When Equivalence and Bisimulation Join Forces in Probabilistic Automata[*]

Yuan Feng[1,2] and Lijun Zhang[3]

[1] University of Technology Sydney, Australia
[2] Department of Computer Science and Technology, Tsinghua University, China
[3] State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences

Abstract. Probabilistic automata were introduced by Rabin in 1963 as language acceptors. Two automata are equivalent if and only if they accept each word with the same probability. On the other side, in the process algebra community, probabilistic automata were re-proposed by Segala in 1995 which are more general than Rabin's automata. Bisimulations have been proposed for Segala's automata to characterize the equivalence between them. So far the two notions of equivalences and their characteristics have been studied most independently. In this paper, we consider Segala's automata, and propose a novel notion of distribution-based bisimulation by joining the existing equivalence and bisimilarities. Our bisimulation bridges the two closely related concepts in the community, and provides a uniform way of studying their characteristics. We demonstrate the utility of our definition by studying distribution-based bisimulation metrics, which gives rise to a robust notion of equivalence for Rabin's automata.

1 Introduction

In 1963, Rabin [29] introduced the model *probabilistic automata* as language acceptors. In a probabilistic automaton, each input symbol determines a stochastic transition matrix over the state space. Starting with the initial distribution, each word (a sequence of symbols) has a corresponding probability of reaching one of the final states, which is referred to the accepting probability. Two automata are equivalent if and only if they accept each word with the same probability. The corresponding decision algorithm has been extensively studied, see [29, 31, 25, 26].

Markov decision processes (MDPs) were known as early as the 1950s [3], and are a popular modeling formalism used for instance in operations research, automated planning, and decision support systems. In MDPs, each state has a set of enabled actions and each enabled action leads to a distribution over successor states. MDPs have been widely used in the formal verification of randomized concurrent systems, and are now supported by probabilistic model checking tools such as PRISM [27], MRMC [24] and IscasMC [20].

On the other side, in the context of concurrent systems, probabilistic automata were re-proposed by Segala in 1995 [30], which extend MDPs with internal nondeterministic choices. Segala's automata are more general than Rabin's automata, in the sense that each input symbol corresponds to one, or more than one, stochastic transition matrices. Various behavioral equivalences are defined, including strong bisimulations, strong probabilistic bisimulations, and weak bisimulation extensions [30]. These behavioral equivalences are used as powerful tools for state space reduction and hierarchical verification of complex systems. Thus, their decision algorithms [4, 2, 23] and logical characterizations [28, 12, 22] are widely studied in the literature.

Equivalences are defined for the specific initial distributions over Rabin's automata, whereas bisimulations are usually defined over states. For Segala's automata, state-based bisimulations have arguably too strong distinguishing power, thus in the recent literature, various relaxations have been proposed. The earliest such formulation is a distribution-based bisimulation in [14], which is defined for Rabin's automata. This is essentially an equivalent characterization of the equivalence in the coinductive manner, as for bisimulations. Recently, in [15], a distribution-based weak bisimulation has been proposed, and the induced distribution-based strong bisimulation is further studied in [21]. It is shown that the distribution-based strong bisimulation agrees with the state-based bisimulations when lifted to distributions.

To the best of the authors' knowledge, even the two notions are closely related, so far their characteristics have been studied independently. As the main contribution of this paper, we consider Segala's probabilistic automata, and propose a novel notion of distribution-based bisimulation by joining the existing equivalence and bisimilarities. We show that for Rabin's probabilistic automata it coincides with equivalences, and for Segala's probabilistic automata, it is reasonably weaker than the existing bisimulation relation. Thus, our bisimulations bridge the two closely related concepts in the community, and provide a uniform way of studying their characteristics.

We demonstrate the utility of our approach by studying distribution-based bisimulation metrics. Bisimulations for probabilistic systems are known to be very sensitive to the transition probabilities: even a tiny perturbation of the transition probabilities will destroy bisimilarity. Thus, bisimulation metrics have been proposed [19]: the distance between any two states are measured, and the smaller the distance is, the more similar they are. If the distance is zero, one then has the classical bisimulation. Because of the nice property of robustness, bisimulation metrics have attracted a lot attentions on MDPs and their extension with continuous state space, see [10, 8, 11, 13, 17, 6, 18, 1, 7].

All of the existing bisimulation metrics mentioned above are state-based. On the other side, as states lead to distributions in MDPs, the metrics must be lifted to distributions. In the second part of the paper, we propose a distribution-based bisimulation metric; we consider it being more natural as no lifting of distances is needed. We provide a coinductive definition as well as a fixed point characterization, both of which are used in defining the state-based bisimulation metrics in the literature. We provide a logical characterization for this metric as well, and discuss the relation of our definition and the state-based ones.

A direct byproduct of our bisimulation-based metrics is the notion of equivalence metrics for Rabin's probabilistic automata. As for bisimulation metrics, the equivalence

metric provides a robust solution for comparing Rabin's automata. To the best of our knowledge, this has not been studied in the literature. We anticipate that more solution techniques developed in one area can inspire solutions for the corresponding problems in the other.

Organization of the paper. We introduce some notations in Section 2. Section 3 recalls the definitions of probabilistic automata, equivalence, and bisimulation relations. We present our distribution-based bisimulation in Section 4, and bisimulation metrics and their logical characterizations in 5. Section 6 concludes the paper. Due to the lack of space, most proofs were omitted in this paper; please refer to the full version at [16] for the details.

2 Preliminaries

Distributions. For a finite set S, a distribution is a function $\mu : S \to [0,1]$ satisfying $|\mu| := \sum_{s \in S} \mu(s) = 1$. We denote by $Dist(S)$ the set of distributions over S. We shall use $s, r, t, \ldots$ and $\mu, \nu \ldots$ to range over S and $Dist(S)$, respectively. Given a set of distributions $\{\mu_i\}_{1 \leq i \leq n}$, and a set of positive weights $\{p_i\}_{1 \leq i \leq n}$ such that $\sum_{1 \leq i \leq n} p_i = 1$, the *convex combination* $\mu = \sum_{1 \leq i \leq n} p_i \cdot \mu_i$ is the distribution such that $\mu(s) = \sum_{1 \leq i \leq n} p_i \cdot \mu_i(s)$ for each $s \in S$. The support of μ is defined by $supp(\mu) := \{s \in S \mid \mu(s) > 0\}$. For an equivalence relation R defined on S, we write $\mu R \nu$ if it holds that $\mu(C) = \nu(C)$ for all equivalence classes $C \in S/R$. A distribution μ is called *Dirac* if $|supp(\mu)| = 1$, and we let δ_s denote the Dirac distribution with $\delta_s(s) = 1$.

Note that when S is finite, the distributions $Dist(S)$ over S, when regarded as a subset of $\mathbb{R}^{|S|}$, is both convex and compact. In this paper, when we talk about convergence of distributions, or continuity of relations such as transitions, bisimulations, and pseudometrics between distributions, we are referring to the normal topology of $\mathbb{R}^{|S|}$. For a set $F \subseteq S$, we define the (column) characteristic vector η_F by letting $\eta_F(s) = 1$ if $s \in F$, and 0 otherwise.

Pseudometric. A pseudometric over $Dist(S)$ is a function $d : Dist(S) \times Dist(S) \to [0,1]$ such that

1. $d(\mu, \mu) = 0$;
2. $d(\mu, \nu) = d(\nu, \mu)$;
3. $d(\mu, \nu) + d(\nu, \omega) \geq d(\mu, \omega)$.

