Available online at www.sciencedirect.com

SCIENCE DIRECT®

Electronic Notes in Theoretical Computer Science

ELSEVIER Electronic Notes in Theoretical Computer Science 87 (2004) 157–203

www.elsevier.com/locate/entcs

Labelled Markov Processes: Stronger and Faster Approximations

Vincent Danos

Université Paris 7, Paris

Josée Desharnais

Université Laval, Québec

Prakash Panangaden

University of Oxford, Oxford and
McGill University, Montréal

Abstract

This paper reports on and discusses three notions of approximation for Labelled Markov Processes that have been developed last year. The three schemes are improvements over former constructions [11,9] in the sense that they define approximants that capture more properties than before and that converge faster to the approximated process. One scheme is constructive and the two others are driven by properties on which one wants to focus. All three constructions involve quotienting the state-space in some way and the last two are quotients with respect to sets of temporal properties expressed in a simple logic with a greatest fixed point operator. This gives the possibility of customizing approximants with respect to properties of interest and is thus an important step towards using automated techniques intended for finite state systems, *e.g.*, model checking, for continuous state systems. Another difference between the schemes is how they relate approximants with the approximated process. The requirement that approximants should be simulated by the approximated process has been abandoned in the last scheme.

Keywords: Probability theory, Labelled Markov Processes, Approximation.

doi:10.1016/j.entcs.2004.09.018

Contents

1 Introduction

Labelled Markov Processes (LMPs) are probabilistic transition systems where the state space might be any general measurable space, in particular this includes situations where the state space may be continuous. They are essentially traditional discrete-time Markov processes enriched with a notion of interaction by synchronization on labels, familiar from process-algebras. They have been studied intensively in the last few years especially in relation with the question of bisimulation [7,10,11,21,20,12]. This is because they embody simple probabilistic interactive behaviours, and yet are rich enough to encompass many examples and to suggest interesting mathematics.

The initial motivation was the inclusion of continuous state spaces with a view towards eventual applications involving stochastic hybrid systems. An unexpected benefit of this additional generality has been the discovery that a simple temporal probabilistic logic, $\mathcal{L}_0$, captures a natural notion of equivalence between such processes, namely strong bisimulation. Remarkably this logic needs neither infinite conjunction, even though the systems may have even uncountable branching, nor negation nor any kind of negative construct (like the "must" modality). With this logical view, it became natural to think of the interplay between discrete structures (the logic) and the continuous mathematics of LMPs (measure and probability theory). This led to the important question of understanding what it means to be an approximation of a given LMP and especially of a "finite" approximant.

The approximation theory has developed along two lines. Desharnais et. al. [10] have developed a metric between LMPs which can be viewed as a "relaxation" of the notion of strong bisimulation. This metric can be used to say that one LMP "comes close to" behaving like another. The other direction was to develop a notion of "finite" approximant [11,9] and cast this in a domain theoretic setting. The papers just cited established that even a system with an uncountable state space could be approximated by a family of finite state processes. The family of approximants converge to the system being approximated in both metric and domain-theoretic senses. The approximations interact smoothly with the logic in the following sense. Any formulas of $\mathcal{L}_0$ that are satisfied by any approximant of P are satisfied by the process P itself and any formula satisfied by P is satisfied by some approximant.

At that point, there remained two soft spots in the approximation theory. First, while an approximant clearly ought to be some sort of finite quotient by temporal properties of the process being approximated, nobody so far was able to lay his hands on a precise way of phrasing just this intuition. Previous results state that every approximant satisfies some subset of the $\mathcal{L}_0$ properties that the process being approximated satisfies, but one does not have a way of saying in advance what these properties are. Second, another motivation for developing the theory further is that when fed with a finite process the approximation machinery was unable to retrieve the process itself in the limit. Instead, a bisimilar process was obtained. For instance, the pure loop process, with one state and one a-transition to itself, was approximated by all its finite unfoldings, *i.e.*, by chains of a-transitions; this seems spectacularly not what one would like to have intuitively, even if it is acceptable on the technical side (the infinite chain of a-transitions is bisimilar to the loop, after all).

The first approximation scheme that we present in this paper is an improvement of an unfolding scheme [9] that circumvent the second limitation. The two other approximation notions that we present are variants that overcome both limitations. They have been introduced in recent papers by Danos and Desharnais [5] and by Danos, Desharnais and Panangaden [6]. Concerning the former, we take a here a slightly different route and correct a mistake in the original paper. We also strengthen the latter notion by considering $\mathcal{L}_0$ extended with fixed points operators.

Specifically the first approach is based on the idea that the approximations can be "guided" by a family of formulas of interest. In other words, if there is a set of formulas of particular importance, one can construct a specific finite approximant geared towards these formulas. One can then be sure that the process in question satisfies a formula of interest if and only if the approximant does. Second, a much more compact representation is used so that loops are not unwound and convergence is attained more rapidly. A disadvantage is that the approximations obtained are not LMPs because the transition "probabilities" are not measures. Instead they are close to the measure-theoretic notion of capacity [2]. Capacities are not additive but they have instead a continuity property and are sub (or super) additive. Our LMP approximants will use superadditive and co-continuous set maps.

Actually, one can have the best of both worlds, keeping the flexibility of a customizable approach to approximation and staying at the same time within the realm of LMPs. This what the second approach does. It is based on a radical departure from the ideas of the previous approaches [5,9]. In these approaches one always approximated a system by ensuring that the transition probabilities in the approximant were below the corresponding transition in the full system. Here we approximate a system by taking a coarse-grained discretization (pixellization) of the state space and then using *average* values. This new notion is not based on the natural simulation ordering between LMPs as were the previous approaches.

Instead of simulation we use *conditional expectation*. This is a traditional construction in probability theory which, given a probability triple (S, Σ, p)

(sample space), a Σ-measurable random variable X (observation), and a sub-σ algebra Λ (pixellization of the sample space), returns the conditional expectation of X with respect to p and Λ. This conditional expectation is written $\mathbb{E}_p(X|\Lambda)$ and in some suitable sense is the best possible Λ-measurable approximation of X. The best will prove to be enough in our case, in that conditional expectations will construct for us low-resolution averages of any given LMP. Furthermore, an LMP will be known completely, up to bisimilarity, from its finite-resolution (meaning finite state) averages.

Moreover the new construction gives closer approximants in a sense that we will have to make precise later. They are also likely to be more robust to numerical variations in the system that one wants to approximate, since they are based on averages. Of course this is a speculative remark and needs to be thrashed out in subsequent work. To summarize, the new approximants are customizable, probabilistic and more accurate and possibly more robust as well. Beyond this construction, we would like to convey the idea that probability theory and its toolkit - especially the uses of averages and expectation values - are remarkably well adapted to a computationally-minded approach to probabilistic processes. It has a way of meshing finite and continuous notions of computations which is not unlike domain-theory. We expect far more interaction in the future between these theories than what is reported here. Work on probabilistic powerdomains [16] and integration on domains [13,14] provides a beginning. Curiously enough the bulk of work in probabilistic process algebra rarely ever mentions averages or expectation values. We hope that the present paper stimulates the use of these methods by others.

Acknowledgement

The first author wishes to thank Prakash Panangaden for the lecture series on Labelled Markov Processes given during his stay in Paris in May 2002 at the PPS lab, and for numerous instructive discussions about the subject as a whole. He also wishes to thank François Laviolette for suggesting many improvements to the ideas developed in this paper. Josée Desharnais and Prakash Panangaden have benefitted immensely from discussions with Vineet Gupta and Radha Jagadeesan. All of us have benefitted from the stimulating atmosphere at the Bellairs Institute of McGill University on Barbados and would like to thank all the participants of the workshop on Domain Theoretic Methods for Probabilistic Processes. Josée Desharnais would like to acknowledge the support of NSERC and FCAR. Prakash Panangaden would like to thank NSERC for support.

2 Preliminaries

This section is a brief reminder of the main objects of the trade with definitions slightly optimized for the development we have in mind. The paper is self-contained, though the reader might find useful to consult a book on basic probability theory, such as David Williams' book [22].

2.1 Notations

When S is a set and $A \subseteq S$, we write $\mathbf{1}_A$ for A's indicator function (this is sometimes called the characteristic function of A). When A, B are disjoint sets, we sometimes write $A + B$ for the (disjoint) union, and conversely each time we write $A + B$ it is understood that A and B are indeed disjoint. We write $\downarrow A_n$ when A_n is a decreasing sequence of sets, that is $A_n \supseteq A_{n+1}$ and $\cap A_n$ for the limit. Similarly we write $\uparrow A_n$ for an increasing sequence of sets A_n, *i.e.*, $A_n \subseteq A_{n+1}$ and $\cup A_n$ for the limit.

When $\mathfrak{R}$ is an equivalence relation over S, and $s \in S$, the equivalence class of s is written either $[s]_{\mathfrak{R}}$ or simply $[s]$, when $\mathfrak{R}$ is clear from the context. If A is a set of equivalence classes, one uses the usual set-theoretic notation for union $\cup A := \{s \in S \mid [s] \in A\}$. When $\mathfrak{R}$ is a binary relation over S, not necessarily an equivalence relation, one writes $\mathfrak{R}(s)$ for $\{t \mid (s,t) \in \mathfrak{R}\}$. One also says that a set A is $\mathfrak{R}$-*closed*, if whenever $s \in A$ and $(s,t) \in \mathfrak{R}$, $t \in A$, or in other words, if for all $s \in A$, $\mathfrak{R}(s) \subseteq A$.

2.2 Measurable spaces and Probabilities

A *measurable space* is a pair (S, Σ) where S is a set and $\Sigma \subset 2^S$ is a σ-*algebra* over S, that is, a set of subsets of S, containing S and closed under countable intersection and complement. Well-known examples are $[0,1]$ and $\mathbb{R}$ equipped with their respective *Borel* σ-algebras generated by the intervals which we will both denote by $\mathcal{B}$.

A map f between two measurable spaces (S, Σ) and (S', Σ') is said to be *measurable* if for all $A' \in \Sigma'$, $f^{-1}(A') \in \Sigma$. Writing $\sigma(f)$ for the σ-algebra generated by f, namely the set of subsets of the form $f^{-1}(A')$ with $A' \in \Sigma'$, one can rephrase this by saying $\sigma(f) \subseteq \Sigma$.

The set of measurable maps from (S, Σ) to $(\mathbb{R}, \mathcal{B})$ will be denoted $m\Sigma$. It is easily seen that a map f from (S, Σ) to $(\mathbb{R}, \mathcal{B})$ is in $m\Sigma$, if and only if for all $r \in \mathbb{R}$, $f^{-1}((r, +\infty)) \in \Sigma$ (or $f^{-1}([r, +\infty)) \in \Sigma$). This latter set is sometimes written $\{f > r\}$ ($\{f \geq r\}$). One says a map f from (S, Σ) to $(\mathbb{R}, \mathcal{B})$ is *simple*, if it can be written as $\sum_{i \leq k} a_i \mathbf{1}_{A_i}$, with $a_i \in \mathbb{R}$, and $A_i \in \Sigma$. Any such map can be rewritten $\sum_{i \leq k} a'_i \mathbf{1}_{A'_i}$ with the $A'_i \in \Sigma$ chosen to be pairwise disjoint, in which case $\{f > r\} = \cup_{\{i|a_i>r\}} A'_i$ which is in Σ, and one sees that all simple maps are in $m\Sigma$. We also take note, for further use, that countable infima preserves measurability. Indeed if f_n is a sequence in $m\Sigma$, then for all r, $\{\inf_n f_n \geq r\} = \cap_n \{f_n \geq r\}$

A *subprobability* on (S, Σ) is a map $p : \Sigma \to [0, 1]$, such that for any countable collection (A_n) of pairwise disjoint sets in Σ, $p(\bigcup_n A_n) = \sum_n p(A_n)$. A subprobability is an actual probability when in addition $p(S) = 1$. The condition on p is called *σ-additivity* and can be conveniently broken up into two parts:
— *additivity*: $p(A \cup A') = p(A) + p(B)$, for A, B disjoint,
— *continuity*: $\forall \uparrow A_n \in \Sigma : p(\cup A_n) = \sup_n p(A_n)$.

Let (S, Σ, p) be a probability triple, that is to say a measurable space (S, Σ) together with a probability p. A subset $N \subset S$ is said to be *negligible* if there exists a $A \in \Sigma$ such that $N \subseteq A$ and $p(A) = 0$.

We write $\mathcal{N}_p$ for p-negligible subsets. Two functions X, Y on (S, Σ, p) are said to be *almost surely equal*, written $X = Y$ a.s., if $\{s \in S \mid X(s) \neq Y(s)\} \in \mathcal{N}_p$. Sometimes we say p-a.s. equal if we wish to emphasize which measure we are talking about.

The subset of $m\Sigma$ consisting of the functions that are integrable with respect to p will be denoted by $\mathcal{L}^1(S, \Sigma, p)$. A last piece of notation that we will use is to write $X_n \uparrow X$ when X_ns and X are in $m\Sigma$, meaning that $X_n \leq X_{n+1}$ with respect to the pointwise ordering and X_n converges pointwise to X.

2.3 Labelled Markov Processes

We begin by defining the objects of interest:

Definition 2.1 [LMP] $\mathcal{S} = (S, \Sigma, h : L \times S \times \Sigma \to [0, 1])$ is a *Labelled Markov Process* (LMP) if (S, Σ) is a measurable space, and:
— for all $a \in L$, $A \in \Sigma$, $h(a, s, A)$ is Σ-measurable as a function of s;
— for all $s \in S$, $h(a, s, A)$ is a subprobability as a function of A.
A *pointed LMP* is an LMP with a distinguished state i, called the *initial state*, and is written $\mathcal{S} = (S, i, \Sigma, h)$.

Given a state property, one says a pointed LMP has this property if its initial state has it. For instance, one says two pointed LMPs are bisimilar when their two initial states are. After the traditional terminology in Markov chains, the map h is called the *kernel* or the transition probability function of $\mathcal{S}$. Most of the time, we will write $h(a, s, A)$ simply as $h_a(s, A)$. It is a measure of the likelihood that being at s and receiving a the LMP will jump to a state in A.

Some particular cases: 1) when S is finite and $\Sigma = 2^S$ we have the familiar probabilistic transition system, 2) when $h(a, s, A)$ does not depend on s or on a, we have the familiar (sub)probability triple. An example of the latter situation is $([0, 1], \mathcal{B}, h)$ with $h(a, s, B) = \lambda(B)$ with λ the Lebesgue measure on the collection $\mathcal{B}$ of Borel sets.

Equivalently LMPs can be defined as follows:

Definition 2.2 [LMP2] A Labelled Markov Process consists of a measurable space (S, Σ) and a family of Σ-measurable functions $(h(a, A))_{a\in L,A\in\Sigma}$ with values in $[0, 1]$, such that:
— additivity: for all disjoint A, B in Σ: $h(a, A \cup B) = h(a, A) + h(a, B)$;
— continuity: for all increasing sequence $\uparrow A_n$ in Σ: $h(a, \cup A_n) = \sup h(a, A_n)$.

From the definition it follows that for all a, s, one has $h(a, S)(s) \leq 1$.

In this second definition we view an LMP as a Σ-indexed family of Σ-measurable functions, namely the random variables "probability of jumping to A in one step labelled with a", instead of an S-indexed family of probabilities on Σ. Both definitions are related by $h(a, s, A) = h(a, A)(s)$ and we will use whichever is more convenient in the following sections. Another license we will take is not to mention actions when, as is often the case, they are not relevant in a particular example or proof, and simply write $h(s, A)$ or $h(A)(s)$.

2.4 Aside: analytic state spaces

In previous treatments, the LMP state space was required to be an analytic topological space [7]. Bisimulation and simulation can be defined either directly through behavioural conditions on kernels, or, indirectly, by using logical characterizations. These logical and behavioural definitions only coincide when the state space is analytic. In the present paper, one mostly uses the latter form of definition, and therefore one has no need for the analytic structure.

