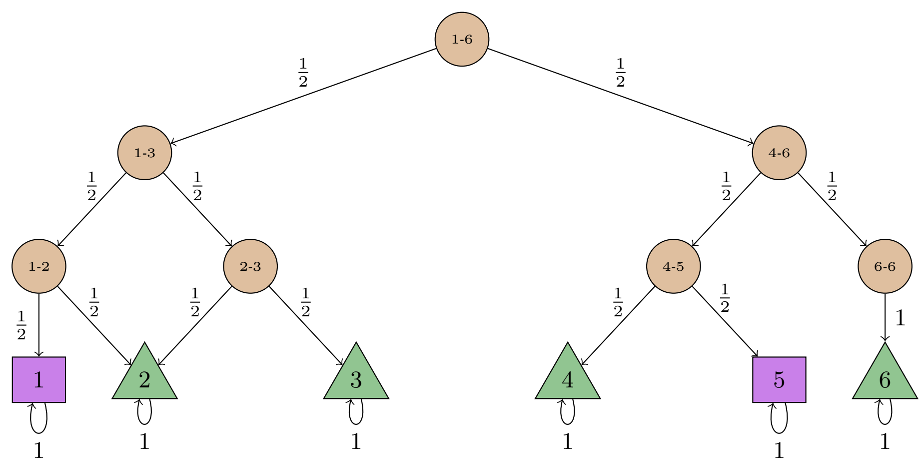
Probabilistic Bisimilarity Distances

Franck van Breugel,
DisCoVeri Group,
Department of Electrical Engineering and Computer Science,
York University, Toronto, Canada

In a recent column, Panangaden discusses probabilistic bisimilarity. Here, we show how probabilistic bisimilarity can be quantitatively generalized leading to probabilistic bisimilarity distances.

1. INTRODUCTION

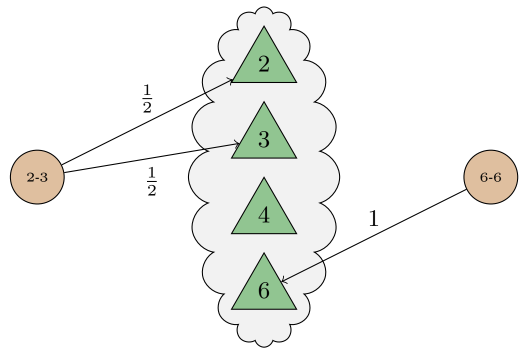
Many different models have been proposed for systems in which probabilities play a role. Here, we consider one of the simplest such models, namely *labelled Markov chains*. These are ordinary Markov chains in which each state has a label. The labels are used to abstractly capture known facts about the states. For example, if the model represents code, then a label may capture whether a particular variable in the code has value zero in a state. In this paper, we consider the labelled Markov chain depicted below as our running example. It uses a fair coin, is small and allows us to illustrate all key notions. In the example we use coloured shapes to represent the labels.



(1)

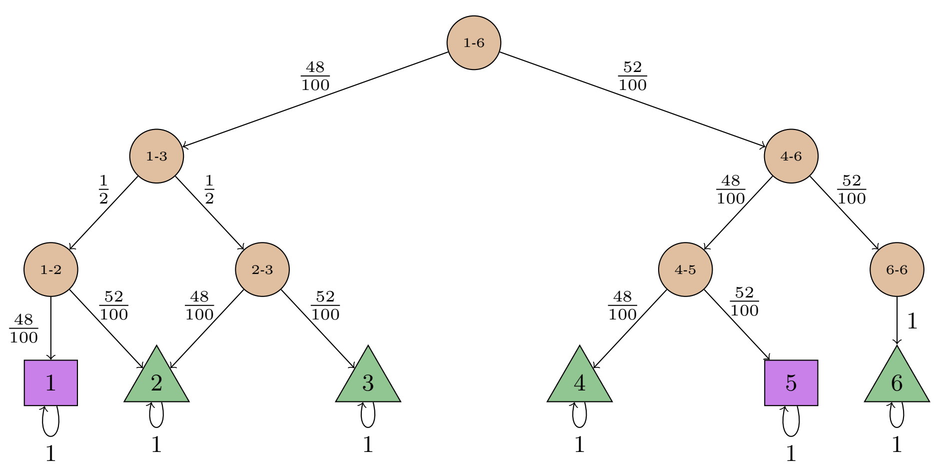
A *behavioural equivalence* is an equivalence relation on the states of a model that captures which states behave the same. Numerous behavioural equivalences have been proposed in the literature. Here, we focus on the behavioural equivalence, known as *probabilistic bisimilarity*, introduced by Larsen and Skou in [Larsen and Skou 1989]. Panangaden discusses this behavioural equivalence in detail in a recent column [Panangaden 2015]. Note that Panangaden considers a slightly different model. Whereas we focus on a model in which the states are labelled, Panangaden considers a model where the transitions are labelled. However, each model can be expressed in terms of the other (see, for example, [Reniers et al. 2014]).

In general, for two states of a model to be considered behaviourally equivalent, the same facts should hold for both states. Furthermore, their transitions should preserve the equivalence. As we will sketch in Section 4, the states 2, 3, 4 and 6 of the labelled Markov chain (1) are probabilistic bisimilar. For all four states the same facts hold, that is, all four are green triangles. Furthermore, with probability one each transitions to itself, that is, to a state that behaves the same. The states 2-3 and 6-6 are both brown circles. Although state 6-6 has only one transition whereas state 2-3 has two transitions, the probabilities of the latter two can be added as they reach states that are behaviourally indistinguishable. Hence, both state 6-6 and state 2-3 transition to the equivalence class containing states 2, 3, 4 and 6 with probability one.



Therefore, as we will sketch in Section 4, the states 6-6 and 2-3 are probabilistic bisimilar as well.

Instead of using a fair coin, let us employ a biased one in all states but state 1-3. Assume we flip our biased coin one million times and we see $485,421$ times heads.[1] Based on this experiment, we model the biased system by the following labelled Markov chain.

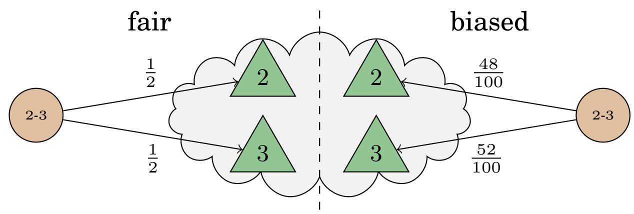


(2)

The fair and biased systems (1) and (2) together can be viewed as one larger system. The states 2 and 3 in the fair component are probabilistic bisimilar to the states 2 and 3 in the biased component, since all four states are green triangles and each transitions with probability one to itself. In both components, state 2-3 is a brown circle. Note

[1] From this experiment we can conclude that the odds of heads lies in the interval $[0.484, 0487]$ with probability 0.99.

that their individual transition probabilities do not match exactly ($\frac{1}{2}$ versus $\frac{48}{100}$ and $\frac{1}{2}$ versus $\frac{52}{100}$), but we can add these probabilities since the states 2 and 3 are considered behaviourally equivalent. Hence, in both components state 2-3 transitions to the equivalence class containing the states 2 and 3 with probability one.



Therefore, state 2-3 in the fair component is probabilistic bisimilar to state 2-3 in the biased component. Let us finally compare the behaviour of state 1-2 in the two components. Both states are brown circles. Again, their individual transition probabilities do not match exactly. However, in this case we cannot add them since the transitions reach states with different coloured shapes, which are therefore not behaviourally equivalent. Hence, state 1-2 in the fair component is not probabilistic bisimilar to state 1-2 in the biased component.

The last example shows that the notion of probabilistic bisimilarity is sensitive to the exact values of the transition probabilities. Note though that these probabilities were obtained experimentally. Assume we flip the coin 100 million times and obtain $48,542,167$ times heads.[2] As a result, we can replace the probabilities $\frac{48}{100}$ and $\frac{52}{100}$ with $\frac{485}{1000}$ and $\frac{515}{1000}$, respectively. In the resulting system, state 1-2 is neither probabilistic bisimilar to state 1-2 in the original system (1) nor to state 1-2 in the biased system (2).

