Rend. Istit. Mat. Univ. Trieste
Vol. 53 (2021), Art. No. 27, pp. 1–25 (electronic)
DOI: 10.13137/2464-8728/33313

Congruences for stochastic automata

Ernst-Erich Doberkat

Abstract. *Congruences for stochastic automata are defined, the corresponding factor automata are constructed and investigated for automata over analytic spaces. We study the behavior under finite and infinite streams. Congruences consist of multiple parts, it is shown that factoring can be undertaken in multiple steps, guided by these parts.*

Keywords: Stochastic automata,Giry functor,congruences,factoring,stepwise reduction.
MS Classification 2020: 68Q87, 18A32, 68Q70.

1. Introduction

Stochastic automata [1, 6] are the natural generalization to non-deterministic Mealy automata; they take an input while being in an internal state, change their state and return an output. Both the new state and the output are distributed according to the automaton's transition law. The basic scenario may be finite or infinite, in the infinite case one may deal with countable or uncountable carrier sets for input, outputs, and states, resp. The finite and the countably infinite case is usually delt with through methods from linear algebra, since matrices with a finite or countable number of entries are manipulated, the uncountable case required methods from measure theory. This is so since the events an automata is assumed to handle are not all possible subsets of the contributing spaces. They rather come from Boolean σ-algebras of events; this is so because using all possible events, i.e., defining the probabilities on the respective power sets, will lead to foundational problems, see [19, p. 125–127].

This kind of automata – without the bells and whistles one finds in later extensions – has been used, e.g., for modelling simple learning processes along a behavioral taxonomy from psychology [6, 20, 26]. In such a scenario, in which the automaton models a learner. The automaton receives inputs from the environment while being in a specific state, it makes a state transition and responds with an output. This happens in a sequential fashion. We are interested mainly in the single-step behavior. Typical for a learning situation is the observation that equivalent inputs may lead to equivalent outputs, and that there may be equivalent states as well; note that the set of states represents an abstraction obtained through a modelling process, hence is not accessible from the outside. For conceptual clarity, and for minimizing the machine at least conceptually, one is interested in these equivalences, i.e., one wants to form equivalence classes and have the transition law respect these classes. This leads to the notion of a *congruence*, well known in (universal) algebra. But we must not ignore a slightly inconvenient fact: while a congruence, say, on a group, relates group elements to each other, an automaton congruence relates pairs of inputs and states to pairs of states and outputs, so we have a slightly heterogeneous situation at hand. One might be reminded of bisimilarity, where carrier sets of two possibly different transition systems are related to each other.

The latter problem is resolved by introducing the notion of *friendship* for two equivalence relations, comparing their probabilistic behavior in a straightforward manner. This yields a notion of congruence for automata, which is exploited here by relating it to morphisms and their kernels and constructing factor automata. For comparison we provide a contrasting view from this vantage point to the corresponding developments for discrete systems. We establish that the discrete approach is contained in the present one, albeit somewhat disguised, but still recognizable with some effort.

An automaton works sequentially, so we study also the automaton's behavior for finite and for infinite input sequences. Here we adopt a black box point of view, hiding state changes from the outside world. This is studied first for finite sequences, then we construct a limit which permits us also to specify behavior under an infinite input stream. Using the powerful Kolmogorov Consistency Theorem, it turns out that friendship is a surprisingly stable relationship which can be maintained also for infinite streams.

Finally we want to know whether we can form longer chains of reduced automata, and it turns out that this is not possible: factoring a factored automaton yields an automaton which can be obtained through one-step factoring by means of a suitably modified congruence. The result also enables us to reduce automata in a stepwise fashion along its components.

Most of the material depends heavily on what we know about Markov transition systems over measurable spaces, i.e., on the coalgebraic approach to stochastic relations [8, 10, 11, 22]. The present paper illustrates the well-known fact that the very old problem of reducing an automaton may be solved in a more general fashion without much effort making use of tools from coalgebras [10, 18]. The setting is much more general than the one for discrete stochastic automata. We widen the scope from state equivalence to friendship, thus considering equivalence relations on the inputs, the outputs, and on the states, rather than on the states only, which would be much more involved to treat in the discrete case by conventional means. In this way the coalgebraic contexts permits harvesting more general results, but at the cost of not posing questions pertaining to computability or complexity, let alone producing answers.

Notation and all that

A *measurable space* $(F, \mathcal{F})$ is a set F together with a Boolean σ-algebra $\mathcal{F}$ of subsets of F. Measurable spaces form a category, taking measurable maps as morphism. A map $f : F \to H$ for the measurable spaces $(F, \mathcal{F})$ and $(H, \mathcal{H})$ is said to be $\mathcal{F}$-$\mathcal{H}$ *measurable* iff $f^{-1}[\mathcal{H}] \subseteq \mathcal{F}$, i.e., iff $f^{-1}[Q] \in \mathcal{F}$ holds for every $Q \in \mathcal{H}$; we will omit the σ-algebras from the notation of maps whenever possible.

This category is closed under finite products: $(X, \mathcal{A}) \otimes (Y, \mathcal{B})$ has the Cartesian product $X \times Y$ as a carrier set and $\sigma(\{A \times B \mid A \in \mathcal{A}, B \in \mathcal{B}\}) =: \mathcal{A} \otimes \mathcal{B}$ as a σ-algebra. In Section 5 we will also need that this category is closed under countably infinite products. So let $\big((F_i, \mathcal{F}_i)\big)_{i\in\mathbb{N}}$ be a countable family of measurable spaces, then their product has $\prod_{i\in\mathbb{N}} F_i$ as a carrier set, and $\sigma\big(\{\prod_{i\in\mathbb{N}} A_i \mid A_i \in \mathcal{F}_i, A_j = F_j$ except for a finite number of indices $j\}\big)$ as the σ-algebra. The generators $\prod_{i\in\mathbb{N}} A_i$ are called *cylinder sets*. This category is also closed under coproducts.

We will frequently apply the *principle of good sets*, viz., we will have a look at the set of all sets which satisfy a property, and we want to show that this set comprises the σ-algebra we are interested in. The famous π-λ-Theorem of Dynkin is helpful in implementing this principle; we quote the theorem here for the reader's convenience and refer to [10, Theorem 1.6.30, p. 85–87] for a discussion. Denote to this end by $\sigma(\{\dots\})$ the smallest σ-algebra on the carrier containing the generator $\{\dots\}$.

Theorem 1.1 (Dynkin's π-λ-Theorem). *Let $\mathcal{P}$ be a family of subsets of F that is closed under finite intersections (this is called a π-class). Then $\sigma(\mathcal{P})$ is the smallest λ-class containing $\mathcal{P}$ (where a family of subsets of F is called a λ-class iff it is closed under complements and countable disjoint unions).*

We will not use the terms *π-class* or *λ-class* below. They are mentioned here because they illustrate nicely the name of this important theorem.

The *Giry functor* $\mathcal{G}$ acts as an endofunctor on this category. It assigns to each measurable space $(F, \mathcal{F})$ the set $\mathcal{G}(F, \mathcal{F})$ of all subprobabilities on $\mathcal{F}$ equipped with the smallest σ-algebra rendering the evaluations $\mu \mapsto \mu(Q)$ for all $Q \in \mathcal{F}$ measurable. To complete the definition of the functor $\mathcal{G}$, map the measurable map $f : F \to H$ to the measurable map $\mathcal{G}(f)$ which assigns each subprobability μ on $\mathcal{F}$ its image $\lambda P.\mu(f^{-1}[P])$ on $\mathcal{H}$, so that we have

$$\mathcal{G}(f)(\mu)(P) = \mu(f^{-1}[P]) \tag{1}$$

for all $P \in \mathcal{H}$.

For a finite or countable set F we simply omit the power set $\mathcal{P}(F)$ as the trivial σ-algebra and write

$$\mathcal{G}(F) = \{p : F \to [0,1] \mid \sum_{x\in F} p(x) \leq 1\}$$

of all subprobability vectors on F. A map $f : F \to H$ for another countable set H is mapped to

$$\mathcal{G}(f)(p)(y) = \sum\{p(x) \mid f(x) = y\}$$

through the Giry functor $\mathcal{G}$.

Assume an equivalence relation ξ on the measurable space $(F, \mathcal{F})$. The map $\eta_\xi : x \mapsto [x]_\xi$ sends an element to its ξ-class. Denote as usual the set of ξ-classes by F/ξ. This set will be furnished with the σ-algebra $\mathcal{F}/\xi$ which is the final σ-algebra on F/ξ with respect to $\mathcal{F}$ and η_ξ, thus $V \in \mathcal{F}/\xi$ iff $\eta_\xi^{-1}[V] \in \mathcal{F}$. We denote the measurable space $\big(F/\xi, \mathcal{F}/\xi\big)$ by $(F, \mathcal{F})/\xi$. 1_F denotes the identity relation on F.

2. Stochastic Automata

A *stochastic relation* $K : (X, \mathcal{A}) \Rightarrow (Y, \mathcal{B})$ is a measurable map $K : X \to \mathcal{G}(Y, \mathcal{B})$, thus $K(x)$ is a subprobability measure on $(Y, \mathcal{B})$ for each $x \in X$, and the map $x \mapsto K(x)(B)$ is $\mathcal{A}$-measurable for each $B \in \mathcal{B}$. Actually – but inconsequentially for the present note – a stochastic relation is a Kleisli morphism for the Giry monad, the functorial part of which is the Giry functor $\mathcal{G}$ [10, 23].

A σ-algebra is *countably generated* iff it has a countable generator, and it *separates points* iff given two distinct points there is a measurable set containing exactly one of them. It is well known that countably generated, point separating σ-algebras are precisely the Borel sets for second countable metric spaces [27].

Definition 2.1. *A* stochastic automaton $\mathbf{K} = \big((X, \mathcal{A}), (Y, \mathcal{B}), (Z, \mathcal{C}), K\big)$ *is a stochastic relation* $K : (X \times Z, \mathcal{A} \otimes \mathcal{C}) \Rightarrow (Z \times Y, \mathcal{C} \otimes \mathcal{B})$.

Thus the new state and the output of $\mathbf{K}$ is a member of the measurable set $D \in \mathcal{C} \otimes \mathcal{B}$ with probability $K(x, z)(D)$ upon input $x \in X$ in state $z \in Z$. Because we work in the realm of subprobabilities, mass may get lost, so that we cannot always reckon with $K(x, z)(Z \times Y) = 1$. This suggests the possibility that events cannot be accounted for.

