A Testing Scenario for Probabilistic Automata*

Mariëlle Stoelinga[1] and Frits Vaandrager[2]

[1] Dept. of Computer Engineering, University of California, Santa Cruz
marielle@soe.ucsc.edu
[2] Nijmegen Institute for Computing and Information Sciences,
University of Nijmegen, The Netherlands
fvaan@cs.kun.nl

Abstract. Recently, a large number of equivalences for probabilistic automata has been proposed in the literature. Except for the probabilistic bisimulation of Larsen & Skou, none of these equivalences has been characterized in terms of intuitive testing scenarios. In our view, this is an undesirable situation: in the end, the behavior of an automaton is what an external observer perceives. In this paper, we propose and study a simple and intuitive testing scenario for probabilistic automata. We prove that the equivalence induced by this scenario coincides with the trace distribution equivalence proposed by Segala. A technical result that we need to establish on the way is an *Approximation Induction Principle (AIP)* for probabilistic processes.

AMS Subject Classification (1991): 68Q05, 68Q10, 68Q55, 68Q75.
CR Subject Classification (1991): F.1.1, F.1.2, F.4.3.
Keywords & Phrases: probabilistic automata, testing, button pushing scenario, trace distributions, approximation induction principle.

1 Introduction

A fundamental idea in concurrency theory is that two systems are deemed to be equivalent if they cannot be distinguished by observation. Depending on the power of the observer, different notions of behavioral equivalence arise. For systems modeled as labeled transition systems, this idea has been thoroughly explored and a large number of behavioral equivalences has been characterized operationally, algebraically, denotationally, logically, and via intuitive "testing scenarios" (also called "button pushing experiments"). We refer to Van Glabbeek [Gla01] for an excellent overview of results in this area of *comparative concurrency semantics*.

Testing scenarios provide an intuitive understanding of a behavioral equivalence via a machine model. A process is modeled as a black box that contains as

* Research 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 first author [Sto02a].

its interface to the outside world (1) a display showing the name of the action that is currently carried out by the process, and (2) some buttons via which the observer may attempt to influence the execution of the process. A process autonomously chooses an execution path that is consistent with its position in the labeled transition system contained in the black box. Trace semantics, for instance, is explained in [Gla01] with the *trace machine*, depicted in Figure 1 on the left. As one can see, this machine has no buttons at all. A slightly less

Fig. 1. The trace machine (left) and the failure trace machine (right).

trivial example is the *failure trace machine*, depicted in Figure 1 on the right, which, apart from the display, contains as its interface to the outside world a switch for each observable action. By means of these switches, an observer may determine which actions are *free* and which are *blocked*. This situation may be changed at any time during a run of a process. The display becomes empty if (and only if) a process cannot proceed due to the circumstance that all actions are blocked. If, in such a situation, the observer changes her mind and allows one of the actions the process is ready to perform, an action will become visible again in the display. Figure 2 gives an example of two labeled transition sys-

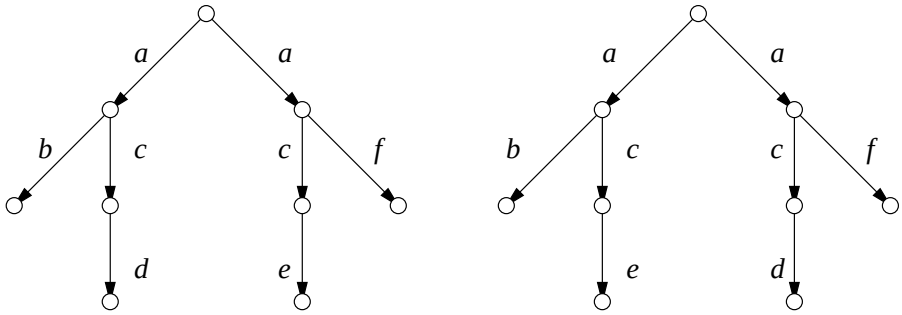

Fig. 2. Trace equivalent but not failure trace equivalent.

tems that can be distinguished by the failure trace machine but not by the trace machine. Since both transition systems have the same traces (ε, a, ab, ac, af, acd and ace), no difference can be observed with the trace machine. However, via the failure trace machine an observer can see a difference by first blocking

actions c and f, and only unblocking action c if the display becomes empty. In this scenario an observer of the left system may see an e, whereas in the right system the observer may see a d, but no e. We refer to [Gla01] for an overview of testing scenarios for labeled transition systems.

Probabilistic automata have become a popular mathematical framework for the specification and analysis of probabilistic systems. They have been developed by Segala [Seg95b,SL95,Seg95a] and serve the purpose of modeling and analyzing asynchronous, concurrent systems with discrete probabilistic choice in a formal and precise way. We refer to [Sto02b] for an introduction to probabilistic automata, and a comparison with related models. In this paper, we propose and study a simple and intuitive testing scenario for probabilistic automata: we just add a *reset* button to the trace machine. The resulting *trace distribution machine* is depicted in Figure 3. By resetting the machine it returns to its initial state and

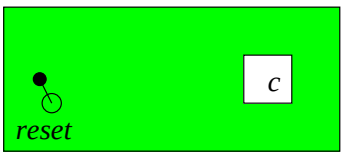

Fig. 3. The trace distribution machine.

starts again from scratch. In the non-probabilistic case the presence of a *reset* button does not make a difference[1], but in the probabilistic case it does: we can observe probabilistic behavior by repeating experiments and applying methods from statistics. Consider the two probabilistic automata in Figure 4. Here the

Fig. 4. Probabilistic automata representing a fair and an unfair coin.

arcs indicate probabilistic choice (as opposed to the nondeterministic choice in Figure 2), and probabilities are indicated next to the edges. These automata represent a fair and an unfair coin, respectively. We assume that the trace distribution machine has an "oracle" at its disposal which resolves the probabilistic

[1] For this reason a *reset* button does not occur in the testing scenarios of [Gla01]. An obvious alternative to the *reset* button would be a *on/off* button.

choices according to the probability distributions specified in the automaton. As a result, an observer can distinguish the two systems of Figure 4 by repeatedly running the machine until the display becomes empty and then restart it using the *reset* button. For the left process the number of occurrences of trace ab will approximately equal the number of occurrences of trace ac, whereas for the right process the ratio of the occurrence of the two traces will converge to $1:2$. Elementary methods from statistics allow one to come up with precise definitions of distinguishing tests.

The situation becomes more interesting when both probabilistic and nondeterministic choices are present. Consider the probabilistic automaton in Figure 5. If we repeatedly run the trace distribution machine with this automaton in-

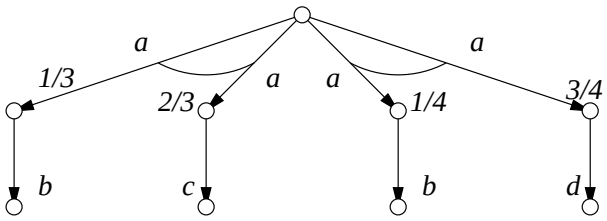

Fig. 5. The combination of probabilistic and nondeterministic choice.

side, the ratio between the various traces does not need to converge to a fixed value. However, if we run the machine sufficiently often we will observe that a weighted sum of the number of occurrences of traces ac and ad will approximately equal the number of occurrences of traces ab. Restricting attention to the cases where the left transition has been chosen, we observe $\frac{1}{2}\#[ac] \approx \#[ab]$. Restricting attention to the cases where the right transition has been chosen, we observe $\frac{1}{3}\#[ad] \approx \#[ab]$. Since in each execution either the left or the right transition will be selected, we have:

$$\frac{1}{2}\#[ac] + \frac{1}{3}\#[ad] \approx \#[ab].$$

Even though our testing scenario is simple, the combination of nondeterministic and probabilistic choice makes it far from easy to characterize the behavioral equivalence on probabilistic automata which it induces. The main technical contribution of this paper is a proof that the equivalence (preorder) on probabilistic automata induced by our testing scenario coincides with the trace distribution equivalence (preorder) proposed by Segala [Seg95a]. A result that we need to establish on the way is an *Approximation Induction Principle (AIP)* (cf. [BK86,BBK87]) for probabilistic processes. This principle says that if two finitely branching processes are equivalent up to any finite depth, then they are equivalent.

Being a first step, this paper limits itself to a simple class of probabilistic processes and to observers with limited capabilities. First of all, only *sequential*

processes are investigated: processes capable of performing at most one action at a time. Furthermore, we only study *concrete* processes in which no internal actions occur. Finally, observers can only interact with machines in an extremely limited way: apart from observing termination and the occurrence of actions, the only way in which they can influence the course of events is via the *reset* button[2]. It will be interesting to extend our result to richer classes of processes and more powerful observers, and to consider for instance a probabilistic version of the failure trace machine described earlier in this introduction.