Hence, we can conclude that the notion of probabilistic bisimilarity is *not robust* as minuscule changes in the probabilities of the transitions may impact which states are probabilistic bisimilar. This lack of robustness was first observed by Giacalone, Jou and Smolka in [Giacalone et al. 1990].

In their paper [Giacalone et al. 1990], Giacalone et al. also propose to generalize the qualitative notion of behavioural equivalence — in which two states are either behavioural equivalent or they are inequivalent — to a quantitative notion. An equivalence relation can be viewed as a function that maps each state pair to a boolean. This boolean captures whether the two states are equivalent. Giacalone et al. suggest to use the real numbers in the unit interval $[0, 1]$ instead of the booleans. Hence, we arrive at a function mapping pairs of states to real numbers. As we will see in Section 2, such a function assigns a *distance* to each pair of states. This distance should be used to capture the behavioural similarity of two states. The smaller their distance, the more alike the states should behave. In particular, distance zero should exactly capture that states are behaviourally equivalent.

Not only do these distances provide a robust alternative to behavioural equivalences. These distances can also be used to reduce the size of the model of a system. For example, in [Murthy et al. 2012], Murthy et al. use a quantitative generalization of a behavioural equivalence to reduce the size of a model of a biological system. Similarly, Sen, Deshpande and Getoor [Sen et al. 2009] exploit a quantitative generalization of a behavioural equivalence for high-level reasoning in the setting of probabilistic relational databases. Bacci, Bacci, Larsen and Mardare [Bacci et al. 2017] show how these distances can be used to reduce the state space of a labelled Markov chain.

[2] In this case we can deduce that with probability 0.99 the odds of heads lies in the interval $[0.4853, 0.4857]$.

As Panangaden discusses in [Panangaden 2015], probabilistic bisimilarity can be characterized by a *logic*. Adjusted to our slightly different model, this logic contains a formula that expresses whether a state of the model has a particular label. For the labelled Markov chain (1), one can express in the logic whether a state is a purple square. This formula holds in state 1 but not in state 2. Furthermore, one can also express that from a given state the probability of transitioning to a state satisfying a particular formula of the logic is greater than some threshold. For example, transitioning to a green triangle with probability greater than $\frac{1}{2}$ holds in state 6-6 but not in state 1-2. Finally, formulae can be combined by means of conjunction. For example, one can express that a state is a brown circle and that the state can transition to a green triangle with probability greater than $\frac{1}{2}$. This formula holds in state 2-3 but not in state 1-2. This logic characterizes probabilistic bisimilarity in the following way. States are probabilistic bisimilar if and only if those states satisfy the same formulae of the logic. As a consequence, if two states are not bisimilar then there exists a formula that distinguishes them, that is, it holds in one state but not in the other.

In [Desharnais et al. 1999], Desharnais, Gupta, Jagadeesan and Panangaden present a *quantitative generalization* of probabilistic bisimilarity for labelled Markov chains. Their generalization is based on a small variation on the logic described above. In this modified logic one can express that a given state transitions to states that satisfy a particular formula of the logic (without a threshold). For example, one can express that a state transitions to a green triangle. The usual interpretation of a logic can be viewed as a function that maps a pair consisting of a formula of the logic and a state of the model to a boolean. The pair is mapped to true if the formula holds in the state and it is mapped to false otherwise. Desharnais et al. provide a real valued interpretation of the logic, that is, they map formula-state pairs to real numbers in the unit interval $[0, 1]$. Given a formula and a state, the corresponding real number measures the validity of the formula in the state. Roughly, the larger this number, the higher the probability that the formula holds in the state. For example, the formula that expresses that a state is a green triangle has value zero in state 1 and value one in state 2. The formula that captures transitioning to a green triangle has value one in state 6-6 and value $\frac{1}{2}$ in state 1-2.

Desharnais et al. assign a distance to each pair of states in terms of the real valued interpretation of the logic. In particular, the distance of two states is defined by means of a formula that distinguishes the states the most. That is, this distance is defined as the absolute difference of the values of such a formula in the two states. For example, the formula that expresses that a state is a green triangle distinguishes the states 1 and 2 the most. Since the values of this formula in those states are zero and one, respectively, we can conclude that the distance of the states 1 and 2 is $|0 - 1| = 1$. The formula that captures transitioning to a green triangle distinguishes the states 1-2 and 6-6 the most. The values of this formula in the states 1-2 and 6-6 are $\frac{1}{2}$ and one, respectively. Hence, the distance of the states 1-2 and 6-6 is $|\frac{1}{2} - 1| = \frac{1}{2}$.

In the biased system (2), the probabilities are slightly different. Let us compare the behaviour of the states in the fair component with their biased counterparts. The states 1, 2, 3, 4, 5, 6, 2-3 and 6-6 in the fair component are probabilistic bisimilar to their biased counterpart and, hence, have distance zero. The fair and biased versions of state 1-2 are distinguished most by the formula that captures transitioning to a green triangle. The values of this formula in the fair and biased versions of state 1-2 are $\frac{1}{2}$ and $\frac{52}{100}$, respectively. Therefore, their distance is $|\frac{1}{2} - \frac{52}{100}| = \frac{2}{100}$. Note that small differences in probabilities result in small distances. Hence, these distances give rise to a robust generalization of probabilistic bisimilarity.

In this paper, we define a quantitative generalization of probabilistic bisimilarity for labelled Markov chains in a different way. Although the definition is different from the one described above, it gives rise to the same distances as shown in joint work with Hermida, Makkai and Worrell [van Breugel et al. 2005, Theorem 6 and 7]. Our definition is based on ideas first presented in joint work with Worrell [van Breugel and Worrell 2001] and work by Desharnais, Gupta, Jagadeesan and Panangaden [Desharnais et al. 2002] and Deng, Chothia, Palamidessi and Pang [Deng et al. 2005]. We will provide the details in Section 5. For some historical background, we refer the reader to joint work with Worrell [van Breugel and Worrell 2014, Section 1].

2. FROM EQUIVALENCES TO DISTANCES

In [Giacalone et al. 1990], Giacalone et al. propose a quantitative generalization of the notion of behavioural equivalence. Before presenting their generalization, let us recall the definition of an equivalence relation.

Definition 2.1. A relation $\mathcal{R} \subseteq S \times S$ is an *equivalence relation* if

— for all $s \in S$, $(s, s) \in \mathcal{R}$,
— for all s, $t \in S$, if $(s, t) \in \mathcal{R}$ then $(t, s) \in \mathcal{R}$, and
— for all s, t, $u \in S$, if $(s, t) \in \mathcal{R}$ and $(t, u) \in \mathcal{R}$ then $(s, u) \in \mathcal{R}$.

An equivalence relation can be viewed as a function mapping a pair to zero if the elements are equivalent and to one if the elements are not equivalent. That is, an equivalence relation $\mathcal{R}$ can be viewed as a function $r : S \times S \to \{0, 1\}$ defined by

$$r(s,t) = \begin{cases} 0 & \text{if } (s,t) \in \mathcal{R} \\ 1 & \text{otherwise} \end{cases}$$

The above three conditions can be captured as

— for all $s \in S$, $r(s, s) = 0$,
— for all s, $t \in S$, $r(s, t) = r(t, s)$, and
— for all s, t, $u \in S$, if $r(s, u) \leq r(s, t) + r(t, u)$.

Rather than using the booleans or $\{0, 1\}$, Giacalone et al. propose to employ the real numbers in the unit interval $[0, 1]$. That is, instead of a function $r : S \times S \to \{0, 1\}$ they suggest to use a function $d : S \times S \to [0, 1]$ satisfying the above conditions. Such a function is known as a pseudometric in topology.

Definition 2.2. A function $d : S \times S \to [0, 1]$ is a *pseudometric* if

— for all $s \in S$, $d(s, s) = 0$,
— for all s, $t \in S$, $d(s, t) = d(t, s)$, and
— for all s, t, $u \in S$, if $d(s, u) \leq d(s, t) + d(t, u)$.