2.5 Temporal properties and simulation

LMPs differ from standard Markov chains in that the kernels are only asked to be subprobabilities and also depend on an auxiliary set L of actions. This seemingly small difference leads to a very different interpretation for them. They are construed as interactive processes which synchronize on labels and therefore one is interested in various notions of bisimulations and simulations as in non-deterministic process algebras [18].

The following "bisimulation logic" $\mathcal{L}_0$ is a central tool for asserting properties of LMPs:

$$\theta := \top \mid \ \theta \wedge \theta \mid \ \langle a \rangle_r \theta.$$

The *depth* $|\theta|$ of a formula θ is defined as: $|\top| = 0$, $|\theta_0 \wedge \theta_1| = \max(|\theta_0|, |\theta_1|)$ and $|\langle a\rangle_r \theta| = |\theta| + 1$.

Definition 2.3 Given an LMP $\mathcal{S}$, one may inductively define the map $[\![.]\!]_\mathcal{S} : \mathcal{L}_0 \to \Sigma$ as:
— $[\![\top]\!]_\mathcal{S} = S$,
— $[\![\theta_0 \wedge \theta_1]\!]_\mathcal{S} = [\![\theta_0]\!]_\mathcal{S} \cap [\![\theta_1]\!]_\mathcal{S}$,
— $[\![\langle a\rangle_r \theta]\!]_\mathcal{S} = \{s \in S \mid h_a(s, [\![\theta]\!]_\mathcal{S}) \geq r\}$

Sometimes, one needs a strict form of $\langle a\rangle_r \theta$, written $\langle a\rangle_{>r} \theta$, the semantics of which is defined as $[\![\langle a\rangle_{>r} \theta]\!]_\mathcal{S} = \{s \in S \mid h_a(s, [\![\theta]\!]_\mathcal{S}) > r\}$. This variant is used in the approximation construction of the next section but not otherwise. The logic $\mathcal{L}_0$ can also be added a disjunction, interpreted as the union, *i.e.* $[\![\theta_0 \vee \theta_1]\!]_\mathcal{S} = [\![\theta_0]\!]_\mathcal{S} \cup [\![\theta_1]\!]_\mathcal{S}$, and the resulting logic is called $\mathcal{L}_\vee$.

We write $s \models \theta$ to mean $s \in [\![\theta]\!]_\mathcal{S}$ and $\theta' \leq \theta$ to mean that θ' is a subformula of θ. Monoidal equations: $\theta_0 \wedge (\theta_1 \wedge \theta_2) = (\theta_0 \wedge \theta_1) \wedge \theta_2$, $\theta_0 \wedge \theta_1 = \theta_1 \wedge \theta_0$, $\theta \wedge \top = \theta$ all clearly preserve $[\![.]\!]_\mathcal{S}$.

Given a set $\mathcal{F}$ of formulas of $\mathcal{L}_0$, one can compare two states with respect to this set. We write sometimes $s \approx_\mathcal{F} t$ to mean that s and t satisfy the exact same formulas of $\mathcal{F}$.

The logic also induces a form of *simulation* between states in the sense that a state can be said to simulate another one if it satisfies at least the same formulas as the other does. The concept can be cast in behavioural terms as in the following definition.

Definition 2.4 [11] Let $\mathcal{S} = (S, \Sigma, h)$ be a LMP. A relation $\mathfrak{R}$ on S is a *simulation* if whenever $s\mathfrak{R}s'$, we have that for all $a \in L$ and every $\mathfrak{R}$-closed set $A \in \Sigma$, $h_a(s, A) \leq h_a(s', A)$. We say s is simulated by s' if $s\mathfrak{R}s'$ for some simulation relation $\mathfrak{R}$. If $\mathcal{S}$ and $\mathcal{T}$ are pointed LMPs, we say that $\mathcal{S}$ simulates $\mathcal{T}$ if the initial state of $\mathcal{S}$ simulates the initial state of $\mathcal{T}$.

This definition can be extended easily to simulation between states of different LMPs.

The notion of simulation meshes properly with the logic in the sense of the following proposition.

Proposition 2.5 [11] *If s simulates s', then for all formulas $\theta \in \mathcal{L}_0$, $s' \models \theta$ implies $s \models \theta$.*

If one adds disjunction to $\mathcal{L}_0$, the converse of this result is also true; that is, the simulation induced by the logic $\mathcal{L}_\vee$ is equivalent to Definition 2.4 (but this result uses analyticity of the state space [11]).

One can construct a family of metrics, d^c for $c \in (0, 1)$, all of them being closely related to $\mathcal{L}_0$, and which one can think of each as measuring the complexity of the simplest distinguishing formula between two states, if any.

We don't give here the precise definition of these metrics here, but we do want to use them to have neat convergence statement for approximants. For that matter, it is enough to use the following result, which is a direct consequence of results relating the logic and the metrics [10].

Proposition 2.6 *Let $(\mathcal{F}_i)_{i\in\mathbf{N}}$ be an increasing sequence of sets of formulas converging to the set of all formulas of $\mathcal{L}_0$. Let $\mathcal{S}$ be an LMP and $(\mathcal{S}_i)_{i\in\mathbf{N}}$ a sequence of LMPs. If $\mathcal{S}_i \approx_{\mathcal{F}_i} \mathcal{S}$ for every i, then for all $c \in (0,1)$*

$$d^c(\mathcal{S}_i, \mathcal{S}) \longrightarrow_{i\to\infty} 0.$$

With these preliminary definitions in place, we move on to the first approximation scheme.

3 Improved Constructive Approximation

In this section and the following, we propose two ways of approximating LMPs from below. The goal is to determine a family of finite processes that are simulated by the original LMP, and converge to it. We will work with pointed LMPs, because this is convenient when it comes to comparing an LMP with its approximations. The first approach, explained in the present section, depends on two parameters, the depth of observation and the accuracy of the probabilities. It improves on a former LMP approximation scheme based on an "unfolding" construction [9]. In this original scheme, as the approximation was refined, there were more and more transitions possible. We follow almost the same idea in the new construction. Some limitations of this first approach will be overcome by the second approach, given in the next section, and based on quotients with respect to sets of formulas.

The state-space is constructed as in the original scheme, but there will be "more" transitions possible, specifically some transitions will introduce cycles. As said, there are two parameters to the approximation: one is a natural number n, and the other is a positive rational ϵ. The number ϵ measures the accuracy with which the transition probabilities of the approximant approximate the transition probabilities of the original process. The parameter n represents the depth of our observation; in the original scheme, n was also the depth of the (acyclic) transition graph of the approximant itself, but it is no longer the case in the improved scheme, since it obtains cyclic graphs.

Given a labelled Markov process $\mathcal{S} = (S, \Sigma, h)$, a natural number n and a rational number $\epsilon > 0$, we define $\mathcal{S}^*(n,\epsilon)$ as an n-step unfolding approximation of $\mathcal{S}$. Its state-space is divided into $n+1$ levels which are numbered $0, 1, \ldots, n$. Bisimulation is the greatest fixed point of a suitable operator on relations and that one has - for each n - a level n approximation to bisimulation [9]. At each level, say n, the states of the approximant is a partition of S corresponding to what one might call n-bisimulation up to precision ϵ. The initial state of $\mathcal{S}^*(n, \epsilon)$ is at level n and transitions from a state of level l go to a state of level l or $l-1$. Thus, in particular, the unique state of level 0 either has no outgoing transitions or has a transition to itself. The main difference between the improved scheme and the original one is that transitions to states at the *same* level are now allowed.

In the following we omit the curly brackets around singletons.

Definition 3.1 Let (S, i, Σ, h) be a labelled Markov process, $n \in \mathbf{N}$ and ϵ a positive rational. We define the finite-state approximation $\mathcal{S}^*(n, \epsilon)$ as the tuple $(P, p_0, 2^P, \rho)$ where:

- P is a finite subset of $\Sigma \times \{0, \ldots, n\}$; the numbers from 0 to n correspond to the level of the states. States are defined by induction on their level.
  — At level 0 there is one state $(S, 0)$.
  — Now, given the states $(C_1, l), (C_2, l), \ldots, (C_m, l)$ at level l, we define states of level $l+1$ as follows. Let $(B_j)_{j \in I}$ be the partition
  $$\{\{0\}, (0, \epsilon/m], (\epsilon/m, 2\epsilon/m], \ldots\}$$
  of the interval $[0, 1]$ into intervals of size ϵ/m, where m is the number of states at level l. States of level $l+1$ are obtained by forming the coarsest common refinement of the partition $\{C_i\}_{i=1}^m$ and the partition generated by the sets $h_a(\cdot, C_i)^{-1}(B_j)$, for every set C_i and every $a \in L$, $j \in I$. If a set X is in this partition of S, $(X, l+1)$ is a state of level $l+1$.
- 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 $\mathcal{S}$.
- Transitions can happen between states of the same level, or from a state to a state of the preceding level, and the transition probability function is given as follows. Let $(X, l+1), (Y, l+1), (Z, l)$ be states of level $l+1$ and l, where $l \geq 0$. Then we set:
  $$\begin{aligned} \rho_a((X, l+1), (Y, l+1)) &:= \inf_{x \in X} h_a(x, Y) \\ \rho_a((X, l+1), (Z, l)) &:= \inf_{x \in X} h_a(x, Z) \\ &\qquad - \textstyle\sum_{i=1}^k \rho_a((X, l+1), (Z_i, l+1)) \end{aligned}$$
  where $\{Z_i\}_{i=1}^k$ is the unique partition of Z such that $(Z_i, l+1)$ is a state for

every i. Unspecified transitions are given the value 0.

The partition of S at level $l+1$ is defined in such a way that every state $x \in X$ (where X is a member of the partition) has probability within ϵ/m to jump to every set in the partition of level l (not necessarily true for transitions to states of level $l+1$). Intuitively, transitions are filled as follows: from a given state $(X, l+1)$, transitions to states at the same level are given the maximum possible probability (compatible with the condition of staying below all *simulating* states $x \in X$). This would not be sufficient to guarantee that the transition stays close to the corresponding transition of $\mathcal{S}$ because the partition of level $l+1$ is constructed with respect to states of level l. Since this condition is essential to preserve the accuracy of the approximation — and the statement of the lemma below reflects this — we complete the probability by adding transitions to states (Z, l).

Let us introduce here a few further notations. If $s \in S$, we denote by (X_s, l) the unique state at level l such that $s \in X_s$. We will write (Y, l) for the set $\{(Y_1, l), (Y_2, l), \ldots\}$, where $Y = \cup Y_j$; in this case, we often say that Y is a *union of sets at level* l and that the Y_i's *correspond to states of level* l. By extension, we will write $\rho_a((X, l+1), (Y, l))$ to mean $\sum_j \rho_a((X, l+1), (Y_j, l))$. The same notation will be used when we work with states of consecutive levels corresponding to the same subset of S: for example, we will write $(Y, l \cup l+1)$ to mean $\{(Y_1, l), (Y_2, l), \cdots, (Y'_1, l+1), (Y'_2, l+1), \cdots\}$, with $\cup Y_i = \cup Y'_i = Y$. Note that every set of level $l-1$ is a union of sets of level l because the partition of S at level l is a refinement of the partition at level $l-1$.

The following lemma uses crucially the fact that the partition of $[0,1]$ depends on the number of states m at the preceding level. This is because the kernels ρ_a are defined as infima, and therefore introduce a default of additivity, which one has to keep under control by refining the precision. The next section will offer a direct treatment of approximation via superadditive kernels.

Lemma 3.2 *Let $\mathcal{S}$ be a labelled Markov process, and $s \in S$. In $\mathcal{S}^*(n, \epsilon)$, if Y is a union of sets appearing at level l, then:*

$$0 < h_a(s, Y) - \rho_a((X_s, l+1), (Y, l \cup l+1)) \leq \epsilon.$$

Proof. The first inequality is trivial. Before proving the second one, note that the lemma is not necessarily true if Y is a union of sets appearing at the same level as $(X_s, l+1)$.

Let $s \in S$ and $(X_s, l+1)$, $(Y_i, l+1), (Y'_j, l)$, $i = 1, \ldots, k$, $j = 1, \ldots, k'$ be states of $\mathcal{S}^*(n, \epsilon)$ such that $Y = \cup_{i=1}^{k} Y_i = \cup_{j=1}^{k'} Y'_j$. Let m be the number of states at level l. Then for all $j = 1, \ldots, k'$ and $t \in X_s$ we have

$$|h_a(s, Y'_j) - h_a(t, Y'_j)| < \epsilon/m,$$

because of the way S is partitioned on level $l + 1$. Moreover, we have

$$\begin{aligned}
&\rho_a((X_s, l+1), (Y, l \cup l+1)) \\
&\quad = \rho_a((X_s, l+1), (Y, l)) + \rho_a((X, l+1), (Y, l+1)) \\
&\quad = \sum_{j=1}^{k'} \rho_a((X_s, l+1), (Y'_j, l)) + \sum_{i=1}^{k} \rho_a((X_s, l+1), (Y_i, l+1)) \\
&\quad = \sum_{j=1}^{k'} \inf_{s \in X_s} h_a(s, Y'_j))
\end{aligned}$$

and hence

$$\begin{aligned}
&|h_a(s, Y) - \rho_a((X_s, l+1), (Y, l \cup l+1))| \\
&\quad = |\sum_{j=1}^{k'} h_a(s, Y'_j) - \sum_{j=1}^{k'} \inf_{s \in X_s} h_a(s, Y'_j)| \\
&\quad \le \sum_{j=1}^{k'} |h_a(s, Y'_j) - \inf_{s \in X_s} h_a(s, Y'_j)| \\
&\quad \le \sum_{j=1}^{k'} \epsilon/m \\
&\quad \le \epsilon.
\end{aligned}$$

□

Since every transition probability of $\mathcal{S}^*(n, \epsilon)$ is smaller than in the corresponding transition in $\mathcal{S}$, then every state (X, l) in $\mathcal{S}^*(n, \epsilon)$ is simulated by every state $s \in X$ in $\mathcal{S}$.

Proposition 3.3 *Every labelled Markov process $\mathcal{S}$ simulates all its approximations of the form $\mathcal{S}^*(n, \epsilon)$. More precisely, every state (X, l) of $\mathcal{S}^*(n, \epsilon)$ $(l \le n)$ is simulated in $\mathcal{S}$ by every $s \in X$.*

Proof. The proof is conceptually easy but the notation necessary for the bookkeeping makes it hard to read. Let $\mathcal{S}^*(n, \epsilon) = (P, p_0, \rho)$ and let $\mathcal{U} = (U, u_0, \Omega, \nu)$ be the direct sum of $\mathcal{S}^*(n, \epsilon)$ and $\mathcal{S}$. Now let $\mathfrak{R}$ be the relation on U that relates a state (X, l) from $\mathcal{S}^*(n, \epsilon)$ to every state $s \in X$ from $\mathcal{S}$. We prove that $\mathfrak{R}$ is a simulation. Consider two related states, (X, l) and $s \in X$ and let $Z \in \Omega$ be $\mathfrak{R}$-closed, that is, $Z \cap S \in \Sigma$ and $\mathfrak{R}(Z \cap P) \subseteq Z$. We want to prove that $\nu_a((X, l), Z \cap P) \le \nu_a(s, Z \cap S)$. We will prove the inequality for Z^* a set containing $Z \cap P$ and defined as follows: Z^* is the smallest set containing $Z \cap P$ and satisfying the property that if it contains a state of level $l-1$, it contains every corresponding state of level l. This is possible because the partition of level l is finer than the one of level $l-1$. Of course, Z^* may contain some other state of level l that do not intersect states of level $l-1$.