Related work Several testing preorders and equivalences for probabilistic processes have been proposed in the literature [Chr90,Seg96,GN98,CDSY99,JY01]. All these papers study testing relations (i.e. testing equivalences or preorders) in the style of De Nicola and Hennesy [DNH84]. That is, they define a test as a (probabilistic) process that interacts with a system via shared actions and that reports success or failure in some way, for instance via success states or success actions. When a test is run on a system, the probability on success is computed, or if nondeterminism is present in either the test or the system, a set of these. By comparing the probabilities on success, one can say whether or not two systems are in the testing equivalence or preorder. For instance, two systems $\mathcal{A}$ and $\mathcal{B}$ are in the testing preorder of [JY01] if and only if for all tests T the maximal probability on success in $\mathcal{A} \parallel T$ is less than or equal to the maximal probability on success in $\mathcal{B} \parallel T$. The different testing relations in the mentioned papers arise by considering different kinds of probabilistic systems, by studying tests with different power (purely nondeterministic tests, finite trees or any probabilistic process) and by using different ways to compare two systems under test (e.g. may testing versus must testing). All of the mentioned papers provides alternative characterizations of their testing relation in terms of trace–based relations.

Thus, these testing relations are button pushing experiments in the sense that a test interacts with a system via synchronization on shared actions. However, in our opinion these relations are not entirely observational, because it is not described how the *probability* on success can be observed. In our view, this is an undesirable situation: in the end, the behavior of an automaton is what an external observer perceives. Therefore, we believe that any behavioral equivalence should either be characterized via some plausible testing scenario, or be strictly finer than such an equivalence and be justified via computational arguments.

The only other paper containing a convincing testing scenario for probabilistic systems is by Larsen & Skou [LS91]. They define a notion of tests for *reactive* probabilistic processes, that is, processes in which all outgoing transitions of a state have different labels. Furthermore, the observer is allowed to make arbitrary many copies of any state. For those tests, a fully observable characterization of probabilistic bisimulation based on hypothesis testing is given. (We note that copies of tests can both serve to discover the branching structure of a system –

[2] This ensures that our testing scenario truly is a "button pushing experiment" in the sense of Milner [Mil80]!

as in the nondeterministic case – and to repeat a certain experiment a number of times.)

More precisely, each test T in [LS91] gives rise to a set of observations O_T. Tests allow certain properties to be tested with arbitrary confidence $\alpha \in [0,1]$, the so–called level of significance. More precisely, a property Φ is said to be *testable* if for every level of significance α, there is a test T and a partition of observations O_T into $(E_\Phi, O_T \setminus E_\Phi)$ such that (1) if Φ holds in a state s and T is run in s, then it is likely that we observe an element from E_Φ, i.e. $\mathbf{P}_\Phi[E_\Phi] \geq 1-\alpha$ and (2) if Φ does not hold, then the probability to observe an element in E_Φ is small: $\mathbf{P}_{\neg\Phi}[E_\Phi] \leq \alpha$. Thus, by checking whether the outcome of the test is in E_Φ or not, we can find out whether s satisfies Φ and probability that the judgment is wrong is less than α. Using the terminology from hypothesis testing, Φ is the null hypothesis and E_Φ is the critical section.

Then is it shown that two states in a system that satisfies the minimal derivation assumption are probabilistically bisimilar if and only if they satisfy exactly the same testable properties. Here the minimal derivation assumption requires that any probability occurring in the system is an integer multiple of some value ε. Thus, although not explicitly phrased in these terms, one can say Larsen & Skou present a button pushing scenario for probabilistic processes.

Our work differs from the approach in [LS91] in the following aspects.

- We present our results in the more general PA model, whereas [LS91] considers the reactive model. As a consequence, the composition of a system and a test in [LS91] is purely probabilistic, that is, it does not contain nondeterministic choices, and theory from classical hypothesis testing applies. In contrast to this, the probabilistic automata that we consider do contain nondeterministic choices. To distinguish between likely and unlikely outcomes in these automata, we have to extend (some parts of) hypothesis testing with nondeterminism, which is technically quite involved.
- The main result of this paper, which is the characterization of trace distribution inclusion as a testing scenario, is established for all finitely branching systems, which is much more general than the minimal derivation assumption needed for the results in [LS91].
- The possibility in the testing scenario of Larsen & Skou to make copies of processes in any state (at any moment), is justified for instance in the case of a sequential system where one can make core dumps at any time. But for many distributed systems, it is not possible to make copies in any but the initial state. Therefore, it makes sense to study scenarios in which copying is not possible, as done in this paper.

Overview Even though readers may not expect this after our informal introduction, the rest of this paper is actually quite technical. We start in Section 2 with some mathematical preliminaries concerning functions, sequences and probability theory. In Section 3 we recall the definitions of probabilistic automata and their behavior. Section 4 is entirely devoted to the proof of the AIP for probabilistic processes. Section 5, finally, presents the characterization of the testing

preorder induced by the trace distribution machine as trace distribution inclusion.

2 Preliminaries

Functions If f is a function, then we denote the domain of f by $Dom(f)$. The *range* of f, notation $Ran(f)$, is the set $\{f(u) \mid u \in U\}$. If U is a set, then the *restriction* of f to U, notation $f \restriction U$, is the function g with $Dom(g) = Dom(f) \cap U$ satisfying $g(u) = f(u)$ for each $u \in Dom(g)$. We say that a function f is a *subfunction* of a function g, and write $f \subseteq g$, if $Dom(f) \subseteq Dom(g)$ and $f = g \restriction Dom(f)$. A function f is called *finite* if $Dom(f)$ is finite.

Sequences Let U be any set. A *sequence* over U is a function σ from a downward closed subset of the natural numbers to U. So the domain of a sequence is either the set $\mathbb{N}$ of natural numbers, or of the form $\{0, \ldots, k\}$, for some $k \in \mathbb{N}$, or the empty set. In the first case we say that the sequence is infinite, otherwise we say it is finite. The sets of finite and infinite sequences over U are denoted by U^* and U^∞, respectively. We will sometimes write σ_n rather than $\sigma(n)$. The symbol ε denotes the empty sequence, and the sequence containing one element $u \in U$ is denoted by u. Concatenation of a finite sequence with a finite or infinite sequence is denoted by juutaposition. We say that a sequence σ is a *prefix* of a sequence ρ, denoted by $\sigma \sqsubseteq \rho$, if $\sigma = \rho \restriction Dom(\sigma)$. Thus $\sigma \sqsubseteq \rho$ if either $\sigma = \rho$, or σ is finite and $\rho = \sigma\sigma'$ for some sequence σ'. If σ is a nonempty sequence then $first(\sigma)$ denotes the first element of σ and, if σ is also finite, then $last(\sigma)$ denotes the last element of σ. Finally, $length(\sigma)$ denotes the length of a finite sequence σ. A *subsequence* of an infinite sequence σ is an infinite sequence ρ that is obtained by removing finitely or infinitely many elements from σ. Formally, ρ is a subsequence of σ if there is an *index function*, that is a function $\kappa : \mathbb{N} \to \mathbb{N}$ such that (a) κ is strictly monotone (i.e., $n < m$ implies $\kappa(n) < \kappa(m)$), and (b) $\rho = \sigma \circ \kappa$.

An elementary (but fundamental) result from Analysis is the following theorem from Bolzano–Weierstraß.

Theorem 1 (Bolzano–Weierstraß). *Every bounded infinite sequence in $\mathbb{R}^n$ has a convergent subsequence.*

Let f_0, f_1, $f_2, \ldots$ be an infinite sequence of functions in $U \to [0,1]$, where U is a finite set. Then this sequence can be seen as a sequence over $[0,1]^n$, where n is the cardinality of U. Applying the Bolzano–Weierstraß Theorem to f_0, f_1, $f_2 \ldots$ yields that this sequence has a convergent subsequence, i.e. there exists an index function κ such that $f_{\kappa(0)}$, $f_{\kappa(1)}$, $f_{\kappa(2)}, \ldots$ has a limit (in $[0,1]^n$).

Probability Theory We recall a few basic notions from probability theory and introduce some notation.

Definition 1. *A* probability distribution *over a set U is a function $\mu : U \to [0,1]$ such that $\sum_{u \in U} \mu(u) = 1$. We define the* support *of μ by* $\mathrm{supp}(\mu) = \{u \in U \mid \mu(u) > 0\}$. *It follows straightforwardly from the definitions that this is a countable set. We denote the set of all probability distributions over U by* $\mathrm{Distr}(U)$.

We denote a probability distribution μ on a countable domain by enumerating it as a set of pairs. So, if $Dom(\mu) = \{u_1, u_2 \ldots\}$ then denote μ by $\{u_1 \mapsto \mu(u_1), u_2 \mapsto \mu(u_2) \ldots\}$. If the domain of μ is known, then we often leave out elements of probability 0. For instance, the probability distribution assigning probability 1 to an element $u \in U$ is denoted by $\{u \mapsto 1\}$, irrespective of U. Such distribution is called the *Dirac distribution* over u.

Definition 2. *A* probability space *is a triple $(\Omega, \mathcal{F}, \mathbf{P})$, where*

- *Ω is a set, called the* sample space,
- *$\mathcal{F} \subseteq 2^\Omega$ is* σ-field, *i.e. a collection of subsets of Ω which is closed under countable*[3] *union and complement, and which contains Ω,*
- *$\mathbf{P} : \mathcal{F} \to [0,1]$ is a* probability measure *on $\mathcal{F}$, which means that $\mathbf{P}[\Omega] = 1$ and for any countable collection $\{C_i\}_i$ of pairwise disjoint subsets in $\mathcal{F}$ we have $\mathbf{P}[\cup_i C_i] = \sum_i \mathbf{P}[C_i]$.*

3 Probabilistic Automata

Now, we recall the notion of a probabilistic automaton from Segala and Lynch [Seg95a,SL95]. Basically, a probabilistic automaton is a non-probabilistic automaton with the only difference that, rather than a single state, the target of a transition is a probability distribution over next states. We consider systems with only external actions, taken from a given, finite set Act. For technical reasons, we assume that Act contains a special element δ, referred to as the *halting* action.