A *behavioural pseudometric* generalizes the notion of a behavioural equivalence. It maps each pair of states of a model to their distance. Such a pseudometric should be defined so that the distance $d(s, t)$ measures the similarity of the behaviour of the states s and t. The more alike the states behave, the smaller this distance should be. In particular, distance zero should capture exactly those state pairs that are behaviourally equivalent.

As we mentioned before, behavioural equivalences such as probabilistic bisimilarity are not robust. Small changes to the probabilities may cause behaviourally equivalent states to become inequivalent or vice versa. Behavioural pseudometrics provide a robust generalization as small changes to the probabilities should only cause small changes in the distances.

3. LABELLED MARKOV CHAINS

To formally define the model of interest, labelled Markov chain, we first recall the notions of a probability distribution and its support, and we provide a few examples. Given a finite set S, a function $\mu : S \to [0,1]$ is a *probability distribution* on S if $\sum_{s \in S} \mu(s) = 1$. The set of all probability distributions on S is denoted by $\mathit{Distr}(S)$. For $s \in S$, the *Dirac distribution* concentrated at s is the function $\delta_s : S \to [0,1]$ defined by

$$\delta_s(t) = \begin{cases} 1 & \text{if } s = t \\ 0 & \text{otherwise} \end{cases}$$

Dirac distributions and their convex combinations, such as $\frac{1}{2}\delta_s + \frac{1}{2}\delta_t$, are all examples of probability distributions. The *support* of a probability distribution μ on S is defined by

$$\mathrm{support}(\mu) = \{\, s \in S \mid \mu(s) > 0 \,\}.$$

For example, we have that $\mathrm{support}(\frac{1}{2}\delta_s + \frac{1}{2}\delta_t) = \{s, t\}$. Next, we define the model of interest.

Definition 3.1. A *labelled Markov chain* is a tuple $\langle S, L, \tau, \ell \rangle$ consisting of

— a finite set S of states,
— a finite set L of labels,
— a transition function $\tau : S \to \mathit{Distr}(S)$, and
— a labelling function $\ell : S \to L$.

Consider the labelled Markov chain (1) presented in the introduction. It has 13 states such as 1, 1-2 and 1-3. It has three labels which are represented by brown circles, green triangles and purple squares. Its transition function τ maps state 1 to the Dirac distribution δ_1 and state 4-5 to the convex combination of Dirac distributions $\frac{1}{2}\delta_4 + \frac{1}{2}\delta_5$. Its labelling function ℓ maps state 1-2 to a brown circle, state 2 to a green triangle, and state 5 to a purple square.

4. PROBABILISTIC BISIMILARITY

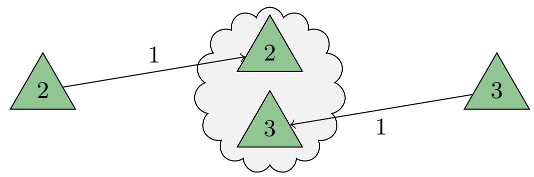
In [Larsen and Skou 1989], Larsen and Skou introduce the notion of probabilitic bisimilarity. This behavioural equivalence captures which states of a labelled Markov chain are considered to behave the same. Equivalent states should have the same label and the transition function should preserve the equivalence. This is captured by the following definition.

Definition 4.1. An equivalence relation $\mathcal{R} \subseteq S \times S$ is a *probabilistic bisimulation* if for all $s, t \in S$, if $(s, t) \in \mathcal{R}$ then $\ell(s) = \ell(t)$ and $(\tau(s), \tau(t)) \in \mathcal{R}{\uparrow}$.

The relation $\mathcal{R}{\uparrow}$ is defined below. A probabilistic bisimulation is an equivalence relation on the states of a labelled Markov chain. The transition function takes a state to a probability distribution on states. Since this transition function should preserve the equivalence relation, we have to lift an equivalence relation on the states to a relation on the probability distributions over those states. Larsen and Skou define this lifting as follows.

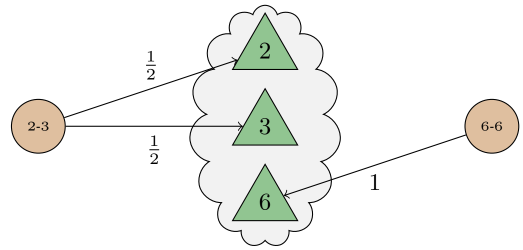
Definition 4.2. The *lifting* of an equivalence relation $\mathcal{R} \subseteq S \times S$ is the relation $\mathcal{R}{\uparrow} \subseteq \mathit{Distr}(S) \times \mathit{Distr}(S)$ defined by $(\mu, \nu) \in \mathcal{R}{\uparrow}$ if $\mu(C) = \nu(C)$ for all $\mathcal{R}$-equivalence classes C.

Note that we write $\mu(C)$ for $\sum_{s \in C} \mu(s)$. Consider the labelled Markov chain (1) of the introduction. To argue that the states 2 and 3 behave the same, let $\mathcal{R}$ be the smallest equivalence relation containing the pair $(2, 3)$. Note that the states 2 and 3 are green triangles and, hence, have the same label. To deduce that the transitions of the states 2 and 3 preserve the equivalence relation $\mathcal{R}$, we need to show that $(\delta_2, \delta_3) \in \mathcal{R}{\uparrow}$. Let C be an $\mathcal{R}$-equivalence class. Since $(2, 3) \in \mathcal{R}$, either C contains both 2 and 3 or it contains neither of them. Hence, either $\delta_2(C) = 1 = \delta_3(C)$ or $\delta_2(C) = 0 = \delta_3(C)$.



Therefore, it follows that $(\delta_2, \delta_3) \in \mathcal{R}{\uparrow}$.

To show that the states 2-3 and 6-6 are behaviourally equivalent, let $\mathcal{R}$ be the smallest equivalence relation containing the pairs (2-3, 6-6), $(2, 3)$ and $(3, 6)$. Note that related states have the same label, since the states 2-3 and 6-6 are both brown circles and the states 2, 3 and 6 are all green triangles. To conclude that the transitions of the states 2-3 and 6-6 preserve the equivalence relation $\mathcal{R}$, we have to prove that $(\frac{1}{2}\delta_2 + \frac{1}{2}\delta_3, \delta_6) \in \mathcal{R}{\uparrow}$. Let C be an $\mathcal{R}$-equivalence class. Then either C contains 2, 3 and 6, or it contains none of them. In both cases, $(\frac{1}{2}\delta_2 + \frac{1}{2}\delta_3)(C) = \delta_6(C)$.



Hence, we have that $(\frac{1}{2}\delta_2 + \frac{1}{2}\delta_3, \delta_6) \in \mathcal{R}{\uparrow}$.

To show that a pair of states behave the same, in general, several different probabilistic bisimulations containing the pair can be constructed. For example, the smallest equivalence relation containing $(2, 3)$ and also the smallest equivalence relation containing $(2, 3)$ and $(3, 6)$ are probabilistic bisimulations, and both show that the states 2 and 3 behave the same. As is often mentioned in the literature, there exists a largest probabilistic bisimulation. However, a proof of this result cannot easily be found in the literature. Therefore, we present one.

PROPOSITION 4.3. *There exists a largest probabilistic bisimulation.*

PROOF. In this proof we assume that the reader is familiar with some concepts of order theory which can, for example, be found in [Davey and Priestley 2002]. We denote the set of equivalence relations on S by $Equiv(S)$. This set equipped with the partial order $\subseteq$ forms a complete lattice [Davey and Priestley 2002, Example 2.34]. The function $\Phi : Equiv(S) \to 2^{S \times S}$ is defined by

$$\Phi(\mathcal{R}) = \{\, (s, t) \in S \times S \mid \ell(s) = \ell(t) \wedge (\tau(s), \tau(t)) \in \mathcal{R}{\uparrow} \,\}.$$

Let $\mathcal{R}$ be an equivalence relation. As a consequence, $\mathcal{R}{\uparrow}$ is also an equivalence relation. Hence, $\Phi(\mathcal{R})$ is an equivalence relation as well. Therefore, Φ is a function on the complete lattice $Equiv(S)$.