The only transitions with positive probability from (X, l) are to states of level l and $l-1$ so we can assume that Z^* is a union of states of these levels, and hence it must be of the form

$$Z^* = (Y', l \cup l-1) \cup (Y, l),$$

where, as before, the notation (Y, l) may refer to a union of sets of level l. By the way Z^* is constructed, $Y \cup Y' \subseteq Z \cap S$. Then we have, by (the first inequality of) the preceding lemma

$$\begin{aligned}
\nu_a((X,l), Z \cap P) &\leq \rho_a((X,l), Z^*) \\
&= \rho_a((X,l), (Y', l \cup l-1)) + \rho_a((X,l), (Y,l)) \\
&\leq h_a(s, Y') + \sum_{i=1}^{k} \rho_a((X,l), (Y_i, l)) \qquad \text{where } \cup_{i=1}^{k} Y_i = Y \\
&= h_a(s, Y') + \sum_{i=1}^{k} \inf_{s \in X} h_a(s, Y_i) \\
&\leq h_a(s, Y') + h_a(s, Y) \\
&= h_a(s, Y' \cup Y) \\
&\leq \nu_a(s, Z \cap S)
\end{aligned}$$

and hence the result. □

The next theorem is the main result of this section. The proof is exactly the same as for the previous version of the construction except for the very last sequence of inequalities, which is adapted to the fact that transitions can happen between states of the same level. Notice that here we use a semantics for $\mathcal{L}_\vee$ with strict inequality in the modal formula.

Theorem 3.4 *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 $(X_s, n) \models \phi$.*

Proof. The proof is by induction on the structure of formulas. We prove the following stronger induction hypothesis. We prove that for all formulas ϕ there is an increasing sequence $(X_n)_{n \geq |\phi|}$ of sets in Σ which satisfy:

(i) $\cup_{n \geq |\phi|} X_n = [\![\phi]\!]_{\mathcal{S}}$;

(ii) $X_n = \cup_{s \in X_n} C_s$, where $(C_s, l) \in \mathcal{S}^*(n, 1/2^n)$ and $l \geq |\phi|$;

(iii) the states (C_s, l) satisfy ϕ in $\mathcal{S}^*(n, 1/2^n)$.

It is obvious for T with $X_n = S$ for all n.

Consider $\phi = \phi_1 \wedge \phi_2$. Assume the claim is true for ϕ_j, $j = 1, 2$. Let $(X^j_n)_{n \geq |\phi_j|}$ be the sequence for ϕ_j. Now define for $n \geq |\phi|$, the sequence

$$X_n = X^1_n \cap X^2_n.$$

Note that this is an increasing sequence of sets in Σ. We first prove (i): for all $s \models \phi$, there is some n such that $s \in X_n$. Choose $n = \max(n_1, n_2)$ where n_j is such that $s \in X^j_{n_j}$. Now for (ii) and (iii), let $s \in X_n$, for a fixed $n \geq |\phi|$. Then because all states (C_s, l) satisfy ϕ_j and $C_s \subseteq X^j_n$, we have $(C_s, l) \models \phi_1 \wedge \phi_2$ and $X_n = \cup_{s \in X_n} C_s$. The proof for the case $\phi_1 \vee \phi_2$ is similar.

Consider $\phi' = \langle a \rangle_{>q} \phi$, and assume the claim is true for ϕ. Let $d = |\langle a \rangle_{>q} \phi|$, $\epsilon_n = 1/2^n$ and let $(X_n)_{n \geq d-1}$ be the sequence for ϕ.

Now define for $n \geq d$, the sequence

$$Y_n = \cup \{C : (C, d) \in \mathcal{S}^*(n, \epsilon_n), \text{ and } \forall s \in C, h_a(s, X_n) > q + \epsilon_n\}.$$

This is an increasing sequence of sets in Σ because if $(C, d) \in \mathcal{S}^*(n, \epsilon_n)$ and $C \subseteq Y_n$, then for all $s \in C$ we have $h_a(s, X_{n+1}) \geq h_a(s, X_n) \geq q + \epsilon_n$. Moreover, if (C', d) is a state of $\mathcal{S}^*(n, \epsilon_{n+1})$ and $s, t \in C'$, then $h_a(t, X_{n+1}) > h_a(s, X_{n+1}) - \epsilon_{n+1} \geq q + \epsilon_n - \epsilon_{n+1} = q + \epsilon_{n+1}$.

We now prove (i), that is, for all $s \models \phi'$, there is some n such that $s \in Y_n$. So assume $h_a(s, [\![\phi]\!]) > q$. Then there is some n such that $h_a(s, X_n) - q > 2\epsilon_n$ because $h_a(s, \cdot)$ is a measure and X_n is an increasing sequence which converges to $[\![\phi]\!]$ and ϵ_n $(= 1/2^n)$ is decreasing to 0. Now since X_n is a union of states of level $l - 1 \geq d - 1$, then for every $t \in C_s$, with (C_s, l) a state of $\mathcal{S}^*(n, \epsilon_n)$ we have

$$|h_a(s, X_n) - h_a(t, X_n)| < \epsilon_n$$

and hence $h_a(t, X_n) - q > \epsilon_n$. Thus $C_s \subseteq Y_n$ and (i) and (ii) are proved. Note that the inequality sign in the meaning of the modal formula was crucial to this part of the proof.

We now prove (iii). Let $s \in Y_n$, for a fixed $n \geq d$. Then because all states $(X, l - 1)$, where $X \subseteq X_n$ and $l - 1 \geq d - 1$, satisfy ϕ and by Lemma 3.2, we have

$$\begin{aligned} \rho_a((C_s, l), ([\![\phi]\!]_{\mathcal{S}^*(n,\epsilon_n)}, l \cup l - 1)) &\geq \rho_a((C_s, l), (X_n, l \cup l - 1)) \\ &\geq h_a(s, X_n) - \epsilon_n \\ &> q + \epsilon_n - \epsilon_n = q, \end{aligned}$$

and hence, $(C_s, l) \models \phi'$ for all $l \geq d$ as wanted in (iii). □

The following results shows that a finite process is eventually approximated by itself. This is the main reason why we have introduced this new construction.

Corollary 3.5 *For every finite process there exists a bisimilar approximation.*

Proof. Since the process $\mathcal{S}$ is finite, the partition at the highest level of $\mathcal{S}^*(n, 1/2^n)$ must stabilize when n increases. In fact, it must converge to the bisimulation equivalence classes. Indeed, if two states are not bisimilar they must be distinguished by a formula ϕ. Then by the (proof of the) previous theorem there is some n such that the two states are not in the same set of $\mathcal{S}^*(n, 1/2^n)$. Thus the partition at the highest level corresponds exactly to the bisimulation equivalence classes. By construction of approximants, transitions from states of this level will only happen to states of this same level and hence the result. □

Corollary 3.6 *Let $\mathcal{S}$ be an LMP. Then for $c < 1$ we have*

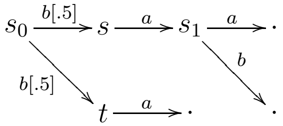
$$d^c(\mathcal{S}, \mathcal{S}^*(n, 1/2^n)) \to 0$$

and it is also true for $c = 1$ if the set of infinite sequences of non-bisimilar states starting in the initial state of $\mathcal{S}$ is of measure 0.

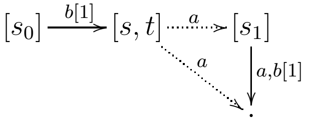
4 Abstract Approximations

Looking again at the construction above, one might wonder why Theorem 3.4 is so hard to prove. Indeed, we partition the state-space with respect to some depth and some accuracy of transition probability in such a way that one could think that the construction is faithful to formulas of the right depth and with probabilities that are a multiple of the accuracy. However, by taking the infimum, we lose some probabilities to unions of sets. This is because the infimum over a disjoint union is greater than or equal to the sum of infima over its parts. By not taking this into account, *i.e.*, by underestimating the transition probabilities to sets of states, we get slightly away from the logic. This is one reason why the logic must have a strict inequality sign in the modal formula. In the approximation scheme presented in this section, we will take all transitions into account and we will show how to quotient an LMP by a set of $\mathcal{L}_0$ formulas. A natural candidate for the quotient kernel is to take the infimum of the original kernel over equivalent states. This is what we have done in the preceding section, by defining the state-to-state transition using the infimum and then defining the transition probability to a set of states to be the sum of the transition probabilities to the individual states in the set; this way we manifestly have additivity. This leads to underestimating the transition probabilites to sets of states quite drastically. We can try to use the infimum to sets of states and get a perhaps better estimate of the approximate transition probabilities. Unfortunately this destroys additivity; if we take a set of probability distributions $\{\mu_i | i \in I\}$ and attempt to define a "measure" by $\mu(A) = \inf_{i \in I} \mu_i(A)$ we lose additivity. The following example illustrates why and also shows that taking infima over sets of states can give very bad estimates.

Example 4.1 Consider the following LMP, where unweighted transitions are of probability 1.



We want to quotient it with respect to the equivalence defined by all formulas of the form $\langle a \rangle_r \top$ and $\langle b \rangle_r \top$ for all $r \in [0, 1]$. The result is as follows:



Both dotted transitions are given value 0, for

$$\inf_{t \in [s]} h_a(t, \cup[s_1]) = \inf_{t \in [s]} h_a(t, \{s_1\}) = 0$$

and similarly for dead states. However $\inf_{t \in [s]} h_a(t, \{s_1, \cdot\}) = 1$. Hence the resulting transition probability function is not a measure and hence the quotient is not an LMP.

This example illustrates that if we want to take infima, we will have to weaken something in the objects of study. In the preceding section, the weakening was on the logical requirements, in this section we will weaken the notion of LMPs. This can be done because even if infima do not preserve additivity of measures, they do preserve super-additivity.

4.1 Pre-LMPs

The difference between a pre-LMP and an LMP lies in the following definition.

Definition 4.2 Given a measurable space (S, Σ), a function $f : \Sigma \to [0, 1]$ is called a *pre-measure* if:
— $\forall A, B \in \Sigma$ disjoint: $f(A + B) \geq f(A) + f(B)$;
— $\forall \downarrow A_n \in \Sigma : f(\cap A_n) = \inf_n f(A_n)$.

Easy consequences of the first condition are $f(\varnothing) = 0$ and monotonicity: $A \subseteq B \Rightarrow f(A) \leq f(B)$. The second property is a (co)continuity property. If one replaces the inequation in the first clause by an equation, the definition is equivalent to f being a sub-probability. Choquet introduced a similar notion under the name "capacity" [2] and realized the importance of keeping continuity while giving up additivity. This definition is weaker than Choquet's since he required both (upwards) continuity and (downwards) co-continuity.

Definition 4.3 A pre-LMP is a triple $\mathcal{S} = (S, \Sigma, h : L \times S \times \Sigma \to [0,1])$ where (S, Σ) is a measurable space, and for all $a \in L$, $s \in S$, $A \in \Sigma$: $h_a(s, .)$ is a pre-measure, and $h_a(., A)$ is measurable.

The intent of this definition is to use pre-LMPs as estimators for LMPs. It is not necessary that the estimation engine be of the same nature as what it tries to estimate. What we are interested in is how easy it is to handle and how well it estimates. Pre-LMPs turn out to be better estimators than LMPs as will be illustrated in Proposition 4.19.

4.2 Temporal properties

Semantics of $\mathcal{L}_0$ still makes sense with pre-LMPs.

Lemma 4.4 *For all pre-LMP $\mathcal{S}$ and $\theta \in \mathcal{L}_0$: $[\![\theta]\!]_{\mathcal{S}} \in \Sigma$.*

Proof. Easy induction on $\mathcal{L}_0$. □

To the modal operator of $\mathcal{L}_0$, namely $\langle a \rangle_r$, a family of maps is naturally associated, still written $\langle a \rangle_r : \Sigma \to \Sigma$ and called the *shifts*:

$$\langle a \rangle_r(A) := \{s \in S \mid h_a(s, A) \geq r\}.$$

Clearly $\langle a \rangle_r(A) = h_a(., A)^{-1}([r, 1])$, and $h_a(., A)$ being measurable for all $A \in \Sigma$, one has that $\langle a \rangle_r(A) \in \Sigma$ (actually, $h_a(., A)$ is measurable iff for all r, $\langle a \rangle_r(A) \in \Sigma$). As said in the preceding section, one can also define the strict shifts as $\langle a \rangle_{>r}(A) := \{s \mid h_a(s, A) > r\}$, which are endomaps of Σ as well.

With this new notation: $[\![\langle a \rangle_r \theta]\!]_{\mathcal{S}} = \langle a \rangle_r([\![\theta]\!]_{\mathcal{S}})$.

Actually a much stronger statement than the lemma above can be made:

Theorem 4.5 *[3] Let (S, Σ, h) be an LMP, the σ-algebra generated by $([\![\theta]\!]_{\mathcal{S}})_{\theta \in \mathcal{L}_0}$ is the smallest sub-σ-algebra of Σ which is closed under the shifts $\langle a \rangle_r$.*

The theorem deserves mention because it gives purely measure-theoretic status to $\mathcal{L}_0$ and besides, we will use it again in the next section.

4.3 Co-simulation morphisms

The following notion of morphism between pre-LMPs will witness the relation between a process and its approximant. Recall that our goal is to define approximants as quotients of pre-LMPs under equivalence relations. Such quotients are usually related to the original process $\mathcal{S}$ through a measurable map from $\mathcal{S}$ to its quotient; this map will be proven to be a co-simulation morphism.

Definition 4.6 Given $\mathcal{S}$, $\mathcal{S}'$ two pre-LMPs, a map $q : S \to S'$ is said to be a *co-simulation* iff it is surjective, measurable and for all $a \in L$, $s \in S$, $A' \in \Sigma'$:

$$h_a(s, q^{-1}A') \geq h'_a(q(s), A').$$

If $\mathcal{S}$ and $\mathcal{S}'$ are pointed LMPs with initial states i and i', then one asks additionally that $q(i) = i'$.

Caveat: we are changing the original definition of simulation morphisms [8], reversing the inequation and requiring surjectivity. Nevertheless, we can use the proof of the dual result with simulation morphisms [8, Proposition 3.6.7] because it does not use the additivity property.

Proposition 4.7 *If $q : \mathcal{S} \to \mathcal{S}'$ is a co-simulation morphism, then every $s \in S$ simulates $q(s)$.*

This proposition will allow us to make sure that the approximant is simulated by (or is *below*) $\mathcal{S}$.

Proposition 2.5 can also be extended to pre-LMPs.

Corollary 4.8 *Let $q : \mathcal{S} \to \mathcal{S}'$ be a co-simulation, then for all $\theta \in \mathcal{L}_0$, $s \in S$: $q(s) \in [\![\theta]\!]_{\mathcal{S}'} \Rightarrow s \in [\![\theta]\!]_{\mathcal{S}}$.*

Proof. The statement can be restated as $q^{-1}[\![\theta]\!]_{\mathcal{S}'} \subseteq [\![\theta]\!]_{\mathcal{S}}$. The proof is by induction on $\mathcal{L}_0$:
— for $\top$, one has $[\![\theta]\!]_{\mathcal{S}'} = S'$ and $q^{-1}S' = S = [\![\theta]\!]_{\mathcal{S}}$;
— $q^{-1}[\![\theta \wedge \psi]\!]_{\mathcal{S}'} = q^{-1}([\![\theta]\!]_{\mathcal{S}'} \cap [\![\psi]\!]_{\mathcal{S}'}) = q^{-1}([\![\theta]\!]_{\mathcal{S}'} \cap [\![\psi]\!]_{\mathcal{S}'}) = q^{-1}[\![\theta]\!]_{\mathcal{S}'} \cap q^{-1}[\![\psi]\!]_{\mathcal{S}'} \subseteq [\![\theta]\!]_{\mathcal{S}} \cap [\![\psi]\!]_{\mathcal{S}}$;
— if $q(s) \in [\![\langle a\rangle_r \theta]\!]_{\mathcal{S}'}$, then $s \in [\![\langle a\rangle_r \theta]\!]_{\mathcal{S}}$ because:

$$r \leq h'_a(q(s), [\![\theta]\!]_{\mathcal{S}'}) \leq h_a(s, q^{-1}[\![\theta]\!]_{\mathcal{S}'}) \leq h_a(s, [\![\theta]\!]_{\mathcal{S}}).$$

□

4.4 The infimum construction

Proposition 4.11 below, which says when "one can take infima" over equivalence classes, is important in the sense that without it we could not construct any quotient.

