On the Relationship between Bisimulation and Trace Equivalence in an Approximate Probabilistic Context

Gaoang Bian[1][2] and Alessandro Abate[2]

[1] Google Inc, United States
[2] Dept. of Computer Science, University of Oxford, UK

Abstract. This work introduces a notion of approximate probabilistic trace equivalence for labelled Markov chains, and relates this new concept to the known notion of approximate probabilistic bisimulation. In particular this work shows that the latter notion induces a tight upper bound on the approximation between finite-horizon traces, as expressed by a total variation distance. As such, this work extends corresponding results for exact notions and analogous results for non-probabilistic models. This bound can be employed to relate the closeness in satisfaction probabilities over bounded linear-time properties, and allows for probabilistic model checking of concrete models via abstractions. The contribution focuses on both finite-state and uncountable-state labelled Markov chains, and claims two main applications: firstly, it allows an upper bound on the trace distance to be decided for finite state systems; secondly, it can be used to synthesise discrete approximations to continuous-state models with arbitrary precision.

1 Introduction

Often in formal verification one is interested in approximations of concrete models. Models are often built from experimental data that are themselves approximate, and taking approximations can reduce the size and complexity of the state space. Markov models in particular can be defined either syntactically as a transition structure (with states and matrices), or semantically as a random process whose trajectory satisfy the Markov property. Each representation gives rise to its own notions of approximation [1]: "the transition matrices have similar numbers and/or structure" vs "the trajectories have similar probability distributions", respectively. While the syntactic representation is used for computations and model checking with concrete numbers, often one is interested in results in terms of the semantics, e.g. "what is the probability of reaching a failure state within 100 steps". This gives practical value to studying how approximations in terms of transition matrices translate into approximations in terms of traces of the random process.

In this paper we build on the notion of ε-approximate probabilistic bisimulation, introduced in [13] as a natural extension to exact probabilistic bisimulation [12]. There, the notion of ε-approximate probabilistic bisimulation (or just
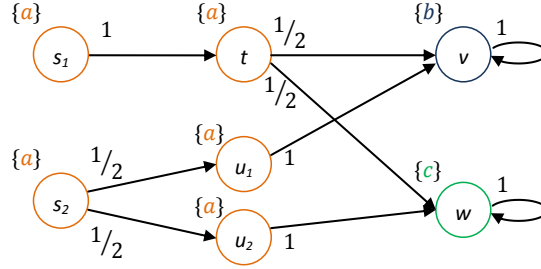
Fig. 1. Branching vs Linear Time Behaviour. In the Labelled Markov Chain below (cf. Section 2 for the LMC model), the states s_1, s_2 both emit traces $\langle \{a\}, \{a\}, \{b\} \rangle$ and $\langle \{a\}, \{a\}, \{c\} \rangle$ with probability 0.5 each, and hence s_1, s_2 have the same linear time behaviour. However, s_1, s_2 have different branching behaviour, since exclusively s_1 satisfies the PCTL formula $P_{=1}[\, X\, P_{=0.5}[\, X\, b\,]\,]$. Conversely, only s_2 satisfies $P_{=0.5}[\, X\, P_{=1}[\, X\, b\,]\,]$.

ε-bisimulation) is defined in terms of the transition structure, and given ε the maximal ε-bisimulation relation can be computed for finite state systems with n states in $\mathcal{O}(n^7)$ time [13].

It is on the other hand of interest to explore what ε-bisimulation means in terms of trajectories. While ε-bisimulation does have characterizations (on countable state spaces) in terms of logics and games [12], this logic is branching in nature, and does not directly relate to the trajectory of the model as it leaks information about the state space (similarly to the difference between CTL and LTL), as illustrated in Figure 1.

In this paper, we investigate what ε-approximate probabilistic bisimulation means in terms of trajectories. We will prove that for Labelled Markov Chains (over potentially uncountable state spaces), ε-bisimulation between two states places the tight upper bound of $1-(1-\varepsilon)^k$ (which is $\leq k\varepsilon$) on the total variation [18] between the distributions of length $k+1$ traces starting from those states, for all $k \in \mathbb{N}$. We will formulate these bounds by introducing the notion of $f(k)$-trace equivalence. As such, we extend the well known result that bisimulation implies trace equivalence in non-probabilistic systems to the context of approximate and probabilistic models (the exact probabilistic case having been considered in [7]).

One direct repercussion of our result is that it provides a method to efficiently bound the total variation of length k traces from two finite-state LMCs (or two states in an LMC), since the aforementioned result in [13] can be used to decide or to compute ε-bisimulation between two states in polynomial time. We will also apply our results to the quantitative verification of continuous-state Markov models [2, 3, 16], improving on the current class of properties approximated and the corresponding approximation errors.

Related Work. Literature on approximations of (finite-state) Markov models can be distinguished into two main branches: one focusing on one-step similarity,

the other dealing with trace distances. One-step similarity can be studied via the notion of probabilistic bisimulation, introduced in the context of finite-state models by [20], and related to lumpability in [23]. [13] discusses a notion of approximate bisimulation, related to quasi-lumpability conditions in [6, 22]. From the perspective of process algebra, [17] studies operators on probabilistic transition systems that preserve the approximate bisimulation distance. The work in [13, 21] is seminal in introducing notions of (exact) probabilistic simulation, much extended in subsequent literature.

On the other hand, there are a few papers studying the total variation distance over traces. [9] presents an algorithm for approximating the total variation of infinite traces of labelled Markov systems and prove the problem of deciding whether it exceeds a given threshold to be NP-hard. [8] shows that the undiscounted bisimilarity pseudometric is a (non-tight) upper bound on the total variation of infinite traces (like ε-bisimulation, the bisimilarity pseudometric is defined on the syntax of the model and there are efficient algorithms for computing it [8, 10, 24]). The contribution in this paper, focusing on finite traces rather than infinite traces, is that the total variation of finite traces is much less conservative, and moreover allows manipulating models under specific error bounds on length k traces, as we will show in Section 5.

[14] studies notions based on the total variation of finite and infinite traces: employing a different notion of ε-bisimulation than ours, it proves error bounds on trace distances, which however depend on additional properties of the structure of the transition kernel (as shown in Section 7 the error bound on reachability probabilities could go to 1 in two steps, for any $\varepsilon > 0$). Finite abstractions of continuous-state Markov models can be synthesised by notions that are variations of the ε-bisimulation in this work [1, 3, 16]. Tangential to our work, [11] shows that the total variation of finite traces can be statically estimated via repeated observations. [26] investigates ways of compressing Hidden Markov Models by searching for a smaller model that minimises the total variation of length-k traces of the two models.

Structure of this article. In Section 2, we introduce the reference model (labelled Markov chains – LMC – over general state spaces) and provide a definition of ε-bisimulation for LMCs. In Section 3, we introduce the notion of approximate probabilistic trace equivalence (and the derived notion of probabilistic trace distance), and discuss how it relates to bounded linear time properties, and to the notion of distinguishability. In Section 4, we present the main result: we will derive a tight upper bound on the probabilistic trace distance between ε-bisimilar states. In Section 5, we show how these results can be used to approximately model check continuous state systems, and Section 6 discusses a case study. In Section 7, we discuss an alternative notion of approximate probabilistic bisimulation that appears in literature and show that it cannot be used to effectively bound probabilistic trace distance.

