Under consideration for publication in Math. Struct. in Comp. Science

Semi-pullbacks and Bisimulation in Categories of Markov Processes

Abbas Edalat

Department of Computing, Imperial College,
London SW7 2BZ

Received

We show that the category whose objects are famidclies of Markov processes on Polish spaces, with a given transition kernel, and whose morphisms are transition probability preserving, surjective continuous maps has semi-pullbacks, i.e. for any pair of morphisms $f_i : S_i \to S$ $(i = 1, 2)$, there exists an object V and morphisms $\pi_i : V \to S_i$ $(i = 1, 2)$ such that $f_1 \circ \pi_1 = f_2 \circ \pi_2$. This property holds for various full subcategories, including that of families of Markov processes on locally compact second countable spaces and also in the larger category where the objects are families of Markov processes on analytic spaces and morphisms are transition probability preserving surjective Borel maps. It also follows that the category of probability measures on Polish spaces and measure-preserving continuous surjective maps has semi-pullbacks. A main consequence of our result is that probabilistic bisimulation for labelled Markov processes on all these categories, defined in terms of spans of morphisms, is an equivalence relation.

Contents

1. Introduction

We say that a category has *semi-pullbacks*, or has the *Ore* property, if for any pair of morphisms $f_1 : S_1 \to S$ and $f_2 : S_2 \to S$ there exists an object V and morphisms

$\pi_i : V \to S_i$ $(i = 1, 2)$ such that $f_1 \circ \pi_1 = f_2 \circ \pi_2$ as in the following diagram.

$$\begin{tikzcd} & V \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\ S_1 \arrow[dr, "f_1"'] & & S_2 \arrow[dl, "f_2"] \\ & S & \end{tikzcd}$$

Consider the following category. An object of the category is given by a tuple $(S, \mathcal{S}, k)$, where S is a Polish space (a topologically complete separable space), $\mathcal{S}$ is the Borel σ-algebra on S and the map $k : S \times \mathcal{S} \to [0, 1])$ is a Markov kernel, i.e. a transition probability function, on S. A morphism $f : (S, \mathcal{S}, k) \to (S', \mathcal{S}', k')$ is a continuous, surjective map $f : S \to S'$ which is transition probability preserving i.e. for all $s \in S$ and all $B \in \mathcal{S}'$, we have $k(s, f^{-1}(B)) = k'(f(s), B)$. The question is if this category has semi-pullbacks.

Clearly this can be regarded as a question in pure probability theory. In fact, before stating the main result of the paper and explaining its application in computer science, we give an example of its relevance in probability theory. Consider two stationary Markov processes in $\mathbb{R}^2$ given by transition kernels $k_1(.,.)$ and $k_2(.,.)$ which have the same vertical distribution, i.e. for all $x, y_1, y_2 \in \mathbb{R}$ and all Borel subsets $A \subseteq \mathbb{R}$ we have: $k_1((x, y_1), \pi_1^{-1}(A)) = k_2((x, y_2), \pi_1^{-1}(A))$ where π_1 is the projection of $\mathbb{R}^2$ to the first coordinate. The existence of semi-pullbacks is equivalent here to the existence of a stationary Markov kernel $h(.,.)$ on $\mathbb{R}^3$ which simulates both of the two original processes in the sense that, for all $x, y, z \in \mathbb{R}$ and all Borel sets $B, C \subseteq \mathbb{R}^2$, we have: $h((x, y, z), \pi_{12}^{-1}(B)) = k_1((x, y), B)$ and $h((x, y, z), \pi_{13}^{-1}(C)) = k_2((x, z), C)$, where $\pi_{ij} : \mathbb{R}^3 \to \mathbb{R}^2$ is the projection to the plane of the ith and jth coordinates.

In this paper, we show that the answer to the above question is positive. More specifically, we give a canonical construction for semi-pullbacks in the category whose objects are families of Markov processes, with given transition kernels, on Polish spaces and whose morphisms are transition probability preserving surjective continuous maps. One immediate consequence is that the category of probability measures on Polish spaces with measure-preserving continuous maps has semi-pullbacks.

Our construction gives semi-pullbacks for various full subcategories, including that of Markov processes on locally compact second countable spaces and also in the larger category where the objects are Markov processes on analytic spaces (i.e. continuous images of Polish spaces) and morphisms are transition probability preserving surjective Borel maps. It also applies to the corresponding categories of ultrametric spaces. Finally, our result also holds in the larger categories with Markov processes which are given by subprobability distributions, i.e. the total probability of transition from a state can be strictly less than one.

We now explain the relevance of our result in computer science. The consequences of our mathematical result in the theory of probabilistic bisimulation has been investigated in (Blute et al., 1997; Desharnais et al., 1998). We will briefly review this here. Following the work of Joyal, Nielsen and Winskel (Joyal et al., 1996) on the notion of bisimulation using open maps, define two objects A and B in a category to be *bisimular* if there exists an object C and morphisms $f : C \to A$ and $g : C \to B$, i.e.,

$$\begin{tikzcd} & C \arrow[dl, "f"'] \arrow[dr, "g"] & \\ A & & B \end{tikzcd}$$

The existence of semi-pullbacks in a category implies that bisimulation is a transitive and hence an equivalence relation. The category of partial, labelled Markov processes on Polish spaces is defined as follows. Given a fixed set of labels $\mathcal{L}$, a partial, labelled Markov process is a tuple $(S, \mathcal{S}, \{k_l | l \in \mathcal{L}\})$ where S is a Polish space with the σ-algebra of Borel subsets $\mathcal{S}$ and for each $l \in \mathcal{L}$, $k_l : S \times \mathcal{S} \to [0, 1]$ is a transition sub-probability function. A morphism between two partial, labelled Markov processes $(S, \mathcal{S}, \{k_l | l \in \mathcal{L}\})$ and $(S', \mathcal{S}', \{k_l | l \in \mathcal{L}\})$ is a continuous, surjective map $f : S \to S'$ which is transition probability preserving for each $l \in \mathcal{L}$. This category generalizes that of Larsen and Skou (Larsen and Skou, 1991) for the probabilistic discrete processes. One can similarly define the category of partial, labelled Markov processes on analytic spaces with Borel, surjective and transition probability preserving morphisms. An immediate corollary of the existence of semi-pullbacks is that bisimulation in these categories is an equivalence relation; this allows the theory of labelled Markov processes to be generalized to continuous state spaces. Furthermore, the existence of semi-pullbacks in the category of Markov processes on analytic spaces has been used to show that bisimulation in this setting can be characterized by a weak, negation-free logic (Desharnais et al., 1998).

In order to properly state and deduce this result, we go beyond these categories, which are similar to the category considered by Giry (Giry, 1981), to a more foundational framework in probability theory and stochastic processes. We first introduce, in Section 2, the notion of a universally measurable regular conditional probability distribution and obtain a necessary and sufficient condition for it to be preserved under a surjective measurable map. Then, we note the existence of regular conditional probability distributions on analytic spaces and we verify a general result on the measurability of integrals of regular conditional probability distributions. In Section 3, we properly define our category and show that Markov processes are invariant under surjective, measurable and transition probability preserving maps. In Section 4, we start to construct the semi-pullback of $f_1 : (S_1, \mathcal{S}_1, k_1) \to (S, \mathcal{S}, k)$ and $f_2 : (S_2, \mathcal{S}_2, k_2) \to (S, \mathcal{S}, k)$. This semi-pullback will be a Markov kernel on the pullback $V = \{(s_1, s_2) \in S_1 \times S_2 | f_1(s_1) = f_2(s_2)\}$ of $f_1 : S_1 \to S$ and $f_2 : S_2 \to S$ in the category of Polish spaces and continuous maps, i.e. the set V with the subspace topology of $S_1 \times S_2$. We first construct for each $(s_1, s_2) \in V$ a probability measure on the product space $S_1 \times S_2$. Then in Section 5, we show that this indeed gives a probability measure on V from which we obtain the required Markov kernel.

For an element a of a set X, sets Ω, Y, subsets $A \subseteq X$, $B \subseteq Y$ and functions $x : \Omega \to X$,

$f : X \to Y$, we sometimes write fa instead of $f(a)$, fA instead of $f(A)$, $f^{-1}B$ instead of $f^{-1}(B)$ and fx for the composition $f \circ x$. The identity map on X is denoted by Id_X, the characteristic function of a A by 1_A and the point measure at a by δ_a. The extended real line is denoted by $\overline{\mathbb{R}}$.

2. Conditional probability distributions

In order to properly define our category and deduce our main theorem, we will need some preliminary work. We first fix our notations by recalling the basic definitions of random variables and conditional probability distributions we need in this paper. See foundational texts on probability theory (Breiman, 1968; Hoffmann-Jørgensen, 1970; Arveson, 1976; Dudley, 1989; Hoffmann-Jørgensen, 1994). We will use the terminology of the comprehensive text (Hoffmann-Jørgensen, 1994) which is the most recent treatise on probability theory.

A measurable space $(S, \mathcal{S})$ is given by a set S and a σ-algebra $\mathcal{S}$ of subsets of S. Given two measurable spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$, a map $f : (S, \mathcal{S}) \to (T, \mathcal{T})$ is *measurable* if $f^{-1}(A) \in \mathcal{S}$ for all $A \in \mathcal{T}$; we also call f a $\mathcal{S}$*-measurable map* from S to T, in particular when there are other σ-algebras on S but only a natural σ-algebra $\mathcal{T}$ on T. A probability space $(S, \mathcal{S}, \mu)$ is given by a probability measure μ on a measurable space $(S, \mathcal{S})$. A measurable map $f : (S, \mathcal{S}, \mu) \to (T, \mathcal{T}, \nu)$ of probability spaces is called *measure-preserving* if $\nu(A) = \mu(f^{-1}(A))$ for all $A \in \mathcal{T}$. Let $(\Omega, \mathcal{F}, P)$ be a probability space. Let $(\Omega, \mathcal{F}, P)$ be a probability space. A *random variable* on Ω with values on a measurable space $(S, \mathcal{S})$ is a measurable map $x : \Omega \to S$. For $A \in \mathcal{S}$, the probability that x takes its value in A is given by $P(x \in A) = P(x^{-1}(A)) = P(\{\omega \in \Omega | x(\omega) \in A\})$; we often write $P_x(A) = P(x \in A)$. Hence, $P_x(.)$ is a probability measure on S which is called the *distribution* of x, and $(S, \mathcal{S}, P_x)$ is itself a probability space. Since the two random variables $x : \Omega \to S$ and the identity map $1 : S \to S$ have the same distribution, one can in practice forget about Ω and consider $1 : S \to S$ as the random variable. In this paper, however, we will often define a random variable with respect to some probability space $(\Omega, \mathcal{F}, P)$. Whenever there are several random variables we usually assume that they are defined using the same probability space; this can always be done by taking Ω to be large enough.

