same level (i.e. $(B_j, l) \in \mathcal{S}(n, \epsilon)$), we will write $\rho_a((X, l+1), (B, l))$ to mean $\sum_{j \in I} \rho_a((X, l+1), (B_j, l))$. If $s \in S$, we denote by (X_s, l) the unique state at level l such that $s \in X_s$.

Proposition 3.2 ([DGJP03]) *Every labelled Markov process $\mathcal{S}$ simulates all its approximants $\mathcal{S}(n, \epsilon)$. More precisely, every state (X, l) of $\mathcal{S}(n, \epsilon)$ $(l \leq n)$ is simulated by every state $s \in X$ of $\mathcal{S}$.*

The next theorem is a key result about our concrete approximations.

Theorem 3.3 *If a state $s \in S$ satisfies a formula $\phi \in \mathcal{L}_\vee$, then there is some approximation $\mathcal{S}(n, \epsilon)$ such that $(X_s, n) \models \phi$.*

This result can be used to prove that the space of all labelled Markov processes has a countable subset, the rational trees, which serves to approximate all labelled Markov processes. For brevity we will just say "rational tree" when we mean a finite-state process with a tree-like transition graph and rational transition probabilities. Moreover, this countable family of rational trees capture all properties of $\mathcal{L}_\bigvee$ of labelled Markov processes

Theorem 3.4 ([DGJP03]) *Given any labelled Markov process $\mathcal{S}$ there is a directed set of rational trees $\mathcal{T}_i$ with each $\mathcal{T}_i$ being strictly simulated by $\mathcal{S}$ and such that any logical formula satisfied by $\mathcal{S}$ is satisfied by some $\mathcal{T}_i$.*

4 A real-valued logic on labelled Markov processes

In this section, following Kozen [Koz85], we present an alternate characterization of probabilistic bisimulation using functions into the reals instead of the logic $\mathcal{L}$. We define a set of functions which are sufficient to characterize bisimulation. It is worth clarifying our terminology here. We define a set of *functional expressions* by giving an explicit syntax. A functional expression becomes a function when we interpret it in a system. Thus we may loosely say "the same function" when we move from one system to another. What we really mean is the "same functional expression"; obviously it cannot be the same function when the domains are different. This is no different from having syntactically defined formulas of some logic which become boolean-valued functions when they are interpreted on a structure.

Definition 4.1 *For each $c \in (0,1]$, we consider a family $\mathcal{F}^c$ of functional expressions generated by the following grammar.*

$$f := \mathbf{1} \mid \mathbf{1} - f \mid \langle a \rangle f \mid \min(f_1, f_2) \mid \sup_{i \in \mathbf{N}} f_i \mid f \ominus q,$$

where q is a rational. $\mathcal{F}^c_+$ is the sub-collection of $\mathcal{F}^c$ that does not use the negation functional $\mathbf{1} - f$ and $\mathcal{F}in^c_+$ is the sub-collection of $\mathcal{F}^c_+$ that uses finite sup*.*

The interpretation is as follows. Let $\mathcal{S} = (S, s_0, \Sigma, \tau)$ be a labelled Markov process. We write $f_{\mathcal{S}} : S \longrightarrow [0,1]$ for the interpretation of $f \in \mathcal{F}^c$ on $\mathcal{S}$ and drop the subscript when no confusion can arise. Let $s \in S$. Then

$$\begin{aligned}
&\mathbf{1}(s) = 1,\\
&(\mathbf{1} - f)(s) = 1 - f(s),\\
&\langle a \rangle f(s) = c \int_S f(t) \tau_a(s, dt),\\
&(f \ominus q)(s) = \max(f(s) - q, 0),
\end{aligned}$$

and min *and* sup *are defined in the obvious way.*

In the interpretation of $\langle a \rangle f$, the c refers to the constant in $\mathcal{F}^c$; this is the only place where an explicit mention of c occurs. It is useful to emphasis concisely the semantics we want to use, since we never work with functionals that have different parameters. The role of c is to discount the effect of future actions. For $c = 1$ all transitions are counted equally even if they are far into the future.

Note that in [DGJP99], $f \ominus q$ was written $\lfloor f \rfloor_q$ and that we had an additional functional written $\lceil f \rceil^q = \min(f, q)$. The latter is not necessary since it can be represented by using the functional min and the constant function $q := \mathbf{1} \ominus (1 - q)$. We use $\langle a \rangle^n f$ to represent $\langle a \rangle \cdots \langle a \rangle f$ where $\langle a \rangle$ appears n times.

One can informally associate functional expressions with every connective of the logic $\mathcal{L}$ in the following way. T is represented by the functional $\mathbf{1}$ and conjunction by min. The contents of the connective $\langle a \rangle_q$ is split up into two expressions: $\langle a \rangle f$, which intuitively corresponds to prefixing, and $f \ominus q$, which captures the "greater than q" idea.

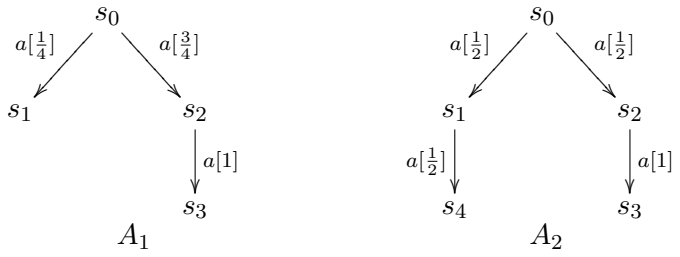


Figure 2: Labelled Markov chains

Example 4.2 Consider the finite processes A_1 and A_2 of Figure 2. The functional expression $(\langle a \rangle \mathbf{1})$ of $\mathcal{F}^c$ evaluates to c at states s_0, s_2 of both A_1 and A_2; it evaluates to 0 at states s_1, s_3 of A_1 and s_3, s_4 of A_2, and it evaluates to $c/2$ at state s_1 of A_2. The functional expression $(\langle a \rangle . \langle a \rangle \mathbf{1})$

evaluates to $3c^2/4$ at states s_0 of A_1, A_2 and to 0 elsewhere. The functional expression $(\langle a\rangle(\langle a\rangle \mathbf{1} \ominus \frac{c}{2}))$ evaluates to $3c^2/8$ at state s_0 of A_1 and to $c^2/4$ at state s_0 of A_2. This example shows the need for the connective $\ominus$ in the functional expressions. Without it there would be no way of distinguishing these two. Note, however, that this example relies on the fact that one can have subprobability distributions associated with a labelled transition. If we insisted that all probability distributions had to be normalized then functional expressions without $\ominus$ would suffice[dA].

A_3 A_4

Figure 3: Labelled Markov chains

Example 4.3 Consider the finite process A_3 of Figure 3 and functionals of $\mathcal{F}^c$. A functional expression of the form $(\langle a\rangle^n.\mathbf{1})$ evaluates to c^n at state s_0. On state s_0 of process A_4 the same functional expression evaluates to $(c \times 0.4)^n$.

A routine induction on the structure of the functional expression $f \in \mathcal{F}^c_+$, shows:

Lemma 4.4 *If $\mathcal{S}$ is simulated by $\mathcal{S}'$, then $\forall s, s'$ such that s and s' are related by the simulation relation we have*

$$(\forall f \in \mathcal{F}^c_+)\ [f_{\mathcal{S}}(s) \leq\ f_{\mathcal{S}'}(s')].$$

The next several lemmas and their corollaries - from 4.5 to 4.8 - are aimed at proving that the functional expressions characterize bisimulation. The proof below uses our earlier results [DEP98] on logical characterization of bisimulation. It is also possible to proceed directly, essentially using the same techniques readapted to functional expressions. However, we will not pursue such a proof here because the proof below also shows how the functional expressions and the logical formulas are related.

For any finite process $\mathcal{P}$ and any formula, there is a functional from $\mathcal{F}^c_+$ which distinguishes between states of $\mathcal{P}$ that do or do not satisfy the formula. This functional furthermore gives a zero value to any state of any process that does not satisfy the formula.