In this paper, we assume that a pseudometric is continuous.

3 Probabilistic Automata and Bisimulations

3.1 Probabilistic Automata

Let AP be a finite set of atomic propositions. We recall the notion of probabilistic automata introduced by Segala [30].

Definition 1 (Probabilistic Automata). *A* probabilistic automaton *is a tuple* $\mathcal{A} = (S, Act, \rightarrow, L, \alpha)$ *where* S *is a finite set of states, Act is a finite set of actions,* $\rightarrow \subseteq S \times Act \times Dist(S)$ *is a transition relation,* $L : S \rightarrow 2^{AP}$ *is a labeling function, and* $\alpha \in Dist(S)$ *is an initial distribution.*

As usual we only consider image-finite probabilistic automata, i.e. for all $s \in S$, the set $\{\mu \mid (s, a, \mu) \in \rightarrow\}$ is finite. A transition $(s, a, \mu) \in \rightarrow$ is denoted by $s \xrightarrow{a} \mu$. We denote by $Act(s) := \{a \mid s \xrightarrow{a} \mu\}$ the set of enabled actions in s. We say $\mathcal{A}$ is *input enabled*, if $Act(s) = Act$ for all $s \in S$. We say $\mathcal{A}$ is an MDP if Act is a singleton.

Interestingly, a subclass of probabilistic automata were already introduced by Rabin in 1963 [29]; Rabin's probabilistic automata were referred to as *reactive automata* in [30]. We adopt this convention in this paper.

Definition 2 (Reactive Automata). *We say $\mathcal{A}$ is* reactive *if it is input enabled, and for all s, $L(s) \in \{\emptyset, AP\}$, and $s \xrightarrow{a} \mu \wedge s \xrightarrow{a} \mu' \Rightarrow \mu = \mu'$.*

Here the condition $L(s) \in \{\emptyset, AP\}$ implies that the states can be partitioned into two equivalence classes according to their labeling. Below we shall identify $F := \{s \mid L(s) = AP\}$ as the set of *accepting states*, a terminology used in reactive automata. In a reactive automaton, each action $a \in Act$ is enabled precisely once for all $s \in S$, thus inducing a stochastic matrix $M(a)$ satisfying $s \xrightarrow{a} M(a)(s, \cdot)$.

3.2 Probabilistic Bisimulation and Equivalence

First, we recall the definition of (strong) probabilistic bisimulation for probabilistic automata [30]. Let $\{s \xrightarrow{a} \mu_i\}_{i \in I}$ be a collection of transitions, and let $\{p_i\}_{i \in I}$ be a collection of probabilities with $\sum_{i \in I} p_i = 1$. Then $(s, a, \sum_{i \in I} p_i \cdot \mu_i)$ is called a *combined transition* and is denoted by $s \xrightarrow{a}_{\mathsf{P}} \mu$ where $\mu = \sum_{i \in I} p_i \cdot \mu_i$.

Definition 3 (Probabilistic bisimulation [30]). *An equivalence relation $R \subseteq S \times S$ is a probabilistic bisimulation if sRr implies that $L(s) = L(r)$, and for each $s \xrightarrow{a} \mu$, there exists a combined transition $r \xrightarrow{a}_{\mathsf{P}} \nu$ such that $\mu R \nu$.*

We write $s \sim_{\mathsf{P}} r$ whenever there is a probabilistic bisimulation R such that sRr.

Recently, in [15], a distribution-based weak bisimulation has been proposed, and the induced distribution-based strong bisimulation is further studied in [21]. Their bisimilarity is shown to be the same as $\sim_{\mathsf{P}}$ when lifted to distributions. Below we recall the definition of equivalence for reactive automata introduced by Rabin [29].

Definition 4 (Equivalence for Reactive Automata [29]). *Let $\mathcal{A}_i = (S_i, Act_i, \rightarrow_i, L_i, \alpha_i)$ with $i = 1, 2$ be two reactive automata with $Act_1 = Act_2 =: Act$, and $F_i = \{s \in S_i \mid L(s) = AP\}$ the set of final states for $\mathcal{A}_i$. We say $\mathcal{A}_1$ and $\mathcal{A}_2$ are equivalent if $\mathcal{A}_1(w) = \mathcal{A}_2(w)$ for each $w \in Act^*$, where $\mathcal{A}_i(w) := \alpha_i M_i(a_1) \ldots M_i(a_k) \eta_{F_i}$ provided $w = a_1 \ldots a_k$.*

Stated in plain english, $\mathcal{A}_1$ and $\mathcal{A}_2$ with the same set of actions are equivalent iff for an arbitrary input w, the probabilities of *absorbing* in F_1 and F_2 are the same.

So far bisimulations and equivalences were studied most independently. The only exception we are aware is [14], in which for Rabin's probabilistic automata, a distribution-based bisimulation is defined that generalizes both equivalence and bisimulations.

Definition 5 (Bisimulation for Reactive Automata [14]). *Let $\mathcal{A}_i = (S_i, Act_i, \rightarrow_i, L_i, \alpha_i)$ with $i = 1, 2$ be two given reactive automata with $Act_1 = Act_2 =: Act$, and F_i the set of final states for $\mathcal{A}_i$. A relation $R \subseteq Dist(S_1) \times Dist(S_2)$ is a bisimulation if for each $\mu R \nu$ it holds (i) $\mu \cdot \eta_{F_1} = \nu \cdot \eta_{F_2}$, and (ii) $(\mu M_1(a)) R (\nu M_2(a))$ for all $a \in Act$.*

We write $\mu \sim_d \nu$ whenever there is a bisimulation R such that $\mu R \nu$.

It is shown in [14] that two reactive automata are equivalent if and only if their initial distributions are distribution-based bisimilar according to the definition above.

4 A Novel Bisimulation Relation

In this section we introduce a notion of distribution-based bisimulation for Segala's automata by extending the bisimulation defined in [14]. We shall show the compatibility of our definition with previous ones in Subsection 4.1, and some properties of our bisimulation in Subsection 4.2.

For the first step of defining a distribution-based bisimulation, we need to extend the transitions starting from states to those starting from distributions. A natural candidate for such an extension is as follows: for a distribution μ to perform an action a, each state in its support *must* make a combined a-move. However, this definition is problematic, as in Segala's general probabilistic automata, action a may not always be enabled in any support state of μ. In this paper, we deal with this problem by first defining the distribution-based bisimulation (resp. distances) for input enabled automata, for which the transition between distributions can be naturally defined, and then reducing the equivalence (resp. distances) of two distributions in a general probabilistic automata to the bisimilarity (resp. distances) of these distributions in an input enabled automata which is obtained from the original one by adding a *dead* state.

To make our idea more rigorous, we need some notations. For $A \subseteq AP$ and a distribution μ, we define $\mu(A) := \sum \{\mu(s) \mid L(s) = A\}$, which is the probability of being in those state s with label A.

Definition 6. *We write $\mu \xrightarrow{a} \mu'$ if for each $s \in supp(\mu)$ there exists $s \xrightarrow{a}_P \mu_s$ such that $\mu' = \sum_s \mu(s) \cdot \mu_s$.*

We first present our distribution-based bisimulation for input enabled probabilistic automata.