If μ is any probability measure on $(S, \mathcal{S})$ then a subset $A \subseteq S$ is *measurable w.r.t.* μ *or* μ*-measurable* if there exist $B, C \in \mathcal{S}$ with $B \subseteq A \subseteq C$ such that $\mu(C \setminus B) = 0$. The μ-measurable subsets of S form a σ-algebra denoted by $\mathcal{M}(\mu)$. We say $A \subseteq S$ is *universally measurable* if A is measurable w.r.t. any probability measure on $(S, \mathcal{S})$. The set of all universally measurable subsets of S forms a σ-algebra denoted by $\mathcal{U}(\mathcal{S})$; in other words $\mathcal{U}(\mathcal{S}) = \bigcap\{\mathcal{M}(\mu) | \mu \text{ a probability measure on } (S, \mathcal{S})\}$. Given measurable spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$, a map $f : S \to T$ is called *universally measurable* if $f : (S, \mathcal{U}(\mathcal{S})) \to (T, \mathcal{T})$ is measurable; it then follows that $f : (S, \mathcal{U}(S)) \to (T, \mathcal{U}(\mathcal{T}))$ is also measurable and the composition of universally measurable maps is universally measurable (Hoffmann-Jørgensen, 1970, page 12).

If f is a measurable map from $(S, \mathcal{S})$ to $(T, \mathcal{T})$, then the composition $fx : \Omega \to T$ is a random variable with values on $(T, \mathcal{T})$. We denote by $\sigma(f; \mathcal{T})$ the smallest σ-subalgebra of $\mathcal{S}$ which makes f measurable, i.e., $\sigma(f;\mathcal{T}) = \{f^{-1}A | A \in \mathcal{T}\}$. If $(S_i, \mathcal{S}_i)$ $(1 \leq i \leq n)$ are measurable spaces then $\prod_{i=1}^n S_i$ with the product σ-algebra $\otimes_{i=1}^n \mathcal{S}_i$ is a measurable space. For any set of random variables $x_i : \Omega \to S_i$ we have a random variable $x = (x_1, \cdots, x_n) : \Omega \to \prod_{i=1}^n S_i$, also called a *random vector*, defined by $x(\omega) = (x_1(\omega), \cdots, x_n(\omega))$. The distribution of x is called the *joint distribution* of x_i $(1 \leq i \leq n)$.

If $\mathcal{G} \subseteq \mathcal{F}$ is a σ-subalgebra, then the *conditional probability of x given $\mathcal{G}$* is, for each $A \in \mathcal{S}$, the P-almost unique random variable $P(x \in A|\mathcal{G}) : \Omega \to \overline{\mathbb{R}}$ which is integrable with respect to P and, for each $G \in \mathcal{G}$, satisfies:

$$\int_G P(x \in A|\mathcal{G})\, dP = P(G \cap x^{-1}(A)).$$

Given measurable spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$ with random variables $x : \Omega \to S$ and $y : \Omega \to T$, the *conditional distribution of x given y* is, for all $A \in \mathcal{S}$, the P_y-almost surely unique measurable function

$$t \mapsto P(x \in A|y = t) : T \to \overline{\mathbb{R}}$$

such that $\int_B P(x \in A|y = t) P_y(dt) = P(x \in A, y \in B)$. It follows that $P(x \in A|y = .) \circ y = P(x \in A|\sigma(y;\mathcal{T}))$. One often writes $P(x \in A|y = t)$ simply as $P(A|t)$. Let $\mathcal{D}$ be a σ-subalgebra of $\mathcal{M}(P_y)$. A *$\mathcal{D}$-measurable regular conditional probability distribution* of x given y is a map

$$P_x^y(.|.) : \mathcal{S} \times T \to [0, 1]$$

which satisfies:

(i) $P_x^y(.|t)$ is a probability measure on S for all $t \in T$.
(ii) $P_x^y(A|.)$ is $\mathcal{D}$ measurable for all $A \in \mathcal{S}$.
(iii)$P(x \in A, y \in B) = \int_B P_x^y(A|t) P_y(dt)$ for all $A \in \mathcal{S}$ and $B \in \mathcal{T}$.

A $\mathcal{T}$-measurable regular conditional probability distribution is simply called a *r.c.p.d.* and a $\mathcal{M}(P_y)$-measurable regular conditional probability distribution an *almost surely regular conditional probability distribution (a.s.r.c.p.d.)*. These notions are consistent with the definitions in (Hoffmann-Jørgensen, 1994). Moreover, we call a $\mathcal{U}(\mathcal{T})$-measurable regular conditional probability distribution a *universally measurable regular conditional probability distribution (u.m.r.c.p.d.)*. Note that if $Q(.|.) : \mathcal{S} \times T \to [0, 1]$ also satisfies the above three conditions, then $Q(A|t) = P_x^y(A|t)$ for P_y-almost all $t \in T$. Each such function is called a *version* of $P_x^y(.|.)$. Regular conditional probability distributions exist for a basic class of metrizable and separable spaces S above which we will now describe.

2.1. *Analytic spaces*

Recall that, for any topological space, the class of Borel subsets is the smallest σ-algebra which contains all open subsets. A map $f : X \to Y$ of topological spaces is a *Borel map* if $f^{-1}B$ is Borel for each Borel subset $B \subseteq Y$.

An analytic space (Dudley, 1989, page 388) is the continuous image (or, equivalently, the image under a Borel map) of a Polish space (or, equivalently, a Borel subset of a Polish space) into another Polish space and is equipped with the subspace topology of the latter space. When the underlying topological space is clear, one simply refers to an analytic space as an analytic set. There are analytic sets which are not Borel. Analytic spaces have many useful closure properties; see (Hoffmann-Jørgensen, 1970; Arveson, 1976; Dudley, 1989) for a comprehensive exposition.. Any Borel subset of a Polish space is analytic, and an analytic set whose complement is also analytic is Borel. Analytic sets are closed under taking countable unions or intersections (Dudley, 1989, page 393). It follows that the Borel subsets of an analytic space A are analytic, since such a subset is of the form $A \cap B$ where B is a Borel subset of the underlying Polish space. It follows immediately from the definition that the image of an analytic space under a Borel map into a Polish space is analytic; see (Arveson, 1976, page 71). Furthermore,

Corollary 2.1. The image of an analytic space by a Borel map into an analytic space is analytic.

We need the following version of the existence theorem for cross sections.

Lemma 2.1. Let $f : S \to T$ be a Borel surjective map of analytic spaces. Then, there exists a universally measurable map $g : T \to S$ with $f \circ g = \mathrm{Id}_T$ such that $g(T) \in \mathcal{U}(\mathcal{S})$.

Proof. By (Arveson, 1976, Theorem 3.4.3), there exists a universally measurable map $g : T \to S$ with $f \circ g = \mathrm{Id}_T$. We will show that $g(T)$ is universally measurable. Consider the map $F : x \mapsto (x, g(f(x))) : S \to S \times S$ which is easily seen to be universally measurable since the composition of two universally measurable maps is universally measurable. Note that the set of fixed points of the map $g \circ f : S \to S$ is $g(T)$ and that the diagonal $D = \{(x,x) | x \in S\}$ is a closed subset of $S \times S$. Hence, the inverse image $F^{-1}(D) = \{x \in S | x = g(f(x))\} = g(T)$ is universally measurable. □

Analytic sets have also important measure-theoretic properties. Recall that a *Borel measure* on a topological space is a measure defined on the σ-algebra of Borel subsets of the space. We then have the following:

Theorem 2.1. (Dudley, 1989, page 391) An analytic subset of a Polish space is measurable with respect to any finite Borel measure on the Polish space.

In other words, an analytic subset is universally measurable. It follows that any Borel measure μ on a Polish space P induces a Borel measure μ_A on any analytic subspace $A \subseteq P$. Indeed any Borel subset of A is of the form $B \cap A$ where A is a Borel subset of P. Since $B \cap A$ is analytic, it is μ-measurable and we can therefore put $\mu_A(B \cap A) = \mu(B \cap A)$.

Corollary 2.2. An analytic subset $A_1 \subseteq A_2$ of an analytic space A_2 is measurable with respect to any finite Borel measure μ on A_2.

Proof. Let P be the underlying Polish space and define the finite Borel measure μ_P on P by $\mu_P(B) = \mu(B \cap A_2)$ for any Borel subset $B \subseteq P$. By Theorem 2.1, there exist Borel subsets $B_1, B_2 \subseteq P$ with $B_1 \subseteq A_1 \subseteq B_2$ such that $\mu_P(B_2 \setminus B_1) = 0$. It follows that $B_1 \cap A_2 \subseteq A_1 \subseteq B_2 \cap A_2$ and

$$\mu((B_2 \cap A_2) \setminus (B_1 \cap A_2)) = \mu((B_2 \setminus B_1) \cap A_2) = \mu_P(B_2 \setminus B_1) = 0$$

as required.

□

Proposition 2.1. Let $f : A \to B$ be a surjective Borel map of analytical spaces A and

B. Suppose μ is a Borel measure on B. Then there exists a Borel measure ν on A with $\mu = \nu \circ f^{-1}$.

Proof. By Lemma 2.1, there exists a universally measurable map $g : B \to A$ with $f \circ g = \mathrm{Id}_B$ such that $g(B)$ is universally measurable. Put $\nu = \mu \circ g^{-1}$. $\square$

Finally, we state a main result on the existence of regular conditional probability distributions. The following property of analytic spaces is folklore in the probability theory community; its proof can be deduced from a general result by (Pachl, 1978).

Theorem 2.2. Suppose S is an analytic space, and $x : \Omega \to S$ and $y : \Omega \to T$ are random variables. Then a regular conditional probability distribution $P_x^y(.|.)$ of x given y exists.

Proof. See the appendix. $\square$

2.2. *Measurability and regular conditional probability distributions*

We need the following general results on the integrals of regular conditional probability distributions and their measurability later in this paper. We first provide the proofs of the following routine lemmas. In this paper, the unit interval $[0, 1]$ is always equipped with its Borel σ-algebra. Let $(S, \mathcal{S})$ and $(T, \mathcal{T})$ denote measurable spaces for the rest of this section.

Lemma 2.2. If $f : S \to T$ is measurable then for each $n \geq 0$ the map

$$f_n : (s_0, \ldots, s_n) \mapsto (fs_0, \ldots, fs_n) : (S^{n+1}, \otimes_{i=0}^n \mathcal{U}(\mathcal{S})) \to (T^{n+1}, \otimes_{i=0}^n \mathcal{U}(\mathcal{T}))$$

is measurable.

