# The Metric Analogue of Weak Bisimulation for Probabilistic Processes

Josée Desharnais*
Département d'Informatique
Université Laval
Québec, Canada, G1K 7P4

Vineet Gupta
Stratify Inc.
501 Ellis Street
Mountain View CA 94043, USA

Radha Jagadeesan†
Dept. of Computer Science
Loyola University-Lake Shore Campus
Chicago IL 60626, USA

Prakash Panangaden‡
School of Computer Science
McGill University
Montréal, Canada, H3A 2A7

April 15, 2002

**Abstract**

We observe that equivalence is not a robust concept in the presence of numerical information - such as probabilities - in the model. We develop a metric analogue of weak bisimulation in the spirit of our earlier work on metric analogues for strong bisimulation. We give a fixed point characterization of the metric. This makes available coinductive reasoning principles and allows us to prove metric analogues of the usual algebraic laws for process combinators. We also show that quantitative properties of interest are continuous with respect to the metric, which says that if two processes are close in the metric then observable quantitative properties of interest are indeed close. As an important example of this we show that nearby processes have nearby channel capacities - a quantitative measure of their propensity to leak information.

*Research supported by NSERC.
†Research supported by NSF.
‡Research supported in part by NSERC and MITACS.

# 1 Introduction

The starting point and conceptual basis for classical investigations in concurrency are the notions of equivalence and congruence of processes — when can two processes be considered the same and when can they be inter-substituted for each other. Most investigations into probabilistic concurrent processes are also based on equivalences of one kind or another, e.g. [HJ90, JY95, LS91, HS86, BBS95, JS90, CSZ92, Seg95, WSS97, PLS00].

As is now recognized, this style of reasoning is too fragile in the sense of being too dependent on the exact numerical values of the probabilities[1]. This is particularly unfortunate for two reasons: firstly, the probabilities appearing in models really cannot be viewed as exact numbers; rather, they should be viewed as numbers with some error estimate. So, reasoning principles based on the exact value of numbers are of dubious practical value. Secondly, probability distributions over uncountably many states arise in even superficially discrete paradigms. For example, in the presence of recursion or even iteration a binary choice inside a nonterminating computation determines an uncountable tree of possible outcomes with a probability distribution defined on it. Certainly in models of stochastic hybrid systems one has to deal with continuous state spaces. In such continuous-state systems with general probability distributions, an effective handle on the model is recovered by approximating with discrete probabilistic systems [DGJP00]. Clearly, these approximants do not match the continuous state model exactly. While our earlier approximation results show that working with finite-state systems is a good basis for dealing with general systems, it forces us to think about approximate reasoning principles.

Thus, we really want an "approximate" notion of equality of processes. Furthermore, we want a compositional version of such an approximate equality, i.e., we would like to move from the classical study of "exactly inter-substitutable" processes to the study of the more robust notion of "approximately inter-substitutable."

A natural pathway to approximate reasoning is to work with a relaxed notion of truth - for example using the interval $[0, 1]$ as the collection of truth values instead of $\{0, 1\}$. This is precisely Kozen's seminal idea on logics in the context of probability [Koz81, Koz85]— moving from truth-valued boolean functions to real-valued (measurable) functions qua logical formulas. This demands that we take a more sophisticated view of the numerical probability values. This idea of the importance of numerical quantities also guides the search for a relaxation of the notion of equivalence of probabilistic processes. Jou and Smolka [JS90] note that the idea of saying that processes that are close should have probabilities that are close does not yield a transitive relation. This leads them to propose that the correct formulation of the "nearness" notion is via a metric.

Similar reasons motivate the study of Lincoln, Mitchell, Mitchell and Scedrov [LMMS98], our earlier studies of metrics for real-time systems [GHJ97] and our study of metrics for labelled Markov processes and strong bisimulation [DGJP99, DGJP00] and the study of the fine structure of these metrics by van Breugel and Worrell [vBW01b, vBW01a]. In contrast to these investigations, this paper will be carried out in the context of internal nondeterminism and weak bisimulation. The importance of weak bisimulation comes from the need for abstraction. In order to construct larger programs from smaller programs one works with the composition mechanisms of the language. When doing so it is necessary to hide internal actions and work with weak (rather than strong) bisimulation.

[1] Indeed this is a criticism of the use of equivalences for any model that uses quantitative information in a serious way.

## 1.1 An Informal Summary of Results

We define a pseudometric on mixed nondeterministic and probabilistic processes such that zero distance corresponds to weak bisimilarity. Our main results are

- a fixed point characterization of this metric which permits coinductive proofs.
- a real-valued modal logic characterizing the metric.
- we show that several process combinators including parallel composition are non-expansive; this is an analogue of the congruence properties of weak bisimulation.
- a demonstration that nearness in the metric corresponds to nearness in quantitative properties of the processes - for example, we are able to generalize some common theorems about Markov chains and information theory to the case of concurrent Markov chains. This is of interest in its own right.

Working with nondeterminism and probability together (and with the metric analogue of weak bisimulation) forces us to confront some new technical obstacles. More specifically:

- Nondeterminism destroys the additivity property of the probability distributions. For each action we can now have multiple probabilistic distributions — letting the probability of a set of states be the maximum (or minimum) of these probabilities invalidates additivity.
- The algebraic laws of weak bisimulation disable the basic technique used to achieve contractive maps in the work on metric semantics, namely discounting future transitions and making them quantitatively less important than the current transitions. For example, $\tau.P$ is weakly bisimilar to $P$. Clearly, the "corresponding" transitions of $P$ in both $P, \tau.P$ have to be weighted equally, even though they are delayed by a $\tau$ step in $\tau.P$. We overcome this hurdle by providing a novel explicit construction of the required (maximum) fixed point.

**Criteria on Metrics for probabilistic processes** A pseudometric[2] $d$ on processes yields a real number distance for each pair of processes, i.e., $d(P,Q) = d(Q,P)$ and $d(P,R) \leq d(P,Q) + d(Q,R)$. In our approach, we expect that the pseudometric $d$ should satisfy:

- $d(P,Q) = 0 \Leftrightarrow P$ is weakly bisimilar to $Q$

This constraint can be viewed as a soundness criterion on the distance notion identified by the pseudometric.

The interesting cases of the distance function $d$ are when it assigns a non-zero distance between two processes. We expect the pseudometric distance between processes to be roughly determined by the differences in the probabilities of similar transitions. For example, we expect $d(P_{\epsilon_1}, P_{\epsilon_2}) \leq |\epsilon_1 - \epsilon_2|$, where $P_\epsilon = a_{1-\epsilon}.Q$, is the process that does an $a$ with probability $1-\epsilon$ and then behaves like $Q$. Such criteria have good pedigree in the form of the properties of the Hutchinson metric on probability measures [Hut81] on $[0,1]$ defined as $h(\mu,\nu) = \sup\{|\int f d\mu - f d\nu| \mid f$ is a Lipschitz (non-expansive) function on [0,1]$\}$. This metric plays a fundamental role in fractals [Edg98] and also in the work of van Breugel and Worrell [vBW01b, vBW01a].

[2]Really we are working with pseudometrics on processes because $d(P,Q) = 0$ does not mean that $P = Q$.

We define such a pseudometric $d$. We identify $d$ as the maximum fixed point of a certain functional on pseudometrics. This approach exploits the associated good (co)universal properties and the associated coinduction proof rule permits us to recover much of the elegant proof techniques for (weak) bisimilarity. We follow up with an explicit construction of $d$ — we consider a class $\mathcal{F}$ of $[0,1]$ valued functions on states, inspired by the logics for probabilistic transition systems [DEP98]. $\mathcal{F}$ is to be viewed as a real-valued modal logic. These functions enable the definition of $d$ as $d(P,Q) = \sup\{|f(s_P) - f(s_Q)| \mid f \in \mathcal{F}\}$ ($s_p, s_Q$ are start states of $P, Q$ respectively). The key technical tool in both these routes is a linear programming characterization of pseudometric distances inspired by the work of van Breugel and Worrell [vBW01a].

Process algebras provide the link to the desired compositional reasoning about approximate equality in such a pseudometric framework. If $d$ is a pseudometric on processes, we would like *non-expansiveness*, expressed below as a desired compositional proof rule. For every $n$-ary process combinator $C[X_1, \ldots, X_n]$:

$$\frac{d(P_i, Q_i) \leq \epsilon_i,\ i = 1 \ldots n}{d(C(P_1, \ldots, P_n), C(Q_1, \ldots, Q_n)) \leq \sum_i \epsilon_i}.$$

If we set $\epsilon_i = 0$ in the above equation, we get the condition that weak bisimulation is a congruence with respect to the operator $C[\cdot]$.

We show that parallel composition, restriction, CSP style internal choice $\oplus$ and guarded sum $a.P + b.Q$ - where $a, b$ are not $\tau$ - in this language, are non-expansive in the sense of the above proof rule[3]. Our proofs are formally similar to the usual ones for pure labelled transition systems, showing that approximate reasoning (in our sense) is no harder than exact reasoning. Since weak bisimulation is *not* a congruence for CCS + even in purely nondeterministic contexts, such a result does not hold for CCS +. Our extension of the theory to handle the CCS + operator follows the spirit of the definition of observational congruence from weak bisimulation for labelled transition systems, and involves stricter matching of initial $\tau$ transitions. We show that the + operator is non-expansive for this extension.

**An application: Secure Substitution.** The context for our investigations is exemplified by mobile code applications where programs (such as software for submitting income taxes) are downloaded as needed, executed on a trusted host (the home computer), require access to sensitive local data (such as financial information) and yet should not be permitted to "leak information". Thus, we are in the situation where a "mole" may have penetrated inside the system, and the analysis of the system has to account for the fact that this mole may be attempting to leak information to the outside world. Our interest is in *secure substitution* — when can one component (say $C_2$) can be substituted for another (say $C_1$) in the system without affecting the usual observations and the capability of the mole to leak information.

The definition of channel capacity in information theory offers a route to quantifying "information leakage". Channel capacity attempts to characterize the maximum amount of information that can be transmitted by a communication channel. In the context that we are considering, we are interested in measuring the information flow of the channel that is identified by the interface of the downloaded untrustworthy program. The (sequences of) labels on the internal interactions of

[3]Such a language is a probabilistic variant of languages for which weak bisimulation is known to be a congruence, for example languages in the ISOS format of Ulidowski [Uli92, Uli94].

this component with the system constitute the input symbols of the channel. The "mole" is viewed as attempting to leak information about the sequences of these input symbols by interacting with the outside world, the (sequences of) labels on these interactions constitute the output symbols. The outside observer is attempting to deduce input traces while only viewing the output traces.

The first technical obstacle that we need to overcome is that in the presence of nondeterminism, we are working with sets of distributions rather than single distributions. We model worst case assumptions by permitting the mole to control the (nondeterministic) scheduler, by permitting the mole to serve as (a possibly probabilistic) oracle to guide the scheduler[4]. Our definition of channel capacity captures these worst case assumptions.

Given this modelling assumption, we describe the following results. Using basic results from information theory and exploiting the non-expansive properties, we show that there is a constant $k$ only dependent on $C_1$ such that: $d(C_1, C_2) < \epsilon$ implies that the difference in the channel capacities of $C_1, C_2$ is at most $k\epsilon$. Thus, the channel capacity is a continuous function of the pseudometric. This analysis provides an à posteriori justification for the criteria on pseudometrics discussed earlier. This result assumes particular force in conjunction with the non-expansiveness of the basic process combinators and gives us the basic rudiments of a methodology to determine if a component can be securely substituted for another.

**An application: Quantitative reasoning** Very often one wants not just logical properties but quantitative properties of systems. For example, one more likely wants to know the average size of a queue rather than, say, the fact that it is nonempty with probability one. Similarly one often wants to know large deviation probabilities away from the mean.

Consider a queueing system with multiple servers and a stream of customers arriving with a given distribution. There are two basic strategies: one can have all the customers in a single queue with each server taking the next customer from the common queue (as is typical in banks) or one could have a queue for each server with the customer committing to a queue upon entry to the system (as in supermarkets). Given distributions for arrival times and service times we would like to know if these systems are close in their quantitative properties such as expected waiting times, average throughput. Clearly the two systems are quite different in terms of their exact logical properties and one would get no sense of how different they are if we were to compare them on that basis. In terms of modelling with a probabilistic process algebra, we cannot really make comparisons using a metric analogue of strong bisimulation because the transitions are very different right away. However, if we hide the internal details of how processes are allocated to different queues and discount these internal transitions we would have the basis for a quantitative comparison.