Let $\mathcal{R}$ and $\mathcal{S}$ be equivalence relations with $\mathcal{R} \subseteq \mathcal{S}$. Then, the partition into equivalence classes induced by $\mathcal{R}$ is a refinement of the partition induced by $\mathcal{S}$. As a consequence, each $\mathcal{S}$-equivalence class is a disjoint union of $\mathcal{R}$-equivalence classes. Hence, $\mathcal{R}{\uparrow} \subseteq \mathcal{S}{\uparrow}$. Therefore, $\Phi(\mathcal{R}) \subseteq \Phi(\mathcal{S})$, that is, the function Φ is order-preserving.

From the definitions of Φ and probabilistic bisimulation we can conclude that an equivalence relation $\mathcal{R}$ is a probabilistic bisimulation if and only if $\mathcal{R} \subseteq \Phi(\mathcal{R})$, that is, $\mathcal{R}$ is a post-fixed point of Φ (see [Davey and Priestley 2002, Definition 8.14]). Since Φ is an order-preserving function on a complete lattice, we can conclude from the dual of [Davey and Priestley 2002, Theorem 8.20] that Φ has a greatest post-fixed point, which is the largest probabilistic bisimulation. □

Definition 4.4. *Probabilistic bisimilarity* is the largest probabilistic bisimulation.

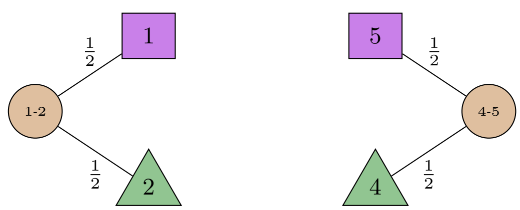
This equivalence relation captures exactly which states of a labelled Markov chain behave the same. In the labelled Markov chain of the introduction, the states 2, 3, 4 and 6 are probabilistic bisimilar. Also the states 1 and 5 are probabilistic bisimilar. Furthermore, the states 1-2, 2-3 and 1-3, are probabilistic bisimilar to the states 4-5, 6-6 and 4-6, respectively.

Towards a quantitative generalization of probabilistic bisimilarity, we first characterize the lifting of an equivalence relation by means of a coupling. This notion from probability theory was introduced by Doeblin in 1936 and first published in [Doeblin 1938].

Definition 4.5. Let $\mu, \nu \in \mathit{Distr}(S)$. The set $\Omega(\mu, \nu)$ of *couplings* of μ and ν is defined by

$$\Omega(\mu, \nu) = \{\, \omega \in \mathit{Distr}(S \times S) \mid \forall s, t \in S : \omega(s, S) = \mu(s) \wedge \omega(S, t) = \nu(t) \,\}.$$

Note that we write $\omega(s, S)$ for $\sum_{t \in S} \omega(s, t)$. Next, we provide a visual representation of these couplings in case the probability distributions are the transitions of a labelled Markov chain. Consider, for example, the transitions of the states 1-2 and 4-5 of the labelled Markov chain (1) presented in the introduction. These transitions can be represented as follows.



A coupling of the probability distributions $\tau(1\text{-}2)$ and $\tau(4\text{-}5)$ has to satisfy the following equations.

$$\begin{aligned} \omega(1,4) + \omega(1,5) &= \tfrac{1}{2} \\ \omega(2,4) + \omega(2,5) &= \tfrac{1}{2} \end{aligned} \tag{3}$$

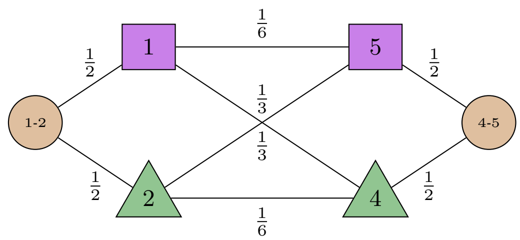
$$\begin{aligned} \omega(1,4) + \omega(2,4) &= \tfrac{1}{2} \\ \omega(1,5) + \omega(2,5) &= \tfrac{1}{2} \end{aligned} \tag{4}$$

Such a coupling can be viewed as a transportation plan of moving one unit from state 1-2 to state 4-5. For example, $\omega(1, 4)$ is the amount transported from state 1 to state 4. Equation (3) tells us that the amount entering state 1, which is $\frac{1}{2}$, also leaves it and is transported to states 4 and 5. Equation (4) captures that the amount leaving state 4, which is $\frac{1}{2}$, also enters it and is transported from states 1 and 2.

For example,

$$\omega(1,4) = \tfrac{1}{3}, \omega(1,5) = \tfrac{1}{6}, \omega(2,4) = \tfrac{1}{6} \text{ and } \omega(2,5) = \tfrac{1}{3} \tag{5}$$

satisfies the above equations and can be depicted as

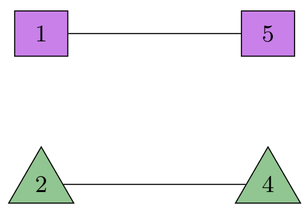


(6)

In [Jonsson and Larsen 1991, Theorem 4.6], Jonsson and Larsen characterize the lifting as presented in Definition 4.2 by means of couplings.

THEOREM 4.6. *Let $\mathcal{R} \subseteq S \times S$ be an equivalence relation and let μ, $\nu \in \mathit{Distr}(S)$. Then $(\mu, \nu) \in \mathcal{R}{\uparrow}$ if and only if there exists $\omega \in \Omega(\mu, \nu)$ such that* $\mathrm{support}(\omega) \subseteq \mathcal{R}$.

Assume that $\mathcal{R}$ is the smallest equivalence relation containing the pairs $(1, 5)$ and $(2, 4)$. This relation can be depicted as follows.

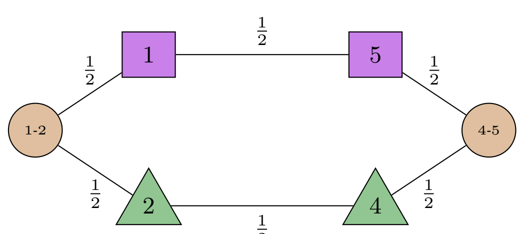


(7)

To conclude that the probability measures $\tau(\text{1-2})$ and $\tau(\text{4-5})$ are in the lifting of $\mathcal{R}$, we have to show that there exists a coupling of $\tau(\text{1-2})$ and $\tau(\text{4-5})$, that is, a transportation plan, that only transports between sources and targets related by $\mathcal{R}$. The coupling defined in (5) and depicted in (6) transports between source 1 and target 4 which are not related by $\mathcal{R}$. Hence, from this coupling we cannot conclude that $\tau(\text{1-2})$ and $\tau(\text{4-5})$ are in the lifting of $\mathcal{R}$. However,

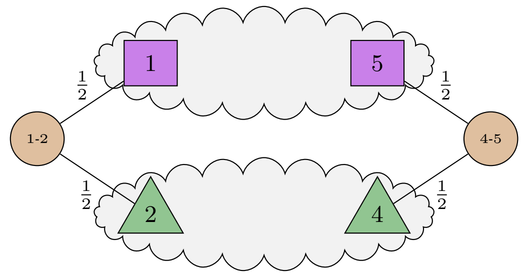
$$\pi(1,5) = \tfrac{1}{2} \text{ and } \pi(2,4) = \tfrac{1}{2}$$

is also a coupling of $\tau(\text{1-2})$ and $\tau(\text{4-5})$. The corresponding transportation plan, which is depicted below, only transports between related sources and targets. As a result, we can conclude that $\tau(\text{1-2})$ and $\tau(\text{4-5})$ are in the lifting of $\mathcal{R}$.



(8)

Note the similarity of the above diagram and the following one, which captures that $\tau(\text{1-2})$ and $\tau(\text{4-5})$ are in the lifting of $\mathcal{R}$ according to Definition 4.2. The transportation plan is replaced with equivalence classes.