Definition 3. *A* probabilistic automaton (PA) *is a triple $\mathcal{A} = (S, s^0, \Delta)$ with*

- *S a set of states,*
- *$s^0 \in S$ the initial state, and*
- *$\Delta \subseteq S \times Act \times \mathrm{Distr}(S)$ a transition relation.*

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

For the remainder of this section, we fix a PA $\mathcal{A} = (S, s^0, \Delta)$ and assume that Δ contains no transition labeled with δ.

Definition 4. *A PA $\mathcal{A}$ is* finitely branching *if for each state s, the set $\{(a, \mu, t) \mid s \overset{a,\mu}{\rightsquigarrow} t\}$ is finite.*

[3] In our terminology, countable objects include finite ones.

Thus, each state in a finitely branching PA has finitely many outgoing transitions and the target distribution of each transition has a finite support. As in the non-probabilistic case, an execution of $\mathcal{A}$ is obtained by resolving the nondeterministic choices in $\mathcal{A}$. This choice resolution is described by an adversary, a function which in each state of the system determines the next transition to be taken. Adversaries can be randomized, i.e. make choices probabilistically, history-dependent, i.e. make choices depending on on the path leading to the current state, and partial, i.e. they may choose to halt the execution at any point in time. Since we want adversaries to produce infinite sequences only, even when the execution is halted, we define adversaries of a PA via its halting extention.

Definition 5. *A* path *of $\mathcal{A}$ is an alternating, finite or infinite sequence*

$$\pi = s_0 a_1 \mu_1 s_1 a_2 \mu_2 s_2 \dots$$

of states, actions, and distributions over states such that (1) π starts with the initial state,[4] *i.e. $s_0 = s^0$, (2) if π is finite, it ends with a state, (3) $s_i \overset{a_{i+1},\mu_{i+1}}{\rightsquigarrow} s_{i+1}$, for each nonfinal i. We set the* length *of π, notation $|\pi|$, to the number of actions occurring in it and denote the set of all finite paths of $\mathcal{A}$ by $Path^*(\mathcal{A})$. For $n \in \mathbb{N} \cup \{\infty\}$, the set of all paths of $\mathcal{A}$ of length n by $Path^n(\mathcal{A})$. We define the associated* trace *of π, notation $trace(\pi)$, by $trace(\pi) = a_1 a_2 a_3 \dots$.*

Definition 6. *The* halting extension *of $\mathcal{A}$ is the PA $\delta\mathcal{A} = (S \cup \{\bot\}, s^0, \Delta')$, where Δ' is the least relation such that*

1. $s \xrightarrow{\delta}_{\delta\mathcal{A}} \{\bot \mapsto 1\}$,
2. $s \xrightarrow{a}_{\mathcal{A}} \mu \implies s \xrightarrow{a}_{\delta\mathcal{A}} (\mu \cup \{\bot \mapsto 0\})$.

Here we assume that $\bot$ is fresh. The transitions with label δ are referred to as halting *transitions.*

Definition 7. *A* (partial, randomized, history-dependent) adversary *E of $\mathcal{A}$ is a function*

$$E : Path^*(\delta\mathcal{A}) \to \mathrm{Distr}(Act \times \mathrm{Distr}(S_{\delta\mathcal{A}}))$$

such that, for each finite path π, if $E(\pi)(a, \mu) > 0$ then $last(\pi) \xrightarrow{a}_{\delta\mathcal{A}} \mu$.

We say that E is deterministic *if, for each π, $E(\pi)$ is a Dirac distribution. An adversary E* halts *on a path π if it extends π with the halting transition, i.e.,*

$$E(\pi)(\delta, \{\bot \mapsto 1\}) = 1.$$

For $k \in \mathbb{N}$, we say that the adversary E halts after k steps *if it halts on all paths with length greater than or equal to k. We denote by $Adv(\mathcal{A}, k)$ the set of all adversaries of $\mathcal{A}$ that halt after k steps and by $Dadv(\mathcal{A}, k)$ the set of deterministic adversaries in $Adv(A, k)$. Finally, we call E* finite *if $E \in Adv(\mathcal{A}, k)$, for some $k \in \mathbb{N}$.*

[4] Here we deviate from the standard definition, as we do not need paths starting from non-initial states.

The probabilistic behavior of an adversary is summarized by its associated probability space. First we introduce the function $\mathbf{Q}^E$, which yields the probability that E assigns to finite paths.

Definition 8. *Let E be an adversary of $\mathcal{A}$. The function $\mathbf{Q}^E : Path^*(\delta\mathcal{A}) \to [0,1]$ is defined inductively by*

$$\mathbf{Q}^E(s_0) = 1,$$
$$\mathbf{Q}^E(\pi a \mu s) = \mathbf{Q}^E(\pi) \cdot E(\pi)(a,\mu) \cdot \mu(s).$$

Definition 9. *Let E be an adversary of $\mathcal{A}$. The* probability space *associated to E is the probability space given by*

1. *$\Omega_E = Path^\infty(\delta\mathcal{A})$,*
2. *$\mathcal{F}_E$ is the smallest σ-field that contains the set $\{C_\pi \mid \pi \in Path^*(\delta\mathcal{A})\}$, where $C_\pi = \{\pi' \in \Omega_E \mid \pi \sqsubseteq \pi'\}$,*
3. *$\mathbf{P}_E$ is the unique measure on $\mathcal{F}_E$ such that $\mathbf{P}_E[C_\pi] = \mathbf{Q}^E(\pi)$, for all $\pi \in Path^*(\delta\mathcal{A})$.*

The fact that $(\Omega_E, \mathcal{F}_E, \mathbf{P}_E)$ is a probability space follows from standard measure theory arguments, see for instance [Coh80]. Note that Ω_E and $\mathcal{F}_E$ do not depend on E but only on $\mathcal{A}$, and that $\mathbf{P}_E$ is fully determined by the function $\mathbf{Q}^E$. For $E \in Adv(\mathcal{A},k)$, $\mathbf{P}_E$ is fully determined by $\mathbf{Q}^E \upharpoonright Path^k(\mathcal{A})$, i.e., the weight function restricted to paths of length k.

As for non-probabilistic automata, the visible behavior of $\mathcal{A}$ is obtained by removing the non-visible elements (in our case, the states) from an execution (adversary). This yields a trace distribution of $\mathcal{A}$, which assigns a probability to (certain) sets of traces.

Definition 10. *The* trace distribution *H of an adversary E, denoted $trd(E)$, is the probability space given by*

1. *$\Omega_H = Act^\infty$,*
2. *$\mathcal{F}_H$ is the smallest σ-field that contains the sets $\{C_\beta \mid \beta \in Act^*\}$, where $C_\beta = \{\beta' \in \Omega_H \mid \beta \sqsubseteq \beta'\}$,*
3. *$\mathbf{P}_H$ is the unique measure on $\mathcal{F}_H$ such that $\mathbf{P}_H[X] = \mathbf{P}_E[trace^{-1}(X)]$.*

Standard measure theory arguments [Coh80] ensure again that that $trd(E)$ is well-defined. Note that Ω_H and $\mathcal{F}_H$ do not depend on $\mathcal{A}$. This means that trace distributions are fully characterized by their probability measure. The set of trace distributions of adversaries of $\mathcal{A}$ is denoted by $trd(\mathcal{A})$. We write $\mathcal{A} \equiv_{\mathrm{TD}} \mathcal{B}$ if $trd(\mathcal{A}) = trd(\mathcal{B})$ and $\mathcal{A} \sqsubseteq_{\mathrm{TD}} \mathcal{B}$ if $trd(\mathcal{A}) \subseteq trd(\mathcal{B})$. The set of trace distributions of that arise from adversaries of $\mathcal{A}$ that halt after k steps is denoted $trd(\mathcal{A},k)$. If $trd(\mathcal{A},k) \subseteq trd(\mathcal{B},k)$ then we write $\mathcal{A} \sqsubseteq^k_{\mathrm{TD}} \mathcal{B}$.

Lemma 1. *Let X and Y be non-empty sets and $f : X \to \mathrm{Distr}(Y)$ a function. If X is finite then*

$$\sum_{g:X\to Y} \prod_{x\in X} f(x)(g(x)) = 1.$$

Proof: By induction on $\# X = n$. Let $x_0 \in X$ and write $X_1 = X \setminus \{x_0\}$.

