Weak Bisimulation is Sound and Complete for PCTL*

Josée Desharnais[1⋆], Vineet Gupta[2], Radha Jagadeesan[3∗], and Prakash Panangaden[4∗]

[1] Département d'Informatique,Université Laval,Québec, Canada, G1K 7P4
[2] Stratify Inc., 501 Ellis Street, Mountain View CA 94043 USA
[3] Dept. of Computer Science, Loyola University-Lake Shore Campus, Chicago IL 60626, USA
[4] School of Computer Science, McGill University, Montreal, Quebec, Canada

Abstract. We investigate weak bisimulation of probabilistic systems in the presence of nondeterminism, i.e. labelled concurrent Markov chains (LCMC) with silent transitions. We build on the work of Philippou, Lee and Sokolsky [1] for £nite state LCMCs. Their de£nition of weak bisimulation destroys the additivity property of the probability distributions, yielding instead *capacities*. The mathematics behind capacities naturally captures the intuition that when we deal with nondeterminism we must work with estimates on the possible probabilities.
Our analysis leads to three new developments:

- We identify an axiomatization of "image £niteness" for countable state systems and present a new de£nition of weak bisimulation for these LCMCs. We prove that our de£nition coincides with that of Philippou, Lee and Sokolsky for £nite state systems.
- We show that bisimilar states have matching computations. The notion of matching involves *linear combinations* of transitions. This idea is closely related to the use of randomized schedulers.
- We study a minor variant of the probabilistic logic pCTL* — the variation arises from an extra path formula to address action labels. We show that bisimulation is sound and complete for this variant of pCTL*.

1 Introduction

The main object of this paper is to study systems that combine probability, concurrency and nondeterminism. We focus in particular on weak bisimulation. The importance of weak bisimulation comes from the need for abstraction. In order to construct larger programs from smaller programs one works with the composition mechanisms of the language. When doing so it is necessary to hide internal actions and work with weak (rather than strong) bisimulation.

In the purely probabilistic context, the study of strong bisimulation was initiated by Larsen and Skou [2], and an equivalence notion was developed, similar to the queuing theory notion of "lumpability" [3]. This theory has been extended to continuous state spaces and continuous distributions [4–6] and, in the discrete setting, to weak bisimulation [7].

The study of weak bisimulation for systems with probability and non-determinism is sensitive to the underlying model. The two principal models are the *alternating*

⋆ Research supported by NSERC, NSF and MITACS.

model [8] - where there are two disjoint classes of states, probabilistic states and nondeterministic states - and the nonalternating model [9]. Weak bisimulation for £nite-state systems in the alternating model with distinct nondeterministic and probabilistic states was de£ned by Philippou, Lee and Sokolsky [1] whereas weak bisimulation for the non-alternating model was studied by Segala and Lynch [9]. Our study is set in the context of the alternating model and follows [1].

We explore the subtle consequences of the benign looking de£nitions of [1]. The most signi£cant change from ordinary probability theory is that the "probabilities" no longer satisfy additivity[5]. In the presence of nondeterminism, we are describing a *set* of probability distributions $\{Q_i\}$ for a given state s and a given weak transition label a. The "probabilities" ascribed by [1] arise by majorizing over this set, i.e. $P(s, a, E)$, the probability of reaching a set of states E from state s on weak transition labelled a, is given by $\max_i Q_i(E)$ for any subset of states E.

The second important change is that the notion of matching has changed radically. The essence of any bisimulation notion is that transitions of one process can be matched with transitions in the bisimilar process. In order to match computation paths on given weak labels one can (and needs to!) take linear combinations of computations. The "computations" (to be de£ned precisely later) now have a vector space structure. In example 2 we discuss this point in detail. Essentially randomized schedulers allow one to take just such linear combinations.

The three main points that we make can be summarized as follows.

- First, we generalize the de£nitions of [1] to a large class of in£nite-state systems satisfying a compactness property. Informally, compactness is a topological formalization of £nite branching. In this context, compactness enables us to capture a robust notion of "image £niteness" for weak transitions that hide internal actions. The compact systems that we consider include all £nite state systems (including those with cycles).
- Second, we adapt the ideas on randomized schedulers from Segala's work on probabilistic IO automata [10]. On the one hand, randomized schedulers do not change the semantics (the sups that one computes are the same). On the other hand, these schedulers enable us to perform a £ne-grained analysis of the structure of computations in bisimilar systems. This analysis permits us to establish that bisimilar states s, t satisfy a familiar property: "for every distribution of states induced by a resolution of non-deterministic choices from s, there exists a resolution of non-deterministic choices from t that results in a matching distribution on states". We show simple examples that demonstrate that this matching property *requires* the presence of linear combinations.
- Third, we analyze the structure that arises by majorizing over a set of probability distributions. This structure is called a capacity — for our purposes, capacities are monotone functions from a Borel algebra to the reals that preserve sups (resp. infs) of increasing (resp. decreasing) sequences of sets. Capacities are not necessarily additive. Indeed, the capacities induced by the de£nitions of [1] only satisfy: $P(s, a, A) + P(s, a, B) \geq P(s, a, A \cup B)$ for disjoint sets of states A, B.

[5] Additivity: P is additive if for disjoint sets A, B, $P(A \cup B) = P(A) + P(B)$.

This loss of additivity has already been recognized in various situations in mathematics [11–13] and in economics [14][6], and a rich theory was already available for our use. This theory meshes very well with the idea that uncertainty in probability distributions should be captured by giving upper and lower bounds on probabilities and expectation values. We show that the key equations that are demanded by this theory are met by the capacities that arise in the context of weak bisimulation.