In this work we offer sketches of the proof of some theorems, and omit the proof of other results: the complete proofs, as well as the details on the implementation of the Case Study, can be found in the Appendix of [4].

2   Preliminaries

2.1   Labelled Markov Chains

We will work with discrete-time Labelled Markov Chains (LMCs) over general state spaces. Known definitions of countable- or finite-state LMCs represent special instances of the general models we introduce next.

Definition 1 (LMC syntax). *A Labelled Markov Chain (LMC) is a structure* $\mathscr{M} = (S, \Sigma, \tau, L)$ *where:*

- *S is a (potentially uncountable) set of states.*
- *$\Sigma \subseteq \mathcal{P}(S)$ is a Σ-algebra over S representing the set of measurable subsets of S.*
- *$\tau : S \times \Sigma \to [0,1]$ is a transition kernel. That is, for all $s \in S$, $\tau(s, \cdot)$ is a probability measure on the measure space (S, Σ), and for all $A \in \Sigma$ we require $\tau(\cdot, A)$ to be Σ-measurable.*
- *$L : S \to \mathcal{O}$ labels each state $s \in S$ with a subset of atomic propositions from* AP*, where $\mathcal{O} = 2^{\mathrm{AP}}$. L is required to be Σ-measurable, and we will assume* AP *to be finite.*

$L(s)$ captures all the observable information at state $s \in S$: this drives our notion relating pairs of states, and we characterise properties over the codomain of this function.

Definition 2 (LMC semantics). *Let $\mathscr{M} = (S, \Sigma, \tau, L)$ be a LMC. Given an initial distribution p_0 over S, the state of $\mathscr{M}$ at time $t \in \mathbb{N}$ is a random variable $\mathscr{M}_t^{p_0}$ over S, such that*

$$\mathbb{P}\left[\mathscr{M}_0^{p_0} \in A_0\right] = p_0(A_0),$$

$$\mathbb{P}\left[\mathscr{M}_0^{p_0} \in A_0, \cdots, \mathscr{M}_k^{p_0} \in A_k\right] = \int_{y_0 \in A_0} p_0(\mathrm{d}y_0)$$

$$\cdot \int_{y_1 \in A_1} \tau(y_0, \mathrm{d}y_1) \cdots \int_{y_{k-1} \in A_{k-1}} \tau(y_{k-2}, \mathrm{d}y_{k-1}) \cdot \tau(y_{k-1}, A_k),$$

for all $k \in \mathbb{N} \backslash \{0\}$, $A_k \in \Sigma$, where of course $\tau(y_{k-1}, A_k) = \int_{y_k \in A_k} \tau(y_{k-1}, \mathrm{d}y_k)$.

Models in related work. A body of related literature works with labelled MDPs, which are more general models allowing a non-deterministic choice $u \in \mathcal{U}$ (for some finite $\mathcal{U}$) of the transition kernel τ_u at each step. This choice is made by a "policy" that probabilistically selects u based on past observations of the process. Whilst we will ignore non-determinism and work with LMCs for simplicity, our results can be adapted to labelled MDPs by quantifying over all policies, or over all choices u for properties like ε-bisimulation, in order to remove the non-determinism. The seminal work on bisimulation and ε-bisimulation dealt with models known as LMPs [12]. LMPs allow for non-determinism (like labelled MDPs) but their states are unlabelled and at each step they have a probability of halting. For the study of bisimulation in this work, LMPs can be considered as a simplification of labelled MDPs to the case $\mathcal{O} = \{\emptyset, \{\text{halted}\}\}$.

2.2 Exact and Approximate Probabilistic Bisimulations

The notion of approximate probabilistic bisimulation (in this work just ε-bisimulation) is a structural notion of closeness, based on the stronger notion of exact probabilistic bisimulation [12]. We discuss both next. Considering a binary relation R over set X, we say that a subset $\tilde{S} \subseteq X$ is R-closed if $\tilde{S}$ contains its own image under R. That is, if $R(\tilde{S}) \coloneqq \{y \in X \mid x \in \tilde{S},\ xRy\} \subseteq \tilde{S}$.

Definition 3 (Exact probabilistic bisimulation). *Let $\mathscr{M} = (S, \Sigma, \tau, L)$ be a LMC. For $\varepsilon \in [0,1]$, an equivalence relation $R \subseteq S \times S$ over the state space is an exact probabilistic bisimulation relation if*

$$\forall (s_1, s_2) \in R, \qquad \text{we have that } L(s_1) = L(s_2),$$
$$\forall (s_1, s_2) \in R,\ \forall \tilde{T} \in \Sigma \text{ s.t. } \tilde{T} \text{ is } R\text{-closed}, \qquad \text{we have that } \tau(s_1, \tilde{T}) = \tau(s_2, \tilde{T}).$$

A pair of states $s_1, s_2 \in S$ are said to be (exactly probabilistically) bisimilar if there exists an exact probabilistic bisimulation relation R such that $s_1 R s_2$.

Note that since R is an equivalence relation, R-closed sets are exactly the unions of whole equivalence classes.

Next, we adapt the notion of ε-bisimulation (as discussed in [13] for LMPs over countable state spaces) to LMCs over general spaces.

Definition 4 (ε-bisimulation). *Let $\mathscr{M} = (S, \Sigma, \tau, L)$ be a LMC. For $\varepsilon \in [0,1]$, a symmetric binary relation $R_\varepsilon \subseteq S \times S$ over the state space is an ε-approximate probabilistic bisimulation relation (or just ε-bisimulation relation) if*

$$\forall T \in \Sigma, \qquad \text{we have } R_\varepsilon(T) \in \Sigma, \tag{1}$$
$$\forall (s_1, s_2) \in R_\varepsilon, \qquad \text{we have } L(s_1) = L(s_2), \tag{2}$$
$$\forall (s_1, s_2) \in R_\varepsilon,\ \forall T \in \Sigma, \qquad \text{we have } \tau(s_2, R_\varepsilon(T)) \geq \tau(s_1, T) - \varepsilon. \tag{3}$$

Two states $s_1, s_2 \in S$ are said to be ε-bisimilar if there exists an ε-bisimulation relation R_ε such that $s_1 R_\varepsilon s_2$.

The condition raised in (3) could be understood intuitively as "for any move that s_1 can take (say, into set T), s_2 can match it with higher likelihood over the corresponding set $R_\varepsilon(T)$, up to ε tolerance." Notice that (1) is not a necessary requirement for countable state models, but for uncountable state models it is needed to ensure that $R_\varepsilon(T)$ is measurable and $\tau(s_2, R_\varepsilon(T))$ is defined in (3).