Lemma 4.5 *Given $\phi \in \mathcal{L}_\bigvee$, a finite process $\mathcal{P}$, and $c \in (0, 1]$, there is a functional expression $f \in \mathcal{F}^c_+$ such that*

1. *$\forall p \in P$ we have $f_{\mathcal{P}}(p) > 0$ iff $p \models_{\mathcal{P}} \phi$;*
2. *for any state s of any labelled Markov process $\mathcal{S}$, we have $f_{\mathcal{S}}(s) > 0 \Rightarrow s \models_{\mathcal{S}} \phi$.*

Proof The proof is by induction on the structure of ϕ. The key case is $\phi = \langle a\rangle_q \psi$, let g be the functional expression corresponding to ψ yielded by induction. Let $x = \min\{g(t) \mid t \in [\![\psi]\!]_{\mathcal{P}}\}$. By induction hypothesis, $x > 0$. Recall that a constant function $1 - q$ on processes can be obtained with the functional $\mathbf{1} \ominus q$: consequently we can legitimately use the notation $\min(g, x)$ to mean $\min(g, \mathbf{1} \ominus (1 - x))$. Consider the functional expression f given by $(\langle a\rangle \min(g, x)) \ominus cxq$. For all $t \in [\![\psi]\!]_{\mathcal{P}}$, $\min(g, x)(t) = x$. Now for any state $p \in P$,

$$\langle a\rangle \min(g, x)(p) = cx \sum_{t \in [\![\psi]\!]_{\mathcal{P}}} \tau_a(p, t) = cx\tau_a(p, [\![\psi]\!]_{\mathcal{P}}).$$

Now the last expression is $> cxq$ if and only if $p \in [\![\langle a \rangle_q . \psi]\!]_{\mathcal{P}}$. Thus f satisfies the first condition.

The second condition holds because for any $s \in S$, $\langle a \rangle \min(g, x)(s) \leq cx\tau_a(s, [\![\psi]\!]_{\mathcal{S}})$, so if $s \not\models \phi$ then $\tau_a(s, [\![\psi]\!]_{\mathcal{S}}) \leq q$ and hence $f(s) = 0$. ∎

Note that if the formula is finite, then the corresponding functional lies in $\mathcal{F}in^c_+$. The previous lemma can be partially extended to arbitrary labelled Markov processes. In this case, the functional corresponding to a formula does not work for *every* state of the process. The functional will depend on the states and the formula must be finite. We give different proofs for the two cases.

Corollary 4.6 *Given $\phi \in \mathcal{L}_\vee$, a labelled Markov chain $\mathcal{P}$, $c \in (0, 1]$ and a state $p \in P$, if $p \models_{\mathcal{P}} \phi$, then there exists $f \in \mathcal{F}in^c_+$ such that*

1. *$f_{\mathcal{P}}(p) > 0$ and*
2. *for any state s of any labelled Markov process $\mathcal{S}$, we have $f_{\mathcal{S}}(s) > 0 \Rightarrow s \models_{\mathcal{S}} \phi$.*

Proof Let p be a state in P such that $p \models_{\mathcal{P}} \phi$. By Lemma 2.9, there is a finite sub-process $\mathcal{Q}^p_\phi$ of $\mathcal{P}$ such that $p \models_{\mathcal{Q}^p_\phi} \phi$. By Lemma 4.5, $\exists f \in \mathcal{F}in^c_+$ such that $f_{\mathcal{Q}^p_\phi}(p) > 0$ and for any process $\mathcal{S}$, $\forall s \in S$, $f_{\mathcal{S}}(s) > 0 \Rightarrow s \models \phi$. By Lemma 4.4, $f_{\mathcal{P}}(s) > f_{\mathcal{Q}^p_\phi}(s) > 0$, so f satisfies the conditions required by the lemma. ∎

Corollary 4.7 *Given $\phi \in \mathcal{L}_\vee$, a labelled Markov process $\mathcal{S}$, $c \in (0, 1]$ and a state $s \in S$, if $s \models \phi$, then there exists $f \in \mathcal{F}in^c_+$ such that*

1. *$f_{\mathcal{S}}(s) > 0$;*
2. *for any state s' of any other labelled Markov process $\mathcal{S}'$, we have $f_{\mathcal{S}'}(s') > 0 \Rightarrow s' \models \phi$.*

Proof Let $\mathcal{S}$ be an arbitrary process and $\phi \in \mathcal{L}_\vee$. Let s be a state in $\mathcal{S}$ such that $s \models \phi$. By Theorem 3.3 there is a finite approximation $\mathcal{P}$ of $\mathcal{S}$ and a state $p_s \in P$ such that $p_s \models \phi$. By Lemma 4.5, $\exists f \in \mathcal{F}in^c_+$ such that $f_{\mathcal{P}}(p_s)0$ and for any process $\mathcal{S}'$, $\forall s' \in S'.s' \not\models \phi \Rightarrow f_{\mathcal{S}'}(s') = 0$. By Lemma 4.4, $f_{\mathcal{S}}(s) > f_{\mathcal{P}}(p_s) > 0$, thus f satisfies above conditions 1. and 2. ∎

Corollary 4.8 *Given $\phi \in \mathcal{L}_\vee$ and $c \in (0, 1]$, there exists $f_\phi \in \mathcal{F}^c_+$ such that for every state s of any labelled Markov process $\mathcal{S}$,*

$$f_\phi(s) > 0 \Leftrightarrow s \models \phi.$$

Proof Recall from Theorem 3.4 that approximations can be replaced by a countable family of finite trees. Take the sup of the functions given by Lemma 4.5 corresponding to ϕ for the (countably many) rational trees. This function has the desired property for every state of the rational trees and hence, by Lemma 4.4, it also works for every state of every labelled Markov process. ∎

Example 4.9 f_ϕ satisfies:

- $f_{\mathsf{T}} = \mathbf{1}$
- For any state s in process $\mathcal{S}$, $f_{\langle a\rangle_q \mathsf{T}}(s) = \max(\tau_a(s,S) - q, 0)$.
- $f_{\phi\wedge\psi} = \min(f_\phi, f_\psi)$
- $f_{\phi\vee\psi} = \max(f_\phi, f_\psi)$

The next result says that functions are sound and complete for bisimulation.

Theorem 4.10 *For any labelled Markov processes $\mathcal{S}$, $\mathcal{S}'$, $\forall c \in (0,1]$,*

$$s \in S \text{ and } s' \in S' \text{ are bisimilar iff } (\forall f \in \mathcal{F}in^c_+)\ [f_{\mathcal{S}}(s) = f_{\mathcal{S}}(s')]$$

Note that the left-to-right direction is also true for any functional of $\mathcal{F}^c$ but $\mathcal{F}in^c_+$ is enough for the other direction.

Proof ($\Rightarrow$) : We show that for any bisimulation R, sRs' implies that $(\forall f \in \mathcal{F}^c)\ [f_{\mathcal{S}}(s) = f_{\mathcal{S}'}(s')]$. The proof proceeds by induction on the structure of the functional expression f. The key case is when f is of the form $(\langle a\rangle g)$. Then we would like to show that $\int_{t\in S} g(t)\tau_a(s,dt) = \int_{t\in S'} g(t)\tau'_a(s',dt)$. Consider any simple function h approximating g, with values $v_i, i = 1\ldots n$, defined by $h(s) = max\{v_i \mid v_i \leq g(s)\}$. Then the set $S_i = h^{-1}(v_i) \subseteq S \cup S'$ is measurable because it is $g^{-1}([v_i, v_{i+1}))$ and it is R-closed because if $t \in S_i$ and tRt' then by induction $g(t) = g(t')$, so $t' \in S_i$. Thus $\tau_a(s,S_i) = \tau'_a(s',S_i)$, which shows the result.

($\Leftarrow$) : Assume that s and s' are not bisimilar. Then there is a formula ϕ of $\mathcal{L}$ such that $s \models \phi$ and $s' \not\models \phi$ (or the converse). ByCorollary 4.7,there is a functional expression $f \in \mathcal{F}in^c_+$ such that $f_{\mathcal{S}}(s) > 0$ and $f_{\mathcal{S}'}(s') = 0$. ∎

Given that we now know that functional expressions characterize bisimulation and that logical formulas also characterize bisimulation we immediately get:

Corollary 4.11 *For any process $\mathcal{S}$, $(\forall c \in (0,1])$, $\forall s, s' \in S$*

$$[(\forall \phi \in \mathcal{L})\ s \models_{\mathcal{S}} \phi \Leftrightarrow s' \models_{\mathcal{S}'} \phi] \Leftrightarrow (\forall f \in \mathcal{F}^c)\ [f_{\mathcal{S}}(s) = f_{\mathcal{S}'}(s')].$$

Note that for the $\mathcal{L}$ sub-fragment of the logic, the resulting function is in $\mathcal{F}in^c_+$.