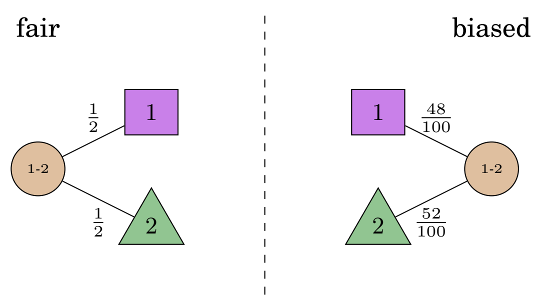
From the above discussion we can conclude that for probabilistic bisimilar states there exists a transportation plan that transports one unit from the one state to the other while only transporting between related states.

5. PROBABILISTIC BISIMILARITY DISTANCES

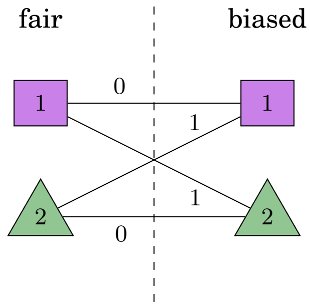
Probabilistic bisimilarity lacks robustness because it requires related states to transition with the *exact* same probability to states that behave *exactly* the same. In our quantitative generalization we not only want to relax these two exactness conditions, we also want to measure how much these conditions need to be relaxed.

We use pseudometrics instead of equivalence relations in our quantitative generalization of probabilistic bisimilarity. As we already discussed in Section 2, a behavioural pseudometric maps each pair of states to a real number in the interval $[0, 1]$. These distances of states capture the similarity of their behaviour.

Consider, for example, the fair and biased version of state 1-2. Both transition to the fair and biased versions of state 1 and 2. The fair and biased versions of the states 1 and 2 behave exactly the same. Hence, in this example only the probabilities of the transitions differ slightly, but the reachable states behave exactly the same.



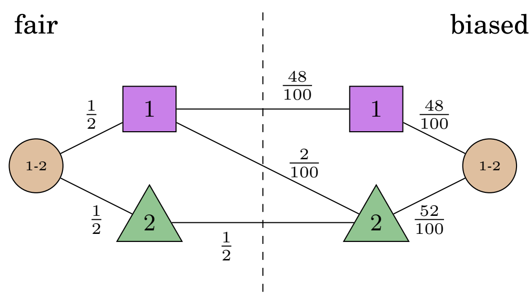
Since the fair and biased versions of the states 1 and 2 are probabilistic bisimilar, their distance should be zero. Let us compare the behaviour of the fair version of state 1 and the biased version of state 2. Since even the basic facts known about these states differ, since their labels are different, their behaviour should be considered very different. This is reflected by the fact that their distance is maximal, that is, it is one. These distances are depicted in the diagram below.



(9)

With probability $\frac{48}{100} + \frac{1}{2}$ the fair and biased version of state 1-2 transition to states that behave exactly the same. With the remaining probability, that is, $\frac{2}{100}$, they transition to states that behave very differently. Hence, the similarity in behaviour of the fair and biased versions of state 1-2 is measured as the weighted sum $\frac{48}{100} \times 0 + \frac{1}{2} \times 0 + \frac{2}{100} \times 1 = \frac{2}{100}$. This captures that their behaviour is very similar, yet they are not probabilistic bisimilar.

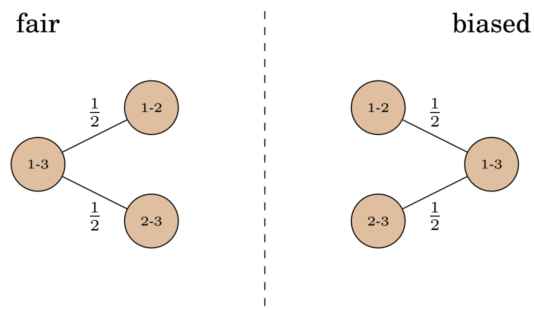
Recall from Section 4 that probabilistic bisimilarity can be captured in terms of transportation plans. As we will discuss next, our distances can also be defined in terms of transportation plans. Consider again the above example. We can transport $\frac{48}{100}$ from the fair state 1 to the biased state 1 and $\frac{1}{2}$ from the fair state 2 to the biased state 2. The remaining $\frac{2}{100}$ from the fair state 1 can only be transported to the biased state 2. This transportation plan can be depicted as follows.



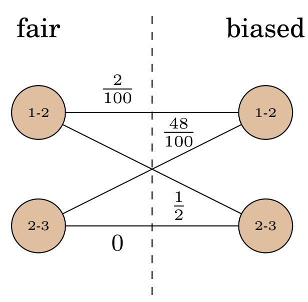
Since the fair states 1 and 2 are probabilistic bisimilar to the biased states 1 and 2, their distances are zero. As a result, transporting between these pairs contributes zero to the distance of the fair state 1-2 and the biased state 1-2. Because the fair state 1 behaves very differently from biased state 2, their distance is one. Consequently, transporting between this pair of states contributes $\frac{2}{100} \times 1$ to the distance of the fair state 1-2 and the biased state 1-2. Hence, the distance of the fair state 1-2 and the biased state 1-2 can be viewed as the cost of transporting one unit from the fair state 1-2 to the biased state 1-2, where the costs are given by the distances in (9). As the reader can easily verify, the above transportation plan gives rise to the minimal cost of transporting one unit from the fair state 1-2 to the biased state 1-2.

Recall from Section 4 that for probabilistic bisimilar states there exists a transportation plan that transports one unit from the one state to the other while only transporting between related states. In our quantitative generalization, we replace the relation with distances (compare (7) with (9)). In the quantitative setting, these distances play the role of costs and we are interested in the minimal cost of transporting one unit from the one state to the other.

Next, let us consider the fair and biased versions of state 1-3. In this example, the probabilities of the transitions match exactly, but the reachable states do not behave exactly the same.

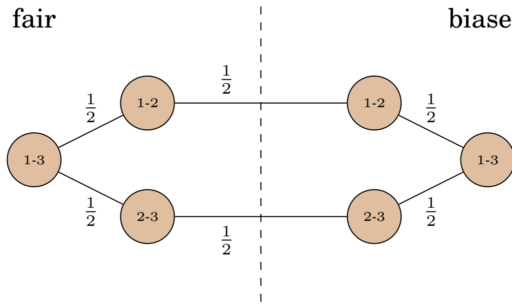


The fair and biased state 2-3 are probabilistic bisimilar and, therefore, have distance zero. As we already discussed earlier, the fair and biased state 1-2 have distance $\frac{2}{100}$. The other distances can be found in the following diagram.



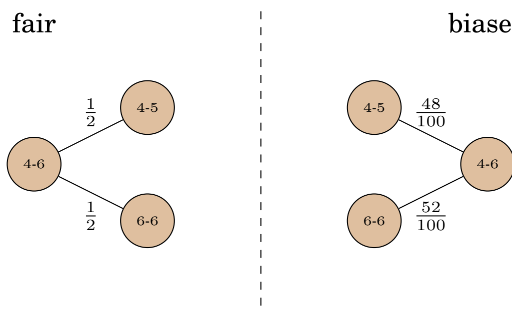
From the above two diagrams we can conclude the following. With probability $\frac{1}{2}$ the fair (biased) state 1-3 transitions to the fair (biased) state 1-2. Although the fair and biased states 1-2 are not probabilistic bisimilar, their behaviour is very similar, reflected by their small distance. With the remaining probability $\frac{1}{2}$ the fair (biased) state 1-3 transitions to the fair (biased) state 2-3. The latter two states are probabilistic bisimilar and, hence, have distance zero. Therefore, the similarity of the behaviour of the fair and biased states 1-3 is captured by the weighted sum $\frac{1}{2} \times \frac{2}{100} + \frac{1}{2} \times 0 = \frac{1}{100}$.

Given the above distances, a transportation plan that minimizes the cost of transporting one unit between the fair and biased states 1-3 is captured by the following diagram.