Soundness and Completeness of weak bisimulation for probabilistic logics. A fundamental application of these ideas and the original impetus for these investigations is the analysis of soundness and completeness of bisimulation for probabilistic logics. We study a minor variant of the probabilistic logic pCTL* [15] – the variation arises from an extra path formula to address action labels – and is inspired by the variants of probabilistic logics that deal with action labels [9, 8] . We show that bisimulation is sound and complete for this variant of pCTL*. Our soundness and completeness proofs relies crucially on all three developments identi£ed above.

Organization of this paper. The rest of this paper is organized as follows. First, in section 2, we review the basic de£nitions of the model (the "alternating model") and weak probabilistic bisimulation and associated results to make the paper self-contained. Section 3 identi£es the class of countable systems to which our study applies. In section 4 we show that our de£nition is equivalent to that of Philippou, Lee and Sokolsky [1]. In section 5 we show that the capacities de£ned in the development of weak bisimulation satisfy the axioms required of capacities. Finally, in section 6, we use the machinery that has been developed to prove soundness and completeness results for the logic.

2 Background and De£nitions

We begin with a review of the underlying framework — our de£nitions are adapted from [1]. We work in the context of the "alternating model" for labelled concurrent Markov chains [8], labelled transition systems with non-determinism and probability.

De£nition 1. *A labelled concurrent Markov chain (henceforth LCMC), is a tuple* $\mathcal{K} = (K, \mathtt{Act}, \longrightarrow, k_0)$*, where*
(1) $K = K_p \cup K_n$*, a countable set, is partitioned into the probabilistic states,* K_p*, and the nondeterministic states* K_n*.* k_0 *is the start state.*
(2) $\mathtt{Act}$ *is a £nite set of action symbols that contains a special action* τ*.*
(3) The transition relation $\longrightarrow = \longrightarrow_p \cup \longrightarrow_n$ *is partitioned into probabilistic and nondeterministic transitions.* $\longrightarrow_n \subseteq K_n \times \mathtt{Act} \times K_p$ *is* image-£nite*, i.e. for each* $s \in K_n$ *and* $a \in \mathtt{Act}$*, the set* $\{s' \in K_p \mid s \xrightarrow{a} s'\}$ *is £nite.* $\longrightarrow_p \subseteq K_p \times (0, 1] \times K_n$ *satis£es that for each* $s \in K_p$*,* $\sum_{(s,\pi,t) \in \longrightarrow_p} \pi = 1$.

[6] Economic studies distinguish risk (the relative probabilities of the events are known) from uncertainty (there is no unique assignment of probabilities to events) - this is what computer scientists call nondeterminism. Risk is modelled using probability. The modelling of uncertainty is via a set of probability measures that are consistent with the known information. The structure obtained by majorizing this set of probability measures does not satisfy additivity and is a capacity.

A state is either probabilistic - in which case the transitions are probabilistic and unlabelled - or nondeterministic, in which case the transitions are £nite-branching and labelled (possibly by a τ). The probabilistic branching can be countable at a state. In this paper, we will work with countable state systems.

Every probabilistic state s induces a probability distribution Q on K_n given by $Q(t) = \sum_{(s,\pi,t)\in\longrightarrow_p} \pi$ for every $t \in K$. We sometimes write $s \rightarrow_p Q$ to emphasize this distribution. Indeed, one can take the view that the "real" states are the nondeterministic states and the probabilistic states are really just names for certain probability distributions.

The LCMC model does not need to be strictly alternating. One can work with a model that only restricts states to be either purely nondeterministic or purely probabilistic and does not enforce strict alternation.

We use some notation for sequences (of states or transitions). We use ε for the empty sequence and $\cdot$ for concatenation. Every sequence, say σ, of transitions has as an associated probability $\mathtt{prob}(\sigma)$, obtained by multiplying the probabilities occurring on the path. Thus, we attribute 1 to a nondeterministic transition in a path, and multiply together probabilities of all the probabilistic transitions. Similarly, every sequence σ of transitions has an associated weak sequence of labels $\mathtt{Weak}(\sigma) \in (\mathtt{Act} - \{\tau\})^*$, obtained by removing the labels of τ-transitions. Thus, probabilistic transitions and nondeterministic transitions with label τ do not contribute to the weak label. We use τ for the empty sequence as well as for the empty transition. Thus we will say that a path of τ transitions and probabilistic transitions has weak label τ.

We de£ne *computations* of an LCMC as transition trees obtained by unfolding the LCMC from the root, resolving the nondeterministic choices (i.e. each nondeterministic state has at most one transition coming out of it) and taking all probabilistic choices at a probabilistic state. A computation can thus be viewed as a purely (sub)probabilistic labelled Markov chain. We refer to the set of all the probabilistic transitions from a probabilistic state as a *fan*.

De£nition 2. *A* computation *of an LCMC is a (possibly in£nite) subtree of the tree obtained by partially unfolding the LCMC. In a computation every nondeterministic state has at most one transition coming out of it and if a probabilistic transition is included then the entire fan of that probabilistic transition is included.*

We are interested in transitions with particular weak labels.

De£nition 3. *Let $\mathcal{K}$ be a LCMC, $a \in$ $\mathtt{Act}$. An a-computation from $s \in K$ is a computation such that every path from the root has weak label a or ε.*

It may seem peculiar to allow an a-computation to have paths labelled by ε. This is done to allow for a computation where the a transition has not happened yet (or may never happen). However, when we associate probability distributions with computations we will not count the paths labelled with ε, we insist that the paths that contribute to the distribution have weak label a.