Example 2.2. Before we loose ourselves in the jungle of measurable spaces, let us have a look at the finite situation. Assume that X, Y and Z are finite sets, then a finite stochastic automaton (X, Y, Z, p) is characterized by the probability $p(x, z)(z', y)$ with which the automaton reacts with an output $y \in Y$ and goes into a new state $z' \in Z$ upon receiving an input $x \in X$ in state $z \in Z$. Again, we assume that mass may get lost, so $p(x, z)$ is a subprobability on $Z \times Y$ rather than a probability. We model such an automaton also in the finite case through a map $p : X \times Z \to \mathcal{G}(Z \times Y)$.

The machine is assumed to work sequentially, so p is extended to a map $p^* : X^* \times Z \to \mathcal{G}\left(Z \times Y^*\right)$ with e.g., X^* denoting the set of all finite words over X such that

$$\forall x \in X \forall z \in Z : p^*(x, z) = p(x, z)$$

and

$$p^*(vv', z)(z', ww') := \sum_{z'' \in Z} p^*(v, z)(z'', w) \cdot p^*(v', z'')(z', w') \tag{2}$$

always hold. The map p^* will be used later on.

Sometimes the automaton is assumed to start in an initial state, or, more general, that the initial state follows some initial distribution. We will pursue this idea further when we have the automaton process finite or infinite input words, see Eq. (8).

The automata may work in different environments, so different input and output spaces have to be taken into account. Morphisms are used for relating automata. Assume that we have another stochastic automaton $\mathbf{K}' = \big((X', \mathcal{A}'), (Y', \mathcal{B}'), (Z', \mathcal{C}'), K'\big)$. A morphism $\mathfrak{f} : \mathbf{K} \to \mathbf{K}'$ is a triplet $\mathfrak{f} = (f, g, h)$ of surjective measurable map $f : X \to X'$, $g : Y \to Y'$ and $h : Z \to Z'$ rendering this diagram commutative (with, e.g., $f \times h : \langle x, z \rangle \mapsto \langle f(x), h(z) \rangle$):

$$\begin{tikzcd}
X \times Z \arrow[r, "K"] \arrow[d, "f \times h"'] & \mathcal{G}\left((Z \times Y, \mathcal{C} \otimes \mathcal{B})\right) \arrow[d, "\mathcal{G}(h \times g)"] \\
X' \times Z' \arrow[r, "K'"'] & \mathcal{G}\left((Z' \times Y', \mathcal{C}' \otimes \mathcal{B}')\right)
\end{tikzcd}$$

Thus

$$\begin{aligned}K'(f(x), h(z))(E) = \big(K' \circ (f \times h)\big)(x, z)(E) = \\ \big(\mathcal{G}\left(h \times g\right) \circ K\big)(x, z)(E) = K(x, z)((h \times g)^{-1}\left[E\right])\end{aligned}$$

whenever $E \in \mathcal{C}' \otimes \mathcal{B}'$ indicates the operation of automaton $\mathbf{K}'$.

Example 2.3. Continuing Example 2.2, let (X, Y, Z', p') be another automaton over the same alphabets X abd Y. Then a map $h : Z \to Z'$ induces a morphism $(1_X, 1_Y, h) : (X, Y, Z, p) \to (X, Y, Z', p')$ iff

$$p'(x, h(z)) = \sum \{p(x, z') \mid h(z') = h(z)\}$$

holds for each $x \in X, z \in Z$. Traditionally input and output alphabets are left untouched, on the assumption that both are provided by the environment, and that merely the set of states is subject to modelling.

3. Congruences

Before we define congruences for stochastic automata, we need to talk about friendly relations, i.e., relations on different states which behave nevertheless like congruences. To be specific: Given a stochastic relation $K : (F, \mathcal{F}) \Rightarrow (H, \mathcal{H})$ and equivalence relations ξ and ϑ on F resp. H, call ξ *friendly to* ϑ iff there exists a stochastic relation $K_{\xi,\vartheta} : (F, \mathcal{F})/\xi \Rightarrow (H, \mathcal{H})/\vartheta$ rendering this diagram commutative:

$$\begin{tikzcd} F \arrow[r, "K"] \arrow[d, "\eta_\xi"'] & \mathcal{G}\left(H, \mathcal{H}\right) \arrow[d, "\mathcal{G}(\eta_\vartheta)"] \\ F/\xi \arrow[r, "K_{\xi,\vartheta}"'] & \mathcal{G}\left((H, \mathcal{H})/\vartheta\right) \end{tikzcd} \tag{3}$$

We observe for friendly ξ, ϑ that

$$K_{\xi,\vartheta}([x]_\xi)(T) = \big(\mathcal{G}\left(\eta_\vartheta\right) \circ K\big)(x)(T) = K(x)(\eta_\vartheta^{-1}\left[T\right])$$

so that ξ and ϑ indeed cooperate in a congruential manner.

We will also need the concept of a small equivalence relation, given that *equivalence* is a very broad notion. It needs to be restricted somewhat for being useful in our context.

Again, assume an equivalence relation ξ on the measurable space $(F, \mathcal{F})$. Call the measurable set $Q \in \mathcal{F}$ *ξ-invariant* iff Q is the union of equivalence classes, thus iff $x \in Q$ and $x \; \xi \; x'$ entails $x' \in Q$. It is not difficult to see that

$$[\mathcal{F}, \xi] := \{Q \in \mathcal{F} \mid Q \text{ is } \xi\text{-invariant}\} \tag{4}$$

is a σ-algebra, the *σ-algebra of ξ-invariant sets.* Observe that $\eta_\xi\left[U\right] \in \mathcal{F}/\xi$ for $U \in [\mathcal{F}, \xi]$, because $\eta_\xi^{-1}\left[\eta_\xi\left[U\right]\right] = U$.

Example 3.1. Let us have a look at the finite situation again, so we look at the measurable space $(F, \mathcal{P}\left(F\right))$ with the powerset $\mathcal{P}\left(F\right)$ of the finite set F as the σ-algebra. Then the σ-algebra $\left[\mathcal{P}\left(F\right), \xi\right]$ of ξ-invariant sets can be described explicitly as

$$\left[\mathcal{P}\left(F\right), \xi\right] = \{\bigcup_{x \in A} [x]_\xi \mid A \subseteq F\},$$

because the equivalence classes for ξ form a partition of F.

A measurable map $m_\xi : (F, \mathcal{P}\left(F\right)) \to \big(F, \left[\mathcal{P}\left(F\right), \xi\right]\big)$ is induced by the identity $x \mapsto x$. Just for fun, let us have a look at what the Giry functor does to this map. By additivity is apparently enough to compute for

$q \in \mathcal{G}(F) = \mathcal{G}(F, \mathcal{P}(F))$ the value $\mathcal{G}(m_\xi)(q)([x]_\xi)$ on a class $[x]_\xi$. We obtain from Eq. (1)

$$\begin{aligned}\mathcal{G}(m_\xi)(q)([x]_\xi) &= q(\{x' \in F \mid x' \in [x]_\xi\}) \\ &= \sum\{q(x') \mid x' \ \xi \ x\}.\end{aligned}$$

So $\mathcal{G}(m_\xi)(q)$ collects the q-probabilities for the elements of an equivalence class, and assigns this as the probability for that class.

Call the equivalence relation ξ on the measurable space $(F, \mathcal{F})$ *small* iff there exists a countable family $\left(U_n\right)_{n\in\mathbb{N}} \subseteq \mathcal{F}$ such that

$$x \ \xi \ x' \text{ iff } \forall n \in \mathbb{N} : x \in U_n \Leftrightarrow x' \in U_n.$$

$\left(U_n\right)_{n\in\mathbb{N}}$ is said to *create* relation ξ. Then $[\mathcal{F}, \xi] = \sigma(\{U_n \mid n \in \mathbb{N}\})$ is countably generated, so is $\mathcal{F}/\xi$, which also separates points.

Example 3.2. Let $f : (F, \mathcal{F}) \to (H, \mathcal{H})$ be measurable, and assume that $\mathcal{H}$ is countably generated and separates points. Then the kernel relation

$$\ker(f) := \{\langle x, x'\rangle \mid f(x) = f(x')\}$$

is small. In fact, let $(U_n)_{n\in\mathbb{N}}$ be the generator for $\mathcal{H}$, then we show that $\{U_n \mid n \in \mathbb{N}\}$ separates points. Take $y, y' \in H$ such that $y \in U_n$ iff $y' \in U_n$ for all $n \in \mathbb{N}$. Since $\{U \subseteq H \mid \forall n \in \mathbb{N} : y \in U \Leftrightarrow y' \in U\}$ is a σ-algebra which contains the generator, it contains $\mathcal{H}$. From this we conclude that $y = y'$. But this means that $\left(f^{-1}[U_n]\right)_{n\in\mathbb{N}}$ creates $\ker(f)$.

The following observation helps characterizing friendly equivalence relations.

Lemma 3.3. *Let $K : (F, \mathcal{F}) \Rightarrow (H, \mathcal{H})$ be a stochastic relation and assume equivalence relations ξ and ϑ on F resp. H, are given. Then these conditions are equivalent:*

1. *ξ is friendly to ϑ.*
2. *$\mathcal{G}(m_\vartheta) \circ K : (F, [\mathcal{F}, \xi]) \Rightarrow (H, [\mathcal{H}, \vartheta])$ with $m_\vartheta : (H, \mathcal{H}) \to (H, [\mathcal{H}, \vartheta])$ as the identity.*
3. *$\ker(\mathcal{G}(m_\vartheta) \circ K) \supseteq \xi$.*

Proof. Abbreviate the map $\mathcal{G}(m_\vartheta) \circ K$ by L, and note that $\mathcal{G}(m_\vartheta)$ restricts measures on $\mathcal{H}$ to its sub σ-algebra $[\mathcal{H}, \vartheta]$.

