Logical Characterization of Bisimulation for Transition Relations over Probability Distributions with Internal Actions*

Matias David Lee[1] and Erik P. de Vink[2]

1 Univ Lyon, ENS de Lyon, CNRS, UCB Lyon 1, LIP, France and FaMAF, UNC–CONICET, Córdoba, Argentina
matias-david.lee@ens-lyon.fr
2 Eindhoven University of Technology, The Netherlands and CWI, Amsterdam, The Netherlands
evink@win.tue.nl.

Abstract

In recent years the study of probabilistic transition systems has shifted to transition relations over distributions to allow for a smooth adaptation of the standard non-probabilistic apparatus. In this paper we study transition relations over probability distributions in a setting with internal actions. We provide new logics that characterize probabilistic strong, weak and branching bisimulation. Because these semantics may be considered too strong in the probabilistic context, Eisentraut et al. recently proposed weak distribution bisimulation. To show the flexibility of our approach based on the framework of van Glabbeek for the non-deterministic setting, we provide a novel logical characterization for the latter probabilistic equivalence as well.

1998 ACM Subject Classification F.3.2 Semantics of Programming Languages, F.4.1 Mathematical Logic

Keywords and phrases Probabilistic Transition Systems, Weak Bisimulations, Logical Characterization, Transition Relation over Distributions, Modal Logics

Digital Object Identifier 10.4230/LIPIcs.MFCS.2016.29

1 Introduction

Labeled transition systems (LTS) are a standard way of modeling processes. To verify processes, process theory has embraced two related lines of research, viz. *behavioral equivalences* and *modal logics*. Behavioral equivalences state when two processes present the same behavior. Corresponding minimization algorithms facilitate, e.g., state space reduction. On the other hand, modal logics allow to express extensional properties of processes to be used, for example, in formal specification and verification of systems. In classical cases, a logic characterizes a particular equivalence; two processes are equivalent precisely when they satisfy the same logical formula. In such a situation, when two processes are not equivalent, the logic has a formula that is only satisfied by one of the processes. In a way, the particular formula provides an explanation why the two processes do not have the same behavior.

The introduction of probabilistic transition systems called for an extension of the results known for the non-probabilistic context to the probabilistic one. In [10], Hennessy takes *"a*

* M.D. Lee has been supported by EU7FP grant agreement 295261 (MEALS), SECyT–UNC and the project ANR 12IS02001 PACE

fresh look at strong probabilistic bisimulation for processes which exhibit both non-determinism and probabilistic behavior". In his view a process is not in a single state of a probabilistic LTS, but control is spread over a distribution of states. Operationally, this makes sense, because after the execution of an action, a process reaches a set of states with a particular distribution. Thus, transitions of the shape $s \xrightarrow{a} \mu$ are replaced by transitions of the shape $\mu \xrightarrow{a} \mu'$, where s is a state and μ and μ' are distributions over states. This underlies the smooth adaptation of many results for strong bisimulation to probabilistic strong bisimulation in [10].

In this paper we study transitions over distributions in the context of *weak semantics*, i.e. semantics that allow to abstract from internal actions, and the logical characterization of these semantics. We present two concrete enhancements w.r.t. earlier works. First, we do not require processes to be *divergence-free*, i.e. processes may execute infinite τ-actions. Second, we do not focus our attention on weak simulation and bisimulation only; our goal is provide a framework to deal with more general weak semantics. Key for this development is to define what constitutes a transition between two distributions. Following [14, 9], we opt for *hyper-transitions*. Different alternatives and variations of hyper-transitions appear in the literature, e.g. [6, 20, 5, 2]. In [2] also provides a comparison. Given a notion of transitions over distributions, there is a natural way to adapt the relational definition of a semantics from the state-based context to the distributions-based context: just take the standard definition and replace states by distributions.

We discuss *probabilistic strong bisimulation*, *probabilistic weak bisimulation* and *probabilistic branching bisimulation*. For each semantics we propose a logical characterization. These characterizations follow the set-up for the non-deterministic context of van Glabbeek presented in [21]. However, we add in particular the modal operator $[\cdot]_{\geq q}$ to measure probabilities. Because probabilistic weak bisimulation, as one can argue, may be considered too strong, [9] introduces a variant of weak bisimulation, so-called weak distribution bisimulation. We also introduce a logical characterization for this semantics. The peculiarity of this logic is the way probabilities are measured, for which the modal operator $\oplus_q$ is used as introduced in [10]. However, the semantics of $[\cdot]_{\geq q}$ and $\oplus_q$ are rather different: $[\cdot]_{\geq q}$ governs the support of a distribution, while $\oplus_q$ concerns decomposition of a distribution.

Related work on logics for distributions includes [17, 7, 18, 10, 2]. In [17, 2], the idea of transitions over sets of states or distributions does not appear. There, the semantics of the prefix operator depends on the actions that can be executed by the states in the support of the distribution rather than the distribution itself. In addition, in [17] the relational characterization is given for a probabilistic LTS and not for transition relation over distributions. Inspired by the different logics, [2] introduces relations over distributions. On the other hand, [10] takes into account the idea of transitions over set of states but it only focuses on probabilistic strong bisimilarity. Also [7, 18] deal with logics over distributions. The logics introduced in [7] are variants of the probabilistic μ-calculus. Two of these variants, that do not use fix-point operators, characterize weak and strong bisimulation. In [18] only extensions of HML logics are considered. The main difference between [7, 18] and our work lies in the definition of hyper-transitions; in [7, 18], only divergence-free processes are considered. Moreover, they only consider logics to characterize weak (bi)similarity and the approach cannot be uniformly extended other weak semantics as in our case. Specific details on the relation relating to the work mentioned are discussed throughout the paper.

Since the seminal work of [13] on the logical characterization of probabilistic strong bisimulation many extensions have been presented. Work on the logical characterization of weak bisimulation in the probabilistic setting includes [8, 19]. In [8], Desharnais et al. prove that, for the alternating model, PCTL* is sound and complete with respect to weak bisimulation. Song et al. do a similar study taking into account probabilistic automata, see [19]. First, they show that the logic is sound but not complete for strong bisimulation. For this reason, they introduce a variant of the semantics such that PCTL* is complete too. This new semantics relaxes the transfer property: existence of a matching (combined) transition is replaced by existence of a transition of at least the same weight on downward closed sets. Along the same lines a variant for weak bisimulation is obtained. A counterexample in [19] shows that the results for the alternating model in [8] do not hold for probabilistic automata.

The remainder of the paper is organized as follows. In Section 2, following [9], we review *probabilistic automata*, the model used to formalize probabilistic LTS, and transitions between two distributions, so-called *hyper-transitions*, taking into account both visible and internal actions. In Section 3, we introduce the relational characterizations of the various semantics and their logical characterization. Section 4 collects concluding remarks.

2 Preliminaries

For a set X, we denote by $\mathit{SubDisc}(X)$ the set of discrete sub-probability distributions over X. Given $\varrho \in \mathit{SubDisc}(X)$, we denote by $\mathit{spt}(\varrho)$ the support of ϱ, i.e. the set $\{ x \in X \mid \varrho(x) > 0 \}$, by $\varrho(\perp)$ the value $1 - \varrho(X)$, for a distinguished symbol $\perp \notin X$. For $x \in X$, we use $\delta(x)$ to denote the *Dirac* distribution of x given by $\delta(x)(y) = 1$ for $y = x$, 0 otherwise; $\delta_\perp$ represents the empty distribution with $\delta_\perp(X) = 0$. We call a distribution a *probability* distribution if $\varrho(X) = 1$. The set of all discrete probability distributions over X is denoted by $\mathit{Disc}(X)$. Given $\{x_1, \ldots, x_n\} \subseteq X$ and $p_1, \ldots, p_n > 0$ such that $p_1 + \cdots + p_n = 1$, we write $\sum_{i=1}^{n} p_i x_i$ to denote the distribution that assigns probability p_i to x_i, for $i = 1, \ldots, n$. In addition, given distributions $\mu_1, \ldots, \mu_n \in \mathit{Disc}(X)$, the distribution $\sum_{i=1}^{n} p_i \mu_i$ is called a convex combination of μ_1 to μ_n. If $n = 2$, we may write $\mu_1 \oplus_p \mu_2$ where $p = p_1$ instead of $p_1\mu_1 + p_2\mu_2$.