The following example shows that the conditional functional expressions are necessary.

Example 4.12 Consider the processes A_1, A_2 of Figure 2. The calculations of Example 4.2 show that the s_0 states of A_1, A_2 are distinguishable. Furthermore, the states are indistinguishable if we use only the functionals $\mathbf{1}, \mathbf{1} - f, \langle a\rangle f, \min(f_1, f_2), \sup_{i\in\mathbf{N}} f_i$. Thus, Example 4.2 shows that the functional expression $f \ominus q$ is indeed necessary.

So far we have shown that functional expressions are just as good for characterizing bisimulation as were logical formulas. We are now in a position to use the extra information in the functions to define a metric.

5 Metrics on Processes

In the present section we introduce the notion of metrics between processes. Intuitively the metrics measure how "visibly" different the processes are. In terms of logic one can say that two processes are very close if the formulas that tell them apart are very complex. To capture this intuition quantitatively we use the functions introduced in the last section. There is now a second notion of how far apart processes are; the distinguishing functions could have values which are very different or only slightly different. We actually study a family $\{d^c \mid c \in (0,1]$ of definitions which assign different weights to these differences[7]. The main results of this section are:

- We show that each $d^c, c \in (0,1]$ is a metric. In particular, processes at 0 distance are bisimilar. The finite representation results of [DGJP00, DGJP03] show that the space of processes is a separable metric space for each of these metrics.
- We describe some perturbation results — informally, we show that small perturbations of the probabilities in a process yields a process that is within a small distance of the unperturbed process.
- The definition of the metric has a quantification over all functional expressions. To ease working with metrics, we show that for $c < 1$, there is a single function that characterizes the ϵ balls around a given state.
- For $c < 1$, we show that that the problem $d(\mathcal{S}, \mathcal{S}') < \epsilon$ is decidable.

Definition 5.1 *Each collection $\mathcal{F}^c$ of functional expressions induces a distance function as follows:*

$$d^c(\mathcal{P}, \mathcal{Q}) = \sup_{f \in \mathcal{F}^c} |f_{\mathcal{P}}(p_0) - f_{\mathcal{Q}}(q_0)|.$$

Theorem 5.2 *For all $c \in (0,1]$, d^c is a metric.*

Proof The transitivity and symmetry of d^c are immediate. $d^c(\mathcal{S}, \mathcal{S}') = 0$ iff $\mathcal{S}$ and $\mathcal{S}'$ are bisimilar follows from Theorem 4.10. ∎

This definition is close in form to the definition of the Kantorovich metric [Hut81] which is used in the theory of optimal transport problems and also in the theory of fractals by Hutchinson. The difference is in the class of functions used. In the Kantorovich metric one uses the family of Lipschitz[8] functions. In our case the underlying state space is not a metric space[9] so we cannot really talk about Lipschitz functions. However - in a sense - these functions are really close to being Lipschitz. In suitable situations, one can show that our functions are dense in the class of Lipschitz functions.

We study the family of metrics $\{d^c \mid c \in (0,1]\}$. These metrics support the spectrum of possibilities of relative weighting of the two factors that contribute to the distance between processes: the complexity of the functions distinguishing them versus the amount by which each function distinguishes them. d^1 captures only the differences in the probability numbers; probability differences at the first transition are treated on par with probability differences that arise very deep in the evolution of the process. In contrast, d^c for $c < 1$ give more weight to the probability differences

[7] There are other interesting notions of metric that we do not address here.

[8] With Lipschitz constant 1 these are just the contractive functions.

[9] It is of course metrizable being analytic.

that arise earlier in the evolution of the process, i.e. differences identified by simpler functions. As c approaches 0, the future gets discounted more.

As is usual with metrics, the actual numerical values of the metric are less important than the notions of convergence that they engender. Thus, we take the uniformity view of metrics, e.g. see [Ger85][10], and will view the metric via properties like the significance of zero distance, relative distance of processes, contractivity and the notion of convergence rather than a detailed justification of the exact numerical values.

Example 5.3 The analysis of Example 4.12 yields $d^c(A_1, A_2) = c^2/4$. This is witnessed by the functional $\langle a\rangle \min(\langle a\rangle \mathbf{1}, (\mathbf{1} - \langle a\rangle \mathbf{1}) \ominus (\mathbf{1} - c))$.

Example 5.4 (Analysis of Example 1.4) Consider the family of processes $\{\mathcal{P}_\epsilon \mid 0 \leq \epsilon < r\}$ where $\mathcal{P}_\epsilon = a_{r-\epsilon}.\mathcal{Q}$, i.e. $\mathcal{P}_\epsilon$ is the process that does an a with probability $r - \epsilon$ and then behaves like $\mathcal{Q}$. The function expression $(\langle a\rangle \mathbf{1})$ evaluates to $(r - \epsilon)c$ at $\mathcal{P}_\epsilon$. This functional expression witnesses the distance between any two $\mathcal{P}$'s (other functions will give smaller distances). Thus, we get $d(\mathcal{P}_{\epsilon_1}, \mathcal{P}_{\epsilon_2}) = c|\epsilon_1 - \epsilon_2|$. This furthermore ensures that $\mathcal{P}_\epsilon$ converges to $\mathcal{P}_0$ as ϵ tends to 0.

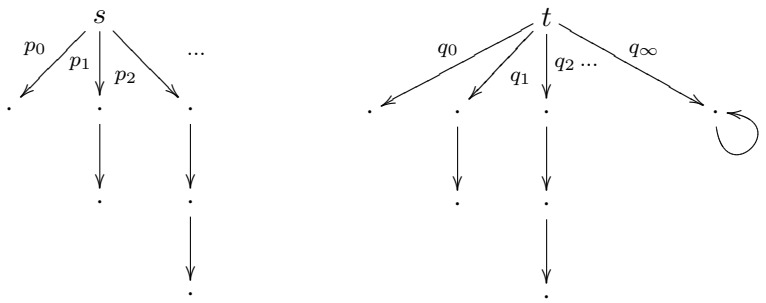


Figure 4:

Example 5.5 (from [DEP98]) Consider processes s and t of Figure 4. t is just like s except that there is an additional transition to a state which then has an a-labelled transition back to itself. The probability numbers are as shown. If both processes have the same values on all functional expressions we will show that $q_\infty = 0$, i.e. it really cannot be present. The functional expression $(\langle a\rangle \mathbf{1})$ yields $c(\sum_{i\geq 0} p_i)$ on s and $c(q_\infty + \sum_{i\geq 0} q_i)$ on t. The functional expression $(\langle a\rangle\langle a\rangle \mathbf{1})$ yields $c^2(\sum_{i\geq 1} p_i)$ on s and $c^2(q_\infty + \sum_{i\geq 2} q_i)$ on t. Thus, we deduce that $p_0 = q_0$. Similarly, considering functional expressions $(\langle a\rangle\langle a\rangle\langle a\rangle \mathbf{1})$ etc, we deduce that $p_n = q_n$. Thus, $q_\infty = 0$.

6 Metric Convergence of the Approximants

In this section we show that - in our metric - the approximants introduced in Section 3 converge to the labelled Markov process being approximated.

In order to prove convergence of the approximants we start with the following lemma. All lemmas for which we do not provide proofs here are proved in Desharnais' Ph.D. thesis [Des99a].

[10]Intuitively, a uniformity captures relative distances, eg. is x is closer to z than y; it does not tell us what the actual distances are. For example, a uniformity on a metric space M is induced by the collection of all ϵ balls S_ϵ where $S_\epsilon = \{\{y \mid d(x, y) < \epsilon\} \mid x \in M\}$.

Lemma 6.1 *If $\mathcal{S}$ involves a finite number of labels, $\mathcal{S}(n, c^n/n)$ converges to $\mathcal{S}$ in the metric d_c with $c < 1$.*

The condition $c < 1$ is important in the calculation. However, it has been pointed out to us that the restriction to finite action sets could be weakened to countable sets if we adapted the semantics for $\langle a_n \rangle f$ to be $\langle a_n \rangle f(s) = c^n \int_S f(t) \tau_{a_n}(s, dt)$. The approximation algorithm does capture every label of a countable set thanks to parameter n, which not only refers to the depth of the approximant but also to the labels that are considered in approximation, $a_1, \ldots, a_n$.