1 $\Rightarrow$ 2: It is clear that $L : (F, \mathcal{F}) \Rightarrow (H, [\mathcal{H}, \vartheta])$, because $\mathcal{G}(m_\vartheta)$ acts as restriction to $[\mathcal{H}, \vartheta]$. So it has to be shown that $x \mapsto L(x)(G)$ is $[\mathcal{F}, \xi]$-measurable for each $G \in [\mathcal{H}, \vartheta]$. Let $G_0 := \eta_\vartheta[G] \in \mathcal{H}/\vartheta$, then $L(x, G) =$

$L(x, \eta_\vartheta^{-1}[G_0]) = (\mathcal{G}(\eta_\vartheta) \circ K)(x)(G_0)$, thus $L(x)(G) < r$ iff $K_{\xi,\vartheta}([x]_\xi)(G_0) < r$, which implies measurability of $x \mapsto L(x)(G)$.

2 $\Rightarrow$ 3: The assumption that there exists $T \in [\mathcal{H}, \vartheta]$ such that $K(x)(T) < r < K(x')(T)$ for some x, x' with $x \;\xi\; x$ gives immediately a contradiction.

3 $\Rightarrow$ 1: Define $K_{\xi,\vartheta}([x]_\xi) := (\mathcal{G}(\eta_\vartheta) \circ K)(x)$, then $K_{\xi,\vartheta}$ is well-defined, satisfies the measurability conditions and renders diagram (3) commutative. □

This useful characterization permits testing friendship without actually constructing the factors. It extends to bounded, measurable functions:

Corollary 3.4. *Under the assumptions of Lemma 3.3, these statements are equivalent*

1. *ξ is friendly to ϑ.*
2. *For each bounded and $[\mathcal{H}, \vartheta]$-measurable $f : H \to \mathbb{R}$*

$$x \;\xi\; x' \Rightarrow \int_H f \; dK(x) = \int_H f \; dK(x').$$

Proof. The implication 1 $\Rightarrow$ 2 follows from part 3 in Lemma 3.3 together with the observation that a bounded measurable function is the pointwise limit of a sequence of step functions, and Lebesgue's Convergence Theorem. The converse implication observes that the indicator function of a measurable set is a bounded measurable function. An application of part 3 in Lemma 3.3 yields the result. □

An interesting example for friendship is given by kernels of morphisms for stochastic relations. Recall that finality of a measurable map $f : (F, \mathcal{F}) \to (H, \mathcal{H})$ may be characterized by the property that $\mathcal{H} = \{R \subseteq H \mid f^{-1}[R] \in \mathcal{F}\}$. Thus we may conclude from $f^{-1}[R] \in \mathcal{F}$ that $R \in \mathcal{H}$, provided f is final and onto.

Example 3.5. Let $K_i : (F_i, \mathcal{F}_i) \Rightarrow (H_i, \mathcal{H}_i)$ be stochastic relations for $i = 1, 2$, and assume that $(f, g) : K_1 \to K_2$ is a morphism, which means $K_2 \circ f = \mathcal{G}(g) \circ K_1$ for the surjective measurable maps $f : F_1 \to F_2$ and $g : H_1 \to H_2$. We claim that $\ker(f)$ is friendly to $\ker(g)$, provided g is final and onto.

In fact, let $f(x) = f(x')$, then we have to show that $K_1(x)(G) = K_1(x')(G)$ for all $G \in [\mathcal{H}_1, \ker(g)]$. Fix such a set G; we know that $G = \eta_{\ker(g)}^{-1}\left[\eta_{\ker(g)}[G]\right]$ with $\eta_{\ker(g)}[G] \in \mathcal{H}_1/\ker(g)$. Factoring $g = g_\bullet \circ \eta_{\ker(g)}$ with $g_\bullet : H_1/\ker(g) \to H_2$ measurable, final and injective yields the surjective map $g_\bullet^{-1}$ between powersets. We find therefore $H_0 \subseteq H_2$ with $g_\bullet^{-1}[H_0] = \eta_{\ker(g)}[G]$. Because

$$g^{-1}[H_0] = \eta_{\ker(g)}^{-1}\left[g_\bullet^{-1}[H_0]\right] = \eta_{\ker(g)}^{-1}\left[\eta_{\ker(g)}[G]\right] = G \in [\mathcal{H}_1, \ker(g)] \subseteq \mathcal{H}_1$$

we conclude from finality of $g_\bullet$ that $H_0 \in \mathcal{H}_2$, so that

$$K_1(x)(G) = K_1(x)(g^{-1}[H_0]) = \\ (\mathcal{G}(g) \circ K_1)(x)(H_0) = K_2(f(x))(H_0) = \\ K_2(f(x'))(H_0) = K_1(x')(G).$$

This gives the assertion.

After these somewhat lenghty preparations we are in a position to define congruences for stochastic automata.

DEFINITION 3.6. *Let* $\mathbf{K} = \big((X,\mathcal{A}),(Y,\mathcal{B}),(Z,\mathcal{C}),K\big)$ *be a stochastic automaton, then a triplet* $\mathfrak{c} = (\alpha,\beta,\gamma)$ *of equivalence relations on* X, Y *resp.* Z *is called a* congruence for $\mathbf{K}$ *iff* $\alpha \times \gamma$ *is friendly to* $\gamma \times \beta$.

A congruence $\mathfrak{c}$ for stochastic automaton $\mathbf{K}$ is characterized by the existence of a stochastic relation

$$K_\mathfrak{c} : \big((X,\mathcal{A}) \otimes (Z,\mathcal{C})\big)/(\alpha \times \gamma) \Rightarrow \big((Z,\mathcal{C}) \otimes (Y,\mathcal{B})\big)/(\gamma \times \beta) \tag{5}$$

which renders this diagram commutative:

$$\begin{tikzcd}
(X,\mathcal{A}) \otimes (Z,\mathcal{C}) \arrow[r, "K"] \arrow[d, "\eta_{\alpha\times\gamma}"'] & \mathcal{G}\left((Z,\mathcal{C}) \otimes (Y,\mathcal{B})\right) \arrow[d, "\mathcal{G}(\eta_{\gamma\times\beta})"] \\
\big((X,\mathcal{A}) \otimes (Z,\mathcal{C})\big)/(\alpha \times \gamma) \arrow[r, "K_\mathfrak{c}"'] & \mathcal{G}\left(\big((Z,\mathcal{C}) \otimes (Y,\mathcal{B})\big)/(\gamma \times \beta)\right)
\end{tikzcd}$$

This is an immediate consequence:

PROPOSITION 3.7. *In the notation of Definition 3.6,* $(\eta_\alpha, \eta_\beta, \eta_\gamma) : \mathbf{K} \to \mathbf{K}_\mathfrak{c}$ *is a morphism.* $\dashv$

The classic case of state reduction, studied originally for finite automata, by a relation γ for automaton $\mathbf{K} = \big((X,\mathcal{A}),(Y,\mathcal{B}),(Z,\mathcal{C}),K\big)$ is captured through the triplet $\mathfrak{s} = (1_X, 1_Y, \gamma)$; that $\mathfrak{s}$ is a congruence for $\mathbf{K}$ is characterized through

$$\forall B \in \mathcal{B} : K(x,z)(E \times B) = K(x,z')(E \times B),$$

whenever $z \;\gamma\; z'$, and $E \in [\mathcal{C},\gamma]$ is a γ-invariant measurable subset of Z. This is quite close to the intuition of a (state-) congruence for an automaton: equivalent states behave in the same way on those measurable sets which cannot separate equivalent states.

On the other hand, one probably wants to leave the states alone and cater only for inputs and outputs. Here one would work with $\mathfrak{t} = (\alpha, \beta, 1_Z)$, and $\mathfrak{t}$ is a congruence iff

$$\forall C \in \mathcal{C} : K(x,z)(C \times B) = K(x',z)(C \times B),$$

whenever $x \, \alpha \, x'$ and $B \in [\mathcal{B}, \beta]$, so the behavior of $\mathbf{K}$ on inputs which are identified through α is the same on those sets which cannot separate β-equivalent outputs. Certainly other combinations are possible.

It is noted that the behavior of an automaton is completely characterized by its assigning values to sets of the form $C \times B$. This is so because these sets determine the respective product σ-algebras uniquely, and their collection is closed under intersections and, by the very definition of a finite measure, under countable disjoint unions as well as complements [10, Lemma 1.6.31].

Example 3.8. Let us have a look at what happens in the case of a finite automaton (X, Y, Z, p). Extend $p : X \times Z \to \mathcal{G}(Z \times Y)$ to $p^* : X^* \times Z \to \mathcal{G}(Z \times Y^*)$ as in Example 2.2. Call states z and z' *equivalent* iff

$$\forall v \in X^* \forall w \in Y^* : \sum_{z'' \in Z} p^*(v, z)(z'', w) = \sum_{z'' \in Z} p^*(v, z')(z'', w),$$

see [5, p. 24] or [4, p. 13], and denote this equivalence relation by Γ.

We fit this definition into the framework considered here. Curry $p^* : X^* \times Z \to \mathcal{G}(Z \times Y^*)$ to a map $\widehat{p^*} : Z \to X^* \to \mathcal{G}(Z \times Y^*)$ and define $M : \mathcal{G}(Z \times Y^*) \to \mathcal{G}(Y^*)$ by

$$M(q) : w \mapsto \sum_{z \in Z} q(z, w).$$

Probabilistically speaking, $M(q)$ is the marginal distribution of q on Y^*. Then[1]

$$\Gamma = \ker\left(M \circ \widehat{p^*}\right),$$

because

$$\left(M \circ \widehat{p^*}\right)(z)(v)(w) = \sum_{z' \in Z} p^*(v, z)(z', w)$$

by expanding definitions.

We claim that $1_X \times \Gamma$ is friendly to $\Gamma \times 1_Y$. We know from [5, Satz 1] or [4, Theorem 4.2] that

$$P\big(x, [z]_\Gamma\big)\big([z']_\Gamma, y\big) := \sum_{z'' \; \Gamma \; z'} p(x, z)(z'', y)$$

$(x \in X, y \in Y, z, z' \in Z)$ defines the factor automaton $(X, Y, Z/\Gamma, P)$. Translating this into our scenario, we see that this exactly what is stated in part 2 of Lemma 3.3.

[1] The reader is invited to reflect on J. W. v. Goethe' quote: *Die Mathematiker sind eine Art Franzosen: redet man zu ihnen, so übersetzen sie es in ihre Sprache und dann ist es alsbald ganz etwas Anderes.* [24, p. 5]

4. Factoring

We will restrict the class of measurable spaces to analytic spaces now, and we will deal only with small equivalence relations.