Proof. The case $n = 0$ follows since any $\mathcal{S}$-measurable map is trivially $\mathcal{U}(\mathcal{S})$-measurable. Suppose $n > 0$. For any rectangle $A_0 \times \ldots \times A_n$ with $A_i \in \mathcal{U}(\mathcal{T})$ $(i = 0, \ldots, n)$ we have $f_n^{-1}(A_0 \times \ldots \times A_n) = (f^{-1}A_0) \times \ldots \times (f^{-1}A_n) \in \otimes_{i=0}^n \mathcal{U}(\mathcal{S})$, by the result for $n = 0$. Since the σ-algebra $\otimes_{i=0}^n \mathcal{U}(\mathcal{T})$ is generated by such rectangles, the result follows. $\square$

Lemma 2.3. Let $g : S \to [0, 1]$ and $f : S \to T$ be measurable maps with f a surjection. Then, g is $\sigma(f; \mathcal{T})$-measurable iff there exists a measurable map $h : T \to [0, 1]$ such that $g = h \circ f$.

Proof. Suppose g is $\sigma(f; \mathcal{T})$-measurable. If $f(s_1) = f(s_2)$ for some $s_1, s_2 \in S$, then $s_1 \in g^{-1}\{g(s_1)\} = f^{-1}(B)$ for some measurable subset $B \subseteq T$; hence, $s_2 \in f^{-1}(B)$ and it follows that $g(s_1) = g(s_2)$. Therefore, we define $h : T \to [0, 1]$ by $h(t) = g(s)$ for any $s \in S$ with $f(s) = t$. If $A \subseteq [0, 1]$ is a Borel subset, then there exists a measurable subset $C \subseteq T$ such that $g^{-1}(A) = f^{-1}(C)$. It follows that $h^{-1}(A) = f(g^{-1}(A)) = f(f^{-1}(C)) = C$ and, hence, h is measurable. Conversely, if there exists a measurable map $h : T \to [0, 1]$ such that $g = h \circ f$, then, for any Borel subset $A \subseteq [0, 1]$, we have: $g^{-1}(A) = f^{-1}(h^{-1}(A)) \in \sigma(f; \mathcal{T})$. $\square$

Theorem 2.3. Suppose x and y are random variables with values on $(S, \mathcal{S})$ and $f : S \to T$ is a measurable surjective function from S to the measurable space $(T, \mathcal{T})$. Assume a u.m.r.c.p.d. (respectively, r.c.p.d.) $P_x^y(.|.) : \mathcal{S} \times S \to [0, 1]$ exists. Then the following conditions are equivalent.

(i) $\forall A \in \mathcal{T}. P_x^y(f^{-1}A|.) : S \to [0,1]$ is $\sigma(f;\mathcal{U}(\mathcal{T}))$-measurable (respectively, $\sigma(f;\mathcal{T})$-measurable).

(ii) A u.m.r.c.p.d. (respectively, r.c.p.d.) $P_{fx}^{fy}(.|.) : \mathcal{T} \times T \to [0,1]$ exists and $\forall A \in \mathcal{T}. P_{fx}^{fy}(A|f(.)) = P_x^y(f^{-1}A|.)$

Proof. The proofs for u.m.r.c.p.d. and r.c.p.d. are similar. We present the proof for the u.m.r.c.p.d.

(i) $\Rightarrow$ (ii). Let $Q(.|.) : \mathcal{T} \times T \to [0,1]$ be given by $Q(A|t) = P_x^y(f^{-1}A|s)$ where $f(s) = t$. Our assumption implies, by Lemma 2.3 (with $\mathcal{T}$ in the lemma replaced with $\mathcal{U}(\mathcal{T})$), that $Q(.|.)$ is well-defined and for each $A \in \mathcal{T}$, the map $Q(A|.) : T \to [0,1]$ is $\mathcal{U}(\mathcal{T})$-measurable. We show that it satisfies the two other conditions for $P_{fx}^{fy}(.|.)$. Clearly, for each $t \in T$, $Q(.|t)$ is a probability measure on T. Also, for any $B \in \mathcal{T}$, we have:

$$\begin{array}{rcl} \int_B Q(A|t)P_{fy}(dt) & = & \int_{f^{-1}B} P_x^y(f^{-1}A|s)P_y(ds) \\ & = & P(x \in f^{-1}A, y \in f^{-1}B) = P(fx \in A, fy \in B). \end{array}$$

Hence, $Q(.|.)$ is a version of $P_{fx}^{fy}(.|.)$ and the result follows.

(ii) $\Rightarrow$ (i). By Lemma 2.2, $f : (S,\mathcal{U}(S)) \to (T,\mathcal{U}(\mathcal{T}))$ is measurable. Let C be a Borel subset of $[0,1]$. Then, we have:

$$\{s|P_x^y(f^{-1}A|s) \in C\} = \{s|P_{fx}^{fy}(A|f(s)) \in C\} = f^{-1}\{t|P_{fx}^{fy}(A|t) \in C\} \in \sigma(f;\mathcal{U}(\mathcal{T})).$$

$\square$

Theorem 2.3 has a straightforward extension to the case of conditioning on a random vector as follows. Given $f : S \to T$ and $n \geq 0$, we let $f_n : S^n \to T^n$ be defined by $f_n(s_1, \ldots, s_n) = (fs_1, \ldots, fs_n)$. By Lemma 2.2, $f_n : (S^{n+1}, \otimes_{i=0}^n \mathcal{U}(S)) \to (T^{n+1}, \otimes_{i=0}^n \mathcal{U}(\mathcal{T}))$ is measurable.

Theorem 2.4. Suppose x and y_i $(0 \leq i \leq n)$ are random variables with values on $(S,\mathcal{S})$ and $f : S \to T$ is a measurable surjective function from S to the measurable space $(T,\mathcal{T})$. Assume a $\mathcal{U}(\otimes_{i=0}^n \mathcal{S})$-measurable r.c.p.d. (respectively, a r.c.p.d.) $P_x^{(y_0,\ldots,y_n)}(.|.) : \mathcal{S} \times S^{n+1} \to [0,1]$ exists. Then the following conditions are equivalent.

(i) $\forall A \in \mathcal{T}. P_x^{(y_0,\ldots,y_n)}(f^{-1}A|.) : S^{n+1} \to [0,1]$ is $\sigma(f_{n+1}; \otimes_{i=0}^n \mathcal{U}(\mathcal{T}))$-measurable (respectively, $\sigma(f_{n+1}; \otimes_{i=0}^n \mathcal{T})$-measurable).

(ii) A $\otimes_{i=0}^n \mathcal{U}(\mathcal{T})$-measurable r.c.p.d. (respectively a r.c.p.d.) $P_{fx}^{(fy_0,\ldots,fy_n)}(.|.) : \mathcal{T} \times T^{n+1} \to [0,1]$ exists and $\forall A \in \mathcal{T}. P_{fx}^{(fy_0,\ldots,fy_n)}(A|f_{n+1}(.)) = P_x^{(y_0,\ldots,y_n)}(f^{-1}A|.)$.

Definition 2.1. A surjective measurable function $f : (S,\mathcal{S}) \to (T,\mathcal{T})$ satisfying the equivalent conditions in Theorem 2.3 with respect to the random variables x and y is called $P_x^y(.|.)$-*preserving*.

Proposition 2.2. Suppose x and y are random variables with values on $(S,\mathcal{S})$ and $f : S \to T$ is a measurable surjective map to the measurable space $(T,\mathcal{T})$. Assume $P_x^y(.|.)$ and $P_x^{(y,fx)}(.|.)$ exist and f is $P_x^y(.|.)$-preserving. Then, for all $B \in \mathcal{T}$, all $A \in \mathcal{S}$ and for P_y-almost all $s \in S$ we have:

$$\int_B P_x^{(y,fx)}(A|(s,t))P_{fx}^{fy}(dt|f(s)) = P_x^y(A \cap f^{-1}B|s).$$

Proof. We first claim $\int_B P_x^{(y,fx)}(A\setminus f^{-1}B|(s,t))P_{fx}^y(dt|s)=0$, P_y-almost everywhere. Let 1_B be the characteristic function of B. Then,

$$\begin{aligned}
&\textstyle\int_S P_y(ds)\int_B P_x^{(y,fx)}(A\setminus f^{-1}B|(s,t))P_{fx}^y(dt|s) \\
&= \textstyle\int_S P_y(ds)\int_T P_x^{(y,fx)}(A\setminus f^{-1}B|(s,t))1_B(t)P_{fx}^y(dt|s) \\
&= E(P_x^{(y,fx)}(A\setminus f^{-1}B|(.,.))1_B)(y,fx) && \text{(by 10.3.2 of Hof94, vol II)} \\
&= \textstyle\int_{T\times S}(P_x^{(y,fx)}(A\setminus f^{-1}B|(s,t))1_B(t))P_{y,fx}(ds,dt) && \text{(by definition of expected value)} \\
&= \textstyle\int_{B\times S} P_x^{(y,fx)}(A\setminus f^{-1}B|(s,t))P_{y,fx}(ds,dt) \\
&= P(x\in A\setminus f^{-1}B\,,\,(y,fx)\in S\times B) \\
&= P(x\in A\setminus f^{-1}B\,,\,fx\in B) \\
&= 0
\end{aligned}$$

which proves the claim.

Now, since $P_{fx}^{fy}(.|f(s))=P_x^y(f^{-1}(.)|s)=P_{fx}^y(.|s)$ we get, for P_y-almost all $s\in S$:

$$\begin{aligned}
&\textstyle\int_B P_x^{(y,fx)}(A|(s,t))P_{fx}^{fy}(dt|f(s)) \\
&= \textstyle\int_B P_x^{(y,fx)}(A|(s,t))P_{fx}^y(dt|s) \\
&= \textstyle\int_B P_x^{(y,fx)}(A\cap f^{-1}B|(s,t))P_{fx}^y(dt|s) && \text{by the above claim} \\
&= P_x^y(A\cap f^{-1}B|s) && \text{by (Hoffmann-Jørgensen, 1994, 10.4.5,vol II),}
\end{aligned}$$

as required. □

Next, we show a stronger version of the above result if, in addition, S and T are analytic spaces and $f:S\to T$ is a Borel map.

