A Testing Scenario for Probabilistic Processes

LING CHEUNG
Radboud University Nijmegen
and
MARIËLLE STOELINGA
University of Twente
and
FRITS VAANDRAGER
Radboud University Nijmegen

We introduce a notion of finite testing, based on statistical hypothesis tests, via a variant of the well-known trace machine. Under this scenario, two processes are deemed observationally equivalent if they cannot be distinguished by any finite test. We consider processes modeled as image finite probabilistic automata and prove that our notion of observational equivalence coincides with the trace distribution equivalence proposed by Segala. Along the way, we give an explicit characterization of the set of probabilistic generalize the *Approximation Induction Principle* by defining an also prove limit and convex closure properties of trace distributions in an appropriate metric space.

Categories and Subject Descriptors: F.1.1 [**Computation by abstract devices**]: Models of Computation—*Automata*; F.1.2 [**Computation by abstract devices**]: Modes of Computation—*Probabilistic Computation*; F.4.3 [**Mathematical logic and formal languages**]: Formal languages—*Classes defined by grammars or automata*; G.3 [**Probability and statistics**]: Probabilistic algorithms; stochastic processes

General Terms: Theory, Verification

Additional Key Words and Phrases: Concurrency theory, probabilistic automata, testing, button pushing scenario, observational equivalence, trace distributions, approximation induction principle, CPO, metric spaces

First author's address: L. Cheung, Massachusetts Institute of Technology, CSAIL, Cambridge, MA, USA, lcheung@theory.csail.mit.edu. Second author's address: M.I.A. Stoelinga, Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, P.O. Box 217, 7500 AE Enschede, The Netherlands, marielle@cs.utwente.nl. Third author's address: F.W. Vaandrager, Institute for Computing and Information Sciences, Radboud University Nijmegen, P.O. Box 9010, 6500 GL Nijmegen, The Netherlands, F.Vaandrager@cs.ru.nl.

This research was supported by project DN 62-600 of the DFG/NWO bilateral cooperation programme (VOSS2). The research of the second author has been supported in part by EU grant IST-004527 (ARTIST2) and by NWO under FOCUS/BRICKS grant 642.000.505 (MOQS). The research of the third author was supported by PROGRESS Project TES4199, Verification of Hard and Softly Timed Systems (HaaST).

A preliminary version of this paper appeared in the PhD thesis of the second author [Stoelinga 2002]. An extended abstract [Stoelinga and Vaandrager 2003] received the Best Paper Award at ICALP'03 (Track B).

1. INTRODUCTION

A fundamental idea in concurrency theory is that two processes are deemed equivalent if they cannot be distinguished by external observation. Varying the power of the external observer, different notions of behavioral equivalence arise. For processes modeled as labeled transition systems (LTSs), this idea has been thoroughly explored: a large number of behavioral equivalences have been characterized via intuitive *testing scenarios*, also called *button-pushing experiments* [Milner 1980].

In a typical button-pushing experiment, we envision a machine equipped with a display and a series of buttons. The process under observation resides within this machine and its activities, represented by action symbols, are shown on the display. An external observer may influence the execution of this process by pressing one or more buttons at various times. The simplest example of such an experiment is the *trace machine*, which has an action display but no buttons. It turns out to be sufficient for characterizing the well-known *trace equivalence* for LTSs.

Button-pushing experiments are desirable for a number of reasons. First, they provide a simple and intuitive way to understand behavioral equivalences that are defined more abstractly, e.g. via process algebras or in terms of satisfaction of logical formulas. Second, they provide a unified setting for comparing these behavioral equivalences. We refer to Van Glabbeek [Glabbeek 2001] for an excellent overview of results in this area of *comparative concurrency semantics*. Finally, in a button-pushing experiment, interactions between a process and an observer take place exclusively via the predefined interface, namely, display and buttons. This is in keeping with the tradition of modular reasoning, which requires that processes evolve independently from their environments, aside from explicit inputs.

The present paper proposes such a testing scenario for probabilistic processes. (For our purposes, a *probabilistic process* may make discrete random choices as well as nondeterministic choices.) This task calls for a nontrivial extension of existing testing scenarios for LTSs, because one must specify a means to "observe" probability distributions. For that end, we devise a *trace distribution machine* and use the theory of *null hypothesis testing* to provide a link between

—probability distributions derived in an abstract semantics and

—sample observations collected from the trace distribution machine.

The distinguishing feature of our trace distribution machine is a *reset* button, which restarts the machine from its initial state. This allows an observer to record traces from multiple runs of the machine. These runs are assumed to be independent; that is, random choices in one run are not correlated with those in another run. However, we do not assume that nondeterministic choices are resolved in exactly the same way, therefore each run is governed by a possibly different probability distribution.

The semantics of this reset button poses a challenge in designing our hypothesis tests. Even though we can compute frequencies of traces from a sample of m runs, it is not immediately clear what information we have obtained about the m possibly distinct probability distributions. As it turns out, this frequency statistic provides a very natural estimator for the *average* of the m distributions. Thus we reason about these m distribution collectively: a typical null hypothesis states that a sample consisting of m runs is generated by a particular sequence of m distributions.

Another challenging issue is infinite behaviors of the probabilistic processes. These may include infinite branching and non-terminating runs. In contrast, experiments on the trace distribution machine are of a finite character: an observer can record only finitely many symbols from a single run and can observe only finitely many runs. To overcome this discrepancy, we prove an Approximation Induction Principle, stating that every infinite probabilistic behavior can be approximated by its finite "sub-behaviors". In addition, we introduce an *extended trace distribution machine* for processes with an infinite action alphabet. This machine allows the observer to suppress all but a finite number of actions, so that the sample space of each experiment remains finite.

Our work is carried out in the framework of *probabilistic automata (PA)*, which augments the LTS model with discrete probability distributions [Segala 1995]. This framework has seen many applications in the analysis of distributed algorithms [Aggarwal 1994; Lynch et al. 1994; Pogosyants et al. 2000; Stoelinga and Vaandrager 1999]. In the present paper, we prove that the observational equivalence induced by our testing scenario coincides with the trace distribution equivalence of [Segala 1995]. Therefore, our testing scenario can be viewed as an intuitive justification of the more abstract notion of trace distribution equivalence.

We have chosen the PA framework in part for its simplicity, so that we are free from particular features that may hamper the portability of our results. Indeed, we focus on semantic objects induced by PAs, as opposed to the automata themselves. These objects are probability distributions on computation paths (here called *probabilistic executions*) and probability distributions on traces (here called *trace distributions*). They can be viewed very naturally as trees with probabilistic branching, so that our technical developments quickly migrate towards the more fundamental settings of ordered sets and metric spaces. We believe these developments can be easily adapted to other settings, where the semantic objects of interest are such probabilistic trees, regardless of the particular framework under which these trees are induced.

Finally, many of our results are of independent interests, outside the context of the current testing scenario. For instance, we define an ordering $\leq_\flat$ on ω-sequences over the unit interval and thus on the set of trace distributions. We favor $\leq_\flat$ over the pointwise ordering induced by the usual $\leq$ relation on $[0, 1]$, because the resulting CPO structures are algebraic, with a very natural characterization of compact elements. In addition, we give an explicit characterization of the set of probabilistic executions of an arbitrary PA $\mathcal{A}$, as well as a generic construction of limits. These are in turn used to show that the set of trace distributions induced by $\mathcal{A}$ is a closed set in an appropriate metric space and is closed under convex combinations. All such results are useful tools in formal verification.

Related Work. Several testing preorders and equivalences for probabilistic processes have been proposed in the literature [Christoff 1990; Segala 1996; Gregorio-Rodrígez and Núñez 1998; Cleaveland et al. 1999; Jonsson and Yi 2002; Deng et al. 2007a; 2007b]. All these papers study testing relations in the style of De Nicola and Hennesy [Nicola and Hennessy 1984]. That is, a *test* is defined as a (probabilistic) process that interacts with a system via shared actions and reports either success or failure. The various testing relations are then obtained by comparing success probabilities. Unlike our testing scenario, these papers do not describe how success probabilities can be observed from an external point of view. Therefore, in our opinion, these relations are not completely observational. In that sense, our work is more closely related to the seminal paper of Larsen and Skou [Larsen and Skou 1991], where probabilistic bisimulation is characterized by a testing scenario based on hypothesis testing. Technically, the setting in [Larsen and Skou 1991] is more restrictive than ours because of their minimal deviation assumption, which imposes a uniform lower bound on all transition probabilities and hence an upper bound on the probabilistic branching degree.

Also closely related is the fast emerging field of *statistical model checking* [Younes and Simmons 2002; Younes et al. 2004; Sen et al. 2004; Younes 2005]. Traditionally, a probabilistic model checker does its job by exploring the state space and computing numerically all relevant probabilities. In statistical model checking, the idea is instead to collect sample runs from the model. Properties of interest are formulated as test hypotheses and, by increasing the number of sample runs, one can control the probability of producing an erroneous answer to the model checking question. So far, statistical model checking techniques have been developed for discrete and continuous time Markov chains [Younes et al. 2004; Sen et al. 2004], semi-Markov processes [Sen et al. 2004] and stochastic discrete event systems [Younes and Simmons 2002; Younes 2005]. In most of these models, the notions of delay and relative timing are treated explicitly, whereas in our approach nondeterminism is used to model timing uncertainty. Much of our effort goes to show that standard techniques in hypothesis testing can be used to distinguish processes even in the presence of nondeterminism, as long as all nondeterministic choices are within a closed set.

Our development differs in another way from many other works on stochastic systems (e.g. [Edalat 1995; Baier and Kwiatkowska 1998; Desharnais et al. 2002]), which focus more on functional behaviors of these processes and hence probability distributions on the state space. These distributions are *conditional* upon occurrences of events, which are often interpreted as inputs to a system. In contrast, we focus on probability distributions on computation paths and traces, therefore we must take into account probability distributions on events, in addition to distributions on states. In this respect, our development is closer to [Vatan 2001], which studies properties of distribution functions (a generalized notion of language) generated by finite-state probabilistic automata. One may argue that this distinction between state-based and action-based reasonings is inconsequential, yet our experience suggests the slight difference in interpretation can lead to divergence in the methods of analysis and eventually in the types of application domains.

Organization. We start in Section 2 with an informal presentation of our testing scenario. Section 3 provides some mathematical preliminaries, while Section 4 recalls the definitions of probabilistic automata and their behaviors. In Section 5, we introduce in detail the design and motivation of our test scenario and, in Section 6, we provide an explicit characterization of the set of probabilistic executions and use that to prove convex closure properties and to construct limiting adversaries. Section 7 gives a formal treatment of finite approximations on three levels:

adversaries, probabilistic executions and trace distributions. Section 8 deals with technical results regarding metric convergence and Section 9 presents a proof of our main theorem. Concluding remarks and discussions of future work follow in Section 10.

2. PREVIEW: BUTTON-PUSHING EXPERIMENTS

Before presenting our results at a technical level, we give an informal overview of the proposed testing scenario. As described in Section 1, a typical button-pushing experiment consists of a process operating inside a black box. Given a process $\mathcal{S}$, such an experiment induces a set $Obs(\mathcal{S})$ of all observations that are possible/acceptable under $\mathcal{S}$. This in turn yields an observational equivalence: two LTSs $\mathcal{S}_1$ and $\mathcal{S}_2$ are equivalent if and only if $Obs(\mathcal{S}_1) = Obs(\mathcal{S}_2)$.

For instance, trace semantics for image finite[1] LTSs can be characterized by the *trace machine* [Glabbeek 2001], depicted in Figure 1 on the left. This machine has no buttons at all, thus the observer cannot influence its execution.

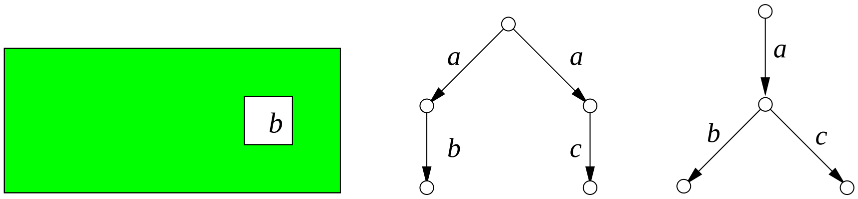

Fig. 1. The trace machine (left), and LTSs $\mathcal{S}_1$ and $\mathcal{S}_2$.

During a single experiment, the observer records the contents of the display over time, yielding a finite trace of the process inside the machine. Gathering all possible observations, we obtain a testing scenario that corresponds to trace equivalence. Indeed, the LTSs $\mathcal{S}_1$ and $\mathcal{S}_2$ in Figure 1 are trace equivalent and have the same observations under this testing scenario: ε (the empty sequence), a, ab and ac.

To obtain a testing scenario for probabilistic processes, we add to the trace machine a *reset* button, which brings the machine back to its initial state. The resulting *trace distribution machine* is depicted in Figure 2.

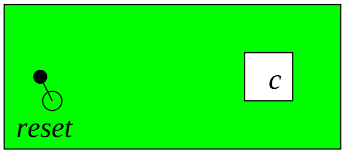

Fig. 2. The trace distribution machine.

An experiment on the trace distribution machine is carried out as follows.

[1] This means, for each state s and action a, only finitely many a-transitions are enabled in s (cf. Section 4).

(1) First, the observer fixes the *type* of the experiment: two natural numbers k and m. The first specifies the maximum length of each run and is referred to as the *depth* of the experiment. The second specifies the number of runs to be executed and is referred to as the *width*.

(2) The observer then starts the machine by pushing the reset button.

(3) As the machine executes, the action symbols appearing on the display are recorded in succession.

(4) When the display becomes empty, or when the observer has recorded k actions, the machine is reset and recording starts in a fresh column.

(5) The experiment stops when m runs of the machine have been recorded.

Table I illustrates a sample that *may* be obtained in a type-$\langle 2, 6\rangle$ experiment conducted on the process $\mathcal{S}_1$ from Figure 1. (In our setting, LTSs are degenerate probabilistic processes.)

1	2	3	4	5	6
a	a	a	a	a	a
c	b	c	b	c	c

Table I. Sample obtained in type-$\langle 2, 6\rangle$ experiment conducted on process $\mathcal{S}_1$.

So far, we have described how to collect a sample from the trace distribution machine. The next step is to use hypothesis testing to define the set of *type-$\langle k, m\rangle$ acceptable observations of $\mathcal{S}$*, denoted $\mathit{Obs}(\mathcal{S}, k, m)$, for a given process S and sample type $\langle k, m\rangle$. Then $\mathit{Obs}(\mathcal{S})$ is defined to be the union $\bigcup_{k,m} \mathit{Obs}(\mathcal{S}, k, m)$. In this way, two processes $\mathcal{S}_1$ and $\mathcal{S}_2$ are distinguished in our semantics if and only if there exists sample type $\langle k, m\rangle$ such that $\mathit{Obs}(\mathcal{S}_1, k, m) \neq \mathit{Obs}(\mathcal{S}_2, k, m)$.

As we mentioned in Section 1, this task is complicated by the semantics of our reset button. Namely, nondeterministic choices may be resolved differently in the various runs of an experiment, so that the traces recorded from these runs need not be identically distributed. These nondeterministic choices are said to be *demonic*, because we have no control over them.

To facilitate understanding, we first consider hypothesis tests in the weaker setting of *angelic* nondeterministic choices, where we do assume control. In Section 2.2, we explain how we adapt these tests to the original setting of demonic choices.

2.1 Hypothesis Testing: Angelic Nondeterminism

Consider a type-$\langle k, m\rangle$ experiment on a probabilistic process $\mathcal{S}$ with finite action alphabet[2] Act. Let $\mathit{Act}^{\leq k}$ denote the set of traces with length at most k. Suppose we can make sure that nondeterministic choices are resolved in the same way in all m runs, so that every run is associated with the same discrete probability distribution D on $\mathit{Act}^{\leq k}$.

[2] This finiteness restriction on Act can be replaced by a *finite branching* condition on processes [Stoelinga and Vaandrager 2003]. In Section 2.3 of the present paper, we introduce the extended trace distribution machine, which accommodates image finite processes with countably infinite action alphabet.

Fix such a trace β. We can view the m runs of this experiment as m independent *Bernoulli trials* as follows: during each run, a *success* occurs if the record for that run contains exactly β; otherwise, we have a *failure*. By assumption, these trials are identically distributed and the common parameter θ is precisely $D(\beta)$.

It is well-known that the frequency of successes from a Bernoulli sample is a *sufficient statistic* for the parameter θ. Intuitively, the number of successes in a sample contains all the information about θ that is present in the sample. This suggests we define our hypothesis test in terms of the frequency of successes. In fact, since $Act^{\leq k}$ is finite, we can do so for all traces β simultaneously, by devising a test with this null hypothesis: "the underlying probability distribution is D." This hypothesis is *accepted* if, for every β, the frequency of successes in the actual outcome is in the interval $[D(\beta) - r, D(\beta) + r]$; otherwise, it is *rejected*. Here r is some appropriate real number between 0 and 1. To discuss how we choose r, we need to bring in some terminology.

Since hypothesis tests are concerned with yes/no questions, there are two possible types of errors: *false rejection* and *false acceptance*. A good test should guarantee that the probability of committing either error is low. However, it is often hard to control these errors independently[3], therefore one typically starts with tests that control false rejections, while keeping false acceptance small. We adopt the same approach, namely, given any $\alpha \in [0, 1]$, we define tests with probability of false rejection at most α. These tests are said to have *level* α.

It may seem desirable to have tests that never commit false rejection errors (i.e., level 0). However, this strategy leads to rather uninteresting tests, because it forces acceptance whenever the actual outcome has nonzero probability under the null hypothesis. To avoid such triviality, one typically fixes a small but nonzero level, e.g. $\alpha = 0.05$. This quantity α determines the size of the *acceptance region*, which is the set of outcomes that lead to acceptance of the null hypothesis. In particular, an acceptance region should contain just enough possible outcomes so that the probability of false rejection is below α. A smaller acceptance region would violate the level-α requirement, while a larger one would lead to higher probability of false acceptance errors.

In our case, the size of the acceptance region depends on the value r and we choose the smallest r that give rise to a level-α test. Now we can define $Obs(D, k, m)$ to be this acceptance region, namely, the set of possible outcomes such that the frequency of successes for every β is in the interval $[D(\beta) - r, D(\beta) + r]$. The set of acceptable type-$\langle k,\ m\rangle$ observations for $\mathcal{S}$ is in turn given as $\bigcup_D Obs(D, k, m)$, where D ranges over all possible distributions induced by $\mathcal{S}$. The following example illustrate such hypothesis tests for a fair coin and a biased coin, respectively.

Example 2.1. Consider the two probabilistic processes in Figure 3. We interpret the symbol a as the action of flipping a coin, while b and c announce on which side the coin lands. Then $\mathcal{A}_1$ models a fair coin, i.e., the uniform distribution on the set $\{ab, ac\}$. Similarly, $\mathcal{A}_2$ models a coin with bias $\frac{1}{3}$ for heads, i.e., a distribution assigning probability $\frac{1}{3}$ to the trace ab and $\frac{2}{3}$ to the trace ac.

[3] In some cases, it is proven to be impossible to control false acceptance uniformly among all alternative parameters, while conforming to a certain tolerance of false rejection. We refer to Chapter 8 of [Casella and Berger 1990].

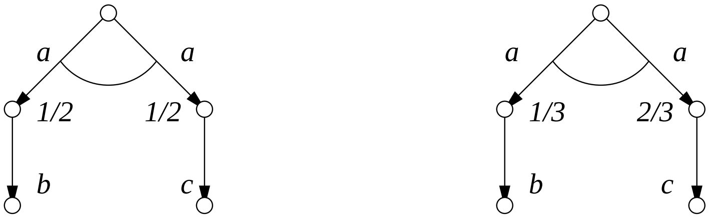

Fig. 3. Probabilistic processes $\mathcal{A}_1$ and $\mathcal{A}_2$.

Suppose α is set at 0.05 and we consider experiments of type $\langle 2, 100\rangle$. In other words, we observe 100 runs of length 2 each. The acceptance region for $\mathcal{A}_1$ consists of sequences in which the traces *ab* occurs between 41 and 59 times, while in the acceptance region for $\mathcal{A}_2$ the trace *ab* occurs between 24 and 42 times. If *ab* is actually observed 45 times, we answer "yes" in the test for $\mathcal{A}_1$ and "no" in the test for $\mathcal{A}_2$. Therefore, $\mathcal{A}_1$ and $\mathcal{A}_2$ are *distinguished* in our semantics.

Intuitively, the distinguishing power of this testing scenario is a direct consequence of the well-known (weak) law of large numbers. Given any small ϵ, we can toss a coin sufficiently many times so that it is extremely unlikely to observe a sample mean that deviates from the true bias by more than ϵ. This allows us to "separate" the acceptance regions of two coins with different biases.