Recall that an *analytic space* is the measurable image of a *Polish space*, i.e., of a second countable, completely metrizable topological space. Analytic spaces are topological spaces in their own right with a countable and point separating base for their topology. As topological spaces they carry the σ-algebra of Borel sets. For the rest of the paper we will assume that analytic spaces are equipped with just these Borel sets. This will render notation lighter as well, because it will permit us to omit the σ-algebra for an analytic space from notation. Measurability refers to the Borel sets, unless otherwise noted.

Analytic spaces have a number of desirable technical properties [10, 27], among them the closure under countable products; we note also that $\mathfrak{B}(F \times H) = \mathfrak{B}(F) \otimes \mathfrak{B}(H)$ for analytic spaces F and H, $\mathfrak{B}(\dots)$ denoting the Borel sets. Alas, that the product of Borel sets equals the Borel sets of a product is far from being common among topological spaces. In general this requires some additional assumptions. Just to emphasize this property, we have for the analytic spaces F and H

$$\begin{aligned}\mathfrak{B}(F \times H) &\overset{(*)}{=} \sigma\bigl(\{W \mid W \subseteq F \times H \text{ is open}\}\bigr)\\ &\overset{(+)}{=} \sigma\bigl(\{U \times V \mid U \in \mathfrak{B}(F), V \in \mathfrak{B}(H)\}\bigr)\\ &= \mathfrak{B}(F) \otimes \mathfrak{B}(H)\end{aligned}$$

Here equation $(*)$ derives from the definition of the Borel sets as the smallest σ-algebra containing the open sets, and equation $(+)$ derives from the definition of the product σ-algebra. Analytic spaces are also closed under factoring through small equivalence relations ([27, Exercise 5.1.14], [10, Proposition 4.4.22]).

A first witness to usefulness is given by the following observation (cp. [9, Corollary 2.11]).

LEMMA 4.1. *Assume that ξ and ζ are small equivalence relations on the analytic spaces F esp. H. Then*

1. $[\mathfrak{B}(F \times H), \xi \times \zeta] = [\mathfrak{B}(F), \xi] \otimes [\mathfrak{B}(H), \zeta]$.
2. *The measurable spaces $(F \times H)/(\xi \times \zeta)$ and $F/\xi \times H/\zeta$ are isomorphic.*

The second assertion, in its full beauty, means that the factor space $\bigl(F \times H, \mathfrak{B}(F \times H)\bigr)/(\xi \times \zeta)$ is isomorphic to $\bigl(F, \mathfrak{B}(F)\bigr)/\xi \otimes \bigl(H, \mathfrak{B}(H)\bigr)/\zeta$.

Proof. 1. Assume that ξ and ζ have the respective generators $(U_n)_{n\in\mathbb{N}}$ and $(V_m)_{m\in\mathbb{N}}$. Since $\langle x, y\rangle$ $(\xi \times \zeta)$ $\langle x', y'\rangle$ iff

$$\forall n \in \mathbb{N}\, \forall m \in \mathbb{N} : \bigl[x \in U_n \Leftrightarrow x' \in U_n\bigr] \wedge \bigl[z \in V_m \Leftrightarrow z' \in V_m\bigr],$$

we see that

$$\begin{aligned}[\mathfrak{B}(F \times H), \xi \times \zeta] &= \sigma(\{U_n \times V_m \mid n, m \in \mathbb{N}\}) \\ &= \sigma(\{U_n \mid n \in \mathbb{N}\}) \otimes \sigma(\{V_m \mid m \in \mathbb{N}\}) \\ &= [\mathfrak{B}(F), \xi] \otimes [\mathfrak{B}(H), \zeta]\end{aligned}$$

2. It is not difficult to see that $[\langle x, y\rangle]_{\xi\times\zeta} \mapsto \langle [x]_\xi, [y]_\zeta\rangle$ is a bijection and measurable. Now look a the inverse ℓ. We want to show that $\ell^{-1}[E] \in \big(F, \mathfrak{B}(F)\big)/\xi \otimes \big(H, \mathfrak{B}(H)\big)/\zeta$ for each $E \in \big(F \times H, \mathfrak{B}(F \times H)\big)/(\xi \times \zeta)$. By the observation following (4) it is sufficient to show that $\ell^{-1}[\eta_{\xi\times\zeta}[D]] \in \mathfrak{B}(F/\xi) \otimes \mathfrak{B}(H/\zeta)$ for every $D \in [\mathfrak{B}(F \times H), \xi \times \zeta]$. Now the principle of good sets kicks in helpfully: The set

$$\mathcal{D} := \big\{D \in [\mathfrak{B}(F \times H), \xi \times \zeta] \mid \ell^{-1}[\eta_{\xi\times\zeta}[D]] \in \mathfrak{B}(F/\xi) \otimes \mathfrak{B}(H/\zeta)\big\}$$

certainly contains all rectangles $P \times Q$ with $P \in [\mathfrak{B}(F), \xi]$ and $Q \in [\mathfrak{B}(H), \zeta]$ and, because the complement of an invariant set is invariant again, it is closed under complementation. Also, $\mathcal{D}$ is closed under disjoint countable unions. Since the set of rectangles with invariant sides is closed under intersection, Dynkin's celebrated π-λ-Theorem together with part 1 tells us that $\mathcal{D} = [\mathfrak{B}(F \times H), \xi \times \zeta]$. □

This result is not only of structural importance, as we will see in a moment. It will also permit us to use, e.g., $\langle [x]_\xi, [y]_\zeta\rangle$ and $[\langle x, y\rangle]_{\xi\times\zeta}$ interchangeably, similarly with maps. This will simplify notation somewhat and thus make life a bit easier.

From now on all automata are working over analytic spaces.

A decent morphism generates a congruence via its kernel [3, 13]. The following counterpart to Proposition 3.7 shows that this is also the case for stochastic automata.

Proposition 4.2. *Given the stochastic automata* $\mathbf{K}$ *and* $\mathbf{K}'$ *with* $(f, g, h) : \mathbf{K} \to \mathbf{K}'$ *an automata morphism. Then* $(\ker(f), \ker(g), \ker(h))$ *is a congruence for* $\mathbf{K}$*, provided* g *and* h *are final.*

Proof. 1. Write $\mathbf{K} = \big(X, Y, Z, K\big)$ and $\mathbf{K}' = \big(X', Y', Z', K'\big)$. We establish first that we find for $V \in [\mathfrak{B}(Z), \ker(h)]$ a Borel set $V_0 \in \mathfrak{B}(Z')$ such that $V = h^{-1}[V_0]$, and for $W \in [\mathfrak{B}(Y), \ker(g)]$ another Borel set $W_0 \in \mathfrak{B}(Y')$ with $W = g^{-1}[W_0]$. This is done exactly as in Example 3.5 using finality of the respective maps.

2. Assume $f(x) = f(x')$ and $h(z) = h(z')$, and take $G \in [\mathfrak{B}(Z\times Y), \ker(h)\times \ker(g)]$. We want to show that $K(x, z)(G) = K(x', z')(G)$ holds. Assume first that $G = V \times W$ with $V \in [\mathfrak{B}(Z), \ker(h)]$ and $W \in [\mathfrak{B}(Y), \ker(g)]$ and

determine V_0, W_0 as above, so that $G = (h \times g)^{-1}[V_0 \times W_0]$. But now

$$K(x,z)(G) = K(x,z)\big((h \times g)^{-1}[V_0 \times W_0]\big) = K'(f(x), h(z))\big(V_0 \times W_0\big)$$
$$= K(x', z')(G).$$

This argument shows that

$$\mathcal{D} := \big\{G \in [\mathfrak{B}(Z \times Y), \ker(h) \times \ker(g)] \mid K(x,z)(G) = K(x',z')(G)\big\}$$

contains all rectangles $V \times W$ with $V \in [\mathfrak{B}(Z), \ker(h)]$ and $W \in [\mathfrak{B}(Y), \ker(g)]$. The set of these rectangles is closed under finite intersections, and $\mathcal{D}$ is closed under complementation as well as under countable disjoint unions. By Dynkin's π-λ-Theorem, $\mathcal{D}$ equals $[\mathfrak{B}(Z), \ker(h)] \otimes [\mathfrak{B}(Y), \ker(g)]$, which in turn is equal to $[\mathfrak{B}(Z \times Y), \ker(h) \times \ker(g)]$ by the first part of Lemma 4.1.

3. We have shown that $\ker(f) \times \ker(h) \subseteq \ker\big(\mathcal{G}\left(m_{\ker(h)\times\ker(g)}\right) \circ K\big)$, which establishes the claim by Lemma 3.3. ☐

Recall that a map $f : F \to H$ has an em-factorization $f = f_\bullet \circ \eta_{\ker(f)}$. If f is measurable, so are the components (but this does not entail the em-factorization living in the category of measurable spaces!). We obtain a similar decomposition for stochastic automata: Let $\mathfrak{f} = (f, g, h) : \mathbf{K} \to \mathbf{K}'$ be a morphism, and put $\eta_{\ker(\mathfrak{f})} := (\eta_{\ker(f)}, \eta_{\ker(g)}, \eta_{\ker(h)})$ and $\mathfrak{f}_\bullet := (f_\bullet, g_\bullet, h_\bullet)$ for conciseness.

This is an immediate consequence of Proposition 4.2:

Corollary 4.3. *In the notation of Proposition 4.2,* $\eta_{\ker(\mathfrak{f})} : \mathbf{K} \to \mathbf{K}_{\ker(\mathfrak{f})}$ *and* $\mathfrak{f}_\bullet : \mathbf{K}_{\ker(\mathfrak{f})} \to \mathbf{K}'$ *are morphisms, and* $\mathfrak{f} = \mathfrak{f}_\bullet \circ \eta_{\ker(\mathfrak{f})}$.

Proof. The first part follows from Proposition 4.2 together with Proposition 3.7. As for the second part, a somewhat lengthy but straightforward computation shows that $\big(\mathcal{G}\,(h_\bullet \times g_\bullet) \circ K_{\ker(\mathfrak{f})}\big)([x]_{\ker(f)}, [z]_{\ker(h)})(E')$ equals $\big(K' \circ (f_\bullet \times h_\bullet)\big)([x]_{\ker(f)}, [z]_{\ker(h)}\,(E')$ whenever $E' \in \mathfrak{B}(Z' \times Y')$. The asserted equality is obvious. ☐

5. Sequential Work

A stochastic automaton works sequentially and synchronously: input is fed into it, in each step an output is produced, then a new input is given, a new output is produced, etc. Of course, state changes occur as part of these operations. For the finite case, this was already anticipated in Example 2.2, see Eq. (2).