It might appear, at first sight, that abstracting from the internal transitions would destroy the temporal information. However, time really passes while waiting in the queues and while interacting with the servers. One can set up a timer process to keep track of the times. What one is really abstracting over are the internal transitions that govern the interactions with the queue(s). It is not hard to show that such average quantitative properties are continuous which means that if two

[4]Our modelling differs from the study of Lincoln, Mitchell, Mitchell and Scedrov [LMMS98] in this aspect. They use a probabilistic process algebra to explore the power of arbitrary probabilistic polynomial time processes in the context of security protocols. In their model, the system is probabilistically determinate and the adversary is "outside" the system and thus has no useful control over the scheduler of the system. Rather, it derives its power from the ability to intercept and alter messages.

systems are metrically close their quantitative properties will be close as well. Thus the metric distance gives a good measure of how closely the quantitative properties of two systems match.

## 2 Definitions

We begin with a review of the underlying framework — our definitions are adapted from [PLS00]. We work in the context of the "alternating model" for labelled concurrent Markov chains [Han94], labelled transition systems with non-determinism and probability.

**Definition 2.1** *A labelled concurrent Markov chain (henceforth LCMC), is a tuple* $\mathcal{K} = (K, \mathtt{Act}, \longrightarrow, k_0)$, *where*
*(1)* $K = K_p \cup K_n$, *a countable set, is partitioned into the probabilistic states,* $K_p$, *and the nondeterministic states* $K_n$. $k_0$ *is the start state.*
*(2)* $\mathtt{Act}$ *is a finite set of action symbols that contains a special action* $\tau$.
*(3) The transition relation* $\longrightarrow = \longrightarrow_p \cup \longrightarrow_n$ *is partitioned into probabilistic and nondeterministic transitions.* $\longrightarrow_n \subseteq K_n \times \mathtt{Act} \times K_p$ *is* image-finite, *i.e. for each* $s \in K_n$ *and* $a \in \mathtt{Act}$, *the set* $\{s' \in K_p \mid s \xrightarrow{a} s'\}$ *is finite.* $\longrightarrow_p \subseteq K_p \times (0,1] \times K_n$ *satisfies that for each* $s \in K_p$ $\sum_{(s,\pi,t)\in\longrightarrow_p} \pi \leq 1$.

Thus, transitions from nondeterministic states are finitely branching[5] and labelled. Transitions from probabilistic states are un-labelled and are associated with numbers that are interpreted as probabilities. *In this paper, we will work with finite state systems,* the domain of definition of the weak bisimulation definition of Lee, Philippou and Sokolsky [PLS00].

Every probabilistic state $s$ induces a distribution $P$ on $K$ given by $P(t) = \sum_{(s,\pi,t)\in\longrightarrow_p} \pi$ for every $t \in K$. We sometimes write $s \rightarrow_p P$ to emphasize this distribution.

The LCMC model does not need to be strictly alternating. One can work with a model that only restricts states to be either purely nondeterministic or purely probabilistic and does not enforce strict alternation. We discuss this variant at the end of this section.

**Example 2.2 (Constructions on LCMCs)**

*Given LCMC* $\mathcal{K}$, *we construct the LCMC for restriction,* $\mathcal{K}\backslash a$, *for some* $a \neq \tau$. *Formally,* $\mathcal{K}\backslash a = (K, \mathtt{Act}, \longrightarrow', k)$, *where* $\longrightarrow'$ *has all probabilistic transitions from* $\longrightarrow$ *and all nondeterministic transitions whose labels are not* $a$.

*Given LCMCs* $\mathcal{K}_1, \mathcal{K}_2$, *the LCMC for internal choice,* $\mathcal{K}_1 \oplus \mathcal{K}_2$ *is constructed by adding a new start state with* $\tau$ *transitions to the start states of* $\mathcal{K}_1, \mathcal{K}_2$. *Formally, assume that* $K_1$ *and* $K_2$ *have disjoint sets of states and their start states are probabilistic and let* $\longrightarrow_i$ *be the transition relations of* $\mathcal{K}_i$, $i = 1, 2$. *Then,* $\mathcal{K}_1 \oplus \mathcal{K}_2 = (K_1 \cup K_2 \cup \{s\}, \mathtt{Act}, \longrightarrow_1 \cup \longrightarrow_2 \cup \{s \xrightarrow{\tau} k_1, s \xrightarrow{\tau} k_s\}, s)$.

*Given LCMCs* $\mathcal{K}_i$, *the LCMC* $\sum_i a_i.\mathcal{K}_i$, *where* $a_i, i = 1, \ldots n$ *is a finite collection of labels, is constructed by adding an* $a_i$*-transition to* $\mathcal{K}_i$. *Assume that the* $\mathcal{K}_i$*'s have pairwise disjoint state spaces and let* $\longrightarrow_i$ *be their transition relations and* $s_i$ *their start states, which are assumed to be probabilistic. Let* $s \notin \cup K_i$. *Then,* $\sum_i a_i.\mathcal{K}_i = (\cup K_i \cup \{s\}, \mathtt{Act}, \bigcup_i \longrightarrow_i \ \cup \{s \xrightarrow{a_i} s_i \mid i = 1, \ldots n\}, s)$.

*Given finitely many LCMCs* $\mathcal{K}_i$ *and* $\pi_i$ *such that* $\sum_i \pi_i \leq 1$, *the LCMC for probabilistic choice,* $\sum_i(\pi_i, \mathcal{K}_i)$ *is constructed as follows. Assume that the* $\mathcal{K}_i$*'s have pairwise disjoint state spaces and let* $\longrightarrow_i$ *be their transition relations and* $s_i$ *their start states, which are assumed to be non-deterministic. Let* $s \notin \cup K_i$. *Then,* $\sum_i(\pi_i, \mathcal{K}_i) = (\cup K_i \cup \{s\}, \mathtt{Act}, \{s \xrightarrow{\pi_i} s_i \mid i\} \cup \bigcup_i \longrightarrow_i, s)$.

[5] Since we have a finite action set, "image finite" and "finitely branching" the same thing.

*Given LCMCs $\mathcal{K}_1, \mathcal{K}_2$, the LCMC $\mathcal{K}_1 + \mathcal{K}_2$ is constructed as follows. Formally, assume that $K_1$ and $K_2$ have disjoint sets of states except for the start state $k$ which is nondeterministic and let $\longrightarrow_i$ be the transition relations of $\mathcal{K}_i$, $i = 1, 2$. Then, $\mathcal{K}_1 + \mathcal{K}_2 = (K_1 \cup K_2, \mathtt{Act}, \longrightarrow_1 \cup \longrightarrow_2, k)$.*

We use some notation for sequences (of states or transitions). We use $\varepsilon$ for the empty sequence and $\cdot$ for concatenation. Every sequence, say $\sigma$, of transitions has an associated probability $\mathtt{prob}(\sigma)$, obtained by multiplying the probabilities occurring on the path. Thus, we attribute 1 to a nondeterministic transition in a path, and multiply in the probability of a probabilistic transition. Similarly, every sequence $\sigma$ of transitions has an associated weak sequence of labels $\mathtt{Weak}(\sigma) \in (\mathtt{Act} - \{\tau\})^*$, obtained by removing the $\tau$'s. Thus probabilistic transitions and nondeterministic transitions with label $\tau$ do not contribute to the weak label.

We define *computations* of an LCMC as transition trees obtained by unfolding the LCMC from the root, resolving the nondeterministic choices (i.e. each nondeterministic state has at most one transition coming out of it) and taking all probabilistic choices at a probabilistic state. A computation can thus be viewed as a purely (sub)probabilistic labelled Markov chain. We elide standard definitions of trees.

**Definition 2.3** *Let $\mathcal{K}$ be a LCMC, $a \in \mathtt{Act}$. An $a$-computation from $s \in K$ is a computation such that every path from the root has weak label $a$ or $\varepsilon$.*

We allow a $\varepsilon$ label in order to allow paths that could be extended to have an $a$-transition in an extended $a$-computation.

Each computation induces a distribution on its leaf states in the standard way — the probability of a leaf node is the probability of the (unique) path going to it. Here we insist that the paths contributing to the distribution do have an $a$-label.

**Definition 2.4** *Let $\mathcal{K}$ be a LCMC, $s \in K$, and let $Q$ be a distribution on states.*
*We write $s \stackrel{a}{\Rightarrow} Q$, if there is an $a$-computation such that for all $s_i \in K$, $Q(s_i) \leq \sum_\sigma \mathtt{prob}(\sigma)$ where the summation is taken over paths $\sigma$ with weak label $a$ that start in $s$ and end in the leaf $s_i$.*

*We extend this notation to linear combinations of $Q$'s. Let $s_i \stackrel{a}{\Rightarrow} Q_i$, and let $\sum_i \lambda_i = 1$. Then, we write: $\sum_i \lambda_i \times (s_i \stackrel{a}{\Rightarrow} Q_i)$. In the special case where all $s_i = s$, we write $s \stackrel{a}{\Rightarrow} \sum_i \lambda_i \times Q_i$.*

We sometimes refer to the transitions $s \stackrel{a}{\Rightarrow} Q$ that are not linear combinations as "basic" to distinguish them from transitions that are of the form $\sum_i \lambda_i \times (s_i \stackrel{a}{\Rightarrow} Q_i)$. For $s \neq t$, the notation $\lambda \times (s \stackrel{a}{\Rightarrow} P) + (1-\lambda) \times (t \stackrel{a}{\Rightarrow} Q)$ is merely notational convenience. However, $s \stackrel{a}{\Rightarrow} [\lambda \times P + (1-\lambda) \times Q]$ is reminiscent of the randomized schedulers [Seg95].

We define the "probability" from a state $s$ to a subset of states via a path with weak label $a$ by taking the supremum over all possible computations.

**Definition 2.5** *Let $\mathcal{K}$ be a LCMC, $s \in K, E \subseteq K$. Then, the probability of going from $s$ to $E$ via $a$, denoted by $P(s, a, E)$, is defined as:*

$$P(s, a, E) = \sup\{\sum_{s \in E} Q(s) \mid s \stackrel{a}{\Rightarrow} Q\}.$$

The supremum in this definition is the source of the subtlety of weak bisimulation — $P(s, a, .)$ does not satisfy additivity. The computation yielding the maximum probability is constructed by choosing at every non-deterministic state, the transition that maximizes the probability. Thus, we get:

**Lemma 2.6** *$P(s,a,E) = \sum_{t \in E} Q(t)$ for some $s \stackrel{a}{\Rightarrow} Q$.*

**Proof.** We provide a recipe to pick out the computation that reaches $P(s,a,E)$. Let $Q_i$ be such that $s \stackrel{a}{\Rightarrow} Q_i$. Clearly, if there are only finitely many of these there is nothing to prove. According let us suppose that there are infinitely many such $Q_i$. Let $p_i = Q_i(E), p = \sup p_i$. Since the nondeterministic branching at any state is finite, there is at least one branch with a given source state that is chosen by infinitely many $Q_i$. The computation constructed by making these choices must give a probability of $p$ or one of the finitely many other computations attains the probabilty $p$. ∎

We are now ready to define weak bisimulation. Our presentation of this definition is different from [PLS00] — we will prove later that it does indeed coincide with [PLS00] for finite state systems. We consider equivalence relations on the set of states. Given an equivalence relation $R$, we say a set $E$ is $R$-closed if $E = \{s \mid \exists t \in E \text{ such that } tRs\}$.

**Definition 2.7** *An equivalence relation $R$ on $K$ is a weak bisimulation if for all $s,t \in K$ such that $s\ R\ t$ and all $R$-closed $E \subseteq K$, we have:*

$$(\forall a \in \mathtt{Act})\ [P(s,a,E) = P(t,a,E)].$$

*There is a maximum weak bisimulation, denoted by $\approx$.*

The equational laws supported by this definition extend the usual ones for nondeterministic labelled transition systems or purely probabilistic transition systems. Indeed, the usual relations that witness the bisimulation are carried over essentially unchanged, for example, $\tau.\mathcal{K} \approx \mathcal{K}$, and unfolding a LCMC yields a weakly bisimilar system. See [BS01] for a full axiomatization of equational laws for finite processes.