It is interesting to note that the observational equivalence, thus obtained, is independent of the choice of α, because we have the freedom to vary the number of runs. In general, as α decreases, we must enlarge the acceptance regions for the two processes in question, possibly increasing the overlap between them. Therefore more runs need to be performed so that we can find sample points residing in the difference of the two acceptance regions.

2.2 Hypothesis Testing: Demonic Nondeterminism

In the angelic case, a width-m experiment on the trace distribution machine can be likened to tossing the same coin m times. Our testing scenario thus boils down to the problem of distinguishing two coins with different biases. In the demonic case, a width-m experiment can be likened to tossing a sequence of m coins with possibly different biases, and our testing scenario reduces to the following (slightly more complicated) problem.

Suppose we have a sequence S of coins with biases $p_0, p_1, p_2, \ldots$ such that every p_i is in a closed interval $I \subseteq [0, 1]$. Given any m, we devise a hypothesis test for the first m coins in S as follows: a length-m sequence of heads and tails leads to a "yes" answer if and only if the frequency of heads falls in the interval $[\overline{p}-r, \overline{p}+r]$. Here $\overline{p}$ is the average of $p_0, \ldots, p_{m-1}$ and r is chosen as before to guarantee a level-α test.

Suppose there is another coin with bias $q \notin I$ and, for each m, we construct a test for m tosses of the new coin in exactly the same way. (Here the midpoint of the interval is simply q.) The question we try to answer is: is there an m for which there exists a sample point that leads to a "yes" answer in the test for $p_0, \ldots, p_{m-1}$ but a "no" answer in the test for $q, \ldots, q$?

Again, we can appeal to the weak law of large numbers in the second test, with repeated tosses of the same coin. As it turns out, the same intuition also applies in the first test, despite the fact that the p_i's are possibly different. In Section 9, we prove an analog of the weak law of large numbers for independent Bernoulli variables, replacing the bias of a single coin with the average bias of m different coins (Lemma 9.2). This key observation, together with the fact that $\overline{p}$ and q are separated by the closed interval I, allows us to separate two acceptance regions just as in the angelic case.

Using the same trick of treating all traces in $Act^{\leq k}$ simultaneously, we generalize the above argument on coin tosses to trace distributions. It is therefore important that the set of all trace distributions of a probabilistic process forms a convex closed set.

2.3 Extension to Countably Infinite Action Alphabet

So far we have worked with processes with finite action alphabet, so that each length-k run has finitely many possible outcomes (namely, traces in $Act^{\leq k}$). This is an important property because our separation argument only works in finite-dimensional metric spaces. To preserve this property in the case of countably infinite action alphabet, we add buttons $0, 1, 2, \ldots$ to the trace distribution machine in Figure 2, yielding the *extended* trace distribution machine. This is depicted in Figure 4.

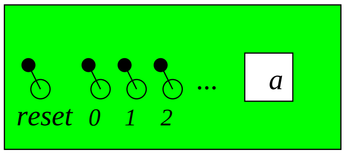

Fig. 4. The extended trace distribution machine.

At the start of each experiment (i.e., Step (1)), the observer fixes not only the depth and width of the experiment, but also the *breadth*. This is done by pressing exactly one of the buttons $l \in \mathbb{N}$, indicating that only the first l actions $\{b_0, b_1, \ldots, b_{l-1}\}$ of the alphabet[4] are enabled during the entire experiment. We then proceed exactly as before.

Notice that the type of an experiment now has three arguments: k, l and m. Given a process $\mathcal{S}$, $Obs(\mathcal{S})$ is defined as the union $\bigcup_{k,l,m} Obs(\mathcal{S}, k, l, m)$, where $Obs(\mathcal{S}, k, l, m)$ is the set of type-$\langle k,\, l,\, m\rangle$ acceptable outcomes of $\mathcal{S}$. This induces an observational equivalence that coincides with trace distribution equivalence, provided the processes are image finite. The image-finite requirement is necessary for the various convergence properties that are essential in our proofs. (This is very much analogous to the situation of LTSs and the trace machine.)

We can think of this new feature of action switches as a "finite testing policy": each experiment focuses on a finite number of possibilities. Since the observer may free an arbitrarily large number of actions, this is a sufficient method of exploring the entire structure.

[4] We assume a fixed enumeration of Act (cf. Section 4).

The rest of this paper studies image finite processes and the extended trace distribution machine. For brevity, we omit from now on the word ”extended”.

3. PRELIMINARIES

In this section we provide a summary of basic mathematical notions necessary for our development. In particular, we review materials from real analysis [Kolmogorov and Fomin 1970; Rudin 1987], probability theory [Cohn 1980; Rudin 1987], statistics [Casella and Berger 1990; Trivedi 2002] and order theory [Davey and Priestley 1990]. Our reader is encouraged to skip (portions of) this section as he sees fit.

3.1 Metric Spaces

We encounter many times in this paper the notion of “limits”. They come in two flavors: (i) limit of a sequence of points in some metric space, and (ii) limit of an increasing sequence in a partially ordered set. We now recall the former, while the latter is treated in Section 3.4.

Let $\mathbb{P}$ denote the set of non-negative real numbers. A *metric space* is a pair $\langle X,\ \mathit{dist}\rangle$ where X is a set and the function $\mathit{dist} : X \times X \to \mathbb{P}$ satisfies the following: for all $x,\ y \in X$,

(1) identity: $\mathit{dist}(x, y) = 0$ if and only if $x = y$;

(2) symmetry: $\mathit{dist}(x, y) = \mathit{dist}(y, x)$; and

(3) triangle inequality: $\mathit{dist}(x, z) \leq \mathit{dist}(x, y) + \mathit{dist}(y, z)$.

We give two familiar examples of metric spaces.

Example 3.1. The n-dimensional space $\mathbb{R}^n$ $(n \in \mathbb{N})$ together with the Euclidean distance function:

$$\mathit{dist}(\vec{x}, \vec{y}) := \sqrt{\sum_{i=0}^{n}(x_i - y_i)^2}.$$

Example 3.2. The infinite dimensional space $[l, u]^\omega$ $(l, u \in \mathbb{R}$ with $l < u)$ together with the distance function:

$$\mathit{dist}(\vec{x}, \vec{y}) := \sup_{i \in \mathbb{N}} |x_i - y_i|.$$

Given an arbitrary metric space $\langle X,\ \mathit{dist}\rangle$, we define the usual notion of an (open) *ϵ-ball* around a point x:

$$B_\epsilon(x) := \{y \in X \mid \mathit{dist}(x, y) < \epsilon\}.$$

A sequence of points $\{x_i \mid i \in \mathbb{N}\}$ in X *converges* to a *limit* $x \in X$ if, for every $\epsilon > 0$, there is $N_\epsilon \in \mathbb{N}$ such that $x_i \in B_\epsilon(x)$ for all $i \geq N_\epsilon$. Equivalently, we may require $\lim_{i\to\infty} \mathit{dist}(x, x_i) = 0$. It is trivial to check that limits must be unique and that all subsequences converge to the same limit.

The following is a special case of the famous Bolzano-Weierstraß Theorem.

Theorem 3.3. *Every bounded infinite sequence over $\mathbb{R}$ has a convergent subsequence.*

3.2 Probability Spaces

Let Ω be a set. A collection $\mathcal{F}$ of subsets of Ω is said to be a *σ-field* over Ω if $\mathcal{F}$ satisfies the following properties:

(1) $\Omega \in \mathcal{F}$;

(2) if $X \in \mathcal{F}$, then $\Omega \setminus X$ is also in $\mathcal{F}$ (closure under complement); and

(3) if $\{X_i \mid i \in \mathbb{N}\} \subseteq \mathcal{F}$, then $\bigcup_{i\in\mathbb{N}} X_i$ is also in $\mathcal{F}$ (closure under countable union).

We have the following familiar theorem about σ-fields.

THEOREM 3.4. *Let $\mathcal{S}$ be any family of subsets of Ω. There exists a smallest σ-field $\mathcal{F}$ over Ω such that $\mathcal{S} \subseteq \mathcal{F}$. In that case, we say that $\mathcal{F}$ is* generated *by $\mathcal{S}$.*

A *probability measure* on a σ-field $\mathcal{F}$ is a countably additive function $\mathbf{m} : \mathcal{F} \to [0, 1]$ such that $\mathbf{m}(\Omega) = 1$. *Countable additivity* says that, given any disjoint family $\{X_i \mid i \in \mathbb{N}\} \subseteq \mathcal{F}$, it must be the case that

$$\mathbf{m}(\bigcup_{i\in\mathbb{N}} X_i) = \sum_{i\in\mathbb{N}} \mathbf{m}(X_i).$$

If $\mathbf{m}$ is a probability measure, the triple $\langle\Omega, \mathcal{F}, \mathbf{m}\rangle$ is said to form a *probability space*. The set Ω is called the *sample space* and members of $\mathcal{F}$ are called *events*.

Example 3.5. The powerset of Ω, $\mathcal{P}(\Omega)$, is a σ-field over Ω. Consider a function $\mu : \Omega \to [0, 1]$ such that $\sum_{s\in\Omega} \mu(s) = 1$. Then μ induces a function $\mathbf{m} : \mathcal{P}(\Omega) \to [0, 1]$ as follows:

$$\mathbf{m}(X) := \sum_{s\in X} \mu(s).$$

It is easy to check that $\mathbf{m}$ is countably additive, hence a probability measure on $\mathcal{P}(\Omega)$.

Such a function μ is often called a *discrete probability distribution* over the set Ω. The *support* of μ is defined to be the set $\mathrm{supp}(\mu) := \{s \in \Omega \mid \mu(s) \neq 0\}$. Note that the support of a discrete probability distribution is a countable set. If $\mathrm{supp}(\mu)$ is a singleton $\{s\}$, then μ is called a *Dirac distribution* and is often written as $\{s \mapsto 1\}$. The set of all discrete probability distributions over Ω is denoted by $\mathrm{Distr}(\Omega)$.

Similarly, we define a *sub-probability measure* to be a countably additive function $\mathbf{m} : \mathcal{F} \to [0, 1]$ such that $\mathbf{m}(\Omega) \leq 1$. Thus a *discrete sub-distribution* is a function $\mu : \Omega \to [0, 1]$ such that $\sum_{s\in\Omega} \mu(s) \leq 1$. The set of all such sub-distributions is denoted $\mathrm{SubDistr}(\Omega)$.

Example 3.6. Let Ω be the two element set $\{0, 1\}$ and let μ be a discrete probability distribution over Ω. Write p for $\mu(1)$. This describes a *Bernoulli distribution* with parameter p. The two possible outcomes 1 and 0 are often referred to as *success* and *failure*, respectively.

3.3 Statistics

Let $\langle\Omega, \mathcal{F}, \mathbf{m}\rangle$ be a discrete probability space generated by the function $\mu : \Omega \to [0, 1]$. A *random variable* is a function $X : \Omega \to \mathbb{R}$. Intuitively, it is a rule that assigns a numerical value to each possible outcome of an experiment. Given $x \in \mathbb{R}$, let $[X = x]$ denote the event $\{s \in \Omega \mid X(s) = x\}$. The *probability mass function (pmf)* associated with X is defined by

$$p_X(x) := \mathbf{m}([X = x]) = \sum_{s \in [X=x]} \mu(s).$$

Often we write $\mathbf{P}[X = x]$ for $p_X(x)$. Similarly, we let $[X \geq x]$ denote the event $\{s \in \Omega \mid X(s) \geq x\}$ and write $\mathbf{P}[X \geq x]$ for $\sum_{s \in [X \geq x]} \mu(s)$.

The *expectation* (or *expected value*) of X, denoted $\mathbf{E}[X]$, is given by the sum

$$\mathbf{E}[X] := \sum_{\{x \in \mathbb{R} \mid [X=x] \neq \emptyset\}} x \; \mathbf{P}[X = x].$$

The *variance* of X, denoted $\mathrm{Var}[X]$, is defined as

$$\mathrm{Var}[X] := \mathbf{E}[(X - \mathbf{E}[X])^2] = \sum_{\{x \in \mathbb{R} \mid [X=x] \neq \emptyset\}} (x - \mathbf{E}[X])^2 \; \mathbf{P}[X = x].$$

Example 3.7. A *Bernoulli variable* is a random variable X with range $\{0, 1\}$. Intuitively, it classifies each outcome of an experiment as either success or failure. The value $\mathbf{P}[X = 1] = p$ is called the parameter of the Bernoulli variable. It is routine to derive $\mathbf{E}[X] = p$ and $\mathrm{Var}[X] = p(1 - p)$.

We have the following important inequality.

Theorem 3.8. *(Chebyshev's inequality). For every random variable X and $t > 0$,*

$$\mathbf{P}[|X - \mathbf{E}[X]| \geq t] \leq \frac{\mathrm{Var}[X]}{t^2}.$$

Next we consider hypothesis testing. This is a common method of *statistical inference*, which refers broadly to the practice of estimating characteristics of an entire population based on evidence produced by a sample drawn from that population. The starting point is a pair of complementary hypotheses: the *null* hypothesis and the *alternative* hypothesis. These are complementary statements about the probability distribution in question. A *hypothesis test* is a rule that specifies which sample values lead to the decision that the null hypothesis is accepted (thus the alternative hypothesis is rejected). This subset of the sample space is called the *acceptance region*, while its complement is called the *rejection region.* We say that a *false negative* (or *false rejection, type I*) error is committed if the null hypothesis is true but the test procedure concludes otherwise. Dually, a *false positive* (or *false acceptance*, *type II*) error is committed if the null hypothesis is false but is accepted by the test procedure. A test is said to be of *level* α ($\alpha \in [0, 1]$) if the probability of committing a type I error is at most α.

3.4 Partial Orders

A *partially ordered set* (or *poset*) is a set P endowed with a binary relation $\leq$, which is reflexive, (weakly) antisymmetric and transitive. Given a subset $X \subseteq P$, we write $\bigvee X$ for the least upperbound of X, if it exists.

A non-empty subset D of P is *directed* if every finite subset D' of D has an upperbound in D. The least upperbound of a directed set (if it exists) is often called a *directed limit.* The poset P forms a *complete partial order (CPO)* if it has a bottom element $\bot$ and all directed limits. A function $f : P \to Q$ between CPOs P and Q is *monotone* if, for all $p, p' \in P$, $p \leq p'$ implies $f(p) \leq f(p')$. Such a function is said to be *continuous* if it is monotone and, for every directed set D in P, we have $f(\bigvee D) = \bigvee f(D)$.

An increasing sequence of elements $p_0 \leq p_1 \leq p_2 \leq \ldots$ in P is called a *chain.* Chains are typical examples of directed sets and we write $\lim C$ for the least upperbound of a chain C. In fact, any directed limit can be converted to the limit of a chain with the same cardinality.

THEOREM 3.9. *A poset P with $\bot$ is a CPO if and only $\lim C$ exists for every non-empty chain C.*

Finally, an element $c \in P$ is *compact* if, for every directed set D such that $c \leq \bigvee D$, there exists $p \in D$ with $c \leq p$. A CPO P is said to be *algebraic* if, for all p, the set $\{c \mid c \leq p \text{ and } c \text{ compact}\}$ is directed and p is in fact the limit of this set.

Example 3.10. Let $X^{<\omega}$ (resp., X^{ω}) denote the set of finite (resp., infinite) sequences over a set X. Then the union of these two sets, denoted $X^{\leq\omega}$, forms an algebraic CPO under the prefix ordering $\sqsubseteq$. The compact elements are precisely the finite sequences.

Example 3.11. Let $X \rightharpoonup Y$ denote the set of partial functions from X to Y. We define the *information ordering* on $X \rightharpoonup Y$ as follows: $f \subseteq g$ if and only if (i) $Dom(f) \subseteq Dom(g)$ and (ii) for all $x \in Dom(f)$, $f(x) = g(x)$. In other words, the graph of f is a subset of the graph of g, hence the relation is also called the *subset ordering.* This gives rise to an algebraic CPO whose compact elements are partial functions with finite domain.

3.5 Infinite Sequences over $[0, 1]$

We define a *flat* ordering on $[0, 1]^{\omega}$ as follows: $\sigma \leq_{\flat} \sigma'$ if and only if, for all $i \in \mathbb{N}$, $\sigma_i \neq 0$ implies $\sigma_i = \sigma'_i$. This ordering is very much analogous to the subset ordering in Example 3.11, since infinite sequences over $[0, 1]$ can be viewed as functions from $\mathbb{N}$ to $[0, 1]$ and we can interpret $\sigma_i = 0$ as "σ undefined at i". Given an arbitrary directed limit in this poset, we can always convert it to the limit of an ω-chain. This is a strengthening of Theorem 3.9 for the special case of $[0, 1]^{\omega}$.

LEMMA 3.12. *Let $\mathcal{D}$ be an arbitrary (not necessarily countable) directed subset of $[0, 1]^{\omega}$. There is an ω-chain $\{\sigma_0, \sigma_1, \ldots\} \subset \mathcal{D}$ such that $\lim_{k\to\infty} \sigma_k = \bigvee \mathcal{D}$.*

PROOF. First we construct a sequence $\sigma'_0, \sigma'_1, \ldots$ as follows: for each $i \in \mathbb{N}$, choose $\sigma'_i \in \mathcal{D}$ such that $\sigma'_i(i) = (\bigvee \mathcal{D})(i)$. This is possible due to the definition of $\leq_{\flat}$. Then

—set σ_0 to be σ'_0;

—for $i + 1$, set σ_{i+1} to be any upperbound of $\{\sigma_0, \ldots, \sigma_i, \sigma'_{i+1}\}$ in $\mathcal{D}$.

Since $\mathcal{D}$ is directed, this ω-chain is well-defined. One can easily check that its limit in fact equals the least upperbound of $\mathcal{D}$. □

Lemma 3.12 is used to prove Lemma 3.13 about infinite sums. Let I be an arbitrary index set and let $\{\{c_{i,j}\}_{j\in\mathbb{N}} \mid i \in I\}$ be a set of ω-sequences over $[0,1]$. Assuming the infinite sums converge, it is true in general that

$$\bigvee_{i\in I}\sum_{j\in\mathbb{N}} c_{i,j} \leq \sum_{j\in\mathbb{N}}\bigvee_{i\in I} c_{i,j}.$$

We claim that equality holds under the assumption that $\{\{c_{i,j}\}_{j\in\mathbb{N}} \mid i \in I\}$ is directed with respect to $\leq_\flat$. This can be seen as a special form of the well-known Monotone Convergence Theorem.

Lemma 3.13. *Assume that* $\{\{c_{i,j} \mid j \in \mathbb{N}\} \mid i \in I\}$ *is a directed subset of* $[0,1]^\omega$ *and for all i,* $\sum_{j\in\mathbb{N}} c_{i,j}$ *converges to a limit in* $[0,1]$*. Then the sum* $\sum_{j\in\mathbb{N}}\bigvee_{i\in I} c_{i,j}$ *converges and*

$$\bigvee_{i\in I}\sum_{j\in\mathbb{N}} c_{i,j} = \sum_{j\in\mathbb{N}}\bigvee_{i\in I} c_{i,j}.$$

An obvious corollary of Lemma 3.13 concerns the set of discrete probabilistic sub-distributions.

Corollary 3.14. *Let S be a countable set. The set* SubDistr(S) *of discrete probabilistic sub-distributions over S is a CPO with respect to the flat ordering.*

Proof. Via an enumeration of S, we view SubDistr(S) as a subset of $[0,1]^\omega$. Clearly the everywhere-0 distribution is a bottom element. Given any directed subset Δ, we apply Lemma 3.13 to

$$\{\{D(j) \mid j \in \mathbb{N}\} \mid D \in \Delta\}$$

and conclude that the join of Δ is also a sub-distribution. □

4. PROBABILISTIC AUTOMATA

As described in Section 2, our trace distribution machine contains a probabilistic process which interacts with its environment via an action display and a collection of buttons. This section makes precise what we mean by a *probabilistic process* and its *behaviors*.

As far as we are concerned, a probabilistic process is a *(simple) probabilistic automaton* as introduced by Segala and Lynch [Segala 1995; Segala and Lynch 1995]. This extends the usual nondeterministic automata model by allowing probabilistic information at the target of each transition. More precisely, every transition leads to a probability distribution over possible next states, rather than a single state.

For simplicity, we consider systems with no internal actions. All external actions are taken from a countable set Act, which has a fixed enumeration $\{b_i \mid i \in \mathbb{N}\}$ throughout this paper. Given $l \in \mathbb{N}$, we write Act_l for the list $b_0, \ldots, b_{l-1}$. The set of finite (resp. infinite) traces is denoted $Act^{<\omega}$ (resp. Act^ω), while the set of all traces is $Act^{\leq\omega}$. Also, we write ε for the empty trace.

