A general framework for probabilistic characterizing formulae

Joshua Sack[1] and Lijun Zhang[2]

[1] Department of Mathematics and Statistics, California State University Long Beach
[2] DTU Informatics, Technical University of Denmark

Abstract. Recently, a general framework on characteristic formulae was proposed by Aceto et al. It offers a simple theory that allows one to easily obtain characteristic formulae of many non-probabilistic behavioral relations. Our paper studies their techniques in a probabilistic setting. We provide a general method for determining characteristic formulae of behavioral relations for probabilistic automata using fixed-point probability logics. We consider such behavioral relations as simulations and bisimulations, probabilistic bisimulations, probabilistic weak simulations, and probabilistic forward simulations. This paper shows how their constructions and proofs can follow from a single common technique.

1 Introduction

Probabilistic automata have been extensively used in systems involving both stochastic and nondeterministic choice. To combat the state space explosion problem, various reduction techniques have been introduced and applied to probabilistic automata. These techniques include bisimulation and simulation relations [21, 20], partial order reductions [3, 12], symbolic data structures [13], and game-based abstractions [15].

Bisimulation and simulation relations are particularly useful, because they enable us to use compositional minimization [21]. Briefly, each of the constituting components can be minimized first before being composed with other interacting components. This idea is extended to a probabilistic setting in [6]. Various logics have been considered to reason about probabilistic automata. In [5], a model checking algorithm is presented for probabilistic automata with respect to the logic PCTL, and in [9, 14], Hennessy-Milner logics are used to characterize behavioral relations.

A characteristic formula for a behavioral relation is associated with each state in a model; the formula for a given state characterizes the set of states that the given state is related to according to the behavioral relation. In the case that the behavioral relation is simulation, one state is related to another if the first can be simulated by the other, that is, can be mimicked by the other. In effect, a characteristic formula allows us to reduce the problem of determining whether one state is simulated by another to the problem of model checking. Instead of directly checking whether the first state is simulated by the other, we check if the other state satisfies the characteristic formula of the first. In a more theoretical setting, some modal completeness and decidability theorems can be proved by constructing a finite satisfying model whose elements are normal forms, which are characteristic formulae for bisimulation or approximations to such formulae [18].

This paper focuses on behavioral relations over probabilistic automata and their characteristic formulae. The semantics of our languages involve fixed-points, which provide us with a natural facility for expressing various kinds of infinite behavior, such as those that are infinite or have loops. We present a single method, adapted from [1], that allows one to easily obtain characteristic formulae of many behavioral relations, including simulations and bisimulations, probabilistic bisimulations, probabilistic weak simulations, and probabilistic forward simulations. The strength of this technique is its generality: we can construct a variety of characteristic formulae and prove their correctness using a single simple method.

Relation to Related Work: Our theory builds on a recent paper by Aceto et al. [1], where a general framework is introduced for constructing non-probabilistic characteristic formulae over transition systems. It allows one to directly obtain the characteristic formulae for many behavioral relations, which have traditionally involved technical – even if not difficult – proofs. Their main result (an earlier version of Theorem 1 in this paper), in its generality, can be used for all the behavioral relations we consider, except for probabilistic forward simulation. We thus provide a modest generalization of this theorem to address forward simulation.

A more universally relevant extension to the overall setting of [1] is to involve in its applications (previously developed) liftings of relations. Liftings are discussed in [10, 23], and employed in [14] for fixed-point characterizations of (bi)simulations and probabilistic (bi)simulations. As they are central to probabilistic behavioral relations, liftings play a key role in adapting the framework of [1] to a probabilistic setting.

Another difference between our work and [1] is with the language used. The languages in [1] are fixed-point variants of Hennessy-Milner logic. For all our behavioral relations except the probabilistic forward simulation, we use a fixed-point variant of a two-sorted probability logic given in [16]. This allows us to interpret the characteristic formulae over states, as in [1], but to also have formulae over distributions that better fit with the setting of probabilistic automata. For probabilistic forward simulation, we involve a language, as in [19], only interpreted over distributions rather than states.

In [7], Deng and van Glabbeek study characteristic formulae for all the behavioral relations over probabilistic automata that we consider, though they restrict their automata to being finite. For all their behavioral relations, their characteristic formulae use a more complex one-sorted language over distributions than the one we use for probabilistic forward simulation, and the form of their formulae are different (reflecting their different but equivalent approach to lifting) and somewhat simpler (our characteristic formula for probabilistic bisimulation involve an infinitary disjunction). But the difference that we emphasize is that they use a separate technique for proving correctness of characteristic formulae for each preorder considered, while our framework provides characteristic formulae which are correct by construction.

Organization of the paper: In Section 2, we provide definitions to be used later in the paper. In Section 3, we present a slight adaptation of the framework developed in [1]. In Section 4, we recall the definition of probabilistic automata, the fixed-point characterization of bisimulation and simulation relations, and the weak bisimulations, and then we clarify the relationship between liftings used in [7] and in [14]. In Section

5, we present the language that we use for all our formulae except those from a language introduced in Section 6 that is defined specifically to characterize forward simulations. In Section 6, we illustrate how the characteristic formulae for all the behavioral relations that we consider can be constructed by applying the general framework. In Section 7, we describe some possible extensions of our work. Finally, Section 8 concludes the paper.

2 Preliminaries

Distributions. Let S be a set. A *distribution* over S is a function $\mu\colon S \to \mathbb{R}^{\geq 0}$ such that the *support* of μ, defined by $\mathrm{supp}(\mu) := \{s \mid \mu(s) > 0\}$, is countable, and $\sum_{s\in S}\mu(s) = 1$. We let $\mu(A)$ denote the sum $\sum_{s\in A}\mu(s)$, for all $A \subseteq S$. We denote by $\mathit{Dist}(S)$ the set of discrete probability distributions over S and, given an element $s \in S$, we denote by δ_s the *Dirac distribution* on s that assigns probability 1 to $\{s\}$.

Given a countable set of distributions $\{\mu_i\}_{i\in I}$ and a set $\{p_i\}_{i\in I}$ of real numbers in $[0, 1]$ such that $\sum_{i\in I} p_i = 1$, we define the *convex combination* $\sum_{i\in I} p_i\mu_i$ of $\{\mu_i\}_{i\in I}$ as the probability distribution μ such that, for each $s \in S$, $\mu(s) = \sum_{i\in I} p_i\mu_i(s)$.

Given a distribution over distributions ($\mu \in \mathit{Dist}(\mathit{Dist}(S))$), define the flattening of μ by the function $\mathit{flatten}$, that maps μ to a distribution ν, defined by

$$\nu(s) = \sum_{\nu'\in\mathrm{supp}(\mu)} \mu(\nu')\nu'(s). \tag{1}$$

Complete lattices. A *partially ordered set* (poset) is a set A together with a relation $\sqsubseteq_A$ that is reflexive ($a \sqsubseteq_A a$ for every $a \in A$), anti-symmetric ($a \sqsubseteq_A b$ and $b \sqsubseteq_A a$ implies $a = b$), and transitive ($a \sqsubseteq_A b$ and $b \sqsubseteq_A c$ implies $a \sqsubseteq_A c$). We omit the subscript A when it should be clear from context. A *complete lattice* is a partially ordered set $(A, \sqsubseteq)$, such that every subset $B \subseteq A$ has a least upper bound in A, written $\sqcup B$, and consequently a greatest lower bound $\sqcap B$ in A as well.