**Minor extensions to the model.** The LCMC model does not need to be strictly alternating. One can work with a model that restricts states to be either purely nondeterministic or purely probabilistic. Any such transition system $\mathcal{U} = (U, \mathtt{Act}, \longrightarrow, u_0)$ has a (weak) bisimulation preserving translation into $\mathcal{K} = (K, \mathtt{Act}, \longrightarrow, k_0)$, a strictly alternating transition system as follows. The states $K = U^p \cup U^n$ are a disjoint union of two copies of the states of $U$. For all $s \in U$ such that $s$ has only nondeterministic transitions, define $s^p \stackrel{1}{\rightarrow} s^n$ and $s^n \stackrel{a}{\rightarrow} t^p$ if $s \stackrel{a}{\rightarrow} t$ in $\mathcal{U}$. Similarly, for all $s \in U$ such that $s$ has only probabilistic transitions, define $s^n \stackrel{\tau}{\rightarrow} s^p$ and $s^p \stackrel{\pi}{\rightarrow} t^n$ if $s \stackrel{\pi}{\rightarrow} t$ in $\mathcal{U}$. There is clearly a weak bisimulation relating $\mathcal{U}$ and $\mathcal{K}$.

**Coincidence with the definition of Lee, Philippou and Sokolsky.** Our definition of weak bisimulation, Definition 2.7 coincides with [PLS00]. The key structural properties exploited in the proof that our definition implies their definition are:

- If $t$ is a nondeterministic state, and $s$ is a probabilistic state, such that $t$ is weakly bisimilar to $s$, then there is a tau transition from $t$ to some $t'$ such that $t'$ is weakly bisimilar to $s$.
- A linear programming criterion that foreshadows our later development for pseudometrics.

**Theorem 2.8** *Given an LCMC which satisfies the property that the total of all the probabilities from any probabilistic state is 1, if states $s$ and $t$ in it are bisimilar then they are bisimilar according to the definition of Lee, Philippou and Sokolsky [PLS00].*

**Proof.** See Appendix. ∎

The converse of this theorem is also true. The converse relies on the ability to mimic computations at bisimilar states.

**Lemma 2.9** *Let $s \approx t$. Let $s \stackrel{a}{\Rightarrow} P$. Then, there exists $t \stackrel{a}{\Rightarrow} Q$ such that for all states $u$: $P([u]) = Q[u])$.*

**Proof.** The result for linear combinations follows from the result for basic transitions $s \stackrel{a}{\Rightarrow} P$. We prove this case below.

Formally, let us be given $s \stackrel{a}{\Rightarrow} P$ and a matching transition $t \stackrel{a}{\Rightarrow} Q$ such that $P, Q$ agree, ie. for all states $u$: $P([u]) = Q[u])$.

Let $s \stackrel{a}{\Rightarrow} P'$ extend $s \stackrel{a}{\Rightarrow} P$ by adding a one-step transition at one of the leaves of $P$, say $u$, by a nondeterministic transition $u \stackrel{b}{\rightarrow} u'$. Let $P(u) = p$. In this case, consider $t \stackrel{a}{\Rightarrow} Q'$, the extension of $Q$ by matching transitions $v \stackrel{b}{\Rightarrow} Q_i$ from all the $v \approx u$ that are leaves. The required transition from $t$ is obtained by a linear combination $t \Rightarrow [\lambda \times Q' + (1 - \lambda) \times Q]$, where $\lambda = p/P([u])$.

The case when $s \stackrel{a}{\Rightarrow} P'$ extends $s \stackrel{a}{\Rightarrow} P$ by adding a one-step probabilistic transition at one of the leaves of $P$ is similar and is omitted. ∎

This lemma provides the tools to establish that the definition of [PLS00] implies our definition.

**Theorem 2.10** *Given a finite state LCMC which satisfies the property that the total of all the probabilities from any probabilistic state is 1, if states $s$ and $t$ in it are bisimilar according to the definition of Lee, Philippou and Sokolsky [PLS00], then they are bisimilar.*

**Proof.**

Let $s, t$ be bisimilar by the definition of Lee, Philippou and Sokolsky. To show that $s \approx t$, it suffices to show that for every finite $L$-computation rooted at $s$, say $C$, there is a $L$-computation rooted at $t$ that assigns the same probabilities to the leaves of $C$. This follows easily from the lemma 2.9. ∎

# 3 The Pseudometric as a maximum fixed point

Our first presentation of the pseudometric is as a maximum fixed point of a certain functional.

We fix an $LCMC$ and consider pseudometrics on its set of states.

**Definition 3.1** *$\mathcal{M}$ is the class of 1-bounded pseudometrics on states with the ordering*

$$m_1 \preceq m_2 \text{ if } (\forall s, t)\ [m_1(s, t) \geq m_2(s, t)]$$

**Lemma 3.2** *$(\mathcal{M}, \preceq)$ is a complete lattice.*

**Proof.** The least element is given by: $\perp(s, t) = 0$ if $s = t, 1$ otherwise. The top element is given by $(\forall s, t)\top(s, t) = 0$. Greatest lower bounds are given by: $(\sqcap\{m_i\})(s, t) = \sup_i m_i(s, t)$. Note that:

$$\begin{aligned}
(\sqcap\{m_i\}(s, t) &= \sup_i m_i(s, t) \\
&\leq \sup_i [m_i(s, u) + m_i(t, u)] \\
&\leq \sup_i [m_i(s, u)] + \sup_i [m_i(t, u)] \\
&\leq (\sqcap\{m_i\}(s, u) + (\sqcap\{m_i\}(u, t)
\end{aligned}$$

∎

$m \in \mathcal{M}$ is extended to distributions on sets of states as follows. The definition is based on the Hutchinson metric on probability measures — we have merely simplified the definition for our context of discrete finite state distributions.

**Definition 3.3** *Let $m \in \mathcal{M}$. Let $P, Q$ be distributions on states such that the total mass of $P$ is not less than the total mass of $Q$. Then $m(P, Q)$ is given by the solution to the following linear program:*

$$\begin{aligned} & \max \sum_i (P(s_i) - Q(s_i))a_i \\ \textit{subject to}: \ & \forall i. 0 \leq a_i \leq 1 \\ & \forall i, j.\ a_i - a_j \leq m(s_i, s_j). \end{aligned}$$

We need the constraints $a_i \leq 1$ if the distributions do not have equal total probability, without them the maximum is unbounded. Equivalently, following the analysis of [vBW01a], $m(P, Q)$ is given by the solution to the following dual linear program:

$$\begin{aligned} \min & \sum_{i,j} l_{ij} m(s_i, s_j) + \sum_i x_i + \sum_j y_j \\ \text{subject to}: \ & \forall i. \textstyle\sum_j l_{ij} + x_i = P(s_i) \\ & \forall j. \textstyle\sum_i l_{ij} + y_j = Q(s_j) \\ & \forall i, j.\ l_{ij}, x_i, y_j \geq 0. \end{aligned}$$

The following lemma shows that this extension to distributions satisfies the triangle inequality and is consistent with the ordering on pseudometrics. The proof of the first item is an elementary exercise using the primal linear program. The proof of the second item uses the dual linear program — every solution to the (dual) linear program $m'(P, Q)$ is also a solution to the (dual) linear program for $m(P, Q)$.

**Lemma 3.4**

- *Let $m \in \mathcal{M}$. Then, $(\forall P, Q, R)\ m(P, Q) \leq m(P, R) + m(Q, R)$.*
- *Let $m, m' \in \mathcal{M}$ such that $m \preceq m'$. Then, for all distributions on states $P, Q$, $m(P, Q) \geq m'(P, Q)$.*

**Proof.**

- For the first item, we proceed as follows. We first prove that if total mass of $P$ is less than the total mass of $Q$, then

$$\max \sum_i (P(s_i) - Q(s_i))a_i < \max \sum_i (Q(s_i) - P(s_i))a_i$$

where the $a_i$ are subject to the usual constraints $0 \leq a_i \leq 1, a_i - a_j \leq m(s_i, s_j)$.

$$\begin{aligned} \sum_i (P(s_i) - Q(s_i))a_i &= \sum_i (Q(s_i) - P(s_i))(1 - a_i) - \sum_i Q(s_i) + \sum_i P(s_i) \\ &< \sum_i (Q(s_i) - P(s_i))(1 - a_i) \end{aligned}$$

Now $b_i = 1 - a_i$ also satisfy the constraints on the $a_i$ above, proving the result.

To prove the triangle inequality, given distributions $P, Q, R$, $\sum_i (P(s_i) - Q(s_i))a_i = \sum_i (P(s_i) - R(s_i))a_i + \sum_i (R(s_i) - Q(s_i))a_i$. Taking the maximum over the $a_i$ for the left side, we get $m(P,Q) \leq m(P,R) + m(R,Q)$.

- For the second item, note that every solution to the linear program defining $m'(P,Q)$ is also a solution to the linear program defining $m(P,Q)$. So, the maximum value $m(P,Q)$ is $\geq m'(P,Q)$.

∎

The dual linear program is a key tool to move from distributions to states in the following sense. Given close-by distributions $P$, $Q$, the dual linear program permits us to construct a matching of states (that may include "splitting" of the probabilities assigned by $P, Q$), in such a way that exactly the distance between $P, Q$ can be recovered.

**Lemma 3.5** *Let $P$ and $Q$ be probability distributions on a set of states $K$. Let $P_1$ and $P_2$ be such that: $P = P_1 + P_2$. Then, there exist $Q_1, Q_2$, such that $Q_1 + Q_2 = Q$ and*

$$m(P,Q) = m(P_1,Q_1) + m(P_2,Q_2).$$

**Proof.** Let $\{l_{ij}\}$, $\{x_i\}$ and $\{y_j\}$ be such that the minimum is attained in the dual linear program above: that is, $m(P,Q) = \sum_{i,j} l_{ij} m(s_i,s_j) + \sum_i x_i + \sum_j y_j$. Define: $Q_k(s_j) = \sum_i l_{ij} P_k(s_i)/P(s_i) + y_j P_k(K)/P(K)$, for $k = 1, 2$. Then, $Q_1 + Q_2 = Q$.

Furthermore, setting $l^1_{ij} = [P_1(s_i)/P(s_i)]l_{ij}, y^1_j = [P_1(K)/P(K)]y_j, x^1_i = [P_1(s_i)/P(s_i)]x_i$, we get:

$$\begin{aligned}
\sum_i l^1_{ij} + y^1_j &= Q_1(s_j) \\
\sum_j l^1_{ij} + x^1_i &= \sum_j [P_1(s_i)/P(s_i)]l_{ij} + [P_1(s_i)/P(s_i)]x_i \\
&= [P_1(s_i)/P(s_i)](\sum_j l_{ij} + x_i) \\
&= [P_1(s_i)/P(s_i)]P(s_i) \\
&= P_1(s_i).
\end{aligned}$$

Thus, $m(P_1,Q_1) \leq \sum_{i,j} l^1_{ij} m(s_i,s_j)$. Similarly, $m(P_2,Q_2) \leq \sum_{i,j} l^2_{ij} m(s_i,s_j)$.

Thus:

$$\begin{aligned}
m(P_1,Q_1) + m(P_2,Q_2) &\leq \sum_{i,j} l^1_{ij} m(s_i,s_j) + \sum_i x^1_i + \sum_j y^1_j + \sum_{i,j} l^2_{ij} m(s_i,s_j) + \sum_i x^2_i + \sum_j y^2_j \\
&= \sum_{i,j} [l^1_{ij} + l^2_{ij}] m(s_i,s_j) + \sum_i (x^1_i + x^2_i) + \sum_j (y^1_j + y^2_j) \\
&= \sum_{i,j} l_{ij} * m(s_i,s_j) + \sum_i x_i + \sum_j y_j \\
&= m(P,Q).
\end{aligned}$$

To show that $m(P_1, Q_1) + m(P_2, Q_2) \geq m(P, Q)$, consider the $a_i$ which achieves the maximum in the definition of $m(P, Q)$. Then (note that if $P(K) \geq Q(K)$, then $P_i(K) \geq Q_i(K)$)

$$\begin{aligned} m(P_1, Q_1) + m(P_2, Q_2) &\geq \sum_i (P_1(s_i) - Q_1(s_i))a_i + \sum_i (P_2(s_i) - Q_2(s_i))a_i \\ &= \sum_i (P(s_i) - Q(s_i))a_i \\ &= m(P, Q). \end{aligned}$$

∎

As a straightforward corollary, we get a complete matching on individual states.

**Corollary 3.6** *Given distributions $P, Q$ there exist distributions $P_i, Q_i$ such that*

- *$P_i$ are point distributions that are non-zero at only one state.*
- *$P = \sum P_i$, $Q = \sum Q_i$*
- *$m(P, Q) = \sum m(P_i, Q_i)$.*

We now define a functional $F$ on $\mathcal{M}$ that closely resembles the usual functional for weak bisimulation.

