Scientific Annals of Computer Science vol. 32 (2), 2022, pp. 211–254
doi: 10.7561/SACS.2022.2.211

Behavioural Pseudometrics for Nondeterministic Probabilistic Systems

Wenjie Du[1], Yuxin Deng[2], and Daniel Gebler[3]

Abstract

For the model of probabilistic labelled transition systems that allow for the co-existence of nondeterminism and probabilities, we present two notions of bisimulation metrics: one is state-based and the other is distribution-based. We provide a sound and complete modal characterisation for each of them, using real-valued modal logics based on Hennessy-Milner logic. The logic for characterising the state-based metric is much simpler than an earlier logic proposed by Desharnais et al. as it uses only two non-expansive operators rather than the general class of non-expansive operators. For the kernels of the two metrics, which correspond to two notions of bisimilarity, we give a comprehensive comparison with some typical distribution-based bisimilarities in the literature.

Keywords: Probabilistic labelled transition systems; Behavioral pseudometrics; Real-valued modal logics

1 Introduction

Bisimulation is an important proof technique for establishing behavioural equivalences of concurrent systems. In probabilistic concurrency theory,

[1]Shanghai Normal University, China, Email: wenjiedu@shnu.edu.cn
[2]Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, China, Email: yxdeng@sei.ecnu.edu.cn, Corresponding author
[3]VU University Amsterdam, The Netherlands, Email: e.d.gebler@vu.nl

there are roughly two kinds of bisimulations: one is state-based that is directly defined over states and then lifted to distributions, and the other is distribution-based as it is a relation between distributions. The former is originally defined in [37] to represent a branching time semantics; the latter, as defined in [31, 24, 14], represents a linear time semantics.

In correspondence with those bisimulations, there are two notions of behavioural pseudometrics (simply called metrics in the current work). They are more robust ways of formalising behavioural similarity between formal systems than bisimulations because, particularly in the probabilistic setting, bisimulations are too sensitive to probabilities (a very small perturbation of the probabilities would render two systems non-bisimilar). A metric gives a quantitative measure of the distance between two systems and distance 0 usually means that the two systems are bisimilar. A logical characterisation of the state-based bisimulation metric for labelled Markov processes is given in [17]. For a more general model of labelled concurrent Markov chains (LCMCs) that allow for the co-existence of nondeterminism and probabilities, a weak bisimulation metric is proposed in [18]. Its logical characterisation uses formulae like $h \circ f$, which is the composition of formula f with any non-expansive operator h on the interval $[0,1]$, i.e. $|h(x) - h(y)| \leq |x - y|$ for any $x, y \in [0,1]$. A natural question then arises: instead of the general class of non-expansive operators, is it possible to use only a few simple non-expansive operators without losing the capability of characterising the bisimulation metric?

In the current work, we give a positive answer to the above question. For simplicity of presentation, we focus on strong bisimulation metrics. But the proof idea can be generalised to the weak case. We work in the framework of probabilistic labelled transition systems (pLTSs) that are essentially the same as LCMCs, so the interplay of nondeterminism and probabilities is allowed. We provide a modal characterisation of the state-based bisimulation metric closely in line with the classical Hennessy-Milner logic (HML) [30]. Our variant of HML makes use of state formulae and distribution formulae, which are formulae evaluated at states and distributions, respectively, and yield success probabilities. We use merely two non-expansive operators: negation ($\neg\phi$) and testing ($\phi \ominus p$). Negation is self-explanatory and the testing operator checks if a state satisfies a property with certain threshold probability. More precisely, if state s satisfies formula ϕ with probability q, then it satisfies $\neg\phi$ with probability $1-q$, and satisfies $\phi \ominus p$ with probability $q - p$ if $q > p$ and 0 otherwise. In other words, we do not need the general class of non-expansive operators because negation and testing, together with other modalities inherited from the classical HML, are expressive enough to characterise bisimulation metrics[4]. As regards to the characterisation of distribution-based bisimulation metric, we drop state formulae and use distribution formulae only. In addition, we show that the distribution-based metric is a lower bound of the state-based metric when the latter is lifted to distributions.

The kernels of the two metrics generate two notions of bisimilarity: one is state-based and the other is distribution-based. The state-based bisimilarity is widely accepted by the community of probabilistic concurrency theory, and it admits elegant characterisations from metric, logical, and algorithmic perspectives [11]. On the contrary, there is no general agreement on what is a good notion of distribution-based bisimilarity. We compare the two bisimilarities induced by our metrics with some typical notions of distribution-based bisimilarities proposed in the literature. Our distribution-based bisimilarity turns out to coincide with the one defined in [24] and they constitute the coarsest bisimilarity for distributions.

The rest of this paper is organised as follows. Section 2 provides some basic concepts on pLTSs. Section 3 defines a two-sorted modal logic that leads to a sound and complete characterisation of the state-based bisimulation metric. Section 4 gives a similar characterisation for the distribution-based bisimulation metric. In Section 5 we compare the two metrics discussed in the previous two sections. In Section 6 we compare the two bisimilarities generated by the two metrics with some distribution-based bisimilarities that appeared in the literature. In Section 7 we review some related work. Finally, we conclude in Section 8.

An extended abstract of this paper has appeared as [19]. All the proofs omitted there are now given in great detail.

2 Preliminaries

Let S be a countable set. A *(discrete) probability subdistribution* over S is a function $\Delta : S \to [0,1]$ with $\sum_{s\in S} \Delta(s) \leq 1$. It is a *(full) distribution* if $\sum_{s\in S} \Delta(s) = 1$. Its *support*, written $\lceil \Delta \rceil$, is defined to be the set $\{s \in S \mid$

[4]Notice that we do not claim that negation and testing operators, plus some constant functions, suffice to approximate all the non-expansive operators on the unit interval. That claim is too strong to be true. For example, the operator $f(x) = \frac{1}{2}x$ cannot be represented by those operators.

$\Delta(s) > 0\}$. Let $\mathcal{D}_{sub}(S)$ (resp. $\mathcal{D}(S)$) denote the set of all subdistributions (resp. distributions) over S. We use ε to stand for the empty subdistribution, that is $\varepsilon(s) = 0$ for any $s \in S$. We write $\overline{s}$ for the point distribution, satisfying $\overline{s}(t) = 1$ if $t = s$, and 0 otherwise. The *total mass* of subdistribution Δ, written $|\Delta|$, is defined as $\sum_{s\in S} \Delta(s)$. A *weight function*[5] $\omega \in \mathcal{D}(S \times S)$ for $(\Delta, \Theta) \in \mathcal{D}(S)\times\mathcal{D}(S)$ is given if it satisfies the two conditions: $\sum_{t\in S} \omega(s,t) = \Delta(s)$ and $\sum_{s\in S} \omega(s,t) = \Theta(t)$ for all $s, t \in S$. We denote the set of all weight functions for (Δ, Θ) by $\Omega(\Delta, \Theta)$. If $\{\Delta_i\}_{i\in I}$ is a finite collection of subdistributions and $\{p_i\}_{i\in I}$ is a collection of probabilities with $\sum_{i\in I} p_i \leq 1$, then $\sum_{i\in I} p_i \cdot \Delta_i$ is also a subdistribution with $(\sum_{i\in I} p_i \cdot \Delta_i)(s) = \sum_{i\in I} p_i \cdot \Delta_i(s)$ for any $s \in S$.

A *metric* d over a space $\mathtt{S}$ is a distance function $d : \mathtt{S} \times \mathtt{S} \to \mathbb{R}_{\geq 0}$ satisfying: (i) $d(s,t) = 0$ iff $s = t$ (isolation), (ii) $d(s,t) = d(t,s)$ (symmetry), (iii) $d(s,t) \leq d(s,u) + d(u,t)$ (triangle inequality), for any $s, t, u \in \mathtt{S}$. If we replace (i) with $d(s,s) = 0$, we obtain a *pseudometric*. In this article we are interested in pseudometrics because two distinct states can still be at distance zero if their behaviour is similar. But for simplicity, we often use the term metrics though we really mean pseudometrics. Let $c \in \mathbb{R}_{\geq 0}$ be a positive real number. A metric d over $\mathtt{S}$ is c-bounded if $d(s,t) \leq c$ for any $s, t \in \mathtt{S}$. In the rest of this article, we restrict ourselves to 1-bounded metrics.

Let $d\colon \mathtt{S} \times \mathtt{S} \to [0,1]$ be a metric over $\mathtt{S}$. We can lift it to be a metric over $\mathcal{D}(\mathtt{S})$ by using the *Kantorowich metric* [34] $K(d)\colon \mathcal{D}(\mathtt{S}) \times \mathcal{D}(\mathtt{S}) \to [0,1]$ defined via a linear programming problem as follows:

$$K(d)(\Delta, \Theta) = \min_{\omega\in\Omega(\Delta,\Theta)} \sum_{s,t\in\mathtt{S}} d(s,t) \cdot \omega(s,t) \tag{1}$$

for $\Delta, \Theta \in \mathcal{D}(\mathtt{S})$. The dual of the above linear programming problem is the following

$$\max\ \textstyle\sum_{s\in\mathtt{S}} (\Delta(s) - \Theta(s))x_s, \text{ subject to } \quad \begin{array}{l} 0 \leq x_s \leq 1 \\ \forall s, t \in \mathtt{S}\colon\ x_s - x_t \leq d(s,t)\ . \end{array} \tag{2}$$

The duality theorem in linear programming guarantees that both problems have the same optimal value.

Let $\hat{d}\colon \mathcal{D}(\mathtt{S}) \times \mathcal{D}(\mathtt{S}) \to [0,1]$ be a metric over $\mathcal{D}(\mathtt{S})$. We can lift it to be a metric over the powerset of $\mathcal{D}(\mathtt{S})$, written $\mathcal{P}(\mathcal{D}(\mathtt{S}))$, in the standard way by using the *Hausdorff metric* $H(\hat{d})\colon \mathcal{P}(\mathcal{D}(\mathtt{S})) \times \mathcal{P}(\mathcal{D}(\mathtt{S})) \to [0,1]$ given as

[5] A weight function is also known as a coupling in some literature [46].

follows

$$H(\hat{d})(\Pi_1, \Pi_2) = \max\{\sup_{\Delta \in \Pi_1} \inf_{\Theta \in \Pi_2} \hat{d}(\Delta, \Theta), \quad \sup_{\Theta \in \Pi_2} \inf_{\Delta \in \Pi_1} \hat{d}(\Theta, \Delta)\}$$

for all $\Pi_1, \Pi_2 \subseteq \mathcal{D}(\mathtt{S})$, whereby $\inf \emptyset = 1$ and $\sup \emptyset = 0$.

Probabilistic labelled transition systems (pLTSs) generalise labelled transition systems by allowing for probabilistic choices in the transitions. They are essentially *simple probabilistic automata* [42] without initial states.

Definition 2.1 *A* probabilistic labelled transition system *is a triple* $(S, A, \rightarrow)$, *where* S *is a countable set of states,* A *is a countable set of actions, and the relation* $\rightarrow \subseteq S \times A \times \mathcal{D}(S)$ *is a transition relation.*

We write $s \xrightarrow{a} \Delta$ for $(s, a, \Delta) \in \rightarrow$ and $s \not\xrightarrow{a}$ if there is no Δ satisfying $s \xrightarrow{a} \Delta$. Let $der(s, a) = \{\Delta \mid s \xrightarrow{a} \Delta\}$ be the set of all a-successor distributions of s. A pLTS is *image-finite* (resp. *deterministic* or *reactive*) if for any state s and action a the set $der(s, a)$ is finite (resp. has at most one element). In the current work, we focus on image-finite pLTSs with finitely many states.

3 State-Based Bisimulation Metrics

We consider the complete lattice $([0,1]^{S \times S}, \sqsubseteq)$ defined by $d \sqsubseteq d'$ iff $d(s,t) \leq d'(s,t)$, for all $s, t \in S$. For any $D \subseteq [0,1]^{S \times S}$ the least upper bound is given by $(\bigsqcup D)(s,t) = \sup_{d \in D} d(s,t)$, and the greatest lower bound is given by $(\bigsqcap D)(s,t) = \inf_{d \in D} d(s,t)$ for all $s, t \in S$. The bottom element $\mathbf{0}$ is the constant zero function $\mathbf{0}(s,t) = 0$ and the top element $\mathbf{1}$ is the constant one function $\mathbf{1}(s,t) = 1$ for all $s, t \in S$.

Definition 3.1 *A* 1*-bounded metric* d *on* S *is a* state-based bisimulation metric *if for all* $s, t \in S$ *with* $d(s,t) < 1$, *whenever* $s \xrightarrow{a} \Delta$ *then there exists some* $t \xrightarrow{a} \Delta'$ *with* $K(d)(\Delta, \Delta') \leq d(s,t)$.

The smallest (wrt. $\sqsubseteq$) state-based bisimulation metric, denoted by $\mathbf{d}_s$, is called *state-based bisimilarity metric.* Its kernel is the state-based bisimilarity as defined in [37, 42]. Note that $\mathbf{0}$ does not satisfy Definition 3.1 for general pLTSs, thus is not a state-based bisimulation metric in general.

Example 3.1 *Let us calculate the distance between states* s *and* t *in Figure 1. Firstly, it is clear that* $\mathbf{d}_s(s_4, t_5) = 0$ *because both* s_4 *and* t_5 *are deadlock*

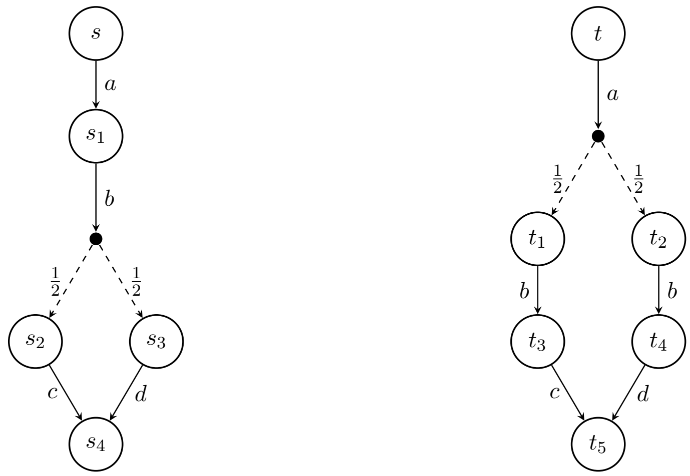

Figure 1: $\mathbf{d}_s(s,t) = \frac{1}{2}$

states. It follows that $\mathbf{d}_s(s_2,t_3) = 0$ *because* s_2 *has a unique c-transition to* s_4 *and* t_3 *has a unique c-transition to* t_5*. On the contrary,* $\mathbf{d}_s(s_3,t_3) = 1$ *because the two states* s_3 *and* t_3 *perform completely different actions. Secondly, let* $\Delta = \frac{1}{2}\overline{s_2} + \frac{1}{2}\overline{s_3}$ *and* $\Theta = \overline{t_3}$*. We see that*

$$\begin{array}{rcl} K(\mathbf{d}_s)(\Delta,\Theta) & = & \min_{\omega\in\Omega(\Delta,\Theta)} \mathbf{d}_s(s_2,t_3)\cdot\omega(s_2,t_3) + \mathbf{d}_s(s_3,t_3)\cdot\omega(s_3,t_3) \\ & = & \min_{\omega\in\Omega(\Delta,\Theta)} 0\cdot\omega(s_2,t_3) + 1\cdot\omega(s_3,t_3) \\ & = & 0\cdot\frac{1}{2} + 1\cdot\frac{1}{2} \\ & = & \frac{1}{2} \end{array}$$

Here the only weight function is ω *with* $\omega(s_2,t_3) = \omega(s_3,t_3) = \frac{1}{2}$*. It follows that* $\mathbf{d}_s(s_1,t_1) = \frac{1}{2}$*. Similarly, we get* $\mathbf{d}_s(s_1,t_2) = \frac{1}{2}$*. Then it is not difficult to see that*

$$K(\mathbf{d}_s)(\overline{s_1}, \frac{1}{2}\overline{t_1} + \frac{1}{2}\overline{t_2}) \;=\; \mathbf{d}_s(s_1,t_1)\cdot\frac{1}{2} + \mathbf{d}_s(s_1,t_2)\cdot\frac{1}{2} \;=\; \frac{1}{2}$$

from which we finally obtain $\mathbf{d}_s(s,t) = \frac{1}{2}$*.*

The above coinductively defined bisimilarity metric can be reformulated as a fixed point of a monotone functional operator. Let us define the functional operator $F_{\mathrm{s}}\colon [0,1]^{S\times S} \to [0,1]^{S\times S}$ for $d\colon S\times S \to [0,1]$ and $s,t\in S$ by

$$F_{\mathrm{s}}(d)(s,t) = \sup_{a\in A}\{H(K(d))(der(s,a), der(t,a))\}\,. \tag{3}$$

It can be shown that $F_{\rm s}$ is monotone and its least fixed point is given by $\bigsqcup d_i$, where $d_0 = \mathbf{0}$ and $d_{i+1} = F_{\rm s}(d_i)$ for all $i \in \mathbb{N}$.

Proposition 3.1 $\mathbf{d}_s$ *is the least fixed point of* $F_{\rm s}$. □

Essentially the same property as Proposition 3.1 has appeared in [18].