We reserve the symbol τ to denote the silent action. For a set X with $\tau \notin X$ we write X_τ for $X \cup \{\tau\}$.

▶ **Definition 1.** A *probabilistic automaton* or PA $\mathcal{A}$ is a tuple (S, Σ_τ, D), where S is the finite set of states, Σ_τ is the set of actions, and $D \subseteq S \times \Sigma_\tau \times \mathit{Disc}(S)$ is the transition relation.

For the rest of the paper we assume that a PA $\mathcal{A} = (S, \Sigma_\tau, D)$ is given. Moreover, $\mathcal{A}$ is *image-finite*, i.e. for all $a \in \Sigma_\tau$ and $s \in S$, the set $\{ \mu \mid (s, a, \mu) \in D \}$ is finite. We write $s \xrightarrow{a} \mu$ for $(s, a, \mu) \in D$. We write $D(a)$ for the set of transitions with label a and $D(s)$ for the set of transitions with source s.

An *execution fragment* $\alpha = s_0 a_1 s_1 a_2 s_2 \ldots$ of $\mathcal{A}$ is a finite or infinite alternating sequence of states and actions such that for each $i > 0$ there exists a transition $(s_{i-1}, a_i, \mu_i) \in D$ with $\mu_i(s_i) > 0$. We say, α is starting from $\mathit{fst}(\alpha) = s_0$, and in case the sequence is finite, ending in $\ell\mathit{st}(\alpha)$. We use $\mathit{frags}(\mathcal{A})$ to denote the set of execution fragments of $\mathcal{A}$, and by $\mathit{ffrags}(\mathcal{A})$ the set of finite execution fragments of $\mathcal{A}$. An execution fragment α is a prefix of an execution fragment α', notation $\alpha \preccurlyeq \alpha'$, if the sequence α is a prefix of the sequence α'. The trace $\mathit{trace}(\alpha)$ of α is the subsequence of non-silent actions of α. We use ε to denote the empty trace. Thus, $\mathit{trace}(a) = a$ for $a \in \Sigma$ and $\mathit{trace}(\tau) = \varepsilon$.

A *scheduler* for $\mathcal{A}$ is a map $\sigma \colon \mathit{ffrags}(\mathcal{A}) \to \mathit{SubDisc}(D)$ with $\sigma(\alpha) \in \mathit{SubDisc}(D(\ell\mathit{st}(\alpha)))$ for every finite execution fragment α. The scheduler is *deterministic* if for every α, $\sigma(\alpha)$ is a Dirac distribution or $\delta_\perp$. Note that by using sub-probability distributions, it is possible that with non-zero probability no transition is chosen after α, that is, the computation stops after α with probability $\sigma(\alpha)(\perp)$. Given a scheduler σ and a finite execution fragment α, the

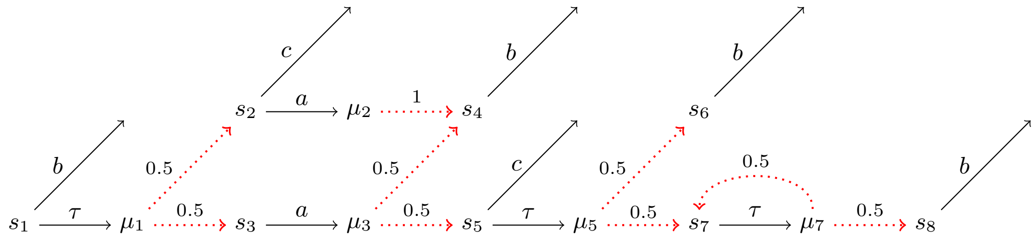

Figure 1 Example probabilistic automaton $\mathcal{A}$.

distribution $\sigma(\alpha)$ describes how transitions are chosen to move on from $\ell st(\alpha)$. A scheduler σ and a state s induce a probability distribution $\mu_{\sigma,s}$ over execution fragments as follows.

The cone C_α of a finite fragment α is the set $\{\, \alpha' \in \mathit{frags}(\mathcal{A}) \mid \alpha \preccurlyeq \alpha' \,\}$. Given a scheduler σ and states s and t, the distribution $\mu_{\sigma,s}$ on cones C_α is recursively defined by

$$\mu_{\sigma,s}(C_t) = \delta(s)(t) \qquad \mu_{\sigma,s}(C_{\alpha a t}) = \mu_{\sigma,s}(C_\alpha) \cdot \textstyle\sum_{\ell st(\alpha) \xrightarrow{a} \mu} \sigma(\alpha)\big(\ell st(\alpha) \xrightarrow{a} \mu\big) \cdot \mu(t)$$

For a finite execution fragment α, the *probability* $\mu_{\sigma,s}(\alpha)$ *of executing* α *(and stop)* based on σ and s is defined as $\mu_{\sigma,s}(\alpha) = \mu_{\sigma,s}(C_\alpha) \cdot \sigma(\alpha)(\bot)$.

A state s can execute a combined weak transition for an action $a \in \Sigma$ if there is a scheduler σ such that with probability 1 the action a is executed once while no other visible action is executed. After a is executed, a state t will be reached with probability $\mu_{\sigma,s}(\{\, \alpha \in \mathit{ffrags}(\mathcal{A}) \mid \ell st(\alpha) = t \,\})$. If $a = \tau$, we have a similar definition but with probability 1 no visible action is executed. Definition 2 takes both cases into account. As usual, $\hat{a} = a$ if $a \in \Sigma$ and $\hat{a} = \varepsilon$ if $a = \tau$

▶ **Definition 2.** Let $s \in S$ and $a \in \Sigma_\tau$. A transition $s \overset{\hat{a}}{\Longrightarrow}_c \mu$ is called a *weak combined transition* if there exists a scheduler σ such that $\mu_{\sigma,s}$ satisfies the following:

1. $\mu_{\sigma,s}(\mathit{ffrags}(\mathcal{A})) = 1$,
2. for each $\alpha \in \mathit{ffrags}(\mathcal{A})$, if $\mu_{\sigma,s}(\alpha) > 0$, then $\mathit{trace}(\alpha) = \mathit{trace}(a)$,
3. for each state t, $\mu_{\sigma,s}(\{\, \alpha \in \mathit{ffrags}(\mathcal{A}) \mid \ell st(\alpha) = t \,\}) = \mu(t)$.

Occasionally we want to make reference to the scheduler σ underlying a weak combined transition $s \overset{\hat{a}}{\Longrightarrow}_c \mu$. We do so by writing $s \overset{\hat{a}}{\Longrightarrow}_\sigma \mu$. For execution fragment α, let $\ell gt(\alpha) = n$ in case $\alpha = s_0 a_1 s_1, \ldots a_n s_n$ is finite, and $\ell gt(\alpha) = \infty$ if α is infinite. We define the length of a scheduler σ with respect to a state s by $\ell gt_s(\sigma) = \sup\{\, \ell gt(\alpha) \mid \mathit{fst}(\alpha) = s, \sigma(\alpha) = \delta_\bot \,\}$.

▶ **Example 3.** Let $\mathcal{A}$ be the PA in Fig. 1. The state s_1 can execute the following weak combined transitions:

(i) $s_1 \overset{\varepsilon}{\Longrightarrow}_c \nu_0$ with $\nu_0(s_1) = 0.5$, $\nu_0(s_2) = 0.25$ and $\nu_0(s_3) = 0.25$;
(ii) $s_1 \overset{a}{\Longrightarrow}_c \nu_1$ with $\nu_1(s_4) = 0.75$ and $\nu_1(s_5) = 0.25$;
(iii) $s_1 \overset{a}{\Longrightarrow}_\sigma \nu_2$ with $\nu_2(s_4) = 0.75$, $\nu_2(s_5) = 0.05$ and $\nu_2(s_6) = \nu_2(s_7) = 0.1$, where σ stops at state s_5 with probability 0.2 and selects the transition $s_5 \xrightarrow{\tau} \mu_5$ with probability 0.8;
(iv) $s_7 \overset{\varepsilon}{\Longrightarrow}_\sigma \delta(s_8)$ where σ is such that $\ell gt(\sigma) = \infty$.