Definition 4.9 [Compatible equivalences] Let (S, Σ) be a measurable space, $\mathfrak{R}$ be an equivalence relation on S, and A be in Σ. One defines the *closure* of A as:

$$[A]^+ := \{s \mid [s] \cap A \neq \varnothing\},$$

and one says the relation $\mathfrak{R}$ is *compatible with* (S, Σ) if:

$$\forall A \subseteq S : A \in \Sigma \rightarrow [A]^+ \in \Sigma.$$

Take note that this compatibility condition is weaker than asking all $\mathfrak{R}$-closed subsets of S to be in Σ. For instance, if $\mathfrak{R}$ is the identity, then the latter condition is satisfied only if the measurable space is discrete (*i.e.*, $\Sigma = 2^S$), whereas the former is always trivially true, since $[A]^+ = A$.

The closure of A provides the best upper approximation of A, within the $\mathfrak{R}$-closed subsets of S. One may also define a best lower approximation $[A]^- := \{s \mid [s] \subseteq A \neq \varnothing\}$. Since $[A]^- = ([A^c]^+)^c$, it is also measurable and both approaches are equivalent.

Example 4.10 There is no reason in general why $\mathfrak{R}$ should be compatible, but sometimes it is. An important example is when $\mathfrak{R}$ has countably many classes, all in Σ. Then, since $[A]^-$ is the union of all classes contained in A, it is measurable.

As another example consider $([0, 1], \mathcal{B})$, and take $\mathcal{R}$-closed sets to be the subsets of $[0, 1]$ closed under some $\phi : [0, 1] \to [0, 1]$ such that $\phi(\Sigma) \subseteq \Sigma$. Then $[A]^+ = \cup A_n$, with $A_0 = A$, and $A_{n+1} = A_n \cup \phi(A_n)$, the sequence of successive one-step closures. For instance, if one takes ϕ to be the symmetry $\lambda x.(1 - x)$, $[A]^+$ is just the closure under symmetry and is obtained in one step.

One now wants to extend this notion of lower approximation of sets in Σ to functions in $m\Sigma$.

Proposition 4.11 *Let (S, Σ) be a measurable space, $\mathfrak{R}$ be an equivalence compatible with (S, Σ), and g be a bounded function in $m\Sigma$, the function $[g]^-(s) := \inf_{t \in [s]} g(t)$ is itself in $m\Sigma$.*

Proof. The argument decomposes in two parts. We suppose first g is a simple function. As such, it can always be written as $\sum_{i \leq k} a_i \mathbf{1}_{A_i}$, with a_i

strictly increasing, and A_is pairwise disjoint and all in Σ. Define the decreasing sequence $B_i := \{s \mid [s] \subseteq \cup_{i\leq j\leq k} A_j\}$. Since $B_i = [\cup_{i\leq j\leq k} A_j]^-$, and $\mathfrak{R}$ is compatible, B_i is in Σ. Set now $C_i = B_i \setminus B_{i+1}$, one has:

$$[g]^- = \textstyle\sum_{i\leq k} a_i \mathbf{1}_{C_i}$$

which indeed is a simple measurable function, since again C_is are all in Σ.

For the general case of a bounded function, we may suppose without loss of generality, that $0 < g \leq 1$. We then define the following sequence of simple functions:

$$g_n := \textstyle\sum_{i=0}^{2^n-1} (i+1)2^{-n} \mathbf{1}_{\{i2^{-n} < g \leq (i+1)2^{-n}\}}$$

This sequence is decreasing and converging pointwise to g, that is to say, for all s, $g(s) = \inf_n g_n(s)$. One has:

$$\begin{aligned} [g]^-(s) &:= \inf_{t\in[s]} \inf_n g_n(t) \\ &= \inf_n \inf_{t\in[s]} g_n(t) \\ &= \inf_n [g_n]^-(t) \end{aligned}$$

and now $[g]^-$ is expressed as an infimum of a countable family of functions, the $[g_n]^-$, which we know from the first part of the argument are all measurables, and is therefore itself measurable. □

Note that the second part of the argument uses boundedness to approach g from above. We don't know if that additional assumption about g can be lifted. Anyway, it is not a constraint in the application to LMP kernels, since these have values in $[0, 1]$. Apart from that, the argument uses no assumption on g. A similar argument can be made for the supremum based dual construction, obtaining a measurable $[g]^+$.

To understand the necessity of the compatibility condition, let us consider another example. Take $g = \mathbf{1}_A$, and $\mathfrak{R}$ an equivalence relation on S. Then it is readily seen that $[g]^- = \mathbf{1}_{[A]^-}$. So $[\mathbf{1}_A]^-$ will be measurable if and only if $[A]^- \in \Sigma$.

4.5 Aside: another infimum construction

A comparable construction was given in the paper where the notion of pre-LMP was first defined [5]. That one did not rely on a compatibility assumption on the equivalence relation, but on the assumption that the equivalence is countably generated.

Specifically, one says an equivalence $\mathfrak{R}$ on (S, Σ) is countably generated, if there exists a countable family $\mathcal{F} \subseteq \Sigma$, such that $s\mathfrak{R}t$ if and only if for all $A \in \mathcal{F}$, $s \in A$ if and only if $t \in A$. This is the case for all equivalence relations generated by a choice of formulas in $\mathcal{L}_0$, and thus seems a good working assumption regarding the particular application we have in mind.

Suppose now given a countably generated equivalence relation $\mathfrak{R}$ on (S, Σ). One can always exhibit an increasing sequence of finite families $\mathcal{F}_i \subseteq \Sigma$, such that $\cup_i \mathcal{F}_i = \mathcal{F}$ and $\mathcal{F}$ generates $\mathfrak{R}$. Each $\mathcal{F}_i$ finitely generates an equivalence $\mathfrak{R}_i$, which is compatible as said earlier, and therefore the $[.]^-$ construction works fine. One may then define, for any $g \in m\Sigma$:

$$\begin{aligned} g_i &:= [g]^-_{\mathfrak{R}_i} \\ g_\star &:= \sup_i g_i \end{aligned}$$

It is readily seen that: 1) $g_\star$ does not depend on the particular choice of the increasing sequence $\mathcal{F}_i$, that 2) it is Σ-measurable and 3) it is constant on $\mathfrak{R}$ classes. Clearly when both $g_\star$ and $[g]^-$ exist, $g_i \leq [g]^-$, so $g_\star \leq [g]^-$.

It was claimed wrongly in the original paper [5, Prop.13] that $g_\star = [g]^-$. Here is an example showing that sometimes this might be a strict inequality. Take $([0,1], \mathcal{B})$ as measurable space, $\mathcal{F}_i = \{[0, 1/j]; i \geq j > 0\}$ as the generating set, and $g(0) = 1$ and $g(s \neq 0) = 0$. One has $g_i = 0$, so $g_\star = 0$, while 0 is alone in its class and therefore $[g]^- = g$.

This new lower approximation of g could be an alternative to $[g]^-$ when $\mathfrak{R}$ is not compatible but countably generated. The bad news about it is that if we use it in Definition 4.12 below, we do not get Lemma 4.13, as co-continuity is false.

Here we choose to work with compatible relations, and will restrict to quotients induced by finite sets of formulas. Accordingly, in the rest of the paper, we will simply say that $\mathfrak{R}$ is an equivalence on a given pre-LMP (S, Σ, h), to actually mean that $\mathfrak{R}$ is compatible with (S, Σ).

4.6 Quotients and Simulations

Definition 4.12 Given an equivalence $\mathfrak{R}$ on a pre-LMP $\mathcal{S} = (S, \Sigma, h)$, we define the *quotient* pre-LMP, written $\mathcal{S}_\mathfrak{R}$, as the following triple $(S_\mathfrak{R}, \Sigma_\mathfrak{R}, h_\mathfrak{R})$:
—$S_\mathfrak{R}$ is the set of $\mathfrak{R}$ equivalence classes,
—$\Sigma_\mathfrak{R}$ is the quotient σ-algebra of $\mathfrak{R}$-closed sets of Σ,
—$h_\mathfrak{R}(a, [s], A) := \inf_{t \in [s]} h_a(t, \cup A)$ for $a \in L$, $s \in S$ and $A \in \Sigma_\mathfrak{R}$. If $\mathcal{S}$ has an initial state, then its equivalence class is the quotient initial state.

When the kernel and the equivalence matches exactly (in the sense that equivalent states have equal transition probabilities to unions of equivalence classes), then $h_{\mathfrak{R}}([s], A) = h(t, \cup A)$ for all $t \in [s]$ and the construction boils down to an ordinary bisimulation quotient.

We have seen in Example 4.1 that this quotient does not always define an LMP. However it does define a pre-LMP.

Lemma 4.13 *$\mathcal{S}_{\mathfrak{R}}$ as defined above is a pre-LMP.*

Proof. We have three things to verify according to Definition 4.3 above. The first is obvious. For the second condition, the verification that $h_{\mathfrak{R}}(a, [s], .)$ is a pre-measure breaks down in two subconditions. (We drop the labels since they play no role in the argument.)
Super-additivity. If A, B are disjoint sets in $\Sigma_{\mathfrak{R}}$:

$$\begin{aligned} h(s, \cup(A + B)) &= h(s, \cup A + \cup B) \\ &\geq h(s, \cup A) + h(s, \cup B) \\ &\geq h_{\mathfrak{R}}([s], A) + h_{\mathfrak{R}}([s], B). \end{aligned}$$

Co-continuity. Let $\downarrow A_n$ be a decreasing sequence of sets in $\Sigma_{\mathfrak{R}}$, then $\downarrow \cup A_n$ is also a decreasing sequence of $\mathfrak{R}$-closed sets of Σ and:

$$\begin{aligned} h_{\mathfrak{R}}([s], \cap A_n) &:= \inf_{t\in[s]} h(t, \cap(\cup A_n)) \\ &= \inf_{t\in[s]} \inf_n h(t, \cup A_n) \\ &= \inf_n \inf_{t\in[s]} h(t, \cup A_n) \\ &=: \inf_n h_{\mathfrak{R}}([s], A_n) \end{aligned}$$

so indeed $h_{\mathfrak{R}}([s], .)$ is a pre-measure.

Finally for the third, we verify that for all $A \in \Sigma_{\mathfrak{R}}$ and $r \in \mathbb{R}$, the set $\{h_{\mathfrak{R}}(., A) \geq r\}$ is in $\Sigma_{\mathfrak{R}}$. Writing q for the canonical projection from S to $S_{\mathfrak{R}}$, we can write our set as:

$$\{[s] \mid h_{\mathfrak{R}}([s], A) \geq r\} = q(\{s \mid \inf_{t\in[s]} h(t, q^{-1}A) \geq r\})$$

i.e., as the projection of a set which is clearly $\mathfrak{R}$-closed and, by Proposition 4.11 applied to $h(., q^{-1}A)$ (which indeed is a measurable function, since q is measurable and therefore $q^{-1}A \in \Sigma$), belongs to Σ. □

Clearly:

Proposition 4.14 *$\mathcal{S}_{\mathfrak{R}}$ is simulated by $\mathcal{S}$. Specifically, the canonical surjection $q : \mathcal{S} \to \mathcal{S}_{\mathfrak{R}}$ is a co-simulation.*

4.7 Quotients and Logical Properties

Now that we know the quotient $\mathcal{S}_{\mathfrak{R}}$ exists, we need to bring up the properties it might share with $\mathcal{S}$. Combining Proposition 4.14 with Corollary 4.8, we get that each property that $\mathcal{S}_{\mathfrak{R}}$ satisfies is also satisfied by $\mathcal{S}$.

Corollary 4.15 *Let $\mathfrak{R}$ be an equivalence on $\mathcal{S}$, then for all $\theta \in \mathcal{L}_0$, and $s \in S$: $[s] \in [\![\theta]\!]_{\mathcal{S}_{\mathfrak{R}}} \Rightarrow s \in [\![\theta]\!]_{\mathcal{S}}$.*

We now need a converse to this, that will quantify how good the approximation given by the quotient is, and say how much of the $\mathcal{L}_0$ properties of s in $\mathcal{S}$ are still properties of $[s]$ in $\mathcal{S}_{\mathfrak{R}}$.

Definition 4.16 We will say $\mathfrak{R}$ *refines* a property θ if and only if all interpretations of subformulas of θ are $\mathfrak{R}$-closed.

In other words: for all $\theta' \leq \theta$ and all $(s, t) \in \mathfrak{R}$, if $s \models \theta'$ then $t \models \theta'$.

Proposition 4.17 *Let $\mathfrak{R}$ be an equivalence on $\mathcal{S}$, then $\forall \theta \in \mathcal{L}_0$ that $\mathfrak{R}$ refines, $s \in S$: $s \in [\![\theta]\!]_{\mathcal{S}} \Rightarrow [s] \in [\![\theta]\!]_{\mathcal{S}_{\mathfrak{R}}}$.*

Proof. The lemma can be rephrased as $q^{-1}[\![\theta]\!]_{\mathcal{S}_{\mathfrak{R}}} \supseteq [\![\theta]\!]_{\mathcal{S}}$. We prove it by induction on θ. The only interesting case is when $\theta = \langle a \rangle_r \psi$, and then one has for all a, s:

$$\begin{aligned} h_{\mathfrak{R}}(a, [s], [\![\psi]\!]_{\mathcal{S}_{\mathfrak{R}}}) &= \inf_{t \in [s]} h(a, t, q^{-1}[\![\psi]\!]_{\mathcal{S}_{\mathfrak{R}}}) \\ &= \inf_{t \in [s]} h(a, t, [\![\psi]\!]_{\mathcal{S}}) \end{aligned}$$

where the second equation is by induction (since $\mathfrak{R}$ refines also ψ, $q^{-1}[\![\psi]\!]_{\mathcal{S}_{\mathfrak{R}}} \supseteq [\![\psi]\!]_{\mathcal{S}}$ and by the corollary above, these two subsets of S are actually equal). It follows that if $s \models \theta$ and $[s] \not\models \theta$, there must be a $t \in [s]$ close enough to the infimum, such that $t \not\models \theta$ either, which means $\mathfrak{R}$ actually does not refine θ. □

Subformulas *have* to be included in the refinement condition and this can be seen on a small example. Say $r \geq u$, here are $\mathcal{S}$ and a quotient $\mathcal{S}_{\mathfrak{R}}$:

$$\begin{tikzcd}
s_0 \arrow[rr, "{a[r]}"] \arrow[dr, dotted] & & s \arrow[rr, "{a[u]}"] \arrow[dr, dotted] & & t \arrow[dl, dotted] \\
& {[s_0]} \arrow[rr, "{a[r]}"] & & {[s, t]} &
\end{tikzcd}$$

Now set $\theta = \langle a\rangle_r\langle a\rangle_u\top$; one has:

$$[\![\langle a\rangle_u\top]\!]_{\mathcal{S}} = \{s, s_0\},\ [\![\theta]\!]_{\mathcal{S}} = \{s_0\},$$
$$[\![\langle a\rangle_u\top]\!]_{\mathcal{S}_{\mathfrak{R}}} = \{[s_0]\},\ [\![\theta]\!]_{\mathcal{S}_{\mathfrak{R}}} = \varnothing,$$

so though $[\![\theta]\!]_{\mathcal{S}}$ is $\mathfrak{R}$-closed and $s_0 \in [\![\theta]\!]_{\mathcal{S}}$, yet $q_{\mathfrak{R}}(s_0) = [s_0] \notin [\![\theta]\!]_{\mathcal{S}_{\mathfrak{R}}}$.

