Bisimulation and Cocongruence for Probabilistic Systems

Vincent Danos*
Équipe PPS
Université Paris 7 and CNRS
Paris, France
Vincent.Danos@pps.jussieu.fr

Josée Desharnais†
Département d'Informatique
Université Laval
Québec, Canada
josee.desharnais@ift.ulaval.ca

François Laviolette†
Département d'Informatique
Université Laval
Québec, Canada
francois.laviolette@ift.ulaval.ca

Prakash Panangaden‡
Computing Laboratory
University of Oxford
United Kingdom
prakash@cs.mcgill.ca

January 20, 2005

Abstract

We introduce a new notion of bisimulation, called *event* bisimulation on labelled Markov processes (LMPs) and compare it with the, now standard, notion of probabilistic bisimulation, originally due to Larsen and Skou. Event bisimulation uses a sub σ-algebra as the basic carrier of information rather than an equivalence relation. The resulting notion is thus based on measurable subsets rather than on points: hence the name. Event bisimulation applies smoothly for general measure spaces; bisimulation, on the other hand, is known only to work satisfactorily for analytic spaces. We prove the logical characterization theorem for event bisimulation without having to invoke any of the subtle aspects of analytic spaces that feature prominently in the corresponding proof for ordinary bisimulation. These complexities only arise when we show that on analytic spaces the two concepts co-incide.

We show that the concept of event bisimulation arises naturally from taking the co-congruence point of view for probabilistic systems. We show that the theory can be given a pleasing categorical treatment in line with general coalgebraic principles. As an easy application of these ideas we develop a notion of "almost sure" bisimulation; the theory comes almost "for free" once we modify Giry's monad appropriately.

1 Introduction

Markov processes with continuous state spaces or continuous time evolution (or both) arise naturally in several fields of physics, biology, economics and computer science. Examples of such systems are brownian motion, gas diffusion, population growth models, noisy control systems and communications systems.

Labelled Markov Processes (LMPs) were formulated [BDEP97, DEP02] to study such general *interacting* Markov processes. In an LMP, instead of one transition probability function (or Markov

*In love and promoted!

†Research supported by NSERC.

‡Research supported by NSERC (Canada) and EPSRC (U.K.). On leave from McGill University.

kernel) there are several, each associated with a distinct label[1]. Each such transition probability function represents the (stochastic) response of the system to an external stimulus represented by the label. In our work we do not associate probabilities with these external stimuli; in other words, we do not intend to quantify the behaviour of the environment. Thus, for those familiar with process algebra terminology, an LMP is a labelled transition system with probabilistic transitions. The interaction is captured by synchronizing on labels in the manner familiar from process algebra.

The following example, taken from [DGJP03], illustrates these ideas.

Example 1.1 Consider the flight management system of an aircraft. It is responsible for monitoring the state of the aircraft – the altitude, windspeed, direction, roll, yaw etc. – periodically (usually several times a second), it also monitors navigational data from satellites and makes corrections, as needed, by issuing commands to the engines and the wing flaps. The physical system is a complex continuous real-time stochastic system; stochastic because the response of the physical system to commands cannot be completely deterministic and also because of unexpected situations like turbulence. From the point of view of the flight management system, however, the system is discrete-time and has continuous space. The time unit is the sampling rate. The entire system consists of many interacting concurrent components and programming it correctly – letting alone verifying that the system works – is very challenging. A formal model of this type of software brings us into the realm of process algebra, because of the concurrent interacting components, stochastic processes and real-time systems, the last because the responses have hard deadlines.

This study was initiated by Larsen and Skou [LS91] for *discrete* processes in a style similar to the queueing theory notion of "lumpability" invented in the late 1950s [KS60]. In a series of previous papers [BDEP97, DEP98, DEP02] such Markov processes with continuous state spaces and independently acting components were studied, and the phrase "labelled Markov processes" appeared in print explicitly referring to the continuous state space case. Of course, closely related concepts were already around: for example, Markov decision processes [Put94]. The papers by Desharnais, Edalat and Panangaden gave a definition of bisimulation between LMPs, and gave a logical characterization of this bisimulation. Subsequently an approximation theory was developed [DGJP00, DGJP03, DD03] and metrics were defined [DGJP99, DGJP04, vBW01b, vBW01a].

Before we begin the present paper we will briefly review the prior results. The notion of probabilistic bisimulation - henceforth just "bisimulation" - was based on the idea that if two states are bisimilar then their transition probabilities to bisimulation equivalence classes should match. This notion works well for the discrete case, but has to be generalized appropriately to the continuous case. One idea was to mimic this definition exactly with a few measure-theoretic conditions imposed to deal with the fact that not all sets need be measurable. This was the approach followed in [DGJP00, DGJP03]. However, in an earlier approach [BDEP97, DEP98, DEP02] the authors had defined a bisimulation - they called it a "zig-zag" - morphism and then defined a bisimulation relation as a span of such morphisms. This also generalizes the discrete case but it turned out to be very painful to prove that one gets a transitive relation and one had to restrict to Polish or analytic spaces. For such cases (analytic subsumes Polish) the two notions coincide.

One of the nice things about the theory was the logical characterization of bisimulation. There was already such a theorem for the discrete case in the paper of Larsen and Skou [LS91] but it was not clear that such a theorem would work in the continuous state case[2]. Not only did such a theorem exist but logical characterization worked with a much more parsimonious logic than one

[1]We do not consider *internal* nondeterminism in the present paper.

[2]In retrospect this seems like a misplaced worry but it was a real worry at the time.

was led to expect from the discrete case. There turned out to be a very spartan logic:

$$\phi ::== \top | \phi_1 \wedge \phi_2 | \langle a \rangle_q \phi$$

which characterises bisimulation for both continuous and discrete systems. There are two striking things about this logic: *there is no negative construct at all* and one only needs binary conjunction *even though the branching may be uncountable.* The proof heavily uses special properties of analytic spaces that à priori have nothing to do with anything logical.

This is an irritating fact: one has to restrict to state-spaces that were analytic. In one sense this is not very restrictive: almost every process that one can imagine has an analytic state space. In particular $\mathbf{R}^n$ with the usual Borel sets is analytic, indeed any manifold is analytic. In another sense it is conceptually unsatisfying that these notions specific to measure theory *on metric spaces* should turn out to be so crucial. Why doesn't the theory work for general measure spaces? Certainly the *statement* of the logical characterization theorem does not suggest anything about analytic spaces.

2 The road to event bisimulation

The first attempt to define LMPs for continuous systems [BDEP97] did not have any assumptions about the σ-algebra on the state space. LMPs were organized in a category and bisimulation was defined in terms of spans of particular morphisms of this category, called zig-zag morphisms. However, buried in the proofs was an alternative view of bisimulation as a cospan: in fact this is the germ of the co-congruence idea that we develop in the present paper. With this definition, one could prove that the logic $\mathcal{L}_0$ characterizes bisimulation. This was, however, viewed as an intermediate step at the time: firstly, it was a compromise to use cospans instead of spans of morphisms as is usually done when one wants to define a *relation* between objects of a category. Spans could not be used because one could not show that bisimulation was transitive; indeed, this is equivalent to constructing a span given a cospan, and to this day, only analytic spaces have been proven to satisfy this property [Eda99, Dob03]. Secondly, the definition of bisimulation was not given on the state-space of a process but, rather, through morphisms in the category: this is not what one was used to working with in the finite case or in the non-probabilistic case. Consequently, a new relational definition of bisimulation was formulated that looked like a nice generalization of non-probabilistic bisimulation as well as of finite probabilistic bisimulation. However, for this definition as well, characterization of bisimulation by the logic was only proven for analytic spaces.

In this paper, we give a new definition of bisimulation, called *event-bisimulation*, that it equivalent with the cospan definition and hence is characterized by the logic. This is proved without any analyticity assumption: it works for arbitrary measure spaces. Moreover, it has the nice property that the logic yields the biggest possible event-bisimulation on any process. We compare this definition with the relational bisimulation mentioned above, which we call *state-bisimulation* if we want to emphasize the difference, and show that roughly speaking, the latter is finer that the latter, and hence equates fewer states than the former in general.

We also show that the two notions are exactly the same for countable processes and, more generally, for analytic spaces. Consequently, we believe that event-bisimulation is the correct generalization of state-bisimulation to the category of all LMPs defined for arbitrary measure spaces.