Lemma 6.2 *Given any process of the form $\mathcal{S}(n, \epsilon)$ we can construct a sequence of rational trees $\mathcal{T}_i$ such that $\mathcal{T}_i$ is strictly simulated by $\mathcal{T}_{i+1}$ and all of them are strictly simulated by $\mathcal{S}(n, \epsilon)$ and with $\lim_{i \to \infty} d(\mathcal{T}_i, \mathcal{S}(n, \epsilon)) = 0$.*

Proof Given a finite acyclic process like $\mathcal{S}(n, \epsilon)$ we can construct a finite depth tree, say $\mathcal{T}$, that is bisimilar to it by duplicating states as necessary. The transition probabilities of this tree will not necessarily be rational numbers. We can construct out required family of trees by making all the $\mathcal{T}_i$ have the same shape as $\mathcal{T}$ but by choosing the transition probabilites in the $\mathcal{T}_i$ to be rational numbers converging to the corresponding transition probabilities of $\mathcal{T}$. Since these proabilities are all strictly increasing we get the desired strict simulation. The convergence is immediate from the definition of the metric. ∎

The result that we want can be stated as follows.

Theorem 6.3 *For all $c \in (0, 1]$, the metric d^c yields a separable metric space.*

Proof We show that the rational trees form a countable dense subset. Given any process $\mathcal{S}$ we have a countable family of finite approximations given by $\mathcal{S}(n, 2^{-n})$. For each of these finite approximations we have a countable sequence of rational trees, $\mathcal{T}_j^{(n)}$ that converge to it by the previous lemma. This doubly indexed family of rational trees forms a directed set so we can extract a countable sequence of rational trees that converge to $\mathcal{S}$. ∎

Thus we have a situation analogous to the rationals where there is a countable family that serves to aproximate all the processes as limits of Cauchy sequences. What we do not know is whether the metric space is complete; in other words we do not know whether we have a Polish space.

7 Examples of metric reasoning principles

In this section, we use a process algebra and an example coded in the process algebra to illustrate the type of reasoning provided by our study. We also show that small perturbations of a process results in a nearby process.

7.1 A process algebra

The process algebra describes probabilistically determinate processes. The processes are input-enabled [LT89, Dil88, Jos92] in a weak sense ($(\forall p \in P)\ (\forall a \in \mathcal{A})\ \tau_{a?}(p, P) > 0$) and communication is via CSP style broadcast. The process combinators that we consider are parallel composition, prefixing and probabilistic choice. We do not consider hiding since this paper focuses on strong probabilistic bisimulation. Though we do not enforce the fact that output actions do not block, this assumption can safely be added to the algebra to make it an IO calculus [Vaa91].

We assume an underlying set of labels $\mathcal{A}$. Let $\mathcal{A}? = \{a? \mid a \in \mathcal{A}\}$ be the set of input labels, and $\mathcal{A}! = \{a! \mid a \in \mathcal{A}\}$ the set of output labels. Every process $\mathcal{P}$ is associated with a subset of labels: $\mathcal{P}_O \subseteq \mathcal{A}!$, the set of relevant output labels. This signature is used to constrain parallel composition.

Prefixing $\mathcal{P} = a?_r.\mathcal{Q}$ where r is a rational number, is the process that accepts input a and then performs as $\mathcal{Q}$. The number r is the probability of accepting $a?$. With probability $(1-r)$ the process $\mathcal{P} = a?_r.\mathcal{Q}$ will *block* on an $a?$ label. P is given by adding a new initial state, p_0 to Q. Add a transition labelled $a?$ from p_0 to the start state of $\mathcal{Q}$ with probability r. For all other labels l, add a $l?$ labelled self-loop at p_0 with probability 1.

Output prefixing, $\mathcal{P} = a!_r.\mathcal{Q}$, where r is a rational number, the process that performs output action $a!$ and then functions as $\mathcal{Q}$, is defined analogously. In this case, $\mathcal{P}_O = \mathcal{Q}_O \cup \{a!\}$. For both input and output prefixing, we have: $d^c(a_r.\mathcal{P}, a_u.\mathcal{P}) \leq c \mid r - u \mid$.

Probabilistic Choice $\mathcal{P} = \mathcal{Q} +_r \mathcal{Q}'$ is the probabilistic choice combinator [JP89] that chooses $\mathcal{Q}$ with probability r and $\mathcal{Q}'$ with probability $1-r$. $\mathcal{P}_O = \mathcal{Q}_O \cup \mathcal{Q}'_O$. $P = Q \uplus Q'$. Now $\tau_a^{\mathcal{P}}(q, X \uplus X') = \tau_a(q, X)$ if $q \in Q$, and $\tau_a^{\mathcal{P}}(q, X \uplus X') = \tau'_a(q, X')$ if $q \in Q'$. We define an initial distribution μ: $\mu(\{q_0\}) = r, \mu(\{q'_0\}) = 1 - r$, referring the reader to Section 2 for a way to convert to an initial state format.

We have: $d^c(\mathcal{P} +_r \mathcal{Q}, \mathcal{P} +_u \mathcal{Q}) \leq \mid r - u \mid d^c(\mathcal{P}, \mathcal{Q})$; $d^c(\mathcal{P} +_r \mathcal{Q}, \mathcal{P}' +_r \mathcal{Q}) \leq r d^c(\mathcal{P}, \mathcal{P}')$.

Parallel Composition $\mathcal{P} = \mathcal{Q} \mid\mid \mathcal{Q}'$ is permitted if the output actions of $\mathcal{Q}, \mathcal{Q}'$ are disjoint, i.e. $\mathcal{Q}_O \cap \mathcal{Q}'_O = \emptyset$. The parallel composition synchronizes on all labels in $\mathcal{Q}_L \cap \mathcal{Q}'_L$. $\mathcal{P}_O = \mathcal{Q}_O \uplus \mathcal{Q}'_O$. $P = Q \times Q'$. The $\tau_a^{\mathcal{P}}$ definition is motivated by the following idea. Let s (resp. s') be a state of $\mathcal{Q}$ (resp. $\mathcal{Q}'$). We expect the following synchronized transitions from the product state (s, s').

$$\frac{s \xrightarrow{c?} t \quad s' \xrightarrow{c?} t'}{(s,s') \xrightarrow{c?} (t,t')} \qquad \frac{s \xrightarrow{c!} t \quad s' \xrightarrow{c?} t'}{(s,s') \xrightarrow{c!} (t,t')} \qquad \frac{s \xrightarrow{c?} t \quad s' \xrightarrow{c!} t'}{(s,s') \xrightarrow{c!} (t,t')}.$$

The disjointness of the output labels of $\mathcal{Q}, \mathcal{Q}'$ ensures that there is no non-determinism. Formally, if $l = a! \in \mathcal{Q}_O$, then $\tau^{\mathcal{P}}_{a?}((s,s'),(t,t')) = \tau^{\mathcal{P}}_{a!}((s,s'),(t,t')) = \tau_{a!}(s,t) \times \tau'_{a?}(s',t')$. The case when $a! \in \mathcal{Q}'_O$ and $l = a?$ is similar.

To fix terminology, let us use the same symbol $\mathcal{P}$ to stand for the syntactic expression for a process and for the labelled transition system generated by $\mathcal{P}$. When a process, say $\mathcal{P}$, has an a-transition we cannot say that it results in a process $\mathcal{P}'$; instead, we must say that it results in some distribution of possible states of $\mathcal{P}$ - these states are, of course, denoted in the syntax by derivatives of the syntactic expression for $\mathcal{P}$.