Combining the last two statements in the particular case of logically generated approximations, we get:

Theorem 4.18 (abstract approximants) *Let $\mathcal{F}$ be a finite downward closed subset of $\mathcal{L}_0$, and $\mathfrak{R}$ be the associated equivalence on $\mathcal{S}$:*

$$\forall\theta \in \mathcal{F}, \forall s \in S : s \in [\![\theta]\!]_{\mathcal{S}} \Leftrightarrow [s]_{\mathfrak{R}} \in [\![\theta]\!]_{\mathcal{S}_{\mathfrak{R}}}.$$

Proof. $\mathfrak{R}$ is an equivalence on $\mathcal{S}$ because it is has finitely many classes, which are all measurable, and hence Corollary 4.15 applies. Now $\mathfrak{R}$ refines $\mathcal{F}$ and is even the coarsest such equivalence, and one may apply Proposition 4.17. □

A particular case is $\mathcal{F} = \{\top\}$, and then $\mathcal{S}_{\mathfrak{R}} = (\{*\}, \{\varnothing, \{*\}\}, h_a)$ with $h_a(*, \varnothing) = 0$ and $h_a(*, \{*\}) = \inf_{s\in S} h_a(s, S) =: \alpha_a$. So $\mathcal{S}_{\mathfrak{R}}$ is the loop with coefficients $(\alpha_a)_{a\in L}$. Of course very few properties are retained here, namely the combinations of $\langle a\rangle_r$ with $r \leq \alpha_a$. This trivial approximation can be thought of as a quite blunt abstract interpretation of $\mathcal{S}$. The theorem above explains, in essence, how to construct arbitrarily sharper ones. Note also that the set of formulas that are satisfied by a state of the quotient is not necessarily included in $\mathcal{F}$: it may satisfy more formulas.

To go beyond quotients by finite sets of formulas, one has to take care of compatibility of the generated equivalence. This has to be verified on a case-by-case study. Here is an example of a non-finite quotient. Go back to the example where the underlying measurable space is $([0, 1], \mathcal{B})$, and take $h(s, A) = s.\lambda(A)$.

4.8 Abstract approximants are optimal

A noteworthy observation is that if one wants the quotient map q to generate a simulation, the choice made above for $h_{\mathfrak{R}}$ is optimal, $\Sigma_{\mathfrak{R}}$ is the largest σ-algebra that will make q measurable and all other kernels would be pointwise smaller:

Proposition 4.19 *Given $\mathcal{S}$, $\mathcal{S}'$ two pre-LMPs and $q : \mathcal{S} \to \mathcal{S}'$ a co-simulation morphism, and defining the equivalence relation generated by q on S as $(s, t) \in \mathfrak{R}$ if $q(s) = q(t)$, one has:*

1) Σ' *is a sub-*σ*-algebra of* $\Sigma_{\mathfrak{R}}$*;*
2) for all $s' \in S'$*,* $A' \in \Sigma'$*:* $h'(s', A') \leq h_{\mathfrak{R}}(s', A')$*.*

Yet another way of saying this is: the identity $\iota : \mathcal{S}_{\mathfrak{R}} \to \mathcal{S}'$ is a co-simulation which decomposes q as $\iota \circ q_{\mathfrak{R}}$. The proof of 1) is left to the reader. Point 2) is obvious. Note that $\mathfrak{R}$, the equivalence associated to q, is not in general an equivalence on $\mathcal{S}$, since it might not be compatible with (S, Σ). All we are saying here, is that when $h_{\mathfrak{R}}$ is measurable, it is the best pre-LMP for the ordering generated by co-simulations. To make this a more satisfying statement, one would have to add some topological or domain-theoretic structure on state spaces, to make sure the construction is always possible.

To wrap up, we now know for one thing, that pre-LMP support what seems the natural construction, as summarized in Theorem 4.18, whereas with plain LMPs one has to restrict to finite quotients. And with this last proposition, we see that pre-LMPs also give more accurate finite predictors.

4.9 The sup-quotient

We have shown how to construct quotients of pre-LMPs using infima of measurable functions. One could be interested in the dual construction using suprema. All the results above can be dualized to their supremum counterpart with little modification. Basically, one has to reverse inequality signs and replace co-continuity with continuity. The resulting model could be called conveniently *sub-pre-LMP*, since suprema generate subadditive kernels (and our pre-LMPs should then be called super-pre-LMPs since they have superadditive kernels as we know). The quotient of a sub-pre-LMP is above the original process instead of below. Consequently, we have a simulation morphism instead of a co-simulation in the equivalent of Proposition 4.14. The semantics has to be adapted as well, replacing $\langle a \rangle_r(A)$ with the strict version $\langle a \rangle_{>r}(A)$.

4.10 Stronger Approximants

We now extend the results of this section to a logic with fixed points. A more detailed account of this fixed point logic was given elsewhere [4]. This extension will allow us to approximate with respect to a much richer class of properties. More to the point, we have seen in section 3 how the introduction of loops in the approximants allows for quicker convergence when there are loops in the transition graph of the original process. In the present section we have just shown how the approximations may be guided by formulas. However, the formulas used only capture one step transitions and one needs richer formulas to capture looping behaviour. In fact one needs exactly the fixed point logic of this section.

4.10.1 Extended logic

We introduce an extended logic $\mathcal{L}_0^*$ to capture cyclic temporal properties. To deal with mutual fixed point equations, it is convenient to present the extended formulas as automata.

Definition 4.20 [cyclic temporal properties] An $\mathcal{L}_0^*$ formula is a pair (I, λ), with I a finite indexing set and λ a partial map from $L \times I \times I$ to $[0, 1]$.

We write $\mathrm{dom}(\lambda)$ for the domain of λ; working with total maps, by extending λ to be zero outside $\mathrm{dom}(\lambda)$, turns out to be inconvenient. We will use freely the automaton terminology and talk about I as the state space and λ as the transition map. Notice that there is no condition on the transition function: it need not be a subprobability distribution. One should understand the transitions as if they were non-deterministic.

First of all we show how to present our usual $\mathcal{L}_0$ formulas as automata.

Definition 4.21 [mapping $\mathcal{L}_0$ to $\mathcal{L}_0^*$] One defines a map $(.)^*$ from $\mathcal{L}_0$ to $\mathcal{L}_0^*$ as follows:
— I is the set of θ's (occurrences of) maximal conjunctive sub-formulas,
— $\lambda(a, \theta_0, \theta_1) = r$ iff $\theta_0 = \langle a\rangle_r \theta_1 \wedge \theta'$ for some θ', up to the monoidal equations associated to $\wedge$.

The simplest example is $\top^* = (\{\top\}, \varnothing)$. The next simplest is $(\langle a\rangle_{.5}\top)^* = (\{\langle a\rangle_{.5}\top, \top\}, \{(a, \langle a\rangle_{.5}\top, \top, .5)\})$, or in graphical automata notation:

$$\langle a\rangle_{.5}\top \xrightarrow{a[.5]} \top\,.$$

A more complicated formula is $\langle a\rangle_1\theta \wedge \langle b\rangle_{.5}\top$, which translates (again in graphical automata notation) to:

$$\theta^* \xleftarrow{a[1]} \langle a\rangle_1\theta \wedge \langle b\rangle_{.5}\top \xrightarrow{b[.5]} \top\,.$$

This correspondence is one-one, up to monoidal equations, and θ^* is always a tree.

Now, given $\mathcal{S}$ an LMP, we would like to extend the map $[\![.]\!]_{\mathcal{S}}$ to $\mathcal{L}_0^*$-formulas, or in other words, to make sense of $s \models \theta$ for our new formulas. This will be done using two independent approaches that will turn out to be equivalent. One will be the definition of a suitable fixed point in the category $\mathbf{C}_{\mathcal{S}}$ defined below, and the other one will be in terms of simulation relations.

4.11 Semantics of $\mathcal{L}_0^*$ via fixed points

Let $\mathbf{C}_\mathcal{S}$ be the sub-Cartesian category of **Set** generated by:
— cartesian powers of Σ: $1 = \Sigma^0$, Σ, …, Σ^n, …
— shifts $\langle a\rangle_r : \Sigma \to \Sigma$,
— intersections $\cap : \Sigma \times \Sigma \to \Sigma$. If one restricts to shifts with rational coefficients, there are only countably many arrows in $\mathbf{C}_\mathcal{S}$.

Note that products in $\mathbf{C}_\mathcal{S}$ are *ordinary* set-theoretic products, not products of measurable spaces. The objects of the category $\mathbf{C}_\mathcal{S}$, *i.e.*, the Σ^ns, are equipped with a partial order in the following way:

$$(A_1, \ldots, A_n) \leq (B_1, \ldots, B_n) \text{ if } \forall i,\, 1 \leq i \leq n : A_i \subseteq B_i.$$

The key to the extension of $[\![.]\!]_\mathcal{S}$ is the following:

Lemma 4.22 *Morphisms of $\mathbf{C}_\mathcal{S}$ are all monotone and co-continuous; endomorphisms of $\mathbf{C}_\mathcal{S}$ all have greatest fixed points.*

Proof. First of all, we observe that shifts are indeed returning results in Σ by definition of a pre-LMP. Secondly, all generators are clearly monotone. Thirdly, if $\downarrow A_n$ is a decreasing sequence in Σ then:

$$\begin{aligned}\langle a\rangle_r(\cap A_n) &= \{s \mid h_a(s, \cap A_n) \geq r\} \\ &= \{s \mid \inf_n h_a(s, A_n) \geq r\} \\ &= \cap\langle a\rangle_r(A_n)\end{aligned}$$

where the second equation uses co-continuity of $h_a(s, .)$ on Σ, given by definition of pre-LMP kernels. So shifts are co-continuous, and so are evidently projections, intersections and all cartesian combinations of them.

Lastly, suppose ψ is an endomorphism, since it is monotone, it has a greatest fixed point in $(2^\mathcal{S})^n$, and since ψ is also co-continuous, this fixed point can be written as $\cap_n \psi^n(S, \ldots, S)$ and hence is in Σ^n. □

In fact $\mathbf{C}_\mathcal{S}$ has a structure of traced Cartesian category (or Cartesian category with fixed points [15]). We will write $\mathsf{Y}\psi$ for the fixed point of ψ. This is, of course, what we use for interpreting fixed points in the logic.

More generators could be added to the collection while keeping the key lemma above. For example, we could have added unions, countable unions and countable intersections to the generators (and therefore the countable power $\Sigma^\mathbb{N}$ as an object of $\mathbf{C}_\mathcal{S}$). This might indeed prove useful at some later stage, but for now we do not do this. We could *not* have added maps such as:

$$\psi(A) = \langle a \rangle_r(A) \setminus \langle a \rangle_{r'}(A) = \{s \mid h_a(s, A) \in [r, r')\}$$

which is not monotone; having only positive operators in the basic logic is crucial here. More subtly, strict shifts though they are monotone, cannot be added because they are not co-continuous and we need greatest fixed points (as made clear below).

So, strict shifts are not co-continuous and neither are shifts continuous. Here is an example: $([0,1], \mathcal{B}, h)$ with $h_a(s, B) = \lambda(B)$, where λ is the Lebesgue measure on $\mathcal{B}$, and while $\cup_n[0, 1{-}1/n] = [0,1]$, but for no n can we be Lebesgue sure to hit $[0, 1 - 1/n]$ there is always a $1/n$ chance that we do not, and so $\langle a \rangle_1([0, 1 - 1/n]) = \varnothing$. A similar case can be made that strict shifts are not co-continuous using intervals $\downarrow(0, 1/n]$.

Definition 4.23 Given a formula $\theta = (I, \lambda) \in \mathcal{L}_0^*$, we define in turn $\{\!\!\{\theta\}\!\!\}_{\mathcal{S}} \in \mathbf{C}_{\mathcal{S}}[\Sigma^I, \Sigma^I]$ and $[\![\theta]\!]_{\mathcal{S}} \in \Sigma^I$ by:

$$\{\!\!\{\theta\}\!\!\}_{\mathcal{S}}(\tau)(i) := \bigcap_{(a,i,j) \in \mathrm{dom}(\lambda)} \langle a \rangle_{\lambda(a,i,j)}(\tau(j))$$

with $\tau \in \Sigma^I$ a I-indexed tuple in Σ, and:

$$[\![\theta]\!]_{\mathcal{S}} := \mathsf{Y}\{\!\!\{\theta\}\!\!\}_{\mathcal{S}} = \cap_p \downarrow\!\{\!\!\{\theta\}\!\!\}^p_{\mathcal{S}}(S, \ldots, S)$$

where the p stands for the number of iterations.

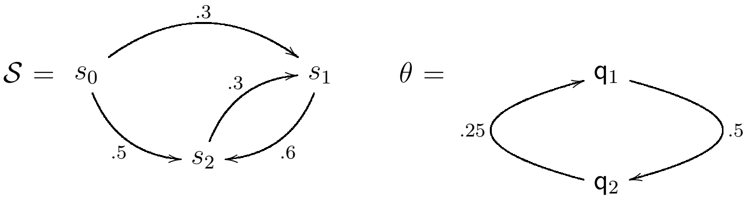
This somewhat pedantic notation comes handy when one wants to access states by their names, not their indices. We will use concrete tuple notation in examples, but not in proofs. Symbol τ sounds like "tuple" and is supposed to be suggestive of what τ is, a tuple. When $\mathrm{dom}(\lambda)$ is empty (which happens exactly when the corresponding state is dead in θ), we take the convention that the intersection is equal to the full set S.

Each component map $\lambda\tau.\{\!\!\{\theta\}\!\!\}_{\mathcal{S}}(\tau)(i)$ is in $\mathbf{C}_{\mathcal{S}}[\Sigma^I, \Sigma]$ indeed, since it is clearly expressed as a finite intersection of shifts; therefore the lemma above applies, and $[\![\theta]\!]_{\mathcal{S}} = \mathsf{Y}\{\!\!\{\theta\}\!\!\}_{\mathcal{S}}$ is well-defined and lies in Σ^I.

Lemma 4.24 *For all $\theta = (I, \lambda) \in \mathcal{L}_0^*$, $[\![\theta]\!]_{\mathcal{S}} \in \Sigma^I$.*

Least fixed points are not interesting here, that is in the absence of specific atomic properties on the state space, since one has to use strict shifts to have them in Σ, but $\langle a \rangle_{>r}(\varnothing) = 0$ for all pre-LMPs, so these would always be empty.

Example 4.25 Here is an LMP example with state space $S = \{s_0, s_1, s_2\}$ followed by a cyclic formula θ in $\mathcal{L}_0^*$:



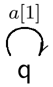
The operator converges to a fixed point in two steps:

$$\begin{aligned}
&\{\!|\theta|\!\}_{\mathcal{S}}(A_1, A_2) = (\langle\rangle_{.5}A_2, \langle\rangle_{.25}A_1),\\
&\mathsf{Y}\{\!|\theta|\!\}_{\mathcal{S}} = \{\!|\theta|\!\}^2_{\mathcal{S}}(S, S) = (\{s_0, s_1\}, \{s_0, s_2\}),\\
&[\![\theta]\!]_{\mathcal{S}}(q_1) = \{s_0, s_1\}.
\end{aligned}$$

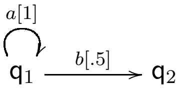
Intuitively $\mathsf{Y}\{\!|\theta|\!\}_{\mathcal{S}}$ is finding the biggest state-sets in $\mathcal{S}$ satisfying the specification described by θ.

The following example might be helpful for people used to the μ-calculus notation.

Example 4.26 Suppose that we have the $\nu X.\langle a\rangle_1 X$ formula and we want to express this in $\mathcal{L}_0^*$. We think of this automata theoretically. There is a state where X is satisfied and the system can do an a transition with probability 1 and return to this state. Thus in $\mathcal{L}_0^*$ we write: $\{q\}, \lambda(a, q, q) = 1$.

$\nu X.\langle a\rangle_1 X =$


Suppose we want to write the usual temporal logic formula with “until”: say, $a_1\mathcal{U}b_{.5}$ meaning that the system can keep doing a with probability 1 until it does a b with probability .5. In this case the automaton has two states and the $\mathcal{L}_0^*$ version would be $(\{q_1, q_2\}, \{\lambda(a, q_1, q_1) = 1, \lambda(b, q_1, q_2) = .5\})$.

$a_1\mathcal{U}b_{.5} =$


Now with Definition 4.23 and a formula $\theta \in \mathcal{L}_0$, we can build both $[\![\theta]\!]_{\mathcal{S}}$ and $[\![\theta^*]\!]_{\mathcal{S}}$, so obviously we have to say something ! (Reminder: formulas are used as their own indexing sets when coerced in $\mathcal{L}_0^*$.)