Each computation induces a distribution on its leaf states in the standard way — the probability of a leaf node is the probability of the (unique) path going to it. We actually use a somewhat looser correspondence between computations and distributions. We allow many distributions to be induced by a given computation; the requirement of matching is weakened to an inequality. This will turn out to be very convenient when constructing certain sequences of weak transitions, for example in proving Lemma 1.

Definition 4. *Let $\mathcal{K}$ be a LCMC, $s \in K$, and let Q be a distribution on states. We write $s \stackrel{a}{\Rightarrow} Q$, if there is an a-computation such that for all $s_i \in K$, $Q(s_i) \leq \sum_\sigma \mathtt{prob}(\sigma)$ where the summation is taken over paths σ with weak label a that start in s and end in the leaf s_i.*

We extend this notation to linear combinations of distributions. $s \stackrel{a}{\Rightarrow} \sum_i \lambda_i \times Q_i$ is an a-transition from s to the distribution $\sum_i \lambda_i \times Q_i$. This is where the linear structure becomes explicit. Such a transition can be viewed as the "weighted superposition" of the transitions $s \stackrel{a}{\Rightarrow} Q_i$.

Definition 5. *Let $s_i \stackrel{a}{\Rightarrow} Q_i$ and let $\sum_i \lambda_i \leq 1$, where all $\lambda_i \geq 0$. Then we write: $\sum_i \lambda_i \times (s_i \stackrel{a}{\Rightarrow} Q_i)$. In the special case where all $s_i = s$, we write $s \stackrel{a}{\Rightarrow} \sum_i \lambda_i \times Q_i$ to represent an a-transition from s to the distribution $\sum_i \lambda_i \times Q_i$.*

We thus have linear (vector-space) structure on the space of computations and on the space of distributions. Note that when we write $s \stackrel{a}{\Rightarrow} Q$ we refer to the general case of transitions of the form $s \stackrel{a}{\Rightarrow} \sum_i \lambda_i \times Q_i$: when we want to refer to transitions that are not weighted combinations we use the term "basic". For $s \neq t$, the notation $\lambda \times (s \stackrel{a}{\Rightarrow} Q_1) + (1-\lambda) \times (t \stackrel{a}{\Rightarrow} Q_2)$ is merely notational convenience. Note that $s \stackrel{a}{\Rightarrow} [\lambda \times Q_1 + (1-\lambda) \times Q_2]$ is reminiscent of the randomized schedulers [10].

Transitions from states to distributions as above are one way to the defintion of bisimulation. Another way is through transitions from states to sets of states, which is how strong bisimulation is defined for labelled Markov processes in [4, 6]. The "probability" from a state s to a subset of states via a path with weak label a is defined by taking the supremum over all possible a-computations.

Definition 6. *Let $\mathcal{K}$ be a LCMC, $s \in K, E \subseteq K$. Then, the probability of going from s to $E \subseteq K$ via a, denoted by $P(s, a, E)$, is defined as:*

$$P(s, a, E) = \sup\{\sum_{t \in E} Q(t) \mid s \stackrel{a}{\Rightarrow} Q\}.$$

The supremum in this definition is the source of the subtlety of weak bisimulation — $P(s, a, .)$ does not satisfy additivity.

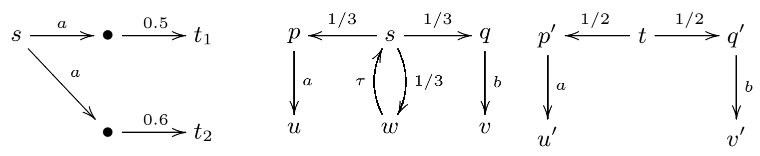

Fig. 1. (a)Additivity Fails (b) Matching with linear combinations

Example 1. Consider the transition system in £gure 1(a). Then $P(s, a, \{t_1\}) = 0.5$, $P(s, a, \{t_2\}) = 0.6$, $P(s, a, \{t_1, t_2\}) = 0.6$. Thus additivity does not hold. This example also illustrates in a trivial way why we must take the sup over all computations in the de£nition of $P(s, a, E)$.

The next example shows the importance of allowing linear combinations when matching computations with given weak labels.

Example 2. Consider the transition systems of £gure 1(b). Intuitively we would like to say that the states s and t are weakly bisimilar. We would also like to say p, p' and q, q' are weakly bisimilar.

The probability of starting from s and reaching u on a weak a label is $1/2$ and the same is true for reaching u' from t. Note that we need to sum over all possible paths that include the τ-loop if we want to get the answer $1/2$ starting from s. Thus the a-computation from t that includes u' gives a probability of $1/2$ to u' and can be matched by the in£nite computation from s that loops in£nitely through w and gives probability $1/2$ to u. However, we have absolutely no way of matching the distribution induced by the computation including only one step from s. Indeed, this computation induces the distribution that gives probability $1/3$ to each one of u, w and v. The only way to match it is to take a linear combination, namely the distribution δ_t induced by the trivial computation consisting only of state t, and the distribution P induced by the one-step computation. The required combination is thus $1/3 \times \delta_t + 2/3 \times P$.

We are now ready to de£ne weak bisimulation. Given an equivalence relation R, we say a set E is R-closed if $E = \mathcal{C}l_R(E) := \{s \mid \exists t \in E \text{ such that } tRs\}$.

De£nition 7. *An equivalence relation R on K is a weak bisimulation iff for all $s, t \in K$ such that $s\ R\ t$ and all R-closed $E \subseteq K$, we have:*

$$(\forall a \in \texttt{Act})\ [P(s, a, E) = P(t, a, E)].$$

There is a maximum weak bisimulation, denoted by $\approx$. We write $[u]$ for the bisimulation class of the state u.

A LCMC $\mathcal{K}$ is *bisimulation collapsed* if for any state, the targets of all transitions are in distinct bisimulation classes.