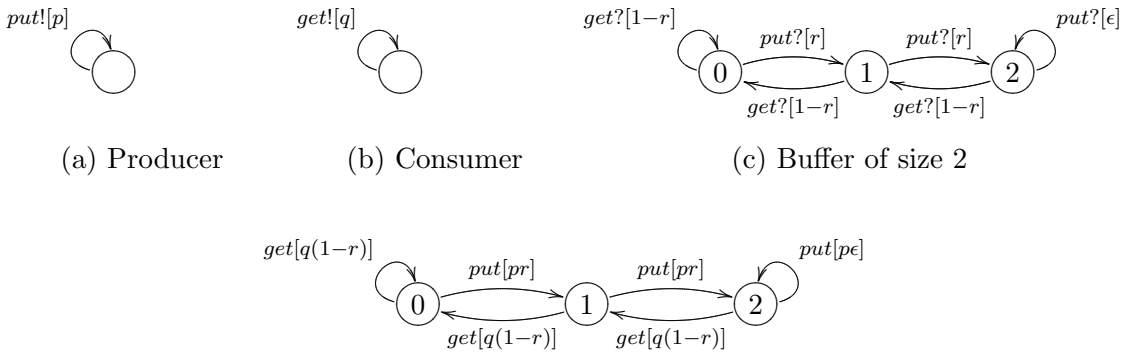
Definition 7.1 *Let $\mathcal{P}$ be a process. Then $\mathcal{P}$* **after** *a is the same process but with start distribution given by $\nu(t) = \tau_a(p_0, t)$. We perform some normalization based on the total probability of the resulting initial configuration $\nu(P)$: If $\nu(P) > 0$, it is normalized to be 1; if $\nu(P) = 0$, it is left untouched. This definition extends inductively to $\mathcal{P}$* **after** *α, where α is a finite sequence of labels $(a_0, a_1, a_2, \ldots, a_k)$.*

Note that $\mathcal{P}$ **after** α is identical to $\mathcal{P}$ - i.e. it denotes the same labelled transition system - except that its initial configuration may be different.

Lemma 7.2 *Let $h \in \mathcal{F}^c$, let $\mathcal{P}$ be a process and let $a \in \mathcal{A}$. Then*

$$\langle a \rangle h(p_0) = c \times h(\mathcal{P} \textbf{ after } a).$$

Here $h(\mathcal{P}$ **after** *$a)$ means $h(p'_0)$ where p'_0 is the initial state of $\mathcal{P}$* **after** *a.*



(a) Producer (b) Consumer (c) Buffer of size 2

(d) Producer || Consumer || Buffer$_2$

Figure 5: The producer consumer example

Theorem 7.3 (Contractivity of process combinators)

- $d(l_r.\mathcal{P}, l_r.\mathcal{Q}) \leq cd(\mathcal{P},\mathcal{Q})$ *for any label* l
- $d(\mathcal{P} +_r \mathcal{R}, \mathcal{Q} +_r \mathcal{R}) \leq d(\mathcal{P},\mathcal{Q})$ *for any* $\mathcal{R}$
- $d(\mathcal{P} \mid\mid \mathcal{R}, \mathcal{Q} \mid\mid \mathcal{R}) \leq d(\mathcal{P},\mathcal{Q})$ *for any* $\mathcal{R}$ *for which* $\mathcal{P} \mid\mid \mathcal{R}, \mathcal{Q} \mid\mid \mathcal{R}$ *are defined.*

Proof The proof proceeds by induction on functional expressions. Let $f^-(\mathcal{P},\mathcal{Q})$ mean $|f(p_0) - f(q_0)|$ where p_0 (q_0) is the initial state of $\mathcal{P}$ ($\mathcal{Q}$). We show that for any f, there exists a g such that f^- of the LHS is less than or equal to some g^- of the RHS. We omit the detailed calculations and prove the result for the key case where f is $\langle a\rangle h$, for parallel composition. Let $\mathcal{P}' = \mathcal{P}$ **after** $b?$, $\mathcal{Q}' = \mathcal{Q}$ **after** $b?$ and $\mathcal{R}' = \mathcal{R}$ **after** $b!$. By induction, we know that there is some functional g such that $h^-(\mathcal{P}' \mid\mid \mathcal{R}', \mathcal{Q}' \mid\mid \mathcal{R}') \leq g^-(\mathcal{P}',\mathcal{Q}')$. Now suppose $a = b!$, and $b! \in \mathcal{R}_O$, then $\mathcal{P} \mid\mid \mathcal{R}$ **after** $a = \mathcal{P}' \mid\mid \mathcal{R}'$. Now we calculate as follows:

$$\begin{aligned}
(\langle a\rangle h)^-(\mathcal{P} \mid\mid \mathcal{R}, \mathcal{Q} \mid\mid \mathcal{R}) &= c \times h^-((\mathcal{P} \mid\mid \mathcal{R}) \textbf{ after } a, (\mathcal{Q} \mid\mid \mathcal{R}) \textbf{ after } a) \\
&= c \times h^-(\mathcal{P}' \mid\mid \mathcal{R}', \mathcal{Q}' \mid\mid \mathcal{R}') \\
&\leq c \times g^-(\mathcal{P}',\mathcal{Q}') \\
&= (\langle a\rangle g)^-(\mathcal{P},\mathcal{Q}).
\end{aligned}$$

∎

Thus,Theorem 4.10 allows us to conclude that bisimulation is a congruence with respect to these operations.

7.2 A bounded buffer example

We specify a producer consumer process with a bounded buffer (along the lines of [PS85]). The producer is specified by the 1 state finite automaton shown in Figure 5(a) — it outputs a *put*, corresponding to producing a packet, with probability p. To keep the figure uncluttered, we omit the input-enabling arcs, all of which have probability 1. The consumer (Figure 5(b)) is analogous — it outputs a *get* with probability q, corresponding to consuming a packet. The buffer is an n-state automaton, the states are merely used to count the number of packets in the buffer, while

the probabilities code up the probability of scheduling either the producer or the consumer (thus the producer gets scheduled with probability r, and then produces a packet with probability p). Upon receiving a *put* in the last state, the buffer accepts it with a very small probability ϵ, modeling a blocked input. The parallel composition of the three processes is shown in Figure 5(d). Notice that the behavior of this process is very similar to a random walk — the process moves to the next state with probability $r = p(1-q)/(p+q-pq)$, corresponding to a *put*, and the previous state with probability $1-r$, corresponding to a *get* — we ignore the transitions back to the same state, regarding them as no-ops. It is easy to show that in any run of this process with a large number of *put* actions, the expected fraction of discarded packets is approximately $(1-r/r)^{-n}$ — we compute the stationary distribution for this process, and since it is ergodic, this stationary distribution is reached after a large number of steps. Then the *put* actions in the last state result in lost packets.

As the buffer size increases, the distance between the bounded buffer and the unbounded buffer decreases to 0. Let $\mathcal{P}_k$ = Producer || Consumer || Buffer_k, where Buffer_k denotes the process Buffer with k states. Then by looking at the structure of the process, we can compute that $d(\mathcal{P}_k, \mathcal{P}_\infty) \propto (cpr)^k$. Thus we conclude the following:

- As the bounded buffer becomes larger, it approximates an infinite buffer more closely: if $m > k$ then $d^c(\mathcal{P}_k, \mathcal{P}_\infty) > d^c(\mathcal{P}_m, \mathcal{P}_\infty)$.
- As the probability of a put decreases, the bounded buffer approximates an infinite buffer more closely. Thus if $p < p'$, $d^c(\mathcal{P}^p, \mathcal{P}^p_\infty) < d^c(\mathcal{P}^{p'}, \mathcal{P}^{p'}_\infty)$, where the superscripts indicate the producer probability.
- Similarly, as the probability of scheduling the Producer process (r) decreases, the buffer approximates an infinite buffer more closely.

7.3 Perturbation

One of the major criticisms of process equivalences is that they are not robust. The results of this section show that if one slightly perturbs the probabilities in a process the result is close.

Definition 7.4 *Let $\mathcal{S} = (S, s_0, \Sigma, \tau)$ be a labelled Markov process. Define $\mathcal{S}' = (S, s_0, \Sigma, \tau')$ to be an ϵ-perturbation of $\mathcal{S}$ if for all labels a,*

$$\forall s \in S.\ \forall X \in \Sigma.\ |\tau_a(s, X) - \tau'_a(s, X)| < \epsilon.$$

Our metric accommodates the notion of small perturbations of probabilities.