A function $f : A \to B$ between lattices is *monotone* if $a \sqsubseteq_A a'$ implies $f(a) \sqsubseteq_B f(a')$ for each $a, a' \in A$. A function $f : A \to B$ is an *isomorphism* if it is bijective, monotone, and f^{-1} is monotone, and consequently maps least upper bounds to least upper bounds. We call a function f from A to itself an *endofunction*. We call a point $a \in A$ a *post-fixpoint* of f if $f(a) \geq a$, and a *fixed-point* of f if $f(a) = a$. By Tarski's fixed-point theorem [22], every monotone endofunction f on a complete lattice A has a least upper bound $\mathsf{gfp}\, f$ given by $\sqcup\{a \mid a \sqsubseteq f(a)\}$.

3 General Framework

In this section we present some background behind our technique for finding characteristic formulae for behavioral relations. We involve languages $\mathcal{L}$ consisting of a set of formulae with variables. We often use I for the index set of the variables. The formulae will be interpreted over a set P, such as a set of states or distributions. In [1], $I = P$. We find that in order to apply this general framework to forward simulations (Section 6.4), it is helpful to distinguishing the index set I from the set P over which formulas will be interpreted, in particular setting I to be the set of states and P the set of distributions.

Variables are interpreted by a function $\sigma : I \to \mathcal{P}(P)$, called a *variable interpretation*. Here $\sigma(i)$ is viewed as the set of elements of P where the variable is considered to be true. This is similar to a valuation of atomic propositions in modal logic. The variable interpretation can be extended to a full fledged semantics $\sigma^* : \mathcal{L} \to \mathcal{P}(P)$, using rules such as $\sigma^*(\varphi \wedge \psi) = \sigma^*(\varphi) \cap \sigma^*(\psi)$. For all formulae φ and $p \in P$, we generally write $p \in [\![\varphi]\!]\sigma$ or $(\sigma, p) \models \varphi$ for $p \in \sigma^*(\varphi)$.

We call a function $E : I \to \mathcal{L}$ a *declaration*. Such a function characterizes an equational system of formulae, equating the variable X_i with the formula $E(i)$. As formulae can contain variables, a declaration is effectively recursive. Involving recursive features of a language allows us to characterize some infinite or looping behavior without the need for infinitary formulae. Lanugages with recursion generally involve fixed-points of an endofunction.[1] Hence we extend a declaration E to an endofunction $[\![E]\!] : \mathcal{P}(P)^I \to \mathcal{P}(P)^I$ on variable intepretations (here we write $\mathcal{P}(P)^I$ for the set of functions that map I to $\mathcal{P}(P)$), given by

$$([\![E]\!]\sigma)(i) = [\![E(i)]\!]\sigma.$$

The endofunction $[\![E]\!]$ has a greatest fixed point if the language is monotone.[2] A language is monotone if whenever $\sigma_1 \sqsubseteq \sigma_2$ (pointwise set inclusion), then for all formulae φ and elements p, it holds that $(\sigma_1, p) \models \varphi \Rightarrow (\sigma_2, p) \models \varphi$.

Behavioral relations, such as bisimulation, are often defined as the greatest fixed-point of a monotone endofunction F on $\mathcal{P}(I \times P)$, where I and P are typically set to be the set of states. The following definition clarifies our formulation of a characteristic formula, which in our setting is really a declaration.[3]

Definition 1 (Declaration characterizing a relation). *A declaration* $E : I \to \mathcal{L}$ characterizes *the greatest fixed-point of an endofunction* $F : \mathcal{P}(I \times P) \to \mathcal{P}(I \times P)$ *if for all* $i \in I$ *and* $p \in P$*,*

$$\mathsf{gfp}[\![E]\!], p \models E(i) \textit{ iff } (i, p) \in \mathsf{gfp}\, F \quad .$$

[1] Although we do not involve fixed-points operators directly in the language, we make use of fixed-points of a function induced by the declaration. A fixed-point sematics based on this equational system is equivalent to a fragment of the μ-calculi. The equational system provides us with a more intuitive way of handling what is equivalent to multiple nestings of fixed-point operators.

[2] This is because variable interpretations form a complete lattice, ordered under pointwise set inclusion, and the function $[\![E]\!]$ is monotone if the language is. Hence we can apply Tarski's fixed-point theorem.

[3] As we do not involve fixed-point operators directly in the formulae of the language, our recursive features come from the equational system given by the declaration. A *formula together with a declaration* contains the information we would normally obtain from a formula in the sufficiently expressive fragment of μ-calculus. Given a declaration E and $i \in I$, we always set the formula component to $E(i)$ when providing a characteristic formula-with-declaration for i. With this convension, the declaration is all we need to specify.

We link variable interpretations with subsets of $I \times P$, using the function $\varphi : \mathcal{P}(I \times P) \to \mathcal{P}(P)^I$ given by

$$\varphi(R) = (i \mapsto R(i)) \tag{2}$$

where $i \mapsto R(i)$ is the function mapping element $i \in I$ to the set $\{p \mid iRp\}$.

Definition 2 (Declaration expressing an endofunction). *A declaration E* expresses *a monotone endofunction $F : \mathcal{P}(I \times P) \to \mathcal{P}(I \times P)$ if*

$$\varphi(R), p \models E(i) \textit{ iff } (i, p) \in F(R)$$

for every relation $R \subseteq I \times P$.

More formally, the theorem from [1] is as follows.

Theorem 1. *If a declaration E expresses a monotone endofunction F, then E characterizes its greatest fixed-point* $\mathsf{gfp}\, F$.

This theorem and the prior two definitions differ from the one in [1] in that they set $I = P$. Our generalization of distinguishing I from P does not affect the proof in [1] of the main theorem.

4 Probabilistic automata, simulations, and bisimulations

We first discuss lifting of relations, followed by the definition of probabilistic automata and simulation relations.

4.1 Lifting of relations

A relation lifting transforms a relation between two sets into a relation between two sets related to the first two. Having two levels of relations is central to definitions of probabilistic behavioral relations. Liftings of relations from $S \times \mathit{Dist}(S)$ to $\mathit{Dist}(S) \times \mathit{Dist}(S)$ were introduced by Jonssen & Larsen [17] using *weight functions* to define simulations for Markov chains. Later, Desharnais [8] gave a definition of liftings that did not involve weight functions. We prove (Theorem 2 below) that these characterizations of liftings are equivalent, by using recent key insights (Lemma 1 below) from [10, 23].