Now we proceed by defining a real-valued modal logic based on Hennessy-Milner logic [30], called metric HML, to characterise the bisimilarity metric. It is motivated by [33, 17, 18, 5].

Definition 3.2 *Our metric HML is two-sorted and has the following syntax:*

$$\begin{array}{rl}\varphi ::= & \top \mid \neg\varphi \mid \varphi \ominus p \mid \varphi_1 \wedge \varphi_2 \mid \langle a \rangle \psi \\ \psi ::= & [\varphi] \mid \neg\psi \mid \psi \ominus p \mid \psi_1 \wedge \psi_2\end{array}$$

with $a \in A$ *and* $p \in [0, 1]$.

Let $\mathcal{L}$ denote the set of all metric HML formulae, φ range over the set of all *state formulae* $\mathcal{L}^{\rm S}$, and ψ range over the set of all *distribution formulae* $\mathcal{L}^{\rm D}$. The two kinds of formulae are defined simultaneously. The operator $\varphi \ominus p$ tests if a state passes φ with probability at least p. Each state formula φ immediately induces a distribution formula $[\varphi]$. Sometimes we abbreviate $\langle a \rangle[\varphi]$ as $\langle a \rangle \varphi$. Other operators such as negation, conjunction, and the diamond operator come from the classical HML, but will be given a quantitative interpretation.

Definition 3.3 *A state formula* $\varphi \in \mathcal{L}^{\rm S}$ *evaluates in* $s \in S$ *as follows:*

$$\begin{array}{rcl}[\![\top]\!](s) & = & 1 \\ [\![\neg\varphi]\!](s) & = & 1 - [\![\varphi]\!](s) \\ [\![\varphi \ominus p]\!](s) & = & \max([\![\varphi]\!](s) - p,\ 0) \\ [\![\varphi_1 \wedge \varphi_2]\!](s) & = & \min([\![\varphi_1]\!](s), [\![\varphi_2]\!](s)) \\ [\![\langle a \rangle \psi]\!](s) & = & \max_{s \xrightarrow{a} \Delta} [\![\psi]\!](\Delta)\end{array}$$

with $\max \emptyset = 0$ *and a distribution formula* $\psi \in \mathcal{L}^{\rm D}$ *evaluates in* $\Delta \in \mathcal{D}(S)$ *as follows:*

$$\begin{array}{rcl}[\![[\varphi]]\!](\Delta) & = & \sum_{s \in S} \Delta(s) \cdot [\![\varphi]\!](s) \\ [\![\neg\psi]\!](\Delta) & = & 1 - [\![\psi]\!](\Delta) \\ [\![\psi \ominus p]\!](\Delta) & = & \max([\![\psi]\!](\Delta) - p, 0) \\ [\![\psi_1 \wedge \psi_2]\!](\Delta) & = & \min([\![\psi_1]\!](\Delta), [\![\psi_2]\!](\Delta)).\end{array}$$

We often use constant formulae e.g. $\underline{p}$ for any $p \in [0,1]$ with the semantics $[\![\underline{p}]\!](s) = p$, which is derivable in the above logic by letting $\underline{p} = \top \ominus (1-p)$. Moreover, we write $\varphi \oplus p$ for $\neg((\neg\varphi) \ominus p)$ with the semantics $[\![\varphi \oplus p]\!](s) = \min([\![\varphi]\!](s)+p,\ 1) = 1-\max(1-[\![\varphi]\!](s)-p, 0)$. In the presence of negation and conjunction we can derive disjunction by letting $\varphi_1 \vee \varphi_2$ be $\neg(\neg\varphi_1 \wedge \neg\varphi_2)$. Intuitively, $[\![\varphi]\!](s)$ measures the degree that formula φ is satisfied by state s; similarly for distribution formulae. Therefore, negation is naturally interpreted as complement, conjunction as minimum and disjunction as maximum[6]. The formula $\langle a\rangle\psi$ specifies the property for a state to perform action a and result in a possible distribution to satisfy ψ. In the presence of nondeterminism, from state s there may be several outgoing transitions labelled by the same action a, e.g. $s \xrightarrow{a} \Delta_i$ with $i \in I$. We take the optimal case by taking $[\![\langle a\rangle\psi]\!](s)$ to be the maximal $[\![\psi]\!](\Delta_i)$ when i ranges over I.

The above metric HML induces two natural logical metrics $\mathbf{d}_s^{\mathrm{ls}}$ and $\mathbf{d}_s^{\mathrm{ld}}$ on states and distributions respectively, by letting

$$\begin{array}{rcl}\mathbf{d}_s^{\mathrm{ls}}(s,t) & = & \sup_{\varphi\in\mathcal{L}^{\mathrm{S}}} |[\![\varphi]\!](s) - [\![\varphi]\!](t)| \\ \mathbf{d}_s^{\mathrm{ld}}(\Delta,\Theta) & = & \sup_{\psi\in\mathcal{L}^{\mathrm{D}}} |[\![\psi]\!](\Delta) - [\![\psi]\!](\Theta)|.\end{array}$$

Remark 3.1 *In the above definition, we can also write*

$$\mathbf{d}_s^{\mathrm{ls}}(s,t) = \sup_{\varphi\in\mathcal{L}^{\mathrm{S}}} ([\![\varphi]\!](s) - [\![\varphi]\!](t)) \tag{4}$$

because if $[\![\varphi]\!](s) < [\![\varphi]\!](t)$ *then we can take the negation of* φ *so as to obtain* $|[\![\varphi]\!](s) - [\![\varphi]\!](t)|$.

$$[\![\neg\varphi]\!](s) - [\![\neg\varphi]\!](t) = (1-[\![\varphi]\!](s)) - (1-[\![\varphi]\!](t)) = |[\![\varphi]\!](s) - [\![\varphi]\!](t)|\ .$$

However, this heavily relies on our semantic interpretation of the negation operator, and we decide not to use (4) as a definition. Similarly for $\mathbf{d}_s^{\mathrm{ld}}(\Delta,\Theta)$.

Example 3.2 *Consider the two probabilistic systems depicted in Figure 2. We have the formula* $\varphi = \langle a\rangle\psi$ *where* $\psi = [\langle a\rangle\top] \wedge [\langle b\rangle\top]$ *and would like to know the difference between* s *and* t *given by* φ. *Let*

$$\begin{array}{rcl}\Delta_1 & = & 0.2\cdot\overline{s_1} + 0.8\cdot\overline{s_2} \\ \Delta_2 & = & 0.8\cdot\overline{s_5} + 0.2\cdot\overline{s_6} \\ \Delta_3 & = & 0.5\cdot\overline{s_3} + 0.5\cdot\overline{s_4}\end{array}$$

[6] Since we will compare our logic with that in [18], it is better for our semantic interpretation to be consistent with that in the aforementioned work. In the literature, there are also other ways of interpreting conjunction and disjunction in probabilistic settings, see e.g. [32, 4].

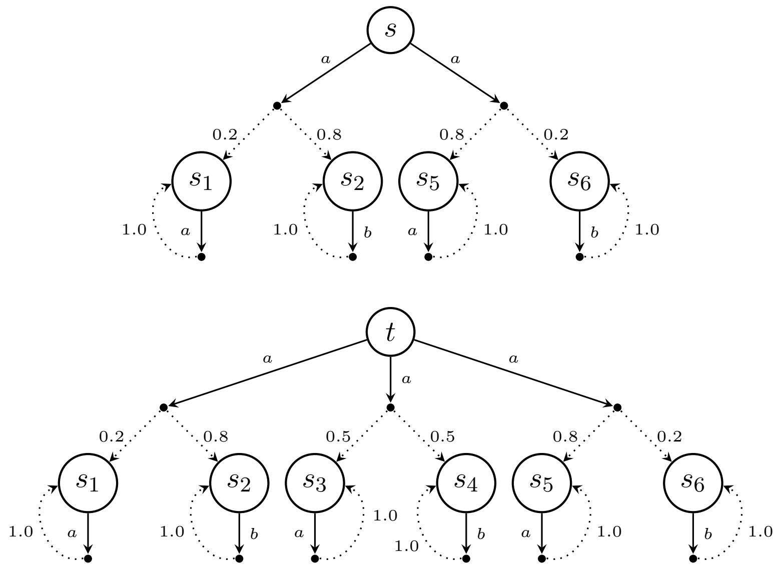

Figure 2: $\mathbf{d}_s^{\mathrm{ls}}(s,t) = 0.3$

Note that $[\![\langle a\rangle\top]\!](s_1) = 1$ *and* $[\![\langle a\rangle\top]\!](s_2) = 0$. *Then*

$$[\![[\langle a\rangle\top]]\!](\Delta_1) = 0.2 \cdot [\![\langle a\rangle\top]\!](s_1) + 0.8 \cdot [\![\langle a\rangle\top]\!](s_2) = 0.2.$$

Similarly, $[\![[\langle b\rangle\top]]\!](\Delta_1) = 0.8$. *It follows that*

$$[\![\psi]\!](\Delta_1) = \min([\![[\langle a\rangle\top]]\!](\Delta_1),\ [\![[\langle b\rangle\top]]\!](\Delta_1)) = 0.2.$$

With similar arguments, we see that $[\![\psi]\!](\Delta_2) = 0.2$ *and* $[\![\psi]\!](\Delta_3) = 0.5$. *Therefore, we can calculate that*

$$\begin{aligned}[\![\varphi]\!](s) &= \max([\![\psi]\!](\Delta_1), [\![\psi]\!](\Delta_2)) = 0.2\\ [\![\varphi]\!](t) &= \max([\![\psi]\!](\Delta_1), [\![\psi]\!](\Delta_2), [\![\psi]\!](\Delta_3)) = 0.5.\end{aligned}$$

So the difference between s *and* t *with respect to* φ *is* $|[\![\varphi]\!](s) - [\![\varphi]\!](t)| = 0.3$. *In fact we also have* $\mathbf{d}_s^{\mathrm{ls}}(s,t) = 0.3$.

In the presence of testing operators in state formulae, one might wonder if the testing operators in distribution formulae can be removed. Unfortunately, this is not the case, as indicated by the following example.

Example 3.3 *At first sight the following two equations seem to be sound.*

$$[\![[\varphi] \ominus p]\!](\Delta) = [\![[\varphi \ominus p]]\!](\Delta) \qquad \text{and} \qquad [\![\psi]\!](\sum_i p_i \Delta_i) = \sum_i p_i([\![\psi]\!](\Delta_i))$$

However, in general they do not hold, as witnessed by the counterexamples below. Let $\varphi = \langle b \rangle \top$, $\psi = [\varphi] \ominus 0.5$ and the distribution Δ_1 be the same as in Example 3.2. Then we have

$$\begin{array}{rcl} [\![[\varphi] \ominus 0.5]\!](\Delta_1) & = & \max([\![[\varphi]]\!](\Delta_1) - 0.5,\ 0) \\ & = & \max(0.2[\![[\langle b \rangle \top]]\!](\overline{s_1}) + 0.8[\![[\langle b \rangle \top]]\!](\overline{s_2}) - 0.5,\ 0) \\ & = & \max(0.2 \cdot 0 + 0.8 \cdot 1 - 0.5,\ 0) \\ & = & 0.3 \end{array}$$

$$\begin{array}{rcl} [\![[\varphi \ominus 0.5]]\!](\Delta_1) & = & 0.2[\![\varphi \ominus 0.5]\!](s_1) + 0.8[\![\varphi \ominus 0.5]\!](s_2) \\ & = & 0.2 \max([\![\varphi]\!](s_1) - 0.5,\ 0) \\ & + & 0.8 \max([\![\varphi]\!](s_2) - 0.5,\ 0) \\ & = & 0.2 \max(0 - 0.5,\ 0) + 0.8 \max(1 - 0.5,\ 0) \\ & = & 0.4 \end{array}$$

$$\begin{array}{rcl} 0.2[\![\psi]\!](\overline{s_1}) + 0.8[\![\psi]\!](\overline{s_2}) & = & 0.2[\![[\varphi] \ominus 0.5]\!](\overline{s_1}) + 0.8[\![[\varphi] \ominus 0.5]\!](\overline{s_2}) \\ & = & 0.2 \max([\![[\varphi]]\!](\overline{s_1}) - 0.5,\ 0) \\ & + & 0.8 \max([\![[\varphi]]\!](\overline{s_2}) - 0.5,\ 0) \\ & = & 0.2 \max(0 - 0.5,\ 0) + 0.8 \max(1 - 0.5,\ 0) \\ & = & 0.4 \end{array}$$

So we see that $[\![[\varphi] \ominus 0.5]\!](\Delta_1) \neq [\![[\varphi \ominus 0.5]]\!](\Delta_1)$ and $[\![\psi]\!](\Delta_1) \neq 0.2[\![\psi]\!](\overline{s_1}) + 0.8[\![\psi]\!](\overline{s_2})$.

It turns out that the logic $\mathcal{L}$ precisely captures the bisimilarity metric $\mathbf{d}_s$: the metric $\mathbf{d}_s^{\mathrm{ls}}$ defined by state formulae coincides with $\mathbf{d}_s$ and the metric $\mathbf{d}_s^{\mathrm{ld}}$ defined by distribution formulae coincides with $K(\mathbf{d}_s)$, the lifted form of $\mathbf{d}_s$.

Theorem 3.1 $\mathbf{d}_s = \mathbf{d}_s^{\mathrm{ls}}$ *and* $K(\mathbf{d}_s) = \mathbf{d}_s^{\mathrm{ld}}$

The two properties in Theorem 3.1 are coupled and should be proved simultaneously because state formulae and distribution formulae are defined reciprocally. The proof is carried out in three steps:

(i) We show $\mathbf{d}_s^{\mathrm{ls}} \sqsubseteq \mathbf{d}_s$ and $\mathbf{d}_s^{\mathrm{ld}} \sqsubseteq K(\mathbf{d}_s)$ simultaneously by structural induction on formulae.

(ii) We establish $K(\mathbf{d}_s^{\rm ls}) \sqsubseteq \mathbf{d}_s^{\rm ld}$ by exploiting the dual form of the Kantorovich metric in (2). Here it is crucial to require the state space of the pLTS under consideration to be finite in order to use binary conjunctions rather than infinitary conjunctions. The negation and testing operators in state formulae play an important role in the proof.

(iii) We verify that $\mathbf{d}_s^{\rm ls}$ is a state-based bisimulation metric and so obtain $\mathbf{d}_s \sqsubseteq \mathbf{d}_s^{\rm ls}$. This part is based on (ii) and requires the pLTS to be image-finite. Its proof makes use of the negation and testing operators in distribution formulae.

We follow the above guideline and decompose Theorem 3.1 into three technical lemmas.

Lemma 3.1 *1.* $\mathbf{d}_s^{\rm ls} \sqsubseteq \mathbf{d}_s$

2. $\mathbf{d}_s^{\rm ld} \sqsubseteq K(\mathbf{d}_s)$

Proof: We show the two statements simultaneously by structural induction on formulae. For any two states $s, t \in S$ and distributions $\Delta_1, \Delta_2 \in \mathcal{D}(S)$, we prove that

(i) $|[\![\varphi]\!](s) - [\![\varphi]\!](t)| \leq \mathbf{d}_s(s,t)$ for all $\varphi \in \mathcal{L}^{\rm S}$;

(ii) $|[\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2)| \leq K(\mathbf{d}_s)(\Delta_1, \Delta_2)$ for all $\psi \in \mathcal{L}^{\rm D}$.

We first analyze the structure of φ in (i).

- $\varphi \equiv \top$. Then it is trivial to see that $|[\![\varphi]\!](s) - [\![\varphi]\!](t)| = |1-1| = 0 \leq \mathbf{d}_s(s,t)$.
- $\varphi \equiv \neg\varphi'$. Then $|[\![\varphi]\!](s) - [\![\varphi]\!](t)| = |[\![\varphi']\!](t) - [\![\varphi']\!](s)| \leq \mathbf{d}_s(s,t)$ where the inequality holds by induction.
- $\varphi \equiv \varphi' \ominus p$. There are four subcases and we consider one of them. Suppose $[\![\varphi']\!](s) > p$ and $[\![\varphi']\!](t) \leq p$, then $|[\![\varphi]\!](s) - [\![\varphi]\!](t)| = |[\![\varphi']\!](s) - p| \leq |[\![\varphi']\!](s) - [\![\varphi']\!](t)| \leq \mathbf{d}_s(s,t)$ by induction.
- $\varphi \equiv \varphi_1 \wedge \varphi_2$. Without loss of generality we assume that $[\![\varphi]\!](s) \geq [\![\varphi]\!](t)$. There are two possibilities:
  - If $[\![\varphi_1]\!](t) \leq [\![\varphi_2]\!](t)$, then $[\![\varphi]\!](s) - [\![\varphi]\!](t) \leq [\![\varphi_1]\!](s) - [\![\varphi_1]\!](t) \leq \mathbf{d}_s(s,t)$, where the last inequality holds by induction.

  - Symmetrically, if $[\![\varphi_2]\!](t) \leq [\![\varphi_1]\!](t)$, then $[\![\varphi]\!](s) - [\![\varphi]\!](t) \leq [\![\varphi_2]\!](s) - [\![\varphi_2]\!](t) \leq \mathbf{d}_s(s,t)$.