Notice there is no combined transition from s_1 that executes an action c, since from s_1, there is no scheduler that allows to execute this action with probability 1.

A *weak hyper-transition* is a linear combination of weak combined transitions with the same label, see [9]. The weight of each weak combined transition is defined by a distribution μ. The notion of a weak hyper-transition allows to work with transitions over distributions.

▶ **Definition 4.** Given $\mu, \mu' \in \mathit{Disc}(S)$ and $a \in \Sigma_\tau$, there is a *weak hyper-transition* $\mu \overset{\hat{a}}{\Longrightarrow}_c \mu'$ if there exists a family of weak combined transitions $\{s \overset{\hat{a}}{\Longrightarrow}_c \mu_s\}_{s \in spt(\mu)}$ such that $\mu' = \sum_{s \in spt(\mu)} \mu(s) \cdot \mu_s$.

Given a scheduler σ, the scheduler σ_n is such that $\sigma_n(\alpha) = \sigma(\alpha)$ if $\ell gt(\alpha) \leqslant n$, otherwise $\sigma_n(\alpha) = \delta_\perp$. For $a \in \Sigma_\tau$, we write $s \xrightarrow{a}_c \mu$ if there is a scheduler σ of length 1 such that $s \overset{\hat{a}}{\Longrightarrow}_\sigma \mu$. Schedulers of length 1 induce the notion of a *one-step transition*.

▶ **Definition 5.** Let $\mu, \mu' \in \mathit{Disc}(S)$ and $a \in \Sigma_\tau$. For $a \neq \tau$, a *one-step transition* $\mu \xrightarrow{a}_c \mu'$ is a weak hyper-transition $\mu \overset{a}{\Longrightarrow}_c \mu'$ for $\{s \overset{a}{\Longrightarrow}_c \mu_s\}_{s \in spt(\mu)}$ such that $s \xrightarrow{a}_c \mu_s$. A *one-step transition* $\mu \xrightarrow{\tau}_c \mu'$ is a weak hyper-transition $\mu \overset{\varepsilon}{\Longrightarrow}_c \mu'$ for $\{s \overset{\varepsilon}{\Longrightarrow}_c \mu_s\}_{s \in spt(\mu)}$ such that either $\mu_s = \delta(s)$ or $s \xrightarrow{\tau}_c \mu_s$, for $s \in spt(\mu)$, and $s \xrightarrow{\tau}_c \mu_s$ for at least one $s \in spt(\mu)$.

The definition of a one-step transition for a visible action requires that each state in the support of μ executes the visible action. On the other hand, if the action is not visible, we require that at least one state executes a τ-transition.

▶ **Example 6.** Consider again Example 3. Because of $s_1 \xrightarrow{a}_c \nu_1$ and $s_2 \xrightarrow{a}_c \mu_2$, $0.5\delta(s_1) + 0.5\delta(s_2) \xrightarrow{a}_c 0.5\nu_1 + 0.5\mu_2$. Since $s_1 \xrightarrow{\tau}_c 0.5\delta(s_1) + 0.5\mu_1$, we have $\delta(s_1) \xrightarrow{\varepsilon}_c 0.5\delta(s_1) + 0.5\mu_1$. Notice that this target distribution cannot be reached from $\delta(s_1)$ by a one-step transition with a deterministic scheduler. Definition 5 does not allow to split the state s_1. Finally, notice that $\delta(s_1) \overset{\varepsilon}{\Longrightarrow}_c \delta(s_1)$ but $\delta(s_1) \xrightarrow{\tau}_c \delta(s_1)$ is not a valid one-step hyper-transition.

We have used the notion of a hyper-transition of [14] to define transitions over distributions because of its clear operational intuition.

3 Semantics for Transitions Over Distributions and their Logical Characterizations

In this section we present four different semantics. First we treat probabilistic strong bisimulation. Here we present the main results to deal with probabilities. The second semantics is probabilistic weak bisimulation. The key point is how to define a logic that characterizes the relation in such a way that the modal operators can be reused for characterizing other semantics that abstract from internal behavior. For this we follow the framework defined by van Glabbeek in [21]. Additionally, we have to introduce a number of properties satisfied by our definitions of combined and hyper-transitions (Lemma 19). We also recall why probabilistic weak bisimulation may be considered too tight. Next we cover probabilistic branching bisimulation. Defining a logic and proving that it characterizes the process equivalence is straightforward given the definitions and results gathered already. We also provide a stuttering lemma for the probabilistic context. Finally, we discuss weak distribution bisimulation [9] and its logical characterization.

3.1 Probabilistic strong bisimulation

Generally, approaches to define behavioral equivalences for probabilistic transition systems provide a relation over states and a lifting to distributions over states, by means of, for example, *weight functions* [20] or *closed sets* [3]. We follow a different approach and will directly define relations over distributions that satisfy the *decomposability* condition of [10].

▶ **Definition 7.** A symmetric relation $\mathcal{R} \subseteq \mathit{Disc}(S) \times \mathit{Disc}(S)$ is *decomposable* if $\mu \,\mathcal{R}\, \nu$ and $\mu = \mu_1 \oplus_p \mu_2$ imply there are $\nu_1, \nu_2 \in \mathit{Disc}(S)$ s.t. $\nu = \nu_1 \oplus_p \nu_2$, $\mu_1 \,\mathcal{R}\, \nu_1$ and $\mu_2 \,\mathcal{R}\, \nu_2$.

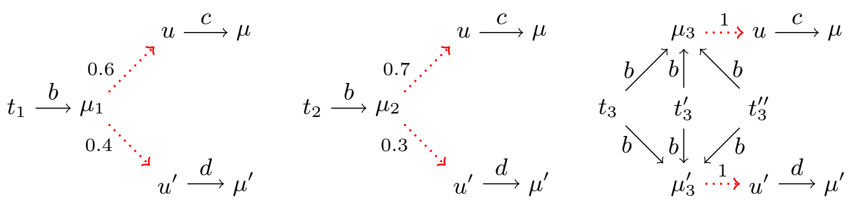

Figure 2 The decomposability condition is needed – Only considering deterministic schedulers is too strong.

In Definition 8, we introduce the notion of a *probabilistic strong bisimulation* for transition relations over probability distributions. Then we will explain why the decomposability condition is needed. We also show that a variant of strong bisimulation for transitions over distributions that only considers deterministic schedulers is too strong. For this reason we will not consider one-step transitions for deterministic schedulers further.

▶ **Definition 8.** A decomposable relation $\mathcal{R} \subseteq Disc(S) \times Disc(S)$ is called a *probabilistic strong bisimulation* if, for every $a \in \Sigma_\tau$, $\mu \,\mathcal{R}\, \nu$ and $\mu \xrightarrow{a}_c \mu'$ imply $\nu \xrightarrow{a}_c \nu'$ and $\mu' \,\mathcal{R}\, \nu'$ for some ν'. *Probabilistic strong bisimilarity*, notation $\approx_{ps}$, is defined as the union of all probabilistic strong bisimulations.

Figure 2 illustrates the need for the decomposability condition. Suppose we remove the condition, therefore μ_1 and μ_2 should be consider equivalent since they execute no transition, see Def. 5; therefore $\delta(t_1)$ and $\delta(t_2)$ are also equivalent. The condition ensures that distributions that are related by $\approx_{ps}$ assign the same weight to equivalence classes. Note that $\mu_1 = 0.6\delta(u) + 0.4\delta(u')$ and there are no μ_2' and μ_2'' such that $\mu_2 = 0.6\mu_2' + 0.4\mu_2''$, $\delta(u) \approx_{ps} \mu_2'$ and $\delta(u') \approx_{ps} \mu_2''$. Thus $\mu_1 \not\approx_{ps} \mu_2$ and therefore $\delta(t_1) \not\approx_{ps} \delta(t_2)$.