- If $n = 1$, then $\sum_{g:X\to Y} \prod_{x\in X} f(x)(g(x)) = \sum_{y\in Y} f(x_0)(y) = 1$ because $f(x_0)$ is a distribution function.
- Assume that the proposition holds for all X' with $\# X' = n$ and let $\# X = n+1$. Then

$$\sum_{g:X\to Y} \prod_{x\in X} f(x)(g(x)) =$$
$$\sum_{g:X\to Y} f(x_0)(g(x_0)) \cdot \prod_{x\in X_1} f(x)(g(x)) =$$
$$\sum_{g_1:X_1\to Y} \sum_{y\in Y} f(x_0)(y) \cdot \prod_{x\in X_1} f(x)(g(x)) =$$
$$\sum_{g_1:X_1\to Y} 1 \cdot \prod_{x\in X_1} f(x)(g(x)) = 1.$$

□

The following lemma shows that each finite adversary in a finitely branching PA can be written as a convex combination of deterministic adversaries.

Lemma 2. *Let $k \in \mathbb{N}$, let $\mathcal{A}$ be a finitely branching PA and let E be an adversary in $Adv(\mathcal{A}, k)$.*

1. *Then E can be written as a convex combination of deterministic adversaries in $Dadv(\mathcal{A}, k)$, i.e., there exists a probability distribution ν over $Dadv(\mathcal{A}, k)$ such that, for all π, a and μ,*

$$E(\pi)(a, \mu) = \sum_{D\in Dadv(\mathcal{A},k)} \nu(D) \cdot D(\pi)(a, \mu).$$

2. *If $E = \sum_{D\in Dadv(\mathcal{A},k)} \nu(D) \cdot D(\pi)(a, \mu)$ for some $\nu \in \mathrm{Distr}(Dadv(\mathcal{A}, k))$ then*

$$\mathbf{Q}^E(\sigma) = \sum_{D\in Dadv(\mathcal{A},k)} \nu(D) \cdot \mathbf{Q}^{E_i}(\sigma).$$

Proof:

1. First, observe that the set $Dadv(\mathcal{A}, k)$ is finite, because $\mathcal{A}$ is finitely branching.
   The idea in the proof is as follows. Let $D \in Dadv(\mathcal{A}, k)$ be an adversary such that

$$D(\sigma)(a, \mu) = 1 \implies E(\sigma)(a, \mu) > 0, \text{ for all } \sigma \qquad (*)$$

   Then D can be seen as an adversary within E: among all the steps that E schedules with a positive probability, D schedules one with probability one.

Now, multiply all the probabilities $E(\pi)(a,\mu)$ that E assigns to steps (a,μ) taken in D, i.e. steps with $D(\pi)(a,\mu) = 1$. This yields a value p_D and we show that E can be obtained by selecting the adversary D with probability $p_D = \nu(D)$. Furthermore, note that $p_D = 0$ if D does not meet (*). Hence, define ν by

$$\nu(D) = \prod_{\sigma \in Path^*(\mathcal{A})} E(\sigma)(D(\sigma)),$$

where, as before, we write $D(\sigma) = (a,\mu)$ for $D(\sigma)(a,\mu) = 1$. Moreover, write $\mathcal{D}$ for $Dadv(\mathcal{A},k)$ and P for $Path^*(\mathcal{A})$. Then $\nu \in \mathrm{Distr}(()Dadv\mathcal{A},k)$ because

$$\sum_{D\in\mathcal{D}} \nu(D) \;=\; \sum_{D\in\mathcal{D}} \prod_{\sigma\in P} E(\sigma)(D(\sigma)) \;=\; 1.$$

Since $\mathcal{D}$ is finite, the last step is justified by Lemma 1. For the same reason we have for all ρ, a, μ that

$$\begin{aligned}
&\sum_{D\in\mathcal{D}} \nu(D)\cdot D(\rho)(a,\mu) \;= \\
&\sum_{D\in\mathcal{D}} \prod_{\sigma\in P} E(\sigma)(D(\sigma))\cdot D(\rho)(a,\mu) \;= \\
&\sum_{D\in\mathcal{D}, D(\rho)=(a,\mu)} \prod_{\sigma\in P} E(\rho)(a,\mu)\cdot 1 \;= \\
&E(\rho)(a,\mu).
\end{aligned}$$

2. By induction on the length of σ.
   - If $\sigma = s^0$ has length 0, then $\mathbf{Q}^E(s^0) = 1 = \sum_{D\in\mathcal{D}} \nu(D) = \sum_{D\in\mathcal{D}} \nu(D)\cdot \mathbf{Q}^D(s^0)$.
   - Let $\sigma = \sigma' a\mu t$, then

$$\begin{aligned}
\mathbf{Q}^E(\sigma' a\mu t) &= \mathbf{Q}^E(\sigma')\cdot E(\sigma')(a,\mu)\cdot\mu(t) = \\
&= \sum_{D\in\mathcal{D}} \nu(D)\cdot\mathbf{Q}^D(\sigma')\cdot \sum_{D'\in\mathcal{D}} \nu(D')\cdot D'(\sigma')(a,\mu)\cdot\mu(t) \;= \\
&= \sum_{D\in\mathcal{D}} \nu(D)\cdot\mathbf{Q}^D(\sigma')\cdot \sum_{D'\in\mathcal{D}} \nu(D')\cdot D(\sigma')(a,\mu)\cdot\mu(t) \;= \\
&= \sum_{D\in\mathcal{D}} \nu(D)\cdot\mathbf{Q}^D(\sigma')\cdot D(\sigma')(a,\mu)\cdot\mu(t) \;= \\
&= \sum_{D\in\mathcal{D}} \nu(D)\cdot\mathbf{Q}^D(\sigma' a\mu t) = \\
&= \sum_{D\in\mathcal{D}} \nu(D)\cdot\mathbf{Q}^D(\sigma).
\end{aligned}$$

□

4 The Approximation Induction Principle

This section is entirely devoted to a proof of an *Approximation Induction Principle (AIP)* (cf. [BK86,BBK87]) for probabilistic processes. We need this result to characterize the equivalence on probabilistic automata induced by the trace distribution machine in Section 5.

Theorem 2 (Approximation Induction Principle). *Let $\mathcal{A}$ and $\mathcal{B}$ be PAs and let $\mathcal{B}$ be finitely branching. Then*

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

Proof: Assume that $\mathcal{A} \sqsubseteq^k_{\mathrm{TD}} \mathcal{B}$, for all k. In order to prove $\mathcal{A} \sqsubseteq_{\mathrm{TD}} \mathcal{B}$, let H be a trace distribution of $\mathcal{A}$ and let E be an adversary of $\mathcal{A}$ with $H = \mathit{trd}(E)$. Via a number of subclaims, we prove that $H \in \mathit{trd}(\mathcal{B})$.

For each $k \in \mathbb{N}$, define E_k by

$$E_k(\pi) = \begin{cases} E(\pi) & \text{if } |\pi| < k, \\ \{(\delta, \{\bot \mapsto 1\}) \mapsto 1\} & \text{otherwise.} \end{cases}$$

Clearly, $E_k \in \mathit{Adv}(\mathcal{A}, k)$, so $\mathit{trd}(E_k) \in \mathit{trd}(\mathcal{A}, k)$. By assumption, there is an adversary F_k of $\mathcal{B}$ such that $\mathit{trd}(E_k) = \mathit{trd}(F_k)$. We view F_k as a function in

$$\mathit{Path}^*(\delta\mathcal{B}) \times \mathit{Act} \times \mathrm{Distr}(S_{\delta\mathcal{B}}) \to [0,1].$$

We will construct an adversary G of $\mathcal{B}$ with $\mathit{trd}(G) = H$ from the sequence of functions $F = F_0, F_1, F_2 \ldots$. The idea is that, since only the paths of length k matter, F_k is essentially a finite function and we can use the Bolzano–Weierstraß Theorem to obtain G as the limit of a convergent subsequence of F. However, this theorem cannot be applied immediately, because the finite domains of these functions are growing. Therefore, we will operate in several stages. The basic idea is to construct at stage $n+1$ a convergent subsequence with index function κ_{n+1} of $F^n_0, F^n_1, F^n_2 n \ldots$, where F^n_k is the restriction of F_k to paths of length n. This sequence consists of finite functions with the same, finite domain and a bounded range (viz. $[0,1]$) and has therefore a convergent subsequence. We define G_n as the limit of κ_n. Thus, we will obtain a sequence of increasing subfunctions $G_1 \subseteq G_2 \subseteq G_3 \ldots$ and we take G to be its limit. We will need several technical lemmas to ensure that everything is as expected and to prove finally that $\mathit{trd}(G) = \mathit{trd}(E)$.