The equational laws supported by this de£nition extend the usual ones for nondeterministic labelled transition systems or purely probabilistic transition systems. Indeed, the usual relations that witness the bisimulation are carried over essentially unchanged, for example, $\tau.\mathcal{K} \approx \mathcal{K}$, and unfolding a LCMC yields a weakly bisimilar system. See [16] for a full axiomatization of equational laws for £nite processes (without loops, so the transition system is a tree).

We present a second de£nition of bisimulation which is similar to the one found in the non-probabilistic setting. It will be shown to be equivalent to the one above in Section 4 for *compact* LCMCs. We use $[u]_R$ to stand for the equivalence class of a state u for an equivalence relation R.

De£nition 8. *An equivalence relation R on K is a weak-$*$ bisimulation iff for all $s, t \in K$ such that $s\ R\ t$, we have:*
if $s \stackrel{a}{\Rightarrow} Q$, there exists $t \stackrel{a}{\Rightarrow} Q'$ such that for all states u: $Q([u]_R) = Q'([u]_R)$. We will denote it $\approx_*$.

3 The compactness condition

We consider countable-state LCMCs that satisfy a compactness condition. Intuitively speaking, the compactness condition can be viewed as the right generalization of "image-£niteness" for countable state LCMCs in the context of weak transitions that hide τ-labels.

We £rst consider some preliminary motivation for considering such a condition. In general, it is not the case – even for £nitely branching systems – that there is a single computation that attains the supremum of de£nition 6.

Example 3. Let $\mathcal{K}$ be the LCMC described by the following diagram.

$$\begin{tikzcd}
s_0 \arrow[r, "\tau"] \arrow[d, "a"] & u_0 \arrow[r, "1"] & s_1 \arrow[r, "\tau"] \arrow[d, "a"] & u_1 \cdots \arrow[r, "1"] & s_i \arrow[r, "\tau"] \arrow[d, "a"] & u_i \cdots \\
t_0 \arrow[drr, "1-\frac{1}{2^0}"] & & t_1 \arrow[d, "1-\frac{1}{2^1}"] & & t_i \arrow[dll, "1-\frac{1}{2^i}"] & \cdots \\
& & s_a & & &
\end{tikzcd}$$

Clearly, $P(s_0, a, \{s_a\}) = 1$, but there is no single computation to witness this.

We diagnose the reason as the in£nite (weak) branching at the state s_0. We now identify a large class of countable systems the class of systems that we will work with. Intuitively, this is a "compactness" condition that captures the essence of a "£nite weak branching" requirement.

We begin with the de£nition of a metric d on distributions of states.

De£nition 9. *The metric d on distributions of the states of a LCMC $\mathcal{K}$ is de£ned by $d(Q_1, Q_2) = sup_{A \subseteq K} |Q_1(A) - Q_2(A)|$.*

In this metric, any computation is the limit of £nite depth computations.

Lemma 1. *Given any weak transition $s \stackrel{a}{\Rightarrow} P$, one can £nd a sequence of £nite-depth computations with corresponding weak transitions $s \stackrel{a}{\Rightarrow} P_i$ with the P_i distributions converging to P in the metric d.*

De£nition 10. *Let $\mathcal{K}$ be the LCMC and s be a state and a any label. We say s is a-compact if the set $\{Q \mid s \stackrel{a}{\Rightarrow} Q\}$ is compact*[7] *under metric d.*

A bisimulation collapsed LCMC $\mathcal{K}$ is compact *if all states s are a-compact for all labels a (including τ).*

A LCMC $\mathcal{K}'$ is compact *if its bisimulation collapse is compact.*

For labelled transition systems, the compactness condition is an image-£niteness condition. Here the probability of all paths is 1 and d is the discrete metric. So, an LTS

[7] A subset of a metric space is compact if every in£nite set has a limit point, ie. given any in£nite S, $(\exists p)$ $(\forall \epsilon)$ $(\exists x \in S)$ $d(p, x) < \epsilon$.

is compact iff for all states s and all labels a, the set of states reachable on a weak transition labelled a is £nite.

The de£nition is general enough to include all £nite state systems. Weighted combinations of computations are crucial to this proof. The proof builds on the idea of Example 2. It shows that for any state s, there is a £nite set of computations rooted at s such that any computation rooted at s can be built as a weighted combination of the elements of this set.

Theorem 1. *All £nite state systems are compact.*

For compact countable-state systems, there is a single computation yielding the maximum probability, thus resolving the issue raised by Example 3.

Lemma 2. $P(s, a, E) = \sum_{s \in E} Q(s)$ *for some* $s \overset{a}{\Rightarrow} Q$.

4 Coincidence with the de£nition of Philippou, Lee and Sokolsky

Our formulation of bisimulation (De£nition 7) is different from the de£nition in [1]. However, the two de£nitions are equivalent.

We begin by presenting their de£nition below — we have recast it in terms of computations rather than schedulers. Recall that $[u]_R$ stands for the equivalence class of a state u for an equivalence relation R. Let C be an a-computation starting from s, we write $P^C(s, a, \cdot)$ for the distribution induced on the leaves of C.

De£nition 11. *An equivalence relation R on K is a PLS-weak bisimulation if for all $s, t \in K$ such that whenever sRt, then*

- *if $s \in K_n, a \in Act$ and $(s, a, s') \in \longrightarrow$, then there exists a computation C such that $P^C(t, a, [s']_R) = 1$.*
- *if $s \in K_p$ with $s \rightarrow_p Q$, then there exists a computation C such that*

$$\forall M \in K/R - [s]_R, P^C(t, \varepsilon, M) = \frac{Q(M)}{1 - Q([s]_R)}.$$

There is a maximum weak bisimulation, denoted by $\approx_{PLS}$.