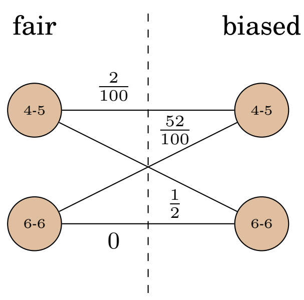


The cost of the above transportation plan is $\frac{1}{2} \times \frac{2}{100} + \frac{1}{2} \times 0 = \frac{1}{100}$. Hence, also in this example the distance between the fair and biased states 1-3 can be viewed as the minimal cost of transporting one unit from the fair state 1-3 to its biased counterpart.

As a third and final example, let us consider the fair and biased versions of state 4-6. In this case, neither do the probabilities of the transitions match, nor do the reachable states behave exactly the same.

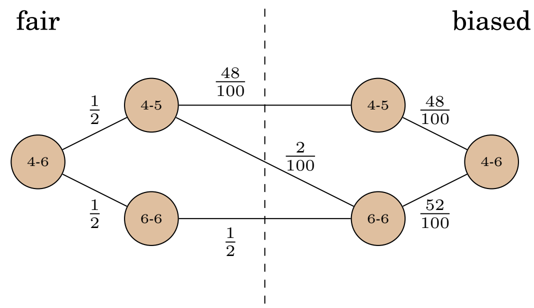


The distances between the relevant states are given in the following diagram.



From the above two diagrams we can conclude the following. With probability $\frac{1}{2}$ the fair and biased state 4-6 transition to states that behave exactly the same, namely the fair and biased states 6-6. With probability $\frac{48}{100}$ the states transition to the fair and biased states 4-5 which behave almost the same, reflected by the fact that their distance is $\frac{2}{100}$. With the remaining probability, namely probability $\frac{2}{100}$, the states transition to the fair state 4-5 and the biased state 6-6. Therefore, the weighted sum $\frac{1}{2} \times 0 + \frac{48}{100} \times \frac{2}{100} + \frac{2}{100} \times \frac{1}{2} = \frac{196}{10000}$ captures the expected similarity in behaviour of the fair and biased states 4-6.

Also in this example the distance can be viewed as the minimum cost of transporting one unit from the fair state 4-6 to its biased counterpart. An optimal transportation plan is depicted below.



The cost of the above transportation plan is $\frac{196}{10000}$.

In the remainder of this section, we will present the technical details of our quantitative generalization of probabilistic bisimilarity. Using pseudometrics, we will generalize the notion of lifting in terms of couplings (Theorem 4.6), the notion of probabilistic bisimulation (Definition 4.1), and the notion probabilistic bisimilarity (Definition 4.4).

A key ingredient of the definition of probabilistic bisimulation is the lifting of an equivalence relation on states to a relation on distributions over those states. In our quantitative setting we lift a pseudometric on states to a distance function on probability distributions over those states. This quantitative notion of lifting is defined as follows.

Definition 5.1. The *lifting* of a pseudometric $d : S \times S \to [0, 1]$ is the function $d{\uparrow} : \mathit{Distr}(S) \times \mathit{Distr}(S) \to [0, 1]$ defined by

$$d{\uparrow}(\mu, \nu) = \min_{\omega \in \Omega(\mu,\nu)} \sum_{u,v \in S} \omega(u, v)\, d(u, v).$$

The above definition is the dual representation of the distance function introduced by Kantorovich in [Kantorovich 1942] (see the proof of Proposition 5.3 for the original definition). Let s, $t \in S$. Recall from Section 4 that a coupling $\omega \in \Omega(\tau(s), \tau(t))$ can be viewed as a transportation plan. Provided that the pseudometric d captures the transportation costs, $d{\uparrow}(\tau(s), \tau(t))$ is the minimum cost of transporting one unit from state s to state t. The distance function due to Kantorovich is also known as the earth mover's distance (see, for example, [Deng and Du 2009]). This distance function was first used to define a behavioural pseudometric in joint work with Worrell [van Breugel and Worrell 2001].

Earlier in this section we considered three examples. For each case, we constructed a transportation plan with minimal cost. These are all examples of liftings.

Next, we generalize the notion probabilistic bisimulation relation to our quantitative setting. The following definition is a special case of [Deng et al. 2005, Definition 2.5], which considers systems that generalize labelled Markov chains.

Definition 5.2. A pseudometric $d : S \times S \to [0, 1]$ is a *probabilistic bisimulation* if for all $s, t \in S$ and $\epsilon \in [0, 1)$, if $d(s, t) \leq \epsilon$ then $\ell(s) = \ell(t)$ and $d{\uparrow}(\tau(s), \tau(t)) \leq \epsilon$.

Consider the labelled Markov chain (1) in the introduction. Let the pseudometric d_1 is defined by

$$d_1(s, t) = \begin{cases} 0 & \text{if } s = t \\ 1 & \text{otherwise} \end{cases}$$

This distance function is known as the discrete pseudometric. We leave it to the reader to verify that it is a probabilistic bisimulation. Note that this pseudometric only captures that each state behaves the same as itself.

Now consider the pseudometric d_2 is defined by

$$d_2(s,t) = \begin{cases} 0 & \text{if } s \text{ and } t \text{ are probabilistic bisimilar} \\ 1 & \text{otherwise} \end{cases}$$

The pseudometric d_2 captures that probabilistic bisimilar states behave the same and, hence, provides more information than d_1. To conclude that the pseudometric d_2 is a probabilistic bisimulation, it suffices to show that if the states s and t are probabilistic bisimilar then $\ell(s) = \ell(t)$ and $d_2{\uparrow}(\tau(s), \tau(t)) = 0$. Assume that the states s and t are probabilistic bisimilar. From the definition of probabilistic bisimulation relation and Theorem 4.6 we can conclude that $\ell(s) = \ell(t)$ and there exists $\omega \in \Omega(\tau(s), \tau(t))$ such that for all u, $v \in \text{support}(\omega)$, states u and v are probabilistic bisimilar and, therefore, $d_2(u,v) = 0$. Hence, $\sum_{u,v \in S} \omega(u,v)\, d_2(u,v) = 0$ and as a consequence $d_2{\uparrow}(\tau(s), \tau(t)) = 0$.

Let us slightly modify the above pseudometric d_2 by redefining

$$d_3(\text{1-2}, \text{2-3}) = d_3(\text{2-3}, \text{1-2}) = d_3(\text{1-2}, \text{6-6}) = d_3(\text{6-6}, \text{1-2}) = \tfrac{1}{2}.$$

To deduce that this modified pseudometric d_3 is a probabilistic bisimulation it remains to show that $d_3{\uparrow}(\tau(\text{1-2}), \tau(\text{2-3})) \leq \frac{1}{2}$ and $d_3{\uparrow}(\tau(\text{1-2}), \tau(\text{6-6})) \leq \frac{1}{2}$. Consider the coupling defined by

$$\omega(1,2) = \tfrac{1}{2} \text{ and } \omega(2,3) = \tfrac{1}{2}.$$

Since $\sum_{u,v \in S} \omega(u,v)\, d_3(u,v) = \frac{1}{2} \times 1 + \frac{1}{2} \times 0 = \frac{1}{2}$, we can conclude $d_3{\uparrow}(\tau(\text{1-2}), \tau(\text{2-3})) \leq \frac{1}{2}$. The fact that $d_3{\uparrow}(\tau(\text{1-2}), \tau(\text{6-6})) \leq \frac{1}{2}$ can be proved similarly. Since four of the distances of the pseudometric d_3 are smaller than the corresponding ones of the pseudometric d_2, d_3 provides more information about the behavioural similarity of the state pairs than d_2.

Finally, consider the pseudometric d_4 defined by the following table.

	1	2	3	4	5	6	1-2	2-3	4-5	6-6	1-3	4-6	1-6
1	0	1	1	1	0	1	1	1	1	1	1	1	1
2	1	0	0	0	1	0	1	1	1	1	1	1	1
3	1	0	0	0	1	0	1	1	1	1	1	1	1
4	1	0	0	0	1	0	1	1	1	1	1	1	1
5	0	1	1	1	0	1	1	1	1	1	1	1	1
6	1	0	0	0	1	0	1	1	1	1	1	1	1
1-2	1	1	1	1	1	1	0	$\frac{1}{2}$	0	$\frac{1}{2}$	1	1	1
2-3	1	1	1	1	1	1	$\frac{1}{2}$	0	$\frac{1}{2}$	0	1	1	1
4-5	1	1	1	1	1	1	0	$\frac{1}{2}$	0	$\frac{1}{2}$	1	1	1
6-6	1	1	1	1	1	1	$\frac{1}{2}$	0	$\frac{1}{2}$	0	1	1	1
1-3	1	1	1	1	1	1	1	1	1	1	0	0	1
4-6	1	1	1	1	1	1	1	1	1	1	0	0	1
1-6	1	1	1	1	1	1	1	1	1	1	1	1	0

We leave it to the reader to verify that this pseudometric is a probabilistic bismulation as well.