Proposition 7.5 *If $c < 1$, and $\mathcal{S}'$ is an ϵ-perturbation of $\mathcal{S}$, then $d^c(\mathcal{S}, \mathcal{S}') < k\epsilon$ where $k = \sup_n nc^n$.*[11]

Proof The proof is by induction on the formulas. The sole non-trivial case is $\langle a \rangle f$. We write f for $f_{\mathcal{S}}$ and f' for $f_{\mathcal{S}'}$. Let $depth(f) = n$, and $|f(t) - f'(t)| < \epsilon n c^n$. Then $f(s) \leq c^n$ and

[11] *e.g.* $k = 1$ for $c \leq 1/2$.

$$
\begin{aligned}
c[\int_t f(t)\tau_a(s,dt) - \int_t f'(t)\tau'_a(s,dt)] & \\
&= c\int f(t)[\tau_a(s,dt) - \tau'_a(s,dt)] + c\int \tau'_a(s,dt)[f(t) - f'(t)] \\
&< c^{n+1}|\tau(s,X) - \tau'(s,X)| + nc^{n+1}\epsilon \int \tau'_a(s,t) \\
&< c^{n+1}\epsilon + nc^{n+1}\epsilon \\
&= (n+1)c^{n+1}\epsilon.
\end{aligned}
$$

Here X is the set on which the measure $\tau_a(s,.) - \tau'_a(s,.)$ is positive. ∎

For $c = 1$, nc^n increases without limit, and Example 4.3 shows that the above lemma does not hold for $c = 1$. However in this case we can still perturb the process $\mathcal{S}$ in the following way — let $\mathcal{S}$ be unfolded, so it has no loops. Let $\epsilon_i, i \in \mathbf{N}$ be non-negative rationals such that $\sum_i \epsilon_i = \epsilon < 1/3$. Now we obtain $\mathcal{S}'$ by taking the same state set as $\mathcal{S}$, and for each state s at depth n, $|\tau_a(s,X') - \tau'_a(s,X')| < \epsilon_n$ for each label a and each measurable set X'. Then we can show by a similar calculation as above that $d^1(\mathcal{S},\mathcal{S}') < 1 - e^{-2\epsilon}$, thus as $\epsilon \longrightarrow 0$, $d^1(\mathcal{S},\mathcal{S}') \longrightarrow 0$.

Example 7.6 Consider "straight line" formulas generated by

$$\phi ::= \mathsf{T} \mid \langle a \rangle_q \phi$$

Consider one such $\phi = \langle a_1 \rangle_{q_1} \ldots \langle a_n \rangle_{q_n} \mathsf{T}$. Let $\mathcal{P}$ be a finite-state process unfolded to the depth of the formula such that p_0, the start state of $\mathcal{P}$, satisfies the formula. An easy induction, using the proof of Lemma 4.5, shows that

$$f_\phi(p_0) \geq c^n \prod_i (r_i - q_i)$$

where $r_i = \inf_{s \in X_i} \tau_{a_i}(s, X_{i+1})$ and X_{i+1} is the set of all the states in level $i+1$ which satisfy the suffix formula $\langle a_{i+1} \rangle_{q_{i+1}} \ldots \langle a_n \rangle_{q_n} \mathsf{T}$. Note that this bound is achieved by the n-length chain automaton which has transition probabilities r_i.

The form of the expression $f(p_0) \geq c^n \prod_i (r_i - q_i)$ tells us that if $f(p_0) > \epsilon$, we can perturb the probabilities at some level by up to $\epsilon^{\frac{1}{n}}/c$, and the resulting process will continue to satisfy the formula.

Finally we close with an important example that shows the importance of the connectivity of the transition graph.

Example 7.7 Consider the systems shown in Fig. 6. The states s_0 and t_0 appear to be very similar and are clearly metrically close - or are they? In system (A) there is no steady state distribution (the Markov chain fails to be aperiodic) whereas in system (B) there is a steady state, namely all the mass eventually leaks into state t_2 and stays there. How is it that the asymptotic behaviour can be so drastically different when the state are so close?

The short answer is that the states *are not at all close*. If one computes the distance a routine calculation shows that the states s_0 and t_0 are at distance 1 for the metrics with $c = 1$ - the maximum possible distance! Even with $c < 1$ the distance is large though less than 1.

(a) System A (b) System B

Figure 6: The effect of topology change

8 The asymptotic metric

Define the j distance between $\mathcal{P}, \mathcal{Q}$, $d^c_j(\mathcal{P}, \mathcal{Q}) = \sup\{d^c(\mathcal{P}\ \textbf{after}\ \alpha, \mathcal{Q}\ \textbf{after}\ \alpha) \mid length(\alpha) = j\}$. We define the asymptotic distance between processes $\mathcal{P}$ and $\mathcal{Q}$, $d^c_\infty(\mathcal{P}, \mathcal{Q})$ to be

$$d^c_\infty(\mathcal{P}, \mathcal{Q}) = \limsup_{i \to \infty\ j > i} d^c_j(\mathcal{P}, \mathcal{Q}).$$

The fact that d^c_∞ satisfies the triangle inequality and is symmetric immediately follows from the same properties for d.

Example 8.1 For any process $\mathcal{P}$, $d^c_\infty(a_q.\mathcal{P}, a_r.\mathcal{P}) = 0$, where $q, r > 0$. Consider A_3 from Figure 3. Without the normalization in the definition of A_3 **after** α, we would have got $d^c_\infty(a_q.A_3, a_r.A_3) = c|q - r|$

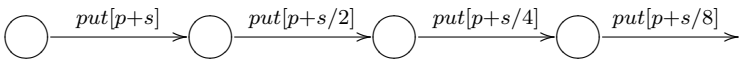


Figure 7: A producer with transient behavior

Example 8.2 Consider the producer process $\mathcal{P}_2$ shown in Figure 7. This is similar to the producer $\mathcal{P}_1$ in Figure 5, except that initially the probability of producing *put* is more than q, however as more *put*'s are produced, it asymptotically approaches q. If we consider the asymptotic distance between these two producers, we see that $d^c(\mathcal{P}_2\ \textbf{after}\ put^n, \mathcal{P}_1\ \textbf{after}\ put^n) \propto 2^{-(n+1)}$. Thus $d^c_\infty(\mathcal{P}_1, \mathcal{P}_2) = 0$. Now by using the compositionality of parallel composition (see below) , we see that $d^c_\infty(\mathcal{P}_1 \mid\mid \text{Consumer} \mid\mid \text{Buffer}_k, \mathcal{P}_2 \mid\mid \text{Consumer} \mid\mid \text{Buffer}_k) = 0$, which is the intuitively expected result.

Asymptotic equivalence is preserved by parallel composition and prefixing.

Theorem 8.3 *1.* $d^c_\infty(l_r.\mathcal{P}, l_r.\mathcal{Q}) \leq d^c_\infty(\mathcal{P}, \mathcal{Q})$ *for any label* l.

2. $d^c_\infty(\mathcal{P} \mid\mid \mathcal{R}, \mathcal{Q} \mid\mid \mathcal{R}) \leq d_\infty(\mathcal{P}, \mathcal{Q})$.

For the key case of parallel composition, the proof is based on:

$$(\mathcal{P} \mid\mid \mathcal{Q})\ \textbf{after}\ \alpha = (\mathcal{P}\ \textbf{after}\ \alpha_1) \mid\mid (\mathcal{Q}\ \textbf{after}\ \alpha_2),$$

where α_1 has the $a!$ labels of α replaced by $a?$ where $a! \notin \mathcal{P}_O$, and similarly for α_2.

9 Related work

The study of the interaction of probability and concurrency - largely in the context of exact equivalences of probabilistic processes - has been explored extensively in the context of different models of concurrency. Probabilistic process algebras add a notion of randomness to the standard process algebra model and have been studied extensively in the traditional framework of semantic theories of process algebras. A representative sample of such work are the following papers: [HJ90, JY95, LS91, HS86, BBS95, vGSS95, CSZ92]. These papers study concepts like probabilistic bisimulation [LS91]probabilistic testing [JY95] and the relationship with (probabilistic) modal logics [HS86]. Probabilistic Petri nets [Mar89, VN92] add Markov chains to the underlying Petri net model. This area has a well developed suite of algorithms for performance evaluation. Investigations into the behaviour of probabilistic systems have also been carried out in the context of IO Automata [Seg95, WSS97].

In contrast to the above body of research the primary theme of this paper is the study of intersubstitutivity of *approximately* equivalent processes. As a minor theme we have also initiated the study of *asymptotic* approximate equivalence. The ideas of approximate substitutivity in this paper are inspired by the work of Jou and Smoka [JS90] and also to the ideas in the area of performance modeling as exemplified in on the work on process algebras for compositional performance modeling (see for example [Hil94]). The extension of the methods of this paper to systems which have both probability and traditional nondeterminism remains open and is the subject of active research at the moment.

