Behavioural Pseudometrics for Nondeterministic Probabilistic Systems

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

[1] Shanghai Normal University
[2] Shanghai Key Laboratory of Trustworthy Computing, East China Normal University
[3] VU University Amsterdam

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 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.

1 Introduction

Bisimulation is an important proof technique for establishing behavioural equivalences of concurrent systems. In probabilistic concurrency theory, 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 [34] to represent a branching time semantics; the latter, as defined in [13, 21, 28], 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 [16]. 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 [17]. 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

⋆ Corresponding author. Partially supported by the National Natural Science Foundation of China (61672229, 61261130589), Shanghai Municipal Natural Science Foundation (16ZR1409100), and ANR 12IS02001 PACE.

is allowed. We provide a modal characterisation of the state-based bisimulation metric closely in line with the classical Hennessy-Milner logic (HML) [27]. 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 bisimilarities: 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 [10]. 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 [21] 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 appeared in the literature. In Section 7 we review some related work. Finally, we conclude in Section 8.

2 Preliminaries

Let S be a countable set. A *(discrete) probability subdistribution* over S is defined as 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 the set $\{s \in S \mid \Delta(s) > 0\}$. 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* $\omega \in \mathcal{D}(S \times S)$ for $(\Delta, \Theta) \in \mathcal{D}(S) \times \mathcal{D}(S)$ is given if $\sum_{t\in S} \omega(s,t) = \Delta(s)$ 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 $\Omega(\Delta, \Theta)$.

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 paper 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 $s,t \in \mathtt{S}$.

[4] Notice that we do not claim that negation and testing operators, plus some constant functions, suffice to represent 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.

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* [31] $K(d)\colon\ \mathcal{D}(\mathtt{S}) \times \mathcal{D}(\mathtt{S}) \to [0,1]$ 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

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

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 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) generalize labelled transition systems by allowing for probabilistic choices in the transitions. They are essentially *simple probabilistic automata* [39] without initial states.

Definition 1. *A* probabilistic labelled transition system *is a triple $(S, A, \to)$, where S is a countable set of states, A is a countable set of actions, and the relation $\to\ \subseteq S \times A \times \mathcal{D}(S)$ is a transition relation.*

We write $s \xrightarrow{a} \Delta$ for $(s,a,\Delta) \in\ \to$ and $s \not\xrightarrow{a}$ if there is no Δ satisfying $s \xrightarrow{a} \Delta$. We 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 $d(s,t) \le d'(s,t)$, 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 2. *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 $K(d)(\Delta,\Delta') \le d(s,t)$.*

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 [34, 39]. Note that $\mathbf{0}$ does not satisfy Definition 2 for general pLTSs, thus is not a state-based bisimulation metric.

Example 3. Let us calculate the distance between states s and t in Figure 1. Firstly, observe that $\mathbf{d}_s(s_2,t_3) = 0$ because s_2 is bisimilar to t_3 while $\mathbf{d}_s(s_3,t_3) = 1$ because the two states s_3

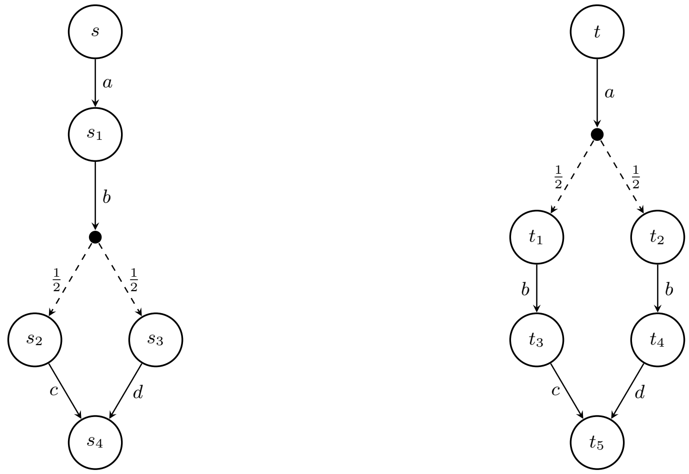

Fig. 1: $\mathbf{d}_s(s,t) = \frac{1}{2}$

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

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

Here the only legitimate 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_{s} is monotone and its least fixed point is given by $\bigsqcup d_i$, where $d_0 = \mathbf{0}$ and $d_{i+1} = F_{\mathrm{s}}(d_i)$ for all $i\in\mathbb{N}$.