As we have seen in Section 4, the larger a probabilistic bisimulation relation, the more state pairs it contains that behave the same. Dually, the smaller a probabilistic bisimulation pseudometric, the more information about the behavioural similarity of the states it gives us. Let d and e be pseudometrics on S. Then d is smaller than e, denoted $d \sqsubseteq e$, if $d(s,t) \leq e(s,t)$ for all s, $t \in S$. The four pseudometrics introduced above are related as follows: $d_4 \sqsubseteq d_3 \sqsubseteq d_2 \sqsubseteq d_1$. As a consequence, of the four d_1 provides the least amount of information and d_4 gives us the most.

As we have shown in Proposition 4.3, there exists a largest probabilistic bisimulation relation. Dually, there exists a smallest probabilistic bisimulation pseudometric as we will prove next. A similar result, without proof, can be found in [Deng et al. 2005].

PROPOSITION 5.3. *There exists a smallest probabilistic bisimulation pseudometric.*

PROOF. We denote the set of pseudometrics on S by $Pseudo(S)$. This set equipped with the partial order $\sqsubseteq$ forms a complete lattice (see, for example, [Desharnais et al. 2002, Lemma 3.2]). The function $\Delta : Pseudo(S) \to [0,1]^{S \times S}$ is defined by

$$\Delta(d)(s,t) = \begin{cases} 1 & \text{if } \ell(s) \neq \ell(t) \\ d{\uparrow}(\tau(s), \tau(t)) & \text{otherwise} \end{cases}$$

Let d be a pseudometric. According to the Kantorovich-Rubinstein duality theorem [Kantorovich and Rubinstein 1958], we have that

$$d{\uparrow}(\mu, \nu) = \max_{f \in (S,d) \nrightarrow [0,1]} \left| \sum_{s \in S} f(s)\,(\mu(s) - \nu(s)) \right|$$

where $f \in (S,d) \nrightarrow [0,1]$ if $|f(s) - f(t)| \leq d(s,t)$ for all $s, t \in S$. From this dual representation, which is Kantorovich's original definition, we can immediately conclude that $d{\uparrow}$ is a pseudometric. Hence, $\Delta(d)$ is a pseudometric as well. Therefore, Δ is a function on the complete lattice $Pseudo(S)$.

Let d and e be pseudometrics with $d \sqsubseteq e$. Then $d{\uparrow} \sqsubseteq e{\uparrow}$ and, therefore, $\Delta(d) \sqsubseteq \Delta(e)$. That is, the function Δ is order-preserving.

Next, we prove that a pseudometric d is a probabilistic bisimulation if and only if $\Delta(d) \sqsubseteq d$, that is, d is a pre-fixed point of Δ (see [Davey and Priestley 2002, Definition 8.14]). We prove two implications. Assume that d is a probabilistic bisimulation. Let s, $t \in S$. We distinguish two cases.

— If $\ell(s) \neq \ell(t)$ then we can conclude that $d(s,t) = 1$ from the definition of probabilistic bisimulation. Hence,

$$\Delta(d)(s,t) \leq 1 = d(s,t).$$

— If $\ell(s) = \ell(t)$ then

$$\Delta(d)(s,t) = d{\uparrow}(\tau(s), \tau(t)) \leq d(s,t)$$

by the definition of probabilistic bisimulation.

To prove the other implication, assume that $\Delta(d) \sqsubseteq d$. Let s, $t \in S$ and $\epsilon \in [0,1)$. Suppose that $d(s,t) \leq \epsilon$. Since

$$\Delta(d)(s,t) \leq d(s,t) \leq \epsilon < 1$$

we can conclude that $\ell(s) = \ell(t)$ and $d{\uparrow}(\tau(s), \tau(t)) = \Delta(d)(s,t) \leq \epsilon$ from the definition of Δ. Hence, d is a probabilistic bisimulation.

Since Δ is an order-preserving function on a complete lattice, we can conclude from [Davey and Priestley 2002, Theorem 8.20] that Δ has a least pre-fixed point, which is the smallest probabilistic bisimulation pseudometric. □

Desharnais et al. [Desharnais et al. 2002] were the first to define a behavioural pseudometric as a least fixed point.

Definition 5.4. The *probabilistic bisimilarity pseudometric* is the smallest probabilistic bisimulation pseudometric.

This pseudometric captures the behavioural similarity of the states of a labelled Markov chain. The pseudometric d_4 presented above is the probabilistic bisimilarity pseudometric for the labelled Markov chain (1) in the introduction. Desharnais et al. [Desharnais et al. 1999, Theorem 1] proved the following fundamental result.

THEOREM 5.5. *The probabilistic bisimilarity distance of states s and t is zero if and only if s and t are probabilistic bisimilar.*

Hence, the probabilistic bisimilarity pseudometric provides a quantitative generalization of probabilistic bisimilarity.

6. CONCLUSION

As we already mentioned several times, probabilistic bisimilarity is not robust. This, however, does not imply that the behavioural equivalence is useless. On the contrary, probabilistic bisimilarity still plays a central role even in the presence of robust alternatives such as probabilistic bisimilarity distances. First of all, it provides a natural notion of behavioural equivalence for labelled Markov chains (see, for example, [Panangaden 2015]). As a consequence, it provides a soundness check for behavioural pseudometrics as we want distance zero to exactly capture behavioural equivalence. Secondly, probabilistic bisimilarity also plays a role in the policy iteration algorithm to compute the probabilistic bisimilarity distances by Bacci et al. [Bacci et al. 2013]. In particular, probabilistic bisimilarity needs to be decided before we can use policy iteration to compute the distances, as shown in joint work with Tang [Tang and van Breugel 2016]. Thirdly, probabilistic bisimilarity also impacts the polynomial time algorithm to compute the probabilistic bisimilarity distances, presented in joint work with Chen and Worrell [Chen et al. 2012]. In this algorithm the distances are computed by means of Khachiyan's ellipsoid method [Khachiyan 1979]. As shown in joint work with Tang [Tang and van Breugel 2017], if we first decide probabilistic bisimilarity then we can use a simpler version of the ellipsoid method to compute the distances.

In Section 5 we argued that the probabilistic bisimilarity distances provide a natural quantitative generalization of probabilistic bisimilarity. As mentioned in the introduction, the probabilistic bisimilarity distances can be characterized by means of a real valued interpretation of a logic, a result due to Desharnais et al. [Desharnais et al. 1999]. As shown in joint work with Shalit and Worrell [van Breugel et al. 2002], these distances can also be characterized in terms of tests. In joint work with Worrell [van Breugel and Worrell 2014], we show that the distances can be captured as the values of a game. Furthermore, the distances can be captured by a coalgebra as first shown in joint work with Worrell [van Breugel and Worrell 2001].

Hopefully, we have convinced the reader that the probabilistic bisimilarity distances defined in Section 5 are a natural generalization of probabilistic bisimilarity. We find it compelling that these distances are defined in terms of the Kantorovich metric, a natural distance function on probability distributions, and that they can be elegantly characterized in terms of a logic, tests, a game and a coalgebra.

6.1. Acknowledgements

This paper was written while the author was visiting the Simons Institute for the Theory of Computing. This research was also supported by the Natural Sciences and Engineering Research Council of Canada. The author is thankful to Giorgio Bacci, Giovanni Bacci, Michael Mislove, Prakash Panangaden, Qiyi Tang and James Worrell for their detailed and constructive feedback on a draft version of this paper, and to Di Chen, Claudio Hermida, Michael Makkai, Mike Mislove, Joel Ouaknine, Steven Shalit, Babita Sharma, Qiyi Tang, Eugene Voronov, Hongming Wu, and in particular James Worrell for their collaborative research on behavioural pseudometrics.