Proposition 2.3. Suppose $(S,\mathcal{S})$ and $(T,\mathcal{T})$ are analytic spaces and that x and y are random variables with values on $(S,\mathcal{S})$ and $f:S\to T$ is a Borel surjective map. Assume f is $P_x^y(.|.)$-preserving. Then, there exists a $\mathcal{U}(\mathcal{S})\otimes\mathcal{U}(\mathcal{T})$-measurable r.c.p.d. $P_x^{(y,fx)}(.|.)$ such that for all $s\in S$ and all $A\in\mathcal{S}$ we have:

$$\int_B P_x^{(y,fx)}(A|(s,t))P_{fx}^{fy}(dt|f(s))=P_x^y(A\cap f^{-1}B|s).$$

Proof. By Theorem 2.2 a r.c.p.d. $P_x^{(y,fx)}(C|(.,.))$ exists. By the proof of Proposition 2.2, it is, therefore, sufficient to show that there exists a version $Q_x^{(y,fx)}(C|(.,.))$ of a $\mathcal{U}(\mathcal{S})\otimes\mathcal{U}(\mathcal{T})$-measurable r.c.p.d. such that we actually get

$$\int_B Q_x^{(y,fx)}(A\setminus f^{-1}B|(s,t))P_{fx}^y(dt|s)=0$$

for all $s\in S$. By Proposition 2.2, there exists a P_y-null Borel subset $N\subseteq S$ such that

$$\int_B P_x^{(y,fx)}(A\setminus f^{-1}B|(s,t))P_{fx}^y(dt|s)=0$$

for all $s\in S\setminus N$. Using Lemma 2.1, let $g:T\to S$ be a universally measurable map with $f\circ g=\mathrm{Id}_T$ such that $g(T)\in\mathcal{U}(\mathcal{S})$. Let

$$Q_x^{(y,fx)}(C|(s,t))=\begin{cases}\delta_{g(t)}(C) & \text{if } s\in N\\ P_x^{(y,fx)}(C|(s,t)) & \text{otherwise.}\end{cases}$$

We check that $Q_x^{(y,fx)}(C|(.,.)) : S \times T \to [0,1]$ is a u.m.r.c.p.d. as follows:

(1) The probability condition: Clearly for each $(s,t) \in S \times T$, $Q_x^{(y,fx)}(.|(s,t))$ is a probability distribution on S.

(2) The measurability condition: For $s \in N$, $Q_x^{(y,fx)}(C|(s,t))$ takes values 0 or 1 and is independent of s; we have $\{t \in T | Q_x^{(y,fx)}(C|(s,t)) = 1\} = \{t | \delta_{g(t)}(C) = 1\} = g^{-1}(C) \in \mathcal{U}(\mathcal{T})$ and hence $\{t \in T | Q_x^{(y,fx)}(C|(s,t)) = 0\} = T \setminus g^{-1}(C) \in \mathcal{U}(\mathcal{T})$. It follows that the map $(s,t) \mapsto \delta_{g(t)}(C) : S \times T \to [0,1]$ is $\mathcal{U}(\mathcal{S}) \otimes \mathcal{U}(\mathcal{T})$-measurable and, hence, the whole function $Q_x^{(y,fx)}(C|(.,.)) : S \times T \to [0,1]$ is also $\mathcal{U}(\mathcal{S}) \otimes \mathcal{U}(\mathcal{T})$-measurable. In fact any function $g : X \to Y$ of measurable spaces X and Y is measurable if it is defined by a measurable set $Z \subseteq X$ and measurable maps $h : X \to Y$ and $k : X \to Y$ such that: $g(x) = h(x)$ if $x \in X \setminus Z$ and $g(x) = k(x)$ if $x \in Z$. This is because for any measurable set $D \subseteq Y$ we have $g^{-1}(D) = (h^{-1}(D) \setminus Z) \cup (k^{-1}(D) \cap Z)$ which is measurable.

(3) The integral condition: By the usual argument, it is sufficient to show that the integral condition

$$\int_{D \times E} Q_x^{y,fx}(C|(s,t)) P_{y,fx}(ds,dt) = P(x \in C \ , \ (y,fx) \in D \times E)$$

holds for rectangles $D \times E$ where $D \subseteq S$ and $E \subseteq T$ are Borel subsets. We have:

$$\begin{array}{ll}
P(x \in C \ \& \ (y,fx) \in D \times E) & \\
= \int_{D \times E} P_x^{y,fx}(C|(s,t)) P_{y,fx}(ds,dt) & \text{r.c.p.d. condition} \\
= \int_D P_y(ds) \int_E P_x^{y,fx}(C|(s,t)) P_{fx}^y(dt|s) & \text{as in Proposition 2.2} \\
= \int_{D \setminus N} P_y(ds) \int_E P_x^{y,fx}(C|(s,t)) P_{fx}^y(dt|s) & \\
+ \int_{D \cap N} P_y(ds) \int_E P_x^{y,fx}(C|(s,t)) P_{fx}^y(dt|s) & \\
= \int_{D \setminus N} P_y(ds) \int_E P_x^{y,fx}(C|(s,t)) P_{fx}^y(dt|s) & \text{as } 0 \leq \int_E P_x^{y,fx}(C|(s,t)) P_{fx}^y(dt|s) \leq 1 \text{ and} \\
+0 & D \cap N \text{ is a } P_y\text{-null set} \\
= \int_{D \setminus N} P_y(ds) \int_E Q_x^{y,fx}(C|(s,t)) P_{fx}^y(dt|s) & \text{as } P_x^{y,fx}(C|(s,t)) = Q_x^{y,fx}(C|(s,t)) \text{ for } s \notin N \\
= \int_{D \setminus N} P_y(ds) \int_E Q_x^{y,fx}(C|(s,t)) P_{fx}^y(dt|s) & \\
+ \int_{D \cap N} P_y(ds) \int_E Q_x^{y,fx}(C|(s,t)) P_{fx}^y(dt|s) & \\
= \int_D P_y(ds) \int_E Q_x^{y,fx}(C|(s,t)) P_{fx}^y(dt|s) & \\
= \int_{D \times E} Q_x^{y,fx}(C|(s,t)) P_{y,fx}(ds,dt) & \text{as in Proposition 2.2.}
\end{array}$$

Therefore, $Q_x^{y,fx}(C|(.,.))$ is a version of $P_x^{y,fx}(C|(.,.))$. Furthermore, for $s \in N$ we have:

$$\int_B Q_x^{(y,fx)}(A \setminus f^{-1}B|(s,t)) P_{fx}^y(dt|s) = \int_B 1_{A \setminus f^{-1}B}(g(t)) P_{fx}^y(dt|s) = 0$$

since $g(t) \notin A \setminus f^{-1}B$ for $t \in B$.

□

Theorem 2.5. Suppose x and y are random variables with values on $(S, \mathcal{S})$ and $f : S \to T$ is a measurable function from S to the measurable space $(T, \mathcal{T})$. Assume a u.m.r.c.p.d. $P_{fx}^{fy}(.|.) : \mathcal{T} \times T \to [0,1]$ exists. Then, for all $\mathcal{U}(\mathcal{S}) \otimes \mathcal{U}(\mathcal{T})$-measurable functions $r : S \times T \to \overline{\mathbb{R}}$, the map $K(r) : S \to \overline{\mathbb{R}}$ with $K(r)s = \int_T r(s,t) P_{fx}^{fy}(dt|f(s))$ is $\mathcal{U}(\mathcal{S})$-measurable.

We first show the theorem when r is a characteristic function.

Lemma 2.4. The map $K(1_C)$ is $\mathcal{U}(\mathcal{S})$-measurable for each $C \in \mathcal{U}(\mathcal{S}) \otimes \mathcal{U}(\mathcal{T})$.

Proof. Let $\mathcal{A} \subseteq \mathcal{U}(\mathcal{S}) \otimes \mathcal{U}(\mathcal{T})$ be the collection of all $C \in \mathcal{U}(\mathcal{S}) \otimes \mathcal{U}(\mathcal{T})$ such that $K(1_C)$ is $\mathcal{U}(\mathcal{S})$-measurable. We show that $\mathcal{A}$ is a monotone class and contains the finite disjoint union of all rectangles $A \times B$ with $A \in \mathcal{U}(\mathcal{S})$ and $B \in \mathcal{U}(T)$. It then follows (Rudin, 1966, page 146) that $\mathcal{A} = \mathcal{U}(\mathcal{S}) \otimes \mathcal{U}(\mathcal{T})$. We have, by a short calculation, $K(1_{A\times B})(s) = P^{fy}_{fx}(B|f(s))1_A(s)$. By Lemma 2.2, $f : (S,\mathcal{U}(\mathcal{S})) \to (T,\mathcal{U}(\mathcal{T}))$ is measurable. Therefore, the composition $P^{fy}_{fx}(B|f(.)) : S \to [0,1]$ is $\mathcal{U}(\mathcal{S})$-measurable. It follows that $K(1_{A\times B})$ is the product of two $\mathcal{U}(\mathcal{S})$-measurable functions and, hence, $\mathcal{U}(\mathcal{S})$-measurable. For a finite disjoint union of rectangles $A_i \times B_i$ $(1 \le i \le n)$ we get $K(1_{\cup_i A_i\times B_i}) = \sum_i K(1_{A_i\times B_i})$ which is $\mathcal{U}(\mathcal{S})$-measurable. To show that $\mathcal{A}$ is a monotone class, let $C_0 \subseteq C_1 \subseteq C_2 \subseteq \cdots$ be an increasing sequence with $C_i \in \mathcal{A}$. For fixed $s \in S$ the maps $1_{C_i}(s,.) : T \to \overline{\mathbb{R}}$ form an increasing sequence of $\mathcal{U}(\mathcal{T})$-measurable functions with supremum $1_{\cup_i C_i}(s,.)$. Hence, by monotone convergence theorem $K(1_{\cup_i C_i})(s) = \sup_i K(1_{C_i})(s)$. Since $K(1_{C_i})$ $(i = 0,1,2\ldots)$ is an increasing sequence of $\mathcal{U}(\mathcal{S})$-measurable functions, it follows that $K(1_{\cup_i C_i}) = \sup_i K(1_{C_i})$ is $\mathcal{U}(\mathcal{S})$-measurable. Therefore, $\cup_i C_i \in \mathcal{A}$. Similarly, if $C_0 \supseteq C_1 \supseteq C_2 \supseteq \cdots$ is a decreasing sequence with $C_i \in \mathcal{A}$, we have that $K(1_{\cap_i C_i})$ is $\mathcal{U}(\mathcal{S})$-measurable. $\square$

Corollary 2.3. $K(r)$ is $\mathcal{U}(\mathcal{S})$-measurable for any simple $\mathcal{U}(\mathcal{S}) \otimes \mathcal{U}(\mathcal{T})$-measurable function $r : S \times T \to \mathbb{R}$.

Proof of Theorem 2.5. By splitting r into its positive and negative parts, we can assume that r is non-negative. Let $r = \sup_i r_i$ where $r_i : S \times T \to \mathbb{R}$ are simple non-negative $\mathcal{U}(\mathcal{S}) \otimes \mathcal{U}(\mathcal{T})$-measurable functions. Then $K(r_i)$ $(i = 0,1,2\cdots)$ is an increasing sequence and, by Corollary 2.3, $K(r_i)$ is $\mathcal{U}(\mathcal{S})$-measurable for each $i \ge 0$. Since, for fixed $s \in S$, $r_i(s,.) : T \to \mathbb{R}$ $(i = 0,1,2,\cdots)$ is an increasing sequence with $\sup_i r_i(s,.) = r(s,.)$, by monotone convergence theorem we get: $\sup_i K(r_i)(s) = \int_T \sup r_i(s,t)P^{fy}_{fx}(dt|f(s)) = \int_T r(s,t)P^{fy}_{fx}(dt|f(s)) = K(r)(s)$. Therefore, $K(r)$ is the supremum of an increasing sequence of $\mathcal{U}(\mathcal{S})$-measurable functions and hence $\mathcal{U}(\mathcal{S})$-measurable.