Definition 7. *Let $\mathcal{A} = (S, Act, \rightarrow, L, \alpha)$ be an input enabled probabilistic automaton. A symmetric relation $R \subseteq Dist(S) \times Dist(S)$ is a (distribution-based) bisimulation if $\mu R \nu$ implies that*

1. *$\mu(A) = \nu(A)$ for each $A \subseteq AP$, and*
2. *for each $a \in Act$, whenever $\mu \xrightarrow{a} \mu'$ then there exists a transition $\nu \xrightarrow{a} \nu'$ such that $\mu' R \nu'$.*

We write $\mu \sim^{\mathcal{A}} \nu$ if there is a bisimulation R such that $\mu R \nu$.

Obviously, the bisimilarity $\sim^{\mathcal{A}}$ is the largest bisimulation relation.

For probabilistic automata which are not input enabled, we define distribution-based bisimulation with the help of *input enabled extension* specified as follows.

Definition 8. *Let* $\mathcal{A} = (S, Act, \rightarrow, L, \alpha)$ *be a probabilistic automaton over* AP. *The* input enabled extension *of* $\mathcal{A}$, *denoted by* $\mathcal{A}_{\perp}$, *is defined as an (input enabled) probabilistic automaton* $(S_{\perp}, Act, \rightarrow^{\perp}, L_{\perp}, \alpha)$ *over* $AP_{\perp}$ *where*

1. $S_{\perp} = S \cup \{\perp\}$ *where* $\perp$ *is a* dead *state not in* S;
2. $AP_{\perp} = AP \cup \{dead\}$ *with* $dead \notin AP$;
3. $\rightarrow^{\perp} = \rightarrow \cup \{(s, a, \delta_{\perp}) \mid a \notin Act(s)\} \cup \{(\perp, a, \delta_{\perp}) \mid a \in Act\}$;
4. $L_{\perp}(s) = L(s)$ *for any* $s \in S$, *and* $L_{\perp}(\perp) = \{dead\}$.

Definition 9. *Let* $\mathcal{A}$ *be a probabilistic automaton which is not input enabled. Then* μ *and* ν *are bisimilar, denoted by* $\mu \sim^{\mathcal{A}} \nu$, *if* $\mu \sim^{\mathcal{A}_{\perp}} \nu$ *in* $\mathcal{A}_{\perp}$.

We always omit the superscript $\mathcal{A}$ in $\sim^{\mathcal{A}}$ when no confusion arises.

4.1 Compatibility

In this section we instantiate appropriate labeling functions and show that our notion of bisimilarity is a conservative extension of both probabilistic bisimulation [29] and equivalence relations [14].

Lemma 1. *Let* $\mathcal{A}$ *be a probabilistic automaton where* $AP = Act$, *and* $L(s) = Act(s)$ *for each* s. *Then,* $\mu \sim_{\mathsf{P}} \nu$ *implies* $\mu \sim \nu$.

Probabilistic bisimulation is defined over states inside one automaton, whereas equivalence and distribution for reactive automata are defined over two automata. However, they can be connected by the notion of direct sum of two automata, which is the automaton obtained by considering the disjoint union of states, edges and labeling functions respectively.

Lemma 2. *Let* $\mathcal{A}_1$ *and* $\mathcal{A}_2$ *be two reactive automata with the same set of actions* Act. *Let* $F_i = \{s \in S_i \mid L(s) = AP\}$. *Then, the following are equivalent:*

1. $\mathcal{A}_1$ *and* $\mathcal{A}_2$ *are equivalent,*
2. $\alpha_1 \sim_d \alpha_2$,
3. $\alpha_1 \sim \alpha_2$ *in their direct sum.*

Proof. The equivalence between (1) and (2) is shown in [14]. The equivalence between (2) and (3) is straightforward, as for reactive automata our definition degenerates to Definition 5. □

To conclude this section, we present an example to show that our bisimilarity is *strictly* weaker than $\sim_{\mathsf{P}}$.

Example 1. Consider the example probabilistic automaton depicted in Fig. 1, which is inspired from an example in [14]. Let $AP = Act = \{a\}$, $L(s) = Act(s)$ for each s, and $\varepsilon_1 = \varepsilon_2 = 0$. We argue that $q \nsim_P q'$. Otherwise, note $q \xrightarrow{a} \frac{1}{2}\delta_{r_1} + \frac{1}{2}\delta_{r_2}$ and $q' \xrightarrow{a} \delta_{r'}$. Then we must have $r' \sim_P r_1 \sim_P r_2$. This is impossible, as $r_1 \xrightarrow{a} \frac{2}{3}\delta_{s_1} + \frac{1}{3}\delta_{s_2}$ and $r' \xrightarrow{a} \frac{1}{2}\delta_{s'_1} + \frac{1}{2}\delta_{s'_2}$, but $s_1 \sim_P s'_1 \nsim_P s_2 \sim_P s'_2$.

However, by our definition of bisimulation, the Dirac distributions δ_q and $\delta_{q'}$ are indeed bisimilar. The reason is, we have the following transition

$$\frac{1}{2}\delta_{r_1} + \frac{1}{2}\delta_{r_2} \xrightarrow{a} \frac{1}{3}\delta_{s_1} + \frac{1}{6}\delta_{s_2} + \frac{1}{6}\delta_{s_3} + \frac{1}{3}\delta_{s_4},$$

and it is easy to check $\delta_{s_1} \sim \delta_{s_3} \sim \delta_{s'_1}$ and $\delta_{s_2} \sim \delta_{s_4} \sim \delta_{s'_2}$. Thus we have $\frac{1}{2}\delta_{r_1} + \frac{1}{2}\delta_{r_2} \sim \delta_{r'}$, and finally $\delta_q \sim \delta_{q'}$.
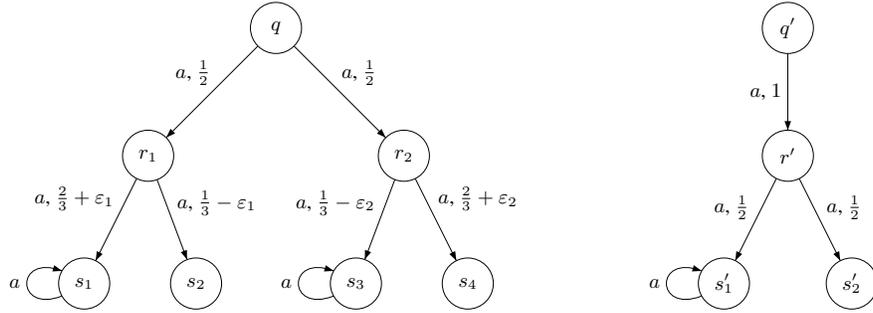

Fig. 1. An illustrating example in which state labelings are defined by $L(s) = Act(s)$.

4.2 Properties of the Relations

In the following, we show that the notion of bisimilarity is in harmony with the linear combination and the limit of distributions.