Throughout this proof, we use the following notations.

$$P_n = \bigcup_{1 \le n} \mathit{Path}^i(\delta\mathcal{B})$$
$$D_n = \{\mu \in \mathrm{Distr}(S_{\delta\mathcal{B}}) \mid \mu \text{ occurs in some } \pi \in P_{n+1}\}$$

$$P = \mathit{Path}^*(\delta\mathcal{B})$$
$$D = \mathrm{Distr}(S_{\delta\mathcal{B}})$$

Note that $P_n \subseteq P_{n+1}$, $D_n \subseteq D_{n+1}$, $P = \cup_n P_n$ and $D \supseteq \cup_n D_n$. In fact, D may contain distributions that are not contained in any D_n. Observe also that $\pi \in P_n$ and $\pi \stackrel{a,\mu}{\rightsquigarrow} s$ implies $\mu \in D_n$. Since $\mathcal{B}$ is finitely branching, there are only finitely many paths of length at most n and hence P_n and D_n are both finite. Recall that Act is finite by definition. Therefore, the following function F_k^n is finite:

$$F_k^n : P_n \times Act \times D_n \to [0,1]$$
$$F_k^n = F_k \upharpoonright P_n \times Act \times D_n.$$

Claim 1 *$F_k^n \subseteq F_k^{n+1}$ for all k, n.*

PROOF: Easy verification. ⊠

For each n, let ρ_n be the sequence

$$\rho_n \quad = \quad F_0^n \;\; F_1^n \;\; F_2^n \;\; F_3^n \ldots$$

and let κ_n be the index function defined inductively as follows:

- κ_0 is the identity function.
- Let κ be the index function of a convergent subsequence of $\rho_n \circ \kappa_n$ (such a subsequence exists by the Bolzano–Weierstraß Theorem). Then $\kappa_{n+1} = \kappa_n \circ \kappa$.

We define function $G_n : P_n \times Act \times D_n \to [0,1]$ by

$$G_n = \lim(\rho_n \circ \kappa_{n+1}),$$

i.e. , G_n is the limit of the convergent subsequence just defined.

Claim 2 *$G_n \subseteq G_{n+1}$.*

PROOF: Clearly, $\mathrm{Dom}(G_n) \subseteq \mathrm{Dom}(G_{n+1})$. Let $(\pi, a, \mu) \in \mathrm{Dom}(G_n)$. Then

$$\begin{aligned} G_n(\pi)(a,\mu) &= \lim_{k\to\infty} F^n_{\kappa_{n+1}(k)}(\pi)(a,\mu) && \{\mathrm{Ran}(\kappa_{n+2}) \subseteq \mathrm{Ran}(\kappa_{n+1}\} \\ &= \lim_{k\to\infty} F^n_{\kappa_{n+2}(k)}(\pi)(a,\mu) && \{\text{Claim 1}\} \\ &= \lim_{k\to\infty} F^{n+1}_{\kappa_{n+2}(k)}(\pi)(a,\mu) \\ &= G_{n+1}(\pi)(a,\mu). \end{aligned}$$

⊠

Let $G' = \cup_n G_n$, i.e. for $\pi \in P_n$, $a \in Act$ and $\mu \in D_n$, $G'(\pi)(a,\mu) = G_n(\pi)(a,\mu)$. Then G' is a function in $\cup_n P_n \times Act \times D_n \to [0,1]$. We extend G' to a function G in $P \times Act \times D \to [0,1]$ as follows

$$G(\pi)(a,\mu) = \begin{cases} G'(\pi)(a,\mu) & \text{if } \exists n.\pi \in P_n \wedge \mu \in D_n, \\ 0 & \text{otherwise.} \end{cases}$$

The rest of this proof is concerned with showing that G is an adversary with $trd(G) = H$, which is exactly what we are after.

Claims 3 and 4 together imply that G is an adversary of $\mathcal{B}$. Claim 3 states that G respects the transition relation of $\delta\mathcal{B}$, and Claim 4 establishes that G has the required type, i.e.

$$G : Path^*(\delta\mathcal{B}) \to \mathrm{Distr}(Act \times \mathrm{Distr}(S_{\delta\mathcal{B}})).$$

Claim 3 *Suppose $\pi \in P$, $a \in Act$, $\mu \in D$ and $G(\pi)(a,\mu) > 0$. Then $last(\pi) \xrightarrow{a} \mu$ is a transition of $\delta\mathcal{B}$.*

PROOF: Since $G(\pi)(a,\mu) > 0$, it follows from the definition of G that $G(\pi)(a,\mu) = G'(\pi)(a,\mu)$. Hence, by definition of G', there is an n such that $G'(\pi)(a,\mu) = G_n(\pi)(a,\mu)$. Then

$$\begin{aligned} 0 &< G_n(\pi)(a,\mu) && \{\text{def. } G_n\} \\ &= \lim_{k\to\infty} F^n_{\kappa_{n+1}(k)}(\pi)(a,\mu) && \{\text{def. } F^n_i\} \\ &= \lim_{k\to\infty} F_{\kappa_{n+1}(k)}(\pi)(a,\mu) \end{aligned}$$

This implies that $F_{\kappa_{n+1}(k)}(\pi)(a,\mu) > 0$ for large k. Since $F_{\kappa_{n+1}(k)}$ is an adversary of $\mathcal{B}$, $last(\pi) \xrightarrow{a} \mu$ is a transition of $\mathcal{B}$. ⊠

Claim 4 *For all $\pi \in P$, $\sum_{a\in Act,\mu\in D} G(\pi)(a,\mu) = 1$.*

PROOF: Choose $\pi \in P$ and let $n = |\pi|$. Then $\pi \in P_n$ and

$$\begin{aligned} &\sum_{a\in Act,\mu\in D} G(\pi)(a,\mu) && \{\text{def. } G\} \\ &= \sum_{a\in Act,\mu\in D_n} G'(\pi)(a,\mu) && \{\text{def. } G'\} \\ &= \sum_{a\in Act,\mu\in D_n} G_n(\pi)(a,\mu) && \{\text{def. } G_n\} \\ &= \sum_{a\in Act,\mu\in D_n} \lim_{k\to\infty} F^n_{\kappa_{n+1}(k)}(\pi)(a,\mu) && \{\text{def. } F^n_i\} \\ &= \sum_{a\in Act,\mu\in D_n} \lim_{k\to\infty} F_{\kappa_{n+1}(k)}(\pi)(a,\mu) && \{\text{finite sum}\} \\ &= \lim_{k\to\infty} \sum_{a\in Act,\mu\in D_n} F_{\kappa_{n+1}(k)}(\pi)(a,\mu) && \{\text{def. } Act,\ D_n\} \\ &= \lim_{k\to\infty} \sum_{a\in Act,\mu\in D} F_{\kappa_{n+1}(k)}(\pi)(a,\mu) && \{F_i \text{ adversary}\} \\ &= \lim_{k\to\infty} 1 \\ &= 1 \end{aligned}$$

⊠

Note that the following claim concerns G and F_i, which are adversaries. In contrast, G_n and F_k^n are just functions, not adversaries.

Claim 5 $\mathbf{Q}^G(\pi) = \lim_{k\to\infty} \mathbf{Q}^{F_{\kappa(n+1)(k)}}(\pi)$ *for all* $\pi \in Path^*(\delta\mathcal{B})$ *with* $|\pi| = n$.

PROOF: By induction on n.

- Then case $n = 0$ follows immediately from the fact that $\mathbf{Q}^E(s_0) = 1$ for all adversaries E.
- Case $n+1$. Let π' be a path of length $n+1$ and write $\pi' = \pi \overset{a,\mu}{\rightsquigarrow} s$. Then $\pi \in P_n$, $a \in Act$, $\mu \in D_n$ and

$$
\begin{aligned}
&\mathbf{Q}^G(\pi') \\
&= \mathbf{Q}^G(\pi \overset{a,\mu}{\rightsquigarrow} s) && \{\text{def. } \mathbf{Q}\} \\
&= \mathbf{Q}^G(\pi) \cdot G(\pi)(a,\mu) \cdot \mu(s) && \{\text{IH, } |\pi| = n\} \\
&= \lim_{k\to\infty} \mathbf{Q}^{F_{\kappa(n)(k)}}(\pi) \cdot G_n(\pi)(a,\mu) \cdot \mu(s) && \{\text{def. } G_n\} \\
&= \lim_{k\to\infty} \mathbf{Q}^{F_{\kappa(n)(k)}}(\pi) \cdot \lim_{k\to\infty} F_{\kappa(n)(k)}(\pi)(a,\mu) \cdot \mu(s) \\
&= \lim_{k\to\infty} \mathbf{Q}^{F_{\kappa(n)(k)}}(\pi) \cdot F_{\kappa(n)(k)}(\pi)(a,\mu) \cdot \mu(s) && \{\text{def. } \mathbf{Q}\} \\
&= \lim_{k\to\infty} \mathbf{Q}^{F_{\kappa(n+1)(k)}}(\pi \overset{a,\mu}{\rightsquigarrow} s) \\
&= \lim_{k\to\infty} \mathbf{Q}^{F_{\kappa(n+1)(k)}}(\pi')
\end{aligned}
$$

$\boxtimes$

Claim 6 $\mathbf{Q}^E(\pi) = \lim_{k\to\infty} \mathbf{Q}^{E_{\kappa(n)(k)}}(\pi)$ *for all* n *and* π.

PROOF: Since $\kappa(n)$ is an index function, we have $\lim_{k\to\infty} \kappa(n)(k) = \infty$ and therefore $E_{\kappa(n)(k)}[\pi] = E[\pi]$ for $\kappa(n)(k) \geq |\pi|$. So, $\lim_{k\to\infty} \mathbf{Q}^{E_{\kappa(n)(k)}}(\pi) = \mathbf{Q}^E(\pi)$. $\boxtimes$

The following is an immediate consequence of the previous claim.