Formally, suppose the automaton $\mathbf{K} = (X, Y, Z, K)$ is in state z and receives first x_1, then x_2 as the input. Quite apart from the salient state changes, an output of length two is produced, and the probability $K(x_1x_2, z)(E)$ for the measurable set $E \subseteq Z \times Y \times Y$ is computed as

$$\int_{Z\times Y} K(x_2, z')(\{\langle z'', y_2\rangle \mid \langle z'', y_1y_2\rangle \in E\})\ dK(x_1, z)(\langle z', y_1\rangle). \tag{6}$$

After input x_1 in state z the automaton makes a transition to state z' and gives an output y_1 with probability $dK(x_1, z)(\langle z', y_1\rangle)$. The new input x_2 is met in state z' and produces a new state z'' as well as an output y_2 so that $\langle z'', y_1y_2\rangle \in E$ with probability $K(x_2, z')(\{\langle z'', y_2\rangle \mid \langle z'', y_1y_2\rangle \in E\})$. We have to average over z' and y_1. Standard arguments [6] show that we have extended the transition law to a stochastic relation $K : X^2 \times Z \Rightarrow Z \times Y^2$ (we could use indices showing the length of the automaton's work so far, but there is already enough notation around).

Let $v \in X^n$ be an input word of length n, and assume that we have extended the transition law already to a stochastic relation $K : X^n \times Z \Rightarrow Z \times Y^n$, all products carrying the corresponding product σ-algebras. Define for input $x \in X$, and state z the probability for the Borel set $E \subseteq Z \times Y^{n+1}$ as

$$K(vx, z)(E) := \int_{Z\times Y} K(x, z')(\{\langle z'', y\rangle \mid \langle z'', wy\rangle \in E\})\ dK(v, z)(\langle z', w\rangle). \tag{7}$$

Then it is shown in [6] that $K : X^{n+1} \times Z \Rightarrow Z \times Y^{n+1}$ is a stochastic relation. In this way we extend the probabilistic transition law to finite input sequences in a natural manner. It is readily seen that Eq. (2) is the discrete version of Eq. (7).

Now assume that $\mathfrak{c} = (\alpha, \beta, \gamma)$ is a congruence for $\mathbf{K}$. We will show now that friendship is not lost during the automata's sequential work as outlined above. Define for the equivalence relation α on X and for $n \in \mathbb{N}$ the extension α^n of α to X^n in the obvious manner

$$\langle x_1, \ldots, x_n\rangle\ \alpha^n\ \langle x'_1, \ldots, x'_n\rangle \Leftrightarrow x_i\ \alpha\ x'_i \text{ for } i = 1, \ldots, n,$$

similarly for the other equivalence relations, and for, e.g., α^∞ when dealing with infinite sequences. We claim that $\alpha^n \times \gamma$ is friendly to $\gamma \times \beta^n$ for each $n \in \mathbb{N}$, so that congruence $\mathfrak{c}$ induces an infinite sequence of friendships. This will be demonstrated for $n = 2$ now, the general case is shown exactly in the same way using induction and Eq. (7).

We do these steps:

Step 1: The set $\{\langle z, y'\rangle \mid \langle z', y, y'\rangle \in E\}$ is a member of $[\mathfrak{B}(Z) \otimes \mathfrak{B}(Y), \gamma \times \beta]$ for each $E \in [\mathfrak{B}(Z) \otimes \mathfrak{B}(Y^2), \gamma \times \beta^2]$ and for each $y \in Y$. It is easy to see that the set in question is a Borel set. Because E is $\gamma \times \beta^2$ invariant, and β is a reflexive relation, the set is also $\gamma \times \beta$-invariant.

Step 2: Let $E \in [\mathfrak{B}(Z) \otimes \mathfrak{B}(Y^2), \gamma \times \beta^2]$ and fix $\bar{x} \in X, \bar{y} \in Y$, then the map

$$\langle z, y\rangle \mapsto K(\bar{x}, z)\big(\{\langle z'', \bar{y}\rangle \mid \langle z'', y, \bar{y}\rangle \in E\}\big)$$

is $[\mathfrak{B}(Z) \otimes \mathfrak{B}(Y), \gamma \times \beta]$-measurable.

Assume first that E is a measurable rectangle, say, $E = C_1 \times B_1 \times B_2$, then

$$K(\bar{x}, z)\big(\{\langle z'', \bar{y}\rangle \mid \langle z'', y, \bar{y}\rangle \in E\}\big) = K(\bar{x}, z)(C_1 \times B_2) \cdot I_{B_1}(x)$$

with I_{B_1} the indicator function of the set B_1. This constitutes certainly a $[\mathfrak{B}(Z) \otimes \mathfrak{B}(Y), \gamma \times \beta]$-measurable function by Lemma 3.3. Applying next the principle of good sets, Dynkin's π-λ-Theorem shows that the set of all E for which the claim is true is all of $[\mathfrak{B}(Z) \otimes \mathfrak{B}(Y^2), \gamma \times \beta^2]$, because the latter σ-algebra is generated by these rectangles, and because of the first part of Lemma 4.1.

Now we are poised to show that $\alpha^2 \times \gamma$ and $\gamma \times \beta^2$ are friends. For this, take $E \in [\mathfrak{B}(Z) \otimes \mathfrak{B}(Y^2), \gamma \times \beta^2]$ and assume that $\langle x_1 x_2, z\rangle \ \alpha^2 \times \gamma \ \langle \bar{x}_1 \bar{x}_2, \bar{z}\rangle$, then we have according to Eq. (6)

$$\begin{aligned}
&K(x_1x_2, z)(E)\\
&= \int_{Z\times Y} K(x_2, z')\big(\{\langle z'', y_2\rangle \mid \langle z'', y_1y_2\rangle \in E\}\big)\ dK(x_1, z)(\langle z', y_1\rangle)\\
&= \int_{Z\times Y} K(x_2, z')\big(\{\dots\}\big)\ dK(\bar{x}_1, \bar{z})(\langle z', y_1\rangle)\\
&\quad \text{(Corollary 3.4, since the integrand}\\
&\qquad\qquad\qquad\qquad \text{is } [\mathfrak{B}(Z) \otimes \mathfrak{B}(Y), \gamma \times \beta]\text{-measurable)}\\
&= \int_{Z\times Y} K(x_2, z')\big(\{\dots\}\big)\ dK(\bar{x}_1, \bar{z})(\langle z', y_1\rangle)\\
&\quad \text{(by Step 1, because } \{\dots\} \text{ is in } [\mathfrak{B}(Z) \otimes \mathfrak{B}(Y), \gamma \times \beta])\\
&= K(\bar{x}_1\bar{x}_2, \bar{z})(E)\,.
\end{aligned}$$

Now the claim is established by Lemma 3.3.

Summarizing, we obtain

Proposition 5.1. *Let (α, β, γ) be a countably generated congruence for the stochastic automaton $\mathbf{K}$ over analytic spaces. Then $\alpha^n \times \gamma$ is friendly to $\gamma \times \beta^n$ for every $n \in \mathbb{N}$.*$\dashv$

In what follows, we will deal with finite or infinite sequences of inputs resp. outputs. Denote as usual for a set M by M^+ the set of all finite non-empty words with letters taken from M, $|v|$ denotes the length of word $v \in M^+$. M^∞ is the set of all infinite sequences, and $M^{\leq\infty} := M^+ \cup M^\infty$ are all non-empty finite or infinite sequences over M. For $\tau \in M^\infty$ the first n letters are denotes by τ_n. If M carries a σ-algebra $\mathcal{M}$, M^n carries for $n \leq \infty$ the n-fold product $\mathcal{M}^n$, and M^+ the coproduct $\mathcal{M}^+$ of $(\mathcal{M}^n)_{n\in\mathbb{N}}$, finally $M^{\leq\infty}$ has the coproduct of $\mathcal{M}^+$ and $\mathcal{M}^\infty$.

Having thus fixed notation, we turn to automata again. Viewed from the outside, a learning system, or a reactive one, receives an input and responds through an output, the internal states being hidden from the observer. They are usually assumed to follow some *initial distribution* $\mu \in \mathcal{G}(Z)$. So we put

$$K_\mu^{|v|}(v)(G) := \int_Z K(v,z)(Z \times G)\ d\mu(z) \tag{8}$$

with $v \in X^+, G \in \mathfrak{B}(Y^{|v|})$, thus $K_\mu^n(v)$ specifies the probability distribution of outputs of length n given input v with $|v| = n$, provided the initial states are distributed according to μ. If μ is concentrated on the point $z_0 \in Z$, we adopt z_0 as an *initial state* for the automaton. Note that the state changes after each input are recorded through K, but are kept hidden behind a kind of smoke screen (indicated by computing the probability $K(v,z)(Z \times G)$, hence not betraying which new state is specifically adopted). Finally, define

$$K_\mu^+(v)(G) := K_\mu^{|v|}(v)(G \cap Y^{|v|})$$

(with $G \in \mathfrak{B}(Y^+)$) as the *black box* associated with the stochastic automaton $\mathbf{K}$.