This new definition also has the advantage to reconcile the two fields of theory of processes and probability theory; it allows one to use the σ-algebra as a vector of information to define event-bisimulation between processes.

3 Background on LMPs

Labelled Markov processes are probabilistic versions of labelled transition systems. Corresponding to each label a Markov process is defined. The transition probability is given by a *Markov kernel*. In brief, a labelled Markov process can be described as follows. There is a set of states and a set of labels. The system is in a state at each point in time. The environment selects an action, and the system reacts by moving to another state. The transition to another state is governed by a probabilistic law. For each label there is a transition probability distribution which gives the probability distribution of the possible final states given the initial state. For discrete state spaces, this is essentially the model developed by Larsen and Skou [LS91].

We extended this to continuous-state systems, thus forcing our formalism to be couched in measure-theoretic terms. For instance, we cannot ask for the transition probability to any set of states — we need to restrict ourselves to measurable sets. The classical theory of Markov processes is typically carried out in the setting of Polish spaces rather than on abstract measure spaces. In previous papers *analytic spaces* - which generalize Polish spaces - were used. However, in this paper we eliminate the need for analytic spaces.

A key ingredient in the theory is the *Markov kernel*, which we sometime call *transition kernel*.

Definition 3.1 *A **Markov kernel** on a measurable space (S, Σ) is a function $\tau : S \times \Sigma \to [0, 1]$ such that for each fixed $s \in S$, the set function $\tau(s, \cdot)$ is a (sub-) probability measure, and for each fixed $X \in \Sigma$ the function $\tau(\cdot, X)$ is a measurable function.*

One interprets $\tau(s, X)$ as the probability of the process starting in state s making a transition into one of the states in X. The Markov kernel is really a *conditional probability*: it gives the probability of the process being in one of the states of the set X after the transition, *given* that it was in the state s before the transition.

We will work with transition kernels where $\tau(s, S) \leq 1$ rather than $\tau(s, S) = 1$. The mathematical results go through in this extended case. We view processes where the transition kernels are only sub-probabilities as being *partially defined*.

Definition 3.2 *A **labelled Markov process (LMP)** $\mathcal{S}$ with label set $\mathcal{A}$ is a structure $(S, \Sigma, \{\tau_a | a \in \mathcal{A}\})$, where S is the set of states, and Σ is a σ-algebra on S, and*

$$\forall a \in \mathcal{A}, \tau_a : S \times \Sigma \longrightarrow [0, 1]$$

is a Markov kernel.

We will fix the label set to be $\mathcal{A}$ once and for all. We will use the following notational convention: we write $\mathcal{S} = (S, \Sigma, \tau)$, using the calligraphic font to stand for the LMP and the ordinary capital for the state space. We often drop the subscript of τ when convenient since it does not restrict the results.

The all important notion is that of a zig-zag morphism.

Definition 3.3 *A **zig-zag** morphism f from $\mathcal{S}$ to $\mathcal{S}'$ is a surjective measurable function $f : S \to S'$ satisfying*

$$\forall a \in \mathcal{A}, s \in S, B \in \Sigma'. \tau_a(s, f^{-1}(B)) = \tau'_a(f(s), B).$$

We originally defined a bisimulation in terms of spans of zig-zags. In order to show transitivity we had to use a subtle construction due to Edalat [Eda99].

Later we gave a direct relational definition and finessed the use of that lemma. Here is the direct relational definition. If R is a binary relation on a set S, we say that $A \subseteq S$ is *R-closed* if $\{s \in S : \exists a \in A. aRs\} \subseteq A$. We denoted by $\Sigma(R)$ the set of R-closed sets in Σ.

Definition 3.4 *Given an LMP $\mathcal{S}$ we define a* bisimulation *relation R as a binary relation on S such that whenever sRt and $C \in \Sigma(R)$, then for all labels a, $\tau_a(s,C) = \tau_a(t,C)$. We say that s and t are* ***bisimilar*** *if there is any bisimulation R such that sRt.*

One can define a simple modal logic and prove that two states are bisimilar if and only if they satisfy exactly the same formulas. Indeed for finite-state processes one can decide whether two states are bisimilar and effectively construct a distinguishing formula in case they are not [DGJP02].

As before we assume that there is a fixed set of "actions" $\mathcal{A}$. The logic is called $\mathcal{L}$ and has the following syntax:

$$\mathsf{T} \mid \phi_1 \wedge \phi_2 \mid \langle a \rangle_q \phi$$

where a is an action and q is a rational number between 0 and 1. This is the basic logic with which one can establish the logical characterization. The last formula is interpreted as follows. We say $s \models \langle a \rangle_q \phi$ if s can make an a-transition with probability greater than or equal to the rational number q and end up in a state satisfying ϕ.

In the analysis of simulation one needs a logic with disjunction, $\mathcal{L}_\vee$:

$$\mathcal{L} \mid \phi_1 \vee \phi_2.$$

The logical characterization theorem uses a remarkably parsimonious logic: no negation, no infinitary constructs. It works only when the state space is an analytic space. In the present paper we will show that the new notion of bisimulation leads to the logical characterization theorem for LMPs defined on arbitrary measure spaces.

4 Event and state bisimulation

In defining a notion of probabilistic bisimulation one is forced to add transition probabilities. It makes no sense to compare labels and transition probabilities between individual transitions. If one were to do so then obvious examples would fail to be bisimilar. Consider a three-state system with a-transitions from state 1 to states 2 and 3, each with a probability $\frac{1}{2}$, and compare it with a two-state system with a probability 1 a-transition from state 1 to state 2. The question is how should one aggregate the states?

Equivalence relations are the first thing that leaps to mind and this leads to the relational definition of bisimulation [DGJP03]. One can think of relations as spans of appropriate functions; in this case we have the zig-zags [DEP02] described in the last section. On the other hand there is a natural family of subsets at hand: the measurable sets. One can think of the σ-algebra and sub-σ-algebras as defining families of interesting sets of states. It is curious that measurability does not crop up as a restriction on the possible bisimulation relations. To be sure, one has to look at R-closed measurable sets but there is no restriction such as that R should be an equivalence relation with measurable equivalence classes (it would be hard to prove transitivity if measurability was imposed on R). The notion of event bisimulation puts measurability front and centre.

It will be helpful to reformulate the relational notion of bisimulation in order to facilitate the switch to the new proposed definition.

Lemma 4.1 *R is a bisimulation iff $(S, \Sigma(R), \tau)$ is an LMP.*

Proof: $(S, \Sigma(R), \tau)$ is an LMP iff $\forall A \in \Sigma(R)$, $\tau(\cdot, A)$ is $\Sigma(R)$-measurable. By definition, R is a bisimulation iff $\forall A \in \Sigma(R)$, $\tau(\cdot, A)$ is constant on R-classes. We show that this is equivalent to saying that $\tau(\cdot, A)$ is $\Sigma(R)$-measurable, which will imply the result. Let $A \in \Sigma(R)$. If $\tau(\cdot, A)$ is constant on R-classes, $\{s|\tau(s,A) \geq r\} \in \Sigma$ because $\tau(\cdot,A)$ is measurable. Also $\{s|\tau(s,A) \geq r\}$ is R-closed because $\tau(\cdot,A)$ is constant. For the converse, if $\tau(\cdot,A)$ is $\Sigma(R)$-measurable, then for every $r \in [0,1]$, $\{s|\tau(s,A) \geq r\}$ is R-closed: this implies that $\tau(\cdot,A)$ is constant on R-classes. ∎

This shows that one can work with a smaller σ-algebra closely connected with bisimulation equivalence. In fact one has a new LMP with the same kernel and state space but with a reduced σ-algebra. It is easy to see that

Lemma 4.2 *R is a bisimulation iff the identity map $i : (S,\Sigma,\tau) \to (S,\Sigma(R),\tau)$ is a zig-zag.*

Ideally one would like it to be the case that any zig-zag induces a bisimulation on its domain. However this is false. Given a zig-zag $f : (S,\Sigma,\tau) \to (S',\Sigma',\tau')$ we can conclude that if $f(s_1) = f(s_2)$ then $\tau(s_1,A) = \tau(s_2,A)$ for every A in $f^{-1}(\Sigma')$ and not for every A in $\Sigma(R)$, where R is the relation induced on S by f. Thus the equivalence induced by a zig-zag is too fine in general. The crucial point is that we are making the equivalence relation primary and the σ-algebra secondary. Instead we should work with the structure naturally associated with a σ-algebra: in other words we should look for a sub-σ-algebra.