Claim 7 $\mathbf{P}_{trd(E)}[C_\alpha] = \lim_{k\to\infty} \mathbf{P}_{trd(E_{\kappa(n)(k)})}[C_\alpha]$, *for all* α.

Claim 8 $trd(G) = trd(E)$.

PROOF: Let $H_1 = trd(G)$ and $H_2 = trd(E)$. It suffices to show that $\mathbf{P}_{H_1}[C_\alpha] = \mathbf{P}_{H_2}[C_\alpha]$ for all $\alpha \in Act^*$. Let $n = |\alpha|$.

$$\begin{aligned}
\mathbf{P}_{H_1}[C_\alpha] &= \sum_{\pi\,|\,trace(\pi)\in C_\alpha} \mathbf{P}_G[\pi] \\
&= \sum_{\pi\,|\,trace(\pi)=\alpha\wedge|\pi|=n} \mathbf{P}_G[C_\pi] && \{\text{def. } C_\pi\} \\
&= \sum_{\pi\,|\,trace(\pi)=\alpha\wedge|\pi|=n} \mathbf{Q}^G(\pi) && \{\text{Claim 5, } |\pi| = n\} \\
&= \sum_{\pi\,|\,trace(\pi)=\alpha\wedge|\pi|=n} \lim_{k\to\infty} \mathbf{Q}^{F_{\kappa(n)(k)}}(\pi) && \{\text{finite sum}\} \\
&= \lim_{k\to\infty} \sum_{\pi\,|\,trace(\pi)=\alpha\wedge|\pi|=n} \mathbf{Q}^{F_{\kappa(n)(k)}}(\pi) && \{\text{def. } C_\alpha,\ \mathbf{P}_{F_i}\} \\
&= \lim_{k\to\infty} \mathbf{P}_{F_{\kappa(n)(k)}}[C_\alpha] \\
&= \lim_{k\to\infty} \mathbf{P}_{E_{\kappa(n)(k)}}[C_\alpha] && \{\text{Claim 7}\} \\
&= \mathbf{P}_E[C_\alpha]\{trd(F_i) = trd(E_i)\}
\end{aligned}$$

Note that the set $\{\pi \mid trace(\pi) = \alpha \wedge |\pi| = n\}$ is finite, because $\mathcal{A}$ is finitely branching and hence the summations above are all finite. ⊠

□

5 Characterization of Testing Preorder

The operational behavior of a trace distribution machine described in Section 1 is specified accurately by the notion of a (partial, randomized, history-dependent) adversary, introduced in Definition 7. Hence, when operating, the trace distribution machine chooses an execution path within some probabilistic automaton $\mathcal{A}$, using some adversary E.

Define a *sample O of depth k and width m* to be an element of $(Act^k)^m$, i.e., a sequence consisting of m sequences of actions of length k. A sample describes what an observer potentially may record when running m times an experiment of length k on the trace distribution machine. Note that if, during a run, the machine halts before k observable actions have been performed, we can still obtain a sequence of k actions by attaching a number of δ actions at the end.

We write $freq(O)$ for the function in $Act^k \to \mathbb{Q}$ that assigns to each sequence β in Act^k the frequency with which β occurs in O. That is, for $O = \beta_1\beta_2, \ldots, \beta_m$

$$freq(O)(\beta) = \frac{\#\{i \mid \beta_i = \beta, 1 \le i \le m\}}{m}.$$

Note that $freq(O)$ is a probability distribution over $(Act^k)^m$. We base our statistical analysis on $freq(O)$ rather than just O. This means we ignore some of

the information contained in samples, which more advanced statistical methods may want to explore. If, for instance, we consider the sample O of depth one and width 2000 that consists of 1000 *head* actions followed by 1000 *tail* actions, then it is quite unlikely that this will be a sample of a trace distribution machine implementing a fair coin. However, the frequency function $\mathit{freq}(O)$ can very well be generated by a fair coin.

For a trace distribution $H \in \mathit{trd}(\mathcal{A}, k)$, we denote by $\mu_H : \mathit{Act}^k \to [0,1]$ the probability distribution given by $\mu_H(\beta) = \mathbf{P}_H[C_\beta]$. Since H halts after k steps, $\mu_H(\beta)$ yields the probability that β is picked when we generate a sequence according to H. In other words, $\mu_H(\beta)$ yields the probability that during a run, the trace distribution machine produces the action sequence β, when it resolves its nondeterministic choices according to an adversary E with $\mathit{trd}(E) = H$. Therefore, the probability that the sample $O = \beta_1\beta_2, \ldots, \beta_m$ is generated when we successively and independently choose sequences β_i according to distributions $H_i \in \mathit{trd}(\mathcal{A}, k)$, is given by

$$\mathbf{P}_{H_1,\ldots,H_m}[O] \;=\; \prod_{i=1}^{m} \mu_{H_i}(\beta_i).$$

Finally, the probability that an element from a set $\mathcal{O} \subseteq (\mathit{Act}^k)^m$ is picked, equals

$$\mathbf{P}_{H_1,\ldots,H_m}[\mathcal{O}] \;=\; \sum_{O \in \mathcal{O}} \mathbf{P}_{H_1,\ldots,H_m}[O].$$

Given $H_1, H_2, \ldots, H_m$, we want to distinguish between outcomes that are likely to be generated by $H_1, H_2, \ldots, H_m$, and those which are not. To do so, we first fix an $\alpha \in (0,1)$ as the desired level of significance. Our goal is to define a set $\mathcal{K}_{H_1,H_2,\ldots,H_m}$, the likely outcomes, such that

1. $\mathbf{P}_{H_1,\ldots,H_m}[\mathcal{K}_{H_1,H_2,\ldots,H_m}] > 1 - \alpha$,
2. $\mathcal{K}_{H_1,H_2,\ldots,H_m}$ is, in some sense, minimal.

Condition (1) will ensures that the probability that we believe that O is not generated by $H_1, \ldots, H_m$ while it is so, is at most α. Condition (2) will ensure that $\mathbf{P}_{H'_1,\ldots,H'_m}[\mathcal{K}_{H_1,H_2,\ldots,H_m}]$ is as small as possible for sequences $(H'_1, \ldots, H'_m)$ different from $(H_1, \ldots, H_m)$. (How small this probability is highly depends on which H'_i's we take.) Therefore, the probability that we consider O to be an execution while it is not, is as small as possible. In terminology from hypothesis testing: our null hypothesis states that O is generated by $H_1, \ldots, H_m$ and condition (1) bounds the probability on false rejection and (2) minimizes the probability on false acceptance. The set $\mathcal{K}_{H_1,H_2,\ldots,H_m}$ is the complement of the critical section. Note that in classical hypothesis testing all subsequent experiments $\beta_1, \ldots \beta_m$ are drawn from the same probability distribution, whereas in our setting, each experiment is governed by a different probability mechanism given by H_i.

The idea behind the definition of $\mathcal{K}_{H_1,\ldots,H_m}$ is as follows. The *expected frequency* of a sequence β in a sample generated by $H_1, \ldots, H_m$ is given by

$$\mathbf{E}_{H_1,\ldots,H_m}(\beta) = \frac{1}{m}\sum_{i=1}^{m} \mu_{H_i}(\beta).$$

Since fluctuations around the expected value are likely, we allow deviations of at most ε from the expected value. Here, we choose ε as small as possible, but large enough such that the probability on a sample whose frequency deviates at most ε from $\mathbf{E}_{H_1,\ldots,H_m}$ is bigger than α. Then, conditions (1) and (2) above are met. Formally, define the *ε-sphere $B_\varepsilon(\mu)$ around μ* as

$$B_\varepsilon(\mu) = \{\nu \in \mathrm{Distr}(Act^k) \mid dist(\mu,\nu) \leq \varepsilon\},$$

where let *dist* is the standard distance on $\mathrm{Distr}(Act^k)$ given by $dist(\mu,\nu) = \sqrt{\sum_{\beta \in Act^k} |\mu(\beta) - \nu(\beta)|^2}$.

Definition 11. *For a sequence $H_1, H_2, \ldots H_m$ of trace distributions in $trd(\mathcal{A}, k)$, we define $\mathcal{K}_{H_1,\ldots H_m}$ as the smallest*[5] *sphere $B_\varepsilon(\mathbf{E}_{H_1,\ldots H_m})$ such that*

$$\mathbf{P}_{H_1,\ldots,H_m}[\{O \in (Act^k)^m \mid freq(O) \in B_\varepsilon(\mathbf{E}_{H_1,\ldots H_m})\}] > 1 - \alpha.$$

We say that O is an observation of $\mathcal{A}$ (of depth k and width m) *if*

$$O \in \mathcal{K}_{H_1,\ldots,H_m}.$$

We write $Obs(\mathcal{A})$ for the set of observations of $\mathcal{A}$.

Example 1. We take $\alpha = 0.05$ as the level of significance. First, consider the leftmost PA in Figure 4 and samples of depth 2 and width 100. This means that the probabilistic trace machine is run 100 times and each time we get a trace of length 2.