3. Markov processes

A *stochastic process* on $(\Omega,\mathcal{F},P)$ with values on $(S,\mathcal{S})$ is a sequence $x_n : \Omega \to S$ $(n = 0,1,2,\cdots)$ of random variables. One can regard the stochastic process as a single random variable x with values on $(\prod_{i\ge0} S, \otimes_{i\ge0}\mathcal{S})$. The distribution of the stochastic process is defined to be the distribution of x which is completely determined if, for all $n \ge 0$, the joint distributions of x_i $(1 \le i \le n)$ is given. A *Markov process* is a stochastic process x_n $(n = 0,1,2,\cdots)$ for which $P(x_{n+1} \in A|x_0 = s_0,\cdots,x_n = s_n) = P(x_{n+1} \in A|x_n = s_n)$ for all $n \ge 0$ and $A \in \mathcal{S}$. The initial distribution, i.e. P_{x_0}, and the conditional probabilities $P(x_{n+1} \in A|x_n = s_n)$ $(n \ge 0)$ determine the joint distributions of x_i $(1 \le i \le n)$ for all $n \ge 0$ and hence the distribution P_x of the the Markov process (Breiman, 1968, Proposition 7.3). Therefore the conditional probabilities determine the dynamics of the process. A Markov process is *stationary* if $P(x_{n+1} \in A|x_n = s)$ is independent of n; in this case we can write the transition probabilities $P(x_{n+1} \in A|x_n = s)$ simply as $P(A|s)$.

Theorem 3.1. Let x_n be a Markov process with values on the measurable space $(S, \mathcal{S})$. Assume a u.m.r.c.p.d. (respectively, r.c.p.d.) $P^{x_n}_{x_{n+1}}(.|.)$ exists for each $n \geq 0$ and let $f : (S, \mathcal{S}) \to (T, \mathcal{T})$ be measurable and surjective and $P^{x_n}_{x_{n+1}}(.|.)$-preserving for all $n \geq 0$. Then $fx_0, fx_1, \cdots$ is a Markov process with a u.m.r.c.p.d. (respectively, r.c.p.d) $P^{fx_n}_{fx_{n+1}}(A|f(s)) = P^{x_n}_{x_{n+1}}(f^{-1}A|s)$.

Proof. We first show that, for all $n \geq 0$, a $\otimes^n_{i=0}\mathcal{U}(\mathcal{T})$-measurable r.c.p.d. (respectively a r.c.p.d.) $P^{(fx_0,\cdots,fx_n)}_{fx_{n+1}}(.|.)$ exists and we have $\forall A \in \mathcal{T}.\, P^{(fx_0,\cdots,fx_n)}_{fx_{n+1}}(A|f_n(s_0,\ldots,s_n)) = P^{x_n}_{x_{n+1}}(f^{-1}A|s_n)$. For each $A \in \mathcal{T}$, we have, by the Markov property and by our assumption respectively, $P^{(x_0,\cdots,x_n)}_{x_{n+1}}(f^{-1}A|(s_0,\ldots,s_n)) = P^{x_n}_{x_{n+1}}(f^{-1}A|s_n) = P^{fx_n}_{fx_{n+1}}(A|f(s_n))$. It follows that $P^{(x_0,\cdots,x_n)}_{x_{n+1}}(f^{-1}A|.)$ is $\sigma(f_n; \otimes^n_{i=0}\mathcal{U}(\mathcal{T}))$-measurable (respectively, $\sigma(f_n; \otimes^n_{i=0}\mathcal{T})$-measurable). We conclude, by Theorem 2.4, that for all $n \geq 0$, $P^{(fx_0,\cdots,fx_n)}_{fx_{n+1}}(.|.)$ exists and $P^{(fx_0,\cdots,fx_n)}_{fx_{n+1}}(A|f_{n+1}(s_0,\ldots,s_n)) = P^{(x_0,\cdots,x_n)}_{x_{n+1}}(f^{-1}A|(s_0,\ldots,s_n)) = P^{x_n}_{x_{n+1}}(f^{-1}A|s_n)$. Now, by applying part (ii) of Theorem 2.3, we have

$$P^{(fx_0,\cdots,fx_n)}_{fx_{n+1}}(A|f_{n+1}(s_0,\ldots,s_n)) = P^{x_n}_{x_{n+1}}(f^{-1}A|s_n) = P^{fx_n}_{fx_{n+1}}(A|f(s_n)).$$

□

We can say something more in the case of stationary Markov processes on analytical spaces.

Proposition 3.1. Let x_n be a stationary Markov process on an analytic space $(S, \mathcal{S})$ and let $f : S \to T$ be a Borel map onto the analytic space $(T, \mathcal{T})$ which is $P^{x_n}_{x_{n+1}}(.,.)$-preserving. Then, there exists a version of a u.m.r.c.p.d. $P^{x_n,fx_{n+1}}_{x_{n+1}}(.|(.,.))$ such that

(i) for all $s \in S$, $t \in T$, $A \in \mathcal{S}$ and $B \in \mathcal{T}$:

$$\int_B P^{(x_n,fx_{n+1})}_{x_{n+1}}(A|(s,t))P^{fx_n}_{fx_{n+1}}(dt|f(s)) = P^{x_n}_{x_{n+1}}(A \cap f^{-1}B|s), \tag{1}$$

(ii) $P^{(x_n,fx_{n+1})}_{x_{n+1}}(.|(.,.))$ is almost surely independent of n, more specifically, there exists a function $Z(.|(.,.)) : \mathcal{S} \times (S \times T) \to [0,1]$, independent of n, such that for all $A \in \mathcal{S}$, $s \in S$ and for $P_{x_{n+1}}$-almost all $t \in T$, we have:

$$P^{(x_n,fx_{n+1})}_{x_{n+1}}(A|(s,t)) = Z(A|(s,t)). \tag{2}$$

Proof. Let $P^{x_n,fx_{n+1}}_{x_{n+1}}(.|.)$ be the version of a u.m.r.c.p.d., given by Proposition 2.3, which satisfies Equation 1. Fix $s \in S$ and let $R(.) = P^{x_n}_{x_{n+1}}(.|s)$ which is a probability measure on S and is independent of n. Consider $(S, \mathcal{S}, R)$, as a probability space, and the random variables $\mathrm{Id} : S \to S$ and $f : S \to T$. Then, the u.m.r.c.p.d. $R^f_{\mathrm{Id}}(.|.)$ is independent of n. Put $Z(A|(s,t)) = R^f_{\mathrm{Id}}(A|t)$. To show Equation 2, it is sufficient to show that, for our fixed $s \in S$, $P^{(x_n,fx_{n+1})}_{x_{n+1}}(.|(s,.))$ is a version of $R^f_{\mathrm{Id}}(.|.)$. Since $P^{(x_n,fx_{n+1})}_{x_{n+1}}(.|(.,.))$ is a u.m.r.c.p.d., for each $t \in T$, $P^{(x_n,fx_{n+1})}_{x_{n+1}}(.|(s,t))$ is a probability measure on S and for each $A \in \mathcal{S}$, the map $P^{(x_n,fx_{n+1})}_{x_{n+1}}(A|(s,.)) : T \to [0,1]$ is $\mathcal{U}(\mathcal{T})$-measurable. It only remains to check the integral condition. Note that $R_f(.) = P^{x_n}_{x_{n+1}}(f^{-1}(.)|s) = P^{fx_n}_{fx_{n+1}}(.|f(s))$.

Therefore, for all $s \in S$, $A \in \mathcal{S}$ and $B \in \mathcal{T}$ we have:

$$\begin{aligned}\int_B P^{(x_n, fx_{n+1})}_{x_{n+1}}(A|(s,t))R_f(dt) &= \int_B P^{(x_n, fx_{n+1})}_{x_{n+1}}(A|(s,t))P^{fx_n}_{fx_{n+1}}(dt|f(s)) \\ &= P^{x_n}_{x_{n+1}}(A \cap f^{-1}B|s) = R(A \cap f^{-1}B)\end{aligned}$$

by Equation 1. □

Assume now that we are working with measurable spaces $(S, \mathcal{S})$ on which u.m.r.c.p.d.'s exist. For analytic space this is guaranteed by Theorem 2.2. Let x_n $(n = 0, 1, 2, \ldots)$ be a stationary Markov process on $(S, \mathcal{S})$. Recall that a u.m.r.c.p.d. $P^{x_n}_{x_{n+1}}(.|.)$ is P_{x_n}-almost surely unique. For each $n \geq 0$, we fix a version of $P^{x_n}_{x_{n+1}}(.|.)$, denote it by $k(.,.)$ with $k(s, A) = P^{x_n}_{x_{n+1}}(A|s)$, for all $n \geq 0$, and call it the *universally measurable transition probability function.* (In case $P^{x_n}_{x_{n+1}}(.|.)$ are actually r.c.p.d.'s, we simply call $k(.,.)$ the *transition probability function.*) Note that $k : S \times \mathcal{S} \to [0, 1]$ satisfies the following properties:

— For each $A \in \mathcal{S}$ the map $k(., A) : S \to [0, 1]$ is $\mathcal{U}(\mathcal{S})$-measurable.
— For each $s \in S$, $k(s, .)$ is a probability measure on S.

Given $k(.,.)$ and any initial distribution x_0, one can obtain a Markov process whose transition probability functions are given by $k(.,.)$. We can therefore change our viewpoint and take the transition probability functions $k(.,.)$ as the given data, forgetting about the probability space $(\Omega, \mathcal{F}, P)$. This means that we are concerned with a family of Markov processes on $(S, \mathcal{S})$ having the same conditional probabilities $k(.,.)$ but different initial distributions; we denote this family by $(S, \mathcal{S}, k(.,.))$. Given two such families $(S, \mathcal{S}, k(.,.))$ and $(S', \mathcal{S}', k'(.,.))$, a surjective measurable map $f : S \to S'$ is called *transition probability preserving* if for all all $A \in \mathcal{S}'$ and all $s \in S$ we have: $k(s, f^{-1}A) = k'(f(s), A)$.

Theorem 3.2. Suppose $f : (S, \mathcal{S}, k(.,.)) \to (S', \mathcal{S}', k'(.,.))$ is surjective, measurable and transition probability preserving and x_n $(n = 0, 1, \ldots)$ is a stationary Markov process on $(S, \mathcal{S})$ with universally measurable transition probability function (respectively, transition probability function) $k(.,.)$.