[13] showed that in countable state spaces, 0-approximate probabilistic bisimulation corresponds to exact probabilistic bisimulation. On uncountable state spaces, not every exact probabilistic bisimulation relation is a 0-bisimulation relation because of the additional measurably requirement, but we still have that 0-bisimulation implies exact probabilistic bisimulation.

Theorem 1. *Let $\mathscr{M} = (S, \Sigma, \tau, L)$ be a LMC, and let $s_1, s_2 \in S$. If s_1, s_2 are 0-bisimilar, then they are (exactly, probabilistically) bisimilar.*

Although above s_1, s_2 are required to belong to the state space of a given LMC, the notions of exact- and ε-bisimulation can be extended to hold over pairs of LMCs by combining their state spaces, as follows.

Definition 5 (ε-bisimulation of pairs of LMCs). *Consider two LMCs $\mathscr{M}_1 = (S_1, \Sigma_1, \tau_1, L_1)$ and $\mathscr{M}_2 = (S_2, \Sigma_2, \tau_2, L_2)$. Without loss of generality, assume that their state spaces S_1, S_2 are disjoint. The direct sum $\mathscr{M}_1 \oplus \mathscr{M}_2$ of $\mathscr{M}_1$ and $\mathscr{M}_2$ is the LMC formed by combining the state spaces of $\mathscr{M}_1$ and $\mathscr{M}_2$. Formally, $\mathscr{M}_1 \oplus \mathscr{M}_2 = (S_1 \uplus S_2, \sigma(\Sigma_1 \times \Sigma_2), \tau_1 \oplus \tau_2, L_1 \uplus L_2)$, where:*

- *$S_1 \uplus S_2$ is the union of S_1 and S_2 where we have assumed wlog (by relabelling if necessary) that S_1, S_2 are disjoint;*
- *$\sigma(\Sigma_1 \times \Sigma_2)$ is the smallest σ-algebra containing $\Sigma_1 \times \Sigma_2$;*
- *$\tau_1 \oplus \tau_2(s, T) := \begin{cases} \tau_1(s, T \cap S_1) & \text{if } s \in S_1 \\ \tau_2(s, T \cap S_2) & \text{if } s \in S_2 \end{cases}$ for $s \in S_1 \uplus S_2$, $T \in \sigma(\Sigma_1 \times \Sigma_2)$;*

- *$L_1 \uplus L_2(s) := \begin{cases} L_1(s) & \text{if } s \in S_1 \\ L_2(s) & \text{if } s \in S_2 \end{cases}$ for $s \in S_1 \uplus S_2$.*

Let $s_1 \in S_1$, $s_2 \in S_2$. We say that s_1, s_2 are ε-bisimilar iff s_1, s_2 are ε-bisimilar as states in the direct sum LMC $\mathscr{M}_1 \oplus \mathscr{M}_2$.

Other Notions of ε-Bisimulation in Literature. There is an alternative, more direct, extension of exact probabilistic bisimulation in literature [1, 3, 14], which simply requires $|\tau(s_1, \tilde{T}) - \tau(s_2, \tilde{T})| \leq \varepsilon$ instead of the conditions in Def. 4. However, this requirement alone is too weak to guarantee properties that we later discuss (cf. Section 7).

3   Approximate Probabilistic Trace Equivalence for LMCs

In this section we introduce the concept of approximate probabilistic trace equivalence (or just $f(k)$-trace equivalence) to represent closeness of observable linear time behaviour. Based on the likelihood over traces of a given LMC, this notion depends on its operational semantics (cf. Definition 2), rather than on the structure of its transition kernel (as in the case of approximate bisimulation). The notion can alternatively be thought of inducing a distance among traces, as elaborated below.

Definition 6 (Trace likelihood). *Let $\mathscr{M} = (S, \Sigma, \tau, L)$ be an LMC, $s_0 \in S$, and $k \in \mathbb{N}$. Let TRACE denote a set of traces (each of length $k + 1$), taking values in time over 2^{AP}, so that TRACE $\subseteq \mathcal{O}^{k+1}$. Denote with $P_k(s_0, \mathrm{TRACE})$ the probability that the LMC $\mathscr{M}$, given an initial state s_0, generates any of the runs $\langle \alpha_0, \cdots, \alpha_k \rangle \in \mathrm{TRACE}$, namely*

$$P_k(s_0, \mathrm{TRACE}_k) = \sum_{\substack{\langle \alpha_0, \cdots, \alpha_k \rangle \\ \in \mathrm{TRACE}_k}} \mathbb{P}\big[\mathscr{M}_0^{s_0} \in L^{-1}(\{\alpha_0\}), \cdots, \mathscr{M}_k^{s_0} \in L^{-1}(\{\alpha_k\})\big],$$

where $\mathscr{M}_t^{s_0}$ is the state of $\mathscr{M}$ at step t, with a degenerate initial distribution p_0 that is concentrated on point s_0 (cf. Definition 2).

As intuitive, we consider traces of length $k+1$ (rather than of length k) because a length $k+1$ trace is produced by one initial state and precisely k transitions. Notice that the set of sequences of states generating TRACE is measurable, being defined via a measurable map L over a finite set of traces.

Definition 7 (Total variation [15]). *Let $(Z, \mathcal{G})$ be a measure space where $\mathcal{G}$ is a σ-algebra over Z, and let μ_1, μ_2 be probability measures over $(Z, \mathcal{G})$. The total variation between μ_1, μ_2 is $d_{\mathrm{TV}}(\mu_1, \mu_2) \coloneqq \sup_{A \in \mathcal{G}} |\mu_1(A) - \mu_2(A)|$.*

Definition 8 ($f(k)$-trace equivalence). *Let $\mathscr{M} = (S, \Sigma, \tau, L)$ be a LMC. For a non-decreasing function $f : \mathbb{N} \to [0,1]$, we say that states $s_1, s_2 \in S$ are $f(k)$-approximate probabilistic trace equivalent if for all $k \in \mathbb{N}$,*

$$d_{\mathrm{TV}}\big(P_k(s_1, \cdot), P_k(s_2, \cdot)\big) \leq f(k),$$

or alternatively if over $\mathrm{TRACE} \subseteq \mathcal{O}^{k+1}$,

$$|P_k(s_1, \mathrm{TRACE}) - P_k(s_2, \mathrm{TRACE})| \leq f(k).$$

The condition on monotonicity follows from the requirement on the total variation distance, which is defined over a product output space and necessarily accumulates over time. The notion of $f(k)$-trace equivalence can be used to relate states from two different LMCs, much in the same way as ε-bisimulation.

One can introduce the notion of *probabilistic trace distance* between pairs of states s_1, s_2 as

$$\min\{f(k) \geq 0 \mid s_1 \text{ is } f(k)\text{-trace equivalent to } s_2\} = d_{\mathrm{TV}}\big(P_k(s_1, \cdot), P_k(s_2, \cdot)\big).$$

Notice that the RHS is clearly a pseudometric. We discuss the development of tight bounds on the probabilistic trace distance in Section 4.