First we present the following characterization [10, 23] of the lifting of a relation $R \subseteq S \times P$ (with S and P both arbitrary sets) to a relation $\widehat{R} \subseteq \mathit{Dist}(S) \times \mathit{Dist}(P)$:

$$\mu \widehat{R} \nu \Leftrightarrow \forall (A \subseteq \operatorname{supp} \mu).\ \mu(A) \leq \nu(R(A)). \tag{3}$$

When $P = \mathit{Dist}(S)$, we can define from $R \subseteq \mathit{Dist}(S) \times \mathit{Dist}(P)$ a relation $\overline{R} \subseteq \mathit{Dist}(S) \times \mathit{Dist}(S)$ by flattening the elements (see Eq. (1)) of $\mathit{Dist}(P)$: for $\mu, \nu \in \mathit{Dist}(S)$,

$$\mu \overline{R} \nu \Leftrightarrow \exists \nu' \in \mathit{Dist}(P).\ \nu = \mathit{flatten}(\nu')\ \&\ \mu R \nu'. \tag{4}$$

We next present the following characterization, based on weight functions, of a lifting from $R \subseteq S \times \mathit{Dist}(S)$ to $\widetilde{R} \subseteq \mathit{Dist}(S) \times \mathit{Dist}(S)$, given by

$$\begin{aligned} \mu \widetilde{R} \nu \Leftrightarrow \exists \{s_i\}_{i \in \mathbb{N}} \in S.\ \exists \{\nu_i\}_{i \in \mathbb{N}} \in \mathit{Dist}(S) \text{ such that} \\ \mu = \textstyle\sum_{i=1}^{\infty} p_i \delta_{s_i} \text{ and } \nu = \sum_{i=1}^{\infty} p_i \nu_i, \text{ for} \\ \text{some } p_i \geq 0, \textstyle\sum_{i=1}^{\infty} p_i = 1, \text{ and } s_i R \nu_i. \end{aligned} \qquad (5)$$

Note that $\delta_x \xrightarrow{a} \mu$ if and only if $x \overset{a}{\rightsquigarrow} \mu$. We will, as in [14], use the lifting (3) in our formulations of behavioral relations. The form (5) was used in [7] to define weak transitions, and will be used by us in the corresponding section (Definition 6). The two characterizations of relation liftings are equivalent in the following sense.

Theorem 2. *Given a relation $R \subseteq S \times \mathit{Dist}(S)$, $\widetilde{R} = \overline{\widehat{R}}$.*

Before proving this, we define weight functions [17] and networks, which will be useful in the proof.

Definition 3 (Weight function). *Let S and P be arbitrary sets. Let $\mu \in \mathit{Dist}(S), \nu \in \mathit{Dist}(P)$ and $R \subseteq S \times P$. A* weight function *for (μ, ν) with respect to R is a function $\Delta : S \times P \to [0, 1]$, such that*

1. *$\Delta(s, p) > 0$ implies $s\ R\ p$,*
2. *$\mu(s) = \sum_{p \in P} \Delta(s, p)$, for $s \in S$ and*
3. *$\nu(p) = \sum_{s \in S} \Delta(s, p)$, for $p \in P$.*

We only make sense of sums that have countably many non-zero terms. The conditions of Definition 3 ensure that $\Delta(s, p) = 0$ whenever either $s \notin \operatorname{supp} \mu$ or $p \notin \operatorname{supp} \nu$. Thus as an uncountable sum, only countably many terms would be non-zero, and hence it is safe to formulate this as an uncountable sum.

Definition 4 (The network for μ, ν and R). *Let $R \subseteq S \times P$, and let $\mu \in \mathit{Dist}(S), \nu \in \mathit{Dist}(P)$ be distributions. A* network *$\mathcal{N}(\mu, \nu, R)$ is a tuple (V, E, c), where*

1. *$V = \{\nearrow, \searrow\} \cup \operatorname{supp}(S) \cup \operatorname{supp}(P)$, with $\nearrow, \searrow \notin S, P$,*
2. *$E = \{(s, p) \mid (s, p) \in R\} \cup \{(\nearrow, s) \mid s \in \operatorname{supp}(\mu)\} \cup \{(p, \searrow) \mid p \in \operatorname{supp}(\nu)\}$,*
3. *c, the* capacity function, *is defined by:*
   (a) *$c(\nearrow, s) = \mu(s)$ for all $s \in \operatorname{supp} \mu$,*
   (b) *$c(p, \searrow) = \nu(p)$ for all $p \in \operatorname{supp}(\nu)$, and*
   (c) *$c(s, p) = \infty$ for all other $(s, p) \in E$.*

Lemma 1. *Let $R \subseteq S \times P$, and let $\mu_1 \in \mathit{Dist}(S), \mu_2 \in \mathit{Dist}(P)$. The following statements are equivalent:*

1. *There exists a weight function for (μ_1, μ_2) with respect to R.*
2. *The maximum flow of the network $\mathcal{N}(\mu_1, \mu_2, R)$ is 1.*
3. *$\mu_1(A) \leq \mu_2(R(A))$ for all $A \subseteq S$.*
4. *$\mu_1(A) \leq \mu_2(R(A))$ for all $A \subseteq \operatorname{supp}(\mu_1)$.*

The above lemma has been proposed in [10, 23], and used in [14]. The formal proof for countable systems makes use of a recent result in [2]. Thus, for completeness, the proof of this lemma for countable systems is given below.

Proof. The equivalence between 1 and 2 is from Lemma 5.1 in [4]. The equivalence between 3 and 4 is straight forward. We will show that 1 implies 3 and that 4 implies 2.

($1 \Longrightarrow 3$): Let Δ denote the corresponding weight function for (μ_1, μ_2) with respect to R. Now we want to prove that for every $A \subseteq S$: $\mu_1(A) \leq \mu_2(R(A))$. First, letting $Dom(R)$ represent the set of first coordinates of the relation R, we have

$$\mu_1(A) = \sum_{u \in A} \sum_{v \in P} \Delta(u, v) = \sum_{u \in A} \sum_{v \in R(A)} \Delta(u, v) = \sum_{u \in A \cap Dom(R)} \sum_{v \in R(A)} \Delta(u, v),$$

which follows from the properties of a weight function (Definition 3), especially that $\Delta(u, v) = 0$ if $u \notin Dom(R)$ or $v \notin R(u)$. Similarly, from the first and third conditions of a weight function, we have that $\mu_2(R(A)) = \sum_{u \in R^{-1}(R(A))} \sum_{v \in R(A)} \Delta(u, v)$. From basic set theory, we see that $A \cap Dom(R) \subseteq R^{-1}(R(A))$. Thus by comparing $\mu_1(A)$ and $\mu_2(R(A))$, we have our desired result: $\mu_1(A) \leq \mu_2(R(A))$.

($4 \Longrightarrow 2$): Assume that the fourth clause is true. We show that the maximum flow of the network $\mathcal{N}(\mu_1, \mu_2, R)$ has value 1. To construct such a maximum flow, we borrow the proof idea of Theorem 7.3.4 from Desharnais [8]. According to the *Maximum Flow Minimum Cut Theorem* [2], the maximum flow equals the capacity of a minimal cut. Therefore, it suffices to show that there exists a minimal cut of capacity 1. Cut $\{\upharpoonright\}$ has capacity 1, but we still have to show that it is minimal. Let C be some minimal cut (not necessarily $\{\upharpoonright\}$). We let $B = C \cap S$. The capacity of C is the sum: $c(C) = \sum\{c(i, j) \mid i \in C, j \notin C\}$. Cut C has to fulfill $s \in B \implies R(s) \subseteq C$ because otherwise it would have infinite capacity. Hence the capacity of C is: $c(C) = \mu_1(S \setminus B) + \mu_2(R(B))$. By construction of the network $\mathcal{N}$, it holds that $B \subseteq \mathrm{supp}(\mu_1)$. Since $\mu_1(B) \leq \mu_2(R(B))$, we have: $c(C) \geq \mu_1(S \setminus B) + \mu_1(B) = \mu_1(S) = 1$. Hence, the capacity of C is greater than or equal to 1, implying that the minimum cut has value 1. □

Proof. (Proof of Theorem 2)

Suppose that $\mu \widetilde{R} \nu$. Then $\mu = \sum_{i=1}^{\infty} p_i \delta_{s_i}$ and $\nu = \sum_{i=1}^{\infty} p_i \nu_i$, where $p_i \geq 0$, $\sum_{i=1}^{\infty} p_i = 1$, and $s_i R \nu_i$. Define $\nu' \in Dist(Dist(S))$, such that $\nu'(\nu_i) = p_i$. Then the p_i are the weights $\Delta(s_i, \nu_i)$ in the weight function for μ and ν' (Definition 3). By Lemma 1, $\mu(A) \leq \nu'(R(A))$, for all $A \in \mathrm{supp}(\mu)$. Thus $\mu \widehat{R} \nu'$, and hence $\mu \overline{\widehat{R}} \nu$.