We explain the problem with deterministic schedulers. Consider state t_3, t_3' and t_3'' in Figure 2. It is clear that $\delta(t_3) \approx_{ps} \delta(t_3')$, $\delta(t_3') \approx_{ps} \delta(t_3'')$ and $\delta(t_3) \approx_{ps} \delta(t_3'')$. If we would consider only deterministic schedulers to define one-step transitions we have the transition $0.5\delta(t_3') + 0.5\delta(t_3'') \xrightarrow{b} 0.5\mu_3 + 0.5\mu_3'$ and this transition cannot be mimicked by $\delta(t_3)$ in the restricted setting. However, since we consider arbitrary schedulers, $\delta(t_3) \xrightarrow{b}_c 0.5\mu_3 + 0.5\mu_3'$.

Definition 9 introduces a logic that characterizes $\approx_{ps}$ (Theorem 15).

▶ **Definition 9.** The logic $\mathcal{L}_{ps}$ is defined by

$$\psi := \top \mid \bigwedge_{i \in I} \psi_i \mid \neg\psi \mid a\psi \mid \tau\psi \mid [\psi]_{\geq q}$$

for $a \in \Sigma$, $q \in \mathbb{Q}$ and possibly infinite index sets I. The *satisfiability of an $\mathcal{L}_{ps}$-formula for $\mu \in Disc(S)$* is defined by the following clauses:

$(\top)$	$\mu \models \top$	for all μ	(a)	$\mu \models a\psi$	if $\mu \xrightarrow{a}_c \mu'$ and $\mu' \models \psi$
$(\wedge)$	$\mu \models \bigwedge_{i \in I} \psi_i$	if $\mu \models \psi_i$ for all $i \in I$	(τ)	$\mu \models \tau\psi$	if $\mu \xrightarrow{\tau}_c \mu'$ and $\mu' \models \psi$
$(\neg)$	$\mu \models \neg\psi$	if $\mu \not\models \psi$	$(\geq q)$	$\mu \models [\psi]_{\geq q}$	if $\mu(\{s \in S \mid \delta(s) \models \psi\}) \geqslant q$

We denote by $\models_{ps}$ the satisfiability relation of $\mathcal{L}_{ps}$.

Recall that a one-step transition with an action a different from τ requires that all states in the support of the distribution execute the action a. This is not the case for a τ action. See Definition 5.

In [10], Hennessy introduces the logic pHML that also characterizes probabilistic strong bisimulation. The difference with this logic is the modality used to measure probabilities. The logic pHML uses a modal operator $\oplus_q$ and its validity, in contrast to $[\psi]_{\geq q}$, does not depend on the support of the distribution. Instead, it depends on how the distribution can be decomposed. The quantitative modal operators of [7, 18] are defined similarly. We do not follow this approach because it does not fit well for weak semantics. We explain this in more detail in the next subsection.

Parma and Segala [17] have introduced a distribution-based logic to characterize state-based probabilistic strong bisimulation, the logic $\mathcal{L}_p^N$. In this setting, states s and t are bisimilar iff $\delta(s)$ and $\delta(t)$ satisfy the same set of formulas. If we compare $\models_{ps}$ with $\mathcal{L}_p^N$, we see that clause (a) of Definition 9 is different: they have $\mu \models a\psi$ if for all $s \in \mathit{spt}(\mu)$, $s \xrightarrow{a}_c \mu_s$ and $\mu_s \models \psi$. In this case, the modal operator refers to the transitions that can be executed by the states underlying the distribution instead of the transitions of the distribution itself.

Theorem 15 states that $\mathcal{L}_{ps}$ characterizes $\approx_{ps}$. To prove the theorem we need a number of auxiliary results.

▶ **Lemma 10.** *Let $\mathcal{R}$ be a decomposable relation. For all μ, ν with $\mu \,\mathcal{R}\, \nu$, there is a finite index set K such that*
1. $\mu = \sum_{k \in K} p_k \cdot \delta(s_k)$.
2. $\nu = \sum_{k \in K} p_k \cdot \delta(t_k)$.
3. $\delta(s_k) \,\mathcal{R}\, \delta(t_k)$ *for all* $k \in K$.

In the following three lemmas $\mathcal{L}$ indicates a sublogic (of the logic at hand, here $\mathcal{L}_{ps}$).

▶ **Lemma 11.** *Let $\hat{s} \in S$ and $\psi \in \mathcal{L}$. If $\delta(\hat{s}) \models \psi$ then it holds that $\delta(\hat{s})(\{\, s \mid \delta(s) \models \psi \,\}) = 1$.*

Note that the last result does not work for arbitrary μ and ψ such that $\mu \models_s \psi$. Referring to Figure 1, $0.5\delta(s_2) + 0.5\delta(s_3) \models_s a[b]_{\geq 0.75}$. However, $\delta(s_3) \not\models_s a[b]_{\geq 0.75}$, and therefore it holds that $0.5\delta(s_2) + 0.5\delta(s_3)(\{\, s \mid \delta(s) \models \psi \,\}) < 1$.

▶ **Lemma 12.** *Let $\mathcal{L}$ be a logic containing $\neg$ and $\bigwedge$. Then there is a formula ψ_C for each $C \in \{\pi \mid \pi \in \mathit{Disc}(S)$ is a Dirac distribution$\}/{\approx_{\mathcal{L}}}$ s.t. for all $s \in S$, $\delta(s) \models \psi_C$ iff $\delta(s) \in C$.*

▶ **Lemma 13.** *Let $\mathcal{L}$ be a logic containing $\bigwedge$, $\neg$, and $[\cdot]_{\geq q}$ for $q \in \mathbb{Q}$. For all $\mu, \nu \in \mathit{Disc}(S)$ with $\mu \approx_{\mathcal{L}} \nu$ there is a finite index set K such that*
1. $\mu = \sum_{k \in K} p_k \cdot \delta(s_k)$.
2. $\nu = \sum_{k \in K} p_k \cdot \delta(t_k)$.
3. $\delta(s_k) \approx_{\mathcal{L}} \delta(t_k)$ *for all* $k \in K$.

▶ **Lemma 14.** *Suppose $\sum_{k \in K} p_k = 1$ for some index finite set K. If $\mu_k \approx_{\mathcal{L}_{ps}} \nu_k$ for $k \in K$, then $\sum_{k \in K} p_k \mu_k \approx_{\mathcal{L}_{ps}} \sum_{k \in K} p_k \nu_k$.*

▶ **Theorem 15.** *Let $\mu, \nu \in \mathit{Disc}(S)$, then $\mu \approx_{ps} \nu$ iff $\mu \approx_{\mathcal{L}_{ps}} \nu$.*

Sketch. To prove $(\Rightarrow)$ we show that $\mu \approx_{ps} \nu$ and $\mu \models_{ps} \psi$ implies $\nu \models_{ps} \psi$. This goes by structural induction on ψ. Cases $\top, \bigwedge_{i \in I} \psi_i$, $\neg\psi$ and $a\psi$ follow [21]. The case $[\psi]_{\geq q}$ follows by Lemmas 10 and 11. To prove $(\Leftarrow)$ we show that $\approx_{\mathcal{L}_{ps}}$ is a probabilistic strong bisimulation. The check of the transfer property follows [21]. To prove that $\approx_{\mathcal{L}_{ps}}$ is decomposable we use Lemmas 13 and 14. ◀

3.2 Probabilistic weak bisimulation

Transition relations over distributions allow to introduce straightforwardly a notion of weak bisimulation. Moreover, the discussion of the previous subsection applies here as well. Then, after the definition of weak bisimulation, we can focus on defining a corresponding logic.