3.1 Interpretation and Application of ε-Trace Equivalence

The notion of ε-trace equivalence subsumes closeness of finite-time traces, and can be interpreted in two different ways. Firstly, ε-trace equivalence leads to closeness of satisfaction probabilities over bounded-horizon linear time properties, e.g. bounded LTL formulae, as follows.

Theorem 2. *Let $\mathscr{M} = (S, \Sigma, \tau, L)$ be an LMC, and let $s_1, s_2 \in S$ be ε-trace equivalent. Let ψ be any bounded LTL property over a k-step time horizon, defined within the LTL fragment $\phi = \phi_1 \mathrm{U}^{\leq t} \phi_2 \mid a \mid \phi_1 \wedge \phi_2 \mid \phi_1 \vee \phi_2 \mid \neg \phi_1$ for $t \leq k$. Then,*

$$\Big| \mathbb{P}\left[s_1 \models \psi\right] - \mathbb{P}\left[s_2 \models \psi\right] \Big| \leq f(k),$$

where $P\left[s \models \psi\right]$ is the probability that starting from state s, the LMC satisfies property ψ.

Proof. Formula ψ is satisfied by a specific set of length $k+1$ traces.   □

Alternatively, via its connection to the notion of total variation, ε-trace *distance* leads to the notion of distinguishability of the underlying LMC, namely the ability (of an agent external to the model) to distinguish a model by observing its traces.

Theorem 3. *Let s_1, s_2 be two states of an LMC. Suppose one of them is selected by a secret fair coin toss. An external agent guesses which one has been selected by observing a trace of length $k+1$ emitted from the unknown state. Then, an optimal agent guesses correctly with probability*

$$\frac{1}{2} + \frac{1}{2}f(k),$$

with $f(k) = d_{\mathrm{TV}}\big(P_k(s_1, \cdot), P_k(s_2, \cdot)\big)$ being the probabilistic trace distance.

4 ε-Probabilistic Bisimulation induces Approximate Probabilistic Trace Equivalence

In this section we present the main result: we show that ε-bisimulation induces a tight upper bound on the probabilistic trace distance, quantifiable as $(1 - (1-\varepsilon)^k)$. This translates to a guarantee on all the properties implied by ε-trace equivalence, such as closeness of satisfaction probabilities for bounded linear time properties. In addition, since for finite state LMPs the maximal ε-bisimulation relation can be computed in $\mathcal{O}(|S|^7)$ time [13], this result allows to establish an upper bound on the probabilistic trace distance with the same time complexity.
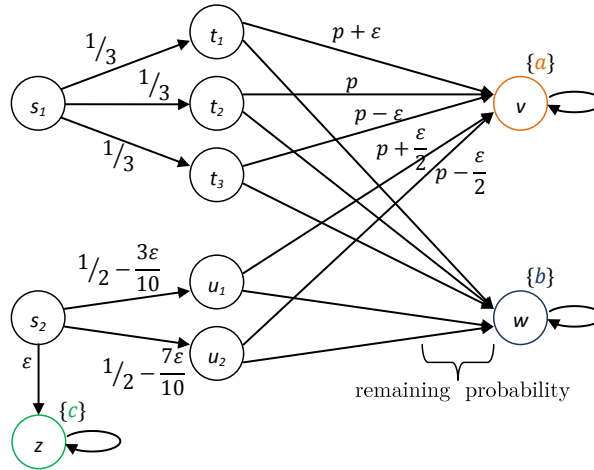

Fig. 2. LMC for the proof of Theorem 4.

Theorem 4 (ε-bisimulation implies $(1-(1-\varepsilon)^k)$-trace equivalence). *Let $\mathscr{M} = (S, \Sigma, \tau, L)$ be a LMC. If $s_1, s_2 \in S$ are ε-bisimilar, then s_1, s_2 are $(1-(1-\varepsilon)^k)$-trace equivalent.*

Proof (Sketch). The full proof, developed for LMCs over uncountable state spaces, can be found in Appendix C of [4]. Here we offer a sketch of proof, employing the finite-state LMP in Figure 2 as an illustrating example (where for simplicity we have omitted the labels for internal states, which can as well be labelled with $\emptyset \in \mathcal{O}$).

The maximal (i.e. coarsest) ε-bisimulation relation R_ε is obtained by pairs of states within the sets

$$\{t_1, t_2, u_1\}, \{t_2, t_3, u_2\}, \{s_1, s_2\}, \{v\}, \{w\}, \{z\}.$$

We would like to prove that these ε-bisimilar states are also ε-trace equivalent. In the full proof, we will show this by induction on the length of the trace, for all ε-bisimilar states at the same time. In this sketch proof, we aim to illustrate the induction step by showing how to bound

$$\left| P_3\big(s_1, \Diamond^{\leq 3}a\big) - P_3\big(s_2, \Diamond^{\leq 3}a\big) \right|,$$

where $\Diamond^{\leq k}a$ is the set of traces of length $k+1$, which reach a state labelled with a (which in this case is just state v). The idea is to match each of the outgoing transitions from s_1 to an outgoing transition from s_2 and to an ε-bisimilar state. Specifically, we explicitly write

$$\begin{aligned}
P_3\big(s_1, \Diamond^{\leq 3}a\big) = {} & \frac{1}{3}P_2\big(t_1, \Diamond^{\leq 2}a\big) + \frac{1}{6}P_2\big(t_2, \Diamond^{\leq 2}a\big) \\
& + \frac{1}{6}P_2\big(t_2, \Diamond^{\leq 2}a\big) + \frac{1}{3}P_2\big(t_3, \Diamond^{\leq 2}a\big),
\end{aligned} \tag{4}$$

and respectively

$$\begin{aligned}
P_3\big(s_2, \Diamond^{\leq 3}a\big) = {} & (\frac{1}{3} - \frac{3\varepsilon}{10}) \cdot P_2\big(u_1, \Diamond^{\leq 2}a\big) + \frac{1}{6}P_2\big(u_1, \Diamond^{\leq 2}a\big) \\
& + \frac{1}{6}P_2\big(u_2, \Diamond^{\leq 2}a\big) + (\frac{1}{3} - \frac{7\varepsilon}{10}) \cdot P_2\big(u_2, \Diamond^{\leq 2}a\big).
\end{aligned} \tag{5}$$

We then match-off the terms in the expansions for $P_3\big(s_1, \Diamond^{\leq 3}a\big)$ and $P_3\big(s_2, \Diamond^{\leq 3}a\big)$, one term at a time. We use the induction hypothesis to argue that the probabilities in the matched terms are $(1-(1-\varepsilon)^k)$-close to each other (here $k = 1$), since they concern ε-bisimilar states. That is,

$$\left| P_2\big(t_1, \Diamond^{\leq 2}a\big) - P_2\big(u_1, \Diamond^{\leq 2}a\big) \right| \leq 1 - (1-\varepsilon)^k$$
$$\left| P_2\big(t_2, \Diamond^{\leq 2}a\big) - P_2\big(u_1, \Diamond^{\leq 2}a\big) \right| \leq 1 - (1-\varepsilon)^k$$
$$\dots$$

The total amount of difference between the matching coefficients is no more than ε. It can be shown (Lemma ?? in Appendix C of [4]) that these conditions guarantee the required bound on $\left| P_3\big(s_1, \Diamond^{\leq 3}a\big) - P_3\big(s_2, \Diamond^{\leq 3}a\big) \right|$.