Then any sample O_1 in which the sequence ab occurs 42 times and ac 58 times is an observation of $\mathcal{A}$; samples in which ab occurs 38 times and ac 62 times are not. Let E be the adversary that, in each state, schedules with probability one the unique transition, if available, in that state and otherwise it schedules the halting transition with probability one. For $H = trd(E)$, we have $\mu_H(ab) = \mu_H(ac) = \frac{1}{2}$ and $\mu_H(\beta) = 0$ for all other sequences. Let $H^{100} = (H_1, \ldots H_{100})$ be sequence of adversaries with $H_i = H$. Then $\mathbf{E}_{H^{100}} = \mu_H$ and, since μ_H assigns a positive probability only to ab and ac, we have that $\mathbf{P}_{H^{100}}[B_\varepsilon(\mu_H)] = \mathbf{P}_{H^{100}}[\{O_1 \mid \frac{1}{2} - \varepsilon < freq(O_1)(ab) < \frac{1}{2} + \varepsilon\}]$. One can show that then smallest sphere such that $\mathbf{P}_{H^{100}}[B_\varepsilon(\mu_H)] \geq 0.95$ is obtained by taking $\varepsilon = \frac{1}{10}$. Since $freq(O_1) \in B_\varepsilon(\mu_H)$, O_1 is an observation.

Then, a sample O_2 containing with 20 $\delta\delta$'s, 42 ab's and 58 ac's is an observation of depth 2 and width 120. It arises from taking 100 times adversary H as above and 20 adversaries that halt with probability one on every path.

[5] This minimum exists, because there finitely many samples.

Now, consider the automaton in Figure 5. Consider the scheduler E_3 that in the initial state, schedules both a transitions with probability $\frac{1}{2}$; with probability one the unique b, c or d transition whenever it is avaible and otherwise the halting transition with probability one. Let K^{120} be the sequence consisting of 120 times the adversary K. The expected frequency of K^{120} is $\frac{7}{24}$ for ab, $\frac{8}{24}$ for ac, and $\frac{9}{24}$ for ad and then $\mathcal{K}_{K^120} = B_{\frac{1}{11}}(\mathbf{E}_{K^{100}})$, so for instance, the sequence with 40 ab's, 40 ac's and 40 ad's is an observation of the mentioned PA.

5.1 The characterization theorem

We can now prove our main characterization theorem, which states that the preorder induced by the testing scenario coincides with trace distribution inclusion.

Theorem 3. *$Obs(\mathcal{A}) \subseteq Obs(\mathcal{B}) \iff \mathcal{A} \sqsubseteq_{\mathrm{TD}} \mathcal{B}$.*

The following corollary is immediate.

Corollary 1. *$Obs(\mathcal{A}) = Obs(\mathcal{B}) \iff \mathcal{A} \equiv_{\mathrm{TD}} \mathcal{B}$.*

Before presenting the proof of the main theorem, we introduce four auxilliary results and some notation.

The first result states that, in a large number of Bernouilli trials (with different parameters), the set of outcomes with large deviations from the expected frequency of the number of 1's has a small probability. In fact, by choosing the number of trials large enough, we can get the probability on deviations of ε or more as small as we want.

Proposition 1. *Let $\alpha \in (0,1)$ and $\varepsilon > 0$. Then there exists an $m' \in \mathbb{N}$ such that the following holds. For all $m \geq m'$, and all sequences $X_1, X_2, \ldots, X_m$ of m independent random variables, where X_i has a Bernouilli distribution with parameter p_i, for some $p_i \in [0,1]$ (i.e. $\mathbf{P}[X_i = 1] = p_i, \mathbf{P}[X_i = 0] = 1 - p_i$), we have that*

$$\mathbf{P}\left[\, |Z_m - \mathbf{E}[Z_m]| \;\geq\; \varepsilon \,\right] \leq \alpha.$$

Here, $Z_m = \frac{1}{m}\sum_{i=1}^{m} X_i$ yields the frequency of the number of 1's that have been drawn in $(X_1, \ldots, X_m)$.

Proof: Take $m \geq m' \geq \frac{1}{4\varepsilon^2\alpha}$ and let $X_1, X_2, \ldots, X_m$ be a sequence of m independent random variables, where X_i has a Bernouilli distribution with parameter $p_i \in [0,1]$. First note that

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

Hence, by Chebychev's Inequality, we have

$$\mathbf{P}\big[|Z_m - \mathbf{E}[Z_m]| \geq \varepsilon\big] \;\leq\; \frac{1}{\varepsilon^2}\mathrm{Var}(Z_m) \;\leq\; \frac{1}{\varepsilon^2}\frac{1}{4m} \;\leq\; \frac{4\varepsilon^2\alpha}{4\varepsilon^2} = \alpha.$$

□

The second result reformulates the proposition above in terms of trace distributions rather than random variables: we can choose the number of runs of the trace distribution machine large enough so that, with high probability (i.e. 1-α), for each sequence β, the number of β's in the sample we draw deviates no more than ε from its expected value.

Proposition 2. *Let $\alpha \in (0,1)$, $\varepsilon > 0$ and $k \in \mathbb{N}$. Then there exists an $m' \in \mathbb{N}$ such that for all $m \geq m'$, all trace distributions $H_1, H_2, \ldots, H_m \in trd(\mathcal{A}, k)$ and all $\beta \in Act^k$*

$$\mathbf{P}_{H_1,\ldots,H_m}\big[\ \{O \in (Act^k)^m \mid\ |freq(O)(\beta) - \mathbf{E}_{H_1,\ldots,H_m}(\beta)| \geq \varepsilon\}\ \big] \leq \alpha.$$

Proof: This statement is merely a reformulation of Proposition 1. Given a sequence of trace distributions $H_1, \ldots, H_m$ and an sequence $\beta \in Act^k$, we define a sequence of independent randoms variables $X_1^\beta, \ldots, X_m^\beta$, where $X_i^\beta : (Act^k)^m \to \{0,1\}$ indicates whether the sequence β is drawn by H_i. Then, X_i has a Bernouilli distribution with parameter $\mu_{H_i}(\beta)$. As before, we put $Z_m^\beta = \frac{1}{m}\sum_{i=1}^m X_i^\beta$. We note that

- $freq(O)(\beta) = Z_m^\beta(O)$ yields the number of β's in O.
- $\mathbf{E}_{H_1,\ldots H_m}(\beta) = \frac{1}{m}\sum_{i=1}^m \mu_{H_i}(\beta) = \mathbf{E}[Z_m^\beta]$.
- $\mathbf{P}[Z_m^\beta = q] = \mathbf{P}_{H_1,\ldots H_m}[\{O \in (Act^k)^m \mid freq(O)(\beta) = q\}]$.

Now, the desired result follows easily from Proposition 1. □

Then, we consider all sequences $\beta \in Act^k$ at the same time. Then, the probability that the vector of the frequencies $freq(O)$ in a sample O deviates a lot from the expected frequency vector $\mathbf{E}_{H_1,\ldots,H_m}$ is small.

Proposition 3. *Let $\alpha \in (0,1)$, $\varepsilon > 0$ and $k \in \mathbb{N}$. Then there exists an $m' \in \mathbb{N}$ such that for all $m \geq m'$ and all trace distributions $H_1, H_2, \ldots, H_m \in trd(\mathcal{A}, k)$*

$$\mathbf{P}_{H_1,\ldots,H_m}[\{O \in (Act^k)^m \mid freq(O) \in B_\varepsilon(\mathbf{E}_{H_1,\ldots,H_m})] \geq 1 - \alpha.$$

Proof: Let $n = \#\, Act^k$. By Proposition 2, there exists an m' such that for all $m \geq m'$ and all β

$$\mathbf{P}_{H_1,\ldots,H_m}\big[\{O \in (Act^k)^m \mid |freq(O)(\beta) - \mathbf{E}_{H_1,\ldots,H_m}(\beta)| \geq \tfrac{\varepsilon}{n}\big] \leq \tfrac{\alpha}{n}. \qquad (*)$$

But then

$$\begin{aligned}
&\mathbf{P}_{H_1,\ldots,H_m}\left[\{O \mid \mathit{freq}(O) \in B_\varepsilon(\mathbf{E}_{H_1,\ldots,H_m})\}\right] = && \{\text{def } B_\varepsilon\}\\
&\mathbf{P}_{H_1,\ldots,H_m}\left[\{O \mid |\mathit{freq}(O) - \mathbf{E}_{H_1,\ldots,H_m}| < \varepsilon\}\right] \geq && \{\text{see below}\}\\
&\mathbf{P}_{H_1,\ldots,H_m}\left[\{O \mid \forall\beta.|\mathit{freq}(O)(\beta) - \mathbf{E}_{H_1,\ldots,H_m}(\beta)| < \tfrac{\varepsilon}{n}\}\right] = && \{\text{prob. th.}\}\\
&1 - \mathbf{P}_{H_1,\ldots,H_m}\left[\{O \mid \exists\beta.|\mathit{freq}(O)(\beta) - \mathbf{E}_{H_1,\ldots,H_m}(\beta)| \geq \tfrac{\varepsilon}{n}\}\right] \geq && \{\text{prob. th.}\}\\
&1 - \sum_{\beta\in Act^k} \mathbf{P}_{H_1,\ldots,H_m}\left[\{O \mid |\mathit{freq}(O)(\beta) - \mathbf{E}_{H_1,\ldots,H_m}(\beta)| \geq \tfrac{\varepsilon}{n}\}\right] \geq && \{\text{from (*)}\}\\
&1 - n\tfrac{\alpha}{n} = 1 - \alpha.
\end{aligned}$$

The inclusion $\{O \mid |\mathit{freq}(O) - \mathbf{E}_{H_1,\ldots,H_m}| < \varepsilon\} \supseteq \{O \mid \forall\beta.|\mathit{freq}(O)(\beta) - \mathbf{E}_{H_1,\ldots,H_m}(\beta)| < \frac{\varepsilon}{n}\}$ is just another formulation of the fact that an n-dimensional sphere with radius r and center μ contains the n-diminsional hypercube with edge length $\frac{r}{2n}$ and center μ. □

Finally, we need some elementary observations about the function μ_H.