▶ **Definition 16.** A decomposable relation $\mathcal{R} \subseteq Disc(S) \times Disc(S)$ is a *probabilistic weak bisimulation* if given $\mu \, \mathcal{R} \, \nu$, for every $a \in \Sigma_\tau$, $\mu \xrightarrow{a}_c \mu'$ implies there is ν' such that $\nu \overset{\hat{a}}{\Longrightarrow}_c \nu'$ and $\mu' \, \mathcal{R} \, \nu'$. *Probabilistic weak bisimilarity*, notation $\approx_{pw}$, is defined as the union of all probabilistic weak bisimulations.

The logic that characterizes $\approx_{pw}$ uses many of the operators of the logic for $\approx_{ps}$, but also adds new features to deal with internal behavior. The logic $\mathcal{L}_{pw}$ will be defined using a new modality ε, and two new clauses, $(\vec{\varepsilon})$ and $(\overleftarrow{\varepsilon})$ in the satisfiability relation. Because internal transitions cannot be observed, the clause (τ) is removed. First, we introduce the syntax and explain the intuition of the modality ε.

▶ **Definition 17.** The *logic* $\mathcal{L}_{pw}$ is defined by

$$\psi := \top \mid \bigwedge_{i \in I} \psi_i \mid \neg\psi \mid a\varepsilon\psi \mid \varepsilon\psi \mid [\psi]_{\geq q}$$

for $a \in \Sigma$, $q \in \mathbb{Q}$ and possibly infinite index sets I.

The modality ε is introduced to encode that internal behavior (zero or more τ's) can happen. In addition, we shall assume that some internal behavior can happen before the execution of any action. For example, $\varepsilon[c\varepsilon\top]_{\geq 0.5}$ encodes that after some internal behavior, with probability at least 0.5, an action c can be executed and the observation terminates; because the assumption and the modality ε, before and after the execution of c, some internal behavior can happen. This behavior is present in state s_1 in Figure 1. Also in Figure 1, notice that states s_2 and s_3 satisfy $a\varepsilon\top$. Because s_1 can reach both states with probability 1 via internal behavior, then $\delta(s_1)$ should also satisfy $a\varepsilon\top$. These ideas are modeled by clauses $(\vec{\varepsilon})$ and $(\overleftarrow{\varepsilon})$, which follow a phrasing of [21], in the following definition.

▶ **Definition 18.** The *satisfiability* of an $\mathcal{L}_{pw}$-formula is defined by the clauses $(\top)$, $(\bigwedge)$, $(\neg)$, $[\cdot]_{\geq q}$, (a) together with the following two:

$(\vec{\varepsilon})$ $\mu \models \varepsilon\psi$ if $\mu \models \psi$. $(\overleftarrow{\varepsilon})$ $\mu \models \psi$ if $\mu \overset{\varepsilon}{\Longrightarrow}_c \mu'$, $\mu' \models \psi$ and the outermost operator of ψ is neither $\neg$, $\bigwedge$ nor $[\cdot]_{\geq q}$.

We write $\models_{pw}$ to denote the satisfiability relation of $\mathcal{L}_{pw}$.

Figures 3 and 4, corresponding to the example in Figure 1, illustrates these clauses. Moreover, Figure 4 shows their interaction. Notice that it is possible to infer that some internal behavior can or cannot happen in a state. For instance, $\delta(s_1) \models \varepsilon[c\varepsilon\top]_{\geq 0.5}$, but $\delta(s_1) \not\models [c\varepsilon\top]_{\geq 0.5}$. The two formulas confirm that an internal transition for s_1 will change the equivalence class of the process for s_1 of Figure 1.

The condition "*the outermost operators of* ψ *is not* $\neg$, $\bigwedge$ *nor* $[\cdot]_{\geq q}$" of Definition 18 is needed for $(\overleftarrow{\varepsilon})$ because operators $\neg$ and $[\cdot]_{\geq q}$ give information about the current distribution. We use an example to explain this. See Figure 1: distribution μ_1 is such that $\mu_1 \models_{pw} \neg b\varepsilon\top$. If the restriction is not present, given that $\delta(s_1) \overset{\varepsilon}{\Longrightarrow} \mu_1$, one has $\delta(s_1) \models_{pw} \neg b\varepsilon\top$. This is inconsistent with the fact that $s_1 \xrightarrow{b}$. Similar reasoning is in place for the modality $[\cdot]_{\geq q}$. Operator $\bigwedge$ is also restricted to take into account the recursive case, for example, $\neg b\varepsilon\top \wedge \top$.

1. $\delta(s_2) \models_{pw} a\varepsilon\top$ and $\delta(s_3) \models_{pw} a\varepsilon\top$
2. $0.5\delta(s_2) + 0.5\delta(s_3) \models_{pw} a\varepsilon\top$ by (a)
3. $\delta(s_1) \overset{\varepsilon}{\Longrightarrow}_c 0.5\delta(s_2) + 0.5\delta(s_3)$
4. $\delta(s_1) \models_{pw} a\varepsilon\top$ by $(\overleftarrow{\varepsilon})$ and 3.

Figure 3 $(\overleftarrow{\varepsilon})$ allows to take into account the formula that are satisfied after an internal hyper-transition with a probabilistic scheduler.

(i) $\delta(s_2) \models_{pw} c\varepsilon\top$.
(ii) $0.5\delta(s_2) + 0.5\delta(s_3) \models_{pw} [c\varepsilon\top]_{\geq 0.5}$ by $(\geq q)$
(iii) $0.5\delta(s_2) + 0.5\delta(s_3) \models_{pw} \varepsilon[c\varepsilon\top]_{\geq 0.5}$ by $(\overrightarrow{\varepsilon})$
(iv) $\delta(s_1) \models_{pw} \varepsilon[c\varepsilon\top]_{\geq 0.5}$ by $(\overleftarrow{\varepsilon})$, 3. and 3

Figure 4 $(\overrightarrow{\varepsilon})$ concerns the current probability measure. Then $(\overleftarrow{\varepsilon})$ can be used for backward propagation. Notice $\delta(s_1) \not\models_{pw} [c\varepsilon\top]_{\geq 0.5}$.

In addition, in the following sections, we will use the same definition to define other logics that characterize other weak semantics.

In [17], Parma and Segala also introduce a logic that characterizes probabilistic weak bisimulation, the logic $\mathcal{L}_w^N$. This logic is the logic $\mathcal{L}_p^N$ where the modality a is replaced by a modality that considers weak combined transitions. Similarly, [7, 18] consider weak hyper-transitions. In comparison with other logics characterizing probabilistic weak bisimulation, $\mathcal{L}_{pw}$ looks more complex because of the modalities ε and the clauses $(\overleftarrow{\varepsilon})$ and $(\overrightarrow{\varepsilon})$. This complexity is needed to extend the logic to other semantics. See the next section.

It has been argued that probabilistic weak bisimulation is too strong [9, 5]: Consider Figure 5 and assume that states $\frownie$ and $\smiley$ are such that $\frownie \not\approx_{pw} \smiley$. State s_1 does not add any behavior to the system represented by the state s, and the probabilities of reaching states $\frownie$ and $\smiley$ from s are, respectively, 0.25 and 0.75. Then it is plausible to consider the distributions $\delta(s)$ and $\delta(t)$ weakly bisimilar, but in fact they are not. Notice there is no matching for the transition $\delta(s) \xrightarrow{\tau} \mu_s$. In [9], a variant of weak bisimulation is introduced to deal with this problem. We study the logic characterization of this variant in Subsection 3.4.