We note for later use

Lemma 5.2. *For every $\mu \in \mathcal{G}(Z)$, $K_\mu^+ : X^+ \Rightarrow Y^+$; given a countably generated congruence (α, β, γ) on $\mathbf{K}$, α^n is a friend to β^n with respect to K_μ^n for each $n \in \mathbb{N}$.*

Proof. It is shown first that $K_\mu^n : X^n \Rightarrow Y^n$ is a stochastic relation for each $\mu \in \mathcal{G}(Z)$ [10, Example 2.4.8, Exercise 4.14]. Since X^+ is the coproduct of the measurable spaces $(X^n)_{n\in\mathbb{N}}$, the first assertion follows. For the second one, fix $G \in [\mathfrak{B}(Y^n), \beta^n]$, and assume $v\ \alpha^n\ v'$. We observe $\langle v, z\rangle\ \alpha^n \times \gamma\ \langle v', z\rangle$ for all $z \in Z$, so in particular $K(v,z)(Z \times G) = K(v',z)(Z \times G)$, because $\alpha^n \times \gamma$ is friendly to $\gamma \times \beta^n$ by Proposition 5.1. Integrating with respect to $\mu \in \mathcal{G}(Z)$ yields $K_\mu^n(v)(G) = K_\mu^n(v')(G)$. So the second assertion follows from Lemma 3.3. $\square$

In fact, we may educate our black box to work on infinite sequences in such a way that the finite initial parts are respected. To be specific, we claim that we find a stochastic relation K_μ^∞ between X^∞ and Y^∞ such that for the cylinder set $G = G_n \times \prod_{m>n} Y$ with $G_n \in \mathfrak{B}(Y^n)$

$$K_\mu^\infty(\tau)(G) = K_\mu^n(\tau_n)(G_n)$$

holds. Consequently, we intend to find $K_\mu^\infty : X^\infty \Rightarrow Y^\infty$ with $K_\mu^n \circ \pi_n^X = \mathcal{G}\left(\pi_n^Y\right) \circ K_\mu^\infty$ for all n, with $\pi_n^\bullet : \tau \mapsto \tau_n$ as the projection of an infinite sequence to its first n letters. Thus we want to close the gap in this diagram

$$\begin{tikzcd} X^\infty \arrow[r, dashed] \arrow[d, "\pi_n^X"'] & \mathcal{G}\left(Y^\infty\right) \arrow[d, "\mathcal{G}\left(\pi_n^Y\right)"] \\ X^n \arrow[r, "K_\mu^n"'] & \mathcal{G}\left(Y^n\right) \end{tikzcd}$$

Evidently this requires the automaton to be fully probabilistic, i.e., that we have always $K(x,z)(Z \times Y) = 1$. For measure-theoretic reasons, we need also a topological assumption.

Proposition 5.3. *Let* $\mathbf{K} = (X, Y, Z, K)$ *be a stochastic automaton such that* X *and* Y *are Polish spaces, and* Z *is an analytic space. Then there exists for each initial distribution* $\mu \in \mathcal{G}(Z)$ *with* $\mu(Z) = 1$ *a uniquely determined stochastic relation* $K_\mu^\infty : X^\infty \Rightarrow Y^\infty$ *such that* $K_\mu^n \circ \pi_n = \mathcal{G}(\pi_n) \circ K_\mu^\infty$ *for all* $n \in \mathbb{N}$*, provided* $K(x,z)(Z \times Y) = 1$ *for all* $x \in X, z \in Z$.

Proof. Fix $\mu \in \mathcal{G}(Z)$ with $\mu(Z) = 1$. Define $\pi_{m,n} : y_1 \ldots y_m \mapsto y_1 \ldots y_n$ as the projection $Y^m \to Y^n$ for $m > n$, and put for $\tau \in X^\infty$

$$L_k(\tau)(G) := K_\mu^k(\tau_k)(G),$$

whenever $k \in \mathbb{N}$ and $G \in \mathfrak{B}(Y^k)$. Then $L_k : X^\infty \Rightarrow Y^k$ with $L_k(\tau)(Y^k) = 1$ for all τ, and

$$L_n(\tau) = \mathcal{G}(\pi_{m,n})\left(L_m(\tau)\right)$$

holds for $m > n$. Thus $\left(L_n(\tau)\right)_{n\in\mathbb{N}}$ is a projective system in the sense of [10, Definition 4.9.18] for every $\tau \in X^\infty$. The assertion now follows from [10, Corollary 4.9.21], a mild variant of the famous Kolmogorov Consistency Theorem. □

Because of its genesis, the stochastic relation K_μ^∞ might be called the *projective limit* associated with automaton $\mathbf{K}$ and distribution μ. It comes in usefully, e.g., when interpreting continuous time stochastic logics, see [7].

Example 5.4. Let (X, Y, Z, p) be a finite stochastic automaton such that $p(x,z)$ is always a probability on $Z \times Y$, and let μ be a distribution of initial states. Topologically speaking, finite sets with their power sets are compact metric spaces, thus they are Polish, hence analytic spaces. Proposition 5.3 states in this case that there exists a unique map $p^{\infty}_{\mu} : X^{\infty} \to \mathcal{G}(Y^{\infty})$ such that

$$\forall \kappa \in X^{\infty} : p^{*}_{\mu}(v)(w) = \sum \{p^{\infty}_{\mu}(v\kappa)(w\lambda) \mid \lambda \in Y^{\infty}\}.$$

Thus, fixing the distribution of the initial state, for any finite word v over X and any finite word over Y the probability that the response of the automaton to v is w can be obtained from projecting all those probabilities over infinite sequences the initial pieces of which are v resp. w.

Our black box works also for an infinite sequence of inputs, responding to it with a uniquely determined distribution on the set of output sequences. The price to pay for this is on one hand the full probabilistic nature of the underlying stochastic relation (given the requirement, this is only too obvious), and on the other hand the assumption of working in Polish rather in the considerably more general analytic spaces. This topological assumption, however, cannot be relaxed, as [2, Example 7.7.3] shows.

The distribution on infinite output sequences is consistent with its initial pieces. Suppose you stop the input sequence at point n, then you obtain the corresponding distribution on the outputs of length n. This observation permits us to decorate trees. Call a subset $\mathcal{T}$ of $X^{\leq\infty}$ a *tree* iff it is prefix free (so if $p \in \mathcal{T}$ and q is a prefix of p, then $q = p$). We interpret the elements of $\mathcal{T}$ and their prefixes as paths, and we associate to each path in the tree a probability: If $\vec{x} = x_1 \dots x_n$ is a path of length n, then there is some $p \in \mathcal{T}$ such that $\vec{x}$ is the prefix of length n to p. So assign to $\vec{x}$ the distribution

$$T(\vec{x}) := \mathcal{G}\left(\pi^{Y}_{n}\right)(K^{|p|}_{\mu})(p)$$

(with $|p| = \infty$, if p is infinitely long). This is well defined: if $\vec{x}$ is the prefix of $q \in \mathcal{T}$ as well, we have by construction $\mathcal{G}\left(\pi^{Y}_{n}\right)(K^{|p|}_{\mu})(p) = \mathcal{G}\left(\pi^{Y}_{n}\right)(K^{|q|}_{\mu})(q)$. In fact, being the prefix of more than one path may occur in case the tree branches out at some node later on.

The so constructed $T(v)(G)$ is the probability that the output is a member of $G \in \mathfrak{B}(G^{|v|})$ after input of the finite sequence v into the tree. If the finite path associated with v ends in the leaf x (so that $v = wx$ for some w, and v is not a prefix of another word in $\mathcal{T}$), then the probability that the final output is a member of $G_0 \in \mathfrak{B}(Y)$ is just $T(v)\big(Y^{|v|-1} \times G_0\big)$.

Friendship is maintained also for infinite sequences. We first show that the extension ξ^{∞} of a small equivalence relation ξ on the measurable space $(F, \mathcal{F})$ is small again. Assume that ξ is created by the countable set $\mathcal{U} := \{U_n \mid n \in \mathbb{N}\} \subseteq \mathcal{F}$, which we may assume to be closed under finite intersections (otherwise take $\{\bigcap_{i \in S} U_i \mid \emptyset \neq S \subseteq \mathbb{N} \text{ finite}\}$ as a countable creator). Put

$$\mathcal{D}_{\mathcal{U},\xi} := \left\{ \prod_{i=1}^{k} U_{n_i} \times \prod_{m>k} F \mid n_i \in \mathbb{N} \text{ for } 1 \leq i \leq k, k \in \mathbb{N} \right\},$$

then it is easy to see that this countable set creates ξ^∞. Note that $\mathcal{D}_{\mathcal{U},\xi}$ is also closed under finite intersections. Now let $\mathcal{V}$ be a countable creator for β. Assume that $\tau\ \alpha^\infty\ \tau'$, fix $\mu \in \mathcal{G}(Z)$ with $\mu(Z) = 1$ as before, and let $G \in \mathcal{D}_{\mathcal{V},\beta}$, then $G = \prod_{i=1}^k V_i \times \prod_{m>k} Y$ for some $V_1, \dots, V_k \in \mathcal{V}$. Hence $G_0 := \prod_{i=1}^k V_i \in [\mathfrak{B}(Y^k), \beta^k]$, and $\tau_k\ \alpha^k\ \tau'_k$, so that we have

$$K_\mu^\infty(\tau)(G) = K_\mu^k(\tau_k)(G_0) \stackrel{(\ddagger)}{=} K_\mu^k(\tau'_k)(G_0) = K_\mu^\infty(\tau')(G).$$

Equality ($\ddagger$) is implied by the friendship of α^k to β^k (Lemma 5.2). Thus $K_\mu(\tau)$ agrees with $K_\mu(\tau')$ on $\mathcal{D}_{\mathcal{V},\beta}$, so these measures agree on $[\mathfrak{B}(Y^\infty), \beta^\infty] = \sigma(\mathcal{D}_{\mathcal{V},\beta})$ by Dynkin's π-λ-Theorem (see [10, Lemma 1.6.31]).

We have shown

Proposition 5.5. *Let $\mathbf{K} = (X, Y, Z, K)$ be a stochastic automaton such that $K(x, z)(Z \times Y) = 1$ for all $x \in X, z \in Z$. Assume that X and Y are Polish spaces, Z is an analytic space and that $\mu \in \mathcal{G}(Z)$ with $\mu(Z) = 1$ is the initial distribution. If (α, β, γ) is a countably generated congruence on $\mathbf{K}$, then α^n is friendly to β^n with respect to K_μ^n for every $n \in \mathbb{N} \cup \{\infty\}$.* $\dashv$

So friendship turns out to be a surprisingly stable relationship, maintained even through finite and infinite streams.

6. Play it again, Sam

Assume that we want to factor a factored automaton. It will turn out that the resulting automaton may be obtained by factoring once the automaton from which we started, albeit with a modified congruence. This result will also enable us to do the reduction iteratively along the multiple components.

Given an equivalence relation ξ on a set F, and an equivalence relation ζ on the set F/ξ, define

$$x\ (\xi * \zeta)\ x' \text{ iff } [x]_\xi\ \zeta\ [x']_\xi\,,$$

hence x is related to x' through the new relation $\xi * \zeta$ iff the class $[x]_\xi$ of x is related to the class $[x']_\xi$ through relation ζ. We may think of $\xi * \zeta$ as a lifting operation (visually, a ζ-class may be seen as a sea in which ξ-classes swim; the $*$ operator lifts these classes to the level of the base space). It is clear that $\xi * \zeta$ is countably generated if both ξ and ζ are.