The verification community has been active in developing model checking tools for probabilistic systems, for example [BLL$^+$96, BdA95, BCHG$^+$97, CY95, HK97]. Approximation techniques in the spirit of those of this paper have been explored for hybrid systems [GHJ97]. Since the first appearance of the present work [DGJP99] we have developed a theory of approximation for labelled Markov processes [DGJP00, DGJP03].

Before we discuss work specifically related to metrics for probabilistic processes it is worth discussing work from the probability theory community on metrics on spaces of measures. An excellent mathematical resource is "Probability Measures on Metric Spaces" by K. R. Parthasarathy [Par67] which deals with the dual topic i.e. measures on metric spaces rather than metrics on spaces of measures. Various measures of "distance" have appeared in the pattern theory and statistics communities - see [DGL96, Chapter 3] for a survey - but most of these are not metrics i.e. they do not obey the triangle inequality. Perhaps the most interesting of these is the relative entropy [CT91] which measures how much uncertainty about a random variable exists when another one is known.

The Kantorovich metric [Kan40] was introduced by Kantorovich in his study of the transshipment problem and was also used later by Hutchinson [Hut81, Edg98] to analyze fractals. His definition is

$$d(\mu,\nu) := \sup_{f\in\mathcal{L}} |\int f \mathrm{d}\mu - \int f \mathrm{d}\nu|$$

where $\mathcal{L}$ is the class of bounded Lipschitz functions. This is a metric as opposed to just being a "distance" function.

Metrics for probabilistic processes have been investigated by a few researchers: deVink and Rutten [dVR99], Kwiatkowska and Norman [KN96, KN98] and very recently by van Breugel and Worrell [vBW01b, vBW01a]. As remarked before, the suggestion that one should look for a metric is due to Giacalone, Jou and Smolka [GJS90]. DeVink and Rutten use *ultrametrics* as a technical tool for defining probabilistic transition systems as coalgebras. Their main interest is in bisimilarity and they did not investigate the idea of using the metric as an alternative to bisimulation. The work

of Kwiatkowska and Norman is very interesting but more motivated by semantical considerations. They are also not interested in using the metric to capture an approximation to bisimulation. For example, there are distance zero processes that are not bisimilar.

The work closest to ours in aim and techniques is that of van Breugel and Worrell. Their metric appears to be the same as our metric for $c < 1$ but the way in which it is defined is quite different. Their construction is based on finding a final coalgebra for a certain functor on the category of metric spaces and nonexpansive maps. This final coalgebra comes with a metric and thus naturally gives a metric on the state spaces of any labelled Markov process through the unique map induced by finality. Their functor is closely based on the Kantorovich metric.

There are a host of topologies that seem to be relevant. First of all the metrics with $c < 1$ and with $c = 1$ are clearly different. Consider the family of processes $\{P_n | n \in \mathbb{N}\}$ where P_n is defined as the process that makes n transitions, each with probability 1, and then terminates. The process P is defined as a countable-state process with the states labelled by natural numbers. There is a transition in P from n to $n + 1$ with probability 1. Using the metrics with $c < 1$, the processes P_n form a Cauchy sequence converging to P. Using the metric with $c = 1$ the P_n do not form a Cauchy sequence. In our study of approximation [DGJP03] we defined a domain of processes. This domain comes equipped with the Scott topology and the Lawson topology. Recently van Breugel at. al. [vBMOW03] have shown that the Lawson topology coincides with the weak topology and with the metric topology for $c < 1$. The Scott topology is, of course, not even metrizable.

A very important contribution of van Breugel and Worrell [vBW01a] is the discovery of a polynomial-time algorithm for the metric. The algorithm makes clever use of linear programming ideas - the transshipment problem - and is actually implemented. Note that their algorithm works for the metrics with $c < 1$. For $c = 1$ we do not know if the metric is decidable; though there is strong experimental evidence [vBb] and heuristic arguments suggesting that it is decidable. In the conference version of this paper we published a decision procedure for the metric when $c < 1$. This was a very crude algorithm and had exponential running time. The beautiful algorithm of van Breugel and Worrell renders our old algorithm obsolete.

Since that time we have developed a metric analogue of weak bisimulation [DGJP02]. This is essentially based on the $c = 1$ metric of the present paper and uses a fixed-point approach mimicking - at the lattice level - the categorical construction of van Breugel and Worrell [vBW01b]. We also use linear programming ideas in a crucial way in that construction.

Our work on the asymptotic metric is closely related to, at least in spirit, the work of Lincoln, Mitchell, Mitchell and Scedrov [LMMS98] in the context of security protocols. Both [LMMS98] and this paper consider the asymptotic behavior of a single process, rather than the limiting behavior of a probabilistically described family of processes as is performed in some analysis performed in Markov theory.

10 Conclusions and Future Work

In this paper, we deal with probabilistic nondeterminism. In a probabilistic analysis, quantitative information is recorded and used in the reasoning. In contrast, a purely qualitative nondeterministic analysis does not require and does not yield quantitative information. In particular, when one has no quantitative information at all, one has to work with indeterminacy — using a uniform probability distribution is not the same as expressing complete ignorance about the possible outcomes or their distribution. Thus nondeterminism is more appropriate for a specification formalism where certain probabilities are left unspecified by the system designer. In cases - such as experimental statistics or performance evaluation - where one is comparing a model against experimental data or analyzing

a system, the probabilistic model is more appropriate.

Our main contribution has been to come up with metric definitions that serve as a weakening of the usual notion of bisimulation. The move was partly inspired by Kozen's work on viewing measurable functions as the formulas of a "quantitative" logic. We have shown that our functions characterize bisimulation in the same way that modal logics do [DEP02]. The metrics based on these functions have the important property that when the distance is zero the processes are bisimilar. We also showed that - for our simple process algebra - the metrics are contractive. This allows us to use the metric for compositional reasoning.

We are working in two different directions. In the first we are looking at the mixture of probability and nondeterminism. This leads naturally to think about weak bisimulation and to characterize weak bisimulation in metric terms. The second direction that we are investigating is trying to understand the quantitative significance of the metric in terms of information flow. It seems plausible that there should be a strong correlation between the behaviours of "close" processes. To make this more concrete we are thinking about information flow in the context of probabilistic processes and the role of relative entropy or channel capacity. It would be fascinating if the distance between processes were linked to information theory notions. In recent work [DGJP02] we have been able to make such links and apply it to the notion of secure substitution.

Acknowledgements We have benefited greatly from discussions with Franck van Breugel and James (Ben) Worrell. We would like to acknowledge support from NSERC (Panangaden and Desharnais), FCAR (Desharnais) and NSF (Jagadeesan).

References

[AJKvO97] R. Alur, L. Jagadeesan, J. J. Kott, and J. E. von Olnhausen. Model-checking of real-time systems: A telecommunications application. In *Proceedings of the 19th International Conference on Software Engineering*, pages 514–524, 1997.

[BBS95] J.C.M. Baeten, J.A. Bergstra, and S.A. Smolka. Axiomatizing probabilistic processes: ACP with generative probabilities. *Information and Computation*, 121(2):234–255, 1995.

[BCHG+97] Christel Baier, Ed Clark, Vasiliki Hartonas-Garmhausen, Marta Kwiatkowska, and Mark Ryan. Symbolic model checking for probabilistic processes. In *Proceedings of the 24th International Colloquium On Automata Languages And Programming*, number 1256 in Lecture Notes In Computer Science, pages 430–440, 1997.

[BdA95] A. Bianco and L. de Alfaro. Model checking of probabilistic and nondeterministic systems. In P. S. Thiagarajan, editor, *Proceedings of the 15th Annual Conference on Foundations of Software Technology and Theoretical Computer Science*, number 1026 in Lecture Notes In Computer Science, pages 499–513, 1995.

[BDEP97] R. Blute, J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labelled Markov processes. In *Proceedings of the Twelfth IEEE Symposium On Logic In Computer Science, Warsaw, Poland.*, 1997.

[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.

[CSZ92] R. Cleaveland, S. Smolka, and A. Zwarico. Testing preorders for probabilistic processes. In W. Kuich, editor, *Automata, Languages and Programming (ICALP 92)*, number 623 in Lecture Notes in Computer Science, pages 708–719. Springer-Verlag, 1992.