Suppose that $\mu \overline{\widehat{R}} \nu$. Then there is a $\nu' \in Dist(Dist(S))$, such that $\nu = \mathit{flatten}(\nu')$ and $\mu \widehat{R} \nu'$, i.e., for all $A \in \mathrm{supp}\, \mu$, $\mu(A) \leq \nu'(R(A))$. By Lemma 1, there is a weight function Δ for μ and ν' with respect to R. Enumerate the pairs (s, ν), using a bijective function $f : (\mathrm{supp}(\mu) \times \mathrm{supp}(\nu')) \to \mathbb{N}$ (replace $\mathbb{N}$ with $\{1, 2, \ldots, N\}$ if $|\mathrm{supp}(\mu) \times \mathrm{supp}(\nu')| = N < \infty$). Let $g = f^{-1}$, $p_i = \Delta(g(i))$, $s_i = \pi_1(g(i))$ (where π_1 is the projection onto the first coordinate), and let $\nu_i = \pi_2(g(i))$. We then obtain the desired condition of (5) from the conditions of the weight function by plugging in an arbitrary s into the right hand side of the equation for μ in (5), and applying second condition of the weight function to see that we indeed get $\mu(s)$; and then note that the third condition of the weight function collapses the right hand side of the equation for ν in (5) into the right hand side of the equation for flattening of ν' into ν (recall that we used equation (4) to obtain ν'). □

4.2 Probabilistic automata

Now recall the definition of probabilistic automaton [21], or PA for short.

Definition 5. *A* probabilistic automaton *is a triple* $\mathcal{M} = (S, \mathit{Act}, \mathit{Steps})$*, where* S *is a countable set of* states*,* Act *is a countable set of* actions*, and the relation* $\mathit{Steps} \subseteq S \times \mathit{Act} \times \mathit{Dist}(S)$ *is the* transition relation.

Obviously, PAs comprise labeled transition systems (LTS) for the special case that for all $(s, a, \mu) \in \mathit{Steps}$, μ is a Dirac distribution. Denote a transition $(s, a, \mu) \in \mathit{Steps}$ by $s \overset{a}{\rightarrow} \mu$, which is also referred to as an a-transition of s. We denote the set of distributions leaving a state s by action a by $\mathit{Steps}_a(s) = \{\mu \mid s \overset{a}{\rightarrow} \mu\}$.

Given a probabilistic automaton $(S, \mathit{Act}, \mathit{Steps})$, we can augment the transition relation Steps (which maps states via actions to distributions) to another transition relation Comb (which also maps states via actions to distributions), such that each transition in Comb for any action corresponds to a convex combination of transitions in Steps for that action. Precisely, if $\{s \overset{a}{\rightarrow} \mu_i\}_{i \in I}$ is a set of transitions, then

$$s \overset{a}{\rightsquigarrow} \mu \text{ iff } \mu = \sum_{i \in I} p_i \mu_i \text{ for some } p_i \text{ where } \sum_{i \in I} p_i = 1. \qquad (6)$$

The a transitions in Step are denoted by $\overset{a}{\rightarrow}$ and those in Comb are denoted by $\overset{a}{\rightsquigarrow}$. Note that as $\overset{a}{\rightarrow}$ may represent a finite relation over states, $\overset{a}{\rightsquigarrow}$ typically represents a relation that is uncountable.

4.3 Simulations and bisimulations

In the following exposition, we fix some PA $\mathcal{M} = (S, \mathit{Act}, \mathit{Steps})$ and observe that the set of relations over S, denoted by $2^{S \times S}$, is a complete lattice with set inclusion as the partial order. We review in this section how some notions of simulation and bisimulation can be presented in terms of suitable monotone functions over this lattice [14].

Simulation. Consider the function $F_{\precsim} : 2^{S \times S} \rightarrow 2^{S \times S}$ defined as follows:

$$R \mapsto \{(s, t) \in S \times S \mid \forall s \overset{a}{\rightarrow} \mu .\ \exists t \overset{a}{\rightarrow} \mu' : \mu \widehat{R} \mu'\} \qquad (7)$$

We say that a relation $R \in 2^{S \times S}$ is a *simulation relation* if R is a post-fixpoint of $F_{\precsim}$, i.e. $R \subseteq F_{\precsim}(R)$. Note that the function $F_{\precsim}$ is monotone. Recall that Tarski's fixed-point theorem [22] says that the fixed-points of a monotone function form a complete lattice and that the greatest fixed-point is the union of all post-fixpoints. *Similarity*, denoted $\precsim$, is defined as the greatest fixed point of $F_{\precsim}$, and hence must be the union of all simulation relations, the greatest simulation relation.

Example 1. Let $\mathcal{M}$ be such that for every $(s, a, \mu) \in \mathit{Steps}$, μ is a Dirac distribution. Then

$$\begin{aligned}
F_{\precsim} : R \mapsto \; & \{(s,t) \in S \times S \mid \forall s \xrightarrow{a} \delta_{s'}.\, \exists t \xrightarrow{a} \delta_{t'} : \delta_{s'} \widehat{R} \delta_{t'}\} \\
= \; & \{(s,t) \in S \times S \mid \forall s \xrightarrow{a} \delta_{s'}.\, \exists t \xrightarrow{a} \delta_{t'} : \delta_{s'}(\{s'\}) \leq \delta_{t'}(R(\{s'\}))\} \\
= \; & \{(s,t) \in S \times S \mid \forall s \xrightarrow{a} \delta_{s'}.\, \exists t \xrightarrow{a} \delta_{t'} : s' R t'\}.
\end{aligned}$$

By replacing the Dirac distributions δ_s by states s in the definition of $F_{\precsim}$ over the LTS $\mathcal{M}$, we obtain the same definition that is given in [1] for an endofunction, whose post-fixpoints are simulations.

A coarser relation, called *probabilistic simulation*, is defined in the same way by replacing transitions with combined transitions so that the greatest probabilistic simulation is the greatest fixed-point of the function $F_{\precsim^p} : 2^{S \times S} \to 2^{S \times S}$ defined by:

$$R \;\mapsto \{(s, s') \in S \times S \mid \forall s \xrightarrow{a} \mu.\, \exists s' \overset{a}{\rightsquigarrow} \mu' :\; \mu \widehat{R} \mu'\} \;. \tag{8}$$

A relation $R \subseteq S \times S$ is a probabilistic simulation if it is a post-fixpoint of $F_{\precsim^p}$. The greatest probabilistic simulation preorder $\precsim^p$ is defined as the greatest fixed-point of $F_{\precsim^p}$.