- $\varphi \equiv \langle a \rangle \psi$. If either s or t cannot perform action a, the expected result is straightforward. So we consider the non-trivial case that both s and t can perform action a. Let Δ_1 be a distribution such that $s \xrightarrow{a} \Delta_1$ and $[\![\langle a \rangle \psi]\!](s) = [\![\psi]\!](\Delta_1)$. Since $\mathbf{d}_s$ is a state-based bisimulation metric, by definition there exists some Δ_2 such that $t \xrightarrow{a} \Delta_2$ and

$$K(\mathbf{d}_s)(\Delta_1, \Delta_2) \leq \mathbf{d}_s(s,t) \; . \tag{5}$$

Without loss of generality we assume that $[\![\varphi]\!](s) \geq [\![\varphi]\!](t)$. It follows that

$$\begin{array}{ll} & [\![\varphi]\!](s) - [\![\varphi]\!](t) \\ = & [\![\psi]\!](\Delta_1) - \max_{s \xrightarrow{a} \Delta'} [\![\psi]\!](\Delta') \\ \leq & [\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2) \\ \leq & K(\mathbf{d}_s)(\Delta_1, \Delta_2) \qquad \text{by induction on } \psi \\ \leq & \mathbf{d}_s(s,t) \qquad \text{by (5)} \end{array}$$

Then we analyze the structure of ψ in (ii).

- $\psi \equiv [\varphi]$ for some $\varphi \in \mathcal{L}^{\mathrm{S}}$. Without loss of generality we assume that $[\![\psi]\!](\Delta_1) \geq [\![\psi]\!](\Delta_2)$. We infer that

$$\begin{array}{ll} & [\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2) \\ = & [\![[\varphi]]\!](\Delta_1) - [\![[\varphi]]\!](\Delta_2) \\ = & \sum_{u \in S} (\Delta_1(u) - \Delta_2(u)) [\![\varphi]\!](u) \\ \leq & \max\{\sum_{u \in S} (\Delta_1(u) - \Delta_2(u)) x_u \mid x_u, x_{u'} \in [0,1] \wedge x_u - x_{u'} \leq \mathbf{d}_s(u,u')\} \\ = & K(\mathbf{d}_s)(\Delta_1, \Delta_2) \end{array}$$

where the last equality holds because of the Kantorovich-Rubinstein duality theorem [34, 48] and the last inequality holds because for any states $u, u' \in S$ we have $[\![\varphi]\!](u), [\![\varphi]\!](u') \in [0,1]$ and $|[\![\varphi]\!](u) - [\![\varphi]\!](u')| \leq \mathbf{d}_s(u,u')$ by induction.

- $\psi = \psi_1 \wedge \psi_2$. Without loss of generality we assume that $[\![\psi]\!](\Delta_1) \geq [\![\psi]\!](\Delta_2)$. There are two possibilities:

  - If $[\![\psi_1]\!](\Delta_2) \leq [\![\psi_2]\!](\Delta_2)$, then $[\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2) \leq [\![\psi_1]\!](\Delta_1) - [\![\psi_1]\!](\Delta_2) \leq K(\mathbf{d}_s)(\Delta_1, \Delta_2)$, where the last inequality holds by induction.

– Symmetrically, if $[\![\psi_2]\!](\Delta_2) \leq [\![\psi_1]\!](\Delta_2)$, then $[\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2) \leq [\![\psi_2]\!](\Delta_1) - [\![\psi_2]\!](\Delta_2) \leq K(\mathbf{d}_s)(\Delta_1, \Delta_2)$.

- $\psi \equiv \neg\psi'$ or $\psi' \ominus p$. Similar to the proof by induction of the last case.

□

Lemma 3.2 $K(\mathbf{d}_s^{\mathrm{ls}}) \sqsubseteq \mathbf{d}_s^{\mathrm{ld}}$

Proof: Let Δ_1, Δ_2 be any two distributions in $\mathcal{D}(S)$. We aim to show that

$$K(\mathbf{d}_s^{\mathrm{ls}})(\Delta_1, \Delta_2) \ \leq \ \sup_{\psi \in \mathcal{L}^{\mathrm{D}}} |[\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2)| \ . \tag{6}$$

Let $L(\Delta_1, \Delta_2)$ be the optimal value of the following linear program

$$\begin{array}{ll} \max \sum_{s \in S} (\Delta_1(s) - \Delta_2(s)) x_s, & \\ \qquad\qquad \text{subject to} & 0 \leq x_s \leq 1 \\ & \forall s,t \in S\colon \ x_s - x_t \leq \mathbf{d}_s^{\mathrm{ls}}(s,t) \end{array} \tag{7}$$

Let $\{k_s\}_{s \in S}$ be a set of real numbers in the interval $[0, 1]$ that maximize the above linear program to reach $L(\Delta_1, \Delta_2)$. We first consider the special case that $k_s = 1$ for all $s \in S$. Then the maximum value of the linear program in (7) is

$$\sum_{s \in S} (\Delta_1(s) - \Delta_2(s)) \cdot 1 = \sum_{s \in S} \Delta_1(s) - \sum_{s \in S} \Delta_2(s) = 1 - 1 = 0 \ .$$

It follows that $K(\mathbf{d}_s^{\mathrm{ls}})(\Delta_1, \Delta_2) = 0$ and this immediately implies (6).

Now consider the general case that $k_s < 1$ for at least one $s \in S$. We are going to show (6) by using an idea inspired by [18]. Let

$$e = \min\{1 - k_t \mid k_t < 1 \text{ and } t \in S\}$$

and $\epsilon > 0$ be any positive real number smaller than e. Hence, if $t \in S$ and $k_t < 1$ then

$$k_t + \epsilon < 1. \tag{8}$$

We construct some formula ψ such that

$$L(\Delta_1, \Delta_2) - \epsilon \ < \ [\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2). \tag{9}$$

For any $s, t \in S$, we distinguish two cases:

1. If $k_s > k_t$, then $0 < k_s - k_t \leq \mathbf{d}_s^{\mathrm{ls}}(s,t)$. It is easy to see that there exists some formula φ_{st} such that
$$k_s - k_t < [\![\varphi_{st}]\!](s) - [\![\varphi_{st}]\!](t) + \epsilon. \tag{10}$$
or equivalently $[\![\varphi_{st}]\!](t) - [\![\varphi_{st}]\!](s) + k_s < k_t + \epsilon$. We define a new formula
$$\varphi'_{st} = \begin{cases} \varphi_{st} \ominus ([\![\varphi_{st}]\!](s) - k_s) & \text{if } [\![\varphi_{st}]\!](s) > k_s \\ \varphi_{st} \oplus (k_s - [\![\varphi_{st}]\!](s)) & \text{otherwise.} \end{cases}$$
Let us compare $[\![\varphi'_{st}]\!](t)$ with k_t.
   (a) If $[\![\varphi_{st}]\!](s) > k_s$, then
$$\begin{array}{rcl} [\![\varphi'_{st}]\!](t) & = & \max([\![\varphi_{st}]\!](t) - [\![\varphi_{st}]\!](s) + k_s, 0) \\ & < & \max(k_t + \epsilon, 0) \qquad \text{by (10)} \\ & = & k_t + \epsilon \end{array}$$
   (b) Otherwise, we have $[\![\varphi'_{st}]\!](t) = \min([\![\varphi_{st}]\!](t) + k_s - [\![\varphi_{st}]\!](s), 1)$ by definition. By (10) we infer the inequality that $[\![\varphi_{st}]\!](t) + k_s - [\![\varphi_{st}]\!](s) < k_t + \epsilon$. It follows that $[\![\varphi'_{st}]\!](t) < k_t + \epsilon$.

   In both (a) and (b) we have $[\![\varphi'_{st}]\!](t) < k_t + \epsilon$, and it is also easy to see that $[\![\varphi'_{st}]\!](s) = k_s$.
2. If $k_s \leq k_t$, then we simply set φ'_{st} to be the formula $\underline{k_s}$. As in the last case, we have $[\![\varphi'_{st}]\!](s) = k_s$ and $[\![\varphi'_{st}]\!](t) = k_s \leq k_t < k_t + \epsilon$.

In summary, the above reasoning says that for any $s, t \in S$ we can construct a formula φ'_{st} such that $[\![\varphi'_{st}]\!](s) = k_s$ and $[\![\varphi'_{st}]\!](t) < k_t + \epsilon$. Now let us define $\varphi'_s = \bigwedge_{t \in S} \varphi'_{st}$. It is easy to see that $[\![\varphi'_s]\!](s) = k_s$ and $[\![\varphi'_s]\!](t) < k_t + \epsilon$ for all $t \in S$. The latter implies $\max\{[\![\varphi'_s]\!](t) \mid s, t \in S\} < k_t + \epsilon$. Then define $\varphi = \bigvee_{s \in S} \varphi'_s$. For all $t \in S$, we have
$$k_t \ = \ [\![\varphi'_t]\!](t) \ \leq \ [\![\varphi]\!](t) \ = \ \max\{[\![\varphi'_s]\!](t) \mid s, t \in S\} \ < \ k_t + \epsilon.$$
Finally, we define $\psi = [\varphi]$. It follows that
$$\begin{array}{rcl} [\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2) & = & [\![[\varphi]]\!](\Delta_1) - [\![[\varphi]]\!](\Delta_2) \\ & = & \sum_{t \in S} \Delta_1(t) \cdot [\![\varphi]\!](t) - \sum_{t \in S} \Delta_2(t) \cdot [\![\varphi]\!](t) \\ & \geq & \sum_{t \in S} \Delta_1(t) \cdot k_t - \sum_{t \in S} \Delta_2(t) \cdot [\![\varphi]\!](t) \\ & > & \sum_{t \in S} \Delta_1(t) \cdot k_t - \sum_{t \in S} \Delta_2(t) \cdot (k_t + \epsilon) \\ & = & \sum_{t \in S} (\Delta_1(t) - \Delta_2(t)) \cdot k_t - \sum_{t \in S} \Delta_2(t) \cdot \epsilon \\ & = & L(\Delta_1, \Delta_2) - \epsilon \end{array}$$

as required in (9). □

The above property will be used to prove the following lemma.

Lemma 3.3 $\mathbf{d}_s \sqsubseteq \mathbf{d}_s^{\mathrm{ls}}$

Proof: We show that $\mathbf{d}_s^{\mathrm{ls}}$ is a state-based bisimulation metric. Let s, t be any two states in S and ϵ be any real number in the interval $[0, 1)$ with $\mathbf{d}_s^{\mathrm{ls}}(s, t) \leq \epsilon$. Assume that $s \xrightarrow{a} \Delta_1$ is an arbitrarily chosen transition from s. Then state t must be able to perform action a too. Otherwise it is easy to see that $\mathbf{d}_s^{\mathrm{ls}}(s, t) = 1 > \epsilon$, which contradicts our assumption above. We need to show that there exists some transition $t \xrightarrow{a} \Delta_2$ with $K(\mathbf{d}_s^{\mathrm{ls}})(\Delta_1, \Delta_2) \leq \epsilon$. Suppose for a contradiction that no a-transition from t satisfies this condition. In other words, for each Δ_2^i with $t \xrightarrow{a} \Delta_2^i$ we have $K(\mathbf{d}_s^{\mathrm{ls}})(\Delta_1, \Delta_2^i) > \epsilon$. By Lemma 3.2, this means $\mathbf{d}_s^{\mathrm{ld}}(\Delta_1, \Delta_2^i) > \epsilon$. Then there must exist some formula $\psi_2^i \in \mathcal{L}^{\mathrm{D}}$ such that $|[\![\psi_2^i]\!](\Delta_1) - [\![\psi_2^i]\!](\Delta_2^i)| > \epsilon$. Furthermore, we can strengthen this condition to the following one

$$[\![\psi_2^i]\!](\Delta_1) - [\![\psi_2^i]\!](\Delta_2^i) > \epsilon \tag{11}$$

because we can take the formula $\neg\psi_2^i$ in place of ψ_2^i in the case that $[\![\psi_2^i]\!](\Delta_1) < [\![\psi_2^i]\!](\Delta_2)$. Let

$$\varphi = \langle a \rangle \bigwedge_i (\psi_2^i \ominus [\![\psi_2^i]\!](\Delta_2^i)) \ .$$

We infer that

$$\begin{array}{rcl}
[\![\varphi]\!](s) & = & \max_{s \xrightarrow{a} \Delta} [\![\bigwedge_i \psi_2^i \ominus [\![\psi_2^i]\!](\Delta_2^i)]\!](\Delta) \\
& \geq & [\![\bigwedge_i (\psi_2^i \ominus [\![\psi_2^i]\!](\Delta_2^i))]\!](\Delta_1) \\
& = & \min_i [\![\psi_2^i \ominus [\![\psi_2^i]\!](\Delta_2^i)]\!](\Delta_1) \\
& = & [\![\psi_2^k \ominus [\![\psi_2^k]\!](\Delta_2^k)]\!](\Delta_1) \qquad \text{for some } k \\
& = & \max([\![\psi_2^k]\!](\Delta_1) - [\![\psi_2^k]\!](\Delta_2^k),\ 0) \\
& > & \epsilon \qquad \text{by (11)}
\end{array}$$

On the other hand, we have

$$\begin{array}{rcl}
[\![\varphi]\!](t) & = & \max_{t \xrightarrow{a} \Delta_2^i} [\![\bigwedge_j (\psi_2^j \ominus [\![\psi_2^j]\!](\Delta_2^j))]\!](\Delta_2^i) \\
& = & \max_{t \xrightarrow{a} \Delta_2^i} \min_j [\![\psi_2^j \ominus [\![\psi_2^j]\!](\Delta_2^j)]\!](\Delta_2^i) \\
& = & \max_{t \xrightarrow{a} \Delta_2^i} \min_j \max(([\![\psi_2^j]\!](\Delta_2^i) - [\![\psi_2^j]\!](\Delta_2^j)),\ 0) \\
& = & 0
\end{array}$$

It follows that $\mathbf{d}_s^{\mathrm{ls}}(s,t) \geq [\![\varphi]\!](s) - [\![\varphi]\!](t) > \epsilon$, which gives rise to a contradiction. □

Finally, we obtain a proof of Theorem 3.1.

Proof: By combining the last three technical lemmas. □

Remark 3.2 *In the proof of Lemma 3.3 we have constructed the formula*

$$\varphi = \langle a \rangle \bigwedge_i (\psi_2^i \ominus [\![\psi_2^i]\!](\Delta_2^i)) \tag{12}$$

by making use of conjunction and minus connectives for distribution formulae. This happens because in the presence of non-determinism state t may perform action a and then evolves into one of several successor distributions Δ_2^i. If we confine ourselves to deterministic pLTSs, then state t will have a unique successor distribution Δ_2^i and therefore (12) can be simplified as $\varphi = \langle a \rangle \psi_2^i$. In this case, there is no need of conjunction and minus connectives for distribution formulae. That is, distribution formulae are in the form $[\varphi]$ or $\neg[\varphi]$. Furthermore, if we fold them into state formulae in Definition 3.2, distribution formulae can be completely dropped. In other words, for deterministic pLTSs, the state-based bisimilarity metric can be characterised by the following one-sorted metric logic

$$\varphi ::= \top \mid \neg\varphi \mid \varphi \ominus p \mid \varphi_1 \wedge \varphi_2 \mid \langle a \rangle \varphi \ . \tag{13}$$

Therefore, for deterministic pLTSs, the two-sorted logic in Definition 3.2 degenerates into the logic considered in [17, 50, 26], as expected. In the one-sorted logic, the formula $\langle a \rangle(\varphi \ominus p)$ will be interpreted the same as the formula $\langle a \rangle[\varphi \ominus p]$ in $\mathcal{L}^{\mathrm{S}}$, but no formula has the same interpretation as $\langle a \rangle([\varphi] \ominus p)$ in $\mathcal{L}^{\mathrm{S}}$; the subtlety has already been discussed in Example 3.3.

In [8, 3] a bisimulation metric for game structures is characterised by a quantitative μ-calculus where formulae are evaluated also on states and no distribution formula is needed. This is not surprising because the considered 2-player games are deterministic: at any state s, if two players have chosen their moves, say a_1 and a_2, then there is a unique distribution $\delta(s, a_1, a_2)$ to determine the probabilities of arriving at a set of destination states.

4 Distribution-Based Bisimulation Metric

The bisimilarity metric given in Definition 3.1 measures the distance between two states. Alternatively, it is possible to directly define a metric that measures subdistributions. In order to do so, we first define a transition relation between subdistributions.

Definition 4.1 *With a slight abuse of notation, we also use the notation $\xrightarrow{a}$ to stand for the transition relation between subdistributions, which is the smallest relation satisfying the three rules given in Figure 3.*

$$\frac{s \xrightarrow{a} \Delta}{\overline{s} \xrightarrow{a} \Delta} \qquad \frac{s \not\xrightarrow{a}}{\overline{s} \xrightarrow{a} \varepsilon}$$

$$\frac{\forall i \in I.\ p_i > 0 \wedge \Delta_i \xrightarrow{a} \Theta_i \quad I \text{ is finite} \quad \sum_{i\in I} p_i \leq 1}{(\sum_{i\in I} p_i \cdot \Delta_i) \xrightarrow{a} (\sum_{i\in I} p_i \cdot \Theta_i)}$$

Figure 3: Rules for transitions between subdistributions

Note that if $\Delta \xrightarrow{a} \Delta'$ then not necessarily all the states in the support of Δ can perform action a. For example, consider the two states s_2 and s_3 in Figure 1. Since $s_2 \xrightarrow{c} \overline{s_4}$ and s_3 cannot perform action c, the distribution $\Delta = \frac{1}{2}\overline{s_2} + \frac{1}{2}\overline{s_3}$ can make the transition $\Delta \xrightarrow{c} \frac{1}{2}\overline{s_4}$ to reach the subdistribution $\frac{1}{2}\overline{s_4}$.