**Definition 3.7** *Define $F$, a functional on $\mathcal{M}$ as follows. $F(m)(s, t) < \epsilon$ if:*

- $(\forall s \stackrel{a}{\Rightarrow} P)\ (\exists t \stackrel{a}{\Rightarrow} Q)\ [m(P, Q) < \epsilon]$.
- $(\forall t \stackrel{a}{\Rightarrow} Q)\ (s \stackrel{a}{\Rightarrow} P)\ [m(P, Q) < \epsilon]$.

$F(m)$ is well-defined because of the following lemma. The triangle inequality on $F(m)$ follows from the triangle inequality on $m$ extended to distributions.

**Lemma 3.8** *$F(m)$ is a pseudometric given by:*

$$F(m)(s, t) = \max(\max_{a \in \mathtt{Act}} \sup_{s \stackrel{a}{\Rightarrow} P} \inf_{t \stackrel{a}{\Rightarrow} Q} m(P, Q), \max_{a \in \mathtt{Act}} \sup_{t \stackrel{a}{\Rightarrow} Q} \inf_{s \stackrel{a}{\Rightarrow} P} m(P, Q)).$$

**Proof.** We prove the triangle inequality. Let $F(m)(s, t) < \epsilon_1, F(m)(t, u) < \epsilon_2$. Let $s \stackrel{a}{\Rightarrow} P$ . Since $F(m)(s, t) < \epsilon_1$, there exists a $t \stackrel{a}{\Rightarrow} Q$ such that $m(P, Q) < \epsilon_1$. Since $F(m)(t, u) < \epsilon_2$, there exists a $u \stackrel{a}{\Rightarrow} R$ such that $m(Q, R) < \epsilon_2$. From the triangle inequality on $m$ (extended to distributions), $m(P, R) < \epsilon_1 + \epsilon_2$. ∎

$F$ is monotone on $\mathcal{M}$.

**Lemma 3.9** *$F$ is monotone on $\mathcal{M}$.*

**Proof.** Let $m_2 \preceq m_1$. We need to show that $F(m_2) \preceq F(m_1)$, i.e., $(\forall s, t) F(m_1)(s, t) \leq F(m_2)(s, t)$.

Let $F(m_2)(s, t) < \epsilon$. Then,

- For all transitions $s \stackrel{a}{\Rightarrow} P$, there exists $t \stackrel{a}{\rightarrow} Q$ such that $m_2(P, Q) < \epsilon$.
- For all transitions $t \stackrel{a}{\Rightarrow} Q$, there exists a transition $s \stackrel{a}{\Rightarrow} P$ such that $m_2(P, Q) < \epsilon$.

Since, $m_2(P, Q) \geq m_1(P, Q)$, $F(m_1)(s, t) < \epsilon$ as required. ∎

Using Tarski's fixed point theorem, $F$ has a maximum fixed point. Using the image finiteness of LCMC, we can show that the closure ordinal of $F$ is $\omega$. The proof proceeds standardly, by showing that the maximum fixed point $m$ is given by $m = \sqcup_i m_i$, where $m_0 = \top$ and $m_{i+1} = F(m_i)$.

**Lemma 3.10** *The closure ordinal of $F$ is $\omega$.*

**Proof.** Let $m(s,t) < \epsilon$. Let $s \stackrel{a}{\Rightarrow} P$. Then, for each $m_i$ there is a $Q_i$ such that $t \stackrel{a}{\Rightarrow} Q_i$ and $m_i(P, Q_i) < \epsilon$. Since the LCMC is image finite, there is a $Q_i$ (say $Q$) such that for all but finitely many $i$, $t \stackrel{a}{\Rightarrow} Q$ and $m_i(P,Q) < \epsilon$. ∎

The maximal fixed point of $F$ is sound with respect to bisimulation. The forward implication of the proof uses the pseudometric $m'$ defined as: $m'(s,t) = 0$ iff s and t are bisimilar and 1 otherwise. $m'$ satisfies $m' \preceq F(m')$. The converse proceeds by showing that the equivalence relation $R$ induced by 0 distance is a bisimulation.

**Lemma 3.11** *$s \approx t \Leftrightarrow m(s,t) = 0$, where $m$ is the maximum fixed point of $F$.*

**Proof.** For the forward direction, consider the pseudometric $m'$ defined as: $m'(s,t) = 0$ iff $s$ and $t$ are bisimilar and 1 otherwise. Using lemma 2.9, $F(m') \leq m'$.

For the converse, consider the relation R induced by 0 metric distance. Clearly, this is an equivalence relation. We show that this equivalence relation is a bisimulation. Let $m(s,t) = 0$. Let $P(s,a,E) = p$, for some $R$-closed set $E$. Then, using lemma 2.6, there is a computation with root $s$ that assigns the maximum probability $p$ to $E$. Thus, there exists a transition $s \stackrel{a}{\Rightarrow} P$, such that $P(E) = \sum_{s \in E} P(s) = p$. Given any $\epsilon > 0$, since $m(s,t) = 0$, we get a transition (or a linear combination of transitions) $t \stackrel{a}{\Rightarrow} Q$ such that $m(P,Q) < d\epsilon/n^2$, where $n$ is the number of states and $d = \min\{m(s_i,s_j) | s_i, s_j$ are states in the system $, m(s_i,s_j) > 0\}$. Then in the dual linear program, $l_{ij} < \epsilon/n^2$ for all $i,j$ such that $m(s_i,s_j) > 0$, and so are $x_i$ and $y_j$. Now $|P(s_i) - Q(s_i)| = \sum_j l_{ij} + x_i - \sum_k l_{ki} - y_i < \epsilon/n$ as $l_{ii}$ cancels out, and there are at most $n$ positive and $n$ negative terms on the RHS. Thus $|P(E) - Q(E)| < \epsilon$, so $Q$ witnesses $p - \epsilon \leq P(t,a,E)$. This calculation holds for any $\epsilon$, ensuring $p \leq P(t,a,E)$. ∎

Finally, we show that it suffices to consider one-step transitions for the hypothesis in the definition of the functional $F$. $F'$ demands only matching of "one-step" transitions. However, by using corollary 3.6 to move from distributions to states, we can show that $F'$ enforces the matching required by $F$. Let $\delta_s$ means the Dirac measure at $s$

**Lemma 3.12** *Define $F' : \mathcal{M} \to \mathcal{M}$ as follows: $F'(m)(s,t) < \epsilon$ if:*

- $(\forall s \stackrel{a}{\rightarrow} s')\ (\exists t \stackrel{a}{\Rightarrow} Q)\ [m(\delta_{s'}, Q) < \epsilon]$. $(\forall s \rightarrow_p P)\ (\exists t \stackrel{\tau}{\Rightarrow} Q)\ [m(P,Q) < \epsilon]$.
- $(\forall t \stackrel{a}{\rightarrow} t')\ (\exists s \stackrel{a}{\Rightarrow} P)\ [m(P, \delta_{t'}) < \epsilon]$. $(\forall t \rightarrow_p Q)\ (\exists s \stackrel{\tau}{\Rightarrow} P)\ [m(P,Q) < \epsilon]$.

*Then, the maximum fixed points of $F, F'$ coincide.*

**Proof.** Clearly $F$ imposes more conditions than $F'$, since the one step computations considered are also weak transitions. So, any fixed point of $F$ is a fixed point of $F'$.

For the converse, given a fixed point of $F'$, and a transition $s \stackrel{a}{\Rightarrow} P$, we need to construct a matching transition $t \stackrel{a}{\Rightarrow} Q$. The result for transitions of the form $s \stackrel{a}{\Rightarrow} \sum_i P_i$ follows from that for basic transitions $s \stackrel{a}{\Rightarrow} P$ by considering linear combinations.

Let $m(s,t) < \epsilon$. For basic transitions $s \stackrel{a}{\Rightarrow} P$, we build the required transition $t \stackrel{a}{\Rightarrow} Q$ by mimicking each step in $s \stackrel{a}{\Rightarrow} P$. Formally, let us be given $s \stackrel{a}{\Rightarrow} P$ and a matching transition $t \stackrel{a}{\Rightarrow} Q$ such that $d(P,Q) < \epsilon$. Let $s \stackrel{a}{\Rightarrow} P'$ extend $s \stackrel{a}{\Rightarrow} P$ by adding a one-step transition at one of the leaves of $P$, say $u$, by a nondeterministic transition $u \stackrel{b}{\rightarrow} u'$[6]. We show how to construct a matching $t \stackrel{a}{\Rightarrow} Q'$ such that $m(P',Q') < \epsilon$. Let $s_1, \ldots, s_n$ be all the states in the two transition systems. Let $P_i, i = 1 \ldots n$ be the distributions such that $P_i(s_i) = P(s_i), P_j(s_i) = 0, j \neq i$. Then $P = \sum_i P_i$. Using corollary 3.6, we deduce $Q_i, i = 1 \ldots n$ such that $\sum_i m(P_i, Q_i) = m(P,Q)$. Wlog, let $u = s_1$. Consider $Q_1$. In this special case where $P_1$ is a point-state distribution, the dual linear program reduces to:

$$\begin{array}{rcl} \min & & \sum_j l_j m(s_1, s_j) + x + \sum_j y_j \\ \text{subject to} & : & \forall i. \sum_j l_j + x = P_1(s_1) \\ & & \forall j. l_j + y_j = Q_1(s_j) \\ & & \forall i, j.\ l_{ij}, x_i, y_j \geq 0. \end{array}$$

For each $s_j, j = 1 \ldots n$, for any $\delta > 0$, the hypothesis on matching one step transitions from $s_1$ yields $s_j \stackrel{b}{\Rightarrow} Q'_j$ such that $m(u', Q'_j) < m(s_1, s_j)$. Choosing $\delta$ sufficiently small, the required matching transition for the one step transition $u \stackrel{b}{\rightarrow} u'$ is yielded by the transition $\sum_j l_j (s_j \stackrel{b}{\Rightarrow} Q'_j)$.

The transition matching $s \stackrel{a}{\Rightarrow} P'$ is given by $t \stackrel{a}{\Rightarrow} Q'$ arises for $Q'(s_j) = Q(s_j) - (l_j + y_j) + \sum_i l_i \times Q'_i(s_j)$. ∎

# 4 Explicit construction of pseudometric

In this section, we provide an explicit construction of the maximum fixed point by considering a class of $[0,1]$ valued functions. These functions ought to be viewed as the analogues of formulas in a real-valued modal logic.

**Definition 4.1** $\mathcal{F}$ is a set of function expressions whose syntax is given by:

$$f ::= \mathbf{1} \mid \max(f,f) \mid h \circ f \mid a.f$$

where $a$ is a label, possibly $\tau$ and $h$ is any non-expansive operator on [0,1] ($|h(x) - h(y)| \leq |x - y|$). Given an LCMC $(K, \mathtt{Act}, \longrightarrow, k_0)$, these function expressions are evaluated on $K$ as follows:

$$\begin{array}{rcl} \mathbf{1}(s) & = & 1 \\ \max(f_1, f_2)(s) & = & \max(f_1(s), f_2(s)) \\ h \circ f(s) & = & h(f(s)) \\ \tau.f(s) & = & \begin{cases} \max(f(s), \{\tau.f(s') | s \stackrel{\tau}{\rightarrow}_n s'\}) & \text{if } s \in K_n \\ \max(f(s), \sum_i P(s_i) \tau.f(s_i)) & \text{if } s \rightarrow_p P \end{cases} \\ \text{if } a \neq \tau & & \\ a.f(s) & = & \begin{cases} \max(\{\tau.f(s') | s \stackrel{a}{\rightarrow}_n s'\} \cup \{a.f(s') | s \stackrel{\tau}{\rightarrow}_n s'\}) & \text{if } s \in K_n \\ \sum_i P(s_i) a.f(s_i) & \text{if } s \rightarrow_p P. \end{cases} \end{array}$$

[6] the proof for the case when the one step extension is a probabilistic transition is similar and is omitted.

These functions are inspired by a simple modal logic. $\mathbf{1}$ corresponds to the formula `true`, $\max(f_1, f_2)$ corresponds to disjunction, and $a.f$ corresponds to the next modality operator. $h \circ f$ encompasses both testing and negation — $h(x) = 1 - x$ corresponds to negation, whereas $h(x) = x - q$ if $x > q$, 0 otherwise allows testing the probability values. We define a pseudometric $d$ as follows.

**Definition 4.2** $d(s,t) = \sup_{f \in \mathcal{F}} |f(s) - f(t)|$

In the rest of this section, we examine the structure of $d$. Our aim is to prove that $d$ coincides with the maximum fixed point of the functional $F$ of the previous section.