Proposition 4. $\mathbf{d}_s$ *is the least fixed point of* F_{s}. □

Essentially the same property as Proposition 4 has appeared in [17].

Now we proceed by defining a real-valued modal logic based on Hennessy-Milner logic [27], called metric HML, to characterize the bisimilarity metric. It is motivated by [4, 16, 17, 30].

Definition 5. *Our metric HML is two-sorted and has the following syntax:*

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

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* $\mathcal{L}^{\mathrm{S}}$, and ψ range over the set of all *distribution formulae* $\mathcal{L}^{\mathrm{D}}$. 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 $\langle a\rangle\varphi$. All other operators are standard.

Definition 6. *A state formula $\varphi \in \mathcal{L}^{\mathrm{S}}$ evaluates in $s \in S$ as follows:*

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

and a distribution formula $\psi \in \mathcal{L}^{\mathrm{D}}$ evaluates in $\Delta \in \mathcal{D}(S)$ as follows:

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

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[5]. 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. Because 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

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

Example 7. 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

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

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

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

[5] Since we will compare our logic with that in [17], 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. [3, 29].

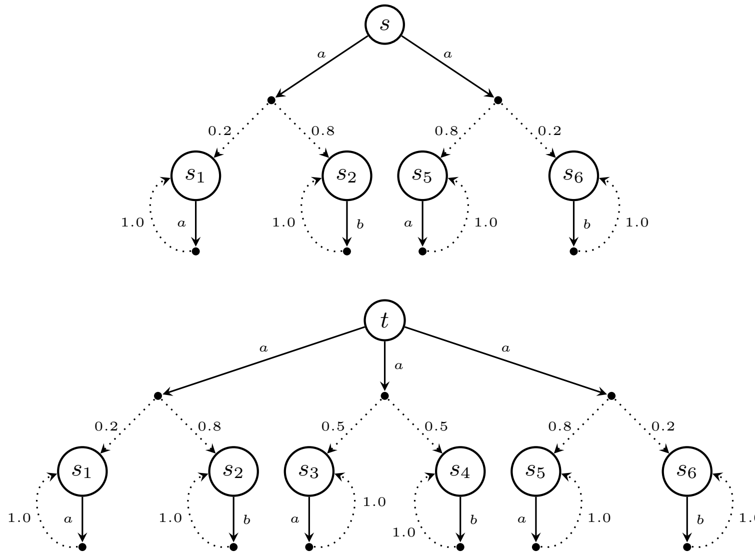

Fig. 2: $\mathbf{d}_s^{\mathrm{ls}}(s,t) = 0.3$

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

$$[\![\varphi]\!](s) = \max([\![\psi]\!](\Delta_1), [\![\psi]\!](\Delta_2)) = 0.2$$
$$[\![\varphi]\!](t) = \max([\![\psi]\!](\Delta_1), [\![\psi]\!](\Delta_2), [\![\psi]\!](\Delta_3)) = 0.5.$$

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 8. 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 7. Then we have

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

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

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

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 9. $\mathbf{d}_s = \mathbf{d}_s^{\mathrm{ls}}$ *and* $K(\mathbf{d}_s) = \mathbf{d}_s^{\mathrm{ld}}$ □

The two properties in Theorem 9 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 $K(\mathbf{d}_s^{\mathrm{ls}}) \sqsubseteq \mathbf{d}_s^{\mathrm{ld}}$ 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 $\mathbf{d}_s^{\mathrm{ls}}$ is a state-based bisimulation metric and so obtain $\mathbf{d}_s \sqsubseteq \mathbf{d}_s^{\mathrm{ls}}$. 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.

Remark 10. For deterministic pLTSs, the proof of Theorem 9 can be greatly simplified. In that case, we can even fold distribution formulae into state formulae and then the state-based bisimilarity metric can be characterised by the following one-sorted metric logic

$$\varphi ::= \top \mid \neg\varphi \mid \varphi \ominus p \mid \varphi_1 \wedge \varphi_2 \mid \langle a \rangle \varphi \ . \tag{4}$$

Therefore, for deterministic pLTSs, the two-sorted logic in Definition 5 degenerates into the logic considered in [16, 23, 45], 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 8.