Lemma 4.1 *For any subdistribution $\Delta \in \mathcal{D}_{sub}(S)$ and action a, if $\Delta \xrightarrow{a} \Delta'$ then there exisits some subdistributions Δ_s such that*

1. $\Delta' = \sum_{s\in\lceil\Delta\rceil} \Delta(s) \cdot \Delta_s$;
2. $\overline{s} \xrightarrow{a} \Delta_s$ *for each* $s \in \lceil\Delta\rceil$;
3. *if* $s \not\xrightarrow{a}$ *then* $\Delta_s = \varepsilon$.

Proof: By induction on the rules of inferring $\Delta \xrightarrow{a} \Delta'$. As displayed in Figure 3, there are three rules. The first two are straightforward, so we assume that $\Delta \xrightarrow{a} \Delta'$ is derived from the last one. Suppose $\Delta = \sum_{i\in I} p_i \cdot \Delta_i$, $\Delta' = \sum_{i\in I} p_i \cdot \Delta'_i$, $\sum_{i\in I} p_i \leq 1$ and for all $i \in I$ we have $\Delta_i \xrightarrow{a} \Delta'_i$. By induction hypothesis, for each $i \in I$, the subdistribution Δ'_i can be decomposed as $\Delta'_i = \sum_{s\in\lceil\Delta_i\rceil} \Delta_i(s) \cdot \Delta_{is}$ with $\overline{s} \xrightarrow{a} \Delta_{is}$ for each $s \in \lceil\Delta_i\rceil$, and $\Delta_{is} = \varepsilon$ if $s \not\xrightarrow{a}$. Note that $\Delta(s) = \sum_{i\in I} p_i \cdot \Delta_i(s)$. It follows that

$$\overline{s} = \sum_{i\in I} \frac{p_i \Delta_i(s)}{\Delta(s)} \cdot \overline{s} \ \xrightarrow{a}\ \sum_{i\in I} \frac{p_i \Delta_i(s)}{\Delta(s)} \cdot \Delta_{is}$$

Now let $\Delta_s = \sum_{i\in I} \frac{p_i\Delta_i(s)}{\Delta(s)} \cdot \Delta_{is}$. The above transition can be written as $\overline{s} \xrightarrow{a} \Delta_s$. We also observe that

$$\Delta' = \sum_{i\in I} p_i \cdot \sum_{s\in\lceil\Delta_i\rceil} \Delta_i(s) \cdot \Delta_{is} = \sum_{s\in\lceil\Delta\rceil} \sum_{i\in I} p_i\Delta_i(s) \cdot \Delta_{is} = \sum_{s\in\lceil\Delta\rceil} \Delta(s) \cdot \Delta_s$$

Moreover, if $s \not\xrightarrow{a}$ then $\Delta_{is} = \varepsilon$ and thus $\Delta_s = \varepsilon$ as required. □

Definition 4.2 *A* 1*-bounded pseudometric d on $\mathcal{D}_{sub}(S)$ is a* distribution-based bisimulation metric *if, for all $\Delta_1, \Delta_2 \in \mathcal{D}_{sub}(S)$, the following two conditions are satisfied:*

1. $|\,|\Delta_1| - |\Delta_2|\,| \leq d(\Delta_1, \Delta_2)$*;*
2. *whenever $d(\Delta_1, \Delta_2) < 1$ and $\Delta_1 \xrightarrow{a} \Delta'_1$ then there is some transition $\Delta_2 \xrightarrow{a} \Delta'_2$ such that $d(\Delta'_1, \Delta'_2) \leq d(\Delta_1, \Delta_2)$.*

The condition $|\,|\Delta_1| - |\Delta_2|\,| \leq d(\Delta_1, \Delta_2)$ says that the distance between two subdistributions should be at least the difference between their total masses. The smallest (wrt. $\sqsubseteq$) distribution-based bisimulation metric, notation $\mathbf{d}_d$, is called *distribution-based bisimilarity metric*. Distribution-based bisimilarity [14] is the kernel of the distribution-based bisimilarity metric.

Let $der(\Delta, a) = \{\Delta' \mid \Delta \xrightarrow{a} \Delta'\}$. We define the functional operator

$$F_{\mathrm{d}} \colon [0,1]^{\mathcal{D}_{sub}(S)\times\mathcal{D}_{sub}(S)} \to [0,1]^{\mathcal{D}_{sub}(S)\times\mathcal{D}_{sub}(S)}$$

for $d\colon \mathcal{D}_{sub}(S) \times \mathcal{D}_{sub}(S) \to [0,1]$ and $\Delta, \Theta \in \mathcal{D}_{sub}(S)$ by

$$F_{\mathrm{d}}(d)(\Delta, \Theta) = \max(\sup_{a\in A}\{H(d)(der(\Delta, a), der(\Theta, a))\},\ |\,|\Delta| - |\Theta|\,|)\,. \quad (14)$$

It can be shown that F_{d} is monotone and its least fixed point is given by $\bigsqcup d_i$, where $d_0(\Delta, \Theta) = |\,|\Delta| - |\Theta|\,|$ for any $\Delta, \Theta \in \mathcal{D}_{sub}(S)$ and $d_{i+1} = F_{\mathrm{d}}(d_i)$ for all $i \in \mathbb{N}$. The property below is analogous to Proposition 3.1.

Proposition 4.1 $\mathbf{d}_d$ *is the least fixed point of F_{d}.* □

It is not difficult to see that $\mathbf{d}_s$ is different from $\mathbf{d}_d$, as witnessed by the following example. A more accurate comparison is given in Section 5.

Example 4.1 *Consider the states in Figure 1. We first observe that* $\mathbf{d}_d(\overline{s_2}, \overline{t_3}) = 0$ *because* s_2 *and* t_3 *can match each other's action exactly. Similarly, we have* $\mathbf{d}_d(\overline{s_3}, \overline{t_4}) = 0$*. Then it is straightforward to see that* $\mathbf{d}_d(\frac{1}{2}\overline{s_2}+\frac{1}{2}\overline{s_3}, \frac{1}{2}\overline{t_3}+\frac{1}{2}\overline{t_4}) = 0$*. Since* $\overline{s_1} \xrightarrow{b} \frac{1}{2}\overline{s_2}+\frac{1}{2}\overline{s_3}$ *and* $\frac{1}{2}\overline{t_1}+\frac{1}{2}\overline{t_2} \xrightarrow{b} \frac{1}{2}\overline{t_3}+\frac{1}{2}\overline{t_4}$*, we infer that* $\mathbf{d}_d(\overline{s_1}, \frac{1}{2}\overline{t_1} + \frac{1}{2}\overline{t_2}) = 0$*. This, in turn, implies* $\mathbf{d}_d(\overline{s}, \overline{t}) = 0$*. We have already seen in Example 3.1 that* $\mathbf{d}_s(s, t) = \frac{1}{2}$*. Therefore, the two distance functions* $\mathbf{d}_s$ *and* $\mathbf{d}_d$ *are indeed different.*

We now turn to the logical characterisation of $\mathbf{d}_d$. Consider the metric logic $\mathcal{L}^{D*}$ whose formulae are defined below:

$$\psi ::= \top \mid \neg\psi \mid \psi \ominus p \mid \psi_1 \wedge \psi_2 \mid \langle a\rangle\psi\ . \tag{15}$$

This logic is the same as that defined in (13) except that now we only have distribution formulae. The semantic interpretation of formulae comes with no surprise.

Definition 4.3 *A formula* $\psi \in \mathcal{L}^{D*}$ *evaluates in* $\Delta \in \mathcal{D}_{sub}(S)$ *as follows:*

$$\begin{array}{rcl} [\![\top]\!](\Delta) & = & |\Delta| \\ [\![\neg\psi]\!](\Delta) & = & 1 - [\![\psi]\!](\psi) \\ [\![\psi \ominus p]\!](\Delta) & = & \max([\![\psi]\!](\Delta) - p,\ 0) \\ [\![\psi_1 \wedge \psi_2]\!](\Delta) & = & \min([\![\psi_1]\!](\Delta), [\![\psi_2]\!](\Delta)) \\ [\![\langle a\rangle\psi]\!](\Delta) & = & \max_{\Delta \xrightarrow{a} \Delta'} [\![\psi]\!](\Delta'). \end{array}$$

This induces a natural logical metric $\mathbf{d}_d^{\mathrm{ld}}$ over subdistributions defined by

$$\mathbf{d}_d^{\mathrm{ld}}(\Delta, \Theta) = \sup_{\psi \in \mathcal{L}^{D*}} |[\![\psi]\!](\Delta) - [\![\psi]\!](\Theta)|$$

It turns out that $\mathbf{d}_d^{\mathrm{ld}}$ coincides with $\mathbf{d}_d$. Below we show that one metric is dominated by the other and vice versa.

Lemma 4.2 $\mathbf{d}_d^{\mathrm{ld}} \sqsubseteq \mathbf{d}_d$

Proof: Similar to the proof of Lemma 3.1. We proceed by structural induction on formulae. For any two subdistributions $\Delta_1, \Delta_2 \in \mathcal{D}(S)$, we prove that

$$|[\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2)| \leq \mathbf{d}_d(\Delta_1, \Delta_2)$$

for all $\psi \in \mathcal{L}^{D*}$.

Let us analyze the structure of ψ.

- $\psi \equiv \top$. Then it is trivial to see that $|[\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2)| = |\,|\Delta_1| - |\Delta_2|\,| \leq \mathbf{d}_d(\Delta_1, \Delta_2)$.

- $\psi \equiv \neg\psi'$. Then $|[\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2)| = |[\![\psi']\!](\Delta_2) - [\![\psi']\!](\Delta_1)| \leq \mathbf{d}_d(\Delta_1, \Delta_2)$ where the inequality holds by induction.

- $\psi \equiv \psi' \ominus p$. There are four subcases and we consider one of them. Suppose $[\![\psi']\!](\Delta_1) > p$ and $[\![\psi']\!](\Delta_2) \leq p$, then $|[\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2)| = |[\![\psi']\!](\Delta_1) - p| \leq |[\![\psi']\!](\Delta_1) - [\![\psi']\!](\Delta_2)| \leq \mathbf{d}_d(\Delta_1, \Delta_2)$ by induction.

- $\psi \equiv \psi_1 \wedge \psi_2$. Without loss of generality we assume that $[\![\psi]\!](\Delta_1) \geq [\![\psi]\!](\Delta_2)$. There are two possibilities:

  - If $[\![\psi_1]\!](\Delta_2) \leq [\![\psi_2]\!](\Delta_2)$, then $[\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2) \leq [\![\psi_1]\!](\Delta_1) - [\![\psi_1]\!](\Delta_2) \leq \mathbf{d}_d(\Delta_1, \Delta_2)$, where the last inequality holds by induction.
  - Symmetrically, if $[\![\psi_2]\!](\Delta_2) \leq [\![\psi_1]\!](\Delta_2)$, then $[\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2) \leq [\![\psi_2]\!](\Delta_1) - [\![\psi_2]\!](\Delta_2) \leq \mathbf{d}_d(\Delta_1, \Delta_2)$.

- $\psi \equiv \langle a \rangle \psi'$. Let Δ_1' be a distribution such that $\Delta_1 \xrightarrow{a} \Delta_1'$ and $[\![\langle a \rangle \psi']\!](\Delta_1) = [\![\psi']\!](\Delta_1')$. Since $\mathbf{d}_d$ is a distribution-based bisimulation metric, by definition there exists some Δ_2' such that $\Delta_2 \xrightarrow{a} \Delta_2'$ and $\mathbf{d}_d(\Delta_1', \Delta_2') \leq \mathbf{d}_d(\Delta_1, \Delta_2)$. Without loss of generality we assume that $[\![\psi]\!](\Delta_1) \geq [\![\psi]\!](\Delta_2)$. It follows that

$$\begin{array}{cll} & [\![\psi]\!](\Delta_1) - [\![\psi]\!](\Delta_2) & \\ = & [\![\psi']\!](\Delta_1') - \max_{\Delta_2 \xrightarrow{a} \Delta_2''} [\![\psi']\!](\Delta_2'') & \\ \leq & [\![\psi']\!](\Delta_1') - [\![\psi']\!](\Delta_2') & \\ \leq & \mathbf{d}_d(\Delta_1', \Delta_2') & \text{by induction on } \psi' \\ \leq & \mathbf{d}_d(\Delta_1, \Delta_2) & \end{array}$$

□

Lemma 4.3 $\mathbf{d}_d \sqsubseteq \mathbf{d}_d^{\mathrm{ld}}$

Proof: The proof is similar to that of Lemma 3.3, so we omit it. □

By combining the previous two lemmas, we obtain the logical characterisation of $\mathbf{d}_d$.

Theorem 4.1 $\mathbf{d}_d = \mathbf{d}_d^{\mathrm{ld}}$ □

5 Comparison of the Bisimilarity Metrics

In this section, we compare the state-based bisimilarity metric $\mathbf{d}_s$ with the distribution-based bisimilarity metric $\mathbf{d}_d$. More precisely, we show that $\mathbf{d}_d$ is a lower bound of $K(\mathbf{d}_s)$ when measuring full distributions[7]. The proof makes use of fully enabled pLTSs as a stepping stone. Let us first fix an overall set of actions Act and a special action $\bot \notin Act$. Let $EA(s) = \{a \mid \exists \Delta.\ s \xrightarrow{a} \Delta\}$ be the set of actions that are enabled at state s. We also use $\bot$ to stand for a special state when there is no confusion.

Definition 5.1 *A pLTS with state set S is* fully enabled *if for any state $s \in S \backslash \{\bot\}$ we have $EA(s) = Act$. Given any pLTS $\mathcal{A} = (S, A, \rightarrow)$ with $A \subseteq Act$, we can convert it into a fully enabled pLTS $\mathcal{A}^\bot = (S_\bot, Act \cup \{\bot\}, \rightarrow_\bot)$ as follows:*

$$\begin{array}{rcl} S_\bot & = & \{s^\bot | s \in S\} \cup \{\bot\} \\ \rightarrow_\bot & = & \{(s^\bot, a, \Delta^\bot) \mid (s, a, \Delta) \in \rightarrow\} \\ & & \cup \{(s^\bot, a, \overline{\bot}) \mid s \not\xrightarrow{a} \text{ and } a \in Act\} \\ & & \cup \{(\bot, a, \overline{\bot}) \mid a \in Act \cup \{\bot\}\}. \end{array}$$

where $\Delta^\bot(s^\bot) = \Delta(s)$ for each $s \in S$ and $\Delta^\bot(\bot) = 1 - |\Delta|$. In other words, each state s in $\mathcal{A}$ corresponds to a state $s^\bot$ in $\mathcal{A}^\bot$ such that $s^\bot$ keeps all the transitions of s and can evolve into the absorbing state $\bot$ by performing any action in Act not enabled by s. As a consequence, each subdistribution Δ on the states of $\mathcal{A}$ has a corresponding full distribution $\Delta^\bot$ on the states of $\mathcal{A}^\bot$.

For any pLTS, let s, t be two states and Δ, Θ two subdistributions. It can be shown that $\mathbf{d}_s(s, t) = \mathbf{d}_s(s^\bot, t^\bot)$ and $\mathbf{d}_d(\Delta, \Theta) = \mathbf{d}_d(\Delta^\bot, \Theta^\bot)$. Moreover, for fully enabled pLTSs, the metric $\mathbf{d}_d$ turns out to be a lower bound of $K(\mathbf{d}_s)$ as far as distributions are concerned. Before proving those properties, we first present the following technical lemma.