Definition 4.3 *An **event bisimulation** on an LMP (S,Σ,τ) is a sub-σ-algebra Λ of Σ such that (S,Λ,τ) is an LMP.*

We will refer to Λ and the associated equivalence $\mathfrak{R}(\Lambda)$ as event bisimulation. The phrase "event bisimulation" is meant to suggest that the focus of interest has shifted from the individual points of S to the measurable sets, or - to emphasize the probabilistic interpretation - the events. What has happened here is that the notion of bisimulation qua relation has been replaced by an arbitrary sub-σ-algebra rather than $\Sigma(R)$, the sub-σ-algebra generated by a relation R. The key point is that the transition kernels have to have the appropriate measurability properties with respect to Λ. We can more sensibly say that Λ, rather than $\mathfrak{R}(\Lambda)$, is an event bisimulation.

The similarity with state bisimulation can be made quite striking.

Lemma 4.4 *If $\mathfrak{R}(\Lambda)$ is an event bisimulation, then the identity function on S defines a zig-zag morphism from (S,Σ,τ) to (S,Λ,τ).*

If we compare this result to Lemma 4.2, we can see that it fits more nicely in the category of LMPs in that it does not talk about relations. Moreover, we get a perfect correspondence with zig-zags.

Lemma 4.5 *If $f : (S,\Sigma,\tau) \to (S',\Sigma',\tau')$ is a zig-zag morphism, then $\mathfrak{R}(f^{-1}(\Sigma'))$ is an event bisimulation on S.*

In order to facilitate the study of the relation between state and event bisimulation, we need some elementary mathematical observations connecting σ-algebras and binary relations. Let R be a relation on S, which is a set equipped with a σ-algebra Σ. We write $\Upsilon(R)$ for the R-closed sets of R. Then $\Sigma(R) = \Sigma \cap \Upsilon(R)$. This is a sub-$\sigma$-algebra of Σ. If $\Lambda \subseteq \Sigma$ is a sub-σ-algebra of Σ we define $\mathfrak{R}(\Lambda)$ an equivalence relation on S by $(s,t) \in \mathfrak{R}(\Lambda)$ if and only if $\forall A \in \Lambda . s \in A \Leftrightarrow t \in A$.

We have two maps back and forth between sub-σ-algebras and equivalence relations: $\Lambda \mapsto \mathfrak{R}(\Lambda)$ and $R \mapsto \Sigma(R)$.

Lemma 4.6 *$\Lambda \subseteq \Sigma(\mathfrak{R}(\Lambda))$.*
$R \subseteq \mathfrak{R}(\Sigma(R))$.
If R-equivalence classes are in Σ, then $R = \mathfrak{R}(\Sigma(R))$.

Proof . First let A be in Λ, then $A \in \Sigma$; also A is $\mathfrak{R}(\Lambda)$-closed. Thus $A \in \Sigma \cap \Upsilon(\mathfrak{R}(\Lambda)) = \Sigma(\mathfrak{R}(\Lambda))$.

Next, we have $[s]_{\mathfrak{R}(\Sigma(R))} = \cap A \ni s$ with $A \in \Sigma$ and R-closed; and $[s]_R = \cap B \ni s$ with B R-closed; so $\subseteq$ follows. Now for $\supseteq$, if $[s]_{\mathfrak{R}}$ itself is in Σ then it qualifies as an A in the big intersection above and the result follows. ∎

Lemma 4.7 $\mathfrak{R}(\Lambda) = \mathfrak{R}(\Upsilon(\mathfrak{R}(\Lambda)))$.

Proof . $\mathfrak{R}(\Lambda) \subseteq \mathfrak{R}(\Upsilon(\mathfrak{R}(\Lambda)))$ is given is by Lemma 4.6 and the reverse inclusion follows from $\Lambda \subseteq \Upsilon(\mathfrak{R}(\Lambda))$. ∎

Proposition 4.8 *If* $\Lambda \subseteq \Sigma$ *then* $\mathfrak{R}(\Lambda) = \mathfrak{R}(\Sigma(\mathfrak{R}(\Lambda)))$.

Proof . Since $\Lambda \subseteq \Sigma \cap \Upsilon(\mathfrak{R}(\Lambda)) \subseteq \Upsilon(\mathfrak{R}(\Lambda))$ and by Lemma 4.7, we have
$\mathfrak{R}(\Lambda) \supseteq \mathfrak{R}(\Sigma \cap \Upsilon(\mathfrak{R}(\Lambda))) \supseteq \mathfrak{R}(\Upsilon(\mathfrak{R}(\Lambda))) = \mathfrak{R}(\Lambda)$. ∎

These lemmas and proposition show how to transfer results between sub-σ-algebras and equivalence relations.

We presented event-bisimulation as a weakening of state-bisimulation. Consequently, we would like to prove that the latter is always an event-bisimulation. However, this is not actually the case in general: essentially because there is not enough measure theoretic control on the relations used to define state bisimulation.

We know from Lemma 4.6 that $R \subseteq \mathfrak{R}(\Sigma(R))$. The following lemma shows that if a state bisimulation satisfies the reverse inclusion, then it is an event bisimulation.

Lemma 4.9 *If relation R is a state bisimulation, then it is an event bisimulation iff* $R = \mathfrak{R}(\Sigma(R))$.

Proof . If R is a state bisimulation and an event bisimulation, then $R = \mathfrak{R}(\Lambda)$ for some Λ. We have that $\Lambda \subseteq \Sigma(R)$ because every $A \in \Lambda$ is R-closed. Thus $R = \mathfrak{R}(\Lambda) \supseteq \mathfrak{R}(\Sigma(R))$. The reverse inclusion is given by Lemma 4.6 . Conversely, if R is a state bisimulation we know that $(S, \Sigma(R), \tau)$ is an LMP, so R is an event bisimulation for $\Lambda = \Sigma(R)$. ∎

Corollary 4.10 *If relation R is a state bisimulation with equivalence classes in Σ, then R is an event bisimulation.*

Proof . If R is a state bisimulation, then by lemma 4.1 $(S, \Sigma(R), \tau)$ is an LMP; so set $\Lambda = \Sigma(R)$ and we have that $R = \mathfrak{R}(\Sigma(R))$ by lemma 4.6, so R is an event bisimulation, by definition. ∎

This immediately implies that a state bisimulation is an event bisimulation in the countable case, where every set is measurable.

Corollary 4.11 *If S is countable and $\Sigma = \mathcal{P}(S)$, any state bisimulation is an event bisimulation.*

We cannot expect that every state bisimulation satisfies the equality in Lemma 4.9, as the following example shows.

Example 4.12 There exists an equivalence relation R - even defined on the *analytic* space $([0,1], \mathcal{B}, \tau)$ - such that $R \subset \mathfrak{R}(\Sigma(R))$. Let $V \subseteq S$ be non Lebesgue-measurable, suppose R to be given by two equivalence classes, V and V^c: $\Sigma(R)$ is the trivial σ-algebra, so $\mathfrak{R}(\Sigma(R))$ equates everything and therefore is different from R.

One could say $R = \mathfrak{R}(\Lambda)$ for some Λ but then Λ cannot separate V and therefore $\Lambda = \{\varnothing, V, V^c, S\}$ so we can pick two s and s' (there must be two of them in the case of $[0,1]$ and the Borel sets) in V and define τ such that $\tau(s,S) = \tau(s',S) = 1$, but $\tau(s,\{s\}) = 1$ and $\tau(s',\{t\}) = 1$ with $t \notin V$ (this way R is a state bisimulation because the only non-empty measurable in $\Sigma(S)$ is $S = [0,1]$), but R is not an event bisimulation for Λ because the kernel τ is not Λ-measurable.

The conclusion is that, even on an analytic space, not every state bisimulation is an event bisimulation. This may seem to kill any hope to show that a state bisimulation is an event bisimulation. However, one can observe that when this is the case, the state bisimulation distinguishes too many states, in particular it distinguishes states that are not separable by Σ. This is a liberty that an event bisimulation never has. The following result shows that assuming the fact that "a bigger state bisimulation is better", we do have that result.