Bisimulation. The function corresponding to bisimulation is a symmetric variation of the function for simulation, such that $F_{\sim} : 2^{S \times S} \to 2^{S \times S}$ is defined by:

$$R \mapsto \left\{ (s,t) \in S \times S \;\middle|\; \begin{array}{l} \forall s \xrightarrow{a} \mu.\, \exists t \xrightarrow{a} \mu' :\; \mu \widehat{R} \mu' \\ \forall t \xrightarrow{a} \mu'.\, \exists s \xrightarrow{a} \mu :\; \mu \widehat{R} \mu' \end{array} \right\}$$

We say that a relation $R \in 2^{S \times S}$ is a *bisimulation relation* if R is a post-fixpoint of $F_{\sim}$, i.e. $R \;\subseteq F_{\sim}(R)$. The greatest bisimulation $\sim$ is defined as the greatest fixed-point $\mathit{gfp}\ F_{\sim}$.

Similarly, we introduce *probabilistic bisimulation*. The function $F_{\sim^p} : 2^{S \times S} \to 2^{S \times S}$ for probabilistic bisimulation is defined analogously, however using combined transitions:

$$R \mapsto \left\{ (s,t) \in S \times S \;\middle|\; \begin{array}{l} \forall s \xrightarrow{a} \mu.\, \exists t \overset{a}{\rightsquigarrow} \mu' : \mu \widehat{R} \mu' \\ \forall t \xrightarrow{a} \mu'.\, \exists s \overset{a}{\rightsquigarrow} \mu : \mu \widehat{R} \mu' \end{array} \right\}$$

A relation $R \subseteq S \times S$ is a probabilistic bisimulation if it is a post-fixpoint of $F_{\sim^p}$. The greatest bisimulation $\sim^p$ is defined as the greatest fixed-point $\mathit{gfp}\ F_{\sim^p}$.

It is easy to see that $\sim$ and $\sim^p$ are equivalence relations. It is not difficult to see that, restricting to LTSs, (bi-)simulation and probabilistic (bi-)simulation coincide.

Weak simulation. We say that an automaton $(S, \mathit{Act}_\tau, \mathit{Steps})$ is divergent if there is an infinite sequence (s_i, μ_i), such that $s_i \xrightarrow{\tau} \mu_i$ and s_{i+1} is in the support of μ_i. An automaton that is not divergent is convergent.

Let *Act* be a non-empty set of actions, and let $\mathit{Act}_\tau = \mathit{Act} \cup \{\tau\}$, where τ is an element not appearing in *Act* and is regarded as an internal step. We define weak transitions similarly to those in [7, 20]:

Definition 6 (Weak transitions). *Given a convergent countable probabilistic automaton* $(S, \mathit{Act}_\tau, \mathit{Steps})$*, we define the following relations:*

- *define* $x \xrightarrow{\widehat{\tau}} \mu$ *iff* $x \xrightarrow{\tau} \mu$ *or* $\mu = \delta_x$*, and define* $x \xrightarrow{\widehat{a}} \mu$ *iff* $x \xrightarrow{a} \mu$*.*
- *define* $\overset{\widehat{\tau}}{\rightarrowtail}$ *and* $\overset{\hat{a}}{\rightarrowtail}$ *from respectively* $\xrightarrow{\widehat{\tau}}$ *and* $\xrightarrow{\hat{a}}$ *according to* (5)*.*
- *for all* $a \in \mathit{Act}_\tau$*, define* $\mu \overset{\hat{a}}{\Rightarrow} \nu$ *iff there are* μ' *and* ν'*, such that* $\mu \overset{\widehat{\tau}}{\rightarrowtail}^{*} \mu'$*,* $\mu' \overset{\hat{a}}{\rightarrowtail} \nu'$*,* $\nu' \overset{\widehat{\tau}}{\rightarrowtail}^{*} \nu$*, where* $\overset{\widehat{\tau}}{\rightarrowtail}^{*}$ *is the reflexive transitve closure of* $\overset{\widehat{\tau}}{\rightarrowtail}$*.*

A *weak simulation relation* is defined as a post-fixpoint of the endofunction $F_{\underset{\approx}{\precsim}} : 2^{S \times S} \to 2^{S \times S}$ defined by:

$$R \mapsto \{(s,t) \in S \times S \mid \forall a \in \mathit{Act}_\tau.\ \forall s \xrightarrow{a} \mu.\ \exists t.\ \delta_t \overset{\widehat{a}}{\Rightarrow} \mu' :\ \mu \widehat{R} \mu'\}\ .$$

Weak similarity, denoted $\underset{\approx}{\precsim}$, is defined is the greatest fixed-point of $F_{\underset{\approx}{\precsim}}$.

5 Hennessy-Milner logic for probabilistic automata

Here we present our basic language $\mathcal{L}_{\mathsf{bas}}$, a two-sorted language, similar to one in [16], consisting of state formulae (interpreted over the states S of the automaton) and distribution formulae (to be interpreted over $\mathit{Dist}(S)$). It is suggested in [14] that such a two-sorted language could be useful for a coalgebraic approach, but we leave coalgebraic characteristic formulae for future work.

Of the two sorts, we are ultimately interested in the formulae over states, as the simulation and bisimulation relations we have seen so far are defined over states. Formally, given a set *Act*$_\tau$ of actions augmented with a silent action τ, we define the language $\mathcal{L}_{\mathsf{bas}}(\mathit{Act}_\tau)$ by the following two-sorted syntax. State formulae are given by:

$$\varphi ::= X_z \mid \top \mid \bot \mid \bigwedge_{k \in K} \varphi_k \mid \bigvee_{k \in K} \varphi_k \mid \langle T \rangle \psi \mid [T]\psi$$

where $T \in \{\xrightarrow{a}, \overset{a}{\rightsquigarrow}, \overset{a}{\Rightarrow} \mid a \in \mathit{Act}_\tau\}$, $k \in K$ for some cardinal K, and $z \in I$ for some index set I, which we will typically set equal to the set S of states; distribution formulae are given by:

$$\psi ::= \top \mid \bot \mid \bigwedge_{k \in K} \psi_k \mid \bigvee_{k \in K} \psi_k \mid \mathsf{L}_p \varphi$$

where $p \in [0,1]$ and $k \in K$ for some cardinal K.[4]

Semantics. Let $\mathcal{M} = (S, \mathit{Act}, \mathit{Steps})$ be a PA. The formula φ is interpreted on states and ψ on distributions over. Both will make use of a variable interpretation $\sigma : I \to \mathcal{P}(P)$, where P is the set of states S. Select components of the semantics are given by:

[4] It may be desirable to restrict p to rational numbers so as to have a countable language, but doing so would require we add a countable conjunction to many of our characteristic formulae.

$\sigma, s \models X_z$	iff $s \in \sigma(z)$
$\sigma, s \models \langle T\rangle\psi$	iff $\sigma, \mu \models \psi$ for some μ such that $sT\mu$
$\sigma, s \models [T]\psi$	iff $\sigma, \mu \models \psi$ for all μ such that $sT\mu$
$\sigma, \mu \models \mathsf{L}_p\varphi$	iff $\mu(\{s \mid \sigma, s \models \varphi\}) \geq p$

where $T \in \{\xrightarrow{a}, \overset{a}{\rightsquigarrow}, \overset{a}{\Rightarrow} \mid a \in \mathit{Act}_\tau\}$. To be clear, we take $\xrightarrow{a}$ to be the primitive relation component in the probabilistic automaton, $\overset{a}{\rightsquigarrow}$ to be derived from $\xrightarrow{a}$ according to (6), and $\overset{a}{\Rightarrow}$ to be defined according to Definition 6.