Lemma 5.1 *For any subdistribution Δ on $\mathcal{A}$ and any $a \in Act$,*

$$\Delta \xrightarrow{a} \Delta_1 \textit{ iff } \Delta^\bot \xrightarrow{a} \Delta_1^\bot.$$

Proof: ($\Rightarrow$) Suppose Δ is a subdistribution on $\mathcal{A}$ and $\Delta \xrightarrow{a} \Delta_1$ for some $a \in Act$ and subdistribution Δ_1. By Lemma 4.1 we can decompose Δ_1 such

[7]Although $\mathbf{d}_d$ can measure the distance between two subdistributions, the Kantorovich lifting of $\mathbf{d}_s$ can only measure the distance between full distributions or subdistributions of equal mass, which can easily be normalized to full distributions.

that $\Delta_1 = \sum_{s\in\lceil\Delta\rceil} \Delta(s)\cdot\Delta_s$, $\overline{s} \xrightarrow{a} \Delta_s$ for each $s \in S'$, where S' is the set of states in the support of Δ that can enable action a, and $\Delta_s = \varepsilon$ if $s \notin S'$. For each $s \in S'$, the state $s^\perp$ keeps all the transitions of s, so we have some distribution $\Delta_s^\perp$ with $\overline{s^\perp} \xrightarrow{a} \Delta_s^\perp$. For each $s \notin S'$, we have $s^\perp \xrightarrow{a} \overline{\perp}$. It follows that

$$\Delta^\perp \xrightarrow{a} (\sum_{s\in S'} \Delta(s)\cdot\Delta_s^\perp + \sum_{s\in\lceil\Delta\rceil\setminus S'} \Delta(s)\cdot\overline{\perp} + (1-|\Delta|)\cdot\overline{\perp}) = \Delta_1^\perp\ .$$

($\Leftarrow$) Suppose $\Delta^\perp \xrightarrow{a} \Delta_1^\perp$ for some subdistribution Δ on $\mathcal{A}$ and some action $a \in Act$. We have that $\Delta^\perp = \sum_{s\in\lceil\Delta\rceil} \Delta(s)\cdot\overline{s^\perp} + (1-|\Delta|)\cdot\overline{\perp}$. By Lemma 4.1 we have that $\Delta_1^\perp = (\sum_{s\in\lceil\Delta\rceil} \Delta(s)\cdot\Theta_s) + (1-|\Delta|)\cdot\overline{\perp}$, where $\Theta_s = \Delta_s^\perp$ if s enables a and $\overline{s} \xrightarrow{a} \Delta_s$ for some distribution Δ_s, or $\Theta_s = \overline{\perp}$ if s cannot enable action a. Let S' be the set of states in the support of Δ that can enable action a. We have that

$$\Delta_1^\perp \ = \ (\sum_{s\in S'} \Delta(s)\cdot\Delta_s^\perp) + (1-\sum_{s\in S'} \Delta(s))\cdot\overline{\perp} \ = \ (\sum_{s\in S'} \Delta(s)\cdot\Delta_s)^\perp\ .$$

By setting $\Delta_1 = \sum_{s\in S'} \Delta(s)\cdot\Delta_s$, we indeed have that $\Delta \xrightarrow{a} \Delta_1$. □

Lemma 5.2 *1. Let s, t be any two states of $\mathcal{A}$. Then $\mathbf{d}_s(s,t) = \mathbf{d}_s(s^\perp, t^\perp)$*

2. Let Δ,Θ be two distributions on $\mathcal{A}$. Then $K(\mathbf{d}_s)(\Delta,\Theta) = K(\mathbf{d}_s)(\Delta^\perp,\Theta^\perp)$.

3. Let Δ,Θ be two subdistributions on $\mathcal{A}$. Then $\mathbf{d}_d(\Delta,\Theta) = \mathbf{d}_d(\Delta^\perp,\Theta^\perp)$.

Proof:

1. By Proposition 3.1 we see that $\mathbf{d}_s = \bigsqcup d_i$, where $d_0 = \mathbf{0}$ and $d_{i+1} = F_{\mathrm{s}}(d_i)$ for all $i \in \mathbb{N}$. We show by induction on i that $d_i(s,t) = d_i(s^\perp, t^\perp)$ for all $i \in \mathbb{N}$. The base case is trivial. Let us consider the inductive step.

$$\begin{array}{rcl} d_{i+1}(s,t) & = & \sup_{a\in Act}\{H(K(d_i))(der(s,a), der(t,a))\} \\ d_{i+1}(s^\perp,t^\perp) & = & \sup_{a\in Act\cup\{\perp\}}\{H(K(d_i))(der(s^\perp,a), der(t^\perp,a))\} \end{array}$$

If $X = \{\Delta_1, ..., \Delta_n\}$ is a set of distributions on $\mathcal{A}$, we denote by $(X)^\perp$ the set $\{\Delta_1^\perp, ..., \Delta_n^\perp\}$. For any $a \in Act$, we distinguish four cases:

(a) Both s and t can enable action a in $\mathcal{A}$. That is, neither $der(s,a)$ nor $der(t,a)$ is empty. Then $s \xrightarrow{a} \Delta$ iff $s^{\perp} \xrightarrow{a} \Delta^{\perp}$. That is, each a-successor distribution of s, say Δ, has a corresponding a-successor distribution of $s^{\perp}$, say $\Delta^{\perp}$, and vice-versa. Similarly, for each $\Theta \in der(t,a)$, we have $\Theta^{\perp} \in der(t^{\perp},a)$. This means that $der(s^{\perp},a) = (der(s,a))^{\perp}$ and $der(t^{\perp},a) = (der(t,a))^{\perp}$. By induction, we have that $d_i(u,v) = d_i(u^{\perp},v^{\perp})$ for any $u,v \in S$. It follows that

$$K(d_i)(\Delta,\Theta) = K(d_i)(\Delta^{\perp},\Theta^{\perp})$$

and moreover,

$$H(K(d_i))(der(s,a), der(t,a)) \;=\; H(K(d_i))(der(s^{\perp},a), der(t^{\perp},a))\,.$$

(b) State s cannot enable action a but state t can enable action a. Then $der(s,a) = \emptyset$, $der(s^{\perp},a) = \{\overline{\perp}\}$, $der(t,a) \neq \emptyset$ and $\overline{\perp} \notin der(t^{\perp},a)$. Clearly, $K(d_i)(\overline{\perp},\Theta^{\perp}) = 1$ for any $\Theta^{\perp} \in der(t^{\perp},a)$ because here Θ is a full distribution and $t' \stackrel{\perp}{\not\longrightarrow}$ for any $t' \in \lceil\Theta\rceil$. It follows that

$$H(K(d_i))(der(s,a), der(t,a)) \;=\; 1 \;=\; H(K(d_i))(der(s^{\perp},a), der(t^{\perp},a))\,.$$

(c) The symmetric case of (b) by exchanging the rules of s and t. We also have

$$H(K(d_i))(der(s,a), der(t,a)) \;=\; 1 \;=\; H(K(d_i))(der(s^{\perp},a), der(t^{\perp},a))\,.$$

(d) Neither s nor t can enable action a. Then $der(s,a) = der(t,a) = \emptyset$ and $der(s^{\perp},a) = der(t^{\perp},a) = \{\overline{\perp}\}$. It follows that

$$H(K(d_i))(der(s,a), der(t,a)) \;=\; 0 \;=\; H(K(d_i))(der(s^{\perp},a), der(t^{\perp},a))\,.$$

In all the cases above, we always have the following equation

$$H(K(d_i))(der(s,a), der(t,a)) \;=\; H(K(d_i))(der(s^{\perp},a), der(t^{\perp},a)) \tag{16}$$

for any $a \in Act$. For the action $\perp$, we have $der(s^{\perp},\perp) = der(t^{\perp},\perp) = \emptyset$. Hence,

$$H(K(d_i))(der(s^{\perp},\perp), der(t^{\perp},\perp)) = 0\,. \tag{17}$$

By using (16) and (17), we can reason that

$$\begin{array}{rcl} d_{i+1}(s,t) & = & \sup_{a\in Act}\{H(K(d_i))(der(s,a), der(t,a))\} \\ & = & \sup_{a\in Act}\{H(K(d_i))(der(s^{\perp},a), der(t^{\perp},a)) \\ & = & \sup_{a\in Act\cup\{\perp\}}\{H(K(d_i))(der(s^{\perp},a), der(t^{\perp},a)) \\ & = & d_{i+1}(s^{\perp},t^{\perp}) \ . \end{array}$$

2. Since Δ and Θ are full distributions, then so are $\Delta^{\perp}$ and $\Theta^{\perp}$. Then Clause 2 follows from Clause 1 immediately.

3. The proof is similar to that of Clause 1. By Proposition 4.1 we see that $\mathbf{d}_d = \bigsqcup d_i$, where $d_0 = \mathbf{0}$ and $d_{i+1} = F_{\mathrm{d}}(d_i)$ for all $i \in \mathbb{N}$. For distributions on $\mathcal{A}^{\perp}$, we need to consider the special action $\perp$ too. We show by induction on i that $d_i(\Delta,\Theta) = d_i(\Delta^{\perp},\Theta^{\perp})$ for all $i \in \mathbb{N}$. The base case is trivial. Let us consider the inductive step.

$$d_{i+1}(\Delta,\Theta) = \max(\sup_{a\in Act}\{H(d_i)(der(\Delta,a), der(\Theta,a))\},\ |\,|\Delta| - |\Theta|\,|) \ .$$

Since $\Delta^{\perp}$ and $\Theta^{\perp}$ are full distributions, there is no need of comparing their masses. Hence,

$$d_{i+1}(\Delta^{\perp},\Theta^{\perp}) = \sup_{a\in Act\cup\{\perp\}}\{H(d_i)(der(\Delta^{\perp},a), der(\Theta^{\perp},a))\} \ .$$

By Lemma 5.1, for any Δ on $\mathcal{A}$ and $a \in A$, we have the correspondence of transitions $\Delta \xrightarrow{a} \Delta_1$ iff $\Delta^{\perp} \xrightarrow{a} \Delta_1^{\perp}$. Similarly, for each $\Theta_1 \in der(\Theta,a)$, we have $\Theta_1^{\perp} \in der(\Theta^{\perp},a)$, and vice-versa. This means that $der(\Delta^{\perp},a) = (der(\Delta,a))^{\perp}$ and $der(\Theta^{\perp},a) = (der(\Theta,a))^{\perp}$. By induction, $d_i(\Delta_1,\Theta_1) = d_i(\Delta_1^{\perp},\Theta_1^{\perp})$ for any $\Delta_1,\Theta_1 \in \mathcal{D}_{sub}(S)$. It follows that

$$H(d_i)(der(\Delta,a), der(\Theta,a)) \ = \ H(d_i)(der(\Delta^{\perp},a), der(\Theta^{\perp},a)) \ .$$

Therefore,

$$\sup_{a\in A}\{H(d_i)(der(\Delta,a), der(\Theta,a))\} \ = \ \sup_{a\in A}\{H(d_i)(der(\Delta^{\perp},a), der(\Theta^{\perp},a))\} \ .$$

Observe that in $\Delta^{\perp}$ no state except for $\perp$ can enable action $\perp$, which means that the following equality holds: $der(\Delta^{\perp},\perp) = \{(1-|\Delta|)\cdot\overline{\perp}\}$. Similarly, $der(\Theta^{\perp},\perp) = \{(1-|\Theta|)\cdot\overline{\perp}\}$. We then have that

$$\begin{array}{rcl} H(d_i)(der(\Delta^{\perp},\perp), der(\Theta^{\perp},\perp))\} & = & d_i((1-|\Delta|)\cdot\overline{\perp},\ (1-|\Theta|)\cdot\overline{\perp}) \\ & = & |\,(1-|\Delta|) - (1-|\Theta|)\,| \\ & = & |\,|\Delta| - |\Theta|\,| \end{array}$$

Now it is easy to see that $d_{i+1}(\Delta, \Theta) = d_{i+1}(\Delta^\perp, \Theta^\perp)$ as required. $\square$

Theorem 5.1 *Let Δ, Θ be two distributions on a fully enabled pLTS. Then*

$$\mathbf{d}_d(\Delta, \Theta) \leq K(\mathbf{d}_s)(\Delta, \Theta).$$

Proof: We will prove that $K(\mathbf{d}_s)$ is a distribution-based bisimulation metric for fully enabled pLTSs. Since $\mathbf{d}_d$ is the smallest distribution-based bisimulation metric, it follows that $\mathbf{d}_d(\Delta, \Theta) \leq K(\mathbf{d}_s)(\Delta, \Theta)$.

By assumption, both Δ and Θ are distributions. It is trivial to see that

$$\left| |\Delta| - |\Theta| \right| \;=\; 0 \;\leq\; K(\mathbf{d}_s)(\Delta, \Theta).$$

Suppose $K(\mathbf{d}_s)(\Delta, \Theta) < 1$ and $\Delta \xrightarrow{a} \Delta'$. Then for any $s \in \lceil \Delta \rceil$, there exists some Δ_s such that $s \xrightarrow{a} \Delta_s$ and $\Delta' = \sum_{s \in \lceil \Delta \rceil} \Delta(s) \cdot \Delta_s$ because the pLTS under consideration is fully enabled. Let S be the state set of the pLTS excluding the special state $\perp$. For any $t \in S$, we observe that $\mathbf{d}_s(s,t) < 1$ because in a fully enabled pLTS no two states different from $\perp$ have distance 1. So by the definition of $\mathbf{d}_s$, there exists some Θ_t such that $t \xrightarrow{a} \Theta_t$ and $K(\mathbf{d}_s)(\Delta_s, \Theta_t) \leq \mathbf{d}_s(s,t)$. We define $\Theta' = \sum_{t \in S} \Theta(t) \cdot \Theta_t$ and it is easy to see that $\Theta \xrightarrow{a} \Theta'$.

Let $\omega \in \Omega(\Delta, \Theta)$ be a weight function satisfying

$$K(\mathbf{d}_s)(\Delta, \Theta) = \sum_{s,t \in S} \omega(s,t) \cdot \mathbf{d}_s(s,t).$$

Similarly, let $\omega_{s,t} \in \Omega(\Delta_s, \Theta_t)$ be a weight function satisfying

$$K(\mathbf{d}_s)(\Delta_s, \Theta_t) = \sum_{u,v \in S} \omega_{s,t}(u,v) \cdot \mathbf{d}_s(u,v).$$

Define $\omega' \in \mathcal{D}(S \times S)$ as follows:

$$\omega'(u,v) = \sum_{s,t \in S} \omega(s,t) \cdot \omega_{s,t}(u,v)$$

for any $u, v \in S$. We check that ω' is a weight function for Δ' and Θ'.

$$\begin{array}{rcl}
\sum_{u \in S} \omega'(u,v) & = & \sum_{u \in S} \sum_{s,t \in S} \omega(s,t) \cdot \omega_{s,t}(u,v) \\
& = & \sum_{s,t \in S} \omega(s,t) \sum_{u \in S} \omega_{s,t}(u,v) \\
& = & \sum_{s,t \in S} \omega(s,t) \cdot \Theta_t(v) \\
& = & \sum_{t \in S} \Theta(t) \cdot \Theta_t(v) \\
& = & \Theta'(v)
\end{array}$$

for any $v \in S$. Similarly, we can infer that $\sum_{v \in S} \omega'(u,v) = \Delta'(u)$ for any $u \in S$. Therefore, we have $\omega' \in \Omega(\Delta', \Theta')$, from which we can do the following reasoning:

$$\begin{array}{rcl} K(\mathbf{d}_s)(\Delta', \Theta') & \leq & \sum_{u,v \in S} \omega'(u,v) \cdot \mathbf{d}_s(u,v) \\ & = & \sum_{u,v \in S} \sum_{s,t \in S} \omega(s,t) \cdot \omega_{s,t}(u,v) \cdot \mathbf{d}_s(u,v) \\ & = & \sum_{s,t \in S} \omega(s,t) \sum_{u,v \in S} \omega_{s,t}(u,v) \cdot \mathbf{d}_s(u,v) \\ & = & \sum_{s,t \in S} \omega(s,t)\, K(\mathbf{d}_s)(\Delta_s, \Theta_t) \\ & \leq & \sum_{s,t \in S} \omega(s,t)\, \mathbf{d}_s(s,t) \\ & = & K(\mathbf{d}_s)(\Delta, \Theta). \end{array}$$

In summary, we have shown that $K(\mathbf{d}_s)$ is a distribution-based bisimulation metric. This completes the proof. □

Then we arrive at the following theorem.

Theorem 5.2 *Let Δ, Θ be two distributions on a pLTS. Then $\mathbf{d}_d(\Delta, \Theta) \leq K(\mathbf{d}_s)(\Delta, \Theta)$.*

Proof: Let Δ, Θ be two distributions on a pLTS $\mathcal{A}$. Let $\Delta^\perp, \Theta^\perp$ be the corresponding distributions on the fully enabled pLTS $\mathcal{A}^\perp$. It follows from Lemma 5.2(2)-(3) and Theorem 5.1 that

$$\mathbf{d}_d(\Delta, \Theta) = \mathbf{d}_d(\Delta^\perp, \Theta^\perp) \leq K(\mathbf{d}_s)(\Delta^\perp, \Theta^\perp) = K(\mathbf{d}_s)(\Delta, \Theta).$$

□

6 Bisimulations

The kernel of $\mathbf{d}_s$ (resp. $\mathbf{d}_d$) is the state-based (resp. distribution-based) bisimilarity, denoted by $\sim_s$ (resp. $\sim_d$). They can be defined in a more direct way. The definition of $\sim_s$ requires us to lift a relation on states to be a relation on distributions. There are several different but equivalent formulations of the lifting operation, and they are closely related to the Kantorovich metric; see [11] for more details. The following one is taken from [16].

Definition 6.1 *Given two sets S, T and a binary relation $\mathcal{R} \subseteq S \times T$, we define the lifted binary relation $\mathcal{R}^\dagger \subseteq \mathcal{D}_{sub}(S) \times \mathcal{D}_{sub}(T)$ as the smallest relation satisfying the following two rules:*

1. *$s\, \mathcal{R}\, t$ implies $\overline{s}\, \mathcal{R}^\dagger\, \overline{t}$;*

2. *$\Delta_i \mathcal{R}^\dagger \Theta_i$ for all $i \in I$ implies $(\sum_{i\in I} p_i \cdot \Delta_i) \mathcal{R}^\dagger (\sum_{i\in I} p_i \cdot \Theta_i)$, where I is a finite index set and $\sum_{i\in I} p_i \leq 1$.*

The state-based bisimilarity $\sim_s$ is essentially Larsen and Skou's probabilistic bisimilarity [37], which is originally defined for deterministic systems.