Definition 4.1. A *probabilistic automaton (PA)* is a triple $\mathcal{A} = (S, s^0, \Delta)$ where

—S is the set of states,

—$s^0 \in S$ is the initial state, and

—$\Delta \subseteq S \times \mathit{Act} \times \mathrm{Distr}(S)$ is the transition relation.

We write $s \xrightarrow{a} \mu$ for $(s, a, \mu) \in \Delta$. Also, we write $s \overset{a,\mu}{\rightsquigarrow} t$ whenever $s \xrightarrow{a} \mu$ and $\mu(t) > 0$. To avoid confusion, we sometimes refer to the components of $\mathcal{A}$ as $S_{\mathcal{A}}$, $s^0_{\mathcal{A}}$ and $\Delta_{\mathcal{A}}$.

Intuitively, we can view target distributions in the transition relation Δ as a form of probabilistic branching; that is, we think of $s \overset{a,\mu}{\rightsquigarrow} t$ as a nondeterministic transition $s \xrightarrow{a} \mu$ followed by a probabilistic transition $\mu \xrightarrow{\mu(t)} t$. In this way, we obtain an informal notion of the *underlying nondeterministic automaton* of $\mathcal{A}$, where we "forget" probabilistic information (i.e., $\mu(t)$) at each probabilistic transition. Thus inspired, we define paths in a probabilistic automaton $\mathcal{A}$ as follows.

Definition 4.2. A *path* π of $\mathcal{A}$ is a (finite or infinite) sequence of the form $s_0 a_1 \mu_1 s_1 a_2 \mu_2 s_2 \ldots$ such that:

—each s_i (resp., a_i, μ_i) denotes a state (resp., action, distribution over states);

—s_0 is the initial state[5];

—if π is finite, then it ends with a state;

—$s_i \overset{a_{i+1},\mu_{i+1}}{\rightsquigarrow} s_{i+1}$, for each non-final i.

The *length* of finite path π is the number of transitions occurring in it.

The set of all paths (finite and infinite) of $\mathcal{A}$ is denoted $\mathit{Path}(\mathcal{A})$, while the set of finite paths is denoted $\mathit{Path}^{<\omega}(\mathcal{A})$. We write $\mathit{Path}^{\leq k}(\mathcal{A})$ for the set of paths with length at most k. The last state of a finite path π is written $\mathit{last}(\pi)$. The *trace* of π, notation $\mathit{Tr}(\pi)$, is defined to be the sequence of actions appearing along π, that is, $a_1 a_2 a_3 \ldots$. Given $F \subseteq \mathit{Path}^{<\omega}\mathcal{A}$ and $a \in \mathit{Act}$, we write $\mathit{Succ}(F, a)$ for the set of paths π' of the form $\pi a \mu s$ with $\pi \in F$. Similarly for $\mathit{Succ}(F, \beta)$ where $\beta \in \mathit{Act}^{<\omega}$.

As in the case of nondeterministic automata, we are interested in certain finiteness properties in branching structure.

Definition 4.3. A PA $\mathcal{A}$ is *finitely (resp. countably) branching* if, for each state s, the set $\{\langle a, \mu\rangle \mid s \xrightarrow{a} \mu\}$ is finite (resp. countable). It is *image finite* if for each state s and action a, the set $\{\mu \mid s \xrightarrow{a} \mu\}$ is finite.

Thus, each state in a finitely branching PA has finitely many outgoing transitions, while a state in an image finite PA may have infinitely many. In both cases, the set $\{t \mid s \overset{a,\mu}{\rightsquigarrow} t \text{ for some } a, \mu\}$ could be infinite, since a target distribution μ may have infinite support. As a result, given a finite trace $\beta \in \mathit{Act}^{<\omega}$, a finitely branching (or image finite) PA may have infinitely many paths with trace β. This is different from the case of nondeterministic automata.

Throughout this paper, we focus on image finite probabilistic automata. Since Act is countable, it is immediate that every image finite probabilistic automaton is also countably branching. Moreover, each transition leads to a discrete distribution on states, which has a countable support. Therefore, $\mathit{Path}^{<\omega}(\mathcal{A})$ remains countable and we often take advantage of this fact by imposing an enumeration.

[5] In other terminology, paths may start from non-initial states.

4.1 Adversaries and Probabilistic Executions

We now turn to behaviors of probabilistic automata. In the non-probabilistic case, an execution (or path) is obtained by resolving all nondeterministic choices in a deterministic fashion. For a probabilistic automaton, we resolve nondeterministic choices by means of an *adversary* (or *scheduler*). Given any finite history leading to the current state, an adversary returns a discrete sub-distribution over the set of available next transitions. Therefore, our adversaries are (i) randomized, (ii) history-dependent, and (iii) partial, in the sense that they may choose to halt the execution at any time.

Definition 4.4. A *(randomized, history-dependent and partial) adversary* E of $\mathcal{A}$ is a function

$$E : \mathit{Path}^{<\omega}(\mathcal{A}) \to \mathrm{SubDistr}(\mathit{Act} \times \mathrm{Distr}(S_{\mathcal{A}}))$$

such that, for each finite path π, $E(\pi)(a, \mu) > 0$ implies $\mathit{last}(\pi) \xrightarrow{a} \mu$.

We write $\mathit{Adv}(\mathcal{A})$ for the set of all adversaries of $\mathcal{A}$. Intuitively, an adversary E tosses a coin to choose the next transition at every step of the computation of $\mathcal{A}$. Thus E induces a purely probabilistic "computation tree". This idea is captured by the notion of a probabilistic execution.

Definition 4.5. Let E be an adversary of $\mathcal{A}$. The *probabilistic execution* induced by E, denoted $\mathbf{Q}_E$, is the function from $\mathit{Path}^{<\omega}(\mathcal{A})$ to $[0, 1]$ defined recursively by

$$\begin{aligned} \mathbf{Q}_E(s_0) &= 1, \\ \mathbf{Q}_E(\pi a \mu s) &= \mathbf{Q}_E(\pi) \cdot E(\pi)(a, \mu) \cdot \mu(s). \end{aligned}$$

The set of all probabilistic executions of $\mathcal{A}$ is written as $\mathit{ProbExec}(\mathcal{A})$. Essentially, the function $\mathbf{Q}_E$ assigns probabilities to finite paths according to decisions made by the adversary E. We shall interpret "$\mathbf{Q}_E(\pi) = p$" as: under the control of adversary E, the automaton $\mathcal{A}$ follows path π with probability p. Notice that it need *not* be the case that $\mathcal{A}$ halts after π. Moreover, if $\pi \sqsubseteq \pi'$, then the event "$\mathcal{A}$ follows π'" implies the event "$\mathcal{A}$ follows π". Therefore $\mathbf{Q}_E$ is *not* a discrete distribution on the set of finite paths. However, $\mathbf{Q}_E$ does induce a probability space over the sample space $\mathit{Path}(\mathcal{A})$ as follows.

Definition 4.6. Let $\pi \in \mathit{Path}^{<\omega}(\mathcal{A})$ be given. The *cone* generated by π is the following set of paths: $C_\pi := \{\pi' \in \mathit{Path}(\mathcal{A}) \mid \pi \sqsubseteq \pi'\}$.

Let $\Omega_{\mathcal{A}} := \mathit{Path}(\mathcal{A})$ be the sample space and let $\mathcal{F}_{\mathcal{A}}$ be the smallest σ-field generated by the collection $\{C_\pi \mid \pi \in \mathit{Path}^{<\omega}(\mathcal{A})\}$. The following theorem states that $\mathbf{Q}_E$ induces a unique probability measure on $\mathcal{F}_{\mathcal{A}}$ [Segala 1995].

Theorem 4.7. *Let E be an adversary of $\mathcal{A}$. There exists a unique measure $\mathbf{m}_E$ on $\mathcal{F}_{\mathcal{A}}$ such that $\mathbf{m}_E[C_\pi] = \mathbf{Q}_E(\pi)$ for all $\pi \in \mathit{Path}^{<\omega}(\mathcal{A})$.*

The measure $\mathbf{m}_E$ in Theorem 4.7 gives rise to a probability space $(\Omega_{\mathcal{A}}, \mathcal{F}_{\mathcal{A}}, \mathbf{m}_E)$. In the literature, many authors define probabilistic executions to be such probability spaces. In this paper, we find it more natural to reason with the function $\mathbf{Q}_E$, rather than the induced probability space. Our choices in the definition of paths and the use of $\mathbf{Q}_E$ simplifies the technical development, for instance in Section 6. By virtue of Theorem 4.7, the two approaches are equivalent.

4.2 Trace Distributions

External behaviors of a probabilistic automaton $\mathcal{A}$ are obtained by removing the non-visible elements from probabilistic executions. Since we do not deal with internal actions, we remove states and distributions of states. This yields a trace distribution of $\mathcal{A}$, which assigns probabilities to certain sets of traces.

We define trace distributions via a lifting of the trace operator $Tr : Path^{<\omega}(\mathcal{A}) \to Act^{<\omega}$. The following lemma is needed to show that the lifting is well defined.

Lemma 4.8. *For all $\beta \in Act^{<\omega}$ and $E \in Adv(\mathcal{A})$, $\sum_{\pi \in Tr^{-1}(\beta)} \mathbf{Q}_E(\pi) \leq 1$.*

Proof. Induction on the length of β. If β is the empty sequence then $Tr^{-1}(\beta)$ consists of the singleton set $\{s^0\}$ and we have $\sum_{\pi \in Tr^{-1}(\beta)} \mathbf{Q}_E(\pi) = \mathbf{Q}_E(s^0) = 1$. Consider βa.

$$\begin{aligned}
&\sum_{\pi \in Tr^{-1}(\beta a)} \mathbf{Q}_E(\pi) \\
&= \sum_{\pi' \in Tr^{-1}(\beta)} \sum_{\mu : last(\pi') \xrightarrow{a} \mu} \sum_{s \in \mathrm{supp}(\mu)} \mathbf{Q}_E(\pi' a \mu s) \\
&= \sum_{\pi' \in Tr^{-1}(\beta)} \sum_{\mu : last(\pi') \xrightarrow{a} \mu} \sum_{s \in \mathrm{supp}(\mu)} \mathbf{Q}_E(\pi') \cdot E(\pi')(a,\mu) \cdot \mu(s) \\
&= \sum_{\pi' \in Tr^{-1}(\beta)} \sum_{\mu : last(\pi') \xrightarrow{a} \mu} \mathbf{Q}_E(\pi') \cdot E(\pi')(a,\mu) \cdot \sum_{s \in \mathrm{supp}(\mu)} \mu(s) \\
&= \sum_{\pi' \in Tr^{-1}(\beta)} \sum_{\mu : last(\pi') \xrightarrow{a} \mu} \mathbf{Q}_E(\pi') \cdot E(\pi')(a,\mu) \\
&= \sum_{\pi' \in Tr^{-1}(\beta)} \mathbf{Q}_E(\pi') \cdot \sum_{\mu : last(\pi') \xrightarrow{a} \mu} E(\pi')(a,\mu) \\
&\leq \sum_{\pi' \in Tr^{-1}(\beta)} \mathbf{Q}_E(\pi') \leq 1 \qquad \text{by induction hypothesis}
\end{aligned}$$

□

Definition 4.9. Let an adversary E of $\mathcal{A}$ be given and consider the probabilistic execution $\mathbf{Q}_E : Path^{<\omega}(\mathcal{A}) \to [0,1]$. The *trace distribution* induced by E is the function $Tr(\mathbf{Q}_E) : Act^{<\omega} \to [0,1]$ given by

$$Tr(\mathbf{Q}_E)(\beta) := \sum_{\pi \in Tr^{-1}(\beta)} \mathbf{Q}_E(\pi).$$

We usually write $\mathbf{D}_E$ for $Tr(\mathbf{Q}_E)$ and, when it is desirable to leave the adversary E implicit, we use variables D, K, etc. The set of trace distributions of $\mathcal{A}$ is denoted by $TrDist(\mathcal{A})$.

Similarly to the case of probability executions, each $\mathbf{D}_E$ induces a probability measure on the sample space $\Omega := Act^{\leq\omega}$. There the σ-field $\mathcal{F}$ is generated by the collection $\{C_\beta \mid \beta \in Act^{<\omega}\}$, where $C_\beta := \{\beta' \in \Omega \mid \beta \sqsubseteq \beta'\}$.

Theorem 4.10. *Let E be an adversary of $\mathcal{A}$. There exists a unique measure $\mathbf{m}^E$ on $\mathcal{F}$ such that $\mathbf{m}^E[C_\beta] = \mathbf{D}_E(\beta)$.*

Again, $\mathbf{m}^E$ gives rise to a probability space $\langle \Omega, \mathcal{F}, \mathbf{m}^E \rangle$, which is elsewhere called the trace distribution induced by E. We refer to [Segala 1995] for these alternative definitions and the proofs of Theorems 4.7 and 4.10.

Finally, we define trace distribution inclusion as follows:

$$\mathcal{A} \sqsubseteq_{\mathrm{TD}} \mathcal{B} \text{ if and only if } \mathit{TrDist}(\mathcal{A}) \subset \mathit{TrDist}(\mathcal{B}).$$

Trace distribution equivalence is thus: $\mathcal{A} \equiv_{\mathrm{TD}} \mathcal{B}$ if and only if $\mathit{TrDist}(\mathcal{A}) = \mathit{TrDist}(\mathcal{B})$.

4.3 Finite Adversaries

Let E be an adversary of a probabilistic automaton $\mathcal{A}$. Given a finite path π, we say that π is *E-reachable* if $\mathbf{Q}_E(\pi) \neq 0$. Recall that adversaries may choose to halt an execution at any point. This is reflected by the fact that $E(\pi)$ is a sub-distribution on the set of possible next transitions, so the probability of E halting after π is

$$1 - \sum_{\langle a,\, \mu \rangle \in \mathrm{supp}(E(\pi))} E(\pi)\langle a,\, \mu \rangle.$$

If $E(\pi)$ has empty support, then we say E *halts* after path π. In that case, $\mathbf{Q}_E(\pi') = 0$ for any proper extension π' of π. We say that E has *depth* (at most) k if E halts after every path of length k. This implies that every E-reachable path has length at most k.

The notion of depth gives a bound on how far an adversary follows each path. We also wish to talk about the degree of branching in an adversary. A typical approach is to give a bound on the cardinality of $\mathrm{supp}(E(\pi))$ for all π. Here we propose a different definition: E has *breadth* (at most) l if, for all E-reachable paths π, we have $\mathit{Tr}(\pi) \in (\mathit{Act}_l)^{<\omega}$.

For all $k, l \in \mathbb{N}$, let $\mathit{Adv}(\mathcal{A}, k, l)$ denote the set of adversaries of depth k and breadth l. We say that E is a *finite* adversary if there exists $k, l \in \mathbb{N}$ such that $E \in \mathit{Adv}(\mathcal{A}, k, l)$. In other words, E is finite if it has both finite depth and finite breadth. The following lemma follows immediately from the relevant definitions.

Lemma 4.11. *Let $E \in \mathit{Adv}(\mathcal{A}, k, l)$ and $\pi \in \mathit{Path}^{<\omega}(\mathcal{A})$ be given. If π is E-reachable then $\mathit{Tr}(\pi) \in (\mathit{Act}_l)^{\leq k}$.*

Finite adversaries are extremely important in our development, because we focus on reduction of infinite behavior to its finite approximations. This idea will become clear in Sections 5 and 7. In the meantime, we make some simple observations.

Lemma 4.12. (*1*) *If $\mathcal{A}$ is an image finite probabilistic automaton and E is an adversary of $\mathcal{A}$ with finite breadth, then* $\mathrm{supp}(E(\pi))$ *is finite for every E-reachable π.*

(*2*) *There exist image finite probabilistic automaton $\mathcal{A}$ and adversary E of $\mathcal{A}$ such that* $\mathrm{supp}(E(\pi))$ *is finite for all π but E has infinite breadth.*

Proof. For the first claim, suppose π is an E-reachable path in $\mathcal{A}$. By image finiteness, there are only finitely many a-transitions available at π for each $a \in \mathit{Act}$.

By finite breadth of E, there are only finitely many $a \in Act$ such that E assigns non-zero probability to transitions labeled a. Therefore, $\mathrm{supp}(E(\pi))$ is finite.

For the second claim, consider a single-state automaton with countably many loops such that no two loops carry the same label. Let E be an adversary that always chooses (with probability 1) a transition carrying a fresh label. Then $\mathrm{supp}(E(\pi))$ is a singleton for all π and yet E has infinite breadth. □

We extend the notion of finiteness to probabilistic executions: $\mathbf{Q}_E$ is finite if there is an E' such that E' is finite and $\mathbf{Q}_E = \mathbf{Q}_{E'}$. The set of probabilistic executions induced by adversaries from $Adv(\mathcal{A}, k, l)$ is denoted $ProbExec(\mathcal{A}, k, l)$.

We define finite trace distributions analogously: $\mathbf{D}_E$ is finite just in case there is a finite E' such that $\mathbf{D}_E = \mathbf{D}_{E'}$. The set of trace distributions induced by adversaries from $Adv(\mathcal{A}, k, l)$ is denoted $TrDist(\mathcal{A}, k, l)$. Also, we write $\mathcal{A} \sqsubseteq^{k,l}_{\mathrm{TD}} \mathcal{B}$ whenever $TrDist(\mathcal{A}, k, l) \subseteq TrDist(\mathcal{B}, k, l)$.

Finally, we use $Adv(\mathcal{A}, k, -)$ to denote the set of all adversaries with depth k (and arbitrary breadth). The same convention applies also to $Adv(\mathcal{A}, -, l)$, $ProbExec(\mathcal{A}, k, -)$, etc.

5. OBSERVATIONS

Having defined trace distributions, we move on to the other side of our story: observations. We begin this section by recalling the procedure of sample collection from a trace distribution machine. Then we identify samples that are *acceptable* if the trace distribution machine operates as specified by a probabilistic automaton $\mathcal{A}$. A sample O falls into this category just in case there exists a possible sequence of trace distributions $D_0, \ldots, D_{m-1}$ under which O is an acceptable outcome. Such samples will constitute the set of observations of $\mathcal{A}$. To save space, we use $\vec{D}$ to denote (syntactically) $D_0, \ldots, D_{m-1}$. Similarly for $\vec{D}'$, $\vec{K}$, etc.

5.1 Sampling

We associate with each experiment a triple $\langle k, l, m \rangle$ of natural numbers. We call this the *type* of the experiment, which specifies some parameters in the data collection procedure. More precisely, an observer conducts a *depth-k*, *breadth-l* and *width-m* experiment on a trace distribution machine as follows.

(1) First, the observer presses the button labeled by l, activating the actions in Act_l.

(2) The observer then starts the machine by pushing the reset button.

(3) As the machine executes, the action symbols appearing on the display are recorded in succession.

(4) When the display becomes empty, or when the observer has recorded k actions, the machine is reset and recording starts in a fresh column.

(5) The experiment stops when m runs of the machine have been recorded.

During such an experiment, an observer records a sequence $\beta_0, \ldots, \beta_{m-1}$, where each β_i is a sequence of actions symbols from Act_l and has length at most k. We call such a record O a *sample* of depth k, breadth l and width m (or simply a sample of *type* $\langle k, l, m \rangle$). A trace β is said to *appear* in $\beta_0, \ldots, \beta_{m-1}$ if $\beta = \beta_i$ for some i. When k, l and m are clear from context, we will write $\mathcal{U}$ for the universe of all possible samples of type $\langle k, l, m\rangle$; that is, $\mathcal{U} := ((Act_l)^{\leq k})^m$.

We assume the trace distribution machine is governed by a PA $\mathcal{A}$. During each run, the trace distribution machine chooses a trace β according to some trace distribution D of $\mathcal{A}$. When the observer presses the *reset* button, the machine returns to the initial state of $\mathcal{A}$ and starts over with a possibly different trace distribution of $\mathcal{A}$. Since all actions outside Act_l are blocked, and each time the machine is allowed to perform at most k steps, a run of the trace distribution machine is essentially governed by a trace distribution from $\mathit{TrDist}(\mathcal{A}, k, l)$. Thus, each sample O of width m is generated by a sequence of m trace distributions from $\mathit{TrDist}(\mathcal{A}, k, l)$.

Let us focus for a moment on a single run. It is possible to record a trace β with length strictly less than k. This happens whenever the machine halts after displaying the sequence β. Therefore, given traces $\beta_0 \neq \beta_1$, the two events "observing exactly β_0" and "observing exactly β_1" are mutually exclusive. This holds even when β_0 is a prefix of β_1. Based on this interpretation, the probability of recording exactly β (written $\mathbf{P}_{D,k}[\beta]$) equals:

—$D(\beta)$, if the length of β is exactly k;

—$D(\beta) - \sum_{a\in Act_l} D(\beta a)$, otherwise.