Definition 10. *A binary relation $R \subseteq Dist(S) \times Dist(S)$ is said to be*

- linear, *if for any finite set I and any probabilistic distribution $\{p_i\}_{i \in I}$, $\mu_i R \nu_i$ for each i implies $(\sum_{i \in I} p_i \cdot \mu_i) R (\sum_{i \in I} p_i \cdot \nu_i)$;*
- continuous, *if for any convergent sequences of distributions $\{\mu_i\}_i$ and $\{\nu_i\}_i$, $\mu_i R \nu_i$ for each i implies $(\lim_i \mu_i) R (\lim_i \nu_i)$;*
- left-decomposable, *if $(\sum_{i \in I} p_i \cdot \mu_i) R \nu$, where $0 < p_i \leq 1$ and $\sum_{i \in I} p_i = 1$, then ν can be written as $\sum_{i \in I} p_i \cdot \nu_i$ such that $\mu_i R \nu_i$ for every $i \in I$.*
- left-convergent, *if $(\lim_i \mu_i) R \nu$, then for any i we have $\mu_i R \nu_i$ for some ν_i with $\lim_i \nu_i = \nu$.*

We prove below that our transition relation between distributions satisfies these properties.

Lemma 3. *For an input enabled probabilistic automata, the transition relation $\xrightarrow{a}$ between distributions is linear, continuous, left-decomposable, and left-convergent.*

Theorem 1. *The bisimilarity relation $\sim$ is both linear and continuous.*

In general, our definition of bisimilarity is not left-decomposable. This is in sharp contrast with the bisimulations defined by using the lifting technique [9]. However, this should not be regarded as a shortcoming; actually it is the key requirement we abandon in this paper, which makes our definition reasonably weak. This has been clearly illustrated in Example 1.

5 Bisimulation Metrics

We present distribution-based bisimulation metrics with discounting factor $\gamma \in (0, 1]$ in this section. Three different ways of defining bisimulation metrics between states exist in the literature: one coinductive definition based on bisimulations [35, 33, 34, 13], one based on the maximal logical differences [10, 11, 32], and one on fixed point [8, 32, 17]. We propose all the three versions for our distribution-based bisimulations with discounting. Moreover, we show that they coincide. We fix a discount factor $\gamma \in (0, 1]$ throughout this section. For any $\mu, \nu \in Dist(S)$, we define the distance

$$d_{AP}(\mu, \nu) := \frac{1}{2} \sum_{A \subseteq AP} |\mu(A) - \nu(A)| .$$

Then it is easy to check that

$$d_{AP}(\mu, \nu) = \max_{\mathcal{B} \subseteq 2^{AP}} \left| \sum_{A \in \mathcal{B}} \mu(A) - \sum_{A \in \mathcal{B}} \nu(A) \right| = \max_{\mathcal{B} \subseteq 2^{AP}} \left[\sum_{A \in \mathcal{B}} \mu(A) - \sum_{A \in \mathcal{B}} \nu(A) \right] .$$

5.1 A Direct Approach

Definition 11. *Let $\mathcal{A} = (S, Act, \rightarrow, L, \alpha)$ be an input enabled probabilistic automaton. A family of symmetric relations $\{R_\varepsilon \mid \varepsilon \geq 0\}$ over $Dist(S)$ is a (discounted) approximate bisimulation if for any $\varepsilon \geq 0$ and $\mu R_\varepsilon \nu$, we have*

1. *$d_{AP}(\mu, \nu) \leq \varepsilon$;*
2. *for each $a \in Act$, $\mu \xrightarrow{a} \mu'$ implies that there exists a transition $\nu \xrightarrow{a} \nu'$ such that $\mu' R_{\varepsilon/\gamma} \nu'$.*

We write $\mu \sim_\varepsilon^{\mathcal{A}} \nu$ whenever there is an approximate bisimulation $\{R_\varepsilon \mid \varepsilon \geq 0\}$ such that $\mu R_\varepsilon \nu$. For any two distributions μ and ν, we define the bisimulation distance of μ and ν as

$$D_b^{\mathcal{A}}(\mu, \nu) = \inf\{\varepsilon \geq 0 \mid \mu \sim_\varepsilon^{\mathcal{A}} \nu\}. \tag{1}$$

Again, the approximate bisimulation and bisimulation distance of distributions in a general probabilistic automaton can be defined in terms of the corresponding notions in the input enabled extension; that is, $\mu \sim_\varepsilon^\mathcal{A} \nu$ if $\mu \sim_\varepsilon^{\mathcal{A}_\perp} \nu$, and $D_b^\mathcal{A}(\mu, \nu) := D_b^{\mathcal{A}_\perp}(\mu, \nu)$. We always omit the superscripts for simplicity if no confusion arises.

It is standard to show that the family $\{\sim_\varepsilon | \varepsilon \geq 0\}$ is itself an approximate bisimulation. The following lemma collects some more properties of $\sim_\varepsilon$.

Lemma 4. *1. For each ε, the ε-bisimilarity $\sim_\varepsilon$ is both linear and continuous.*
2. If $\mu \sim_{\varepsilon_1} \nu$ and $\nu \sim_{\varepsilon_2} \omega$, then $\mu \sim_{\varepsilon_1 + \varepsilon_2} \omega$;
3. $\sim_{\varepsilon_1} \subseteq \sim_{\varepsilon_2}$ whenever $\varepsilon_1 \leq \varepsilon_2$.

The following theorem states that the infimum in the definition Eq. (1) of bisimulation distance can be replaced by minimum; that is, the infimum is achievable.

Theorem 2. *For any $\mu, \nu \in Dist(S)$, $\mu \sim_{D_b(\mu,\nu)} \nu$.*

Proof. By definition, we need to prove $\mu \sim_{D_b(\mu,\nu)} \nu$ in the extended automaton. We first prove that for any $\varepsilon \geq 0$, the symmetric relations $\{R_\varepsilon \mid \varepsilon \geq 0\}$ where

$$R_\varepsilon = \{(\mu, \nu) \mid \mu \sim_{\varepsilon_i} \nu \text{ for each } \varepsilon_1 \geq \varepsilon_2 \geq \cdots \geq 0, \text{ and } \lim_{i \to \infty} \varepsilon_i = \varepsilon\}$$

is an approximate bisimulation. Suppose $\mu R_\varepsilon \nu$. Since $\mu \sim_{\varepsilon_i} \nu$ we have $d_{AP}(\mu, \nu) \leq \varepsilon_i$ for each i. Thus $d_{AP}(\mu, \nu) \leq \varepsilon$ as well. Furthermore, if $\mu \xrightarrow{a} \mu'$, then for any $i \geq 1$, $\nu \xrightarrow{a} \nu_i$ and $\mu' \sim_{\varepsilon_i/\gamma} \nu_i$. Since $Dist(S)$ is compact, there exists a subsequence $\{\nu_{i_k}\}_k$ of $\{\nu_i\}_i$ such that $\lim_k \nu_{i_k} = \nu'$ for some ν'. We claim that

- $\nu \xrightarrow{a} \nu'$. This follows from the continuity of the transition $\xrightarrow{a}$, Lemma 3.
- For each $k \geq 1$, $\mu' \sim_{\varepsilon_{i_k}/\gamma} \nu'$. Suppose conversely that $\mu' \not\sim_{\varepsilon_{i_k}/\gamma} \nu'$ for some k. Then by the continuity of $\sim_{\varepsilon_{i_k}/\gamma}$, we have $\mu' \not\sim_{\varepsilon_{i_k}/\gamma} \nu_j$ for some $j \geq i_k$. This contradicts the fact that $\mu' \sim_{\varepsilon_j/\gamma} \nu_j$ and Lemma 4(3). Thus $\mu' R_{\varepsilon/\gamma} \nu'$ as required.