Definition 6.2 *Let $\sim_s \subseteq S \times S$ be the largest symmetric relation such that if $s \sim_s t$ and $s \xrightarrow{a} \Delta$ then there exists some $t \xrightarrow{a} \Theta$ with $\Delta\ (\sim_s)^\dagger\ \Theta$.*

The distribution-based bisimilarity $\sim_d$ is proposed in [14] as a sound and complete coinductive proof technique for linear contextual equivalence, a natural extensional behavioural equivalence for functional programs.

Definition 6.3 *Let $\sim_d \subseteq \mathcal{D}_{sub}(S) \times \mathcal{D}_{sub}(S)$ be the largest symmetric relation such that if $\Delta \sim_d \Theta$ then $|\Delta| = |\Theta|$ and $\Delta \xrightarrow{a} \Delta'$ implies the existence of some Θ' such that $\Theta \xrightarrow{a} \Theta'$ and $\Delta' \sim_d \Theta'$.*

Notice that, for any states $s, t \in S$, the following three statements are equivalent:

(i) $s \sim_s t$;

(ii) $\mathbf{d}_s(s,t) = 0$;

(iii) $[\![\varphi]\!](s) = [\![\varphi]\!](t)$ for any formula $\varphi \in \mathcal{L}^{\mathrm{S}}$.

Similarly, for any subdistributions $\Delta, \Theta \in \mathcal{D}_{sub}(S)$, the following three statements are equivalent:

(i) $\Delta \sim_d \Theta$;

(ii) $\mathbf{d}_d(\Delta, \Theta) = 0$;

(iii) $[\![\psi]\!](\Delta) = [\![\psi]\!](\Theta)$ for any formula $\psi \in \mathcal{L}^{D*}$.

Although the state-based bisimilarity is widely accepted, there is no general agreement on what is a good notion of distribution-based bisimilarity. In the literature [29, 15, 24, 20, 23, 31], several variations of distribution-based bisimulations have been proposed. Some of them are defined for pLTSs with states labelled by atomic propositions. We adapt them to our setting so as to compare them with $\sim_d$.

In a pLTS $(S, L, \rightarrow)$, a transition goes from a state to a distribution, e.g. $s \xrightarrow{a} \Delta$. In order to lift $\rightarrow$ to be a relation between distributions, e.g. $\Delta \xrightarrow{a} \Theta$, usually we need to decide whether

(i) to require all the states in the support of Δ to perform action a;

(ii) to combine transitions with the same label, which we explain below.

In [24, 20, 23] both (i) and (ii) are imposed, while in [31] and also in our definition of $\sim_d$ (i) is not used. The condition (ii) is built in Definition 4.1 but partially used in [31], as we will see in the sequel. Let $\{s \xrightarrow{a} \Delta_i\}_{i\in I}$ be a collection of transitions, and $\{p_i\}_{i\in I}$ be a collection of probabilities with $\sum_{i\in I} p_i = 1$. Then $s \xrightarrow{a}_C (\sum_{i\in I} p_i \cdot \Delta_i)$ is called a *combined transition* [43]. Let us write $\Delta \xrightarrow{a}_C \Theta$ if $s \xrightarrow{a}_C \Delta_s$ for each $s \in \lceil\Delta\rceil$ and $\Theta = \sum_{s\in\lceil\Delta\rceil} \Delta(s)\cdot\Delta_s$.

Remark 6.1 *An equivalent way of defining combined transitions is to use Definition 4.1. We have that* $s \xrightarrow{a}_C \Delta$ *iff* $\overline{s} \xrightarrow{a} \Delta$ *and* $|\Delta| = 1$; $\Delta \xrightarrow{a}_C \Theta$ *iff* $\Delta \xrightarrow{a} \Theta$ *and* $|\Delta| = |\Theta|$.

Note that a simple way of comparing subdistributions is to lift the state-based bisimilarity and use the relation $(\sim_s)^\dagger$. That relation can be slightly weakened by using the combined transition $t \xrightarrow{a}_C \Theta$ in place of $t \xrightarrow{a} \Theta$ in Definition 6.2 to get a coarser notion of state-based bisimilarity called strong probabilistic bisimulation in [43], written $\sim_s'$, and then lifting it to subdistributions to finally obtain $(\sim_s')^\dagger$. This is essentially the relation investigated in [29]. However, most distribution-based bisimilarities proposed in the literature directly compare the transitions between (sub)distributions, so there is no need of defining certain relations on states and then lift them to subdistributions. Below we recall four typical proposals.

Firstly, we adapt the bisimulation of [24] to our setting. Let $(S, A, \rightarrow)$ be a pLTS, we extend it to be a fully enabled pLTS $(S_\perp, Act \cup \{\perp\}, \rightarrow_\perp)$ according to Definition 5.1.

Definition 6.4 *Let* $\sim_1 \subseteq \mathcal{D}(S_\perp) \times \mathcal{D}(S_\perp)$ *be the largest symmetric relation such that* $\Delta \sim_1 \Theta$ *implies*

1. $\Delta(S) = \Theta(S)$,

2. for each $a \in A$, *whenever* $\Delta \xrightarrow{a}_C \Delta'$, *there exists* Θ' *with* $\Theta \xrightarrow{a}_C \Theta'$ *and* $\Delta' \sim_1 \Theta'$.

Secondly, we adapt the bisimulation in [29, 15] for subdistributions.

Definition 6.5 *Let* $\sim_2 \subseteq \mathcal{D}_{sub}(S) \times \mathcal{D}_{sub}(S)$ *be the largest symmetric relation such that* $\Delta \sim_2 \Theta$ *implies, for all finite sets of probabilities* $\{p_i \mid i \in I\}$ *satisfying* $\sum_{i\in I} p_i \leq 1$, *1.* $|\Delta| = |\Theta|$,

2. whenever $\Delta \xrightarrow{a}_C \Delta'$, *there exists* Θ' *with* $\Theta \xrightarrow{a}_C \Theta'$ *and* $\Delta' \sim_2 \Theta'$,

3. whenever $\Delta = \sum_{i \in I} p_i \cdot \Delta_i$, *for any subdistributions* Δ_i, *there are some subdistributions* Θ_i *such that* $\Theta = \sum_{i \in I} p_i \cdot \Theta_i$ *and* $\Delta_i \sim_2 \Theta_i$ *for each* $i \in I$.

Thirdly, we adapt the bisimulation given in [20] to pLTSs. A subdistribution is *consistent*, if $EA(s) = EA(t)$ for any $s, t \in \lceil \Delta \rceil$. That is, all the states in the support of Δ have the same set of enabled actions.

Definition 6.6 *Let* $\sim_3 \subseteq \mathcal{D}_{sub}(S) \times \mathcal{D}_{sub}(S)$ *be the largest symmetric relation such that* $\Delta \sim_3 \Theta$ *implies*

1. $|\Delta| = |\Theta|$,

2. whenever $\Delta \xrightarrow{a}_C \Delta'$, *there exists* Θ' *with* $\Theta \xrightarrow{a}_C \Theta'$ *and* $\Delta' \sim_3 \Theta'$,

3. if Δ *is not consistent, there exist decompositions* $\Delta = \sum_{i \in I} p_i \cdot \Delta_i$ *and* $\Theta = \sum_{i \in I} p_i \cdot \Theta_i$ *such that* $\Delta_i \sim_3 \Theta_i$ *for each* $i \in I$.

Finally, we adapt the bisimulation of [31]. Let $\mathtt{A}$ be a set of labels. We write $s \xrightarrow{\mathtt{A}} \Delta$ if $s \xrightarrow{a}_C \Delta$ for some $a \in \mathtt{A}$ and denote by $S_{\mathtt{A}} = \{s \mid \exists \Delta .\ s \xrightarrow{A} \Delta\}$ the set of states that can perform some action from $\mathtt{A}$. Then we define a transition relation for distributions by letting $\Delta \xrightarrow{\mathtt{A}} \Theta$ if $s \xrightarrow{\mathtt{A}} \Delta_s$ for each $s \in S_{\mathtt{A}} \cap \lceil \Delta \rceil$ and $\Theta = \frac{1}{\Delta(S_{\mathtt{A}})} \sum_{s \in S_{\mathtt{A}} \cap \lceil \Delta \rceil} \Delta(s) \cdot \Delta_s$.

Definition 6.7 *Let* $\sim_4 \subseteq \mathcal{D}_{sub}(S) \times \mathcal{D}_{sub}(S)$ *be the largest symmetric relation such that* $\Delta \sim_4 \Theta$ *implies*

1. $|\Delta| = |\Theta|$ *and* $\Delta(S_{\mathtt{A}}) = \Theta(S_{\mathtt{A}})$ *for any* $\mathtt{A} \subseteq L$,

2. for each $\mathtt{A} \subseteq L$, *whenever* $\Delta \xrightarrow{\mathtt{A}} \Delta'$, *there exists* Θ' *with* $\Theta \xrightarrow{\mathtt{A}} \Theta'$ *and* $\Delta' \sim_4 \Theta'$.

The lifting operation given in Definition 6.1 enjoys a few useful properties [11, Section 3.3].

Proposition 6.1 *Let* Δ *and* Θ *be two subdistributions over* S *and* T, *respectively, and* $\mathcal{R} \subseteq S \times T$. *Then* $\Delta\ \mathcal{R}^\dagger\ \Theta$ *if and only if there are two collections of states,* $\{s_i\}_{i \in I}$ *and* $\{t_i\}_{i \in I}$, *and a collection of probabilities* $\{p_i\}_{i \in I}$, *for some finite index set* I, *such that* $\sum_{i \in I} p_i \leq 1$ *and* Δ, Θ *can be decomposed as follows:*

1. $\Delta = \sum_{i\in I} p_i \cdot \overline{s_i}$
2. $\Theta = \sum_{i\in I} p_i \cdot \overline{t_i}$
3. *for each* $i \in I$ *we have* $s_i \mathcal{R} t_i$. □

Proposition 6.2 *If* $\mathcal{R}_1 \subseteq \mathcal{R}_2$ *then* $(\mathcal{R}_1)^\dagger \subseteq (\mathcal{R}_2)^\dagger$. □

Proposition 6.3 *Suppose* $\mathcal{R} \subseteq S \times T$ *and* $\sum_{i\in I} p_i \leq 1$. *If* $(\sum_{i\in I} p_i \cdot \Delta_i) \mathcal{R}^\dagger \Theta$ *then* $\Theta = \sum_{i\in I} p_i \cdot \Theta_i$ *for some set of distributions* $\{\Theta_i\}_{i\in I}$ *such that* $\Delta_i \mathcal{R}^\dagger \Theta_i$ *for each* $i \in I$. □

Proposition 6.4 $\sim_4 \subset \sim_d$.

Proof: Let us construct the following relation

$$\mathcal{R} = \{(p \cdot \Delta,\ p \cdot \Theta) \mid p \in [0,1] \wedge \Delta \sim_4 \Theta\}$$

and show that it is a distribution-based bisimulation in the sense of Definition 6.3. Suppose $(p \cdot \Delta,\ p \cdot \Theta) \in \mathcal{R}$ for some subdistributions Δ, Θ with $\Delta \sim_4 \Theta$ and $p \in [0,1]$. We observe that $|p \cdot \Delta| = p \cdot |\Delta| = p \cdot |\Theta| = |p \cdot \Theta|$. Now let $p \cdot \Delta \xrightarrow{a} \Delta'$. It is necessarily the case that $\Delta' = p \cdot \Delta''$ and $\Delta \xrightarrow{a} \Delta''$ for some Δ''. Then for each $s \in \lceil \Delta \rceil$ there exists some Δ_s such that $\Delta'' = \sum_{s \in \lceil \Delta \rceil} \Delta(s) \cdot \Delta_s$ with $\overline{s} \xrightarrow{a} \Delta_s$, i.e. either $s \xrightarrow{a}_C \Delta_s$ or $\Delta_s = \varepsilon$ if $s \not\xrightarrow{a}$. Note that $s \xrightarrow{a}_C \Delta_s$ if and only if $s \xrightarrow{\{a\}} \Delta_s$. It follows that

$$\Delta \xrightarrow{\{a\}} \Delta''' = \frac{1}{\Delta(S_{\{a\}})} \sum_{s \in S_{\{a\}} \cap \lceil \Delta \rceil} \Delta(s) \cdot \Delta_s = \frac{1}{\Delta(S_{\{a\}})} \Delta''.$$

Since $\Delta \sim_4 \Theta$, there exists some Θ''' with $\Theta \xrightarrow{\{a\}} \Theta'''$ and $\Delta''' \sim_4 \Theta'''$. By definition Θ''' must be in the form

$$\frac{1}{\Theta(S_{\{a\}})} \sum_{s \in S_{\{a\}} \cap \lceil \Theta \rceil} \Theta(s) \cdot \Theta_s$$

with $s \xrightarrow{\{a\}} \Theta_s$, i.e. $s \xrightarrow{a}_C \Theta_s$, for any $s \in S_{\{a\}} \cap \lceil \Theta \rceil$. By taking $\Theta_s = \varepsilon$ for any s with $s \notin S_{\{a\}}$, it follows that $\Theta \xrightarrow{a} \Theta'' = \sum_{s \in \lceil \Theta \rceil} \Theta(s) \cdot \Theta_s = \Theta(S_{\{a\}}) \cdot \Theta'''$. We see from $\Delta \sim_4 \Theta$ that $\Delta(S_{\{a\}}) = \Theta(S_{\{a\}})$. In summary, we can infer that

$$p \cdot \Delta \xrightarrow{a} p \cdot \Delta'' = p \cdot \Delta(S_{\{a\}}) \cdot \Delta'''$$
$$p \cdot \Theta \xrightarrow{a} p \cdot \Theta'' = p \cdot \Delta(S_{\{a\}}) \cdot \Theta'''$$

It follows that from $\Delta''' \sim_4 \Theta'''$ that $(p\cdot\Delta(S_{\{a\}})\cdot\Delta''',\ p\cdot\Delta(S_{\{a\}})\cdot\Theta''') \in\mathcal{R}$. Therefore, we have verified that $\mathcal{R}\subseteq\sim_d$, thus $\sim_4\subseteq\sim_d$.

We now prove that $\sim_d\not\subseteq\sim_4$. Consider the example in Figure 4. From state s there is a unique transition $s \xrightarrow{a} \Delta$ with $\Delta = \frac{1}{3}\overline{s_1} + \frac{1}{3}\overline{s_2} + \frac{1}{3}\overline{s_3}$. This can be matched by $t \xrightarrow{a} \Theta$, where $\Theta = \frac{2}{3}\overline{t_1} + \frac{1}{3}\overline{t_2}$, because $\Delta \sim_d \Theta$ holds. To see this, we observe that Δ can initiate three transitions: $\Delta \xrightarrow{a} \frac{2}{3}\overline{s_a}$, $\Delta \xrightarrow{b} \frac{2}{3}\overline{s_b}$, and $\Delta \xrightarrow{c} \frac{1}{3}\overline{s_c}$; they can be matched by $\Theta \xrightarrow{a} \frac{2}{3}\overline{t_a}$, $\Theta \xrightarrow{b} \frac{2}{3}\overline{t_b}$, and $\Theta \xrightarrow{c} \frac{1}{3}\overline{t_c}$, respectively. Similarly, the three outgoing transitions from Θ can be matched by the three transitions of Δ. Therefore, we have verified that $\overline{s} \sim_d \overline{t}$. However, we have $\Delta \not\sim_4 \Theta$. From Δ we have the transition $\Delta \xrightarrow{\{a,b\}} \Delta' \equiv \frac{2}{3}\cdot\overline{s_a} + \frac{1}{3}\cdot\overline{s_b}$. From Θ there are two transitions labelled with $\{a,b\}$, namely $\Theta \xrightarrow{\{a,b\}} \overline{t_a}$ and $\Theta \xrightarrow{\{a,b\}} \overline{t_b}$. Neither of them is able to match the transition from Δ. To see this, we observe that $\Delta'(S_{\{b\}}) = \frac{1}{3}$ while $\overline{t_a}(S_{\{b\}}) = 0$, and $\Delta'(S_{\{a\}}) = \frac{2}{3}$ while $\overline{t_b}(S_{\{a\}}) = 0$. It follows that $\overline{s} \not\sim_4 \overline{t}$. □

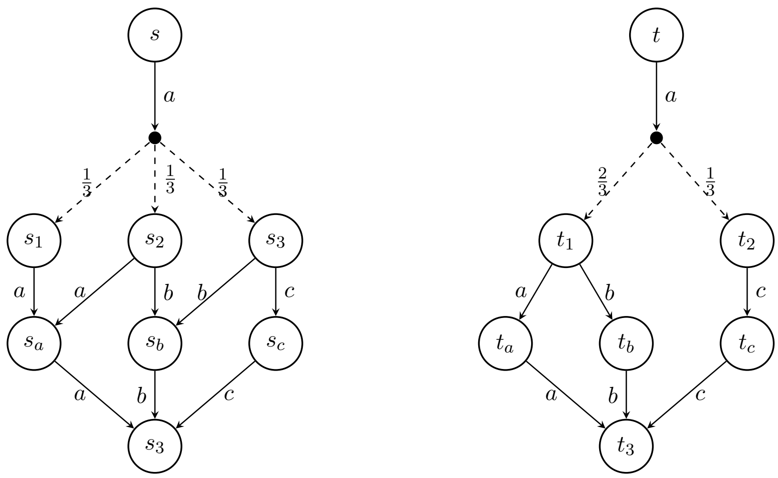

Figure 4: $\overline{s} \sim_d \overline{t}$ but $\overline{s} \not\sim_3 \overline{t}$ and $\overline{s} \not\sim_4 \overline{t}$.