We observe that this language is monotone:

Proposition 1. *if $\sigma_1 \sqsubseteq \sigma_2$ (pointwise set inclusion), then for all state formulae φ and states s, we have $\sigma_1, s \models \varphi \Rightarrow \sigma_2, s \models \varphi$ and for all distribution formulae ψ and distributions μ, we have $\sigma_1, \mu \models \psi \Rightarrow \sigma_2, \mu \models \psi$.*[5]

Proof. This is by induction on the structure of formulae:

IH suppose for every subformula ψ of φ, we have that whenever $\sigma_1 \sqsubseteq \sigma_2$, if ψ were a state formula, we have for each state s, $\sigma_1, s \models \varphi \Rightarrow \sigma_2, s \models \varphi$ and if ψ were a distribution formula, we have for each distribution $\sigma_1, \mu \models \psi \Rightarrow \sigma_2, \mu \models \psi$.

base case $\varphi = X_z$ immediate from definition.

Case booleans: these may be either state or distribution formulae, but the proof is straight forward.

Case $\varphi = \langle T\rangle\psi$, suppose that $\sigma_1, s \models \langle T\rangle\psi$. Then there is a μ such that $sT\mu$ and $\sigma_1, \mu \models \psi$. Then by the IH, $\sigma_2, \mu \models \psi$, and hence $\sigma_2, s \models \langle T\rangle\psi$.

Case $\varphi = [T]\psi$, this is almost identical to the $\langle T\rangle\psi$ case.

Case $\varphi = \mathsf{L}_p\psi$. Suppose that $\sigma_1, \mu \models \mathsf{L}_p\psi$. Then $\mu(\{s \mid \sigma_1, s \models \psi\}) \geq p$. But then by the IH, $\mu(\{s \mid \sigma_2, s \models \psi\}) \geq \mu(\{s \mid \sigma_1, s \models \psi\}) \geq p$. Thus $\sigma_2, \mu \models \mathsf{L}_p\psi$. □

6 Characteristic formulae

In this section, we illustrate how the characteristic formulae for all the behavioral relations that we consider can be constructed by using our adaptation of the general framework of [1].[6]

6.1 Simulations

We express in $\mathcal{L}_{\mathsf{bas}}$ the endofunction $F_{\precsim}$ with the endodeclaration

$$E_{\precsim} : s \mapsto \bigwedge_{a \in \mathit{Act}} \bigwedge_{\mu : s \xrightarrow{a} \mu} \langle \xrightarrow{a} \rangle \bigwedge_{A \subseteq \operatorname{supp}\mu} \mathsf{L}_{\mu(A)} \bigvee_{z \in A} X_z .$$

Recall that $[\![E_{\precsim}]\!]$ is an endofunction on variable interpretations, and is monotone since the language is. Had we restricted our language to only allowing rational subscripts p in L_p, then we could replace $\mathsf{L}_{\mu(A)}$ by $\bigwedge_{p \in \mathbb{Q} \cap [0, \mu(A)]} \mathsf{L}_p$.

We see that $E_{\precsim}$ expresses $F_{\precsim}$ as follows:

[5] Note that this formulation of a monotone language is slightly stronger than the definition of a monotone language given in Section 3.

[6] The general framework in [1] should apply to most of the behavioral relations as presented in that paper; our adaptation is only needed for forward simulation.

1. $(s,t) \in F_{\precsim}(R)$
2. $\forall a \in \mathit{Act},\ \forall s \xrightarrow{a} \mu,\ \exists t \xrightarrow{a} \mu',\ \mu\widehat{R}\mu'$
3. $\forall a \in \mathit{Act},\ \forall s \xrightarrow{a} \mu,\ \exists t \xrightarrow{a} \mu', \forall A \subseteq \mathrm{supp}(A), \mu(A) \leq \mu'(R(A))$.
4. $\varphi(R), t \models \bigwedge_{a\in \mathit{Act}} \bigwedge_{\mu : s \xrightarrow{a} \mu} \langle \xrightarrow{a} \rangle \bigwedge_{A \subseteq \mathrm{supp}\,\mu} \mathsf{L}_{\mu(A)} \bigvee_{z\in A} X_z$.
5. $\varphi(R), t \models E_{\precsim}(s)$

To see the relationship between Items (3) and (4), note that $\llbracket \bigvee_{z\in A} X_z \rrbracket \varphi(R) = R(A)$, and hence the formula $\mathsf{L}_{\mu(A)} \bigvee_{z\in A} X_z$ holds whenever $\mu(A) \leq \mu'(R(A))$.

Then by Theorem 1, $E_{\precsim}$ characterizes $\mathsf{gfp}\, F_{\precsim}$.

Opsim: Toward investigating the opposite of simulation (which we abbreviate *opsim* or *o*), we express the endofunction

$$F_{\precsim^o} : R \mapsto \{(s,t) \in S \times S \mid \forall a \in \mathit{Act}.\ \forall t \xrightarrow{a} \mu'.\ \exists s \xrightarrow{a} \mu : \mu\widehat{R}\mu'\}$$

with the endodeclaration

$$E_{\precsim^o} : s \mapsto \bigwedge_{a\in \mathit{Act}} [\xrightarrow{a}] \bigvee_{\mu : s \xrightarrow{a} \mu} \bigwedge_{A \subseteq \mathrm{supp}\,\mu} \mathsf{L}_{\mu(A)} \bigvee_{z\in A} X_z.$$

We see that $E_{\precsim^o}$ expresses $F_{\precsim^o}$ as follows:

1. $(s,t) \in F_{\precsim^o}(R)$
2. $\forall a \in \mathit{Act},\ \forall t \xrightarrow{a} \mu',\ \exists s \xrightarrow{a} \mu,\ \mu\widehat{R}\mu'$
3. $\forall a \in \mathit{Act},\ \forall t \xrightarrow{a} \mu',\ \exists s \xrightarrow{a} \mu, \forall A \subseteq \mathrm{supp}(A), \mu(A) \leq \mu'(R(A))$.
4. $\varphi(R), t \models \bigwedge_{a\in \mathit{Act}} [\xrightarrow{a}] \bigvee_{\mu : s \xrightarrow{a} \mu} \bigwedge_{A \subseteq \mathrm{supp}\,\mu} \mathsf{L}_{\mu(A)} \bigvee_{z\in A} X_z$.
5. $\varphi(R), t \models E_{\precsim^o}(s)$

Then by Theorem 1, $E_{\precsim^o}$ characterizes $\mathsf{gfp}\, F_{\precsim^o}$. Note that $E_{\precsim} \wedge E_{\precsim^o}$ is the characteristic formula for bisimulation $\sim$.

6.2 Probabilistic simulations and probabilistic bisimulation

Using the same argument as for simulation and opsimulation, we see that the endofunction

$$E_{\precsim^p} : s \mapsto \bigwedge_{a\in \mathit{Act}} \bigwedge_{\mu : s \xrightarrow{a} \mu} \langle \overset{a}{\rightsquigarrow} \rangle \bigwedge_{A \subseteq \mathrm{supp}\,\mu} \mathsf{L}_{\mu(A)} \bigvee_{z\in A} X_z.$$

expresses $F_{\precsim^p}$, and that the endofunction

$$F_{\precsim^{po}} : R \mapsto \{(s,t) \in S \times S \mid \forall a \in \mathit{Act}.\ \forall t \xrightarrow{a} \mu'.\ \exists s \overset{a}{\rightsquigarrow} \mu : \mu\widehat{R}\mu'\}$$

is expressed by the endodeclaration

$$E_{\precsim^{po}} : s \mapsto \bigwedge_{a\in \mathit{Act}} [\xrightarrow{a}] \bigvee_{\mu : s \overset{a}{\rightsquigarrow} \mu} \bigwedge_{A \subseteq \mathrm{supp}\,\mu} \mathsf{L}_{\mu(A)} \bigvee_{z\in A} X_z.$$

Hence $E_{\precsim^p}$ and $E^{\precsim^{po}}$ characterize $\mathsf{gfp}\, F_{\precsim^p}$ and $\mathsf{gfp}\, F_{\precsim^{po}}$ respectively. Note that $E_{\precsim^{po}}$ is typically infinitary, since the disjunction may be over an uncountable set. Similar to the case for ordinary bisimulation, $E_{\precsim} \wedge E_{\precsim^o}$ is the characteristic formula for probabilistic bisimulation $\sim^p$.