Finally, it is direct from definition that there exists a decreasing sequence $\{\varepsilon_i\}_i$ such that $\lim_i \varepsilon_i = D_b(\mu, \nu)$ and $\mu \sim_{\varepsilon_i} \nu$ for each i. Then the theorem follows. $\square$

A direct consequence of the above theorem is that the bisimulation distance between two distributions vanishes if and only if they are bisimilar.

Corollary 1. *For any $\mu, \nu \in Dist(S)$, $\mu \sim \nu$ if and only if $D_b(\mu, \nu) = 0$.*

Proof. Direct from Theorem 2, by noting that $\sim = \sim_0$. $\square$

The next theorem shows that D_b is indeed a pseudometric.

Theorem 3. *The bisimulation distance D_b is a pseudometric on $Dist(S)$.*

5.2 Modal Characterization of the Bisimulation Metrics

We now present a Hennessy-Milner type modal logic motivated by [10, 11] to characterize the distance between distributions.

Definition 12. *The class $\mathcal{L}_m$ of modal formulae over AP, ranged over by φ, φ_1, φ_2, etc, is defined by the following grammar:*

$$\varphi ::= \mathcal{B} \mid \varphi \oplus p \mid \neg\varphi \mid \bigwedge_{i \in I} \varphi_i \mid \langle a \rangle \varphi$$

where $\mathcal{B} \subseteq 2^{AP}$, $p \in [0, 1]$, $a \in Act$, and I is an index set.

Given an input enabled probabilistic automaton $\mathcal{A} = (S, Act, \rightarrow, L, \alpha)$ over AP, instead of defining the satisfaction relation $\models$ for the qualitative setting, the (discounted) semantics of the logic $\mathcal{L}_m$ is given in terms of functions from $Dist(S)$ to $[0, 1]$. For any formula $\varphi \in \mathcal{L}_m$, the satisfaction function of φ, denoted by φ again for simplicity, is defined in a structural induction way as follows:

- $\mathcal{B}(\mu) := \sum_{A \in \mathcal{B}} \mu(A)$;
- $(\varphi \oplus p)(\mu) := \min\{\varphi(\mu) + p, 1\}$;
- $(\neg\varphi)(\mu) := 1 - \varphi(\mu)$;
- $(\bigwedge_{i \in I} \varphi_i)(\mu) := \inf_{i \in I} \varphi_i(\mu)$;
- $(\langle a \rangle \varphi)(\mu) := \sup_{\mu \xrightarrow{a} \mu'} \gamma \cdot \varphi(\mu')$.

Lemma 5. *For any $\varphi \in \mathcal{L}_m$, $\varphi : Dist(S) \to [0, 1]$ is a continuous function.*

From Lemma 5, and noting that the set $\{\mu' \mid \mu \xrightarrow{a} \mu'\}$ is compact for each μ and a, the supremum in the semantic definition of $\langle a \rangle \varphi$ can be replaced by maximum; that is, $(\langle a \rangle \varphi)(\mu) = \max_{\mu \xrightarrow{a} \mu'} \gamma \cdot \varphi(\mu')$. Now we define the logical distance for distributions.

Definition 13. *The* logic *distance of μ and ν in $Dist(S)$ of an input enabled automaton is defined by*

$$D_l^{\mathcal{A}}(\mu, \nu) = \sup_{\varphi \in \mathcal{L}_m} |\varphi(\mu) - \varphi(\nu)| . \tag{2}$$

The logic distance for a general probabilistic automaton can be defined in terms of the input enabled extension; that is, $D_l^{\mathcal{A}}(\mu, \nu) := D_l^{\mathcal{A}_\perp}(\mu, \nu)$. We always omit the superscripts for simplicity.

Now we can show that the logic distance exactly coincides with bisimulation distance for any distributions.

Theorem 4. $D_b = D_l$.

Proof. As both D_b and D_l are defined in terms of the input enabled extension of automata, we only need to prove the result for input enabled case. Let $\mu, \nu \in Dist(S)$. We first prove $D_b(\mu, \nu) \geq D_l(\mu, \nu)$. It suffices to show by structural induction that for any $\varphi \in \mathcal{L}_m$, $|\varphi(\mu) - \varphi(\nu)| \leq D_b(\mu, \nu)$. There are five cases to consider.

- $\varphi \equiv \mathcal{B}$ for some $\mathcal{B} \subseteq 2^{AP}$. Then $|\varphi(\mu) - \varphi(\nu)| = |\sum_{A \in \mathcal{B}} [\mu(A) - \nu(A)]| \leq d_{AP}(\mu, \nu) \leq D_b(\mu, \nu)$ by Theorem 2.
- $\varphi \equiv \varphi' \oplus p$. Assume $\varphi'(\mu) \geq \varphi'(\nu)$. Then $\varphi(\mu) \geq \varphi(\nu)$. By induction, we have $\varphi'(\mu) - \varphi'(\nu) \leq D_b(\mu, \nu)$. Thus

$$|\varphi(\mu) - \varphi(\nu)| = \min\{\varphi'(\mu) + p, 1\} - \min\{\varphi'(\nu) + p, 1\} \leq \varphi'(\mu) - \varphi'(\nu) \leq D_b(\mu, \nu).$$

- $\varphi \equiv \neg\varphi'$. By induction, we have $|\varphi'(\mu) - \varphi'(\nu)| \leq D_b(\mu, \nu)$, thus $|\varphi(\mu) - \varphi(\nu)| = |1 - \varphi'(\mu) - 1 + \varphi'(\nu)| \leq D_b(\mu, \nu)$ as well.
- $\varphi \equiv \bigwedge_{i \in I} \varphi_i$. Assume $\varphi(\mu) \geq \varphi(\nu)$. For any $\varepsilon > 0$, let $j \in I$ such that $\varphi_j(\nu) \leq \varphi(\nu) + \varepsilon$. By induction, we have $|\varphi_j(\mu) - \varphi_j(\nu)| \leq D_b(\mu, \nu)$. Then

$$|\varphi(\mu) - \varphi(\nu)| \leq \varphi_j(\mu) - \varphi_j(\nu) + \varepsilon \leq D_b(\mu, \nu) + \varepsilon,$$

and $|\varphi(\mu) - \varphi(\nu)| \leq D_b(\mu, \nu)$ from the arbitrariness of ε.
- $\varphi \equiv \langle a \rangle \varphi'$. Assume $\varphi(\mu) \geq \varphi(\nu)$. Let $\mu'_* \in Dist(S)$ such that $\mu \xrightarrow{a} \mu'_*$ and $\gamma \cdot \varphi'(\mu'_*) = \varphi(\mu)$. From Theorem 2, we have $\mu \sim_{D_b(\mu,\nu)} \nu$. Thus there exists ν'_* such that $\nu \xrightarrow{a} \nu'_*$ and $\mu'_* \sim_{D_b(\mu,\nu)/\gamma} \nu'_*$. Hence $\gamma \cdot D_b(\mu'_*, \nu'_*) \leq D_b(\mu, \nu)$, and

$$|\varphi(\mu) - \varphi(\nu)| \leq \gamma \cdot [\varphi'(\mu'_*) - \varphi'(\nu'_*)] \leq \gamma \cdot D_b(\mu'_*, \nu'_*) \leq D_b(\mu, \nu)$$

where the second inequality is from induction.