The main difficulty is choosing a suitable decomposition of $P_3(s_1, \Diamond^{\leq 3}a)$ and $P_3(s_2, \Diamond^{\leq 3}a)$. This is non-trivial since in (4), the $1/3$ probability of transitioning into t_2 had to be broken up into two terms with $1/6$ probability each. However, we can tackle this issue using an extension of Hall's Matching Theorem [5] (cf. Appendix C of [4]). It is then relatively straight forward to adapt this proof to LMCs on uncountable state spaces by converging on integrals with simple functions. □

We now show that the expression for the induced bound on probabilistic trace distance proved in Theorem 4, namely $(1 - (1 - \varepsilon)^k)$, is tight, in the sense that for any k and ε, the bound can be attained by some pair of ε-bisimilar states in some LMC. In other words, it is not possible to provide bounds on the induced approximation level for traces, that are smaller than the expression discussed above and that are valid in general.

Theorem 5. *For any $\varepsilon \geq 0$, there exists a LMC $\mathscr{M} = (S, \Sigma, \tau, L)$ and states $s_1, s_2 \in S$ such that s_1, s_2 are ε-bisimilar, and for all $k \in \mathbb{N}$ there exists a set TRACE of length $k + 1$ traces s.t. $|P_k(s_1, \mathrm{TRACE}) - P_k(s_2, \mathrm{TRACE})| = 1 - (1 - \varepsilon)^k$.*

Proof. Select $\varepsilon \geq 0$ and consider the following LMC:
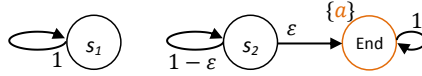
Here s_1, s_2 are ε-bisimilar, and for all $k \in \mathbb{N}$, $P_k(s_1, \Diamond^{\leq k}a) = 0$, whereas $P_k(s_2, \Diamond^{\leq k}a) = \sum_{i=1}^{k}(1 - \varepsilon)^{i-1}\varepsilon = 1 - (1 - \varepsilon)^k$. □

The result in Theorem 4 can be viewed as an extension of the known fact that bisimulation implies trace equivalence in non-probabilistic transition systems. Similar to the deterministic case, the converse of Theorem 4 does not hold.

Theorem 6. *$(1 - (1 - \varepsilon)^k)$-trace equivalence does not imply ε-bisimulation.*

Proof. In Figure 1, states s_1, s_2 are not ε-bisimilar for any $\varepsilon < 1/2$, yet their probabilistic trace distance is equal to 0. □

This example shows that ε-bisimulation cannot be used to effectively estimate the probabilistic trace distance between individual states. In particular, while the $(1 - (1 - \varepsilon)^k)$-bound on probabilistic trace distance discussed above is tight as a uniform bound, it is not tight for individual pairs of states.

5 Application to Model Checking of Continuous-State LMCs

Suppose we are given an LMC $\mathscr{M}^{\mathcal{C}} = \left(S^{\mathcal{C}}, \Sigma^{\mathcal{C}}, \tau^{\mathcal{C}}, L^{\mathcal{C}}\right)$, which we shall refer to as the "concrete" model, possibly over a continuous state space. We are interested in calculating its probability of satisfying a given LTL formula, starting from certain initial states. One approach is to construct a finite-state LMC $\mathscr{M}^{\mathcal{A}} = \left(S^{\mathcal{A}}, \Sigma^{\mathcal{A}}, \tau^{\mathcal{A}}, L^{\mathcal{A}}\right)$ (the "abstract model") that can be related to $\mathscr{M}^{\mathcal{C}}$ (in a way to be made precise shortly). Probabilistic model checking can then be run over $\mathscr{M}^{\mathcal{A}}$ using standard tools for finite-state models such as PRISM [19], and since $\mathscr{M}^{\mathcal{A}}$ is related to $\mathscr{M}^{\mathcal{C}}$, this leads to approximate outcomes that are valid for $\mathscr{M}^{\mathcal{C}}$. The above approach has been studied in several papers [2, 3, 16], and the method for constructing $\mathscr{M}^{\mathcal{A}}$ from $\mathscr{M}^{\mathcal{C}}$ is to raise smoothness assumptions on the kernel $\tau^{\mathcal{C}}$ of $\mathscr{M}^{\mathcal{C}}$, and to partition the state space $S^{\mathcal{C}}$, thus obtaining $S^{\mathcal{A}}$ and $\tau^{\mathcal{A}}$ (the sigma algebra and labels being directly inherited).

In this section we will demonstrate the application of our results. We will employ ε-bisimulation to relate $\mathscr{M}^{\mathcal{A}}$ and $\mathscr{M}^{\mathcal{C}}$, and use our results to bound their trace distance. This method produces tighter error bounds, for a broader class of properties, than are currently established in literature. The first step is to establish simpler conditions that guarantee ε-bisimulation between $\mathscr{M}^{\mathcal{C}}$ and $\mathscr{M}^{\mathcal{A}}$.

Theorem 7. *Let $\varepsilon \in [0, 1]$, and suppose there exists a finite measurable partition $\mathcal{Q}_{\varepsilon} = \{P_1, \cdots, P_N\}$ of $S^{\mathcal{C}}$ such that for all $P \in \mathcal{Q}_{\varepsilon}$, $s_1, s_2 \in P$, we have that $L^{\mathcal{C}}(s_1) = L^{\mathcal{C}}(s_2)$ and*[3]

$$\max_{J \subseteq \{1, \cdots, N\}} \left| \tau^{\mathcal{C}}\left(s_1, \bigcup_{j \in J} P_j\right) - \tau^{\mathcal{C}}\left(s_2, \bigcup_{j \in J} P_j\right) \right| \leq \varepsilon.$$

Assume wlog $P_i \neq \emptyset$, and for each $i \in \{1, \cdots, N\}$ choose a representative point $s_i^{\mathcal{C}} \in P_i$. Consider the abstract model to be $\mathscr{M}^{\mathcal{A}} = \left(S^{\mathcal{A}}, \Sigma^{\mathcal{A}}, \tau^{\mathcal{A}}, L^{\mathcal{A}}\right)$ formed by merging each P_i into $s_i^{\mathcal{C}}$. Formally,

- $S^{\mathcal{A}} = \{s_1^{\mathcal{A}}, \cdots, s_N^{\mathcal{A}}\}$.
- $\Sigma^{\mathcal{A}} = \mathcal{P}\left(S^{\mathcal{A}}\right)$.
- $\tau^{\mathcal{A}}$ *is such that* $\tau^{\mathcal{A}}(s_i^{\mathcal{A}}, \{s_j^{\mathcal{A}}\}) = \tau^{\mathcal{C}}(s_i^{\mathcal{C}}, P_j)$.
- $L^{\mathcal{A}}(s_i^{\mathcal{A}}) = L^{\mathcal{C}}(s_i^{\mathcal{C}})$.