The term $\frac{Q(M)}{1-Q([s]_R)}$ represents the conditional probability of reaching M from s in one step given that the system leaves the equivalence class of s in its £rst step.

For compact LCMCs (and hence all £nite state LCMCs), $\approx$ and $\approx_{PLS}$ coincide. The proof of this theorem requires weighted combinations of computations, as illustrated by Example 2. The role of these weighted linear combinations is seen in the case $(2) \Rightarrow (3)$ in the following proof.

Theorem 2. *The following are equivalent for compact LCMCs.*

1. $s \approx t$.
2. $s \approx_{PLS} t$.
3. $s \approx_* t$.

Proof. We sketch the main ideas below.

– (1) $\Rightarrow$ (2): The key structural properties exploited in the proof are:
  - If t is a nondeterministic state, and s is a probabilistic state, such that t is weakly bisimilar to s, then there is a τ-transition from t to some t' such that t' is weakly bisimilar to s.
  - we can show that $\approx$-bisimilar probabilistic states have identical (upto $\approx$) probabilistic fans.

– (2) $\Rightarrow$ (3): We show this with $\approx_{PLS}$ as the equivalence relation in the De£nition 8. Using Lemma 1, it suf£ces to prove the result for £nite-depth computations Q. In this case, the proof proceeds by induction on depth.
  - Let C extend $s \stackrel{a}{\Rightarrow} Q$ by a nondeterministic transition $u \stackrel{b}{\rightarrow} u'$ at a leaf u. Let $Q(u) = p$. In this case, consider $t \stackrel{a}{\Rightarrow} Q'$, the extension of Q by matching transitions $v \stackrel{b}{\Rightarrow} Q_i$ from all the $v \approx_{PLS} u$ that are leaves.
  - The case when C extends $s \stackrel{a}{\Rightarrow} Q$ by adding a one-step probabilistic transition $u \rightarrow Q$ at a leaf u uses the ideas from example 2. There are two cases depending on whether $Q([u]) = 0$ or not.
    If $Q([u]) = 0$, $u \rightarrow Q$ can be matched by computations from all the $v \approx_{PLS} u$.
    If $Q([u]) = r > 0$, consider the transition from u to Q' where: $Q'[v] = \frac{Q[v]}{1-r}$, if $u \notin [v]$ and $Q'([u]) = 0$. For any $v \approx_{PLS} u$, this computation reaches its leaves with weak label τ and assign probabilities in accordance with Q'. The required transition to Q from v is given by a linear combination (with coef£cient $1 - r$) of this computation with the computation consisting only of v (with coef£cient r).
    Consider $t \stackrel{a}{\Rightarrow} Q'$, the extension of Q by matching transitions $v \stackrel{b}{\Rightarrow} Q_i$ from all the $v \approx_{PLS} u$ that are leaves.

  In either case, the required transition from t is obtained by a linear combination $t \Rightarrow [\lambda \times Q' + (1 - \lambda) \times Q]$, where $\lambda = p/Q([u])$.

– (3) $\Rightarrow$ (1): This is immediate.

5 From measures to capacities

5.1 Background

In this section we £rst review the basic theory of capacities [11]. The original context that Choquet was interested in led him to impose several conditions that need not concern us here. We will present a simpli£ed treatment and omit proofs of any results available in the literature.

We begin by recalling that the basic example 1(a) shows that we lose the additivity property crucial to the de£nition of a measure. We omit a few of the details in the following de£nitions[8].

De£nition 12. *Let S be a set and let Σ be an algebra of subsets of S. A* ***capacity*** *on Σ is a non-negative real-valued set function $\nu : \Sigma \rightarrow \mathcal{R}$ such that*

[8] Like the exact de£nition of the family of sets on which a capacity is de£ned.

– $\nu(\emptyset) = 0$
– *if* $A \subseteq B$ *in* Σ *then* $\nu(A) \leq \nu(B)$,
– *if* $E_1 \subseteq E_2 \subseteq \ldots \subseteq E_n \subseteq \ldots$ *with* $\cup_i E_i = E$ *then* $\lim_{i\to\infty} \nu(E_i) = \nu(E)$,
– *if* $E_1 \supseteq E_2 \supseteq \ldots \supseteq E_n \supseteq \ldots$ *with* $\cap_i E_i = E$ *then* $\lim_{i\to\infty} \nu(E_i) = \nu(E)$.

If, in addition, it satis£es $\nu(A \cup B) \leq \nu(A) + \nu(B)$*, it is said to be* ***subadditive****.*

For measures the two continuity properties are consequences of countable additivity. If we have a family of measures μ_i de£ned on Σ we can get subadditive capacities as follows[9].

$$\overline{\nu}(A) := \sup_i \mu_i(A)$$

We establish the key properties of the functions $\nu(E) = P(s, a, E)$ showing that they are capacities.

Lemma 3. *Let* $s \in K$*,* $a \in$ `Act`*. Then the function* ν *on the* $\approx$*-closed subsets of* K *de£ned as above is a subadditive capacity as per de£nition 12.*

Proof. We sketch the proof. Recall that for any C that is an a-computation from s, we write P^C for the induced distribution on the leaves. We have:

– $E_1 \subseteq E_2 \Rightarrow P^C(s, a, E_1) \subseteq P^C(s, a, E_2)$.
– Let $\{E_i\}$ be an increasing sequence of $\approx$-closed sets of states. Then $P^C(s, a, \cup_i E_i) = \sup_i P^C(s, a, E_i)$.
– If $E_1 \cap E_2 = \emptyset$, $P^C(s, a, E_1 \cup E_2) = P^C(s, a, E_1) + P^C(s, a, E_2)$.