In [2, 7] 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 2 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 11. *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:*

1. *if* $s \xrightarrow{a} \Delta$ *then* $\overline{s} \xrightarrow{a} \Delta$*;*
2. *if* $s \not\xrightarrow{a}$ *then* $\overline{s} \xrightarrow{a} \varepsilon$*;*
3. *if* $\Delta_i \xrightarrow{a} \Theta_i$ *for all* $i \in I$ *then* $(\sum_{i\in I} p_i \cdot \Delta_i) \xrightarrow{a} (\sum_{i\in I} p_i \cdot \Theta_i)$*, where* I *is a finite index set and* $\sum_{i\in I} p_i \leq 1$*.*

Note that if $\Delta \xrightarrow{a} \Delta'$ then some (not necessarily all) 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 $\frac{1}{2}\overline{s_4}$.

Definition 12. *A* 1*-bounded pseudometric* d *on* $\mathcal{D}_{sub}(S)$ *is a* distribution-based bisimulation metric *if* $|\,|\Delta_1| - |\Delta_2|\,| \leq d(\Delta_1,\Delta_2)$ *and for all* $\Delta_1,\Delta_2 \in \mathcal{D}_{sub}(S)$ *with* $d(\Delta_1,\Delta_2) < 1$*, whenever* $\Delta_1 \xrightarrow{a} \Delta_1'$ *then there exists some transition* $\Delta_2 \xrightarrow{a} \Delta_2'$ *such that* $d(\Delta_1',\Delta_2') \leq d(\Delta_1,\Delta_2)$*.*

The condition $|\,|\Delta_1| - |\Delta_2|\,| \leq d(\Delta_1,\Delta_2)$ is introduced to ensure 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 [13] 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

$$F_{\mathrm{d}}(d)(\Delta,\Theta) = \max(\sup_{a\in A}\{H(d)(der(\Delta,a), der(\Theta,a))\},\ |\,|\Delta| - |\Theta|\,|)\,. \tag{5}$$

It can be shown that F_{d} is monotone and its least fixed point is given by $\bigsqcup d_i$, where we set $d_0(\Delta,\Theta) = |\,|\Delta| - |\Theta|\,|$ 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 4.

Proposition 13. $\mathbf{d}_d$ *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 14. 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 easy 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 $s_1 \xrightarrow{b} \frac{1}{2}\overline{s_2}+\frac{1}{2}\overline{s_3}$ 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 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:

$$\psi ::= \top \mid \neg\psi \mid \psi \ominus p \mid \psi_1 \wedge \psi_2 \mid \langle a \rangle \psi \ . \tag{6}$$

This logic is the same as that defined in (4) except that now we only have distribution formulae. The semantic interpretation of formulae comes with no surprise.

Definition 15. *A formula $\psi \in \mathcal{L}^{D*}$ evaluates in $\Delta \in \mathcal{D}_{sub}(S)$ as follows:*

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

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 $\mathbf{d}_d$; the proof is similar to but easier than that of Theorem 9.

Theorem 16. $\mathbf{d}_d = \mathbf{d}_d^{\mathrm{ld}}$ □

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[6]. 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.

Definition 17. *A pLTS is* fully enabled *if $\forall s.\ EA(s) = Act$. 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:*

- $S_{\bot} = S \cup \{\bot\}$
- $\rightarrow_{\bot} = \rightarrow \cup \{(s, a, \overline{\bot}) \mid s \not\xrightarrow{a}$ *and* $a \in Act\} \cup \{(\bot, a, \overline{\bot}) \mid a \in Act \cup \{\bot\}\}$.

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 $\mathcal{A}^{\bot}$ such that $\Delta^{\bot}(s^{\bot}) = \Delta(s)$ and $\Delta^{\bot}(\bot) = 1 - |\Delta|$.

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. Then we arrive at the following theorem.

Theorem 18. *Let Δ, Θ be two distributions on a pLTS. Then* $\mathbf{d}_d(\Delta, \Theta) \leq K(\mathbf{d}_s)(\Delta, \Theta)$. □

[6] 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.

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 [10] for more details. The following one is taken from [15].