We explain why the approach of Hennessy [10] to define the measure modality does not fit well for weak bisimilarity. Let $\psi_{\frownie}$ and $\psi_{\smiley}$ be the characteristic formula of $\frownie$ and $\smiley$. Then

$$\delta(s_1) \models_{pw} \varepsilon([\psi_{\frownie}]_{\geq 0.5} \wedge [\psi_{\smiley}]_{\geq 0.5}) \qquad \delta(s) \models_{pw} \varepsilon([\varepsilon([\psi_{\frownie}]_{\geq 0.5} \wedge [\psi_{\smiley}]_{\geq 0.5}]_{\geq 0.5} \wedge [\psi_{\smiley}]_{\geq 0.5}))$$

Let $\hat{\psi}$ be the last formula, then $\delta(t) \not\models_{pw} \hat{\psi}$. In case we had used the approach used by Hennessy [10], we would replace the measure modality $[\,\cdot\,]_{\geq q}$ by $\oplus_q$ with clause

$(\oplus) \quad \mu \models \psi_1 \oplus_q \psi_2 \quad$ if $\mu = \mu_1 \oplus_q \mu_2$ and $\mu_i \models \psi_i$ for $i = 1, 2$

In the new setting, the formula analogous to $\hat{\psi}$ is $\varepsilon((\varepsilon(\psi_{\frownie} \oplus_{0.5} \psi_{\smiley})) \oplus_{0.5} \psi_{\smiley})$. This formula is satisfied by $\delta(t)$ because $\mu_t = (\frownie \oplus_{0.5} \smiley) \oplus_{0.5} \smiley$. Then $\delta(s)$ and $\delta(t)$ would not be distinguished by the logic in the new setting.

Theorem 21 states $\mathcal{L}_{pw}$ characterizes $\approx_{pw}$. To prove the result, we reuse the results for probabilistic strong bisimulation regarding probabilities. In addition, we need to introduce technical properties of weak transitions, see Lemma 19, and to recast Lemma 14 for $\mathcal{L}_{pw}$.

▶ **Lemma 19.** *Let $s \in S$, $\mu, \mu', \nu, \mu_i, \mu_i' \in Disc(S)$, where $i \in I$, and σ be a scheduler. Then*
1. *$\mu \overset{\varepsilon}{\Longrightarrow}_c \nu$ and $\nu \overset{\varepsilon}{\Longrightarrow}_c \mu'$ imply $\mu \overset{\varepsilon}{\Longrightarrow}_c \mu'$. [14, Prop. 3.6]*
2. *$s \overset{\hat{a}}{\Longrightarrow}_c \mu_i$, $i \in I$, $a \in \Sigma_\tau$ and $\sum_{i\in I} p_i = 1$ imply $s \overset{\hat{a}}{\Longrightarrow}_c \sum_{i \in I} p_i\mu_i$ [14, Prop. 3.4].*

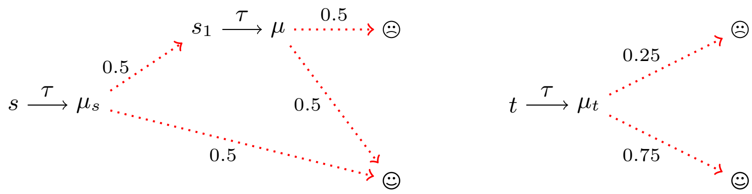

Figure 5 Probabilistic weak bisimulation is too strong.

3. $\mu_i \overset{\hat{a}}{\Longrightarrow}_c \mu'_i$, $i \in I$, $a \in \Sigma_\tau$ *and* $\sum_{i\in I} p_i = 1$ *imply* $\sum_{i\in I} p_i\mu_i \overset{\hat{a}}{\Longrightarrow} \sum_{i\in I} p_i\mu'_i$.
4. *For* $a \in \Sigma$, $\mu \overset{a}{\Longrightarrow}_c \mu'$ *iff there are* $\nu, \nu' \in Disc(S)$ *such that* $\mu \overset{\varepsilon}{\Longrightarrow}_c \nu \xrightarrow{a}_c \nu' \overset{\varepsilon}{\Longrightarrow}_c \mu'$.
5. $\sigma = \lim_{n\to\infty} \sigma_n$ *pointwise. Moreover,* $\mu_{s,\sigma}(\alpha) = \lim_{n\to\infty} \mu_{s,\sigma_n}(\alpha)$ *[20, Prop. 5.3.22].*
6. *If* $\ell gt(\sigma) > n$, $s \overset{\varepsilon}{\Longrightarrow}_{\sigma_n} \mu_n$ *and* $s \overset{\varepsilon}{\Longrightarrow}_{\sigma_{n+1}} \mu_{n+1}$ *then* $\mu_n \xrightarrow{\tau}_c \mu_{n+1}$.
7. *If* $\mu \approx_{pw} \nu$ *and* $\mu \overset{\varepsilon}{\Longrightarrow}_c \mu'$, *there is* $\nu' \in Disc(S)$ *such that* $\nu \overset{\varepsilon}{\Longrightarrow}_c \nu'$ *and* $\mu' \approx_{pw} \nu'$.
8. *For* $a \in \Sigma$, *if* $\mu \approx_{pw} \nu$ *and* $\mu \overset{a}{\Longrightarrow}_c \mu'$, *there is* $\nu' \in Disc(S)$ *such that* $\nu \overset{a}{\Longrightarrow}_c \nu'$ *and* $\mu' \approx_{pw} \nu'$.

▶ **Lemma 20.** *Suppose* $\sum_{k\in K} p_k = 1$, *for some finite index set* K. *If* $\mu_k \approx_{\mathcal{L}_{pw}} \nu_k$ *for* $k \in K$, *then* $\sum_{k\in K} p_k\mu_k \approx_{\mathcal{L}_{pw}} \sum_{k\in K} p_k\nu_k$.

▶ **Theorem 21.** *Let* $\mu, \nu \in Disc(S)$. *Then* $\mu \approx_{pw} \nu$ *iff* $\mu \approx_{\mathcal{L}_{pw}} \nu$.

The proofs of Lemma 20 and Theorem 21 strongly depend on the properties of Lemma 19. The proof of these statements are intricate, because of the definitions of combined and hyper-transitions considering schedulers of infinite length. These chedulers are needed, e.g., to distinguish between distributions $\delta(s_7)$ and $\delta(s_8)$ in Figure 1, i.e. $\delta(s_7) \approx_{pw} \delta(s_8)$. Note, state s_8 executes a transition with action b and this transition can be mimicked by s_7 only using a weak hyper-transition defined by a scheduler of infinite length. The variant of weak bisimulation of [7, 18] does not relate $\delta(s_7)$ and $\delta(s_8)$, because the relation $\overset{\varepsilon}{\Longrightarrow}_c$ is defined as the reflexive and transitive closure of $\xrightarrow{\hat{\tau}}_c$. This way of defining hyper-transition is sufficient in the context of [7, 18], because they deal with divergence-free PA. Notice, $\mathcal{A}$ in Figure 1 is not divergence-free.

3.3 Probabilistic branching bisimulation

We discuss probabilistic branching bisimilarity and its logical characterization. We remark that for the correspondence result we only need to add one new auxiliary result, a lemma analogous to Lemma 20.

▶ **Definition 22.** A decomposable relation $\mathcal{R} \subseteq Disc(S) \times Disc(S)$ is a *probabilistic branching bisimulation* if given $\mu \,\mathcal{R}\, \nu$, for every $a \in \Sigma_\tau$, $\mu \xrightarrow{a}_c \mu'$ implies

- $a = \tau$ and $\mu' \,\mathcal{R}\, \nu$, or
- there are $\tilde{\nu}$ and ν' such that $\nu \overset{\varepsilon}{\Longrightarrow}_c \tilde{\nu} \xrightarrow{a}_c \nu'$ with $\mu \,\mathcal{R}\, \tilde{\nu}$ and $\mu' \,\mathcal{R}\, \nu'$.

Probabilistic branching bisimilarity, notation $\approx_{pb}$, is defined as the union of all probabilistic branching bisimulations.

For the logic $\mathcal{L}_{pb}$ for probabilistic branching bisimulation we include binary operators $\_a\_$, for $a \in \Sigma$, and $\_\tau\_$ replacing $a\varepsilon$ and ε.