The distance between two probabilistic states is closely related to their outgoing distributions as will be shown in Corollary 4.5, this observation yields a natural definition for the distance between two distributions on the set of states of a LCMC.

**Definition 4.3** *Let $P$ and $Q$ be probability distributions on a set of states. Then:*

- *We write $f(P)$ to mean $\sum_i P(s_i) f(s_i)$, the expectation of $f$ under $P$.*
- *the distance between $P$ and $Q$ is defined as*

$$d(P,Q) = \sup_{f \in \mathcal{F}} |f(P) - f(Q)|.$$

The following lemma shows that when computing the distance between two states, we can restrict to function expressions that are of the form $a.f$.

**Lemma 4.4** *For every pair of states $s, t$,*

$$d(s,t) = \sup_{a \in \mathtt{Act}, f \in \mathcal{F}} \{|a.f(s) - a.f(t)|\}.$$

**Proof.** We prove by induction on the structure of $f$ that there exists a function $a.g$ such that $|a.g(s) - a.g(t)| \geq |f(s) - f(t)|$. The base case $f = \mathbf{1}$ is trivial, and so is the case $f = a.f'$.

Let $f = h \circ f'$. Since $h$ is non-expansive, we have that

$$|f(s) - f(t)| \leq |f'(s) - f'(t)| \leq |a.g(s) - a.g(t)|$$

where $g$ is given by induction from $f'$.

Let $f = \max(f_1, f_2)$. By induction, we have $g_1, g_2$ such that:

$$|f_j(s) - f_j(t)| \leq |a_j.g_j(s) - a_j.g_j(t))|, \quad j = 1, 2.$$

But $|f'(s) - f'(t)| \leq \max(|f_1(s) - f_1(t)|, |f_2(s) - f_2(t)|)$. So the required function expression $g$ is one of $g_1, g_2$. ∎

The following result relates the distance between two probabilistic states and the distance between their probabilistic transitions distributions.

**Corollary 4.5** *Let $s, t$ be probabilistic states such that $s \rightarrow_p P$ and $t \rightarrow_p Q$. Then*

$$d(s,t) = d(P,Q).$$

**Proof.** We want to show that $d(s,t) = \sup_{f\in\mathcal{F}}\{|\sum_i (P(s_i) - Q(s_i))f(s_i)|\}$. By the preceding lemma, we have that $d(s,t) = \sup_{a\in\mathtt{Act}, f\in\mathcal{F}}\{|a.f(s) - a.f(t)|\}$. For $a \neq \tau$, we have that for all $f \in \mathcal{F}$, $|a.f(s) - a.f(t)| = |\sum_i (P(s_i) - Q(s_i))a.f(s_i)|$, as wanted, by definition of $a.f$. For $a = \tau$, we have that for all $f \in \mathcal{F}$

$$\begin{aligned} |\tau.f(s) - \tau.f(t)| &= |\max(f(s), \sum_i P(s_i)\tau.f(s_i)) - \max(f(t), \sum_i Q(s_i)\tau.f(s_i))| \\ &\leq \max(|f(s) - f(t)|, |\sum_i P(s_i)\tau.f(s_i)) - \sum_i Q(s_i)\tau.f(s_i))|. \end{aligned}$$

If the maximum is obtained from the second argument, then the result follows. If the maximum is obtained from $|f(s) - f(t)|$, we know from the proof of the preceding lemma that there exist $a$ and $g$ such that $|f(s) - f(t)| \leq |a.g(s) - a.g(t)| = |\sum_i (P(s_i) - Q(s_i))a.g(s_i)|$, as wanted. ∎

The pseudometric $d$ satisfies the requirements of definition 3.3.

**Lemma 4.6 ((Adapted from [vBW01a]))** *Let $\mathcal{K}$ be a LCMC and $P$ and $Q$ be two probability distributions on its set of states. Then the distance between $P$ and $Q$ is given by the solution to the linear program (assuming $P(K) \geq Q(K)$):*

$$\max \sum_i (P(s_i) - Q(s_i))a_i, \text{ subject to: } \begin{array}{l} 0 \leq a_i \leq 1 \\ \forall i,j \;\; a_i - a_j \leq d(s_i, s_j) \end{array}.$$

**Proof.** Let us write $L(P,Q)$ for the solution of the given linear program. For any function $f$, $a_i = f(s_i)$ satisfies the constraints. Let us prove that $|f(P) - f(Q)| \leq L(P,Q)$ for every $f \in \mathcal{F}$. If $f(P) \geq f(Q)$, then $|f(P) - f(Q)| \leq L(P,Q)$. Otherwise consider the $a_i$'s be given by the function $\mathbf{1} - f$. Then $|f(P) - f(Q)| = f(Q) - f(P) = (\mathbf{1} - f)(P) - (\mathbf{1} - f)(Q) \leq L(P,Q)$.

Now consider a set of real numbers $\{a_i\}$ which maximizes the expression of $L(P,Q)$. For every $\epsilon > 0$, we want to find a function expression $f$ such that $\sum_i (P(s_i) - Q(s_i))a_i - \epsilon \leq \sum_i (P(s_i) - Q(s_i))f(s_i)$ Let $\epsilon > 0$; we can determine for each $s_i$ and $s_j$ a function $f_{ij}$ such that $f_{ij}(s_i) = a_i$, and $f_{ij}(s_j) \leq a_i - d(s_i, s_j) + \epsilon$. Define $f_i = \min_j (f_{ij})$. Then $f_i(s_i) = a_i$, and for all $j \neq i$, $f_i(s_j) \leq a_i - d(s_i, s_j) + \epsilon \leq a_j + \epsilon$. Define $f = \max_i f_i$. Thus for all $i$, $a_i \leq f(s_i) \leq a_i + \epsilon$. Thus $d(P,Q) \geq L(P,Q) - \epsilon$. ∎

We first relate the evaluation of $a.f$ over states with its evaluation over the distributions to which these states have a weak $a$-transition.

**Lemma 4.7** *For any label $a \in \mathtt{Act}$, $a.f(s) = \sup_{s \stackrel{a}{\Rightarrow} P} f(P)$*

**Proof.** We first prove that $a.f(s) \geq f(P)$ for all $P$ such that $s \stackrel{a}{\Rightarrow} P$. The result for linear combinations $s \stackrel{a}{\Rightarrow} \sum_i \lambda_i \times P_i$ follows from the result for basic transitions $s \stackrel{a}{\Rightarrow} P$, since

$f(\lambda_i \times P_i) = \sum_i \lambda_i f(P_i)$ from definition 4.3. It suffices to prove this for computation trees of finite depth, since $f(P) = \sup_i f(P_i)$, where $P_i$ is a distribution on the leaves of the computation tree which assigns value 0 to all leaves of depth $> i$.

The proof proceeds by induction on the depth of the computation tree $C$ underlying $s \stackrel{a}{\Rightarrow} P$. If its depth is 0, then $C$ has a single node $s$, and $a = \tau$. The result follows since $\tau.f(s) \geq f(s)$. Assume the claim is true for computations of depth $n$ or less, and let $C$ be a computation of depth $n+1$.

If $s \in K_n$, then the result follows easily by definitions. Now assume that $s \in K_p$. For every $s_i$ such that $s \rightarrow_p s_i$ we have by induction that $a.f(s_i) \geq f(P_i)$ where $P_i$ given by the branch of the computation that goes through $s_i$. Let $s \rightarrow_p Q$. Then we have $P(s_j) = \sum_i Q(s_i)P_i(s_j)$. Then

$$\begin{aligned} a.f(s) &\geq a.f(Q) = \sum_i Q(s_i) a.f(s_i) \\ &\geq \sum_i Q(s_i) f(P_i) \\ &\geq \sum_i Q(s_i)(\sum_j P_i(s_j) f(s_j)) = f(P). \end{aligned}$$

The first inequality is by definition of $a.f$.

We now prove that for every $s \in K$ there is a distribution $P$ such that $s \stackrel{a}{\Rightarrow} P$ and $f(P) = a.f(s)$. We construct the (possibly infinite) corresponding $C$ as follows.

If $a = \tau$ and $\tau.f(s) = f(s)$ or 0, the required computation is the single node $s$.

Otherwise, if $s$ is a probabilistic state we simply choose the full probabilistic transition as the first transition out of $s$. If $s$ is a non-deterministic state, then the definition of $a.f$ yields either a transition $s \stackrel{a}{\rightarrow}_n s'$ such that $a.f(s) = \tau.f(s')$ or a transition $s \stackrel{\tau}{\rightarrow}_n s'$ such that $a.f(s) = a.f(s')$. We choose this transition as the first transition out of $s$. ∎

Next, we establish $d$ as a fixed point of the functional $F$ from the previous section.

**Lemma 4.8** *Let $s$, $t$ be two states, and $d(s,t) < \delta$. Let $s \stackrel{a}{\Rightarrow} P$. Then, there exists a transition $t \stackrel{a}{\Rightarrow} Q$ such that $d(P,Q) < \delta$.*

**Proof.** We prove by contradiction. Let $\epsilon = \delta - d(s,t)$. Let $Q_1, Q_n$, be all the distributions such that $t \stackrel{a}{\Rightarrow} Q_i$. If none of these satisfy the condition, then for each $i$ there exists a function $f_i$ such that $f_i(P) = d(s,t) + \epsilon$, $f_i(Q_i) = 0$. We will show how construct a finite subset of functions $f_i$ such that $f = \min f_i$, $f(P) = d(s,t) + \epsilon$, and $f(Q_i) \leq \epsilon/2$ for all $i$. Thus using lemma 4.7, $a.f(s) - a.f(t) > d(s,t)$, which is a contradiction.

The finite subset can be chosen as follows: if distributions $q = \sum q_i s_i$ and $q' = \sum q'_i s_i$ are such that $|q_i - q'_i| < \epsilon/rn^2$ (where $r = \min d(s_i, s_j)$, and $n$ is the number of states), we get $d(q,q') \leq \epsilon/rn^2 \sum a_i \leq \epsilon/rn^2 * rn^2/2 = \epsilon/2$. So if we regard each distribution as a vector in a finite dimensional cube of side 1, then the whole cube can be partitioned into finitely many cubes of sides $\epsilon/rn^2$. If we choose the $f_i$ corresponding to a distribution $q$ in each tiny cube, then for all the other distributions $q'$ in the cube, $f_i(q') \leq f_i(q) + d(q',q) = \epsilon/2$. Thus we have finitely many $f_i$ satisfying the property. ∎

**Corollary 4.9**

$$d(s,t) = \max(\max_{a\in \mathtt{Act}} \sup_{s\stackrel{a}{\Rightarrow}P} \inf_{t\stackrel{a}{\Rightarrow}Q} d(P,Q), \max_{a\in \mathtt{Act}} \sup_{t\stackrel{a}{\Rightarrow}Q} \inf_{s\stackrel{a}{\Rightarrow}P} d(P,Q))).$$

**Proof.** Let $M$ stand for the maximum on the right hand side of the claim. $d(s,t) \geq M$ by Lemma 4.7. We now show that $d(s,t) \leq M$. We want to show that for all $f$ and $\epsilon > 0$, there exists $a \in \mathtt{Act}$, $P \in A$, $Q \in B$, $g \in \mathcal{F}$ such that $|f(s) - f(t)| \leq |g(P) - g(Q)| + \epsilon \leq M + \epsilon$. $a$ and $g$ are given by Lemma 4.4, which yields $|f(s) - f(t)| \leq |a.g(s) - a.g(t)|$. Then by Lemma 4.7, $a.g(s) = \sup\{g(P) | s \stackrel{a}{\Rightarrow} P\}$ and similarly for $t$; this implies that for every $\epsilon > 0$ there are $P \in A$, $Q \in B$, such that $|f(s) - f(t)| \leq |g(P) - g(Q)| + \epsilon$, as wanted. ∎

The maximum fixed point of $F$ is the pseudometric $d$ that we have already studied. Corollary 4.9 already shows that d is a fixed point of $F$. For the converse, we define the depth of functional expressions as follows:

$$\begin{aligned} \mathit{depth}(\mathbf{1}) &= 0 \\ \mathit{depth}(h \circ f) &= \mathit{depth}(f) \\ \mathit{depth}(\max(f_1, f_2)) &= \max(\mathit{depth}(f_1), \mathit{depth}(f_2)) \\ \mathit{depth}(\langle a\rangle.f) &= \mathit{depth}(f) + 1 \end{aligned}$$

and show that $m_i \preceq d_i$ for all $i$, where $d_i$ is the pseudometric induced by functions of depth $\leq i$.