Define the bijections

$$\varphi_{\xi,\zeta} : \begin{cases} F/(\xi * \zeta) & \to (F/\xi)/\zeta \\ [x]_{\xi*\zeta} & \mapsto \Big[[x]_\xi\Big]_\zeta \end{cases} \quad \text{and} \quad \psi_{\xi,\zeta} : \begin{cases} (F/\xi)/\zeta & \to F/(\xi * \zeta) \\ \Big[[x]_\xi\Big]_\zeta & \mapsto [x]_{\xi*\zeta} \end{cases}$$

We obtain this diagram

$$\begin{tikzcd} F \arrow[r, "\eta_\xi"] \arrow[d, "\eta_{\xi*\zeta}"'] & F/\xi \arrow[d, "\eta_\zeta"] \\ F/(\xi*\zeta) \arrow[r, shift left, "\varphi_{\xi,\zeta}"] & (F/\xi)/\zeta \arrow[l, shift left, "\psi_{\xi,\zeta}"] \end{tikzcd}$$

with

$$\varphi_{\xi,\zeta} \circ \eta_{(\xi*\zeta)} = \eta_\zeta \circ \eta_\xi \text{ and } \eta_{\xi*\zeta} = \psi_{\xi,\zeta} \circ \eta_\zeta \circ \eta_\xi \tag{9}$$

Assume our stage is a measurable space, then we obtain

LEMMA 6.1. *Let* $(F, \mathcal{F})$ *be a measurable space, then the measurable spaces* $(F, \mathcal{F})/(\xi * \zeta)$ *and* $\big((F, \mathcal{F})/\xi\big)/\zeta$ *are isomorphic. Moreover* $\eta_\xi[G] \in [\mathcal{F}/\xi, \zeta]$, *provided* $G \in [\mathcal{F}, \xi * \zeta]$.

Proof. 1. We show that the bijections $\varphi_{\xi,\zeta}$ and $\psi_{\xi,\zeta}$ from above are measurable. From $\varphi_{\xi,\zeta} \circ \eta_{\xi*\zeta} = \eta_\zeta \circ \eta_\xi$ we see that $\varphi_{\xi,\zeta} \circ \eta_{\xi*\zeta}$ is measurable, and since $\eta_{\xi*\zeta}$ is final, we conclude measurability of $\varphi_{\xi,\zeta}$. Similarly, since the composition of final morphisms is final again, measurability of $\psi_{\xi,\zeta} \circ \eta_\zeta \circ \eta_\xi$ by (9) implies measurability of $\psi_{\xi,\zeta}$.

2. For establishing the second part, we have to show that $\eta_\xi[G]$ is ζ-invariant, since clearly $\eta_\xi[G] \in \mathcal{F}/\xi$ on account of $G \in \mathcal{F}$ being ξ-invariant. Given $t \in \eta_\xi[G]$ and t' with $t\ \zeta\ t'$, we find $x \in G'$ and some $x' \in F$ with $t = [x]_\xi$ and $t' = [x']_\xi$. Is $x' \in G$? Well, $t\ \zeta\ t'$ translates to $\Big[[x]_\xi\Big]_\zeta = \Big[[x']_\xi\Big]_\zeta$, equivalently $[x]_{\xi*\zeta} = [x']_{\xi*\zeta}$. Since $x \in G$, and because G is $\xi * \zeta$-invariant by assumption, we find that $x' \in G$ holds indeed, which means $t' \in \eta_\xi[G]$, so that the latter set is ζ-invariant. □

Thus we may and do identify ξ with $1_F * \xi$, and with $\xi * 1_{F/\xi}$.

Given a congruence on a factor automaton, each equivalence relation on a factored component space individually generates a new equivalence on the component proper through lifting, as we have seen above. These new equivalences are countably generated, if their components are. Combining all these lifted equivalences will yield a congruence, as will be shown now.

In a slight abuse of terminology, call a congruence *countably generated* (abbreviated *cg*) iff all its components are.

PROPOSITION 6.2. *Let* $\mathfrak{c} = (\alpha, \beta, \gamma)$ *be a cg congruence on the stochastic automaton* $\mathbf{K}$, *and* $\mathfrak{c}' = (\alpha', \beta', \gamma')$ *a cg congruence on the factor automaton* $\mathbf{K}_\mathfrak{c}$. *Then* $\mathfrak{c} * \mathfrak{c}' := (\alpha * \alpha', \beta * \beta', \gamma * \gamma')$ *is a cg congruence on* $\mathbf{K}$.

Proof. 1. Write $\mathbf{K} = \big(X, Y, Z, K\big)$. We know already that $\mathfrak{c} * \mathfrak{c}'$ is countably generated, so we have to show that $\alpha * \alpha' \times \gamma * \gamma'$ is friendly to $\gamma * \gamma' \times \beta * \beta'$ ($*$ binds stronger than $\times$). This is done through Lemma 3.3.

2. Let $\langle x, z\rangle$ $(\alpha * \alpha' \times \gamma * \gamma')$ $\langle x', z'\rangle$, we want to show $K(x, z)(G) = K(x', z')(G)$ for all $G \in [\mathfrak{B}(Z \times Y), \gamma * \gamma' \times \beta * \beta']$. Fix such a set G. Because $\alpha' \times \gamma'$ is friendly to $\gamma' \times \beta'$, we know that

$$K_{\mathfrak{c}}\big([x]_{\alpha}, [z]_{\gamma}\big)(H) = K_{\mathfrak{c}}\big([x']_{\alpha}, [z']_{\gamma}\big)(H) \tag{10}$$

for all $H \in [\mathfrak{B}(Z/\gamma) \otimes \mathfrak{B}(Z/\beta), \gamma' \times \beta']$. From the second part of Lemma 6.1 we see that $\eta_{\gamma\times\beta}[G] \in [\mathfrak{B}(Z/\gamma) \otimes \mathfrak{B}(Y/\beta), \gamma' \times \beta']$. Thus

$$\begin{aligned}
K(x, z)(G) &= K(x, z)\big(\eta^{-1}_{\gamma\times\beta}[\eta_{\gamma\times\beta}[G]]\big)\\
&= K_{\mathfrak{c}}\big([x]_{\alpha}, [z]_{\gamma}\big)\big(\eta_{\gamma\times\beta}[G]\big)\\
&\overset{(10)}{=} K_{\mathfrak{c}}\big([x']_{\alpha}, [z']_{\gamma}\big)\big(\eta_{\gamma\times\beta}[G]\big)\\
&= K(x', z')(G),
\end{aligned}$$

and we are done. $\square$

Factoring twice, each time with a countably generated congruence, has – up to isomorphism – the same effect as factoring once through a suitably constructed congruence. This observation is similar to the Third Isomorphism Theorem in Group Theory [17, Corollary 5.10], which tells us what happens when factoring a group iteratively through normal subgroups.

PROPOSITION 6.3. *Let $\mathbf{K}$ be a stochastic automaton with a countably generated congruence $\mathfrak{c}$, and $\mathfrak{c}'$ a countably generated congruence on the factor automaton $\mathbf{K}_{\mathfrak{c}}$. The factor automaton of $\mathbf{K}$ for the congruence $\mathfrak{c} * \mathfrak{c}'$, and the factor automaton of $\mathbf{K}_{\mathfrak{c}}$ for the congruence $\mathfrak{c}'$ are isomorphic.*

We could write $\big(\mathbf{K}/\mathfrak{c}\big)/\mathfrak{c}'$ more suggestively as $\mathbf{K}/(\mathfrak{c} * \mathfrak{c}')$.

Proof. 0. We assume that the automaton $\mathbf{K}$ is defined over the analytic spaces X, Y, and Z, and that the congruences are $\mathfrak{c} = (\alpha, \beta, \gamma)$ resp. $\mathfrak{c}' = (\alpha', \beta', \gamma')$. Denote by $\mathbf{K}_1$ the factor automaton of $\mathbf{K}$ for the congruence $\mathfrak{c} * \mathfrak{c}'$, and by $\mathbf{K}_2$ the factor automaton of $\mathbf{K}_{\mathfrak{c}}$ for the congruence $\mathfrak{c}'$. K is assumed to be the transition law for automaton $\mathbf{K}$, K_i the one for $\mathbf{K}_i$, $i = 1, 2$.

1. The candidates for the isomorphism are the suspects already indicated in the equations (9), specifically

$$\begin{aligned}
a^{\sharp} &: [x]_{\alpha*\alpha'} \mapsto [[x]_{\alpha}]_{\alpha'}, & \qquad & & a^{\flat} &: [[x]_{\alpha}]_{\alpha'} \mapsto [x]_{\alpha*\alpha'},\\
b^{\sharp} &: [y]_{\beta*\beta'} \mapsto \Big[[y]_{\beta}\Big]_{\beta'}, & & \text{and} & b^{\flat} &: \Big[[y]_{\beta}\Big]_{\beta'} \mapsto [y]_{\beta*\beta'},\\
c^{\sharp} &: [z]_{\gamma*\gamma'} \mapsto \Big[[x]_{\gamma}\Big]_{\gamma'}, & & & c^{\flat} &: \Big[[x]_{\gamma}\Big]_{\gamma'} \mapsto [z]_{\gamma*\gamma'}.
\end{aligned}$$

2. We show that this diagram commutes

$$\begin{tikzcd} X/(\alpha * \alpha') \times Z/(\gamma * \gamma') \arrow[r, "K_1"] \arrow[d, "a^\sharp \times c^\sharp"'] & \mathcal{G}\left(Z/(\gamma * \gamma') \times y/(\beta * \beta')\right) \arrow[d, "\mathcal{G}(c^\sharp \times b^\sharp)"] \\ (X/\alpha)/\alpha' \times (Z/\gamma)/\gamma' \arrow[r, "K_2"'] & \mathcal{G}\left((Z/\gamma)/\gamma' \times (y/\beta)/\beta'\right) \end{tikzcd}$$

For this, fix $J \in \mathfrak{B}((Z/\gamma)/\gamma' \times (y/\beta)/\beta')$. A not particularly exciting manipulation shows that

$$(\eta_{\gamma*\gamma'} \times \eta_{\beta*\beta'})^{-1}\left[(c^\sharp \times b^\sharp)^{-1}[J]\right] = (\eta_\gamma \times \eta_\beta)^{-1}\left[(\eta_{\gamma'} \times \eta_{\beta'})^{-1}[J]\right]. \tag{11}$$