[CT91] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley, New York, 1991.

[CY95] C. Courcoubetis and M. Yannakakis. The complexity of probabilistic verification. *Journal of the ACM*, 42(4):857–907, 1995.

[dA] L. de Alfaro. Private communication.

[DEP98] J. Desharnais, A. Edalat, and P. Panangaden. A logical characterization of bisimulation for labelled Markov processes. In *proceedings of the 13th IEEE Symposium On Logic In Computer Science, Indianapolis*, pages 478–489. IEEE Press, June 1998.

[DEP02] J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labeled Markov processes. *Information and Computation*, 179(2):163–193, Dec 2002.

[Des99a] J. Desharnais. *Labelled Markov Processes*. PhD thesis, McGill University, November 1999.

[Des99b] J. Desharnais. Logical characterization of simulation for Markov chains. In M. Kwiatkowska, editor, *Probmiv99, Proceedings of the Second International Workshop on Probabilistic Methods in Verification*, pages 33–48. The University of Birmingham, 1999. Available as TR: CSR-99-8.

[DGJP99] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled Markov systems. In *Proceedings of CONCUR99*, Lecture Notes in Computer Science. Springer-Verlag, 1999.

[DGJP00] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Approximation of labeled Markov processes. In *Proceedings of the Fifteenth Annual IEEE Symposium On Logic In Computer Science*, pages 95–106. IEEE Computer Society Press, June 2000.

[DGJP02] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. The metric analogue of weak bisimulation for labelled Markov processes. In *Proceedings of the Seventeenth Annual IEEE Symposium On Logic In Computer Science*, pages 413–422, July 2002.

[DGJP03] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Approximating labeled Markov processes. *Information and Computation*, 2003. To appear. Available from `http://www.ift.ulaval.ca/~jodesharnais` and `http://rl.cs.mcgill.ca/~prakash/pubs.html`.

[DGL96] L. Devroye, L. Györfi, and G. Lugosi. *A Probabilistic Theory of Pattern Recognition*. Springer-Verlag, 1996.

[Dil88] D. Dill. *Trace Theory for Automatic Hierarchical Verification of Speed-Independent Circuits*. ACM Distinguished Dissertations. MIT Press, 1988.

[dKW89] Johan de Kleer and B. C. Williams. Diagnosis with behavioral modes. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, pages 1324–1330, August 1989.

[dVR99] E. de Vink and J. J. M. M. Rutten. Bisimulation for probabilistic transition systems: A coalgebraic approach. *Theoretical Computer Science*, 221(1/2):271–293, June 1999.

[Edg98] Gerald A. Edgar. *Integral, Probability and Fractal Measures*. Springer-Verlag, 1998.

[Ger85] Robert Geroch. *Mathematical Physics*. Chicago Lectures in Physics. University of Chicago Press, 1985.

[GHJ97] V. Gupta, T. A. Henzinger, and R. Jagadeesan. Robust timed automata. In Oded Maler, editor, *Hybrid and Real-Time Systems*, volume 1201 of *Lecture Notes In Computer Science*, pages 331–345. Springer Verlag, March 1997.

[GJS90] A. Giacalone, C. Jou, and S. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Proceedings of the Working Conference on Programming Concepts and Methods*, IFIP TC2, 1990.

[Hil94] J. Hillston. *A Compositional Approach to Performance Modelling*. PhD thesis, University of Edinburgh, 1994. Published as a Distinguished Dissertation by Cambridge University Press in 1996.

[HJ90] H. Hansson and B. Jonsson. A calculus for communicating systems with time and probabilities. In *Proceedings of the 11th IEEE Real-Time Systems Symposium*, pages 278–287. IEEE Computer Society Press, 1990.

[HK97] M. Huth and M. Kwiatkowska. Quantitative analysis and model checking. In *proceedings of the 12 IEEE Symposium On Logic In Computer Science*, pages 111–122. IEEE Press, 1997.

[HS86] S. Hart and M. Sharir. Probabilistic propositional temporal logics. *Information and Control*, 70:97–155, 1986.

[Hut81] J. E. Hutchinson. Fractals and self-similarity. *Indiana Univ. Math. J.*, 30:713–747, 1981.

[Jos92] M. B. Josephs. Receptive process theory. *Acta Informatica*, 29(1):17–31, February 1992.

[JP89] C. Jones and G. D. Plotkin. A probabilistic powerdomain of evaluations. In *Proceedings of the Fourth Annual IEEE Symposium On Logic In Computer Science*, pages 186–195, 1989.

[JS90] C.-C. Jou and S. A. Smolka. Equivalences, congruences, and complete axiomatizations for probabilistic processes. In J.C.M. Baeten and J.W. Klop, editors, *CONCUR 90 First International Conference on Concurrency Theory*, number 458 in Lecture Notes In Computer Science. Springer-Verlag, 1990.

[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.

[Kan40] L. V. Kantorovich. A new method for solving some classes of extremal problems. *Comptes Rendus (Doklady) Acad. Sci. USSR*, 28:211–214, 1940.

[KN96] M. Kwiatkowska and G. J. Norman. Probabilistic metric semantics for a simple language with recursion. In W. Penczek and A. Szalas, editors, *Proceedings of the 21st International Symposium on Mathematical Foundations of Computer Science*, number 1113 in Lecture Notes in Computer Science, pages 419–430, 1996.

[KN98] M. Kwiatkowska and G. J. Norman. A testing equivalence for reactive probabilistic processes. In I. Castelliani and C. Palamidessi, editors, *Proceedings of the Fifth International Workshop on Expressiveness in Concurrency*, volume 16 of *Electronic Notes In Theoretical Computer Science*. Elsevier, 1998.

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

[LMMS98] P. D. Lincoln, J.C. Mitchell, M. Mitchell, and A. Scedrov. A probabilistic poly-time framework for protocol analysis. In *ACM Computer and Communication Security (CCS-5)*, 1998.

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

[LT89] N. A. Lynch and M. R. Tuttle. An introduction to input/output automata. *CWI Quarterly*, 2(3):219–246, 1989.

[Mar89] M. Ajmone Marsan. Stochastic petri nets: an elementary introduction. In *Advances in Petri Nets 1989*, pages 1–29. Springer-Verlag, 1989.

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

[PS85] J. L Peterson and A. Silberschatz. *Operating System Concepts*. Addison-Wesley Inc., 1985.

[Seg95] R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, MIT, Dept. of Electrical Engineering and Computer Science, 1995. Also appears as technical report MIT/LCS/TR-676.

[Vaa91] F. W. Vaandrager. On the relationship between process algebra and input/output automata. In *Proceedings, Sixth Annual IEEE Symposium on Logic in Computer Science*, pages 387–398. IEEE Computer Society Press, July 1991.

[vBa] Franck van Breugel. private communication.

[vBb] Franck van Breugel. private communication.

[vBMOW03] Franck van Breugel, Michael Mislove, Joël Ouaknine, and James Worrell. An intrinsic characterization of approximate probabilistic bisimilarity. In *Proceedings of FOSSACS 03*, volume 2620 of *Lecture Notes In Computer Science*. Springer-Verlag, 2003.

[vBW01a] Franck van Breugel and James Worrell. An algorithm for quantitative verification of probabilistic systems. In K. G. Larsen and M. Nielsen, editors, *Proceedings of the Twelfth International Conference on Concurrency Theory - CONCUR'01*, number 2154 in Lecture Notes In Computer Science, pages 336–350. Springer-Verlag, 2001.

[vBW01b] Franck van Breugel and James Worrell. Towards quantitative verification of probabilistic systems. In *Proceedings of the Twenty-eighth International Colloquium on Automata, Languages and Programming*. Springer-Verlag, July 2001.

[vGSS95] R. van Glabbeek, S.A. Smolka, and B.U. Steffen. Reactive, generative, and stratified models of probabilistic processes. *Information and Computation*, 121(1):59–80, 1995.

[VN92] N. Viswanadham and Y. Narahari. *Performance Modeling of Automated Manufacturing Systems*. Prentice-Hall Inc, 1992.

[WSS97] S.-H. Wu, S.A. Smolka, and E. W. Stark. Composition and behaviors for probabilistic I/O automata. *Theoretical Computer Science*, 176(1–2):1–36, April 1997.