Thus, the £rst three properties and sub-additivity follow from basic properties of $\sup$.

The proof of the fourth property crucially uses compactness. First note that ν is the sup of a family of measures, say Q_i. Measures are down-continuous - considered as functions from the σ-algebra - as an easy consequence of σ-additivity. Since the space is compact the convergence is uniform and the limit of a uniformly convergent family is continuous.

6 pCTL*

We now examine the relation between our processes and a minor variant of pCTL* [17, 15], a standard modal logic used for expressing properties of probabilistic systems. We will largely elide formal de£nitions, instead focusing on explaining the key differences from the treatment of de Alfaro [15] for Markov decision processes (that lack τ and associate *unique* probability distributions with each label at a state).

[9] There are examples showing that not all capacities arise in this way.

The logic. There are two kinds of formulas — state formulas, denoted $\phi, \phi', \ldots$, and sequence formulas, denoted $\psi, \psi', \ldots$. These are generated by the following grammar:

$$\phi ::= \bot \mid \neg\phi \mid \phi \wedge \phi' \mid E\psi \mid P_{\bowtie q}\psi$$
$$\psi ::= a \mid \phi \mid \neg\psi \mid \psi \wedge \psi' \mid \bigcirc\phi \mid \Diamond\phi \mid \psi\mathcal{U}\psi'$$

In the above, $\bowtie$ is drawn from $\{=, \leq, \geq, <, >\}$ q is a rational in $[0, 1]$, and $a \in \mathtt{Act}$.

We ignore *atomic formulas* which are £rst order logic formulas over some £xed sets of variables, functions and predicate symbols. One can assume that bisimilar states satisfy the same atomic formulas.

Silent transitions and behaviors We handle the presence of silent transitions by considering a "saturation" of the set of paths from a state, in the spirit of closure under "stuttering".

We de£ne a *behavior* (adapting the de£nition of de Alfaro [15] to weak transition sequences) from a state s to be a sequence of states and labels $s = s_0, l_0, s_1, l_1, s_2, \ldots$ where $l_i \in \mathtt{Act}$ is the weak label for the transition from s_i to s_{i+1} and the probability of this transition is non-zero. Thus, we are permitting state repetition and skipping of intermediate states reached by τ transitions.

The non-probabilistic formulas For $a \in \mathtt{Act}$, the path formula a is true of behaviors $s_0, a, s_1, \ldots$ whose £rst weak label is a. Following standard de£nitions, the state formula $E\psi$ is true at a state s if there is a behavior $s = s_0, a, s_1, \ldots$ at s that satis£es the path formula ψ.

Policies and the probabilistic quanti£er A basic policy [15], say η, is a partial function from state sequences to states — thus a policy resolves the non-determinism completely. We also permit linear combinations of policies $\sum_i \lambda_i \eta_i$, where $\lambda_i > 0, \sum_i \lambda_i = 1$. Each policy η de£nes a computation $C(\eta, s)$ starting from each state s. We denote by $\mu_{\eta,s}$ the measure on the paths of $C(\eta, s)$ which is induced in a standard way[10].

The path formulas of pCTL* are interpreted on behaviours. We de£ne an operation $\mathcal{C}\updownarrow$ from paths to sets of behaviours by closing under repetition of states and under replacing subsequences of the form $s \xrightarrow{\tau} u \xrightarrow{\tau} t$ with $s \xrightarrow{\tau} t$. This is lifted to give a map from sets of paths to sets of behaviours. Now we de£ne $\mu_{\eta,s}$ on behaviours (using the same name as on paths) by $\mu_{\eta,s}(B) = \mu_{\eta,s}(\mathcal{C}\updownarrow^{-1}(B))$, where B is a set of behaviours.

Fix a policy η. A set of behaviors is measurable if the set of the corresponding paths in η is measurable. By a routine structural induction, we can show that the sets of behaviours that satisfy path formulas are measurable.

Following standard de£nitions, the state formula $P_{\bowtie q}\psi$ is true at a state s if for all policies η, the set B of behaviours that satisfy ψ satis£es $\mu_{\eta,s}(B) \bowtie q$.

Soundness of bisimulation

The key to the proof, as might be expected, is to show that the paths and computations out of bisimilar states "match" suf£ciently.

[10] We elide well-known measure-theoretic details in this paper.

First, we consider behaviors. The following lemma is a standard use of the co-inductive de£nition of bisimulation.

Lemma 4. *Let $s \approx t$. Then, for any behavior $s = s_0, l_0, s_1, l_1, s_2, \ldots$ from s, there is a behavior, $t = t_0, l'_0, t_1, l'_1, t_2, \ldots$, from t such that: $(\forall i)$ $[s_i \approx t_i]$ and $(\forall i)$ $[l_i = l'_i]$.*

Based on this we de£ne two behaviours to be equivalent if they satisfy the conclusions of Lemma 4.

Next, we move to policies and induced computations. For this, we follow the proof of Theorem 2 (in particular the implication $(2) \Rightarrow (3)$). This proof has already shown that given a computation C from a state s, and given t bisimilar to s, there is a computation C' from t that assigns the same probabilities to the leaves of C. We will now generalize this to all paths — given a computation C_η induced by a policy η from a state s, we show that for any bisimilar state t, there is a policy η' that assigns at least the probabilities assigned by η to all the paths in C_η. We use the equivalence of our definitions with those of Philippou, Lee and Sokolsky [1]. The £rst case of their de£nition permits the simulation of non-deterministic edges. The second case of their de£nition permits the simulation of probabilistic branches.