(i) For all $A \in \mathcal{S}'$ the map $k(., f^{-1}A) : S \to [0, 1]$ is $\sigma(f; \mathcal{U}(\mathcal{S}'))$-measurable (respectively, $\sigma(f; \mathcal{S}')$-measurable).
(ii) fx_n $(n = 0, 1, \ldots)$ is a stationary Markov process on $(S', \mathcal{S}')$ with universally measurable transition probability function (respectively, transition probability function) $k'(.,.)$.

Proof. (i) This is the consequence of the fact that $k(., f^{-1}A) = k'(f(.), A)$ and f is $\mathcal{U}(\mathcal{S})$-measurable (respectively, $\mathcal{S}$-measurable).

(ii) Since $P^{x_n}_{x_{n+1}}(f^{-1}A|.) = k(., f^{-1}A)$ is $\sigma(f; \mathcal{U}(\mathcal{S}'))$ measurable (respectively, $\sigma(f; \mathcal{S}')$ measurable), f is $P^{x_n}_{x_{n+1}}(.|.)$-preserving. It follows by Theorem 3.1 that fx_n $(n = 0, 1, \ldots)$ is a Markov process and $P^{fx_n}_{fx_{n+1}}(A|f(s)) = P^{x_n}_{x_{n+1}}(f^{-1}A|s) = k(s, f^{-1}A) = k'(f(s), A)$. □

We now consider the following category. An object $(S, \mathcal{S}, k(.,.))$ of our category is given by a family of stationary Markov processes with a given universally measurable transition probability functions $k(.,.) : S \times \mathcal{S} \to [0, 1]$ on a Polish space S with its set $\mathcal{S}$ of Borel subsets. A morphism of our category is a continuous transition probability preserving surjective map. The composition of two morphisms is given by the composition of the two continuous maps whenever the latter exists. A diagram of morphisms will commute if and only if the corresponding diagram commutes in the category **Set** of sets and maps. If in an object $(S, \mathcal{S}, k(.,.))$ the transition probability function $k(.,.)$ is independent of the first argument, then we have $k(.,.) = \mu(.)$ where μ is a probability measure on S. Therefore, this object is simply given by a probability measure on S. Given two such objects $(S, \mathcal{S}, \mu)$ and $(T, \mathcal{T}, \nu)$, a morphism $f : (S, \mathcal{S}, \mu) \to (T, \mathcal{T}, \nu)$ is simply a measure-preserving map. Therefore, our category contains as a full subcategory the category of probability measures on Polish spaces and measure-preserving continuous surjective maps.

Three objects

$$S_1 = (S_1, \mathcal{S}_1, k_1 : S_1 \times \mathcal{S}_2 \to [0,1]),$$

$$S_2 = (S_2, \mathcal{S}_2, k_2 : S_2 \times \mathcal{S}_2 \to [0,1]),$$

$$S = (S, \mathcal{S}, k : S \times \mathcal{S} \to [0,1])$$

and morphisms $f_1 : S_1 \to S$ and $f_2 : S_2 \to S$ are given. We will construct a semi-pullback, i.e an object $(V, \mathcal{V}, h : V \times \mathcal{V} \to [0,1])$ with morphisms $\pi_i : V \to S_i$ $(i = 1, 2)$ such that the following diagram commutes.

$$\begin{tikzcd}
 & (V, \mathcal{V}, h : V \times \mathcal{V} \to [0,1]) \arrow[dl, "\pi_1"'] \arrow[dr, "\pi_2"] & \\
(S_1, \mathcal{S}_1, k_1 : S_1 \times \mathcal{S}_1 \to [0,1]) \arrow[dr, "f_1"'] & & (S_2, \mathcal{S}_2, k_2 : S_2 \times \mathcal{S}_2 \to [0,1]) \arrow[dl, "f_2"] \\
 & (S, \mathcal{S}, k : S \times \mathcal{S} \to [0,1]) &
\end{tikzcd}$$

Let $V = \{(s_1, s_2) \in S_1 \times S_2 \mid f_1(s_1) = f_2(s_2)\}$ equipped with the subspace topology of the product topology on $S_1 \times S_2$. The Borel σ-algebra $\mathcal{V}$ on V is given by $\{C \cap V | C \in \mathcal{S}_1 \otimes \mathcal{S}_2\}$ and is generated by the set $\{(A_1 \times A_2) \cap V \mid A_1 \in \mathcal{S}_1, A_2 \in \mathcal{S}_2\}$. Let $\pi_1 : V \to S_1$ and $\pi_2 : V \to S_2$ be the projection maps. We will construct $h : V \times \mathcal{V} \to [0,1]$ so that $(V, \mathcal{V}, h(.,.))$ is an object and $\pi_1 : V \to S_1$ and $\pi_2 : V \to S_2$ are morphisms with $f_1 \circ \pi_1 = f_2 \circ \pi_2$.

We fix $(s_1, s_2) \in V$, i.e. $s_1 \in S_1$ and $s_2 \in S_2$ with $f_1 s_1 = f_2 s_2$, and $n \in \mathbb{N}$ in the rest of this paper. The index i always takes the values 1 and 2. The variable s runs through S.

4. Construction of a probability measure on the product space

In order to define $h((s_1, s_2), -) : \mathcal{V} \to [0,1]$ we first define a probability measure $g((s_1, s_2), -) : \mathcal{S}_1 \otimes \mathcal{S}_2 \to [0,1]$ on the product space $(S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2)$. We will then show in the next section that $g((s_1, s_2), -)$ is supported on V, i.e. $g((s_1, s_2), V) = 1$. We will finally define $h((s_1, s_2), C \cap V) = g((s_1, s_2), C)$ for $C \in \mathcal{S}_1 \otimes \mathcal{S}_2$.

Assume in this and the next section that $x_n : \Omega \to S_1$ and $y_n : \Omega \to S_2$ are, respectively, stationary Markov processes on S_1 and S_2 with universally measurable transition probability functions $k_1(.,.)$ and $k_2(.,.)$. Note that for the measurable space $(\Omega, \mathcal{F})$ we can take, for example, the product space $((\prod_{i\geq 0} S_1) \times (\prod_{i\geq 0} S_2), (\otimes_{i\geq 0} S_1) \otimes (\otimes_{i\geq 0} S_2))$; the mappings x_n and y_n would then be just the nth coordinates of the first and second projections respectively. By Theorem 3.2, we get:

Corollary 4.1. $f_1 x_n$ and $f_2 y_n$ are Markov processes on S both with universally measurable transition probability functions $k(.,.)$.

By Proposition 2.3, we can and will work with a $\mathcal{U}(\mathcal{S}_1) \otimes \mathcal{U}(\mathcal{S})$-measurable r.c.p.d. $P_{x_{n+1}}^{x_n, f_1 x_{n+1}}(.|.)$ and a $\mathcal{U}(\mathcal{S}_2) \otimes \mathcal{U}(\mathcal{S})$-measurable r.c.p.d. $P_{y_{n+1}}^{y_n, f_2 y_{n+1}}(.|.)$ which both satisfy the conditions of Proposition 3.1 and, hence, are almost surely independent of n. Fix n for the rest of the paper. Let $\mathcal{R} = \mathcal{S}_1 \times \mathcal{S}_2$ be the semi-ring of all "rectangles" $A_1 \times A_2 \in \mathcal{S}_1 \otimes \mathcal{S}_2$. For $A_1 \times A_2 \in \mathcal{R}$ and $s \in S$, put

$$\rho(A_1 \times A_2 | ((s_1, s_2), s)) = P_{x_{n+1}}^{(x_n, f_1 x_{n+1})}(A_1 | (s_1, s)) P_{y_{n+1}}^{(y_n, f_2 y_{n+1})}(A_2 | (s_2, s)). \tag{3}$$

Note that the map $r : (s_1, s_2, s) \mapsto \rho(A_1 \times A_2 | ((s_1, s_2), s) : S_1 \times S_2 \times S \to [0,1]$ is $\mathcal{U}(\mathcal{S}_1) \otimes \mathcal{U}(\mathcal{S}_2) \otimes \mathcal{U}(\mathcal{S})$-measurable. Since $P_{x_{n+1}}^{(x_n, f_1 x_{n+1})}(.|(s_1, s))$ and $P_{y_{n+1}}^{(y_n, f_2 y_{n+1})}(.|(s_2, s))$ are, for all $s_1 \in S_1, s_2 \in S_2, s \in S$, probability measures on S_1 and S_2 respectively, we have:

Proposition 4.1. For all $s \in S$, $\rho(.|((s_1, s_2), s)$ is countably additive on $\mathcal{R}$.

Proof. See (Dudley, 1989, Theorem 4.4.1). □

For $A_1 \times A_2 \in \mathcal{R}$, put

$$g((s_1, s_2), A_1 \times A_2) = \int_S \rho(A_1 \times A_2 | ((s_1, s_2), s)) k(f_1 s_1, ds). \tag{4}$$

Proposition 4.2. $g((s_1, s_2), -)$ has a unique extension to a probability measure on $(S_1 \times S_2, \mathcal{S}_1 \otimes \mathcal{S}_2)$.

Proof. We check that $g((s_1, s_2), -)$ is countably additive on $\mathcal{R}$. Let $A_1 \times A_2 = \bigcup_{j\in J} A_{j1} \times A_{j2}$ be the disjoint union of a countable set of rectangles. Then,

$$\begin{aligned}
& g((s_1, s_2), A_1 \times A_2) \\
&= \int_S \rho(A_1 \times A_2 | ((s_1, s_2), s)) k(f_1 s_1, ds) \\
&= \int_S \sum_{j\in J} \rho(A_1 \times A_2 | ((s_1, s_2), s)) k(f_1 s_1, ds) \qquad \text{by Proposition 4.1} \\
&= \textstyle\sum_{j\in J} \int_S \rho(A_1 \times A_2 | ((s_1, s_2), s)) k(f_1 s_1, ds) \quad \text{monotone convergence theorem} \\
&= \textstyle\sum_{j\in J} g((s_1, s_2), A_{j1} \times A_{j2}).
\end{aligned}$$

This establishes countable additivity. We next extend $g((s_1, s_2), -)$ in the usual way (Dudley, 1989, page 102) to the ring $\mathcal{A}$ generated by the rectangles in $\mathcal{R}$, i.e. to the collection of all unions of finitely many disjoint elements of $\mathcal{R}$: For a finite set of disjoint rectangles $A_{j1} \times A_{j2}$, $j = 1, \cdots, n$, we put

$$g((s_1, s_2), \bigcup_{1 \le j \le n} A_{j1} \times A_{j2}) = \sum_{1 \le j \le n} g((s_1, s_2), A_{j1} \times A_{j2}).$$

In fact $\mathcal{A}$ is an algebra of subsets of $S_1 \times S_2$ since it contains $S_1 \times S_2$. Furthermore,

$$\begin{aligned} g((s_1, s_2), S_1 \times S_2) &= \int_S P_{x_{n+1}}^{(x_n, f_1 x_{n+1})}(S_1 | (s_1, s)) P_{y_{n+1}}^{(y_n, f_2 y_{n+1})}(S_2 | (s_2, s)) k(f_1 s_1, ds) \\ &= \int_S k(f_1 s_1, ds) = 1. \end{aligned}$$

Therefore by standard results in measure theory (Dudley, 1989, Theorems 3.1.4 and 3.1.10), $g((s_1, s_2), -)$ has a unique extension to $\mathcal{S}_1 \otimes \mathcal{S}_2$, i.e., the σ-algebra generated by $\mathcal{R}$ or $\mathcal{A}$. □

Next we need to check the required measurability of $g(., C)$ for any $C \in \mathcal{S}_1 \otimes \mathcal{S}_2$.