Proposition 4.13 *If R is a state bisimulation, then $\mathfrak{R}(\Sigma(R))$ is a state bisimulation and an event bisimulation.*

Proof . By Lemma 4.2, if R is a state bisimulation, then the identity morphism $i : (S, \Sigma, \tau) \to (S, \Sigma(R), \tau)$ is a zig-zag. Since $i^{-1}(\Sigma(R)) = \Sigma(R)$, by Lemma 4.5 we have that $\mathfrak{R}(\Sigma(R))$ is an event bisimulation. ∎

Another route to the result is without going through zig-zag morphisms.

Lemma 4.14 *If R is a state bisimulation, then $\Sigma(R) = \Sigma(\mathfrak{R}(\Sigma(R)))$.*

Proof . $\Sigma(R) \supseteq \Sigma(\mathfrak{R}(\Sigma(R)))$ because $R \subseteq \mathfrak{R}(\Sigma(R))$. For inclusion, let $X \in \Sigma(R)$, then X is $\mathfrak{R}(\Sigma(R))$-closed and in Σ. Thus it is in $\Sigma(\mathfrak{R}(\Sigma(R)))$. ∎

Alternate proof of Proposition 4.13: An easy application of Proposition 4.8 and Lemma 4.9 gives us that $\mathfrak{R}(\Sigma(R))$ is an event bisimulation. The lemma just above and Lemma 4.1 proves that it is a state bisimulation. ∎

Proposition 4.13 implies that if a state bisimulation is not an event bisimulation, then it can be expanded to one that is. One example is the identity relation, which is not an event bisimulation when Σ does not separate points, but it is a state bisimulation. This is an example of a relation that sees more differences than Σ can see.

Lemma 4.15 *The identity relation I is a state bisimulation; it is an event bisimulation iff Σ separates points.*

Proof . $\Sigma(I) = \Sigma$ and lemma 4.1 proves the first point. For the second one, suppose $I = \mathfrak{R}(\Lambda)$ for some $\Lambda \subseteq \Sigma$; then Λ separates points in S, so Σ also does; in fact $I = \mathfrak{R}(\Sigma)$ in this case.

Conversely if Σ separates points, then $I = \mathfrak{R}(\Sigma)$ and obviously (S, Σ, τ) is an LMP, so I is an event bisimulation. ∎

One interpretation of this is that the correct "identity relation" in measure spaces is the one generated by the sigma-algebra rather than the usual one defined on the points.

4.1 Analytic spaces

We have already shown that for countable spaces the two notions, state and event bisimulation coincide. In fact they coincide for the vastly larger class of analytic spaces. Indeed the proofs of logical characterization of bisimulation given in previous papers essentially establish this fact though it is hidden in the proofs. They hinge on the special properties of countably generated σ-algebras. The following lemma from [DP01] uses the unique structure theorem of analytic spaces. If $\mathcal{C} \subseteq \Sigma$, we write $\sigma(\mathcal{C})$ for the smallest σ-algebra containing $\mathcal{C}$.

Lemma 4.16 ([DP01]) *Let (S, Σ) be an analytic space. Let $\mathcal{C} \subseteq \Sigma$ be countable and assume $S \in \mathcal{C}$. Then $\Sigma(\mathfrak{R}(\mathcal{C})) = \sigma(\mathcal{C})$.*

Lemma 4.17 (Bridge lemma) *If Σ is analytic and Λ an event bisimulation such that $\Lambda = \sigma(\mathcal{C})$ for some countable $\mathcal{C} \subseteq \Sigma$, then $\mathfrak{R}(\Lambda)$ is a state bisimulation.*

Proof . By Lemma 4.16, we have $\Sigma(\mathfrak{R}(\Lambda)) = \sigma(\Lambda) = \Lambda$. Since Λ is an event bisimulation (S, Λ, τ) is an LMP. Consequently, $(S, \Sigma(\mathfrak{R}(\Lambda)), \tau)$ is an LMP and hence $\mathfrak{R}(\Lambda)$ is a state bisimulation. ∎

This implies that in analytic spaces, the maximal state bisimulation and event bisimulation are equal.

Corollary 4.18 *In the countable case state bisimulation is exactly event bisimulation whenever the sigma-algebra considered is the powerset.*

5 Logical characterization

The logical characterization of probabilistic bisimulation [DEP98, DEP02] as it was originally proved worked with what we are now calling state bisimulation and was established for analytic spaces. In the present section we establish the logical characterization of bisimulation for event bisimulation for LMPs defined on general measure spaces. In conjunction with the results of the previous section - that the two notions are essentially the same on analytic spaces - it implies the earlier logical characterization of bisimulation result. Moreover, it shows that the role of analytic spaces can be confined to a single lemma, namely Lemma 4.16. At the end of the section we explicate the measure theoretic significance of the particular logic $\mathcal{L}_0$ that we used.

We recall the logic

$$\top \mid \phi_1 \wedge \phi_2 \mid \langle a \rangle_q \phi.$$

Given a formula ϕ we write $[\![\phi]\!]$ for the set of states satisfying the formula ϕ. It is easy to see that these are all measurable sets. We write $[\![\mathcal{L}_0]\!]$ for the collection of sets of the form $[\![\phi]\!]$. We write $\sigma(\mathcal{L}_0)$ for the σ-algebra generated by $[\![\mathcal{L}_0]\!]$: we call this the σ-algebra generated by the logic. The key point that we shall establish is that the σ-algebra generated by the logic is an event bisimulation; moreover it is the maximal event bisimulation. From this the logical characterization of event bisimulation is immediate.

The proofs depend on properties of π-systems and d-systems. We recall the basic definitions from the literature [Wil91].

Definition 5.1 *Let S be a set: (1) a π-system on S is a subset of $\mathcal{P}(S)$ closed under intersections and containing S, (2) a d-system $\mathcal{D}$ on S is a subset of $\mathcal{P}(S)$ containing S, closed under increasing unions and relative complements (i.e., if $A, B \in \mathcal{D}$ and $A \subseteq B$, then $B \setminus A \in \mathcal{D}$).*

The point of π-systems is that one can often work with them instead of the σ-algebras that they generate; usually the sets of a π-system are much simpler than the sets in the generated σ-algebra.

Another key concept, and the one that brings out the special role of the logic $\mathcal{L}_0$ is stability.

Definition 5.2 *Let (S, Σ, τ) be an LMP and $\Lambda \subseteq \Sigma$, we say that Λ is **stable** with respect to (S, Σ, τ) if for all $A \in \Lambda, r \in [0,1], a \in \mathcal{A}$,*

$$\{s : \tau_a(s, A) > r\} \in \Lambda.$$

Note that Λ is an event bisimulation if and only if it is stable and that the condition of measurability of a kernel $\tau(\cdot, A)$ is exactly that Σ be stable. We will use the following notation in the rest of the paper. We write $\langle a \rangle_q A$ where A is a (measurable) set to mean $\{s : \tau_a(s, A) > q\}$ and also $\{a\}_q A$ to mean $\{s : \tau_a(s, A) \leq q\}$. Clearly both these sets are measurable if A is and τ is a Markov kernel. We also write $A_n \uparrow A$ to mean that the family of sets A_n is nested increasing and that $\cup_n A_n = A$.

Proposition 5.3 *$[\![\mathcal{L}_0]\!]$ is the smallest stable π-system of (S, Σ, τ).*

Proof . By construction, $[\![\mathcal{L}_0]\!]$ contains S=$[\![\top]\!]$. It is a π-system because $\wedge \in \mathcal{L}_0$. It is stable because $\langle a \rangle_r \phi \in \mathcal{L}_0$ whenever $\phi \in \mathcal{L}_0$. It is the smallest because if $\mathcal{C}$ is another stable π-system:
(i) $[\![\top]\!] = S \in \mathcal{C}$;
(ii) if $[\![\phi]\!]$, $[\![\phi']\!] \in \mathcal{C}$, then $[\![\phi \wedge \phi']\!] = [\![\phi]\!] \cap [\![\phi']\!] \in \mathcal{C}$, since $\mathcal{C}$ is closed under intersection;
(iii) if $[\![\phi]\!] \in \mathcal{C}$, then $[\![\langle a \rangle_q \phi]\!] = \langle a \rangle_q([\![\phi]\!]) \in \mathcal{C}$, since $\mathcal{C}$ is stable. Thus, inductively, we have that $[\![\mathcal{L}_0]\!] \subseteq \mathcal{C}$. ∎

Lemma 5.4 *If $\mathcal{C}$ is a stable π-system of (S, Σ, τ), then $\sigma(\mathcal{C})$ is also stable.*