REFERENCES

Giorgio Bacci, Giovanni Bacci, Kim Larsen, and Radu Mardare. 2013. On-the-fly exact computation of bisimilarity distances. In *Proceedings of the 19th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (Lecture Notes in Computer Science)*, Nir Piterman and Scott Smolka (Eds.), Vol. 7795. Springer-Verlag, Rome, Italy, 1–15.

Giorgio Bacci, Giovanni Bacci, Kim Larsen, and Radu Mardare. 2017. On the metric-based approximate minimization of Markov chains. In *Proceedings of the 44th International Colloquium on Automata, Languages and Programming (Leibniz International Proceedings in Informatics)*, Ioannis Chatzigiannakis, Piotr Indyk, Fabian Kuhn, and Anca Muscholl (Eds.), Vol. 80. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Warsaw, Poland, 104:1–104:14.

Di Chen, Franck van Breugel, and James Worrell. 2012. On the complexity of computing probabilistic bisimilarity. In *Proceedings of the 15th International Conference on Foundations of Software Science and Computational Structures (Lecture Notes in Computer Science)*, Lars Birkedal (Ed.), Vol. 7213. Springer-Verlag, Tallinn, Estonia, 437–451.

Brian Davey and Hilary Priestley. 2002. *Introduction to lattices and order*. Cambridge University Press, Cambridge, United Kingdom.

Yuxin Deng, Tom Chothia, Catuscia Palamidessi, and Jun Pang. 2005. Metrics for action-labelled quantitative transition systems. In *Proceedings of the 3rd Workshop on Quantitative Aspects of Programming Languages (Electronic Notes in Theoretical Computer Science)*, Antonio Cerone and Herbert Wiklicky (Eds.), Vol. 153(2). Elsevier, Edinburgh, Scotland, 79–96.

Yuxin Deng and Wenjie Du. 2009. The Kantorovich metric in computer science: a brief survey. In *Proceedings of the 7th Workshop on Quantitative Aspects of Programming Languages (Electronic Notes in Theoretical Computer Science)*, Christel Baier and Alessandra di Pierro (Eds.), Vol. 253(3). Elsevier, York, United Kingdom, 73–82.

Josée Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. 1999. Metrics for labeled Markov systems. In *Proceedings of the 10th International Conference on Concurrency Theory (Lecture Notes in Computer Science)*, Jos Baeten and Sjouke Mauw (Eds.), Vol. 1664. Springer-Verlag, Eindhoven, The Netherlands, 258–273.

Josée Desharnais, Vineet Gupta, Radha Jagadeesan, and Prakash Panangaden. 2002. The metric analogue of weak bisimulation for probabilistic processes. In *Proceedings of the 17th Annual IEEE Symposium on Logic in Computer Science*. IEEE, Copenhagen, Denmark, 413–422.

Wolfgang Doeblin. 1938. Exposé de la théorie des chaînes simples constantes de Markov à un nombre fini d'états. *Revue mathématique de l'union interbalkanique* 2 (1938), 77–105.

Alessandro Giacalone, Chi-Chang Jou, and Scott Smolka. 1990. Algebraic reasoning for probabilistic concurrent systems. In *Proceedings of the IFIP WG 2.2/2.3 Working Conference on Programming Concepts and Methods*. North-Holland, Sea of Gallilee, Israel, 443–458.

Bengt Jonsson and Kim Larsen. 1991. Specification and refinement of probabilistic processes. In *Proceedings of the 6th Annual Symposium on Logic in Computer Science*. IEEE, Amsterdam, The Netherlands, 266–277.

Leonid Kantorovich. 1942. On the transfer of masses (in Russian). *Doklady Akademii Nauk* 5, 1 (1942), 1–4. Translated in *Management Science*, 5, 1 (October 1958), 1–4.

Leonid Kantorovich and Gennadi Rubinstein. 1958. On the space of completely additive functions (in Russian). *Vestnik Leningradskogo Universiteta* 3, 2 (1958), 52–59.

Leonid Khachiyan. 1979. A polynomial algorithm in linear programming. *Soviet Mathematics Doklady* 20, 1 (1979), 191–194.

Kim Larsen and Arne Skou. 1989. Bisimulation through probabilistic testing. In *Proceedings of the 16th Annual ACM Symposium on Principles of Programming Languages*. ACM, Austin, TX, USA, 344–352.

Abhishek Murthy, Md. Ariful Islam, Ezio Bartocci, Elizabeth Cherry, Flavio Fenton, James Glimm, Scott Smolka, and Radu Grosu. 2012. Approximate bisimulations for sodium channel dynamics. In *Proceedings of 10th International Conference on Computational Methods in Systems Biology (Lectures Notes in Computer Science)*, David Gilbert and Monika Heiner (Eds.), Vol. 7605. Springer-Verlag, London, United Kingdom, 267–287.

Prakash Panangaden. 2015. Probabilistic bisimulation. *ACM SIGLOG News* 2, 3 (July 2015), 72–84.

Michel Reniers, Rob Schoren, and Tim Willemse. 2014. Results on embeddings between state-based and event-based systems. *The Computer Journal* 57, 1 (Jan. 2014), 73–92.

Prithviraj Sen, Amol Deshpande, and Lise Getoor. 2009. Bisimulation-based approximate lifted inference. In *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence*, Jeff Bilmes and Andrew Ng (Eds.). AUAI Press, Montreal, QC, Canada, 496–505.

Qiyi Tang and Franck van Breugel. 2016. Computing probabilistic bisimilarity distances via policy iteration. In *Proceedings of the 27th International Conference on Concurrency Theory (Leibniz International Proceedings in Informatics)*, Josée Desharnais and Radha Jagadeesan (Eds.), Vol. 59. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Quebec City, QC, Canada, 22:1–22:15.

Qiyi Tang and Franck van Breugel. 2017. Algorithms to compute probabilistic bisimilarity distances for labelled Markov chains. In *Proceedings of the 28th International Conference on Concurrency Theory (Leibniz International Proceedings in Informatics)*, Roland Meyer and Uwe Nestmann (Eds.), Vol. 85. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Berlin, Germany, 27:1–27:16.

Franck van Breugel, Claudio Hermida, Michael Makkai, and James Worrell. 2005. An accessible approach to behavioural pseudometrics. In *Proceedings of the 32nd International Colloquium on Automata, Languages and Programming (Lecture Notes in Computer Science)*, Luís Caires, Giuseppe Italiano, Luís Monteiro, Catuscia Palamidessi, and Moti Yung (Eds.), Vol. 3580. Springer-Verlag, Lisbon, Portugal, 1018–1030.

Franck van Breugel, Steven Shalit, and James Worrell. 2002. Testing labelled Markov processes. In *Proceedings of the 29th International Colloquium on Automata, Languages, and Programming (Lecture Notes in Computer Science)*, Peter Widmayer, Francisco Triguero Ruiz, Rafael Morales-Bueno, Matthew Hennessy, Stephan Eidenbenz, and Ricardo Conejo (Eds.), Vol. 2380. Springer-Verlag, Malaga, Spain, 537–548.

Franck van Breugel and James Worrell. 2001. Towards quantitative verification of probabilistic systems. In *Proceedings of 28th International Colloquium on Automata, Languages and Programming (Lecture Notes in Computer Science)*, Fernando Orejas, Paul Spirakis, and Jan van Leeuwen (Eds.), Vol. 2076. Springer-Verlag, Heraklion, Crete, 421–432.

Franck van Breugel and James Worrell. 2014. The complexity of computing a bisimilarity pseudometric on probabilistic automata. In *Horizons of the Mind – A Tribute to Prakash Panangaden*, Franck van Breugel, Elham Kashefi, Catuscia Palamidessi, and Jan Rutten (Eds.). Lecture Notes in Computer Science, Vol. 8464. Springer-Verlag, Oxford, United Kingdom, 191–213.