Lemma 4.27 *Definitions 4.23 and 2.3 of $[\![\theta]\!]_{\mathcal{S}}$ agree, in the sense that for all $\theta \in \mathcal{L}_0$: $[\![\theta^*]\!]_{\mathcal{S}}(\theta) = [\![\theta]\!]_{\mathcal{S}}$.*

Proof. The proof is an induction on $\mathcal{L}_0$, where we prove in addition that the fixed point $\mathsf{Y}\{\!|\theta|\!\}_{\mathcal{S}}$ is obtained in $|\theta|$ steps (and therefore "convergence time" for $\mathcal{L}_0$ formulas is independent of $\mathcal{S}$).
— $\theta = \top$: then $I = \{\top\}$, $\lambda = \varnothing$ and $\{\!|\top^*|\!\}_{\mathcal{S}}(\tau)(\top) = S$, $[\![\top^*]\!]_{\mathcal{S}}(\top) = S$ which is the correct answer obtained in $0 = |\top|$ steps;
— $\theta = \theta_0 \wedge \theta_1$: $I = I_0 \cdot I_1$ is the smashed sum of I_0 and I_1, obtained by fusing the initial states θ_0 and θ_1 into θ (since θ_0 and θ_1 are no longer maximal conjunctive) and taking the disjoint union otherwise; the only state where λ changes value is precisely θ itself, and $\lambda(a, \theta, i) = \lambda_0(a, \theta_0, i) + \lambda_1(a, \theta_1, i)$; so that, by definition:

$$\{\!|\theta^*|\!\}_{\mathcal{S}}(\tau_0 \cdot \tau_1)(\theta) = \{\!|\theta_0^*|\!\}_{\mathcal{S}}(\tau_0)(\theta_0) \cap \{\!|\theta_1^*|\!\}_{\mathcal{S}}(\tau_1)(\theta_1)$$
$$[\![\theta^*]\!]_{\mathcal{S}}(\theta) = [\![\theta_0^*]\!]_{\mathcal{S}}(\theta_0) \cap [\![\theta_1^*]\!]_{\mathcal{S}}(\theta_1)$$

and the answer is obtained in $\max(|\theta_0|, |\theta_1|)$.
— $\theta = \langle a\rangle_r \theta_0$: $I = I_0 + \{\theta\}$; λ takes now one more value, namely $\lambda(a, \theta, \theta_0) = r$ and:

$$\{\!|\theta^*|\!\}_{\mathcal{S}}(\tau)(\theta) = \langle a\rangle_r(\tau(\theta_0)),$$

again the correct answer, and obtained in $|\theta_0| + 1$ steps, as expected. □

4.12 Semantics of $\mathcal{L}_0^*$ via simulations

The fixed point definition of $\mathcal{L}_0^*$'s semantics, while being convenient for measure-theoretic considerations, is a bit clumsy when it comes to understanding what is going on. To rectify this we introduce a more perspicuous semantics $s \in [\![\theta]\!]_{\mathcal{S}}(i)$ which will turn out to be equivalent to the one just given.

Observe that when we say that a state satisfies a logical property, we expect this state to satisfy *at least* this property, and that it may satisfy other properties as well. Now that our properties are stated in a labelled transition setting, it is tempting to use the corresponding algebraic notion, that is, simulation. Indeed, if we look back to Example 4.25, we can observe that (the reflexive and transitive closure of) the relation $(q_1, s_0), (q_1, s_1), (q_2, s_0), (q_2, s_2)$ is a simulation relation.

The definition 2.4 of simulation must be extended to include systems that are not pre-LMPs. Recall that even if we view formulas of $\mathcal{L}_0^*$ as automata, they are not pre-LMPs because of the fact that a-transitions probabilities may sum up to some number > 1. This problem is easily disposed of by considering formulas as non-deterministic systems, and thus every transition as defining a distinct sub-probability distribution.

Definition 4.28 Given $\theta = (I, \lambda) \in \mathcal{L}_0^*$, $\mathcal{S} = (S, \Sigma, h)$ a pre-LMP, a relation $\mathfrak{S} \subseteq I \times S$ is a *non-deterministic simulation* if for all $a \in L$, $(i, s) \in \mathfrak{S}$ and $j \in I$: $\lambda(a, i, j) \leq h_a(s, \mathfrak{S}(j))$.

It is understood above that $\forall i$, $\mathfrak{S}(i) \in \Sigma$. However, this requirement is not in Definition 2.4 of simulation between LMPs, for if it was, bisimulation would not be a simulation. This issue is technically complex and not addressed here. The reader must keep in mind that this definition of simulation is safe only if we deal with countable state-space processes or if we manipulate simulated processes related by a co-simulation morphism, as will be argued in Lemma 4.30 below.[1]

Proposition 4.29 *Given $\theta = (I, \lambda) \in \mathcal{L}_0^*$, $\mathcal{S} = (S, \Sigma, h)$ a pre-LMP, $\mathfrak{S} \subseteq I \times S$ is a simulation if and only if for all i:*
— $\mathfrak{S}(i) \in \Sigma$,
— $\mathfrak{S}(i) \subseteq \{\!|\theta|\!\}_{\mathcal{S}}(\lambda j.\mathfrak{S}(j))(i)$.

Proof. The proof is by trivial manipulation of the various definitions. □

Now we say that s *simulates* i, when there exists a non-deterministic simulation $\mathfrak{S}$, with $(i, s) \in \mathfrak{S}$. Here is the rephrasing: $s \in [\![\theta]\!]_{\mathcal{S}}(i)$, or $s \models \theta(i)$ in shorthand notation, if and only if s simulates i. We also observe that $[\![\theta]\!]_{\mathcal{S}}$, regarded as a relation on $I \times S$, is the coarsest simulation.

4.13 Quotients with $\mathcal{L}_0^*$

We can now prove the analog of Corollary 4.8: co-simulation morphisms preserve formulas of $\mathcal{L}_0^*$.

Lemma 4.30 *Let $\mathcal{S}$, $\mathcal{S}'$ be pre-LMPs and $q : \mathcal{S} \to \mathcal{S}'$ be a co-simulation morphism, then for all $\theta = (I, \lambda) \in \mathcal{L}_0^*$, $i \in I$, $s \in S$: $q(s) \in [\![\theta]\!]_{\mathcal{S}'}(i) \Rightarrow s \in [\![\theta]\!]_{\mathcal{S}}(i)$.*

Proof. Composing a non-deterministic simulation on $I \times S'$ with the co-simulation q gives a simulation for $\mathcal{S}$. □

Finally, it remains now to prove the analog of Proposition 4.17, that is, that quotient states satisfy the same formulas of $\mathcal{F}$ as the states they $\mathcal{F}$-

[1] We conjecture that requiring that $\mathfrak{S}(i) \in \Sigma$ be an analytic set in $\mathcal{S}$ would solve the problem.

approximate. But before we have to explain what it means now for an equivalence $\mathfrak{R}$ over $\mathcal{S}$ to refine a formula $\theta \in \mathcal{L}_0^*$.

Definition 4.31 Let $\mathcal{S}$ be a pre-LMP, $\mathfrak{R}$ be an equivalence over $\mathcal{S}$, and $\theta = (I, \lambda) \in \mathcal{L}_0^*$, then $\mathfrak{R}$ refines θ if for all $i \in I$, $[\![\theta]\!]_{\mathcal{S}}(i)$ is $\mathfrak{R}$-closed.

By Lemma 4.27, this second definition coincides with the definition given before for $\mathcal{L}_0$ (to be exact, only maximal conjunctive subformulas have to be $\mathfrak{R}$-closed, so next proposition is marginally better).

With our definition in place we can home in on our proposition:

Proposition 4.32 *Let $\mathcal{S}$ be a pre-LMP and $\mathfrak{R}$ be an equivalence on $\mathcal{S}$ which refines θ, then for all $i \in I$, $s \in S$: $s \in [\![\theta]\!]_{\mathcal{S}}(i) \Rightarrow [s]_{\mathfrak{R}} \in [\![\theta]\!]_{\mathcal{S}_{\mathfrak{R}}}(i)$.*

Proof. Let $\mathfrak{R}$ be an equivalence relation refining θ, and assume that $s \models \theta(i)$. Then there is an associated simulation relation $\mathfrak{S}$ between θ and $\mathcal{S}$ such that for all $i\mathfrak{S}s$, if $\lambda(a, i, j) = r$ then $h_a(s, \mathfrak{S}(j)) \geq r$. Now let us prove that the corresponding relation between θ and $\mathcal{S}_{\mathfrak{R}}$ is a simulation relation. This relation $\mathfrak{R}^*$ is defined as $i\mathfrak{R}^*[s]$ if $i\mathfrak{R}'t$ for all $t \in [s]$. Since $\mathfrak{R}$ refines θ, and by definition of $\mathfrak{R}^*$ in terms of $\mathfrak{S}$, we have that $\mathfrak{S}(j) = q^{-1}\mathfrak{R}^*(j)$ where q is the quotient function. But now if $\lambda(a, i, j) = r$ then $h_a(t, \mathfrak{S}(j)) \geq r$ for all $t \in [s]$ and hence $h_{\mathfrak{R}}(a, [s], \mathfrak{R}^*(j)) = \inf_{t\in[s]} h_a(t, q^{-1}\mathfrak{R}^*(j)) = \inf_{t\in[s]} h_a(t, \mathfrak{S}(j)) \geq r$. □

We can pithily summarize the results of this section in a statement paralleling Theorem 4.18:

Theorem 4.33 (strong approximants) *Let $\mathcal{S}$ be a pre-LMP, $\mathcal{F}$ be a finite subset of $\mathcal{L}_0^*$, and $\mathfrak{R}$ the associated equivalence on $\mathcal{S}$, then for all $\theta = (I, \lambda) \in \mathcal{F}$, $s \in S$ and $i \in I$:*

$$s \in [\![\theta]\!]_{\mathcal{S}}(i) \Leftrightarrow [s]_{\mathfrak{R}} \in [\![\theta]\!]_{\mathcal{S}_{\mathfrak{R}}}(i).$$

Proof. As in the parallel statement, $\mathfrak{R}$ has finitely many equivalence classes which are all in Σ, because $\mathfrak{R}$ has finitely many generators, namely the $[\![\theta]\!]_{\mathcal{S}}(i)$, for $\theta \in \mathcal{F}$, $i \in I_\theta$. So again $\mathfrak{R}$ is an equivalence on $\mathcal{S}$, the quotient $\mathcal{S}_{\mathfrak{R}}$ is well-defined by Lemma 4.13, and the projection is a co-simulation morphism by Proposition 4.14, so Lemma 4.30 applies, and this gives the left to right implication. Besides and by definition, $\mathfrak{R}$ is the coarsest equivalence on $\mathcal{S}$ refining all θs in $\mathcal{F}$, so one may apply Proposition 4.32 and obtain the other implication. □

Note that, even if $\mathcal{S}$ is itself infinite state, the quotient will be finite, as soon as $\mathcal{F}$ is, just as in the $\mathcal{L}_0$ case.

The following result, which now follows easily, is one of the main motivations for using a logic with loops. We first need to prove that simulation relation between finite LMPs preserve formulas of $\mathcal{L}_0^*$.

Lemma 4.34 *If two states s and t of a finite LMP are related by a simulation relation, then every formula of $\mathcal{L}_0^*$ that s satisfies is also satisfied by t.*

Proof. The proof lies on simple manipulations of relations and inequalities and on the fact that every set in a finite LMP is measurable. □

Theorem 4.35 *For every finite-state LMP, there is a finite set of formulas $\mathcal{F}$ of $\mathcal{L}_0^*$ such that the quotient with respect to $\mathcal{F}$ is bisimilar to the process itself.*

Proof. The logic $\mathcal{L}_0^*$ clearly characterizes bisimulation of LMPs.[2] Indeed, it is an extension of $\mathcal{L}_0$ and since simulation preserves satisfaction of formulas of $\mathcal{L}_0^*$ (by the preceding lemma), so does bisimulation. This implies that if two states are not bisimilar, then there is a formula of $\mathcal{L}_0^*$ that will distinguish them. There are finitely many pairs, and taking all formulas that distinguish pairs of non-bisimilar states and closing this set under subformulas yields a finite set of formulas. This set defines a quotient which is bisimilar to the original finite-state process. Indeed, since non-bisimilar states belong to different equivalence classes, we have that every state of the quotient is made of bisimilar states of the original process. These states have the same transition probability to every bisimulation-closed set, and hence to every equivalence class. □

5 Approximation through average

In this section, we present a customizable approach to approximation and stay within the realm of LMPs. The approach is based on a radical departure from the ideas of the previous approaches. In the previous approaches one always approximated a system by ensuring that the transition probabilities in the approximant were below the corresponding transition in the full system. Here we approximate a system by taking a coarse-grained discretization (pixellization) of the state space and then using *average* values. This new notion is not based on the natural simulation ordering between LMPs as were the previous approaches.

Instead we use *conditional expectation*, which will construct for us low-resolution averages of any given LMP. Furthermore, an LMP will be known

[2] Note that for uncountable processes, this result needs an assumption that the state-space is analytic.

completely, up to bisimilarity, from its finite-resolution (meaning finite state) averages.

We first recall the definition of conditional expectation, then we identify circumstances in which the conditional expectation is actually defined pointwise and not only "almost everywhere". We construct an adaptation of the Lebesgue measure on any given LMP that will serve as the ambient probability which we need to drive the construction home. With all this in place we may turn to the definition of approximants. This conditional expectation will be made with respect to a σ-algebra generated by a set of formulas of $\mathcal{L}_0^*$. We will prove that the approximant satisfies exactly the same formulas of the given set as does the process being approximated. This will prove that they are correct, but we will also show the precise relation in which they stand with the order-theoretic approximants given in Section 4.

5.1 Conditional expectation

The expectation $\mathbb{E}_p(X)$ of a random variable X is the average computed by $\int X dp$ and therefore it is just a number. The *conditional* expectation is not a mere number but a random variable. It is meant to measure the expected value in the presence of additional information. The conditional expectation is typically thought of in the form: "if I know in advance that the outcome is in the set A then my revised estimate of the expectation is $\mathbb{E}_p(X|A)$." However additional information may take a more subtle form than merely stating that the result is in or not in a set.

The additional information takes the form of a sub-σ algebra, say Λ, of Σ. In what way does this represent "additional information"? The idea is that an experimenter is trying to compute probabilities of various outcomes of a random process. The process is described by (S, Σ, p). However she may have partial information in advance by knowing that the outcome is in a measurable set A. Now she may try to recompute her expectation values based on this information. To know that the outcome is in A also means that it is *not* in the complement A^c. Note that $\{\varnothing, A, A^c, S\}$ is in fact a (tiny) sub-σ-algebra of Σ. Thus one can generalize this idea and say that for some given sub-σ-algebra Λ of Σ she knows for every $A \in \Lambda$ whether the outcome is in A or not. Now she can recompute the expectation values given this information.

How can she actually express this revised expectation when the σ-algebra Λ is large? It is presented as a density function so that for every Λ-measurable set B one can compute the conditional expectation by integration over B. Thus instead of a number we get a Λ-measurable function called the *conditional* *expectation given* Λ and written $\mathbb{E}_p(\_|\Lambda)$.[3]

It is not at all obvious that such a function should exist and is indeed a fundamental result of Kolmogorov [22, p.84].