Proof . We show that $\mathcal{D} = \{A \in \Sigma : \forall a \forall q \, \langle a \rangle_q(A) \in \sigma(\mathcal{C})\}$ is a d-system.
(i) $S \in \mathcal{D}$ because $S \in \mathcal{C}$ and $\mathcal{C}$ is stable;
(ii) if A, $B \in \mathcal{D}$ and $A \subset B$, then

$$\langle a \rangle_q(B \setminus A) = [\cup_{r \geq q}(\{a\}_r(B) \cap (\langle a \rangle_{r-q}(A)))]^c,$$

because $\tau_a(s, B \setminus A) \leq q$ iff $\exists r \geq q \, (\tau_a(s, B) \leq r \wedge \tau_a(s, A) \geq r - q)$, where r can always be chosen rational, so we have a countable union of measurable sets and hence $\langle a \rangle_q(B \setminus A) \in \sigma(\mathcal{C})$, implying that $B \setminus A \in \mathcal{D}$;
(iii) if $A_n \in \mathcal{D}$ and $A_n \uparrow A$, then $\langle a \rangle_r(\cup_n A_n) = \cap_n \langle a \rangle_r(A_n)$, because $\tau_a(s, A) = lim \uparrow \tau_a(s, A_n)$, by the standard continuity property of measures, and hence $A \in \mathcal{D}$. This shows that $\mathcal{D}$ is a d-system.

Moreover, $\mathcal{C} \subset \mathcal{D}$, because $\mathcal{C}$ is stable, and by a well known theorem in measure theory known as the monotone class theorem [Wil91], we have that $\sigma(\mathcal{C}) \subset \mathcal{D}$. In other words, $\sigma(\mathcal{C})$ is stable. ∎

This gives us characterization of event bisimulation by the logic $\mathcal{L}_0$.

Proposition 5.5 *$\sigma([\![\mathcal{L}_0]\!])$ is the smallest stable σ-algebra included in Σ.*

Proof . Let Σ_m be the smallest stable σ-algebra included in Σ.
By Lemma 5.3, $[\![\mathcal{L}_0]\!] \subset \Sigma_m$, because Σ_m is a stable π-system, and hence $\sigma([\![\mathcal{L}_0]\!]) \subset \Sigma_m$.
Conversely, $[\![\mathcal{L}_0]\!]$ is a stable π-system by Lemma 5.3, and hence, by Lemma 5.4, $\sigma([\![\mathcal{L}_0]\!])$ is stable and hence contains Σ_m. ∎

Stability tells us that $\sigma([\![\mathcal{L}_0]\!])$ is an event bisimulation and the fact that it is the smallest gives that any event bisimulation preserves $\mathcal{L}_0$ formulas.

Moreover, this proposition yields an interesting definition of $\mathcal{L}_0$ on a pure measurability basis. Finally, we also know that we cannot do better since $\sigma([\![\mathcal{L}_0]\!])$ is the smallest stable sub-σ-algebra.

$$
\begin{tikzcd}
 & U \arrow[dl, "f"'] \arrow[d, "\gamma"'] \arrow[dr, "g"] & \\
X \arrow[d, "\alpha"'] & \Pi U \arrow[dl, "\Pi f"'] \arrow[dr, "\Pi g"] & Y \arrow[d, "\beta"] \\
\Pi X & & \Pi Y
\end{tikzcd}
$$

Figure 1: Bisimulation as a span.

6 Event bisimulation as probabilistic cocongruence

In this section we give the categorical - more precisely, coalgebraic - description of event bisimulation. We recall the categorical treatment of LMPs [Gir81, Pan99] first. For simplicity we will elide labels. In talking about bisimulation categorically it makes more sense to think of it as a relation between different LMPs rather than as a relation on the state space of a single LMP. This is a very slight shift in point of view. At the beginning we will mention the connection but as the discussion proceeds we will just talk about state and event bisimulation between different LMPs.

The base category is the category **Mes**: the objects are sets equipped with a σ-algebra and the morphisms are measurable functions. This category has pullbacks and finite products constructed just as in **Set**. More importantly, it has coequalizers and finite coproducts - also constructed just as in **Set** - and hence, all finite colimits; in particular, it has pushouts.

Giry [Gir81] defined a monad Π on **Mes** taking (X, Σ_X) to $\Pi X := \{\nu | \nu : \Sigma_X \to [0,1]\}$ where the ν are (sub)probability measures on X. One has the canonical evaluation maps $\forall A \in \Sigma_X . e_A : \Pi X \to [0,1]$ given by $e_A(\nu) = \nu(A)$. The set ΠX is equipped with the initial σ-algebra making all the e_A measurable. The arrow part of Π is $\Pi(f) = \nu \circ f^{-1}$. The monad multiplication $\mu : \Pi^2 \to \Pi$ is given by

$$\mu(\Omega)(A \in \Sigma_X) = \int_{\Pi X} e_A d\Omega$$

and the unit is $\eta_X : X \to \Pi X$ is $\eta(x) = \delta_x$, the Dirac measure concentrated at x.

Labelled Markov processes are just coalgebras of Π. A measurable function $\tau : X \to \Pi X$ is - if curried appropriately - exactly a Markov kernel. A coalgebra homomorphism is precisely a zig-zag morphism. Thus the category **LMP** of LMPs and zig-zag morphisms is just the category of coalgebras of Π. This coalgebraic presentation was developed by de Vink and Rutten [dVR97, dVR99] and noted in passing in [BDEP97].

Usually bisimulation is defined as a span in the coalgebra category as shown in Fig. 1. Here α and β define Π-coalgebras on X and Y respectively; in other words they define LMPs. The span of zig-zags given by f and g - with f and g both surjective - define a bisimulation relation between (X, α) and (Y, β). The relational notion of bisimulation viewed as an equivalence relation on the state space of an LMP is defined as the existence of the special span shown in Fig. 2. In the other direction we can talk about bisimulation between two LMPs $\mathcal{S}$ and $\mathcal{S}'$ as a relation on the state space of the co-product $\mathcal{S} + \mathcal{S}'$, which is the disjoint union $S \uplus S'$ equipped with the evident σ-algebra.

We now think of bisimulation (state and event) as a relation between two LMPs. One needs to show that bisimulation is transitive. That is given spans from $\mathcal{S}_1$ to $\mathcal{S}_2$ and from $\mathcal{S}_2$ to $\mathcal{S}_3$ we

$$
\begin{tikzcd}
 & (S, \Sigma, \tau) \arrow[dl, "i"'] \arrow[dr, "i"] & \\
(S, \Sigma, \tau) & & (S, \Sigma(R), \tau)
\end{tikzcd}
$$

Figure 2: Relational bisimulation as a span.

$$
\begin{tikzcd}
 & (U_1, \eta_1) \arrow[dl, "f_1"'] \arrow[dr, "f_2"] & & (U_2, \eta_2) \arrow[dl, "g_1"'] \arrow[dr, "g_2"] & \\
(S_1, \tau_1) & & (S_2, \tau_2) & & (S_3, \tau_3)
\end{tikzcd}
$$

Figure 3: Composing spans

would like to construct a span from $\mathcal{S}_1$ to $\mathcal{S}_3$. Given this situation we have a co-span formed by $\mathcal{U}_1$, $\mathcal{U}_2$, $\mathcal{S}_2$ and the zig-zags f_2 and g_1 as shown in Fig. 3.

Usually one postulates the existence of pullbacks, or at least weak pullbacks, in order to complete the square as shown in Fig. 4.

However, these weak pullbacks need to exist in the category of coalgebras, not just in the category **Mes**. In order for this to happen one needs to have that Π preserves weak pullbacks. This rarely happens. In the original de Vink and Rutten paper [dVR97, dVR99] the existence of weak pullbacks in the category of coalgebras was shown in the discrete setting. Edalat [Eda99] produced a much weaker construction - he called it a "semi-pullback" - which allows one to complete the square, but has no universal properties, in the category of coalgebras of Π over the base category of analytic spaces equipped with their Borel σ-algebra. This was subsequently used to show that bisimulation is transitive [BDEP97, DEP02] given the span definition of bisimulation.