Notice that the second clause corresponds to the case in which $\mathcal{A}$ halts after β. The following lemma justifies our definition of $\mathbf{P}_{D,k}$.

LEMMA 5.1. *For every $D \in \mathit{TrDist}(\mathcal{A}, k, l)$, the function $\mathbf{P}_{D,k} : (Act_l)^{\leq k} \to [0, 1]$ is a discrete probability distribution over $(Act_l)^{\leq k}$.*

PROOF. First we check that the range of $\mathbf{P}_{D,k}$ is included in $[0, 1]$. Let $\beta \in (Act_l)^{\leq k}$ be given and let $\mathbf{m}$ be the unique probability measure associated with D (see Theorem 4.10). We have $D(\beta) = \mathbf{m}[C_\beta] \in [0, 1]$. Moreover, the set $\{C_{\beta a} \mid a \in Act_l\}$ is a countable family of pairwise-disjoint members of the σ-algebra $\mathcal{F}$, therefore the set $C_\beta \setminus \bigcup_{a\in Act_l} C_{\beta a}$ is measurable. Thus

$$D(\beta) - \sum_{a\in Act_l} D(\beta a) = \mathbf{m}[C_\beta \setminus \bigcup_{a\in Act_l} C_{\beta a}] \in [0, 1].$$

It remains to verify $\sum_{\beta\in(Act_l)^{\leq k}} \mathbf{P}_{D,k}[\beta] = 1$. Without loss, assume $k \geq 1$.

$$\begin{aligned}
&\sum_{\beta\in(Act_l)^{\leq k}} \mathbf{P}_{D,k}[\beta] \\
&= \sum_{i=0}^{k} \sum_{\beta\in(Act_l)^i} \mathbf{P}_{D,k}[\beta] = \sum_{i=0}^{k-1} \sum_{\beta\in(Act_l)^i} \mathbf{P}_{D,k}[\beta] + \sum_{\beta\in(Act_l)^k} \mathbf{P}_{D,k}[\beta] \\
&= \sum_{i=0}^{k-1} \sum_{\beta\in(Act_l)^i} (D(\beta) - \sum_{a\in Act_l} D(\beta a)) + \sum_{\beta\in(Act_l)^k} D(\beta)
\end{aligned}$$

$$= \sum_{i=0}^{k-1} (\sum_{\beta \in (Act_l)^i} D(\beta) - \sum_{\beta \in (Act_l)^i} \sum_{a \in Act_l} D(\beta a)) + \sum_{\beta \in (Act_l)^k} D(\beta)$$

$$= \sum_{i=0}^{k-1} (\sum_{\beta \in (Act_l)^i} D(\beta) - \sum_{\beta \in (Act_l)^{i+1}} D(\beta)) + \sum_{\beta \in (Act_l)^k} D(\beta)$$

$$= \sum_{\beta \in (Act_l)^0} D(\beta) = D(\varepsilon) = 1$$

□

Now we put together the m runs in an experiment. Note that each run involves two distinct types of choices: first the machine chooses a trace distribution D, then D in turn chooses a trace β. We do not make any assumptions on the first type of choices. However, once D_i is chosen for run i, D_i is solely responsible for selecting a trace β_i. That is, for any $i \neq j$, the choice of β_i by D_i is independent from the choice of β_j by D_j. Therefore, assuming trace distributions $\vec{D}$ are chosen, the probability of generating a depth-k sample $O = \beta_0, \ldots, \beta_{m-1}$ can be expressed as:

$$\mathbf{P}_{\vec{D},k}[O] := \prod_{i=0}^{m-1} \mathbf{P}_{D_i,k}[\beta_i].$$

For a set $\mathcal{O}$ of such samples, we have $\mathbf{P}_{\vec{D},k}[\mathcal{O}] := \sum_{O \in \mathcal{O}} \mathbf{P}_{\vec{D},k}[O]$.

Finally, we make a quick remark about $\mathbf{P}_{D,k}$. Namely, if two trace distributions from $\mathit{TrDist}(\mathcal{A}, k, l)$ induce the same discrete distribution on $(Act_l)^{\leq k}$, then they must be identical.

LEMMA 5.2. *The function* $\mathbf{P}_{-,k} : \mathit{TrDist}(\mathcal{A}, k, l) \to \mathit{Disc}((Act_l)^{\leq k})$ *is one-to-one.*

PROOF. We will give a left inverse of $\mathbf{P}_{-,k}$. Let $D \in \mathit{TrDist}(\mathcal{A}, k, l)$ be given. Define a function $D' : (Act_l)^{\leq k} \to [0, 1]$ as follows:

$$D'(\beta) = \sum_{\beta \sqsubseteq \beta'; \beta' \in (Act_l)^{\leq k}} \mathbf{P}_{D,k}[\beta'].$$

Using a (backwards) inductive argument on the length of $\beta \in (Act_l)^{\leq k}$, it is easy to check that $D = D'$. □

5.2 Frequencies

Our statistical analysis is based on the frequencies with which finite traces from $(Act_l)^{\leq k}$ appear in a sample O. Formally, the *frequency* of β in O is given by:

$$\mathit{freq}(O)(\beta) := \frac{\#\{i \mid 0 \leq i < m \text{ and } \beta = \beta_i\}}{m}.$$

Although each run is governed by a possibly different distribution, we can still obtain useful information from frequencies of traces. This is done as follows. Fix k, l, m, $\vec{D}$ and $\beta \in (Act_l)^{\leq k}$. For each $0 \leq i \leq m - 1$, we say that a *success* occurs at the i-th run just in case the observer records exactly β at the i-th run. Thus, the probability of a success at the i-th run is given by $\mathbf{P}_{D_i,k}[\beta]$. This can be viewed as a Bernoulli distribution with parameter $\mathbf{P}_{D_i,k}[\beta]$. Let X_i denote such a random variable. Then the random variable $Z := \frac{1}{m}\sum_{i=1}^{m} X_i$ represents the frequency of successes in the m trials governed by $\vec{D}$. Moreover, the expected value of this frequency is:

$$\mathbf{E}_{\beta}^{\vec{D},k} := \mathbf{E}Z = \mathbf{E}(\frac{1}{m}\sum_{i=0}^{m-1} X_i) = \frac{1}{m}\sum_{i=0}^{m-1}\mathbf{E}(X_i) = \frac{1}{m}\sum_{i=0}^{m-1}\mathbf{P}_{D_i,k}[\beta].$$

Notice that both $\mathit{freq}(O)$ and $\mathbf{E}^{\vec{D},k}$ can be viewed as points in the metric space $[0,1]^{(Act_l)^{\leq k}}$ with distance function[6] $\mathit{dist}(\vec{u},\vec{v}) := \sup_{\beta\in(Act_l)^{\leq k}} |u_\beta - v_\beta|$. Thus $\mathit{dist}(\mathit{freq}(O),\mathbf{E}^{\vec{D},k})$ provides a very natural way to quantify the deviation between $\mathit{freq}(O)$ and $\mathbf{E}^{\vec{D},k}$. This plays a central role in classifying acceptable outcomes of $\vec{D}$.

5.3 Acceptable Outcomes: Motivation

Returning to our original goal, we would like to define a set of acceptable outcomes of $\mathcal{A}$. This is done by defining a set of acceptable outcomes for each sequence $\vec{D}$ of trace distributions. Thus, in the terminology of hypothesis testing, we develop a test with this null hypothesis: the sample O is generated by the sequence $\vec{D}$.

Fix an $\alpha \in (0,1)$ as the desired level of the test. Also fix the sample type $\langle k, l, m\rangle$. The set $\mathit{Obs}(\vec{D},k,l,m,\alpha)$ of acceptable outcomes should then satisfy the following:

(1) $\mathbf{P}_{\vec{D},k}[\mathit{Obs}(\vec{D},k,l,m,\alpha)] \geq 1-\alpha$, and

(2) $\mathbf{P}_{\vec{D}',k}[\mathit{Obs}(\vec{D},k,l,m,\alpha)]$ is minimized for different choices of $\vec{D}'$.

Condition 1 says the probability of false rejection (i.e., rejecting O as a sample generated by $\vec{D}$ while it is so) is at most α. Condition 2 says the probability of false acceptance (i.e., accepting O as a sample generated by $\vec{D}$ while it is not) should be reasonably small. Note that the probability of false acceptance depends highly upon the choice of $\vec{D}'$. Loosely speaking, if $\vec{D}$ and $\vec{D}'$ are very close to each other, then the probability of false acceptance becomes very high.

The design of our test stems from the concept of interval estimation. After each experiment, we try to make an educated guess about the trace distributions governing our machine, based on the sample just observed.

In case the m trials are identically distributed, i.e., controlled by the same trace distribution D, one typically uses $\mathit{freq}(O)(\beta)$ as an estimator for the value $\mathbf{P}_{D,k}[\beta]$. (By virtue of Lemma 5.2, this also gives an estimator for D.) Since the probability of making exactly the right guess is small, an interval around $\mathit{freq}(O)(\beta)$ is used to guarantee that the guess is correct with probability $1-\alpha$, where α is the prescribed level. That is, if $\mathit{freq}(O)(\beta)$ is observed, then our guess is $\mathbf{P}_{D,k}[\beta]$ falls in the interval $[\mathit{freq}(O)(\beta) - r, \mathit{freq}(O)(\beta) + r]$, where r depends on the level α.

[6] This metric is chosen (instead of the usual Euclidean metric) because it generalizes easily to higher dimensional cases. For instance, consider the space $[0,1]^{Act^{<\omega}}$ with $\mathit{dist}'(\vec{u},\vec{v}) := \sup_{\beta\in Act^{<\omega}} |u_\beta - v_\beta|$. Then, given any two points $\vec{u},\vec{v} \in [0,1]^{(Act_l)^{\leq k}}$, the distance between them in $[0,1]^{(Act_l)^{\leq k}}$ coincides with the distance in $[0,1]^{Act^{<\omega}}$.

Inverting this interval around $\mathbf{P}_{D,k}[\beta]$, we obtain a set of values for $\mathit{freq}(O)(\beta)$, namely, the interval $[\mathbf{P}_{D,k}[\beta]-r, \mathbf{P}_{D,k}[\beta]+r]$. If a frequency from this interval is actually observed, then our guess about $\mathbf{P}_{D,k}[\beta]$ would be correct. Thus, a frequency vector $\mathit{freq}(O)$ is deemed acceptable if, for all β, $\mathit{freq}(O)(\beta)$ is within the appropriate interval around $\mathbf{P}_{D,k}[\beta]$.

In the formal definitions that follow, the situation is slightly different: we do not always have the same trace distribution in all m trials. Thus we cannot give an estimate to the value $\mathbf{P}_{D,k}[\beta]$ for a single trace distribution D. Instead, we use $\mathit{freq}(O)(\beta)$ as an estimator for $\mathbf{E}_{\beta}^{\vec{D},k} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{P}_{D_i,k}[\beta]$, an average from the m trace distributions.

5.4 Acceptable Outcomes: Definition

As explained above, we accept a sample O if $\mathit{freq}(O)$ is within some distance r of the value $\mathbf{E}^{\vec{D},k}$. Our task is to find an appropriate $r \in [0,1]$ such that Condition 1 is satisfied. Moreover, for Condition 2, we need to minimize r in order to reduce the probability of false acceptance.

Recall that the (closed) ball centered at $\mathbf{E}^{\vec{D},k}$ with radius r is given by:

$$B_r(\mathbf{E}^{\vec{D},k}) := \{v \in [0,1]^{(Act_l)^{\leq k}} \mid \forall \beta \in (Act_l)^{\leq k}, |v(\beta) - \mathbf{E}_{\beta}^{\vec{D},k}| \leq r\}.$$

Then $\mathit{freq}^{-1}(B_r(\mathbf{E}^{\vec{D},k}))$ is the set of samples whose frequencies deviate from the average $\mathbf{E}^{\vec{D},k}$ by at most r.

Definition 5.3. Fix $k, l, m \in \mathbb{N}$ and a sequence $\vec{D}$ of trace distributions from $\mathit{TrDist}(\mathcal{A}, k, l)$. Let

$$\bar{r} := \inf\{r \mid \mathbf{P}_{\vec{D},k}[\mathit{freq}^{-1}(B_r(\mathbf{E}^{\vec{D},k}))] > 1-\alpha\}.$$

The set of type-$\langle k,\, l,\, m\rangle$ *acceptable outcomes* of $\vec{D}$ (with level α) is defined to be:

$$\mathit{Obs}(\vec{D}, k, l, m, \alpha) := \mathit{freq}^{-1}(B_{\bar{r}}(\mathbf{E}^{\vec{D},k})) = \{O \mid \mathit{dist}(\mathit{freq}(O), \mathbf{E}^{\vec{D},k}) \leq \bar{r}\}.$$

The set of type-$\langle k,\, l,\, m\rangle$ *acceptable outcomes* of $\mathcal{A}$ (with level α) is then:

$$\mathit{Obs}(\mathcal{A}, k, l, m, \alpha) := \bigcup_{\vec{D} \in (\mathit{TrDist}(\mathcal{A},k,l))^m} \mathit{Obs}(\vec{D}, k, l, m, \alpha).$$

Example 5.4. Let Act be $\{a, b, c\}$ and α be 0.05. Consider the automaton of Figure 5 with a nondeterministic choice between two branches and let $\vec{D}$ be a sequence of 10 trace distributions generated by: 4 adversaries that choose the left branch with probability 1 and 6 that choose the right branch with probability 1.

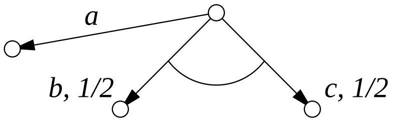

Figure 5

Then the average of the 10 induced trace distributions assign the value 0.4 to a and 0.3 to each of b and c. Notice that the frequency of a in every possible outcome is 0.4. Thus the following two outcomes have the greatest distance from the average: the one in which b never occurs and the one in which c never occurs. It is easy to verify that $\mathit{Obs}(\vec{D}, 1, 3, 10, 0.05)$ contains all but these two outcomes.

It is interesting to note that, while our notion of acceptable outcomes captures the clustering of samples around the expected value, it often fails to capture individual outcomes with relatively high probability. We illustrate this point with the following example.

Example 5.5. Consider an almost fair coin, say, with 0.51 for heads and 0.49 for tails. Suppose we toss this coin 10 times. The most likely outcome, all heads, has frequency vector $\langle 1, 0\rangle$, which lies very far from the expected frequency of $\langle 0.51, 0.49\rangle$. In fact, it is easy to check that for $\alpha = 0.005$, this most likely outcome is rejected.

Finally, we define our notion of observation preorder based on acceptable outcomes.

Definition 5.6. Let $\mathcal{A}, \mathcal{B}$ be probabilistic automata and let $\alpha \in (0, 1)$ be given. We write $\mathcal{A} \leq_\alpha \mathcal{B}$ if, for all $k, l, m \in \mathbb{N}$, $\mathit{Obs}(\mathcal{A}, k, l, m, \alpha) \subset \mathit{Obs}(\mathcal{A}, k, l, m, \alpha)$. We say that $\mathcal{A}$ and $\mathcal{B}$ are *observationally indistinguishable* up to level α just in case $\mathcal{A} \leq_\alpha \mathcal{B}$ and $\mathcal{B} \leq_\alpha \mathcal{A}$.

6. MORE ON PROBABILISTIC EXECUTIONS AND TRACE DISTRIBUTIONS

This section presents some basic results on probabilistic executions and trace distributions. First we give an explicit characterization of probabilistic executions. This characterization is then used to prove that the set of trace distributions, $\mathit{TrDist}(\mathcal{A})$, is closed under convex combinations. Finally, we describe a method of constructing an adversary from an infinite sequence of adversaries.

6.1 Characterizing Probabilistic Executions

By definition, a probabilistic execution $\mathbf{Q}_E$ is a mapping from $\mathit{Path}^{<\omega}(\mathcal{A})$ to $[0, 1]$, induced by some adversary E of a probabilistic automaton $\mathcal{A}$. Hence we can view $\mathbf{Q}$ as an operator from the set of adversaries of $\mathcal{A}$ to the function space $\mathit{Path}^{<\omega}(\mathcal{A}) \to [0, 1]$. This section provides an explicit characterization of the image of $\mathbf{Q}$. In other words, given an arbitrary function $Q : \mathit{Path}^{<\omega}(\mathcal{A}) \to [0, 1]$, we determine whether $Q = \mathbf{Q}_E$ for some adversary E of $\mathcal{A}$.

Clearly, if Q is induced by some E, it must satisfy the following properties.

(1) $Q(s^0) = 1$ and, whenever π is a prefix of π', $Q(\pi) \geq Q(\pi')$ (i.e., Q is antitone with respect to the prefix ordering).

(2) Given π, a, μ, s_0, s_1 such that $\mathit{last}(\pi) \xrightarrow{a} \mu$ and $s_0, s_1 \in \mathrm{supp}(\mu)$, we have

$$\frac{Q(\pi a \mu s_0)}{\mu(s_0)} = \frac{Q(\pi a \mu s_1)}{\mu(s_1)}.$$

We call this property the *consistency* of Q.

(3) Given $\pi \in \mathit{Path}^{<\omega}(\mathcal{A})$ with $Q(\pi) \neq 0$, let S_π denote the set of (a, μ) such that $\mathit{last}(\pi) \xrightarrow{a} \mu$. For each $(a, \mu) \in S_\pi$, fix any $s_{a,\mu} \in \mathrm{supp}(\mu)$. Then

$$\sum_{(a,\mu)\in S_\pi} \frac{Q(\pi a \mu s_{a,\mu})}{Q(\pi) \cdot \mu(s_{a,\mu})} \leq 1.$$

(Notice that, if Q is consistent, the choice of $s_{a,\mu}$ does not affect the summand.)

To see that these conditions are not only necessary but also sufficient to characterize the set of probabilistic executions, we note the following. Condition (1) expresses that, if $\pi \sqsubseteq \pi'$, then the event "$\mathcal{A}$ follows π'" is included in the event "$\mathcal{A}$ follows π'". Also, any probabilistic execution begins at the start state s^0 with probability 1. Condition (2) is more subtle. Recall that $\mathbf{Q}_E(\pi a \mu s) = \mathbf{Q}_E(\pi) \cdot E(\pi)(a, \mu) \cdot \mu(s)$. If $Q(\pi) > 0$, we can recover the value $E(\pi)(a, \mu)$ from Q by setting $\frac{Q(\pi a \mu s)}{Q(\pi) \cdot \mu(s)}$ for some state $s \in \mathrm{supp}(\mu)$, provided any choice of s yields the same quotient. This is precisely Condition(2). Condition (3) then says the sum of $E(\pi)(a, \mu)$ over all possible transitions $\mathit{last}(\pi) \xrightarrow{a} \mu$ must be under 1, i.e., $E(\pi)$ is a discrete sub-distribution on S_π.

Given a function Q with these properties, we construct an adversary E_Q as follows: for π, a and μ, define $E_Q(\pi)(a, \mu)$ to be

—0, in case $Q(\pi) = 0$ or $\mathit{last}(\pi) \xrightarrow{a} \mu$ is not a transition in $\mathcal{A}$;
—$\frac{Q(\pi a \mu s)}{Q(\pi) \cdot \mu(s)}$ otherwise, where s is some state in $\mathrm{supp}(\mu)$.

By Conditions 2 and 3, E_Q is well-defined and $E_Q(\pi)$ is a discrete sub-distribution for every π. Moreover, $E_Q(\pi)(a, \mu) \neq 0$ only if $\mathit{last}(\pi) \xrightarrow{a} \mu$ is a transition in $\mathcal{A}$, therefore E_Q is an adversary for $\mathcal{A}$. It remains to prove $Q = \mathbf{Q}_{E_Q}$ (so that we have a right inverse of the operation $\mathbf{Q}$).

Lemma 6.1. *For all $\pi \in \mathit{Path}^{<\omega}(\mathcal{A})$, we have $Q(\pi) = \mathbf{Q}_{E_Q}(\pi)$.*

Proof. By induction on the length of π. If π consists of just the initial state, then $Q(\pi) = 1 = \mathbf{Q}_{E_Q}(\pi)$.

Now consider π' of the form $\pi a \mu s$. If $Q(\pi) = 0$, then $Q(\pi') = 0$ by Condition 1. Also by induction hypothesis, $\mathbf{Q}_{E_Q}(\pi) = Q(\pi) = 0$. Hence

$$\mathbf{Q}_{E_Q}(\pi') = \mathbf{Q}_{E_Q}(\pi) \cdot E_Q(\pi)(a, \mu) \cdot \mu(s) = 0 = Q(\pi'),$$

regardless of the values of $E_Q(\pi)(a, \mu)$ and $\mu(s)$.