But now we obtain

$$\begin{aligned} K(a^\sharp([x]_{\alpha*\alpha'}), c^\sharp([z]_{\gamma*\gamma'})(J) &= K_2(x,z)\big((\eta_\gamma \times \eta_\beta)^{-1}\left[(\eta_{\gamma'} \times \eta_{\beta'})^{-1}[J]\right]\big) \\ &\overset{(11)}{=} K(x,z)\big((\eta_{\gamma*\gamma'} \times \eta_{\beta*\beta'})^{-1}\left[(c^\sharp \times b^\sharp)^{-1}[J]\right]\big) \\ &= K_1([x]_{\alpha*\alpha'}, [z]_{\gamma*\gamma'})\big((c^\sharp \times b^\sharp)^{-1}[J]\big). \end{aligned}$$

So the diagram in question commutes indeed. A similar diagram for $(a^\flat, b^\flat, c^\flat)$ is shown to commute in exactly the same manner. Because all contributing maps are bijective and measurable by the remarks at the beginning of this section, we have found the desired isomorphisms. □

This result indicates that a stepwise reduction is possible. Suppose that we want to first reduce states according to γ, and then reduce inputs and outputs through α resp. β. We observe that up to isomorphism

$$(\alpha, \beta, \gamma) = (1_X, 1_Y, \gamma) * (\alpha, \beta, 1_{Z/\gamma})$$

holds. Reducing inputs and outputs first and then dealing with states gives rise to a similar isomorphism:

$$(\alpha, \beta, \gamma) = (\alpha, \beta, 1_Z) * (1_{X/\alpha}, 1_{Y/\beta}, \gamma).$$

7. Conclusion and Discussion

The notion of a congruence for stochastic automata is defined and investigated, the interplay of congruences with the kernels of morphisms is briefly shed light on. The central notion is the friendship of equivalence relations with respect to stochastic relations, which is studied extensively. We investigate also the behavior of an automaton when the input comes from a finite or infinite stream; this permits the automaton to work on trees with possibly infinite paths. Some topological assumptions had to be made in order to face measure theoretic problems adequately. Finally an isomorphism result is stated which permits the reduction of an automaton in a stepwise fashion.

A historic digression. The results here generalize among others well know constructions from discrete stochastic automata. This is akin to the situation when discussing bisimulations for Markov transition systems. There is a well established approach to constructing bisimulations for discrete Markov transition systems, see, e.g., the seminal paper [21] by Larsen and Skou. Panangaden's careful analysis [22, 23], however, pointed at some difficulties when trying to transplant this approach to non-discrete Markov transition systems, i.e., to probabilistic systems which work over general measurable spaces. Here one asks for methods based on coalgebras, and Azcel's Theorem [25] rapidly became the gateway, because it permitted expressing bisimilarity in terms of coalgebra morphisms. This opened the toolbox of categories including the results on the intriguing Giry monad, which captures probabilistic arguments through a categorical framework.

If the reader is not yet fully convinced that the generalization from finite stochastic systems to infinite ones does not entail a simple extension of well-proven established methods, it may be worthwhile having a look at stochastic dynamic programming [12, 16]. Optimization in a finite scenario usually applies techniques from convex optimization, which in turn are based on the geometrically appealing realm of convex polytopes in Euclidean, hence finite dimensional, spaces. Migrating to non-discrete measurable spaces, however, necessitates the employment of other tools, since finite dimensional solution spaces are no longer available. Here measurable selections [14, 15, 27] play a fundamental rôle as an important tool for investigating solutions. While convex polytopes live on compact convex sets in a finite dimensional space, measurable selections thrive on measurable relations with closed values in a Polish space. Clearly, the continuous case is no straighforward generalization of the discrete one.

What could be done next. An extension to these ideas lets equivalence relations act on subprobabilities by the well-known technique of randomization (see, e.g., [8, Chapter 3]). To be specific, let $(F, \mathcal{F})$ be a measurable space, ξ an equivalence relation on F with the σ-algebra $[\mathcal{F}, \xi]$ of ξ-equivalent sets. Define for $\mu, \nu \in \mathcal{G}(F, \mathcal{F})$

$$\mu \ \xi^{\diamond} \ \nu \text{ iff } \forall E \in [\mathcal{F}, \xi] : \mu(E) = \nu(E).$$

This is the *randomization* of ξ [8]; note that $x \ \xi \ x'$ iff $\delta_x \ \xi^{\diamond} \ \delta_{x'}$ with $\delta_x \in \mathcal{G}(F, \mathcal{F})$ the point mass on x. Furthermore, extend the stochastic relation $K : (F, \mathcal{F}) \Rightarrow (H, \mathcal{H})$ to a measurable map $K^* : \mathcal{G}(F, \mathcal{F}) \to \mathcal{G}(H, \mathcal{H})$ upon setting

$$K^*(\mu)(E) := \int_F K(x)(E)\mu(dx)$$

for $E \in \mathcal{H}$ (remember, stochastic relation K is really a Kleisli morphism for the Giry monad, K^* is its Kleisli extension). Call then the equivalence relation ξ a *random friend*[2] to the equivalence relation ζ iff we have $K^*(\mu)\ \zeta^\diamond\ K^*(\nu)$ provided $\mu\ \xi^\diamond\ \nu$ with $\mu, \nu \in \mathcal{G}\left(F, \mathcal{F}\right)$. An equivalent formulation without explicit randomization reads

$$\ker\left(\mathcal{G}\left(m_\xi\right)\right) \subseteq \left(K^* \times K^*\right)^{-1}\left[\ker\left(\mathcal{G}\left(m_\zeta\right)\right)\right],$$

where m is defined in Lemma 3.3. Transporting these ideas to automata, one would have to decide whether one wants friendship of the level of, say, $(\alpha \times \gamma)^\diamond$, or of $\alpha^\diamond \times \gamma^\diamond$; the latter one indicates a much tighter pairing than the former one (recall that a finite measure on a product space in not necessarily a product measure).

Acknowledgements

The author wants to thank Dr. Jan Bessai (TU Dortmund) for his curiosity regarding the behavior of an automaton on trees, and for his suggestions which improved the paper's readability.

The author, Professor Emeritus, held the Chair for Software Technology in the Department of Computer Science and was co-opted with the Department of Mathematics at the Technische Universität Dortmund.

References

[1] J. Asendorpf, *Reduktion stochastischer Automaten*, Master's thesis, Fachbereich Mathematik Christan Albrechts Universiät Kiel, 1973.

[2] V. I. Bogachev, *Measure theory*, Springer, 2007.

[3] S. Burris and H. P. Sankappanavar, *A course in universal algebra*, Springer, New York, 1981.

[4] M. N. Cakir, M. Saleemi, and K.-H. Zimmermann, *On the theory of stochastic automata*, arXiv:2103.14423v1, March 2021.

[5] V. Claus, *Stochastische Automaten*, B. G. Teubner Verlagsgesellschaft, Stuttgart, 1971.

[6] E.-E. Doberkat, *Stochastic automata – nondeterminism, stability, and prediction*, Lecture Notes in Comput. Sci., vol. 113, Springer, Berlin, 1981.

[7] E.-E. Doberkat, *The Hennessy-Milner equivalence for continuous-times stochastic logic with mu-operator* J. Appl. Logic **5** (2007), 519–544.

[8] E.-E. Doberkat, *Stochastic coalgebraic logic.* EATCS Monographs in Theoretical Computer Science. Springer, Berlin, 2009.

[9] E.-E. Doberkat, *Algebraic properties of stochastic effectivity functions*, J. Log. Algebr. Methods Progr. **83** (2014), 339–358.

[10] E.-E. Doberkat, *Special Topics in mathematics for computer scientists: Sets, categories, topologies and measures.* Springer, Cham, 2015.

[2]The present author does not know whether a random friend is a casual acquaintance, or a friend for life, or something in between.

[11] E.-E. Doberkat and Ch. Schubert, *Coalgebraic logic over general measurable spaces - a survey*, Math. Structures Comput. Sci. (2011), 175–234.
[12] L. E. Dubins and L. J. Savage, *How to Gamble if you Must: Inequalities for Stochastic Processes*, McGraw-Hill, New York, 1965.
[13] G. Grätzer, *Universal algebra*, The University Series in Higher Mathematics, Van Nostrand, Princeton, N. J., 1968.
[14] C. J. Himmelberg, *Measurable relations*, Fund. Math. **87** (1975), 53–72.
[15] C. J. Himmelberg and F. Van Vleck, *Some selection theorems for measurable functions*, Canad. J. Math. **21** (1969), 394–399.
[16] K. Hinderer, *Foundations of non-stationary dynamic programming with discrete time parameter*, Lect. Notes Op. Res. Math. Syst., vol. 33, Springer, Berlin, 1970.
[17] T. W. Hungerford, *Algebra*, Springer, Berlin, 1974.
[18] B. Jacobs, *Introduction to coalgebra: towards mathematics of states and observations*, Cambridge University Press, Cambridge, UK, 2017.
[19] T. J. Jech, *Set theory*, Springer, Berlin, 2006.
[20] F. Klix, *Information und Verhalten*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1972.
[21] K. G. Larsen and A. Skou, *Bisimulation through probabilistic testing*, Inform. and Comput. **94** (1991), 1–28.
[22] P. Panangaden, *Probabilistic relations*, School of Computer Science, McGill University, Montreal, 1998, pp. 59 – 74.
[23] P. Panangaden, *Labelled Markov processes*, World Scientific, 2009.
[24] H. Rasiowa and R. Sikorski, *The mathematics of metamathematics*, Monografie Matematyczne, vol. 41, PWN, Warschau, 1963.
[25] J. J. M. M. Rutten, *Universal coalgebra: a theory of systems*, Theoret. Comput. Sci. **249** (2000), 3–80.
[26] B. F. Skinner, *The Behavior of Organisms*, Appleton-Century-Crofts, New York, 1938.
[27] S. M. Srivastava, *A course on Borel sets*, Graduate Texts in Mathematics, Springer, Berlin, 1998.

Author's address:

Ernst-Erich Doberkat
Walther-von-der-Vogelweide-Str. 46
97422, Schweinfurt, Germany
E-mail: eed@doberkat.de

Received January 10, 2021
Revised November 13, 2021
Accepted November 15, 2021