Proposition 4. *1. For all trace distributions $H, K \in \mathit{trd}(\mathcal{A}, k)$, we have $H = K \iff \mu_H = \mu_K$.*
2. For all m and $H_i \in \mathit{trd}(\mathcal{A}, k)$, $\mathbf{E}_{(H_1,\ldots H_m)} = \mu_K$, where $K = \sum_{i=1}^m H_i$.

Proof: Immediately from the definitions. □

Proposition 5. *For every $H \in \mathit{trd}(\mathcal{A}, k)$, μ_H can be written as a convex combination of distributions μ_{H_i}, where H_i is generated by a deterministic adversary. That is, there exists a probability distribution ν over the set $\mathit{Dadv}(A, k)$ such that, for all $\sigma \in \mathit{Act}^k$, $\mu_K(\sigma) = \sum_{D\in\mathcal{D}} \nu(D) \cdot \mu_{\mathit{trd}(D)}(\sigma)$.*

Proof: Immediately from Lemma 2. □

We can now prove the main theorem.

Proof: (of Theorem 1) The "$\Longleftarrow$"-direction follows immediately from the definitions. To prove the "$\Longrightarrow$"-direction, assume that $\mathcal{A} \not\sqsubseteq_{\mathrm{TD}} \mathcal{B}$. We show that $\mathit{Obs}(\mathcal{A}) \not\subseteq \mathit{Obs}(\mathcal{B})$.

By Theorem 2, there exists a k such that $\mathcal{A} \not\sqsubseteq^k_{\mathrm{TD}} \mathcal{B}$, i.e. $\mathit{trd}(\mathcal{A}, k) \not\subseteq \mathit{trd}(\mathcal{B}, k)$. Let H be a trace distribution in $\mathit{trd}(\mathcal{A}, k)$ that is not a trace distribution in $\mathit{trd}(\mathcal{B}, k)$. We write H^m for the sequence $(H_1, H_2, \ldots, H_m)$ with $H_i = H$. Then, Proposition 4(1) concludes that there is no $K \in \mathit{trd}(\mathcal{B}, k)$ such that $\mu_H = \mu_K$. Moreover, Proposition 5 states that the set $\{\mu_K \mid K \in \mathit{trd}(\mathcal{B}, k)\}$ is a polyhedron. Therefore, there is minimal distance $d > 0$ between μ_H and *any* μ_K with K in $\mathit{trd}(\mathcal{B}, k)$.

Consider the trace distribution H. By Proposition 3, we can find $m_{\mathcal{A}}$ such that for all $m \geq m_{\mathcal{A}}$

$$\mathbf{P}_{H^m}[\{O \in (\mathit{Act}^k)^m \mid \mathit{freq}(O) \in B_{\frac{d}{3}}(\mathbf{E}_{H^m})\}] \geq 1 - \alpha.$$

This means that for all $m \geq m_{\mathcal{A}}$, $\mathcal{K}_{H^m} \subseteq B_{\frac{d}{3}}(\mathbf{E}_{H^m})$ and since Proposition 4(1) states that $\mathbf{E}_{H^m} = \mu_H$, we have $\mathcal{K}_{H^m} \subseteq B_{\frac{d}{3}}(\mu_H)$.

Now, consider a sequence trace distributions in $trd(\mathcal{B}, k)$. By Proposition 3, we can find $m_{\mathcal{B}}$ such that for all $m \geq m_{\mathcal{B}}$ and all trace distributions $K_1, K_2, \ldots, K_m$ of $\mathcal{B}$

$$\mathbf{P}_{K_1,\ldots,K_m}[\{O \in (Act^k)^m \mid freq(O) \in B_{\frac{d}{3}}(\mathbf{E}_{K_1,\ldots,K_m})\}] \geq 1 - \alpha.$$

Let $m \geq \max(m_{\mathcal{A}}, m_{\mathcal{B}})$. Then $\mathcal{K}_{K_1,\ldots,K_m} \subseteq B_{\frac{d}{3}}(\mathbf{E}_{K_1,\ldots,K_m})$ and by Proposition 4(2) we have $\mathcal{K}_{K_1,\ldots,K_m} \subseteq B_{\frac{d}{3}}(\mathbf{E}_{K_1,\ldots,K_m}) = B_{\frac{d}{3}}(\mu_K)$, where $K = \sum_{i=1}^{m} K_i$.

Since $|\mu_H - \mu_K| \geq d$, we have $B_{\frac{d}{3}}(\mu_H) \cap B_{\frac{d}{3}}(\mu_K) = \emptyset$, and therefore, and $\mathcal{K}_{H^m} \cap \mathcal{K}_{K_1,\ldots,K_m} = \emptyset$. Hence, none of the observations generated by H^m is an observation of $\mathcal{B}$ and therefore $Obs(\mathcal{A}) \not\subseteq Obs(\mathcal{B})$. □

Acknowledgement The ideas worked out in this paper were presented in preliminary form at the seminar "Probabilistic Methods in Verification", which took place from April 30 – May 5, 2000, in Schloss Dagstuhl, Germany. We thank the organizers, Moshe Vardi, Marta Kwiatkowska, Christoph Meinel and Ulrich Herzog, for inviting us to participate in this inspiring meeting.

References

[BBK87] J.C.M. Baeten, J.A. Bergstra, and J.W. Klop. On the consistency of Koomen's fair abstraction rule. *Theoretical Computer Science*, 51(1/2):129–176, 1987.

[BK86] J.A. Bergstra and J.W. Klop. Verification of an alternating bit protocol by means of process algebra. In W. Bibel and K.P. Jantke, editors, *Math. Methods of Spec. and Synthesis of Software Systems '85, Math. Research 31*, pages 9–23, Berlin, 1986. Akademie-Verlag.

[CDSY99] R. Cleaveland, Z. Dayar, S. A. Smolka, and S. Yuen. Testing preorders for probabilistic processes. *Information and Computation*, 154(2):93–148, 1999.

[Chr90] I. Christoff. Testing equivalence and fully abstract models of probabilistic processes. In J.C.M. Baeten and J.W. Klop, editors, *Proceedings CONCUR 90,* Amsterdam, volume 458 of *Lecture Notes in Computer Science*. Springer-Verlag, 1990.

[Coh80] D.L. Cohn. *Measure Theory*. Birkhaeuser, Boston, 1980.

[DNH84] R. De Nicola and M. Hennessy. Testing equivalences for processes. *Theoretical Computer Science*, 34:83–133, 1984.

[Gla01] R.J. van Glabbeek. The linear time — branching time spectrum I. The semantics of concrete, sequential processes. In J.A. Bergstra, A. Ponse, and S.A. Smolka, editors, *Handbook of Process Algebra*, pages 3–99. North-Holland, 2001.

[GN98] C. Gregorio-Rodrígez and M. Núñez. Denotational semantics for probabilistic refusal testing. In M. Huth and M.Z. Kwiatkowska, editors, *Proc. ProbMIV'98*, volume 22 of *Electronic Notes in Theoretical Computer Science*, 1998.

[JY01] B. Jonsson and W. Yi. Compositional testing preorders for probabilistic processes. *Theoretical Computer Science*, 2001.

[LS91] K.G. Larsen and A. Skou. Bisimulation through probabilistic testing. *Information and Computation*, 94:1–28, 1991.

[Mil80] R. Milner. *A Calculus of Communicating Systems*, volume 92 of *Lecture Notes in Computer Science*. Springer-Verlag, 1980.

[Seg95a] R. Segala. Compositional trace–based semantics for probabilistic automata. In *Proc. CONCUR'95*, volume 962 of *Lecture Notes in Computer Science*, pages 234–248, 1995.

[Seg95b] R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, June 1995. Available as Technical Report MIT/LCS/TR-676.

[Seg96] R. Segala. Testing probabilistic automata. In *Proc. CONCUR'96*, volume 1119 of *Lecture Notes in Computer Science*, pages 299–314, 1996.

[SL95] R. Segala and N.A. Lynch. Probabilistic simulations for probabilistic processes. *Nordic Journal of Computing*, 2(2):250–273, 1995.

[Sto02a] M.I.A. Stoelinga. *Alea jacta est: verification of probabilistic, real-time and parametric systems*. PhD thesis, University of Nijmegen, the Netherlands, April 2002. Available via `http://www.soe.ucsc.edu/~marielle`.

[Sto02b] M.I.A. Stoelinga. An introduction to probabilistic automata. In G. Rozenberg, editor, *EATCS bulletin*, volume 78, 2002.