▶ **Definition 23.** The *logic* $\mathcal{L}_{pb}$ is defined by

$$\psi := \top \mid \bigwedge_{i\in I}\psi_i \mid \neg\psi \mid \psi a \psi' \mid \psi\tau\psi' \mid [\psi]_{\geq q}$$

for $a \in \Sigma$, $q \in \mathbb{Q}$ and possibly infinite index sets I. The satisfiability of an $\mathcal{L}_{pb}$-formula is defined by the clauses $(\top)$, $(\bigwedge)$, $(\neg)$, $(\geq q)$, $(\overleftarrow{\varepsilon})$ and

(η) $\quad \mu \models \psi a \psi'$ $\quad$ if $\mu \xrightarrow{a}_c \mu'$, $\mu \models \psi$ and $\mu' \models \psi'$.

(η_τ) $\quad \mu \models \psi\tau\psi'$ $\quad$ if, $\mu = \mu'$ or $\mu \xrightarrow{\tau}_c \mu'$, $\mu \models \psi$ and $\mu' \models \psi'$.

As in the non-probabilistic context [22, 21], the modality $\psi a \psi'$ (based on the notion of η-replication) allows to observe ψ' after an execution of action a that is preceded by the observation of ψ. This modality generalizes $a\psi$. A similar meaning for $\psi\tau\psi'$. These two modalities allow to check the branching of a process. Notice that $\_a\_$ does not force using ε, because of this, it is possible to check the branching after the execution of the action a.

▶ **Lemma 24.** *Suppose* $\sum_{k\in K} p_k = 1$ *for some finite index set* K.

1. *If* $\mu_k \approx_{pb} \nu_k$ *for* $k \in K$, *then* $\sum_{k\in K} p_k\mu_k \approx_{pb} \sum_{k\in K} p_k\nu_k$.
2. *If* $\mu_k \approx_{\mathcal{L}_{pb}} \nu_k$ *for* $k \in K$, *then* $\sum_{k\in K} p_k\mu_k \approx_{\mathcal{L}_{pb}} \sum_{k\in K} p_k\nu_k$.

▶ **Theorem 25.** *Let* $\mu, \nu \in \mathit{Disc}(S)$, $\mu \approx_{pb} \nu$ *iff* $\mu \approx_{\mathcal{L}_{pb}} \nu$.

3.4 Probabilistic weak bisimulation with sloppy probabilities

We have argued that weak bisimulation may be considered too strong. To deal with this problem, Eisentraut and co-workers introduced in [9] a notion called *weak distribution bisimulation*. We recall this process equivalence in Definition 27. Our presentation slightly differs from the original because we build on the notion of a weak decomposable relation (see Definition 26). In line with the nomenclature used in [21] for the global testing variants, we refer to our notion as *probabilistic weak bisimulation with sloppy probabilities*. We will motivate this further after the presentation of the logic that characterizes the semantics.

▶ **Definition 26.** A symmetric relation $\mathcal{R} \subseteq \mathit{Disc}(S) \times \mathit{Disc}(S)$ is called a *weak decomposable relation* if $\mu \,\mathcal{R}\, \nu$ and $\mu = \mu_1 \oplus_p \mu_2$ implies there are ν_1 and ν_2 such that $\nu \xRightarrow{\varepsilon}_c \nu_1 \oplus_p \nu_2$, $\mu_1\mathcal{R}\nu_1$ and $\mu_2\mathcal{R}\nu_2$.

Next we define for a weak decomposable relation when it is called a probabilistic weak bisimulation with sloppy probabilities.

▶ **Definition 27.** A weak decomposable relation $\mathcal{R} \subseteq \mathit{Disc}(S)\times\mathit{Disc}(S)$ is a called *probabilistic weak bisimulation with sloppy probabilities* if, for every $a \in \Sigma_\tau$, $\mu\mathcal{R}\nu$ and $\mu \xrightarrow{a}_c \mu'$ imply there is $\nu' \in \mathit{Disc}(S)$ s.t. $\nu \xRightarrow{\hat{a}}_c \nu'$ and $\mu' \,\mathcal{R}\, \nu'$. *Probabilistic weak bisimilarity with sloppy probabilities*, notation $\approx_{spw}$, is defined as the union of all the probabilistic weak bisimulations with sloppy probabilities.

The single difference of Definition 16 and Definition 27 is that the former uses a decomposable relation while the latter requires it to be weakly decomposable. However, this change is sufficient to capture $\delta(s) \approx_{spw} \delta(t)$ in Figure 5. Notice, $\mu_s \approx_{spw} \mu_t$ and these distributions are weakly decomposable because $\delta(s) \approx_{spw} \mu$.

In Figure 5 we have seen how the modal operator $[\cdot]_{\geq q}$ can be used to distinguish $\delta(s)$ and $\delta(t)$. We have argued that the approach of Hennessy does not differentiate between $\delta(s)$ and $\delta(t)$. However, to characterize the new semantics we only need to push his approach a little forward. The extra subtlety is this: Recall $\mu_s \approx_{spw} \mu_t$ and take into account the operator $\oplus_p$ with clause $(\oplus)$. Then $\mu_t \models \psi_{☹} \oplus_{0.25} \psi_{☺}$, but there are no μ_1 and μ_2 such that

$\mu_s = \mu_1 \oplus_{0.25} \mu_2$, $\mu_1 \models \psi_{☹}$ and $\mu_2 \models \psi_{☺}$. On the other hand, $\mu_s \stackrel{\varepsilon}{\Longrightarrow}_c 0.5\mu + 0.5\delta(☺)$ and $0.5\mu + 0.5\delta(☺) \models \psi_{☹} \oplus_{0.25} \psi_{☺}$. Therefore, in order to achieve $\mu_s \models \psi_{☹} \oplus_{0.25} \psi_{☺}$, we will allow a distribution to observe the measuring that is done after some internal behavior. This can be arranged 'for free', because the constraint that "*the outermost operator of ψ is not $\neg$, $\bigwedge$ nor $[\cdot]_{\geq q}$*" in the clause $(\overleftarrow{\varepsilon})$ does not concern the operator $\oplus_p$.

▶ **Definition 28.** The logic $\mathcal{L}_{spw}$ is defined by

$$\psi := \top \mid \bigwedge_{i\in I} \psi_i \mid \neg\psi \mid a\varepsilon\psi \mid \varepsilon\psi \mid \psi_1 \oplus_q \psi_2$$

for $a \in \Sigma$ and $q \in \mathbb{Q}$. The *satisfiability* of an $\mathcal{L}_{spw}$ formula is defined by clauses $(\top)$, $(\bigwedge)$, $(\neg)$, (a), $(\overleftarrow{\varepsilon})$, $(\overrightarrow{\varepsilon})$, and

$(\oplus)$ $\quad \mu \models \psi_1 \oplus_p \psi_2 \quad$ if $\mu = \mu_1 \oplus_p \mu_2$ and $\mu_i \models \psi_i$ for $i = 1, 2$.

We write $\models_{spw}$ to denote the satisfiability relation of $\mathcal{L}_{spw}$.

▶ **Theorem 29.** *For $\mu, \nu \in Disc(S)$, it holds that $\mu \approx_{spw} \nu$ iff $\mu \approx_{\mathcal{L}_{spw}} \nu$.*

4 Concluding remarks

In this paper we studied various behavioral equivalences for transitions systems over distributions in the presence of internal actions. An important contribution of our work is that we have consider weak hyper-transitions that deal with schedulers of infinite length. This allows to avoid the divergence-free condition for processes. Led by van Glabbeek's framework for the non-deterministic setting, we considered various ways to deal with τ-moves and provide logical characterizations for distribution-based probabilistic bisimulations. Moreover, we gave new characterization results following a uniform framework. The logics and axioms derive from the step-based behaviour encounter in the transfer conditions of the underlying bisimulation relation. The approach to prove correspondence results is the same for all notions of bisimulations considered. Crucial is the technical treatment of decomposable relations for weak combined and weak hyper-transitions (see Lemma 19). The uniform set-up allows to extend the results presented here to other semantics of the probabilistic branching-time spectrum without significantly more effort. Examples of this are η-bisimulation [1, 22] and delay bisimulation [22, 20].

▶ **Definition 30.** Let $\mathcal{R} \subseteq Disc(S) \times Disc(S)$ be a decomposable relation.