**Theorem 4.10** *$d$ is the maximum fixed point of $F$.*

**Proof.** Corollary 4.9 shows that $d$ is a fixed point of $F$. So $d \preceq m$.

We prove the converse now. It suffices to show that $m_i \preceq d_i$ for all $i$, where $d_i$ is the pseudometric induced by functions of depth $\leq i$. Proof proceeds by an induction on $i$. The proof is immediate for base case $i = 0$. For the inductive case $i = k + 1$, using lemma 3.8,

$$F(m_i)(s,t) = \max(\max_{a\in \mathtt{Act}} \sup_{s\stackrel{a}{\Rightarrow}P} \inf_{t\stackrel{a}{\Rightarrow}Q} m_i(P,Q), \max_{a\in \mathtt{Act}} \sup_{t\stackrel{a}{\Rightarrow}Q} \inf_{s\stackrel{a}{\Rightarrow}P} m_i(P,Q)).$$

By induction, for all functions $f$ of depth $k$, forall $P, Q$ $m_i(P,Q) \geq |f(P) - f(Q)|$. Thus, for all functions $\langle a\rangle.f$, $m_{i+1}(s,t) \geq |\langle a\rangle.f(s) - \langle a\rangle.f(Q)|$. The results for the cases max and $h\circ$ are immediate since they are non-expansive. ∎

# 5 Process algebra

**Prefixing, Internal choice, Probabilistic choice, Restriction.** The constructions have already been described in example 2.2.

**Lemma 5.1** *Prefixing, internal choice, probabilistic choice and restriction are non-expansive.*

**Proof.**

- *Prefixing:* Let $m$ be the pseudometric witnessing $d(s_1, s_2) < \epsilon$. We will construct a prefixed point $m'$ such that $m'(u_1, u_2) < \epsilon$, where $u_i$ are the start states of $a.s_i$. Define $m'$ as follows: Define $m'(u_1, u_2) = m(s, t)$. For all other states $x \neq u_1, u_2$, $m'(u_1, x) = m'(u_2, x) = 1$. Finally, $m'(x, y) = m(x, y)$ for all states $x, y \notin \{u_1, u_2\}$. Using lemma 3.12, we only need to worry about the $a$ transition out of $u_1, u_2$. For this case, $u_1 \xrightarrow{a} s, u_2 \xrightarrow{a} s_2$ serve for the required match since $m'(u_1, u_2) = m(s, t)$.

- *Internal choice:* Let $m$ be the pseudometric witnessing $d(s_1, s_2) < \epsilon$. Let $u_i$ be the start state of $s_i \oplus t, i = 1, 2$. We will construct a prefixed point $m'$ such that $m'(u_1, u_2) < \epsilon$. Define $m'(u1, u2) = m(s1, s2)$. For all other states $x \neq u_1, u_2$, $m'(u_1, x) = m'(u_2, x) = 1$. Finally, $m'(x, y) = m(x, y)$ for all states $x, y \notin \{u_1, u_2\}$. Using lemma 3.12, we only need to worry about the $\tau$ transitions out of $u_1, u_2$. For this case, $u_1 \xrightarrow{\tau} s_1, u_2 \xrightarrow{\tau} s_2$ serve for the required match since $m'(u_1, u_2) = m(s, t)$. Similarly, $u_1 \xrightarrow{\tau} t, u_2 \xrightarrow{\tau} t$ serve for the required match since $m'(t, t) = 0$.

- *Probabilistic choice:* Let $m$ be the pseudometric witnessing $d(s_1, s_2) < \epsilon$. Let $u_i$ be the start state of $rs_i + (1-r)t, i = 1, 2$. We will construct a prefixed point $m'$ such that $m'(u_1, u_2) < \epsilon$. Define $m'(u1, u2) = r \times m(s1, s2)$. For all other states $x \neq u_1, u_2$, $m'(u_1, x) = m'(u_2, x) = 1$. Finally, $m'(x, y) = m(x, y)$ for all states $x, y \notin \{u_1, u_2\}$. Using lemma 3.12, we only need to worry about the probabilistic transitions out of $u_1, u_2$. First, note that the dual form of the linear program defining $m$ for distributions has a solution $r\epsilon$ [by setting all $x_i, y_j$ to zero, and setting $l_{tt} = 1 - r, l_{s_1,s_2} = r$]. So, $m(\{(r, s_1), (1-r, t)\}, \{(r, s_2), (1-r, t)\}) \leq r\epsilon$. This immediately gives us the required match for the probabilistic transitions out of $u_1, u_2$.

- *Restriction:* Let $m$ be the pseudometric witnessing $d(s_1, s_2) < \epsilon$. Then, $m$ also witnesses $d(s_1 \backslash a, s_2 \backslash a) < \epsilon$.

∎

**Parallel composition.** For any LCMC, we can construct a bisimilar "saturated" transition system, that has the property that every nondeterministic state has a bisimilar copy in probabilistic states and vice versa. Given $\mathcal{K} = (K, \mathtt{Act}, \longrightarrow, k_0)$, construct $\mathcal{U} = (U, \mathtt{Act}, \longrightarrow, u_0)$ as follows: replace each transition $s \xrightarrow{a} t$ from a nondeterministic state $s$ by a sequence $s \xrightarrow{a} s_1$, $s_1 \xrightarrow{1} s_2$, $s_2 \xrightarrow{\tau} t$, where $s_1, s_2$ are new states not in $K$. Then, for every nondeterministic state $s$ add the following self-loop: $s \xrightarrow{\tau} s^p$, $s^p \xrightarrow{1} s$, where $s^p$ is again a new state.

We will now define the parallel composition of LCMCs. Assume that both LCMCs are saturated and that their start states are nondeterministic. When restricted to LTSs, our definition agrees upto weak bisimilarity. On saturated LCMCs, our definition agrees upto strong bisimulation with the definition of Hansson and Johnson [HJ90] as modified in the thesis of Hansson [Han94].

Given saturated LCMCs $\mathcal{K}, \mathcal{L}$ whose start states are nondeterministic, we define $\mathcal{M} = \mathcal{K}||\mathcal{L}$ as follows. For any nondeterministic state $s$ (in either of $\mathcal{K}, \mathcal{L}$), we use $s^p$ to refer to the corresponding bisimilar probabilistic state to which $s$ has a $\tau$ transition. The nondeterministic (resp. probabilistic) states of $\mathcal{M}$ are $\{(s, t) \mid s \in K, t \in L\}$, both nondeterministic} (resp.

$\{(s,t) \mid s \in K, t \in L\}$, both probabilistic$\}$). The transition relation is:

$$\begin{array}{ll} (s,t) \xrightarrow{a} (s', t^p), (t,s) \xrightarrow{a} (t^p, s') & \text{if } s \xrightarrow{a} s' \\ (s,t) \xrightarrow{\tau} (s',t') & \text{if } s \xrightarrow{a} s', t \xrightarrow{\bar{a}} t' \\ (s,t) \xrightarrow{p \times q} (s',t') & \text{if } s \xrightarrow{p} s', t \xrightarrow{q} t' \end{array}$$

This definition preserves saturation and strict alternation of nondeterministic and probabilistic transitions.

**Lemma 5.2** *Parallel composition is non-expansive.*

**Proof.**(Sketch) Consider $\mathcal{K}_1||\mathcal{L}$ and $\mathcal{K}_2||\mathcal{L}$. Let $m$ be the pseudometric witnessing $d(s_1, s_2) < \epsilon$, where $s_i \in \mathcal{K}_i$. Define $m'((s_1,t),(s_2,t)) = m(s_1,s_2)$, where $s_i \in \mathcal{K}_i$, both $s_1, s_2$ nondeterministic or probabilistic. The required witness $m'$ such that $m'(u_1,u_2) < \epsilon$ is defined by $m'((s_1,t),(s_2,t)) = 1$, where $s_i \in \mathcal{K}_i$, one of $s_1, s_2$ nondeterministic and the other probabilistic. ∎

**Handling CCS +.** We now sketch the extension of the theory to handle the CCS + operator. Weak bisimulation is not a congruence for CCS + operator even in LTSs, a situation remedied by a stricter matching of initial $\tau$ transitions.

We follow this standard trick in the metric context. We write $s \stackrel{\tau^+}{\Rightarrow} P$ for a computation $P$ from $s$ such that every path from $s$ has at least one $\tau$ transition. The $d^c$ pseudometric agrees with $d$ wrt matching of all non-$\tau$ transitions. For $\tau$-transitions, we demand a match by a transition of form $s \stackrel{\tau^+}{\Rightarrow} P$.

**Definition 5.3**

$$d^c(s,t) \;=\; \max \left\{ \begin{array}{ll} d(s,t), & \\ \max & (\sup_{P \in A} \inf_{Q \in B} d(P,Q), \\ & \sup_{Q \in B} \inf_{P \in A} d(P,Q))) \end{array} \right.$$

*where* $A = \{P \mid s \stackrel{\tau^+}{\Rightarrow} P\}, B = \{Q \mid t \stackrel{\tau^+}{\Rightarrow} Q\}$.

Zero distance in $d^c$ agrees with the largest congruence contained in weak bisimulation [BS01].

**Lemma 5.4** *Let* $d^c(s,t) = 0$. *Then:*

- *For all one-step transitions* $s \xrightarrow{a} s'$, *there exists* $t \stackrel{a}{\Rightarrow} Q$ *such that* $d(s', Q) = 0$, *if* $a \neq \tau$.
- *For all one-step* $\tau$ *transitions* $s \xrightarrow{\tau} P$, *there exists* $t \stackrel{\tau^+}{\Rightarrow} Q$ *such that* $d(P,Q) = 0$.
- *If* $s$ *is a probabilistic state with targets of the probabilistic fan given by the distribution* $P$, *there exists a transition* $t \stackrel{\tau^+}{\Rightarrow} Q$ *such that* $d(P,Q) = 0$.

**Proof.** The only non-trivial case to consider is $s \stackrel{\tau^+}{\Rightarrow} P$. In this case, since $d^c(s,t) = 0$, there is a $t \stackrel{\tau^+}{\Rightarrow} Q$ such that $d(P,Q) = 0$. ∎

**Lemma 5.5** $+$ *is non-expansive wrt* $d^c$.

**Proof.** Let $d^c(s_1, s_2) < \epsilon$. Let $u_i = s_i + t, i = 1, 2$. We will prove that $d^c(u_1, u_2) < \epsilon$. Let $u_1 \xrightarrow{a} u'$. There are three cases:

- The transition is caused by a transition $t \xrightarrow{a} u'$. In this case, the required matching transition $u_2 \xrightarrow{a} u'$ is immediate.
- $a \neq \tau$ and the transition is caused by a transition $s_1 \xrightarrow{a} u'$. In this case the required transition $s_2 \stackrel{a}{\Rightarrow} Q$ such that $d(s', Q) < \epsilon$ is yielded by the hypothesis $d^c(s_1, s_2) < \epsilon$ since $d(s_1, s_2) \leq d^c(s_1, s_2)$.
- The key case is when $a = \tau$ and this transition is caused by $s_1 \xrightarrow{\tau} u'$. The required transition $s_2 \stackrel{a}{\Rightarrow} Q$ such that $d(s', Q) < \epsilon$ is yielded by the hypothesis $d^c(s_1, s_2) < \epsilon$.

∎

$d^c$ can be captured in the explicit construction of the metric by adding a function expression $\tau^+.f$ that is evaluated at a state $s$ of an LCMC as:

$$\tau^+.f(s) \quad = \quad \begin{cases} \max(\{\tau.f(s') | s \xrightarrow{\tau}_n s'\}) & \text{if } s \in K_n \\ \max(\sum_i P(s_i)\tau.f(s_i)) & \text{if } s \rightarrow_p P. \end{cases}$$

The addition of a top-level $\tau^+$ test is exactly what is needed to enable the real-valued modal logic to capture $d^c$. The proof proceeds by demonstrating that the new function $\tau^+$ satisfies lemmas 4.7 and that $d^c, \tau^+$ satisfy the characterization of lemma 4.8. This enables us to adapt the proof of lemma 4.9 to this case.

**Lemma 5.6**

$$d^c(s,t) = \max \begin{cases} \sup_{f \in \mathcal{F}} |\tau^+.f(s) - \tau^+.f(t)| \\ \sup_{f \in \mathcal{F}} |f(s) - f(t)| \end{cases}$$

**Proof.** It suffices to prove that

$$\sup_{f \in \mathcal{F}} |\tau^+.f(s) - \tau^+.f(t)| = \max(\sup_{P \in A} \inf_{Q \in B} d(P,Q), \sup_{Q \in B} \min_{P \in A} d(P,Q)))$$

where $A$ is the set of distributions $P$ such that $s \stackrel{\tau^+}{\Rightarrow} P$, and $B$ is the set of distributions $Q$ such that $t \stackrel{\tau^+}{\Rightarrow} Q$.