Definition 19. *Let S and T be two sets and $\mathcal{R} \subseteq S \times T$ be a binary relation. The lifted relation $\mathcal{R}^\dagger \subseteq \mathcal{D}_{sub}(S) \times \mathcal{D}_{sub}(T)$ is the smallest relation that satisfies:*

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 [34], which is originally defined for deterministic systems.

Definition 20. *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 [13] as a sound and complete coinductive proof technique for linear contextual equivalence, a natural extensional behavioural equivalence for functional programs.

Definition 21. *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 $\Delta' \sim_d \Theta'$.*

It is obvious that $s \sim_s t$ iff $\mathbf{d}_s(s,t) = 0$ iff $[\![\varphi]\!](s) = [\![\varphi]\!](t)$ for any states s, t and formula $\varphi \in \mathcal{L}^{\mathrm{S}}$. Similarly, $\Delta \sim_d \Theta$ iff $\mathbf{d}_d(\Delta,\Theta) = 0$ iff $[\![\psi]\!](\Delta) = [\![\psi]\!](\Theta)$ for any subdistributions Δ, Θ and 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 [14, 18, 20, 21, 26, 28], 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 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 [18, 20, 21] both (i) and (ii) are imposed, while in [28] and also in our definition of $\sim_d$ (i) is not used. The condition (ii) is built in Definition 11 but partially used in [28], 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* [40]. 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 22. An equivalent way of defining combined transitions is to use Definition 11. 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 20 to get a coarser notion of state-based bisimilarity called strong probabilistic bisimulation in [40], written $\sim'_s$, and then lifting it to subdistributions to finally obtain $(\sim'_s)^\dagger$. This is essentially the relation investigated in [26]. 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 [21] 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 17.

Definition 23. *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 [14, 26] for subdistributions.

Definition 24. *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 with $\Theta = \sum_{i\in I} p_i \cdot \Theta_i$, such that $\Delta_i \sim_2 \Theta_i$ for each $i \in I$.*

Thirdly, we adapt the bisimulation in [18] 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 25. *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 [28]. 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 26. *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 $\Delta' \sim_4 \Theta'$.*

Theorem 27. *Figure 3 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 3 remains unchanged.

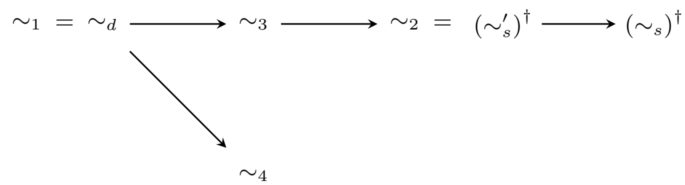

Fig. 3: Relationship between the seven bisimilarities for distributions. An arrow pointing from one relation to another means that the former relation is strictly coarser than the latter. Two relations are incomparable if there is no path from one to the other.

7 Other Related Work

Metrics for probabilistic transition systems are first suggested by Giacalone *et al.* [25] to formalize a notion of distance between processes. They are used also in [33,36] to give denotational semantics for deterministic models. De Vink and Rutten [8] 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 [9]. In [46] Ying proposes a notion of bisimulation index for the usual labelled transition systems, by using ultrametrics on actions instead of using pseudometrics on states. A quantitative linear-time-branching-time spectrum for non-probabilistic systems is given in [19].

Metrics for deterministic systems are extensively studied. Desharnais *et al.* [16] 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 [44] 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 [23]. The metric of [16, 23, 45] 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 [43] 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 [42]. In [22] a sampling algorithm for calculating bisimulation distances in Markov decision processes is shown to have good performance. In [6, 7] the probabilistic bisimulation metric on game structures is characterised by a quantitative μ-calculus. Algorithms for game metrics are proposed in [2, 38]. A notion of bisimulation distance for distributions is proposed in [21]. 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 [17], where Desharnais *et al.* deal with labelled concurrent Markov chains (similar to pLTSs, this model can be captured by the simple probabilistic automata of [39]). They show that the greatest fixed point of a monotonic function on pseudometrics corresponds to the weak probabilistic bisimilarity of [37]. In [24] 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 [41] a notion of trace metric is proposed for pLTSs and a tool is developed to compute the trace metric. In [1] the boolean-valued logic from [12] 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 [11] a notion of bisimulation metric is proposed that extends the approach of [16, 17] to a more general framework called action-labelled quantitative transition systems. In [5] 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 [35] and the μ-calculus [32].