6.3 Weak simulations

A weak simulation is defined as the greatest fixed-point of the endofunction $F_{\underset{\approx}{\precsim}} : 2^{S\times S} \to 2^{S\times S}$ defined by

$$R \mapsto \{(s,t) \in S \times S \mid \forall a \in \mathit{Act}_\tau .\ \forall s \xrightarrow{a} \mu .\ \exists t.\ \delta_t \overset{\widehat{a}}{\Rightarrow} \mu' :\ \mu \widehat{R} \mu' \}\ .$$

Letting $s \overset{\hat{a}}{\rightrightarrows} \mu$ be defined by $\delta_s \overset{\hat{a}}{\Rightarrow} \mu$, we express this endofunction with the endodeclaration

$$E_{\underset{\approx}{\precsim}} : s \mapsto \bigwedge_{a\in \mathit{Act}_\tau} \bigwedge_{\mu : s \xrightarrow{a} \mu} \langle \overset{\hat{a}}{\rightrightarrows} \rangle \bigwedge_{A \subseteq \operatorname{supp} \mu} \mathsf{L}_{\mu(A)} \bigvee_{z\in A} X_z .$$

Note that this is the same as for simulation, but with $\xrightarrow{a}$ replaced by $\overset{a}{\rightrightarrows}$. The proof that $E_{\underset{\approx}{\precsim}}$ expresses $F_{\underset{\approx}{\precsim}}$ is essentially the same as the proof for simulation. Thus by Theorem 1, $E_{\underset{\approx}{\precsim}}$ characterizes $\mathsf{gfp}\, F_{\underset{\approx}{\precsim}}$.

6.4 Probabilistic forward simulation for probabilistic automata

Given a distribution $\mu \in \mathit{Dist}(S)$, we define $\breve{\mu} \in \mathit{Dist}(\mathit{Dist}(S))$ by

$$\breve{\mu}(\nu) = \begin{cases} \mu(s) & \nu = \delta_s \\ 0 & \text{otherwise} \end{cases} .$$

Note that $\mathit{flatten}(\breve{\mu}) = \mu$. In this section we consider the probabilistic forward simulation, defined by:

$$F_{\underset{\approx}{\precsim}^f} : R \mapsto \{(s,\mu) \in S \times \mathit{Dist}(S) \mid \forall a \in \mathit{Act}_\tau .\ \forall s \xrightarrow{a} \nu .\ \exists \mu' . \mu \overset{\hat{a}}{\Rightarrow} \mu' : \nu \widehat{R} \breve{\mu}' \}$$

Note also that $F_{\underset{\approx}{\precsim}^f}$ is monotone, as increasing the size of R will in turn increase the size of $\widehat{R}$, and hence $F_{\underset{\approx}{\precsim}^f}(R)$ will not shrink.

As before, we want to express the endofunction $F_{\underset{\approx}{\precsim}^f}$. We employ a “distribution” language $\mathcal{L}_{\mathsf{dst}}$, define as follows. Given a set *Act* of actions, the language $\mathcal{L}_{\mathsf{dst}}(\mathit{Act}_\tau)$ is given by:

$$\varphi ::= X_z \mid \top \mid \bot \mid \bigwedge_{k\in K} \varphi_k \mid \bigvee_{k\in K} \varphi_k \mid \langle \overset{\hat{a}}{\Rightarrow} \rangle \varphi \mid [\overset{\hat{a}}{\Rightarrow}] \varphi \mid \mathsf{L}_p \varphi$$

where $a \in \mathit{Act}_\tau$, $k \in K$ for some cardinal K, and $z \in I$ for some index set I, (which we will typically, or maybe always, make the set of distributions), $p \in [0,1]$.

We interpret all formulae φ on distributions, and will use a variable interpretation $\sigma : I \to \mathcal{P}(P)$, where $P = \mathit{Dist}(S)$. Select components of the semantics are:

$\sigma, \mu \models X_z$	iff $\mu \in \sigma(z)$
$\sigma, \mu \models \langle \overset{\hat{a}}{\Rightarrow} \rangle \psi$	iff $\sigma, \nu \models \psi$ for some ν where $\mu \overset{\hat{a}}{\Rightarrow} \nu$
$\sigma, \mu \models [\overset{\hat{a}}{\Rightarrow}] \psi$	iff $\sigma, \nu \models \psi$ for all ν where $\mu \overset{\hat{a}}{\Rightarrow} \nu$
$\sigma, \mu \models \mathsf{L}_p \varphi$	iff $\breve{\mu}(\{\nu \mid \sigma, \nu \models \varphi\}) \geq p$

Note that $\mathsf{L}_p\varphi$ is defined differently here as it was in $\mathcal{L}_{\mathsf{bas}}$: in $\mathcal{L}_{\mathsf{dst}}$, we take the probabilities to be over sets of distributions, while in $\mathcal{L}_{\mathsf{bas}}$ we take them to be over sets of states. Also, although the variables are indexed by states in both languages, their interpretations are also different. One can check that $\mathcal{L}_{\mathsf{dst}}$ is monotone.

Then the endofunction

$$E_{\precapprox^f} : s \mapsto \bigwedge_{a \in \mathit{Act}_\tau} \bigwedge_{\nu : s \xrightarrow{a} \nu} \langle \overset{a}{\Rightarrow} \rangle \bigwedge_{A \subseteq \operatorname{supp} \nu} \mathsf{L}_{\nu(A)} \bigvee_{z \in A} X_z .$$

expresses $F_{\precapprox^f}$, which can be seen as follows:

1. $(s, \mu) \in F_{\precapprox^f}(R)$
2. $\forall a \in \mathit{Act}_\tau,\ \forall s \xrightarrow{a} \nu,\ \exists \mu \overset{\hat{a}}{\Rightarrow} \mu', \nu \widehat{R} \breve{\mu}'$
3. $\forall a \in \mathit{Act}_\tau,\ \forall s \xrightarrow{a} \nu,\ \exists \mu \overset{\hat{a}}{\Rightarrow} \mu', \forall A \subseteq \operatorname{supp}(\nu), \nu(A) \leq \breve{\mu}'(R(A))$.
4. $\varphi(R), \mu \models \bigwedge_{a \in \mathit{Act}_\tau} \bigwedge_{\nu : s \xrightarrow{a} \nu} \langle \overset{\hat{a}}{\Rightarrow} \rangle \bigwedge_{A \subseteq \operatorname{supp} \nu} \mathsf{L}_{\nu(A)} \bigvee_{z \in A} X_z .$
5. $\varphi(R), \mu \models E_{\precapprox^f}(s)$

Thus by Theorem 1, $E_{\precapprox^f}$ characterizes $\mathsf{gfp}\, F_{\precapprox^f}$.