The proof that the LHS $\leq$ RHS follows the proof of lemma 4.8. The proof that the RHS $\leq$ LHS follows the proof of lemma 4.9. Following the proof of lemma 4.7, we can show that: $\tau^+.f(s) = \sup_{s \stackrel{\tau^+}{\Rightarrow} P} f(P)$. Thus: $\tau^+.g(s) = \sup\{g(P) | s \stackrel{\tau^+}{\Rightarrow} P\}$, $\tau^+.g(t) = \sup\{g(Q) | t \stackrel{\tau^+}{\Rightarrow} Q\}$. For any $P$, there is a $Q$ such that: $|g(P) - g(Q)| \leq$ LHS. Thus, $|\tau^+.g(s) - \tau^+.g(t)| \leq$ LHS. ∎

# 6 Secure substitution

The context for our investigations is exemplified by mobile code applications where programs (such as tax software) are downloaded as needed, executed on a trusted host (the home computer), require access to sensitive local data (such as financial information) and yet should not be permitted to "leak information". Thus, we are in the situation where a "mole" has been permitted inside the system and we are interested in measuring the information that the mole can leak to the outside world. We use the definition and basic results about channel capacity from information theory — see [CT91] for details.

We first formalize the interface of this channel. Fix an LCMC, $\mathcal{K} = (K, \mathtt{Act}, \longrightarrow, k_0)$. Fix $O \subset \mathtt{Act} - \{\tau\}$, a subset of labels not including $\tau$ that is intended to model the "output labels" visible to the external observer. The remaining labels, $I = (\mathtt{Act} - \{\tau\}) \setminus O$ is the set of "input labels". The (sequences of) labels from $I$ constitute the input symbols of the channel. The "mole" is viewed as attempting to leak information using these input symbols by influencing the (sequences of) $O$ labels. These sequences of output labels constitute the output symbols. Thus, the outside observer is attempting to deduce input traces (elements of $I^\star$) while only viewing the output traces (elements of $O^\star$).

Next, we identify the probabilistic transducers from input symbols to output symbols associated with this channel. Inevitably, in the presence of nondeterminism, we cannot identify a single transducer, rather we are forced to accommodate a set of probabilistically determinate transducers. The following definition of $O$-determinate subtransition systems captures the conditions on the elements of this set. The conditions reflect two aims. First, we want to ensure liveness conditions that prevent the adversary from blocking the system from progress, if it is possible. Secondly, we want to ensure that the nondeterminism is resolved enough to get a (probability) distribution on output symbols once the input symbols are fixed.

First some notation. For purely notational convenience, assume that the LCMC is completely unrolled into a tree. We say that two sequences of transitions $\sigma, \sigma'$ are *consistent*, if the choices at nondeterministic states that occur in both $\sigma, \sigma'$ are consistent. For a state $s$ and a subset of transitions $T$ , define $Enabled^T(s) = \{a | a \neq \tau$, there is a path with weak label $a$ from $s$ in $T\}$. We will consider subsets $T$ of transitions that satisfy:

1. Liveness: For all $s$, $|Enabled^T(s) \cap \mathtt{Act}| = 0 \Rightarrow |Enabled^{\longrightarrow}(s) \cap \mathtt{Act}| = 0$.
2. $I$-liveness: For all $s$ such that $|Enabled^T(s) \cap O| = 0$, $Enabled^T(s) \cap I = Enabled^{\longrightarrow}(s) \cap I$.
3. $O$-determinacy: Let $s$ be such that $|Enabled^T(s) \cap O| \geq 1$. Let $a, b \in \mathtt{Act} \setminus \{\tau\}$. Let $\sigma_1 = s \xrightarrow{\tau^*} s_1' \xrightarrow{a} s_1$ and $\sigma_2 = s \xrightarrow{\tau^*} s_2' \xrightarrow{b} s_2$ be sequences of transitions in $T$. Then, $\sigma_1, \sigma_2$ are consistent.
4. $I$-determinacy: Let $a \in I$, $s$ a state. Let $\sigma_1 = s \xrightarrow{\tau^*} s_1' \xrightarrow{a} s_1$ and $\sigma_2 = s \xrightarrow{\tau^*} s_2' \xrightarrow{a} s_2$ be sequences of transitions in $T$. Then, $\sigma_1, \sigma_2$ are consistent.

The first condition ensures that $T$ cannot reject all labels. The second condition states that the only reason for $T$ to reject an input symbol (that might otherwise be accepted) at a state is a nondeterministic choice leading to an output symbol. The third condition ensures that the choices at states that can perform a weak $O$-labelled transition are purely probabilistic. The last condition ensures that at any state, there is only one computation for a given $I$-label.

The transition systems satisfying these conditions induce probabilistic transformers from sequences (say of length $m$) of $I$-labels to sequences of $O$-labels (say of length $n$). Let $\sigma \restriction A, A \subseteq \texttt{Act}$ denote the subsequence of $A$-actions in $\texttt{Weak}(\sigma)$.

**Definition 6.1** *Let $T \subseteq \longrightarrow$ satisfy the above conditions. Let $y \in O^m, x \in I^n$. Define:*

$$\begin{aligned} p_T^{m,n}(y|x) &= \sum\{prob(\sigma) \mid \sigma = s \to s_1 \to \ldots \xrightarrow{a} t, \\ &\qquad a \neq \tau, \sigma \restriction O = y, \sigma \restriction I = x\} \\ p_T^{m,n}(\delta|x) &= 1 - \sum_{y \in O^m} p_T^{m,n}(y|x) \end{aligned}$$

$\delta$ is a special symbol to indicate absence of output from $O^m$. $p_T^{m,n}(.|x)$ is a probability distribution on $O^m \cup \{\delta\}$.

The channel capacity of a single probabilistic transformer such as the one above is defined in a standard fashion — see [CT91] for detailed intuitions.

**Definition 6.2** *Given a joint probability distribution $f(x, y)$, define mutual information, written $I(f)$ as:*

$$I(f) = -\sum_{x \in X} \sum_{y \in Y} f(x, y) \log[\frac{f(x, y)}{f(x) \times f(y)}].$$

*The channel capacity of $p(y|x)$, a probabilistic transducer from inputs $x$ to outputs $y$, written $Ch(p)$, is $\max_r I(g)$ where $r(x)$ is a probability distribution on inputs $x$ and $g(x, y) = r(x) \times p(y|x)$.*

The general theorems of information theory guarantee that the channel capacity calculated as per the above definition can be achieved operationally by repeated use of the transducer $p$. This definition can of course be used on $p_T^{m,n}((.)|x)$ defined above since it is a pure probabilistic transducer.

Recall that we model worst case assumptions about the mole by allowing the mole to control the (nondeterministic) scheduler and by permitting it to serve as oracle to guide the scheduler.

**Definition 6.3** *The $O$-channel capacity of a state $s$, written $Ch(s)$ is defined as the supremum over all $m, n, T$ of the channel capacity of $p_T^{m,n}((.)|x)$.*

As per this definition, the mole can choose a nondeterministic scheduler to derive a purely probabilistic computation that gets arbitrarily close to the value of the $O$-channel capacity of a state $s$. The mole does not gain anything by using probabilities to combine several such purely probabilistic computations because of the convexity of mutual information:

**Theorem 6.4 ( [CT91])** *Let $p_1(y|x), p_2(y|x)$ be probabilistic transducers from inputs $x$ to outputs $y$. Let $r(x)$ be any distribution on inputs. Let $g_i(x, y) = r(x) \times p_i(y|x)$, for $i = 1, 2$. Let $0 \leq \lambda \leq 1$. Then $I(\lambda g_1 + (1 - \lambda)g_2)) \leq \lambda I(g_1) + (1 - \lambda)g_2$*

Thus, mixing probabilistic transducers diminishes the channel capacity, validating our definition 6.3 as the correct modelling of the worst case assumptions.

Basic information theory also provides bounds on changes in channel capacity as a function of changes in probability for purely probabilistic transducers.

**Lemma 6.5** *Let $p(y|x), p'(y|x)$ be probabilistic transducers from inputs to outputs such that $\max_{x,y} |p(y|x) - p'(y|x)| < \epsilon$. Then: $|Ch(p) - Ch(p')| < k\epsilon$, for some constant $k$ that depends only on $p$.*

The coinduction infrastructure that we have described earlier permits us to match nondeterministic choices between nearby processes. Thus, if $s, t$ are close, every probabilistically determinate computation from $s$ that contributes to the channel-capacity of $s$ can be matched by one from $t$. This permits us to lift the bound described above for single probabilistic transducers to sets of probabilistic transducers and reach our goal of showing that nearby states have almost the same channel capacities.

**Theorem 6.6** *Let $d(s,t) < \epsilon$. Then $|Ch(s) - Ch(t)| < k\epsilon$, for some constant $k$ that depends only on $s$. Thus, $Ch(\cdot)$ is a continuous function w.r.t. pseudometric $d$.*

**Proof.** Let $T \subseteq \longrightarrow$ be a subset of transitions rooted at $s$ satisfying conditions liveness, $O$-determinacy, $I$-determinacy and $I$-liveness. Let $d(s,t) < \epsilon$. Then, since $F(d) \preceq d$, following the proof of lemma 3.12, we can construct $T' \subseteq \longrightarrow$, a subset of transitions rooted at $t$ satisfying conditions $O$-determinacy, $I$-determinacy and $I$-liveness, such that $(\forall m, n)$, $p_T^{m,n}((.)|x)$ and $p_{T'}^{m,n}((.)|x)$ satisfy the hypothesis of lemma 6.5. Result follows from the conclusion of lemma 6.5. ∎

# 7 Conclusions

We have described a fixed-point approach to a metric analogue for weak bisimulation. This is the first time to our knowledge that the "$\epsilon$-analogue" of the usual process algebra equations for weak (or strong for that matter) bisimulation have been developed. The fixed point approach was crucial to this development since it made coinductive techniques possible. We were able to make use of the beautiful duality principle from linear programming.

We have also explored the quantitative meaning of the metric. In particular we have shown that the analogue of the information theory analysis of Markov chains can be extended to this setting (i.e.to concurrent Markov chains). This involved an extension of the usual concepts like channel capacity.

It may appear that our use of linear programming forces us into the setting of finite state systems. However, the notion of duality works well for infinite-dimensional spaces [AN87] and we are exploring this extension. We have not stressed the logical characterization of weak bisimulation in this paper though that can also be done. In another paper being written the logical version of the theory is being developed as well as the proof that we get a sound and complete model for pCTL*.

## References


[AN87] E. J. Anderson and P. Nash. *Linear Programming in Infinite-dimensional Spaces*. Discrete Mathematics and Computation. Wiley-Interscience, 1987.

[BBS95] J.C.M. Baeten, J.A. Bergstra, and S.A. Smolka. Axiomatizing probabilistic processes: ACP with generative probabilities. *Information and Computation*, 121(2):234–255, 1995.

[BS01] E. Bandini and R. Segala. Axiomatizations for probabilistic bisimulation. In *Proceedings of the 28th International Colloquium on Automata, Languages and Programming*, number 2076 in Lecture Notes In Computer Science, pages 370–381. Springer-Verlag, July 2001.

[CSZ92] R. Cleaveland, S. Smolka, and A. Zwarico. Testing preorders for probabilistic processes. In W. Kuich, editor, *Automata, Languages and Programming (ICALP 92)*, number 623 in Lecture Notes in Computer Science, pages 708–719. Springer-Verlag, 1992.

[CT91] T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley, New York, 1991.

[DEP98] J. Desharnais, A. Edalat, and P. Panangaden. A logical characterization of bisimulation for labeled Markov processes. In *proceedings of the 13th IEEE Symposium On Logic In Computer Science, Indianapolis*, pages 478–489. IEEE Press, June 1998.

[DGJP99] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled Markov processes. In *Proceedings of CONCUR99*, Lecture Notes in Computer Science. Springer-Verlag, 1999.

[DGJP00] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Approximation of labeled markov processes. In *Proceedings of the Fifteenth Annual IEEE Symposium On Logic In Computer Science*, pages 95–106. IEEE Computer Society Press, June 2000.

[Edg98] Gerald A. Edgar. *Integral, Probability and Fractal Measures*. Springer-Verlag, 1998.