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. A similar idea is used to characterise the distribution-based bisimulation metric. 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 [10]. This entails that the logical characterisation of weak bisimulation metrics would be similar to those presented here.

Acknowledgement We thank the anonymous referees for their helpful comments.

References

1. V. Castiglioni, D. Gebler, and S. Tini. Logical characterization of bisimulation metrics. In *Proceedings of QAPL 2016*. EPTCS, 2016.
2. K. Chatterjee, L. De Alfaro, R. Majumdar, and V. Raman. Algorithms for game metrics. *Logical Methods in Computer Science*, 6(3), 2010.
3. R. Cleaveland, S. P. Iyer, and M. Narasimha. Probabilistic temporal logics via the modal mu-calculus. *Theor. Comput. Sci.*, 342(2-3):316–350, 2005.
4. P. R. D'Argenio, P. Sánchez Terraf, and N. Wolovick. Bisimulations for non-deterministic labelled Markov processes. *Mathematical Structures in Computer Science*, 22(1):43–68, 2012.
5. L. De Alfaro, M. Faella, and M. Stoelinga. Linear and branching system metrics. *IEEE Transactions on Software Engineering*, 35(2):258–273, 2009.
6. L. De Alfaro, R. Majumdar, V. Raman, and M. Stoelinga. Game relations and metrics. In *Proc. LICS'07*, pages 99–108. IEEE, 2007.
7. L. De Alfaro, R. Majumdar, V. Raman, and M. Stoelinga. Game refinement relations and metrics. *Logical Methods in Computer Science*, 4(3), 2008.
8. E. P. de Vink and J. J. M. M. Rutten. Bisimulation for probabilistic transition systems: A coalgebraic approach. *Theor. Comput. Sci.*, 221(1/2):271–293, 1999.
9. J. I. den Hartog. *Probabilistic Extensions of Semantical Models.* PhD thesis, Free University Amsterdam, 2002.