- $\mathcal{R}$ is a *probabilistic η-bisimulation* if given $\mu \,\mathcal{R}\, \nu$, for every $a \in \Sigma_\tau$, $\mu \xrightarrow{a} \mu'$ implies
  1. $a = \tau$ and $\mu' \,\mathcal{R}\, \nu$, or
  2. there are $\tilde{\nu}$, $\hat{\nu}$ and ν' such that $\nu \stackrel{\varepsilon}{\Longrightarrow}_c \tilde{\nu} \xrightarrow{a}_c \hat{\nu} \stackrel{\varepsilon}{\Longrightarrow}_c \nu'$ with $\mu \,\mathcal{R}\, \tilde{\nu}$ and $\mu' \,\mathcal{R}\, \nu'$.
- $\mathcal{R}$ is a *probabilistic delay bisimulation* if given $\mu \mathcal{R} \nu$, for every $a \in \Sigma_\tau$, $\mu \xrightarrow{a} \mu'$ implies
  1. $a = \tau$ and $\mu' \,\mathcal{R}\, \nu$, or
  2. there are $\tilde{\nu}$ and ν' such that $\nu \stackrel{\varepsilon}{\Longrightarrow}_c \tilde{\nu} \xrightarrow{a}_c \nu'$ with $\mu' \,\mathcal{R}\, \nu'$.

Probabilistic η-bisimilarity is defined as the union of all probabilistic η-bisimulations. *Probabilistic delay bisimilarity* is defined as the union of all probabilistic delay bisimulations.

We claim that the logics that characterize probabilistic η-bisimilarity and delay bisimilarity have the following syntax (using formulas ψ and φ for η and delay bisimulation, respectively).

$$\psi := \top \mid \textstyle\bigwedge_{i\in I} \psi_i \mid \neg\psi \mid \psi a\varepsilon\psi' \mid \psi\tau\psi' \mid [\psi]_{\geq q} \qquad \varphi := \top \mid \textstyle\bigwedge_{i\in I} \varphi_i \mid \neg\varphi \mid a\varphi \mid [\varphi]_{\geq q}$$

In the first logic, adding the modality ε in $\psi a\varepsilon\psi'$ does not allow anymore to check the branching of a process after the execution of a visible action. In the second logic, because the modalities $\psi a \psi'$ and $\psi \tau \psi'$ are removed, it is no longer possible to check the branching of a process before the execution of a visible action. We also remark that the problem presented in Figure 5 for probabilistic weak bisimulation is also present for these two new semantics. Then, one may also consider variants of the semantics with sloppy probabilities.

A decomposable relation $\mathcal{R} \subseteq \mathit{Disc}(S) \times \mathit{Disc}(S)$ straightforwardly induces a relation $\mathcal{S}$ over states just by putting $\mathcal{S} = \{ \ (s,t) \mid \delta(s) \, \mathcal{R} \, \delta(t) \ \}$. For distributions μ and ν with $\mu \mathcal{R} \nu$ we can define a weight function w for μ and ν with respect to $\mathcal{S}$ using the decompositions of μ and ν given by Lemma 10. Then the lifting of $\mathcal{S}$ to distributions agrees with $\mathcal{R}$. On the other hand, we expect that the approach based on *preserving* transitions used in [9] to give a state-based characterization of weak bisimulation with sloppy probabilities can be generalized to any weak decomposable relation. We have not studied this so far.

The probabilistic linear-time branching-time spectrum contains many more equivalences besides the ones discussed above. In [20] different types of combined transitions have been defined, each of which may potentially yield a new variant of a particular weak semantics. Alternatively, one can relax the condition over distributions, such as the variant of weak bisimulation of [19], or the variants of abstract probability bisimulation of [4]. The study of transition relations over distributions with internal actions can also be extended in other directions. We have considered image-finite relations, but this condition could be dropped, cf. [12]. Another interesting direction of future work considers relations over distributions with internal actions for uncountable state spaces, both regarding states and labels, as studied in [11]. Finally, it would be interesting to study how other probabilistic logics (like PCTL* [19] or variants of the probabilistic μ-calculus [15, 16]) behave in the distribution-based approach.

References

1 J.C.M. Baeten and R. J. van Glabbeek. Another look at abstraction in process algebra. In T. Ottmann, editor, *Proc. ICALP'87*, pages 84–94. LNCS 267, 1987.

2 S. Crafa and F. Ranzato. Logical characterizations of behavioral relations on transition systems of probability distributions. *TOCL*, 16(1):2:1–24, 2014.

3 P.R. D'Argenio and M.D. Lee. Probabilistic transition system specification: Congruence and full abstraction of bisimulation. In L. Birkedal, editor, *Proc. FOSSACS 2012*, 2012.

4 P.R. D'Argenio, M.D. Lee, and D. Gebler. SOS rule formats for convex and abstract probabilistic bisimulations. In S. Crafa et al., editor, *Proc. EXPRESS/SOS 2015*, 2015.

5 Y. Deng and M. Hennessy. On the semantics of Markov automata. *Information and Computation*, 222:139–168, 2013.

6 Y. Deng, R. J. van Glabbeek, M Hennessy, and C. Morgan. Characterising testing preorders for finite probabilistic processes. *Logical Methods in Computer Science*, 4(4), 2008.

7 Y. Deng and R.J. van Glabbeek. Characterising probabilistic processes logically. *CoRR*, abs/1007.5188, 2010.

8 J. Desharnais, V. Gupta, R. Jagadeesan, and P Panangaden. Weak bisimulation is sound and complete for PCTL*. *Information and Computation*, 208(2):203–219, 2010.

9 C. Eisentraut, H. Hermanns, J. Krämer, A. Turrini, and L. Zhang. Deciding bisimilarities on distributions. In Joshi, K.R. et al., editor, *Proc. QEST 2013*, pages 72–88, 2013.

10 M. Hennessy. Exploring probabilistic bisimulations, part I. *Formal Aspects of Computing*, 24(4-6):749–768, 2012.

11 H. Hermanns, J. Krcál, and J. Kretínský. Probabilistic bisimulation: Naturally on distributions. In P. Baldan and D. Gorla, editors, *Proc. CONCUR 2014*, pages 249–265, 2014.

12 H. Hermanns, A. Parma, R. Segala, B. Wachter, and L. Zhang. Probabilistic logical characterization. *Information and Computation*, 209(2):154–172, 2011.

13 K.G. Larsen and A. Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94(1):1–28, 1991.

14 N.A. Lynch, R. Segala, and F.W. Vaandrager. Observing branching structure through probabilistic contexts. *SIAM Journal of Computing*, 37(4):977–1013, 2007.

15 M. Mio. Upper-expectation bisimilarity and Łukasiewicz μ-calculus. In A. Muscholl, editor, *Proc. FoSSaCS*, pages 335–350, 2014. doi:10.1007/978-3-642-54830-7_22.

16 M. Mio and A.K. Simpson. Łukasiewicz μ-calculus. *CoRR*, abs/1510.00797, 2015.

17 A. Parma and R. Segala. Logical characterizations of bisimulations for discrete probabilistic systems. In H. Seidl, editor, *Proc. FOSSACS 2007*, pages 287–301. LNCS 4423, 2007.

18 J. Sack and L. Zhang. A general framework for probabilistic characterizing formulae. In *VMCAI 2012, Philadelphia, USA, January, 2012. Proceedings*, pages 396–411, 2012.

19 L. Song, L. Zhang, and J.C. Godskesen. Bisimulations meet PCTL equivalences for probabilistic automata. *Logical Methods in Computer Science*, 9(2), 2013.

20 M.I.A. Stoelinga. *Alea jacta est: verification of probabilistic, real-time and parametric systems.* PhD thesis, Radboud Universiteit Nijmegen, 2002.

21 R. J. van Glabbeek. The linear time–branching time spectrum II. In E. Best, editor, *Proc. CONCUR '93*, pages 66–81. LNCS 715, 1993.

22 R. J. van Glabbeek and W. P. Weijland. Branching time and abstraction in bisimulation semantics. *Journal of the ACM*, 43(3):555–600, 1996.