Otherwise, we may choose π'' as in the definition of $E_Q(\pi)(a, \mu)$. Let s' denote $\mathit{last}(\pi'')$. Then

$$\begin{aligned}
\mathbf{Q}_{E_Q}(\pi') &= \mathbf{Q}_{E_Q}(\pi) \cdot E_Q(\pi)(a, \mu) \cdot \mu(s) && \text{definition } \mathbf{Q}_{E_Q} \\
&= Q(\pi) \cdot \frac{Q(\pi'')}{Q(\pi) \cdot \mu(s')} \cdot \mu(s) && \text{I.H. and definition of } E(\pi')(a, \mu) \\
&= \frac{Q(\pi'') \cdot \mu(s)}{\mu(s')} \\
&= Q(\pi'). && \text{consistency of } Q
\end{aligned}$$

□

This completes the proof of the following characterization theorem.

Theorem 6.2 Characterization of probabilistic executions. *For all $Q : \mathit{Path}^{<\omega}(\mathcal{A}) \to [0,1]$, Q is the probabilistic execution induced by some adversary E of $\mathcal{A}$ if and only if Q satisfies Conditions (1), (2) and (3).*

6.2 Convex Combinations

Recall that probabilistic executions are mappings from $\mathit{Path}^{<\omega}(\mathcal{A})$ to $[0,1]$. Thus it makes sense to talk about convex combinations of two (or finitely many) of them. Similarly for trace distributions, which are mappings from $\mathit{Act}^{<\omega}$ to $[0,1]$.

Lemma 6.3. *Let $p \in [0,1]$ be given and let E_0 and E_1 be adversaries of $\mathcal{A}$. There exists an adversary E of $\mathcal{A}$ such that $\mathbf{Q}_E = p \cdot \mathbf{Q}_{E_0} + (1-p) \cdot \mathbf{Q}_{E_1}$.*

Proof. Define $Q := p \cdot \mathbf{Q}_{E_0} + (1-p) \cdot \mathbf{Q}_{E_1}$. By Theorem 6.2, it suffices to verify Conditions (1), (2) and (3). The first two are straightforward. For Condition (3), let π, S_π and $\{s_{a,\mu} \mid \langle a,\, \mu\rangle \in S_\pi\}$ be given as stated. Then

$$\begin{aligned}
&\sum_{(a,\mu)\in S_\pi} \frac{Q(\pi a \mu s_{a,\mu})}{Q(\pi)\cdot \mu(s_{a,\mu})}\\
&= \sum_{(a,\mu)\in S_\pi} \frac{p\cdot \mathbf{Q}_{E_0}(\pi a\mu s_{a,\mu}) + (1-p)\cdot \mathbf{Q}_{E_1}(\pi a \mu s_{a,\mu})}{Q(\pi)\cdot \mu(s_{a,\mu})}\\
&= \sum_{(a,\mu)\in S_\pi} \frac{p\cdot \mathbf{Q}_{E_0}(\pi)\cdot E_0(\pi)(a,\mu) + (1-p)\cdot \mathbf{Q}_{E_1}(\pi)\cdot E_1(\pi)(a,\mu)}{Q(\pi)}\\
&= \frac{p\cdot \mathbf{Q}_{E_0}(\pi)\cdot \sum_{(a,\mu)\in S_\pi} E_0(\pi)(a,\mu) + (1-p)\cdot \mathbf{Q}_{E_1}(\pi)\cdot \sum_{(a,\mu)\in S_\pi} E_1(\pi)(a,\mu)}{Q(\pi)}\\
&\le \frac{p\cdot \mathbf{Q}_{E_0}(\pi) + (1-p)\cdot \mathbf{Q}_{E_1}(\pi)}{Q(\pi)} = 1
\end{aligned}$$

□

The next lemma says that Tr preserves convex combinations. This follows immediately from the definition of $\mathit{Tr} : (\mathit{Path}^{<\omega}(\mathcal{A}) \to [0,1]) \to (\mathit{Act}^{<\omega} \to [0,1])$ (cf. Section 4.2).

Lemma 6.4. *Let $p \in [0,1]$ be given and let E_0 and E_1 be adversaries of $\mathcal{A}$. Then*

$$\mathit{Tr}(p \cdot \mathbf{Q}_{E_0} + (1-p)\cdot \mathbf{Q}_{E_1}) = p \cdot \mathit{Tr}(\mathbf{Q}_{E_0}) + (1-p)\cdot \mathit{Tr}(\mathbf{Q}_{E_1}).$$

Corollary 6.5. *The set of trace distributions of $\mathcal{A}$ is closed under convex combinations.*

Proof. By Lemma 6.3 and Lemma 6.4. □

We have one more corollary, which concerns the discrete probability distribution $\mathbf{P}_{D,k}$ (cf. Section 5.1).

Corollary 6.6. *For all $k, l \in \mathbb{N}$, the set $\{\mathbf{P}_{D,k} \mid D \in \mathit{TrDist}(\mathcal{A},k,l)\}$ is closed under convex combinations.*

Proof. By Corollary 6.5 and the definition of $\mathbf{P}_{D,k}$. □

6.3 Limit Construction

Suppose we have an infinite sequence $\{E_i\}_{i\in\mathbb{N}}$ of adversaries. From this, we construct an infinite decreasing sequence of sequences: (i) set the initial sequence $\{E_j^0\}_{j\in\mathbb{N}}$ to be $\{E_i\}_{i\in\mathbb{N}}$; (ii) for each $n \in \mathbb{N}$, define a subsequence $\{E_j^{n+1}\}_{j\in\mathbb{N}}$ of $\{E_j^n\}_{j\in\mathbb{N}}$. While choosing the appropriate subsequences, we obtain a function $Q : \mathit{Path}^{<\omega}(\mathcal{A}) \to [0,1]$ such that Q is the probabilistic execution induced by some adversary E. Once we specify our notion of convergence, such E is an obvious candidate for the limit of $\{E_i\}_{i\in\mathbb{N}}$.

By assumption, $\mathcal{A}$ is countably branching, hence $\mathit{Path}^{<\omega}(\mathcal{A})$ is countable. Let $\{\pi_n\}_{n\in\mathbb{N}}$ be an enumeration of that set. Given $n \in \mathbb{N}$, the sequence $\{\mathbf{Q}_{E_j^n}(\pi_n)\}_{j\in\mathbb{N}}$ is an infinite sequence in $[0,1]$. By Theorem 3.3, there is a convergent subsequence. Let $\{E_j^{n+1}\}_{j\in\mathbb{N}}$ be a subsequence of $\{E_j^n\}_{j\in\mathbb{N}}$ such that $\{\mathbf{Q}_{E_j^{n+1}}(\pi_n)\}_{j\in\mathbb{N}}$ converges. Define

$$Q(\pi_n) := \lim_{j\to\infty} \mathbf{Q}_{E_j^{n+1}}(\pi_n).$$

Given an adversary E_j^n as above, let $\mathit{index}(E_j^n)$ denote the index of E_j^n in the original sequence $\{E_i\}_{i\in\mathbb{N}}$.

The idea here is, at each stage n, we decide the value of Q at path π_n. Moreover, we remove those adversaries whose probabilistic executions (evaluated at π_n) fail to converge to $Q(\pi_n)$, taking care that we still have infinitely many adversaries left. As a consequence, at every stage after n, the probabilistic executions of remaining adversaries converge to the same limit at π_n. This claim is formalized in the following lemma.

Lemma 6.7. *For all $n < n'$, $\{\mathbf{Q}_{E_j^{n'}}(\pi_n)\}_{j\in\mathbb{N}}$ converges to $Q(\pi_n)$.*

Proof. For all $n < n'$, $\{E_j^{n'}\}_{j\in\mathbb{N}}$ is a subsequence of $\{E_j^n\}_{j\in\mathbb{N}}$. Hence sequence $\{\mathbf{Q}_{E_j^{n'}}(\pi_n)\}_{j\in\mathbb{N}}$ converges to the same limit as $\{\mathbf{Q}_{E_j^n}(\pi_n)\}_{j\in\mathbb{N}}$, namely, to $Q(\pi_n)$. □

Corollary 6.8. *Let $S \subseteq \mathbb{N}$ be finite. For all $n \in S$, $\{\mathbf{Q}_{E_j^{\max(S)+1}}(\pi_n)\}_{j\in\mathbb{N}}$ converges to $Q(\pi_n)$.*

The meaning of Corollary 6.8 is best explained by: "finitely many is the same as just one." Instead of taking the defining sequence of $Q(\pi_n)$ for each n, we can simply go to a much later stage in the construction where, for each $n \in S$, the weight on π_n is guaranteed to converge to the right value. Notice that it is essential that S is finite. With this idea in mind, we prove that Q satisfies Conditions (1), (2) and (3) in Section 6.1; then we apply Theorem 6.2 to conclude there is an adversary E with $\mathbf{Q}_E = Q$.

By definition, $Q(s^0) = 1$; moreover, the next lemma shows that Q is antitone with respect to prefix ordering on $\mathit{Path}^{<\omega}(\mathcal{A})$. Therefore Q satisfies Condition (1).

Lemma 6.9. *Let $\pi, \pi' \in \mathit{Path}^{<\omega}(\mathcal{A})$ be given. Suppose π is a prefix of π', then $Q(\pi) \geq Q(\pi')$.*

Proof. Choose $n, n' \in \mathbb{N}$ such that $\pi = \pi_n$ and $\pi' = \pi_{n'}$. Let $N := \max(n, n')$. Recall that for every j, we have $\mathbf{Q}_{E_j^{N+1}}(\pi) \geq \mathbf{Q}_{E_j^{N+1}}(\pi')$. Therefore, by Corollary 6.8,

$$Q(\pi) = \lim_{j\to\infty} \mathbf{Q}_{E_j^{N+1}}(\pi) \geq \lim_{j\to\infty} \mathbf{Q}_{E_j^{N+1}}(\pi') = Q(\pi').$$

□

The following lemmas verify Conditions (2) and (3).

Lemma 6.10 Condition (2). *Let $n, n_1, n_2 \in \mathbb{N}$ be given. Suppose $\pi_{n_1} = \pi_n a \mu s_1$, $\pi_{n_2} = \pi_n a \mu s_2$, $last(\pi_n) \xrightarrow{a} \mu$ and $s_1, s_2 \in \mathrm{supp}(\mu)$. Then*

$$\frac{Q(\pi_{n_1})}{\mu(s_1)} = \frac{Q(\pi_{n_2})}{\mu(s_2)}.$$

Proof. Let $N := \max(n_1, n_2)$. By Corollary 6.8 and the consistency of $\mathbf{Q}_{E_j^{N+1}}$, we have

$$\frac{Q(\pi_{n_1})}{\mu(s_1)} = \lim_{j\to\infty} \frac{\mathbf{Q}_{E_j^{N+1}}(\pi_{n_1})}{\mu(s_1)} = \lim_{j\to\infty} \frac{\mathbf{Q}_{E_j^{N+1}}(\pi_{n_2})}{\mu(s_2)} = \frac{Q(\pi_{n_2})}{\mu(s_2)}.$$

□

Lemma 6.11 Condition (3). *Let π be a path in $Path^{<\omega}(\mathcal{A})$ such that $Q(\pi) \neq 0$. Recall that S_π denotes the set $\{(a,\mu) \mid last(\pi) \xrightarrow{a} \mu\}$. For each $(a,\mu) \in S_\pi$, let $s_{a,\mu} \in \mathrm{supp}(\mu)$ be given. Then*

$$\sum_{(a,\mu)\in S_\pi} \frac{Q(\pi a \mu s_{a,\mu})}{Q(\pi)\cdot \mu(s_{a,\mu})} \leq 1.$$

Proof. Let $\{(a_k, \mu_k)\}_{k\in\mathbb{N}}$ be a (possibly finite) enumeration of S_π. It suffices to show that all finite partial sums are below 1. Let $K \in \mathbb{N}$ be given. For each $0 \leq k \leq K$, let n_k be the index of $\pi a_k \mu_k s_{a_k,\mu_k}$ in the enumeration $\{\pi_n\}_{n\in\mathbb{N}}$. Similarly, let n be the index of π. Define N to be $\max\{n_0, \ldots, n_K, n\} + 1$. Then by Corollary 6.8 we have

$$\sum_{k=0}^{K} \frac{Q(\pi_{n_k})}{Q(\pi)\cdot \mu_k(s_{a_k,\mu_k})} = \sum_{k=0}^{K} \frac{\lim_{j\to\infty} \mathbf{Q}_{E_j^N}(\pi_{n_k})}{Q(\pi)\cdot \mu_k(s_{a_k,\mu_k})}$$

By the definition of $\mathbf{Q}_{E_j^N}$, this becomes

$$\begin{aligned}
&\sum_{k=0}^{K} \lim_{j\to\infty} \frac{\mathbf{Q}_{E_j^N}(\pi)\cdot E_j^N(\pi)(a_k,\mu_k)\cdot \mu_k(s_{a_k,\mu_k})}{Q(\pi)\cdot \mu_k(s_{a_k,\mu_k})} \\
&= \sum_{k=0}^{K} \lim_{j\to\infty} \frac{\mathbf{Q}_{E_j^N}(\pi)\cdot E_j^N(\pi)(a_k,\mu_k)}{Q(\pi)} \\
&= \lim_{j\to\infty} \frac{\mathbf{Q}_{E_j^N}(\pi)}{Q(\pi)} \sum_{k=0}^{K} E_j^N(\pi)(a_k,\mu_k) && \text{finite sum} \\
&\leq \lim_{j\to\infty} \frac{\mathbf{Q}_{E_j^N}(\pi)}{Q(\pi)} && E_j^N(\pi) \text{ sub-distribution} \\
&= 1. && \text{Corollary 6.8}
\end{aligned}$$

□

So far we have presented a construction that yields an adversary from any given countable sequence of adversaries. Let us now consider two examples in which this construction fails to provide a sensible "limit".

Example 6.12. Consider the infinitely branching automaton $\mathcal{A}$ drawn in Figure 6, where all transitions are labeled with symbol a and all target distributions are Dirac distributions. Consider this sequence $\{E_k\}_{k\in\mathbb{N}}$ of adversaries: each E_k follows the kth branch of $\mathcal{A}$ with probability 1 and halts at the end of that branch. Thus, each E_k induces the trace distribution $\{a^k \mapsto 1\}$, where a^k is the length-k trace containing all a's. Intuitively, the limit of this sequence of trace distributions should assign probability 1 to the infinite trace $aa\ldots$; yet this is not possible, simply because $\mathcal{A}$ has no infinite paths. In this case, our limit construction yields the everywhere-halting adversary.

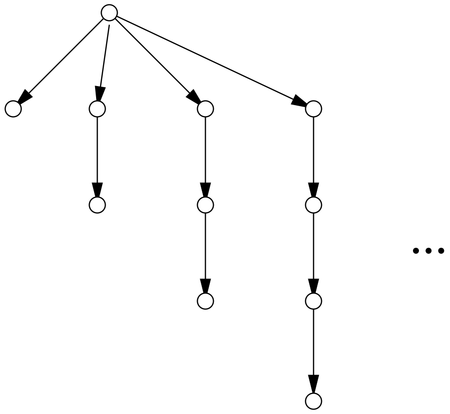

Figure 6

Example 6.13. Consider automaton $\mathcal{A}$ as in Example 6.12. Take the following sequence $\{E_k\}_{k\in\mathbb{N}}$ of adversaries: (i) at the start state, each E_k schedules the k-th transition with probability $\frac{2^k-1}{2^k}$ and halts with probability $\frac{1}{2^k}$; (ii) item every E_k halts completely after one step. This sequence of adversaries induce the following sequence of trace distributions:

$$\{\{a \mapsto \frac{2^k-1}{2^k}\} \mid k \in \mathbb{N}\}.$$

Intuitively, this is a converging sequence with limit $\{a \mapsto 1\}$. However, the limit of $\{E_k\}_{k\in\mathbb{N}}$, as constructed in the present section, is again the everywhere-halting adversary.

In Section 7, we will prove CPO properties of *ProbExec*($\mathcal{A}$) and *TrDist*($\mathcal{A}$) for image finite $\mathcal{A}$. In particular, our results imply that image finiteness is sufficient to remove Counterexample 6.12. In Section 8, we prove that image finiteness implies *TrDist*($\mathcal{A}, k, l$) forms a closed set in the metric space $[0,1]^{Act^{<\omega}}$, thus Counterexample 6.13 is also removed.

7. CPO PROPERTIES

In an earlier version of this paper [Stoelinga 2002], we proved the following *Approximation Induction Principle (AIP)* (cf. [Bergstra and Klop 1986; Baeten et al. 1987]) for probabilistic processes. A very similar result was observed by Segala [Segala 1996], who presented an informal proof sketch.

Theorem 7.1 AIP. *Let $\mathcal{A}$ and $\mathcal{B}$ be PAs and let $\mathcal{B}$ be finitely branching. Then*

$$\forall k[\mathcal{A} \sqsubseteq^k_{\mathrm{TD}} \mathcal{B}] \to \mathcal{A} \sqsubseteq_{\mathrm{TD}} \mathcal{B}.$$

The AIP provides a useful strategy for proving trace inclusion between probabilistic automata. The goal of this section is to strengthen it in a more abstract setting, thus obtaining the original Theorem 7.1 as a corollary. In particular, we relax the finite branching requirement to image finiteness.

Given an image finite probabilistic automaton $\mathcal{A}$, we define partial orders on these three sets: $\mathit{Adv}(\mathcal{A})$, $\mathit{ProbExec}(\mathcal{A})$ and $\mathit{TrDist}(\mathcal{A})$. We show that, in the case of $\mathit{TrDist}(\mathcal{A})$, we obtain an algebraic CPO whose compact elements are precisely the finite trace distributions defined in Section 4.3. We also prove that the operator $\mathbf{Q} : \mathit{Adv}(\mathcal{A}) \to \mathit{ProbExec}(\mathcal{A})$ is continuous and bottom preserving, and present an example to illustrate that the operator $\mathit{Tr} : \mathit{ProbExec}(\mathcal{A}) \to \mathit{TrDist}(\mathcal{A})$ is not continuous.

7.1 Image Finite Automata

Every adversary E for an image finite automaton $\mathcal{A}$ is *bounded* in the following sense: given any finite trace β and a small, positive error ϵ, it is possible to find a finite set $F \subseteq \mathit{Tr}^{-1}(\beta)$ such that $\mathbf{Q}_E$ assigns probability at least $\mathbf{D}_E(\beta) - \epsilon$ on F. The finite set F is a *uniform* bound, in that it depends only on β and ϵ, but not on the choice of adversary E. Existence of such a uniform bound is the key to avoiding counterexamples such as that in Example 6.12.

We now give a formal proof of this boundedness claim. Notice that Lemma 7.2 does not require image finiteness.

Lemma 7.2. *For all $F \subset \mathit{Path}^{<\omega}(\mathcal{A})$ and $\beta \in \mathit{Act}^{<\omega}$, we have*

$$\sum_{\pi \in F} \mathbf{Q}_E(\pi) \geq \sum_{\pi' \in \mathit{Succ}(F,\beta)} \mathbf{Q}_E(\pi'),$$

provided both sums converge.