Theorem 5.1 (Kolmogorov) *Let* (S, Σ, p) *be a probability triple,* X *be in* $\mathcal{L}^1(S, \Sigma, p)$ *and* Λ *be a sub-*σ*-algebra of* Σ*, then there exists a* $Y \in \mathcal{L}^1(S, \Lambda, p)$ *such that*

$$\forall B \in \Lambda. \int_B X dp = \int_B Y dp. \tag{1}$$

Not only does the conditional expectation exist, but it has a lot of properties. As a functional of type

$$\mathbb{E}_p(\_|\Lambda) : \mathcal{L}^1(S, \Sigma, p) \to \mathcal{L}^1(S, \Lambda, p)$$

it is *linear*, *positive*, *monotone* with respect to the pointwise ordering and *continuous* in the sense that for any sequence (X_n) with $0 \leq X_n \uparrow X$ and X_n, $X \in \mathcal{L}^1(S, \Sigma, p)$, then $\mathbb{E}_p(X_n|\Lambda) \uparrow \mathbb{E}_p(X|\Lambda) \ldots$ but it is *not* uniquely defined !

All candidate conditional expectations are called *versions* of the conditional expectation. It is easy to prove that any two Λ-measurable functions satisfying the characteristic property (1) given above may differ only on a negligible set (a set of p-probability zero).

5.2 The finite case

As we have said before, the basic intuition of $\mathbb{E}_p(X|\Lambda)$ is that it averages out all variations in X that are below the resolution of Λ, *i.e.*, which do not depend on Λ. In particular, if X is independent of Λ, then $\mathbb{E}_p(X|\Lambda) = \mathbb{E}_p(X)$,[4] and X is completely averaged out.[5] On the other hand, if X is fully dependent on Λ, in other words if X is Λ-measurable, then $\mathbb{E}_p(X|\Lambda) = X$.

This intuition is exact in the case that the sample space S is *finite*. We may suppose then that $\Sigma = 2^S$, and Λ will be generated by a set of equivalence classes. But then $Y = \mathbb{E}_p(X|\Lambda)$ has to be constant on equivalence classes

[3] Take note that, in the same way as $\mathbb{E}_p(X)$ is constant on S, the conditional expectation will be constant on every "pixel" or smallest observable set in Λ. In the above "tiny" sub-σ-algebra, this means constant on both A and A^c. This will turn out to be exactly what we need later when pixels are defined by sets of formulas.

[4] Recall that in this equation the left-hand side is a function while the right-hand side is a number; we mean to say that the function on the left is a constant function whose value is given by the right-hand side.

[5] Given a probability triple (S, Σ, p), a random variable $X \in m\Sigma$ is said to be independent of a sub-σ-algebra Λ if for any event $A \in \sigma(X)$ and $B \in \Lambda$, $p(A \cap B) = p(A)p(B)$. In particular, as one can easily verify, X is always independent of the trivial σ-algebra $\Lambda_0 = \{\varnothing, S\}$ and by the remark above, $\mathbb{E}_p(X|\Lambda_0) = \mathbb{E}_p(X)$ the ordinary unconditional expectation of X.

(else it is not Λ-measurable) and by the characteristic property, with B an equivalence class $[s]$, we get:

$$Y(s).p([s]) = \int_{[s]} Y dp = \int_{[s]} X dp = \sum_{t \in [s]} X(t)p(\{t\})) = \mathbb{E}(1_{[s]}X),$$

where $1_{[s]}$ is the indicator function of the measurable set $[s]$.

When $p([s]) > 0$ we see that Y is exactly the *p-average* of X over equivalence classes associated to Λ:

$$Y(s) = \frac{1}{p([s])} \cdot \mathbb{E}(1_{[s]}X).$$

5.3 The example that says it all

Now that it is understood that in the finite state-space case conditional expectations are averages over equivalence classes, we can consider a revealing example. Put $S = \{x, y, 0, 1\}$, $\Sigma = 2^S$, $L = \{a\}$ (there is only one label, so we will not even bother to write a in the kernels); $h(\{0\})(x) = h(\{1\})(y) = 1$ and every other state-to-state transition is of probability zero. Suppose Λ identifies x and y, and call the resulting class z.

One can conceive of three ways to define a kernel k on the quotient space $\{z, 0, 1\}$. The first two are already familiar from the two previous sections. One can define k as the *infimum* over $\{x, y\}$ or dually one can take it to be the *supremum*:

$$k_i(\{0\})(z) = 0,\ k_i(\{1\})(z) = 0,\ k_i(\{0,1\})(z) = 1,$$
$$k_s(\{0\})(z) = 1,\ k_s(\{1\})(z) = 1,\ k_s(\{0,1\})(z) = 1.$$

or one can also define k as an *average* (using here the uniform probability on the underlying state space):

$$k_a(\{0\})(z) = 1/2,\ k_a(\{1\})(z) = 1/2,\ k_a(\{0,1\})(z) = 1.$$

As we said earlier, the use of the infimum results in super-additive kernels while the use of a supremum results in sub-additive kernels:

$$k_i(\{0,1\})(z) = 1 > k_i(\{0\})(z) + k_i(\{1\})(z) = 0$$
$$k_s(\{0,1\})(z) = 1 < k_s(\{0\})(z) + k_s(\{1\})(z) = 2.$$

Of the three options, only the third preserves additivity:

$$k_a(\{0,1\})(z) = 1 = k_a(\{0\})(z) + k_a(\{1\})(z).$$

Besides we observe that, perhaps not surprisingly, in all cases the kernel obtained by using averages is sandwiched between the others, *e.g.*:

$$0 = k_i(\{0\})(z) \leq k_a(\{0\})(z) = 1/2 \leq k_s(\{0\})(z) = 1.$$

The rest of this section is essentially about structuring this nice concrete notion of approximant by averages as a general construction and explaining in what sense these approximants are actually approximating what they are supposed to be approximants of.

5.4 When $\mathbb{E}_p(\_|\Lambda)$ *is unique*

There is one thing we have to confront. As we noted before, conditional expectations are unique only "almost surely." Now we want to use them to average our family of $h(a, A)$ and, from the definition of an LMP, we need these averages to be defined *pointwise*, not only up to p. Yet, in the case of finite systems, one option is to choose for p the uniform probability on S, in which case "almost surely" actually means "surely," since only the empty set is in $\mathcal{N}_p$. This, intuitively, is because points are big enough chunks to be seen by the probability distribution. This leads to the following two definitions.

Definition 5.2 [pixels] Let (S, Σ) be a measurable space, one says s and $t \in S$ are Σ-indistinguishable if $\forall A \in \Sigma$, $s \in A \leftrightarrow t \in A$.

This is an equivalence on S and we write $[s]_\Sigma$, or sometimes simply $[s]$ to denote the equivalence class of s. One has $[s]_\Sigma = \cap\{A \mid s \in A \in \Sigma\}$. So equivalence classes might not be measurable themselves, unless Σ is countably generated, which is the case we are interested in.

Definition 5.3 [granularity] Let (S, Σ, p) be a probability triple and $\Lambda \subseteq \Sigma$ be a sub-σ-algebra of Σ; p is said to be *granular* over Λ if for all $s \in S$, $[s]_\Lambda \notin \mathcal{N}_p$.

In other words, p is granular over Λ if no Λ equivalence class is negligible. What this means intuitively is that the "pixellization" of Λ is always seen by p. It may be instructive to point out that there are at most countably many equivalence classes in this case.

As an example, we can take the probability triple $([0, 1)^2, \mathcal{B}_2, \lambda_2)$, where λ_2 is the Lebesgue measure on the square, and $\Lambda = \mathcal{B} \times [0, 1)$. Then $[s]_\Lambda = \{s\} \times$

$[0,1) \in \Lambda$ and $\lambda_2([s]) = 0$ so our p is not granular over this Λ. The measurable sets of Λ are very thin strips. They are too fine to be granular. But if we take a cruder Λ, namely that generated by the squares $[k/n, k+1/n) \times [h/n, h+1/n)$ for k, $h < n$ (with n fixed), then $[s]_\Lambda$ is such a square of λ_2-measure $1/n^2$, so here p is granular.

The big payoff of granularity is the following:

Lemma 5.4 (Uniqueness lemma) *Let (S, Σ, p) be a probability triple, $\Lambda \subseteq \Sigma$, p granular over Λ, X and Y both Λ-measurable, then:*

$$X = Y \ a.s. \Rightarrow X = Y.$$

So in this case "almost surely" does mean "surely !"

Proof. Set $A := \{s \in S \mid X(s) = \alpha \wedge Y(s) = \beta\}$ and $t \in A$. One has $A \in \Lambda$, by Λ-measurability of X and Y, but then $[t]_\Lambda \subseteq A$ (otherwise A splits $[t]_\Lambda$). So by granularity $p(A) > 0$ (else $[t]_\Lambda$ is negligible), and therefore $\alpha = \beta$ or else X and Y differ on a non negligible set A. □

So in this favourable circumstances we can do away with versions. If $X \in \mathcal{L}^1(S, \Sigma, p)$, and p is granular over Λ:

$$\mathbb{E}_p(X|\Lambda) : \mathcal{L}^1(S, \Sigma, p) \to \mathcal{L}^1(S, \Lambda, p)$$

is uniquely defined and we can proceed to the main definition.

5.5 Projecting LMPs

Definition 5.5 [projection of an LMP] Given (S, Σ) a measurable space, Λ a sub-σ-algebra of Σ, p a probability on (S, Σ) granular over Λ, and $\mathcal{S} = (h(a, A))_{a \in L, A \in \Sigma}$ an LMP on (S, Σ), one defines the *p-projection* of $\mathcal{S}$ on Λ, written $(\mathcal{S}|\Lambda)_p$ as:

$$h'(a, A) = \mathbb{E}_p(h(a, A)|\Lambda), \text{ for } a \in L, \ A \in \Lambda.$$

Take note that this is *the* version of the conditional expectation. Existence follows from the fact that the $h(a, A)$ evidently are integrable with respect to p (they are measurable, positive and bounded by 1), in other words they are in $\mathcal{L}^1(S, \Sigma, p)$.

Proposition 5.6 (Staying within LMPs) $(\mathcal{S}|\Lambda)_p$ *is an LMP.*

Proof. All maps $h'(a, A)$ are Λ-measurable by definition of the conditional expectation; $h'(a, A)$ has values $[0, 1]$, because conditional expectation is monotone, and from $0 \leq h(a, A) \leq 1$, one gets $0 = \mathbb{E}_p(0|\Lambda) \leq h'(a, A) \leq \mathbb{E}_p(1|\Lambda) = 1$; additivity is because $\mathbb{E}_p(\_|\Lambda)$ is linear; continuity follows from the fact that $\mathbb{E}_p(\_|\Lambda)$ is itself continuous (a property known as the conditional form of the monotone convergence theorem). □

We may now round off the construction by changing the state space.

Let us write $[\_]_\Lambda : S \to [S]_\Lambda$ for the canonical surjection to the set of equivalence classes and denote accordingly the quotient σ-algebra by $[\Lambda]_\Lambda$. Then one can define the *quotient* LMP $([S]_\Lambda, [\Lambda]_\Lambda, k)$ with:

$$k(a, B)([s]_\Lambda) := h'(a, \cup B)(t) := \mathbb{E}_p(h(a, \cup B)|\Lambda)(t),$$

with $t \in [s]$. Take note that the right hand side is independent of the choice of $t \in [s]_\Lambda$ since $h'(a, A)$ is Λ-measurable, and therefore $h'(a, A)$ has to be constant on $[s]_\Lambda$ (else the equivalence is split by an event in Λ). Moreover, $[\_]_\Lambda$ is a bisimulation morphism (which was formerly called a "zig-zag" [7]) from $(\mathcal{S}|\Lambda)_p$ to $([S]_\Lambda, [\Lambda]_\Lambda, k)$ and as such it preserves all $\mathcal{L}_0$ properties.

So far we have a quotient theory for LMPs when pixels are big enough, but everything hinges on the choice of an ambient p. This is the second problem we have to deal with.

5.6 A "uniform" probability on $(S, \sigma(\mathcal{L}_0))$

The key is to construct an appropriate measure, and we will use $\mathcal{L}_0$ to do this. So, given an LMP $\mathcal{S} = (S, \Sigma, h)$, and a *fixed* enumeration (θ_n) of $\mathcal{L}_0$, we first define a sequence (S, Λ_n) of measurable spaces:[6]

$$\Lambda_0 := \{\varnothing, S\},\ \Lambda_n := \sigma([\![\theta_i]\!]_\mathcal{S}; i < n).$$

Then for each n, we set $\tau_n := \mathbf{1}_{[\![\theta_n]\!]_\mathcal{S}}$ and define $\alpha_n : \{0, 1\}^n \to \Lambda_n$ as:

$$\alpha_n(\boldsymbol{x}) = \cap_{i<n}\{s \mid \tau_i(s) = \boldsymbol{x}_i\},$$

with the convention that $\{0, 1\}^0 = \{*\}$ and $\alpha_0(*) = S$.

Each Λ_n is a finite boolean algebra and so has atoms (non empty sets in Λ_n with no proper subsets); each atom of Λ_n is the image by α_n of a unique sequence $\boldsymbol{x} \in \{0, 1\}^n$, but not all sequences are mapped to atoms, some are mapped to the empty set.

Now the idea is to construct p stagewise and at each stage to divide evenly the mass of an atom $\alpha_n(\boldsymbol{x}) \in \Lambda_n$ between its proper subsets in Λ_{n+1} if there

[6] For each n, $\Lambda_n \subseteq \Lambda_{n+1}$, this is usually called a filtration.

are some. Specifically, we define inductively p_n on Λ_n-atoms as:

$$\begin{array}{l} p_0(\varnothing) = 0,\ p_0(S) = 1 \\ \alpha_{n+1}(\boldsymbol{x}0) \neq \varnothing,\ \alpha_{n+1}(\boldsymbol{x}1) \neq \varnothing \Rightarrow p_{n+1}(\alpha_{n+1}(\boldsymbol{x}0)) = p_{n+1}(\alpha_{n+1}(\boldsymbol{x}1)) = \frac{1}{2} \cdot p_n(\alpha_n(\boldsymbol{x})) \\ \alpha_{n+1}(\boldsymbol{x}0) = \varnothing,\ \alpha_{n+1}(\boldsymbol{x}1) \neq \varnothing \Rightarrow p_{n+1}(\alpha_{n+1}(\boldsymbol{x}0)) = 0,\ p_{n+1}(\alpha_{n+1}(\boldsymbol{x}1)) = p_n(\alpha_n(\boldsymbol{x})) \\ \alpha_{n+1}(\boldsymbol{x}0) \neq \varnothing,\ \alpha_{n+1}(\boldsymbol{x}1) = \varnothing \Rightarrow p_{n+1}(\alpha_{n+1}(\boldsymbol{x}0)) = p_n(\alpha_n(\boldsymbol{x})),\ p_{n+1}(\alpha_{n+1}(\boldsymbol{x}1)) = 0 \end{array}$$

Clearly each p_n extends to a unique probability on (S, Λ_n) since it is defined on Λ_n-atoms and the p_n are compatible in the sense that $p_{n+1} \restriction \Lambda_n = p_n$; the sequence p_n converges to an additive set map on the union $\cup_n \Lambda_n$.

In most cases, including the case of finite state spaces, this p will be extendable to a sort of "skewed" Lebesgue measure, also written p, and defined on $\sigma(\mathcal{L}_0)$, the σ-algebra generated by our temporal formulas. [7] But, and contrary to what was said in a former version of this construction [6], in the absence of further structure on the state space, one *cannot guarantee* this. On the other hand, if such an extension exists, it is unique. From now on, we will take the conservative assumption that this extension exists.

We take note, for future use, that for any finite set of formulas $\mathcal{F} \subset \mathcal{L}_0$, writing $\Lambda_{\mathcal{F}}$ the associated σ-algebra, one has:

$$p([s]_{\Lambda_{\mathcal{F}}}) \geq 2^{-N} \tag{2}$$

where $N = \max\{i \mid \theta_i \in \mathcal{F}\}$ and $s \in S$.

Second, we observe that the p obtained here will depend on the original enumeration, and we leave for future investigation the question of whether there is a principled way of choosing p. In our case, all choices will work equally well.

As an example we can consider the transition system with only state s, only one letter a and $h(a, \{s\})(s) = 1/2$. Then $s \models \theta$ if and only if all coefficients used in θ are below $1/2$. In this case, and as with all one-state systems, at any stage there will be at most one atom namely $\{s\}$ and therefore $p(\{s\}) = 1$.

