Conditional Expectation and the Approximation of Labelled Markov Processes

Vincent Danos[1,⋆], Josée Desharnais[2], and Prakash Panangaden[3]

[1] Université Paris 7 & CNRS
[2] Université Laval, Québec
[3] McGill University, Montréal

Abstract. We develop a new notion of approximation of labelled Markov processes based on the use of conditional expectations. The key idea is to approximate a system by a coarse-graining of the state space and using averages of the transition probabilities. This is unlike any of the previous notions where the approximants are simulated by the process that they approximate. The approximations of the present paper are customizable, more accurate and stay within the world of LMPs. The use of averages and expectations may well also make the approximations more robust. We introduce a novel condition – called "granularity" – which leads to unique conditional expectations and which turns out to be a key concept despite its simplicity.

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 the process-algebra based notion of interaction by synchronization on labels. These have been studied intensively in the last few years ([6,7,8,15]). 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

⋆ *Corresponding author:* Équipe PPS, Université Paris 7 Denis Diderot, Case 7014, 2 Place Jussieu 75251 Paris Cedex 05, Vincent.Danos@pps.jussieu.fr

R. Amadio, D. Lugiez (Eds.): CONCUR 2003, LNCS 2761, pp. 477–491, 2003.

(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. [7] 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 [8,9] 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.

In a recent paper Danos and Desharnais [5] have developed a variant of the approximation that has two important advantages. First, the approximations can be "guided" by a family of formulas of interest. In other words, if there is a set of formulas of particular interest one can construct a specific finite approximant geared towards these formulas. One can then be sure that the process in question satisfied a formula of interest if and only if the approximant did. Second, a much more compact representation was used so that loops were not unwound and convergence was attained more rapidly. A disadvantage was that the approximations obtained were not LMPs because the transition "probabilities" are not measures. Instead they were capacities [2]. Capacities are not additive but they have instead a continuity property and are sub (or super) additive. The variants of LMPs obtained by using capacities instead of measures are called pre-LMPs.

In the present paper we show that we can have the best of both worlds in the sense that we can have the flexibility of 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 [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 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, written $\mathbb{E}_p(X|\Lambda)$, which 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 the construction given here, 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 [12] and integration on domains [10,11] 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.

Outline. First we recall the definitions of our two basic objects of concern, LMPs and conditional expectations. Then we identify circumstances in which the conditional expectation is actually defined pointwise and not only "almost everywhere". We construct an adaptation of 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. We show they are correct both by a direct argument and by showing the precise relation in which they stand with the order-theoretic approximants given in [5].

2 Preliminaries

2.1 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 $Q' \in \Sigma'$, $f^{-1}(Q') \in \Sigma$. Writing $\sigma(f)$ for the σ-algebra generated by f, namely the set of sets of the form $f^{-1}(Q')$ with $Q' \in \Sigma'$, 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 $m\Sigma$.

A *subprobability* on (S, Σ) is a map $p : \Sigma \to [0, 1]$, such that for any countable collection (Q_n) of pairwise disjoint sets in Σ, $p(\bigcup_n Q_n) = \sum_n p(Q_n)$. An actual probability is when in addition $p(S) = 1$. The condition on p is called σ*-additivity* and can be conveniently broken in two parts:

— *additivity*: $p(Q \cup Q') = p(Q) + p(Q')$, for Q, Q' disjoint,
— *continuity*: $\forall \uparrow Q_n \in \Sigma : p(\cup Q_n) = \sup_n p(Q_n)$.[1]

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 $Q \in \Sigma$ such that $N \subseteq Q$ and $p(Q) = 0$.

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.

2.2 Labelled Markov Processes

We need to define the objects of study:

Definition 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 $a \in L$, $Q \in \Sigma$, $h(a, s, Q)$ is Σ-measurable as a function of s;*
— *for all $s \in S$, $h(a, s, Q)$ is a subprobability as a function of Q.*

Some particular cases: 1) when S is finite and $\Sigma = 2^S$ we have the familiar probabilistic transition system, 2) when $h(a, s, Q)$ 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.

Second we see that equivalently LMPs can be defined as follows:

Definition 2 (LMP2). *A Labelled Markov Process consists of a measurable space (S, Σ) and a family of Σ-measurable functions $(h(a, Q))_{a \in L, Q \in \Sigma}$ with values in $[0,1]$, such that:*
— *additivity: for all disjoint Q, Q': $h(a, Q \cup Q') = h(a, Q) + h(a, Q')$;*
— *continuity: for all increasing sequence $\uparrow Q_n$: $h(a, \bigcup Q_n) = \sup h(a, Q_n)$.*

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

In this second definition we see an LMP as a Σ-indexed family of Σ-measurable functions, namely the random variables "probability of jumping to Q in one step labelled with a", instead of an S-indexed family of probabilities on Σ. Both definitions are related by $h'(a, s, Q) = h(a, Q)(s)$. The functions h, h' are commonly referred to as transition probability functions or Markov kernels (or stochastic kernels).