Lemma 4.1. Given any two measurable spaces $(S, \mathcal{S})$ and $(T, \mathcal{T})$, we have $\mathcal{U}(\mathcal{S}) \otimes \mathcal{U}(\mathcal{T}) \subseteq \mathcal{U}(\mathcal{S} \otimes \mathcal{T})$.

Proof. Let $A \in \mathcal{U}(\mathcal{S})$ and $B \in \mathcal{U}(\mathcal{T})$. We claim that $A \times T \in \mathcal{U}(\mathcal{S} \otimes \mathcal{T})$. Suppose μ is a probability measure on $S \times T$ and define a measure μ_1 on S by $\mu_1(R) = \mu(C \times T)$ for $R \in \mathcal{S}$. By assumption, there exist $C, D \in \mathcal{S}$ with $C \subseteq A \subseteq D$ and $\mu_1(D \setminus C) = 0$. Hence, $C \times T \subseteq A \times T \subseteq D \times T$ with $C \times T, D \times T \in \mathcal{S} \otimes \mathcal{T}$ and $\mu((D \times T) \setminus (C \times T) = \mu((D \setminus C) \times T) = \mu_1(D \setminus C) = 0$, which proves the claim. Similarly, $S \times B \in \mathcal{U}(\mathcal{S} \otimes \mathcal{T})$. Therefore, $A \times B = (A \times T) \cap (S \times B) \in \mathcal{U}(\mathcal{S} \otimes \mathcal{T})$. The result now follows since $\mathcal{U}(\mathcal{S}) \otimes \mathcal{U}(\mathcal{T})$ is the smallest σ-algebra which contains all rectangles $A \times B$ as above. □

Theorem 4.1. For any $C \in \mathcal{S}_1 \otimes \mathcal{S}_2$, the mapping $g(., C) : S_1 \times S_1 \to [0, 1]$ is $\mathcal{U}(\mathcal{S}_1 \otimes \mathcal{S}_2)$-measurable.

Proof. Let $\mathcal{A} \subseteq \mathcal{S}_1 \otimes \mathcal{S}_2$ be the set of all $C \in \mathcal{S}_1 \otimes \mathcal{S}_2$ such that $g(., C) : S_1 \times S_2 \to [0, 1]$ is measurable. We show that $\mathcal{A}$ contains the disjoint finite union of rectangles and is a monotone class; it will then follow that $\mathcal{A} = \mathcal{S}_1 \otimes \mathcal{S}_2$. Let $f : S_1 \times S_2 \to S$ with $f(a_1, a_2) = f_1 a_1$. Consider the rectangle $A_1 \times A_2$ and $g(., A_1 \times A_2)$ given by Equation 4. Note that, for our random variables x_j and y_j $(j = 0, 1, 2 \ldots)$, we have $f(x_j, y_j) = f_1 x_j$ for all $j \ge 0$. Since the map

$$\begin{aligned} r : \quad (S_1 \times S_2) \times S &\to [0, 1] \\ ((s_1, s_2), s) &\mapsto P_{x_{n+1}}^{(x_n, f_1 x_{n+1})}(A_1 | (s_1, s)) P_{y_{n+1}}^{(y_n, f_2 y_{n+1})}(A_2 | (s_2, s)), \end{aligned}$$

is $\mathcal{U}(\mathcal{S}_1) \otimes \mathcal{U}(\mathcal{S}_2) \otimes \mathcal{U}(\mathcal{S})$-measurable, by Lemma 4.1, it is $\mathcal{U}(\mathcal{S}_1 \otimes \mathcal{S}_2) \otimes \mathcal{U}(\mathcal{S})$-measurable. By Theorem 2.5 applied to the $\mathcal{S}$-measurable map $f : S_1 \times S_2 \to S$ and the $\mathcal{U}(\mathcal{S}_1 \otimes \mathcal{S}_2) \otimes \mathcal{U}(\mathcal{S})$-measurable map $r : (S_1 \times S_2) \times S \to [0, 1]$ with $x = (x_{n+1}, y_{n+1})$ and $y = (x_n, y_n)$, it follows that $K(r) = g(., A_1 \times A_2)$ is $\mathcal{U}(\mathcal{S}_1 \otimes \mathcal{S}_2)$-measurable. Since $g(((s_1, s_2), .)$ is additive, $\mathcal{A}$ contains the disjoint union of all rectangles. Finally, $\mathcal{A}$ is a monotone class: For an increasing sequence $C_0 \subseteq C_1 \subseteq \ldots$ we have $g(((s_1, s_2), \bigcup_j C_j) = \sup_j g(((s_1, s_2), C_j)$ and hence $g(.,\bigcup_j C_j) = \sup_j g(.,C_j)$ is $\mathcal{U}(\mathcal{S}_1 \otimes \mathcal{S}_2)$-measurable if each $g(.,C_j)$ is $\mathcal{U}(\mathcal{S}_1 \otimes \mathcal{S}_2)$-measurable; similarly for decreasing sequences of subsets.

$\square$

By applying the technique in the proof of Theorem 2.5 directly we can indeed show the stronger result that $g(.,C) : S_1 \times S_1 \to [0,1]$ is in fact $\mathcal{U}(\mathcal{S}_1) \otimes \mathcal{U}(\mathcal{S}_2)$-measurable.

5. Construction of the semi-pullback

We will now show that $g((s_1,s_2),V) = 1$. First note that the diagonal $D = \{(s,s) \mid s \in S\}$ is a closed subset of $S \times S$ equipped with the product topology; hence $(S \times S) \setminus D = \bigcup_{i \geq 0} O_{1i} \times O_{2i}$, where O_{1i} and O_{2i} are open subsets of S. Furthermore, the function $(f_1,f_2) : S_1 \times S_2 \to S \times S$ with $(f_1,f_2)(a_1,a_2) = (f_1 a_1, f_2 a_2)$ is continuous. It follows that $(S_1 \times S_2) \setminus V = \bigcup_{i \geq 0} (f_1^{-1} O_{1i}) \times (f_2^{-1} O_{2i})$ where for each $i \geq 0$ the sets $f_1^{-1} O_{1i}$ and $f_2^{-1} O_{2i}$ are open. Hence, in order to show that V has full $g((s_1,s_2),-)$-measure, it is sufficient to prove:

Proposition 5.1. For any $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$ with $V \cap (A_1 \times A_2) = \emptyset$ we have $g((s_1,s_2), A_1 \times A_2) = 0$.

We prove the above Proposition by using the following technical lemma. For the time being, we fix $A_1 \in \mathcal{S}_1$ and $A_2 \in \mathcal{S}_2$ as well as $s_1 \in S_1$ and $s_2 \in S_2$. For convenience we use the notation $\mu(-) = k(f_1 s_1, -)$. Recall that $P_{x_{n+1}}^{x_n, f_1 x_{n+1}}(.|.)$ and $P_{y_{n+1}}^{y_n, f_2 y_{n+1}}(.|.)$ have been both chosen to satisfy Equation 1.

Lemma 5.1. There exists, for $i = 1,2$, measurable subsets $B_i \subseteq S \setminus f_i A_i$ with $\mu(B_i) = \mu(S \setminus f_i A_i)$ such that for all $s \in B_1$ we have $P_{x_{n+1}}^{(x_n, f_1 x_{n+1})}(A_1|(s_1,s)) = 0$ and for all $s \in B_2$ we have $P_{y_{n+1}}^{(y_n, f_2 y_{n+1})}(A_2|(s_2,s)) = 0$.

Proof. By Corollaries 2.1 and 2.2, $f_1 A_1$ and hence $S \setminus f_1 A_1$ are μ-measurable. Therefore, there exists a Borel subset $C_1 \in \mathcal{S}_1$ with $C_1 \subseteq S \setminus f_1 A_1$ such that $\mu(C_1) = \mu(S \setminus f_1 A_1)$. Hence,

$$
\begin{aligned}
&\int_{C_1} P_{x_{n+1}}^{(x_n, f_1 x_{n+1})}(A_1|(s_1,s))\mu(ds) \\
&= \int_{C_1} P_{x_{n+1}}^{(x_n, f_1 x_{n+1})}(A_1|(s_1,s))k(f_1 s_1, ds) \\
&= \int_{C_1} P_{x_{n+1}}^{(x_n, f_1 x_{n+1})}(A_1|(s_1,s))P_{f_1 x_{n+1}}^{f_1 x_n}(ds|f_1 s_1) \\
&= P_{x_{n+1}}^{x_n}(A_1 \cap f_1^{-1} C_1 | s_1) && \text{by Equation 1} \\
&= k_1(s_1, A_1 \cap f_1^{-1} C_1) = 0
\end{aligned}
$$

since $(A_1 \cap f_1^{-1} C_1) \subseteq (A_1 \cap f_1^{-1}(S \setminus f_1 A_1)) = \emptyset$. Therefore, there exists a μ-measurable subset B_1 with $B_1 \subseteq C_1$ and $\mu(B_1) = \mu(C_1) = \mu(S \setminus f_1 A_1)$ such that $P_{x_{n+1}}^{(x_n, f_1 x_{n+1})}(A_1|(s_1,s)) = 0$ for all $s \in B_1$. Similarly, we obtain B_2. $\square$

Proof of Proposition 5.1 Since $(A_1 \times A_2) \cap V = \emptyset$, we have $f_1 A_1 \cap f_2 A_2 = \emptyset$ or equivalently

$$(S \setminus f_1 A_1) \cup (S \setminus f_2 A_2) = S.$$

It follows, by Lemma 5.1, that $\mu(S \setminus (B_1 \cup B_2)) = 0$. Therefore, for μ-almost all s we have $P^{(x_n, f_1 x_{n+1})}_{x_{n+1}}(A_1|(s_1, s)) = 0$ or $P^{(y_n, f_2 y_{n+1})}_{y_{n+1}}(A_2|(s_2, s)) = 0$. In other words, for μ-almost all s, we have $P^{(x_n, f_1 x_{n+1})}_{x_{n+1}}(A_1|(s_1, s)) P^{(y_n, f_2 y_{n+1})}_{y_{n+1}}(A_2|(s_2, s)) = 0$. Hence,

$$g((s_1, s_2), A_1 \times A_2) = \int_S P^{(x_n, f_1 x_{n+1})}_{x_{n+1}}(A_1|(s_1, s)) P^{(y_n, f_2 y_{n+1})}_{y_{n+1}}(A_2|(s_2, s)) \mu(ds) = 0,$$

as required.