5.7 Compressing Σ

But the reader might protest that to apply the projection, one needs a probability on an arbitrary Σ not just on $\sigma(\mathcal{L}_0)$. Well, in fact, it is enough to consider the latter case because of Theorem 4.5 (saying that $\sigma(\mathcal{L}_0)$ is the smallest σ-algebra closed under shifts).

Therefore, $\sigma(\mathcal{L}_0)$ is always included in Σ, since Σ has to be closed by shifts (this is equivalent to asking that $h(a, A)$ are all Σ-measurable) and one can

[7] To be exact, by $\sigma(\mathcal{L}_0)$ we mean $\sigma([\![\theta]\!]_{\mathcal{S}}; \theta \in \mathcal{L}_0)$.

always “compress” an LMP to $\sigma(\mathcal{L}_0)$. The obtained LMP is obviously bisimilar to the first since by construction states are the same and their temporal properties remain the same as well. Without loss of generality, we may and will suppose thereafter that $\Sigma = \sigma(\mathcal{L}_0)$.

5.8 Approximations

Now we can complete the approximation construction.

Let $\mathcal{S}$ be a compressed LMP $\mathcal{S} = (S, \Sigma, h)$ with $\Sigma = \sigma(\mathcal{L}_0)$, and $\mathcal{F} \subseteq \mathcal{L}_0^*$ be a *finite* set of formulas, set Λ to be the σ-algebra, $\sigma(\mathcal{F})$, generated by $\mathcal{F}$ on S.

We observe that by inequation (2), p is granular over Λ, so the machinery gets us a *finite-state* LMP approximant:

$$\mathcal{S} = (S, \Sigma, h) \xrightarrow{[\cdot]_\Lambda} \mathcal{S}_\mathcal{F} = ([S]_\Lambda, [\Lambda]_\Lambda, k)$$

which is the quotient constructed above after the appropriate projection.

There are at most $2^{|\mathcal{F}|}$ states in $\mathcal{S}_\mathcal{F}$, in particular it is a finite-state probabilistic transition system.

5.9 Convergence

We need to say how the obtained $\mathcal{S}_\mathcal{F}$ approximates $\mathcal{S}$. In the previous approaches, approximants were always below the approximated process and hence simulated by it. It is not the case here since approximants are neither above nor below $\mathcal{S}$. However, $\mathcal{S}_\mathcal{F}$ does converge to $\mathcal{S}$. This is what the following proposition says; it improves on the analog proposition in the original paper which only concerned the smaller logic $\mathcal{L}_0$.

Proposition 5.7 *For every finite subformula-closed set of formulas $\mathcal{F} \subset \mathcal{L}_0^*$: $\mathcal{S}_\mathcal{F} \approx_\mathcal{F} \mathcal{S}$.*

Proof.

Let $\mathfrak{R}$ be the coarsest simulation between $\theta = (I, \lambda) \in \mathcal{L}_0^*$ and $\mathcal{S}$. Define $\mathfrak{R}^*$ to be the composition of $\mathfrak{R}$ with the quotient morphism from $\mathcal{S}$ to $\mathcal{S}_\mathcal{F}$. This is well defined since $\mathfrak{R}$ is the coarsest simulation and because equivalent states satisfy the same formulas of $\mathcal{F}$. We prove that $\mathfrak{R}^*$ is a simulation. Let $(i, [s]) \in \mathfrak{R}^*$ and $j \in I$. Then $\lambda(a, i, j) \leq h(a, \mathfrak{R}(j))(t)$ for all $t \in [s]$ because it is true for at least one $t \in [s]$ by definition of $\mathfrak{R}^*$ and because all states in $[s]$ satisfy the same formulas of $\mathcal{F}$. This implies that $\lambda(a, i, j) \leq \mathbb{E}_p(h(a, \mathfrak{R}(j))|\Lambda)(t)$ for all $t \in [s]$, which shows that $\mathfrak{R}^*$ is a simulation since $\cup\mathfrak{R}^*(j) = \mathfrak{R}(j)$.

Now let $\mathfrak{R}$ be a simulation between $\theta = (I, \lambda) \in \mathcal{L}^*$ and $\mathcal{S}_\mathcal{F}$. Define $\mathfrak{R}^*$ to be the composition of $\mathfrak{R}$ with the inverse of the quotient morphism from $\mathcal{S}$ to $\mathcal{S}_\mathcal{F}$. We prove that $\mathfrak{R}^*$ is a simulation. Let $(i, s) \in \mathfrak{R}^*$ and $j \in I$, that is, $(i, [s]) \in \mathfrak{R}$. Thus $\lambda(a, i, j) \leq \mathbb{E}_p(h(a, \cup\mathfrak{R}(j))|\Lambda)(t)$ for all $t \in [s]$. Then at least one $t \in [s]$ satisfies $\lambda(a, i, j) \leq h(a, \cup\mathfrak{R}(j))(t) = h(a, \mathfrak{R}^*(j))(t)$. This equation is true for all $t \in [s]$ because they all satisfy the same formulas of $\mathcal{F}$, and hence it is true for s, as wanted. □

From Proposition 2.6, it follows now easily that:

Theorem 5.8 *If $(\mathcal{F}_i)$ is an increasing sequence of subformula-closed sets of formulas converging to the set of all formulas $\mathcal{L}_0^*$, then for all $c \in (0, 1)$:*

$$d^c(\mathcal{S}_{\mathcal{F}_i}, \mathcal{S}) \longrightarrow_{i\to\infty} 0.$$

We could have taken another route to prove Proposition 5.7. As the example 5.3 suggested, quotients constructed with conditional expectations do lie between the inf- and the sup- approximants [5]:

$$\begin{aligned}
k(a, [A])([s]_\Lambda) &:= h'(a, A)(s) \\
&= \tfrac{1}{p([s]_\Lambda)} \int_{[s]_\Lambda} h'(a, A)\, dp \qquad && h'(a, A) \text{ constant on } [s]_\Lambda \\
&= \tfrac{1}{p([s]_\Lambda)} \int_{[s]_\Lambda} h(a, A)\, dp && [s]_\Lambda \in \Lambda \\
&\geq \inf_{t \in [s]_\Lambda} h(a, A)
\end{aligned}$$

The second equation holds both because $h'(a, A)$ is constant on equivalence classes *and* because p is granular and therefore $p([s]_\Lambda) > 0$. The third equation is the characteristic property of conditional expectations. A similar type of argument allows one to reason analogously for the supremum case.

Thus another, indirect, way to prove the previous proposition, is to use this sandwiching effect and the fact that the infimum and supremum were proven to give approximations in the same sense as proposition 5.7 [5]. This also makes clear in which sense the average-based approximants are better than the order-theoretic ones.

6 Conclusion

We have presented a constructive approximation which is an improvement of the original one [9], and also two new abstract notions of approximation. The first is based on customizing the approximation with respect to certain formulas of interest, the second is based on averaging techniques, or - more precisely - on the use of conditional expectations.

For the first abstract construction, we have added two simple ideas to the theory of LMPs: first, LMP approximants should be quotients with respect to the LMP bisimulation logic $\mathcal{L}_0$, yielding *stronger* approximants; second, the same quotient construction, supposing there is one, should be possible with a logic enriched with greatest fixed points and produce families of approximants sharing cyclic behaviours with the approximation target, resulting in a *faster* approximation construction, since finite processes are approximated by themselves at some finite stage.

Not only do these two ideas carry through, but despite their apparent independence they work together fruitfully. Some of the known constructions and definitions have to be relaxed in so doing but the resulting theory is in many ways more pleasing than the original.

We believe that the present work is an important step towards model-checking LMPs. For example, if one knows what are the properties that a given continuous process should satisfy, one would prefer to check for these properties on a finite *faithful* approximant of the process instead of checking each property on the process itself. Our construction achieves this goal since it theoretically ensures exact satisfaction of formulas.

Observe that in Example 4.1, if one was interested specifically in the initial state s_0 one could live with the approximant: $[s_0] \xrightarrow{b} [s,t]$ because $[s_0]$ is equivalent to s_0, if we consider only formulas of depth 1 —of course there is a loss for other states like s and t which are *not* equivalent to $[s,t]$. This suggests that there may still be a way of quotienting with formulas and obtain an LMP. We want to investigate this possibility, which we think will be a fairly easy task. More interestingly, observe that the quotient we have defined does not depend only on the satisfied formulas, we crucially use probability information from the system itself. This implies that two processes that are $\mathcal{F}$-equivalent may not have the same quotient. We plan to investigate the possibility of using the quotient construction without using the actual values of the transition probabilities in the original process, but only values provided by formulas that are satisfied. Instead, we would use only formulas in $\mathcal{F}$ and take infima over the formulas satisfied by equivalent states. Every state in the resulting pre-LMP would be the representant for every $\mathcal{F}$-equivalent state. An important application of this would be a way to construct a process by using only the formulas that it has to satisfy, that is, the automated design of probabilistic models from specifications. We believe that in this case, there will be no LMP that will satisfy the same property (even for finite quotients), showing that pre-LMPs are essential for the design of probabilistic systems.

On the practical side, the effective construction of these pre-LMPs could be costly in time or inconvenient. One has to choose a set of formulas that will be used to quotient the state space. A pre-LMP is then produced by computation of an infimum from every equivalence class to possibly *every* logically definable union of states. This last step increases complexity significantly.

However, we also presented an even faster version of approximants which parallels a former construction [9] and introduces additional loops. The inconvenient of this concrete approximation scheme is that one does not have the choice of formulas, except for their depth and a desired precision. The same properties are satisfied: every formula satisfied by a state is eventually satisfied by the state approximant, and finite processes are eventually approximated by themselves.

Ongoing research is also trying to apply this theory of approximants to other probabilistic models such as continuous time Markov chains and to extend it to a richer logic. One potential application are Markov Decision Processes that one finds in the field of machine learning. Approximants have been studied in this field, but always with a focus on partitioning the state-space without taking account of the behaviour of processes, that is, of the actual transitions that states can take. As a result, bisimilar or behaviourally close states can be split in the process, whereas our constructions always partition the state-space with respect to satisfaction of formulas.

The last approach to approximation is more probabilistically-minded. It is based on conditional expectations. Given a probability p on (S, Σ), and a sub-σ-algebra Σ' of Σ, it is possible to define the conditional expectation given Σ' of any integrable function according to p. Applied to finite-state systems, the idea downs to taking the quotient kernel to be an average rather than an infimum.

This technique for LMPs shares a number of good properties with our first abstract approach. It can be customized in the same sense; however, one can use it and also stay within the framework of traditional LMPs and avoid having to work with pre-measures.

We feel that, beyond the properties of the construction, there are some new directions implicit in this probabilistic approximation work. First, the idea of granularity is, we feel, significant. One of the big obstacles to the applicability of modern probability theory on general spaces to the computational setting has been the curse of non uniqueness embodied in the phrases "almost everywhere" and "almost surely" seen almost everywhere in probability theory. One can even argue that the bulk of the computer science community has worked with discrete systems to try and avoid this non uniqueness. Our use of granularity shows a new sense in which the discrete can be used to dispel the non uniqueness that arises in measure theory.

The second important direction that we feel should be emphasized is the use of averages rather than infima. This should lead to better numerical properties. More striking than that however is the fact that the simulation order is not respected by the approximants. Perhaps it suggests that some sort of non monotone approximation occurs. Similar phenomena have been observed by Martin [17] - which was the first departure from Scott's ideas of monotonicity as being one of the key requirements of computability - and also in the context of non determinate dataflow [19].

Let us conclude with a further comment on why we do not mention any properties of analytic space, in contrast to what is done in previous papers on LMPs. In fact, analyticity is needed if one wants to use the fact that the relational definition of bisimulation is characterized by the logic. If one is happy with only the logic or the metric in order to compare or work with LMPs, there is no need for analyticity of the state space in the definition. However, if one indeed needs the analytic property of processes, the results of the present paper carry through since the quotient of an analytic space under countably many conditions is analytic. This follows essentially from well known facts about analytic spaces, see for example chapter 3 of "Invitation to C^*-algebras" by Arveson [1].

References

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

[2] G. Choquet. Theory of capacities. *Ann. Inst. Fourier (Grenoble)*, 5:131–295, 1953.

[3] Vincent Danos and Josée Desharnais. Note sur les chaînes de Markov étiquetées. Unpublished (in French), 2002.

[4] Vincent Danos and Josée Desharnais. A fixpoint logic for Labelled Markov Processes. In *Proceedings of the International Workshop on Fixed Points in Computer Science (FICS'03)*, pages 413–422, Warsaw, 2003.

[5] Vincent Danos and Josée Desharnais. Labeled Markov Processes: Stronger and faster approximations. In *Proceedings of the 18^{th} Symposium on Logic in Computer Science*, pages 341–350, Ottawa, 2003. IEEE.

[6] Vincent Danos, Josée Desharnais, and Prakash Panangaden. Conditional expectation and the approximation of labelled Markov processes. In *CONCUR 2003 - Concurrency Theory*, volume 2761 of *Lecture Notes in Computer Science*, pages 477 – 491. Springer-Verlag Heidelberg, December 2003.

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

[8] Josée Desharnais. *Labelled Markov Processes.* PhD thesis, McGill University, November 1999.

[9] Josée Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Approximating labeled Markov processes. *Information and Computation*, 184(1):160–200, July 2003.

[10] Josée Desharnais, Vineet Gupta, R. Jagadeesan, and Prakash Panangaden. Metrics for labeled Markov processes. In *Proceedings of CONCUR99*, Lecture Notes in Computer Science, pages 258–273. Springer-Verlag, 1999.

[11] Josée Desharnais, Vineet Gupta, R. Jagadeesan, and Prakash Panangaden. Approximating continuous Markov processes. In *Proceedings of the 15th Annual IEEE Symposium On Logic In Computer Science, Santa Barbara, Californie, USA*, pages 95–106, 2000.

[12] E.-E. Doberkat. Semi-pullbacks and bisimulations in categories of stochastic relations. In *Proceedings of the 30th International Colloquium on Automata, Languages, and Programming (ICALP)*, number 2719 in Lecture Notes In Computer Science, pages 996–1007. Springer-Verlag, 2003.

[13] Abbas Edalat. Domain of computation of a random field in statistical physics. In C. Hankin, I. Mackie, and R. Nagarajan, editors, *Theory and Formal Methods 1994: Proceedings of the second Imperial College Department of Computing Workshop on Theory and Formal Methods*, pages 11–14. IC Press, 1994.

[14] Abbas Edalat. Domain theory and integration. *Theoretical Computer Science*, 151:163–193, 1995.

[15] Masahito Hasegawa. Recursion from cyclic sharing: traced monoidal categories and models of cyclic lambda calculi. In *3rd International Conference on Typed Lambda Calculi and Applications (TLCA'97)*, volume 1210 of *LNCS*, pages 196–213. Springer, 1997.

[16] 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.

[17] Keye Martin. The measurement process in domain theory. In *International Colloquium on Automata, Languages and Programming*, pages 116–126, 2000.

[18] R. Milner, J. Parrow, and D. Walker. A calculus of mobile processes i and ii. *Information and Computation*, 100:1–77, 1992.

[19] Prakash Panangaden and V. Shanbhogue. The expressive power of indeterminate dataflow primitives. *Information and Computation*, 98(1):99–131, 1992.

[20] Franck van Breugel, Michael Mislove, Joël Ouaknine, and James Worrell. An intrinsic characterization of approximate probabilistic bisimilarity. In *Proceedings of the 6th International Conference on Foundations of Software Science and Computation Structures (FOSSACS)*, number 2620 in Lecture Notes In Computer Science, pages 200–215, 2003.

[21] Franck van Breugel, Steven Shalit, and James Worrell. Testing labelled markov processes. In *Proceedings of the 29th International Colloquium on Automata, Languages, and Programming (ICALP)*, number 2380 in Lecture Notes In Computer Science, pages 537–548. Springer-Verlag, 2002.

[22] David Williams. *Probability with Martingales.* CUP, Cambridge, 1991.