In previous treatments [6] LMPs were required to have an analytic state space. This was needed for the proof of the logical characterization of bisimulation. We will not mention this again in the present paper since we will not need the analytic structure. In fact it is hard to give examples of spaces that are not analytic, let alone one that might be useful in an example.

[1] Where $\uparrow Q_n$ denotes an increasing sequence of sets Q_n, *i.e.*, for all n, $Q_n \subseteq Q_{n+1}$.

2.3 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 Q then my revised estimate of the expectation is $\mathbb{E}_p(X)$ is $\mathbb{E}_p(X|Q)$." 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 Q. Now she may try to recompute her expectation values based on this information. To know that the outcome is in Q also means that it is *not* in Q^c. Note that $\{\varnothing, Q, Q^c, S\}$ 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 $Q \in \Lambda$ whether the outcome is in Q 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 is written $\mathbb{E}_p(\_|\Lambda)$.[2]

It is not at all obvious that such a function should exist and is indeed a fundamental result of Kolmogorov (see for instance [16], p.84).

Theorem 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*, *increasing* 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 $\mathbb{E}_p(X_n|\Lambda) \uparrow \mathbb{E}_p(X|\Lambda)$ ... but it is *not* uniquely defined !

[2] 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 Q and Q^c. This will turn out to be exactly what we need later when pixels are defined by sets of formulas.

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 set of p-probability zero.

2.4 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 $\mathbb{E}_p(X|\Lambda) = \mathbb{E}_p(X)$,[3] and X is completely averaged out. On the other hand, if X is fully dependent on Λ, in other words if X is Λ-measurable, then $\mathbb{E}_p(X|\Lambda) = X$.[4]

Actually 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 (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 characteristic 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:

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

2.5 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 $\{z, 0, 1\}$. One can define the kernel as the *infimum* over $\{x, y\}$ or dually one can

[3] 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.

[4] 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.

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 else one can *average* (using here the uniform probability):

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

As we said in the introduction, 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 using averages preserve 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 the paper 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.

2.6 Logic and Metric

The other goal of having approximants that are customizable with respect to formulas of interest will be achieved by using the notion of expectation above with Λ a σ-algebra generated by a set of formulas of a suitable logic. We will prove that the approximant satisfies exactly the same formulas of the given set as does the process being approximated.

The following logic $\mathcal{L}_0$ is a central tool for asserting properties of LMPs, since it characterizes strong bisimulation between them [6].

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

The parameter r above can be any rational in $[0,1]$.

Definition 3. *Given an LMP $\mathcal{S}$, one defines inductively the map $[\![.]\!]_{\mathcal{S}} : \mathcal{L}_0 \to \Sigma$ as:*
— $[\![\top]\!]_{\mathcal{S}} = S$,
— $[\![\theta_0 \wedge \theta_1]\!]_{\mathcal{S}} = [\![\theta_0]\!]_{\mathcal{S}} \cap [\![\theta_1]\!]_{\mathcal{S}}$,
— $[\![\langle a \rangle_r \theta]\!]_{\mathcal{S}} = \{s \in S \mid h(a, [\![\theta]\!]_{\mathcal{S}})(s) \geq r\}$.

Let $\mathcal{S}$ be an LMP, one says $s \in S$ satisfies θ, written $s \models \theta$, if $s \in [\![\theta]\!]_{\mathcal{S}}$; one says $\mathcal{S}$ satisfies θ, still written $\mathcal{S} \models \theta$, if there exists an $s \in S$ such that $s \models \theta$. Finally given another LMP $\mathcal{S}'$ and $\mathcal{F}$ a subset of formulas of $\mathcal{L}_0$, we write $\mathcal{S} \approx_{\mathcal{F}} \mathcal{S}'$ if $\mathcal{S}$ and $\mathcal{S}'$ satisfy the same formulas of $\mathcal{F}$.

As we have already said, we will take the simplifying stance that two LMPs are bisimilar iff they satisfy the same formulas. This was proven in the case of analytic state spaces and that is a general enough class to encompass any conceivable physical example.