Then, for all $i \in \{1, \cdots, N\}$, $s^{\mathcal{C}} \in P_i$, we have that $s^{\mathcal{C}}$ is ε-bisimilar to $s_i^{\mathcal{A}}$, and hence $1 - (1 - \varepsilon)^k$-trace equivalent.

In practical terms the partition $\mathcal{Q}_{\varepsilon}$ can be straightforwardly constructed in many cases. As shown in Theorem 8, the approach in [2, 3, 25] generates a partition of $S^{\mathcal{C}}$ satisfying the conditions of Theorem 7. Thus, ε-bisimulation can be seen as the underlying reason for the closeness of probabilities of events.

---

[3] The left hand side is just $d_{\mathrm{TV}}(\mu_1, \mu_2)$ where for $i = 1, 2$, for $A \subseteq \mathcal{Q}_{\varepsilon}$, $\mu_i(A) \coloneqq \tau^{\mathcal{C}}\left(s_i, \bigcup A\right)$

Theorem 8. *Consider an LMC $\mathscr{M}^{\mathcal{C}} = \left(S^{\mathcal{C}}, \Sigma^{\mathcal{C}}, \tau^{\mathcal{C}}, L^{\mathcal{C}}\right)$ where $S^{\mathcal{C}}$ is a Borel subset of $\mathbb{R}^d$. Suppose that $\tau^{\mathcal{C}}(s, T)$ is of the form $\int_{t \in T} f(s, t) \mathrm{d}t$, so that for each state $s \in S^{\mathcal{C}}$, $f(s, \cdot) : S^{\mathcal{C}} \to \mathbb{R}_0^+$ is the probability density of the next state. Suppose further that $f(\cdot, t)$ is uniformly K-Lipschitz continuous for all $t \in S^{\mathcal{C}}$. That is, for some $K \in \mathbb{R}$, for all $s_1, s_2, t \in S^{\mathcal{C}}$,*

$$|f(s_1, t) - f(s_2, t)| \leq K \cdot \|s_1 - s_2\|.$$

For $A \in \Sigma^{\mathcal{C}}$ (so $A \subseteq \mathbb{R}^d$), let $\lambda(A)$ be the volume of A and $\delta(A) \coloneqq \sup_{x_1, x_2 \in A}\{\|x_1 - x_2\|\}$ be the diameter of A. For any $\varepsilon \in [0, 1]$, finite $\lambda(S^{\mathcal{C}})$, suppose partition $\mathcal{Q} = \{P_1, \cdots, P_N\}$ of $S^{\mathcal{C}}$ is such that

$$\max_{j \in \{1, \cdots, N\}} \delta(P_j) \leq \frac{2\varepsilon}{K\lambda(S^{\mathcal{C}})}.$$

Then, we have that $\mathcal{Q}$ satisfies the conditions of Theorem 7 and can be used to construct the abstract model $\mathscr{M}^{\mathcal{A}}$.

There are a number of adaptations that could be made to this result. [16] improves a related approach by varying the size of each partition in response to the local Lipschitz constant, rather than enforcing a globally uniform K. Similarly to this paper, [3] also discusses the relation of approximate probabilistic bisimulation to the problem of generating the abstract model, but a strictly weaker definition of approximate probabilistic bisimulation is employed (cf. Section 7). Finally, note that using algorithms in [13], we can compute ε-bisimulation relations on $\mathscr{M}^{\mathcal{A}}$: this allows $\mathscr{M}^{\mathcal{A}}$ to be further compressed (at the cost of an additional ε_2 approximation), by merging the states that are ε_2-bisimilar to each other.

6   Case Study

Concrete Model. Consider the concrete model $\mathscr{M}^{\mathcal{C}} = \left(S^{\mathcal{C}}, \Sigma^{\mathcal{C}}, \tau^{\mathcal{C}}, L^{\mathcal{C}}\right)$, describing the weather forecast for a resort. Here $S^{\mathcal{C}} = \{0, 1\} \times [0, 1)$, $\Sigma^{\mathcal{C}} = \mathcal{B}(S^{\mathcal{C}})$, and the state at time t is (R_t, H_t), where

- $R_t \in \{0, 1\}$ is a random variable representing whether it rains on day t,
- $H_t \in [0, 1)$ is a random variable representing the humidity after day t.

Raining on day t causes it to become more likely to rain on day $t + 1$, but it also tends to reduce the humidity, which causes it to become gradually less likely to rain in the future. The meteorological variations are encompassed by

$\tau^{\mathcal{C}}$, which is such that the model evolves according to

$$\mathbb{P}(R_{t+1} \mid R_0, \cdots, R_t, H_0, \cdots, H_t) = \mathbb{P}(R_{t+1} \mid R_t, H_t)$$

$$\sim \begin{cases} \mathrm{B}(\frac{1}{4} + \frac{3}{4}H_t) & \text{if } R_t = 1 \\ \mathrm{B}(\frac{3}{4}H_t) & \text{if } R_t = 0 \end{cases},$$

$$\mathbb{P}(H_{t+1} \mid R_0, \cdots, R_t, H_0, \cdots, H_t, R_{t+1}) = \mathbb{P}(H_{t+1} \mid H_t, R_{t+1})$$

$$\sim \begin{cases} \mathrm{U}\left[0, \frac{1+H_t}{2}\right) & \text{if } R_{t+1} = 1 \\ \mathrm{U}\left[\frac{H_t}{2}, 1\right) & \text{if } R_{t+1} = 0 \end{cases},$$

where $\mathrm{B}(p)$ is the Bernoulli distribution with probability p of producing 1, and $\mathrm{U}\left[a, b\right)$ is the uniform distribution over the real interval $[a, b)$. Finally, the states of the model are labelled according to whether it rains on that day, namely

$$L^{\mathcal{C}}((r, h)) = \begin{cases} \{\mathrm{RAIN}\} & \text{if } r = 1 \\ \emptyset & \text{if } r = 0. \end{cases}$$

Given $\mathscr{M}^{\mathcal{C}}$ we are interested in computing the likelihood of events expressing meteorological predictions, given knowledge of present weather conditions.

Synthesis of the Abstract Model. Notice that $\mathscr{M}^{\mathcal{C}}$ does not directly satisfy the smoothness assumptions of Theorem 8, in view of the discrete/continuous structure of its state space and the discontinuous probability density resulting from the uniform distribution. Nonetheless, we can still construct $\mathscr{M}^{\mathcal{A}}$ by taking a sensible partition of $S^{\mathcal{C}}$ and proving that it satisfies the conditions of Theorem 7. Let $\mathcal{Q}_\varepsilon \coloneqq \left\{ P_{r,h} \mid r \in \{0, 1\}, \ h \in \{0, \cdots, N-1\} \right\}$, where $P_{r,h} = \{r\} \times [h/N, h+1/N)$.