Proposition 6.5 *Suppose* $\Delta, \Theta \in \mathcal{D}(S)$. *Then* $\Delta \sim_1 \Theta$ *if and only if* $\Delta \sim_d \Theta$.

Proof: Let us define the following two relations:

$$\begin{aligned}\mathcal{R}_1 &= \{(\Delta,\Theta) \mid \Delta,\Theta\in\mathcal{D}_{sub}(S) \wedge p\in[0,1] \wedge (\Delta + p\cdot\overline{\bot} \sim_1 \Theta + p\cdot\overline{\bot})\}\\ \mathcal{R}_2 &= \{(\Delta + p\cdot\overline{\bot}, \Theta + p\cdot\overline{\bot}) \mid \Delta,\Theta\in\mathcal{D}_{sub}(S) \wedge p\in[0,1] \wedge \Delta \sim_d \Theta\}.\end{aligned}$$

We can prove that $\mathcal{R}_1 \subseteq \sim_d$ and $\mathcal{R}_2 \subseteq \sim_1$. As an example, let us consider $\mathcal{R}_1$ and suppose $(\Delta, \Theta) \in \mathcal{R}_1$ with $\Delta \xrightarrow{a} \Delta'$. Then there is some probability p with $\Delta + p \cdot \overline{\bot} \sim_1 \Theta + p \cdot \overline{\bot}$. By Lemma 4.1 $\Delta' = \sum_{s \in \lceil \Delta \rceil} \Delta(s) \cdot \Delta_s$ and for each $s \in \lceil \Delta \rceil$ we have $\overline{s} \xrightarrow{a} \Delta_s$, i.e. either $s \xrightarrow{a}_C \Delta_s$ or $\Delta_s = \varepsilon$. Let $\Delta'' = \Delta' + (|\Delta| - |\Delta'|) \cdot \overline{\bot}$. It is easy to see that

$$\Delta + p \cdot \overline{\bot} \xrightarrow{a}_C \Delta'' + p \cdot \overline{\bot}.$$

It follows from $\Delta + p \cdot \overline{\bot} \sim_1 \Theta + p \cdot \overline{\bot}$ that

$$\Theta + p \cdot \overline{\bot} \xrightarrow{a}_C \Theta'' + p \cdot \overline{\bot}$$

and $\Delta'' + p \cdot \overline{\bot} \sim_1 \Theta'' + p \cdot \overline{\bot}$ for some Θ''. Observe that Θ'' must take the form $\Theta' + (|\Theta| - |\Theta'|) \cdot \overline{\bot}$ with $\Theta' = \sum_{s \in \lceil \Theta \rceil} \Theta(s) \cdot \Theta_s$ and for each $s \in \lceil \Theta \rceil$ either $s \xrightarrow{a}_C \Theta_s$ or $\Theta_s = \varepsilon$. It follows that $\Theta \xrightarrow{a} \Theta'$.

Since Δ, Θ are subdistributions over S and $\Delta + p \cdot \overline{\bot} \sim_1 \Theta + p \cdot \overline{\bot}$, we know that

$$|\Delta| = \Delta(S) = (\Delta + p \cdot \overline{\bot})(S) = (\Theta + p \cdot \overline{\bot})(S) = \Theta(S) = |\Theta|.$$

It follows from $\Delta'' + p \cdot \overline{\bot} \sim_1 \Theta'' + p \cdot \overline{\bot}$ that

$$(\Delta' + (|\Delta| - |\Delta'|) \cdot \overline{\bot} + p \cdot \overline{\bot}) \;\sim_1\; (\Theta' + (|\Theta| - |\Theta'|) \cdot \overline{\bot} + p \cdot \overline{\bot}).$$

As a result, we obtain $(\Delta', \Theta') \in \mathcal{R}_1$ and

$$\begin{aligned} |\Delta'| &= \Delta'(S) = (\Delta' + (|\Delta| - |\Delta'|) \cdot \overline{\bot} + p \cdot \overline{\bot})(S) \\ &= (\Theta' + (|\Theta| - |\Theta'|) \cdot \overline{\bot} + p \cdot \overline{\bot})(S) = \Theta'(S) = |\Theta'|. \end{aligned}$$

Therefore, we have established that $\mathcal{R}_1 \subseteq \sim_d$. By similar arguments, it can be shown that $\mathcal{R}_2 \subseteq \sim_1$. □

Proposition 6.6 $(\sim_s)^\dagger \subset (\sim'_s)^\dagger = \sim_2 \subset \sim_3 \subset \sim_d$.

Proof: Since $\sim'_s$ allows for combined transitions while $\sim_s$ uses plain transitions only, it is obvious that $\sim_s \subseteq \sim'_s$. By the monotonicity of the lifting operation, Proposition 6.2, we can infer that $(\sim_s)^\dagger \subseteq (\sim'_s)^\dagger$. Moreover, the inclusion is strict. For example, consider the two point distributions $\overline{s}$ and $\overline{t}$ in Figure 5, we have $\overline{s}\ (\sim'_s)^\dagger\ \overline{t}$ but not $\overline{s}\ (\sim_s)^\dagger\ \overline{t}$ because neither $t \xrightarrow{a} \overline{t_1}$ nor $t \xrightarrow{a} \overline{t_2}$ matches the transition $s \xrightarrow{a} (\frac{1}{2}\overline{s_1} + \frac{1}{2}\overline{s_2})$, but a combination of them will do.

Next, let us prove $(\sim'_s)^\dagger \subseteq \sim_2$. Suppose Δ, Θ are two subdistributions with $\Delta\ (\sim'_s)^\dagger\ \Theta$. By Proposition 6.1 we can decompose Δ and Θ as follows:

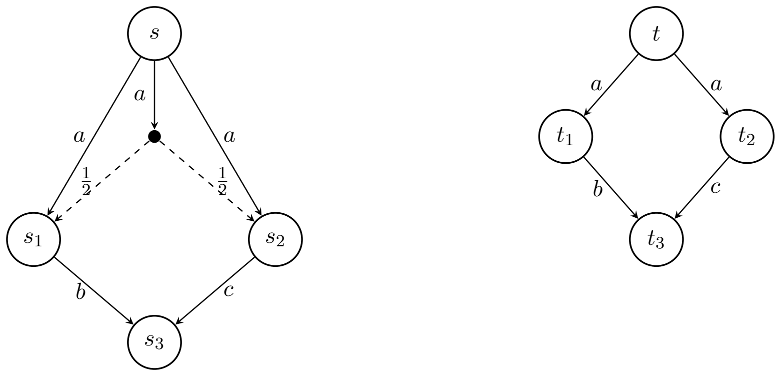

Figure 5: $\overline{s}\ (\sim'_s)^\dagger\ \overline{t}$ but not $\overline{s}\ (\sim_s)^\dagger\ \overline{t}$.

- $\Delta = \sum_{i\in I} p_i \cdot \overline{s_i}$
- $\Theta = \sum_{i\in I} p_i \cdot \overline{t_i}$
- for each $i \in I$ we have $s_i \sim'_s t_i$.

It is obvious that $|\Delta| = \sum_{i\in I} p_i = |\Theta|$. It remains to check that Δ and Θ can match each other's transitions. Suppose $\Delta \xrightarrow{a}_C \Delta'$. Then $\Delta' = \sum_{s\in\lceil\Delta\rceil} \Delta(s)\cdot \Delta_s$, where for each $s \in \lceil\Delta\rceil$ we have $s \xrightarrow{a}_C \Delta_s$. By Proposition 6.3 we can decompose Θ as

$$\Theta = \sum_{s\in\lceil\Delta\rceil} \Delta(s)\cdot\Theta_s \tag{18}$$

such that $\overline{s}\ (\sim'_s)^\dagger\ \Theta_s$ for each $s \in \lceil\Delta\rceil$. By Proposition 6.1 we can derive that $s \sim'_s t_s$ for each $t_s \in \lceil\Theta_s\rceil$. From $s \xrightarrow{a}_C \Delta_s$, we can find some matching transition $t_s \xrightarrow{a}_C \Theta_{t_s}$ with $\Delta_s\ (\sim_s)^\dagger\ \Theta_{t_s}$ for each $t_s \in \lceil\Theta'_s\rceil$. It follows that

$$\Delta_s\ (\sim'_s)^\dagger\ (\sum_{t_s\in\lceil\Theta_s\rceil} \Theta_s(t_s)\cdot\Theta_{t_s}). \tag{19}$$

Let $\Theta'_s = \sum_{t_s\in\lceil\Theta_s\rceil} \Theta_s(t_s)\cdot\Theta_{t_s}$ and $\Theta' = \sum_{s\in\lceil\Delta\rceil} \Delta(s)\cdot\Theta'_s$. Then $\Theta_s \xrightarrow{a}_C \Theta'_s$ is clearly a valid transition for each distribution Θ_s where $s \in \lceil\Delta\rceil$. Combine this with (18), we obtain

$$\Theta \xrightarrow{a}_C \Theta'. \tag{20}$$

From (19), we derive that

$$\Delta'\ (\sim'_s)^\dagger\ \Theta'. \tag{21}$$

By Proposition 6.3, any decomposition of Δ as $\sum_{i\in I} p_i \cdot \Delta_i$ can be matched by some decomposition of Θ as $\sum_{i\in I} p_i \cdot \Theta_i$ with $\Delta_i \ (\sim'_s)^\dagger \ \Theta_i$ as desired. Therefore, we have completed the proof of $(\sim'_s)^\dagger \subseteq \sim_2$.

In order to prove the other inclusion, $\sim_2 \subseteq (\sim'_s)^\dagger$, it suffices to construct the following relation

$$\mathcal{R} = \ \{(s,t) \mid \overline{s} \sim_2 \overline{t}\}$$

and show that it is a strong probabilistic bisimulation, which means $\mathcal{R} \subseteq \sim'_s$. The proof makes use of the property that $\Delta \sim_2 \Theta$ implies $\Delta \ \mathcal{R}^\dagger \ \Theta$.

It is easy to see that $\sim_3$ is a relaxation of $\sim_2$ by requiring decompositions for inconsistent subdistributions rather than for any subdistribution in general. Furthermore, $\sim_3$ is strictly coarser than $\sim_2$. Consider the two states s and t in Figure 6. We see that $\overline{s} \not\sim_2 \overline{t}$ because the point distribution $\overline{s_1}$ reachable from s is not related to the subdistribution $(\frac{1}{2} \cdot \overline{t_1} + \frac{1}{2} \cdot \overline{t_2})$ reachable from t if their decompositions need to be compared: the former cannot be decomposed into two subdistributions that can mimic $\overline{t_1}$ and $\overline{t_2}$, respectively. However, it is straightforward to check that $\overline{s} \sim_3 \overline{t}$.

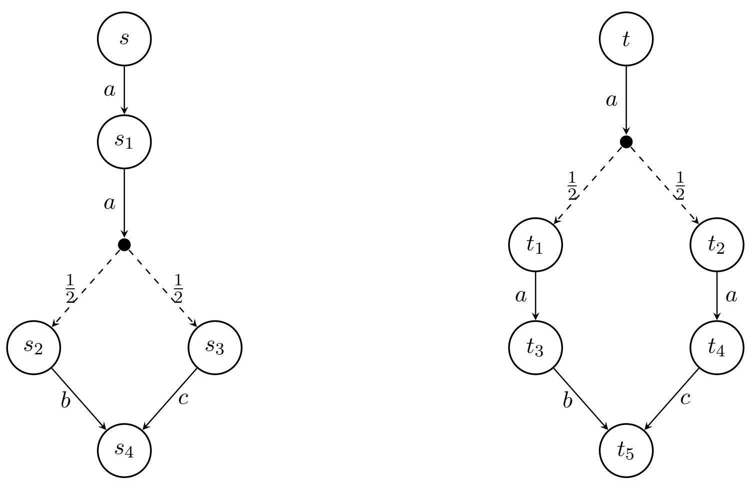

Figure 6: $\overline{s} \sim_3 \overline{t}$ and $\overline{s} \sim_4 \overline{t}$ but $\overline{s} \not\sim_2 \overline{t}$.

In [22, Theorem 7.1.1] it is proven that $\sim_3$ is strictly included in $\sim_1$. But Proposition 6.5 says that $\sim_1$ coincides with $\sim_d$. Therefore, $\sim_3$ is strictly included in $\sim_d$. As a matter of fact, it is also not difficult to give a direct proof. □

Proposition 6.7 $\sim_4$ *is incomparable with the three relations* $(\sim_s)^\dagger$, $\sim_2$ *and* $\sim_3$.

Proof: In [22, Theorem 7.1.2] it is shown that $\sim_3$ is incomparable with $\sim_4$. Note that $\sim_4 \not\subseteq \sim_3$ implies $\sim_4 \not\subseteq \sim_2$ and $\sim_4 \not\subseteq (\sim_s)^\dagger$ by Proposition 6.6.

Let us consider the diagram in Figure 7. Let $\Delta = \frac{1}{2} \cdot \overline{s_1} + \frac{1}{2} \cdot \overline{s_2}$. Observe that $s_1 \sim_s s_2$ and thus $\overline{s_1}\ (\sim_s)^\dagger\ \Delta$. By Proposition 6.6, we also have $\overline{s_1} \sim_2 \Delta$ and $\overline{s_1} \sim_3 \Delta$. However, we have $\overline{s_1} \not\sim_4 \Delta$ because the transition $\Delta \xrightarrow{\{a,b\}} \frac{1}{2}\overline{s_3} + \frac{1}{2}\overline{s_6}$ can be matched by neither $\overline{s_1} \xrightarrow{\{a,b\}} \overline{s_3}$ nor $\overline{s_1} \xrightarrow{\{a,b\}} \overline{s_4}$, the only two $\{a, b\}$-labelled transitions from $\overline{s_1}$. □

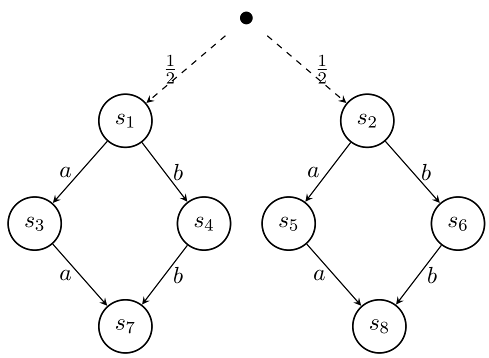

Figure 7: $\overline{s_1}\ (\sim_s)^\dagger\ (\frac{1}{2} \cdot \overline{s_1} + \frac{1}{2} \cdot \overline{s_2})$ but $\overline{s_1} \not\sim_4 (\frac{1}{2} \cdot \overline{s_1} + \frac{1}{2} \cdot \overline{s_2})$.

The last four propositions can be recapitulated by the following theorem.

Theorem 6.1 *Figure 8 depicts the relationship between the seven bisimilarities for distributions mentioned above.* □

If we confine ourselves to deterministic pLTSs, then combined transitions add nothing new to ordinary transitions and thus $\sim_s'$ degenerates into $\sim_s$, but the rest of Figure 8 remains unchanged.

7 Other Related Work

Metrics for probabilistic transition systems are first suggested by Giacalone *et al.* [28] to formalize a notion of distance between processes. They are used also in [36, 39] to give denotational semantics for deterministic models. De

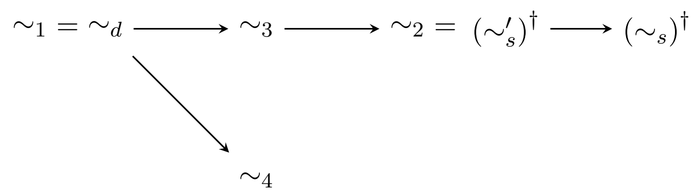

Figure 8: Relationship between the seven bisimilarities for distributions. An arrow pointing from one relation to another means that the former relation is strictly coarser than the latter. Two relations are incomparable if there is no path from one to the other.

Vink and Rutten [9] show that discrete probabilistic transition systems can be viewed as coalgebras. They consider the category of complete ultrametric spaces. Similar ultrametric spaces are considered by den Hartog in [10]. In [51] Ying proposes a notion of bisimulation index for the usual labelled transition systems, by using ultrametrics on actions instead of using pseudo-metrics on states. A quantitative linear-time-branching-time spectrum for non-probabilistic systems is given in [21].