With cospans everything works much more smoothly. To begin, we observe that if we work with cospans then we can compose using pushouts and *this does not require* Π *to preserve anything.* We consider the situation shown in Fig. 5 where we have omitted labels for some of the arrows, for example $\Pi f : \Pi X \to \Pi U$, where they can be inferred by functoriality. The arrows f, g, h and k are zig-zags.

$$
\begin{tikzcd}
 & & (W, \rho) \arrow[dl, "k_1"'] \arrow[dr, "k_2"] & & \\
 & (U_1, \eta_1) \arrow[dl, "f_1"'] \arrow[dr, "f_2"] & & (U_2, \eta_2) \arrow[dl, "g_1"'] \arrow[dr, "g_2"] & \\
(S_1, \tau_1) & & (S_2, \tau_2) & & (S_3, \tau_3)
\end{tikzcd}
$$

Figure 4: Composing spans given weak pullbacks.

$$
\begin{tikzcd}
X \arrow[d, "\alpha"'] \arrow[dr, "f"] & & Y \arrow[dl, "g"'] \arrow[d, "\beta"'] \arrow[dr, "h"] & & Z \arrow[dl, "k"'] \arrow[d, "\gamma"'] \\
\Pi X \arrow[dr] & U \arrow[d, "\tau"'] & \Pi Y \arrow[dl] \arrow[dr] & V \arrow[d, "\eta"'] & \Pi Z \arrow[dl] \\
& \Pi U & & \Pi V &
\end{tikzcd}
$$

Figure 5: Composing cospans.

$$
\begin{tikzcd}
& & Y \arrow[dl] \arrow[d, "\beta"'] \arrow[dr] & & \\
& g \arrow[dl] & \Pi Y \arrow[dl] \arrow[dr] & h \arrow[dr] & \\
U \arrow[d, "\tau"'] \arrow[dr] & \Pi g \arrow[dl] & & \Pi h \arrow[dr] & V \arrow[dl] \arrow[d, "\eta"] \\
\Pi U \arrow[dr] & i \arrow[dr] & & j \arrow[dl] & \Pi V \arrow[dl] \\
& \Pi i \arrow[dr] & W \arrow[d, dashed, "\rho"'] & \Pi j \arrow[dl] & \\
& & \Pi W & &
\end{tikzcd}
$$

Figure 6: Pushouts to compose cospans.

One way to construct a cospan from $\mathcal{X}$ to $\mathcal{Z}$ is to construct a pushout in the coalgebra category. In **Mes** we can construct a pushout for the arrows g and f to obtain the situation shown in Fig. 6. Here W is the object constructed as the tip of the pushout in **Mes**. In order to have a pushout in the category of coalgebras we need to put a coalgebra structure on W, i.e. we need to construct a morphism $\rho : W \to \Pi W$, shown dotted in the diagram. Consider the following calculation:

$g;\tau;\Pi i$		
$=$	$\beta;\Pi g;\Pi i$	g is a zig-zag,
$=$	$\beta;\Pi(g;i)$	functoriality,
$=$	$\beta;\Pi(h;j)$	pushout,
$=$	$\beta;\Pi h;\Pi j$	functoriality,
$=$	$h;\eta;\Pi j$	h is a zig-zag.

Thus, the outer square formed by Y, U, V and ΠW commutes and couniversality implies the existence of the morphism ρ from W to ΠW. It is a routine calculation that this gives a pushout in the category of coalgebras. This does not require any special properties of Π; it holds in the most general case, i.e. in **Mes**.

The categorical definitions can now be given.

Definition 6.1 *An* ***event bisimulation*** *on* $\mathcal{S} = (S, \Sigma, \tau)$ *is a surjection in the category of coalgebras of* Π *to some* $\mathcal{T}$.

We have already noted that such arrows are zig-zags and that a zig-zag induces an event bisimulation on its source. For the case of an event bisimulation between two different LMPs we have.

Definition 6.2 *An* ***event bisimulation*** *between* $\mathcal{S}$ *and* $\mathcal{S}'$ *is a cospan of surjections in the category of coalgebras to some object* $\mathcal{T}$.

This can be viewed as ordinary event bisimulation on $\mathcal{S} + \mathcal{S}'$.

Since cospans compose it is clear that viewed as a relation between LMPs event bisimulation (properly called probabilistic cocongruence) is an equivalence relation. We do not quite have a category of LMPs with cospans as the morphisms since associativity only holds up to isomorphism: we have a bicategory. One can define a nice epi-mono factorization on the category **Mes** that carries over to the category of coalgebras.

7 Almost sure bisimulation

A major virtue of the categorical presentation of the previous section is that one can modify the monad to deal with other, closely related situations. One does not then have to develop the theory again from scratch; one can just use the abstract machinery with a slightly different instantiation. In this section we give an example of this.

In one of the early papers on LMPs [BDEP97] the question of negligible sets was raised. Consider the LMPs $\mathcal{U} = ([0,1], \mathcal{B}, \tau)$, where $\mathcal{B}$ are the Borel sets and $\tau(x, A) = \lambda(A)$ where λ is Lebesgue measure; and $\mathcal{U}'$ with the same state space and σ-algebra but with Markov kernel given by $\tau'(x, A) = \tau(x, A)$ if x is irrational and $\tau'(x, A) = 0$ if x is rational. These two processes behave identically "almost always": except for a set of measure zero they are bisimilar. If we were to observe these systems the probability that we would detect the difference is zero. We would like to formalize this concept by introducing a notion of *almost sure* bisimulation.

There is an obvious gap in the discussion of the last paragraph. According to what measure should one say that the rationals have measure zero? One immediately thinks of Lebesgue measure in this example but what of other examples? Should the state space come equipped with additional structure - perhaps a measure - in order to define what the sets of measure zero are? We introduce such additional structure, but one does not need to introduce a measure just to define the sets of measure zero. Instead we introduce an axiomatically defined class of negligible sets. This can be smoothly incorporated into Giry's monad and then one has the notion of almost sure bisimulation "almost free" with the discussion of the previous section.

We introduce a new category $\mathbf{Mes}'$ refining the structure of **Mes**. The *purpose* of this refinement is to give a means of taking morphisms differing only on negligibly many points to be equal. In order to express this, one needs to add to each object a notion of negligible sets $\mathcal{N}$ and define:

$$f \sim g \quad := \quad \exists N \in \mathcal{N} : \{f \neq g\} \subseteq N \tag{1}$$

We write $A \subseteq\in \mathcal{N}$ to mean, there is an $N \in \mathcal{N}$ such that $A \subseteq N$. For instance, $f \sim g$ can be rewritten as $\{f \neq g\} \subseteq\in \mathcal{N}$.

7.1 The category Mes′

An *object* is a triple $(S, \Sigma, \mathcal{N})$ where:

- (S, Σ) is an object in **Mes**,
- $\mathcal{N} \subseteq \Sigma$ is a distinguished set of measurable subsets of S.

An *arrow* $(S, \Sigma, \mathcal{N}) \xrightarrow{f} (T, \Lambda, \mathcal{M})$ is:

- is an arrow in **Mes**,
- such that in addition:

$$\forall M \in \mathcal{M}\, \exists N \in \mathcal{N} : f^{-1}(M) \subseteq N \tag{2}$$

This additional condition is obviously stable by composition, so our data actually defines a category[3].

The additional component in objects, $\mathcal{N}$, is to be thought of intuitively as a set of *negligible sets*, and will have *some* closure properties to be specified later on. For the present we assume nothing about $\mathcal{N}$.

Note that one does not demand the stronger condition that $f^{-1}(M) \in \mathcal{N}$; it is enough for $f^{-1}(M)$ to be included in a negligible set. One could write equivalently $f^{-1}(\mathcal{M}) \subseteq\in \mathcal{N}$.

This condition will result in a serious restriction on **Mes** arrows. For instance, random variables with a finite number of values, say $x_1, \ldots, x_n$, from $(S, \Sigma, \mathcal{N})$ to $([0,1], \mathcal{B}, \mathcal{N}(\lambda))$, where λ is the Lebesgue measure, will all violate this condition unless for all i, $\{X = x_i\} \subseteq\in \mathcal{N}$. In particular, no discrete random variable from $([0,1], \mathcal{B}, \mathcal{N}(\lambda))$ to itself is in **Mes′**.