Theorem 9. *For any $\varepsilon \in [0, 1]$, by taking $N \geq 2/\varepsilon$, we have that $\mathcal{Q}_\varepsilon$ satisfies the conditions of Theorem 7.*

Therefore, we may construct the abstract model using Theorem 7. We choose the abstract state $(r^{\mathcal{A}}, h^{\mathcal{A}}) \in S^{\mathcal{A}} \coloneqq \{0, 1\} \times \{0, \cdots, N-1\}$ to correspond to the partition $P_{r^{\mathcal{A}}, h^{\mathcal{A}}} \in \mathcal{Q}_\varepsilon$, and within each partition we select the concrete state with the lowest H_t-coordinate to be the representative state. This produces the abstract model $\mathscr{M}^{\mathcal{A}} = \left(S^{\mathcal{A}}, \Sigma^{\mathcal{A}}, \tau^{\mathcal{A}}, L^{\mathcal{A}}\right)$, where $\Sigma^{\mathcal{A}} = \mathcal{P}(S^{\mathcal{A}})$, $L^{\mathcal{A}}((r, h)) = L^{\mathcal{C}}((r, h))$, and

$$\tau^{\mathcal{A}}\left((h_0, r_0), \{(h_1, r_1)\}\right) = \begin{cases} p_{\mathrm{R}_1 \mid \mathrm{R}_0}(h_0) \cdot p_{\mathrm{H}_1 \mid \mathrm{R}_1}(h_0, h_1) & \text{if } r_0 = 1, r_1 = 1 \\ \left(1 - p_{\mathrm{R}_1 \mid \mathrm{R}_0}(h_0)\right) \cdot p_{\mathrm{H}_1 \mid \neg \mathrm{R}_1}(h_0, h_1) & \text{if } r_0 = 1, r_1 = 0 \\ p_{\mathrm{R}_1 \mid \neg \mathrm{R}_0}(h_0) \cdot p_{\mathrm{H}_1 \mid \mathrm{R}_1}(h_0, h_1) & \text{if } r_0 = 0, r_1 = 1 \\ \left(1 - p_{\mathrm{R}_1 \mid \neg \mathrm{R}_0}(h_0)\right) \cdot p_{\mathrm{H}_1 \mid \neg \mathrm{R}_1}(h_0, h_1) & \text{if } r_0 = 0, r_1 = 0, \end{cases}$$

where

$-\ p_{\mathrm{R}_1 \mid \mathrm{R}_0}(h_0) = \frac{1}{4} + \frac{3}{4}\frac{h_0}{N}$, and $p_{\mathrm{R}_1 \mid \neg \mathrm{R}_0}(h_0) = \frac{3}{4}\frac{h_0}{N}$,

- $p_{H|\neg \mathrm{R}}(h_0, h_1) = \frac{2}{2N - h_0} \cdot \max\left(\min\left(h_1 + 1 - {h_0}/2, 1\right), 0\right)$,
- $p_{H|\mathrm{R}}(h_0, h_1) = \frac{2}{N + h_0} \cdot \max\left(\min\left(\frac{N + h_0}{2} - h_1, 1\right), 0\right)$.

Computation of Approximate Satisfaction Probabilities. Suppose that at the end of day 0, we have $R_0 = 0, H_0 = 0.5$, and a travel agent wants to know the risk of there being 2 consecutive days of rain over the next three days.

This probability can be computed algorithmically according to $\mathcal{M}^{\mathcal{A}}$, and for $N = 1000$, this is 0.365437 (see Appendix G of [4]). Since point $(0, 500) \in S^{\mathcal{A}}$ is 0.001-bisimilar with $(0, 0.5) \in S^{\mathcal{C}}$, this means that according $\mathcal{M}^{\mathcal{C}}$ with initial state $(0, 0.5) \in S^{\mathcal{C}}$, the probability of there being two consecutive days of rain over the next three days is 0.365437 ± 0.003.

Analytical Validation of the Result. In this setup it is possible to evaluate the exact result for $\mathcal{M}^{\mathcal{C}}$ analytically:

$$\mathbb{P}\left[R_1 = 1, R_2 = 1\right] = \mathbb{P}\left[R_1 = 1\right] \int_0^1 f_{H_1 | R_1 = 1}(h_1) \cdot \mathbb{P}\left[R_2 = 1 \mid H_1 = h_1, R_1 = 1\right] \mathrm{d}h_1,$$

which amounts to 0.199219, and similarly $\mathbb{P}\left[R_1 = 0, R_2 = 1, R_3 = 1\right] = 0.166626$. This yields

$$\mathbb{P}\left[\text{Two consecutive rainy days over next 3 days}\right] =$$
$$\mathbb{P}\left[R_1 = 1, R_2 = 1z\right] + \mathbb{P}\left[R_1 = 0, R_2 = 1, R_3 = 1\right] = 0.365845,$$

which is within the error bounds guaranteed by ε-trace equivalence, as expected.

7   Other Notions of ε-Bisimulation

The following condition appears in literature [1,3,14] as the definition of approximate probabilistic bisimulation. For simplicity, let us restrict our attention to finite state spaces.

Definition 9 (Alternative notion of approximate probabilistic bisimulation, adapted from [3]). *Let $\mathcal{M} = (S, \Sigma, \tau, L)$ be a LMC, where S is finite and $\Sigma = \mathcal{P}(S)$. For $\varepsilon \in [0, 1]$, a binary relation R_ε on S satisfies Definition 9 if:*

$$\forall (s_1, s_2) \in R_\varepsilon, \qquad \text{we have } L(s_1) = L(s_2),$$

$$\forall (s_1, s_2) \in R_\varepsilon, \ \forall \tilde{T} \subseteq S \text{ s.t. } \tilde{T} \text{ is } R_\varepsilon\text{-closed}, \qquad \text{we have } \left| \tau(s_1, \tilde{T}) - \tau(s_2, \tilde{T}) \right| \leq \varepsilon.$$

This is different from our notion of approximate probabilistic bisimulation because $\tilde{T}$ ranges over R_ε-closed sets rather than all (measurable) sets. This definition is closer to exact probabilistic bisimulation (cf. Definition 3), but it is too weak to effectively bound probabilistic trace distance.