In [7] a family of metrics, d^c for $c \in (0,1)$, has been introduced that is closely related to this logic. Indeed one can think of the metric as measuring the complexity of the distinguishing formula between two states if any.

We do not need to give the precise definition of these metrics here, but we do want to use it to show convergence of approximants. This will be done using the following result which is a direct consequence of results relating the logic and the metric that can be found in [7].

Proposition 4. *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. Then if $\mathcal{S}_i \approx_{\mathcal{F}_i} \mathcal{S}$ for every set $\mathcal{F}_i$ of formulas of $\mathcal{L}_0$, then for all $c \in (0,1)$:*

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

3 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 $h(a,Q)$ 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 (pixels). *Let (S,Σ) be a measurable space, one says s and $t \in S$ are Σ-indistinguishable if $\forall Q \in \Sigma$, $s \in Q \leftrightarrow t \in Q$.*

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

Definition 6 (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 containing 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 7 (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 $Q := \{s \in S \mid X(s) = \alpha \wedge Y(s) = \beta\}$ and $t \in Q$. One has $Q \in \Lambda$, by Λ-measurability of X and Y, but then $[t]_\Lambda \subseteq Q$ (otherwise Q splits $[t]_\Lambda$). So by granularity $p(Q) > 0$ (else $[t]_\Lambda$ is negligible), and therefore $\alpha = \beta$ or else X and Y differ on a non negligible set Q. □

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.

4 Projecting LMPs

Definition 8 (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 $\mathcal{S} = (h(a,Q))_{a \in L, Q \in \Sigma}$ an LMP on (S, Σ), one defines the* p-projection *of $\mathcal{S}$ on Λ, written $(\mathcal{S}|\Lambda)_p$ as:*

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

Take note that this is *the* version of the conditional expectation. Existence follows from the fact that the $h(a,Q)$ 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 9 (Staying within LMPs). *$(\mathcal{S}|\Lambda)_p$ is an LMP.*

Proof. All maps $h'(a,Q)$ are Λ-measurable by definition of the conditional expectatioin; additivity is because $\mathbb{E}_p(\_|\Lambda)$ is linear; continuity follows because $\mathbb{E}_p(\_|\Lambda)$ is continuous as can be seen by using 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 $h'(a,Q)$ is Λ-measurable, and therefore $h'(a,Q)$ 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" [6]) 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 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:[5]

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

Then for each n, we set $\tau_n := 1_{[\![\theta_n]\!]_\mathcal{S}}$ 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 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 some "skewed" Lebesgue measure p on $\sigma(\mathcal{L}_0)$, the σ-algebra generated by our temporal formulas.[6]

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

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

First, we have to remind for future use that for any finite set of formulas $\mathcal{F} \subset \mathcal{L}_0$ and $\Lambda_\mathcal{F}$ the associated σ-algebra:

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

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

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 sytem with only state s, only one letter a and $h(a, \{s\})(s) = 1/2$. Then $s \models \theta$ iff 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 $p(\{s\}) = 1$.

5.1 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:

Proposition 10. *$\sigma(\mathcal{L}_0)$ is the smallest σ-algebra which is closed under the shifts:*

$$\langle a \rangle_r(Q) = \{s \mid h(a, s)(Q) \geq r\}$$

That it is the smallest is obvious, but that it is closed is not [3].

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

6 Approximations

Now we can complete the approximation construction.

6.1 Finite-State Approximants

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:

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

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 sytem.

6.2 Convergence

We need to say how the obtained $\mathcal{S}_{\mathcal{F}}$ approximates $\mathcal{S}$. In the previous approach [8], approximants were always below the approximated process and hence simulated by it. It was shown that they converge in the domain of all LMPs. It is not the case here since approximants are neither above nor below $\mathcal{S}$. However, $\mathcal{S}_{\mathcal{F}}$ does converge to $\mathcal{S}$.

Proposition 11. *For every finite subformula-closed set of formulas $\mathcal{F} \subset \mathcal{L}_0$: $\mathcal{S}_{\mathcal{F}} \approx_{\mathcal{F}} \mathcal{S}$.*