One now refines Giry's monad to take into account the collection of negligible sets. First the object part:
one chooses the additional component $\mathcal{N}(\mathbf{\Pi}(S, \Sigma, \mathcal{N}))$ to be generated (in the sense of the closure properties to be defined later) by the following particular subsets of measures:

$$\forall A \in \mathcal{N}. N_A \quad := \quad \{\nu \mid \nu(A) > 0\}. \tag{3}$$

Since $A \in \mathcal{N} \subseteq \Sigma$, $N_A \in \sigma(\mathbf{\Pi}(S, \Sigma))$ so our collections of negligible sets are adapted to the σ-algebra structure. The intent of this definition is that we want to neglect the measures that ascribe nonzero weight to the negligible sets.

Second the arrow part:
The definition of $\Pi(f)$ remains the same, but one has to check that condition (2) is satisfied by $\mathbf{\Pi}(f)$. Suppose then $f : (S, \Sigma, \mathcal{N}) \to (S', \Sigma', \mathcal{N}')$ and let A be in $\mathcal{N}'$:

$$\begin{aligned} \nu \in \mathbf{\Pi}(f)^{-1}(N_A) \quad &\Leftrightarrow \quad \mathbf{\Pi}(f)(\nu) \in N_A \\ &\Leftrightarrow \quad \mathbf{\Pi}(f)(\nu)(A) > 0 \\ &\Leftrightarrow \quad \nu(f^{-1}(A)) > 0 \end{aligned}$$

so that $\mathbf{\Pi}(f)^{-1}(N_A) = \{\nu \mid \nu(f^{-1}(A)) > 0\}$. Now by (2), for some $B \in \mathcal{N}$, $f^{-1}(A) \subseteq B$, hence $\mathbf{\Pi}(f)^{-1}(N_A) \subseteq N_B$, which is what we wanted to prove.

Verifying functoriality of Π is as before. We now have to lift the monad to the refined setting with the negligible sets.

We need to prove that the natural transformations also are arrows in the new sense.

[3] for $N'' \in \mathcal{N}''$, $(gf)^{-1}(N'') = f^{-1}(g^{-1}(N'')) \subseteq f^{-1}(N' \in \mathcal{N}') \subseteq N \in \mathcal{N}$.

Lemma 7.1 *For all $A \in \Sigma$:*
— (i) $\eta^{-1}(N_A) = A$;
— (ii) $\mu^{-1}(N_A) = N_{N_A}$.

Proof . (i) $a \in \eta^{-1}(N_A) \Leftrightarrow \delta_a(A) > 0 \Leftrightarrow a \in A$.
(ii) $P \in \mu^{-1}(N_A) \Leftrightarrow \mu(P)(A) > 0 \Leftrightarrow \int e_A dP > 0 \Leftrightarrow \int \mathbf{1}_{N_A} dP > 0$ since $N_A := \{e_A > 0\}$, the support of e_A (trivial integration lemma). And the last statement says precisely $P(N_A) > 0$, *i.e.*, $P \in N_{N_A}$. ∎

From this lemma one deduces readily that both η and μ satisfy (2).

We now have constructed a refined version $\mathbf{\Pi}'$ of $\mathbf{\Pi}$. Naturality requirements and other conditions are still valid, because they are commutative diagrams and these will continue to hold.

The idea of the definition of $\mathcal{N}(\mathbf{\Pi}(S, \Sigma, \mathcal{N}))$ is that a set of measures is considered to be negligible if all measures in it "see" a given negligible set. Why are we working with abstract negligibles, as opposed to those given by an actual measure? Well, one doesn't know how to define such a probability on $\mathbf{\Pi}(S, \Sigma)$. When is an $\mathcal{N}$ obtainable as an $\mathcal{N}_p$ for some p is not so clear. In any case, such a p would only intervene in the theory through sets of p measure 0, so any two equivalent p would define isomorphic objects. That said, we still work with the (S, Σ, p) notation when convenient.

One now defines the quotient category, $\mathbf{Mes}'$ obtained by identifying arrows that disagree only on a negligible set. Objects are as in $\mathbf{Mes}'$; arrows are $\sim$ equivalence classes.

Lemma 7.2 *The equivalence $\sim$ defined in (1) is stable by composition.*

Proof . $g \sim h \Rightarrow gf \sim hf$: one has $\{g \neq h\} \subseteq N'$ for some $N' \in \mathcal{N}'$, so:

$$\{gf \neq hf\} = f^{-1}(\{g \neq h\}) \subseteq f^{-1}(N')$$

and for some $N \in \mathcal{N}$, $f^{-1}(N') \subseteq N \in \mathcal{N}$, because f satisfies condition (2).
$g \sim h \Rightarrow fg \sim fh$: again one has $\{g \neq h\} \subseteq N'$ for some $N' \in \mathcal{N}'$, so:

$$\{fg \neq fh\} \subseteq \{g \neq h\} \subseteq N' \in \mathcal{N}'.$$

∎

Lemma 7.3 ($\mathbf{\Pi}$ respects $\sim$) *Let f, g be arrows from $\mathcal{S} = (S, \Sigma, \mathcal{N})$ to $\mathcal{T} = (T, \Lambda, \mathcal{M})$, then: $f \sim g \Rightarrow \mathbf{\Pi}(f) \sim \mathbf{\Pi}(g)$.*

Proof . Let f, g be as above, p be in $\mathbf{\Pi}(\mathcal{S})$. By definition of $\sim$, there exists an $N \in \mathcal{N}(\mathcal{S})$, such that $\{f \neq g\} \subseteq N$. Then one has:

$$\begin{aligned}
\mathbf{\Pi}(f)(p) \neq \mathbf{\Pi}(g)(p) &\Leftrightarrow \exists B \in \Lambda : p(f^{-1}(B)) \neq p(g^{-1}(B)) \\
&\Rightarrow \exists B \in \Lambda : p(f^{-1}(B) \triangle g^{-1}(B)) > 0 \\
&\Rightarrow p(N) > 0 \\
&\Leftrightarrow p \in \{e_N > 0\} \in \mathcal{N}(\mathbf{\Pi}(S))
\end{aligned}$$

Second step: obvious. [4]
Third step: $f^{-1}(B) \triangle g^{-1}(B) \subseteq \{f \neq g\} \subseteq N$.
So $\{\mathbf{\Pi}(f) \neq \mathbf{\Pi}(g)\} \subseteq\in \mathcal{N}(\mathbf{\Pi}(\mathbf{S}))$, or, in other words, $\mathbf{\Pi}(f) \sim \mathbf{\Pi}(g)$. ∎

[4]By additivity, $p(A \cup B) = p(A \triangle B) + p(A \cap B)$, so $p(A \triangle B) = 0$ implies $p(A \cup B) = p(A \cap B)$, which implies $p(A) = p(B)$ by sandwiching.

As an illustration, we may compute what it means for a kernel $k : (S, \Sigma, \mathcal{N}) \to \mathbf{\Pi}(S, \Sigma, \mathcal{N})$ to be a morphism in $\mathbf{Mes}'$. It has to satisfy two conditions:

1. usual stability wrt measurables: Σ stable (this is the usual measurability condition);

2. stability wrt negligibles: $\mathcal{N}$ (sub-) stable under $\{a\}_0$.

Proof . Only the second part needs a proof. By definition τ has to satisfy the condition that for all $A \in \mathcal{N}$, $\tau(\cdot, A)^{-1}(N_A) \subseteq\in \mathcal{N}$, which means $\{s \mid \tau(s, A) > 0\} = \{a\}_0(A) \subseteq\in \mathcal{N}$. ∎

When negligibles are given via some measure ν defined on (S, Σ), that is to say $\mathcal{N} = \mathcal{N}_\nu$, then the condition amounts to

$$\nu(A) = 0 \quad \Rightarrow \quad \nu(\{a\}_0(A)) = 0,$$

which says that “negligiby many points may jump to a negligible set”.

Lemma 7.4 (domination) *Suppose τ is a kernel, and ν dominates τ in the sense that for all s, $\tau(s, \cdot) \ll \nu$, then $\mathcal{N}_\nu$ is stable under τ, that is to say:*

$$\forall B \in \mathcal{N}_\nu : \{a\}_0(B) \in \mathcal{N}_\nu.$$

Proof . If $\nu(B) = 0$, then for all s, $\tau(s, B) = 0$, so that $\{a\}_0(B) = \varnothing \in \mathcal{N}_\nu$. ∎

The second condition of “stability with respect to negligibles” is seen to be a weaker form of “domination”.