Proof. By induction on the length of β. If β is the empty sequence, then $\mathit{Succ}(F, \beta) = F$ and the inequality trivially holds. Consider βa and let $\pi' \in$

$Succ(F, \beta a)$ be given. By definition of $\mathbf{Q}_E$, we have the following.

$$\begin{aligned}
&\sum_{\pi' \in Succ(F,\beta a)} \mathbf{Q}_E(\pi') \\
&= \sum_{\pi'' \in Succ(F,\beta)} \sum_{\mu : last(\pi'') \xrightarrow{a} \mu} \sum_{s \in \mathrm{supp}(\mu)} \mathbf{Q}_E(\pi'') \cdot E(\pi'')(a,\mu) \cdot \mu(s) \\
&= \sum_{\pi'' \in Succ(F,\beta)} \mathbf{Q}_E(\pi'') \cdot (\sum_{\mu : last(\pi'') \xrightarrow{a} \mu} E(\pi'')(a,\mu) \cdot \sum_{s \in \mathrm{supp}(\mu)} \mu(s)) \\
&= \sum_{\pi'' \in Succ(F,\beta)} \mathbf{Q}_E(\pi'') \cdot (\sum_{\mu : last(\pi'') \xrightarrow{a} \mu} E(\pi'')(a,\mu))
\end{aligned}$$

Since E is a discrete sub-distribution, the inner sum is at most 1 and the whole expression is at most $\sum_{\pi'' \in Succ(F,\beta)} \mathbf{Q}_E(\pi'')$. Applying the induction hypothesis, this is at most $\sum_{\pi \in F} \mathbf{Q}_E(\pi)$. □

Lemma 7.3. *Assume $\mathcal{A}$ is image finite. Let $\epsilon > 0$ be given. For every finite path π and action symbol a, there exists finite $F \subseteq Succ(\pi, a)$ such that for every adversary E, $\sum_{\pi' \in Succ(\pi,a) \setminus F} \mathbf{Q}_E(\pi') \leq \epsilon$.*

Proof. Since $\mathcal{A}$ is image finite, there are finitely many μ's such that $last(\pi) \xrightarrow{a} \mu$. Call them $\mu_0, \ldots, \mu_{n-1}$. For each $0 \leq i \leq n-1$, choose a finite subset $F_i \subseteq \mathrm{supp}(\mu_i)$ such that

$$\sum_{s \in \mathrm{supp}(\mu_i) \setminus F_i} \mu_i(s) \leq \frac{\epsilon}{n}.$$

Define F to be $\bigcup_{0 \leq i \leq n-1} \{\pi a \mu_i s \mid s \in F_i\}$. Clearly F is finite. For any adversary E, we have

$$\begin{aligned}
&\sum_{\pi' \in Succ(\pi,a) \setminus F} \mathbf{Q}_E(\pi') \\
&= \sum_{0 \leq i \leq n-1} \sum_{s \in \mathrm{supp}(\mu_i) \setminus F_i} \mathbf{Q}_E(\pi) \cdot E(\pi)(a,\mu_i) \cdot \mu_i(s) \\
&\leq \sum_{0 \leq i \leq n-1} \sum_{s \in \mathrm{supp}(\mu_i) \setminus F_i} \mu_i(s) \qquad\qquad \mathbf{Q}_E(\pi) \leq 1; E(\pi)(a,\mu_i) \leq 1 \\
&\leq n \cdot \frac{\epsilon}{n} = \epsilon.
\end{aligned}$$

□

Lemma 7.4. *Assume $\mathcal{A}$ is image finite. Let $\epsilon > 0$ and $\beta \in Act^{<\omega}$ be given. There exists finite $F_\beta \subseteq Tr^{-1}(\beta)$ such that for all adversaries E, $\sum_{\pi \in Tr^{-1}(\beta) \setminus F_\beta} \mathbf{Q}_E(\pi) \leq \epsilon$.*

Proof. We proceed by induction on the length of β. If β is the empty sequence, then take F_β to be the singleton $\{s^0\}$.

Consider a finite trace βa and assume the induction hypothesis holds for β. Choose finite F_β such that for all E, $d_E := \sum_{\pi \in Tr^{-1}(\beta) \setminus F_\beta} \mathbf{Q}_E(\pi) \leq \frac{\epsilon}{2}$. By

Lemma 7.2, we have for all E,

$$\sum_{\pi' \in Succ(Tr^{-1}(\beta) \setminus F_\beta, a)} \mathbf{Q}_E(\pi') \leq d_E \leq \frac{\epsilon}{2}.$$

If F_β is empty, then

$$\begin{aligned}\sum_{\pi' \in Tr^{-1}(\beta a) \setminus \emptyset} \mathbf{Q}_E(\pi') &= \sum_{\pi' \in Tr^{-1}(\beta a)} \mathbf{Q}_E(\pi') \\ &= \sum_{\pi' \in Succ(Tr^{-1}(\beta), a)} \mathbf{Q}_E(\pi') \\ &= \sum_{\pi' \in Succ(Tr^{-1}(\beta) \setminus F_\beta, a)} \mathbf{Q}_E(\pi') \\ &\leq \frac{\epsilon}{2} \leq \epsilon.\end{aligned}$$

Otherwise, let $\pi_0, \ldots, \pi_n$ be an enumeration of F_β and let $0 \leq i \leq n$ be given. By Lemma 7.3, we may choose $F_i \subseteq Succ(\pi_i, a)$ such that for all E, $c_{E,i} := \sum_{\pi' \in Succ(\pi_i, a) \setminus F_i} \mathbf{Q}_E(\pi') \leq \frac{\epsilon}{2(n+1)}$. Let F be $\bigcup_{0 \leq i \leq n} F_i$. We have for all E,

$$\begin{aligned}\sum_{\pi \in Tr^{-1}(\beta a) \setminus F} \mathbf{Q}_E(\pi) &= \sum_{0 \leq i \leq n} \sum_{\pi' \in Succ(\pi_i, a) \setminus F_i} \mathbf{Q}_E(\pi') + \sum_{\pi' \in Succ(Tr^{-1}(\beta) \setminus F_\beta, a)} \mathbf{Q}_E(\pi') \\ &\leq (\sum_{0 \leq i \leq n} c_{E,i}) + d_E \\ &\leq (n+1) \cdot \frac{\epsilon}{2(n+1)} + \frac{\epsilon}{2} = \epsilon.\end{aligned}$$

□

7.2 Adversaries

We define the *flat* ordering on $Adv(\mathcal{A})$: $E \leq_\flat E'$ if, for all finite executions π, action symbols a and distributions μ, $E(\pi)(a, \mu) \neq 0$ implies $E(\pi)(a, \mu) = E'(\pi)(a, \mu)$. As the name suggests, this is essentially the same ordering on $[0, 1]^\omega$ defined in Section 3.5.

Let $\mathcal{D}$ be a directed subset of $Adv(\mathcal{A})$. Given $\pi \in Path^{<\omega}(\mathcal{A})$, $a \in Act$ and $\mu \in \text{Distr}(S_\mathcal{A})$, define $\widehat{E}(\pi)(a, \mu) := \bigvee_{E \in \mathcal{D}} E(\pi)(a, \mu)$. In other words, $\widehat{E}$ is the pointwise join of $\mathcal{D}$ in the function space $Path^{<\omega}(\mathcal{A}) \times Act \times \text{Distr}(S_\mathcal{A}) \longrightarrow [0, 1]$. Our task is to show that $\widehat{E}$ is an adversary.

Notice that $\widehat{E}(\pi)$ assigns non-zero probability to $\langle a, \mu \rangle$ if and only if some E in $\mathcal{D}$ does. Hence

$$\langle a, \mu \rangle \in \text{supp}(\widehat{E}(\pi)) \Rightarrow \exists E \in \mathcal{D}, \langle a, \mu \rangle \in \text{supp}(E(\pi)) \Rightarrow last(\pi) \xrightarrow{a} \mu.$$

Fix $\pi \in Path^{<\omega}(\mathcal{A})$; we need to show $\widehat{E}(\pi)$ is a sub-distribution. By the countable branching assumption, we may choose a countable subset X_π of $Act \times \text{Distr}(S_\mathcal{A})$ such that $E(\pi)$ is a sub-distribution over X_π for every adversary E. Since $\mathcal{D}$ is directed, the set $\{E(\pi) \mid E \in \mathcal{D}\}$ is also directed. We can now apply Corollary 3.14 to conclude that $\widehat{E}(\pi)$ is also a sub-distribution. This gives the following lemma.

Lemma 7.5. *For all finite executions π, $\widehat{E}(\pi)$ is a probabilistic sub-distribution over $Act \times \mathrm{Distr}(S_{\mathcal{A}})$.*

Hence the set $Adv(\mathcal{A})$ equipped with the flat ordering is a CPO.

Theorem 7.6. *For a countably branching probabilistic automaton $\mathcal{A}$, the set of adversaries for $\mathcal{A}$ forms a CPO.*

Proof. Apply Lemma 7.5 and take the everywhere-0 adversary to be the bottom element. □

7.3 Probabilistic Executions

Again we consider the *flat* ordering: given $Q_1, Q_2 \in ProbExec(\mathcal{A})$, we say that $Q_1 \leq_\flat Q_2$ if for all $\pi \in Path^{<\omega}(\mathcal{A})$, $Q_1(\pi) \neq 0$ implies $Q_1(\pi) = Q_2(\pi)$.

Let $\mathcal{D}$ be a directed subset of $ProbExec(\mathcal{A})$. We claim that the pointwise join of $\mathcal{D}$ in the function space $Path^{<\omega}(\mathcal{A}) \longrightarrow [0,1]$ is also a probabilistic execution. By Theorem 6.2, it suffices to show $\bigvee \mathcal{D}$ satisfies the three properties in Section 6.1.

Conditions (1) and (2) follow directly from the definition of pointwise joins. To verify Condition (3), we first apply Lemma 3.12 to obtain an increasing ω-chain $C = \{Q_i\}_{i \in \mathbb{N}} \subseteq \mathcal{D}$ such that $\bigvee C = \bigvee \mathcal{D}$.

Lemma 7.7. *The function $\bigvee C$ satisfies Condition (3).*

Proof. Since C is increasing, $\bigvee C(\pi) = \lim_{i\to\infty} Q_i(\pi)$ for all $\pi \in Path^{<\omega}(\mathcal{A})$. Suppose $\bigvee C(\pi) \neq 0$ and, by monotonicity, we may assume without loss of generality $Q_i(\pi) \neq 0$ for all i. For each $\langle a,\, \mu\rangle \in X_\pi$, fix $s_{a,\mu} \in \mathrm{supp}(\mu)$. Then

$$\begin{aligned}
&\sum_{\langle a,\,\mu\rangle \in X_\pi} \frac{\bigvee C(\pi a \mu s_{a,\mu})}{\bigvee C(\pi) \cdot \mu(s_{a,\mu})} \\
&= \sum_{\langle a,\,\mu\rangle \in X_\pi} \frac{\lim_{i\to\infty} Q_i(\pi a \mu s_{a,\mu})}{\lim_{i\to\infty} Q_i(\pi) \cdot \mu(s_{a,\mu})} \\
&= \sum_{\langle a,\,\mu\rangle \in X_\pi} \lim_{i\to\infty} \frac{Q_i(\pi a \mu s_{a,\mu})}{Q_i(\pi) \cdot \mu(s_{a,\mu})} && \text{non-zero denominator} \\
&= \bigvee_{F \in \mathcal{P}_{<\omega}(X_\pi)} \sum_{\langle a,\,\mu\rangle \in F} \lim_{i\to\infty} \frac{Q_i(\pi a \mu s_{a,\mu})}{Q_i(\pi) \cdot \mu(s_{a,\mu})} \\
&= \bigvee_{F \in \mathcal{P}_{<\omega}(X_\pi)} \lim_{i\to\infty} \sum_{\langle a,\,\mu\rangle \in F} \frac{Q_i(\pi a \mu s_{a,\mu})}{Q_i(\pi) \cdot \mu(s_{a,\mu})} && \text{finite sum} \\
&\leq 1
\end{aligned}$$

□

Therefore the set $ProbExec(\mathcal{A})$ equipped with the flat ordering is also a CPO.

Theorem 7.8. *For a countably branching probabilistic automaton $\mathcal{A}$, the set of probabilistic executions of $\mathcal{A}$ forms a CPO whose bottom element is that generated by the everywhere-halting adversary.*

7.4 Continuity of Operator $\mathbf{Q}$

Recall that $\mathbf{Q}$ is an operator from *Adv*$(\mathcal{A})$ to *ProbExec*$(\mathcal{A})$, both of which have a CPO structure. Naturally, we proceed with a proof that $\mathbf{Q}$ is continuous. In fact, $\mathbf{Q}$ is also *strict*, i.e., bottom preserving.

Lemma 7.9. *The operator $\mathbf{Q}$ is monotone.*

Proof. Let $E_1 \leq_\flat E_2$ be given. We show that $\mathbf{Q}_{E_1} \leq_\flat \mathbf{Q}_{E_2}$, by induction on the length of execution π. The base case is trivial. Take an execution π' of the form $\pi a \mu s$ and assume $\mathbf{Q}_{E_1}(\pi') \neq 0$. Then $\mathbf{Q}_{E_1}(\pi) \neq 0$; applying I.H., we have $\mathbf{Q}_{E_1}(\pi) = \mathbf{Q}_{E_2}(\pi)$. On the other hand, we have $E_1(\pi)(a,\mu) \neq 0$, thus $E_1(\pi)(a,\mu) = E_2(\pi)(a,\mu)$. Hence

$$\mathbf{Q}_{E_1}(\pi') = \mathbf{Q}_{E_1}(\pi) \cdot E_1(\pi)(a,\mu) \cdot \mu(s) = \mathbf{Q}_{E_2}(\pi) \cdot E_2(\pi)(a,\mu) \cdot \mu(s) = \mathbf{Q}_{E_2}(\pi').$$

□

Lemma 7.10. *Let $\mathcal{D}$ be a directed set of adversaries. We have $\bigvee_{E \in \mathcal{D}} \mathbf{Q}_E = \mathbf{Q}_{\bigvee \mathcal{D}}$.*

Proof. Induction on the length of execution π. Since $\mathbf{Q}_E(s^0) = 1$ for every adversary E, the base case is trivial. For the inductive step, take an execution of the form $\pi a \mu s$ and let $\widehat{E}$ denote $\bigvee \mathcal{D}$. The following holds:

$$\begin{aligned}
\mathbf{Q}_{\widehat{E}}(\pi a \mu s) &= \mathbf{Q}_{\widehat{E}}(\pi) \cdot \widehat{E}(\pi, a, \mu) \cdot \mu(s) \\
&= \bigvee_{E \in \mathcal{D}} \mathbf{Q}_E(\pi) \cdot \bigvee_{E' \in \mathcal{D}} E'(\pi, a, \mu) \cdot \mu(s) && \text{I.H. and definition } \widehat{E} \\
&= \bigvee_{E, E' \in \mathcal{D}} \mathbf{Q}_E(\pi) \cdot E'(\pi, a, \mu) \cdot \mu(s) \\
&= \bigvee_{E \in \mathcal{D}} \mathbf{Q}_E(\pi) \cdot E(\pi, a, \mu) \cdot \mu(s) && \mathcal{D} \text{ directed and Lemma 7.9} \\
&= \bigvee_{E \in \mathcal{D}} \mathbf{Q}_E(\pi a \mu s).
\end{aligned}$$

□

Theorem 7.11. *The map $\mathbf{Q} : Adv(\mathcal{A}) \to ProbExec(\mathcal{A})$ is strictly continuous.*

7.5 Trace Distributions

Finally, we treat the case of trace distributions. Define $\leq_\flat$ in exactly the same way: given $D_1, D_2 \in$ *TrDist*$(\mathcal{A})$, we say that $D_1 \leq_\flat D_2$ if for all $\beta \in Act^{<\omega}$, $D_1(\beta) \neq 0$ implies $D_1(\beta) = D_2(\beta)$.

First we show the join of an ω-chain of trace distributions is again a trace distribution. Let $\{E_i\}_{i \in \mathbb{N}}$ be a sequence of adversaries for $\mathcal{A}$ such that the set $\mathcal{C} := \{\mathbf{D}_{E_i} \mid i \in \mathbb{N}\}$ forms a chain. We need to find a adversary E such that $\mathbf{D}_E = \bigvee \mathcal{C}$. For convenience, let D_i denote $\mathbf{D}_{E_i}$ and let $\widehat{D}$ denote $\bigvee \mathcal{C}$.

Let $\{\pi_n\}_{n \in \mathbb{N}}$ be an enumeration of $Path^{<\omega}(\mathcal{A})$. We apply the construction of Section 6.3 to $\{E_i\}_{i \in \mathbb{N}}$ and $\{\pi_n\}_{n \in \mathbb{N}}$ to obtain a sequence $\{\{E^n_j\}_{j \in \mathbb{N}}\}_{n \in \mathbb{N}}$ of sequences of adversaries for $\mathcal{A}$ and $Q \in$ *ProbExec*$(\mathcal{A})$. We claim that the trace distribution associated with Q is precisely $\widehat{D}$, thus any adversary E inducing Q also induces $\widehat{D}$.

Lemma 7.12. *For all* $\beta \in Act^{<\omega}$, $\sum_{\pi \in Tr^{-1}(\beta)} Q(\pi) \leq \widehat{D}(\beta)$.

Proof. Let $\beta \in Act^{<\omega}$ be given. Let S be the set of n such that $Tr(\pi_n) = \beta$. It suffices to prove for all finite $Y \subseteq S$, $\sum_{n \in Y} Q(\pi_n) \leq \widehat{D}(\beta)$.

Let $N := \max(Y)$. By definition of Q and Corollary 6.8, we have

$$Q(\pi_n) = \lim_{j \to \infty} \mathbf{Q}_{E_j^{n+1}}(\pi_n) = \lim_{j \to \infty} \mathbf{Q}_{E_j^{N+1}}(\pi_n).$$

Thus, moving the finite sum into the limit, we have

$$\sum_{n \in Y} Q(\pi_n) = \sum_{n \in Y} \lim_{j \to \infty} \mathbf{Q}_{E_j^{N+1}}(\pi_n) = \lim_{j \to \infty} \sum_{n \in Y} \mathbf{Q}_{E_j^{N+1}}(\pi_n).$$

For each $j \in \mathbb{N}$, we have $\sum_{n \in Y} \mathbf{Q}_{E_j^{N+1}}(\pi_n) \leq \mathbf{D}_{E_j^{N+1}}(\beta) \leq \widehat{D}(\beta)$, hence the limit is also below $\widehat{D}(\beta)$. □

Lemma 7.13. *For all* $\beta \in Act^{<\omega}$, $\sum_{\pi \in Tr^{-1}(\beta)} Q(\pi) \geq \widehat{D}(\beta)$.

Proof. Let $\beta \in Act^{<\omega}$ be given. Without loss of generality, assume that $\widehat{D}(\beta) \neq 0$. It suffices to show, for arbitrary $0 < \epsilon < \widehat{D}(\beta)$, $\sum_{\pi \in Tr^{-1}(\beta)} Q(\pi) \geq \widehat{D}(\beta) - \epsilon$. Let such ϵ be given. By Lemma 7.4, choose finite $F \subseteq Tr^{-1}(\beta)$ such that for all $i \in \mathbb{N}$, $D_i(\beta) - \sum_{\pi \in F} \mathbf{Q}_{E_i}(\pi) \leq \epsilon$.

Clearly, $\sum_{\pi \in Tr^{-1}(\beta)} Q(\pi) \geq \sum_{\pi \in F} Q(\pi)$. We will prove that the latter is greater than or equal to $\widehat{D}(\beta) - \epsilon$. Since F is finite, we may choose $N \in \mathbb{N}$ such that $F \subseteq \{\pi_0, \ldots, \pi_N\}$. Now we have

$$\begin{aligned}
\sum_{\pi \in F} Q(\pi) = \sum_{\{n \mid \pi_n \in F\}} Q(\pi_n) &= \sum_{\{n \mid \pi_n \in F\}} \lim_{j \to \infty} \mathbf{Q}_{E_j^{n+1}}(\pi_n) && \\
&= \sum_{\{n \mid \pi_n \in F\}} \lim_{j \to \infty} \mathbf{Q}_{E_j^{N+1}}(\pi_n) && \text{Lemma 6.7} \\
&= \lim_{j \to \infty} \sum_{\{n \mid \pi_n \in F\}} \mathbf{Q}_{E_j^{N+1}}(\pi_n) && F \text{ finite} \\
&\geq \lim_{j \to \infty} (D_{index(E_j^{N+1})}(\beta) - \epsilon) && \text{choice of } F \\
&= (\lim_{j \to \infty} D_{index(E_j^{N+1})}(\beta)) - \epsilon && \mathcal{C} \text{ increasing chain} \\
&= \widehat{D}(\beta) - \epsilon
\end{aligned}$$

□

Corollary 7.14. *For all* $\beta \in Act^{<\omega}$, $\sum_{\pi \in Tr^{-1}(\beta)} Q(\pi) = \widehat{D}(\beta)$.

The following lemma summarizes the results we have obtained so far.

Lemma 7.15. *Let* $\mathcal{C}$ *be an increasing* ω*-chain of trace distributions of an image finite probabilistic automaton* $\mathcal{A}$. *Then* $\bigvee \mathcal{C}$ *is also a trace distribution of* $\mathcal{A}$.

Theorem 7.16. *Let* $\mathcal{D}$ *be an arbitrary directed subset of* $TrDist(\mathcal{A})$ *for an image finite probabilistic automaton* $\mathcal{A}$. *Then* $\bigvee \mathcal{D}$ *is also a trace distribution of* $\mathcal{A}$.

Proof. By Lemma 7.15 and Lemma 3.12. □

Corollary 7.17. *Given an image finite probabilistic automaton $\mathcal{A}$, $\mathit{TrDist}(\mathcal{A})$ is a CPO whose bottom element is generated by the everywhere-halting adversary.*

Recall that the trace function $\mathit{Tr} : \mathit{Path}^{<\omega}(\mathcal{A}) \to \mathit{Act}^{<\omega}$ induces a map $\mathit{Tr} : \mathit{ProbExec}(\mathcal{A}) \to \mathit{TrDist}(\mathcal{A})$. The example below shows that this map is not continuous.