Now we turn to the proof of $D_b(\mu, \nu) \leq D_l(\mu, \nu)$. We will achieve this by showing that the symmetric relations $R_\varepsilon = \{(\mu, \nu) \mid D_l(\mu, \nu) \leq \varepsilon\}$, where $\varepsilon \geq 0$, constitute an approximate bisimulation. Let $\mu R_\varepsilon \nu$ for some $\varepsilon \geq 0$. First, for any $\mathcal{B} \subseteq 2^{AP}$ we have

$$\left| \sum_{A \in \mathcal{B}} \mu(A) - \sum_{A \in \mathcal{B}} \nu(A) \right| = |\mathcal{B}(\mu) - \mathcal{B}(\nu)| \leq D_l(\mu, \nu) \leq \varepsilon.$$

Thus $d_{AP}(\mu, \nu) \leq \varepsilon$ as well. Now suppose $\mu \xrightarrow{a} \mu'$ for some μ'. We have to show that there is some ν' with $\nu \xrightarrow{a} \nu'$ and $D_l(\mu', \nu') \leq \varepsilon/\gamma$. Consider the set

$$\mathcal{K} = \{\omega \in Dist(S) \mid \nu \xrightarrow{a} \omega \text{ and } D_l(\mu', \omega) > \varepsilon/\gamma\}.$$

For each $\omega \in \mathcal{K}$, there must be some φ_ω such that $|\varphi_\omega(\mu') - \varphi_\omega(\omega)| > \varepsilon/\gamma$. As our logic includes the operator $\neg$, we can always assume that $\varphi_\omega(\mu') > \varphi_\omega(\omega) + \varepsilon/\gamma$. Let $p = \sup_{\omega \in \mathcal{K}} \varphi_\omega(\mu')$. Let

$$\varphi'_\omega = \varphi_\omega \oplus [p - \varphi_\omega(\mu')], \quad \varphi' = \bigwedge_{\omega \in \mathcal{K}} \varphi'_\omega, \quad \text{and} \quad \varphi = \langle a \rangle \varphi'.$$

Then from the assumption that $D_l(\mu, \nu) \leq \varepsilon$, we have $|\varphi(\mu) - \varphi(\nu)| \leq \varepsilon$. Furthermore, we check that for any $\omega \in \mathcal{K}$,

$$\varphi'_\omega(\mu') = \varphi_\omega(\mu') \oplus [p - \varphi_\omega(\mu')] = p.$$

Thus $\varphi(\mu) \geq \gamma \cdot \varphi'(\mu') = \gamma \cdot p$.

Let ν' be the distribution such that $\nu \xrightarrow{a} \nu'$ and $\varphi(\nu) = \gamma \cdot \varphi'(\nu')$. We are going to show that $\nu' \notin \mathcal{K}$, and then $D_l(\mu', \nu') \leq \varepsilon/\gamma$ as required. For this purpose, assume conversely that $\nu' \in \mathcal{K}$. Then

$$\varphi(\nu) = \gamma \cdot \varphi'(\nu') \leq \gamma \cdot \varphi'_{\nu'}(\nu') \leq \gamma \cdot [\varphi_{\nu'}(\nu') + p - \varphi_{\nu'}(\mu')]$$
$$< \gamma \cdot p - \varepsilon \leq \varphi(\mu) - \varepsilon,$$

contradicting the fact that $|\varphi(\mu) - \varphi(\nu)| \leq \varepsilon$.

We have proven that $\{R_\varepsilon \mid \varepsilon \geq 0\}$ is an approximate bisimulation. Thus $\mu \sim_\varepsilon \nu$, and so $D_b(\mu, \nu) \leq \varepsilon$, whenever $D_l(\mu, \nu) \leq \varepsilon$. So we have $D_b(\mu, \nu) \leq D_l(\mu, \nu)$ from the arbitrariness of ε. $\qquad\square$

5.3 A Fixed Point-Based Approach

In the following, we denote by $\mathcal{M}$ the set of pseudometrics over $Dist(S)$. Denote by $\mathbf{0}$ the zero pseudometric which assigns 0 to each pair of distributions. For any $d, d' \in \mathcal{M}$, we write $d \leq d'$ if $d(\mu, \nu) \leq d'(\mu, \nu)$ for any μ and ν. Obviously $\leq$ is a partial order, and $(\mathcal{M}, \leq)$ is a complete lattice.

Definition 14. *Let $\mathcal{A} = (S, Act, \to, L, \alpha)$ be an input enabled probabilistic automaton. We define the function $F : \mathcal{M} \to \mathcal{M}$ as follows. For any $\mu, \nu \in Dist(S)$,*

$$F(d)(\mu, \nu) = \max_{a \in Act} \{\ d_{AP}(\mu, \nu),$$
$$\sup_{\mu \xrightarrow{a} \mu'} \inf_{\nu \xrightarrow{a} \nu'} \gamma \cdot d(\mu', \nu'),\ \sup_{\nu \xrightarrow{a} \nu'} \inf_{\mu \xrightarrow{a} \mu'} \gamma \cdot d(\mu', \nu')\}.$$

Then, F is monotonic with respect to $\leq$, and by Knaster-Tarski theorem, F has a least fixed point, denoted $D_f^{\mathcal{A}}$, given by

$$D_f^{\mathcal{A}} = \bigvee_{n=0}^{\infty} F^n(\mathbf{0})\ .$$

Once again, the fixed point-based distance for a general probabilistic automaton can be defined in terms of the input enabled extension; that is, $D_f^{\mathcal{A}}(\mu, \nu) := D_f^{\mathcal{A}_\perp}(\mu, \nu)$. We always omit the superscripts for simplicity.

Similar to Lemma 5, we can show that the supremum (resp. infimum) in Definiton 14 can be replaced by maximum (resp. minimum). Now we show that D_f coincides with D_b.

Theorem 5. $D_f = D_b$.

5.4 Comparison with State-Based Metrics

In this section, we prove that our distribution-based bisimulation metric is lower bounded by the state-based game bisimulation metrics [8] for MDPs. This game bisimulation metric is particularly attractive as it preserves probabilistic reachability, long-run, and discounted average behaviors [5]. We first recall the definition of state-based game bisimulation metrics [8] for MDPs:

Definition 15. *Given $\mu, \nu \in Dist(S)$, $\mu \otimes \nu$ is defined as the set of* weight *functions $\lambda : S \times S \to [0, 1]$ such that for any $s, t \in S$,*

$$\sum_{s \in S} \lambda(s, t) = \nu(t) \quad and \quad \sum_{t \in S} \lambda(s, t) = \mu(s).$$

Given a metric d defined on S, we lift it to $Dist(S)$ by defining

$$d(\mu, \nu) = \inf_{\lambda \in \mu \otimes \nu} \left(\sum_{s, t \in S} \lambda(s, t) \cdot d(s, t) \right).$$

Actually the infimum in the above definition is attainable.

Definition 16. *We define the function $f : \mathcal{M} \to \mathcal{M}$ as follows. For any $s, t \in S$,*

$$f(d)(s, t) = \max_{a \in Act} \left\{ 1 - \delta_{L(s), L(t)}, \sup_{s \xrightarrow{a}_P \mu} \inf_{t \xrightarrow{a}_P \nu} \gamma \cdot d(\mu, \nu), \sup_{t \xrightarrow{a}_P \nu} \inf_{s \xrightarrow{a}_P \mu} \gamma \cdot d(\mu, \nu) \right\}$$

where $\delta_{L(s), L(t)} = 1$ if $L(s) = L(t)$, and 0 otherwise. We take $\inf \emptyset = 1$ and $\sup \emptyset = 0$. Again, f is monotonic with respect to $\leq$, and by Knaster-Tarski theorem, F has a least fixed point, denoted d_f, given by

$$d_f = \bigvee_{n=0}^{\infty} f^n(\mathbf{0}).$$

Now we can prove the following theorem which may be regarded as the quantitative extension of Lemma 1.