Metrics for deterministic systems are extensively studied. Desharnais *et al.* [17] propose a logical pseudometric for labelled Markov chains, which is a deterministic model of probabilistic systems. A similar pseudometric is defined by van Breugel and Worrell [49] via the terminal coalgebra of a functor based on a metric on the space of Borel probability measures. Essentially the same metric is investigated in the setting of continuous Markov decision processes [26]. The metric of [17, 50, 26] works for continuous probabilistic transition systems, while in this work we concentrate on discrete systems with nondeterminism. In the future it would be interesting to see how to generalise our results to continuous systems. In [48] van Breugel and Worrell present a polynomial-time algorithm to approximate their coalgebraic distances. Furthermore, van Breugel *et al.* propose an algorithm to approximate a behavioural pseudometric without discount [47]. In [25] a sampling algorithm for calculating bisimulation distances in Markov decision processes is shown to have good performance. Later on, more efficient algorithms for computing probabilistic bisimilarity distances for probabilistic automata have been developed in, e.g., [45, 1]. In [7, 8] the probabilistic bisimulation metric on game structures is characterised by a quantitative μ-calculus. Algorithms for game metrics are proposed in [3, 41]. A notion of bisimulation distance for distributions is proposed in [24]. It is defined for full distributions only and the definition itself has to be given in terms of fully enabled transition systems. Our distribution-based bisimulation metric generalises it to subdistributions, and allowing transitions between subdistributions has the advantage of allowing our definition to be more direct.

Metrics for nondeterministic probabilistic systems are considered in [18], where Desharnais *et al.* deal with labelled concurrent Markov chains (similar to pLTSs, this model can be captured by the simple probabilistic automata of [42]). They show that the greatest fixed point of a monotonic function on pseudometrics corresponds to the weak probabilistic bisimilarity of [40]. In [27] a notion of uniform continuity is proposed to be an appropriate property of probabilistic processes for compositional reasoning with respect to $\mathbf{d}_s$. In [44] a notion of trace metric is proposed for pLTSs and a tool is developed to compute the trace metric. In [2] the boolean-valued logic from [13] is used to characterise state-based bisimulation metrics. It crucially relies on distribution formulae of the form $\bigoplus_{i\in I} p_i\varphi_i$, which is demanding in the sense that if Δ satisfies that formula then there is some decomposition $\Delta = \sum_{i\in I} p_i \cdot \Delta_i$ such that for each $i \in I$ all the states in the support of Δ_i must satisfy φ_i.

Metrics for other quantitative models are also investigated. In [12] a notion of bisimulation metric is proposed that extends the approach of [18, 17] to a more general framework called action-labelled quantitative transition systems. In [6] de Alfaro *et al.* consider metric transition systems in which the propositions at each state are interpreted as elements of metric spaces. In that setting, trace equivalence and bisimulation give rise to linear and branching distances that can be characterised by quantitative versions of linear-time temporal logic [38] and the μ-calculus [35].

8 Concluding Remarks

We have considered two behavioural pseudometrics for probabilistic labelled transition systems where nondeterminism and probabilities co-exist. They correspond to state-based and distribution-based bisimulations. Our modal characterisation of the state-based bisimulation metric is much simpler than an earlier proposal by Desharnais *et al.* since we only use two non-expansive operators, negation and testing, rather than the general class of non-expansive operators. Our modal characterisation of the distribution-based bisimulation metric is new. The characterisations are shown to be sound and complete. We have also shown that the distribution-based bisimulation metric is a lower bound of the state-based bisimulation metric lifted to distributions. In addition, we have compared the bisimilarities entailed by the two metrics with a few other distribution-based bisimilarities.

In the current work we have not distinguished internal actions from external ones. But it is not difficult to make the distinction and abstract away internal actions so as to introduce weak versions of bisimulation metrics. In a finite-state and finitely branching pLTS, the set of subdistributions reachable from a state by weak transitions may be infinite but can be represented by the convex closure of a finite set [11]. This entails that the logical characterisation of weak bisimulation metrics would be similar to those presented here.

Acknowledgement We thank Yuan Feng for helpful discussions on the proof of Theorem 5.1. We also thank the anonymous referees of SETTA 2016 and SACS for their constructive comments. Deng would like to acknowledge the support of the National Natural Science Foundation of China under Grants Nos. 62072176 and 61832015, and Shanghai Trusted Industry Internet Software Collaborative Innovation Center.

References

[1] Giorgio Bacci, Giovanni Bacci, Kim G. Larsen, Radu Mardare, Qiyi Tang, and Franck van Breugel. Computing probabilistic bisimilarity distances for probabilistic automata. *Logical Methods in Computer Science*, 17(1):9:1–9:36, 2021. doi:10.23638/LMCS-17(1:9)2021.

[2] Valentina Castiglioni, Daniel Gebler, and Simone Tini. Logical characterization of bisimulation metrics. In Mirco Tribastone and Herbert Wiklicky, editors, *14th International Workshop Quantitative Aspects of Programming Languages and Systems, QAPL 2016*, volume 227 of *Electronic Proceedings in Theoretical Computer Science*, pages 44–62, 2016. doi:10.4204/EPTCS.227.4.

[3] Krishnendu Chatterjee, Luca De Alfaro, Rupak Majumdar, and Vishwanath Raman. Algorithms for game metrics. *Logical Methods in Computer Science*, 6(3), 2010. doi:10.2168/LMCS-6(3:13)2010.

[4] Rance Cleaveland, S. Purushothaman Iyer, and Murali Narasimha. Probabilistic temporal logics via the modal mu-calculus. *Theoretical Computer Science*, 342(2-3):316–350, 2005. doi:10.1016/j.tcs.2005.03.048.

[5] Pedro R. D'Argenio, Pedro Sánchez Terraf, and Nicolás Wolovick. Bisimulations for non-deterministic labelled Markov processes. *Mathematical Structures in Computer Science*, 22(1):43–68, 2012. doi:10.1017/S0960129511000454.

[6] Luca De Alfaro, Marco Faella, and Mariëlle Stoelinga. Linear and branching system metrics. *IEEE Transactions on Software Engineering*, 35(2):258–273, 2009. doi:10.1109/TSE.2008.106.

[7] Luca de Alfaro, Rupak Majumdar, Vishwanath Raman, and Mariëlle Stoelinga. Game relations and metrics. In *22nd IEEE Symposium on Logic in Computer Science, LICS 2007*, pages 99–108. IEEE Computer Society, 2007. doi:10.1109/LICS.2007.22.

[8] Luca De Alfaro, Rupak Majumdar, Vishwanath Raman, and Mariëlle Stoelinga. Game refinement relations and metrics. *Logical Methods in Computer Science*, 4(3), 2008. doi:10.2168/LMCS-4(3:7)2008.

[9] Erik P. de Vink and Jan J. M. M. Rutten. Bisimulation for probabilistic transition systems: A coalgebraic approach. *Theoretical Computer Science*, 221(1-2):271–293, 1999. doi:10.1016/S0304-3975(99)00035-3.

[10] Jeremy Ian den Hartog. *Probabilistic Extensions of Semantical Models*. PhD thesis, Free University Amsterdam, 2002.

[11] Yuxin Deng. *Semantics of Probabilistic Processes: An Operational Approach*. Springer, 2014. doi:10.1007/978-3-662-45198-4.

[12] Yuxin Deng, Tom Chothia, Catuscia Palamidessi, and Jun Pang. Metrics for action-labelled quantitative transition systems. *Electronic Notes in Theoretical Computer Science*, 153(2):79–96, 2006. doi:10.1016/j.entcs.2005.10.033.

[13] Yuxin Deng and Wenjie Du. Logical, metric, and algorithmic characterisations of probabilistic bisimulation. Technical Report CMU-CS-11-110, Carnegie Mellon University, 2011. arXiv:1103.4577.

[14] Yuxin Deng, Yuan Feng, and Ugo Dal Lago. On coinduction and quantum lambda calculi. In Luca Aceto and David de Frutos-Escrig, editors, *26th International Conference on Concurrency Theory, CONCUR 2015*, volume 42 of *LIPIcs*, pages 427–440. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2015. doi:10.4230/LIPIcs.CONCUR.2015.427.

[15] Yuxin Deng and Matthew Hennessy. On the semantics of Markov automata. *Information and Computation*, 222:139–168, 2013. doi:10.1016/j.ic.2012.10.010.

[16] Yuxin Deng, Rob J. van Glabbeek, Matthew Hennessy, and Carroll Morgan. Testing finitary probabilistic processes. In Mario Bravetti and Gianluigi Zavattaro, editors, *20th International Conference on Concurrency Theory, CONCUR 2009*, volume 5710 of *Lecture Notes in Computer Science*, pages 274–288. Springer, 2009. doi:10.1007/978-3-642-04081-8_19.

[17] Josée Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. Metrics for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, 2004. doi:10.1016/j.tcs.2003.09.013.

[18] Josée Desharnais, Radha Jagadeesan, Vineet Gupta, and Prakash Panangaden. The metric analogue of weak bisimulation for probabilistic processes. In *17th IEEE Symposium on Logic in Computer Science, LICS 2002*, pages 413–422. IEEE Computer Society, 2002. doi:10.1109/LICS.2002.1029849.

[19] Wenjie Du, Yuxin Deng, and Daniel Gebler. Behavioural pseudometrics for nondeterministic probabilistic systems. In Martin Fränzle, Deepak Kapur, and Naijun Zhan, editors, *2nd International Symposium on Dependable Software Engineering: Theories, Tools, and Applications, SETTA 2016*, volume 9984 of *Lecture Notes in Computer Science*, pages 67–84, 2016. doi:10.1007/978-3-319-47677-3_5.

[20] Christian Eisentraut, Jens Chr. Godskesen, Holger Hermanns, Lei Song, and Lijun Zhang. Probabilistic bisimulation for realistic schedulers. In Nikolaj S. Bjørner and Frank S. de Boer, editors, *20th International Symposium on Formal Methods, FM 2015*, volume 9109 of *Lecture Notes in Computer Science*, pages 248–264. Springer, 2015. doi:10.1007/978-3-319-19249-9_16.

[21] Uli Fahrenberg and Axel Legay. The quantitative linear-time–branching-time spectrum. *Theoretical Computer Science*, 538:54–69, 2014. doi:10.1016/j.tcs.2013.07.030.

[22] Yuan Feng, Lei Song, and Lijun Zhang. Distribution-based bisimulation and bisimulation metric in probabilistic automata, 2015. arXiv:1512.05027.

[23] Yuan Feng and Mingsheng Ying. Toward automatic verification of quantum cryptographic protocols. In Luca Aceto and David de Frutos-Escrig, editors, *26th International Conference on Concurrency Theory, CONCUR 2015*, volume 42 of *LIPIcs*, pages 441–455. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2015. doi:10.4230/LIPIcs.CONCUR.2015.441.

[24] Yuan Feng and Lijun Zhang. When equivalence and bisimulation join forces in probabilistic automata. In Cliff B. Jones, Pekka Pihlajasaari, and Jun Sun, editors, *19th International Symposium on Formal Methods, FM 2014*, volume 8442 of *Lecture Notes in Computer Science*, pages 247–262. Springer, 2014. doi:10.1007/978-3-319-06410-9_18.

[25] Norm Ferns, Prakash Panangaden, and Doina Precup. Bisimulation metrics for continuous Markov decision processes. *SIAM Journal on Computing*, 40(6):1662–1714, 2011. doi:10.1137/10080484X.

[26] Norm Ferns, Doina Precup, and Sophia Knight. Bisimulation for Markov decision processes through families of functional expressions. In Franck van Breugel, Elham Kashefi, Catuscia Palamidessi, and Jan Rutten, editors, *Horizons of the Mind. A Tribute to Prakash Panangaden*, volume 8464 of *Lecture Notes in Computer Science*, pages 319–342. Springer, 2014. doi:10.1007/978-3-319-06880-0_17.

[27] Daniel Gebler, Kim Guldstrand Larsen, and Simone Tini. Compositional metric reasoning with probabilistic process calculi. In Andrew M. Pitts, editor, *18th International Conference on Foundations of Software Science and Computation Structures, FoSSaCS 2015*, volume 9034 of *Lecture Notes in Computer Science*, pages 230–245. Springer, 2015. doi:10.1007/978-3-662-46678-0_15.

[28] Alessandro Giacalone, Chi-Chang Jou, and Scott A. Smolka. Algebraic reasoning for probabilistic concurrent systems. In Manfred Broy

and Cliff B. Jones, editors, *IFIP Working Group 2.2, 2.3 Working Conference on Programming Concepts and Methods*, pages 443–458. North-Holland, 1990.

[29] Matthew Hennessy. Exploring probabilistic bisimulations, part I. *Formal Aspects of Computing*, 24(4-6):749–768, 2012. doi:10.1007/s00165-012-0242-7.

[30] Matthew Hennessy and Robin Milner. Algebraic laws for nondeterminism and concurrency. *Journal of the ACM*, 32:137–161, 1985. doi:10.1145/2455.2460.

[31] Holger Hermanns, Jan Krcál, and Jan Kretínský. Probabilistic bisimulation: Naturally on distributions. In Paolo Baldan and Daniele Gorla, editors, *25th International Conference on Concurrency Theory, CONCUR 2014*, volume 8704 of *Lecture Notes in Computer Science*, pages 249–265. Springer, 2014. doi:10.1007/978-3-662-44584-6_18.

[32] Michael Huth and Marta Z. Kwiatkowska. Quantitative analysis and model checking. In *12th Annual IEEE Symposium on Logic in Computer Science*, pages 111–122. IEEE Computer Society, 1997. doi:10.1109/LICS.1997.614940.

[33] Bengt Jonsson and Wang Larsen, Kim G.and Yi. Probabilistic extensions of process algebras. In Jan A. Bergstra, Alban Ponse, and Scott A. Smolka, editors, *Handbook of Process Algebra*, pages 685–710. Elsevier, 2001. doi:10.1016/b978-044482830-9/50029-1.

[34] Leonid Kantorovich and Gennady S. Rubinstein. On a space of totally additive functions. *Vestnik Leningrad Universitet*, 13:52–59, 1958.

[35] Dexter Kozen. Results on the propositional mu-calculus. *Theoretical Computer Science*, 27:333–354, 1983. doi:10.1016/0304-3975(82)90125-6.

[36] Marta Z. Kwiatkowska and Gethin Norman. Probabilistic metric semantics for a simple language with recursion. In Wojciech Penczek and Andrzej Szalas, editors, *21st International Symposium on Mathematical Foundations of Computer Science 1996, MFCS'96*, volume 1113 of *Lecture Notes in Computer Science*, pages 419–430. Springer, 1996. doi:10.1007/3-540-61550-4_167.

[37] Kim Guldstrand Larsen and Arne Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94(1):1–28, 1991. doi:10.1016/0890-5401(91)90030-6.

[38] Zohar Manna and Amir Pnueli. *The Temporal Logic of Reactive and Concurrent Systems - Specification*. Springer, 1992. doi:10.1007/978-1-4612-0931-7.

[39] Gethin Norman. *Metric Semantics for Reactive Probabilistic Systems*. PhD thesis, University of Birmingham, 1997.

[40] Anna Philippou, Insup Lee, and Oleg Sokolsky. Weak bisimulation for probabilistic systems. In Catuscia Palamidessi, editor, *11th International Conference on Concurrency Theory, CONCUR 2000*, volume 1877 of *Lecture Notes in Computer Science*, pages 334–349. Springer, 2000. doi:10.1007/3-540-44618-4_25.

[41] Vishwanath Raman. *Game Relations, Metrics and Refinements*. PhD thesis, University of California, 2010.

[42] Roberto Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, MIT, 1995.

[43] Roberto Segala and Nancy A. Lynch. Probabilistic simulations for probabilistic processes. In Bengt Jonsson and Joachim Parrow, editors, *5th International Conference on Concurrency Theory, CONCUR '94*, volume 836 of *Lecture Notes in Computer Science*, pages 481–496. Springer, 1994. doi:10.1007/978-3-540-48654-1_35.

[44] Lin Song, Yuxin Deng, and Xiaojuan Cai. Towards automatic measurement of probabilistic processes. In *7th International Conference on Quality Software, QSIC 2007*, pages 50–59. IEEE Computer Society, 2007. doi:10.1109/QSIC.2007.65.

[45] Qiyi Tang and Franck van Breugel. Deciding probabilistic bisimilarity distance one for probabilistic automata. *Journal of Computer and System Sciences*, 111:57–84, 2020. doi:10.1016/j.jcss.2020.02.003.

[46] Hermann Thorisson. *Coupling, Stationarity, and Regeneration*. Probability and Its Applications. Springer, 2000.

[47] Franck van Breugel, Babita Sharma, and James Worrell. Approximating a behavioural pseudometric without discount for probabilistic systems. *Logical Methods in Computer Science*, 4(2), 2008. doi:10.2168/LMCS-4(2:2)2008.

[48] Franck van Breugel and James Worrell. An algorithm for quantitative verification of probabilistic transition systems. In Kim Guldstrand Larsen and Mogens Nielsen, editors, *12th International Conference on Concurrency Theory, CONCUR 2001*, volume 2154 of *Lecture Notes in Computer Science*, pages 336–350. Springer, 2001. doi:10.1007/3-540-44685-0_23.

[49] Franck van Breugel and James Worrell. Towards quantitative verification of probabilistic transition systems. In Fernando Orejas, Paul G. Spirakis, and Jan van Leeuwen, editors, *28th International Colloquium on Automata, Languages and Programming*, volume 2076 of *Lecture Notes in Computer Science*, pages 421–432. Springer, 2001. doi:10.1007/3-540-48224-5_35.

[50] Franck van Breugel and James Worrell. A behavioural pseudometric for probabilistic transition systems. *Theoretical Computer Science*, 331(1):115–142, 2005. doi:10.1016/j.tcs.2004.09.035.

[51] Mingsheng Ying. Bisimulation indexes and their applications. *Theoretical Computer Science*, 275(1-2):1–68, 2002. doi:10.1016/S0304-3975(01)00124-4.