Lemma 5. *Let s, t be bisimilar states. Let η be a policy and let $C(\eta, s)$ be the induced η-computation from s. Then, there is a policy η' such that every path in $C(\eta, s)$ is equivalent to a behaviour in $C(\eta', t)$ with the same probability.*

Proof. It suf£ces to prove this for the case where η is a basic policy.

The proof is a routine induction. We write C_η for $C(\eta, s)$ and $C_{\eta'}$ for $C(\eta', t)$. C_η has countably many transitions. Consider any ordering o of these transitions such that a transition occurs after all the transitions leading upto it. We construct $C_{\eta'}$ by mimicking transitions in the order prescribed by o. Our induction hypothesis is that at the i'th stage: every path in the subtree induced by the £rst i transitions (as per o) is a behavior in $C^i_{\eta'}$ computation from t with the same probability.

Let the $i+1$'st transition be a transition at u. Let p be the probability of the path from t to u in C_η. Let V be the set of leaves in $C^i_{\eta'}$ such that:

- $v \approx u$
- The path from s to u in C_η is a behavior corresponding to the path from t to v in $C^i_{\eta'}$

The measure of V in $C^i_{\eta'}$, say q, is at least p by the induction hypothesis.

There are two cases based on the kind of the $(i+1)$st transition.

1. The $(i+1)$st transition is a nondeterministic transition $u \xrightarrow{b} u'$. This transition can be matched by computations from all elements of V: by de£nition these computations reach $[u']$ with probability 1 on weak label b.
2. The $(i+1)$st transition is a probabilistic transition $u \to Q$. There are two cases depending on whether $Q([u]) = 0$ or not.
   If $Q([u]) = 0$, this transition can be matched by computations from all elements of V: by theorem 2 these computations reach the leaves with weak label τ and assign probabilities in accordance with Q.

If $Q([u]) = r > 0$, consider the transition from u to Q' where: $Q'[v] = \frac{Q[v]}{1-r}$, if $u \notin [v]$ and $Q'([u]) = 0$. Pick any element $v \in V$. Since $v \approx u$, by theorem 2, this computation reaches the leaves with weak label τ and assign probabilities in accordance with Q'. The required transition to Q from v is given by a linear combination (with coef£cient $1 - r$) of this computation with the computation consisting only of v (with coef£cient r).

In either case, let $C^{i'}_{\eta'}$ be the extension of $C^{i}_{\eta'}$ by these matching transitions. $C^{i+1}_{\eta'}$ is got by a linear combination $t \Rightarrow [\lambda \times C^{i'}_{\eta'} + (1-\lambda) \times C^{i}_{\eta'}]$, where $\lambda = p/q$.

Lemmas 4 and 5 yield the desired theorem by a standard induction on the structure of formulas.

Theorem 3. *If $s \approx t$, then for all pCTL* state formulas ϕ, $s \models \phi$ iff $t \models \phi$.*

Proof. We sketch the case of $P_{\geq q}\psi$. Let s satisfy $P_{\geq q}\psi$. Every policy induces a set of computations from s. For every computation from s, using lemma 5, there is a computation from t that attributes a larger measure to the set of behaviors from t that satisfy ψ. Hence, t satis£es $P_{\geq q}\psi$.

Completeness

We proceed now to completeness. Here the fact that we have a capacity plays a key role, as we use the downward continuity property of capacities.

We identify $\mathcal{L}$, a sub-fragment of the state formulas of the pCTL* variant above, that suf£ces for completeness. These are generated by the following grammar:

$$\phi ::= \top \mid \phi_1 \wedge \phi_2 \mid \phi_1 \vee \phi_2 \mid \neg\phi \mid \langle a \rangle_{\geq q}\phi$$

where $a \in \mathtt{Act}$ (including τ), q is a rational and $\langle a \rangle_{\geq q}\phi$ is shorthand for $\neg P_{<1-q}[a \wedge \bigcirc\phi]$. Thus, a state s satis£es $\langle a \rangle_{\geq q}\phi$ iff there is a policy η such that the computation induced by η assigns probability greater than q to the states satisfying ϕ reachable on a weak a transition. More succinctly, s satis£es $\langle a \rangle_{\geq q}\phi$ if $P(s, a, \{t \mid t \text{ satis£es } \phi\}) \geq q$.

Theorem 4. *If two states satisfy the same formulas of $\mathcal{L}$, then they are bisimilar.*

Proof. Let R be the equivalence relation de£ned by the formulas of $\mathcal{L}$. Let s and t be two R-related states. We need to prove that for every R-closed set X, $P(s, \{a\}, X) = P(t, \{a\}, X)$, where $a \neq \tau$. By using formulas of the form $\langle a \rangle_{\geq q}\phi$, we obtain the required equality for sets of states X that are denotations of formulas, i.e. $X = \{s' \mid s' \text{ satis£es } \phi\}$, $\phi \in \mathcal{L}$.

Since the state space is countable every R-closed set is a countable union of equivalence classes. Every equivalence class is described by countably many formulas and - since we have negation - can be described as the intersection of countably many sets of the form $\{s | s \text{ satis£es } \phi\}$. Thus every R-closed set, say Y, is of the form

$$Y = \cup_{i=1}^{\infty} \cap_{j=1}^{\infty} X_{ij}$$

where the X_{ij} are the denotations of formulas.

We de£ne

$$Y_i := \cap_{j=1}^{\infty}[\cup_{k=1}^{i} X_{kj}].$$

Note that Y_i forms an increasing family in the subset ordering. Furthermore $\cup_{i=1}^{\infty} Y_i = Y$ by distributivity. Now, for each i, the sets $Z_i^{(l)} := \cap_{j=1}^{l} \cup_{k=1}^{i} X_{kj}$ are a decreasing family as l increases and they are the denotations of formulas, since there is conjunction and disjunction in the logic. Thus the two capacities will agree on each $Z_i^{(l)}$ and by up continuity they will agree on Y_i and thus, by down continuity, they agree on Y.