Theorem 10. *For any $\varepsilon > 0$, there exists an LMC $\mathcal{M} = (S, \Sigma, \tau, L)$, a binary relation R_ε on S, and a pair of states $(s_1, s_2) \in R_\varepsilon$, such that R_ε satisfies the conditions in Definition 9, but the 2-step reachability probabilities from s_1, s_2 differ by 1 for some destination states.*

Proof. For $\varepsilon > 0$, let $N \in \mathbb{Z}^+$, $1/N \leq \varepsilon$. Consider the following LMC. Let $R_\varepsilon := \{(s_1, s_2)\} \cup \{(t_k, t_{k+1}) \mid k \in \{0, \cdots, N-1\}\}$.
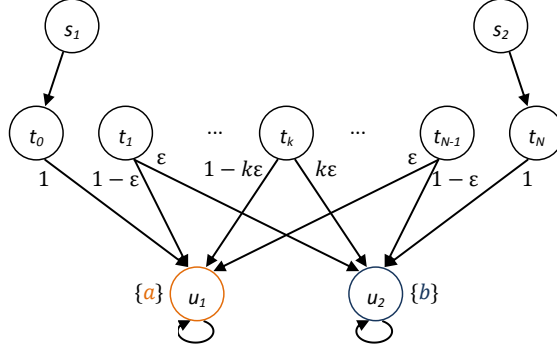
The only R_ε-*closed sets are* $\{s_1, s_2\}$, $\{t_0, \ldots, t_N\}$, $\{u_1\}$, $\{u_2\}$ *and unions of these sets, and so* R_ε *satisfies Definition 9.*

We have $s_1 R_\varepsilon s_2$, *and yet* $P_2(s_1, \lozenge^{\leq 2} a) = 1$ *but* $P_2(s_2, \lozenge^{\leq 2} a) = 0$, *where* $\lozenge^{\leq 2} a$ *is the set of length 3 traces that reach a state labelled with* a.

As shown in [14] however, there is still some relationship between the probabilities of specific traces, which hinges on additional details of the structure of the transition kernel.

8   Conclusions and Extensions

In this paper we have developed a theory of $f(k)$-trace equivalence. We derived the minimum $f(k)$ such that ε-bisimulation implies $f(k)$-trace equivalence, thus extending the well known result for the exact non-probabilistic case. By linking error bounds on the total variation of length k traces to a notion of approximation based on the underlying transition kernel, we provided a means of computing upper bounds for the total variation and of synthesising abstract models with arbitrarily small total variation to a given concrete model.

It is of interest to extend our results to allow the states of the LMC to be labelled with bounded real-valued rewards, and then to limit the difference in expected reward between approximately bisimilar states.

Acknowledgments

References

1. Abate, A.: Approximation metrics based on probabilistic bisimulations for general state-space Markov processes: a survey. Electronic Notes in Theoretical Computer Science 297, 3–25 (2013)
2. Abate, A., Katoen, J.P., Lygeros, J., Prandini, M.: Approximate model checking of stochastic hybrid systems. European Journal of Control 16(6), 624–641 (2010)
3. Abate, A., Kwiatkowska, M., Norman, G., Parker, D.: Probabilistic model checking of labelled Markov processes via finite approximate bisimulations. In: Horizons of the Mind. A Tribute to Prakash Panangaden, pp. 40–58. Springer (2014)
4. Bian, G., Abate, A.: On the relationship between bisimulation and trace equivalence in an approximate probabilistic context (extended version). arXiv.org (2017), http://arxiv.org/abs/1701.04547
5. Bollobás, B., Varopoulos, N.: Representation of systems of measurable sets. Mathematical Proceedings of the Cambridge Philosophical Society 78(02), 323–325 (1975)
6. Buchholz, P.: Exact and ordinary lumpability in finite Markov chains. Journal of applied probability pp. 59–75 (1994)
7. Castro, P.S., Panangaden, P., Precup, D.: Notions of state equivalence under partial observability. In: IJCAI (2009)
8. Chen, D., van Breugel, F., Worrell, J.: On the complexity of computing probabilistic bisimilarity. In: Foundations of Software Science and Computational Structures, pp. 437–451. Springer (2012)
9. Chen, T., Kiefer, S.: On the total variation distance of labelled Markov chains. In: Proceedings of the Joint Meeting of the Twenty-Third EACSL Annual Conference on Computer Science Logic (CSL) and the Twenty-Ninth Annual ACM/IEEE Symposium on Logic in Computer Science (LICS). p. 33. ACM (2014)
10. Comanici, G., Panangaden, P., Precup, D.: On-the-fly algorithms for bisimulation metrics. In: Quantitative Evaluation of Systems (QEST), 2012 Ninth International Conference on. pp. 94–103. IEEE (2012)
11. Daca, P., Henzinger, T.A., Kretínský, J., Petrov, T.: Linear distances between markov chains. In: 27th International Conference on Concurrency Theory, CONCUR 2016, August 23-26, 2016, Québec City, Canada. pp. 20:1–20:15 (2016)
12. Desharnais, J., Edalat, A., Panangaden, P.: Bisimulation for labelled Markov processes. Information and Computation 179(2), 163–193 (2002)
13. Desharnais, J., Laviolette, F., Tracol, M.: Approximate analysis of probabilistic processes: Logic, simulation and games. In: Quantitative Evaluation of Systems, 2008. QEST'08. Fifth International Conference on. pp. 264–273. IEEE (2008)
14. D'Innocenzo, A., Abate, A., Katoen, J.P.: Robust PCTL model checking. In: Proceedings of the 15th ACM international conference on Hybrid Systems: Computation and Control. pp. 275–286. ACM (2012)
15. Durrett, R.: Probability: Theory and Examples - Third Edition. Duxbury Press (2004)
16. Esmaeil Zadeh Soudjani, S., Abate, A.: Adaptive and sequential gridding procedures for the abstraction and verification of stochastic processes. SIAM Journal on Applied Dynamical Systems 12(2), 921–956 (2013)
17. Gebler, D., Tini, S.: Compositionality of approximate bisimulation for probabilistic systems. In: EXPRESS/SOS 2013. pp. 32–46 (2013)
18. Gibbs, A.L., Su, F.E.: On choosing and bounding probability metrics. International statistical review 70(3), 419–435 (2002)

19. Kwiatkowska, M., Norman, G., Parker, D.: PRISM 4.0: Verification of probabilistic real-time systems. In: Computer aided verification. pp. 585–591. Springer (2011)
20. Larsen, K.G., Skou, A.: Bisimulation through probabilistic testing. In: in Conference Record of the 16th ACM Symposium on Principles of Programming Languages (POPL. pp. 344–352 (1989)
21. Segala, R., Lynch, N.: Probabilistic simulations for probabilistic processes. Nordic Journal of Computing (1995)
22. Spears, W.M.: A compression algorithm for probability transition matrices. SIAM journal on matrix analysis and applications 20(1), 60–77 (1998)
23. Sproston, J., Donatelli, S.: Backward bisimulation in Markov chain model checking. Software Engineering, IEEE Transactions on 32(8), 531–546 (2006)
24. Tang, Q., van Breugel, F.: Computing probabilistic bisimilarity distances via policy iteration. In: 27th International Conference on Concurrency Theory, CONCUR 2016, August 23-26, 2016, Québec City, Canada. pp. 22:1–22:15 (2016)
25. Tkachev, I., Abate, A.: Formula-free finite abstractions for linear temporal verification of stochastic hybrid systems. In: Proceedings of the 16th international conference on Hybrid systems: computation and control. pp. 283–292. ACM (2013)
26. Wu, H., Noé, F.: Probability distance based compression of hidden Markov models. Multiscale Modeling & Simulation 8(5), 1838–1861 (2010)