Proof. We prove something stronger, namely that if $\theta \in \mathcal{F}$, then $\cup[\![\theta]\!]_{\mathcal{S}_{\mathcal{F}}} = [\![\theta]\!]_{\mathcal{S}}$ or equivalently that $[\![\theta]\!]_{\mathcal{S}_{\mathcal{F}}} = [[\![\theta]\!]_{\mathcal{S}}]_{\Lambda}$ (recall that $\Lambda = \sigma(\mathcal{F})$). This is done by induction on the structure of formulas in $\mathcal{F}$, which is why we ask $\mathcal{F}$ to be closed by subformulas.
The only interesting case is when $\theta = \langle a \rangle_r \phi$. If all states in class $[s]$ satisfy θ (equivalently if one state in $[s]$ satisfies θ), that is to say if $h(a, [\![\phi]\!]_{\mathcal{S}})(t) \geq r$ for all $t \in [s]$, then obviously the conditional expectation is also above r and hence $[s]$ satisfies θ since:

$$k(a, [\![\phi]\!]_{\mathcal{S}_{\mathcal{F}}})[s] := \mathbb{E}_p(h(a, \cup[\![\phi]\!]_{\mathcal{S}_{\mathcal{F}}})|\Lambda)(t) = \mathbb{E}_p(h(a, [\![\phi]\!]_{\mathcal{S}})|\Lambda)(t) \geq r,$$

where the first equation is by definition of k, and the second equation is by induction hypothesis. Conversely, if the conditional expectation on $[s]$ (recall that it is constant on this set) is $\geq r$, then at least one $t \in [s]$ must satisfy $h(a, [\![\phi]\!]_{\mathcal{S}})(t) \geq r$. Since all states in $[s]$ satisfy the same formulas of $\mathcal{F}$, then they all satisfy formula θ, as required. □

Notice that this proposition is also true for a logic extended with a greatest fixpoint operator [4,5].[7]

From Proposition 4, it follows now easily that:

Theorem 2. *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 11. As the example 2.5 suggested, quotients constructed with conditional expections do lie between the inf- and the sup- approximants [5]:

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

[7] The proof can be found in a survey of LMP approximation (to appear in ENTCS).

The second equation holds both because $h'(a, Q)$ 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 11 [5]. This also makes clear that the average-based approximants are better than the order-theoretic ones.

7 Conclusion

We have given an approximation technique for LMPs that has a number of good properties. It can be customized in the sense that if one is interested in a special set of formulas one can arrange the approximation so that one obtains a finite system (assuming that one had finitely many formulas) with the property that the formulas of interest are satisfied by the original system if and only if they are satisfied by the finite approximant. This brings the work much closer to the goal of using automated verification tools on continuous state space systems. This property is shared by the infima technique [5] however, unlike that result, we can also stay within the framework of traditional LMPs and avoid having to work with capacities.

The results of this paper give yet another approximation construction and one may well wonder if this is just one more in a tedious sequence of constructions that are of interest only to a small group of researchers. In fact, we feel that there are some new directions in this work whose significance extends beyond the properties of the construction. 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 [13] – 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 [14].

One might ask why we do not mention any analytic space property contrarily 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, as reported in [9]. This follows essentially from well known facts about analytic spaces, see for example chapter 3 of "Invitation to C^*-algebras" by Arveson [1].

Acknowledgements. Josée Desharnais would like to acknowledge the support of NSERC and FCAR. Prakash Panangaden would like to thank NSERC for support.

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 labeled Markov Processes. In Zoltan Esik and Igor Walukiewicz, editors, *Proceedings of an international Workshop FICS'03 (Fixed Points in Computer Science)*, 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*, Ottawa, 2003. IEEE.
6. J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labelled Markov processes. *Information and Computation*, 179(2):163–193, Dec 2002. Available from `http://www.ift.ulaval.ca/~jodesharnais`.
7. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled Markov processes. In *Proceedings of CONCUR99*, Lecture Notes in Computer Science. Springer-Verlag, 1999.
8. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Approximating continuous Markov processes. In *Proceedings of the 15th Annual IEEE Symposium On Logic In Computer Science, Santa Barbara, Californie, USA*, 2000. pp. 95-106.
9. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Approximating labeled Markov processes. *Information and Computation*, 2003. To appear. Available from `http://www.ift.ulaval.ca/~jodesharnais`.
10. 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.
11. Abbas Edalat. Domain theory and integration. *Theoretical Computer Science*, 151:163–193, 1995.
12. 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.
13. Keye Martin. The measurement process in domain theory. In *International Colloquium on Automata, Languages and Programming*, pages 116–126, 2000.

14. P. Panangaden and V. Shanbhogue. The expressive power of indeterminate dataflow primitives. *Information and Computation*, 98(1):99–131, 1992.
15. Franck van Breugel, Michael Mislove, Joël Ouaknine, and James Worrell. Probabilistic simulation, domain theory and labelled Markov processes. To appear in the Proceedings of FOSSACS'03, 2003.
16. David Williams. *Probability with Martingales*. CUP, Cambridge, 1991.