7.2 Abstract Completion

In measure theory there is a standard notion of completion of a σ-algebra with respect to a measure. The point is that many σ-algebras contain non measurable subsets of sets of measure zero. This is very annoying; one would like to say that a subset of a negligible set is also negligible, but one cannot, in general, for silly reasons. The standard example is the Borel algebra and the Lebesgue measure. There are subsets of sets of Lebesgue measure zero that are not Borel measurable. The completion process adds in all these sets and yields a bigger σ-algebra[5]. We can describe this completion process in terms of our abstract notion of negligible set: it is here that we need to impose some axioms on $\mathcal{N}$.

Definition 7.5 (σ-ideal) *A σ-ideal $\mathcal{N}$ on S is a set of subsets of S which is 1) downward closed (i.e., $\mathcal{N}\downarrow = \mathcal{N}$), and 2) closed under countable unions.*

Definition 7.6 (abstract completion) *Given a measurable space (S, Σ), and a σ-ideal $\mathcal{N}$, one defines the completion of Σ by $\mathcal{N}$:*

$$\Sigma_{\mathcal{N}} \quad := \quad \{K \mid \exists A, B \in \Sigma : A \subseteq K \subseteq B \,\&\, B \setminus A \in \mathcal{N}\}$$

Lemma 7.7 *$\Sigma_{\mathcal{N}}$ is a σ-algebra that contains Σ, and $(\Sigma \cap \mathcal{N})\downarrow$. Let p be a subprobability defined on (S, Σ), if $p(\mathcal{N}) = \{0\}$, p has a unique extension to $\Sigma_{\mathcal{N}}$.*

[5]Actually “completion” is a terrible word; there is no reasonable sense in which the resulting σ-algebra is complete except that the completion process when applied again adds nothing new.

Proof . $\Sigma \subseteq \Sigma_{\mathcal{N}}$: any $A \in \Sigma$ verifies $A \subseteq A \subseteq A$ and $A \setminus A = \varnothing \in \mathcal{N}$, since $\mathcal{N}$ is downward closed. $(\Sigma \cap \mathcal{N})\downarrow \subseteq \Sigma_{\mathcal{N}}$: any $N \in (\Sigma \cap \mathcal{N})\downarrow$ verifies $\varnothing \subseteq N \subseteq N'$ for some $N' \in \Sigma \cap \mathcal{N}$, and $N' \setminus \varnothing = N' \in \mathcal{N}$.
$\Sigma_{\mathcal{N}}$ closed under complements: suppose $A \subseteq K \subseteq B$, with $A, B \in \Sigma$ and $B \setminus A \in \mathcal{N}$, then $\bar{B} \subseteq \bar{K} \subseteq \bar{A}$, and $\bar{A} \setminus \bar{B} = B \setminus A \in \mathcal{N}$.
$\Sigma_{\mathcal{N}}$ closed under countable unions: suppose $A_i \subseteq K_i \subseteq B_i$, with $A_i, B_i \in \Sigma$ and $B_i \setminus A_i \in \mathcal{N}$, then $\cup_I A_i \subseteq \cup_I K_i \subseteq \cup_I B_i$, and

$$\cup_I B_i \setminus \cup_I A_i \subseteq \cup_I (B_i \setminus A_i) \in \mathcal{N}$$

The rhs is in $\mathcal{N}$ because $\mathcal{N}$ is closed under countable unions, and therefore the left hand side also is in $\mathcal{N}$, since $\mathcal{N}$ is downward closed[6] . ∎

This completion can also be done when $\mathcal{N}$ is not downward closed, by asking that $B \setminus A \subseteq\in \mathcal{N}$. Then $\Sigma_{\mathcal{N}\downarrow} = \Sigma_{\mathcal{N}}$, so that the downward closure is happening during the completion.

8 Conclusions

The main point of this paper is to argue that one should work with probabilistic cocongruence rather than with probabilistic bisimulation. The theory works smoothly for LMPs on general measure spaces: one need not work with analytic spaces. The proof of the logical characterization of cocongruence is much simpler and more general than the proof of the logical characterization of bisimulation. One only needs to invoke the theory of analytic spaces to show that cocongruence and bisimulation coincide: which they indeed do on analytic spaces.

Indeed, it seems to us that bisimulation defined categorically is a historical anomaly. In the discrete case bisimulation and cocongruence coincide and one can argue that it makes no difference. However transition systems are coalgebras and cospans should fit the theory better than spans; as indeed is our experience in the probabilistic case. It would be interesting to reformulate the general theory in terms of cocongruences.

In the case of probabilistic systems it has been argued that one should use metrics rather than equivalence relations [DGJP99, DGJP04]. In the present context a pressing problem is to understand the metric analogue of the theory of probabilistic cocongruence.

Acknowledgements

We have benefitted from discussions with Samson Abramsky, Martin Escardo, Gordon Plotkin and each other. Vincent Danos enjoyed the company of Elham Kashefi during the course of this research. Prakash Panangaden acknowledges the crucial role played by bad weather during his holidays which allowed him to complete this paper. We all like to thank each other for being such a fun bunch.

References

[BDEP97] R. Blute, J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labelled Markov processes. In *Proceedings of the Twelfth IEEE Symposium On Logic In Computer Science, Warsaw, Poland.*, 1997.

[6] The two hand sides are not equal in general: $B_i = S$, then the lhs is $B \setminus \cup_I A_i = \cap_I \bar{A}_i$ and the rhs is $\cup_I \bar{A}_i$; *e.g.* , $I = \{0, 1\}$, $B_0 = B_1 = \{a, b\}$, $A_0 = \{a\}$, $A_1 = \{b\}$, lhs is $\varnothing$, while rhs is $\{a, b\}$.

[DD03] Vincent Danos and Josée Desharnais. Labeled Markov Processes: Stronger and faster approximations. In *Proceedings of the 18th Symposium on Logic in Computer Science*, Ottawa, 2003. IEEE.

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

[DEP02] J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labeled Markov processes. *Information and Computation*, 179(2):163–193, Dec 2002.

[DGJP99] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Metrics for labeled Markov systems. 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.

[DGJP02] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. The metric analogue of weak bisimulation for labelled Markov processes. In *Proceedings of the Seventeenth Annual IEEE Symposium On Logic In Computer Science*, pages 413–422, July 2002.

[DGJP03] J. Desharnais, V. Gupta, R. Jagadeesan, and P. Panangaden. Approximating labeled Markov processes. *Information and Computation*, 184(1):160–200, July 2003.

[DGJP04] Josée Desharnais, Vineet Gupta, Radhakrishnan Jagadeesan, and Prakash Panangaden. A metric for labelled Markov processes. *Theoretical Computer Science*, 318(3):323–354, June 2004.

[Dob03] E.-E. Doberkat. Semi-pullbacks and bisimulations in categories of stochastic relations. In J. C. M. Baeten, J. K. Lenstra, J. Parrow, and G. J. Woeinger, editors, *Proceedings of the 27th International Colloquium On Automata Languages And Programming, ICALP'03*, number 2719 in Lecture Notes In Computer Science, pages 996–1007. Springer-Verlag, July 2003.

[DP01] J. Desharnais and P. Panangaden. Continuous stochastic logic characterizes bisimulation for continuous-time Markov processes. Available from `http://rl.cs.mcgill.ca/~prakash/pubs.html`, 2001.

[dVR97] E. de Vink and J. J. M. M. Rutten. Bisimulation for probabilistic transition systems: A coalgebraic approach. In *Proceedings of the 24th International Colloquium On Automata Languages And Programming*, 1997.

[dVR99] E. de Vink and J. J. M. M. Rutten. Bisimulation for probabilistic transition systems: A coalgebraic approach. *Theoretical Computer Science*, 221(1/2):271–293, June 1999.

[Eda99] Abbas Edalat. Semi-pullbacks and bisimulation in categories of Markov processes. *Mathematical Structures in Computer Science*, 9(5):523–543, 1999.

[Gir81] M. Giry. A categorical approach to probability theory. In B. Banaschewski, editor, *Categorical Aspects of Topology and Analysis*, number 915 in Lecture Notes In Mathematics, pages 68–85. Springer-Verlag, 1981.

[KS60] J. G. Kemeny and J. L. Snell. *Finite Markov Chains.* Van Nostrand, 1960.

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

[Pan99] Prakash Panangaden. The category of markov processes. *ENTCS*, 22:17 pages, 1999. `http://www.elsevier.nl/locate/entcs/volume22.html`.

[Put94] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley, 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.

[Wil91] David Williams. *Probability with Martingales.* CUP, Cambridge, 1991.