[GHJ97] V. Gupta, T. A. Henzinger, and R. Jagadeesan. Robust timed automata. In Oded Maler, editor, *Hybrid and Real-Time Systems*, volume 1201 of *Lecture Notes In Computer Science*, pages 331–345. Springer Verlag, March 1997.

[Han94] Hans A. Hansson. *Time and Probability in Formal Design of Distributed Systems*, volume 1 of *Real-time Safety-critical Systems*. Elseiver, 1994.

[HJ90] H. Hansson and B. Jonsson. A calculus for communicating systems with time and probabilities. In *Proceedings of the 11th IEEE Real-Time Systems Symposium*, pages 278–287. IEEE Computer Society Press, 1990.

[HS86] S. Hart and M. Sharir. Probabilistic propositional temporal logics. *Information and Control*, 70:97–155, 1986.

[Hut81] J. E. Hutchinson. Fractals and self-similarity. *Indiana Univ. Math. J.*, 30:713–747, 1981.

[JS90] C.-C. Jou and S. A. Smolka. Equivalences, congruences, and complete axiomatizations for probabilistic processes. In J.C.M. Baeten and J.W. Klop, editors, *CONCUR 90 First International Conference on Concurrency Theory*, number 458 in Lecture Notes In Computer Science. Springer-Verlag, 1990.

[JY95] B. Jonsson and W. Yi. Compositional testing preorders for probabilistic processes. In *Proceedings of the 10th Annual IEEE Symposium On Logic In Computer Science*, pages 431–441, 1995.

[Koz81] D. Kozen. Semantics of probabilistic programs. *Journal of Computer and Systems Sciences*, 22:328–350, 1981.

[Koz85] D. Kozen. A probabilistic PDL. *Journal of Computer and Systems Sciences*, 30(2):162–178, 1985.

[LMMS98] P. D. Lincoln, J.C. Mitchell, M. Mitchell, and A. Scedrov. A probabilistic poly-time framework for protocol analysis. In *ACM Computer and Communication Security (CCS-5)*, 1998.

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

[PLS00] A. Philippou, I. Lee, and O. Sokolsky. Weal bisimulation for probabilistic processes. In C. Palamidessi, editor, *Proceedings of CONCUR 2000*, number 1877 in Lecture Notes In Computer Science, pages 334–349. Springer-Verlag, 2000.

[Seg95] R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems.* PhD thesis, MIT, Dept. of Electrical Engineering and Computer Science, 1995. Also appears as technical report MIT/LCS/TR-676.

[Uli92] I. Ulidowski. Equivalences on observable processes. In *Proceedings of the Seventh IEEE Symposium On Logic In Computer Science*, pages 148–159. IEEE Press, 1992.

[Uli94] I. Ulidowski. *Local Testing and Implementable Concurrent Processes.* PhD thesis, Imperial College, 1994.

[vBW01a] Franck van Breugel and James Worrell. An algorithm for quantitative verification of probabilistic systems. In K. G. Larsen and M. Nielsen, editors, *Proceedings of the Twelfth International Conference on Concurrency Theory - CONCUR'01*, number 2154 in Lecture Notes In Computer Science, pages 336–350. Springer-Verlag, 2001.

[vBW01b] Franck van Breugel and James Worrell. Towards quantitative verification of probabilistic systems. In *Proceedings of the Twenty-eighth International Colloquium on Automata, Languages and Programming*. Springer-Verlag, July 2001.

[WSS97] S.-H. Wu, S.A. Smolka, and E. W. Stark. Composition and behaviors for probabilistic I/O automata. *Theoretical Computer Science*, 176(1–2):1–36, April 1997.

# A Proof of Lemma A.1

**Lemma A.1** *Given a countable set of states $A$, with every pair of states in $A$ non-bisimilar, there exists $s \in A$ such that $P(s, \emptyset, A \setminus \{s\}) < 1$.*

**Proof.** Let $A = \{s_i \mid i = 1, 2, \ldots\}$. We first prove that if $((\forall s_i \in A)\ [P(s_i, \emptyset, A \setminus \{s_i\}) = 1])$ then the same statement is true for $A \setminus \{s_j\}$, for any $s_j \in A$.

Define an "$E$-maximal" $L$-computation as follows. Consider computations under the prefix ordering on trees. Given $E \subseteq K$, an $L$-computation from $s \in K$ is $E$-maximal if it is maximal among computations that satisfy: any node $n$ such that $\texttt{State}(n) \in E$ is a leaf. Thus an $E$-maximal $L$-computation intuitively "does its best" to reach an $E$ node — a node $n$ with $\texttt{State}(n) \notin E$ can be a leaf only if every possible extension of the path from the root to $n$ leads to a weak label not in $L$.

Let $C_i$ be the computation that induces the maximum value, namely 1, of $P(s_i, \emptyset, A \setminus \{s_i\})$, and let $P_i$ be the distribution induced by this computation on $A \setminus \{s_i\}$. We can assume that $C_i$ is $A \setminus \{s_i\}$-maximal. Then

$$\begin{aligned} 1 = P(s_1, \emptyset, [A \setminus \{s_1\}]) &= P_1(s_2) + P_1(A \setminus \{s_1, s_2\}) \text{ and} \\ 1 = P(s_2, \emptyset, [A \setminus \{s_2\}]) &= P_2(s_1) + P_2(A \setminus \{s_1, s_2\}) \end{aligned}$$

We want to prove that for all $s_j \in A \setminus \{s_1\}$, $P(s_j, [A \setminus \{s_1, s_j\}]) = 1$.

$$\begin{aligned} P(s_j, \emptyset, [A \setminus \{s_1, s_j\}]) &\geq P_j([A \setminus \{s_1, s_j\}]) + P_j(s_1)P(s_1, \emptyset, [A \setminus \{s_1, s_j\}]) \\ &\geq P_j([A \setminus \{s_1, s_j\}]) \\ &\quad + P_j(s_1)(P_1([A \setminus \{s_1, s_j\}]) + P_1(s_j)P(s_j, \emptyset, [A \setminus \{s_1, s_j\}])). \end{aligned}$$

Thus since $P_j(s_1)P_1(s_j) < 1$ (otherwise $s_1 R s_j$), we have

$$P(s_j, \emptyset, [A \setminus \{s_1, s_j\}]) \geq \frac{P_j([A \setminus \{s_1, s_j\}]) + P_j(s_1)P_1([A \setminus \{s_1, s_j\}])}{1 - P_j(s_1)P_1(s_j)}.$$

But this fraction is equal to 1 because of the two equalities above. Thus it follows that $P(s_j, \emptyset, [A \setminus \{s_1, s_j\}]) = 1$.

Thus we can remove any finite set of states from $A$, and still have it satisfy the property $((\forall s_i \in A)\ [P(s_i, \emptyset, A \setminus \{s_i\}) = 1])$. In particular, if $A_i \stackrel{d}{=} A \setminus \{s_1, \ldots, s_i\}$, then $P(s_1, \emptyset, A_i) = 1$ for all $i \geq 1$. Thus $P(s_1, \emptyset, \cap_i A_i) = 1$, or $P(s_1, \emptyset, \emptyset) = 1$, which is a contradiction. ∎

# B Complete Proof of Theorem 2.8

**Theorem B.1** *Given an LCMC which satisfies the property that the total of all the probabilities from any probabilistic state is 1, if states $s$ and $t$ in it are bisimilar then they are bisimilar according to the definition of Lee, Philippou and Sokolsky [PLS00].*

**Proof.**

The definition of [PLS00] is as follows: (we have recast the definitions in terms of computations rather than schedulers)

An equivalence relation $R \subseteq S \times S$ is a weak bisimulation iff whenever $sRt$, then

- if $s, t \in S_n, \alpha \in Act$ and $(s, \alpha, s') \in \longrightarrow$, then there exists a computation $C$ such that $P^C(t, \{\alpha\}, [s']) = 1$.
- there exists a computation $C$ such that for all $M \in S/R - [s]$, $\mu_R(s, M) = P^C(t, \emptyset, M)$.

$\mu_R$ is the probability distribution from $s \in S_p$ "normalized" by weighting by the probability of exiting $[s]$:

$$\mu_R(s, M) \quad = \quad \begin{cases} \mu(s, M), \text{ if } \mu(s, [s]) = 1 \\ \frac{\mu(s,M)}{1-\mu(s,[s])}, \text{ otherwise} \end{cases}$$

Now we will show that the relation $\approx$ satisfies both these conditions. The first condition is satisfied easily: If $(s, \alpha, s') \in \longrightarrow$, $P(s, \{\alpha\}, [s']) = 1$. Since $s \approx t$, $P(t, \{\alpha\}, [s']) = 1$, and using Lemma 2.6, we have an $\{\alpha\}$-computation $C$ such that $P^C(t, \{\alpha\}, [s']) = 1$.

For the second condition, let $s \in S_p$. Let $s_i$, $i = 1, \ldots$ be the targets of the probabilistic transition from $s$ that are not $\approx$-related to $s$. Our proof for this case proceeds in the following steps.

If $t \in S_n$, we show that there exists $t' \approx t$ such that $(t, \tau, t') \in \longrightarrow$, thus reducing this case to the case when $t$ is probabilistic.

If $t \in S_p$,we show that the targets of the probabilistic transition from $t$ are precisely the $s_i$ with identical "normalized" probabilities.

- Suppose $t \in S_n$. In this case, we will show that there exists $t' \approx t$ such that $(t, \tau, t') \in \longrightarrow$. For each $\approx$-closed set $E$ there is a state $t_E$ belonging to the targets of $\tau$-transitions from $t$ such that $P(t, \emptyset, E) = P(t_E, \emptyset, E)$; this state is given by the computation obtained from Lemma 2.6. Let $A = [s_1, \ldots, s_n]$. Then $P(t_A, \emptyset, A) = P(t, \emptyset, A) = P(s, \emptyset, A) = 1$. Also for all $E$, $1 = P(t, \emptyset, [t_E]) = P(s, \emptyset, [t_E])$. If for *any* $E$, $[t] = [t_E]$, then we have $t_E \approx t$, and we can apply case 2. Otherwise it follows that for all $E$, $P(s_i, \emptyset, [t_E]) = 1$ for all $s_i$ and hence for every element of $A$. Thus $P(t_A, \emptyset, [t_E]) = 1$ and hence

  $$P(t_A, \emptyset, E) \geq P(t_E, \emptyset, E) = P(t, \emptyset, E) \geq P(t_A, \emptyset, E).$$

  From this we can show that $t_A \approx s \approx t$ and we can apply the following case.

- Let $t \in S_p$. If $t$ has a probability 1 transition to another state, then it is bisimilar to that state, reducing us to the case above. Otherwise, w.l.o.g., assume that: (i) none of the $s_i$ are bisimilar, or bisimilar to $s$.
  (ii) The targets of the probabilistic transition from $t$ are also $s_1, s_2, \ldots$. This is possible because there are only countably many bisimulation classes, and these could be the $s_i$'s, some with 0 probability.

  Suppose that the "normalized" probability assigned by $s$ (resp. $t$) to $s_i$ is $p_i$ (resp. $p'_i$). Let $A$ be a set of states such that for any state $s_j \notin A$, $p_i = p'_i$ ($A$ can be the empty set). Then $P(s, \emptyset, A) = \sum_{s_i \in A} p_i + \sum_{s_j \notin A} p_j P(s_j, A)$. Using the same equality on the $t$ side, and using $P(s, \emptyset, A) = P(t, \emptyset, A)$, we have that $0 = \sum_{s_i \in A}(p_i - p'_i)$.

  Now by Lemma A.1 there exists an $s_k \in A$ such that $P(s_k, \emptyset, A \setminus \{s_k\}) < 1$. Now $P(s, \emptyset, A \setminus \{s_k\}) = \sum_{s_i \in A, i \neq k} p_i + p_k P(s_k, \emptyset, A \setminus \{s_k\}) + \sum_{s_j \notin A} p_j P(s_j, A \setminus \{s_k\})$. By using a similar

equality for $t$, we have $0 = \sum_{s_i \in A, i \neq k}(p_i - p'_i) + (p_k - p'_k)P(s_j, A \setminus \{s_k\})$. Subtracting this equation from the previous equation, we have $(p_k - p'_k)(1 - P(s_j, A \setminus \{s_k\}) = 0$, which means that $p_k = p'_k$, as $P(s_j, A \setminus \{s_k\}) < 1$.

Thus if the set $\{s_i \mid p_i \neq p'_i\}$ was non-empty, we can derive a contradiction as shown above. So $p_i = p'_i$ for all $i$.

∎