7 Extensions

For simplicity of presentation, we have chosen probabilistic automata, as they are one of the most important types of stochastic models studied in the literature. We want to note, however, that the general framework can be easily extended to other types of stochastic models.

Let us briefly discuss the model called continuous-time Markov chains (CTMC). In CTMCs, we do not have nondeterministic choices, whereas transitions are governed by a negative exponential distribution. Briefly, from each state s we have a unique transition of the form $s \xrightarrow{\lambda} \mu$, where λ is a positive constant characterizing the negative exponential distribution, and μ is the distribution (as in probabilistic automata). Then, starting from s, the probability of triggering the transition within time $t > 0$ is given by $1 - e^{-\lambda t}$, and once the transition is triggered, t is reached with probability $\mu(s')$.

As for probabilistic automata, the important preparation steps are to (i) provide a fixed-point based definition of bisimulation and simulation relations, and (ii) define appropriate logic and semantics, such as those in the Hennessy-Milner style. Indeed, both can be done for CTMCs in a straightforward way. The fixed-point based definition of simulation is based on the function: $R \mapsto \{(s,t) \mid E(s) \leq E(t) \wedge \mu \widehat{R} \mu'\}$where $E(s)$ is such that $s \xrightarrow{E(s)} \mu$ (which is unique as we mentioned), and similarly for $E(t)$. The only additional information is that the exit rate $E(t)$ from t is larger than that of s, meaning that t is *faster* than s. The logic is also simple because of the lack of nondeterministic choices: the only modal operator for state formulae is of the form $\langle \lambda \rangle \psi$, and the distribution formulae are the same as for PAs. The semantics for the modal operator is: s satisfies $\langle \lambda \rangle \psi$ if and only if $E(s) \geq \lambda$ and μ satisfies ψ with $s \xrightarrow{E(s)} \mu$ (as for probabilistic automata). In this way, characteristic formulae can be obtained for CTMCs, with respect to simulations, and also bisimulations. Moreover, further extensions to Markov automata [11], an orthogonal extension of CTMCs and PAs, can also be obtained along the same line.

8 Conclusion

This paper shows how the general theory in [1] for finding characteristic formulae can be adapted and applied to forward simulation and other behavioral relations in a setting for probabilistic automata. Although the characteristic formulae constructed using this method may differ from ones developed using other methods (such as those in [7]), it is helpful to see how a single method can be used to find characteristic formulae for these probabilistic behavioral relations in general, and that this technique can likely be used for far more probabilistic behavioral relations. Thus the main thrust of this paper is not in the results themselves, but in highlighting a method the research community should be aware of.

In [10], Desharnais et al. have considered a relaxation of (bi)-simulations in which the weight functions may differ by as much as ε. The case $\varepsilon = 0$ reduces to the traditional bisimulation relations considered in this paper, whereas the case $\varepsilon > 0$ is particularly useful for reasoning about systems that *nearly* match each other. Extending our results to such ε-bisimulations would be an interesting line of future work.

Acknowledgements

Joshua Sack was partly supported by the project “Processes and Modal Logic” (project nr. 100048021) of The Icelandic Research Fund. Lijun Zhang is partly supported by MT-LAB, a VKR Centre of Excellence.

We thank Luca Aceto, Holger Hermanns, Anna Ingolfsdottir and Björn Wachter for insightful discussions on an early version of the paper.

References

1. L. Aceto, A. Ingolfsdottir, P. Levy, J. Sack. Characteristic formulae for fixed-point semantics: a general approach. To appear in *Mathematical Structures in Computer Science*, 2010.
2. R. Aharonia, E. Bergerb, A. Georgakopoulosc, A. Perlsteina, and P. Sprüssel. The Max-Flow Min-Cut theorem for countable networks. *Journal of Combinatorial Theory, Series B*, 101(1):1–17, 2011.
3. C. Baier, P. R. D'Argenio, and M. Größer. Partial order reduction for probabilistic branching time. *Electr. Notes Theor. Comput. Sci.*, 153(2):97–116, 2006.
4. C. Baier, B. Engelen, M. E. Majster-Cederbaum. Deciding bisimilarity and similarity for probabilistic processes. *J. Comput. Syst. Sci.* 60(1)187–231, 2000.
5. A. Bianco and L. de Alfaro. Model Checking of Probabilistic and Nondeterministic Systems. In *FSTTCS, LNCS 1026:499-513*. Springer, 1995.
6. H. Boudali, P. Crouzen, and M. Stoelinga. A rigorous, compositional, and extensible framework for dynamic fault tree analysis. *IEEE Transactions on Dependable and Secure Computing*, 99(1), 2009.
7. Y. Deng and R. van Glabeek. Characterising Probabilistic Processes Logically. *Lecture Notes in Computer Science*, 6397:287–293, 2010.
8. J. Desharnais. *Labelled Markov processes*. Ph.D. thesis, McGill University, 1999.
9. J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Weak bisimulation is sound and complete for pctl*. *Inf. Comput.*, 208(2):203–219, 2010.

10. J. Desharnais, F. Laviolette, and M. Tracol. Approximate analysis of probabilistic processes: Logic, simulation and games. In *QEST*, pages 264–273, 2008.
11. C. Eisentraut, H. Hermanns, and L. Zhang. On probabilistic automata in continuous time. In *LICS*, pages 342–351, 2010.
12. S. Giro, P. R. D'Argenio, and L. M. F. Fioriti. Partial order reduction for probabilistic systems: A revision for distributed schedulers. In *CONCUR*, pages 338–353, 2009.
13. H. Hermanns, M. Z. Kwiatkowska, G. Norman, D. Parker, and M. Siegle. On the use of mtbdds for performability analysis and verification of stochastic systems. *J. Log. Algebr. Program.*, 56(1-2):23–67, 2003.
14. H. Hermanns, A. Parma, R. Segala, B. Wachter, and L. Zhang. Probabilistic logical characterization. *Inf. Comput.*, 209(2):154–172, 2011.
15. M. Z. Kwiatkowska, G. Norman, and D. Parker. Game-based abstraction for markov decision processes. In *QEST*, pages 157–166, 2006.
16. B. Jonsson, K. Larsen, and W. Yi. Probabilistic Extensions of Process Algebras. In the *Handbook of Process Algebra*. Editors Jan A. Bergstra, Alban Ponse, and Scott A. Smolka. pp. 685–710. Elsevier, 2001
17. B. Jonsson and K. Larsen. Specification and Refinement of Probabilistic Processes. *LICS*, pages 266–277, 1991.
18. L. Moss. Finite Models Constructed from Canonical Formulas. *Journal of Philosophical Logic*, 36:605–740, 2007.
19. A. Parma, R. Segala, Logical Characterizations of Bisimulations for Discrete Probabilistic Systems. In *FOSSACS*, pages 287-301, April 2007.
20. R. Segala. *Modeling and Verification of Randomized Distributed Realtime Systems*. PhD thesis, MIT, 1995.
21. R. Segala and N. A. Lynch. Probabilistic simulations for probabilistic processes. *Nord. J. Comput.*, 2(2):250–273, 1995.
22. A. Tarski. A Lattice-Theoretical Fixpoint Theorem and its Applications. *Pacific Journal of Mathematics 5, pp. 285–309.*, 1955.
23. L. Zhang. *Decision Algorithms for Probabilistic Simulations*. PhD thesis, Universität des Saarlandes, 2008.