Example 7.18. Consider an automaton with two outgoing a transitions from the initial state. Let Q_1 be a probabilistic execution that with probability $\frac{1}{2}$ does the first transition and with probability $\frac{1}{2}$ halts. Let Q_2 be a probabilistic execution that does the first transition with probability $\frac{1}{2}$ and the second transition with probability $\frac{1}{2}$. Then $Q_1 \leq Q_2$. However, it is not the case that $\mathit{Tr}(Q_1) \leq \mathit{Tr}(Q_2)$, since $\mathit{Tr}(Q_1)(a) = \frac{1}{2} \neq 1 = \mathit{Tr}(Q_2)(a)$. Therefore Tr is not monotone.

7.6 Algebraicity

In Segala's proposal of the Approximation Induction Principle [Segala 1996], trace distributions are ordered pointwise by the usual ordering on $\mathbb{R}$, rather than our flat ordering. In fact, this alternative also gives rise to a CPO on $\mathit{TrDist}(\mathcal{A})$, but the resulting structure is not algebraic.

Example 7.19. Consider an automaton with a single a-transition and an adversary E that assigns probability 1 to that transition. Consider the sequence $E_0, E_1, \ldots$ of adversaries where each E_k chooses the a-transition with probability $1 - 2^{-k}$ and halts with probability 2^{-k}. Clearly, this infinite sequence converges monotonically to E under Segala's ordering; yet $E \neq E_k$ for all k. Therefore E is not a compact element. Similarly, one can show that every non-trivial trace distribution is *not* compact.

We now give a proof that $\mathit{TrDist}(\mathcal{A})$ forms an algebraic CPO under our flat ordering. (In fact, the same holds for $\mathit{Adv}(\mathcal{A})$ and $\mathit{ProbExec}(\mathcal{A})$, but the characterizations of compact elements are different.) Recall from Section 4.3 the definition of finite trace distributions. Essentially, $\mathbf{D}_E$ is finite if it assigns zero probability to all but a finite number of traces. The following lemma says that all finite trace distributions are compact in the CPO $\langle \mathit{TrDist}(\mathcal{A}), \leq_\flat \rangle$.

Lemma 7.20. *Let $\mathbf{D}_E$ be a finite trace distribution and let $\mathcal{D}$ be a directed set of trace distributions such that $\mathbf{D}_E \leq_\flat \bigvee \mathcal{D}$. Then there exists adversary E' with $\mathbf{D}_{E'} \in \mathcal{D}$ and $\mathbf{D}_E \leq_\flat \mathbf{D}_{E'}$.*

Proof. Let F denote the finite set of traces $\{\beta \in \mathit{Act}^{<\omega} \mid \mathbf{D}_E(\beta) \neq 0\}$. For each $\beta \in F$, choose E_β with $\mathbf{D}_{E_\beta} \in \mathcal{D}$ and $\mathbf{D}_{E_\beta}(\beta) = \mathbf{D}_E(\beta)$. This is possible by the definitions of $\leq_\flat$ and $\bigvee$. Since $\mathcal{D}$ is directed and F is finite, we may choose E' such that $\mathbf{D}_{E'}$ is in $\mathcal{D}$ and is an upperbound of $\{\mathbf{D}_{E_\beta} \mid \beta \in F\}$. Clearly $\mathbf{D}_E \leq_\flat \mathbf{D}_{E'}$. □

Lemma 7.21. *Let E be an adversary for $\mathcal{A}$ with $\mathbf{D}_E$ not finite. There exists a directed set $\mathcal{D}$ of trace distributions of $\mathcal{A}$ such that $\mathbf{D}_E = \bigvee \mathcal{D}$ and yet $\mathbf{D}_{E'} < \mathbf{D}_E$ for all $\mathbf{D}_{E'} \in \mathcal{D}$.*

Proof. Let $\{\beta_0, \beta_1, \ldots\}$ be a prefix-preserving enumeration of $\mathit{Act}^{<\omega}$. That is, if β_m is a prefix of β_n, then $m \leq n$. This is always possible for the set of finite words over a countable alphabet.

For each $n \in \mathbb{N}$, construct an adversary E_n as follows: for all π, a and μ,

—$E_n(\pi)(a, \mu) = E(\pi)(a, \mu)$ if $Tr(\pi)a$ is in $\beta_0, \dots, \beta_n$;

—$E_n(\pi)(a, \mu) = 0$ otherwise.

Informally, each E_n makes the same decisions as E until it reaches a trace not in $\beta_0, \dots, \beta_n$, at which point it halts. Since $\{\beta_n\}_{n \in \mathbb{N}}$ preserves prefix, it is easy to verify that $\{\mathbf{D}_{E_n} \mid n \in \mathbb{N}\}$ satisfies these two conditions:

—for all $m \leq n$, $\mathbf{D}_{E_n}(\beta_m) = \mathbf{D}_E(\beta_m)$;

—for all $m > n$, $\mathbf{D}_{E_n}(\beta_m) = 0$.

Clearly, each $\mathbf{D}_{E_n}$ is finite. Since $\mathbf{D}_E$ is infinite, we have $\mathbf{D}_{E_n} < \mathbf{D}_E$ for all $n \in \mathbb{N}$. Also $\{\mathbf{D}_{E_n} \mid n \in \mathbb{N}\}$ is an increasing chain whose limit is precisely $\mathbf{D}_E$, hence $\mathbf{D}_E$ must not be compact. □

LEMMA 7.22. *Let E be an adversary of $\mathcal{A}$. Let K_E denote the set of compact elements below $\mathbf{D}_E$, i.e., $K_E := \{\mathbf{D}_{E'} \mid \mathbf{D}_{E'}$ finite and $\mathbf{D}_{E'} \leq_\flat \mathbf{D}_E\}$. Then K_E is directed and $\mathbf{D}_E = \bigvee K_E$.*

PROOF. Again we make use of the prefix-preserving enumeration $\{\beta_n\}_{n \in \mathbb{N}}$. Take $\{E_n\}_{n \in \mathbb{N}}$ as in the proof of Lemma 7.21. Given a finite subset F of K_E, we can find $N \in \mathbb{N}$ such that for all $\mathbf{D}_{E'} \in F$ and $n \geq N$, $\mathbf{D}_{E'}(\beta_n) = 0$. This is because F is finite and each $\mathbf{D}_{E'}$ is finite. Then $\mathbf{D}_{E_N}$ is an upperbound of F. Moreover, $\mathbf{D}_{E_N}$ is finite, hence in K_E. This shows K_E is directed.

Finally, by the definition of $\leq_\flat$, we have for all n:

$$\bigvee K_E(\beta_n) = \mathbf{D}_{E_n}(\beta_n) = \mathbf{D}_E(\beta_n).$$

□

THEOREM 7.23. *Given an image finite probabilistic automaton $\mathcal{A}$, the structure $\langle TrDist(\mathcal{A}), \leq_\flat \rangle$ is an algebraic CPO and the compact elements are precisely the finite trace distributions.*

8. BOUNDEDNESS AND CONVERGENCE

The main result we establish in this section is that $TrDist(\mathcal{A}, k, l)$ forms a closed set in the metric space $[0,1]^{Act^{<\omega}}$, where $dist(\vec{u}, \vec{v}) := \sup_{\beta \in Act^{<\omega}} |u_\beta - v_\beta|$.

Let $\{E_i\}_{i \in \mathbb{N}}$ be a sequence of adversaries for elements of $TrDist(\mathcal{A}, k, l)$ for some $k, l \in \mathbb{N}$. For convenience, we write D_i for $\mathbf{D}_{E_i}$, the trace distribution generated by E_i. Each D_i can be viewed as a point in the metric space $[0,1]^{Act^{<\omega}}$. We say that $\{E_i\}_{i \in \mathbb{N}}$ is a *trace convergent* sequence of adversaries whenever $\{D_i\}_{i \in \mathbb{N}}$ is a convergent sequence in the space $[0,1]^{Act^{<\omega}}$. That is, there exists $D \in [0,1]^{Act^{<\omega}}$ such that

$$\forall \epsilon \; \exists N \; \forall i \geq N \; dist(D_i, D) \leq \epsilon.$$

Equivalently, we have

$$\forall \epsilon \; \exists N \; \forall i \geq N \; \forall \beta \in Act^{<\omega} \; |D_i(\beta) - D(\beta)| \leq \epsilon.$$

We claim that D is also a trace distribution (i.e., there is an adversary E such that $\mathbf{D}_E = D$). In particular, let E be the adversary constructed from $\{E_i\}_{i \in \mathbb{N}}$ by the procedure described in Section 6.3. We will show that $\mathbf{D}_E$ is in fact the limit of $\{D_i\}_{i\in\mathbb{N}}$.

First we prove a modification of Lemma 7.4. We restrict our attention to adversaries from $Adv(\mathcal{A}, -, l)$ and strengthen the conclusion to the existence of a uniform bound for all $\beta \in Act^{\leq k}$.

LEMMA 8.1. *Let $k, l \in \mathbb{N}$ and $\epsilon > 0$ be given. There exists finite, non-empty $P_{k,\epsilon} \subseteq Path^{\leq k}(\mathcal{A})$ such that for all $E \in Adv(\mathcal{A}, -, l)$ and for all $\beta \in Act^{\leq k}$, $\sum_{\pi\in Tr^{-1}(\beta)\setminus P_{k,\epsilon}} \mathbf{Q}_E(\pi) \leq \epsilon$.*

PROOF. We proceed by induction on k. For every ϵ, take $P_{0,\epsilon}$ to be the singleton $\{s^0\}$. Now suppose the claim holds for k. Let $\epsilon > 0$ be given and choose a finite, nonempty set $P_{k,\frac{\epsilon}{2}}$ as stated. Let $m > 0$ be its cardinality. Consider the set

$$S := \bigcup_{|\pi|=k, \pi\in P_{k,\frac{\epsilon}{2}}} \{last(\pi) \xrightarrow{a} \mu \mid a \in Act_l\}.$$

Since $\mathcal{A}$ is image finite, S is a finite union of finite sets, hence also finite. If S is empty, set $P_{k+1,\epsilon}$ to be $P_{k,\frac{\epsilon}{2}}$. Otherwise, let $n > 0$ be its cardinality. For each μ occurring in S, choose a finite set $X_\mu \subseteq \mathrm{supp}(\mu)$ such that

$$\sum_{s\in\mathrm{supp}(\mu)\setminus X_\mu} \mu(s) \leq \frac{\epsilon}{2mn}.$$

Then set $P_{k+1,\epsilon}$ to be $P_{k,\frac{\epsilon}{2}} \cup \{\pi a \mu s \mid (last(\pi) \xrightarrow{a} \mu) \in S \text{ and } s \in X_\mu\}$. We will prove that $P_{k+1,\epsilon}$ satisfies the desired condition.

Let $E \in Adv(\mathcal{A}, -, l)$ and $\beta \in Act^{\leq k+1}$ be given. Notice that, if β contains a symbol not in Act_l, then $\mathbf{Q}_E(\pi) = 0$ for all $\pi \in Tr^{-1}(\beta)$. Thus we may assume that $\beta \in (Act_l)^{\leq k+1}$. Moreover, if β has length at most k, then $Tr^{-1}(\beta) \setminus P_{k+1,\epsilon} = Tr^{-1}(\beta) \setminus P_{k,\frac{\epsilon}{2}}$. This is because every path $\pi \in P_{k+1,\epsilon} \setminus P_{k,\frac{\epsilon}{2}}$ (if it exists) must have length $k + 1$. Therefore, we have

$$\sum_{\pi\in Tr^{-1}(\beta)\setminus P_{k+1,\epsilon}} \mathbf{Q}_E(\pi) = \sum_{\pi\in Tr^{-1}(\beta)\setminus P_{k,\frac{\epsilon}{2}}} \mathbf{Q}_E(\pi) \leq \frac{\epsilon}{2} \leq \epsilon.$$

Now we focus on the case in which $\beta \in (Act_l)^{k+1}$. Suppose β is of the form $\beta' a$. We partition $Y := Tr^{-1}(\beta) \setminus P_{k+1,\epsilon}$ into two sets:

$$Y_0 := \{\pi a \mu s \in Y \mid \pi \notin P_{k,\frac{\epsilon}{2}}\},$$

$$Y_1 := \{\pi a \mu s \in Y \mid \pi \in P_{k,\frac{\epsilon}{2}} \text{ and } s \notin X_\mu\}.$$

Then by Lemma 7.2 and the induction hypothesis, we have

$$\sum_{\pi\in Y_0} \mathbf{Q}_E(\pi) \leq \sum_{\pi\in Tr^{-1}(\beta')\setminus P_{k,\frac{\epsilon}{2}}} \mathbf{Q}_E(\pi) \leq \frac{\epsilon}{2}.$$

On the other hand,

$$\begin{aligned}\sum_{\pi\in Y_1}\mathbf{Q}_E(\pi) &= \sum_{\pi a\mu s\in Y_1}\mathbf{Q}_E(\pi)\cdot E(\pi)(a,\mu)\cdot\mu(s)\\ &\leq \sum_{\pi a\mu s\in Y_1}\mu(s)\\ &\leq \sum_{\pi\in P_{k,\frac{\epsilon}{2}}}\ \sum_{last(\pi)\xrightarrow{a}\mu\in S}\ \sum_{s\in\mathrm{supp}(\mu)\setminus X_\mu}\mu(s)\\ &\leq m\cdot n\cdot\sum_{s\in\mathrm{supp}(\mu)\setminus X_\mu}\mu(s)\\ &\leq m\cdot n\cdot\frac{\epsilon}{2mn}=\frac{\epsilon}{2}.\end{aligned}$$

Therefore,

$$\sum_{\pi\in Tr^{-1}(\beta)\setminus P_{k+1,\epsilon}}\mathbf{Q}_E(\pi)=\sum_{\pi\in Y_0}\mathbf{Q}_E(\pi)+\sum_{\pi\in Y_1}\mathbf{Q}_E(\pi)\leq\frac{\epsilon}{2}+\frac{\epsilon}{2}=\epsilon.$$

□

Lemma 8.2. *Let $\mathcal{A}$ be an image finite probabilistic automaton and let $k,l\in\mathbb{N}$ be given. Let $\{E_i\}_{i\in\mathbb{N}}$ be a sequence of trace convergent adversaries from $Adv(\mathcal{A},k,l)$ and write D_i for $\mathbf{D}_{E_i}$. Let E be constructed as in Section 6.3. Then $\mathbf{D}_E$ is the limit of $\{D_i\}_{i\in\mathbb{N}}$ in the space $[0,1]^{Act^{<\omega}}$. That is,*

$$\forall\epsilon\ \exists N\ \forall i>N\ \forall\beta\in Act^{<\omega}\ |D_i(\beta)-\mathbf{D}_E(\beta)|\leq\epsilon.$$

Proof. First note that, for all $\beta\notin(Act_l)^{\leq k}$ and $i\in\mathbb{N}$, we have $D_i(\beta)=0=\mathbf{D}_E(\beta)$. Hence we may focus on traces in $(Act_l)^{\leq k}$. Let $\epsilon>0$ be given. Choose finite, non-empty $P_{k,\frac{\epsilon}{3}}$ as in Lemma 8.1 and let $m:=|P_{k,\frac{\epsilon}{3}}|$. Moreover, by trace convergence of $\{E_i\}_{i\in\mathbb{N}}$, we may choose M_0 such that for all $i,j>M_0$, $dist(D_i,D_j)<\frac{\epsilon}{3}$.

Recall from Section 6.3 that we have an enumeration $\{\pi_n\}_{n\in\mathbb{N}}$ of $Path^{<\omega}(\mathcal{A})$. Let $M:=\max\{n\mid\pi_n\in P_{k,\frac{\epsilon}{3}}\}+1$. Then by Corollary 6.8, we have

$$\forall\pi\in P_{k,\frac{\epsilon}{3}}\lim_{j\to\infty}\mathbf{Q}_{E_j^M}(\pi)=\mathbf{Q}_E(\pi).$$

For each $\pi\in P_{k,\frac{\epsilon}{3}}$, choose j_π such that

$$\forall j>j_\pi\ |\mathbf{Q}_{E_j^M}(\pi)-\mathbf{Q}_E(\pi)|<\frac{\epsilon}{3m}.$$

Let L be the least number such that $L>\max\{j_\pi\mid\pi\in P_{k,\frac{\epsilon}{3}}\}$ and $index(E_L^M)>M_0$. Take $N:=index(E_L^M)$. Write Y_0 for $Tr^{-1}(\beta)\cap P_{k,\frac{\epsilon}{3}}$ and Y_1 for $Tr^{-1}(\beta)\setminus P_{k,\frac{\epsilon}{3}}$. Then for all $i > N$ and $\beta \in (Act_l)^{\leq k}$,

$$\begin{aligned}
&|D_i(\beta) - \mathbf{D}_E(\beta)| \\
&\leq |D_i(\beta) - D_N(\beta)| + |D_N(\beta) - \mathbf{D}_E(\beta)| \\
&\leq \frac{\epsilon}{3} + |\sum_{\pi \in Tr^{-1}(\beta)} \mathbf{Q}_{E_L^M}(\pi) - \sum_{\pi \in Tr^{-1}(\beta)} \mathbf{Q}_E(\pi)| \\
&\leq \frac{\epsilon}{3} + |\sum_{\pi \in Y_0} \mathbf{Q}_{E_L^M}(\pi) - \sum_{\pi \in Y_0} \mathbf{Q}_E(\pi) + \sum_{\pi \in Y_1} \mathbf{Q}_{E_L^M}(\pi) - \sum_{\pi \in Y_1} \mathbf{Q}_E(\pi)| \\
&\leq \frac{\epsilon}{3} + \sum_{\pi \in Y_0} |\mathbf{Q}_{E_L^M}(\pi) - \mathbf{Q}_E(\pi)| + |\sum_{\pi \in Y_1} \mathbf{Q}_{E_L^M}(\pi) - \sum_{\pi \in Y_1} \mathbf{Q}_E(\pi)| \\
&\leq \frac{\epsilon}{3} + m \cdot \frac{\epsilon}{3m} + \frac{\epsilon}{3} = \epsilon.
\end{aligned}$$

□

Corollary 8.3. *For all $k, l \in \mathbb{N}$, the set $TrDist(\mathcal{A}, k, l)$ is a closed subset of $[0,1]^{Act^{<\omega}}$.*

Next we prove the analogous result for induced probability distributions (as defined in Section 5.1).

Lemma 8.4. *Let $\{P_i\}_{i \in \mathbb{N}} \subseteq \{\mathbf{P}_{D,k} \mid D \in TrDist(\mathcal{A}, k, l)\}$ be a convergent sequence in $Act^{<\omega}$ with limit point P. Then P is a discrete distribution on $Act^{<\omega}$.*

Proof. Clearly, $P[\beta] = 0$ for all $\beta \notin (Act_l)^{\leq k}$. On the other hand, since $(Act_l)^{\leq k}$ is a finite set, we have

$$\sum_{\beta \in (Act_l)^{\leq k}} P[\beta] = \sum_{\beta \in (Act_l)^{\leq k}} \lim_{i \to \infty} P_i[\beta] = \lim_{i \to \infty} \sum_{\beta \in (Act_l)^{\leq k}} P_i[\beta] = 1.$$

□

Lemma 8.5. *Let $k, l \in \mathbb{N}$ and $\{P_i\}_{i \in \mathbb{N}} \subseteq \{\mathbf{P}_{D,k} \mid D \in TrDist(\mathcal{A}, k, l)\}$ be given. Suppose $\{P_i\}_{i \in \mathbb{N}}$ is a convergent sequence in $Act^{<\omega}$ with limit point P. For each i, choose D_i so that $P_i = \mathbf{P}_{D_i,k}$. Then $\{D_i\}_{i \in \mathbb{N}}$ is also a convergent sequence in $Act^{<\omega}$. Moreover, $P = \mathbf{P}_{D,k}$, where D is the limit of $\{D_i\}_{i \in \mathbb{N}}$.*

Proof. Recall from Lemma 5.2 that for each $i \in \mathbb{N}$ and $\beta \in (Act_l)^{\leq k}$, we have

$$D_i(\beta) = \sum_{\beta \sqsubseteq \beta'; \beta' \in (Act_l)^{\leq k}} P_i[\beta'].$$

Define D from P with the same formula. Notice that this is a finite sum, therefore D is the limit of $\{D_i\}_{i \in \mathbb{N}}$. □

Corollary 8.6. *For all $k, l \in \mathbb{N}$, the set $\{\mathbf{P}_{D,k} \mid D \in TrDist(\mathcal{A}, k, l)\}$ is also a closed subset of $[0,1]^{Act^{<\omega}}$.*

Proof. By Corollary 8.3 and Lemma 8.5. □

9. THE CHARACTERIZATION THEOREM

Let us briefly recapitulate our development. Our goal is to show that the testing preorder defined in Section 5.4 coincides with trace distribution inclusion, as defined in Section 4.2. In Section 7.6, we established that the set of trace distributions of an image finite automaton forms an algebraic CPO. Therefore the following are equivalent for image finite automata $\mathcal{A}$ and $\mathcal{B}$:

—$\mathcal{A} \sqsubseteq_{\mathrm{TD}} \mathcal{B}$;
—for all $k, l \in \mathbb{N}$, $\mathcal{A} \sqsubseteq_{\mathrm{TD}}^{k,l} \mathcal{B}$.