Theorem 6. *Let $\mathcal{A}$ be a probabilistic automaton. Then $D_f \leq d_f$.*

Example 2. Consider Fig. 1, and assume $\varepsilon_1 \geq \varepsilon_2 > 0$. Applying the definition of D_b, it is easy to check that $D_b(\delta_q, \delta_{q'}) = 0.5(\varepsilon_1 - \varepsilon_2)\gamma$. By our results, we have $D_l(\delta_q, \delta_{q'}) = D_f(\delta_q, \delta_{q'}) = D_b(\delta_q, \delta_{q'})$. Note that for the discounting case $\gamma < 1$, difference far in the future will have less influence in the distance.

We further compute the distance under state-based bisimulation metrics (see [17] for example). Assume that $\gamma = 1$. One first compute the distance between r_1 and r' being $\frac{1}{6} + \varepsilon_1$, between r_2 and r' being $\frac{1}{6} + \varepsilon_2$. Then, the state-based bisimulation metric between q and q' is $\frac{1}{6} + 0.5(\varepsilon_1 + \varepsilon_2)$, which can be obtained by lifting the state-based metrics.

5.5 Comparison with Equivalence Metric

Note that we can easily extend the equivalence relation defined in Definition 5 to a notion of equivalence metric:

Definition 17 (Equivalence Metric). *Let $\mathcal{A}_i = (S_i, Act_i, \rightarrow_i, L_i, \alpha_i)$ with $i = 1, 2$ be two reactive automata with $Act_1 = Act_2 =: Act$, and $F_i = \{s \in S_i \mid L(s) = AP\}$ the set of final states for $\mathcal{A}_i$. We say $\mathcal{A}_1$ and $\mathcal{A}_2$ are ε-equivalent, denoted $\mathcal{A}_1 \sim_\varepsilon^d \mathcal{A}_2$, if for any input word $w = a_1 a_2 \ldots a_n$, $|\mathcal{A}_1(w) - \mathcal{A}_2(w)| \leq \varepsilon$. Furthermore, the equivalence distance between $\mathcal{A}_1$ and $\mathcal{A}_2$ is defined by $D_d(\mathcal{A}_1, \mathcal{A}_2) := \inf\{\varepsilon \geq 0 \mid \mathcal{A}_1 \sim_\varepsilon^d \mathcal{A}_2\}$.*

Now we show that for reactive automata, the equivalence metric D_d coincide with our undiscounted bisimulation metric D_b, which may be regarded as a quantitative extension of Lemma 2.

Proposition 1. *Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two reactive automata with the same set of actions Act. Let the discount factor $\gamma = 1$. Then $D_d(\mathcal{A}_1, \mathcal{A}_2) = D_b(\alpha_1, \alpha_2)$ where D_b is defined in the direct sum of $\mathcal{A}_1$ and $\mathcal{A}_2$.*

Proof. We first show that $D_d(\mathcal{A}_1, \mathcal{A}_2) \leq D_b(\alpha_1, \alpha_2)$. For each input word $w = a_1 a_2 \ldots a_n$, it is easy to check that $\mathcal{A}_i(w) = \varphi(\alpha_i)$ where $\varphi = \langle a_1 \rangle \langle a_2 \rangle \ldots \langle a_n \rangle (F_1 \cup F_2)$. As we have shown that $D_b = D_l$, it holds $|\mathcal{A}_1(w) - \mathcal{A}_2(w)| \leq D_b(\alpha_1, \alpha_2)$, and hence $\mathcal{A}_1 \sim_{D_b(\alpha_1, \alpha_2)}^d \mathcal{A}_2$. Then $D_d(\mathcal{A}_1, \mathcal{A}_2) \leq D_b(\alpha_1, \alpha_2)$ by definition.

Now we turn to the proof of $D_d(\mathcal{A}_1, \mathcal{A}_2) \geq D_b(\alpha_1, \alpha_2)$. First we show that

$$R_\varepsilon = \{(\mu, \nu) \mid \mu \in Dist(S_1), \nu \in Dist(S_2), \mathcal{A}_1^\mu \sim_\varepsilon^d \mathcal{A}_2^\nu\}$$

is an approximate bisimulation. Here for a probabilistic automaton $\mathcal{A}$, we denote by $\mathcal{A}^\mu$ the automaton which is the same as $\mathcal{A}$ except that the initial distribution is replaced by μ. Let $\mu R_\varepsilon \nu$. Since $L(s) \in \{\emptyset, AP\}$ for all $s \in S_1 \cup S_2$, we have $\mu(AP) + \mu(\emptyset) = \nu(AP) + \nu(\emptyset) = 1$. Thus

$$d_{AP}(\mu, \nu) = |\mu(AP) - \nu(AP)| = |\mu(F_1) - \nu(F_2)|.$$

Note that $\mu(F_1) = \mathcal{A}_1^\mu(e)$ and $\nu(F_2) = \mathcal{A}_2^\nu(e)$, where e is the empty string. Then $d_{AP}(\mu, \nu) = |\mathcal{A}_1^\mu(e) - \mathcal{A}_2^\nu(e)| \leq \varepsilon$.

Let $\mu \xrightarrow{a} \mu'$ and $\nu \xrightarrow{a} \nu'$. We need to show $\mu' R_\varepsilon \nu'$, that is, $\mathcal{A}_1^{\mu'} \sim_\varepsilon^d \mathcal{A}_2^{\nu'}$. For any $w \in Act^*$, note that $\mathcal{A}_1^{\mu'}(w) = \mathcal{A}_1^\mu(aw)$. Then

$$|\mathcal{A}_1^{\mu'}(w) - \mathcal{A}_2^{\nu'}(w)| = |\mathcal{A}_1^\mu(aw) - \mathcal{A}_2^\nu(aw)| \leq \varepsilon,$$

and hence $\mathcal{A}_1^{\mu'} \sim_\varepsilon^d \mathcal{A}_2^{\nu'}$ as required.

Having proven that R_ε is an approximate bisimulation, we know $\mathcal{A}_1 \sim_\varepsilon^d \mathcal{A}_2$ implies $\alpha_1 \sim_\varepsilon \alpha_2$. Thus $D_d(\mathcal{A}_1, \mathcal{A}_2) = \inf\{\varepsilon \mid \mathcal{A}_1 \sim_\varepsilon^d \mathcal{A}_2\} \geq \inf\{\alpha_1 \sim_\varepsilon \alpha_2\} = D_b(\alpha_1, \alpha_2)$. $\square$

6 Discussion and Future Work

In this paper, we considered Segala's automata, and proposed a novel notion of bisimulation by joining the existing notions of equivalence and bisimilarities. We have demonstrated the utility of our definition by studying distribution-based bisimulation metrics, which have been extensively studied for MDPs.

As future work we would like to identify further solutions and techniques developed in one area that could inspire solutions for the corresponding problems in the other area. This includes for instance decision algorithm developed for equivalence checking [31, 25], extensions to simulations, and compositional verification for probabilistic automata.