10. Y. Deng. *Semantics of Probabilistic Processes: An Operational Approach.* Springer, 2015.
11. Y. Deng, T. Chothia, C. Palamidessi, and J. Pang. Metrics for action-labelled quantitative transition systems. *ENTCS*, 153(2):79–96, 2006.
12. Y. Deng and W. Du. Logical, metric, and algorithmic characterisations of probabilistic bisimulation. Technical Report CMU-CS-11-110, Carnegie Mellon University, March 2011.
13. Y. Deng, Y. Feng, and U. Dal Lago. On coinduction and quantum lambda calculi. In *Proc. CONCUR'15*, pages 427–440. LIPIcs, 2015.
14. Y. Deng and M. Hennessy. On the semantics of Markov automata. *Information and Computation*, 222:139–168, 2013.
15. Y. Deng, R. van Glabbeek, M. Hennessy, and C. Morgan. Testing finitary probabilistic processes (extended abstract). In *Proc. CONCUR'09*, volume 5710 of *Lecture Notes in Computer Science*, pages 274–288. Springer, 2009.
16. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labelled Markov processes. *Theor. Comput. Sci.*, 318(3):323–354, 2004.
17. J. Desharnais, R. Jagadeesan, V. Gupta, and P. Panangaden. The metric analogue of weak bisimulation for probabilistic processes. In *Proc. LICS'02*, pages 413–422. IEEE, 2002.
18. C. Eisentraut, J. C. Godskesen, H. Hermanns, L. Song, and L. Zhang. Probabilistic bisimulation for realistic schedulers. In *Proc. FM'15*, volume 9109 of *Lecture Notes in Computer Science*, pages 248–264. Springer, 2015.
19. U. Fahrenberg and A. Legay. The quantitative linear-timebranching-time spectrum. *Theoretical Computer Science*, 538:54–69, 2014.
20. Y. Feng and M. Ying. Toward automatic verification of quantum cryptographic protocols. In *Proc. CONCUR'15*, volume 42 of *LIPIcs*, pages 441–455, 2015.
21. Y. Feng and L. Zhang. When equivalence and bisimulation join forces in probabilistic automata. In *Proc. FM'14*, volume 8442 of *LNCS*, pages 247–262. Springer, 2014.
22. N. Ferns, P. Panangaden, and D. Precup. Bisimulation metrics for continuous Markov decision processes. *SIAM J. Comput.*, 40(6):1662–1714, 2011.
23. N. Ferns, D. Precup, and S. Knight. Bisimulation for Markov decision processes through families of functional expressions. In *Panangaden Festschrift*, volume 8464 of *LNCS*, pages 319–342. Springer, 2014.
24. D. Gebler, K. G. Larsen, and S. Tini. Compositional metric reasoning with probabilistic process calculi. In *Proc. FOSSACS'15*, volume 9034 of *LNCS*, pages 230–245. Springer, 2015.
25. A. Giacalone, C. Jou, and S. Smolka. Algebraic reasoning for probabilistic concurrent systems. In *Proc. IFIP TC2 Working Conf. on Prog. Concepts and Methods*, pages 443–458, 1990.
26. M. Hennessy. Exploring probabilistic bisimulations, part I. *Formal Aspects of Computing*, 24(4-6):749–768, 2012.
27. M. Hennessy and R. Milner. Algebraic laws for nondeterminism and concurrency. *J. ACM*, 32:137–161, 1985.
28. H. Hermanns, J. Krcál, and J. Kretínský. Probabilistic bisimulation: Naturally on distributions. In *Proc. CONCUR'14*, volume 8704 of *LNCS*, pages 249–265. Springer, 2014.
29. M. Huth and M. Z. Kwiatkowska. Quantitative analysis and model checking. In *Proceedings of the 12th Annual IEEE Symposium on Logic in Computer Science*, pages 111–122. IEEE Computer Society, 1997.
30. B. Jonsson and W. Larsen, Kim G.and Yi. Probabilistic extensions of process algebras. In *Handbook of Process Algebra*, pages 685–710. Elsevier, 2001.
31. L. Kantorovich and G. Rubinshtein. On a space of totally additive functions. *Vestn Len. Univ.*, 13(7):52–59, 1958.
32. D. Kozen. Results on the propositional mu-calculus. *Theoretical Computer Science*, 27:333–354, 1983.
33. M. Z. Kwiatkowska and G. J. Norman. Probabilistic metric semantics for a simple language with recursion. In *Proc. MFCS'96*, volume 1113 of *LNCS*, pages 419–430. Springer, 1996.
34. K. G. Larsen and A. Skou. Bisimulation through probabilistic testing. *Inf. Comput.*, 94:1–28, 1991.
35. Z. Manna and A. Pnueli. *The Temporal Logic of Reactive and Concurrent Systems: Specification.* Springer-Verlang, 1991.
36. G. J. Norman. *Metric Semantics for Reactive Probabilistic Systems.* PhD thesis, University of Birmingham, 1997.

37. A. Philippou, I. Lee, and O. Sokolsky. Weak bisimulation for probabilistic systems. In *Proc. CONCUR'00*, volume 1877 of *LNCS*, pages 334–349. Springer, 2000.
38. V. Raman. *Game Relations, Metrics and Refinements*. PhD thesis, University of California, 2010.
39. R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, MIT, 1995.
40. R. Segala and N. Lynch. Probabilistic simulations for probabilistic processes. *Proc. CONCUR'94*, pages 481–496, 1994.
41. L. Song, Y. Deng, and X. Cai. Towards automatic measurement of probabilistic processes. In *Proc. QSIC'07*, pages 50–59. IEEE, 2007.
42. F. van Breugel, B. Sharma, and J. Worrell. Approximating a behavioural pseudometric without discount for probabilistic systems. *Logical Methods in Computer Science*, 4(2), 2008.
43. F. van Breugel and J. Worrell. An algorithm for quantitative verification of probabilistic transition systems. In *Proc. CONCUR'01*, volume 2154 of *LNCS*, pages 336–350. Springer, 2001.
44. F. van Breugel and J. Worrell. Towards quantitative verification of probabilistic transition systems. In *Proc. ICALP'01*, volume 2076 of *LNCS*, pages 421–432. Springer, 2001.
45. F. van Breugel and J. Worrell. A behavioural pseudometric for probabilistic transition systems. *Theor. Comput. Sci.*, 331(1):115–142, 2005.
46. M. Ying. Bisimulation indexes and their applications. *Theoretical Computer Science*, 275(1/2):1–68, 2002.