By virtue of this observation, it suffices to prove the following finitary characterization theorem.

THEOREM 9.1. *Let $\mathcal{A}$ and $\mathcal{B}$ be image finite probabilistic automata. Let $\alpha \in (0,1)$ and $k, l \in \mathbb{N}$ be given. We have $\mathit{TrDist}(\mathcal{A},k,l) \subseteq \mathit{TrDist}(\mathcal{B},k,l)$ if and only if, for all m, $\mathit{Obs}(\mathcal{A},k,l,m,\alpha) \subseteq \mathit{Obs}(\mathcal{B},k,l,m,\alpha)$.*

Since $\mathit{Obs}(\mathcal{A},k,l,m,\alpha)$ is entirely defined in terms of $\mathit{TrDist}(\mathcal{A},k,l)$ and parameters k, l, m and α, the "only if" direction of Theorem 9.1 is trivial. For the converse, we assume there is $D \in \mathit{TrDist}(\mathcal{A},k,l) \setminus \mathit{TrDist}(\mathcal{B},k,l)$ and our goal is to find $m \in \mathbb{N}$ and a sample $O \in \mathit{Obs}(\mathcal{A},k,l,m,\alpha) \setminus \mathit{Obs}(\mathcal{B},k,l,m,\alpha)$.

Intuitively, we obtain such O by running the trace distribution machine repeatedly under D. For each $m \in \mathbb{N}$, let D^m denote the length-m sequence in which every element is D. Recall from Section 5.4 that an outcome is acceptable if its frequency vector deviates minimally from the expected frequency vector. Our claim is, as the number of trials increases, the amount of deviation allowed decreases to 0. In other words, given any small $\delta > 0$, we can find $m \in \mathbb{N}$ such that any acceptable outcome of a width-m experiment must have a frequency vector within distance δ of the expectation. This claim, together with the fact that we can always separate the point $\mathbf{P}_{D^m,k}$ from the set $\{\mathbf{P}_{\vec{K},k} \mid \vec{K} \in \mathit{TrDist}(\mathcal{B},k,l)\}$ (Corollary 8.6), allows us to distinguish acceptable outcomes of D^m from those generated by trace distributions in $\mathit{TrDist}(\mathcal{B},k,l)$.

Before presenting the formal proofs, let us further motivate our approach by considering again the coin-flipping example. Suppose $\mathcal{A}$ is the fair coin and we conduct 100 experiments on $\mathcal{A}$. In this case, every outcome is just as likely as every other outcome. Yet a frequency vector close to $\langle 0.5, 0.5\rangle$ (for example $\langle 0.49, 0.51\rangle$) is much more likely to be observed than a frequency vector far away from $\langle 0.5, 0.5\rangle$ (for example $\langle 0.01, 0.99\rangle$). This is because there are many more outcomes with frequency $\langle 0.49, 0.51\rangle$ than there are outcomes with $\langle 0.01, 0.99\rangle$. As we increase the number of trials, this clustering effect intensifies and the probability of observing a frequency vector with large deviation becomes very small.

This simple idea also applies in the case of m independent coin flips, where each coin may have a different bias. This is formalized in the following lemma, which is an analog of the weak law of large numbers for independent Bernoulli variables.

LEMMA 9.2. *Let $\alpha \in (0,1)$ and $\delta > 0$ be given. There exists $M \in \mathbb{N}$ such that for all $m \geq M$ and sequences $X_1, \ldots, X_m$ of independent Bernoulli variables,*

$$\mathbf{P}[|Z - \mathbf{E}Z| \geq \delta] \leq \alpha,$$

where $Z = \frac{1}{m}\sum_{i=1}^{m} X_i$ *represents the success frequency in these* m *trials.*

Proof. Take $M \geq \frac{1}{4\delta^2\alpha}$ and let $m, X_1, \ldots, X_m$ be given as stated. Assume that each Bernoulli variable X_i has parameter $p_i \in [0,1]$. First note that for all $p \in [0,1]$, $p(1-p) \leq \frac{1}{4}$. Then

$$\begin{aligned}\mathrm{Var}[Z] &= \mathrm{Var}[\frac{1}{m}\sum_{i=1}^{m} X_i] = \frac{1}{m^2}\sum_{i=1}^{m}\mathrm{Var}[X_i]\\ &= \frac{1}{m^2}\sum_{i=1}^{m} p_i(1-p_i) \leq \frac{1}{m^2}\sum_{i=1}^{m}\frac{1}{4} = \frac{1}{4m}.\end{aligned}$$

By Chebychev's inequality (Theorem 3.8), we have

$$\mathbf{P}[|Z - \mathbf{E}[Z]| \geq \delta] \leq \frac{1}{\delta^2}\mathrm{Var}[Z] \leq \frac{1}{\delta^2}\cdot\frac{1}{4m} \leq \frac{1}{\delta^2}\cdot\frac{1}{4M} \leq \frac{4\delta^2\alpha}{4\delta^2} = \alpha.$$

□

In our case, successes correspond to occurrences of a particular trace β: if the machine operates according to trace distributions $\vec{D}$, then each run i corresponds to a Bernoulli variable with parameter $\mathbf{P}_{D_i,k}[\beta]$ (see Section 5.2). Thus Lemma 9.2 gives the following corollary.

Corollary 9.3. *Given any* $\delta > 0$, *there exists* $M \in \mathbb{N}$ *such that for all* $m \geq M$, $\beta \in Act^{\leq k}$ *and sequences* $\vec{D}$ *of trace distributions in* $TrDist(\mathcal{A})$,

$$\mathbf{P}_{\vec{D},k}[\{O \in \mathcal{U} \mid |\mathit{freq}(O)(\beta) - \mathbf{E}_{\beta}^{\vec{D},k}| \geq \delta\}] \leq \alpha.$$

Now we consider all sequences $\beta \in (Act_l)^{\leq k}$ at the same time. This is where we must restrict to sequences over Act_l (rather than Act), since otherwise we are concerned with infinitely many β's.

Lemma 9.4. *Given any* $\delta > 0$, *there exists* $M \in \mathbb{N}$ *such that for all* $m \geq M$ *and sequences* $\vec{D}$ *of trace distributions in* $TrDist(\mathcal{A}, k, l)$,

$$\mathbf{P}_{\vec{D},k}[\mathit{freq}^{-1}(B_\delta(\mathbf{E}^{\vec{D},k}))] \geq 1 - \alpha.$$

Proof. Let n be the cardinality of $(Act_l)^{\leq k}$. By Corollary 9.3, we may choose M such that for all $m \geq M$, $\beta \in Act^{\leq k}$ and sequences $\vec{D}$ of trace distributions in $TrDist(\mathcal{A})$,

$$\mathbf{P}_{\vec{D},k}[\{O \in \mathcal{U} \mid |\mathit{freq}(O)(\beta) - \mathbf{E}_{\beta}^{\vec{D},k}| \geq \delta\}] \leq \frac{\alpha}{n}.$$

Then for all $m \geq M$ and sequences $\vec{D}$, we have

$$\begin{aligned}
&\mathbf{P}_{\vec{D},k}[\mathit{freq}^{\text{-}1}(B_\delta(\mathbf{E}^{\vec{D},k}))] \\
&= \mathbf{P}_{\vec{D},k}[\{O \in \mathcal{U} \mid \forall\beta\ |\mathit{freq}(O)(\beta) - \mathbf{E}^{\vec{D},k}_\beta| < \delta\}] && \text{definition of } \mathit{dist} \\
&= 1 - \mathbf{P}_{\vec{D},k}[\{O \in \mathcal{U} \mid \exists\beta\ |\mathit{freq}(O)(\beta) - \mathbf{E}^{\vec{D},k}_\beta| \geq \delta\}] \\
&\geq 1 - \sum_{\beta \in (Act_l)^{\leq k}} \mathbf{P}_{\vec{D},k}[\{O \in \mathcal{U} \mid |\mathit{freq}(O)(\beta) - \mathbf{E}^{\vec{D},k}_\beta| \geq \delta\}] \\
&\geq 1 - n\frac{\alpha}{n} = 1 - \alpha && \text{choice of } M
\end{aligned}$$

□

We are now ready for the proof of Theorem 9.1.

Proof Theorem 9.1. The "only if" direction is trivial. For the converse, assume there is $D \in \mathit{TrDist}(\mathcal{A}, k, l) \setminus \mathit{TrDist}(\mathcal{A}, k, l)$. Let δ denote the distance between the point $\mathbf{P}_{D^m,k}$ and the set $\{\frac{1}{m}\sum_0^{m-1}\mathbf{P}_{\vec{K},k} \mid \vec{K} \in \mathit{TrDist}(\mathcal{B}, k, l)\}$. By Corollaries 6.6 and 8.6, δ must be non-zero.

By Lemma 9.4, we can find $M_\mathcal{A}$ and $M_\mathcal{B}$ such that for all $m \geq \max(M_\mathcal{A}, M_\mathcal{B})$ and all sequences of trace distributions $\vec{K}$ in $\mathit{TrDist}(\mathcal{B}, k, l)$,

$$\begin{aligned}
\mathbf{P}_{D^m,k}[\mathit{freq}^{\text{-}1}(B_{\frac{\delta}{3}}(\mathbf{E}^{D^m,k}))] &\geq 1 - \frac{\alpha}{2} > 1 - \alpha \\
\mathbf{P}_{\vec{K},k}[\mathit{freq}^{\text{-}1}(B_{\frac{\delta}{3}}(\mathbf{E}^{\vec{K},k}))]]) &\geq 1 - \frac{\alpha}{2} > 1 - \alpha.
\end{aligned}$$

Therefore, we have

$$\mathit{Obs}(D^m, k, \alpha) \subseteq \mathit{freq}^{\text{-}1}(B_{\frac{\delta}{3}}(\mathbf{E}^{D^m,k})) = \mathit{freq}^{\text{-}1}(B_{\frac{\delta}{3}}\mathbf{P}_{D^m,k})$$

and, for all sequences $\vec{K}$ in $\mathit{TrDist}(\mathcal{B}, k, l)$,

$$\mathit{Obs}(\vec{K}, k, \alpha) \subseteq \mathit{freq}^{\text{-}1}(B_{\frac{\delta}{3}}(\mathbf{E}^{\vec{K},k})) = \mathit{freq}^{\text{-}1}(B_{\frac{\delta}{3}}(\sum_0^{m-1}\frac{1}{m}\mathbf{P}_{\vec{K},k})).$$

Since $\mathit{dist}(\mathbf{P}_{D^m,k}, \sum_0^{m-1}\frac{1}{m}\mathbf{P}_{\vec{K},k}) \geq \delta$, we have $B_{\frac{\delta}{3}}\mathbf{P}_{D^m,k} \cap B_{\frac{\delta}{3}}(\sum_0^{m-1}\frac{1}{m}\mathbf{P}_{\vec{K},k}) = \emptyset$. Therefore $\mathit{Obs}(D^m, k, \alpha) \not\subseteq \mathit{Obs}(\mathcal{B}, k, l, \alpha)$. □

Theorem 9.5. *Let $\mathcal{A}$ and $\mathcal{B}$ be image finite probabilistic automata and let $\alpha \in (0,1)$ be given. We have $\mathcal{A} \sqsubseteq_{\mathrm{TD}} \mathcal{B}$ if and only if $\mathcal{A} \leq_\alpha \mathcal{B}$.*

Proof. We have the following chain of equivalences:

$$\begin{aligned}
&\mathcal{A} \sqsubseteq_{\mathrm{TD}} \mathcal{B} \\
&\Leftrightarrow \mathcal{A} \sqsubseteq^{k,l}_{\mathrm{TD}} \mathcal{B} \text{ for all } k, l \in \mathbb{N} && \text{Theorem 7.23} \\
&\Leftrightarrow \mathit{Obs}(\mathcal{A}, k, l, m, \alpha) \subseteq \mathit{Obs}(\mathcal{B}, k, l, m, \alpha) \text{ for all } k, l, m \in \mathbb{N} && \text{Theorem 9.1} \\
&\Leftrightarrow \mathcal{A} \leq_\alpha \mathcal{B} && \text{definition of } \leq_\alpha
\end{aligned}$$

□

10. CONCLUDING DISCUSSIONS

The theory of stochastic processes finds many applications in the area of performance analysis of computer systems. In such applications, randomness is typically used to model uncertainties in the computation environment; for example, the arrival rate of jobs and processing time required for each job. We are then interested in calculating or estimating parameters such as expected waiting time and percentage of missed deadlines over a given period.

The model considered in this paper is developed in a different tradition, namely, the analysis of distributed algorithms. Here randomness is used by the processes themselves to achieve certain goals. For instance, processes cast randomly generated votes to reach consensus, or they choose a neighbor at random to propagate information. In this setting, the computation environment is extremely unpredictable and it does not always makes sense to assume a fixed pattern of events (e.g. exponential distribution on message delay). Nondeterminism is therefore a more reasonable alternative for modeling timing uncertainties. Moreover, nondeterministic choices are extremely useful in specification and verification, allowing us to abstract away from inessential temporal ordering of events.

This paper presents a first step in developing statistical testing techniques for systems with nondeterministic behavior. We show that, under some appropriate finiteness assumptions, nondeterministic choices are "harmless". The rationale behind this statement is that we can view a nondeterministic choice among events as a weighted sum of those events, but with unknown weights. Therefore the behavior of a process is represented by a convex closed set of distributions, rather than a single distribution. This retains many of the nice properties of purely probabilistic processes and we are able to use hypothesis tests to characterize an existing semantic equivalence.

We see much potential in applying our ideas to "black-box" verification, where we have little or no control over the system of interest. Given such a system, one can construct a probabilistic automaton as the test hypothesis and use samples generated from the actual system to either accept or reject the hypothesis. This method provides rigorous guarantees regarding error probabilities.

We define very simple hypothesis tests in this paper, because we do not have a special problem in mind and thus cannot make use of any domain knowledge. In practice, one can design more powerful tests (i.e., those that also control false positive errors) using specific properties of the distributions involved. Also, it may be sufficient to consider simple or one-sided hypotheses, for which standard methods exist for finding uniformly most powerful tests. (In contrast, our tests have composite and two-sided alternative hypotheses.)

REFERENCES

Aggarwal, S. 1994. Time optimal self-stabilizing spanning tree algorithms. M.S. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. Available as Technical Report MIT/LCS/TR-632.

Baeten, J., Bergstra, J., and Klop, J. 1987. On the consistency of Koomen's fair abstraction rule. *Theoretical Computer Science 51,* 1/2, 129–176.

Baier, C. and Kwiatkowska, M. 1998. Model checking for a probabilistic branching time logic with fairness. *Distributed Computing 11,* 3, 125–155.

Bergstra, J. and Klop, J. 1986. Verification of an alternating bit protocol by means of process algebra. In *Mathematical Methods of Specification and Synthesis of Software Systems '85, Math.* Mathematical Research, vol. 31. Akademie-Verlag, Berlin, 9–23.

Casella, G. and Berger, R. 1990. *Statistical Inference.* Duxbury Press, Belmont.

Christoff, I. 1990. Testing equivalence and fully abstract models of probabilistic processes. In *Proceedings CONCUR 90.* Lecture Notes in Computer Science, vol. 458. Springer-Verlag, New York, 126–140.

Cleaveland, R., Dayar, Z., Smolka, S., and Yuen, S. 1999. Testing preorders for probabilistic processes. *Information and Computation 154,* 2, 93–148.

Cohn, D. 1980. *Measure Theory.* Birkhäuser, Boston.

Davey, B. and Priestley, H. 1990. *Introduction to Lattices and Order.* Cambridge University Press, Cambridge.

Deng, Y., Glabbeek, R. v., Hennessy, M., Morgan, C., and Zhang, C. 2007a. Characterising testing preorders for finite probabilistic processes. In *LICS'07: Proceedings of the 22nd Annual IEEE Symposium on Logic in Computer Science.* IEEE Computer Society, 313–325.

Deng, Y., Glabbeek, R. v., Hennessy, M., Morgan, C., and Zhang, C. 2007b. Remarks on testing probabilistic processes. *Electronic Notes in Theoretical Computer Science 172*, 359–397.

Desharnais, J., Edalat, A., and Panangaden, P. 2002. Bisimulation for labeled Markov processes. *Information and Computation 179,* 2, 163–193.

Edalat, A. 1995. Domain theory in stochastic processes. In *Proceedings LICS 95.* IEEE Computer Society Press, Los Alamitos, CA, USA, 244–254.

Glabbeek, R. v. 2001. The linear time - branching time spectrum I. The semantics of concrete, sequential processes. In *Handbook of Process Algebra*, J. Bergstra, A. Ponse, and S. Smolka, Eds. North-Holland, 3–99.

Gregorio-Rodríguez, C. and Núñez, M. 1998. Denotational semantics for probabilistic refusal testing. In *Proceedings ProbMIV 98.* Electronic Notes in Theoretical Computer Science, vol. 22.

Jonsson, B. and Yi, W. 2002. Testing preorders for probabilistic processes can be characterized by simulations. *Theoretical Computer Science 282,* 1, 33–51.

Kolmogorov, A. and Fomin, S. 1970. *Introductory Real Analysis.* Dover Publications, Inc., New York.

Larsen, K. and Skou, A. 1991. Bisimulation through probabilistic testing. *Information and Computation 91*, 1–28.

Lynch, N., Saias, I., and Segala, R. 1994. Proving time bounds for randomized distributed algorithms. In *Proceedings of the 13th Annual ACM Symposium on the Principles of Distributed Computing.* 314–323.

Milner, R. 1980. *A Calculus of Communicating Systems.* Lecture Notes in Computer Science, vol. 92. Springer-Verlag.

Nicola, R. and Hennessy, M. 1984. Testing equivalences for processes. *Theoretical Computer Science 34*, 83–133.

Pogosyants, A., Segala, R., and Lynch, N. 2000. Verification of the randomized consensus algorithm of Aspnes and Herlihy: a case study. *Distributed Computing 13,* 3, 155–186.

Rudin, W. 1987. *Real and Complex Analysis.* McGraw-Hill, Inc., Boston.

Segala, R. 1995. Modeling and verification of randomized distributed real-time systems. Ph.D. thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology. Available as Technical Report MIT/LCS/TR-676.

Segala, R. 1996. Testing probabilistic automata. In *Proceedings CONCUR 96.* Lecture Notes in Computer Science, vol. 1119. Springer-Verlag, 299–314.

Segala, R. and Lynch, N. 1995. Probabilistic simulations for probabilistic processes. *Nordic Journal of Computing 2,* 2, 250–273.

Sen, K., Viswanathan, M., and Agha, G. 2004. Statistical model checking of black-box probabilistic systems. In *Computer-Aided Verification.* LNCS, vol. 3114. Springer-Verlag, 202–215.

Stoelinga, M. 2002. An introduction to probabilistic automata. *Bulletin of the European Association for Theoretical Computer Science 78*, 176–198.

Stoelinga, M. and Vaandrager, F. 1999. Root contention in IEEE 1394. In *Proceedings 5th International AMAST Workshop on Formal Methods for Real-Time and Probabilistic Systems*, J.-P. Katoen, Ed. Lecture Notes in Computer Science, vol. 1601. Springer-Verlag, 53–74.

Stoelinga, M. and Vaandrager, F. 2003. A testing scenario for probabilistic automata. In *Proceedings 30 ICALP*. Lecture Notes in Computer Science, vol. 2719. Springer-Verlag, 407–418.

Trivedi, K. 2002. *Probability and Statistics with Reliability, Queuing and Computer Science Applications*. John Wiley & Sons, Inc., New York.

Vatan, F. 2001. Distribution functions of probabilistic automata. In *Proceedings STOC 01*. 684–693.

Younes, H. 2005. Probabilistic verification for "black-box" systems. In *Proceedings CAV 2005*. 253–265.

Younes, H., Kwiatkowska, M., Norman, G., and Parker, D. 2004. Numerical vs. statistical probabilistic model checking: an empirical study. In *Tools and Algorithms for the Construction and Analysis of Systems*. Lecutre Notes in Computer Science, vol. 2988. Springer-Verlag, 46–60.

Younes, H. and Simmons, R. 2002. Probabilistic verification of discrete event systems using acceptance sampling. In *Computer-Aided Verification*. Lecture Notes in Computer Science, vol. 2404. Springer-Verlag, 223–235.

Received Month year; revised Month Year; accepted Month Year