References

1. G. Bacci, G. Bacci, K. G. Larsen, and R. Mardare. On-the-Fly Exact Computation of Bisimilarity Distances. In N. Piterman and S. A. Smolka, editors, *TACAS*, volume 7795 of *Lecture Notes in Computer Science*, pages 1–15. Springer, 2013.
2. C. Baier, B. Engelen, and M. E. Majster-Cederbaum. Deciding Bisimilarity and Similarity for Probabilistic Processes. *J. Comput. Syst. Sci.*, 60(1):187–231, 2000.
3. R. Bellman. *Dynamic Programming*. Princeton University Press, 1957.
4. S. Cattani and R. Segala. Decision Algorithms for Probabilistic Bisimulation. In L. Brim, P. Jancar, M. Kretínský, and A. Kucera, editors, *CONCUR*, volume 2421 of *Lecture Notes in Computer Science*, pages 371–385. Springer, 2002.
5. K. Chatterjee, L. de Alfaro, R. Majumdar, and V. Raman. Algorithms for game metrics (full version). *Logical Methods in Computer Science*, 6(3), 2010.
6. D. Chen, F. van Breugel, and J. Worrell. On the Complexity of Computing Probabilistic Bisimilarity. In L. Birkedal, editor, *FoSSaCS*, volume 7213 of *Lecture Notes in Computer Science*, pages 437–451. Springer, 2012.
7. G. Comanici, P. Panangaden, and D. Precup. On-the-Fly Algorithms for Bisimulation Metrics. In *QEST*, pages 94–103. IEEE Computer Society, 2012.
8. L. de Alfaro, R. Majumdar, V. Raman, and M. Stoelinga. Game relations and metrics. In *LICS*, pages 99–108. IEEE Computer Society, 2007.
9. Y. Deng, R. J. van Glabbeek, M. Hennessy, and C. Morgan. Testing finitary probabilistic processes. In M. Bravetti and G. Zavattaro, editors, *CONCUR*, volume 5710 of *Lecture Notes in Computer Science*, pages 274–288. Springer, 2009.
10. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for Labeled Markov Systems. In J. C. M. Baeten and S. Mauw, editors, *CONCUR*, volume 1664 of *Lecture Notes in Computer Science*, pages 258–273. Springer, 1999.
11. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labelled markov processes. *Theor. Comput. Sci.*, 318(3):323–354, 2004.
12. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Weak bisimulation is sound and complete for pCTL*. *Inf. Comput.*, 208(2):203–219, 2010.
13. J. Desharnais, F. Laviolette, and M. Tracol. Approximate Analysis of Probabilistic Processes: Logic, Simulation and Games. In *QEST*, pages 264–273. IEEE Computer Society, 2008.
14. L. Doyen, T. A. Henzinger, and J.-F. Raskin. Equivalence of Labeled Markov Chains. *Int. J. Found. Comput. Sci.*, 19(3):549–563, 2008.
15. C. Eisentraut, H. Hermanns, and L. Zhang. On Probabilistic Automata in Continuous Time. In *LICS*, pages 342–351. IEEE Computer Society, 2010.
16. Y. Feng and L. Zhang. When equivalence and bisimulation join forces in probabilistic automata. *CoRR*, abs/1311.3396, 2013.
17. N. Ferns, P. Panangaden, and D. Precup. Bisimulation Metrics for Continuous Markov Decision Processes. *SIAM J. Comput.*, 40(6):1662–1714, 2011.
18. H. Fu. Computing Game Metrics on Markov Decision Processes. In A. Czumaj, K. Mehlhorn, A. M. Pitts, and R. Wattenhofer, editors, *ICALP (2)*, volume 7392 of *Lecture Notes in Computer Science*, pages 227–238. Springer, 2012.

19. A. Giacalone, C. Jou, and S. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *IFIP TC2 Working Conference on Programming Concepts and Methods*, page 443458. North-Holland, 1990.
20. E. M. Hahn, Y. Li, S. Schewe, A. Turrini, and Z. Lijun. IsCASMC: A web-based probabilistic model checker. In *FM*, 2014. To appear.
21. M. Hennessy. Exploring probabilistic bisimulations, part I. *Formal Asp. Comput.*, 24(4-6):749–768, 2012.
22. H. Hermanns, A. Parma, R. Segala, B. Wachter, and L. Zhang. Probabilistic Logical Characterization. *Inf. Comput.*, 209(2):154–172, 2011.
23. H. Hermanns and A. Turrini. Deciding Probabilistic Automata Weak Bisimulation in Polynomial Time. In D. D'Souza, T. Kavitha, and J. Radhakrishnan, editors, *FSTTCS*, volume 18 of *LIPIcs*, pages 435–447. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2012.
24. J.-P. Katoen, I. S. Zapreev, E. M. Hahn, H. Hermanns, and D. N. Jansen. The ins and outs of the probabilistic model checker mrmc. *Perform. Eval.*, 68(2):90–104, 2011.
25. S. Kiefer, A. S. Murawski, J. Ouaknine, B. Wachter, and J. Worrell. Language Equivalence for Probabilistic Automata. In G. Gopalakrishnan and S. Qadeer, editors, *CAV*, volume 6806 of *Lecture Notes in Computer Science*, pages 526–540. Springer, 2011.
26. S. Kiefer, A. S. Murawski, J. Ouaknine, B. Wachter, and J. Worrell. On the Complexity of the Equivalence Problem for Probabilistic Automata. In L. Birkedal, editor, *FoSSaCS*, volume 7213 of *Lecture Notes in Computer Science*, pages 467–481. Springer, 2012.
27. M. Z. Kwiatkowska, G. Norman, and D. Parker. PRISM 4.0: Verification of Probabilistic Real-Time Systems. In G. Gopalakrishnan and S. Qadeer, editors, *CAV*, volume 6806 of *Lecture Notes in Computer Science*, pages 585–591. Springer, 2011.
28. A. Parma and R. Segala. Logical Characterizations of Bisimulations for Discrete Probabilistic Systems. In H. Seidl, editor, *FoSSaCS*, volume 4423 of *Lecture Notes in Computer Science*, pages 287–301. Springer, 2007.
29. M. Rabin. Probabilistic automata. *Information and Control*, 6(3):230245, 1963.
30. R. Segala. *Modeling and Verification of Randomized Distributed Realtime Systems*. PhD thesis, MIT, 1995.
31. W. Tzeng. A polynomial-time algorithm for the equivalence of probabilistic automata. *SIAM Journal on Computing*, 21(2):216227, 1992.
32. F. van Breugel, B. Sharma, and J. Worrell. Approximating a Behavioural Pseudometric Without Discount for Probabilistic Systems. In H. Seidl, editor, *FoSSaCS*, volume 4423 of *Lecture Notes in Computer Science*, pages 123–137. Springer, 2007.
33. M. Ying. *Topology in Process Calculus: Approximate Correctness and Infinite Evolution of Concurrent Programs*. Springer-Verlag, New York, 2001.
34. M. Ying. Bisimulation indexes and their applications. *Theoretical Computer Science*, 275:1–68, 2002.
35. M. Ying and M. Wirsing. Approximate bisimilarity. In *Algebraic Methodology and Software Technology*, pages 309–322. Springer-Verlag, Berlin, 2000.