Corollary 5.1. $g((s_1, s_2), V) = 1$.

Proof. By Proposition 5.1, $g((s_1, s_2), (S_1 \times S_2) \setminus V) = 0$, since $(S_1 \times S_2) \setminus V$ is the countable union of rectangles disjoint from V. □

An element of the subspace σ-algebra $\mathcal{V}$ induced on V is of the form $C \cap V$ with $C \in \mathcal{S}_1 \otimes \mathcal{S}_2$. We now define

$$h((s_1, s_2), C \cap V) = g((s_1, s_2), C), \tag{5}$$

for $C \in \mathcal{S}_1 \otimes \mathcal{S}_2$.

Lemma 5.2. Let $(S, \mathcal{S})$ be a measurable space and $(A, \mathcal{A})$ a subspace with $A \in \mathcal{S}$ and $\mathcal{A} = \{A \cap B | B \in \mathcal{S}\}$. Suppose $B \in \mathcal{U}(\mathcal{S})$, then $A \cap B \in \mathcal{U}(\mathcal{A})$.

Proof. Let μ be a measure on $(A, \mathcal{A})$. Then we have a measure μ' on $(S, \mathcal{S})$ defined, for $C \in \mathcal{S}$, by $\mu'(C) = \mu(A \cap C)$. Therefore, there exist $D, E \in \mathcal{S}$ with $C \subseteq B \subseteq D$ such that $\mu'(D \setminus C) = 0$. It follows that $A \cap D, A \cap E \in \mathcal{A}$ with $A \cap C \subseteq A \cap B \subseteq A \cap D$ such that $\mu(A \cap D \setminus A \cap C) = \mu'(D \setminus C) = 0$. □

Proposition 5.2. $h(., C \cap V) : V \to [0, 1]$ is $\mathcal{U}(\mathcal{V})$-measurable for each $C \in \mathcal{S}_1 \otimes \mathcal{S}_2$.

Proof. This follows immediately from Theorem 4.1 and Lemma 5.2. □

Theorem 5.1. $(V, \mathcal{V}, h : V \times \mathcal{V} \to [0, 1])$ is an object of the category and $\pi_i : V \to S_i$ are morphisms with $f_1 \circ \pi_1 = f_2 \circ \pi_2$.

Proof. Clearly V is a Polish space as it is a closed subset of the Polish space $S_1 \times S_2$. We have already shown that $h((s_1, s_2), -)$ is a probability measure on V for any $(s_1, s_2) \in V$

and $h(.,C\cap V)$ is $\mathcal{U}(\mathcal{V})$-measurable. Finally, for any $A_1\in\mathcal{S}_1$ we have:

$$\begin{aligned}
& h((s_1,s_2),\pi_1^{-1}A_1) \\
&= h((s_1,s_2),(A_1\times S_2)\cap V) \\
&= g((s_1,s_2),A_1\times S_2) \\
&= \int_S P_{x_{n+1}}^{(x_n,f_1x_{n+1})}(A_1|(s_1,s))P_{y_{n+1}}^{(y_n,f_2y_{n+1})}(S_2|(s_2,s))k(f_1s_1,ds) \\
&= \int_S P_{x_{n+1}}^{(x_n,f_1x_{n+1})}(A_1|(s_1,s))k(f_1s_1,ds) \\
&= \int_S P_{x_{n+1}}^{(x_n,f_1x_{n+1})}(A_1|(s_1,s))P_{f_1x_{n+1}}^{f_1x_n}(ds|f_1s_1) \\
&= P_{x_{n+1}}^{x_n}(f_1^{-1}A_1|s_1) \\
&= P_{x_{n+1}}^{x_n}(A_1|s_1) \qquad\qquad \text{by Equation 1} \\
&= k_1(s_1,A_1)=k_1(\pi_1(s_1,s_2),A_1).
\end{aligned}$$

Therefore π_1 is a morphism. Similarly π_2 is a morphism and we have $f_1\circ\pi_1=f_2\circ\pi_2$. $\square$

Corollary 5.2. The category of families of stationary Markov processes with given universally measurable transition probability functions on Polish spaces and surjective, continuous, transition probability preserving maps has semi-pullbacks.

For non-stationary Markov processes, ρ in Equation 3 and consequently $g(.,.)$ in Equation 4 and $h(.,.)$ in Equation 5 will depend on n and we get:

Corollary 5.3. The category of families of Markov processes with given universally measurable transition probability functions on Polish spaces and surjective, continuous, transition probability preserving maps has semi-pullbacks.

Furthermore, by considering the full subcategory whose objects have state-independent transition probability functions we get:

Corollary 5.4. The category of probability measures on Polish spaces and measure-preserving continuous surjective maps has semi-pullbacks.

Remark 5.1. If we are dealing with discrete processes, i.e. if S_1,S_2 and S are countable discrete spaces, then Equation 4 reduces to a sum and we get

$$\begin{aligned}
h((s_1,s_2),(A_1\times A_2)\cap V) &= g((s_1,s_2),A_1\times A_2) \\
&= \sum_{s\in S}\rho(A_1\times A_2|((s_1,s_2),s))k(f_1s_1,s).
\end{aligned}$$

Since, for $k(f_is_i,s)\neq 0$, we have

$$\begin{aligned}
P_{x_{n+1}}^{(x_n,f_1x_{n+1})}(A_1|(s_1,s)) &= \frac{k_1(s_1,A_1\cap f_1^{-1}s)}{k(f_1s_1,s)} \\
P_{y_{n+1}}^{(y_n,f_2y_{n+1})}(A_2|(s_2,s)) &= \frac{k_2(s_2,A_2\cap f_2^{-1}s)}{k(f_2s_2,s)}
\end{aligned}$$

we obtain

$$\rho(A_1 \times A_2|((s_1, s_2), s)) = \begin{cases} \dfrac{k_1(s_1, A_1 \cap f_1^{-1}s)k_2(s_2, A_2 \cap f_2^{-1}s)}{(k(f_1 s_1, s))^2} & \text{if } k(f_1 s_1, s) \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

6. Other categories and concluding remarks

Since regular conditional probability distributions exist on analytic spaces, we can easily check that our results hold in various other categories. Firstly, our construction of semi-pullbacks in the category of families of Markov processes, with given universally measurable transition probability functions, and transition probability preserving surjective maps goes through for various full subcategories including that of families of Markov processes on second countable locally compact Hausdorff spaces and for compact separable metric spaces. The only thing to check is that in Theorem 5.1 the space V given by a closed subset of the product space has the right type. But this is clear since, for example, a closed subset of a second countable locally compact space is another such space. The construction also goes through in the larger category of families of Markov processes on analytic spaces with surjective transition probability preserving Borel maps. A Borel subset of an analytic space with the subspace topology is an analytic space. This means that all the results in Section 5 hold in this larger category by replacing the terms Polish space, continuous map, and closed set respectively with analytic space, Borel map and Borel set. Note finally that all the above spaces are metrizable; our construction also goes through for all the corresponding categories of ultrametric spaces.

On the other hand pullbacks do not exist in these categories; they do not even exist in categories of Markov processes on finite discrete spaces (Desharnais et al., 1998). Since from a measure-theoretic viewpoint the construction of the semi-pullback is canonical, the question remains what universal property, if any, it satisfies.

Acknowledgements

I am grateful to Prakash Panangaden who initially raised the question if semi-pullbacks exist in any category of Markov processes on non-discrete spaces. I would also like to thank him and Josée Desharnais for pointing out a correction and a gap in the earlier versions of this paper and for discussions on analytic spaces. A careful proof-reading of the final version of the paper was carried out by Mauricio Alvarez-Manilla. Finally, I would like to acknowledge discussions on conditional probabilities with Terry Lyons in the early stage of this work. This work has been supported by EPSRC.

References

Arveson, W. (1976). *An Invitation to C*-Algebras.* Springer-Verlag.

Blute, R., Desharnais, J., Edalat, A., and Pananagden, P. (1997). Bisimulation for labelled Markov processes. In *Twelfth Annual IEEE Symposium on Logic in Computer Science.* IEEE.

Breiman, L. (1968). *Probability.* Addison-Weseley.

Desharnais, J., Edalat, A., and Pananagden, P. (1998). A logical characterization of bisimulation for labelled Markov processes. In *Thirteenth Annual IEEE Symposium on Logic in Computer Science (LICS)*. IEEE.

Dudley, R. M. (1989). *Real Analysis and Probability.* Wadsworth & Brooks/Cole.

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

Hoffmann-Jørgensen, J. (1970). The theory of analytic spaces. Various Publication Series 10, Mathematics Institute, Aarhus University.

Hoffmann-Jørgensen, J. (1994). *Probability with a view toward statistics.* Chapman & Hall. In two volumes.

Joyal, A., Nielsen, M., and Winskel, G. (1996). Bisimulation from open maps. *Information and Computation*, 127(2):164–185.

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

Pachl, J. (1978). Disintegration and compact measures. *Math. Scand.*, 43:157–168.

Rudin, W. (1966). *Real and Complex Analysis.* McGraw-Hill.

Appendix

The existence of regular conditional probability distribution (r.c.p.d.) on an analytic space follows form the work in (Pachl, 1978) and is folklore in certain sections of the probability theory community. For completeness, we will provide an accessible proof of this result which has been extracted from (Hoffmann-Jørgensen, 1994). We first need one more definition.

Recall that a finite Borel measure μ on a metric space is called *Radon* if for all Borel subsets B we have:

$$\mu(B) = \sup\{\mu(K) | K \subseteq B, K \text{ compact}\}.$$

We can now give a proof that r.c.p.d. exist on analytic spaces:

Proof of Theorem 2.2

There exists an a.s.r.c.p.d. for any Radon measure on any metric space (Hoffmann-Jørgensen, 1994, statement 10.29 and remark 2, vol. 2, page 155). Furthermore, any σ-finite Borel measure on an analytic space is Radon (Hoffmann-Jørgensen, 1994, P.5.19, vol. 2, page 560). Hence, as an analytic space can be considered as a metric space, there exists an a.s.r.c.p.d. for any probability Borel measure on an analytic space. Since an analytic space has a separable sigma-algebra, the existence of an a.s.r.c.p.d. gives, by (Hoffmann-Jørgensen, 1994)[10.26.3, vol. 2, page 152], the existence of a r.c.p.d. as desired.