The proof for $P(s, \varepsilon, X) = P(t, \varepsilon, X)$ is similar except for the use of the formulas $\langle\tau\rangle\phi$ and is omitted.

7 Conclusions

The main thrust of the present paper has been to elucidate the interaction between probability and nondeterminism. The de£nition of weak bisimulation that we have used generalizes the elegant treatment of Philippou, Lee and Sokolsky from £nite state to countable systems. We have emphasized two features of their de£nition that were left implicit by them, namely the loss of additivity and the need for considering linear structure when matching weak transitions. The main new result of our analysis is that weak bisimulation is sound and complete for (a minor variant of) pCTL*.

It is worth taking a retrospective view of some of the mathematical ideas in the proofs. The basic loss that we have had to struggle with is the loss of σ-additivity. The heart of any completeness proof of this type is arguing that equality of the transition probabilities to sets of states de£ned by the logic forces equality of all the transition probabilities. Such an argument rests on theorems that guarantee equality of measures given equality on a suitable generating set for the σ-£eld. These uniqueness theorems heavily rely on σ-additivity. Thus we were led to consider what structure we do have given that we do not have a probability measure. The fact that we have capacities and in particular that capacities satisfy strong continuity properties (both upward and downward) turns out to be strong enough to rescue the uniqueness theorems that we need. What remains to argue is that we really have the property of a capacity. Here the compactness property turns out to be crucial.

In closely related work [18] we have shown that one can develop a metric for weak bisimulation analogous to our previous treatment of metrics for strong bisimulation [19]. In this work we heavily use linear programming and duality.

The present treatment is for discrete systems, we are considering two new directions: continuous state spaces and continuous time. We have preliminary results on continuous time, namely we have shown completeness for continuous stochastic logic [20].

References

1. Philippou, A., Lee, I., Sokolsky, O.: Weal bisimulation for probabilistic processes. In Palamidessi, C., ed.: Proceedings of CONCUR 2000. Number 1877 in Lecture Notes In Computer Science, Springer-Verlag (2000) 334–349

2. Larsen, K.G., Skou, A.: Bisimulation through probablistic testing. Information and Computation **94** (1991) 1–28
3. Kemeny, J.G., Snell, J.L.: Finite Markov Chains. Van Nostrand (1960)
4. Blute, R., Desharnais, J., Edalat, A., Panangaden, P.: Bisimulation for labelled Markov processes. In: Proceedings of the Twelfth IEEE Symposium On Logic In Computer Science, Warsaw, Poland. (1997)
5. Desharnais, J., Edalat, A., Panangaden, P.: A logical characterization of bisimulation for labeled Markov processes. In: proceedings of the 13th IEEE Symposium On Logic In Computer Science, Indianapolis, IEEE Press (1998) 478–489
6. Desharnais, J., Edalat, A., Panangaden, P.: Bisimulation for labeled Markov processes. Information and Computation (2002)
7. Baier, C., Hermanns, H.: Weak bisimulation for fully probabilistic processes. In: Proceedings of the 1997 International Conference on Computer Aided Veri£cation. Number 1254 in Lecture Notes In Computer Science, Springer-Verlag (1997)
8. Hansson, H.A.: Time and Probability in Formal Design of Distributed Systems. Volume 1 of Real-time Safety-critical Systems. Elseiver (1994)
9. Segala, R., Lynch, N.: Probabilistic simulations for probabilistic processes. In Jonsson, B., Parrow, J., eds.: Proceedings of CONCUR94. Number 836 in Lecture Notes In Computer Science, Springer-Verlag (1994) 481–496
10. Segala, R.: Modeling and Veri£cation of Randomized Distributed Real-Time Systems. PhD thesis, MIT, Dept. of Electrical Engineering and Computer Science (1995) Also appears as technical report MIT/LCS/TR-676.
11. Choquet, G.: Theory of capacities. Ann. Inst. Fourier (Grenoble) **5** (1953) 131–295
12. Dellacherie, C.: Capacités et Processus Stochastiques. Springer-Verlag (1972)
13. Meyer, P.A.: Probability and Potentials. Blaisdell (1966)
14. Schmeidler, D.: Subjective probability without additivity. Technical report, Foerder Institute of Economic Research (1984)
15. de Alfaro, L.: Formal Veri£cation of Probabilistic Systems. PhD thesis, Stanford University (1997) Technical Report STAN-CS-TR-98-1601.
16. Bandini, E., Segala, R.: Axiomatizations for probabilistic bisimulation. In: Proceedings of the 28th International Colloquium on Automata, Languages and Programming. Number 2076 in Lecture Notes In Computer Science, Springer-Verlag (2001) 370–381
17. Aziz, A., Singhal, V., F.Balarin, Brayton, R.K., Sangiovanni-Vincentelli, A.L.: It usually works:the temporal logic of stochastic systems. In: Proceedings of the Conference on Computer-Aided Veri£cation. (1995)
18. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: The metric analogue of weak bisimulation for labelled Markov processes. In: Proceedings of the Seventeenth Annual IEEE Symposium On Logic In Computer Science. (2002)
19. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Metrics for labeled Markov processes. In: Proceedings of CONCUR99. Lecture Notes in Computer Science, Springer-Verlag (1999)
20. Desharnais, J., Panangaden, P.: Continuous stochastic logic characterizes bisimulation for continuous-time markov processes. Available from `http://www-acaps.cs.mcgill.ca/~prakash/pubs.html` (2001)